Thick cloud removal of Landsat time-series using convolutional LSTM with embedded residual modules

Lanxing Wang, Qunming Wang, and Peter M. Atkinson

***Abstract*—Extensive cloud contamination severely hinders the interpretation of optical remote sensing images. Existing cloud removal methods focus primarily on the reconstruction of individual cloudy images, with few studies addressing the reconstruction of cloudy time-series images. Furthermore, current methods tend to prioritize using cloud-free auxiliary images while overlooking valuable information present in the cloudy auxiliary images that are temporally closer to the target cloudy image. In this paper, we proposed a deep network called Res-cLSTM to reconstruct cloudy time-series images. Res-cLSTM processes time-series images sequentially using convolutional LSTM, synthesizing long- and short-term memory streams to match the complex temporal relationships amongst them. Then, Res-cLSTM further decodes the feature maps using a refined residual module with skip connections, resulting in the final output. Simulated and real cloud removal experiments on Landsat 8 OLI time-series data across five different regions demonstrated that Res-cLSTM is an effective cloud removal method, which can produce more accurate predictions than three benchmark approaches. For example, for reconstruction of the cloudy time-series of three simulated cloudy areas, the average CC of the Res-cLSTM prediction is about 0.01, 0.04 and 0.04 larger than that of the second most accurate method (i.e., autoencoder (AE)). As a lightweight network, Res-cLSTM does not require global sampling of training data and can fully exploit the valuable information in the non-cloud regions of cloudy time-series images to facilitate cloud removal. Moreover, Res-cLSTM demonstrates robustness to thin cloud omission and exhibits a faster convergence rate, thus, holds great potential for practical applications requiring real-time processing.**

***Index Terms*—Time-series, cloud removal, thick clouds, deep learning, Landsat 8.**

I. INTRODUCTION

Optical remote sensing images provide important data support for continuous land surface monitoring [1-4], and provide unprecedented opportunities for exploring natural [5-8] and anthropogenically-affected processes [9-13]. However, the number of effective optical satellite sensor images available is considerably less than the number of images measured according to the satellite's stated nominal revisit frequency. The main reason for the data gap is that clouds (and their shadows) obscure the observation scene. Studies have shown that clouds obscure an average of 67% of the Earth's surface at any one time [14]. Cloud detection is often used as a pre-process to identify cloud regions in the image, commonly leading to the generation of cloud masks to aid subsequent reconstruction [15-19]. In reality, clouds vary in thickness, from thin to thick. Thin cloud can be removed with reference to partially visible information under the cloud [20-24]. Thick clouds, however, block the propagation of light across all bands, resulting in a complete loss of land surface information and a greater challenge to recover the information below. This paper focuses on thick cloud removal, and reconstruction of the information under thick clouds based on defined cloud masks.

Due to the mobility of clouds, images acquired on other dates at the same location may contain valid information about the target cloud region, and such images are called auxiliary images. Current mainstream methods for cloud removal often employ temporally auxiliary images, which can be categorized into two groups: those based on a single auxiliary image and those based on multiple temporal auxiliary images [25-28]. The two groups of methods can be further classified into methods utilizing homologous auxiliary data and methods employing heterogeneous auxiliary data. Given the temporal dependence as well as temporal variation, the auxiliary images need to be sufficiently close to the target cloudy image temporally.

The single auxiliary image-based methods construct the mapping relationship between the auxiliary image and the target cloudy image based on the common non-cloud data between them, and utilize the valid data in the auxiliary image to reconstruct the target cloud region [29]. The auxiliary data can be classified into homologous and heterologous sources. The homologous data refer to the temporally close data from the same sensor or satellite series. Considering the complexity of the mapping relationship between the auxiliary image and the target cloudy image, machine learning-based methods have been increasingly investigated for cloud removal [30]. For example, Melak et al. [31] employed an autoencoder (AE) neural network to remove clouds, which consists of fully connected layers and ReLU functions, for training and prediction based on individual pixels or patches. Gao et al. [32] proposed a deep code regression (DCR) model, combined with an autoregressive structure, to reconstruct the target cloud region. In Zi et al. [33], an auxiliary image was initially employed to derive preliminary predictions of the target cloud region using linear regression, which were subsequently refined using a convolutional neural network to achieve the final predictions. Tao et al. [34] sorted the training samples based on texture complexity and used self-paced learning to train a generative adversarial network for cloud removal. The aforementioned methods based on homologous auxiliary image capitalize on the spectral consistency. However, frequent cloud occlusion and inherent satellite revisit periods often render the available homologous auxiliary images temporally distant from the target cloudy image. Consequently, researchers have

This research was supported by the National Natural Science Foundation of China under Grants 42222108 and 42171345. (*Corresponding author: Q. Wang*.)

L. Wang and Q. Wang are with the College of Surveying and Geo-Informatics, Tongji University, 1239 Siping Road, Shanghai 200092, China (e-mail: wqm11111@126.com).

P.M. Atkinson is with the Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YR, UK and also with Geography and Environment, University of Southampton, Highfield, Southampton SO17 1BJ, UK.

explored the use of Synthetic Aperture Radar (SAR) images, which can penetrate clouds, as a source of auxiliary images (i.e., heterologous auxiliary data) [35-39]. Deep learning methods have demonstrated competent performance in establishing cross-spectral mapping between SAR and optical images. For example, Xiang et al. [40] proposed a two-step cloud removal method that first converts the SAR image into the corresponding optical image, and subsequently refines the cloud region of the optical image using a cloud-guided fusion network to achieve the final predictions. For methods that utilize only SAR images as auxiliary information, it remains a great challenge to mitigate the effects of severe speckle noise while effectively extracting relevant information from the original data [41-42]. Overall, single auxiliary image-based methods are generally simple to implement as they do not necessitate the collection of extensive training datasets.

Methods based on multi-temporal auxiliary images also involve the use of either homologous or heterologous auxiliary images. Zhu et al. [43] developed a reconstruction model by using the Julian day of the time-series acquisition date as the independent variables and the corresponding pixel values as the dependent variable, enabling the reconstruction of cloud pixels for any given date. To characterize the complex mapping relationship between the time-series, deep learning based-methods have also been developed. Conventional methods focus on fitting explicit relationships based on known information to construct prediction models, making them difficult to handle complex mapping relationships in large-scale cloud-contaminated scenes and time-series data [44-45]. Therefore, deep learning-based methods have been developed rapidly. Chen et al. [46] developed a spatiotemporal information-based neural network known as STnet, which effectively reconstructs a target cloudy image by leveraging spatiotemporal feature fusion modules for spatiotemporal information learning. In addition, deep learning techniques have also been used as post-processing to further optimize the results of tensor completion algorithms. Zhang et al. [47] combined model-driven and data-driven approaches by employing third-order tensor singular value decomposition along with 3D convolutional neural networks (CNNs) to reconstruct time-series cloudy images. Zheng et al. [48] introduced tensor network decomposition and integrated the initial known mask iteration into the optimization process to refine the cloud mask further while using multi-temporal data for cloudy image reconstruction. With respect to the use of heterologous auxiliary data, the powerful learning ability of deep learning models facilitates the fusion of SAR and optical auxiliary images [49]. Specifically, time-series optical images and SAR images can be explored simultaneously for cloud removal of temporally neighboring cloudy images [49]. For cloud removal from Landsat series data, Li et al. [50] first applied a spatiotemporal nonlocal filtering model to fuse a homologous auxiliary image with coarse-resolution time-series images (i.e., MODIS) to obtain a cloud-free image. Then, the cloud-contaminated areas in the target cloudy image were removed using the cloud-free image through nonnegative matrix factorization.

Despite the above progress, existing cloud removal methods exhibit several limitations. Specifically, single auxiliary image-based methods typically do not require a large number of training data, but often necessitate the use of cloud-free auxiliary images [29, 31-32]. For the methods based on single auxiliary images, the accuracy of the final prediction hinges on the validity of the cloud-free auxiliary images temporally closest to the target cloudy image [51]. However, clouds remain prevalent. Due to the limited temporal resolution of satellite sensors, there can be substantial time gaps (sometimes exceeding a year) between the available cloud-free auxiliary images and the target cloudy images. In such cases, a cloudy auxiliary image that is temporally closer to the target cloudy image may provide more valuable information. Moreover, since the cloud-contaminated areas in the auxiliary image are likely to overlap with the target cloud regions, utilizing multi-temporal cloudy auxiliary images could be an effective approach. However, cloud removal methods that rely on time-series images typically require extensive data for model training, creating high demands for both the quantity and quality of training datasets. Without sufficient training data, the models may struggle with generalization [52-54]. Additionally, current research in cloud removal focuses largely on reconstructing individual cloudy images, with comparatively few methods addressing the reconstruction of time-series cloudy images jointly. Yet, time-series data contain critical information regarding temporal changes [55]. By reconstructing time-series images, dynamic phenomena such as landscape alterations, plant growth and urban expansion can be more comprehensively monitored and analyzed [56-57], which cannot be achieved using a single remote sensing image.

In light of the limitations discussed, this paper proposed a Residual block-enhanced convolutional Long Short-Term Memory network (Res-cLSTM) for reconstructing time-series cloudy images. Originally designed for precipitation prediction, convolutional LSTM [58] excels at capturing complex temporal relationships within time-series data. Res-cLSTM operates by sequentially inputting cloudy time-series images into the convolutional LSTM module, which decodes them while fitting the intricate temporal change relationships based on both long-term and short-term memory flows. In the Res-cLSTM model, convolution operations are employed to integrate spatial information. Moreover, the short-circuit connections of the residual module can enhance the generalization ability, reduce the risk of overfitting and stabilize the optimization process [59]. Taking this into consideration, Res-cLSTM then uses a residual module with a short-circuit connection to further refine the description of the feature map and output the final result.

When using cloudy auxiliary images, the cloud areas may overlap with the target cloud areas. In such cases, methods based on single auxiliary image need to combine reconstruction results from multiple auxiliary images in turn to achieve a complete prediction. This accumulates the uncertainty and computational cost step by step. Methods that utilize time-series auxiliary images typically input the data simultaneously into the network, limiting their ability to fully learn the inherent temporal variations. In contrast, Res-cLSTM processes time-series auxiliary images sequentially in chronological order, thereby enhancing its ability to learn temporal variations. Moreover, when time-series are cloudy, the integration of long-term and short-term memory in Res-cLSTM effectively facilitates the complementarity of valuable information in non-cloud areas across the time-series.

In summary, the Res-cLSTM network offers several advantages:

1) Different from traditional cloud removal methods that focus solely on reconstructing individual cloudy images, Res-cLSTM aims to reconstruct the entire cloudy time-series.
2) Res-cLSTM is a lightweight network with a relatively small number of parameters, allowing it to be trained solely on non-cloud data from cloudy time-series images. Thus, Res-cLSTM eliminates the need for extensive training datasets, significantly reducing training time compared to existing models that require a large number of training data.

The remainder of the paper is structured as follows. The proposed Res-cLSTM is detailed in Section II. In Section III, experiments based on simulated and real clouds are conducted to demonstrate the effectiveness of the proposed Res-cLSTM. Section IV further discusses the effectiveness of Res-cLSTM, along with its potential capabilities and limitations. Section V concludes the paper.

II. METHODS

A. Overview of the network design of Res-cLSTM

In this paper, we proposed Res-cLSTM for reconstructing cloud occlusion data in time-series cloudy images. The effective non-cloud data in time-series cloudy images contain valuable information of temporal variation and also spatial structure. Res-cLSTM utilizes valid non-cloud data from time-series cloudy images to train the models, thereby eliminating the need for extensive training datasets. As illustrated in Fig. 1, Res-cLSTM is an end-to-end model composed primarily of two modules: the multi-temporal decoding convolutional LSTM module and the refining residual module. The convolutional LSTM module processes multi-temporal data by inputting them sequentially and capturing the complex temporal relationships amongst them through the integration of long- and short-term memory. The refining residual module serves to further refine the feature maps for the final predictions. This research focuses on six bands of Landsat 8 OLI imagery: blue, green, red, near-infrared (NIR), shortwave infrared 1 (SWIR 1) and shortwave infrared 2 (SWIR 2). Each Landsat 8 OLI image fed into the network includes these six bands. In Fig. 1, $\mathbf{T}_1$, $\mathbf{T}_2$, …, $\mathbf{T}_{p-1}$, $\mathbf{T}_{p+1}$, …, $\mathbf{T}_{n-1}$, $\mathbf{T}_n$ represent the input time-series auxiliary cloudy images, while $\mathbf{T}_p$ denotes the target cloudy image to be predicted. n is the number of images in the time-series, including the target cloudy image itself. Res-cLSTM reconstructs the cloudy images at successive time points, ultimately generating a series of reconstructed cloud-free images. The components and functionalities of each Res-cLSTM module are detailed in Sections II-B and II-C.

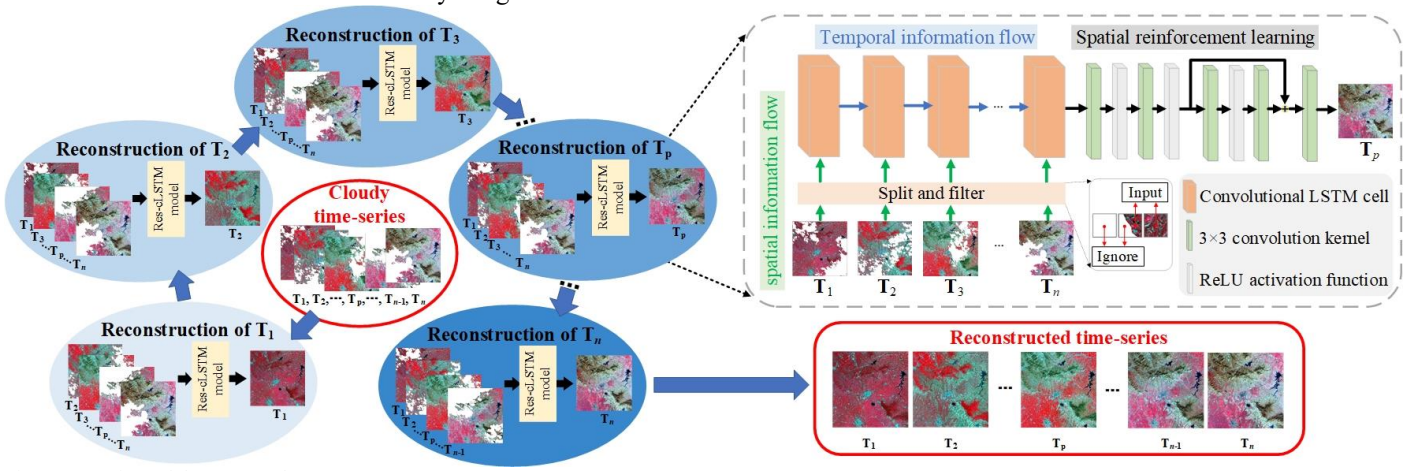

Fig. 1. Overview of the proposed Res-cLSTM.

B. Multi-temporal decoding module

Shi et al. [58] proposed the convolutional LSTM structure based on the fully connected layer-based LSTM, and converted the precipitation prediction problem into the spatiotemporal series prediction problem from the perspective of machine learning. The original LSTM, based on fully connected layers, is a type of neural network characterized by memory capabilities and was designed primarily for processing one-dimensional time-series. However, it struggles to effectively capture spatial features. On this basis, the convolutional LSTM ingeniously replaces the fully connected layers with convolutional layers, enabling it to capture additional spatial structural information from the input data.

Given the satisfactory performance of convolutional LSTM in predicting time-series data, we proposed to utilize it to reconstruct multi-spectral time-series cloudy images. The first component of Res-cLSTM is the convolutional LSTM module, which consists of n-1 convolutional LSTM cells for the n-1 auxiliary time-series images. In this paper, n was set to 5 (i.e., 4 auxiliary images were used). The time-series data are input into the convolutional LSTM module sequentially in chronological order. The number of channels in the hidden layer is configured to match the number of channels in the input layer, denoted as b. In this article, six bands of data were utilized (i.e., blue, green, red, NIR, SWIR 1 and SWIR 2) and, thus b was set to six. Unlike LSTM, which employs fully connected layers, convolutional LSTM utilizes convolutional kernels, enhancing the model's ability to extract spatial information. This modification greatly increases the operational efficiency of the network and significantly reduces computational time. Moreover, by retaining long-term memory and integrating it with short-term memory for each new input, the convolutional

LSTM module effectively captures temporal-varying relationships within the time-series data.

C. Refining residual module

The residual structure was first proposed in [59], which contains a short-circuit connection. Short-circuit connections can facilitate the effective reuse of features extracted from previous layers. By simplifying the optimization process, short-circuit connections often enable the network to converge more rapidly, resulting in more efficient training. Furthermore, short-circuit connections provide greater flexibility in adjusting the mapping between layers during the learning process, which improves the fitting capacity of network. Additionally, they can help mitigate the risk of overfitting, ultimately enhancing the model's generalization ability.

After processing through the convolutional LSTM module, the spatiotemporal information of the time-series data is learned preliminarily. However, we expect the network to further refine the results. Considering the above advantages of the residual structure, we designed a post-processing module based on it, that is, the refining residual module. The output feature map is fed into the refining residual module to further decompose and learn the feature information of the target cloudy images. As illustrated in Fig. 1, the refining residual module comprises five convolutional layers, three of which are activated by the ReLU function. During the model training process, both the images and their corresponding masks are segmented into patches before being fed into the network. A loss function is formulated between the final output and the reference patches based on the L1 norm:

$$Loss = \left\| \left(\mathbf{1} - \mathbf{M}_p\right) \odot \left(f\left(\mathbf{T}_1, \mathbf{T}_2, ..., \mathbf{T}_{p-1}, \mathbf{T}_{p+1}, ..., \mathbf{T}_{n-1}, \mathbf{T}_n\right) - \mathbf{T}_p \right) \right\|_1 \quad (1)$$

where f represents the proposed Res-cLSTM network, ($\mathbf{T}_1$, $\mathbf{T}_2$, …, $\mathbf{T}_{p-1}$, $\mathbf{T}_{p+1}$, …, $\mathbf{T}_{n-1}$, $\mathbf{T}_n$) and $\mathbf{T}_p$ represent the input time-series auxiliary images and the corresponding target cloudy image, respectively. $\mathbf{M}_p$ refers to the known cloud mask corresponding to the target cloudy image, presented as a binary matrix where 0 indicates non-cloud regions and 1 signifies cloud regions. $\mathbf{1}$ is an all-one matrix with the same dimensions as $\mathbf{M}_p$. $\| \|_1$ signifies the L1 norm, and $\odot$ denotes the matrix dot product operation. The Res-cLSTM model was trained based on Eq. (1) to reconstruct sequentially the entire set of time-series cloudy images. In this paper, Res-cLSTM underwent training for 120 epochs, starting with an initial learning rate of 0.01, which halved every 50 epochs.

D. Model training and predicting

Let $\mathbf{T}_p$ be the target cloudy image, $\mathbf{M}_p$ be its cloud mask, and ($\mathbf{T}_1$, $\mathbf{T}_2$, …, $\mathbf{T}_{p-1}$, $\mathbf{T}_{p+1}$, …, $\mathbf{T}_{n-1}$, $\mathbf{T}_n$) be the corresponding time-series auxiliary data. The implementation of the proposed Res-cLSTM method is outlined as follows:

(1) Training:

1) Partitioning $\mathbf{T}_p$, $\mathbf{M}_p$ and ($\mathbf{T}_1$, $\mathbf{T}_2$, …, $\mathbf{T}_{p-1}$, $\mathbf{T}_{p+1}$, …, $\mathbf{T}_{n-1}$, $\mathbf{T}_n$) into $m \times m$ patches: $\mathbf{t}_j^{m \times m \times b}$, $\mathbf{m}_j^{m \times m \times 1}$ and ($\mathbf{t}_1^{m \times m \times b}$, $\mathbf{t}_2^{m \times m \times b}$, ..., $\mathbf{t}_{p-1}^{m \times m \times b}$, $\mathbf{t}_{p+1}^{m \times m \times b}$, …, $\mathbf{t}_{n-1}^{m \times m \times b}$, $\mathbf{t}_n^{m \times m \times b}$)$_j$. Here, b represents the number of spectral bands in each image, and j = 1, 2, ..., J, where J is the total number of patches resulting from the segmentation of each image.

2) Inputting the patches of time-series auxiliary data ($\mathbf{t}_1^{m \times m \times b}$, $\mathbf{t}_2^{m \times m \times b}$, ..., $\mathbf{t}_{p-1}^{m \times m \times b}$, $\mathbf{t}_{p+1}^{m \times m \times b}$, …, $\mathbf{t}_{n-1}^{m \times m \times b}$, $\mathbf{t}_n^{m \times m \times b}$)$_j$ into the Res-cLSTM network. Firstly, the convolutional LSTM module processes the time-series data sequentially in chronological order, yielding an intermediate result $\mathbf{h}_j^{m \times m \times b}$. This intermediate output is then fed into the residual module to generate the final output $\mathbf{p}_j^{m \times m \times b}$ of Res-cLSTM. The loss is computed using $\mathbf{p}_j^{m \times m \times b}$, $\mathbf{t}_j^{m \times m \times b}$ and $\mathbf{m}_j^{m \times m \times 1}$ based on Eq. (1) to guide network training.

(2) Predicting:

Using the trained Res-cLSTM model f obtained from the training steps, the time-series data ($\mathbf{T}_1$, $\mathbf{T}_2$, …, $\mathbf{T}_{p-1}$, $\mathbf{T}_{p+1}$, …, $\mathbf{T}_{n-1}$, $\mathbf{T}_n$) are fed to derive the prediction $\mathbf{P}_p$. Finally, the non-cloud areas of $\mathbf{T}_p$ are concatenated with the reconstructed cloud area from $\mathbf{P}_p$ to yield the final prediction. The whole process is repeated for each cloudy image to produce results ($\mathbf{P}_1$, …, $\mathbf{P}_n$) in turn.

III. Experiments

A. Data and experimental design

This paper focuses on reconstructing the cloud occlusion areas in the cloudy time-series images. A total of 25 Landsat 8 OLI images from five regions (five images for each region) were used for the experiments. Fig. 2 shows the geographical distribution of the five study regions. Fig. 3 shows all the images, each covered by 2000 ×2000 Landsat pixels, corresponding to an area of 60 ×60 km^2. As shown in Fig. 3, Regions 1-3 are based on simulated thick clouds, and the real land surface information under cloud are originally known. Therefore, the real surface reflectance values under the simulated clouds can be used for both quantitative and qualitative evaluation of the reconstructions. Regions 4-5 in Fig. 3 are real cloudy regions, where the accuracy of the reconstructions can be assessed visually only. Regions 1-5 all contain different forms of farmland, mountains and water bodies, as well as urban areas. The types of ground objects are rich and diverse in Regions 1-5, especially the farmland with complex time changes, which poses a great challenge to accurate reconstruction of the time-series images.

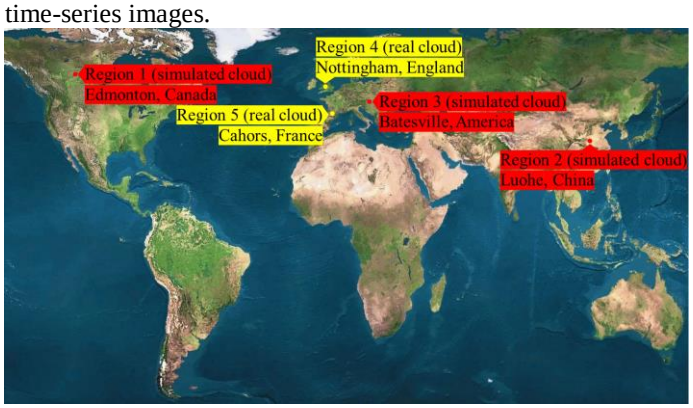

Fig. 2 Geographical locations of the study regions.

B. Comparison with benchmark methods

In this section, the proposed Res-cLSTM was compared with four benchmark methods, namely AE [31], STnet [41], DCR [32] and CR-former [50]. Amongst them, both Res-cLSTM and STnet used time-series data for training and prediction. The difference is that STnet inputs all the time-series auxiliary

images simultaneously, while Res-cLSTM uses the convolutional LSTM module to feed them into the network in chronological order. Notably, AE, DCR and CR-former require that the target cloud area corresponds to a completely cloudless region on the auxiliary images. Therefore, when there is overlap between the cloud area of the auxiliary images and the target image, we stitched the predictions from multiple auxiliary images to obtain the final reconstruction for these three methods. Consequently, the training and prediction time costs increase proportionally with the number of spliced scenes. The reconstructions of one image (with two subareas zoomed on the right) from each region (that is, L11, L25 and L31) is shown in Fig. 4. It can be seen that the five methods can generally produce predictions with good consistency, especially in Region 2. The more detailed reconstructions of each method can be seen from the zoomed subareas in each region, whose locations are marked by yellow and black boxes in Fig. 4. For L11, STnet and Res-cLSTM predicted the black objects in Subarea 1 and the green objects in Subarea 2 more accurately than AE and DCR that use only single auxiliary images. In the two subareas of L25, it can also be seen that Res-cLSTM and STnet reconstructed the distribution of red objects (i.e. vegetation) more accurately. This may be because the cloud region in the temporally closest auxiliary image overlaps with the target cloud region. As a result, AE and DCR must rely on a temporally further auxiliary image that does not contain overlapping cloud regions, significantly diminishing the effectiveness of the auxiliary image. In contrast, the method based on time-series auxiliary data can utilize temporally closer cloudy auxiliary images, which is conducive to obtaining more accurate predictions. For CR-former, the predictions exhibit tonal anomalies in the two subareas of L11, although it accurately predicts the green features in the second subarea. Moreover, CR-former fails to restore the red features in both subareas of L25. In contrast, for L31, the cloud region of the temporally closest auxiliary image does not overlap with its cloud region, and AE, DCR and CR-former can reconstruct the cloud-contaminated information more satisfactorily than STnet in this case. For L31, Res-cLSTM can reconstruct the texture details of ground objects most accurately and its prediction is the closest to the reference image in tone. Overall, amongst the five methods, the proposed Res-cLSTM consistently produces predictions that are the closest in tone and texture to the reference images.

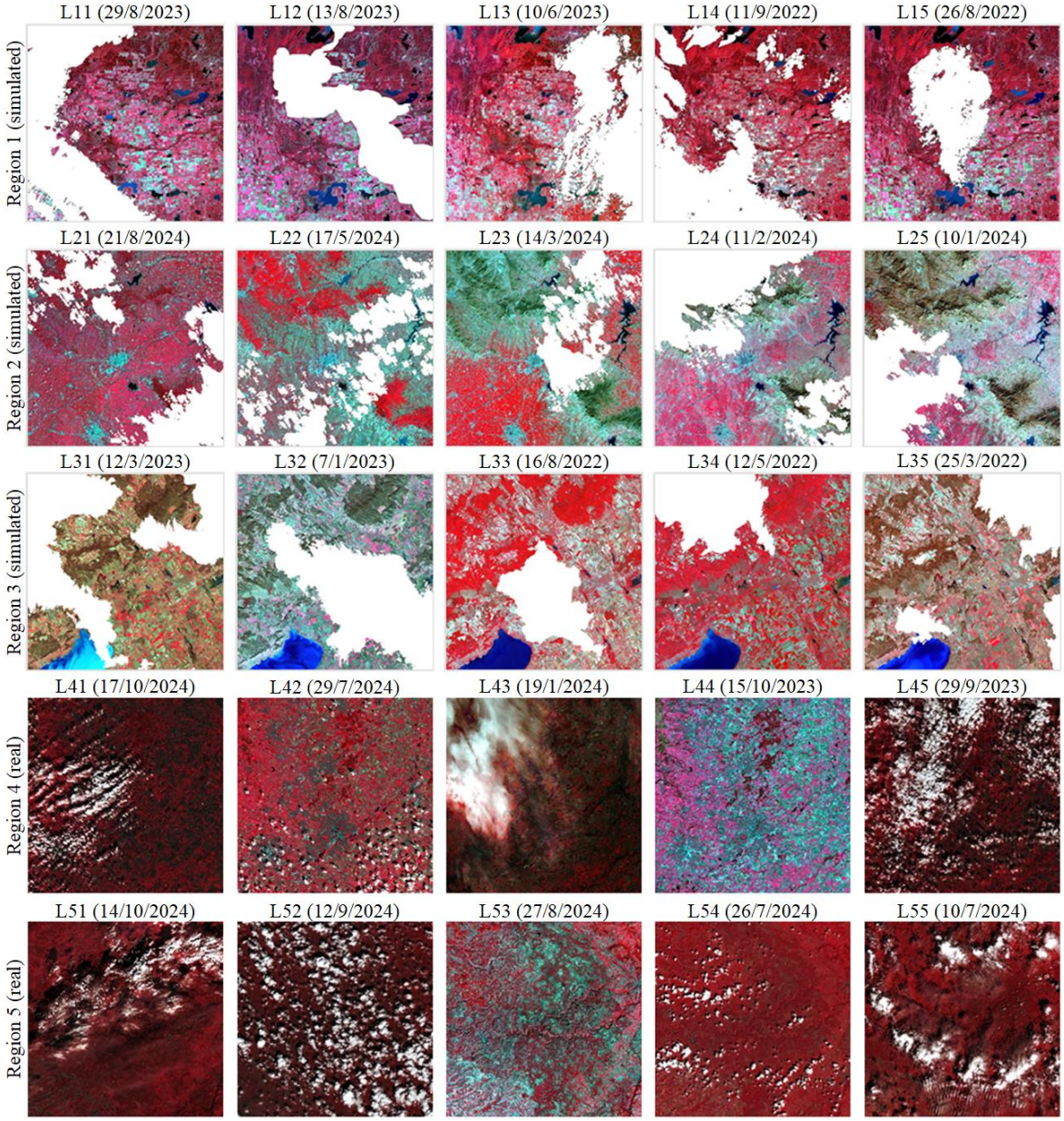

Fig. 3. The simulated and real cloudy time-series images (each with a size of 2000×2000 Landsat pixels) in Regions 1-5. The images are arranged in chronological order from right to left. Above each image is its number and acquisition date (day/month/year).

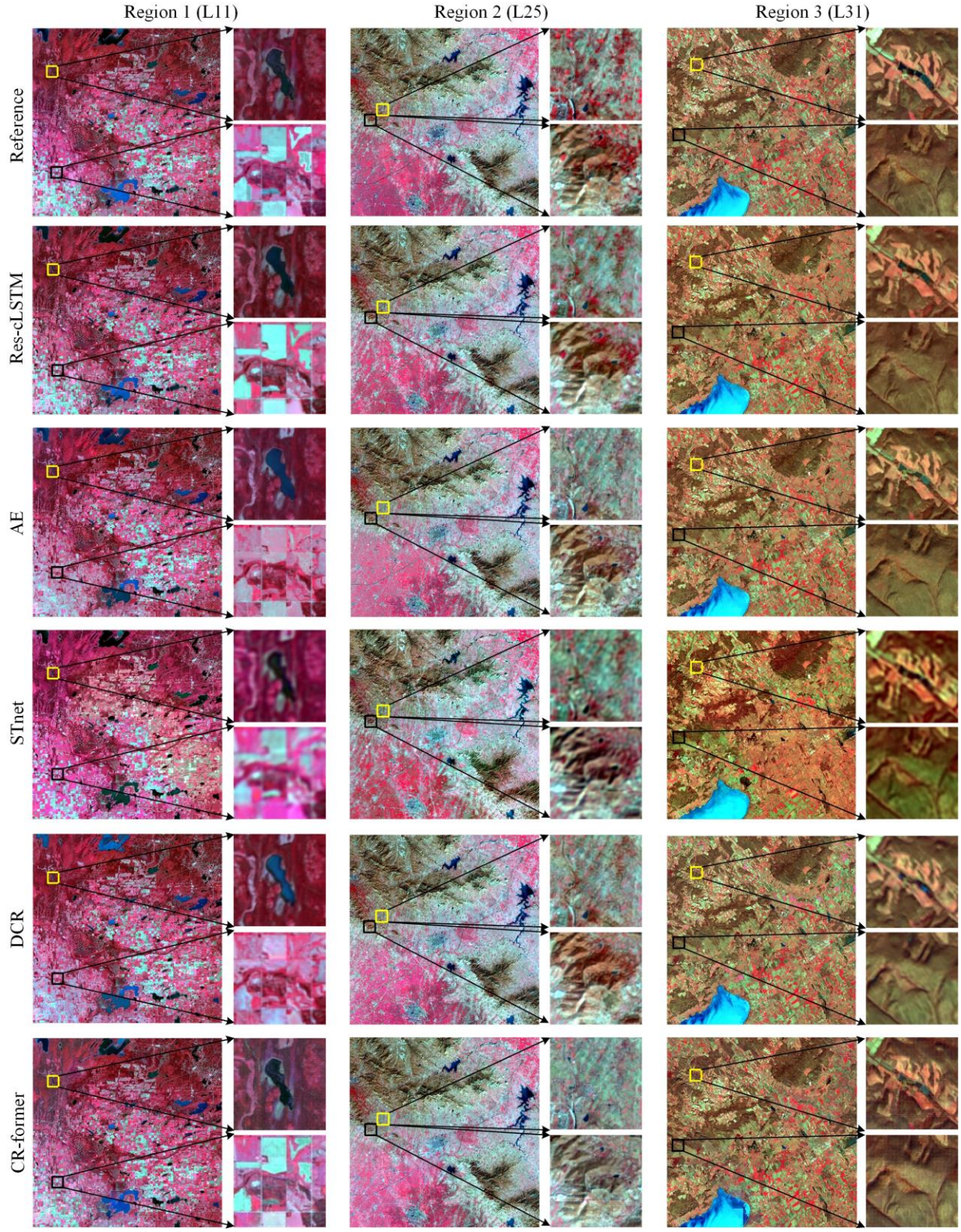

Fig. 4. Cloud removal results of the five methods in Regions 1-3 (with two subareas shown on the right; NIR, red, and green bands as RGB).

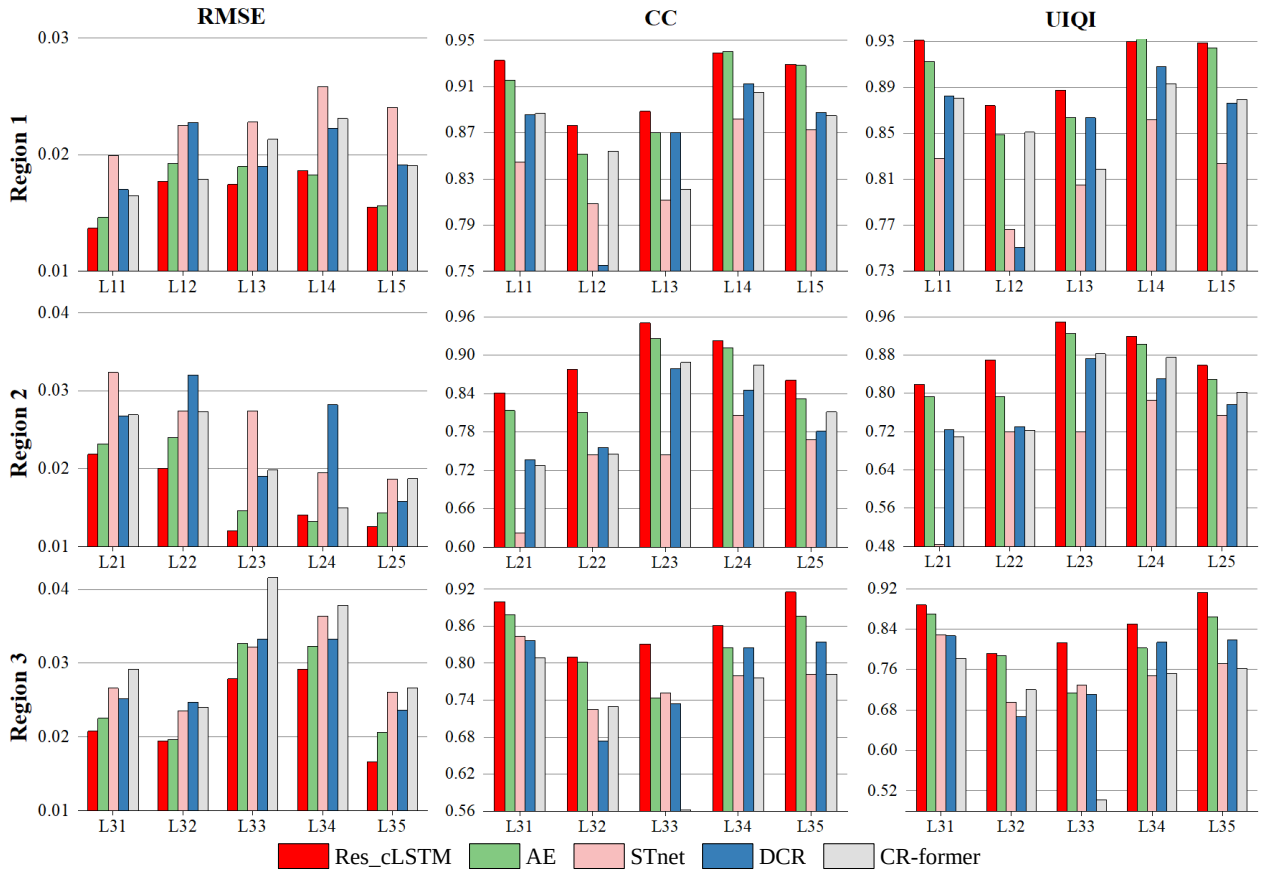

Fig. 5. Accuracies of the five cloud removal methods in Regions 1-3.

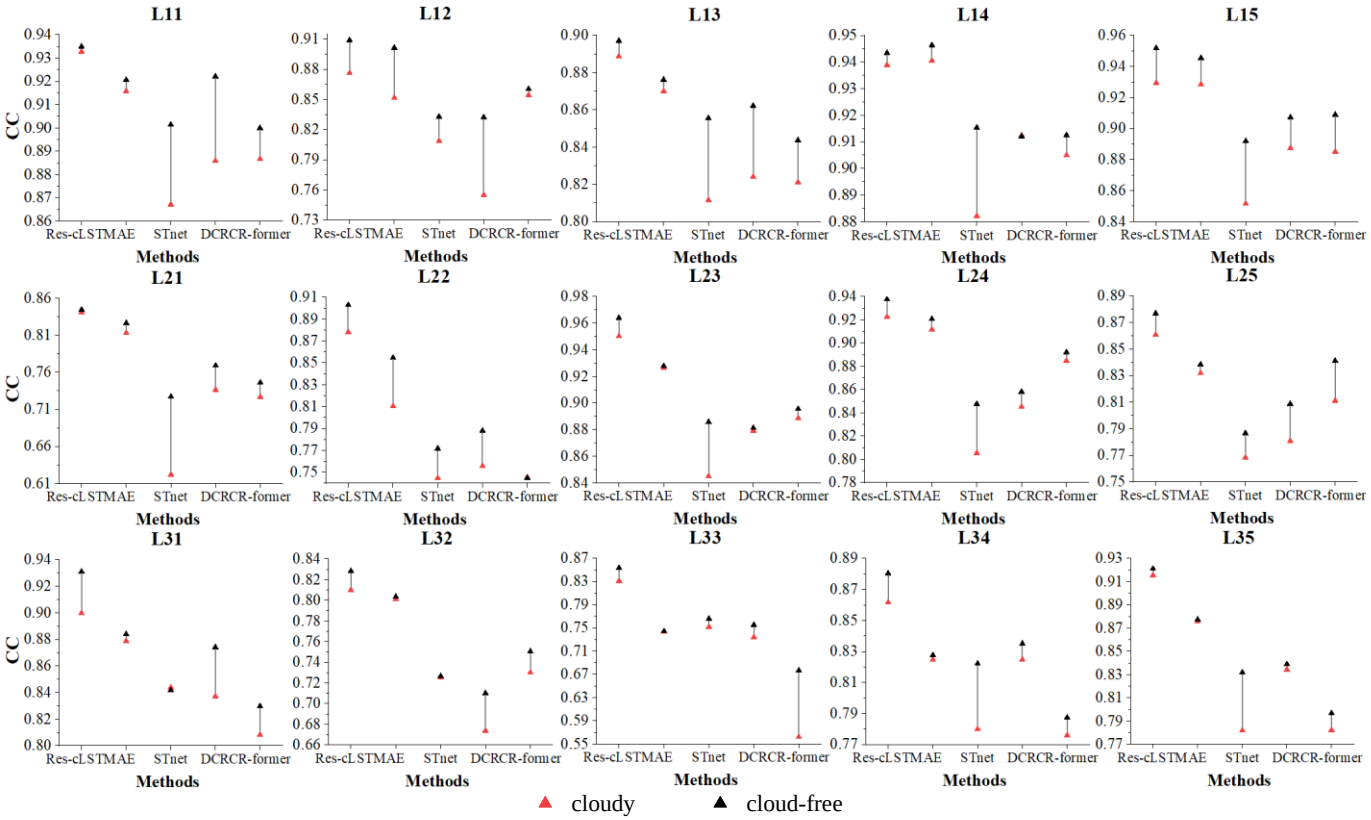

Fig. 6. CCs (averages of the six bands) of the five cloud removal methods using cloudy or cloud-free time-series auxiliary images in Regions 1-3.

In this paper, the root mean square error (RMSE), correlation coefficient (CC) and universal image quality index (UIQI) were used to evaluate quantitatively the prediction of each method. The average accuracy of each prediction in six bands is shown in Fig. 5, which shows that the proposed Res-cLSTM generally produces the most accurate predictions. For example, for L11, the average RMSE of Res-cLSTM prediction is 0.0009, 0.0063, 0.0033 and 0.0028 smaller than that of the AE, STnet, DCR and CR-former predictions, respectively, the average CC is 0.0170, 0.0878, 0.0469 and 0.0461 larger, and the average UIQI is 0.0182, 0.1026, 0.0488 and 0.0506 larger, respectively. In most cases, the AE predictions are second only to that of the Res-cLSTM predictions. Taking Region 2 as an example, the average CCs of the Res-cLSTM predictions of five temporal images were 0.0276, 0.0678, 0.0239, 0.0109 and 0.0290 greater than that of AE predictions, respectively. The results of quantitative evaluation show that Res-cLSTM can stably produce the most accurate predictions.

C. Influence of thick cloud in the auxiliary time-series

Cloud occlusion on the auxiliary image will reduce the available valid data and may affect the accuracies of the final predictions. This section aims to test the effect of cloud data in the auxiliary image on the performance of each method. Specifically, we reconstructed the same target cloudy image using the simulated cloudy auxiliary time-series images (the same as in Section III-B) and their corresponding original cloud-free data, respectively, and compared the prediction accuracies in the two cases.

Fig. 6 shows the average CC of all predictions across six bands, from which it can be seen that the cloud-contaminated data in the auxiliary time-series images generally reduced prediction accuracy noticeably. For STnet, the clouds significantly affect the performances in most cases (e.g., the predictions of L11, L13 and L23 in Fig. 6). The reason may be that STnet input time-series images simultaneously for model training and prediction, which fails to fully utilize the temporal variation information. In contrast, Res-cLSTM processes the time-series data sequentially in chronological order and is capable of transmitting the valid information of the previous image to the next image. The mode of combining long-term and short-term memory enables Res-cLSTM to effectively complement the effective information in the non-cloud areas of different auxiliary images. Thus, Res-cLSTM is relatively less affected by time-series cloud contamination than STnet. In Region 1, the accuracies of DCR predictions are significantly reduced when cloudy time-series images are used in most cases. This may be due to the drastic temporal changes in this region, and the cloud region in the temporally closest auxiliary image overlaps with the target cloud region, resulting in the use of a temporally further auxiliary image. The prediction accuracy of CR-former also shows obvious decreases in some results, such as L13 and L33. In contrast, AE shows relatively stable performance, especially in Region 3, since the cloud region of the temporally closest auxiliary image does not overlap with the target cloud region. Meanwhile, Res-cLSTM is also relatively less affected by clouds in the time-series images and consistently achieves the greatest performance. This leads to the conclusion that Res-cLSTM is more robust to cloud contamination in auxiliary time-series images than the four benchmark methods.

D. Influence of cloud overlap rate

For cloudy time-series images, the accuracy of predictions may be affected when their cloud area overlaps with the target cloud area. In the cloud overlap area, some cloud removal methods may even fail completely (e.g., the method based on single auxiliary image). The purpose of this section is to explore the effect of the overlap rate between the auxiliary image cloud region and the target cloud region. As shown in Fig. 7, different cloud overlap rates were obtained by simultaneously moving a fixed-size simulated cloud region in the direction indicated by the blue arrows. Amongst them, masks 1-4 are the cloud masks (500×500 Landsat pixel) of the four time-series images, respectively, and mask-target is the cloud mask (1000×1000 Landsat pixel) of the target cloudy image. For methods using single auxiliary images (i.e. AE and DCR), the auxiliary images acquired on two different dates were used to reconstruct the same target cloudy image, and then the two results were spliced to obtain the final predictions, to avoid complete failure zones. In this section, the cloud overlap rates were calculated based on all auxiliary time-series images. For example, when the overlap rate is 80%, the proportion of cloud overlap area in each auxiliary image is 20% of the target cloud area. For DCR and AE, they only use the predictions of two auxiliary images to concatenate, and the cloud overlap rate on them is only 40%.

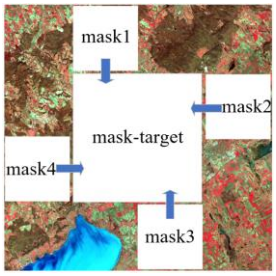

Fig. 7. Schematic diagram of cloud mask position under different cloud overlap rates. The fixed-size masks 1-4 (corresponding to the four auxiliary temporal images) move in the direction indicated by the blue arrow to obtain different overlap rates with the mask of target cloudy image (mask-target).

The results for the first image in each region (i.e., L11, L21 and L31) are shown in Fig. 8. It can be seen that the prediction accuracies of the five methods generally decrease as the cloud overlap rate increases. Amongst them, the accuracies of the STnet predictions fluctuate greatly in all three regions, and decrease significantly in general, which may be because all auxiliary time-series images are inputted simultaneously to reconstruct the target cloud region. The influence of the cloud overlap areas on the DCR predictions in Regions 1 and 3 is relatively small, which may be due to the relatively smooth temporal changes in the two regions. However, in Region 2, the obvious decreases in the DCR prediction accuracies may be due to the large land cover changes in the temporally further auxiliary images. In contrast, AE and Res-cLSTM are less affected by cloud overlap areas, and the changes in the prediction accuracies are relatively stable in the three regions. Moreover, in almost all cases, Res-cLSTM performs more satisfactorily than AE. Thus, it can be concluded that Res-cLSTM can produce more accurate predictions consistently

when there is varying degree of cloud overlap between the auxiliary images and the target cloudy image.

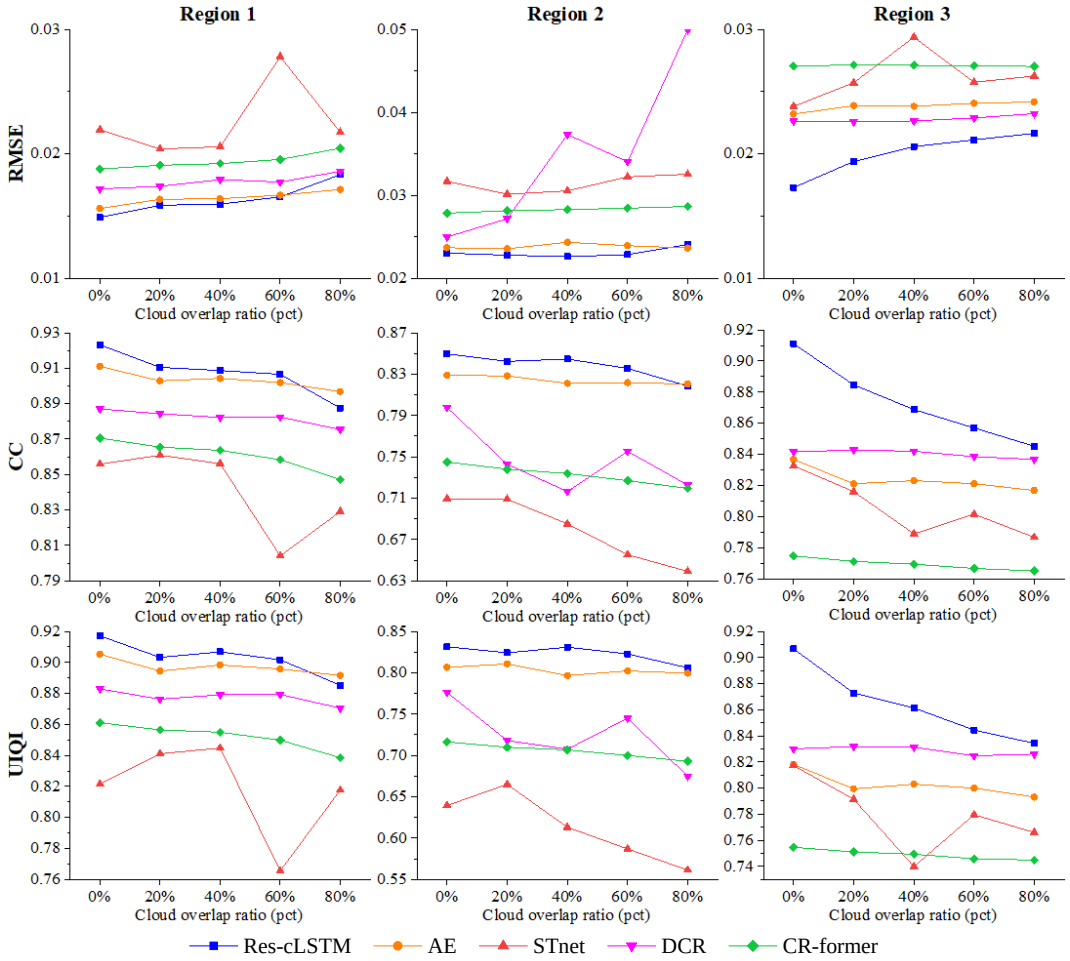

Fig. 8. Accuracies (averages of the six bands) of the five cloud removal methods with different cloud overlap ratio.

E. The impact of thin cloud omission

The previous experiments were all based on simulated cloud masks, which were known perfectly. Practically, cloud detection results commonly contain uncertainties, resulting in omission of thin clouds that are treated as cloud-free data during model training and prediction. In this section, we tested the influence of undetected thin cloud on the methods. Specifically, we used the method proposed by Guo et al. [60] to generate simulated thin clouds by mimicking the real situation of light penetrating clouds. Fig. 9 illustrates the distribution of simulated thin and thick clouds on L34, with the yellow coil indicating the simulated thin clouds. We incorporated varying proportions of these simulated thin clouds into 15 cloudy images across Regions 1-3 (as outlined in Section III-B). The percentages of thin clouds in total cloud cover ranging from 9.23% to 43.33%, as detailed in Table 1.

Fig. 10 presents the accuracies of predictions with and without thin cloud omission for each method. The accuracies present a general decrease across all methods due to thin cloud omission, particularly for STnet, DCR, and CR-former. In contrast, AE and Res-LSTM demonstrated greater robustness against the effects of thin cloud omission. Overall, Res-cLSTM consistently achieved the greatest accuracy, maintaining a relatively low decrease in performance. This suggests that Res-cLSTM offers considerable promise for practical applications, remaining effective even in the presence of undetected thin clouds.

Fig. 9 Distribution of the simulated thick and thin clouds (taking L34 as an example). The yellow line delineates the simulated thin cloud.

Table 1 Ratio of thin clouds to the total cloud cover percentage (x represents the region number).

	Lx1	Lx2	Lx3	Lx4	Lx5
Region 1	9.63%	13.52%	12.69%	11.93%	9.23%
Region 2	17.67%	43.33%	39.32%	25.43%	29.92%
Region 3	9.84%	10.42%	13.69%	20.73%	14.75%

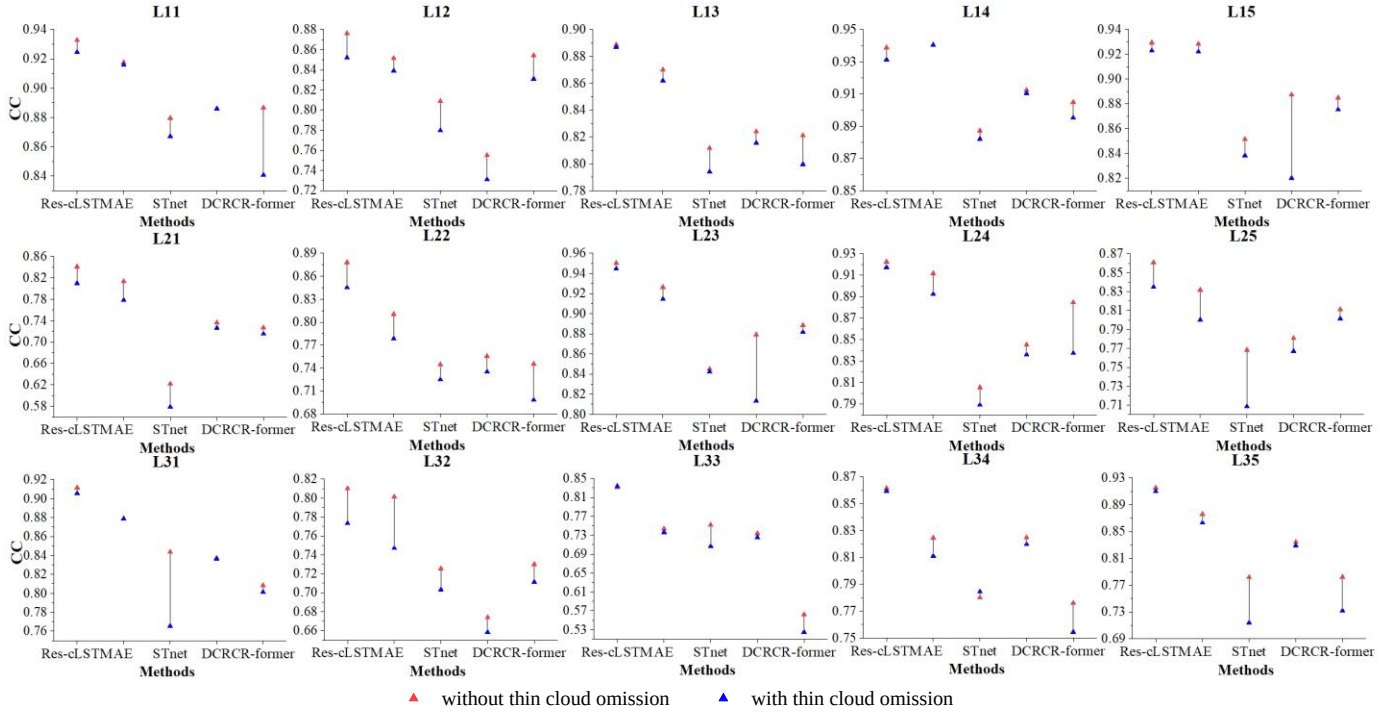

Fig. 10. CCs (averages of the six bands) of the five cloud removal methods based on images with or without thin cloud omission in Regions 1-3.

F. Network convergence rate

The convergence rate is a crucial metric for deep learning networks, as it influences directly their effectiveness in practical applications. In this section, we evaluated the convergence rates of the five methods by analyzing the increase in accuracy over time. The proposed Res-cLSTM was implemented using Python 3.7 on a personal computer equipped with an Intel Core i9@3.70 GHz and an NVIDIA GeForce RTX 3070.

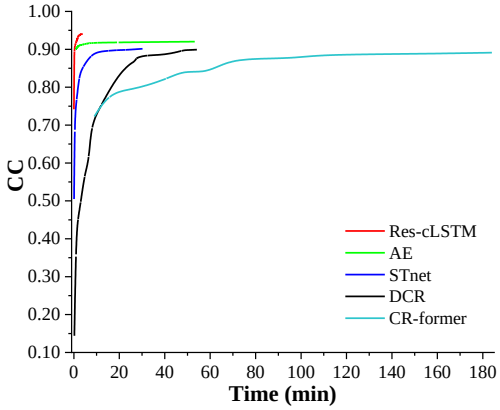

Fig. 11. Changes in CCs (averages of the six bands) of the five cloud removal methods over computing time.

The results are presented in Fig. 11. Each curve begins with the accuracy achieved after training each network for one epoch and ends at the point where the increase in accuracy for each method significantly slows down. Notably, the prediction accuracy of AE reaches 95% of its final level after just one epoch. This can be attributed to its reliance on multi-band information from single pixels for model training and prediction, resulting in a relatively straightforward learning relationship and a limited maximum attainable accuracy. In contrast, DCR and CR-former features a deeper network structure, which restricts its learning rate and prolongs training time per epoch. Both Res-cLSTM and STnet demonstrate comparable and rapid operating speeds (approximately 5-6 seconds per epoch), with their prediction accuracies increasing quickly and significantly at the outset. However, Res-cLSTM ultimately achieves substantially greater prediction accuracy than STnet. These results indicate that Res-cLSTM exhibits a faster convergence rate and is capable of delivering more accurate predictions in a shorter timeframe. Consequently, Res-cLSTM shows considerable potential for development and real-world applications, with its rapid convergence facilitating real-time generation of up-to-date predictions.

G. Experiments on real clouds

In this section, real cloudy images in Regions 4 and 5 were employed to evaluate the performance of the five cloud removal methods. As illustrated in Fig. 3, L44 and L53 are cloud-free images, while the remaining images contain clouds. The predictions of the five methods are presented in Fig. 12. Notably, when the cloud-contaminated area of the target image is excessively large (for example, in L43 and L55, where both have cloud cover exceeding 70%), the DCR and CR-former predictions exhibit significant tonal anomalies. This phenomenon may be attributed to the limited availability of training data, which hampers the effective training of the DCR models. In contrast, predictions generated by Res-cLSTM, AE, and STnet show satisfactory tonal consistency. Fig. 13 presents a detailed view of a subarea from each reconstructed time-series, with their locations indicated by red boxes on the cloud masks in Fig. 12. The subareas for L42 and L54 are cloud-free. It is evident that predictions produced by STnet are somewhat blurry. Compared with STnet and DCR, Res-cLSTM and AE

effectively recover more texture details. In addition, Res-cLSTM can recover effective temporal variation, as characterized by the reconstructed time-series.

IV. Discussion

A. Advantages of the proposed Res-cLSTM

In this paper, we proposed a deep network model called Res-cLSTM for reconstructing missing data in time-series cloudy images. The effectiveness of Res-cLSTM was demonstrated in Section III, where it achieves more accurate cloudy time-series data reconstruction results compared to several benchmark methods. Unlike traditional methods that require cloud-free auxiliary images, Res-cLSTM effectively utilizes the available non-cloud data in cloudy time-series images. Moreover, while most existing methods focus on reconstructing single cloudy images, Res-cLSTM excels at efficiently reconstructing time-series cloudy images jointly, enabling more effective monitoring of dynamic land surface changes and offering greater practical significance. Furthermore, Res-cLSTM is a relatively lightweight network that can be trained solely using non-cloud data from time-series cloudy images, eliminating the need to gather extensive training datasets globally and greatly simplifying data collection. Additionally, compared with the three benchmark methods, Res-cLSTM exhibits faster convergence and provides more accurate predictions in a shorter time frame, showcasing significant practical application potential, particularly in scenarios that require real-time processing.

B. Generalization ability of Res-cLSTM

In this paper, the Res-cLSTM network leverages a known cloud mask to exclude cloud region data from the input cloudy images during the training process. Notably, Res-cLSTM has potential applicability in handling other types of missing data, such as quantitative remote sensing data. Employing the convolutional LSTM originally designed for precipitation prediction, Res-cLSTM is expected to reconstruct datasets with clear temporal relationships, including Land Surface Temperature (LST), Fractional Vegetation Cover (FVC) and nighttime lights. Given that Res-cLSTM was applied to multi-spectral time-series data in this research, the rich spectral information may also be a major reason to promote accurate predictions. For quantitative remote sensing data that may contain limited spectral information, integrating auxiliary data from other sources could be a more effective approach. For example, when reconstructing missing LST data, the incorporation of elevation and geographical information (such as latitude and longitude) could provide valuable context. However, how to optimally leverage this effective information requires further experimental investigation. Additionally, Res-cLSTM does not impose stringent requirements on data resolution, allowing for flexibility in its application. By utilizing the adaptable nature of convolution operations across different spatial scales, the Res-cLSTM model can be readily developed and employed to reconstruct data at various spatial resolutions.

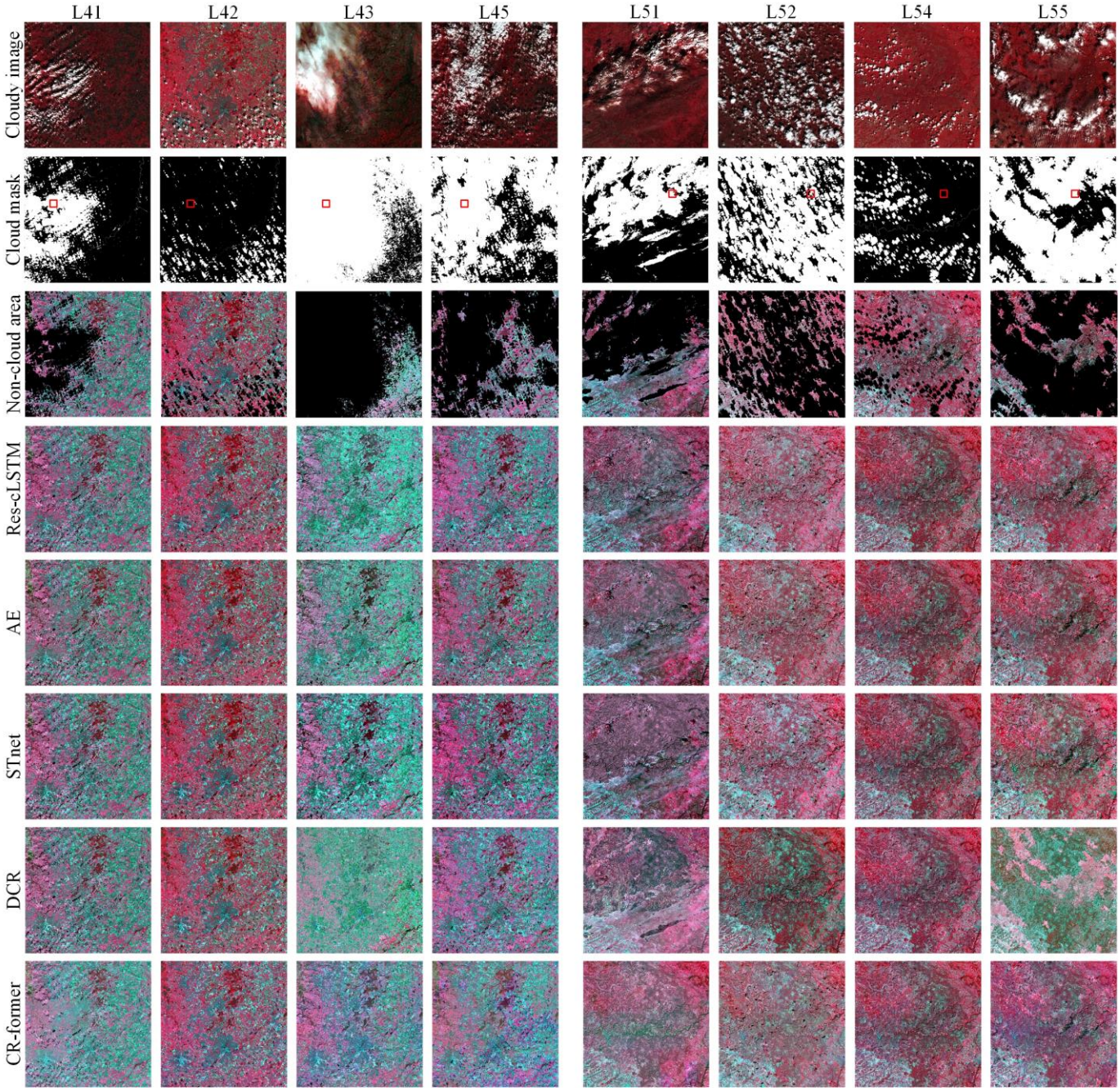

Fig. 12. Cloud removal results of the five methods in Regions 4 and 5 (NIR, red, and green bands as RGB).

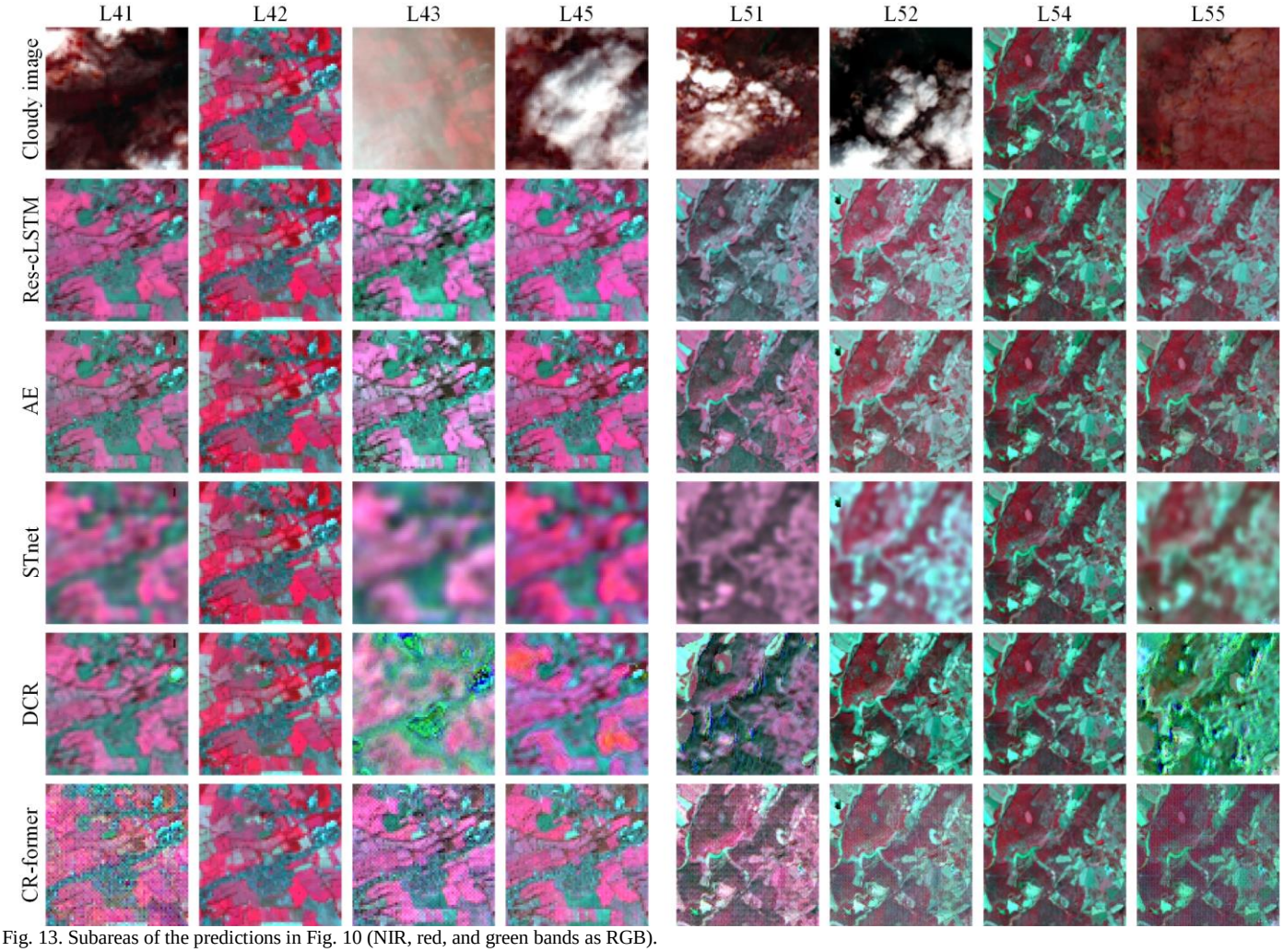

Fig. 13. Subareas of the predictions in Fig. 10 (NIR, red, and green bands as RGB).

C. *Differentiated removal of thick and thin clouds*

The proposed Res-cLSTM removes clouds from cloudy time-series images using known cloud masks for both simulated and real cloudy images, without differentiating between thick and thin clouds. However, thin clouds often allow some underlying land surface information to be visible, and treating them the same as thick clouds does not fully leverage this valuable information. In practical applications, cloud masks are generated using specialized cloud detection methods [61-62]. However, thick and thin clouds frequently coexist and can be difficult to distinguish. Given the powerful learning capabilities of neural networks, there is potential to develop models that can differentiate between thin and thick clouds during the removal process. For example, by utilizing the strong learning capabilities of deep networks, specific loss functions can be designed to guide the automatic learning of thick and thin cloud features within the network, enabling targeted reconstruction of both types simultaneously.

D. *Computational cost*

Fig. 14 illustrates the prediction accuracy and corresponding training time of the Res-cLSTM model for a target cloudy image with varying numbers of auxiliary images. The target cloudy images for the three regions are L11, L21, and L31. In Fig. 14, CC1-CC3 represent the average CC of the reconstructions for L11, L21 and L31 across six bands, while Time1-Time3 indicate the corresponding time consumption. Notably, when the number of images exceeds 4, the time consumption increases sharply, but the prediction accuracy decreases in Regions 2 and 3. In Region 1, the differences in prediction accuracy with varying numbers of auxiliary images are slight, likely due to the small temporal variations amongst the time-series images. Consequently, four auxiliary images were used in this paper.

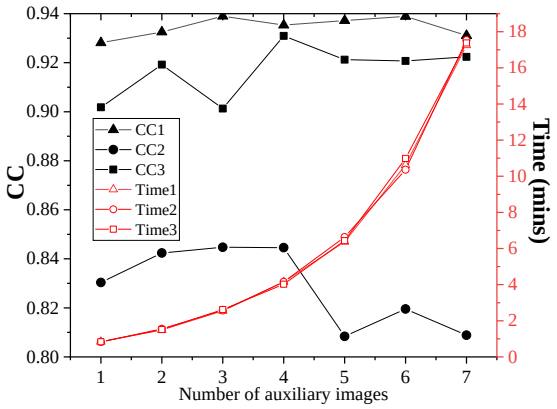

Fig. 14. Computational cost and accuracy of Res-cLSTM under different numbers of auxiliary images in Regions 1-3.

E. Uncertainty in Res-cLSTM

The proposed Res-cLSTM reconstructs time-series cloudy images by modeling the temporal relationships between them. In practice, the acquisition dates of time-series data are random. In regions consistently covered by clouds, Res-cLSTM is constrained by the scarcity of auxiliary information. Moreover, when adjacent images are captured in different seasons, significant changes in land surface characteristics may occur, increasing the risk of model failure. Given that Res-cLSTM is flexible in the number of time-series data it utilizes, future implementations could incorporate embedding modules or formulae to assess the severity of the temporal changes. This enhancement may increase the accuracy of predictions for other time-series images by eliminating those with large temporal changes. Subsequently, the reconstructed time-series images could be used to target and reconstruct the eliminated cloudy images that exhibit large temporal changes. Moreover, since Res-cLSTM reconstructs time-series cloudy images based on known cloud masks, the accuracy of its outputs may be influenced by the precision of the mask in practical applications. Integrating interactive modules capable of learning and refining the cloud masks during Res-cLSTM training could provide an alternative approach to enhance predictions while reducing reliance on known cloud masks. Additionally, auxiliary images from other sensors (such as SAR) could be integrated to further optimize prediction.

F. Comparison with non-deep learning-based methods

In this section, we compared the proposed Res-cLSTM with three traditional (i.e., non-deep learning-based) cloud removal methods that utilize time-series auxiliary data: (1) a modified neighborhood similar pixel interpolator approach for removing thick clouds based on Landsat time-series auxiliary data, abbreviated as multi-NSPI [44]; (2) a missing observation prediction based on spectral-temporal metrics (MOPSTM) gap-filling method [45]; (3) an algorithm for generating synthetic Landsat images based on all available Landsat data, referred to as AGSL [43]. We evaluated the performance of each method using both cloudy and cloud-free auxiliary images, with the results presented in Tables 2 and 3.

As shown in Table 2, the predictions produced by Res-cLSTM exhibit obviously larger accuracy compared to the three non-deep learning methods. This discrepancy may stem from the reliance of Multi-NSPI and MOPSTM on the spatial neighborhood information of the target cloud pixels, which poses challenges for reconstructing large-scale missing areas, such as the extensive cloud occlusion present in the 2000 ×2000 Landsat pixel experimental region here. Furthermore, our experiments utilized only four auxiliary temporal images, which falls short of the minimum recommended by AGSL (at least eight scenes). This limitation could be a significant factor affecting the accuracy of AGSL. Additionally, traditional approaches often emphasize establishing explicit relationships based on spatially and temporally adjacent information, which may lack sufficient descriptive power for addressing the complex mapping relationships in real-world scenarios.

Moreover, as indicated in Table 3, the prediction accuracy of the three conventional methods declines sharply when the time-series auxiliary images are affected by clouds. This decline is primarily due to the further reduction in available effective data, which exacerbates the limitations mentioned above. In contrast, Res-cLSTM consistently delivers more accurate predictions, regardless of whether the time-series data are cloudy or cloud-free, demonstrating superior robustness against cloud occlusion in the auxiliary data.

Table 2 Accuracies of the four cloud removal methods (Res-cLSTM and non-deep learning-based) produced using cloud-free auxiliary time-series data (Region 1 as an example; the value in bold means the most accurate result in each case)

		Res-cLSTM	Multi-NSPI	MOPSTM	AGSL
L11	RMSE	**0.0131**	0.0162	0.0420	0.0218
	CC	**0.9349**	0.8975	0.8702	0.8510
	UIQI	**0.9337**	0.8959	0.8131	0.8442
L12	RMSE	**0.0149**	0.0179	0.0557	0.0341
	CC	**0.9089**	0.8587	0.7116	0.7553
	UIQI	**0.9076**	0.8556	0.5999	0.7073
L13	RMSE	**0.0170**	0.0199	0.0294	0.0392
	CC	**0.8971**	0.8554	0.8137	0.8076
	UIQI	**0.8958**	0.8504	0.7855	0.7352
L14	RMSE	**0.0176**	0.0229	0.0388	0.0278
	CC	**0.9435**	0.9039	0.8921	0.8869
	UIQI	**0.9408**	0.9006	0.8255	0.8465
L15	RMSE	**0.0162**	0.0174	0.0381	0.0300
	CC	**0.9230**	0.9131	0.8604	0.8514
	UIQI	**0.9216**	0.9100	0.8279	0.8191

Table 3 Accuracies of the four cloud removal methods (Res-cLSTM and non-deep learning-based) produced using cloudy auxiliary time-series data (Region 1 as an example; the value in bold means the most accurate result in each case)

		Res-cLSTM	Multi-NSPI	MOPSTM	AGSL
L11	RMSE	**0.0137**	0.0243	0.0244	0.0245
	CC	**0.9328**	0.8264	0.8100	0.8464
	UIQI	**0.9309**	0.8142	0.7772	0.8157
L12	RMSE	**0.0177**	0.0212	0.0297	0.0441
	CC	**0.8762**	0.7888	0.7008	0.4832
	UIQI	**0.8742**	0.7841	0.6676	0.4716
L13	RMSE	**0.0174**	0.0256	0.0384	0.0415
	CC	**0.8887**	0.7828	0.8227	0.7603
	UIQI	**0.8876**	0.7731	0.7560	0.6984
L14	RMSE	**0.0186**	0.0290	0.0311	0.0300
	CC	**0.9389**	0.8528	0.8874	0.8725
	UIQI	**0.9300**	0.8408	0.7487	0.8400
L15	RMSE	**0.0155**	0.0222	0.0228	0.0283
	CC	**0.9295**	0.8585	0.8377	0.8375
	UIQI	**0.9287**	0.8514	0.8131	0.8033

V. CONCLUSION

In this paper, a deep learning method (i.e., Res-cLSTM) was proposed for reconstructing cloudy time-series images. Res-cLSTM leverages effective data from the non-cloud regions of cloudy time-series images for reconstruction. First, Res-cLSTM employs a convolutional LSTM to sequentially encode the cloudy time-series images in their temporal order, effectively capturing the complex temporal changes through the integrated long- and short-term memory flow. Then, Res-cLSTM utilizes a refined residual module to further decode the feature map and generate the final predictions. Experiments were conducted using Landsat 8 OLI time-series data across five different regions, yielding the following conclusions:

1) Compared to the four benchmark methods, Res-cLSTM produced more accurate predictions that are similar to the reference images, confirming its effectiveness for reconstructing cloudy time-series images.

2) Res-cLSTM can achieve the most accurate predictions even when the time-series auxiliary images contain clouds, showing resilience against the influence of cloud data in the auxiliary time-series.
3) Under varying degrees of overlap (up to 80%) between the cloud regions in the auxiliary time-series images and the target cloud regions, Res-cLSTM consistently produced predictions with greater accuracies.
4) Res-cLSTM demonstrates robustness against thin cloud omission and shows greater potential for practical implementations.
5) Res-cLSTM is a relatively lightweight network boasting high computational efficiency and a convergence speed that significantly exceeds that of the four benchmark methods, showing significant potential for real-time processing in practical applications.

REFERENCES

[1] K. P. Davies, J. Duncan, R. Varea, D. Ralulu, S. Nagaunavou, N. Wales, E. Bruce, and B. Boruff, "An intercomparison of national and global land use and land cover products for Fiji," *International Journal of Applied Earth Observation and Geoinformation*, vol. 135, pp. 104260, 2024.

[2] C. Giri, B. Pengra, J. Long, and T.R. Loveland, "Next generation of global land cover characterization, mapping, and monitoring," *International Journal of Applied Earth Observation and Geoinformation*, vol. 25, pp. 30–37, 2013.

[3] L. Graf, Q. Merz, A. Walter, and H. Aasen, "Insights from field phenotyping improve satellite remote sensing based in-season estimation of winter wheat growth and phenology," *Remote Sensing of Environment*, vol. 299, pp. 113860, 2023.

[4] W. Zhao, R. Lyu, J. Zhang, J. Pang, and J. Zhang, "A fast hybrid approach for continuous land cover change monitoring and semantic segmentation using satellite time series," *International Journal of Applied Earth Observation and Geoinformation*, vol. 134, pp. 104222, 2024.

[5] E. Amin, L. Pipia, S. Belda, G. Perich, L. Graf, H. Aasen, S. Wittenberghe, J. Moreno, and J. Verrelst, "In-season forecasting of within-field grain yield from Sentinel-2 time series data.," *International Journal of Applied Earth Observation and Geoinformation*, vol. 126, pp. 103636, 2024.

[6] M. Chen, X. Xu, X. Wu, and C. Mi, "Centennial-scale study on the spatial-temporal evolution of riparian wetlands in the Yangtze River of China," *International Journal of Applied Earth Observation and Geoinformation*, vol. 113, pp. 102874, 2022.

[7] Y. Li, M. Liu, X. Liu, W. Yang, and W. Wang, "Characterising three decades of evolution of forest spatial pattern in a major coal-energy province in northern China using annual Landsat time series," *International Journal of Applied Earth Observation and Geoinformation*, vol. 95, pp. 102254, 2021.

[8] S. V. Stehman, B. W. Pengra, J. A. Horton, and D. F. Wellington, "Validation of the U.S. Geological Survey's Land Change Monitoring, Assessment and Projection (LCMAP) Collection 1.0 annual land cover products 1985–2017," *Remote Sensing of Environment*, vol. 265, pp. 112646, 2021.

[9] C. Deng and Z. Zhu, "Continuous subpixel monitoring of urban impervious surface using Landsat time series," *Remote Sensing of Environment*, vol. 238, pp. 110929, 2020.

[10] H. He, J. Yan, D. Liang, Z. Sun, J. Li, and L. Wang, "Time-series land cover change detection using deep learning-based temporal semantic segmentation," *Remote Sensing of Environment*, vol. 305, pp. 114101, 2024.

[11] Y. Liang, S. Cao, Y. Mo, M. Du, and X. Wang, "Characterizing annual dynamics of two- and three-dimensional urban structures and their impact on land surface temperature using dense time-series Landsat images," *International Journal of Applied Earth Observation and Geoinformation*, vol. 134, pp. 104162, 2024.

[12] S. Wu, X. Lin, Z. Bian, M. Lipson, R. Lafortezza, Q. Liu, S. Grimmond, E. Velasco, A. Christen, V. Masson, B. Crawford, H. Ward, N. Chrysoulakis, K. Fortuniak, E. Parlow, W. Pawlak, N. Tapper, J. Hong, J. Hong, M. Roth, J. An, C. Lin, and B. Chen, "Satellite observations reveal a decreasing albedo trend of global cities over the past 35 years," *Remote Sensing of Environment*, vol. 303, pp. 114003, 2024.

[13] S. Xiong, X. Zhang, Y. Lei, G. Tan, H. Wang, and S. Du, "Time-series China urban land use mapping (2016–2022): An approach for achieving spatial-consistency and semantic-transition rationality in temporal domain," *Remote Sensing of Environment*, vol. 312, pp. 114344, 2024.

[14] M. D. King, S. Platnick, W.P. Menzel, S. A. Ackerman, and P. A. Hubanks, "Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua Satellites," *Transactions on Geoscience and Remote Sensing*, vol. 51, no. 7, pp. 3826–3852, 2013.

[15] J. Li, Z. Wu, Z. Hu, C. Jian, S. Luo, L. Mou, X. Zhu, and M. Molinier, 2022, "A lightweight deep learning-based cloud detection method for Sentinel-2A imagery fusing multiscale spectral and spatial features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–19.

[16] J. Li, Z. Wu, Q. Sheng, B. Wang, Z. Hu, S. Zheng, G. Camps-Valls, and M. Molinier, "A hybrid generative adversarial network for weakly-supervised cloud detection in multispectral images," *Remote Sensing of Environment*, vol. 280, pp. 113197, 2022.

[17] C. Luo, S. Feng, X. Yang, Y. Ye, X. Li, B. Zhang, Z. Chen, and Y. Quan, "LWCDnet: A lightweight network for efficient cloud detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[18] J. Zhang, H. Wang, Y. Wang, Q. Zhou, and Y. Li, "Deep network based on up and down blocks using wavelet transform and successive multi-scale spatial attention for cloud detection," *Remote Sensing of Environment*, vol. 261, pp. 112483, 2021.

[19] X. Zhou, X. Xie, H. Huang, Z. Shao, and X. Huang, "WodNet: Weak object discrimination network for cloud detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–20, 2024.

[20] M. Li, Q. Xu, J. Guo, and W. Li, "DecloudNet: Cross-patch consistency is a nontrivial problem for thin cloud removal from wide-swath multispectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.

[21] Z. Xu, K. Wu, L. Huang, Q. Wang, and P. Ren, "Cloudy image arithmetic: A cloudy scene synthesis paradigm with an application to deep-learning-based thin cloud removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[22] Q. Xu, J. Chen, X. Yan, and W. Li, "MRF-Net: An infrared remote sensing image thin cloud removal method with the intra-inter coherent constraint," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–19, 2024.

[23] Y. Zi, F. Xie, X. Song, Z. Jiang, and H. Zhang, "Thin cloud removal for remote sensing images using a physical-model-based cycleGAN with unpaired data," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[24] Y. Zhang, W. Zhang, J. Zhang, and J. Yin, "Double rank-one prior: Thin cloud removal by visible bands," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–10, 2024.

[25] Y. Chen, M. Chen, W. He, J. Zeng, M. Huang, and Y. Zheng, "Thick cloud removal in multitemporal remote sensing images via low-rank regularized self-supervised network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–13, 2024.

[26] L. Li, T. Huang, Y. Zheng, W. Zheng, J. Lin, G. Wu, and X. Zhao, "Thick cloud removal for multitemporal remote sensing images: When tensor ring decomposition meets gradient domain fidelity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[27] H. Peng, T. Huang, X. Zhao, J. Lin, W. Wu, and L. Li, "Deep domain fidelity and low-rank tensor ring regularization for thick cloud removal of multitemporal remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.

[28] J. Wang, X. Zhao, H. Li, K. Cao, J. Miao, and T. Huang, "Unsupervised domain factorization network for thick cloud removal of multitemporal remotely sensed images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023.

[29] X. Zhu, F. Gao, D. Liu, and J. Chen, "A modified neighborhood similar pixel interpolator approach for removing thick clouds in Landsat images," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 3, pp. 521–525, 2012.

[30] Q. Wang, L. Wang, X. Zhu, Y. Ge, X. Tong, P. M. Atkinson, "Remote sensing image gap filling based on spatial-spectral random forests," *Science of Remote Sensing*, vol. 5, pp. 100048, 2022.

[31] S. Malek, F. Melgani, Y. Bazi, and N. Alajlan, "Reconstructing cloud-contaminated multispectral images with contextualized autoencoder neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2270–2282, 2018.

[32] J. Gao, Q. Yuan, J. Li, and X. Su, "Unsupervised missing information reconstruction for single remote sensing image with Deep Code

Regression," *International Journal of Applied Earth Observation and Geoinformation*, vol. 105, pp. 102599, 2021.
[33] Y. Zi, X. Song, F. Xie, and Z. Jiang, "Thick cloud removal in multitemporal remote sensing images using a coarse-to-fine framework," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.
[34] C. Tao, S. Fu, J. Qi, and H. Li, "Thick cloud removal in optical remote sensing images using a texture complexity guided self-paced learning method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.
[35] F. Darbaghshahi, M. Mohammadi, and M. Soryani, "Cloud removal in remote sensing images using generative adversarial networks and SAR-to-optical image translation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–9, 2022.
[36] C. Duan, M. Belgiu, and A. Stein, "Efficient cloud removal network for satellite images using SAR-optical image fusion," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.
[37] W. Li, Y. Li, and J. Chan, "Thick cloud removal with optical and SAR imagery via convolutional-mapping-deconvolutional network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2865–2879, 2020.
[38] A. Meraner, P. Ebel, X. Zhu, and M. Schmitt, "Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 166, pp. 333–346, 2020.
[39] J. Pan, J. Xu, X. Yu, G. Ye, M. Wang, Y. Chen, and J. Ma, "HDRSA-Net: Hybrid dynamic residual self-attention network for SAR-assisted optical image cloud and shadow removal," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 218-B, pp. 258–275, 2024.
[40] X. Xiang, Y. Tan, and L. Yan, "Cloud-guided fusion with SAR-to-optical translation for thick cloud removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024.
[41] Z. Huang, Z. Zhu, Z. Wang, Y. Shi, H. Fang, and Y. Zhang, "DGDNet: Deep gradient descent network for remotely sensed image denoising," *IEEE Geoscience of Remote Sensing Letters*, vol. 20, pp. 6002405, 2023.
[42] Z. Huang, Z. Wang, Z. Zhu, Y. Zhang, H. Fang, Y. Shi, and T. Zhang, "DLRP: Learning Deep Low-Rank Prior for Remotely Sensed Image Denoising," *IEEE Geoscience of Remote Sensing Letters*, vol. 19, no. 6508905, pp. 1–5, 2022.
[43] Z. Zhu, C. E. Woodcock, C. Holden, and Z. Yang, "Generating synthetic Landsat images based on all available Landsat data: Predicting Landsat surface reflectance at any given time," *Remote Sensing of Environment*, vol. 162, pp. 67–83, 2015.
[44] X. Zhu, E. H. Helmer, J. Chen, and D. Liu, "An automatic system for reconstructing high-quality seasonal Landsat time series," *Remote Sensing Time Series Image Processing, CRC Press*, 2018.
[45] Z. Tang, H. Adhikari, P.K.E. Pellikka, and J. Heiskanen, "A method for predicting large-area missing observations in Landsat time series using spectral-temporal metrics," *International Journal of Applied Earth Observations and Geoinformation*, vol. 99, p. 102319, 2021.
[46] Y. Chen, Q. Weng, L. Tang, X. Zhang, M. Bilal, and Q. Li, "Thick clouds removing from multitemporal Landsat images using spatiotemporal neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.
[47] Q. Zhang, Q. Yuan, Z. Li, F. Sun, and L. Zhang, "Combined deep prior with low-rank tensor SVD for thick cloud removal in multitemporal images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 177, pp. 161–173, 2021.
[48] W. Zheng, X. Zhao, Y. Zheng, J. Lin, L. Zhuang, and T. Huang, "Spatial-spectral-temporal connective tensor network decomposition for thick cloud removal," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 199, pp. 182–194, 2023.
[49] A. Sebastianelli, E. Puglisi, M. Rosso, J. Mifdal, A. Nowakowski, P. Mathieu, F. Pirri, and S. Ullo, "PLFM: Pixel-level merging of intermediate feature maps by disentangling and fusing spatial and temporal data for cloud removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.
[50] X. Li, L. Wang, Q. Cheng, P. Wu, W. Gan, and L. Fang, "Cloud removal in remote sensing images using nonnegative matrix factorization and error correction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 148, pp. 103–113, 2019.
[51] Q. Wang, L. Wang, X. Zhu, Y. Ge, X. Tong, and P. M. Atkinson, "Remote sensing image gap filling based on spatial-spectral random forests," *Science of Remote Sensing*, vol. 5, pp. 100048, 2022.
[52] Y. Li, F. Wei, Y. Zhang, W. Chen, and J. Ma, "HS2P: Hierarchical spectral and structure-preserving fusion network for multimodal remote sensing image cloud and shadow removal," *Information Fusion*, vol. 94, pp. 215–228, 2023.
[53] Y. Wu, Y. Deng, S. Zhou, Y. Liu, W. Huang, and J. Wang, "CR-former: Single-image cloud removal with focused taylor attention," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.
[54] L. Arp, H. Hoos, P. van Bodegom, A. Francis, J. Wheeler, D. van Laar, and M. Baratchi, "Training-free thick cloud removal for Sentinel-2 imagery using value propagation interpolation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 216, pp. 168–184, 2024.
[55] S. Qian, Z. Xue, M. Jia, and H. Zhang, "Streamlined multilayer perceptron for contaminated time series reconstruction: A case study in coastal zones of southern China," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 221, pp. 193–209, 2025.
[56] H. Calatrava, B. Duvvuri, H. Li, R. Borsoi, E. Beighley, D. Erdoğmuş, P. Closas, and T. Imbiriba, "Recursive classification of satellite imaging time-series: An application to land cover mapping," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 218, part A, pp. 447–465, 2024.
[57] L. Fan, L. Xia, J. Yang, X. Sun, S. Wu, B. Qiu, J. Chen, W. Wu, and P. Yang, "A temporal-spatial deep learning network for winter wheat mapping using time-series Sentinel-2 imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 214, pp. 48–64, 2024.
[58] X. Shi, Z. Chen, H. Wang, and D. Yeung, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," Annual Conference on Neural Information Processing Systems, pp. 802–810, 2015.
[59] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
[60] J. Guo, J. Yang, H. Yue, H. Tan, C. Hou and K. Li, "RSDehazeNet: Dehazing network with channel refinement for multispectral remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2535–2549, 2021.
[61] Y. Li, W. Chen, Y. Zhang, C. Tao, R. Xiao, and Y. Tan, "Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning," *Remote Sensing of Environment*, vol. 250, 112045, 2020.
[62] B. Zhang, Y. Zhang, Y. Li, Y. Wan, and Y. Yao, "CloudViT: A lightweight vision transformer network for remote sensing cloud detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.

Lanxing Wang received the M.S. degree from Tongji University, Shanghai, China, in 2022. She is currently working toward the Ph.D. degree at Tongji University, Shanghai, China. Her research interests include remote sensing data reconstruction.

Qunming Wang received the Ph.D. degree from the Hong Kong Polytechnic University, Hong Kong, in 2015.

He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. He was a Lecturer (Assistant Professor) with Lancaster Environment Centre, Lancaster University, Lancaster, U.K., from 2017 to 2018. His 3-year Ph.D. study was supported by the hypercompetitive Hong Kong Ph.D. Fellowship and his Ph.D. thesis was

awarded as the Outstanding Thesis in the Faculty. He has authored or coauthored over 100 peer-reviewed articles in international journals such as *Remote Sensing of Environment*, *IEEE Transactions on Geoscience and Remote Sensing*, and *ISPRS Journal of Photogrammetry and Remote Sensing*. His research interests include remote sensing, image processing, and geostatistics.

Dr. Wang serves as Associate Editor for *Science of Remote Sensing* (sister journal of *Remote Sensing of Environment*) and *Photogrammetric Engineering & Remote Sensing*, and was Associate Editor for *Computers and Geosciences* (2017−2020).

Peter M. Atkinson received the Ph.D. degree from the University of Sheffield (NERC CASE award with Rothamsted Experimental Station) in 1990. More recently, he received the MBA degree from the University of Southampton in 2012.

He is currently Distinguished Professor of Spatial Data Science and Executive Dean of the Faculty of Science and Technology at Lancaster University, UK. He was previously Professor of Geography at the University Southampton, where he is currently Visiting Professor. He is also Visiting Professor at the Chinese Academy of Sciences, Beijing and previously held the Belle van Zuylen Chair at Utrecht University, the Netherlands.

The main focus of Peter's research is in remote sensing, geographical information science and spatial (and space-time) statistics applied to a range of environmental science and socio-economic problems. He has published over 400 peer-reviewed articles on these topics in international scientific journals and over 50 refereed book chapters. He has also edited 14 journal special issues and eight books. Peter is currently listed as an ISI highly-cited researcher.

Peter has received several awards for his research including the Cuthbert Peek Award of the Royal Geographical Society-Institute of British Geographers and the Peter Burrough Award of the International Spatial Accuracy Research Association. He is also a Fellow of the Learned Society of Wales.

Professor Atkinson is Editor-in-Chief of *Science of Remote Sensing*, a sister journal of *Remote Sensing of Environment*. He is also Associate Editor for *Environmetrics*.