TidalMark: a Scalable Benchmark for Coastal Water Level Forecasting

Lucas A. Raicu[1] Daniel Grzenda[1] Ian Foster[1] Kyle Chard[1]

[1]University of Chicago

THE UNIVERSITY OF CHICAGO

globus labs

Abstract

Accurate forecasting of water levels is essential for flood mitigation. Traditionally, predictions have been based on harmonic analysis and sensor networks maintained by the National Oceanographic and Atmospheric Administration. However, these methods struggle with high-variance events. **TidalMark** evaluates deep learning models on coastal water-level forecasting. Our results show that tuned models consistently outperform harmonic approaches between 2.1X and 4.7X (between 7 day to 1 day predictions), enabling adaptive, scalable, and more accurate forecasts.

Harmonic Analysis

- NOAA [4] derives constituents from **>21 years** of continuous data: frequencies, amplitudes, phases via long-term averaging [5, 7]
- Assumes linear superposition of tides and quasi-stationarity
- Effective under stable conditions, but not designed for dynamic forcing (e.g. storms)

Motivation

Challenge [1, 6]	Impact
Rising flood risk	Infrastructure damage, safety risks
Harmonic limits	Miss storm surges, regime shifts
Need adaptivity	Capture aperiodic + periodic patterns

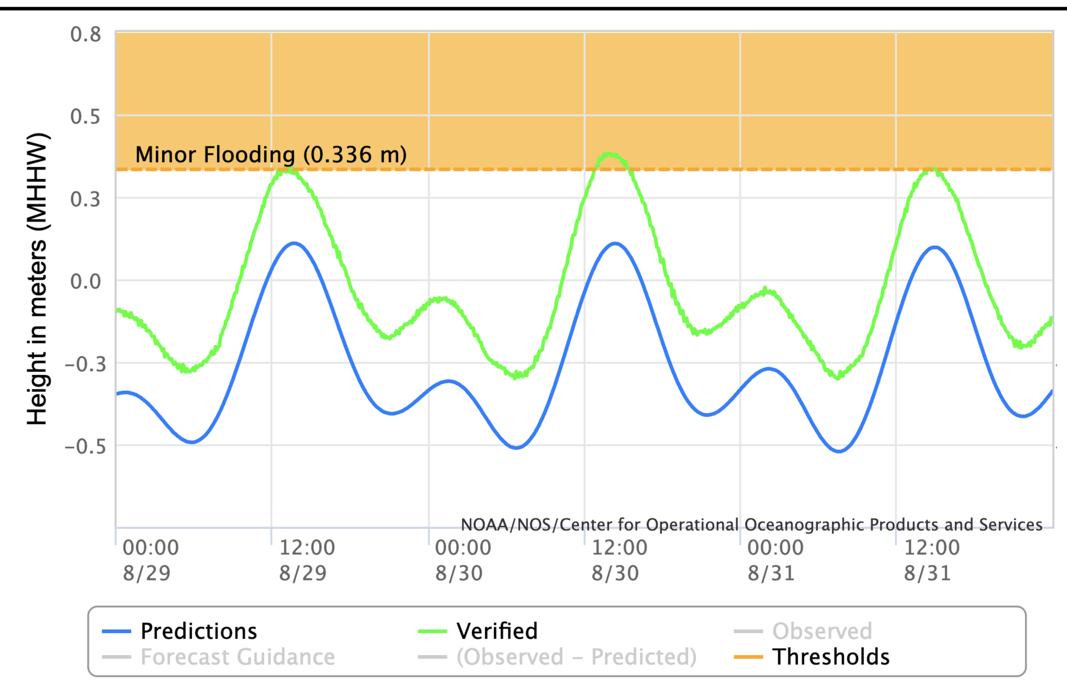
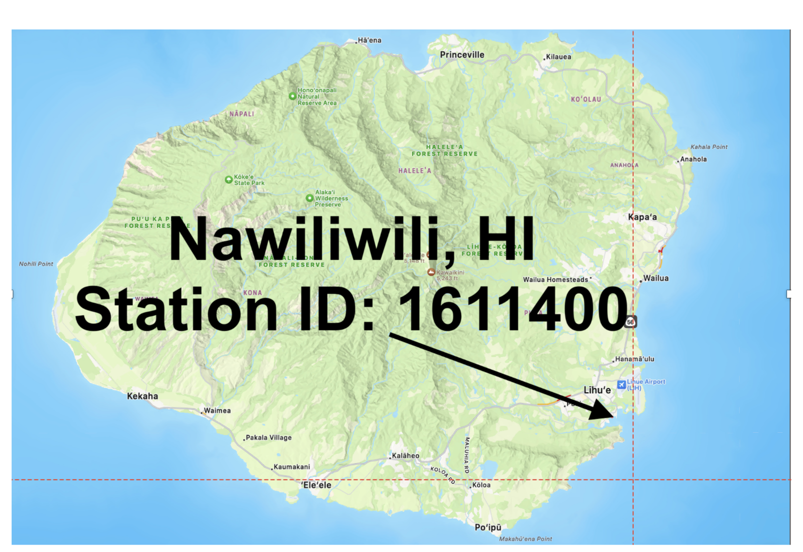


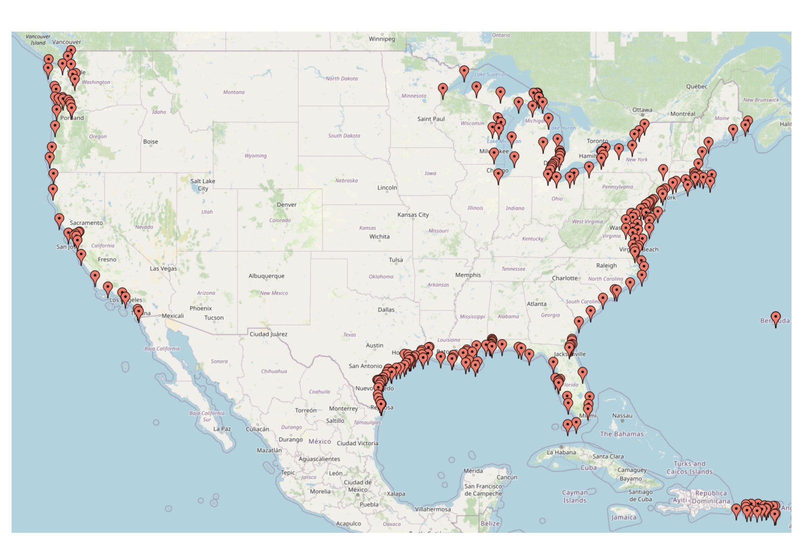
Figure 1. Nawiliwili (2024): observed levels severely exceeded harmonic-based guidance during a major weather system.

Dataset

- Source: NOAA National Water Level Observation Network (NWLON) [4, 3]
- Coverage: **217 stations** (U.S.), **2019–2024**, 6-minute sampling; **>127M** measurements
- Focus station for results: **Nawiliwili, HI** (Station ID: 1611400) due to completeness



(a) Target station (Nawiliwili)



(b) U.S. station coverage

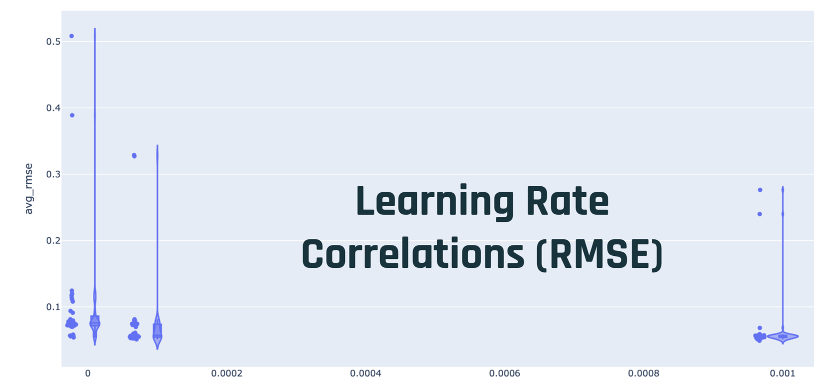
Proposed Work

We developed a modular pipeline in PyTorch with a Long Short-Term Memory (LSTM) [2] neural network for training and evaluating water-level forecasting models. Each model receives a fixed-length window of prior water levels (7 or 14 days) as input and predicts future water levels at multiple time horizons (1, 3, 5, and 7 days). To understand the importance of exogenous variables in accurate water level forecasting, we examined univariate models at first.

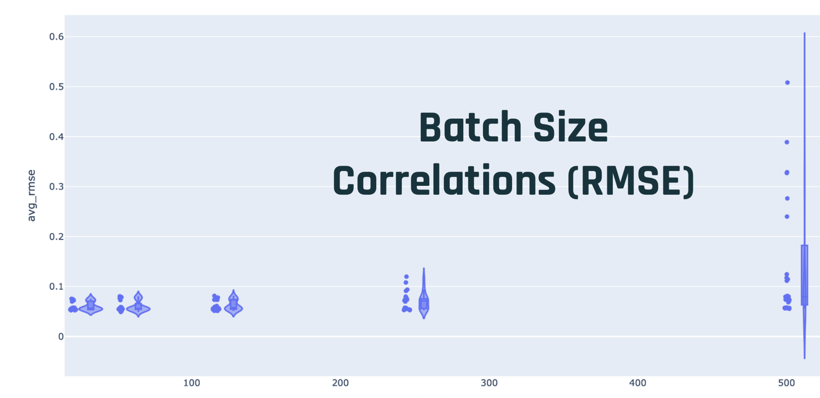
Hyperparameter Sweep

Model Type	Variants Tested	Parameter Ranges
LSTM	Vanilla, BiLSTM	Hidden sizes: 32, 64
Conv-LSTM	1D convolution	Learning rates: 1e-3, 5e-4, 1e-4, 1e-5
GRU	Single, stacked	Layers: 1, 2
Attention-LSTM	Scaled dot-product	Batch sizes: 32, 64, 128, 256, 512

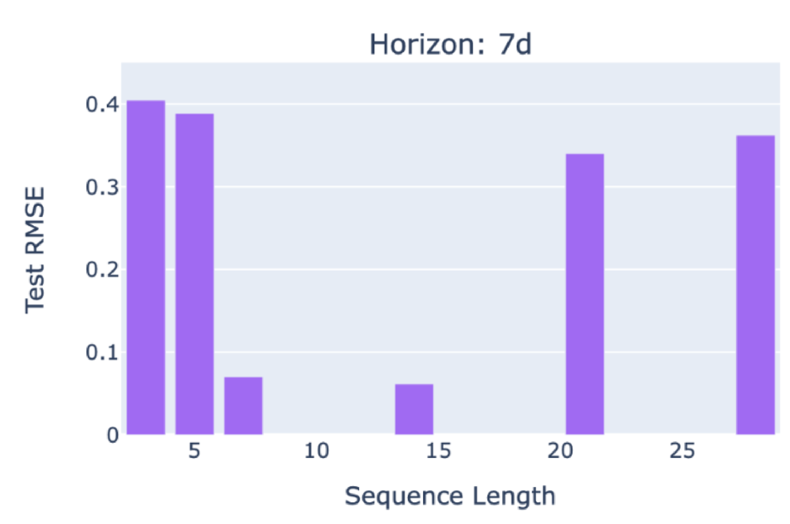
We also extended the LSTM input space to include additional sources of temporal information (neighboring station water levels, NOAA tidal predictions, all 37 harmonic constituents) to determine whether richer inputs improve model performance.



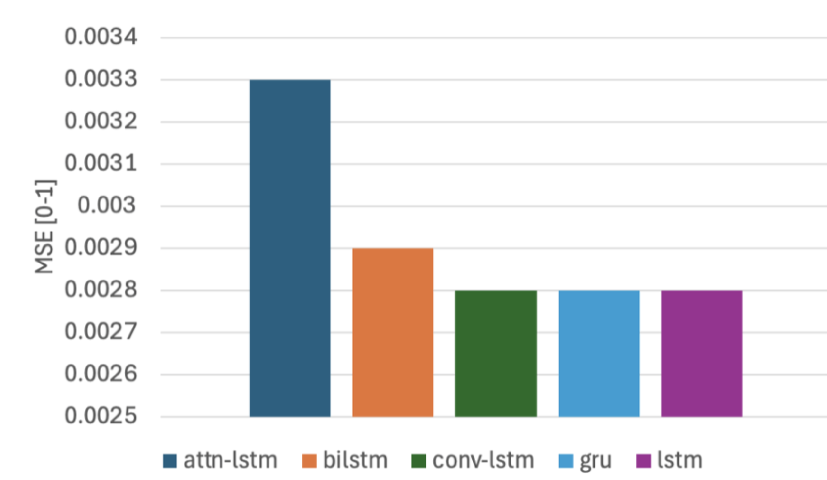
(a) Learning rate vs RMSE



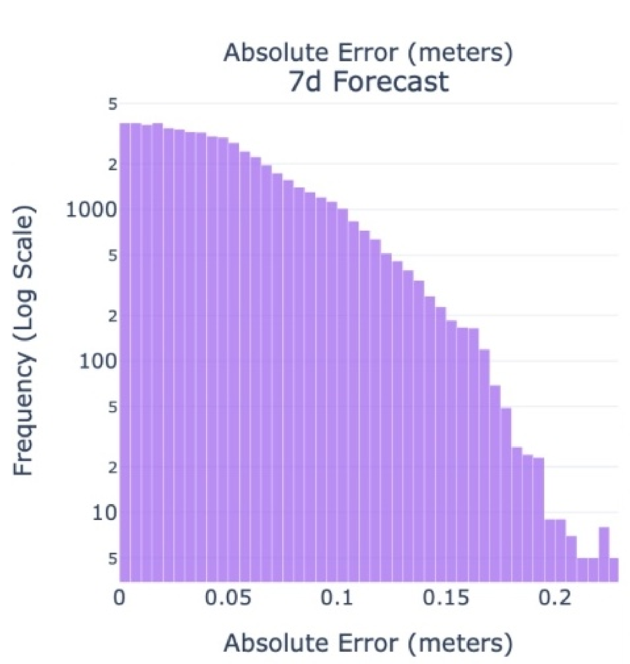
(b) Batch size vs RMSE



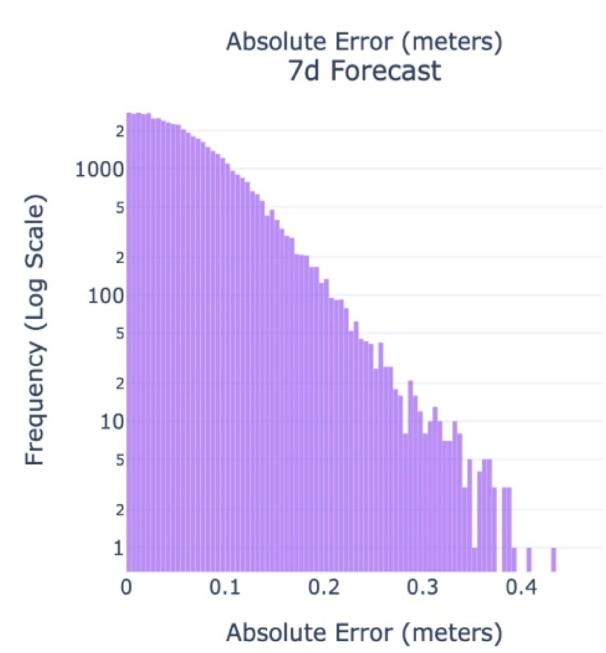
(c) LR dominates; small/medium batches stable; 14d windows help but cost more



(d) MSE distribution across architectures



(e) Best Univariate LSTM – MaxAE: 0.241



(f) Multivariate w/ Neighbor – MaxAE: 0.445

Best LSTM Models vs NOAA Predictions

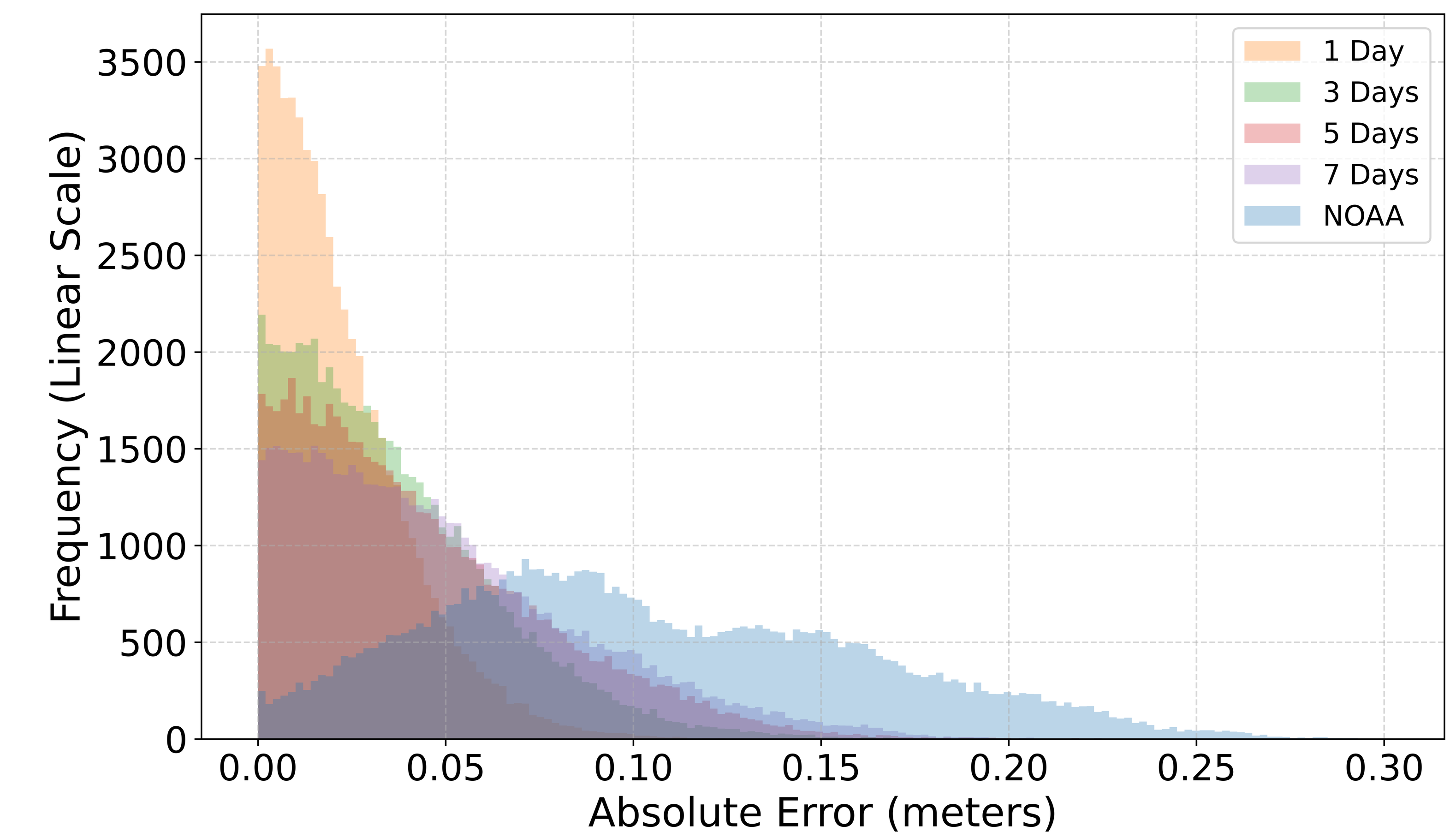


Figure 4. LSTM vs NOAA errors (Nawiliwili, 2024) at 1/3/5/7-day horizons. Best LSTM: seq 14, hs64, lr 0.001, lay1, bs32

Challenges

- **Non-stationarity:** Regimes shift → retraining & drift monitoring
- **Overfitting:** Multivariate LSTMs overloaded → regularization & feature selection
- **Data Quality:** NOAA gaps/outliers → QC & gap filling
- **Compute Costs:** Long training times → GPU accel., pruning, early stopping

Conclusion

- **Tuning discipline > architecture novelty**
- **LSTM models** outperform harmonic forecasts in dynamic coastal environments
- **Performance gain**: LSTMs achieve 2.1x–4.7x better accuracy than harmonic methods, depending on forecast horizon (7-day to 1-day)
- **Broader trend**: Supports evidence that deep learning is surpassing classical statistical and physical models in Earth system modeling

Future Work

- **Spatiotemporal GNNs:** Joint temporal + station dependencies
- **Auto HPO:** Bayesian / PBT search for accuracy vs. cost
- **Feature Fusion:** Weather + harmonic + attention & uncertainty
- **Deployment:** Edge-based early warning for coastal resilience

References

[1] Sergio Consoli, Diego Reforgiato Recupero, and Vanni Zavarella. A survey on tidal analysis and forecasting methods for tsunami detection. *arXiv preprint arXiv:1403.0135*, 2014. Traditional tidal forecasting methods based on harmonic analysis require long-term data and fail under non-astronomical influences such as weather.

[2] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[3] National Oceanic and Atmospheric Administration (NOAA). National oceanic and atmospheric administration. `https://www.noaa.gov/`, 2025. Accessed: 2025-08-08.

[4] NOAA National Water Level Observation Network. Coastal station data, station 1611400. Online, 2025.

[5] David Pugh. *Tides, Surges and Mean Sea-Level*. Wiley, 2nd edition, 2004.

[6] NOAA Ocean Service. What threats do coastal communities face? Online facts page, 2024. Lists threats including extreme natural events, sea level rise and coastal erosion.

[7] Wikipedia contributors. Tide — harmonic analysis historical introduction by william thomson. Wikipedia, accessed 2025-08-09, 2025. Describes development of harmonic analysis in the 1860s for tidal prediction.