Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2022-125-RC1, 2022

Comment on hess-2022-125

Anonymous Referee #1

Referee comment on "Accuracy of five ground heat flux empirical simulation methods in the surface-energy-balance-based remote-sensing evapotranspiration models" by Zhaofei Liu, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-125-RC1, 2022

Remote sensing-based surface energy balance typically requires G simulation to close the surface energy balance, which is often a challenge given that G could not be easily sensed from the surface. Hence, most remote sensing-based ET models use an empirical approach to scale G between the two extreme limits of % or fraction of G/Rn within the open surface and full canopy. This % or fraction G/Rn is characterized by vegetation and remotely sensed indices like NDVI, LAI, albedo, LST, etc using simple empirically derived values. This paper aims to study the spatiotemporal variations of this empirical relationship (G, Rn, H) and evaluate some of the remote sensing-based empirical methods using half-hourly global flux observations data. While, I think it is important to improve remote sensing approaches to simulate G, as it will also improve remote sensing and surface energy balance-based ET models, the results and discussion, as presented in the paper, are a little challenging to follow with not many insights into how G simulation in remote sensing-based ET models could be improved. So, I have some major issues (and some minor issues) that the author needs to consider before the paper can be reevaluated.

Major Comments

- The paper is more focused on the assessment of Rn and G relationships than the evaluation of simulated G within the existing remote sensing-based ET models. So I wonder if this should be reflected in the title of the paper, which suggests that the paper is focused on the evaluation of the existing methods. Note that the empirical nature of G simulations and their uncertainty in remote sensing-based ET models is a well-known issue. So while the optimization of regression coefficients (e.g., those in the LC methods) is nice, the finding that the coefficients differ across different parts of the world is obvious. What is more important is to present an idea about how this empiricism and uncertainty can be reduced and a globally applicable model could be developed. The paper falls short on this part.

- The paper acknowledges the limitation of existing G derivation methods in remote sensing-based ET models but does not consider some of the widely used approaches, such as the one used in the SEBAL model, which would require albedo and LST. The author acknowledged that SEBAL based G method was found to be working better than other approaches in another study (Saadi et al. 2018). The G models evaluated in this paper are very similar in nature. Hence, it is important to incorporate G models with different structures/inputs. Note that obtaining albedo and LST for these sites is as easy as obtaining NDVI. The author should have incorporated some additional G derivation methods used in the common remote sensing-based ET model.
- The author mentioned that observed G is taken as the residual of the energy balance to evaluate different G models, assuming that all other components are perfectly derived. While the author acknowledges this in section 4.1 (Line 364-365), I think still problematic because no attempt has been made to address this issue. Here, there is no information on how the energy balance was closed (or was not unclosed) or corrected. The observed G used in this paper and all error metrics presented hence could be highly biased and uncertain.
- Note that typically in a remote sensing-based ET model, Rn is calculated using radiation balance using remote sensing and meteorological inputs, and G is estimated as a fraction of Rn. Hence, the uncertainty in Rn calculation is also a source of error in G. In some cases when G may be biased available energy (Rn-G, where Rn is coming from remote sensing-based radiation balance) may be reasonable. In this paper, the author used observed Rn in calibrating G, so when you compare coefficients, the uncertainty in Rn (even better when remote sensing-based Rn is used) needs to be mentioned too. Given that Rn is the key input used in all G methods considered in this study, additional assumptions (assuming that Rn is perfectly simulated by the remote sensing-based ET model) and uncertainties need to be discussed.
- It is not clear how the coefficients of the LC methods are calibrated in this study. Are these just the regression coefficients or other optimization methods used? Was any calibration/validation approach used (using independent sets of data)?
- I am surprised why the author did not test the actual LC methods (i.e., the original coefficients) used in different ET models considered in this study. In addition, it is important to mention how these different ET models come up with different empirical coefficients.
- I find no difference between the contents in the abstract and the conclusion. Both summarize key results with no discussion on the key reasons for differences in model performances and insights into how future remote sensing-based G models can be improved. I couldn’t find the main objective of the paper in the abstract.

Minor comments:

Line 7: Instead of saying “According to 230 flux site observations” better say Based on the assessment from 230….

Line 8-9: Based on the previous statement, it shows that G accounts for a significant proportion of the daily surface energy balance.

Line 19: It’s not the accuracy of the sites. It’s rather the accuracy of the models in these sites.

Line 31-42: It’s better to differentiate “ground heat flux” or “soil heat flux” by providing their physical meanings and with more detailed descriptions. The author defines soil heat flux as the heat flux measured by the flux plates near the surface.

Line 74-75: Suggest citing Roerink et al., 2000 and Merlin et al., 2014 right after the corresponding model names

Lines 101-110: Given the numbers of towers from different networks, could you please indicate how you came up with the number “230” (i.e., 230 sites used in this study).

Line 270-272: I do not think you can say NSE is suitable but RE and KGE for evaluation. Yet, you are using RE, RMSE, and KGE for model evaluation. Maybe you need to rephrase the sentence. It is better to justify the choice of model evaluation metrics in the methods section.

Line 340: Please mention the optimization process in the Methods section

Line 329: How can daily G be simulated at 6:30? Shouldn’t this be G only or half-hourly G?

Line 383: MODIS is not used at 10:30 and 13:30. MODIS data represents conditions around these times.

Line 393-398: redundant information in the paper

Line 416-417: These data are easy to get. It may not be a good idea to ignore Bastiaanssen (1995) Method when it was found to be working better than other approaches in another study (Saadi et al. 2018).

Line 420: The difference among different methods was not significant because NDVI and fc are highly correlated (in fact NDVI is likely used to derive fc) and they are calibrated similarly.

Line 430: there may be a case when a large error in G may be canceled by a large error in

Rn leading to reasonable estimates of available energy (Rn-G), which is further partitioned into sensible and latent heat fluxes.