For-backward LSTM-based missing data reconstruction for time-series Landsat images

Ya’nan Zhou a,1, Shunying Wang a, Tianjun Wu b, Li Feng a, Wei Wu a, Jiancheng Luo a,d,e, Xin Zhang a,d,e and Na’na Yan a,d,e

+College of Hydrology and Water Resources, Hohai University, Nanjing, China; *School of Science, Chang’an University, Xi’an, China; †College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China; State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China; *University of Chinese Academy of Sciences, Beijing, China

ABSTRACT

Reconstructing the missing data for cloud/ shadow-covered optical satellite images has great significance for enhancing the data availability and multi-temporal analysis. In this study, we proposed a deep-learning-based method for cloud/ shadow-covered missing data reconstruction for time-series Landsat images. Central to this method is the combined use of autoencoder, long-term memory (AE-LSTM)-based similar pixel clustering and for backward LSTM-based time-series prediction. First, manually delineated cloud/ shadow-covered masks were overlaid onto multi-temporal satellite images to produce pixel-wise time-series data with masking values. Second, these pixel-wise time-series were clustered by an AE-LSTM-based unsupervised method into multiple clusters, for searching similar pixels. Third, for each cluster of target images, a backward-LSTM-based model was established to restore missing values in time series data. Finally, reconstructed data were merged with cloud-free (image) regions to produce cloud-free time-series images. The proposed method was applied onto three datasets of multi-temporal Landsat-8 OLI images to restore cloud/ shadow-covered images. The reconstruction results, showing an improvement of greater than 10% in normalized mean-square error compared to the state-of-the-art methods, demonstrated the effectiveness of the proposed method in time-series reconstruction for satellite images.

ARTICLE HISTORY

Received 29 March 2021
Revised 20 July 2021
Accepted 20 July 2021

KEYWORDS

Missing data reconstruction; cloud and shadow; Landsat image; time series; deep learning

1. Introduction

As one of the most powerful tools for monitoring earth surface, remote-sensing images have been applied in a variety of applications, such as land-cover classification (Gomez, White, and Wulder 2016; Ma et al. 2017), waterbody mapping (Zhou et al. 2014), urbanization monitoring (Feng et al. 2017), and crop forecasting (Zhou et al. 2019). Especially, multi-temporal images can further demonstrate the changes of earth surface and evolutions of ground objects (Zhu and Woodcock 2014). However, due to the influence of the weather, (optical) images acquired from satellite sensors are often contaminated (or covered) by clouds and shadows, especially in the humid tropical areas (Tseng, Tseng, and Chien 2008). In cloudy and shadowy regions, the true ground information is occluded by clouds and shadows and difficult to obtain, which strongly limits the application of optical images (Pouliot and Latifovic 2018). Further, for multi-temporal remote-sensing applications (such as deforestation, agricultural planting, urbanization, and wetland loss), the cloud- and shadow-covered (for convenience, hereafter, we simply group them into cloud-covered) data will produce irregular time intervals, increasing difficulties in further time-series analysis (Wu et al. 2018). Therefore, reconstructing the missing data (regions) of such cloud-covered images has great significance for enhancing the data availability and multi-temporal analysis (Gao and Gu 2017).

Much research effort has been devoted to reconstructing missing data for cloud-covered images, which can be classified into three major categories (Chen, Huang, and Chen et al. 2016): self-complementation-based, reference-complementation-based, and multi-temporal-complementation-based methods. In the self-complementation-based approaches, without the incorporation of other complementary (image...
data, missing data in the cloud-covered regions is reconstructed using the remaining cloud-free regions in the same image. Based on the assumption that the missing regions and remaining regions in the image share the same (or similar) statistical and geometrical structures, the cloud-covered regions are reconstructed by propagating the geometrical structure from the local or nonlocal cloud-free regions. Under theories of spatial interpolation and error propagation, missing pixel interpolation methods (Van der Meer 2012; Pringle, Schmidt, and Muir 2009; Zhu, Liu, and Chen 2012; Zhang, Li, and Travis 2009), image inpainting methods (Lorenzi, Melgani, and Mercier 2011; Cheng et al. 2013), and model fitting methods (Maalouf et al. 2009) have been widely used in reconstructing cloud-covered regions. Although these methods yield visually plausible results, they are sensitive to the land-cover types underneath clouds (Zhang, Wen, and Gao et al. 2019) and make the results unsuitable for data analysis in further applications (Cheng et al. 2014). Furthermore, as uncertainty and error accumulate along with propagation, these methods can hardly deal with large-region or heterogeneous data missing (Zhang, Wen, and Gao et al. 2019).

To overcome the bottleneck of the self-complementation approaches, reference-complementation methods reconstruct missing data by modeling the mapping relationship from auxiliary cloud-free images (or bands) to cloud-covered images (or bands). This category of methods relies on the fact that there is usually high correlation between different spectral data. When cloud-insensitive spectral bands available in multispectral or hyperspectral images, they are utilized to reconstruct the missing data of cloud-covered bands. Based on the close correlation between band 7 and band 6 in moderate resolution imaging spectroradiometer (MODIS) data, various fitting methods (Wang et al. 2006; Shen, Zeng, and Zhang 2010; Rakwatin, Takeuchi, and Yasuoka 2008; Gladkova et al. 2011) were proposed to reconstruct the missing data of MODIS band 6 by using MODIS band 7. Ji (2008) utilized the near infrared band to estimate the spatial distribution of haze intensity in visible bands of Landsat images through a linear regression model established over deep water area. Some studies attempted to employ cloud-free images from different sensors to reconstruct cloud-cover images. For example, Roy et al. (2008) proposed to predict the missing data of Landsat images by incorporating the corresponding MODIS image (with close acquired dates). With the advantage of insensitivity to cloud and rain, synthetic aperture radar (SAR) data were employed to reconstruct cloud-covered optical images (Hoan and Tateishi 2009; Huang et al. 2015; Grohfeldt, Schmitt, and Zhu 2018). Reference images provide much information for reconstructing missing data, but these approaches are usually constrained by the spectral compatibility (Lin et al. 2012), spatial resolution, and temporal coherence, which are incapable of restoring the quantitative products and modeling change information over time (Cheng et al. 2014; Malambo and Heatwole 2015).

With or without reference images, the previous two approaches are limited in that they are formulated with the assumption of no or gradual change in the images reconstructed (Mariethoz, McCabe, and Renard 2012). This inherent stationarity assumption is clearly a weakness in time-series applications when documenting land-cover change and monitoring crop growth (Mariethoz, McCabe, and Renard 2012). Since remote-sensing satellites with a fixed repeat cycle can regularly acquire images in the same area where it is not always cloud-covered all the time, it is easy to get multi-temporal images for the same region. These multi-temporal (cloud-covered or cloud-free) images provide the potential to address the cloud-covered missing data for time-series analysis (Cheng et al. 2014). For reconstructing time-series images, multi-temporal-complementation approaches (most relevant to this work) involve two main steps: searching cloud-free pixels with similar land-cover types to cloud-covered pixels, and predicting cloud-covered pixels by using these cloud-free pixels. In searching similar pixels, incorporating spatial, spectral, and temporal relationships were widely investigated to calculate the similarity between cloud-free pixels and cloud-covered pixels (Gao and Gu 2017; Melgani 2006; Benabdellaker and Melgani 2008; Zhang et al. 2018). In reconstructing cloud-covered pixels, various models were developed. Assuming small change between cloud-covered pixels and cloud-free pixels, Tseng, Tseng, and Chien (2008) simply and directly replaced cloud-covered pixels with cloud-free pixels of the same location from reference images. By solving a group of constrained Poisson equations, Lin et al. (2012) cloned information from cloud-free regions to fill corresponding cloud-covered regions.
Based on the geo-statistical techniques, Chen, Huang, and Chen et al. (2016) introduced a spatially and temporally weighted regression model to reconstruct the missing data of a target scene from both the self-target scene and multi-temporal reference scenes. Padhee and Dutta (2019) proposed a moving offset method, restoring missing values in time-series NDVI at a pixel by using coefficients of linear regression with reference NDVI at another pixel having identical conditions. Furthermore, some multi-temporal methods designed for filling gaps caused by sensor failure can be used for reconstructing cloud-covered pixels, such as the neighborhood similar pixel interpolator (NSPI) method (Chen et al. 2011; Zhu et al. 2011) and the weighted linear regression (WLR) approach (Zeng, Shen, and Zhang 2013).

Recently, some deep-learning-based reconstruction methods have been developed to reconstruct cloud-covered remote-sensing images. Grohfeldt, Schmitt, and Zhu (2018) used generative adversarial networks to fuse SAR data and optical images to generate cloud-free optical images. Malek et al. (2018) attempted to use an autoencoder neural network to find the essential mapping function from cloud-free reference images to cloud-covered target images. Dai, Ji, and Zhang (2020) proposed a gated convolutional neural network (CNN) for cloud removal from bi-temporal images. Zhang et al. (2018), (Zhang et al. 2020) constructed and improved CNN-based frameworks for reconstructing missing data in multi-temporal satellite images caused by sensor failure and thick cloud. For further exploit temporal tendency of multi-temporal observation for series reconstruction, some studies introduced recurrent neural networks (RNN) into fields of geoscience for temporal-spatial interpolation (Ma et al. 2019; Liu et al. 2019; Yoon, Yeeh, and Byun 2020), and prediction (Chang et al. 2020; Qi et al. 2019; Zhou et al. 2020).

Although encouraging reconstruction results have been produced in the previous studies, there are several limitations to these approaches. (1) In addition to spectral and spatial similarity, temporal tendency in multi-temporal images representing land-cover change, was underutilized and hard to be used in previous methods (especially the traditional methods). (2) Available deep-learning-based methods focused mainly on CNNs to capture spectral and spatial information, only a few studies employed RNNs to learn temporal tendency across multi-temporal images. (3) Since clouds may appear anywhere and anytime, pixel-wise time series always has irregular time intervals, which makes it hard to be used in time-series analysis.

The objective of this article is to develop a deep-learning-based method for time-series reconstruction in Landsat images. Inspired by the promising results of RNNs, we tried to use RNNs to learn multi-scale temporal features and tendency for reconstruction. And to overcome the problems of unequal time intervals and unaligned time steps, we introduced the masking layers for time-series reconstruction. We implemented this method through the combined use of autoencoder-LSTM (long-short-term memory, an implementation of RNN)-based similar pixel clustering and for-backward LSTM-based missing data reconstruction. Compared to previous approaches, this study combines two principal contributions. The first contribution is establishing a for-backward LSTM-based framework to incorporate temporal tendencies both from previous and subsequent images for reconstruction. The second one is introducing masking layers to utilize incomplete and unaligned time series to improve reconstruction results.

The proposed method is further discussed and validated through missing data reconstruction on datasets of multi-temporal cloud-covered Landsat-8 OLI images in two study areas, compared to the state-of-the-art methods.

2. Study areas and datasets

The first study area is in the southwestern region of Iberian Peninsula, Europe, straddling Spain and Portugal, with central coordinates of 7°22′W and 38°54′N (as presented in Figure 1), which is covered by World Reference System 2 (WRS2) with Path 203 and Row 33. This region is dominated by Mediterranean climate of hot and dry summer and mild and rainy winter. The vegetation is mainly subtropical evergreen hardwood forest. The study area covers most of landforms and land uses types, including waterbody, forest, grassland, farmland, and urban areas, which is suitable to test the proposed method.

All available 22 Landsat-8 OLI images (hereafter referred to as Landsat images) with P/R 203/33 during 2015 (as listed in Figure 2) were employed to validate the proposed method. Landsat images with a 16-day
revisit comprise seven multi-spectral bands with spatio-spectral resolution of 30 m, from which the widely-used blue, green, red, and near-infrared (NIR) bands (band 2 ~ 5) were selected for experiments. All images were obtained from the USGS website at Level 1 G (the image with DOY 42 was unavailable from the USGS website). There are 15 images contaminated by clouds and shadows (cloud-covered percentage...
from 1% to 100%). And the mean cloud-covered percentage is approximately 33%. Especially, the image with DOY 362 is totally covered by clouds and shadows. Thus, we excluded images with DOY 42 and 362 from reconstruction in experiments, since there is not any reference information.

The second study area is in Haiyang County, Shandong Province, China. Being different from the first study area, it has typical temperate monsoonal climate. It is covered by typical land-cover types including artificial area, water, wetland, farmland, forest land, and grass land. In this area, all available 23 Landsat images with P/R 120/34 during 2017 were employed. They have a mean cloud-covered percentage of 38%.

### 3. Methodology

A for-backward LSTM-based reconstruction method was proposed to restore cloud-covered missing data for time-series Landsat images, as illustrated in Figure 3. The method involved three main steps: (1) pixel-wise time-series data with masking values, (2) autoencoder-LSTM-based time-series pixel clustering, and (3) for-backwards-LSTM-based time-series reconstruction.

Before the main process, data pre-processing was conducted, which involved the geographic registration, pixel alignment and common-region clipping of multi-temporal images. (1) Firstly, all cloud-covered regions were manually delineated to produce cloud-covered masks for each Landsat image (Subsection 3.1). Then these masks were overlaid onto corresponding satellite images to generate pixel-wise time-series data with masking values (Subsection 3.2). (2) These pixel-wise time series were clustered by an AE-LSTM-based unsupervised method into multiple clusters, for searching similar pixels (Subsection 3.3). (3) First, the pixel-wise time-series set was split into training sample set and predicting sample set (Subsection 3.4.1). Then, a for-backward LSTM-based reconstruction model trained by training samples was employed to predict the time-series missing values in predicting samples (Subsection 3.4.2). Finally, reconstructed (predicted) data were merged with cloud-free (image) regions to produce cloud-free time-series Landsat images, and verified by simulated actual images (Subsection 3.5).

#### 3.1. Data pre-processing and cloud mask

Landsat images were employed in this study. Since these data have well-characterized radiometric quality, data pre-processing including geographic registration, pixel alignment and common-region clipping

![Flowchart](image)

**Figure 3.** Flowchart of the LSTM-based reconstruction method for restoring cloud-covered missing data for time-series Landsat images.
were conducted. Firstly, a polynomial model with 2 degrees and 9 manually selected and evenly distributed ground control points were employed to geometrically correct the Landsat image to a Google satellite map (produced from Google Tile Map Service with a spatial resolution of 2.5 m), with registration errors amounted to less than 10 m. Secondly, taking image with DOY 10 as reference, we resampled other images into the same pixel grids as image with DOY 10, for pixel alignment in coordinate. Finally, the common parts of the multi-temporal images were clipped for time-series analysis.

For reconstructing cloud-covered pixels of time-series images, cloudy and shadowy regions were delineated for each image. Firstly, according to the aerosol band and the NIR-red-green band composed image (as presented in Figure 2), we manually delineated cloudy and shadowy regions to generate initial cloud-covered masks for Landsat images. Then, the cloud-covered masks were properly extended to further cover those ambiguous pixels, which were hard to determine being cloud-contaminated or not, around delineated regions. Finally, the cloud-covered mask was converted to a mask band and appended to the corresponding Landsat image, for the following reconstruction.

3.2. Pixel-wise time series with missing values

In the Landsat image, there was a mask band m to indicate whether a pixel p is cloud-covered or not. To produce pixel-wise time-series data, we defined the masked pixel value \( m_p \) of the \( j \) (\( 1 \leq j \leq T \)) time step (\( T \) is the number of time steps), in the \( i \) (\( 1 \leq i \leq B \)) band (\( B \) is the number of bands) and at \((x, y)\) coordinates as:

\[
\begin{align*}
mp_{x,y}^j &= \begin{cases} 
p_{x,y}^j, & if m_{x,y} = 0 \\
mv, & if m_{x,y} = 1.
\end{cases}
\end{align*}
\]

(1)

where \( mv \) is the masking value (a value much more than the biggest pixel value, and we set it to be 19,999 in this study) in time-series data, indicating a missing value caused by clouds.

Then, we constructed pixel-wise time series \( TS \) for the \( i \) band at \((x, y)\) coordinates as:

\[
TS_{x,y}^i = \{ mp_{x,y}^1, mp_{x,y}^2, \ldots, mp_{x,y}^T \}.
\]

(2)

This pixel-wise time series with masking values was the basic element in reconstruction. And just the masking value filled up the missing data and aligned time steps in pixel-wise time series, which prepare data for the following AE-LSTM-based pixel clustering and LSTM-based prediction.

3.3. Autoencoder-LSTM-based pixel clustering

An autoencoder-LSTM-based network was employed for clustering time-series data with masking values (masked pixels), as shown in Figure 4. In the AE-based framework for time-series clustering, we firstly fed pixel-wise time-series from all bands \([TS_{x,y}^1, TS_{x,y}^2, \ldots, TS_{x,y}^B]\) into a masking layer and a stacked LSTM network (with four LSTM layers) to produce the latent high-level features \( z \) of input time-series data (referred as encoder). Then, on one hand, the latent features \( z \) were fed into another stacked LSTM network (with four LSTM layers) to produce reconstructions (referred as decoder), under the constraints of reconstruction loss; on the other hand, the latent features \( z \) were fed into a clustering layer to calculate clustering probabilities of the latent features (referred as clustering), under the constraints of clustering loss. Finally, the reconstruction loss and the clustering loss were incorporated for joint training, as \( L = L_r + aL_c \), where \( L_r \) and \( L_c \) are reconstruction loss and clustering loss, respectively, and \( a \geq 0 \) is the coefficient of clustering loss, to balance \( L_r \) and \( L_c \).

The clustering loss (Xie, Girshick, and Farhadi 2016) is defined as Kullback-Leibler (KL) divergence (measuring the non-symmetric difference) between distributions \( P \) and \( Q \), where \( Q \) is the distribution of soft labels measured by Student’s t-distribution and \( P \) is the target distribution derived from \( Q \). So, the clustering loss is defined as:

\[
L_c = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}
\]

(3)

\[
q_{ij} = \frac{(1 + z_i - \mu_j^2)^{-1}}{\sum_j (1 + z_i - \mu_j^2)^{-1}}
\]

(4)

\[
p_{ij} = \frac{q_{ij}^2 / \sum_j q_{ij}}{\sum_i (q_{ij}^2 / \sum_j q_{ij})}
\]

(5)
where $p_{ij}$ is target distribution, $q_{ij}$ is the similarity between latent features $z$ and cluster center $\mu$ measured by Student’s $t$-distribution.

The reconstruction loss is measured by mean-squared error (MSE):

$$L_r = \frac{1}{n} \sum_{i=1}^{n} (x_i - g(f(x_i)))^2$$  \hspace{1cm} (6)

where $f$ and $g$ are encoder and decoder mappings, respectively. More details can be found in (Xie, Girshick, and Farhadi 2016; Guo, Gao, and Liu et al. 2017; Min et al. 2018).

In details, we set the number of hidden neurons of LSTM layers to be 32, and the coefficient of clustering loss $a$ to be its default value 0.1.

The number of clusters is an open parameter. On one hand, more clusters may improve the homogeneity of clusters, facilitating reconstruction; on the other hand, too many clusters would reduce the pixel numbers in clusters, hampering training reconstruction models. When determining this parameter, we took the practical land-cover types of study areas into consideration. This was, how many land-cover types can be interpreted from Landsat-8 images. For example, there are 10 types in the GLOBELAND30 dataset (Chen, Ban, and Li 2015). For a special region, there are often less than 10 types of land cover. In this study, we set it to be 10, which was also found to produce more better results in simple series experiments with various numbers of clusters.

### 3.4. For-backward LSTM-based reconstruction

#### 3.4.1. Sample sets for model training and prediction

Cloud may appear in any region of the image of any time step. We had to reconstruct missing values at any time step $t$, for any band $b$. For a special time-series curve $TS_{x,y}^b$ at pixel $(x, y)$ and band $b$, we grouped it into the predicting sample set $PS_{x,y}^b$ when $mp_{x,y}^b$ was equal to the masking value $mv$, or into the training sample set $TS_{x,y}^b$. Pixel-wise time series in the training sample set $TS_{x,y}^b$ were used to train the reconstruction model at time step $t$, for band $b$, and the predicting sample set $PS_{x,y}^b$ was time-series data to be reconstructed.

Further, for the for-backward reconstruction model, a time-series curve $TS_{x,y}^b$ from the training or predicting sample set was split at the target time step $t$, into the forward time-series sub curve $ftS_{x,y}^b$ with a length of $t$, the corresponding backward time-series sub curve $btS_{x,y}^b$ with a length of $T - t - 1$, and the target value $mp_{x,y}^b$. Here, we produced a sample for reconstruction models as:
3.4.2. For-backward LSTM-based model structure

Using the traditional LSTM networks, the forward tendency in time-series images can be easily captured to predict cloud-covered pixels. While, it failed to utilize the backward trend information when targets in the middle section. Thus, incorporating the forward and backward trend information, a forward-backward-fusion (for-backward) LSTM network was constructed for better time-series reconstruction, as presented in Figure 5. Firstly, a pixel-wise time series $TS$ was split into a forward time series $fTS$ and a backward time series $bTS$, as demonstrated in the above Subsection. Then, in the forward part, the forward time series $fTS$ was successively fed into a masking layer, a stacked LSTM network, and a dense layer, to produce a forward time-series feature $ff$; and following the similar network structure, the backward time series $bTS$ (in reverse order) produced a backward time-series feature $bf$. Finally, the forward time-series feature $ff$ and the backward time-series feature $bf$ were concatenated and fed into a dense layer, to predict the missing value.

In detail, in the staked LSTM network, four LSTM layers with 32 hidden neurons were stacked to transfer raw time-series curves into high-level features. The length of the forward and backward time-series feature are eight times that of the forward and backward time series, respectively.

3.5. Performance evaluation

3.5.1. Comparative methods

To make a comparative analysis, the proposed LSTM-based method was compared with the state-of-the-art methods of spatially and temporally weighted regression (STWR-based) method (Chen, Huang, and Chen et al. 2016), spectral–temporal patch-based missing area reconstruction (STP-based) (Wu et al. 2018), and pixel-based contextualized AE neural network (PCAE-based) method (Zeng, Shen, and Zhang 2013). The STWR-based method makes full use of cloud-free pixels in the reference time-series cloud-covered scenes and optimally integrates both spatial and temporal information (using spatially and temporally weighted regression) to recover missing pixels (using inverse distance weighted interpolator) in the target scene, which is selected as a traditional geostatistical method for experimental comparison. The STP-based method takes spectral-temporal patches (retrieved from multi-temporal image segmentation) as basic units for searching similar structures (scenes) from cloud-free regions to cloud-covered regions and injecting textural information to preserve spatial and spectral continuities and suppress salt-and-pepper noises and false edges, which is selected as a patch-based method for experimental comparison. The PCAE-based

![Figure 5. The for-backward LSTM network for reconstructing missing data.](image-url)
method employs an autoencoder neural network to calculate the transformation (mapping) function from the cloud-free reference to the target images, to restore missing pixels for multispectral images, which is selected as a deep-learning-based approach for experimental comparison.

3.5.2. Validation datasets and evaluation metrics
Missing data reconstruction performance was assessed using simulated cloud-covered regions, with the advantage of the ability to evaluate the reconstruction against actual observed data. Firstly, we randomly removed patches (with varied shapes, areas, and distributed positions) from target images to generate the simulated cloud-covered images. Then, we compared the reconstructed pixels with the actual pixels in the simulated cloud-covered regions pixel-by-pixel, to calculate reconstruction accuracy.

For pixel \( i \) with an actual value \( x_i \) at a special band, we restored it to be predicted value \( \hat{x}_i \). Its reconstruction error can be calculated as \( e_i = |x_i - \hat{x}_i| \). The smaller the value of \( e \), the higher the accuracy. For the total reconstruction accuracy of cloud-covered regions, we adopted the normalized mean square error (NMSE) index, mean absolute error (MAE) index, mean absolute percentage error (MAPE) index, and correlation coefficient (CC) index for quantitative evaluation.

\[
NMSE = \frac{1}{N} \sum_{i \in R} \frac{(x_i - \hat{x}_i)^2}{u(x)u(\hat{x})} \tag{11}
\]

\[
MAE = \frac{1}{N} \sum_{i \in R} |x_i - \hat{x}_i| \tag{12}
\]

\[
MAPE = \frac{1}{N} \sum_{i \in R} \frac{|x_i - \hat{x}_i|}{x_i} \tag{13}
\]

\[
CC = \frac{\sum_{i \in R} (x_i - u(x))(\hat{x}_i - u(\hat{x}))}{\sqrt{\sum_{i \in R} (x_i - u(x))^2 \sum_{i \in R} (\hat{x}_i - u(\hat{x}))^2}} \tag{14}
\]

where \( N \) is the total number of missing pixels in cloud-covered region \( R \), \( u(x) \) and \( u(\hat{x}) \) are the mean values of the actual and the reconstructed values in the cloud-covered regions, respectively.

NMSE indicates the deviation between the actual observation and the reconstructed results, MAE reflects the actual errors of predicted values, MAPE provides a measure of relative comparison even when data vary in magnitude. CC depicts the textural similarity between the actual observation and the recovered result. For NMSE, MAE and MAPE, lower values indicate better reconstruction and are preferable. A larger CC indicates a closer consistency between the regions of data, and the data will be identical when CC equals 1.

Furthermore, an improvement ratio (IR) index (Shen et al. 2013) is introduced to evaluate the degree of the improvement of the proposed method with respect to comparative methods.

\[
IR = \frac{V_{sota} - V_{prop}}{V_{sota}} \text{ for NMSE, MAE, and MAPE; } IR = \frac{V_{prop} - V_{sota}}{V_{sota}} \text{ for CC} \tag{15}
\]

where \( V_{sota} \) is the best evaluation value in comparative (state-of-the-art) methods, \( V_{prop} \) is the evaluation value of the proposed method.

4. Experiments and discussion
Using multi-temporal Landsat images, experiments were carried out to test the effectiveness and efficiency of the proposed reconstruction method, compared to three state-of-the-art methods. Accuracy assessments were conducted on the reconstruction results of time-series images to discuss the promotion of the proposed LSTM-based model over the comparative methods, the reconstruction performance along cloud-covered percentages and target time steps, and reconstruction error distributions over various land-cover types.

4.1. Reconstruction results in Iberian Peninsula
The time-series reconstructed images (using the NIR-red-green band composition) produced by the proposed LSTM-based method were presented in Figure 6. Since images of DOY 170 and 330 have small cloud-covered regions (smaller than 0.01 in cloud-covered percentages), we do not demonstrate their reconstruction results.

As the whole, the multi-temporal reconstructed images can demonstrate the landscape change process of the study area. For any image in time series, the proposed method has reconstructed missing data in cloud-covered regions with various cloud-covered percentages and distribution positions. Taking a close look at textural and spectral presentation of reconstructed images, we found out that the proposed
method achieved better reconstruction performance on the center images (such as the images with DOY 186, 218, 234 and 250) that on the beginning (such as the images with DOY 90) and the end (such as the image with DOY 346) images of time series. We argued that the for-backwards network can incorporate temporal tendency information both from previous and subsequent time-series images. So, the center image (in time series) could employ more tendency and change information than the beginning and the end images to reconstruct missing data.

Furthermore, we observed that there was serious spectral distortion in the reconstructed images with DOY 106, 122, and 298. The possible reason was that their cloud-covered percentages were too big (cloud-free regions are too small) (approximate 0.92, 0.89,
and 0.91 for the images with DOY 106, 122, and 298, respectively) to provide enough information for restoring cloud-covered regions.

4.2. Comparison and evaluation

4.2.1. Compared results

For comparisons, the three state-of-the-art reconstruction methods (including the STWR-based, the STP-based and the PCALE-based methods) were applied to the multi-temporal image to restore missing data of cloud-covered regions. Since the STWR-based and STP-based methods were multi-temporal based, the whole time series were used to reconstruct the target images. In the PCALE-based method, the image with closer acquired date to that of the image to be reconstructed was taken as the reference (source) image. For example, the image with DOY 266 was referred to the image with DOY 314. Reconstructed images with DOY 74, 202, 266 and error maps of DOY 202 were presented in Figure 7. The error maps indicated the sum of the absolute values of differences between the predicting values and the actual values in the four bands.

Generally, all comparative methods and the proposed method yielded visually plausible images, compared to the actual observation. With a close comparison with the actual cloud-free images and error maps, the proposed method can capture the landscape change (especially for vegetation-covered regions in yellow boxes) over time. Our restored results were shown to be quite close to the actual ones in both spatial details and spectral comparability.

4.2.2. Quantitative evaluation

To quantitatively evaluate reconstruction performance of the proposed LSTM-based approach (comparing with the STWR-based, STP-based, and PCALE-based methods), we carried out four experiments to reconstruct the image with DOY 186 (with a cloud-covered percentage 0.69). In the STWR-based, STP-based and LSTM-based experiments, we employed the whole time series as reference images to restore the image with DOY 186. In the PCALE-based experiment, the image with DOY 202 was taken as reference image. The accuracy assessment of reconstruction results was listed in Table 1.

Overall, the LSTM-based method achieved the best reconstruction accuracy, compared with the three state-of-the-art methods. This was expected, when comparing to the PCALE-based method (only using one reference image), the LSTM-based method incorporated more landscape change information from the time-series images to reconstruct cloud-covered regions. When comparing to the STWR-based and STP-based methods, the LSTM-based method used the for-backward network to learn multi-scale and nonlinear time-series information to reach the landscape change patterns.

Furthermore, reconstruction accuracies of the LSTM-based method varied greatly in various spectral bands. The green band presented the best reconstruction accuracies (0.6029 in the NMSE, 32.1560 in the MAE, 4.88% in MAPE, and 0.9572 in CC), followed by the blue and red bands, and the NIR band produced the worst results. This was consistent with the results of previous studies (Chen, Huang, and Chen et al. 2016). On one hand, clouds make the greatest impact on the blue band with the shortest wavelength. On the other hand, red band and NIR band mainly demonstrate more complex temporal tendency information of vegetation (widely appearing in the study area) growth, which is hard to precisely capture by reconstruction models. And the LSTM-based method achieved the greatest promotion in the red band (with 20%, 27%, 7%, 3% IR in the NMSE, MAE, MAPE, and CC, respectively). It further indicated the proposed method could extract and model the vegetation tendency and change patterns, compared to the comparative methods.

4.2.3. Computational complexity

Since the four reconstruction methods are different from principles, technological processes, and data preparation, it is hard to evaluate and compare their computational complexity of the whole produces in theory. So, in this study, we just focused on their key steps, and compared their time consumptions and resource occupancies on memory and GPU. In the PCALE-based method, a global mapping function (an AE network) from reference images to target images was established for reconstruction. And in the proposed LSTM-based method, there were 10 global time-series predicting functions (LSTM networks) for the 10 clusters. So, the training and predicting computation of the AE network and the 10 LSTM networks...
represented the computation complexities of the PCAE-based and LSTM-based methods, respectively. While, in the STWR-based and STP-based methods, each cloud-covered pixel or patch was reconstructed by a local regression function. Thus, we summed the training and predicting consumption of all cloud-covered pixels or patches, as the total.

Following the same experimental settings as in the above subsection, we applied the four methods to reconstruct the missing data of the image of DOY 186. Table 2 presented their time consumptions and resource occupancies, on a computer with 32GB memory and 4GB memory of GPU. Values before and after “+” were training and predicting consumptions, respectively. It was noted that the training times of the PCAE-based and LSTM-based were time consumptions for one epoch, and approximately 20 and 30 epochs were taken to achieve stable performance, respectively.

Since the PCAE-based and LSTM-based methods employed deep learning algorithms, they used more time (especially the training time), memory and GPU resources. And the STWR-based and STP-based methods did not use GPU. Furthermore, the PCAE-based
method utilized more pixels from neighbors and multi-temporal images, than the LSTM-based method (only using pixels from multi-temporal images), for reconstruct a cloud-covered pixel. Thus, the PCAE-based method occupied more resources than the LSTM-based method.

4.3. Discussion on reconstruction performance

The for-backward LSTM network was used to reconstruct cloud-covered images. To further understand the performance of this method, we conducted a series of experiments on whether the for-backward structure was useful, the minimum cloud-free percentage (or the maximum cloud-covered percentage) for reconstructing missing pixels for the image with a special step, whether the center image in time series presented better reconstruction results, and reconstruction performance on various land-cover types.

### 4.3.1. Forward vs. backward vs. for-backward LSTM

In this study, a for-backward LSTM model was designed to extract both forward and backward time-series information to improve reconstruction performance. Here, we carried out a comparative experiment to test the advantage of the for-backward model over a forward and a backward model (as presented in Figure 3, just using the forward or backward part). Aim to reconstruct the cloud-covered image with DOY 186 (this step time is at the center of the time series), we fed a forward time series from DOY 10 to DOY 170 into the forward LSTM model, fed a backward time series from DOY 362 to DOY 202 into the backward LSTM model, and fed the whole time series into the for-backward LSTM model. Accuracy assessment of reconstruction results produced from the three models were presented in Table 3.

Overall, the for-backward LSTM model achieved the best performance, in all evaluation index, for all spectral bands, compared to the forward and backward models. This was expected, since the for-backward model can incorporate the previous (from the forward time series) and the subsequent tendency information (from backward time series) to model the landscape change pattern for reconstructing missing pixels.

### 4.3.2. Cloud-covered percentage vs. reconstruction performance

Landsat images may be covered by clouds of various areas (cloud-covered percentage from 0% to 100%). We employed cloud-free regions to reconstruct cloud-covered regions. Here, we carried out a series of experiments (taking the cloud-free image with DOY 170 as target image) to discuss whether various

---

**Table 1.** Accuracy assessment of reconstruction results in the comparative experiments.

|      | STWR | STP | PCAE | LSTM | IR (%) |
|------|------|-----|------|------|--------|
| Blue | NMSE | 1.0860 | 1.1063 | 1.0961 | 0.9593 | 11.67 |
|      | MAE  | 38.4256 | 40.0724 | 39.1762 | 34.1562 | 11.11 |
|      | MAPE | 7.0539 | 6.9451 | 6.8646 | 6.1283 | 10.73 |
|      | CC   | 0.9032 | 0.8807 | 0.8915 | 0.9346 | 3.48 |
| Green| NMSE | 0.7494 | 0.9171 | 0.6831 | 0.6029 | 11.74 |
|      | MAE  | 40.3169 | 42.5354 | 38.4329 | 32.1560 | 16.33 |
|      | MAPE | 5.3901 | 5.7463 | 5.4308 | 4.8841 | 9.39 |
|      | CC   | 0.9217 | 0.8867 | 0.9261 | 0.9572 | 3.36 |
| Red  | NMSE | 1.8850 | 1.8484 | 2.1013 | 1.4821 | 19.82 |
|      | MAE  | 69.8642 | 78.3371 | 67.8327 | 49.2581 | 27.38 |
|      | MAPE | 8.1597 | 8.7937 | 8.9738 | 8.1894 | 6.87 |
|      | CC   | 0.8808 | 0.8423 | 0.8951 | 0.9216 | 2.96 |
| NWIR | NMSE | 7.6751 | 6.3309 | 6.2370 | 5.7624 | 7.61 |
|      | MAE  | 156.4461 | 154.7382 | 176.4823 | 118.2696 | 23.57 |
|      | MAPE | 17.3284 | 16.5881 | 16.7976 | 14.8089 | 10.73 |
|      | CC   | 0.7961 | 0.8187 | 0.7896 | 0.8537 | 4.285 |

---

**Table 2.** Computational complexity of the four reconstruction methods. * indicating times in one epoch.

|      | STWR | STP | PCAE | LSTM |
|------|------|-----|------|------|
| Timing/s | 378 + 78 | 297 + 116 | 1870*+582 | 570*+196 |
| Memory/GB | 0.3 + 0.4 | 0.5 + 0.8 | 2.1 + 2.5 | 1.8 + 2.1 |
| GPU/GB | / | / | 3.2 + 3.6 | 1.0 + 1.2 |

---

**Table 3.** Accuracy assessment of reconstruction results produced from the forward, backward, and for-backward LSTM models.

|      | Backward | Forward | For-backward |
|------|----------|---------|--------------|
|      | NMSE | MAE | MAPE | CC | NMSE | MAE | MAPE | CC | NMSE | MAE | MAPE | CC |
| Blue | 2.8026 | 36.3328 | 9.6273 | 0.8673 | 3.3875 | 37.3584 | 10.8450 | 0.8805 | 0.9771 | 34.1094 | 7.0427 | 0.9027 |
| Green| 1.1994 | 39.0973 | 7.1844 | 0.9148 | 1.4682 | 38.9068 | 7.6234 | 0.9076 | 0.5735 | 32.6492 | 5.1876 | 0.9481 |
| Red  | 2.6458 | 51.0496 | 10.8735 | 0.8647 | 3.7065 | 53.2561 | 11.4183 | 0.8514 | 1.6908 | 46.7392 | 9.3291 | 0.9294 |
| NIR  | 27.0126 | 149.8232 | 29.2901 | 0.7949 | 20.5649 | 168.5934 | 30.2819 | 0.7683 | 9.4307 | 115.9650 | 17.4318 | 0.8419 |
cloud-covered percentages impacted the reconstruction performance, and the maximum cloud-covered percentage that can be restored. Since the image with DOY 170 is cloud-free, we firstly generated a series of simulated cloud-covered images with various cloud-covered percentage from 10% to 100% at an interval of 10% (and 5% for better presenting the change of performance). Then, we used the cloud-free regions to restore the cloud-covered regions. Finally, we compared the reconstructed image with the actual image to test the reconstruction performance. Figure 8 presented the change of reconstruction accuracy (NMSE and MAPE) in the blue, green, and red bands (since the NIR band produced the worst reconstruction results, we didn’t discuss it here), along cloud-covered percentage increasing.

Along with decreasing of cloud-free percentages, reconstruction performance gradually decreased. When cloud-free percentages decreasing from 90% to 40%, accuracies kept stable (for example, MAPE of green band was around 5.0) and presented very small decrease (less than 0.2 for MAPE in green band). When cloud-free percentages less than 40% (from 40% to 5%), reconstruction accuracies decreased sharply (from 5.2 to 14.0 in MAPE of green band). Taking a close look at the accuracy change of the red band, its decrease appeared earlier (approximate around 50% in cloud-free percentage), and showed sharper decreases from 20% to 5%, compared to the green band and the blue band. This further verified the conclusion in subsection 3.2, that too small cloud-free regions can’t provide enough information for reconstructing cloud-covered regions. Here, we concluded that the maximum cloud-covered percentage for producing good reconstruction was 60%. When the cloud-covered percentage was greater than 60%, it was hard for the proposed method to achieve satisfactory results. Thus, in the following discussion, we focused on reconstruction of images with cloud-free percentages greater than 40%.

4.3.3. Reconstruction performance along time steps
The proposed for-backward method employed a sub forward time series and a sub backward time series to predict cloud-covered missing data. The number of images was 22 in this study, and the total number of forward and backward steps was 21. For a special step, the numbers of forward and backward time-series images were $t$ and $21-t$. While for different time steps, how about the reconstruction accuracy of the proposed method with various forward and backward steps? Here, we carried out a series of experiments to reconstruct images at different time steps. Under the conclusion from the previous experiments that the minimum cloud-free percentage of the target image for reconstruction is 0.4, we selected those images with cloud-free percentages greater than 0.4 as target images, including images with DOY 26, 58, 74, 90, 138, 154, 170, 186, 202, 218, 234, 250, 266, 314, 330, and 346. Figure 9 presented the reconstruction performance at different time steps.

Overall, when the time step was close to the center of the time series (such as DOY 186 and 234), the for-backward model achieved better reconstruction.

![Figure 8. Reconstruction accuracy as a function of the cloud-covered percentages of target image. NMSE-B, NMSE-G, NMSE-R, indicate the NMSE values in the blue, green, red band of the reconstruction results, respectively. MAPE-B, MAPE-G, and MAPE-R indicate the MAPE values in the blue, green, red band, respectively.](image-url)
performance. When the step was away from the center (such as DOY 74 and 330), the reconstruction accuracy got worse. We argued that the change of ground surface was continuous and gradual, and the for-backward model can incorporate the trend information from both forward and backward steps in time-series images. When the target image was at the beginning or the end of time series, the for-backward model degenerated into backward or forward models, and produced worse reconstruction results. Thus, when using the for-backward model to reconstruct cloud-covered images, we had better place the target image in the center of the time series for better reconstruction. For example, if we want to reconstruct images in October, multi-temporal images from April to March of next year will be better than that from January to December. Taking a close look at the reconstruction performance before and after DOY 186 (the center of the time series), we found out that reconstruction after DOY 186 (from DOY 202 to 234) were slightly better than that before DOY 186 (from DOY 138 to 170). One possible reason was that there were more continuous and usable (cloud-free percentage greater than 0.4) images after DOY 186.

Furthermore, determining the minimum number of time steps that are required for stable results was beneficial to increase the repeatability of the proposed method. Here, the change pattern of performance in Figure 9 inspired us to search the suitable time steps of time-series images. In the proposed method, we only utilized the trend relationship (without periodic component in multiple-year observations) in time-series images. If we took the minimum length of forward time series and backward time series as the length of time series, the accuracy changes decreased when the length greater than 8 (DOY from 170 to 234) in Figure 9. From this perspective, we could argue the minimum length for better performance is 17 (8 + 1 + 8), in Iberian Peninsula. However, it was still noted that land-cover types, cloud-covered percentages and spatial-temporal distribution can also affect the reconstruction performance, in addition to the length of time series.

4.3.4. Error distributions vs. land cover types

Various land-cover types present different change patterns in multi-temporal images and different sensitivity to cloud covering (Wu et al. 2018). In this experiment, we attempted to assess the reconstruction performance of the proposed method for various land-cover types (including water, forest, grassland, farmland, and urban area). To reconstruct the image with DOY 202, we firstly randomly removed grid patches (covering a preponderant land cover, such as forest) to simulate cloud-covered regions (approximate 20% areas for each land-cover type), and then we employed the remaining cloud-free regions to reconstruct the cloud-covered regions. Finally, we compared the actual image with the reconstructed results pixel by pixel, to calculate error distributions (sum of errors from four bands) across images.

Figure 10 presented the sub actual images and
The proposed method produced more and greater reconstruction errors for artificial objects (farmland and urban area) than natural objects (water, forest, and grassland). Specifically, water area presented smaller and uniform reconstruction errors across images, which was consistent with the conclusion of Wu’s study (Wu et al. 2018). Since waterbody appeared low spectral radiance in satellite images and small changes during one year, the proposed method produced smaller absolute errors for water areas. Due to less affected by human activities, forest (especially in alpine and gorge regions, as presented in Figure 10) is homogeneous in spectral band. Furthermore, forest in satellite images was mainly dominated by the green band, which showed better

**Figure 10.** Reconstruction error distributions for various land-cover types.
reconstruction performance than other bands. Thus, forest shows smaller errors. Grassland almost followed the similar rules from forest. In theory, mainly dominated by vegetation, farmland should present the similar error distribution with forest. However, farmland is artificial, and agricultural activities (such as multiple cropping) would change or even destroy the temporal tendency of vegetation. Thus, farmland presented greater reconstruction error. What’s more, error distribution was parcel-based, pixels from one parcel always presented similar error values. Compared with water area and forest, urban areas in middle-resolution Landsat images have more mixed pixels (with water, vegetation, or bare land) and more land-cover changes resulted from human activities. So, urban areas (especially in the development areas) presented more greater errors that other land-cover types.

4.4. Compared results in Haiyang County

To further verify the performance of the proposed LSTM-based method, one more experiment was carried out in Haiyang County, Shandong Province, China. The proposed and three comparative reconstruction methods were applied to the multi-temporal images to restore cloud-covered images. Reconstruction results and their actual images with DOY 10, 250, and 298 were presented in Figure 11.

Generally, the four methods restored the missing data in cloud-covered regions, and the proposed LSTM-based method produced the most similar results with the actual images, compared with the comparative methods. Furthermore, the proposed method presented the time-series change of water areas, in the middle of image of DOY 10, and in the right bottom of image of DOY 298. We argued that using the whole time series could effectively capture the dynamic changes of ground surface, which produced the better reconstruction. This proved the performance of the proposed method.

Also, following the procedure of quantitative evaluation on reconstruction results, we used MAPE to conduct a simple accuracy assessment on the restored green bands from the four methods, as listed in Table 4.

Similarly, the proposed LSTM-based method achieved a great improvement, comparing to the three comparative methods, which was constant with the study in Iberian Peninsula. This further proved the performance of the proposed method. In MAPE of time steps, reconstruction results in DOY 250 had the highest scores, followed by results in DOY 298, and the results in DOY 10 had the lowest scores. This is expected, since DOY 10 was on the beginning of time-series images. When further comparing these results to that in Iberian Peninsula, it was found, this experiment achieved worse performance. The possible reason was this study area had a greater mean cloud-covered percentage than that in Iberian Peninsula.

5. Conclusion

Reconstructing the missing data for cloud/shadow-covered optical images is great important for enhancing the data availability and multi-temporal analysis. We proposed a deep-learning-based method for reconstructing cloud/shadow-covered missing data for time-series images, through the combined use of the AE-LSTM-based similar pixel clustering and the for-backward LSTM-based missing data reconstruction. The method was applied onto a dataset of multi-temporal cloud-covered Landsat-8 OLI images. The reconstruction results, showing a greater improvement relative to the state-of-the-art methods, demonstrated the effectiveness of the proposed method in missing data reconstruction for time-series satellite images. From further discussions on the performance of the proposed framework with various directional tendency, with various cloud/shadow-covered percentages, on different time steps, and for various land-cover types, this study concludes: (1) the for-backward reconstruction model presents better results, compared to the forward and backward models; (2) cloud/shadow-covered percentage of the target image had better be less than 60% for a good reconstruction; (3) the target image around the center of time series could achieve better reconstruction results than that at the beginning and end; and (4) the reconstruction method achieves better performance in natural ground object (such as water area and forest) and the blue band.

Despite these encouraging results, more work is needed to further explore the use of deep learning for time-series reconstruction for satellite images, such as (1) incorporating satellite data
Figure 11. Results of missing data reconstruction in the proposed and comparative experiments. Panel A lists the overviews of reconstruction results (in the NIR-red-green band composition) of the LSTM-based method, and the text below the subfigures indicates the acquisition date (in Day of Year) and cloud-covered percentage of the image. Panel B lists the actual cloud-free Landsat images acquired on Jan. 10, Sep. 7, and Oct. 25, in 2017, highlighted by yellow boxes. Panel C-F list the corresponding detailed results using the STWR-based, STP-based, PCAE-based, and the proposed LSTM-based methods, respectively.

Table 4. MAPE of reconstruction results in the comparative experiments in Haiyang County, China.

| DOY | STWR     | STP     | PCAE    | LSTM    |
|-----|----------|---------|---------|---------|
| 10  | 20.7682  | 19.9618 | 17.2657 | 13.9486 |
| 250 | 7.6931   | 9.2637  | 8.7032  | 7.0108  |
| 298 | 11.3263  | 10.8734 | 11.7589 | 8.8549  |
from other sensors to further reconstruct completely cloud-covered images (such as the image with DOY 362); (2) designing an integrated deep-learning-based network to join the similar pixel clustering and the missing data reconstruction together.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

This work was supported in part by the National Key Research and Development Program under Grant No.s [2021YFB3901300, 2019YFC1804301], the National Natural Science Foundation of China under Grant No. [42071316], Open Fund of State Key Laboratory of Remote Sensing Science under Grant No. [OFSLRSS201919], the Fundamental Research Funds for the Central Universities under Grant No. [B20202008], the Opening Foundation of Key Lab of Spatial Data Mining & Information Sharing, Ministry of Education (Fuzhou University) under Grant No. [2019LSDMIS04], the Chongqing Agricultural Industry Digital Map Project under Grant No. 21C00346.

**ORCID**

Ya’nan Zhou http://orcid.org/0000-0002-4880-6439
Xin Zhang http://orcid.org/0000-0003-0394-7972

**Data availability statement**

The data that support the findings of this study are available upon request by contact with the corresponding author, or accessed through https://earthexplorer.usgs.gov.

**References**

Benabdellkader, S., and F. Melgani. 2008. “Contextual Spatiostemporal Postreconstruction of Cloud-contaminated Images.” *IEEE Geoscience and Remote Sensing Letters* 5 (2): 204–208. doi:10.1109/LGRS.2008.915596.

Chang, Y. S., H. T. Chiao, S. Abimannan, Y.-P. Huang, Y.-T. Tsai, K.-M. Lin. 2020. “An LSTM-based Aggregated Model for Air Pollution Forecasting.” *Atmospheric Pollution Research* 11 (8): 1451–1463. doi:10.1016/j.apr.2020.05.015.

Chen, J., Y. Ban, and S. Li. 2015. “China: Open Access to Earth Land-cover Map.” *Nature* 514 (7523): 434.

Chen, B., B. Huang, L. Chen, Xu, B. 2016. “Spatially and Temporally Weighted Regression: A Novel Method to Produce Continuous Cloud-Free Landsat Imagery.” *IEEE Transactions on Geoscience and Remote Sensing* 55 (1): 27–37.

Chen, J., X. Zhu, J. E. Vogelmann, F. Gao, S. Jin. 2011. “A Simple and Effective Method for Filling Gaps in Landsat ETM+ SLC-off Images.” *Remote Sensing of Environment* 115 (4): 1053–1064. doi:10.1016/j.rse.2010.12.010.

Cheng, Q., H. Shen, L. Zhang, P. Li. 2013. “Inpainting for Remotely Sensed Images with a Multichannel Nonlocal Total Variation Model.” *IEEE Transactions on Geoscience and Remote Sensing* 52 (1): 175–187. doi:10.1109/TGRS.2012.2237521.

Cheng, Q., H. Shen, L. Zhang, Q. Yuan, C. Zeng. 2014. “Cloud Removal for Remotely Sensed Images by Similar Pixel Replacement Guided with a Spatio-temporal MRF Model.” *ISPRS Journal of Photogrammetry and Remote Sensing* 92: 54–68. doi:10.1016/j.isprsjprs.2014.02.015.

Dai, P., S. Ji, and Y. Zhang. 2020. “Gated Convolutional Networks for Cloud Removal from Bi-Temporal Remote Sensing Images.” *Remote Sensing* 12 (20): 3427. doi:10.3390/rs12203427.

Feng, L., S. Guo, L. Zhu, Z. Yanan, L. Di. 2017. “Urban Vegetation Phenology Analysis Using High Spatio-temporal NDVI Time Series.” *Urban Forestry & Urban Greening* 25: 43–57. doi:10.1016/j.ufug.2017.05.001.

Gao, G., and Y. Gu. 2017. “Multitemporal Landsat Missing Data Recovery Based on Tempo-spectral Angle Model.” *IEEE Transactions on Geoscience and Remote Sensing* 55 (7): 3656–3668. doi:10.1109/TGRS.2017.2656162.

Gladkova, I., M. D. Grossberg, F. Shahriar, G. Bonev, P. Romanov. 2011. “Quantitative Restoration for MODIS Band 6 on Aqua.” *IEEE Transactions on Geoscience and Remote Sensing* 50 (6): 2409–2416. doi:10.1109/TGRS.2011.2173499.

Gomez, C., J. C. White, and M. A. Walder. 2016. “Optical Remotely Sensed Time Series Data for Land Cover Classification: A Review.” *ISPRS Journal of Photogrammetry and Remote Sensing* 116: 55–72. doi:10.1016/j.isprsjprs.2016.03.008.

Grohnfeldt, C., M. Schmitt, and X. Zhu. 2018. “A Conditional Generative Adversarial Network to Fuse SAR and Multispectral Optical Data for Cloud Removal from Sentinel-2 Images.” *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain. IEEE, 1726–1729.*

Guo, X., L. Gao, X. Liu, Yin, J. 2017. “Improved Deep Embedded Clustering with Local Structure Preservation.” *IJCAI, Melbourne, Australia, pp. 1753–1759.*

Hoan, N. T., and R. Tateishi. 2009. “Cloud Removal of Optical Image Using SAR Data for ALOS Applications. Experimenting on Simulated ALOS Data.” *Journal of the Remote Sensing Society of Japan* 29 (2): 410.

Huang, B., Y. Li, X. Han, Y. Cui, W. Li, R. Li. 2015. “Cloud Removal from Optical Satellite Imagery with SAR Imagery Using Sparse Representation.” *IEEE Geoscience and Remote Sensing Letters* 12 (5): 1046–1050. doi:10.1109/LGRS.2014.2377476.
Ji, C. Y. 2008. "Haze Reduction from the Visible Bands of Landsat TM and ETM+ Images over a Shallow Water Reef Environment." Remote Sensing of Environment 112 (4): 1773–1783. doi:10.1016/j.rse.2007.09.006.

Lin, C. H., P. H. Tsai, K. H. Lai, J.-Y. Chen. 2012. "Cloud Removal from Multitemporal Satellite Images Using Information Cloning." IEEE Transactions on Geoscience and Remote Sensing 51 (1): 232–241. doi:10.1109/TGRS.2012.2197682.

Liu, H., Z. Liu, H. Dong, J. Ge, Z. Yuan, J. Zhu, H. Zhang, et al. 2019. "Recurrent Neural Network-based Approach for Sparse Geomagnetic Data Interpolation and Reconstruction." IEEE Access 7: 33173–33179. doi:10.1109/ACCESS.2019.2903599.

Lorenzi, L., F. Melgani, and G. Mercier. 2011. "Inpainting Strategies for Reconstruction of Missing Data in VHR Images." IEEE Geoscience and Remote Sensing Letters 8 (5): 914–918. doi:10.1109/LGRS.2011.2141112.

Ma, J., Y. Ding, J. C. P. Cheng, F. Jiang, Z. Wan. 2019. "A Temporal-spatial Interpolation and Extrapolation Method Based on Geographic Long Short-Term Memory Neural Network for PM2.5." Journal of Cleaner Production 237: 117729. doi:10.1016/j.jclepro.2019.117729.

Ma, L., M. Li, X. Ma, L. Cheng, P. Du, Y. Liu. 2017. "A Review of Supervised Object-based Land-cover Image Classification". ISPRS Journal of Photogrammetry and Remote Sensing 130: 277–293. doi:10.1016/j.isprsjprs.2017.06.001.

Maitouf, A., P. Carré, B. Augereau, C. Fernandez-Maloigne. 2009. "A Bandelet-Based Inpainting Technique for Clouds Removal from Remotely Sensed Images." IEEE Transactions on Geoscience and Remote Sensing 47 (7): 2363–2371. doi:10.1109/TGRS.2008.2010454.

Malambo, L., and C. D. Heatwole. 2015. "A Multitemporal Profile-Based Interpolation Method for Gap Filling Nonstationary Data." IEEE Transactions on Geoscience and Remote Sensing 54 (1): 252–261.

Malek, S., F. Melgani, Y. Bazi, N. Alajlan. 2018. "Reconstructing Cloud-contaminated Multispectral Images with Contextualized Autoencoder Neural Networks." IEEE Transactions on Geoscience and Remote Sensing 56 (4): 2270–2282. doi:10.1109/TGRS.2017.2777886.fg

Mariazgho, M., F. M. McCabe, and P. Renard. 2012. "Spatiotemporal Reconstruction of Gaps in Multivariate Fields Using the Direct Sampling Approach." Water Resources Research 48 (10). doi:10.1029/2012WR012115.

Melgani, F. 2006. "Contextual Reconstruction of Cloud-contaminated Multitemporal Multispectral Images." IEEE Transactions on Geoscience and Remote Sensing 44 (2): 442–455. doi:10.1109/TGRS.2005.861929.

Min, E., X. Guo, Q. Liu, G. Zhang, J. Cui, J. Long. 2018. "A Survey of Clustering with Deep Learning: From the Perspective of Network Architecture". IEEE Access 6: 39501–39514. doi:10.1109/ACCESS.2018.2855437.

Padhee, S. K., and S. Dutta. 2019. "Spatio-temporal Reconstruction of MODIS NDVI by Regional Land Surface Phenology and Harmonic Analysis of Time-series." GIScience & Remote Sensing 56 (8): 1261–1288. doi:10.1080/15481603.2019.1646977.

Poulhiet, D., and R. Latifovic. 2018. "Reconstruction of Landsat Time Series in the Presence of Irregular and Sparse Observations: Development and Assessment in North-eastern Alberta, Canada." Remote Sensing of Environment 204: 979–996. doi:10.1016/j.rse.2017.07.036.

Pringle, M. J., M. Schmidt, and J. S. Muir. 2009. "Geostatistical Interpolation of SLC-off Landsat ETM+ Images." ISPRS Journal of Photogrammetry and Remote Sensing 64 (6): 654–664. doi:10.1016/j.isprsjprs.2009.06.001.

Qi, Y., Q. Li, H. Karimian, D. Liu. 2019. "A Hybrid Model for Spatiotemporal Forecasting of PM2.5 Based on Graph Convolutional Neural Network and Long Short-term Memory". Science of the Total Environment 664: 1–10. doi:10.1016/j.scitotenv.2019.01.033.

Rakwatin, P., W. Takeuchi, and Y. Yasuoka. 2008. "Restoration of AQUA MODIS Band 6 Using Histogram Matching and Local Least Squares Fitting." IEEE Transactions on Geoscience and Remote Sensing 47 (2): 613–627. doi:10.1109/TGRS.2008.2003436.

Roy, D. P., J. Ju, P. Lewis, C. Schaaf, F. Gao, M. Hansen, E. Lindquist, et al. 2008. "Multi-temporal MODIS–Landsat Data Fusion for Relative Radiometric Normalization, Gap Filling, and Prediction of Landsat Data." Remote Sensing of Environment 112 (6): 3112–3130. doi:10.1016/j.rse.2008.03.009.

Shen, H., P. Wu, Y. Liu, T. Ai, Y. Wang, X. Liu. 2013. "A Spatial and Temporal Reflectance Fusion Model considering Sensor Observation Differences." International Journal of Remote Sensing 34 (12): 4367–4383. doi:10.1080/01431161.2013.777488.

Shen, H., C. Zeng, and L. Zhang. 2010. "Recovering Reflectance of AQUA MODIS Band 6 Based on Within-class Local Fitting." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 4 (1): 185–192. doi:10.1109/JSTARS.2010.2077620.

Tseng, D. C., H. T. Tseng, and C. L. Chien. 2008. "Automatic Cloud Removal from Multi-temporal SPOT Images." Applied Mathematics and Computation 205 (2): 584–600. doi:10.1016/j.amc.2008.05.050.

Van der Meer, F. 2012. "Remote-sensing Image Analysis and Geostatistics." International Journal of Remote Sensing 33 (18): 5644–5676. doi:10.1080/01431161.2012.666363.

Wang, L., J. J. Qu, X. Xiong, X. Hao, Y. Xie, N. Che. 2006. "A New Method for Retrieving Band 6 of Aqua MODIS." IEEE Geoscience and Remote Sensing Letters 3 (2): 267–270. doi:10.1109/LGRS.2006.869966.

Wu, W., L. Ge, J. Luo, R. Huan, Y. Yang. 2018. "A Spectral–Temporal Patch-Based Missing Area Reconstruction for Time-Series Images." Remote Sensing 10 (10): 1560. doi:10.3390/rs10101560.

Xie, J., R. Girshick, and A. Farhadi. 2016. "Unsupervised Deep Embedding for Clustering Analysis." International conference on machine learning, New York City, NY, USA, pp. 478–487.

Yoon, D., Z. Yeeh, and J. Byun. 2020. "Seismic Data Reconstruction Using Deep Bidirectional Long Short-term Memory with Skip Connections." IEEE Geoscience and Remote Sensing Letters 18 (7): 1298–1302.
Zeng, C., H. Shen, and L. Zhang. 2013. “Recovering Missing Pixels for Landsat ETM+ SLC-off Imagery Using Multi-temporal Regression Analysis and a Regularization Method.” Remote Sensing of Environment 131: 182–194. doi:10.1016/j.rse.2012.12.012.

Zhang, C., W. Li, and D. J. Travis. 2009. “Restoration of Clouded Pixels in Multispectral Remotely Sensed Imagery with Cokriging.” International Journal of Remote Sensing 30 (9): 2173–2195. doi:10.1080/01431160802549294.

Zhang, Y., F. Wen, Z. Gao, Ling, X. 2019. “A Coarse-to-Fine Framework for Cloud Removal in Remote Sensing Image Sequence. IEEE Transactions on Geoscience and Remote Sensing 57 (8): 5963–5974.

Zhang, Q., Q. Yuan, J. Li, Z. Li, H. Shen, L. Zhang. 2020. “Thick Cloud and Cloud Shadow Removal in Multitemporal Imagery Using Progressively Spatio-temporal Patch Group Deep Learning”. ISPRS Journal of Photogrammetry and Remote Sensing 162: 148–160. doi:10.1016/j.isprsjprs.2020.02.008.

Zhang, Q., Q. Yuan, C. Zeng, X. Li, Y. Wei. 2018. “Missing Data Reconstruction in Remote Sensing Image with a Unified Spatial–temporal–spectral Deep Convolutional Neural Network.” IEEE Transactions on Geoscience and Remote Sensing 56 (8): 4274–4288. doi:10.1109/TGRS.2018.2810208.

Zhou, Y., J. Luo, L. Feng, Y. Yang, Y. Chen, W. Wu. 2019. “Long-short-term-memory-based Crop Classification Using High-resolution Optical Images and Multi-temporal SAR Data.” GIScience & Remote Sensing 56 (8): 1170–1191. doi:10.1080/15481603.2019.1628412.

Zhou, Y., J. Luo, Z. Shen, X. Hu, H. Yang. 2014. “Multiscale Water Body Extraction in Urban Environments from Satellite Images.” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7 (10): 4301–4312. doi:10.1109/JSTARS.2014.2360436.

Zhou, Y., X. Yang, L. Feng, W. Wu, T. Wu, J. Luo, X. Zhou, et al. 2020. “Superpixel-based Time-series Reconstruction for Optical Images Incorporating SAR Data Using Autoencoder Networks.” GIScience & Remote Sensing 57 (8): 1005–1025. doi:10.1080/15481603.2020.1841459.

Zhu, X., F. Gao, D. Liu, J. Chen. 2011. “A Modified Neighborhood Similar Pixel Interpolator Approach for Removing Thick Clouds in Landsat Images.” IEEE Geoscience and Remote Sensing Letters 9 (3): 521–525. doi:10.1109/LGERS.2011.2173290.

Zhu, X., D. Liu, and J. Chen. 2012. “A New Geostatistical Approach for Filling Gaps in Landsat ETM+ SLC-off Images.” Remote Sensing of Environment 124: 49–60. doi:10.1016/j.rse.2012.04.019.

Zhu, Z., and C. E. Woodcock. 2014. “Continuous Change Detection and Classification of Land Cover Using All Available Landsat Data.” Remote Sensing of Environment 144: 152–171. doi:10.1016/j.rse.2014.01.011.