

Review of paper by Senay et al.: Evaluating the SSEBop approach for evapotranspiration mapping with landsat data using lysimetric observations in the semi-arid Texas High Plains.

General comments:

The paper applies the SSEBop approach (presented and applied to MODIS data in Senay et al., 2013) to 14 Landsat-5 thermal infrared images for estimation of daily evapotranspiration, and the results are evaluated using daily large-scale lysimeter measurements established in four plots. The approach appears interesting due to a pixelwise estimation of cold and warm temperature boundary conditions. However, it is surprising that a pixelwise estimation of $(R_n - G)$ is not included (it is not described how R_n is estimated), and another serious weakness is that the atmospheric resistance between source height (the surface where T_s is measured) and air (r_{ah}) is given a fixed value completely ignoring impacts of spatio-temporal variations in weather, land cover, wetness and/or T_s . In reality, r_{ah} is highly dynamic and difficult to estimate. Furthermore, the choice of value for the crop coefficient k is not explained (in theory, it varies throughout the season due to LAI variation), and data sources for atmospheric correction are not presented. These are potential sources of bias error and should be clearly assessed and described.

Overall SSEBop results show a bias error, and this bias should be further examined, discussed and, if possible, reduced. The bias error quantifies a systematic error of the SSEBop approach, however the authors choose to correct this error empirically by simply fitting SSEBop calculated ET to ET observations, thereby increasing SSEBop calculated ET empirically by 12 %. This is however an unacceptable solution, and this part of the analysis must be completely removed from the paper. Instead, the authors should try to identify the reasons for the systematic error which may be related to the simplified representations of $(R_n - G)$, r_{ah} and k .

The authors aggregate ET over different periods and conclude that the bias error contribution to total error is increasing. However, this is a misleading (and unfavorable!) conclusion. Mean bias error in percentage of mean observation remains constant irrespective of the time period considered (this is illustrated in the below table) whereas the total error decreases. The reason for higher bias error contribution to total error (rmse or mse) is therefore that total error decreases. The reason for total error to decrease can be that the ranges of variability (in data and estimations) reduce due to aggregation. To further examine this, I advice the authors to assess the relative rmse ($rmse/mean_obs$ or $mse/mean_obs$) in relation to the coefficients of variance (standard deviation divided by mean) for data and estimations. If relative rmse exceeds the coefficient of variation for observations, then the SSEBop estimations would be no better than the mean of ET observations. The coefficients of variation should be reported for the different aggregation levels for both observations and estimations.

Example: MBE/mean_obs remains constant irrespective of aggregation:

Day	Obs	estimate	MBE	obs-2day	est-2day	MBE-2day
1	5	4	1			
2	6	5	1	11	9	2
3	7	6	1			
4	8	7	1	15	13	2
mean:	6.5	5.5	1	13	11	2

Daily: MBE/mean=1/6.5= 0.153846

2-Day MBE/mean=2/13= 0.153846

The calculation of R^2 quantifies how well SSEBop estimates the observed ET variability, though I don't think that the small differences in R^2 for the different aggregation periods are statistically significant (all correlations are statistically significant). Therefore, it seems that the aggregation may not play any role at all for the accuracy of the SSEBop approach. This would also be strange since the inputs for error analysis are the same, but summarized over different periods. This can however be further assessed by comparing the coefficients of variance and relative rmse's, as recommended above.

Specific comments:

- 1) p. 726, l. 11: replace "exit" by "exists"
- 2) p. 726. Section 1.1 refers to the application and performance of the SSEBop and SSEB approaches, as presented in previous studies. Please include a short explanation of how these two methods work (and how they are different) before summarizing on the results.
- 3) p. 727, l. 25-forward: what types of crops are being cultivated on the lysimeter fields in the study period?
- 4) p.728, l. 25. A value for k equal to 1.25 is chosen and a reference is given to Allen et al. (2011a), however the choice of this parameter setting should be explained: Does it represent a typical crop grown in the area, or how was the value chosen? Normally, a seasonal variation in k is expected due seasonal variations in LAI and related variables. Does the chosen k represent the maximum k? Please clarify. The reference to Allen et al. (2011a) should also be included or corrected. It is not in the reference list.
- 5) p. 729, equation 3. This is a nice approach, however theoretically R_n should be replaced by (R_n-G) , otherwise dT will be overestimated. Please consider to correct this or provide a comment/explanation about this assumption. It is also necessary to explain how R_n is estimated. Another serious weakness is the fixed value of r_{ah} for a bare and dry soil surface. A more realistic (dynamic) estimation of r_{ah} should be used or at least the uncertainty caused by the assumption of fixed r_{ah} should be assessed, eg. what is the sensitivity of dT to a realistic range of variation in r_{ah} . I expect that it is very high.

6) p. 730, l. 5. I assume that the reason for using $T_{a(max)}$ instead of T_a is that $T_{a(max)}$ is the only available gridded weather dataset. Please include this information, or explain why the hourly weather data (used for calculating ET_0) are not used instead of $T_{a(max)}$.

7) p. 730, l.20-24. Earlier it was informed that ET_0 , T_a and T_s are the only data needed for application of SSEBop. However, since ET_0 is calculated from hourly weather data (including wind speed and air humidity?), the data needed for the ET_0 calculation using REF-ET should also be described. Please include this information.

8) p. 731, l.19-20. Information should be included about the data sources and method used for atmospheric correction of Landsat-5 data. This information and the corrections are important when working with satellite time series data.

9) p. 734, l. 4. Insert “mean” before “square of the random error”

10) p. 735, 9. I don't understand the sentence “These dT values do not change from year to year for a given day of year”. Is dT fixed for a certain day of year?? As already remarked above, it was not explained how R_n was estimated (used for calculating dT in equation 3). This is needed. Accurate estimates of both R_n , G and r_{ah} should be important because they impact dT and thus the SSEBop estimated evapotranspiration.

11) p. 736, l. 1-2. It is surprising to read that the pixelwise estimates of the cold and warm boundary temperatures are similar for the four lysimeter fields even though two fields are irrigated and two fields are dryland crop fields. It is to be expected that albedo and surface temperature differ between at least the dryland and irrigated plots, and that this should be reflected by differences in R_n and thus also by the warm boundary temperature. Please explain why this is not the case.

12) p. 736, l. 19. Observed T_s exceeds the warm boundary temperature in the early part of the season, and the discussion here should refer to the various simplifications of the SSEBop method in order to explain possible reasons for this problem: 1) Information is missing on how R_n is calculated/measured for dT calculation (Eq. 3), but it seems that it does not consider spatial variations in albedo (line 23) – why not calculate R_n from the satellite data using albedo and T_s ? I also do not understand why there is no attempt to estimate the soil heat flux. It is better to include an (uncertain) soil heat flux estimate than completely ignoring it (though it would tend to further reduce the calculated upper boundary temperature in this case). My recommendation is to include pixelwise estimates of (R_n -G).

Another weakness to be discussed (or improved) is that the atmospheric resistance r_{ah} is represented by a fixed value. If weather data exists for estimating r_{ah} on this day, these should be used to assess whether too high r_{ah} could be a reason for the upper boundary temperature being lower than T_s . Atmospheric correction of Landsat could be another source of error (information is missing on data sources for atmospheric correction). Furthermore, the calculated dT applies a simple “big-leaf” approximation whereas the land surface (soil and vegetation) would be better represented by a two-source approach in the beginning of the growing season where LAI is low. Overall, reasons for underestimation of the upper boundary temperature should be assessed and discussed in more detail.

13) p. 737, l. 10. The underestimation of ET by SSEBop should also be discussed in relation to the simplifications of the SSEBop approach. ET is underestimated despite the use of a very high k coefficient. Why?

14) p. 737, l. 20 . “squared mean” should be inserted before “bias errors” to be precise.

15) p. 738, l. 4-6. The ET bias error is corrected by simply linearly fitting the estimated ET to observed ET (thereby increasing model ET by 12 %), and then error statistics are calculated again. However, the model bias quantifies a systematic error of SSEBop that should at least be thoroughly examined and discussed. It makes no sense at all to empirically fit the calculations to observed ET. This is unacceptable, and the analyses related to the empirical fitting of model results must be removed from the paper. Instead the authors should examine reasons for the bias error and, if possible, try to improve the SSEBop approach to give better results.

16) p. 738, l. 11-15. It is misleading to refer to increasing bias errors contributions due to aggregating. Mean bias error in percentage of mean observed ET remains the same irrespective of aggregation, but total error reduces, perhaps due to a lower range of variability in observations and estimations. This can be quantified and examined by calculating the coefficient of variance for both observations and estimations. If the rmse exceeds the coefficient of variance (sd/mean) for observations, then the SSEBop results are no better than the mean of observations. How does sd/mean vary for the different aggregation periods. Another comment is that higher bias error is a disadvantageous result that should be avoided. Bias error indicates a problem with the approach whereas random error may theoretically be reduced by increasing the number of observations.

17) p.739, l. 17-18. I disagree. This is not a scientifically defensible approach. The fitting of final ET model estimates to data should be deleted from the paper.

18) p. 748, Table 3. Numbers in brackets of the two last rows are not “errors expressed in percent of observed ETa” (as stated), but they represent percentages of MSE. Please correct. It is also stated that “MSE, MBE² and MSE_e are variances in units of mm²”, but they are “squared errors in units of mm²”. This should also be corrected.

19) p.749, Table 4. Same comments as for Table 3.

20) p. 750, Table 5. This table should be completely removed from results and discussion. The authors remove bias error by empirical adjustment of SSEBop results using a regression slope factor. Instead, the authors should test or at least discuss how to improve the SSEBop method in order to reduce or avoid bias errors.

21) p. 751, Fig. 1. Instead of showing the center location of lysimeters with dots, please indicate the area extent of these large lysimeters.