ISPRS Journal of Photogrammetry and Remote Sensing 177 (2021) 161–173

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Combined deep prior with low-rank tensor SVD for thick cloud removal in multitemporal images

Qiang Zhang [a], Qiangqiang Yuan [b,*], Zhiwei Li [c,*], Fujun Sun [d], Liangpei Zhang [a]

[a] State Key Laboratory of Information Engineering, Survey Mapping and Remote Sensing, Wuhan University, Wuhan, China
[b] School of Geodesy and Geomatics, Wuhan University, Wuhan, China
[c] School of Resource and Environmental Sciences, Wuhan University, Wuhan, China
[d] Beijing Electro-mechanical Engineering Institute, Beijing, China

ARTICLE INFO

Keywords:
Thick cloud removal
Multitemporal images
Spatio-temporal
Low-rank tensor SVD
Deep prior

ABSTRACT

The thick cloud coverage phenomenon severely disturbs optical satellite observation missions (covering approximately 40–60% areas in the global scale). Therefore, the manner by which to eliminate thick cloud in remote sensing imagery is greatly significant and indispensable. In this study, we combine the deep spatio-temporal prior with low-rank tensor singular value decomposition (DP-LRTSVD) for thick cloud removal in multitemporal images. On the one hand, DP-LRTSVD utilizes the low-rank characteristic of multitemporal images via the third-order tensor SVD and completion. On the other hand, DP-LRTSVD employs the deep spatio-temporal feature expression ability by 3D convolutional neural network. The proposed framework can effectively eliminate thick cloud in multitemporal images through combining the model-driven and data-driven strategies. Moreover, DP-LRTSVD outperforms on thick cloud removal in the simulated and real multitemporal Sentinel-2/GF-1 experiments compared with model-driven or data-driven methods. In contrast with most methods that can only use a single reference image for thick cloud removal, the proposed method can simultaneously eliminate thick cloud in time-series images.

1. Introduction

Optical satellites play a crucial role for earth observation mission, such as global land covering (Qiu et al., 2019), ocean monitoring (Liu et al., 2019), and atmospheric pollutant retrieval (Wang et al., 2020). However, almost all of the bands in optical satellites are inevitably polluted by thick cloud and cloud shadow due to the physical limitation of the electromagnetic spectrum (Weng, 2009; Watmough et al., 2011). The original information under thick cloud and shadow is directly covered or polluted. This issue seriously disturbs the subsequent remote sensing imagery processing and applications (Li et al., 2019; Meraner et al., 2020; Zhang et al., 2020).

The distribution of thick cloud is widespread in most obtained optical satellite data (Qiu et al., 2019). Cloud-free remote sensing images are rare and difficult to acquire, especially for the desired time and location (Li et al., 2019; Wang et al., 2020; Xu et al., 2019). Therefore, the mechanism by which to eliminate the thick cloud in optical images is significant for remote sensing interpretations and applications (Zhu et al., 2011; Zhang et al., 2019a).

To date, a variety of methods have been presented for thick cloud removal in remote sensing images. From the perspective of information domain (Shen et al., 2015; Zhang et al., 2019), these methods could be classified as the three types: 1) Spatial domain; 2) temporal domain; and 3) spatio-temporal domain methods. The details of these algorithms are listed as follows:

1) Spatial domain: The spatial-based method is the basic strategy for thick cloud removal in single remote sensing imagery. This strategy transforms the thick cloud removal task into the image completion or inpainting task (Watmough et al., 2011; Xu et al., 2021). The spatial-based methods rely on just a single image for missing data reconstruction without other auxiliary information. These methods can utilize diffusion-based, variation-based, and learning-based strategy to eliminate the thick cloud in remote sensing data by means of several mathematical tools. Chan and Shen (2001) developed a nonlinear partial differential equation image completion algorithm. Maalouf et al. (2009) employed bandelet transform and total variation for thick cloud removal in remote sensing image. Zheng et al. (2020) introduced a generative network for single image thick removal. The spatial-based methods are

* Corresponding author.
E-mail addresses: yqiang86@gmail.com (Q. Yuan), lizw@whu.edu.cn (Z. Li).

https://doi.org/10.1016/j.isprsjprs.2021.04.021
Received 1 January 2021; Received in revised form 21 April 2021; Accepted 29 April 2021

convenient for eliminating small thick cloud in a single image. Nevertheless, these methods are unstable and incredible, especially for large area cloud covering (Shen et al., 2015).

2) Temporal domain: The temporal-based methods utilize the time-series information for filling the cloudy regions (Bayad et al., 2020). These methods take advantage of the sequentiality in multi-temporal images, especially for long-term remote sensing data. The time-series filters are typical methods for the temporal-based methods. Chen et al. (2004) introduced the Savitzky-Golay filter for time-series NDVI reconstruction under the cloudy or poor atmospheric conditions. Sedano et al. (2014) proposed a recurrent Kalman filter for rebuilding continuous and dense 30 m NDVI long-term results through MODIS and Landsat images. Kong et al. (2019) developed a weighted Whittaker filter for global EVI time-series smoothing and gap-filling, especially for thick cloud contaminated pixels. Overall, the temporal-based methods can effectively reflect the smooth and sequential 1D temporal information. However, these methods ignore the 2D spatial information, which is crucial for large area thick cloud removal, especially for the cloud boundary regions (Mateo-García et al., 2020).

3) Spatio-temporal domain: The spatio-temporal methods have been gradually presented for thick cloud removal in recent years to overcome the above-mentioned shortages of the spatial-based and temporal-based models. These methods jointly employed spatial and temporal information (Weng et al., 2014), and they could simultaneously consider the spatial consistency and temporal sequentiality. Zhu et al. (2011) proposed an improved neighborhood homogeneous value interpolation spatio-temporal model that could greatly fill the gap or thicken the cloud regions in Landsat reflectivity images. Chen et al. (2016) introduced the spatio-temporal weighted regression strategy to reconstruct time-series cloud-free Landsat images. Zhang et al. (2019b) utilized the spatio-temporal group-sparsity and robust principal component analysis theory to remove the thick cloud in multi-temporal images. Chen et al. (2019) developed a blind thick cloud and shadow removal method by means of the low-rank property of multi-temporal images. Cao et al. (2020) presented a spatio-temporal autoregression model for thick cloud removal in Landsat time-series images, and it also considered the effects of cloudy temporal images. In addition to the above-mentioned model-based methods, several data-driven-based methods have been proposed for thick cloud removal, benefiting from the rapid development of deep learning. Zhang et al. (2018) built a unified spatio-temporal–spectral network to deal with the various incomplete information filling issues, such as deadline and thick cloud elimination. Zhang et al. (2020) further developed a deep progressive spatio-temporal model (named PSTCR) for multi-temporal image thick cloud removal. The PSTCR method could use multiple cloudy images rather than just a single cloud-free image as the reference information.

Recently, the spatio-temporal methods have increasingly become the mainstream strategy for thick cloud removal (Cao et al., 2020; Duan et al., 2020; Li et al., 2019). These methods can be classified as two types: model-driven and data-driven methods (Luo et al., 2011; Yuan et al., 2019; Zhang et al., 2019). Nevertheless, pros and cons exist in the model-driven or data-driven strategy for spatio-temporal thick cloud removal (Zhang et al., 2020). Some shortages and limitations still exist in the two strategies for thick cloud removal. The detailed descriptions are provided below.

For model-driven methods: These methods can accurately depict the inherent characteristics of multi-temporal images, such as low-rank tensor completion (Ji et al., 2018) and sparse representation (Xu et al., 2016). However, these methods are sensitive to the setting parameters, such as rank value and iteration number (Chen et al., 2011; Gao and Gu, 2017). Besides, the model-driven methods are time consuming due to the complex iterative optimizations (Jiang et al., 2020; Zhao et al., 2020; Zhang et al., 2020).

For data-driven methods: These methods efficiently execute due to the deep learning strategy and large sample training. Nevertheless, the simulated cloud distribution for multi-temporal cloud-free samples is not consistent with the real thick cloud distribution because of the complicated atmospheric flow procedure. However data-driven methods poorly perform when the data distribution of the training samples are dissimilar with the testing cloudy images (Liang et al., 2020; Ouyang et al., 2020; Yu et al., 2019). Lastly, data-driven methods still have limitations to fully utilize the spatial consistency between the cloudy and the neighboring cloud-free regions (Paudel and Andersen, 2011). These issues greatly affect the generality of data-driven methods.

Apart from above-mentioned shortcomings of the data-driven and model-driven methods, another practical issue also needs to be emphasized. Most spatio-temporal methods can recover just a single cloudy image rather than time-series cloudy images (Zhang et al., 2021). In addition, the reference image must be cloud-free to act as the temporal complementary information in most cloud removal methods. However, this assumption is Utopian and unpractical due to the actual imaging environment (Zhu and Woodcock, 2012). Almost all full-width images are polluted by thick cloud in optical satellite images to a certain degree. A cloud-free reference image is difficult and rare to acquire (Qiu et al., 2019; Zhu et al., 2015). Hence, the mechanism by which to simultaneously deal with time-series cloudy images is also significant for thick cloud removal.

From the above perspectives, can we simultaneously utilize the advantages of model-driven methods with data-driven methods for thick cloud removal in time-series cloudy images?

On this basis, we develop a novel spatio-temporal strategy for multitemporal image thick cloud removal via combining the model-driven and data-driven strategies. The main contributions of this study are listed as follows:

- In this study, we propose a novel framework: Deep Spatio-temporal Prior with Low-rank Tensor Singular Value Decomposition (SVD) (DP-LRTSVD) for thick cloud removal in multitemporal images via combining the model-driven and data-driven strategies. The deep 3D spatio-temporal prior and low-rank tensor completion are integrated in the proposed method to benefit from each other.
- DP-LRTSVD jointly utilizes the low-rank characteristic of time-series images via the third-order tensor SVD and the deep spatio-temporal feature expression ability via the 3D convolutional neural network under the ADMM iteration optimization framework. This strategy can effectively accelerate and boost the convergence procedure.
- In contrast with most methods that can only use a single reference image for thick cloud removal, DP-LRTSVD can simultaneously deal with time-series cloudy images. Experimental results testify the practicability of the proposed method for GF-1/Sentinel-2 images.

The rest of this paper is organized as follows. Section 2 defines the problem formulation and describes the algorithmic details of the proposed method. Section 3 shows the multitemporal thick cloud removal results of the simulated and real experiments. Section 4 presents the discussion to investigate the parameter sensitivity and iteration optimization of the proposed method. Finally, Section 5 summarizes the conclusion and prospect.

2. Methodology

2.1. Problem formulation

Before the proposed cloud removal model is illustrated, the notation declarations are defined as follows: The tensor value is set as the curlicue format, such as $\mathcal{X}$. The matrix value is fixed in boldface capital format,

such as **X**. The vector value is determined in boldface lowercase format, such as **x**. The scalar value is identified in italic lowercase format such as x.

In consideration of the thick cloud covering as the information missing issue in multitemporal images, the general degraded model can be simplified as follows:

$$\mathcal{Y} = \mathcal{X} \otimes \mathcal{M} \tag{1}$$

where $\mathcal{Y} \in \mathbb{R}^{w \times h \times t}$ represents the time-series cloudy images for each band. $\mathcal{X}$ stands for the corresponding time-series cloud-free images of $\mathcal{Y}$. $\mathcal{M}$ refers to the time-series cloud and shadow masks of $\mathcal{Y}$. $\otimes$ denotes the cloud and shadow pollution procedure. The solving procedure from cloudy $\mathcal{Y}$ to cloud-free images $\mathcal{X}$ is an ill-posed problem. Therefore, suitable priors must be introduced to interfere the latent clean images $\mathcal{X}$. Multitemporal images have a strong correlation in temporal dimension. We depict the statistical distribution of the singular values of the multitemporal Sentinel-2 images, as shown in Fig. 1. The image shows the decaying trends of the curve, thus indicating the low-rank prior of multitemporal images.

An extensive thick cloud removal framework for multitemporal images could be formulated as follows taking the low-rank characteristic of the third-order tensor via stacking the time-series cloudy images:

$$\min_{\mathcal{X}} \frac{1}{2}\|\mathcal{Y} - \mathcal{X} \otimes \mathcal{M}\|_F + \lambda \cdot \mathcal{T}\left(\mathcal{X}\right) \tag{2}$$

where $\mathcal{T}(\cdot)$ could be denoted as any tensor low-rank regularization term to utilize the low-rank prior of multitemporal images, and $\lambda$is the balancing factor of the regularization prior term. (2) could be rewritten through introducing the tenor nuclear norm theory (Zhang and Aeron, 2017):

$$\mathcal{X}\min_{\mathcal{X}} \frac{1}{2}\|\mathcal{Y} - \mathcal{X} \otimes \mathcal{M}\|_F + \lambda \cdot \|\mathcal{X}\|_{TNN} \tag{3}$$

where $\|\mathcal{X}\|_{TNN}$ refers to the tubal nuclear norm of tensor $\mathcal{X}$, which is the sum of singular values of all the frontal slices of $\overline{\mathcal{X}}$. This expression can be defined as follows:

$$\|\mathcal{X}\|_{TNN} = \|\overline{\mathcal{X}}\|_* \tag{4}$$

2.2. DP-LRTSVD model

The flowchart of the proposed DP-LRTSVD framework for thick cloud removal in multitemporal images is depicted in Fig. 2. First, the time-series cloudy images and corresponding cloud masks $\mathcal{M} \in \mathbb{R}^{w \times h \times t}$ are stacked as the initialized third-order tensor $\mathcal{Y} \in \mathbb{R}^{w \times h \times t}$. Second, this third-order tensor is decomposed by three components, namely, $\mathcal{U} \in \mathbb{R}^{h \times h \times t}$, $\mathcal{S} \in \mathbb{R}^{w \times h \times t}$, and $\mathcal{V} \in \mathbb{R}^{w \times w \times t}$ through tensor-SVD in Fig. 2. Third, the reorganized tensor $\mathcal{Z} \in \mathbb{R}^{w \times h \times t}$ is generated through tensor-product and then updated as the updated tensor $\mathcal{P} \in \mathbb{R}^{w \times h \times t}$ via (13) and cloud masks. Fourth, tensor $\mathcal{P}$ and the corresponding cloud masks are simultaneously imported as the input data of a deep 3D spatio-temporal network to further adjust the reconstructing cloudy regions. Finally, if the current result satisfies the convergence condition, then the immediate result is exported as the final outputs in Fig. 2. Otherwise, the optimizing iteration is carried out through alternating the direction method of multipliers (ADMM) (Boyd et al., 2011) until the convergence condition or the maximum number of iterations is met. The details of these steps are described as follows:

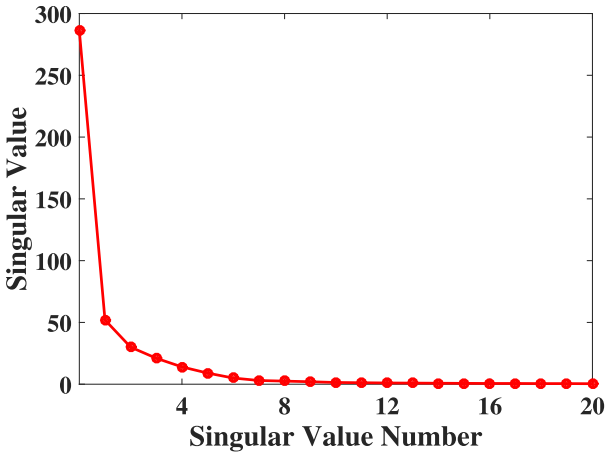

Fig. 1. Statistical distribution of the singular values of the multitemporal Sentinel-2 images.

a) Preprocessing: Before thick cloud removal in multitemporal images, the mechanism by which to estimate the precise cloudy regions is extremely significant for the subsequent procedure. Therefore, cloud detection in remote sensing imagery is an indispensable preprocessing step for cloud removal. In this work, we use the MSCFF (Li et al., 2019) method to obtain the time-series cloud masks $\mathcal{M}$, as shown in Fig. 2. The cloud masks simultaneously include cloud and cloud shadow. Cloud mask in a single date is a binary matrix, in which the cloudy and cloud-free pixels are denoted as one and zero, respectively. The initialized input $\mathcal{Y}$ is generated via the point-wise product operation between the original time-series cloudy images and the corresponding cloud masks to better solve the optimal value in tensor decomposition.

b) Tensor-product: Tensor-product refers to the $*$ operation between two third-order tensors, namely, $\mathcal{B}_1 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B}_2 \in \mathbb{R}^{n_2 \times n_4 \times n_3}$. The definition of tensor-product $\mathcal{B}_1 * \mathcal{B}_2$ is determined as follows:

$$\mathcal{B}_3\left(i,j,:\right) = \sum_{k=1}^{n_2} \mathcal{B}_1\left(i,k,:\right) \odot \mathcal{B}_2\left(k,j,:\right) \tag{5}$$

where $\mathcal{B}_3 \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ and $\odot$represent the circular convolution operation.

c) Tensor-SVD: After the cloud masks are obtained, we stack the multiple temporal cloudy images for each band as a three-order tensor. Then, tensor-SVD (Zheng et al., 2020) is utilized to constrain the low-rank characteristic of multitemporal images. The third-order tensor $\mathcal{Y} \in \mathbb{R}^{w \times h \times t}$ can be decomposed as three components, namely, $\mathcal{U} \in \mathbb{R}^{h \times h \times t}$, $\mathcal{S} \in \mathbb{R}^{w \times h \times t}$, and $\mathcal{V} \in \mathbb{R}^{w \times w \times t}$ in Fig. 2:

$$\mathcal{Y} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\top} \tag{6}$$

where $*$ stands for the tensor product operation in the next step. In the third-order tensor-SVD, we primarily obtain the three components as follows:

$$\left(\mathbf{U}_i, \mathbf{S}_i, \mathbf{V}_i^{\top}\right) = SVD\left(\mathbf{Y}_i\right) \rightarrow i = 1, 2, 3 \tag{7}$$

where $SVD(\cdot)$ represents the matrix SVD function. $\mathbf{Y}_i$ refers to the frontal slice matrix of the third-order tensor $\mathcal{Y}(:,:,i)$, the same as $\mathbf{U}_i, \mathbf{S}_i$, and $\mathbf{V}_i$. The fast Fourier transformation (FFT) and inverse fast Fourier transformation (IFFT)[1] are introduced to these matrix SVD operations to enhance the mathematical solving efficiency (Lu et al., 2015). Then, the tensor tubal rank r of a third-order tensor is determined to be the maximum number of non-zero tubes of each $\overline{\mathbf{S}}_i$ in (7):

$$r = rank_{tubal}\left(\mathcal{Y}\right) = \max\left(D\left(\overline{\mathbf{S}}_1\right), D\left(\overline{\mathbf{S}}_2\right), D\left(\overline{\mathbf{S}}_3\right)\right) \tag{8}$$

where $D(\cdot)$ stands for the non-zero number of the diagonal matrix operation for $\overline{\mathbf{S}}_1, \overline{\mathbf{S}}_2$, and $\overline{\mathbf{S}}_3$. To further exploit the low-rank and sparse characteristic of the third-order tensor, the original $\mathcal{U}, \mathcal{S}$, and $\mathcal{V}$ are

[1] We implement the matrix FFT and IFFT operation through Python scipy package

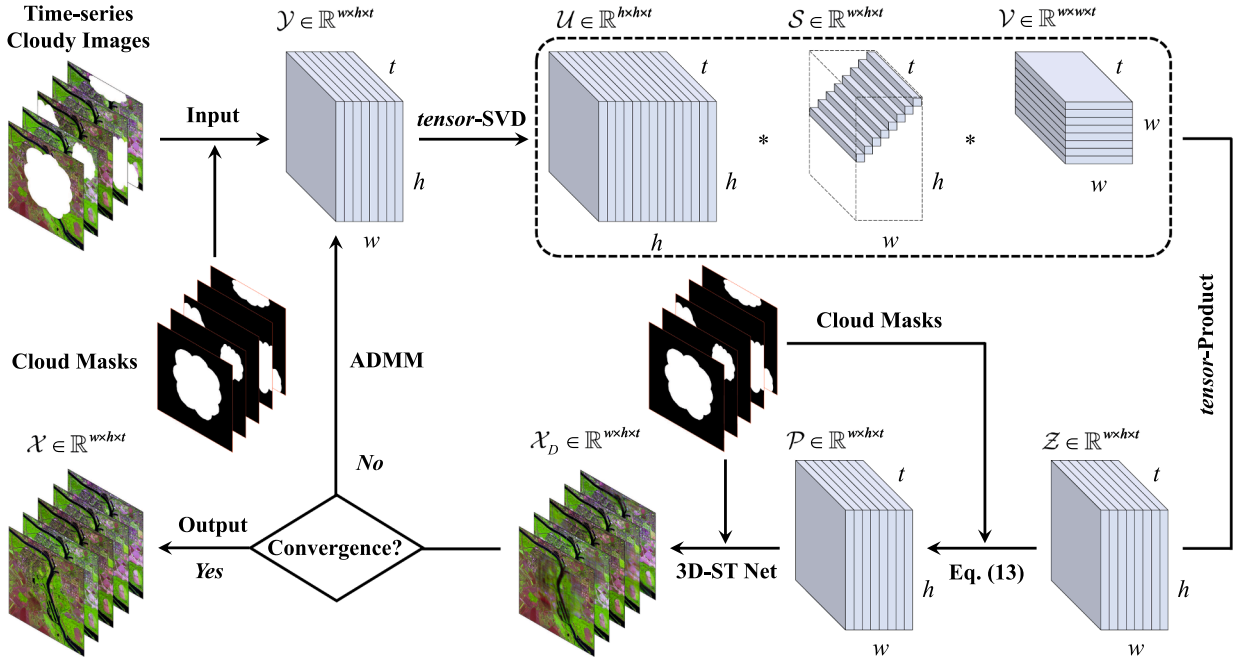

Fig. 2. Flowchart of the proposed DP-LRTSVD framework.

simplified through the tubal rank interception as follows:

$$\overline{\mathcal{U}} = \mathcal{U}\left(:,1:r,:\right) \tag{9}$$

$$\overline{\mathcal{S}} = \mathcal{S}\left(1:r,1:r,:\right) \tag{10}$$

$$\overline{\mathcal{V}} = \mathcal{V}\left(:,1:r,:\right) \tag{11}$$

After tensor-SVD and tubal rank simplification, the proposed model reorganizes the new third-order tensor $\mathcal{Z} \in \mathbb{R}^{w\times h\times t}$ below:

$$\mathcal{Z} = \overline{\mathcal{U}} * \overline{\mathcal{S}} * \overline{\mathcal{V}}^{\top} \tag{12}$$

Then, we introduce a Lagrange multiplier $\mathcal{Q} \in \mathbb{R}^{w\times h\times t}$ and penalty parameter β (Lu et al., 2019) to optimize the proposed framework:

$$\mathcal{P} = \mathcal{M} \odot \left(\mathcal{Z} + \beta \cdot \mathcal{Q}\right) + \left(\mathbf{1} - \mathcal{M}\right) \odot \mathcal{Y}^0 \tag{13}$$

where cloud masks $\mathcal{M}$ are employed to restrict the cloud-free and reconstruction regions. $\mathcal{Y}^0$ is the initial input of the proposed model in Fig. 2.

d) Deep 3D Spatio-temporal Prior: Considering the context consistency between the cloudy and the adjacent cloud-free regions, the spatio-temporal information can be utilized for thick cloud removal. We introduce a deep 3D spatio-temporal prior into the proposed method to boost the reconstruction accuracy in Fig. 2. The architecture of the deep 3D spatio-temporal network (3D-ST Net) is displayed in Fig. 3, taking the reorganized third-order tensor $\mathcal{P}$ and cloud masks $\mathcal{M}$ as the inputs.

In Fig. 3, 3D-ST Net uses the 3D CNN with seven layers to exact the spatial and temporal information for cloudy region reconstruction. The thick cloud removal result could be further enhanced in terms of context consistency between the cloudy and the adjacent cloud-free regions through this step. The mathematical procedure is defined as follows:

$$\mathcal{X}_D = \mathcal{M} \odot 3DConv_{depth=7}\left(\mathcal{P}\right) + \left(\mathbf{1} - \mathcal{M}\right) \odot \mathcal{Y}^0 \tag{14}$$

where $3DConv(\cdot)$ represents the cascade-connected 3D CNN layer in Fig. 3. The specific descriptions of the network optimization and training are provided in Sections 2.3 and 3.1, respectively.

e) ADMM Optimization: After the reconstruction tensor $\mathcal{X}_D$ via tensor-SVD, tensor-Product, and 3D-ST Net were obtained, we use the alternating direction method of multipliers (ADMM) to iteratively approximate the optimal solution (Ji et al., 2018). If the current iteration meets the condition of convergence in Fig. 2, then the result is exported as the final reconstruction output $\mathcal{X} \in \mathbb{R}^{w\times h\times t}$. Otherwise, the proposed framework executes the ADMM iteration strategy, as shown in Fig. 2. The iterative Lagrange multipliers $\mathcal{Q}^k, \mathcal{Y}^k$ and penalty coefficient β^k of the k-th iteration follow the rules for updating in the proposed method to boost the convergence solution through ADMM:

$$\mathcal{Y}^k = \mathcal{X}_D^{k-1} - 1/\beta^{k-1} \cdot \mathcal{Q}^{k-1} \tag{15}$$

$$\mathcal{Q}^k = \mathcal{Q}^{k-1} + \beta^{k-1} \cdot \left(\mathcal{Y}^k - \mathcal{X}_D^k\right) \tag{16}$$

$$\beta^k = \min\left(\eta \cdot \beta^{k-1}, \beta_{\max}\right) \tag{17}$$

where $\eta(\eta > 1)$ is the scaling threshold to enhance the robustness of the iteration solving procedure. $\beta_{\max}$ stands for the maximum number of ADMM iteration. In the first iteration, the initialized $\mathcal{X}_D^0, \beta^0$, and $\mathcal{Q}^0$ are equal to $\mathcal{Y}^0$, 0.02, and zero-tensor, respectively. The whole workflow of the ADMM algorithm in the proposed framework is listed as follows:

Algorithm 1. Combined Deep 3D Spatio-temporal Prior with Low-rank Tensor SVD for Thick Cloud Removal via ADMM

Input: Time-series cloudy images $\mathcal{Y}$, corresponding cloud masks $\mathcal{M}$
Initialization: $\mathcal{Y}^0 = (\mathbf{1} - \mathcal{M}) \odot \mathcal{Y}, \mathcal{X}_D^0 = \mathcal{Y}^0, \mathcal{Q}^0 = \mathbf{0}, \beta^0 = 0.02, \beta_{\max} = 1, \eta = 1.3, \varepsilon = 1e{-}5, k = 1, k_{\max} = 20$
1: **while** not converged and $k \leqslant k_{\max}$ **do**
2: Updating $\overline{\mathcal{U}}^k, \overline{\mathcal{S}}^k$, and $\overline{\mathcal{V}}^k$ via Eqs. (7)–(11)
3: Updating $\mathcal{Z}^k$ via (12)
4: Updating $\mathcal{P}^k$ via (13)
5: Updating $\mathcal{X}_D^k$ via (14)
6: Updating $\mathcal{Y}^k, \mathcal{Q}^k$, and β^k via (15)–(17), respectively
7: If $\|\mathcal{X}_D^k - \mathcal{X}_D^{k-1}\|_F / \|\mathcal{X}_D^{k-1}\|_F < \varepsilon$, stop iteration
8: $k = k + 1$
9: **end while**
Output: The construction cloud-free result $\mathcal{X} = \mathcal{X}_D^k$

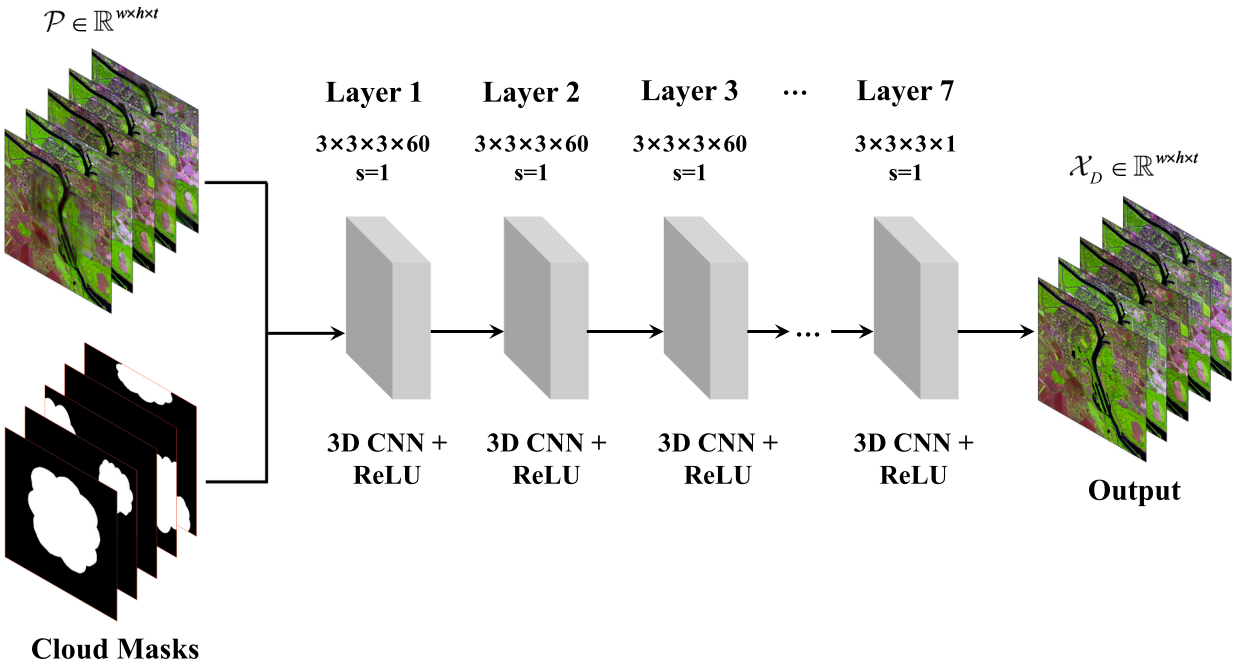

Fig. 3. Architecture of 3D-ST Net in the proposed model.

2.3. Network optimization

In the proposed DP-LRSVD model, we employ the 3D-ST Net to further enhance the context consistency between the cloudy and the adjacent cloud-free regions in Fig. 3. Therefore, the mechanism by which to optimize this network is crucial for thick cloud removal in multitemporal images. The reconstruction network develops a jointly global-regional loss $\mathcal{L}$ for the network optimization to attain holistic consistency and regional specificity:

$$\mathcal{L} = \mu_1 \cdot \mathcal{L}_g + \mu_2 \cdot \mathcal{L}_r + (1 - \mu_1 - \mu_2) \cdot \mathcal{L}_{TV} \tag{18}$$

where μ_1 and μ_2 represent the balancing thresholds for the global $\mathcal{L}_g$, regional $\mathcal{L}_r$, and total variational (TV) losses $\mathcal{L}_{TV}$. In this global–regional loss $\mathcal{L}$, the global $\mathcal{L}_g$, regional $\mathcal{L}_r$, and TV losses $\mathcal{L}_{TV}$ are defined as follows:

$$\mathcal{L}_g = \frac{1}{2N} \sum_{n=1}^{N} \left\| \mathcal{X}_D^{(n)} - \mathcal{X}^{(n)} \right\|_2^2 \tag{19}$$

$$\mathcal{L}_r = \frac{1}{2N} \sum_{n=1}^{N} \frac{1}{sum(\mathcal{M}^{(n)})} \left\| \mathcal{M}^{(n)} \odot \mathcal{X}_D^{(n)} - \mathcal{M}^{(n)} \odot \mathcal{X}^{(n)} \right\|_2^2 \tag{20}$$

$$\mathcal{L}_{TV} = \frac{1}{2N} \sum_{n=1}^{N} \sum_{ij} \frac{1}{sum(\mathcal{M}^{(n)})} \sqrt{\left(\mathcal{X}_{D(i,j+1,:)}^{(n)} - \mathcal{X}_{D(i,j,:)}^{(n)} \right)^2 + \left(\mathcal{X}_{D(i+1,j,:)}^{(n)} - \mathcal{X}_{D(i,j,:)}^{(n)} \right)^2} \tag{21}$$

where (n) stands for the n-th training sample with the whole number size N. Global loss $\mathcal{L}_g$ is employed to reconstruct the holistic results. Regional loss $\mathcal{L}_r$ is introduced to ensure the context consistency between the cloudy and the adjacent cloud-free regions. TV loss $\mathcal{L}_{TV}$ is utilized to promote the spatial smoothness in the reconstruction cloudy regions. The detailed descriptions in terms of the network training and parameter setting are provided in Section 3.1.

3. Experiments

In this section, experimental setting explanations and different experimental outcomes are provided to validate the practicability of the presented DP-LRSVD. First, Section 3.1 depicts the details of network training and parameter setting. Afterward, the simulated and real experiments for thick cloud removal in multitemporal images are carried out in Sections 3.2 and 3.3, respectively.

3.1. Network training and parameter setting

Network training is important to enhance the context consistency between the cloudy and the adjacent cloud-free regions in 3D-ST Net of the proposed method. Time-series of the five different clean Landsat TM[2] images in the same covering areas is selected in this work for the training data. The acquisition times start from November 9, 2004 to January 30, 2005. These images comprised four bands from blue to NIR spectrum. The spatial size of each image is equal to 2400 × 3200 and with the 30 m spatial resolution. After the normalization for each band, we crop these time-series images as the patch-tensors with the size of 80 × 80 × 5. Number N of the whole labels is equal to 19,200 in the training dataset. The additional cloud mask patches are also employed to generate the simulated cloudy time-series data. Then, the original clean and cloudy patches are fixed as the label and data samples in our training dataset, respectively.

The specific settings in network training in terms of the network hyperparameters are determined below. The batch size is fixed as 32 for the training samples. The initial learning rate in the network training procedure is varied from 0.01 and gradually reduced via product parameter 0.5 for every 50 epochs (Zhang et al., 2018). The whole epoch number is set as 300 for network training. ADAM (Zhang et al., 2020) strategy is employed as the gradient descent algorithm in back-propagation. We use the Python language and Pytorch platform to train and test our network under Windows 10 environment and GPU accelerating mode.

The related parameters for the parameter setting of the proposed framework are listed as follows. Factor β^0 in (15) and (16) is initialized as 0.02, whose upper limit $\beta_{\max}$ is equal to one in (17). The scaling threshold $\eta$is set as 1.3 to enhance the robustness of the iteration solving procedure in (17). In terms of the network optimization, the thresholds

[2] Data download link: https://earthexplorer.usgs.gov/.

μ_1 and μ_2 are fixed as 0.7 and 0.25, respectively, to balance the global, regional, and TV losses in (18).

3.2. Simulated experiments

In the simulated experiments, we employ the time-series cloud-free images as the testing data. The multitemporal Sentinel-2 and GF-1 images are utilized to verify the reliability of the presented thick cloud removal framework. The simulated cloud masks are imposed on the clean images. Then, the time-series simulated cloudy images are reconstructed and displayed in this section through different cloud removal algorithms. The model-driven methods HaLRTC (Liu et al., 2013), BCPF (Zhao et al., 2015), TNN (Zhang and Aeron, 2017), and the data-driven method PSTCR (Zhang et al., 2020) are determined as the comparing algorithms. In terms of the quantitative evaluation index, we use the average Correlation Coefficient (CC), Structural Similarity Index

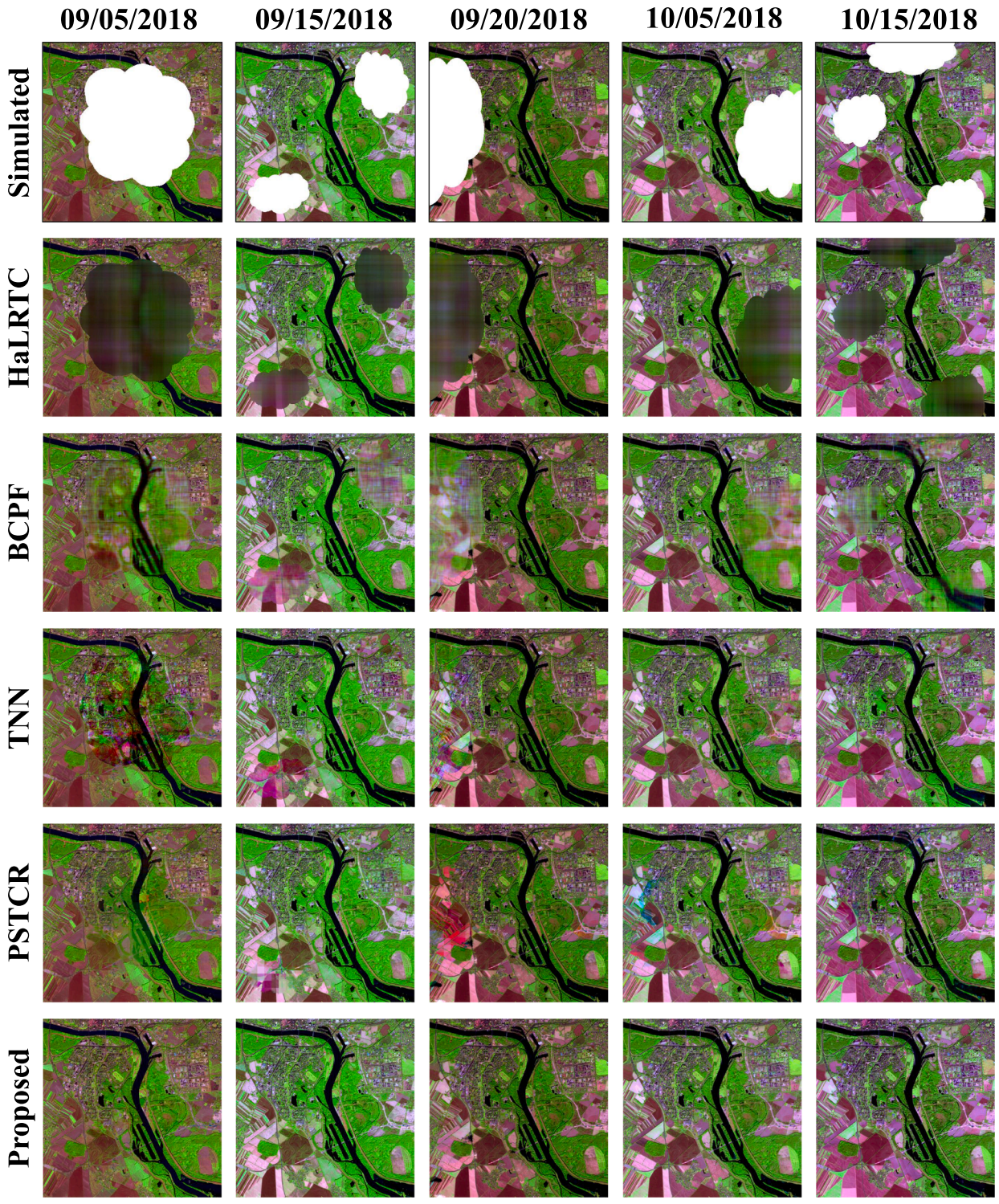

Fig. 4. Simulated cloud removal experiments for the multitemporal Sentinel-2 images (09/05, 09/15, 09/20, 10/05, and 10/15/2018) through HaLRTC, BCPF, TNN, PSTCR, and the proposed method.

Measure (SSIM), Root Mean Square Error (RMSE), and Spectral Angle Mapper (SAM) for multitemporal reconstructed images. The detailed descriptions and results are provided as follows:

a) Multitemporal Sentinel-2 results: Five time-series Sentinel-2 images (acquired times: 09/05, 09/15, 09/20, 10/05, and 10/15/2018 via Sentinel-2A/B L1C over Mechelen, Belgium) are imposed with the simulated cloud, as shown in the first row of Fig. 4. The second, third, fourth, fifth, and sixth rows are the reconstructed multitemporal images by HaLRTC, BCPF, TNN, PSTCR, and the proposed method in Fig. 4, respectively. We validate on 20 m spatial resolution bands (B12, B11, B8a, B07, B06, and B05) in these time-series Sentinel-2 images. The false-color 500 × 500 results of bands B11, B07, and B05 are displayed in Fig. 4. Evaluation indexes CC, SSIM, RMSE, and SAM of the Sentinel-2 simulated experiments are listed in Table 1. The optimal values in Table 1 are marked in bold format for comparison.

In the fourth row of Fig. 4, the results of TNN show a serious spectral distortion issue at date 09/05/2018, especially for large missing area reconstruction. TNN exploits only the low-rank tensor property of the 3D time-series Sentinel-2 images. Nevertheless, TNN does not consider the global consistency and regional particularity for thick cloud removal. In the fifth row of Fig. 4, the reconstruction outcomes of PSTCR are better than those of TNN due to the powerful nonlinear mapping ability via deep learning. However, PSTCR ignores the low-rank characteristic of the multitemporal Sentinel-2 images. This model generates substantial artifacts and blurry textures in the cloudy regions of dates 09/05 and 09/15/2018. For instance, the river in the simulated cloudy regions seems to be blurred by PSTCR in Fig. 4, especially for Sentinel-2 large area cloud removal.

The proposed method outperforms on Sentinel-2 thick cloud removal for the time-series images compared with model-driven method TNN and data-driven method PSTCR. In the sixth row of Fig. 4, the proposed method can not only better maintain the spatial consistency but also ensure the detailed texture without an obvious blurry issue than TNN and PSTCR. The proposed method simultaneously utilizes the low-rank prior of the third-order tensor and the prominent feature expression ability of deep spatio-temporal prior for the multitemporal Sentinel-2 image cloud removal. This notion also testifies the effectiveness of the combined strategy in DP-LRSVD for time-series images reconstruction. In addition, the quantitative evaluation indexes of the Sentinel-2 simulated experiments are listed in Table 1. The proposed method in this work also validates the practicability for multitemporal images cloud removal. This method achieves the best CC, SSIM, RMSE, and SAM compared with the Sentinel-2 cloudy data, HaLRTC, BCPF, TNN, and PSTCR.

b) Multitemporal GF-1 WFV results: Four time-series GF-1 images (acquired times: 07/30, 08/03, 08/07, and 09/09/2015 via GF-1 WFV over Wuhan, China) are imposed with the simulated cloud, as shown in the first row of Fig. 5. The second, third, fourth, fifth, and sixth rows are the reconstructed multitemporal images by HaLRTC, BCPF, TNN, PSTCR, and the proposed method in Fig. 5, respectively. We validate on 16 m spatial resolution bands (NIR, red, green, and blue) in these time-series GF-1 WFV images. The false-color 800 × 800 results of band NIR, red, and green are displayed in Fig. 5. Evaluation indexes CC, SSIM, RMSE, and SAM of the GF-1 WFV simulated experiments are listed in Table 2. The optimal values in Table 2 are marked in bold format for comparison.

Table 1
Evaluation indexes (CC, SSIM, RMSE, and SAM) of the Sentinel-2 simulated experiments through HaLRTC, BCPF, TNN, PSTCR, and the proposed method.

Method	Evaluation Index (average value)			
	CC	SSIM	RMSE	SAM
Cloudy	0.6628	0.7845	0.1983	9.6431
HaLRTC	0.7857	0.8563	0.1246	6.2878
BCPF	0.8486	0.8971	0.0952	1.5613
TNN	0.9553	0.9386	0.0571	1.4984
PSTCR	0.9648	0.9412	0.0509	1.2375
Proposed	**0.9817**	**0.9658**	**0.0383**	**0.9424**

TNN can recover the detailed textures for multitemporal images in the fourth row of Fig. 5. However, TNN poorly performs in terms of spatial context consistency between the simulated cloudy and the neighboring cloud-free regions. This phenomenon indicates that the global consistency and regional particularity are significant for time-series image thick cloud removal. PSTCR behaves better than TNN in terms of spatial context consistency in the fifth row of Fig. 5. Nevertheless, the spectral aberration and spatial blurry details disturb the reconstructing GF-1 results, especially at dates 08/03 and 08/07/2015. These issues also demonstrate the common limitations of deep learning-based methods for thick cloud removal in multitemporal images.

The proposed method can effectively reconstruct the missing area for time-series GF-1 images in the sixth row of Fig. 5. This method performs better in terms of detailed texture recovery and spatial context consistency compared with TNN and PSTCR. DP-LRSVD can fully utilize the low-rank property of the third-order tensor and the feature expression ability of deep spatio-temporal prior by means of the combined model-driven and data-driven methods. Besides, the proposed method achieves the best CC, SSIM, RMSE, and SAM compared with the GF-1 WFV cloudy data, HaLRTC, BCPF, TNN, and PSTCR in Table 2. This notion also validates the reliability of the presented framework for multitemporal image cloud removal.

3.3. Real experiments

In the real experiments, we select the Sentinel-2 and GF-1 cloudy time-series images as the testing objects. Similar to the simulated experiments, the model-driven method TNN and data-driven method PSTCR are also utilized as the contrast algorithms. The detailed descriptions and results are provided as follows:

a) Multitemporal Sentinel-2 results: In the first row of Fig. 6, four time-series cloudy Sentinel-2 images (08/16, 08/31, 09/10, and 09/20/2018 via Sentinel-2A/B L1C over Mechelen, Belgium) are polluted by thick cloud and shadow to a certain degree. We validate on 20 m spatial resolution bands (B12, B11, B8a, B07, B06, and B05) in these time-series Sentinel-2 images. The false-color 500 × 500 results of bands B11, B07, and B05 are displayed in Fig. 6. TNN shows serious artifacts and distortions in the second row of Fig. 6. Besides, TNN poorly performs in terms of spatial context consistency between the cloudy and the neighboring clean regions for the circled area. This notion illustrates the instability for the model-driven method, especially for complex scenarios. PSTCR can well reconstruct the cloudy region in the third row of Fig. 6. Nevertheless, the cloud removal results reveal obvious “block” effects, such as Sentinel-2 data in 08/31 and 09/10/2018 for the circled area. The main reason is that PSTCR employs the patch group strategy for reconstruction. Therefore, the “block” effects are inevitably for PSTCR, especially for the circled area in the third row of Fig. 6.

The proposed method outperforms on time-series image cloud removal in the fourth row of Fig. 6. The proposed method can not only better maintain the spatial consistency but also ensure the detailed texture without obvious blurry issue or “block” effects than TNN and PSTCR compared with the model-driven method TNN and data-driven method PSTCR. Meanwhile, the proposed method simultaneously utilizes the low-rank prior of the third-order tensor and the prominent feature expression ability of deep spatio-temporal prior for multitemporal Sentinel-2 image cloud removal. This notion also testifies the effectiveness of the combined strategy in DP-LRSVD for Sentinel-2 time-series image reconstruction.

b) Multitemporal GF-1 WFV results: Three time-series GF-1 WFV cloudy images (acquired times: 07/30, 08/03, and 08/07/2015 via GF-1 WFV over Wuhan, China) are shown in the first row of Fig. 7. We validate on 16 m spatial resolution bands (NIR, red, green, and blue) in these time-series GF-1 WFV cloudy images. The false-color results of bands

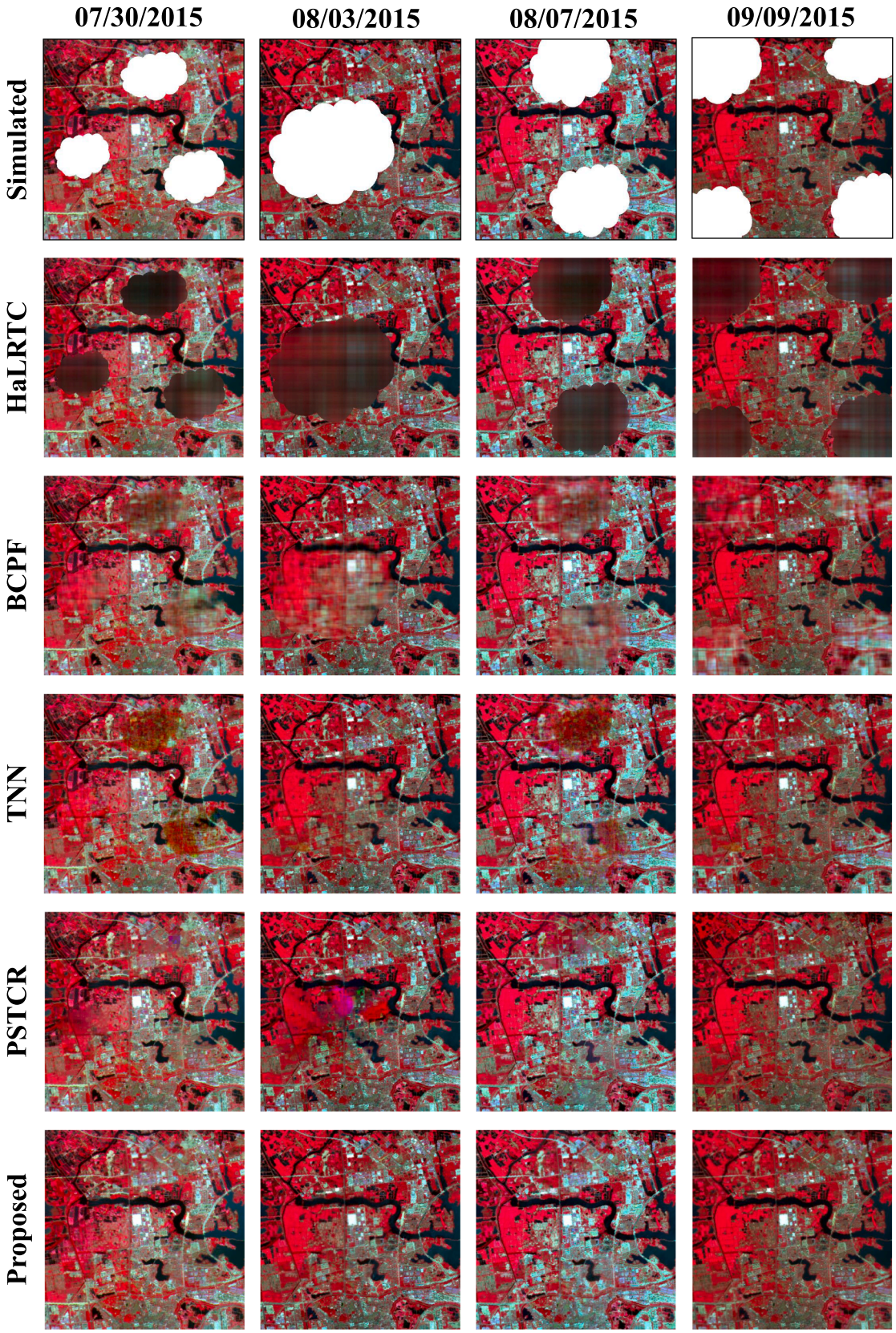

Fig. 5. Simulated cloud removal experiments for the multitemporal GF-1 WFV images (07/30, 08/03, 08/07, and 09/09/2015) through HaLRTC, BCPF, TNN, PSTCR, and the proposed method.

Table 2
Evaluation indexes (CC, SSIM, RMSE, and SAM) of the GF-1 simulated experiments through HaLRTC, BCPF, TNN, PSTCR, and the proposed method.

Method	Evaluation Index (Average value)			
	CC	SSIM	RMSE	SAM
Cloudy	0.6448	0.7535	0.2129	8.2129
HaLRTC	0.7689	0.8346	0.1453	5.2129
BCPF	0.8175	0.8627	0.1264	1.8743
TNN	0.9163	0.8826	0.0837	1.6856
PSTCR	0.9675	0.8943	0.0558	1.5294
Proposed	**0.9842**	**0.9359**	**0.0426**	**1.1828**

NIR, red, and green are displayed in Fig. 7. TNN can recovery the detailed textures for the multitemporal images in the second row of Fig. 7. However, TNN poorly performs in terms of spatial context consistency between the simulated cloudy and the neighboring cloud-free regions, especially in the circled area. PSTCR behaves better than TNN in terms of spatial context consistency in the third row of Fig. 7. Nevertheless, the "block" effects and spatial blurry details disturb the reconstruction GF-1 results, especially at dates 07/30 and 08/07/2015 in the circled area. The proposed method can effectively reconstruct the cloudy area for time-series GF-1 images in the fourth row of Fig. 7. The proposed method performs better in terms of detailed texture recovery and spatial context consistency compared with TNN and PSTCR.

4. Discussions

In this section, discussions in terms of the parameter sensitivity and model availability of the proposed method are presented. Then, the ADMM iteration optimization and 3D-ST Net effects are analyzed for thick cloud removal in time-series images. Finally, we discuss the correlation between the number of multi-temporal scenes and reconstruction effects. The specific discussions are described as follows:

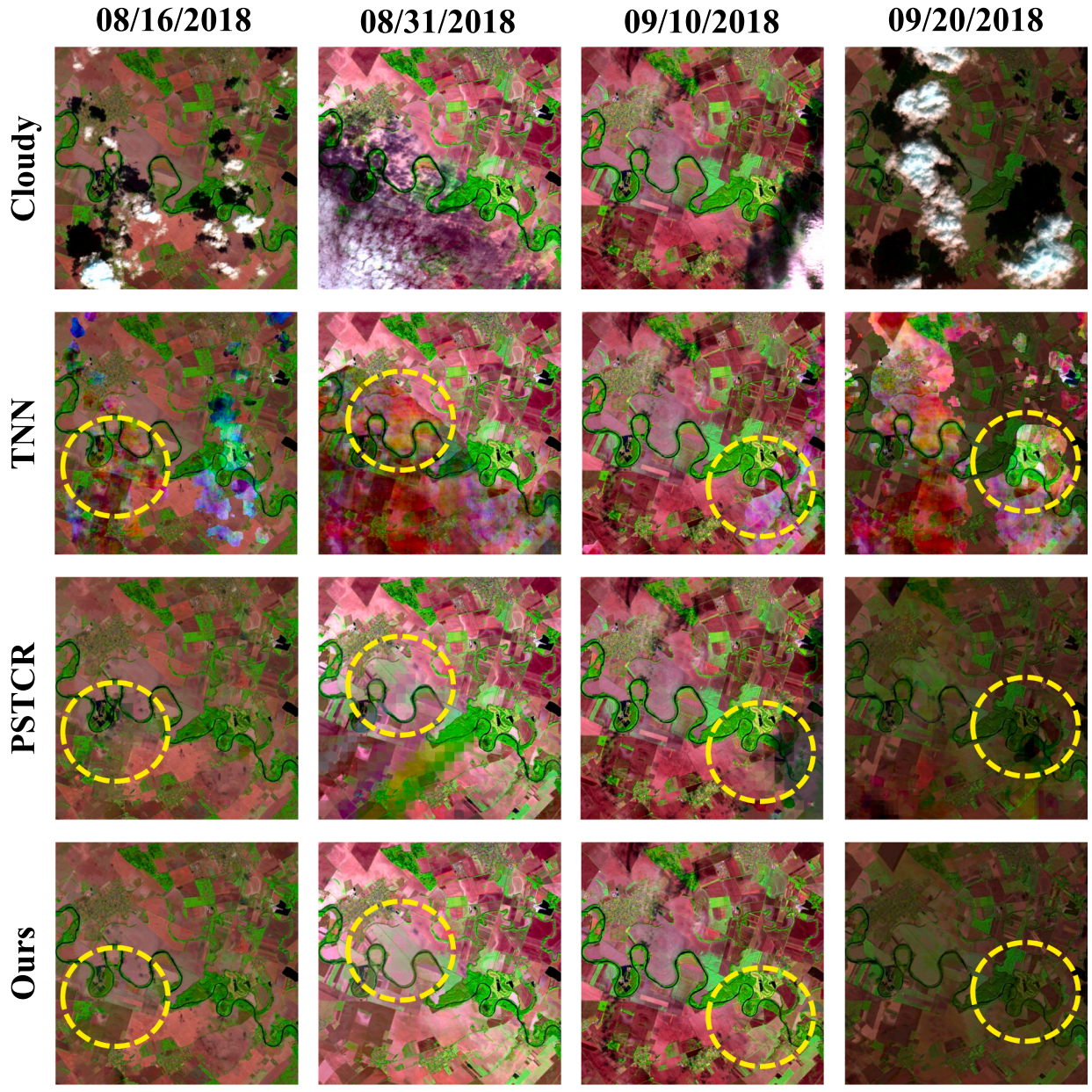

Fig. 6. Real cloud removal experiments for the multitemporal Sentinel-2 images (08/16, 08/31, 09/10, and 09/20/2018) through HaLRTC, BCPF, TNN, PSTCR, and the proposed method.

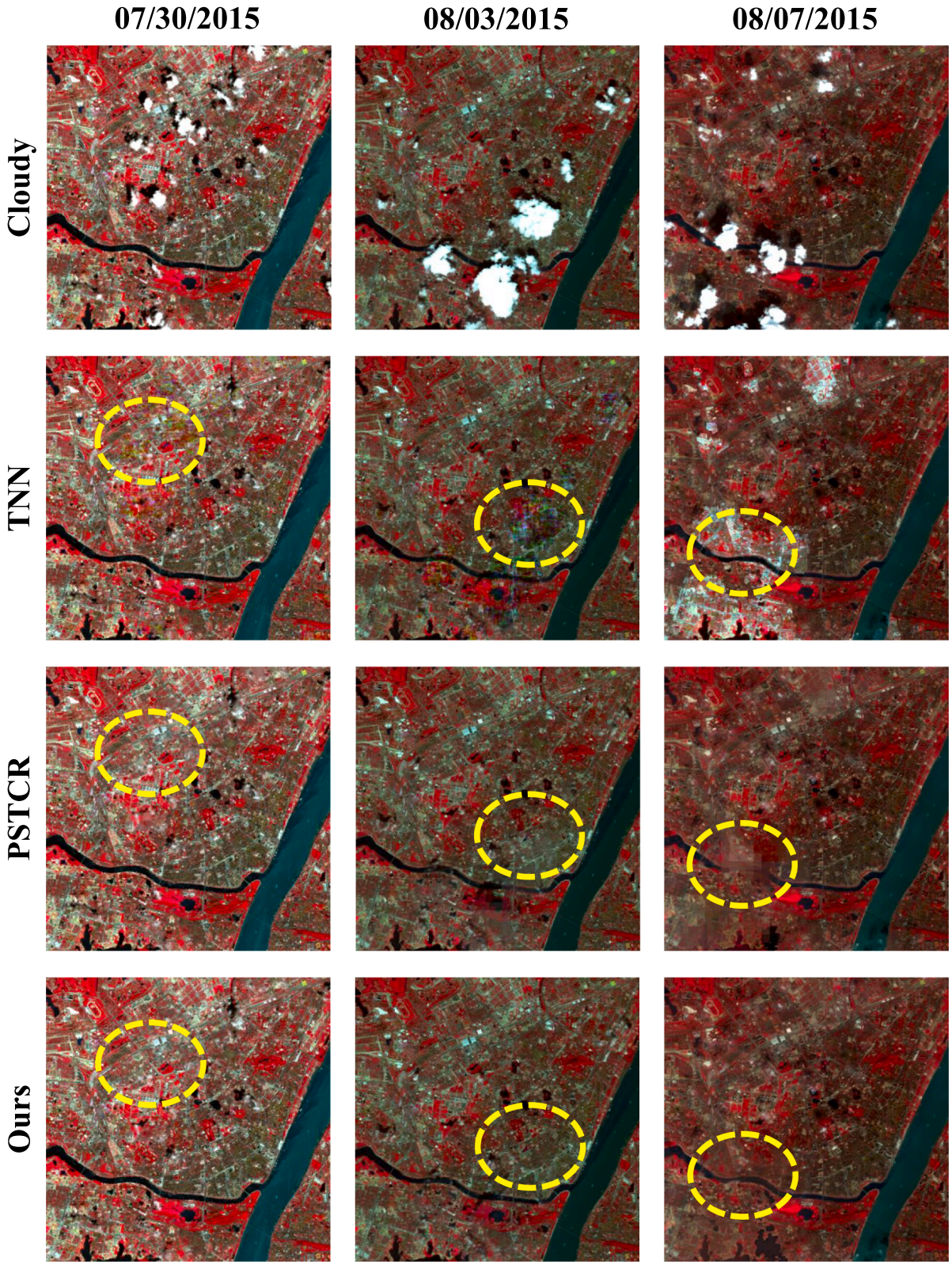

Fig. 7. Real cloud removal experiments for the multitemporal GF-1 WFV images (07/30, 08/03, and 08/07/2015) through HaLRTC, BCPF, TNN, PSTCR, and the proposed method.

4.1. Parameter sensitivity

In the proposed method, we utilize two manual parameter scaling factors, namely, β^0 and step threshold η, for the ADMM iteration framework. Therefore, the parameter sensitivity of the presented method must be discussed. In Fig. 8(a) and (b), the RSME indexes of increasing parameters β^0 and ηin the simulated Sentinel-2 and GF-1 experiments are recorded, respectively. The red lines refer to the outcomes of Sentinel-2, and the blue lines stands for the results of GF-1. Scaling factor β^0 and step threshold η are varied from 0.005 to 0.05

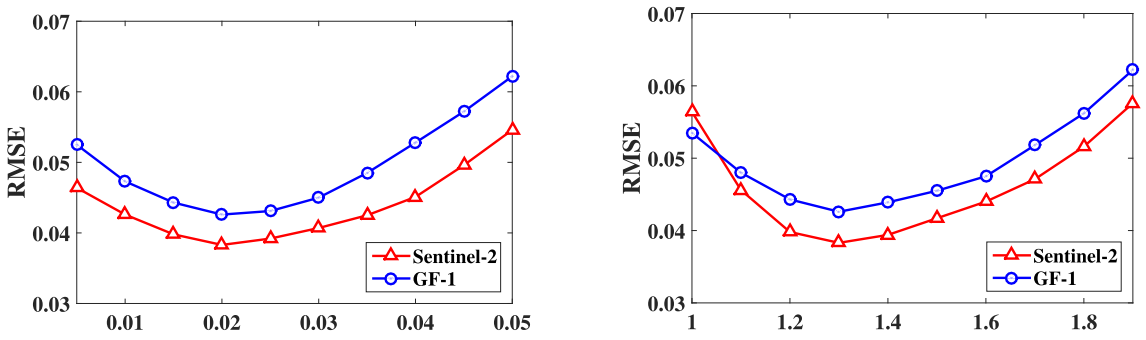

Fig. 8. Parameter sensitivity (scaling factor β^0 and step threshold η) of the proposed method in the two simulated GF-1 and Sentinel-2 experiments.

and 1 to 1.9 in Fig. 8(a) and (b), respectively. Our parameter sensitivity analysis adopts the principle of controlled variables.

The RMSE index first rapidly decreases before $\beta^0 = 0.02$ and $\eta = 1.3$ with the scaling factor β^0 and step threshold ηincreasing in Fig. 8. Afterward, the RMSE index gradually increases in Sentinel-2 and GF-1 experiments. Considering these parameter sensitivities, scaling factor β^0 and step threshold $\eta$are set as 0.02 and 1.3, respectively, in the cloud removal procedure. This situation also demonstrates the significance of the ADMM optimization for the proposed method.

4.2. Iteration optimization

The proposed DP-LRSVD utilizes the iteration optimization strategy and deep spatio-temporal prior for multitemporal image cloud removal. We need to discuss the meaning of ADMM iteration optimization and 3D-ST Net. In Fig. 9(a), the RSME indexes of the increasing iteration number in the simulated Sentinel-2 and GF-1 experiments are recorded. The red lines denote the outcomes of Sentinel-2, and the blue lines represents the results of GF-1. The comparison results with/without 3D-ST Net in the simulated Sentinel-2 images are also displayed in Fig. 9(b).

The RMSE index first rapidly decreases before the five iterations with the increasing iteration number and then gradually decreases until the final iterations in Fig. 9(a). The presented framework could quickly achieve the convergence condition through combing the deep spatio-temporal prior with low-rank tensor factorization. In contrast with the traditional tensor completion, the proposed method develops 3D-ST Net into the ADMM iteration optimization. This method can not only further enhance the context consistency between the cloudy and the adjacent cloud-free regions but also accelerate the ADMM iteration optimization without a time-consuming iteration in Fig. 9(b). This discussion also demonstrates the effectiveness of the proposed method for time-series image reconstruction.

4.3. Number of multitemporal scenes

We supplement a discussion to explore the correlation between the number of multi-temporal scenes and the reconstruction effects when 50% of the areas for each Sentinel-2 image are covered by thick cloud to examine the performance of the proposed method according to the cloud cover rate. The number of multi-temporal scenes is set from 2 to 6, and the reconstruction evaluation indexes (CC) are illustrated in Table 3.

Table 3 illustrates that the reconstruction accuracy gradually improves with the increase in temporal scene number. When 50% of the areas of each Sentinel-2 image are covered by thick cloud, the CC of the proposed method can achieve 0.9519 through inputting three time-series images. The more temporal images are used, the better the thick cloud removal performs. This notion also verifies that the proposed method depends on the temporal complementarity and redundancy for thick cloud removal in multitemporal images. The low rank prior can be effectively reflected through more temporal images in the proposed framework.

5. Conclusion

We combine the deep prior with low-rank tensor decomposition in this study for thick cloud removal in time-series images. The low-rank

Table 3
Correlation between the number of multi-temporal scenes and reconstruction evaluation indexes when 50% of the areas for each Sentinel-2 image are covered by thick cloud.

Number	2	3	4	5	6
CC	0.9136	0.9428	0.9519	0.9532	0.9547

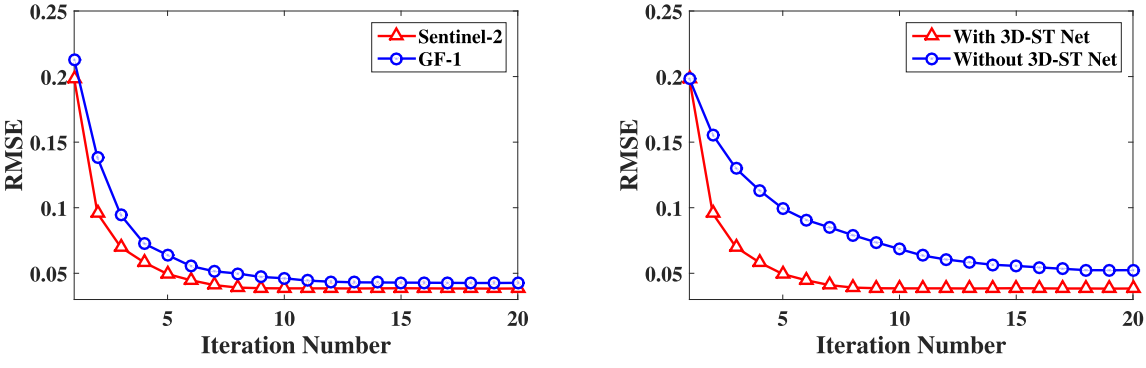

Fig. 9. Discussions for ADMM iteration optimization on the GF-1 and Sentinel-2 images and with/without 3D-ST Net in the proposed method.

tensor SVD and tensor-product are employed for exploiting the structural property of multitemporal images. A deep 3D spatio-temporal CNN is utilized to ensure the context consistency between the cloudy and the adjacent cloud-free regions. ADMM iteration optimization is carried out until the reconstruction result meets the convergence condition. The proposed DP-LRTSVD can effectively eliminate the thick cloud in multitemporal images through combining the model-driven and data-driven strategies. DP-LRTSVD outperforms on thick cloud removal in simulated and real time-series Sentinel-2/GF-1 experiments compared with the model-driven or data-driven methods.

In our future work, more third-order tensor decomposition strategy and spatio-temporal priors will be exploited and introduced. We will also distinguish the cloud types and consider the integrative thin-thick cloud removal through introducing the available information within thin cloud-covered regions.

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.

Acknowledgement

This work was supported in part by the National Key Research and Development Program of China (No. 2016YFB0501403), in part by the National Natural Science Foundation of China (No. 41922008), and in part by China Postdoctoral Science Foundation (No. 2020TQ0229).

References

Bayad, M., Chau, H.W., Trolove, S., Müller, K., Condron, L., Moir, J., Yi, L., 2020. Time series of remote sensing and water deficit to predict the occurrence of soil water repellency in New Zealand pastures. ISPRS J. Photogramm. Remote Sens. 169, 292–300.

Boyd, S., Parikh, N., Chu, E., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. Now Publishers Inc.

Cao, R., Chen, Y., Chen, J., Zhu, X., Shen, M., 2020. Thick cloud removal in Landsat images based on autoregression of Landsat time-series data. Remote Sens. Environ. 249, 112001.

Chan, T.F., Shen, J., 2001. Nontexture inpainting by curvature-driven diffusions. J. Vis. Commun. Image. R. 12 (4), 436–449.

Chen, B., Huang, B., Chen, L., Xu, B., 2016. Spatially and temporally weighted regression: A novel method to produce continuous cloud-free Landsat imagery. IEEE Trans. Geosci. Remote Sens. 55 (1), 27–37.

Chen, J., Jønsson, P., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. Remote Sens. Environ. 91 (3), 332–344.

Chen, J., Zhu, X., Vogelmann, J.E., Gao, F., Jin, S., 2011. A simple and effective method for filling gaps in Landsat ETM+ SLC-off images. Remote Sens. Environ. 115 (4), 1053–1064.

Chen, Y., He, W., Yokoya, N., Huang, T.-Z., 2019. Blind cloud and cloud shadow removal of multitemporal images based on total variation regularized low-rank sparsity decomposition. ISPRS J. Photogramm. Remote Sens. 157, 93–107.

Duan, C., Pan, J., Li, R., 2020. Thick cloud removal of remote sensing images using temporal smoothness and sparsity regularized tensor optimization. Remote Sens. 12 (20), 3446.

Gao, G., Gu, Y., 2017. Multitemporal landsat missing data recovery based on tempo-spectral angle model. IEEE Trans. Geosci. Remote Sens. 55 (7), 3656–3668.

Ji, T.-Y., Yokoya, N., Zhu, X.X., Huang, T.-Z., 2018. Nonlocal tensor completion for multitemporal remotely sensed images inpainting. IEEE Trans. Geosci. Remote Sens. 56 (6), 3047–3061.

Jiang, T., Ng, M.K., Zhao, X., Huang, T., 2020. Framelet representation of tensor nuclear norm for third-order tensor completion. IEEE Trans. Image Process. 29, 7233–7244.

Kong, D., Zhang, Y., Gu, X., Wang, D., 2019. A robust method for reconstructing global MODIS EVI time series on the Google Earth Engine. ISPRS J. Photogramm. Remote Sens. 155, 13–24.

Li, W., Li, Y., Chen, D., Chan, J.C.-W., 2019. Thin cloud removal with residual symmetrical concatenation network. ISPRS J. Photogramm. Remote Sens. 153, 137–150.

Li, Z., Shen, H., Cheng, Q., Li, W., Zhang, L., 2019. Thick cloud removal in high-resolution satellite images using stepwise radiometric adjustment and residual correction. Remote Sens. 11 (16), 1925.

Li, Z., Shen, H., Cheng, Q., Liu, Y., You, S., He, Z., 2019. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. ISPRS J. Photogramm. Remote Sens. 150, 197–212.

Liang, Y., Ouyang, K., Wang, Y., Liu, Y., Zhang, J., Zheng, Y., Rosenblum, D.S., 2020. Revisiting convolutional neural networks for citywide crowd flow analytics. arXiv, 2003.00895.

Liu, H., Zhou, Q., Li, Q., Hu, S., Shi, T., Wu, G., 2019. Determining switching threshold for NIR-SWIR combined atmospheric correction algorithm of ocean color remote sensing. ISPRS J. Photogramm. Remote Sens. 153, 59–73.

Liu, J., Musialski, P., Wonka, P., Ye, J., 2013. Tensor completion for estimating missing values in visual data. IEEE Trans. Pattern Anal. Mach. Intell. 35 (1), 208–220.

Lu, C., Feng, J., Chen, Y., Liu, W., Lin, Z., Yan, S., 2019. Tensor robust principal component analysis with a new tensor nuclear norm. IEEE Trans. Pattern Anal. Mach. Intell. 42 (4), 925–938.

Lu, C., Tang, J., Yan, S., Lin, Z., 2015. Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm. IEEE Trans. Image Process. 25 (2), 829–839.

Luo, D., Huang, H., Ding, C., 2011. Discriminative high order SVD: Adaptive tensor subspace selection for image classification, clustering, and retrieval. In: International Conference on Computer Vision, pp. 1443–1448.

Maalouf, A., Carr, P., Augereau, B., Fernandez, C., 2009. A bandelet-based inpainting technique for clouds removal from remotely sensed images. IEEE Trans. Geosci. Remote Sens. 47 (7), 2363–2371.

Mateo-García, G., Laparra, V., López-Puigdollers, D., Gómez-Chova, L., 2020. Transferring deep learning models for cloud detection between Landsat-8 and Proba-V. ISPRS J. Photogramm. Remote Sens. 160, 1–17.

Meraner, A., Ebel, P., Zhu, X.X., Schmitt, M., 2020. Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. ISPRS J. Photogramm. Remote Sens. 166, 333–346.

Ouyang, K., Liang, Y., Liu, Y., Tong, Z., Ruan, S., Rosenblum, D., Zheng, Y., 2020. Fine-grained urban flow inference. IEEE Trans. Knowl. Data Eng. 1–13.

Paudel, K.P., Andersen, P., 2011. Monitoring snow cover variability in an agropastoral area in the Trans Himalayan region of Nepal using MODIS data with improved cloud removal methodology. Remote Sens. Environ. 115 (5), 1234–1246.

Qiu, C., Mou, L., Schmitt, M., Zhu, X.X., 2019. Local climate zone-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network. ISPRS J. Photogramm. Remote Sens. 154, 151–162.

Qiu, S., Zhu, Z., He, B., 2019. Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. Remote Sens. Environ. 231, 111205.

Sedano, F., Kempeneers, P., Hurtt, G., 2014. A Kalman filter-based method to generate continuous time series of medium-resolution NDVI images. Remote Sens. 6 (12), 12381–12408.

Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., Zhang, L., 2015. Missing information reconstruction of remote sensing data: A technical review. IEEE Geosci. Remote Sens. Mag. 3 (3), 61–85.

Wang, L., Bi, J., Meng, X., Geng, G., Huang, K., Li, J., Tang, L., Liu, Y., 2020. Satellite-based assessment of the long-term efficacy of PM2.5 pollution control policies across the Taiwan Strait. Remote Sens. Environ. 251, 112067.

Wang, Q., Tang, Y., Atkinson, P.M., 2020. The effect of the point spread function on downscaling continua. ISPRS J. Photogramm. Remote Sens. 168, 251–267.

Watmough, G.R., Atkinson, P.M., Hutton, C.W., 2011. A combined spectral and object-based approach to transparent cloud removal in an operational setting for Landsat ETM+. Int. J. Appl. Earth. Obs. Geo. 13 (2), 220–227.

Weng, Q., 2009. Thermal infrared remote sensing for urban climate and environmental studies: Methods, applications, and trends. ISPRS J. Photogramm. Remote Sens. 64 (4), 335–344.

Weng, Q., Fu, P., Gao, F., 2014. Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. Remote Sens. Environ. 145, 55–67.

Xu, H., Tang, X., Ai, B., Gao, X., Yang, F., Wen, Z., 2021. Missing data reconstruction in vhr images based on progressive structure prediction and texture generation. ISPRS J. Photogramm. Remote Sens. 171, 266–277.

Xu, M., Jia, X., Pickering, M., Jia, S., 2019. Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform. ISPRS J. Photogramm. Remote Sens. 149, 215–225.

Xu, M., Jia, X., Pickering, M., Plaza, A.J., 2016. Cloud removal based on sparse representation via multitemporal dictionary learning. IEEE Trans. Geosci. Remote Sens. 54 (5), 2998–3006.

Yu, P., Froyd, K.D., Portmann, R.W., Toon, O.B., Freitas, S.R., Bardeen, C.G., Brock, C., Fan, T., Gao, R.-S., Katich, J.M., et al., 2019. Efficient in-cloud removal of aerosols by deep convection. Geophys. Res. Lett. 46 (2), 1061–1069.

Yuan, Q., Zhang, Q., Li, J., Shen, H., Zhang, L., 2019. Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network. IEEE Trans. Geosci. Remote Sens. 57 (2), 1205–1218.

Zhang, J., Li, Z., Jing, P., Liu, Y., Su, Y., 2019. Tensor-driven low-rank discriminant analysis for image set classification. Multimedia Tools Appl. 78 (4), 4001–4020.

Zhang, Q., Yuan, Q., Li, J., Li, Z., Shen, H., Zhang, L., 2020. Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning. ISPRS J. Photogramm. Remote Sens. 162, 148–160.

Zhang, Q., Yuan, Q., Li, J., Liu, X., Shen, H., Zhang, L., 2019. Hybrid noise removal in hyperspectral imagery with a spatial-spectral gradient network. IEEE Trans. Geosci. Remote Sens. 57 (10), 7317–7329.

Zhang, Q., Yuan, Q., Li, J., Sun, F., Zhang, L., 2020. Deep spatio-spectral Bayesian posterior for hyperspectral image non-i.i.d. noise removal. ISPRS J. Photogramm. Remote Sens. 164, 125–137.

Zhang, Q., Yuan, Q., Li, J., Wang, Y., Sun, F., Zhang, L., 2021. Generating seamless global daily AMSR2 soil moisture (SGD-SM) long-term products for the years 2013–2019. Earth Syst. Sci. Data 13, 1385–1401.

Zhang, Q., Yuan, Q., Li, J., Yang, Z., Ma, X., 2018. Learning a dilated residual network for SAR image despeckling. Remote Sens. 10 (2), 196.

Zhang, Q., Yuan, Q., Zeng, C., Li, X., Wei, Y., 2018. Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network. IEEE Trans. Geosci. Remote Sens. 56 (8), 4274–4288.
Zhang, X., Xiao, P., Feng, X., 2020. Object-specific optimization of hierarchical multiscale segmentations for high-spatial resolution remote sensing images. ISPRS J. Photogramm. Remote Sens. 159, 308–321.
Zhang, X., Xiao, P., Feng, X., He, G., 2019a. Another look on region merging procedure from seed region shift for high-resolution remote sensing image segmentation. ISPRS J. Photogramm. Remote Sens. 148, 197–207.
Zhang, Y., Wen, F., Gao, Z., Ling, X., 2019b. A coarse-to-fine framework for cloud removal in remote sensing image sequence. IEEE Trans. Geosci. Remote Sens. 57 (8), 5963–5974.
Zhang, Z., Aeron, S., 2017. Exact tensor completion using t-SVD. IEEE Trans. Signal Process. 65 (6), 1511–1526.
Zhao, Q., Zhang, L., Cichocki, A., 2015. Bayesian CP factorization of incomplete tensors with automatic rank determination. IEEE Trans. Pattern Anal. Mach. Intell. 37 (9), 1751–1763.
Zhao, X.-L., Xu, W.-H., Jiang, T.-X., Wang, Y., Ng, M.K., 2020. Deep plug-and-play prior for low-rank tensor completion. Neurocomputing 400, 137–149.
Zheng, J., Liu, X., Wang, X., 2020. Single image cloud removal using U-Net and generative adversarial networks. IEEE Trans. Geosci. Remote Sens. 1–15.
Zheng, Y.-B., Huang, T.-Z., Zhao, X.-L., Jiang, T.-X., Ji, T.-Y., Ma, T.-H., 2020. Tensor N-tubal rank and its convex relaxation for low-rank tensor recovery. Infor. Sci. 532, 170–189.
Zhu, X., Gao, F., Liu, D., Chen, J., 2011. A modified neighborhood similar pixel interpolator approach for removing thick clouds in Landsat images. IEEE Geosci. Remote Sens. Lett. 9 (3), 521–525.
Zhu, Z., Wang, S., Woodcock, C.E., 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, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens. Environ. 118, 83–94.