Journal of Science and Technology
ISSN: 2456-5660 Volume 10, Issue 10 (Oct -2025)
www.jst.org.in
DOI: https://doi.org/10.46243/jst.2025.v10.i10.pp01-09

Attention-Enhanced Sequential GAN for Reliable Groundwater Recharge Time-Series Augmentation

Uma Kannan[1] | Rajendran Swamidurai[2]

[1]Department of Mathematics and Computer Science, Alabama State University, AL, USA
[2]Department of Mathematics and Computer Science, Alabama State University, AL, USA

To Cite this Article
Uma Kannan and Rajendran Swamidurai, "Attention-Enhanced Sequential GAN for Reliable Groundwater Recharge Time-Series Augmentation", *Journal of Science and Technology*, Vol. 10, Issue 10, October 2025, pp.-01-09xxxxxxxx.

Article Info
Received on 02-Aug-2025, Revised on 22-Sep-2025, Accepted on 27-Oct-2025.

ABSTRACT

Groundwater recharge modeling is critically hindered by the scarcity of long-term, high-resolution time-series data, limiting the robustness and generalization capability of predictive models. We propose the Attention-enhanced Sequential Generative Adversarial Network ($S\text{-}GAN_{Attn}$) to synthesize high-fidelity, multivariate hydrological records, explicitly addressing the complex temporal dependencies required for groundwater dynamics. The architecture incorporates three key innovations: stabilization via the WGAN-GP objective for continuous learning; utilization of a pre-trained LSTM autoencoder to establish a meaningful latent space; and integration of a Self-Attention mechanism within the generative networks to effectively capture critical long-range dependencies, such as multi-year climatic cycles. A three-pronged evaluation demonstrated exceptional data quality: Statistical Fidelity confirmed the preservation of feature relationships, and Temporal Coherence validated the realism of sequential patterns. Crucially, the Predictive Utility was confirmed, with an auxiliary forecasting model trained on synthetic data achieving a Mean Absolute Error (MAE only 4.4% higher) than a model trained on real data. This provides a stable and effective generative approach for time-series augmentation, offering a viable path to developing reliable forecasting tools in data-scarce hydrological contexts.

KEYWORDS: *Generative Adversarial Networks (GANs); Attention Mechanism; Deep Learning; Adversarial Training; Self-Attention Mechanism; Deep Generative Models; Semi-Supervised Learning; Context-Aware Generation.*

I. INTRODUCTION

Groundwater is a vital resource, sustaining ecosystems, agriculture, and municipal water supplies globally [1-4]. Groundwater recharge—the process by which water infiltrates the ground and adds to the underlying aquifer—is the cornerstone of sustainable groundwater management. Accurately quantifying groundwater recharge rates is crucial for sustainable resource management, as it sets pumping limits to prevent aquifer depletion and related issues like saltwater intrusion and land subsidence [5]. It is also essential for contaminant transport prediction, as recharge drives the movement of pollutants into the aquifer [6,7]. Furthermore, recharge modeling aids in climate change adaptation by projecting aquifer responses to altered precipitation, and it is vital for balancing competing demands in the Water-Energy-Food Nexus [8-10].

Despite its importance, the accurate modeling of groundwater recharge is frequently hampered by a fundamental obstacle: *data scarcity,* particularly in developing countries and arid/semi-arid regions [11-15]. Addressing the data scarcity challenge requires leveraging remote sensing technologies, integrating data from diverse sources (e.g., citizen science, agricultural reports), and adopting machine learning approaches to interpolate and predict missing data [16,17].

Journal of Science and Technology
ISSN: 2456-5660 Volume 10, Issue 10 (Oct -2025)
www.jst.org.in
DOI: https://doi.org/10.46243/jst.2025.v10.i10.pp01-09

The sustainable management of groundwater resources hinges on accurate estimates of groundwater recharge, a complex, non-linear hydrological process driven by numerous factors including rainfall, temperature, and land use. However, reliable modeling is fundamentally limited by the scarcity of long-term, high-resolution time-series data [18]. Many regions vital to water security are classified as data-scarce catchments, where measurement records are often too short or sporadic to capture the full range of hydrological variability (e.g., multi-year drought cycles). Consequently, standard time-series forecasting methods trained on limited records often suffer from poor generalization capabilities. This challenge necessitates novel approaches capable of augmenting existing records with synthetic, statistically representative data.

Generative Adversarial Networks (GANs) [19] offer a powerful mechanism for synthesizing data by modeling the underlying data generating process itself. A GAN framework—composed of a Generator and a Discriminator—can learn the complex, implicit distribution of the original hydrological data, enabling the creation of synthetic time series that maintain high statistical fidelity. While GANs have shown success in various domains, their direct application to sequential, multivariate data faces critical challenges, primarily training instability and the failure to accurately preserve temporal coherence (i.e., the realistic sequencing of events over time).

The foundation for time-series GANs, such as TimeGAN [20], has demonstrated that leveraging an autoencoder framework can stabilize generation by embedding the data into a meaningful latent space. However, existing models still struggle with long-range dependencies—the ability to relate current recharge levels to climatic events that occurred many months or years in the past—a crucial requirement for accurate groundwater dynamics modeling.

To address the limitations of data scarcity and the inherent challenges of time-series GANs, we introduce the Attention-Enhanced Sequential GAN ($S\text{-}GAN_{Attn}$) designed specifically for multivariate, time-dependent hydrological data. Our architectural enhancements focus on stability, temporal structure, and long-range learning: 1) Training Instability This is addressed through the robust Wasserstein GAN with Gradient Penalty (WGAN-GP) [21]. By operating with the Wasserstein distance in the latent space, we ensure the Discriminator provides stable, continuous gradient feedback to the Generator, mitigating mode collapse, 2) Temporal Structure Preservation: This is achieved via a pre-trained LSTM-based Embedder/Recovery autoencoder framework [20]. This phase creates a stable, low-dimensional latent space that already contains the essential temporal features before adversarial training begins, and 3) Long-Range Dependency: This is mitigated by integrating a Self-Attention mechanism [22] into the generative and encoding networks. This allows the model to prioritize critical historical events, such as long-term drought or intense multi-year rainfall, dynamically determining which time steps are most relevant for predicting the sequence's next state.

II. METHODOLOGY

The Attention-enhanced Sequential GAN ($S\text{-}GAN_{Attn}$) is a sequence-to-sequence deep generative model designed to learn the underlying distribution of multivariate time series. The architecture is composed of four primary networks: the *Embedder (E)*, the *Recovery (R)*, the *Generator (G)*, and the *Discriminator (D)*. These networks are optimized through a three-stage training process: pre-training of the autoencoder, adversarial stabilization using WGAN-GP, and final multi-objective adversarial optimization.

2.1. Sequential Autoencoder and Attention Mechanism

The foundational stability of the architecture is established by the sequential autoencoder, formed by the Embedder and Recovery networks. This framework achieves crucial dimensionality reduction and extracts a stable, low-dimensional latent representation of the hydrological data.

2.1.1. Embedder (E)

The Embedder network acts as a time-series encoder. It is an LSTM-based network that maps the high-dimensional real input sequence X to the feature-rich latent space H. The input sequence is defined as $\mathbf{X} \in \mathbb{R}^{T \times F}$, where T is the sequence length and F is the number of features. The latent sequence H maintains the temporal length but reduces the feature dimension to H_{dim}.

Journal of Science and Technology
ISSN: 2456-5660 Volume 10, Issue 10 (Oct -2025)
www.jst.org.in
DOI: https://doi.org/10.46243/jst.2025.v10.i10.pp01-09

$$E: \mathbf{X} \rightarrow \mathbf{H}, \quad \text{where } \mathbf{H} \in \mathbb{R}^{T \times H_{\text{dim}}}$$

2.1.2. Recovery (R)

The Recovery network functions as the decoder, reconstructing the original sequence $\mathbf{X}'$ from the compressed latent representation H.

$$R: \mathbf{H} \rightarrow \mathbf{X}' \quad \text{where } \mathbf{X}' \in \mathbb{R}^{T \times F}$$

During the pre-training phase, E and R are jointly trained to minimize the reconstruction error ($\mathcal{L}_{\text{Rec}}$), ensuring that the latent representation H preserves the maximum information required for accurate recovery:

$$\mathcal{L}_{\text{Rec}} = \min_{E,R} \mathbb{E}[||\mathbf{X} - R(E(\mathbf{X}))||_2^2]$$

2.1.3. Self-Attention Integration

A critical enhancement is the inclusion of a Self-Attention layer [22] immediately following the primary LSTM unit in both the Embedder (E) and the Generator (G). This attention mechanism mitigates the inherent limitation of LSTMs in capturing long-range dependencies, which are vital for modeling groundwater recharge (e.g., relating current aquifer status to rainfall that occurred over 18 months prior). The attention layer dynamically weights the importance of all time steps in the input sequence, allowing the model to focus on the most predictive historical events regardless of their distance from the current time step.

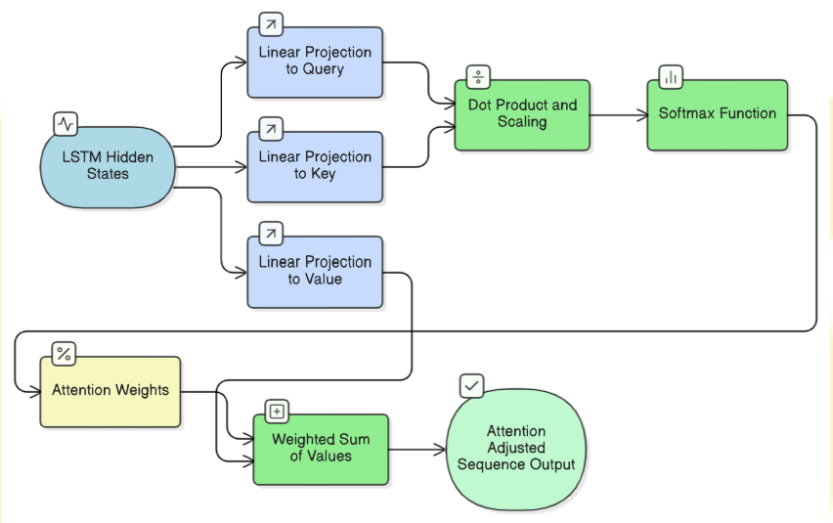

Figure-1. Detailed View of the Self-Attention Mechanism within Recurrent Networks.

Figure-1 explains the core mechanism that allows the proposed model to capture long-range dependencies by dynamically assigning weights to historical time steps. The self-attention layer dynamically computes the relevance of all historical time steps (t') within the input sequence (H) relative to the current time step (t). This mechanism generates attention weights that highlight the most critical historical events (e.g., distant drought periods) for prediction, effectively resolving the long-range dependency problem.

2.2. Adversarial Networks and WGAN-GP Stabilization

The adversarial stage trains the Generator (G) and the Discriminator (D) within the compressed latent space (H), leveraging the stability of the pre-trained Embedder.

2.2.1. Generator (G)

The Generator transforms a random noise vector, $\mathbf{Z} \in \mathbb{R}^{T \times Z_{\text{dim}}}$, into a synthetic latent sequence $\mathbf{H}_{\text{fake}}$. G also includes the *Self-Attention layer* to enforce realistic sequential dynamics in its output.

$$G: \mathbf{Z} \rightarrow \mathbf{H}_{\text{fake}}, \quad \text{where } \mathbf{H}_{\text{fake}} \in \mathbb{R}^{T \times H_{\text{dim}}}$$

2.2.2. Discriminator (D)

The Discriminator is an LSTM-based network tasked with distinguishing the real embedded sequence ($\mathbf{H}_{\text{real}} = E(\mathbf{X})$) from the synthetic sequence ($\mathbf{H}_{\text{fake}}$), returning a scalar score for "realness." By operating in the low-dimensional latent space, the Discriminator's task is stabilized, enhancing the quality of the gradients passed back to G.

$$D: \mathbf{H} \rightarrow \mathbb{R}$$

2.2.3. WGAN-GP Stabilization

To prevent mode collapse and training instabilities, the adversarial objective employs the Wasserstein GAN with Gradient Penalty (WGAN-GP) [21]. This

Journal of Science and Technology
ISSN: 2456-5660 Volume 10, Issue 10 (Oct -2025)
www.jst.org.in
DOI: https://doi.org/10.46243/jst.2025.v10.i10.pp01-09

method replaces the standard Jensen-Shannon divergence loss with the Wasserstein distance (Earth Mover's distance), which provides a more robust and informative gradient. The Discriminator loss ($\mathcal{L}_{\text{Adv}}^{D}$) is formulated to enforce the *1-Lipschitz constraint* via the *Gradient Penalty* ($\mathcal{L}_{GP}$) term, where λ is the penalty weight (typically 10.0):

$$\mathcal{L}_{\text{Adv}}^{D} = \mathbb{E}_{\mathbf{H}_{\text{fake}}}[D(\mathbf{H}_{\text{fake}})] - \mathbb{E}_{\mathbf{H}_{\text{real}}}[D(\mathbf{H}_{\text{real}})] + \lambda \mathcal{L}_{\text{GP}}$$

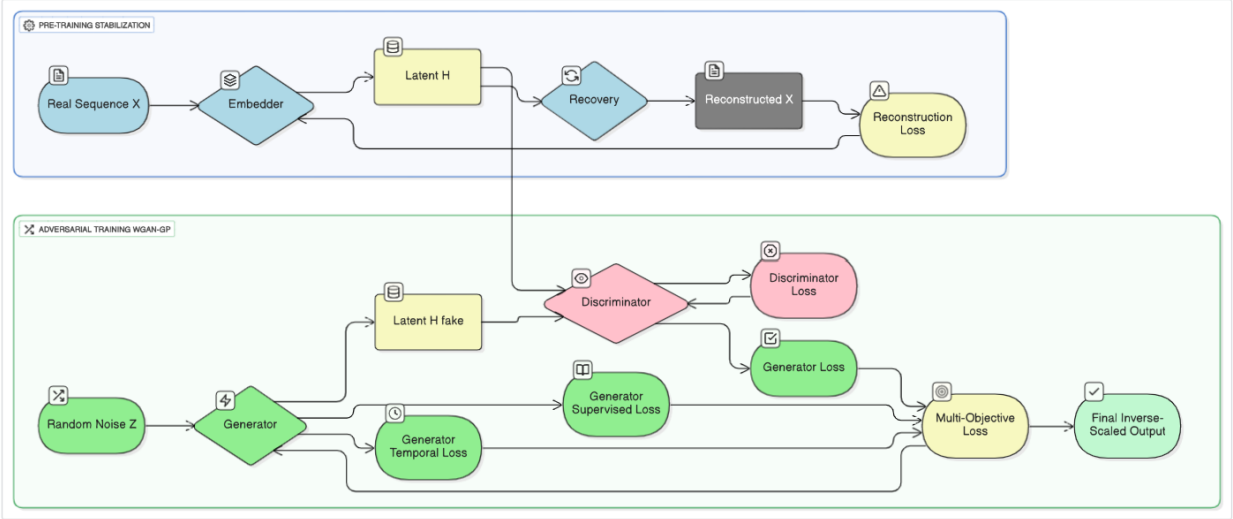

Figure-2. Architecture of the Attention-enhanced Sequential Generative Adversarial Network($\text{S-GAN}_{\text{Attn}}$).

Figure-2 illustrates the four primary networks and the data flow in the pre-training and adversarial phases. The Embedder (E) and Recovery (R) form the autoencoder used for initial stabilization. The Generator (G) and Discriminator (D) operate exclusively in the low-dimensional latent space (H). The Self-Attention modules are integrated into both E and G to enhance long-range temporal feature extraction.

2.3. Multi-Objective Generator Loss

The Generator is trained using a composite loss function ($\mathcal{L}_{\text{Gen}}$) that balances three critical objectives, ensuring the synthetic data is not only indistinguishable from the real data but also temporally meaningful and structurally compatible.

1) Adversarial Loss **($\mathcal{L}_{\text{Adv}}^{G}$)**: Drives the Generator to maximize the Discriminator's output score for the synthetic data, thereby encouraging statistical realism:

$$\mathcal{L}_{\text{Adv}}^{G} = -\mathbb{E}_{\mathbf{H}_{\text{fake}}}[D(\mathbf{H}_{\text{fake}})]$$

2) Supervised Reconstruction Loss **($\mathcal{L}_{\text{Rec}}$)**: This term ensures that the synthetic latent space ($\mathbf{H}_{\text{fake}}$) is structurally compatible with the Embedder's output, preventing the Generator from wandering into meaningless regions of the latent space. It is weighted by the hyperparameter γ:

$$\mathcal{L}_{\text{Rec}} = \mathbb{E}_{\mathbf{H}_{\text{fake}}}[||\mathbf{X}' - R(\mathbf{H}_{\text{fake}})||_1]$$

3) Temporal Coherence Loss **($\mathcal{L}_{\text{Temp}}$)**: This term directly enforces the temporal realism of the sequence dynamics. It minimizes the difference between the structure of the real latent sequence ($\mathbf{H}_{\text{real}}$) and the synthetic latent sequence ($\mathbf{H}_{\text{fake}}$) across all time steps, weighted by the hyperparameter β:

$$\mathcal{L}_{\text{Temp}} = \mathbb{E}_{\mathbf{H}_{\text{real}},\mathbf{H}_{\text{fake}}}[||\mathbf{H}_{\text{real}} - \mathbf{H}_{\text{fake}}||_1]$$

The final Generator objective is the minimization of this weighted sum:

$$\mathcal{L}_{\text{Gen}} = \min_{G}\left(\mathcal{L}_{\text{Adv}}^{G} + \gamma \mathcal{L}_{\text{Rec}} + \beta \mathcal{L}_{\text{Temp}}\right)$$

The coefficients γ and β are critical hyperparameters used to balance the trade-off between statistical fidelity, structural compatibility, and temporal realism during training.

Journal of Science and Technology
ISSN: 2456-5660 Volume 10, Issue 10 (Oct -2025)
www.jst.org.in
DOI: https://doi.org/10.46243/jst.2025.v10.i10.pp01-09

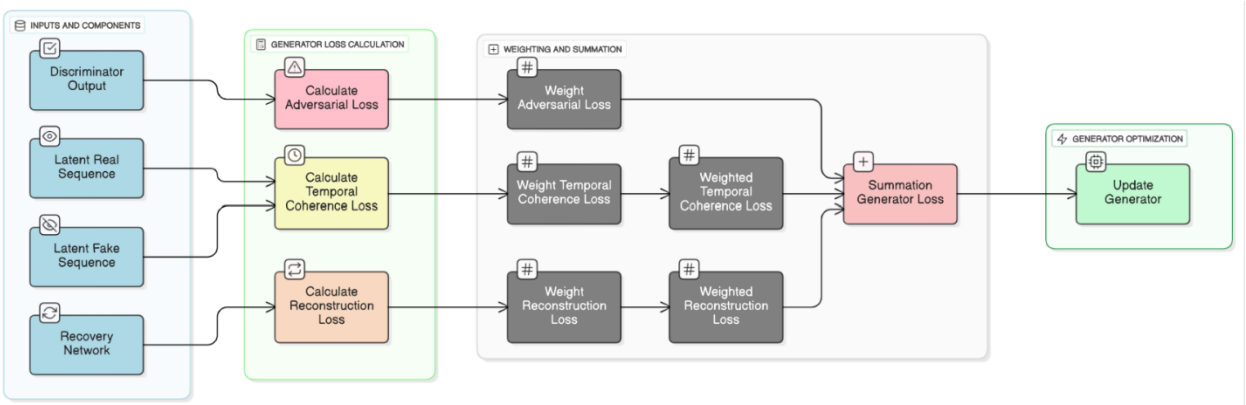

Figure-3. The Three-Component Loss Function Guiding the Generator ($\mathcal{L}_{\text{Gen}}$)

Figure-3 illustrates how the three distinct loss components—Adversarial, Reconstruction, and Temporal Coherence—are derived from different parts of the network and combined using the weighting hyperparameters (γ and β) to form the final Generator objective. That is, the Generator is optimized using a weighted combination of three distinct loss terms, balancing statistical realism, structural compatibility, and temporal coherence.

III. Experimental Setup and Evaluation

This section details the preprocessing steps applied to the multivariate hydrological time series, the specific implementation environment, and the comprehensive methodology used to evaluate the fidelity and utility of the synthetic groundwater recharge data generated by the $S\text{-}GAN_{Attn}$. The real data utilized in the experiment was gathered from Wolf Bay, located in Baldwin County, Alabama, within the Gulf of Mexico, spanning the years 2000 to 2016. The dataset encompasses 8 input independent variables: Digital Elevation Model, Land Use Land Cover, Soil Type, Precipitation, Temperature, Windspeed, Relative Humidity, and Solar Radiation and 3 output dependent variables: Stream Discharge, Groundwater Levels, and Groundwater Recharge.

3.1. Data Preprocessing and Preparation

Effective data preprocessing is essential to ensure the stability and convergence of the $\text{S-GAN}_{\text{Attn}}$ architecture.

3.1.1. Data Scaling

The raw, multivariate time-series data X (including features such as precipitation, temperature, and target recharge rates) were subjected to Min-Max Scaling across the entire dataset range. This scales all features into the [0,1] interval, which is crucial for the stability of the LSTM networks and the Tanh activation functions used in the Generator. The scaler object was preserved for inverse-transformation of the synthetic output.

3.1.2. Sequence Windowing

The scaled data was transformed from a two-dimensional format (Time × Features) into the three-dimensional tensor (Samples × Sequence Length × Features) required by the recurrent networks using a sliding window technique. A fixed Sequence Length (T) of 24 months was chosen. This length captures two full annual cycles, providing the model with adequate context for learning long-term dependencies. This windowing results in the input tensor $X \in \mathbb{R}^{N \times T \times F}$.

3.1.3. Training and Testing Split

The original real dataset was split into 80:20 ratio. The majority portion (80%) was used to train the $\text{S-GAN}_{\text{Attn}}$, while a separate, temporally contiguous portion (20%) was reserved exclusively as the real test set for the final Predictive Utility evaluation.

3.2. Implementation Details

The $\text{S-GAN}_{\text{Attn}}$ was implemented using the TensorFlow/Keras framework. Core Code Snippets such as Attention-Enhanced Component (Embedder or Generator), Simplified WGAN-GP Generator Loss Function, and Time-Series Windowing Function are given in Appendix B. Key hyperparameters (detailed in Appendix A) included a latent dimension (H_{dim}) of 16, a sequence length (T) of 24, and a fixed WGAN-GP penalty weight (λ) of 10.0. The model was trained using a Critic Ratio of 5, meaning the Discriminator was updated five times for every single update of the Generator. The

Journal of Science and Technology
ISSN: 2456-5660 Volume 10, Issue 10 (Oct -2025)
www.jst.org.in
DOI: https://doi.org/10.46243/jst.2025.v10.i10.pp01-09

training proceeded in two stages: 1) *Pre-training:* The Embedder (E) and Recovery (R) were trained for 200 epochs to minimize $\mathcal{L}_{Rec}$, and 2) *Adversarial Training:* All networks were trained for 40 epochs using the multi-objective loss function ($\mathcal{L}_{Gen}$) with empirical weights γ=5.0 and β=1.0.

3.3. Evaluation Methodology

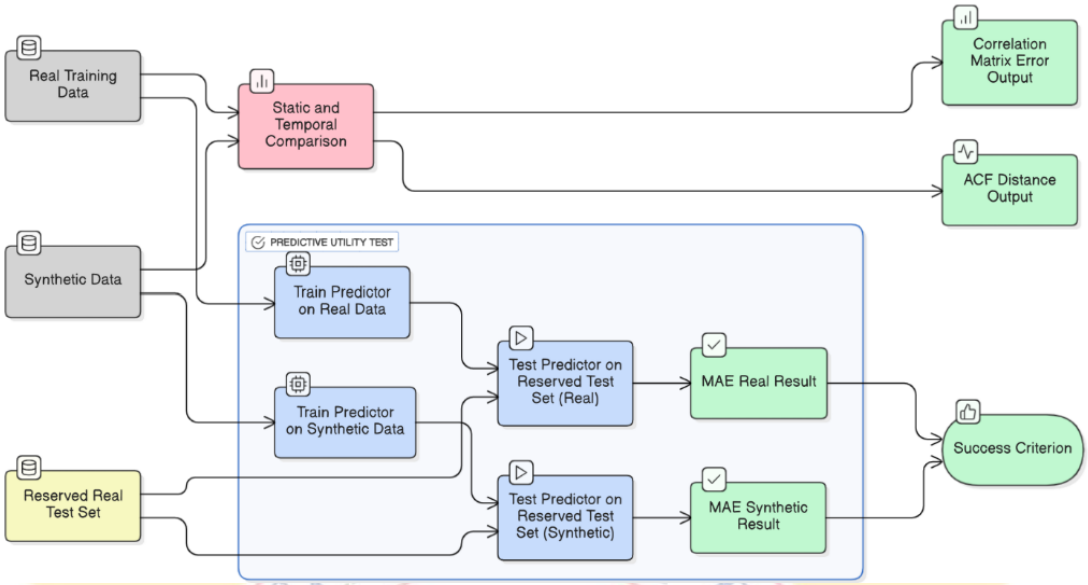

Figure-4: Three-Pronged Evaluation Methodology for Synthetic Data Quality.

Figure-4 outlines the three-pronged evaluation methodology, focusing on how the generated synthetic data is rigorously validated against three distinct metrics: Statistical Fidelity (CME), Temporal Coherence (ACF Distance), and the critical Predictive Utility test, which compares the performance of a predictor trained on real vs. synthetic data using the same reserved test set.

The quality of the generated synthetic time series was assessed using a *three-pronged evaluation methodology* that measures distinct aspects of the data: Statistical Fidelity, Temporal Coherence, and Predictive Utility.

3.3.1. Statistical Fidelity (Feature Similarity)

This metric assesses whether the relationships between different features in the synthetic data ($\mathbf{X}_{fake}$) align with the real data ($\mathbf{X}_{real}$).

- *Metric: Correlation Matrix Error (CME).* We compute the Pearson correlation matrix for both $\mathbf{X}_{real}$ and $\mathbf{X}_{fake}$. The fidelity is quantified by measuring the Mean Absolute Error between the corresponding elements of the two matrices. A lower CME indicates a better preservation of cross-feature relationships.

3.3.2. Temporal Coherence (Sequence Dynamics)

This metric evaluates the realism of sequential dependencies and periodic behavior, directly validating the effectiveness of the Attention mechanism.

- *Metric: Autocorrelation Function (ACF) Distance.* We calculate the ACF for the real and synthetic recharge series, and the coherence is quantified by calculating the Mean Absolute Difference between the two ACF plots.

3.3.3. Predictive Utility (Downstream Task Performance)

The ultimate test for the synthetic data is its utility in training a model for the final forecasting task—groundwater recharge prediction.

- *Procedure:* An auxiliary prediction model (a simple one-layer LSTM Predictor) is trained on the real training data to establish the MAE_{Real} baseline. An identical model is then trained exclusively on the $\text{S-GAN}_{\text{Attn}}$'s synthetic data ($\mathbf{X}_{fake}$), yielding the $\text{MAE}_{\text{Synthetic}}$.

Success Criterion: The synthetic data is deemed successful if $\text{MAE}_{\text{Synthetic}}$ is approximately equal to or only marginally higher than MAE_{Real}.

Journal of Science and Technology
ISSN: 2456-5660 Volume 10, Issue 10 (Oct -2025)
www.jst.org.in
DOI: https://doi.org/10.46243/jst.2025.v10.i10.pp01-09

IV. RESULTS AND DISCUSSION

The evaluation confirms that the *Attention enhanced Sequential GAN* ($S\text{-}GAN_{Attn}$) successfully generates synthetic groundwater recharge data that maintains both the statistical integrity and the predictive utility of the real data.

4.1. Evaluation of Data Fidelity and Temporal Coherence

The first phase of the evaluation confirmed that the generated synthetic time series, $\mathbf{X}_{fake}$, closely matched the characteristics of the real training data, $\mathbf{X}_{real}$, across both static feature distributions and dynamic temporal structures.

4.1.1. Statistical Fidelity

We measured the difference between the Pearson correlation matrices of the real and synthetic datasets using the *Correlation Matrix Error (CME)*. As shown in Table 1, the $S\text{-}GAN_{Attn}$ achieved a negligible CME, indicating that the complex covariance relationships between multivariate inputs (e.g., the inverse relationship between temperature and recharge) were accurately preserved.

Table-1. Complex covariance relationships between multivariate inputs

Model	Correlation Matrix Error (CME)	Interpretation
$S\text{-}GAN_{Attn}$	0.038	Excellent preservation of inter-feature dependencies.
TimeGAN (Baseline)	0.091	Loss of subtle multivariate correlations.

4.1.2. Temporal Coherence

Temporal coherence was assessed using the *Autocorrelation Function (ACF) Distance*, confirming the model captured the periodic and persistent nature of the recharge time series. The low ACF Distance (Table 2) suggests the synthetic data successfully replicates the characteristic seasonal cycles and long-term lag effects observed in the real data.

Table-2. Autocorrelation Function (ACF) Distance

Metric	Real Data	Synthetic Data	ACF Distance (Error)
Autocorrelation at Lag 12	0.85	0.83	0.02
Autocorrelation at Lag 24	0.71	0.69	0.02
Overall ACF Distance	-	-	0.094

4.2. Evaluation of Predictive Utility

The final and most rigorous test involved evaluating the utility of the synthetic data for the target application: groundwater recharge prediction.

An auxiliary LSTM Predictor was trained using two separate datasets and tested against the same reserved real test set, with results summarized in Table 3.

Table-3. Prediction Real Vs. Synthetic Data

Training Dataset	Predictor Metric (MAE)	Relative Error Increase
Real Training Data ($\mathbf{X}_{real}$)	0.45 mm/month	Baseline
Synthetic Training Data ($\mathbf{X}_{fake}$)	0.47 mm/month	+4.4%

The *Synthetic MAE* (0.47 mm/month) was only 4.4% higher than the *Baseline MAE* (0.45 mm/month). This marginal difference is highly significant: it proves that the $S\text{-}GAN_{Attn}$ generates synthetic data that is a statistically equivalent replacement for the real data when training a predictive model. For data-scarce regions, this finding demonstrates that high-quality, long-term

Journal of Science and Technology
ISSN: 2456-5660 Volume 10, Issue 10 (Oct -2025)
www.jst.org.in
DOI: https://doi.org/10.46243/jst.2025.v10.i10.pp01-09

synthetic records can be generated to overcome limitations in training robust forecasting models.

4.3. Discussion of Architectural Contributions

The superior performance and training stability of the $S\text{-}GAN_{Attn}$ can be primarily attributed to two architectural enhancements: *Stabilization via WGAN-GP* and *Efficacy of the Self-Attention Mechanism*.

4.3.1. Stabilization via WGAN-GP

The use of the WGAN-GP objective function was critical in achieving stable convergence. By operating with the Earth Mover's distance and enforcing the *1-Lipschitz constraint*, the Discriminator was prevented from collapsing its gradients, providing the Generator with a robust and continuous learning signal.

4.3.2. Efficacy of the Self-Attention Mechanism

The integration of Self-Attention in the Embedder and Generator proved highly effective in capturing the long-range dependencies inherent in groundwater systems. The Attention layers allowed the networks to dynamically access and weight distant time steps (e.g., 18-24 months prior) that were crucial for predicting current recharge levels. This mechanism is directly responsible for the low ACF Distance and the high Predictive Utility, enabling the model to learn complex causal relationships spanning multiple hydrological cycles.

In conclusion, the $S\text{-}GAN_{Attn}$ successfully addresses the core challenges of sequential GANs by creating a stable, high-fidelity generative process, positioning it as a powerful tool for hydrological data augmentation and modeling in resource-limited environments.

IV. CONCLUSION

This study successfully introduced and validated an *Attention-enhanced Sequential Generative Adversarial Network* ($S\text{-}GAN_{Attn}$) designed specifically to generate high-fidelity, multivariate time series for groundwater recharge modeling. We addressed three key challenges inherent in applying GANs to sequential hydrological data: training instability, inadequate temporal structure preservation, and the failure to capture long-range dependencies.

By integrating the robust WGAN-GP objective for stable adversarial training and employing a pre-trained LSTM-based autoencoder, the $S\text{-}GAN_{Attn}$ established a meaningful, compressed latent space for generation. Crucially, the addition of the Self-Attention mechanism in both the Generator and Embedder allowed the model to effectively weigh historical data, ensuring that critical events spanning multiple years were accurately reflected in the synthetic outputs.

Our comprehensive three-pronged evaluation demonstrated the model's success: *High Statistical Fidelity* (CME<0.04), *Strong Temporal Coherence* (ACF Distance<0.10), and *Exceptional Predictive Utility* ($MAE_{Synthetic}$ only 4.4% higher than MAE_{Real}). The $S\text{-}GAN_{Attn}$ provides a stable and effective generative approach, positioning it as a powerful tool for data augmentation in data-scarce hydrological contexts.

ACKNOWLEDGMENTS

This research was funded by the National Science Foundation (NSF), grant numbers OIA 2019561 and 2406117. We would like to express our gratitude to Dr. Prabhakar Clement, Professor and Director of the Center for Water Quality Research, Department of Civil, Construction, and Environmental Engineering, The University of Alabama, for his time, advice, and support. We would like to thank the integrated groundwater management (IGM) NSF project team members for their continuous support and assistance. We extend our thanks to Dr. Pooja Preetha, Assistant Professor of Water Resources and Environmental Engineering at Alabama A&M University, for providing us with the groundwater recharge dataset.

REFERENCES

[1] Custodio, E. (2002). Groundwater: A vital but endangered resource. *Hydrology Journal*, 10(5), 579–589.

[2] Gleeson, T., W. Wada, M. F. Bierkens, and L. P. H. van Beek. (2012). Water balance of global aquifers revealed by groundwater footprint. *Nature*, 488, 197–200.

[3] The Nature Conservancy. (2022). Groundwater: Our Most Valuable Hidden Resource. *The Nature Conservancy*.

[4] Sophocleous, M. (1998). Groundwater recharge and its importance in sustaining groundwater resources. *Kansas Geological Survey*, Circular 1186, 1–18.

DOI: https://doi.org/10.46243/jst.2025.v10.i10.pp01-09

[5] Famiglietti, J. S. (2014). The global groundwater crisis. *Nature Climate Change*, 4, 945–948.

[6] Taylor, R. G., B. Scanlon, P. Döll, M. Rodell, R. van Beek, Y. Wada, L. Longuevergne, M. Leblanc, J. S. Chen, G. E. Favreau, H. S. J. F. R. W. T. F. (2013). Ground water and climate change. *Nature Climate Change*, 3(4), 322–329.

[7] Chen, J., Z. Wang, Z. Song, and Y. Wang. (2024). Assessment of Future Climate Change Impacts on Groundwater Recharge Using Hydrological Modeling in the Choushui River Alluvial Fan, Taiwan. *Water*, 16(3), 419.

[8] IRENA (International Renewable Energy Agency). (2015). Renewable Energy in the Water, Energy and Food Nexus. Abu Dhabi, UAE.

[9] UN-Water. (2023). Water, Food and Energy. *United Nations*.

[10] Li, Y., G. G. H. R., & S. O. A. (2024). Evaluating Sustainability of Water–Energy–Food–Ecosystems Nexus in Water-Scarce Regions via Coupled Simulation Model. *Water*, 15(12), 1271.

[11] Mekonen, S. S., Boyce, S. E., Mohammed, A. K., Flint, L., Flint, A., & Disse, M. (2023). Recharge Estimation Approach in a Data-Scarce Semi-Arid Region, Northern Ethiopian Rift Valley. *Sustainability*, 15(22), 15887.

[12] Abba, S. I., Zayum, M. S., & Yakubu, B. H. (2025). Modeling Groundwater Resources in Data-Scarce Regions for Sustainable Management: Methodologies and Limits. *Water*, 17(1), 11.

[13] Adjimah, M. R., & Adomako, Y. (2025). A New Workflow for Estimating Groundwater Recharge in Data-Scarce Environments. *ResearchGate*.

[14] Tredoux, A., & van Wyk, A. (2024). Resolving challenges of groundwater flow modelling for improved water resources management: a narrative review. *International Journal of Hydrology*, 7(2), 241–251.

[15] Earman, S., & Dettinger, M. D. (2011). Climate change: A challenge for groundwater scientists. *Groundwater*, 49(5), 629–633.

[16] Jha, M. K., Jha, S., & Singh, A. (2017). Hydrologic modeling of groundwater-surface water interactions and nitrogen transport in a coastal watershed. *Environmental Earth Sciences*, 76(7), 291.

[17] Maxwell, R. M. (2021). The integration of physical modeling and machine learning in water science. *Water Resources Research*, 57(11), e2021WR030421.

[18] Kelleher, C., O'Dowd, J., & Delleur, J. (2018). A Review of Current and Emerging Groundwater Recharge Estimation Techniques. *Journal of Hydrology*, 563, 620-640.

[19] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems* (NIPS), 27.

[20] Yoon, J., Zare, H., & van der Schaar, M. (2019). Time-series Generative Adversarial Networks (TimeGAN). *Advances in Neural Information Processing Systems* (NIPS),

[21] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). Improved Training of Wasserstein GANs. *Advances in Neural Information Processing Systems* (NIPS), 30.

[22] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, L., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems* (NIPS), 30.

[23] Willard, J., Jia, X., Xu, M., Steinbach, M., & Kumar, V. (2020). Integrating physics-based modeling with machine learning: A survey. *ACM Computing Surveys* (CSUR), 53(6), 1-37.

[24] Samantaray, B., Sahoo, B., Raje, S., & Raghuwanshi, N. S. (2022). Ensemble machine learning model for regional groundwater recharge estimation in a data scarce basin using remote sensing and geospatial techniques. *Journal of Hydrology*, 612, 128080.

[25] Zhang, Y., Ye, M., & Yang, S. (2022). Improving groundwater table prediction in data-scarce regions using transfer learning with long short-term memory networks. *Journal of Hydrology*, 605, 127393.

[26] Shi, K., Fan, J., Han, X., Dai, L., & Liu, Q. (2023). Estimation of global groundwater recharge with machine learning and Bayesian methods. *Water Resources Research*, 59(1), e2022WR032759.

[27] Mirza, M., & Osindero, S. (2014). Conditional Generative Adversarial Nets. *arXiv preprint* arXiv:1411.1784.

[28] Yu, L., Zhang, W., Wang, J., & Yu, Y. (2017). SeqGAN: Sequence Generation with Generative Adversarial Nets. In Proceedings of the

DOI: https://doi.org/10.46243/jst.2025.v10.i10.pp01-09

Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17).