Comments from Reviewer #2:

This study developed the first nationwide 30 m resolution Soil and Water Conservation Terrace Measures Dataset (SWCTMD) of China from 2000 to 2020. Terrace is the most important type of soil conservation engineering measures for most areas in China, given the critical role of terraces in mitigating soil erosion, enhancing water retention, this work fills a critical data gap that has long constrained national-scale erosion modeling and conservation planning for decades. A two-stage classification framework was proposed, first distinguishing terraces from non-terraces and then classifying four terrace types: level terrace, slope terrace, zig terrace, and slope-separated terrace. The framework integrates time-series Landsat imagery, DEM data, and GlobeLand30 cropland data, employing the Random Forest classifier on the Google Earth Engine platform. Key findings include:

(1) Total terrace area increased from 400,896 km² in 2000 to 496,934 km² in 2020;

(2) Terraces were estimated to reduce about 818 million tons of soil erosion on croplands in 2020;

(3) The dataset outperforms previous products in spatial completeness and terrace type granularity at national scale;

(4) It supports applications in soil erosion modeling, ecosystem service assessment, and land

management planning.

While the study is of high quality and makes a significant contribution, I would like to offer the following suggestions for improvement.

Suggestions and Minor Revisions Required :

Response: We appreciate you very much for your positive comments concerning our manuscript. Those comments are valuable and helpful for improving our manuscript. We followed all comments and made revision and responses carefully. A point-by-point reply to the comments are listed below.

1. Some sentences in the manuscript require more precise wording and would benefit from a more rigorous formulation. For instance, Lines 74-75 currently states:

"Using this mapping framework, the first Soil and Water Conservation Terrace Measures Dataset of China (SWCTMD) was produced using time-series Landsat satellite imagery and digital elevation model data, covering the period from 2000 to 2020."

In fact, several nationwide terrace mapping efforts have been conducted prior to this study, such as the work by Cao et al. (2020), which generated a 30-meter resolution terrace map for China. While those studies were not based on long-term time series data, they do represent national-scale mapping efforts. Therefore, the sentence should be revised to clarify that the SWCTMD is the first long-term (two-decade) national terrace dataset, rather than the first national terrace dataset overall. Suggested revision: *"Using this mapping framework, we developed the first long-term (2000– 2020) national Soil and Water Conservation Terrace Measures Dataset (SWCTMD) of China."*

Lines 370, Reference: Cao B, Yu L, Naipal V, et al. A 30-meter terrace mapping in China using Landsat 8 imagery and digital elevation model based on the Google Earth Engine. Earth System Science Data

Discussions, 2020: 1– 35.

Similarly, Lines 108-109 "Compared to other DEM data, SRTM DEM is the most quality-controlled, broadest coverage, and highest accuracy DEM among open-source data"

While SRTM has historically been one of the most widely used and quality-controlled global DEM datasets, it is worth noting that in recent years, several newer open-access DEM products, such as ALOS AW3D30 and Copernicus DEM, have demonstrated higher spatial resolution and improved accuracy in many regions, including mountainous and vegetated areas.

Response: Thank you for your insightful comment. We apologize for the unclear expression. The original sentence was: *"Using this mapping framework, the first Soil and Water Conservation Terrace Measures Dataset of China (SWCTMD) was produced using time-series Landsat satellite imagery and digital elevation model data, covering the period from 2000 to 2020."* We revised it to the suggested version: *"Using this mapping framework, we developed the first long-term (2000 to 2020) national Soil and Water Conservation Terrace Measures Dataset (SWCTMD) of China."* (Lines 75-77).

We recognized that several newer open-source DEM data products in recent years offer higher accuracy than SRTM DEM, particularly in mountainous and vegetated areas. Therefore, we replaced the SRTM DEM data with Copernicus DEM data and regenerated the SWCTMD dataset.

Accordingly, the sentences (Lines 105-110) have been revised as follows:

"Topographical features are essential characteristics that differentiate regular cropland and terrace, playing a crucial role in the identification of terraces. We used the Copernicus DEM data to calculate these topographical features. The Copernicus DEM is a Digital Surface Model with 30 m resolution, derived from radar satellite data acquired from 2010 to 2015 during the TanDEM-X mission. Compared to other DEM data (SRTM, ASTER GDEM, ALOS World 3D, and NASADEM), Copernicus DEM has the highest accuracy among open-source data (Guth and Geoffroy, 2021), exhibiting the greatest detail of terrain (Li et al., 2022a). The GEE platform provides access to the Copernicus DEM at 30 m resolution."

2. According to the manuscript, all four types of terraces used in the study are assigned dimensionless conservation factor (P or E-values) below 0.343. This implies that, under the RUSLE/CSLE framework, terrace implementation on a given slope would reduce potential soil loss by at least 65.7%. Furthermore, since terraces are predominantly distributed on sloping croplands, areas generally subject to higher erosion rates, the expected relative reduction in soil loss should arguably be even more significant when terraces are applied. However, In Section 3.4, the authors state that, "In comparison to the scenario without terrace measures, the amount of soil erosion in the regions of Yunnan, Sichuan, Chongqing, Guizhou, Gansu, Shanxi, and Shaanxi regions decreased by 47.47%, 46.02%, 45.57%, 45.25%, 35.48%, 29.75%, and 27.80%, respectively (Fig. 8b)". Therefore, I would appreciate it if the authors could clarify the reason behind.

Response: We assigned a dimensionless conservation factor (P or E value) below 0.343 for each type of terrace. Under the RUSLE/CSLE framework, terrace implementation on a given slope would reduce potential soil loss by at least 65.7%. However, this reduction is achievable only when the terrace area is close to or approximately equal to the cropland area. For instance, when the terrace area in a province is significantly smaller than its cropland area, the total amount of soil erosion reduced by terraces is limited.

Compared to scenarios without terrace measures, the proportion of soil erosion reduction would be far below 65.7%. Furthermore, the soil and water conservation effectiveness of terraces also depends on their spatial distribution. If a province's terraces are mainly distributed in areas with moderate and strong soil erosion intensity, with few in areas of very strong and severe soil erosion intensity. Since the region with very strong and severe soil erosion intensity may have contributed to the majority of the province's total soil erosion, the proportion of soil erosion reduction would be also far below 65.7%.

3. While the overall accuracy (OA) and Kappa coefficient reported for the classification results appear high, this does not necessarily imply satisfactory classification performance. Notably, the Producer's Accuracy (PA) for zig terraces and level terraces was particularly low (as low as 15–25%), and their corresponding F1-scores were generally below 40%, indicating considerable misclassification and omission errors. Such results suggest that these terrace types may have been substantially underestimated in the dataset. It is important to note that high OA/Kappa values may be misleading in imbalanced classification tasks, especially when majority classes dominate the confusion matrix. The low performance metrics for certain terrace types likely reflect a combination of factors, including:

Class imbalance between dominant and minority categories during model training;

Limited or unrepresentative training samples for fragmented or narrow terraces;

Intrinsic heterogeneity in real-world spatial distribution, especially in mountainous regions;

Fragmented and narrow terraces are especially prone to omission;

Spectral confusion and mixed pixels in medium-resolution imagery.

These aspects should be discussed in the manuscript to provide a more nuanced interpretation of the classification results and to guide future refinement. The authors may also refer to relevant studies for in-depth discussions on these issues.

Response: Thank you for your insightful comment. In the revised manuscript Section 4.3, we discussed the impact of sample imbalance, uneven distribution of terraces, spectral confusion, and mixed pixels on classification accuracy, and proposed recommendations for improving the classification of terraces in the future. It now stands as follows (Lines 464-489):

"The spatial heterogeneity of land types frequently leads to class imbalance in remote sensing classification, consequently diminishing classification accuracy for minority classes that occupy a smaller area (Xiao et al., 2024). The models tend to favor majority classes during training, reducing their ability to accurately identify minority classes (Chen et al., 2025). When the ratio of samples across different classes remains balanced, classification performance typically falls short of optimal accuracy thresholds (Deng et al., 2025). A common strategy to alleviate this negative effect is to divide the study area into multiple sub-regions for localized classification, thereby reducing the impact of sample imbalance on model accuracy (Zhang et al., 2020). In this study, we employed a partitioned two-stage RF approach to reduce the effects of sample imbalance on classification accuracy. The results demonstrated that classification for terrace and different terrace types achieved satisfactory accuracy in both the entire study area and individual subregions. However, the accuracy metrics of the majority class were still higher than those of the minority class. In future studies, sample optimization techniques and

more advanced classification methods could be combined to further improve the accuracy of minority class classification."

"The complex and diversity diverse landform types have resulted in differences in the spectral information and topographic features of terraces in different regions. In Southwest and Northwest China, terraces exhibit concentrated distributions with clearly defined characteristics, making them easily identifiable. However, South China, Central China, and East China have relatively low topographic relief. Some terraces have spectral and topographic features similar to those of sloping farmland and flatland. This similarity, combined with the presence of mixed pixels in medium-resolution imagery (Wang et al., 2021a), makes it challenging to detect the terrace patches. Although the classification used 30-meter Landsat imagery in this study was generally robust, some fragmented and narrow terraces were omitted. Future research could employ high-resolution remote sensing images to effectively identify fragmented and narrow terraces. Previous small-scale studies have demonstrated that the use of high-resolution remote sensing imagery, combined with object-based classification methods and deep learning approaches, can significantly enhance classification accuracy and reduce the impact of spectral confusion and mixed pixels on terrace identification (Diaz-Varela et al., 2014; Wang et al., 2023; Kan et al., 2025). To improve classification accuracy and efficiency, cropland data were used as the basis for terrace identification. Inevitably, the accuracy of cropland data impacts the terrace mapping process, as errors in cropland data propagate into the terrace maps. In summary, future studies could utilize high-resolution remote sensing imagery and more accurate cropland datasets, and adopt sample optimization techniques and more advanced classification algorithms to improve the detection of subpixel terrace distributions."

4. Since the manuscript involves the quantitative assessment of soil and water conservation benefits of terraces, particularly through the estimation of soil erosion reduction, it is important to critically examine the assumption of using a uniform conservation factor (P-value or E-value) for the same type of terrace measures across different regions of China. While assigning a single value to each terrace type simplifies the model and facilitates national-scale analysis, it may overlook important spatial heterogeneity in terrace structure, maintenance conditions, climatic regimes, and land management practices. For example:

The same "slope terrace" may perform diferently in terms of erosion control in Yunnan's highlands compared to the Loess Plateau, due to diferences in rainfall erosivity, soil properties, and vegetationcover; Engineering design standards and actual field implementation of terraces may vary significantly between provinces, leading to divergence in functional efectiveness; The topographic context (e.g., slope gradient, curvature) strongly influences the conservation outcome, even for structurally similar erraces.

Therefore, I recommend that the authors acknowledge and briefly discuss the limitations of using fixed terrace factor values in erosion modeling across diverse environmental settings.

Response: Thank you for your insightful comment. We recognize the limitations of using fixed terrace factor values in erosion modeling across diverse environmental settings. We discuss this in Section 4.3.

It now stands as follows (Lines329-332):

"In this study, each terrace type was assigned a fixed E value to facilitate the estimation of large-scale soil erosion. However, this approach overlooks spatial heterogeneity in terrace structure, maintenance status, field management, climate, and topography. Future research should incorporate regional characteristics and adjust the E-value accordingly."