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.

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 1200x1200 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.

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.

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.

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.

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.

Additional comments:
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.

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).

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?

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

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

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

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

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.

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

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.

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.

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

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

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

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

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

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

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.

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