The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume L-4/W1-2026
Free and Open Source Software for Geospatial (FOSS4G) 2026 – Academic Track “Bridging Geospatial Technology and Humanity”,
30 August–5 September 2026, Hiroshima, Japan

Eliminating Temporal Misalignment in SAR Flood Detection with a ConvLSTM-Siamese Approach Using Sentinel-1 Time Series

Tatsuya Nakajima[1], Narumasa Tsutsumida[1,2]

[1] Graduate School of Science and Engineering, Saitama University, Saitama, Japan - nakajima.t.419@ms.saitama-u.ac.jp
[2] Space Data Frontiers Research Center, Fujitsu Research, Fujitsu Limited, Kanagawa, Japan - rsnaru.jp@gmail.com

Keywords: Flood Detection, Time-Series Analysis, ConvLSTM, Synthetic Aperture Radar, Change Detection, Sentinel-1.

Abstract

Flood risk has been increasing worldwide due to climate change and rapid urbanization. Rapid and accurate flood mapping is essential for reducing damage and supporting rescue activities. Synthetic Aperture Radar (SAR) has been widely used for flood monitoring because it can observe the Earth’s surface regardless of weather conditions or time of day. Conventional flood detection methods based on change detection between pre- and post-flood SAR images, however, often suffer from false detections caused by seasonal vegetation changes and speckle noise. This study proposes a flood detection method that generates a predicted SAR image representing the normal ground condition using a ConvLSTM model and compares it with an observed SAR image acquired during flooding. To preserve structural information while suppressing noise, the prediction model was trained using the Structural Similarity Index Measure (SSIM) as the loss function. In addition, a Siamese model was employed to model the correlation between predicted and observed images for flood change detection. Experimental results demonstrated that the proposed method reduced false detections and improved overall flood detection accuracy compared with conventional approaches. The results indicate that reducing the temporal gap between comparison images is effective for improving flood detection performance in SAR imagery.

1. Introduction

Large-scale flood disasters caused by climate change and rapid urbanization have been increasing worldwide (Febrian et al., 2025; Chen et al., 2021), and rapid and accurate flood mapping is essential to reduce the human and economic losses (Portalés-Julià et al., 2023). Satellite remote sensing has been widely used for flood monitoring because it enables large-scale observation of the Earth’s surface within a short period of time. Among remote sensing technologies, Synthetic Aperture Radar (SAR) has the advantage of observing the ground surface regardless of weather conditions or time of day (Chini et al., 2019; Bernardino et al., 2024). Water surfaces generally exhibit low backscatter intensity in SAR images. Based on this characteristic, numerous flood detection methods using thresholding and change detection have been proposed (Chen et al., 2021; Liang and Liu, 2020). However, threshold-based flood mapping using a single Sentinel-1 SAR image often struggles to distinguish flooded areas from permanent water bodies because both typically exhibit low backscatter values caused by specular reflection (Chen et al., 2021). Consequently, rivers, lakes, and reservoirs may be misclassified as floodwater, leading to false detections and reduced mapping accuracy (Liang and Liu, 2020). As a practical alternative, bi-temporal change detection methods that compare pre- and post-flood SAR images have been widely adopted (Lin et al., 2019; Risling et al., 2024). By directly identifying backscatter changes associated with inundation, these methods can effectively identify flooded areas. Nevertheless, they often suffer from false detections caused by temporal gaps between observations, such as seasonal vegetation dynamics, land-cover changes, and variations in speckle patterns resulting from satellite revisit intervals (Li et al., 2023). Another research direction is the use of interferometric SAR (InSAR) techniques for flood detection. These methods exploit phase information between SAR acquisitions and can improve the identification of inundated areas (Chini et al., 2019; Chaabani et al., 2018). However, their reliance on complex interferometric processing limits their applicability to rapid operational flood monitoring (Chini et al., 2019; Ansari et al., 2017). In this sense, it is worth considering analyzing trends in dynamic land surface changes under normal conditions using long-term backscatter intensity time-series data, rather than limiting the analysis to a comparison of two specific periods (Li et al., 2023; Moskolai et al., 2020).

In recent years, time series analysis using deep learning has become popular in the remote sensing community. In particular, ConvLSTM is well-suited for processing two-dimensional data, such as images, and has the advantage of simultaneously analyzing spatial information, such as pixel layout, alongside temporal changes (Mou et al., 2018; Tiwari et al., 2024; Lu et al., 2023). Consequently, ConvLSTM has demonstrated high accuracy in predicting images from time series data, as well as in predicting dynamic ground surface changes caused by land subsidence (Lu et al., 2023; Leng et al., 2023). If a model can accurately predict the expected normal-state SAR image at the time of observation, flood detection can be formulated as the identification of deviations from the predicted normal conditions rather than as a comparison between two observations acquired at different times. Such an approach may have the potential to reduce false detections caused by seasonal vegetation changes and long-term surface variations. Although previous studies have demonstrated the feasibility of predicting SAR images at the next time step of observation, they have mainly focused on prediction accuracy only (Moskolai et al., 2020), and little attention has been paid to whether the predicted images can serve as effective references for real applications, such as flood detection. The usefulness of spatiotemporal prediction information derived from normal conditions for flood mapping has not yet been sufficiently validated.

Therefore, to achieve accurate flood detection from SAR backscatter intensity time series data, we develop a novel method

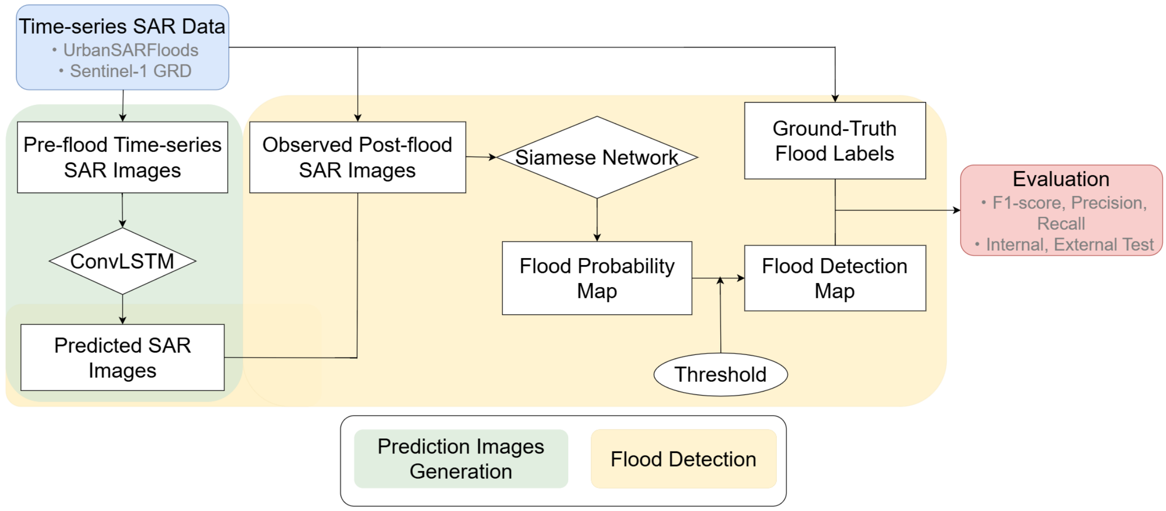

Figure 1. Workflow of the proposed flood detection method.

that (i) predicts the most recent SAR observation image from the prior time series behavior at the timing of flooding, and (ii) detects flood extents by comparing the predicted image with the actual latest observation. We first construct a ConvLSTM model that generates predicted normal-state images using the “UrbanSARFloods” dataset (Zhao et al., 2024) and Sentinel-1 time-series SAR backscatter intensity data. Next, we construct a Siamese network to detect flooded areas by comparing the generated prediction images with actual flood-time SAR observations. Flood detection performance is then evaluated using the reference data included in UrbanSARFloods. The proposed method is expected to improve flood detection accuracy by reducing temporal inconsistencies in satellite SAR image comparisons.

2. Materials

The workflow of the proposed method consists of four main components: (1) preparation of time-series SAR datasets, (2) generation of a predicted normal-state SAR image using a ConvLSTM model, (3) flood detection by comparing the predicted image with an observed post-flood SAR image through a Siamese network, and (4) performance evaluation using ground-truth flood labels (Figure 1). The framework aims to detect flood-induced changes by identifying deviations between predicted normal conditions and actual post-flood observations.

2.1 Data

This study utilized two types of datasets: UrbanSARFloods and Sentinel-1 GRD data. UrbanSARFloods is a publicly available SAR image dataset developed for global-scale flood detection (Zhao et al., 2024). The dataset covers 18 flood events from five continents worldwide (Figure 2) and was generated from Sentinel-1 Single Look Complex (SLC) data and annotated flood-label data. It consists of 8,879 image chips with a spatial resolution of 20 m and a size of 512×512 pixels, covering a total area of approximately 807,500 km^2. In this study, pre- and post-flood Sentinel-1 SAR backscatter intensity data in VV and VH polarizations were extracted from the dataset and used for flood detection. The labels are classified into three categories: (1) non-flooded areas, (2) urban flooded areas, and (3) other flooded areas. To simplify, we merged all flooded categories into ”flood” areas. In this study, 628 image chips from 15 cities worldwide were used. Among them, 578 chips, excluding Weihui and Jubba, were divided into training, validation, and test datasets at a ratio of 7:2:1 for model development. The remaining 50 chips from Weihui and Jubba were used as independent test data to evaluate the generalization capability of the proposed method.

Sentinel-1 is a C-band SAR satellite operated by the European Space Agency (ESA) under the Copernicus Earth observation program (Torres et al., 2012; European Space Agency, 2024). In this study, we used the backscatter intensity for the VV and VH polarization modes from Ground Range Detected (GRD) products. Sen1Floods11 is a benchmark dataset for flood mapping based on Sentinel-1 imagery and provides metadata associated with each flood-event image, including the acquisition orbit direction (ascending or descending) (Bonafilia et al., 2020). For the flood events included in UrbanSARFloods, time-series data covering one year prior to each flood event were acquired using Google Earth Engine (GEE) (Gorelick et al., 2017). To ensure geometric consistency within each time series, only Sentinel-1 GRD images acquired from the same orbit direction as the corresponding flood-event image were selected. GRD products are generated by processing Single Look Complex (SLC) data and applying multilook processing using multiple observations to reduce SAR-specific speckle. Although phase information is lost through this processing, the resulting data preserves backscatter intensity information.

2.2 Model

This study utilized two deep learning models: a ConvLSTM model for image prediction and a Siamese model for flood detection. ConvLSTM is a hybrid model that integrates convolutional operations into the gate structures of Long Short-Term Memory (LSTM), a type of recurrent neural network (RNN). LSTMs have three gate structures—input, output, and forget gates —that selectively retain past information in memory cells when processing time-series data. Although LSTM, which relies on matrix operations, is effective for one-dimensional sequential data, it cannot effectively process the spatial arrangement of pixels in two-dimensional image data (Lu et al., 2023; Shi et al., 2015). To address this limitation, ConvLSTM enables processing while preserving the three-dimensional structure of the data, including temporal information and spatial dimensions such as height and width. As a result, ConvLSTM has demon-

Figure 2. Locations and numbers of image chips used in this study.

strated high performance in complex prediction tasks such as spatiotemporal simulation of time-series imagery and land subsidence prediction (Moskolai et al., 2020). We develop the ConvLSTM model by taking a five-dimensional tensor ($Batch$, $Time$, $Height$, $Width$, $Channels$) as input and output a four-dimensional tensor ($Batch$, $Height$, $Width$, $Channels$). Specifically, $Time$ sequential SAR images are provided as input, and a predicted SAR image is generated through three ConvLSTM2D layers for feature extraction, followed by a final layer for reconstruction into the original image format.

The Siamese Network applies CNNs with shared weights to two input images and determines the presence of flooding on a pixel-by-pixel basis by comparing the extracted features. The Siamese Network maps the two input images into a shared feature space by applying the same feature extraction network to both. This shared-weight structure enables direct comparison between a predicted non-flood image and the corresponding post-flood SAR observation at the same time point. Zhou et al. (2022) demonstrated that Siamese Networks are effective for image change detection and that difference features can effectively emphasize changed regions (Zhou et al., 2022). We used this characteristic to interpret the differences between the predicted and observed images as flood-induced changes and to identify flooded areas. The model adopts an encoder–decoder architecture as its basic structure. The encoder extracts high-dimensional features from the input images, while the decoder reconstructs the feature maps to the original image size. In addition, skip connections are introduced between corresponding encoder and decoder layers to preserve spatial information lost during downsampling and to enable accurate boundary detection. The model takes two types of four-dimensional tensor inputs ($Batch$, $Channels$, $Height$, $Width$) corresponding to the predicted and observed images. First, feature maps are generated independently for each image using shared weights. Next, pixel-wise differences between the two generated feature maps are calculated to emphasize changed regions. Finally, a sigmoid activation function is applied to output a flood probability map, where the change intensity at each pixel is represented as the probability of being a flooded area (0–1).

3. Proposed Method

3.1 Prediction Images Generation

Prediction images representing the ground surface condition in the absence of flooding are generated using the UrbanSAR-Floods dataset and the ConvLSTM model. First, preprocessing was applied to the pre-flood SAR time series, excluding post-flood observations. To improve learning efficiency, a 3×3 median filter was used for speckle reduction, followed by Min-Max normalization for each band. As some time-series images contained missing bands (Figure 3), missing-data interpolation was performed. In this process, three-dimensional spline interpolation was applied on a pixel-wise, approximating temporal variations in backscatter intensity across the entire time series with smooth curves, and the missing band images were estimated and interpolated from these curves.

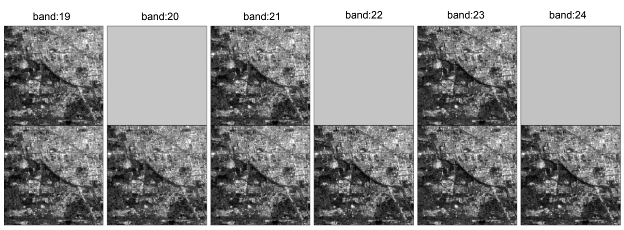

Figure 3. Example of data interpolation using a three-dimensional spline. The top and bottom panels show the data before and after interpolation, respectively.

Next, data preparation was performed for input into the ConvLSTM model. We adopted a site-specific learning approach to capture temporal surface changes unique to each observation site. To ensure a consistent temporal input length across all sites, the number of time-series images used for ConvLSTM training was limited to a maximum of 40 images per site. Furthermore, a dataset was constructed for each image chip at each observation site, with six consecutive images in chronological order forming a single set. Among them, the first five images were used as input for prediction, and the sixth image was used as ground-truth data for model training. The sequence length of five input images for image prediction was determined based on a previous study by (Moskolai et al., 2020), which demonstrated that this temporal range is sufficient to capture surface changes in time-series SAR imagery. Next, multiple image chip pairs consisting of six temporally consecutive observations were generated from the time-series data at each site by shifting the sequence one time step at a time. These datasets were then split into training and validation sets at an 8:2 ratio for each site, and an independent ConvLSTM model was trained for each Sen1Flood11 chip site to learn local surface characteristics. After training each model, the five most recent image chips prior to a flood event were fed into the trained model to generate a prediction image for the next time step, assuming a flood event.

A hybrid loss function combining Mean Absolute Error (MAE) and Structural Similarity Index Measure (SSIM) was used for the model (Equation 1).

$$\text{Loss} = \alpha \cdot \text{MAE} + (1-\alpha) \cdot (1 - \text{SSIM}) \tag{1}$$

Here, MAE directly measures pixel-wise intensity differences, whereas SSIM evaluates structural similarity between two images using statistical and structural information. The weighting coefficient was fixed at $\alpha = 0.70$. MAE is defined as follows:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |x_i - y_i| \tag{2}$$

where N denotes the number of pixels, and x_i and y_i represent the pixel values of the reference and predicted images, respectively. SSIM is defined as follows:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{3}$$

where μ_x and μ_y denote the mean pixel values of the predicted image x and ground-truth image y, respectively; σ_x^2 and σ_y^2 denote the variances; σ_{xy} denotes the covariance; and C_1 and C_2 are small constants for numerical stability ($C_1 = 0.0001$, $C_2 = 0.0009$).

The hyperparameters of the ConvLSTM model were configured as shown in Table 1. Early stopping was applied to prevent overfitting, where training was terminated if the validation loss did not improve for 10 consecutive epochs. In addition, the learning rate was reduced to 0.2 times its current value when the validation loss failed to improve for 5 consecutive epochs.

Hyperparameter	Value
Epochs	100
Patience (EarlyStopping)	10
Batch Size	2
Optimizer	Adam
Learning Rate	0.001
Patience (ReduceLROnPlateau)	5
Loss Function	MAE + SSIM

Table 1. Hyperparameter settings of the ConvLSTM model.

These processes were applied to all study sites, yielding a total of 628 predicted images.

3.2 Flood Detection

Flood areas were detected by comparing the predicted images generated in Section 3.1 with the actual post-flood SAR images. First, the predicted VV and VH polarization images were concatenated along the channel dimension to create two-channel input data compatible with the Siamese model. Of the 628 generated samples, 578, excluding those from Jubba and Weihui, were divided into training, validation, and test datasets at a ratio of 7:2:1. The remaining 50 samples were used as independent test data for evaluating the generalization capability of the proposed method. As pre-processing, Z-score normalization was applied to each channel, followed by clipping to $[-3, 3]$ to remove outliers. Next, the Siamese model described above was trained using the prepared training and validation datasets. The model takes the predicted and observed images as inputs, generates feature maps for each, and outputs a flood probability map with values ranging from 0 to 1 via a sigmoid activation function based on the difference between the extracted features. To obtain a binary flood map, a threshold was applied to the output probabilities. Pixels with values greater than or equal to the threshold were classified as flooded areas, while the remaining pixels were classified as non-flooded areas.

To address the class imbalance between flood and non-flood pixels in the dataset, Binary Focal Crossentropy (BFCE) was used as the loss function. BFCE extends the standard binary cross-entropy loss by introducing a modulation term that emphasizes difficult samples, and is defined as follows:

$$\text{BFCE}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{4}$$

where p_t denotes the predicted probability for the ground-truth class, and $-\log(p_t)$ represents the standard cross-entropy loss. The modulation term $(1-p_t)^{\gamma}$ reduces the contribution of easily classified samples and emphasizes difficult ones. The focusing parameter was set to $\gamma = 3.0$. To compensate for the imbalance between flood and non-flood pixels, the balancing parameter α_t was determined using the ratio between non-flood pixels (N_{neg}) and flood pixels (N_{pos}) in the training dataset:

$$\text{pos\_weight} = \frac{N_{neg}}{N_{pos}} \tag{5}$$

$$\alpha_t = \frac{\text{pos\_weight}}{\text{pos\_weight} + 1.0} \tag{6}$$

Because the UrbanSARFloods dataset is dominated by non-flood pixels, the resulting α_t becomes close to 1, thereby increasing the contribution of rare flood pixels during training and mitigating class imbalance. The hyperparameters of the Siamese Network model were configured as shown in Table 2.

4. Experiments

4.1 Comparative Methods

To verify the effectiveness of the proposed method, we evaluated the following two methods for flood detection.

Hyperparameter	Value
Epochs	100
Patience (EarlyStopping)	20
Batch Size	8
Optimizer	Adam
Learning Rate	0.001
Patience (ReduceLROnPlateau)	5
Loss Function	BFCE

Table 2. Hyperparameter settings of the Siamese Network model.

4.1.1 Pre- and Post-Flood Comparison Method (Pre/Post)

As a conventional baseline, flood detection was performed by directly comparing pre-flood and post-flood SAR images. SAR images processed with a median filter were input into the Siamese model to generate feature maps, and flood areas were detected from differences between the extracted features.

4.1.2 Prediction-Based Flood Detection Method (Baseline)

A prediction-based flood detection method was also implemented as an ablation baseline. This method follows the same prediction and comparison framework as the proposed approach, but without the proposed preprocessing and training configurations. Specifically, the ConvLSTM model was trained without speckle filtering, missing-data interpolation, or restrictions on the temporal sequence length. In addition, mean squared error (MSE) was used as the loss function instead of the proposed hybrid loss combining MAE and SSIM. MSE is defined as follows:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \tag{7}$$

where N denotes the number of pixels, and x_i and y_i represent the pixel values of the reference and predicted images, respectively. Because MSE tends to smooth spatial details and edges, resulting in visually blurred predictions (Mathieu et al., 2016), this baseline enables evaluation of the contribution of the proposed pre-processing and loss-function design to the final flood detection accuracy.

Finally, the thresholds used for each method are summarized in Table 3. Threshold optimization was performed using the training and validation datasets. Candidate thresholds ranging from 0.00 to 1.00 were evaluated at intervals of 0.05, and the threshold yielding the highest F1-score was selected. The optimized threshold was subsequently fixed and applied to the test dataset for the final performance evaluation.

Method	Threshold
Pre/Post	0.70
Baseline	0.65
Proposed	0.70

Table 3. Threshold values used for pre- and post-flood SAR image comparison, baseline, and proposed methods.

4.2 Evaluation Metrics

To evaluate the flood detection performance of the baseline and proposed methods, four pixel-wise classification metrics were used: Accuracy, Precision, Recall, and F1-score. These metrics were calculated based on the confusion matrix consisting of true positives (N_{TP}), true negatives (N_{TN}), false positives (N_{FP}), and false negatives (N_{FN}). The definitions of these terms are as follows.

- **True Positive** (N_{TP}): Number of pixels correctly classified as flooded areas.
- **True Negative** (N_{TN}): Number of pixels correctly classified as non-flooded areas.
- **False Positive** (N_{FP}): Number of non-flooded pixels incorrectly classified as flooded areas (false detection).
- **False Negative** (N_{FN}): Number of flooded pixels incorrectly classified as non-flooded areas (missed detection).

The evaluation metrics are defined as follows.

$$\text{Accuracy} = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}} \tag{8}$$

$$\text{Precision} = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{9}$$

$$\text{Recall} = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{10}$$

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

Accuracy represents the overall classification accuracy of the entire image. However, in highly imbalanced datasets such as flood detection tasks, where flooded pixels occupy only a small portion of the image, Accuracy tends to be dominated by correctly classified non-flooded pixels. Therefore, greater emphasis was placed on Precision, which evaluates the correctness of detected flooded areas, Recall, which evaluates the ability to avoid missed detections, and the F1-score, which represents the harmonic mean of Precision and Recall.

5. Results

5.1 Flood Detection on the Internal Test Dataset

The trained Siamese model was evaluated on a previously held-out test set to verify flood-detection performance, along with two baseline approaches (Table 4). Among the three methods, the proposed method achieved the highest F1-score of 0.605, improving by 0.324 over the baseline method and by 0.178 over the pre/post method. The proposed method achieved the highest Precision value of 0.556. In terms of Recall, the proposed method achieved the highest Recall (0.661), although the difference from the pre/post method (0.610) was relatively small. Figure 4 shows representative flood detection results from five regions (Hebei, Canada, Iran, Sydney, and Houston) included in the test dataset. The pre/post method successfully detected the general flooded regions at all locations; however, speckle-like false detections were observed throughout the images, particularly in Canada and Iran. The baseline method showed substantial over-detection in several locations, particularly in Canada, Iran, Sydney, and Houston, leading to inaccurate flood delineation regardless of flood size. In contrast, the proposed method produced flood maps that visually matched the ground truth while effectively suppressing speckle-like false detections. In addition, small and isolated flooded regions, such as those observed in Canada and Houston, were successfully detected with accurate locations and shapes while minimizing missed detections. These results indicate that our method effectively suppresses false detections while preserving the ability to detect flooded regions.

Method	Accuracy	Precision	Recall	F1-score
Pre/Post	0.970	0.328	0.610	0.427
Baseline	0.939	0.188	0.560	0.281
Proposed	**0.989**	**0.556**	**0.661**	**0.605**

Table 4. Flood detection performance of pre- and post-flood SAR image comparison, baseline method, and proposed method.

5.2 Generalization Performance of the External Test Dataset

To evaluate generalization performance, flood detection was conducted using unseen sites (Jubba and Weihui) that were excluded from training. Table 5 compares the detection performance of the proposed, pre/post, and baseline approaches. The proposed method achieved the highest Precision (0.466) and F1-score (0.483) among all methods, while maintaining a Recall comparable to that of the pre/post method. In contrast, the baseline method performed worst across all evaluation metrics, indicating limited generalization to unseen regions. These results suggest that the proposed method effectively suppresses false detections while maintaining robust flood-detection performance across different regions. Figure 5 presents flood detection results for four representative test sites. For Jubba-2 and Weihui, all methods successfully captured the overall extent of the flooded regions. However, the proposed method provided more accurate delineation of flood boundaries and fine-scale flood patterns. In contrast, flood detection errors were more evident in Jubba-1 and Jubba-3, where all methods exhibited both false detections and missed flooded areas. Although our method generally produced flood maps that better matched the ground truth, accurate flood delineation remained challenging in Jubba-3, indicating the need for further improvements in complex terrain.

Method	Accuracy	Precision	Recall	F1-score
Pre/Post	0.983	0.293	**0.541**	0.380
Baseline	0.962	0.212	0.386	0.274
Proposed	**0.988**	**0.466**	0.536	**0.483**

Table 5. Flood detection performance on unseen sites.

6. Discussion

6.1 Effect of the Loss Function

The flood detection results in Table 4 indicate that the loss function used for ConvLSTM prediction considerably affected the final detection accuracy. Figure 6 compares prediction images generated using MSE, MAE, and the proposed MAE+SSIM hybrid loss. The MSE-based model produced noisy predictions because squared errors are sensitive to speckle noise and outliers in SAR imagery. Although MAE improved intensity reconstruction, fine textures and edge information became spatially blurred. In contrast, the MAE+SSIM loss generated prediction images with improved structural consistency and edge preservation. Because SSIM evaluates luminance, contrast, and structural similarity, the proposed loss better preserves rivers, roads, and urban structures. These improvements contributed to the higher flood detection accuracy of the proposed method.

6.2 False Detections and Missed Regions

To investigate false detections by the proposed approach, we selected representative examples from mountainous and urban environments (Figure 7). In the mountainous region of Hebei, scattered false detections were frequently observed. These errors were likely caused by SAR geometric distortions, including layover and radar shadow, as well as rainfall-induced backscatter changes in vegetation areas (Wu et al., 2021). In contrast, the urban region of Houston showed several missed flood regions. This tendency is considered to be related to double-bounce scattering effects in built-up areas, where flooded regions can still exhibit high backscatter intensity (Chini et al.,

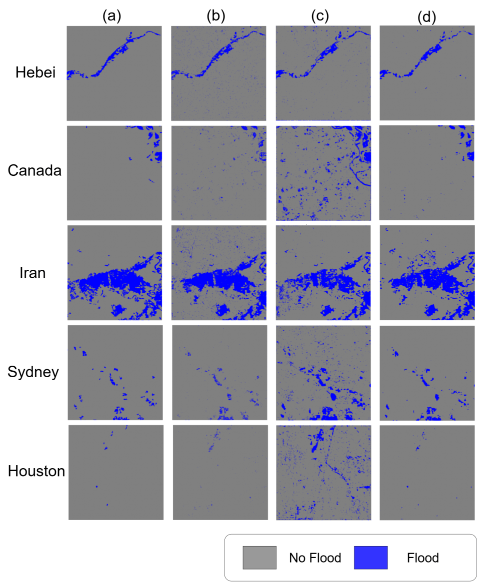

Figure 4. Flood detection results for the test data. Rows represent different test regions and columns show: (a) Ground Truth, (b) pre- and post-flood SAR image comparison, (c) baseline, and (d) proposed methods.

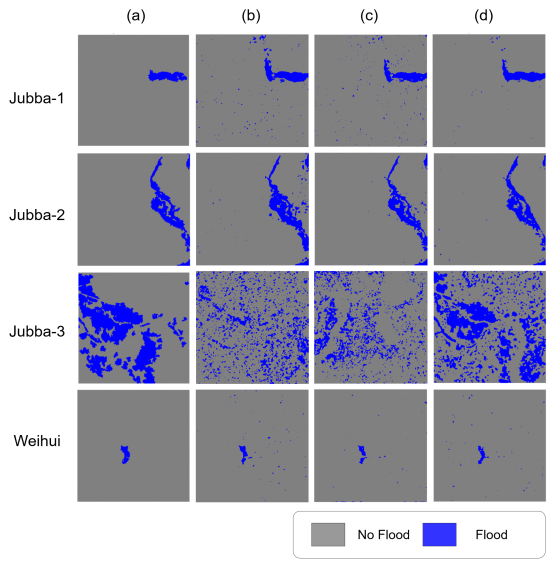

Figure 5. Flood detection results for unseen sites. Rows represent different test regions and columns show: (a) Ground Truth, (b) pre- and post-flood SAR image comparison, (c) baseline, and (d) proposed methods.

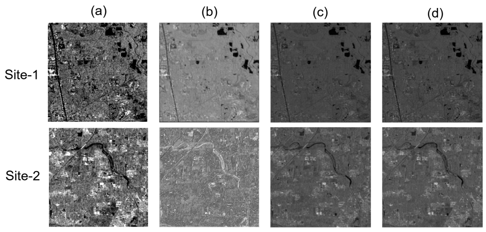

Figure 6. Predicted images generated by the ConvLSTM model using different loss functions. Rows represent individual data samples, while columns show: (a) Reference, (b) MSE, (c) MAE, and (d) MAE+SSIM.

2019; Delgado Blasco et al., 2020). To address these issues, future studies should develop more robust image registration techniques to reduce spatial inconsistencies between predicted and observed SAR images. Furthermore, incorporating coherence information and digital elevation model (DEM) data may improve the discrimination of flood-induced changes from terrain- and vegetation-related variations in backscatter intensity. In addition, fine-tuning the model with more samples from mountainous and urban regions could improve its performance in complex scattering environments, thereby reducing both false detections and missed flood areas.

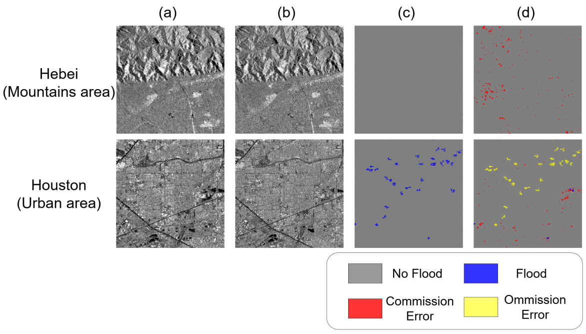

Figure 7. Examples of flood detection results on the test dataset. Rows represent different test samples, and columns correspond to: (a) observed post-flood SAR image, (b) predicted SAR image, (c) ground-truth flood map, and (d) flood detection result of the proposed method.

6.3 Effect of Temporal Sequence Length

In this study, we fixed the temporal sequence length for ConvLSTM prediction at 40 images. The temporal sequence length for ConvLSTM prediction should be selected based on the climatic characteristics and revisit intervals at each observation site. The objective of the prediction model is to learn normal land-surface dynamics in the absence of flooding. However, when excessively long temporal sequences are used, observations acquired under substantially different seasonal conditions may be included in the input data. In many regions, rainy seasons occur annually and can cause considerable changes in surface conditions even without flooding. Consequently, observations from a previous rainy season may introduce unnecessary variability into the training data, potentially degrading prediction performance. These findings suggest that adaptive selection of temporal sequence length based on regional environmental conditions may improve both predictive quality and subsequent flood-detection accuracy.

7. Conclusions

This study proposed a flood detection method that compares predicted normal-state SAR images, generated from past spatiotemporal changes, with observed post-flood SAR images. The proposed framework generated prediction images from time-series SAR datasets using a deep learning-based ConvLSTM model and subsequently extracted change regions between the two images using a Siamese network. As a result, the proposed method was the best for detecting flood regions across the compared approaches. Compared with the conventional pre-/post-flood comparison approach, the proposed method reduced false detections without substantially increasing missed flood areas. This result suggests that the proposed framework has the potential to alleviate the influence of non-flood-related temporal variations that commonly affect bi-temporal flood detection methods. These findings indicate that spatiotemporal prediction of normal surface conditions provides useful information for flood detection and can effectively mitigate the temporal-gap problem inherent in pre-/post-flood image comparison. Furthermore, the proposed method maintained moderate performance on unseen sites, suggesting its potential applicability to diverse geographic regions. However, further improvements in detection accuracy are required for practical applications. This study utilized only UrbanSARFloods and Sentinel-1 backscatter intensity data. Future work should investigate the integration of SAR coherence and DEM data. Coherence data are expected to improve the identification of flood-induced changes that are difficult to distinguish using backscatter intensity alone, particularly in urban areas where strong double-bounce scattering can affect flood detection. In addition, DEM data are expected to provide information on terrain characteristics and physical constraints on water flow, thereby helping to suppress unrealistic detections in mountainous regions. Integrating these complementary data sources is likely to contribute to further improvements in the robustness and accuracy of flood mapping. Recent advances in time-series prediction models have also been promising, such as Transformer-based architectures of PatchTST (Nie et al., 2022), EarthFormer (Gao et al., 2022), and Mamba (Gu and Dao, 2024), as well as large-scale pre-trained foundation models such as Chronos (Ansari et al., 2024) and SatMAE (Cong et al., 2022). These models may help capture longer, more complex spatiotemporal dependencies while also improving robustness to diverse environmental conditions and complex surface change patterns through more efficient learning frameworks.

Acknowledgements

This work was funded by the Disaster Prevention Research Institute's Implementation Science Research for Regional Communities (Specific) project at Kyoto University (Project No. 2024RS-01) and NEDO (23200859-0).

References

Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P. et al., 2024. Chronos: Learning the Language of Time Series. *arXiv preprint arXiv:2403.07815*.

Ansari, H., De Zan, F., Bamler, R., 2017. Sequential estimator: Toward efficient InSAR time series analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 55(10), 5637–5652.

Bernardino, P. N., Oliveira, R. S., Van Meerbeek, K., Hirota, M., Furtado, M. N., Sanches, I. A., Somers, B., 2024. Estimating Vegetation Water Content from Sentinel-1 C-Band SAR

Data over Savanna and Grassland Ecosystems. *Environmental Research Letters*, 19(3), 034019.

Bonafilia, D., Tellman, B., Anderson, T., Issenberg, E., 2020. Sen1Floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 210–211.

Chaabani, C., Chini, M., Abdelfattah, R., Hostache, R., Chokmani, K., 2018. Flood mapping in a complex environment using bistatic TanDEM-X/TerraSAR-X InSAR coherence. *Remote Sensing*, 10(12), 1873.

Chen, S., Huang, W., Chen, Y., Feng, M., 2021. An Adaptive Thresholding Approach Toward Rapid Flood Coverage Extraction from Sentinel-1 SAR Imagery. *Remote Sensing*, 13(23), 4899.

Chini, M., Pelich, R., Pulvirenti, L., Pierdicca, N., Hostache, R., Matgen, P., 2019. Sentinel-1 InSAR Coherence to Detect Floodwater in Urban Areas: Houston and Hurricane Harvey as a Test Case. *Remote Sensing*, 11(2), 107.

Cong, Y., Khanna, S., Meng, C., Liu, P., Rozi, E. et al., 2022. SatMAE: Pre-training Transformers for Temporal and Multi-Spectral Satellite Imagery. *Advances in Neural Information Processing Systems*, 35, 197–211.

Delgado Blasco, J. M., Fitrzyk, M., Patruno, J., Ruiz-Armenteros, A. M., Marconcini, M., 2020. Effects on the double bounce detection in urban areas based on SAR polarimetric characteristics. *Remote Sensing*, 12(7), 1187.

European Space Agency, 2024. Sentinel-1. European Space Agency, www.esa.int (accessed 2 February 2026).

Febrian, R. D., Kim, W., Lee, Y., Kim, J., Choi, M., 2025. Estimation of Flood Inundation Area Using Soil Moisture Active Passive Fractional Water Data with an LSTM Model. *Sensors*, 25(8), 2503.

Gao, Z., Shi, X., Wang, H., Zhu, Y., Wang, Y. B. et al., 2022. Earthformer: Exploring space-time transformers for earth system forecasting. *Advances in Neural Information Processing Systems*, 35, 25390–25403.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. et al., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27.

Gu, A., Dao, T., 2024. Mamba: Linear-time sequence modeling with selective state spaces. *First Conference on Language Modeling*.

Leng, J., Gao, M., Gong, H., Chen, B., Zhou, C., Shi, M., Chen, Z., Li, X., 2023. Spatio-temporal prediction of regional land subsidence via ConvLSTM. *Journal of Geographical Sciences*, 33(10), 2131–2156.

Li, Z., Xiang, Z., Demiray, B. Z., Sit, M., Demir, I., 2023. MA-SARNet: A one-shot nowcasting framework for SAR image prediction with physical driving forces. *ISPRS Journal of Photogrammetry and Remote Sensing*, 205, 176–190.

Liang, J., Liu, D., 2020. A Local Thresholding Approach to Flood Water Delineation Using Sentinel-1 SAR Imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159, 53–62.

Lin, Y. N., Yun, S.-H., Bhardwaj, A., Hill, E. M., 2019. Urban flood detection with Sentinel-1 multi-temporal synthetic aperture radar (SAR) observations in a Bayesian framework: a case study for Hurricane Matthew. *Remote Sensing*, 11(15), 1778.

Lu, M., Li, Y., Yu, M., Zhang, Q., Zhang, Y., Liu, B., Wang, M., 2023. Spatiotemporal prediction of radar echoes based on ConvLSTM and multisource data. *Remote Sensing*, 15(5), 1279.

Mathieu, M., Couprie, C., LeCun, Y., 2016. Deep multi-scale video prediction beyond mean square error. *International Conference on Learning Representations (ICLR)*.

Moskolai, W., Abdou, W., Dipanda, A., Kolyang, D. T., 2020. Application of LSTM architectures for next frame forecasting in Sentinel-1 images time series. *arXiv preprint arXiv:2009.00841*.

Mou, L., Bruzzone, L., Zhu, X. X., 2018. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2), 924–935.

Nie, Y., Nguyen, N. H., Sinthong, P., Kalagnanam, J., 2022. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. *arXiv preprint arXiv:2211.14730*.

Portalés-Julià, E., Mateo-García, G., Purcell, C., Gómez-Chova, L., 2023. Global Flood Extent Segmentation in Optical Satellite Images. *Scientific Reports*, 13(1), 20316.

Risling, A., Lindersson, S., Brandimarte, L., 2024. A comparison of global flood models using Sentinel-1 and a change detection approach. *Natural Hazards*, 120(12), 11133–11152.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-C., 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 802–810.

Tiwari, A., Sadhashivam, N., Ohenhen, L. O., Lucy, J., Shirzaei, M., 2024. Leveraging power of deep learning for fast and efficient elite pixel selection in time series SAR interferometry. *arXiv preprint arXiv:2402.17069*.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M. et al., 2012. GMES Sentinel-1 mission. *Remote Sensing of Environment*, 120, 9–22.

Wu, L., Wang, H., Li, Y., Guo, Z., Li, N., 2021. A novel method for layover detection in mountainous areas with SAR images. *Remote Sensing*, 13(23), 4882.

Zhao, J., Xiong, Z., Zhu, X. X., 2024. UrbanSARFloods: Sentinel-1 SLC-based benchmark dataset for urban and open-area flood mapping. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 419–429.

Zhou, H., Ren, Y., Li, Q., Yin, J., Lin, Y., 2022. MASNet: Improve performance of Siamese networks with mutual-attention for remote sensing change detection tasks. *arXiv preprint arXiv:2206.02331*.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-L-4-W1-2026-181-2026 |