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

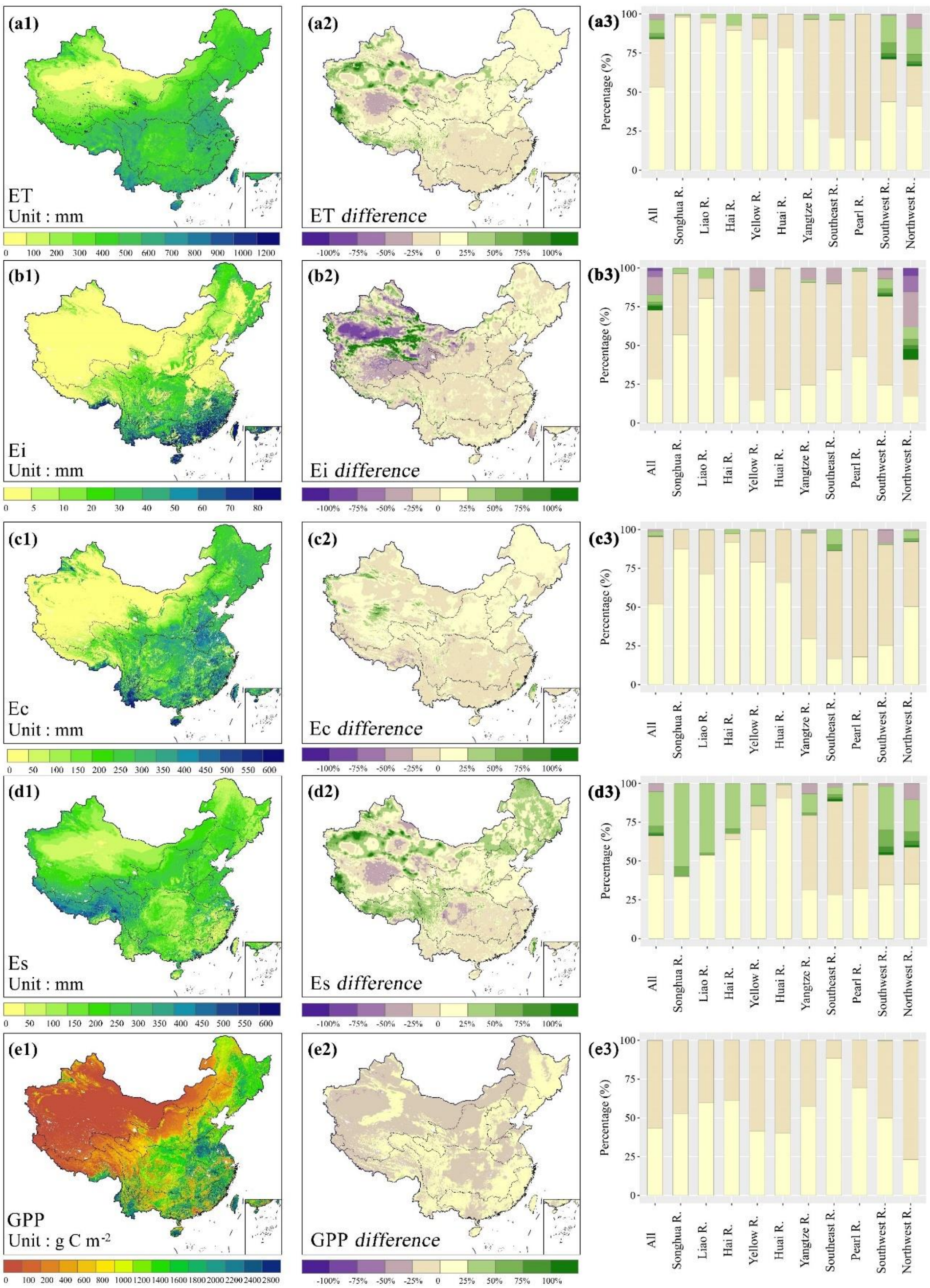

Figure S1: The modelling results using GLDAS-2.1 meteorological forcing data during 2001-2018 and comparison with the PML-V2(China) product using CMFD: (a1-e1) Spatial distribution of the 18-year mean of five variables; (a2-e2) Spatial distribution of the difference using two forcing datasets, calculated 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

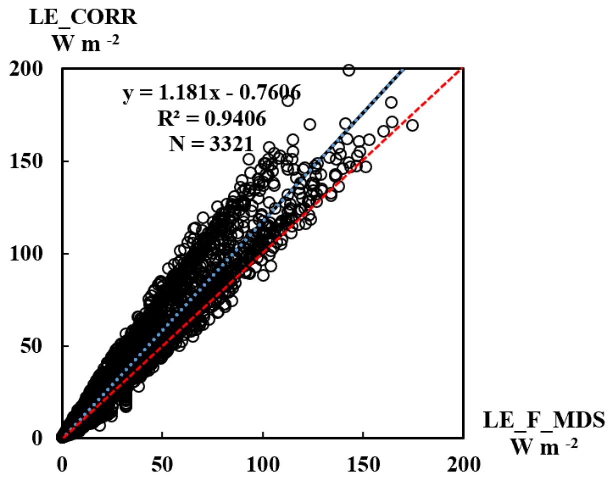

Figure R1: Scatterplot between the latent heat flux item (LE_F_MDS) and the latent heat flux corrected by energy balance closure correction factor (LE_CORR) at the daily scale.

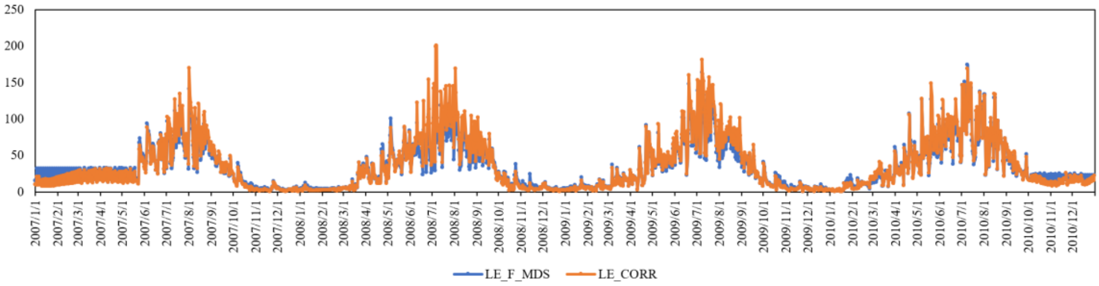

Figure R2: Comparison of LE_F_MDS and LE_CORR from January 2007 to December 2010 at the CN-Cng site on a daily scale.

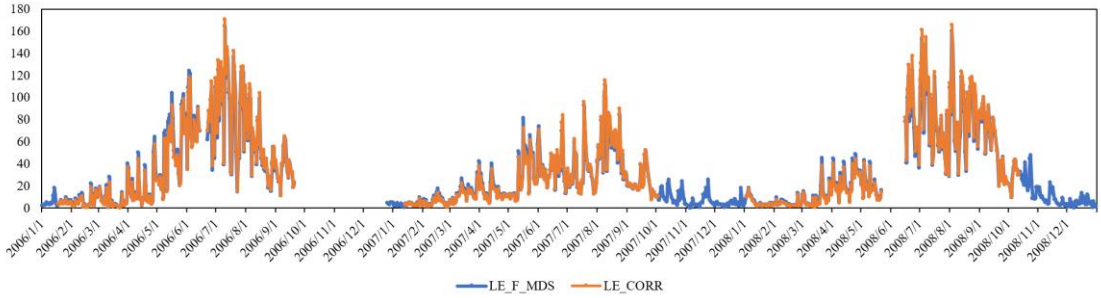

Figure R3: Comparison of LE_F_MDS and LE_CORR from January 2006 to December 2008 at the CN-Du2 site on a daily scale.

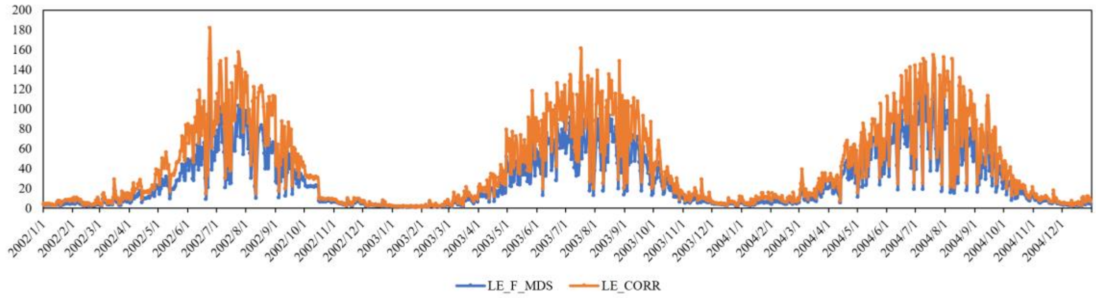

Figure R4: Comparison of LE_F_MDS and LE_CORR from January 2002 to December 2004 at the CN-HaM site on a daily scale.

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 population size 1000 and number of generations 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:

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

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

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

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

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.

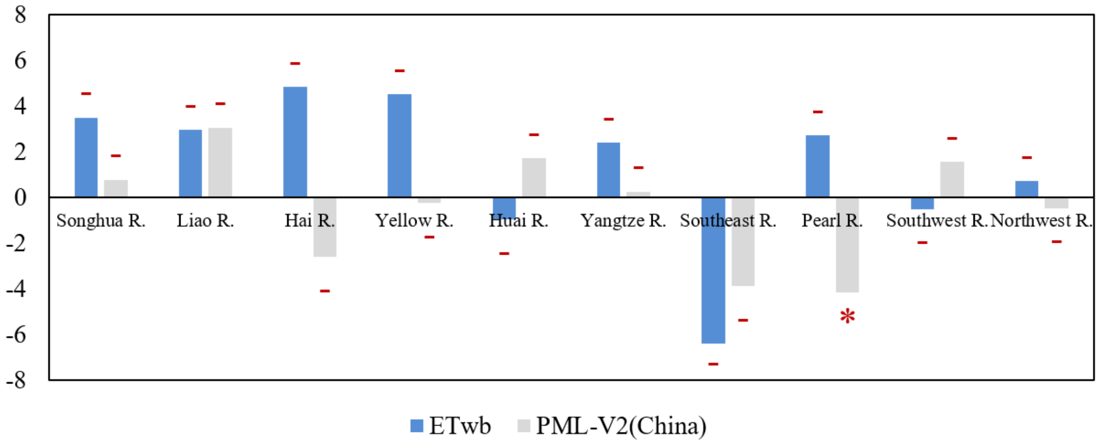

Figure R5: The linear trend bars between with PML-V2(China) ET and the water balance-based ET on the ten river basins. Note that "-" indicates that the *p*-value of the t-test for trend analysis is not less than 0.05. Similarly, "*" means 0.001≤*p*<0.05, and "" means *p*<0.001.**