Point-by-point Responses to two Referees for # hess-2021-164

Referee 1 for # hess-2021-164

The authors have used regular fonts for the Referee's comments and blue fonts for our responses and red fonts with quotation marks to show the revised text.

This manuscript focuses on a very important subject --- how to improve the model in better predicting soil temperature at the soil surface layer. The topic is well within the scientific scopes of HESS because soil temperature data are critical in many research areas, such as meteorological, hydrological and ecosystem modeling, agricultural, soil and plant studies, and so on. The paper itself is generally well written and structured. Although the proposed method and model were developed for and tested in the region of U.S. Great Plains only, they can be easily applied to other regions in the world. The conclusions were sound and based on their data and figures. I have made some comments that may help authors in improving their manuscript. I strongly recommend that this manuscript should be accepted and published on HESS after the following comments are addressed.

Response: Thank you for your review and insight, which improved our paper. We responded to all of your comments.

Major comments:

1. On page 7 authors stated "the current empirical model was selected because it was more accurate than the process-based model". Then on page 14 (lines 261-262), they chose not to show the evidence to support the above statement. Theoretically, the energy balance models are physically sound and should predict the soil temperature more accurately if they are used properly. If no evidence or explanation is provided in this manuscript, it is difficult to convince readers that the statement on page 7 is true. If the paper length is a concern, authors at least should offer some explanations in the discussion session why the process-based models (energy balance models) failed in their studied region, or in other words, why their proposed model was a better choice than the energy balance model in this region.

Response: We appreciate your insights and suggested revisions. The authors initially studied both an empirical model and a surface energy-balance model (EBM, Chalhoub et al., 2017) for all selected stations in this study. Fig. R1.1 below presents the modeling results for both EBM and EM02 when compared to our improved model (iEM02). When we consider the paper's length and its focus, we decided not to include the energy balance model. We described the reason why we excluded the EBM modeling results in our original manuscript (lines 59-62).

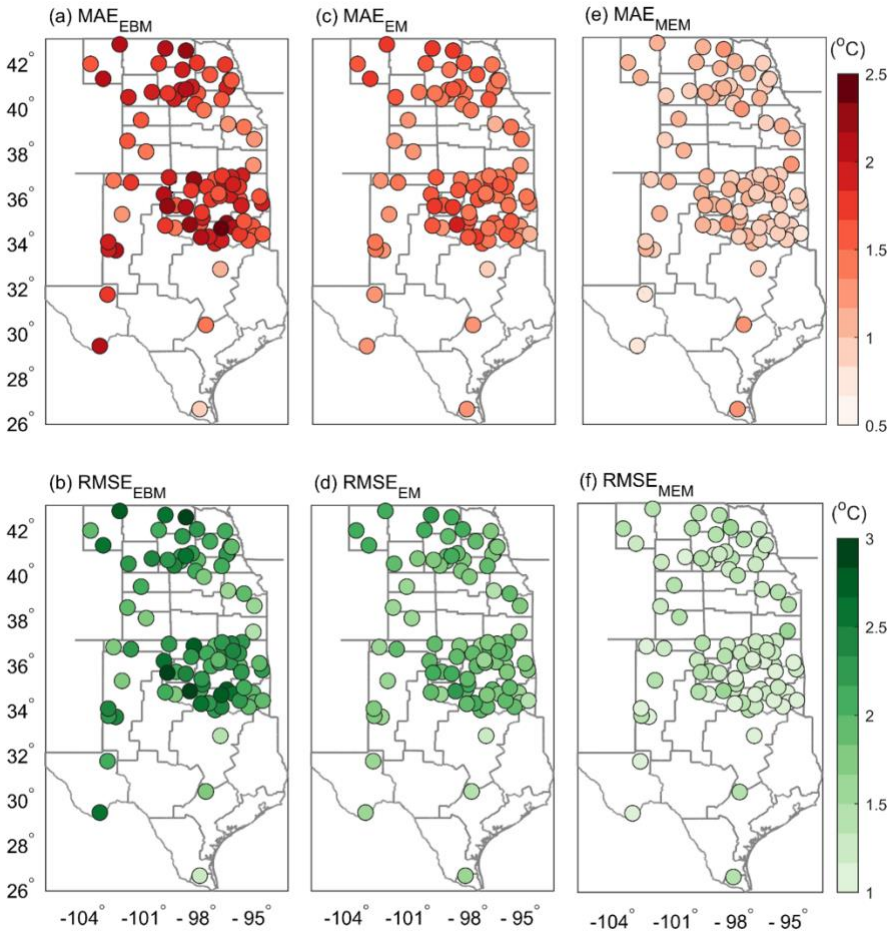

Figure R1.1: Spatial distribution of mean absolute error (MAE) (a, c, and e) and RMSE (b, d, and f) for the energy balance model (Chalhoub et al., 2017) (EBM, a, b), the empirical model (EM02, c, d), and an improved empirical model (iEM02, e, f), respectively. The colorbar defines values of MAE (°C) and RMSE (°C).

2. Section 2.2.2: authors may want to provide more rationales or explanations in designing their improved model. They introduced several new parameters to the model to represent the physical processes related to soil temperature. Some of those can be easily understood, such as soil thermal conductivity, thermal diffusivity and snow depth while the reasons for others were not very obvious. For example, why was it necessary to create a fictive environmental temperature (Tenv, °C) in replacement of air temperature (Ta)? What was the reason to include Tenv from an extra prior day (j-3)? Their results in Figure 4 did not show that this extra inclusion led to a significant improvement of model outputs.

Response: Thanks for these useful suggestions. We agree. We added and modified sentences in our revision to clarify T_{env} and $T_{a,\ j-3}$. The modeling improvement by using the fictive environmental temperature (T_{env}) was significant in northern areas of NE and KS (See Fig. 3b). For example, an approximately 15% improvement in simulated soil temperature was achieved in NE and KS when T_a was replaced by T_{env}. We have modified the text to explain the importance of T_{env} in the model, as follows:

"where a fictive environmental temperature (T_{env}) is assumed to be the weighted mean of air temperature (T_a) at 2 m and surface temperature (T_{sfc}). β is a partitioning coefficient, which defines the relative weight of the air temperature. This weighted fictive temperature will help weigh surface cooling and heating processes due to radiative and convective process (Dolschak et al., 2015)."

The temperature $T_{env,\ j-3}$ is added because soil texture changes across our study area and soil temperature usually has a longer memory.

3. Section 2.3: authors evaluated the model by quantifying the errors (RMSE and MAE) or the deviations of magnitude between the model outputs and observations. Another important characteristic of daily mean soil temperature is its seasonal cycle. This can also be an important metric for model validation (whether model outputs are in phase with the observations). One way of doing this can be to test the correlation (or lagged correlation) between the model outputs and observations.

Response: Thank you for your suggestions. We analyzed the linear correlation coefficient between predicted and observed soil temperature (Fig. R1.2). The linear correlation is less robust than the RMSE metric in this application.

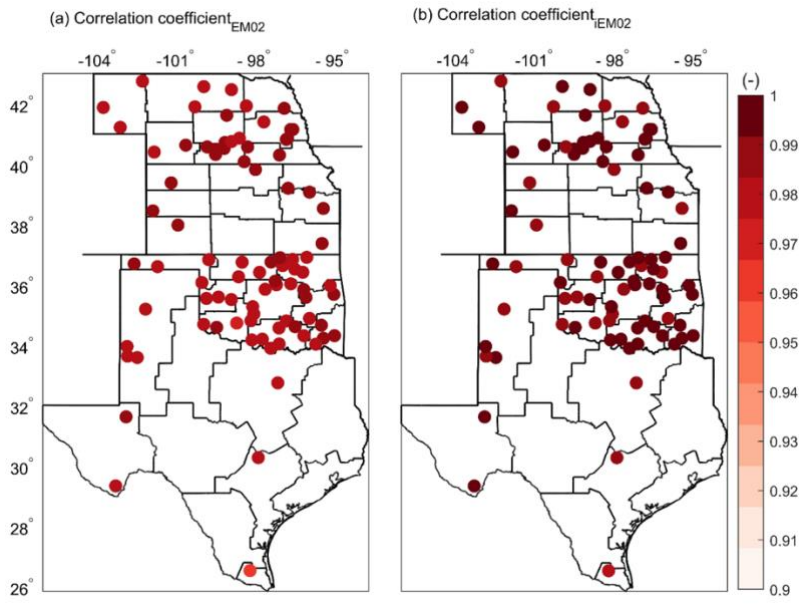

Figure R1.2: Spatial distribution of the correlation coefficient between predicted and observed soil temperature an empirical model (EM02, a) and the improved modelEM02 (iEM02, b). The colorbar defines the correlation coefficient (-).

Minor comments:
1. Line 62 on page 4: 'Kutikoff et al., 2021' and 'Dhungel et al., 2021' were not found in the reference list.

Response: Thank you. We have added these two references.

"Kutikoff, S., Lin, X., Evett, S. R., Gowda, P., Brauer, D., Moorhead, J., Marek, G., Colaizzi, P., Aiken, R., Xu, L. K., and Owensby, C. Water vapor density and turbulent fluxes from three generations of infrared gas analyzers. Atmospheric Measurement Techniques, 14(2), 1253-1266, 2021."

"Dhungel, R., Aiken, R., Evett, S. R., Colaizzi, P. D., Marek, G., Moorhead, J. E., Baumhardt, R. L., Brauer, D., Kutikoff, S., and Lin, X. Energy imbalance and evapotranspiration hysteresis under an advective environment: Evidence from lysimeter, eddy covariance, and energy balance modelling. Geophysical Research Letters, 2021."

2. Lines 112-117 on pages 6 and 7: the writing in this part can be confusing and should be rewritten in the following format --- 'In this study, three surface climate datasets were obtained: (1) …; (2)…; and (3)…' or 'In this study, three surface climate datasets were obtained. The first one … . The second one … . The third one …'.

Response: Thank you for the suggestion. We have adjusted the sentences in our revision, as shown below:

"In this study, three surface climate datasets were obtained from: (1) Automated Weather Data Networks (AWDN) (https://hprcc.unl.edu/awdn/), commissioned in the 1980s for Nebraska and Kansas; (2) The Oklahoma Mesonet, which is a daily climate data source for Oklahoma and, which started in the 1990s (http://www.mesonet.org/); and (3) the Soil and Climate Analysis Network, which gives daily climate observations (https://www.wcc.nrcs.usda.gov/scan/) that we selected for Texas due to limited quality data available in its automated weather station network."

3. Is it more proper to refer to Figure 2 somewhere between line# 167 and 177 (pages 9 and 10) instead of line# 148 on page 8?

Response: Agreed. We added this.

4. Lines 210-213 on page 11: How can soil types at different locations explain the different levels of modeling improvement by including Ta, j-3? Please be more specific.

Response: Thank you for your careful review. Thermal conductivity of clay and silt soil is lower than that of the sand soil. In our manuscript, we described the importance of soil types in Lines 212-213. Here we slightly modified them as below:

"The soil types in OK are more clay and silt compared to NE and KS (Fig. 1). However, the improvement by using the fictive environmental temperature was significant in northern areas of NE and KS (sandy soil) but not in the southern area of OK and part of TX (clay and silt soil) (Fig. 3b)."

5. Lines 240-251 on page 13 and Figure 5: first, you stated 'the α0 term was removed (Fig.5, a-d)' in the text but in caption of Figure 5 you stated '(a-d) with vs. without α4'. This is inconsistent (α0

vs. α4). Second, in Figure 5 shown on the x- and y-axes were simply the difference (absolute error) between the simulated and the observed data. However, in both the text and figure caption you used phrases RMSE and absolute mean errors (neither is a correct description of Figure 5). Please correct them thoroughly.

Response: Thank you for your careful review. First, we replaced α_4 in the caption by α_0. Second, the RMSE was calculated individually in completed and reduced models. We simply displayed this RMSE ratio in Figure 5 to contain both absolute error and RMSE. To clarify, we modified the caption for Figure 5 in our revision as follow:

"Figure 5: One-to-one plots of absolute mean errors between the complete model (EM02) and reduced model where one independent variable term was removed in the improved empirical model (iEM02): (**a-d**) with vs. without α_0 in Nebraska (NE), Kansas (KS), Oklahoma (OK), and Texas (TX), respectively, (**e-h**) with vs. without b_1 in Nebraska (NE), Kansas (KS), Oklahoma (OK), and Texas (TX), respectively; (**i-l**) with vs. without d_1 in NE, KS, OK, and TX, respectively. The ratio of the root mean square error (RMSE) shown includes the reduced RMSE and RMSE complete that refer to root mean square error for reduced and complete models, respectively. The colorbar indicates the number of observed data points."

6. Lines 242-243 on page 13: 'indicating a 20% drop in RMSE' --- did you mean 'a 20% raise'?

Response: Thank you for your review. Yes, you are correct. It indicates a 20% increase. We replaced "drop" with "raise" in our revision.

7. Line 268 on page 14: "Nebraska and Oklahoma Similar results" --- insert a period (.) between "Nebraska and Oklahoma" and "Similar results"

Response: Done.

8. Figures 8 and 9: with the current settings in these two figures, those lines (EM02, iEM02 and Obs) almost overlaid on each other and the difference between them was hardly identified. This reduced the values of these plots. Consider re-doing them or converting them into tables.

Response: To compare modeling performance clearly, we calculated the RMSE for three sub-growing seasons: October-November, December-January-February, and March-April-May-June. The values are shown in Table R1.1. We added these RMSE values inside the revised Figures 8 and 9. We revised the Figure 8's caption, as below:

"Figure 8: Daily soil temperature comparison between observed (grey line), complete model (EM02, green line), and improved model (iEM02, blue line) in western (>100°W), central (between 97° and 100°W), and eastern (<97°W) Nebraska (a-c) and Kansas (d-f) during the winter wheat growing seasons from 2015 to 2019. RMSE is the root mean square error (°C). Three values in brackets refer to the RMSE for the periods of October-November, December-January-February, and March-April-May-June. Shaded areas indicate the winter season (December-February)."

Table R1.1. RMSE values we added in revised Figures 8 and 9.

Regions	Model performance	
	EM02	**iEM02**
NE_west	RMSE = [0.9 1.1 1.1]	RMSE = [0.5 0.7 0.8]
NE_central	RMSE = [1.0 0.9 0.9]	RMSE = [0.6 0.6 0.5]
NE_east	RMSE = [1.1 1.1 1.1]	RMSE = [0.7 0.7 0.6]
KS_west	RMSE = [1.5 1.2 1.0]	RMSE = [1.1 0.9 0.7]
KS_central	RMSE = [1.6 1.8 2.1]	RMSE = [1.4 1.5 1.3]
KS_east	RMSE = [1.6 1.2 0.8]	RMSE = [1.2 1.0 0.5]
OK_west	RMSE = [1.2 1.3 1.4]	RMSE = [0.8 0.8 0.9]
OK_central	RMSE = [1.2 0.8 0.7]	RMSE = [0.7 0.5 0.4]
OK_east	RMSE = [1.1 0.7 0.6]	RMSE = [0.6 0.4 0.4]
TX_west	RMSE = [0.8 0.7 0.6]	RMSE = [0.5 0.6 0.5]
TX_central	RMSE = [0.7 1.1 1.2]	RMSE = [0.8 1.0 0.9]
TX_east	RMSE = [1.0 1.4 1.6]	RMSE = [0.9 1.4 1.3]

--- The END of point-by-point response for referee #1