Ref.: Dr. No. ESSD-2020-345

Title: Long time series of daily evapotranspiration in China based on the SEBAL model and multisource images and validation

Author: Minghan Cheng, Xiyun Jiao, Binbin Li, Xun Yu, Mingchao Shao, Xiuliang Jin

Research Paper

Earth System Science Data

Cover letter

Dear Editor and Reviewers

I am submitting here a manuscript entitled "*Long time series of daily evapotranspiration in China based on the SEBAL model and multisource images and validation*". We submitted this manuscript in November 2020. Three reviewers gave us good advices. First we would like to thank the reviewers for their constructive and helpful suggestions and improvements to our manuscript (**ESSD-2020-345**). We revised the manuscript by following the suggestions of the reviewers. Our response to each suggestion or comment are given one by one in the following Pages of this letter. For details, please refer to the responses as follows (Reviewer comments are in black font, responses are in blue or red font)

Looking forward to your favorable decision.

Thanks too much.

With best regards,

Minghan Cheng and co-authors

Responses to Reviewers

Reviewer 2#:

The authors have addressed most of my previous comments (mostly in the methods section). A few further modification/classifications suggested in the revised version:

1. Line 399: .....caused the latent heat flux underestimation?

Response: Thank you for your help and suggestions in improving our manuscript. It has been revised (**Line 404, Page 21**).

2. Line 410-411: if the assumption of constant Bowen ratio was not correct, the authors must have used the correct/suitable method for closing the energy balance. Please provide more clarification or rephrase the sentence.

Response: Thank you for your help and suggestions in improving our manuscript. In this study, the eddy covariance system measured value was filtered and corrected. First, the data with Energy Balance Closure Ratio (ECR, Eq. 2) less than 80% were not selected for validation, and then, the remaining data with ECR more than 80% were corrected by using Bowen Ratio energy balance correction (Eq. 3) (**Lines 152-156, Page 7**).

Reviewer 3#:

General comments:

The manuscript introduces a newly generated ET data set with 1km spatial resolution and daily temporal resolution over China based on the SEBAL model. Given that this manuscript is a contribution to ESSD, with a focus on the newly provided data product, I am missing information on and/or discussion of critical issues such as the selection of the extreme pixel values for the SEBAL model, the amount of missing data in the LST time series, the derivation of ET from the water balance. The comparison with MOD16 is a bit lengthy, and much of the information could be put together in tables, rather than listing all performance scores for each vegetation class, terrain class etc in the running text. Since I am not a native speaker myself, I do not comment on language at all, but the manuscript needs rigorous English proofreading.

Response: First we would like to thank the reviewers for their constructive and helpful

suggestions and improvements to our manuscript. Our response to each suggestion or comment are given one by one in the following Pages of this letter. MOD16 products is one of widely used ET dataset, so we selected it as a typical for comparing with the generated SEBAL ET product in this paper. If the SEBAL ET showed a comparable performance with MOD16 or even better, which could indicate the SEBAL ET have an acceptable accuracy and can be used for related studies.

Major comments:

1. The selection of the pixels that define the extreme hot and cold conditions is a critical step in the application of the SEBAL model. In the Appendix the authors describe their routine for the extreme pixel selection: they select a single hot and cold pixel over the MODIS scene. I am missing a discussion on the justification of this approach; given that a single MODIS scene covers an area of 1200 x 1200 km, with differences in elevation, weather conditions etc., I don't think that two extreme pixels coming maybe from points very far apart from each other could be related to each other in a reasonable way. Because in the SEBAL method it is assumed that changes in LST are mainly due to the evaporative cooling effect, rather than elevation variation, shadows etc. I would be interested to see an analysis showing the sensitivity of the extreme LST pixels to different selection methods.

Response: Thank you for your help and suggestions in improving our manuscript. The hot/cold pixel selection method in this study was referred to Long et al. (2011), which has been referred in many studies. Domain size (defined as the actual size of the modeling domain/satellite imagery being used) is an important effect for hot/cold pixel selection, in the study of Long et al. (2011), this issue has been deeply discussed. In this study, we further compared the different domain sizes' performance in ET estimation (**Fig. 12 Lines 475, Page 24**), overall, the domain size employed in this study (1200 km × 1200 km) performed an acceptable accuracy (**Lines 457-475, Pages 23-24**).

2. If I understand the authors correctly, they calculate a yearly water balance ET for nine primary water resources divisions. My first question here would be whether they selected hydrological years or calendar years? Second, from my own experience with ET derived from the water balance, an averaging period of a year is not enough to ensure the assumption of ignorable storage changes. However, I have no experience with such large basins. Figure 10 shows that the variation of ETwb

is quite significant for some of the basins. I would therefore encourage the authors to discuss their approach and its implications.

Response: Thank you for your help and suggestions in improving our manuscript. We used average ET_{WB} over multiple years instead of in one year for regional-scale validation (**Lines 320-327, Page 17**). ΔS over multiple years can be ignored (Liu et al., 2016; Senay et al., 2011) (**Lines 174-176, Page 8**). The ET_{WB} was calculated using observed data recorded in calendar years.

3. LST data availability is often a major limitation when applying LST-based ET algorithms. The statement 'it should be noted that there are several missing or unreliable pixels in MODIS images' is a bit vague, in my opinion. I would prefer some quantification of the share of valid to invalid pixel values, e.g. in the form of a percentage of valid data points in the time series per pixel, or a table with similar information further categorized into seasons, etc. The authors apply a very simple data imputation method and it would be interesting for the reading to know how much of the modelled ET values are based on these interpolated LST data. Even a flag in the data set could be considered.

Response: Thank you for your help and suggestions in improving our manuscript. We supplemented two figures to describe the ratio of interpolated pixels of land surface temperature (MOD11) data (**Lines 640-643, Page 31**). Fig. A2a describes the time series of interpolated pixels per month over 2001-2018 and Fig.A2b describes the histogram of ratio of interpolated pixels.

4. For the comparison with the EC data, it would be interesting to also include the other energy balance components, sensible and ground heat flux and net radiation.

Response: Thank you for your help and suggestions in improving our manuscript. The source of ET estimation errors is a subject worthy of further study. In this paper, we referred to previous studies to further discussed this question: *'Besides sensible heat flux, the errors of SEBAL ET may derived from net radiation or soil heat flux as well (Li et al., 2017; Teixeira et al., 2009). For net radiation, which is computed using surface albedo and Stephen Boltzmann law (Eq. 2 in appendix), generally performed a relatively agreement with flux tower observed value, while soil heat flux, which computed using empirical formula related to net radiation and NDVI (Eq. 9 in appendix), has a poor performance (Li et al., 2017; Song et al., 2016). In the study of Li et al.*

(2017), soil heat flux estimation showed a clear overestimation in higher ET area, e.g., wetland, which may further cause the sensible and latent heat flux underestimation in higher ET rates. In the most SEB-based algorithms, the similar net radiation and soil heat flux estimation methods are used, and various sensible heat flux estimation methods are the main sources of the difference among the various SEB-based algorithms. However, the causes of the net radiation and soil heat flux estimation errors have not been clearly discussed, e.g., the effect of satellite transmitted time or land cover types. These issues could be the focus of our follow-up research, for example, high frequency geostationary satellite and flux tower observations may be helpful for this research.' (**Section 4.3.2, Page 24, Lines 476 - 486**). However, due to the limited of flux tower data, we could not study the instantaneous energy component in this paper, and this object will be conducted in our follow-up research. Thank you for your help and good suggestions in improving our manuscript.

5. Given that only eight EC towers are available for model evaluation, I am wondering whether a differentiation of model performance according to land cover (three types) but even more climate zones (five zones) and terrain classes (four classes) makes sense. I am not sure how well suited the available data are to draw general conclusions on the performance in the different climate zones, etc.

Response: Thank you for your help and suggestions in improving our manuscript. EC flux tower has been proved that could be used for regional ET validation, even better than most of other approaches (Wang et al., 2012). For example, Hu et al. (2015) used 15 flux towers to validate the performance of MOD16 and LSA-SAF MSG evapotranspiration products over Europe; Aguilar et al. (2018) used five flux towers to validate MOD16 ET product over Northwestern Mexico; Ramoelo et al. (2014) used two flux towers to validate MOD16 ET products over parts area of South Africa; Yang et al. (2017) used eight towers to validate GLEAM ET products over China; Li et al. used 12 towers to validate GLEAM and GLDAS ET products over China; Kim et al. (2012) used 20 towers to validate MOD16 products over Asia. In general, the density of flux towers in this study (eight towers in China) is comparable with previous studies, even better than some of them.

Moreover, the performance assessment of the product under different conditions (climate zones,

ecosystems and terrain) could make the results more comprehensive (Velpuri et al., 2013), in this study, the eight towers basically cover most of climate zones, ecosystems and terrain in China, and the observed period is long-time-series (each tower both have more than 1000 samples with total available samples of 9896). Therefore, the validation of MOD16 in this study is convincingness to some extent.

References:

Wang, K. and Dickinson, R.E., 2012. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. Reviews of Geophysics, 50(2).

Hu, G., Jia, L. and Menenti, M., 2015. Comparison of MOD16 and LSA-SAF MSG evapotranspiration products over Europe for 2011. Remote Sensing of Environment, 156: 510-526

Aguilar, A. et al., 2018. Performance Assessment of MOD16 in Evapotranspiration Evaluation in Northwestern Mexico. Water, 10(7).

Ramoelo, A. et al., 2014. Validation of Global Evapotranspiration Product (MOD16) using Flux Tower Data in the African Savanna, South Africa. Remote Sensing, 6(8).

Yang, X., Yong, B., Ren, L., Zhang, Y. and Long, D., 2017. Multi-scale validation of GLEAM evapotranspiration products over China via ChinaFLUX ET measurements. International Journal of Remote Sensing.

Kim, H.W., Hwang, K., Mu, Q., Lee, S.O. and Choi, M., 2012. Validation of MODIS 16 global terrestrial evapotranspiration products in various climates and land cover types in Asia. KSCE Journal of Civil Engineering, 16(2).

Velpuri, N.M., Senay, G.B., Singh, R.K., Bohms, S., & Verdin, J.P., 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. Remote Sensing of Environment.

Additional comments:

1. P. 2, line 38: what are traditional methods in this context? Remote sensing models rely on very traditional approaches (Penman-Monteith or surface energy balance residual models are very

traditional approaches.）

Response: Thank you for your help and suggestions in improving our manuscript. The traditional methods indicate the methods based on point-scale or small-area-scale analysis, such as lysimeter and eddy covariance. In order to make it, this sentence was rephased: "*However, the methods for the estimation of ET based on point-scale or small-area-scale analysis, such as lysimeter and eddy covariance, cannot meet the requirement of global climate change research and regional water resource management*" (**Page 2, Lines 38 - 40**).

2. P. 3, line 63: In my opinion the classification of ET models into SEB and SEF models is a bit subjective. The Penman-Monteith equation e.g. is also physically-based (as SEB approaches).

Response: Thank you for your help and suggestions in improving our manuscript. In the revised version of manuscript, we divided into three types according to their mechanism: those based on surface energy balance residual (SEBR), those based on semi-empirical formulas (SEFs) and statistic methods by referring the paper of Wang et al. (2012) and Zhang et al. (2016). The P-M and P-T equation, which is partly physical, were divided into semi-empirical formula method. Moreover, we add a category - statistic methods to avoid confusion with SEF-based method (**Pages 2-3, Lines 47 - 66**).

3. P. 3, line 69: The authors state that the temporal resolution of eight days is not sufficient for search on water resources management? How do the authors come to this conclusion and what temporal resolution would be sufficient?

Response: Thank you for your help and suggestions in improving our manuscript. It indicates that higher temporal resolution could use for finer water resources management, e.g., irrigation regime making. In order to avoid misunderstanding, we deleted this sentence (**Page 3, Line 71**).

4. P. 5, line 125: see major comment 3

Response: Thank you for your help and suggestions in improving our manuscript. We supplemented an appendix to show the ratio of interpolated pixels of land surface temperature (MOD11) data (**Pages 6, Lines 133-135 in new version**). The details were described in the response of major comment 3.

5. P. 6, line 141: The authors state that the EC method measures ET using the covariance between vapor and heat fluxes. This is wrong! The EC method measures the covariance of the vertical wind velocity (!) and concentration of the entity of interest.

Response: Thank you for your help and suggestions in improving our manuscript. We have rewritten this sentence by referring the paper of Wang et al. (2012): "*The eddy covariance method measures λET from the covariance of the heat and moisture fluxes, respectively, with vertical velocity using rapid response sensors at frequencies typically equal to or greater than 10 Hz*" (**Page 7, Lines 147- 148**).

6. P. 7, line 163: see major comment 2)

Response: Thank you for your help and suggestions in improving our manuscript. We used average ET_{WB} over multiple years instead of in one year for regional-scale validation (**Page 17, Lines 320-326**). ΔS over multiple years can be ignored (Liu et al., 2016; Senay et al., 2011) (**Page 8, Lines 174-176**).

7. P. 8, line 180: RMSE is not suited to describe model bias.

Response: Thank you for your help and suggestions in improving our manuscript. RMSE is the most indicator to describe the model accuracy. We rephased this sentence by replacing "*bias*" with "*the performance of the model*" (**Page 8, Line 188**).

8. P. 9, line 225: The authors conclude that 'ET_SEBAL is relatively reliable for daily-scale Application'. I am wondering how they justify that statement and if they have some references to define what a relatively reliable model performance is.

Response: Thank you for your help and suggestions in improving our manuscript. We deleted this sentence in this location (**Page 10, Line 230**), and added it in Section 4.1: "*Overall, the SEBAL ET showed an acceptable performance in China by comparing previous studies.*" (**Page 20, Line 374**)

9. P. 11, line 240: see major comment 5)

Response: Thank you for your help and suggestions in improving our manuscript. In this study,

the eight towers basically cover most of climate zones, ecosystems and terrain in China, and the observed period is long-time-series (each tower both have more than 1000 samples with total available samples of 9896). Therefore, the validation of MOD16 in this study is convincingness to some extent. The details were described in the response of major comment 5.

10. P.21, line 390: The authors state that a decrease in surface temperature corresponds to a reduced evapotranspiration. I think this needs rephrasing because in general low surface temperature at similar meteorological forcing would indicate that more of the available energy is dissipated via ET than sensible heat flux.

Response: Thank you for your help and suggestions in improving our manuscript. We rephased this sentence to make it clearer: "*Moreover, it should be noted, due to the decrease of surface available radiation energy which was caused by cloud cover, the ET (both actual and modeled value) is also less than that of nearest date (Cheng et al., 2020)*" (**Page 21, Lines 394-395**)

11. P. 22, line 428: I am wondering why the authors decided for the described upscaling method, if they explain in this section why other methods would be preferable.

Response: Thank you for your help and suggestions in improving our manuscript. In this study, we would like to discuss the sources of ET estimation errors, upscaling method may be one of the sources of errors. The upscaling method which was used in this study (constant evaporative fraction) will cause a negative bias of 10–20% in the estimation of daily ET (Delogu et al., 2012; Ryu et al., 2012; Van Niel et al., 2012), be that as it may, this method has also been widely used. Moreover, although other methods have been proposed, however, they also have a certain error and uncertainties (Gentine et al. 2007). (**Pages 22-23, Lines 434-445**)

12. P. 23, line 444: see major comment 4)

Response: Thank you for your help and suggestions in improving our manuscript. Due to the limited of flux tower data, we could not study the instantaneous energy component in this paper, and this object will be conducted in our follow-up research. The details were described in the response of major comment 4.

13. P. 23, line 454: What do the authors mean by ‘a low domain size’ in this sentence?

Response: Thank you for your help and suggestions in improving our manuscript. The domain size is defined here as the actual size of the modeling domain/satellite imagery being used. (**Page 23, Lines 457-458**)

14. P. 26, line 538: the use of the arrows for indicating up- and downwelling radiation is inconsistent between equations and the text.

Response: Thank you for your help and suggestions in improving our manuscript. We modified them to R_{s_down}, R_{l_up}, and R_{l_down}. (**Page 27, Lines 550-556,**)

15. P. 27: line equation (12) is true for neutral conditions only.

Response: Thank you for your help and suggestions in improving our manuscript. The computation of r_a is modified based on Eq. 30, when it is not neutral conditions, the Eq. 30 will be adjusted by Monin–Obkhov length which could judge whether it is stable or unstable conditions. The Eq. 12 was used at the first time of the loop only, as the neutral conditions was assumed. The Eqs. 19 – 30 could describe this process (**Pages 29-30, Lines 610-626**).

16. P. 28, line 580: see major comment 1)

Response: Thank you for your help and suggestions in improving our manuscript. We supplemented parts of content to compare the different domain sizes’ performance in ET estimation (**Fig. 12, Page 24, Lines 475**). The details were described in the response of major comment 1.

17. P. 28: some of the equations (20) to (28) are redundant.

Response: Thank you for your help and suggestions in improving our manuscript. Eqs. 20-30 clearly describe the process of r_a computation at different conditions (neutral, stable or unstable), so we prefer to remain it.

18. Zenodo homepage: The authors state ‘The products were evaluated using the eight flux towers observation data for point validation and water balance method for regional validation and showed

R value of 0.79 and 0.88, respectively, which indicated the products have a great Performance'. In my opinion a rRMSE of > 40 % might maybe not indicate great performance.

Response: Thank you for your help and suggestions in improving our manuscript. As described in Section 4.1, SEBAL ET showed a comparable performance in China with previous studies, and it is better than MOD16 products. "great performance" may not very suit, so this sentence was changed to "*The products were evaluated using the eight flux towers observation data for point validation and water balance method for regional validation and showed R value of 0.79 and 0.98, respectively, which indicated the products have a better performance than MOD16 products which have been widely used*". In general, the RMSE (0.92 mm/d for point scale and 48.99 mm/year for regional scale) and rRMSE (42.04% for point scale and 13.57% for regional scale) of SEBAL ET are an acceptable accuracy at current, which due to the validation methods still have errors, e.g., EC tower have a typical error of 5-20% in ET observation, be that as it may, EC tower is still a widely used in situ measuring method.

19. Zenodo homepage: the coordinate system should be stated in the text

Response: Thank you for your help and suggestions in improving our manuscript. The coordinate system is GCS_WGS_1984, and have added in the statement of Zenodo homepage:

"*The dataset named SEBAL evapotranspiration in China (SEBAL ET) characterized the daily evapotranspiration (in millimeter) of vegetation in China from 2001 to 2018, the spatial resolution is 1 km × 1 km and the temporal resolution is 1 day with the coordinate system of GCS_WGS_1984. The products were generated using Surface Energy Balance Algorithm of Land (SEBAL) and multi-sources remote sensing data, including MOD43A1 daily surface albedo, MOD11A1 daily surface temperature and MOD13 vegetation indices (obtained from NASA: https://ladsweb.modaps.eosdis.nasa.gov/search/), the meteorological data obtained from GMAO (https://gmao.gsfc.nasa.gov/research/highlights/2013-2015.php), the input variables were all aggregated of resampled to 1 km × 1km. The products were evaluated using the eight flux towers observation data for point validation and water balance method for regional validation and showed R value of 0.79 and 0.98, respectively, which indicated the products have a better performance than MOD16 products which have been widely used. SEBAL ET can be used for several geoscience studies, especially for global change, water resources management and*

agricultural drought monitoring, etc.”