Machine learned reconstruction of tsunami waves from sparse observations

Edward McDugald
University of Arizona
emcdugald@arizona.edu

Arvind Mohan
Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory

Darren Engwirda
Theoretical Division, Los Alamos National Laboratory

Javier Santos
Earth and Environmental Sciences, Los Alamos National Laboratory

Agnese Marcato
Earth and Environmental Sciences, Los Alamos National Laboratory

Abstract

Accurate forecasting of tsunami waves is critical to a functioning early warning system. While physical models of tsunami waves are well-understood and solvable on a computer, simulation of such models at high resolution is computationally expensive and time consuming. The primary source of practical wave-height data is supplied by the DART (Deep-ocean Assessment and Reporting of Tsunami) Network, a series of buoys that provide wave-height measurements throughout the ocean. The challenge we address in this work is in accurately estimating densely sampled wave-height fields given sparse measurements obtained from the DART network. We use an attention-based neural network designed for sparse sensing problems in the physical sciences, and test its reconstruction accuracy on realistic tsunami simulations with mixed initial conditions. Our experiments demonstrate a promising new tool for obtaining densely sampled observation networks for tsunami forecasting.

1   Introduction

Tsunamis are a devastating natural hazard, causing more than 250,000 deaths globally between 1998 and 2017, and US$280 billion in damages [4]. Each year, about 60,000 people and US$4 billion in assets are exposed to tsunami hazard [3]. The majority of existing tsunami forecast systems employ numerical solvers, in which an initial ocean surface displacement is derived based on estimates of the underlying earthquake dynamics, and a shallow water type PDE model is integrated to predict the propagation of wave fronts over the ocean basin. Such approaches provide so-called 'full-field' information to forecasters. Various PDE-based models are currently employed both operationally and throughout the research communities, including NOAA's MOST system [15], the GeoCLAW solver [2, 7], and the Gerris/Basilisk frameworks [11, 12]. Integrating these models is an expensive computational task, requiring massively parallel implementations and resources.

Alternatively, data assimilation methods seek to forecast tsunamis without the need for initial source parameters or sea surface height displacements, instead using real sensor measurements as input. In [9], a data assimilation method is proposed that combines bottom-pressure gauges with 2D linear long wave equations to derive wave height and velocity estimates for forecasting. While this method can be made computationally efficient with the use of pre-computed Green's functions [16], it relies on a dense observation network to be effective. To address this need for dense observations, [18] proposed an algorithm based on the Huygens-Fresnel principal to generate virtual observation points from a sparse set of real observation points.

We propose an attention-based machine learning architecture known as the Senseiver to generate an arbitrarily dense sample of observations from sparse measurements derived from DART buoys. While some notable works exist in the use of basic machine learning (ML) methods [8] and Convolutional Neural Networks (CNN) [13] to *predict* tsunami wave heights at fixed sensor locations, we are not aware of any previous works where machine learning is used to solve the sparse sensing problem encountered in the tsunami data assimilation method. To the best of our knowledge, the most similar work to the present study is [1], where sea surface height (SSH) measurements supported on lines are coupled with densely sampled sea surface temperature (SST) data to train an attention-based encoder-decoder model that generates densely sampled sea surface height reconstructions in order to estimate geostrophic currents. While their approach uses a similar ML architecture, it differs significantly from our work in that they use full field sea temperature information in addition to height measurements supported on lines as input, and estimate sea surface currents rather than tsunami waves.

The code for our experiments is available on github, and the data is available on zenodo.

2 Methods

The tsunami forecast problem can be posed as a sparse sensing challenge: given a discrete set of observations of the ocean state $s = \{s_1, s_2, \ldots, s_N\}$ at locations $\mathbf{x}^s = \{\mathbf{x}_1^s, \mathbf{x}_2^s, \ldots, \mathbf{x}_N^s\}$, reconstruct a set of state outputs $\hat{s} = \{\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_M\}$ at a set of query points $\mathbf{x}^q = \{\mathbf{x}_1^q, \mathbf{x}_2^q, \ldots, \mathbf{x}_M^q\}$. In this work, the input state $s(\mathbf{x}^s)$ is taken to be the ocean surface height measured at DART buoy locations, and the query state $\hat{s}(\mathbf{x}^q)$ is the ocean surface height throughout the remainder of the global ocean domain. Noting that DART buoys number only a few dozen, and that even relatively coarse coverage of the global ocean requires at least several hundred thousand discrete points, the reconstruction problem is seen to be extremely sparse, with $M \gg N$. The Senseiver [14], an attention-based neural network, has been developed to solve such problems — employing a multi-level encoder-decoder architecture to reconstruct dense output from sparse observations. In the context of sparse sensing, the key benefits of the attention mechanicsm are (1) the use of positional information as a feature, (2) the ability to learn long-range spatial associations immediately, and (3) flexibility to process arbitrarily structured data. The drawback of attention is that it scales quadratically with its input. The Senseiver resolves this issue by mapping its input to latent arrays, and performing the bulk of computation in a latent space. In brief, the Senseiver operates as follows: A positional encoder P_E makes a trigonometric encoding of the spatial coordinates, which are then passed, along with the sensor values, to an attention-based encoder E that yields a compressed representation of the system. Finally, an attention-based decoder D outputs field values at arbitrary query locations. The reader is referred to [14] for in-depth discussions of the Senseiver architecture. The weights of the model are optimized using Adam [6] with learning rate 1×10^{-4}, by minimizing the mean-squared error

$$\mathcal{L} = \sum \left(s(\mathbf{x}^q, t^q) - \hat{s}(\mathbf{x}^q) \right)^2, \tag{1}$$

$$\text{where} \quad \hat{s}(\mathbf{x}^q) = D\left(E\left(s(\mathbf{x}^q), P_E(\mathbf{x}^s) \right), P_E(\mathbf{x}^q) \right). \tag{2}$$

Noting that the loss (1) is a summation over the query points $\mathbf{x}^q$, here taken to be a dense set covering the global ocean domain, training the model requires that a dense set of state estimates $s(\mathbf{x}^q)$ be available, provided in this work by a PDE-based surrogate model. While previous experiments of the Senseiver used data residing on a structured grid, the flexibility of the architecture allows one to seamlessly build a model on unstructured data. Further, the choice of coordinates for the positional encodings are not restricted to the underlying spatial coordinates. For this problem, we

added encodings for the ocean bathymetry in addition to latitude and longitude. We also excluded land pixels, training only on data corresponding to ocean pixels. Each model was trained for 10 hours on 8 GPUs, however we found that the loss had converged after $\sim$ 6 hours of training.

3 Data generation

To generate data, we use numerical solutions of the shallow water equations (SWE), in which the dynamics of a layer of fluid is represented by a coupled system of PDEs that govern the evolution of the depth-averaged horizontal velocity $u = u(\mathbf{x}, t)$ and the fluid thickness $h = h(\mathbf{x}, t)$

$$\frac{\partial u}{\partial t} + (u \cdot \nabla)u + fu^\perp = -(1 - \beta)g\nabla(h + z_b) + c_d \frac{|u|}{h}u + \nu_k^u \nabla^k u \,, \tag{3}$$

$$\frac{\partial h}{\partial t} + \nabla \cdot (uh) = 0 \,. \tag{4}$$

Here, $z_b = z(\mathbf{x})$ is the height of the bathymetry at the base of the fluid domain, $g = 9.80665 \,\mathrm{m\,s^{-2}}$ is the acceleration due to gravity, and $f = 2\Omega \sin \phi$ is the Coriolis parameter, with $\Omega = 7.292 \times 10^{-5} \,\mathrm{rad\,s^{-1}}$ being the rotation rate of the Earth and ϕ the angle of latitude. $u^\perp = \mathbf{k} \times u$, where $\mathbf{k}$ is the local vertical direction, is a tangential velocity field used to express the acceleration due to the Coriolis tendency. $\beta = 0.015$ is a reduced gravity correction that approximates the self-attraction and loading effects associated with tsunami waves [5].

For the mesh used in the SWE solver, each frame is output as an unstructured list of 163,842 tuples, consisting of a longitude/latitude coordinate and a wave height. We subsample each frame by a factor of 2, yielding frames consisting of 81,921 pixels. For each training epicenter, the simulation is run for 288 time steps, with a time step of 50s, effectively giving 4 hour simulations. The epicenter coordinates used in training were randomly selected from a list of historic seismic events restricted to the coast of Japan. We split each four hour simulation into two simulations of two hours, giving two models (one for each time regime).

To test the model, we perturbed the training epicenters to generate eight new simulations, each consisting of frames not observed in training. We refer to this data as the "unseen epicenters". These epicenters are depicted visually in relation to the training epicenters in the appendix. The distances of the unseen epicenters from their closest training epicenter ranges from 2.78 miles to 119.97 miles, such that the wave dynamics are qualitatively different due to the change in ocean bathymetry. The error metric we use to evaluate our models is scaled by the absolute max wave height for each frame, and is computed only over pixels achieving a value of at least 1×10^{-4} in the true field. This pixel restriction mitigates the issue of artificially low cumulative error as the pixels in the rest of the ocean unaffected by the tsunami wave will show near-zero error. For each frame we compute the mean absolute difference between the true and reconstructed wave height fields, and divide by the maximum absolute true wave height for that frame. As a formula, this is:

$$\mathrm{Error}(t) = \frac{\overline{|h(t) - \hat{h}(t)|}}{\max(\overline{|h(t)|})} \,. \tag{5}$$

4 Results

Evaluating the 0-2 hour model across all epicenters and times of the *training* set yields an error of 5.6×10^{-2}, whereas the same evaluation for the 2-4 hour model yields an error of 7.22×10^{-3}. This order of magnitude difference in error between the two models derives from the extremely ill-posed reconstruction problem occurring for the early time frames in the 0-2 hour model. The wave height fields at $t = 0$ consist of a Gaussian supported over a small neighborhood of the epicenter, resulting in null to negligible information to the sensors. We note that this issue is common to the original tsunami data assimilation method [9], where surface height fields are not updated until at least one sensor has received a non-zero signal. The same issue arises in the machine learning based forecasting framework discussed in [8]; sensor measurements can not be used for prediction until the wave front has interacted with the sensor.

Across the eight epicenters in the training data, we measured a mean lag period of 30.4 minutes for the tsunami waves to register sufficient sensor information to generate a reconstruction error no greater than 1×10^{-1}. In contrast, the 2-4 hour models never exceed a reconstruction error of 2×10^{-2}. For the *unseen* epicenters, the overall error for the 0-2 hour model is 8.64×10^{-2}, with a mean lag period to achieve a reconstruction error no greater than 1×10^{-1} of 59.5 minutes. For the 2-4 hour model, the overall error is 2.78×10^{-2}.

In the first two rows of Figure 4, we contrast the reconstructions at times 12.5 minutes and 70.8 minutes for unseen epicenter $(142.5°E, 36.2°N)$. The closest epicenter in the training set is separated from this epicenter by a distance of 64.8 miles. We note that at 12.5 minutes, only two sensors register the tsunami wave, which grows to five sensors at 70.8 minutes. Inspecting the reconstruction at 12.5 minutes, we see qualitatively a close estimate of the epicenter location, as well as a close estimate of the support of the first spatial period of the phase. The max amplitude of the reconstruction is $\approx 80\%$ of the true max amplitude, and the reconstruction struggles to capture the two-period nature of the spatial phase in the true field. For the reconstruction at 70.8 minutes, we observe much improved reconstruction accuracy, with remarkable agreement in phase field extent. In the bottom two rows, we present sample reconstructions from the 2-4 hour model for the unseen epicenter at $(138.9°E, 28.1°N)$ at times 133.3 minutes and 225.0 minutes. In this case, we see remarkable resolution of fine phase field and amplitude features.

In the fourth column, we show a time series of the error for the reconstructions (top), and the waveform estimate at a DART sensor location (bottom). The sensor corresponding to the waveform is marked by a black circle in the corresponding reconstruction plots. Unlike the 0-2 hour model, every frame in the 2-4 hour model has suitably informative sensor measurements to yield accurate reconstruction, so there is no noticeable decay in the error profile. Notably, both models yield accurate estimates for the wave forms at fixed sensor locations. As no pixels are favored during the training process, the accurate wave form recoveries at DART locations are attributed to the attention mechanism used in the Senseiver.

We also measured the physical consistency of the reconstructions as defined by equation 4, and plotted the results for the epicenter at $(136.7°E, 33.1°N)$ in the $2-4$ hour time regime in Figure 2. The error is defined by balancing the wave height time derivatives with the negative of the divergence of the product of wave height and the horizontal velocity. We found that the Senseiver reconstructions yielded lower physical consistency errors than the inherent numerical error of the SWE solver. For the 0-2 hour time frame, the mean physical consistency error is 6.8×10^{-2}, and for the 2-4 hour time frame, the mean error is 4.0×10^{-2}.

5 Discussion and conclusion

Our results demonstrate remarkable full field reconstruction accuracy for unseen epicenters deriving from incredibly sparse measurements, provided sufficiently rich information content has propagated to the sensors. Future work can include the addition of temporal encodings, modifying the training data to increase spatial epicenter coverage, and exploring alternative observation networks such as the S-net [17]. Experiments can also be conducted to probe optimal placement of DART buoys, by making the sensor locations differentiable parameters of the network [10]. We also suggest a study to test the use of the Senseiver to generate dense observation networks in the tsunami data assimilation method.

In conclusion, this study tested the utility of the Senseiver architecture in a more realistic, non-academic problem. We found the Senseiver architecture capable of yielding accurate high resolution reconstructions of tsunami waves from incredibly sparse measurements. Notably, this study showcased the ability of the Senseiver to extrapolate to initial conditions not seen in training, and took advantage of its ability to handle unstructured data.
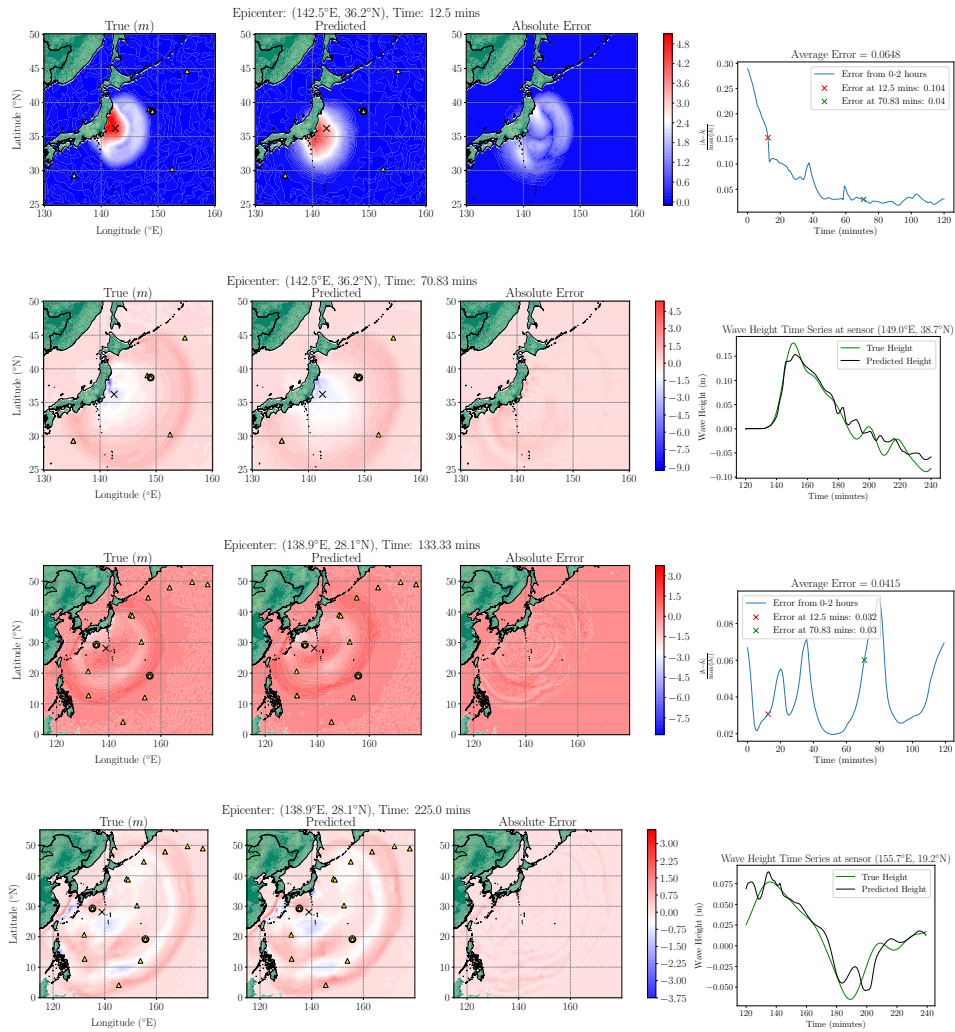
6 Acknowledgements

Figure 1: Evaluation of model performance on unseen epicenter $(142.5°E, 36.2°N)$ (top two rows) in the 0-2 hour time regime and unseen epicenter $(138.9°E, 28.1°N)$ (bottom two rows) in the 2-4 hour regime.
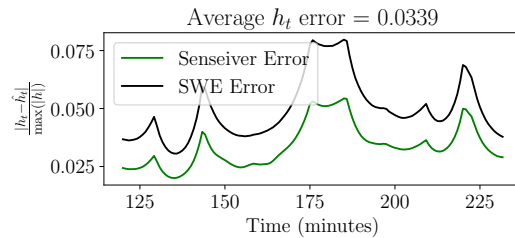

Figure 2: Temporal continuity equation error from 2-4 hours (bottom) for epicenter $(136.7°E, 33.1°N)$ for unregularized model.

References

[1] Théo Archambault, Arthur Filoche, Anastase Charantonis, Dominique Béréziat, and Sylvie Thiria. Learning sea surface height interpolation from multi-variate simulated satellite observations. *Journal of Advances in Modeling Earth Systems*, 16(6):e2023MS004047, 2024. e2023MS004047 2023MS004047.

[2] Marsha J Berger, David L George, Randall J LeVeque, and Kyle T Mandli. The GeoClaw software for depth-averaged flows with adaptive refinement. *Advances in Water Resources*, 34(9):1195–1206, 2011.

[3] E. Bernard and Vasily Titov. Evolution of tsunami warning systems and products. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 373, 10 2015.

[4] Fumihiko Imamura, Sébastien Boret, Anawat Suppasri, and Abdul Muhari. Recent occurrences of serious tsunami damage and the future challenges of tsunami disaster risk reduction. *Progress in Disaster Science*, 1:100009, 05 2019.

[5] Daisuke Inazu and Tatsuhiko Saito. Simulation of distant tsunami propagation with a radial loading deformation effect. *Earth, Planets and Space*, 65:835–842, 2013.

[6] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[7] Randall J LeVeque, David L George, and Marsha J Berger. Tsunami modelling with adaptively refined finite volume methods. *Acta Numerica*, 20:211–289, 2011.

[8] Christopher Liu, Donsub Rim, Robert Baraldi, and Randall Leveque. Comparison of machine learning approaches for tsunami forecasting from sparse observations. *Pure and Applied Geophysics*, 178, 12 2021.

[9] Takuto Maeda, Kazushige Obara, Masanao Shinohara, Toshihiko Kanazawa, and Kenji Uehira. Successive estimation of a tsunami wavefield without earthquake source data: A data assimilation approach toward real-time tsunami forecasting. *Geophysical Research Letters*, 42, 10 2015.

[10] Agnese Marcato, Daniel O'Malley, Hari Viswanathan, Eric Guiltinan, and Javier E. Santos. Reconstruction of fields from sparse sensing: Differentiable sensor placement enhances generalization, 2023.

[11] Stéphane Popinet. Adaptive modelling of long-distance wave propagation and fine-scale flooding during the Tohoku tsunami. *Natural Hazards and Earth System Sciences*, 12(4):1213–1227, 2012.

[12] Stéphane Popinet. A vertically-Lagrangian, non-hydrostatic, multilayer model for multiscale free-surface flows. *Journal of Computational Physics*, 418:109609, 2020.

[13] Donsub Rim, Robert Baraldi, Christopher M Liu, Randall J LeVeque, and Kenjiro Terada. Tsunami early warning from global navigation satellite system data using convolutional neural networks. *Geophysical Research Letters*, 49(20):e2022GL099511, 2022.

[14] Javier E Santos, Zachary R Fox, Arvind Mohan, Daniel O'Malley, Hari Viswanathan, and Nicholas Lubbers. Development of the Senseiver for efficient field reconstruction from sparse observations. *Nature Machine Intelligence*, pages 1–9, 2023.

[15] Vasily V Titov and Frank I Gonzalez. Implementation and testing of the method of splitting tsunami (MOST) model. 1997.

[16] Yuchen Wang. *Green's Function-Based Tsunami Data Assimilation (GFTDA)*, pages 19–43. 10 2022.

[17] Yuchen Wang, Kentaro Imai, Takuya Miyashita, Keisuke Ariyoshi, Narumi Takahashi, and Kenji Satake. Coastal tsunami prediction in tohoku region, japan, based on s-net observations using artificial neural network. *Earth Planets and Space*, 75:154, 10 2023.

[18] Yuchen Wang, Takuto Maeda, Kenji Satake, Mohammad Heidarzadeh, Heng-Yi Su, A. Sheehan, and Aditya Gusman. Tsunami data assimilation without a dense observation network. *Geophysical Research Letters*, 46, 02 2019.

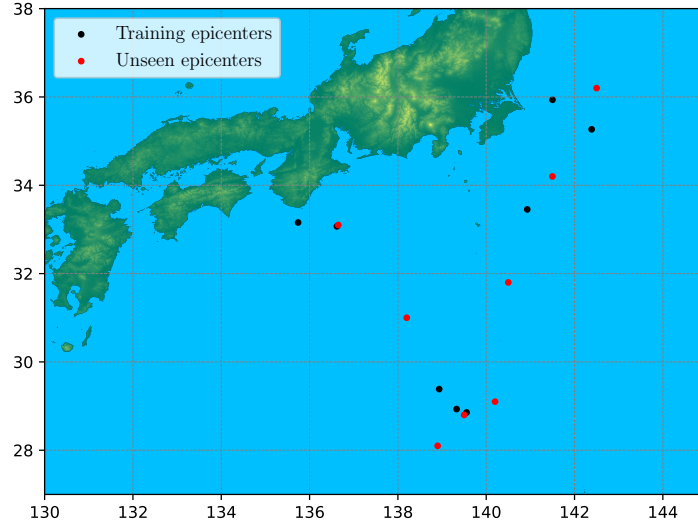## A Visual Reference for Training and Unseen Epicenters

Figure 3: Training (black) and Unseen (red) epicenters used for synthetic tsunami generation.

B Table of Training Epicenters

Simulation Coordinates								
Coordinate	Epi 1	Epi 2	Epi 3	Epi 4	Epi 5	Epi 6	Epi 7	Epi 8
Longitude	136.62	139.56	139.33	138.94	140.93	135.74	141.50	142.39
Latitude	33.07	28.86	28.93	29.38	33.45	33.16	35.94	35.27

C Table Unseen Epicenters

Simulation Coordinates								
Coordinate	Epi 1	Epi 2	Epi 3	Epi 4	Epi 5	Epi 6	Epi 7	Epi 8
Longitude	136.65	138.20	138.90	139.50	140.20	140.50	141.50	142.50
Latitude	33.10	31.00	28.10	28.80	29.10	31.80	34.20	36.20

D Discretized Physical Consistency Equation

$$\frac{h_n - h_{n-1}}{\Delta t} = -\frac{(\nabla \cdot (hu)_n + \nabla \cdot (hu)_{n-1})}{2}. \tag{6}$$
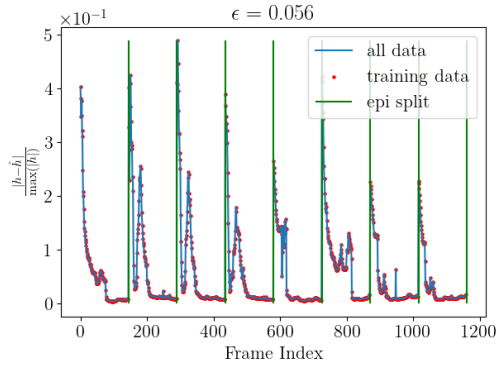
E Error Plots

Figure 4: Average per-pixel error of the training data for the $0 - 2$ hour model. The large spikes indicate the beginning of a new epicenter simulation in the training data. The total number of frames in the data is 1160, consisting of 8 epicenter simulations of 145 times steps each. The metric at the top of the figure is the average over all frames.
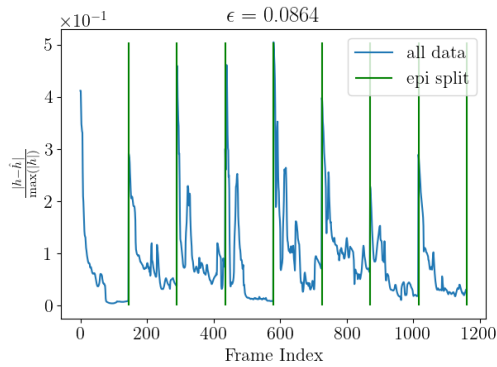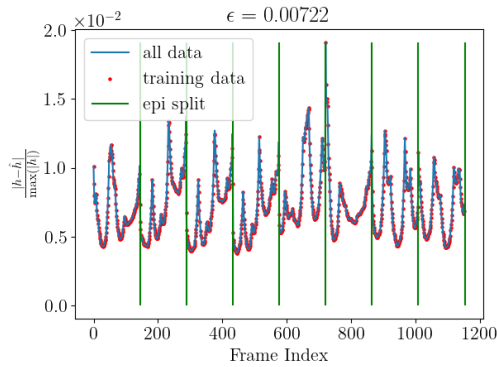

Figure 5: Average per-pixel error of the unseen data for the $0 - 2$ hour model. The large spikes indicate the beginning of a new epicenter simulation in the training data. The total number of frames in the data is 1160, consisting of 8 epicenter simulations of 145 times steps each. The metric at the top of the figure is the average over all frames.

Figure 6: Average per-pixel error of the training data for the $2 - 4$ hour model. The total number of frames in the data is 1152, consisting of 8 epicenter simulations of 144 times steps each. The metric at the top of the figure is the average over all frames.
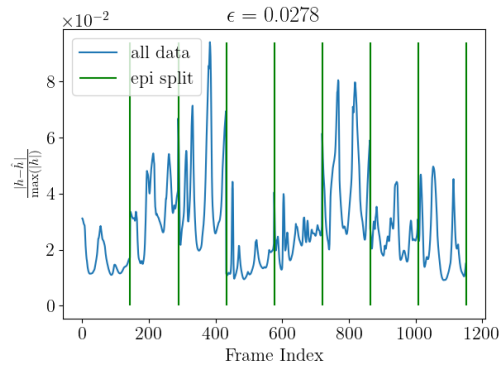
Figure 7: Average per-pixel error of the unseen data for the $2 - 4$ hour model. The total number of frames in the data is 1152, consisting of 8 epicenter simulations of 144 times steps each. The metric at the top of the figure is the average over all frames.
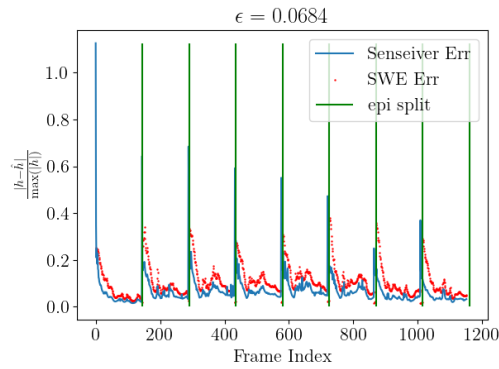

Figure 8: Average per-pixel physical consistency error of the unseen data for the $0 - 2$ hour model. The large spikes indicate the beginning of a new epicenter simulation in the training data. The total number of frames in the data is 1160, consisting of 8 epicenter simulations of 145 times steps each. The metric at the top of the figure is the average over all frames.
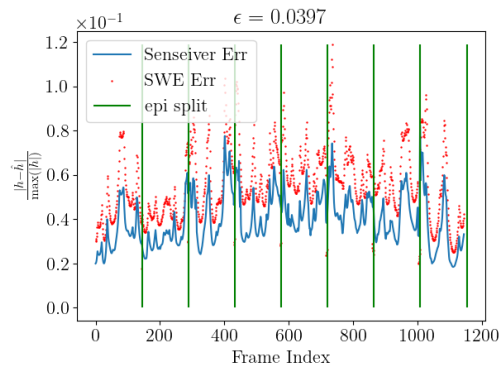

Figure 9: Average per-pixel physical consistency error of the unseen data for the $2 - 4$ hour model. The total number of frames in the data is 1152, consisting of 8 epicenter simulations of 144 times steps each. The metric at the top of the figure is the average over all frames.