Towards a Seismology Foundation Model

Waleed Esmail[1†], Alexander Kappes[1], Stuart Russell[2] and Christine Thomas[2]

1 Institut für Kernphysik, Universität Münster, Wilhelm-Klemm-Straße 9, 48149, Münster
2 Institut für Geophysik, Universität Münster, Corrensstraße 24, 48149, Münster

† wesmail@uni-muenster.de

The 2nd European AI for Fundamental Physics Conference (EuCAIFCon2025) Cagliari, Sardinia, 16-20 June 2025

Abstract

In this article we introduce SeismoGPT, a step toward a foundation model for seismology, developed to forecast three-component seismic waveforms with direct application to future gravitational wave detectors such as the Einstein Telescope. SeismoGPT learns to predict the next token in an autoregressive framework. It can be applied to both single-station and multi-station seismic array, learning both temporal and spatial dependencies directly from raw waveform. Beyond immediate forecasting, SeismoGPT represents a step toward a general-purpose framework for seismology, one that could support Newtonian-noise mitigation, real-time observatory control, and ultimately broader seismic monitoring and prediction tasks.

Publication information to appear upon publication.

Received Date
Accepted Date
Published Date

Contents

1 Introduction 2

2 Model Architecture 2
2.0.1 Single-Station Model 2
2.0.2 Array-Based Model 2

3 Results 3

4 Conclusion 4

References 4

---

1 Introduction

Seismic noise is one of the key challenges for ground-based gravitational wave (GW) detectors, as it strongly limits sensitivity, particularly at low frequencies.

In GW detectors, seismic noise couples to the optical system in two main ways. The first is mechanical coupling, where ground motion travels through the mirror suspension and isolation systems, ultimately shaking the optical components, which is an effect that is many orders of magnitude stronger than a GW signal [1]. The second is Newtonian Noise (NN), a subtler process in which seismic waves change the density of the surrounding ground and air, creating gravitational field fluctuations that directly act on the test masses [2].

To address this challenge, especially in the low frequency band between 0.1–10 Hz, GW detectors rely on sophisticated isolation systems [1]. These rely on a mix of passive methods, like pendulums and springs, which filter out ground motion, and active control, where sensors and actuators detect vibrations and cancel them in real time. In addition, signal processing techniques, such as Wiener filtering [2] are used to further suppress it. However, these methods rely on linear assumptions and dense sensor arrays, which limit their flexibility. Here, deep learning techniques come into play, as they can capture non-linear relationships and adapt more easily to changing conditions. By directly predicting seismic waveforms, we take a first step toward integrating learning-based solutions into future control and subtraction pipelines.

In this work, we introduce a transformer-based approach for predicting seismic waveforms, envisioned as a step toward a foundation model for seismology. Although we highlight future third-generation GW observatories such as the Einstein Telescope (ET) [3] as a motivating use case, the approach is general and transferable to other physical systems.

2 Model Architecture

SeismoGPT is a transformer-based architecture for seismic waveform forecasting. The model comes in two variants—one for single-station input and another for seismic arrays—with a schematic overview provided in [4].

The *SeismoGPT* uses the transformer encoder [5], where self-attention captures long-range dependencies. In our autoregressive setup we use *causal attention*, so each token attends only to the present and past tokens.

2.0.1 Single-Station Model

The single-station model forecasts three-component waveforms from one seismic station. Input waveforms $X \in \mathbb{R}^{T\times 3}$ (Z, N, E components) are split into N non-overlapping tokens of length L, yielding $X_{\text{tok}} \in \mathbb{R}^{N\times L\times 3}$. Each token is flattened and embedded with a 1D convolutional block, producing $Z \in \mathbb{R}^{N\times d}$.

In addition, a sinusoidal positional encoding [5] has been added and processed the sequence with stacked transformer encoder layers using causal attention. The output $H \in \mathbb{R}^{N\times d}$ is projected back into waveform space:

$$\hat{Y}_{\text{tok}} \in \mathbb{R}^{N\times L\times 3},$$

representing the predicted waveform tokens.

2.0.2 Array-Based Model

The array-based model leverages spatial correlations across multiple stations. Input data has shape $X \in \mathbb{R}^{B\times S\times N\times L\times 3}$, where B is batch size, S the number of stations, N the number of

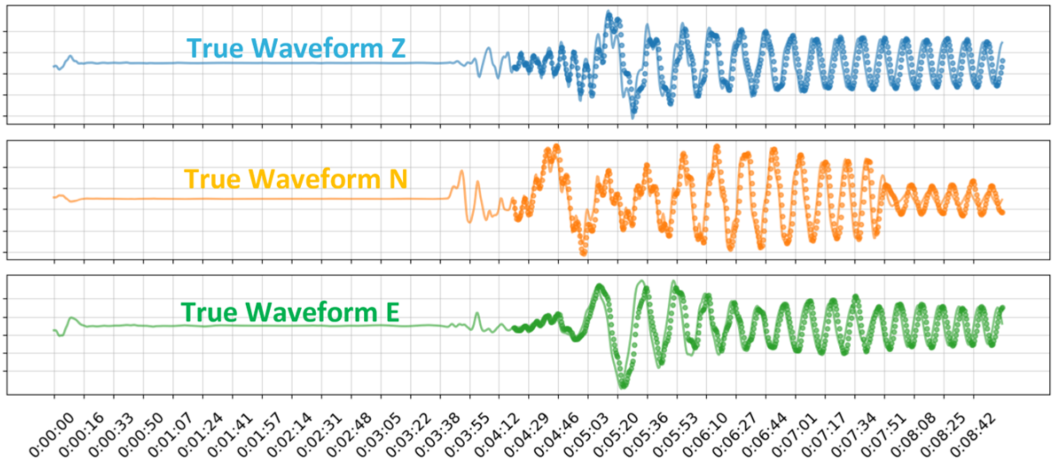

Figure 1: Prediction results from the single-station *SeismoGPT* model. The predicted waveform (dotted) is overlaid on the ground truth (solid) for each component (Z, N, E).

tokens per station, and L the token length. Tokens are flattened and embedded via a 2D convolutional network:

$$X_{\text{flat}} \in \mathbb{R}^{B\times S\times N\times(L\cdot 3)} \rightarrow Z \in \mathbb{R}^{B\times S\times N\times d}.$$

The embeddings are reshaped into two branches: a temporal branch, where each station is modeled independently $Z_{\text{temp}} \in \mathbb{R}^{(B\cdot S)\times N\times d}$, and a spatial branch, where time steps across stations are grouped $Z_{\text{spat}} \in \mathbb{R}^{(B\cdot N)\times S\times d}$. Positional encodings are added only to the temporal embeddings. The temporal branch applies causal self-attention, while the spatial branch uses full self-attention. Their outputs are then reshaped, merged, and projected back into waveform space, producing $\hat{Y}_{\text{tok}} \in \mathbb{R}^{B\times S\times(N\cdot L)\times 3}$, which represents the predicted three-component waveforms for all stations.

3 Results

We trained *SeismoGPT* on noise-free synthetic seismograms generated with Instaseis [6]. In this simulation, the seismic stations were placed at teleseismic distances to ensure stable waveform characteristics. For the array-based setup, a 16-station network was simulated near the proposed ET site in the Euregio Meuse-Rhine region. All waveforms were bandpass filtered, tapered, and aligned to theoretical P and S arrivals for consistency. The models were trained autoregressively to predict the next token from a context of 64 tokens, where each token is 16 samples, using 1D convolutional embeddings and a 6-layer transformer encoder with 8 attention heads. Optimization was performed with Adam [8] and learning rate of 5×10^{-4}.

For evaluation, we tested both the single-station and array-based models on a representative setup where the input consists of 32 tokens (16 samples each), corresponding to about 269 seconds (1.9 Hz sampling rate) and the model forecasts the next 32 tokens.

As an example, Figure 1 shows predictions from the single-station model. The forecast closely follows the ground-truth waveform in the early steps, when the prediction window is still near the observed context. As the horizon extends, uncertainty grows and small errors accumulate due to the autoregressive decoding, yet the model still captures key features such as phase arrivals and oscillatory patterns with reasonable fidelity. Overall, both model variants achieved strong accuracy in the initial forecast steps. The array-based model benefited from spatial redundancy across stations, producing more stable predictions than the single-station variant, particularly at longer horizons.

4 Conclusion

In this work, we presented *SeismoGPT*, a transformer-based model for seismic waveform forecasting with both single-station and array-based inputs. Beyond demonstrating its potential for seismic noise mitigation and Newtonian-noise subtraction at GW observatories, our results point toward broader applications in real-time systems such as earthquake early warning. This work is a step toward developing a general-purpose foundation model for seismology.

Acknowledgments

This work is funded by the ErUM-WAVE project 05D2022, “ErUM-Wave: Antizipation 3-dimensionaler Wellenfelder”, which is supported by the German Federal Ministry of Education and Research (BMBF).

References

[1] L. Trozzo, F. Badaracco, *Seismic and Newtonian Noise in the GW Detectors*, Galaxies **10**, 20 (2022), doi:10.3390/galaxies10010020.

[2] J. Harms, *Terrestrial gravity fluctuations*, Living Rev. Rel. **22**, 6 (2019), doi:10.1007/s41114-019-0022-2.

[3] M. Punturo, M. Abernathy, F. Acernese, B. Allen, N. Andersson, K. Arun, F. Barone, B. Barr, M. Barsuglia, M. Beker *et al.*, *The Einstein Telescope: A third-generation gravitational wave observatory*, Class. Quant. Grav. **27**, 194002 (2010), doi:10.1088/0264-9381/27/19/194002.

[4] W. Esmail, A. Kappes, S. Russell and C. Thomas, *Forecasting Seismic Waveforms: A Deep Learning Approach for Einstein Telescope*, doi:10.22323/1.501.0924 [arXiv:2509.21446 [cs.LG]].

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, *Attention Is All You Need*, Adv. Neural Inf. Process. Syst. **30**, 6000–6010 (2017), arXiv:1706.03762 [cs.CL].

[6] M. van Driel, L. Krischer, S. C. Stähler, K. Hosseini, T. Nissen-Meyer, *Instaseis: instant global seismograms based on a broadband waveform database*, Solid Earth **6**, 701–717 (2015), doi:10.5194/se-6-701-2015.

[7] B. L. N. Kennett, E. R. Engdahl, R. Buland, *Constraints on seismic velocities in the Earth from traveltimes*, Geophys. J. Int. **122**, 108–124 (1995), doi:10.1111/j.1365-246X.1995.tb03540.x.

[8] D. P. Kingma, J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv:1412.6980 [cs.LG].