

Comparing the Performance of Two Road Weather Models in the Netherlands

VIRVE KARSISTO

Finnish Meteorological Institute, and Department of Physics, University of Helsinki, Helsinki, Finland

SANDER TIJM

Royal Netherlands Meteorological Institute, De Bilt, Netherlands

PERTTI NURMI

Finnish Meteorological Institute, Helsinki, Finland

(Manuscript received 29 August 2016, in final form 7 February 2017)

ABSTRACT

High-quality road condition forecasts are a prerequisite for road authorities to ensure wintertime road safety. Harsh winter conditions can cause problems for traffic not only in countries where snowy winters are common but also in regions where the temperature drops below the freezing point occasionally. This study reports on the evaluation of the Royal Netherlands Meteorological Institute's (KNMI) new road weather forecasting model by comparing it with the Finnish Meteorological Institute's (FMI) road weather model, both run for 321 Dutch road weather stations, four times daily (0300, 0900, 1500, and 2100 UTC) during the test period, 15 January–28 February 2015. Road surface temperature forecasts by both models were evaluated against observations. The KNMI model produced slightly more accurate forecasts than the FMI model. The main reason for the difference is probably due to the optimization of the physical properties of the KNMI model for the Netherlands, whereas the FMI model is designed for quite different Finnish wintertime meteorological conditions. However, in general the road surface temperature forecasts were of quite comparable quality.

1. Introduction

High quality road weather forecasts are needed to optimize wintertime road maintenance operations and services. The plowing and salting of roads consumes resources and are costly operations. As one example, around 100 million euros are spent annually for winter road maintenance in Finland (Venäläinen and Kangas 2003). A comparable amount is spent in the Netherlands even with much less frequent tough wintry weather conditions. Neglecting timely maintenance operations would lead to slippery roads, increasing the number of accidents, which would become even more expensive for society. In addition to injuries, casualties, and damaged vehicles, traffic congestion can cause long delays in transportation. Winter tires are not commonly used in the Netherlands, causing trucks to get stuck in steep access and exit areas of highways and blocking them

under icy conditions. Salting and plowing can be planned well ahead and thus the costs can be minimized by making use of accurate road weather forecasts.

Many road weather models (hereafter RWMs) have been developed during the past 30 years (Rayer 1987; Jacobs and Raatz 1996; Chapman et al. 2001; Crevier and Delage 2001; Fujimoto et al. 2012; Yang et al. 2012; Kangas et al. 2015). The Finnish Meteorological Institute (FMI) initiated road surface temperature modeling activities in 1979 (Nysten 1980). The resulting model was in operational use during the early 1980s but was later discontinued. The model was also tested in the Netherlands within the European Cooperation in Science and Technology (COST) 30 bis project (David and Portal 1985). Data from road weather stations were collected, and an automatic system produced forecasts and warnings for a few hours in the future. The project also covered road/vehicle communications, automatic incident detection, and variable traffic signals.

The current operational RWM in FMI was developed in the late 1990s and has been operational since 2000

Corresponding author e-mail: Virve Karsisto, virve.karsisto@fmi.fi

DOI: 10.1175/WAF-D-16-0158.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](http://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

(Kangas et al. 2015). Several model improvements and developments have been made thereafter, including a pavement condition forecast application for pedestrians (Kangas et al. 2015) and a perfect prog-type statistical application to forecast road surface friction (Juga et al. 2013).

The Royal Netherlands Meteorological Institute (KNMI) RWM model was developed during 2014–15. This paper reports on the assessment of forecast quality of this brand new model by comparing its output with the FMI RWM, which has a long history of operational use. Both models' results were evaluated against road surface temperature observations. Model comparisons can be truly beneficial in finding out good properties as well as weaknesses of the models, which then leads to potential model improvements. There have been very few earlier comparative studies like this work despite the relatively high number of RWMs in use. Thornes and Shao (1991) compared three ice prediction models developed in the United Kingdom: the ICEBREAK model (Shao and Lister 1996), the Met Office model (Rayer 1987) and Thornes' model (Thornes 1984). All of these three models used the same input data and were run for a single test site in 24-h cycles. The ICEBREAK model showed the best performance based on model bias, standard deviation, and root-mean-square error.

Having a separate RWM in addition to a general numerical weather prediction (NWP) model is important since the physical processes can be evaluated in more detail with separate models. An RWM can model conditions specific to the road surface, whereas NWP uses the generalized land-use types. In addition, the effect of traffic on the amount of water and snow can be taken into account in an RWM. Since RWMs are usually one-dimensional, they can also use observations made at certain road points in their initialization rather than using interpolated observations. The present study compares the road surface temperature forecasts made by the KNMI and the FMI road weather models. Both models used the same observations and NWP forecast data as input to make the results comparable. The aim of the study is to assess the performance of the new KNMI model using the FMI model as a reference. Therefore, other road weather models are not included in the present study, but comparison with other models could be an important research topic in the future. Section 2 defines the physical and technical properties of both the FMI and KNMI models. Section 3 introduces the observations and the forcing datasets. The results are reported and analyzed in section 4 using standard verification metrics, and section 5 concludes the paper with the final discussion. Finally, the appendix gives a

more detailed description of the physical equations used in the models.

2. Model descriptions

a. Initialization

Both the KNMI and FMI RWMs are one-dimensional heat balance models that require as their input forecasted parameters from a three-dimensional NWP system. The input includes the following parameters interpolated to the respective road points: air temperature, dewpoint temperature/relative humidity, wind speed, incoming long- and shortwave radiation, and precipitation. The models also make use of observations from road weather stations (RWSs) when defining the models' initial temperature profiles. However, the two models adapt different procedures in determining the initial state of the forecast. The FMI model is run for 2 days prior to the latest measurements using observations from road weather stations as the forcing. The temperature of the first surface layer is set to be the observed road surface temperature at each time step, and the temperature profile evolves according to a heat transfer equation. The model then includes a 3-h period during which the forecasted radiation is adjusted to the observed road surface temperature. This method is called *coupling* and is explained in more detail by Crevier and Delage (2001) and Karsisto et al. (2016). The coupling phase starts at the time when the input forecast from the NWP model is initiated. The model calculates the temperature profile during this phase based on the heat balance equation using the NWP forecast as the forcing. The method determines a correction coefficient iteratively either for forecasted longwave (LW) or shortwave (SW) radiation so that at the end of the period the forecasted road surface temperature fits the observed road surface temperature. This correction coefficient is consequently used during the actual forecast phase. It approaches exponentially unity (1.0) as the forecast evolves and, after 6 h, the correction is typically about 20% of its original value. The correction coefficient is given for the radiation variable that has a higher intensity at the end of the 3-h period.

The KNMI model applies a quite different initialization procedure. The model run starts 1 h before the beginning of the actual RWM forecast, and the initial temperature profile is taken from the previous forecast rather than running the model with observations. However, the profile is adjusted according to the temperature difference between the observed surface temperature and the modeled surface temperature from the

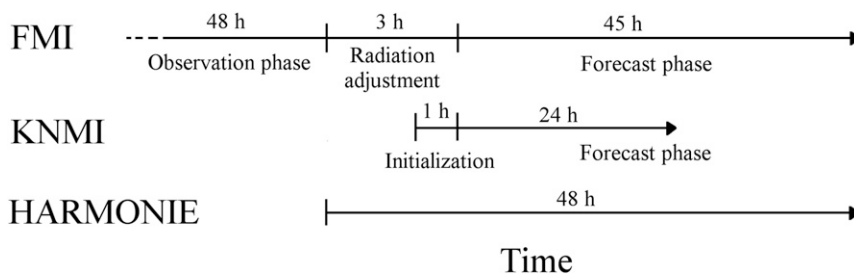


FIG. 1. The phases of FMI and KNMI model runs and their relation to the driving HARMONIE forecast.

previous forecast. The adjustment is 100% at the top layer and decreases linearly to 0% at the bottom layer. Then, the model also has a period that is used to correct the forecasted radiation according to the forecasted road surface temperature, but the length of the period is 1 h and the adjustment is not done iteratively. Instead, it is based on general calculations of how much energy is needed to change the surface temperature. The model is run for 1 h using the latest available forecast from an NWP model as the forcing, and the forecasted surface temperature is compared with the observed temperature at the end of the period. If the difference is more than 0.05 K, the model calculates coefficients either for LW or SW radiation based on the general calculations, so that the change compensates for the temperature difference. The coefficient is given for SW radiation if its intensity is larger than 100 W m^{-2} in the 1-h period and otherwise for longwave radiation. In the actual forecast phase the chosen radiation parameter is corrected using this coefficient. However, the correction used is only 50% of the original coefficient, because the adjustment of the 100% correction did not yield results that were as good as in the sensitivity tests. The correction coefficient remains the same for 3 h and after that is scaled linearly back to 1.0 in 9 h. The temperature profile is modified again before the start of the forecast phase using an observed surface temperature similar to that at the start of the model runs.

Due to the different initialization methods, the model runs were organized in such a way that the actual forecast phase started at the same time with both the KNMI and FMI models (Fig. 1). The necessary input forecasts were taken from the HIRLAM-ALADIN Research on Mesoscale Operational NWP in Euromed (HARMONIE; Bengtsson et al. 2017) model run by KNMI. Further details about this version of HARMONIE are given in section 3a. The FMI model requires 3 h during which observations and forecasts are available simultaneously. The starting time of the actual forecasts is therefore always 3 h after the HARMONIE starting time. To get the same starting time for the KNMI model, the model run

must start 2 h after the HARMONIE forecast run, because the initialization period of the KNMI model is only 1 h. For example, if a HARMONIE forecast starts at 0000 UTC, the actual forecast phase in both models begins at 0300 UTC. The forecast length was 45 h for the FMI model and 24 h for the KNMI model.

b. Physical properties

The physical properties of the surface and the road are quite different for the KNMI and FMI models (Table 1). The ground is divided into separate layers in both models, and the heat transfer is calculated between each layer at each time step. The KNMI model has 20 layers and the FMI model 16. The first two layers in the FMI model are considered to be asphalt and the rest have soil properties, whereas all of the layers are considered to be asphalt in the KNMI model. The layers of the KNMI model are much thinner than the FMI model layers, and the thickness of the road surface layer is only 0.3 cm, when it is 1.5 cm in the FMI model. The difference is highlighted in Fig. 2. The output surface temperature in the KNMI model is given as the temperature of the uppermost layer, whereas the FMI model uses the average of the top two layers as the output temperature. The depth of the lowest layer in the KNMI model is 0.33 m, but it is much deeper in the FMI model, with the middle point of the bottom layer being as deep as 4.28 m in the ground. The FMI model has a relatively long initialization period partly because it takes time to adjust the temperature of the lower layers. The density and specific heat of asphalt are larger in the KNMI model than in the FMI model, compensating for the thinner model layers. Also the asphalt heat conductivity is higher in the KNMI model. Moreover, the KNMI model has a separate mode for bridges, in which the lowest model-layer temperature is influenced by the air temperature. This mode was used when running the model for road weather stations on bridges.

The density, specific heat, albedo, absorption, and emissivity parameters of the KNMI model were determined before the start of the comparison study by

TABLE 1. The surface and ground properties of the FMI and KNMI models. The KNMI model has only an asphalt layer and thus the soil properties are not given in the table. In addition, the KNMI model does not use porosity in the model calculations.

	FMI (Kangas et al. 2015)	KNMI
No. of ground layers	16	20
Depth of the bottom layer (m)	4.28	0.33
Thickness of the first layer (cm)	1.5	0.3
Specific heat, asphalt ($\text{J kg}^{-1} \text{K}^{-1}$)	919	2000
Heat conductivity, asphalt ($\text{W K}^{-1} \text{m}^{-1}$)	0.5	1.25
Density, asphalt (kg m^{-3})	2110	3000
Porosity, asphalt	0.1	—
Specific heat, soil ($\text{J kg}^{-1} \text{K}^{-1}$)	813	—
Heat conductivity, soil ($\text{W K}^{-1} \text{m}^{-1}$)	1.4	—
Density, soil (kg m^{-3})	1600	—
Porosity, soil	0.4	—
Surface albedo	0.1, bare road; 0.6, snow; 0.1–0.6, ice	0.20
Emissivity	0.95	0.94
Absorption of long wave radiation	0.95	0.99
Roughness length for momentum (m)	0.4	0.001
Roughness length for heat (m)	0.001	0.001
Temperature reference height (m)	2	2
Wind reference height (m)	10	10

performing sensitivity tests for roughly 15 stations during the time period 1–28 February. Different parameter values were tested to find the best combination. Varying the parameters has a significant effect on the model bias values; for example, the negative road surface temperature bias for runs starting at 1800 UTC increases from -0.8° to -1.3°C at 0600 UTC (+12-h forecast) when the density is decreased from 3000 to 2000 kg m^{-3} . The 1200 UTC (+18-h forecast) positive bias increases from 1.3° to 2.0°C . These tests were done without initial surface temperature correction. As a result of the optimization of the physical properties, the KNMI parameters are heavily tuned toward values that correspond to observations. The high density and heat capacity values make the KNMI model slower to react to radiation and air temperature changes than it would be otherwise. This aims to correct for the effects of shading, which greatly affects the road surface temperature (Bogren et al. 2000). The corresponding parameters are defined in the FMI model in an attempt to produce reliable results for the Finnish roads. The KNMI model uses a sky-view factor of 0.9 that is the same at all locations, because not enough sky-view factor data were available at the time of this project. The option to use a station-specific sky-view factor is available in the KNMI model but not been used yet. The FMI model did not use the sky-view factor in this study. Some sensitivity tests were performed with the KNMI model to estimate the effect of the sky-view factor on the surface temperature. The model was run for the second half of February 2016 with sky-view factors varying from 0.4 to 1.0. Decreasing the sky-view factor to 0.1 caused the model bias to be

$0.2\text{--}0.3\text{ K}$ more positive in the forecasts that started at 1800 UTC after 10 h of nighttime running. It must be noted that the test was run at lower density and conductivity values for asphalt than in the operational model, which caused the model to be a bit more sensitive.

c. Output

The main output variable in both models is the road surface temperature. In the KNMI model it is the temperature at 1.5-mm depth inside the road, whereas it is the average temperature of the top two layers with thicknesses of 1.5 and 3.25 cm in the FMI model. The ground temperatures at depths of 3 and 20 cm are also produced by the KNMI model. The whole temperature profile of the FMI model was saved every hour for further analysis during this study. However, only the temperatures at depths of 3.0 and 6.5 cm are produced by the FMI model under normal operational forecast

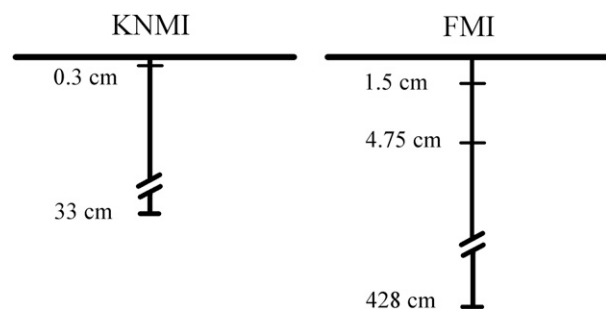


FIG. 2. The lower boundaries of the surface layers and the middle-point heights of the bottom layers in the KNMI and FMI models.

conditions. Both models also provide estimates of the amounts of water and ice on the road. The FMI model produces, in addition, separate values for snow and frost, whereas they are all considered as ice output in the KNMI model. The snow and ice also have some effect on the surface temperature values, causing them to remain near zero in the models during melting. There were some cases during the test period when snow was forecasted in the models and also at least one case when there was actually snow on the roads.

A road surface condition index is an additional output value of both models defining whether the road is dry, wet, icy, snowy, etc. Moreover, the sensible and latent heat fluxes and the net surface radiation were included as model output. The FMI model does not normally produce these variables but they were included in this study to enable more detailed comparison between models. All output parameters were produced every full hour. The FMI operational model version also calculates the surface friction (Juga et al. 2013) as well as indices for pedestrians and drivers depicting whether the conditions are normal, difficult, or very difficult (Kangas et al. 2015).

3. Data

a. NWP forecast

The high-resolution HARMONIE model is based on the AROME model developed by Météo-France and is described in more detail by Seity et al. (2011) and Bengtsson et al. (2017). HARMONIE has been in operational use at KNMI since summer 2012 (Baas and Van den Brink 2014), where the model domain extends roughly from 42° to 60°N and from 10°W to 17°E for an area of 2000 × 2000 km². HARMONIE version 36h1.4 with a resolution of 2.5 km was used in this study during 15 January–28 February 2015 to provide input forecasts for both RWMs. Within HARMONIE, three-dimensional variational data assimilation (3DVAR), in addition to blending of the large-scale High Resolution Limited Area Model (HIRLAM; Undén et al. 2002), is used to improve the initial conditions in the atmosphere. More information about data assimilation can be found in the work by Seity et al. (2011) and Bengtsson et al. (2017). HARMONIE uses a separate externalized surface model (SURFEX) library (Masson et al. 2013) to model surface processes. The SURFEX library uses four tile types (land, town, sea, and inland water) to describe the grid-box area and the physical parameterizations used are different for each type. Output values are calculated as weighted averages of the results for each tile according to their relative areas in the grid box. The parameterizations used are described in more detail by Masson et al. (2013), Seity et al. (2011), and Bengtsson et al. (2017).

The HARMONIE runs were initiated at 0000, 0600, 1200, and 1800 UTC each day. Because the initialization procedure of the FMI model takes 3 h, the RWM forecasts started at 0300, 0900, 1500, and 2100 UTC. The local time in the Netherlands is UTC + 1 h in winter, meaning that the starting hours correspond to 0400, 1000, 1600, and 2200 local time (LT), respectively. In total there were over 50 000 forecasts considered when all road weather stations were included; that is a large enough dataset to determine the behavior and the quality of the different models. The studied period contains multiple days with a large daily temperature cycle as the sun rises high enough in the sky to cause significant heating of the surface. Many days in February were very sunny, but there were also several cases during the test period with very cloudy conditions. The minimum temperatures in De Bilt, near the center of the Netherlands, were around −5°C and the maximum temperatures were around 10°C during the test period. In the center of the Netherlands there were more than 20 days with a minimum 2-m temperature below 0°C. In De Bilt there was one case of freezing rain turning into snow on 24 January and one case of light freezing rain on 7 February. This gives enough of a variety of conditions for the models to be tested, allowing for reactions to the daytime heating and nighttime cooling and the behavior for temperatures around 0°C.

b. Observations

Observation data were obtained from 321 road weather stations scattered across the country and maintained by the Dutch Rijkswaterstaat. Each station provides up to 12 road surface temperature sensors and 12 conductivity sensors at a single location. The sensors are typically located at slightly different places near the station (e.g., on different lanes). The surface temperature sensors are installed 2 mm below the surface, which is close to the middle point of the uppermost KNMI model layer at 1.5 mm. The stations also measure air temperature, dewpoint temperature, and, at some locations, soil temperature. The observation frequency is 5 min. Before producing the forecasts, the observations underwent an automatic quality control procedure to remove suspicious values. In total there were 298 stations where the RWMs could be run with proper initialization.

Because of several surface temperature observations being available at one location, there was the need to decide which sensor should be used as the RWM input. Road surface temperatures can vary significantly across the width of a road profile (Chapman and Thornes 2011). It is most relevant for the road maintenance authorities to get information on the overall lowest surface temperature at all locations to be able to determine

potential areas prone to freezing. The model can be adjusted to best predict the minimum temperature value at the station area when initialized with the lowest temperatures. Consequently, data from the sensor with the lowest temperature were used in the model initialization procedure, and this temperature was selected at each observation time. Therefore, the input data can include observations from several different sensors rather than being a full time series originating from one single sensor. The differences between sensors are usually less than 2 K during nighttime, but can be as large as 6 K at noon, highlighting the effect of the station's location on road surface temperatures.

4. Results

a. Comparison between HARMONIE and KNMI RWM

Before focusing only on the RWMs, the error statistics of the surface temperature forecasts made by the KNMI RWM and HARMONIE model were compared. Around 15 road weather stations were selected and the 0000 UTC forecast runs were analyzed for the period 1–28 February. The HARMONIE model has a negative surface temperature bias from around -0.5 to -1.0 K throughout the forecast, whereas the KNMI model bias is mostly positive except during the morning, when the most negative value is about -0.3 K. The KNMI model's positive bias varies from around 0.1 – 0.3 K during the nighttime to 0.4 – 1.0 K during the daytime. In the RWM the heat fluxes and ground properties are specified for the road, which explains the smaller bias values during the night. However, the asphalt seems to become too warm during the day in the model.

During the first forecast hours the root-mean-square error (RMSE) values of the KNMI RWM are more than 1 K better than those of the HARMONIE model. This considerable difference is expected because the RWM uses road surface temperature observations in the initialization. During the daytime the RMSE difference between models is around 0.0 – 0.3 K, but the difference increases again during the evening, and the KNMI RWM has considerably better RMSE values throughout the rest of the 24-h forecast. Overall, the HARMONIE model can predict the afternoon temperatures a little better than the KNMI RWM, but the KNMI model is considerably better during the rest of the day.

b. Bias

This and the following sections contain the verification results of the road surface temperature forecasts from the model runs starting at 0300 and 1500 UTC,

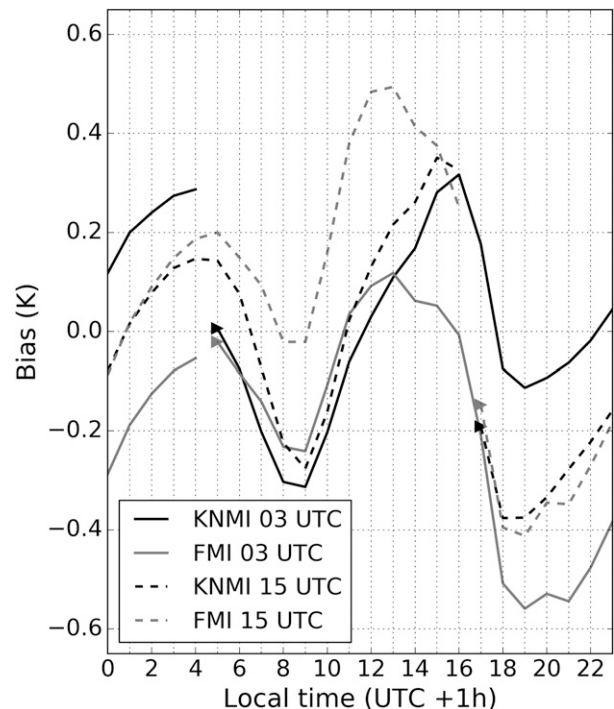


FIG. 3. Bias of road surface temperature as a function of local time for 0300 UTC (continuous line) and 1500 UTC (dashed line) forecast runs. Values are calculated for all stations and for the whole test period (15 Jan–28 Feb 2015). KNMI (FMI) results are shown by the black (gray) line. The triangles show the first forecast hour.

because they are the most relevant values for road maintenance (like salting the roads). Bias values were calculated separately for each forecast hour (Fig. 3). The forecasted values are very close to each other during the first 8–12 h from the start of the forecasts, with differences of less than 0.1 K. The differences become larger beyond 9 h when the biases show different signs.

The bootstrap method was used to determine whether the bias difference between the models is statistically significant (Efron and Tibshirani 1993; Hogan and Mason 2011). Some 10 000 bootstrap samples were generated with replacement for each lead time using the sample size of the original data. Bias values of both models' were calculated from each sample. Then, the 1st and 99th percentiles were determined from the distribution of the bias differences between the models. If zero is not included in the obtained range, the differences between the models are considered to be statistically significant (corresponding to a p value of 0.02). For the forecasts starting at 0300 UTC, the differences are statistically significant for all times shown in the Fig. 3, except for the 1300 LT forecast. However, for the 0600 LT forecast the significance level is just barely attained. For the 1500 UTC forecasts the results are statistically

significant for all times except for 2000, 0000, 0100, and 0200 LT.

The model biases have a daily cycle with the largest positive values during the day around 1300–1600 LT and during the night around 0500 LT, reaching their lowest negative values in the morning at 0900 LT and in the evening at 1900 LT. One reason for the high positive daytime biases is the shading effects, which are not taken into account in the models. This causes significant temperature overestimation at a number of stations during the time of the dominant shortwave radiation. Ignoring the sky-view factor may be one fundamental reason for the negative nighttime bias, because the longwave radiation from the surrounding objects is not taken into account in the model. The midday temperature bias maximum shows up about 3 h later in the KNMI results than in the FMI model results. The bias maximum tends to occur at the same time as the surface temperature maximum in the FMI model, but in the KNMI model it occurs after the temperature has started decreasing. This means that the KNMI model usually cools more slowly during afternoon than the FMI model. This must be considered to be a net effect of the differences in the model physical properties, and it is hard to find an individual reason causing the behavior. The considerably thinner surface layers in the KNMI model would presumably lead to faster cooling than in the FMI model, but the results show that other differences between the models are more dominant. First, the slower cooling in the KNMI model is supported by the larger heat capacity of the road material. Second, the sensible heat fluxes and latent heat fluxes tend to be more negative in the FMI model during the daytime as a result of the larger roughness length for momentum, and also the net radiation has smaller values, which supports faster cooling. Consequently, the slower reaction in the KNMI model must be considered to be net effect of the heat capacity, conductivity, layer thickness, and differences in the model fluxes. More information about the fluxes and their calculation in the models can be found in the [appendix](#).

As mentioned above, the daytime maximum temperatures tend to be too high in both models. In addition, the nighttime minimum temperatures tend to be too warm, except the FMI 0300 UTC forecasts, where the bias remains on the negative side. One major difference between the models is that the FMI model becomes much more negative in the early evening in the 0300 UTC run, and it remains negative throughout the night. Again there are multiple factors causing this difference between the models. One reason could be that in the KNMI model the heat stored to the ground during the daytime is transferred more efficiently to the surface

during nighttime as a result of the larger heat conductivity value of the asphalt. Also, the net radiation and latent heat flux are less negative in the KNMI model during the nighttime in general.

Contrarily, the KNMI model 1500 UTC run has a more negative road surface temperature bias at 0900 LT, whereas the FMI model bias is close to zero. The KNMI model seems to react more slowly to the increasing shortwave radiation during the early morning than does the FMI model. This must be considered again to be a net effect of the model differences. The behavior is partially caused by the larger heat capacity of the surface material in the KNMI model, although the results contradict the fact that the surface layers are thinner in the KNMI model. In addition, the sensible heat flux in the FMI model is larger during the early morning, which supports the faster warming. However, it needs to be highlighted that the values are averaged over all stations, and that there are huge variations between stations even within the relatively small area of the Netherlands. As an example, [Fig. 4](#) shows the biases of one forecast hour calculated separately for all stations. The values reveal that although the FMI model bias is about zero in [Fig. 3](#), there are in reality many stations with either positive or negative biases canceling each other. The KNMI model biases are mainly negative, but there are also stations with positive values.

A part of the bias in the RWMs is caused by the errors in the HARMONIE forecasts. Their effects could not be validated thoroughly because of the lack of the radiation and cloudiness observations at the road weather stations, but the effect of removing the 2-m temperature bias was tested by running the KNMI model with observed values. This reduced the nighttime bias by 50%, so the HARMONIE forecast errors clearly have a significant effect on the accuracy of the RWM forecasts.

c. RMSE

[Figure 5](#) shows the RMSE for the same road surface temperature forecasts as in [Fig. 3](#). The FMI model has on average a slightly larger RMSE than the KNMI model but the differences are mostly around 0.1 K. The differences were determined to be statistically significant with all lead times using the same bootstrap method described in the previous section. The RMSE values are greatest at midday, and the difference between the models grows to approximately 0.3 K in the 0300 UTC run and up to 0.5 K in the 1500 UTC model run. Daily maximum temperatures are usually difficult to predict, because they depend so much on the total radiation budget. In addition, the observational data used in the verification originates from a sensor giving the lowest temperature at each station. It is the sensor that,

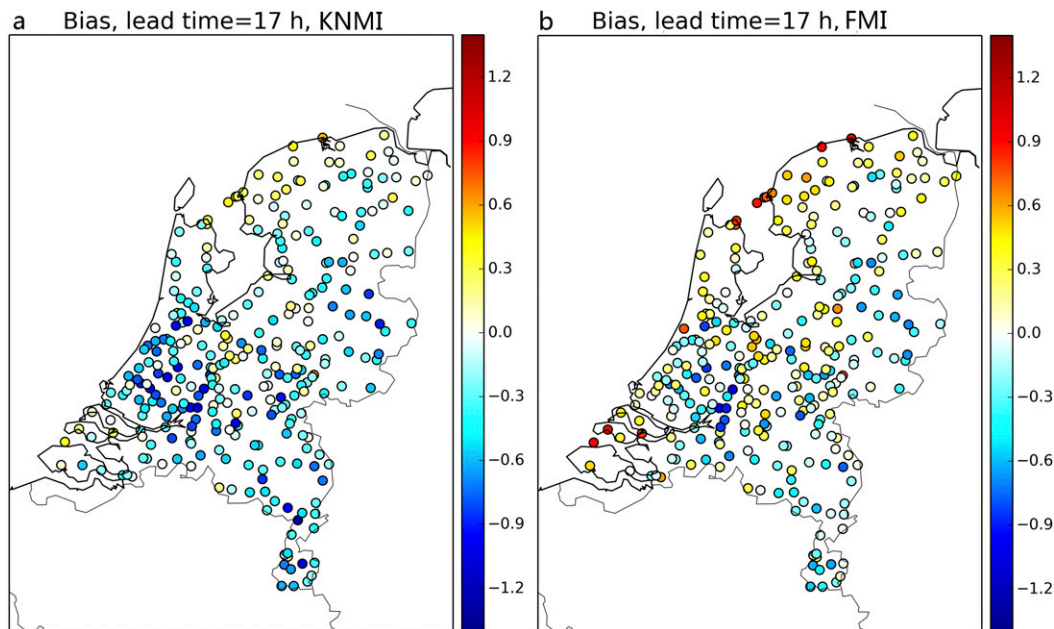


FIG. 4. Biases (K) of road surface temperature forecasts in the Netherlands for the (a) KNMI and (b) FMI models at 0900 LT. The forecasts were initiated at 1500 UTC and the lead time is 17 h, averaged over the whole time period from 15 Jan to 28 Feb 2015.

on average, has the largest influence of shading, which is not taken into account in the RWMs. This produces larger RMSE values at midday than would be obtained from observational data consisting of the maximum observations among the sensors. There are small reductions in the RMSEs at around 1000 and 1800 LT. The 0300 UTC model run produces lower RMSE values than the 1500 UTC model run for forecast lengths of a few hours. The reason can be the radiation adjustment. The radiation changes rapidly around 1500 UTC (1600 LT), so the radiation correction factor determined during the initialization does not fit that well during later hours in either model. During the early morning, at 0300 UTC (0400 LT) the radiation does not change that much and the correction is more appropriate. When studying other model runs, it was noted that forecasts initiated at 0900 UTC produced the largest RMSE values in the short range, the error being approximately 0.8 K in the first forecast hour. This is reasonable because it is hard to give an accurate radiation adjustment because of the unsteady radiation around 0900 UTC (1000 LT). Similarly to the bias values, there was much variation in the RMSE values between individual stations.

d. Categorical performance

One of the most important issues in road weather forecasting involves making accurate predictions around 0°C. To verify this, the hit and false alarm ratios were computed within five different temperature ranges

as follows: $T < 0^{\circ}\text{C}$, $-5.0^{\circ} < T < -1.0^{\circ}\text{C}$, $-1.0^{\circ} < T < 0.0^{\circ}\text{C}$, $0.0^{\circ} < T < 1.0^{\circ}$, and $1.0^{\circ} < T < 5.0^{\circ}\text{C}$, where T refers to the road surface temperature. The whole dataset was categorized utilizing the common contingency table shown in Table 2, followed by the computation of the probability of detection (POD) and the false alarm ratio (FAR) within these categories (WMO 2014):

$$\text{POD} = \frac{a}{a + c} \quad \text{and} \quad (1)$$

$$\text{FAR} = \frac{b}{a + b}. \quad (2)$$

The POD defines how frequently an event is correctly forecasted in relation to the number of cases when the event is observed. The FAR, on the other hand, indicates the number of false alarms in relation to the number of cases when the event is forecasted. The results are shown collectively in Fig. 6 in the form of a categorical performance diagram (Roebber 2009; Ebert et al. 2013). The total number of forecast cases was 13 000. The y axis shows the POD values and the x axis the FAR values in a reversed scale. A perfect forecast would fall in the top-right corner of the diagram, where $\text{POD} = 1$ and $\text{FAR} = 0$. The dotted lines represent the frequency bias $[(a + b)/(a + c)]$, which describes whether there was over- or underforecasting of the event in the given category. Values higher (lower) than 1 indicate overforecasting (underforecasting). Figure 6

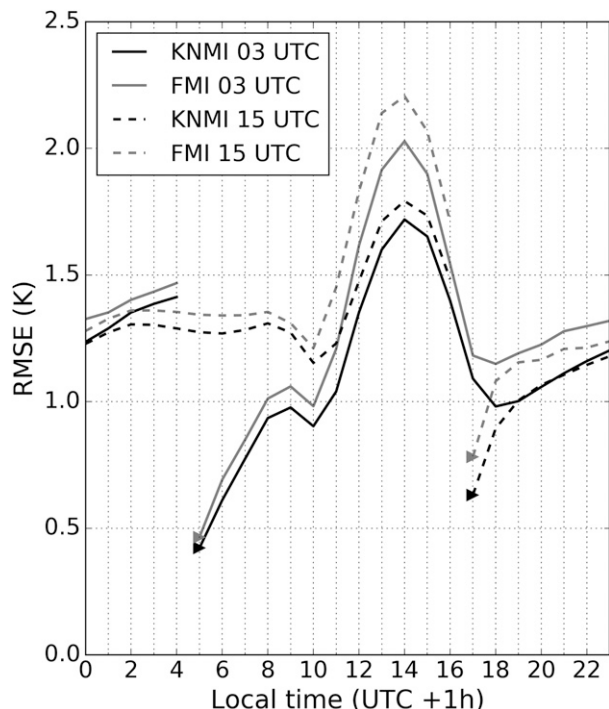


FIG. 5. RMSE (K) of road surface temperature forecasts as a function of local time for 0300 UTC (continuous line) and 1500 UTC (dashed line) forecast runs. KNMI (FMI) results are shown by the black (gray) line. The triangles show the first forecast hour.

further shows, with the continuous line, the so-called critical success index [CSI; $a/(a + b + c)$], which expresses the relation of hits to the total number of cases where the event was either observed or forecasted. In an ideal case, CSI would be equal to 1. The error bars in Fig. 6 represent the 95% confidence interval that is calculated using the error variance as described by Hogan and Mason (2011).

Figure 6 highlights that the scores for both of the models are quite similar. The same bootstrap method as described in section 4b was used to find out the statistical significance of the differences. The FMI model has typically slightly larger FAR values than the KNMI model. The differences were significant except for the forecasts started at 0300 UTC in the range $-1.0^{\circ} < T < 0.0^{\circ}\text{C}$. The differences in POD values were statistically significant only for temperature ranges below 0°C of the 0300 UTC model run and for the range $1.0^{\circ} < T < 5.0^{\circ}\text{C}$ of the 1500 UTC model run. In the 0300 UTC model runs the KNMI model has a somewhat higher POD for ranges $T < 0.0^{\circ}\text{C}$ and $-5.0^{\circ} < T < -1.0^{\circ}\text{C}$, but the POD of FMI is slightly better in the range $-1.0^{\circ} < T < 0.0^{\circ}\text{C}$. In the 1500 UTC model run the KNMI model has a little higher POD than the FMI model with range $1.0^{\circ} < T < 5.0^{\circ}\text{C}$. The scores for a larger hit range give better results than scores calculated for a range of 1°C , because the

TABLE 2. The contingency table.

		Event observed	
		Yes	No
Event forecast	Yes	<i>a</i>	<i>b</i>
	No	<i>c</i>	<i>d</i>

probability of a correct forecast is higher with a larger temperature range. Within ranges $-1.0^{\circ} < T < 0.0^{\circ}$ and $0.0^{\circ} < T < 1.0^{\circ}\text{C}$, both the POD and FAR results are around 0.5 for the 0300 UTC run, and the verification markers are even closer in the bottom-left corner for the 1500 UTC run, indicating lower forecast quality. Moreover, the 0300 UTC forecasts produce in general better results than the forecasts initiated at 1500 UTC. This is expected because the surface temperature usually varies more between 1500 and 1800 UTC in the Netherlands than between 0300 and 0600 UTC, and thus the values from the 0300 UTC model run do not differ that much from the observations used in the initialization and are easier to predict.

Some of the POD and FAR values are quite dependent on the time of day, as was the case with the bias and RMSE. This can be seen in Fig. 7, which represents the mean POD and FAR values as a function of local time. In the temperature range $1.0^{\circ} < T < 5.0^{\circ}\text{C}$, the smallest POD value and the largest FAR value are detected at 1400 LT. This is in agreement with the RMSE values, where the maxima were also found around midday as a result of the difficulties in predicting the daily maximum temperatures. This feature cannot be seen within temperature ranges $T < 0.0^{\circ}$ and $-5.0^{\circ} < T < -1.0^{\circ}\text{C}$, since there were so few observed and forecasted values at midday that the POD and FAR values could not be calculated. Instead, these temperature ranges are overpredicted in the evening, when there is a peak in the FAR values. This may also be seen in Fig. 6, where the 1500 UTC runs with both models produced relatively large FAR values within these categories. Both models have a cold bias in the evening, so the reason for this behavior is probably that the surfaces in the models cool too fast. During the nighttime the FAR is considerably smaller. The time dependency is not clear within temperature ranges $-1.0^{\circ} < T < 0.0^{\circ}\text{C}$ and $0.0^{\circ} < T < 1.0^{\circ}\text{C}$, in which both the FAR and POD values are worse compared to all other temperature categories.

e. Relative difference between models

The median differences of the surface temperature forecasts of the two models were finally analyzed to better understand the dissimilarities in their performance. Figure 8 shows the results. Overall, the differences are

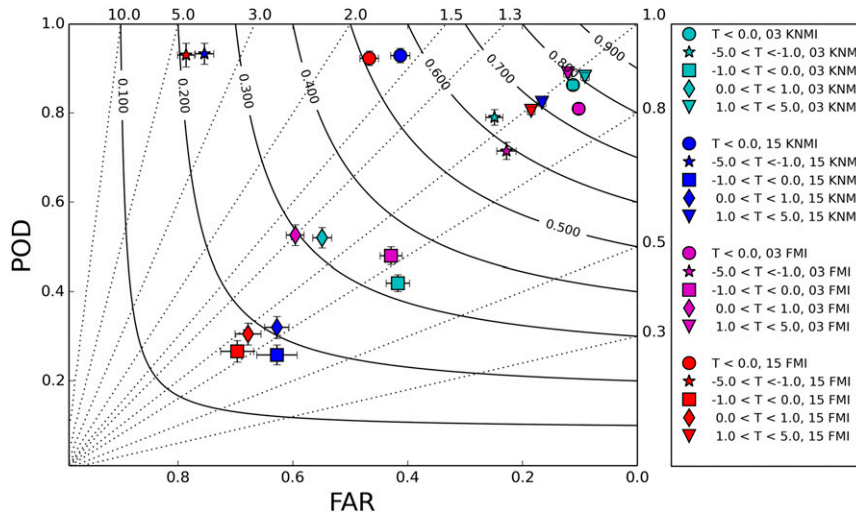


FIG. 6. Categorical performance diagram for forecasts initiated at 0300 and 1500 UTC for 3-h lead times. KNMI results are shown with cyan markers for 0300 UTC forecasts and with blue markers for 1500 UTC forecasts. FMI results for the same initialization times are shown with magenta and red markers, respectively. Circles represent results for $T < 0^{\circ}\text{C}$, stars for the range $-5.0^{\circ} < T < -1.0^{\circ}\text{C}$, squares for $-1.0^{\circ} < T < 0.0^{\circ}\text{C}$, diamonds for $0.0^{\circ} < T < 1.0^{\circ}\text{C}$, and triangles for $1.0^{\circ} < T < 5.0^{\circ}\text{C}$. FAR runs along the x axis and POD along the y axis. Dotted lines represent the frequency bias and continuous lines the CSI.

relatively small, and the median absolute difference is always less than 0.7 K. The median difference is close to zero during the first 8 h of the 0300, 1500, and 2100 UTC forecast runs. However, in the 0900 UTC run the KNMI model is relatively warmer than the FMI model at the beginning of the forecast. The radiation changes rapidly in the morning and, consequently, the different initialization methods generate larger differences between the models. Results also show that the FMI model is usually a bit warmer in the morning for the 0300, 1500, and 2100 UTC runs, and the difference is largest at 0900 UTC. The surface temperature in the KNMI model usually rises more slowly in the morning, which is seen also in the model bias results and is caused as a net effect of the many model differences, as discussed in the section 4b. As the day advances, the KNMI model becomes warmer, and the difference becomes largest in the evening around 2000 UTC. It was seen also in the bias results that the FMI model tends to be colder during the nighttime, and the reasons for this were also discussed in the section 4b.

The standard deviations of model differences were also calculated (Fig. 9). The results follow the same pattern as for the RMSE, being largest around 1300 UTC and dropping around 0800 and 1700 UTC. A comparison of Figs. 8 and 9 shows that the KNMI model is usually a little warmer than the FMI model at midday, but the discrepancy between forecasts is large. In other words, there are also many cases where the FMI model is warmer during daytime. In the morning, when the

KNMI model is generally colder, the standard deviation is smaller, so there are relatively fewer cases when the KNMI model is warmer in the morning. Correspondingly, it is not very common for the FMI model to be warmer than the KNMI model in the evening.

5. Discussion

The quality of the new (2015) KNMI road weather model was assessed by comparing it with the well-established road weather model of the FMI. The KNMI model generated somewhat smaller forecast errors across the Netherlands than the FMI model, confirming the applicability of its operational use for Dutch highways. The reason for the somewhat better performance is its optimization of the physical properties of local Dutch roads. The FMI model, on the other hand, has been designed by default for Finnish roads, whose physical properties are not considered totally suitable for the Netherlands. This study highlights the importance of the optimization of model physical properties when being implemented in new climatological and physiological environments. In the Netherlands the asphalt properties may vary across different areas; so, further studies are needed where physical properties are individually optimized for relevant road weather stations. Overall, the surface temperature forecasts of the models are quite similar, although the surface-layer thicknesses are very different in the two models. The

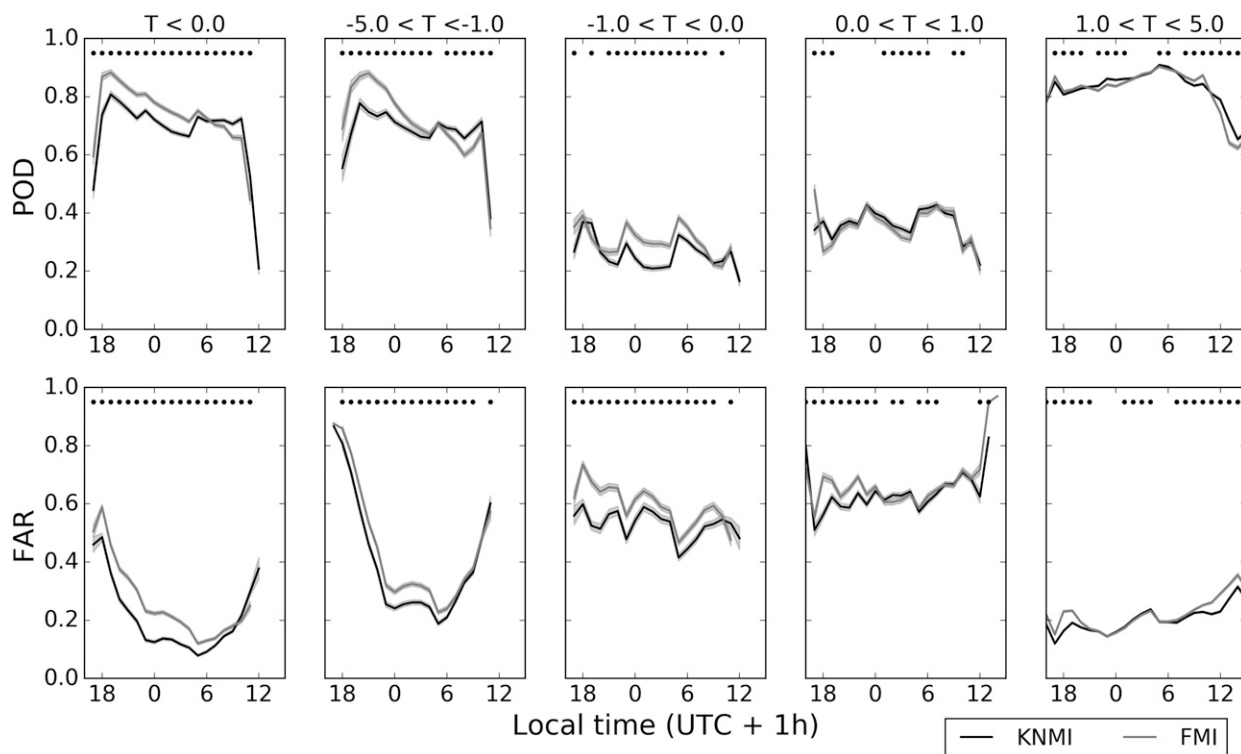


FIG. 7. Mean (top) POD and (bottom) FAR as a function of local time in the five different temperature range categories. The time values are from 1600 LT to 1500 LT from left to right. Averages are taken from the POD and FAR values calculated for 0300, 0900, 1500, and 2100 UTC model runs, i.e., each value corresponds to the same time of day regardless of the lead time. The means for the KNMI (FMI) model are shown by black (gray) lines. Not shown are cases where there were fewer than 100 hit values in the *a* category for POD and, correspondingly, fewer than 100 miss values in the *b* category for FAR. The 95% confidence intervals are shown by gray shading. The black dots show the cases where the differences between the models were statistically significant at the 98% confidence level.

net effect caused by the differences in the heat fluxes and physical parameters like asphalt heat capacity caused the KNMI model to react more slowly to the temperature changes during the morning and evening, despite the fact that the thickness of its surface layers was much thinner than those of the FMI model.

The use of the lowest surface temperature measurements at each station made the forecasting of daily maximum temperatures a challenge, since the possible shadowing effects at these locations can make the surface colder than forecasted. Shadowing was not taken into account except in the initialization process, because shadow factors have not been determined for the station locations. Doing this would have been too time consuming of a task in the present context. In the KNMI model the optimal values for each station are currently tested by running simulations with different heat conductivity and sky-view factor values. It is planned that in the future the sky-view and shading factors would be determined from a very high-resolution (25 cm) height map of the Netherlands. In the FMI the use of sky-view factors is currently tested in the in a small area of Norway as part of the Advanced Snow Plough and Salt

Spreader Based on Innovative Space Technologies—Winter Road Maintenance (ASSIST WRM) project, where they are determined from 100-m-resolution height maps. Plans include testing different heat capacity and conductivity values for Finnish road weather stations to find the best combinations.

To develop RWMs, comparing results from different models is highly beneficial. However, there are very few recent published road weather model comparison studies. [Thornes and Shao \(1991\)](#) stated in the beginning of the 1990s that commercial reasons prevented the comparison of other than the three models that were analyzed in their study. However, thanks to the development of communication networks and scientific collaboration, it is now easy to share large datasets between countries. It has become possible for collaborating institutes to run their models with commonly shared input data and without necessarily providing access to local model codes if that should be the case for preventing collaboration. Further comparison studies similar to what has been performed here but with more participants would be highly interesting. However, even with only two partners both parties benefited greatly

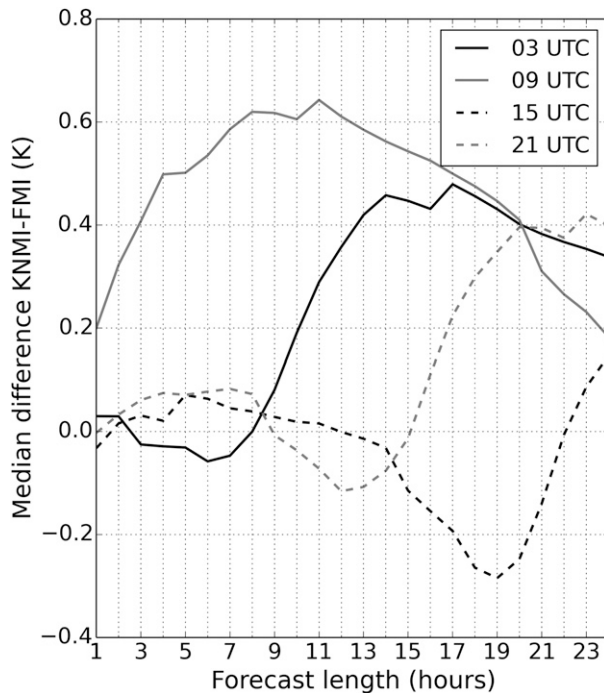


FIG. 8. Median difference of road surface temperature forecasts between KNMI and FMI models (KNMI – FMI) as a function of forecast length. Included are all stations where road surface temperature observations were available during the full test period (15 Jan–28 Feb 2015). Forecasts initiated at 0300 UTC are shown by the continuous black line, the 0900 UTC run by the continuous gray line, the 1500 UTC run by the dashed black line, and the 2100 UTC run by the dashed gray line.

from the collaboration, gaining valuable guidance and information for further development of their local weather models.

Acknowledgments. We thank Suomen Kulttuurirahasto (the Finnish Cultural Foundation) for financial support of road weather forecasting research.

APPENDIX

Model Physics

Tables A1 and A2 give a summary of the variables and physical equations used in the models. Heat flux into the ground is calculated as in Brutsaert (1984), except that the KNMI model takes into account the freezing and melting energies and the FMI model has its own parameter for traffic-caused heating. This parameter has values of 10 W m^{-2} during daytime traffic (0400–1900 UTC) and 5 W m^{-2} during nighttime traffic (1900–0400 UTC). The FMI model uses a simpler approach to take into account the energy needed to melt ice and snow compared

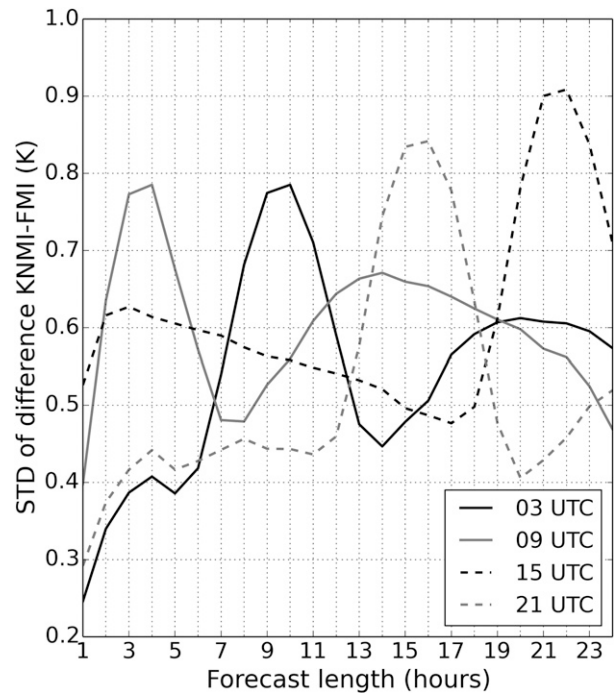


FIG. 9. As in Fig. 8, but for the standard deviation (STD).

with the KNMI model. In the FMI model the surface temperature remains at 0.25°C when melting occurs instead of taking it into account in the flux calculation. The remaining energy is used to warm up the surface after all the snow and ice has melted.

Net radiation is also calculated as in Brutsaert (1984) in both models, except the KNMI model takes into account the sky-view factor (0.9). The use of the sky-view factor reduces the amount of longwave radiation from the atmosphere but takes into account the longwave radiation emitted from the surroundings. In the initialization both models calculate a correction factor for either long- or shortwave radiation. This also has an effect on the net radiation, which is explained in more detail in section 2a. Figure A1 shows the average net radiation, sensible heat, and latent heat fluxes in the 0300 UTC model runs. Other model runs also identify similar behavior. In general, the KNMI model has more positive net radiation than the FMI model.

The boundary layer conductance and stability parameters are calculated using an iterative procedure in both models. Although the equations for these parameters are rather different, the results with the same input values produce boundary layer conductance values of similar quantity when tested with $T_s = 5.0^\circ\text{C}$, $T_a = 0.0^\circ\text{C}$, $z_m = 0.001 \text{ m}$, $z_t = 0.001 \text{ m}$, and varying the wind speed from 1 to 11.5 m s^{-1} . However, the FMI model uses a larger roughness length for momentum ($z_m = 0.4 \text{ m}$),

TABLE A1. Definitions of variables used in Table A2.

Variable	Definition
G	Heat flux into the ground
I_{net}	Net radiation at the surface
H	Sensible heat flux
LE	Latent heat flux
PC	Heat flow due to melting
PF	Heat flow due to freezing
Tr	Heating caused by traffic
α_s	Surface albedo
I_g	Downwelling shortwave radiation
ϵ_l	Absorption of longwave radiation
I_L	Downwelling longwave radiation
ϵ_s	Surface emittance
σ	Stefan–Boltzmann constant
T_s	Surface temperature
BLC	Boundary layer conductance
θ_s	Potential temperature at surface
θ_a	Potential temperature at 2 m
T_a	Air temperature
c_a	Specific heat of air
W_s	Water on the road
W_{max}	Maximum amount of water on the road
ρ_a	Density of air
k	von Kármán's constant
u^*	Friction velocity
z_T	Temperature measurement height
d	Zero-plane displacement
z_h	Roughness length for heat
Ψ_h	Stability correction factor for heat
Ψ_{h0}	Stability factor for heat at the height of z_h
z_W	Wind speed measurement height
z_m	Roughness length for momentum
Ψ_m	Stability factor for momentum
Ψ_{m0}	Stability factor for momentum at the height of z_m
ζ	Stability parameter
g	Gravitational constant
L_v	Specific heat of evaporation
q_a	Specific humidity for air at 2 m
q_s	Specific humidity for air at surface
ρ_m	Density of moist air
γ	Psychrometric constant
e_s	Water vapor pressure of the surface
e_a	Water vapor pressure of the air
r_o	Aerodynamic resistance
T	Ground temperature
z	Vertical distance in the ground
t	Time
κ_g	Thermal diffusivity
k_g	Heat conductivity
ρ_g	Ground density
c_g	Specific heat capacity of the ground
T_{clim}	Climatological temperature deep in the ground (10.0°C)
A	0.6
ω	$2\pi/365$
z_b	Depth of the bottom layer
z_d	2.7 m

which leads to much stronger coupling of the road to the atmosphere. Consequently, the absolute sensible heat flux values are larger in the FMI model than in the KNMI model in general. Another reason for this behavior is that the temperature of the uppermost layer in the FMI model rises much faster than the surface-layer temperature in the KNMI model, which causes a larger temperature difference between the surface and the air. The difference is not that great in the verification scores because the output surface temperature in the FMI model is given as average of the top two layers. This average temperature is also used when stability parameters and the boundary layer conductance are calculated.

The equations for latent heat flux are also quite different in the two models. In general, the absolute values of the latent heat flux are greater in the FMI model than in the KNMI model. The main reason for this is again the larger roughness length in the FMI model. With similar input values and wind speeds greater than 1 m s^{-1} , the equations give latent heat flux values that are closer to each other when tested with $T_s = 5.0^\circ\text{C}$, $T_a = 3.0^\circ\text{C}$, $z_m = 0.001 \text{ m}$, $z_t = 0.001 \text{ m}$, $\text{Rh} = 50\%$, $W_s = 0.1 \text{ mm}$, and varying the wind speed from 1.5 to 11.5 m s^{-1} . However, the FMI model equations tend to still give larger absolute values. The FMI model also allows thicker layers of water and ice on the surface; so, more energy is required for evaporation. In the FMI model, the maximum limit for water storage is 2 mm, for snow it is 100 mm, for ice it is 20 mm, and for frost it is 2 mm. The values are given in water equivalent millimeters. In the KNMI model the maximum storage values for water and ice are 0.2 mm. The value for the psychrometric constant γ in the FMI model has been developed using values from Calder (1990). The value for aerodynamic resistance r_o in the FMI model is determined by a modified version of the equation given by Tourula and Heikinheimo (1998). Restrictions for low wind speeds in the FMI model are used because the divider in the equation becomes small with low wind speeds and gives quite large values for aerodynamic resistance. With wind speeds of 1 m s^{-1} and with the other input values mentioned above, the FMI model equation gives a much larger absolute latent heat flux value than the KNMI model equation because of the usage of a constant value of 30 m^{-1} for the aerodynamic resistance.

Heat transfer in the ground is calculated in the same way in both models except the FMI models uses a different solving method for the differential equation in the initialization phase. In this phase the FMI model uses an algorithm obtained by solving the heat transfer equation by a time-centered Crank–Nicholson scheme, and the

TABLE A2. Summary of the main physical equations used in the KNMI and FMI models.

	KNMI	FMI
G	$I_{\text{net}} - H - LE + PC + PF$ (Brutsaert 1984)	$I_{\text{net}} - H - LE + \text{Tr}$ (Brutsaert 1984)
I_{net}	$(1 - \alpha_s)I_g + 0.9\epsilon_l I_L + 0.1\epsilon_s \sigma T_a^4 - \epsilon_s \sigma T_s^4$ (Brutsaert 1984)	$(1 - \alpha_s)I_g + \epsilon_l I_L - \epsilon_s \sigma T_s^4$ (Brutsaert 1984)
H	$\text{BLC}(\theta_s - \theta_a)$ (Garratt 1992)	$\text{BLC}(T_s - T_a)$ (Campbell 1985, 1986)
BLC	$\frac{c_a \rho_a k^2 u^*}{\left[\ln\left(\frac{z_W}{z_m}\right) - \Psi_m + \Psi_{m0} \right] \left[\ln\left(\frac{z_l}{z_h}\right) - \Psi_h + \Psi_{h0} \right]}$ (De Bruin 1994)	$\frac{c_a \rho_a k u^*}{\ln\left(\frac{z_T - d + z_h}{z_h}\right) + \Psi_h}$ (Campbell 1985, 1986)
u^*	$\frac{ku}{\ln\left(\frac{z_W}{z_m}\right) - \Psi_m + \Psi_{m0}}$ (Van Ulden and Holtslag 1985)	$\frac{ku}{\ln\left(\frac{z_W - d + z_m}{z_m}\right) + \Psi_m}$ (Campbell 1985, 1986)
Ψ_h, Ψ_m stable conditions	-5ζ (Van Ulden and Holtslag 1985)	4.7ζ (Campbell 1985, 1986)
Ψ_h, Ψ_m unstable conditions	$x = (1 - 16\zeta)^{1/4}$	$\Psi_h = -2 \ln\left(\frac{1 + \sqrt{1 - 16\zeta}}{2}\right)$
	$\Psi_h = 2 \ln\left(\frac{1 + x^2}{2}\right)$	$\Psi_m = 0.6\Psi_h$ (Campbell 1985, 1986)
	$\Psi_m = 2 \ln\left(\frac{1 + x}{2}\right) + \ln\left(\frac{1 + x^2}{2}\right) - 2\arctan(x) + \pi/2$ (Van Ulden and Holtslag 1985)	
ζ	$-\frac{kz_T g H}{c_a \rho_a T_a u^{*3}}$ (Campbell 1985, 1986)	$-\frac{kz_T g H}{c_a \rho_a T_a u^{*3}}$ (Campbell 1985, 1986)
LE	$\frac{-\rho_a k^2 L_a u (q_a - q_s)(W_s/W_{\text{max}})^{1/2}}{\left[\ln\left(\frac{z_W}{z_m}\right) - \Psi_m + \Psi_{m0} \right] \left[\ln\left(\frac{z_l}{z_h}\right) - \Psi_h + \Psi_{h0} \right]}$ (De Bruin 1994)	$\frac{\rho_m c_a \epsilon_s - \epsilon_a}{\gamma} \frac{r_o}{r_o}$ (Calder 1990)
γ		$0.1 \times (0.00063 T_a + 0.47496)$
$r_o, u > 1 \text{ m s}^{-1}$		$\frac{\left[\ln\left(\frac{z_W + z_m}{z_m}\right) + \Psi_m \right] \left[\ln\left(\frac{z_W + z_h}{z_h}\right) + \Psi_h \right]}{k^2 u}$ (Tourula and Heikinheimo 1998)
$r_o, u \leq 1 \text{ m s}^{-1}$		30 sm^{-1}
Heat transfer in the ground	$\frac{\delta T(z, t)}{\delta t} = \kappa_g \frac{\delta^2 T(z, t)}{\delta z^2}, \kappa_g = \frac{k_g}{\rho_g c_g}$ (Garratt 1992)	$\frac{\delta T(z, t)}{\delta t} = \kappa_g \frac{\delta^2 T(z, t)}{\delta z^2}, \kappa_g = \frac{k_g}{\rho_g c_g}$ (Garratt 1992)
Temp of the bottom layer	Free boundary condition for normal road	$T_{\text{chim}} + A \sin\left(\omega t - \frac{z_b}{z_d}\right)$ (Campbell 1985, 1986)
Numerical solving method for heat transfer	Exchange with 2-m temperature for bridges Explicit scheme, forward Euler method	Initialization phase before coupling: time-centered Crank-Nicholson scheme used to solve equation numerically; Thomas algorithm (Campbell 1985) used to solve the obtained tridiagonal matrix system Coupling phase and forecast phase: explicit scheme, forward Euler method

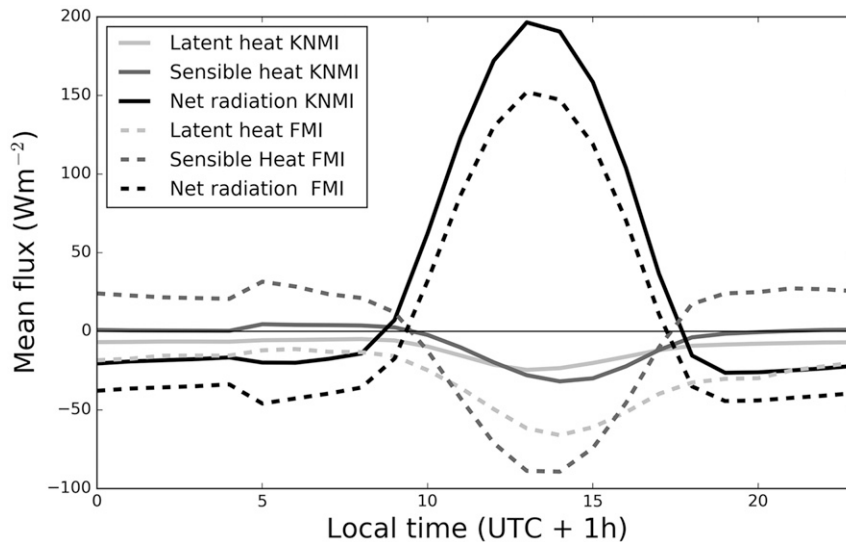


FIG. A1. The averages of net radiation (black lines), latent heat flux (light gray lines), and sensible heat flux (dark gray lines) in the 0300 UTC model runs as a function of local time. The values for the KNMI model are shown by continuous lines and the values for FMI model by dashed lines.

resulted tri-diagonal matrix system is solved iteratively by the Thomas algorithm (Campbell 1985). As the lower boundary conditions, the model uses a climatological average that changes depending on the time of the year. The model was modified to use a simpler forward Euler method as the numerical solution to the heat transfer when coupling was added to the model. This method is used during the coupling phase and onward because the coupling did not work well with the Thomas algorithm-based solving method. The time step is also changed in the FMI model when the coupling phase starts. Before this change is made, the FMI model uses a time step of 5 min in the initialization, but afterward the time step is reduced to 30 s. The KNMI model implements the forward Euler method during the whole model run with a time step of 10 s and the bottom-layer temperature can evolve freely. On bridges the heat transfer from the air below also affects the bottom-layer temperature.

REFERENCES

- Baas, P., and H. Van den Brink, 2014: The added value of the high-resolution HARMONIE model for deriving the HBCs: Contribution to WP 1 of the WTI 2017 Wind Modelling Project. KNMI Final Rep., 54 pp. [Available online at publicaties.miniennm.nl/download-bijlage/61534/report-dec13-added-value-20140207-final.pdf.]
- Bengtsson, L., and Coauthors, 2017: The HARMONIE-AROME model configuration in the ALADIN-HIRLAM NWP system. *Mon. Wea. Rev.*, doi:10.1175/MWR-D-16-0417.1, in press.
- Bogren, J., T. Gustavsson, M. Karlsson, and U. Postgård, 2000: The impact of screening on road surface temperature. *Meteor. Appl.*, **7**, 97–104, doi:10.1017/S135048270000150X.
- Brutsaert, W., 1984: *Evaporation into the Atmosphere: Theory, History, and Applications*. D. Reidel, 299 pp.
- Calder, I. R., 1990: *Evaporation in the Uplands*. John Wiley and Sons, 148 pp.
- Campbell, G. S., 1985: *Soil Physics with BASIC*. Elsevier, 150 pp.
- , 1986: *An Introduction to Environmental Biophysics*. Springer-Verlag, 159 pp.
- Chapman, L., and J. E. Thornes, 2011: What spatial resolution do we need for a route-based road weather decision support system? *Theor. Appl. Climatol.*, **104**, 551–559, doi:10.1007/s00704-011-0433-9.
- , —, and A. V. Bradley, 2001: Modeling of road surface temperature from a geographical parameter database. Part 2: Numerical. *Meteor. Appl.*, **8**, 421–436, doi:10.1017/S1350482701004042.
- Crevier, L. P., and Y. Delage, 2001: A new model for road-condition forecasting in Canada. *J. Appl. Meteor.*, **40**, 2026–2037, doi:10.1175/1520-0450(2001)040<2026:MANMFR>2.0.CO;2.
- David, Y., and M. L. Portal, Eds., 1985: Electronic traffic aids on major roads. Commission of the European Communities Rep. 9835, Science and Technology Policy Series, 68 pp. [Available online at <https://bookshop.europa.eu/en/electronic-traffic-aids-on-major-roads-cost-30-bis-pbCDNW85003/>.]
- De Bruin, H., 1994: *Micrometeorologie*. Landbouwniversiteit Wageningen, 159 pp.
- Ebert, E., and Coauthors, 2013: Progress and challenges in forecast verification. *Meteor. Appl.*, **20**, 130–139, doi:10.1002/met.1392.
- Efron, B., and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap*. Chapman and Hall, 436 pp.
- Fujimoto, A., A. Saida, and T. Fukuhara, 2012: A new approach to modeling vehicle-induced heat and its thermal effects on road surface temperature. *J. Appl. Meteor. Climatol.*, **51**, 1980–1993, doi:10.1175/JAMC-D-11-0156.1.
- Garratt, J. R., 1992: *The Atmospheric Boundary Layer*. Cambridge University Press, 316 pp.
- Hogan, R. J., and I. B. Mason, 2011: Deterministic forecasts of binary events. *Forecast Verification: A Practitioner's Guide in*

- Atmospheric Science*, 2nd ed., I. T. Jolliffe and D. B. Stephenson, Eds., John Wiley and Sons, 31–59, doi:10.1002/9781119960003.ch3.
- Jacobs, W., and W. E. Raatz, 1996: Forecasting road-surface temperatures for different site characteristics. *Meteor. Appl.*, **3**, 243–256, doi:10.1002/met.5060030306.
- Juga, I., P. Nurmi, and M. Hippi, 2013: Statistical modelling of wintertime road surface friction. *Meteor. Appl.*, **20**, 318–329, doi:10.1002/met.1285.
- Kangas, M., M. Heikinheimo, and M. Hippi, 2015: RoadSurf—A modelling system for predicting road weather and road surface conditions. *Meteor. Appl.*, **22**, 544–533, doi:10.1002/met.1486.
- Karsisto, V., P. Nurmi, M. Kangas, M. Hippi, C. Fortelius, S. Niemelä, and H. Järvinen, 2016: Improving road weather model forecasts by adjusting the radiation input. *Meteor. Appl.*, **23**, 503–513, doi:10.1002/met.1574.
- Masson, V., and Coauthors, 2013: 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, doi:10.5194/gmd-6-929-2013.
- Nysten, E., 1980: Determination and forecasting of road surface temperature in the COST 30 automatic road station. Finnish Meteorological Institute Tech Rep. 23, 32 pp.
- Rayer, P. J., 1987: The Meteorological Office road surface temperature model. *Meteor. Mag.*, **116**, 180–191.
- Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, doi:10.1175/2008WAF2222159.1.
- Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, **139**, 976–991, doi:10.1175/2010MWR3425.1.
- Shao, J., and P. J. Lister, 1996: An automated nowcasting model of road surface temperature and state for winter road maintenance. *J. Appl. Meteor.*, **35**, 1352–1361, doi:10.1175/1520-0450(1996)035<1352:AANMOR>2.0.CO;2.
- Thornes, J. E., 1984: The prediction of ice formation on motorways in Britain. Ph.D. thesis, University of London, 404 pp. [Available online at <http://discovery.ucl.ac.uk/1348922/1/336707.pdf>.]
- , and J. Shao, 1991: A comparison of UK ice prediction models. *Meteor. Mag.*, **118**, 93–99.
- Tourula, T., and M. Heikinheimo, 1998: Modelling evapotranspiration from a barley field over the growing season. *Agric. For. Meteorol.*, **91**, 237–250, doi:10.1016/S0168-1923(98)00065-3.
- Undén, P., and Coauthors, 2002: HIRLAM-5 scientific documentation. HIRLAM-5 Project, 144 pp. [Available online at http://hirlam.org/index.php/hirlam-documentation/doc_download/270-hirlam-scientific-documentation-december-2002.]
- Van Ulden, A. P., and A. A. M. Holtslag, 1985: Estimation of atmospheric boundary layer parameters for diffusion applications. *J. Climate Appl. Meteorol.*, **24**, 1196–1207, doi:10.1175/1520-0450(1985)024<1196:EOABLP>2.0.CO;2.
- Venäläinen, A., and M. Kangas, 2003: Estimation of winter road maintenance costs using climate data. *Meteor. Appl.*, **10**, 69–73, doi:10.1017/S1350482703005073.
- WMO, 2014: Forecast verification for the African severe weather forecasting demonstration projects. WMO 1132, 31 pp. [Available at https://www.wmo.int/pages/prog/www/Documents/1132_en.pdf.]
- Yang, C. H., D. G. Yun, and J. G. Sung, 2012: Validation of road surface temperature prediction model using real-time weather forecasts. *KSCE J. Civ. Eng.*, **16**, 1289–1294, doi:10.1007/s12205-012-1649-7.