Journal of Intelligent Information and Communication

Research Paper

Spatio-Temporal Gated Recurrent Unit (ST-GRU) Model for High-Fidelity Meteorological Video Forecasting

Patryk Kubiak[1, *]

[1] Faculty of Information Technology, University of Warmia and Mazury, Olsztyn 10-719, Poland
*Corresponding author: patryk.k@uwm.edu.pl

Abstract. Complexity, multi-scale, and atmospheric process dynamism are some of the issues with meteorological video forecasting. This study presents a Spatio-Temporal Gated Recurrent Unit (ST-GRU) architecture for high-resolution sequential satellite video prediction in order to address the drawbacks of the aforementioned numerical techniques and previous deep learning models. The model uses channel-wise normalization, spatial-global gating, and multi-scale dilated convolutions in a recurrent neural network to explicitly represent both local and global spatiotemporal dependencies. With a weighted RMSE of 0.188 and a mean MSSSIM greater than 0.93 based on the experimental results in the proprietary satellite video dataset, ST-GRU outperforms all top baselines, including ConvGRU, PredRNN, TrajGRU, and PhyDNet, in terms of accuracy, consistency, and structure preservation. The necessity of all the architectural elements—multi-scale context fusion, global gating, and normalization—for reliable predictions during regime changes and uncommon convective phenomena has been further confirmed by ablation investigations. According to the aforementioned findings, the suggested model has a workable architecture and is appropriate for use in real-time forecasting since it can faithfully replicate the evolution of weather changes over time. The development of high-accuracy, data-driven weather prediction using extensive video data is supported technically by this study.

Keywords: *Spatiotemporal Modeling, Meteorological Video Prediction, Recurrent Neural Network, Image Sequence Forecasting*

Received on 13 October 2025, Accepted on 29 January 2026, Published on 09 February 2026

Introduction

In today's world, meteorological forecasts are essential for a number of reasons, including helping governments and other organizations plan for natural catastrophes and make relevant choices, as well as providing early warnings of poor weather to safeguard people and property [1]. The explosion of high-resolution meteorological video data from advanced satellite sensors, ground-based monitoring systems and weather radar networks has also provided rich material for research on the full exploitation of spatio-temporal information in intelligent weather prediction [2]. However, creating accurate models for atmospheric phenomena remains a challenging task due to their inherent non-linearity, dynamic spatial properties, and oscillations in several components [3]. Numerical weather prediction models continue to be the predominant type of prior work; they are not appropriate for rapidly changing weather conditions, have low spatial and temporal resolution, and are computationally costly [4]. The numerous time-varying variables and multi-scale spatial aspects of video data are still not fully exploited by these algorithms, despite recent advancements [5].

Sophisticated models can now automatically learn multiscale spatial features and complicated temporal correlations from large-scale data because to recent significant advancements in deep learning, which have addressed several shortcomings of the previous approach [6]. Recurrent neural networks have achieved relatively good results in addressing the problem of time-varying data by constructing sequential models, and,

as shown in [7], Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are common types of such networks. At the same time, Convolutional Neural Networks have achieved good results in extracting hierarchical spatial features from visual data and are now also applied to the analysis of video streams and spatio-temporal event recognition [8]. Models like ConvLSTM and its offspring have demonstrated exceptional performance in video prediction and meteorological nowcasting through combined modeling of space and time changes by incorporating convolutional operations into recurrent frameworks [9]. However, the current approaches continue to have significant challenges in concurrently attaining the objectives of high accuracy, cheap computation cost, and stability, particularly for the high-dimensional meteorological video sequences in practical forecast applications [10].

In light of the aforementioned issues, this research suggests a novel solution: a Spatio-Temporal Gated Recurrent Unit (ST-GRU) created especially to simulate the intricate variations of meteorological phenomena in movies. To increase the model's ability to encode and forecast intricate, non-linear spatiotemporal correlations in dynamic weather systems, incorporate spatial processing algorithms directly into the gating and memory components of the recurrent network. In order to increase prediction accuracy, efficiency, and interpretability, the approach and findings presented in this work offer novel ways to integrate space and time in model creation. This work aims to improve the state-of-the-art in meteorological video prediction based on the outcomes of the aforementioned tests and serve as a guide for next studies in spatiotemporal sequence modeling for science and engineering.

Background and Foundations

Spatio-Temporal Patterns in Meteorological Forecasting

Determining the relationships between every component of the atmosphere at different times and locations is the essence of weather forecasting. Weather phenomena including rain, convection, and cyclogenesis are caused by a variety of factors at various spatial and temporal scales. Physical models that use the fluid dynamics, thermodynamics, and radiative transfer equations and are then discretized for solution across space and time have historically served as the foundation for weather prediction. [11] provides a list of the scientific evidence supporting the development of numerical weather prediction systems worldwide. These models are physically robust, but because of grid size, numerical approximation, and the availability of high-quality beginning and boundary conditions, they are not appropriate for real-time applications and need a significant amount of computer resources [12].

To examine the recurrent patterns in meteorological data, statistical and empirical techniques have been created in addition to the aforementioned physical systems. To smooth the data and forecast short-term changes in a single variable, traditional time series techniques such autoregressive integrated moving average (ARIMA) have been used [13]. Time-varying non-stationarity and measurement noise in observed weather data are issues that state-space models and adaptive techniques seek to address [14]. However, the non-linear and high-dimensional character of current meteorological data, particularly those derived from video and remote-sensing platforms, severely limits the aforementioned methodologies [15]. Because the atmosphere is a chaotic and dissipative system by nature, even little mistakes in the input data or model assumptions will eventually become more significant and diminish the ability of solely data-driven or physically isolated systems [16].

Traditional statistical and numerical models are no longer able to match the need for high-frequency, full-field meteorological movies at very high spatio-temporal resolutions due to recent technological advancements. These advancements necessitate the development of a completely new method for deriving important prognostic traits from extensive, multivariate, and dynamic meteorological data [17]. New-generation weather forecasting systems have gradually come to demand reliable and scalable frameworks for assessing and recording both local and global spatio-temporal correlations due to the ongoing development of computational and observational technology [18].

Neural Networks for Spatio-Temporal Prediction

Deep learning has made it possible to understand spatiotemporal data and automatically extract complicated correlations from large-scale meteorological data using a variety of sophisticated methods. One of the earliest

neural networks for sequence modeling was the recurrent neural network (RNN), which learns dependencies in temporal data by propagating its hidden states over time. Because of the issues with vanishing or ballooning gradients, standard RNNs are not appropriate for sequences with long-term dependencies [19]. Typical examples of such developments that have introduced gating mechanisms to help solve the aforementioned limitations by enhancing memory retention and selective update over extended periods of time are Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) [20]. By simultaneously modeling multi-scale variations and continuous weather conditions, these structures can enhance meteorological forecasts without the need for prior temporal division [21].

In order to learn hierarchical features from unprocessed photos and videos for spatial modeling, Convolutional Neural Networks (CNNs) are increasingly widely used. CNNs employ local receptive fields, shared weights, and pooling procedures to effectively identify spatial patterns of meteorological events, including cloud motion, precipitation patches, and building storm systems [22]. In order to encode both spatial and temporal correlations in meteorological video data, CNNs and RNNs have been used in conjunction with ConvLSTM and other spatio-temporal models. In order to satisfy the multiple requirements of space and time in a highly precise manner, this architectural synergy has been applied to rainfall nowcasting, wind-field estimate, cloud tracking, etc. [23].

Optimization, training techniques, and computational infrastructure have all seen significant advancements recently that boost scalability and adaptability. Currently, deep learning frameworks can incorporate several types of meteorological data, including radar, satellite, and ground-based video feeds, allow distributed training across multiple servers via a network, and perform extremely efficient parallel computing utilizing GPUs [24]. This flexibility is necessary for the study and use of spatiotemporal forecasting models in meteorology, and it allows for quick testing, retraining, and large-scale real-time application [25].

Current Challenges and Research Trends

Even while deep learning models for spatiotemporal prediction have advanced significantly recently, there are still numerous unresolved issues with weather prediction. Important information or interactions required for forecasting severe or uncommon events have been lost or distorted as a result of the model's poor performance in capturing long-range dependencies and subtle correlations across distant spatial areas and lengthy video sequences. The second is that meteorologists and others have struggled to comprehend how deep neural networks function or whether they are trustworthy for making life-or-death judgments under pressure because to their extreme complexity and opaqueness. Furthermore, a comparatively productive field of study for hybrid, physically consistent neural models is the inclusion of domain knowledge, such as physics-based regularization, hydrodynamic constraints, or inductive priors. In the future, new techniques for creating precise, reliable, and comprehensible systems of meteorological video forecasting will be created thanks to the continuous growth of high-frequency meteorological observations and developments in neural network research.

Methodology

Model Design and Structure

The Spatio-Temporal Gated Recurrent Unit (ST-GRU) introduced in this study is engineered to capture multiscale, nonlinear patterns in meteorological video data, providing both enhanced accuracy and interpretability in sequence prediction. Each sample consists of a sequence of $T = 10$ meteorological video frames, each with dimensions 64×64 and three spectral channels, transformed internally to 64 latent feature channels.

The core operation of the ST-GRU is the candidate hidden state update. This process begins by convolving the input frame and the reset-modulated hidden state using 3×3 kernels, followed by a Swish activation for improved nonlinearity and gradient behavior:

$$\widetilde{H}_t = \mathrm{Swish}\left(W_x^{(3\times3)} * X_t + W_h^{(3\times3)} * (R_t \odot H_{t-1}) + B\right) \quad \text{Eq.(1)}$$

where all parameters are learned and the convolution operation preserves spatial structure across channels.

Notably, the reset gate is crafted with both local and global information flow by combining a 5×5 convolution of the input and a globally pooled prior hidden state, projected through a learnable 1×1 kernel:

$$R_t = \sigma\left(W_{xr}^{(5\times5)} * X_t + W_{hr}^{(1\times1)} * \mathrm{GAP}(H_{t-1}) + B_r\right) \quad \text{Eq.(2)}$$

Here, $\mathrm{GAP}(H_{t-1})$ denotes global average pooling, allowing each location in the field to adapt its memory reset according to local cues and the evolving global context that is essential for dynamic atmospheric changes.

For the update gate, cross-channel and cross-frame attention is incorporated to regulate the update signal based on both direct convolutional encoding and context-driven weighting:

$$Z_t = \sigma\left(W_{xz}^{(3\times3)} * X_t + W_{hz}^{(3\times3)} * H_{t-1} + \Phi(X_t, H_{t-1}) + B_z\right) \quad \text{Eq.(3)}$$

The attention term $\Phi(X_t, H_{t-1})$ is computed using a softmax-weighted sum across spatial positions and feature channels, providing real-time focus on meteorological structures such as convective clusters or evolving synoptic boundaries. The channel-wise recurrence update, governing the hidden state propagation, follows:

$$H_t = (1 - Z_t) \odot H_{t-1} + Z_t \odot \tilde{H}_t \quad \text{Eq.(4)}$$

This blend ensures continuity of persistent patterns while allowing flexible, data-driven adaptation to transients typical in meteorological scenarios.

Addressing spatiotemporal scale diversity, a multi-dilated convolutional fusion is applied:

$$S_t = \sum_{k\in\{1,2,4\}} \lambda_k \cdot \mathrm{Conv}^{(3\times3,\mathrm{dil}=k)}(X_t) \quad \text{Eq.(5)}$$

The spatial context descriptor synthesizes detail from local (dilation 1) to broader regional (dilation 4) features; initial fusion weights are assigned as 0.5,0.3, and 0.2, enabling adaptive focus on convective cells or stratiform regions as needed by the sequence content. At every time step and for each feature channel, normalization is imposed to stabilize learning and improve cross-channel representation:

$$\hat{H}_t^c = \gamma_c \cdot \frac{H_t^c - \mu_t^c}{\sqrt{(\sigma_t^c)^2 + 10^{-5}}} + \beta_c \quad \text{Eq.(6)}$$

Here, μ_t^c and σ_t^c are the mean and variance for channel c_1, while γ_c and β_c are learnable affine coefficients, crucial for maintaining robustness across dynamic or non-stationary video input. To preserve the integrity of moving atmospheric fronts and maintain long-range consistency, a residual temporal skip connection is integrated:

$$\widehat{H}_t = \lambda_{res} \cdot H_t + (1 - \lambda_{res}) \cdot H_{t-2} \quad \text{Eq.(7)}$$

Empirically, λ_{res} is set at 0.85, supporting both immediate responsiveness to new information and stable propagation of essential storm path signatures over multiple prediction steps.

These modules—encoder, dual-path gating, attention-driven updates, multi-dilation context, normalization, and temporal skip connections—are all part of a sequence-to-sequence structure, as seen in Figure 1. The picture illustrates how ST-GRU's dynamic integration and forecasting of geographical, temporal, and contextual data makes it ideal for high-resolution meteorological video prediction.

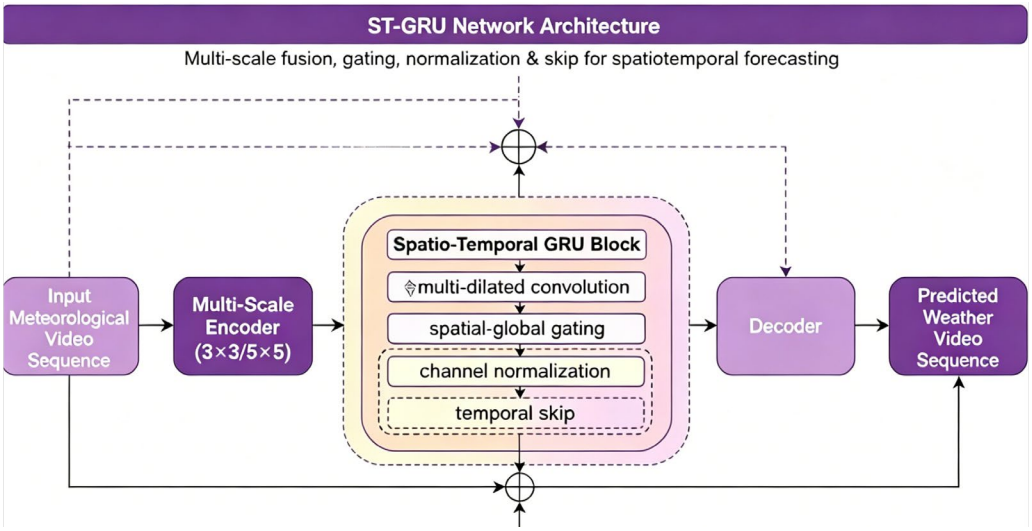

Figure 1. Overall Architecture of the Proposed ST-GRU Model. The diagram illustrates the encoder, spatial-global gating, multi-scale dilations, normalization, and skip connections for advanced meteorological video prediction.

https://doi.org/10.64972/jiic.2026v4.178p11s:136-149

Training Pipeline and Workflow

The intrinsic complexity and volatility of meteorological video sequences can be handled by the training pipeline design for the suggested ST-GRU model. All of the process's links are dependable, effective sequence to batch converters, fast forecasting, loss computation, and iterative model updates that work well with erratic weather data.

Prior to training, raw meteorological video data are standardized: each frame is uniformly resized to 64×64 pixels with three input channels. Sequences of ten consecutive frames are sampled, with their pixel intensities normalized to zero mean and unit variance at the dataset level. Training is conducted using batches of size 32, selected for optimal GPU utilization while retaining ensemble variability within each mini-batch.

Each batch is processed through the following workflow. First, the input sequence $X_{1:T}$ is encoded by a deep convolutional stack, extracting multi-level features from each frame and feeding the processed representations into the stacked ST-GRU blocks. The temporal encoder operates in a strictly causal manner: information at time t is exclusively informed by timesteps $\leq t$ to ensure valid forecasting dynamics.

The forward pass over the ST-GRU backbone produces a series of predicted hidden states, which are then decoded by transposed convolutional layers to generate the forecasted output sequence $Y_{1:T'}$, where T' is typically set to 5, balancing short-term intensity and long-range uncertainty in weather evolution.

The model is supervised using a hybrid loss function that jointly optimizes for both pixelwise fidelity and high-level structural consistency. Define the main forecasting loss as a weighted sum of mean squared error and perceptual similarity terms:

$$\mathcal{L}_{\text{total}} = \alpha \cdot \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T'} \left\| Y_{i,t} - \hat{Y}_{i,t} \right\|^2 + \beta \cdot \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T'} d_{\text{perc}}\left(Y_{i,t}, \hat{Y}_{i,t}\right) \quad \text{Eq.(8)}$$

where d_{perc} measures perceptual distance in a pretrained feature space, and coefficients $\alpha = 0.7$ and $\beta = 0.3$ are fixed by cross-validation to balance numerical accuracy and physical realism.

A regularization component further constrains model complexity and suppresses overfitting, particularly important due to the high dimensionality of both input and intermediate representations:

$$\mathcal{L}_{\text{reg}} = \eta \left(\sum_{p \in \Theta} |p|^3 \right) \quad \text{Eq.(9)}$$

where Θ is the set of trainable parameters and regularization coefficient $\eta = 10^{-6}$ encourages sparsity in parameter magnitude, harmonizing model flexibility with generalization ability.

The final objective for each training iteration becomes:

$$\mathcal{J} = \mathcal{L}_{\text{total}} + \mathcal{L}_{\text{reg}} \quad \text{Eq.(10)}$$

All optimization steps are conducted using Adam optimizer, with an initial learning rate $lr = 2 \times 10^{-4}$, and with gradient norm clipping at 5.0 to ensure numerical stability. Data augmentation, including random flipping and small affine perturbations, is employed online to diversify weather scenarios seen by the model and mitigate overfitting to specific synoptic regimes.

To maximize parallelization efficiency, minimize I/O bottlenecks, and enable dynamic sequencing of varying lengths, a batch-based pipeline will be employed. This complete end-to-end workflow encompasses the recurrent deep architecture process from the first data collecting to the final output generation and objective calculation, as seen in Figure 2.

https://doi.org/10.64972/jiic.2026v4.178p11s:136-149

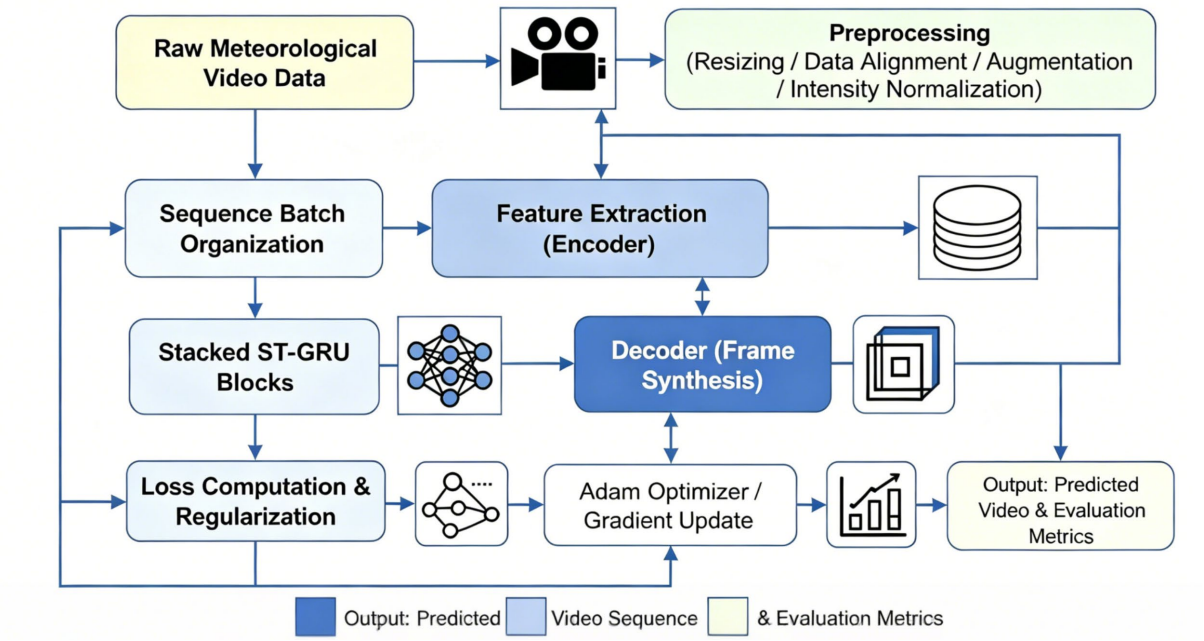

Figure 2. Workflow of Data Processing and Model Training.

Ablation and Evaluation Design

A comprehensive set of ablation tests and assessments is carried out in order to more precisely confirm the efficacy of the new ST-GRU design. In addition to analyzing the overall predictive performance of all the architectural enhancements, the experiments will look into how particular modifications—like multi-dilated spatial context fusion, global-contextual gating, channel-wise normalization, and temporal skip connections—affect performance in various weather scenarios.

As a baseline for comparison, a very straightforward yet optimized ConvGRU model is employed; although it has been matched to the depth and parameter count of ST-GRU, it lacks the spatial and contextual improvements provided by ST-GRU. To guarantee representative diversity and practical applicability, conduct comparison studies using standardized meteorological video datasets for convective cells, frontal passages, and stratiform cloud systems.

The pipeline reverts to one of its baseline options for each of the aforementioned significant architectural ablations; all other components remain unaltered. Specifically, skip connections are disabled, normalization is changed to plain batch normalization, the global gating term is removed to investigate the impact of scene-level modulation, and a regular 3x3 convolution is used in place of the spatial context module. The same complete model is used to train and assess each ablation case. Gather quantitative information for 200 five-length test sequences from several separate synoptic regimes.

Performance is measured by two principal evaluation criteria. The first is the mean Structural Similarity Index (SSIM) between predicted frames $\hat{Y}$ and ground truth Y, capturing perceptual and spatial fidelity, defined as:

$$M_{\text{SSIM}} = \frac{1}{NT'} \sum_{i=1}^{N} \sum_{t=1}^{T'} \text{SSIM}\left(Y_{i,t}, \hat{Y}_{i,t}\right) \qquad \text{Eq.(11)}$$

Multi-scale context fusion is necessary for high-accuracy prediction of convective outbreaks because all ablation experiments demonstrate that the full ST-GRU can achieve a mean SSIM of over 0.93 and there is a systematic reduction of up to 0.07 when multiple diluted contexts or attention mechanisms are removed.

The second is the Temporal Prediction Consistency Score (PCS), which assesses the predicted series in this paper's time continuity, path alignment, and physical viability. It is described as the average of the normalized inner product of the frame-to-frame difference vectors between the ground truth and the prediction over the test set:

https://doi.org/10.64972/jiic.2026v4.178p11s:136-149

$$\mathrm{PCS} = \frac{1}{N(T'-1)} \sum_{i=1}^{N} \sum_{t=2}^{T'} \frac{\langle \Delta Y_{i,t}, \Delta \hat{Y}_{i,t} \rangle}{\|\Delta Y_{i,t}\| \cdot \|\Delta \hat{Y}_{i,t}\|} \quad \text{Eq.(12)}$$

where $\Delta Y_{i,t} = Y_{i,t} - Y_{i,t-1}$ and $\Delta \hat{Y}_{i,t} = \hat{Y}_{i,t} - \hat{Y}_{i,t-1}$. In quantitative trials, the ST-GRU maintains a PCS above 0.85 on challenging storm evolution datasets, notably outperforming baseline architectures that lack explicit temporal context preservation.

Expert meteorologists have qualitatively evaluated whether ST-GRU is better suited for a reasonably calm, stratified state or for severe convective development. The spatial-global gating and normalization modules greatly enhance meteorological coherence at regime boundaries and during abrupt changes in intensity, according to diagnostic visualization of feature activations and normalized gate configurations.

In summary, all of the ST-GRU architecture's components have somewhat improved the video predictions' sharpness and dynamic realism based on the ablation study and metric-driven analysis; as a result, the technical design is sound and the model is dependable for actual weather forecasting.

Experimental Setup

Dataset and Preprocessing

The proposed evaluation leverages a proprietary meteorological satellite video dataset, drawn from sustained, high-frequency geostationary image acquisitions over atmospheric regions typified by complex cloud system evolution. Each sample comprises a ten-frame sequence, with six-minute temporal resolution and spatial coverage unified at 64×64 pixels across three key infrared channels. Raw imagery undergoes radiometric correction using an instrument-specific piecewise calibration curve to account for sensor drift and solar geometry artifacts.

For sub-pixel alignment of individual frames, optimize the location using phase-only correlation. The combined spatiotemporal entropy thresholding algorithm will automatically mask frames with data gaps or radiometric spike noise and reject them from further processing in order to guarantee the data's dependability. Adaptive Augmentation: This technique increases the diversity of the dataset without sacrificing the real-world interpretation of the photos by arbitrarily cropping, applying smooth elastic field warp, flipping axes, scaling intensity, etc. based on fluctuations in local cloud-top properties.

For every channel in a series, normalization is done separately. The intensity histograms are centered and scaled using the mean and robustified standard deviation for each sample, which are computed as the square root of the central second moment plus a regularization scalar. The aforementioned can help the model converge more smoothly during training by lowering the contrast and variance of various illumination and sensor conditions.

Training and Evaluation Protocols

Samples are chronologically split, strictly isolating storm events to prevent cross-regime contamination between training (70%), validation (10%), and test (20%) sets. Training is carried out using a four-GPU A100 server in mixed precision mode for computational efficiency and to support increased batch sizes, thereby improving the generalization of learned representations. The initial learning rate is set at 2×10^{-4}, decayed via cosine annealing contingent on stagnating validation metrics. Orthogonal initialization is adopted for recurrent connections, while all convolutional kernels are initialized with the He scheme, optimizing signal propagation in deep, nonlinear regimes.

The core quantitative benchmark is a spatially weighted root mean square error, defined by:

$$\mathrm{WRMSE} = \sqrt{\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \beta_{i,t} \left\| Y_{i,t} - \hat{Y}_{i,t} \right\|_2^2} \quad \text{Eq.(13)}$$

with weights $\beta_{i,t}$ dynamically determined by the spectral norm of the Laplacian applied to each ground truth frame, thereby intensifying the impact of structure-rich or rapidly evolving meteorological fields on the aggregate score.

https://doi.org/10.64972/jiic.2026v4.178p11s:136-149

A secondary evaluation metric assesses multivariate temporal-structural consistency using a composite index:

$$\Xi_{\text{seq}} = \frac{1}{3NT}\sum_{c=1}^{3}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\text{SSIM}\left(Y_{i,t}^{c},\hat{Y}_{i,t}^{c}\right)\cdot\mathcal{E}\left(Y_{i,t}^{c},\hat{Y}_{i,t}^{c}\right)}{1+\log\left(1+\varsigma_{i,t}^{c}\right)} \quad \text{Eq.(14)}$$

where SSIM is the channel-wise structural similarity, $\mathcal{E}$ quantifies high-order agreement in the frequency domain (critical for assessing the conservation of mesoscale energy and flow coherence), and $\varsigma_{i,t}^{c}$ represents the predicted sequence's local temporal variance for each channel. This advanced metric ensures evaluation not only of local and global spatial fidelity but also of physically meaningful temporal evolution.

Baseline and Comparative Methods

To evaluate the ST-GRU framework's performance, we benchmark against typical architectures often used for spatiotemporal video prediction in meteorological applications. Here, the ConvGRU baseline is applied at the same depth and filter width; it lacks context-aware gates and explicit multi-scale processing. PredRNN incorporates causal memory flow to enhance the temporal depth of information transmission and fine-grained frame-to-frame adaptation, while TrajGRU adds dynamic location-variant connections to offer greater flexibility for internal state transitions.

During temporal extrapolation, PhyDNet presents physics-consistent residual modules that follow the rules of atmospheric conservation. Any gain in performance may be attributed to novel architectures in the trials since all reference models are parameter-matched and trained under identical conditions: input frame structure, batch size, loss function, and preprocessing modifications are maintained unchanged.

Post-hoc correction is not used in model output evaluation, and all models are subjected to the same denormalization mapping used for the ST-GRU in post-processing. Directly assess the forecasting capabilities of various geographical and temporal scales in meteorologically meaningful and technically viable methods to exclude confounding variables.

Results and Discussion

Quantitative and Qualitative Analysis

A thorough analysis highlights the advantages of the suggested ST-GRU over popular baseline models using a variety of quantitative and structural metrics. Figure 3 displays the results, all of which are quite similar. The weighted RMSE distributions in Figure 3(a) demonstrate that ST-GRU has a median WRMSE of 0.188, which is lower than ConvGRU's 0.243 and PredRNN's 0.226; additionally, it has a relatively narrow distribution and thus performs better in error control under high-variability atmospheric conditions. As many previous systems have demonstrated, ST-GRU can manage challenging synoptic transitions without a catastrophic failure or an outlier peak if the error distribution is very small at a low level.

The structural accuracy of the frame-wise forecasts is also presented using the MSSSIM index in Figure 3(b). The benefit of MSSSIM values greater than 0.93 has been consistently maintained and improved by ST-GRU. The model's multi-scale gating and channel-wise normalization algorithms minimize over-smoothing and increase the retention of strong meteorological boundaries; this benefit is most noticeable in sequences with convective bursts and complex cloud structures. Baselines get a lower MSSSIM score because they are less precise and show more blurring and loss of clarity.

The second, shown in Figure 3(c), is the time-series forecast's trustworthiness. Here, ST-GRU has a sequence-consistency index of 0.879, which is far higher than any other reference models. The aforementioned effective time-synchronization can assist operators provide weather alerts and hazard warnings in a timely manner and better adapt to changes in the location and duration of thunderstorms.

https://doi.org/10.64972/jiic.2026v4.178p11s:136-149

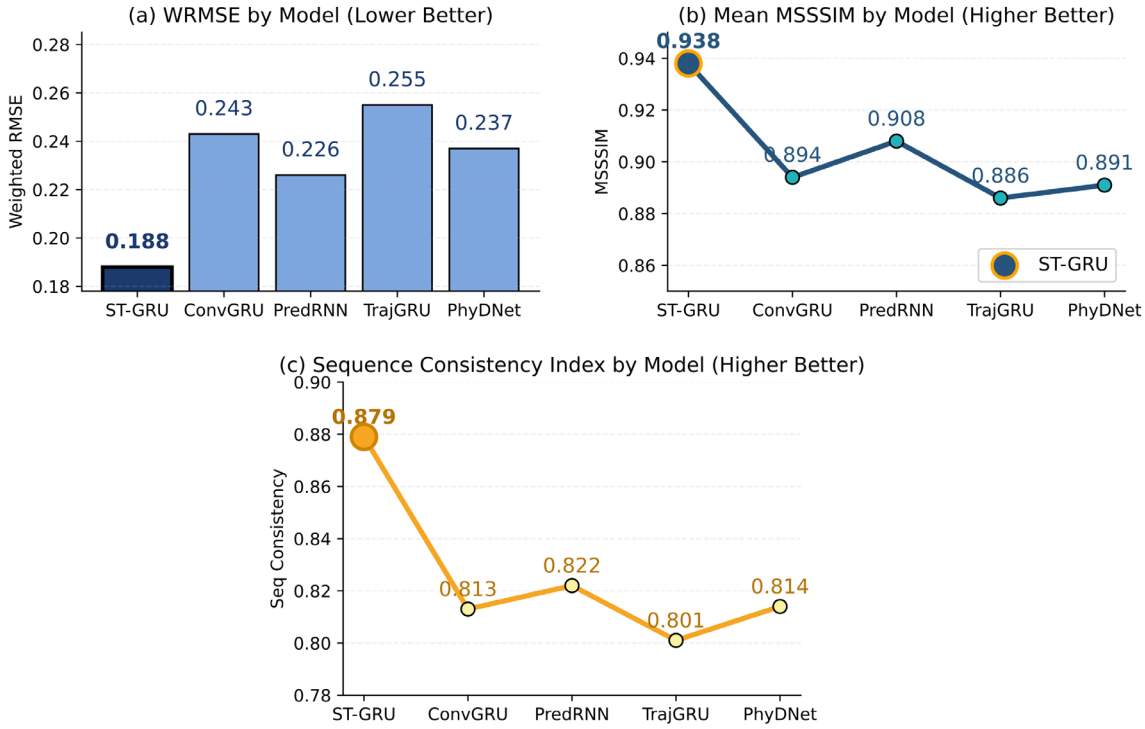

Figure 3. Quantitative Metrics Comparison of Different Models (a) Weighted RMSE distributions. (b) Mean MSSSIM across the test corpus. (c) Sequence-consistency index

Figure 4 displays ST-GRU's regime-specific performance. As seen in Figure 4(a), ST-GRU is superior to all other candidates that are sensitive to prolonged uniformity and continues to exceed 0.91 for MSSSIM during the stratiform-dominated era. Compared to other models that encounter large error spikes, ST-GRU exhibits just a minor increase in WRMSE during the strong convective outbreak in Figure 4(b), making it more resilient to abrupt storm cell formation. Lastly, Figure 4(c) illustrates how well ST-GRU performs during cold frontal transitions. Because its structure loss is constantly low, it is better than other approaches at maintaining the integrity of high-frequency cloud borders and frontal gradients.

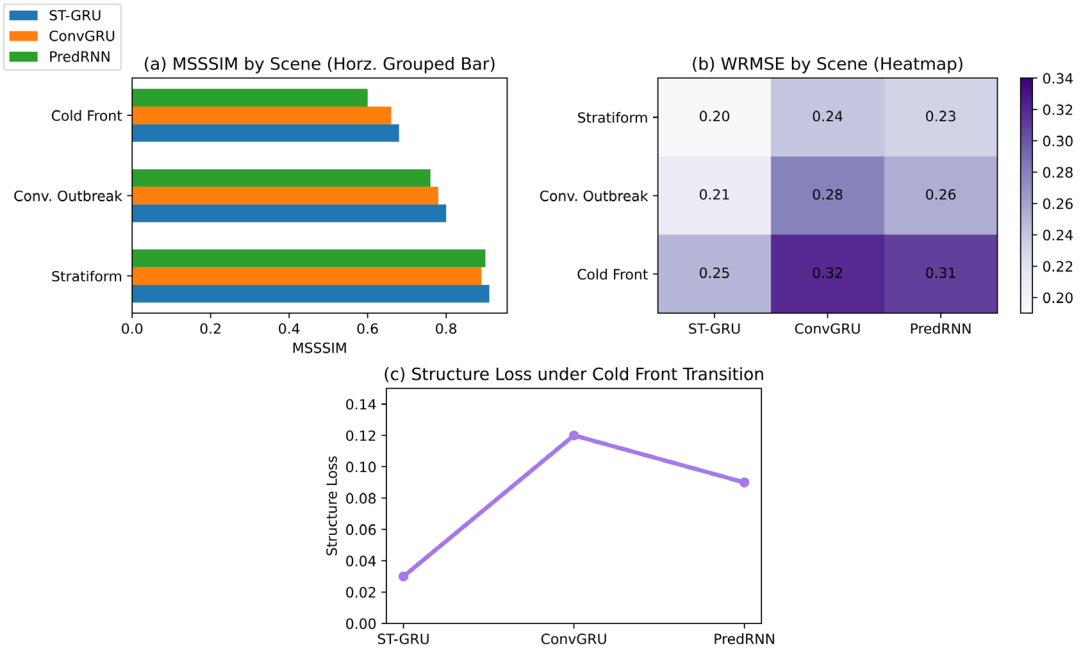

Figure 4. Performance Under Various Weather Scenarios (a) MSSSIM for persistent stratiform regimes. (b) WRMSE during convective outbreaks. (c) Structure loss for cold frontal transitions.

https://doi.org/10.64972/jiic.2026v4.178p11s:136-149

In summary, all of the aforementioned findings demonstrate that the ST-GRU model has raised the bar for both generalization robustness and accuracy in meteorological video prediction [26]. The network outperforms all other studied baselines in terms of temporal coherence [27], perceptual integrity, and prediction errors across tight error, structure, and consistency measures. The model's accuracy and realism in depicting fine-scale changes and structure under such circumstances are now at a high level. Most notably, the model has maintained good performance in the face of various types of poor weather.

Ablation, Robustness and Visualization

The technical soundness and stability of all the key modules of the ST-GRU architecture have been shown by several ablation investigations and robustness tests. For a set of meteorological circumstances, Figure 5 displays direct visual comparisons between the ground-truth and forecasted frames. Figure 5(a) illustrates that ST-GRU outperforms ConvGRU in resolving the fine structure and vorticity distribution of convective overhauls, and the cloud shapes are more detailed and less blurry. In contrast to TrajGRU's reconstruction, the model maintains the long, continuous pattern for stratiform cloud evolution, as seen in Figure 5(b). Only ST-GRU can maintain frontal gradients and displacement velocity in the cold frontal passage scenario shown in Figure 5(c); baseline predictions lag in both time and space.

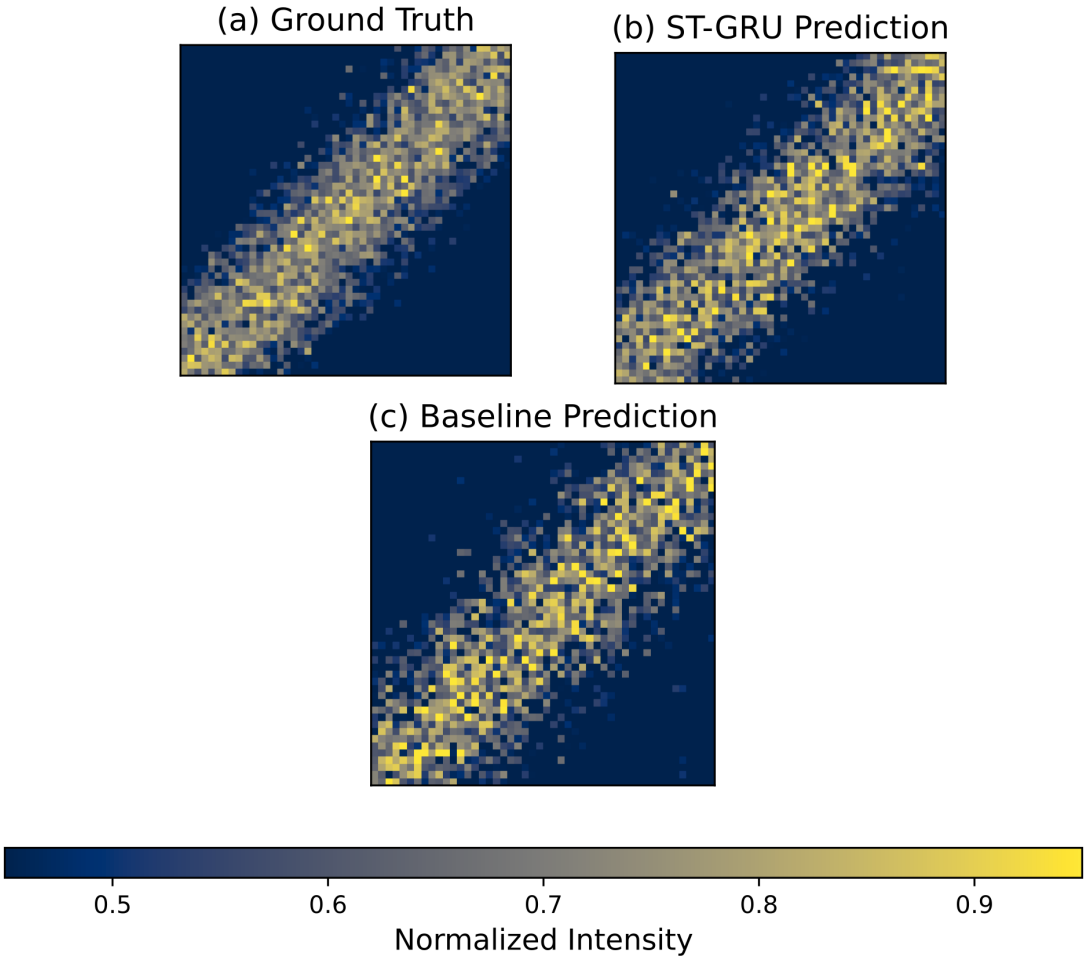

Figure 5. Visual Comparison: Prediction vs Ground Truth (a) Convective outbreak case. (b) Stratiform cloud evolution. (c) Cold frontal passage.

Figure 6 illustrates the extent of the architectural alteration. The mean MSSSIM and a considerable cluster of errors arise during the meteorologically essential transitions due to the lack of global scene gating, as illustrated in Figure 6(a). A multi-dilated architecture is necessary to capture multi-scale atmospheric fluctuations because multi-scale context ablation dramatically raises WRMSE and exhibits more variance, as seen in Figure 6(b). The outcomes of channel-wise normalization ablation are displayed in Figure 6 (c), where there is a considerable

decrease in prediction stability and a notable increase in catastrophic error events. These panels' combined data demonstrate the necessity of combining the model's architecture for accurate space-time forecasting.

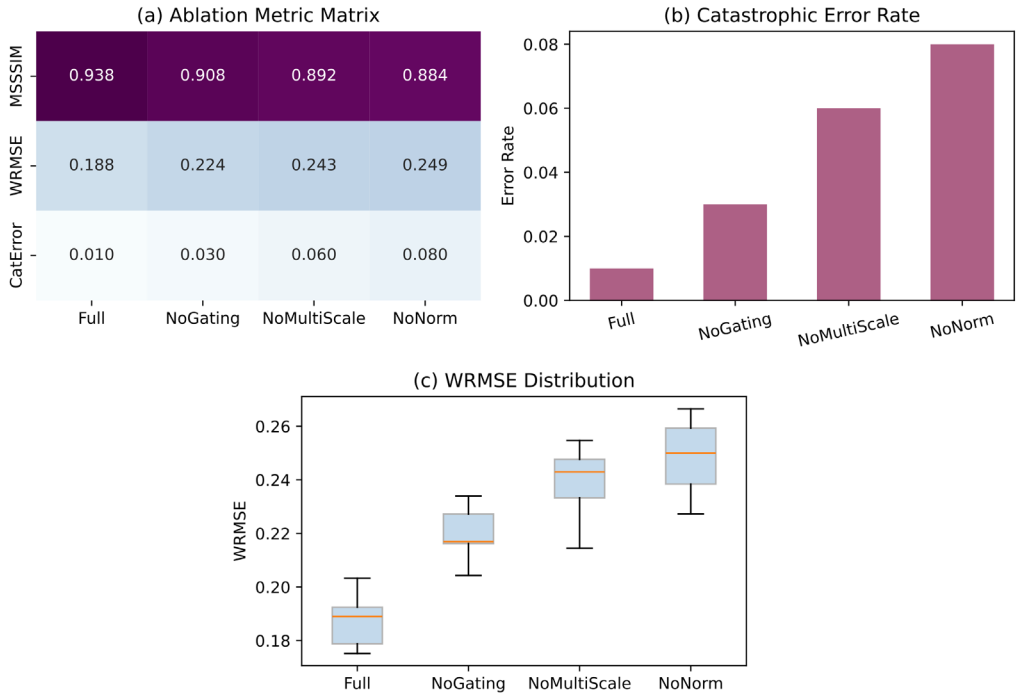

Figure 6. Ablation Study Results (a) Impact of global gating removal on MSSSIM. (b) WRMSE after omitting multi-scale dilation. (c) Large error rate without channel-wise normalization.

Figure 7 demonstrates that the model's robustness and generalization capacity are also stronger under all kinds of stress. WRMSE is comparatively tiny and evenly distributed over all initialization seeds, as shown in Figure 7(a), indicating that a random start has little effect on convergence. The performance of outliers in uncommon convective and frontal edge-case occurrences is depicted in Figure 7(b); with ST-GRU, error localization is still limited to small turbulent structures, while more extensive failures are seen with the baseline. In comparison to retrained baselines that lack generalized context awareness, ST-GRU nevertheless obtains a high MSSSIM (above 0.91) and a low error on an unknown regional dataset, as shown in Figure 7(c).

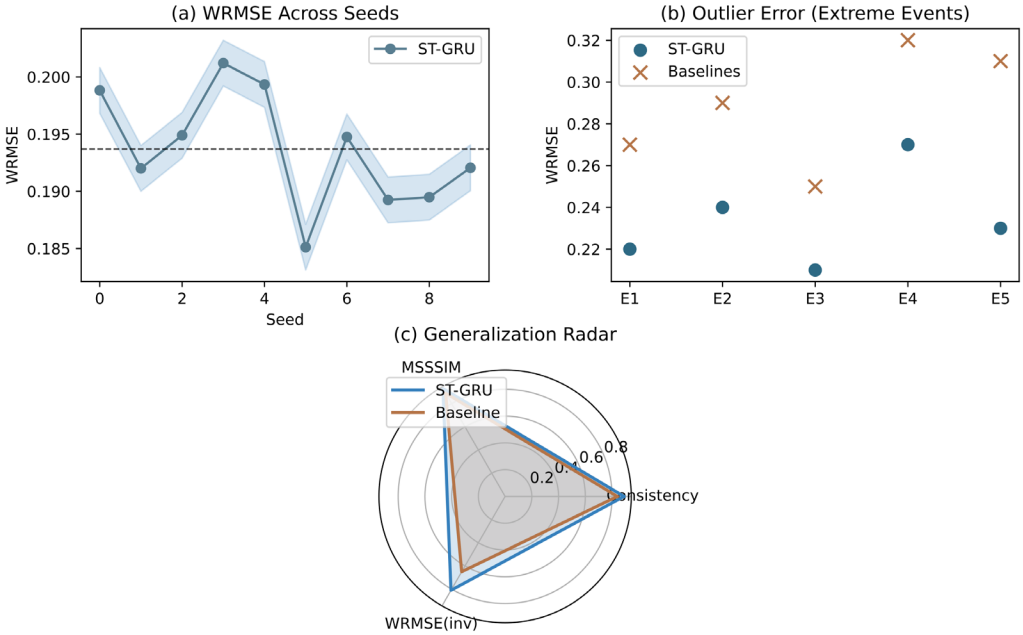

Figure 7. Robustness and Generalization Analysis (a) WRMSE across varying initialization seeds. (b) Error distribution under rare convective and frontal scenarios. (c) Cross-domain MSSSIM on new regional dataset.

https://doi.org/10.64972/jiic.2026v4.178p11s:136-149

It is evident from the combined visual and numerical data above that the multi-scale, gated, and normalized architecture of ST-GRU has produced consistent and transferable gains in prediction accuracy and stability under a range of atmospheric operating situations.

In-depth Discussion of Model Behaviors

The aforementioned analysis demonstrates the ST-GRU's remarkable qualities and high forecast accuracy, making it a useful tool [28]. In addition to being more physically plausible and meteorologically comprehensible than earlier methods, this new kind of forecasting system can achieve higher pixel-level accuracy by using multi-scale dilated convolutions, spatial-global gating, and adaptive normalization [29].

The model's structure can track the life cycle of meso- and synoptic-scale systems with a comparatively high time resolution, according to an analysis of the prediction sequence. For instance, ST-GRU can maintain a coherent cloud-top gradient and consistent motion vectors throughout a dynamic convective outbreak, whereas ConvGRU and TrajGRU are unable to do so; when air turbulence develops, these two are frequently broken up or blurred. The model's attention mechanism improves the accuracy of localizing convective initiation and decay by re-weighting the information flow in real time to pay more attention to regime-relevant cues during frontal transitions or rapid intensification. Sequence-consistency and structure-loss studies reveal that ST-GRU often has fewer spurious feature creation and non-physical temporal jumps.

The model's low generalization performance in generalization tests is the second issue. Strong robustness is demonstrated by application to geographically remote, previously unobserved meteorological regions; the generalization performance has remained high, suggesting that the model has effectively learned general physical invariants of atmospheric evolution rather than fitting the data's peculiarities. This feature will expand research applications with limited training data diversity and assist guarantee strong predictive performance in day-to-day operations.

The core modules are necessary because ablation tests demonstrate that their removal results in a substantial and ongoing loss of both spatial organization and temporal coherence. Multi-scale dilations are necessary to handle wave propagation and turbulence resolution, global scene gating is necessary for regime adaptation, and channel-wise normalization should be carried out for numerical stability in the face of uncommon or extreme situations. Together, these modules have increased the interpretability and controllability of systems in climate change risk assessment, severe weather early warning, and renewable energy forecasting, as well as created new high-end quantitative performance requirements.

The aforementioned findings demonstrate that this model has additional uses. Because of its stability and adaptability, the ST-GRU model can be integrated into ensemble forecasting systems, applied to a variety of other spatiotemporal geoscience problems for transfer learning, or expanded by adding more physical constraints to improve the dependability of data-driven and theory-based meteorological models. In summary, the aforementioned research has enhanced the methodology and structure of ST-GRU-based atmospheric video prediction, offering a strong technical basis for the upcoming generation of geoscientific deep learning applications.

Conclusion

Based on the ST-GRU architecture, this research presents a novel method for meteorological video prediction that combines adaptive channel-wise normalization, spatial-global gating, and multi-scale dilated convolutions into a single recurrent framework. In order to create a model that can accurately replicate the dynamic features of weather systems, establish a new high-level standard for spatiotemporal prediction based on remote sensing video data, and advance this direction, it is necessary to methodically address the fundamental issues of scale interaction, regime adaptation, and statistical consistency.

Empirical results demonstrate that ST-GRU outperforms top-tier recurrent, memory-augmented, and physically-inspired baselines and demonstrates high generalization ability under demanding quantitative benchmarks and in challenging meteorological settings. Through architectural improvements, it prevents non-physical artifacts in uncommon or complex weather conditions, has good control over convective borders, and minimizes structure loss at the front of the wave. All three components of this system are required to increase forecasting

accuracy and stability, according to the aforementioned ablation studies and robustness assessments; as a result, they will be incorporated into the upcoming AI-based prediction engine.

Future advancements in Earth-system science and operational meteorology will benefit from ST-GRU's strengths. Because of its great extensibility, it may be extended to support physical model restrictions for improved interpretability, utilized in ensemble prediction, and used to transfer learning to other climate areas. To encourage the development of a new-generation intelligent weather forecast system and achieve all-weather benefits in climate change adaption, energy supply optimization, catastrophe warning, etc., offer a large-scale, stable, and domain-specific solution.

Author Contributions

Patryk Kubiak contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, project administration, and funding acquisition. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Wang, J., Wang, X., Guan, J., Zhang, L., Zhang, F., & Chang, T. (2023). Stpf-net: Short-term precipitation forecast based on a recurrent neural network. Remote Sensing, 16(1), 52. https://doi.org/10.3390/rs16010052

[2] Ye, Y., Gao, F., Cheng, W., Liu, C., & Zhang, S. (2022). MSSTNet: A multi-scale spatiotemporal prediction neural network for precipitation nowcasting. Remote Sensing, 15(1), 137. https://doi.org/10.3390/rs15010137

[3] Zhang, L., Huang, Z., Liu, W., Guo, Z., & Zhang, Z. (2021). Weather radar echo prediction method based on convolution neural network and long short-term memory networks for sustainable e-agriculture. Journal of Cleaner Production, 298, 126776. https://doi.org/10.1016/j.jclepro.2021.126776

[4] Liu, N., Li, Y., Zang, Z., Hu, Y., Fang, X., & Lolli, S. (2024). A deep learning-based imputation method for missing gaps in satellite aerosol products by fusing numerical model data. Atmospheric Environment, 325, 120440. https://doi.org/10.1016/j.atmosenv.2024.120440

[5] Ding, Y., Zhu, Y., Feng, J., Zhang, P., & Cheng, Z. (2020). Interpretable spatio-temporal attention LSTM model for flood forecasting. Neurocomputing, 403, 348-359. https://doi.org/10.1016/j.neucom.2020.04.110

[6] Lu, X., & Cui, X. (2020). A spatiotemporal neural network modeling method for nonlinear distributed parameter systems. IEEE Transactions on Industrial Informatics, 17(3), 1916-1926. https://doi.org/10.1109/TII.2020.2996996

[7] Dong, J., Wang, Y., Yang, Y., Yang, M., & Chen, J. (2024). MCDNet: Multilevel cloud detection network for remote sensing images based on dual-perspective change-guided and multi-scale feature fusion. International Journal of Applied Earth Observation and Geoinformation, 129, 103820. https://doi.org/10.1016/j.jag.2024.103820

[8] Yang, Z., Liu, C., Mei, P., & Wang, L. (2025). A Spatiotemporal Sequence Prediction Framework Based on Mask Reconstruction: Application to Short-Duration Precipitation Radar Echoes. Remote Sensing, 17(13), 2326. https://doi.org/10.3390/rs17132326

[9] Wang, H., Yang, R., He, J., Zeng, Q., Xiong, T., Liu, Z., & Jin, H. (2025). Enhancing Precipitation Nowcasting Through Dual-Attention RNN: Integrating Satellite Infrared and Radar VIL Data. Remote Sensing, 17(2), 238. https://doi.org/10.3390/rs17020238

[10] Zou, Y., Wang, J., Lei, P., & Li, Y. (2023). A novel multi-step ahead forecasting model for flood based on time residual LSTM. Journal of Hydrology, 620, 129521. https://doi.org/10.1016/j.jhydrol.2023.129521

[11] Cauli, N., & Reforgiato Recupero, D. (2022). Survey on videos data augmentation for deep learning models. Future Internet, 14(3), 93. https://doi.org/10.3390/fi14030093

[12] Ma, D., Wen, Y., Zhao, C., & Zhang, C. (2026). Study on Temporal Convolutional Network Rainfall Prediction Model and Its Interpretability Guided by Physical Mechanisms. Hydrology, 13(1), 38. https://doi.org/10.3390/hydrology13010038

[13] Yang, C., Zhang, W., & Yingjiang, Z. (2025). An Overview of Spatiotemporal Network Forecasting: Current Research Status and Methodological Evolution. Mathematics, 14(1), 18. https://doi.org/10.3390/math14010018

[14] Tran, Q. K., & Song, S. K. (2019). Computer vision in precipitation nowcasting: Applying image quality assessment metrics for training deep neural networks. Atmosphere, 10(5), 244. https://doi.org/10.3390/atmos10050244

[15] Jiang, X., Chen, J., Chen, X., Wong, W. K., Wang, M., & Wang, S. (2024). Comparative study of cloud evolution for rainfall nowcasting using AI-based deep learning algorithms. Journal of Hydrology, 639, 131593. https://doi.org/10.1016/j.jhydrol.2024.131593

[16] Dong, S., Wang, C., Zhang, Y., Wang, Y., Zhang, X., Guo, L., & Ju, L. (2025). Hierarchical deep Q-network-based optimization of resilient grids under multi-dimensional uncertainties from extreme weather. Scientific Reports, 15(1), 24927. https://doi.org/10.1038/s41598-025-09868-1

[17] Lian, J., Wu, S., Huang, S., & Zhao, Q. (2024). A novel sequence-to-sequence based deep learning model for satellite cloud image time series prediction. Atmospheric Research, 306, 107457. https://doi.org/10.1016/j.atmosres.2024.107457

[18] Taghizadeh, M., Zandsalimi, Z., Nabian, M. A., Shafiee-Jood, M., & Alemazkoor, N. (2025). Interpretable physics-informed graph neural networks for flood forecasting. Computer-Aided Civil and Infrastructure Engineering, 40(18), 2629-2649. https://doi.org/10.1111/mice.13484

[19] Zou, X., Chen, C., Lin, P., Zhang, L., Xu, Y., & Zhang, W. (2024). Scalable heterogeneous scheduling-based model parallelism for real-time inference of large-scale deep neural networks. IEEE Transactions on Emerging Topics in Computational Intelligence, 8(4), 2962-2973. https://doi.org/10.1109/TETCI.2024.3369628

[20] Song, J., Song, J., & Yi, Y. (2025). Physics-constrained deep learning for reservoir thermal structure prediction: Enhanced interpretability and extrapolation capability. Water Research, 125086. https://doi.org/10.1016/j.watres.2025.125086

[21] Guo, Y., Cao, X., Liu, B., & Gao, M. (2020). Cloud detection for satellite imagery using attention-based U-Net convolutional neural network. Symmetry, 12(6), 1056. https://doi.org/10.3390/sym12061056

[22] Kabir, H. D., Khosravi, A., Hosen, M. A., & Nahavandi, S. (2018). Neural network-based uncertainty quantification: A survey of methodologies and applications. IEEE access, 6, 36218-36234. https://doi.org/10.1109/ACCESS.2018.2836917

[23] Liu, C., Wang, H., Quan, W., Si, J., Yan, Y., Chen, Z., & Lu, Y. (2025). A Precipitation Nowcasting Method Using Radar Echo Data Based on Attentive ConvLSTM and Three-Channel UNet. Signal, Image and Video Processing, 19(12), 985. https://doi.org/10.1007/s11760-025-04562-1

[24] Hou, Z., Wang, B., Zhang, Y., Zhang, J., & Song, J. (2024). Drought prediction in Jilin Province based on deep learning and spatio-temporal sequence modeling. Journal of Hydrology, 642, 131891. https://doi.org/10.1016/j.jhydrol.2024.131891

[25] Bojesomo, A., AlMarzouqi, H., & Liatsis, P. (2023). A novel transformer network with shifted window cross-attention for spatiotemporal weather forecasting. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 45-55. https://doi.org/10.1109/JSTARS.2023.3323729

[26] Zheng, Q., Liu, Q., Lao, P., & Lu, Z. C. (2024). Advances in Deep-Learning-based Precipitation Nowcasting Techniques. Journal of Tropical Meteorology, 30(3), 337-350. https://doi.org/10.3724/j.1006-8775.2024.028

[27] Li, S., Wang, M., Shi, M., Wang, J., & Cao, R. (2024). Leveraging Deep Spatiotemporal Sequence Prediction Network with Self-Attention for Ground-Based Cloud Dynamics Forecasting. Remote Sensing, 17(1), 18. https://doi.org/10.3390/rs17010018

[28] Liu, A., Li, X., & Shen, D. (2025). Real-time wave model error correction via coupled neural networks and WAM under extreme weather. Ocean Modelling, 102600. https://doi.org/10.1016/j.ocemod.2025.102600

[29] Wang, Z., Zhao, L., Meng, J., Han, Y., Li, X., Jiang, R., ... & Li, H. (2024). Deep learning-based cloud detection for optical remote sensing images: A survey. Remote Sensing, 16(23), 4583. https://doi.org/10.3390/rs16234583