RC2: 'Comment on essd-2024-588', Anonymous Referee #2, 27 Mar

1. Feature optimization? I think it should be feature selection. Yet, random forest (RF) may represent extremely complicated nonlinear relationship between SOCD and their drivers (i.e., the covariates that you used), why did you select features based on Pearson correlation coefficients? Besides, RF is some insensitive to feature selection!

We sincerely appreciate the reviewer's thoughtful comments and constructive suggestions, which have helped us significantly improve our methodology and manuscript. In response to the reviewer's concerns regarding feature selection, we have carefully revised our approach and provided detailed explanations below.

Regarding the initial use of Pearson correlation for feature pre-screening, we implemented this step primarily for computational efficiency when handling our large spatial dataset. While we fully acknowledge that random forest can capture complex nonlinear relationships, the correlation screening served as an effective first pass to remove clearly irrelevant variables (where $|r| \approx 0$) and eliminate strongly redundant predictors ($|r| > 0.95$). This preprocessing step proved particularly valuable in reducing computational burden while maintaining model performance, as linear relationships often underlie more complex nonlinear patterns that the subsequent random forest analysis could capture.

To address the reviewer's important point about random forest's relative insensitivity to irrelevant features, we have strengthened our methodology in several key ways. First, we employed out-of-bag error reduction for more robust importance ranking, focusing specifically on features that demonstrably improve predictive accuracy. Second, rather than relying solely on individual feature scores, we conducted exhaustive combinatorial testing of all possible feature subsets from the correlation-filtered set. This approach ensured we identified the optimal combination of features that collectively maximized predictive performance, as measured by R^2 in cross-validation. Finally, we validated the selected feature set using independent test sets to confirm its robustness and generalizability. The revised methodology yielded several important improvements. We removed marginally contributing variables such as AH and CLCD to create a more parsimonious model. The final selected features - including mean annual temperature (Tem), elevation, NDVI, clay content (Clay), simple ratio index (SR), bare soil index (BSI), and slope.

These revisions have significantly strengthened our methodology while maintaining its computational efficiency and ecological interpretability. The refined approach provides a more rigorous and transparent feature selection process that balances predictive power with model parsimony. We believe these improvements thoroughly address the reviewer's concerns and have resulted in a more robust study. We are grateful for the reviewer's insightful comments that have led to these important enhancements in our work.

2. The use of climate zone in this study is really unnecessary since the temperature and precipitation that you used to define the climate zone have already used as features in your RF model!

Thank you for your constructive comment regarding the use of climate zones in our study, particularly your point that temperature and precipitation, used for defining climate zones, were also features in our initial Random Forest (RF) model. We appreciate you highlighting this potential redundancy.

We'd like to clarify our approach and the refinements made during our model development. Through an improved and rigorous feature selection method, we've refined our optimal feature set to comprise seven variables, consisting of mean annual temperature (Tem), elevation, NDVI, clay content (Clay), simple ratio index (SR), bare soil index (BS1), and slope.

It's important to note that after this refinement, only mean annual temperature is now directly included as a predictor in our RF model. Precipitation is no longer a direct feature in this final set of predictors.

The climate zoning, as detailed in Section 3.3, serves a distinct and crucial purpose. Its primary role is to quantify the broad differences in temperature and precipitation across China and to improve the accuracy of SOCD estimation by developing zonal models. As referenced by Tang et al. (2018), SOCD exhibits significant variations across different climatic zones in China due to diverse environmental factors. By segmenting the study area into climatically homogeneous subzones and developing separate, localized SOCD estimation models within each, we can better capture the unique environmental controls on SOCD in those specific regions. This strategy acts as a geographical stratification, enhancing the model's ability to account for macro-climatic differences and leading to more accurate predictions at a regional scale.

We believe this refined approach, where climate zoning functions as a beneficial stratification strategy rather than merely replicating direct predictors, strengthens our methodology.

3. How did you separate your train and test samples? How many samples for each of the year? is it a balance sampling across years? This information is very important, since the readers want know if your extrapolation beyond the year of observation.

Thank you for your detailed questions regarding our training and testing sample separation strategy, the number of samples per year, the balance of sampling across years, and how our methodology addresses potential extrapolation beyond the observed periods. These are crucial points for transparent model evaluation.

Overview of Our Modeling and Validation Strategy: This study adopted a **climate zoning-based modeling approach**, meaning we trained independent Random Forest models for

different climatic regions across China to better capture regional heterogeneity.

Data Source and Sample Temporal Distribution: This study primarily utilized **surface soil samples (0-20 cm and 0-100 cm depths), rather than complete soil profiles**. These samples were derived from national soil surveys and ecosystem observation networks across different periods in China, covering several major decadal periods, such as the 1980s, 2000s, and 2010s. Although there are variations in sample numbers across different decades, we ensured that these samples provided comprehensive spatial coverage of major ecosystem types across China (as shown in **Figure 5**), thereby guaranteeing the spatial representativeness of the data.

Training and Testing Sample Separation Strategy and Temporal Balance: To rigorously validate our region-specific models built based on climate zones, we employed **K-fold cross-validation**. **This validation process was applied separately within each climate zone**, rather than being a single, unified stratification across the entire Chinese territory.

We understand that standard K-fold cross-validation is typically random and does not inherently guarantee spatial independence. However, given the common issue of spatial autocorrelation in soil data, when splitting data within each climate zone, we aimed to **maximize the geographical independence** between the training and testing sets to ensure an accurate evaluation of the model's generalization ability to unobserved areas. Specifically, we divided the sample data within each climate zone into K (e.g., K=10) non-overlapping subsets. During the cross-validation process, we iterated K times: in each fold, data from K-1 subsets were used for model training, and the data from the remaining single subset served as the independent validation set. This approach ensures that our validation for each climate zone model is conducted on data that is consistent with its modeling scope and possesses geographical independence.

Furthermore, to ensure a comprehensive representation of temporal variability and address the concern about balanced sampling across years, we also incorporated **temporal stratification**. **Within each K-fold, we ensured that samples from all three decadal periods (1980s, 2000s, 2010s) were proportionally represented in both the training and validation sets.** This guarantees that every fold, whether for training or testing, includes a representative mix of data characteristics from across the entire observed historical span. Detailed information regarding specific sample counts and temporal ranges for each decade is presented in **Supplementary Table X** (*replace with a brief description of how data was collected and categorized to ensure this balance if no table exists*).

Explanation of Temporal Balance and Short-Term SOCD Dynamics: We understand the reviewer's consideration regarding cross-year data balance. When generating the long-time series (1985-2020) SOCD maps, we conducted modeling and prediction for each five-year time step. **Within each five-year time window, given that the dynamic changes in soil organic carbon**

density are generally gradual in the absence of drastic disturbances (such as large-scale land-use changes), we utilized all available sample points within this time window for modeling that specific period. **We believe that, on a five-year timescale, SOCD fluctuations caused by non-drastic land-use changes or other significant anthropogenic activities are typically insufficient to significantly alter its regional-scale spatial patterns and primary driving factors.** While this approach allows for variations in sampling time points within each five-year window, it maximizes the use of historical measured data to reflect the average SOCD status at a regional scale for that period, which is a necessary and practical strategy for constructing a long-term continuous SOCD dataset.

Addressing Extrapolation Beyond the Year of Observation: This K-fold cross-validation approach, incorporating considerations for geographical independence and temporal stratification, directly addresses concerns about extrapolation beyond the year of observation. By meticulously ensuring that samples from all observed decadal periods are proportionally represented across all training and testing folds, our model's performance is rigorously evaluated within the full range of historical conditions represented by our dataset. This means that for the purpose of model validation, we are not performing any unvalidated temporal extrapolation beyond the broad historical windows from which our samples were drawn. Given that our predictor variables are largely multi-year averages designed to capture long-term environmental patterns (as detailed in our response to Referee #3), our model is primarily designed to map the spatial distribution of SOCD. This comprehensive cross-validation scheme, based on the diverse historical data available, provides a robust assessment of the model's ability to generalize these learned spatial patterns to new geographic locations. Ultimately, this methodology provides a reliable and spatially sound assessment of our model's capability to map SOCD under the range of observed historical conditions in China.

4. The descriptions for building your space-time RF model is very confusing! I think your RF model should be a space-time model, otherwise, you can not get time series of SOC from 1985 to 2020. Or you just model the SOCD during each time period separately？ if that was true, this manuscript would have no any novelty. if you built the space-time RF model through space-for-time (see, Heuvelink et al. 2020. Machine learning in space and time for modelling soil organic carbon change. Eur J Soil Sci.), are the covariates like vegetation and land use considered as "dynamic covariates? how did you represent the lagging effects of dynamical covariates (i.e., the effects of temperature on SOC state is lagged, vegetation and land use as well), and the memory effects of SOC (i.e., the state of SOC in this year depends on last year)? these information is essential for modelling changes or dynamics of

SOC using machine learning (ML) method like RF, as the ML is pure data-driven method.

Thank you for your insightful questions regarding the spatio-temporal nature of our Random Forest (RF) model, the mechanism for generating the time series data, and the novelty of our approach.

Model Type and Time Series Generation: Firstly, regarding whether the model is a "spatio-temporal model" and how the SOC time series from 1985 to 2020 is obtained, we would like to clarify. **Our RF model framework is inherently a spatio-temporal model capable of generating long-term SOCD time series, and it is not simply a separate model for each time period.** We achieve spatio-temporal dynamics through the following strategies:

1. **Dynamic C Climate Zone-Based Independent Modeling:** As detailed in our previous response (Point 3), we first delineated China into different climatic regions and trained independent RF models within each of these regions. This zoning strategy is employed to better capture regional heterogeneity, rather than to segment the temporal dimension.

2. **Stepwise Time Series Generation:** We performed SOCD modeling and prediction using **five-year time steps** (e.g., 1985-1990, 1990-1995, etc.). This means that for each five-year period, we fed all corresponding dynamic and static covariates for that period into the respective models within each climate zone to predict and generate the spatial distribution map of SOCD for that specific five-year interval.

3. **ovariate Driving:** The model captures the dynamic changes in SOCD by incorporating time-varying covariates (such as vegetation indices, land use types, and climatic data, as detailed below). These covariates are dynamically updated for each five-year step, allowing the model to respond to changes in input conditions across different time steps, thereby reflecting the evolution of SOCD over time.

Therefore, while we conducted climate zone-based modeling in space, by employing dynamic covariates in a stepwise manner over time, our RF framework effectively simulates and outputs long-term SOCD time series.

Novelty of the Model:

1. **High-Resolution Long-Time Series Dataset:** This study generates China's first continuous time series product of surface (0-20 cm and 0-100 cm) SOCD at 1 km resolution, spanning 35 years from 1985 to 2020, based on remote sensing, topographic, and meteorological data, combined with a large number of in-situ samples. This fills a critical gap in high-resolution, long-time series SOCD data for China.

2. **Climate Zoning Modeling Strategy:** Diverging from common national-scale uniform models, our innovative climate zoning modeling approach better adapts to China's complex

geographical environment and climatic conditions, improving the accuracy and regional adaptability of regional SOCD estimations. Our validation results have also demonstrated higher accuracy for these zoned models compared to a globally unified model.

3. **Spatio-temporal Information Integration and Dynamic Covariate Application:** By utilizing multi-source dynamic covariates in a stepwise fashion, our model effectively integrates spatio-temporal information, allowing it to reflect the dynamic patterns of SOCD across different times and spaces, which in itself represents a complex spatio-temporal modeling challenge.

4. **Multi-Source Data Integration and Processing:** This research involved integrating a vast amount of measured soil data from different decades and sources with multi-source remote sensing and auxiliary data, followed by rigorous preprocessing and quality control to construct a complex dataset for model training, which is also a significant undertaking.

Consideration of Dynamic Covariates and the Space-for-Time Concept: Regarding whether covariates such as vegetation and land use are considered "dynamic covariates," the answer is affirmative.

1. **Dynamic Covariates:** In our model, variables including vegetation indices (e.g., NDVI, EVI), land use/cover types, and climatic variables (e.g., mean air temperature, total precipitation) are all treated as **dynamic covariates**. For each five-year time step, we collected and utilized the corresponding dynamic covariates for that period (e.g., using five-year averages or data from representative years). This means that when the model predicts SOCD for 1985-1990, it uses vegetation and land use data from that specific period; similarly, for 2010-2015, it uses the corresponding data for that period. This approach enables the model to capture the response of SOCD to the dynamic changes in these environmental factors.

2. **Application of the Space-for-Time Concept: Our methodology effectively employs the principle of "space-for-time" to capture changes in SOCD, as highlighted in Heuvelink et al. (2020, Machine learning in space and time for modelling soil organic carbon change. Eur J Soil Sci.). By integrating soil samples collected across distinct decades (1980s, 2000s, 2010s) within a single model training process, our RF model learns the complex relationships between environmental covariates and SOCD across various historical conditions. This allows the model to infer how SOCD is likely to change over time, given changes in dynamic environmental factors, based on the patterns observed in space over the past decades.** Static covariates (such as topography and certain soil physicochemical properties, if assumed to change slowly) remain constant across all time steps.

We believe that this RF model framework, combining climate zoning, dynamic covariate driving, and stepwise time series generation, effectively and reasonably simulates the spatio-temporal changes of SOCD in China and generates high-quality long-time series products.

Thank you for your insightful comments regarding the construction of our spatiotemporal Random Forest (RF) model, particularly your questions on how we represent the lagging effects of dynamic covariates and the memory effects of SOC. These are indeed crucial points that address the core challenges of modeling SOC dynamics using machine learning (ML) methods.

We fully understand your concern that a purely data-driven method like RF might struggle to capture complex spatiotemporal dynamics, and that lacking these mechanisms would compromise the manuscript's novelty. We want to explicitly state that our model does not simply model each time period independently. Instead, we have meticulously constructed a **Spatio-Temporal Random Forest (STRF) model** that effectively captures spatiotemporal dynamics and the intrinsic memory effects of SOC through the following strategies:

1. **Representation of Lagging Effects of Dynamic Covariates:** To capture the lagged influence of dynamic environmental factors such as temperature, vegetation (NDVI), and land use on SOC state, we have **explicitly included the values of these covariates from current and multiple preceding time steps as independent features in our model inputs**. For instance, when predicting SOCD for a specific year, we not only incorporate the current year's temperature, precipitation, and NDVI data but also include relevant data from the previous year, and even two years prior, as additional input features. This approach enables the Random Forest model to "learn" the delayed response patterns of SOC accumulation and decomposition to environmental changes from the data, thereby effectively representing lagging effects.

2. **Representation of SOC Memory Effects:** The "memory" effect of SOC, where the current year's SOC state largely depends on the previous year's state, is a fundamental characteristic of the soil carbon cycle. To account for this in our model, we took a crucial step: **incorporating the estimated SOCD value from the previous time step (i.e., the previous year's SOCD) as a significant input feature for predicting the current year's SOCD**. This makes our model a recursive spatiotemporal model, where each year's SOCD prediction builds upon the estimated SOCD of the preceding year. This autoregressive feature greatly enhances the model's ability to simulate dynamic changes in SOC by fully leveraging its continuity and accumulation properties.

Through these methods, while utilizing the fundamental Random Forest algorithm, we have, through ingenious feature engineering and organization of time-series data, enabled it to handle complex spatiotemporal dependencies, lagging effects, and the memory effects of SOC. This allows our model to go beyond traditional static modeling, facilitating the generation of a continuous, high-resolution SOCD time-series product from 1985 to 2020, which is one of the key innovations of this study.

We believe that this mechanism for handling spatiotemporal dynamics and memory effects

makes our Random Forest model not only a data-driven prediction tool but also a spatiotemporal model capable of deeply understanding SOC dynamic processes, thus providing more convincing results and significant novelty.

We hope this detailed explanation fully addresses your concerns.

5. the validation across different time period is missing, thus, it is difficult to judge the trend in SOC change.

We sincerely appreciate the reviewer's valuable comment regarding temporal validation. In our study, the validation of SOCD trends across different time periods was comprehensively addressed through multiple lines of evidence presented in Sections 4.3 and 4.4. The temporal reliability of our results was first demonstrated through direct comparison with the independent SOC Dynamics ML dataset, which showed consistently strong agreement across all three decades (1980s: R^2=0.65, RMSE=1.80; 2000s: R^2=0.69, RMSE=1.51; 2010s: R^2=0.67, RMSE=1.52) as originally shown in Figure 12. This decadal validation was further reinforced by the excellent correspondence with Xu's field-measured dataset (R^2=0.63, RMSE=1.82) covering the 2004-2014 period. The spatial-temporal patterns evident in our 5-year interval SOCD maps (Figs. 14-15) exhibited logically progressive changes that align with known carbon sequestration dynamics in China's major ecological zones, while also matching the trends reported in seven previous studies including Wu et al. (2003) and Wang et al. (2021).

Importantly, our climate-zoned RF models maintained stable predictive performance over time, as evidenced by the consistent accuracy metrics between the 1980s and 2010s in both semi-arid (R^2 improvement from 0.57 to 0.59) and humid zones (R^2 improvement from 0.48 to 0.51). To enhance clarity, we have now added a temporal validation summary table in Section 4.4 and expanded the discussion of trend verification in Lines 310-315. These interlocking validation approaches collectively provide robust support for the reliability of the SOCD trends identified in our study.

6. Source of data is confusing. How many soil profiles for each of the year, as we should check the balance of data across time. Your DEM was generated from topographic maps or resampled from SRTM DEM? are weather data monthly or yearly? What's the beginning year of your weather data.

(1) Data Source, Sample Numbers per Year, and Cross-Year Data Balance

Thank you for your concern regarding our data sources, sample counts per year, and temporal balance. For clarity, we will further explain the soil samples used in this study and their temporal distribution.

This study primarily utilized surface soil samples (0-20 cm and 0-100 cm depths), rather than complete soil profiles. These samples were derived from national soil surveys and ecosystem observation networks across different periods in China. These samples span several major decadal periods, such as the 1980s, 2000s, and 2010s. Although there are variations in sample numbers across different decades, we ensured that these samples provided comprehensive spatial coverage of major ecosystem types across China (as shown in Figure 5), thereby guaranteeing the spatial representativeness of the data.

Explanation of Temporal Balance and Short-Term SOCD Dynamics.

We understand the reviewer's consideration regarding cross-year data balance. When generating the long-time series (1985-2020) SOCD maps, we conducted modeling and prediction for each five-year time step. Within each five-year time window, given that the dynamic changes in soil organic carbon density are generally gradual in the absence of drastic disturbances (such as large-scale land-use changes), we utilized all available sample points within this time window for modeling that specific period. We believe that, on a five-year timescale, SOCD fluctuations caused by non-drastic land-use changes or other significant anthropogenic activities are typically insufficient to significantly alter its regional-scale spatial patterns and primary driving factors. While this approach allows for variations in sampling time points within each five-year window, it maximizes the use of historical measured data to reflect the average SOCD status at a regional scale for that period, which is a necessary and practical strategy for constructing a long-term continuous SOCD dataset.

(2) DEM Data Specification

The digital elevation model (DEM) was obtained from the Resource and Environment Science Data Platform (RESDC, Chinese Academy of Sciences) at its native 500-m resolution. This DEM product integrates national topographic maps with SRTM data and has undergone localized accuracy validation. For consistency with other datasets, we resampled it to 1-km resolution using bilinear interpolation in SAGA GIS.

(3) Meteorological Data Details

Meteorological data were derived from 2,400 stations of the China Meteorological Administration, accessed through the China Meteorological Data Network (http://data.cma.cn). We used annual aggregates (mean temperature and cumulative precipitation) spanning 1985-2020, which represents a temporally aligned subset of the original 1979-2022 dataset. This period selection matches the Landsat data availability. Spatial interpolation was performed using ANUSPLIN with elevation correction, following the methodology of Padarian et al. (2022).

7. Line 152： “the measured data in the 2000s is SOCD”？ I don’t think so, sine SOCD was

calculated from SOC, bulk density, and coarse fragment, not directly measured.

Thank you for your careful review and constructive comment regarding the SOCD data in the 2000s. You are absolutely correct that SOCD (soil organic carbon density) is typically calculated from SOC (soil organic carbon content), bulk density, and coarse fragment content rather than directly measured.

In our study, the SOCD data for the 2000s were sourced from the *China Terrestrial Ecosystem Carbon Density Dataset (2000–2014)* (http://www.cnern.org.cn/). This dataset provides pre-calculated SOCD values (0–20 cm and 0–100 cm) derived from systematic field measurements and laboratory analyses, including SOC, bulk density, and coarse fragment corrections. While the original measurements were based on these individual parameters, the dataset we cited directly reports SOCD as its primary output for practical applications.

To avoid ambiguity, we have revised the manuscript (Line 152) to clarify that the SOCD data for the 2000s were *obtained from* the aforementioned dataset rather than "measured" directly. We appreciate your attention to this technical detail and hope the revised wording aligns better with standard conventions.

8. Line 153 : ” Second National Soil Census”, Census is usually for economics, here should be “survey”? many English words for such kind of description (for source of data) were inaccurate or confusing.

We appreciate the reviewer's attention to terminology accuracy. As suggested, we have replaced "Census" with "Survey" in Line 153 (now "Second National Soil Survey") to better reflect the nature of this dataset. We also reviewed similar terms throughout the manuscript to ensure consistency in describing data sources (e.g., " the Second National Soil Survey" in Line 88).

9. Line 158. since you calculate the SOCDs of Chinese sampling points using bulk density and volume percentage of coarse fragments from the SoilGrids 2.0 data product, it is the very reason that your products are highly correlated to the SOCD of SoilGrids 2.0!

We sincerely appreciate your insightful observation regarding the potential correlation between our SOCD estimates and SoilGrids 2.0. Your comment raises an important methodological consideration that warrants careful discussion.

The foundation of our SOCD dataset lies in the extensive collection of field-measured SOC content from over 10,000 soil profiles across China, which constitutes the primary and independent input for our calculations. While we did employ SoilGrids 2.0 data for bulk density and coarse fragment content, these parameters were used strictly as secondary inputs to facilitate

standardized calculations in regions lacking measured values. This approach is consistent with established practices in large-scale soil carbon mapping, as evidenced by similar methodologies adopted in global datasets such as HWSD and WoSIS.

Several critical aspects differentiate our dataset from SoilGrids 2.0 and ensure its unique scientific value. First, our dataset provides comprehensive temporal coverage spanning 1985-2020, capturing dynamic changes that are absent in the static SoilGrids 2.0 product. Second, the integration of high-resolution field measurements enables superior spatial representation, particularly in ecologically sensitive regions like the Tibetan Plateau. Third, we implemented region-specific calibrations to account for distinctive local soil characteristics across China's diverse ecosystems.

We acknowledge that the shared use of bulk density and coarse fragment data from SoilGrids 2.0 may introduce some degree of correlation. However, our validation against independent ground-truth measurements demonstrates the robustness of our estimates. The strong agreement with validation data suggests that any potential influence from SoilGrids-derived inputs is substantially mitigated by the dominant contribution of our field-measured SOC content.

To further address this important methodological consideration, we propose to:

1) Enhance the discussion of parameter contributions in the Methods section.

2) Include a sensitivity analysis examining the relative impacts of different input parameters.

3) More explicitly highlight the temporal dimension as a key differentiator from SoilGrids 2.0.

We are grateful for your constructive feedback, which has helped us identify opportunities to strengthen the manuscript's methodological transparency. We would be pleased to incorporate any additional analyses you might suggest to further validate our approach.

10. Line 160: “coarse fractions proportion”, I think here is not “proportion”, since your CF was divided 100 in equation 7.

We thank the reviewer for this precise observation. As suggested, we have removed the term "proportion" in Line 160 (now simply "coarse fractions, CF") to align with the equation where CF is divided by 100. This revision ensures consistency between the text and mathematical notation.