Multisource Topographic-Enhanced Cloud Removal for Remote Sensing in Mountainous Landscapes

Meryeme Boumahdi, Consuelo Gonzalo-Martin, Mario Lillo-Saavedra, Marcelo Somos-Valenzuela, and Angel García-Pedrero

***Abstract*—In mountainous landscapes, integrating topographical information is crucial for effective analysis and understanding. Remote sensing becomes indispensable in studying mountains, enabling the monitoring of critical aspects such as grassland degradation, snow depth, glacier dynamics. The intricate nature of mountain ecosystems requires a thorough understanding of their dynamics, given their vital role in providing habitats for unique species, influencing hydrology patterns, and indicating the impact of climate change. However, the persistent challenge of cloud cover obstructs surface observations, limiting the effectiveness of optical sensors. To overcome this obstacle, various cloud removal techniques have been developed, although they tend to struggle in such landscapes. In this study, we introduce CRT-UNet, a UNet-based cloud removal model that incorporates topographical information from digital elevation model (DEM) data as input to enhance performance in mountainous regions. By integrating synthetic aperture radar Sentinel-1 (S1) data, DEM information, and topographic insights into Sentinel-2 (S2) data, our model aims to enhance cloud removal capabilities, particularly in challenging terrains characterized by thick cloud coverage and significant elevation variations. This integration of topographical information enriches the cloud removal process, enabling more accurate restoration of obscured terrain features. The results demonstrate the superior performance of CRT-UNet in cloud removal across varying cloud cover levels and complex terrain features, such as rugged peaks and deep valleys. The proposed model outperforms state-of-the-art cloud removal models quantitatively and qualitatively. This underscores the importance of incorporating topographical information in remote sensing applications, particularly in mountainous regions, to improve data accuracy.**

Received 12 November 2024; revised 27 March 2025; accepted 17 May 2025. Date of publication 22 May 2025; date of current version 6 June 2025. The work of Meryeme Boumahdi was supported in part by GMV and in part by the Fundación Mujeres por Africa under the Learn Africa program. This work was supported in part by the ESA under Grant 4000126706/19/I-NB through the Fire CCi+ project, in part by the Chilean Science Council (ANID) through the Anillo under Grant ACT210080, and in part by the Water Research Center for Agriculture and Mining, CRHIAM under Grant ANID/FONDAP/1523A0001. *(Corresponding author: Angel García-Pedrero.)*

Meryeme Boumahdi, Consuelo Gonzalo-Martin, and Angel García-Pedrero are with Center for Biomedical Technology, Universidad Politécnica de Madrid, 28223 Pozuelo de Alarcón, Spain, and also with the Department of Computer Architecture and Technology, School of Computer Engineering, Universidad Politécnica de Madrid, 28660 Boadilla del Monte, Spain (e-mail: m.boumahdi@alumnos.upm.es; consuelo.gonzalo@upm.es; angelmario.garcia@upm.es).

Mario Lillo-Saavedra is with the Facultad de Ingeniería Agrícola, Universidad de Concepción, Chillán 3812120, Chile (e-mail: malillo@udec.cl).

Marcelo Somos-Valenzuela is with the Department of Forest Sciences, Faculty of Agriculture and Environmental Sciences, University of La Frontera, Temuco 4811230, Chile (e-mail: marcelo.somos@ufrontera.cl).

This article has supplementary downloadable material available at https://doi.org/10.1109/JSTARS.2025.3572379, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2025.3572379

***Index Terms*—Cloud removal, mountainous area, Sentinel-1 (S1), Sentinel-2 (S2), topographic data, UNet.**

I. Introduction

REMOTE sensing images are crucial for Earth monitoring, providing valuable insights into land use patterns, atmospheric conditions, forest cover, and various other aspects of the Earth's surface [1], [2]. Open data policies provide access to large quantities of satellite images at no cost, fostering a cost-effective approach to data collection and analysis that broadens its applications [3].

Mountainous regions present distinct challenges and opportunities for environmental analysis and understanding due to their rugged terrain and diverse ecosystems. Remote sensing technology emerges as a powerful tool for studying mountains, providing insight into critical aspects such as vegetation health, snow accumulation, glacier movement, and natural hazards. It also facilitates the detection of underground targets, such as pipelines, within mountains, thereby aiding urban management and the development of underground spaces [4]. Additionally, it is vital to monitor the depth of snow in mountainous regions, which is essential for hydrology, water resource management, and climate research [5]. Furthermore, it contributes to the assessment of vegetation parameters in mountain grasslands, allowing the evaluation of ecosystem stability, overgrazing risks, and conservation strategies for endemic species [6]. Moreover, it assists in studying rainfall patterns in mountainous areas, which are crucial to climate, environment, and water resource management, particularly in regions prone to disasters like hurricanes [7]. Overall, remote sensing plays a pivotal role in understanding and managing various aspects of mountain environments for sustainable development and conservation efforts.

However, the presence of persistent clouds poses challenges to optical sensors employed in mountain monitoring, as they are sensitive to cloud and atmospheric conditions. Consequently, the effectiveness of optical sensors is compromised, hindering the processing and analysis of data for tasks such as target monitoring. These limitations stem from the obstructive nature of clouds, which block direct visibility of the Earth's surface for space-based optical sensors. This obstruction not only affects the clarity of captured images, but also results in gaps in data collection, diminishing the temporal resolution of monitoring systems.

For this reason, cloud removal methods play an important role as an indispensable preprocessing step for continuous remote

sensing data applications [8]. These methods can significantly improve the quality of images and the temporal frequency of clear images by removing clouds and haze from optical images.

Cloud removal can be considered as a special type of technique to reconstruct missing information from remote sensing data affected by cloud cover. This process involves filling the contaminated areas with new and relevant content. Traditional methods for cloud removal can be categorized into four groups: spatial-based, spectral-based, temporal-based, and hybrid approaches. Spatial-based methods assume that both clear and cloudy pixels share similar statistical and geometric structures. For example, Zhang et al. [9] proposed an interpolation method to fill gaps in Landsat ETM+ images using an ordinary kriging technique. Maalouf et al. [10] used a tool called "bandelet" to accurately determine the propagation directions. The values obtained were then propagated into the pixels contaminated with clouds, contributing to an effective cloud removal method. However, spatial-based methods face limitations in reconstructing large areas because of the absence of prior information. In contrast, spectral-based methods extract additional information from the spectral domain to reconstruct missing data in specific bands. To remove thin clouds in cloudy images, Hu et al. [11] employed a dual tree complex wavelet transform and a transfer least squares support vector. Meanwhile, Xu et al. [12] employed signal transmission and spectral mixture analysis for pixel correction. However, these methods are limited in thick cloudy conditions, as all spectral bands are affected to varying degrees, confining their effectiveness primarily to haze and thin clouds. To address this limitation, temporal-based approaches have been explored, in which information from the same location at different dates is integrated to reconstruct missing data. Lin et al. [13] introduced a patch-based method leveraging temporal correlation among multitemporal images with similar temporal variations to select the optimal patch for replacement. Similarly, Tseng et al. [14] replaced cloudy areas by using cloud-free zones from other dates, adjusting the pixel values based on color statistics. A multitemporal dictionary learning method was also used in [15]. However, these methods struggle when there is a rapid change between the dates used for substitution and the reference dates. Hybrid approaches aim to leverage the strengths of existing methods to address their respective limitations. In the study presented in [16], spatial and temporal correlations were exploited to replace contaminated pixels. This was achieved by identifying similar pixels within a window using a multitemporal image as a guide for locating similar pixels. Another hybrid method, as discussed in [17], involved the development of a contextual spatio-spectral postprediction system. This system integrated spatial and spectral information to reconstruct the cloudy pixels.

In contrast to traditional approaches for cloud removal, deep learning has demonstrated its efficacy by leveraging advanced neural network architectures [18]. Traditional methods often rely on rule-based algorithms or image processing techniques that may struggle to adapt to the diverse and complex patterns associated with clouds in various scenarios. On the other hand, deep learning techniques utilize intricate and nonlinear changes, employing numerous layers of neural networks. This nonlinearity allows deep learning models to capture intricate relationships and patterns in images that may be difficult for linear methods to discern. One approach is to consider only surface reflections and cloud top reflections as a linear mixture of image elements and use a two-step convolutional neural network (CNN) to extract transparency information from clouds and recover surface information [19]. Another approach involves integrating spatial detail in a reference image and a coarse spectral pattern in a near-target-date image to generate missing data with land cover change information, using a spatiotemporal attention network (STAN) that adopts the self-attention mechanism, residual learning and high-pass features [20]. Sarukkai et al. [21] have framed cloud removal as a conditional image synthesis challenge, presenting a network capable of removing clouds from a single optical image or a triplet of optical images. Zhang et al. [22] proposed a method utilizing a spatial–temporal-spectral CNN to fill data gaps in Landsat TM data. To address cloud interference, the fusion of multiple sensors is considered, incorporating synthetic aperture radar (SAR) for its cloud-penetrating advantage and measureing the backscatter. Meraner et al. [23] introduced a ResNet model called DSen-CR to reconstruct cloudy optical images using information from a single SAR. They also presented the "SEN12MS-CR" data set, which includes global distributed satellite observations with multitemporal and multimodal characteristics. Generative adversarial networks (GANs) have been employed to translate SAR images into optical images, allowing the subsequent removal of clouds using the translated images [24]. In addition, a CycleGAN that integrates saliency enhancement and high-level feature enhancement modules has been developed for cloud removal [25]. Xu et al. [26] presented a GAN-based model with attention mechanism called AMGAN to capture the distribution of cloud thickness and that take advantage of multispectral information to remove clouds. Multimodal and multitemporal models were also used to improved cloud removal performance [27]. For instance, Sebastianelli et al. [28] introduced a multimodal deep learning model that combines temporal-sequence blending and SAR-to-optical translation to restore optical scenes across diverse atmospheric conditions.

Despite advancements in cloud removal techniques aimed at mitigating the challenge of cloud cover, their effectiveness in mountainous regions remains constrained [29]. The intricate relationship between elevation variations and dense cloud cover presents a significant obstacle for traditional methods. The effectiveness of cloud removal in satellite imagery of mountainous regions significantly affects the accuracy of streamflow estimation and other hydrological parameters. Consequently, there is a need for innovative approaches that can handle the unique characteristics of mountainous terrain. However, effectively mitigating cloud contamination in these regions remains challenging due to factors such as fragmented cloud cover and highly variable topographic conditions. While some existing cloud removal models incorporate digital elevation model (DEM) data, they often overlook other critical topographic characteristics, such as slope and aspect, which significantly influence environmental processes [30]. Integrating these factors

into cloud removal methodologies could enhance the precision of remote sensing analyses, ultimately improving hydrological modeling and resource management in mountainous landscapes. Advanced cloud removal models tailored to these regions could facilitate the acquisition of more comprehensive time series data, including snow cover, soil moisture, and evapotranspiration variables, which are critical inputs for hydrological models [31]. Furthermore, these techniques would allow for more continuous and precise monitoring of glaciers and glacial lakes, which are vital for effective hydrological risk management in mountainous terrains [32]. Cloud removal models hold considerable promise for enhancing our ability to model and manage water resources in complex mountain ecosystems. However, improving the accuracy of cloud removal in these regions requires the development of models that account for specific challenges, such as fragmented cloud cover and varying elevations. Considering the data from the DEM, the slope and aspect in mountainous regions are crucial because of their significant impacts on various aspects. These insights provide valuable information on the topographic features of mountainous regions, helping to assess the ruggedness of the terrain, identify landforms and types of land cover, analyze hydrological processes, and predict natural hazards such as landslides or avalanches. Consequently, the integration of topographic information plays a pivotal role in understanding the complex dynamics of these landscapes.

In response to this necessity, we introduce CRT-UNet, a UNet-based cloud removal model specifically tailored for mountainous landscapes. At the core of our approach lies the integration of topographical information derived from DEM data. By leveraging SAR Sentinel-1 (S1) data, DEM insights, and Sentinel-2 (S2) imagery, our model aims to enhance cloud removal capabilities, particularly in regions characterized by rugged terrain and frequent cloud cover. Experimental findings highlight the significant efficacy of the proposed model in reconstructing S2 images obscured by thick clouds. The inclusion of topographic information proves indispensable in enabling the model to comprehend the image structure, thereby enhancing its ability to discern intricate details. In particular, the model exhibits remarkable proficiency in effectively distinguishing between clouds and snow cover while accurately identifying cloud shadows. Even in scenarios of complete cloud cover with zero visibility, the model demonstrates resilience by reconstructing spatial details that encapsulate the location's topographical features, thereby showcasing its robustness and efficacy in remote sensing applications.

The rest of this article is organized as follows. Section II details the data and the generation of the dataset. Section III provides an explanation of the proposed methodology, encompassing the designed neural network architecture. The experimental set-up is presented in Section IV and the obtained results are presented in Section V, with further discussion in Section VI. Finally, Section VII concludes this article.

II. Data and Study Area

A. Data Description

1) Sentinel-1: Sentinel-1 (S1) synthetic aperture radar ground range detected (SAR GRD) data is a crucial asset in Earth observation and remote sensing as a part of the European Space Agency (ESA)'s Copernicus program. Comprising two satellites, Sentinel-1A launched on 3 April 2014, followed by Sentinel-1B on 25 April 2016, both are sharing the same orbital plane. Equipped with a C-band SAR sensor operating at 5.4 GHz, it provides valuable information with a ground range spatial resolution of 10 m $\times$ 10 m [33]. GRD data are indispensable for various applications, including monitoring land deformation, tracking vegetation changes, detecting surface water, and assessing disaster impacts [34]. S1 SAR GRD data has become a cornerstone in environmental monitoring, scientific research, and disaster management, significantly contributing to understanding the dynamic processes of the Earth [35]. Its ability to acquire data regardless of weather conditions or daylight makes it particularly valuable.

2) Sentinel-2: The ESA, within the Copernicus program, has developed the Sentinel-2 (S2) satellite imaging mission, characterized by its wide swath and fine spatial resolution. Launched on 23 June 2015, Sentinel-2A was followed by the launch of Sentinel-2B on 7 March 2017, with the aim of improving the temporal resolution from 10 to 5 days [36], [37]. Positioned in the same orbit, the twin satellites maintain a phased arrangement of 180°. Equipped with multispectral imaging instruments (MSI), S2 satellites are capable of capturing imagery in 13 wide-swath bands. This includes four bands at 10 m, six bands at 20 m, and three bands at 60 m resolution [38]. The synchronized operation of the twin satellites ensures comprehensive and frequent observation of the Earth's surface. Covering regions from $-56°$ to 84° latitude, the S2 satellites contribute mainly data for global land services. These services include monitoring vegetation, soil, and water cover, as well as inland waterways and coastal areas.

3) Digital Elevation Model: The Shuttle Radar Topographic Mission (SRTM) was launched in 2000 by NASA and the National Geospatial-Intelligence Agency (NGA) with the aim of collecting global elevation data. The mission generates a global high-resolution DEM with a coverage of 80% of Earth's land surface and a spatial resolution of 1-arc second (30 m). The SRTM elevation data are derived from *X*-bands and *C*-band interferometric synthetic aperture radar (InSAR) sensor (5.6 cm wavelength and 5.3 GHz frequency). Serving as a valuable resource for scientists, geographers, and researchers, the SRTM DEM provides a foundation for a multitude of applications [39]. It aids in geological studies, environmental monitoring, and urban planning, providing crucial information on landforms and surface features.

4) Topographic Information: Among the relevant data extracted from DEMs are slope, aspect, and hillshade.

Slope: It represents the steepness of the terrain and is typically measured in degrees or as a percentage. This information is essential for assessing terrain stability, potential erosion, and suitability for various activities such as agriculture, infrastructure development, or recreation. In mountainous areas, slope gradients can vary significantly, impacting land use, accessibility, and ecosystem dynamics [40].

Aspect: It refers to the direction a slope faces, measured in degrees from 0° to 360°, where 0° represents north, 90° east, 180° south, and 270° west. Understanding the aspect is crucial

for grasping how solar radiation interacts with the terrain. In mountainous regions, the aspect influences temperature, moisture distribution, vegetation patterns, and even snow accumulation and melt rates [41].

Hillshade: It is a visualization technique employed to produce shaded relief maps, mimicking the effect of sunlight on the terrain's surface. This technique enhances the perception of terrain features by incorporating shadows and highlights based on the slope and aspect of the surface. Hillshade maps offer a clear depiction of landforms, including ridges, valleys, and slopes, facilitating visual interpretation and terrain analysis [42].

B. Dataset

1) Preprocessing: The data preparation workflow involved the acquisition, preprocessing and the preparation of S2 and S1 satellite imagery. S2 1C level data were downloaded from PEPS platform[1] (Plateforme d'Exploitation des Produits Sentinel) using EODAG library.[2] The atmospheric correction of S2 images was carried out using the MACCS-ATCOR joint algorithm (MAJA) algorithm [43]. MAJA proved instrumental in mitigating atmospheric effects, including aerosols, water vapor, and ozone, resulting in precise surface reflectance values. It effectively addresses adjacency and topographic effects, which are critical for accurate surface reflectance measurements. MAJA was also used to generate an extensive cloud mask that includes not only thin and thick clouds but also shadows, ensuring the use of high-quality images. S1 images were procured utilizing the S1Tiling tool,[3] which seamlessly automated both the download and processing stages using ORFEO ToolBox. This tool facilitated orthorectification and calibration in sigma nought for both the vertical–vertical(VV) and vertical–horizontal (VH) SAR image channels. Moreover, the tool performed the crucial task of reprojecting and resampling the final S1 images onto an identical coordinate reference system and pixel grid as the S2 images, ensuring spatial coherence at a 10-m physical spacing resolution. Furthermore, the DEM was downloaded and resampled to a spatial resolution of 20 m, with topographic information such as aspect, slope, and hillshade subsequently extracted from it. This process not only enriched the dataset but also provided essential terrain attributes for such a landscape. Fig. 1 summarizes the preprocessing used to generate the dataset.

2) Study Area: The study area is positioned in northern Chilean Patagonia, located in the south of Chile, and encompasses seven S2 tiles covering from 41.5°S to 45.5°S and -71°W to -74°W (see Fig. 2). The region comprises several inhabited towns and critical infrastructure, including the "Carretera Austral" (Route 7), a vital north–south thoroughfare that traverses the primary valley and connects Chilean Patagonia with central Chile [44]. The contemporary climate in Patagonia is mainly shaped by persistent westerly winds, commonly referred to as westerlies, which drive storms against the western slopes of the southern Andes, resulting in local precipitation rates exceeding

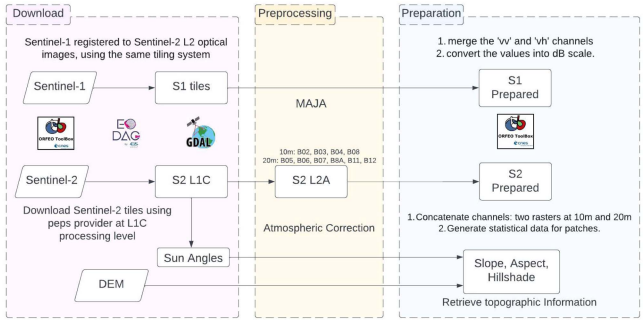


Fig. 1. Workflow for preprocessing and preparing S1 and S2 data. This workflow involves atmospheric correction of S2 images, resampling from 20 to 10 m resolution, and merging of bands postcorrection. It also includes topographic information extraction from the DEM, such as slope, aspect, and hillshade. For S1 data, the workflow includes tiling to align with the S2 grid system and dB normalization of the VV and VH polarization bands.

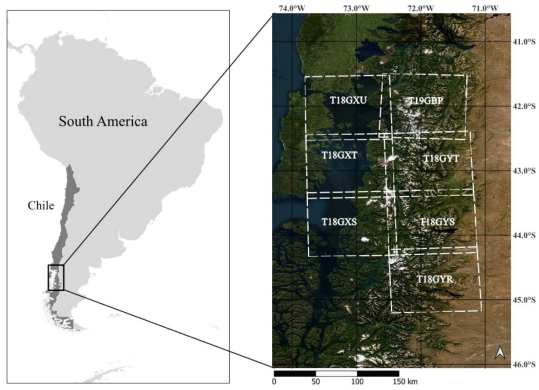


Fig. 2. Study area covers seven S2 tiles in northern Chilean Patagonia.

4000 mm/year. Mean annual temperatures in the region range from 14 °C in the northeast to 10 °C in the southwest, with high-elevation areas of the Andes experiencing minimum temperatures of approximately 6 °C. The presence of the Andes Mountain range and its proximity to the Southern Ocean create atmospheric conditions favorable for the formation of cloud masses, leading to a significant increase in the frequency of both precipitating and nonprecipitating clouds.

3) Dataset Generation: S2A and S2B are twin satellites that are part of the Copernicus program. These satellites share similar objectives and are equipped with almost identical instruments. However, there are slight discrepancies in their spectral response functions (SRF), which give rise to observable differences in channel reflectance [45]. This poses a challenge when it comes to ensuring consistency in observations and retrieved variables between the two satellites. Additionally, while S2A and S2B have similar orbits, they are not perfectly synchronized. Consequently, there is a time discrepancy in the capture of images of the same geographical area. This time offset results in variations in lighting conditions and shadow positions. The position of the sun in the sky becomes an influential factor in determining the shadows cast by the terrain. Therefore, images taken at different times of the day may exhibit disparities in the lengths,

[1]https://peps.cnes.fr/ (last accessed November 2024)
[2]https://eodag.readthedocs.io/en/stable/ (last accessed March 2025)
[3]https://gitlab.orfeo-toolbox.org/s1-tiling/s1tiling (last accessed March 2025)

orientations, and overall appearance of shadows, particularly in regions with hills and slopes.

Because of the aforementioned differences in SRFs and the temporal misalignment between S2A and S2B, we exclusively employed S2A data alongside the ascending S1. This deliberate strategy is aimed at reducing potential errors and inconsistencies that stem from spectral and temporal disparities between the two satellites, thus enhancing the reliability and accuracy and precision of our model.

Preprocessed S2 bands at 20 m were resampled to 10 m using the nearest-neighbor technique. Two supplementary output rasters are computed, one that incorporates the cloud mask generated via the MAJA atmospheric correction algorithm and another depicting the pixel validity information. The cloud mask and pixel validity data are then used to identify cloud-free and cloudy pixels for image generation.

The preparation pipeline for S1 mirrors certain aspects of the S2 process and only ascending S1 was considered in the generation of the dataset. Initially, the VV and VH channels are merged, resulting in a single raster. Subsequently, the raw intensity values are transformed into decibel (dB) scale. Additionally, an output raster is generated to compute the count of valid and nonvalid pixels within the dataset.

The topographic data were derived from the DEM, taking into account the sun's position (zenith and azimuth angles) for generating the hillshade. Rather than relying on default values, the sun angles were obtained from the S2 metadata, ensuring accurate representation of shadows specific to the location. The construction of the data set involves the selection of S2 cloudy, S2 clear (target), and S1 tuples.

The selection parameters are carefully chosen to ensure dataset integrity, with a maximum 72-hour gap between the acquisition dates of the S1 and S2 images and a maximum 7-day temporal gap between cloud-free and cloudy S2 images [30]. The rationale behind this strategy is predicated on the assumption that, given the nature of the study location, negligible changes occur within this short timeframe, facilitating the consideration of S2 and S1 images as if they were captured on the same day. This approach not only augments the dataset's patch count but also guarantees the inclusion of real cloud images in the training set, a crucial consideration for accurate cloud and shadow identification, particularly in mountainous regions. Fig. 3 shows examples of patch triplets from the dataset including cloudy S2, cloud free, S1 images as well as DEM and topographic data.

Following the extraction of S2 cloudy, S2 cloud-free, and S1 images, patch selection for the dataset incorporates specific criteria. Exclusively, only land pixels are considered, excluding water bodies such as seas or lakes. The selected patches consist of 100% land pixels with dimensions of 256 × 256. To ensure dataset integrity, no overlap is permitted between training, testing, and validation patches. We used the seven tiles of S2 images acquired from the Chilean Patagonia area from January 2018 to December 2022 and a total of 34 k patches were generated. The study site is partitioned with 15% and 5% randomly allocated for the validation and test datasets, respectively, while the remaining 80% form the training dataset.

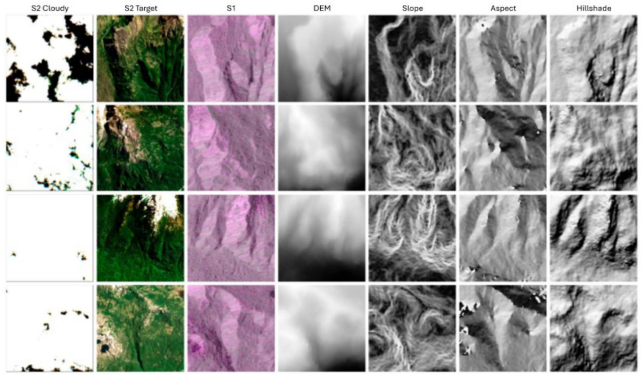


Fig. 3. Examples 256 × 256 px patch input from the dataset. The columns exhibit various images: RGB S2 cloudy images and S2 target cloud-free optical images, and a composite of the two polarization channels of S1 (where R = VV, G = VH, B = VV). Additionally, the patch includes grayscale representations of DEM, slope, sspect, and hillshade, with white color indicating higher values.

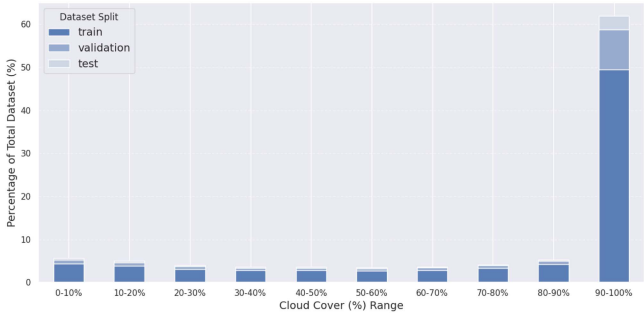


Fig. 4. Distribution of cloud coverage percentages in the dataset. The dataset spans a wide range of cloud cover levels, from 0% –10% to 90% –100%, ensuring diverse evaluation scenarios.

4) Cloud Type Distribution: Fig. 4 shows the distribution of cloud cover percentages in the dataset across ten cloud cover intervals (from 0% –10% to 90% –100%). The stacked bars represent the proportions of the training, validation, and test sets. The figure highlights the high prevalence of cloud contamination, with over 60% of the images falling within the 90% –100% cloud coverage range. In contrast, only about 6% of the images exhibit minimal cloud cover (0% –10%). The remaining cloud coverage intervals are more evenly distributed. This variability supports a comprehensive evaluation under diverse cloud conditions, ranging from heavily clouded to nearly cloud-free scenes.

To better understand the distribution of cloud conditions within the dataset, we analyzed the occurrence of thick clouds, thin clouds, and shadows across the same ten cloud cover intervals and in the training, validation, and test sets. The proportions of cloud types were derived from the cloud mask produced by the MAJA algorithm, which enables accurate detection and categorization of cloud types and shadows. This allowed for a reliable characterization of cloud conditions throughout the entire dataset. Fig. 5 presents the average percentage of each cloud type in these intervals for each dataset split. In low cloud coverage intervals (e.g., 0% –10%), thick clouds still account

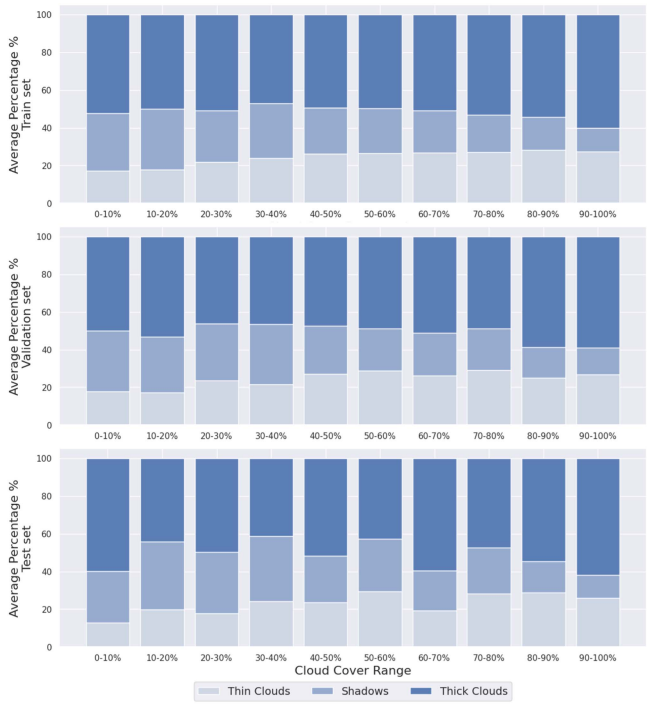


Fig. 5. Average percentage of cloud types (thick clouds, thin clouds, and shadows) across ten cloud cover intervals for the training, validation, and test sets (top to bottom).

for the majority (around 55%), followed by shadows (around 27%) and thin clouds (13% –20%). As cloud coverage increases to medium levels (50% –60%), the proportion of thin clouds increases, while thick clouds remain prominent. At the highest cloud coverage range (90% –100%), thick clouds reach their peak (around 60%), whereas shadows and thin clouds decrease Thick clouds consistently dominate across all subsets and coverage ranges, especially when a high cloud coverage is presented. Shadow, though less dominant, presents unique challenges for cloud removal, especially in mountainous regions and areas with subtle atmospheric variations.

These patterns reflect the variability and complexity of cloud structures within the dataset, which is essential for training and evaluating models like CRT-UNet. The presence of diverse cloud types across all splits and intervals ensures that the model is exposed to a wide range of atmospheric conditions, supporting the development of more generalizable and terrain-aware reconstruction strategies.

III. U-Net Architecture for Cloud Removal

A. U-Net Overview

The U-Net is a fully CNN model originally developed for medical image segmentation in 2015 [46]. This architecture consists of two essential paths: the contraction path (encoder) and the symmetric expansion path (decoder). The contraction path, resembling a typical convolutional network, involves repeated 3×3 convolutions, each followed by a rectified linear unit (ReLU), and a 2×2 max-pooling operation with a stride of 2 for downsampling. At each downsampling step, the number of feature channels is doubled, creating a feature map and reducing its size to minimize network parameters. The symmetric expansion path, or decoder, facilitates accurate localization through transposed convolution. Deconvolution on the decoder side mitigates the bottleneck problem seen in autoencoder architectures, preventing loss of features. In the expansive path, each step involves upsampling the feature map, a 2×2 convolution ("up-convolution") halving the number of feature channels, concatenation with the correspondingly cropped feature map from the contraction path to ensure the compatibility of the spatial size, and two 3×3 convolutions followed by a ReLU. Cropping compensates for the loss of border pixels in every convolution.

The U-Net architecture has diverse strengths that make it adaptable to various applications. One of its key elements is the encoder–decoder architecture, which effectively preserves both local and global information, making it suitable for tasks like image reconstruction. Another notable aspect of U-Net is the integration of skip connections, also known as shortcut or residual connections. These connections allow the model to skip one or more layers during encoding and decoding, helping to retain intricate details in the reconstructed images. Additionally, U-Net incorporates multiscale feature extraction through the use of skip connections, enabling the model to capture features at different levels of abstraction. This integration enhances the model's ability to understand and reconstruct complex structures. The optimized learning of deep models is achieved through the substantial representational capacity provided by numerous layers and filters in convolutional neural networks (CNNs). This attribute is particularly advantageous in applications such as cloud removal, where the network must reconstruct signals beneath thick clouds and restore intricate details in complex structures.

B. Proposed Model

To enhance cloud removal performance in complex landscapes, we propose CRT-UNet, a U-Net-based architecture that integrates DEM and topographic information including slope, aspect, and hillshade to improve reconstruction accuracy in mountainous regions. While the standard U-Net effectively captures spatial and spectral features, it lacks explicit terrain awareness, which is essential for reconstructing cloud-covered regions in areas with significant elevation variations. CRT-UNet addresses this limitation by incorporating DEM, slope, aspect, and hillshade as additional inputs in the second layer of the encoder, enabling the model to learn terrain-influenced patterns early in the feature extraction process. This choice also reduces computational load by avoiding the need to process high-dimensional terrain features in the early layers.

This midlevel integration strategy is inspired by the work in [30], which demonstrated that injecting auxiliary information (DEM) into intermediate layers improves learning efficiency and model performance. CRT-UNet introduces terrain context at a meaningful depth—after the initial low-level feature extraction but before higher level semantic abstraction, by fusing DEM, slope, aspect, and hillshade in the second encoder layer. That

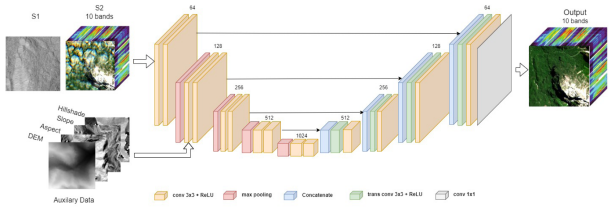


Fig. 6. Structure diagram of CRT-UNet model.

balances the preservation of fine texture details with the integration of broader terrain semantics, making it more effective than architectures that introduce auxiliary data only at the input layer or in deeper layers.

Fig. 6 illustrates the overall architecture, highlighting the integration of topographic features into the early encoder layers. The model retains the classical U-Net structure, with an encoder-decoder design and skip connections that help preserve spatial information across scales. This approach enhances spatial coherence, prevents terrain misclassification, and allows the model to reconstruct occluded areas with greater accuracy. The terrain-aware modification improves the model's ability to restore missing details, enhances spectral consistency, and mitigates common challenges associated with cloud removal in rugged environments.

IV. Experimental Set-Up

The implementation details of the proposed method include considerations for the execution environment, network training and testing configurations, as well as the preprocessing and postprocessing hyperparameters. In this experiment, a computer equipped with an AMD Ryzen Threadripper 3960X 24-Core CPU, 64GB of memory storage, and an NVIDIA GeForce RTX 3090 GPU was utilized. The deep learning framework used for all model implementations was PyTorch version 2.1. The normalization was performed on the images to ensure the range [0, 1], thus facilitating stable training. Throughout the training phase, the network utilized the Adam optimization algorithm [47] in conjunction with L1 loss. The initial learning rate was set at 0.0002, while the batch size was established to 8. The training processes were conducted for a total of 150 epochs.

To enhance dataset diversity and model robustness, data augmentation techniques were applied. These included vertical and horizontal flipping and random 90° rotations, each with a 50% probability, to simulate varying spatial configurations. The Albumentations library[4] was used to ensure efficient processing of multichannel input while preserving spatial alignment across all modalities. Augmentations were applied only during training, while validation and testing datasets remained unchanged for consistent model evaluation.

A. Computational Complexity

To evaluate the computational complexity of CRT-UNet in comparison with existing cloud removal methods, we examined the parameter count and total training time for each model considered in the study, namely a ResNet model called DSen2-CR introduced in [23], the baseline UNet-based model described in [30] and a GAN-based model called AMGAN presented in [26] and a multimodal model presented in [28]. Table I presents a summary of the key computational characteristics. The table indicates that the proposed model has a parameter count comparable to that of the UNet, and both models require less training time compared to the other models.

TABLE I
Number of Parameters and Training Time (In Hours) for the Trained Models on a NVIDIA GeForce RTX 3090 GPU With a Batch Size of 8

Model	Params (M)	Training time(h)
UNet	22.325	10
DSen2-CR	18.933	58
AMGAN	6.977	30
PLFM	73.241	53
CRT-UNet	22.328	11

B. Evaluation Metrics

The assessment of model performance involved quantitative analysis employing various metrics. In this context, we present four metrics designed for quantitative evaluation, aiming to provide a thorough assessment of cloud removal: root-mean-square error (RMSE) [48], structural similarity index measure (SSIM) [49], and spectral angle mapper (SAM) [50]. RMSE provides an overall indication of the error at the pixel level between the generated image and the ground truth. SSIM gauges perceptual similarity in structural information rather than pixel-wise absolute errors. SAM quantifies the spectral similarity, revealing the spectral distortion between two images.

We computed the following metrics between the ground true cloud free S2 (x) and the cloud free image generated by the models (y):

$$\text{RMSE(x,y)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

$$\text{SSIM(x,y)} = \frac{(2\mu_x\mu_y + (k_1L)^2)(2\sigma_{xy} + (k_2L)^2)}{(\mu_x^2 + \mu_y^2 + (k_1L)^2)(\sigma_x^2 + \sigma_y^2 + (k_2L)^2)} \quad (2)$$

$$\text{SAM(x,y)} \triangleq \arccos\left(\frac{\langle x, y\rangle}{\|x\|_2 \cdot \|y\|_2}\right). \quad (3)$$

The total number of samples is denoted as n, while d represents the maximum possible values of both x and y. The mean values of x and y are represented by μ_x and μ_y, respectively, and the variances of x and y are denoted by σ_x and σ_y. The covariance between x and y is represented by σ_{xy}. The constant parameters k_1 and k_2 default to 0.01 and 0.03, respectively. Lastly, L symbolizes the dynamic range of x and y.

V. Results

We evaluated the performance of our model by comparing it with state-of-the-art models: UNet, DSen2-CR, AMGAM, PLFM, and CRT-Unet. For that comparison, the complete test

[4]https://albumentations.ai (last accessed March 2025)

TABLE II
QUANTITATIVE EVALUATION OF THE PROPOSED METHOD CRT-UNET AGAINST STATE-OF-THE-ART APPROACHES IN TERMS OF RMSE, SSIM, AND SAM METRICS

	UNet	DSen2-CR	AMGAN	PLFM	CRT-UNet
RMSE ↓	0.080	0.086	0.176	0.138	**0.069**
SAM ↓	0.279	0.299	0.322	0.411	**0.267**
SSIM ↑	0.741	0.726	0.466	0.514	**0.823**

* Best values are in bold.

TABLE III
ABLATION STUDY RESULTS FOR THE PROPOSED CLOUD REMOVAL MODEL

Method	RMSE ↓	SAM ↓	SSIM ↑
Zeroing out DEM	0.122	0.377	0.743
Zeroing out Aspect	0.099	0.381	0.762
Zeroing out Slope	0.087	0.353	0.774
Zeroing out Hillshade	0.099	0.355	0.757
Proposed model	**0.069**	**0.267**	**0.823**

* Best values are in bold.

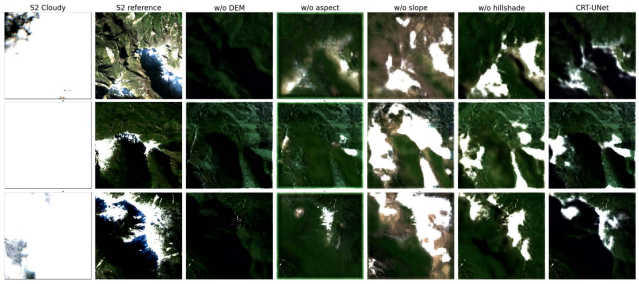


Fig. 7. Qualitative ablation study across three scenes. For each scene, the images are displayed from left to right as follows: the original cloudy image, the cloud-free reference image, followed by images with the DEM, aspect, slope, and hillshade individually zeroed out (labeled as "w/o DEM," "w/o aspect," "w/o slope," and "w/o hillshade," respectively), and finally, the output from the proposed model. Each image is sized at 256 × 256 pixels.

set is utilized. This test set consists of 1789 S2 images that cover a wide range of cloud coverage levels (from 1% to 100%).

The results generated by our proposed approach are compared with those obtained by state-of-the-art methodologies, as presented in Table II, where the average performance of the test set is illustrated. CRT-UNet stands out with higher SSIM values compared to other models, along with lower RMSE and SAM. This superiority can be attributed to the effective utilization of topographic data to extract both global features and spatial details. The proposed CRT-UNet network demonstrates significant enhancements when compared to state-of-the-art methods.

A. Ablation Study

To evaluate the impact of different input features on our cloud removal model, we conducted an ablation study, systematically removing specific inputs such as DEM, aspect, slope, and hillshade. The model's performance was evaluated using the previously described metrics, which are essential for measuring the accuracy and quality of the cloud removal process. The results of this analysis are presented in both Table III and Fig. 7 that visually depicts the effect of each feature's removal on the reconstructed images.

Fig. 7 demonstrates the visual degradation in image quality when each topographical input is excluded. For instance, when DEM is omitted, the resulting images are significantly darker with a notable loss of terrain detail especially in the existance of snow, emphasizing the critical role of DEM in elevation-dependent corrections. Similarly, removing the aspect input leads to blurred images, particularly in areas where topography significantly influences sunlight distribution, highlighting the importance of aspect data for capturing directional lighting effects. Excluding the slope results in washed-out images with less defined features in regions with steep elevation changes, underlining the slope's role in accurately modeling terrain steepness. Without hillshade, the images appear flat and less realistic, indicating the importance of hillshade in providing depth and distinguishing landforms. The proposed model, which incorporates all these inputs, produces images that closely match the ground truth, demonstrating the importance of each component. The ablation study reveals the importance of each topographical feature in the accurate reconstruction of cloud-free Sentinel-2 images. When key inputs like DEM, aspect, slope, or hillshade are removed, the model's output quality degrades, leading to darker, blurred, or flattened images. This indicates that these features are crucial for capturing the complex interactions between terrain and illumination in satellite imagery.

Quantitatively, Table III reveals that excluding DEM led to the most significant decline in performance, with RMSE increasing from 0.069 to 0.122, and SSIM dropping from 0.823 to 0.743, indicating a substantial loss in accuracy and image quality. Setting the aspect to zero resulted in a slight increase in RMSE to 0.099, with corresponding drops in SSIM. Although this scenario outperformed the DEM removal, it still fell short compared to the model utilizing all topographic features, underscoring the role of aspect in structural preservation, albeit less critical than DEM. Excluding slope resulted in an RMSE of 0.087, and an SSIM of 0.774, which were closer to the performance achieved with all topographic features, yet still indicated that slope is crucial for minimizing errors and improving image quality. The removal of hillshade led to an RMSE of 0.099, an SSIM of 0.757, showing a decline in model performance, although not as pronounced as with the exclusion of DEM.

Overall, the proposed model, which incorporates all input features, achieved the best performance with an RMSE of 0.069, and an SSIM of 0.823. This ablation study underscores the critical role each input feature plays in the cloud removal model's performance. By systematically zeroing out each input, we observed that the absence of any single feature leads to a decrease in performance, though the extent of the impact varies. Among all features, DEM had the most substantial impact, particularly in capturing elevation-related variations and supporting illumination-aware reconstruction. However, the results also demonstrate that aspect, slope, and hillshade play complementary and nonnegligible roles. Each contributes uniquely to maintaining structural integrity, minimizing reconstruction errors, and improving the realism of shaded regions. Although the performance drop was most severe when DEM was excluded, the removal of any single feature consistently degraded

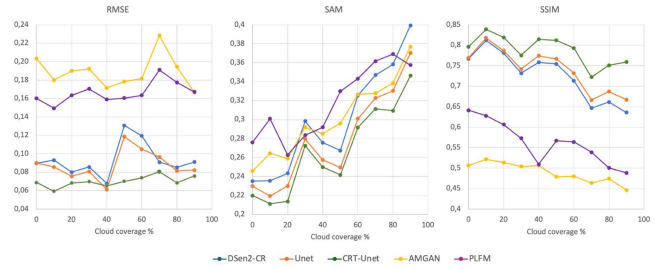


Fig. 8. Quantitative comparisons of proposed CRT-UNet to state-of-the-art methods on different cloud cover levels.

model performance, confirming that the full combination of topographic inputs is essential for optimal reconstruction quality in complex terrains.

B. Influence of Cloud Coverage Level

Additionally to the overall performance of the model, we conducted a comprehensive comparative analysis of the CRT-UNet approach at various levels of cloud cover. The evaluation involved assessing the performance of cloud removal in images with cloud cover ranging from 0% to 100% with a step of 10%. The comparison results in terms of the RMSE, SAM, and SSIM quality metrics are presented in Fig. 8.

Fig. 8 demonstrates that the proposed method outperforms other state-of-the-art approaches across all levels of cloud coverage. A noticeable trend is the negative correlation between model performance and the extent of cloud coverage, as depicted in the graph. Lower cloud coverage, such as in the [0%, 20%] interval, consistently yields better metrics, while performance declines with higher cloud cover. Across all metrics for the three models, this range exhibits the poorest outcomes. However, an anomaly is observed in the results for images within the range of cloud coverage from 60% to 70%. This exception may be attributed to the specific characteristics of clouds within this range, which influence the performance of the models. As shown in Fig. 5, this interval has one of the highest proportions of thick clouds (59.55%) compared to other ranges of cloud coverage. Thick clouds pose significant challenges for cloud removal due to their high optical density, which completely hides both spatial and spectral information from the underlying landscape. This forces the model to rely heavily on the learned patterns and contextual cues of the surrounding regions, increasing the likelihood of reconstruction errors.

Fig. 9 illustrates the outcomes of cloud removal models at varying cloud coverage levels. The analysis of these results emphasizes the consistent superiority of our proposed method in achieving higher accuracy in cloud removal across different percentages of cloud cover compared to other models. The CRT-UNet network, leveraging global features and contextual information from topographic data as input, significantly enhances the performance of the cloud removal task. Fig. 9(a) displays an example with a low percentage of clouds in the upper left corner of the image. The results indicate that all models produce cloud-free images, though with varying levels of performance. Notably, AMGAN, PLFM, and DSen2-CR introduce some spectral artifacts, particularly in the shadowed areas of the hill.

In Fig. 9(b), the outcomes of the cloud removal model in a complex area, featuring hills and a water body (river) with different types of clouds (thick and thin), and shadow are presented. DSen2-CR along with AMGAN struggled to remove all thick clouds and exhibited haziness in the water body. Moreover, AMGAN showed some artifacts and blurring in locations where clouds were removed. PLFM, on the other hand, generated some spectral distortion especially in the top left corner of the image, giving the effect of the cloud. In contrast, CRT-UNet effectively removed all types of clouds, including their shadows in the left bottom corner of the image, and mitigated pixelization issues apparent in the UNet model. Fig. 9(c) illustrates a case where both thick clouds and shadows are present. The DSen2-CR model exhibited artifacts and was unable to fully remove the clouds or restore the spatial details. AMGAN performed the worst, showing significant spectral distortion across most of the image and a loss of spatial information. In contrast, the proposed CRT-UNet successfully recovered the majority of the spatial characteristics. Handling high-level cloud cover poses a greater challenge, as exemplified in Fig. 9(d) and (e), showing extreme cases of thick clouds. DSen2-CR introduced spectral distortion and lost spatial details, including hills and snow coverage. PLFM failed to remove all the clouds from the completely cloudy seances. Conversely, CRT-UNet successfully recovered important spatial details, slopes, and hills, maintaining greater spectral consistency in the output image without distortion.

The importance of topographic data, particularly extracted from the DEM, becomes evident in mountainous regions with substantial elevation differences. Information derived from slope, aspect, and hillshade enables CRT-UNet to distinguish between flat and sloped terrains, accurately identify hill locations, and help in the comprehension of elevation variations and spatial attributes of the landscape. This underscores that the rich complementary information encoded in the DEM and topographic data images effectively improves cloud removal performance.

C. Performance on Challenging Mountainous Area

The inclusion of results from challenging mountainous locations such as rugged peaks test the applicability and the robustness of our model, which offer valuable insights into the performance under diverse and demanding conditions. Mountainous and high-elevation terrains pose a significant challenge for cloud removal models. This is primarily due to the intricate interplay between rugged topography and atmospheric conditions. The cloud patterns in these regions are characterized by their dynamic and fast-paced nature, making it difficult for traditional algorithms to precisely differentiate between clouds and the complex terrain. To address these difficulties, cloud removal models must be equipped with advanced capabilities to identify spatial and spectral features that are unique to mountainous areas. Furthermore, they should be able to adapt to the various atmospheric phenomena that occur at higher altitudes.

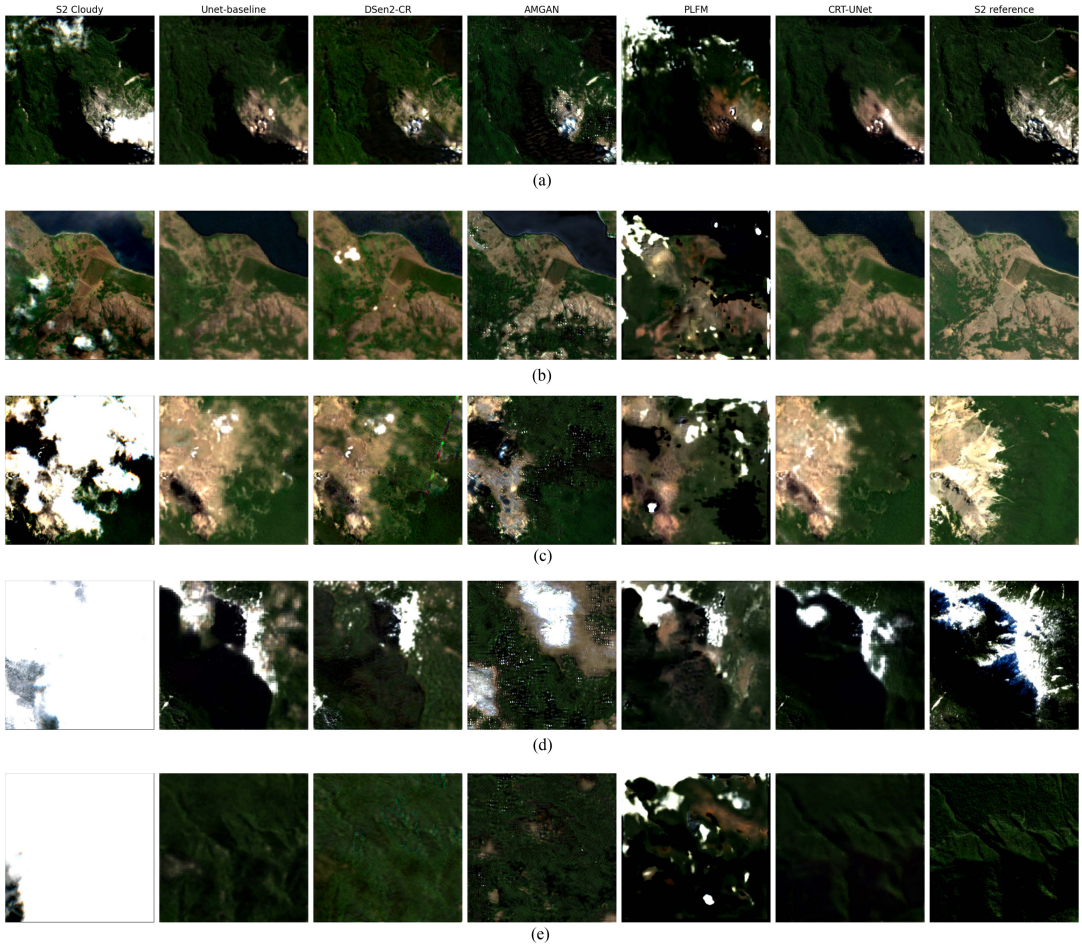


Fig. 9. Qualitative results of cloud removal models for different scenes with different cloud cover levels. Rows (a)–(e) depict cloud cover percentages from top to bottom 36%, 54%, 71%, 89%, and 100%, respectively. Within each scene, the horizontal sequence showcases the cloudy image, UNet, DSen2-CR, AMGAN, PLFM, CRT-UNet output, and the corresponding reference cloud-free image. All images have a size of 256 x 256 pixels.

TABLE IV
AVERAGE OF QUANTITATIVE EVALUATION METRICS FOR THE MOUNTAINOUS AREA OF THE TEST SET

Metric	UNet	DSen2-CR	AMGAN	PLFM	CRT-UNet
RMSE ↓	0.084	0.090	0.200	0.142	**0.074**
SAM ↓	0.255	0.266	0.264	0.555	**0.241**
SSIM ↑	0.746	0.734	0.484	0.545	**0.837**

*Best values are in bold.

Additionally, the presence of shadows caused by mountains further complicates the task at hand.

Visual and quantitative results, illustrated in Fig. 10 and Table IV, showcase the effectiveness of cloud removal in mountainous scenes, considering the varying positions of the sun, which generate shadow and snow coverage. The comparative analysis indicates the superior performance of the proposed CRT-UNet in both visual and quantitative assessments. In the first and second scenes, DSen2-CR, AMGAN and PLFM introduced spectral distortions, particularly in areas with hills and its shadows with a worst case for AMGAN. Moreover, PLFM failed to remove all the clouds from the seance, especially in location with high elevation and snow coverage. Furthermore, they failed to preserve spatial details, resulting in an incomplete reconstruction of terrain information. In contrast, CRT-UNet successfully restored essential topographical details, reconstructing hills along with their shadows. Notably, significant snow coverage in the first scene [see Fig. 10(a)] was effectively retained. The third scene [see Fig. 10(c)], characterized by a dense and complete cloud coverage, represents a highly intricate scenario. DSen2-CR encountered challenges in restoring the image, as it faced issues such as spectral inconsistency and the preservation of a significant portion of the cloud. The UNet-baseline performed better, recovering some spatial details but retaining thin clouds, especially in the upper part of the image. AMGAN demonstrated the poorest performance among the models, failing to reconstruct the hills and their 3-D structure, and showing no snow cover. PLFM also suffered from recovering spatial details and removing the clouds, it introduces noise and spectral distortion in the hills. CRT-UNet exhibited clear superiority by eliminating a substantial portion of the cloud, preserving snow coverage, and reducing the pixelization effect introduced by UNet-baseline. In particular, terrain information was accurately detected and reconstructed. Quantitative assessments, detailed in Table IV,

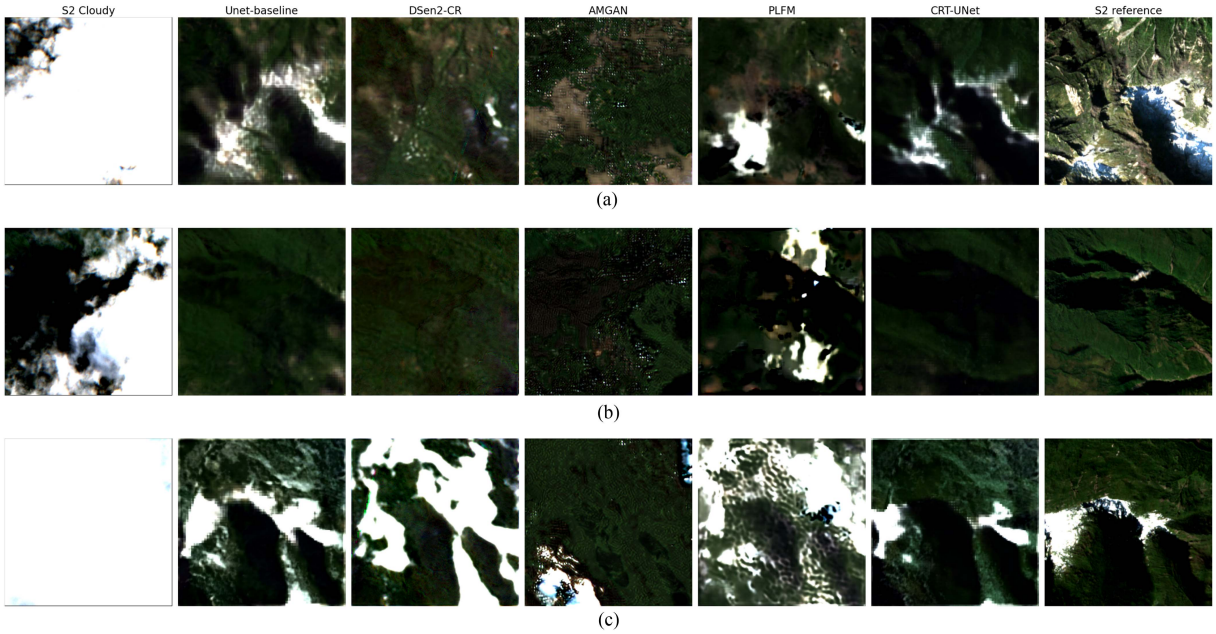


Fig. 10. Exemplary results of three different mountainous scenes. Rows (a)–(c) represent three different samples with mountain. Columns: S2 Cloudy, UNet, DSen2-CR, AMGAN, PLFM, CRT-UNet outputs, and S2 reference.

further support the efficacy of CRT-UNet, demonstrating remarkable improvements over state-of-the-art methods. The proposed approach excels in addressing the complexities inherent in mountainous scenes, establishing itself as a promising solution for cloud removal and terrain restoration in remote sensing applications.

D. Application on Large Scenes

To ensure the model robustness and reliability in real-world applications, where the extent of cloud coverage can vary significantly, we conducted extensive testing of the CRT-UNet on large-scale scenes in order to provide a more realistic representation of the challenges faced in remote sensing data analysis. Moreover, large scenes allow for the assessment of the models' ability to handle variations in cloud coverage, landscape temporal changes, and the density and thickness of clouds. Fig. 11 shows the performance of the CRT-UNet model on large-scale S2 images ($400\,km^2$), shown in both RGB (top row) and false-color compositions using bands B12, BA8, and B06 (bottom row). The first column represents the input cloudy image, where extensive cloud coverage obscures much of the underlying landscape, making surface features such as vegetation and water bodies difficult to discern. The second column shows the model's predicted output after cloud removal. In the RGB composition, the model successfully clears the cloud cover, revealing key landscape features with a high level of detail. The predicted image closely resembles the target image, shown in the third column, with visible improvements in clarity and detail, particularly in areas previously covered by clouds. Similarly, in the false-color composition, the model effectively removes clouds and recovers important surface information. The predicted output accurately reconstructs the landscape, highlighting features such as vegetation and water bodies that align closely with the target image. Across both compositions, the CRT-UNet model demonstrates its ability to generate all 10 bands of S2 imagery, managing dense cloud coverage and landscape variability. These results illustrate the model's robustness in producing cloud-free images that preserve fine landscape details, closely matching the ground truth and supporting its application in large-scale remote sensing tasks.

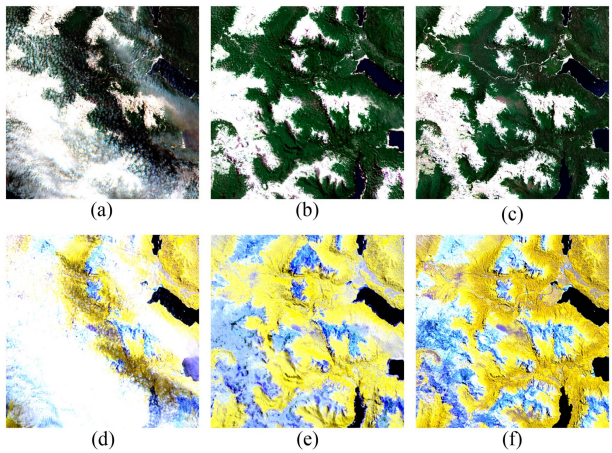


Fig. 11. Example results using CRT-UNet on large-scale S2 images in RGB (a)–(c) and false-color (B12, BA8, B06) compositions (d)–(f). Columns represent the cloudy input (left), predicted cloud-free output (middle), and target image (right). Size of the image is 20 km $\times$ 20 km.

E. Snow Cover Application

Accurate detection of snow cover is essential for effective water resource management, climate change studies, and snowpack assessment. However, cloud cover often obstructs satellite imagery, complicating snow detection, particularly in mountainous

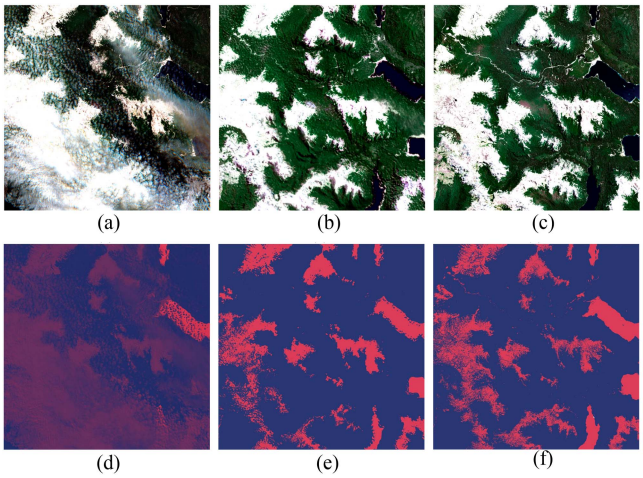


Fig. 12. Example results using CRT-UNet for snow cover estimation. (a) Top row displays RGB composites of the original cloudy image. (b) Predicted cloud-free image. (c) Target cloud-free Sentinel-2 image. The bottom row presents the corresponding snow cover maps, derived using the snow index from the images in the top row.

regions where clouds are frequent. The CRT-UNet model addresses this challenge by generating cloud-free images, thereby enabling reliable snow cover monitoring even under adverse weather conditions.

Fig. 12 illustrates an application of the cloud removal model. The top row presents RGB composites for: the original cloudy S2 image, the predicted cloud-free image, and the target cloud-free S2 image. The bottom row displays snow cover maps derived from the snow index (NDSI) for the corresponding images, with snow highlighted in red. The top row showcases the RGB imagery stages. Image (a) is the original RGB composite of S2 data, where significant cloud cover obscures the landscape. The opaque white clouds mask large portions of the surface, making snow detection challenging. The similarity between clouds and snow in visible spectral bands highlights the need for advanced processing techniques. Image (b) presents the RGB composite predicted by the model after cloud removal. This cloud-free image provides a clearer view of the landscape, allowing for a more accurate assessment of snow cover. The final image in the top row, (c), shows the RGB composition of the target cloud-free S2 image. In the bottom row, snow cover maps are shown. Map (d) corresponds to the snow cover derived from the original cloudy image (a). Due to cloud interference, this map may contain inaccuracies and potential misclassifications, with some snow-covered areas potentially underrepresented or confused with clouds. Map (e) is based on the snow mask derived from the predicted cloud-free image (b). By removing the clouds, the snow index can be applied more effectively, resulting in a more accurate estimation of snow cover. This map demonstrates a significant improvement over the cloudy scenario, as it eliminates cloud-induced errors and better isolates snow-covered pixels. Map (f) represents the snow cover derived from the target cloud-free image (c). This map provides the most refined and accurate snow cover estimation, highlighting the model's ability to differentiate between snow and clouds. Having cloud-free images is crucial for correctly estimating snow cover. In areas where cloud cover is persistent due to atmospheric conditions, obtaining a significant number of cloud-free images can be extremely challenging. The CRT-UNet model's ability to generate cloud-free imagery is therefore vital for accurate snow cover detection in such environments.

Quantitative analysis further supports the model's effectiveness. Based on the number of positive pixels in the snow masks, the calculated snow cover areas are as follows. For the original cloudy Sentinel-2 image (a), the snow cover area is 38.6 km^2. The predicted cloud-free image (b) yields a snow cover area of 50.9 km^2. The target cloud-free image (c), processed with the snow index, provides the most accurate snow cover estimation with an area of 55.5 km^2. These results indicate that the model significantly improves snow cover assessment, with the final snow cover map reflecting the closest approximation of true snow extent despite initial cloud cover. Overall, the CRT-UNet model offers a promising approach for snow cover estimation under challenging conditions. By combining cloud removal with snow index analysis, the model enhances estimation accuracy and facilitates continuous snow monitoring. This approach is adaptable to various regions and seasons, making it a valuable tool for hydrological applications, snowmelt forecasting, and water resource management.

VI. Discussion

In this study, we introduced CRT-UNet, a novel model based on the UNet architecture to remove clouds in challenging mountainous areas. Our method focusses on leveraging Sentinel-2A data exclusively to address spectral and temporal disparities between Sentinel-2A and Sentinel-2B satellites. The results of the application of CRT-UNet for cloud removal, utilizing S1 and S2 imagery along with topographic data including DEM, slope, aspect, and hillshade, show significant progress in handling the complexities of mountainous landscapes in remote sensing applications. The superior performance of CRT-UNet is attributed to its robust image reconstruction capabilities, particularly noticeable in terrains characterized by varying sun positions, shadows, and snow coverage.

CRT-UNet's architecture has demonstrated remarkable proficiency in image reconstruction. The encoder and decoder architecture allows the model to capture both global context and detailed information, which is crucial for image reconstruction tasks, where preserving fine details while understanding the overall structure is critical. Its distinctive architecture, which includes skip connections and multiscale feature fusion, maintains detailed information while effectively capturing features, resulting in precise image reconstruction, as demonstrated in tasks like cloud removal. CRT-UNet's strength lies in its effective integration of multisource data, including S1 and S2 imagery, DEM, and topographic data, to produce precise reconstructions of mountainous scenes. This information plays a crucial role in understanding the characteristics of the terrain. It offers a digital representation of elevation, which serves as the basis for identifying key features such as peaks, valleys, ridges, and slopes. By integrating these data, the model establishes a foundational understanding of the land surface shape and elevation.

Leveraging this information enables effective identification, characterization, and interpretation of terrain features even when obscured by cloud cover. It also provides a detailed depiction of the terrain's morphology, orientation, and spatial variations in illumination, thus facilitating the reconstruction process. By leveraging complementary information from these datasets, CRT-UNet improves its understanding of terrain topography, resulting in more accurate cloud removal and terrain restoration. This capability is especially valuable for snow cover estimation, where accurate representation of terrain features is crucial for differentiating snow from other surface elements. The model's proficiency in integrating detailed terrain information enhances the reliability of snow detection and monitoring, providing a solid foundation for subsequent snow cover analysis and related applications.

Comparing our proposed method with existing approaches in the literature underscores the critical importance of selecting the correct inputs for cloud removal, particularly in remote sensing applications. In our model, the inclusion of topographic data and Sentinel-1 inputs were crucial in enhancing the cloud removal process, especially in complex terrains. The unique characteristics of the landscape, such as elevation variations and specific surface features, often necessitate tailored inputs to achieve optimal results. This comparison highlights that choosing the appropriate combination of input data, informed by the area's unique geography, significantly improves the accuracy and reliability of cloud removal.

Our proposed methodology, which integrates topographic data, aligns seamlessly with prior research methodologies that have focused on analyzing intricate mountainous terrain. For instance, Hess et al. [51] utilized topographic data to investigate glacier changes in mountain regions. Similarly, other studies [52], [53] have leveraged this information to explore its correlations with precipitation patterns and vegetation dynamics in such landscapes. These findings collectively underscore the significant influence of topography in mountainous areas. The significance of topographic data is underscored in mountainous regions with significant elevation differences. Information derived from slope, aspect, and hillshade aids CRT-UNet in distinguishing between flat and sloped terrains and in identifying hill locations. This highlights the importance of using rich complementary information encoded in DEM and topographic data images to improve cloud removal performance. In conclusion, the effectiveness of the CRT-UNet model in cloud removal and terrain restoration in mountainous areas stems from its robust architecture and effective integration of multisource data.

Further analysis of results reinforces the efficacy and superiority of the proposed CRT-UNet method in cloud removal tasks at varying levels of cloud cover and complex terrain features. Leveraging global features and contextual information from topographic data enhances the robustness and accuracy of the cloud removal model, showcasing its potential for various remote sensing applications, especially in mountainous regions.

The results of snow cover estimation and large-scale scene applications underscore the CRT-UNet model's effectiveness in overcoming cloud cover challenges in satellite imagery. For snow cover estimation, the model accurately reconstructs snow-covered terrain from cloudy Sentinel-2 images, which is useful for streamflow estimation, water management, and related fields. By generating cloud-free images, the CRT-UNet ensures precise snow detection, which is critical for applications such as water resource management and avalanche prediction.

In large-scale scene applications, the CRT-UNet demonstrates its robustness in handling varying levels of cloud coverage across expansive landscapes. The model effectively removes dense clouds from the 10 Sentinel-2 bands, revealing key surface features like vegetation and water bodies with accuracy. Its ability to produce all 10 bands of Sentinel-2 imagery while preserving fine landscape details supports a wide range of remote sensing tasks, including agricultural monitoring and climate modeling. These results highlight the model's capacity to manage real-world complexities, such as cloud thickness and landscape variability, making it a valuable tool for generating high-quality, time-series cloud-free images.

However, it is essential to address the limitations inherent in our model. While our implementation demonstrates remarkable efficacy, we encountered challenges primarily associated with pixelization effects in intricate scenes with complex landscapes. One of the primary challenges we encountered was pixelization, stemming from the resolution mismatch between input bands. The model utilizes 4 inputs with 10 m resolution and 10 inputs with 20 m resolution, leading to pixelation artifacts in the final output. This issue is particularly noticeable in scenes with intricate textures or detailed landscapes. Despite the robustness of the architecture, further refinement is imperative to enhance the spatial quality and fidelity of reconstructed images, particularly in scenarios with intricate details or fine textures. These shortcomings underscore the need for continued research and development efforts to improve the performance of CRT-UNet. Strategies such as incorporating more sophisticated image processing techniques, leveraging higher resolution datasets, and exploring novel architectural modifications could potentially mitigate these limitations and elevate the model's capacity to faithfully reconstruct complex visual scenes.

VII. Conclusion

Despite the progress made in cloud removal models within the existing literature, the mountainous landscape continues to present challenges due to its distinctive topographic and atmospheric features. In our study, we have shown the feasibility of enhancing image quality through a thick cloud removal process tailored specifically for mountainous regions. Introducing CRT-UNet, a UNet-based multisource model to address the challenges posed by cloud cover in rugged terrains in such landscapes. By harnessing information from multiple sources including S2 imagery, S1 data, and DEM-derived topographic data, CRT-UNet offers a robust solution to the persistent issue of cloud obstruction in optical remote sensing data. We validated our approach using real-world data comprising S2 images with varying degrees of cloud coverage. Our results demonstrated that CRT-UNet surpasses existing methods such as DSen2-CR,

AMGAN, and PLFM in a range of evaluation metrics. Notably, our model exhibits superior spectral consistency compared to other cloud removal techniques. Visual evaluations further confirmed the efficacy and superiority of our proposed method, particularly in complex mountainous regions with snow coverage. We highlight the significance of leveraging topographic data, particularly slope, aspect, and hillshade, in areas with substantial elevation variations. CRT-UNet effectively discerns between different terrain types and accurately identifies features such as hills and valleys. This underscores the importance of utilizing rich complementary information encoded in DEM and topographic data to improve cloud removal performance, especially in challenging terrain conditions.

However, it is important to acknowledge the limitations of our model. We observed challenges related to pixelization effects in some complex scenes, as well as the need for further refinement to enhance spatial details in reconstructed images. Future research efforts should focus on addressing these limitations to improve the overall performance of CRT-UNet in real-world applications.

REFERENCES

[1] G. Srivastava and K. Shankar, "Advances in hyperspectral remote sensing for earth monitoring and mapping," *Can. J. Remote Sens.*, vol. 48, pp. 575–578, 2022.

[2] R. R. Navalgund, V. Jayaraman, and P. Roy, "Remote sensing applications: An overview," *Curr. Sci.*, vol. 93, pp. 1747–1766, 2007.

[3] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, and M. Ranagalage, "Sentinel-2 data for land cover/use mapping: A review," *Remote Sens.*, vol. 12, no. 14, 2020, Art. no. 2291.

[4] D. Malakhov and O. Dolbnya, "Remote sensing as a tool of biological conservation and grassland monitoring in mountain areas of Southeastern Kazakhstan," *J. Appl. Sci. Technol. Trends*, vol. 4, no. 1, pp. 72–79, 2023.

[5] W. Deng, K. Wang, X. Liu, T. Zhang, H. Liu, and J. Liu, "Research on remote sensing detection method for distributed subsurface targets inside mountain bodies," in *Proc. Int. Conf. Artif. Intell. Comput. Inf. Technol.*, 2022, pp. 1–6.

[6] Z. Zhang, "Retracted article: Mountain rainfall forecast and regional environmental economic development model based on remote sensing images," *Arabian J. Geosciences*, vol. 14, no. 12, 2021, Art. no. 1176.

[7] C. Xiong, J. Yang, J. Pan, Y. Lei, and J. Shi, "Mountain snow depth retrieval from optical and passive microwave remote sensing using machine learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 2001705.

[8] R. Jing, F. Duan, F. Lu, M. Zhang, and W. Zhao, "Cloud removal for optical remote sensing imagery using the SPA-CycleGAN network," *J. Appl. Remote Sens.*, vol. 16, no. 3, pp. 34520–34520, 2022.

[9] C. Zhang, W. Li, and D. Travis, "Gaps-fill of SLC-OFF landsat ETM satellite image using a geostatistical approach," *Int. J. Remote Sens.*, vol. 28, no. 22, pp. 5103–5122, 2007.

[10] A. Maalouf, P. Carré, B. Augereau, and C. Fernandez-Maloigne, "A bandelet-based inpainting technique for clouds removal from remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2363–2371, Jul. 2009.

[11] G. Hu, X. Li, and D. Liang, "Thin cloud removal from remote sensing images using multidirectional dual tree complex wavelet transform and transfer least square support vector regression," *J. Appl. Remote Sens.*, vol. 9, no. 1, pp. 95053–95053, 2015.

[12] M. Xu, M. Pickering, A. J. Plaza, and X. Jia, "Thin cloud removal based on signal transmission principles and spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1659–1669, Mar. 2016.

[13] C.-H. Lin, K.-H. Lai, Z.-B. Chen, and J.-Y. Chen, "Patch-based information reconstruction of cloud-contaminated multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 163–174, Jan. 2013.

[14] D.-C. Tseng, H.-T. Tseng, and C.-L. Chien, "Automatic cloud removal from multi-temporal spot images," *Appl. Math. Comput.*, vol. 205, no. 2, pp. 584–600, 2008.

[15] X. Li, H. Shen, L. Zhang, H. Zhang, Q. Yuan, and G. Yang, "Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7086–7098, Nov. 2014.

[16] Q. Cheng, H. Shen, L. Zhang, Q. Yuan, and C. Zeng, "Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal MRF model," *ISPRS J. Photogrammetry Remote Sens.*, vol. 92, pp. 54–68, 2014.

[17] S. Benabdelkader and F. Melgani, "Contextual spatiospectral postreconstruction of cloud-contaminated images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 204–208, Apr. 2008.

[18] B. Jiang et al., "A deep-learning reconstruction method for remote sensing images with large thick cloud cover," *Int. J. Appl. Earth Observation Geoinformation*, vol. 115, 2022, Art. no. 103079.

[19] D. Ma, R. Wu, D. Xiao, and B. Sui, "Cloud removal from satellite images using a deep learning model with the cloud-matting method," *Remote Sens.*, vol. 15, no. 4, 2023, Art. no. 904.

[20] H. Liu, B. Huang, and J. Cai, "Thick cloud removal under land cover changes using multisource satellite imagery and a spatiotemporal attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5601218.

[21] V. Sarukkai, A. Jain, B. Uzkent, and S. Ermon, "Cloud removal from satellite images using spatiotemporal generator networks," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2020, pp. 1796–1805.

[22] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4274–4288, Aug. 2018.

[23] A. Meraner, P. Ebel, X. X. Zhu, and M. Schmitt, "Cloud removal in Sentinel-2 imagery using a deep residual neural network and sar-optical data fusion," *ISPRS J. Photogrammetry Remote Sens.*, vol. 166, pp. 333–346, 2020.

[24] F. N. Darbaghshahi, M. R. Mohammadi, and M. Soryani, "Cloud removal in remote sensing images using generative adversarial networks and SAR-to-optical image translation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4105309.

[25] X. Ma, Y. Huang, X. Zhang, M.-O. Pun, and B. Huang, "Cloud-EGAN: Rethinking CycleGAN from a feature enhancement perspective for cloud removal by combining CNN and transformer," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 4999–5012, 2023.

[26] M. Xu, F. Deng, S. Jia, X. Jia, and A. J. Plaza, "Attention mechanism-based generative adversarial networks for cloud removal in Landsat images," *Remote Sens. Environ.*, vol. 271, 2022, Art. no. 112902.

[27] C. Long, X. Li, Y. Jing, and H. Shen, "Bishift networks for thick cloud removal with multitemporal remote sensing images," *Int. J. Intell. Syst.*, vol. 2023, no. 1, 2023, Art. no. 9953198.

[28] A. Sebastianelli et al., "PLFM: Pixel-level merging of intermediate feature maps by disentangling and fusing spatial and temporal data for cloud removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5412216.

[29] R. Wu et al., "An innovative approach for effective removal of thin clouds in optical images using convolutional matting model," *Remote Sens.*, vol. 15, no. 8, 2023, Art. no. 2119.

[30] R. Cresson et al., "Comparison of convolutional neural networks for cloudy optical images reconstruction from single or multitemporal joint SAR and optical images," in *Proc. XXIV ISPRS Congr.*, vol. 43, pp. 1317–1326, 2022.

[31] W. W. Immerzeel et al., "Importance and vulnerability of the world's water towers," *Nature*, vol. 577, no. 7790, pp. 364–369, 2020.

[32] R. Hugonnet et al., "Accelerated global glacier mass loss in the early twenty-first century," *Nature*, vol. 592, no. 7856, pp. 726–731, 2021.

[33] M. El Hajj, N. Baghdadi, M. Zribi, and H. Bazzi, "Synergic use of sentinel-1 and sentinel-2 images for operational soil moisture mapping at high spatial resolution over agricultural areas," *Remote Sens.*, vol. 9, no. 12, 2017, Art. no. 1292.

[34] M. Zhang, F. Chen, D. Liang, B. Tian, and A. Yang, "Use of sentinel-1 GRD SAR images to delineate flood extent in Pakistan," *Sustainability*, vol. 12, no. 14, 2020, Art. no. 5784.

[35] P. Potin et al., "Copernicus Sentinel-1 constellation mission operations status," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 5385–5388.

[36] ESA Sentinel, "Missions-Sentinel Online," European Space Agency, 2014.

[37] Z. Malenovsky et al., "Sentinels for science: Potential of sentinel-1,-2, and-3 missions for scientific observations of ocean, cryosphere, and land," *Remote Sens. Environ.*, vol. 120, pp. 91–101, 2012.

[38] M. Drusch et al., "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012.
[39] Z. E. Mashimbye and K. Loggenberg, "A scoping review of landform classification using geospatial methods," *Geomatics*, vol. 3, no. 1, pp. 93–114, 2023.
[40] W. Peng, J. Zhou, C.-j. Yang, and Z. He, "Analysis on slope uncertainty based on different resolution level DEM: A case study," in *Proc. Geoinformatics Joint Conf. GIS Built Environ. Adv. Spatial Data Models Analyses*, 2008, pp. 295–306.
[41] C. B. Marsh, J. W. Pomeroy, and R. J. Spiteri, "Implications of mountain shading on calculating energy for snowmelt using unstructured triangular meshes," *Hydrological Processes*, vol. 26, no. 12, pp. 1767–1778, 2012.
[42] M. Van Den Eeckhaut et al., "The effectiveness of hillshade maps and expert knowledge in mapping old deep-seated landslides," *Geomorphol.*, vol. 67, no. 3/4, pp. 351–363, 2005.
[43] J. Colin et al., "Assessment of the performance of the atmospheric correction algorithm Maja for Sentinel-2 surface reflectance estimates," *Remote Sens.*, vol. 15, no. 10, 2023, Art. no. 2665.
[44] B. Morales, E. Lizama, M. A. Somos-Valenzuela, M. Lillo-Saavedra, N. Chen, and I. Fustos, "A comparative machine learning approach to identify landslide triggering factors in northern Chilean Patagonia," *Landslides*, vol. 18, no. 8, pp. 2767–2784, 2021.
[45] F. Chen, C. Ming, J. Li, C. Wang, and M. Claverie, "A comparison of Sentinel-2a and Sentinel-2b with preliminary results," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Valencia, Spain, 2018, pp. 22–28.
[46] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*,, 2015, pp. 234–241.
[47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, 2015.
[48] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)," *Geoscientific Model Devt. Discuss.*, vol. 7, no. 1, pp. 1525–1534, 2014.
[49] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[50] F. A. Kruse et al., "The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data," *Remote Sens. Environ.*, vol. 44, no. 2-3, pp. 145–163, 1993.
[51] K. Hess, S. Schmidt, M. Nüsser, C. Zang, and J. Dame, "Glacier changes in the semi-arid Huasco valley, Chile, between 1986 and 2016," *Geosciences*, vol. 10, no. 11, 2020, Art. no. 429.
[52] C. Prudhomme and D. W. Reed, "Relationships between extreme daily precipitation and topography in a mountainous region: A case study in Scotland," *Int. J. Climatol., A J. Roy. Meteorological Soc.*, vol. 18, no. 13, pp. 1439–1453, 1998.
[53] Y. Deng, X. Chen, E. Chuvieco, T. Warner, and J. P. Wilson, "Multi-scale linkages between topographic attributes and vegetation indices in a mountainous landscape," *Remote Sens. Environ.*, vol. 111, no. 1, pp. 122–134, 2007.

Meryeme Boumahdi received the bachelor's degree in mathematics and computer science from the University of Sultan Moulay Slimane, Beni Mellal, Morocco, in 2015, and the master's degree in information systems and networking from Abdelmalek Essaadi University, Tangier, Morocco, in 2017. She is currently working toward the Ph.D. degree in software, systems, and computing with the School of Computer Engineering, Universidad Politécnica de Madrid, Madrid, Spain.

Her research interests include image processing and artificial intelligence applied to Earth observation.

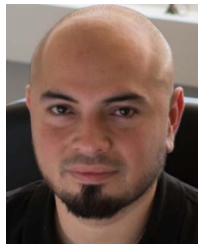
Consuelo Gonzalo-Martin received the B.A degree from the Salamanca University and the Ph.D from the Complutense University of Madrid, both in physics, in 1986 and 1989 respectively. She is a Full Professor with the Department of Computer Systems Architecture and Technology, Universidad Politécnica de Madrid (UPM), Madrid, Spain. She has worked in the development of different algorithms for optical image fusion and ART and SOM artificial neural networks. Since September 2012, she has been a Member of the MIDAS (Data Mining and Simulation) research group, Center for Biomedical Technology, UPM, where she is involved in research and development projects for text and image mining in the health care domain. As a result of her research, she has led more than 20 research projects and she has participated in 65. With 53 publications indexed in the JCR (26 Q1, 19 Q2, 5 Q3 and 3 Q4), another 28 indexed in Scopus, and SJR, 5 book chapters. She was awarded the mention of Distinguished Visit from the University of Concepcion, Concepcion, Chile, in 2015. Her main research interests include image processing, based on pixels and objects, and artificial neural networks for application in remote sensing, medical images, and face recognition.

Mario Lillo-Saavedra was born in Chile. He received the B.E. and M.Sc. degrees in electrical engineering from the University of Concepción, Concepción, Chile, in 1997 and 2000, respectively, and the Ph.D. degree in computer sciences from the Universidad Politécnica de Madrid, Madrid, Spain, in 2005.

Since 2000, he has been a Full Professor with the Faculty of Agricultural Engineering, University of Concepción. His role as a valued Member of the Water Research Center for Agriculture and Mining (CHRIAM) underscores his dedication to propelling advancements in water-related research. His research interests include remote sensing applied to agricultural water management, in addition to machine learning and data analytics.

Marcelo Somos-Valenzuela received the Ph.D. degree in water resources engineering from the University of Texas at Austin, TX, USA, in 2014.

He is currently an Associate Professor with the University of La Frontera, Temuco, Chile, with a demonstrated history of working on projects related to adaptation to climate change in diverse regions such as the Himalayas in Nepal, the Cordillera Blanca in Peru, the Continental US, and Patagonia in Chile. His research interests includes hydro-climatology, field observations and regional modeling, water resource management, and climate change.



Angel García-Pedrero received the degree in computer science from Instituto Tecnológico de Villahermosa, Villahermosa, Mexico, in 2006, the M.Sc. degree in computer science from Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, Mexico, in 2009, and the Ph.D. degree in advanced computing for science and engineering from Universidad Politécnica de Madrid, Madrid, Spain, in 2016.

He is currently an Associate Professor with the Department of Computer Systems Architecture and Technology, e Universidad Politécnica de Madrid and Researcher with the Center for Biomedical Technology. More specifically, he examines automatic techniques based on artificial intelligence for image analysis. His research experience includes the publication of more than 60 scientific papers (30 JCR-indexed journal articles and 2 nonindexed journal articles; 32 conference papers; 2 book chapters). His main research interests include data analysis, image understanding, image interpretation (remotely sensed and medical images), and deep learning.