Reply to Reviewers and Editor

Dear editorial board members and reviewers,

The authors would like to thank the editors and three reviewers reviewing this manuscript. The authors sincerely thank the reviewers' insightful and constructive comments and suggestions. We appreciate your time and effort in considering this manuscript for publication. According to your nice suggestions, we have checked carefully and incorporated these comments to the revised version. Follows are the key changes in this submission:

1. Bias correction of meteorological forcing inputs during 2019-2020 and remove the spatial and trend analysis in that period because the CMFD and the GLDAS-2.1 are different forcing datasets for driving the PML-V2 model;
2. The spatial analysis showing the ability of the PML-V2(China) product to identify the crop phenology has been added in the revised version;
3. The calibration using the genetic algorithm has been supplemented in materials and methods part;
4. The details of the PML-V2 model has been well-organized and put in the supplement; and
5. Minor issues have been corrected according to the comments from the reviewers and editors.

We reply to every comment point-by-point in the response letter. All comments are shown in blue. Sentences from the manuscript are in italics and the revised contents are indicated in red. We believe that the concerns from the reviewers have been addressed. Please let us know if there are any questions and queries. Thanks again for the editors and the reviewers for their valuable time, suggestions and comments.

Editor

Comments:

The study did a generally good job in producing GPP and ET simultanesly over China. After reading through the reviewers' comments and the authors' respones, I feel there are still two issues needing to be resolved:

The first is the inconsistency between GLDAS and CMFD meteorological data. If the authors would like to publish the GPP and ET products during 2019-2020, I would suggest to at least bias-correct GLDAS data in the period using CMFD data before 2018 as the differences in GPP and ET produced by the two datasets are considerable.

The second is that the authors should avoid statements that PMLv2 performs better than other models. As reviewers pointed out, the results that PMLv2 derived GPP and ET products showed better performance may arise from its calibration using observations at 27 flux sites, which are maybe inaccessble to other models. Meanwhile, I cannot agree to conclude that PMLv2(China) performs better than PMLv2(Global) simply because the former runs using daily inputs but without process improvements. I suggest that the authors only state the GPP and ET products produced in this study are better than other currently availabe data of the same type.

Response:

We appreciate your thoughtful and positive comments on our work. With the help of your constructive suggestions, we believe that this manuscript will be improved substantially. Following are our responses to your two questions:

1. Bias correction of GLDAS data in 2019-2020. Yes, we agree that using bias correction of GLDAS forcings can eliminate the subsequent bias in estimating ET and GPP. After a comprehensive comparison of various bias correction methods, a widely used methodology, delta change (i.e., DC, also called change factor methodology), was selected in this study (Anandhi et al., 2011; Teutschbein and Seibert, 2012; Rasmussen et al., 2012; Hempel et al., 2013; Beck et al., 2018; Haro-Monteagudo et al., 2020). The underlying idea of the DC method is to use simulated future anomalies (i.e., GLDAS-2.1 in this study) for a perturbation of observed data (i.e., CMFD) rather than to use the simulations of future conditions directly. For each grid cell, we bias-corrected the daily meteorological data during 2019-2020 by monthly scaling factors. The details for bias correction have also been added to the manuscript.
2. Internal comparison of PML-V2 versions. We agree that it is not appropriate to claim that PML-V2(China) is better than PML-V2(Global) simply because the former runs using daily local inputs but without process improvements. As such, we have changed the description of model performance in this revision based on your suggestions and the reviewers' comments.

References

Anandhi, A., Frei, A., Pierson, D. C., Schneiderman, E. M., Zion, M. S., Lounsbury, D., and Matonse, A. H.: Examination of change factor methodologies for climate change impact assessment, Water Resources Research, 47, W03501, https://doi.org/10.1029/2010WR009104, 2011.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future Köppen-Geiger climate classification maps at 1-km resolution, Sci Data, 5, 180214, https://doi.org/10.1038/sdata.2018.214, 2018.

Haro-Monteagudo, D., Palazón, L., and Beguería, S.: Long-term sustainability of large water resource systems under climate change: A cascade modeling approach, Journal of Hydrology, 582, 124546, https://doi.org/10.1016/j.jhydrol.2020.124546, 2020.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F.: A trend-preserving bias correction – the ISI-MIP approach, Earth Syst. Dynam., 4, 219–236, https://doi.org/10.5194/esd-4-219-2013, 2013.

Rasmussen, J., Sonnenborg, T. O., Stisen, S., Seaby, L. P., Christensen, B. S. B., and Hinsby, K.: Climate change effects on irrigation demands and minimum stream discharge: impact of bias-correction method, Hydrology and Earth System Sciences, 16, 4675–4691, https://doi.org/10.5194/hess-16-4675-2012, 2012.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, Journal of Hydrology, 456–457, 12–29, https://doi.org/10.1016/j.jhydrol.2012.05.052, 2012.

Reviewer 1

General Comments:

This study with a title of "A daily and 500m coupled evapotranspiration and gross primary production product across China during 2000-2020" has been seriously reviewed. Overall, this paper is well organized, including written English, structures, and the conclusions. Importantly, I believe that the PML-V2(China) product could provide a great opportunity for academic communities and various agencies for scientific studies and applications. However, before acceptance the authors should give the reasonable explanations to the following questions. So, I would like to recommend this paper to be conducted a major revision.

Response:

We appreciated tremendously your thoughtful comments and positive review on our article. According to your nice suggestions, we have checked and re-edited the original manuscript carefully. In the following, we reply to all comments in a point-by-point response. All comments are shown in blue. Sentences from the manuscript are in italics and the revised contents are indicated in red.

Specific Comments:

1. In the section 2.2.2, I found that the different meteorological forcings were used here, i.e., CMFD during 2000 to 2018, but GLDAS during 2019 to 2020. Although the authors compare the difference between PML-V2(China)$_{GLDAS-2.1}$ and PML-V2(China)$_{CMFD}$ at the national scale. However, the author did not compare the liner trends of these simulations. Maybe, the authors could add the evaluations of the linear trends of ML-V2(China)$_{GLDAS-2.1}$ and PML-V2(China)$_{CMFD}$ GPP and ET during 2000-2018 at different spatial scales (i.e., grid and national scales). Mainly because this product has a great potential to use for study the linear trends of GPP and ET by the scholars.

Response:

The China Meteorological Forcing Dataset (CMFD) was constructed by merging in situ measurements at 753 China Meteorological Administration stations with advanced retrospective analyses data from five remoting sensing or reanalysis data including Global Land Data Assimilation System (GLDAS) (He et al., 2020). We chose to use the CMFD dataset as meteorological inputs, because it shows much more accuracy and superior quality than other meteorological datasets in China, such as GLDAS meteorological data (He et al., 2020). We compared the magnitude and variability of the products using different meteorological forcing inputs, i.e., PML-V2(China)$_{GLDAS-2.1}$ and PML-V2(China)$_{CMFD}$, at the grid and national scale in section 4.4.2 of the first draft, as follows:

To extend the simulation period, we used GLDAS-2.1 meteorological forcing data during 2019-2020 since the CMFD dataset is only up to 2018. To check if using these two datasets generates a systematic bias, we reran the PML-V2(China) in 2001-2018 using GLDAS-2.1 and compared the modelling results with those obtained using CMFD (Fig. S1). At the *national scale, the mean difference, calculated by* $(PML\text{-}V2(China)_{GLDAS\text{-}2.1} - PML\text{-}V2(China)_{CMFD})/PML\text{-}V2(China)_{CMFD}$, *varied from -1.22% to 1.62% among* E_i, E_c, *and GPP, and was 13.72% for* E_s *and 7.78% for ET. The difference is within -25% ~ 25% in more than 66% of the research region for all five variables (Fig. S1b2-e2), specifically 100% for GPP, 95% for* E_c, *84% for ET, 73% for* E_i, *and 66% for* E_s *(Fig. S1b3-e3). This illustrates that PML-V2(China) using the GLDAS-2.1 in 2019-2020 does not generate a noticeable systematic deviation.*

The PML-V2(China) product of 2019-2020 is the interim data as the supplement of PML-V2(China) after 2018. We suggest that users do spatial variability analysis instead of trend analysis if they want to use the PML-V2(China) from 2019 to 2020. With the release of the meteorological dataset, we will continue to update the PML-V2(China) using the CMFD inputs. Moreover, we have removed the description and figures about the trend analysis of PML-V2(China) for 2019-2020 in the manuscript.

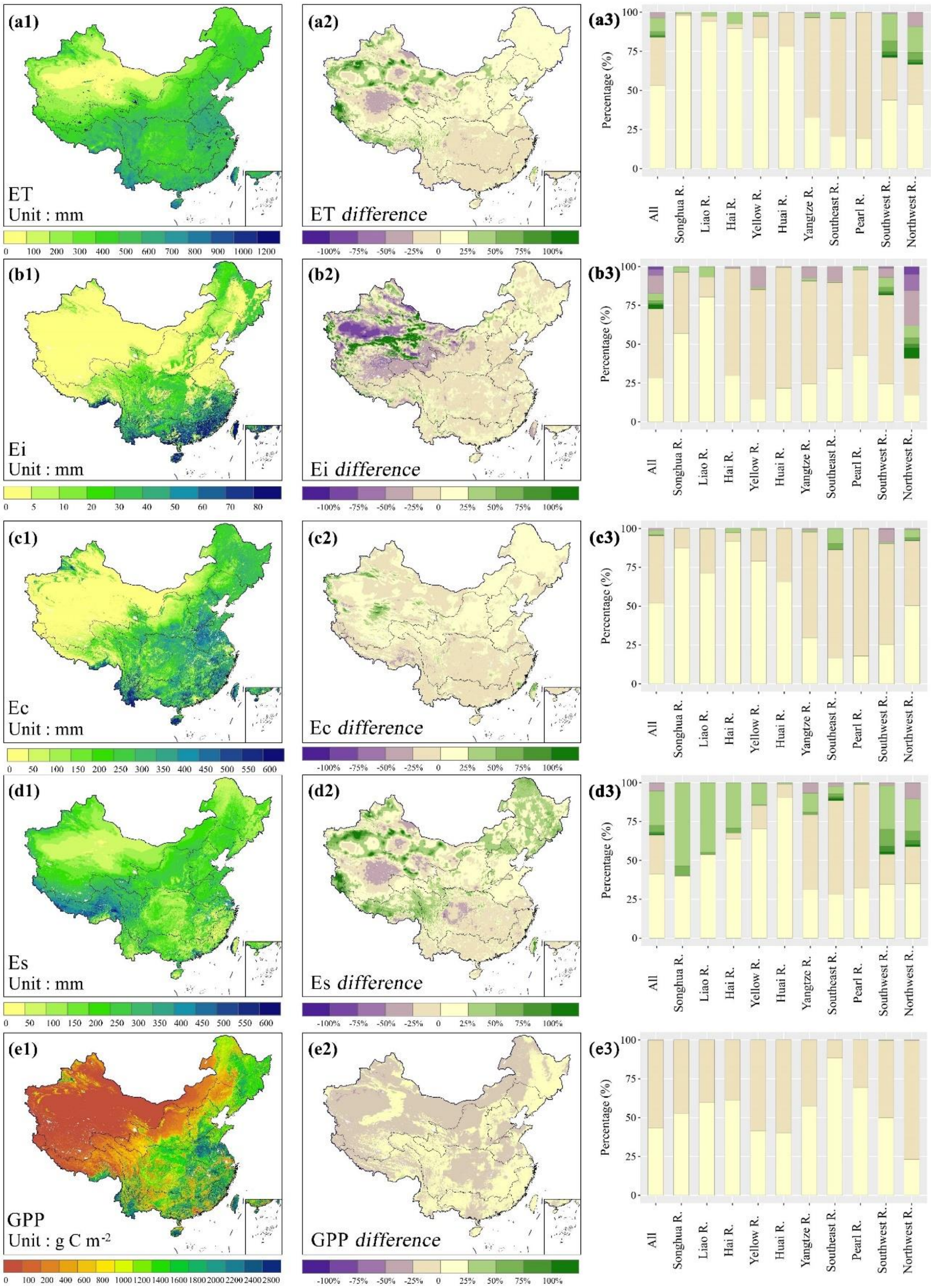

Figure S1: The modelling results using GLDAS-2.1 meteorological forcing data during 2001-2018 and comparison with the PML-V2(China) product using CMFD: (a1-e1) Spatial distribution of the 18-year mean of five variables; (a2-e2) Spatial distribution of the difference using two forcing datasets, calculated by $(PML\text{-}V2(China)_{GLDAS\text{-}2.1} - PML\text{-}V2(China)_{CMFD})/PML\text{-}V2(China)_{CMFD}$***; and (a3-e3) Proportion of difference in each river basin. 'ALL' represents the whole study area. The legends for (a3-e3) are the same as that for (a2-e2). Taking Fig.(a3) as an example, the area percentage of ET difference in 0 ~ 25% in the Songhua River Basin is about 99%.***

References

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., and Li, X.: The first high-resolution meteorological forcing dataset for land process studies over China, Sci Data, 7, 25, https://doi.org/10.1038/s41597-020-0369-y, 2020.

2. Line 180-181: The authors did not correct the energy imbalance issues within the EC observations? Although the authors stated that "correcting such a problem may also introduce more uncertainties (Foken, 2008)", I insist to think that not correcting the energy imbalance issues would like to greatly impact the estimated ET. Because to date there were so many studies to do the correction before using the EC observations.

Response:

In this study, we didn't correct the energy imbalance issues within the EC observations after considering the following facts:

- First, only 16 EC datasets with the soil heat flux (G) are open to access among the 26 EC observed shown in Table R1. In addition, a large number of net radiation (Rn) observations are missing. For instance, Rn data are missing during 2015 at the QZ-QOMS site around 72.8% (Ma et al., 2020). Considering the consistency of all the EC observed used for model calibration and validation, we use the observed data without energy balance correction.
- Second, we evaluated the difference between the latent heat flux (LE) without the energy balance correction and LE after the energy balance correction at the daily scale. There was little difference. For example, three site-dataset: the CN-Cng, CN-Du2, and CN-HaM, are posted in FLLUXNET (https://fluxnet.org/data/fluxnet2015-dataset/), not only having the latent heat flux item (LE_F_MDS) but also having the latent heat flux corrected by energy balance closure correction factor (LE_CORR). The overall determinable coefficient and bias are 0.94 and 11.08%, respectively (Fig. R1). Fig. R2-4 show the comparison of the two variables at the three sites on a daily scale. Both LE_F_MDS and LE_CORR replicate the seasonal variations well. Overall, LE_F_MDS can reasonably represent the magnitude and the seasonal variations of the latent heat flux on a daily scale.
- Third, observation errors coming from G and Rn may be introduced in calculating the latent heat flux and the sensible heat flux if we force the energy balance closure. Also, many studies used the LE without closing the energy balance (Zhang et al., 2019; Ma and Zhang, 2022). In that case, we chose to use the EC observations without the energy imbalance correction.
- Fourth, the independent water balance validations in the 10 large basins in China show that the PML-V2 model has no obvious bias in estimating annual evapotranspiration at a basin scale. This gives more confidence to us that our parameterization is reasonable.

Table R1: Details of 26 EC flux towers employed in this study.

Site code	Site name	IGBP	Time cover	Includes G?
ARCJZ	Arou	GRA	2013-2017	Yes

BNXJL	Xishuangbanna rubber	EBF	2013	No
CF-CBF	Chinaflux Changbai forest	MF	2003-2010	No
CF-HBG_S01	Chinaflux Haibei grassland	OSH	2003-2010	No
CF-HBG_W01	Chinaflux Haibei wetland	WET	2004-2006	No
CF-NMG	Chinaflux Neimengu grassland	GRA	2004	No
CF-QYF	Chinaflux Qianyanzhou forest	ENF	2004-2006	No
CF-YCA	Chinaflux Yucheng	CRO	2006-2007	No
CN-Cng	Changling	GRA	2007-2010	Yes
CN-Du2	Duolun_grassland (D01)	GRA	2006-2008	Yes
CN-HaM	Haibei Alpine Tibet site	GRA	2002-2004	Yes
DMCJZ	Daman	CRO	2017	Yes
DSLZ	Dashalong	WET	2015-2018	Yes
DXZ	Daxing	CRO	2010	Yes
DYKGTSLZ	Dayekouguantan forest	ENF	2010-2011	Yes
GTZ	Guantao	CRO	2008	No
HLZ	Huailai	CRO	2014	Yes
HZZHMZ	Huazhaizi Desert Steppe	BSV	2017	Yes
MYZ	Miyun	CRO	2008	Yes
QZ-BJ	Tibetan Plateau BJ	GRA	2011-2013	Yes
QZ-NAMORS	Tibetan Plateau NAMORS	GRA	2008-2009	No
QZ-QOMS	Tibetan Plateau QOMS	BSV	2015	Yes
YJGRHG	Yuanjiang dry-hot valley	SAV	2014	No
YKGQLZZ	Yingke	CRO	2011	Yes
YKZ	Yakou	GRA	2016-2018	Yes
ZYSDZ	Zhangye wetland	WET	2013-2018	Yes

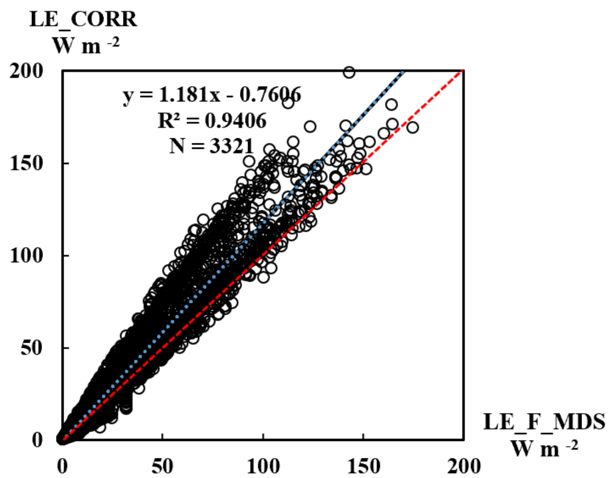

Figure R1: Scatterplot between the latent heat flux item (LE_F_MDS) and the latent heat flux corrected by energy balance closure correction factor (LE_CORR) at the daily scale.

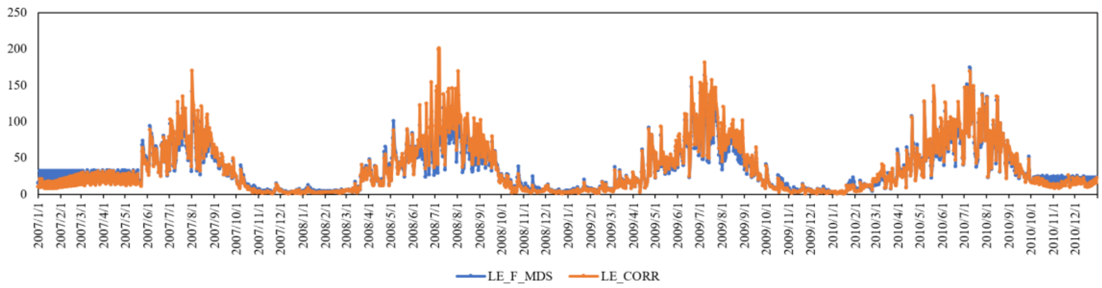

Figure R2: Comparison of LE_F_MDS and LE_CORR from January 2007 to December 2010 at the CN-Cng site on a daily scale.

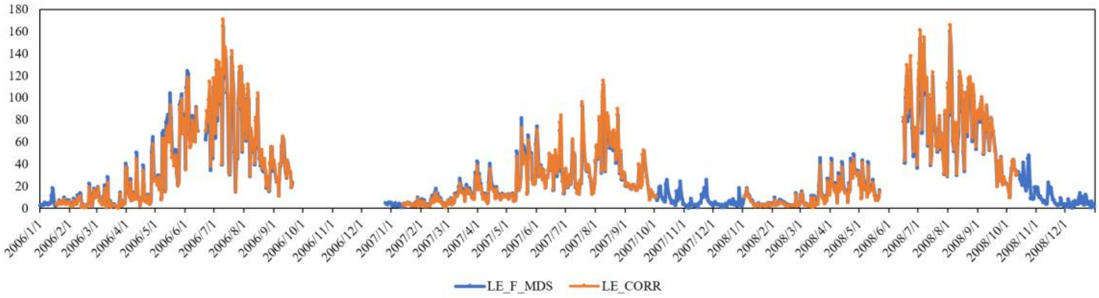

Figure R3: Comparison of LE_F_MDS and LE_CORR from January 2006 to December 2008 at the CN-Du2 site on a daily scale.

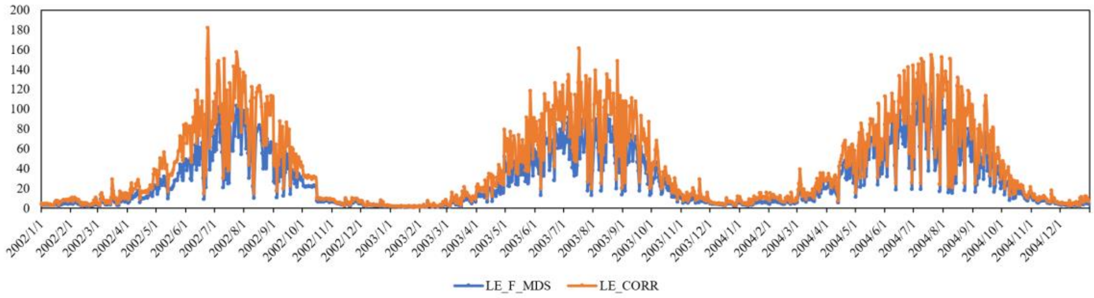

Figure R4: Comparison of LE_F_MDS and LE_CORR from January 2002 to December 2004 at the CN-HaM site on a daily scale.

References

Ma, Y., Hu, Z., Xie, Z., Ma, W., Wang, B., Chen, X., Li, M., Zhong, L., Sun, F., Gu, L., Han, C., Zhang, L., Liu, X., Ding, Z., Sun, G., Wang, S., Wang, Y., and Wang, Z.: A long-term (2005–2016) dataset of hourly integrated land-atmosphere interaction observations on the Tibetan Plateau, Earth Syst. Sci. Data, 12, 2937-2957, https://doi.org/10.5194/essd-12-2937-2020, 2020.

Ma, N. and Zhang, Y.: Increasing Tibetan Plateau terrestrial evapotranspiration primarily driven by precipitation, Agricultural and Forest Meteorology, 317, https://doi.org/10.1016/j.agrformet.2022.108887, 2022.

Zhang, Y., Kong, D., Gan, R., Chiew, F. H. S., McVicar, T. R., Zhang, Q., and Yang, Y.: Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017, Remote Sensing of Environment, 222, 165–182, https://doi.org/10.1016/j.rse.2018.12.031, 2019.

3. In section 2.6, the authors simply describe the calibration for the model. I think that the authors should added some necessary description about the calibration. For example, how did you determine the final parameters for each PFT? Please clarify.

Response:

We have revised sections 2.5 and 2.6 to supplement the calibration part, as follows:

2.5 Model calibration and model validation

The 11 parameters of the PML-V2 model for each PFT were calibrated and cross-validated against 26 EC sites by a global optimization method - genetic algorithm (GA). The GA generates a randomly initialized population and then evaluates the fitness of solutions according to its objective function. As generations iterate, the population includes more appropriate solutions, and eventually, it will converge (Holland, 1992; Konak et al., 2006). Specifically, we applied the GA algorithm with a population size of 1000 and number generations of 50. All EC-observed ET and GPP data within a PFT are used to minimize the following objective function (F_{opt}):

$$F_{opt} = 2 - NSE_{ET} - NSE_{GPP} = \frac{\sum_{i=1}^{N}(ET_{est}-ET_{obs})^2}{\sum_{i=1}^{N}\left(ET_{obs}-\overline{ET_{obs}}\right)^2} + \frac{\sum_{i=1}^{N}(GPP_{est}-GPP_{obs})^2}{\sum_{i=1}^{N}\left(GPP_{obs}-\overline{GPP_{obs}}\right)^2} \quad (8)$$

where NSE_{ET} and NSE_{GPP} are the Nash-Sutcliffe Efficiency of the daily ET and the daily GPP, respectively. The subscripts est and obs stand for the estimated and the observed, respectively. In this way, each of the nine PFTs gained a unique set with 11 calibrated parameter values, illustrated in Table S1.

The 'leave-one-out' cross-validation method was utilized to evaluate the robustness of the PML-V2 model (Zhang et al., 2019). For each PFT, the data from one "ungauged" observation was excluded from the optimization while the data from all other observations at the same PFT were used for model calibration to obtain the simulated at the "ungauged" position. All nine PFTs were actualized in this way. Note that the PFT including EBF, MF, OSH, and SAV only has one ground site (Table 2). Therefore, it is appropriate to divide the data in each of the four sites into two sub-groups for cross-validation. The CF-CBF and the CF-HBG_S01 covering from 2003 to 2010, were divided into two sub-groups, each of which had 4 years: 2003-2006 and 2007-2010. While both the BNXJL and YJGRHG only covered one year and were divided into two sub-groups by a two-day time step, separately. After that, the daily estimates in the cross-validation mode were against the daily observation from the 26 stations to explore the model transferability from known observations to any location.

2.6 Model performance metrics

We assessed the performance of calibration and cross-validation of PML-V2 (and other seven mainstream ET and GPP products) against the observed sites or water-balance basins utilizing the following four metrics:

$$NSE_X = 1 - \frac{\sum_{i=1}^{N}(X_{est}-X_{obs})^2}{\sum_{i=1}^{N}\left(X_{obs}-\overline{X_{obs}}\right)^2}, \quad (9)$$

$$R_X = \frac{\sum_{i=1}^{N}\left(X_{est}-\overline{X_{est}}\right)\left(X_{obs}-\overline{X_{obs}}\right)}{\sqrt{\sum_{i=1}^{N}\left(X_{est}-\overline{X_{est}}\right)^2 \times \sum_{i=1}^{N}\left(X_{obs}-\overline{X_{obs}}\right)^2}}, \quad (10)$$

$$RMSE_X = \sqrt{\frac{\sum_{i=1}^{N}(X_{est}-X_{obs})^2}{N}}, \quad (11)$$

$$Bias_X = \frac{\sum_{i=1}^{N}(X_{est}-X_{obs})}{N \times \overline{X_{obs}}}, \quad (12)$$

where NSE, R, $RMSE$, and $Bias$ are the Nash-Sutcliffe Efficiency, the correlation coefficient, the Root Mean Square Error, and the ratio of the difference between the estimated and the observed to the observed average. The subscript X represents ET or GPP; the subscripts est and obs stand for the estimated and the observed, respectively.

References here are the same as those in the manuscript.

4. In section 3.1, was the estimated ET and GPP based on the EC observational meteorological variables? How the calibrated model perform at EC sites when the model was run with the CMFD forcings?

Response:

Thank you very much for your careful reading.

(1) The ET and GPP were estimated based on the model with the parameters, which were calibrated by the observed ET and GPP from the EC station, LST data from ERA5-Land and other meteorological variables from CMFD, and MODIS inputs for LAI, albedo, and emissivity. For clarity, we have expanded the first sentence in Section 2.3 as follows:

We collated EC flux towers and automatic weather stations (AWSs) data from 26 sites across China (Fig. 2 and Table 2) and generated the high-quality ET and GPP observed for calibration and validation of PML-V2.

(2) We posted the model calibration performance in row 1 and row 3 of figure 3 and described it in section 3.1 as follows:

Overall, PML-V2(China) shows an excellent performance in estimating daily ET and daily GPP, as evidenced by the NSE (0.75 and 0.82, respectively), R (0.88 and 0.9, respectively), RMSE (0.69 mm d^{-1} and 1.71 g C m^{-2} d^{-1}, respectively), and Bias (-5.81% and -2.3%, respectively). For the mean values of each site, the simulated daily ET and daily GPP show higher NSE (≥ 0.87) and R (≥ 0.93) values (Fig. 3).

5. The used hydrological sites should shown in figure 1. Considering the high spatial resolution, the validation may be better at the small basins rather the water resources regions (i.e., Yangtze River Basin, Yellow River Basin, and so on). Additionally, the linear trends of the PML-V2 ET should be compared with the water balance-based ET at the basin scale.

Response:

We didn't use hydrological sites in this study, because the basin-wide runoff data and basin boundaries have been provided in the National Water Resources Bulletin. We chose to use the water resources regions for the validation of the five ET products since (i) the PML-V2(China) product is country-wide, so it is more comprehensive to use the ten river basins covering most of China to test its performance; and (ii) although water-balance data can be tested in small basins, based on the fact that the change of terrestrial water storage is more accurate in large basins, it may be misleading for validation in small basins.

We compared the linear trends between PML-V2(China) ET and the water balance-based ET on the ten river basins, shown in Fig. R5. The ET trend of the Liao basin has the best consistency with ET_{wb} demonstrated as a bias of 3.73%, followed by Southeast, Songhua, and Yangtze. The Southwest basin gets the worst result with a bias of -403.92%, followed by Huai, Pearl, Northwest, Hai, and Yellow. But it is not statistically significant (p >0.05) among the linear trends of the ET based on the water balance of the ten river basins (Fig. R5). The span of only 11 years is too short to analyze its long-term trend, resulting in huge uncertainty. In that case, we didn’t add the linear trends of the PML-V2 ET and the water balance-based ET at the basin scale.

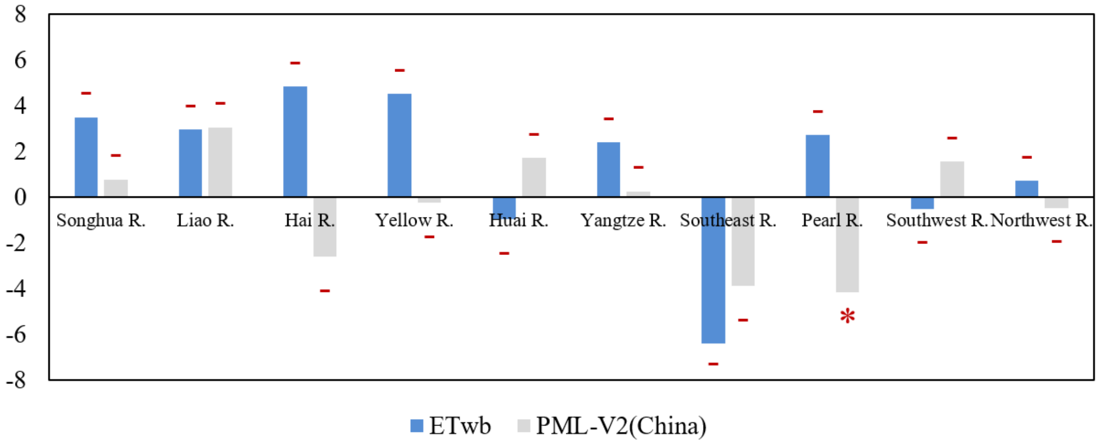

Figure R5: The linear trend bars between with PML-V2(China) ET and the water balance-based ET on the ten river basins. Note that “-” indicates that the *p*-value of the t-test for trend analysis is not less than 0.05. Similarly, “*” means 0.001≤*p*<0.05, and “” means *p*<0.001.**

Reviewer 2

General Comments:

He et al. constructed daily and 500m ET and GPP datasets in China using PML-V2. Compared with previous products, this model outputs improved in several aspects, including 26 EC sites being used for model calibration and validation, country-specific meteorological forcing, daily data, and intra-annual dynamics for multiple ecosystems. This ambitious work provides valuable data products for assessing the carbon and water cycles in China. They may also provide guidance in agricultural production and ecosystem management. The authors may consider the following suggestions to improve the robustness of this manuscript.

Response:

Thank you for appreciating our work and considering that the products are very valuable. We have carefully checked and re-edited the original manuscript. In the following, we reply to all comments in a point-by-point response. All comments are shown in blue. Sentences from the manuscript are in italics and the revised contents are indicated in red.

Specific Comments:

1. Line 99, the whole name for CMFD should be provided when it is first mentioned in the text.

Response:

We have added the whole name - the China Meteorological Forcing Dataset for the CMFD dataset.

2. Line 144-146, this sentence is not appropriate. You may use the MODIS land cover product, but it is debatable if it has the highest accuracy in China since there are many recently released land use/covered datasets with a high spatial resolution (30m and 10m). Many MODIS products based on the MODIS land use dataset may have low credibility in regions with complex terrain such as in the Loess Plateau.

Response:

We revised the sentence as follows:

Here we used the International Geosphere-Biosphere Program (IGBP) layer of MCD12Q1.006 land cover product (Sulla-Menashe et al., 2019) during 2000-2020 since IGBP classification is annually continuous and has acceptable accuracy in China when compared with other land cover products (Feng and Bai, 2019).

3. Did you test the continuity between GLDAS-2.1 and CMFD?

Response:

Currently, we compared the magnitude and variability of the products using different meteorological forcing inputs, i.e., PML-V2(China)$_{\text{GLDAS-2.1}}$ and PML-V2(China)$_{\text{CMFD}}$, at the grid and national scale in section 4.4.2 of the first draft, as follows:

To extend the simulation period, we used GLDAS-2.1 meteorological forcing data during 2019-2020 since the CMFD dataset is only up to 2018. To check if using these two datasets *generates a systematic bias, we reran the PML-V2(China) in 2001-2018 using GLDAS-2.1 and compared the modelling results with those obtained using CMFD (Fig. S1). At the national scale, the mean difference, calculated by* $(PML\text{-}V2(China)_{GLDAS\text{-}2.1} - PML\text{-}V2(China)_{CMFD})/PML\text{-}V2(China)_{CMFD}$*, varied from -1.22% to 1.62% among* E_i*,* E_c*, and GPP, and was 13.72% for* E_s *and 7.78% for ET. The difference is within -25% ~ 25% in more than 66% of the research region for all five variables (Fig. S1b2-e2), specifically 100% for GPP, 95% for* E_c*, 84% for ET, 73% for* E_i*, and 66% for* E_s *(Fig. S1b3-e3). This illustrates that PML-V2(China) using the GLDAS-2.1 in 2019-2020 does not generate a noticeable systematic deviation.*

The PML-V2(China) product of 2019-2020 is the interim data as the supplement of PML-V2(China) after 2018. We suggest that users do spatial variability analysis instead of trend analysis if they want to use the PML-V2(China) of 2019-2020. With the release of the meteorological dataset, we will continue to update the PML-V2(China) using the CMFD inputs. Moreover, we have removed the description and figures about the trend analysis of PML-V2(China) for 2019-2020 in the manuscript.

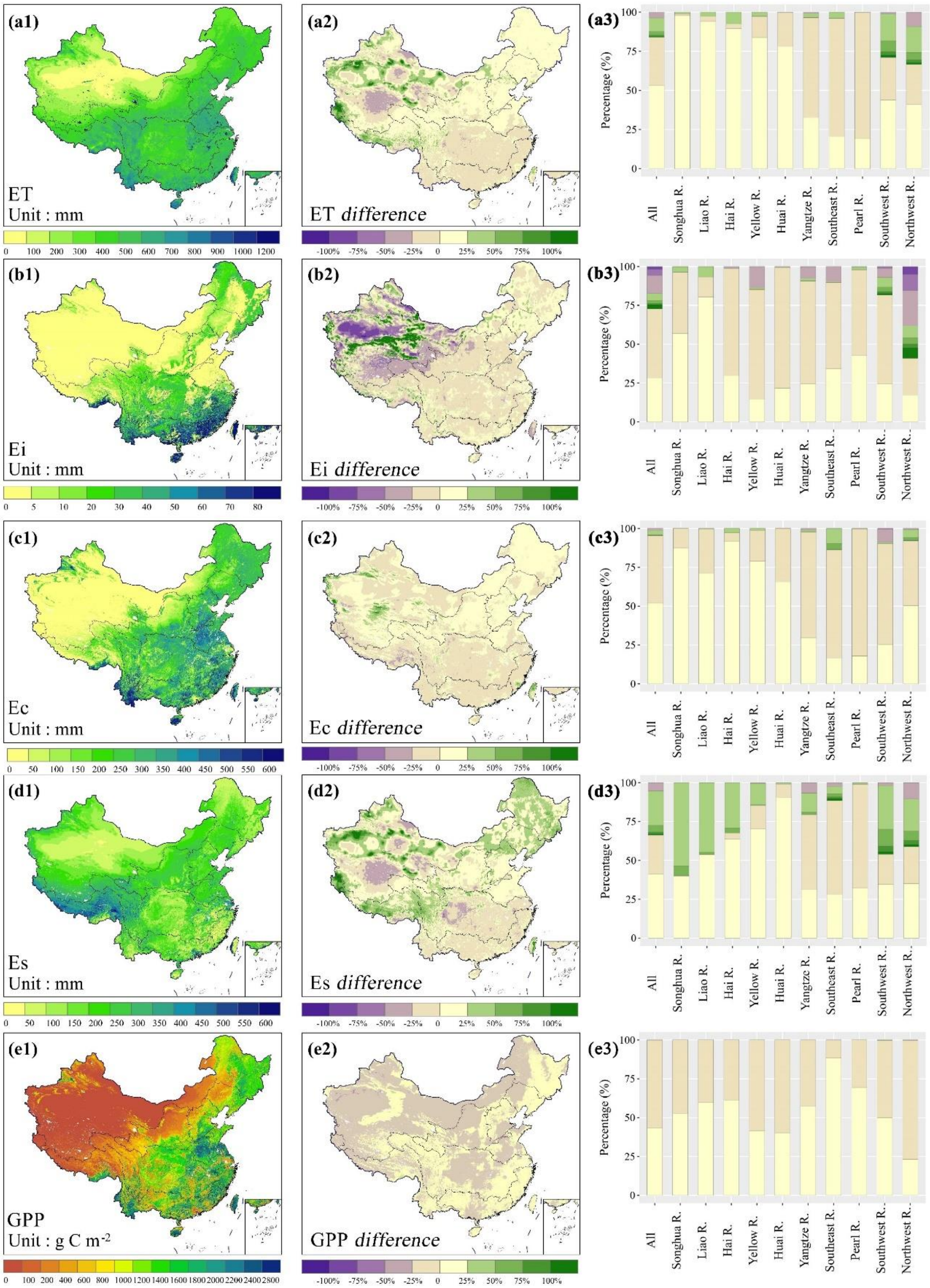

Figure S1: The modelling results using GLDAS-2.1 meteorological forcing data during 2001-2018 and comparison with the PML-V2(China) product using CMFD: (a1-e1) Spatial distribution of the 18-year mean of five variables; (a2-e2) Spatial distribution of the difference using two forcing datasets, calculated by $(PML\text{-}V2(China)_{GLDAS\text{-}2.1} - PML\text{-}V2(China)_{CMFD})/PML\text{-}V2(China)_{CMFD}$***; and (a3-e3) Proportion of difference in each river basin. 'ALL' represents the whole study area. The legends for (a3-e3) are the same as that for (a2-e2). Taking Fig.(a3) as an example, the area percentage of ET difference in 0 ~ 25% in the Songhua River Basin is about 99%.***

References

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., and Li, X.: The first high-resolution meteorological forcing dataset for land process studies over China, Sci Data, 7, 25, https://doi.org/10.1038/s41597-020-0369-y, 2020.

4. Section 2.6, the simulated model outputs were validated at the EC site-level, and compared with the publicly available dataset. How did you get the parameter set for a certain land use type? Did all the land use types have a unique parameter set? Did you run the model at each site?

Response:

(1) For each land use type, we used a global optimization method - genetic algorithm to gain the optimal solution by setting population size 1000 and number of generations 50 by minimizing an objective function including ET_{obs} and GPP_{obs}. (2) Each of the nine land use types has a unique parameter set, so they are nine parameter sets. (3) Yes, we run the model at each site. We revise sections 2.5 and 2.6 to make the model calibration and model validation parts clearer, as follows:

2.5 Model calibration and model validation

The 11 parameters of the PML-V2 model for each PFT were calibrated and cross-validated against 26 EC sites by a global optimization method - genetic algorithm (GA). The GA generates a randomly initialized population and then evaluates the fitness of solutions according to its objective function. As generations iterate, the population includes more appropriate solutions, and eventually, it will converge (Holland, 1992; Konak et al., 2006). Specifically, we applied the GA algorithm with a population size of 1000 and number generations of 50. All EC-observed ET and GPP data within a PFT are used to minimize the following objective function (F_{opt}):

$$F_{opt} = 2 - NSE_{ET} - NSE_{GPP} = \frac{\sum_{i=1}^{N}(ET_{est}-ET_{obs})^2}{\sum_{i=1}^{N}(ET_{obs}-\overline{ET_{obs}})^2} + \frac{\sum_{i=1}^{N}(GPP_{est}-GPP_{obs})^2}{\sum_{i=1}^{N}(GPP_{obs}-\overline{GPP_{obs}})^2} \quad (8)$$

where NSE_{ET} and NSE_{GPP} are the Nash-Sutcliffe Efficiency of the daily ET and the daily GPP, respectively. The subscripts est and obs stand for the estimated and the observed, respectively. In this way, each of the nine PFTs gained a unique set with 11 calibrated parameter values, illustrated in Table S1.

The 'leave-one-out' cross-validation method was utilized to evaluate the robustness of the PML-V2 model (Zhang et al., 2019). For each PFT, the data from one "ungauged" observation was excluded from the optimization while the data from all other observations at the same PFT were used for model calibration to obtain the simulated at the "ungauged" position. All nine PFTs were actualized in this way. Note that the PFT including EBF, MF, OSH, and SAV only has one ground site (Table 2). Therefore, it is appropriate to divide the data in each of the four sites into two sub-groups for cross-validation. The CF-CBF and the CF-HBG_S01 covering from 2003 to 2010, were divided into two sub-groups, each of which had 4 years: 2003-2006 and 2007-2010. While both the BNXJL and YJGRHG only covered one year and were divided into two sub-groups by a two-day time step, separately. After that, *the daily estimates in the cross-validation mode were against the daily observation from the 26 stations to explore the model transferability from known observations to any location.*

2.6 Model performance metrics

We assessed the performance of calibration and cross-validation of PML-V2 (and other seven mainstream ET and GPP products) against the observed sites or water-balance basins utilizing the following four metrics:

$$NSE_X = 1 - \frac{\sum_{i=1}^{N}(X_{est}-X_{obs})^2}{\sum_{i=1}^{N}\left(X_{obs}-\overline{X_{obs}}\right)^2}, \quad (9)$$

$$R_X = \frac{\sum_{i=1}^{N}\left(X_{est}-\overline{X_{est}}\right)\left(X_{obs}-\overline{X_{obs}}\right)}{\sqrt{\sum_{i=1}^{N}\left(X_{est}-\overline{X_{est}}\right)^2 \times \sum_{i=1}^{N}\left(X_{obs}-\overline{X_{obs}}\right)^2}}, \quad (10)$$

$$RMSE_X = \sqrt{\frac{\sum_{i=1}^{N}(X_{est}-X_{obs})^2}{N}}, \quad (11)$$

$$Bias_X = \frac{\sum_{i=1}^{N}(X_{est}-X_{obs})}{N \times \overline{X_{obs}}}, \quad (12)$$

where NSE, R, $RMSE$, and $Bias$ are the Nash-Sutcliffe Efficiency, the correlation coefficient, the Root Mean Square Error, and the ratio of the difference between the estimated and the observed to the observed average. The subscript X represents ET or GPP; the subscripts est and obs stand for the estimated and the observed, respectively.

References here are the same as those in the manuscript.

Reviewer 3

General Comments:

1. This study used the PML-V2 model to develop ET and GPP datasets in China. The PML-V2 is calibrated and validated based on the data from 26 eddy covariance flux towers. The GPP and ET data developed in this study are compared with other global ET and GPP products and water balance data at the regional level. This study did a good job on model validation, but there still exist some issues in this stage.

Response:
Thank you for your very positive overall evaluation of the manuscript. We have carefully checked and re-edited it. In the following, we reply to all comments in a point-by-point response. All comments are shown in blue. Sentences from the manuscript are in italics and the revised contents are indicated in red.

2. As a data description paper, the methodology is an important section to let the audience know how the data is developed. However, the model description is not very clear and well organized in this paper. Although the PML-V2 model is already described in other papers, I think more details are still needed and could be put in the supplementary. Whether the code of the model is open source? If yes, a link to the model program should be provided. Why the PML-V2 (China) can simulate daily scale data while PML-V2 (Global) cannot? Are there any improvements in the model?

Response:
All the details of the PML-V2 model have now been reorganized and put in the supplement part. The PML-V2(China) source code is available through the public GitHub repository (https://github.com/SylviaHeee/PML-V2-China). First, PML-V2(China) uses a new parameter set for the country-wide simulation based on the daily EC observed, while PML-V2 (Global) uses the global parameters resolution that performs not well compared to PML-V2(China) at the plot scale and also at the basin-scale, shown in this manuscript. Second, PML-V2(China) uses daily input data while PML-V2 (Global) uses those at the 8-day scale. Third, the country-specific meteorological forcing, i.e., the China Meteorological Forcing Dataset (CMFD), is used to drive the PML-V2 in China, which is more accurate than those forcings extracted from global forcing products. Fourth, PML-V2(China) uses land surface temperature data, ERA5-Land, as the input surface temperature instead of air temperature like PML-V2 (Global) choosing to calculate the outgoing longwave radiation. Firth, PML-V2(China) utilizes the MODIS leaf area index data after the improved Whittaker filter and reveals the characteristics of the planting system.

3. The authors claimed the PML-V2 model performed better than other products. The evidence of the high accuracy of the ET and GPP mainly comes from the validation results at 26 EC sites. The 26 EC sites were used to calibrate and validate the model, while other global products did not calibrate and validate based on the same EC sites. If the PML-V2 and other products were used to compare against other new EC sites (not the 26 sites), can it still be the best one? It seems a little bit unfair to claim that this dataset is better than others when other models cannot

access these EC data. I encourage authors to also publish these EC data that are used in validation.

Response:

Here we use the water-balance ET to compare the accuracy of several products and it can be regarded as an independent validation since the PML-V2 (China) is not calibrated against the water-balance ET. The PML-V2 (China) model estimates get the smallest Bias of 6.28% and the highest NSE of 0.82 against water-balance annual ET estimates across 10 major river basins in China among five ET products.

The copyright of the EC data used in this study belongs to principal investigators of the EC stations. Therefore, we have no right to publish them. The EC data are all free to access from platforms on the Internet and their download links are provided in Table 2 and the references part of the manuscript.

4. According to the distribution of 26 EC sites (Fig 2), most of them are located in arid regions where ET may be low. The total estimated ET in China may be controlled by the ET estimated in the south region where few EC sites are located. There may exist large uncertainties in quantifying total ET.

Response:

We used the 26 EC-observed classified according to various plant function types (PETs) to get the 11 calibrated parameters for each PFTs, not based on climate types. Every PFTs have no less than one EC site. Besides, China has one of the largest dryland areas worldwide about 6.6 million km^2 which covers 68.8% of the country (Prăvălie, 2016; Li et al., 2021). This shows the rationality of using more EC sites in arid areas. In addition, we have reselected the color ramps for the aridity index (AI) map based on United Nations Convention to Combat Desertification (as shown in Fig. R1 below), because the original AI colors between 0.6 and 1 were set as different degrees of yellow which may mislead readers that China has too much dryland (Fig 2 in the original manuscript).

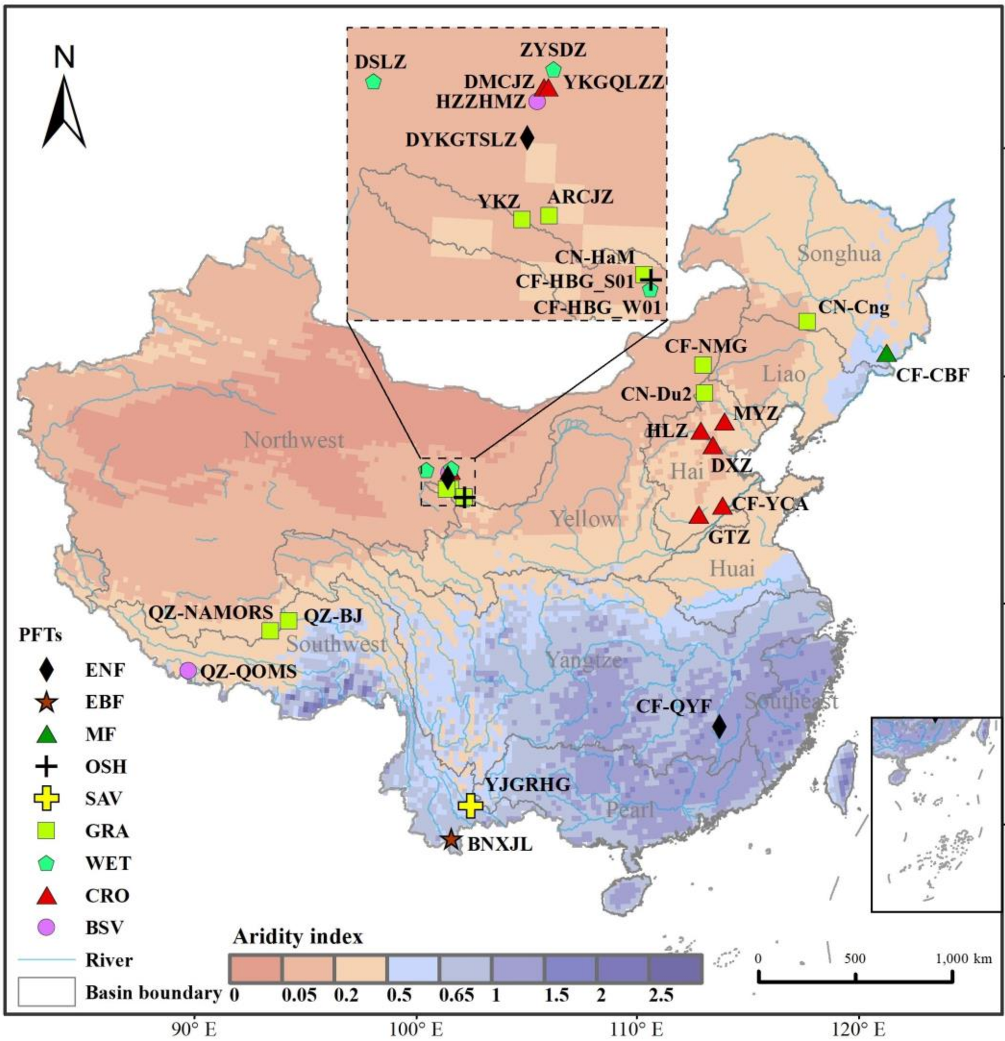

Figure R1: Geographical locations of 26 EC flux towers for nine major IGBP PFTs, the main rivers, and the ten major river basins in China. Overlain are 20-year mean annual aridity index (AI) values during 2001-2020 using GLDAS-2.1, that is, the ratio of annual precipitation to Penman potential evapotranspiration. PFTs shown in legend are ENF (Evergreen Needleleaf Forests), EBF (Evergreen Broadleaf Forests), MF (Mixed Forests), OSH (Open Shrublands), SAV (Savannas), GRA (Grasslands), WET (Permanent Wetlands), CRO (Croplands), and BSV (Barren Sparse Vegetation).

References

Li, C., Fu, B., Wang, S., Stringer, L. C., Wang, Y., Li, Z., Liu, Y., and Zhou, W.: Drivers and impacts of changes in China's drylands, Nat Rev Earth Environ, 2, 858–873, https://doi.org/10.1038/s43017-021-00226-z, 2021.

Prăvălie, R.: Drylands extent and environmental issues. A global approach, Earth-Science Reviews, 161, 259–278, https://doi.org/10.1016/j.earscirev.2016.08.003, 2016.

5. In the discussion section, two advantages of this new dataset are provided, one is the water-carbon coupled process, and the other is more EC data help constrain the parameters. How are water and carbon coupled in the model? And why does the coupled carbon process help advance

the model? There are many land surface models that couple water and carbon processes, but it is not always the case that these models performed better in simulating ET.

Response:

PML-V2 adopted coupling a photosynthesis model (Farquhar et al., 1980) and an improved canopy stomatal conductance model (Yu et al., 2004) with the Penman-Monteith (P-M) equation to estimate GPP and transpiration from the plant canopy (E_c) collectively (Gan et al., 2018). Detailed descriptions of PML-V2 have been provided in the revised supplement. The most important fact is that E_c and GPP processes should be coupled through stomata. Not coupling these two processes can cause the following issues: (i) internal inconsistency between ET and GPP estimates if their forcing data are not the same; (ii) inaccurate causality analysis for mean annual values, trends/variation, and water use efficiency. To better understand the influence of carbon-constrained impacts on evapotranspiration, it is critical to credibly couple ET and GPP products at moderate spatial resolution (Zhang et al., 2019; Ma et al., 2022). PML-V2 is a water-carbon coupled but a parsimonious model with only 11 parameters. For most land surface models, they contain much more parameters that are hard to calibrate, which may cause uncertainties in simulating ET.

References

Ma, N. and Zhang, Y.: Increasing Tibetan Plateau terrestrial evapotranspiration primarily driven by precipitation, Agricultural and Forest Meteorology, 317, https://doi.org/10.1016/j.agrformet.2022.108887, 2022.

Zhang, Y., Kong, D., Gan, R., Chiew, F. H. S., McVicar, T. R., Zhang, Q., and Yang, Y.: Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017, Remote Sensing of Environment, 222, 165–182, https://doi.org/10.1016/j.rse.2018.12.031, 2019.

6. The daily data is one important advantage of this dataset. But there are no details of how daily data is better than the data at the 8-day scale.

Response:

Here we emphasized the advantage and implications of daily data against the data at the 8-day scale is an improvement of temporal resolution of PML-V2(China) compared to current mainstream products with higher simulation accuracy. The cross-validated statistical indicators of the daily GPP estimated by PML-V2(China) at 26 EC flux towers have been added in Table 3. It is evident that PML-V2(China) at a daily scale excels its global version at the 8-day scale, rendered by *RMSE* being 0.48 mm d^{-1} lower for ET, and 1.30 g C m^{-2} d^{-1} lower for GPP, *NSE* being 0.04 higher for ET and 0.08 higher for GPP and *R* being 0.04 higher for ET and 0.05 higher for GPP. Then in the 4.3 section, we discussed the implications of the daily products. For instance, daily outputs from PML-V2(China) can be better used by the agricultural and water sectors for operational applications. Timely access to daily data at the regional or national scale helps the Ministry of Agriculture and Water Resources to develop better policies. Indeed, there is a remarkable relationship between soil water content and ET (Graf et al., 2014; Brust et al., 2021), so getting daily ET information accurately is of great significance for soil water depletion assessment, irrigation system design, and water resources management in agricultural areas, such as in the North China Plain.

Table 3: Statistical indicators of PML-V2(China) and other models for simulating ET and GPP at 26 EC flux towers. NSE and R values are unitless. The unit of RMSE for ET is mm d^{-1} while it is g C m^{-2} d^{-1} for GPP. The unit of Bias is %.

Scale	Variable	Models	NSE	R	RMSE	Bias
daily	ET	PML-V2(China)	0.66	0.84	0.33	-7.97
		GLEAM	0.44	0.69	1.04	-14.45
		SEBAL	-7.10	0.16	3.95	5.31
8-day	ET	PML-V2(China)	0.74	0.87	0.66	-11.54
		PML-V2(Global)	0.62	0.80	0.81	-5.05
		MOD16A2	0.37	0.63	1.07	-10.90
daily	GPP	PML-V2(China)	0.76	0.87	0.87	-0.82
8-day	GPP	PML-V2(China)	0.75	0.87	1.93	-6.51
		PML-V2(Global)	0.68	0.82	2.17	-1.74
		MOD17A2H	0.49	0.78	2.74	-38.79
		EC-LUE	-0.04	0.35	3.91	-41.91
		VPM	0.21	0.60	3.41	-8.21

Specific Comments:

1. L60. The 8-day scale data is enough to detect seasonal changes.

Response:

We agree with you. The sentence has been modified and an appropriate reference has also been added below.

For instance, products with low temporal resolutions are erratic to detect subtle seasonal changes in areas seriously affected by human activities and in arid regions, such as irrigated farmland with a dry climate (Bodner et al., 2015) and an evergreen broad-leaf Mediterranean forest during severe summer drought (Liu et al., 2015).

References

Bodner, G., Nakhforoosh, A., and Kaul, H.-P.: Management of crop water under drought: a review, Agron. Sustain. Dev., 35, 401–442, https://doi.org/10.1007/s13593-015-0283-4, 2015.

Liu, J., Rambal, S., and Mouillot, F.: Soil Drought Anomalies in MODIS GPP of a Mediterranean Broadleaved Evergreen Forest, Remote Sensing, 7, 1154–1180, https://doi.org/10.3390/rs70101154, 2015.

2. L68-70. Whether this dataset has a better performance in simulating WUE. Different data sources of GPP and ET do not necessarily mean high uncertainties. If a water-carbon model is used to estimate GPP and ET, other information such as nutrient limitation may be lost, therefore, the estimated GPP may not be more accurate than directly observed data.

Response:

Figure R2 summarizes PML-V2 performance when estimating annual WUE for whole ecosystems at the 95 global flux sites, in comparison to other model performance, i.e., FluxCom GPP/GLEAM ET, VPM GPP/GLEAM ET, and MOD17 GPP/MOD16 ET (Zhang et al., 2019). PML-V2 performs reasonably well in estimating annual total WUE, indicated by the statistical metrics: NSE = 0.48, R^2 = 0.49, RMSE = 0.86 g C mm^{-1} H_2O, Bias = 3.3%. Furthermore, PML-V2 is much better than the combinations of other products for estimating ecosystem WUE. This result indicates the benefit of using the coupled PML-V2 model for estimating ecosystem WUE as the use of the coupled GPP/ET models avoids internal inconsistencies between independent ET and GPP models which provides strong motivation for this research.

We agree that the estimated GPP may not be more accurate than directly observed data. But there are sparse and short-period ground observations in China, continuous gridded GPP products are needed to understand the spatial and temporal pattern of GPP. The paragraph from lines 68 to 70 has a chief sentence at first: *secondly, the phenomenon of ignoring the water-carbon coupling process frequently appearing in the existing products has brought systematic errors.* So, in this paragraph, we just introduced a problem in the existing ET and GPP gridded models. Moreover, the nutrient limitation for estimating GPP and ET deserves further study in the future.

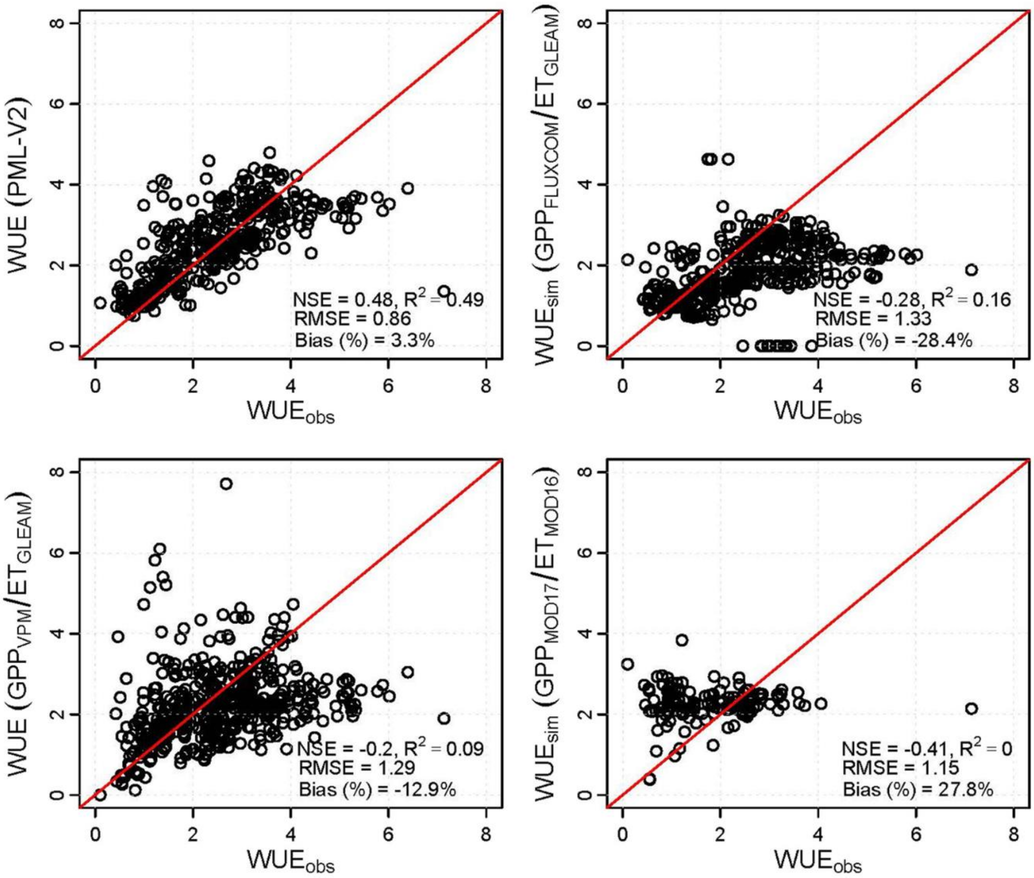

Figure R2: Scatterplots between the observed annual WUE (GPP/ET, g C mm^{-1} H_2O) and simulated by PML-V2 model at the 95 global flux sites.

References

Zhang, Y., Kong, D., Gan, R., Chiew, F. H. S., McVicar, T. R., Zhang, Q., and Yang, Y.: Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017, Remote Sensing of Environment, 222, 165–182, https://doi.org/10.1016/j.rse.2018.12.031, 2019.

3. L86. "While" is not proper here.

Response:
It has now been revised to "On the other hand".

4. L109. Please provide the name of Ec, Es and Ei.

Response:
Added as you suggested: plant transpiration (E_c), evaporation from the soil (E_s), and canopy evaporation from precipitation interception (E_i).

5. Equation 1-6. Reorganize the description of these equations. Separate equations and descriptions rather than list them together.

Response:
We have reorganized the detailed descriptions of PML-V2 and provided them in the revised supplement.

6. L151. Is the climate data publicly available? Maybe provide a data source link.

Response:
Yes, and the CMFD data source link has been provided in the acknowledgements part like other data sources or links, as follows:
We appreciate the China Meteorological Forcing Dataset shared by He et al. at https://doi.org/10.6084/m9.figshare.c.4557599.

7. L188. Provide the time period information of these datasets.

Response:
The time period information has been provided, as follows.
Among them, $Prcp$ and Q are the annual values of ten major river basins in China from 2003 to 2013, including the Hai, Huai, Liao, Northwest, Pearl, Songhua, Southeast, Southwest, Yangtze, and Yellow (Fig. 2), from the National Water Resources Bulletin, which is extensively used in water resources calculation (Miao et al., 2022) and assessment (Yang et al., 2004; Xie et al., 2018).

8. L190. The data source link of the National Water Resources Bulletin.

Response:
The data source link has been provided in the acknowledgements part like other data sources and links, as follows.

Thanks to the Ministry of Water Resources of the People's Republic of China for providing the basin-wide precipitation and runoff data from the National Water Resources Bulletin at http://szy.mwr.gov.cn/gbsj/index.html.

9. Section 3.1. Instead of the detailed description of NSE change between calibration and validation, it may be better to explain the model performance in different PFTs. For example, why PML-V2 doesn't perform well on wetland.

Response:

This part of the manuscript has been revised, as follows:

For daily ET, the declines in NSE values are less than 0.14 in most PFTs except BSV and ENF, whose NSE decreased by 0.36 and 0.33, respectively. As expected, RMSE values all increased to some extent in all PFTs (ranging from 0.002 to 0.305 mm d^{-1}) when compared with those in calibration mode. The Bias values in the cross-validation mode were almost identical to those in the calibration mode for most PFTs except WET and ENF of which the absolute value of Bias increased by 10.59% and 17.42%, respectively (Fig. 4a). From calibration to cross-validation, the degradation of BSV, ENF, and WET is more serious than that for the remaining PFTs, which is mainly caused by the small samples (2, 2, and 3, respectively) for ET estimates.

10. L293. PML-V2 performs well when compared with other mainstream ET or GPP products in China. Please check through the manuscript and make it clear the model only performed better in China.

Response:

We have revised the description of the model's performance, as follows.

In summary, PML-V2(China) performs well when compared with other mainstream ET or GPP products in China.

11. L299-300. Performs better in simulating GPP. Add GPP in this sentence.

Response:

We have revised the description of the model's performance, as follows.

As shown in Fig. 8, PML-V2(China) performs significantly better than other advanced methods in simulating the GPP of CRO, MF, ENF, EBF, SAV, and BSV, producing higher NSE, R, and lower RMSE and Bias.

12. L333. The units of change rates could be yr-2, mm/yr /yr.

Response:

We have revised the units of change rates throughout the manuscript.

13. L338. Use the same decimal digits.

Response:

We have edited the decimal digits to make sure that they are the same.

14. Fig 10. What does the SD represent? SD of spatial data or temporal data within a year.

Response:

The SD represents the standard deviation of the annual simulated values during the study years. We have made the SD description clear and changed the shaded area in each sub-part for representing the 95% confidence interval as follows:

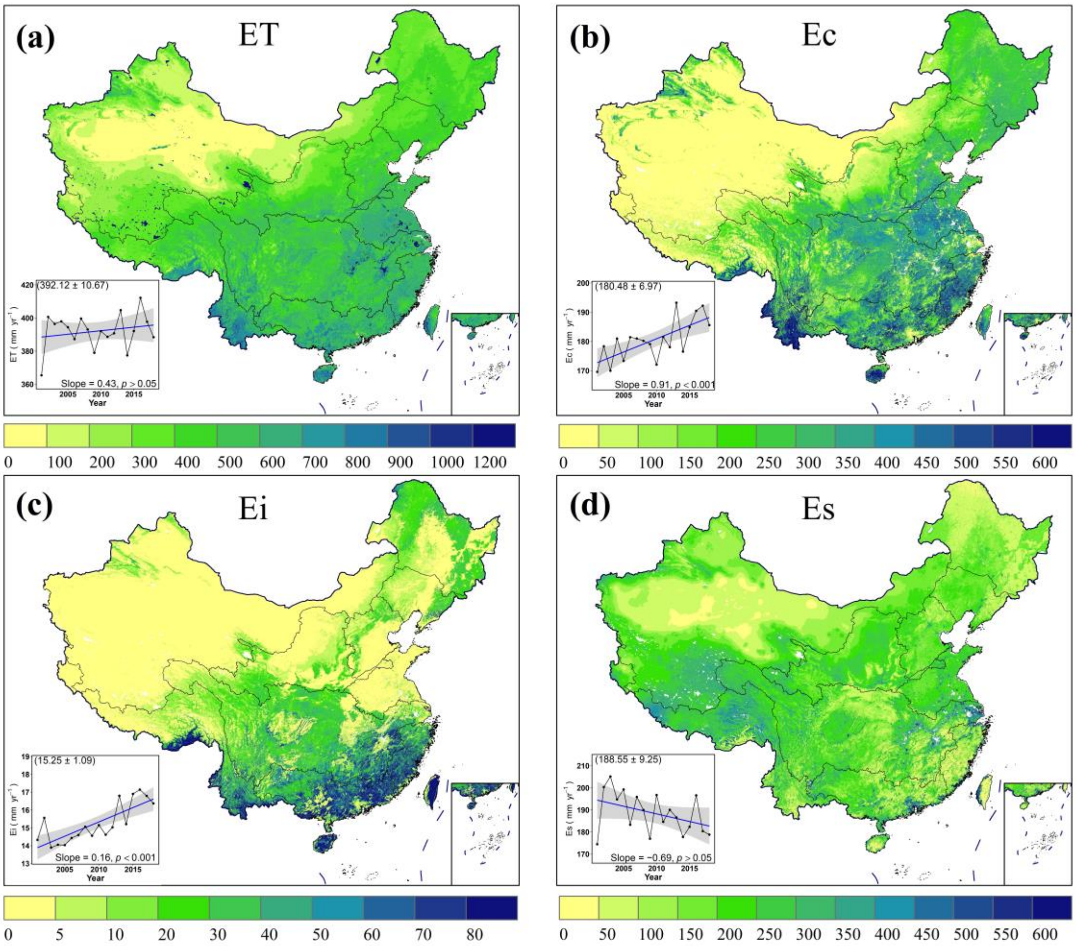

Figure 10: Spatial pattern of mean annual ET, E_c, E_i, E_s and their annual variation during 2001–2018. In all insets, the shaded areas represent the 95% confidence interval based on the linear regression modelling. The number in the parentheses of each inset is mean ± standard deviation of the annual simulated variables during the 18 years.

15. L347. What are the time periods of these previous estimations?

Response:

The sentence has been revised, as follows.

For annual ET over China, the multi-year (2001-2018) mean annual ET from PML-V2(China) is 392.12 ± 10.67 mm yr^{-1} (Fig. 10a). This result is overall consistent with the country-wide averaged annual ET estimated by the machine learning method (Yin et al., 2021: 397.65 mm yr^{-1} for 2000-2018) and land surface models (Ma et al., 2019a: 395.34 mm yr^{-1} for 2001-2012), and slightly higher than MOD16A2 ET about 359.61 ± 59.52 mm yr^{-1} for 2001-2018 (Cheng et al., 2021). But they are all less than the annual ET of about 482.27 ± 192.31 mm yr^{-1} from SEBAL for 2001-2018 (Cheng et al., 2021).

References

Cheng, M., Jiao, X., Li, B., Yu, X., Shao, M., and Jin, X.: Long time series of daily evapotranspiration in China based on the SEBAL model and multisource images and validation, Earth System Science Data, 13, 3995–4017, https://doi.org/10.5194/essd-13-3995-2021, 2021.

Ma, N., Szilagyi, J., Zhang, Y., and Liu, W.: Complementary-Relationship-Based Modeling of Terrestrial Evapotranspiration Across China During 1982–2012: Validations and Spatiotemporal Analyses, Journal of Geophysical Research: Atmospheres, 124, 4326–4351, https://doi.org/10.1029/2018jd029850, 2019a.

Yin, L., Tao, F., Chen, Y., Liu, F., and Hu, J.: Improving terrestrial evapotranspiration estimation across China during 2000–2018 with machine learning methods, Journal of Hydrology, 600, https://doi.org/10.1016/j.jhydrol.2021.126538, 2021.

16. L354. Rephrase this sentence.

Response:

The sentence has been rephrased, as follows.

The annual ET displays a statistically insignificant increasing trend from 2001 to 2018, which is consistent with the calculated ET using the Budyko equation (Feng et al., 2018; Su et al., 2022).

References

Feng, T., Su, T., Ji, F., Zhi, R., and Han, Z.: Temporal Characteristics of Actual Evapotranspiration Over China Under Global Warming, Journal of Geophysical Research: Atmospheres, 123, 5845-5858, https://doi.org/10.1029/2017JD028227, 2018.

Su, T., Feng, T., Huang, B., Han, Z., Qian, Z., Feng, G., Hou, W., and Dong, W.: Long-term mean changes in actual evapotranspiration over China under climate warming and the attribution analysis within the Budyko framework, International Journal of Climatology, 42, 1136–1147, https://doi.org/10.1002/joc.7293, 2022.

17. L369-370. The global data also have many EC observation data, but not in China. These models were calibrated on a global scale, not only in China, and pursued a global optimal solution.

Response:

The sentence has been rephrased, as follows.

This indicates that more local observations will facilitate the improvement of ET and GPP estimates at regional and national scales.

18. L374. More spatial analysis should be conducted to prove the strong ability of the model in simulating ET on the double-cropping system. Show some regional analysis rather than only using EC sites. Why PML-V2 can have this ability?

Response:

We have supplemented the spatial analysis of the simulated ET of PML-V2(China) on the double-cropping system. We extracted the cropland with peaks and identified the dates of peaks appearing within a year at each pixel by a faster peak detection algorithm (Liu et al., 2020). Here, we quantified the cropping intensity (e.g., double-cropping system) in croplands (Fig. S2a1) and identified the dates of the first peak and the second peak appearing in 2015 (Fig. S2a2, a3). To verify the reliability of the results, we mapped the double-cropping cropland areas of winter wheat and summer maize rotations (Fig. S2b1), the heading dates distribution of winter wheat and summer maize (Fig. S2b2, b3) in 2015 based on the crop phenological dataset (Luo et al., 2020). The croplands with a double-cropping show similar spatial patterns, as indicated by Fig. S2a1 and b1. In particular, we also compared the first ET peak date (Fig. S2a2) with the heading date of winter wheat (Fig. S2b2) in 2015. The first ET peak date (i.e., day of the year (DOY)) is mainly between DOY 120 and 150, occurring after the heading date of winter wheat about DOY 100 to 130. Similarly, the second ET peak also occurs slightly later than the heading of summer maize (Fig. S2a3, b3). The ET intensity was the highest of the entire growth period from the heading date to the filling date (He et al., 2022), which means the ET peak appears slightly later than the crop heading. Moreover, we have added this figure in the Supplementary material. Compared to the old version of PML-V2, PML-V2(China) utilized the MODIS leaf area index data after the improved Whittaker filter. The filtered LAI carrying more accurate phenology information as model inputs, is not only the reason why the PML-V2(China) product can reveal the characteristics of the water consumption from the croplands but also the reason why it is well estimated in most plant function types.

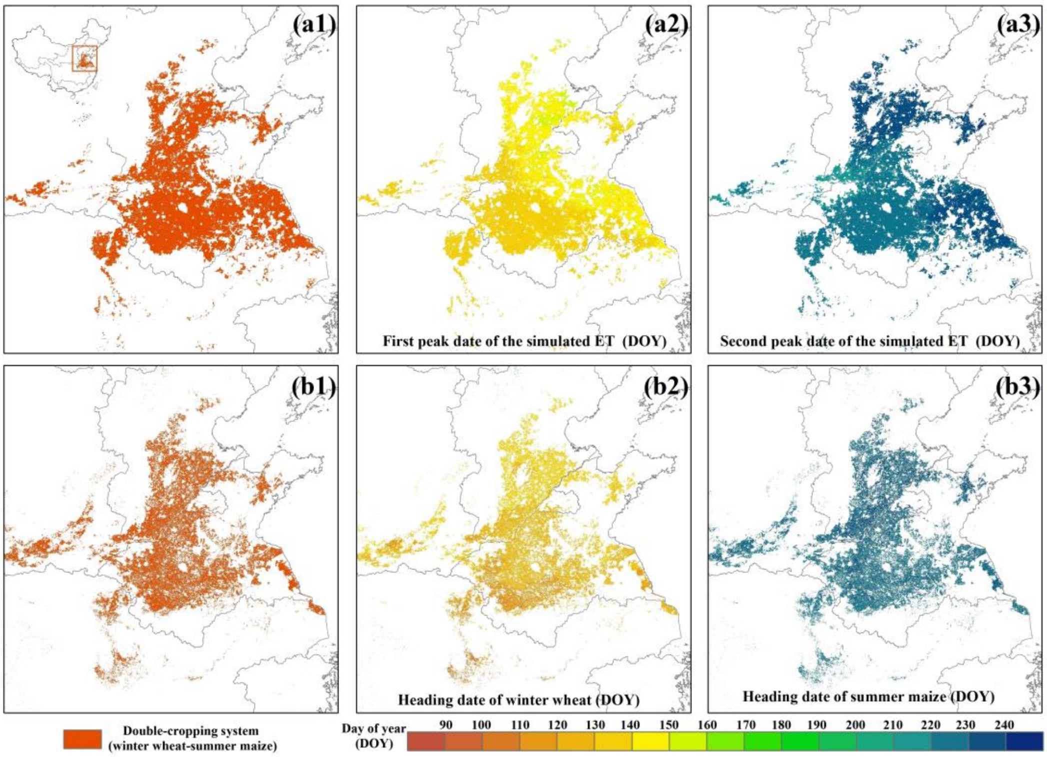

Figure S2: Spatial patterns of the PML-V2(China) ET with double peaks in 2015 (a1) and the double-cropping croplands in 2015 from a crop phenological dataset (ChinaCropPhen1km) (b1); spatial patterns of the first peak dates (a2) and the second peak dates (a3) from the PML-V2(China) ET in 2015; and spatial

patterns of the heading dates of winter wheat (b2) and those of summer maize (b3) from the crop phenological dataset in 2015.

References

He, H., Wu, Z., Li, D., Zhang, T., Pan, F., Yuan, H., Jiang, S., Shi, Z., Yang, S., and Wang, F.: Characteristics of Winter Wheat Evapotranspiration in Eastern China and Comparative Evaluation of Applicability of Different Reference Evapotranspiration Models, J Soil Sci Plant Nutr, 22, 2078–2091, https://doi.org/10.1007/s42729-022-00795-y, 2022.

Liu, L., Xiao, X., Qin, Y., Wang, J., Xu, X., Hu, Y., and Qiao, Z.: Mapping cropping intensity in China using time series Landsat and Sentinel-2 images and Google Earth Engine, Remote Sensing of Environment, 239, 111624, https://doi.org/10.1016/j.rse.2019.111624, 2020.

Luo, Y., Zhang, Z., Chen, Y., Li, Z., and Tao, F.: ChinaCropPhen1km: a high-resolution crop phenological dataset for three staple crops in China during 2000–2015 based on leaf area index (LAI) products, Earth System Science Data, 12, 197–214, https://doi.org/10.5194/essd-12-197-2020, 2020.

19. L395. Which results in this study exhibit a huge potential for carbon sequestration of vegetation in China?

Response:

Gross primary productivity (GPP), the gross uptake of carbon dioxide (CO_2) by plant photosynthesis, is the primary driver of the land carbon sink, which presently removes around one-quarter of the anthropogenic CO_2 emissions each year (Spielmann et al., 2019). GPP is usually recognized as an important positive factor to gain carbon sequestration by terrestrial ecosystems (Liu et al., 2014; Chen et al., 2021). Using the PML-V2(China) dataset, we find that the trend of annual GPP has a significant increase of about 8.99 g C m^{-2} yr^{-2} for 2001-2018, which indicates that vegetation in China exhibit a huge potential for carbon sequestration in China. The paragraph in this manuscript has been revised, as follows:

On the other hand, this dataset has better simulations of carbon consequences and water use efficiency, which is important for carbon-neutron policy (Yang et al., 2022). Specifically, for 2001-2018, the annual GPP and water use efficiency experienced a significant increase (8.99 g C m^{-2} yr^{-2} and 0.02 g C mm^{-1} H_2O yr^{-1}, respectively), but annual ET showed a non-significant increase (0.43 mm yr^{-2}). This indicates that vegetation in China exhibits a huge potential for carbon sequestration with little cost in water resources, which plays an important role in the global carbon cycle.

References

Chen, Y., Feng, X., Tian, H., Wu, X., Gao, Z., Feng, Y., Piao, S., Lv, N., Pan, N., and Fu, B.: Accelerated increase in vegetation carbon sequestration in China after 2010: A turning point resulting from climate and human interaction, Global Change Biology, 27, 5848–5864, https://doi.org/10.1111/gcb.15854, 2021.

Liu, Y., Zhou, Y., Ju, W., Wang, S., Wu, X., He, M., and Zhu, G.: Impacts of droughts on carbon sequestration by China's terrestrial ecosystems from 2000 to 2011, Biogeosciences, 11, 2583–2599, https://doi.org/10.5194/bg-11-2583-2014, 2014.

Spielmann, F. M., Wohlfahrt, G., Hammerle, A., Kitz, F., Migliavacca, M., Alberti, G., Ibrom, A., El-Madany, T. S., Gerdel, K., Moreno, G., Kolle, O., Karl, T., Peressotti, A., and Delle Vedove, G.: Gross Primary Productivity of Four European Ecosystems Constrained by Joint CO2 and COS Flux Measurements, Geophysical Research Letters, 46, 5284–5293, https://doi.org/10.1029/2019GL082006, 2019.

Yang, Y., Shi, Y., Sun, W., Chang, J., Zhu, J., Chen, L., Wang, X., Guo, Y., Zhang, H., Yu, L., Zhao, S., Xu, K., Zhu, J., Shen, H., Wang, Y., Peng, Y., Zhao, X., Wang, X., Hu, H., Chen, S., Huang, M., Wen, X., Wang, S., Zhu, B., Niu, S., Tang, Z., Liu, L., and Fang, J.: Terrestrial carbon sinks in China and around the world and their contribution to carbon neutrality, Sci. China Life Sci., 65, 861–895, https://doi.org/10.1007/s11427-021-2045-5, 2022.

20. L454. How this misclassification issue was dealt with in this study.

Response:

Thank you for your great comments. To extend the estimated ET and GPP from the observed sites to gridded country-wide maps, we used the Annual International Geosphere-Biosphere Programme (IGBP) classification of MCD12Q1.006 dataset as land covers or plant function types. Among the MCD12Q1.006 IGBP land cover types, the Cropland/Natural Vegetation Mosaics (CNVM) type is usually recognized as a part of cropland (Estel et al., 2016; Odongo et al., 2019; Zhang et al., 2019). Hence, we used the same 11 parameters as cropland in CNVM. Besides, we used the published MCD12Q1.006 IGBP land cover types without handling some misclassification coming from spectral confusion.

References

Estel, S., Kuemmerle, T., Levers, C., Baumann, M., and Hostert, P.: Mapping cropland-use intensity across Europe using MODIS NDVI time series, Environ. Res. Lett., 11, 024015, https://doi.org/10.1088/1748-9326/11/2/024015, 2016.

Odongo, V. O., van Oel, P. R., van der Tol, C., and Su, Z.: Impact of land use and land cover transitions and climate on evapotranspiration in the Lake Naivasha Basin, Kenya, Science of The Total Environment, 682, 19–30, https://doi.org/10.1016/j.scitotenv.2019.04.062, 2019.

Zhang, Y., Kong, D., Gan, R., Chiew, F. H. S., McVicar, T. R., Zhang, Q., and Yang, Y.: Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017, Remote Sensing of Environment, 222, 165–182, https://doi.org/10.1016/j.rse.2018.12.031, 2019.