

Overview

- We assess water quality to understand climate change and human impacts, focusing on indices like temperature, pH, and oxygen levels.
- Traditional in-situ monitoring is accurate but costly, time-consuming, and limited in coverage.
- Our research develops forecasting models for water temperatures using a combination of in-situ and simulated data.
- We integrate physics-informed neural networks, Koopman embeddings, and LSTM to improve long-term temperature forecasting.
- Our model outperforms conventional LSTM models, showing enhanced forecasting performance.
- We aim to create efficient forecasting models for key water quality indices to improve water body management.

Physics-Informed Neural Networks

Model

- $t = (t_1, t_2, \dots, t_N)$ represents multi-depth lake temperatures over N time steps, where each $t_i \in \mathbb{R}^D$ contains temperatures at D depth levels.
- The Koopman embedding encoder K transforms each t_i into a higher-dimensional vector $k_i \in \mathbb{R}^M$, with M as a tunable hyperparameter.
- After pre-training, the encoder converts the data into a higher-dimensional time series, which is used to train an LSTM.
- The model predicts multi-depth temperatures t_n at time step n using a window of previous W time steps $t_{n-W} : t_{n-1}$.
- The encoder transforms $t_{n-W} : t_{n-1}$ into an embedding matrix, combined with periodic date features to form the full feature matrix.
- The feature matrix is input to the LSTM model, which predicts \hat{t}_n for the next time step.

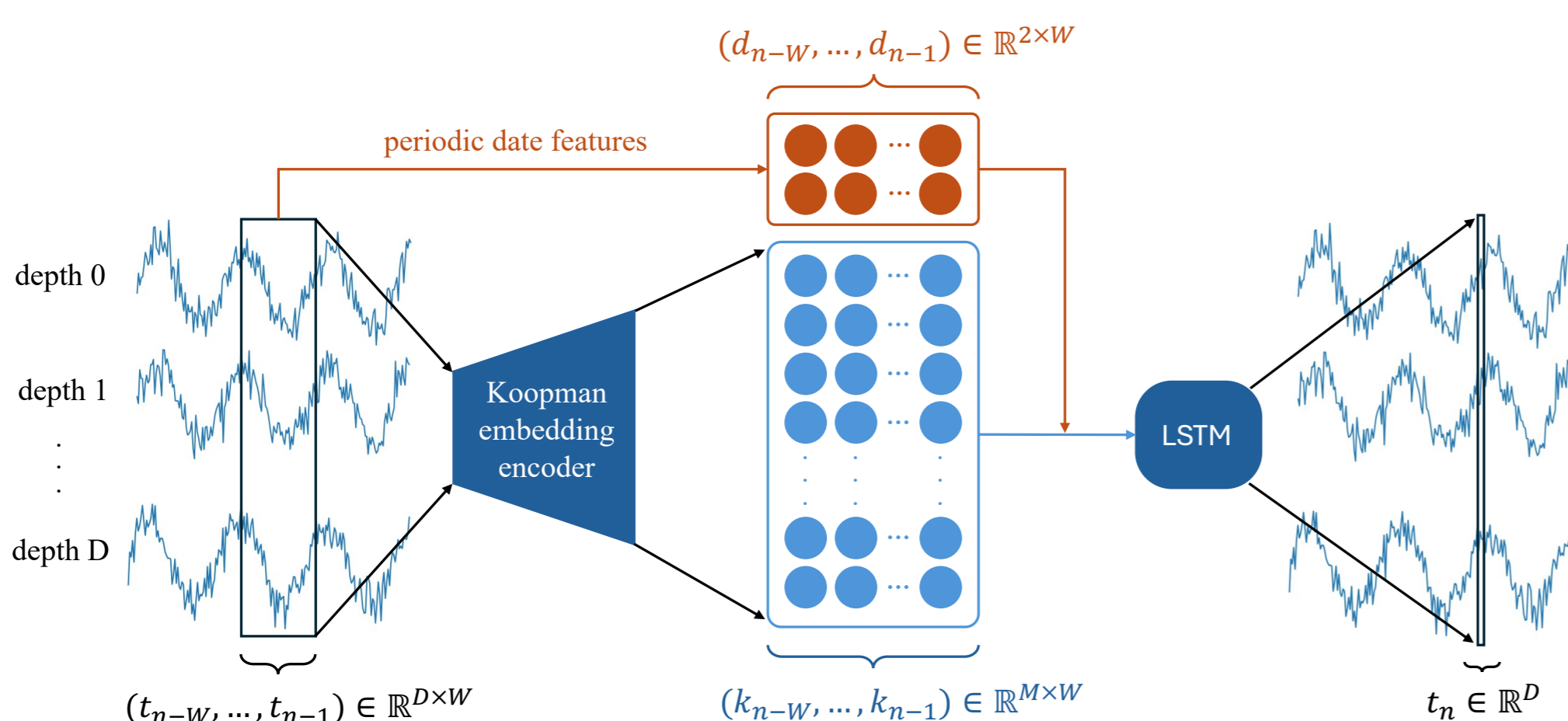


Figure 1. An illustration of our physics-informed model for multi-depth lake temperature prediction.

Training & Prediction

- We train the model by minimizing the loss function $\mathcal{L}(\theta) = \sum_{i=W+1}^N \|\hat{t}_i - t_i\|^2 / (N - W)$ using optimizers like SGD or Adam.
- After training, the model predicts a new time series $(\tilde{t}_1, \tilde{t}_2, \dots)$ by using the previous W predictions.
- The first W values are given as seed values for prediction.

Koopman embedding encoder

- The Koopman embedding encoder is part of a model with an encoder and decoder neural network.
- It transforms physical state data into high-dimensional embedding vectors for deep learning models.
- The model uses the linear Koopman operator to simplify nonlinear systems and capture dynamic evolution.
- The encoder maps physical states (e.g., multi-depth lake temperatures) into an M -dimensional embedding space.
- The decoder reconstructs the original states, but only the encoder is used in our physics-informed LSTM model.

Dataset

- We use data from Willard et al. (2021) on 450 Midwestern U.S. lakes, including attributes like lake depth, surface area, and water temperature at multiple depths.
- The dataset spans 40 years (1980–2019), but we use a subset with 14,600 daily records for a single lake due to its extensive temperature records.
- We generate temperature data using the PBO model, which estimates thermal dynamics based on meteorological inputs and lake attributes.
- While the dataset includes temperature profiles at 50 depths, we use only data at 0, 1, and 2 meters in this study.
- We use the first 32 years of data for model building, the next 4 years for validation, and the last 4 years for testing.

Experiments & Results

Experiment Setup

- Training: 32 years (1980–2011).
- Validation: 4 years (2012–2015).
- Testing: 4 years (2016–2019).
- Each year has 365 time steps, with February 29 removed from leap years.

Koopman Embedding Model

- Architecture: 3D input to 32D embedding space with an encoder and decoder.
- Optimized using Adam for 70 epochs with learning rate decay.
- Loss function includes reconstruction loss, Koopman dynamics loss, and a decay term.

LSTM Training

- Frozen Koopman encoder used to train LSTM with four layers of 64 hidden nodes each.
- Optimized with Adam (learning rate 0.01) for 250 epochs.
- First 4 days' temperatures used as seed for predictions on the test series.

Baseline Comparison

- LSTM without Koopman embeddings (learning rate 0.001) trained as a baseline.
- Both models evaluated with mean squared error (MSE) on the test series.

Results

- LSTM+Koopman outperforms the baseline, reducing MSE from 35.682 to 10.800 (70% improvement).

Table 1. Comparison of test MSEs of the LSTM baseline and our model.

Depth (m)	LSTM	LSTM+Koopman (ours)
0	36.306	11.120
1	35.871	10.890
2	34.869	10.390
Overall	35.682	10.800

Conclusion & Future Work

- Used physics-informed neural networks with Koopman embeddings to predict lake water temperatures at different depths.
- Achieved a 70% reduction in mean squared errors.
- Future work will combine in-situ and simulated data, explore transformers (Vaswani et al., 2017), and study other water quality indices.

Acknowledgements

Our work is supported by the following grants: NSF (IIS-2024733, IIS-2331908), ONR (N00014-23-1-2789), DHS (23STSLA00016-01-00), DoD (78170-RT-REP), FDEP (INV31).