MAT-MS

A mask-aware transformer for constructing gap-free MODIS normalized difference snow index products

Xu, Jiahui; Hua, Ruiyang; Wang, Shujie; Lhermitte, Stef; Gu, Qingyu; Yu, Bailang; Wu, Jianping; Huang, Yan

DOI
10.1016/j.isprsjprs.2025.07.004

Publication date
2025

Document Version
Final published version

Published in
ISPRS Journal of Photogrammetry and Remote Sensing

Citation (APA)
Xu, J., Hua, R., Wang, S., Lhermitte, S., Gu, Q., Yu, B., Wu, J., & Huang, Y. (2025). MAT-MS: A mask-aware transformer for constructing gap-free MODIS normalized difference snow index products. *ISPRS Journal of Photogrammetry and Remote Sensing*, *227*, 775-788. https://doi.org/10.1016/j.isprsjprs.2025.07.004

ISPRS Journal of Photogrammetry and Remote Sensing 227 (2025) 775–788

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

MAT-MS: A mask-aware transformer for constructing gap-free MODIS normalized difference snow index products

Jiahui Xu [a,b,c,1], Ruiyang Hua [a,b,c,1], Shujie Wang [d], Stef Lhermitte [e,f], Qingyu Gu [a,b,c], Bailang Yu [a,b,c], Jianping Wu [a,b,c], Yan Huang [a,b,c,*]

[a] Key Laboratory of Geographic Information Science, Ministry of Education, East China Normal University, Shanghai, China
[b] School of Geographic Sciences, East China Normal University, Shanghai, China
[c] Key Laboratory of Spatial-temporal Big Data Analysis and Application of Natural Resources in Megacities, Ministry of Natural Resources, Shanghai, China
[d] Department of Geography, Earth and Environmental Systems Institute, Pennsylvania State University, University Park, PA, USA
[e] Department of Earth & Environmental Sciences, KU Leuven, Leuven, Belgium
[f] Department of Geosciences & Remote Sensing, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Keywords:
Snow cover
MODIS
NDSI
Cloud cover
Transformer
Mask-aware

ABSTRACT

The Normalized Difference Snow Index (NDSI) is essential for accurate snow monitoring, but the widely used MODIS NDSI products generally have significant data gaps mainly due to cloud cover. Existing gap-filling methods often introduce artifact issue in regions with extensive and persistent cloud cover, where gap areas produce inaccurate results influenced by cloud shapes. To address NDSI gap-filling issue, we developed a mask-aware Transformer integrating multi-source data (MAT-MS) to effectively fill these gaps in MODIS NDSI data. The MAT-MS model leverages spatiotemporal information related to meteorology, topography, and geographic location. By incorporating a mask-aware technique, the MAT-MS can learn cloud shapes and patterns, helping to mitigate the common artifact issue. Validation using data from the Tibetan Plateau demonstrated the superior performance of the MAT-MS model, with averaged MAE, RMSE, and R^2 of 1.585, 5.531, and 0.868, respectively. The model reduced RMSE by over 30 % compared to traditional spatiotemporal interpolation methods, and by 9 % compared to mainstream deep learning models. Using MAT-MS, we generated a daily gap-free NDSI dataset for the Tibetan Plateau spanning from 2003 to 2020. This spatiotemporally continuous dataset is critical for detailed snow identification, enabling enhanced estimates of snow cover area, fractional snow cover, and snow depth. The flexibility of the MAT-MS model also makes it applicable to a wide range of continuous remote sensing datasets affected by data gaps.

1. Introduction

Snow cover is a crucial component of Earth's surface, playing a key role in the global climate system (Fyfe et al., 2017). In recent decades, climate warming has led to substantial changes in snow cover patterns worldwide (Pepin et al., 2015; Pulliainen et al., 2020), including reduced snow cover, earlier snowmelt onset, and increased snowmelt runoff (Musselman et al., 2021). These changes have significant implications for water resource management (Kraaijenbrink et al., 2021), ecosystem functions (Shen et al., 2022), and human health (Gottlieb and Mankin 2024). Effective snow cover monitoring is essential for understanding these impacts and making informed decisions in response.

Due to the sparse and uneven distribution of *in situ* observations, satellite remote sensing has become an indispensable tool for continuous snow cover monitoring. The Normalized Difference Snow Index (NDSI) is an important snow detection index for optical satellite imagery. NDSI identifies snow based on its high reflectance in the visible spectrum and low reflectance in the shortwave infrared spectrum, producing a continuous numerical value that effectively indicates the presence of snow within a pixel (Hall et al., 1995). Currently, Moderate Resolution Imaging Spectroradiometer (MODIS) snow products are widely used for snow cover monitoring. In the latest Collection 6 (C6) product, snow cover is reported as NDSI snow cover, rather than as binary Snow Cover Area (SCA) and Fractional Snow Cover (FSC) as in Collection 5 (C5)

* Corresponding author at: Key Laboratory of Geographic Information Science, Ministry of Education, East China Normal University, Shanghai, China.
E-mail address: yhuang@geo.ecnu.edu.cn (Y. Huang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.isprsjprs.2025.07.004
Received 8 December 2024; Received in revised form 24 May 2025; Accepted 5 July 2025
Available online 11 July 2025

(Riggs et al., 2019), addressing several limitations of the previous product collection. While an NDSI threshold value of 0.4 has been commonly used for mapping SCA as a binary snow map (Hall et al., 2002), increasing evidence suggests that the optimal NDSI threshold for SCA depends on various factors, such as landscape conditions, topography, and satellite viewing conditions (Hao et al., 2022; Härer et al., 2018; Zhang et al., 2020). Therefore, applying a uniform threshold of 0.4 for global NDSI binary classification may result in snow detection biases. Similarly, the determination of FSC in the MODIS C5 algorithm is based on a linear regression relationship of NDSI to FSC, developed from empirical relationships between MODIS and Landsat TM data (Salomonson and Appel 2004, 2006). However, due to regional differences in these relationships, using a uniform empirical model can also introduce errors in product accuracy (Huang et al., 2022; Pan et al., 2024). Consequently, MODIS C6 has replaced both SCA and FSC with direct NDSI data (Riggs et al., 2017, 2019; Zhang et al., 2019). Accurate NDSI data enables the derivation of more precise snow metrics, including snowlines (Xiao and Liang 2024) and snow phenology (Notarnicola 2020), which are crucial for understanding snow cover variations in the context of climate change. Recent studies have employed continuous NDSI curves to determine the snowmelt onset date, a critical parameter for evaluating the effect of snowmelt on vegetation, agricultural growing seasons, and other ecological processes (Zheng et al., 2022). The wide applications of NDSI highlight its significance as a tool for monitoring snow cover. However, the effectiveness of these applications relies on the availability of high-quality and spatiotemporally continuous NDSI data.

Numerous studies have reported that MODIS NDSI products achieve an accuracy of over 90 % under clear-sky conditions (Bousbaa et al., 2024; Li et al., 2019). However, due to frequent cloud cover, daily MODIS NDSI products have significant data gaps, resulting in spatiotemporal discontinuities that significantly hinder their application (Hou et al., 2022; Huang et al., 2022; Muhammad and Thapa 2021). Therefore, there is an urgent need to fill these data gaps caused by cloud cover. Nevertheless, filling data gaps in continuous NDSI values presents a greater challenge than filling gaps in binary SCA products. Current methods for addressing this challenge can be broadly categorized into two approaches: traditional spatiotemporal modeling and machine learning approaches. Traditional spatiotemporal methods primarily employ two strategies: one involves interpolating the weight function relationship between the spatial or temporal distances of the data-gap pixels and their surrounding gap-free pixels (Deng et al., 2024; Jing et al., 2022), while the other employs pattern matching to identify and replace data-gap pixels with similar gap-free pixels (Li et al., 2020). Machine learning methods have evolved from ensemble learning approaches like random forest (RF) (Luo et al., 2022) to deep learning techniques including Long Short-Term Memory networks (LSTM) (Hou et al., 2022) and U-Net architectures (Xing et al., 2022).

Despite the effectiveness of traditional spatiotemporal modeling and machine learning methods for filling MODIS NDSI data gaps, both encounter significant challenges in regions with extensive and persistent cloud cover. Traditional spatiotemporal modeling often fails when there are insufficient surrounding pixels for effective interpolation. More critically, both approaches are susceptible to the influence of cloud shapes and extent, leading to inaccurate results and unnatural transitions at the data gap boundaries—commonly referred to as the ''artifact issue'' (Zhang et al., 2022). Xing et al. (2022) mitigated this limitation by integrating partial convolution into the U-Net architecture and reinforcing boundary continuity through a smoothing loss function. However, the effectiveness of this approach was constrained by the inherent limitations of partial convolution—most notably, the use of a binary mask, which tends to discard fine-grained pixel-level information—thereby only partially resolving the artifact issue (Yu et al. 2019). Filling data gaps in MODIS NDSI under conditions of extensive cloud cover still remains a significant challenge. Additionally, NDSI is influenced by various factors, including meteorological conditions, topography, geographical location, and time variations (Xu et al., 2024; You et al., 2020). Integrating this information could substantially improve gap-filled accuracy. Therefore, developing advanced gap-filling models that leverage multi-source data is crucial for overcoming these challenges in complex environments.

Recently, Transformer models have gained significant attention for their ability to capture long-range dependencies (Dosovitskiy et al., 2021; Vaswani et al., 2017). Their embedding structure is particularly well-suited for integrating multi-source data. Meanwhile, the mask-aware technique shows great potential in mitigating artifact issue by more effectively focusing on data-gap regions, while preserving the original gap-free information and preventing model error propagation (Li et al., 2022; Motamed et al., 2023). Building upon these advancements, this study proposes a Mask-Aware Transformer integrating Multi-Source data (MAT-MS) model to address the limitations of existing gap-filling methods. We selected the Tibetan Plateau as the case study area to validate the effectiveness of the MAT-MS model. The Tibetan Plateau's complex topography, heterogeneous snow cover distribution, and extensive cloud cover make it an ideal yet challenging environment for gap-filling model validation.

The paper is organized as follows: we begin by introducing the study area and datasets used. Next, we provide a detailed description of the MAT-MS model and evaluate its performance, with a particular focus on its ability to address artifact issue and fill data gaps across varying conditions. Finally, we present a 20-year daily gap-free NDSI dataset for the Tibetan Plateau, generated using the MAT-MS model, and explore its potential applications of this spatiotemporally continuous NDSI product.

2. Case study area and datasets

2.1. Case study area

The Tibetan Plateau encompasses the world's largest nonpolar terrestrial cryosphere and is one of the most sensitive regions to climate change (Yao et al., 2022). It is characterized by complex topography and extreme climatic conditions, including low precipitation, cold temperatures, and strong solar radiation. Snow cover on the Tibetan Plateau is typically shallow, patchy, and changes rapidly over time (You et al., 2020; Zhang et al., 2023), resulting in uneven and rapidly fluctuating distributions of NDSI. Additionally, due to the monsoon circulation in Oceania, warm and moist air from low-latitude oceans is continuously transported inland, supplying abundant water vapor for cloud formation. This leads to extensive and irregular cloud cover over the Tibetan Plateau (Wu et al., 2024), with over 58 % of the area under cloud cover for more than 120 days per year (Fig. 1a). These factors make filling MODIS NDSI data gaps in the Tibetan Plateau particularly challenging.

2.2. Datasets

2.2.1. MODSI NDSI products

We used daily MODIS C6 Terra (MOD10A1) and Aqua (MYD10A1) snow products to fill data gaps, with a spatial resolution of 500 m, covering the period from 1 January 2003 to 31 December 2020. The data were accessed via the Google Earth Engine cloud platform (https://earthengine.google.com/, last accessed 15 September 2023) and reprojected to Universal Transverse Mercator (UTM) zone 45. The original MODIS snow products include raw NDSI data, NDSI snow cover, snow albedo, and quality control flags, with the raw NDSI data and NDSI snow cover used in this study. The preprocessing steps were as follows:

Firstly, we used the NDSI snow cover as a mask to identify pixels in the raw NDSI data corresponding to cloud, missing data, detector saturated, night, no decision, and reclassified them as data-gap pixels. Pixels unaffected by these issues retained their original values. We then combined the reclassified MOD10A1 and MYD10A1 data for the same day using the following rules (Deng et al., 2024): if a pixel was available in both the MOD10A1 and MYD10A1 products, the MOD10A1 NDSI was

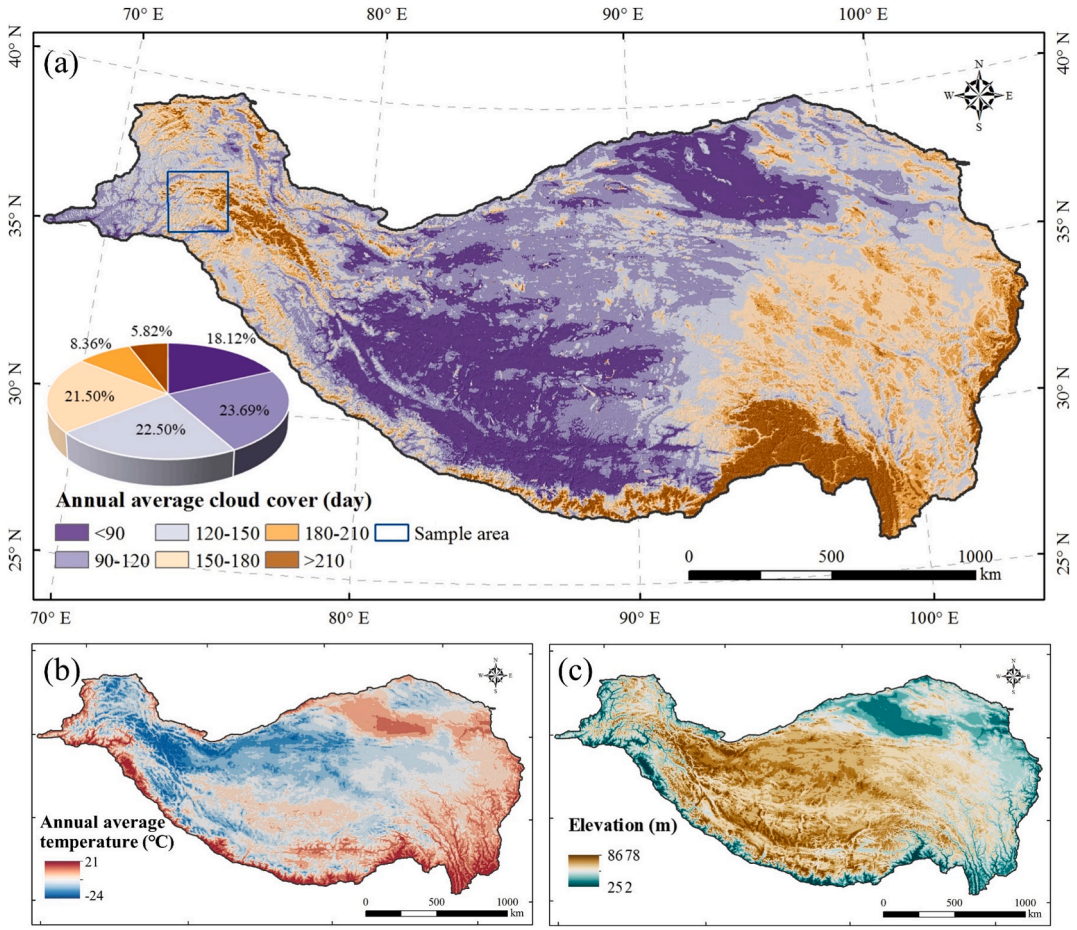

Fig. 1. Distribution of annual average cloud cover days (a), annual average temperature from 2003 to 2020 (b), and elevation (c) on the Tibetan Plateau. Note: the blue rectangle in panel (a) marks the sample area analyzed in Section 5.2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

used; if a pixel was available in only one product, the available observation was retained. This approach efficiently reduced cloud coverage by 5 %–20 % with minimal loss in precision. Finally, the combined NDSI values, originally ranging from −10,000 to 10,000, were normalized to a range of −1 to 1.

2.2.2. *Auxiliary data*

Since extensive studies have identified temperature as the primary factor influencing snow cover variability (Musselman et al., 2021; Zhao et al., 2022), we selected temperature as the key meteorological variable in our modeling. We used daily temperature data from the National Tibetan Plateau Data Center (https://data.tpdc.ac.cn/, last accessed 21 December 2023) for the Tibetan Plateau at a resolution of 1/30°. This dataset, generated by integrating long-term ERA5 reanalysis, short-term high-resolution atmospheric simulations, and *in situ* observations, has been validated as more accurate than widely used reanalysis datasets (He et al., 2020; Yang et al., 2023a). All daily temperature data from 2003 to 2020 was first reprojected and resampled to a 500 m resolution to align with the MODIS snow product, and then normalized.

In addition, elevation is recognized as a critical factor influencing the timing of snow accumulation and melting, with higher elevations typically experiencing earlier accumulation and later melting compared to lower elevations (Ma et al., 2023). To incorporate topographic information, we used elevation data from the 90 m Shuttle Radar Topography Mission (SRTM) gridded digital elevation model (DEM). The DEM data was reprojected and resampled to 500 m to maintain consistency with the MODIS snow product.

3. Mask-aware Transformer integrating multi-source data (MAT-MS)

The flowchart of the MAT-MS model is described in Fig. 2. The model incorporates temperature, topography, geographic location (latitude and longitude), time variations (date, ranging from 1 to 365), and the spatiotemporal information of NDSI values derived from traditional spatiotemporal interpolation. The MAT-MS effectively integrates this information while also learning cloud patterns using masks that indicate data gaps. Consequently, our MAT-MS model input consists of four channels: spatiotemporal interpolated NDSI, cloud mask, DEM, and temperature, along with a three-dimensional vector including date, latitude, and longitude (Fig. 2). The input slice size is set to 128 × 128 for computational efficiency. The MAT-MS architecture integrates a Transformer-based encoder with a CNN-based decoder for NDSI reconstruction. Then, the model's accuracy was compared with mainstream models across varying conditions. Finally, we produced a 20-year daily gap-free NDSI dataset for the Tibetan Plateau using the MAT-MS model.

3.1. *Model structure*

The Transformer model has achieved significant advancements in image reconstruction (Motamed et al., 2023; Yang et al., 2023b), which is analogous to filling data gaps in remote sensing images. Its core strength lies in the self-attention mechanism, which evaluates correlation scores across different regions of an input image to model global contextual similarity (e.g., spatial correlations between distant regions) (Vaswani et al., 2017). It also exhibits superior scalability and

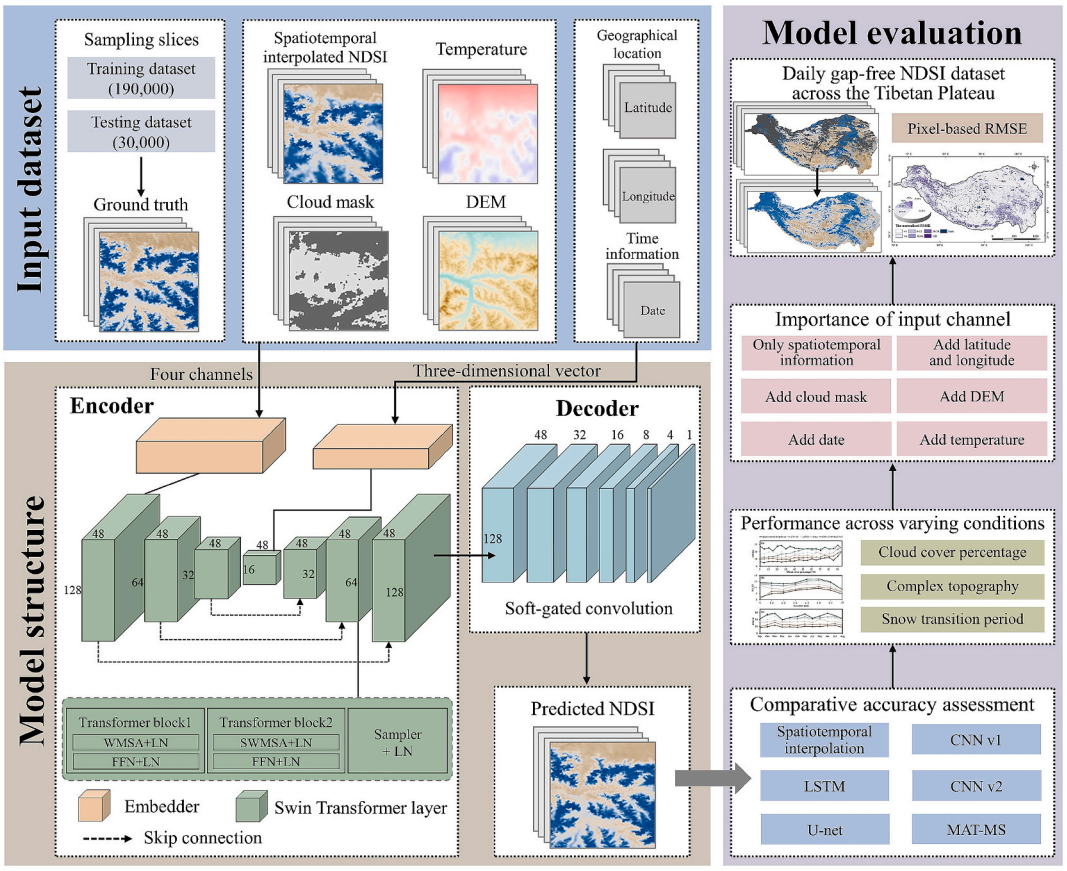

Fig. 2. Flowchart of the mask-aware Transformer integrating multi-source data (MAT-MS) model for filling data gaps in MODIS NDSI.

compatibility, making it particularly suitable for processing remote sensing datasets (Zhao et al., 2023; Zhou et al., 2023). In addition, the Convolutional Neural Networks (CNN) model is more effective at capturing local spatial hierarchies within an image through shared convolutional kernels (Redmon et al., 2016). Therefore, we employed the Transformer as the encoder to extract global contextual similarity, and the CNN as the decoder to reconstruct localized spatial details to combine their complementary strengths.

To address the cloud-induced artifact issue, we introduce a novel mask-aware technique integrated into the architecture through three key modifications (Li et al., 2022). Specifically, the cloud mask is processed as an independent input channel in the encoder, while the decoder employs a soft-gated convolution instead of traditional convolution (Yu et al., 2019), coupled with a weighted composite loss function that differentially handles data-gap and gap-free regions. These modifications collectively enhance the model's recognition of cloud shapes and patterns, while improving boundary precision between data-gap and gap-free regions.

Compared to prior approaches, our methodology emphasizes leveraging a robust encoder for novel representation of multi-source

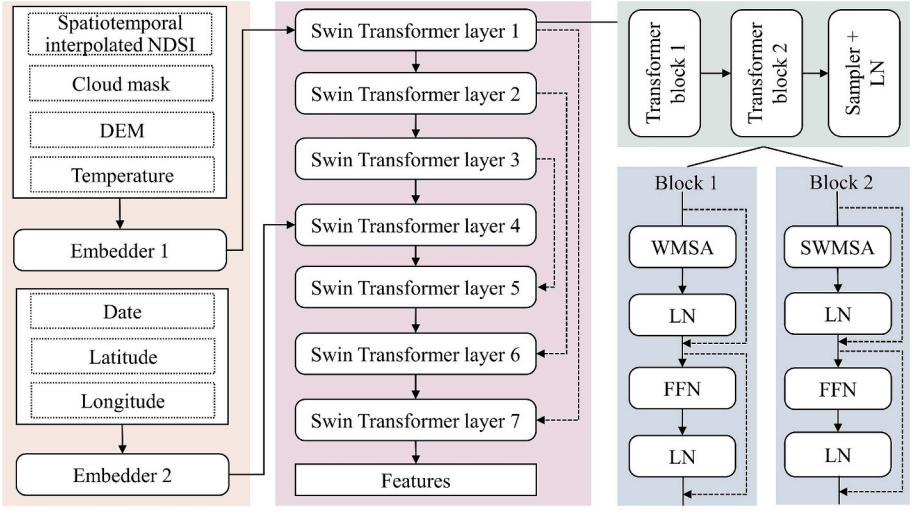

Fig. 3. The schematic diagram of the encoder in the MAT-MS model.

information, while employing a mask-aware technique to further mitigate artifact issue.

3.1.1. Encoder

We employ the widely used Swin Transformer architecture as the encoder, which effectively balances model performance and computational complexity (Liu et al., 2021). The encoder first consists of embedders (Fig. 3), which map low-dimensional input data into high-dimensional representations, enabling features to form more complex and nonlinear combinations. This is followed by a series of Transformer layers that further refine these representations.

During the embedding stage, the cloud mask is also incorporated as an independent input channel, enabling explicit feature differentiation between data-gap and gap-free regions (Fig. 3). Along with spatiotemporal interpolated NDSI, temperature, and DEM, this four-channel data was first embedded to a 48-dimensional feature maps through a 1×1 convolutional embedder (Fig. 3). The three-dimensional vector (date, latitude, and longitude) was embedded into a 48-dimensional vector using a three-layer linear projection with Gaussian Error Linear Unit (GELU) activation (Zheng et al., 2022a). GELU is a smooth activation function that scales inputs according to their cumulative probability under the standard normal distribution. It has been widely adopted in deep learning as a modern alternative to the ReLU activation function (Liu et al., 2022).

Each of our Transformer layers contains two consecutive blocks (Fig. 3, Liu et al., 2021): window-based multi-head self-attention (*WMSA*), and its shifted version (*SWMSA*). *WMSA* is a computational block that associates features within the windows by dividing the feature map into windows of an appropriate size. *SWMSA* is a similar block, except that it shifts the window by half of its size. Therefore, the features within the windows can also compute attention with adjacent windows. We applied layer normalization (*LN*) at the output of each block following Liu et al. (2022):

$$attn_1 = LN\left(WMSA\left(x^{l-1}\right)\right) \tag{1}$$

$$\widehat{x}_1^l = LN\left(FFN\left(attn_1 + x^{l-1}\right)\right) \tag{2}$$

$$attn_2 = LN\left(SWMSA\left(\widehat{x}_1^l\right)\right) \tag{3}$$

$$\widehat{x}_2^l = LN\left(FFN\left(attn_2 + \widehat{x}_1^l\right)\right) \tag{4}$$

$$x^l = LN\left(sampler\left(\widehat{x}_2^l\right)\right) \tag{5}$$

where $\widehat{x}_1^l$ and $\widehat{x}_2^l$ represent the outputs from the l^{th} layer's *WMSA* and *SWMSA*, respectively, while x^l denotes the final output of the l^{th} layer. *FFN* represents a fully connected feed-forward network, including two linear projections and a GELU activation function. *sampler* after blocks represents the resampling of the size of the feature map $\widehat{x}^l$.

We set seven Transformer layers (Figs. 2 and 3): the first three layers use downsampling to extract high-dimensional features, the middle layer concatenates the embedded vector, and the last three layers use upsampling to restore the size of the input slices, with skip connection to reduce the loss of feature information (Li et al., 2022). The attention heads patch size, window size, and expansion scale of the FFN were set to 3, 1, 8, and 3, respectively.

3.1.2. Decoder

The decoder is designed to dynamically select the channels of critical features extracted from the aforementioned encoder. Compared to traditional convolutional layers, soft-gated convolution enables the model to capture feature interactions while maintaining relatively low computational complexity (Yu et al., 2019). More importantly, it dynamically adjusts the weighting of pixel-level information, effectively suppressing the propagation of errors from inaccurately reconstructed regions and preserving feature continuity, thereby alleviating the artifact issue.

Our decoder has six layers of soft-gated convolution, with the number of channels gradually decreasing from 48 to 1 (Fig. 2). The output of the l^{th} soft-gated convolution ($Layer_l$) is calculated as follows:

$$Layer_l = \sigma(Conv_1(Layer_{l-1})) \times \phi(Conv_2(Layer_{l-1})) \tag{6}$$

where σ and ϕ are the sigmoid function and exponential linear unit (ELU) function, respectively. $Conv_1$ and $Conv_2$ represent two convolution filters with the same channels of input and output, but with different parameters.

3.1.3. Loss function

Conventional loss functions typically compute mean squared error (MSE) solely within data-gap pixels during backpropagation. However, this approach often results in boundary discontinuities between data-gap and gap-free regions. To address this limitation, a weighted composite loss function within both data-gap and gap-free pixels was introduced to enhance the smoothness of cloud mask boundaries:

$$Loss = \alpha \bullet MSE(G, \widehat{x} - x) + \beta \bullet MSE(GF, \widehat{x} - x) \tag{7}$$

$$MSE = \sum M \bullet (\widehat{x} - x)^2 \Big/ \sum M \tag{8}$$

where x and $\widehat{x}$ represent the ground truth and the predicted NDSI value, respectively. G denotes the data-gap regions, GF denotes the gap-free regions. Based on experimental validation and practical considerations, the weighting coefficients are set to $\alpha = 10$ for the data-gap regions, and $\beta = 1$ for the gap-free regions. The *MSE* function computes the average squared difference between x and $\widehat{x}$, where M denotes the number of pixels in the respective region.

3.2. Dataset generation

3.2.1. Input dataset generation

Given the challenges of obtaining actual NDSI values under cloud-covered conditions, we established ground truth for both training and testing using gap-free NDSI slices sampled across the Tibetan Plateau. Data from 2003 to 2018 were used for training, while data from 2019 to 2020 were used for testing. Due to the highly uneven spatial distribution of cloud cover across the Tibetan Plateau (Fig. 1a), we balanced the dataset by ensuring an approximately equal number of samples across different regions of the plateau. Using this approach, we sampled 190,000 NDSI slices for training and 30,000 NDSI slices for testing (hereafter referred to as the baseline test dataset). Additionally, we generated a snow-intensive test dataset to evaluate the model's performance in snow-covered areas. This snow-intensive test dataset consisted of 2,336 NDSI slices, selected from the baseline test dataset, where at least 50 % of the pixels have NDSI values greater than 0.1 (Zhang et al., 2020; Zhang et al., 2019).

We then overlaid a range of cloud masks onto ground truth to simulate data gaps caused by cloud cover (Fig. 4). These cloud masks were constructed directly from actual data gaps, ensuring that the cloud distribution accurately reflected real-world conditions. A total of 76,000 cloud mask slices were obtained. Subsequently, we implemented a dynamic cloud mask matching strategy to reduce the model's dependency on specific cloud shapes. Compared to static cloud masks strategy, this adaptive data generation strategy continuously produced various training inputs, thereby enhancing model robustness to heterogeneous cloud cover shapes and patterns.

3.2.2. Spatiotemporal interpolation

In computer vision, data-gap pixels are often initialized with fixed values (e.g., grayscale intensities) before being processed by deep learning models. Building on this strategy, while leveraging the spatio-

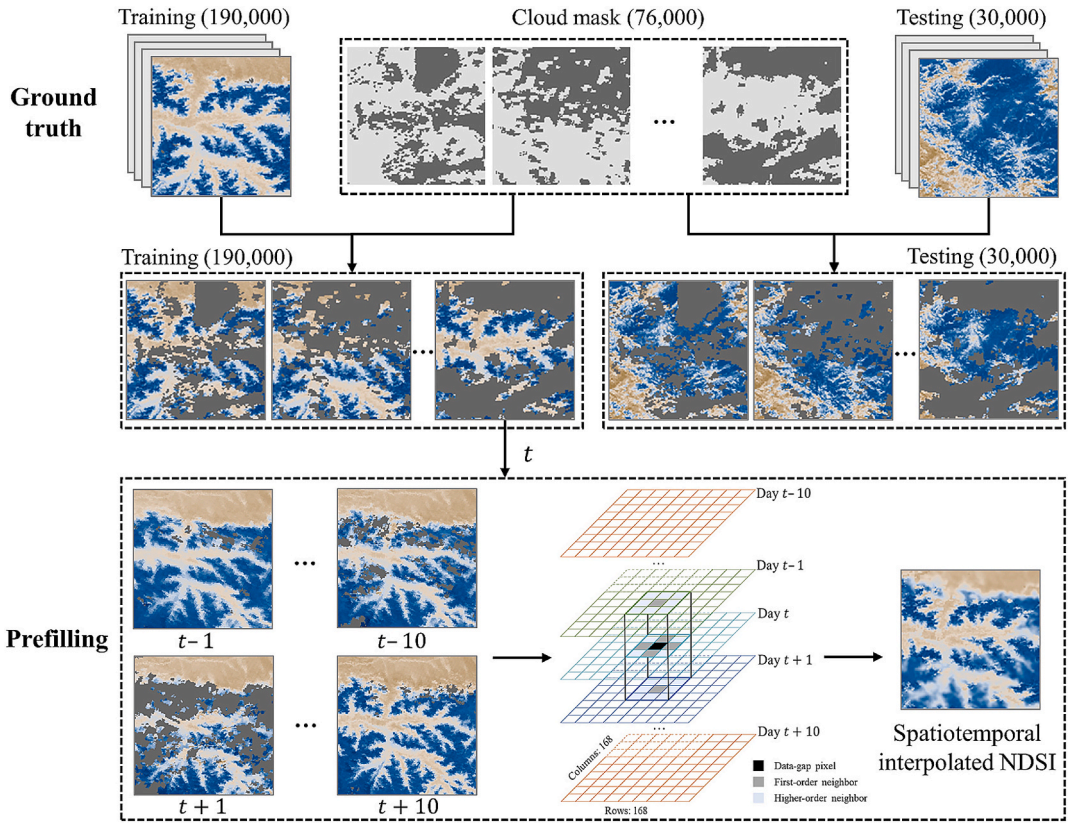

Fig. 4. The construction and spatiotemporal interpolation for data-gap NDSI.

temporal characteristics inherent in remote sensing data (Huang et al., 2022; Jing et al., 2022), we first employed a traditional spatiotemporal interpolation model to prefill the gaps, followed by refinement using the MAT-MS model. To effectively capture the relevant spatiotemporal information, the spatial and temporal distances were set to 20 and 10, respectively, generating a 168 × 168 × 21 spatiotemporal cube (Fig. 4). We then applied a sliding 3 × 3 × 3 spatiotemporal cube across this larger structure, using an inverse distance-weighted (IDW) averaging approach to fill data gaps by incorporating information from valid neighboring pixels (Deng et al., 2024):

$$NDSI_{i,j,t} = \sum_{e} W_{e(i,j,t)} NDSI_{e(i,j,t)} \Big/ \sum_{e} W_{e(i,j,t)} \tag{9}$$

$$W_{e(i,j,t)} = Dist^{-2}(e(i,j,t),(i,j,t)) \tag{10}$$

where e is an enumeration function of the coordinates of valid pixels around the center pixel of the spatiotemporal coordinate (i, j, t). *Dist* means Euclidean Distance.

This initial spatiotemporal interpolated NDSI, along with the corresponding cloud mask, DEM, temperature, and a three-dimensional vector including date, latitude, and longitude, served as the input for model training (Fig. 2).

3.3. Model evaluation

To better illustrate the variations in NDSI, we rescaled the original NDSI values from the range of −1 to 1 to a new range of −100 to 100. Following Jing et al. (2022) for accuracy calculations, NDSI values below 0 were reclassified as 0, and accuracy was computed using values greater than or equal to 0. We evaluated our model's performance using mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2):

$$MAE = \frac{\sum_{n=1}^{N} |x_{p,n} - x_{t,n}|}{N} \tag{11}$$

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N} (x_{p,n} - x_{t,n})^2}{N}} \tag{12}$$

$$R^2 = \frac{\sum_{n=1}^{N} (x_{p,n} - \bar{x}_t)^2}{\sum_{n=1}^{N} (x_{t,n} - \bar{x}_t)^2} \tag{13}$$

where $x_{p,n}$ and $x_{t,n}$ represent predicted NDSI value and ground truth in pixel n, respectively. N is the total number of data-gap pixels. $\bar{x}_t$ is the average of all ground truth NDSI values.

Considering the capabilities of LSTM (Hou et al., 2022) and U-Net (Xing et al., 2022) in NDSI gap-filling tasks, we also constructed models based on architectures from those previous studies for comparative analysis. Additionally, we included a comparison with CNN models, given their widespread use in image processing tasks (Redmon et al., 2016). The first version, CNN v1 model, is similar to our decoder but without the encoder output (Yu et al., 2019). This comparison highlights the role of the encoder in enhancing the model's ability to understand and process the inputs. The second version, CNN v2 model, uses the same decoder but incorporates contextual attention in the encoder (Yu et al., 2018). This comparison directly demonstrates the strength of the MAT-MS model's ability to capture global contextual similarity.

4. Results

4.1. Accuracy assessment of the MAT-MS model

Our MAT-MS model achieved an MAE of 1.585, an RMSE of 5.531, and an R^2 of 0.868 on the baseline test dataset, and an MAE of 5.760, an RMSE of 9.899, and an R^2 of 0.879 on the snow-intensive test dataset

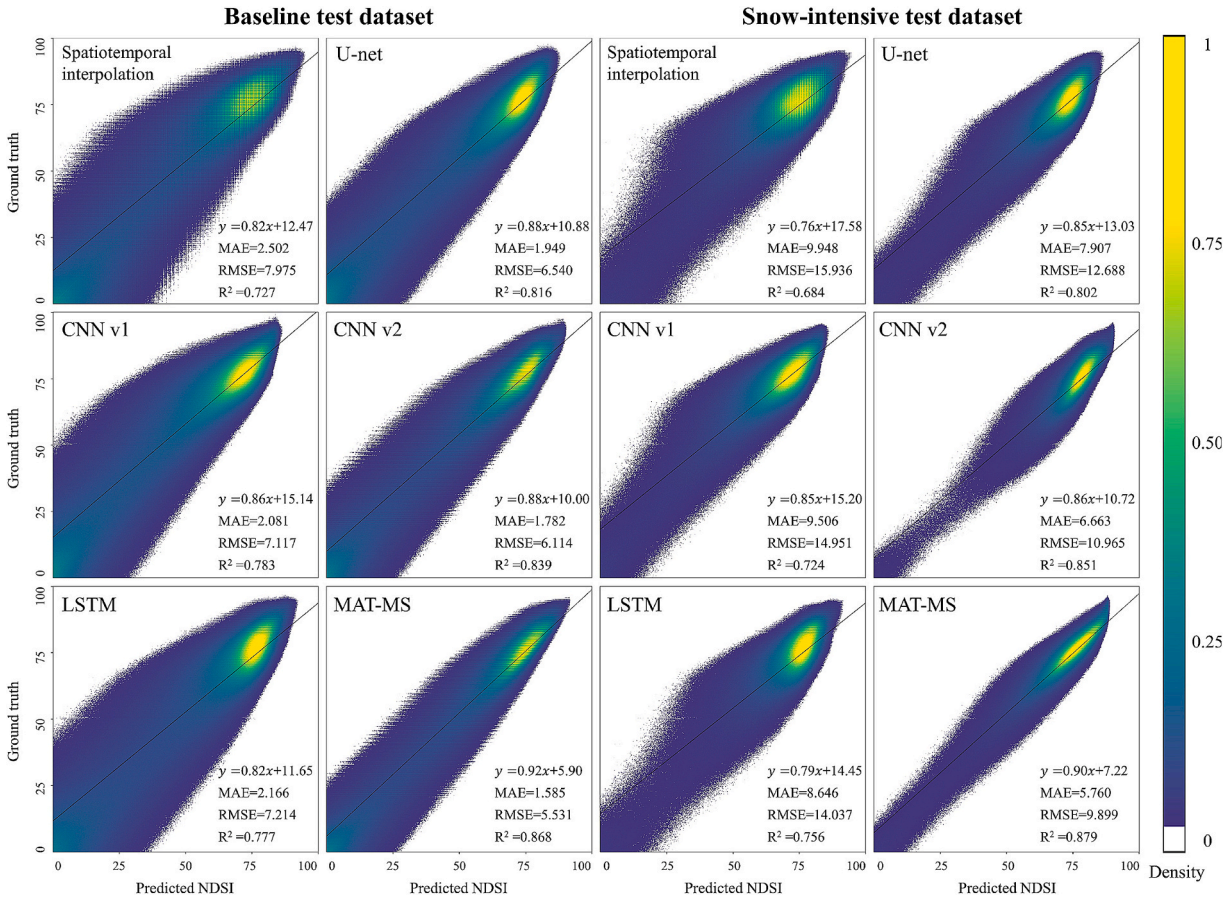

Fig. 5. Accuracy evaluation of different models for filling data gaps in MODIS NDSI. Note: the values at 0 are not displayed on the scatter plot due to the large number of points at that position, but they are still included in the accuracy metrics calculations.

(Fig. 5). Compared to the traditional spatiotemporal interpolation, our model reduced MAE and RMSE by 36.65 % and 30.65 %, respectively, while increasing R^2 by 19.39 % on the baseline test dataset (Fig. 5). These improvements were even more pronounced on the snow-intensive test dataset, with a 37.88 % reduction in RMSE, demonstrating the MAT-MS's ability to effectively refine data and maintain strong robustness against potential errors in spatiotemporal interpolation inputs. Comparative analysis with mainstream deep learning architectures also revealed improvements. On the snow-intensive dataset, our model achieved RMSE reductions of 33.79 % (vs CNN v1), 29.48 % (vs LSTM), 21.98 % (vs U-net), and 9.72 % (vs CNN v2), respectively (Fig. 5).

To further illustrate these improvements, we visualized the results of different models (Fig. 6). The spatiotemporal interpolation (Fig. 6b1-b4), CNN v1 (Fig. 6c1-c4), LSTM (Fig. 6d1-d4), and U-net (Fig. 6e1-e4) models, which lack a mask-aware technique, often suffer from cloud-induced artifact issue, resulting in unnatural transitions and inaccurate gap-filled results. In contrast, both the CNN v2 (Fig. 6f1-f4) and MAT-MS (Fig. 6g1-g4) models, which incorporate mask-aware processing, significantly mitigated artifact issue, producing NDSI values that more accurately approximate true NDSI value while preserving edge coherence. The MAT-MS model outperformed CNN v2 by enhancing the texture and patterns of snow cover, owing to the Swin Transformer encoder's ability to model global context effectively. Additionally, under extreme cloud cover conditions (Fig. 6a5), where comparative models failed to generate high precision gap-filled results (Fig. 6b5-f5), the MAT-MS outputs demonstrated remarkable alignment with the ground truth (Fig. 6g5 and 6 h5). These advancements -including artifact reduction, improved edge coherence, better spatial texture preservation, and superior performance under extreme cloud conditions- enable the MAT-MS model to achieve precise NDSI prediction, effectively reducing both overestimation and underestimation biases.

4.2. *Performance across varying conditions*

4.2.1. *Cloud cover testing*

We further evaluated the models' performance across varying cloud cover percentages based on the snow-intensive test dataset (Fig. 7a). While all models experienced accuracy degradation with increasing cloud cover, the MAT-MS model consistently demonstrated superior performance. Specifically, MAT-MS maintained RMSE values below 9 for cloud cover less than 45 % and remained under 11 for cloud coverage between 50 % and 85 %. Even under dense cloud conditions (85 %–95 %), the model maintained great performance, yielding RMSE values within 11–12. Under complete cloud obstruction (100 %), the MAT-MS model has its highest RMSE of 14.35, yet still outperformed spatiotemporal interpolation (15.50), CNN v1 (16.92), LSTM (15.61), U-Net (15.90), and CNN v2 (14.70).

4.2.2. *Complex topography and snow transition period*

Given the challenges posed by the Tibetan Plateau's complex topography and seasonal snow cover variations, we conducted additional evaluations to assess model performance across different elevations and months. The results indicate that performance trends remain generally consistent across these factors. Statistical analyses show that the MAT-MS model maintains stable accuracy even under extreme high-elevation conditions (>5 km, Fig. 7b), consistently achieving lower RMSE values than other models. During critical snow transition periods, particularly in October-November (snow accumulation) and February-April (snowmelt) periods, the MAT-MS model outperforms its

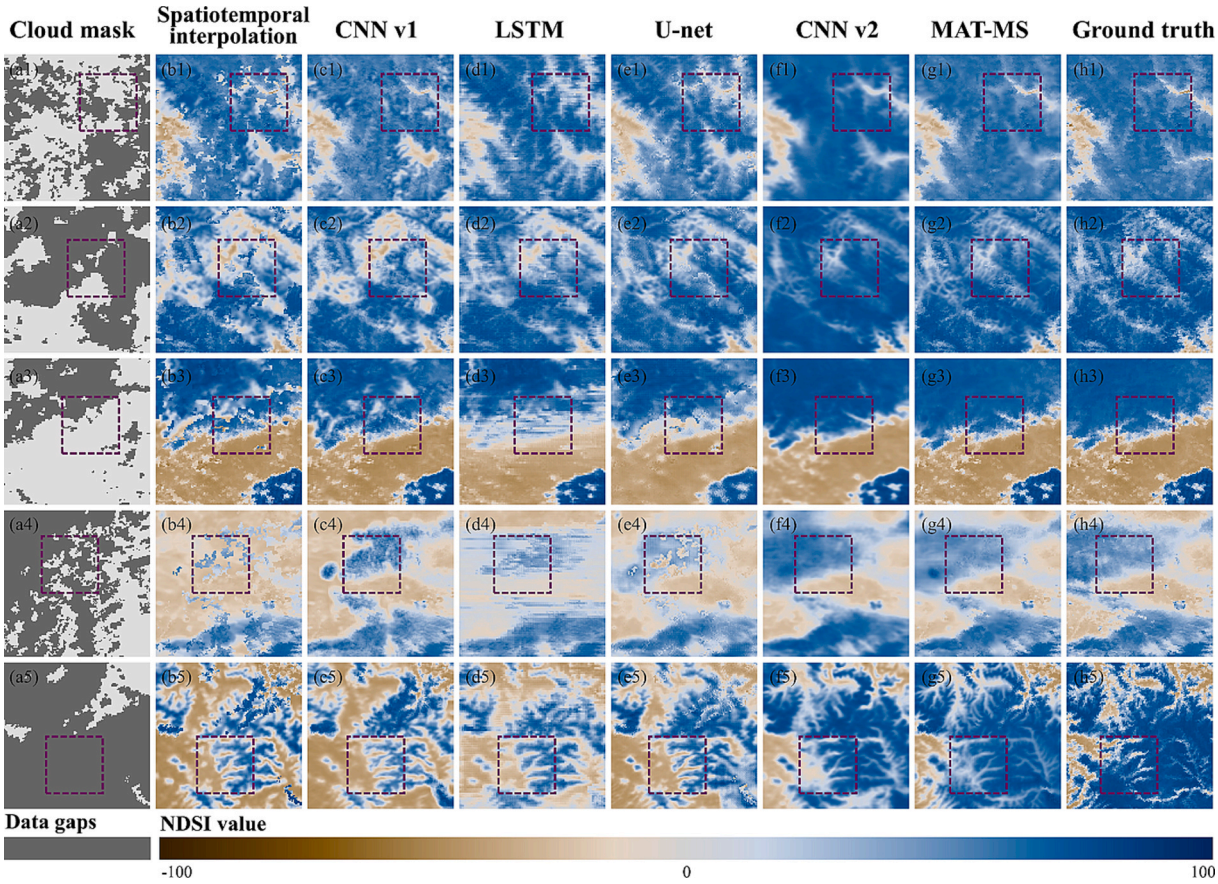

Fig. 6. Comparison of the predicted results of different models at 128 × 128 pixel level: cloud mask for input (a1-a5), spatiotemporal interpolated NDSI (b1-b5), CNNv1 outputs (c1-c5), LSTM outputs (d1-d5), U-Net outputs (e1-e5), CNN v2 outputs (f1-f5), MAT-MS outputs (g1-g5), and ground truth (h1-h5). Note: the purple-boxed areas highlight regions of significant improvement. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

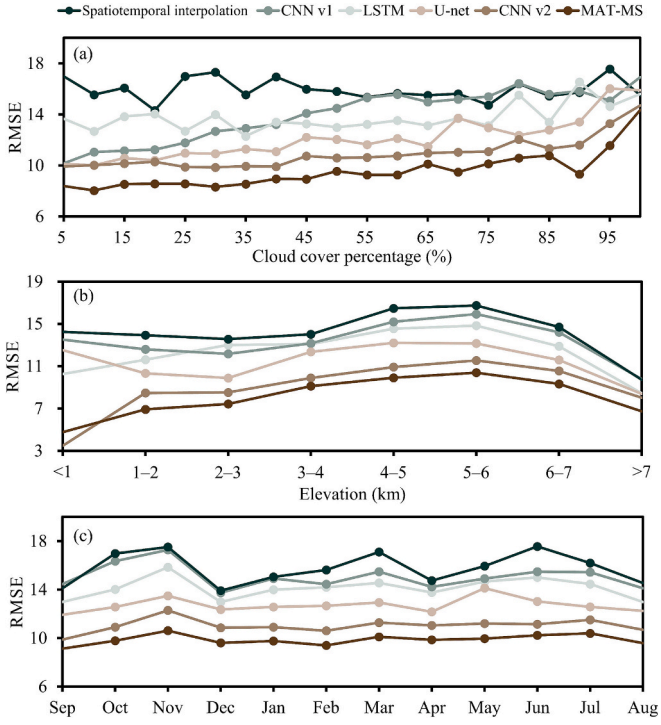

Fig. 7. The models' performance across varying cloud cover percentages (a), elevation (b), and month (c) based on the snow-intensive test dataset.

counterparts in both accuracy and robustness (Fig. 7c). These findings suggest that MAT-MS is better suited to capturing the dynamic evolution of snow cover and mitigating the challenges posed by complex topographical influences.

4.2.3. *Importance of input channel*

To demonstrate the importance of input channels, we conducted an ablation experiment by progressively adding different data sources to the MAT-MS model and comparing the resulting accuracy. To reduce the computational costs, we employed a light-weight version of the model and trained it to convergence based on the snow-intensive test dataset. We began with an initial model using only spatiotemporal information and then sequentially added cloud mask, date, latitude and longitude, DEM, and temperature data.

As illustrated in Fig. 8, incorporating DEM led to the most substantial improvement, reducing the RMSE by 0.611 (from 14.913 to 14.301), highlighting the critical role of topography. Including the cloud mask as an independent input channel allowed the model to identify data-gap regions, leading to the second-largest RMSE reduction of 0.457. The addition of date, latitude, and longitude resulted in modest improvements, reducing the RMSE by 0.027 and 0.152, respectively. While these vectors contain relatively less information compared to other data sources, they provide important macro-level spatiotemporal constraints.

These results indicate that relying solely on spatiotemporal information is insufficient to address the complexities associated with extensive data gaps. Instead, integrating multi-source information proves to be a more effective approach, particularly for regions with extensive gaps. Future research should explore incorporating additional information to further enhance accuracy.

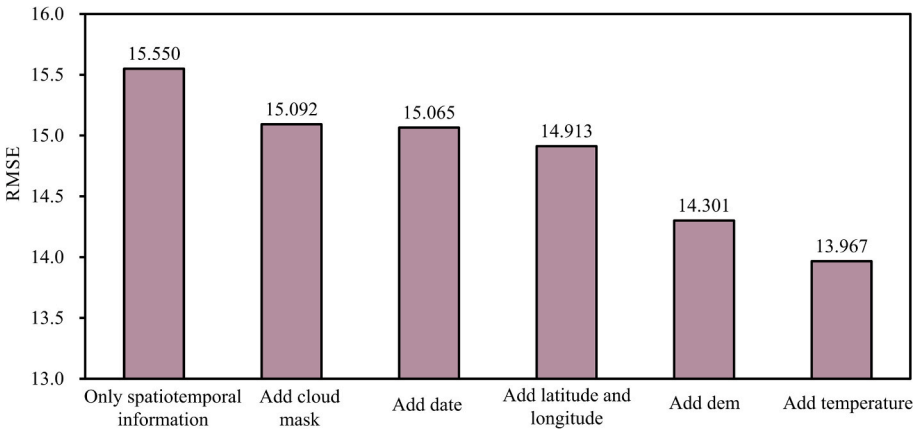

Fig. 8. The evaluation of incorporating different input channels using a converged light-weight MAT-MS model based on the snow-intensive test dataset.

4.3. Daily gap-free NDSI dataset across the Tibetan Plateau

The comprehensive evaluations confirm the effectiveness of the MAT-MS model in addressing the challenges posed by persistent cloud cover and complex topography over the Tibetan Plateau. Using this trained model, we reconstructed the original MODIS NDSI dataset by

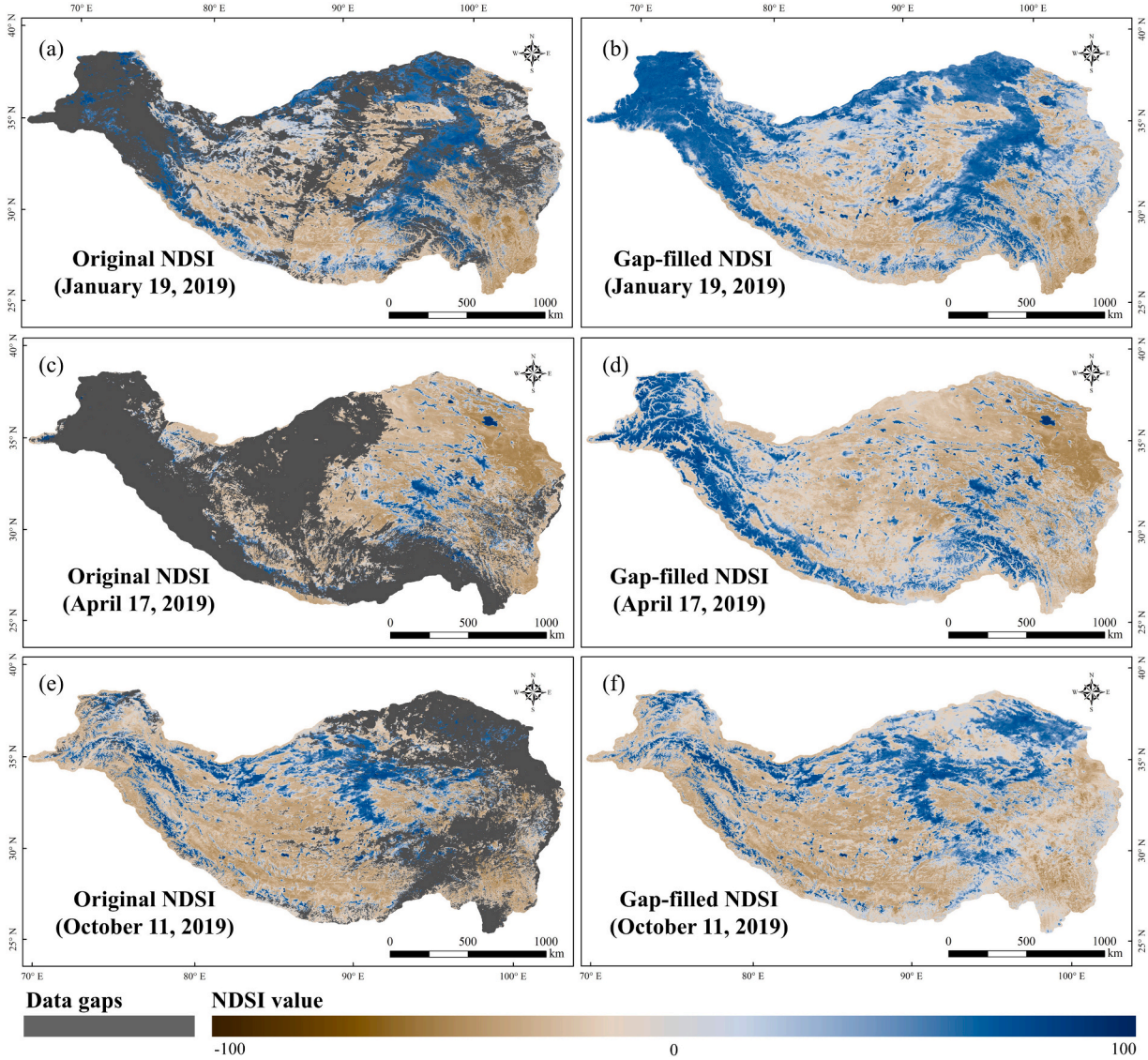

Fig. 9. Comparison of original and gap-filled MODIS NDSI from the MAT-MS model during three snow phenological stages on the Tibetan Plateau: snow-stable (a-b), melting (c-d), and accumulation (e-f).

filling data gaps across the entire plateau, generating a daily gap-free NDSI dataset at 500 m resolution spanning from 2003 to 2020. To illustrate the gap-filled results, we randomly selected three days in 2019, each representing a distinct snow phenological stage: snow-stable (January 19, 2019), melting (April 17, 2019), and accumulation (October 11, 2019) (Fig. 9). Our results demonstrate robust gap-filling capabilities across all critical snow phenological stages, ensuring seamless spatiotemporal continuity in the reconstructed data while effectively mitigating cloud-induced artifact issue. This improvement is particularly valuable for applications requiring high-precision snow cover mapping, as it enables accurate capturing of fine-scale variations in snow distribution.

Furthermore, we provide the spatial distribution of RMSE across the entire plateau as a reference for users. All test datasets were spatially aligned to obtain pixel-based model accuracy (Fig. 10). The results indicate high model accuracy, with 58.16 % of the area exhibiting an RMSE below 4. Only 0.90 % of the region recorded an RMSE exceeding 20, primarily in regions with high cloud cover. However, a comparison between the RMSE spatial distribution and average annual cloud cover (Fig. 1a) reveals no significant increase in errors in regions with persistent cloud cover. These findings demonstrate the effectiveness of the sampling strategy described in Section 3.2.1 and highlight the model's robustness even under challenging conditions.

5. Discussion

5.1. Generalizability of MAT-MS model

This study developed a MAT-MS model for snow cover reconstruction, achieving high accuracy in filling data gaps in MODIS NDSI. Our model effectively leverages spatiotemporal information as input while avoiding interference from it, excelling particularly in challenging regions characterized by extensive cloud cover and complex topography – areas where traditional spatiotemporal interpolation methods exhibit critical limitations. By incorporating a spatiotemporal interpreted NDSI as input and employing Transformer-based architecture, we were able to characterize short-term snow dynamics while mitigating temporal discontinuity in the original data (Figs. 5 and 6).

Additionally, previous studies have often overlooked the cloud-induced artifact issue, leading to unnatural transitions and inaccurate gap-filled results (Hou et al., 2022; Xing et al., 2022). To overcome these limitations, we developed a novel mask-aware technique integrated into the architecture through three key modifications: (1) processing the cloud mask as an independent encoder input, (2) implementing a soft-gated convolution in the decoder, and (3) incorporating a weighted composite loss function that operates on both data-gap and gap-free pixels. To validate the generalizability of the mask-aware technique, we implemented it into the existing deep learning models for NDSI gap-filling (LSTM and U-Net). Cross-model validation showed RMSE reductions of 16.81 % and 6.23 % compared to their original implementations without the mask-aware technique based on the snow-intensive test dataset (Fig. S1), demonstrating consistent improvements across different deep learning architectures.

The superior performance of the MAT-MS model also stems from its effective use of the Transformer's self-attention mechanism, which captures long-range spatial dependencies. Comparative experiments with the CNN v2 model, which employs the same mask-aware technique and decoder architecture but lacks the self-attention mechanism, revealed that the MAT-MS model better preserves the spatial texture of snow cover (Fig. 6).

Another strength of the MAT-MS model is its flexibility and generalizability. The architecture and code are freely available to the research community (see Data availability section for details), and the model can be adapted to include auxiliary information based on specific study areas. For instance, in our case study of the Tibetan Plateau, topographic data was essential due to the region's complexity. However, when applying this model to other regions, such as Northeastern China, where forest cover significantly impacts the gap-filling accuracy (Hao et al., 2022), researchers can easily adjust the auxiliary inputs to better suit local environmental conditions. Moreover, our model is not limited to NDSI datasets. Its flexibility makes it applicable to a wide range of continuous remote sensing datasets with data gaps, such as Land Surface Temperature (LST) (Wang et al., 2024), Leaf Area Index (LAI) (Zhu et al., 2022), and Aerosol Optical Depth (AOD) (Bai et al., 2024). The encoder-decoder structure of the model is also designed for scalability, allowing it to be expanded or simplified based on available computational resources and data requirements.

However, the current framework has certain limitations. While the model significantly improves temporal continuity, its reliance on spatiotemporal interpolation-based prefilling assumes linear

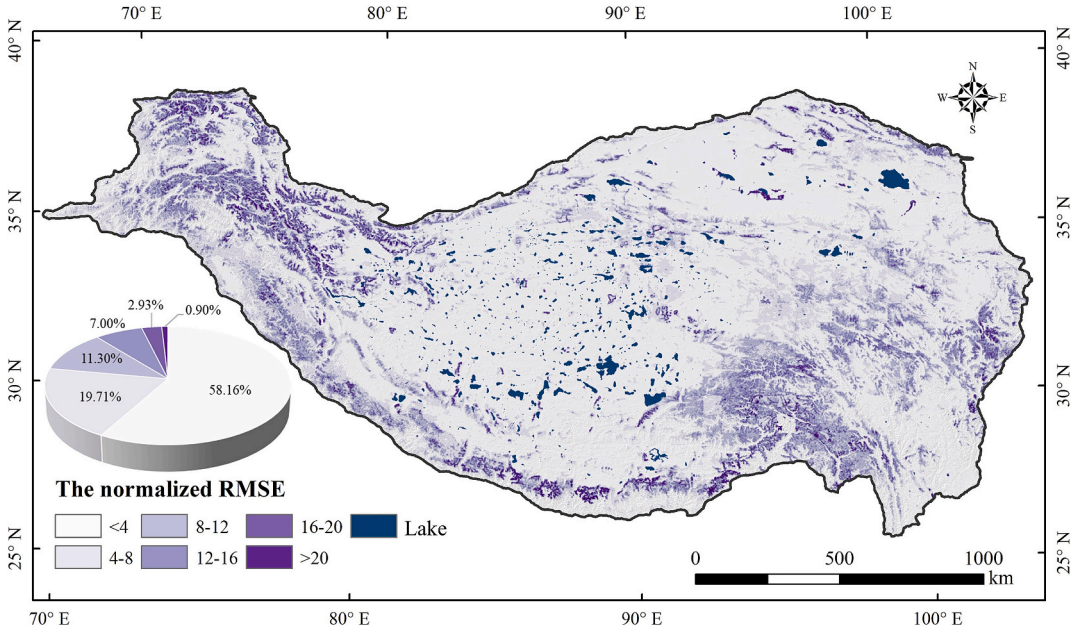

Fig. 10. The spatial distribution of the root mean square error (RMSE) across the Tibetan Plateau.

relationships that may not fully capture the nonlinear response mechanisms inherent in snowmelt dynamics. Future research should focus on developing nonlinear temporal modeling approaches to better represent progressive snow ablation patterns through advanced temporal feature extraction.

5.2. Application potential of the gap-free NDSI product

Using the novel MAT-MS model, we generated a daily gap-free NDSI dataset for the Tibetan Plateau spanning from 2003 to 2020. These high-quality NDSI products have significant applications across various fields. One common application is the generation of binary SCA products based on NDSI, which provide clear visualizations of snow extent and its temporal variations. While an NDSI threshold of 0.4 is often used to derive SCA, some studies have suggested that a threshold of 0.1 may be more appropriate for the unique environmental conditions of the Tibetan Plateau (Zhang et al., 2020; Zhang et al., 2019). To evaluate the effectiveness of gap-filling, we compared MODIS SCA based on a threshold of 0.1, both before and after gap-filling, against high-resolution Landsat-derived SCA on the Tibetan Plateau (Fig. 11). The Landsat-derived SCA was obtained using NDSI and snowline extraction based on elevation data (Gascoin et al., 2019). The results showed a strong agreement between the MODIS SCA and the Landsat-derived SCA. Specifically, the accuracy of the MODIS SCA based on original NDSI was 84.54 %, while the MODIS SCA based on our gap-free NDSI achieved a slightly higher accuracy of 86.02 % (Table 1). These results demonstrate that our gap-free NDSI dataset effectively fills data gaps while maintaining comparable accuracy. The SCA based on this high-quality NDSI data can support the derivation of more detailed snow metrics, such as snow lines (Xiao and Liang 2024) and snow phenology (Notarnicola 2020; Xu et al., 2024), enabling more precise analysis of snow cover trends in the context of global warming.

Table 1
Confusion matrices comparing MODIS SCA from original MODIS NDSI products, gap-filled NDSI products, and Landsat-derived SCA based on Landsat-5 image from November 30, 2006 (path/row: 134/37).

Landsat observation	Original MODIS NDSI products			Our gap-free NDSI products		
	Snow	Non-snow	Total	Snow	Non-snow	Total
Snow	60,815 (85.31 %)	10,470 (14.69 %)	71,285	90,574 (87.31 %)	13,170 (12.69 %)	103,744
Non-snow	6026 (17.02 %)	29,389 (82.98 %)	35,415	6383 (17.65 %)	29,780 (82.35 %)	36,163
Overall accuracy	**84.54 %**			**86.02 %**		

Moreover, binary SCA products have some inherent limitations due to the mixed pixel problem, where a single pixel contains both snow and other land surface types. This can affect subsequent applications, leading to unreliable analyses. To address these limitations, the FSC was introduced as a more refined metric derived from continuous NDSI values (Salomonson and Appel 2004). The FSC represents the percentage of snow cover within a pixel, reducing the impact of the mixed pixel problem and providing a more detailed view of snow distribution compared to binary products. As a result, this metric provides more

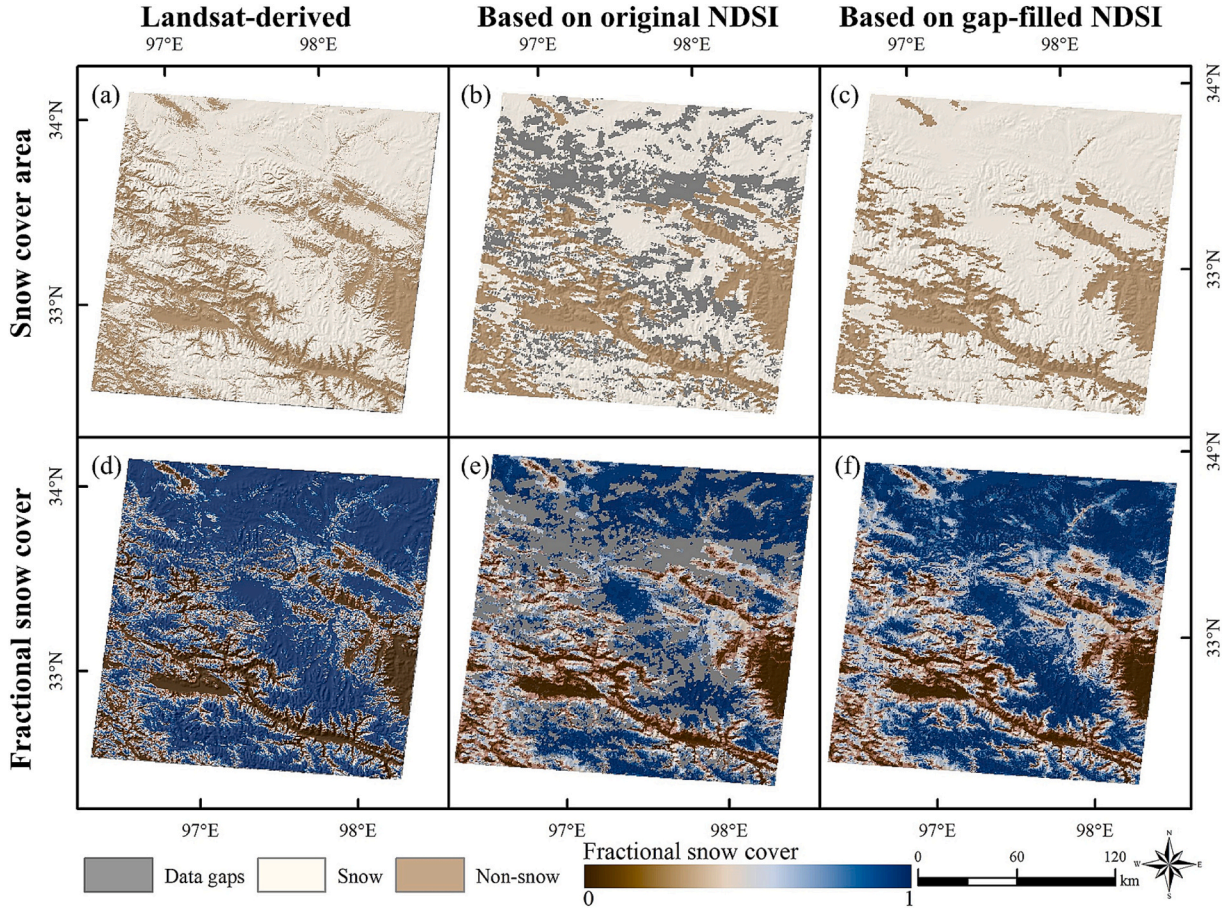

Fig. 11. The Landsat-derived snow cover area (SCA) based on Landsat-5 image (path/row: 134/37) from November 30, 2006 (a), the SCA based on original MODIS NDSI (b), and the SCA based on our gap-free MODIS NDSI (c). The Landsat-derived fractional snow cover (FSC) (d), the simulated FSC from original MODIS NDSI (e), and the simulated FSC from our gap-free MODIS NDSI (f).

accurate inputs for various models and analyses (Xiao et al., 2021). However, the variability in landscape types and environmental conditions across different regions presents significant challenges to the accuracy of globally applied FSC algorithms. When FSC values are derived from NDSI calculations on a global scale, substantial errors often occur in complex regions, leading to discrepancies in snow cover estimates (Huang et al., 2022; Pan et al., 2024). Consequently, the latest MODIS C6 products provide NDSI data rather than FSC. To assess the contribution of gap-free NDSI to FSC calculation, we re-simulated FSC using an FSC equation specifically developed for the Tibetan Plateau (Eq.14, Huang et al., 2022). The results showed that the simulated FSC exhibited strong consistency with the high-resolution Landsat-derived FSC (Fig. 11). This confirms that our gap-free NDSI product enables more detailed and consistent FSC simulations, thus providing more reliable snow data for further analyses and applications.

$$FSC = 1.222 \times NDSI + 0.038 \quad (14)$$

In addition, the temporal dynamics of continuous NDSI curves can be used to extract key phenological metrics, such as the snowmelt onset date (Zheng et al., 2022b). During the snowmelt process, changes in snow grain size and water content primarily affect the spectral properties of snow, particularly in the shortwave infrared band. These changes are effectively captured by NDSI, which is calculated based on the visible (0.545–0.565 μm) and shortwave infrared (1.628–1.652 μm) bands (Roy et al., 2013). In contrast, a reduction in snow depth occurs later in the snowmelt process, as liquid water is generated and begins to flow out. As a result, NDSI is capable of detecting the onset of snowmelt earlier than snow depth measurements, providing an earlier indication of melt timing. To test this, we selected a sample area (Fig. 1a) to explore the relationship between variations in NDSI and snow depth. The snow depth data we used was the 5 km data produced by Yan et al. (2022) based on the SMMR, SSM/I, and SSMI/S downscaling. Fig. 12a illustrates the temporal trends of gap-free NDSI and snow depth throughout a snow season. The result shows that as snow accumulation begins, the NDSI value increases first, followed by an increase in snow depth. Similarly, during the onset of snowmelt, the NDSI value decreases initially, and snow depth gradually declines thereafter. Further analysis reveals a strong correlation between the normalized curves of NDSI and snow depth, with a correlation coefficient of 0.877. This suggests that changes in snowpack are first reflected in the spectral signals (i.e., NDSI), which are subsequently observed in snow depth data derived from passive microwave sensors. Therefore, by combining NDSI curves with snow depth data, it is possible to more accurately determine the onset of snowmelt in a given region.

This finding has practical implications for snow depth studies. Snow depth is crucial but challenging to measure accurately due to the low resolution of passive microwave sensors and the mixed pixel problem in complex topography like the Tibetan Plateau. Existing downscaling methods often include snow cover days as an input factor, which represents optical snow information from SCA products. For instance, Gu et al. (2024) used a LightGBM model with *in situ* snow depth observations, incorporating 9 critical factors (including snow cover days), to downscale AMSR2 data (10 km) and obtain 500 m snow depth estimates. When we added our gap-filled NDSI as an additional factor to this model, the RMSE decreased by 7.5 %, and the R^2 increased by 8.1 % (Fig. 12b). This improvement may be due to NDSI's ability to capture early snowmelt signals through spectral changes in snow grain size and water content—information that snow cover days alone do not provide. Our spatiotemporally continuous NDSI data thus provides a robust dataset for snow depth downscaling studies on the Tibetan Plateau.

6. Conclusions

This study presents a MAT-MS model for filling data gaps in MODIS NDSI products by integrating multi-source data. A mask-aware technique was also integrated into the architecture to mitigate cloud-induced artifact issue. We applied the model to the Tibetan Plateau and demonstrated its effectiveness, achieving an MAE of 1.585, an RMSE of 5.531, and an R^2 of 0.868. Specifically, the model improved RMSE by over 30 % compared to traditional spatiotemporal interpolation methods and by 9 % compared to mainstream deep learning models. The model's performance remained stable even in regions with extensive cloud cover and complex topography, highlighting its robustness and accuracy.

Using this model, we generated a daily gap-free NDSI dataset for the Tibetan Plateau spanning from 2003 to 2020. This dataset enables more detailed and continuous snow detection, providing high-quality input data for research and applications in hydrology, climatology, and ecology. The flexible and generalized MAT-MS model is also applicable to a wide range of continuous remote sensing datasets that suffer from data gaps.

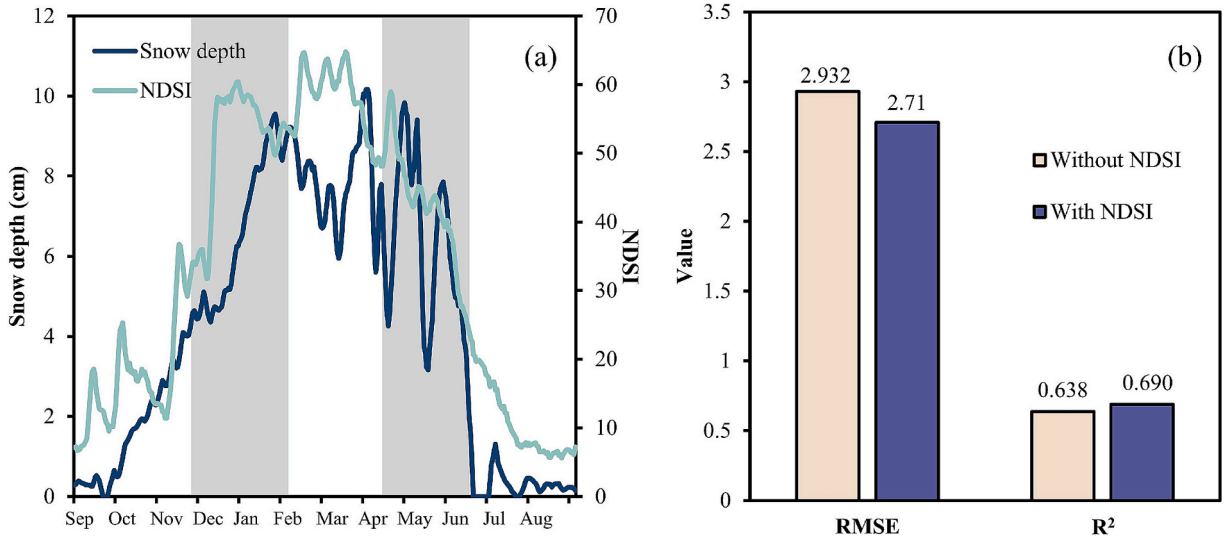

Fig. 12. The temporal trends of gap-free NDSI and snow depth on a sample area within a snow season from September 1, 2017, to August 1, 2018 (a), the accuracy of the LightGBM snow depth downscaled model with and without the NDSI factor (b).

Data availability

The original MODIS NDSI and dem data were accessed via the Google Earth Engine cloud platform (https://earthengine.google.com/). The daily temperature was accessed from the National Tibetan Plateau Data Center (https://cstr.cn/18406.11.Atmos.tpdc.300398) (Yang et al., 2023a). The Python code for the MAT-MS model developed in this study can be available at https://github.com/YanHuang-ECNU/MAT-MS. The daily gap-free NDSI dataset for the Tibetan Plateau from 2003 to 2020, generated using the MAT-MS model can be assessed at https://doi.org/10.11888/Cryos.tpdc.301533.

CRediT authorship contribution statement

Jiahui Xu: Methodology, Conceptualization, Writing – original draft, Software, Formal analysis. **Ruiyang Hua:** Conceptualization, Writing – original draft, Methodology, Formal analysis, Software. **Shujie Wang:** Supervision, Writing – review & editing, Conceptualization. **Stef Lhermitte:** Supervision, Writing – review & editing, Methodology. **Qingyu Gu:** Writing – review & editing, Formal analysis. **Bailang Yu:** Writing – review & editing, Supervision. **Jianping Wu:** Supervision, Writing – review & editing. **Yan Huang:** Writing – original draft, Funding acquisition, Methodology.

Declaration of competing interest

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 42071306 and 42471143).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.isprsjprs.2025.07.004.

References

Bai, K., Li, K., Shao, L., Li, X., Liu, C., Li, Z., Ma, M., Han, D., Sun, Y., Zheng, Z., Li, R., Chang, N.-B., Guo, J., 2024. LGHAP v2: a global gap-free aerosol optical depth and PM2.5 concentration dataset since 2000 derived via big Earth data analytics. Earth Syst. Sci. Data 16, 2425–2448. https://doi.org/10.5194/essd-16-2425-2024.

Bousbaa, M., Boudhar, A., Kinnard, C., Elyoussfi, H., Karaoui, I., Eljabiri, Y., Bouamri, H., Chehbouni, A., 2024. An accurate snow cover product for the Moroccan Atlas Mountains: optimization of the MODIS NDSI index threshold and development of snow fraction estimation models. Int. J. Appl. Earth Obs. Geoinf. 129, 103851. https://doi.org/10.1016/j.jag.2024.103851.

Deng, G., Tang, Z., Dong, C., Shao, D., Wang, X., 2024. Development and evaluation of a cloud-gap-filled MODIS Normalized Difference Snow Index Product over high mountain Asia. Remote Sens. 16, 192. https://doi.org/10.3390/rs16010192.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An image is worth 16x16 words: Transformers for image recognition at scale. In: Proc. Int. Conf. Learn. Represent., Vienna, Austria, https://doi.org/10.48550/arXiv.2010.11929.

Fyfe, J.C., Derksen, C., Mudryk, L., Flato, G.M., Santer, B.D., Swart, N.C., Molotch, N.P., Zhang, X., Wan, H., Arora, V.K., Scinocca, J., Jiao, Y., 2017. Large near-term projected snowpack loss over the western United States. Nat. Commun. 8, 14996. https://doi.org/10.1038/ncomms14996.

Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G., Hagolle, O., 2019. Theia Snow collection: high-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data. Earth Syst. Sci. Data 11, 493–514. https://doi.org/10.5194/essd-11-493-2019.

Gottlieb, A., Mankin, J., 2024. Evidence of human influence on Northern Hemisphere snow loss. Nature 625, 293–300. https://doi.org/10.1038/s41586-023-06794-y.

Gu, Q., Xu, J., Ni, J., Peng, X., Zhou, H., Dong, L., Yu, B., Wu, J., Zheng, Z., Huang, Y., 2024. Improved snow depth estimation on the Tibetan Plateau using AMSR2 and ensemble learning models. Int. J. Appl. Earth Obs. Geoinf. 133, 104102. https://doi.org/10.1016/j.jag.2024.104102.

Hall, D., Riggs, G., Salomonson, V., 1995. Development of methods for mapping global snow cover using Moderate Resolution Imaging Spectroradiometer (MODIS) data. Remote Sens. Environ. 54, 127–140. https://doi.org/10.1016/0034-4257(95)00137-P.

Hall, D., Riggs, G., Salomonson, V., DiGirolamo, N., Bayr, K., 2002. MODIS snow-cover products. Remote Sens. Environ. 83, 181–194. https://doi.org/10.1016/S0034-4257(02)00095-0.

Hao, X., Huang, G., Zheng, Z., Sun, X., Ji, W., Zhao, H., Wang, J., Li, H., Wang, X., 2022. Development and validation of a new MODIS snow-cover-extent product over China. Hydrol. Earth Syst. Sci. 26, 1937–1952. https://doi.org/10.5194/hess-26-1937-2022.

Härer, S., Bernhardt, M., Siebers, M., Schulz, K., 2018. On the need for a time- and location-dependent estimation of the NDSI threshold value for reducing existing uncertainties in snow cover maps at different scales. Cryosph 12, 1629–1642. https://doi.org/10.5194/tc-12-1629-2018.

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., Li, X., 2020. The first high-resolution meteorological forcing dataset for land process studies over China. Sci. Data 7. https://doi.org/10.1038/s41597-020-0369-y.

Hou, J., Huang, C., Zhang, Y., You, Y., 2022. Reconstructing a gap-free MODIS Normalized Difference Snow Index product using a Long Short-Term Memory Network. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/10.1109/tgrs.2022.3178421.

Huang, Y., Xu, J., Xu, J., Zhao, Y., Yu, B., Liu, H., Wang, S., Xu, W., Wu, J., Zheng, Z., 2022. HMRFS-TP: long-term daily gap-free snow cover products over the Tibetan Plateau from 2002 to 2021 based on hidden Markov random field model. Earth Syst. Sci. Data 14, 4445–4462. https://doi.org/10.5194/essd-14-4445-2022.

Jing, Y., Li, X., Shen, H., 2022. STAR NDSI collection: a cloud-free MODIS NDSI dataset (2001–2020) for China. Earth Syst. Sci. Data 14, 3137–3156. https://doi.org/10.5194/essd-14-3137-2022.

Kraaijenbrink, P., Stigter, E., Yao, T., Immerzeel, W., 2021. Climate change decisive for Asia's snow meltwater supply. Nat. Clim. Change 11, 591–597. https://doi.org/10.1038/s41558-021-01074-x.

Li, M., Zhu, X., Li, N., Pan, Y., 2020. Gap-filling of a MODIS Normalized Difference Snow Index product based on the similar pixel selecting algorithm: a case study on the Qinghai–Tibetan Plateau. Remote Sens. 12. https://doi.org/10.3390/rs12071077.

Li, W., Lin, Z., Zhou, K., Qi, L., Wang, Y., Jia, J., 2022. MAT: mask-aware transformer for Large Hole Image Inpainting. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., New Orleans, LA, USA, pp. 10748-10758, https://doi.org/10.1109/Cvpr52688.2022.01049.

Li, X., Jing, Y., Shen, H., Zhang, L., 2019. The recent developments in cloud removal approaches of MODIS snow cover product. Hydrol. Earth Syst. Sci. 23, 2401–2416. https://doi.org/10.5194/hess-23-2401-2019.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. IEEE Int. Conf. Comput. Vis., Seoul, Korea (South), pp. 10012-10022, https://doi.org/10.48550/arXiv.2103.14030.

Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L., Wei, F., Guo, B., 2022. Swin transformer v2: scaling up capacity and resolution. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., New Orleans, LA, USA, pp. 11999-12009, https://doi.org/10.1109/Cvpr52688.2022.01170.

Luo, J., Dong, C., Lin, K., Chen, X., Zhao, L., Menzel, L., 2022. Mapping snow cover in forests using optical remote sensing, machine learning and time-lapse photography. Remote Sens. Environ. 275. https://doi.org/10.1016/j.rse.2022.113017.

Ma, Q., Keyimu, M., Li, X., Wu, S., Zeng, F., Lin, L., 2023. Climate and elevation control snow depth and snow phenology on the Tibetan Plateau. J. Hydrol. 617. https://doi.org/10.1016/j.jhydrol.2022.128938.

Motamed, S., Xu, J., Wu, C., Häne, C., Bazin, J., De La Torre, F., 2023. PATMAT: person aware tuning of mask-aware transformer for face inpainting. In: Proc. IEEE Int. Conf. Comput. Vis., Paris, France, pp. 22721-22730, https://doi.org/10.1109/Iccv51070.2023.02082.

Muhammad, S., Thapa, A., 2021. Daily Terra–Aqua MODIS cloud-free snow and Randolph Glacier Inventory 6.0 combined product (M*D10A1GL06) for high-mountain Asia between 2002 and 2019. Earth Syst. Sci. Data 13, 767–776. https://doi.org/10.5194/essd-13-767-2021.

Musselman, K., Addor, N., Vano, J., Molotch, N., 2021. Winter melt trends portend widespread declines in snow water resources. Nat. Clim. Change 11, 418–424. https://doi.org/10.1038/s41558-021-01014-9.

Notarnicola, C., 2020. Hotspots of snow cover changes in global mountain regions over 2000–2018. Remote Sens. Environ. 243. https://doi.org/10.1016/j.rse.2020.111781.

Pan, F., Jiang, L., Wang, G., Pan, J., Huang, J., Zhang, C., Cui, H., Yang, J., Zheng, Z., Wu, S., Shi, J., 2024. MODIS daily cloud-gap-filled fractional snow cover dataset of the Asian Water Tower region (2000–2022). Earth Syst. Sci. Data 16, 2501–2523. https://doi.org/10.5194/essd-16-2501-2024.

Pepin, N., Bradley, R., Diaz, H., Baraer, M., Caceres, E., Forsythe, N., Fowler, H., Greenwood, G., Hashmi, M., Liu, X., Miller, J., Ning, L., Ohmura, A., Palazzi, E., Rangwala, I., Schoner, W., Severskiy, I., Shahgedanova, M., Wang, M., Williamson, S., Yang, D., 2015. Elevation-dependent warming in mountain regions of the world. Nat. Clim. Change 5, 424–430. https://doi.org/10.1038/nclimate2563.

Pulliainen, J., Luojus, K., Derksen, C., Mudryk, L., Lemmetyinen, J., Salminen, M., Ikonen, J., Takala, M., Cohen, J., Smolander, T., Norberg, J., 2020. Patterns and trends of Northern Hemisphere snow mass from 1980 to 2018. Nature 581, 294–298. https://doi.org/10.1038/s41586-020-2258-0.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: unified, real-time object detection. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Las Vegas, NV, USA, pp. 779-788, https://doi.org/10.48550/arXiv.1506.02640.

Riggs, G., Hall, D., Román, M., 2017. Overview of NASA's MODIS and Visible Infrared Imaging Radiometer Suite (VIIRS) snow-cover earth system data records. Earth Syst. Sci. Data 9, 765–777. https://doi.org/10.5194/essd-9-765-2017.

Riggs, G., Hall, D., Román, M., 2019. MODIS snow products Collection 6.1 user guide https://modis-snow-ice.gsfc.nasa.gov/uploads/snow_user_guide_C6.1_final_revised_april.pdf.

Roy, A., Picard, G., Royer, A., Montpetit, B., Dupont, F., Langlois, A., Derksen, C., Champollion, N., 2013. Brightness temperature simulations of the canadian seasonal snowpack driven by measurements of the snow specific surface area. IEEE Trans. Geosci. Remote Sens. 51, 4692–4704. https://doi.org/10.1109/tgrs.2012.2235842.

Salomonson, V., Appel, I., 2004. Estimating fractional snow cover from MODIS using the normalized difference snow index. Remote Sens. Environ. 89, 351–360. https://doi.org/10.1016/j.rse.2003.10.016.

Salomonson, V., Appel, I., 2006. Development of the Aqua MODIS NDSI fractional snow cover algorithm and validation results. IEEE Trans. Geosci. Remote Sens. 44, 1747–1756. https://doi.org/10.1109/Tgrs.2006.876029.

Shen, M., Wang, S., Jiang, N., Sun, J., Cao, R., Ling, X., Fang, B., Zhang, L., Zhang, L., Xu, X., Lv, W., Li, B., Sun, Q., Meng, F., Jiang, Y., Dorji, T., Fu, Y., Iler, A., Vitasse, Y., Steltzer, H., Ji, Z., Zhao, W., Piao, S., Fu, B., 2022. Plant phenology changes and drivers on the Qinghai–Tibetan Plateau. Nat. Rev. Earth Env. 3, 633–651. https://doi.org/10.1038/s43017-022-00317-5.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. In: Proc. Adv. Neural Inf. Process. Syst., Long Beach, CA, USA, https://doi.org/10.48550/arXiv.1706.03762.

Wang, Q., Tang, Y., Tong, X., Atkinson, P., 2024. Filling gaps in cloudy Landsat LST product by spatial-temporal fusion of multi-scale data. Remote Sens. Environ. 306. https://doi.org/10.1016/j.rse.2024.114142.

Wu, Y., Gao, J., Zhao, A., 2024. Cloud properties and dynamics over the Tibetan Plateau – a review. Earth Sci. Rev. 248. https://doi.org/10.1016/j.earscirev.2023.104633.

Xiao, X., Liang, S., 2024. Assessment of snow cover mapping algorithms from Landsat surface reflectance data and application to automated snowline delineation. Remote Sens. Environ. 307. https://doi.org/10.1016/j.rse.2024.114163.

Xiao, X., Liang, S., He, T., Wu, D., Pei, C., Gong, J., 2021. Estimating fractional snow cover from passive microwave brightness temperature data using MODIS snow cover product over North America. Cryosph 15, 835–861. https://doi.org/10.5194/tc-15-835-2021.

Xing, D., Hou, J., Huang, C., Zhang, W., 2022. Spatiotemporal reconstruction of MODIS Normalized Difference Snow Index products using U-Net with partial convolutions. Remote Sens. 14. https://doi.org/10.3390/rs14081795.

Xu, J., Tang, Y., Dong, L., Wang, S., Yu, B., Wu, J., Zheng, Z., Huang, Y., 2024. Temperature-dominated spatiotemporal variability in snow phenology on the Tibetan Plateau from 2002 to 2022. Cryosph 18, 1817–1834. https://doi.org/10.5194/tc-18-1817-2024.

Yan, D., Ma, N., Zhang, Y., 2022. Development of a fine-resolution snow depth product based on the snow cover probability for the Tibetan Plateau: validation and spatial–temporal analyses. J. Hydrol. 604. https://doi.org/10.1016/j.jhydrol.2021.127027.

Yang, K., Jiang, Y., Tang, W., He, J., Shao, C., Zhou, X., Lu, H., Chen, Y., Li, X., Shi, J., 2023a. A high-resolution near-surface meteorological forcing dataset for the Third Pole region (TPMFD, 1979-2022). In, National Tibetan Plateau/Third Pole Environment Data Center, https://doi.org/10.11888/Atmos.tpdc.300398.

Yang, S., Chen, X., Liao, J., 2023b. Uni-paint: a unified framework for multimodal image inpainting with pretrained diffusion model. In: Proc. Int. Conf. Learn. Represent., Kigali, Rwanda, https://doi.org/10.1145/3581783.3612200.

Yao, T., Bolch, T., Chen, D., Gao, J., Immerzeel, W., Piao, S., Su, F., Thompson, L., Wada, Y., Wang, L., Wang, T., Wu, G., Xu, B., Yang, W., Zhang, G., Zhao, P., 2022. The imbalance of the asian water tower. Nat. Rev. Earth Env. 3, 618–632. https://doi.org/10.1038/s43017-022-00299-4.

You, Q., Wu, T., Shen, L., Pepin, N., Zhang, L., Jiang, Z., Wu, Z., Kang, S., AghaKouchak, A., 2020. Review of snow cover variation over the Tibetan Plateau and its influence on the broad climate system. Earth Sci. Rev. 201. https://doi.org/10.1016/j.earscirev.2019.103043.

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T., 2018. Generative image inpainting with contextual attention. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Salt Lake City, UT, USA, pp. 5505-5514, https://doi.org/10.1109/Cvpr.2018.00577.

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T., 2019. Free-form image inpainting with gated convolution. In: Proc. IEEE Int. Conf. Comput. vis., Seoul, Korea (south) 4470–4479. https://doi.org/10.1109/Iccv.2019.00457.

Zhang, H., Zhang, F., Che, T., Wang, S., 2020. Comparative evaluation of VIIRS daily snow cover product with MODIS for snow detection in China based on ground observations. Sci. Total Environ. 724. https://doi.org/10.1016/j.scitotenv.2020.138156.

Zhang, H., Zhang, F., Zhang, G., Che, T., Yan, W., Ye, M., Ma, N., 2019. Ground-based evaluation of MODIS snow cover product V6 across China: Implications for the selection of NDSI threshold. Sci. Total Environ. 651, 2712–2726. https://doi.org/10.1016/j.scitotenv.2018.10.128.

Zhang, L., Zhou, Y., Barnes, C., Amirghodsi, S., Lin, Z., Shechtman, E., Shi, J., 2022. Perceptual artifacts localization for inpainting. In: Proc. Eur. Conf. Comput. Vis., Tel Aviv, Israel, pp. 146-164, https://doi.org/10.1007/978-3-031-19818-2_9.

Zhang, Y., Hong, S., Liu, D., Piao, S., 2023. Susceptibility of vegetation low-growth to climate extremes on Tibetan Plateau. Agr. Forest Meteorol. 331. https://doi.org/10.1016/j.agrformet.2023.109323.

Zhao, D., Heidler, K., Asgarimehr, M., Arnold, C., Xiao, T., Wickert, J., Zhu, X.X., Mou, L., 2023. DDM-Former: Transformer networks for GNSS reflectometry global ocean wind speed estimation. Remote Sens. Environ. 294, 113629. https://doi.org/10.1016/j.rse.2023.113629.

Zhao, Q., Hao, X., Wang, J., Luo, S., Shao, D., Li, H., Feng, T., Zhao, H., 2022. Snow cover phenology change and response to climate in China during 2000–2020. Remote Sens. 14. https://doi.org/10.3390/rs14163936.

Zheng, H., Lin, Z., Lu, J., Cohen, S., Shechtman, E., Barnes, C., Zhang, J., Xu, N., Amirghodsi, S., Luo, J., 2022a. Image inpainting with cascaded modulation GAN and object-aware training. In: Proc. Eur. Conf. Comput. Vis., Tel Aviv, Israel, pp. 277-296, https://doi.org/10.1007/978-3-031-19787-1_16.

Zheng, J., Jia, G., Xu, X., 2022b. Earlier snowmelt predominates advanced spring vegetation greenup in Alaska. Agr. Forest Meteorol. 315. https://doi.org/10.1016/j.agrformet.2022.108828.

Zhou, W., Persello, C., Li, M., Stein, A., 2023. Building use and mixed-use classification with a transformer-based network fusing satellite images and geospatial textual information. Remote Sens. Environ. 297, 113767. https://doi.org/10.1016/j.rse.2023.113767.

Zhu, X., Li, J., Liu, Q., Yu, W., Li, S., Zhao, J., Dong, Y., Zhang, Z., Zhang, H., Lin, S., 2022. Use of a BP neural network and meteorological data for generating spatiotemporally continuous LAI time series. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/10.1109/tgrs.2021.3095535.