scientific data

Check for updates

OPEN

DATA DESCRIPTOR

All-day cloud property and occurrence probability dataset based on satellite remote sensing data

Longfeng Nie[1,2,3], Yuntian Chen[3,4] ✉ & Dongxiao Zhang[1,2,3,4] ✉

The cloud property database and different type cloud occurrence probability datasets are helpful for meteorological research and application, as well as climate comprehension. Building on the foundation of CldNet with all-day cloud type recognition capability, CldNet Version 2.0 (CldNetV2) is proposed. This enhanced version leverages transfer learning and model parameter sharing techniques to not only classify cloud types but also predict additional cloud properties. Datasets with multiple cloud properties obtained by CldNetV2 make up for the lack of current Himawari cloud product at nighttime. Meanwhile, the dataset of different cloud type occurrence probabilities is statistically obtained on three time scales including annual, seasonal, and monthly, and more importantly, the dataset distinguishes between all day, daytime, and nighttime. In addition, the reliability of our cloud product is independently validated by the cloud properties from CALIPSO trajectories, ERA5 cloud cover fraction and the visualization of cloud property distribution and typhoon eye track during typhoons. Further more, the all-day cloud property and occurrence probability dataset for meteorological environment assessment has been publicly released.

Background & Summary

The occurrence of extreme weather has aroused people's attention to global climate change[1–5], especially extreme rainfall and drought[6–11]. Clouds play an extremely important role in global climate change[12–16], which can weaken or reinforce the thermodynamic effect regionally and modify the amplification of extreme precipitation and drought. Extreme rainfall can cause floods on the ground in a short period of time, and rapidly advancing floods lead to landslides in unprotected mountains, causing mudslides to damage nearby villages' houses, and even posing a threat to the safety of people's lives and property[17–19]. Urban pluvial flooding[20–23] is also a threatening natural disaster in urban region around the world, especially in recent years, when it has been occurring with increasing frequency. Not only do extreme rainfalls cause disasters, but droughts can also exacerbate food shortages and lead to armed conflict[24,25]. The distribution characteristics of clouds are of great help in evaluating changes in rainfall, typhoons, heatwaves, water resources, floods and droughts, food security and vegetation cover, as well as land use[26–33]. Therefore, the retrieval of cloud property is a prerequisite for in-depth exploration of cloud distribution characteristics.

Research on cloud property retrieval using satellite remote sensing data is advantageous due to the wide coverage, multispectral capability, and high spatiotemporal resolution of geostationary satellite sensors[34–36], providing real-time and extensive data for cloud property analysis. According to the processing of satellite spectral data, cloud property retrieval can be categorized into physics-based cloud property retrieval and machine learning-based cloud property retrieval. Traditionally, physics-based cloud property retrieval algorithms for satellite observation require the assistance of microscopic physical and optical models[37,38], and sometimes even the

[1]Pengcheng Laboratory, Shenzhen, 518000, P. R. China. [2]School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, 518055, P. R. China. [3]Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, 315200, P. R. China. [4]Zhejiang Key Laboratory of Industrial Intelligence and Digital Twin, Eastern Institute of Technology, Ningbo, 315200, P. R. China. ✉e-mail: ychen@eitech.edu.cn; dzhang@eitech.edu.cn

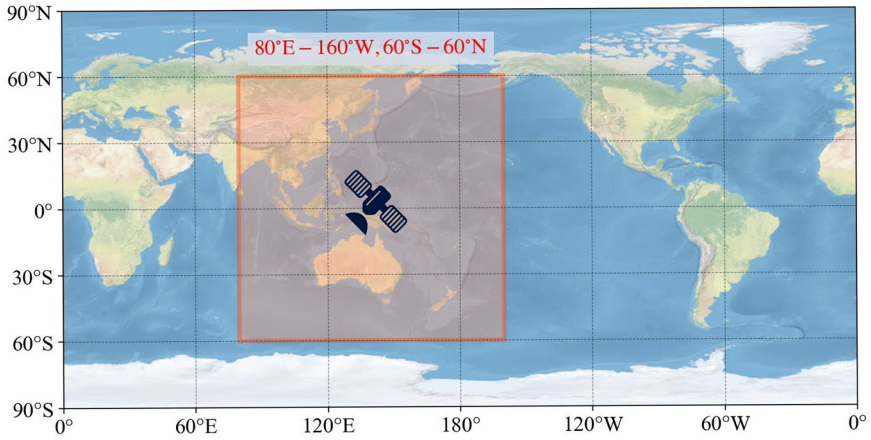


Fig. 1 The scope of the satellites Himawari-8/9.

setting of thresholds based on experience[39–43]. In recent years, machine learning-based cloud property retrieval algorithms have been developed to compensate for the shortcomings of physical models[44,45].

With the growth of computer hardware and massive satellite remote sensing data, deep learning-centered methods have become one of the hottest research directions[46–50]. In order to effectively utilize high-resolution geostationary observations of the Himawari-8/9 satellite, the physical algorithms[51,52] proposed by Ishida and Nakajima are applied to produce the Japan Aerospace Exploration Agency (JAXA) cloud property product. However, the Himawari cloud product is limited to daytime areas as physical algorithms involve the use of visible and near-infrared spectra[44,53]. Only providing cloud properties in the daytime area by JAXA is not conducive to the observation and in-depth study of cloud microphysical characteristics during nighttime. For example, during nighttime, meteorological departments can make warnings based on the cloud microphysical characteristics of the area where the landfalling typhoon is located to protect the safety of buildings and humans when the cloud properties of the nighttime area are provided. The Himawari cloud product provided by JAXA can only count the probability of cloud occurrence during the daytime, but not the probability of cloud occurrence during the nighttime and throughout the day. However, the probability of cloud occurrence is useful for some climatic or meteorological understanding, which should not be limited to daytime conditions, but should also focus on nighttime conditions. Therefore, it is important to realize all-day retrieval of cloud properties. If satellite spectral information is selectively utilized based on diurnal differences, all-day cloud properties can be integrated over the entire satellite observation area. Even if both day and night exist simultaneously in the entire area, the daytime area uses all remote sensing channels while the nighttime area only uses infrared remote sensing channels. In this study, a high-fidelity, all-day, spatiotemporal cloud property database for meteorological disaster assessments is published and the datasets of different type cloud occurrence probability for climate comprehension are produced. Therefore, researchers can not only obtain datasets of cloud properties and occurrence probabilities under daytime conditions for in-depth study of the meteorological environment, but also the corresponding datasets under nighttime conditions for a comprehensive assessment of changes in the meteorological environment.

Methods

Data. In this study, remote sensing data of the satellites Himawari-8/9 is used to predict cloud properties including cloud types (CTYPE), cloud optical thickness (COT), cloud top temperature (CTT), cloud top height (CTH), and cloud effective radius (CER). The scope of the satellites Himawari-8/9 is shown in Fig. 1, whose coverage area is from 80°E to 160°W, and from 60°N to 60°S. Satellite spectral channels include albedo from B01 to B06 and brightness temperature (BT) from B07 to B16, whose specific information is shown in Table 1[54–56]. The Himawari L1 gridded data is downloaded through the JAXA Himawari Monitor P-Tree System (available at https://www.eorc.jaxa.jp/ptree/index.html).

The JAXA Himawari Monitor P-Tree System provides Himawari L2 gridded data cloud properties (daytime only) with spatial resolution $0.05° \times 0.05°$[57–59]. The Himawari cloud product is mainly developed based on the physical algorithms[44,51,52] proposed by Ishida and Nakajima, and is released as the official product of JAXA. There are many evaluation studies on the performance of Himawari cloud product. For example, Lai *et al.*[58] compared cloud mask, cloud phase, COT, and CER of Himawari cloud product with the Moderate Resolution Imaging Spectroradiometer (MODIS) cloud product to evaluate the reliability of Himawari cloud product. Liu *et al.*[60] found that CTH, COT, and CER retrieved from the Himawari cloud product are consistent with passive Moderate Resolution Imaging Spectroradiometer (MODIS) cloud product; Huang *et al.*[61] evaluated CTH and CTT of the Himawari cloud product using shipborne radar-lidar data, and there was good agreement between them; The inter-comparison of CTH from Himawari-8, MODIS, and CALIPSO in Min *et al.*'s study[62] and comparisons of the CTH and COT retrieval from the JAXA Himawari Monitor, integrated cloud analysis system (ICAS), and CloudSat track in Wang *et al.*'s study[44] also indicate the reliability of Himawari cloud product. Therefore, it is feasible to use the Himawari cloud product as model reference labels in this study.

Channel	Central wavelength	Bandwidth	SNR or NEΔT @ specified input	Resolution	Primary application	Value range
B01	455 nm	50 nm	≥300 @ albedo (%)	1.0 km	Aerosol	[0, 100]
B02	510 nm	20 nm	≥300 @ albedo (%)	1.0 km	Aerosol	[0, 100]
B03	645 nm	30 nm	≥300 @ albedo (%)	0.5 km	Low cloud, Fog	[0, 100]
B04	860 nm	20 nm	≥300 @ albedo (%)	1.0 km	Vegetation, Aerosol	[0, 100]
B05	1610 nm	20 nm	≥300 @ albedo (%)	2.0 km	Cloud phase, SO2	[0, 100]
B06	2260 nm	20 nm	≥300 @ albedo (%)	2.0 km	Particle size	[0, 100]
B07	3.85 μm	0.22 μm	≤0.16 @ BT (K)	2.0 km	Low cloud, Fog, Forest fire	[220, 335]
B08	6.25 μm	0.37 μm	≤0.40 @ BT (K)	2.0 km	Upper level moisture	[200, 260]
B09	6.95 μm	0.12 μm	≤0.10 @ BT (K)	2.0 km	Mid-upper level moisture	[200, 270]
B10	7.35 μm	0.17 μm	≤0.32 @ BT (K)	2.0 km	Mid-level moisture	[200, 275]
B11	8.60 μm	0.32 μm	≤0.10 @ BT (K)	2.0 km	Cloud phase, SO2	[200, 320]
B12	9.63 μm	0.18 μm	≤0.10 @ BT (K)	2.0 km	Ozone content	[210, 295]
B13	10.45 μm	0.30 μm	≤0.10 @ BT (K)	2.0 km	Cloud imagery, Information of cloud top	[200, 330]
B14	11.20 μm	0.20 μm	≤0.10 @ BT (K)	2.0 km	Cloud imagery, Sea surface temperature	[200, 330]
B15	12.35 μm	0.30 μm	≤0.10 @ BT (K)	2.0 km	Cloud imagery, Sea surface temperature	[200, 320]
B16	13.30 μm	0.20 μm	≤0.30 @ BT (K)	2.0 km	Cloud top height	[200, 295]

Table 1. Specific information for each spectral channel of the satellites Himawari-8/9.

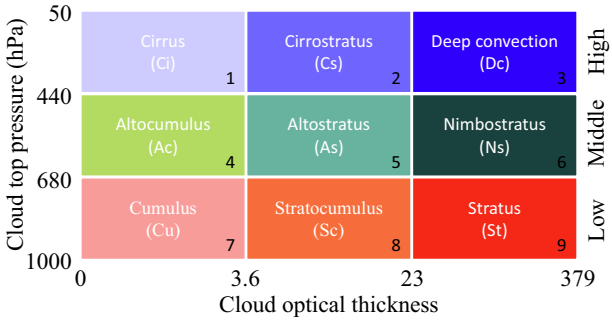


Fig. 2 The ISCCP definition of cloud types.

The cloud types of the Himawari cloud product are classified according to the International Satellite Cloud Climatology Project (ISCCP). The ISCCP definition of cloud types is shown in Fig. 2, which is detailed at https://isccp.giss.nasa.gov. Under cloudy skies, cloud types are classified as cirrus (Ci), cirrostratus (Cs), deep convection (Dc), altocumulus (Ac), altostratus (As), nimbostratus (Ns), cumulus (Cu), stratocumulus (Sc), and stratus (St), depending on cloud optical thickness and cloud top pressure[63,64]. Otherwise, the sky is clear (Cl). Among them, Ci, Cs, and Dc belong to high-level clouds, Ac, As and Ns belong to mid-level clouds, and Cu, Sc and St belong to low-level clouds.

In the work of Nie *et al.*[65], it is found that the proportion of Himawari cloud product (only daytime) covering the entire observation area peaks around 3:00 (UTC+0), and the cloud distributions do not change rapidly. Therefore, the Himawari L1 gridded data with spatial resolution 0.05° × 0.05° and the Himawari cloud product (daytime only) at 3:00 (UTC+0) every 3 days from January 1, 2020 to December 31, 2022 are downloaded for model training and evaluation. The data from 2020 and 2021 is used for the training process of the model, with 80% used to update model parameters and 20% used to validate the model to prevent overfitting by early stopping of the model training[66]. The data from 2022 is used to evaluate the performance of the model.

Model. Building upon the peer-reviewed and formally published CldNet model[65] with only 0.46 M total parameters, it was found that compared to other widely used semantic segmentation networks (e.g., SegNet, PSPNet, DeepLabV3+, UNet, and ResUnet), the accuracy of the CldNet in cloud type classification improved by 32%, 46%, 22%, 2%, and 39%, respectively. Meanwhile, CldNet has the ability to accurately detect clouds at daytime and nighttime, and can be directly generalized to higher resolution satellite spectral input data without any fine-tuning. In order to better expand the capabilities of the CldNet, CldNet Version 2.0 (CldNetV2) is proposed and developed to predict more cloud properties, as shown in Fig. 3. CldNetV2 mainly consists of KnowledgeModule, addition of auxiliary information, MaskModule, Block-<cloud property> and Head-<cloud property>.

The KnowledgeModule is used to obtain satellite spectral channels and their combinations. From previous research knowledge[67,68], it has been found that cloud properties are not only related to single spectral channel, but also to the differences between two different channels. Here, it should be noted that satellite spectral

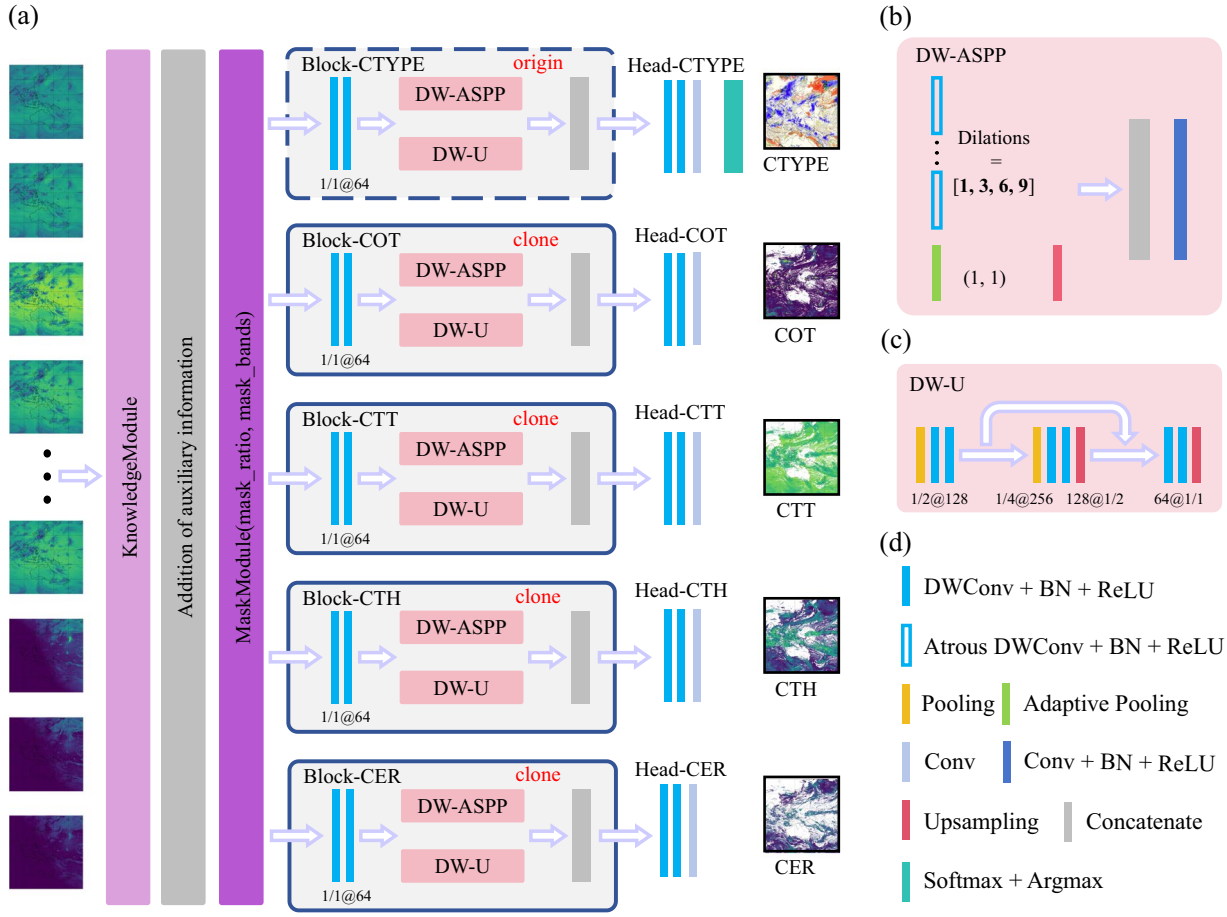


Fig. 3 The specific architecture of CldNet Version 2.0 (CldNetV2).

channels B01 ~ B06 are characterized by albedo, while satellite spectral channels B07 ~ B16 are characterized by BT. Since albedo and BT are distinct variables and their units are different, the pairwise combinations must have the same dimension in order to make a difference. The KnowledgeModule completed the reorganization of the satellite spectral information of B01 ~ B16, the differences between the pairwise combinations of B01 ~ B06, and the differences between the pairwise combinations of B07 ~ B16.

The auxiliary information including satellite zenith angle (SAZ), satellite azimuth angle (SAA), solar zenith angle (SOZ), and solar azimuth angle (SOA) is added as model inputs, improving the accuracy of the model. The spectral channels of the satellites Himawari-8/9 are divided into visible light (B01 ~ B03), near-infrared (B04 ~ B06), and infrared (B07 ~ B16). At daytime, the satellites Himawari-8/9 are able to receive visible and near-infrared spectral information by reflecting sunlight from the Earth's surface. At nighttime, the satellites Himawari-8/9 are unable to obtain visible and near-infrared spectral information in the absence of sunlight. In order to enable the model to accurately predict cloud properties at nighttime, the inputs to the model need to exclude the visible and near-infrared channels (B01 ~ B06). The MaskModule is used to determine whether to delete the satellite spectral information of B01 ~ B06 and the differences between the pairwise combinations of B01 ~ B06 through the parameters mask ratio and mask bands. The model using the satellite spectral channels B01 ~ B06 is labeled as CldNetV2-W, and vice versa, labeled as CldNetV2-O. It should be noted that CldNetV2-O only uses the infrared spectral channel information without considering the auxiliary information (SOZ, SOA, SAZ and SAA), mainly for the following reasons: (1) CldNetV2 is utilizing the Himawari cloud product, which is only available in daytime, as targets to achieve cloud property retrieval in both daytime and nighttime regions; (2) The range of SOZ and SOA during nighttime is different from that during daytime, and if SOZ and SOA are used, it may lead to the SOZ and SOA information inconsistency between the inference and training stages for CldNetV2-O; (3) The spatial distributions of SAZ and SAA remain constant over time, and not using SAZ and SAA can simplify CldNetV2-O's dependence on this information.

Block-<cloud property> mainly includes DW-ASPP (atrous spatial pyramid pooling (ASPP) using depth-wise (DW) separable convolution) and DW-U (U-shaped network structure using DW separable convolution). DW-ASPP can enhance the network structure in the receptive field of the feature maps, and DW-U can extract feature maps at different levels. The fusion of the two extracted feature maps is conducive to mining the internal relationship between model inputs and cloud properties. Head-<cloud property> is used to predict the corresponding cloud property. Furthermore, the network structure between CldNetV2-W and CldNetV2-O is different in the first layer due to the different number of input channels, the rest of the network structure is the same.

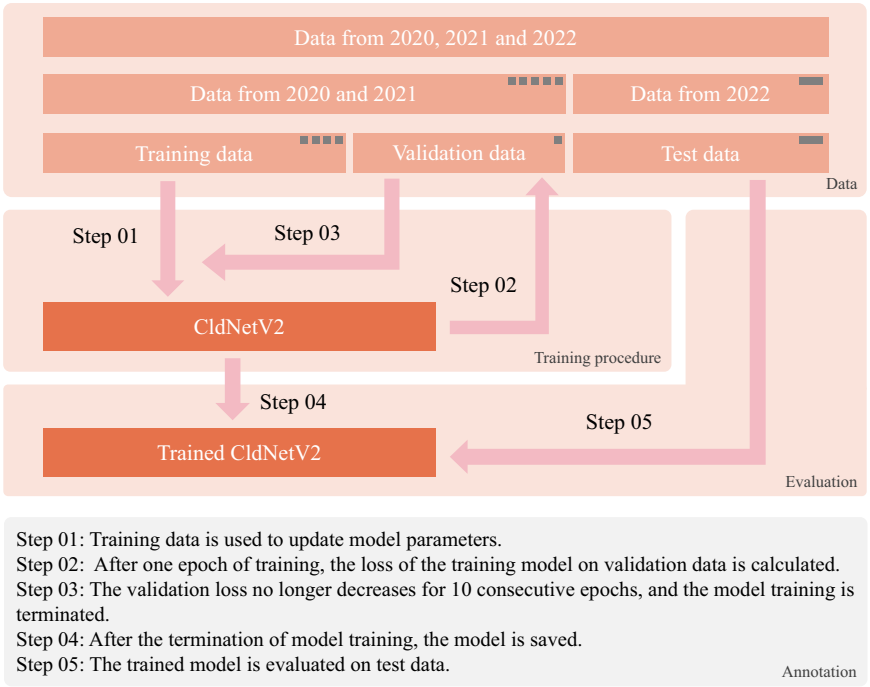


Fig. 4 The training and evaluation procedure of CldNetV2.

AI-based cloud retrieval has been studied in the past, for example, Yang *et al.*[69] developed a cloud property retrieval model for the Advanced Himawari Imager (AHI) onboard the Himawari-8 satellite based on the XGBoost machine learning algorithm, which utilized the AHI thermal data to achieve the retrieval of cloud mask, CTT, and CTH for both daytime and nighttime. In addition, Wang *et al.*[45] proposed a convolutional neural network with UNet-like structure based on deep learning with strong nonlinear expressive ability, and achieved cloud property retrieval including cloud mask, COT, CTH, and CER from thermal infrared radiometry. The difference is that our model CldNetV2 integrates the features extracted from both DW-ASPP and DW-U, and selectively inputs satellite spectral channels for daytime and nighttime, achieving all-day cloud property retrieval. Meanwhile, techniques such as parameter sharing and fine-tuning are also utilized to save computational time cost.

The training and evaluation procedure of CldNetV2 is depicted in Fig. 4, and each cloud property adopts the same training and evaluation procedure. Inspired by transfer learning, CldNetV2 has two training execution schemes. Scheme 01 is to train each cloud property using Block-<cloud property> and Head-<cloud property>, and the network parameters of each cloud property are parallel and not shared with each other. Scheme 02 is to train the network parameters of CTYPE to achieve loss convergence first, which include Block-CTYPE and Head-CTYPE. For COT, CTT, CTH and CER, during the training process, Block-COT, Block-CTT, Block-CTH and Block-CER directly share the trained Block-CTYPE parameters and freeze them. And only Head-COT, Head-CTT, Head-CTH, and Head-CER need to be trained to update the parameters for loss convergence.

Experimental setup. All models in this study are trained on single NVIDIA GPU A100. In each training batch, single remote sensing image is segmented into 5×5 slices, and its dimension is $25 \times 76 \times 480 \times 480$ through the KnowledgeModule. For CldNetV2-W, the dimension of the model input data is $25 \times 80 \times 480 \times 480$ after the addition of auxiliary information. For CldNetV2-O, after deleting satellite spectral channels B01 ~ B06 and the differences between their pairwise combinations, the dimension of the model input data is $25 \times 55 \times 480 \times 480$ without the addition of auxiliary information. The batch size is 25. Optimization algorithm Adam is used and its learning rate (lr) is controlled through a multi-step learning rate approach. If epoch < 5, lr $= 0.01$; if $5 \leq$ epoch < 10, lr $= 0.001$; if $10 \leq$ epoch < 20, lr $= 0.0001$; if $20 \leq$ epoch < 30, lr $= 0.00001$; and if epoch ≥ 30, lr $= 0.000001$.

For CTYPE, the loss function uses the cross entropy loss in Equation (1); For other cloud properties, the loss function adopts the mean squared loss in Equation (4).

$$CrossEntropyLoss = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{C} - Cld_{ref,ij}^{prob}\ln\left(Cld_{pre,ij}^{prob}\right) = -\frac{1}{N}\sum_{i}^{N}\ln\left(Cld_{pre,i[label]}^{prob}\right) \quad (1)$$

$$Cld_{pre,i}^{prob} = softmax(x_i) = \frac{e^{x_{ij}}}{\sum_{j=0}^{C} e^{x_{ij}}} \quad (2)$$

$$Cld_{pre,i}^{label} = argmax\left(Cld_{pre,i}^{prob}\right) \tag{3}$$

$$MSELoss = \frac{1}{N}\sum_{i=1}^{N}|Cld_{pre,i} - Cld_{ref,i}| \tag{4}$$

where N is the number of samples per batch; C is the number of CTYPE labels; Cld_{ref}^{prob} and $[label]$ is the probability and label of the reference CTYPE, respectively; Cld_{pre}^{prob} and Cld_{pre}^{label} are the probability and label of the model's predicted CTYPE, respectively; x is the output result of the network structure for CTYPE; and the function *argmax* returns the index of the maximum value of all elements in the input vector. $Cld_{ref,i}$ and $Cld_{pre,i}$ are the reference and predicted values of cloud properties including COT, CTT, CTH, and CER, respectively.

Fusion inference workflow. The trained CldNetV2 will be used to inference the distribution of cloud property over the entire satellite observation region based on Himawari L1 Gridded data. The inference workflow of all-day cloud property is depicted in Algorithm 1. According to the fusion inference process, we can obtain the corresponding cloud property distribution based on Himawari L1 Gridded data at any time. COT, CTT, CTH, and CER only exist when cloud exists (CTYPE $\neq 0$), and these values do not exist under clear sky (CTYPE $= 0$). Within the entire satellite observation area, grids with solar zenith angles greater than 65° will use the results of CldNet-O, and vice versa, the results of CldNet-W will be used.

Algorithm 1 Retrieval of all-day cloud property.

Data: Himawari L1 Gridded data including albedo of B01 ∼ B06, brightness temperature (BT) of B07 ∼ B16, satellite zenith angle (SAZ), satellite azimuth angle (SAA), solar zenith angle (SOZ), and solar azimuth angle (SOA)
Result: All-day cloud properties including cloud types (CTYPE), cloud optical thickness (COT), cloud top temperature (CTT), cloud top height (CTH), and cloud effective radius (CER)

```
RS: satellite spectral channels B01 ∼ B16;
RS-W ← KnowledgeModule(RS);
Cld-W ← CldNetV2-W(RS-W, SAZ, SAA, SOZ, SOA);        ▷ COT, CTT, CTH, and CER = NaN if CTYPE = 0
RS-O ← MaskModule(RS-W);
Cld-O ← CldNetV2-O(RS-O);                            ▷ COT, CTT, CTH, and CER = NaN if CTYPE = 0
for i=0; i<N; i++ do
    for j=0; j<M; j++ do
        if SOZ_ij > 65° then
            Cld_ij = Cld_ij-O
        else
            Cld_ij = Cld_ij-W
        end
    end
end
Cld is saved in NetCDF format.
```

Probability of cloud occurrence. The probability of cloud occurrence over a region has a significant impact on its climate, as clouds play an important role in regulating the energy exchange between the Earth and the atmospheric system. Let P_C represent the probability of cloud appearance in a specific area during a certain time period, defined as:

$$P_C = \frac{N_C}{N_T} \tag{5}$$

where N_C is the number of times clouds are observed during that specific time. Clouds can be nine types of clouds or a total cloud without distinguishing cloud types. N_T is the total number of observations during that time period. Obviously, the probability of clear sky is $1 - P_{total\ cloud}$.

Data Records

This database mainly includes cloud property (e.g. CTYPE, COT, CTT, CTH, and CER), which is saved in NetCDF format. The cloud property records in the NetCDF file are depicted in Table 2. Compared with Himawari cloud product, CldNetV2 cloud product can not only predict cloud property in daytime areas, but also in nighttime areas. Through the publicly available code in this study (see the section Code availability), all cloud property data can be directly generated by inputting corresponding remote sensing data from the satellites Himawari-8/9. Meanwhile, the code for generating all-day cloud property data has been fully open-source on the GitHub page https://github.com/metdai/PMD-CldNetV2 or figshare link https://doi.org/10.6084/m9.figshare.26893474.v3[70]. In order to further meet the direct usage needs of users, multiple figshare accounts have been applied for uploading more data. The CldNetV2 cloud product with a time resolution of 3 hours throughout 2022 will be released on the link https://doi.org/10.6084/m9.figshare.c.7417813.v2[71]. It takes up

Cloud property	Field	Type	Units	Valid Range	Missing	Factor	Offset
Cloud types	CLTYPE	int	—	[0, 9]	255	1.0	0.0
Cloud optical thickness	CLOT	float	—	[0.0, 150.0]	−9999	0.01	0.0
Cloud top temperature	CLTT	float	K	[0.0, 350.0]	−9999	0.01	150.0
Cloud top height	CLTH	float	km	[0.0, 20.0]	−9999	0.001	0.0
Cloud effective radius	CLER_23	float	μm	[0.0, 100.0]	−9999	0.01	0.0

Table 2. Cloud property records in the NetCDF file.

about 62.8 GB of storage space. At the same time, the video tutorial for using the CldNetV2 cloud retrieval system has been recorded so that users can use it immediately. In short, a highly effective all-day cloud property retrieval system capable of providing a real-time, all-day, spatiotemporal cloud property database for meteorological disaster assessments is established.

With the help of CldNetV2 cloud type product, the datasets of the probability of different cloud types occurring at different time scales are produced for climate comprehension. This study mainly calculates the probability of cloud occurrence on three time scales, including annual, seasonal, and monthly scales. The coverage area of this study includes both the Southern and Northern Hemispheres, where the climate seasons may differ in the same month. For the convenience of the following description, March to May is referred to as MAM (March, April and May), June to August is referred to as JJA (June, July and August), September to November is referred to as SON (September, October and November), and December to February of the following year is referred to as DJF (December, January and February). The details are as follows:

- **Annual Scale**: The probability of different types of cloud occurrence was calculated for each year between 2016 and 2022, covering a total of 7 years. The dimension of the resulting data is $7 \times 2401 \times 2401$.
- **Seasonal Scale**: The probability of different types of cloud occurrence for each season between September 2015 and November 2022 was calculated (covering 29 seasons from SON 2015 to SON 2022). The resulting data matrix has dimensions of $29 \times 2401 \times 2401$.
- **Monthly Scale**: The monthly probability of different types of cloud occurrence during the period from August 2015 to December 2022 was calculated, totaling 89 months of data. The resulting data matrix has dimensions of $89 \times 2401 \times 2401$.

In this study, the CldNetV2 cloud product with a time resolution of 3 hours from July 7, 2015 to December 31, 2022 is used to calculate occurrence probability of the nine types of clouds and the total cloud. One significant advantage of our cloud type product is that it contains not only daytime conditions, but also nighttime conditions. Therefore, the datasets of different type cloud occurrence probability at the annual, seasonal, and monthly scales for all day, daytime, and nighttime from 2015 to 2022 are produced and released on the link https://doi.org/10.6084/m9.figshare.27368562.v1[72], which are saved in NetCDF format. It takes up about 6.45 GB of storage space. These data help to provide insight into the distribution patterns of different types of clouds and their trends over different time scales. Here, the mean spatial distributions of the total cloud occurrence probability at the annual and seasonal scales are shown in Fig. 5.

Technical Validation

Quantitative evaluation metrics. The recognition of nine cloud types and clear sky is a multi classification problem. In order to better evaluate the predictive classification performance of the model, accuracy is a commonly used indicator. Accuracy can measure the ratio of correctly classified predictions to the total number, as follows:

$$\text{Accuracy} = \frac{\sum_{j=0}^{C} \text{TP}_j}{N} \tag{6}$$

In order to understand the model's ability to distinguish between clouds and clear skies, the indicator cloud detection (CD) accuracy is defined as follows:

$$\text{CD accuracy} = \frac{\sum_{k=0}^{1} \text{TP}_k}{N} \tag{7}$$

where TP is true positive; true means that the test result is correct; positive means that the intended target is detected. N is the pixel number of valid cloud property references. C is the number of CTYPE labels.

For COT, CTT, CTH, and CER, correlation coefficient (r), root mean square error (RMSE), and mean bias error (MBE) are selected as quantitative error indicators. Among them, r with the range of [-1, +1] is used to measure the correlation between the predicted and reference cloud property. For MBE, a negative value implies underestimation, and vice versa. In summary, the larger is the r, the better is the model; the smaller is the RMSE, the better is the model; and the closer the MBE is to zero, the better is the model. The best possible results are $r = 1$, $\text{RMSE} = 0$, and $\text{MBE} = 0$. These indicators are defined by the following Equations (8) - (10), respectively:

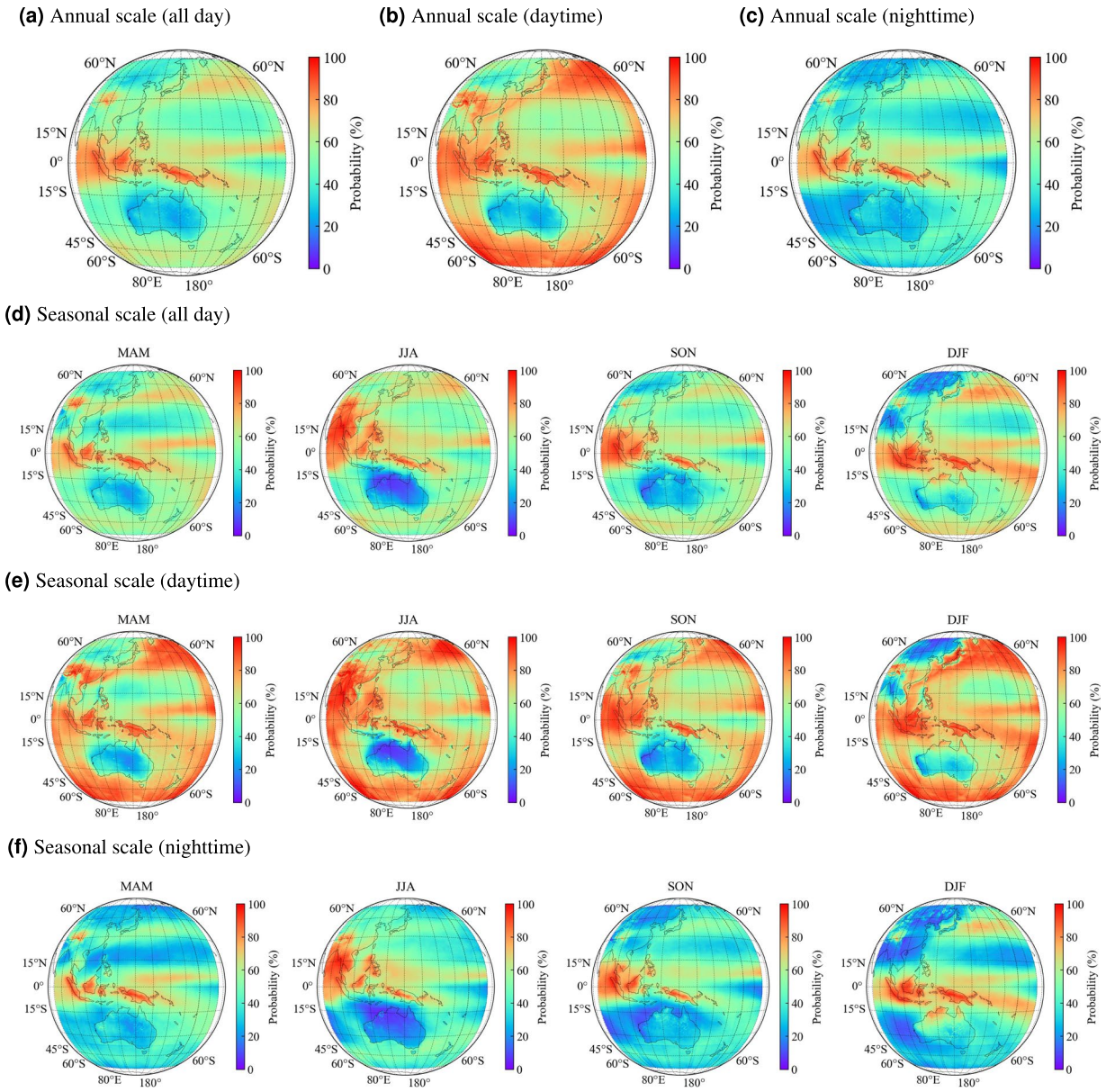


Fig. 5 The mean spatial distributions of total cloud occurrence probability at the annual scale during (**a**) all day, (**b**) daytime, and (**c**) nighttime and seasonal scale during (**d**) all day, (**e**) daytime, and (**f**) nighttime.

$$r = \frac{\sum_{i=1}^{N}(Cld_{pre,i} - Cld_{pre,-})(Cld_{ref,i} - Cld_{ref,-})}{\sqrt{\sum_{i=1}^{N}(Cld_{pre,i} - Cld_{pre,-})^2 \sum_{i=1}^{N}(Cld_{ref,i} - Cld_{ref,-})^2}} \quad (8)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(Cld_{pre,i} - Cld_{ref,i})^2}{N}} \quad (9)$$

$$\mathrm{MBE} = \frac{1}{N}\sum_{i=1}^{N}(Cld_{pre,i} - Cld_{ref,i}) \quad (10)$$

where $Cld_{ref,i}$ denotes the reference cloud property (eg. COT, CTT, CTH, and CER); $Cld_{pre,i}$ denotes the predicted cloud property using CldNetV2; $Cld_{ref,-}$ and $Cld_{pre,-}$ represent the mean of cloud property references and predictions, respectively; and N represents the pixel number of valid cloud property references.

To quantify the overall similarity between the distributions of two variables, cosine similarity is employed. Cosine similarity is a metric used to measure the degree of similarity between two vectors by calculating the

Model	Cloud property	Evaluating indicator	Scheme 01						Scheme 02					
			Mean	Std	Min	Median	Max	Full	Mean	Std	Min	Median	Max	Full
CldNetV2-W	CTYPE	Accuracy	0.82257	0.02141	0.77238	0.82417	0.86168	0.82279	—	—	—	—	—	—
CldNetV2-W	CTYPE	CD accuracy	0.94668	0.00452	0.92897	0.94753	0.95610	0.94653	—	—	—	—	—	—
CldNetV2-W	COT	r	0.98004	0.00464	0.96593	0.98111	0.98681	0.97970	0.96424	0.00767	0.93401	0.96657	0.97439	0.96363
CldNetV2-W	COT	RMSE	4.47263	0.58112	3.50131	4.39155	6.34548	4.50577	6.05660	0.74405	4.79284	5.96512	8.55399	6.08441
CldNetV2-W	COT	MBE	−0.14683	0.32618	−0.62445	−0.30076	0.57557	−0.14776	−0.11938	0.35401	−1.10681	−0.12085	0.60113	−0.11473
CldNetV2-W	CTT (K)	r	0.97048	0.00398	0.96069	0.97095	0.97847	0.97023	0.96253	0.00352	0.95450	0.96292	0.97031	0.96156
CldNetV2-W	CTT (K)	RMSE	5.56587	0.46327	4.69490	5.46942	7.03898	5.57062	6.32266	0.44516	5.60269	6.20909	7.94092	6.32621
CldNetV2-W	CTT (K)	MBE	−0.10641	0.74960	−2.35982	−0.03641	1.01084	−0.09085	0.17520	1.04061	−2.08653	0.10358	3.20033	0.16743
CldNetV2-W	CTH (km)	r	0.97460	0.00360	0.96276	0.97553	0.98063	0.97442	0.97191	0.00345	0.96347	0.97262	0.97817	0.97094
CldNetV2-W	CTH (km)	RMSE	0.88160	0.05403	0.78756	0.87069	1.04546	0.88223	0.94185	0.10108	0.83898	0.91215	1.36346	0.94418
CldNetV2-W	CTH (km)	MBE	−0.03168	0.10489	−0.24927	−0.02969	0.22076	−0.03179	−0.04468	0.16908	−0.63070	−0.01441	0.19907	−0.04188
CldNetV2-W	CER (μm)	r	0.91524	0.00730	0.89920	0.91451	0.93576	0.91494	0.73065	0.01754	0.68400	0.73123	0.77323	0.72840
CldNetV2-W	CER (μm)	RMSE	5.98929	0.30707	5.32775	5.97558	6.73076	5.99263	10.17832	0.50075	9.18853	10.12500	11.56481	10.18132
CldNetV2-W	CER (μm)	MBE	−0.21392	0.28507	−1.19489	−0.13232	0.19308	−0.20907	0.19838	0.79778	−1.07042	0.12129	2.81828	0.17783
CldNetV2-O	CTYPE	Accuracy	0.72895	0.01420	0.68464	0.72928	0.76092	0.72953	—	—	—	—	—	—
CldNetV2-O	CTYPE	CD accuracy	0.93948	0.00579	0.91181	0.94073	0.94935	0.93931	—	—	—	—	—	—
CldNetV2-O	COT	r	0.81126	0.01305	0.76575	0.81253	0.83741	0.80969	0.81464	0.00863	0.78567	0.81533	0.83540	0.81416
CldNetV2-O	COT	RMSE	12.50533	0.71097	10.61644	12.48643	14.32756	12.52638	12.42443	0.67258	10.51528	12.44290	14.34994	12.45202
CldNetV2-O	COT	MBE	1.45261	0.54825	0.13490	1.37242	3.53292	1.44787	1.22896	0.40892	0.23739	1.24914	2.02239	1.23646
CldNetV2-O	CTT (K)	r	0.95086	0.00503	0.93804	0.95142	0.96239	0.95098	0.95307	0.00454	0.93839	0.95358	0.96307	0.95257
CldNetV2-O	CTT (K)	RMSE	7.21467	0.37888	6.39429	7.16909	8.61899	7.21817	7.07831	0.40073	6.20996	7.02284	8.71667	7.08298
CldNetV2-O	CTT (K)	MBE	0.43064	0.50417	−0.57040	0.37411	1.84820	0.42189	0.15558	0.91773	−1.29182	−0.14142	2.96052	0.14164
CldNetV2-O	CTH (km)	r	0.96136	0.00364	0.95201	0.96169	0.96993	0.96094	0.96286	0.00320	0.95604	0.96311	0.97005	0.96281
CldNetV2-O	CTH (km)	RMSE	1.10796	0.05879	0.99806	1.09521	1.31187	1.10878	1.08501	0.04515	0.99409	1.08088	1.24012	1.08545
CldNetV2-O	CTH (km)	MBE	−0.00930	0.14001	−0.43880	0.00962	0.18831	−0.00726	−0.01135	0.09282	−0.26584	−0.00230	0.14415	−0.01011
CldNetV2-O	CER (μm)	r	0.79936	0.01492	0.76403	0.79719	0.84090	0.79548	0.76032	0.01838	0.70692	0.75983	0.80859	0.76004
CldNetV2-O	CER (μm)	RMSE	9.04995	0.49455	8.17379	9.04801	10.68470	9.05689	9.69718	0.47316	8.85392	9.64184	10.84549	9.70569
CldNetV2-O	CER (μm)	MBE	−0.58917	1.30131	−4.53020	−0.34924	1.67985	−0.57309	−0.56453	0.63392	−2.68090	−0.41668	0.51275	−0.56554

Table 3. Evaluation indicators of CldNetV2 using scheme 01 and scheme 02 for cloud properties including CTYPE, COT, CTH, CTT and CER on the test dataset.

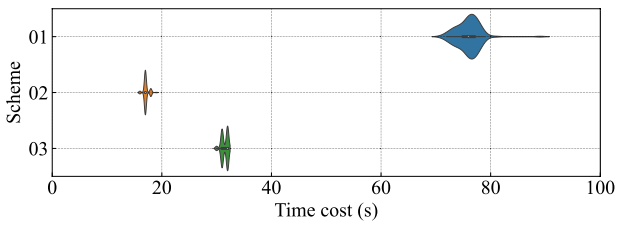


Fig. 6 The time cost of CldNetV2 in inferring cloud properties across the entire observation area.

cosine of the angle between them. Its value ranges from [−1, 1]. A value of 0 indicates no similarity; A negative value signifies a negative correlation; And a positive value denotes a positive correlation. The closer the absolute value of cosine similarity is to 1, the higher the similarity between the two vectors. The mathematical expression for cosine similarity, denoted as COSINE, is given by Equation (11):

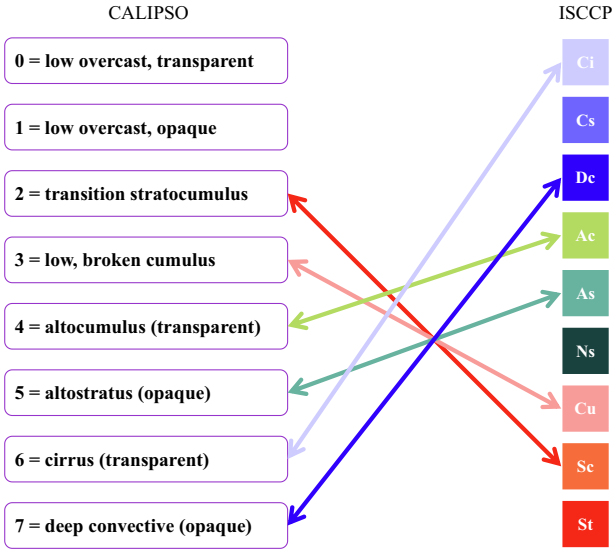


Fig. 7 The mapping relationship between CALIPSO and ISCCP cloud types.

$$\text{COSINE}\,(\mathbf{X},\mathbf{Y}) = \frac{\mathbf{X}\cdot\mathbf{Y}}{\|\mathbf{X}\|\|\mathbf{Y}\|} = \frac{\sum_{i=1}^{N}(\mathrm{X}_i \times \mathrm{Y}_i)}{\sqrt{\sum_{i=1}^{N}\mathrm{X}_i^2} \times \sqrt{\sum_{i=1}^{N}\mathrm{Y}_i^2}} \tag{11}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are the vectors whose similarity is being measured; $\mathbf{X}\cdot\mathbf{Y}$ represents the dot product of $\mathbf{X}$ and $\mathbf{Y}$; $\|\mathbf{X}\|$ and $\|\mathbf{Y}\|$ are the magnitudes of vectors $\mathbf{X}$ and $\mathbf{Y}$, respectively.

Evaluation indicator statistics. From January 1, 2020 to December 31, 2022, there are a total of 363 Himawari cloud property files (multispectral remote sensing files) at 3:00 (UTC+0) every 3 days. Among them, 119 Himawari cloud property files (multispectral remote sensing files) will be used as test data in 2022. Therefore, based on the 119 Himawari multispectral remote sensing files, 119 CldNetV2 cloud property files can be obtained using CldNetV2. The evaluation indicators for each of the 119 cloud property files were calculated, and then their mean, standard deviation, maximum, minimum, and median were calculated. At the same time, the 119 cloud property files are used once to calculate the evaluation indicators, labeled with Full. For each cloud property, the evaluation indicators of CldNetV2 are summarized in Table 3.

For the identification of nine types of clouds and clear skies, CldNetV2-W and CldNetV2-O achieve accuracy of 0.823 and 0.729, respectively. More importantly, CldNetV2-W and CldNetV2-O have strong ability to distinguish between cloud cover and clear sky, with accuracies of 0.947 and 0.939, respectively. For cloud properties including COT, CTT and CTH, the indicator r of CldNetV2-W is above 0.96, regardless of whether scheme 01 or scheme 02 is used. For CER, the indicator r of CldNetV2-W adopting scheme 01 is approximately 0.915, while using scheme 02 is approximately 0.731. The indicator r of CldNetV2-O is above 0.81 for COT and is above 0.95 for CTT and CTH, respectively, regardless of whether scheme 01 or scheme 02 is used. For CER, CldNetV2-O using scheme 01 and scheme 02 achieves r of 0.799 and 0.760, respectively. In aggregate, CldNetV2 has achieved good performance in cloud property retrieval.

Inference time. From the result of Table 3, it is found that for COT, CTT, and CTH, the performance of the model using scheme 01 and using scheme 02 are almost equivalent. However, the performance of the model trained using scheme 02 for CER is much lower than that of scheme 01. More importantly, it should be noted that in scheme 02, all cloud properties COT, CTT, CTH, and CER share the calculation results of Block-CTYPE, reducing the additional calculations for Block-COT, Block-CTT, Block-CTH, and Block-CER in scheme 01. Scheme 02 has certain advantages in terms of time consumption for inferring cloud properties. In order to leverage the advantages of both scheme 01 and scheme 02, scheme 03 is proposed. Scheme 03 adopts scheme 02 for inference of cloud properties including COT, CTT and CTH, and scheme 01 for cloud property CER.

In Fig. 6, the time cost of CldNetV2 in inferring cloud properties across the entire observation area is compared using different schemes. Fusion inference time cost of different schemes on single NVIDIA GPU 3090 is recorded. In the case of adopting scheme 01, the consumption time of single cloud property file is 75.96 ± 2.20 s; in the case of adopting scheme 02, the consumption time of single cloud property file is 17.11 ± 0.46 s; in the case of adopting scheme 03, the consumption time of single cloud property file is 31.46 ± 0.59 s. In practice, scheme 03 is adopted to build real-time retrieval of all-day cloud property.

Comparison of daytime and nighttime performance. In order to further understand the performance of the model CldNetV2, Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO product download link: https://subset.larc.nasa.gov/calipso/login.php) is used to verify the model. In this study, CALIPSO

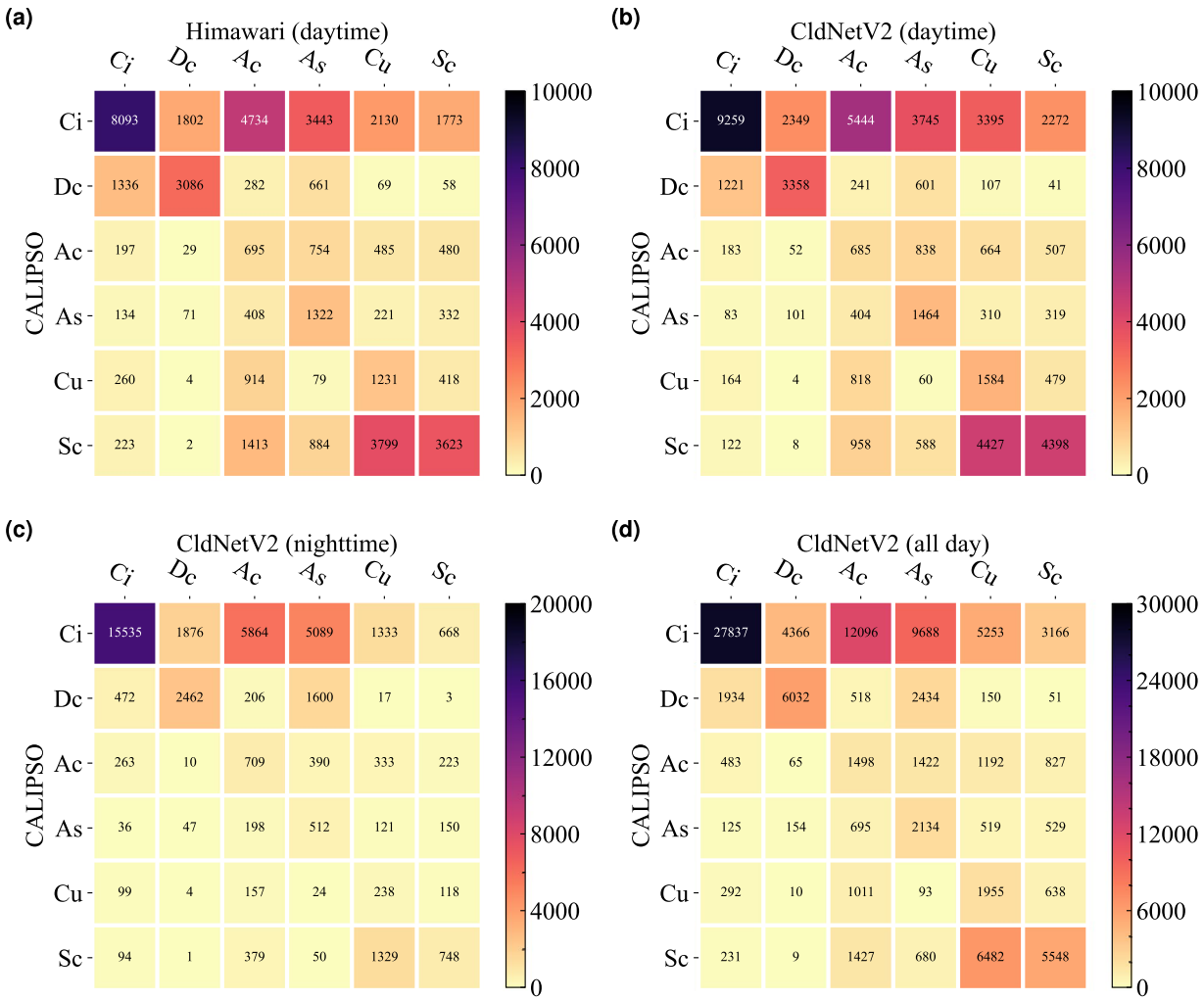


Fig. 8 The performance of Himawari cloud product at daytime and CldNetV2 cloud product at daytime, nighttime and all day (diurnal cycle) compared with the cloud type from CALIPSO trajectories.

Lidar Level-2 5 km Vertical Feature Mask (VFM) Version 4.51 Product is adopted to explore the performance of CTYPE from Himawari cloud product (daytime only) and CldNetV2 cloud product (all day). The CTYPE of the highest cloud in each profile observed by CALIPSO VFM V4.51 product is selected as CALIPSO CTYPE. Meanwhile, CALIPSO Lidar Level-2 5 km Cloud Layer (CLay) Version 4.51 Product is adopted to explore the performance of COT, CTH, and CTT from Himawari cloud product (daytime only) and CldNetV2 cloud product (all day). The CTH of the highest cloud in each profile observed by CALIPSO CLay V4.51 product is selected as CALIPSO CTH, and its corresponding CTT is selected as CALIPSO CTT. In this investigation, the cloud properties on the satellite CALIPSO trajectories need to be spatiotemporal matched with Himawari cloud product (daytime only) and CldNetV2 cloud product (all day). Here, the matching algorithm[62] calculates the minimum distance between CALIPSO trajectories and Himawari/CldNetV2 pixel to determine the specific matching point within a time distance of about ± 20 min.

The performance of Himawari cloud product at daytime and CldNetV2 cloud product at daytime, nighttime and all day compared with the cloud properties including CTYPE, COT, CTH, and CTT from CALIPSO trajectories for January through October 1 and December 9, 2022 is evaluated. The cloud detection accuracy of Himawari cloud product has reached 81.27%, while the cloud detection accuracy of CldNetV2 cloud product has reached 82.65%. This implies that CldNetV2 can achieve the same or even slightly better performance as Himawari in cloud mask recognition. The correct classification of cloud types is an important indicator for verifying the model's capabilities. To some extent, there are differences between the cloud type classification standards of Himawari/CldNetV2 and CALIPSO, but in order to better analyze, the cloud type relationship between them is mapped, as shown in Fig. 7. In Fig. 8, a statistical comparison between the Himawari cloud type and the CldNetV2 cloud type is presented using the CALIPSO cloud type as a reference. Overall, the performance of cloud type recognition between Himawari and CldNetV2 is similar for daytime. Additionally, CldNetV2 at nighttime and all day exhibits daytime-like characteristics. However, it has been observed that the cloud types of CALIPSO are recognized as other types by Himawari and CldNetV2. This phenomenon may be due to

Date	Cloud property	Himawari daytime		CldNetV2 daytime		CldNetV2 nighttime		CldNetV2 all day	
		RMSE	MBE	RMSE	MBE	RMSE	MBE	RMSE	MBE
2022_01_01	COT	3.66	1.22	3.90	1.60	4.01	0.98	3.98	1.42
	CTH (km)	3.59	−1.72	3.62	−1.57	3.82	−2.87	3.67	−2.06
	CTT (K)	37.67	24.62	36.45	23.23	32.82	21.49	34.55	22.32
2022_02_01	COT	4.09	1.76	4.08	1.94	4.03	0.90	4.17	1.65
	CTH (km)	3.20	−1.59	2.96	−1.30	3.71	−2.67	3.24	−1.79
	CTT (K)	31.11	18.52	30.34	18.01	29.95	19.21	30.28	18.60
2022_03_01	COT	3.84	1.35	3.88	1.65	3.93	1.09	3.98	1.54
	CTH (km)	3.30	−1.40	3.15	−1.16	3.86	−2.75	3.40	−1.73
	CTT (K)	32.63	18.16	32.36	18.18	31.92	23.04	32.12	20.33
2022_04_01	COT	3.81	1.18	3.98	1.52	4.12	1.04	4.11	1.40
	CTH (km)	3.82	−2.22	3.92	−2.41	4.64	−3.74	4.25	−2.98
	CTT (K)	29.44	19.15	27.60	17.13	27.29	18.06	27.69	17.79
2022_05_01	COT	3.79	1.53	3.63	1.48	3.91	1.16	3.87	1.44
	CTH (km)	4.16	−2.43	4.08	−2.53	4.11	−3.21	4.07	−2.76
	CTT (K)	38.10	26.42	36.80	26.00	30.36	21.87	33.62	23.47
2022_06_01	COT	3.68	1.43	3.75	1.66	3.89	0.69	3.95	1.35
	CTH (km)	4.09	−2.13	4.20	−2.34	4.29	−3.47	4.23	−2.86
	CTT (K)	34.89	22.26	33.15	21.02	28.18	19.66	30.23	19.78
2022_07_01	COT	3.78	1.46	3.78	1.60	4.11	1.42	4.02	1.61
	CTH (km)	4.00	−1.96	4.04	−2.21	4.17	−3.10	4.05	−2.52
	CTT (K)	35.70	22.82	34.58	22.56	35.82	25.81	34.87	23.83
2022_08_01	COT	3.90	1.33	4.17	1.86	4.02	1.25	4.26	1.75
	CTH (km)	3.83	−2.04	3.82	−2.17	4.05	−3.03	3.81	−2.33
	CTT (K)	31.60	20.45	30.78	20.06	29.93	20.42	29.97	19.56
2022_09_01	COT	3.83	1.34	4.06	1.75	3.88	1.35	4.05	1.65
	CTH (km)	3.67	−2.04	3.70	−2.19	4.25	−3.20	3.94	−2.58
	CTT (K)	35.19	23.22	33.39	21.78	31.18	21.93	32.24	21.75
2022_10_01	COT	3.76	1.10	4.15	1.69	3.93	1.04	4.10	1.52
	CTH (km)	3.44	−1.82	3.51	−1.87	4.22	−3.48	3.75	−2.34
	CTT (K)	30.89	19.09	30.33	19.01	30.10	23.10	31.42	21.44
2022_12_09	COT	3.79	1.10	4.08	1.64	4.14	1.20	4.15	1.54
	CTH (km)	3.80	−1.93	3.86	−1.89	4.06	−3.08	3.88	−2.26
	CTT (K)	35.12	22.57	34.03	21.50	31.14	21.76	32.98	21.87
Full	COT	3.82	1.34	3.97	1.68	4.00	1.10	4.06	1.54
	CTH (km)	3.70	−1.91	3.70	−1.93	4.11	−3.14	3.84	−2.37
	CTT (K)	33.83	21.42	32.73	20.66	30.89	21.47	31.85	20.94

Table 4. The performance of Himawari cloud product at daytime and CldNetV2 cloud product at daytime, nighttime and all day (diurnal cycle) compared with the cloud properties including COT, CTH, and CTT from CALIPSO trajectories.

differences in the criteria for classifying cloud types, as well as the fact that some cloud types are similar in nature and challenging to strictly distinguish.

The evaluation indicators RMSE and MBE of Himawari cloud product at daytime and CldNetV2 cloud product at daytime, nighttime and all day (diurnal cycle) compared with the cloud properties including COT, CTH, and CTT from CALIPSO trajectories are calculated in Table 4, whose Full data is shown in Fig. 9. For COT, Himawari cloud product achieves RMSE = 3.82 and MBE = 1.34 at daytime; CldNetV2 cloud product achieves RMSE = 3.97 and MBE = 1.68 at daytime; CldNetV2 cloud product achieves RMSE = 4.00 and MBE = 1.10 at nighttime; and CldNetV2 cloud product achieves RMSE = 4.06 and MBE = 1.54 at all day. For CTH, Himawari cloud product achieves RMSE = 3.70 km and MBE = -1.91 km at daytime; CldNetV2 cloud product achieves RMSE = 3.70 km and MBE = -1.93 km at daytime; CldNetV2 cloud product achieves RMSE = 4.11 km and MBE = -3.14 km at nighttime; and CldNetV2 cloud product achieves RMSE = 3.84 km and MBE = -2.37 km at all day. For CTT, Himawari cloud product achieves RMSE = 33.83 K and MBE = 21.42 K at daytime; CldNetV2 cloud product achieves RMSE = 32.73 K and MBE = 20.66 K at daytime; CldNetV2 cloud product achieves RMSE = 30.89 K and MBE = 21.47 K at nighttime; and CldNetV2 cloud product achieves RMSE = 31.85 K and MBE = 20.94 K at all day. Upon analysis, it is observed that while Himawari exhibits a slight edge in performance over CldNetV2 during the daytime, the precision levels are comparable, indicating a high degree of consistency in practical applications. Crucially, CldNetV2 demonstrates nighttime performance that is essentially on par with its daytime results, affirming its capability in nighttime scenarios.

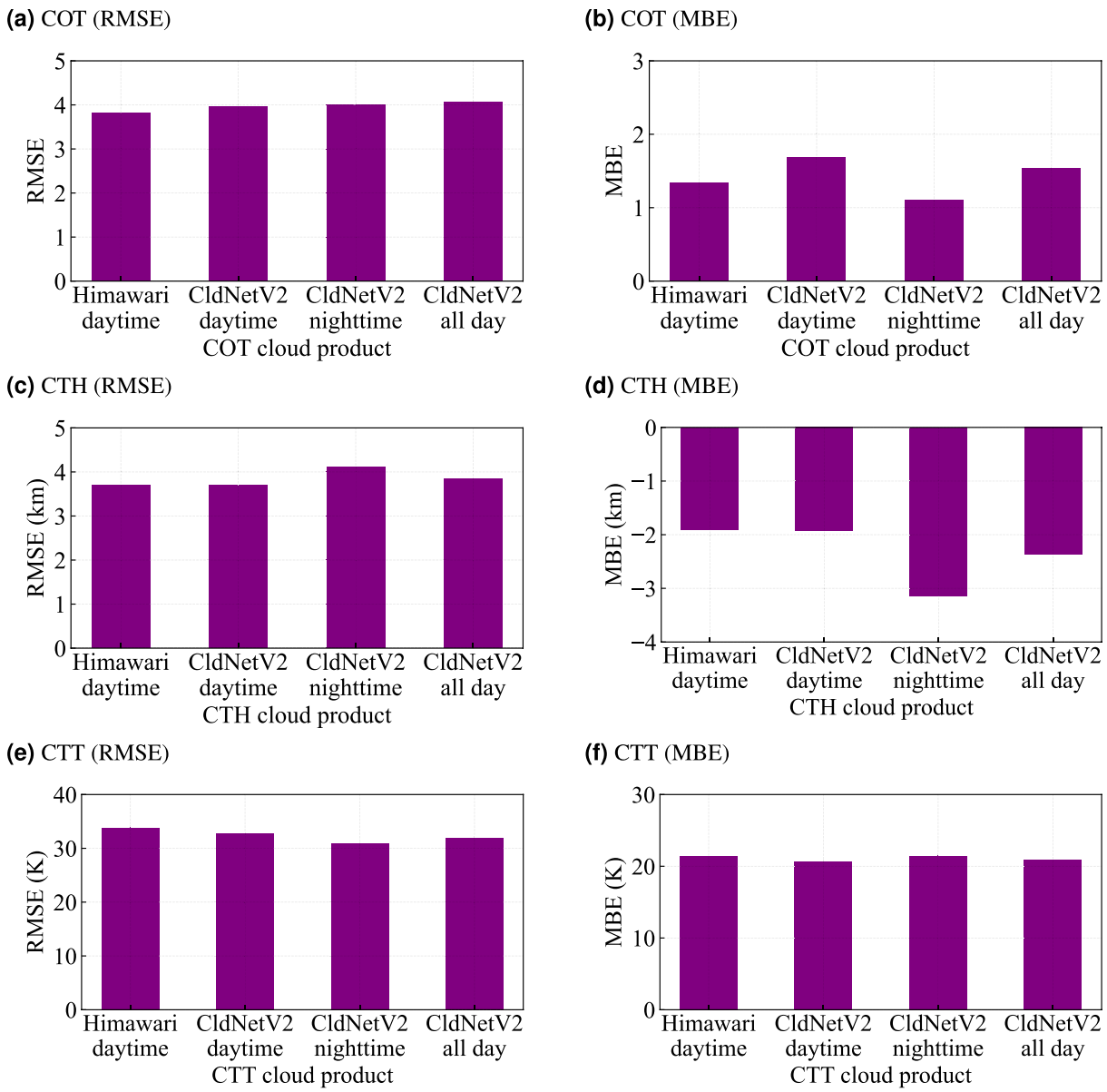


Fig. 9 The performance of Himawari cloud product at daytime and CldNetV2 cloud product at daytime, nighttime and all day (diurnal cycle) compared with the cloud properties including COT, CTH, and CTT from CALIPSO trajectories.

Similarly, the performance of the diurnal cycle (all day) is consistent with that of daytime and nighttime. This not only validates the model's robustness across different lighting conditions but also highlights its unique advantage over Himawari by providing reliable nighttime cloud property retrieval — a feature that Himawari lacks. In essence, CldNetV2 enriches the dataset by extending its coverage to the night, effectively bridging a gap in the existing data landscape.

Comparison with ERA5 cloud cover. In order to further independently validate CldNetV2 cloud product, the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) data is adopted. ERA5 is the fifth generation ECMWF reanalysis for the global climate and weather for the past 8 decades, which is available from 1940 onwards. ERA5 replaces the ERA-Interim reanalysis which stopped being produced on 31 August 2019. Here, cloud cover of ERA5 monthly averaged data on single levels from 1940 to present is used, with the data for 2022 downloaded from the website https://cds.climate.copernicus.eu. Meanwhile, the CldNetV2 cloud product during 2022 with a time resolution of 3 hours is used. The cloud cover fraction data, including high, middle, low, and total clouds, is used to verify the cloud types of CldNetV2 by grouping them by height. The cloud cover fraction refers to the proportion of the area covered by clouds within a specific region relative to the total area of that region, usually expressed as a percentage. The cloud occurrence probability indicates the likelihood of clouds appearing at a specific time and place, which is derived from statistical analyses based on historical data. Therefore,

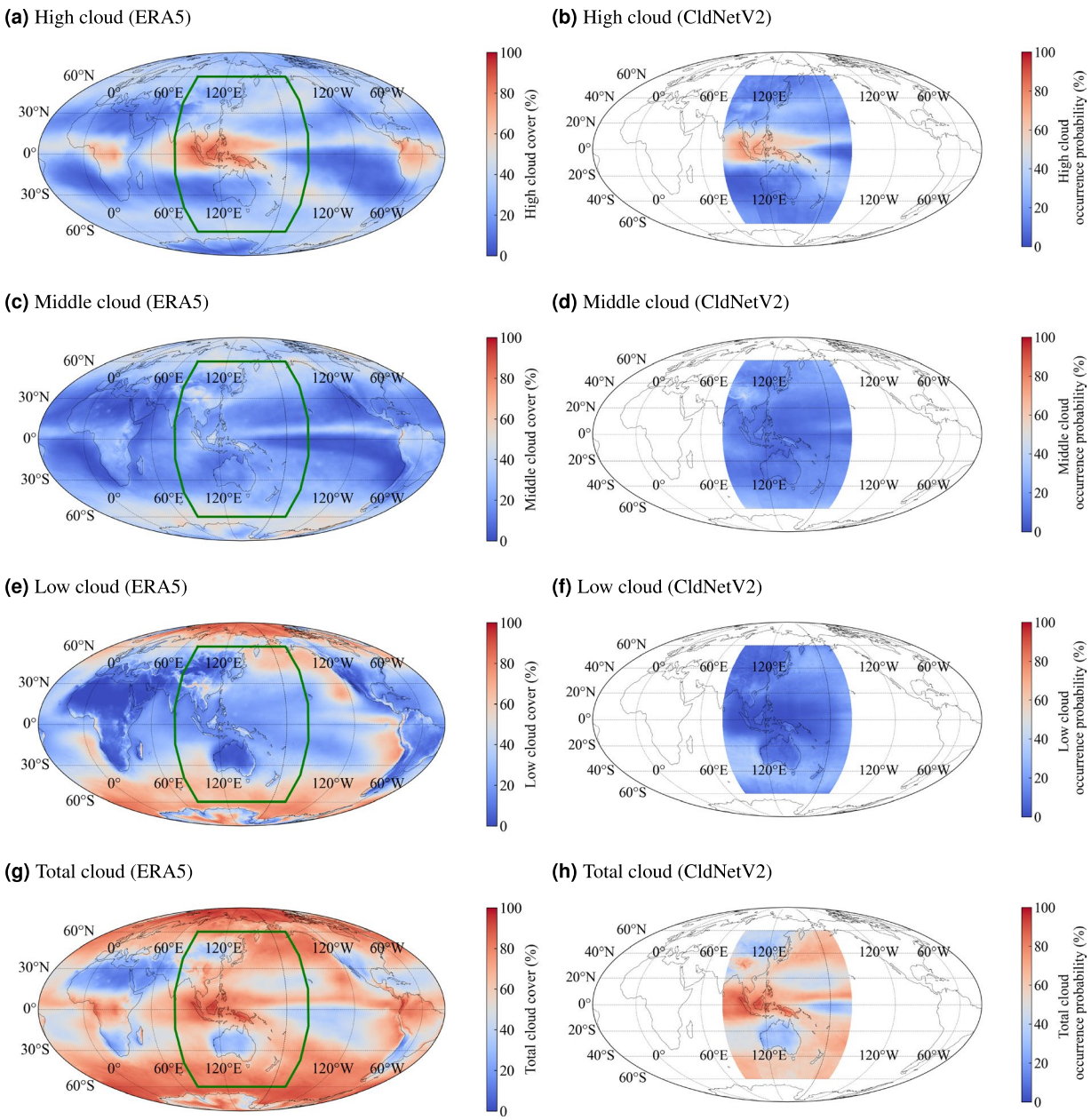


Fig. 10 The spatial distribution of ERA5 cloud cover fraction including (**a**) high cloud, (**c**) middle cloud, (**e**) low cloud, and (**g**) total cloud and CldNetV2 cloud occurrence probability including (**b**) high cloud, (**d**) middle cloud, (**f**) low cloud, and (**h**) total cloud during 2022.

the cloud cover fraction and occurrence probability are not numerically equivalent, but the comparison between the two provides an additional perspective to reflect the accuracy of our cloud product.

The spatial distribution of ERA5 cloud cover fraction and CldNetV2 cloud occurrence probability including high, middle, low, and total clouds during 2022 are shown in Fig. 10. ERA5 provides globally distributed data in Fig. 10a,c,e and g, while CldNetV2 covers data from 80°E to 160°W and from 60°N to 60°S in Fig. 10b,d,f and h. From Fig. 10, it can be observed that high cloud has the highest probability of occurrence among the three height clouds in ERA5 and CldNetV2, which mainly occurs between 15°N and 15°S. The region is located in a tropical climate zone, where high temperatures cause a large amount of seawater to evaporate, forming clouds and generating strong updrafts. In addition, the effect of the Walker circulation makes the region more prone to high cloud distribution. In the mid to high latitude ocean regions, the coverage of low clouds is significantly higher, mainly due to the evaporation of seawater forming clouds, but the evaporation induced updrafts in temperate zone are significantly weaker than those in tropical zone. In Fig. 10g and h, ERA5 total cloud cover fraction and CldNetV2 total cloud occurrence probability have similar distribution characteristics. The cosine similarity between ERA5 cloud cover fraction and CldNetV2 cloud occurrence probability is calculated, where the cosine similarity is

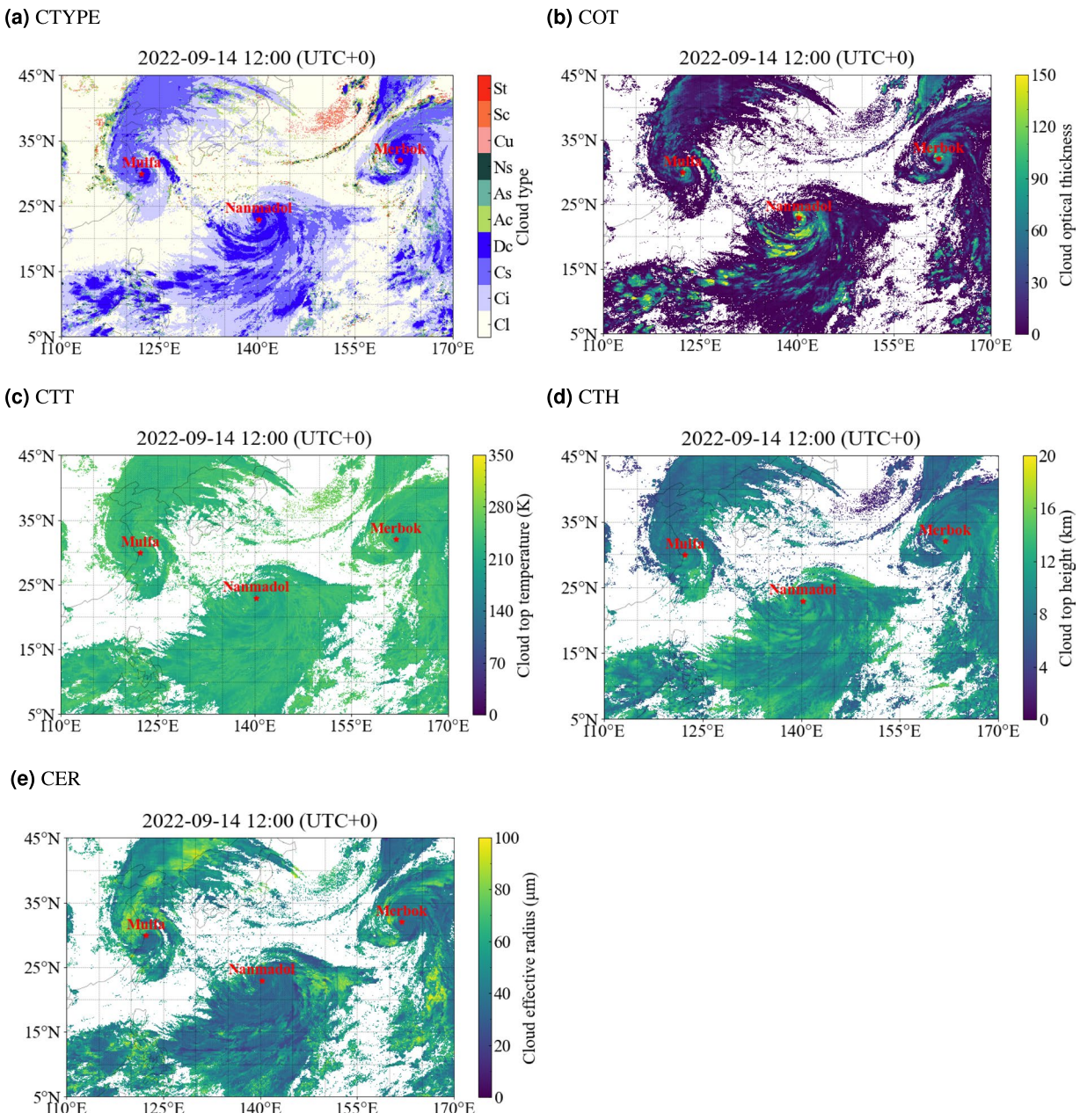


Fig. 11 The overall distribution of cloud properties including (**a**) cloud types, (**b**) cloud optical thickness, (**c**) cloud top temperature, (**d**) cloud top height, and (**e**) cloud effective radius when typhoons Muifa, Nanmadol, and Merbok in 2022 occur simultaneously.

0.964, 0.972, 0.938, and 0.990 for high, middle, low, and total clouds, respectively. This indicates that the performance of CldNetV2 cloud product in terms of reliability and availability is worthy of recognition. Meanwhile, there is a significant discrepancy observed between the cloud cover fraction and occurrence probability for low clouds, mainly because the two are not numerically equivalent. Nevertheless, the similarity between the two spatial distributions measured by the cosine similarity indirectly verifies our reliability as an effective choice. More importantly, the datasets of different type cloud occurrence probability at the annual, seasonal, and monthly scales for all day, daytime, and nighttime from 2015 to 2022 are statistically produced and released.

Cloud properties of typhoons. The Tropical Cyclone Data Center of the China Meteorological Administration (CMA) provides the best typhoon track dataset, which includes the location and intensity of tropical cyclones in the northwest Pacific Ocean (including the South China Sea, north of the equator and west of 180°E) since 1949. The best typhoon track dataset[73,74] is shared on the China Typhoon Network (https://tcdata.typhoon.org.cn). The cloud property distribution characteristics of typhoons Muifa,

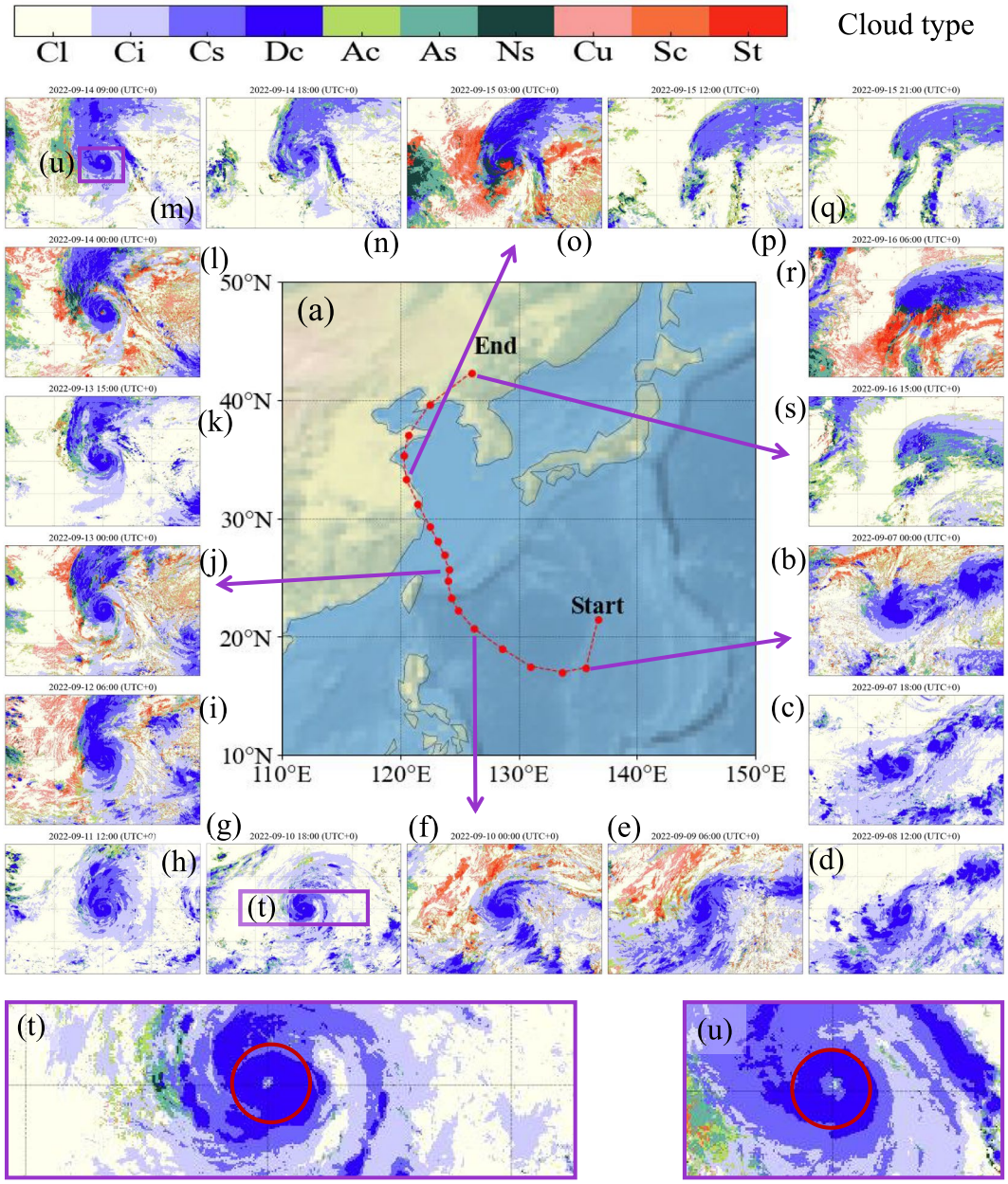


Fig. 12 The spatiotemporal characteristics of cloud types for typhoon Muifa in 2022.

Nanmadol, and Merbok in 2022 are shown in Fig. 11. It can be clearly observed that the distribution of cloud properties conforms to the characteristics of tropical cyclones, with spiral features. In Fig. 11a, deep convection clouds and cirrostratus clouds are distributed around the tropical cyclones, which implies that tropical cyclones are accompanied by heavy rainfall. And this also conforms to the meteorological characteristics of typhoons. The cloud optical thickness at the location of deep convection clouds is significantly larger in Fig. 11b, which meets the cloud type classification criteria. Overall, it is found that our proposed cloud property retrieval system can predict cloud property distribution well based on the cloud property distribution characteristics of the three typhoons. Although cloud product using passive remote sensing information has some shortcomings compared to active remote sensing (e.g., Lidar), which is capable of providing rich spatial structural information of the typhoon, our cloud product can still serve as a beneficial addition to typhoon observation.

The spatiotemporal characteristics of typhoon Muifa. In Fig. 12, the spatiotemporal characteristics of typhoon Muifa are depicted for cloud type. The trajectory of typhoon Muifa is shown in Fig. 12a, which reflects the entire process of typhoon generation and disappearance. The cloud type distribution within the range of 30° × 20° centered on the best typhoon track is selected for visualization. Figure 12b-s show

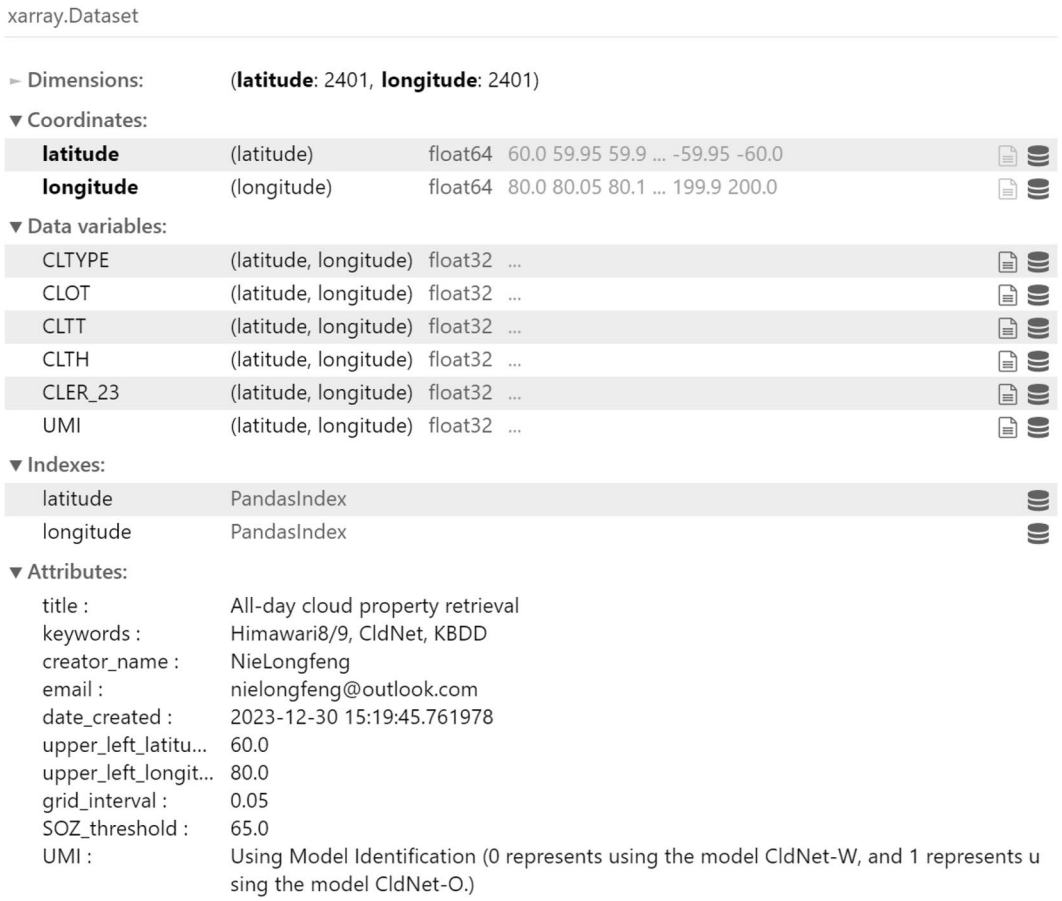


Fig. 13 The content of a cloud property file.

the spatiotemporal variation process of the distribution characteristics of typhoon cloud types. It undergoes a process of tropical depressions → tropical storms → strong tropical storms → typhoons → strong typhoons → typhoons → strong tropical storms → tropical storms → typhoon degeneration. In Fig. 12g and m, the typhoon eyes can be accurately captured, and their enlarged structures are shown in Fig. 12t and u, respectively. The consistency between the position of typhoon eye reflected by cloud type distribution and the best typhoon track confirms the reliability of the current cloud property retrieval system. More importantly, the current cloud property retrieval system can provide detailed cloud property distribution characteristics around the typhoon, which enables us to better understand the spatiotemporal characteristics of typhoon generation and disappearance.

Usage Notes

Based on the inference code and model parameters we provide, the all-day cloud property can be retrieved in real-time by inputting Himawari L1 Gridded data, which is saved in NetCDF format. Its spatial resolution is 0.05° × 0.05°, and the coverage range is from 80°E to 160°W, and from 60°N to 60°S.

These cloud property files can be read using the Xarray package in Python, and the content of the cloud property file is shown in Fig. 13. It contains coordinates, data variables related to cloud properties, and file attributes such as file creation date, grid resolution, and solar zenith angle threshold. Here, it is important to emphasize that CER has not undergone independent validation and that the error, particularly under nighttime conditions, is unknown. But for completeness, the CER is provided as part of our output product. Users should decide whether to apply CER to their research based on this information. Meanwhile, the current cloud property retrieval used a single-homogeneous cloud assumption in this study, and it could be highly problematic for overlapping or inhomogeneous clouds. Therefore, the uncertainty caused by overlapping or inhomogeneous clouds and the development of an overlapping or inhomogeneous cloud property retrieval system will be important topics for discussion and exploration in the future.

For the nine types of clouds and the total cloud, their occurrence probabilities are produced and stored according to all day, daytime, and nighttime conditions at the annual, seasonal, and monthly scales, resulting in a total of 90 NetCDF files. The contents of the three files regarding the total cloud occurrence probability at the annual, seasonal, and monthly scales are shown in Fig. 14. It should be noted that although the cloud type data for July 2015 does not cover the entire month, JJA and July 2015 are still saved in files at both seasonal and monthly scales during the calculation of cloud occurrence probability.

(a)

xarray.DataArray 'CP' (**time**: 7, **latitude**: 2401, **longitude**: 2401)

[40353607 values with dtype=float32]

▼ Coordinates:

time	(time)	int32	2016 2017 2018 2019 2020 2021 2022
latitude	(latitude)	float64	60.0 59.95 59.9 ... -59.95 -60.0
longitude	(longitude)	float64	80.0 80.05 80.1 ... 199.9 200.0

▼ Indexes:

time	PandasIndex
latitude	PandasIndex
longitude	PandasIndex

▼ Attributes:

long_name :	Cloud Probability
units :	Dimensionless
valid_min :	0
valid_max :	1
description :	Probability of cloud occurrence

(b)

xarray.DataArray 'CP' (**time**: 30, **latitude**: 2401, **longitude**: 2401)

[172944030 values with dtype=float32]

▼ Coordinates:

time	(time)	object	'2015-JJA' ... '2022-SON'
latitude	(latitude)	float64	60.0 59.95 59.9 ... -59.95 -60.0
longitude	(longitude)	float64	80.0 80.05 80.1 ... 199.9 200.0

▼ Indexes:

time	PandasIndex
latitude	PandasIndex
longitude	PandasIndex

▼ Attributes:

long_name :	Cloud Probability
units :	Dimensionless
valid_min :	0
valid_max :	1
description :	Probability of cloud occurrence

(c)

xarray.DataArray 'CP' (**time**: 90, **latitude**: 2401, **longitude**: 2401)

[518832090 values with dtype=float32]

▼ Coordinates:

time	(time)	object	'2015-07' '2015-08' ... '2022-12'
latitude	(latitude)	float64	60.0 59.95 59.9 ... -59.95 -60.0
longitude	(longitude)	float64	80.0 80.05 80.1 ... 199.9 200.0

▼ Indexes:

time	PandasIndex
latitude	PandasIndex
longitude	PandasIndex

▼ Attributes:

long_name :	Cloud Probability
units :	Dimensionless
valid_min :	0
valid_max :	1
description :	Probability of cloud occurrence

Fig. 14 The file content of total cloud occurrence probability at the (**a**) annual, (**b**) seasonal, and (**c**) monthly scales.

Code availability

The GitHub page https://github.com/metdai/PMD-CldNetV2 is available, which provides the relevant code, model parameters and CldNetV2 usage tutorial video for this study.

Received: 14 November 2024; Accepted: 14 February 2025;
Published online: 05 March 2025

References

1. Tabari, H. Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports* **10**, 13768, https://doi.org/10.1038/s41598-020-70816-2 (2020).
2. Kang, S. M., Ceppi, P., Yu, Y. & Kang, I.-S. Recent global climate feedback controlled by Southern Ocean cooling. *Nature Geoscience* **16**, 775–780, https://doi.org/10.1038/s41561-023-01256-6 (2023).
3. Gu, L. *et al*. Large anomalies in future extreme precipitation sensitivity driven by atmospheric dynamics. *Nature Communications* **14**, 3197, https://doi.org/10.1038/s41467-023-39039-7 (2023).
4. Tang, S. *et al*. Linkages of unprecedented 2022 Yangtze River Valley heatwaves to Pakistan flood and triple-dip La Nina. *npj Climate and Atmospheric Science* **6**, 44, https://doi.org/10.1038/s41612-023-00386-3 (2023).
5. Gibson, P. B. *et al*. Training machine learning models on climate model output yields skillful interpretable seasonal precipitation forecasts. *Communications Earth and Environment* **2**, 159, https://doi.org/10.1038/s43247-021-00225-4 (2021).
6. Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A. & Maher, N. More extreme precipitation in the world's dry and wet regions. *Nature Climate Change* **6**, 508–513, https://doi.org/10.1038/nclimate2941 (2016).
7. Wasko, C. & Nathan, R. Influence of changes in rainfall and soil moisture on trends in flooding. *Journal of Hydrology* **575**, 432–441, https://doi.org/10.1016/j.jhydrol.2019.05.054 (2019).
8. Vogel, M. M., Zscheischler, J., Wartenburger, R., Dee, D. & Seneviratne, S. I. Concurrent 2018 hot extremes across northern hemisphere due to human-induced climate change. *Earth's Future* **7**, 692–703, https://doi.org/10.1029/2019EF001189 (2019).
9. Ionita, M., Nagavciuc, V., Kumar, R. & Rakovec, O. On the curious case of the recent decade, mid-spring precipitation deficit in central europe. *npj Climate and Atmospheric Science* **3**, 49, https://doi.org/10.1038/s41612-020-00153-8 (2020).
10. Zaitchik, B. F., Rodell, M., Biasutti, M. & Seneviratne, S. I. Wetting and drying trends under climate change. *Nature Water* **1**, 502–513, https://doi.org/10.1038/s44221-023-00073-w (2023).
11. Bevacqua, E., Zappa, G., Lehner, F. & Zscheischler, J. Precipitation trends determine future occurrences of compound hot-dry events. *Nature Climate Change* **12**, 350–355, https://doi.org/10.1038/s41558-022-01309-5 (2022).
12. Yang, J. *et al*. Cloud behaviour on tidally locked rocky planets from global high-resolution modelling. *Nature Astronomy* https://doi.org/10.1038/s41550-023-02015-8 (2023).
13. Yi, B. Diverse cloud radiative effects and global surface temperature simulations induced by different ice cloud optical property parameterizations. *Scientific Reports* **12**, 10539, https://doi.org/10.1038/s41598-022-14608-w (2022).
14. Goldblatt, C., McDonald, V. L. & McCusker, K. E. Earth's long-term climate stabilized by clouds. *Nature Geoscience* **14**, 143–150, https://doi.org/10.1038/s41561-021-00691-7 (2021).
15. Watanabe, M., Kamae, Y., Shiogama, H., DeAngelis, A. M. & Suzuki, K. Low clouds link equilibrium climate sensitivity to hydrological sensitivity. *Nature Climate Change* **8**, 901–906, https://doi.org/10.1038/s41558-018-0272-0 (2018).
16. Cesana, G. V. & Del Genio, A. D. Observational constraint on cloud feedbacks suggests moderate climate sensitivity. *Nature Climate Change* **11**, 213–218, https://doi.org/10.1038/s41558-020-00970-y (2021).
17. Gan, B., Liu, X., Yang, X., Wang, X. & Zhou, J. The impact of human activities on the occurrence of mountain flood hazards: lessons from the 17 August 2015 flash flood/debris flow event in Xuyong county, south-western China. *Geomatics, Natural Hazards and Risk* **9**, 816–840, https://doi.org/10.1080/19475705.2018.1480539 (2018).
18. Li, N. *et al*. Characteristics of the disastrous debris flow of Chediguan gully in Yinxing town, Sichuan Province, on August 20, 2019. *Scientific Reports* **11**, 23666, https://doi.org/10.1038/s41598-021-03125-x (2021).
19. Cabral, V. *et al*. The consequences of debris flows in brazil: a historical analysis based on recorded events in the last 100 years. *Landslides* **20**, 511–529, https://doi.org/10.1007/s10346-022-01984-7 (2023).
20. Ke, Q. *et al*. Urban pluvial flooding prediction by machine learning approaches - a case study of Shenzhen city, China. *Advances in Water Resources* **145**, 103719, https://doi.org/10.1016/j.advwatres.2020.103719 (2020).
21. Wang, Z., Chen, X., Qi, Z. & Cui, C. Flood sensitivity assessment of super cities. *Scientific Reports* **13**, 5582, https://doi.org/10.1038/s41598-023-32149-8 (2023).
22. Choi, Y., Kang, J. & Kim, J. Urban flood adaptation planning for local governments: Hydrology analysis and optimization. *International Journal of Disaster Risk Reduction* **59**, 102213, https://doi.org/10.1016/j.ijdrr.2021.102213 (2021).
23. Dąbrowska, J. *et al*. Between flood and drought: How cities are facing water surplus and scarcity. *Journal of Environmental Management* **345**, 118557, https://doi.org/10.1016/j.jenvman.2023.118557 (2023).
24. Anderson, W. *et al*. Violent conflict exacerbated drought-related food insecurity between 2009 and 2019 in sub-Saharan Africa. *Nature Food* **2**, 603–615, https://doi.org/10.1038/s43016-021-00327-4 (2021).
25. Kennett, D. J. *et al*. Drought-induced civil conflict among the ancient Maya. *Nature Communications* **13**, 3911, https://doi.org/10.1038/s41467-022-31522-x (2022).
26. Perkins-Kirkpatrick, S. E. & Lewis, S. C. Increasing trends in regional heatwaves. *Nature Communications* **11**, 3357, https://doi.org/10.1038/s41467-020-16970-7 (2020).
27. Cai, Y. *et al*. Vegetation cover changes in China induced by ecological restoration-protection projects and land-use changes from 2000 to 2020. *CATENA* **217**, 106530, https://doi.org/10.1016/j.catena.2022.106530 (2022).
28. Zhang, S. *et al*. Evolution of vegetation dynamics and its response to climate in ecologically fragile regions from 1982 to 2020: A case study of the Three Gorges Reservoir area. *CATENA* **219**, 106601, https://doi.org/10.1016/j.catena.2022.106601 (2022).
29. Song, X.-P. *et al*. Global land change from 1982 to 2016. *Nature* **560**, 639–643, https://doi.org/10.1038/s41586-018-0411-9 (2018).
30. Smith, C., Baker, J. C. A. & Spracklen, D. V. Tropical deforestation causes large reductions in observed precipitation. *Nature* **615**, 270–275, https://doi.org/10.1038/s41586-022-05690-1 (2023).
31. Aybar, C. *et al*. CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2. *Scientific Data* **9**, 782, https://doi.org/10.1038/s41597-022-01878-2 (2022).
32. BESSHO, K. *et al*. An introduction to Himawari-8/9 — Japan's new-generation geostationary meteorological satellites. *Journal of the Meteorological Society of Japan. Ser. II* **94**, 151–183, https://doi.org/10.2151/jmsj.2016-009 (2016).
33. Flores-Anderson, A. I. *et al*. Spatial and temporal availability of cloud-free optical observations in the tropics to monitor deforestation. *Scientific Data* **10**, 550, https://doi.org/10.1038/s41597-023-02439-x (2023).
34. Mecikalski, J. R., Rosenfeld, D. & Manzato, A. Evaluation of geostationary satellite observations and the development of a 1-2h prediction model for future storm intensity. *Journal of Geophysical Research: Atmospheres* **121**, 6374–6392, https://doi.org/10.1002/2016JD024768 (2016).
35. Hashimoto, H. *et al*. New generation geostationary satellite observations support seasonality in greenness of the Amazon evergreen forests. *Nature Communications* **12**, 684, https://doi.org/10.1038/s41467-021-20994-y (2021).
36. Zhao, Y. *et al*. Evaluating the potential of H8/AHI geostationary observations for monitoring vegetation phenology over different ecosystem types in northern China. *International Journal of Applied Earth Observation and Geoinformation* **112**, 102933, https://doi.org/10.1016/j.jag.2022.102933 (2022).
37. Iwabuchi, H., Saito, M., Tokoro, Y., Putri, N. S. & Sekiguchi, M. Retrieval of radiative and microphysical properties of clouds from multispectral infrared measurements. *Progress in Earth and Planetary Science* **3**, 32, https://doi.org/10.1186/s40645-016-0108-3 (2016).
38. Loyola, D. G. *et al*. The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor. *Atmospheric Measurement Techniques* **11**, 409–427, https://doi.org/10.5194/amt-11-409-2018 (2018).

39. Zhang, C., Zhuge, X. & Yu, F. Development of a high spatiotemporal resolution cloud-type classification approach using Himawari-8 and CloudSat. *International Journal of Remote Sensing* **40**, 6464–6481, https://doi.org/10.1080/01431161.2019.1594438 (2019).
40. Shang, H. *et al*. Development of a daytime cloud and haze detection algorithm for Himawari-8 satellite measurements over central and eastern China. *Journal of Geophysical Research: Atmospheres* **122**, 3528–3543, https://doi.org/10.1002/2016JD025659 (2017).
41. Kouki, M., Toshiharu, I., Hiroshi, S. & Ryo, Y. *Algorithm Theoretical Basis Document for Cloud Type/Phase Product*, Meteorological satellite center technical note. https://www.data.jma.go.jp/mscweb/technotes/msctechrep61-2.pdf (2016).
42. Hiroshi, S., Takahito, I., Kouki, M. & Ryo, Y.*High-resolution Cloud Analysis Information derived from Himawari-8 data*, Meteorological satellite center technical note. https://www.data.jma.go.jp/mscweb/technotes/msctechrep61-4.pdf (2016).
43. Nakajima, T. Y. *et al*. Theoretical basis of the algorithms and early phase results of the GCOM-C (Shikisai) SGLI cloud products. *Progress in Earth and Planetary Science* **6**, 52, https://doi.org/10.1186/s40645-019-0295-9 (2019).
44. Wang, X., Iwabuchi, H. & Yamashita, T. Cloud identification and property retrieval from Himawari-8 infrared measurements via a deep neural network. *Remote Sensing of Environment* **275**, 113026, https://doi.org/10.1016/j.rse.2022.113026 (2022).
45. Wang, Q. *et al*. Retrieval of cloud properties from thermal infrared radiometry using convolutional neural network. *Remote Sensing of Environment* **278**, 113079, https://doi.org/10.1016/j.rse.2022.113079 (2022).
46. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444, https://doi.org/10.1038/nature14539 (2015).
47. Xu, H., Chen, Y. & Zhang, D. Worth of prior knowledge for enhancing deep learning. *Nexus* **1**, 100003, https://doi.org/10.1016/j.ynexs.2024.100003 (2024).
48. Chen, Y. & Zhang, D. Integration of knowledge and data in machine learning. *arXiv* https://arxiv.org/abs/2202.10337 (2022).
49. Ravuri, S. *et al*. Skilful precipitation nowcasting using deep generative models of radar. *Nature* **597**, 672–677, https://doi.org/10.1038/s41586-021-03854-z (2021).
50. Li, X. *et al*. Big data in earth system science and progress towards a digital twin. *Nature Reviews Earth and Environment* **4**, 319–332, https://doi.org/10.1038/s43017-023-00409-w (2023).
51. Nakajima, T. Y. & Nakajma, T. Wide-area determination of cloud microphysical properties from noaa avhrr measurements for fire and astex regions. *Journal of Atmospheric Sciences* **52**, 4043 – 4059, https://doi.org/10.1175/1520-0469(1995)052<4043:WADOCM>2.0.CO;2 (1995).
52. Ishida, H. & Nakajima, T. Y. Development of an unbiased cloud detection algorithm for a spaceborne multispectral imager. *Journal of Geophysical Research: Atmospheres* **114**, D07206, https://doi.org/10.1029/2008JD010710 (2009).
53. Nakajima, T. Y. & Letu, H.*Algorithm Theoretical Basis Document of CAPCOM for GCOM/SLGI*, https://suzaku.eorc.jaxa.jp/GCOM_C/data/ATBD/ver1/Nakajima_ATBD_SGLI_CAPCOM20110920.pdf (2011).
54. Huang, Y. *et al*. Distribution and variability of satellite-derived signals of isolated convection initiation events over central eastern China. *Journal of Geophysical Research: Atmospheres* **122**, 11357–11373, https://doi.org/10.1002/2017JD026946 (2017).
55. Taniguchi, D., Yamazaki, K. & Uno, S. The great dimming of betelgeuse seen by the Himawari-8 meteorological satellite. *Nature Astronomy* **6**, 930–935, https://doi.org/10.1038/s41550-022-01680-5 (2022).
56. Nie, L., Chen, Y., Zhang, D., Liu, X. & Yuan, W. QIENet: Quantitative irradiance estimation network using recurrent neural network based on satellite remote sensing data. *International Journal of Applied Earth Observation and Geoinformation* **127**, 103584, https://doi.org/10.1016/j.jag.2023.103584 (2024).
57. Letu, H. *et al*. Ice cloud properties from Himawari-8/AHI next-generation geostationary satellite: Capability of the AHI to monitor the DC cloud generation process. *IEEE Transactions on Geoscience and Remote Sensing* **57**, 3229–3239, https://doi.org/10.1109/TGRS.2018.2882803 (2019).
58. Lai, R. *et al*. Comparison of cloud properties from Himawari-8 and FengYun-4A geostationary satellite radiometers with MODIS cloud retrievals. *Remote Sensing* **11**, 1703, https://doi.org/10.3390/rs11141703 (2019).
59. Letu, H. *et al*. High-resolution retrieval of cloud microphysical properties and surface solar radiation using Himawari-8/AHI next-generation geostationary satellite. *Remote Sensing of Environment* **239**, 111583, https://doi.org/10.1016/j.rse.2019.111583 (2020).
60. Liu, C.-Y., Chiu, C.-H., Lin, P.-H. & Min, M. Comparison of cloud-top property retrievals from Advanced Himawari Imager, MODIS, CloudSat/CPR, CALIPSO/CALIOP, and Radiosonde. *Journal of Geophysical Research: Atmospheres* **125**, e2020JD032683, https://doi.org/10.1029/2020JD032683 (2020).
61. Huang, Y. *et al*. Evaluating Himawari-8 cloud products using Shipborne and CALIPSO observations: Cloud-top height and cloud-top temperature. *Journal of Atmospheric and Oceanic Technology* **36**, 2327 – 2347, https://doi.org/10.1175/JTECH-D-18-0231.1 (2019).
62. Min, M., Li, J., Wang, F., Liu, Z. & Menzel, W. P. Retrieval of cloud top properties from advanced geostationary satellite imager measurements based on machine learning algorithms. *Remote Sensing of Environment* **239**, 111616, https://doi.org/10.1016/j.rse.2019.111616 (2020).
63. Rossow, W. B. & Schiffer, R. A. Advances in understanding clouds from ISCCP. *Bulletin of the American Meteorological Society* **80**, 2261–2288, https://doi.org/10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2 (1999).
64. Wang, Z. & Sassen, K. Cloud type and macrophysical property retrieval using multiple remote sensors. *Journal of Applied Meteorology* **40**, 1665–1682, https://doi.org/10.1175/1520-0450(2001)040<1665:CTAMPR>2.0.CO;2 (2001).
65. Nie, L., Chen, Y., Du, M., Sun, C. & Zhang, D. A knowledge-based data-driven (KBDD) framework for all-day identification of cloud types using satellite remote sensing. *Remote Sensing of Environment* **304**, 114054, https://doi.org/10.1016/j.rse.2024.114054 (2024).
66. Dietterich, T. Overfitting and undercomputing in machine learning. *ACM Computing Surveys* **27**, 326–327, https://doi.org/10.1145/212094.212114 (1995).
67. Huang, Y. *et al*. Distribution and variability of satellite-derived signals of isolated convection initiation events over central eastern China. *Journal of Geophysical Research: Atmospheres* **122**, 11,357–11,373, https://doi.org/10.1002/2017JD026946 (2017).
68. Purbantoro, B. *et al*. Comparison of cloud type classification with split window algorithm based on different infrared band combinations of Himawari-8 satellite. *Advances in Remote Sensing* **7**, 218–234, https://doi.org/10.4236/ars.2018.73015 (2018).
69. Yang, Y. *et al*. Machine learning-based retrieval of day and night cloud macrophysical parameters over East Asia using Himawari-8 data. *Remote Sensing of Environment* **273**, 112971, https://doi.org/10.1016/j.rse.2022.112971 (2022).
70. Nie, L., Chen, Y. & Zhang, D. Code and tutorial video for cloud property retrieval system CldNetV2. *figshare* https://doi.org/10.6084/m9.figshare.26893474.v3 (2024).
71. Nie, L., Chen, Y. & Zhang, D. All-day cloud property every 3 hours throughout 2022. *figshare* https://doi.org/10.6084/m9.figshare.c.7417813.v2 (2024).
72. Nie, L., Chen, Y. & Zhang, D. The dataset of different type cloud occurrence probability at the annual, seasonal, and monthly scales for all day, daytime, and nighttime from 2015 to 2022. *figshare* https://doi.org/10.6084/m9.figshare.27368562.v1 (2024).
73. Ying, M. *et al*. An overview of the China meteorological administration tropical cyclone database. *Journal of Atmospheric and Oceanic Technology* **31**, 287 – 301, https://doi.org/10.1175/JTECH-D-12-00119.1 (2014).
74. Lu, X. *et al*. Western north Pacific tropical cyclone database created by the China meteorological administration. *Advances in Atmospheric Sciences* **38**, 690–699, https://doi.org/10.1007/s00376-020-0211-7 (2021).

Acknowledgements
This work was supported and partially funded by the China Meteorological Administration Climate Change Special Program (CMA-CCSP) under Grant QBZ202316, the National Center for Applied Mathematics Shenzhen (NCAMS), the Major Key Project of PCL (Grant No. PCL2022A05, PCL2023A09 & PCL2024A06), the National Key Research and Development Program (Grant No. 2024YFF1500600), and High Performance Computing Centers at Eastern Institute of Technology, Ningbo, and Ningbo Institute of Digital Twin.

Author contributions
Longfeng Nie and Dongxiao Zhang conceived the study. Longfeng Nie and Yuntian Chen carried out and analysed the work. Longfeng Nie wrote the initial manuscript. Longfeng Nie, Yuntian Chen and Dongxiao Zhang reviewed and approved the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Y.C. or D.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025