Supplement of Earth Syst. Sci. Data, 18, 4097–4112, 2026
https://doi.org/10.5194/essd-18-4097-2026-supplement

Open Access Earth System Science Data

Supplement of

A high-resolution gridded dataset of water footprints for China's major food crops from 2001 to 2020

En Hua et al.

Correspondence to: Xuhui Wang (xuhui.wang@pku.edu.cn)

S1 Methods

S1.1 Irrigation water use efficiency

For irrigation water use (WUI), the data primarily comes from the dataset by Zhou Feng et al. (2020), which covers irrigation water use at the municipal level from 2001 to 2013.

Irrigation WUI is modeled as the ratio of $PIRR$ to irrigation efficiency (IE), limited by freshwater availability allocated to irrigation sector ($AIRR$) as well:

$$IRR = Area \times WUI, \quad \text{(S1)}$$

$$WUI = Min\{PIRR/IE\,, AIRR\}, \quad \text{(S2)}$$

$$AIRR = RFR(1-\eta) - (u_2 + u_3), \quad \text{(S3)}$$

$$1/IE = a \times WCI + b, \quad \text{(S4)}$$

where $PIRR$ is derived from an ensemble of six GHMs (i.e., DBH, H08, LPJmL, PCR-GLOBWB, WaterGAP2, and VIC) of ISIMIP. $AIRR$ is defined as renewable freshwater resources (local, upstream import, and transboundary transfer; RFR) minus both of flood water ($\eta \times RFR$) and industrial and domestic water uses ($(u_2 + u_3)$), because flood water is unavailable for human use and $(u_2 + u_3)$ have a high supply reliability (e.g., >95%). Data of RFR, η, u_2, and u_3 are obtained from the first and second National Water Resources Assessment Programs. IE is more difficult to quantify over time and space. WCI, defined as the ratio between the area equipped for WCI and total irrigated area, was taken as a proxy for the variation of IE, as IE is correlated with WCI at the national level. Even if the relationship is only a correlation, and not a causality, we can still use the relationship to determine the irrigation efficiency corresponding to a given level of WCI. Data for WCI were obtained from the China Water Conservancy Yearbook and the China Statistics Yearbook.

Additionally, a multi-source data fusion approach was employed to supplement the municipal irrigation water use database for 2014 to 2020. In the machine learning modeling stage, based on quadratic polynomial feature engineering (degree=2) and Z-score standardized preprocessing, the performance of linear regression, ridge regression, and lasso regression models was evaluated by a five-fold cross-validation system, and the preferred model was selected based on the Mean Squared Error for irrigation water use prediction, and non-negative constraints were implemented for optimization of the predicted values. For spatial downscaling, the provincial-level irrigated area of cropland served as the benchmark. The municipal irrigated area share in 2013 was combined with the average irrigated area share of crops (maize, rice, and wheat) from 2001 to 2013 to construct a spatiotemporal weighting matrix for municipal-scale decomposition. The soybean and potato irrigated area was estimated using a proportionality coefficient method, with parameters derived from the average provincial soybean sown area share for 2014 to 2020, based on statistical yearbook data.

The evaluation of crop irrigation water use efficiency can reveal issues such as water resource waste, outdated irrigation technologies, and uneven water distribution, providing a basis for optimizing irrigation management and improving water use efficiency. To this end, the study introduced the irrigation water use efficiency indicator, which was solved through the crop blue water footprint and irrigation water use.

S1.2 LMDI driver analysis

The data used in this study include blue water footprint (BWF), green water footprint (GWF), and their associated driving factors, including evaporation, transpiration, planting area, and phenology.

This study employed a logarithmic mean weighted decomposition approach inspired by the Logarithmic Mean Divisia Index (LMDI) framework to quantify the relative contributions of multiple driving factors to variations in water footprints. Unlike conventional share-based LMDI formulations, the present study adopts a non-normalized raw-variable framework, in which the water footprint response is evaluated directly using the original values of driving factors without introducing structural share transformations.

The temporal change in water footprint between two consecutive periods is expressed as:

$$\Delta WF = WF_t - WF_{t-1}, \tag{S5}$$

The contribution of each driving factor is estimated using the logarithmic mean weighted formulation:

$$C_i = L(WF_t, WF_{t-1}) \cdot \ln\left(\frac{X_{i,t}}{X_{i,t-1}}\right), \tag{S6}$$

where $X_{i,t}$represents the original (non-normalized) value of the i-th driving factor at time t.

The logarithmic mean function is defined as:

$$L(WF_t, WF_{t-1}) = \frac{WF_t - WF_{t-1}}{\ln(WF_t) - \ln(WF_{t-1})}, \tag{S7}$$

with the special case:

$$L(WF_t, WF_{t-1}) = WF_t \text{if} WF_t = WF_{t-1}, \quad \text{(S8)}$$

A positive value of C_i indicates that the corresponding driving factor promotes an increase in the water footprint, whereas a negative value indicates a mitigating effect. The absolute magnitude of C_i reflects the relative strength of the factor's influence on temporal water footprint variations.

This framework emphasizes the response of water footprint changes to variations in physical driving variables and avoids normalization procedures that may obscure direct process-based interpretations. Therefore, it is particularly suitable for analyzing the dynamic effects of climatic, physiological, and agricultural management factors on crop water use.

S1.3 Water footprint of seasonal water shortages

Based on the traditional water footprint accounting method, this study supplements the water footprint to account for seasonal water shortages, specifically defining the blue water supplemented through irrigation (ΔS_{blue}) and the green water supplemented through precipitation (ΔS_{green}). The specific accounting steps are as follows:

First, the soil saturated water content was estimated. Based on the soil porosity data for China provided by Shi et al. (2025), and considering the spatial heterogeneity of soil texture and structure across different regions, the saturated water content for each layer was calculated using porosity data at standard depth layers with a spatial resolution of 1 km. The soil profile data used by Shi et al. (2025) were primarily derived from the Second National Soil Survey conducted during the 1970s and 1980s. This dataset not only offers new insights into the spatiotemporal variations in soil properties but is also consistent with existing maps developed from soil profiles compiled since the 2010s.

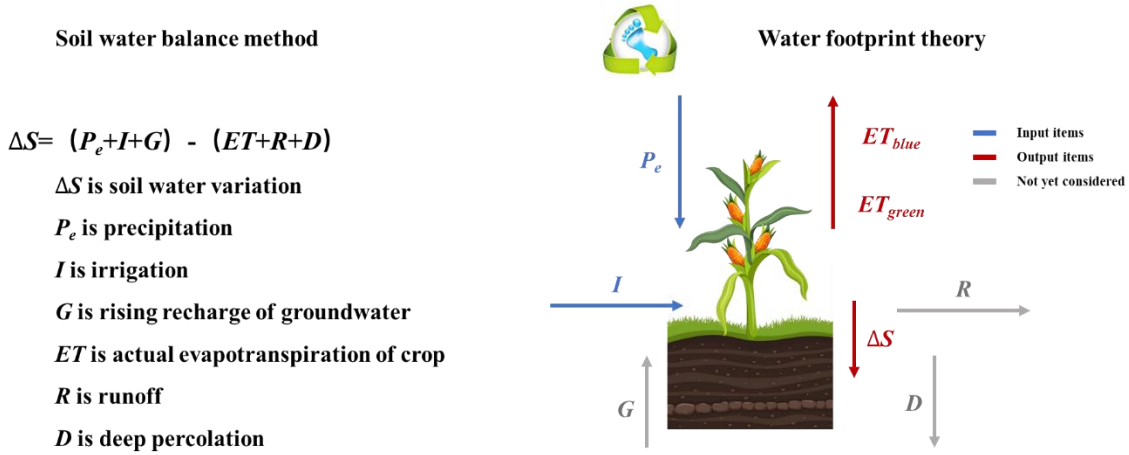

Figure S1: Schematic diagram of the soil water balance method.

Second, the initial soil water content was calculated (Fig. S1). Based on the soil water balance method, daily soil water content dynamics were simulated using daily precipitation and evapotranspiration data from 2001 to 2002. It was assumed that the initial soil water content was at the upper limit of water holding capacity for that period. The multi-day average value after model spin-up was taken as the initial soil water content on January 1, 2001. Due to data limitations, this balance

calculation does not account for deep percolation, capillary rise, or surface runoff, but its impact on the calculation of the crop water footprint is relatively small.

Finally, ΔS was determined based on the soil water balance method. The upper limit of suitable soil moisture during the crop growing period was set as the threshold—based on saturated water content: 100% for rice and 60% for other crops. Based on this, the ΔS_{blue} and ΔS_{green} required at the beginning of the growing period to supplement the soil water to this threshold were calculated.

$$\Delta S_{green} = \begin{cases} \max(P_e - ET, 0), & S_t - S_{t-1} > P_e - ET \\ S_t - S_{t-1}, & S_t - S_{t-1} \le P_e - ET \end{cases} \quad \text{(S9)}$$

$$\Delta S_{blue} = \begin{cases} S_t - S_{t-1} - \Delta S_{green}, & S_t - S_{t-1} > \Delta S_{green} \\ 0, & S_t - S_{t-1} \le \Delta S_{green} \end{cases} \quad \text{(S10)}$$

where S_t is the soil moisture content on day *t*; S_{t-1} is the soil moisture content on day *t-1*.

To test the robustness of the soil moisture threshold settings, we conducted a sensitivity analysis by setting two additional scenarios with thresholds of 45% and 75% based on the baseline value of 60% for non-rice crops, while the threshold for rice remained unchanged at 100% of saturated water content. The formula for sensitivity analysis is:

$$S = [(WF - WF_{base})/WF_{base}]/[(X - X_{base})/X_{base}] \quad \text{(S11)}$$

where X represents soil moisture content under different scenarios, and X_{base} represents soil moisture content under the baseline scenario; WF represents the water footprint under different scenarios, and WF_{base} represents the water footprint under the baseline scenario. S is the sensitivity coefficient.

Due to crop rotation, root depth and water extraction zones differ markedly between crops grown in different years, rendering residual deep soil moisture from previous years largely inaccessible to subsequent crops. Furthermore, unproductive evaporation during fallow periods substantially depletes any remaining soil moisture, while deep percolation beyond the root zone represents an irreversible loss from the field-scale water balance. In addition, actual irrigation management adheres to the principle of in-season replenishment, whereby water demand is satisfied within the same growing season rather than relying on carryover from prior years. Therefore, water replenished in each year to address seasonal water shortages serves only that year's crop production and cannot be reused across years.

S2 Results

S2.1 Spatial distribution pattern of crop water footprint

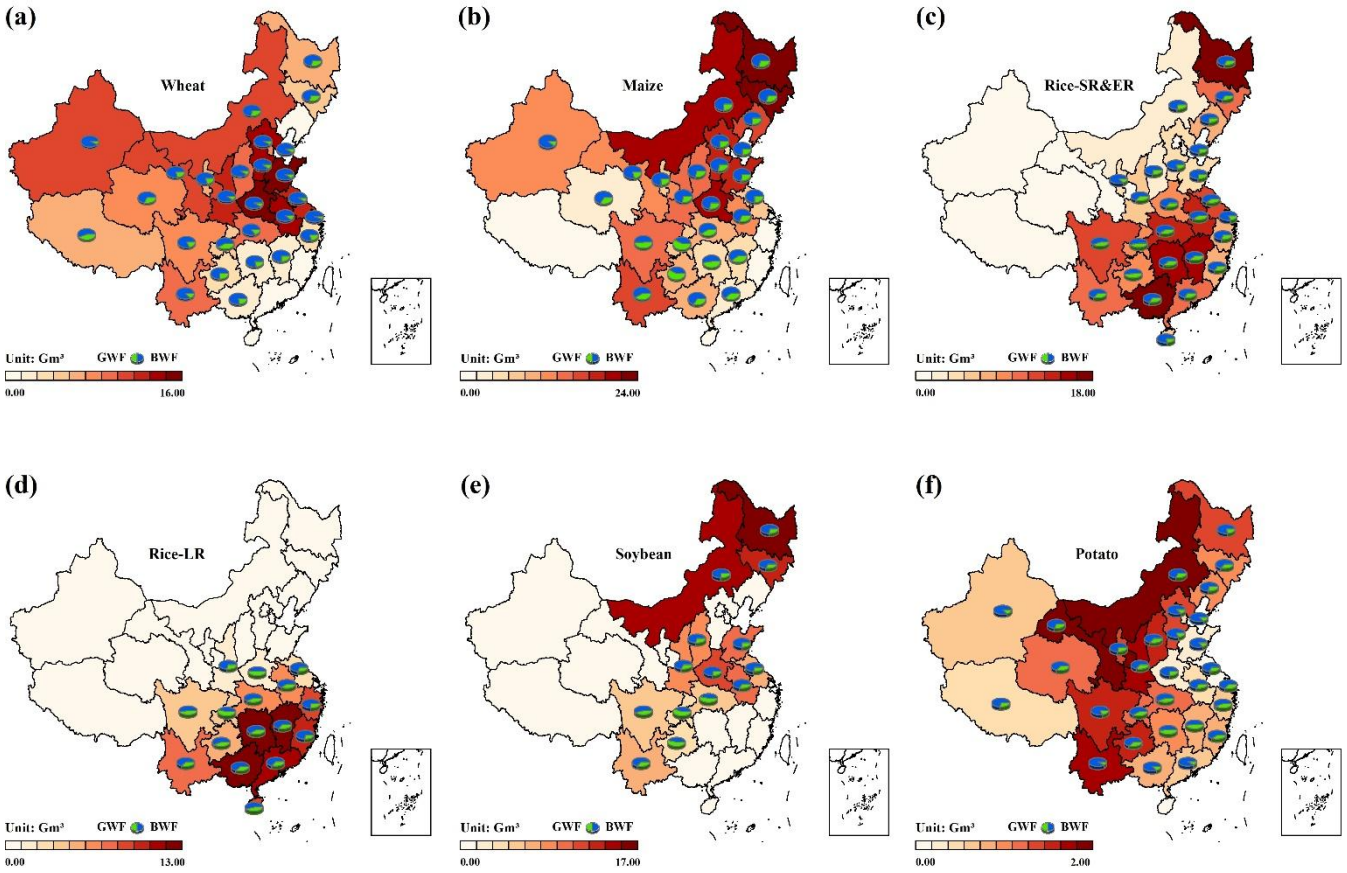

Figure S2: Spatial distribution pattern of crop water footprint in 2020. Pie charts showing the proportions of blue water footprint and green water footprint.

S2.2 Development trend of irrigation water use efficiency

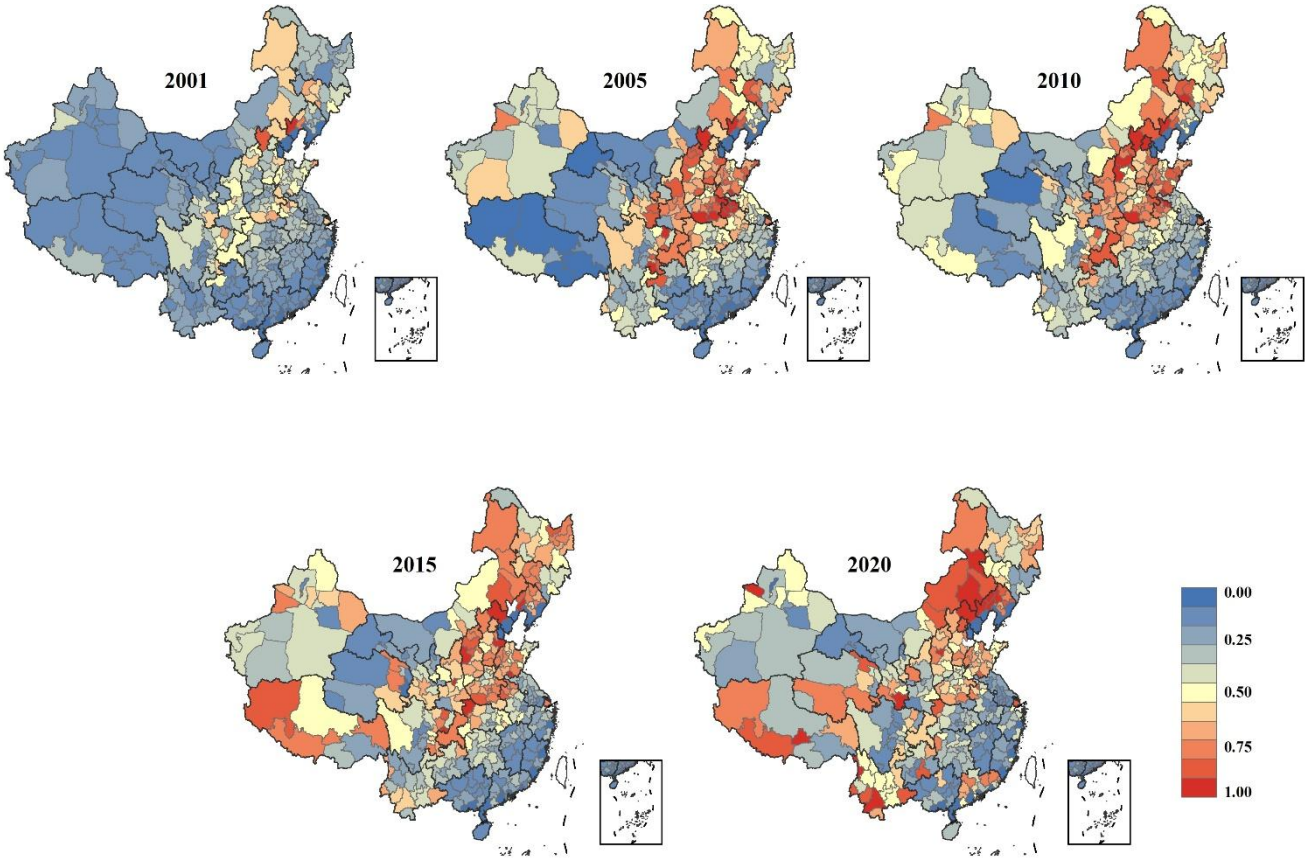

Figure S3: The spatial distribution pattern of irrigation water use efficiency.

S2.3 Data validation of ChinaCropWF

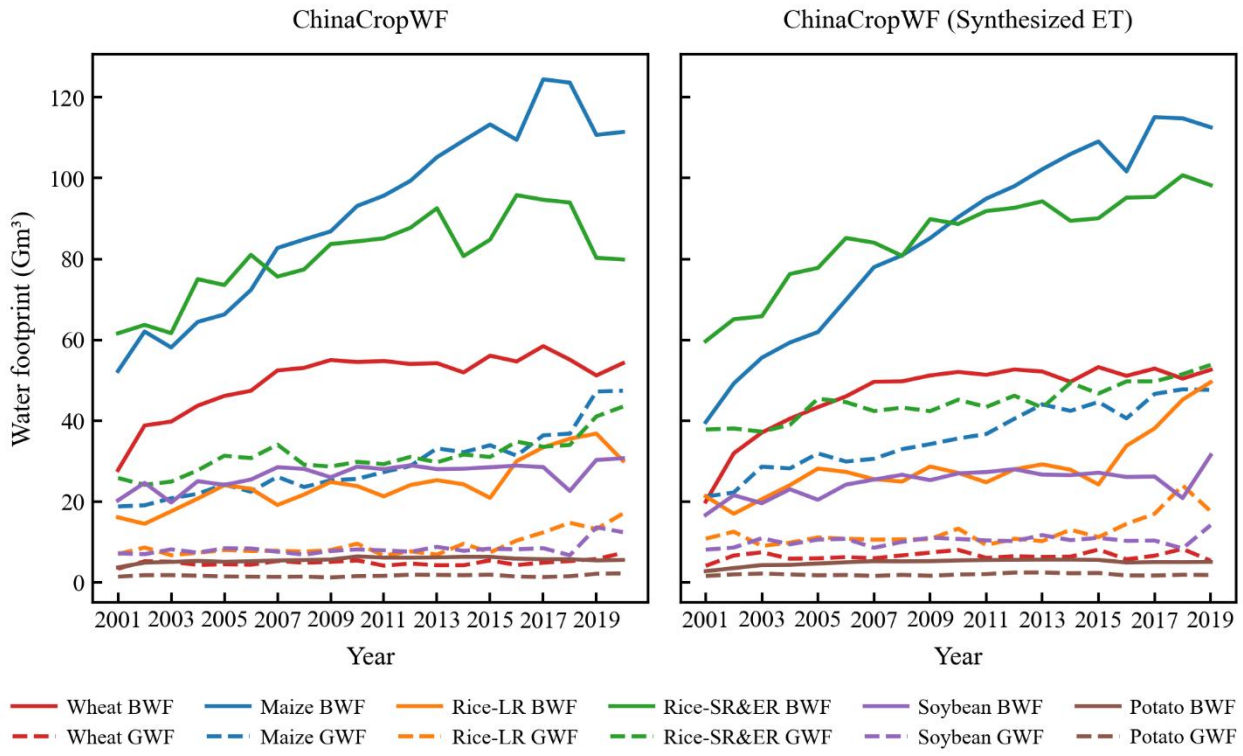

Figure S4: Comparison of crops water footprint based on two evapotranspiration products.

S2.4 Comparison of different water footprint datasets

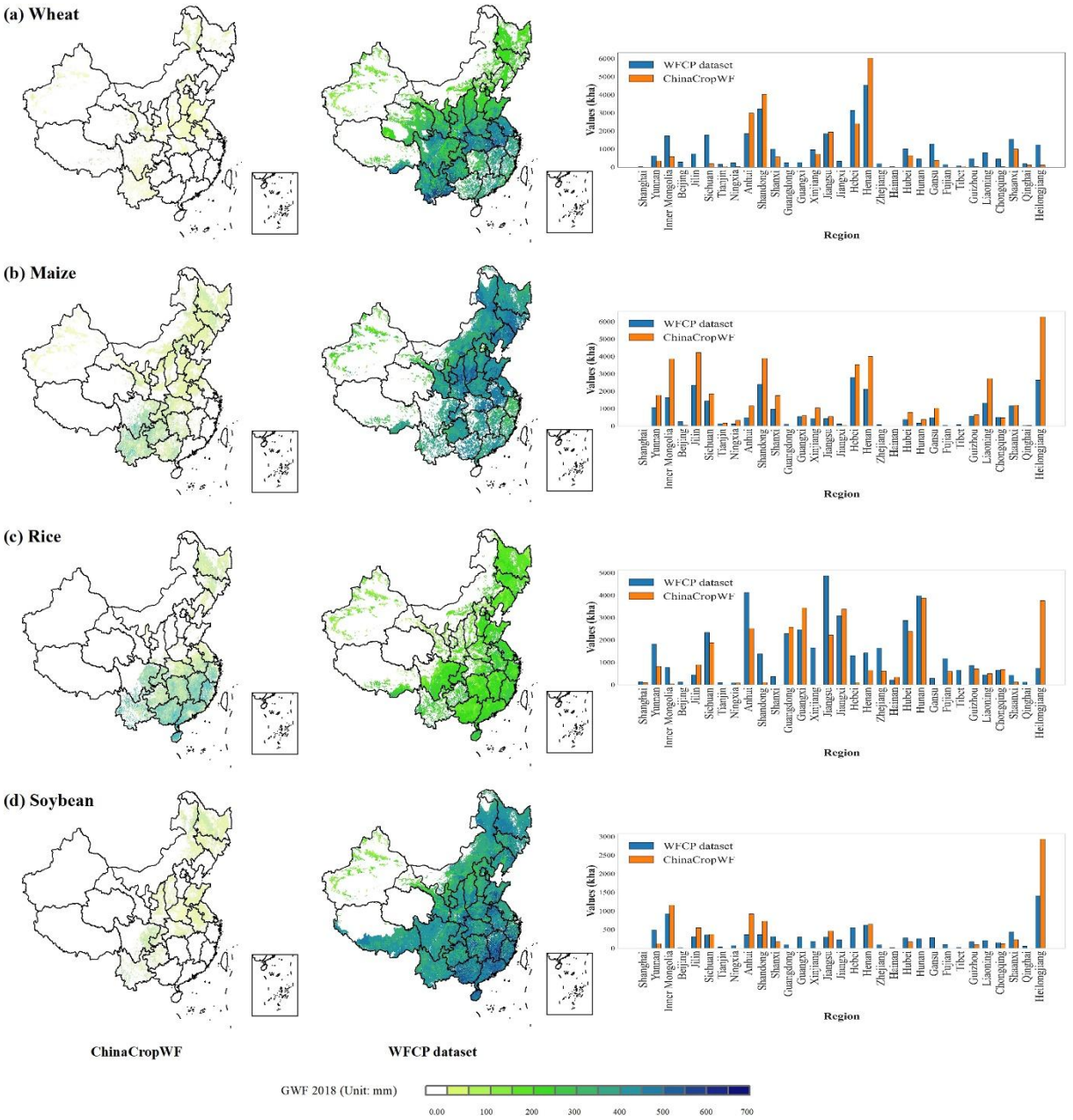

Figure S5: Comparison of planting area of different water footprint datasets.

S2.5 Sensitivity analysis of soil moisture thresholds.

Table S1 Sensitivity coefficient in crop water footprint under the 45% threshold scenario relative to the baseline scenario.

Year	Wheat-BWF	Wheat-GWF	Maize-BWF	Maize-GWF	Soybean-BWF	Soybean-GWF	Potato-BWF	Potato-GWF
2001	51.8%	1.6%	31.3%	0.2%	42.8%	0.2%	17.4%	0.2%
2002	61.5%	0.1%	40.5%	0.1%	52.2%	0.1%	23.5%	3.8%
2003	61.7%	0.0%	45.9%	0.1%	63.4%	2.5%	32.7%	0.1%
2004	58.3%	0.0%	44.0%	0.0%	52.6%	0.6%	32.2%	1.8%
2005	59.7%	0.5%	45.9%	0.1%	56.1%	0.6%	36.7%	0.2%
2006	59.0%	0.1%	44.6%	0.2%	53.9%	3.1%	37.3%	0.1%
2007	56.0%	0.0%	44.5%	0.2%	48.5%	2.9%	37.5%	0.1%
2008	57.2%	0.2%	44.0%	0.1%	51.3%	0.3%	36.7%	0.2%
2009	56.1%	0.3%	45.5%	0.2%	56.8%	0.4%	36.1%	0.1%
2010	57.3%	0.1%	44.8%	0.1%	51.7%	0.4%	38.4%	0.5%
2011	54.4%	0.2%	45.4%	0.1%	52.6%	0.4%	40.5%	0.9%
2012	56.8%	0.1%	45.9%	0.1%	51.8%	0.3%	41.8%	0.1%
2013	57.8%	0.2%	45.5%	0.1%	53.4%	0.2%	40.5%	2.3%
2014	56.5%	0.0%	44.7%	0.2%	53.7%	0.5%	40.3%	2.5%
2015	56.7%	0.0%	44.8%	0.1%	53.0%	0.5%	42.1%	1.0%
2016	56.7%	0.0%	44.4%	0.3%	51.8%	2.0%	41.4%	0.1%
2017	54.7%	0.0%	45.0%	0.0%	53.1%	0.5%	43.6%	0.0%
2018	56.3%	0.0%	46.2%	0.1%	53.5%	0.4%	44.8%	0.2%
2019	59.7%	0.0%	50.4%	0.0%	62.2%	0.2%	46.4%	0.4%
2020	56.4%	0.0%	50.3%	0.1%	60.7%	0.2%	45.4%	1.3%

Table S2 Sensitivity coefficient in crop water footprint under the 75% threshold scenario relative to the baseline scenario.

Year	Wheat-BWF	Wheat-GWF	Maize-BWF	Maize-GWF	Soybean-BWF	Soybean-GWF	Potato-BWF	Potato-GWF
2001	85.9%	0.3%	43.7%	0.4%	60.6%	0.3%	29.4%	1.5%
2002	68.7%	0.1%	44.3%	0.1%	57.5%	0.1%	30.8%	4.7%
2003	65.9%	0.1%	49.4%	0.1%	70.9%	2.1%	36.9%	0.1%
2004	61.6%	0.1%	47.7%	0.1%	57.3%	0.4%	36.5%	2.2%
2005	61.9%	0.3%	49.0%	0.1%	60.3%	0.6%	39.2%	0.1%
2006	61.0%	0.1%	47.5%	0.1%	57.7%	2.7%	39.0%	0.0%
2007	57.7%	0.0%	46.3%	0.2%	51.7%	2.2%	38.0%	0.1%
2008	58.7%	0.2%	45.6%	0.1%	53.7%	0.3%	37.3%	0.2%
2009	57.5%	0.3%	46.9%	0.2%	59.0%	0.4%	36.6%	0.1%
2010	58.6%	0.1%	46.4%	0.1%	53.6%	0.5%	42.1%	0.9%
2011	55.5%	0.2%	47.2%	0.2%	54.5%	0.5%	44.2%	1.5%
2012	57.8%	0.1%	47.3%	0.1%	53.5%	0.4%	45.9%	0.2%
2013	58.8%	0.1%	46.7%	0.1%	55.2%	0.2%	43.7%	4.1%
2014	57.1%	0.0%	46.0%	0.2%	55.3%	0.6%	43.7%	3.8%
2015	57.4%	0.0%	46.0%	0.2%	54.5%	0.7%	46.1%	0.7%
2016	57.0%	0.0%	45.6%	0.4%	53.4%	1.8%	45.9%	0.1%
2017	54.9%	0.0%	46.3%	0.0%	54.7%	0.5%	48.2%	0.0%
2018	56.5%	0.0%	47.3%	0.1%	55.1%	0.3%	49.0%	0.3%
2019	59.9%	0.0%	51.8%	0.0%	64.0%	0.3%	51.1%	0.2%
2020	56.6%	0.0%	51.7%	0.1%	62.1%	0.3%	49.6%	1.5%

References

Wei, S., Dai, Y., Liu, B., Zhu, A., Duan, Q., Wu, L., Ji, D., Ye, A., Yuan, H., Zhang, Q., Chen, D., Chen, M., Chu, J., Dou, Y., Guo, J., Li, H., Li, J., Liang, L., Liang, X., Liu, H., Liu, S., Miao, C., and Zhang, Y.: A China data set of soil properties for land surface modeling, JAMES, 5, 212–224, https://doi.org/10.1002/jame.20026, 2013.

Zhou, F., Bo, Y., Ciais, P., Dumas, P., Tang, Q., Wang, X., Liu, J., Zheng, C., Polcher, J., Yin, Z., Guimberteau, M., Peng, S., Ottle, C., Zhao, X., Zhao, J., Tan, Q., Chen, L., Shen, H., Yang, H., Piao, S., Wang, H., and Wada, Y.: Deceleration of China's human water use and its key drivers, Proc. Natl. Acad. Sci., 117, 7702–7711, https://doi.org/10.1073/pnas.1909902117, 2020.