HIDRA-T – A Transformer-Based Sea Level Forecasting Method

Marko Rus[1,2], Anja Fettich[2], Matej Kristan[1], Matjaž Ličer[2]

[1]*University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, 1000 Ljubljana, Slovenia*
[2]*Slovenian Environment Agency, Vojkova 1b, 1000 Ljubljana, Slovenia*
E-mail: marko.rus@fri.uni-lj.si

HIDRA-T – transformerski model za napovedovanje višine morske gladine

Sea surface height forecasting is critical for timely prediction of coastal flooding and mitigation of is impact on coastal comminities. Traditional numerical ocean models are limited in terms of computational cost and accuracy, while deep learning models have shown promising results in this area. However, there is still a need for more accurate and efficient deep learning architectures for sea level and storm surge modeling. In this context, we propose a new deep-learning architecture HIDRA-T for sea level and storm tide modeling, which is based on transformers and outperforms both state-of-the-art deep-learning network designs HIDRA1 and HIDRA2 and two state-of-the-art numerical ocean models (a NEMO engine with sea level data assimilation and a SCHISM ocean modeling system), over all sea level bins and all forecast lead times. Compared to its predecessor HIDRA2, HIDRA-T employs novel transformer-based atmospheric and sea level encoders, as well as a novel feature fusion and regression block. HIDRA-T was trained on surface wind and pressure fields from ECMWF atmospheric ensemble and on Koper tide gauge observations. Compared to other models, a consistent superior performance over all other models is observed in the extreme tail of the sea level distribution.

1 Introduction

Anthropogenic climate changes are causing a global mean sea level rise. In part, this reflects in a world-wide coastal flooding frequency increase and leads to a variety of negative consequences for coastal communities, civil security, and the economy [2]. Shallow semi-enclosed coastal regional basins like Northern Adriatic (North Central Mediterranean Sea) are thus facing growing threats of coastal inundation and erosion [2], seawater intrusions in freshwater reservoirs and are worsening the conditions for marine traffic. Northern Adriatic ports like Venice, Koper and Trieste, but also other cultural landmark towns like Chioggia or Piran, have been – or will be – forced to take expensive preventive measures to mitigate their exposure.

Local governments and emergency responders thus require accurate prediction of short-term sea level changes such as surges, causing floods, to take proactive measures for protecting communities from flooding. Traditional methods [7, 9] are based on elaborate physics models, which, however, are prone to modeling errors and require substantial computational resources. Machine learning models are also computationally intensive to train, however, the inference is numerically cheap. For instance, single-point Koper sea level prediction from the neural network HIDRA2 [8] is a million times faster than from the full-basin operational NEMO ocean model [7].

This paper presents HIDRA-T, a novel architecture for sea level forecasting. The new model extracts relevant information from different spatial locations in the atmosphere signal and predicts the sea surface height (SSH) with a three-days horizon at significantly better accuracy than its predecessor HIDRA2 [8] as well as two state-of-the-art physics-based numerical ocean models.

The paper is organized as follows. Section 2 presents related work, Section 3 details the new HIDRA-T architecture. Section 4 reports the evaluation of the HIDRA-T architecture and provides comparisons with the state-of-the-art numerical ocean models. Conclusions and outlook are drawn in Section 5.

2 Related Work

The key difficulty of sea level forecasting in the shallow seas like Adriatic arises from high sensitivity of total sea level to the phase lag between the gravitationally generated tides (independent from meteorological forcing) and meteorologically generated basin seiches (independent from gravitational forcing). This requires the simulation of complex models covering the entire basin [7].

To avoid the high numeric cost of ensemble sea level forecasting, computationally efficient machine-learning-based ensemble models have recently been explored [12]. The early machine-learning approaches [4] were based on classic machine learning models such as support vector machines with radial basis function kernels. Authors of [5] used long short-term memory (LSTM) networks together with several atmospheric variables to improve one-hour prediction into the future but did not expand the prediction horizon. Authors of [1] predicted further in time by applying a combination of LSTMs and convolutional neural networks, but at a very coarse level. Autoregressive neural networks were considered in [3] to increase the temporal resolution and the prediction horizon.

In 2021, a convolutional neural network HIDRA1 [12]

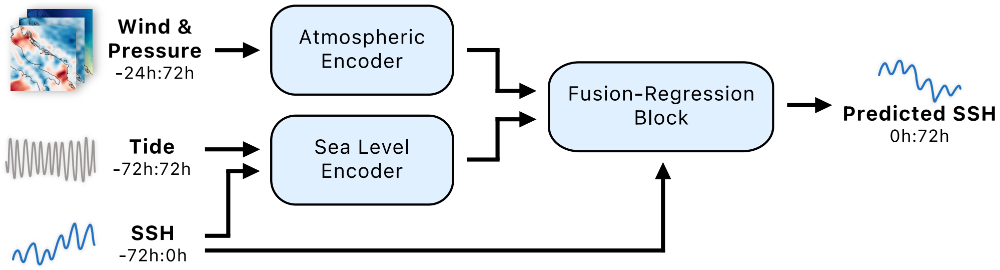

Figure 1: The HIDRA-T architecture. The Atmospheric Encoder and Sea Level Encoder embed the input data with a transformer. Both atmospheric and sea level features are then fused with the past 72 h SSH and regressed into the final SSH predictions by the Fusion-Regression Block.

with a specialized architecture to utilize atmospheric data, sea surface heights and astronomic tides was proposed. But while HIDRA1 performed favorably in comparison to the NEMO [7] model used in that study, it failed to beat the NEMO setup at very high and very low ends of sea level distributions. Recently, we proposed a continuation of HIDRA1, named HIDRA2 [8]. HIDRA2 differs from its predecessor in many ways, HIDRA2 uses 1D convolutional layer to fuse information from different temporal and spatial positions of the atmosphere, enabling weights sharing between different prediction points. HIDRA1 predicts the difference between sea level height and the astronomic tide, while HIDRA2 directly predicts the full sea level height. This and other improvements enabled HIDRA2 to outperform HIDRA1 at high sea levels for over 25 %. Due to the generalization capabilities of HIDRA2, it is also a foundation of HIDRA-T architecture.

Sequence-based models represent an attractive architectural choice for the sea level modeling as the input and output data are sequential. Transformers [10] showed dominance in sequence-based tasks, firstly in natural language processing tasks as text classification, machine translation, and question answering, and have also become the dominant methodology in computer vision. Since the atmospheric data can be interpreted as a sequence of images, we utilize the transformers in HIDRA-T to encode atmospheric as well as sea level input data to get contextually rich features, which are then used to predict sea level within the prediction horizon.

3 HIDRA-T

HIDRA-T architecture is shown in Figure 1. The input data is encoded by two encoders. The wind and pressure sequences are processed by the Atmospheric Encoder (Section 3.1), while the sea level and the tidal signal are encoded by the Sea Level Encoder (Section 3.2). The outputs of both encoders are fused with the past 72 h SSH and regressed into the final SSH hourly predictions for the future 72 h by the Fusion-Regression Block (Section 3.3). A single prediction run of HIDRA-T model creates a 72-hour sea level forecast. Although HIDRA-T would be capable of forecasting further into the future with more atmospheric forecasts, we have decided to maintain the 72 h forecast horizon to ensure that the model's performance remains comparable to that of numerical models that also produce a 72 h forecast. The aforementioned individual blocks are detailed in the following subsections.

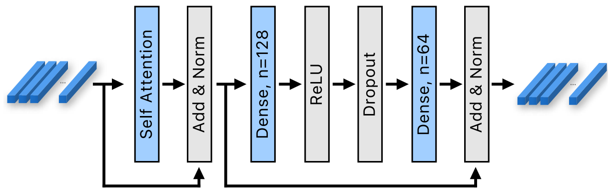

Figure 2: Transformer encoder layer used for encoding atmospheric and sea level data [11]. Tokens are firstly processed by self attention, then by two fully connected layers with 128 and 64 output features (denoted by n). ReLU, drop-out, skip connections and layer normalization are applied as shown.

3.1 Atmospheric Encoder

The atmospheric data for the Adriatic basin at a given time-step is represented by a 57×73 spatial grid with three channels, one for pressure and two for wind. HIDRA-T uses only coarse spatial resolution of atmospheric data, so the data is first downsampled to 9×12 grid by averaging before passing it to the Atmospheric encoder.

The Atmospheric Encoder is composed of two stages. In the first stage, time independent features are extracted from the atmospheric data. The input sequence of 96 h (the past 24 h and future 72 h) is firstly divided into 24 groups of four consecutive hours, which are then processed separately. Each group is processed by the transformer encoder [11] (see Figure 2) as follows. Each spatial position represents one token, which is firstly projected into a 64-dimensional space by a fully connected layer. Fixed spatial positional encoding [11] is added in the form of sine and cosine functions of different frequencies [11]. Then, two layers of single-head self attention and feed-forward network are applied. A drop-out layer with drop-out rate 0.9 is added to both self attention and feed-forward network. As the final step of the first stage, spatial dimensionality is reduced to $1 \times 1 \times 64$ by a fully connected layer, outputting a single vector. Note, however, that 24 independent passes, corresponding to 24 groups of four consecutive hours, are performed in parallel for the entire atmospheric input sequence, so the final output has dimensionality 24×64.

The second stage of the Atmospheric Encoder extracts the temporal atmospheric features by considering

Table 1: Performance of HIDRA-T, HIDRA1 [12], HIDRA2 [8], NEMO [7] and SCHISM [9] over all sea level bins (the *Overall* columns) and only during storm tide events (*Storm tide events* columns). Tidal forecast is included for reference.

	Overall				Storm tide events						
	MAE [cm]	RMSE [cm]	Bias [cm]	Acc [%]	MAE [cm]	RMSE [cm]	Bias [cm]	Acc [%]	R [%]	P [%]	F1 [%]
Tide	13.82	18.86	-5.13	47.45	55.75	59.45	-55.75	0.00	0.00	/	/
NEMO	6.54	8.52	-1.23	79.14	13.03	17.09	-11.24	49.68	63.58	**100.00**	77.73
SCHISM	5.57	7.50	**0.20**	85.06	11.04	14.70	-6.19	57.63	78.81	89.47	83.80
HIDRA1	4.72	6.73	-0.26	90.04	12.95	17.65	-10.66	53.76	74.17	94.12	82.96
HIDRA2	4.12	**5.82**	0.21	**92.89**	9.77	14.07	-5.99	64.52	84.11	91.37	87.59
HIDRA-T	**4.11**	5.83	0.28	92.88	**9.48**	**13.73**	**-5.85**	**66.45**	**86.49**	92.09	**89.20**

the wind-pressure features extracted by the first stage. Two transformer encoder layers (see Figure 2) are applied to the 24 tokens of size 64, fusing information from different time points. Finally, number of features in each token is halved from 64 to 32 by a fully connected layer, and the output is reshaped into the final 1×768 dimensional atmospheric vector $\mathbf{v}_a$.

3.2 Sea Level Encoder

The Sea Level Encoder takes the past 72 h SSH measurements and past 72 h and future 72 h of tide (tidal forecasts) as the input. The inputs are stacked into a single 144×2 tensor, where the unknown future SSH values are set to zero. The input tensor is embedded using a single 1D convolution with kernel size 2×3, stride 2 and output dimension for each time point of 64. The operation produces 71 tokens, which are processed with two transformer encoder layers (see Figure 2), producing a 71×64 output tensor. Dimensionality of the output is reduced in two ways, firstly, number of features is reduced from 64 to 16 by a fully connected network, then the number of tokens is also reduced from 71 to 16 by another fully connected layer. The output is then reshaped into the final 1×256 dimensional sea level encoding vector $\mathbf{v}_s$.

3.3 Fusion-Regression Block

The atmospheric embeddings $\mathbf{v}_a$ and sea-level embeddings $\mathbf{v}_s$ from the Atmospheric and Sea Level Encoders are concatenated and mixed by a fully connected layer, reducing the dimensionality from 1024 to 512. The obtained 512-dimensional domain context feature vector thus contains rich atmospheric and sea-surface height information from all time points and all parts of the input domain.

While the encoding and mixing operations extract the domain context, the explicit surface height information might not be well retained in the extracted feature vector. To re-inject this information, the obtained domain context feature vector is concatenated with the timeseries of past observed SSH before passing to the final regression block. The latter is composed of two fully connected layers with 584 units, SELU activations and residual connections, followed by a fully connected layer with 72 outputs for the 72 h prediction horizon.

4 Experimental Results

4.1 Training and Evaluation Dataset

We used the same training and evaluation datasets as in HIDRA2 [8]. Training is performed using the data from 2006–2018. The evaluation dataset for all models is separate from the training dataset and consists of ECMWF atmospheric forecasts [6] from Adriatic and Koper water levels between June 01, 2019 and December 31, 2020. This period was chosen due to challenging conditions and unusually high flood occurrence. The ECMWF forecasts contain 50 ensemble members with a lead time of three days. The input to HIDRA2 during training is a single forecast, obtained by averaging forecasts from all ensemble members. Floods account for 0.45 % of the training dataset and 1.1 % of the test dataset.

4.2 Implementation Details

HIDRA-T is implemented in PyTorch and is trained end-to-end using mean squared error (MSE) loss between the predictions and the ground truth. We train the model using the AdamW optimizer with standard parameter values (learning rate $3\mathrm{e}{-5}$, and the running average damping parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$), and apply the cosine annealing learning schedule to gradually reduce the learning rate during training by a factor 10. The training batch size is set to 256 data samples, and the model is trained for 60 epochs. Training takes approximately 1.5 hours on a single computer with NVIDIA Geforce RTX 2080 TI graphics card, while the inference of a single 72 h prediction for one member of the atmospheric ensemble takes only 4 ms.

4.3 SSH Forecast Performance

The overall prediction performance and performance restricted to storm events are shown in Table 1, a single forecast example is shown in Figure 3. HIDRA-T outperforms HIDRA1 [12], NEMO [7] as well as SCHISM [9] overall as well as during storms, yielding a lower MAE/RMSE and higher accuracy. While HIDRA1 achieves a lower bias, its RMSE/MAE are substantially higher, HIDRA-T outperforms HIDRA1 in MAE by 12.9 % overall, and by 26.8 % during the storm tide events. HIDRA-T is comparable to HIDRA2 [8] in overall test data, but

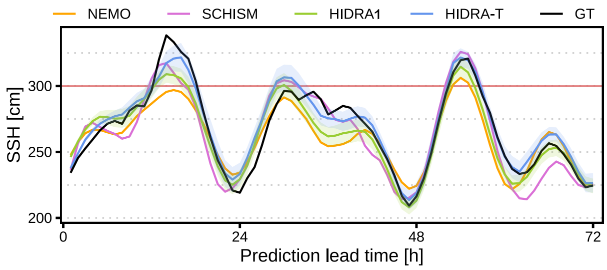

Figure 3: An example of a single forecast on a flooding event from December 2020 by HIDRA-T, HIDRA1 [12], NEMO [7] and SCHISM [9]. The coastal flood threshold is marked with a horizontal red line. Semi-transparent regions depict 2σ envelope of each HIDRA ensemble, where σ is standard deviation of ensemble predictions at each time point.

significantly outperforms it on storm surges in all measures.

NEMO achieves the highest precision of flood detection ($P = 100$ %), meaning that all detected floods are true positives. But while all NEMO's predicted floods were true, not all floods were predicted, resulting in its low recall of $R = 63.58$ %. A similar situation is observed for HIDRA1. The recalls for these two methods (NEMO: 63.58 % and HIDRA1: 74.17 %) are substantially lower than that of HIDRA-T ($R = 86.49$ %), which detects many more floods with fewer false negatives.

The excellent trade-off between the precision and recall of HIDRA-T is reflected in its F1 score (89.20 %), which is substantially higher than that of NEMO (77.73 %), HIDRA1 (82.96 %), SCHISM (83.80 %), or the next best HIDRA2 (87.59 %). This highlights HIDRA-T's exceptional ability to accurately predict sea surface heights, including extreme events such as high and low tides. In fact, HIDRA-T is next to HIDRA2 the first machine learning model to outperform numerical models in predicting extreme events, demonstrating its potential for improving coastal warning systems.

5 Conclusions

We presented a sea level prediction model HIDRA-T. Its performance is benchmarked against the current state-of-the-art Mediterranean forecasting setup of NEMO ocean model [7] (available as part of Copernicus Marine Service) and against a multi-decadal reanalysis run of the SCHISM model [9] on an unstructured grid with very high coastal resolution. We demonstrate that HIDRA-T outperforms HIDRA2 [8] as well as numerical ocean models across extreme sea level bins. With its high recall, HIDRA-T is expected to detect more extreme floods than any other model, making it a valuable tool for operational services. In addition, HIDRA-T is at inference much faster than numerical models as it can run on a personal computer, whereas numerical models require several hours on a supercomputer.

Overall, the development of HIDRA-T represents a significant milestone in the field of sea surface height prediction and provides a highly accurate and cost-effective alternative to traditional numerical models. As the effects of climate change are felt around the world, accurate sea level height prediction models will play an increasingly important role in protecting vulnerable coastal communities. With the promise of models like HIDRA-T, we can hope to have the tools we need to meet this challenge.

Zahvala

Delo je delno financirano iz ARIS programa P2-0214 in P1-0237 ter projekta J2-2506.

References

[1] Braakmann-Folgmann, A., Roscher, R., Wenzel, S., Uebbing, B., Kusche, J.: Sea level anomaly prediction using recurrent neural networks. arXiv preprint arXiv:1710.07099 (2017)

[2] Ferrarin, C., et al.: Integrated sea storm management strategy: the 29 october 2018 event in the Adriatic Sea. Natural Hazards and Earth System Sciences **20**(1), 73–93 (2020)

[3] Hieronymus, M., Hieronymus, J., Hieronymus, F.: On the application of machine learning techniques to regression problems in sea level studies. Journal of Atmospheric and Oceanic Technology **36**(9), 1889–1902 (2019)

[4] Imani, M., Kao, H.C., Lan, W.H., Kuo, C.Y.: Daily sea level prediction at Chiayi coast, Taiwan using extreme learning machine and relevance vector machine. Global and planetary change **161**, 211–221 (2018)

[5] Ishida, K., Tsujimoto, G., Ercan, A., Tu, T., Kiyama, M., Amagasaki, M.: Hourly-scale coastal sea level modeling in a changing climate using long short-term memory neural network. Science of The Total Environment **720**, 137613 (2020)

[6] Leutbecher, M., Palmer, T.: Ensemble forecasting. Tech. rep. (02 2007). https://doi.org/10.21957/c0hq4yg78

[7] Madec, G.: NEMO ocean engine (2016), `https://www.nemo-ocean.eu/wp-content/uploads/NEMO\_book.pdf`

[8] Rus, M., Fettich, A., Kristan, M., Ličer, M.: HIDRA2: deep-learning ensemble sea level and storm tide forecasting in the presence of seiches – the case of the northern Adriatic. Geoscientific Model Development **16**(1), 271–288 (2023)

[9] Toomey, T., Amores, A., Marcos, M., Orfila, A.: Coastal sea levels and wind-waves in the mediterranean sea since 1950 from a high-resolution ocean reanalysis. Frontiers in Marine Science **9** (2022)

[10] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in neural information processing systems. pp. 5998–6008 (2017)

[11] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)

[12] Žust, L., Fettich, A., Kristan, M., Ličer, M.: HIDRA 1.0: deep-learning-based ensemble sea level forecasting in the northern Adriatic. Geoscientific Model Development **14**(4), 2057–2074 (2021)