Zero-Shot Learning for Salinity Forecasting

Khoa Nguyen[1], Khoa Nguyen Tho Anh[1], Vinh Dinh Quang[1], Vuong Nguyen Dinh[2], Ngoc Hong Tran[1,3], and Binh Doan Van[1,3]

[1] Vietnamese-German University, Binh Duong, Vietnam
[2] Southern Institute of Water Resources Research, Ho Chi Minh, Vietnam
[3] Co-corresponding author

Abstract. Salinity forecasting is vital for ensuring sustainable aquatic ecosystems, optimizing agricultural practices, and safeguarding water resources from the adverse impacts of salinization. In time series forecasting, limited data often poses significant challenges. This research specifically addresses this constraint in the context of salinity forecasting. Central to our approach is the use of Zero-Shot Learning, a self-supervised technique that allows models to make predictions in scenarios they have never seen during training. Following this, we employ Transfer Learning, fine-tuning the model using the weights obtained from Zero-Shot Learning to enhance its forecasting accuracy further. Code is available at: https://github.com/neko941/Self-Supervised-Learning-for-Salinity-Forecasting

Keywords: salinity forecasting · time-series forecasting · domain generalization.

1 Introduction

Plants face various abiotic stressors such as salinity, drought, extreme temperatures, and freezing throughout their lifecycle, which can significantly hinder their growth and agricultural yield, sometimes leading to crop failure [1]. Salinity, caused by high soluble salt concentrations in the soil [2], is a particularly pressing issue in arid regions with insufficient rainfall to remove salts from the root zone. Addressing this problem includes the application of deep learning, a subset of machine learning using artificial neural networks. To accurately predict soil salinity fluctuations, salinity forecasting can empower farmers to protect their yields, optimize resource allocation, and mitigate economic losses. However, the limited data availability in this domain emphasizes the importance of employing advanced technological approaches. This work explores the potential of self-supervised learning, specifically zero-shot and transfer learning, to maximize the utility of the limited salinity dataset (SD).

Our contributions can be summarized as:

- Collected, cleaned and transformed the SD into a correct and more conducive format for training purposes.

- First instance where deep learning has been employed for salinity prediction in this specific region (Mekong Delta).
- Produced a pre-trained model for time-series forecasting using the zero-shot technique for later usage.
- Employed transfer learning on the SD.

2 Related Works

While machine learning excels in applications like salinity forecasting, its limitations with sparse datasets pave the way for the introduction of zero-shot learning. Zero-shot learning is a specialized machine learning approach designed to make predictions in scenarios without previously seen data points for certain classes, bridging the gap in traditional models that require extensive data. However, the literature on zero-shot learning for time series forecasting is notably sparse. In [3], the authors introduced an RNN-based model, Memory-endowed Ordinal Regression Deep Neural Networks (MOrdReD), capable of deducing complete predictive distributions for time-series forecasting tasks with minimal or no training data, termed as few-shot and zero-shot learning respectively.

Recent advancements in salinity forecasting have seen the adoption of sophisticated models that integrate deep learning techniques with environmental data analysis. Notably, the T-GCFN model utilizes a graph-based convolutional approach to capture spatial dependencies and address distribution shift challenges in marine time series data [14]. Similarly, the use of Bi-LSTM models, enhanced by partial mutual information selection, offers a robust framework for predicting groundwater salinization levels by effectively managing the complexities inherent in the heterogeneous distribution of fresh-saline groundwater [15]. Furthermore, the introduction of Res-LSTM and Res-GRU models, specifically tailored for the Sacramento–San Joaquin Delta, exemplifies the application of residual neural networks in capturing the spatial and temporal variations of salinity with high accuracy [16]. Additionally, the CPSOCGSA-tuned neural processor presents an innovative approach to river water salinity forecasting by combining data preprocessing techniques with an optimized artificial neural network, showcasing the potential of hybrid models in achieving high precision in salinity predictions [17].

The field of time series analysis has witnessed significant advancements in recent years, particularly with the emergence of Transformer-based solutions. Among the most notable models that address the complex and less explored long-term time series forecasting (LTSF) problem are LogTrans [9] (presented at NeurIPS 2019), Informer [8] (AAAI 2021 Best paper), Autoformer [10] (NeurIPS 2021), Pyraformer [12] (ICLR 2022 Oral), Triformer [11] (IJCAI 2022), and the recent FEDformer [13] (ICML 2022).

Recent breakthroughs in time series forecasting have been marked by the introduction of NLinear and DLinear models, which, despite their linear foundations, have yielded impressively enhanced results in long-term forecasting [4]. NLinear, a variation of the foundational Linear model, incorporates a normalization step to mitigate the impact of distribution changes in the dataset, ensuring predictions align with the actual value distribution. On the other hand, DLinear employs a decomposition scheme, breaking input data into trend and seasonal components, enhancing predictive accuracy, especially in datasets with pronounced trends. These advancements signify the continuous evolution of time series forecasting methods, aiming to address specific challenges and improve prediction accuracy. Opting for NLinear and DLinear models in this work, given their proven superiority in accuracy over traditional Transformer-based methods, is a strategic choice for this work.

3 Proposed Dataset

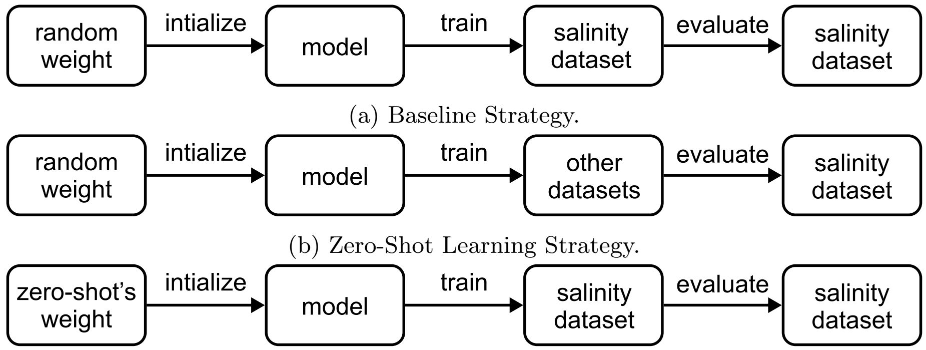
The SD provides a detailed collection from 43 monitoring stations in the Mekong Delta provided by Dr.Binh Doan Van and Dr.Vuong Nguyen Dinh. These stations are strategically dispersed along prominent rivers and waterways, such as the Sai Gon, Đong Nai, Vam Co, Tien, Co Chien, and Hau Rivers. The stations in the Mekong Delta provinces, including Long An, Tien Giang, Tra Vinh, Vinh Long, Soc Trang, Ben Tre, Kien Giang, Bac Lieu, and Ca Mau, provide insights into the delta's unique aquatic ecosystem. To accurately assess the salinity levels in the Mekong Delta, the dataset utilizes a combination of on-site water sampling, direct field measurements with portable devices, and detailed laboratory analyses using automatic salinity meters. The dataset spans from 1995-01-02 to the end date of 2020-06-30, covering every month from January to December from 1995 to 2020, with one-day granularity. Encompassing a comprehensive collection, the dataset contains a total of 60,207 samples. However, this dataset exhibits a notable challenge with a high missing rate of 79.72%. This significant portion of missing data presents a unique opportunity for leveraging the knowledge of other domains by utilizing zero-shot learning.

4 Proposed Training Strategies

Our strategies set our framework apart from traditional supervised learning by emphasizing adaptability and versatility, allowing for integration with a variety of methods. This approach enables application in diverse scenarios, overcoming the limitations of relying on large, labeled datasets. The flexibility of our strategies enhances the potential for innovation in applying supervised learning across different domains.

Baseline Strategy: The baseline approach begins with a model whose randomly initializes weights. This model is then trained exclusively on the training set of the SD. After training, the model's performance is evaluated using the best weights on the test set of the SD.

Zero-shot Learning Strategy (ZSL): In the ZSL approach applied to the SD, the model begins with randomly initialized weights. It undergoes training on multiple datasets (Table 1) from diverse domains such as finance, climate, and electricity, each having a distinct domain from the SD. Despite the varying



(a) Baseline Strategy.

(b) Zero-Shot Learning Strategy.

(c) Transfer Learning Strategy.

Fig. 1: The figure illustrates the three distinct training strategies: a) Baseline, b) Zero-shot Learning, and c) Transfer Learning.

granularity of these datasets, a uniform time step is adopted for training to maintain consistency. Furthermore, given the different number of variables in each dataset, the model is trained using all possible combinations of two variables (from the n variables of each dataset), aligning the training input format with the salinity test set. This strategic approach results in an expansive training framework, generating over 35 million training samples for each test scenario. Remarkably, although the model is not directly trained on the SD, its optimally trained weights are employed to assess its performance on the test set of the SD, showcasing the model's ability to generalize and apply learned knowledge to new, unseen data in the domain of Oceanography.

Table 1: Dataset Metadata After Transformed

Dataset	**Domain**	**Granularity***	**Samples**	**Missing Rate**	**Variates**
Salinity	Oceanography	1,440	60,207	79.72	2
Crypto [5]	Finance	1,440	1,048,575	47.46	8
DTJ [6]	Climate	1,440	67,875	2.46	3
WSB [7]	Climate	10	778,047	0	21
WSS [7]	Climate	10	1,038,630	0	30
ETTm1 [8]	Electricity	15	69,680	0	7
ETTm2 [8]	Electricity	15	69,680	0	7
ETTh1 [8]	Electricity	60	17,420	0	7
ETTh2 [8]	Electricity	60	17,420	0	7

*: in minute(s).

Transfer Learning Strategy (TL): Transfer learning leverages the knowledge gained from the zero-shot learning phase. The model's weights are initialized using those obtained from the zero-shot learning strategy. It is then fine-tuned on the training set of the SD. After this training, the model's performance is assessed using the best weights on the test set of the SD.

To ensure fair and consistent evaluation, all strategies applied to the SD are assessed using the same test set, allowing for accurate comparisons. The methodologies of these strategies are visually represented in Figure 1, providing a clear comparative overview.

5 Experiments

5.1 Data Pre-processing

Data pre-processing is a critical step in machine learning and data analysis pipelines, as it significantly impacts model performance. Raw data often contains noise, outliers, and inconsistencies that can affect results. Pre-processing ensures clean, consistent, and relevant data, leading to accurate outcomes. The technique used in this project is Standard Scaler, which standardizes data by subtracting the mean and dividing by the standard deviation, resulting in a mean-centered distribution. This scaling benefits algorithms like k-means and k-nearest neighbors, particularly in zero-shot learning, where mean-centering improves convergence and comparability between datasets. It also introduces regularization, preventing overfitting and enhancing model generalization. This transformation ensures that the feature values follow a standard normal distribution.

5.2 Implementation Details

We chose linear regression as our foundational model for its simplicity, interpretability, and computational efficiency, making it ideal for time series forecasting. We extended this foundation with NLinear and DLinear models, advanced linear variants that excel in handling complex data patterns. Additionally, we utilize one of the recent models for predicting salinity, represented by the Res-LSTM model. This demonstrates the adaptability of our method across various models within the supervised learning framework. All models have been slightly adjusted to align with our specific output requirements.

Our salinity forecasting project employed the Adam optimizer with a learning rate of 0.1, targeting Mean Squared Error (MSE) loss. The data sorted in chronological order is first transformed into training samples using lag (input or historical information) and ahead (output or future information to be predicted). Then it is continuously split into training, validation, and test sets (the first 70% of training samples of SD, the next 10%, and the last 20% ratios, respectively). The training lasted 10,000,000 epochs with early stopping (patience: 20, min_delta: 0.00001) based on validation loss. Moreover, transfer learning test cases with a patience setting of 100 demonstrate that extending the training duration can enhance performance (in **Transfer*** in Table 2). To enhance learning, we dynamically reduced the learning rate on plateaus (factor:

0.1, patience: 4, min_delta: 0.0001). This adaptive approach prevented overfitting on large datasets and ensured efficient model convergence while capturing essential knowledge.

5.3 Results

Table 2: Results of the Conducted Cases. The **best results** are highlighted in **bold** and the second-best results are highlighted with an underline. All the cases are trained with patience 20 except for the **Transfer*** columns (with patience 100) whose results are highlighted with an *italic*.

Ahead	Model	Baseline		Zero-shot		Transfer		Transfer*	
		R2	MSE	R2	MSE	R2	MSE	R2	MSE
1	Linear	-0.3	17.687	0.911	1.209	**0.931**	**0.932**	*0.94*	*0.82*
	NLinear	-22.157	314.92	0.555	6.047	**0.922**	**1.06**	*0.922*	*1.06*
	DLinear	-18.661	267.379	0.928	0.974	**0.932**	**0.919**	*0.936*	*0.867*
	Res-LSTM	-5.486	88.207	0.795	2.785	**0.827**	**2.357**	*0.842*	*2.154*
2	Linear	-2.771	51.147	0.842	2.126	**0.848**	**2.046**	*0.908*	*1.234*
	NLinear	-0.066	14.371	0.898	1.373	**0.912**	**1.182**	*0.919*	*1.09*
	DLinear	-4.277	70.522	0.899	1.352	**0.92**	**1.069**	*0.921*	*1.061*
	Res-LSTM	-0.309	17.667	0.808	2.586	**0.809**	**2.564**	*0.809*	*2.564*
4	Linear	-67.9117	922.754	0.793	2.741	**0.85**	**1.982**	*0.862*	*1.826*
	NLinear	-12.662	183.731	0.823	2.334	**0.846**	**2.028**	*0.861*	*1.837*
	DLinear	-14.215	203.265	0.814	2.452	**0.856**	**1.907**	*0.873*	*1.683*
	Res-LSTM	-1.255	30.03	0.704	3.953	**0.757**	**3.248**	*0.757*	*3.248*
8	Linear	0.635	5.139	0.633	5.041	**0.732**	**3.717**	*0.772*	*3.15*
	NLinear	-0.065	15.275	0.653	4.721	**0.757**	**3.311**	*0.773*	*3.11*
	DLinear	0.715	3.942	0.659	4.637	**0.77**	**3.155**	*0.781*	*3.006*
	Res-LSTM	-3.46	62.39	0.599	5.65	**0.717**	**4.003**	*0.719*	*3.978*
16	Linear	-0.44	20.739	0.439	7.346	**0.669**	**4.492**	*0.724*	*3.759*
	NLinear	-0.2665	18.606	0.389	7.87	**0.596**	**5.124**	*0.655*	*4.392*
	DLinear	0.678	4.285	0.4	7.699	**0.698**	**4.011**	*0.698*	*4.011*
	Res-LSTM	0.334	9.425	**0.514**	**6.719**	0.496	7.37	*0.62*	*5.283*

Our salinity forecasting study evaluated the baseline Linear model alongside advanced NLinear and DLinear models over various forecasting horizons. We focused on key metrics: R-squared (R2) and MSE. Our experiments spanned several configurations, initially showing that all three models could accurately forecast in diverse settings. However, incorporating a historical lag of 128-time steps significantly reduced their performance. To overcome this, we employed innovative ZSL and TL strategies. Results can be found at Table 2.

For short-term forecasts (1-4 days ahead), the baseline strategy shows a non-positive R2 value and an MSE of at least 14. In contrast, ZSL demonstrates significant improvement, with R2 values mostly above 0.79, except for NLinear at a 1-day ahead prediction, where R2 is 0.555. The MSE for ZSL ranges from 1 to 6, indicating a notable enhancement in prediction accuracy compared to the baseline.

When extending the prediction to 8 and 16 days ahead, ZSL effectively leverages the properties of NLinear, resulting in a substantial improvement over the baseline strategy. However, strategies like Linear and DLinear, which do not inherently normalize within the model, show ZSL results slightly smaller than the baseline. Generally, ZSL outcomes are expected to converge with or surpass the baseline performance.

However, TL introduces a significant advancement, especially considering the integration of generalized knowledge and domain-specific insights (salinity). In the 1, 2, and 4-day ahead predictions, ZSL already demonstrates the strength of generalization. TL does not exhibit a substantially superior performance in these cases compared to ZSL. Conversely, the scenario shifts significantly for longer-term predictions (8 and 16 days ahead). While ZSL struggles with these extended forecasts, TL capitalizes on its existing knowledge base. By leveraging pre-learned and domain-specific information, TL achieves much better results than ZSL. This indicates that the depth and specificity of TL's knowledge become increasingly important for more extended forecasts, allowing it to deliver more accurate and reliable predictions.

Aligned with previous findings, the Res-LSTM model, developed for salinity forecasting in the Sacramento–San Joaquin Delta of California, also demonstrates significant enhancements when applying ZSL and TL. This indicates that the proposed strategies can be adapted for use with any supervised learning model.

Overall, while ZSL shows marked improvements over the baseline, especially in the short term, TL emerges as the most robust strategy, excelling in short- and long-term forecasts by effectively utilizing pre-existing and domain-specific knowledge.

6 Conclusion

Salinity forecasting is crucial for sustainable agriculture and the protection of aquatic ecosystems, particularly in regions like the Mekong Delta. Recognizing the importance of accurate data, we meticulously collected and optimized the SD for this specific area. This research marks the first instance where deep learning has been employed for salinity prediction in the Mekong Delta. Our approach leverages Zero-Shot Learning and Transfer Learning by addressing the challenges of limited data. These innovative self-supervised techniques enable accurate predictions in previously unseen scenarios, providing a significant step forward in the application of advanced learning methods for environmental monitoring and management.

References

1. Krasensky, J., Jonak, C.: Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. Journal of experimental botany, 63(4), 1593–1608 (2012)

2. Ali, Mostafa, F.Z.: The determinants of salinity tolerance in maize (Zea mays L.). Grafimedia, Facilitair Bedrijf RUG (2011)
3. Orozco, B. P., Roberts, S. J.: Zero-shot and few-shot time series forecasting with ordinal regression recurrent neural networks. Clinical Orthopaedics and Related Research (2020)
4. Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are Transformers Effective for Time Series Forecasting? In: Thirty-Seventh Association for the Advancement of Artificial Intelligence, pp. 11121-11128. Washington, DC, USA (2023)
5. Crypto Data Download, `https://www.cryptodatadownload.com/`, last accessed 2023/08/05
6. Temperature time series recorded in Jena since 1821 based on daily observations, `https://www.bgc-jena.mpg.de/~martin.heimann/weather/weather_temperature/`, last accessed 2023/07/08
7. Documentation of the Weather Station on Top of the Roof of the Institute Building of the Max-Planck-Institute for Biogeochemistry, `https://www.bgc-jena.mpg.de/wetter/Weatherstation.pdf`, last accessed 2023/07/08
8. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W.: Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In: Thirty-Third Conference on Innovative Applications of Artificial, pp. 11106-11115. Virtual Event (2021)
9. Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X., Yan, X.: Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. In: Neural Information Processing Systems, pp. 5244-5254. Vancouver, BC, Canada (2019)
10. Xu, J., Wang, J., Long, M., et al.: Autoformer: Decomposition transformers with autocorrelation for long-term series forecasting. Neural Information Processing Systems, pp. 22419–22430, Curran Associates, (2021)
11. Cirstea, R.-G., Guo, C., Yang, B., Kieu, T., Dong, X., Pan, S.: Triformer: Triangular, variable-specific attentions for long sequence multivariate time series forecasting. In: International Joint Conference on Artificial Intelligence, pp. 1994-2001, IJCAI, Vienna (2022)
12. Liu, S., Yu, H., Liao, C., Li, J., Lin, W., Liu, A. X., Dustdar, S.: Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In: The Tenth International Conference on Learning Representations (2022)
13. Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., Jin, R.: Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In: Proceedings of the 39th International Conference on Machine Learning, vol. 162, pp. 27268-27286. PMLR, USA (2022)
14. Li, D., Zhao, W., Hu, J., Zhao, S., Liu, S.: A long-term water quality prediction model for marine ranch based on time-graph convolutional neural network. In: Ecological Indicators, vol. 154, pp. 110782 (2023)
15. Muniappan, A., Jarin, T., Sabitha, R., Ghfar, A.A., Fattah, I.M.R., Bowa, C.K., Mwanza, M.: Bi-LSTM and partial mutual information selection-based forecasting groundwater salinization levels. In: Water Reuse, vol. 13, pp. 525-544 (2023)
16. Qi, S., He, M., Bai, Z., Ding, Z., Sandhu, P., Chung, F., Namadi, P., Zhou, Y., Hoang, R., Tom, B., et al.: Novel Salinity Modeling Using Deep Learning for the Sacramento–San Joaquin Delta of California. In: Water, vol. 14 (2022)
17. Khudhair, Z.S., Zubaidi, S.L., Al-Bugharbee, H., Al-Ansari, N. and Ridha, H.M.: A CPSOCGSA-tuned neural processor for forecasting river water salinity: Euphrates river, Iraq. In: Cogent Engineering, vol. 9, pp. 2150121 (2022)