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

En Hua[1, 4], Ling Huang[1], Liangliang Zhang[1], Quanbo Zhao[1], Yingxuan Wang[1], Bowei Wu[1], Sien Li[2], Yaokui Cui[3], Yubao Wang[4], Xuhui Wang[1]

[1]College of Urban and Environmental Sciences, Peking University, Beijing, 100871, China
[2]College of Water Resources and Civil Engineering, China Agricultural University, Beijing, 100083, China
[3]School of Earth and Space Sciences, Peking University, Beijing, 100871, China
[4]College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi, 712100, China

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

S1 Methods

S1.1 Irrigation water use

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

$$IRR = Area \times WUI, \quad (1)$$

$$WUI = Min\{PIRR/IE\,, AIRR\}, \quad (2)$$

$$AIRR = RFR(1-\eta) - (u_2 + u_3), \quad (3)$$

$$1/IE = a \times WCI + b, \quad (4)$$

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 is 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 is

evaluated by a five-fold cross-validation system, and the preferred model is selected based on the Mean Squared Error for irrigation water use prediction, and non-negative constraints are implemented for optimization of the predicted values. For spatial downscaling, the provincial-level irrigated area of cropland serves as the benchmark. The municipal irrigated area share in 2013 is 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 irrigated area is estimated using a proportionality coefficient method, with parameters derive from the average provincial soybean sown area share for 2014 to 2020, based on statistical yearbook data.

S1.2 Irrigation water use efficiency

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 introduces the irrigation water use efficiency indicator, which is solved through the crop blue water footprint (ET_{blue}) and irrigation water use (WUI).

$$IWUE = ET_{blue}/WUI, \quad (5)$$

where $IWUE$ is the municipal irrigation water use efficiency, dimensionless.

S1.3 LMDI driver analysis

The data used in this study include blue water footprint (BWF), green water footprint (GWF), and their driving factors (evaporation, transpiration, planting area, and phenology). All data were normalized using Min-Max scaling to eliminate the influence of different units. The normalization formula is as follows:

$$x_{i,norm} = \left(x_i - \min\,(x_i)\right)/\left(max(x_i) - \min\,(x_i)\right), \quad (6)$$

where x_i represents the original data, and $x_{i,norm}$ represents the normalized data, which ranges from 0 to 1.

This study employs the Logarithmic Mean Divisia Index (LMDI) decomposition method to quantify the contributions of various driving factors to changes in the water footprint. The LMDI method has advantages such as complete decomposition, no residuals, and additivity, making it suitable for multi-factor driver analysis. Specifically, the change in the water footprint at time t, denoted as ΔWF, can be decomposed as follows:

$$\Delta WF = WF_t - WF_{t-1} = \sum_{i=1}^{n} C_i, \quad (7)$$

where WF_t represents the water footprint (BWF or GWF) in year t, and C_i denotes the contribution of the i-th driving factor. The contribution C_i of each driving factor is calculated using the logarithmic mean weighting function (L function).

$$C_i = L(WF_t, WF_{t-1}) \times \ln\left(x_{i,t}/x_{i,t-1}\right), \quad (8)$$

Here, $L(WF_t, WF_{t-1})$ is the logarithmic mean weighting function, and its calculation formula is as follows:

$$L(WF_t, WF_{t-1}) = \begin{cases} (WF_t - WF_{t-1})/(lnWF_t - lnWF_{t-1}), if WF_t \neq WF_{t-1} \\ WF_t, if WF_t = WF_{t-1} \end{cases}, \quad (9)$$

where $x_{i,t}$ represents the normalized value of the i-th driving factor in year t. A positive contribution ($C_i > 0$) indicates that the factor causes an increase in the water footprint, while a negative contribution ($C_i < 0$) indicates that the factor causes a decrease. The absolute value of the contribution reflects the magnitude of the factor's influence—the larger the value, the more significant the impact.

S2 Results

S2.1 Spatial distribution pattern of crop water footprint

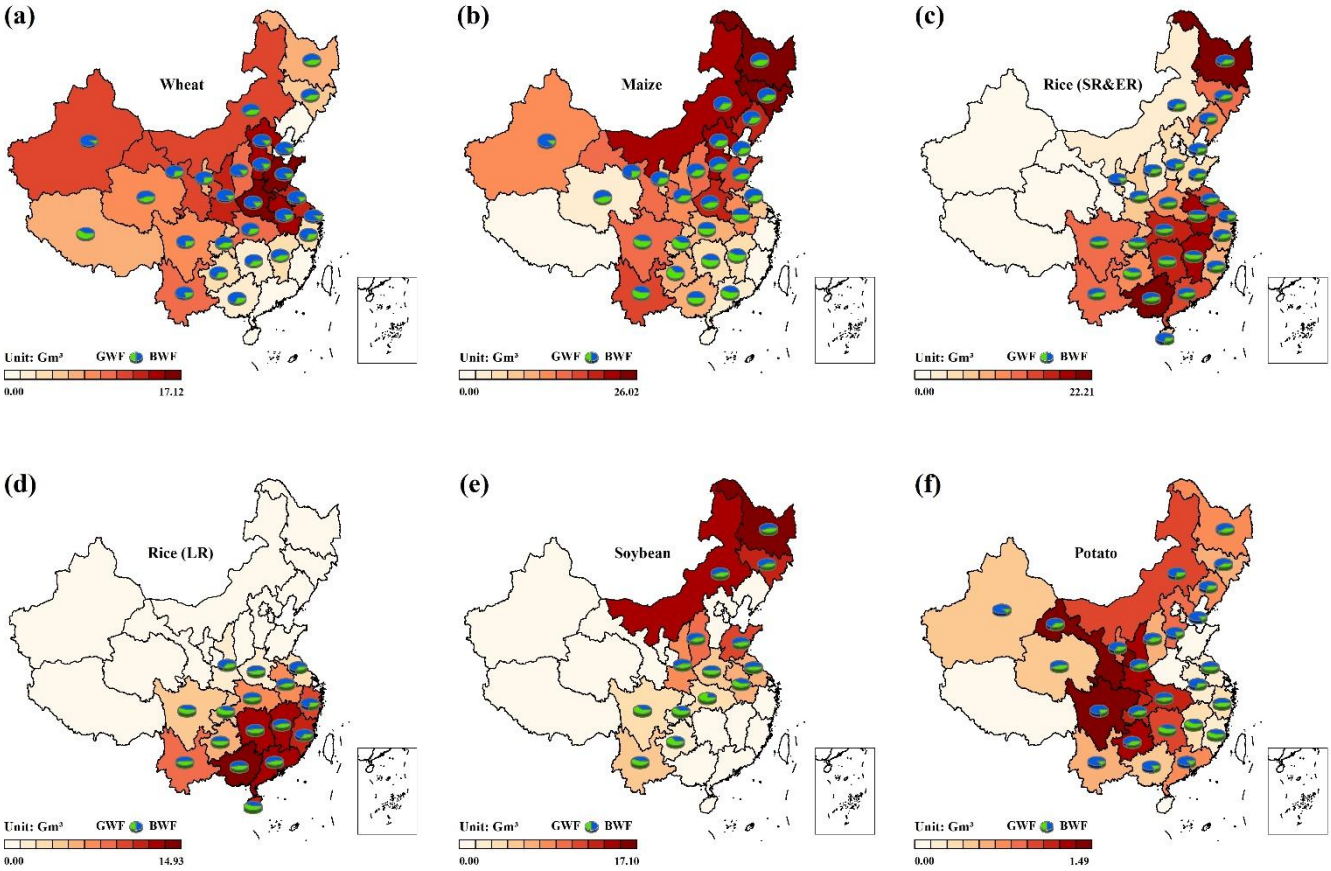

Figure S1: Spatial distribution pattern of crop water footprint in 2020 (pie charts showing the proportions of blue water footprint and green water footprint).

S2.2 Development trend of irrigation water use efficiency

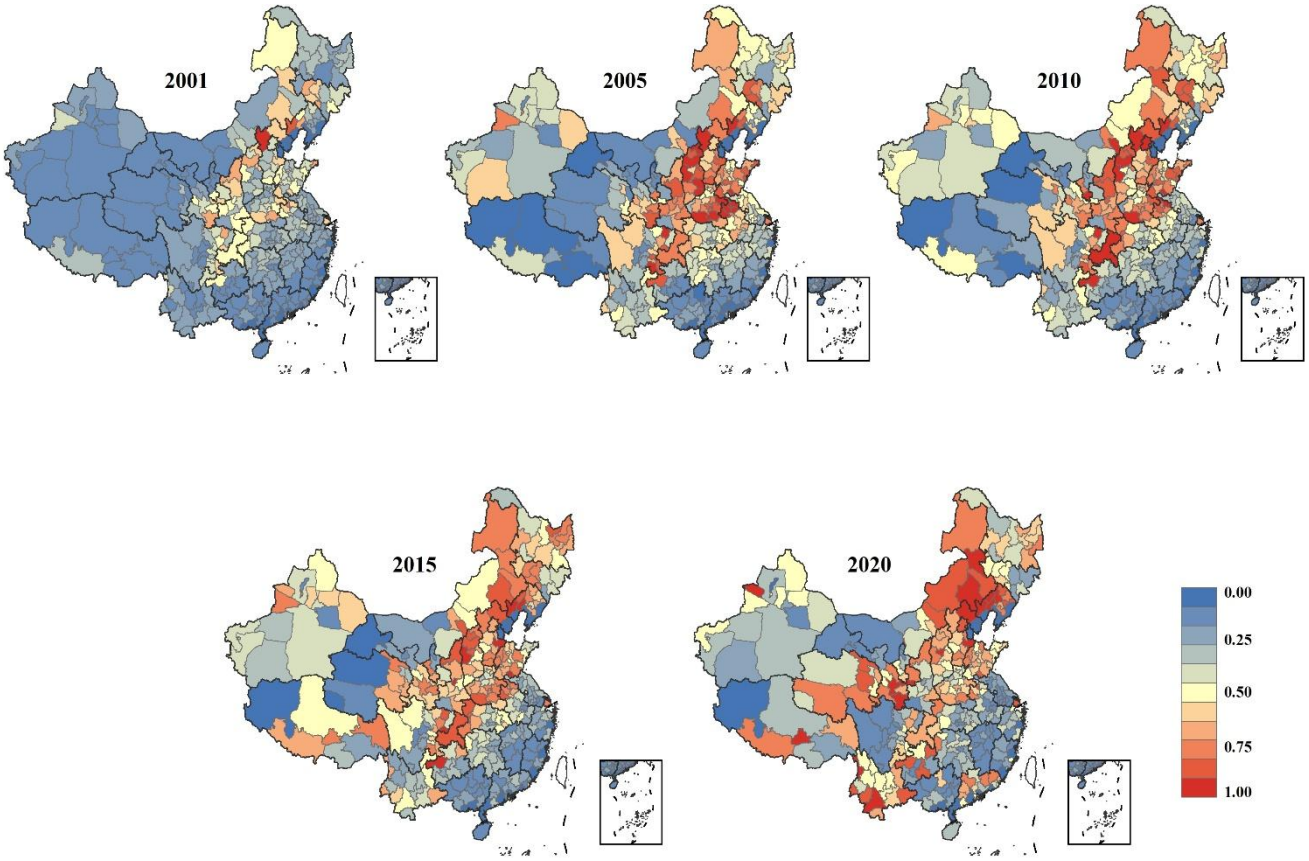

Figure S2: The spatial distribution pattern of irrigation water use efficiency.

S2.3 Data validation of ChinaCropWF

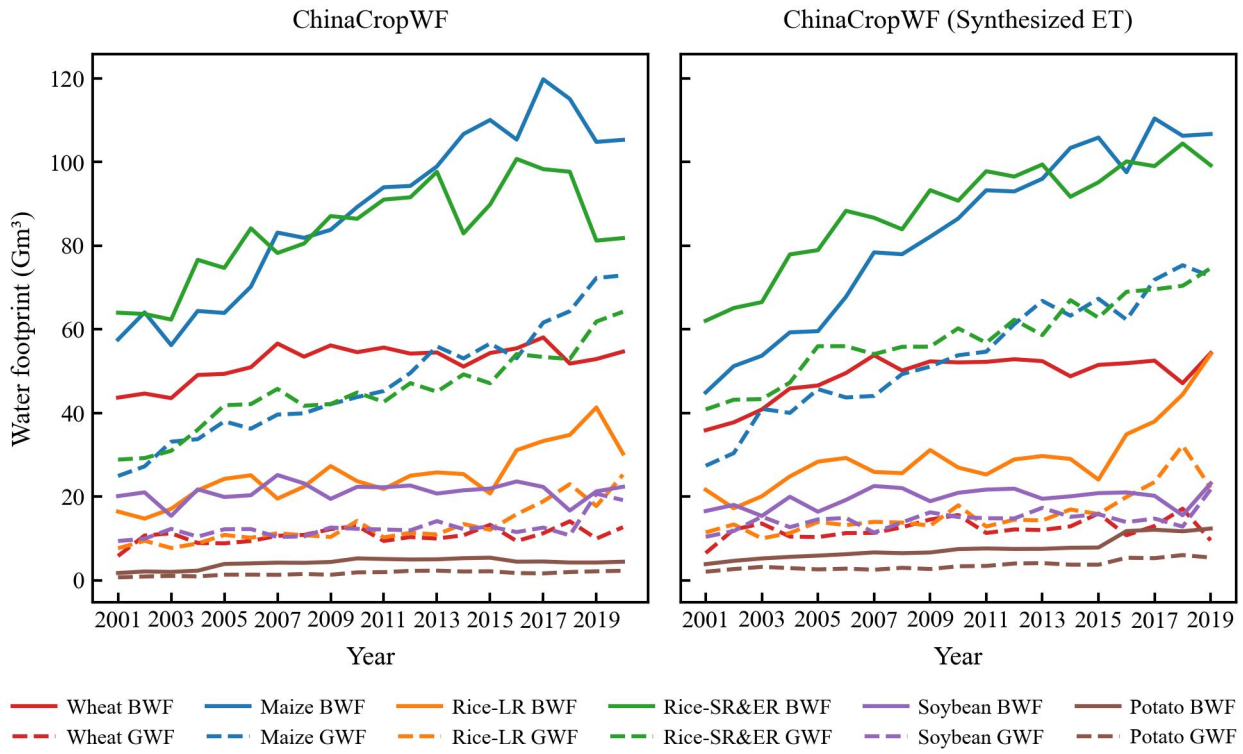

Figure S3: Comparison of crops water footprint based on two evapotranspiration products.

S2.4 Comparison of different water footprint datasets

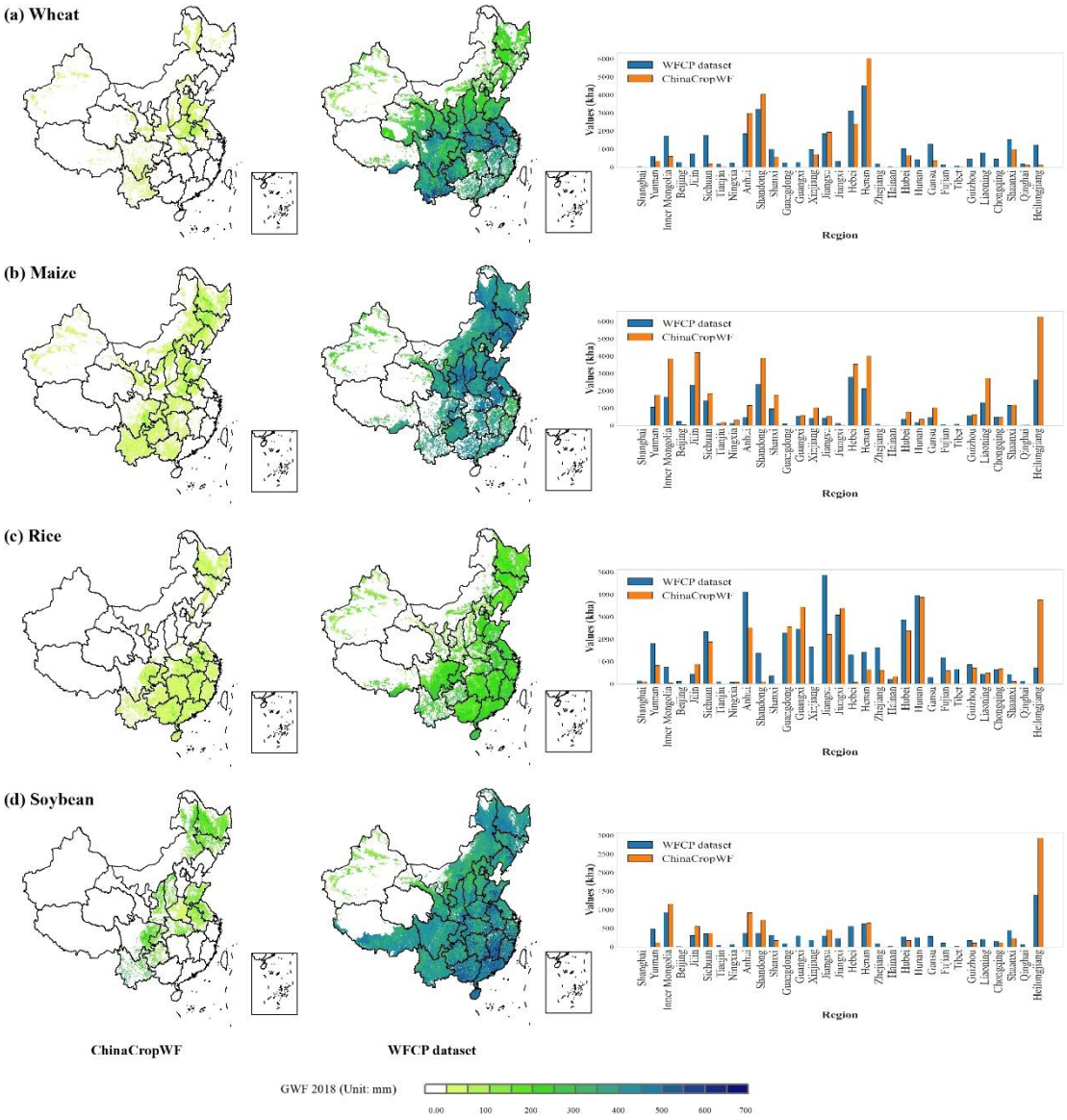

Figure S4: Comparison of planting area of different water footprint datasets.