CHAPTER THIRTY-SEVEN

PROcess-Guided Deep Learning and DAta-Driven Modelling (PRODA)

Feng Tao
Tsinghua University, Beijing, China

Yiqi Luo
Cornell University, Ithaca, USA

CONTENTS

This chapter describes a PROcess-guided deep learning and DAta-driven modeling (PRODA) approach to optimize parameterization of Earth system models (ESMs) using spatio-temporal datasets. PRODA involves both data assimilation to estimate parameter values and deep learning to predict spatial and temporal distributions of parameter values so as to optimize ESM prediction. An application to the Community Land Model version 5 (CLM5) using soil organic carbon (SOC) distributions in the conterminous United States illustrates the potential and utility of the PRODA approach.

THE NEED FOR OPTIMIZING PARAMETERIZATION OF EARTH SYSTEM MODELS

Earth system models (ESMs) are used to simulate historical and potential future states of climate and ecosystems. However, simulations often deviate substantially from observations. For example, soil carbon dynamics simulated by ESMs vary widely among models and often fit poorly with observations. Modeled global soil carbon storage differs by up to six-fold among 11 models of the Coupled Model Intercomparison Project Phase 5 (CMIP5) ensemble (Todd-Brown et al. 2013). None of the models reproduces the spatial distribution of SOC stocks presented in the Harmonized World Soil Database (HWSD) (Luo et al. 2015).

Uncertainty in simulating SOC dynamics with ESMs could stem from poor parameterization, incorrect model structure, or biased external forcing (Luo and Schuur 2020, chapter 33). While model structure represents ecological processes (e.g., decomposition of soil organic matter), parameters in ESMs characterize properties of the processes, such as baseline decomposition rate at reference temperature and moisture content, or sensitivity to these drivers. The choice of parameter values can strongly influence model projections

DOI: 10.1201/9780429155659-47

of SOC dynamics. Parameter values in the current generation of ESMs, however, are mostly determined on an *ad hoc* basis. They may be derived from the results of field experiments, other models, or informed from scientific or grey literature (Luo et al. 2001), but rarely take into account the range of possible values encompassed by such sources.

Data assimilation techniques to estimate parameter values from observations were discussed and illustrated in earlier chapters (units 6, 7, 8). Parameter values constrained by data assimilation can improve SOC simulation in ESMs compared to the default parameter values. For instance, the global representation of SOC distribution in the Community Land Model version 3.5 (CLM3.5) was improved from explaining 27 to 41% of variation in the HWSD database by constraining model parameters with a Bayesian Markov Chain Monte Carlo (MCMC) data assimilation method (Hararuk et al. 2014). The large unexplained variation in observed SOC with ESMs is partly due to a textbook concept that parameter values of a simulation model must be constant in contrast to variables that can vary over the time course of simulation (Forrester 1961). In reality, ecosystem properties, which parameters characterize in models, constantly evolve via acclimation and adaptation. In addition, a model, no matter how complex it is, can never represent all the processes of a system at resolved scales (Luo and Schuur 2020). Interactions of processes at unresolved scales with those at resolved scales should be reflected in model parameters. Therefore, Luo and Schuur (2020) argue that parameter values in ESMs may have to vary over space and time (i.e., heterogeneous parameter values) to represent changing properties of evolving ecosystems and unresolved processes.

The advent of big ecological data provides a golden opportunity to reconcile model representations with observations and quantify the spatial and temporal features of key parameters in soil carbon cycle simulation. Meanwhile, new techniques such as deep learning have been proposed to improve performance of ESMs (Reichstein et al. 2019). By constructing computational models with multiple processing layers and allowing the models to learn representations of data from multiple levels of abstraction (LeCun et al. 2015), deep learning techniques have promising applications in Earth system science, such as pattern classification, anomaly detection, regression, and space- or time-dependent state prediction (Reichstein et al. 2019). Exploration is warranted on how to properly employ deep learning techniques in reducing uncertainties of simulated carbon dynamics in ESMs.

Here, we propose the PROcess-guided deep learning and DAta-driven modelling (PRODA) approach to estimate spatially and temporally heterogeneous parameter values for ESMs from extensive spatio-temporal datasets ('big data') at regional or global scales. The PRODA approach estimates parameter values at individual sites via data assimilation and builds a deep learning model to upscale the site-level estimates of parameters to predict spatially heterogeneous parameters at regional and global scales so that modeled and observed SOC are maximally matched.

In this chapter, we introduce the PRODA approach by using an extensive dataset of vertical soil profiles across the conterminous United States to optimize SOC representation by CLM5. We discuss the PRODA-optimized model performance in representing SOC stock and its vertical and spatial distributions, and compare it with results of the default model simulation and after the data assimilation optimization. In particular, we highlight that the PRODA approach helps the process model to achieve the most precise SOC distribution ever represented in ESMs. An accurate SOC representation in ESMs is critical to fully understand soil carbon feedbacks to future climate change.

THE WORKFLOW OF PRODA

Three fundamental components together formulate the PRODA approach (Figure 37.1a), namely the process-based model, the site-level data assimilation, and the deep learning model. Process-based models with their predefined structure and default parameter values simulate SOC distributions using meteorological forcing data. Data assimilation is used to estimate parameter values of a process-based model with soil carbon data at sites where the observations were made. The deep learning model is used to predict optimized site-level parameter values with their associated environmental variables. Eventually, the process-based model will apply the optimized parameter values upscaled by the deep learning model to simulate SOC distributions at regional or global scales.

Process-based model: We use the matrix representation of the Community Land Model version 5 (CLM5) to facilitate data assimilation and model simulation in the PRODA approach (Figure 37.1b). CLM5 is the latest version of CLM

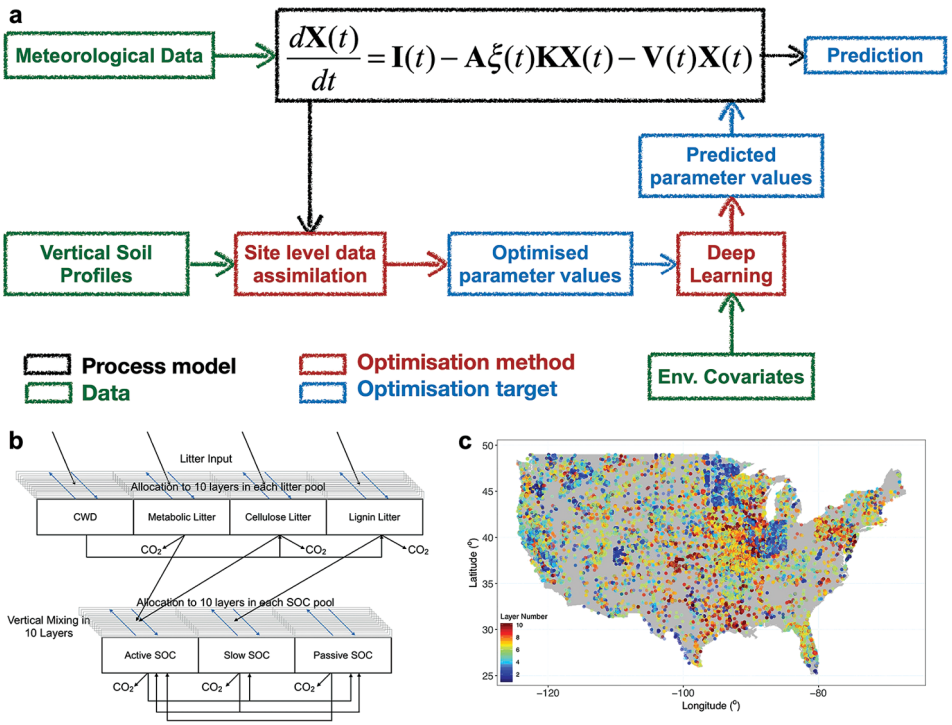

Figure 37.1. Workflow of the PRODA approach. (a) PRODA optimally matches CLM5 as the process-based model (b) with vertical SOC profiles on the conterminous United States (c). We first assimilate data at each site into CLM5 to estimate its parameters through the Markov Chain Monte Carlo method (MCMC). We further assemble the estimated site-level parameter values (i.e., the mean value of the posterior distribution after MCMC) as targets to be predicted by a multilayer neural network with environmental covariates in a deep learning model. The predicted parameters by the deep learning model are applied to CLM5 to optimize model representation of SOC distribution.

models (Lawrence et al. 2019). Its soil carbon module is similar to that in CLM4.5 (Koven et al. 2013), except that it has an option to change the number of soil layers from a default of 20. In this example, we use ten soil layers with a vertical transformation among carbon pools from the surface to a maximum depth of 3.8 m as in CLM4.5. The soil carbon component of CLM5 includes carbon transfer among four litter pools (coarse woody debris, metabolic litter, cellulose litter, and lignin litter) and three soil organic carbon pools (fast, slow, and passive SOC) in each of ten layers, totaling 70 pools. The thickness of soil layers increases exponentially from the surface layer (1.75 cm) to deep layers (151 cm), with a total depth of 3.8 m over the ten layers. Vertical carbon transfer between soil layers only occurs among the adjacent layers and represents both diffusive and advective carbon flux transportation caused by bioturbation and cryoturbation. The baseline advective rate of carbon flux is set to zero in CLM5 as a default, and this is assumed in our example as well.

We have discussed in units 1–5 that carbon balance equations in land carbon models can be unified to a matrix form. For CLM5, we use the matrix equation to describe carbon transfer among the 70 pools with state variables $\mathbf{X}$ (t) as:

$$\frac{d\mathbf{X}(t)}{dt} = \mathbf{B}u(t) + \mathbf{A}\xi(t)\mathbf{K}\mathbf{X}(t) - \mathbf{V}(t)\mathbf{X}(t) \quad (37.1)$$

where $\mathbf{B}$ is a vector (70×1) of partitioning coefficients from C input to each of the pools (unitless), and u(t) is C input rate (gC m^{-3} day^{-1}). $\boldsymbol{A}$ represents the transfer coefficients among litter and soil pools (unitless), including the transfer coefficients from four litter pools to three soil carbon pools as well as the transfer coefficients of SOC among soil carbon pools in the same layer. ξ(t) represents effects of environmental variables on decomposition of litter and soil (unitless). It includes scalars of temperature (ξ_T), soil water (ξ_W), oxygen (ξ_O), nitrogen (ξ_N), and depth (ξ_D). $\mathbf{K}$ indicates

the decomposition rate of SOC in different litter and soil carbon pools (day^{-1}). $\boldsymbol{V}(t)$ represents SOC mixing among vertical soil layers through cryoturbation or bioturbation (day^{-1}). The t in parentheses indicates that the corresponding element is time-dependent. At a steady-state of the carbon cycle ($\frac{d\mathbf{X}(t)}{dt} = 0$), the SOC content of each carbon pool at each layer can be calculated as:

$$\mathbf{X}(t) = \left[\boldsymbol{A}\xi(t)\mathbf{K} - \boldsymbol{V}(t)\right]^{-1}\left(-\mathbf{B}u(t)\right) \quad (37.2)$$

Soil carbon data and site-level data assimilation: We use vertical SOC profiles in the conterminous U.S. from the World Soil Information Service (WoSIS) dataset (www.isric.org) for the site-level data assimilation (Figure 37.1c). The depth of recorded SOC layers ranges from the surface to more than 3 metres. A total of 26,509 soil profiles with a total of 240,148 layers at different depths in the conterminous U.S. are available in this study.

In addition, we use the mean values of global net primary productivity (NPP) from 2000 to 2014 as carbon input (DAAC 2018). After running the CLM4.5 model to a steady-state by the pre-industrial climate forcing (version code of forcing database: I1850CRUCLM45BGC), ten-year records of soil temperature and soil water potential of the conterminous U.S. were obtained from the model outputs.

The site-level data assimilation constrains parameter values of CLM5 with one data set of a vertical SOC profile at each site with the Markov chain Monte Carlo (MCMC) method (as described in chapter 22). Three parallel chains are generated each containing a test run of 20,000 iterations and a formal run of 30,000 iterations. To effectively capture the vertical distribution pattern of soil content along the depths, we put weights to observations at different depths in calculating the discrepancy between modeled and observed SOC content (i.e., cost function). These weights decrease exponentially with the depth (i.e., $weight_i = e^{-|depth_i|}$, where i refers to the layer's soil depth in observations) except for the top layer and the bottom layer, where a weight of ten is assigned to accelerate calibrating the upper and lower bounds of the SOC distribution curve. To monitor the efficiency of the MCMC process, an acceptance rate threshold is set. For Markov chains whose acceptance rate is higher than 50% or lower than 15%, the corresponding data assimilation results are rejected. After the MCMC process, the first half of the accepted parameter values in the formal run are discarded as burn-in. The Gelman-Rubin statistics of each parameter are then calculated for each soil profile to ensure the convergence of these three independent MCMC results. We randomly select one chain after eliminating the burn-in period to generate the posterior distributions of parameters. The mean value of the parameter's posterior distribution is calculated and chosen to serve as the training target in the deep learning model.

We evaluate the effectiveness of the site-level data assimilation by the coefficient of efficiency:

$$E = 1 - \frac{\sum\left(obs_i - \mathrm{mod}_i\right)^2}{\sum\left(obs_i - \overline{obs}\right)^2} \quad (37.3)$$

where obs_i and mod_i are the observed and modeled SOC content at ith soil layer of one soil profile; $\overline{obs}$ is the mean value of observed SOC content of the soil profile. In this study, we take profiles having negative E values as invalid and discard the results from the corresponding deep learning model. Moreover, at those sites where an observation is available at only one soil depth, we do not apply the data assimilation to the data point. After those data sets are excluded, 25,444 out of 26,905 soil profiles, or 94.6% of the entire dataset, are used in the PRODA approach.

Deep learning model: We design a deep learning model with multiple processing layers to predict optimized parameter values with environmental covariates. A total of 60 environmental variables that describe the climatic, edaphic and vegetation features at the observational sites is used. We used 80% of the total dataset to train and validate the neural network. After model training, we use the remaining 20% of the dataset to quantify the prediction accuracy of the deep learning model. The predicted parameter values are first compared with those retrieved in site-level data assimilation and then applied to the matrix CLM5 model to simulate soil organic carbon stock at each observational site. Meanwhile, we used the trained deep learning model to generate parameter maps across the United States based on gridded environmental covariates. The parameter maps are then applied to the matrix CLM5 to simulate the SOC distributions across the United States at a resolution of 0.5 degrees.

SOC distributions optimized by data assimilation: To analyse the significance of the spatially-explicit parameter estimation of PRODA compared with a traditional approach, we perform a batch data assimilation using all the observational dataset as one batch in the MCMC method to estimate parameter values of CLM5 with data assimilation. The estimated parameter values from this method are spatially homogeneous, in contrast with the site-level data assimilation, which is a middle step of the PRODA approach to estimate spatially heterogeneous parameter values. SOC distributions simulated by CLM5, trained by the batch data assimilation versus the results by PRODA, can then be compared.

The batch data assimilation runs three parallel MCMC chains, each containing 50,000 iterations as test run and 200,000 iterations as formal run. Weights at different depth in calculating the cost function and acceptance control are the same as those in the site-level data assimilation. After the MCMC method, we first discard the first half of the accepted parameter values of the formal run as burn-in. The Gelman-Rubin statistics for each parameter are then calculated to ensure the convergence of these three independent MCMC results. We randomly select one Markov chain after eliminating the burn-in period to generate the posterior distribution for each parameter. We then randomly sample parameter values from the posterior distributions 1,000 times and apply the sampled parameter values to the CLM5 matrix model. We estimate SOC content distributions at different sites by calculating the average of the results. The same sampled parameter values are further assigned in CLM5 to estimate SOC content distributions at each grid cell on the map of the conterminous US at a resolution of 0.5 degrees.

Reference SOC data products: We use two sets of SOC data, WISE30sec and SoilGrids250m (Hengl et al. 2017), as references to compare with spatial and vertical distributions of SOC obtained from our study over the United States. WISE30sec is an updated version of the dataset HWSD, generated by using traditional mapping methods at a resolution of 30 × 30 arc sec. SoilGrids250m is a global gridded soil information dataset generated by using machine learning techniques at 250 m resolution. We took data of SOC content over three depth intervals from these two datasets, 0–30 cm, 0–100 cm and 0–200 cm. All the original data with high resolution were resampled to a resolution of 0.5 × 0.5 degrees.

MODEL REPRESENTATION OF SOC CONTENT ACROSS OBSERVATION SITES

The original CLM5 model with default parameterization presents significant geographical biases on the estimation of SOC content in comparison with observations. Modeled SOC in the grid cell in which the site of observation was located is compared with observations (Figure 37.2a). SOC storage is systematically overestimated by the original model near the east and west coasts of the U.S. but underestimated in the Midwest. The consistency between observed and modeled SOC content is low, with $R^2 = 0.32$ and RMSE = 15.9 kgC m^{-3} (Figure 37.2b and Table 37.1).

The batch data assimilation method generates the distribution of SOC from continentally homogeneous posterior distributions of parameters estimated from all the observation data at once in data assimilation. With the batch data assimilation, the mismatch between observed and modeled SOC content in the CLM5 model is moderately reduced in the north and east parts of the U.S. (Figure 37.2c). However, geographical biases in model representation of SOC are not eliminated. CLM5 optimized by the batch data assimilation still underestimates SOC storage in the Intermontane Plateaus and southern Great Plains. Meanwhile, overestimation still exists in the Great Lakes areas and the Northeast. Overall, CLM5 after optimization by the batch data assimilation explains 43% variation in the observed SOC content with RMSE = 11.4 kg C m^{-3} (Figure 37.2d and Table 37.1).

Through the deep learning model, the PRODA approach predicts the optimized parameter values at each site across the conterminous U.S. by its environmental variables. PRODA-optimized CLM5 achieves a better representation of SOC distribution compared to the batch data assimilation. Little systematic geographical biases in estimating SOC storage are observed across the study domain (Figure 37.2e). The modeled and observed SOC content are highly correlated with $R^2 = 0.62$ and RMSE = 9.0 kg C m^{-3} (Figure 37.2f and Table 37.1).

SPATIAL DISTRIBUTION OF SOC ACROSS THE CONTERMINOUS U.S.

We take point observations (Figure 37.3a–c) and estimations from WISE30sec (Figure 37.3d–f) and SoilGrids250m (Figure 37.3g–i) as references

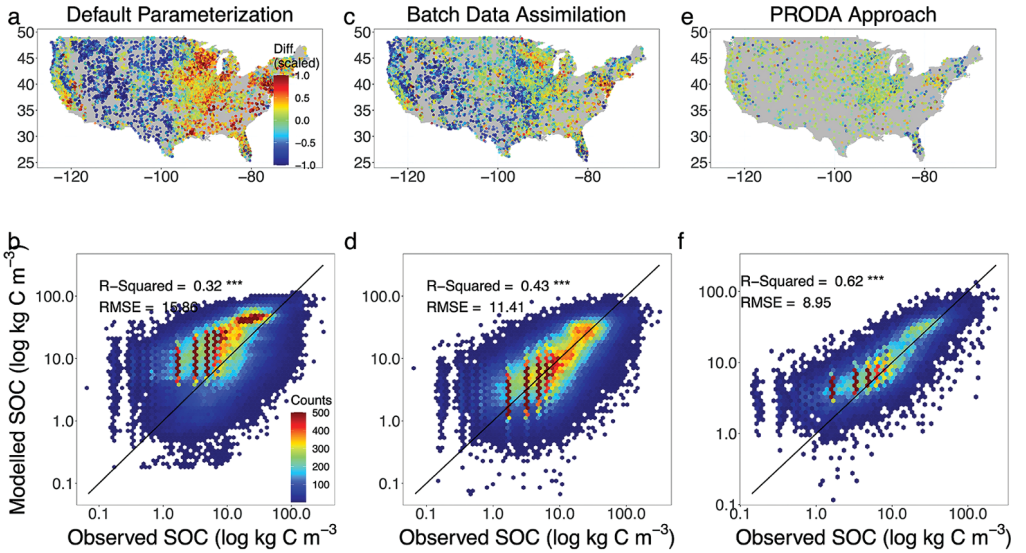

Figure 37.2. The agreement between observed and modeled SOC content with different approaches. SOC estimates modeled by CLM5 were extrapolated to the depths of observations to evaluate model performance. The upper panel indicates the deviation of the modeled SOC storage from the observation of the whole profile for each site. The lower panel shows the results of linear regression between observed and modeled vertical SOC content at different depths in different methods. In calculating the deviation of modeled SOC storage from observations, for better presentation, the positive (overestimation) and negative (underestimation) discrepancy between the observed and modeled SOC content were scaled based on the 95% quantile of the positive discrepancy and 5% quantile of the negative discrepancy, respectively. Meanwhile, only the results of the testing set were presented in PRODA approach.

TABLE 37.1
Performance of CLM5 in representing SOC distribution under different approaches

	Model Performance	
Method	R^2	RMSE (kg C/m³)
Default CLM5	0.32	15.86
Batch Data Assimilation	0.43	11.41
PRODA Approach	0.62	8.95

Note: R^2 is the coefficient of determination from linear regression between the observed and modeled SOC content. RMSE is the root mean square error.

to compare the SOC estimations by CLM5 with default parameterization, optimized parameterization after the batch data assimilation, and the PRODA approach. At the continental scale, the reference data suggest large volumes of SOC in the northeast and northwest of the conterminous U.S. The magnitude of SOC content in these regions can be as high as 30 kg C m^{-2} for the 0–200 cm depth interval. Meanwhile, a decreasing gradient of SOC from the northeast to the southwest is observed. High SOC exists in areas across the Great Plains, extending from Texas to the Great Lakes.

The default CLM5 model (Figure 37.3j–l) captures the continental SOC content gradient from the northeast to the southwest but fails to reproduce sub-regional features of SOC distribution in the Great Plains. Meanwhile, SOC content in the east and northwest estimated by the original CLM5 is significantly higher than that indicated by the reference data. After optimization by the batch data assimilation, CLM5 reproduces the continental SOC gradient from the northeast to the southwest with reasonable values (Figure 37.3m–o). However, high SOC content in the Great Plains is still not well represented. The PRODA approach performs best overall, helping achieve the most realistic spatial SOC distribution (Figure 37.3p–r) in comparison with observations (Figure 37.3a–c) and data products (Figure 37.3d–i). In addition to capturing the continental SOC distribution pattern, the PRODA-optimized CLM5 presents more accurate subregional SOC distribution patterns in the Great Plains.

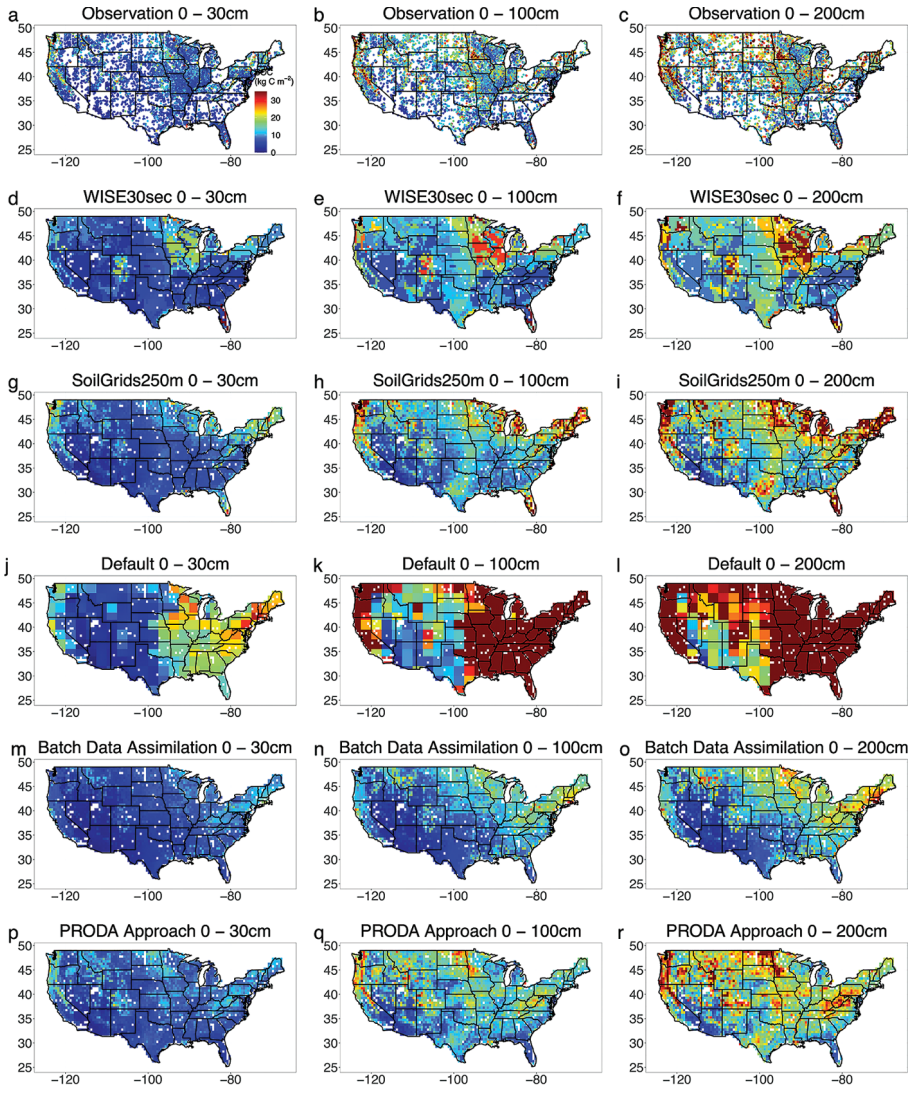

Figure 37.3. Modeled spatial SOC distributions in three depth intervals across the conterminous U.S. by different approaches and datasets.

VERTICAL DISTRIBUTION OF SOC ACROSS THE CONTERMINOUS U.S.

We take results from WISE30sec and SoilGrids250m as references in estimated SOC stocks at different depth intervals (Figure 37.4). For the first 2-meter soil, WISE30sec suggests 243 PgC and SoilGrids250m estimates 269 PgC stored as SOC. Along the soil depth, WISE30sec suggests 98 PgC at 0–30 cm depth, 81 PgC at 30–100 cm, and 64 PgC at 100–200 cm. SoilGrids250m estimates 102, 86, and 81 PgC at the same three depth intervals, respectively.

The original CLM5 model with default parameterization substantially overestimates SOC stocks in comparison with the references at all three soil depths (Figure 37.4). Compared with the references, the overestimation becomes stronger with increasing soil depth. Both the batch data assimilation and the PRODA approach help CLM5 estimate more reasonable SOC storage compared with the original CLM5 model. We estimate 165 PgC using the batch data assimilation and 246 PgC for the first 2-meter soils using the PRODA approach.

For different vegetation types, the PRODA approach presents more accurate estimations of the vertical SOC distribution than the batch data assimilation (Figure 37.5). CLM5 underestimates the SOC content in the evergreen forest, shrubland,

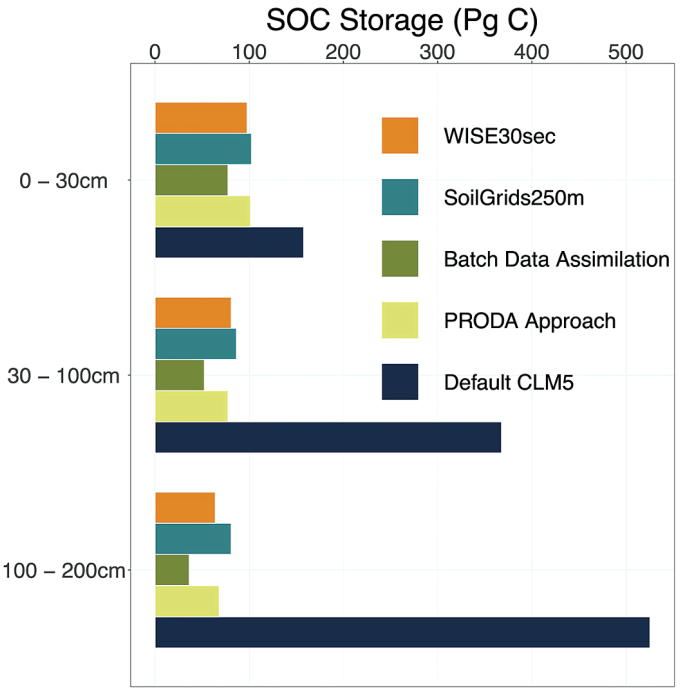

Figure 37.4. SOC storage across the conterminous U.S. at different depths estimated by different approaches and data sources.

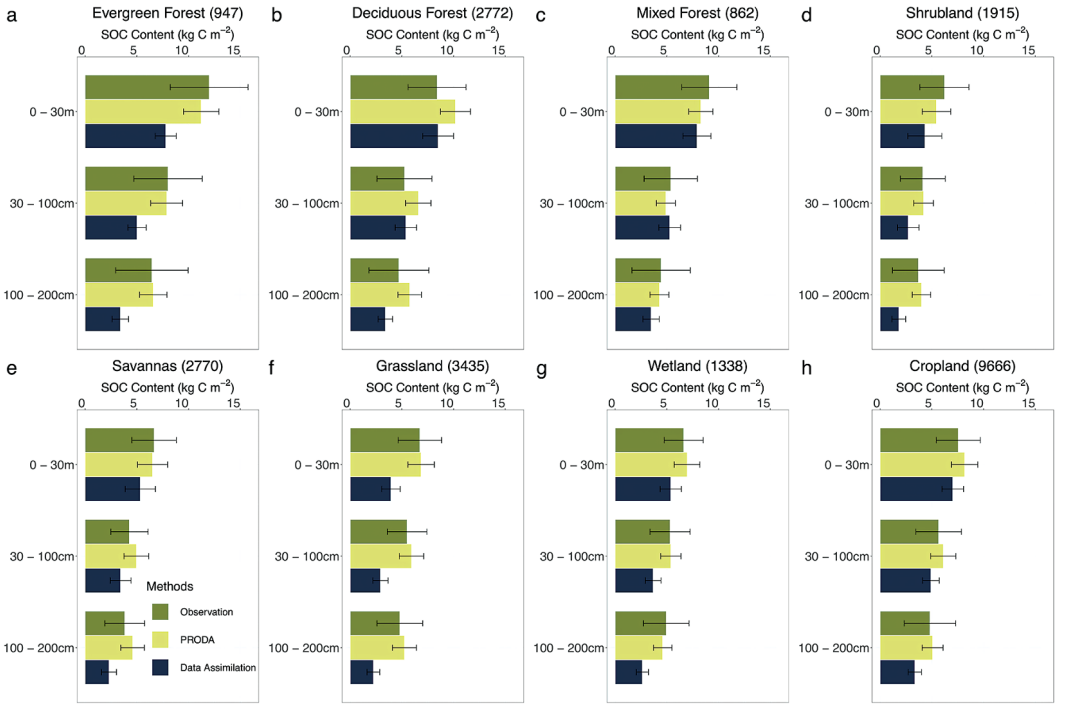

Figure 37.5. SOC storage for different vegetation types across the conterminous U.S. at different depths estimated by different approaches and data sources. The number in parentheses after vegetation type is the number of sites with that vegetation in the dataset used in this study. The error bars indicate ±0.5 standard deviation.

savanna, grassland and wetland regions after the optimization by the batch data assimilation. The PRODA-optimized CLM5, in contrast, presents the least biased estimations in comparison with observations at all depth intervals in the aforementioned regions.

TOWARD MORE REALISTIC REPRESENTATIONS OF SOC DISTRIBUTION

This chapter has systematically explored the significance of spatially heterogeneous parameterization for the adequate prediction of SOC distribution in Earth system models, with CLM5 as a representative case. The results support the PROcess-guided deep learning and DAta-driven modelling (PRODA) as a promising approach to optimize model representation of SOC, utilising the explanatory power implicit in immense observational data. PRODA considers biogeochemical processes in the soil carbon cycle while preserving strong big data analysis ability to integrate soil data into complex models. We compared the PRODA-optimised SOC representation by CLM5 with the default model simulation and the results optimized by batch data assimilation and conclude that PRODA helped CLM5 achieve the most accurate SOC representation. Indeed, no better fit to reference data on SOC has ever been simulated by process-based models.

In the past decades, different approaches have been developed for representation of SOC distribution (Figure 37.6). Soil scientists collect soil data and develop mechanistic understanding of soil carbon cycling from field observations or experiments. The simulation modeling approach conceptualizes those mechanisms into mathematical equations and strives to simulate SOC according to process understanding. Notwithstanding the detailed description of carbon cycle processes, the models struggle to realistically simulate SOC distribution. Such unrealistic model simulations mainly arise from inadequate parameterization. Parameters that represent critical processes of the soil carbon cycle in the real world are not sufficiently constained with widely distributed observational data. Therefore, it is difficult for process-based models to accurately represent SOC distributions. In our example, CLM5 with default parameter values substantially overestimates the total SOC storage of the conterminous U.S. and presents strong geographical biases in the representation of SOC distribution.

Batch data assimilation provides a way of incorporating observational data information into the process model to improve SOC simulation. Such data-driven optimization harmonizes site-level data information as a whole to adjust the parameter values for better representation of the SOC. We

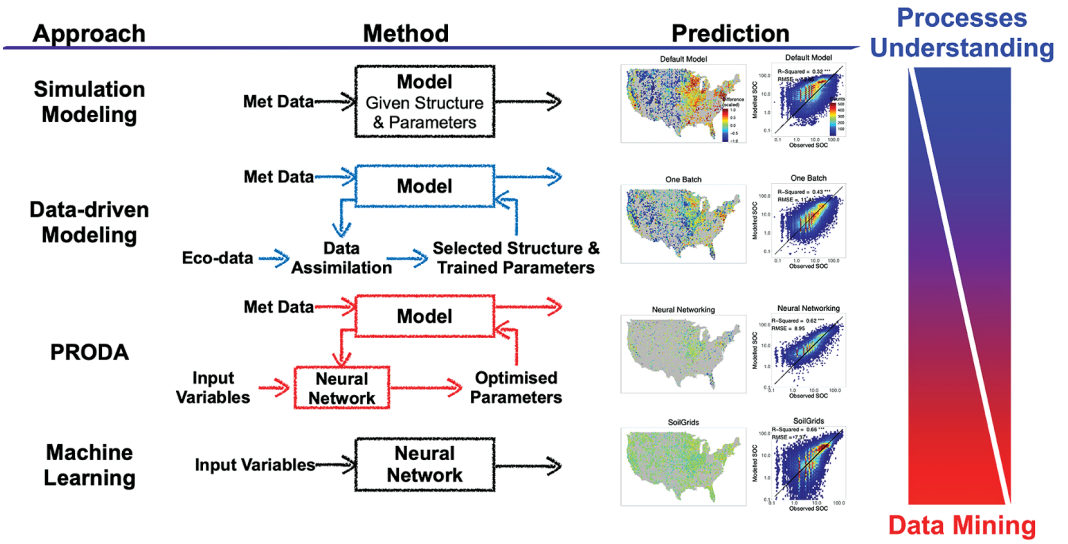

Figure 37.6. Schema of different approaches to represent SOC distributions. The PRODA approach benefits from both process understanding (as featured by simulation modelling) with the real-world information brought out by big data analysis from machine learning. The latter is primarily to obtain accurate representations of the spatial distribution of SOC and its underlying mechanisms.

have shown in the example study that the optimized CLM5 with data assimilation successfully corrects the considerable overestimation of total carbon storage across our study domain.

In terms of representing the spatial variability of SOC, however, batch data assimilation fails to capture the spatial variability of observed SOC. The spatially invariant parameter values optimized from the batch data assimilation approach are insufficient in describing the heterogeneity of SOC distribution at large scales. In our example study, geographical bias still exists after the optimization by the batch data assimilation.

The PRODA approach solves the issue of geographical bias by using a deep learning model to first fully estimate parameters at the site level using the data assimilation and then upscales the site-level estimates of parameters to the whole U.S. continent. The spatially varying parameter values retrieved from the PRODA approach contribute to a more accurate model representation of SOC across the range of ecosystem types (vegetation class, soil type, geology etc) across the continent. PRODA-optimized CLM5 simulates the most realistic SOC distribution ever simulated by process models. The high agreement between observed and modeled SOC content ($R^2 = 0.623$ across the conterminous U.S.) achieved by the PRODA approach is comparable with that for harmonization mapping in SoilGrids250m by machine learning ($R^2 = 0.635$ across the globe) (Hengl et al. 2017), and greater than the agreement between separate gridded empirical data products (Wu et al. 2019).

More importantly, the PRODA approach paves the way for more mechanistic understanding of the soil carbon cycle from big data analysis with machine learning. Machine learning alone is good at accurately describing SOC distribution, yet previous applications used in digital soil mapping focus only on the complex statistical relationship between environmental variables and SOC. The PRODA approach not only precisely maps SOC distributions but also provides the spatial patterns of different mechanisms (as represented by different parameters) of the soil carbon cycle. In the future, disentangling how these mechanisms vary with environments and quantifying their importance to SOC storage will be essential for understanding terrestrial carbon dynamics and their feedbacks to climate change.

SUGGESTED READING

Tao, F., Z. Zhou, Y. Huang, Q. Li, X. Lu, S. Ma, X. Huang, Y. Liang, G. Hugelius, L. Jiang, R. Doughty, Z. Ren, and Y. Luo. 2020. Deep Learning Optimizes Data-Driven Representation of Soil Organic Carbon in Earth System Model Over the Conterminous United States. *Frontiers in Big Data* 3, 17.

QUIZZES

1. What is the main difference, in terms of parameterization scheme, between the batch data assimilation and the PRODA approach as described in this chapter?
2. Describe the input and output of the deep learning model in the PRODA approach?
3. What is the advantage of the PRODA approach in comparison to conventional machine learning methods in representing SOC distributions?