

An improved estimate of soil carbon pool and carbon fluxes in the Qinghai-Tibetan grasslands using data assimilation with an ecosystem biogeochemical model



Ruiying Zhao ^{a,b}, Wenxin Zhang ^b, Zheng Duan ^b, Songchao Chen ^{a,c}, Zhou Shi ^{a,*}

^a College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, China

^b Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

^c ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou, China

ARTICLE INFO

Handling Editor: Budiman Minasny

Keywords:

Soil organic carbon
Ecosystem model
Qinghai-Tibet Plateau
Data assimilation
Climate change

ABSTRACT

The accurate estimation of soil carbon (C) pool and fluxes is a prerequisite to better understand the terrestrial C feedback to climate change. However, recent studies showed considerable uncertainties in soil C estimates. To provide a reliable C estimate in the grasslands of the Qinghai-Tibet Plateau (QTP), we calibrated key parameters in a process-based ecosystem model (the CENTURY model) through data assimilation based on 570 soil samples and 21 sites of eddy covariance measurements. Two assimilating strategies (Opt1 – assimilating C pool observations; Opt2 –assimilating both C pool and C flux) were examined. Compared to default parameterization, our results showed both Opt1 and Opt2 improved the soil organic carbon density (SOCD) estimation, with R^2 increasing from 0.59 to 0.75 and 0.73, respectively. Opt2 was superior to Opt1 in constraint of parameters dominating aboveground processes and yield a better estimation of net ecosystem production (NEP). Based on different parameterization, the spatial variability of SOCD and NEP across the QTP grassland were generated. Both Opt1 and Opt2 ameliorated the overestimation of SOCD by the default model, estimating a total soil C of 6.63 Pg and 6.48 Pg C for the topsoil (0–30 cm) of the QTP grasslands, respectively. Opt2 showed lower uncertainties in the NEP estimation and predicted a net sink of 14.33 Tg C annually. Compared with existing datasets, our study provided a more reliable estimation of carbon storage and fluxes in the QTP grassland with the calibrated ecosystem model. The results highlight that data assimilation with multiple observational data sets is promising to constrain process-based ecosystem models and increase the robustness of model predictions for terrestrial C cycle feedback to future climate change.

1. Introduction

More carbon (C) stays in soil than in the atmosphere and vegetation combined (Lal, 2004; Ciais et al., 2014). Even relatively small changes of soil organic carbon (SOC) in response to rising temperature could exacerbate global climate change due to substantial amounts of greenhouse gases released into the atmosphere (Knorr et al., 2005; Pries et al., 2017). It has been documented that the effect of warming on soil C strongly depends on the initial soil C pool size and its spatial heterogeneity (Crowther et al., 2016; Luo et al., 2017). Therefore, a robust estimate of soil C stocks and their spatial variability is essential to improve our understanding of ecosystem C dynamics and predictions of terrestrial C-cycle feedback to climatic warming.

Cold regions where a tremendous amount of SOC has been stored due to slow soil decomposition rates under long-term cold environmental conditions have recently attracted increasing attention in the context of climatic warming. The Qinghai-Tibet Plateau (QTP), known as ‘The Third Pole’ of the Earth, is the largest high-altitude plateau with an average elevation of over 4000 m a.s.l. and an area of 2.5 million km². Alpine grasslands, including meadow and steppe, are the dominant ecosystems in the QTP, covering nearly 60% of the entire plateau region (Tan et al., 2010; Li et al., 2018). Despite large uncertainties in existing estimates, the SOC stock in the QTP grasslands is estimated over half of the total C in the China’s grassland soils (Wu et al., 2003; Yang et al., 2010; Piao et al., 2012; Zhang et al., 2016). The majority of SOC in the QTP is projected to be stored in the uppermost 30 cm, which accounts

* Corresponding author.

E-mail address: shizhou@zju.edu.cn (Z. Shi).

for over half of the total SOC within the 0–100 cm depth interval (Wang et al., 2021; Yang et al., 2008). Nevertheless, many lines of evidence have demonstrated that such rich C in QTP is being exposed to a warming rate twice the global average (Wang et al., 2008; Piao et al., 2010; You et al., 2016). As the QTP grassland holds a high density of soil C and is more sensitive to climate change, it plays an important role in C cycles on regional and even national scales (Wu et al., 2003; Piao et al., 2012). Therefore, it is vital to accurately quantify the size of SOC storage and spatial patterns across this region.

SOC estimates in earlier studies were mainly based on the stratification method. This method has generated many legacy distribution maps of SOC by stratifying study area according to specific land cover classifications such as vegetation or soil types, averaging point observations of SOC values within each stratum, and finally multiplying the areal extent of the stratum (Batjes, 1996, 2016; Nachtergael et al., 2008; Tang et al., 2018). However, this approach has been reported to be associated with high estimation errors due to a large heterogeneity within each stratum (Meersmans et al., 2008). Empirical models are alternative approaches to estimate the soil C spatial variability. The general idea behind empirical models is to develop a quantitative relationship between soil observations and spatial referenced environmental covariates based on various prediction-oriented regression algorithms such as multiple linear regression and machine learning algorithms (McBratney et al., 2003; Sanchez et al., 2009; Chen et al., 2022). Although empirical models have shown great power for mapping fine-resolution spatial distribution of SOC across scales (Viscarra Rossel et al., 2014; Hengl et al., 2017; Liang et al., 2019), such data-driven approaches strongly depend on the availability of observations. Whereas, due to the harsh environment, many environmental variables and soil measurements are not readily available or uneven distributed in the QTP, resulting in large uncertainties in soil C estimation (Zhou et al., 2019). Moreover, without accounting for ecosystem dynamics and biogeochemical processes, both stratification methods and empirical models have limitations in simulating temporal variations of soil C pools and their association with ecosystem C uptake and soil decomposition. Therefore, these two approaches are mostly used to provide static maps of SOCD.

Alternatively, process-based ecosystem biogeochemistry models such as the CENTURY model, have been widely used to estimate soil C storage and fluxes by explicitly representing C flows and pools in ecosystems. This approach has shown advantages in simulating site, regional and even global scale of C dynamics (Parton et al., 1988; Todd-Brown et al., 2013; Tian et al., 2015; Berardi et al., 2020). Especially in the warming world, process-based models are capable to disentangle the effects of warming on C uptake and release, estimating a more reasonable magnitude and spatial patterns for C pools at any specific time (Lugato et al., 2014; Todd-Brown et al., 2014; Kern et al., 2018; Huang et al., 2018). However, in process-based models, C dynamics are regulated by a large number of parameters related to photosynthesis, biomass allocation, decomposition, etc. These parameters are mostly determined on an *ad hoc* basis and may be derived from the results of specific field experiments, modelling references or expert knowledge (Luo et al., 2001). It is largely untested whether or not the model parameterization is representative of the system properties to be simulated (Tao et al., 2020; Huang et al., 2021), leading to considerable uncertainty in prediction (Xiao et al., 2014; Luo et al., 2015).

Data assimilation (DA), a statistically rigorous methodology to integrate observations and models, has been widely used to calibrate model parameters (Luo et al., 2011; Niu et al., 2014; Hou et al., 2019; Ge et al., 2019). Among various algorithms for DA application, the Bayesian Markov chain Monte Carlo (MCMC) has attracted much attention in parameter estimation and uncertainty evaluation. For one thing, as a global optimization method, MCMC is more suitable to deal with high-dimensional nonlinear optimization and avoid trapping into a local optimum (Gallagher et al., 2009; Lu et al., 2017). For another, MCMC can accurately quantify SOC simulation uncertainty because of model

parameters with informative posterior distribution (Huang et al., 2021). Numerous studies have demonstrated that the application of the MCMC data assimilation method can provide confidence ranges of model parameters and improve the performance of process-based models for SOC simulation (Hararuk et al., 2014; Tao et al., 2020). More recently, different types of observations generated from numerous sources, such as field sampling, eddy covariance measurements and remote sensing, have become available for MCMC data assimilation to calibrate model parameters. Simultaneously assimilating multiple data has shown superiority, because different data streams may contain information about different processes and would be powerful to constrain relevant parameters tightly (Keenan et al., 2013; Du et al., 2015). Despite the advantages of assimilation of multiple data to constrain C stream on the site level (Richardson et al., 2010), very few studies have been carried out at the regional scale and these studies have rarely explored the impact of different types of observations on specific model parameters and processes.

As data assimilation largely relies on the availability of measurements in both space and time, our study for the first time constrains the process-based model based on the largest collection of soil C samples and CO₂ flux in QTP. We collected 570 soil samples and 21 eddy covariance (EC) sites across the QTP grasslands. Based on these datasets, we applied the data assimilation approach to the CENTURY model to calibrate C storage and net ecosystem productivity. The CENTURY is a widely-used ecosystem biogeochemical model and the stereotype of soil models in many land surface models, such as CLM5 (Lawrence et al., 2019), ORCHIDEE (Krinner et al., 2005), CABLE (Wang et al., 2011) and dynamic global vegetation models, such as LPJ-GUESS (Smith et al., 2014), our work can bring much insight to guide the land surface model calibration and uncertainty analysis. Besides, our simulations also demonstrate the advances in the moderate resolution modelling (i.e., 0.05° × 0.05°), which is much higher than (0.5° × 0.5°), commonly used by many land surface model intercomparison projects (e.g., the TRENDY project, <https://blogs.exeter.ac.uk/trendy/>; the Multi-scale Synthesis and Terrestrial Model Intercomparison Project – MsTMIP, <https://nacp.onrl.gov/MsTMIP.shtml>).

The objectives of this study are to (1) use an MCMC data assimilation to constrain the CENTURY model to simulate SOCD and EC-measured C flux; (2) assess posterior distributions of parameters and their relations to model performance; (3) compare our estimates of SOC with other published SOC datasets. The study has explored an alternative approach to provide a reliable estimation of soil C storage and CO₂ fluxes for QTP grasslands by constraining an ecosystem model using multiple observations. It will provide a tool to make a process-based prediction of the future trajectory of the terrestrial C cycle in response to climate change.

2. Materials and methods

2.1. Study area and observational data sets

2.1.1. Study area

The study area was located in the Qinghai-Tibet Plateau (73°18' E – 104°46' E, 26°00' N – 39°46' N), which is the largest and highest plateau in the world (Fig. 1). The QTP is mainly influenced by the Indian monsoon in summer and the westerlies in winter. With the altitude increasing from southeastern QTP to the northwest, the mean annual precipitation decreases from 2000 mm to below 50 mm, and the temperature drops from 20 °C to –6 °C (Yao et al., 2012). The QTP harbors diverse ecosystems, transiting from forests, grasslands to deserts from the southeast to the northwest. In this study, we focused on the alpine meadow and steppe grasslands, which are two major ecosystems in the QTP, accounting for 27.66% and 30.34% of the entire plateau region, respectively (Tan et al., 2010).

2.1.2. Soil samples

A total of 570 soil samples across the alpine meadow (233 samples)

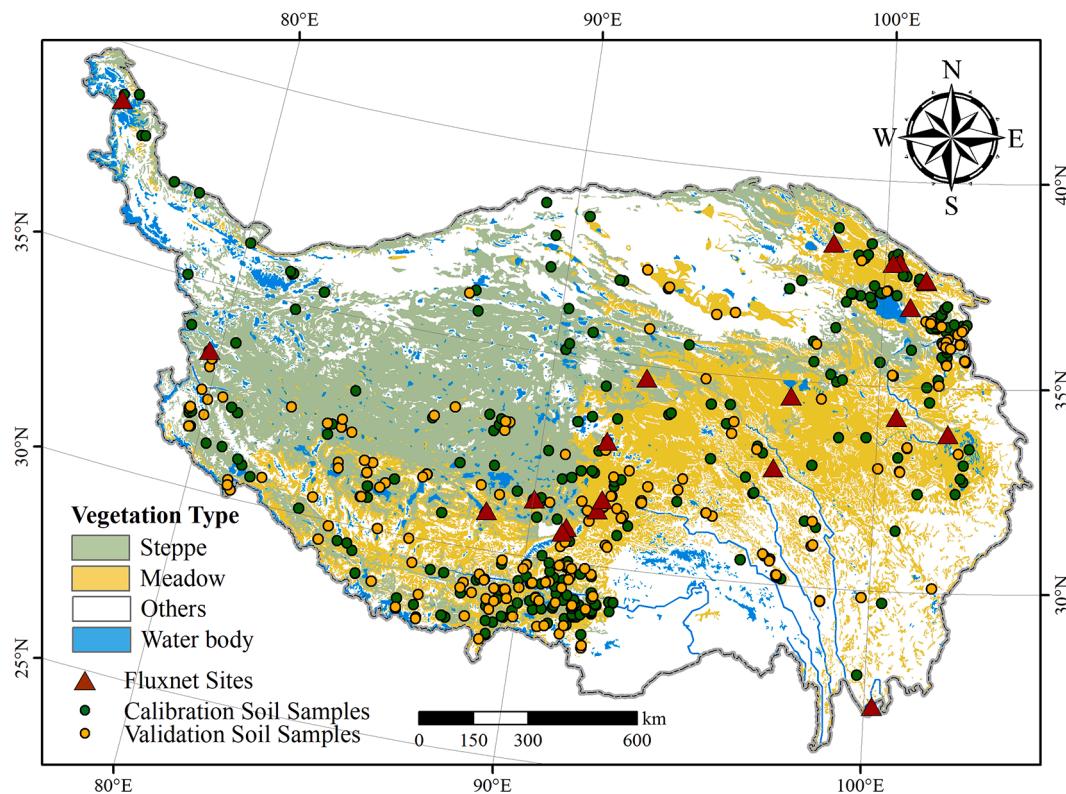


Fig. 1. The locations of soil samples and eddy covariance (EC) sites in the Tibetan Plateau used by this study. The vegetation map is based on the China Vegetation Atlas (<https://www.resdc.cn/data.aspx?DATAID=122>). The details about soil samples and EC measurements can be found in Table S1 and Table S2.

and steppe (337 samples) in the QTP were collected from the Second National Soil Survey of China, which was conducted in the early 1980 s (Li et al., 1992) (Fig. 1). These soil sample records have been documented in a total of 15 books (Table S1). The original soil profile-based records were first transformed to 0–30 cm of soil data using the equal-area smoothing spline method (Malone et al., 2009; Zhou et al., 2019). C density was then calculated based on the following equation,

$$SOCD = SOC \times BD \times H \times \frac{1 - Rv}{100} \quad (1)$$

where, $SOCD$ is soil organic C density (kg m^{-2}) of each sample. SOC , BD and Rv stand for upper layer soil organic C content (g kg^{-1}), bulk density (g cm^{-3}) and gravel content (%), respectively. H is the soil thickness, which is 30 cm in this study.

2.1 Eddy covariance measurements of CO_2 fluxes

Fluxes measurements were obtained from 21 eddy covariance sites (Fig. 1). Among these sites, 16 sites are located in the alpine meadow and 5 sites in the steppe (Table S2). Due to the harsh environment in the western part, the FLUXNET sites are mostly placed in the central and eastern QTP (Fig. 1). Detailed information on EC sites can be found in Wei et al (2021). The original raw data they collected was mainly provided at 30- or 60-min intervals. They used linear interpolation to reconstruct the flux data missing within 2-h intervals. Data missing for more than 2 h were constructed based on the Michaelis-Menten light response relationship (Wei et al., 2021). After gap filling, the flux tower data in each site was aggregated to the monthly total net ecosystem production (NEP, positive means carbon uptake and negative means carbon release) to fit the monthly step of the CENTURY outputs.

We have finally collected a total of 409 site months for meadow ecosystems and 78 site months for steppe ecosystems.

2.2 CENTURY model

2.2.1 Overview

The CENTURY model, initially designed to simulate long-term C changes in grasslands, has become a widely-used process-based model to predict C and nutrient dynamics for different types of ecosystems (Parton et al., 1993). The CENTURY model has also been adopted as the core module of C cycling by many contemporary Earth System Models due to its reasonable multi-compartment framework (Berardi et al., 2020; Boysen et al., 2021). A new version of the CENTURY model (v4.7) was used in this study. This model version has been updated with recent advanced features in representing terrestrial C cycle, for instance, its abiotic response function of soil decomposition accounting for not only temperature and soil moisture but also the C:N ratio and pH.

2.2.2 Model input and configuration

The CENTURY model requires monthly maximum and minimum air temperature and precipitation as climate drivers, and thus the spatial resolution of the simulation is determined by the spatial resolution of climate variables. In this study, the climate variables were taken from a high-resolution (1/24°, ~4 km) monthly climate dataset (TerraClimate) (Abatzoglou et al., 2018, <https://www.climatologylab.org/terraclimate.html>) from 1961 to present. The TerraClimate dataset is developed using three widely-used gridded products, including the WorldClim 2 (Fick and Hijmans, 2017), CRU ts4.0 (Harris et al., 2020) and JRA55 reanalysis data (Kobayashi et al., 2015). The TerraClimate dataset has been well validated with over 3000 station-based observations, with Pearson's correlation coefficient higher than 0.9 for approximately 80% of the stations.

To set up soil properties (i.e. soil texture, pH, bulk density) used in the simulation, we used the 30 × 30 arc-second resolution gridded soil characteristics data set of China (Shangguan et al., 2013). This dataset was constructed based on 8979 soil profiles and specifically used for land surface modelling. In our simulations, the soil data set was

resampled to the same spatial resolution as climate data sets using the bilinear approach.

Before running the historical simulation, a long-term spin-up run was needed to allow the simulated vegetation and soil carbon to reach their equilibrium state. The spin-up run was carried out by using the detrended monthly climate data during 1961–1970 to repeatedly run the CENTURY model for 4000 years. After that, a normal transient simulation from 1961 to 2020 driven by the historical climate dataset started.

2.2.3. Key processes and parameters

In the CENTURY model, ecosystem exchange of CO₂ is mainly regulated by the following ecosystem processes (Figure S1), i.e., photosynthesis, plants decay, soil decomposition and mineralization. First, CO₂ flux enters the biosphere from the atmosphere through plant photosynthesis. The potential production is assumed to be a function of solar radiation and controlled by temperature (PPDF), moisture (P) and CO₂ fertilization effect. After regulated by a 0–1 scalar (PRDX), which is applied to scale the radiation use efficiency on potential production, the assimilated C is then allocated to different compartments of plants according to a dynamic C allocation scheme in which the allocation coefficients vary with soil water and nutrients availability (Poorter and Sack, 2012; De Kauwe et al., 2014). Secondly, when plants die and fall, the above- and below-ground plant residues are partitioned into litter pools. The amount of decayed C is determined by the death rate (FSDETH, RDR) and the fall rate (FALLRATE). Thirdly, through the process of decomposition, C is transformed from the litter C pools to SOM pools and then flows among different SOM pools following a first-order decay function. The theoretical turnover amount of C for each pool is the product of the C pool size (C_i) and maximum decomposition rate (DEC), which is constrained by the effect of temperature (TEFF) and moisture (P). Fourthly, during the decomposition of C pools, organic C is mineralized into CO₂ because of microbial respiration and finally released to the atmosphere. The fraction of C lost to CO₂ is microbe-mediated and depends on the soil texture and relevant parameters (PCO2).

According to the key processes mentioned above, 24 parameters that dominate the C dynamics were identified in Table 1. The default and range of parameters were first assigned based on the field experiments, expert knowledge and previous literature (Wu et al., 2014; Rafique et al., 2014; Necpálová et al., 2015). To constrain model parameters and provide more reliable simulations, the CENTURY model was calibrated through data assimilation, and the details on data assimilation are described in the following section.

2.3. Data assimilation for parameter estimation

2.3.1. The Bayes' theorem

Under the framework of data assimilation, Bayesian probabilistic inversion is applied. The principle of the Bayesian framework is to constrain parameter values and find their posterior probability density functions based on prior knowledge and the bias between model output and observations (Luo et al., 2009; Sanz-Alonso et al., 2018). Bayesian inversion can be expressed by the following formulation,

$$p(\mu|y) = \frac{p(y|\mu)p(\mu)}{p(y)} \quad (2)$$

where, $p(\mu|y)$ represents the posterior probability distributions of parameters μ for a given observed data y . It is calculated from the data likelihood function $p(y|\mu)$, the proposal distribution of the parameters $p(\mu)$ and the normalization constant $p(y)$.

The likelihood function, which is determined by the given model, is negatively related to the cost function J , that is,

$$p(y|\mu) \propto \exp(-J) \quad (3)$$

$$J = \sum_i^n \sum_j^m \frac{(y_{ij} - \hat{y}_{ij})^2}{2\sigma_i^2} \quad (4)$$

The cost function J measures the misfit between the observations y and predictions \hat{y} . In the equation, y_{ij} and \hat{y}_{ij} represent the observed and simulated values of the i^{th} variable at the j^{th} grid, respectively. The parameter n represents the total number of evaluated variables. As we have done two parameter optimizations (Opt1 - optimizing soil C storage and Opt 2 - optimizing both soil C and net ecosystem productivity), we which equals to set n as 1 and 2 for Opt1 and Opt2, respectively. The σ_i is the standard deviation of the i^{th} observation to stand for the measurement uncertainty. The soil samplings from the Second Soil Survey of China have been widely used, but they do not include the uncertainty estimates, which is a crucial aspect of utilizing MCMC methods. To address these uncertainties, we assumed a standard deviation of 30% for the observations at each grid cell (Harmon & Challenor, 1997; Tao et al., 2020).

2.3.2. Markov chain Monte Carlo and convergence test

To obtain the posterior density distribution of parameters, data assimilation repeats two steps (i.e. proposing phase and moving phase) to maximize the likelihood function, i.e., to minimize the cost function. First, during the proposing phase, a sampling algorithm is applied to generate a new random set of parameter values (μ_{new}) from the given range based on default values for the first run or accepted values in the last iteration (μ_{old}). Due to the high efficiency for sampling multi-dimensional distributions, the adaptive Metropolis-Hastings sampling algorithm has been widely applied in previous studies (Hastings, 1970; Hararuk et al., 2014, 2015; Huang et al., 2021).

Second, after sampling, the forward model would be run to obtain current simulations. In the moving phase step, whether the new parameters can be accepted or not would be determined according to the Metropolis criterion (Xu et al., 2006). A probability of acceptance was first calculated as below,

$$p(\mu_{old}|\mu_{new}) = \min \left\{ \frac{p(y|\mu_{new})p(\mu_{new})}{p(y|\mu_{old})p(\mu_{old})}, 1 \right\} \quad (5)$$

If $p(\mu_{old}|\mu_{new})$ is larger than a number randomly selected from a uniform distribution in the range from 0 to 1, the new parameters would be accepted, otherwise, μ_{new} was set to μ_{old} . To monitor the efficiency of MCMC process, the Markov chain with acceptance rate higher than 50% or lower than 15% would be rejected (Gelman et al., 1997; Roberts and Rosenthal, 2001; Tao et al., 2020). Data assimilation will continuously iterate those two steps until the prescribed maximum simulation run is reached.

The proposal distribution (prior distribution) of parameters can strongly determine the efficiency of the adaptive Metropolis-Hastings algorithm (Xu et al., 2006). In this study, two types of proposal distribution were used in the proposing phase. First, a test run of 20,000 simulations using a uniform proposal distribution over the given parameters' range was conducted (Figure S2). According to previous studies (Xu et al., 2006; Hararuk et al., 2014; Tao et al., 2020), the newly parameters were generated following the equation (6):

$$\mu_{new} = \mu_{old} + r \frac{\mu_{max} - \mu_{min}}{D} \quad (6)$$

where μ_{max} and μ_{min} are the upper and lower limits of parameter μ , r is a random number uniformly distributed between –0.5 and 0.5. D is the coefficient controlling the proposing step size. According to Xu et al. (2006) and Tao et al. (2020), D is set to the value of 5 so that the maximum step size is 1/10 of the range between the upper and lower limit.

In the test run, three parallel formal runs were conducted using a Gaussian distribution of parameters as the proposal distribution. The proposal distribution is constructed based on the covariances of the

Table 1

Key parameters and their default and optimized values.

Processes	Parameters	Description	Unit	Default	Prior Range	M_Opt1		M_Opt2		S_Opt1		S_Opt2	
						mean	cv (%)						
Photosynthesis	PRDX(1)	coefficient for calculating potential monthly production as a function of solar radiation outside the atmosphere	unitless	0.5	[0.3 1]	0.52	24.61	0.44	25.22	0.49	29.41	0.53	28.04
	PPDF(1)	optimum temperature for production for parameterization of a Poisson Density Function curve to simulate temperature effect on growth	°C	18, 16 ^a	[10 25]	15.70	20.13	18.59	18.52	16.02	23.38	19.08	13.15
	PPDF(2)	maximum temperature for production for parameterization of a Poisson Density Function curve to simulate temperature effect on growth	°C	37, 31 ^b	[25 50]	40.06	15.19	42.44	11.78	39.83	14.61	34.87	16.03
	PPDF(3)	left curve shape for parameterization of a Poisson Density Function curve to simulate temperature effect on growth	°C	1	[0.5 1.5]	0.99	26.55	1.04	26.15	0.97	27.85	0.97	27.33
	PPDF(4)	right curve shape for parameterization of a Poisson Density Function curve to simulate temperature effect on growth	°C	3, 3.5 ^c	[25]	3.22	23.67	3.16	21.62	3.38	23.24	4.01	14.11
	PPRPTS(3)	the lowest ratio of available water to pet at which there is no restriction on production	unitless	0.8	[0.7 0.9]	0.80	7.09	0.81	6.92	0.81	6.34	0.81	6.26
Decay	FSDETH (1)	maximum shoot death (fraction/month)	unitless	0.2	[0.1 0.25]	0.19	20.43	0.16	26.77	0.15	21.32	0.15	22.22
	FALLRT	fall rate (fraction of standing dead which falls each month)	unitless	0.15	[0.1 0.25]	0.15	24.79	0.16	25.49	0.16	25.56	0.14	24.69
Decomposition	RDR	maximum root death rate (fraction/month)	unitless	0.2	[0.1 0.25]	0.17	22.98	0.19	22.53	0.16	23.16	0.16	22.40
	DEC3(1)	maximum decomposition rate of surface organic matter with active turnover	g C/month	6	[58]	6.53	11.92	6.02	11.54	6.55	12.91	6.56	12.20
	DEC3(2)	maximum decomposition rate of soil organic matter with active turnover	g C/month	7.3	[610]	8.15	13.10	7.32	13.56	7.79	13.24	7.53	12.85
	DEC4	maximum decomposition rate of soil organic matter with slow turnover	g C/month	0.0066	[0.004 0.008]	0.0062	16.44	0.0060	17.64	0.0060	17.11	0.0057	19.17
	DEC5(1)	maximum decomposition rate of surface organic matter with intermediate turnover	g C/month	0.2	[0.1 0.3]	0.20	26.91	0.20	29.68	0.19	28.72	0.17	27.35
	DEC5(2)	maximum decomposition rate of soil organic matter with intermediate turnover	g C/month	0.2	[0.1 0.3]	0.23	21.12	0.20	25.72	0.22	20.72	0.24	15.22
TEFF	TEFF(1)	"x" location of inflection point, for determining the temperature component of decomposition factor	°C	15.4	[10 20]	13.71	16.37	14.83	17.32	14.36	16.08	15.38	13.53
	TEFF(2)	"y" location of inflection point, for determining the temperature component of decomposition factor	°C	11.75	[515]	12.13	11.18	12.06	10.40	12.19	10.79	11.95	10.59
	TEFF(3)	step size (distance from the maximum point to the minimum point), for	unitless	29.7	[25 35]	29.65	8.20	29.90	8.48	29.67	9.04	31.09	8.87

(continued on next page)

Table 1 (continued)

Processes	Parameters	Description	Unit	Default	Prior Range	M_Opt1		M_Opt2		S_Opt1		S_Opt2	
						mean	cv (%)	mean	cv (%)	mean	cv (%)	mean	cv (%)
Mineralization	P1CO2A (1)	determining the temperature component of the decomposition factor	unitless	0.6	[0.4 0.8]	0.61	16.42	0.57	16.85	0.60	17.99	0.62	16.61
	P1CO2A (2)	intercept parameter which controls flow from surface organic matter with fast turnover to CO ₂				0.17	[0.15 0.19]	0.17	6.75	0.17	6.01	0.17	6.11
	P1CO2B (1)	slope parameter which controls flow from surface organic matter with fast turnover to CO ₂				0	[0 0.2]	0.10	52.79	0.09	59.87	0.11	48.06
	P1CO2B (2)	slope parameter which controls flow from soil organic matter with fast turnover to CO ₂				0.68	[0.6 0.75]	0.68	6.23	0.68	5.93	0.67	5.88
	P2CO2(1)	controls flow from surface organic matter with intermediate turnover to CO ₂				0.55	[0.5 0.6]	0.55	4.84	0.53	4.52	0.55	4.66
	P2CO2(2)	controls flow from soil organic matter with intermediate turnover to CO ₂				0.55	[0.5 0.6]	0.56	4.92	0.55	5.06	0.55	4.79
	P3CO2	controls flow from soil organic matter with slow turnover rate to CO ₂				0.55	[0.5 0.6]	0.55	4.79	0.55	5.00	0.55	5.04

Note: a, b, c represent the default values of PPDF(1), PPDF(2) and PPDF(4) for Steppe ecosystem, respectively. M_Opt1, M_Opt2, S_Opt1 and S_Opt2 mean the optimization strategies (Opt1 and Opt2) for meadow (M) and Steppe (S) ecosystems, respectively.

parameters estimated from the initial test run. In the formal runs, parameter values were proposed based on the equation (7):

$$\mu_{new} = \mu_d + \mathcal{N}(0, cov(\mu_{testrun})) \quad (7)$$

9-10 Three runs starting at dispersed initial points in the parameter space and each run contained 10,000 simulations. After the first half of the simulations (as burn-in phase) was discarded, the Gelman-Rubin (G-R) statistics for each parameter was calculated to test whether or not the parameter convergent to stationary distribution (Gelman & Rubin, 1992). The parameters are better convergent when the G-R value is closer to 1.

11 The G-R values for all optimized parameters in our study were approximately-one (Table S3), indicating that all estimated parameters were converged among the MCMC chains. The acceptance rates of parameters sets for both ecosystems based on Opt1 or Opt2 were among 18% ~ 33% for three parallel chains. According to Xu et al., (2006) and Hararuk et al. (2014), the union of the all accepted parameter values after removing their burn-in periods were used to generated the posterior distribution for each parameter.

2.4. Evaluation of model performance with default and optimized parameterization

Before running the model calibration, all the soil samples and EC measurements were randomly split into two sub-sets for calibration (60%) and validation (40%) respectively (Fig. 1) (Smith et al., 2013; Hararuk et al., 2014). The split algorithm was carried out using the "sample" function in R software (Becker et al., 2020). Because the EC measurements from many sites (13 out of the 27 sites) only provided one-year or an even shorter period of measurements (Table S2), the split of EC measurements was based on site months rather than site numbers (Scholze et al., 2019; Norton et al., 2019). Both calibration and

validation are representative for the entire distribution of the observations (Table S4). The calibration dataset was assimilated into the CENTURY model based on Bayes' theorem to constrain model parameters. In present study, we compared two assimilation strategies (Opt1 – assimilating C pool observations; Opt2 – assimilating both C pool and CO₂ flux).

12 In data assimilation, we randomly sampled parameter values 1,000 times from the posterior distributions derived from different strategies, respectively (Tao et al., 2020). These sampled parameter values were further applied to CENTURY model. Comparison between observations and simulations (SOC and NEP for independent validation dataset was carried out. Note that the simulated SOCD was based on the averaged outputs from 1980 to 1985, given that soil samples used in this study were mainly collected in this period. All results and discussion about the agreement of simulated data with observed date referred to the validation subsample of the data.

The coefficient of determination (R²), root mean square error (RMSE) and mean absolute error (MAE) between model simulations and observations were calculated to evaluate the model performance.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - f_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - f_i)^2}{n}} \quad (9)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - f_i|}{n} \quad (10)$$

where, y_i and f_i are the i^{th} observed and simulated values, respectively, and \bar{y} is the mean value of the observations.

In addition, to examine the correlation of individually calibrated parameter with model performance, we first calculated the RMSE

between observed and simulated SOCD from each accepted run. Furthermore, Pearson correlation coefficient (r) between individual calibrated parameter and RMSE values was obtained by the following equation,

$$r = \frac{\sum_{t=1}^T (RMSE_t - \bar{RMSE})(p_t - \bar{p})}{\sqrt{\sum_{t=1}^T (RMSE_t - \bar{RMSE})^2} \sqrt{\sum_{t=1}^T (p_t - \bar{p})^2}} \quad (11)$$

where, $RMSE_t$ and p_t are the values of $RMSE$ and parameter p of t^{th} accepted run. $RMSE$ and \bar{p} are mean values of $RMSE$ and parameter p derived by all the accepted simulations (T).

3. Results

3.1. Comparison of model performance before and after applying data assimilation

The calibration and validation datasets were comparable, with similar mean and span of SOCD and NEP values (Table S4). In general, the average SOCD in the topsoil (0–30 cm) of the QTP grassland were 4.41 kg m^{-2} , ranging from 0.13 to 10.95 kg m^{-2} . Monthly NEP values in the study area varied from -67.56 to $109.67 \text{ g C m}^{-2} \text{ mon}^{-1}$, with an average value as $1.71 \text{ g C m}^{-2} \text{ mon}^{-1}$.

Performance of the model before and after data assimilation was assessed (Fig. 2). Compared to default model, the increased slope and decreased RMSE of Opt1 and Opt2 in both calibration and validation datasets indicated that the application of data assimilation could reduce the simulation biases. The agreement of simulated data with observed independent validation data was depicted in Fig. 3. All regression lines between model estimates and observations in Fig. 3 are statistically significant ($p < 0.001$).

Results in Fig. 3 showed that after applying the data assimilation (either Opt1 or Opt2), the model could dramatically reduce the mismatches between simulated outputs and observations than the simulations using default parameter values. In term of SOCD simulation, data

assimilation by Opt1 and Opt2 strategies resulted in a decrease of MAE from 1.56 to 1.21 and 1.26 kg m^{-2} , respectively. For NEP simulation, MAE dropped from 19.50 to 15.13 and $12.89 \text{ g C m}^{-2} \text{ mon}^{-1}$, respectively. Parameter optimization also narrowed the dispersion between observations and simulations, with data closely distributed around the 1:1 line (Figure S4). Compared with the assimilation of only soil C observations (Opt1), the assimilation with both soil C and fluxes data (Opt2) had little impact on the model's performance of SOCD. However, it dramatically improved the performance of modelling NEP with an over 12% increase of explanation power (R^2) for both meadow and steppe.

After calibrating parameters using Opt2, the CENTURY model can explain more than half variability of total soil C (R^2 of 0.73) and NEP (R^2 of 0.62) observations in validation datasets, which is superior to the default model (R^2 of 0.59 for SOCD and R^2 of 0.16 for NEP).

3.2. Comparison of parameter optimization among different parameterization

The violin-box plots of posterior probability distributions of 24 parameters were depicted in Fig. 4. The interval of Y axis represents the range of prior uniform distribution. Results showed that most parameters controlling the photosynthesis process and decay process were better constrained by Opt2 than Opt1, either resulting in a smaller posterior range or an obvious unimodal distribution. For instance, the posterior scalar of primary production ($PRDX(1)$) for meadow was tightly constrained around the value of 0.4 and optimized optimum temperature ($PPDF(1)$) for steppe was distributed above 15°C after Opt2, narrowed nearly half range than Opt1 (Fig. 4). The sharper skewed distribution of falling rate ($FALLRT$) for steppe after Opt2 also suggests a tight constraint of these parameters. This improvement indicates that adding NEP observations to data assimilation can alter these parameters towards a more realistic projection of C flows in the aboveground processes.

Regarding the parameters dominating the decomposition process,

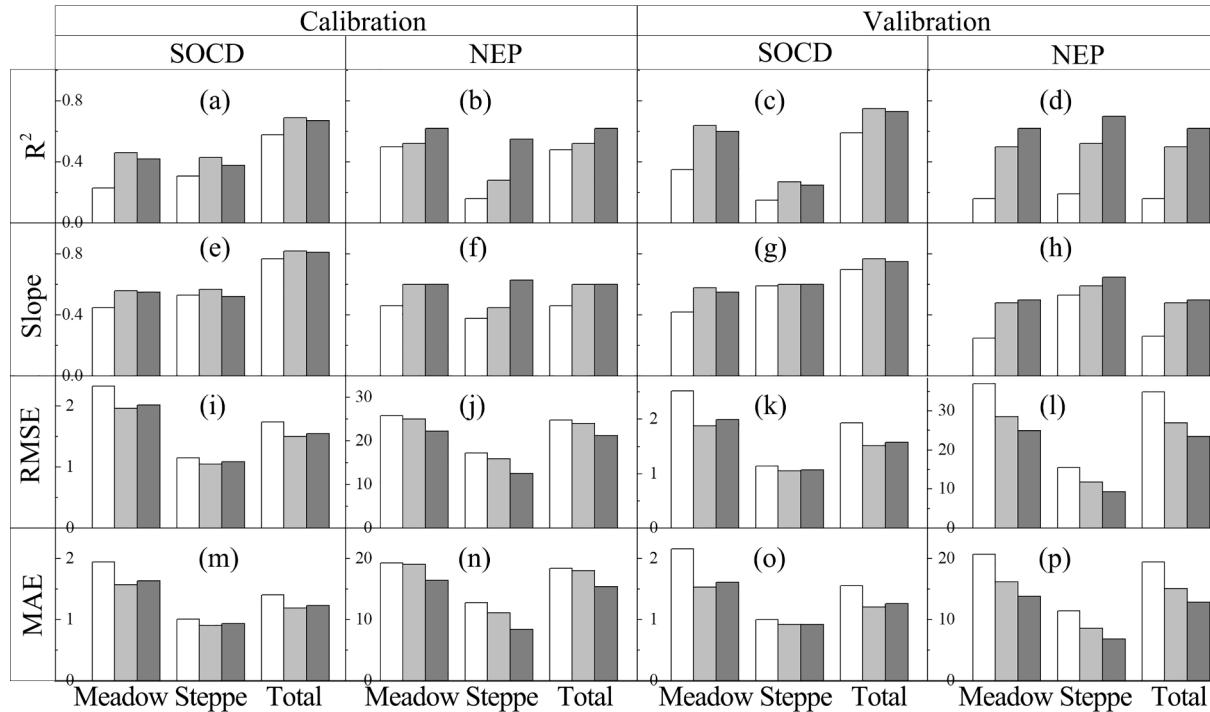


Fig. 2. Accuracy evaluation of default (white), Opt1 (light gray) and Opt2 (dark gray) models for simulation of surface (0–30 cm) soil organic carbon density (SOCD, kg m^{-2}) (a) and net ecosystem production (NEP, $\text{g C m}^{-2} \text{ mon}^{-1}$) in calibration dataset, and validation dataset. Opt1 is constrained by soil organic carbon density (SOCD) and Opt2 is constrained by both C fluxes and SOCD.

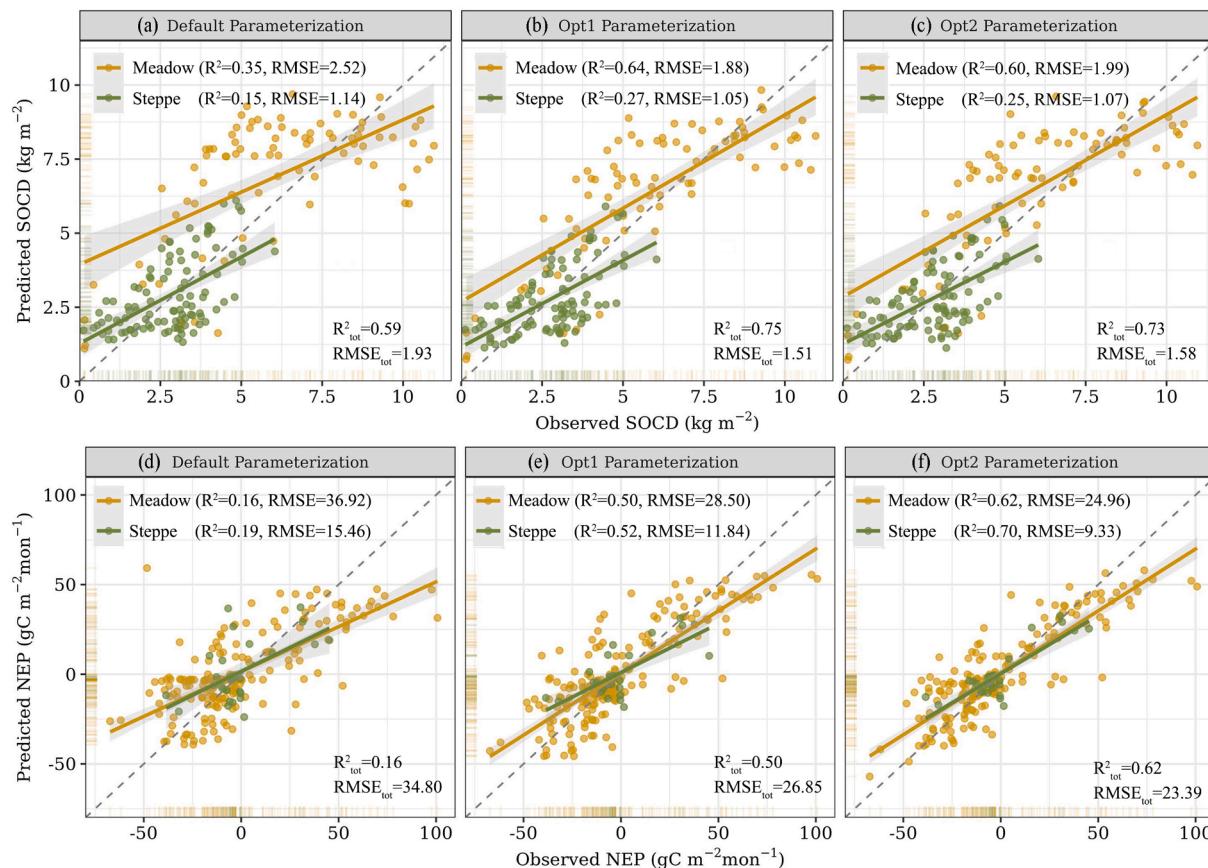


Fig. 3. The scatter plot of observed and simulated surface (0–30 cm) soil organic carbon density (SOCD, kg m⁻²) (a-c) and net ecosystem production (NEP, g C m⁻² mon⁻¹) (d-f) in the validation dataset based on the default, Opt1 (constrained by SOCD) and Opt2 (constrained by both C fluxes and SOCD) parameterizations for the meadow and steppe ecosystems.

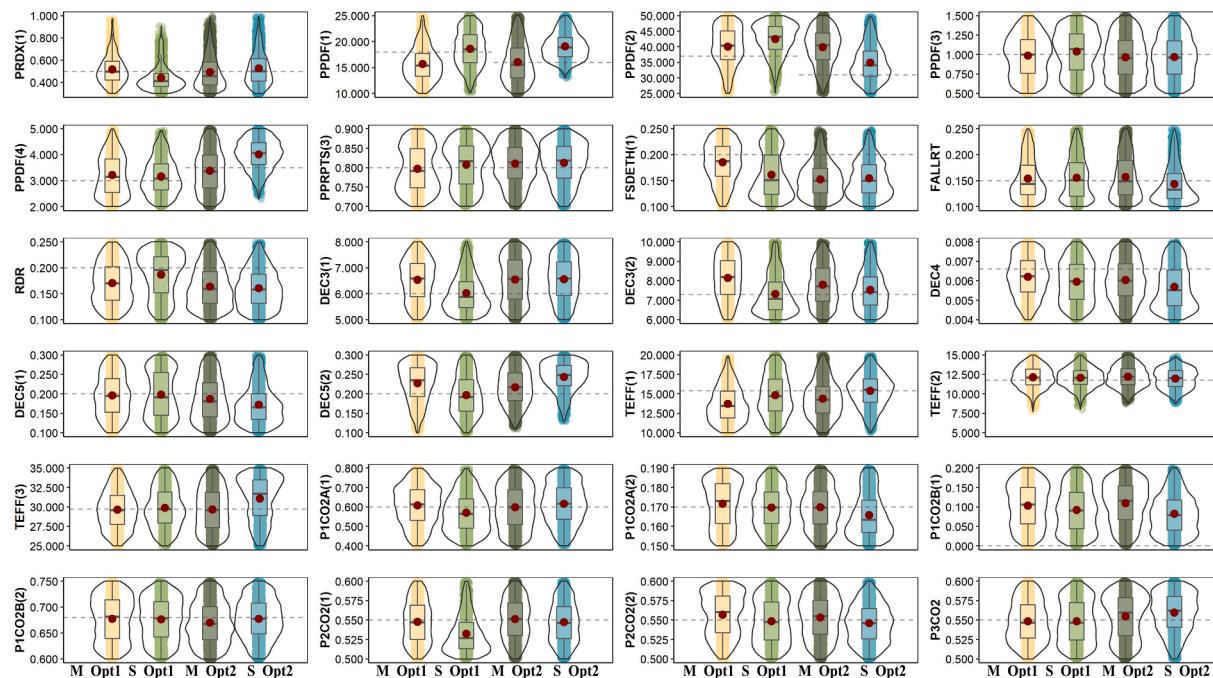


Fig. 4. The posterior probability distributions of the parameters for meadow (M) and steppe (S) derived from Opt1 (constrained by SOCD) and Opt2 (constrained by both C fluxes and SOCD). The dash line represents the mean of prior parameter range. The box plot represents the median (red dot), 25th percentile (box) and 75th percentile (line) of posterior parameters. The parameters can be found in Table 1.

both Opt1 and Opt2 resulted in noticeable constraints on some relevant parameters, especially for *TEFF(2)*, *DEC3(2)* and *DEC5(2)*, as their posterior probability distributed narrowly within the prior uniform distribution after data assimilation. Similarly, some parameters related to the mineralization process were also well constrained. For instance, the highly left-skewed distribution of *P2CO2(1)* for meadow and standard normalized distribution for steppe both meant the parameters were optimized into small ranges after DA. Interestingly, Opt1 and Opt2 resulted in little difference on the parameterization of both decomposition and mineralization process, indicating that assimilation of NEP observations caused a minor effects on constraining belowground process.

3.3. Correlation between optimized parameters and model performances

Fig. 5 shows the correlation coefficients between model performance and calibrated parameters for SOCD and NEP estimation from Opt1 and Opt2. The correlation coefficients from Opt2 were generally higher than that from Opt1, especially for the steppe ecosystem. Only 3 out of the 24

parameters optimized after Opt1 have a low correlation with SOCD estimation for steppe ecosystem ($0.3 < |r| < 0.5$) and only parameter *TEFF(1)* resulted in low correlation with NEP estimation ($|r| = 0.44$). While for Opt2, over 10 parameters has $|r|$ higher than 0.3 for simulating either SOCD or NEP. Among them, the parameters representing the scalars of primary production (*PRDX(1)*) is most related to SOCD modelling accuracy, with the highest correlation coefficient ($r = 0.54$). While for NEP estimation, well-constrained *PPDF(1)* and *PPDF(2)* are ranked as the most negatively ($r = -0.49$) and positively ($r = 0.45$) related parameters, respectively. From a processes perspective, after Opt2 parameters from photosynthesis process have higher correlations with the model performance in simulating NEP. While for soil C estimation, parameters controlling the decay and decomposition process become more related to the model accuracy.

For the meadow ecosystem, Opt2 showed similar advantages over Opt1 in strengthening the correlation relationship between parameters and NEP prediction. A total of five more parameters optimized by Opt2 showed a higher relationship with NEP simulation than that using Opt1, with the values of $|r|$ larger than 0.3. Most of the high correlated

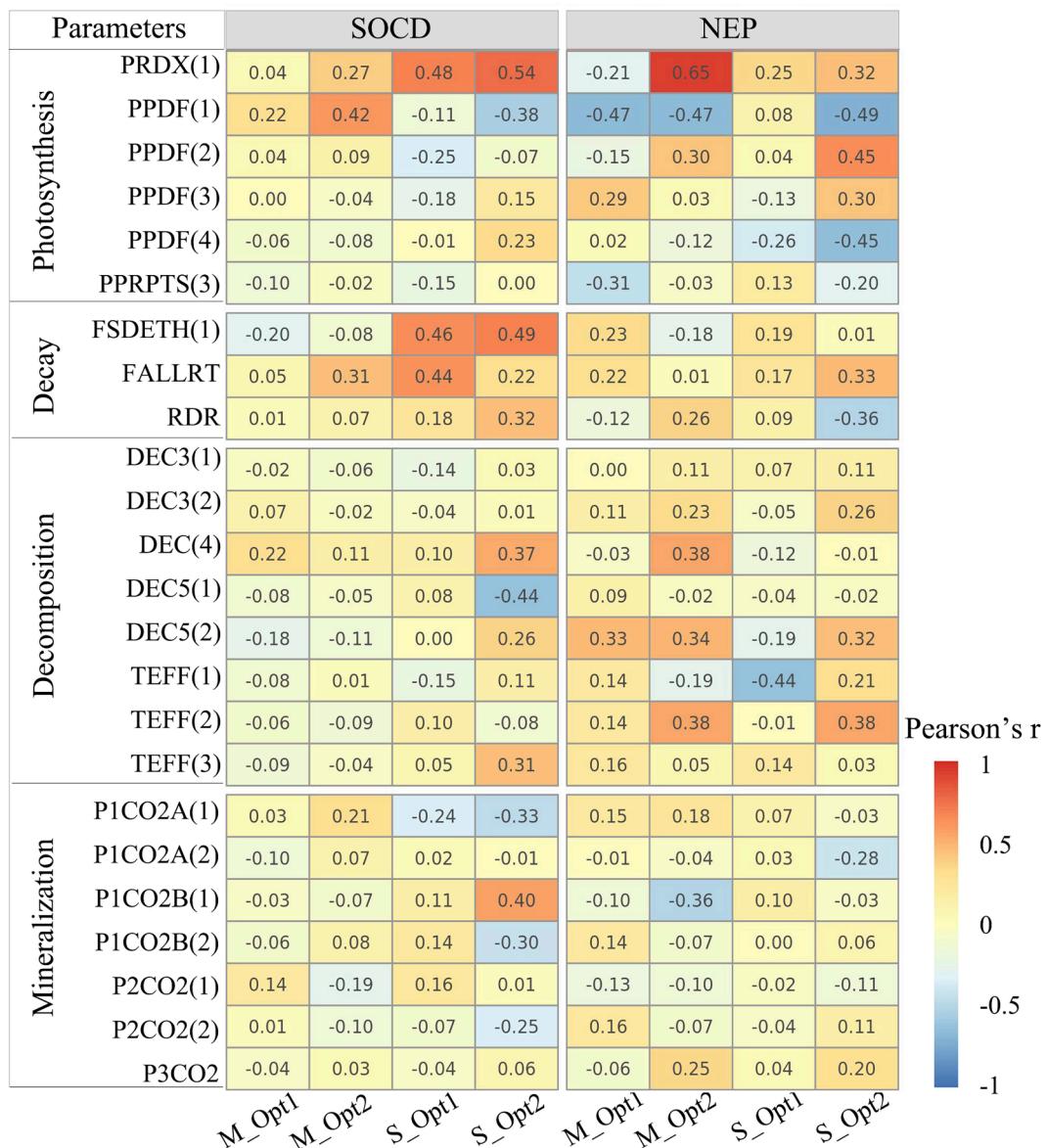


Fig. 5. The Pearson's correlation between individual parameters and model performance (in terms of root mean square error, RMSE) for the estimation of soil organic carbon density (SOCD) and net ecosystem production (NEP) in meadow (M) and steppe (S) derived from Opt1 (constrained by SOCD) and Opt2 (constrained by both C fluxes and SOCD).

parameters are related to the process of photosynthesis and decomposition. Among them, the regulator of photosynthesis (*PRDX(1)*) is listed as the two strongest related parameters for NEP modelling ($r = 0.65$). Conversely, the optimization strategy did not show much effect on the correlation between parameters and the SOCD estimation. Only *PPDF(1)* and *FALLRT* were slightly correlated to the SOCD modelling performance with r of 0.42 and 0.31, respectively.

3.4. Spatial patterns of SOCD and NEP in the QTP grasslands

Based on the default and optimized parameters, the spatial distribution of top layer SOCD (0–30 cm) and yearly NEP across the dominant grasslands of QTP were produced. Fig. 6 showed the simulation outputs for 1980–1985, during which most of the observed soil samples were collected. Spatially, SOCD gradually decreased from the eastern part of QTP to the west (Fig. 6a). The average SOCD in most meadow areas was higher than 5 kg m^{-2} , with a striped distribution along with river courses. The total soil organic carbon stock (SOCS) in the meadow ecosystem was around 4.50 Pg. While in the steppe grasslands, the SOCD was relatively low. Part of the low SOCD was distributed in north central QTP, where the Qaidam Basin is located. The basin is a hyperarid basin and regarded as one of the aridest non-polar locations on earth. Apart from that, the steppe grasslands between southwest QTP and north Gangdisi Range also showed lower SOCD, with an average less than 2 kg m^{-2} . Results showed that nearly 2.50 Pg C was reserved in the steppe soils of QTP.

Compared with the default model, estimated SOCS based on optimized models in both the meadow and steppe grasslands were decreased (Fig. 6b and 6c). Using the Opt1 parameterization, the calibrated model obtained higher predictions in the central meadows and steppes, while SOCD in the eastern meadows and southwest steppes showed a noticeable decrease after calibration. Unlike location-specific change of SOCD in each ecosystem for Opt1, the Opt2 parameterization resulted in a relatively unifying reduction of estimated SOCD in steppes and small changes in the meadow grasslands.

The estimated NEP in most regions of the QTP grasslands were positive and we also derived positive estimations of total NEP for both ecosystems (Fig. 6d), which means the QTP grassland was a C sink with a

total net uptake of 14.15 Tg C annually. C sequestration in the meadow grasslands was stronger with almost twice larger of total yearly NEP than steppes. Fig. 6e and 6f showed Opt1 and Opt2 parameterization resulted in divergent changes in NEP for different ecosystems. NEP estimated by the Opt1 simulation became significantly lower for meadow ecosystems.

4. Discussion

4.1. Improvement of soil carbon modelling

Accurate estimation of soil C using process-based models is important for improving our understanding of the global C cycle under global warming. However, numerous studies have documented large uncertainties in current soil C modelling. Model parameterization has been regarded as one of major sources for uncertainties of soil C projections in the process-based model (Luo et al., 2016). Using data assimilation to constrain parameters is an efficient approach to reduce model uncertainties. For example, Hararuk et al. (2014) and Tao et al. (2020) found that the MCMC data assimilation with soil C observations can improve the model to represent soil variability across global or regional scales by 14% or 11%, respectively. In our study, the similar extent of improvement for soil C estimates in the QTP grasslands was also obtained. All these work demonstrates that data assimilation is a promising technique to reduce model uncertainties. In addition, our study showed that adding C fluxes observations significantly improved NEP estimation but did not help reduce the uncertainty of soil C estimation. Keenan et al. (2013) and Du et al. (2015) reported a similar conclusion that the flux data mainly constrained flux-related parameters but was imbedded little information that can be used to constrain pool-based parameters. Our work suggest that the efficiency of data assimilation also relies on the assimilated dataset which is relevant and can specific purposes of model improvement.

Meanwhile, many studies also demonstrated that there were still considerable variations (around 25% to 73%) in soil C which was not captured by the optimized models. Even other optimization approaches can result in a similar magnitude of modelling uncertainty. Xu et al. (2018) calibrated parameters in a process-based model through a

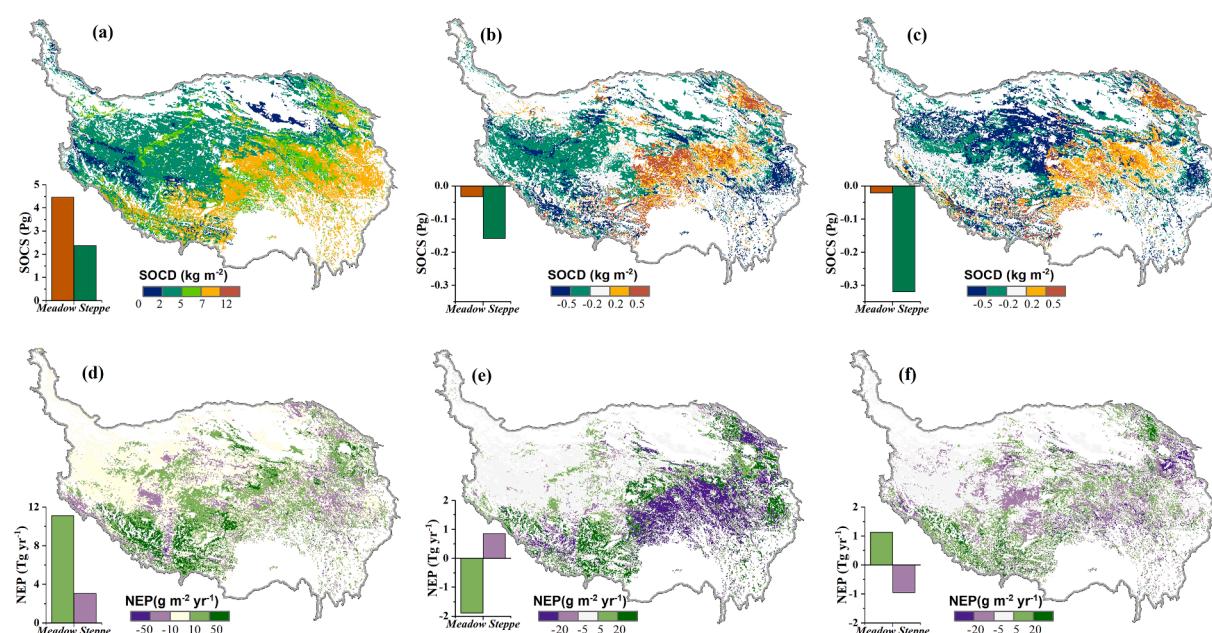


Fig. 6. The spatial pattern of soil organic carbon stock (SOCS, Pg) and net ecosystem productivity (NEP, Tg C/yr) across the upper layer (0–30 cm) of Qinghai-Tibet Plateau (QTP) grasslands derived from default model (a, d) and the difference between optimized and default model (Opt1: b, e; Opt2: c, f). Opt1 is constrained by soil organic carbon density (SOCD) and Opt2 is constrained by both C fluxes and SOCD.

13–14 surrogate-based optimization models and found that only 42% of the soil C variation was explained. The unexplained uncertainty might be partly due to that the static parameters may not account for spatial variability of soil C storage. Some studies have shown that model parameters vary between locations. For instance, using spatially-explicit parameter values, Tao et al. (2020) yielded the least geographical bias in SOC estimation and explained 62% variability of observed soil C. Our study also showed that posterior parameter values were highly different among steppe and meadow grasslands, suggesting that to simulate SOCD for different ecosystems, regional and global models should calibrate their parameterization for each ecosystem.

 The model structure is another major source of modelling uncertainty (Luo et al., 2016). The CENTURY model is a typical soil C dynamic model that simulates soil C decomposition as a first-order decay function of the soil organic matter. Nevertheless, studies have shown that such a framework of modelling ignored the microbial activities, such as effects of priming and thermal acclimation (Kuzyakov et al., 2000; Fontaine et al., 2007) and microbial acclimation to increasing temperatures (Peng et al., 2009). Nonlinear microbe-explicit models based on Michaelis-Menten equations have been developed recently (Wieder et al., 2013, 2015). Using a calibrated microbial model, Hararuk et al. (2015) explained 51% of the variability in the observed SOC, with a 10%

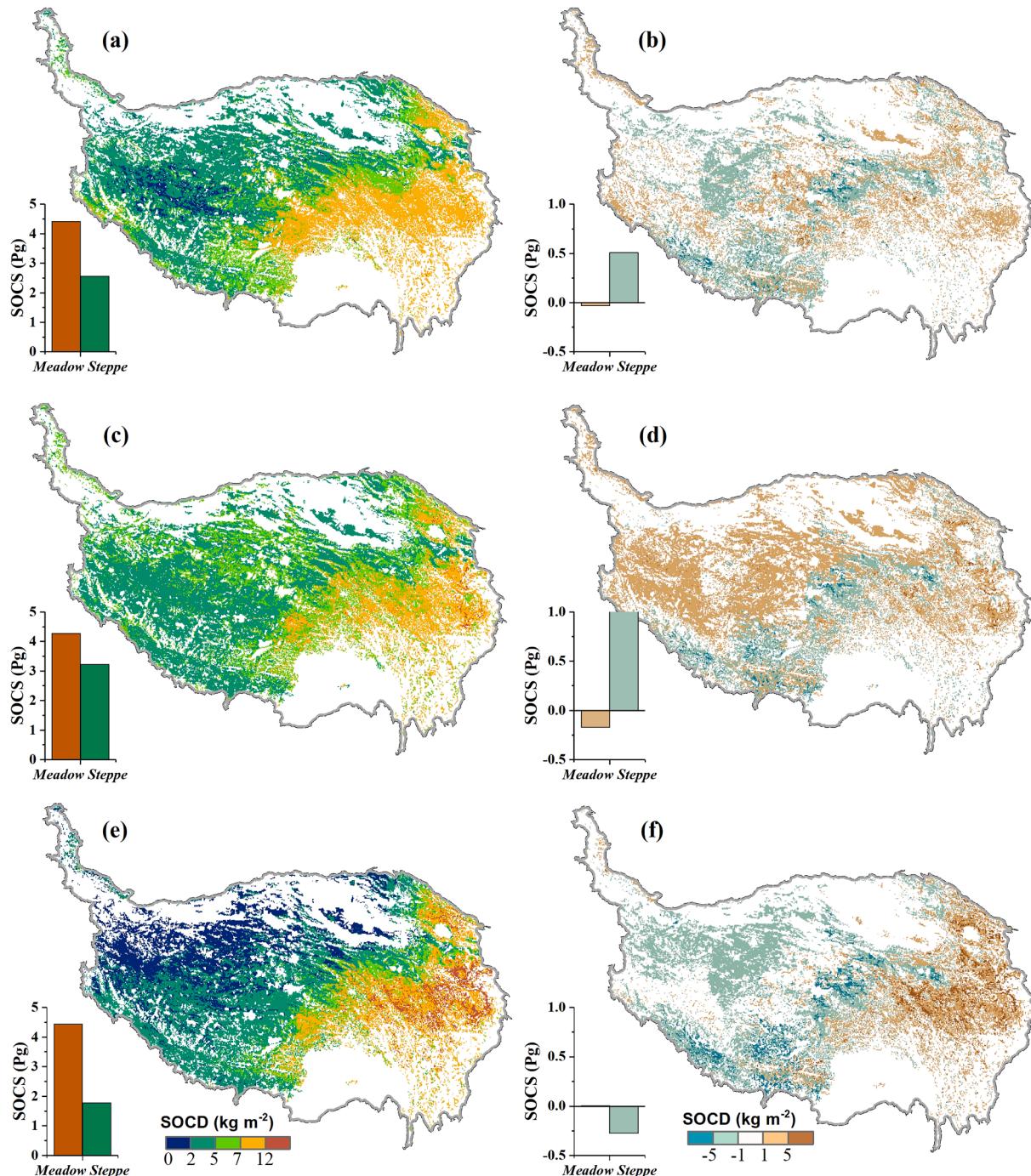


Fig. 7. The spatial distribution of topsoil organic carbon density (SOCD, kg m^{-2}) (0–30 cm) from SoilGrids 2.0 (Hengl et al., 2017), GSOCmap (FAO, 2017) and ESSD_Wang (Wang et al., 2021) (a, c, e) and the corresponding difference between the existing datasets with estimation from this study (derived from Opt2; Opt2 is constrained by SOCD and C fluxes) (b, d, f). The subplots represent the change in soil organic carbon stock (SOCS, Pg) for meadow and steppe ecosystems.

improvement than a calibrated conventional model. Additionally, many disturbance events especially freeze-thawing cycles in permafrost-affected soils have strong effects on soil C dynamics. Evidence showed over the last 50 years, the area of collapsed permafrost has increased by a ratio of about 40 (Gao et al., 2021). Although a decadal experiment indicated that upper active layer (the uppermost 30 cm) of soils experienced a mean accumulation rate of $28.0 \text{ g C m}^{-2} \text{ yr}^{-1}$ (Ding et al., 2017), some studies demonstrated that permafrost thawing in future warming scenarios would release a large amount of deep soil C into the atmosphere along with gradual active layer deepening (Wang et al., 2020). Therefore, to understand and simulate depth-dependent soil C dynamics should encourage to use the models with a multilayer soil profile and explicit representation of microbial decomposition.

4.2. A SOCD map for QTP grasslands

Based on the ensemble mean of Opt2 using the CENTURY model we have constructed a new baseline map of SOCD for QTP grasslands. Our results were evaluated against three published SOCD datasets generated based on different approaches and data input (Fig. 7). Specifically, GSOCmap was generated based on a country-specific stratification and launched by Food and Agriculture Organization (FAO) in 2017. Both SoilGrids 2.0 (Poggio et al., 2021) and ESSD_wang (Wang et al., 2021) were obtained through machine learning methods based on different sampling sites and environmental variables. The spatial variability and magnitude of our estimation are comparable to these datasets, suggesting that the SOCD gradually decreased from the eastern meadow grassland to the western steppe regions.

Among all the datasets, the GSOCmap resulted in a more spatially homogeneous SOCD estimates with the smallest difference of SOCS between the two ecosystems (Fig. 7c). The insignificant spatial heterogeneity and lowest accuracy of SOCD estimation (Fig. 8a) found in GSOCmap indicate that both process-based modelling and machine learning methods have advantages in upscaling over country-specific stratification, which has been confirmed by many previous studies (Cambule et al., 2013; Zhou et al., 2019; Liang et al., 2019).

Although the SoilGrids 2.0 has the highest spatial resolution (250 m) among the study datasets, it showed a noticeable overestimation of SOCD in the Steppe area (Fig. 7b). The overestimation is also reported by Zhou et al. (2019) and Wang et al. (2021) and mainly resulted from overestimated soil bulk density, as well as the neglected influence of

coarse gravel content (Hengl et al., 2017). The ESSD_Wang dataset is derived using the approach of machine learning. Unlike SoilGrids 2.0 which is derived from a global model, ESSD_Wang is generated based on a regional model. The higher accuracy of ESSD_Wang in SOCD estimation (Fig. 8b) indicates that the regional model based on a finer spatiotemporal resolution is more accurate when focusing on regional extent (Vitharana et al., 2019; Liu et al., 2020; Wang et al., 2021). The magnitude of soil C in our estimation is much closer to the result from ESSD_Wang (Fig. 7f), especially for meadow ecosystems.

Machine learning approaches have been widely used to provide fine-resolution soil information globally and regionally. However, the Taylor diagram (Fig. 8a) showed that for both meadows and steppes, the calibrated CENTURY model in this study is superior to other approaches and capable to obtain more accurate estimation of SOCD in the QTP. Our estimations showed a better match with observational soil C (with smaller RMSE and shorter distance to REF) (Fig. 8a). The averaged SOCD for meadows and steppes are also more comparable with results derived from previous studies (Fig. 8b). This might be because machine-learning approaches are data-driven methods, which may lack its accuracy for data-rare regions. The different SOCD spatial patterns derived by SoilGrids 2.0 and ESSD_Wang might result from the adoption of different land use types (Hengl et al., 2017; Wang et al., 2021), which also emphasized the conclusion that results derived by machine learning methods are highly data-dependent. SOCD for poorly-accessible areas should be validated with estimates by other approaches.

4.3. Outlook

Although our data assimilation method estimates SOC for meadow and steppe, which is comparable to other observation data sets, our model has not accounted for some important ecosystem processes such as human-induced land cover change, grazing and permafrost thawing, which could potentially affect long-term ecosystem C dynamics and C turnover rates in responses to future climate change, particularly for the alpine ecosystems. For example, numerous studies have reported that permafrost thawing may significantly influence the C dynamics by not only enhancing plant C fixation and soil C loss but also accelerating the turnover of old soil C (i.e. priming effect) (Koven et al., 2011; Schuur et al., 2015; Chen et al., 2018; Chang et al., 2021). The rhizosphere priming effect may also be further triggered in deep soils with roots reaching the thawing front of the soil (Iversen et al., 2015). Keuper et al.

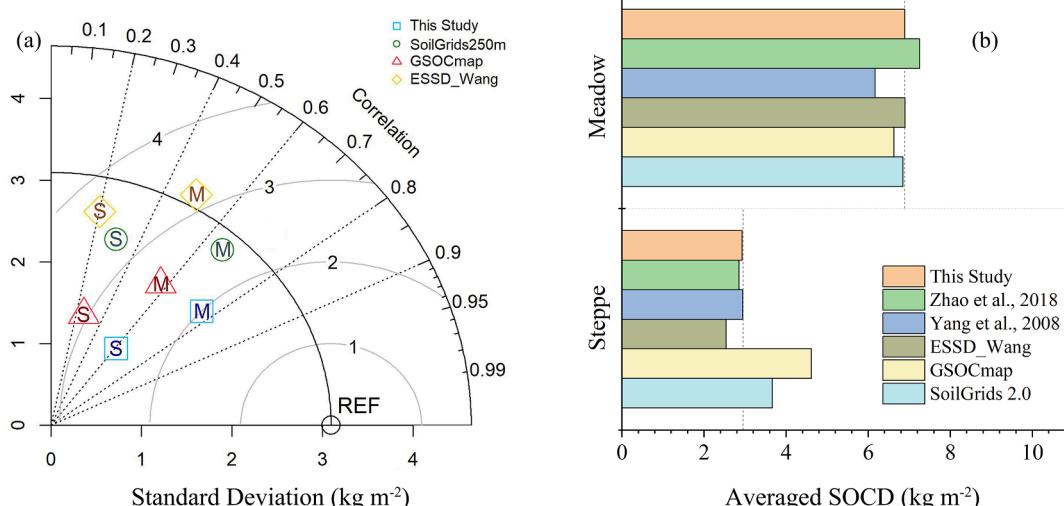


Fig. 8. Comparisons of the estimation of this study (derived from Opt2; Opt2 is constrained by SOCD and C fluxes) with existing datasets and previous studies about site-wise soil organic carbon density (SOCD, kg m^{-2}) estimation (a) and the averaged SOCD (b) stored in the surface soil (0–30 cm) of meadow and steppe ecosystems. The vertical dot line in the subplot b represents the mean value of the observed SOCD. Note, in panel a, M stands for meadow and S stands for steppe. REF is the standard deviation of independent validation datasets.

(2020) has predicted a loss of about 40 Pg soil C amplified by rhizosphere priming from the northern permafrost area by 2100. In addition, livestock grazing as a dominant activity in grassland ecosystems for several millennia has shown a profound impact on C cycles. Evidence showed after biomass was ingested by animals, only a proportion of C returned to grasslands as manure, which would influence nutrient availability, soil structures and root distribution (Jones et al., 2006; Ghab & Kacorzyk, 2011; Matsuura et al., 2021). Zhang et al. (2017) found grazing intensity in the QTP could explain 2.3% variation in SOC. By explicitly representing grassland management modules such as livestock production and the ratio of intensive and extensive management in a Land Surface Model, Chang et al. (2013, 2015) also decreased 53% of bias in the C fluxes. Apart from that, the frequency and intensity of land-cover conversion, wildfire and soil erosion in the QTP are increasing due to rising temperature and human activities (Wang et al., 2019; You et al., 2018; Teng et al., 2018). These events showed a strong impact on the permafrost-affected C cycle (Lal, 2005; Mack et al., 2011; Lugato et al., 2018; Inley, 2021) and should be included in further studies.

15-16

2 notes
and non-linear processes often encounters issues of equifinality.

Equifinality means that different combinations of parameters can lead to a similarly good fit of model outputs to the observations, even though some parameters are based on values beyond the range of the studied context (Beven and Freer, 2001). This issue often occurs when the model is over-parameterized. In our study, we have selected many parameters from the same processes (e.g., photosynthesis and decomposition), simply because we are not certain about which parameter can be the main factor that regulates the process. Therefore, compared to Opt1, Opt2 showed some parameters (e.g., PPDF(4) and DEC5(1)) with a reduced posterior range, but some other parameters (e.g., DEC3 and DEC(4)) with an increased posterior range. To further constrain the parameters that regulate C flows exchanged between the above- and belowground C pools and reduce the equifinality in parameterization, we need more measurements that explicitly reflect C flows and pool turnover rates, such as the litterfall rate, soil respiration of CO₂ fluxes, rate coefficients of decomposition in both the labile and humus C pools.

17

Overall, our data assimilation results have shown the merits of assimilating soil samples and EC measurements to improve C simulation, the latter of which has not been included in any commonly-used C flux measurement networks. As field-based and satellite-derived products are increasing rapidly (Xiao et al., 2019), different types of observations (e.g. such as soil moisture, land surface temperature and leaf area index) have become available to tightly constrain different components of the underlying process model (MacBean et al., 2016; Scholze et al., 2017, 2019). Based on multiple observations, various data assimilation systems have been launched recently and showed advantages to constrain C cycles (Scholze et al., 2007; Peylin et al., 2016). For instance, by simultaneous assimilation of SMOS soil moisture data and CO₂ flask samples within a Carbon Cycle Data Assimilation System (CCDAS), Wu et al. (2020) optimized soil hydrological and biophysical parameters and obtained good agreements of simulated carbon fluxes with measurements. However, existing multiple-data assimilation systems are mostly model-specific, making them inflexible to be applied to different models (Huang et al., 2021). Taking full advantage of using multiple observations and independent data assimilation approach is necessary and promising to improve the model's predictability.

5. Conclusions

Accurate estimation of C pool and CO₂ fluxes is essential for reducing the uncertainty of Earth system modelling and improving understanding of the terrestrial carbon pools in response to climate warming. In this study, an advanced data assimilation technique was applied to calibrate a widely-used process-based ecosystem biogeochemistry model, the CENTURY model. With about 12% more variation in observed soil C and NEP explained, the calibrated model yielded a better simulation of

SOCD and NEP across the QTP grasslands. Additionally, our study demonstrated that different types of observations contain distinct information and can help to tightly constrain multiple components of the ecosystem model. Based on the optimized model by simultaneously assimilating C-pool and C-flux observations, we simulated a total of 6.48 Pg soil carbon stored in the upper layer of the QTP grassland and 14.33 Tg carbon sink annually. Compared with existing datasets, our estimation was more reliable with a better match with observations. Overall, our study highlights that data assimilation with multiple observational data sets is a promising approach to constrain process-based ecosystem models and increase the robustness of model predictions for terrestrial C cycle feedback to future climate change.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We thank Y. Luo and his colleagues for the sharing code of data assimilation used in the present study and all colleagues from Z. Shi's team for collecting the soil samples. This study is financially supported by the National Natural Science Foundation of China (grant no. 41930754) and the China Scholarship Council (CSC). W.Z. was supported by the grants from the Swedish Research Council VR 2020-05338 and State Key Laboratory of Frozen Soil Engineering Open Fund (SKLFSE202002). The authors would also like to thank the Center for Scientific and Technical Computing at Lund University (LUNARC) for providing resources of computation and storage within the Swedish National Infrastructure for Computing project (LU 2021/2-115 and SNIC 2021/6-341). The authors also thank three anonymous reviewers for their constructive feedback.

Data Availability Statement

The EC measurements can be obtained from Wei et al (2021). Soil samples are available from <https://www.resdc.cn/data.aspx?DATAID=185> or on request from the corresponding author. The model input data can be download from <https://www.climatologylab.org/terraclimate.html> and <http://globalchange.bnu.edu.cn/research/data>. The calibrated parameters and model outputs are publicly available from Zenodo at <https://doi.org/10.5281/zenodo.6570584> (Zhao, 2022). Code for carrying out data assimilation and model simulation can also be obtained through this website.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.geoderma.2022.116283>.

References

- Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A., Hegewisch, K.C., 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* 5 (1), 1–12.
- Batjes, N.H., 1996. Total carbon and nitrogen in the soils of the world. *Eur. J. Soil Sci.* 47 (2), 151–163.
- Batjes, N.H., 2016. Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma* 269, 61–68.
- Becker, R.A., Chambers, J.M., Wilks, A.R., 1988. The New S Language. Wadsworth & Brooks/Cole.

- Berardi, D., Brzostek, E., Blanc-Betes, E., Davison, B., DeLucia, E.H., Hartman, M.D., Kent, J., Parton, W.J., Saha, D., Hudiburg, T.W., 2020. 21st-century biogeochemical modeling: challenges for Century-based models and where do we go from here? *GCB Bioenergy* 12 (10), 774–788.
- Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol.* 249 (1–4), 11–29.
- Boysen, L.R., Brovkin, V., Wärldin, D., Peano, D., Lansø, A.S., Delire, C., Burke, E., Poeplau, C., Don, A., 2021. Evaluation of soil carbon dynamics after forest cover change in CMIP6 land models using chronosequences. *Environ. Res. Lett.* 16 (7), 074030.
- Cambule, A., Rossiter, D., Stoorvogel, J., 2013. A methodology for digital soil mapping in poorly-accessible areas. *Geoderma* 192, 341–353.
- Chang, J., Ciais, P., Viovy, N., Vuichard, N., Sultan, B., Soussana, J.F., 2015. The greenhouse gas balance of European grasslands. *Glob. Chang. Biol.* 21 (10), 3748–3761.
- Chang, R., Liu, S., Chen, L., Li, N.a., Bing, H., Wang, T., Chen, X., Li, Y., Wang, G., 2021. Soil organic carbon becomes newer under warming at a permafrost site on the Tibetan Plateau. *Soil Biol. Biochem.* 152, 108074.
- Chang, J.F., Viovy, N., Vuichard, N., Ciais, P., Wang, T., Cozic, A., Lardy, R., Graux, A.-I., Klumpp, K., Martin, R., Soussana, J.-F., 2013. Incorporating grassland management in ORCHIDEE: model description and evaluation at 11 eddy-covariance sites in Europe. *Geosci. Model Dev.* 6 (6), 2165–2181.
- Chen, S., Arrouye, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., Libohova, Z., Lagacherie, P., Shi, Z., Hannam, J., Meersmans, J., Richer-de-Forges, A. C., Walter, C., 2022. Digital mapping of GlobalSoilMap soil properties at a broad scale: a review. *Geoderma* 409, 115567.
- Chen, L., Liu, L.i., Mao, C., Qin, S., Wang, J., Liu, F., Blagodatsky, S., Yang, G., Zhang, Q., Zhang, D., Yu, J., Yang, Y., 2018. Nitrogen availability regulates topsoil carbon dynamics after permafrost thaw by altering microbial metabolic efficiency. *Nat. Commun.* 9 (1).
- Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J. & Heimann, M. (2014). Carbon and other biogeochemical cycles. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* Cambridge University Press, pp. 465–570.
- Crowther, T.W., Todd-Brown, K.E., Rowe, C.W., Wieder, W.R., Carey, J.C., Machmuller, M.B., Snoek, B., Fang, S., Zhou, G., Allison, S.D., 2016. Quantifying global soil carbon losses in response to warming. *Nature* 540 (7631), 104–108.
- De Kauwe, M.G., Medlyn, B.E., Zaehle, S., Walker, A.P., Dietze, M.C., Wang, Y.-P., Luo, Y., Jain, A.K., El-Masri, B., Hickler, T., Wärldin, D., Weng, E., Parton, W.J., Thornton, P.E., Wang, S., Prentice, I.C., Asao, S., Smith, B., McCarthy, H.R., Iversen, C.M., Hanson, P.J., Warren, J.M., Oren, R., Norby, R.J., 2014. Where does the carbon go? A model-data intercomparison of vegetation carbon allocation and turnover processes at two temperate forest free-air CO₂ enrichment sites. *New Phytol.* 203 (3), 883–899.
- Ding, J., Chen, L., Ji, C., Hugelius, G., Li, Y., Liu, L.i., Qin, S., Zhang, B., Yang, G., Li, F., Fang, K., Chen, Y., Peng, Y., Zhao, X., He, H., Smith, P., Fang, J., Yang, Y., 2017. Decadal soil carbon accumulation across Tibetan permafrost regions. *Nat. Geosci.* 10 (6), 420–424.
- Du, Z., Nie, Y., He, Y., Yu, G., Wang, H., Zhou, X., 2015. Complementarity of flux-and biometric-based data to constrain parameters in a terrestrial carbon model. *Tellus B: Chem. Phys. Meteorol.* 67 (1), 24102.
- Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37 (12), 4302–4315.
- Finney, D.L., 2021. Lightning threatens permafrost. *Nature Clim. Change* 11 (5), 379–380.
- Fontaine, S., Barot, S., Barré, P., Bdioui, N., Mary, B., Rumpel, C., 2007. Stability of organic carbon in deep soil layers controlled by fresh carbon supply. *Nature* 450 (7167), 277–280.
- Gallagher, K., Charvin, K., Nielsen, S., Cambridge, M., Stephenson, J., 2009. Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for Earth Science problems. *Mar. Pet. Geol.* 26 (4), 525–535.
- Gao, T., Zhang, Y., Kang, S., Abbott, B.W., Wang, X., Zhang, T., Yi, S., Gustafsson, Ö., 2021. Accelerating permafrost collapse on the eastern Tibetan Plateau. *Environ. Res. Lett.* 16 (5), 054023.
- Ge, R., He, H., Ren, X., Zhang, L.i., Yu, G., Smallman, T.L., Zhou, T., Yu, S.-Y., Luo, Y., Xie, Z., Wang, S., Wang, H., Zhou, G., Zhang, Q., Wang, A., Fan, Z., Zhang, Y., Shen, W., Yin, H., Lin, L., 2019. Underestimated ecosystem carbon turnover time and sequestration under the steady state assumption: a perspective from long-term data assimilation. *Glob. Chang. Biol.* 25 (3), 938–953.
- Gelman, A., Gilks, W.R., Roberts, G.O., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* 7 (1), 110–120.
- Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7 (4), 457–472.
- Głab, T., Kacorzyk, P., 2011. Root distribution and herbage production under different management regimes of mountain grassland. *Soil Tillage Res.* 113 (2), 99–104.
- Hararuk, O., Xia, J., Luo, Y., 2014. Evaluation and improvement of a global land model against soil carbon data using a Bayesian Markov chain Monte Carlo method. *J. Geophys. Res. Biogeogr.* 119 (3), 403–417.
- Hararuk, O., Smith, M.J., Luo, Y., 2015. Microbial models with data-driven parameters predict stronger soil carbon responses to climate change. *Glob. Chang. Biol.* 21 (6), 2439–2453.
- Harmon, R., Challenor, P., 1997. A Markov chain Monte Carlo method for estimation and assimilation into models. *Ecol. Model.* 101 (1), 41–59.
- Harris, I., Osborn, T.J., Jones, P., Lister, D., 2020. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data* 7 (1), 1–18.
- Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. *PLoS One* 12 (2).
- Hou, E., Lu, X., Jiang, L., Wen, D., Luo, Y., 2019. Quantifying soil phosphorus dynamics: a data assimilation approach. *J. Geophys. Res. Biogeogr.* 124 (7), 2159–2173.
- Huang, X., Lu, D., Ricciuto, D.M., Hanson, P.J., Richardson, A.D., Lu, X., Weng, E., Nie, S., Jiang, L., Hou, E., Steinmacher, I.F., Luo, Y., 2021. A model-independent data assimilation (MIDA) module and its applications in ecology. *Geosci. Model Dev.* 14 (8), 5217–5238.
- Huang, Y., Zhu, D., Ciais, P., Guenet, B., Huang, Y., Goll, D.S., Guimberteau, M., Jornet-Puig, A., Lu, X., Luo, Y., 2018. Matrix-based sensitivity assessment of soil organic carbon storage: a case study from the ORCHIDEE-MICT model. *J. Adv. Model. Earth Syst.* 10 (8), 1790–1808.
- Iversen, C.M., Sloan, V.L., Sullivan, P.F., Euskirchen, E.S., McGuire, A.D., Norby, R.J., Walker, A.P., Warren, J.M., Wullschleger, S.D., 2015. The unseen iceberg: plant roots in arctic tundra. *New Phytol.* 205 (1), 34–58.
- Jones, S., Rees, R., Kosmas, D., Ball, B., Skiba, U., 2006. Carbon sequestration in a temperate grassland: management and climatic controls. *Soil Use Manag.* 22 (2), 132–142.
- Keenan, T.F., Davidson, E.A., Munger, J.W., Richardson, A.D., 2013. Rate my data: quantifying the value of ecological data for the development of models of the terrestrial carbon cycle. *Ecol. Appl.* 23 (1), 273–286.
- Kern, J.S., Turner, D.P. & Dodson, R.F. (2018). Spatial patterns of soil organic carbon pool size in the Northwestern United States. In *Soil processes and the carbon cycle*, CRC Press, pp. 29–43.
- Keuper, F., Wild, B., Kummel, M., Beer, C., Blume-Werry, G., Fontaine, S., Gavazov, K., Gentsch, N., Guggenberger, G., Hugelius, G., Jalava, M., Koven, C., Krab, E.J., Kuhry, P., Monteux, S., Richter, A., Shahzad, T., Weedon, J.T., Dorrepaal, E., 2020. Carbon loss from northern circumpolar permafrost soils amplified by rhizosphere priming. *Nat. Geosci.* 13 (8), 560–565.
- Knorr, W., Prentice, I.C., House, J., Holland, E., 2005. Long-term sensitivity of soil carbon turnover to warming. *Nature* 433 (7023), 298–301.
- Kobayashi, S., Ota, Y., Harada, Y., Ebata, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., Takahashi, K., 2015. The JRA-55 reanalysis: general specifications and basic characteristics. *J. Meteorol. Soc. Japan. Ser.* 93 (1), 5–48.
- Koven, C.D., Ringeval, B., Friedlingstein, P., Ciais, P., Cadule, P., Kholostyanov, D., Krinner, G., Tarnocai, C., 2011. Permafrost carbon-climate feedbacks accelerate global warming. *Proc. Natl. Acad. Sci.* 108 (36), 14769–14774.
- Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., Prentice, I.C., 2005. A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochem. Cycles* 19 (1).
- Kuzyakov, Y., Friedel, J.K., Stahr, K., 2000. Review of mechanisms and quantification of priming effects. *Soil Biol. Biochem.* 32 (11–12), 1485–1498.
- Lal, R., 2004. Soil carbon sequestration impacts on global climate change and food security. *Science* 304 (5677), 1623–1627.
- Lal, R., 2005. Soil erosion and carbon dynamics Vol. 81, 137–142.
- Lawrence, D.M., Fisher, R.A., Koven, C.D., Oleson, K.W., Swenson, S.C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P.J., Li, F., Li, H., Lombardozzi, D., Riley, W.J., Sacks, W.J., Shi, M., Vertenstein, M., Wieder, W.R., Xu, C., Ali, A.A., Badger, A.M., Bish, G., van den Broeke, M., Brunke, M.A., Burns, S.P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J.B., Flanner, M., Fox, A.M., Gentile, P., Hoffman, F., Keppel-Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L.R., Lipscomb, W.H., Lu, Y., Pandey, A., Pelletier, J.D., Perker, J., Randerson, J.T., Ricciuto, D.M., Sanderson, B.M., Slater, A., Subin, Z.M., Tang, J., Thomas, R.Q., Val Martin, M., Zeng, X., 2019. The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *J. Adv. Model. Earth Syst.* 11 (12), 4245–4287.
- Li, X.R., Lin, P., Li, C.X., 1992. National Soil Survey Technique of China. Agriculture Press, Beijing. In Chinese.
- Li, L., Zhang, Y., Liu, L., Wu, J., Li, S., Zhang, H., Zhang, B., Ding, M., Wang, Z., Paudel, B., 2018. Current challenges in distinguishing climatic and anthropogenic contributions to alpine grassland variation on the Tibetan Plateau. *Ecol. Evol.* 8 (11), 5949–5963.
- Liang, Z., Chen, S., Yang, Y., Zhou, Y., Shi, Z., 2019. High-resolution three-dimensional mapping of soil organic carbon in China: effects of SoilGrids products on national modeling. *Sci. Total Environ.* 685, 480–489.
- Liu, F., Zhang, G.-L., Song, X., Li, D., Zhao, Y., Yang, J., Wu, H., Yang, F., 2020. High-resolution and three-dimensional mapping of soil texture of China. *Geoderma* 361, 114061.
- Lu, D., Ricciuto, D., Walker, A., Safta, C., Munger, W., 2017. Bayesian calibration of terrestrial ecosystem models: a study of advanced Markov chain Monte Carlo methods. *Biogeosciences* 14 (18), 4295–4314.
- Lugato, E., Panagos, P., Bampa, F., Jones, A., Montanarella, L., 2014. A new baseline of organic carbon stock in European agricultural soils using a modelling approach. *Glob. Chang. Biol.* 20 (1), 313–326.
- Lugato, E., Smith, P., Borrelli, P., Panagos, P., Ballabio, C., Orgiazzi, A., Fernandez-Ugalde, O., Montanarella, L., Jones, A., 2018. Soil erosion is unlikely to drive a future carbon sink in Europe. *Sci. Adv.* 4 (11), eaau3523.

- Luo, Y., Wu, L., Andrews, J.A., White, L., Matamala, R., Schäfer, V.R., Schlesinger, W.H., 2001. Elevated CO₂ differentiates ecosystem carbon processes: deconvolution analysis of Duke Forest FACE data. *Ecol. Monogr.* 71 (3), 357–376.
- Luo, Y., Weng, E., Wu, X., Gao, C., Zhou, X., Zhang, L., 2009. Parameter identifiability, constraint, and equifinality in data assimilation with ecosystem models. *Ecol. Appl.* 19 (3), 571–574.
- Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C., LaDau, S., Clark, J.S., Schimel, D.S., 2011. Ecological forecasting and data assimilation in a data-rich era. *Ecol. Appl.* 21 (5), 1429–1442.
- Luo, Y., Keenan, T.F., Smith, M., 2015. Predictability of the terrestrial carbon cycle. *Glob. Chang. Biol.* 21 (5), 1737–1751.
- Luo, Y., Ahlström, A., Allison, S.D., Batjes, N.H., Brovkin, V., Carvalhais, N., Chappell, A., Ciais, P., Davidson, E.A., Finzi, A., Georgiou, K., Guenet, B., Hararuk, O., Harden, J. W., He, Y., Hopkins, F., Jiang, L., Koven, C., Jackson, R.B., Jones, C.D., Lara, M.J., Liang, J., McGuire, A.D., Parton, W., Peng, C., Randerson, J.T., Salazar, A., Sierra, C. A., Smith, M.J., Tian, H., Todd-Brown, K.E.O., Torn, M., Groenigen, K.J., Wang, Y.P., West, T.O., Wei, Y., Wieder, R.W., Xia, J., Xu, X., Xu, Z., Zhou, T., 2016. Toward more realistic projections of soil carbon dynamics by Earth system models. *Global Biogeochem. Cycles* 30 (1), 40–56.
- Luo, Z., Feng, W., Luo, Y., Baldock, J., Wang, E., 2017. Soil organic carbon dynamics jointly controlled by climate, carbon inputs, soil properties and soil carbon fractions. *Glob. Chang. Biol.* 23 (10), 4430–4439.
- MacBean, N., Peylin, P., Chevallier, F., Scholze, M., Schümann, G., 2016. Consistent assimilation of multiple data streams in a carbon cycle data assimilation system. *Geosci. Model Dev.* 9 (10), 3569–3588.
- Mack, M.C., Bret-Harte, M.S., Hollingsworth, T.N., Jandt, R.R., Schuur, E.A., Shaver, G. R., Verbyla, D.L., 2011. Carbon loss from an unprecedented Arctic tundra wildfire. *Nature* 475 (7357), 489–492.
- Malone, B.P., McBratney, A., Minasny, B., Laslett, G., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* 154 (1–2), 138–152.
- Matsuura, S., Kazama, R., Hibino, H., Funatsu, M., Hojito, M., 2021. Manure application in managed grasslands can contribute to soil organic carbon sequestration: evidence from field experiments across Japan. *Reg. Environ. Chang.* 21 (3), 1–12.
- McBratney, A.B., Santos, M.M., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117 (1–2), 3–52.
- Meersmans, J., De Ridder, F., Canters, F., De Baets, S., Van Molle, M., 2008. A multiple regression approach to assess the spatial distribution of Soil Organic Carbon (SOC) at the regional scale (Flanders, Belgium). *Geoderma* 143 (1–2), 1–13.
- Nachtergaele, F., van Velthuizen, H., Verelst, L., Batjes, N., Dijkshoorn, J., van Engelen, V. & Petri, M. (2008). Harmonized world soil database (Version 1.0), Food and Agric Organization of the UN (FAO). International Inst. for Applied Systems Analysis (IIASA), ISRIC-World Soil ...
- Nećpálová, M., Anex, R.P., Fienien, M.N., Del Gross, S.J., Castellano, M.J., Sawyer, J.E., Iqbal, J., Pantoja, J.L., Barker, D.W., 2015. Understanding the DayCent model: calibration, sensitivity, and identifiability through inverse modeling. *Environ. Model. Softw.* 66, 110–130.
- Niu, S., Luo, Y., Dietze, M.C., Keenan, T.F., Shi, Z., Li, J., Iii, f.s.c., 2014. The role of data assimilation in predictive ecology. *Ecosphere* 5 (5), 1–16.
- Norton, A.J., Rayner, P.J., Koffi, E.N., Scholze, M., Silver, J.D., Wang, Y.P., 2019. Estimating global gross primary productivity using chlorophyll fluorescence and a data assimilation system with the BETHY-SCOPE model. *Biogeosciences* 16 (15), 3069–3093.
- Parton, W.J., Stewart, J.W., Cole, C.V., 1988. Dynamics of C, N, P and S in grassland soils: a model. *Biogeochemistry* 5 (1), 109–131.
- Parton, W.J., Scurlock, J.M.O., Ojima, D.S., Gilmanov, T.G., Scholes, R.J., Schimel, D.S., Kirchner, T., Menaut, J.-C., Seastedt, T., García Moya, E., Kamalrutt, A., Kinyamario, J.I., 1993. Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide. *Global Biogeochem. Cycles* 7 (4), 785–809.
- Peng, S., Piao, S., Wang, T., Sun, J., Shen, Z., 2009. Temperature sensitivity of soil respiration in different ecosystems in China. *Soil Biol. Biochem.* 41 (5), 1008–1014.
- Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P., Kuppel, S., Koffi, E., Kane, A., Maingan, F., Chevallier, F., Ciais, P., Prunet, P., 2016. A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle. *Geosci. Model Dev.* 9 (9), 3321–3346.
- Piao, S., Ciais, P., Huang, Y., Shen, Z., Peng, S., Li, J., Zhou, L., Liu, H., Ma, Y., Ding, Y., 2010. The impacts of climate change on water resources and agriculture in China. *Nature* 467 (7311), 43–51.
- Piao, S., Tan, K., Nan, H., Ciais, P., Fang, J., Wang, T., Vuichard, N., Zhu, B., 2012. Impacts of climate and CO₂ changes on the vegetation growth and carbon balance of Qinghai-Tibetan grasslands over the past five decades. *Global Planet. Change* 98–99, 73–80.
- Poggio, L., De Sousa, L.M., Batjes, N.H., Heuvelink, G., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil* 7 (1), 217–240.
- Poorter, H., Sack, L., 2012. Pitfalls and possibilities in the analysis of biomass allocation patterns in plants. *Front. Plant Sci.* 3, 259.
- Pries, C.E.H., Castanha, C., Porras, R., Torn, M., 2017. The whole-soil carbon flux in response to warming. *Science* 355 (6332), 1420–1423.
- Rafique, R., Kumar, S., Luo, Y., Xu, X., Li, D., Zhang, W., Asam, Z.-u.-Z., 2014. Estimation of greenhouse gases (N₂O, CH₄ and CO₂) from no-till cropland under increased temperature and altered precipitation regime: a DAYCENT model approach. *Global Planet. Change* 118, 106–114.
- Richardson, A.D., Williams, M., Hollinger, D.Y., Moore, D.J.P., Dail, D.B., Davidson, E.A., Scott, N.A., Evans, R.S., Hughes, H., Lee, J.T., Rodrigues, C., Savage, K., 2010. Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints. *Oecologia* 164 (1), 25–40.
- Roberts, G.O., Rosenthal, J.S., 2001. Optimal scaling for various Metropolis-Hastings algorithms. *Stat. Sci.* 16 (4), 351–367.
- Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huisings, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., de Lourdes Mendonça-Santos, M., 2009. Digital soil map of the world. *Science* 325 (5941), 680–681.
- Sanz-Alonso, D., Stuart, A.M. & Taeb, A. (2018). Inverse problems and data assimilation. *arXiv preprint arXiv:1810.06191*.
- Scholze, M., Kaminski, T., Rayner, P., Knorr, W., Giering, R., 2007. Propagating uncertainty through prognostic carbon cycle data assimilation system simulations. *J. Geophys. Res. Atmos.* 112 (D17).
- Scholze, M., Buchwitz, M., Dorigo, W., Guanter, L., Quegan, S., 2017. Reviews and syntheses: systematic earth observations for use in terrestrial carbon cycle data assimilation systems. *Biogeosciences* 14 (14), 3401–3429.
- Scholze, M., Kaminski, T., Knorr, W., Voßbeck, M., Wu, M., Ferrazzoli, P., Kerr, Y., Mialon, A., Richaume, P., Rodríguez-Fernández, N., Vittucci, C., Wigneron, J.-P., Mecklenburg, S., Drusch, M., 2019. Mean European carbon sink over 2010–2015 estimated by simultaneous assimilation of atmospheric CO₂, soil moisture, and vegetation optical depth. *Geophys. Res. Lett.* 46 (23), 13796–13803.
- Schuur, E.A., McGuire, A.D., Schädle, C., Grosse, G., Harden, J.W., Hayes, D.J., Hugelius, G., Koven, C.D., Kuhry, P., Lawrence, D.M., 2015. Climate change and the permafrost carbon feedback. *Nature* 520 (7546), 171–179.
- Shangguan, W., Dai, Y., Liu, B., Zhu, A., Duan, Q., Wu, L., Ji, D., Ye, A., Yuan, H., Zhang, Q., Chen, D., Chen, M., Chu, J., Dou, Y., Guo, J., Li, H., Li, J., Liang, L., Liang, X., Liu, H., Liu, S., Miao, C., Zhang, Y., 2013. A China data set of soil properties for land surface modeling. *J. Adv. Model. Earth Syst.* 5 (2), 212–224.
- Smith, M.J., Purves, D.W., Vanderwel, M.C., Lyutkarev, V., Emmott, S., 2013. The climate dependence of the terrestrial carbon cycle, including parameter and structural uncertainties. *Biogeosciences* 10 (1), 583–606.
- Smith, B., Wärnlund, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., Zaehle, S., 2014. Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. *Biogeosciences* 11 (7), 2027–2054.
- Tan, K., Ciais, P., Piao, S., Wu, X., Tang, Y., Vuichard, N., Liang, S., Fang, J., 2010. Application of the ORCHIDEE global vegetation model to evaluate biomass and soil carbon stocks of Qinghai-Tibetan grasslands. *Global Biogeochem. Cycles* 24 (1), n/a–n/a.
- Tang, X., Zhao, X., Bai, Y., Tang, Z., Wang, W., Zhao, Y., Wan, H., Xie, Z., Shi, X., Wu, B., Wang, G., Yan, J., Ma, K., Du, S., Li, S., Han, S., Ma, Y., Hu, H., He, N., Yang, Y., Han, W., He, H., Yu, G., Fang, J., Zhou, G., 2018. Carbon pools in China's terrestrial ecosystems: new estimates based on an intensive field survey. *Proc. Natl. Acad. Sci.* 115 (16), 4021–4026.
- Tao, F., Zhou, Z., Huang, Y., Li, Q., Lu, X., Ma, S., Huang, X., Liang, Y., Hugelius, G., Jiang, L., Doughty, R., Ren, Z., Luo, Y., 2020. Deep learning optimizes data-driven representation of soil organic carbon in Earth system model over the conterminous United States. *Front Big Data* 3.
- Teng, H., Liang, Z., Chen, S., Liu, Y., Rossel, R.A.V., Chappell, A., Yu, W., Shi, Z., 2018. Current and future assessments of soil erosion by water on the Tibetan Plateau based on RUSLE and CMIP5 climate models. *Sci. Total Environ.* 635, 673–686.
- Tian, H., Lu, C., Yang, J., Banger, K., Huntzinger, D.N., Schwalm, C.R., Michalak, A.M., Cook, R., Ciais, P., Hayes, D., Huang, M., Ito, A., Jain, A.K., Lei, H., Mao, J., Pan, S., Post, W.M., Peng, S., Poulter, B., Ren, W., Ricciuto, D., Schaefer, K., Shi, X., Tao, B.O., Wang, W., Wei, Y., Yang, Q., Zhang, B., Zeng, N., 2015. Global patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: Current status and future directions. *Global Biogeochem. Cycles* 29 (6), 775–792.
- Todd-Brown, K., Randerson, J., Post, W., Hoffman, F., Tarnocai, C., Schuur, E., Allison, S., 2013. Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences* 10 (3), 1717–1736.
- Todd-Brown, K.E.O., Randerson, J.T., Hopkins, F., Arora, V., Hajima, T., Jones, C., Sheviakova, E., Tjiputra, J., Volodin, E., Wu, T., Zhang, Q., Allison, S.D., 2014. Changes in soil organic carbon storage predicted by Earth system models during the 21st century. *Biogeosciences* 11 (8), 2341–2356.
- Viscarra Rossel, R.A., Webster, R., Bui, E.N., Baldock, J.A., 2014. Baseline map of organic carbon in Australian soil to support national carbon accounting and monitoring under climate change. *Glob. Chang. Biol.* 20 (9), 2953–2970.
- Vitharana, U.W.A., Mishra, U., Mapa, R.B., 2019. National soil organic carbon estimates can improve global estimates. *Geoderma* 337, 55–64.
- Wang, B., Bao, Q., Hoskins, B., Wu, G., Liu, Y., 2008. Tibetan Plateau warming and precipitation changes in East Asia. *Geophys. Res. Lett.* 35 (14).
- Wang, C., Gao, Q., Yu, M., 2019. Quantifying trends of land change in Qinghai-Tibet Plateau during 2001–2015. *Remote Sens. (Basel)* 11 (20), 2435.
- Wang, Y.P., Kowalczyk, E., Leuning, R., Abramowitz, G., Raupach, M.R., Pak, B., van Gorsel, E., Luhar, A., 2011. Diagnosing errors in a land surface model (CABLE) in the time and frequency domains. *J. Geophys. Res. Biogeosci.* 116 (G1).
- Wang, D., Wu, T., Zhao, L., Mu, C., Li, R., Wei, X., Hu, G., Zou, D., Zhu, X., Chen, J., 2021. A 1 km resolution soil organic carbon dataset for frozen ground in the Third Pole. *Earth Syst. Sci. Data* 13 (7), 3453–3465.
- Wang, T., Yang, D., Yang, Y., Piao, S., Li, X., Cheng, G., Fu, B., 2020. Permafrost thawing puts the frozen carbon at risk over the Tibetan Plateau. *Sci. Adv.* 6 (19), eaaz3513.
- Wei, D.A., Qi, Y., Ma, Y., Wang, X., Ma, W., Gao, T., Huang, L., Zhao, H., Zhang, J., Wang, X., 2021. Plant uptake of CO₂ outpaces losses from permafrost and plant respiration on the Tibetan Plateau. *Proc. Natl. Acad. Sci. U.S.A.* 118 (33).
- Wieder, W.R., Bonan, G.B., Allison, S.D., 2013. Global soil carbon projections are improved by modelling microbial processes. *Nat. Clim. Chang.* 3 (10), 909–912.

- Wieder, W.R., Allison, S.D., Davidson, E.A., Georgiou, K., Hararuk, O., He, Y., Hopkins, F., Luo, Y., Smith, M.J., Sulman, B., Todd-Brown, K., Wang, Y.-P., Xia, J., Xu, X., 2015. Explicitly representing soil microbial processes in Earth system models. *Global Biogeochem. Cycles* 29 (10), 1782–1800.
- Wu, H., Guo, Z., Peng, C., 2003. Land use induced changes of organic carbon storage in soils of China. *Glob. Chang. Biol.* 9 (3), 305–315.
- Wu, Y., Liu, S., Huang, Z., Yan, W., 2014. Parameter optimization, sensitivity, and uncertainty analysis of an ecosystem model at a forest flux tower site in the United States. *J. Adv. Model. Earth Syst.* 6 (2), 405–419.
- Wu, M., Scholze, M., Kaminski, T., Voßbeck, M., Tagesson, T., 2020. Using SMOS soil moisture data combining CO₂ flask samples to constrain carbon fluxes during 2010–2015 within a Carbon Cycle Data Assimilation System (CCDAS). *Remote Sens. Environ.* 240, 111719.
- Xiao, J., Davis, K.J., Urban, N.M., Keller, K., 2014. Uncertainty in model parameters and regional carbon fluxes: a model-data fusion approach. *Agric. For. Meteorol.* 189, 175–186.
- Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J.A., Huete, A.R., Ichii, K., Ni, W., Pang, Y., Rahman, A.F., Sun, G., Yuan, W., Zhang, Li., Zhang, X., 2019. Remote sensing of the terrestrial carbon cycle: a review of advances over 50 years. *Remote Sens. Environ.* 233, 111383.
- Xu, T., White, L., Hui, D., Luo, Y., 2006. Probabilistic inversion of a terrestrial ecosystem model: analysis of uncertainty in parameter estimation and model prediction. *Global Biogeochem. Cycles* 20 (2), n/a–n/a.
- Xu, H., Zhang, T., Luo, Y., Huang, X., Xue, W., 2018. Parameter calibration in global soil carbon models using surrogate-based optimization. *Geosci. Model Dev.* 11 (7), 3027–3044.
- Yang, Y., Fang, J., Ma, W., Smith, P., Mohammat, A., Wang, S., Wang, W.E.I., 2010. Soil carbon stock and its changes in northern China's grasslands from 1980s to 2000s. *Glob. Chang. Biol.* 16 (11), 3036–3047.
- Yao, T., Thompson, L., Yang, W., Yu, W., Gao, Y., Guo, X., Yang, X., Duan, K., Zhao, H., Xu, B., Pu, J., Lu, A., Xiang, Y., Kattel, D.B., Joswiak, D., 2012. Different glacier status with atmospheric circulations in Tibetan Plateau and surroundings. *Nat. Clim. Chang.* 2 (9), 663–667.
- Yang, Y., Fang, J., Tang, Y., Ji, C., Zheng, C., He, J., Zhu, B., 2008. Storage, patterns and controls of soil organic carbon in the Tibetan grasslands. *Glob. Chang. Biol.* 14 (7), 1592–1599.
- You, Q., Min, J., Kang, S., 2016. Rapid warming in the Tibetan Plateau from observations and CMIP5 models in recent decades. *Int. J. Climatol.* 36 (6), 2660–2670.
- You, C., Yao, T., Xu, C., 2018. Recent increases in wildfires in the Himalayas and surrounding regions detected in central Tibetan ice core records. *J. Geophys. Res. Atmos.* 123 (6), 3285–3291.
- Zhang, W., Zhang, F., Qi, J., Hou, F., 2017. Modeling impacts of climate change and grazing effects on plant biomass and soil organic carbon in the Qinghai-Tibetan grasslands. *Biogeosciences* 14 (23), 5455–5470.
- Zhang, L., Zhou, G., Ji, Y., Bai, Y., 2016. Spatiotemporal dynamic simulation of grassland carbon storage in China. *Sci. China Earth Sci.* 59 (10), 1946–1958.
- Zhou, Y., Webster, R., Rossel, R.V., Shi, Z., Chen, S., 2019. Baseline map of soil organic carbon in Tibet and its uncertainty in the 1980s. *Geoderma* 334, 124–133.

An improved estimate of soil carbon pool and carbon fluxes in the Qinghai-Tibetan grasslands using data assimilation with an ecosystem biogeochemical model

Zhao, Ruiying; Zhang, Wenxin; Duan, Zheng; Chen, Songchao; Shi, Zhou

- 01 mingxi zhang Page 2
23/11/2023 1:02
- 02 mingxi zhang Page 2
19/9/2024 2:33
- 03 mingxi zhang Page 3
23/11/2023 1:30
- 04 mingxi zhang Page 3
29/1/2024 2:53
- 05 mingxi zhang Page 3
29/1/2024 2:53
- 06 mingxi zhang Page 4
29/1/2024 3:00
- 07 mingxi zhang Page 4
1/3/2024 7:22
- 08 mingxi zhang Page 4
1/3/2024 7:27

09 mingxi zhang

Page 6

1/3/2024 7:50

10 mingxi zhang

Page 6

4/4/2024 9:12

11 mingxi zhang

Page 6

12/3/2024 5:46

12 mingxi zhang

Page 10

29/1/2024 2:24

13 mingxi zhang

Page 11

29/1/2024 3:05

14 mingxi zhang

Page 11

19/9/2024 2:54

15 mingxi zhang

Page 13

29/1/2024 3:08

16 mingxi zhang

Page 13

29/1/2024 3:08

17 mingxi zhang

Page 13

29/1/2024 3:09