Response to Referee # 2

This manuscript provides an evaluation of STIC1.2 in estimating actual evapotranspiration at the spatial scale by combining the model with satellite remote sensing data. In addition, the authors also compare the performance of STIC1.2 with other two existing remote sensing algorithm (i.e., SEBS and MOD16). In general, the topic of this MS is of interest to the HESS' readership and the manuscript is well written. However, there are several major issues in this study, which introduce additional uncertainties and preclude a focused evaluation of the models themselves (as described below). In this light, a MAJOR REVISION is needed.

We thank Referee #2 (R#2 hereafter) for finding our work interesting to the HESS community and providing useful criticism, comments, and suggestions. We also acknowledge R #2 for pointing out potential uncertainties associated with the use of input data. Uncertainties arising from the temporal and spatial mismatches of input datasets from different sources are common in remote sensing based studies. Acknowledging this fact, we made best possible efforts to minimize those. First, it is important to clarify that our study aims to provide a scientific basis for the operational use of STIC1.2 model towards estimating regional ET using remotely sensed data. Since the MODIS daily products suffer greatly due to cloud cover, the 8-day MODIS products are more applicable for regional ET model implementation and hence the validation was done at this temporal scale, which is a common practice (Ichii et al., 2009;Zhang et al., 2010;Yang et al., 2006;Xiong et al., 2015). The MOD11A2 product was considered to provide a better temporal and spatial representation (compared to the daily MODIS products) of regions (i.e. a wide range of biome and aridity conditions in the US). Nevertheless, we have now implemented STIC1.2 and SEBS at instantaneous (at point scale using daily MODIS and instantaneous weather data) and presented main validation results in this response. However, this implementation is not the core part of our paper and hence will only be added to the supplementary (1 figure and 1 table) and with a brief discussion in the main text (highlighted in red font, Page 6).

In this response, we not only justify our approach (e.g. use of MODIS 8-day LST) with additional analysis and references but also show results from the evaluation of instantaneous ET from SEBS and STIC1.2 models using instantaneous/daily MODIS products (and weather inputs of the same period). Results suggest that the core findings and conclusions of this study will remain the same regardless of whether instantaneous or 8-day products are used. However, additional knowledge on the potential sensitivity of SEBS to input meteorological forcing and the applicability of the STIC1.2 model across different time scales were gained. In the revised version, we propose to add additional descriptions (highlighted in red font, see pages 6 and 12) on potential uncertainties associated with the application of models at 8-day scales, use of 8-day average (daytime) weather inputs, and the MODIS 16 ET products. We believe that the evidence we provided and detailed answers to your concerns with appropriate references will sufficiently address R #2's major concerns that are mostly associated with the use of input data and uncertainties.

Major: 1. My largest criticism lies in the use of MOD11A2, where LST is reported as the average values of clear-sky LSTs during an 8-day period. As there is no information about

which day (or days) out of each 8-days contributes to the final 8-day averages, this 8-day average LST is highly likely to not correspond to the 8-day averages of meteorological variables (i.e., air temperature, VPD, etc). For example, the 8-day LST might only be a result of day-1 LST, or the average of day-1 and day-7 LSTs. Even they correspond well with each other, using 8-day averages may still lead to additional uncertainties due to differences in the temporal variability between, say, daily LST and air temperature. For example, H_day1+H_day2. . .+H_day8 does not equal to 1/8 * (H_calculated using 8-day average LST and meteorological inputs), as all the responses are non-linear. To focus on evaluating the model itself, it is recommended to work on the instantaneous scale rather than 8-day averages.

Response: The scientific basis for using MOD11A2 comes from an abundance of studies (for e.g. Ichii et al., 2009;Jin et al., 2011;Tian et al., 2013;Garcia et al., 2014;Xiong et al., 2015) that have also used this 8-day product in ET modeling. In our study, SEBS and STIC1.2 models were run at 8-day average scale corresponding to the MODIS daytime overpass time using MOD11A2 (and ancillary MODIS data; Table 2, Page 27) and 8-day average meteorological data corresponding at the MODIS Terra LST overview time (not the entire 8-day average; Page 9, Line 26-Page 10, Line 3 in the MC). The 8-day averaged meteorological data (that considers all hours) are only used in extrapolation of daytime ET to 8-day ET using a constant ET approach (Brutsaert and Sugita, 1992;Crago, 1996;Chávez et al., 2008). We evaluated 8-day cumulative ET to compare model performances against those from available MOD16 8-day ET products (Mu et al., 2011).We provide further justifications for using 8-day MODIS products as below:

Comparison of 8-day vs. daily LST (or T_R), air temperature (T_A), and $T_R - T_A$

We find that the 8-day average LST (or T_R), air temperature (T_A), and difference between T_R and T_A ($T_R - T_A$) were good representative of the corresponding instantaneous values during each of the 8-days within the corresponding MODIS 8-day period (R^2= 0.80-0.92, PBIAS within 3%) (Fig. AR1).

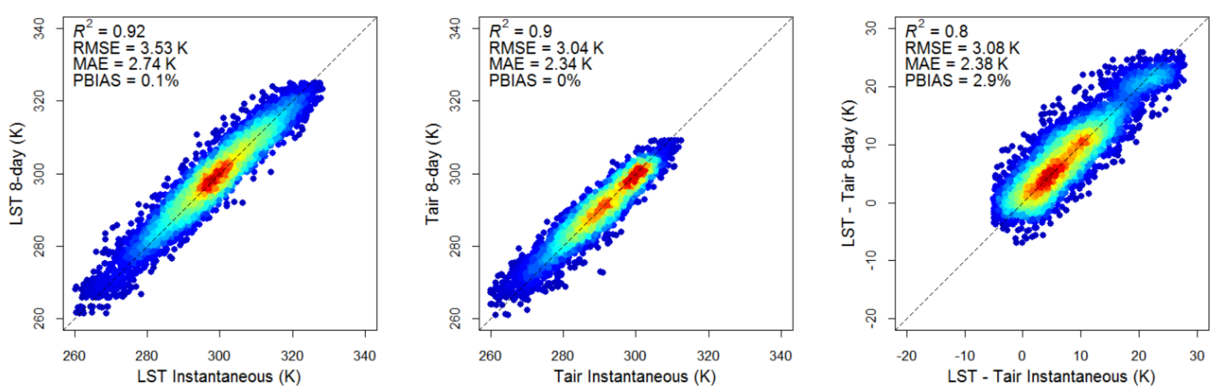

Fig AR1. Scatter plots of 8-day average LST vs. instantaneous LST, 8-day average daytime T_A vs. instantaneous T_A, and 8-day average daytime $T_R - T_A$ vs. instantaneous $T_R - T_A$.

Comparison of 8-day vs. daily ancillary meteorological variables

When 8-day vs. daily ancillary meteorological variables were compared, solar radiation (R^2 = 0.82, PBIAS = -5%) and relative humidity (R^2 = 0.78, PBIAS= 6%) were also found to be in a similar range (Table AR1). However, we noted that 8-day average daily wind speed, which is highly variable with time and space, was not well representative of daytime conditions (R^2 = 0.36). Note that this uncertainty in wind speed could affect instantaneous ET values from the SEBS model, which uses wind speed to determine sensible heat flux (H) and estimate latent heat flux (λE) as a residual of surface energy balance. In addition, this would also slightly affect 8-day average net radiation that used the FAO-based Penman-Monteith (PM) equation (Allen et al., 1998;ASCE-EWRI, 2005) that is used to upscale instantaneous ET to 8-day cumulative ET. Hence, we think the uncertainties associated with differences in actual meteorological conditions will have more of an effect on SEBS than STIC1.2, which is also supported by results from model evaluation at the instantaneous scale (i.e., MODIS TERRA daytime overpass time).

Table AR1. Comparison of 8-day average daytime meteorological and radiative inputs vs. instantaneous inputs to assess how representative the 8-day average values were of each day within the 8-day period.

Variable	R^2	RMSE	MAE	PBIAS (%)
T_R (K)	0.92	3.53	2.74	0.1
T_A (K)	0.900	3.04	2.34	0
T_R - T_A (K)	0.80	3.16	2.42	3.8
RH (%)	0.78	10	8	6
Wind speed (m s^{-1})	0.36	1.61	1.19	2
Incoming shortwave radiation (W m^{-2})	0.82	69	49	-5%

Model implementation and validation at instantaneous scale showed similar results

STIC1.2 and SEBS were implemented using all the available daily MODIS products (MOD11A1, MOD09GA, etc. and instantaneous weather data from NLDAS-2 forcing data[1]) during the study years (Fig. 3, page 32) and evaluated at the instantaneous scale. We noticed similar overestimation and slight underestimation tendencies of SEBS and STIC1.2, respectively (Table AR2 and Fig. AR2). Overall, model performances at the instantaneous scale during dry, normal, and wet years were also consistent with those observed when validation was done at the 8-day scale (see figure 4, Page 33 in the manuscript); there was a slight underestimation tendency of STIC1.2 and overestimation of SEBS (better under wetter conditions). The additional 11% overestimation of SEBS (17% vs. 28%) could be attributed to uncertainties associated with widely varied wind speed and other meteorological variables within the 8 days of a given 8-day period as well as the slightly overestimated 8-day average net radiation (Fig.AR3, also in the supplementary, Figure S1). The uncertainty associated with wind speed should not affect STIC1.2, which does not use wind speed; therefore, STIC1.2 performances were more consistent compared to SEBS. These results suggest that regardless of whether 8-day or daily MODIS products are used, the key findings of our research would largely remain the same.

[1] Note: This dataset is 12.5° (~12.5 km) dataset (https://ldas.gsfc.nasa.gov/nldas/NLDAS2forcing.php). The typo in Table2 (page 27) and Page 26, Line 10, where it was written as 4 km will be corrected in the revised version.

Table AR2.Evaluation of instantaneous ET and 8-day cumulative ET (Table 3, Page 28) from STIC1.2 and SEBS against observed ET from thirteen core AmeriFlux sites in the US combining data from one dry, one wet, and one normal year. Note: the 8-day ET estimates are derived from 8-day MODIS products (MOD11A2, MOD09A2, etc.) and 8-day average weather data (both at the satellite overview time and the 8-day average values; Table AR2).

Scales	Model	R^2	RMSE (mm hr^{-1} or mm 8-day^{-1})	MAE (mm) (mm hr^{-1} or mm 8-day^{-1})	PBIAS (%)
Instantaneous	STIC1.2	0.61	0.12	0.09	-5
	SEBS	0.53	0.14	0.10	17
8-day	STIC1.2	0.66	7.5	5.4	-3
	SEBS	0.53	9.8	7.3	28

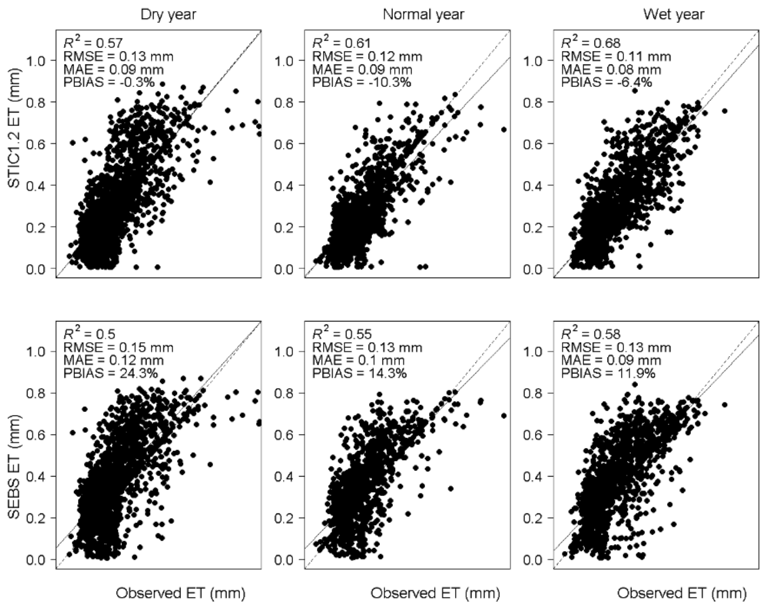

Fig. AR2. Evaluation of ET estimates from the STIC and SEBS models against flux data at the instantaneous scale using daily MODIS products (MOD11A1) during the dry, normal and wet years (see Fig. 4, Page 33).

<u>Good correlation between estimated and observed 8-day average available energy (ϕ)</u>

The 8-day mean (considering all hours) available energy (ϕ = R_n-G) that was used to upscale 8-day average daytime instantaneous ET to 8-day total ET (Eq. 17, page 10) was found to be strongly correlated with the observed 8-day mean ϕ at the flux towers. While the random noises in instantaneous ϕ (Figure AR3), which was also found to be well correlated with observed values (R^2 = 0.65), are removed in the 8-day average ϕ, a small positive bias (i.e. overestimation of 9%) (Fig. AR3, also in the supplementary, Figure S1) was also added. The residual difference in estimated 8-day average ϕ and observed 8-day average ϕ was positively correlated with the residual difference in estimated and observed 8-day cumulative ET (r = 0.27 for SEBS and r = 0.18 for STIC1.2, p-value < 0.001). This positive bias in estimated 8-day mean ϕ could have reduced biases from STIC1.2 from -5% to -3% (positive shift) and led to further increases in SEBS biases (i.e. 28%, Table AR2).

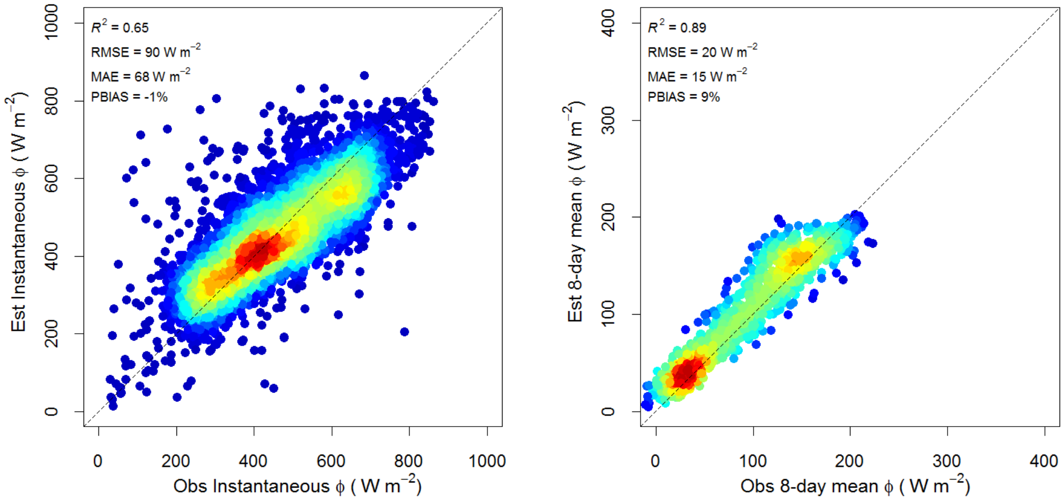

Fig. AR3. Scatter plot of estimated vs. observed available energy (ϕ) at instantaneous (using daily MODIS products) and 8-day averaged conditions (using 8-day products).

Models performed better in areas that are mostly affected by clouds

In the humid regions, all three models performed better than in the arid and semi-arid regions with similar accuracies despite high cloud cover (resulting in a fewer number of cloud-free days in each MODIS 8-day cycle). This is partly because vegetation (forests) is mostly energy-limited in this region and because estimated average 8-day ϕ from the meteorological inputs from the NLDAS-2 forcing (12.5°) were well correlated with observations (R^2=0.91) and with 9% error. In the arid sites, the estimated and observed ϕ relationship was also similar (R^2 = 0.96 and PBIAS within 6%). In our opinion, the effect of cloud cover is smaller in the arid and semi-arid regions compared to the humid and sub-humid regions and hence most of the differences in model performances could be attributed to the physical differences among the models.

MOD11A2 and aggregated meteorological are commonly used in ET modeling

The model implementation and validation scheme used in this study (i.e. use of MODIS aggregated datasets and *n*-day averaged meteorological inputs) have been applied in several other studies (Ichii et al., 2009;Senay et al., 2013;Wu et al., 2010;Xiong et al., 2016). In addition, comparison of daily ET, 8-day average ET or the 8-day with respective flux ET has become a common practice in ET model evaluation studies (Yang et al., 2006;Ichii et al., 2009;Senay et al., 2013;Biggs et al., 2016;Xiong et al., 2015;Ryu et al., 2012) and particularly those that use MODIS 8-day datasets.

Proposed changes in the revised manuscript

While we have provided evidence and justifications for using MOD11A2 and aggregated weather information, we do acknowledge that there are uncertainties associated with the use of 8-day average LST and aggregated meteorological variables which should be mentioned in the manuscript. We propose to add the following text (in section 4, after page 17, Line 5) in the revised manuscript and provide tables and figures in the supplementary (Table AR1-AR2 and Figs. AR1-AR3).

"One of the key sources of uncertainty in the implementation of STIC1.2 and SEBS at the 8-day timescale (using MOD11A2) could be the use of average 8-day daily time meteorological inputs that may not well correspond with LST observation days within each MODIS 8-day cycle. We found all 8-day daytime averaged meteorological variables (those used in SEBS or STIC1.2 models) except wind speed to be well representative of instantaneous measurements within the 8-day period (Supplementary Table S2[2]). This could be a source of additional uncertainty in SEBS since it uses wind speed to parameterize the aerodynamic conductance using MOST theory. Model implementation at an instantaneous scale (i.e. MODIS overpass time and using daily MODIS products including MOD11A1 datasets) showed that the performance of STIC1.2 (R^2 = 0.61, PBIAS = -5%) was similar to its performance at the 8-day scale. However, for SEBS (R^2 = 0.53) the performance was slightly better with a PBIAS of 17% (Table S3[3]). In addition to the wind speed, the slight overestimation of 8-day average ϕ (PBIAS = 9%), and variations in T_R, T_A, T_R -T_A, and other meteorological variables during days within the corresponding 8-day period could have added positive biases to SEBS (increase from 17% to 28%), when evaluated at the 8-day scale. SEBS is sensitive to the meteorological input especially the temperature gradient and its performance is expected to degrade with the use of gridded forcing data (Ershadi et al., 2013;McCabe et al., 2016;van der Kwast et al., 2009). Conversely, the overestimation in ϕ could have slightly reduced STIC biases (increase from -5% to -3%). Overall, the application of STIC1.2 and SEBS at the instantaneous scale showed similar model predictive capability and potential model strengths and weaknesses (e.g. better under wetter conditions). STIC1.2 appears to be slightly more consistent through time, which could be because the PM equation is designed to be applied at different time scales and STIC1.2 does not rely on wind speed to solve for G_A and G_C. Results also suggest that biases from SEBS could be kept well under 20% if uncertainties associated with meteorological and radiative forcing are reduced."

2. Page11 (Line 8-17) Again, validation should be carried out at an instantaneous scale but not daily (or 8-day averages), as upscaling can introduce additional uncertainties. In my opinion, upscaling from satellite overpass to longer time scales is another scientific question.

Response: For the most part, please refer to our response to the previous comment. We have added few more points below.

Should an ET model always be validated at the instantaneous scale only?

While we agree that the validation of ET at the instantaneous scale could help reduce uncertainties associated with upscaling and overall meteorological representation, we also disagree that the validation should always be conducted this way. ET is a hydrological process and like precipitation and runoff, these processes (and the errors) are perceived better when reported at daily or seasonal scales (e.g. 0.01 mm hr^{-1}vs 1 mm/day or 1000 mm $year^{-1}$). For

[2] Table AR1
[3] Table AR2

example, ET at daily or seasonal scales is more meaningful for hydrologists who manage water resources (Tang et al., 2015;Cammalleri et al., 2014;Colaizzi et al., 2006) and for comparing with accumulated precipitation (Baldocchi and Ryu, 2011). There are host of studies (for e.g.Fisher et al., 2008;Senay et al., 2013;Velpuri et al., 2013;Jiang and Ryu, 2016;Bunting et al., 2014) that have conducted ET validation at much larger temporal scales (e.g., 8-day, monthly, seasonal, annual) than the instantaneous scale.

The 8-day cumulative or mean ET corresponds with the cycle of MODIS global coverage (Masuoka et al., 1998) and one of the most widely used forms of ET model implementation and validation (Ryu et al., 2012). The 8-day or other multiday MODIS composites are designed to deal with cloud cover to provide a more routine measurement of Earth's surface than the daily MODIS data. The cloud effect on ET estimation or understanding other physical processes is greatly reduced when a composite 8-day LST product is used (Yang et al., 2013;Hu and Brunsell, 2013).

STIC1.2 has already been validated at an instantaneous scale

STIC1.2 model has been extensively validated at a half-hourly scale using flux tower data (Mallick et al., 2014;Mallick et al., 2015;Mallick et al., 2016), which is a better evaluation than using remotely sensed data, as any bias associated with the spatial and temporal mismatches between input meteorological and land surface variables are distinctly identified. The strength of the present study is to test the ET mapping potential of STIC1.2 at the regional scale using purely remote sensed data and gridded climate data. Therefore, we performed validation of the STIC1.2 model at the 8-day scale and compared our resulting ET estimates with readily available products such as MOD16 or the widely used SEBS model. In addition, our results are consistent with instantaneous scale validation (Table AR2 and Fig. AR2 in this response and Table 3, page 28 and Fig. 4, page 33 in original MS).

Upscaling errors (instantaneous to 8-day ET) are minimized through reliable estimates of 8-day average ϕ

We agree that the upscaling of ET from satellite overpass time to longer time scales (e.g. 8-day, as done in this study) is a different scientific question, as there are several uncertainties associated with it. The approach (Page 10, Equation 17) we used in this paper is a well-established method (Allen et al., 2007;Colaizzi et al., 2006;Ryu et al., 2012;Shuttleworth et al., 1989;Gentine et al., 2007;Chávez et al., 2008). In addition, the estimated 8-day ϕ (the key driving force of ET) was strongly correlated with the 8-day average ϕ at flux towers (R^2 =0.89, RMSE = 20 W m^{-2}, PBIAS = 9%, Fig. AR3, also in the supplementary, Figure S1). Hence, the errors associated with model evaluation at the 8-day scale by upscaling of instantaneous to 8-day cumulative ET should be within 9%. If we had directly used the observed ϕ at the towers, PBIAS from SEBS would have reduced to 18% (from 28%) and STIC biases would have been -10% (from -3%) with increased R^2 (STIC = 0.69 and SEBS= 0.58) and slightly reduced RMSEs (STIC= 6.8 mm and SEBS = 8.6 mm) from both models. However, it should be noted that the evaporative fraction was derived during the image time obtained using the weather information from gridded data, not the flux tower data itself. Hence there are some uncertainties with the use of meteorological data from multiple sources during instantaneous and multiple scales, as data from the same source its typically used to extrapolate instantaneous to daily or other scales (Allen et al., 2007;Chávez et al., 2008;Allen et al., 1998). Overall, we find that the upscaling errors are within 10% for both models.

Major Point 3. Page11 (L29-30): Any explanation of this model performance: overestimation in dry year and underestimation in wet years? Additionally, according to your Figure 6, it seems that this "overestimation in dry and underestimation in wet" pattern persists across sites (i.e., spatially). This may suggest some systematic uncertainty of the model. Given this, I do not agree with the statement given in Page 15 (Line 4-12). First, does any previous study support this wet/dry patches around the studied sites? If not, this is just your speculation. Second, the footprint issue could lead to random errors rather than a systematic underestimation. Finally, it is the authors' responsibility to ensure the footprint of a flux site corresponds (or encompasses) the MODIS footprint so that to eliminate data uncertainties and to allow a focused evaluation of the model.

Response: Overestimation of SEBS could be due to uncertainties associated with the kB^{-1} parameter (Chen et al., 2013), as well as the positive biases in 8-day average ϕ (as discussed earlier). Underestimation of ET from STIC1.2 could be due to an excessive moisture constraint applied during initialization of soil moisture availability (M) using T_R and dew point temperature at source/sink and reference heights. In addition, in the dry years, overestimation errors of ϕ (R^2 = 0.88, RMSE = 22 W m^{-2}, PBIAS = 12%) was slightly more than in the wet year (R^2 = 0.91, RMSE = 18 W m^{-2}, PBIAS = 9%). At the instantaneous scale, we noticed that STIC1.2 did not overestimate ET during the dry year and the SEBS overestimation was within 24%(PBIAS = 24%), which could be due to uncertain conductance parameterization, as in case of 8-day evaluation. We have discussed potential uncertainties in detail in section 4 (Figs. 7 and 13). The manuscript is about first ever regional scale implementation of the STIC1.2 model using remotely sensed data and hence have focused more on the initial validation as well as the comparison with two other commonly used models.

Please check figures AR4-AR7, where annual ET maps from three global ET products: 1) The Global Land Evaporation Amsterdam Model (GLEAM; 0.25 ° spatial resolution) (Martens et al., 2017;Miralles et al., 2011); 2) MPI-BGC or Fluxnet: MTE (0.5° spatial resolution) (Jung et al., 2010;Jung et al., 2011)3) SSEBOp (1km spatial resolution; https://earlywarning.usgs.gov/fews/datadownloads) (Senay et al., 2013;Velpuri et al., 2013) are added. Here, the annual ET map from one of three study years (the year when datasets from the other three models were also available) is shown. While the first two datasets (GLEAM and MPI) are at relatively coarser spatial resolution, most of these maps clearly show a similar spatial pattern of ET as STIC1.2. Hence, the spatial patterns of ET produced by our model seem to be reasonable and not linked with any systematic uncertainty of the model. In addition, scatter plots of estimated vs. observed ET (both instantaneous and 8-day cumulative; Figure AR2, Figure 3 and 4) show points spread uniformly around the 1:1 line.

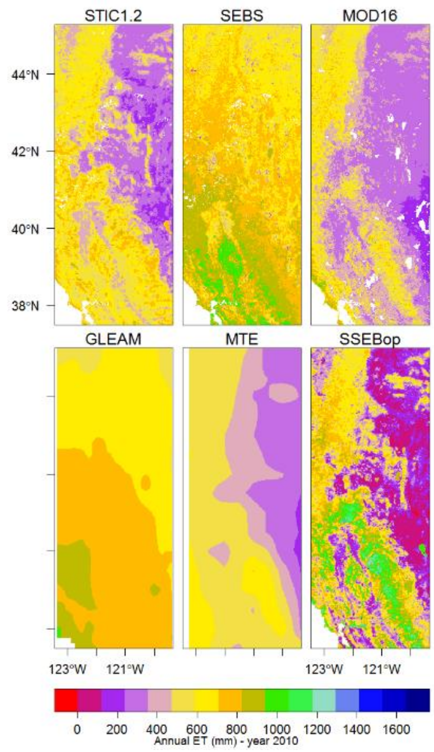

Figure AR4. Annual ET map derived from STIC1.2, SEBS, MOD16, GLEAM, MTE, and SSEBop for the western (W) bounding box covering US-Ton and US-Me2 flux sites (Fig. 2, Page 31).

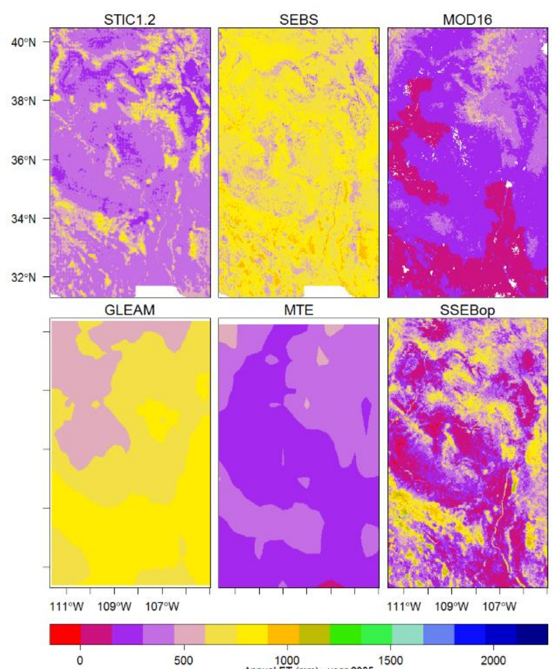

Figure AR5. Annual ET map derived from STIC1.2, SEBS, MOD16, GLEAM, MTE, and SSEBop for the mid-western 2 (MW2) bounding box covering US-ARM, US-SRG, US-Wkg, and US-NR1 flux sites (Fig. 2, Page 31).

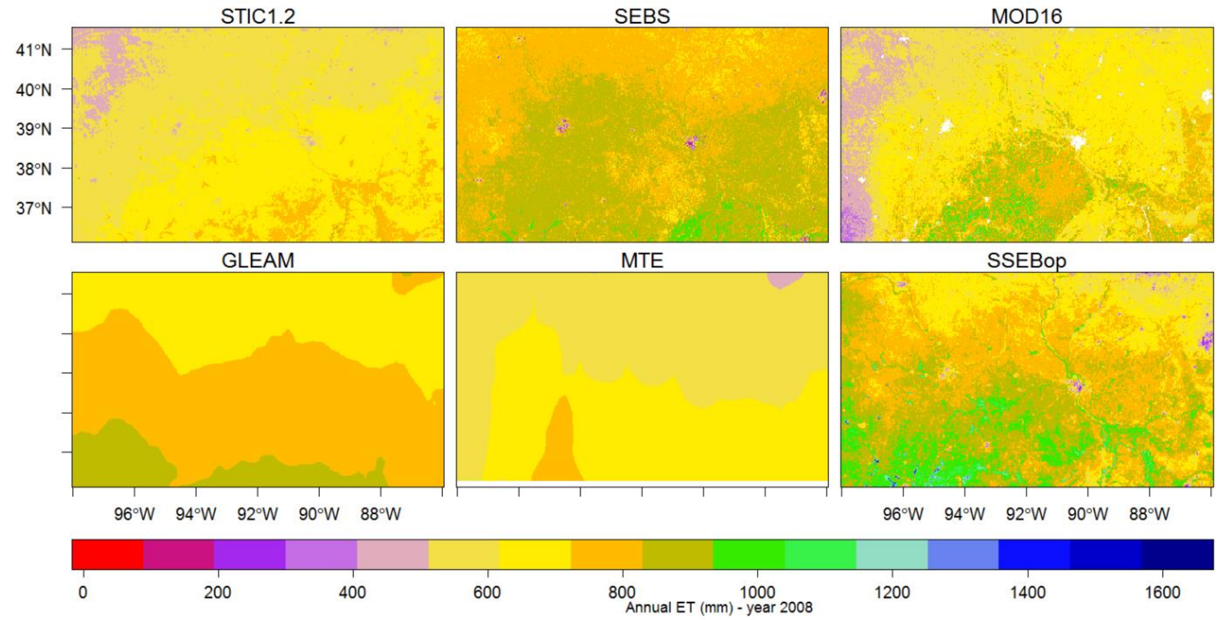

Figure AR6. Annual ET map derived from STIC1.2, SEBS, MOD16, GLEAM, MTE, and SSEBop for mid-western 1 (MW1) bounding box covering US-Kon, US-KFS, US-ARM, US-Ne1, and US-MMS flux sites (Fig. 2, Page 31).

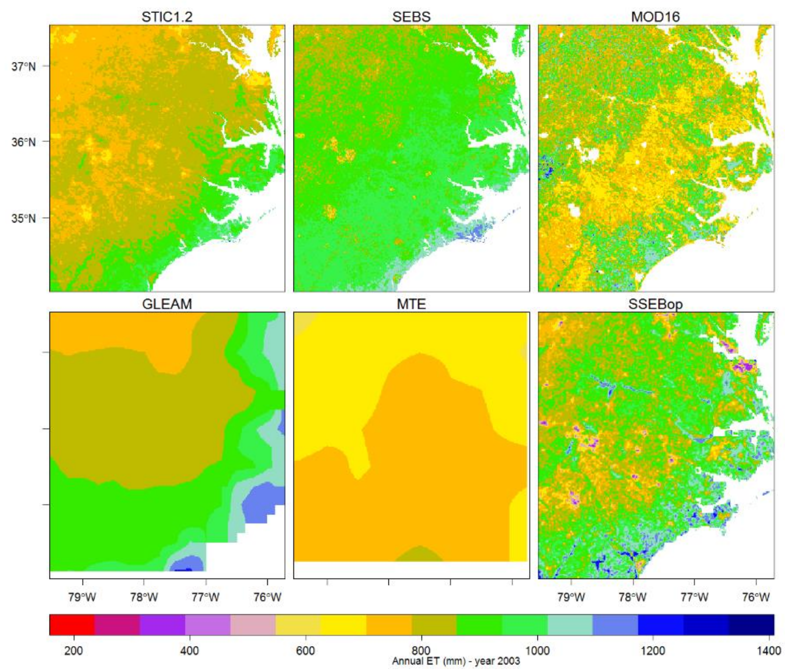

Figure AR7. Annual ET map derived from STIC1.2, SEBS, MOD16, GLEAM, MTE, and SSEBop for the eastern (E) bounding box covering US-NC1 and US-NC2 flux sites (Fig. 2, Page 31).

As demonstrated and discussed in Mallick et al. (2014), although towers are often installed in relatively homogenous terrain, rarely can this be assumed for heterogeneous landscapes in arid and semi-arid environment. The slope of the regression between the observed and estimated λE of individual biome category was significantly related to the average variance of LST surrounding the tower sites (Mallick et al., 2014). The slope of regression varied systematically with the landscape heterogeneity. Similar results are also shown by Stoy et al. (2013), who also found a systematic relationship between the surface energy balance closure, soil moisture variability, and landscape heterogeneity over 173 FLUXNET tower sites.

Currently, there is no consensus on which MODIS footprint size to use to represent fluxes from a flux site and hence any size or method used is subjected to debate. However, most flux sites (other than the arid and some semi-arid sites) used in this study cover vegetated area that is large enough for a 1×1 km^2 MODIS pixel to be represented as a homogenous pixel, which was also verified in Google Earth Engine. Typically, a pixel-to-footprint match is considered adequate if the vegetation and environmental characteristics within the footprint are good representatives of the surrounding area contained by the MODIS pixels (Yuan et al., 2010). The US-Ton site, however, may not be as homogenous as other sites in terms of vegetation type, as the site is dominated by deciduous blue oaks (*Quercusdouglasii sp.*) and the understory and open grassland are mainly cool-season C3 annual species(Ma et al., 2007). This could lead to dry and wet patches of LST, as briefly discussed on Page 15, Line 5-10. Blue oaks and grasses have distinct phenology and MODIS is not sensitive to understory canopy (Ma et al., 2007;Xiao et al., 2010). In addition, the US-Wkg site was classified as either open shrublands or grasslands in different years on MCD12Q1 datasets and was not homogenous beyond a 3×3 neighborhood (i.e. one class was surrounded by pixels of another class).

Nonetheless, the sites considered in our study have been widely used in validation of ET as well as other land surface variables and is currently considered the state of the art observation datasets that can be used as benchmark to assess the performance of the remote sensing based models (Running et al., 2004;Yang et al., 2007;Jung et al., 2010) and several common approaches to extract representative MODIS pixel values include single tower pixel (Yuan et al., 2010;Ryu et al., 2011;Jiang and Ryu, 2016), 3×3 mean value with center pixel as the coordinates of flux towers (Sims et al., 2008;Xiao et al., 2008;Yang et al., 2013), and some foot print analysis (Vinukollu et al., 2011). In this study, we extracted ET values from a single tower pixel located closest to the MODIS pixel, but when a 3×3 mean value of estimated ET with the center pixel as the coordinates of flux towers was considered, only negligible changes in model performances was noticed (Table AR2). In addition, the mean values of 8-day cumulative ET from each model were not significantly different when a single pixel or a 3×3 neighborhood was considered (*p*-value > 0.75). Hence, we think that in this study, footprint uncertainties are minimized by selecting homogenous core AmeriFlux sites and this method is consistent with what has been done in previous studies.

Table AR2. Evaluation of 8-day cumulative ET (Table 3, Page 28) from STIC1.2 and SEBS against observed ET from thirteen core AmeriFlux sites in the US combining data from one dry, one wet, and one normal year when pixel values of estimated ET were taken as a mean of 3×3 neighborhood with center pixel as the coordinates of flux towers. No significant differences in model performance were noticed when a single tower pixel was considered (Table 3, Page 28).

Model	R^2	RMSE (mm 8-day^{-1})	MAE (mm) (mm 8-day^{-1})	PBIAS (%)
STIC1.2	0.66	7.3	5.3	-4
SEBS	0.54	9.7	7.2	27
MOD16	0.58	9	6.3	-27

Minor: 4. Page7(L7): Delete "the" between "both" and "variables"

Response: Necessary change will be incorporated.

Minor: 5. Page9(L20): Please specify the equation for NDVI and/or provide references.

Response: We will provide the following reference for NDVI in page 9 line 24-25.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, *8*(2), 127-150.

Minor 6. Page 16 (Discussion on MOD16): It is worthwhile reading Yang et al. (2016, WRR; doi: 10.1002/2014WR015619) for a more physical explanation on the MOD16 uncertainty.

Response: Thanks for referring this paper and will add the following description in the revised manuscript (After Page 16, Line 30)

"Other studies have also found MOD16 to underestimate λE or ET significantly (Yang et al., 2015;Biggs et al., 2016). Yang et al. (2015) highlighted four key uncertainties associated with the MOD16 algorithm (Mu et al., 2011), which could explain the relatively poor performance of MOD16 in this study. First, the dependency of the MOD16 algorithm on meteorological forcing (and not the T_R) to account for the soil moisture restriction on evaporation and transpiration processes results in a slower response of variations in energy and heat fluxes (Long and Singh, 2010). Second, underestimation of transpiration in MOD16 could occur due to overestimation of environmental stresses on canopy conductance that is expressed as the potential canopy conductance multiplied by two scaling factors that represent influences from T_A and VPD deficit (Yang et al., 2013). Third, the empirical nature of the soil moisture constraint function(Fisher et al., 2008) based on the complementary hypothesis (Bouchet, 1963) using VPD and RH leads to large uncertainties in evaporation from the unsaturated soil. Finally, the coarse resolution meteorological data (1° × 1.25°) used in MOD16 may not be well representative of surfaces with high moisture availability."

References

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56, FAO, Rome, 300, D05109, 1998.

Allen, R. G., Tasumi, M., and Trezza, R.: Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC) - Model, J Irrig Drain Eng, 133, 380-394, 2007.

ASCE-EWRI: The ASCE standardized reference evapotranspiration equation: ASCE-EWRI Standardization of Reference Evapotranspiration Task Committe Report, 2005,

Baldocchi, D. D., and Ryu, Y.: A Synthesis of Forest Evaporation Fluxes – from Days to Years – as Measured with Eddy Covariance, in: Forest Hydrology and Biogeochemistry: Synthesis of Past Research and Future Directions, edited by: Levia, D. F., Carlyle-Moses, D., and Tanaka, T., Springer Netherlands, Dordrecht, 101-116, 2011.

Biggs, T. W., Marshall, M., and Messina, A.: Mapping daily and seasonal evapotranspiration from irrigated crops using global climate grids and satellite imagery: Automation and methods comparison, Water Resour Res, 52, 7311-7326, 10.1002/2016WR019107, 2016.

Bouchet, R.: Evapotranspiration reelle, evapotranspiration potentielle, et production agricole, Annales Agronomiques, 1963, 743-824,

Brutsaert, W., and Sugita, M.: Application of self-preservation in the diurnal evolution of the surface energy budget to determine daily evaporation, Journal of Geophysical Research: Atmospheres, 97, 18377-18382, 10.1029/92JD00255, 1992.

Bunting, D. P., Kurc, S. A., Glenn, E. P., Nagler, P. L., and Scott, R. L.: Insights for empirically modeling evapotranspiration influenced by riparian and upland vegetation in semiarid regions, Journal of Arid Environments, 111, 42-52, https://doi.org/10.1016/j.jaridenv.2014.06.007, 2014.

Cammalleri, C., Anderson, M., and Kustas, W.: Upscaling of evapotranspiration fluxes from instantaneous to daytime scales for thermal remote sensing applications, Hydrology and Earth System Sciences, 18, 1885-1894, 2014.

Chávez, J. L., Neale, C. M. U., Prueger, J. H., and Kustas, W. P.: Daily evapotranspiration estimates from extrapolating instantaneous airborne remote sensing ET values, Irrigation Sci, 27, 67-81, 10.1007/s00271-008-0122-3, 2008.

Chen, X., Su, Z., Ma, Y., Yang, K., Wen, J., and Zhang, Y.: An Improvement of Roughness Height Parameterization of the Surface Energy Balance System (SEBS) over the Tibetan Plateau, Journal of Applied Meteorology and Climatology, 52, 607-622, 10.1175/jamc-d-12-056.1, 2013.

Colaizzi, P., Evett, S., Howell, T., and Tolk, J.: Comparison of five models to scale daily evapotranspiration from one-time-of-day measurements, Transactions of the ASABE, 49, 1409-1417, 2006.

Crago, R. D.: Conservation and variability of the evaporative fraction during the daytime, Journal of Hydrology, 180, 173-194, 1996.

Ershadi, A., McCabe, M. F., Evans, J. P., Mariethoz, G., and Kavetski, D.: A Bayesian analysis of sensible heat flux estimation: Quantifying uncertainty in meteorological forcing to improve model prediction, Water Resour Res, 49, 2343-2358, 2013.

Fisher, J. B., Tu, K. P., and Baldocchi, D. D.: Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites, Remote Sensing of Environment, 112, 901-919, https://doi.org/10.1016/j.rse.2007.06.025, 2008.

Garcia, M., Fernández, N., Villagarcía, L., Domingo, F., Puigdefábregas, J., and Sandholt, I.: Accuracy of the Temperature–Vegetation Dryness Index using MODIS under water-limited vs. energy-limited evapotranspiration conditions, Remote Sensing of Environment, 149, 100-117, https://doi.org/10.1016/j.rse.2014.04.002, 2014.

Gentine, P., Entekhabi, D., Chehbouni, A., Boulet, G., and Duchemin, B.: Analysis of evaporative fraction diurnal behaviour, Agricultural and Forest Meteorology, 143, 13-29, https://doi.org/10.1016/j.agrformet.2006.11.002, 2007.

Hu, L., and Brunsell, N. A.: The impact of temporal aggregation of land surface temperature data for surface urban heat island (SUHI) monitoring, Remote Sensing of Environment, 134, 162-174, https://doi.org/10.1016/j.rse.2013.02.022, 2013.

Ichii, K., Wang, W., Hashimoto, H., Yang, F., Votava, P., Michaelis, A. R., and Nemani, R. R.: Refinement of rooting depths using satellite-based evapotranspiration seasonality for ecosystem modeling in California, Agricultural and Forest Meteorology, 149, 1907-1918, https://doi.org/10.1016/j.agrformet.2009.06.019, 2009.

Jiang, C., and Ryu, Y.: Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from Breathing Earth System Simulator (BESS), Remote Sensing of Environment, 186, 528-547, https://doi.org/10.1016/j.rse.2016.08.030, 2016.

Jin, Y., Randerson, J. T., and Goulden, M. L.: Continental-scale net radiation and evapotranspiration estimated using MODIS satellite observations, Remote Sensing of Environment, 115, 2302-2319, https://doi.org/10.1016/j.rse.2011.04.031, 2011.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., de Jeu, R., Dolman, A. J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A. D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., and Zhang, K.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467, 951, 10.1038/nature09396

https://www.nature.com/articles/nature09396#supplementary-information, 2010.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, Journal of Geophysical Research: Biogeosciences, 116, n/a-n/a, 10.1029/2010JG001566, 2011.

Long, D., and Singh, V. P.: Integration of the GG model with SEBAL to produce time series of evapotranspiration of high spatial resolution at watershed scales, Journal of Geophysical Research: Atmospheres, 115, n/a-n/a, 10.1029/2010JD014092, 2010.

Ma, S., Baldocchi, D. D., Xu, L., and Hehn, T.: Inter-annual variability in carbon dioxide exchange of an oak/grass savanna and open grassland in California, Agricultural and Forest Meteorology, 147, 157-171, https://doi.org/10.1016/j.agrformet.2007.07.008, 2007.

Mallick, K., Jarvis, A. J., Boegh, E., Fisher, J. B., Drewry, D. T., Tu, K. P., Hook, S. J., Hulley, G., Ardö, J., Beringer, J., Arain, A., and Niyogi, D.: A Surface Temperature Initiated Closure (STIC) for surface energy balance fluxes, Remote Sensing of Environment, 141, 243-261, http://dx.doi.org/10.1016/j.rse.2013.10.022, 2014.

Mallick, K., Boegh, E., Trebs, I., Alfieri, J. G., Kustas, W. P., Prueger, J. H., Niyogi, D., Das, N., Drewry, D. T., Hoffmann, L., and Jarvis, A. J.: Reintroducing radiometric surface temperature into the Penman-Monteith formulation, Water Resour Res, 51, 6214-6243, 10.1002/2014wr016106, 2015.

Mallick, K., Trebs, I., Boegh, E., Giustarini, L., Schlerf, M., Drewry, D. T., Hoffmann, L., von Randow, C., Kruijt, B., Araùjo, A., Saleska, S., Ehleringer, J. R., Domingues, T. F., Ometto, J. P. H. B., Nobre, A. D., de Moraes, O. L.

L., Hayek, M., Munger, J. W., and Wofsy, S. C.: Canopy-scale biophysical controls of transpiration and evaporation in the Amazon Basin, Hydrol. Earth Syst. Sci., 20, 4237-4264, 10.5194/hess-20-4237-2016, 2016.

Martens, B., Gonzalez Miralles, D., Lievens, H., van der Schalie, R., de Jeu, R. A., Fernández-Prieto, D., Beck, H. E., Dorigo, W., and Verhoest, N.: GLEAM v3: Satellite-based land evaporation and root-zone soil moisture, Geosci Model Dev, 10, 1903-1925, 2017.

Masuoka, E., Fleig, A., Wolfe, R. E., and Patt, F.: Key characteristics of MODIS data products, IEEE Transactions on Geoscience and Remote Sensing, 36, 1313-1323, 10.1109/36.701081, 1998.

McCabe, M. F., Ershadi, A., Jimenez, C., Miralles, D. G., Michel, D., and Wood, E. F.: The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally gridded forcing data, Geosci Model Dev, 9, 283-305, 10.5194/gmd-9-283-2016, 2016.

Miralles, D., Holmes, T., De Jeu, R., Gash, J., Meesters, A., and Dolman, A.: Global land-surface evaporation estimated from satellite-based observations, Hydrology and Earth System Sciences, 15, 453, 2011.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sensing of Environment, 115, 1781-1800, http://dx.doi.org/10.1016/j.rse.2011.02.019, 2011.

Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., and Hashimoto, H.: A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production, BioScience, 54, 547-560, 10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2, 2004.

Ryu, Y., Baldocchi, D. D., Kobayashi, H., van Ingen, C., Li, J., Black, T. A., Beringer, J., van Gorsel, E., Knohl, A., Law, B. E., and Roupsard, O.: Integration of MODIS land and atmosphere products with a coupled-process model to estimate gross primary productivity and evapotranspiration from 1 km to global scales, Global Biogeochemical Cycles, 25, n/a-n/a, 10.1029/2011GB004053, 2011.

Ryu, Y., Baldocchi, D. D., Black, T. A., Detto, M., Law, B. E., Leuning, R., Miyata, A., Reichstein, M., Vargas, R., Ammann, C., Beringer, J., Flanagan, L. B., Gu, L., Hutley, L. B., Kim, J., McCaughey, H., Moors, E. J., Rambal, S., and Vesala, T.: On the temporal upscaling of evapotranspiration from instantaneous remote sensing measurements to 8-day mean daily-sums, Agricultural and Forest Meteorology, 152, 212-222, https://doi.org/10.1016/j.agrformet.2011.09.010, 2012.

Senay, G. B., Bohms, S., Singh, R. K., Gowda, P. H., Velpuri, N. M., Alemu, H., and Verdin, J. P.: Operational Evapotranspiration Mapping Using Remote Sensing and Weather Datasets: A New Parameterization for the SSEB Approach, J Am Water Resour As, 49, 577-591, 2013.

Shuttleworth, W., Gurney, R., Hsu, A., and Ormsby, J.: FIFE: the variation in energy partition at surface flux sites, IAHS Publ, 186, 1989.

Sims, D. A., Rahman, A. F., Cordova, V. D., El-Masri, B. Z., Baldocchi, D. D., Bolstad, P. V., Flanagan, L. B., Goldstein, A. H., Hollinger, D. Y., Misson, L., Monson, R. K., Oechel, W. C., Schmid, H. P., Wofsy, S. C., and Xu, L.: A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS, Remote Sensing of Environment, 112, 1633-1646, https://doi.org/10.1016/j.rse.2007.08.004, 2008.

Stoy, P. C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., Arain, M. A., Arneth, A., Aurela, M., Bernhofer, C., Cescatti, A., Dellwik, E., Duce, P., Gianelle, D., van Gorsel, E., Kiely, G., Knohl, A., Margolis, H., McCaughey, H., Merbold, L., Montagnani, L., Papale, D., Reichstein, M., Saunders, M., Serrano-Ortiz, P., Sottocornola, M., Spano, D., Vaccari, F., and Varlagin, A.: A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape scale heterogeneity, Agricultural and Forest Meteorology, 171, 137-152, http://dx.doi.org/10.1016/j.agrformet.2012.11.004, 2013.

Tang, R., Tang, B., Wu, H., and Li, Z.-L.: On the feasibility of temporally upscaling instantaneous evapotranspiration using weather forecast information, Int J Remote Sens, 36, 4918-4935, 10.1080/01431161.2015.1029597, 2015.

Tian, F., Qiu, G., Yang, Y., Lü, Y., and Xiong, Y.: Estimation of evapotranspiration and its partition based on an extended three-temperature model and MODIS products, Journal of Hydrology, 498, 210-220, https://doi.org/10.1016/j.jhydrol.2013.06.038, 2013.

van der Kwast, J., Timmermans, W., Gieske, A., Su, Z., Olioso, A., Jia, L., Elbers, J., Karssenberg, D., and de Jong, S.: Evaluation of the Surface Energy Balance System (SEBS) applied to ASTER imagery with flux-measurements at the SPARC 2004 site (Barrax, Spain), Hydrol. Earth Syst. Sci., 13, 1337-1347, 10.5194/hess-13-1337-2009, 2009.

Velpuri, N. M., Senay, G. B., Singh, R. K., Bohms, S., and Verdin, J. P.: 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, 139, 35-49, 2013.

Vinukollu, R. K., Meynadier, R., Sheffield, J., and Wood, E. F.: Multi-model, multi-sensor estimates of global evapotranspiration: climatology, uncertainties and trends, Hydrological Processes, 25, 3993-4010, 2011.

Wu, C., Munger, J. W., Niu, Z., and Kuang, D.: Comparison of multiple models for estimating gross primary production using MODIS and eddy covariance data in Harvard Forest, Remote Sensing of Environment, 114, 2925-2939, https://doi.org/10.1016/j.rse.2010.07.012, 2010.

Xiao, J., Zhuang, Q., Baldocchi, D. D., Law, B. E., Richardson, A. D., Chen, J., Oren, R., Starr, G., Noormets, A., Ma, S., Verma, S. B., Wharton, S., Wofsy, S. C., Bolstad, P. V., Burns, S. P., Cook, D. R., Curtis, P. S., Drake, B. G., Falk, M., Fischer, M. L., Foster, D. R., Gu, L., Hadley, J. L., Hollinger, D. Y., Katul, G. G., Litvak, M., Martin, T. A., Matamala, R., McNulty, S., Meyers, T. P., Monson, R. K., Munger, J. W., Oechel, W. C., Paw U, K. T., Schmid, H. P., Scott, R. L., Sun, G., Suyker, A. E., and Torn, M. S.: Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data, Agricultural and Forest Meteorology, 148, 1827-1847, https://doi.org/10.1016/j.agrformet.2008.06.015, 2008.

Xiao, J., Zhuang, Q., Law, B. E., Chen, J., Baldocchi, D. D., Cook, D. R., Oren, R., Richardson, A. D., Wharton, S., Ma, S., Martin, T. A., Verma, S. B., Suyker, A. E., Scott, R. L., Monson, R. K., Litvak, M., Hollinger, D. Y., Sun, G., Davis, K. J., Bolstad, P. V., Burns, S. P., Curtis, P. S., Drake, B. G., Falk, M., Fischer, M. L., Foster, D. R., Gu, L., Hadley, J. L., Katul, G. G., Matamala, R., McNulty, S., Meyers, T. P., Munger, J. W., Noormets, A., Oechel, W. C., Paw U, K. T., Schmid, H. P., Starr, G., Torn, M. S., and Wofsy, S. C.: A continuous measure of gross primary production for the conterminous United States derived from MODIS and AmeriFlux data, Remote Sensing of Environment, 114, 576-591, https://doi.org/10.1016/j.rse.2009.10.013, 2010.

Xiong, Y., Zhao, S., Yin, J., Li, C., and Qiu, G.: Effects of Evapotranspiration on Regional Land Surface Temperature in an Arid Oasis Based on Thermal Remote Sensing, IEEE Geoscience and Remote Sensing Letters, 13, 1885-1889, 10.1109/LGRS.2016.2616409, 2016.

Xiong, Y. J., Zhao, S. H., Tian, F., and Qiu, G. Y.: An evapotranspiration product for arid regions based on the three-temperature model and thermal remote sensing, Journal of Hydrology, 530, 392-404, https://doi.org/10.1016/j.jhydrol.2015.09.050, 2015.

Yang, F., White, M. A., Michaelis, A. R., Ichii, K., Hashimoto, H., Votava, P., Zhu, A. X., and Nemani, R. R.: Prediction of Continental-Scale Evapotranspiration by Combining MODIS and AmeriFlux Data Through Support Vector Machine, IEEE Transactions on Geoscience and Remote Sensing, 44, 3452-3461, 10.1109/TGRS.2006.876297, 2006.

Yang, F., Ichii, K., White, M. A., Hashimoto, H., Michaelis, A. R., Votava, P., Zhu, A. X., Huete, A., Running, S. W., and Nemani, R. R.: Developing a continental-scale measure of gross primary production by combining MODIS and AmeriFlux data through Support Vector Machine approach, Remote Sensing of Environment, 110, 109-122, https://doi.org/10.1016/j.rse.2007.02.016, 2007.

Yang, Y., Shang, S., Guan, H., and Jiang, L.: A novel algorithm to assess gross primary production for terrestrial ecosystems from MODIS imagery, Journal of Geophysical Research: Biogeosciences, 118, 590-605, 10.1002/jgrg.20056, 2013.

Yang, Y., Long, D., Guan, H., Liang, W., Simmons, C., and Batelaan, O.: Comparison of three dual-source remote sensing evapotranspiration models during the MUSOEXE-12 campaign: Revisit of model physics, Water Resour Res, 51, 3145-3165, 10.1002/2014WR015619, 2015.

Yuan, W., Liu, S., Yu, G., Bonnefond, J.-M., Chen, J., Davis, K., Desai, A. R., Goldstein, A. H., Gianelle, D., Rossi, F., Suyker, A. E., and Verma, S. B.: Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data, Remote Sensing of Environment, 114, 1416-1431, https://doi.org/10.1016/j.rse.2010.01.022, 2010.

Zhang, Y., Leuning, R., Hutley, L. B., Beringer, J., McHugh, I., and Walker, J. P.: Using long-term water balances to parameterize surface conductances and calculate evaporation at 0.05° spatial resolution, Water Resour Res, 46, n/a-n/a, 10.1029/2009WR008716, 2010.