Detailed responses to referees comments

We thank both reviewers for their comments on the study. Following their comments, we made significant changes in the manuscript.

We have changed the scope of this study by focussing on urban climate applications. This has induced a change in the title of the article, which is now: 'Improving winter conditions simulations in SURFEX-TEB v9.0 with a multilayer snow and ice model'. All the introduction has been modified to fit the new focus of the paper. The use of TEB for road weather forecasting is still discussed in the article as an important future application of this work. This answers the reviewers' reservation of the current TEB-ES parameterisation that lacks human activities parameterization for use for road winter maintenance applications. You will see important changes in the article, from the abstract to the conclusion.

Best regards.

Fixed issues in the manuscript for both reviewers:

A.C, author's comments: After further checking, thanks to the reviewers' comments, we have spotted errors in both the text and the code of the ice modelling. It leads to minor modifications on the results. We have modified the manuscript in agreement with the following modifications

A.C Changes in the manuscript:

$$F = \frac{1}{\tau}\min(W_s, \frac{\max(O, T_f - T_{road})}{C_l L_f})$$

$$M = \frac{1}{\tau}\min(W_i, \frac{\max(O, T_{road} - T_f)}{C_l L_f})$$

$$LE^* = \rho_a \frac{1}{R_a L_s}[Q_{sati}(T) - Q_a] \qquad LE^* \leq 0$$

Models	Detection rate %	Missed event rate %	False ~~detection~~ positive rate %	False ~~alarm~~ discovery rate %
TEB	74	26	4	2
TEB-ES	~~90~~ 92	~~9~~ 8	~~5~~ 7	~~2~~ 3
ISBA-Route/CROCUS	97	3	21	9
MLR	70	30	6	4

Table 3. Performance of TEB, TEB-ES and the benchmarks (ISBA-Route/CROCUS, MLR) surface temperature occurrence, below 0.5 C, ~~for~~ at 1 hour time step, at the Col de Porte location during the winter 1998-1999

Scores	Snow height [m])			Road surface temp. [°C]			
Models	TEB	TEB-ES	ISBA-Route/CROCUS	TEB	TEB-ES	ISBA-Route/CROCUS	MLR
RMSE	0.19	0.13	0.14	2.82	~~2.05~~ 2.27	2.53	3.64
MAE	0.12	0.08	0.09	2.10	~~1.33~~ 1.39	1.40	2.45
R^2	0.54	0.84	0.80	0.82	~~0.89~~ 0.86	0.83	0.57
Bias	-0.04	-0.02	-0.02	0.02	~~-0.62~~ -0.75	-0.92	0.00

Table 2. Scores for TEB, TEB-ES and the benchmarks (ISBA-Route/CROCUS, MLR) at the Col de Porte location during winter 1998-1999

Moreover, in the following table it was the MSE that was calculated. We fixed the values for RMSE.

	TEB	TEB-ES	MLR
RMSE	~~2.51~~ 1.58	~~2.27~~ 1.51	1.19
MAE	1.19	~~1.12~~ 1.13	0.83
R^2	0.95	0.95	0.97
Bias	-0.31	~~-0.31~~ -0.32	0

Table 5. Road surface temperature scores for the Hajala site during winter 2017-2018.

Major changes for the first reviewer (R.C 1)

R.C 1: Unfortunately, the improved model falls far short of achieving this aim. This is mainly due to the lack of processes included in the model. These are listed by the authors at the very end of the paper. 'Traffic heating, salting, water splashing, and snow compaction impact the road surface conditions and the road surface temperature.' I would add snow spraying and snow ploughing to this list as well as an improved description of water drainage. Since the authors include none of these processes then the model will not be able to reproduce real world winter road conditions.

Of the two experiments used for validation only the Col de Porte experiment is relevant for this model. However, this site is not a road and so has little relevance to the title of the paper. What the authors seem to have done is to have improved the TEB model so it can better model snow, but not made it relevant at all to roads. The reviewer presumes this to be one of the next steps.

Even so, there is clearly a process involved in this model development and this paper would appear to be the first stage in a general application of the TEB model for applications concerning road weather. This reviewer then suggests that either the authors wait until they have implemented these relevant processes or that the paper be rewritten in a way that makes this article the first step in several model improvements. This can be made clear in the title of the paper, for instance using a phrase like 'Towards improved winter conditions … for road weather applications'. The authors would be wise to state what the process will be, including their next steps in making the model practically usable on real roads and use the Hajala site to demonstrate these needs.

A.C, author's comments: We have changed the scope of this study by focussing on urban climate applications of this work. This has induced a change in the title of the article, which is now: 'Improving winter conditions simulations in SURFEX-TEB v9.0 with a multilayer snow and ice model'. The use of TEB for road weather forecasting remains discussed in the article as an important future application of this work. This answers the reviewers' reservation of the current TEB-ES parameterisation that lacks human activities parameterization for use for road winter maintenance applications.The title, abstract, introduction, results, discussion, conclusion have been changed.This is an important reshape of the article context that focuses on the modelling of winter conditions on artificial surfaces. In TEB, the artificial surfaces at ground level are called: road. But the "road" component represents sidewalk, roads or parking as written in the article section 2.1.

For the current article, we have shown how the new processes improve the simulation of winter conditions on artificial surfaces. Indeed, while almost all the roads are subject to anthropogenic impacts, this is not the case for all artificial surfaces. For cities, the snow cover can hold for months and a better representation of the snow evolution is needed to simulate its impact on the urban climate as introduced in the Introduction. Moreover, it is our purpose to bridge the gap between urban climate and road weather models as we intend to use TEB for both cold urban climate simulations and road weather forecasts. We believe that informing the community about the advantages of using an urban model for simulating road conditions is important and will encourage the community to improve this tool.

We also have added a second appendix to explain the snow cover parameterization in cities to account for winter maintenance operations. We have shortened the analysis of the results, particularly on the Finnish site, and show the difficulties of the model in simulating road state conditions (snow and ice on the road) with road with traffic and winter maintenance activities.

A.C Changes in the manuscript: We have added a new title for this article: “Improving winter conditions simulations in SURFEX-TEB v9.0 with a multi-layer snow model and ice”

Abstract: (l1-l5) we have removed the previous context for the new context of this study: “In winter, snow or ice covered artificial surfaces are important aspect of the urban climate and trigger road maintenance operations. Urban climate and road weather models have specialised in simulating these conditions in cities or in the countryside, respectively. In this study, we intend to bridge the gap between road weather models and urban climate models in terms of cold regions urban modelling and artificial surface condition predictions in any environment”

Introduction : The previous two first paragraph have been removed and we have written two new paragraph (l16 - l45)

Discussion: We have added 3 new paragraphs, the first paragraph and second paragraph (l459-484), MLR analysis (l507 -519), removed some of the previous discussion and focused the discussion on urban climate simulations.

Conclusion: first paragraph, 3 last sentences

New appendix b : snow cover parameterization in cities

R.C 1: Since the model does not include traffic or winter road activities the Hajala site was not suitable for model comparisons, where the aim of the model comparisons was to test the new snow model routine. The only reasonable comparison would be the surface temperature, which the model seems to reproduce quite well. This section can be significantly reduced if the authors use it more to demonstrate that the model is not suitable for road applications yet, rather than to try to find moments in the data which actually are reproduced in the model.

A.C, author’s comments: We thank the reviewer for this comment. We acknowledge that the improvements shown are a significant step, yet further improvements are needed to simulate road states on a road with traffic and winter maintenance operations. In the edited paper, we have reduced the analysis of the Finnish site experiment by focusing on the improvement of road temperature simulation and the difficulties of the model in predicting road state. Indeed, the passage of vehicles and the systematic use of salt to melt slippery conditions strongly modifies the road contaminant. This experiment provides an important result: modeling the effects of human activities on the road is essential for predicting its state. We have added this result to the result section 5.1 and the abstract.

A.C Changes in the manuscript: **(l12 - l13)** Abstract: “For roads with high traffic and/or winter maintenance operations, future modeling work should focus on the representations of anthropogenic effects.”

(l414 - l420) section 5.1 :In this 9-day period, the road surface temperature is well simulated for both models with similar results. But TEB-ES significantly outperforms TEB on January

25th moderate snowfall followed by a rainfall episode. It shows an improved modelling of the snow mantle variables with ES for positive air temperatures. In both models, the simulated SWE occurrence is consistent with the observed snowfall but fails to match the observed mass content on the road. The snow removal parameterisation at 6h UTC in the models appears to be an oversimplification of road maintenance operations such as salting or snow ploughings.

R.C 1: Generally the article is well written, though there were a large number of small mistakes in the text. This should be proof read before the next submission.

A.C, author's comments: Many changes have been made throughout the article from grammar corrections to sentences written more clearly as shown by the latexdiff PDf article.

Specific changes from the first reviewer (R.C 1):

R.C 1, line 42: Only temperature is simulated with the heat balance equation, not surface moisture for example. Rewrite this so it makes sense.

A.C, author's comments: Thank you for this comment, we will modify this sentence and make it clearer. Indeed, the heat balance equation only modifies the temperatures and not the other variables.

A.C Changes in the manuscript: (l47 - l49) "Land Surface Models (LSM) based on the physical heat balance equation are well suited as they can represent various surface types ranging from natural to artificial surfaces in urban environments."

R.C 1, line 70: Do you have a reference/proof for this statement? It is a fairly general statement.

A.C, author's comments: This argument is based on experiments conducted in the urban climate community. Physical models that take little or no account of urban morphology tend to have lower performance in simulating physical variables in cities than those that take it into account (2-tile model or canyon model) [Lipson et al. 2024]. By analogy, the behavior will be the same for road weather models that do not take into account the effect of building geometry on surface conditions.

A.C Changes in the manuscript: (l80-l81) On the other hand, road weather models, with physics comparable to one-tile urban models \citep{Lipson2024} struggle to compute accurate urban conditions

R.C 1, line 82: When modelling individual roads with TEB, can the canyon direction be specified?

A.C, author's comments: The direction of the canyon can indeed be specified since the version of TEB in Surface V9. We will mention this in the text.

A.C Changes in the manuscript: (l108) "The direct solar flux received by the road or the walls is computed according to shadowing effects and the direction of the road. "

R.C 1, line 107: It would be useful for the reader if a small table was included giving the actual values of the parameters used here.

A.C, author's comments: The parameters associated with equation 2 are presented in table 1. We will add to the text the link between this equation and the table.

A.C Changes in the manuscript: (l124-l125) " CR1 is the heat capacity of the road surface depending on the dry surface heat capacity and of the mass contents amount, and dR1 the depth of the first road layer with values as described in Table 1"

R.C 1, line 124: I would assume that albedo is diagnosed from the other prognostic variables, perhaps density as well? Make this clear.

A.C, author's comments: The albedo does not depend on the other variables. Indeed, as in Douville 95 (Douville et al. 1995a), it follows an exponential law that depends on the initial condition and a time constant, whose values are the same as in Lemonsu et al. 2010.

A.C Changes in the manuscript: l140 Simple formulations are used for snow density and snow albedo with exponential evolution laws to represent snow ageing

R.C 1, line 137, line 144: So there is no conduction of heat through the snow layer to the ice layer? This seems strange. Could you explain why this flux is not necessary? Only a thin layer, for example?

I do not follow this. The amount of melting or freezing is a result of energy exchange but here it is parameterised in terms of a temperature difference and a time scale, that is 18 hours which seems very long. Reading Boone 2000 refers to earlier references and that parameterisation was made for ice in soil, not ice on the surface. This part of the model is somewhat unclear so it would be good to provide some clearer explanation of what is actually being done here.

A.C, author's comments: The introduction of conduction between snow and ice, and ice and road surface would strongly modify the way mass reservoirs are taken into account in TEB and more broadly in SURFEX, the mass reservoirs being simulated in a similar way in ISBA.To be in agreement with the TEB modeling choices, the ice layer Wi energy, as the water reservoir Ws energy is not explicitly modeled. Their energies are considered indistinguishable from the road surface.Therefore, there is no conduction flux between the soil and the water, the ice and the water, the ice and the snow. The energy consumption of the processes, while taking into account that the ice energy is indistinguishable from the road surface layer, are taken into account from the road surface heat-balance equation.

Thus, the processes involved in ice evolution, freezing , melting and sublimation, are parametrized by simple energy consumption methods that directly impact the road surface budget (Boone et al. 2000). This approximation remains reasonable and will not much influence the results.The freezing rate, the melting rate and the solid-gas latent heat flux equations for the ice evolution on the natural surface layer from Boone et al. (2000) are adapted for impermeable artificial surfaces. By changing the values of the coefficients involved, we adapt this parameterization for impervious surfaces. Moreover, the insulating effects of the vegetation is removed (Ks=1) (Boone et al. (2000)) as well as the phase change coefficients for subgrid-scales effects (epsilon=1) (Boone et al. (2000)).

A.C Changes in the manuscript: (l151- l155) To be in agreement with the TEB modeling choices, the ice layer Wi energy, like the water reservoir Ws energy is not explicitly modeled. Their temperatures are considered indistinguishable from the road surface. Thus, the processes involved in the ice evolution, freezing , melting and sublimation impact directly the road surface temperature as follows :

(l158- l159) Freezing of water is an exothermic reaction while melting is endothermic. This will affect the energy balance at the surface of the road as shown in Fig.2 . (l169 - l172) The freezing rate, the melting rate and the solid-gas latent heat flux equations for the ice evolution on the natural surface layer from Boone et al. 2000 are adapted to impermeable artificial surfaces. The insulating effects of the vegetation are removed as well as the phase change coefficients for subgrid-scales effects. The adapted water mass rates for impermeable artificial surfaces are therefore defined as :

R.C 1, line 149: You increase the time scale, from Boone 2000, by a factor of 20. Why?

A.C, author's comments: After further analysis thanks to the reviewer's comments, we use the value of 25000 seconds rather than the 3300/0.05 value in the paper, also used in SURFEX/ISBA when the model is forced by the ARPEGE atmospheric model. This value is the most consistent with observations (co-evolution of the ice and water reservoirs) particularly on the Finnish experiment. The value given in (Boone 2000.) of 3300 seconds turned out to be much too low. We recognize that further verification is necessary regarding this value. We will add a comment on this subject to the discussion.

A.C Changes in the manuscript: (l178- 179) The tau parameter is set to 25000 s rather than the 3300 s value used for natural soil in order to get realistic simulations for this study experiments.

First, the ice content modeling could be improved by finding a better estimate of the characteristic timescale for phase change tau set for now to 25000 s. This parameter value should be evaluated more rigorously on more experiments to get a better estimate.

R.C 1, line 150: Why is it assumed to be? You have the equations for it in LE, but it is set to 0? Can you explain why this is done? If it is snow covered OK, but otherwise?

A.C, author's comments: The sublimation of ice is negligible. Indeed, this quantity is very small in all atmospheric conditions, particularly when compared to the melting process.

A.C Changes in the manuscript: (l179 - l181) Ice sublimation is assumed to be negligible because its evolution is small compared to F or M. Thus, when the road surface reaches the saturation specific humidity with Qsati <= Qa, LE* is set to 0. So LE <= 0, and deposition as frost on the road can occur.

R.C 1, line 163: The weakest part of this modelling is that it does not include traffic. Most of the details in the snow modelling become irrelevant when traffic is included. There are no roads of interest for road weather modelling that do not have traffic.

A.C, author's comments: We have changed the scope of the paper and focus on the improvement of winter urban climate processes. The use of TEB is not limited to artificial surfaces with vehicles. It is used for all urban surfaces as sidewalks, parking lots, pedestrian paths, bike paths. But road weather forecasts remain a significant potential application of this work.

The impact of road traffic on the snowpack has been little studied. However, the compaction of snow by vehicles is likely to have a very important effect on the physical variables of the snowpack. Indeed, according to T. Kobayashi et al. 2006, just a few dozen vehicle passages cause the snow density to be multiplied by 4, which probably explains the very low values observed of water quantity by the sensors in Finland.

Some processes modeled in the multi-layer snow model remain essential which are not modeled in the previous model. For example, the use of a temperature profile allows us to model the real temperature of the snow and in particular represent the isothermal episodes that a single-layer model cannot handle correctly. The new model also allows us to take into account the effect of rain on the evolution of the liquid water content of the snowpack and a part of the incoming radiation goes through the snowpack and is absorbed by the road.

A.C Changes in the manuscript: Discussion about Finnish experiment performance: (l473 -l482)"Our study suggests that new developments within TEB are interesting for artificial surface predictions but are flawed for roads impacted by human activities. Indeed, overall model performance for the Finland experiment is poorer than for the Col de Porte experiment, as shown by the experiment's analysis and scores. This inferior performance is caused by several factors caused by human activities: errors in modelling snow removal, salting not modelled, or traffic effects not modelled (snow compaction and heating effects). In fact, traffic has a large effect on snow compaction: it reduces the snow depth and leads to measurement errors. In addition, Finland's winter road maintenance operator salts major roads whenever a slippery road condition is observed or forecast. Snow ploughing and salting is roughly simulated in the models by mechanical snow and ice removal every morning at 6h UTC in the Hajala experiment. The actual effects and timings of winter service vehicles are more complicated and impact the water contents and the surface heat energy. Salting indirectly affects road surface temperature by melting the snow cover that insulates the road from the atmosphere."

R.C 1, line 230: Sorry, but what is downgraded? Or do you mean simplified? Should be clearer what you are comparing against. Does this road weather model include traffic?

A.C, author's comments: The operational version of ISBA-Route/CROCUS in Météo France is the same but the forcings are diffferents compared to the one described in the 2006 study (Bouilloud and Martin 2006). We will remove this word that does not make sense here. ISBA-Route/CROCUS model does not include human activities.

A.C Changes in the manuscript: (l252 -l253) First, the heat-balance model ISBA-Route/CROCUS described in Bouilloud and Martin (2006), in operation at the French national meteorological office, is used in comparison at Col de Porte location

R.C 1, line 298: I would say there is too high a level of detail in the meteorological description, given it is temperature and precipitation that are important. It is not necessary to know for the reader about where these weather systems come from.

A.C, author's comments: In agreement with the comment, we will reduce the synoptic weather descriptions on this subection and those of the Finnish experiments.

A.C Changes in the manuscript: (l359-l361) After small observed snowfall in the afternoon of November the 12th, the snow depth evolution is well computed by TEB-ES with less than a 2 cm difference with the observations.

Deleted: Small persistent snowfalls on January 17th and January 18th in the morning were induced by associated with a a small low surface pressure weakly active quasi-stationary front.

In early April, a low pressure system traveled fast from the Baltic States to Hajala and hit the station

Then, in the night between January 23rd and January 24th, the upstream warm front of the weakening low pressure system brought moderate snowfall

warmer air mass brought in by the low pressure system.

R.C 1, line 330: This period with constant 0 temperature. What is being measured/modelled. Is it the pavement temperature or is it the snow surface temperature? Is the pavement at 0 all the time?

A.C, author's comments: The road surface temperature is at 0°C. The temperature probe is buried in the road. We display in the figure the road surface temperature, so the first layer of the road in contact with the snow mantle bottom layer.

A.C Changes in the manuscript: (l291- l292) The road surface temperature, with a probe inside the artificial structure,and the snow depth were monitored

R.C 1,2 : It would be useful for the reader to see the air temperature in these plots as well.

A.C, author's comments: Thank you for this comment. We have modified the figures accordingly.

R.C 1, line 364: Based on the information you provide I would say that the surface temperature is reasonably modelled but nothing else is. As you state there are salting and snow ploughing activities not included in the model, not to mention the impact of traffic. I would conclude that 'clearly road surface conditions are strongly influenced by these activities and the current models do not adequately predict road surface conditions as a result

A.C, author's comments: We agree with this comment and will change the section accordingly by comments on the road surface temperature simulation performance, then by raw comments on mass contents physics consistency with the atmospheric forcings.

A.C Changes in the manuscript: (l445- 451) As previously seen, the snow removal parameterisation at 6 h UTC in both TEB and TEB-ES is not complex enough to represent road maintenance operations such as salting or snow ploughings which greatly influences the mass content evolution. For instance, freezing is simulated more often than observed with 743 observed ice occurrences and 1368 modelled ice occurrences. As said in the sensor descriptions, the optical sensors are not able to distinguish between snow-covered or ice-covered road conditions on busy lanes. This explains the high missed event rate shown in Table9. The scores are shown in Table9 and Table8, but only a raw analysis can be extracted from these values. Therefore, the mass contents performances can not be compared between the models on this road, because of human activities.

(l452- 457) However, the models simulations are consistent with physics, and tend to accurately represent the road conditions without human activities. The physical consistency of the models with the observed precipitation leads to an high detection rates for SWE but a snow cover that is lower in observations than modelled (706 hours observed, 1394 modelled by TEB and 1417 modelled by TEB-ES) which leads to a >60% false discovery rate on both models as shown Table7 In addition, the snow occurrence detection by the model shows < 25% false detection rate and > 70% detection rate. There are a high number of events without snow and ice.

R.C 1 line 376: If I am reading the figure correctly then both models do not capture the snow evolution at all. There is simply way too much snow in the models.

A.C, author's comments: We have changed the comment on the result section accordingly.

A.C Changes in the manuscript: (l406 - l408)The models fail to match the time span of the event. It could be explained by the strong morning traffic commuting pattern that blew the snow away, removed the thin flake layer on the road, and then delayed the accumulation of the snow cover Denby2013b.

R.C 1, line 376: Why? Is this to simulate snow removal

A.C, author's comments: At 6h UTC in the models, the snow mantle is removed. It is a simple parameterization to take into account the snow ploughing on the Finnish site.

A.C Changes in the manuscript: (l276 - l280) A straightforward parameterisation of snow removal operations is implemented in TEB, TEB-ES, and ISBA-Route/CROCUS. Within these models, snow depth and ice content are set to zero when an operator clears the snow. This parameterisation is adapted for the snow removal procedures carried out at the experimental sites presented in the next section. In the Col de Porte experiment, snow and ice are reset to 0 on the known date of snow removal by a manual operator. In the Hajala experiment, the exact dates of winter maintenance activities are unknown. So, snow and ice are reset to 0 at 6h UTC in the models.

Major changes for the second reviewer (R.C 2)

R.C 2: Why were the two benchmarks not evaluated at Hajala? Given the apparent unsuitability of either TEB or TEB-ES to predict conditions at a real world site, it would be helpful to know if an operational road forecasting model, such as ISBA-Route, performs any better. Ideally results from the two benchmarks should be shown for the Hajala site but if that is not possible it should be briefly explained in the text why not.

A.C, author's comments: We have decided to not include the ISBA-Route/CROCUS benchmark on the Finnish site. Indeed, we have shown that TEB and TEB-ES comparison on the Finnish site is difficult and cannot be completely exploited due to the human activities impacts. ISBA-Route/CROCUS as TEB does not simulate the human activities impact, and then it will lead to the same conclusion. On the other hand, the statistical benchmark could be interesting to see if a statistical model is capable to infer the human activities within the data. So we have added a simple MLR model on the Hajala experiment.

A.C Changes in the manuscript:**(l252-255) In the model configuration:** First, the heat-balance model ISBA-Route/CROCUS described in Bouilloud and Martin (\citeyear{Bouilloud2006}), in operation at the French national meteorological office, is used in comparison at Col de Porte location. The model is not used at the Hajala location. Indeed, it lacks processes to model the impacts of human activities such as salting or traffic heating as in TEB and TEB-ES.

In the results: (l441 - 444) However, the statistical benchmark significantly outperforms the heat-balance models with better RMSE, MAE, R2 and bias. It is able to retrieve an accurate road surface temperature for each season as shown Fig.9. Even if the MLR predictions are mainly unbiased, Fig.9 shows that in snow-covered conditions the road surface temperature simulated is slightly biased.

In the discussion: (l507 - 519) Comparison of heat-balance models with statistical benchmarks provides interesting insight for further studies. The artificial surface is a low inertia and simple enough system with easily modelled behavior as shown by the good in-sample MLR performance in the Finland experiment. Although this behaviour is true in an open environment, more validation is needed with roadside components, trees, or buildings. The in-sample MLR Hajala simulation which has been trained using observed road surface temperature is also capable of correcting the forcing errors and captures the impacts of human activities. So, these components could be systematic and cyclical enough to be easily modelled. It means that there is potential for further studies to take into account these effects in the heat-balance models. In both experiments, MLR models struggle to simulate the road surface temperatures when snow-covered. It leads to poor performance on Col de Porte with a mostly snow-covered road during the 6 month experiment. It suggests that the snow/road coupling is crucial for the heat-balance model performances. Indeed, it is difficult to capture surface physics when the road surface temperature is insulated from the atmosphere by the snow mantle. More complex statistical methods are needed, such as recurrent neural networks, to take into account the long-term system inertia and model the coevolution of road surface temperature with road mass contents. However, training such models is likely to require the acquisition of accurate mass content observations.

New figure and extended table in Appendix1

R.C 2: More detail is needed to explain what data was used to drive the model simulations at Hajala. Did you use the observations directly from the synop station at Hajala or output from either the Kriging process or the HARMONIE-AROME model runs described by Karisisto and Loven 2019? Either way, is this data publicly available and where can it be found (it does not appear to be included at the link provided in your data availability statement)? If the input data is not available then this needs to be made clear in your data availability statement. You also mention that "The hourly atmospheric forcing for the model consisted of a mix of observation data and ERA5 reanalysis" – please give more detail as to how the ERA5 data was mixed with the observations. Was this for gap filling? How much ERA5 data was used? Was it for all or just some of the fields?

A.C, author's comments: We acknowledge the confusion, we have modified and improved the methodology section to clarify the experimental setup. On the Finnish site, we have used the observations from the on-site road weather station as forcing rather than the forecasts from the HARMONIE-AROME model. These data are not in the Karsisto and Loven 2019 Zenodo. They are included in the Zenodo associated with the manuscript. The shortwave and longwave radiation data from ERA5 are used because the radiation data are not measured at the road weather station. Furthermore, the nearest SYNOP station is too far away and does not measure the longwave radiation.

A.C Changes in the manuscript: **(l306 - l308)** Fintraffic has installed numerous road weather stations to monitor atmospheric variables (wind speed, air temperature, humidity and precipitation), road surface temperature and road conditions.

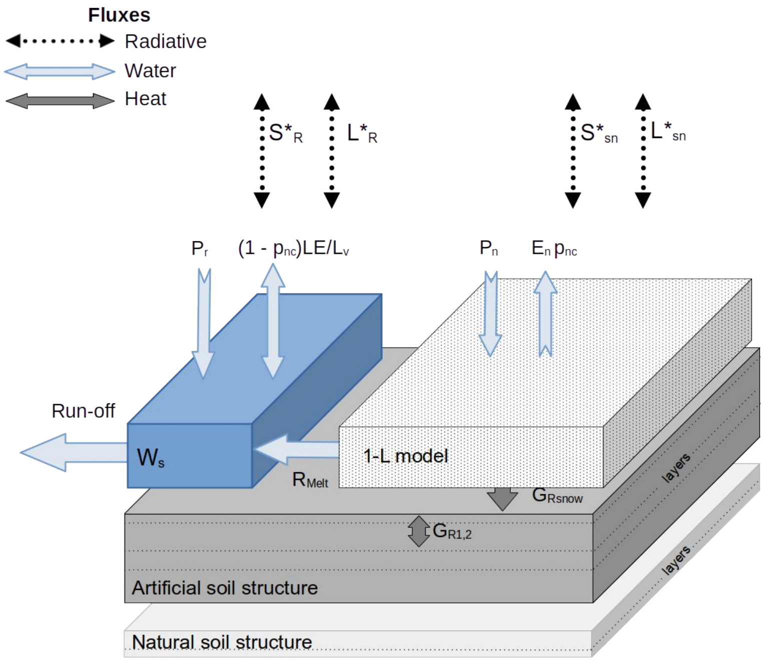
(l313 - l317) To force and validate the model, we used data from a study conducted by Karsisto and Loven Karsisto2009 and from measurements provided by the Finnish Meteorological Institute Colas2024. Observed wind speed, air temperature, humidity and precipitation from the Hajala road weather station are used as atmospheric forcing in the model and processed in the same way as the Col de Porte forcing. Since there is no radiation measurement at the Hajala road weather stations, shortwave and longwave radiation were extracted from ERA5 reanalysis at the closest grid point Hersbach2020.

R.C 2: It is not clear in the text whether section 2.2 is a description of an existing treatment of ice in TEB or a new treatment for TEB-ES. Authors need to make it clear which aspects of the model are novel in this work.

A.C, author's comments: In the original version of TEB, the ice layer was missing, so the water phase was always liquid, even in freezing conditions. In light of your comment, we will improve the clarity of this section and emphasize the relevance of this new component. We will also add on the appendix or in the paper, a summary scheme of the processes modeled on the original version of TEB V9.0 surface.

A.C Changes in the manuscript: (l143 - l145) The road component of the TEB initial version described previously and represented in Fig.1 is then modified to improve the representation of winter processes. The following sections describe the modification within TEB represented in Fig.2 with a new ice content Wi and an improved representation of the snow mantle with the Explicit Snow (ES) model.

New figure:



R.C 2, comment technical N°1: In general there are numerous instances of inaccurate grammar and sentences that could be written more clearly. I suggest that with further proofreading, this manuscript could be edited to become much easier to comprehend.

A.C, author's comments: Many changes have been made throughout the article from grammar corrections to sentences written more clearly as shown by the latexdiff PDf article.

Specific changes for the second reviewer (R.C 2):

R.C 2, comment N°1: Abstract: A major finding of this study was that physical models need to include representation of various effects due to traffic and road maintenance activities to be able to accurately predict real world road surface conditions. This finding should be included in your abstract.

A.C, author's comments: Thank you for this useful comment, we have added this result to the abstract.

A.C Changes in the manuscript: (l12-l13) For roads with high-traffic and/or winter maintenance operations, future modelling work should focus on the representations of anthropogenic effects.

R.C 2, comment N°2: Line 134 and 135: I'm not sure I understand, if ice content growth is limited by water availability then what other limitations are being referred to when it says there are none? Perhaps the first sentence means "Contrary to the liquid water content, ice content can grow on both snow covered and snow free fractions of the road surface”?

A.C, author’s comments: The water content in the model Ws, is constrained by the maximum size of the reservoir, set according to the experimental conditions. In the original version of TEB, it is set to 1 mm. For the Col De Porte experiment both TEB and TEB-ES maximum water content is set to 1 mm. But for the Finnish experiment, the parameter is set to 0.6 mm. This maximum content parameter does not exist for the ice layer. Indeed, the ice layer can grow indefinitely and depends on the available liquid water. If the water content is zero, then the freezing rate F is zero. We will add this explanation to make this section clearer. In practice, this assumption is sufficient. However, one could imagine that at a certain height of ice, the liquid water flows directly to the sides without having time to freeze, i.e., in the model, it would run as runoff.

A.C Changes in the manuscript: (l163- l165) In addition, the ice content can grow without limitation as long as it is supplied by the freezing rate F of the available water content Ws, or by the ice deposition by sublimation on the snow-free fraction (1 - fsn)*LE*/Ls

R.C 2, comment N°3: Figure 1: It might be useful to have a diagram of original TEB infrastructure for comparison.

A.C, author’s comments: We have added a diagram in the Section 2 to understand quickly the improvement of this new TEB version**.** You will see the diagram on the “Major change for the second reviewer”.

R.C 2, comment N°4: Line 165: Is that the correct reference? Vionnet 2012 describes a multi-layer snow model but does not compare it to a single layer one. A need for multiple snow layers is a stated motivation of that paper but is not well explained. They reference Etchevers 2004 which found that more ‘complex models’ (assessed based partially on the number of snow layers) were better able to predict the net longwave radiation flux.

A.C, author’s comments: We appreciate the comment, we have added the 2004 reference which is indeed more relevant in this context.

A.C Changes in the manuscript: (l186 - l 187) The thermal and liquid profiles of the snow mantle cannot be represented by averaged single-layer variables as in one-layer snow model schemes, they require multi-layer models instead (Etchevers et al., 2004).

R.C 2, comment N°5: Line 170: At least 3 layers – why? Is that based on a reference?

A.C, author’s comments: Yes, this argument is based on a reference (Lynch-Stieglitz 1994; Sun et al. 1999).

A.C Changes in the manuscript: (l191-192) The prognostic variables are snow density, heat content, thickness for each snow layer and albedo. Sun et al. 1999 suggested that 3 module layers at least needed to represent a snow thermal profile.

R.C 2, comment N°6: Line 249: Why do you change the way that natural soil is initialised in

TEB-ES? What impacts will this have?

A.C, author's comments: : In the original version of TEB, the natural soil is initialized simply with the dry conductivity. This value is consistent with urban conditions, with mostly artificial surfaces and scarce water infiltration on the entire city area. At our experimental sites, around the artificial road, the soil is natural. So, the water infiltrates and moisten the natural soil under the road. We have done some experiments (not shown here) and this initialization is important for the heat transfer in the soil. The TEB-ES results worsen without this initialization.

A.C Changes in the manuscript: (l271 - 275) The natural soil under the artificial structure is initialized in TEB by the dry soil thermal, whereas for TEB-ES it is initialized with the moist soil thermal conductivity \citep{Bouilloud2006}. Indeed, at our experiment site, water infiltrates beneath the road from surrounding natural soil. For the Finnish experiment in this paper, the TEB-Hydro component is enabled in order to simulate the water run-off on the roads Bernard2020, and Wsmax is set to 0.6 kg m-2 rather than 1 kg m-2 because of the different road properties.

(l350-352) But the moisture conductivity in the natural soil under the pavement added in TEB-ES as seen in Table 1 leads to improved pavement heat restitution and reduces the cold bias by 0.5 °C.

R.C 2, comment N°7: Line 250: It's not clear what is meant by the term 'water wear-off' and it does not appear in the referenced paper. Please can you clarify what this means – is it the same as runoff?

A.C, author's comments: In the article, we have changed the term "wear-off" by the generic term "water drainage" for the reader understability.

R.C 2, comment N°8: Line 261: You mention that six different surfaces are installed at the site but only show one set of results for Col de Porte – did you just use observations from one surface or was it a mix of all six? Was there any performance differences associated with the different pavement types?

A.C, author's comments: We have used the observations from a single surface, the one that corresponds to the French highway equivalent. We do not have run experiments on the other road types. To avoid confusion, we are not talk about the six roads built but simply about the one we have used.

A.C Changes in the manuscript: (l290- 291) An artificial pavement (2m x 3m) equivalent to a French highway shown in Fig.3 was installed at the site Bouilloud2006.

R.C 2, comment N°9: Line 282: Is it possible to quantify the errors associated with these measurements?

A.C, author's comments: We can only quantify the errors with the technical manual from the sensor manufacturers. However, these values are not enough to explain the sensors large errors.In the article we have analyzed the physics consistency of the optical instruments. In the results, the mass contents,except the water, more reliable, are transformed as occurrences for both the simulations and the measurements.

A.C Changes in the manuscript: (l320-328) In many cases, optical instruments are considered unreliable for detecting road conditions with precision subject to anthropic effects and winter maintenance road operations. Indeed, optical sensors always failed to distinguish between ice and snow content on the road surface. In the Hajala experiment, on the 706 snow occurrences and 743 ice occurrences measured, the sensors recorded 706 occurrences of both ice and snow at the same time, and the other 37 hourly occurrences for ice are at the beginning or at the end of a snow event. Anthropic effects such as traffic and winter maintenance directly influence the physical variables. In addition, the optical sensor might only see the top of the snow or ice layer and is unable to measure the actual thickness. For these reasons, the ice and snow mass contents measured by the optical sensor should not be used to validate the models quantitatively but qualitatively as occurrences. They are compared in the Hajala experiment, with the snow and ice output variables from the models transformed as occurrences.

R.C 2, comment N°10: Line 297: I don't understand the second half of this sentence, please clarify.

A.C, author's comments: Thank you for this comment, we have changed the sentence.

A.C Changes in the manuscript: In this reference experience, the new winter processes implemented are evaluated to see whether they have a positive impact on the model performance and the physics consistency. Snow depth and road surface temperature simulations are compared with the observations.

R.C 2, comment N°11: Figure 4: The observed snow depth occasionally appears to show increases when there is no observed precipitation – why is this? (Particularly small blip in observed snow depth on November 12th and large increase on November 17th).

A.C, author's comments: These small increase in snow depth are within the accuracy of the sensor. So, if there is a very thin layer of snow, we can measure 0 or 1cm of snow. The snow surrounding the road, could have been blown by the wind and land on the road, then melted.

R.C 2, comment N°12: Line 324: On what basis do you say that the precipitation forcing was wrong? Is there another observation source for precipitation phase? Presumably the partitioning of precipitation phase on the basis of air temperature isn't always accurate, are you able to comment either here or near line 269 as to how appropriate this assumption is and how often it fails?

A.C, author's comments: In this case, the air temperature is between 0 and 1 degree. We have chosen 1 degree as the criterion for the phase change from rain to snow. Even if in most cases this is true, in this case, it was rather rain that was observed than snow. Not able to quantify the error and no other sources of observations. For more information the reader could refer to Jennings2018.

A.C Changes in the manuscript: **(l300- 301)**Jennings et al. 2018 confirmed that the air temperature of 1 ° C is the average temperature of separation of the precipitation phase in the Alps. But this assumption can often fail Jennings2018.

(l482-484) Discussion: Other measurements errors and sources of uncertainty may decrease the reliability of numerical experiments: lack of precipitation detection by the raingauge, errors in distinguishing between snow and rain, sensor detection errors, and radiation forcing errors from the ERA5 reanalysis.

R.C 2, comment N°13: Tables 3, 4, 6, 7, and 8: What's the difference between 'false detection rate' and 'false alarm rate'? Does one of these columns correspond to the 'true negative' scenario referenced in line 350? It may be worth renaming one of these columns for clarity and/or describing what the headings mean in the table caption.

A.C, author's comments: Ok, clarification of the criterion.

False detection rate also called **false positive rate**: FDR = FP/ (FP + TN) with FP the number of false positives, TN the true negatives

False alarm rate also called **false discovery rate** : FPR = FP/(FP + TP) with FP the number of false positives, TP the number of true positives.

We have made a confusion with the English terms. We will use the new terms **false positive rate** and **false discovery rate** which are more accurate and explain their calculations.

A.C Changes in the manuscript: (l329- 334)Statistical scores are calculated hourly at the Hajala site for the whole simulation, similarly to the Col de Porte experiment. The scores in Table 3, Table 4, Table 6, Table 7 and Table 8 are calculated from the confusion matrices that report the number of true positives (TP), false negatives (FN), false positives (FP), true negatives (TN). They are calculated as follow: detection rate = TP/(TP + PN), missed event rate = FN/(TP + PN), false positive rate = FP/(FP + TN) and false discovery rate = FP/(TP + FP). These metrics help to evaluate the models performance for important thresholds, in particular for decision making in the context of road weather forecasts.

R.C 2, comment N°16: Line 451: Figure 7 does not show observed air temperature, only road surface temperature.

R.C 2, comment N°14: Figures 6b and 7b: The scale is not sufficient to distinguish changes in the observed SWE. The trends described in the text on January 21[st] and 22[nd] cannot be seen in this plot because the values are too small to be distinguishable on this scale. Consider increasing the height of the panel or an alternative method to ensure that the trends described in the text are visible in these figures.

A.C, author's comments: The SWE and Ice in the figures for the Hajala experiment are now displayed as occurrences. You can look on the p19 of this document to see the plots.

R.C 2, comment N°15: Line 388-399: TEB-ES also appears to melt snow considerably slower than the observations. I think this sentence needs to be re-phrased to avoid overstating the performance of TEB-ES in this instance.

A.C, author's comments: We have modified the sentence.

A.C Changes in the manuscript: (l418-420) In both models, the simulated SWE occurrence is consistent with the observed snowfall but fails to match the observed mass content on the road. The snow removal parameterisation at 6 h UTC in the models appears to be an oversimplification of road maintenance operations such as salting or snow ploughings.

R.C 2, comment N°17: Line 460: I agree with the first sentence of this paragraph but it needs more detail to support it and it also seems unrelated to the rest of the paragraph. I would expect a separate paragraph here discussing both the reduced reliability of observations at Hajala and also the impracticalities of using live road observations to validate a model that doesn't represent those conditions.

A.C, author's comments: We have separated this comment from the rest of the paragraph. Now the paragraph three only comments on the Hajala experiment and the fourth about the CDP experiment and what we have learned from it.

R.C 2, comment N°18: In your zenodo dataset, the readme file states that raw observations and validation data can be found in the zenodo dataset associated with Karisto and Loven's paper but I can only see observations for validation data there – the only files containing the atmospheric forcing fields you mention are model output from HARMONIE simulations. You also state that 'raw observations for SW and LW ERA5 radiation are available' but ERA5 is not raw observations – it is a reanalysis product that involves a lot of modelling. Please be very clear, both here and in your paper, what data you are using comes from observations and what comes from models/reanalysis.

A.C, author's comments: As required, we will make the readme file clearer. Indeed, some observations are not available in the associated dataset but directly in the Zenodo dataset associated with this paper, within the experiment folders. To avoid confusions, we have clarified. Similarly, for the reanalyses, we have clarified the description, they are indeed not observations.

A.C Changes in the manuscript: (l313- 314) To force and validate the model, we used data from a study conducted by Karsisto and Loven 2019 and from measurements provided by the Finnish Meteorological Institute Colas2024.

(l316- 317) Since there is no radiation measurement at the Hajala road weather stations, shortwave and longwave radiation were extracted from ERA5 reanalysis at the closest grid point Hersbach2020.

R.C 2, comment N°19: The zipped data file in your zenodo dataset appears to be corrupted – when I try and unzip it I get error messages saying the file is 'invalid' or 'empty'.

A.C, author's comments: A new Zenodo version has been published V3.

Technical corrections from the second reviewer (R.C 2)

R.C 2, comment N°1: In general there are numerous instances of inaccurate grammar and sentences that could be written more clearly. I suggest that with further proofreading, this manuscript could be edited to become much easier to comprehend.

A.C, author's comments: For helping the reader's understanding: We have modified the subsection names from the Method section. Moreover, we have modified the set-up subsection for the reader's understandability and the result section is shorter for the Finnish experiment.

A.C Changes in the manuscript: New subsections titles: 2.1 Initial TEB model / 2.2 Implementation of road ice / 2.3 Improvement of the road snow processes with a multi-layer scheme

R.C 2, comment N°2: remove first instance of 'water'

A.C, author's comments: Okay, thank you for this comment.

R.C 2, comment N°3: Need to distinguish that M is the melting rate of ice as opposed to the melting rate of snow (Rmelt) which is also used in that equation.

A.C, author's comments: We have modified the text accordingly.

A.C Changes in the manuscript: **(l184)** With F, M the freezing and melting rates for ice

R.C 2, comment N°4: Line 162 (eqn. 9): Is d_{r1} the same as d_{R1} used in eqn. 2? If so, these need to be the same symbol, if not, need to specify what d_{r1} refers to.

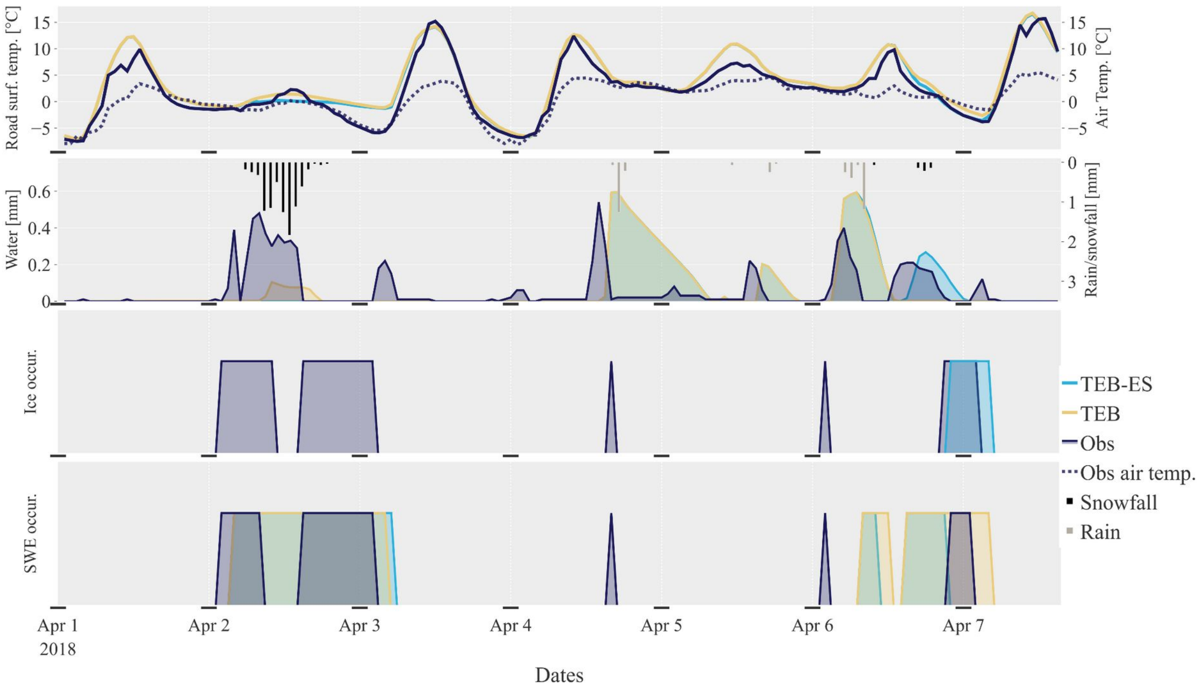
A.C, author's comments: We have modified the symbol in the eqn 9. It is indeed the same symbol of the eqn. 2 representing the road layer depth. Fixed as dR1.

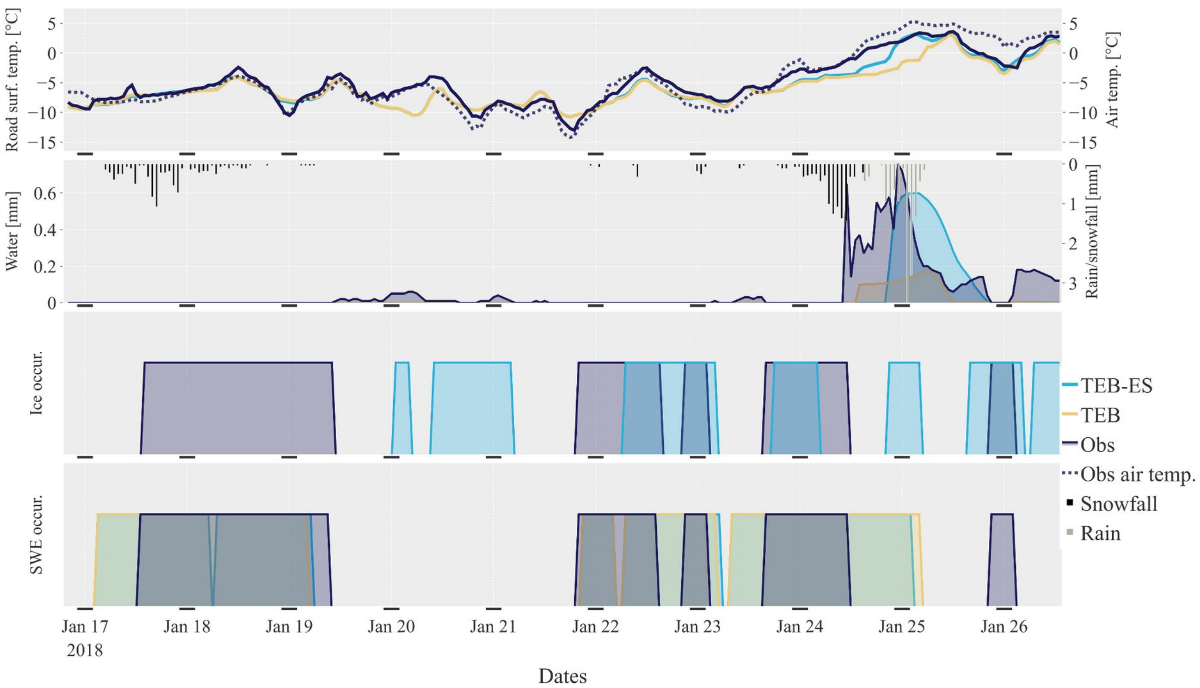
R.C 2, comment N°5: Line 342: Do you mean TEB-ES is more accurate than ISBA-Route/CROCUS rather than TEB? TEB-ES outperforms ISBA on most statistical measures in table 2 but TEB does not.

A.C, author's comments. This comment is an interpretation of the results and has its place on the Discussions. However, similar comments can be found in the discussion section, so this sentence has been just removed.

R.C 2, comment N°6: Figures 5 and 8: Increase font size

A.C, author's comments. Thank you for this comment, we have increased the font by 6 for the ticks, the legend and the x-axis. We have also let it visible the Apr 1 2018 on the x-axis for the reader's understandability.





R.C 2, comment N°7: The figure caption for figure 9 is very difficult to comprehend, especially the last sentence.

A.C, author's comments: We have removed the figure 9. Indeed, this figure was not interesting enough and led to only 2 line comments in the result section. For clarity and not have much added value for the results.

References :

-Jennings, K.S., Winchell, T.S., Livneh, B. *et al.* Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere. *Nat Commun* 9, 1148 (2018). https://doi.org/10.1038/s41467-018-03629-7

– **Douville H, Royer JF, Mahfouf JF :**A new snow parameterization for the Météo-France climate model, Part I : Validation in stand-alone experiments. Climate Dyn 12:21-35, 1995a

- **Boone, A., Masson, V., Meyers, T., and Noilhan, J**.: The Influence of the Inclusion of Soil Freezing on Simulations by a Soil–Vegetation–Atmosphere Transfer Scheme, J Appl Meteorol Clim, 39, 1544–1569, https://doi.org/10.1175/1520 0450(2000)039<1544:TIOTIO>2.0.CO;2, 2000.

- **Bouilloud, L. and Martin,** E.: A Coupled Model to Simulate Snow Behavior on Roads, J Appl Meteorol Clim, 45, 500 – 516, https://doi.org/10.1175/JAM2350.1, 2006

- **Lipson, M. J., Grimmond, S., Best, M., Abramowitz, G., Coutts, A., Tapper, N., Baik, J.-J., Beyers, M., Blunn, L., Boussetta, S., Bou-Zeid,E., De Kauwe, M. G., de Munck, C., Demuzere, M., Fatichi, S., Fortuniak, K., Han, B.-S., Hendry, M. A., Kikegawa, Y., Kondo, H.,Lee, D.-I., Lee, S.-H., Lemonsu, A., Machado, T., Manoli, G., Martilli, A., Masson, V., McNorton, J., Meili, N., Meyer, D., Nice, K. A.,Oleson, K. W., Park, S.-B., Roth, M., Schoetter, R., Simón-Moral, A., Steeneveld, G.-J., Sun, T., Takane, Y., Thatcher, M., Tsiringakis, A.,Varentsov, M., Wang, C., Wang, Z.-H., and Pitman, A. J**.: Evaluation of 30 urban land surface models in the Urban-PLUMBER project: Phase 1 results, Quarterly Journal of the Royal Meteorological Society, 150, 126–169, https://doi.org/10.1002/qj.4589, 2024

- **T. Kobayashi, T. Sato, K. Kosugi, S. Mochizuki and A. Sato** : Consolidation, Melting and Refreezing Process of Snow on Roads by Vehicles, SIRWEC, 2006