Forecasting the 2016-2017 Central Apennines Earthquake Sequence with a Neural Point Process

Sam Stockman[1], **Max Werner**[2], **Dan Lawson**[1]

[1]School of Mathematics, University of Bristol, [2]School of Earth Sciences, University of Bristol

SCEC AN NSF+USGS CENTER

University of BRISTOL School of Mathematics

Introduction

The use of Neural Network based phase pickers to determine P and S-wave arrival times has meant that the size of some earthquake catalogs have increased ten-fold.
Yet, little investigation has been done into what significance these enhanced catalogs have for forecasting ability.
Here we investigate this potential by comparing the well known stationary temporal ETAS model with a highly expressive model; a Neural Point Process [Omi et al. 2019], [Shchur et al. 2019].

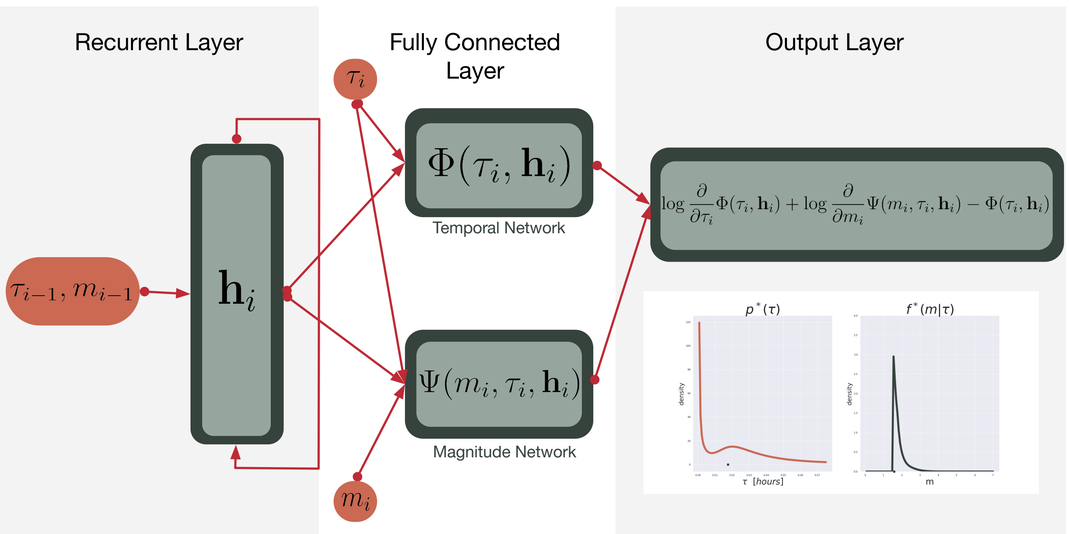
The Neural Point Process acquires its expressivity through approximating the history of a point process with an Recurrent Neural Network,

$$\mathbf{h}_i = f(W^h \mathbf{h}_{i-1} + \mathbf{w}^\tau \tau_{i-1} + \mathbf{w}^m m_{i-1} + \mathbf{b}^h),$$

Followed by two fully connected neural networks which approximate the cumulative intensity function and cumulative magnitude distribution. This allows for the log-likelihood of the point process to be formulated as the output of the neural network:

$$\begin{aligned}\log L(\{t_i, m_i\}) &= \sum_i \left[\log \lambda^*(t_i) + \log f^*(m_i|t_i) - \int_{t_{i-1}}^{t_i} \lambda^*(t)dt\right] \\ &= \sum_i \left[\log \phi(\tau_i, \mathbf{h}_i) + \log \psi(m_i, \tau_i, \mathbf{h}_i) - \int_0^{t_i - t_{i-1}} \phi(t, \mathbf{h}_i)dt\right] \\ &= \sum_i \left[\log \frac{\partial}{\partial \tau_i}\Phi(\tau_i, \mathbf{h}_i) + \log \frac{\partial}{\partial m_i}\Psi(m_i, \tau_i, \mathbf{h}_i) - \Phi(\tau_i, \mathbf{h}_i)\right].\end{aligned}$$

This architecture is a multivariate extension to the one introduced by [Omi et al. 2019] to model the joint intensity function of the next time and magnitude.



Importantly;

1. Unlike ETAS, the magnitude distribution is allowed to be dependent on the history of the point process.
2. By construction, we can pass in lower magnitudes than we are trying to predict. i.e. We can distinguish an input catalog from a target catalog.

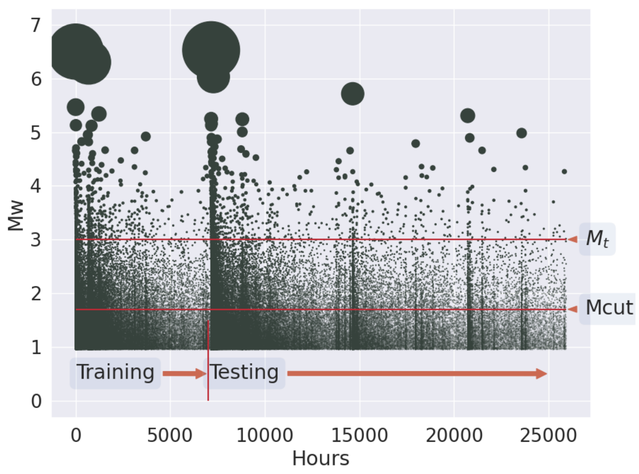
Experimental Set-up

We test both models performance on a real and synthetic earthquake catalog where the magnitude of the input catalog (Mcut) is lowered.

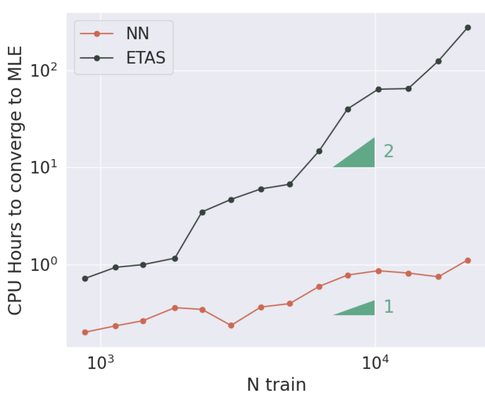
We fix the target magnitude threshold $M_t = 3\mathrm{M_w}$.

Spatial covariates are ignored and the dataset is partitioned in time into a training and testing set.

A second synthetic catalog is created by artificially removing events to emulate short-term aftershock incompleteness.



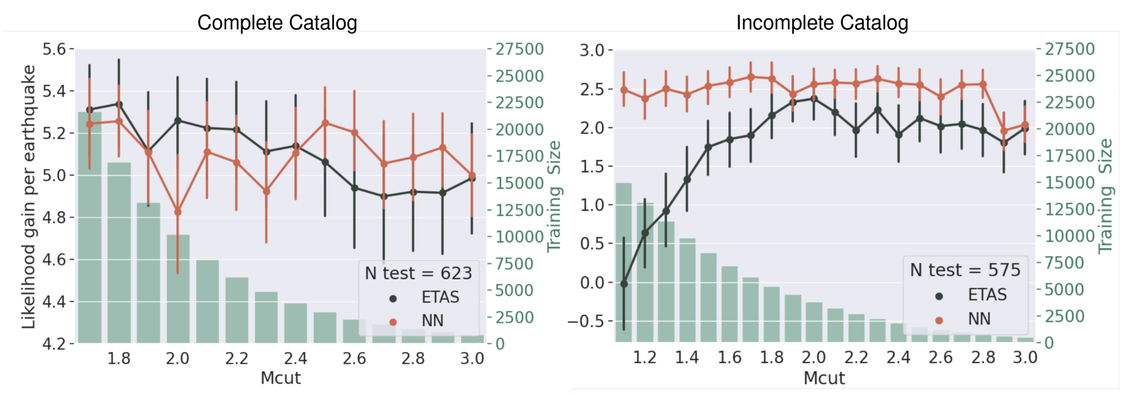
Synthetic Tests



The Neural model is much quicker to train than ETAS. The training time grows linearly with the data size unlike ETAS which grows quadratically.

The Neural model is suitably expressive to fit complete ETAS data.

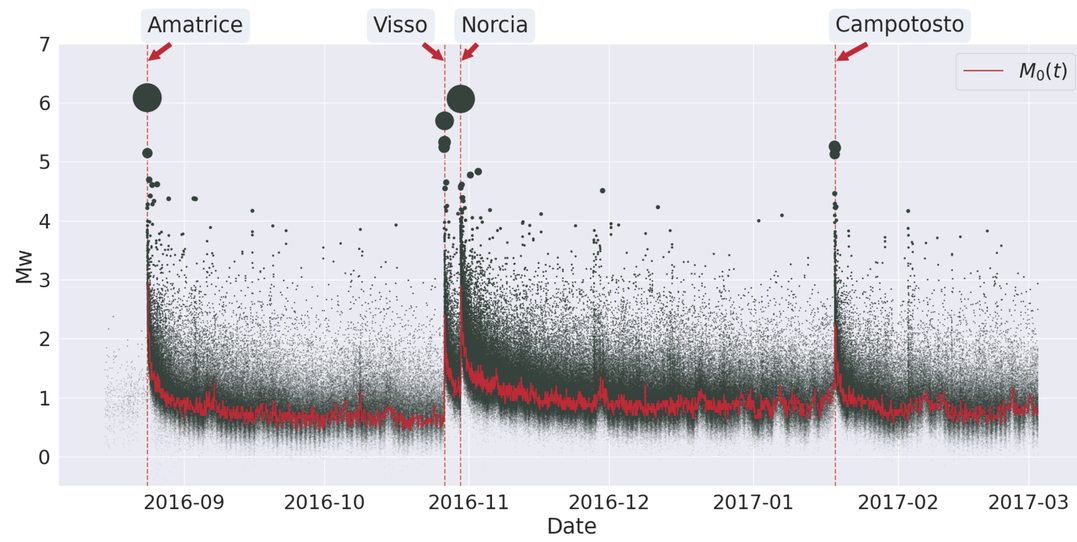
The Neural model is more robust to incomplete data than ETAS. As we lower Mcut, the performance of ETAS decreases unlike the Neural model.



High Resolution Catalog for 2016 Central Apennines

The original catalog for the Central Apennines sequence produced by the INGV contained 82,000 events, complete down to 2.5 $\mathrm{M_w}$. With the use of PhaseNet, Tan et al. 2021, produced a significantly enhanced catalog for the same sequence, with an overall estimate of the completeness 1.0 $\mathrm{M_w}$ and around 900,000 events.

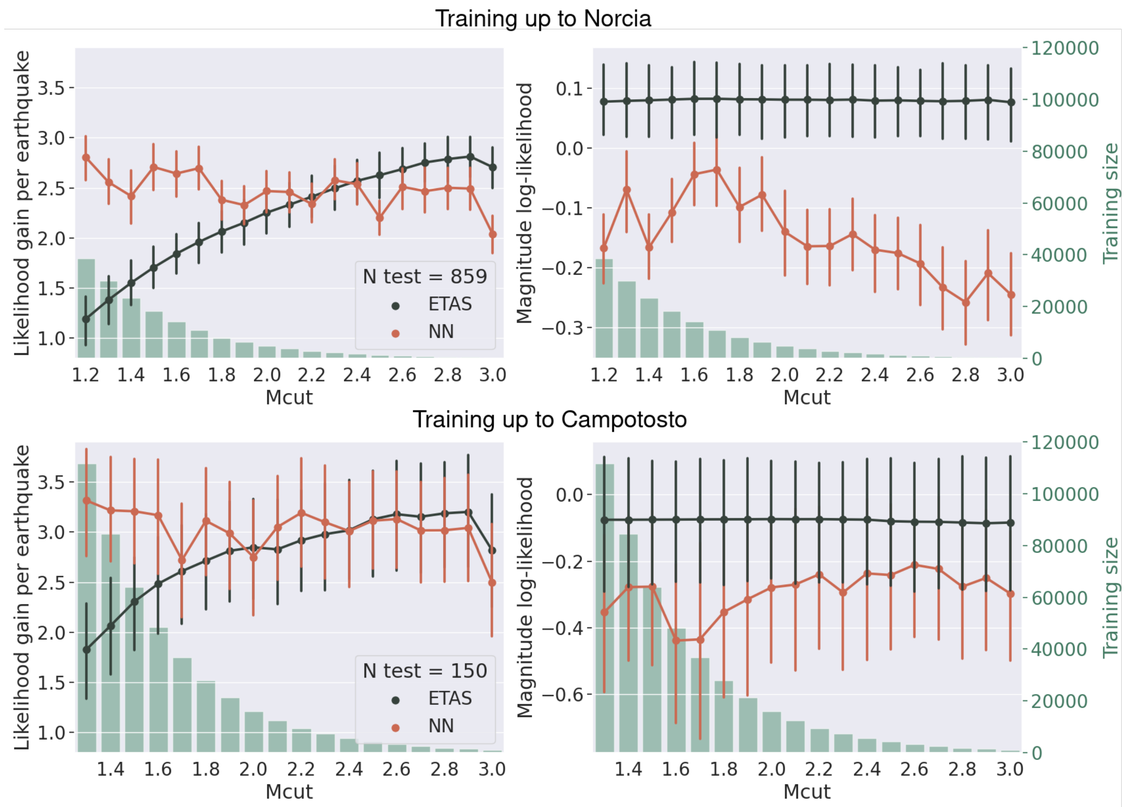
We take 2 partitions of this dataset and look at the performance as we lower Mcut.



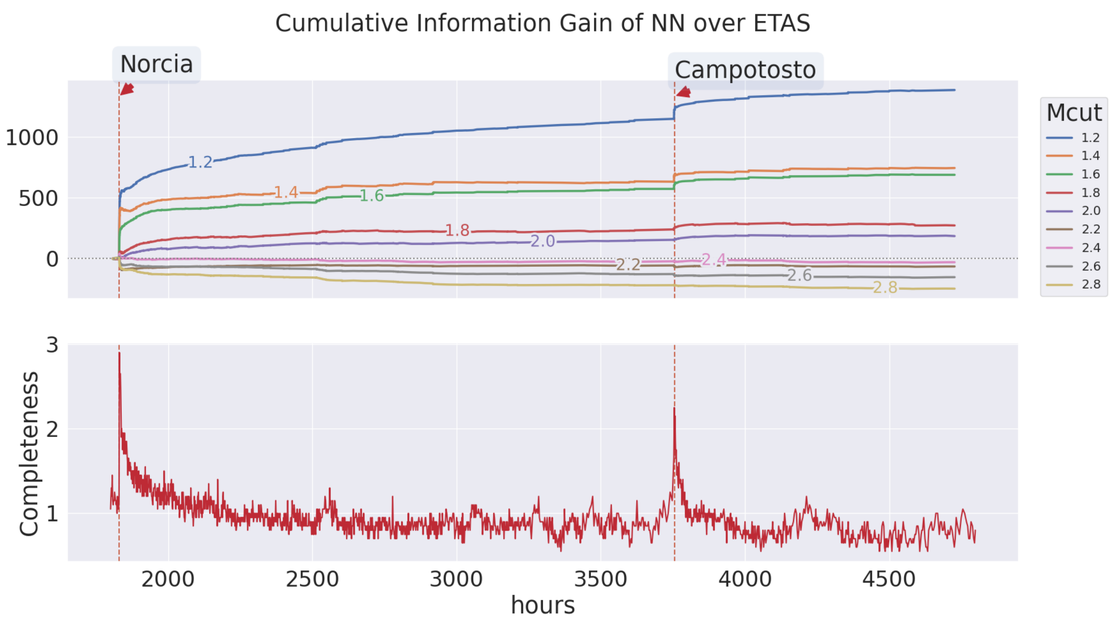
For both of the partitions, the temporal performance of ETAS decreases steadily as Mcut is lowered.

Whereas the Neural model increases steadily as Mcut is lowered. Demonstrating a significantly better performance in the new unexplored thresholds of the enhanced catalog. However, the overall forecasting improvement across thresholds is marginal.

Constructing a history dependant magnitude distribution is less-powerful than the stationary Gutenberg-Richter law.



- The periods of greatest gains are when there is greatest short-term aftershock incompleteness.



Conclusions

- Neural Point Processes are a more appropriate model for the newly exposed magnitude thresholds found in high resolution catalogs.
- They may offer a more efficient solution to modelling short-term aftershock incompleteness than ETAS based methods.

References

Omi, Takahiro et al. (2019). "Fully Neural Network based Model for General Temporal Point Processes". In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc.

Shchur, Oleksandr et al. (2019). "Intensity-Free Learning of Temporal Point Processes". In: *International Conference on Learning Representations*.

Tan, Yen Joe et al. (2021). "Machine-Learning-Based High-Resolution Earthquake Catalog Reveals How Complex Fault Structures Were Activated during the 2016–2017 Central Italy Sequence". In: *The Seismic Record* 1.1, pp. 11–19.