International Journal of Applied Earth Observation and Geoinformation 135 (2024) 104254

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Cloud probability distribution of typical urban agglomerations in China based on Sentinel-2 satellite remote sensing

Jing Ling [a,b,c], Rui Liu [a,c], Shan Wei [a,c], Shaomei Chen [a,c], Luyan Ji [d,e,f], Yongchao Zhao [d,e,f,g], Hongsheng Zhang [a,c,*]

[a] *Department of Geography, The University of Hong Kong, Pokfulam, Hong Kong SAR, China*
[b] *School of Information Engineering, Guangdong University of Technology, Guangzhou, China*
[c] *The University of Hong Kong Shenzhen Institute of Research and Innovation, Shenzhen, China*
[d] *Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China*
[e] *Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Chinese Academy of Sciences, Bejing, China*
[f] *Key Laboratory of Target Cognition and Application Technology, Beijing, China*
[g] *School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China*

A R T I C L E I N F O

Keywords:
Cloud distribution characteristics
Climate change
Remote sensing observation
Urbanization
Land cover
Subtropical

A B S T R A C T

Cloud distribution significantly impacts global climate change, ecosystem health, urban environments, and satellite remote sensing observations. However, past research has primarily focused on the meteorological characteristics of clouds with limitations in scale and resolution, leading to an insufficient understanding of large-scale cloud distribution and its relationship with land surface cover and urbanization. This study investigates the cloud distribution characteristics of typical urban agglomerations in different climatic regions of China using high-resolution Sentinel-2 satellite imagery and the Google Earth Engine platform. A cloud probability descriptor was constructed based on three years of high spatiotemporal resolution observations. The results revealed significant differences in cloud distribution among urban agglomerations, challenging the traditional understanding based on climate zoning. The Northeast urban agglomeration in the temperate zone exhibited high cloud coverage (37.54%), while the Chengdu-Chongqing urban agglomeration in the subtropical zone and the Qinghai-Tibet Plateau urban agglomeration in the plateau climate zone had even higher average cloud probabilities (50.72% and 43.27%, respectively). The analysis suggests land surface cover, urbanization, and other surface factors may influence cloud distribution patterns. These findings emphasize the ubiquity of cloud cover and highlight the importance of considering the complex interactions among geographical factors when characterizing cloud cover diversity. This study contributes to providing new insights for enhancing meteorological models and remote sensing observation strategies in cloudy environments across different climate zones.

1. Introduction

Clouds are an important component of the atmosphere, directly affecting solar radiation and Earth's surface reflection of sunlight, thereby regulating the Earth's heat distribution, temperature, humidity, and influencing multiple key elements such as energy balance, climate systems, and the water cycle (Toll et al., 2019). Cloud cover directly affects the reception of solar radiation, regulation of surface temperature, maintenance of energy balance, and formation of precipitation patterns. Therefore, an in-depth study of cloud distribution characteristics and their influencing factors plays a crucial role in comprehensively understanding global climate change, ecosystem health, and effective environmental management (Shah et al., 2020).

Cloud formation is a complex process involving various physical mechanisms. When water vapor in the air reaches a saturated state, condensation occurs, forming cloud droplets. This typically happens when air rises, cools, and reaches its dew point temperature (Wallace and Hobbs, 2006). On a global scale, the acceleration of urbanization has led to significant changes in local climate and cloud distribution within cities (Shepherd et al., 2010). The unique conditions of urban environments, including impervious surface, urban pollutants and urban heat island can influence the cloud formation process. The spatial

* Corresponding author.
E-mail addresses: jingling@connect.hku.hk (J. Ling), rhysliu@connect.hku.hk (R. Liu), shanw@connect.hku.hk (S. Wei), 2020803203@stu.njau.edu.cn (S. Chen), jily@mail.ustc.edu.cn (L. Ji), ofcours_sure@sina.com.cn (Y. Zhao), zhanghs@hku.hk (H. Zhang).

https://doi.org/10.1016/j.jag.2024.104254
Received 26 July 2024; Received in revised form 18 October 2024; Accepted 3 November 2024
Available online 9 November 2024

heterogeneity of urban surface characteristics may also affect local circulations and atmospheric boundary layer processes, influencing cloud formation and distribution (Shepherd, 2005). Given the complex interactions between cloud formation and urban conditions, investigating the cloud distribution characteristics in urban areas is crucial for understanding urban environments and regional climate and weather.

In addition to the meteorological field, clouds also hold significant value in satellite remote sensing observations. Cloud cover affects the observation of the Earth's surface by remote sensing satellites (Ling et al., 2021). Cloud layers obstruct optical signals, making it challenging to obtain clear, cloud-free remote sensing images over large areas, which greatly impacts subsequent tasks such as object recognition and monitoring (Shen et al., 2014; Zhang et al., 2014; Zhu et al., 2021). Therefore, for cloud-covered regions, satellite remote sensing faces challenges in efficiency and accuracy of surface observations (Ling and Zhang, 2023). In this context, accurately understanding cloud distribution characteristics will help optimize remote sensing observation strategies and improve the efficiency and precision of surface observations.

Although research on clouds has gradually received attention over the past few decades, it has mainly focused on aspects such as the microphysical properties, radiative effects, and precipitation mechanisms of clouds (Stewart et al., 1998). There have also been analyses of cloud climate characteristics (Fu et al., 2020) and studies on cloud condensation nuclei characteristics (Shen et al., 2019). However, these studies tend to focus more on using ground-based cloud data to analyze cloud amounts (Singh and Glennen, 2005) or focus on local-scale cloud distribution analysis (Yang et al., 2020a). The influence of cloud height, thickness, and morphology on near-surface air temperature has also been widely studied (Jiang et al., 2022). Nevertheless, these studies focus more on the physical properties and meteorological functions of clouds and are primarily based on ground observation data such as meteorological station records (Leena et al., 2022), thus having limitations in coverage and spatial resolution. In this context, meteorological satellite data have become an important supplement to ground observation data due to their extensive coverage, rich information content, and high-frequency repeated observations. The use of satellite cloud data for climate analysis and diagnosis has received widespread attention (Norris et al., 2016). Polar-orbiting meteorological satellites observe the global surface at the same local time and can serve climate observations and monitor large-scale natural disasters. Geostationary meteorological satellites, on the other hand, can synchronously rotate with the Earth and perform high-frequency observations of the fixed area they cover, serving weather forecasting and analysis and providing cloud image data (Bessho et al., 2016). However, there are still few studies that use these satellite images for comprehensive quantitative analysis of large-scale cloud distribution characteristics. For example, global cloud distribution analysis using CloudSat satellite data (Hagihara et al., 2010), analysis of China's cloud amount characteristics using Terra and Aqua satellite cloud amount data (Ma et al., 2014), analysis of cloud characteristics using Himawari-8 satellite data (Yang et al., 2020b), detection of convective cloud using FY-2 VISSR satellite data (Liang et al., 2017), and exploration of the spatiotemporal distribution of different cloud types in China using ISCCP cloud data (Sirui et al., 2020). However, due to data availability and computational processing capabilities, these studies often have issues such as low resolution or small research scales. Moreover, past research has focused on the high-altitude distribution and meteorological characteristics of clouds, with insufficient attention to the underlying surface and little exploration of the relationship between clouds and land surface cover, urbanization, and other factors (Ma et al., 2014; Norris et al., 2016). Therefore, a more in-depth understanding of cloud distribution is needed. For example, past research in the field of remote sensing earth observation often considers subtropical regions as cloudy and rainy areas. However, considering the vastness of China's territory and the diversity of its climate, there are significant differences in cloud distribution in different geographical locations and climate zones, and traditional climate zoning may not fully describe the distribution patterns of clouds. Further systematic research on the long-term, large-scale, and high-resolution spatial distribution characteristics of clouds is still needed.

Facing this challenge, this study selected urban agglomerations in different geographical locations, topography, and climate zones in China as research objects. Utilizing high-resolution Sentinel-2 satellite remote sensing imagery and the Google Earth Engine (GEE) data processing platform, this study leveraged three years of continuous high spatiotemporal resolution remote sensing data to establish a cloud probability descriptor. From a remote sensing perspective, we explored the high spatiotemporal resolution cloud distribution characteristics of urban agglomerations in typical climate zones of China. The main contributions of this study include: (1) presenting a novel approach for mapping high-resolution cloud probability distribution using Sentinel-2 imagery and GEE, investigating cloud distribution characteristics from a remote sensing perspective; (2) conducting a comprehensive analysis of cloud distribution characteristics across urban agglomerations in different climatic regions of China, revealing regional differences in cloud distribution and complementing the traditional understanding based on climate zoning; and (3) exploring the relationship between cloud distribution and multiple influencing factors, such as land surface cover and topography. Through visual analysis models and quantitative analysis, this study aims to deepen the understanding of the regional differences in cloud distribution, reveal the complex interactions that shape cloud distribution patterns, and provide scientific support for decision-making in regional climate change, urban planning, and environmental management.

2. Cloud probability distribution mapping method for typical climate zones

This section will introduce the cloud distribution acquisition method for typical climate zones in this study. Starting from data acquisition and processing, extraction of typical climate zone urban agglomerations, cloud probability calculation, cloud distribution mapping, and quantitative analysis, this study systematically explores cloud distribution characteristics and reveals the cloud probability distribution in different climate zones of China. Fig. 1 shows the main workflow of the research method.

2.1. Data acquisition and processing

This study aims to conduct a broad analysis of cloud distribution. To this end, it requires the use of open-source satellite data with rich data volume, and Sentinel-2 satellite data demonstrate unique advantages compared to other satellites due to their excellent performance. Sentinel-2 is a high-resolution multispectral imaging satellite developed and operated by the European Space Agency (ESA), with a revisit period of approximately 5 days. The abundant data brought by their high temporal resolution helps to enhance the accuracy and reliability of cloud information. Compared to traditional remote sensing satellites, Sentinel-2 has a finer spatial resolution, which helps to more accurately analyze cloud information. Moreover, Sentinel-2 satellite data provides rich spectral information, including visible, near-infrared, short-wave infrared, and other bands, which helps to accurately identify clouds in images.

The cloud information band in Sentinel-2 imagery data, known as the QA60 band, is a bitmask band that provides information about cloud presence. Bits 10 and 11 of the QA60 band represent opaque clouds and cirrus clouds, respectively, at a spatial resolution of 60 m. The Sentinel-2 cloud detection algorithm utilizes a series of spectral reflectance thresholds, ratios, and indices (e.g., NDSI, NDVI) to identify potential cloud pixels. Studies have shown that the Sentinel-2 cloud product provides a reliable source of cloud information, with an overall accuracy exceeding 90 % in most cases (Coluzzi et al., 2018). Sentinel-2's cloud

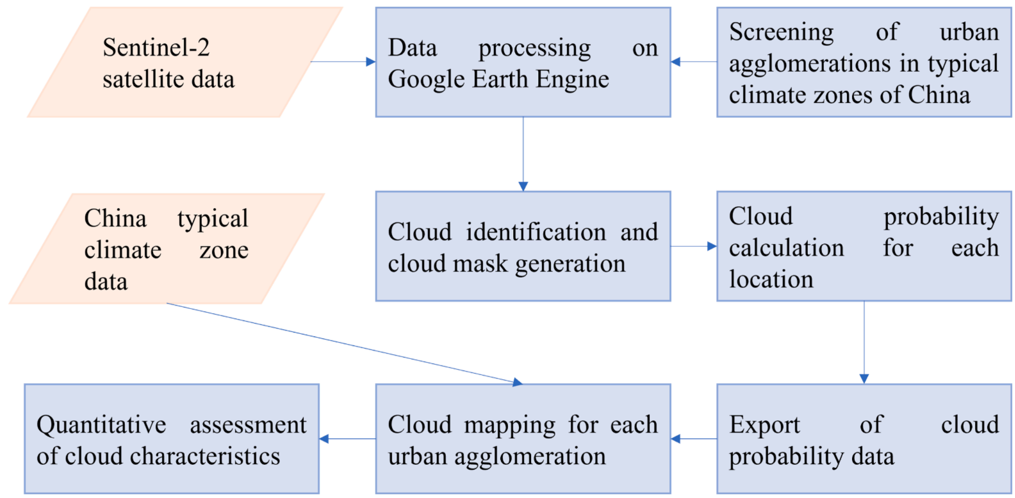

Fig. 1. Workflow of the cloud probability distribution mapping method. Rectangles represent processes, and parallelograms represent input data.

identification product is widely recognized for detecting clouds in original images (Meygret et al., 2009) and is used as a recommended routine cloud detection algorithm on the Google Earth Engine platform (Li et al., 2022a). It is commonly used as a cloud mask for declouding to obtain cloud-free images in various surface monitoring studies (Peterson et al., 2020; Yang et al., 2021). Although there are other cloud detection methods, such as Fmask (Zhu et al., 2015), temporal-based method Tmask (Zhu and Woodcock, 2014), and automatic time-series analysis method ATSA (Zhu and Helmer, 2018), we chose the Sentinel-2 QA band cloud product for its reliability and its convenience, consistency, and computational efficiency in large-scale spatiotemporal analysis, as the cloud detection results are directly included in the Sentinel-2 data.

To ensure the reliability and accuracy of cloud probability calculations, we performed four key data processing steps. First, we acquired Sentinel-2 imagery data covering the continuous three-year period from January 1, 2020, to December 31, 2022, from the Google Earth Engine platform to obtain sufficient data samples and improve statistical significance. We then filtered the imagery data to retain only data covering the geographical boundaries of each urban agglomeration, reducing computational complexity and enabling efficient cloud information extraction. Next, we performed bit operations on the QA60 band of Sentinel-2 to identify opaque and cirrus clouds, distinguishing cloud pixels from non-cloud pixels and generating cloud masks, following existing practices (Yang et al., 2021; Peterson et al., 2020). Finally, for each location, we conducted cloud probability calculation. With the help of the cloud information band of Sentinel-2 satellites, this study can efficiently obtain cloud distribution information and conduct a comprehensive analysis of cloud distribution in typical climate zones across China.

2.2. Extraction of typical climate zone urban agglomerations in China

China has a vast territory and diverse climate types. Based on the distribution patterns of temperature, precipitation, and other meteorological elements, the China Meteorological Administration divided the country into 10 first-level climate regions (Institute of Geographic Sciences and Natural Resources, (2023)). Each region has unique climate characteristics and ecological environments, including the northern temperate zone, mid-temperate zone, southern temperate zone, northern subtropical zone, mid-subtropical zone, southern subtropical zone, northern tropical zone, mid-tropical zone, southern tropical zone, and plateau climate zone.

To further investigate and verify the cloud distribution situation in China, this study is not limited to the subtropical region but includes multiple typical climate zones in China, aiming for a more comprehensive cloud distribution analysis. To achieve this goal, six representative typical climate zones in China were carefully selected, including the mid-temperate zone, southern temperate zone, northern subtropical zone, mid-subtropical zone, southern subtropical zone, and plateau climate zone. Within each typical climate zone, 9 representative urban agglomerations were further delineated (Institute, 2022) to ensure coverage of cities with similar climate characteristics, providing a regional basis for subsequent cloud distribution analysis.

Fig. 2 shows the climate zoning map of China and marks the distribution of the 9 research areas selected in this study. Through the overlay of the climate zone layer and the research area boundaries, it intuitively presents the relationship between the covered urban agglomerations and their corresponding typical climate zones. The selected research areas are relatively evenly distributed across China, covering various typical regions of China and having a relatively comprehensive representative role. It can be observed that the mid-temperate zone, southern temperate zone, and mid-subtropical climate zone each include two urban agglomerations. This is because the same climate zone covers a wide range, and the regions within it may have huge differences in land surface cover, resulting in different cloud characteristics. Therefore, for specific climate zones with diverse topography, this study selected multiple representative urban agglomerations to more comprehensively analyze the relationship between cloud distribution and various factors. For example, although the Beijing-Tianjin-Hebei urban agglomeration is one of the representative urban agglomerations in the southern temperate climate zone, the Southern Xinjiang urban agglomeration, located in the same southern temperate zone, may exhibit significantly different cloud cover performance due to its distinctive desert land surface cover. Therefore, the southern temperate zone includes these two representative urban agglomerations. Table 1 lists the representative urban agglomerations covered in each typical climate zone, as well as the specific cities included.

2.3. Cloud probability calculation

With the help of the GEE platform and Sentinel-2 satellite imagery data, this study performed cloud probability calculations for each geographical location within the urban agglomerations across China. Referring to the practices in existing research (Yang et al., 2021; Peterson et al., 2020), this study performs bitwise operations on the QA60 band of Sentinel-2 to identify clouds. Next, for each pixel location, a time series analysis is performed using all the images covering that specific location. The cloud probability P at each pixel is calculated by dividing the number of times the pixel is labeled as cloudy (N_c) by the total number of images (N_t) covering that pixel location:

Fig. 2. Geographical distribution of urban agglomerations studied in China's typical climate zones.

Table 1
Representative urban clusters covered in typical climatic zones.

Climate Zone	Urban Agglomeration	Cities
Mid-Temperate	Northeast Urban Agglomeration	Changchun, Jilin, Yanbian Korean Autonomous Prefecture, Harbin, Mudanjiang
	Inner Mongolia Urban Agglomeration	Hohhot, Baotou, Ordos
South Temperate	Beijing-Tianjin-Hebei Urban Agglomeration	Beijing, Tianjin, Baoding, Tangshan, Langfang, Shijiazhuang, Qinhuangdao, Zhangjiakou, Chengde, Cangzhou, Hengshui, Xingtai, Handan, Anyang
	South Xinjiang Urban Agglomeration	Kashgar, Hotan
North Subtropical	Yangtze River Delta Urban Agglomeration	Shanghai, Nanjing, Wuxi, Changzhou, Suzhou, Nantong, Yancheng, Yangzhou, Zhenjiang, Taizhou, Hangzhou, Ningbo, Jiaxing, Huzhou, Shaoxing, Jinhua, Zhoushan, Taizhou, Hefei, Wuhu, Ma'anshan, Tongling, Anqing, Chuzhou, Chizhou, Xuancheng
Mid-Subtropical	Chengdu-Chongqing Urban Agglomeration	Chengdu, Zigong, Luzhou, Deyang, Mianyang, Suining, Neijiang, Leshan, Nanchong, Meishan, Yibin, Guang'an, Dazhou, Ya'an, Ziyang, Chongqing
	Yunnan Urban Agglomeration	Kunming, Lijiang, Chuxiong Yi Autonomous Prefecture, Dali Bai Autonomous Prefecture
South Subtropical	Greater Bay Area Urban Agglomeration	Hong Kong, Macau, Guangzhou, Shenzhen, Zhuhai, Foshan, Huizhou, Dongguan, Zhongshan, Jiangmen, Zhaoqing
Plateau Climate	Qinghai-Tibet Plateau Urban Agglomeration	Lhasa, Shigatse, Nagqu

$$P = N_c / N_t \quad (1)$$

This percentage represents the cloud probability value for each pixel location, reflecting the frequency of cloud occurrence at that specific point over the entire time series.

Through the above steps, this study utilized the GEE platform and Sentinel-2 satellite data to achieve cloud probability calculations at a spatial resolution of 60 m for each geographical location within the research scope. This complete process provided remote sensing-based information for this study, enabling the exploration of cloud distribution characteristics in different climate regions of China and providing strong support for cloud distribution analysis.

2.4. Cloud distribution mapping and quantitative analysis of typical climate zones in China

To gain a deeper understanding of the cloud distribution in typical climate zones of China, this study conducted research from the perspectives of visualization and quantitative analysis of large-scale cloud distribution.

By processing the obtained three-year cloud probability data, this study first converted it into cloud cover percentage data, reflecting the degree to which each geographical location was covered by clouds over the three years. Then, a color gradient legend was used to map the cloud cover percentage onto the cloud distribution map, allowing different cloud probability regions to be displayed in different colors. To better locate the positions of various urban agglomerations, this study overlaid the climate zoning map of China on the cloud distribution map to show the spatial geographical distribution of different climate regions and urban agglomerations.

In addition to visual display of cloud distribution, this study conducted quantitative statistical analysis to obtain more in-depth information. Specifically, this study plotted cloud probability distribution histograms for each urban agglomeration, revealing the cloud distribution by counting the number of pixels in different cloud probability intervals. The study further calculated the mean and standard deviation of cloud probability for each urban agglomeration within different climate zones, aiding in the quantitative assessment of the concentration and variability of cloud distribution. These quantitative analysis results will provide accurate data support and further deepen this study's understanding of the cloud distribution characteristics of urban agglomerations in different climate zones.

3. Cloud probability distribution analysis of subtropical regions in China

3.1. Differences in cloud probability distribution between subtropical and other climate zones

Through detailed analysis of the experimental results obtained using the above research methods, Fig. 3 presents the cloud probability distribution of 9 typical climate zone urban agglomerations in China. This visual cloud probability map allows the study to reveal significant differences in cloud distribution among different urban agglomerations. The color gradient approach used in the figure, transitioning from blue to red, intuitively displays the different levels of cloud cover probability, highlighting the diversity and regional differences in cloud distribution among different urban agglomerations.

Previous environmental remote sensing observation studies considered subtropical regions as cloudy and rainy areas. However, the results of this study indicate that although the Chengdu-Chongqing urban agglomeration in the mid-subtropical region exhibits the highest cloud coverage rate, not all subtropical regions are cloudy zones. Take the

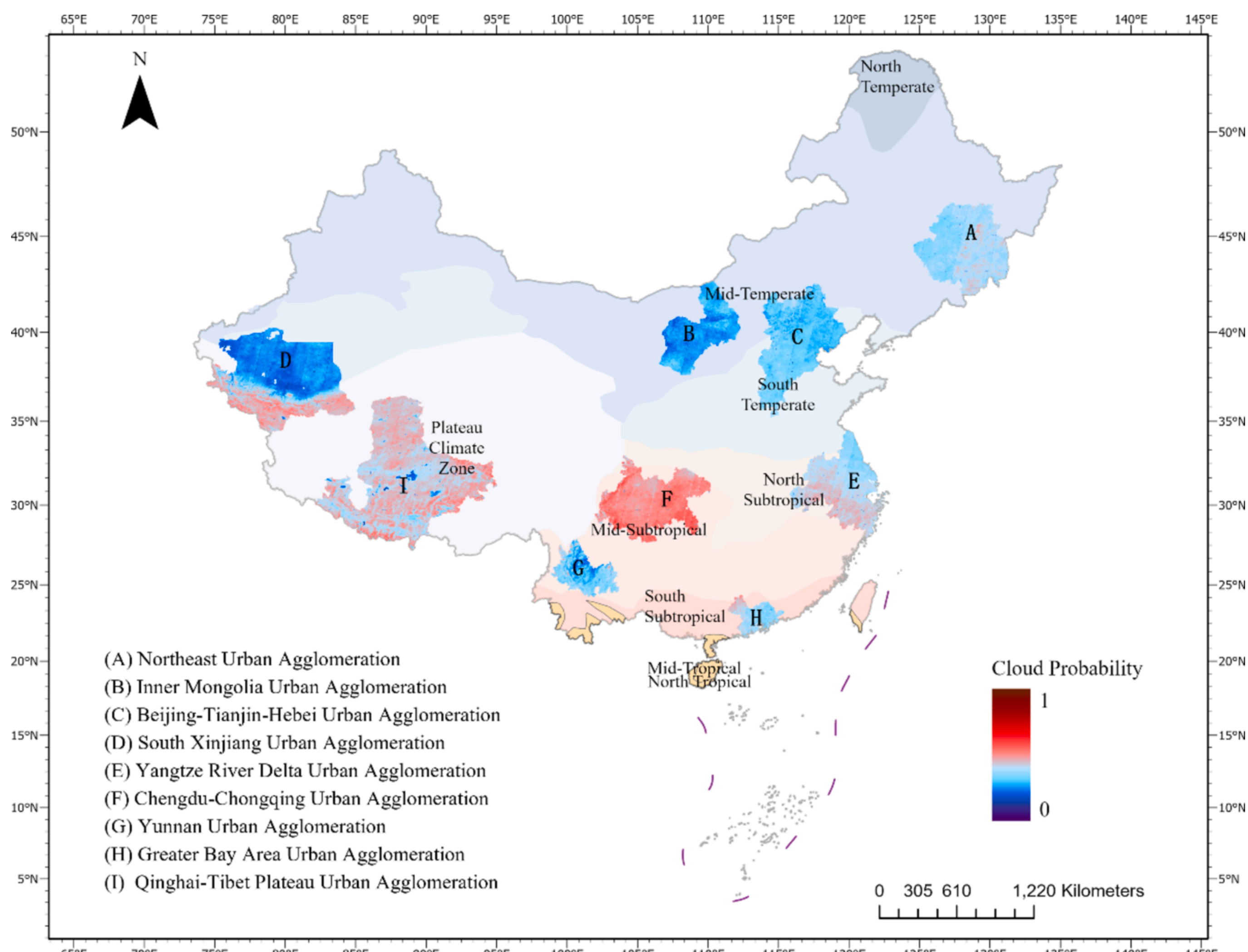

Fig. 3. Cloud probability distribution of urban agglomerations in typical climate regions.

Yunnan urban agglomeration as an example, its cloud coverage probability is relatively low, appearing in dark blue and light blue. In contrast, the Northeast urban agglomeration in the temperate region shows higher cloud coverage on its eastern side, appearing in red. However, the Inner Mongolia urban agglomeration, which also belongs to the mid-temperate zone, exhibits a lower cloud coverage rate. It is worth noting that the Qinghai-Tibet Plateau urban agglomeration, as a plateau climate region, also displays a relatively high cloud coverage rate.

When focusing on urban agglomerations within different climate zones across China, it can be seen that there are huge spatial differences in cloud distribution above the urban agglomerations. Even urban agglomerations with similar geographical locations may present completely different cloud cover situations. Although China is divided into multiple climate zones, urban agglomerations within the same climate zone may still exhibit significant differences. Overall, although climate zoning reflects some cloud distribution characteristics, for example, most temperate regions have relatively low cloud probability while most subtropical regions have relatively high cloud probability, there are still some situations that cannot be explained by a single climatic factor alone. For instance, the Yunnan urban agglomeration, despite being located in the subtropical climate zone, shows low cloud coverage probability. On one hand, Yunnan's location in the monsoon climate zone and the Indian Ocean water vapor channel path results in high wind speeds, hindering atmospheric moisture accumulation and cloud cluster formation (Chen, 2008). On the other hand, central and northern Yunnan's leeward slope location decreases water vapor content and relative humidity during air descent, making cloud formation difficult (Zhu et al., 2022). Moreover, Yunnan's complex terrain and diverse landforms may also influence cloud formation and distribution (Li et al., 2022b). Therefore, cloud distribution may be influenced by multiple factors, and climate zoning alone is not sufficient to fully summarize cloud distribution characteristics.

Furthermore, it is worth noting that regions at higher latitudes tend to exhibit lower cloud coverage rates. For example, compared to the Yangtze River Delta urban agglomeration and the Chengdu-Chongqing urban agglomeration in the south, the Inner Mongolia urban agglomeration and the Beijing-Tianjin-Hebei urban agglomeration have lower cloud coverage rates. This phenomenon may be the result of a

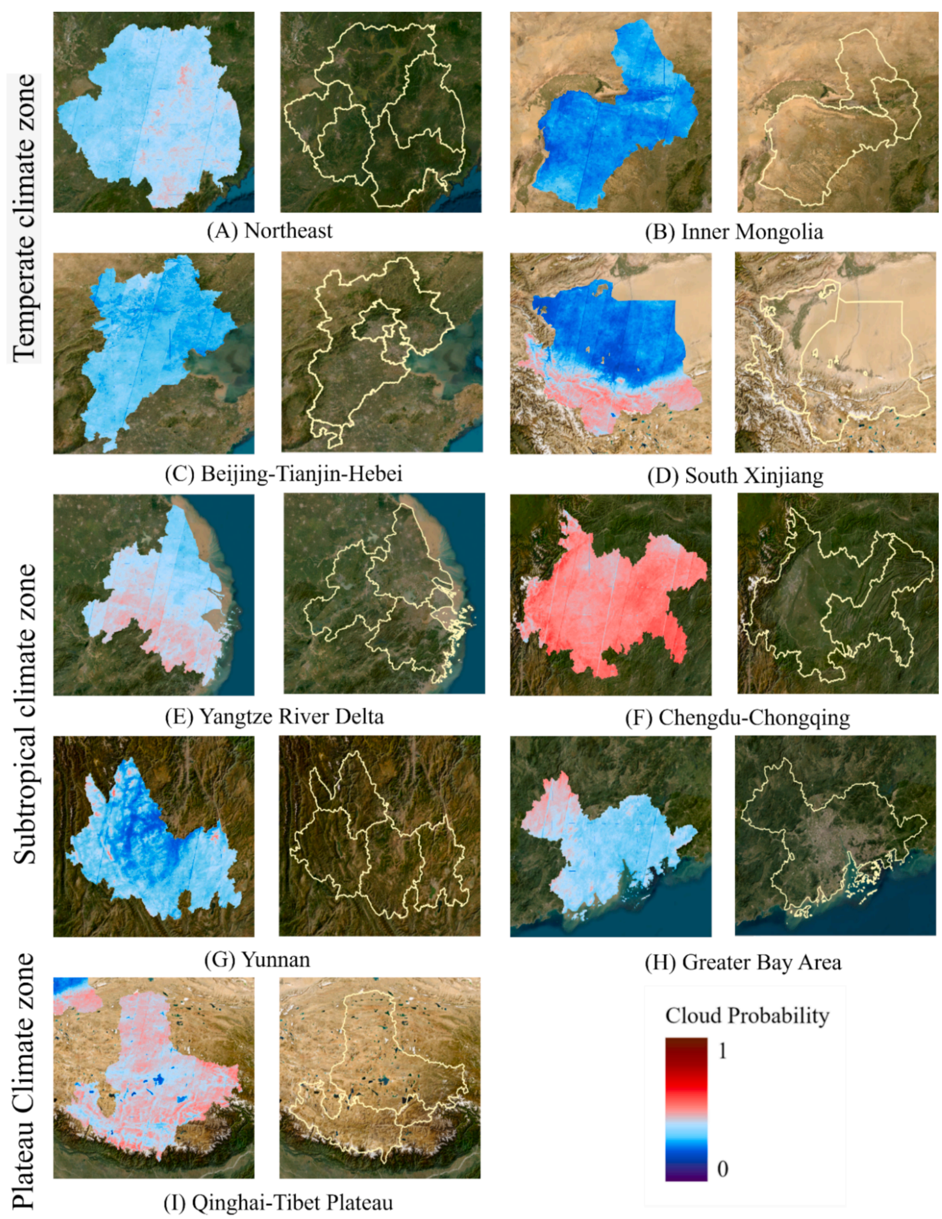

Fig. 4. Urban agglomeration cloud probability detail distribution map and underlying surface cover map.

combination of various factors. Firstly, the cold climate conditions in high-latitude regions lead to lower water vapor content, thereby reducing cloud formation. Secondly, different airflows and wind directions may also influence cloud distribution, and high-latitude regions may be affected by dry airflows. Additionally, high-latitude regions may have large areas of grasslands, deserts, or snow-covered areas, and these surface features may impact local temperatures, humidity, and cloud formation. Low solar radiation angles, surface characteristics, and shorter daylight hours may also influence cloud distribution.

Previous studies have generally suggested that coastal regions have higher cloud coverage due to abundant water vapor supply (Lee et al., 2020). However, our results indicate that in some cases, coastal regions may exhibit lower cloud coverage compared to inland areas, such as in the Yangtze River Delta and the Greater Bay Area urban agglomerations. This phenomenon may be related to the regulating effect of the ocean, which has stable temperature, humidity, and flat terrain, influencing local airflows and water vapor transport.

In summary, the result suggests that the cloud probability of urban agglomeration regions is influenced by a combination of multiple geographical and meteorological factors, resulting in significant differences in cloud distribution between different regions.

3.2. Influence of underlying surface and topographic factors on cloud distribution

To more comprehensively discuss the influence of underlying surface and topographic factors on cloud distribution, Fig. 4 presents cloud probability detail distribution maps for each urban agglomeration, along with corresponding underlying surface cover maps for visual analysis.

For the four urban agglomerations in the temperate climate zone, the Inner Mongolia urban agglomeration has the lowest cloud coverage probability. The terrain here is relatively flat, with land cover types including grasslands, deserts, and water bodies. Some areas may be arid grasslands, and there are desert regions such as the Ordos Desert, one of the deserts in Inner Mongolia. These factors may suppress cloud formation, resulting in lower cloud coverage in the region. The Beijing-Tianjin-Hebei region also exhibits a low cloud coverage probability. The Beijing-Tianjin-Hebei region has relatively flat terrain, with an overall hilly and plain landform, and coexisting land cover types such as cities, farmlands, grasslands, some mountains, and water bodies. It can be observed that highly urbanized areas like Beijing and Tianjin show lower cloud coverage, which might be due to the influence of human activities, buildings, and transportation in urban areas on cloud distribution. In contrast, suburban and farmland areas may have more cloud coverage because large open lands and vegetation contribute to water vapor evaporation and airflow generation, thereby promoting cloud formation. Therefore, the southern open plains and agricultural lands here have relatively higher cloud coverage rates. The Northeast urban agglomeration presents a complex cloud probability distribution, with lower rates in the west and higher rates in the east. This distribution is related to its diverse surface features such as terrain undulations, plains, hills, mountains, and lakes. The west is mainly plains, similar to the Beijing-Tianjin-Hebei region, while the east has higher mountains and hills, higher vegetation coverage, which may influence air movement and humidity distribution, thereby affecting cloud formation. The Southern Xinjiang urban agglomeration shows a huge difference in cloud distribution, with extremely low cloud coverage in the northern desert region and significantly higher cloud probability in the southern plateau region. Despite the adjacent geographical locations in the north and south, they exhibit vastly different cloud distributions. This is due to significant differences in topography and altitude. The northern region is a desert zone with relatively flat terrain, dry climate, and less precipitation, and the air is relatively lacking in water vapor. The southern plateau region has high altitude, complex terrain, frequent air rising and falling, which is conducive to water vapor condensation into clouds. Moreover, the southern plateau region has a relatively humid climate, with water bodies such as lakes, rivers, and alpine glaciers, as well as high mountain vegetation cover, which helps to provide more water vapor and promote cloud formation. The formation of cloud distribution differences in the Southern Xinjiang urban agglomeration is the result of the combined effects of multiple geographical and meteorological factors. Topography, altitude, climate, surface characteristics, and other factors are intertwined, forming the pattern of north–south cloud distribution differences. The Qinghai-Tibet Plateau urban agglomeration, located in the plateau climate zone, is in the same plateau as the southern part of Southern Xinjiang, and also exhibits similar high cloud coverage probability spatial distribution characteristics for the same reasons. In the cloud distribution map of the Qinghai-Tibet Plateau urban agglomeration, it can be clearly observed that the cloud probability above lakes is far lower than the surrounding areas, and the location of lakes can be determined even from the cloud map.

The urban agglomerations in the subtropical climate zone also exhibit different cloud distribution characteristics. The most obvious contrast is between the adjacent Yunnan urban agglomeration and the Chengdu-Chongqing urban agglomeration. The former shows medium to low cloud probability, while the latter has the highest cloud probability among all urban agglomerations, reaching 62.18 %. This difference may be related to various factors such as climate, topography, and land cover. The Yunnan region has a greater terrain undulation, with diverse landforms including high mountains, plateaus, and valleys, and the vegetation coverage is not as high as the Chengdu-Chongqing urban agglomeration. The complexity of its terrain may influence the flow and rise of air currents, thereby affecting cloud formation. In contrast, the Chengdu-Chongqing urban agglomeration is located within the Sichuan Basin, belonging to the subtropical monsoon climate, with relatively humid climate characteristics. Moreover, the Sichuan Basin has relatively flat terrain and low altitude, making it easy for humid air to gather here, which is conducive to cloud formation. Additionally, the Chengdu-Chongqing urban agglomeration is influenced by monsoon airflows, with humid airflows transported from the ocean to the Sichuan Basin, providing sufficient water vapor for cloud formation. Furthermore, the land cover of the Chengdu-Chongqing urban agglomeration is mainly plains and hills, with high vegetation coverage and relatively high surface humidity, which is conducive to water vapor evaporation and release. These factors collectively contribute to the extremely high cloud coverage probability in the region, making it a typical area with frequent cloud and rain cover in China. The Yangtze River Delta urban agglomeration and the Greater Bay Area urban agglomeration show similar distributions, with lower cloud probability in coastal and highly urbanized areas, while the mountainous areas with high vegetation coverage near inland regions exhibit significantly higher cloud probability. Overall, although subtropical regions are considered cloudy and rainy areas in environmental remote sensing observations, in reality, each urban agglomeration often varies due to factors such as topography and land cover, and areas with less clouds and rain are not uncommon.

In general, the underlying surface may have a significant influence on cloud formation. Cloud distribution is not only influenced by climatic and geographical factors but also by the combined effects of various complex surface and meteorological factors, providing a novel perspective for further in-depth research on cloud distribution and its impact on the environment. At the same time, this study also found that not all subtropical regions have high cloud coverage rates, while temperate regions may have more cloud cover, and plateau regions also exhibit high cloud coverage rates. This indicates that the impact of cloud obstruction on remote sensing earth observations should not only be a focus in subtropical regions. In fact, remote sensing surface monitoring activities in most regions are affected by cloud cover. This finding emphasizes the prevalence of cloud cover in remote sensing earth observations and deepens the understanding of cloudy environments in the field of environmental remote sensing observations.

3.3. Quantitative assessment of cloud distribution differences in urban agglomerations

To more accurately analyze the cloud distribution characteristics of urban agglomerations in each climate zone, Fig. 5 presents cloud probability distribution histograms for each urban agglomeration. The horizontal axis represents the cloud probability value, and the vertical axis represents the number of locations with the corresponding cloud probability.

Urban agglomerations in the temperate climate zone show similar histogram distribution shapes but differ in specific values. The cloud probability of the Inner Mongolia urban agglomeration is mainly concentrated around 26 %, with the number of locations with higher or lower cloud probabilities decreasing progressively. The highest cloud probability does not exceed 35 %. In comparison, the cloud probability distribution of the Beijing-Tianjin-Hebei urban agglomeration is higher, with most locations having a cloud probability of around 32 %, fluctuating mainly between 20 % and 40 %, with minimum and maximum cloud probabilities of 10 % and 48 %, respectively. The Northeast urban agglomeration is more prone to cloud cover, with the most common cloud probability value being around 38 %, rarely below 30 %, and the highest value being around 54 %. The huge internal cloud distribution difference in the Southern Xinjiang urban agglomeration is also reflected in the histogram, showing two peaks. The cloud probability in the desert region is generally around 24 %, while the cloud probability in the southern plateau region is mostly as high as 45 %. In the subtropical climate zone, the histogram distribution of the Yunnan urban agglomeration is similar to that of the Beijing-Tianjin-Hebei urban agglomeration, concentrated around 33 %, mainly distributed between 18 % and 51 %. In contrast, the Chengdu-Chongqing urban agglomeration, despite having a similar shape, has a significantly larger cloud probability distribution, with the cloud probability in the vast majority of regions being as high as 48 %, with a maximum of even 63 % and usually not lower than 42 %, indicating that the Chengdu-Chongqing urban agglomeration is one of the regions with high-frequency cloud cover. The Yangtze River Delta urban agglomeration and the Greater Bay Area urban agglomeration also experience frequent cloud cover, with cloud coverage probabilities mostly between 30 % and 48 %, and up to 57 %. At the same time, the plateau region also exhibits high cloud probability, mainly distributed between 36 % and 51 %, with a maximum of 66 %.

Fig. 6 shows the average cloud coverage probability and its standard deviation for each urban agglomeration, intuitively displaying the differences between different urban agglomerations. The Chengdu-Chongqing urban agglomeration ranks first with an average cloud probability value of 50.72 %, closely followed by the Qinghai-Tibet Plateau urban agglomeration with an average cloud probability of 43.27 %. The average cloud probabilities of the Yangtze River Delta urban agglomeration and the Greater Bay Area urban agglomeration do not differ much. Overall, except for the Yunnan urban agglomeration, which has a relatively low average cloud probability of 31.93 %, the urban agglomerations in the subtropical climate zone tend to have slightly higher average cloud probabilities than those in the temperate climate zone. In the temperate climate zone, the Inner Mongolia urban agglomeration has the lowest average cloud probability of 26.32 %, while the Northeast urban agglomeration has the highest average cloud probability of 37.54 %. Although the average cloud probability of the Southern Xinjiang urban agglomeration is relatively high at 31.93 %, its standard deviation is very high, indicating that its cloud probability distribution is the most uneven, consistent with the histogram distribution in Fig. 5.

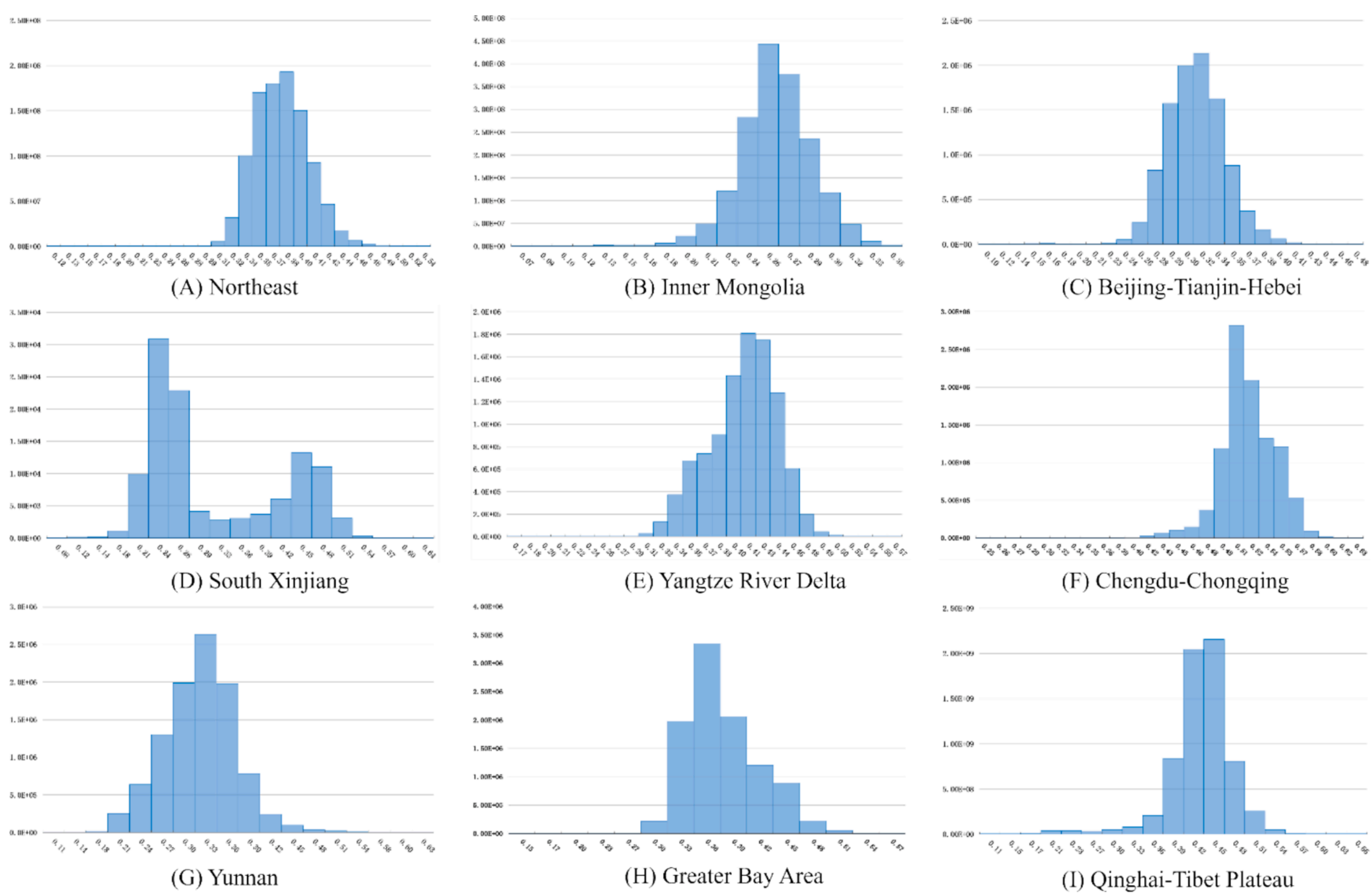

Fig. 5. Cloud probability distribution histogram of urban agglomeration, where the abscissa represents the cloud probability value and the ordinate represents the number of pixels with corresponding cloud probability.

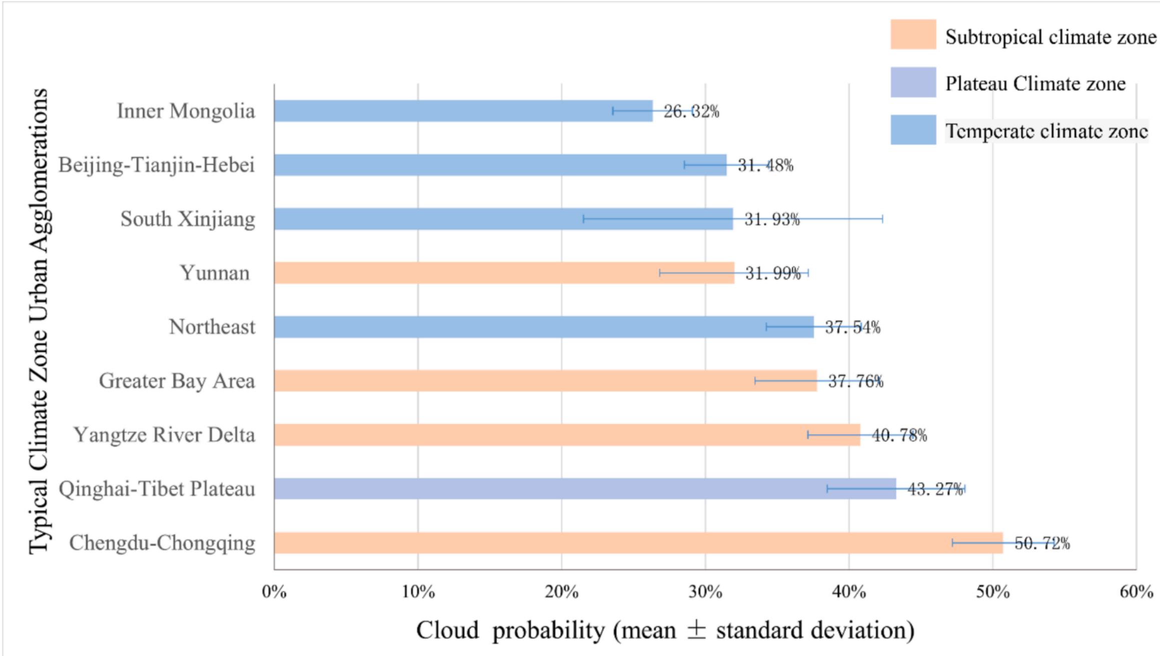

Fig. 6. Average cloud coverage probability and its standard deviation of urban agglomeration.

3.4. Discussion and analysis

This study obtained many findings through qualitative and quantitative assessment analysis of cloud distribution in typical climate zones of China. Firstly, this study noticed significant differences in cloud distribution among urban agglomerations in different climate zones. Although the field of environmental remote sensing observations traditionally considers subtropical regions to have cloudy and rainy characteristics, the results of this study indicate that the cloud cover situation of subtropical urban agglomerations is not entirely consistent. This indicates that cloud distribution is not only influenced by climate zoning but may also be affected by factors such as topography, land cover, and altitude. Urban aerosols, pollutants, and the urban heat island effect may suppress cloud formation, resulting in relatively lower cloud coverage over cities. Vegetation may contribute to water vapor evaporation and airflow generation, thereby promoting cloud formation. These findings are consistent with the mechanistic understanding and observational evidence provided by previous studies (Shepherd et al., 2010; Williams et al., 2015; Xu et al., 2022). This provides some reference for urban planning and construction, as urban layout and vegetation coverage have a certain impact on cloud formation and distribution. Overall, the complexity of cloud distribution is the result of the joint shaping of geographical environments, meteorological conditions, and surface characteristics. Therefore, for the study of cloud distribution characteristics, it is not possible to simply rely on the classification of climate zones, but should comprehensively consider multiple factors such as topography, altitude, land cover, and urbanization degree. It is important to recognize that the relationship between cloud cover and factors such as land surface cover, and urbanization is a complex and multifaceted issue. Qualitative observations in this study suggest potential impacts of the geographical factors. Strict quantification of these relationships would require conducting in-depth quantitative analyses involving extensive data collection, advanced statistical modeling, and consideration of potential confounding factors, which necessitates future research.

The findings of this study have potential applications in enhancing meteorological models, urban solar energy development, and remote sensing observation strategies. The high-resolution cloud probability distribution data can be assimilated into numerical weather prediction models, which incorporates cloud-related parameters such as cloud cover, to improve initial conditions and boundary conditions related to cloud distribution (Forbes et al., 2011; Skamarock, 2008). Furthermore, the insights into the relationship between cloud distribution and geographical factors can inform the development of more accurate parameterization schemes for cloud processes in these models, such as weather research and forecasting model's urban canopy model and land surface model (Chen et al., 2011). Moreover, the revealed cloud distribution characteristics in different climate zones of China provide crucial data support for developing solar energy resources in cities. Cloud cover significantly affects the solar energy potential of urban areas. Therefore, a systematic analysis of cloud distribution patterns over urban areas can provide a scientific basis for estimating the solar radiation intensity in cities. Ji et al. (2024) demonstrated that accurately assessing the solar energy potential of cities is a prerequisite for optimizing the layout of rooftop photovoltaics (Ji et al., 2024). Understanding the patterns of cloud distribution in cities can provide important theoretical support and practical guidance for promoting urban energy transition and building green and low-carbon cities. In addition, the vast majority of urban agglomerations show average cloud probability values exceeding 30 %, highlighting that cloud cover is a common phenomenon in urban environments. While previous research often emphasizes the prevalence of clouds in tropical and subtropical regions, this study reveals spatial variability in cloud cover across different climate zones. This finding can guide the development of adaptive remote sensing observation plans. In regions with high cloud frequency, increasing the temporal resolution of satellite observations, prioritizing radar remote sensing, and integrating multi-source data can help overcome the limitations of optical remote sensing in cloudy conditions. It is essential to consider the unique surface and environmental conditions of different climate zones when formulating effective remote sensing observation strategies.

This study still has certain limitations. Firstly, the Sentinel-2 satellite, which serves as the data source for this study, has a transit time of around 10:30 a.m. local time. Consequently, the analysis may not fully capture the diurnal variations in cloud distribution, as it is based on instantaneous conditions at the satellite's transit time. It is also important to acknowledge that cloud frequency can vary significantly across different seasons, as demonstrated by previous studies (Tian et al., 2021). While the present study focuses on characterizing the overall cloud distribution patterns in urban agglomerations across different climate zones and their potential relationships with geographical factors, we recognize the value of investigating monthly or seasonal cloud probabilities. Future research could build upon the comprehensive understanding of general cloud distribution patterns established in this study to explore the temporal dynamics of cloud distribution at finer scales.

Additionally, the accuracy of the Sentinel-2 cloud product has certain limitations, and air pollution can potentially impact cloud detection. Previous studies have shown that it has an acceptable accuracy, with an overall average accuracy of 86.5 % (Coluzzi et al., 2018). In plateau regions above 2500 m, the accuracy is about 90 % (Wang et al., 2020). For the purpose of exploring cloud probability in this study, the error in cloud identification of a single image is within an acceptable range. Moreover, the QA60 band of Sentinel-2 satellites utilizes multiple spectral bands for cloud detection, and the near-infrared and short-wave infrared bands help mitigate the impact of air pollution on cloud detection, as they are less sensitive to atmospheric aerosols (Zhu et al., 2015). Given that this study utilized all Sentinel-2 images over a three-year time span for statistical analysis of cloud distribution, while severe pollution events at individual time points may affect cloud detection, these short-term pollution events are unlikely to significantly influence the overall cloud probability distribution when considering the entire three-year time scale. Moreover, relevant literature on cloud distribution indicates that its cloud probability results are generally consistent with this study (Li et al., 2022b; Shuai et al., 2022; Wang et al., 2019; Yang et al., 2020b). Considering the ultra-high computational efficiency of the Sentinel-2 cloud product on the Google Earth Engine platform, as well as the requirement for large-scale, high-resolution, and large-data cloud probability calculations, the Sentinel-2 cloud product is still an appropriate choice at present. In the future, the integration of more efficient and accurate cloud detection algorithms and additional data sources to account for the impact of air pollution will further enhance such research.

4. Conclusions

This study provides a comprehensive understanding of the cloud distribution characteristics across urban agglomerations in typical climate zones of China. By utilizing high-resolution Sentinel-2 satellite imagery and the Google Earth Engine platform, we revealed the complex nature of cloud cover patterns and their potential influencing factors. The results challenged the traditional understanding of cloud distribution based on climate zoning alone. Significant variations in cloud probability were observed among urban agglomerations, even within the same climate zone. The analysis suggested that factors such as topography, terrain, and urbanization level may influence cloud distribution patterns. The diversity of cloud cover patterns across different regions is likely the result of complex interactions among these factors. This study contributes to advancing the knowledge of regional climate, environmental management, and remote sensing earth observations by providing new insights into the complex and diverse nature of cloud distribution across urban agglomerations in China. However, it is important to acknowledge the limitations of this study, such as the specific overpass times of Sentinel-2 satellites and potential accuracy issues in cloud detection. Future work could explore the quantitative relationships between cloud distribution and various geographical factors using advanced statistical modeling techniques and integrate cloud probability distribution information into meteorological and remote sensing models. As the complexities of cloud distribution and its driving factors continue to be unraveled, more effective strategies can be developed for monitoring and adapting to changing atmospheric conditions in various contexts.

CRediT authorship contribution statement

Jing Ling: Writing – original draft, Visualization, Validation, Methodology, Data curation. **Rui Liu:** Writing – review & editing, Investigation. **Shan Wei:** Writing – review & editing, Software. **Shaomei Chen:** Software. **Luyan Ji:** Writing – review & editing, Resources. **Yongchao Zhao:** Writing – review & editing, Resources. **Hongsheng Zhang:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was jointly supported by the Research Grants Council (RGC) of Hong Kong, China (HKU27602020, HKU17613022); the National Natural Science Foundation of China (42071390), China; the Shenzhen Science and Technology Program (JCYJ20210324124013037), Shenzhen, China; and the Seed Funding for Strategic Interdisciplinary Research Scheme of The University of Hong Kong, Hong Kong, China.

Data availability

Data will be made available on request.

References

Bessho, K., et al., 2016. An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites. J. Meteorolog. Soc. Jpn. 94, 151–183.
Chen, J., 2008. The analysis of regional climate change features over Yunnan in recent 50 years. Prog. Geogr. 27, 19–26.
Chen, F., et al., 2011. The integrated WRF/urban modeling system: development, evaluation, and applications to urban environmental problems.
Coluzzi, R., et al., 2018. A first assessment of the Sentinel-2 Level 1-C cloud mask product to support informed surface analyses. Remote Sens. Environ. 217, 426–443.
Forbes, R.M., et al., 2011. A new prognostic bulk microphysics scheme for the IFS. ECMWF Reading, UK.
Fu, Y., et al., 2020. Land-surface processes and summer-cloud-precipitation characteristics in the Tibetan Plateau and their effects on downstream weather: a review and perspective. Natl. Sci. Rev. 7, 500–515.
Hagihara, Y., et al., 2010. Development of a combined CloudSat-CALIPSO cloud mask to show global cloud distribution. J. Geophys. Res.: Atmos. 115.
Institute of Geographic Sciences and Natural Resources, C.A.o.S., 2023. China Climate Zoning Data. https://www.resdc.cn/data.aspx?DATAID=243.
Institute, C.D., 2022. Planning scope data of China's top ten urban agglomerations. https://mp.weixin.qq.com/s/KsNBjo3Ehym9P-rbpvmWDQ.
Ji, N., et al., 2024. An urban-scale spatiotemporal optimization of rooftop photovoltaic charging of electric vehicles. Urban Inf. 3, 4.
Jiang, S., et al., 2022. Distinct response of near surface air temperature to clouds in North China. Atmos. Sci. Lett. 23, e1128.
Lee, S., et al., 2020. Comparative analysis of the cloud behavior over inland and coastal regions within single climate characteristics. Atmos. 11, 1316.
Leena, P., et al., 2022. Influence of aerosol physico-chemical properties on cloud microphysical parameters perceived using in-situ high altitude observations. Atmos. Res. 271, 106111.
Li, J., et al., 2022a. An automatic cloud detection model for Sentinel-2 imagery based on Google Earth Engine. Remote Sens. Lett. 13, 196–206.
Li, S., et al., 2022b. Construction and application of comprehensive drought monitoring model considering the influence of terrain factors: a case study of southwest Yunnan. China. Environ. Sci. Pollut. Res. 29, 72655–72669.
Liang, K., et al., 2017. An integrated convective cloud detection method using FY-2 VISSR data. Atmos. 8, 42.
Ling, J., et al., 2021. Improving urban land cover classification in cloud-prone areas with Polarimetric SAR Images. Remote Sens. 13, 4708.
Ling, J., Zhang, H.S., 2023. WCDL: A Weighted Cloud dictionary learning method for fusing cloud-contaminated optical and SAR images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16, 2931–2941.
Ma, J., et al., 2014. Multiyear satellite and surface observations of cloud fraction over China. J. Geophys. Res.: Atmos. 119, 7655–7666.
Meygret, A., et al., 2009. SENTINEL-2 image quality and level 1 processing. Earth Obs. Syst. XIV. SPIE 98–107.

Norris, J.R., et al., 2016. Evidence for climate change in the satellite cloud record. Nat. 536, 72–75.
Peterson, K.T., et al., 2020. Deep learning-based water quality estimation and anomaly detection using Landsat-8/Sentinel-2 virtual constellation and cloud computing. Gisci. Remote Sens. 57, 510–525.
Shah, V., et al., 2020. Global modeling of cloud water acidity, precipitation acidity, and acid inputs to ecosystems. Atmos. Chem. Phys. 20, 12223–12245.
Shen, H., et al., 2014. An effective thin cloud removal procedure for visible remote sensing images. ISPRS J. Photogramm. Remote Sens. 96, 224–235.
Shen, Y., et al., 2019. Estimating cloud condensation nuclei number concentrations using aerosol optical properties: role of particle number size distribution and parameterization. Atmos. Chem. Phys. 19, 15483–15502.
Shepherd, J.M., 2005. A review of current investigations of urban-induced rainfall and recommendations for the future. Earth Interact 9, 1–27.
Shepherd, J., et al., 2010. Urbanization: Impacts on clouds, precipitation, and lightning. Urban Ecosyst. Ecol. 55, 1–28.
Shuai, Y., et al., 2022. Spatial-Temporal distribution of cloud covers over China during 2001–2020, IOP Conf. Ser.: Earth Environ. Sci. IOP Publishing, 012049.
Singh, M., Glennen, M., 2005. Automated ground-based cloud recognition. Pattern Anal. Appl. 8, 258–271.
Sirui, F., et al., 2020. Tempral and spatial distribution characteristics of clouds with different types over China based on ISCCP data. J. Arid. Metrorol. 38, 213.
Skamarock, W., 2008. A description of the advanced research WRF version 3. NCAR/ TN–.-475+ STR NCAR TECHNICAL NOTE 125.
Stewart, R.E., et al., 1998. Midlatitude cyclonic cloud systems and their features affecting large scales and climate. Rev. Geophys. 36, 245–273.
Tian, J., et al., 2021. Improving the accuracy of spring phenology detection by optimally smoothing satellite vegetation index time series based on local cloud frequency. ISPRS J. Photogramm. Remote Sens. 180, 29–44.
Toll, V., et al., 2019. Weak average liquid-cloud-water response to anthropogenic aerosols. Nat. 572, 51–55.
Wallace, J.M., Hobbs, P.V., 2006. Atmospheric science: an introductory survey. Elsevier.
Wang, M., et al., 2019. Temporal and spatial distributions of cloud in china region in recent 60 Years. Clim. Change Res. Lett. 2, 225–233.
Wang, M., et al., 2020. Comparsions of Image Cloud Detection Effect based on Sentinel-2 Bands/Products. Remote Sens. Technol. Appl. 35, 1167–1177.
Williams, A.P., et al., 2015. Urbanization causes increased cloud base height and decreased fog in coastal Southern California. Geophys. Res. Lett. 42, 1527–1536.
Xu, R., et al., 2022. Contrasting impacts of forests on cloud cover based on satellite observations. Nat. Commun. 13, 670.
Yang, Y., et al., 2020b. Spatiotemporal distributions of cloud properties over China based on Himawari-8 advanced Himawari imager data. Atmos. Res. 240, 104927.
Yang, Y., et al., 2021. Testing accuracy of land cover classification algorithms in the qilian mountains based on gee cloud platform. Remote Sens. 13, 5064.
Yang, J., et al., 2020a. Spatial distribution and impacts of aerosols on clouds under meiyu frontal weather background over central China based on aircraft observations. J. Geophys. Res.: Atmos. 125, e2019JD031915.
Zhang, H., et al., 2014. Seasonal effects of impervious surface estimation in subtropical monsoon regions. Int. J. Digital Earth 7, 746–760.
Zhu, Z., et al., 2015. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. Remote Sens. Environ. 159, 269–277.
Zhu, X., et al., 2021. Characterization of dry-season phenology in tropical forests by reconstructing cloud-free Landsat time series. Remote Sens. 13, 4736.
Zhu, D., et al., 2022. Spatiotemporal variations in daytime and night-time precipitation on the Yunnan-Guizhou Plateau from 1960 to 2017. Atmos. 13, 415.
Zhu, X., Helmer, E.H., 2018. An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions. Remote Sens. Environ. 214, 135–153.
Zhu, Z., Woodcock, C.E., 2014. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: an algorithm designed specifically for monitoring land cover change. Remote Sens. Environ. 152, 217–234.