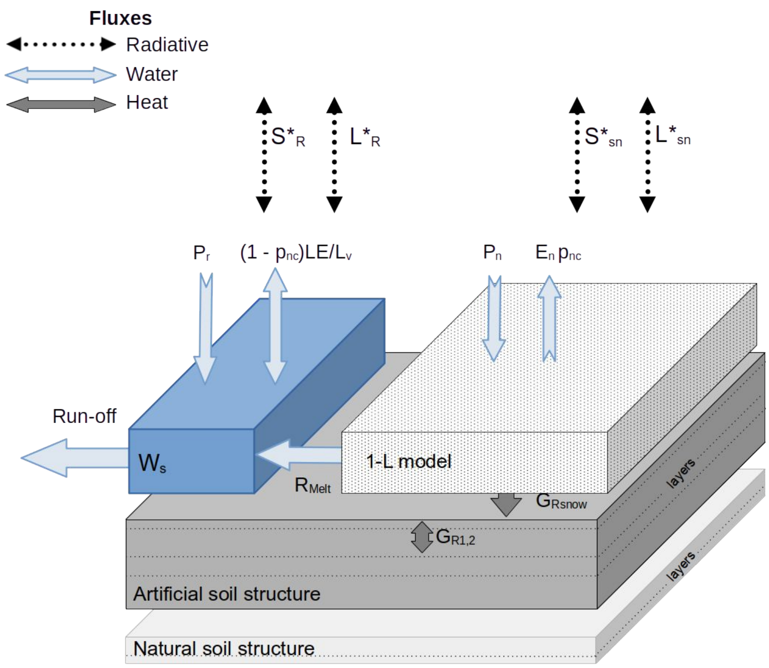
Responses to the Second Reviewer

Major Comment 1: 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 will add a simple MLR model on the

Major Comment 2: We acknowledge the confusion, we will modify and improve the methodology section to clarify the experimental setup. On the Finnish site, we use 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.

Major Comment 3: 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.



Minor Comment 1: Thank you for this useful comment, we will add this result to the abstract.

Minor Comment 2: After further checking, thanks to the reviewers' comments, we spotted errors in both the text and the code of the ice modelling. It leads to minor modifications on the results. We will modify the manuscript in agreement with the following modifications:

Fixed equation:

$$F = \frac{1}{\tau}\min(W_s, \frac{\max(O, T_f - T_{road})}{C_l L_f})$$
$$M = \frac{1}{\tau}\min(W_i, \frac{\max(O, T_{road} - T_f)}{C_l L_f})$$
$$LE^* = \gamma_{ice}\rho_a \frac{1}{R_a L_s}[Q_{sati}(T) - Q_a]$$

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 1mm. 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 suficient. 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.

Minor Comment 3: It would indeed be interesting to add a scheme of the original version of TEB before its modifications to compare. We will add a summary scheme of the processes modeled on the original version of TEB V9.0 surface in the paper or in the appendix.

Comment 4: We appreciate the comment, we will add the 2004 reference which is indeed more relevant in this context.

Comment 5: Yes, this argument is based on a reference (Lynch-Stieglitz 1994; Sun et al. 1999).

Comment 6: 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.

Comment 7: The wear as defined in the RoadSurf user manual (Karsisto and Kangas 2023) is a factor that decrease the amount of snow, ice, deposit and water and transform snow to ice. It is caused by traffic. What we named the water wear-off is all the processes like traffic wear, gravity wear from sloping roads or infiltration that lead to decrease the amount of water on the road.

The hydro option within the TEB model, model the soil water infiltration into the sewerage network. It allows to make an analogy with the water wear-off. The water amount decrease on the road depends on the water amount on the road and a time constant to set. This is approximately a process of the same order and we use it to simulate the water wear-off. We kept the default values of hydro option within TEB for the Finnish site.

Comment 8: We 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 will not talk about the six roads built but simply about the one we used.

Comment 9: We can only quantify the errors with the technical manual from the sensor manufacturers. However, theses values are not enough to explain the sensors large errors.

Comment 11: These small increases in snow height 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.

Comment 12: 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.

Comment 13: 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 neagtives

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

Comment 14: We will modify the plot to clarify the snowpack liquid water amount. In fact, we will not consider the absolute snowpack liquid water amount but only the occurrence.

Comment 15: We will modify the sentence. But as we will also modify the way of representing the liquid water content in the graph (reduced to an occurrence), this will also modify the context of the sentence.

Comment 16: As suggested, we will add the air temperature on all the result figures.

Comment 17: Thank you for this comment, we will separate the first line and put it elsewhere with the rest of the paragraph that does not talk about the same thing.

Comment 18: 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 confusion, we will try to clarify. Similarly, for the reanalyses, we will clarify, they are indeed not observations.

Comment 19: The Zenodo does not seem to be corrupted. We do not have any problem to download or unzip the dataset. If the problem persists on your side, please let us know. However, we will upload a new version, as some minor results will change because of the small ice modeling mistake.

References :

- Karsisto, V. and Lovén, L.: Verification of Road Surface Temperature Forecasts Assimilating Data from Mobile Sensors, Weather and Forecasting, 34, 539 – 558, https://doi.org/10.1175/WAF-D-18-0167.1, 2019.

-M. Lynch-Stieglitz. The development and validation of a simple snow model for the giss gcm. J. Clim., 7(12):1842–1855, December 1994.

-**Sun, S., J. Jin, and Y. Xue, 1999**: A simple snow-atmosphere-soil transfer (SAST) model. J. of Geophys. Res., 104, D16, 19587–19579