Project Report:

Neural Network Prediction of Sea Level Rise Based on Greenhouse Gas Emissions

University of Osnabrück
Faculty 8: Human Science

Submitted in the Course Foundations of Machine Learning: Theory and Practice in Summer Semester 2025

by:

Johanna Pilz
Study Program: Umweltsysteme und Ressourcenmanagement
Semester: 3
Student ID Number: 985733
Contact: jpilz@uni-osnabrueck.de

Martin Iribarren Tortone
Study Program: Cognitive Science
Semester: 4
Student ID Number: 1011094
Contact: miribarrento@uni-osnabrueck.de

Osnabrück, August 22, 2025

Contents

List of Figures

1 Introduction

Global sea level rise is one of the most significant consequences of climate change, posing a critical threat to coastal communities and ecosystems worldwide. The primary driver of this rise is the increasing concentration of greenhouse gases (GHGs) in the atmosphere, resulting from human activities. These gases trap heat, warming the planet and causing both thermal expansion of ocean water and the melting of land ice and glaciers. This project aims to explore the relationship between historical and future GHG emissions and changes in sea level. To do this, a neural network was developed to predict future sea level rise based on different GHG emission scenarios from the IPCC. The goal is to build a simple, predictive model that can serve as a tool for understanding the potential consequences of different emission trajectories.

2 Dataset Description

To build a robust model, an extensive search and consolidation of multiple data sources was performed.

Past Greenhouse Gas (GHG) Emissions Data: Three data sources were evaluated:

- **Source 1 (Intergovernmental Panel On Climate Change, 2024)**:

  Data extracted from the IPCC AR6 Synthesis Report, which also includes future projections.

- **Source 2 (Intergovernmental Panel On Climate Change, 2025)**:

  Another dataset from the IPCC AR6 that details emissions from 1990 to 2019.

- **Source 3 (Jones et al. (2023) from Our World in Data)**:

  A more extensive historical dataset on national and global emissions.

After a visual comparison , it was determined that Source 3 offered the longest historical record and a stronger alignment with both historical trends and future emission scenarios derived from Source 1. A unified historical emissions timeline was constructed by merging data from Source 3 and Source 1. The merging point was determined by identifying the year exhibiting the minimum discrepancy between the two datasets. Additionally, future emission scenarios from the IPCC AR6 were used and then interpolated with a linear relationship to have an annual frequency that matched the historical data.

Sea Level Data: Historical data on sea level rise was obtained from NOAA Climate.gov, using the University of Hawaii Fast Delivery sea level data (UHSLC), sourced via Our World in Data (Climate.gov und Philip Thompson). This data was averaged annually to serve as the target variable for model training.

Sea Level predictions: Predictions on sea level rise were obtained from the NASA Sea Level Change Portal, using the data from the IPCC AR6 Sea Level Projections (Fox-Kemper et al., 2021; Garner et al., 2021; Kopp et al., 2023).

3 Methods

3.1 Data preparation

Both GHG emission and sea level datasets were normalized using standardization to ensure stable and efficient neural network training. To prevent training bias, future scenario data was normalized using statistics derived from the historical emission data. Training data was created by pairing multi-year GHG emission inputs with corresponding historical sea level rise values. Considering the inherent time delay between GHG emissions and sea level rise — primarily due to cumulative effects on global temperature and long-term processes such as glacial/ice sheet melt — an initial input timespan of 15 years was selected. To facilitate realistic prediction of future sea level changes, the data was split chronologically into training (1970-2000), validation (2001-2007), and testing (2008-2014) sets, with a 70/15/15 ratio. This chronological split prevents data leakage and ensures that the test and validation sets represent unseen data.

3.2 Model Setup

This project employs two feedforward neural network architectures for analyzing and predicting sea level rise: a single-layer network and a two-hidden-layer network. Both networks were implemented as custom classes with randomly initialized weights and biases.

Single-Layer Network: This baseline model consists of a single output layer with a linear activation function, providing a direct relationship between input and output.

Two-Hidden-Layer Network: This network incorporates two hidden layers utilizing the Rectified Linear Unit (ReLU) activation function to capture more complex patterns.

Both networks were trained using a gradient descent-based optimization algorithm. Forward propagation involves sequentially calculating the output through each layer. In the single-layer network, this is a direct multiplication of the input by the weights, addition of the bias, and application of the linear activation. The two-hidden-layer network performs this process iteratively through each layer.

The network learns through backpropagation, adjusting weights and biases to minimize the error between predictions and actual values. A dynamic learning rate, reducing by 10% every 100 epochs, was implemented to improve convergence. The training process can be monitored by displaying 20 MSE values at even intervals.

Initial hyperparameters included an input timespan of 15 years, a training duration of 20,000 epochs, a learning rate of 0.001, and an output size of 1.

4 Analysis and Results

4.1 Hyperparameter Optimization

To optimize model performance and mitigate overfitting, we conducted a hyperparameter study focusing on the number of training epochs and GHG emission input timespan, as measured by the Mean Squared Error (MSE) on training and validation data. Initial

training was performed using thoughtfully selected, but assumption-based hyperparameters. Given the relatively small training dataset, we initially set the number of training epochs to 20,000 to prevent overfitting. To account for the variability introduced by random weight and bias initialization, each hyperparameter configuration was evaluated through 100 or 1000 independent training runs, and the MSE was averaged across these runs to provide a robust performance estimate.

4.1.1 Training Epochs - Baseline Model

Analysis of training epochs revealed an optimal point for model performance: the lowest validation MSE was achieved after approximately 35,000 epochs. While the training data MSE continued to decrease throughout the investigated interval (from 10^4 to $10^{6.5}$ epochs), the validation MSE initially decreased but began to increase after 35,000 epochs, indicating the onset of overfitting. Therefore, further training beyond this point is not recommended. Consequently, to mitigate overfitting and optimize generalization performance, subsequent analyses were conducted using a fixed training epoch of 35,000 as a new optimized hyperparameter.

4.1.2 Input Timespan - Baseline Model

The analysis of the input timespan revealed a trend mirroring that of the training epochs: an exponential decrease in training data MSE over the interval of 10 to 100 years. Conversely, the validation data MSE increased with longer input timespans, reaching a minimum at the shortest timespan of 10 years.
This finding—that a shorter timespan yielded the best validation performance—was unexpected, as we initially hypothesized a longer timespan would be necessary to capture the long-term impact of GHG emissions. The interplay between input timespan and training epochs potentially contributes to this result, suggesting the need to identify an optimal combination of both parameters. For this single-layer network, 35,000 training epochs combined with a 10-year input timespan proved most beneficial.
However, it's important to note that this model may be more sensitive to recent acceleration in sea level rise and less responsive to longer-term trends and cumulative past emissions.

4.1.3 Training epochs - Specific Two-Hidden-Layer Model

The Hyperparameter analysis for the two-hidden-layer network yielded more dispersed and less conclusive results compared to the baseline model, likely due to increased sensitivity to random initial weight and bias assignments. Despite this variability, the trend of overfitting with excessive training epochs and large input timespans remained consistent. Given this dependence on random initialization, we focused on preventing overfitting and identifying an optimal epoch count by analyzing a single, specifically initialized network, acknowledging that the findings may not be fully generalizable. This analysis revealed the lowest validation MSE at approximately 6,000 epochs, after which the training MSE continued to decline but the validation MSE began to increase, indicating overfitting. Subsequent analyses utilized a training epoch count of 6,000.

4.2 Experiment with a dynamic learning rate

To assess potential benefits of adaptive learning, we compared a baseline neural network trained with and without a dynamic learning rate. The results showed that a static learning rate consistently yielded a lower Mean Squared Error (MSE) during training. This suggests a static learning rate was more effective for this model, potentially because the implemented decay schedule (10% reduction every 100 epochs) was too aggressive, leading to premature convergence and a suboptimal solution. Further investigation of alternative decay schedules may optimize the training process.

5 Evaluation

The performance of the models was assessed by comparing results before and after hyperparameter optimization, as summarized in Table 1. The results demonstrate a moderate improvement in the performance of the **baseline model** following hyperparameter optimization. The neural network trained with the refined hyperparameters exhibits a lower MSE across the validation and testing datasets compared to the baseline model with initial parameters. Notably, the test MSE decreased substantially from 0.1777 to 0.0418. However, the MSE for the training data slightly increased with the optimization. This suggests the hyperparameter tuning process successfully identified a configuration that better generalizes to unseen data and reduces overfitting. It's important to note that differences in emission input timespans prevented the initialization of identical weights and biases across models, potentially introducing some bias due to the randomization process.

Model	Train MSE	Validation MSE	Test MSE
Baseline (Linear)	0.0134	0.0315	0.1777
2-Hidden-Layer (ReLU)	0.0087	0.0126	0.0444
Baseline new parameters	0.0142	0.0285	0.0418
2-Hidden-Layer new parameters	0.0155	0.0055	0.0778

Table 1: MSE for training, validation and test for the two models before and after hyperparameter optimization

The results of the evaluation for the **Two-Hidden Layer Model** show a complex picture. While the hyperparameter optimization led to a significantly reduced validation MSE (0.0055 compared to 0.0126 in the initial model), the test MSE almost doubled. This suggests that the optimized hyperparameters may be causing some overfitting to the validation set. This outcome is likely due to the hyperparameter selection being based on achieving the lowest possible validation MSE, which possibly led to the model capturing noise or specific patterns within the validation set that do not generalize to the broader population represented by the test set.
Comparing the performance of the baseline model with hyperparameter optimization and the two-hidden layer model without hyperparameter optimization, the results across all datasets are quite similar. The two-hidden layer model exhibits lower training and

validation MSE values compared to the baseline model, while the baseline model with optimized hyperparameters demonstrates a marginally better performance on the test set. This suggests that both models perform comparably, and the choice between them may depend on the specific priorities of the application. The baseline model with optimized hyperparameters provides a slight advantage in generalization, as indicated by its lower test MSE, while the two-hidden layer model appears to fit the training and validation data slightly better.

We further evaluated model performance by visually comparing predictions on the test data to the actual sea level changes (see Figure 1). The curves of the baseline model with hyperparameter optimization and the two-hidden layer model without hyperparameter optimization exhibit a similar trajectory, suggesting a comparable predictive capability. Notably, the standard baseline model demonstrates a particularly poor ability to capture the overall sea level trend, while the optimized baseline and 2-layer models offer marginally improved, though still inaccurate, predictions.

All three models significantly underestimate the sea level rise, indicating a systemic bias in the predictions. While the magnitude of the predicted values is almost consistently lower than the actual data, the models do appear to approximate the general slope of the sea level rise over time. This suggests the models are capturing the direction of the trend, but failing to accurately quantify the magnitude of the change. A systematic underestimation of sea level severely compromises the predictive abilities of these models. Several factors might contribute to this: a potential bias within the training data itself, an insufficient complexity of the models to fully represent the underlying processes driving sea level rise, or a limitation in the chosen input features. It's also plausible that the models struggle to account for non-linear accelerations in sea level rise that are not adequately captured by the linear relationships they are learning.

As a final evaluation step, we compared the predictions of our two-hidden-layer model (with initial hyperparameters) to future sea level projections from the Intergovernmental Panel on Climate Change (IPCC), as shown in Figure 2.

Our model's predictions consistently underestimate future sea level rise compared to both the IPCC's 1.5°C and 2.0°C warming scenarios. This underestimation was also observed during the evaluation of our model on the test set, where it systematically predicted lower sea levels than were actually observed. While our model could approximately predict the correct slope during this test, it fails to do so here. The slope of our prediction notably decreases over time, contrasting with the consistently increasing slopes shown in the IPCC projections. A key factor contributing to this discrepancy is the limited nature of our training data, which consisted exclusively of periods with rising GHG emissions and corresponding rising sea levels. This inherently restricts the model's ability to accurately predict scenarios involving stabilized or declining emissions, as represented by the IPCC's 1.5°C scenario. The model essentially learns a relationship between increasing emissions and increasing sea level, and cannot extrapolate beyond

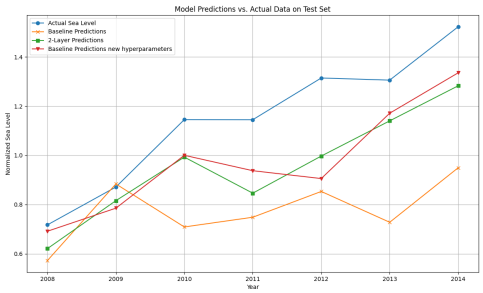

Figure 1: Comparison of model predictions against actual sea level data on the test set

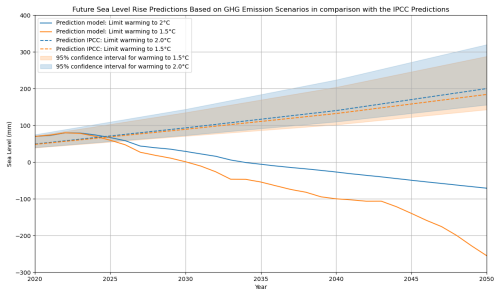

Figure 2: Comparison of model predictions against actual sea level data on the test set

that association. A contributing factor to this limitation is the relatively short input timespan used, which hinders the model's ability to capture the cumulative effects of past emissions. Furthermore, the IPCC projections benefit from more complex climate models that incorporate a wider range of feedback mechanisms, potential emission pathways and mechanistic processes, including scenarios with reduced emissions. Therefore, the observed underestimation is likely a consequence of both the limited scope of our training data and the simplified representation of climate processes compared to the sophisticated models employed by the IPCC.

6 Discussion

Our investigation into hyperparameter optimization demonstrated its potential to improve model performance, but also highlighted the risk of overfitting, particularly with the two-hidden-layer network. A significant disparity between validation and test MSE values in certain configurations served as a clear indicator of this. A key takeaway is the critical role of input data characteristics; models trained on data exhibiting primarily linear relationships struggled to capture the underlying non-linear dynamics of sea level rise. Looking forward, several improvements are possible. Regularization techniques like L2 regularization or dropout could mitigate overfitting, while transitioning to mini-batch gradient descent may enhance computational efficiency and convergence. A systematic hyperparameter search, employing methods like grid or random search, is crucial for optimizing network architecture and training parameters. However, the most impactful next step is likely data enrichment. The current reliance on a single feature limits model capacity. Integrating additional relevant data—such as global average temperature, atmospheric CO2 concentrations, and volcanic/solar activity data from sources like NASA—would significantly enhance predictive power and facilitate a more nuanced understanding of sea level change. In conclusion, while hyperparameter optimization is promising, a holistic approach—combining advanced optimization with richer data sources and sophisticated architectures—is essential for developing robust and accurate sea level rise predictions.

References

Climate.gov, NOAA ; Philip Thompson, University of H.: *Sea Level Data*. https://ourworldindata.org/grapher/sea-level.csv?v=1&csvType=full&useColumnShortNames=true. – Accessed August 29, 2024.

Fox-Kemper, B. ; Hewitt, H.T. ; Xiao, C. ; Aalgeirsdóttir, G. ; Drijfhout, S.S. ; Edwards, T.L. ; Golledge, N.R. ; Hemer, M. ; Kopp, R.E. ; Krinner, G. ; Mix, A. ; Notz, D. ; Nowicki, S. ; Nurhati, I.S. ; Ruiz, L. ; Sallée, J.-B. ; Slangen, A.B.A. ; Yu, Y. ; Masson-Delmotte, V. (Hrsg.) ; Zhai, P. (Hrsg.) ; Pirani, A. (Hrsg.) ; Connors, S.L. (Hrsg.) ; Péan, C. (Hrsg.) ; Berger, S. (Hrsg.) ; Caud, N. (Hrsg.) ; Chen, Y. (Hrsg.) ; Goldfarb, L. (Hrsg.) ; Gomis, M.I. (Hrsg.) ; Huang, M. (Hrsg.) ; Leitzell, K. (Hrsg.) ; Lonnoy, E. (Hrsg.) ; Matthews, J.B.R. (Hrsg.) ; Maycock, T.K. (Hrsg.) ; Waterfield, T. (Hrsg.) ; Yelekçi, O. (Hrsg.) ; Yu, R. (Hrsg.) ; Zhou, B. (Hrsg.): *Ocean, Cryosphere and Sea Level Change*. Cambridge, United Kingdom and New York, NY, USA : Cambridge University Press, 2021. – 1211–1362 S.

Garner, G. G. ; Hermans, T. ; Kopp, R. E. ; Slangen, A. B. A. ; Edwards, T. L. ; Levermann, A. ; Nowicki, S. ; Palmer, M. D. ; Smith, C. ; Fox-Kemper, B. ; Hewitt, H. T. ; Xiao, C. ; Aalgeirsdóttir, G. ; Drijfhout, S. S. ; Golledge, N. R. ; Hemer, M. ; Krinner, G. ; Mix, A. ; Notz, D. ; Nurhati, I. S. ; Ruiz, L. ; Sallée, J-B. ; Yu, Y. ; Hua, L. ; Palmer, T. ; Pearson, B.: *IPCC AR6 Sea Level Projections*. – Accessed [19.08.2025].

Intergovernmental Panel On Climate Change: *IPCC AR6 Synthesis Report LR Figure 2.5 (a): Global GHG emissions of modelled pathways*.

Intergovernmental Panel On Climate Change: *IPCC AR6 Synthesis Report LR Figure 2.2 (c): Global net anthropogenic GHG emissions by region (1990-2019)*.

Jones, Matthew W. ; Peters, Glen P. ; Gasser, Thomas ; Andrew, Robbie M. ; Schwingshackl, Clemens ; Gütschow, Johannes ; Houghton, Richard A. ; Friedlingstein, Pierre ; Pongratz, Julia ; Le Quéré, Corinne: National contributions to climate change due to historical emissions of carbon dioxide, methane, and nitrous oxide since 1850. In: *Scientific data* 10 (2023), Nr. 1, S. 155. – DOI 10.1038/s41597–023–02041–1.

Kopp, R. E. ; Garner, G. G. ; Hermans, T. H. J. ; Jha, S. ; Kumar, P. ; Reedy, A. ; Slangen, A. B. A. ; Turilli, M. ; Edwards, T. L. ; Gregory, J. M. ; Koubbe, G. ; Levermann, A. ; Merzky, A. ; Nowicki, S. ; Palmer, M. D. ; Smith, C.: The Framework for Assessing Changes To Sea-Level (FACTS) v1.0: A platform for characterizing parametric and structural uncertainty in future global, relative, and extreme sea-level change. In: *Geoscientific Model Development* 16 (2023), S. 7461–7489. – DOI 10.5194/gmd–16–7461–2023.