Response to Comments：

Comment 1: Further highlight the core innovation of the manuscript

The authors are encouraged to further highlight the core innovation of the manuscript. The current manuscript provides a relatively complete description of the research background and experimental results, but the main contributions of the study are somewhat scattered. It is suggested that the authors add a more focused summary of the contributions at the end of the Introduction. Specifically, the authors should clearly state that this study is not only intended to construct a GNSS-constrained FY-4A PWV calibration model, but more importantly, to quantitatively evaluate, through controlled experiments, the effects of GNSS training station number and spatial layout on model accuracy, spatial generalization capability, and temporal stability. The proposed station configuration principle, namely prioritizing spatial representativeness and coverage uniformity under limited station availability, should also be emphasized.

Reply: In the concluding section of the article's introduction, the main objective of this study is emphasized: to investigate the influence of the number and spatial distribution of stations on the reconstruction of FY-4A PWV, especially in terms of calibration. In the conclusion section, it is further highlighted that, under the constraints of limited station resources, priority should be given to spatial representativeness and the uniformity of coverage.

For specific details, please refer to lines 74 and 450 in the revised version of the article.

Line 74: However, most studies have concentrated on regions with relatively dense GNSS station coverage, while quantitative understanding of model performance, stability, and generalization behavior under sparse station conditions is still lacking. How the number of training stations and their spatial configuration jointly constrain the spatial consistency and extrapolation capability of calibration models has not been systematically evaluated. To address this gap, we develop a machine learning-based calibration framework for the all-weather FY-4A PWV product over China.

Line 450: Comprehensive analyses across spatial, temporal, and seasonal dimensions reveal a clear performance saturation behavior. Increasing the number of training stations leads to substantial improvements in independent validation accuracy and spatial error homogeneity, particularly when training sample sizes increase from sparse to moderate levels.

Comment 2: Clarify the resolution harmonization and scale consistency of auxiliary variables

In Section 2.2, some variables were subjected to resolution harmonization and scale transformation. The authors are advised to provide further details in the data processing or methodology section regarding how variables such as DEM and NDVI were spatially matched to the FY-4A PWV grid. Adding these explanations would help improve the transparency of the data processing workflow and enhance the rationality and reproducibility of the model input design.

Reply: Thank you for the valuable suggestions from the reviewers. To achieve strict alignment with the FY-4A PWV target grid, we performed spatial resampling on all variables with inconsistent resolutions. For static or quasi-static variables, such as DEM, NDVI, and LCT, bilinear interpolation was uniformly applied to resample them onto the FY-4A PWV grid. During the resampling process, we ensured that all input variables shared identical grid counts, spatial extents, and grid cell center coordinates. This enabled precise pixel-by-pixel spatial matching and prevented errors caused by grid offset or boundary discrepancies.

Please refer to lines 163- 167 of the revised manuscript for specific details.

Lines 163-167: To ensure spatiotemporal consistency across all input data, a rigorous pre-processing procedure was applied. Temporally, the ERA5 data were aligned to match the exact observation hours of the full-disk FY-4A PWV product. Spatially, the FY-4A PWV data and the DEM over the study domain were resampled onto the ERA5's 0.25° × 0.25° grid using a bilinear interpolation method, thereby unifying the spatial resolution and grid alignment for all datasets prior to model training.

Comment 3: Provide more details on the construction of the three spatial layouts

The authors are encouraged to provide more details on how the three station spatial layouts were constructed. Section 3.3 mentions three station distribution patterns: random, surrounding, and clustered. However, it is still unclear how these layouts were generated. For example, it should be clarified whether the random layout was generated through repeated random sampling and whether the averaged results were reported, how the spatial clustering range of the clustered layout was determined, and whether the same number of training stations was strictly maintained among the three layouts. The authors are advised to supplement these methodological details and explain how the corresponding validation stations were determined, in order to enhance the reproducibility of the experiments and the reliability of the conclusions.

Reply: Thank you for the valuable comments from the reviewer. In response to the point raised regarding the unclear definition of the "spatial layout experiments (clustered/surrounding/random)" and the corresponding validation sets, we have carefully reviewed the issue and made revisions and improvements to the original description. The specific clarifications are as follows:

In this study, we aimed to investigate the impact of different spatial distribution patterns of a 200-station network on PWV calibration accuracy. Our baseline network consists of 244 stations, among which 16 stations covering diverse geographical locations and climatic environments are reserved as the validation dataset and are not involved in the modeling process. To construct three representative spatial layouts, the following methods were employed:

Random distribution: A network composed of 200 stations is selected completely at random from all stations. This method simulates a scenario where station locations have no specific spatial pattern.

Surrounding distribution: To simulate the pattern of stations distributed around a central area, the geographic center of all stations is first identified. Then, 28 stations are randomly removed from the central area. The remaining 200 stations, primarily located in peripheral regions, constitute the training set.

Clustered distribution: To simulate the pattern of stations densely concentrated in a certain area, the opposite strategy is adopted: 28 stations are randomly removed from the peripheral areas of the station network, retaining 200 stations in the central and relatively concentrated regions.

In the revised manuscript, we have clearly defined the specific construction methods for these three spatial layouts. It is noted that the "random distribution" serves as the baseline scenario, while the "surrounding distribution" and "clustered distribution" are used to examine the model's performance under two extreme cases of non-uniform sampling: missing central data and missing edge data, respectively. We believe this additional explanation clearly delineates the different experimental setups and clarifies the rationale behind the composition of their respective validation sets.

Comment 4: Improve the presentation of figures, equations, and language

The overall structure of the manuscript is clear, and the figures, tables, and equations generally support the methodological description and experimental analysis. However, there are still some minor issues related to figure presentation, equation formatting, and formatting consistency. The authors are advised to further check and revise these aspects to improve the readability, standardization, and overall consistency of the manuscript.

(1) Some figures contain a large amount of information, especially multi-panel result figures. It is suggested that the authors further standardize the size and layout of the figures. For example, the aspect ratios of the subplots in Figures 6 and 7 could be appropriately adjusted to improve the standardization and consistency of the manuscript.

Reply: Based on the reviewer's suggestion, the size and layout of all figures have been further standardized and organized.

(2) In some figures, the text overlaps with graphical elements. The authors are advised to further check and standardize the font size, legends, color bars, axis labels, and subplot numbering, so that comparisons among different station numbers, model results, and spatial layouts can be presented more clearly.

Reply: Based on the reviewer's suggestion, the size and layout of all figures have been further standardized and organized.

(3) The authors should carefully check whether the equation formatting and variable descriptions are complete. For example, the explanation of the variables in Equation (4) should be further supplemented, and the symbol formatting in Equations (7)–(10) should be kept consistent. This would help avoid ambiguity when readers interpret the evaluation metrics.

Reply: The formula has been verified and the symbol descriptions have been supplemented.