# Estimating Global Soil Heterotrophic Respiration Based on Environmentally Similar Zones and Remote Sensing Data

Luying Zhu, Ni Huang, Li Wang, Zheng Niu, Jinxiao Wang, Yuelin Zhang, Jie Liu

***Abstract*—Accurately estimating global soil heterotrophic respiration ($R_H$) is crucial in evaluating whether terrestrial ecosystems act as carbon sources or sinks. However, current global $R_H$ estimates were significantly restricted by the scarcity of in-situ $R_H$ observations and their biased distribution, leading to considerable uncertainties. This study developed a novel data-driven model of global $R_H$ based on the environmentally similar zones of global in-situ $R_H$ sites with daily and subdaily observations and remote sensing data with high spatial and temporal resolutions. Compared with the unified modelling method using all available data as the training samples of data-driven models, our zone modelling method was more accurate. The relationship between observed and predicted $R_H$ was improved, with the $R^2$ value increasing from 0.41 to 0.53 and the RMSE decreasing from 0.87 to 0.78 g C $m^{-2}$ $d^{-1}$. Our study effectively improved the problem that the data-driven models were highly affected by the spatial representativeness of in-situ $R_H$ observations and achieved a significantly improved accuracy for global $R_H$ estimation entirely based on remote sensing data. Future researches focusing on improving the sparse sampling of in-situ $R_H$ sites and the availability of remote sensing data will help to reduce the uncertainties of our study.**

***Index Terms*—soil heterotrophic respiration, global estimates, remote sensing data.**

## I. INTRODUCTION

SOIL heterotrophic respiration ($R_H$) represents the release of $CO_2$ from the decomposition of soil organic carbon (SOC) by soil microorganisms and fauna [1]. Global total $R_H$ is approximately four times higher than $CO_2$ emissions from fossil fuel combustion [2], and accounts for 54–63% of global total soil respiration ($R_S$) [3]. Therefore, $R_H$ plays an important role in regulating the global carbon cycle and is potentially sensitive to future climate change [4]. However, significant uncertainties exist in the estimation of global $R_H$.

Numerous field experiments on $R_H$ have been conducted during recent decades to investigate its spatiotemporal variations and responses to environmental factors [5-7]. To establish a reliable basis for regional or global scientific research on soil carbon flux, great efforts have been made to integrate past $R_H$ measurements into standardized databases, such as the global Soil Respiration Database (SRDB) [8, 9]. Based on the SRDB, a dataset containing daily and monthly $R_H$ observations was compiled (DGRsD) [10, 11]. In addition, COSORE, which is a continuous, long-term dataset of soil-atmosphere greenhouse gas fluxes from globally distributed sites, offers half-hourly and hourly observations of $R_H$ [12]. However, these in-situ $R_H$ data available from global datasets are not explicitly designed to provide uniform and consistent observations of the global carbon flux. They only represent a sparse and spatially biased sample of global terrestrial ecosystems, particularly tropical ones [13].

Empirical statistical modelling based on in-situ observations can be employed to estimate global $R_H$. For example, Bond-Lamberty, et al. [14] and Subke, et al. [15] obtained empirical relationships between $R_S$ and $R_H$ through analyses of in-situ observations. These empirical relationships, alongside spatially explicit $R_S$ information, have been utilized to drive global estimates of $R_H$ [16-18]. Nevertheless, it's important to note that the estimation results of $R_S$, the empirical relationship between $R_H$ and $R_S$, and the various $R_H$ measurements methods all contributed to sources of uncertainty. Unlike the aforementioned method, machine learning techniques, which are fully data-adaptive and do not require prior assumptions about functional relationships, can capture and establish nonlinear relationships through extensive model training. They enable the upscaling of location- and time-constrained measurements to spatially and temporally explicit datasets [19]and have been gradually applied to estimate global carbon fluxes [20, 21]. Random Forest (RF) is the most prevalent machine learning method and has been used to estimate global $R_H$ [22-24].

Although the advantages are obvious, there were still some limitations in these data-driven global $R_H$ estimates. Firstly, they suffered from low spatial and temporal resolution. On the temporal scale, all datasets were derived from annual $R_H$ observations, ignoring the inter-annual variations of $R_H$.

This research was funded by the National Key Research and Development Program of China (2023YFB3907402) and the National Natural Science Foun dation of China (no. 42371361). *(Corresponding author: Ni Huang)*

Luying Zhu, Zheng Niu, Jie Liu, are with Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute - Chinese Academy of Sciences, Beijing 100094, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China. (e-mail: zhuluying18@mails.ucas.ac.cn; niuzheng@radi.ac.cn; liujie222@mails.ucas.ac.cn).

Ni Huang, Li Wang, Yuelin Zhang are with Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute - Chinese Academy of Sciences, Beijing 100094, China (e-mail: huangni@radi.ac.cn; wangli@radi.ac.cn; zyl08@live.com).

Jinxiao Wang is with Key Laboratory of Target Cognition and Application Technology and Key Laboratory of Network Information System Technology, Aerospace Information Research Institute - Chinese Academy of Sciences, Beijing 100190, China (e-mail: wangjx@radi.ac.cn).

Concerning spatial resolution, previous studies have predominantly utilized meteorological data with coarse spatial resolution as model input parameters, resulting in global $R_H$ estimates with coarse spatial resolution. However, advances in Earth observation technology provide an opportunity to overcome the spatiotemporal resolution limitations of global $R_H$ estimates [25]. In particular, the increasing availability of satellite remote sensing data (i.e., temperature-, water- and vegetation-related datasets) with various spatial and temporal scales [26-28] make it possible to estimate global $R_H$ at finer spatial and temporal scales. Secondly, current research primarily concentrated on modelling $R_H$ at the global scale, which did not account for the high spatial heterogeneity in $R_H$ and the non-uniform distribution of in-situ observation sites. Several recent studies demonstrated that spatiotemporal variations in $R_H$ were comprehensively affected by various environmental factors [29, 30]. Thus, establishing $R_H$ models based on environmentally similar zones may be a more feasible method for estimating global $R_H$. A similar study has been conducted to estimate global $R_S$ based on specific models established for eight climate zones [11].

This study developed a new method for estimating global $R_H$ based on environmentally similar zones of all global available $R_H$ observation records at subdaily and daily time scale. This method used remote sensing data with high spatial and temporal resolutions to establish a data-driven $R_H$ model for each zone. Given the similar environmental factors within each zone, we expect these zonal models to provide greater accuracy than conventional global $R_H$ models based on undivided observation data.

## II. Materials

### *A. Construction of an in-situ $R_H$ database*

Two available $R_H$ databases, DGRsD and COSORE, were selected to generate an 8-day temporal resolution in-situ $R_H$ observation database. To ensure data quality and eliminate mismatches between in-situ $R_H$ data and satellite remote sensing data, we developed a set of criteria for constructing the in-situ database: (1) only observations obtained using infrared gas analyzer or gas chromatography methods were selected; (2) measurements obtained after human intervention, such as fertilization or temperature control, were excluded [10]; (3) low-quality and duplicated data were deleted; (4) the Moderate Resolution Imaging Spectroradiometer (MODIS) used in this study has only been officially releasing data since 2000, thus only measurements taken after 2000 were retained; and (5) to obtain the in-situ $R_H$ observations that match the 8-day temporal resolution remote sensing data, the hourly and half-hourly $R_H$ observations from COSORE, as well as the daily $R_H$ observations from DGRsD, were averaged within the same 8-day compositing intervals periods as remote sensing data. In addition to $R_H$, the dataset included information such as latitude, longitude, observation year, month, cumulative day of the year (DOY), temperature, precipitation, and vegetation type. Based on the above criteria, we constructed an in-situ $R_H$ database containing 3654 observations obtained from 167 stations from 2000 to 2019 (Fig. 1).

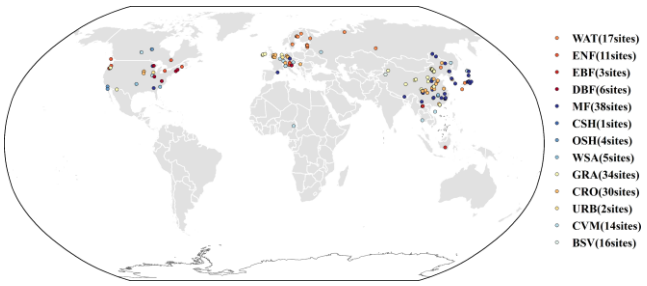


Fig. 1. Spatial distribution of selected soil heterotrophic respiration in-situ measurement sites according to the International Geosphere–Biosphere Programme (IGBP) land cover classification: WAT (Wetland), ENF (Evergreen Needleleaf Forest), EBF (Evergreen Broadleaf Forest), DBF (Deciduous Broadleaf Forest), MF (Mixed Forest), CSH (Closed Shrublands), OSH (Open Shrublands), WSA (Woody Savannas), GRA (Grasslands), CRO (Croplands), URB (Urban), CVM (Cropland/Natural Vegetation Mosaic) and BSV (Barren or Sparsely Vegetated).

### *B. Remote sensing data*

We obtained the 1-km MOD11A2 land surface temperature (LST) product (https://lpdaac.usgs.gov/products/mod11a2v006/), 500-m MOD09A1 surface spectral reflectance product (https://lpdaac.usgs.gov/products/mod09a1v061/) and 1-km MOD12Q1 Land Cover Type (LCT) product (https://lpdaac.usgs.gov/products/mcd12q1v006/) from available MODIS data at an 8-day temporal resolution from 2000 to 2020. The normalized difference vegetation index (NDVI), enhanced vegetation index (EVI) and land surface water index (LSWI) were calculated based on MOD09A1. Gross primary production (GPP), the fraction of absorbed photosynthetically active radiation (FPAR) and leaf area index (LAI) were derived from data from the Global Land Surface Satellite (GLASS) at an 8-day/500-m resolution (accessed at http://www.bnu-datacenter.com/). Penman-Monteith-Leuning Evapotranspiration V2 (PML_V2) products provided evapotranspiration (ET) data at an 8-day/500-m resolution [27] (freely accessible at https://code.earthengine.google.com/?asset=projects/pml_evapotranspiration/PML/OUTPUT/PML_V2_8day_v014).

### *C. Climate and DEM data*

Daily air temperature (TEM) and precipitation (PRE) data at a resolution of 0.5° × 0.5° were sourced from the fifth generation of European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis of the global climate, known as ERA5 (downloadable at https://cds.climate.copernicus.eu/cdsapp#!/home). The Multi-Error-Removed Improved-Terrain Digital Elevation Model (MERIT DEM) used in this study is a high-accuracy global DEM with 3-arc-second resolution. It was generated by eliminating major error components from existing DEM products [31] and is available at http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/.

## III. Methods

### A. Environmentally similar zones among global in-situ $R_H$

#### 1) Selection of environmental factors

$R_H$ represents a comprehensive ecological process characterized by high spatial heterogeneity. It is influenced by various factors such as temperature, precipitation, land cover type, topography and soil properties [29]. Comprehensively considering variable representativeness and data availability, four categories of environmental factors (climate, vegetation, topography, soil property) were selected to assign the global $R_H$ sites to environmentally similar zones. The climate factors were multi-year averaged TEM and PRE, vegetation factors were multi-year averaged NDVI and growing season LSWI, topography was represented by DEM and soil property was represented by SOC.

#### 2) Environmentally similar zones based on k-means clustering

The selected six environmental factors were acquired for each site during the period 2000–2020 based on their latitude and longitude. We used multiyear-averaged variables (2000–2020) for the following analysis. To mitigate the impacts of covariance and differing scales among these environmental factors, principal component transformation and normalization were employed. The initial first four principal components effectively accounted for over 80% of the information present in the original data and were subsequently utilized in place of the original six factors.

*k*-means clustering was applied to assign global in-situ $R_H$-measurement sites to different environmentally similar zones based on the pre-processed environmental factors [32]. To address issues inherent in the *k*-means algorithm, such as sensitivity to initial cluster centres and limited global search ability, the particle swarm optimization (PSO) algorithm was used [33]. The appropriate number of zones and the zone results were determined through experiments and comparative analyses. Subsequently, an analysis of the similarities and differences among the various zones was conducted using multiple comparisons. Due to the unequal sample size and uneven variance in each zone, Tamhane'T2 method was used for multiple comparison among different zones. The zoning process was conducted using MATLAB, and all analyses were performed on IBM SPSS Statistics 20.

### B. $R_H$ estimation in each zone

#### 1) Selection of predictive variables for $R_H$ estimation in each zone

Variables related to temperature, moisture and vegetation productivity were derived from remote sensing data and employed to establish a set of predictor variables that captured the temporal and spatial variations in global $R_H$. Temperature variables encompassed surface temperatures during both daytime (LST_day) and nighttime (LST_night). ET and LSWI served as indicators of soil moisture status. Variables associated with vegetation productivity included GPP, EVI, FPAR and LAI. To gain further insight into the site conditions throughout the observation year, we also selected the maximum, minimum and mean values of these parameters during the observation year as additional predictor variables.

Related studies have demonstrated that $R_H$ exhibits distinct seasonal fluctuations [34, 35]. Thus, we used DOY as a predictor variable to estimate $R_H$. However, it was observed that the DOY variable, when left untreated, exhibited significant differences between the first day of the year (DOY = 1) and the last (DOY = 365). This disparity introduced errors to the modelling because, on DOYs 1 and 365, temperature, moisture, and vegetative growth status were similar at each site, leading to comparable $R_H$ values. For this reason, a sine function was employed to transform DOY.

#### 2) Establishing the data-driven RF model for $R_H$ estimation in each zone

The RF algorithm was applied for $R_H$ modelling. RF is a bagging-integrated machine learning algorithm. It uses decision trees as the base learner, with the final output determined by majority voting on the classification results of each decision tree [36]. The method is algorithmically simple, easy to implement, supports parallel processing, and can effectively reduce overfitting during the modelling process. In the present study, the RF model was trained in MATLAB. To assess the performance of the model, the Leave-One-Site-Out Cross-Validation method was used to ensure that the model was able to predict completely unknown locations. This method set data from one site at a time for validation while the remaining sites are used for training. The two commonly used error metrics, root mean square error (RMSE) and coefficient of determination ($R^2$) were used for model accuracy evaluation.

## IV. Results and Discussions

### A. Environmentally similar zones of global $R_H$ sites

Considering the spatial distribution and environmental characteristics of the in-situ $R_H$ sites, and to ensure a certain amount of data in each zone, we conducted several experiments that allowed us to divide the global in-situ $R_H$ sites into four zones based on the enhanced *k*-means cluster algorithm. The numbers of sites in the four zones were 9, 40, 37 and 81, and the corresponding number of $R_H$ observations were 96, 871, 1052 and 1635, respectively. Their global spatial distribution was shown in Fig. 2.

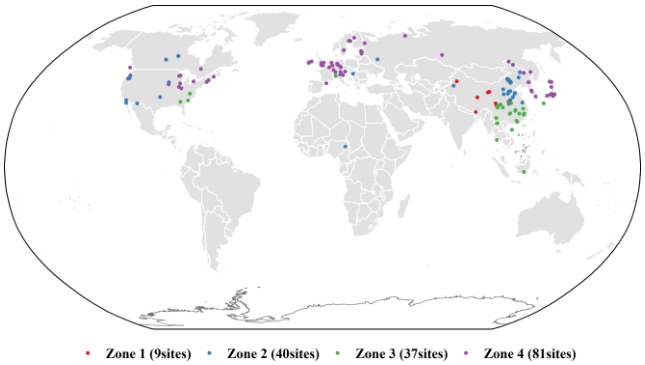


Fig. 2. Locations of soil heterotrophic respiration in-situ sites according to the four environmentally similar zones.

The environmental characteristics of each zone were further analysed. It was evident that all six environmental factors exhibited significant differences across the four zones (Fig. 3). Zone 1 was characterized by having the lowest TEM and highest DEM, and was predominantly composed of GRA sites (Fig. S1a). Zone 2 mainly included GRA and CRO sites (Fig. S1b), and was distinguished by having the lowest PRE, LSWI and SOC. In Zone 3, there was a predominant presence of MF and CRO sites (Fig. S1c), marked by having the highest PRE and TEM. Zone 4 had the highest NDVI, LSWI and SOC, and consisted mainly of forest sites (ENF, DBF and MF), as illustrated in Fig. S1d.

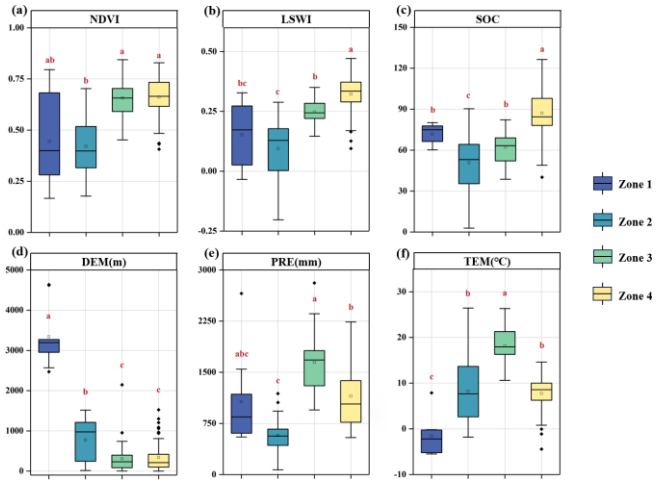


Fig. 3. Comparisons of the six environmental factors across the four zones of similarity: (a) NDVI, (b) LSWI, (c) SOC, (d) DEM, (e) PRE and (f) TEM.

### *B. Modeling of $R_H$ in different zones*

A separate $R_H$ model was constructed for each environmentally similar zone. To avoid the impact of inter-variable collinearity on model performance, Pearson coefficients ($R$) of the correlations between variables were calculated. Variables with extremely strong correlations ($R$ > 0.9) were then removed. The remaining variables were entered into the RF model and their importance calculated. Based on the ranking of variable importance, we sequentially eliminated variables with low importance to determine the final set of variables.

The optimal variable set for each zone was shown in Table I. Leave-One-Site-Out Cross-Validation for each zone showed that our zone modelling method explained 43–66% of the observed $R_H$, with RMSEs of 0.66–0.84 g C $m^{-2}$ $d^{-1}$ (Fig. 4). Importantly, these results were better than those obtained through the globally unified modelling method (Fig. S2), in which $R^2$ ranged from 0.30 to 0.56 with RMSEs of 0.80–0.91 g C $m^{-2}$ $d^{-1}$ across zones. The global-scale validation results also substantiated that our zone modelling method outperformed the global unified modelling method (Fig. S3). These results suggested that our zone approach better reflected spatial distribution differences and provided global $R_H$ estimates with superior accuracy and reliability.

TABLE I
OPTIMAL INPUT VARIABLES USED TO ESTIMATE SOIL HETEROTROPHIC RESPIRATION IN EACH ZONE.

| Zone | Number of variables | Variables |
|---|---|---|
| 1 | 7 | LST_day_max, LSWI, LSWI_min, ET_max, GPP, EVI, DOY |
| 2 | 7 | LST_day, LST_night, LSWI, LSWI_min, EVI, EVI_max, DOY |
| 3 | 9 | LST_day_max, LST_night, LSWI_mean, ET, GPP_max, EVI_max, EVI_mean, NDVI_max, LAI |
| 4 | 8 | LST_day_mean, LST_night, LST_night_mean, LSWI, ET_mean, GPP, GPP_min, GPP_mean |

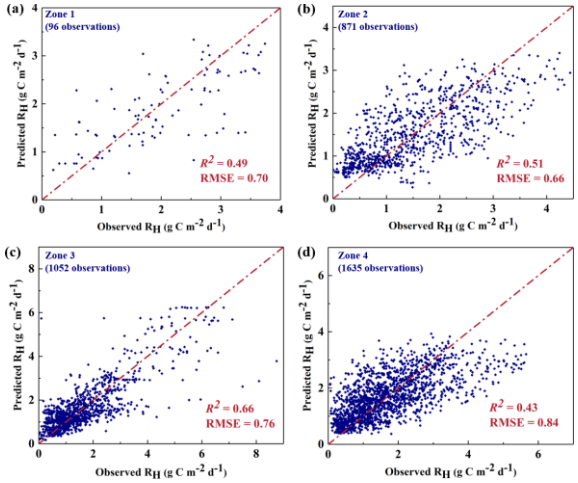


Fig. 4. Comparisons of observed soil heterotrophic respiration ($R_H$) and predicted $R_H$ by the zone modelling method for Zones 1–4 (a-d).

### *C. Variable importance analyses for each zone*

The importance of the variables selected for each zone was illustrated in Fig. 5. Notably, remote sensing variables such as GPP, EVI and LAI, which are known to characterize vegetation productivity [37], exhibited high importance in each zone. This is expected, as vegetation serves as the organic carbon source through fine roots and leaves, providing the main carbon substrate for $R_H$ [38]. In Zone 1, GPP had the highest importance. Previous research has indicated that vegetation-related variables explain most $R_H$ changes in low-temperature and high-altitude grasslands [39, 40]. In Zone 2, LSWI_min had the highest importance. It has been demonstrated that carbon cycling processes in water-limited areas are highly dependent on precipitation variability, with which soil water content is strongly correlated [41]. Additionally, temperature was considered to act in conjunction with soil moisture to influence $R_H$, rendering LST_night a variable of high importance in these regions [42]. LSWI_mean and LST_night had substantial roles in Zone 3, which can be explained by the fact that in humid and hot regions, soil temperature and soil water content significantly affect the seasonal variation in $R_H$ [43]. In Zone 4, temperature-related variables (LST_night_mean and LST_night) showed the highest significance. This result was consistent with previous studies that demonstrated a correlation between $R_H$ and temperature in temperate forests [44, 45].

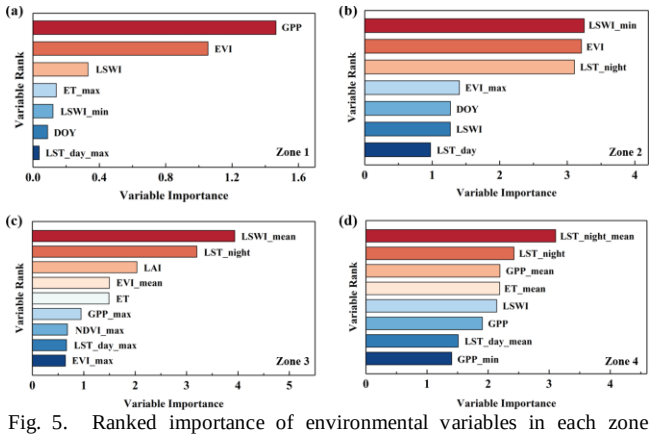


Fig. 5. Ranked importance of environmental variables in each zone of similarity.

### *D. Comparison with Previous Studies*

Compared with previous studies on estimating global $R_H$, the zone model employed in this study had notable advantages. The utilization of remote sensing products with a spatial resolution of 500-m or 1-km allowed land surface information to be expressed at a finer resolution than that of previous studies, which have been based on 0.5° × 0.5° meteorological data [16-18, 22-24]. Temporally, most current global $R_H$ estimates have used annual $R_H$ observations for model construction and validation [16-18, 22-24]. However, more frequent observations are critical for understanding sub-annual $R_H$ variability and the distribution of extremes at temporal scales [46].

Methodologically, we adopted a partitioned modelling approach based on environmental similarity partitioning. Unlike previous data-driven $R_H$ estimations that assumed a globally uniform relationship between $R_H$ and influencing factors [16-18, 22-24], we developed distinct RF model for each environmentally similar zone. This modelling approach effectively addressed the issue of data-driven models being strongly influenced by the spatial representation of the in-situ sites and achieved a model efficiency of 53% (Fig. S3a). Moreover, in comparison with previous studies employing Ten-Fold or Five-Fold Cross-Validation methods for validation [23, 24], this study adopted the Leave-One-Site-Out Cross-Validation method. This validation method enhanced the reliability of our model by ensuring that data from the same site were not simultaneously used for both training and validation. Yao, et al. [22] also utilized the Leave-One-Site-Out Cross-Validation for validation, but their achieved $R^2$ of 0.38 was notably lower than that of our study. Our estimation method contributed to a more accurate representation of global $R_H$.

### *E. Uncertainties*

There were some uncertainties in this study. Firstly, field $R_H$ measurements are inadequately sampled in both time and space, which hampers the ability to upscale $R_H$ data from regional to global scales. Distinguishing $R_H$ from $R_S$ remains challenging with current observational methods [47]. Thus, most current studies on $R_S$ do not consider $R_H$ separately, further contributing to the scarcity of global-scale $R_H$ observations [9]. In Zone 1, only 96 observations from 9 in-situ sites were utilized for modelling, and this small sample size may introduce bias to the $R_H$ estimation [10]. Moreover, in-situ $R_H$ sites lack the capability to provide long-term and continuous observations unlike eddy flux network, such that the understanding of precise rates of change in $R_H$ over various timescales remains limited. $R_H$ observations used for modelling in this study were temporally unevenly distributed, with more observations derived from the warm periods (Figure S4). Meanwhile, within a given region, the controlling factors of $R_H$ may vary seasonally [48], but this variability was not considered in this study. Therefore, there is a strong need for datasets with more observations in both time and space to provide a more detailed description of the spatial and temporal variability of $R_H$ and facilitate better training of data-driven models.

Secondly, the reliability of the estimates made in this study was influenced to some extent by the quality of the remote sensing data used. It is crucial to note that remote sensing products such as MODIS and PML (used to obtain EVI, LSWI and ET) are limited by factors such as weather and the sensors and inversion models used [27], and they may not exhibit complete continuity in both time and space. These discontinuities may result in the true state of the surface not being fully captured when characterizing it with remote sensing products. Based on the constructed in-situ $R_H$ observation database, temperature-related variables inverted from remote sensing data were found to be significantly correlated with TEM in each zone (Table SI). Similarly, LSWI, which was used to characterize soil water content in this study, was significantly correlated with PRE across four zones (Table SI). However, it's noteworthy that their correlation coefficient was relatively low, suggesting that remote sensing data still has some uncertainty in capturing surface water conditions.

## V. Conclusion

This study developed a novel data-driven method for estimating global $R_H$. It was based on grouping worldwide in-situ $R_H$ sites into four environmentally similar zones using an enhanced *k*-means clustering algorithm. In each zone, we established an RF model to estimate 8-day $R_H$ based entirely on remote sensing variables. The zone modelling method significantly improved the accuracy of $R_H$ predictions at both global and zonal scales, with less bias than that of the global unified modelling. It is now possible to accurately estimate global $R_H$ at high spatial and temporal resolutions. Additionally, we expect that our zone modeling method to inspire other in-situ observations-based flux modelling efforts.

## References

[1] S. J. Basile, X. Lin, W. R. Wieder, M. D. Hartman, and G. Keppel-Aleks, "Leveraging the signature of heterotrophic respiration on atmospheric $CO_2$ for model benchmarking," *Biogeosciences*, vol. 17, no. 5, pp. 1293-1308, Mar. 2020, doi: 10.5194/bg-17-1293-2020.

[2] C. Le Quéré *et al.*, "Global Carbon Budget 2018," *Earth Syst. Sci. Data*, vol. 10, no. 4, pp. 2141-2194, Dec. 2018, doi: 10.5194/essd-10-2141-2018.

[3] B. Bond-Lamberty, V. L. Bailey, M. Chen, C. M. Gough, and R. Vargas, "Globally rising soil heterotrophic respiration over recent decades," *Nature*, vol. 560, no. 7716, pp. 80-+, Aug. 2018, doi: 10.1038/s41586-018-0358-x.

[4] S. N. Bhanja and J. Y. Wang, "Estimating influences of environmental drivers on soil heterotrophic respiration in the Athabasca River Basin, Canada," *Environ. Pollut.*, vol. 257, Feb. 2020, Art no. 113630, doi: 10.1016/j.envpol.2019.113630.

[5] N. J. Noh *et al.*, "Responses of Soil, Heterotrophic, and Autotrophic Respiration to Experimental Open-Field Soil Warming in a Cool-Temperate Deciduous Forest," *Ecosystems*, vol. 19, no. 3, pp. 504-520, Apr. 2016, doi: 10.1007/s10021-015-9948-8.

[6] B. Bond-Lamberty and A. Thomson, "Temperature-associated increases in the global soil respiration record," *Nature*, vol. 464, no. 7288, pp. 579-U132, Mar. 2010, doi: 10.1038/nature08930.

[7] J. W. Raich and C. S. Potter, "Global patterns of carbon dioxide emissions from soils," *Glob. Biogeochem. Cycles*, vol. 9, no. 1, pp. 23-36, Mar. 1995, doi: 10.1029/94GB02723.

[8] B. Bond-Lamberty and A. Thomson, "A global database of soil respiration data," *Biogeosciences*, vol. 7, no. 6, pp. 1915-1926, 2010, doi: 10.5194/bg-7-1915-2010.

[9] J. Jian *et al.*, "A restructured and updated global soil respiration database (SRDB-V5)," *Earth Syst. Sci. Data*, vol. 13, no. 2, pp. 255-267, Feb. 2021, doi: 10.5194/essd-13-255-2021.

[10] J. Jian, M. K. Steele, R. Q. Thomas, S. D. Day, and S. C. Hodges, "Constraining estimates of global soil respiration by quantifying sources of variability," *Glob. Change Biol.*, vol. 24, no. 9, pp. 4143-4159, Sep. 2018, doi: 10.1111/gcb.14301.

[11] J. S. Jian, M. K. Steele, S. D. Day, and R. Q. Thomas, "Future Global Soil Respiration Rates Will Swell Despite Regional Decreases in Temperature Sensitivity Caused by Rising Temperature," *Earths Future*, vol. 6, no. 11, pp. 1539-1554, Nov. 2018, doi: 10.1029/2018ef000937.

[12] B. Bond-Lamberty *et al.*, "COSORE: A community database for continuous soil respiration and other soil-atmosphere greenhouse gas flux data," *Glob. Change Biol.*, vol. 26, no. 12, pp. 7268-7283, Dec. 2020, doi: 10.1111/gcb.15353.

[13] X. Xu *et al.*, "Soil properties control decomposition of soil organic carbon: Results from data-assimilation analysis," *Geoderma*, vol. 262, pp. 235-242, Jan. 2016, doi: 10.1016/j.geoderma.2015.08.038.

[14] B. Bond-Lamberty, C. K. Wang, and S. T. Gower, "A global relationship between the heterotrophic and autotrophic components of soil respiration?" *Glob. Change Biol.*, vol. 10, no. 10, pp. 1756-1766, Oct. 2004, doi: 10.1111/j.1365-2486.2004.00816.x.

[15] J. A. Subke, I. Inglima, and M. F. Cotrufo, "Trends and methodological impacts in soil $CO_2$ efflux partitioning: a meta-analytical review," *Glob. Change Biol.*, vol. 12, no. 9, pp. 1813-1813, Sep. 2006, doi: 10.1111/j.1365-2486.2006.01218.x.

[16] S. Hashimoto, N. Carvalhais, A. Ito, M. Migliavacca, K. Nishina, and M. Reichstein, "Global spatiotemporal distribution of soil respiration modeled using a global database," *Biogeosciences*, vol. 12, no. 13, pp. 4121-4132, 2015, doi: 10.5194/bg-12-4121-2015.

[17] D. L. Warner, B. Bond-Lamberty, J. Jian, E. Stell, and R. Vargas, "Spatial Predictions and Associated Uncertainty of Annual Soil Respiration at the Global Scale," *Glob. Biogeochem. Cycles*, vol. 33, no. 12, pp. 1733-1745, Dec. 2019, doi: 10.1029/2019gb006264.

[18] E. Stell, D. Warner, J. Jian, B. Bond-Lamberty, and R. Vargas, "Spatial biases of information influence global estimates of soil respiration: How can we improve global predictions?" *Glob. Change Biol.*, vol. 27, no. 16, pp. 3923-3938, May 2021, doi: 10.1111/gcb.15666.

[19] H. Chu, D. D. Baldocchi, R. John, S. Wolf, and M. Reichstein, "Fluxes all of the time? A primer on the temporal representativeness of FLUXNET," *J. Geophys. Res.-Biogeosci.*, vol. 122, no. 2, pp. 289-307, Feb. 2017, doi: 10.1002/2016jg003576.

[20] M. Jung *et al.*, "Compensatory water effects link yearly global land $CO_2$ sink changes to temperature," *Nature*, vol. 541, no. 7638, pp. 516-520, Jan. 2017, doi: 10.1038/nature20780.

[21] C. Beer *et al.*, "Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate," *Science*, vol. 329, no. 5993, pp. 834-838, Aug. 2010, doi: 10.1126/science.1184984.

[22] Y. T. Yao *et al.*, "A Data-Driven Global Soil Heterotrophic Respiration Dataset and the Drivers of Its Inter-Annual Variability," *Glob. Biogeochem. Cycles*, vol. 35, no. 8, Aug. 2021, Art no. e2020GB006918, doi: 10.1029/2020gb006918.

[23] X. L. Tang *et al.*, "Spatial and temporal patterns of global soil heterotrophic respiration in terrestrial ecosystems," *Earth Syst. Sci. Data*, vol. 12, no. 2, pp. 1037-1051, May 2020, doi: 10.5194/essd-12-1037-2020.

[24] H. B. Lu *et al.*, "Comparing machine learning-derived global estimates of soil respiration and its components with those from terrestrial ecosystem models," *Environ. Res. Lett.*, vol. 16, no. 5, May 2021, Art no. 054048, doi: 10.1088/1748-9326/abf526.

[25] N. Huang *et al.*, "Spatial and temporal variations in global soil respiration and their relationships with climate and land cover," *Sci. Adv.*, vol. 6, no. 41, Oct. 2020, Art no. eabb8508, doi: 10.1126/sciadv.abb8508.

[26] E. Masuoka, A. Fleig, R. E. Wolfe, and F. Patt, "Key characteristics of MODIS data products," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 4, pp. 1313-1323, Jul. 1998, doi: 10.1109/36.701081.

[27] Y. Q. Zhang *et al.*, "Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002-2017," *Remote Sens. Environ.*, vol. 222, pp. 165-182, Mar. 2019, doi: 10.1016/j.rse.2018.12.031.

[28] S. L. Liang *et al.*, "The Global Land Surface Satellite (GLASS) Product Suite," *Bull. Amer. Meteorolog. Soc.*, vol. 102, no. 2, pp. E323-E337, Feb. 2021, doi: 10.1175/bams-d-18-0341.1.

[29] X. L. Tang *et al.*, "Global patterns of soil heterotrophic respiration–A meta-analysis of available dataset," *Catena*, vol. 191, p. 104574, Aug. 2020, doi: 10.1016/j.catena.2020.104574.

[30] M. F. Yan, N. Guo, H. R. Ren, X. S. Zhang, and G. S. Zhou, "Autotrophic and heterotrophic respiration of a poplar plantation chronosequence in northwest China," *Forest Ecol. Manage.*, vol. 337, pp. 119-125, Feb. 2015, doi: 10.1016/j.foreco.2014.11.009.

[31] D. Yamazaki *et al.*, "A high-accuracy map of global terrain elevations," *Geophys. Res. Lett.*, vol. 44, no. 11, pp. 5844-5853, Jun. 2017, doi: 10.1002/2017gl072874.

[32] M. Meilă, "The uniqueness of a good optimum for k-means," *Proc. ICML*, pp. 625-632, 2006.

[33] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proc. Int. Conf. Neural Netw. (ICNN)*, vol. 4, pp. 1942-1948, 1995.

[34] J. W. Tang and D. D. Baldocchi, "Spatial-temporal variation in soil respiration in an oak-grass savanna ecosystem in California and its partitioning into autotrophic and heterotrophic components," *Biogeochemistry*, vol. 73, no. 1, pp. 183-207, Mar. 2005, doi: 10.1007/s10533-004-5889-6.

[35] S. T. Chen *et al.*, "Modeling interannual variability of global soil respiration from climate and soil properties," *Agric. Forest Meteorol.*, vol. 150, no. 4, pp. 590-605, Apr. 2010, doi: 10.1016/j.agrformet.2010.02.004.

[36] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5-32, Oct. 2001, doi: 10.1023/a:1010933404324.

[37] N. Huang, L. H. Gu, T. A. Black, L. Wang, and Z. Niu, "Remote sensing-based estimation of annual soil respiration at two contrasting forest sites," *J. Geophys. Res.-Biogeosci.*, vol. 120, no. 11, pp. 2306-2325, Nov. 2015, doi: 10.1002/2015jg003060.

[38] Y. Luo and X. Zhou, "Soil Respiration and the Environment," *Academic Press*, 2006.

[39] Y. Geng *et al.*, "Soil Respiration in Tibetan Alpine Grasslands: Belowground Biomass and Soil Moisture, but Not Soil Temperature, Best Explain the Large-Scale Patterns," *PLoS ONE*, vol. 7, no. 4, Apr. 2012, Art no. e34968, doi: 10.1371/journal.pone.0034968.

[40] N. Huang, J. S. He, and Z. Niu, "Estimating the spatial pattern of soil respiration in Tibetan alpine grasslands using Landsat TM images and MODIS data," *Ecological Indicators*, vol. 26, pp. 117-125, Mar. 2013, doi: 10.1016/j.ecolind.2012.10.027.

[41] X. H. Zhou, M. Talley, and Y. Q. Luo, "Biomass, Litter, and Soil Respiration Along a Precipitation Gradient in Southern Great Plains, USA," *Ecosystems*, vol. 12, no. 8, pp. 1369-1380, Dec. 2009, doi: 10.1007/s10021-009-9296-7.

[42] L. M. Lai *et al.*, "Soil Respiration in Different Agricultural and Natural Ecosystems in an Arid Region," *PLoS ONE*, vol. 7, no. 10, Oct. 2012, Art no. e48011, doi: 10.1371/journal.pone.0048011.

[43] Y. D. Wang *et al.*, "Precipitation frequency controls interannual variation of soil respiration by affecting soil moisture in a subtropical forest plantation," *Can. J. Forest Res.*, vol. 41, no. 9, pp. 1897-1906, Sep. 2011, doi: 10.1139/x11-105.

[44] Z. Y. Zhou, M. L. Xu, F. F. Kang, and O. J. Sun, "Maximum temperature accounts for annual soil $CO_2$ efflux in temperate forests of Northern China," *Sci. Rep.*, vol. 5, Jul. 2015, Art no. 12142, doi: 10.1038/srep12142.

[45] B. Klimek, M. Chodak, and M. Niklinska, "Soil respiration in seven types of temperate forests exhibits similar temperature sensitivity," *J. Soils Sediments*, vol. 21, no. 1, pp. 338-345, Jan. 2021, doi: 10.1007/s11368-020-02785-y.

[46] J. S. Jian *et al.*, "Leveraging observed soil heterotrophic respiration fluxes as a novel constraint on global-scale models," *Glob. Change Biol.*, vol. 27, no. 20, pp. 5392-5403, Oct. 2021, doi: 10.1111/gcb.15795.

[47] Y. Kuzyakov, "Sources of $CO_2$ efflux from soil and review of partitioning methods," *Soil Biol. Biochem.*, vol. 38, no. 3, pp. 425-448, Mar. 2006, doi: 10.1016/j.soilbio.2005.08.020.

[48] J. Jian, M. K. Steele, S. D. Day, R. Q. Thomas, and S. C. Hodges, "Measurement strategies to account for soil respiration temporal heterogeneity across diverse regions," *Soil Biol. Biochem.*, vol. 125, pp. 167-177, Oct. 2018, doi: 10.1016/j.soilbio.2018.07.003.

**Luying Zhu** received the bachelor's degree in geochemistry from China University of Geosciences (Beijing) in 2018. She is currently studying for the Ph.D. degree in cartography and geographic information system at the Institute of Aerospace Information, Chinese Academy of Sciences, Beijing, China.

Her current research interest is ecological remote sensing.

**Ni Huang** received the Ph.D. degree in GIS and remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, China, 2012. She obtained her master's degree in GIS and remote sensing from Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, China, 2009.

She is a research scientist in carbon cycle remote sensing study in the Aerospace Information Research Institute, Chinese Academy of Sciences. Her research was focused on estimating carbon cycle parameters based on remote sensing data.

**Li Wang** received the Ph.D. degree from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS) in 2008.

He is a research scientist in the Aerospace Information Research Institute, Chinese Academy of Sciences. He has authored or co-authored more than 100 peer-reviewed papers. His research interest includes the fields of land-use and land-cover change, with a focus on remote sensing technologies and their application in eco-environment change (mapping and monitoring).

**Zheng Niu** received his Ph.D. degree in physical geography from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academic of Sciences, China, 1996. He obtained his master's degree in GIS and remote sensing from Peking University, China, 1991.

He is a professor in carbon cycle remote sensing study in the Aerospace Information Research Institute, Chinese Academy of Sciences. His expertise is in the fields of remote sensing related to global change ecosystem.

**Jinxiao Wang** received Ph.D. degree in cartography and geographic information system at the Institute of Aerospace Information, Chinese Academy of Sciences, Beijing, China, 2023. He obtained the bachelor's degree in remote sensing science and technology from China University of Geosciences (Wuhan), Wuhan, China, 2018.

He is currently a research associate in Key Laboratory of Network Information System Technology, Aerospace Information Research Institute, Chinese Academy of Sciences. His research interest is intelligent mining of remote sensing big data.

**Yuelin Zhang** received a bachelor degree in Physical Geography and Resource Environment from the Hebei Agricultural University, China, 2019.

He is a researcher in the Aerospace Information Research Institute, Chinese Academy of Sciences. His main work is remote sensing data processing.

**Jie Liu** received the bachelor's degree in geographic information science from China University of Geosciences (Beijing) in 2022. She is currently studying for the Ph.D. degree in cartography and geographic information system at the institute of Aerospace Information, China Academy of Sciences, Beijing, China.

Her research interest is ecological remote sensing.