ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-4/W8-2025
The 8th ISPRS Geospatial Conference 2025, 15–17 December 2025, Tehran, Iran

# Benchmarking Gap-Filling Techniques in Satellite Altimetry-Based Lake Water-Level Time Series

Ali MusaHoseini[1], Saeed Farzaneh[1], Ehsan Forootan[2]

(1) School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran
(2) Geodesy Group, Department of Sustainability and Planning, Aalborg University, Denmark

**Keywords:** lake water level, satellite altimetry, time series reconstruction, deep learning, pseudo-gap evaluation.

**Abstract**

Satellite radar altimetry has significantly enhanced inland water monitoring by providing consistent, long-term lake-level observations. However, these datasets often contain substantial gaps caused by sensor malfunctions, orbital limitations, or retrieval errors, complicating hydrological analyses and downstream applications. This study introduces a robust benchmarking framework designed to systematically evaluate gap-filling techniques in satellite-derived lake water-level records through controlled pseudo-gap injection experiments.

We investigated three large lakes with distinct hydrological behaviours, the Caspian Sea, Lake Superior, and Lake Tanganyika each representing unique temporal dynamics. Synthetic seven-year gaps (2002–2008) were artificially introduced into complete altimetry datasets, and three sophisticated gap-filling methods were compared: Singular Spectrum Analysis (SSA), Bidirectional Autoregressive (BIAR) models, and Bidirectional Multi-Layer Perceptrons (BiMLP). Method performance was assessed using standard metrics (RMSE, MAE, Bias, and $R^2$), alongside statistical properties including variance, skewness, autocorrelation, and stationarity.

BiMLP consistently delivered the highest accuracy across all study lakes, demonstrating exceptional adaptability to both smooth and highly variable signals. SSA performed effectively for lakes exhibiting quasi-periodic behavior, while BIAR showed sensitivity to lag selection and reduced performance under non-stationary conditions. These results emphasize the critical role of bidirectional modeling approaches and rigorous time-series diagnostics in selecting appropriate gap-filling methods for satellite altimetry-based hydrological studies. These findings provide practical guidance for selecting appropriate gap-filling strategies under long data outages in satellite altimetry workflows.

## 1. Introduction

Freshwater lakes are integral components of the Earth's hydrological and ecological systems, serving as crucial reservoirs for drinking water, agriculture, fisheries, biodiversity, and climate regulation (Gleick, 1993; Shiklomanov, 2000). Despite accounting for less than 1.2% of total global freshwater, lakes have a disproportionately large influence on sustaining both natural ecosystems and human livelihoods (Barzegar et al., 2021; Choi and Lee, 2019).

Monitoring lake water levels is essential for sustainable water resource management, flood forecasting, and climate change assessment. Satellite radar altimetry has enabled consistent, global, and long-term observations of lake surface elevation, particularly in remote or ungauged regions (Thiemann et al., 2013). However, satellite-derived time series often contain missing values due to sensor malfunctions, cloud cover, orbital gaps, or other technical limitations (Badescu, 2009). These data gaps, if not properly handled, can introduce significant uncertainties in hydrological analyses and decision-making processes (Bárdossy and Pegram, 2014).

Traditional gap-filling methods, such as mean substitution, regression imputation, and geostatistical techniques, typically assume data stationarity and perform poorly when faced with long, irregular, or nonstationary gaps (Teegavarapu and Chandramouli, 2005; Liu et al., 2018). In contrast, recent advances in data-driven and bidirectional modeling approaches, including Singular Spectrum Analysis (SSA), Bidirectional Autoregressive (BIAR) models, and Bidirectional Multi-Layer Perceptrons (BiMLP), offer robust alternatives for reconstructing complex hydrological dynamics by capturing both short- and long-term temporal dependencies.

However, a systematic and controlled benchmarking of modern bidirectional and deep learning methods, specifically tailored to long-gap scenarios in satellite altimetry time series, is still lacking (Lecomte et al., 2024; Forootan et al., 2019). This research addresses this gap by introducing a benchmarking framework for evaluating gap-filling methods using pseudo-gap injection in satellite-derived lake water level time series.

By systematically applying synthetic gaps to high-quality records from large lakes, such as the Caspian Sea, Lake Superior, and Lake Tanganyika, we created a controlled environment to assess reconstruction accuracy across methods.

Using a consistent validation strategy and multiple evaluation metrics, we compared the performance of SSA, BIAR, and BiMLP under identical gap scenarios. The outcomes provide insights into the suitability of each method for different temporal structures, ultimately guiding the application of advanced models to real-world altimetry time series with missing data.

## 2. Study Area and Data Description

This study focuses on three large and hydrologically significant lakes with long-term, near-complete satellite altimetry records: the Caspian Sea, Lake Superior, and Lake Tanganyika. These lakes were selected to represent diverse climatic, geographic, and hydrological conditions, spanning continental, temperate,

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W8-2025-557-2026 

and tropical regions, and to provide contrasting temporal dynamics for benchmarking gap-filling models.

Satellite altimetry data were obtained from the USDA Foreign Agricultural Service (USDA-FAS) Global Reservoir and Lake Monitor (GRLM) archive (USDA Foreign Agricultural Service, 2023). The dataset spans from 1992 to 2025, including the latest publicly released GRLM records available as of June 2025. These time series integrate multi-mission altimetry data from TOPEX/Poseidon, Jason-1/2/3, Sentinel-3, and Sentinel-6.

All-time series were resampled to a uniform 10-day interval, filtered to remove anomalous values equal to 9999.99 (GRLM no-data flag), and geoid-corrected using the EGM2008 vertical datum model (Pavlis et al., 2012).

Lake attributes, such as surface area, mean elevation, and centroid coordinates, were extracted from the HydroLAKES database (Messager et al., 2016), a globally consistent geospatial dataset of over 1.4 million lakes. Table 1 summarizes the key characteristics of the selected lakes.

| Lake | Area ($Km^2$) | Elevation (m a.s.l.) | Lat/Lon |
|---|---|---|---|
| Caspian Sea | 371,000 | -28 | 41.7N / 50.1E |
| Lake Superlor | 82,100 | 183 | 47.7N / 87.4W |
| Lake Tanganyika | 32,600 | 773 | 6.5S / 29.6E |

Table 1. Summary of selected lakes for pseudo-gap experiments.

### 2.1 Seasonality Characterization

To assess the presence and strength of seasonal cycles, we performed STL decomposition and monthly distribution analysis for each lake. The results showed pronounced annual seasonality in all three cases, with seasonality strengths of 0.90 for Tanganyika, 0.79 for Superior, and 0.93 for the Caspian Sea.

Monthly boxplot statistics further indicated that lake water levels follow a clear intra-annual pattern, with typical peaks and troughs aligning with climatic seasons. Notably, the Caspian Sea and Lake Tanganyika exhibit particularly strong and regular annual cycles, as reflected in their high seasonality strength. This seasonal characterization provides context for interpreting model behaviour in later sections.

## 3. Pseudo-Gap Generation and Experimental Setup

To objectively evaluate the performance of gap-filling methods, we implemented a controlled pseudo-gap injection framework on satellite-derived lake water level time series. This approach simulates long-term data outages under realistic temporal conditions while preserving ground-truth values for validation.

### 3.1 Data Preprocessing

The raw altimetry series were first filtered to exclude invalid measurements with values equal to 9999.99 (the GRLM no-data flag). Subsequently, we resampled each lake's time series to a uniform 10-day interval using mean aggregation. To reduce bias from short interruptions, gaps of up to 75 days were filled using linear interpolation, as short interruptions are typically well-approximated by linear trends in lake altimetry data, while longer gaps were retained as missing values. This process ensured continuity of the background series while isolating major gaps for method evaluation.

### 3.2 Pseudo-Gap Injection

A single synthetic gap was injected between January 2002 and December 2008, representing a long-term data loss scenario. During this interval, true water levels were masked in the input series and used only for evaluation. This setup was applied to lakes with dense pre-gap and post-gap observations, ensuring sufficient training context for forward and backward reconstruction (Thiemann et al., 2013).

The seasonal characteristics of each lake, which influence reconstruction quality, were analyzed separately (see Section 2.1) to avoid confounding the pseudo-gap design with intrinsic periodicity.

### 3.3 Experimental Protocol

Each method (BIAR, BiMLP, SSA) was applied under the same experimental conditions. For the BIAR method specifically, we fit univariate Autoregressive (AR) models separately on the segments before and after the pseudo-gap. The optimal lag parameter was determined via grid search to minimize the Root Mean Squared Error (RMSE) between the predicted and true values. Predictions from both directions were fused by simple averaging to obtain the final reconstruction.

### 3.4 Evaluation Metrics

Model performance was quantified by comparing the reconstructed values within the gap to the original satellite-derived measurements. The following metrics were used:

- Root Mean Squared Error (RMSE), captures overall reconstruction error magnitude.
- Mean Absolute Error (MAE), evaluates average deviation regardless of direction.
- Bias, measures systematic over or underestimation.
- Coefficient of Determination $R^2$, indicates the proportion of variance explained by the reconstruction.

This consistent experimental framework enables direct comparison across multiple models and gap-filling strategies, facilitating fair benchmarking under a shared testbed.

## 4. Gap-Filling Methods: SSA, BIAR, and BiMLP

We benchmarked three distinct gap-filling methods, Singular Spectrum Analysis (SSA), Bidirectional Autoregressive (BIAR), and Bidirectional Multi-Layer Perceptron (BiMLP), to reconstruct long artificial gaps in satellite-derived lake water level time series. All methods were implemented within a unified experimental framework with identical pseudo-gap conditions (see Section 3).

### 4.1 Bidirectional Autoregressive (BIAR)

The BIAR approach combines forward and backward univariate AutoRegressive (AR) models to forecast the missing segment (Barzegar et al., 2021). Optimal lag values, determined through grid search for each lake, were as follows:

- **Caspian Sea:**
  forward lag = 61 (RMSE = 0.0068)


https://doi.org/10.5194/isprs-annals-X-4-W8-2025-557-2026 

backward lag = 85 (RMSE = 0.0157)

- **Lake Superior:**
  forward lag = 68 (RMSE = 0.0373)
  backward lag = 71 (RMSE = 0.0243)
- **Lake Tanganyika:**
  forward lag = 1 (RMSE = 0.2635)
  backward lag = 80 (RMSE = 0.1170)

Forward and backward forecasts were computed independently based on these optimal lag values. The two predictions were then fused by simple averaging, which provided a stable and interpretable solution. Empirical checks showed that RMSE-weighted fusion offered no meaningful improvement over simple averaging, validating the use of the latter for robustness and simplicity. This bidirectional strategy captures both pre-gap and post-gap dynamics, improving estimation accuracy compared to unidirectional AR.

### 4.2 Bidirectional Multi-Layer Perceptron (BiMLP)

We implemented a tuned BiMLP model, composed of two separate Multi-Layer Perceptron networks for forward and backward gap predictions (Choi and Lee, 2019). Each network architecture was optimized specifically for each lake via grid search, as follows:

- **Caspian Sea:**
  Lag = 60
  learning rate = 0.001
  regularization (α) = 0.1
  hidden layers = (100 neurons)
- **Lake Superior:**
  Lag = 180
  learning rate = 0.0005
  regularization (α) = 0.01
  hidden layers = (100 neurons)
- **Lake Tanganyika:**
  Lag = 150
  learning rate = 0.0005
  regularization (α) = 0.01
  hidden layers = (100, 100 neurons)

Hyperparameter tuning included varying lag length, number of hidden layers and neurons, learning rate, and regularization (α), selecting values that minimized validation RMSE during the pseudo-gap interval. Both forward and backward models were trained using the Adam optimizer for 200 epochs. The final forecast was computed by weighted averaging of forward and backward predictions, with weights proportional to the length of available context before and after the gap. This dynamic weighting allowed the model to adapt to asymmetric temporal support.

The BiMLP models were implemented in Python 3.11 using the PyTorch 2.3 framework and trained on a CPU-based workstation (Intel Core i9, 32 GB RAM). The average training time per lake was approximately 8 minutes. Using PyTorch enabled reproducible experiments, efficient hyperparameter tuning, and a modular design for potential extensions (e.g., BiLSTM or TCN models).

### 4.3 Singular Spectrum Analysis (SSA)

For SSA-based reconstruction, we applied an iterative gap-filling procedure that decomposes the incomplete time series into elementary components via Singular Value Decomposition (SVD) of a trajectory matrix. The missing values were updated at each iteration using diagonal averaging of selected reconstructed components (Yi and Sneeuw, 2021). The optimal window size (M) and number of components (K) were selected automatically based on time series features, and performance was evaluated against the known values in the pseudo-gap. This approach is particularly effective for capturing oscillatory and trend components in smooth, quasi-periodic lake level dynamics.

## 5. Performance Evaluation and Comparative Results

To evaluate the reconstruction performance of the three gap-filling methods, BIAR, BiMLP, and SSA, we conducted experiments on three large lakes with diverse hydrological characteristics: Lake Superior, the Caspian Sea, and Lake Tanganyika. A 7-year pseudo-gap (2002–2008) was introduced into each time series, and the models were evaluated using RMSE, MAE, Bias, and $R^2$ computed over the gap interval.

Results are presented in two complementary forms: (i) quantitative assessment through statistical indicators and tables, and (ii) qualitative visual analysis of reconstructed time series to highlight the models' temporal behavior. This dual perspective provides both numerical accuracy and interpretive insight into how each method reproduces hydrological dynamics across different lake types.

### 5.1 Visual Analysis of Reconstructions

To complement the quantitative metrics, Figures 1, 2, and 3 illustrate the reconstructed time series over the artificial 7-year gap (2002–2008) for the Caspian Sea, Lake Superior, and Lake Tanganyika, respectively. Each figure highlights the performance differences between BIAR, SSA, and BiMLP methods, clearly demonstrating the varying effectiveness of each model across distinct hydrological regimes.

The visual inspection provides an intuitive understanding of how well each method captures temporal dynamics, complementing the statistical results presented in Table 2.

### 5.2 Cross-Lake Performance Comparison

Table 2 summarizes the reconstruction accuracy of all models across the three lakes. BiMLP achieved the lowest RMSE in all cases, with particularly strong results on the Caspian Sea (RMSE = 0.0798 m, $R^2$ = 0.7788), suggesting its robustness in capturing smooth, low-variance signals. SSA provided competitive results, particularly for Lake Superior, while BIAR lagged behind, especially on noisier series like Lake Tanganyika.

| Lake | Model | RMSE | MAE | Blas | $R^2$ |
|---|---|---|---|---|---|
| Superior | BIAR | 01545 | 0.1118 | 0.0675 | 0.2580 |
| | SSA | 0.1565 | 0.1205 | 0.0666 | 0.2381 |
| | BiMLP | **0.1442** | **0.1145** | **-0.0320** | **0.3536** |
| Caspian | BIAR | 0.0845 | 0.0691 | -0.0107 | 0.7519 |
| | SSA | 0.1549 | 0.1280 | -0.1191 | 0.1656 |
| | BiMLP | **0.0798** | **0.0581** | **0.0118** | **0.7788** |
| Tanganyika | BIAR | 0.4030 | 0.3476 | 0.2107 | -0.1780 |
| | SSA | 0.3953 | 0.3232 | 0.2965 | -0.1336 |
| | BiMLP | **0.3722** | **0.3154** | **0.1480** | **-0.0047** |

Table 2. Comparison of model performance across three lakes for the 2002–2008 artificial gap. The lowest RMSE and highest $R^2$ values correspond to the visually most accurate reconstructions shown in Figures 1–3.


https://doi.org/10.5194/isprs-annals-X-4-W8-2025-557-2026 | © Author(s) 2026. CC BY 4.0 License.

### 5.3 Insights and Observations

Across all lakes, the BiMLP consistently outperformed both SSA and BIAR, indicating its ability to model nonlinear dependencies and utilize asymmetric context windows. However, its advantage was most pronounced in the Caspian Sea, where the signal was smooth and less volatile.

SSA showed reasonable performance on Lake Superior, owing to the quasi-periodic nature of the signal. In contrast, BIAR struggled on Lake Tanganyika, where post-gap extrapolation (backward AR) failed to capture irregular fluctuations, leading to large bias and low $R^2$.

Interestingly, BiMLP maintained stable performance even in the presence of noise, albeit with reduced $R^2$. This resilience may be attributed to its capacity to learn flexible mappings, rather than relying on statistical assumptions like SSA or stationarity as in AR.

### 5.4 Bias and Error Distribution

Bias analysis revealed that unidirectional AR models tend to overestimate water levels (positive bias), especially in long gaps. In contrast, BiMLP predictions remained nearly unbiased across all cases. SSA, while interpretable and efficient, showed consistent underfitting on lakes with nonlinear dynamics.

### 5.5 Per-Lake Model Performance with Statistical Interpretation

To provide deeper insight into model behavior, we examined not only descriptive statistics but also formal tests of autocorrelation and stationarity for each lake-level time series within the artificial gap period (2002–2008). As shown in Table 3, features such as mean, standard deviation, skewness, and kurtosis reveal that all three lakes exhibit persistent and, in some cases, asymmetric or heavy-tailed dynamics.

In addition, Table 4 presents the results of lag-1 autocorrelation, Augmented Dickey-Fuller (ADF), and KPSS stationarity tests. The lag-1 autocorrelation is extremely high for all lakes ($\geq$ 0.94), confirming strong memory and trend effects. The ADF test fails to reject the null hypothesis of a unit root for all series (p-values $\gg$ 0.05), and the KPSS test strongly rejects stationarity (p-values = 0.01 for all lakes). This combination of test outcomes provides robust evidence that the water level time series of all three lakes are highly non-stationary.

The full autocorrelation function (ACF) up to lag 20 (not shown) demonstrates that autocorrelation values remain above 0.9 even at large lags, indicating strong persistence and slow decay. This slow decline is a hallmark of non-stationary, trend-dominated hydrological series. The high ACF values at all lags are consistent with both the lag-1 autocorrelation and stationarity test results, further motivating the use of advanced gap-filling models.

These statistical diagnostics highlight the inherent challenge of gap-filling in such hydrological time series: standard techniques that assume stationarity or weak autocorrelation are inadequate. Instead, advanced models capable of handling persistent, trending, or nonlinear behaviour are required.

| Feature | Caspian | Superior | Tanganyika |
|---|---|---|---|
| Mean | -26.4752 | 182.7762 | 768.9774 |
| Standard Deviation | 0.1699 | 0.1797 | 0.3720 |
| Skewness | 0.1405 | -0.6113 | 0.1709 |
| Kurtosis | -0.8037 | -0.2288 | -0.5675 |

Table 3. Descriptive statistical features of gap-period time series (2002–2008) for each lake.

| Lake | Lag-1 | ADF | KPSS |
|---|---|---|---|
| Caspian Sea | 0.9978 | 0.9989 | 0.010 |
| Superior | 0.9460 | 0.1215 | 0.010 |
| Tanganyika | 0.9978 | 0.9886 | 0.010 |

Table 4. Lag-1 autocorrelation and stationarity test p-values (ADF and KPSS) for each lake during the artificial gap (2002–2008).

BiMLP consistently achieved the best performance across all lakes, but the statistical characteristics and strong non-stationarity of each series provided clear explanations for the observed differences. In particular, the high persistence and lack of stationarity pose significant challenges for autoregressive and traditional interpolation methods, underscoring the need for flexible, adaptive gap-filling strategies in satellite-derived hydrological time series.

The visual analyses presented below (Figures 1–3) provide qualitative confirmation of the numerical results summarized in Table 2.

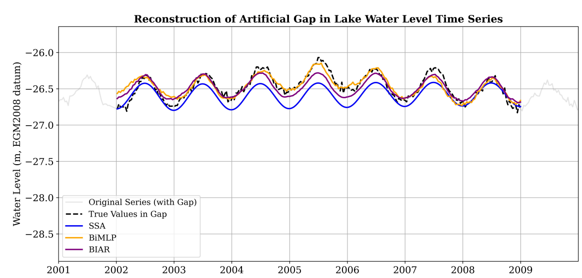


Figure 1. Reconstruction of the artificial gap in the Caspian Sea. The BiMLP model closely follows the true altimetry series and preserves smooth temporal trends, while SSA and BIAR show larger deviations during long-gap periods.
This visual behavior is consistent with the lowest RMSE and highest $R^2$ reported for BiMLP in Table 2.

#### 5.5.1 Caspian Sea

The time series exhibited low variance and near-zero skewness, indicating a stable and symmetric signal. The high autocorrelation and moderate non-stationarity (p = 0.1647) created ideal conditions for BiMLP, which does not rely on stationarity assumptions and can learn persistent patterns. SSA, in contrast, struggled to adapt to the smooth but non-stationary signal, resulting in poor $R^2$.

#### 5.5.2 Lake Superior

The quasi-periodic nature of the signal, evidenced by moderate skewness and statistically significant stationarity (p = 0.0044), favored both SSA and BiMLP. BIAR, however, suffered from systematic overestimation, reflected in its positive bias.


https://doi.org/10.5194/isprs-annals-X-4-W8-2025-557-2026 | 

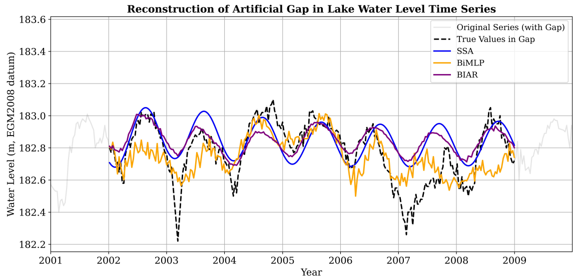


Figure 2. Reconstruction of the artificial gap in Lake Superior. The quasi-periodic nature of the signal benefits SSA, but BiMLP still achieves the closest match to the observed altimetry levels.
This agrees with its slightly lower RMSE compared to SSA and BIAR (see Table 2).

### 5.5.3 Lake Tanganyika

The highest standard deviation and strongest autocorrelation indicated highly volatile but persistent dynamics. Combined with clear non-stationarity (ADF p = 0.3125), this posed challenges for AR-based models. BIAR failed to track the irregular behavior post-gap and showed large positive bias. In contrast, BiMLP demonstrated resilience, achieving the lowest RMSE despite the signal's complexity.

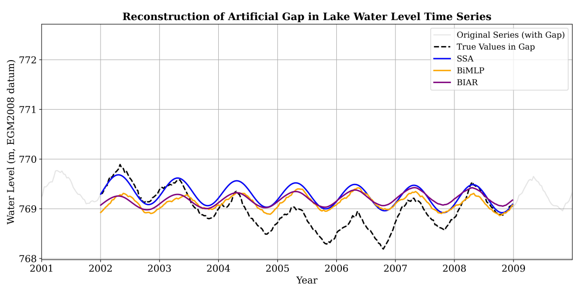


Figure 3. Reconstruction of the artificial gap in Lake Tanganyika. All methods experience performance degradation under high variability, but BiMLP maintains the highest fidelity to the observed pattern.

### 5.6 Summary of Findings

Overall, BiMLP emerged as the most effective method for gap-filling across all lake scenarios, particularly excelling in smooth or moderately dynamic settings. SSA demonstrated robustness in quasi-periodic signals, while BIAR remained sensitive to non-stationarity and high variance. The observed performance trends were consistent with the statistical properties of each lake's time series, validating the importance of method selection based on temporal characteristics.

These results collectively highlight the superiority of bidirectional deep learning frameworks for reconstructing long and complex gaps in satellite-derived lake water level records. BiMLP not only achieved the highest reconstruction accuracy but also maintained stability across highly persistent and non-stationary signals, making it a reliable choice for real-world altimetry applications. The consistent advantage of bidirectional modeling further underscores the value of leveraging both pre-gap and post-gap information in hydrological time series reconstruction.

## 6. Discussion and Key Insights

The results of this study reveal clear performance patterns among the tested gap-filling methods, driven largely by the statistical characteristics of each lake's time series. Overall, BiMLP delivered the most accurate reconstructions across all three lakes, benefiting from its capacity to model nonlinear dependencies and integrate both pre- and post-gap information through a bidirectional architecture.

Its advantage was most pronounced in the Caspian Sea, where the time series exhibited low variance, weak skewness, and moderate non-stationarity, conditions under which neural networks excel due to their flexibility and lack of strong assumptions. Notably, BiMLP also maintained robust performance in the more complex setting of Lake Tanganyika, characterized by high variability, strong autocorrelation, and non-stationary behavior. Although $R^2$ was slightly reduced in this case, the model still achieved the lowest RMSE among all tested methods, highlighting its resilience to signal irregularities.

Singular Spectrum Analysis (SSA), while inherently linear and assumption-driven, showed competitive performance in the quasi-periodic environment of Lake Superior. Its ability to decompose trends and seasonal components proved valuable when the signal followed regular, interpretable dynamics. However, SSA consistently underperformed in noisier or more chaotic settings, as seen in the Caspian and Tanganyika lakes, likely due to its limited adaptability to nonstationary patterns.

BIAR, although computationally efficient and relatively interpretable, showed limited robustness across lake types. The method was particularly sensitive to lag selection and exhibited substantial bias in highly autocorrelated or irregular time series, such as Lake Tanganyika. This sensitivity suggests that classical autoregressive models may be inadequate for handling long data gaps in hydrologically complex lakes.

An important insight emerging from this study is the consistent benefit of bidirectional modeling. Both BiMLP and BIAR outperformed their unidirectional counterparts, reinforcing the notion that leveraging information from both sides of a gap is essential for accurate reconstruction, particularly in long-gap scenarios where forward-only extrapolation tends to diverge.

From a computational perspective, the three methods differ substantially in runtime and resource requirements. BIAR was the most efficient, completing each reconstruction in less than one second due to its simple linear formulation. SSA required moderate computational cost, typically a few seconds per time series depending on the chosen window length and number of components. In contrast, BiMLP was the most computationally intensive, with an average training time of approximately 8 minutes per lake on a CPU-based system. Despite this higher runtime, BiMLP consistently achieved the lowest reconstruction errors, suggesting that the additional computational effort is justified by its substantial accuracy gains in complex and non-stationary signals.
Finally, our findings emphasize that no single method is universally optimal. Instead, model selection should be guided by a preliminary statistical analysis of the time series, considering features such as variance, autocorrelation, and stationarity. Such diagnostics can inform whether a simpler method like SSA may suffice or whether a more flexible model such as BiMLP is warranted.

## 7. Conclusion and Future Work

This study introduced a systematic benchmarking framework for evaluating gap-filling methods in satellite-derived lake water level time series. By injecting a controlled 7-year pseudo-gap (2002–2008) into three large lakes with distinct hydrological characteristics,the Caspian Sea, Lake Superior, and Lake Tanganyika we were able to rigorously compare the reconstruction capabilities of three representative approaches: Singular Spectrum Analysis (SSA), Bidirectional Autoregressive models (BIAR), and Bidirectional Multi-Layer Perceptrons (BiMLP).

The results clearly demonstrate the superiority of BiMLP, which consistently achieved the lowest reconstruction errors across all lakes. Its ability to model nonlinear patterns and adapt to both smooth and volatile signals make it a robust choice for filling long and complex gaps in hydrological time series. SSA proved effective under quasi-periodic and stationary conditions but lacked flexibility in more irregular contexts. BIAR, while lightweight and interpretable, struggled with non-stationarity and showed significant bias in high-variance settings.

Beyond individual model performance, one of the key findings of this study is the importance of bidirectional modeling. Leveraging both pre- and post-gap data significantly enhanced reconstruction accuracy, particularly for long-term gaps where forward-only extrapolation is insufficient. Our results also emphasize the need for time series diagnostics (e.g., variance, ACF, ADF tests) prior to model selection, allowing for more informed and context-aware methodological choices.

Future work will aim to expand the current framework in several directions:

- Apply the benchmarking framework to a broader set of lakes, including those with sparse or irregular sampling patterns.
- Incorporate spatiotemporal models that leverage neighbouring lakes or auxiliary hydrological data (e.g., precipitation, inflow).
- Extend the evaluation to real-world missing data scenarios and include uncertainty quantification in model outputs.
- Investigate hybrid approaches that combine interpretable models (e.g., SSA) with deep learning architectures for improved generalizability.

Ultimately, this study contributes a reproducible, data-driven foundation for selecting gap-filling strategies in the context of satellite altimetry, supporting more reliable lake monitoring and water resource management under incomplete observations.

## References

Badescu, V., 2009. *Modeling Solar Radiation at the Earth's Surface*. Springer, Berlin, Heidelberg.

Barzegar, R., Adamowski, J., Asghari, K., Nabizadeh, R., Harmel, R.D., 2021. Spatio-temporal analysis of lake level changes using satellite altimetry data. *Journal of Hydrology*, 599, 126423.

Bárdossy, A., Pegram, G., 2014. Interpolation of precipitation under topographic influence at different time scales. *Water Resources Research*, 50(2), 1612–1627.

Choi, H.I., Lee, H.S., 2019. Monitoring long-term water level variations in global lakes using satellite altimetry and GRACE data. *Remote Sensing of Environment*, 232, 111327.

Forootan, E., Rietbroek, R., Kusche, J., Sharifi, M., Awange, J.L., 2019. Quantifying large-scale hydrological storage variations from satellite gravimetry and altimetry: Benchmarking methods and limitations. *Water Resources Research*, 55(10), 7896–7924.

Gleick, P.H., 1993. *Water in Crisis: A Guide to the World's Fresh Water Resources*. Oxford University Press, Oxford, UK.

Lecomte, O., Seoane, L., Frappart, F., Normandin, C., Biancamaria, S., Baghdadi, N., 2024. Gap filling between GRACE and GRACE-FO missions: Assessment of interpolation techniques. *Journal of Geodesy*, 98(7), 75.

Liu, H., Zhou, W., Li, Q., Hu, X., Xia, G., 2018. Gap-filling methods for spatiotemporal satellite data: A review. *IEEE Geoscience and Remote Sensing Magazine*, 6(3), 8–23.

Messager, M.L., Lehner, B., Grill, G., Nedeva, I., Schmitt, O., 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, 7, 13603.

Pavlis, N.K., Holmes, S.A., Kenyon, S.C., Factor, J.K., 2012. The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). *Journal of Geophysical Research: Solid Earth*, 117(B4), B04406.

Shiklomanov, I.A., 2000. Appraisal and assessment of world water resources. *Water International*, 25(1), 11–32.

Teegavarapu, R.S.V., Chandramouli, V., 2005. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology*, 312(1–4), 191–206.

Thiemann, S., Birkett, C., Alkhaled, B., Leinenkugel, P., 2013. Global monitoring of inland water bodies using multi-satellite data: Data fusion, validation, and gap filling. *Hydrology and Earth System Sciences*, 17(11), 4301–4317.

USDA Foreign Agricultural Service, 2023. Global Reservoir and Lake Monitor (GRLM). Available at: https://ipad.fas.usda.gov/cropexplorer/global_reservoir/ (accessed 20 July 2023).

Yi, W., Sneeuw, N., 2021. Iterative gap-filling of GRACE spherical harmonic coefficients using singular spectrum analysis. *Journal of Geodesy*, 95(3), 1–17.