remote sensing

MDPI

Article

Cloud Base Height Inversion Based on a Heterogeneous Stacked Integrated Network

Yuanyang Hua [1], Tao Xie [2,3,*], Yixiang Chen [2], Wenmei Li [2], Jian Li [1] and Chao Wang [4]

[1] School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; 202412482424@nuist.edu.cn (Y.H.); lijian@nuist.edu.cn (J.L.)
[2] School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China; chenyixiang@njupt.edu.cn (Y.C.); liwm@njupt.edu.cn (W.L.)
[3] Laboratory for Regional Oceanography and Numerical Modeling, Qingdao Marine Science and Technology Center, Qingdao 266200, China
[4] School of Electronics and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; chaowang@nuist.edu.cn
* Correspondence: xietao@nuist.edu.cn

Highlights

What are the main findings?

- Hetero-CBHNet is a heterogeneous stacked integrated network framework that uses random forests, CNNs, and Transformers as a pool of candidate base models, which are then fused via the RidgeCV meta-learner. For cloud base height inversion from FY-4A and MODIS data, it achieves RMSE values of 2966.12 m and 2797.64 m under strict chronological splitting, with R values of 0.630 and 0.659; under random splitting, the RMSE values are 1510.01 m and 1488.41 m. Under both protocols, Hetero-CBHNet outperforms all single models.

What are the implications of the main findings?

- For cloud base height retrieval, rather than prescribing a single model, it is preferable to determine the model architecture through data-driven selection from a pool of candidate models.
- Existing passive cloud type information is not yet sufficient to replace active-satellite labels, and developing a high-accuracy passive cloud type classifier supervised by active-satellite labels is a key direction for future work.

Abstract

To address the challenge of accurately inverting cloud base height (CBH) from passive remote sensing, this study proposes Hetero-CBHNet, a heterogeneous stacked integrated network framework for CBH inversion. The framework uses random forest (RF), convolutional neural network (CNN), and Transformer models as a pool of candidate base models and employs the RidgeCV algorithm to automatically learn the fusion weights. A total of 1,011,620 samples from FY-4A and MODIS, covering the period from May 2019 to May 2020, were selected (with CloudSat/CALIPSO 2B-CLDCLASS-LIDAR as the ground truth), and the dataset was split chronologically. Across the two data sources under the chronological splitting protocol, Hetero-CBHNet achieves RMSE values of 2966.12 m and 2797.64 m and R values of 0.630 and 0.659 on the test sets, outperforming the individual models. Feature analysis shows that incorporating operational passive cloud type products does not yield a stable gain, indicating that current passive cloud type information is not yet sufficient to replace active satellite labels. In addition, the RMSE values under the random splitting

Check for updates

Academic Editor: Filomena Romano

Received: 14 September 2026
Revised: 14 October 2026
Accepted: 20 October 2026
Published: 22 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

protocol are 1510.01 m and 1488.41 m, substantially better than those under chronological splitting, which confirms the strong dependence of the conclusions on the evaluation protocol.

Keywords: cloud base height; heterogeneous stacked integrated network; passive remote sensing; FY-4A; MODIS

1. Introduction

Clouds are vital components of the Earth's hydrological cycle. They both reflect solar radiation and absorb infrared radiation, thereby playing a significant regulatory role in the Earth's energy balance and temperature variation. Cloud base height (CBH) is one of the key parameters describing the vertical structure of clouds and has a significant impact on the radiative budget of the Earth–atmosphere system [1]. CBH is closely related to visibility. In aviation operations, the high water vapour content of low and middle clouds significantly affects flight visibility; moreover, when aircraft pass through cloud layers, ice accretion on the airframe surface is likely to occur, seriously compromising flight safety [2,3]. Accurate inversion of CBH is of great significance for understanding cloud radiative effects and providing meteorological support [4].

At present, there are two main approaches for obtaining cloud base height (CBH) data: ground-based radar observations and satellite remote sensing. Although ground-based radar observations can accurately determine CBH, they are limited by the physical environment and technical constraints, making it difficult to cover areas such as oceans and deserts; consequently, they cannot fully meet the urgent needs of research fields such as weather analysis and climate change studies [5]. Since the CloudSat and CALIPSO satellites became fully operational, the detection of cloud vertical structure has entered a new phase, and CBH observations have been widely utilised [6,7]. However, because these meteorological satellites operate in sun-synchronous orbits, their active onboard sensors cannot continuously observe fixed regions, and they suffer from issues such as long revisit intervals and narrow observation swaths; In contrast, passive spaceborne sensors, such as the Advanced Geostationary Radiation Imager (AGRI) and the Moderate Resolution Imaging Spectroradiometer (MODIS), can capture the horizontal distribution of clouds over wider areas and at higher temporal resolutions [8,9]; Nevertheless, their cloud products provide only cloud-top information and do not allow direct CBH inversion.

Consequently, CBH inversion using active and passive remote sensing data is of considerable research value. Tan et al. [10] utilised a random forest algorithm, combining FY-4A data with data from the A-Train satellite series, to construct a random forest model for inverting the CBH of the uppermost cloud layer observed by FY-4A; their results indicated some discrepancies between the inverted CBH and observations from active remote sensing satellites. Yu et al. [11] categorised FY-4A cloud products into eight cloud types and utilised random forest and gradient boosting tree algorithms to classify cloud types and invert CBH. They compared these results with those obtained from an inversion that did not distinguish between cloud types, using CBH measured by CloudSat as the ground truth for validation. The results indicated that the cloud-type-classified inversion yielded more accurate CBH estimates, although the inversion results for deep convective clouds were less satisfactory. Tan et al. [12] introduced effective cloud water content (ECWC) to construct a systematic ECWC lookup table as an alternative to fixed water content assumptions, and achieved good agreement with CloudSat Cloud Profiling Radar (CPR) observations when using MODIS and AHI data. Lenhardt et al. [13] proposed a new method called ORABase (Ordered Regression Autoencoder for Cloud Base), which utilised MODIS cloud products to invert

oceanic CBH; however, the inversion results were limited to oceanic regions. Qiu et al. [14] stacked and integrated RF, XGBoost, LightGBM, and MLP via the RidgeCV meta-learner (Stack-CBHNet), and validated the effectiveness of multi-model stacking in improving CBH inversion accuracy using MODIS-CloudSat collaborative data.

At present, most research remains confined to ensemble approaches based primarily on single-tree models or models of the same type, and the potential for collaboration between different architectures has not yet been fully explored. In fact, clouds exhibit distinct multi-scale characteristics and spatial correlations; it is difficult for a single model to capture all aspects of this information. Random forest (RF) models are particularly effective at processing tabular data but cannot directly model or utilise the spatial context between pixels; convolutional neural network (CNN) models excel at extracting local textural features; Transformer models are advantageous for modelling global long-range dependencies [15–17].

This study proposes a heterogeneous stacked integrated network, Hetero-CBHNet, for CBH inversion. It uses RF, CNN, and Transformer models as a pool of heterogeneous base learners, leveraging their distinct modelling capabilities to achieve complementarity, and employs a ridge regression meta-learner to learn the fusion weights. Its effectiveness is evaluated on two data sources, FY-4A and MODIS, under a strict chronological splitting protocol, with a random splitting control and a multi-layer cloud error decomposition also provided.

2. Materials and Methods

2.1. Data Source

The AGRI sensor is a key payload on the FY-4A meteorological satellite, and its derived products—such as cloud-top height, cloud-top temperature, cloud-top pressure, and cloud microphysical and optical properties—have been widely utilised [18–20]. This study constructs a CBH inversion model for FY-4A using cloud top height (CTH), cloud top pressure (CTP), cloud top temperature (CTT), cloud effective radius (CER), cloud optical thickness (COT), ice water path (IWP), and liquid water path (LWP) from the FY-4A/AGRI L2 products, together with the latitude and longitude of each sample point, giving a total of nine dimensions, to construct a CBH inversion model for FY-4A.

The MODIS sensor is a moderate-resolution imaging spectroradiometer aboard the Terra and Aqua sun-synchronous satellites. It supports retrieving various cloud macro- and microphysical parameters and provides information on surface conditions, ocean colour, cloud top height, cloud top temperature, and cloud boundaries. Based on its cloud product CLDPROP_L2_MODIS_Aqua, this study extracts cloud top height, cloud top pressure, cloud top temperature, cloud optical thickness, cloud water path (CWP), and cloud effective radius, which, together with latitude and longitude, constitute the model input for the MODIS data source, giving a total of eight dimensions.

The CloudSat satellite uses its onboard Cloud Profiling Radar (CPR) to obtain data from space on the vertical structure of global cloud cover, as well as liquid water and ice content. The data collected during a single orbit around the Earth constitute one track, comprising 36,383 observation profiles. Both CALIPSO and CloudSat belong to the A-Train satellite constellation; they fly in the same orbital plane, with CloudSat leading and CALIPSO following, separated by only about 15 s, which enables them to observe almost the same area on Earth simultaneously. The CALIPSO satellite uses its onboard Cloud-Aerosol Lidar with Orthogonal Polarisation (CALIOP) to detect extremely thin cloud layers and aerosols from space. The combined CPR and CALIOP product, 2B-CLDCLASS-LIDAR, provides a comprehensive and detailed description of cloud vertical structure [21,22].

Spatiotemporal matching was performed on the FY-4A, MODIS, and CloudSat data, with a temporal interval of no more than 15 min and a spatial interval of no more than 2 km. The cloud type, cloud top height, and number of cloud layers from 2B-CLDCLASS-LIDAR were used only for post hoc error decomposition, validation, and ablation analysis, and were not used as input to any model, thereby ensuring that the model inputs and output variables are mutually independent. The samples were not restricted to a specific subregion; the matched samples are distributed along the CloudSat track, and both FY-4A and MODIS were taken from the same set of triple spatiotemporally matched co-located samples. Therefore, the geographical distribution of the samples is determined by the full-disk coverage of FY-4A/AGRI: approximately 99.7% of the samples lie between 25°E and 180°E, with the remaining approximately 0.3% located at the western edge of the disk between 175°W and 180°W; the latitude range is −76.1°N to 80.3°N. No land–sea or regional screening was applied to the samples: across the full sample set, the land and ocean proportions are approximately 18% and 82%, respectively.

The main protocol of this study is the chronological splitting protocol. Data from May to December 2019 and January 2020 (a total of 689,302 sample points) were used to split the base-model training and validation sets, data from February 2020 (a total of 104,289 sample points) were used as the meta-layer validation set for training the meta-learner weights, and data from March to May 2020 (a total of 218,029 sample points) were used as the test set. Table 1 lists the sensors and products used in the experiments in this study.

Table 1. Sensors and products.

Sensors	Products	Variables
AGRI	L2	CTH, CTT, CTP, CER, COT, IWP, LWP
MODIS	CLDPROP_L2_MODIS_Aqua	CTH, CTT, CTP, COT, CWP, CER
CPR-CALIOP	2B-CLDCLASS-LIDAR	Cloud layer height (top and base), CloudType

2.2. Methods

2.2.1. Overall Framework of Hetero-CBHNet

The metrics used to evaluate model performance were root mean square error (RMSE), mean absolute error (MAE), mean relative error (MRE), and the correlation coefficient (denoted as R hereafter). The data used for accuracy assessment comprised CBH derived from FY-4A and MODIS data, together with CloudSat observations serving as the ground truth.

Figure 1 illustrates the overall inversion workflow of Hetero-CBHNet. To maximise the complementary strengths of different models, the framework incorporates three models with distinct architectures—RF, CNN, and Transformer—into the pool of candidate base learners, each of which is trained independently on its respective base-model training and validation sets. Upon entering the stacking stage, rather than directly fusing all candidate models, Ridge Regression with built-in Cross-Validation (RidgeCV) acts as the meta-learner. It learns the fusion weights of the base models on the meta-layer validation set to determine the optimal combination for CBH inversion; the test set is not involved in any training or selection process. In addition, existing studies generally rely on active-satellite-derived cloud type or cloud-top products as model inputs [11,12]. Such active-satellite-derived variables are homologous to the CBH to be inverted and are unavailable in operational settings, which can make the reported accuracy overly optimistic. Accordingly, this study tightens the task formulation as follows: the inputs consist solely of passive satellite products, while active satellite data are used only for labels, validation, and post hoc analysis.

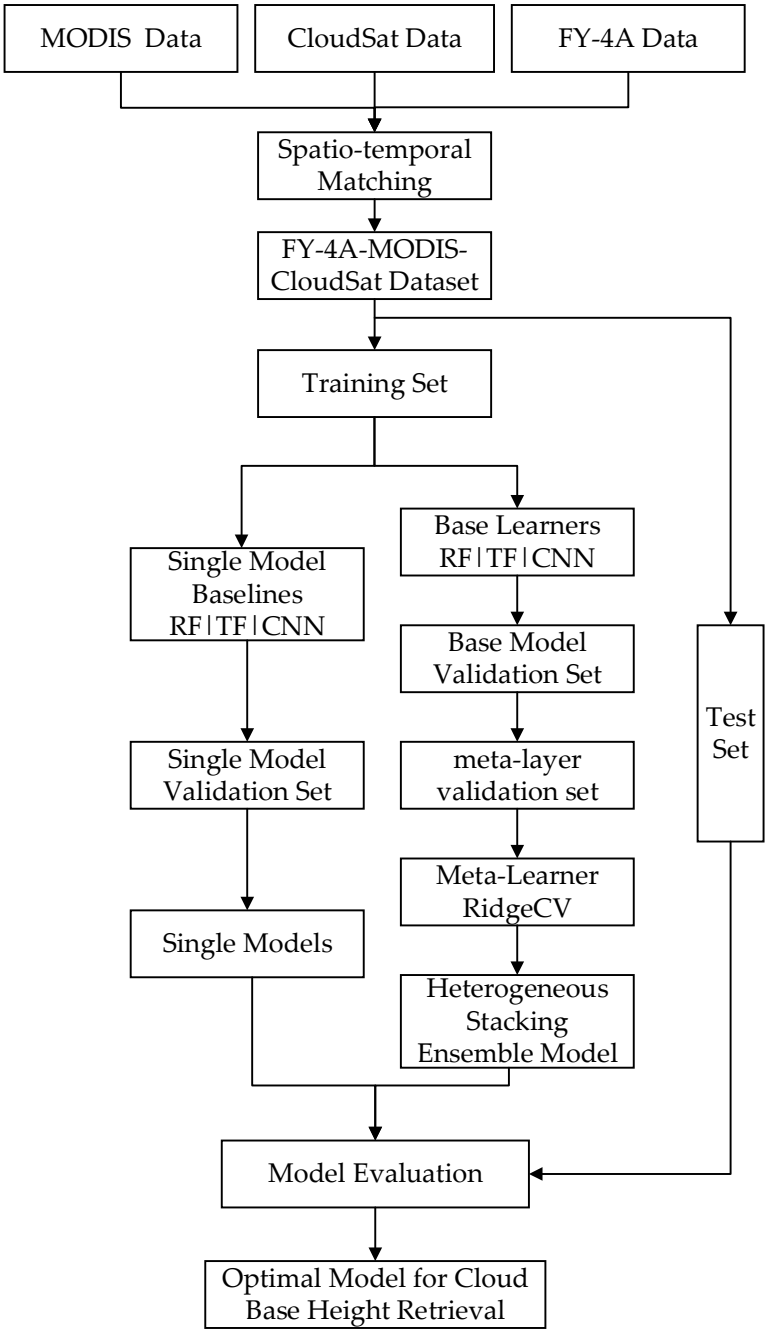


Figure 1. Overall flowchart of CBH inversion using Hetero-CBHNet.

2.2.2. Base Model Design

The random forest (RF) model employs a bagging ensemble strategy, generating multiple training subsets via bootstrap sampling. On this basis, 200 decision trees are constructed, with each tree limited to a maximum depth of 20. When splitting nodes, each tree searches for the optimal split only within a randomly selected subset of features, and the final output is the average of the predictions from all trees. This design makes RF less sensitive to outliers, giving it advantages when processing noisy satellite remote sensing data.

Convolutional neural network (CNN) models perform convolution operations along the feature dimension. To ensure stable training as the network deepens, residual connections are incorporated into the architecture to mitigate the vanishing gradient problem. In addition, a squeeze-and-excitation (SE) channel attention module is embedded, enabling the model to adaptively determine which feature channels are more critical during training

and to assign higher weights to important channels, thereby making fuller use of relevant information.

The Transformer model employs a four-head attention mechanism for neighbour aggregation, enabling it to adaptively learn the importance weights of different neighbouring nodes without relying on predefined edge weights. The importance of nodes within the adjacency structure is derived entirely through multi-head attention; each layer is followed by BatchNorm and the ELU activation function, and, during the decoding stage, a multilayer perceptron (MLP) maps the learned node representations to CBH predictions. As multi-head attention can capture relationships between nodes from different perspectives, the model demonstrates greater flexibility when dealing with complex spatial dependencies and multi-scale cloud field structures.

For the Transformer model, the graph structure between satellite observation samples must be constructed in advance. This study adopts a dual graph reconstruction strategy (Dual KNN Graph) to address this issue: on the one hand, a geospatial KNN graph is constructed based on latitude and longitude coordinates using the Haversine spherical distance metric to represent the continuity of clouds in physical space; on the other hand, after standardising cloud attribute features using RobustScaler based on scikit-learn version 1.5.1, a feature-space KNN graph is constructed using Euclidean distance as the metric and twice the median feature distance as the threshold to reflect the similarity of clouds in terms of their physical characteristics. These two graphs characterise the relationships between samples from different perspectives, providing a structural foundation for subsequent aggregation operations in the graph network. The Haversine formula is calculated as follows:

$$d = 2r\arcsin\left(\sqrt{\sin^2\left(\frac{\Delta lat}{2}\right) + \cos(lat_1)\cos(lat_2)\sin^2\left(\frac{\Delta lon}{2}\right)}\right) \quad (1)$$

In this context, the nodes of the graph structure are the matched sample points i. $i = (lon_1, lat_1), (lon_2, lat_2), \ldots, (lon_n, lat_n)$, two separate KNN subgraphs are constructed: a spatial graph comprising 20 nearest neighbours and a feature graph comprising 15 nearest neighbours. The edge weights are determined by a Gaussian kernel, where r is the radius of the Earth (approximately 6371 km), $\Delta lat = lat_2 - lat_1$, $\Delta lon = lon_2 - lon_1$.

2.2.3. Meta-Learner

This study compares two fusion strategies. The 'Simple Average' method, which calculates the arithmetic mean of the predictions from each base model without weight learning, serves as a non-learning control; the core advantage of RidgeCV lies in its ability to control model complexity through linear constraints, providing greater stability in cross-temporal data transfer and reducing the likelihood of overfitting caused by weights becoming over-adapted to a specific time period [14].

The specific definition of RidgeCV is as follows: let the prediction vector of the base models for the i-th sample in the validation set be $p_i = \left[\hat{y}_i^{RF}, \hat{y}_i^{CNN}, \hat{y}_i^{TF}\right]^T$, and let the corresponding true CBH be y_i. The meta-learner solves for the linear weights $w = [w_1, w_2, w_3]^T$ and bias b, so that the fused prediction is as close as possible to the true value. The optimisation objective is:

$$\min_{w,b} \sum_{i=1}^{N_{val}} \left(y_i - \left(w^T p_i + b\right)\right)^2 + \lambda \|w\|_2^2 \quad (2)$$

Here, λ is the regularisation coefficient. This objective function comprises two terms: The first term is the sum of squared errors, which measures the model's fit to the validation set samples. The second term is the L2 norm penalty, ($\|w\|_2^2$), which is the core constraint of RidgeCV; it restricts the numerical range of the weight parameters, thereby preventing

the weight of any single base model from becoming excessively large and causing unstable fusion results. RidgeCV solves for unconstrained linear weights with a bias; therefore, the sum of the weights need not equal 1. Because the target values are standardised before training, the fused prediction is obtained in the standardised space. When it is transformed back to the original space via the scaler, any deviation of the sum of the weights from 1 is compensated by the centre parameter, and this does not affect the inversion accuracy.

3. Results

3.1. Base Model Inversion Results

Following the workflow shown in Figure 1, CBH inversion results were obtained for the three candidate models in the pool used as a comparison baseline, using FY-4A and MODIS data.

As shown in Table 2, RF achieves the best accuracy on the FY-4A data, with an RMSE of 3000.21 m and an R of 0.6208; on the MODIS data, its RMSE is 2856.57 m, and its R is 0.6414. The Transformer and CNN models are slightly less accurate than RF: on the FY-4A data, the RMSE values of the Transformer and CNN models are 3017.81 m and 3363.62 m, respectively, with R values of 0.6148 and 0.5177. On the MODIS data, the Transformer achieves the best accuracy, with an RMSE of 2829.30 m and an R of 0.6524, while the CNN and RF models have RMSE values of 2843.48 m and 2856.57 m, respectively, with R values of 0.6448 and 0.6414. In terms of data sources, the errors of all models are generally lower on the MODIS data than on the FY-4A data, which is related to differences in product resolution and information content.

Table 2. Single-model CBH inversion results on the FY-4A and MODIS test datasets.

Model	FY-4A				MODIS			
	RMSE/m	MAE/m	MRE/%	R	RMSE/m	MAE/m	MRE/%	R
RF	3000.21	2128.34	157.64	0.6208	2856.57	1966.72	146.88	0.6414
CNN	3363.62	2336.99	136.05	0.5177	2843.48	1904.51	138.08	0.6448
Transformer	3017.81	2094.66	148.33	0.6148	2829.30	1882.28	134.28	0.6524

3.2. Results of the Hetero-CBHNet Integration

3.2.1. Meta-Learner Learning Strategies

An exhaustive search was performed over the combinations within the candidate pool using different fusion strategies on the meta-layer validation set. Table 3 presents the accuracy metrics for representative configurations on the meta-layer validation set, together with the regularisation coefficient λ automatically selected by RidgeCV through its built-in cross-validation. Table 4 lists the fusion weights assigned to each base model by RidgeCV.

Table 3. Results for the best single model and Hetero-CBHNet on the meta-layer validation set.

Model	Ensemble Strategy	FY-4A				MODIS			
		RMSE/m	MAE/m	R	λ	RMSE/m	MAE/m	R	λ
RF	/	3144.12	2278.68	0.6351	/	3091.90	2227.55	0.6401	/
Transformer	/	3159.48	2250.07	0.6289	/	3064.90	2149.49	0.6478	/
Hetero-CBHNet (RF + CNN + Transformer)	RidgeCV	3112.16	2175.09	0.6418	2.976	3034.49	2083.26	0.6539	1.438
RF + Transformer	RidgeCV	3115.12	2177.59	0.6410	2.976	3035.60	2083.48	0.6536	2.976
CNN + RF	RidgeCV	3125.67	2201.33	0.6378	0.336	3059.29	2126.26	0.6467	1.438
CNN + Transformer	RidgeCV	3151.58	2205.20	0.6300	0.336	3046.56	2094.50	0.6504	0.695

Table 4. Fusion weights of the base models in Hetero-CBHNet on the meta-layer validation set.

Model	FY-4A				MODIS			
	RF	**CNN**	**Transformer**	**Weight Sum**	**RF**	**CNN**	**Transformer**	**Weight Sum**
Hetero-CBHNet (RF + CNN + Transformer)	0.6530	0.1296	0.3965	1.1791	0.3969	0.1165	0.6013	1.1147
RF + Transformer	0.6596	/	0.4542	1.1138	0.4532	/	0.6740	1.1273
CNN + RF	0.9951	0.2162	/	1.2113	0.6434	0.4237	/	1.0671
CNN + Transformer	/	0.1456	0.9939	1.1394	/	0.2869	0.7846	1.0715

Based on Tables 3 and 4, learned weight fusion outperforms simple averaging with equal weights. Under the RidgeCV strategy, the three-model combination achieves RMSE values of 3112.16 m and 3034.49 m on the meta-layer validation set for the two data sources, respectively, which are 87.8 m and 12.5 m lower than the corresponding simple averaging results of 3199.98 m and 3046.98 m. The regularisation coefficient λ is automatically selected by cross-validation, with values of 2.976 and 1.438 on the FY-4A and MODIS data, respectively, thereby suppressing overfitting in RidgeCV.

On both the test set and the meta-layer validation set, the best single model varies with the data source; it is RF for FY-4A and Transformer for MODIS. This indicates that relying on a single model for CBH inversion is unstable, and that the choice of the best model depends on the data source and the evaluation protocol. This is precisely the rationale for adopting a heterogeneous ensemble network architecture in this study, which uses multiple models to reduce such instability.

As shown in Table 4, under the RidgeCV strategy, in the three-model Hetero-CBHNet architecture, RF has the largest weight for the FY-4A data source, at 0.6530; the Transformer weight is 0.3965, the CNN weight is 0.1296, and the weight sum is 1.1791. For the MODIS data source, Transformer has the largest weight, at 0.6013; the RF weight is 0.3969, the CNN weight is 0.1165, and the weight sum is 1.1147. Among the two-model combinations, the weight sums of RF + Transformer are 1.1138 and 1.1273, respectively; those of CNN + RF are 1.2113 and 1.0671, respectively; and those of CNN + Transformer are 1.1394 and 1.0715, respectively. All weight sums deviate from 1, which is consistent with RidgeCV solving for unconstrained weights with a bias term. Because the target values are standardised before training, the fused prediction must be transformed back to the original space by the scaler. The deviation of the weight sum from 1 is automatically compensated for by the centre parameter (intercept), and thus no normalisation constraint needs to be imposed on the weights.

Based on the combined results in Tables 3 and 4, the Hetero-CBHNet architecture emerges as the optimal candidate for CBH inversion. The final model will then be evaluated on the test set, thereby further confirming the effectiveness of this architecture.

3.2.2. Hetero-CBHNet's Results on the Test Set

The test set is used only once, after all training and selection processes have been completed, to verify whether the Hetero-CBHNet architecture described in Section 3.2.1 can reliably invert CBH on independent data. Table 5 presents the accuracy metrics for the same set of representative configurations on the test set.

Table 5. Results for the best single model, Hetero-CBHNet, and simple averaging on the test set.

Model	Ensemble Strategy	FY-4A				MODIS			
		RMSE/m	MAE/m	MRE/%	R	RMSE/m	MAE/m	MRE/%	R
RF	/	3000.21	2128.34	157.64	0.6208	2856.57	1966.72	146.88	0.6414
CNN	/	3363.62	2336.99	136.05	0.5177	2843.48	1904.51	138.08	0.6448
Hetero-CBHNet (RF + CNN + Transformer)	RidgeCV	2966.12	2033.13	145.69	0.6303	2797.64	1828.05	131.55	0.6590
RF + Transformer	RidgeCV	2970.55	2038.40	145.83	0.6290	2802.04	1888.12	151.18	0.6577
CNN + RF + Transformer	SimpleAvg	3035.58	2145.55	145.73	0.6258	2804.90	1891.64	138.28	0.6590

Table 5 presents the inversion accuracy of the representative configurations on the test sets. Hetero-CBHNet achieves the best overall performance on both data sources. On the FY-4A test set, Hetero-CBHNet obtains an RMSE of 2966.12 m, an MAE of 2033.13 m, and an R of 0.6303; on the MODIS test set, it obtains an RMSE of 2797.64 m, an MAE of 1828.05 m, and an R of 0.6590. Compared with the best single model for each data source, Hetero-CBHNet reduces the RMSE by 34.09 m on FY-4A, from 3000.21 m to 2966.12 m, and by 31.66 m on MODIS, from 2829.30 m to 2797.64 m. Its MAE and R are also better than those of the best single models, indicating that the ensemble framework maintains a stable accuracy advantage on the independent test sets.

A further comparison of the fusion strategies shows that the three-model combination with RidgeCV-learned weights outperforms simple averaging with equal weights. On the FY-4A test set, the RMSE of simple averaging is 3035.58 m, which is 69.46 m higher than that of Hetero-CBHNet; its MAE is 2145.55 m, also clearly higher than the 2033.13 m of Hetero-CBHNet. On the MODIS test set, the RMSE of simple averaging is 2804.90 m, which is 7.26 m higher than that of Hetero-CBHNet; its R is 0.6590, comparable to that of Hetero-CBHNet, whereas its MAE and MRE remain inferior to those of Hetero-CBHNet. These results indicate that weighted fusion can allocate weights according to the actual contribution of each base model, thereby outperforming equal-weight averaging overall.

The results of the two-model combinations also support this conclusion. On the FY-4A test set, the RMSE of the RF and Transformer two-model combination is 2970.55 m, which is 4.43 m higher than that of Hetero-CBHNet; on the MODIS test set, its RMSE is 2802.04 m, which is 4.40 m higher than that of Hetero-CBHNet. Although the two-model combinations already achieve good accuracy, the three-model heterogeneous combination remains superior on both data sources, indicating that the complementarity among different model paradigms is further exploited by incorporating CNN.

Based on the above analysis, it can be confirmed that the Hetero-CBHNet framework, which uses RF, CNN, and Transformer as its pool of candidate base models and employs RidgeCV to learn the fusion weights, is the optimal model framework for CBH inversion on both the FY-4A and MODIS data sources. This framework not only outperforms every single model on the test sets, but also outperforms simple averaging and the two-model combinations, further validating the effectiveness of heterogeneous ensemble learning with data-driven model selection.

3.3. Sensitivity Analysis

3.3.1. Splitting Protocol Sensitivity Analysis: Random Splitting Control

All sections of the main text adopt a strict forward chronological splitting protocol. To examine whether the conclusions depend on this protocol, this section adds a random splitting control under the same dataset, the same features, and the same model configura-

tion: the training, validation, and test sets are pooled into a single set of 1,011,620 samples and then randomly assigned by sample, with 728,367 sample points in the training and validation set, 111,278 sample points in the meta-layer validation set, and 171,975 sample points in the test set, while all other settings remain unchanged.

Table 6 presents the inversion results of each model on the test sets under the random splitting protocol. Hetero-CBHNet achieves the best performance on both the FY-4A and MODIS data sources, with RMSE values of 1510.01 m and 1488.41 m, MAE values of 691.12 m and 688.77 m, and R values of 0.9168 and 0.9147, respectively, outperforming the three single models, CNN, RF, and Transformer. The RMSE values of CNN are 1556.20 m and 1529.16 m, those of RF are 1712.87 m and 1737.69 m, and those of Transformer are 2734.31 m and 2716.35 m. These results indicate that, under the random splitting protocol, Hetero-CBHNet still outperforms all single models, consistent with the conclusions under the main protocol in the main text.

Table 6. Single-model CBH inversion results on the FY-4A and MODIS test datasets under the random splitting protocol.

Model	FY-4A			MODIS		
	RMSE/m	MAE/m	R	RMSE/m	MAE/m	R
Hetero-CBHNet	1510.01	691.12	0.9168	1488.41	688.77	0.9147
CNN	1556.20	713.63	0.9121	1529.16	704.51	0.9105
RF	1712.87	976.79	0.8983	1737.69	1004.71	0.8891
Transformer	2734.31	1834.82	0.6934	2716.35	1799.75	0.7030

In addition, under random splitting, the best single model is CNN for both data sources, whereas under the chronological splitting protocol in the main text, the best single model is RF for FY-4A and Transformer for MODIS. That is, the ranking of the best single model changes with the splitting protocol, further indicating that the relative superiority of single models is not a stable property but is highly sensitive to the evaluation protocol. This phenomenon is consistent with the core argument of this study, namely that the model architecture should not be fixed in advance but should be selected in a data-driven manner.

In terms of overall accuracy, the results under the random splitting protocol are clearly better than those under the chronological splitting protocol. The RMSE of Hetero-CBHNet decreases from 2966.12 m and 2797.64 m under chronological splitting to 1510.01 m and 1488.41 m under random splitting, with reductions of 1456.11 m and 1309.23 m, respectively, corresponding to an overall accuracy inflation of about 1.3 to 1.5 km. This discrepancy mainly arises because samples from the same orbit, the same time step, and temporally adjacent periods appear in both the training and test sets, allowing the model to exploit the spatial and temporal correlations among samples, which produces an optimistic bias introduced by the evaluation protocol. In contrast, the chronological splitting protocol fully separates the training, validation, and test data in time, which is closer to operational deployment, and its accuracy results better reflect the true generalisation ability of the model.

Therefore, this study adopts the chronological splitting protocol as the main protocol and uses random splitting as a control to illustrate the substantial influence of the evaluation protocol itself on the conclusions regarding CBH inversion.

3.3.2. Passive Cloud Type Sensitivity Analysis

To quantitatively analyse the impact of passive cloud type information on CBH inversion, following the workflow shown in Figure 1, bar chart were used to compare the overall inversion accuracy of the three base models, CNN, RF, and Transformer, as well as

Hetero-CBHNet, under two scenarios: with and without cloud type input. For FY-4A, the operational L2 cloud type product CLT was used, whereas for MODIS, since no operational product of the same family is available, an ISCCP-style cloud type was derived from its own CTP and COT; both are purely passive-source data. The results are shown in Figure 2.

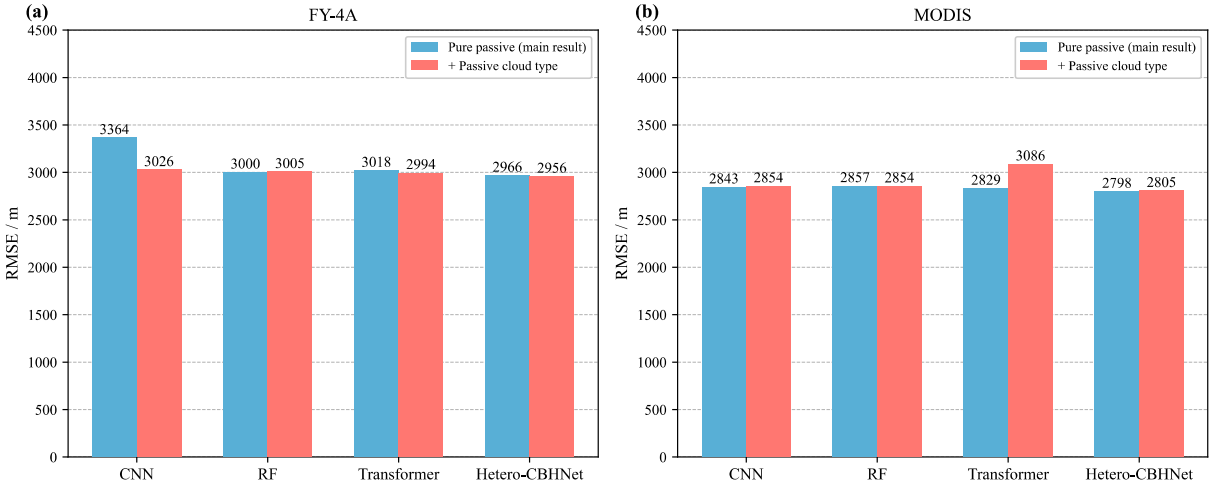


Figure 2. Ablation of the passive cloud type prior: comparison between purely passive input and input with passive cloud type information on the test set: (**a**) results for the FY-4A data source; (**b**) results for the MODIS data source.

The results are shown in Figure 2. After incorporating passive cloud types, the test-set RMSE is 2956.29 m for FY-4A, compared with a purely passive baseline of 2966.12 m, a reduction of 9.8 m; for MODIS, it is 2805.44 m, compared with a baseline of 2797.64 m, an increase of 7.8 m. The direction of change differs between the two data sources, and both magnitudes fall within the range of operational noise. Taking the deterministic RF model as a reference, the corresponding changes are an increase of 4.7 m and a decrease of 3.0 m, respectively, again showing no consistent direction.

This phenomenon indicates that neither the operational passive cloud type product nor the ISCCP-style cloud type derived from MODIS products alone provides a stable gain for CBH inversion. A possible reason is that passive cloud types are themselves inferred from cloud-top radiance signals and thus lack a direct physical constraint linking them to CBH. In addition, passive cloud types have limited ability to discriminate multi-layer, thin, and overlapping clouds, whereas the analysis in Section 3.4.2 of this study has shown that multi-layer clouds are precisely the main source of CBH inversion error. Therefore, passive cloud types still differ substantially from the structural prior required for CBH inversion in terms of information content. At the present stage, passive cloud types are not yet sufficient to replace active satellite labels.

3.4. Error Analysis

3.4.1. Error Analysis Across Cloud Types

To analyse the errors across different cloud types, after the test-set evaluation was completed, the test samples were divided into eight classes, namely Ac, As, Ci, Cu, Dc, Ns, St, and Sc, based on the CloudSat 2B-CLDCLASS-LIDAR cloud type product. This cloud type information was used only for post hoc error analysis and was not used as model input. Bias is defined as the mean difference between the retrieved CBH and the measured CBH, that is, the mean of retrieved CBH minus measured CBH. A positive bias indicates overestimation, while a negative bias indicates underestimation. Table 7 and

Figure 3 present the error decomposition results of Hetero-CBHNet across the eight cloud types for the two data sources.

Table 7. Error of Hetero-CBHNet by cloud type on the test set.

Cloud Type	N	FY-4A		MODIS	
		RMSE/m	Bias/m	RMSE/m	Bias/m
Ci	31,801	5013.2	−4060	4836.4	−3820
As	29,907	2106.0	−416	2007.0	−638
Ac	30,589	2119.6	+196	2090.2	+208
St	3522	936.9	+401	877.5	+566
Sc	60,533	1652.0	+878	1542.7	+894
Cu	29,195	3906.7	+3016	3933.6	+3006
Ns	18,608	1554.3	+942	1247.1	+813
Dc	5222	1647.7	+1226	1597.1	+1238

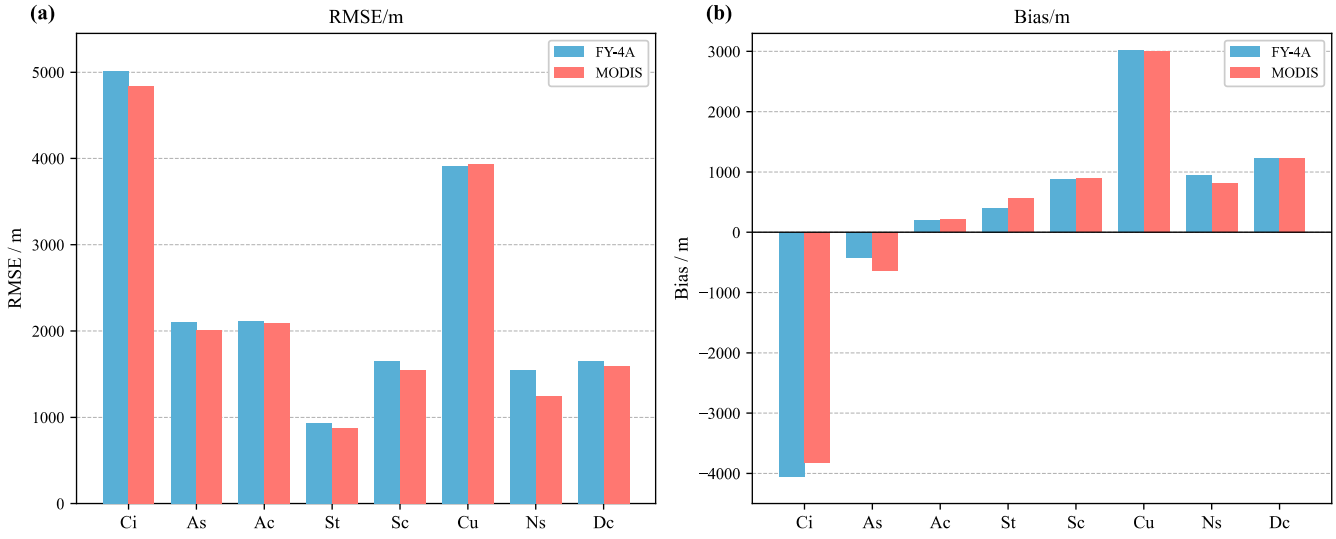


Figure 3. Error decomposition by cloud type under passive input on the test set: (**a**) RMSE for each cloud type; (**b**) bias for each cloud type. Blue and red bars represent the FY-4A and MODIS data sources, respectively.

The errors show a clear dependence on cloud type. Ci and Cu are the main sources of error: the RMSE of Ci reaches 5013.2 m and 4836.4 m, with systematic underestimation, showing biases of −4060 m and −3820 m, because the cloud base of Ci is extremely high, with a mean of about 10.7 km, and passive signals can hardly constrain its geometric thickness; the RMSE of Cu is 3906.7 m and 3933.6 m, with systematic overestimation, showing biases of +3016 m and +3006 m, because 78.3% of the Cu samples have a multi-layer structure, in which the lowest layer is low Cu and an As layer exists above it; the cloud top observed by passive sensors comes from the upper layer, which pulls the estimated cloud base upward.

The low cloud types perform best: the RMSE of St is 936.9 m and 877.5 m, that of Ns is 1554.3 m and 1247.1 m, and that of Sc is 1652.0 m and 1542.7 m. In terms of error contribution, Ci and Cu together account for about 30% of the samples, 31.5% for FY-4A and 29.1% for MODIS, yet they contribute about 73% of the total mean squared error, 73.5% and 73.0% for the two sources, respectively, indicating that the overall accuracy is essentially determined by these two cloud types. The RMSE values of the remaining six types all range from 0.9 to 2.2 km: Dc has the fewest samples but its error remains controllable, at 1647.7 m

and 1597.1 m; among the middle-level clouds, As shows a negative bias of −416 m and −638 m, while the bias of Ac is close to zero, at +196 m and +208 m.

3.4.2. Error Decomposition by Number of Cloud Layers

This study does not exclude multi-layer cloud samples but retains them in full, and provides an error decomposition by cloud layer number on the test set, as shown in Table 8.

Table 8. Error decomposition by cloud layer number on the test set.

Cloud Type	N (Percentage)	FY-4A		MODIS	
		RMSE/m	Bias/m	RMSE/m	Bias/m
Single layer	107,561 (51.4%)	2863.7	−861	2639.4	−685
Two layers	68,717 (32.8%)	2933.9	+684	2808.1	+695
Three or more layers	33,099 (15.8%)	3315.7	+1797	3240.5	+1730

In the test set, the proportion of samples containing two or more cloud layers is 49.3% for FY-4A and 48.6% for MODIS. The results show a clear pattern: single-layer cloud samples are systematically underestimated, with biases of −861 m for FY-4A and −685 m for MODIS, because passive signals are dominated by the cloud top and the model tends to overestimate cloud thickness. As the number of layers increases, the error increases monotonically and shifts from underestimation to overestimation; the bias for two-layer samples is about +0.7 km, and for samples with three or more layers, the RMSE reaches 3315.7 m and 3240.5 m, with a bias of about +1.8 km. This indicates that passive methods have an information ceiling in multi-layer cloud scenarios: the radiance signal is dominated by upper-layer clouds, and the base of the lowest layer cannot be directly observed.

This decomposition provides a quantitative description of the application boundary of the proposed method: the method is applicable to all cloud types, but its accuracy is limited under multi-layer cloud conditions and is best under single-layer cloud conditions. Further decomposing the RMSE into a systematic bias component and a random component shows that the random term is almost constant across the three layer-count bins, at about 2.7 to 2.9 km for FY-4A and about 2.5 to 2.7 km for MODIS, whereas the proportion of the systematic bias in the mean squared error increases from 7% to 9% for single-layer clouds to about 29% for three or more layers. This indicates that the accuracy loss for multi-layer clouds arises almost entirely from directional systematic error; that is, from missing information, rather than from a decline in model stability.

4. Discussion

4.1. Sensitivity of Models to Different Protocols

Model selection is clearly sensitive to the evaluation protocol. The best single model in this study is not consistent across different evaluation settings: on the test set, the best model for FY-4A is RF, with an RMSE of 3000.21 m, while the best model for MODIS is Transformer, with an RMSE of 2829.30 m; under the meta-validation set, the random splitting protocol, and the five-fold cross-validation protocol detailed in Appendix A, the model rankings also differ. For the same task and the same data, the best single model switches among RF, CNN, and Transformer solely because of differences in the evaluation setting, indicating that the conclusion that "a certain model is optimal" depends strongly on the splitting protocol rather than on the intrinsic superiority of the model. Meanwhile, Hetero-CBHNet remains leading or close to leading under all settings, and its advantage does not depend on which single model is optimal.

In addition, the RMSE under strict chronological splitting is 2966.12 m and 2797.64 m, whereas under random splitting it is 1510.01 m and 1488.41 m, a difference of about 1.3 to 1.5 km. Random splitting causes neighbouring samples from the same orbit and the same time step to enter the training and test sets separately, allowing the model to exploit the spatial correlation among nearby samples, which leads to markedly optimistic reported accuracy. The complete random splitting control is provided in Section 3.3.1. On this basis, this study recommends that, for large-scale spatially matched data, results under chronological splitting should be reported preferentially.

4.2. Comparison with Existing Results

This study reviews representative CBH inversion studies in recent years. These studies differ significantly in sensors, target-layer definitions, sample screening, and splitting protocols, and their RMSE values are therefore not directly comparable. Accordingly, when citing previous results, this section also specifies their protocols and provides the results of this study under the corresponding protocols.

Ye et al. [6] used Himawari-8 infrared channels, employed a CNN together with an effective cloud water content lookup table, retained only single-layer clouds, excluded deep convective samples, and required the cloud top height deviation to be less than 2 km, reporting an RMSE of about 1260 m. Jiang et al. [19] used FY-4A top-of-atmosphere radiance channels and L2 cloud top height, adopted a random forest algorithm, used only single-layer cloud samples, and reported an RMSE of about 1520 m under random splitting. Yu et al. [11] used cloud top height, cloud optical thickness, and cloud effective radius from FY-4A, adopted random forest and gradient boosting tree algorithms, and introduced truth cloud types under all cloud types, reporting an RMSE of about 1305 m, while the unified model without cloud types gave 2109 m.

The above results indicate that the reported values on the order of 1.2 to 1.5 km in the literature are generally built upon single-layer cloud screening, active-satellite auxiliary information, or random splitting. Under the main protocol of this study, namely all cloud types, strict chronological splitting, and purely passive input, the RMSE values of FY-4A and MODIS are 2966.12 m and 2797.64 m, respectively, which represents a more difficult benchmark; the higher values stem from the stricter task formulation rather than insufficient model capability.

Under the random splitting protocol, Hetero-CBHNet achieves RMSE values of 1510.01 m and 1488.41 m on FY-4A and MODIS, respectively, without restricting to single-layer clouds and without using any active-satellite prior, which is comparable to existing studies. When employing the same three-stage screening method as in the study by Ye et al., the root mean square error (RMSE) in this study, under the temporal division, was reduced to 1474.3 and 1779.2 m.

It can therefore be seen that the evaluation protocol and sample screening have a much greater impact on the reported accuracy than differences in model architecture. This study therefore does not advocate measuring method superiority solely by a single RMSE value, but instead emphasises comparability after protocol alignment. Under the same protocol, the proposed method achieves accuracy comparable to that of existing studies; under the stricter formulation of all cloud types, purely passive input, and chronological splitting, the results of this study better reflect the difficulty of CBH inversion in real operational scenarios.

4.3. Boundary of Cloud Type Priors and the Information Ceiling of Multi-Layer Clouds

The ablation experiments and the error decomposition by cloud layer number indicate that the accuracy ceiling of passive CBH inversion does not mainly depend on the model

architecture, but is constrained by the information content that passive observations can provide. This is manifested in two aspects: insufficient cloud type priors and the lack of vertical structure information for multi-layer clouds.

Cloud type can provide height-level, phase, and vertical structure information that passive spectral features can hardly encode, and thus serves as an important structural prior. However, the ablation experiments in this study show that current passive cloud types are not yet sufficient to replace active-satellite cloud types: the self-trained passive classifier achieves an accuracy of about 46% across eight cloud types, and the ISCCP-style cloud type derived from MODIS CTP and COT alone achieves a height-level agreement of 60.1% for single-layer clouds, which drops to 21.2% to 36.4% for multi-layer clouds. Consequently, incorporating passive cloud types does not yield a stable gain. Developing a high-accuracy passive cloud type classifier supervised by active-satellite labels is therefore key to unlocking the value of cloud type priors.

Multi-layer clouds reflect another information ceiling. Table 8 shows that, as the number of cloud layers increases, the error increases monotonically and the bias shifts from negative to positive: single-layer clouds are systematically underestimated, the bias for two-layer clouds becomes positive, and for samples with three or more layers, the RMSE increases further, and the bias is about +1.8 km. This pattern corroborates the overestimation of Cu and the underestimation of Ci in Section 3.4.1. Their common root lies in the fact that passive radiance signals mainly originate from the uppermost cloud layer, and the base and geometric thickness of the lowest layer cannot be directly observed. The random term remains almost constant across the three layer-count bins, whereas the proportion of the systematic bias in the mean squared error increases from 7% to 9% for single-layer clouds to about 29% for three or more layers, indicating that the accuracy loss for multi-layer clouds mainly arises from missing information rather than from a decline in model stability.

In summary, cloud type priors and multi-layer cloud structures jointly define the information boundary of passive CBH inversion. Future work can seek incremental constraints from two directions: first, constructing a high-accuracy passive cloud type classifier supervised by active-satellite labels; second, introducing temporal differential motion information, exploiting the fact that clouds at different altitudes have different motion vectors, and combining it with finer profile auxiliary information to provide additional vertical structure constraints for multi-layer cloud scenarios.

5. Conclusions

CBH is an important parameter for describing the vertical structure of clouds, but passive satellites struggle to accurately obtain CBH information. To address this challenge, this study constructs a heterogeneous stacked integrated network, Hetero-CBHNet, which uses RF, CNN, and Transformer models as a pool of candidate base models and employs a RidgeCV meta-learner to learn the fusion weights and select the best combination from all combinations within the pool. Comparative experiments under strict chronological splitting were conducted on matched FY-4A and MODIS datasets. The experimental conclusions are as follows:

(1) Hetero-CBHNet uses RidgeCV to learn the fusion weights of the candidate base model pool, rather than simply stacking the models directly. Under strict chronological splitting, the RMSE values on the FY-4A and MODIS test sets are 2966.12 m and 2797.64 m, respectively, with R values of 0.630 and 0.659; under random splitting, the RMSE values are 1510.01 m and 1488.41 m, respectively. Under both protocols, Hetero-CBHNet outperforms all single models, indicating that the complementary synergy of heterogeneous models provides a stable performance improvement for CBH inversion.

(2) Incorporating passive cloud type products does not yield a stable gain in accuracy: the RMSE for FY-4A is 2956.29 m, compared with a baseline of 2966.12 m; the RMSE for MODIS is 2805.44 m, compared with a baseline of 2797.64 m. The directions of change for the two data sources are inconsistent, and the magnitudes are within the range of operational noise, indicating that existing passive cloud type information is not yet sufficient to replace active-satellite labels.

(3) The error decomposition for multi-layer clouds shows that about 49% of the samples in the test set contain two or more cloud layers. The bias for single-layer cloud samples is about −0.9 km and −0.7 km, indicating systematic underestimation; for samples with three or more layers, the RMSE reaches 3.32 km and 3.24 km, with a bias of about +1.8 km, indicating systematic overestimation. Further decomposed by cloud type, the RMSE of cirrus is about 4.8 to 5.0 km, and that of cumulus is about 3.9 km, making them the main sources of error.

Author Contributions: Conceptualization, T.X. and Y.H.; methodology, Y.H.; software, Y.H.; validation, Y.H. and T.X.; formal analysis, Y.H. and T.X.; investigation, Y.H. and J.L.; resources, Y.C. and W.L.; data curation, Y.H.; writing—original draft preparation, Y.H.; writing—review and editing, Y.H. and C.W.; visualization, Y.H.; supervision, T.X.; project administration, J.L.; funding acquisition, T.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 42176180; the Fengyun Application Pioneering Project, grant number FY-APP-2024.0408; and the Natural Science Research Start-up Foundation of Recruiting Talents of Nanjing University of Posts and Telecommunications, grant number NY226003.

Data Availability Statement: The data used in this study are publicly available. The FY-4A/AGRI Level-2 cloud products were obtained from the Fengyun Meteorological Satellite Remote Sensing Data Service Platform (https://satellite.nsmc.org.cn/DataPortal/cn/data/order.html, accessed on 15 May 2025). The MODIS/Aqua CLDPROP_L2_MODIS_Aqua cloud products were obtained from the NASA Level-1 and Atmosphere Archive and Distribution System (LAADS) Distributed Active Archive Center (DAAC) (https://ladsweb.modaps.eosdis.nasa.gov/search/order/2/CLDPROP_L2_MODIS_Aqua, accessed on 15 May 2025). The CPR-CALIOP 2B-CLDCLASS-LIDAR dataset was obtained from the CloudSat Data Processing Center (https://www.cloudsat.cira.colostate.edu/data-products/2b-cldclass-lidar, accessed on 15 May 2025).

Acknowledgments: The authors would like to sincerely thank the National Satellite Meteorological Center (NSMC) for providing the FY-4A/AGRI data, and the National Aeronautics and Space Administration (NASA) for providing the MODIS and CloudSat data.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Five-Fold Cross-Validation of the Base Models

The main text uses January 2020 as the epoch-selection validation set to determine the training epochs of each base model. To examine whether epoch selection depends on this single-month choice, this study re-performs epoch selection within the training period, namely from May 2019 to January 2020, using five-fold cross-validation: the training-period samples are divided into five folds after being blocked by month, training and validation are performed fold by fold, and the optimal epoch of each base model is recorded; the median of the optimal epochs across the five folds is taken as the final training budget. The meta-layer validation set, namely February 2020, the test set, namely March to May 2020, and all other settings remain unchanged. The results of each model and the ensemble model on the test set under the five-fold protocol are shown in Table A1.

Table A1. Accuracy of each model on the test set under the five-fold cross-validation epoch-selection protocol.

Model	FY-4A				MODIS			
	RMSE/m	MAE/m	MRE/%	R	RMSE/m	MAE/m	MRE/%	R
Hetero-CBHNet	2962.09	2033.79	144.80	0.6319	2800.30	1823.04	130.11	0.6584
CNN	3036.43	2046.07	131.03	0.6134	2865.35	1912.74	133.93	0.6412
RF	3000.21	2128.34	157.64	0.6208	2856.57	1966.72	146.88	0.6414
Transformer	3028.24	2113.30	147.34	0.6152	2831.41	1910.03	137.91	0.6529

The results under the five-fold and single-month epoch-selection protocols are almost identical. Hetero-CBHNet achieves RMSE values of 2962.09 m and 2800.30 m on the FY-4A and MODIS test sets, respectively, which differ from the 2966.12 m and 2797.64 m obtained in the main text using January 2020 alone as the validation set by less than 5 m, within the range of operational noise. The gains of the ensemble over the respective best single model, and the ranking of the best single model, namely RF for FY-4A and Transformer for MODIS, are both consistent with the main text. These results indicate that the conclusions of this study regarding ensemble effectiveness and model ranking do not depend on the validation month used for epoch selection. In addition, five-fold cross-validation is less computationally efficient for epoch selection than using a single validation month, and therefore a single-month validation is adopted as the main protocol in this study.

References

1. Viúdez-Mora, A.; Costa-Surós, M.; Calbó, J.; González, J.A. Modeling Atmospheric Longwave Radiation at the Surface during Overcast Skies: The Role of Cloud Base Height. *J. Geophys. Res. Atmos.* **2015**, *120*, 199–214. [CrossRef]
2. Herzegh, P.; Wiener, G.; Bateman, R.; Cowie, J.; Black, J. Data Fusion Enables Better Recognition of Ceiling and Visibility Hazards in Aviation. *Bull. Am. Meteorol. Soc.* **2015**, *96*, 526–532. [CrossRef]
3. Miller, S.D.; Forsythe, J.M.; Partain, P.T.; Haynes, J.M.; Bankert, R.L.; Sengupta, M.; Mitrescu, C.; Hawkins, J.D.; Vonder Haar, T.H. Estimating Three-Dimensional Cloud Structure via Statistically Blended Satellite Observations. *J. Appl. Meteorol. Climatol.* **2014**, *53*, 437–455. [CrossRef]
4. Luo, E.; Chen, G.; Wang, W.-C.; Feng, J.; Gao, Y. Effects of Cloud Vertical Structure on the Development of Tropical Cyclones: A Case Study Based on In-Fa (2021). *Atmos. Res.* **2025**, *315*, 107904. [CrossRef]
5. Martucci, G.; Milroy, C.; O'Dowd, C.D. Detection of Cloud-Base Height Using Jenoptik CHM15K and Vaisala CL31 Ceilometers. *J. Atmos. Ocean. Technol.* **2010**, *27*, 305–318. [CrossRef]
6. Ye, T.; Tan, Z.; Ai, W.; Ma, S.; Zhao, X.; Hu, S.; Liu, C.; Guo, J. On the Hybrid Algorithm for Retrieving Day and Night Cloud Base Height from Geostationary Satellite Observations. *Remote Sens.* **2025**, *17*, 2469. [CrossRef]
7. Li, R.; Ma, X. Evaluation and Improvement of a CALIPSO-Based Algorithm for Cloud Base Height in China. *Remote Sens.* **2024**, *16*, 2801. [CrossRef]
8. Noh, Y.-J.; Haynes, J.M.; Miller, S.D.; Seaman, C.J.; Heidinger, A.K.; Weinrich, J.; Kulie, M.S.; Niznik, M.; Daub, B.J. A Framework for Satellite-Based 3D Cloud Data: An Overview of the VIIRS Cloud Base Height Retrieval and User Engagement for Aviation Applications. *Remote Sens.* **2022**, *14*, 5524. [CrossRef]
9. Yang, Y.; Sun, W.; Chi, Y.; Yan, X.; Fan, H.; Yang, X.; Ma, Z.; Wang, Q.; Zhao, C. Machine Learning-Based Retrieval of Day and Night Cloud Macrophysical Parameters over East Asia Using Himawari-8 Data. *Remote Sens. Environ.* **2022**, *273*, 112971. [CrossRef]
10. Tan, Z.H.; Ma, S.; Han, D.; Gao, D.; Yan, W. FY-4A Cloud Base Height Estimation Method Based on Random Forest Algorithm. *J. Infrared Millim. Waves* **2019**, *38*, 381–388.
11. Yu, Z.F.; Wang, Y.; Ma, S.; Ai, W.H.; Yan, W. FY-4A Cloud Base Height Retrieval Method Based on Ensemble Learning. *Acta Opt. Sin.* **2023**, *43*, 49–62.
12. Tan, Z.; Ma, S.; Liu, C.; Teng, S.; Letu, H.; Zhang, P.; Ai, W. Retrieving Cloud Base Height from Passive Radiometer Observations via a Systematic Effective Cloud Water Content Table. *Remote Sens. Environ.* **2023**, *294*, 113633. [CrossRef]
13. Lenhardt, J.; Quaas, J.; Sejdinovic, D. Marine Cloud Base Height Retrieval from MODIS Cloud Properties Using Machine Learning. *Atmos. Meas. Tech.* **2024**, *17*, 5655–5677. [CrossRef]

14. Qiu, L.; Cheng, C.; Shi, H.; Wang, X.; Wang, T.; Ye, H. Stack-CBHNet: A Stacked Ensemble Learning Approach for Cloud Base Height Estimation. In Proceedings of the 2025 7th International Conference on Frontier Technologies of Information and Computer (ICFTIC), Qingdao, China, 5–7 December 2025; pp. 129–133. [CrossRef]
15. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022. [CrossRef]
16. Li, T.; Chen, N.; Tao, F.; Hu, S.; Xue, J.; Han, R.; Wu, D. Cloud Top Height Retrieval from FY-4A Data: A Residual Module and Genetic Algorithm Approach. *Atmosphere* **2024**, *15*, 643. [CrossRef]
17. Wang, Q.; Zhou, C.; Letu, H.; Zhu, Y.; Zhuge, X.; Liu, C.; Weng, F.; Wang, M. Obtaining Cloud Base Height and Phase from Thermal Infrared Radiometry Using a Deep Learning Algorithm. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4105914. [CrossRef]
18. Cui, L.L.; Guo, W.; Ge, W.Q.; Yan, Y.F.; Luo, S. Accuracy Verification of FY-4A Satellite Cloud Top Parameters and Research on Typhoon Application. *Plateau Meteorol.* **2020**, *39*, 196–203.
19. Jiang, Y.; Tang, B.-H.; Li, Z.-L.; Zhang, H. A New Cloud Base Height Retrieval Method and Its Application in Cloudy-Sky Surface Downwelling Longwave Radiation Estimation. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4505111. [CrossRef]
20. Lao, L.; Liu, Q.; Ding, Y.H.; Wang, Y. Rainrate Estimation from FY-4A Cloud Top Temperature for Mesoscale Convective Systems by Using Machine Learning Algorithm. *Remote Sens.* **2021**, *13*, 3273. [CrossRef]
21. Yan, W.; Yang, H.L.; Ye, J. Cloud Characteristics Analysis of Spaceborne Millimeter-Wave Cloud Radar Data. *Natl. Remote Sens. Bull.* **2009**, *13*, 575–579.
22. Chi, Y.; Zhao, C.; Yang, Y.; Zhao, X.; Yang, J. Global Characteristics of Cloud Macro-Physical Properties from Active Satellite Remote Sensing. *Atmos. Res.* **2024**, *302*, 107316. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.