Editor's comments:

Referee #1 still has some concerns about the use and exact description of GLEAM, specifically what it is and is not, which I share. I therefore kindly ask you to revise the manuscript once again, carefully considering all remaining comments of the referee.

Thank you for your very thoughtful summary and refinement of the reviewer' comments. We have revised the manuscript carefully according to Referee #1's comments.

Answer to Reviewer 1 ESSD-2021-61

We thank Reviewer 1 for the comments. We provide here our responses to those comments and describe how we addressed them in the revised manuscript. The original reviewer comments are in normal black font while our answers appear in blue font. The corresponding edit in the manuscript are included in red font.

Table 2 and 3: I'd strongly suggest to include GLEAM validation metrics.
Why is no validation of GLEAM included? Surely this is vital as GLEAM is very often used for global studies and the paper should provide insights in how, when or where the merged product surpasses GLEAM. Note that when validating E products it is often standard-practise to exclude days with strong precipitation.
If GLEAM outperforms the others for a given pixel, can the merged product actually achieve the same or better performance than the GLEAM reference data? I'm just wondering, some thoughts and a sentence or two would be helpful.

Response:

Thank you for your very thoughtful comment and suggestions. We have added the validation metrics of GLEAM in table 2 and 3. The descriptions of table 2 and 3 have been modified in the revised manuscript accordingly. Lines 345-377 read: "

Table 2. The verification results including R and RMSD between daily Ground-measured ET and daily ET from different products in different ecosystems. Values in bold indicates the highest quality.

Ecosystem type	*ERA5*		*MERRA2*		*GLDAS*		*GLEAM*		*REA*	
	R	*RMSD*	*R*	*RMSD*	*R*	*RMSD*	*R*	*RMSD*	*R*	*RMSD*
CRO	***0.66***	*1.24*	*0.55*	*1.42*	*0.60*	*1.48*	*0.60*	***1.22***	*0.60*	*1.38*
DBF	*0.76*	***1.06***	*0.71*	*1.23*	*0.74*	*1.19*	*0.67*	*1.16*	***0.77***	*1.07*
DNF	***0.81***	***0.55***	*0.73*	*0.75*	*0.77*	*0.73*	*0.64*	*0.75*	*0.80*	*0.62*
EBF	***0.72***	***1.08***	*0.61*	*1.59*	*0.58*	*1.36*	*0.70*	*1.11*	*0.65*	***1.08***
ENF	*0.66*	*1.03*	*0.66*	*1.21*	*0.67*	*1.05*	*0.66*	*1.04*	***0.73***	***0.88***
GRA	*0.72*	*1.05*	***0.77***	*1.11*	*0.70*	*1.09*	*0.73*	*0.96*	***0.77***	***0.94***
MF	*0.77*	***1.05***	***0.79***	*1.37*	*0.70*	*1.23*	*0.70*	*1.12*	*0.74*	*1.12*
OSH	*0.43*	*1.00*	*0.47*	*0.92*	*0.46*	*0.96*	*0.33*	*1.15*	***0.50***	***0.88***

SAV	*0.61*	*1.23*	*0.62*	*1.40*	*0.63*	*1.22*	*0.58*	*1.25*	***0.66***	***1.16***
WET	***0.57***	***1.40***	*0.44*	*1.66*	*0.47*	*1.56*	*0.52*	*1.44*	*0.46*	*1.59*
WSA	*0.68*	*1.24*	*0.63*	*1.46*	*0.64*	*1.17*	***0.72***	***1.11***	*0.70*	*1.13*

Verification of ET products from different ecosystems has been conducted in order to further evaluate their performances. Table 2 describes quantitatively the performances of the ET products in 11 ecosystem types of site on a daily scale from two indicators, R and RMSD. The values in bold print indicate the best performance of the four products. The results demonstrate that no individual product performs best across all ecosystems. For 42 ENF, 34 GRA, 9 OSH, 8 SAV sites, REA has higher R and lower RMSD than individual products. For 23 DBF and 13 EBF sites, REA has a optimal R or RMSD; specifically with the highest R of 0.77 and second lowest RMSD of 1.07 mm per day for DBF, and the lowest RMSD of 1.08 mm per day and the second highest R of 0.65 for EBF. REA performs worse than at least one individual product at 63 other sites. Specifically, REA has a lower R of 0.60 than ERA5, and a higher RMSD of 1.38 mm per day than GLEAM and ERA5 at 17 CRO sites. For 1 DNF site, REA has lower R of 0.80 and higher RMSD of 0.62 mm per day than ERA5. For 9 MF sites, REA has a lower R of 0.74 than ERA5 and MERRA2, and a higher RMSD of 1.12 mm per day than ERA5. For 19 WET sites, REA has a lower R of 0.46 and a higher RMSD of 1.59 mm per day than ERA5 and GLDAS. For 6 WSA sites, REA has lower R of 0.70 and higher RMSD of 1.13 mm per day than GLEAM. Generally, ERA5, MERRA2, GLEAM and REA show the best performance respectively in four (including CRO, DNF, EBF and WET), two (GRA and MF), one (WSA) and five (DBF, ENF, GRA, OSH and SAV) ecosystems in terms of R. Based on RMSD, both ERA5 and REA performed best in five ecosystems (with the former including DBF, DNF, EBF, MF and WET, and the latter including EBF, ENF, GRA, OSH and SAV). REA does not perform best across all ecosystems, however, it avoids the worst performance in any ecosystem. Taylor Diagram results of daily Ground-measured ET and ET from the different products in 11 ecosystems are put in support information (Fig. S2).

Table 3. The verification results including R and RMSD between monthly Ground-measured ET and monthly ET from different products in different ecosystems. Values in bold indicates the highest quality.

Ecosystem type	*ERA5*		*MERRA2*		*GLDAS*		*GLEAM*		*REA*	
	R	*RMSD*	*R*	*RMSD*	*R*	*RMSD*	*R*	*RMSD*	*R*	*RMSD*
CRO	*0.71*	*31.84*	*0.58*	*38.33*	*0.64*	*39.59*	*0.66*	***30.85***	***0.73***	*31.14*
DBF	*0.84*	*26.61*	*0.77*	*32.40*	*0.83*	*30.27*	*0.75*	*28.43*	***0.86***	***26.17***
DNF	***0.93***	***10.62***	*0.84*	*18.63*	*0.91*	*18.96*	*0.85*	*14.92*	*0.91*	*11.01*
EBF	*0.81*	*25.85*	*0.71*	*40.32*	*0.69*	*32.21*	***0.84***	***22.79***	*0.78*	*27.06*
ENF	*0.74*	*25.01*	*0.72*	*31.69*	***0.76***	*25.38*	***0.76***	*24.24*	***0.76***	***23.91***
GRA	*0.77*	*27.11*	***0.83***	*28.59*	*0.77*	*26.54*	*0.81*	***22.06***	*0.77*	*27.53*
MF	*0.83*	*27.57*	***0.85***	*37.98*	*0.75*	*32.90*	*0.79*	*27.71*	*0.83*	***26.99***
OSH	*0.45*	*25.34*	*0.51*	***23.39***	*0.52*	*23.75*	*0.27*	*32.57*	***0.53***	*24.54*

SAV	*0.65*	*32.05*	*0.65*	*38.08*	*0.67*	*31.88*	*0.59*	*34.33*	***0.68***	***31.45***
WET	*0.61*	*38.30*	*0.46*	*46.92*	*0.49*	*43.78*	*0.55*	*40.38*	***0.64***	***37.35***
WSA	*0.73*	*33.71*	*0.68*	*38.54*	*0.72*	*28.62*	***0.77***	***28.12***	*0.73*	*34.49*

Similar to Table 2, Table 3 shows the performance of ET products in different ecosystems on a monthly scale. Compared with daily scale, the performance of each product has changed, among which all of the R becomes higher. REA has higher R and lower RMSD than individual products at 23 DBF, 42 ENF, 8 SAV and 19 WET sites. It has an optimal R or RMSD at 17 CRO, 9 MF and 9 OSH sites. At other 64 sites, REA has a worse performance than at least one individual product. For 1 DNF site, REA has lower R and higher RMSD than ERA5. For 13 EBF sites, REA has lower R and higher RMSD than GLEAM and ERA5. For 34 GRA sites, ERA5 has a lower R of 0.77 than MERRA2 and GLEAM, and a higher RMSD of 27.53 mm per month than GLEAM, ERA5 and GLDAS. For 6 WSA sites, REA has a lower R of 0.73 than GLEAM, and a higher RMSD of 34.49 mm per day than GLEAM, GLDAS and ERA5. Similar to the daily scale, REA does not have a better performance than any individual product in all ecosystems, but is superior to at least one individual one. Similarly, the Taylor charts of monthly Ground-measured ET and ET from the different products in 11 ecosystems are put in support information (Fig. S3).".

Generally, eddy-covariance measurements are less reliable during days with strong precipitation. We have validated the merged product with Ground-measured ET exclude these days. Fig. S1 shows the taylor diagrams of daily ET from different products and Ground-measured ET including and excluding days with strong precipitation. The table shows the validation metrics including R and RMSD between daily Ground-measured ET and daily ET from different products. As can be seen from the figure and table, there is small difference between the validation results of (a) and (b), which indicating that data on days with strong precipitation have minimal impact on this study. Also, we believe that using the quality control measures provided by the insitu data developers gets us close enough to a representative quality indication. Nonetheless, we have added the some text based on the discussed comparison above to the revised manuscript. Lines 193-195 read: "*Since there is minimal impact of Ground-measured ET on days with strong precipitation on the verification results (Fig. S1), data on these days are not excluded in order to retain more ground data samples for statistical analysis.*".

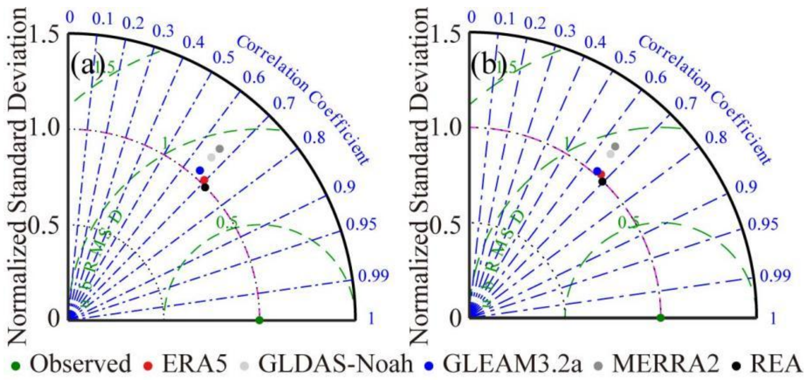

Figure S1. Taylor diagram of daily Ground-measured ET (a) including, (b) excluding days with strong precipitation and ET from different products.

Table. The verification results including R and RMSD between daily Ground-measured ET and daily ET from different products.

	ERA5		GLDAS		GLEAM		MERRA2		REA	
	R	RMSD	R	RMSD	R	RMSD	R	RMSD	R	RMSD
(a)	0.69	0.96	0.66	1.12	0.66	1.03	0.66	1.18	**0.72**	**0.91**
(b)	0.68	0.97	0.65	1.10	0.66	0.99	0.65	1.15	**0.70**	**0.92**

Fig. S6 shows the relationship between the quality of the merged product and GLEAM. As can be seen, the merged product is highly sensitive to the quality of the reference data. If the quality of GLEAM is high, the quality of the merged product will be correspondingly high. We have added the description to the revised manuscript. Lines 519-522 read: "*Consequently, the results also demonstrate that REA is more sensitive to higher qualities in GLEAM. As a result, where GLEAM has lower qualities, REA tends to have higher qualities. Meanwhile, they both have very similar qualities, or even higher in REA, at regions where GLEAM has higher correlations and lower differences with the insitu datasets (Fig S6). Thus REA is more sensitive to the reference data where it is more reliable.*".

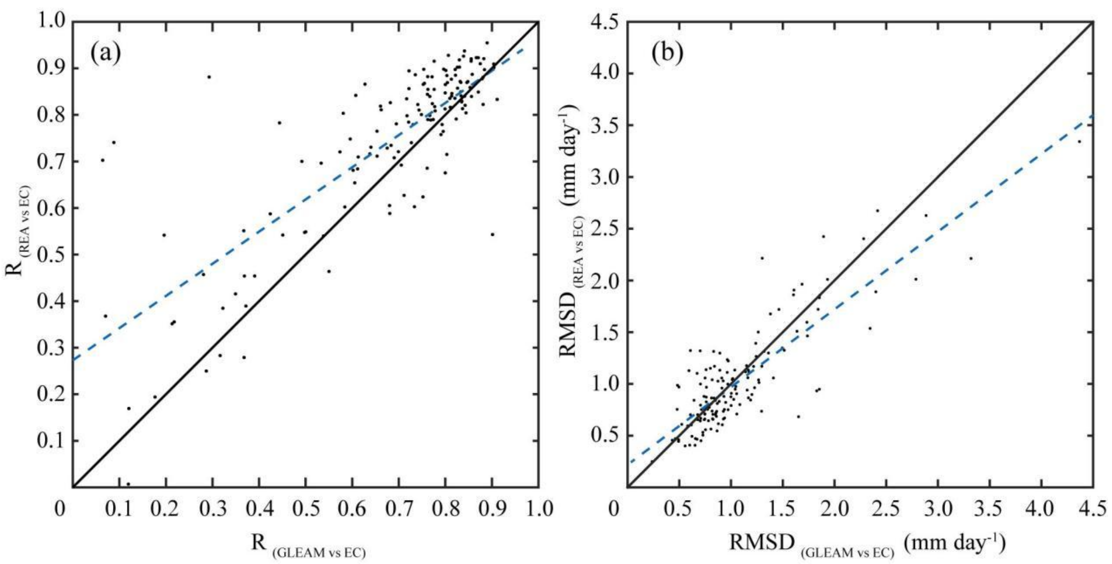

Figure S6. Scatter plots of the correlation coefficients and RMSD between GLEAM, REA and Ground-measured ET. Linear fits are plotted in blue and the 1:1 line is depicted.

Furthermore I'd expect a few sentences on the exact input data GLEAM version a relies on based on the respective GLEAM paper. Net radiation as correctly stated is based on ERA-Interim and GLEAM is actually extremely sensitive to net radiation.

Response:

Thank you for your cogent advice. GLEAM version 3a is produced using satellite observations including soil moisture, vegetation optical depth and snow-water equivalent, a multi-source precipitation product and relies on only radiation and temperature inputs from reanalysis products (Martens et al., 2017). We have added the description to the revised manuscript. Lines 117-120 read: "*Additionally, GLEAM is not a complex terrestrial model as found in land models of ERA5, MERRA2 and GLDAS, but a set of algorithms dedicated to estimating terrestrial evaporation using retrieved satellite observations including soil moisture, vegetation optical depth and snow-water equivalent, a multi-source precipitation product and relies on only radiation and temperature inputs from reanalysis products (Martens et al., 2017).*".

Concerning these sentences in the rebuttal letter/or text:
Please reword/revise parts in the updated manuscript that reflect on the ideas below.
"GLEAM is not a traditional terrestrial model ..." Please reword this, it's not clear what a 'traditional' terrestrial model is or why GLEAM should be any different. Perhaps argue that GLEAM is specifically designed to estimate evaporation whereas the other 'big' models are required to simulate a higher number of variables decently (This is a spontaneous idea, please check carefully with the literature). GLEAM is not part of a larger Earth System model with an atmospheric/sea ice component etc. Perhaps that's more of a difference too?

Response:

Thank you for your very thoughtful comment. We have changed "traditional" to "complex", and added the description of "complex" to the revised manuscript. GLEAM is not a complex terrestrial model as found in land models of ERA5, MERRA2 and GLDAS. It is a set of algorithms dedicated to the estimation of terrestrial evaporation (Martens et al., 2017). Lines 117-120 read: "*Additionally, GLEAM is not a complex terrestrial model as found in land models of ERA5, MERRA2 and GLDAS, but a set of algorithms dedicated to estimating terrestrial evaporation using retrieved satellite observations including soil moisture, vegetation optical depth and snow-water equivalent, a multi-source precipitation product and* *relies on only radiation and temperature inputs from reanalysis products (Martens et al., 2017).*".

"MERRA2 and ERA5 are based on brightness temperatures that are assimilated into their atmospheric models and only indirectly impact the land states." --> A lot more than brightness temperatures are assimilated, e.g. IR radiances, air temperature measurements from aircraft etc. etc. etc.

Response:

Thank you for pointing this out. We have modified the description. This sentence is changed to "MERRA2 and ERA5 are based on brightness temperatures, infrared radiances, air temperature measurements from aircraft and a lot more data that are assimilated into their atmospheric models and only indirectly impact the land states.".

"It is expected that GLEAM's over-reliance on observations states would serve as some sort of benchmark to estimate the weights of the model-based products. Thus, the goal is not based on a superior skill of GLEAM but its added value due to its uniqueness relative to the model-based products, which we believe, does have merits"
One might argue that GLEAM is more directly linked to satellite input but it is no less a model than the other products. The reanalysis products incorporate many many more observations than GLEAM does. It is a rather simple model (in a good way) focusing on soil moisture and evaporation (and computes some more variables required for E and soil moisture).

Response:

Thank you very much for your comment. It definitely makes a lot of sense this looking at it this way. We have modified the description to "Although GLEAM is no less a model than the other products, the ET output from GLEAM is more directly linked to the satellite retrieval inputs within a more simplified model. This peculiar framework of GLEAM could be reliable to serve as some sort of benchmark from which we estimate the weights of the model-based products. Thus, the goal is not based on a superior skill of GLEAM but its added value due to its uniqueness relative to the other model-based products, which we believe does have merits."

"GLEAM is not a traditional terrestrial model as found in ERA5, MERRA2 and GLDAS" See above, not sure about traditional.

Response:

Thank you for your very thoughtful comment. We have changed "traditional" to "complex", and added the description of "complex" to the revised manuscript.

GLEAM is not a complex terrestrial model as found in land models of ERA5, MERRA2 and GLDAS. It is a set of algorithms dedicated to the estimation of terrestrial evaporation (Martens et al., 2017). Lines 117-120 read: "*Additionally, GLEAM is not a complex terrestrial model as found in land models of ERA5, MERRA2 and GLDAS, but a set of algorithms dedicated to estimating terrestrial evaporation using retrieved satellite observations including soil moisture, vegetation optical depth and snow-water equivalent, a multi-source precipitation product and relies on only radiation and temperature inputs from reanalysis products (Martens et al., 2017).*".

"GLEAM (Miralles et al., 2011a) (Global Land-Surface Evaporation: The Amsterdam Methodology) is derived from the inversion of multi-source remote sensing data, meteorological reanalysis data and the improved Priestley-Taylor (P-T) formula"
Surprisingly it is sometimes stated that GLEAM is an inversion or retrieval method but in my view it isn't. Inversion in my understanding is based on forward simulations of something observable from satellite, e.g. brightness temperatures, radiances etc. These forward simulations are based on a model (radiative-transfer) with multiple geophysical input variables. Minimising the difference between forward simulations and a satellite observation by assuming certain geophysical conditions is a retrieval based on inversion. GLEAM does no such thing.
GLEAM is a simple land surface model focusing on the estimation of evaporation and soil moisture. It's a traditional top-down approach with a model being fed with atmospheric input and land surface conditions (e.g. vegetation phenology). The estimation of evaporation is based around the Priestley-Taylor formula.
I'm not sure about the other two evaporation products, I would assume they are also specific models and not inversion schemes at all but please check.

Response:

Thank you for your cogent advice. GLEAM and the other two evaporation products are not inversion schemes. We have deleted the following description.

~~*"In recent years, multiple land evaporation data sets at global scales have become available from in situ observations and satellite inversion, such as MOD16 (Mu et al., 2007), GLEAM (Miralles et al., 2011a) and SSEBop (Senay et al., 2013), etc."*~~

"Thank you very much for your comment. Indeed, GLEAM is not the only one that contains soil moisture, however, GLEAM is the only product that uses satellite retrieved soil moisture to drive the model."
Satellite retrieved soil moisture does not drive GLEAM. GLEAM computes soil moisture at different levels based on soil properties, precipitation input etc. very similarly to the other models (similar in principle, not the exact formulas). Satellite

retrievals are assimilated with a very simple Newtonian Nudging scheme slightly correcting the modelled soil moisture. The impact of this assimilation is however mostly quite low. Therefore you can give equal credit to the other models with their repsective soil water modules.

Response:

Thank you very much for your comment. We will surely give equal credit to the other models with their respective soil water modules. Our own preliminary studies of their soil moisture modules have shown their commendable skill.

Further comments:

L43: I suppose SiF can be used for E although data quality is still not great (I'm no expert on this).

Response:

Thank you for your very thoughtful comment. Recently, solar-induced chlorophyll fluorescence (SIF) has been discovered as an emerging technique to observe the photosynthetic processes of vegetation by quantifying the emission of fluorescent radiation (Joiner et al., 2014). Remotely sensed SIF has potential to empirically track the variation of canopy-level transpiration (Lu et al., 2018; Shan et al., 2019). We have added the description to the revised manuscript. Lines 42-45 read: "*Recently, solar-induced chlorophyll fluorescence (SIF) has been discovered as an emerging technique to observe the photosynthetic processes of vegetation by quantifying the emission of fluorescent radiation (Joiner et al., 2014). Remotely sensed SIF has potential to empirically track the variation of canopy-level transpiration (Lu et al., 2018; Shan et al., 2019).*".

Reference:

Lu, X., Liu, Z., An, S., Miralles, D. G., Maes, W., Liu, Y, and Tang, J.: Potential of solar-induced chlorophyll fluorescence to estimate transpiration in a temperate forest, Agric. For Meteorol., 252, 75-87, https://doi.org/10.1016/j.agrformet.2018.01.017, 2018.

Shan, N., Ju, W., Migliavacca, M., Martini, D., Guanter, L., Chen, J., Goulas, Y., and Zhang, Y.: Modeling canopy conductance and transpiration from solar-induced chlorophyll fluorescence, Agric. For Meteorol., 268, 189-201, https://doi.org/10.1016/j.agrformet.2019.01.031, 2019.

Joiner, J., Yoshida, Y., Vasilkov, A. P., Schaefer, K., Jung, M., Guanter, L., Zhang, Y., Garrity, S., Middleton, E. M., Huemmrich, K. F., Gu, L., and Marchesini, L. B.: The seasonal cycle of satellite chlorophyll fluorescence observations and its relationship to vegetation phenology and ecosystem atmosphere carbon exchange, Remote Sens. Environ., 152, 375-391, https://doi.org/10.1016/j.rse.2014.06.022, 2014.

L47: satellite inversion is incorrect, definitely for GLEAM.

Response:

Thank you for that comment. We have deleted the incorrect description.

L117: "GLEAM is not a traditional terrestrial model as found in ERA5, MERRA2 and GLDAS" See above, I don't understand what is meant by this.

Response:

Thank you for pointing this out. We have changed "traditional" to "complex", and added the description of "complex" to the revised manuscript. GLEAM is not a complex terrestrial model as found in land models of ERA5, MERRA2 and GLDAS. It is a set of algorithms dedicated to the estimation of terrestrial evaporation (Martens et al., 2017). Lines 117-120 read: "*Additionally, GLEAM is not a complex terrestrial model as found in land models of ERA5, MERRA2 and GLDAS, but a set of algorithms dedicated to estimating terrestrial evaporation using retrieved satellite observations including soil moisture, vegetation optical depth and snow-water equivalent, a multi-source precipitation product and relies on only radiation and temperature inputs from reanalysis products (Martens et al., 2017).*".

I think it's still missing a clearer justification of using GLEAM (and the validation of GLEAM itself compared to the merged product and other individual ones).

Response:

Thank you for your very thoughtful comment. Firstly, ERA5, MERRA2 and GLDAS are based on climate models, while GLEAM is not. Secondly, GLDAS is driven by atmospheric observations, while GLEAM contains input data from land surface, such as soil moisture and vegetation optical depth. Therefore, GLEAM is the most independent in these ET data. We have added the validation of GLEAM itself compared to the merged product and other individual ones to the revised manuscript. Lines 345-377 read: "

Table 2. The verification results including R and RMSD between daily Ground-measured ET and daily ET from different products in different ecosystems. Values in bold indicates the highest quality.

Ecosystem type	*ERA5*		*MERRA2*		*GLDAS*		*GLEAM*		*REA*	
	R	*RMSD*	*R*	*RMSD*	*R*	*RMSD*	*R*	*RMSD*	*R*	*RMSD*
CRO	***0.66***	*1.24*	*0.55*	*1.42*	*0.60*	*1.48*	*0.60*	***1.22***	*0.60*	*1.38*
DBF	*0.76*	***1.06***	*0.71*	*1.23*	*0.74*	*1.19*	*0.67*	*1.16*	***0.77***	*1.07*
DNF	***0.81***	***0.55***	*0.73*	*0.75*	*0.77*	*0.73*	*0.64*	*0.75*	*0.80*	*0.62*
EBF	***0.72***	***1.08***	*0.61*	*1.59*	*0.58*	*1.36*	*0.70*	*1.11*	*0.65*	***1.08***
ENF	*0.66*	*1.03*	*0.66*	*1.21*	*0.67*	*1.05*	*0.66*	*1.04*	***0.73***	***0.88***
GRA	*0.72*	*1.05*	***0.77***	*1.11*	*0.70*	*1.09*	*0.73*	*0.96*	***0.77***	***0.94***
MF	*0.77*	***1.05***	***0.79***	*1.37*	*0.70*	*1.23*	*0.70*	*1.12*	*0.74*	*1.12*

OSH	0.43	1.00	0.47	0.92	0.46	0.96	0.33	1.15	**0.50**	**0.88**
SAV	0.61	1.23	0.62	1.40	0.63	1.22	0.58	1.25	**0.66**	**1.16**
WET	**0.57**	**1.40**	0.44	1.66	0.47	1.56	0.52	1.44	0.46	1.59
WSA	0.68	1.24	0.63	1.46	0.64	1.17	**0.72**	**1.11**	0.70	1.13

Verification of ET products from different ecosystems has been conducted in order to further evaluate their performances. Table 2 describes quantitatively the performances of the ET products in 11 ecosystem types of site on a daily scale from two indicators, R and RMSD. The values in bold print indicate the best performance of the four products. The results demonstrate that no individual product performs best across all ecosystems. For 42 ENF, 34 GRA, 9 OSH, 8 SAV sites, REA has higher R and lower RMSD than individual products. For 23 DBF and 13 EBF sites, REA has a optimal R or RMSD; specifically with the highest R of 0.77 and second lowest RMSD of 1.07 mm per day for DBF, and the lowest RMSD of 1.08 mm per day and the second highest R of 0.65 for EBF. REA performs worse than at least one individual product at 63 other sites. Specifically, REA has a lower R of 0.60 than ERA5, and a higher RMSD of 1.38 mm per day than GLEAM and ERA5 at 17 CRO sites. For 1 DNF site, REA has lower R of 0.80 and higher RMSD of 0.62 mm per day than ERA5. For 9 MF sites, REA has a lower R of 0.74 than ERA5 and MERRA2, and a higher RMSD of 1.12 mm per day than ERA5. For 19 WET sites, REA has a lower R of 0.46 and a higher RMSD of 1.59 mm per day than ERA5 and GLDAS. For 6 WSA sites, REA has lower R of 0.70 and higher RMSD of 1.13 mm per day than GLEAM. Generally, ERA5, MERRA2, GLEAM and REA show the best performance respectively in four (including CRO, DNF, EBF and WET), two (GRA and MF), one (WSA) and five (DBF, ENF, GRA, OSH and SAV) ecosystems in terms of R. Based on RMSD, both ERA5 and REA performed best in five ecosystems (with the former including DBF, DNF, EBF, MF and WET, and the latter including EBF, ENF, GRA, OSH and SAV). REA does not perform best across all ecosystems, however, it avoids the worst performance in any ecosystem. Taylor Diagram results of daily Ground-measured ET and ET from the different products in 11 ecosystems are put in support information (Fig. S2).

Table 3. The verification results including R and RMSD between monthly Ground-measured ET and monthly ET from different products in different ecosystems. Values in bold indicates the highest quality.

Ecosystem type	ERA5		MERRA2		GLDAS		GLEAM		REA	
	R	RMSD	R	RMSD	R	RMSD	R	RMSD	R	RMSD
CRO	0.71	31.84	0.58	38.33	0.64	39.59	0.66	**30.85**	**0.73**	31.14
DBF	0.84	26.61	0.77	32.40	0.83	30.27	0.75	28.43	**0.86**	**26.17**
DNF	**0.93**	**10.62**	0.84	18.63	0.91	18.96	0.85	14.92	0.91	11.01
EBF	0.81	25.85	0.71	40.32	0.69	32.21	**0.84**	**22.79**	0.78	27.06
ENF	0.74	25.01	0.72	31.69	**0.76**	25.38	**0.76**	24.24	**0.76**	**23.91**
GRA	0.77	27.11	**0.83**	28.59	0.77	26.54	0.81	**22.06**	0.77	27.53
MF	0.83	27.57	**0.85**	37.98	0.75	32.90	0.79	27.71	0.83	**26.99**

OSH	*0.45*	*25.34*	*0.51*	***23.39***	*0.52*	*23.75*	*0.27*	*32.57*	***0.53***	*24.54*
SAV	*0.65*	*32.05*	*0.65*	*38.08*	*0.67*	*31.88*	*0.59*	*34.33*	***0.68***	***31.45***
WET	*0.61*	*38.30*	*0.46*	*46.92*	*0.49*	*43.78*	*0.55*	*40.38*	***0.64***	***37.35***
WSA	*0.73*	*33.71*	*0.68*	*38.54*	*0.72*	*28.62*	***0.77***	***28.12***	*0.73*	*34.49*

Similar to Table 2, Table 3 shows the performance of ET products in different ecosystems on a monthly scale. Compared with daily scale, the performance of each product has changed, among which all of the R becomes higher. REA has higher R and lower RMSD than individual products at 23 DBF, 42 ENF, 8 SAV and 19 WET sites. It has an optimal R or RMSD at 17 CRO, 9 MF and 9 OSH sites. At other 64 sites, REA has a worse performance than at least one individual product. For 1 DNF site, REA has lower R and higher RMSD than ERA5. For 13 EBF sites, REA has lower R and higher RMSD than GLEAM and ERA5. For 34 GRA sites, ERA5 has a lower R of 0.77 than MERRA2 and GLEAM, and a higher RMSD of 27.53 mm per month than GLEAM, ERA5 and GLDAS. For 6 WSA sites, REA has a lower R of 0.73 than GLEAM, and a higher RMSD of 34.49 mm per day than GLEAM, GLDAS and ERA5. Similar to the daily scale, REA does not have a better performance than any individual product in all ecosystems, but is superior to at least one individual one. Similarly, the Taylor charts of monthly Ground-measured ET and ET from the different products in 11 ecosystems are put in support information (Fig. S3).".

L120: Monthly data ... for what purpose is this monthly data used?

Response:

Thank you for your very thoughtful comment. Previous studies show that there is a close relationship between the quality of land evaporation and vegetation (Miralles et al., 2016). Monthly GIMMS NDVI3g data is used to study how the quality of these land evaporation data sets change with vegetation, according to the correlation coefficients between multiple data sets and station-observed data under different vegetation conditions. We have added the description of the purpose of using this data. Lines 122-125 read: "*Monthly GIMMS NDVI3g data with a spatial resolution of 0.25° from the Global Inventory Modeling and Mapping Studies (GIMMS) was used to study how the quality of land evaporation data sets change with vegetation in our study (Pinzon & Tucker 2014), with the time span from 1982 to 2014, which is available from http://ecocast.arc.nasa.gov/data/pub/gimms/3g/.*".

L227: Is GLDAS a reanalysis? If yes, okay.

Response:

Thank you for pointing this out. GLDAS is a model-based product. We have changed "reanalysis" to "model-based" in the revised manuscript. Lines 229-230 read:

"*Reliability Ensemble Averaging (REA) method (Giorgi and Mearns, 2002; Xu et al., 2010) was used to combine multiple sets of model-based ET data into a single product.*".