Check for updates

Reconstructing long-term global satellite-based soil moisture data using deep learning method

Yifan Hu[1], Guojie Wang[1]*, Xikun Wei[1], Feihong Zhou[1],
Giri Kattel ⓘ [1,2,3], Solomon Obiri Yeboah Amankwah[1],
Daniel Fiifi Tawia Hagan[1] and Zheng Duan[4]

[1]Collaborative Innovation Center on Forecast and Evaluation of Metcorological Disasters, Nanjing University
of Information Science and Technology, Nanjing, China, [2]Department of Infrastructure Engineering, The
University of Melbourne, Melbourne, VIC, Australia, [3]Department of Hydraulic Engineering, Tsinghua
University, Beijing, China, [4]Department of Physical Geography and Ecosystem Science, Lund University, Lund,
Sweden

Soil moisture is an essential component for the planetary balance between land surface water and energy. Obtaining long-term global soil moisture data is important for understanding the water cycle changes in the warming climate. To date several satellite soil moisture products are being developed with varying retrieval algorithms, however with considerable missing values. To resolve the data gaps, here we have constructed two global satellite soil moisture products, i.e., the CCI (Climate Change Initiative soil moisture, 1989–2021; CCI_{ori} hereafter) and the CM (Correlation Merging soil moisture, 2006–2019; CM_{ori} hereafter) products separately using a Convolutional Neural Network (CNN) with autoencoding approach, which considers soil moisture variability in both time and space. The reconstructed datasets, namely $CCIr_{ec}$ and CM_{rec}, are cross-evaluated with artificial missing values, and further againt *in-situ* observations from 12 networks including 485 stations globally, with multiple error metrics of correlation coefficients (R), bias, root mean square errors (RMSE) and unbiased root mean square error (ubRMSE) respectively. The cross-validation results show that the reconstructed missing values have high R (0.987 and 0.974, respectively) and low RMSE (0.015 and 0.032 m^3/m^3, respectively) with the original ones. The *in-situ* validation shows that the global mean R between CCI_{rec} (CCI_{ori}) and *in-situ* observations is 0.590 (0.581), RMSE is 0.093 (0.093) m^3/m^3, ubRMSE is 0.059 (0.058) m^3/m^3, bias is 0.032 (0.037) m^3/m^3 respectively; CM_{rec} (CM_{ori}) shows quite similar results. The added value of this study is to provide long-term gap-free satellite soil moisture products globally, which helps studies in the fields of hydrology, meteorology, ecology and climate sciences.

1 Introduction

Soil moisture is one of the most important variables affecting hydrological and climatic processes by affecting the exchanges of water and energy between the Earth's surface and the atmosphere (Schwingshackl et al., 2017). Soil moisture can influence the short-term weather and long-term climatic processes by altering the physical, chemical, and biological interactions at land-atmosphere interface (Jung et al., 2010; Mittelbach et al., 2012; Yang et al., 2016; Sang et al., 2021). Being the major actor of vegetation photosynthesis, soil moisture affects the carbon cycle processes of land surface (Seneviratne et al., 2010). It can also change the surface albedo

and soil thermal properties, consequently affecting the boundary layer properties, such as cloud, precipitation, and evapotranspiration (Ma et al., 1999; Cook et al., 2006; Guan et al., 2009).

Soil moisture has been widely investigated to understand drought variability (Zhang et al., 2021a; Liu et al., 2021; Lu et al., 2021), flood episodes (Wasko and Nathan, 2019; Eeckman et al., 2020), crop yields (Rossato et al., 2017; Champagne et al., 2019), and forest fires (Rigden et al., 2020; Thomas Ambadan et al., 2020). However, studies have indicated significant soil moisture variability over time and space due to climate, biology, soil properties, topography, and human activities such as cultivation, fertilization, irrigation and consequently new soil-formation (Korres et al., 2015). Therefore, it is of great significance to obtain accurate soil moisture variations at local, regional and global scales for many related hydrological, meteorological and ecological applications.

Soil moisture data can be obtained from model simulations, site measurements and satellite observations (Wang et al., 2019). The model simulation often has strong uncertainty in accuracy, which mostly relies on the land surface models to simulate the soil moisture processes (Hu and Lv, 2015; Raoult et al., 2018). Observational soil moisture is more reliable and can assist correcting biases or evaluating the model simulated data (Huang et al., 2008; Yang et al., 2021). Generally, the *in-situ* measurement outperforms satellite observations with higher accuracy, less affected by atmospheric and vegetation conditions (Zhu et al., 2017). However, the site observations have sparse distribution and are heavily limited by manpower and other resources, which thus cannot well capture the soil moisture variability of large spatial scales (Crow et al., 2012). Presently, the satellite remote sensing technology has become the most effective tool for large-scale soil moisture monitoring (Zeng et al., 2015). Compared with *in-situ* measurements, satellite-based observations can capture the soil moisture variability at regional and global scales with lower costs (Yee et al., 2017). The microwave remote sensing has the advantages of all-weather observation, making it a promising approach for retrieving soil moisture of land surface (Stamenkovic et al., 2017). Nevertheless, the satellite service life limits the temporal coverage of soil moisture products (Karthikeyan et al., 2017). Meanwhile, there are considerable data gaps in satellite soil moisture products due to scan width, dense vegetation, atmospheric conditions, and also sensor failures, etc. (Draper, 2018). There are also different qualities among these products due to the used sensors and retrieval algorithms (Kim et al., 2015a). Some studies have used data merging technology make long-term soil moisture retrievals from different satellites, but there are still considerable data gaps in the merged products (Liu et al., 2009; Liu et al., 2012; Wagner et al., 2012; Kim et al., 2015b).

It is necessary to reconstruct the data gaps in satellite Earth observations, and there are several such studies on hydrological and meteorological variables (Long et al., 2014; Alvera-Azcárate et al., 2016; Liu et al., 2017; Cui et al., 2020; Sun et al., 2020). For instance, Wang et al. (2012) used a three-dimensional discrete cosine transform (DCT-PLS) approach to fill the data gaps in satellite soil moisture product (Wang et al., 2012); and Jing et al. (2018) used a random forest (RF) model to estimate the missing soil moisture values. Methods such as general regression neural network (GRNN) and generalized linear model are also tested in the Tibet Plateau (Cui et al., 2019) and the Midwestern United States (Llamas et al., 2020). Generally, good performances can be achieved when it is a regional study with relatively a small portion of missing values in the original data. However, there are still challenges to effectively reconstruct the

missing values in large scale satellite products especially at global scales (Guevara et al., 2021).

With the rapid increase of the data amount and computing capacity, studies have tried to use the deep learning technology for reconstructing Earth observations (Ma et al., 2019), which can efficiently learn the spatio-temporal features and resolve the non-linear relationships among the data (Nogueira et al., 2018). Although the deep learning approach is believed to have significant advantages, it is mainly applied to reconstruct soil moisture, land or sea surface temperature with coarse resolution and regional scales. For instance, Fang et al. (2017) used a Long-Short-Term-Memory network to reconstruct soil moisture of United States; Wu et al. (2019) used a multi-scale feature connected convolutional neural network to reconstruct land surface temperature of Europe and Eastern China; Barth et al. (2020) used a Data INterpolating Convolutional network to reconstruct sea surface temperature with lower error and higher variability. Although there are several deep learning approaches to fill the data gaps in Earth observations, limited work is done to reconstruct long-term global soil moisture from satellites (Zhang et al., 2021b).

In this study, we have adopted a convolutional neural network (CNN) with autoencoding approach to reconstruct two soil moisture retrievals from satellites for the sake of long-term gap-free daily products. For this, a large database of *in-situ* global soil moisture measurements from 12 networks, consisting of 485 stations totally, is used to evaluate the reconstructed products with multiple error metrics. Such gap-free global products are urgently needed by scientific researches of multi-disciplinary fields including hydrology, meteorology, ecology and climate sciences.

2 Materials and methods

2.1 Data description

In our study, two global daily soil moisture products, namely the climate change initiative soil moisture prodcut (CCI$_{ori}$) (Dorigo, et al., 2017; Gruber et al., 2019; Preimesberger et al., 2020) and the correlation merging soil moisture product (CM$_{ori}$) (Hagan et al., 2020), are used. Both datasets are daily products merged from multiple satellites and gridded at 0.25° horizontal resolution. The applied datasets have been widely validated using *in-situ* observations, and they are indicated to be amongst the best satellite soil moisture products with good data promising (Dorigo, et al., 2017; Gruber et al., 2019; Hagan et al., 2020; Preimesberger et al., 2020). The CCI data is combined with active and passive microwave sensors from multiple satellites. The passive microwave soil moisture data include the Scanning Multichannel Microwave Radiometer (SMMR), the Special Sensor Microwave Imager (SSM/I), the Tropical Rainfall Measuring Mission Microwave Imager (TRMM/TMI) (Owe et al., 2008), the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) (Njoku et al., 2003), the WindSat (Parinussa et al., 2011), the Soil Moisture Ocean Salinity (SMOS) (Kerr et al., 2001), the Advanced Microwave Scanning Radiometer 2 (AMSR2) (Parinussa et al., 2015), the Soil Moisture Active Passive (SMAP) (Entekhabi et al., 2010), the Global Precipitation Measurement Microwave Imager (GMP/GMI) (Hou et al., 2014) and the Fengyun-3 (FY-3) (Yang et al., 2012) respectively. We used the most recent CCI version 7.1 combined soil moisture data. The CM
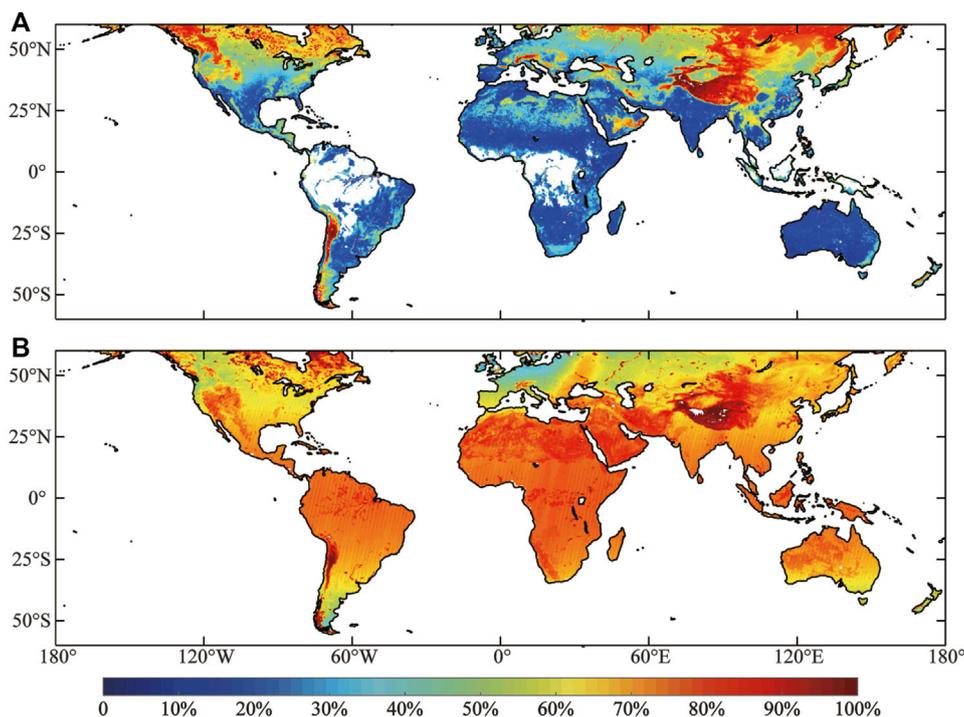
FIGURE 1
Percentages of missing values in **(A)** CCI$_{ori}$ data during 1989−2021, and **(B)** CM$_{ori}$ data during 2006−2019.
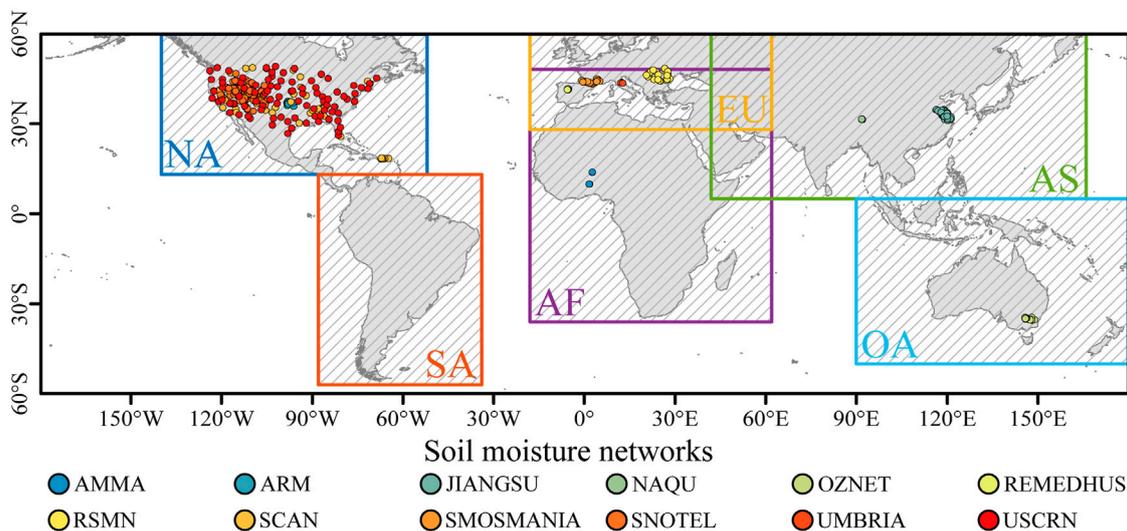

FIGURE 2
The partitioned regions for model training and the used 485 *in-situ* stations monitoring soil moisture. These stations are from 12 networks; JIANGSU network is from China and the other are from the ISMN project.

data is merged from six passive microwave products, including SSM/I, TRMM/TMI, AMSR-E, WindSat, AMSR2 and FY-3B satellites. The CM data is also merged using the CCI scheme with however different merging criteria. CCI merges data by minimizing the error while CM merges data by maximizing the correlation (Hagan et al., 2020). In this study, the CCI ranges from 1989 to 2021 and the CM ranges from

2006 to 2019 respectively. Spatially, we have limited both data to 60°S–60°N, owing to the frozen soil (Ran et al., 2021) and ice cover issues at high latitudes.

The percentages of missing data in CCI$_{ori}$ and CM$_{ori}$ are shown in Figure 1. There are high percentages of missing data in highland mountains and mid-high latitudes, mainly due to frozen soil in winter.

FIGURE 3
Structure of the used CNN model with autoencoding approach.

FIGURE 4
Shown in **(A, B)** are CCI$_{ori}$ and CM$_{ori}$ respectively; **(C, D)** are those with artificial missing values at random and in rectangular regions; CCI$_{rec}$ **(E)** CM$_{rec}$ **(F)** are their respective reconstructions (unit: m^3/m^3).
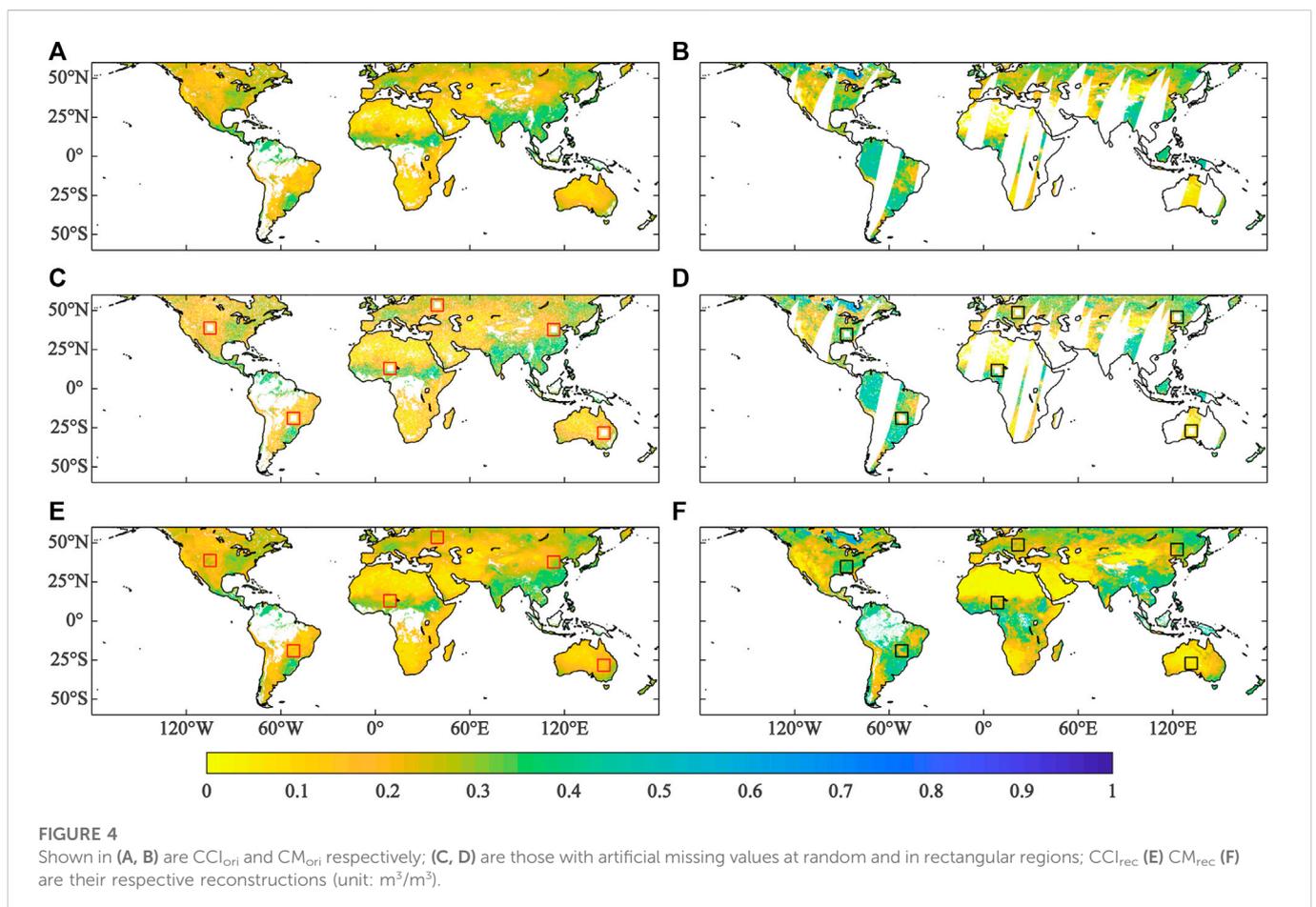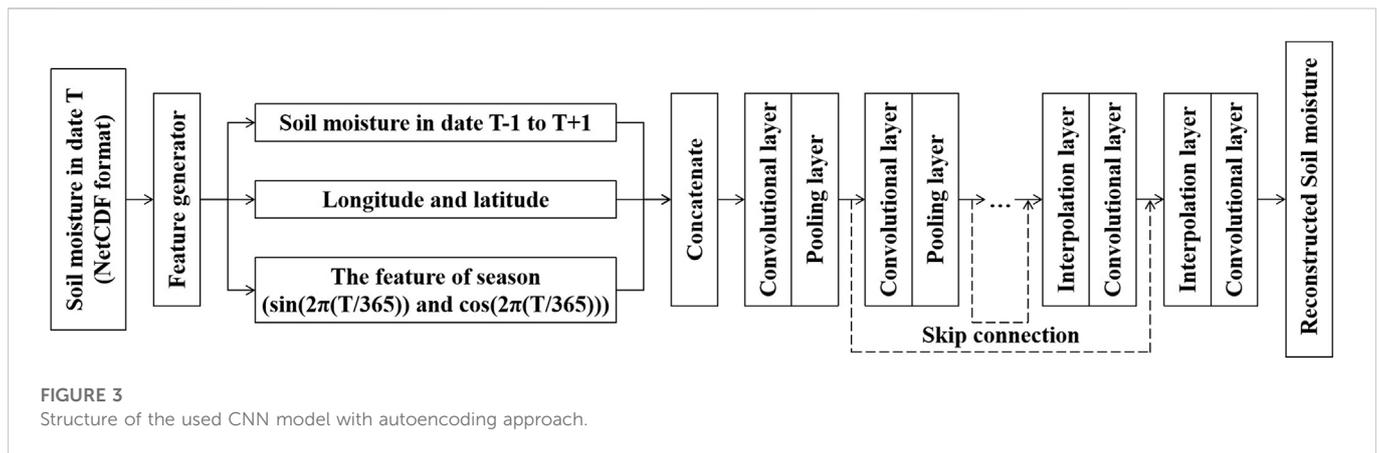
Apparently, there are more missing values in the CM$_{ori}$ than the CCI$_{ori}$ data in most regions of the globe.

2.2 Data preprocessing

Due to different meteorological and terrestial conditions, the spatio-temporal soil moisture patterns are different among continents (Liu et al., 2018). We therefore partition the global land into six regions as shown in Figure 2, considering two key aspects: 1)

The oceans have to be eliminated from the model training so as to minimize its impact on terrestrial soil moisture; 2) these partitions have some overlapping regions, so that the artifacts at edges can be reduced during post-processing (Barth et al., 2020).

2.3 The deep learning (DL) model

Generally, traditional reconstruction methods can only consider temporal information for time series reconstruction or only spatial

FIGURE 5
Selected regions with random and rectangular artificial missing values and their respective reconstructions (unit: m³/m³).
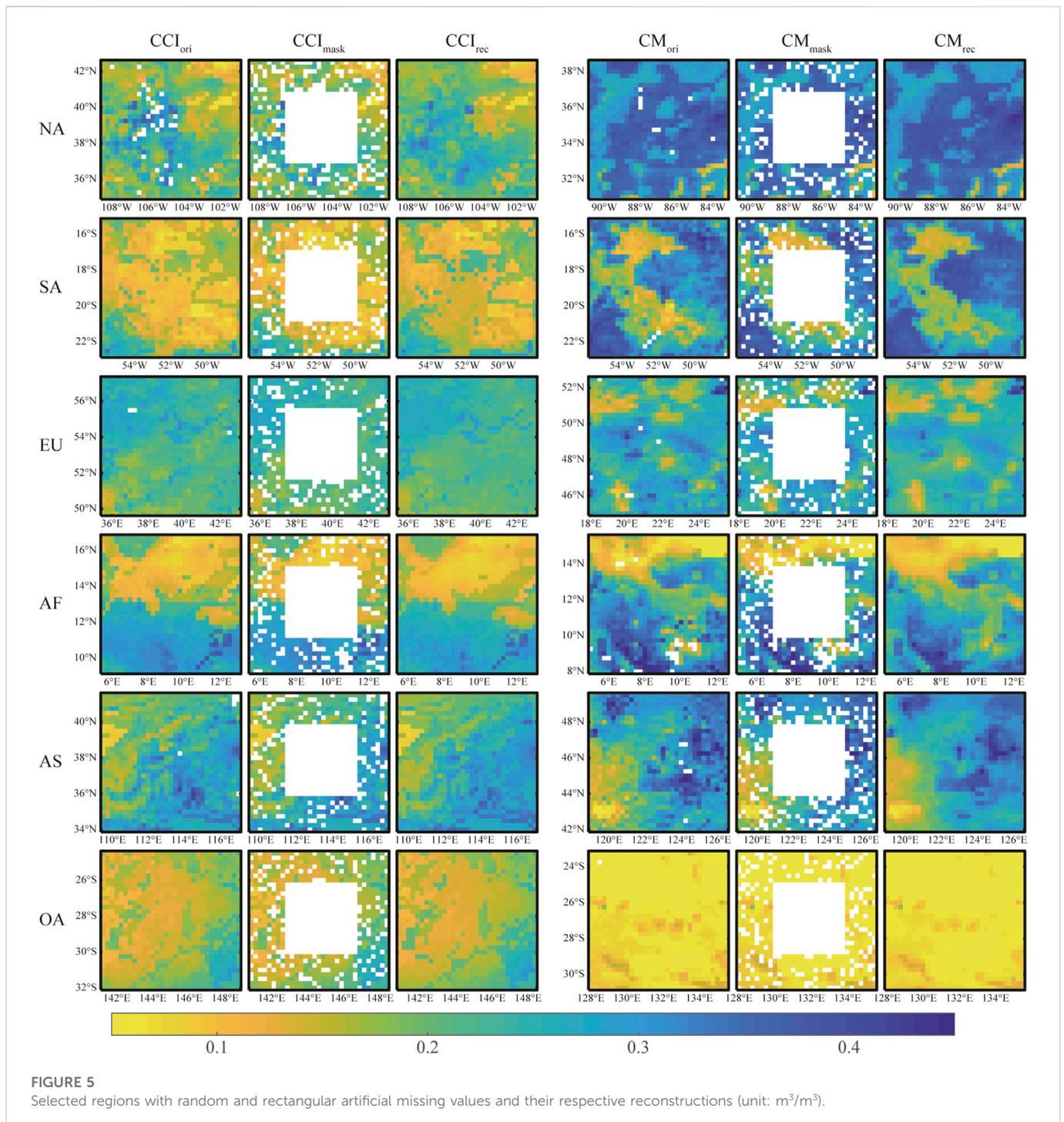
information for image reconstruction in the reconstruction process, while cannot consider both spatial and temporal information of data in the reconstruction process, which leads to a great uncertainty (Zhang et al., 2021c). It has been shown that DL methods generally outperform traditional data reconstruction methods with stronger non-linear modeling capabilities (Nogueira et al., 2018; Ma et al., 2019; Wu et al., 2019), and the CNN method is able to consider both temporal and spatial information of the data in the reconstruction process (Barth et al., 2020; Han et al., 2020; Luo et al., 2022). For the above reasons, this study does not compare the performance of different methods, but focuses on obtaining reconstructed data.

The used model here is a CNN structure with autoencoding approach. The convolutional auto-encoder replaces the full connection layer in the traditional auto-encoder with a convolutional layer and a pooling layer, thus effectively reducing data transfer loss (Yuan et al., 2019). The model contains five encoding and decoding layers, and each layer has a convolutional kernel size of 3 × 3. The pooling layer is used to reduce the size of the model, enhance computational speed, and optimize model robustness during features extraction. After each encoding layer, there is an average pooling layer with a filter size of 2 × 2; they together constitute the encoder which can convert the input feature into a latent-space representation. The
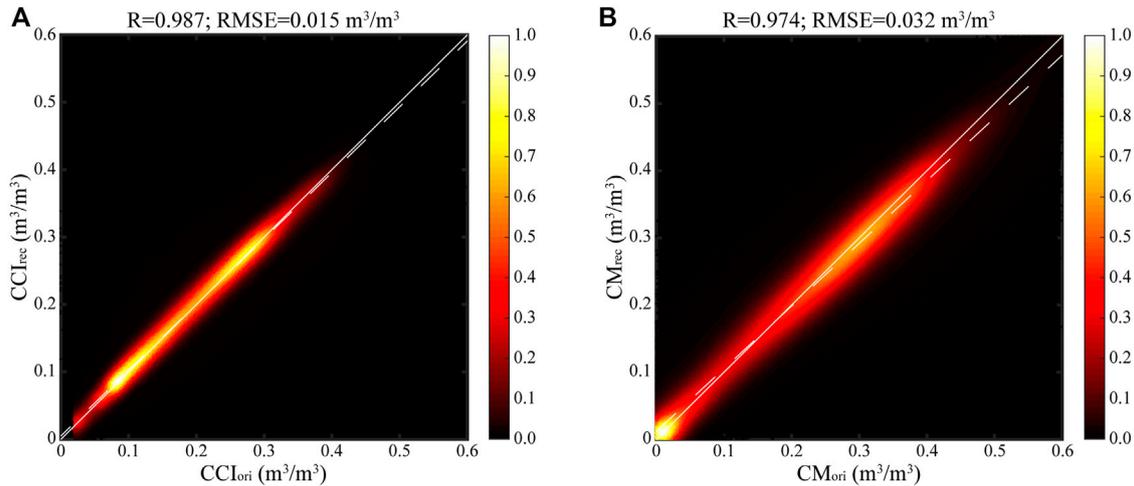
FIGURE 6
The scatter plots of artificial missing values and their reconstructions for CCI **(A)** and CM **(B)** data respectively. The colorbar indicates the data frequency.
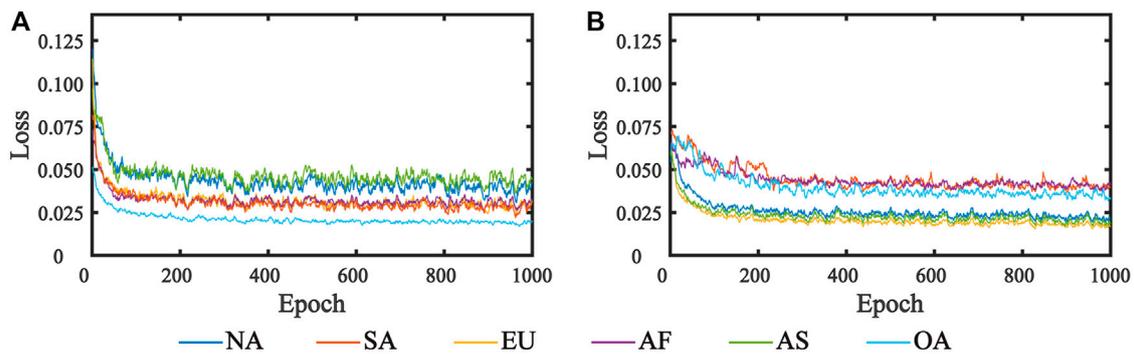

FIGURE 7
The loss function of RMSE for CCI **(A)** and CM **(B)** data. The solid lines with different colors represent the loss functions of different regions, which are consistent with each partition in Figure 2. Each epoch includes 366 iterations.

decoder, consisting of convolutional layers and interpolation layers, reconstructs the input data according to the potential spatial representation obtained by the encoder. Meanwhile, skip connection structures are added to the model to increase feature utilization. The DL model obtains soil moisture values, longitude, latitude and time information from CM_ori and CCI_ori data respectively, from which the input feature is generated. The input feature includes not only soil moisture at date T, but also the values at date T − 1 and T + 1. Moreover, The seasonality calculated at date T is also added to the feature, so that the DL model can consider the seasonal variability. Longitudes and latitudes are added to the input feature, so that the DL model can consider the influence of spatial positions on soil moisture. During the network training process, the convolution operation enables the model to consider the influence of the surrounding pixels on the center pixel soil moisture. In summary, the used DL model can consider information from both spatial and temporal dimensions, which is the main advantage of the model. The model diagram is shown in Figure 3. We chose RMSE as the loss function of the model to assess the error between the predicted and true values, which also provides support for

the parameter adjustment of the model (LeCun et al., 2015). The formula is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \quad (1)$$

where $\hat{y}_i$ represents the predicted value, y_i represents the true value, and n represents the number of data involved in the calculation of the loss function. Adam optimizer is used to optimize the parameters. The learning rate is set to 0.001, the exponential decay rate is set to 0.9 for the first moment and 0.999 for the second moment. The regularization parameter is set to 10^{-8}.

2.4 Data post-processing

The DL model is trained for the six partitioned regions respectively, from which the reconstructed data are then spatially merged with edge artifacts removed. Soil moisture is considered ineffective in densely vegetated areas; therefore we have eliminated

FIGURE 8
Spatial patterns of CCI$_{ori}$ **(A)**, CCI$_{rec}$ **(B)**, CM$_{ori}$ **(C)**, and CM$_{rec}$ **(D)** on 1 August 2012 (unit: m³/m³).

FIGURE 9
Time series from CCI$_{ori}$ and CM$_{ori}$ and their reconstructions. Shown in **(A, C)** are from the same pixel, so are those in **(B, D)**.
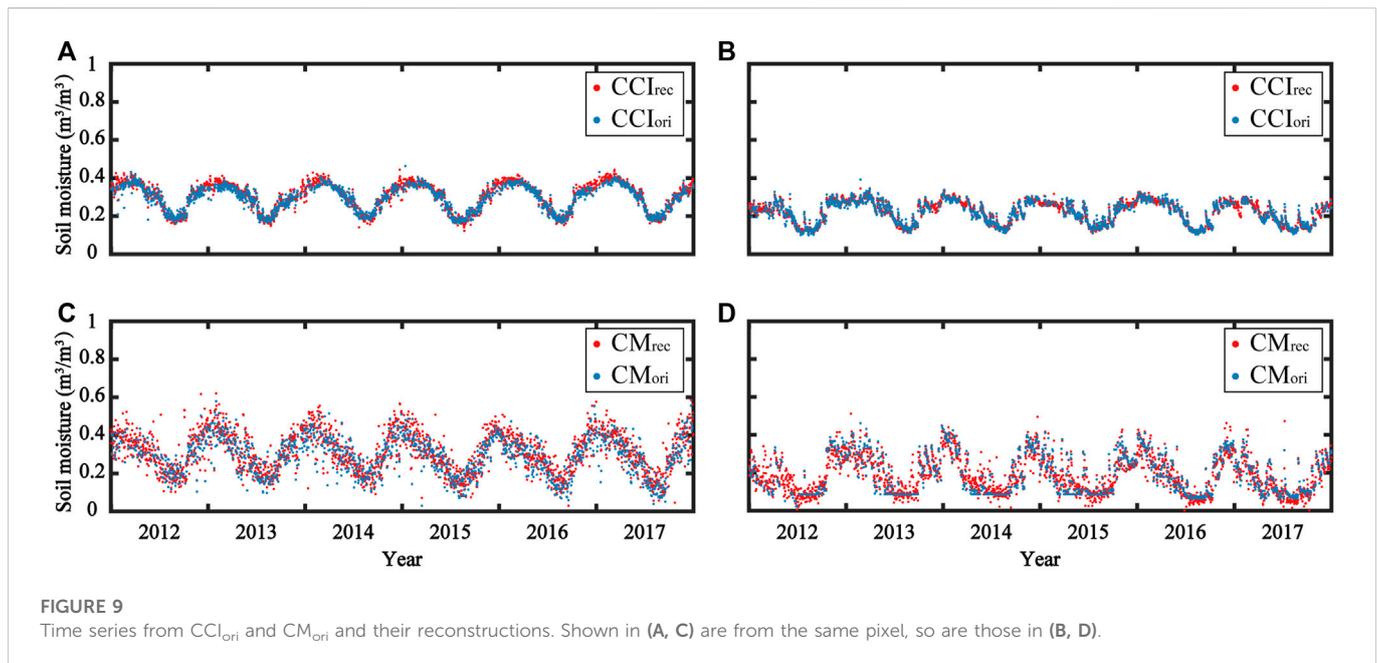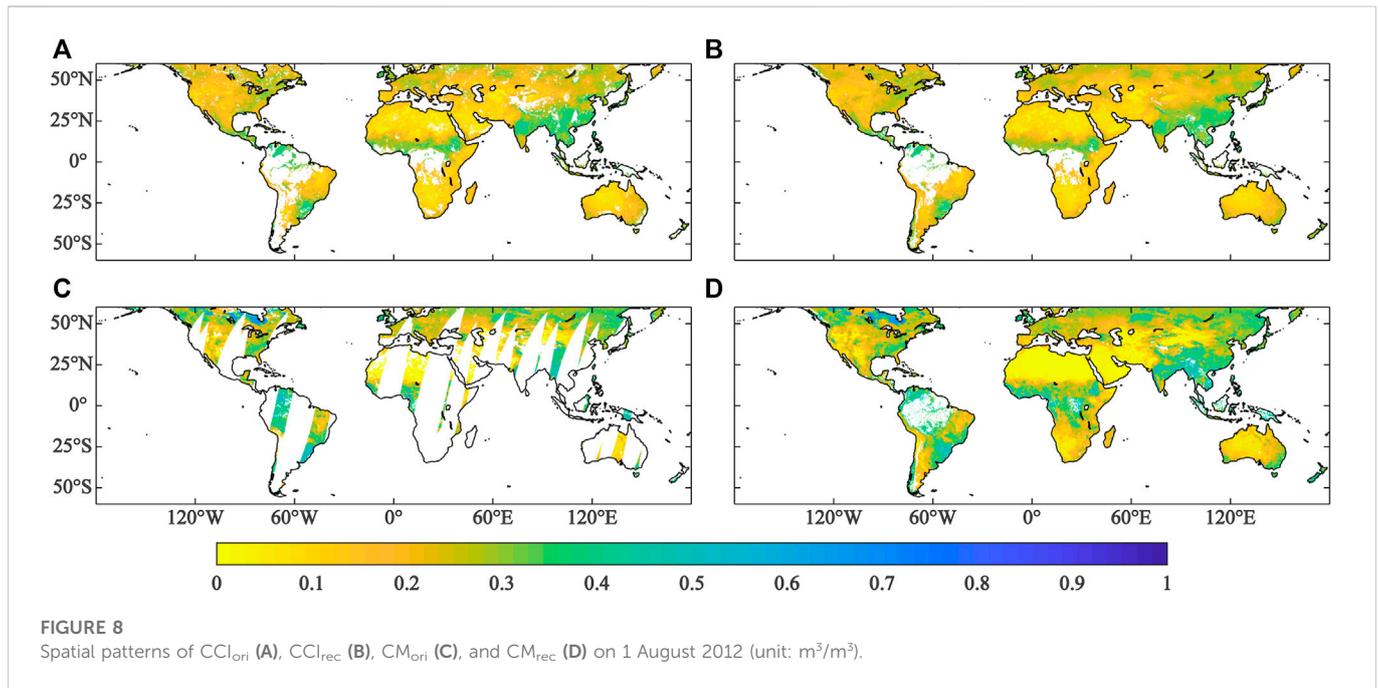
the reconstructed values where the multiple-year averaged NDVI (Normalized Difference Vegetation Index) is higher than 0.8 using the Advanced Very High-Resolution Radiometer (AVHRR) product (doi.org/10.7289/V5PZ56R6). The reconstructed values are also eliminated where soil temperature is below 0°C in cold seasons using ERA5-Land data a reference (doi.org/10.24381/cds.e2161bac).

2.5 Data quality evaluation

We use cross-validation and *in-situ* observations to evaluate the quality of reconstructed data. Several error metrics including the

correlation coefficient (R), bias, root mean square error (RMSE) and unbiased root mean square error (ubRMSE) are used. The formula is as follows:

$$R = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \overline{\hat{y}})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^{n} (\hat{y}_i - \overline{\hat{y}})^2}} \quad (2)$$

$$bias = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i) \quad (3)$$

$$ubRMSE = \sqrt{RMSE^2 - bias^2} \quad (4)$$

where $\hat{y}_i$ represents the original or reconstructed soil moisture, y_i represents the reference soil moisture, $\bar{y}$ represents the mean value of

FIGURE 10
CCI_{rec} against CCI_{ori} **(A)**, CM_{rec} against CM_{ori} **(B)**, CCI_{ori} against *in-situ* observations **(C)**, CM_{ori} against *in-situ* observations **(D)**, CCI_{rec} against *in-situ* observations **(E)**, and CM_{rec} against *in-situ* observations **(F)** from January 2012 to December 2017.
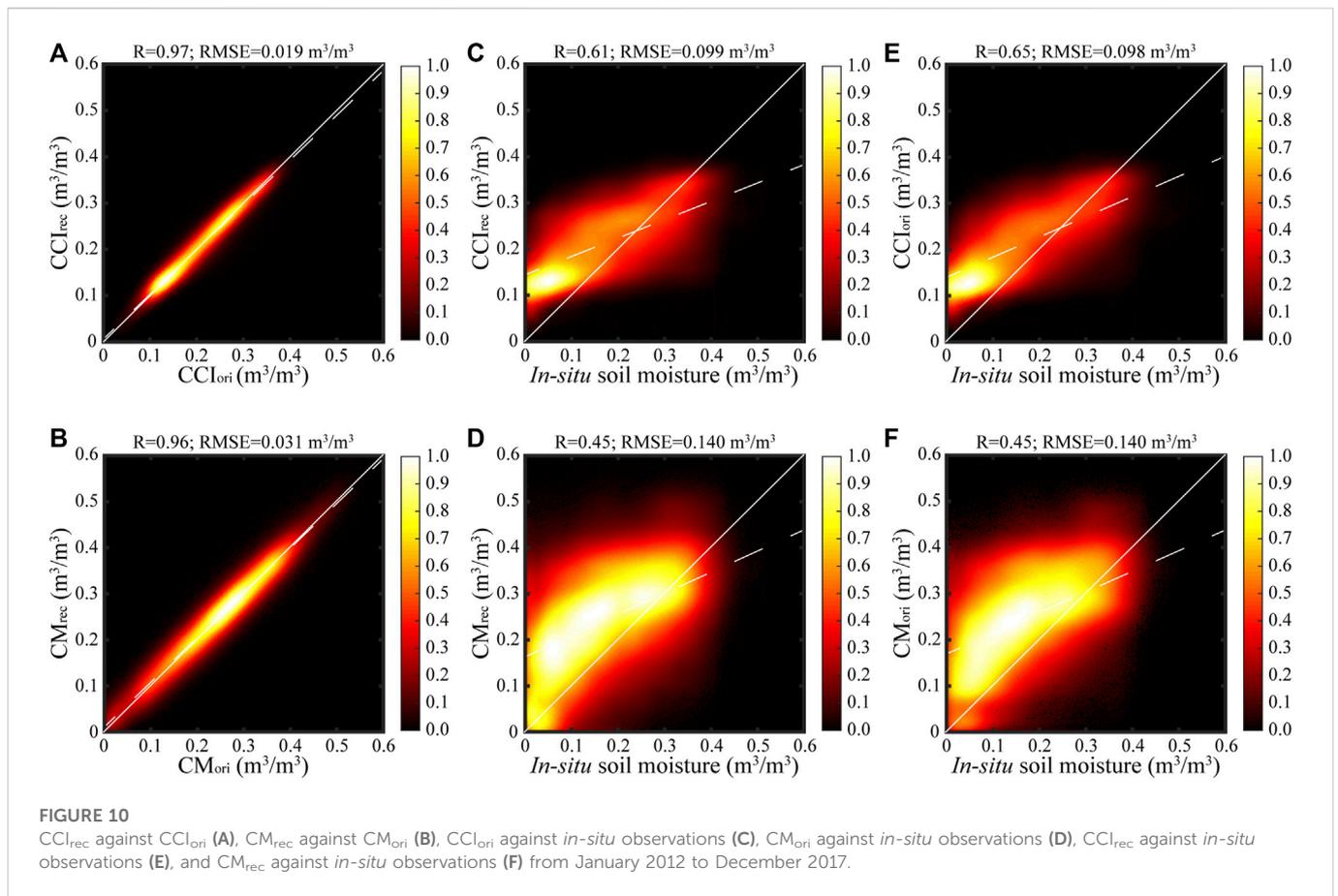
original or reconstructed soil moisture, $\bar{y}$ represents the mean value of reference soil moisture, and the n represents the number of data involved in the calculation.

For the cross-validation, we add 20% artificial missing values in the CCI_{ori} and CM_{ori} data respectively, and also continous missing values in rectangular regions. Then the convolutional auto-encoder is used for reconstruction. Finally, the reconstructed artificial missing values and their corresponding original values are used to evaluate the reconstruction performance. For the *in-situ* validation, we use the measured daily data obtained from multiple soil moisture stations in the International Soil Moisture Network (ISMN) as the *in-situ* data for the evaluation (Dorigo et al., 2013). Totally eleven networks are selected from ISMN, which are distributed in North America and Europe, Africa, Asia, and Oceania. An additional soil moisture network from Jiangsu province, China is also used. Finally we have twelve soil moisture networks which consist *in-situ* soil moisture observations, spanning 2012–2017, from 485 stations for data validation. The soil moisture networks are shown as point in Figure 2 with different colors.
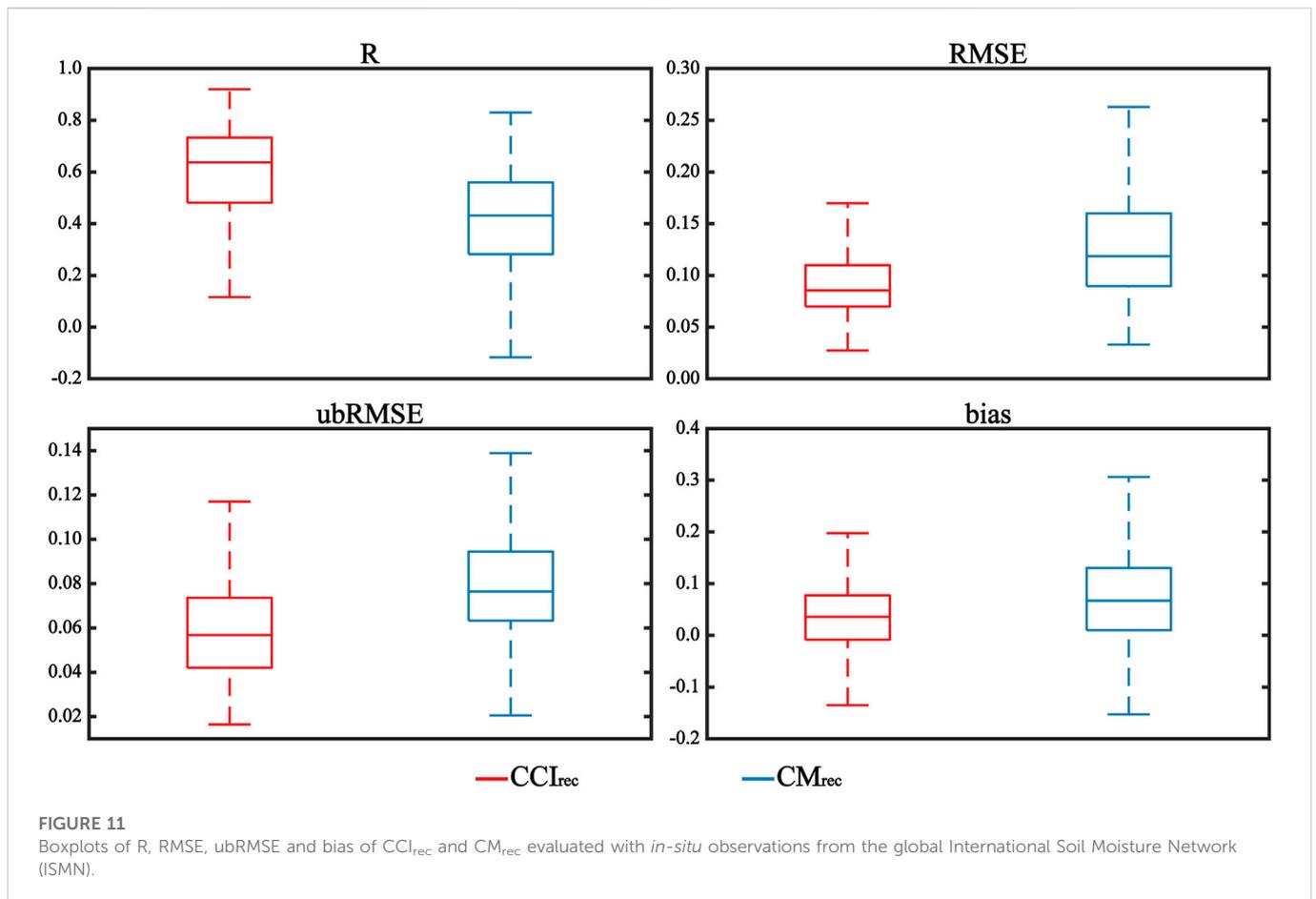
3 Results

In this study, we use a convolutional neural network with autoencoding approach to reconstruct soil moisture datasets. The reconstructed data effectively fills the gaps in the original data and retains the data distribution and characteristics of the original

data. We first conduct cross-validation experiments by simulating missing values to demonstrate the effectiveness of the deep learning method in reconstructing soil moisture data. Then, we use loss function, data spatial pattern, time series, and data distribution to evaluate the data reconstruction results. At the same time, we also choose the *in-situ* data to evaluate the accuracy of the reconstruction data. The specific results for each section are shown below.

3.1 Cross-validation results

We design an experiment of artifical missing values to prove the feasibility and effectiveness of deep learning model in reconstructing soil moisture, using data of the last year of CCI_{ori} (2021) and CM_{ori} (2019). In the experiment, we add 20% missing values in random and spatially continuous manners to the CCI_{ori} and CM_{ori} data respectively. The processed data is put into the trained deep learning model to obtain the reconstructions, denoted as CCI_{rec} and CM_{rec}, respectively. Figure 4 shows the CCI_{ori} and CM_{ori} with artificial missing values on 31 August 2021 and 31 August 2019 as examples, and their reconstructions. It can be seen from Figure 4 that the reconstructed artificial missing values shows quite similar spatial patterns as the original data; particularly, the artificial missing values in the rectangular regions are very well reconstructed. This indicates the used DL model is capable of reconstructing not only random missing values but also those of large regions.

FIGURE 11
Boxplots of R, RMSE, ubRMSE and bias of CCI$_{rec}$ and CM$_{rec}$ evaluated with *in-situ* observations from the global International Soil Moisture Network (ISMN).

To better understand the reconstruction performance, we further select several small regions with artificial random and rectangular missing values to show the results as in Figure 5. Clearly, the reconstructed missing values in CCI$_{rec}$ and CM$_{rec}$ are highly consistent with those in the CCI$_{ori}$ and CM$_{ori}$ data. Particularly, it is found that the artificial missing values in all the rectangular regions are very well constructed with very fine spatial patterns. However, this is not surprising because the used DL model considers soil moisture variability in both spatial and temporal dimensions.

In addition, the scatter plots of the reconstructed artificial missing values against the original ones are shown in Figure 6. It appears clearly that the fited line (white dotted line in Figure 6) is close to 1:1 (white solid line in Figure 6), indicating the consistent density distribution between the reconstructions and original values. Meanwhile, they have high correlation coefficients and low RMSE. R for CCI$_{ori}$ and CCI$_{rec}$ is 0.987 and RMSE for them is only 0.015 m^3/m^3. For CM$_{ori}$ and CM$_{rec}$, R is 0.974 and RMSE is 0.032 m^3/m^3. Obviously, R is higher for the CCI data than the CM data, and RMSE is vise versa. This is reasonable because there are apparently more missing values in the CM data which can largely affect the training of DL model. However, their differences of R and RMSE is relatively limited; this implies that the used DL model can still fulfill the reconstruction task even if there is a large portion of missing values in the original data.

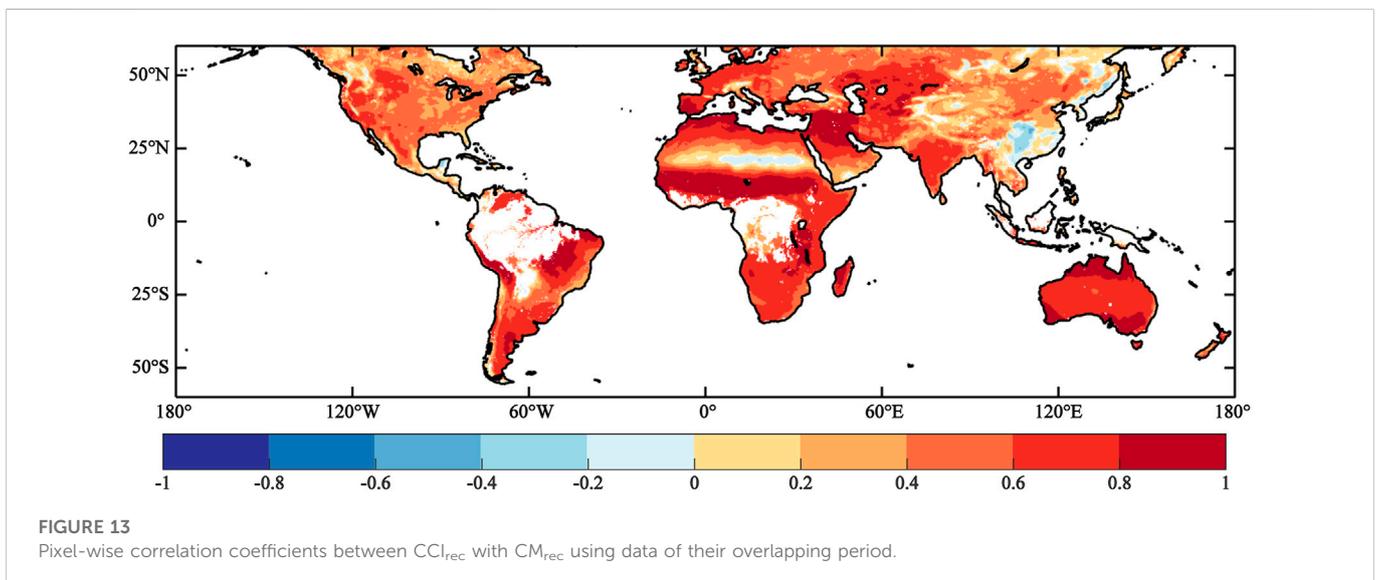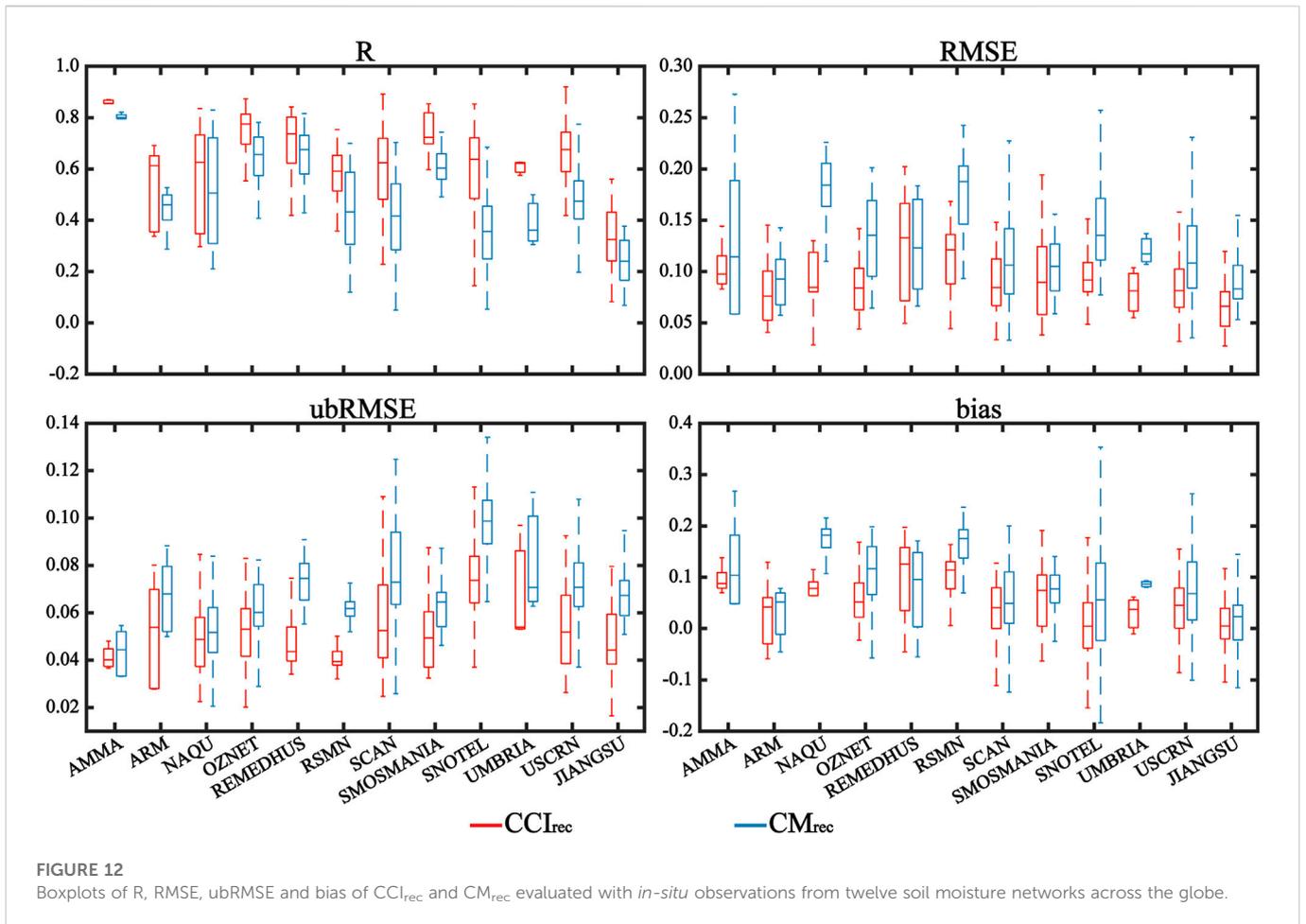The above results of cross-validation indicates that the used DL method is highly capable of reconstructing the missing values in the original soil moisture data, which proves that the soil moisture data reconstructed by this method has good credibility.

3.2 Loss function

To optimize parameters of the DL model, the loss function of each iteration is computed using RMSE, which is shown in Figure 7. Since the global land is partitioned into six regions (Figure 2) for the reconstruction, their loss curves are shown in Figure 7 respectively. The loss values of CCI and CM data in different regions are initially all above 0.05, which gradually decrease with increasing iterations and finally stabilize after 200 epochs. Although all the loss functions are stabilized below 0.05, they are different for the partitioned regions owing to different data properties.

3.3 Spatial patterns

The spatial patterns of CCI$_{ori}$ and CM$_{ori}$ as well as their reconstructions on 1 August 2012 is used as an example to better illustrate the results. Clearly, although there are large regions of missing values in CCI$_{ori}$ (Figure 8A) and CM$_{ori}$ (Figure 8C), their spatial patterns are very well reconstructed as shown in Figures 8B, D respectively. It is worth noting that, values are masked out in CCI$_{rec}$ and CM$_{rec}$ where either soil temperature is below 0°C or Normalized Difference Vegetation Index (NDVI) is larger than 0.8. Clearly, both

FIGURE 12
Boxplots of R, RMSE, ubRMSE and bias of CCI$_{rec}$ and CM$_{rec}$ evaluated with *in-situ* observations from twelve soil moisture networks across the globe.

FIGURE 13
Pixel-wise correlation coefficients between CCI$_{rec}$ with CM$_{rec}$ using data of their overlapping period.

CCI$_{rec}$ and CM$_{rec}$ are rather consistent in their spatial patterns, indicating such a reconstruction with DL method can effectively resolve the data gaps in large space while preserve the spatial features of the original data.

3.4 Time series

To further verify the soil moisture reconstruction in temporal dimension, two pixels are selected to superimpose the original time

series onto the reconstructed ones as shown in Figure 9. The time series spans from 1 January 2012 to 31 December 2017. While time series in Figures 9A, C are derived from the same pixel, those in Figures 9B, D are derived from a second one. It appears that, the temporal changes of original data are well reconstructed considering the variations at seasonal and finer time scales, although they are slightly different in CCI and CM data. It can be seen that, CCI data has smaller day-to-day variations than those in CM data. It is thus indicated the used DL method can reliably reconstruct the temporal variations of both CCI and CM data in presence of considerable portion of missing values.

3.5 Global performance

The above analysis indicates the reconstructed data retains the features of the original data in both spatial and temporal dimensions (Figures 8, 9). To understand the general performance of data reconstruction, we show the scatter plots of the global reconstructions against the original data and *in-situ* observations in Figure 10. The fitted data is quite close to 1:1 line (white solid line in Figure 10), and the points are almost evenly distributed on both sides of the fitted line. It appears R between CCI_{rec} and CCI_{ori} is 0.97, and RMSE is 0.019 m^3/m^3 (Figure 10A); those metrics for CM_{rec} and CM_{ori} is 0.96 and 0.031 m^3/m^3 (Figure 10B) respectively. It is indicated the CCI reconstruction performs better than CM data. The validations of CCI_{rec} and CM_{rec} with *in-situ* observations (Figures 10C, D) are similar to *in-situ* observation verification results for CCI_{ori} and CM_{ori} (Figures 10E, F). The data in Figures 10C, D are all distributed above the 1:1 line (white solid line in Figure 10), which may indicate that the CCI_{rec} and CM_{rec} data overestimate soil moisture to some extent. The data distribution of CCI_{rec} is more concentrated than CM_{rec}. R between CM_{rec} and *in-situ* observations is 0.45 and 0.140 m^3/m^3 (Figure 10D), and those metrics for CCI_{rec} and *in-situ* observations is 0.61 and 0.099 m^3/m^3 (Figure 10C) respectively, indicating better reconstruction results of CCI than CM.

3.6 Comparison with *in-situ* observations

Four metrics of R, RMSE, ubRMSE and bias are used to further evaluate CCI_{rec} and CM_{rec} using global *in-situ* observations (Figure 11). Generally speaking, R between CCI_{rec} and *in-situ* observations appears to be 0.65 on average, and that for CM_{rec} is 0.42 on average. Apparently, CCI_{rec} shows much higher correlations with *in-situ* observations with smaller range among the stations. Conversely, RMSE, ubRMSE and bias of CCI_{rec} is much smaller than those of CM_{rec}, and their ranges are much smaller indicating less uncertainties of these results. It is thus convincing that the derived CCI_{rec} is more consistent with the *in-situ* observations globally.

In Figure 12, we further show the boxplots of R, RMSE, ubRMSE and bias of CCI_{rec} and CM_{rec} against *in-situ* observations from eleven regional soil moisture from ISMN and the JIANGSU network in addition. Metrics of satellite data show considerable variations in different regions due to the factors such as satellite sensors, retrieval algorithms and merging algorithms (Fu et al., 2019). Apparently, R coefficients between CCI_{rec} and *in-situ* observations in all the twelve regions appears to be much higher than those for CM_{rec}; on the

contrary, metrics of RMSE, ubRMSE and bias for CCI_{rec} are largely smaller than those for CM_{rec}. It is thus indicated that CCI_{rec} notably outperforms CM_{rec} in the studied soil moisture networks across the globe.

3.7 Spatio-temporal consistency

It is of interest to understand the spatio-temporal consistency of the reconstructed CCI_{rec} and CM_{rec} data, for which we computed their pixel-wise correlation coefficients as shown in Figure 13. Generally, CCI_{rec} and CM_{rec} are highly and positively correlated across most of the global land, and their correlation coefficients can reach >0.8 in transitional regions between wet and dry climate. However, in some regions such as southwest China, they show slightly negative correlation coefficients, indicating large discrepancy in both data and we have to use them with caution. In densely vegetated regions such as rainforest, the CCI_{rec} and CM_{rec} values are masked out since microwave remote sensing can not penetrate dense vegetation.

4 Conclusion and discussion

The added value of this study is generating of the longest global gap-free soil moisture records from satellites with daily resolution, which is urgently needed in Earth sciences. In this study, the DL method with auto-encoder as the main structure can fully consider the spatio-temporal information of the data itself to effectively reconstruct the satellite-based soil moisture data with missing values. The reconstructed datasets are cross-evaluated with artificial missing values, and further againt *in-situ* observations with multiple error metrics. The main results of this study are as follows: 1) Cross-validation has shown that the DL method used in this study has high reliability in reconstructing the missing values in satellite soil moisture products. 2) The reconstructed data can well retain the pattern of the original data in both space and time dimensions. 3) *In-situ* validations have shown that CCI_{rec} is much better than CM_{rec}. In general, the long-term global gap-free soil moisture data reconstructed in this study can provide credible support for the development of related research.

The contribution of this study is to effectively solve the problem of missing values in long-term satellite-based soil moisture products. At present, the reconstructed CCI_{rec} is the longest gap-free soil moisture data under the premise of daily resolution, which is expected to contribute to water cycle studies in the warming climate. Most of the reconstruction work use single satellite-based soil moisture data for reconstruction (Fang et al., 2017; Zhang et al., 2021c), limited by the service life of the satellite, these data time coverage is short, difficult to obtain long-term reconstruction data. Compared with reconstruction work based on similar satellite merging soil moisture data (Zhang et al., 2021b; Guevara et al., 2021), our reconstructed data have larger spatial scale and higher temporal resolution. At the same time, compared with the global reconstruction work, we partition the global land into six regions for reconstruction, because the local model can better learn regional differences and help improve training efficiency compared with the global-scale model using global data (Chen et al., 2021). In this study, we find that the reconstructed data has a consistency with

original data; but in comparison with the *in-situ* data, the accuracy evaluation index of the reconstructed data are slightly lower than or the same as those of the original data. Because only the soil moisture information is used for reconstruction, so that the accuracy of the reconstructed data is difficult to exceed the accuracy of the original data. This result is consistent with the research results of Zhang et al. (2021c). Barth et al. (2022) pointed out in the latest research that using more variables that are strongly associated with the reconstructed data as input to the data reconstruction process can improve the reconstruction effect to some extent. In future studies, researchers can add more variables closely related to soil moisture as CNN method inputs, such as precipitation, temperature and evapotranspiration, in the process of reconstructing soil moisture to try to improve the reconstruction results of soil moisture.

The main disadvantage of the method used in this study is that although temporal information is considered to some extent in the reconstruction process, it still has some limitations, i.e., when reconstructing soil moisture data at date T, only the effects of date $T + 1$ and $T - 1$ can be considered, while it is difficult to consider the effects of more distant dates on the reconstruction of soil moisture data at date T. At the same time, the present method does not assign different weights according to the distance from date T. Therefore, it results in the same intensity of impact on the date T data reconstruction for dates that are farther away from date T or for dates that are closer. In future research, researchers can try to introduce recurrent neural network modules into the CNN method as a way to achieve the purpose of considering the influence of more distant dates on the reconstructed date T and assigning different weights according to the distance from the reconstructed date T during the reconstruction process.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://dx.doi.org/10.11888/Terre.tpdc.272994.

Author contributions

GW design the study. YH write the paper. XW, FZ, GK, SA, DH, and ZD review the paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alvera-Azcárate, A., Barth, A., Parard, G., and Beckers, J. M. (2016). Analysis of SMOS sea surface salinity data using DINEOF. *Remote Sens. Environ.* 180, 137–145. doi:10.1016/j.rse.2016.02.044

Barth, A., Alvera-Azcárate, A., Licer, M., and Beckers, J. M. (2020). Dincae 1.0: A convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations. *Geosci. Model. Dev.* 13 (3), 1609–1622. doi:10.5194/gmd-13-1609-2020

Barth, A., Alvera-Azcárate, A., Troupin, C., and Beckers, J. M. (2022). Dincae 2.0: Multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations. *Geosci. Model. Dev.* 15 (5), 2183–2196. doi:10.5194/gmd-15-2183-2022

Champagne, C., White, J., Berg, A., Belair, S., and Carrera, M. (2019). Impact of soil moisture data characteristics on the sensitivity to crop yields under drought and excess moisture conditions. *Remote Sens.* 11 (4), 372. doi:10.3390/rs11040372

Chen, Y., Feng, X., and Fu, B. (2021). An improved global remote-sensing-based surface soil moisture (RSSSM) dataset covering 2003–2018. *Earth Syst. Sci. Data* 13 (1), 1–31. doi:10.5194/essd-13-1-2021

Cook, B. I., Bonan, G. B., and Levis, S. (2006). Soil moisture feedbacks to precipitation in southern Africa. *J. Clim.* 19 (17), 4198–4206. doi:10.1175/jcli3856.1

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., et al. (2012). Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Rev. Geophys.* 50 (2). doi:10.1029/2011rg000372

Cui, Y., Zeng, C., Zhou, J., Xie, H., Wan, W., Hu, L., et al. (2019). A spatio-temporal continuous soil moisture dataset over the Tibet Plateau from 2002 to 2015. *Sci. Data* 6 (1), 247–7. doi:10.1038/s41597-019-0228-x

Cui, Y., Ma, S., Yao, Z., Chen, X., Luo, Z., Fan, W., et al. (2020). Developing a gap-filling algorithm using DNN for the Ts-VI triangle model to obtain temporally continuous daily actual evapotranspiration in an arid area of China. *Remote Sens.* 12 (7), 1121. doi:10.3390/rs12071121

Dorigo, W. A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiova, A., Sanchis-Dufau, A. D., et al. (2013). Global automated quality control of *in situ* soil moisture data from the International Soil Moisture Network. *Vadose Zone J.* 12 (3), vzj2012.0097. doi:10.2136/vzj2012.0097

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., et al. (2017). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sens. Environ.* 203, 185–215. doi:10.1016/j.rse.2017.07.001

Draper, D. W. (2018). Radio frequency environment for Earth-observing passive microwave imagers. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 11 (6), 1913–1922. doi:10.1109/jstars.2018.2801019

Eeckman, J., Roux, H., Douinot, A., Bonan, B., and Albergel, C. (2020). A multi-sourced assessment of the spatio-temporal dynamic of soil saturation in the MARINE flash flood model. *Hydrology Earth Syst. Sci. Discuss.* 2020, 1–31. doi:10.5194/hess-2020-311

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., et al. (2010). The soil moisture active passive (SMAP) mission. *Proc. IEEE* 98 (5), 704–716. doi:10.1109/JPROC.2010.2043918

Fang, K., Shen, C., Kifer, D., and Yang, X. (2017). Prolongation of SMAP to spatiotemporally seamless coverage of continental US using a deep learning neural network. *Geophys. Res. Lett.* 44 (21), 11–030. doi:10.1002/2017gl075619

Fu, H., Zhou, T., and Sun, C. (2019). Evaluation and analysis of amsr2 and fy3b soil moisture products by an *in situ* network in cropland on pixel scale in the northeast of China. *Remote Sens.* 11 (7), 868. doi:10.3390/rs11070868

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., and Dorigo, W. (2019). Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. *Earth Syst. Sci. Data* 11 (2), 717–739. doi:10.5194/essd-11-717-2019

Guan, X., Huang, J., Guo, N., Bi, J., and Wang, G. (2009). Variability of soil moisture and its relationship with surface albedo and soil thermal parameters over the Loess Plateau. *Adv. Atmos. Sci.* 26 (4), 692–700. doi:10.1007/s00376-009-8198-0

Guevara, M., Taufer, M., and Vargas, R. (2021). Gap-free global annual soil moisture: 15 km grids for 1991–2018. *Earth Syst. Sci. Data* 13 (4), 1711–1735. doi:10.5194/essd-13-1711-2021

Hagan, D. F. T., Wang, G., Kim, S., Parinussa, R. M., Liu, Y., Ullah, W., et al. (2020). Maximizing temporal correlations in long-term global satellite soil moisture data-merging. *Remote Sens.* 12 (13), 2164. doi:10.3390/rs12132164

Han, Z., He, Y., Liu, G., and Perrie, W. (2020). Application of DINCAE to reconstruct the gaps in chlorophyll-a satellite observations in the south China sea and west philippine sea. *Remote Sens.* 12 (3), 480. doi:10.3390/rs12030480

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., et al. (2014). The global precipitation measurement mission. *Bull. Am. Meteorological Soc.* 95 (5), 701–722. doi:10.1175/bams-d-13-00164.1

Hu, J., and Lv, Y. H. (2015). Research progress on stochastic soil moisture dynamic model. *Prog. Geogr.* 34, 389–400. doi:10.11820/dlkxjz.2015.03.014

Huang, C., Li, X., Lu, L., and Gu, J. (2008). Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter. *Remote Sens. Environ.* 112 (3), 888–900. doi:10.1016/j.rse.2007.06.026

Jing, W., Zhang, P., and Zhao, X. (2018). Reconstructing monthly ECV global soil moisture with an improved spatial resolution. *Water Resour. Manag.* 32 (7), 2523–2537. doi:10.1007/s11269-018-1944-2

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., et al. (2010). Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature* 467 (7318), 951–954. doi:10.1038/nature09396

Karthikeyan, L., Pan, M., Wanders, N., Kumar, D. N., and Wood, E. F. (2017). Four decades of microwave satellite soil moisture observations: Part 2. Product validation and inter-satellite comparisons. *Adv. Water Resour.* 109, 236–252. doi:10.1016/j.advwatres.2017.09.010

Kerr, Y. H., Waldteufel, P., Wigneron, J. P., Martinuzzi, J. A. M. J., Font, J., and Berger, M. (2001). Soil moisture retrieval from space: The soil moisture and Ocean Salinity (SMOS) mission. *IEEE Trans. Geoscience Remote Sens.* 39 (8), 1729–1735. doi:10.1109/36.942551

Kim, S., Liu, Y. Y., Johnson, F. M., Parinussa, R. M., and Sharma, A. (2015a). A global comparison of alternate AMSR2 soil moisture products: Why do they differ? *Remote Sens. Environ.* 161, 43–62. doi:10.1016/j.rse.2015.02.002

Kim, S., Parinussa, R. M., Liu, Y. Y., Johnson, F. M., and Sharma, A. (2015b). A framework for combining multiple soil moisture retrievals based on maximizing temporal correlation. *Geophys. Res. Lett.* 42 (16), 6662–6670. doi:10.1002/2015gl064981

Korres, W., Reichenau, T. G., Fiener, P., Koyama, C. N., Bogena, H. R., Cornelissen, T., et al. (2015). Spatio-temporal soil moisture patterns–A meta-analysis using plot to catchment scale data. *J. hydrology* 520, 326–341. doi:10.1016/j.jhydrol.2014.11.042

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521 (7553), 436–444. doi:10.1038/nature14539

Liu, Q., Zhang, J., Zhang, H., Yao, F., Bai, Y., Zhang, S., et al. (2021). Evaluating the performance of eight drought indices for capturing soil moisture dynamics in various vegetation regions over China. *Sci. Total Environ.* 789, 147803. doi:10.1016/j.scitotenv.2021.147803

Liu, Y., Yang, Y., and Yue, X. (2018). Evaluation of satellite-based soil moisture products over four different continental *in-situ* measurements. *Remote Sens.* 10 (7), 1161. doi:10.3390/rs10071161

Liu, Y. Y., van Dijk, A. I., de Jeu, R. A., and Holmes, T. R. (2009). An analysis of spatiotemporal variations of soil and vegetation moisture from a 29-year satellite-derived data set over mainland Australia. *Water Resour. Res.* 45 (7). doi:10.1029/2008wr007187

Liu, Y. Y., Dorigo, W. A., Parinussa, R. M., de Jeu, R. A., Wagner, W., McCabe, M. F., et al. (2012). Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sens. Environ.* 123, 280–297. doi:10.1016/j.rse.2012.03.014

Liu, Z., Wu, P., Duan, S., Zhan, W., Ma, X., and Wu, Y. (2017). Spatiotemporal reconstruction of land surface temperature derived from fengyun geostationary satellite data. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 10 (10), 4531–4543. doi:10.1109/jstars.2017.2716376

Llamas, R. M., Guevara, M., Rorabaugh, D., Taufer, M., and Vargas, R. (2020). Spatial gap-filling of ESA CCI satellite-derived soil moisture based on geostatistical techniques and multiple regression. *Remote Sens.* 12 (4), 665. doi:10.3390/rs12040665

Long, D., Shen, Y., Sun, A., Hong, Y., Longuevergne, L., Yang, Y., et al. (2014). Drought and flood monitoring for a large karst plateau in Southwest China using extended GRACE data. *Remote Sens. Environ.* 155, 145–160. doi:10.1016/j.rse.2014.08.006

Lu, Z., Peng, S., Slette, I., Cheng, G., Li, X., and Chen, A. (2021). Soil moisture seasonality alters vegetation response to drought in the Mongolian Plateau. *Environ. Res. Lett.* 16 (1), 014050. doi:10.1088/1748-9326/abd1a2

Luo, X., Song, J., Guo, J., Fu, Y., Wang, L., and Cai, Y. (2022). Reconstruction of chlorophyll-a satellite data in Bohai and Yellow sea based on DINCAE method. *Int. J. Remote Sens.* 43 (9), 3336–3358. doi:10.1080/01431161.2022.2090872

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., and Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogrammetry Remote Sens.* 152, 166–177. doi:10.1016/j.isprsjprs.2019.04.015

Ma, Z. G., Wei, H. L., and Fu, C. B. (1999). Progress in the research on the relationship between soil moisture and climate change. *Adv. Earth Sci.* 14, 300–303. doi:10.11867/j.issn.1001-8166.1999.03.0299

Mittelbach, H., Lehner, I., and Seneviratne, S. I. (2012). Comparison of four soil moisture sensor types under field conditions in Switzerland. *J. Hydrology* 430, 39–49. doi:10.1016/j.jhydrol.2012.01.041

Njoku, E. G., Jackson, T. J., Lakshmi, V., Chan, T. K., and Nghiem, S. V. (2003). Soil moisture retrieval from AMSR-E. *IEEE Trans. Geoscience Remote Sens.* 41 (2), 215–229. doi:10.1109/tgrs.2002.808243

Nogueira, K., Fadel, S. G., Dourado, Í. C., Werneck, R. D. O., Muñoz, J. A., Penatti, O. A., et al. (2018). Exploiting convnet diversity for flooding identification. *IEEE Geoscience Remote Sens. Lett.* 15 (9), 1446–1450. doi:10.1109/lgrs.2018.2845549

Owe, M., de Jeu, R., and Holmes, T. (2008). Multisensor historical climatology of satellite-derived global land surface moisture. *J. Geophys. Res. Earth Surf.* 113 (F1), F01002. doi:10.1029/2007jf000769

Parinussa, R. M., Holmes, T. R., and de Jeu, R. A. (2011). Soil moisture retrievals from the WindSat spaceborne polarimetric microwave radiometer. *IEEE Trans. Geoscience Remote Sens.* 50 (7), 2683–2694. doi:10.1109/tgrs.2011.2174643

Parinussa, R. M., Holmes, T. R., Wanders, N., Dorigo, W. A., and de Jeu, R. A. (2015). A preliminary study toward consistent soil moisture from AMSR2. *J. Hydrometeorol.* 16 (2), 932–947. doi:10.1175/jhm-d-13-0200.1

Preimesberger, W., Scanlon, T., Su, C. H., Gruber, A., and Dorigo, W. (2020). Homogenization of structural breaks in the global ESA CCI soil moisture multisatellite climate data record. *IEEE Trans. Geoscience Remote Sens.* 59 (4), 2845–2862. doi:10.1109/tgrs.2020.3012896

Ran, Y., Jorgenson, M. T., Li, X., Jin, H., Wu, T., Li, R., et al. (2021). Biophysical permafrost map indicates ecosystem processes dominate permafrost stability in the Northern Hemisphere. *Environ. Res. Lett.* 16 (9), 095010. doi:10.1088/1748-9326/ac20f3

Raoult, N., Delorme, B., Ottlé, C., Peylin, P., Bastrikov, V., Maugis, P., et al. (2018). Confronting soil moisture dynamics from the ORCHIDEE land surface model with the ESA-CCI product: Perspectives for data assimilation. *Remote Sens.* 10 (11), 1786. doi:10.3390/rs10111786

Rigden, A. J., Powell, R. S., Trevino, A., McColl, K. A., and Huybers, P. (2020). Microwave retrievals of soil moisture improve grassland wildfire predictions. *Geophys. Res. Lett.* 47 (23), e2020GL091410. doi:10.1029/2020gl091410

Rossato, L., Alvala, R. C. D. S., Marengo, J. A., Zeri, M., Cunha, A. P. D. A., Pires, L. B., et al. (2017). Impact of soil moisture on crop yields over Brazilian semiarid. *Front. Environ. Sci.* 5, 73. doi:10.3389/fenvs.2017.00073

Sang, Y., Ren, H. L., Shi, X., Xu, X., and Chen, H. (2021). Improvement of soil moisture simulation in eurasia by the Beijing climate center climate system model from CMIP5 to CMIP6. *Adv. Atmos. Sci.* 38 (2), 237–252. doi:10.1007/s00376-020-0167-7

Schwingshackl, C., Hirschi, M., and Seneviratne, S. I. (2017). Quantifying spatiotemporal variations of soil moisture control on surface energy balance and near-surface air temperature. *J. Clim.* 30 (18), 7105–7124. doi:10.1175/jcli-d-16-0727.1

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., et al. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Rev.* 99 (3-4), 125–161. doi:10.1016/j.earscirev.2010.02.004

Stamenkovic, J., Guerriero, L., Ferrazzoli, P., Notarnicola, C., Greifeneder, F., and Thiran, J. P. (2017). Soil moisture estimation by SAR in Alpine fields using Gaussian process regressor trained by model simulations. *IEEE Trans. Geoscience Remote Sens.* 55 (9), 4899–4912. doi:10.1109/tgrs.2017.2687421

Sun, Z., Long, D., Yang, W., Li, X., and Pan, Y. (2020). Reconstruction of GRACE data on changes in total water storage over the global land surface and 60 basins. *Water Resour. Res.* 56 (4), e2019WR026250. doi:10.1029/2019wr026250

Thomas Ambadan, J., Oja, M., Gedalof, Z. E., and Berg, A. A. (2020). Satellite-observed soil moisture as an indicator of wildfire risk. *Remote Sens.* 12 (10), 1543. doi:10.3390/rs12101543

Wagner, W., Dorigo, W., De Jeu, R., Fernandez, D., Benveniste, J., and Haas, E. (2012). Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* 7, 315–321.

Wang, G., Garcia, D., Liu, Y., De Jeu, R., and Dolman, A. J. (2012). A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil

moisture observations. *Environ. Model. Softw.* 30, 139–142. doi:10.1016/j.envsoft.2011. 10.015

Wang, J., Pang, Y., Zhu, X., and Sun, Z. (2019). A review of researches on inversion of eigenvariance of soil water. *Acta Pedofil. Sin.* 56, 23–35. doi:10.11766/trxb201803090579

Wasko, C., and Nathan, R. (2019). Influence of changes in rainfall and soil moisture on trends in flooding. *J. Hydrology* 575, 432–441. doi:10.1016/j.jhydrol. 2019.05.054

Wu, P., Yin, Z., Yang, H., Wu, Y., and Ma, X. (2019). Reconstructing geostationary satellite land surface temperature imagery based on a multiscale feature connected convolutional neural network. *Remote Sens.* 11 (3), 300. doi:10. 3390/rs11030300

Yang, Z., Lu, N., Shi, J., Zhang, P., Dong, C., and Yang, J. (2012). Overview of FY-3 payload and ground application system. *IEEE Trans. Geoscience Remote Sens.* 50 (12), 4846–4853. doi:10.1109/tgrs.2012.2197826

Yang, K., Wang, C., and Bao, H. (2016). Contribution of soil moisture variability to summer precipitation in the Northern Hemisphere. *J. Geophys. Res. Atmos.* 121 (20), 12,108–12,124. doi:10.1002/2016jd025644

Yang, H., Xiong, L., Liu, D., Cheng, L., and Chen, J. (2021). High spatial resolution simulation of profile soil moisture by assimilating multi-source remote-sensed information into a distributed hydrological model. *J. Hydrology* 597, 126311. doi:10. 1016/j.jhydrol.2021.126311

Yee, M. S., Walker, J. P., Rüdiger, C., Parinussa, R. M., Koike, T., and Kerr, Y. H. (2017). A comparison of SMOS and AMSR2 soil moisture using representative sites of the OzNet monitoring network. *Remote Sens. Environ.* 195, 297–312. doi:10.1016/j.rse.2017.04.019

Yuan, F. N., Zhang, L., Shi, J. T., Xia, X., and Li, G. (2019). Theories and applications of auto-encoder neural networks: A literature survey. *Chin. J. Comput.* 42 (01), 203–230. doi:10.11897/SP.J.1016.2019.00203

Zeng, J., Li, Z., Chen, Q., Bi, H., Qiu, J., and Zou, P. (2015). Evaluation of remotely sensed and reanalysis soil moisture products over the Tibetan Plateau using *in-situ* observations. *Remote Sens. Environ.* 163, 91–110. doi:10.1016/j.rse.2015.03.008

Zhang, G., Su, X., Ayantobo, O. O., and Feng, K. (2021a). Drought monitoring and evaluation using ESA CCI and GLDAS-Noah soil moisture datasets across China. *Theor. Appl. Climatol.* 144 (3), 1407–1418. doi:10.1007/s00704-021-03609-w

Zhang, L., Liu, Y., Ren, L., Teuling, A. J., Zhang, X., Jiang, S., et al. (2021b). Reconstruction of ESA CCI satellite-derived soil moisture using an artificial neural network technology. *Sci. Total Environ.* 782, 146602. doi:10.1016/j.scitotenv.2021.146602

Zhang, Q., Yuan, Q., Li, J., Wang, Y., Sun, F., and Zhang, L. (2021c). Generating seamless global daily AMSR2 soil moisture (SGD-SM) long-term products for the years 2013–2019. *Earth Syst. Sci. Data* 13 (3), 1385–1401. doi:10.5194/essd-13-1385-2021

Zhu, Q., Zhou, Z., Duncan, E. W., Lv, L., Liao, K., and Feng, H. (2017). Integrating real-time and manual monitored data to predict hillslope soil moisture dynamics with high spatio-temporal resolution using linear and non-linear models. *J. Hydrology* 545, 1–11. doi:10.1016/j.jhydrol.2016.12.014