Dear Prof. Tian,

We are very grateful for your detailed comments and constructive suggestions on our manuscript (essd-2025-107). We revised the manuscript according to your comments and provided a point-by-point reply (in blue color) as follows. Our revisions in the manuscript and supplementary materials are marked in blue.

Comment 1: Different models and datasets are used for the production of different ecosystem service maps. A brief justification of how these input data sources and models were selected, and why they are appropriate, would help readers better understand their role in generating the maps.

Response: Thanks for your suggestion. We have now explained the suitability of each model for high-resolution applications and provided relevant citations to support their use at 30 m in the Data and Methods section **(Lines 132-148)**. The detailed content is as follows:

The models and input data were based on the following principles: (1) Widely used of the models: The selected models (e.g., CASA, RUSLE) are well-established within the field of ecosystem service assessment. Their principles are mature and have been extensively validated in applications at global and regional scales, which facilitates the comparison of our results with existing studies. (2) Data availability and model compatibility: The selected models are compatible with the multi-source remote sensing, meteorological, soil, and topographic data collected for this study, ensuring the feasibility of the assessment. (3) Suitability for spatially explicit assessment: All models are capable of spatially explicit calculation, which allows them to fully utilize the 30-meter high-resolution spatial data to generate detailed distribution maps, meeting the accuracy requirements for refined management and policy formulation. The application of these models at this fine resolution is well-supported by previous studies. The CASA model has been successfully applied to estimate China's land net primary productivity (NPP) data with high accuracy (Sun et al., 2021; Zhang et al., 2023). Similarly, both the RUSLE and RWEQ models have been successfully applied at high resolution for soil erosion and sandstorm prevention mapping, respectively, demonstrating their suitability for high-resolution assessment (Zong et al., 2025; Yang et al., 2025). The InVEST has proved to be suitable for large-scale water yield assessment in China (Yin et al., 2020). This capability meets the accuracy requirements for refined management and policy formulation.

Sun, J., Yue, Y., Niu, H.: Evaluation of NPP using three models compared with MODIS-NPP data over China, PLoS ONE., 16(11), e0252149, https://doi.org/10.1371/journal.pone.0252149, 2021.

Zhang, Z., Zhao, W., Liu, Y., Pereira, P.: Impacts of urbanisation on vegetation dynamics in Chinese cities, Environ. Impact Assess. Rev., 103, 107227, https://doi.org/10.1016/j.eiar.2023.107227, 2023.

Zong, R., Fang, N., Zeng, Y., Lu, X., Wang, Z., Dai, W., Shi, Z.: Soil Conservation Benefits of Ecological Programs Promote Sustainable Restoration, Earth's Future., 13, e2024EF005287, https://doi.org/10.1029/2024EF005287, 2025.

Yang, J., Wang, S., Feng, J., He, H., Wang, L., Sun, Z., Zheng, C.: New 30-m resolution dataset reveals declining soil erosion with regional increases across Chinese mainland (1990-2022), Remote Sens. Environ., 323, 114681, https://doi.org/10.1016/j.rse.2025.114681, 2025.

Yin, L., Wang, X., Wang, Y.: Water Yield Product 1-km Grid Yearly Dataset in National Barrier

Zone of China, 1-km resolution dataset of water yield in the National Ecological Barrier Zone (2000-2015), Journal of Global Change Data & Discovery., 4(4), 332-337, https://doi.org/10.3974/geodp.2020.04.03, 2020.

Comment 2: Building on the first point, the results and discussion section would benefit from more detailed validation and interpretation of the different ecosystem service products. Given the comprehensiveness of the dataset—covering multiple services and a large geospatial region-such validation could be presented separately for each service, and, if data permit, also stratified by land cover type. This would improve readers' confidence in the dataset and highlight its applicability.

Response: Thanks for your insightful suggestion. We have developed an indirect cross-validation framework that integrates multiple dataset sources and land cover stratification. The framework systematically leverages diverse, authoritative proxy datasets to triangulate the reliability of the simulations from various perspectives, thereby minimizing dependence on any single observational source. Beyond multi-source datasets, we stratify all evaluations by land cover class (e.g., cropland, forest, grassland, shrubland, and barren), enabling class-specific accuracy diagnostics and revealing class-dependent biases that might be masked in aggregate assessments.

The verification results show that the ecosystem services simulated in this study have high accuracy both overall (Fig. 2, Fig. 3, Fig. 4, Fig. 5) and by land cover type (Fig. S1, Fig. S2, Fig. S3, Fig. S4) compared with the published data products. In the verification section, we discussed the reasons for the deviation between the model simulation service and the published data in the verification of overall and map land cover classification **(Lines 385-529)**.

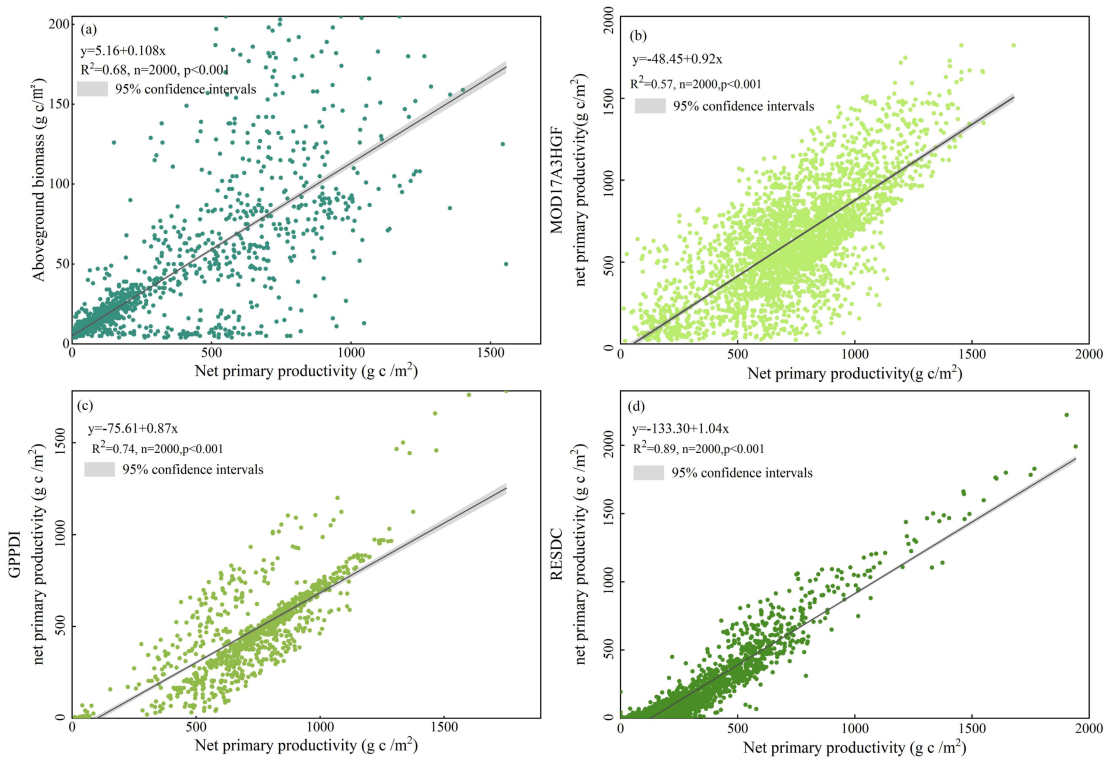

Figure 2: Validation of the NPP in this study, (a) the aboveground biomass and NPP of China in 2010, (b) the NPP estimated in this study and MODIS/Terra Net Primary Production Gap-Filled Yearly L4 (MOD17A3HGF), (c) the NPP estimated in this study and Global Primary Production Data Initiative (GPPDI) NPP data, (d) the NPP estimated in this study and Resource and Environment Science and

Data Center (RESDC) NPP data.

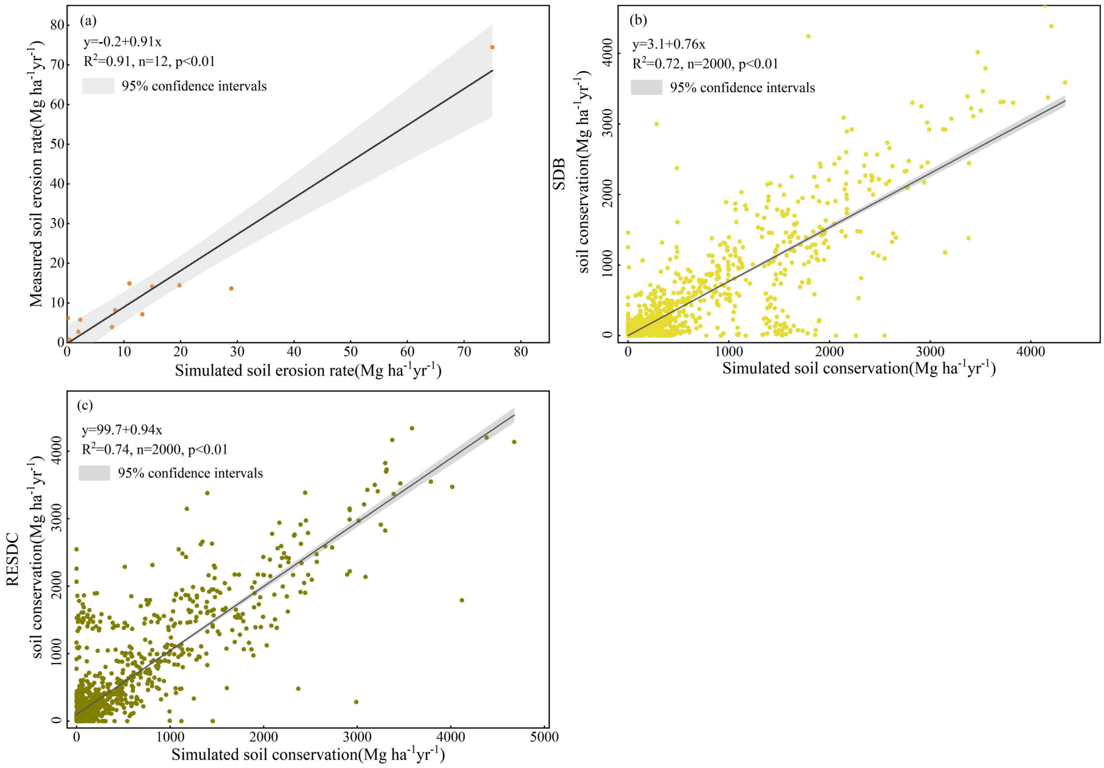

Figure 3: Validation of the soil conservation in this study, (a) the simulations and measurement of annual soil erosion rates for six river basins, including those of the Yangtze, Yellow, Haihe, Huaihe, Pearl, and Songhua and Liaohe in 2000 and 2010, (b) the soil conservation simulated in this study and Science Data Bank (SDB) soil conservation data in 2010, (c) the soil conservation simulated in this study and Resource and Environment Science and Data Center (RESDC) soil conservation data.

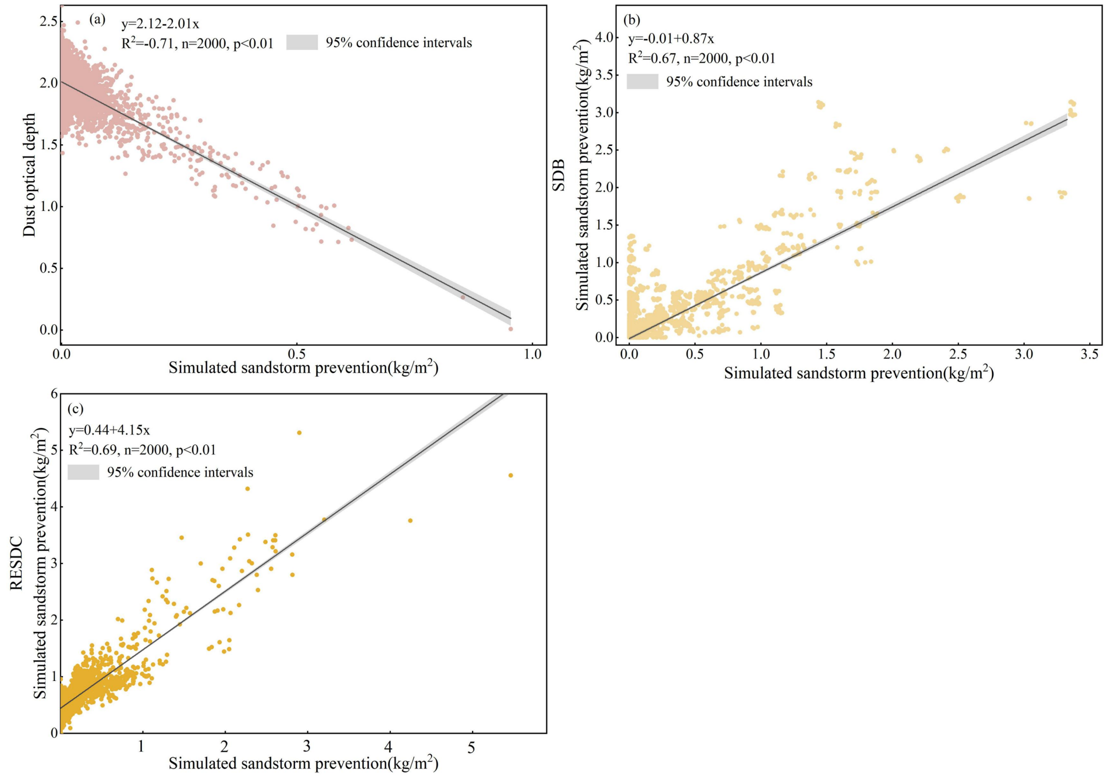

Figure 4: Validation of the sandstorm prevention in this study, (a) the simulated sandstorm prevention and dust optical depth of China in 2010, (b) the sandstorm prevention simulated in this study and Science Data Bank (SDB) soil conservation data in 2010, (c) the sandstorm prevention simulated in

this study and Resource and Environment Science and Data Center (RESDC) soil conservation data.

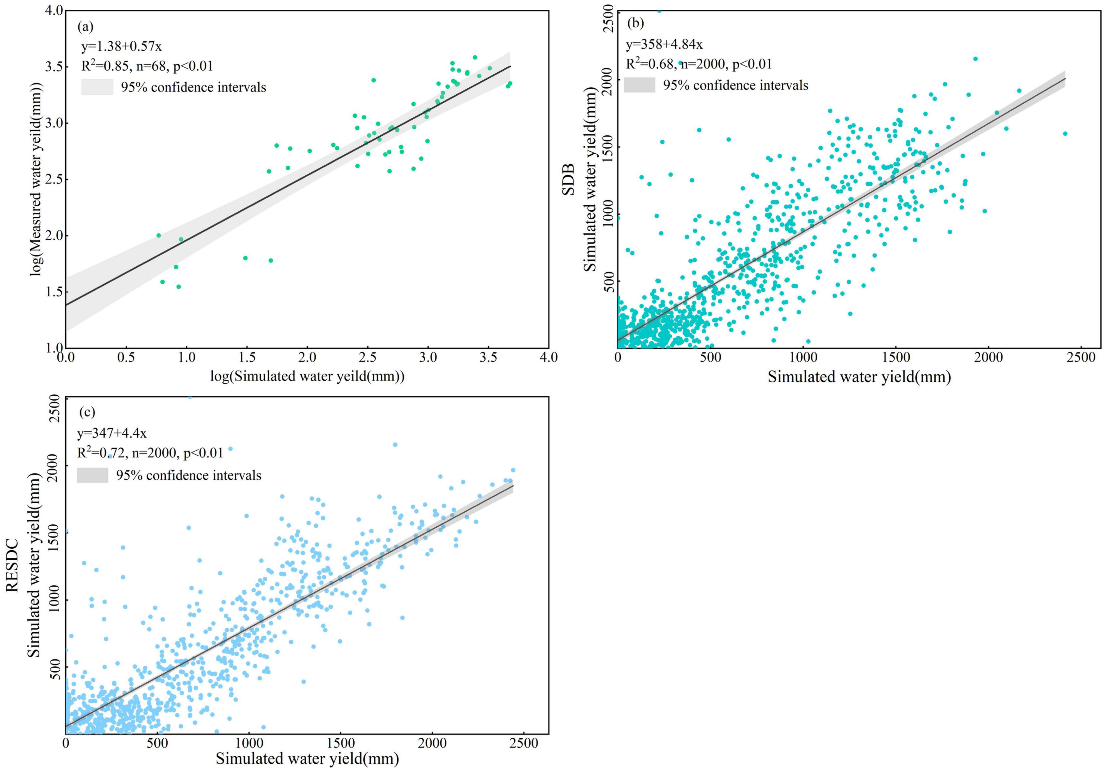

Figure 5: Validation of the water yield in this study, (a) the simulations and measurements of water yield for 34 provinces in 2000 and 2020. (b) the water yield simulated in this study and Science Data Bank (SDB) water yield data in 2010, (c) the water yield simulated in this study and Resource and Environment Science and Data Center (RESDC) water yield data.

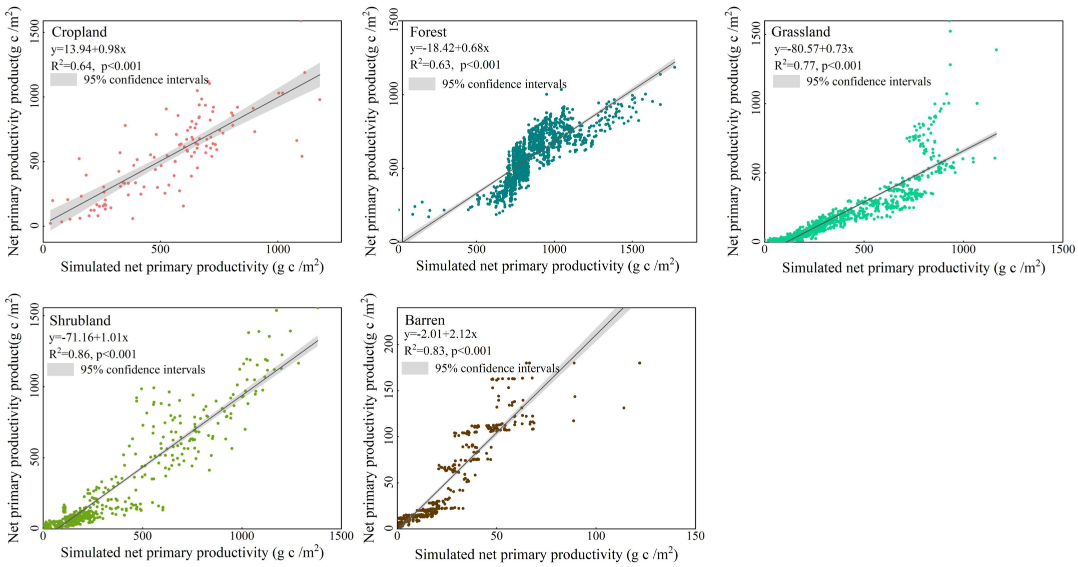

Figure S1: The verification map of the land cover type of NPP simulated by the CASA model and the published NPP products.

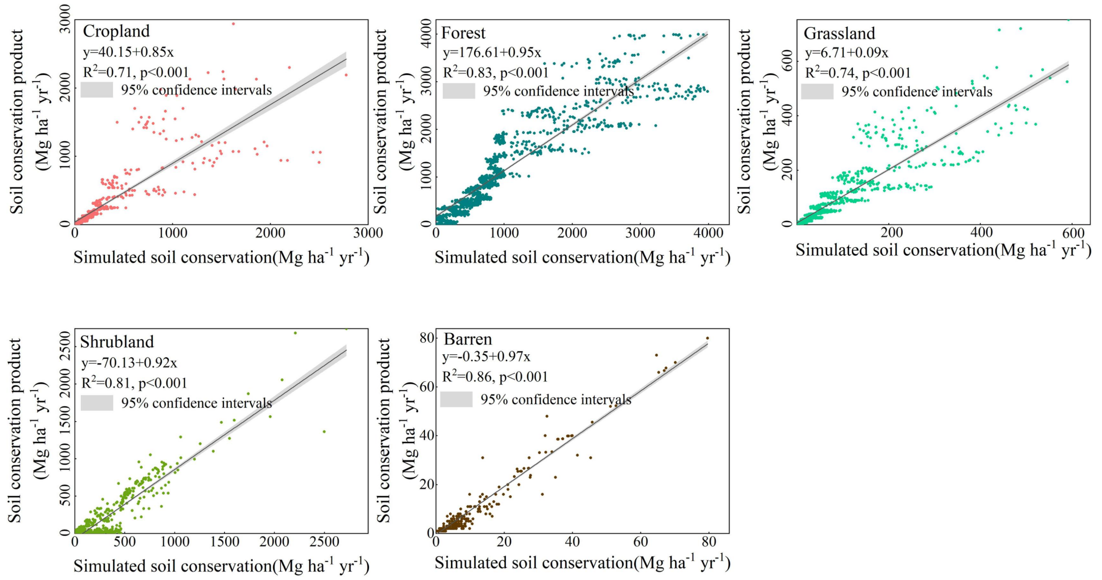

Figure S2: The verification map of the land cover type of soil conservation simulated by the RUSLE model and the published soil conservation products.

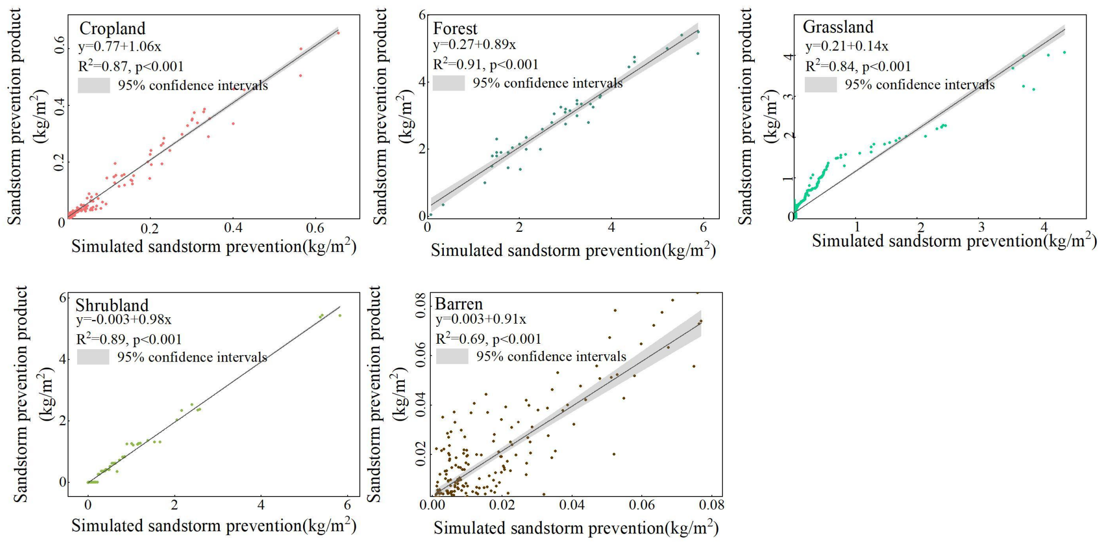

Figure S3: The verification map of the land cover type of sandstorm prevention simulated by the RWEQ model and the published sandstorm prevention products.

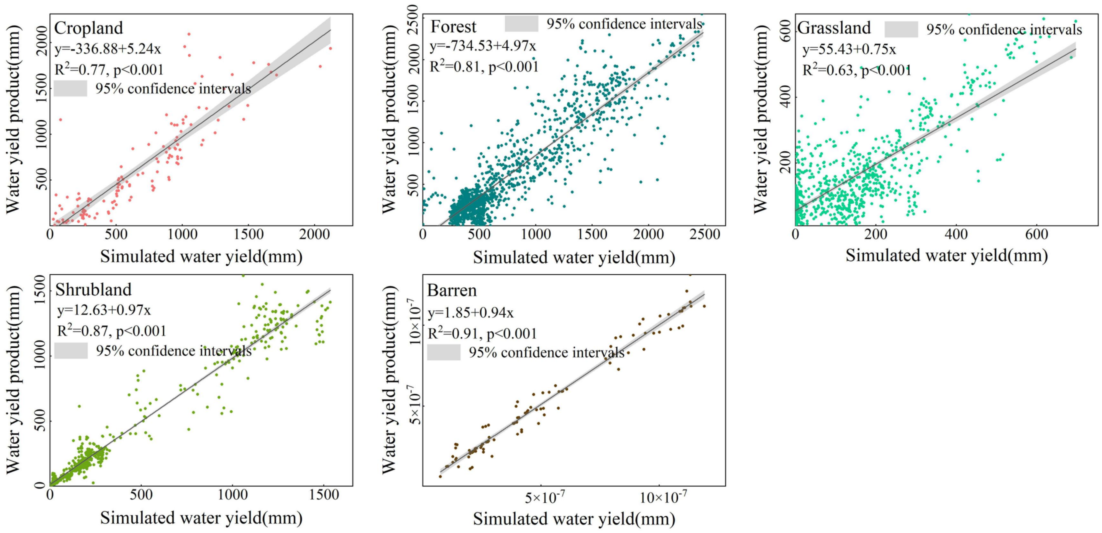

Figure S4: The verification map of the land cover type of water yield simulated by the InVEST model and the published water yield products.

Comment 3: Could the authors clarify the distinction between land cover and land use in this section? In one sentence, the term land use is used, while in the following text land cover appears. If the two concepts are intended to be treated as equivalent here, please provide a brief explanation; if not, it would be helpful to clarify why different terms are used.

Response: Thank you very much for raising this issue. We acknowledge that the terminology used in that section was not precise enough. We confirm that the intended concept throughout our study is the physical description of the Earth's surface, i.e., Land Cover, rather than its socioeconomic function, i.e., Land Use. Following your suggestion, we have thoroughly reviewed the manuscript and use "Land cover" consistently throughout. This ensures greater accuracy and consistency in our terminology.

Comment 4: It would be helpful if the authors could elaborate on how the development of this new validation method contributes to tackling the challenge of insufficient data mentioned previously.

Response: Thank you very much for your helpful suggestion. We fully agree that elaborating on how our validation approach tackles the challenge of data scarcity is essential. We have now added the advantages of our validation method in the Validation section **(Lines 366-372)**. The detailed content is as follows:

We have developed an indirect cross-validation framework that integrates multiple dataset sources and land cover stratification. The framework systematically leverages diverse, authoritative proxy datasets to triangulate the reliability of the simulations from various perspectives, thereby minimizing dependence on any single observational source. Beyond multi-source datasets, we stratify all evaluations by land cover class (e.g., cropland, forest, grassland, shrubland, and barren), enabling class-specific accuracy diagnostics and revealing class-dependent biases that might be masked in aggregate assessments.

Comment 5: Since multiple models and data sources are described in this section, and some inputs represent similar but distinct concepts while others rely on the same data source, it might be helpful to provide a summary table. Such a table could clearly match each model with its corresponding inputs and factors, which would improve clarity for the reader.

Response: Thanks for your detailed consideration. We agree with you that a summary table should be provided to improve clarity for the readers. We have now added the summary table in the ecosystem services assessment parameters to clearly match each model with its corresponding input and factors (Table 1).

Table 1. Assessment model and input data used in this study.

Ecosystem service	Model	Parameter	Dataset	Resolution	Source
NPP	CASA	NDVI	Landsat 5 (2000 and 2010) and Landsat 8 (2020) Level 2, Collection 2, Tier 1 data	30 m	https://earthexplorer.usgs.gov/
		Temperature	A monthly average temperature dataset with a resolution of 1km in China from 1901 to 2024	1 km	http://www.geodata.cn/data/
		Precipitation	A monthly precipitation dataset with a resolution of 1km in China from 1901 to 2024	1 km	http://www.geodata.cn/data/
		Landcover	GlobeLand 30	30 m	http://globeland30.org/
		Evapotranspiration	MOD16A2	500 m	https://modis.gsfc.nasa.gov/
		Potential evapotranspiration	MOD16A2	500 m	https://modis.gsfc.nasa.gov/
Soil conservation	RUSLE	NDVI	Landsat 5 (2000 and 2010) and Landsat 8 (2020) Level 2, Collection 2, Tier 1 data	30 m	https://earthexplorer.usgs.gov/
		Monthly precipitation	A monthly precipitation dataset with a resolution of 1km in China from 1901 to 2024	1 km	http://www.geodata.cn/data/
		Soil properties	SoilGrids V2.0	250 m	https://soilgrids.org/
		DEM	ASTER Global Digital Elevation Model V003	30 m	https://www.earthdata.nasa.gov/
Sandstorm prevention	RWEQ	Wind speed	ERA5 Hourly Data on Single Levels	0.01°	https://developers.google.com/earthengine/datasets/
		Soil properties	SoilGrids V2.0	250 m	https://soilgrids.org/

		Snow cover	MOD10A2	0.05°	https://modis.gsfc.nasa.gov/
		Potential evapotranspiration	MOD16A2	500 m	https://modis.gsfc.nasa.gov/
		Precipitation	A monthly average temperature dataset with a resolution of 1km in China from 1901 to 2024	1 km	http://www.geodata.cn/data/
		Temperature	A monthly precipitation dataset with a resolution of 1km in China from 1901 to 2024	1 km	http://www.geodata.cn/data/
		DEM	ASTER Global Digital Elevation Model V003	30 m	https://www.earthdata.nasa.gov/
Water yield	Invest	Precipitation	A monthly average temperature dataset with a resolution of 1km in China from 1901 to 2024	1 km	http://www.geodata.cn/data/
		Potential evapotranspiration	MOD16A2	500 m	https://modis.gsfc.nasa.gov/
		Soil properties	SoilGrids V2.0	250 m	https://soilgrids.org/
		Landcover	GlobeLand 30	30 m	http://globeland30.org/
		Watersheds	/	/	http://www.mwr.gov.cn/

Comment 6: It may be valuable to conduct validation by land cover type, since many products depend on land cover information. An accuracy assessment stratified by land cover would provide readers with relevant insights into how performance varies across different land cover classes.

Response: Thanks for your suggestion, we have now added the verification results of the ecosystem service dataset simulated in this study and the published datasets by land cover classes, indicating that the four ecosystem service data have high simulation accuracy in cropland, forest, grassland, shrubland, and barren (Fig. S1, Fig. S2, Fig. S3, and Fig. S4). In the manuscript, we expounded the reasons for the deviations in the verification by land type of ecosystem services, specifically in the verification section **(Lines 385-529)**. The detailed content is as follows:

NPP shows a clearly land cover class dependent (Fig. S1). In croplands, strong management signals—such as irrigation, multiple cropping, fertilization, and harvest—are imperfectly captured by generic drivers, resulting in a larger scatter and mismatch. In forests, NDVI saturation and topographic illumination in complex terrain dampen high values and flatten slopes, while differences in disturbance and turnover assumptions add bias. Grasslands are governed by water limitation, so errors in precipitation forcings and residual cloud/snow contamination mainly affect the low-value range. Shrublands show the best agreement, likely because disturbance is weaker and the simulated NPP and NPP products response is closer to linear. In barren lands, sparse vegetation also avoids NDVI saturation, preserving a near-linear radiometric – productivity relationship that reduces slope dampening seen in dense forests. Moreover, the extensive homogeneous patches in these areas ensure higher land cover purity at 30 m resolution, weakening mixed-pixel and misclassification effects. This advantage is further enhanced by the typically low cloudiness in arid regions, which minimizes residual cloud and shadow errors. Together, these conditions foster stronger consistency across datasets.

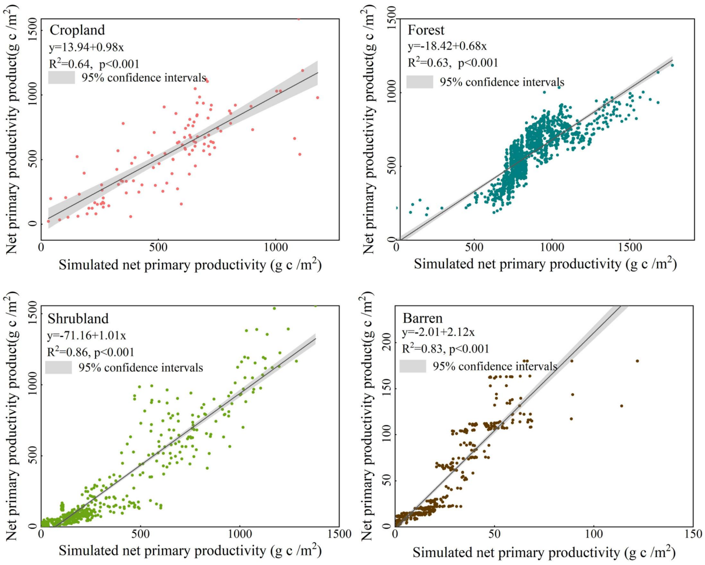

Figure S1: The verification map of the land cover type of NPP simulated by the CASA model and the published NPP products.

The verification accuracy of soil conservation and published products shows significant differences in land cover classes (Fig. S2). The barren areas are generally poorly managed, exhibit homogeneous and blocky patterns, and are primarily influenced by the LS and K factors, leading to the highest consistency across datasets. Forest areas, characterized by low and stable C-factor values, are nevertheless affected by topographic and observational artifacts such as terrain shadows and DEM smoothing. Extreme events such as landslides and gully erosion also introduce outliers. Shrublands maintain a stable structure and thus achieve relatively high estimation precision. In contrast, grasslands are influenced by episodic rainfall events and grazing disturbances, while residual cloud and snow cover increase dispersion in the low-value range. Cropland exhibits the largest uncertainty, mainly due to the high spatiotemporal heterogeneity of the P-factor and the effects of irrigation and tillage practices.

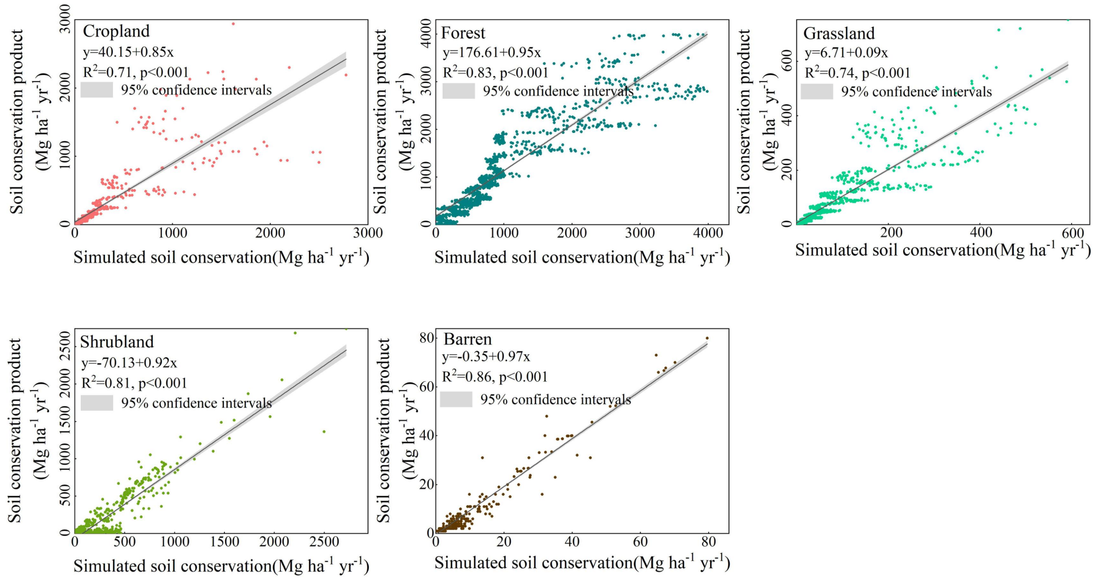

Figure S2: The verification map of the land cover type of soil conservation simulated by the RUSLE model and the published soil conservation products.

Land cover classes validation shows strong agreement for forest, shrubland, grassland, and cropland, whereas barren areas perform less well (Fig. S3). The differences stem from the surface roughness, the timeliness of wind and soil-moisture forcing, and classification/scale effects. Forests and shrublands supply stable roughness elements, so P conservation practices (shelterbelts/barriers) are captured consistently across products. Grasslands and croplands also agree well but exhibit slightly larger scatter at low values due to phenology, irrigation/tillage, and moisture pulses. In barren lands, absolute magnitudes are small and highly sensitive to gust thresholds and fine-fraction composition.

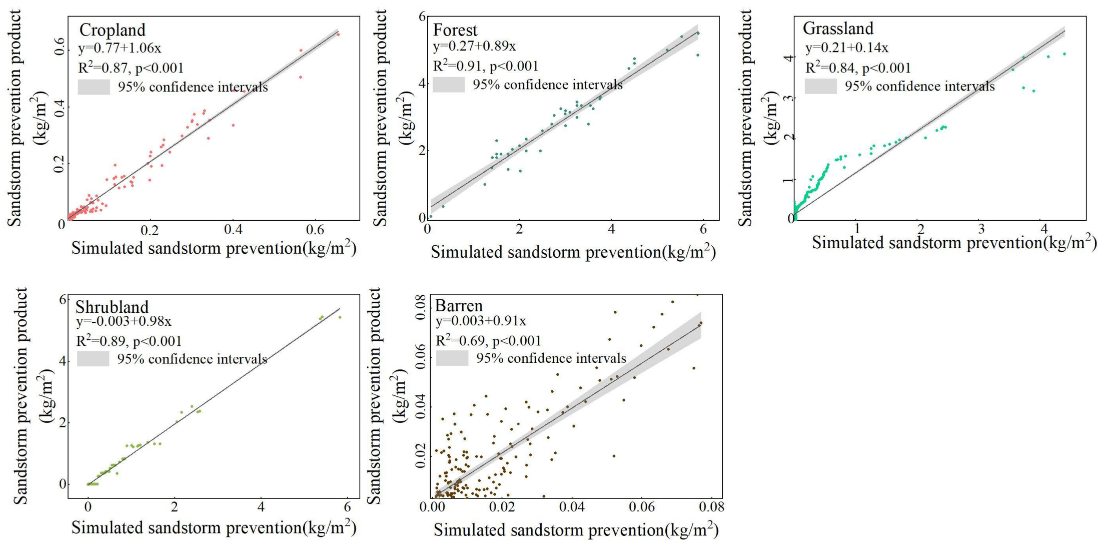

Figure S3: The verification map of the land cover type of sandstorm prevention simulated by the RWEQ model and the published sandstorm prevention products.

Land cover classes validation shows the strongest agreement for shrubland and barren, while forest and cropland correlate well but exhibit steeper slopes and negative intercepts, and grassland performs the weakest (Fig. S4). These differences stem from the Budyko simplifications in

InVEST and scale mismatches. Croplands are strongly affected by irrigation, runoff regulation, and return flows, raising baselines in published products. Forests reflect orographic precipitation biases, snow/ice melt, and baseflow, making external estimates higher. Grasslands show larger dispersion due to water-stress pulses, grazing effects, and heterogeneous PAWC/Kc. By contrast, shrubland and barren areas have simpler processes and weaker management, resulting in closer precipitation and ET partitioning across products.

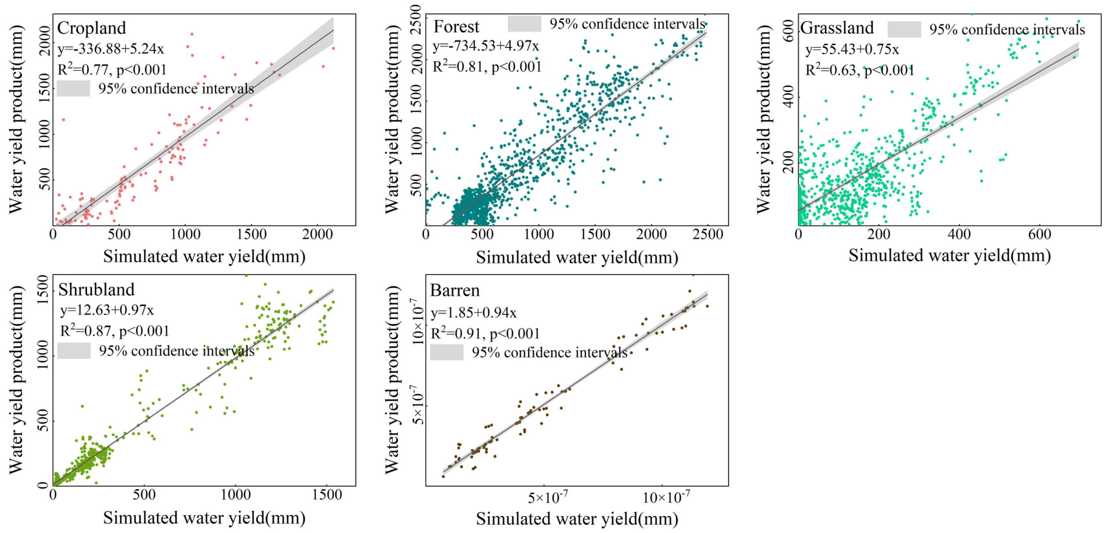

Figure S4: The verification map of the land cover type of water yield simulated by the InVEST model and the published water yield products.

Comment 7: Consider explaining why only 50 points were used for validation, given that for other models larger sample sizes are used (Line 306).

Response: We appreciate the reviewer's comment concerning the validation sample size. We now unified the validation of all ecosystem service models to 2,000 random sampling points in our manuscript. This uniform sample size was chosen to ensure a statistically robust and comparable validation across all services (Fig. 2, Fig. 3, Fig. 4, and Fig. 5). The results show that all four ecosystem services have high simulation accuracy. In the manuscript, we expounded the reasons for the deviations in the verification, specifically in the verification section.

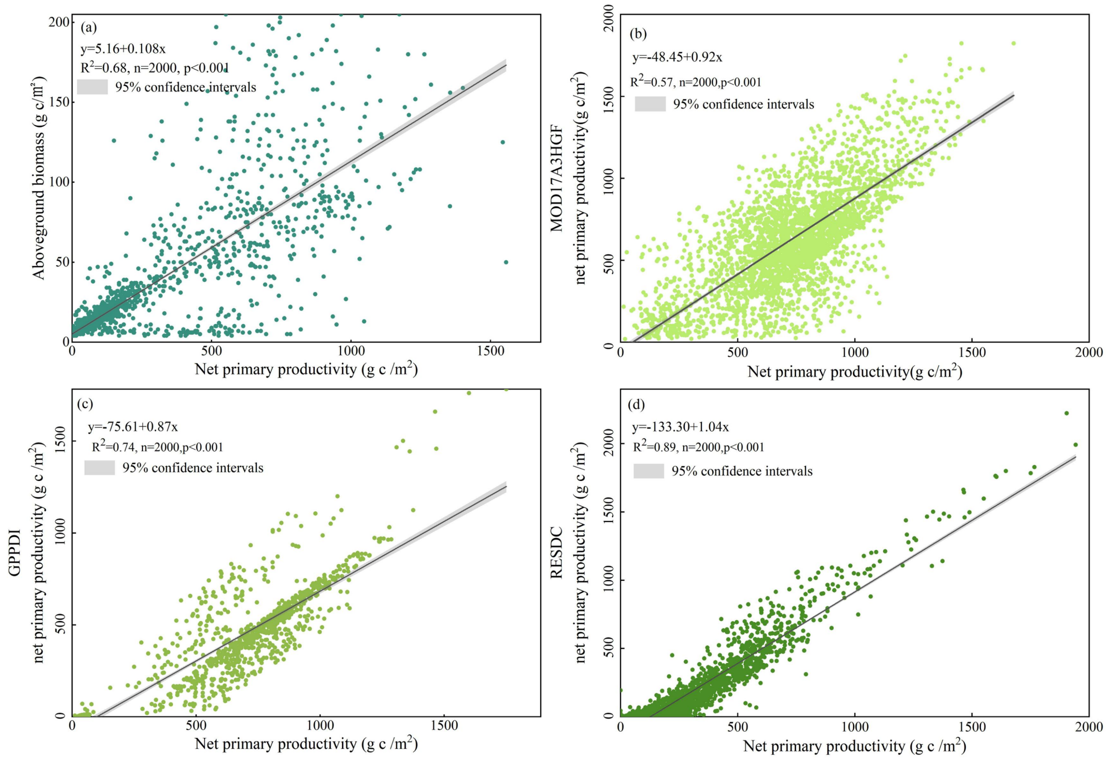

Figure 2: Validation of the NPP in this study, (a) the aboveground biomass and NPP of China in 2010, (b) the NPP estimated in this study and MODIS/Terra Net Primary Production Gap-Filled Yearly L4 (MOD17A3HGF), (c) the NPP estimated in this study and Global Primary Production Data Initiative (GPPDI) NPP data, (d) the NPP estimated in this study and Resource and Environment Science and Data Center (RESDC) NPP data.

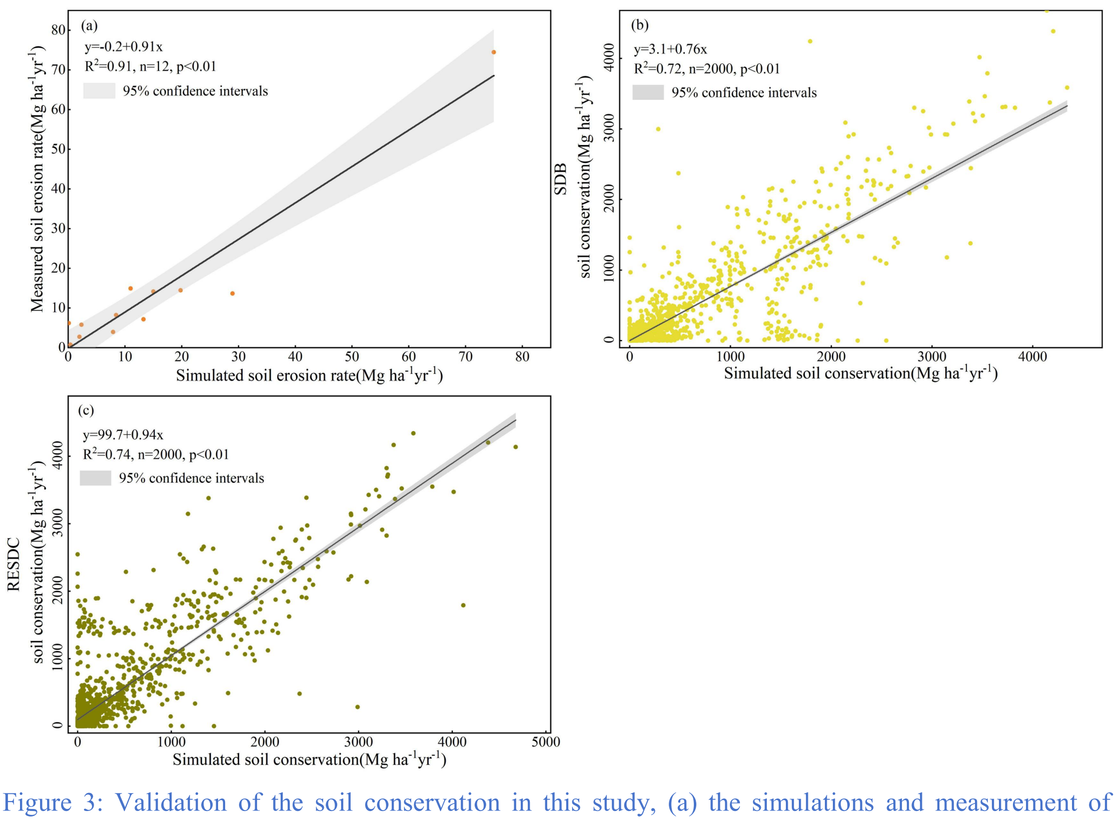

Figure 3: Validation of the soil conservation in this study, (a) the simulations and measurement of annual soil erosion rates for six river basins, including those of the Yangtze, Yellow, Haihe, Huaihe, Pearl, and Songhua and Liaohe in 2000 and 2010, (b) the soil conservation simulated in this study and Science Data Bank (SDB) soil conservation data in 2010, (c) the soil conservation simulated in this

study and Resource and Environment Science and Data Center (RESDC) soil conservation data.

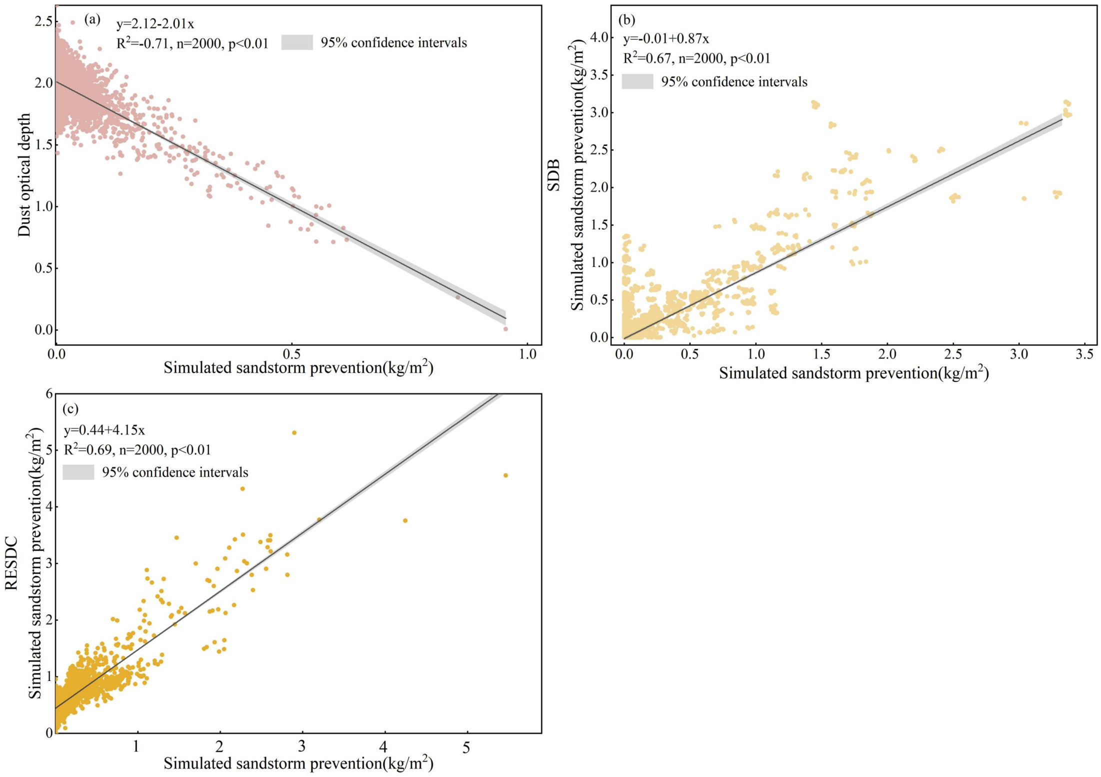

Figure 4: Validation of the sandstorm prevention in this study, (a) the simulated sandstorm prevention and dust optical depth of China in 2010, (b) the sandstorm prevention simulated in this study and Science Data Bank (SDB) soil conservation data in 2010, (c) the sandstorm prevention simulated in this study and Resource and Environment Science and Data Center (RESDC) soil conservation data.

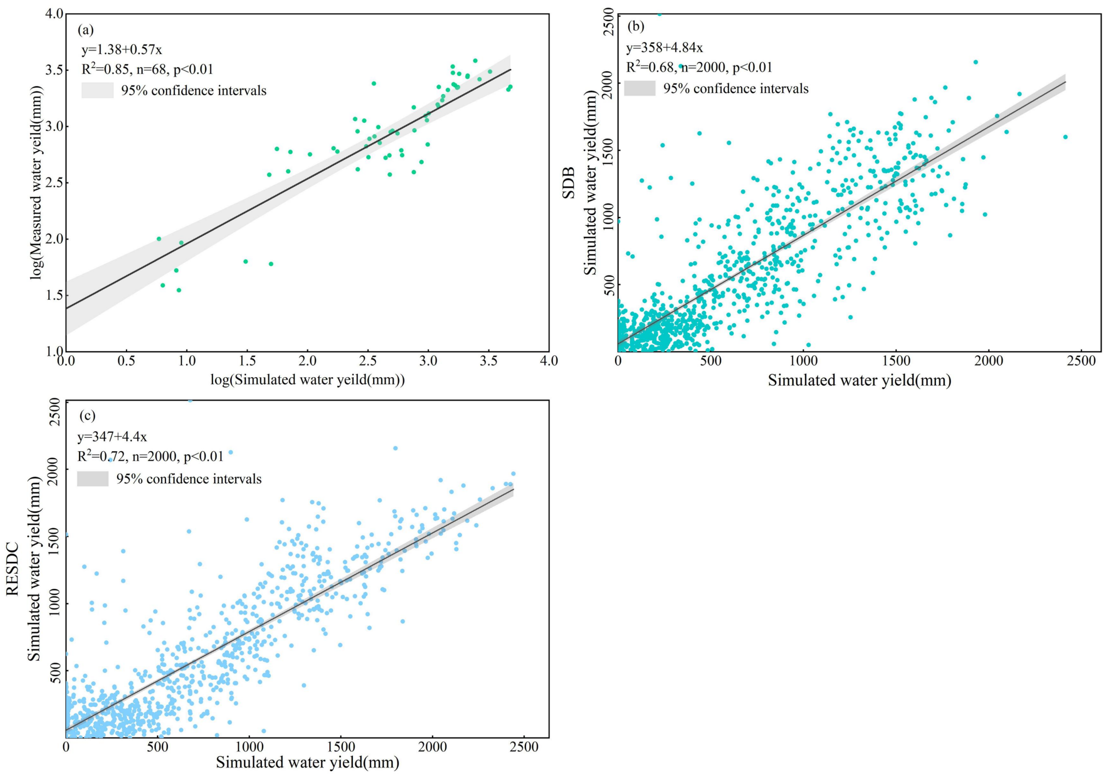

Figure 5: Validation of the water yield in this study, (a) the simulations and measurements of water yield for 34 provinces in 2000 and 2020. (b) The water yield simulated in this study and Science Data Bank (SDB) water yield data in 2010, (c) the water yield simulated in this study and Resource and Environment Science and Data Center (RESDC) water yield data.

Comment 8: It is good to see that this dataset has been validated directly and indirectly in this study (e.g., Figures 2 and 3). To further strengthen the discussion, it would be valuable to address why discrepancies and mismatches with other datasets still appear, and what might explain these differences. For example, when compared with different NPP studies, varying results are observed discussing potential reasons for this would give readers a better impression of the dataset's strengths and limitations. In addition, it may be advisable to phrase conclusions with more caution: rather than stating that 'this dataset showed higher accuracy and reliability,' the validation results suggest that different service datasets perform differently and should be discussed separately.

Response: Thanks for your comment. We have revised the Validation section to explain why discrepancies with other products may occur explicitly. We also expanded the discussion on land cover classes. We replaced the earlier broad statement about "higher accuracy and reliability" with a more cautious conclusion, emphasizing overall agreement while recognizing dataset and class-specific variation **(Lines 366-529)**. The detailed content is as follows:

We have developed an indirect cross-validation framework that integrates multiple dataset sources and land cover stratification. The framework systematically leverages diverse, authoritative proxy datasets to triangulate the reliability of the simulations from multiple perspectives, thereby minimizing dependence on any single observational source. Beyond multi-source datasets, we stratify all evaluations by land cover class (e.g., cropland, forest, grassland, shrubland, and barren), enabling class-specific accuracy diagnostics and revealing class-dependent biases that might be masked in aggregate assessments.

This study utilized simulated Net Primary Productivity (NPP) data and existing remote sensing datasets for cross-validation, addressing the scarcity of large-scale biomass monitoring data. Spawn et al. (2020) provided a global 300 m resolution map of aboveground and belowground biomass carbon density for 2010. This dataset was rigorously validated and quality assessed by its original producers. This study randomly generated 2000 points on the map of China and extracted the values of the simulated NPP and Spawn's datasets in 2010. This study then performed a correlation analysis, with the results shown in Fig. 2a. In addition, the NPP estimated in this study is multi-year monthly data, this study separately cross-validates the NPP for multiple years with remote sensing datasets (MODIS/Terra Net Primary Production Gap-Filled Yearly L4 (MOD17A3HGF) (Fig. 2b), Global Primary Production Data Initiative (GPPDI) (Fig. 2c), and Resource and Environment Science and Data Center (RESDC) (Fig. 2d). The results show that the NPP simulated by the CASA model have good consistency with the available biomass carbon density and NPP datasets.

Despite the overall agreement shown in Fig. 2, differences with other datasets are expected because the compared products diverge in concepts, algorithms, inputs, and scales. Our dataset estimates net primary productivity (NPP). In contrast, the biomass map of Spawn et al. (2020) represents carbon stocks for 2010, stock flux comparisons are sensitive to assumptions about disturbance, harvest, and turnover. CASA model and MOD17A3HGF use different light-use-efficiency parameterization algorithms and environmental data (temperature and precipitation). GPPDI and RESDC further rely on distinct input data and modeling frameworks, which can lead to systematic offsets. Input data also vary (meteorological data, land cover maps, soil/terrain), and the spatial resolution is mismatched (30 m in this study and 1000 m for several products), so resampling and mixed pixels cause scale effects. NPP shows a clearly land cover class dependent (Fig. S1). In croplands, strong management signals—such as irrigation, multiple cropping, fertilization, and harvest—are imperfectly captured by generic drivers, resulting in a larger scatter and mismatch. In forests, NDVI saturation and topographic illumination in complex terrain dampen high values and flatten slopes, while differences in disturbance and turnover assumptions add bias. Grasslands are governed by water limitation, so errors in precipitation forcings and residual cloud/snow contamination mainly affect the low-value range. Shrublands show the best agreement, likely because disturbance is weaker and the simulated NPP and NPP products response is closer to linear. In barren lands, sparse vegetation also avoids NDVI saturation, preserving a near-linear radiometric–productivity relationship that reduces slope dampening seen in dense forests. Moreover, the extensive homogeneous patches in these areas ensure higher land cover purity at 30 m resolution, weakening mixed-pixel and misclassification effects. This advantage is further enhanced by the typically low cloudiness in arid regions, which minimizes residual cloud and shadow errors. Together, these conditions foster stronger consistency across datasets.

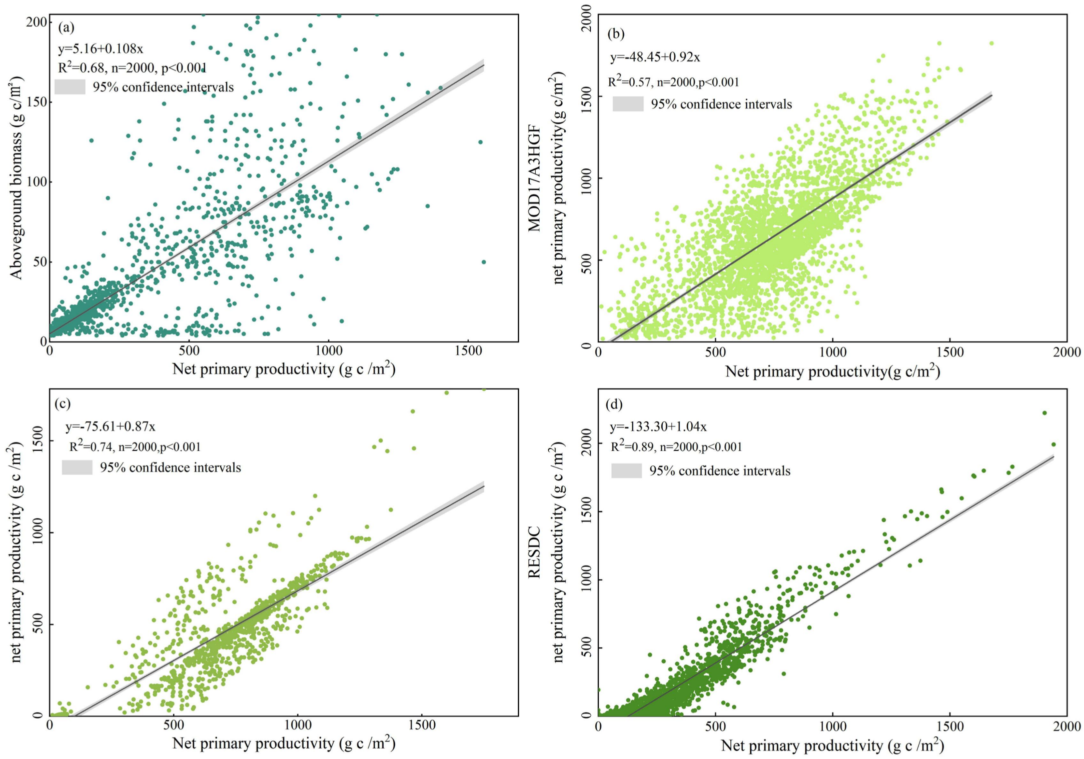

Figure 2: Validation of the NPP in this study, (a) the aboveground biomass and NPP of China in 2010, (b) the NPP estimated in this study and MODIS/Terra Net Primary Production Gap-Filled Yearly L4 (MOD17A3HGF), (c) the NPP estimated in this study and Global Primary Production Data Initiative (GPPDI) NPP data, (d) the NPP estimated in this study and Resource and Environment Science and Data Center (RESDC) NPP data.

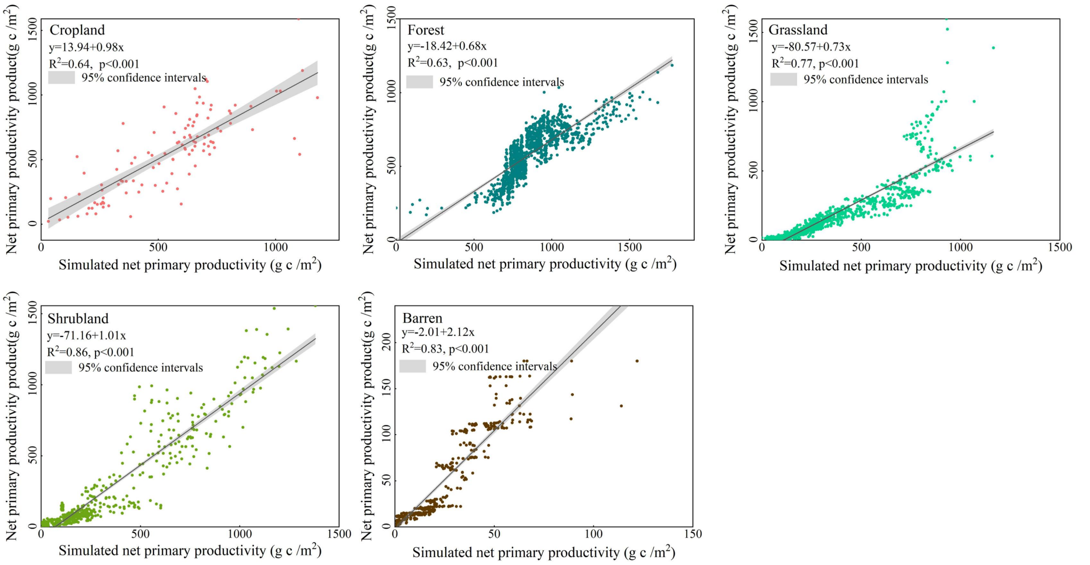

Figure S1: The verification map of the land cover type of NPP simulated by the CASA model and the published NPP products.

Obtaining observed soil conservation data is generally challenging. Since the soil conservation service is derived from soil erosion rates estimated by the RUSLE model, its reliability can be indirectly assessed by validating simulated soil erosion rates (Xiao et al., 2017). Therefore, this study used the watershed soil erosion data to evaluate the model's accuracy (Liu et al., 2023). The watersheds include the Yangtze, Yellow, Huai, and Hai River Basins. This study obtained the soil erosion rates of these watersheds from 2000 to 2020 from the China Soil and Water Conservation Bulletin (http://www.mwr.gov.cn/sj/tjgb/zgstbcgb/), and simulated erosion rates were extracted using basin vectors provided by the Water Resources Department. Based on these two datasets, this study performed a correlation analysis, with the results shown in Fig. 3a. In this study, we additionally cross-validated the simulated soil conservation with two published datasets - the Science Data Bank (SDB) soil conservation product for 2010 (Fig. 3b) and the Resource and Environment Science and Data Center (RESDC) soil conservation dataset (Fig. 3c).

At the basin scale, simulated erosion rates agree well with observations from the China Soil and Water Conservation Bulletin (Fig. 3a), indicating that the RUSLE-driven framework captures the dominant spatial and interannual gradients in water-driven erosion. Cross-comparison with two soil conservation products (SDB and RESDC; Fig. 3b and Fig. 3c) also shows good consistency, but systematic spreads are expected for several reasons. Firstly, RUSLE represents annual hillslope sheet/rill erosion; however, it does not explicitly model gully and bank erosion, landslides/debris flows, snowmelt pulses, and wind erosion. Secondly, parameter/input uncertainty also leads to verification bias. R factor (rainfall erosivity) is derived from station/reanalysis fields that under-resolve short-lived convective storms, K and LS factors depend on soil maps and DEM, C factor comes from NDVI maps and cloud/shadow residuals, and P factor (conservation practices) is often approximated by regional constants, missing local terracing/contouring/residue cover. Thirdly, scale/definition mismatches arise when 30 m maps are compared with 1000 m products. The verification accuracy of soil conservation and published products shows significant differences in land cover classes (Fig. S2). The barren areas are generally poorly managed, exhibit homogeneous and blocky patterns, and are primarily influenced by the LS and K factors, leading to the highest consistency across datasets. Forest areas, characterized by low and stable C-factor values, are nevertheless affected by topographic and observational artifacts such as terrain shadows and DEM smoothing. Extreme events such as landslides and gully erosion also introduce outliers. Shrublands maintain a stable structure and thus achieve relatively high estimation precision. In contrast, grasslands are influenced by episodic rainfall events and grazing disturbances, while residual cloud and snow cover increase dispersion in the low-value range. Cropland exhibits the greatest uncertainty, mainly due to the high spatio-temporal heterogeneity of the P-factor and the effects of irrigation and tillage practices.

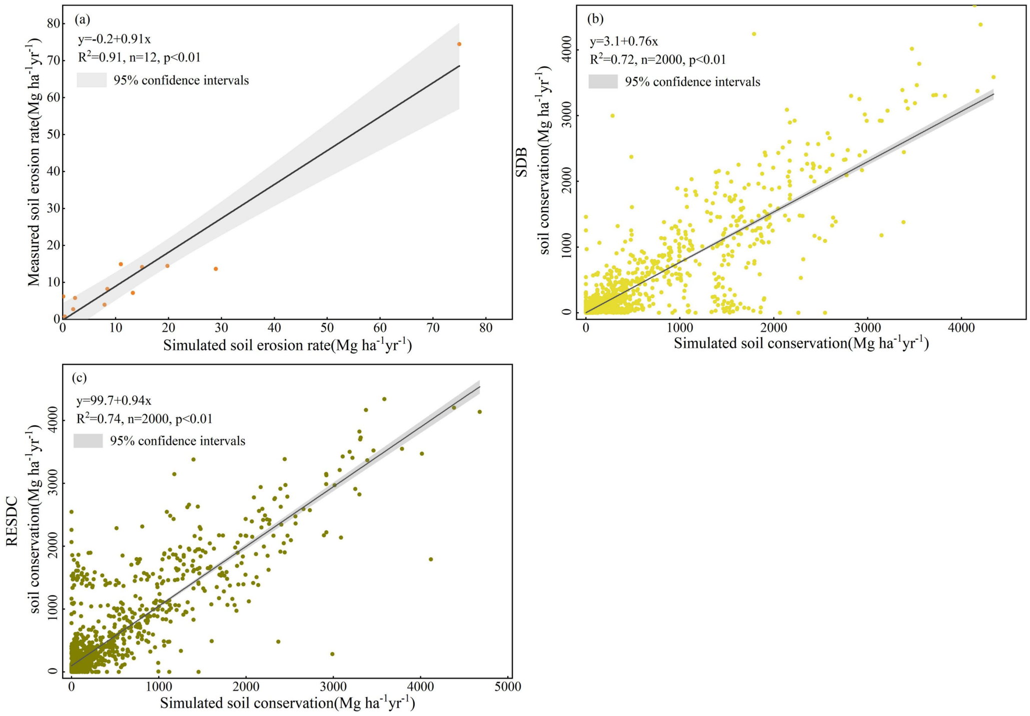

Figure 3: Validation of the soil conservation in this study, (a) the simulations and measurement of annual soil erosion rates for six river basins, including those of the Yangtze, Yellow, Haihe, Huaihe, Pearl, and Songhua and Liaohe in 2000 and 2010, (b) the soil conservation simulated in this study and Science Data Bank (SDB) soil conservation data in 2010, (c) the soil conservation simulated in this study and Resource and Environment Science and Data Center (RESDC) soil conservation data.

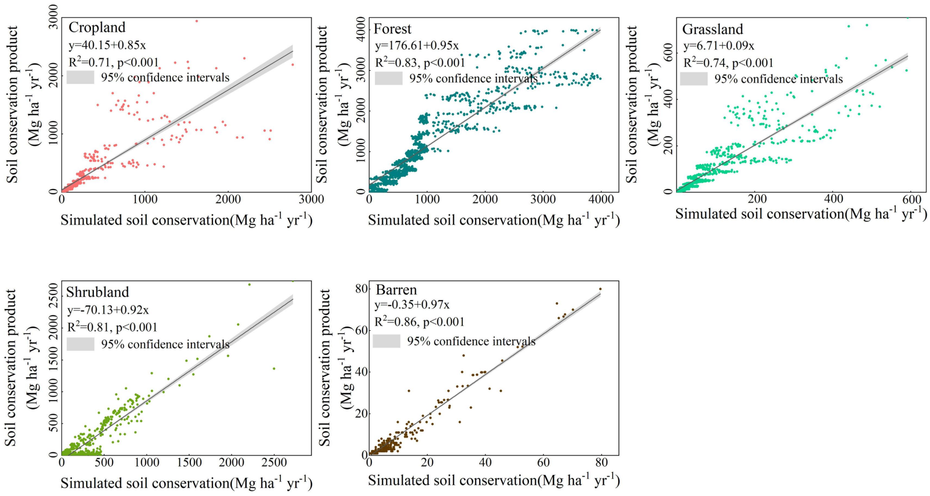

Figure S2: The verification map of the land cover type of soil conservation simulated by the RUSLE model and the published soil conservation products.

This study utilized simulated sandstorm prevention data and a remote sensing dataset for cross-validation, due to the limited availability of monitoring data on sandstorm prevention. Gkikas et al. (2022) quantified the dust optical depth and characterized its monthly and interannual variability at both global and regional scales for the period 2003-2017, using a fine spatial resolution (0.1° x 0.1°). This study randomly generated 2000 points on the map of China and extracted the values of the simulated sandstorm prevention data and Gkikas' datasets in 2010. This study then performed a correlation analysis, with the results shown in Fig. 4a. The simulated sandstorm prevention was also verified with two published datasets, namely the SDB sandstorm prevention data in 2010 (Fig. 4b) and the RESDC sandstorm prevention dataset (Fig. 4c), showing close consistency, thereby enhancing the credibility of the RWEQ model simulation results. The validation results show that simulated sandstorm prevention is negatively correlated with the dust optical depth (DOD) (Fig. 4a) - greater sandstorm prevention implies lower column dust - with residual spread driven by scope and scale mismatches. DOD integrates regional transport, vertical mixing, hygroscopic growth, and advection from remote sources, whereas the RWEQ model quantifies local emission control. Comparisons with other sandstorm prevention datasets (SDB and RESDC) reveal a positive spatial correlation (Fig. 4b and Fig. 4c), indicating a broadly consistent regional distribution. However, systematic offsets in slopes and intercepts are observed due to differences in drivers and parameterizations, such as wind speed, soil erodibility, and vegetation constraints. Land cover classes validation shows strong agreement for forest, shrubland, grassland, and cropland, whereas barren areas perform less well (Fig. S3). The differences stem from the surface roughness, the timeliness of wind and soil-moisture forcing, and classification/scale effects. Forests and shrublands supply stable roughness elements, so P conservation practices (shelterbelts/barriers) are captured consistently across products. Grasslands and croplands also agree well but exhibit slightly larger scatter at low values due to phenology, irrigation/tillage, and moisture pulses. In barren lands, absolute magnitudes are small and highly sensitive to gust thresholds and fine-fraction composition.

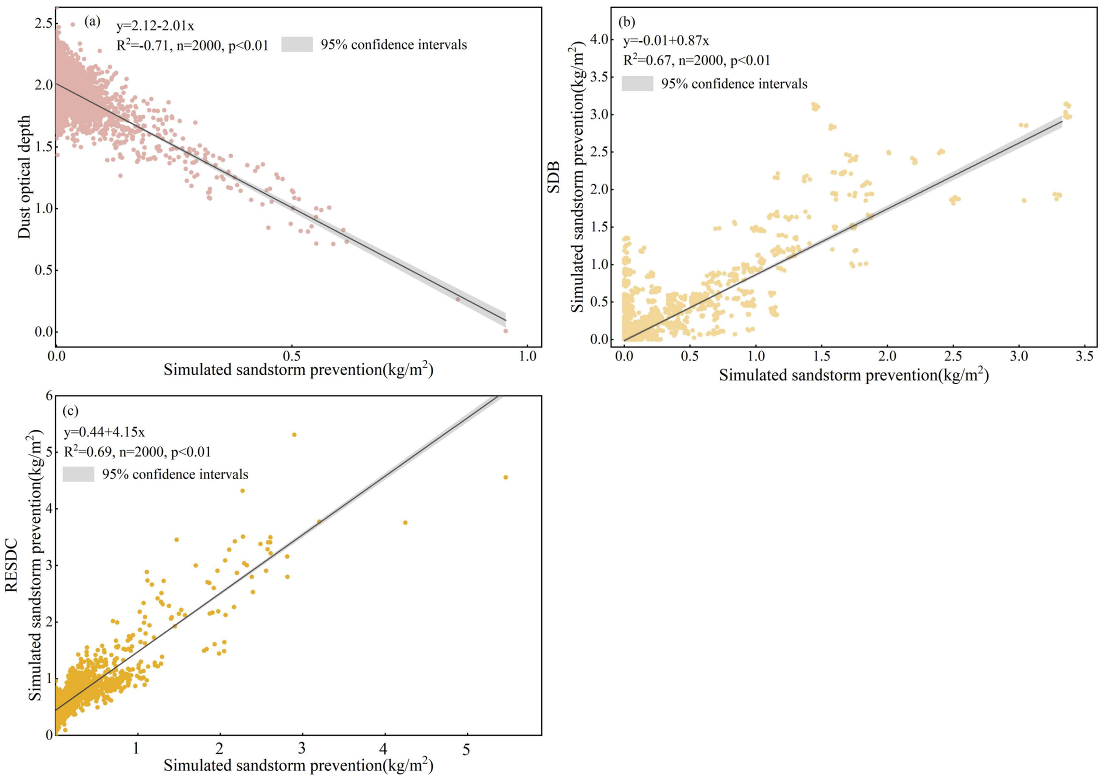

Figure 4: Validation of the sandstorm prevention in this study, (a) the simulated sandstorm prevention and dust optical depth of China in 2010, (b) the sandstorm prevention simulated in this study and

Science Data Bank (SDB) soil conservation data in 2010, (c) the sandstorm prevention simulated in this study and Resource and Environment Science and Data Center (RESDC) soil conservation data.

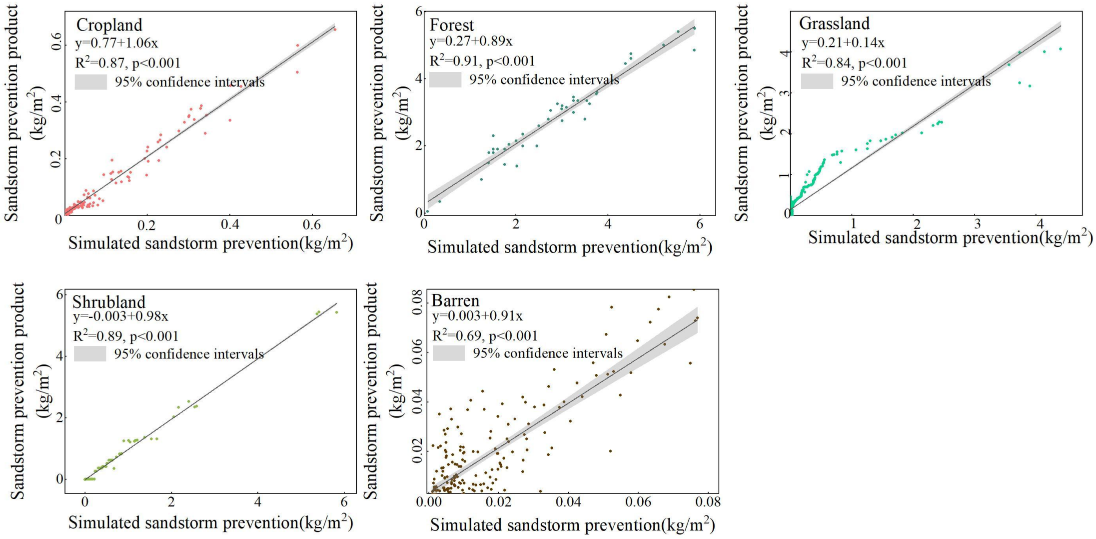

Figure S3: The verification map of the land cover type of sandstorm prevention simulated by the RWEQ model and the published sandstorm prevention products.

Surface water resource data for each province were obtained from the Water Resources Bulletin (http://www.mwr.gov.cn/sj/tjgb/szygb/) from 2000 to 2020, typically through field monitoring and statistical methods conducted by the water conservancy department. This study matched the water yield simulated by the InVEST model with the actual water yield data from the bulletin. To ensure consistency, this study aligned the data based on the same provinces and the same years. Due to missing data for some provinces in the year 2000, this study matched the data for 2010 and 2020 for analysis. This study performed a correlation analysis on the matched datasets. The coefficient of determination (R^2) between the actual water yield and the simulated water yield was calculated to assess the consistency between the two datasets. The results are shown in Fig. 5a. Further validation with the SDB 2010 water yield data in 2010 (Fig. 5b) and the RESDC water yield dataset (Fig. 5c) revealed strong consistency with our simulations, indicating that the InVEST model results are reliable. Fig. 5a shows strong agreement at the provincial scale, yet systematic differences remain because the InVEST model is structurally simplified and several definition/scale mismatches exist. The model estimates water yield from precipitation, reference ET, and vegetation/soil parameters without explicitly representing groundwater and surface water interactions, flow routing and regulation, inter-basin transfers, or human withdrawals/returns. By contrast, provincial Water Resources Bulletin statistics typically include baseflow contributions and management effects and are aggregated by administrative units, which do not perfectly match hydrological boundaries - hence larger deviations in arid or heavily regulated regions. Forcings and parameters add uncertainty (biases in precipitation/ET downscaling, PAWC/root depth, and Kc spatialization, the regional Z parameter), and annual averaging can smooth snow/ice melt or extreme events, affecting slopes and intercepts. Although the comparison between the simulated water yield and the SDB/RESDC dataset shows a good positive correlation (Fig. 5b and Fig. 5c), the intercept is positive and the slope is greater than 1, suggesting that the water yield product has a higher baseline water yield (which may include more base flow/human regulation volume or adopt a more humid meteorological environment). At the same time, these datasets differ in their spatial resolution, land cover, soil inputs, and parametric schemes, while scale effects also intrinsically influence the comparison. Land cover classes validation shows the strongest agreement for shrubland and barren, while forest and cropland correlate well but exhibit steeper slopes and negative intercepts, and grassland performs the weakest (Fig. S4). These differences stem from the Budyko simplifications in InVEST and scale mismatches. Croplands are strongly affected by irrigation, runoff regulation, and return flows, raising baselines in published products. Forests reflect orographic precipitation biases, snow/ice melt, and baseflow, making external estimates higher. Grasslands show larger dispersion due to water-stress pulses, grazing effects, and heterogeneous PAWC/Kc. By contrast, shrubland and barren areas have simpler processes and weaker management, resulting in closer precipitation and ET partitioning across products.

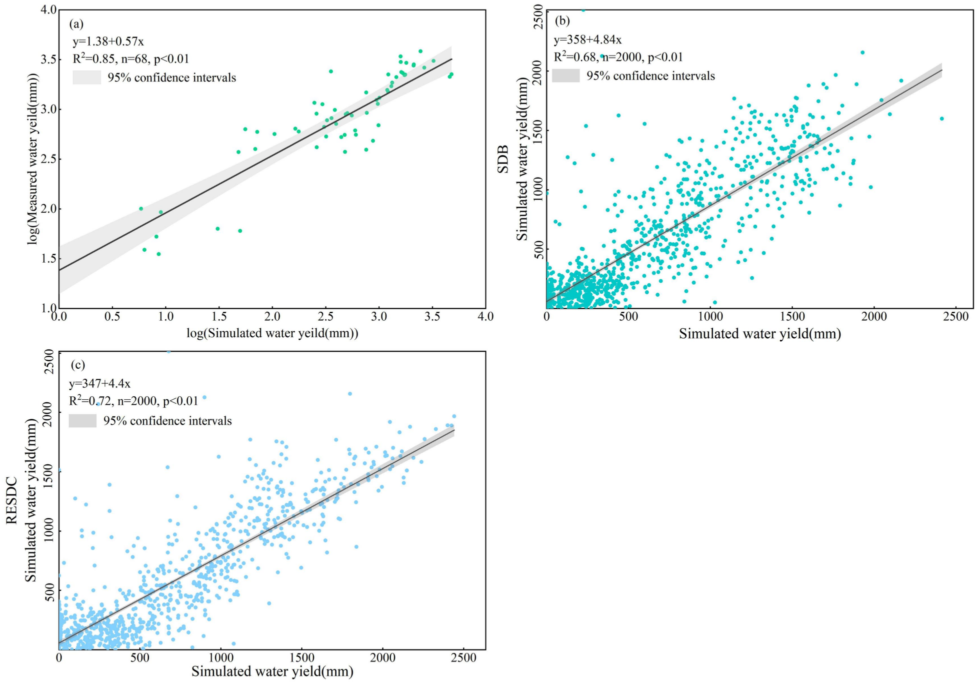

Figure 5: Validation of the water yield in this study, (a) the simulations and measurements of water yield for 34 provinces in 2000 and 2020. (b) the water yield simulated in this study and Science Data Bank (SDB) water yield data in 2010, (c) the water yield simulated in this study and Resource and Environment Science and Data Center (RESDC) water yield data.

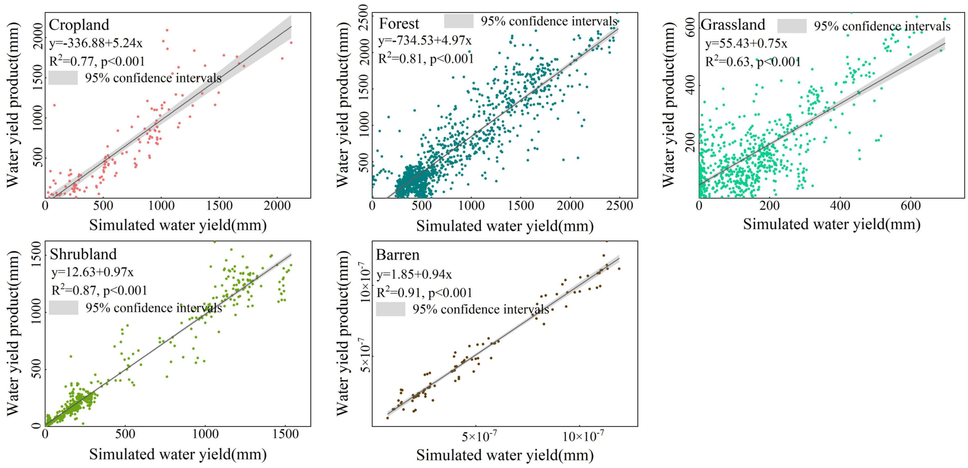

Figure S4: The verification map of the land cover type of water yield simulated by the InVEST model and the published water yield products.

Spawn, S.A., Gibbs, H.K.: Global aboveground and belowground biomass carbon density maps

for the year 2010 [data set], https://doi.org/10.3334/ORNLDAAC/1763, 2020.

Xiao, Q., Hu, D., Xiao, Y.: Assessing changes in soil conservation ecosystem services and causal factors in the Three Gorges Reservoir region of China, J. Clean Prod., 163, S172-S180, https://doi.org/10.1016/j.jclepro.2016.09.012, 2017.

Liu, Y., Zhao, W.W., Zhang, Z.J., Hua, T., Ferreira, C.: The role of nature reserves on conservation effectiveness of ecosystem services in China, J. Environ. Manage., 342, 118228, https://doi.org/10.1016/j.jenvman.2023.118228, 2023.

Gkikas, K., Proestakis1, E., Amiridis, V., Kazadzis, S., Di Tomaso, E.D., Marinou, E., Hatzianastassiou, N., Kok, J. F., Pérez García-Pando, C.: Quantification of the dust optical depth across spatiotemporal scales with the MIDAS global dataset (2003-2017), Atmos. Chem. Phys., 22, 3553-3578, https://doi.org/10.5194/acp-22-3553-2022, 2022.

Comment 9: It is worth noting, however, that only two of the input datasets (Landsat NDVI and GlobeLand30) are at 30 m resolution, while most of the other input data are coarser. Given this, the validity of the final 30 m resolution output may be open to discussion. While it is understandable that fine-resolution input data are difficult to obtain and the use of coarser datasets is often inevitable, this makes it all the more important to acknowledge the limitation in the discussion

Response: Thanks for your detailed consideration. This is indeed a valuable and insightful comment that helps strengthen the methodological rigor and transparency of our manuscript. As the reviewer rightly pointed out, acknowledging and discussing this limitation is essential for the correct interpretation of our high-resolution outputs. We have now added a dedicated discussion in the Limitations section acknowledging that while our final outputs are presented at 30 m resolution, the integration of coarser-resolution datasets introduces uncertainties. We clarify that the 30 m resolution is defined by our finest data layers and advise caution when interpreting fine-scale patterns in heterogeneous areas **(Lines 713-720)**. The main contents are as follows:

The four ecosystem services were assessed using different satellite sources of data, while the ecosystem service maps are presented at 30 m resolution - driven by the highest-resolution data (Landsat NDVI, GlobeLand30, and DEM) - other essential input data (e.g., climate and soil properties) were originally at coarser resolutions. Although these data were resampled to the 30 m resolution, this process inevitably introduces uncertainty (Yan et al., 2025). The fine-resolution output effectively captures spatial patterns defined by the land cover and NDVI. Still, the precision of absolute values in highly heterogeneous areas may be constrained by the inherent information content of the original coarser datasets (Liu et al., 2023).

Yan, J., Wang, S., Feng, J., He, H., Wang, L., Sun, Z., Zheng, C.: New 30-m resolution dataset reveals declining soil erosion with regional increases across Chinese mainland (1990-2022), Remote Sens. Environ., 323, 114681. https://doi.org/10.1016/j.rse.2025.114681, 2025.

Liu, Y., Zhao, W.W., Zhang, Z.J., Hua, T., Ferreira, C.: The role of nature reserves on conservation effectiveness of ecosystem services in China, J. Environ. Manage., 342, 118228, https://doi.org/10.1016/j.jenvman.2023.118228, 2023.

Comment 10: For clarity, it would be helpful if the authors could specify whether this refers to the Pearson correlation coefficient (r) or the coefficient of determination (R^2). Since these represent different statistical measures

Response: Thanks for your detailed consideration. We agree that clarifying the specific statistical metric is crucial for the reader's understanding, as the Pearson correlation coefficient (r) and the coefficient of determination (R^2) represent different measures of agreement. In the manuscript, the “Pearson correlation coefficient” has been corrected to "The coefficient of determination (R^2)". This modification clarifies that we are using the R^2 value, which can better evaluate the consistency and goodness of fit between the actual water yield and the simulated water yield.

Comment 11: For clarity, it would be helpful if the abbreviation DN could be explained here, since it does not seem to be defined earlier in the text.

Response: Thanks for your suggestion. “DN” means “Digital Number,” which refers to the value of a pixel in a remote sensing image. We agree that this term should have been defined upon its first use, and we apologize for this oversight. We have revised the manuscript to include the full term “Digital Number”

Comment 12: Could the authors clarify whether the landcover data used here are derived from GlobeLand30, or whether they are measured data coming in combined with measured NPP?

Response: Thanks for your comment. We have now revised the sentence to specify the "GlobeLand30 dataset" as the source to clarify this point (Lines 172-173).

For a certain land cover from the GlobeLand30 dataset, the error between the measured and the simulated NPP can be expressed by the following formula:

$$E(x)=\sum_{i=1}^{j}(m_i\text{-}n_ix)^2 \quad x \in [l,u], \tag{1}$$

Comment 13: Could the authors specify which version of SoilGrids was used here—V1.0, V2.0, or another digital soil mapping product from ISRIC?

Response: Thanks for your suggestion. The SoilGrids250m v2.0 was used in our study. We have now specified this in the manuscript.

Comment 14: Chang “the” to “The” (Line 189)

Response: Thanks for your comment. This mistake has been corrected. To ensure the accuracy of the language, the manuscript has been thoroughly and comprehensively proofread.

Comment 15: To strengthen the argument, the authors could maybe briefly justify why an empirical equation derived from Australia can also be applied in the context of China, and, for the reader’s benefit, comment on how well it aligns with local conditions (Line 263).

Response: Thanks for your comment, we have now added an explanation of the applicability of the Z-value calculation formula in China in manuscript (Lines 286-291). The detailed content is as follows:

The seasonal parameter Z is an empirical constant that reflects the regional distribution of precipitation and hydrogeological factors. Donohue et al. (2012), through their study of Australia's climatic conditions, found that the seasonal parameter Z can be expressed as Eq. (4). Although this formula originated from Australia, its foundation lies in the globally universal ecological hydrological principle of the water-energy trade-off. Moreover, the extensive climatic gradients spanned by Australia, from humid to arid regions, closely mirror the diverse conditions found across China, thereby providing a robust empirical basis for its application in our study.

Comment 16: Consider avoiding the use of double brackets (Line 290)

Response: Thank you for your suggestion. We have now carefully reviewed the original manuscript and revised the use of all double brackets.

Comment 17: It could further strengthen the paper if, where data permit, the authors provide an accuracy assessment at the national (China) scale, as this may be particularly relevant for readers (Line 325).

Response: Thanks for your detailed consideration. We agree that providing a national-scale accuracy assessment for China would further strengthen the manuscript. We have now added a national-scale accuracy statement for GlobeLand30 over China and cited peer-reviewed sources (Lines 354-357). The detailed content is as follows:

As one of the high-precision global land cover datasets, GlobeLand30 achieves an overall accuracy of over 80.33%, providing detailed ground cover information (Chen et al., 2017). The national-scale independent verification conducted in China (GlobeLand30 2010) indicated that its overall accuracy was 82.39% (Yang et al., 2017).

Chen, W., Liu, W., Geng, Y., Brown, M.T., Gao,C., Wu, R.: Recent progress on energy research: a bibliometric analysis, Renew. Sust. Energ. Rev., 73, 1051-1060, https://doi.org/10.1016/j.rser.2017.02.041, 2017.
Yang, Y., Xiao, P., Feng, X., Li H.: Accuracy assessment of seven global land cover datasets over China. ISPRS J. Photogramm. Remote Sens., 125, 156-173, https://doi.org/10.1016/j.isprsjprs.2017.01.016, 2017.

Comment 18: Considar specifying what the shaded areas stand for (e.g., standard deviation, 95% confidence interval, etc.) (Line 375)

Response: Thanks for your suggestion. The shaded areas in Figs. 2, 3, 4, and 5 represent the 95% confidence intervals. This information was clearly indicated in the figure legend for clarity.

Comment 19: Do you mean Figures 2 and 3 here (Line 532)?

Response: Thank you for noticing this important mistake. We apologize for the confusion caused by the incorrect reference to Fig. 2, Fig. 3, Fig. 4, and Fig. 5. This has now been amended to the correct figure numbers in the manuscript. We have now performed a complete check of all figure citations throughout the manuscript to prevent any similar inconsistencies.