scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Monitoring of carbon-water fluxes at Eurasian meteorological stations using random forest and remote sensing

Mingjuan Xie *et al.*#

Simulating the carbon-water fluxes at more widely distributed meteorological stations based on the sparsely and unevenly distributed eddy covariance flux stations is needed to accurately understand the carbon-water cycle of terrestrial ecosystems. We established a new framework consisting of machine learning, determination coefficient (R^2), Euclidean distance, and remote sensing (RS), to simulate the daily net ecosystem carbon dioxide exchange (NEE) and water flux (WF) of the Eurasian meteorological stations using a random forest model or/and RS. The daily NEE and WF datasets with RS-based information (NEE-RS and WF-RS) for 3774 and 4427 meteorological stations during 2002–2020 were produced, respectively. And the daily NEE and WF datasets without RS-based information (NEE-WRS and WF-WRS) for 4667 and 6763 meteorological stations during 1983–2018 were generated, respectively. For each meteorological station, the carbon-water fluxes meet accuracy requirements and have quasi-observational properties. These four carbon-water flux datasets have great potential to improve the assessments of the ecosystem carbon-water dynamics.

Background & Summary

The eddy-covariance flux stations provide reliable ecosystem-scale measurements of the carbon and energy fluxes at a high temporal resolution[1]. They have become crucial tools to generate observation datasets to verify and benchmark the Earth surface models[2,3]. In particular, it is possible to construct a carbon-water flux simulation model from the station-scale to the regional- or global-scale by means of a large-scale eddy covariance[4] measurement network (e.g. Fluxnet, AmeriFlux and ChinaFlux). However, the existing flux stations are sparsely and unevenly distributed and yield rather discontinuous observation data[1]. This restricts studies on the carbon-water fluxes at a large-scale[3], for example in Eurasia, where a strong spatial heterogeneity is exhibited on complex terrains. The meteorological stations, in contrast, are densely spread around the world with long-term continuous observation data[5], which could have great potential to mine the more extensive carbon-water flux information, particularly combined with machine learning (ML) and remote sensing (RS). This could greatly offset the limitations of the flux station-based observations.

Machine learning is increasingly used to extract the patterns and insights from big geospatial data[6]. Many studies have focused on the comparative evaluation of different ML algorithms and have found the accuracy performance of the same algorithm varies in different research contexts[7–10]. The data-driven ML algorithms are similar to the encapsulated complex empirical algorithms, which demonstrate a high simulation accuracy[3,11]. But the ML algorithms are still influenced by the quality, processing methods, and spatio-temporal representativeness of the data[12–14]. Compared with the process-based land surface or ecosystem models, the ML has a higher carbon-water flux simulation accuracy[1,6]. However, when transferred to other site or regional or spatial (grid) scales, the applicability of both the ML models and process models need to be evaluated due to the distinct spatio-temporal heterogeneity. That is to say, there is no guarantee that these models are applicable to all sites, grids or regions. If this evaluation of the model applicability is not considered, the simulation results will

#A full list of authors and their affiliations appears at the end of the paper.

generate new uncertainties. This issue has become a major problem affecting the simulation accuracy of the carbon-water fluxes at different scales.

In this study, ML with flux observations was used to build carbon-water flux simulation models (random forest model, RFM) to simulate the carbon-water fluxes of the meteorological stations in Eurasia. We proposed a framework to evaluate the applicability of the model transfer and to build a bridge from the flux stations to the meteorological stations. We used this framework to generate four carbon-water flux datasets for the Eurasian meteorological stations. Due to the high precision, these datasets could be regarded as quasi-observational at the site level, which might be used to assess the simulation accuracy of the regional- or global-scale ecosystem carbon-water fluxes based on the ecosystem or land surface or remote sensing or atmospheric inversion models. Our study can, therefore, benefit terrestrial water management and carbon dynamic assessments.

Methods

The RFM was constructed based on the Eurasian flux stations. We built a total of 3,600 RFMs at site scale in accordance with the classification of the flux stations. The simulation accuracy of these RFMs at each flux station in the test set was validated by the spatial cross-validation, thus generating thousands of determination coefficients (R^2) at test stations. According to the third law of geography[15], the factors (variables) similarity between the test station and the training set of the RFM determines the similarity between their fluxes, that is, the R^2 of the RFM at the test station are determined. The similarity between the datasets composed of the same factors could be characterized by the Euclidean distance. Based on the R^2 and Euclidean distance, the R^2 simulation model (RSM) was built by using multiple linear regression (MLR) to evaluate the applicability of RFM on meteorological stations. So that the RFMs can be reasonably transferred to meteorological stations to simulate the carbon-water fluxes. Figure 1 shows the detailed flowchart of the data processing, RFM construction and RFM transfer to the meteorological stations.

Data processing. We selected 156 flux stations in Eurasia from five different landscape types (Fig. 2a), as detailed in the flux station information file[16]. For the flux stations from the National Tibetan Plateau Data Center (NTPDC)[17–133] and European Fluxes Database Cluster (EFDC)[134–137] (http://www.europe-fluxdata.eu/home), the flux data from one hour before (and after) rainfall were excluded. The data collected at 10-min or 30-min intervals were interpolated using the marginal distribution sampling (MDS) method in REddyProc[138]. All final data were converted into daily data. For the flux stations from FLUXNET[139–162], the data were extracted with quality control values ≥ 0.8 for the net ecosystem carbon dioxide exchange (NEE) and latent heat fluxes (LE). The water fluxes (WF) were converted from LE (W/m^2) with a conversion factor of 0.408×10^{-6} [163–165]. For the selected 6856 meteorological stations from the Global Surface Summary of the Day dataset in the National Centers for Environmental Information (https://www.ncei.noaa.gov/metadata/geoportal/rest/metadata/item/gov.noaa.ncdc%3AC00516/html#), the vapour pressure deficit variable was calculated using the air temperature and dew point temperature. The downward shortwave radiation (DSR) of meteorological stations for 2002–2020 and 1983–2018 were extracted from the GLASS dataset[166,167] and the dataset of high-resolution global surface solar radiation[168,169] from the NTPDC, respectively. For the remote sensing (RS) variables (including the fraction of the photosynthetically active radiation extracted from the MCD15A3H data[170], enhanced vegetation index, land surface water index and surface reflectance for the Moderate Resolution Imaging Spectroradiometer bands 1–7 extracted from the MOD09GA data[171]), a linear interpolation was carried out for the missing data with continuous missing days < 8[165,172]. Terrain and soil variables were extracted from the MERIT DEM data[173] and the HWSD data[174], respectively. High quality RS variables, terrain variables and soil variables averaged over a 500-meter spatial extent centered on the station were integrated into the flux stations and meteorological stations (Table S1).

Due to the significant spatial heterogeneity of the earth's surface, the flux stations and meteorological stations were divided into nine categories according to the following four strategies. The first is based on the International Geosphere-Biosphere Programme classification from the MCD12Q1 data[175], including Wetland (i.e. permanent wetlands), Cropland (i.e. croplands and cropland/natural vegetation mosaics), Grassland (i.e. grasslands, savannas, woody savannas) and Forest (i.e. evergreen needleleaf forests, evergreen broadleaf forests, deciduous needleleaf forests, deciduous broadleaf forests, and mixed forests). The second is based on the continents, e.g. Asia and Europe. The third is the arid and non-arid regions classified by the dryland dataset[176]. This dataset identified areas with an aridity index (AI) less than 0.65 as drylands, which were described as arid in this study, and the remaining areas (i.e. $AI \geq 0.65$) were classified as non-arid regions. The fourth comprises entire Eurasia, that is, overall. We used the datasets of nine categories from flux stations as input to RFM, and the detail is presented in the division of flux stations file[177].

RFM construction. In this work, the random forest method[6,178] was used to construct the RFM using the scikit-learn library in Python 3.7.6 (https://pypi.org/project/scikit-learn/1.0.1/). The RFM is a combination model based on independent regression trees[179], of which the predictions were made by averaging the results across all regression trees. The random-search optimizer[180] was applied to identify the optimal hyperparameter settings for the RFM (Table S2). In addition, the most important step in the RFM construction is the k-fold cross-validation (CV)[181]. Suppose the flux station dataset is composed of data from n flux stations and it might generate k sub datasets (D_1, D_2, …, D_k) if it is equally divided into k parts. Here, each sub dataset is a test set, which is composed of data from the $m = n/k$ flux stations, while the remaining sub datasets constitute training sets, which are composed of data from the (n-m) flux stations. Thus, each training set could be used to establish the RFM for the flux simulation; consequently, a total of k models (RFM_1, RFM_2, …, RFM_k) were built. Each RFM is tested and verified at each flux station in the test set. Furthermore, the R^2 is calculated for each case. Hence, the R^2 amount that could

be generated after each RFM training and test is given by $k \times m = k \times n/k = n$. The flux station dataset (n) requires multiple (p) repetitions of the k-fold division to avoid the contingency of the station division. In this way, $k \times p$ RFMs could be constructed and the number of R^2 that could be generated is $n \times p$.

According to the above-mentioned principles, we used the 10-time (split) 10-fold CV (i.e. $p = 10$ and $k = 10$) to set up 100 RFMs for each of the nine categories under two scenarios: with and without RS variables in the model building (i.e. RFM-RS and RFM-WRS), respectively. That is, we constructed 900 RFM-RS models to simulate NEE and WF, respectively. And we also constructed 900 RFM-WRS models to simulate NEE and WF, respectively. Each RFM was validated at each flux station in the corresponding test set, and the R^2 was generated to represent the validation accuracy of the RFM. The R^2 also represents the applicability of these models in the test flux stations. The higher the R^2 of the model on the test flux station, the more suitable the model is for the specific flux station, that is, the more similar the data characteristics of the training set for building the model are to the data characteristics of the flux station according to the third law of geography.

RFM transfer to the meteorological stations. We screened available RFMs for meteorological stations by using the RSM, which was used to evaluate the RFM applicability on the meteorological stations. The framework for the evaluation was designed (Fig. 1) as follows:

① *Euclidean distances of the influencing factors between test flux stations and training sets of RFMs.*

The R^2 of each RFM is determined by the similarity of the influencing factors between the training set and the test set[181]. This could be characterized by the Euclidean distance[182].

For a specific factor affecting the RFM, the Euclidean distance (ds) between a flux station in the test set (test station) and a flux station in the training set (training station) is expressed as:

$$ds(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_t - y_t)^2} = \sqrt{\sum_{(i=1)}^{t}(x_i - y_i)^2}, \; t \le 365 \tag{1}$$

where x represents one of the factors influencing the carbon-water fluxes in the training station (Table S1), y denotes the corresponding influencing factor in the test station, and t is the sample size of the factor. The data of the training station and test station were averaged on the same day (day of the year, DOY) for multiple years, respectively. Then, the two stations could be matched day by day based on DOY to ensure that these have the same daily data sample size.

For the factor j influencing the RFM, the final Euclidean distance (d) is the average of all n-m ds between the test station and each of the n-m training stations, which is the Euclidean distance (d_j) of the factor j between this test station and the training set of the RFM (Eq. 2). In the same way, the Euclidean distances of all influencing factors are produced and denoted as $d_1, \ldots, d_{w\text{-}1}, d_w$. In this way, the R^2 of the RFM tested in the test station from the test set and the Euclidean distances $d_1, \ldots, d_{w\text{-}1}, d_w$ between this test station and the training set for building the RFM constitute a complete data sample (Fig. S3). Similarly, all test stations could generate such samples, which constitute a dataset with a quantity equal to $10 \times n \times 9$. Samples of the same category are integrated into one dataset, that is, nine datasets produced under nine categories, i.e. Dataset 1 in Fig. 1.

$$d(x, y) = \frac{ds_1(x, y) + ds_2(x, y) + \cdots + ds_{(n-m)}(x, y)}{(n - m)} = \frac{\sum_{(i=1)}^{(n-m)} ds_i(x, y)}{(n - m)} \tag{2}$$

where n represents the number of the flux stations of each category and m illustrates the number of the flux stations in each test set.

② *Construction of the RSM.*

Based on Dataset 1, the RSM is constructed using MLR[183] for nine categories under NEE and WF scenarios, of which each one is expressed as:

$$R^2 = a_0 + a_1 d_1 + a_2 d_2 + \cdots + a_{(w-1)} d_{(w-1)} + a_w d_w \tag{3}$$

where $a_0, a_1, a_2, \ldots, a_{w\text{-}1}, a_w$ are regression coefficients and $d_1, d_2, \ldots, d_{w\text{-}1}, d_w$ indicate the Euclidean distances of the factors influencing the carbon-water fluxes between the test station and the training set.

③ *Euclidean distances of the influencing factors between meteorological stations and training sets of RFMs.*

The same processes of ① are applied to the meteorological stations so as to calculate the Euclidean distance for the influencing factors between each RFM training set and meteorological station for each category under two scenarios, which yields a large dataset, i.e. Dataset 2 in Fig. 1.

④ *Prediction of the R^2 of the RFM transfer to the meteorological stations.*

Before a RFM is transferred to a specific meteorological station, the RSM could predict the R^2 value on the station using Dataset 2 in Fig. 1. Only if predicted $R^2 \ge 0.5$, its corresponding RFM might be transferred to the corresponding meteorological stations. Otherwise, the RFM was assumed to be not applicable to the meteorological station. If there was more than one RFM applicable to a meteorological station, the RFM corresponding

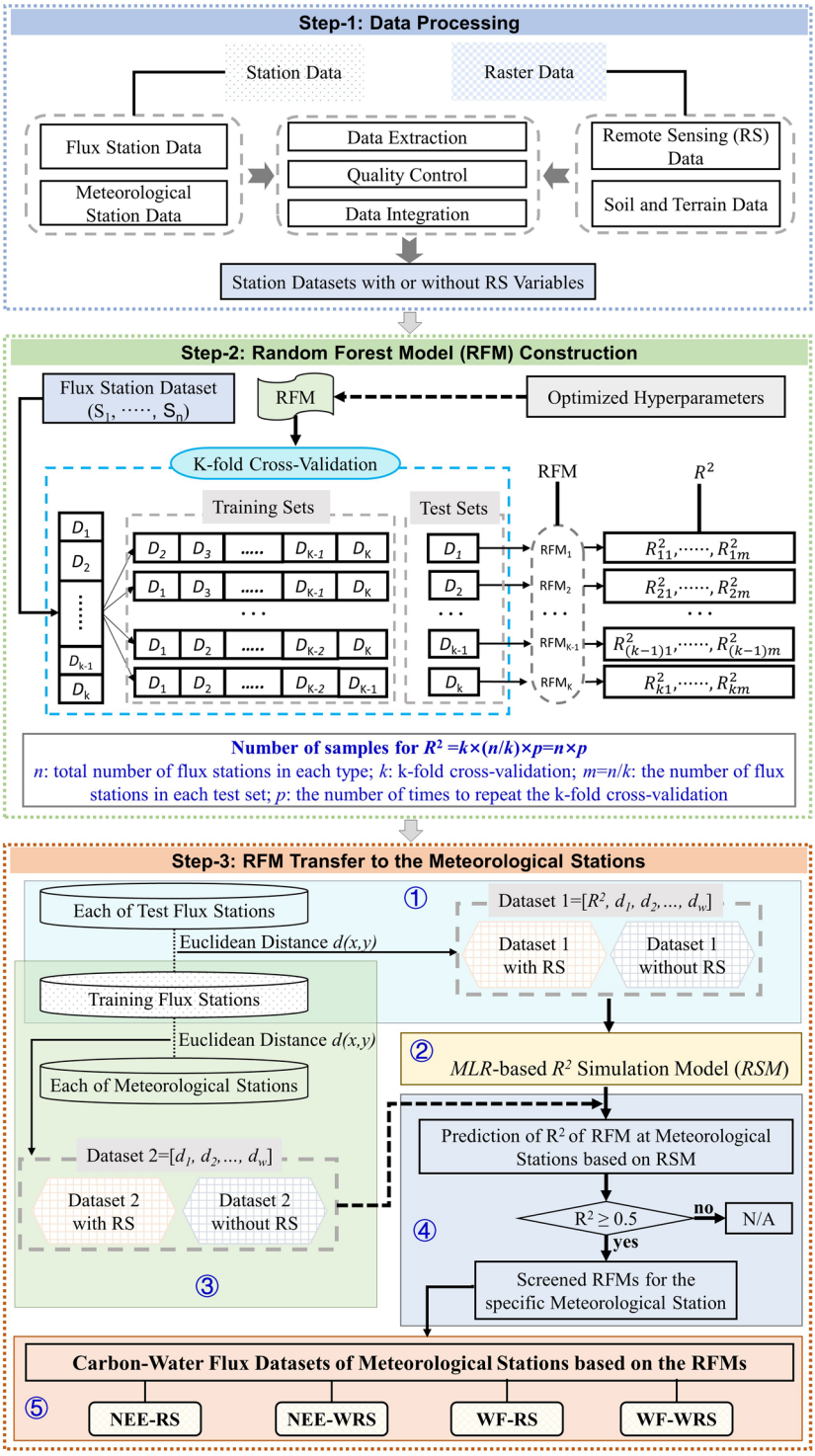

Fig. 1 Research framework. R^2, determination coefficient; MLR, multiple linear regression. N/A (not applicable) indicates that the RFM could not be transferred to the specific meteorological stations. NEE-RS: net ecosystem carbon dioxide exchange (NEE) based on the RFM with remote sensing (RS); WF-RS: water flux (WF) based on the RFM with RS; these explain the fact that the RS variables were used in the RFM construction. NEE-WRS: NEE based on the RFM without RS; WF-WRS: WF based on the RFM without RS; these demonstrate that the RS variables were not applied in the RFM construction. RS variables include the fraction of the photosynthetically active radiation, enhanced vegetation index, land surface water index and surface reflectance for the Moderate Resolution Imaging Spectroradiometer bands 1–7.

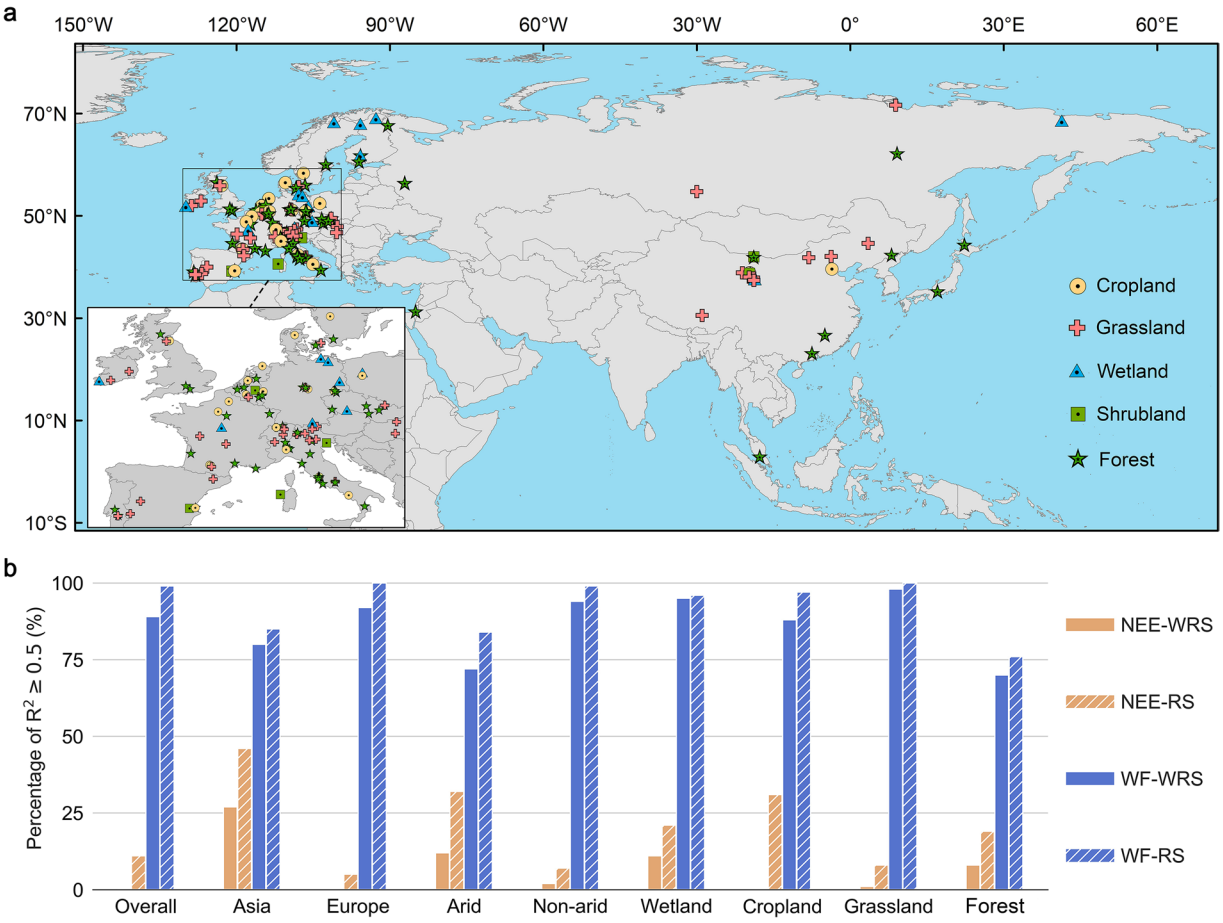

Fig. 2 Study area and the accuracy of carbon-water flux simulation models (random forest model, RFM). (**a**) Distribution of the 156 Eurasian flux stations covering five main landscape types. (**b**) The accuracy assessments of the RFM based on the Eurasian flux stations in the framework of the 10-time 10-fold cross-validation. The figure shows the percentage of the RFMs with $R^2 \geq 0.5$ tested in the test sets for nine categories.

to the maximum predicted R^2 was screened as the model that could be linked to the meteorological station. Not all meteorological stations did have a RFM which is applicable to the meteorological station.

⑤ *Carbon-water flux simulation of the meteorological stations.*

For the meteorological stations in Eurasia that could be linked with an applicable RFM, the corresponding RFM can be used to simulate the daily carbon-water fluxes and to build high-precision carbon-water flux datasets of the Eurasian meteorological stations to analyze the carbon-water dynamics. These datasets[184] consist of two essential building blocks: (i) datasets related to remote sensing, including the net ecosystem carbon dioxide exchange (NEE-RS) and water fluxes (WF-RS) simulated by the RFM-RS; (ii) the net ecosystem carbon dioxide exchange (NEE-WRS) and water fluxes (WF-WRS) simulated by the RFM-WRS.

Data Records

Our carbon-water flux datasets[184] are available at figshare (https://doi.org/10.6084/m9.figshare.21347721.v3). The data record contains two daily carbon dioxide flux datasets (NEE-RS and NEE-WRS) and two daily water flux datasets (WF-RS and WF-WRS) of the Eurasian meteorological stations. The coverage period of the NEE-RS and WF-RS has been recorded from 2002 to 2020 and the one of NEE-WRS and WF-WRS from 1983 to 2018. The data of each meteorological station was deposited separately in the CSV file format under the dataset folders. The file name indicates the identification number of the meteorological station in the meteorological station information file[185] (https://doi.org/10.6084/m9.figshare.23695920.v2). The list of flux stations used in this study and the details of flux station division used for the RFM construction are shown in the flux station information file[16] (https://doi.org/10.6084/m9.figshare.23899701.v1) and the division of flux stations file[177] (https://doi.org/10.6084/m9.figshare.23695980.v2) stored at figshare, respectively. In addition, the details of the RSM construction are presented in the RSMs information file[186] (https://doi.org/10.6084/m9.figshare.23899785.v1) deposited at figshare. The file specific fields are as follows:

Carbon-water flux datasets file (.csv).

(1) id: Identification of the meteorological station.
(2) lon: Longitude of the meteorological station.
(3) lat: Latitude of the meteorological station.
(4) year: Year of the data record.
(5) month: Month of the data record.
(6) day: Day of the data record.
(7) doy: Day of the year.
(8) NEE: Net ecosystem carbon dioxide exchange (g C m^{-2} d^{-1}).
(9) WF: Water fluxes (mm d^{-1}).

Meteorological station information file (.xlsx).

(1) Identification of meteorological station.
(2) Station name: Name of the meteorological station.
(3) Longitude: Longitude of the meteorological station.
(4) Latitude: Latitude of the meteorological station.
(5) Elevation: Elevation of the meteorological station (m).
(6) Continent: Continent of the meteorological station.
(7) Drought situation: Drought situation of the meteorological station.
(8) Landscape: Landscape of the meteorological station.
(9) Data source: Data source of the meteorological station.
(10) Classification of simulated NEE-RS: Accuracy classification of NEE-RS for meteorological stations (1: low quality, $R^2 < 0.5$; 2: moderate quality, $0.5 \leq R^2 < 0.7$; 3: high quality, $R^2 \geq 0.7$).
(11) Classification of simulated NEE-WRS: Accuracy classification of NEE-WRS for meteorological stations (1: low quality, $R^2 < 0.5$; 2: moderate quality, $0.5 \leq R^2 < 0.7$; 3: high quality, $R^2 \geq 0.7$).
(12) Classification of simulated WF-RS: Accuracy classification of WF-RS for meteorological stations (1: low quality, $R^2 < 0.5$; 2: moderate quality, $0.5 \leq R^2 < 0.7$; 3: high quality, $R^2 \geq 0.7$).
(13) Classification of simulated WF-WRS: Accuracy classification of WF-WRS for meteorological stations (1: low quality, $R^2 < 0.5$; 2: moderate quality, $0.5 \leq R^2 < 0.7$; 3: high quality, $R^2 \geq 0.7$).

Flux station information file (.xlsx).

(1) Identification of flux stations.
(2) Flux station: Name of the flux station.
(3) Longitude: Longitude of the flux station.
(4) Latitude: Latitude of the flux station.
(5) Elevation: Elevation of the flux station (m).
(6) Continent: Continent of the flux station.
(7) Drought situation: Drought situation of the flux station.
(8) Landscape: Landscape of the flux station.
(9) Study period: Study period of the flux station used in this study.
(10) Data source: Data source of the flux station.

Division of flux stations file (.xlsx).

(1) Categories: Category of the flux station.
(2) Number of flux station: Number of flux stations under each category.
(3) Split: Identification of divisions for 10-fold cross-validation on flux stations.
(4) Fold: Identification of folds for cross-validation on flux stations.
(5) Identification of flux station: List of identifications for flux stations under each fold.

RSMs information file (.xlsx).

(1) Models: Name of the RSM.
(2) Categories: Category of the RSM.
(3) N: Number of samples used by the RSM.
(4) R^2_{rsm}: Determination coefficient of the RSM.
(5) Adj. R^2_{rsm}: Adjusted determination coefficient of the RSM.
(6) F-statistic: F-statistic of the RSM.
(7) P value: Significance probability of the RSM.
(8) RSMs: Equation of the RSM.

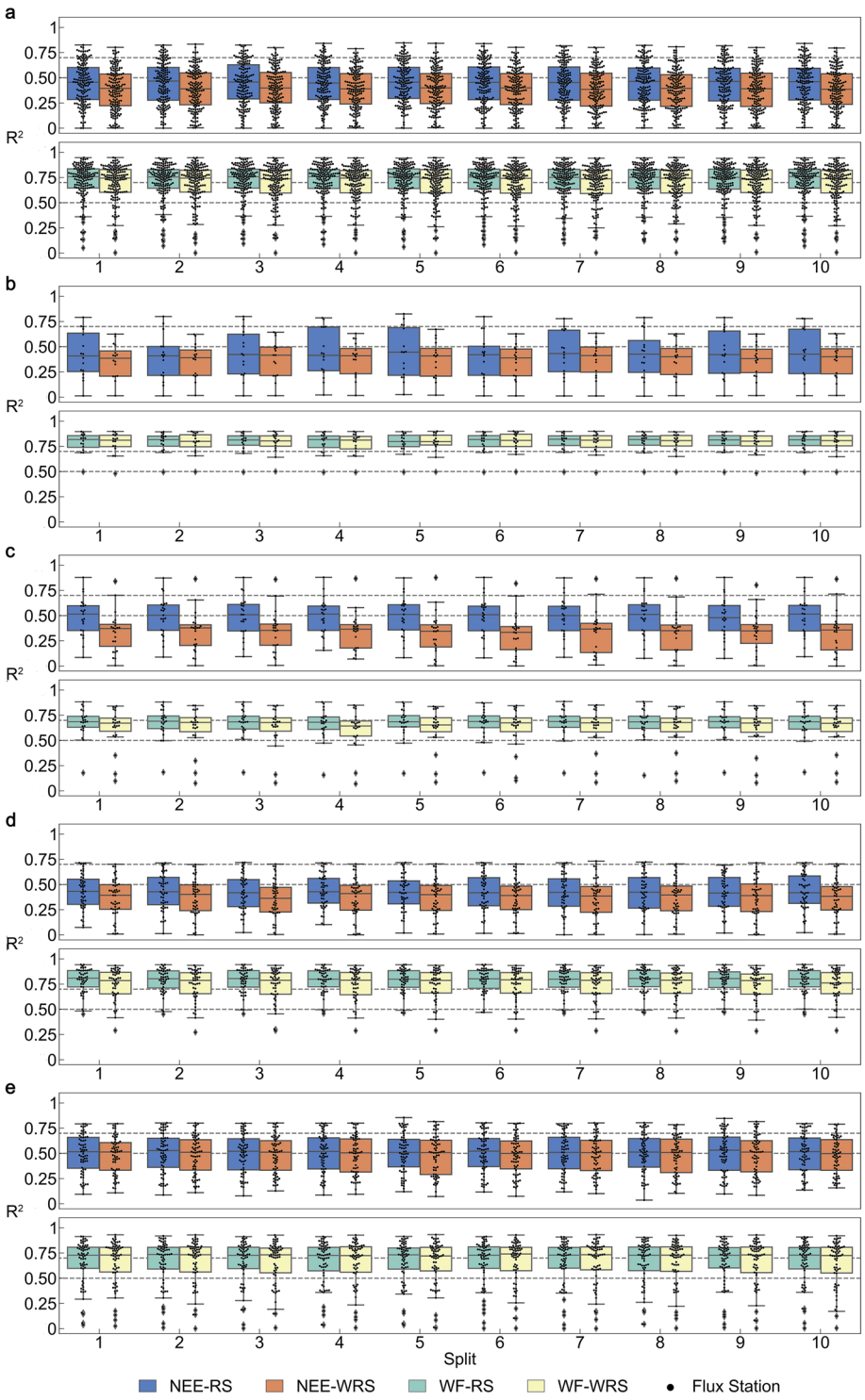

Fig. 3 The accuracy performance of the carbon-water flux simulation models (random forest model, RFM) at test flux stations. The NEE (net ecosystem carbon dioxide exchange) and WF (water flux) R^2-based accuracy performance of the RFM of each split of the 10-time 10-fold cross-validation for (**a**) Overall with 156 stations, (**b**) Wetland with 16 stations, (**c**) Cropland with 23 stations, (**d**) Grassland with 47 stations and (**e**) Forest with 64 stations. The box plots show the R^2 distribution of each flux station of the test set for different categories, in which the whiskers indicate the 1.5 times' interquartile range.

Technical Validation

Model validation. The R^2 and RMSE (root mean square error) were used to evaluate the performance of the RFM to simulate the NEE-RS (NEE simulated by RFM-RS), WF-RS (WF simulated by RFM-RS), NEE-WRS (NEE simulated by RFM-WRS) and WF-WRS (WF simulated by RFM-WRS)[184]. The model's simulation accuracy

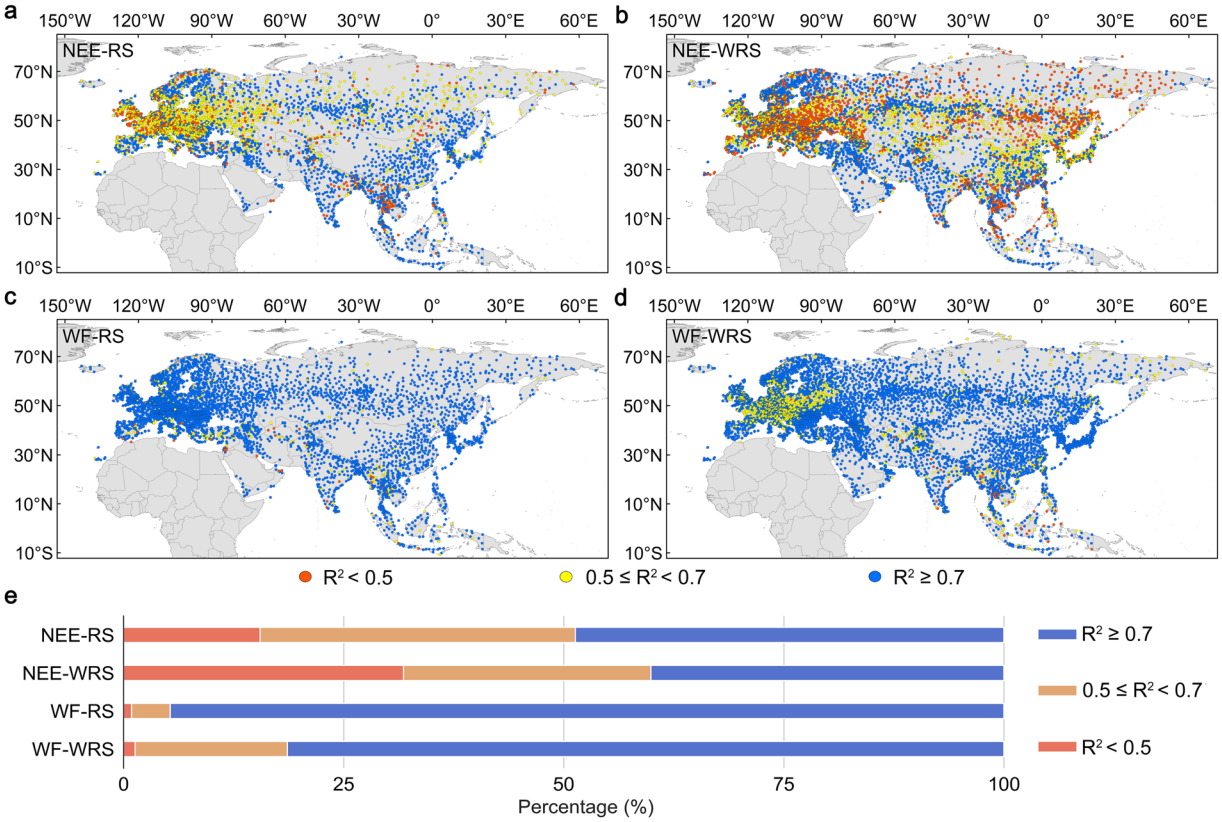

Fig. 4 The distribution of the R^2 predicted by the R^2 simulation model (RSM) at the meteorological stations. Spatial distribution of the R^2 at (**a**) 4466 meteorological stations under the scenario of NEE-RS, (**b**) 6849 meteorological stations under the scenario of NEE-WRS, (**c**) 4466 meteorological stations under the scenario of WF-RS and (**d**) 6849 meteorological stations under the scenario of WF-WRS, respectively. (**e**) The percentage distribution of $R^2 < 0.5$, $0.5 \leq R^2 < 0.7$ and $R^2 \geq 0.7$ of the meteorological stations in the different scenarios.

for the WF was much higher than for NEE under each category and the performance of the RFM-RS was also better than the RFM-WRS (Fig. 2b, Table S3). For the WF simulation of the RFM-RS and RFM-WRS under each category, the percentage of the models with $R^2 \geq 0.5$ in the test sets was larger than 70%, while for the NEE simulation, it was lower than 50% (Fig. 2b). For the category 'overall', the RFM generally indicated a high simulation accuracy (Fig. 2b, Table S3). The simulation accuracy of the RFM was generally higher in Asia and the arid regions than in Europe and the non-arid regions. For the cropland and forest, the RFMs demonstrated the highest simulation accuracy under the scenarios of NEE-RS and NEE-WRS; for the grassland and wetland, the RFMs demonstrated the highest simulation accuracy under the WF-RS and WF-WRS scenarios.

The box plots (Fig. 3, Fig. S1) present the simulation performance of the RFM for NEE and WF in 10-time 10-fold CVs, with each box representing the R^2 distribution for the test flux stations in each split (time). The simulation accuracy of the same RFM for different test flux stations varied widely, indicating that the RFM cannot be applied to all flux stations and that not all stations could match at least one available RFM model. The maximum R^2 distribution for each flux station was observed in a 10-time 10-fold CV (Table S4). The proportion of the flux stations with $R^2 \geq 0.5$ of the RFM test measured 60.9%, 46.2%, 89.7% and 88.5% under the NEE-RS, NEE-WRS, WF-RS and WF-WRS scenarios, respectively.

Accuracy pre-assessment of the carbon-water flux simulation at the meteorological stations. Using the MLR model in which the Euclidean distances were independent variables and R^2 concerned a dependent variable, the RSMs were constructed for different categories under different scenarios, as shown in the RFMs information file[186]. By using the RSM to simulate the R^2 of the RFM at the test flux stations, the overall accuracies (of the RSM) for a correct classification of R^2 under nine categories amounted to 80.1%, 84.0%, 91.0% and 89.1% for the NEE-RS, NEE-WRS, WF-RS and WF-WRS, respectively (Fig. S2). This might prove that the RSM is reliable and could be utilized to predict the accuracy of the RFMs applied to the meteorological stations.

Finally, the RFMs were transferred to all meteorological stations in Eurasia and the R^2 was predicted for every meteorological station for each category under two scenarios (Fig. 4a–d, Table S5). In this study, the criteria for screening the RFM imply that the RFM corresponding to the highest predicted R^2 of a given meteorological station and its $R^2 \geq 0.5$ was screened as the simulation model of the carbon-water fluxes for the meteorological station. The percentages of the meteorological stations in Eurasia were 84.5%, 68.2%, 99.1% and 98.7% for the NEE-RS, NEE-WRS, WF-RS and WF-WRS, respectively, in which the RFMs met the above-mentioned criteria (Fig. 4e). The RFMs have much higher applicable percentages and seem more accurate for the WF simulation

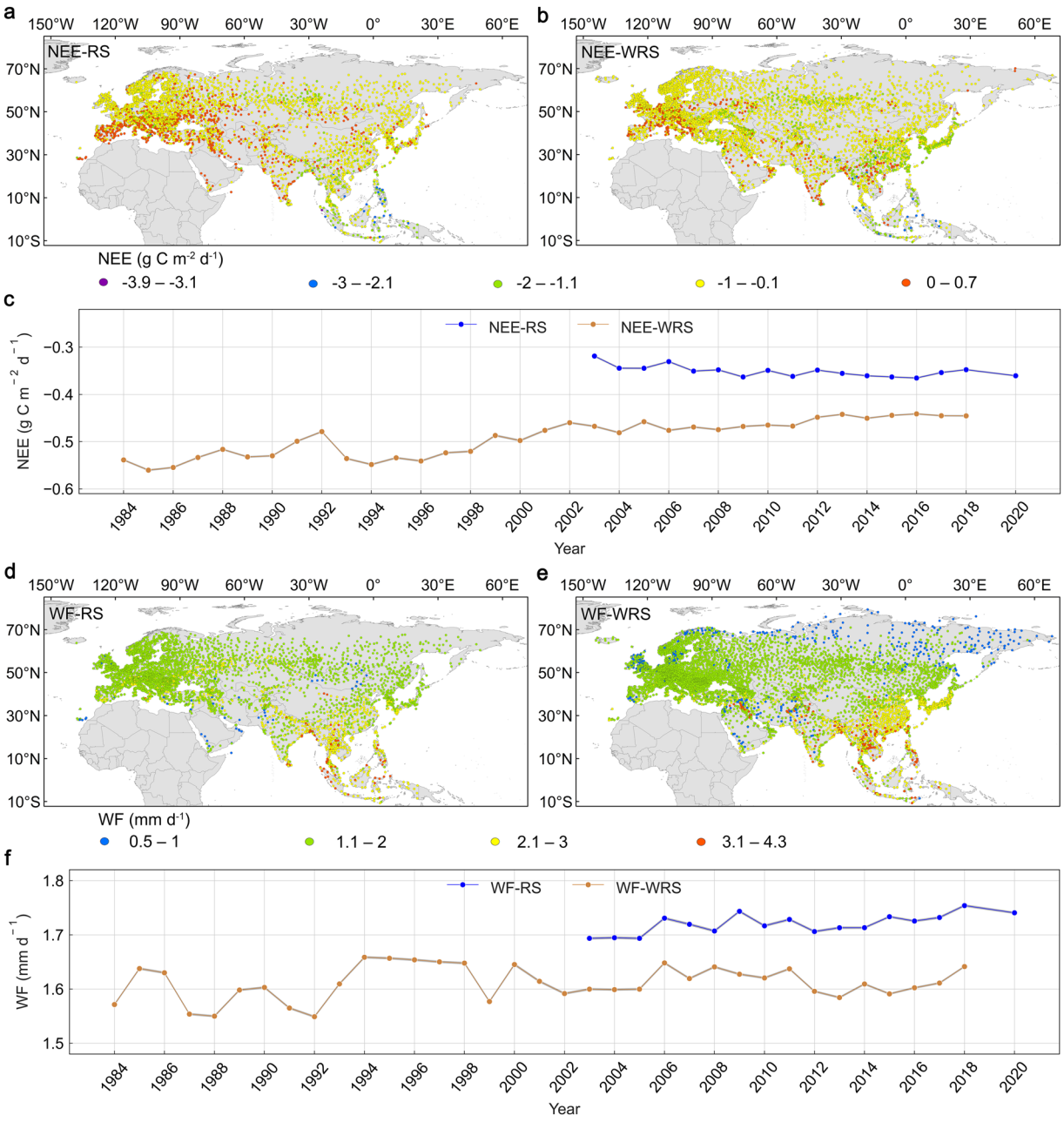

Fig. 5 Spatio-temporal variations of the carbon-water fluxes at the Eurasian meteorological stations. Spatial distribution of the mean daily values during the period March-November of (**a**) NEE-RS from 2003 to 2020 for 3436 meteorological stations and (**b**) the NEE-WRS from 1984 to 2018 for 4352 stations. (**c**) The annual temporal variation of the mean daily NEE (net ecosystem carbon dioxide exchange) values for the meteorological stations and the corresponding 95% confidence interval shown as a shaded band. Spatial distribution of the mean daily values during the period March-November of (**d**) WF-RS from 2003 to 2020 for 3990 stations and (**e**) WF-WRS from 1984 to 2018 for 6302 stations. (**f**) The annual temporal variation of the mean daily WF (water flux) values for the meteorological stations and the corresponding 95% confidence interval shown as a shaded band.

than the NEE simulation at the meteorological stations. The RFM models of the forest and grassland categories were highly applicable and more accurate regarding the NEE and WF simulation of the meteorological stations than for cropland or wetland (Table S5).

The input data for the RFM were primarily derived from the meteorological stations' observations and remote sensing data. Moreover, machine learning models (such as RFM) have the advantage and the predictive ability in the non-linear relation fitting and have been proven in the application research of relevant geoscience[6], which is generally superior to linear regression, ecosystem process models, remote sensing inversion models, etc. Therefore, the carbon-water flux datasets of the meteorological stations generated in this study demonstrate

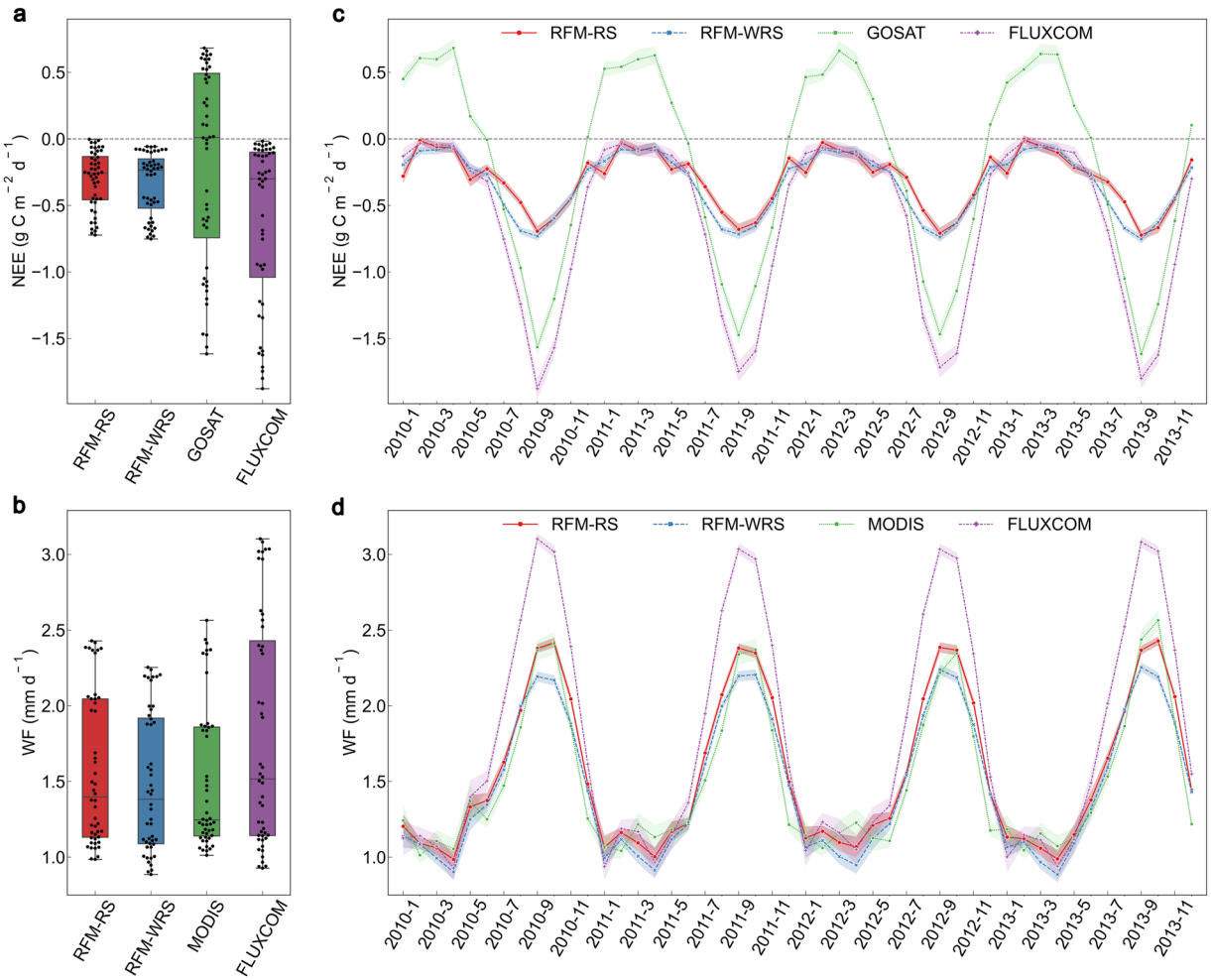

Fig. 6 Comparison of the monthly NEE (net ecosystem carbon dioxide exchange) and WF (water flux) in this study with those from GOSAT, MODIS and FLUXCOM during the period 2010–2013. The box plots of the monthly values (black dots) for (**a**) NEE and (**b**) WF, respectively, in which the whiskers indicate the 1.5 times' interquartile range. The monthly changes in (**c**) NEE and (**d**) WF and the corresponding 95% confidence interval shown as a coloured line and shaded band, respectively. RFM-RS: NEE or WF based on the RFM with remote sensing (RS), representing the fact that the RS variables were used in the RFM construction. RFM-WRS: NEE or WF based on the RFM without RS, illustrating that the RS variables were not used in the RFM construction. The RS variables include a fraction of the photosynthetically active radiation, enhanced vegetation index, land surface water index and surface reflectance for the Moderate Resolution Imaging Spectroradiometer bands 1–7. GOSAT, the GOSAT L4A data; MODIS, the MOD16A2 Version 6 data; FLUXCOM, an initiative to upscale the biosphere-atmosphere fluxes from the FLUXNET sites to the continental and global scales.

a relatively high accuracy and constitute the attribute of quasi-observation, which might be considered as a quasi-observational dataset. They could be applied as benchmark data to verify the simulation results produced by the process-based models or remote sensing inversion models related to the carbon-water fluxes, which overcomes the challenge of insufficient observational data on the carbon-water fluxes[165,187].

Spatio-temporal patterns of the Eurasian NEE and WF. We have further investigated the spatio-temporal distribution of the mean daily values of NEE and WF simulated by the RFM-RS during the period March-November from 2003 to 2020 and simulated by the RFM-WRS from 1984 to 2018. The meteorological stations with at least 30 data volumes of each spring, summer and autumn were used for statistical analysis. The mean daily values of the NEE-RS, NEE-WRS, WF-RS and WF-WRS at the meteorological stations are −3.9~0.7 g C m^{-2} d^{-1}, −2.6~0.4 g C m^{-2} d^{-1}, 0.8~3.8 mm d^{-1} and 0.5~4.3 mm d^{-1}, respectively (Fig. 5a,b,d,e). The spatial distribution of these mean daily NEE fluxes reveals that the ecosystem carbon dioxide loss had increased in Eurasia during 2003–2020, with 457 carbon dioxide loss stations during this period, which means 178 more than from 1984 to 2018 (Fig. 5a,b). The daily average NEE (generally presented as net carbon dioxide uptake) has shown an increasing trend from 1984 to 2002, while a slightly decreasing tendency from 2003 to 2020, with slight fluctuations during these two periods (Fig. 5c). The temporal variation of the WF has demonstrated a rising trend with a

distinct fluctuation from 1984 to 2020 (Fig. 5f). The differences between RS and WRS products might be caused by the differences in the input DSR dataset, the RFMs, and the number of meteorological stations (Fig. 5c,f).

Comparison of the NEE and WF with other carbon-water flux products. We further compared the NEE and WF datasets with those from FLUXCOM[1,3], GOSAT L4A (https://data2.gosat.nies.go.jp/) and MODIS (MOD16A2 Version 6)[188] (Fig. 6, Table S6). The WF from the other products were converted from the LE. Our NEE and WF datasets, as well as the fluxes from the other products, were converted into a monthly scale. The months with the same number of stations for each product were selected for a comparison. All data show a similar seasonal variation, with high carbon-water fluxes in summer and low during winter (Fig. 6c,d). The distributions of the carbon-water fluxes from FLUXCOM and NEE from GOSAT are relatively discrete (Fig. 6a,b). The multi-year monthly averages of the NEE (NEE-RS = $-0.31\,g\,C\,m^{-2}\,d^{-1}$ and NEE-WRS = $-0.34\,g\,C\,m^{-2}\,d^{-1}$) and WF (WF-RS= $+1.57\,mm\,d^{-1}$ and WF-WRS= $+1.48\,mm\,d^{-1}$) simulated herein are less than those from FLUXCOM (NEE = $-0.61\,g\,C\,m^{-2}\,d^{-1}$, and WF= $+1.79\,mm\,d^{-1}$), whereas the averages are larger than those from GOSAT (NEE = $-0.20\,g\,C\,m^{-2}\,d^{-1}$) and MODIS (WF= $+1.51\,mm\,d^{-1}$). The WF from MODIS were almost consistent with our results (Fig. 6d). Because of the fact that the carbon-water flux datasets of the meteorological stations (generated by the RFM in this study) could be considered as "quasi-observational data", the Eurasian carbon-water fluxes from FLUXCOM may be overestimated, while the NEE from GOSAT could rather be underestimated (Fig. 6c,d).

Code availability

The code[189] to generate the carbon-water flux datasets is available at figshare (https://doi.org/10.6084/m9.figshare.21510183.v2).

Received: 9 November 2022; Accepted: 14 August 2023;
Published online: 07 September 2023

References

1. Jung, M. *et al*. Scaling carbon fluxes from eddy covariance sites to globe: Synthesis and evaluation of the FLUXCOM approach. *Biogeosciences* **17**, 1343–1365 (2020).
2. Ciais, P. *et al*. Five decades of northern land carbon uptake revealed by the interhemispheric CO2 gradient. *Nature* **568**, 221–225 (2019).
3. Jung, M. *et al*. The FLUXCOM ensemble of global land-atmosphere energy fluxes. *Sci. Data* **6**, 1–14 (2019).
4. Jung, M. *et al*. Compensatory water effects link yearly global land CO2 sink changes to temperature. *Nature* **541**, 516–520 (2017).
5. Wang, R. *et al*. Recent increase in the observation-derived land evapotranspiration due to global warming. *Environ. Res. Lett.* **17**, 024020 (2022).
6. Reichstein, M. *et al*. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).
7. Li, X. *et al*. Intercomparison of six upscaling evapotranspiration methods: From site to the satellite pixel. *J. Geophys. Res.-Atmos.* **123**, 6777–6803 (2018).
8. Shi, H. *et al*. Evaluation of water flux predictive models developed using eddy-covariance observations and machine learning: a meta-analysis. *Hydrol. Earth Syst. Sci.* **26**, 4603–4618 (2022).
9. Shi, H. *et al*. Variability and uncertainty in flux-site-scale net ecosystem exchange simulations based on machine learning and remote sensing: a systematic evaluation. *Biogeosciences* **19**, 3739–3756 (2022).
10. Xu, T. *et al*. Evaluating different machine learning methods for upscaling evapotranspiration from flux towers to the regional scale. *J. Geophys. Res.-Atmos.* **123**, 8674–8690 (2018).
11. Bzdok, D., Nichols, T. E. & Smith, S. M. Towards algorithmic analytics for large-scale datasets. *Nat. Mach. Intell.* **1**, 296–306 (2019).
12. Fatima, M. & Pasha, M. Survey of machine learning algorithms for disease diagnostic. *J. Intell. Learn. Syst. Appl.* **9**, 1 (2017).
13. Ghahramani, Z. Probabilistic machine learning and artificial intelligence. *Nature* **521**, 452–459 (2015).
14. Jordan, M. I. & Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science* **349**, 255–260 (2015).
15. Zhu, A. X., Lu, G., Liu, J., Qin, C. Z. & Zhou, C. Spatial prediction based on Third Law of Geography. *Ann. GIS* **24**, 225–240 (2018).
16. Xie, M. Flux station information. *figshare* https://doi.org/10.6084/m9.figshare.23899701.v1 (2023).
17. Che, T., Xu, Z., Ren, Z., Tan, J. & Zhang, Y. Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (automatic weather station of Zhangye wetland station, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270677 (2020).
18. Liu, S. *et al*. Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (automatic weather station of Huazhaizi desert steppe station, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270680 (2020).
19. Liu, S. *et al*. Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (an observation system of Meteorological elements gradient of Sidaoqiao Superstation, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270698 (2020).
20. Liu, S. *et al*. Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (automatic weather station of mixed forest station, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270681 (2020).
21. Liu, S. *et al*. Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of mixed forest station, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270689 (2020).
22. Liu, S. *et al*. Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of Huazhaizi station, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270684 (2020).
23. Liu, S. *et al*. Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of Sidaoqiao superstation, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270685 (2020).
24. Liu, S. *et al*. Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of Zhangye wetland station, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270690 (2020).
25. Liu, S. *et al*. Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (automatic weather station of mixed forest station, 2020). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.271403 (2021).

26. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (an observation system of Meteorological elements gradient of Sidaoqiao Superstation, 2020). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.271542 (2021).
27. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of Sidaoqiao superstation, 2020). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Geogra.tpdc.271440 (2021).
28. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (automatic weather station of Zhangye wetland station, 2020). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Geogra.tpdc.271437 (2021).
29. Liu, S., Che, T., Xu, Z., Ren, Z. & Zhang, Y. Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of Zhangye wetland station, 2020). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Geogra.tpdc.271441 (2021).
30. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of mixed forest station, 2020). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.271399 (2021).
31. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (an observation system of Meteorological elements gradient of A'rou Superstation, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270694 (2020).
32. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of A'rou Superstation, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270691 (2020).
33. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (automatic weather station of Dashalong station, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270757 (2020).
34. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of Dashalong station, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270687 (2020).
35. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (automatic weather station of Jingyangling station, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270682 (2020).
36. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of Jingyangling station, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270688 (2020).
37. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (automatic weather station of Yakou station, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270678 (2020).
38. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of Yakou station, 2019). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270686 (2020).
39. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (an observation system of Meteorological elements gradient of A'rou Superstation, 2020). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.271415 (2021).
40. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of A'rou Superstation, 2020). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.271406 (2021).
41. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (automatic weather station of Jingyangling station, 2020). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.271400 (2021).
42. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of Jingyangling station, 2020). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Geogra.tpdc.271409 (2021).
43. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (automatic weather station of Yakou station, 2020). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.271398 (2021).
44. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of Yakou station, 2020). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Geogra.tpdc.271408 (2021).
45. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (automatic weather station of Dashalong station, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270779 (2019).
46. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (an automatic weather station of Sidaoqiao mixed forest station, 2013). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.184.2014.db (2016).
47. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (an automatic weather station of Sidaoqiao mixed forest station, 2014). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.261.2015.db (2016).
48. Liu, S. *et al.* HiWATER: Dataset of Hydrometeorological observation network (an automatic weather station of Sidaoqiao mixed forest station, 2015). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.318.2016.db (2016).
49. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of mixed forest station, 2013). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.197.2014.db (2016).
50. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of mixed forest station, 2014). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.241.2015.db (2016).
51. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of mixed forest station, 2015). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.301.2016.db (2016).
52. Liu, S. *et al.* HiWATER: Dataset of hydro-meteorological observation network (automatic weather station of Huazhaizi Desert Steppe Station, 2014). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.257.2015.db (2016).
53. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (automatic weather station of Huazhaizi desert steppe station, 2015). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.325.2016.db (2016).
54. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Huazhaizi desert Station, 2014). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.240.2015.db (2016).
55. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Huazhaizi desert station, 2015). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.305.2016.db (2016).
56. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (an observation system of meteorological elements gradient of Sidaoqiao superstation, 2013). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.187.2014.db (2016).
57. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (an observation system of Meteorological elements gradient of Sidaoqiao Superstation, 2014). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.264.2015.db (2016).
58. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (an observation system of meteorological elements gradient of Sidaoqiao Superstation, 2015). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.321.2016.db (2016).

59. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Sidaoqiao superstation, 2013). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.200.2014.db (2016).
60. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Sidaoqiao superstation, 2014). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.245.2015.db (2016).
61. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Sidaoqiao superstation, 2015). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.304.2016.db (2016).
62. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (automatic weather station of Zhangye wetland station, 2013). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.192.2014.db (2016).
63. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (automatic weather station of Zhangye wetland station, 2014). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.265.2015.db (2016).
64. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (automatic weather station of Zhangye wetland station, 2015). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.327.2016.db (2016).
65. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Zhangye wetland station, 2013). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.206.2014.db (2016).
66. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Zhangye wetland Station, 2015). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.307.2016.db (2016).
67. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (an automatic weather station of Sidaoqiao mixed forest station, 2016). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.461.2017.db (2017).
68. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of mixed forest station, 2016). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.450.2017.db (2017).
69. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (automatic weather station of Huazhaizi desert steppe station, 2016). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.459.2017.db (2017).
70. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Huazhaizi desert station, 2016). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.448.2017.db (2017).
71. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (an observation system of meteorological elements gradient of Sidaoqiao Superstation, 2016). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.463.2017.db (2017).
72. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Sidaoqiao superstation, 2016). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.451.2017.db (2017).
73. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (automatic weather station of Zhangye wetland station, 2016). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.465.2017.db (2017).
74. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Zhangye wetland Station, 2016). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.453.2017.db (2017).
75. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (an automatic weather station of Sidaoqiao mixed forest station, 2017). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.21.2018.db (2018).
76. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of mixed forest station, 2017). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.10.2018.db (2018).
77. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (automatic weather station of Huazhaizi desert steppe station, 2017). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.19.2018.db (2018).
78. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Huazhaizi desert station, 2017). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.8.2018.db (2018).
79. Liu, S. *et al.* HiWATER: Dataset of Hydrometeorological observation network (an observation system of Meteorological elements gradient of Sidaoqiao Superstation, 2017). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.23.2018.db (2018).
80. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Sidaoqiao superstation, 2017). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.11.2018.db (2018).
81. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (automatic weather station of Zhangye wetland station, 2017). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.25.2018.db (2018).
82. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Zhangye wetland Station, 2017). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.13.2018.db (2018).
83. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (automatic weather station of mixed forest station, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270771 (2019).
84. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of mixed forest station, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270785 (2019).
85. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of the Heihe River Basin integrated observatory network (automatic weather station of Huazhaizi desert steppe station, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270773 (2019).
86. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of Huazhaizi station, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270787 (2019).
87. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of the Heihe River Basin integrated observatory network (automatic weather station of Jingyangling station, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270770 (2019).
88. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (an observation system of meteorological elements gradient of Sidaoqiao superstation, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270780 (2019).
89. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of the Heihe River Basin integrated observatory network (eddy covariance system of Sidaoqiao superstation, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270782 (2019).
90. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (automatic weather station of Zhangye wetland station, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270768 (2019).
91. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of Zhangye wetland station, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270783 (2019).
92. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (an observation system of meteorological elements gradient of A'rou Superstation, 2014). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.248.2015.db (2016).
93. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (an observation system of meteorological elements gradient of A'rou superstation, 2015). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.308.2016.db (2016).
94. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of A'rou Superstation, 2014). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.235.2015.db (2016).

95. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of A'rou Superstation, 2015). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.296.2016.db (2016).
96. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (Dashalong automatic meteorological station, 2013). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.178.2014.db (2016).
97. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (automatic weather station of Dashalong station, 2014). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.251.2015.db (2016).
98. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (automatic weather station of Dashalong station, 2015). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.310.2016.db (2016).
99. Liu, S. *et al.* HiWATER: Dataset of Hydrometeorological observation network (eddy covariance system of Dashalong station, 2013). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.195.2014.db (2016).
100. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Dashalong station, 2014). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.238.2015.db (2016).
101. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Dashalong station, 2015). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.297.2016.db (2016).
102. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (automatic weather station of Yakou station, 2015). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.315.2016.db (2016).
103. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Yakou station, 2015). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.298.2016.db (2016).
104. Liu, S. *et al.* HiWATER: Dataset of hydro-meteorological observation network (automatic weather station of Dashalong station, 2016). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.456.2017.db (2017).
105. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Dashalong station, 2016). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.447.2017.db (2017).
106. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (automatic weather station of Yakou station, 2016). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.464.2017.db (2017).
107. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Yakou station, 2016). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.452.2017.db (2017).
108. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (an observation system of meteorological elements gradient of A'rou Superstation, 2017). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270897 (2018).
109. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of A'rou Superstation, 2017). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.5.2018.db (2018).
110. Liu, S. *et al.* HiWATER: Dataset of hydro-meteorological observation network (automatic weather station of Dashalong station, 2017). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.16.2018.db (2018).
111. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Dashalong station, 2017). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.7.2018.db (2018).
112. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (automatic weather station of Yakou station, 2017). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.14.2018.db (2018).
113. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Yakou station, 2017). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.12.2018.db (2018).
114. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (an observation system of meteorological elements gradient of A'rou Superstation, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270777 (2019).
115. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of A'rou superstation, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270775 (2019).
116. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (eddy covariance system of Dashalong station, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270788 (2019).
117. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of the Heihe River Basin integrated observatory network (eddy covariance system of Jingyangling station, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270784 (2019).
118. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of Heihe integrated observatory network (automatic weather station of Yakou station, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270769 (2019).
119. Liu, S. *et al.* Qilian Mountains integrated observatory network: Dataset of the Heihe River Basin integrated observatory network (eddy covariance system of Yakou station, 2018). *National Tibetan Plateau Data Center* https://doi.org/10.11888/Meteoro.tpdc.270781 (2019).
120. Liu, S., Li, X. & Xu, Z. HiWATER: Dataset of flux observation matrix (automatic meteorological station of No.1) of the MUlti-Scale Observation EXperiment on Evapotranspiration over heterogeneous land surfaces 2012 (MUSOEXE-12). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.059.2013.db (2016).
121. Liu, S., Li, X. & Xu, Z. HiWATER: The multi-scale observation experiment on evapotranspiration over heterogeneous land surfaces 2012 (MUSOEXE-12)-dataset of flux observation matrix (No.1 eddy covariance system). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.080.2013.db (2016).
122. Liu, S., Li, X. & Xu, Z. HiWATER: Dataset of flux observation matrix (automatic meteorological station of No.17) of the MUlti-Scale Observation EXperiment on Evapotranspiration over heterogeneous land surfaces 2012 (MUSOEXE-12). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.075.2013.db (2016).
123. Liu, S., Li, X. & Xu, Z. HiWATER: The multi-scale observation experiment on evapotranspiration over heterogeneous land surfaces (MUSOEXE-12)-dataset of flux observation matrix (No.17 eddy covariance system) from Mar to Sep, 2012. *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.095.2013.db (2016).
124. Liu, S. *et al.* HiWATER: Dataset of hydrometeorological observation network (eddy covariance system of Zhangye wetland Station, 2014). *National Tibetan Plateau Data Center* https://doi.org/10.3972/hiwater.246.2015.db (2016).
125. Liu, S. & Xu, Z. Multi-scale surface flux and meteorological elements observation dataset in the Hai River Basin (Daxing site-automatic weather station) (2008-2010). *National Tibetan Plateau Data Center* https://doi.org/10.3972/haihe.004.2013.db (2016).
126. Liu, S. & Xu, Z. Multi-scale surface flux and meteorological elements observation dataset in the Hai River Basin (Daxing site - eddy covariance system) (2008-2010). *National Tibetan Plateau Data Center* https://doi.org/10.3972/haihe.005.2013.db (2016).
127. Che, T. *et al.* Integrated hydrometeorological, snow and frozen-ground observations in the alpine region of the Heihe River Basin, China. *Earth Syst. Sci. Data* **11**, 1483–1499 (2019).
128. Jia, Z., Liu, S., Xu, Z., Chen, Y. & Zhu, M. Validation of remotely sensed evapotranspiration over the Hai River Basin, China. *J. Geophys. Res.-Atmos.* **117**, D13113 (2012).
129. Liu, S. *et al.* The heihe integrated observatory network: a basin-scale land surface processes observatory in China. *Vadose Zone J.* **17**, 1–21 (2018).
130. Liu, S. *et al.* Upscaling evapotranspiration measurements from multi-site to the satellite pixel scale over heterogeneous land surfaces. *Agric. For. Meteorol.* **230-231**, 97–113 (2016).

131. Liu, S. *et al.* A comparison of eddy-covariance and large aperture scintillometer measurements with respect to the energy balance closure problem. *Hydrol. Earth Syst. Sci.* **15**, 1291–1306 (2011).
132. Liu, S., Xu, Z., Zhu, Z., Jia, Z. & Zhu, M. Measurements of evapotranspiration from eddy-covariance systems and large aperture scintillometers in the Hai River Basin, China. *J. Hydrol.* **487**, 24–38 (2013).
133. Xu, Z. *et al.* Intercomparison of surface energy flux measurement systems used during the HiWATER-MUSOEXE. *J. Geophys. Res.-Atmos.* **118**, 13140–13157 (2013).
134. Dušek, J., Faußer, A., Stellner, S. & Kazda, M. Stems of Phragmites australis are buffering methane and carbon dioxide emissions. *Sci. Total Environ.* **882**, 163493 (2023).
135. Foltýnová, L., Fischer, M. & McGloin, R. P. Recommendations for gap-filling eddy covariance latent heat flux measurements using marginal distribution sampling. *Theor. Appl. Climatol.* **139**, 677–688 (2020).
136. Granier, A., Bréda, N., Longdoz, B., Gross, P. & Ngao, J. Ten years of fluxes and stand growth in a young beech forest at Hesse, North-eastern France. *Ann. For. Sci.* **65**, 704 (2008).
137. Kivalov, S. N. *et al.* Addressing effects of environment on eddy-covariance flux estimates at a Temperate Sedge-Grass Marsh. *Bound.-Layer Meteor.* **186**, 217–250 (2023).
138. Wutzler, T. *et al.* Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences* **15**, 5015–5030 (2018).
139. Aurela, M., Laurila, T., Hatakka, J., Tuovinen, J.-P. & Rainne, J. FLUXNET2015 RU-Tks Tiksi. *FLUXNET* https://doi.org/10.18140/FLX/1440244 (2016).
140. Bernhofer, C. *et al.* FLUXNET2015 DE-Spw Spreewald. *FLUXNET* https://doi.org/10.18140/FLX/1440220 (2016).
141. Bernhofer, C. *et al.* FLUXNET2015 DE-Tha Tharandt. *FLUXNET* https://doi.org/10.18140/FLX/1440152 (2016).
142. Chen, S. FLUXNET2015 CN-Du2 Duolun_grassland. *FLUXNET* https://doi.org/10.18140/FLX/1440140 (2016).
143. Dong, G. FLUXNET2015 CN-Cng Changling. *FLUXNET* https://doi.org/10.18140/FLX/1440209 (2016).
144. Ibrom, A. & Pilegaard, K. FLUXNET2015 DK-Sor Soroe. *FLUXNET* https://doi.org/10.18140/FLX/1440155 (2016).
145. Klatt, J., Schmid, H., Mauder, M. & Steinbrecher, R. FLUXNET2015 DE-SfN Schechenfilz Nord. *FLUXNET* https://doi.org/10.18140/FLX/1440219 (2016).
146. Kosugi, Y. & Takanashi, S. FLUXNET2015 MY-PSO Pasoh Forest Reserve (PSO). *FLUXNET* https://doi.org/10.18140/FLX/1440240 (2016).
147. Kotani, A. FLUXNET2015 JP-MBF Moshiri Birch Forest Site. *FLUXNET* https://doi.org/10.18140/FLX/1440238 (2016).
148. Kotani, A. FLUXNET2015 JP-SMF Seto Mixed Forest Site. *FLUXNET* https://doi.org/10.18140/FLX/1440239 (2016).
149. Li, Y. FLUXNET2015 CN-Ha2 Haibei Shrubland. *FLUXNET* https://doi.org/10.18140/FLX/1440211 (2016).
150. Maximov, T. FLUXNET2015 RU-SkP Yakutsk Spasskaya Pad larch. *FLUXNET* https://doi.org/10.18140/FLX/1440243 (2016).
151. Olesen, J. FLUXNET2015 DK-Fou Foulum. *FLUXNET* https://doi.org/10.18140/FLX/1440154 (2016).
152. Pastorello, G. *et al.* The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Sci. Data* **7**, 225, https://doi.org/10.1038/s41597-020-0534-3 (2020).
153. Pilegaard, K. & Ibrom, A. FLUXNET2015 DK-Eng Enghave. *FLUXNET* https://doi.org/10.18140/FLX/1440153 (2016).
154. Sachs, T., Wille, C., Larmanou, E. & Franz, D. FLUXNET2015 DE-Zrk Zarnekow. *FLUXNET* https://doi.org/10.18140/FLX/1440221 (2016).
155. Schneider, K. & Schmidt, M. FLUXNET2015 DE-Seh Selhausen. *FLUXNET* https://doi.org/10.18140/FLX/1440217 (2016).
156. Shao, C. FLUXNET2015 CN-Du3 Duolun Degraded Meadow. *FLUXNET* https://doi.org/10.18140/FLX/1440210 (2016).
157. Shao, C. FLUXNET2015 CN-Sw2 Siziwang Grazed (SZWG). *FLUXNET* https://doi.org/10.18140/FLX/1440212 (2016).
158. Shi, P., Zhang, X. & He, Y. FLUXNET2015 CN-Dan Dangxiong. *FLUXNET* https://doi.org/10.18140/FLX/1440138 (2016).
159. Tang, Y., Kato, T. & Du, M. FLUXNET2015 CN-HaM Haibei Alpine Tibet site. *FLUXNET* https://doi.org/10.18140/FLX/1440190 (2016).
160. Wang, H. & Fu, X. FLUXNET2015 CN-Qia Qianyanzhou. *FLUXNET* https://doi.org/10.18140/FLX/1440141 (2016).
161. Zhang, J. & Han, S. FLUXNET2015 CN-Cha Changbaishan. *FLUXNET* https://doi.org/10.18140/FLX/1440137 (2016).
162. Zhou, G. & Yan, J. FLUXNET2015 CN-Din Dinghushan. *FLUXNET* https://doi.org/10.18140/FLX/1440139 (2016).
163. Anapalli, S. S. *et al.* Quantifying soybean evapotranspiration using an eddy covariance approach. *Agric. Water Manage.* **209**, 228–239 (2018).
164. Xie, M. *et al.* Simulation of site-scale water fluxes in desert and natural oasis ecosystems of the arid region in Northwest China. *Hydrol. Process.* **35**, e14444 (2021).
165. Zhang, C. *et al.* A framework for estimating actual evapotranspiration at weather stations without flux observations by combining data from MODIS and flux towers through a machine learning approach. *J. Hydrol.* **603**, 127047 (2021).
166. Liang, S. *et al.* The global land surface satellite (GLASS) product suite. *Bull. Amer. Meteorol. Soc.* **102**, 1–37 (2020).
167. Zhang, X., Liang, S., Zhou, G., Wu, H. & Zhao, X. Generating Global Land Surface Satellite incident shortwave radiation and photosynthetically active radiation products from multiple satellite data. *Remote Sens. Environ.* **152**, 318–332 (2014).
168. Tang, W. Dataset of high-resolution (3 hour, 10 km) global surface solar radiation (1983–2018). *National Tibetan Plateau Data Center* https://cstr.cn/18406.11.Meteoro.tpdc.270112 (2019).
169. Tang, W., Yang, K., Qin, J., Li, X. & Niu, X. A 16-year dataset (2000-2015) of high-resolution (3 h, 10 km) global surface solar radiation. *Earth Syst. Sci. Data* **11**, 1905–1915 (2019).
170. Myneni, R., Knyazikhin, Y. & Park, T. MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006. *NASA EOSDIS Land Processes DAAC* https://doi.org/10.5067/MODIS/MCD15A3H.006 (2015).
171. Vermote, E. & Wolfe, R. MOD09GA MODIS/Terra Surface Reflectance Daily L2G Global 1kmand 500m SIN Grid V006. *NASA EOSDIS Land Processes DAAC* https://doi.org/10.5067/MODIS/MOD09GA.006 (2015).
172. Fang, B., Lei, H., Zhang, Y., Quan, Q. & Yang, D. Spatio-temporal patterns of evapotranspiration based on upscaling eddy covariance measurements in the dryland of the North China Plain. *Agric. For. Meteorol.* **281**, 107844 (2020).
173. Yamazaki, D. *et al.* A high-accuracy map of global terrain elevations. *Geophys. Res. Lett.* **44**, 5844–5853 (2017).
174. FAO/IIASA/ISRIC/ISSCAS/JRC. Harmonized World Soil Database (version 1.2). *FAO, Rome, Italy and IIASA, Laxenburg, Austria* https://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML (2012).
175. Friedl, M. & Sulla-Menashe, D. MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006. *NASA EOSDIS Land Processes DAAC* https://doi.org/10.5067/MODIS/MCD12Q1.006 (2019).
176. Sorensen, L. A spatial analysis approach to the global delineation of dryland areas of relevance to the CBD Programme of Work on Dry and Subhumid Lands. *UNEP-WCMC* https://www2.unep-wcmc.org/resources-and-data/a-spatial-analysis-approach-to-the-global-delineation-of-dryland-areas-of-relevance-to-the-cbd-programme-of-work-on-dry-and-subhumid-lands (2007).
177. Xie, M. Division of flux stations. *figshare* https://doi.org/10.6084/m9.figshare.23695980.v2 (2023).
178. Lopatin, J., Dolos, K., Hernández, H., Galleguillos, M. & Fassnacht, F. Comparing generalized linear models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. *Remote Sens. Environ.* **173**, 200–210 (2016).
179. Biau, G. Analysis of a random forests model. *J. Mach. Learn. Res.* **13**, 1063–1095 (2012).
180. Bergstra, J., Komer, B., Eliasmith, C., Yamins, D. & Cox, D. D. Hyperopt: a python library for model selection and hyperparameter optimization. *Comput. Sci. Discov.* **8**, 014008 (2015).
181. Ploton, P. *et al.* Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **11**, 1–11 (2020).

182. Patel, S. P. & Upadhyay, S. H. Euclidean distance based feature ranking and subset selection for bearing fault diagnosis. *Expert Syst. Appl.* **154**, 113400 (2020).
183. Lever, J., Krzywinski, M. & Altman, N. Points of significance: Principal component analysis. *Nat. Methods* **14**, 641–643 (2017).
184. Xie, M. Carbon-water flux datasets of Eurasian meteorological stations. *figshare* https://doi.org/10.6084/m9.figshare.21347721.v3 (2022).
185. Xie, M. Meteorological station information. *figshare* https://doi.org/10.6084/m9.figshare.23695920.v2 (2023).
186. Xie, M. RSMs information. *figshare* https://doi.org/10.6084/m9.figshare.23899785.v1 (2023).
187. Jiang, F. *et al.* A 10-year global monthly averaged terrestrial net ecosystem exchange dataset inferred from the ACOS GOSAT v9 XCO 2 retrievals (GCAS2021). *Earth Syst. Sci. Data* **14**, 3013–3037 (2022).
188. Running, S., Mu, Q. & Zhao, M. MOD16A2 MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid V006. *NASA EOSDIS Land Processes DAAC* https://doi.org/10.5067/MODIS/MOD16A2.006 (2017).
189. Xie, M. Code for mining carbon-water flux information at Eurasian meteorological stations. *figshare* https://doi.org/10.6084/m9.figshare.21510183.v2 (2022).

Acknowledgements

This research was supported by the Tianshan Talent Cultivation (Grant No. 2022TSYCLJ0001), the Key Projects of the Natural Science Foundation of Xinjiang Autonomous Region (Grant No. 2022D01D01), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA20060302), the High-End Foreign Experts Project and the China Scholarship Council. Mingjuan Xie was supported by a grant from the program of China Scholarship Council (ICPIT–International Cooperative Program for Innovative Talents, Grant No. 202110630005) during her stay in Ghent University, Gent, Belgium. Andrej Varlagin was supported by Russian Science Foundation (project 21-14-00209). Iris Feigenwinter was funded by the EU project SUPER-G (Grant No. 774124) and the SNF project ICOS-CH (Grant No. 20F120_198227). Ankit Shekhar acknowledges funding by the ETH Zürich project FEVER ETH-27 19-1. The University of Padova (AP, FM, LT) carried out the study within the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022). The part of this work was carried out within the framework of the state assignment of the Ministry of Education and Science of Russia under the project "Study of biogeochemical cycles and adaptive reactions of plants of boreal and arctic ecosystems of northeastern Russia", AAAA-A21-121012190034-2. We thank the many people who provided the metadata for this research. The MCD15A3H data, the MOD09GA data, the MCD12Q1 data and the MOD16A2 data were provided by the NASA EOSDIS Land Processes DAAC. The GOSAT dataset was obtained via the GOSAT Data Archive Service, and the GOSAT project is a joint effort promoted by the JAXA, the NIES and MOE. The DSR dataset during 1983–2018 and some of the flux station data were provided by the National Tibetan Plateau Data Center (http://data.tpdc.ac.cn). Some of the flux station data were shared by the European Fluxes Database Cluster (http://www.europe-fluxdata.eu/home) and the FLUXNET, whose data providers include Adriano Conte, Adrien Jacotot, Aino Korrensalo, Alexander Knohl, Anders Lindroth, Andrea Pitacco, Andreas Ibrom, Andrej Varlagin, Ankit Shekhar, Annalea Lohila, Anne De Ligne, Arnaud Carrara, Aurore Brut, Axel Don, Ayumi Kotani, Bart Kruijt, Benjamin Loubet, Bernard Heinesch, Bogdan Chojnicki, Carlo Calfapietra, Carole Helfter, Caroline Vincke, Casimiro Pio, Changliang Shao, Christian Bernhofer, Christian Brümmer, Christian Markwitz, Christian Wille, Christoph Ammann, Claire Campbell, Cristina Gimeno, Dan Yakir, Daniel Berveiller, Daniela Franz, Dario Papale, Denis Loustau, Donatella Spano, Edoardo Cremonese, Eeva-Stiina Tuittila, Eiko Nemitz, Eric Ceschia, Eric Grehan, Eric Larmanou, Fabio Turco, Fanny Kittler, Fatima LAGGOUN, Filipe Costa e Silva, Francesco Mazzenga, Franco Meggio, Franco Miglietta, Francois Gastal, Franziska Koebsch, Frédéric Bornet, Frédéric Guibal, Frederik Schrader, Gang Dong, Gary J. Lanigan, Georg Niedrist, Georg Wohlfahrt, Gerald Jurasinski, Gerard Kiely, Giorgio Matteucci, Giovanni Manca, Giuseppe Scarascia Mugnozza, Guillaume Simioni, Guoyi Zhou, Hans Peter Schmid, Huimin Wang, Ignacio Goded, Iris Feigenwinter, Ivan Janssens, Thomas Gruenwald, Jan Elbers, Jan Segers, Janina Klatt, Janusz Olejnik, Jean-Christophe Calvet, Jean-Marc Limousin, Jean-Pierre Delorme, Jiri Dusek, Joachim Jansen, Joao Santos Pereira, Joel Leonard, Jørgen E. Olesen, Johan Neirynck, John Moncrieff, Juha Hatakka, Juha-Pekka Tuovinen, Junhua Yan, Junhui Zhang, Jutta Holst, Juuso Rainne, Kari Minkkinen, Karl Schneider, Katerina Havrankova, Katja KLUMPP, Kim Pilegaard, Ladislav Šigut, Leif Klemedtsson, Lenka Foltynova, Louis Gourlez de la Motte, Luca Tezza, Lukas Hörtnagl, Lutz Merbold, Marek Urbaniak, Mana Gharun, Margaret Anderson-Dunn, Marian Pavelka, Marilyn Roland, Marius Schmidt, Mark A. Sutton, Markus Hehn, Marta Galvagno, Mathias Goeckede, Mathilde Jammet, Matthew Saunders, Matthew Wilkinson, Matthias Cuntz, Matthias Mauder, Michal Heliasz, Michel Vennetier, Mika Aurela, Mika Korkiakoski, Mike Jones, Mingyuan Du, Mirco Migliavacca, Monique Carnol, Nadia Vendrame, Natalia Kowalska, Nelius Foley, Nicola Arriga, Nina Buchmann, Olaf Kolle, Olivier Marloie, Paolo Stefani, Pasi Kolari, Patrick Crill (Dept of Geological Sciences, Stockholm University, Sweden), Paul G. Leahy, Pauline Buysse, Pavel Alekseychik, Peili Shi, Per Weslien, Radek Czerny, Radoslaw Juszczak, Rainer Steinbrecher, Regine Maier, Rémy Soubie, Richard Harding, Rober Falcimagne, Robert Clement, Satoru Takanashi, Sebastien Gogo, Shijie Han, Shiping Chen, Silvano Fares, Sofia Cerasoli, Tanguy Manise, Tarek El-Madany, Thomas Friborg, Tim De Meulder, Timo Vesala, Tiphaine Tallec, Tiziano Sorgi, Tomomichi Kato, Torsten Sachs, Trofim Maximov, Tuomas Laurila, Umberto Morra di Cella, Uta Moderow, Valerio Moretti, Vincenzo Magliulo, Wilma Jans, Xianzhou Zhang, Xiaoli Fu, Yanhong Tang, Yi Wang, Yingnian Li, Yongtao He, Yoshiko Kosugi, Zoltan Nagy, and Zsolt Csintalan.

Author contributions

Geping Luo conceived the work. Mingjuan Xie, Xiaofei Ma and Geping Luo initiated this research and were responsible for the work integrity as a whole. Mingjuan Xie and Xiaofei Ma performed the formal analysis and the calculations and drafted the manuscript. Geping Luo, Mingjuan Xie, Xiaofei Ma, Yuangang Wang, Chaofan Li, Haiyang Shi, Xiuliang Yuan, Chunbo Chen, Wenqiang Zhang, Chen Zhang, Qing Ling, Ruixiang Gao, Yu Zhang and Friday Uchenna Ochege were responsible for the data collection and analysis. Geping Luo, Amaury Frankl,

Philippe De Maeyer, and Olaf Hellwich contributed to the resources and manuscript review. Nina Buchmann, Iris Feigenwinter, Jørgen E. Olesen, Radoslaw Juszczak, Adrien Jacotot, Aino Korrensalo, Andrea Pitacco, Andrej Varlagin, Ankit Shekhar, Annalea Lohila, Arnaud Carrara, Aurore Brut, Bart Kruijt, Benjamin Loubet, Bernard Heinesch, Bogdan Chojnicki, Carole Helfter, Caroline Vincke, Changliang Shao, Christian Bernhofer, Christian Brümmer, Christian Wille, Eeva-Stiina Tuittila, Eiko Nemitz, Franco Meggio, Gang Dong, Gary J. Lanigan, Georg Niedrist, Georg Wohlfahrt, Guoyi Zhou, Ignacio Goded, Thomas Gruenwald, Janusz Olejnik, Joachim Jansen, Johan Neirynck, Juha-Pekka Tuovinen, Junhui Zhang, Katja KLUMPP, Kim Pilegaard, Ladislav Šigut, Leif Klemedtsson, Luca Tezza, Lukas Hörtnagl, Marek Urbaniak, Marilyn Roland, Marius Schmidt, Mark A. Sutton, Markus Hehn, Matthew Saunders, Matthias Mauder, Mika Aurela, Mika Korkiakoski, Mingyuan Du, Nadia Vendrame, Natalia Kowalska, Paul G. Leahy, Pavel Alekseychik, Peili Shi, Per Weslien, Shiping Chen, Silvano Fares, Thomas Friborg, Tiphaine Tallec, Tomomichi Kato, Torsten Sachs, Trofim Maximov, Umberto Morra di Cella, Uta Moderow, Yingnian Li, Yongtao He, and Yoshiko Kosugi contributed to providing the flux station data and manuscript proofreading.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-023-02473-9.

Correspondence and requests for materials should be addressed to G.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Mingjuan Xie[1,2,3,4,5,63], Xiaofei Ma[1,63], Yuangang Wang[1,3], Chaofan Li[6], Haiyang Shi[7], Xiuliang Yuan[1], Olaf Hellwich[8], Chunbo Chen[1], Wenqiang Zhang[1,2,3,4,5], Chen Zhang[1,3], Qing Ling[1,3], Ruixiang Gao[1,3], Yu Zhang[1,2,3,4,5], Friday Uchenna Ochege[1,9], Amaury Frankl[2], Philippe De Maeyer[1,2,3,4,5], Nina Buchmann[10], Iris Feigenwinter[10], Jørgen E. Olesen[11], Radoslaw Juszczak[12], Adrien Jacotot[13], Aino Korrensalo[14,15], Andrea Pitacco[16], Andrej Varlagin[17], Ankit Shekhar[10], Annalea Lohila[18,19], Arnaud Carrara[20], Aurore Brut[21], Bart Kruijt[22], Benjamin Loubet[23], Bernard Heinesch[24], Bogdan Chojnicki[12], Carole Helfter[25], Caroline Vincke[26], Changliang Shao[27], Christian Bernhofer[28], Christian Brümmer[29], Christian Wille[30], Eeva-Stiina Tuittila[31], Eiko Nemitz[25], Franco Meggio[16], Gang Dong[32], Gary Lanigan[33], Georg Niedrist[34], Georg Wohlfahrt[35], Guoyi Zhou[36], Ignacio Goded[37], Thomas Gruenwald[28], Janusz Olejnik[38], Joachim Jansen[39], Johan Neirynck[40], Juha-Pekka Tuovinen[18], Junhui Zhang[41], Katja Klumpp[42], Kim Pilegaard[43], Ladislav Šigut[44], Leif Klemedtsson[45], Luca Tezza[16], Lukas Hörtnagl[10], Marek Urbaniak[38], Marilyn Roland[46], Marius Schmidt[47], Mark A. Sutton[25], Markus Hehn[28], Matthew Saunders[48], Matthias Mauder[28], Mika Aurela[18], Mika Korkiakoski[18], Mingyuan Du[49], Nadia Vendrame[50], Natalia Kowalska[44], Paul G. Leahy[51], Pavel Alekseychik[52], Peili Shi[53], Per Weslien[45], Shiping Chen[54], Silvano Fares[55], Thomas Friborg[56], Tiphaine Tallec[21], Tomomichi Kato[57], Torsten Sachs[30], Trofim Maximov[58], Umberto Morra di Cella[59], Uta Moderow[28], Yingnian Li[60], Yongtao He[53], Yoshiko Kosugi[61] & Geping Luo[1,3,4,62] ✉

[1]State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, Xinjiang, 830011, China. [2]Department of Geography, Ghent University, Ghent, 9000, Belgium. [3]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, 100049, China. [4]Sino-Belgian Joint Laboratory for Geo-Information, Urumqi, China. [5]Sino-Belgian Joint Laboratory for Geo-Information,

Ghent, Belgium. [6]School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing, 210044, China. [7]School of Earth Sciences and Engineering, Hohai University, Nanjing, 211100, China. [8]Department of Computer Vision & Remote Sensing, Technische Universität Berlin, 10587, Berlin, Germany. [9]Department of Geography and Environmental Management, University of Port Harcourt, PMB 5323 Choba, East-West, Port Harcourt, Nigeria. [10]Department of Environmental Systems Science, Institute of Agricultural Sciences, ETH Zürich, 8092, Zürich, Switzerland. [11]Department of Agroecology, Aarhus University, Tjele, Denmark. [12]Laboratory of Bioclimatology, Department of Ecology and Environmental Protection, Faculty of Environmental and Mechanical Engineering, Poznan University of Life Sciences, Piatkowska 94, 60-649, Poznan, Poland. [13]Sol, Agro et hydrosystèmes, Spatialisation (SAS), UMR 1069, INRAE, Institut Agro, 35000, Rennes, France. [14]Department of Environmental and Biological Sciences, University of Eastern Finland, Joensuu campus, P.O Box 111, Joensuu, FI-80101, Finland. [15]Natural Resources Institute Finland, Joensuu, Yliopistokatu 6, FI-80130, Joensuu, Finland. [16]University of Padova - DAFNAE, Viale dell'Università 16, I-35020, Padova, Legnaro (PD), Italy. [17]A.N Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, 119071, Leninsky pr.33, Moscow, Russia. [18]Climate System Research, Finnish Meteorological Institute, P.O Box 503, FI-00101, Helsinki, Finland. [19]University of Helsinki, Institute for Atmospheric and Earth System Research (INAR)/Physics, Faculty of Science, Helsinki, Finland. [20]Fundación CEAM, Parque Tecnológico, C/Charles R. Darwin, 14, Paterna, 46980, Spain. [21]CESBIO, Université de Toulouse, CNES/CNRS/INRAE/IRD/UPS, Toulouse, France. [22]Wageningen Univertsity, Water Systems and Global change group, PO bx 47, 7700AA, Wageningen, Netherlands. [23]ECOSYS, INRAE, AgroParisTech, Université Paris-Saclay, 22 place de l'agronomie, 91120, Palaiseau, France. [24]Terra Teaching and Research Center, University of Liège – Gembloux Agro-Bio Tech, 5030, Gembloux, Belgium. [25]UK Centre for Ecology & Hydrology (UKCEH), Bush Estate, Penicuik, EH26 0QB, UK. [26]Earth and Life Institute, Université Catholique de Louvain, 1348, Louvain-la-Neuve, Belgium. [27]National Hulunber Grassland Ecosystem Observation and Research Station & Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, 100081, China. [28]Institute of Hydrology and Meteorology, TUD Dresden University of Technology, Pienner Str. 23, 01737, Tharandt, Germany. [29]Thünen Institute of Climate-Smart Agriculture, 38116, Braunschweig, Germany. [30]GFZ German Research Centre for Geosciences, Telegrafenberg, 14473, Potsdam, Germany. [31]School of Forest Sciences, University of Eastern Finland, P.O Box 111, FIN-80100, Joensuu, Finland. [32]School of Life Science, Shanxi University, Taiyuan, 030006, China. [33]Teagasc, Johnstown Castle, Wexford, Ireland. [34]Eurac research, Institute for Alpine Environment, Viale Druso 1, 39100, Bolzano, Italy. [35]Institut für Ökologie, Universität Innsbruck, Innrain 52, 6020, Innsbruck, Austria. [36]Institute of Ecology and School of Applied Meteorology, Nanjing University of Information Science & Technology, Nanjing, 210044, China. [37]European Commission, Joint Research Centre (JRC), Ispra, Italy. [38]Laboratory of Meteorology, Department of Construction and Geoengineering, Faculty of Environmental and Mechanical Engineering, Poznan University of Life Sciences, Piatkowska 94, 60-649, Poznan, Poland. [39]Department of Ecology and Genetics/Limnology, Uppsala University, Norbyvägen 18 D, 752 36, Uppsala, Sweden. [40]Research Institute for Nature and Forest, Geraardsbergen, 9500, Belgium. [41]School of life sciences, Qufu Normal University, 57 Jingxuan West Road, Qufu, 273165, Shandong, China. [42]Grassland Ecosystem Research, INRAE, VetAgro-Sup, University of Clermont Auvergne, 5 Chemin de Beaulieu, 63000, Clermont Ferrand, France. [43]Department of Environmental Engineering, Technical University of Denmark (DTU), Kgs, Lyngby, 2800, Denmark. [44]Department of Matter and Energy Fluxes, Global Change Research Institute CAS, Bělidla 986/4a, CZ-603 00, Brno, Czech Republic. [45]Departement of Earth Sciences, Gothenburg University, Guldhedsgatan 5A, Po.Box 460, SE 405 30, Gothenburg, Sweden. [46]Department of Biology, University of Antwerp, Wilrijk, 2610, Belgium. [47]Agrosphere Institute IBG-3, Forschungszentrum Jülich, Jülich, 52425, Germany. [48]School of Natural Sciences, Botany Discipline, Trinity College Dublin, D2, Dublin, Ireland. [49]National Agriculture and Food Research Organization, Tsukuba, Ibaraki, 305-8517, Japan. [50]Center Agriculture Food Environment, University of Trento, Via Edmund Mach 1, I-38010, Trento, San Michele all'Adige (TN), Italy. [51]School of Engineering & Architecture, University College Cork, College Road, Cork, T12 K8AF, Republic of Ireland. [52]Natural Resources Institute Finland, Bioeconomy and environment, 00790, Helsinki, Finland. [53]Lhasa Station, Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China. [54]State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, 100093, China. [55]National Research Council of Italy, Institute for Agriculture and Forestry Systems in the Mediterranean, Portici, Naples, Italy. [56]Department of Geosciences and Natural Resource Management, University of Copenhagen, Oester Voldgade 10, 1350, Copenhagen K, Denmark. [57]Research Faculty of Agriculture, Hokkaido University, Sapporo, Hokkaido, 060-8589, Japan. [58]Institute for Biological Problems of Cryolithozone, Siberian Branch of the Russian Academy of Sciences, Yakutsk, Russia. [59]Climate Change Dept., Environmental Protection Agency of Aosta Valley, Saint-Christophe, I-11020, Italy. [60]Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Qinghai, Xining, 810008, China. [61]Laboratory of Forest Hydrology, Graduate School of Agriculture, Kyoto University, 606-8502, Kyoto, Japan. [62]The National Key Laboratory of Ecological Security and Sustainable Development in Arid Region (proposed), Chinese Academy of Sciences, Urumqi, China. [63]These authors contributed equally: Mingjuan Xie, Xiaofei Ma. ✉e-mail: luogp@ms.xjb.ac.cn