

Geophysical Research Letters®

RESEARCH LETTER

10.1029/2023GL103909

Key Points:

- We introduce a deep learning model for earthquake forecasting and explore its performance on synthetic and regional earthquake data sets
- It is flexible in the sense that a predefined functional form is not required
- It is scalable in two senses: it is efficient on large data sets, and its performance relative to benchmarks improves with more training data

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

K. Dascher-Cousineau,
kdascher@berkeley.edu

Citation:

Dascher-Cousineau, K., Shchur, O., Brodsky, E. E., & Günemann, S. (2023). Using deep learning for flexible and scalable earthquake forecasting. *Geophysical Research Letters*, 50, e2023GL103909. <https://doi.org/10.1029/2023GL103909>

Received 3 APR 2023

Accepted 13 JUL 2023

Author Contributions:

Conceptualization: Kelian Dascher-Cousineau, Oleksandr Shchur, Emily E. Brodsky, Stephan Günemann
Data curation: Kelian Dascher-Cousineau
Formal analysis: Kelian Dascher-Cousineau, Oleksandr Shchur
Funding acquisition: Kelian Dascher-Cousineau, Emily E. Brodsky
Investigation: Kelian Dascher-Cousineau, Oleksandr Shchur, Emily E. Brodsky, Stephan Günemann
Methodology: Kelian Dascher-Cousineau, Oleksandr Shchur
Supervision: Emily E. Brodsky, Stephan Günemann

© 2023. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Using Deep Learning for Flexible and Scalable Earthquake Forecasting



Kelian Dascher-Cousineau^{1,2} , Oleksandr Shchur³, Emily E. Brodsky¹ , and Stephan Günemann³

¹Department of Earth and Planetary Sciences, University of California, Santa Cruz, Santa Cruz, CA, USA, ²Department of Earth and Planetary Sciences, University of California, Berkeley, Berkeley, CA, USA, ³Department of Computer Science and Munich Data Science Institute, Technical University of Munich, Munich, Germany

Abstract Seismology is witnessing explosive growth in the diversity and scale of earthquake catalogs. A key motivation for this community effort is that more data should translate into better earthquake forecasts. Such improvements are yet to be seen. Here, we introduce the Recurrent Earthquake foreCAST (RECAST), a deep-learning model based on recent developments in neural temporal point processes. The model enables access to a greater volume and diversity of earthquake observations, overcoming the theoretical and computational limitations of traditional approaches. We benchmark against a temporal Epidemic Type Aftershock Sequence model. Tests on synthetic data suggest that with a modest-sized data set, RECAST accurately models earthquake-like point processes directly from cataloged data. Tests on earthquake catalogs in Southern California indicate improved fit and forecast accuracy compared to our benchmark when the training set is sufficiently long ($>10^4$ events). The basic components in RECAST add flexibility and scalability for earthquake forecasting without sacrificing performance.

Plain Language Summary We explore the potential for deep learning in earthquake forecasting. Prior work has relied heavily on statistical models that do not scale to fully utilize the currently available large earthquake data sets. Here we build on recent developments in deep learning for forecasting event sequences in general to create an implementation for earthquake data. The new approach allows us to incorporate larger data sets, potentially with more information about each earthquake. We also avoid a specific functional form, so the method naturally adapts to additional information about events, like magnitude or variations in behavior over time. As we add more data, results show continued improvements. This ability to incorporate and improve continually as training data sets increase indicates that there is more information in the earthquake catalogs than has yet been used for earthquake forecasting.

1. Introduction

The diversity and scale of earthquake catalogs has exploded in the past years due to dense seismic networks and increasingly automated data processing techniques (Mousavi & Beroza, 2023; Mousavi et al., 2020; Obara, 2003; Ross et al., 2019; Shelly, 2017; Tan et al., 2021; White et al., 2019). A motivation for this community effort is that more detailed observations should translate into better earthquake forecasts. However, the corresponding improvements in forecasting have yet to materialize (Beroza et al., 2021). One factor here may be the nature of the models used for forecasting. Current operational earthquake forecasts build on seminal work designed for sparse earthquake records based on the canonical statistical laws of seismology (Llenos & Michael, 2017; van der Elst et al., 2022). While the past decades have seen advances in the regionalization of these models (Field et al., 2017; Mai et al., 2016; Ogata, 2017), catalog bias correction (Mizrahi et al., 2021; G. J. Ross, 2021) and spatial forecasts (Ogata, 1998), these advancements have failed to fully capitalize on the wealth of available geophysical data (Mancini et al., 2022; Mousavi & Beroza, 2023). This is largely due to the limitations inherent to current models, which frequently constrain the analysis to only a small portion of the available catalogs.

We turn to recent advances in machine learning using neural temporal point processes to complement existing forecasting capabilities. Such approaches provide the promise of general-purpose flexible and scalable forecasting (Shchur et al., 2021). Here we use the term flexible in the sense that models can both incorporate additional earthquake information and learn complex dependencies in the data without assuming a specific functional form (Gareth et al., 2021). We use the term scalable in the sense that models can efficiently train on large data sets (e.g., Grover & Leskovec, 2016) and improve with additional data (e.g., Kaplan et al., 2020). These characteristics



Visualization: Kelian Dascher-Cousineau
Writing – original draft: Kelian Dascher-Cousineau, Emily E. Brodsky
Writing – review & editing: Kelian Dascher-Cousineau, Oleksandr Shchur, Emily E. Brodsky, Stephan Gunnemann

provide a foundation to explore the contribution of different geophysical data sets, relax model assumptions and incorporate more data into current forecasts. Despite these desirable characteristics, it is still unclear whether deep learning is well suited for earthquake data (Mignan & Broccardo, 2020), wherein the lack of significant progress in machine learning forecasts has been attributed to an earthquake record that is highly stochastic and is strongly influenced by extreme events. Our goal is to implement an earthquake forecasting model that meets the basic requirements of flexibility and scalability and assess whether it is well suited for the task.

This paper introduces a deep learning model: the Recurrent Earthquake foreCAST (RECAST). Section 2 describes the model and the benchmark for performance, a temporal Epidemic Type Aftershock Sequence (ETAS) model. Section 3 shows the results of four experiments comparing model performance on synthetic and observed earthquake catalogs. Finally, Section 4 discusses the advantages and limitations of the modeling approach, considering our results.

2. Model Architecture and Benchmark

RECAST builds on recent developments in machine learning known as neural temporal point processes (Du et al., 2016; Omi et al., 2019; Shchur et al., 2020, 2021; Zhou et al., 2021; Zhu et al., 2021). The model uses a general-purpose encoder-decoder neural network architecture that predicts the timing of the next event given the history of past events (Figure 1a) (Du et al., 2016; Mei & Eisner, 2017; Omi et al., 2019; Shchur et al., 2020, 2021; Upadhyay et al., 2018). Note that to facilitate benchmarking, we restrict our analysis to the timing of events and do not predict the magnitudes. RECAST is based on the Gated Recurrent Unit neural network architecture (Cho et al., 2014) that encodes the variable-length history of past earthquakes, $\mathcal{H}(t_i) = \{(t_1, M_1), \dots, (t_{i-1}, M_{i-1})\}$, where t_i and M_i respectively indicate the time and magnitude of the i th earthquake, into a fixed-dimensional hidden state vector h_i . The hidden state carries the influence of previous events forward. Next, the decoder uses h_i to parametrize the probability density function of the next earthquake origin time.

$$f_{\text{RECAST}}(t_i | \mathcal{H}(t_i)) = f(t_i | h_i).$$

We model the probability density function of the next event origin time with a Weibull Mixture Distribution. This approach results in a model where training and simulation is efficient and exact because it avoids iterative numerical approximations for the evaluation and sampling of the output distribution. A mixture distribution with a sufficiently large number of components can approximate any other probability distribution arbitrarily well provided enough data (Shchur et al., 2020). RECAST output is equivalently expressed continuously in time in terms of the conditional intensity function $\lambda(t|\mathcal{H}(t))$ commonly used in point process literature. Text S5 in Supporting Information S1 explains this equivalence.

The model's architecture follows the application in Shchur et al. (2020). We extend the original implementation by introducing an encoder for continuous event marks, in this case, earthquake magnitudes. We also use a Weibull mixture instead of a log-normal mixture to ensure that $f(t|h)$ can have a non-zero intercept. Early testing suggested we did not need to tune the model hyperparameters for the earthquake forecasting task. See Text S1 in Supporting Information S1 for the full model description.

As a benchmark, we model the recurrence of earthquakes with a temporal ETAS model (Figure 1b), a parametric statistical model that calculates the time-dependent earthquake intensity $\lambda(t|\mathcal{H}(t))$ as a combination of a steady background rate and aftershock sequences (Aalen et al., 2008; Ogata, 1988). It is assumed that all earthquakes can produce aftershocks, and thus the current earthquake rate is an explicit function of the entire event history $\mathcal{H}(t)$. ETAS yields a probability density function for the next event time t , that is similar to Equation 1,

$$f_{\text{ETAS}}(t_i | \mathcal{H}(t_i)) = \lambda(t_i | \mathcal{H}(t_i)) e^{-\int_{t_{i-1}}^{t_i} \lambda(s | \mathcal{H}(s)) ds}.$$

See Text S2 in Supporting Information S1 for further implementation details.

RECAST and ETAS are both temporal point process models. They are trained by maximizing the joint log-likelihood of the cataloged origin times using the same optimization procedure (see Text S3 in Supporting Information S1). However, RECAST introduces practical benefits. Additional event features, such as location

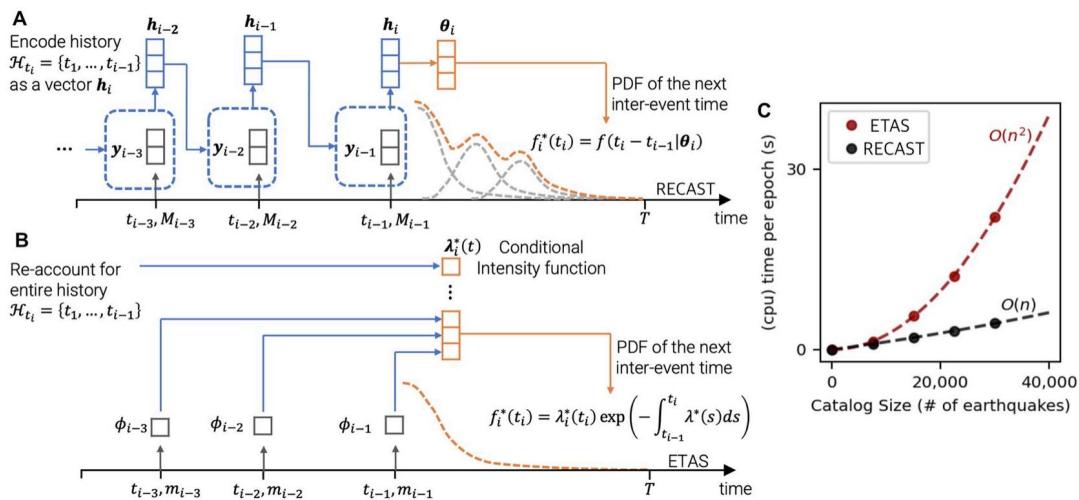


Figure 1. Structurally alike, conceptually distinct earthquake forecasting models (a) Model architecture for the RECAST model. For each likelihood evaluation, the network takes as input the previous hidden state h_{i-1} with features y_{i-1} of the previous event (t_{i-1}, M_{i-1}) and outputs an updated hidden state h_i given the gated recurrent unit update equations. An affine layer connects the hidden state to the three parameters of each Weibull Mixture component which in turn determines the model likelihood given an event t_i (b) Model architecture for the ETAS model. ETAS parametrizes the probability density function of the next origin time given the history of previous events. See Text S1 in Supporting Information S1 for extended model descriptions (c) The computational time and space complexity of evaluating the log-likelihood of an entire earthquake catalog for RECAST is linear, $O(N)$. In contrast, the complexity of the ETAS model is quadratic, $O(N^2)$.

tions, source properties, and geophysical data, can be introduced into the model. Furthermore, the model's components are modular and easily adapted using standard machine learning frameworks (Abadi et al., 2016; Paszke et al., 2019). Another key distinction is computational efficiency. Whereas RECAST processes events sequentially, summarizing the event history into a fixed-dimensional vector (Figure 1a), ETAS references all previous events to determine the intensity function and, in turn, the event likelihood (Figure 1b). As a result, the computational time and space complexity for evaluating the likelihood of a catalog is linear for RECAST and quadratic for ETAS (Figure 1c). To enable the comparisons of this study, we needed to write a new implementation of ETAS that can utilize GPU hardware with a more efficient gradient descent algorithm implemented in PyTorch (see Text S2 in Supporting Information S1). Even with these improvements, the quadratic growth is computationally cumbersome or potentially prohibitive for the real-time evaluation of earthquake forecasts given earthquake catalogs with more than $\sim 10,000$ events. This limitation is mitigated in the RECAST model that can be trained on catalogs with more than 1,000,000 events on a single consumer-grade GPU (see Figure S1 in Supporting Information S1).

We recognize that this comparison to a time-only ETAS poses a minimum performance benchmark. Substantial effort has gone into developing spatial ETAS models (Ogata, 1988, 2017; Zhuang, 2011). Heterogeneous background features complicate consistent benchmarking of RECAST across our multiple data sets. Furthermore, the computational disadvantages of even temporal ETAS required the reimplementations of the algorithm as discussed above and in Text S2 in Supporting Information S1. Spatial ETAS would be even more complex to implement in a computationally tractable framework and thus is beyond the scope of this initial study. Here we retain the more transparent time-only benchmark as a consistent and reproducible measure of performance that allows us to highlight the new model's key features on large catalogs.

3. Results

3.1. Synthetic Data

As a first step, we evaluate the performance of RECAST on a collection of synthetic catalogs generated by an ETAS model. The set of synthetic catalogs are split into training (600), validation (200), and test (200) sets. Each individual realization spans 10,000 days and contains $\sim 1,000$ events on average. For both models, the training

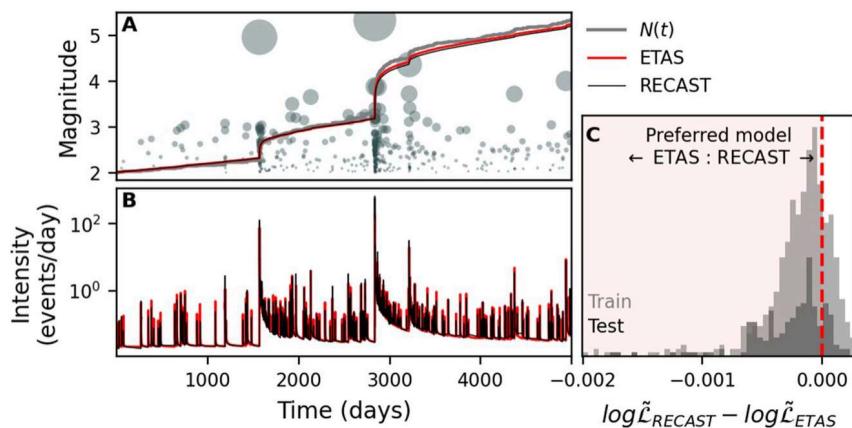


Figure 2. RECAST performance on synthetic earthquake catalogs (a) Example of a synthetic earthquake catalog generated from an ETAS model in the test set. The gray markers indicate events magnitudes and origin times. The gray line tracks the cumulative number of events, $N(t)$. The red (ETAS) and black (RECAST) lines indicate the integral of the conditional intensity function in (b), which corresponds to the expected number of events for each model. (c) For any particular catalog $\log \tilde{\mathcal{L}}_{RECAST} - \log \tilde{\mathcal{L}}_{ETAS}$ greater than zero implies that RECAST performed better than ETAS (and vice-versa). In most cases, $\log \tilde{\mathcal{L}}_{RECAST}$ approaches the $\log \tilde{\mathcal{L}}_{ETAS}$ model, an upper bound in expectation since ETAS is the generative model. Comparable distributions on the training and test sets indicate that the model is not overfitting.

set is used to optimize the parameters. During training, we monitor the performance on the validation set. After training, we recover the weights from the model with the best validation score. Finally, the comparison of model performance is reported for the reserved test set. ETAS will yield, in expectation, the highest possible joint log-likelihood on the test sequences since it is the generative model. If RECAST's performance approaches this limit, it would demonstrate that the model can be trained to capture realistic earthquake temporal clustering from the event data alone.

Figures 2a and 2b shows the conditional intensity function output from the trained RECAST model successfully capturing, for instance, the sharp fluctuations associated with Omori's law (Figures 2a and 2b). The goodness of fit, over both the training and test set, approaches the limiting performance of ETAS (Figure 2c). Increasing the volume of synthetic data for training causes the performance of RECAST to approach ETAS asymptotically (Figure 3a). Without tailoring the RECAST architecture for the task, the model generalizes the statistical laws of seismology composing ETAS from event data alone. The architecture for the input marks, which in this case are magnitudes, is general, and thus the successful recovery of standard magnitude-based behavior is a demonstration of the effectiveness of the approach to incorporating additional inputs into the model.

3.2. San Jacinto Fault Earthquake Catalog and Scaling

We next consider an actual earthquake catalog from 2008 to 2021 bounding the San Jacinto Fault (Figure S4 in Supporting Information S1) (White et al., 2019). This catalog makes a good test case because of its dense station coverage of a particularly seismically active area (See Text S6 in Supporting Information S1 for more detail on data pre-processing). RECAST outperforms ETAS in terms of goodness-of-fit in the reserved test period (Figure 3b).

We again explore the effect of data size. We track performance on the test set while incrementally extending the training period backward in time. In this experiment, the magnitude of completeness is fixed. The benchmark ETAS is the preferred model when trained on fewer than 10,000 events. The new model RECAST is the preferred model beyond 10,000 events. The ETAS benchmark performance saturates with more than ~4,000 training events, whereas the log-likelihood score of RECAST continues to increase in proportion to the log length of the training data. The crossover appears to reflect fundamentally distinct scaling between the size of the training set and performance.

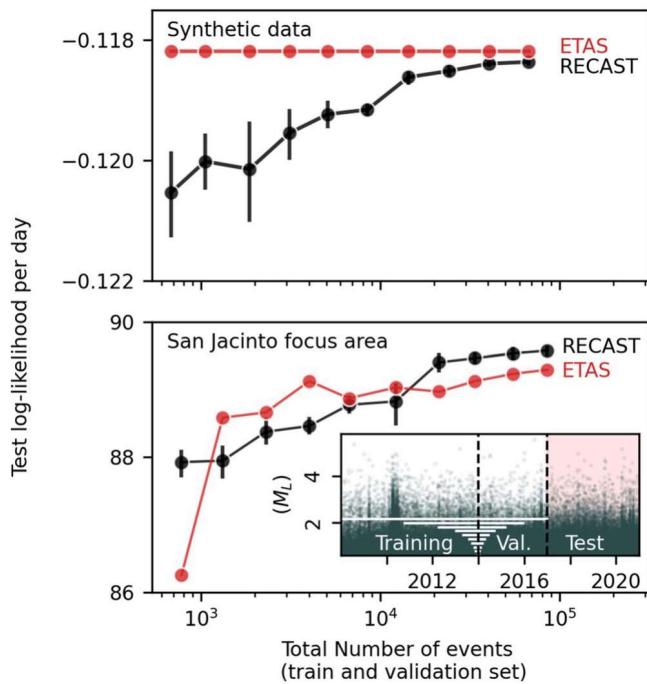


Figure 3. Performance on synthetic and modern earthquake catalogs. Model goodness of fit on an out-of-sample test period as measured by the time-averaged joint log-likelihood for RECAST and ETAS model. We compare performance on (a) a synthetic catalog and on (b) seismicity around the San Jacinto Fault zone in Southern California (White et al., 2019). Error bars indicate the 95% confidence interval of 1,000 bootstrap samples for five random initializations of RECAST. The models were trained with incrementally longer training and validation sets. In panel (b), the inset shows the time series for the seismicity in the San Jacinto area (see Figure S3 in Supporting Information S1). Each white bar shows the corresponding periods used to train and validate ETAS and RECAST. RECAST approaches the theoretically optimal performance of the ETAS model as we increase the training set size. In contrast, for the real catalog, a training and validation catalog in excess of $\sim 10,000$ earthquakes (fourth white bar from the top in the inset) results in a test period best modeled by RECAST.

3.3. Comparative Performance on Other Data Sets

Comparable improvements arise when considering the entire Southern California Earthquake Data Center catalog from 1981 to 2021 (Hutton et al., 2010), and subregions of the Quake Template Matching Catalog (Ross et al., 2019) (Figure S2, S3–S5 and Table S1 in Supporting Information S1). Improvements are most pronounced in the smaller regions we consider. Consistent out-of-sample improvements on these disparate catalogs suggest that RECAST is robust to multiple methods of data production and regional variations in seismicity.

3.4. Southern California Earthquake Catalog and 14-day Forecasts

We finally compare earthquake forecasts generated using ETAS and RECAST on extended time intervals. Once trained, RECAST provides a straightforward way to simulate potential catalog continuations. We generate a sample earthquake by drawing an origin time from the mixture distribution in Equation 1 and a magnitude from the Gutenberg–Richter Distribution (Gutenberg & Richter, 1944). Repeatedly adding events to the catalog yields continuations spanning a range of potential outcomes. For implementation details, refer to Text S4 in Supporting Information S1. We refer to this set as an earthquake forecast. Specific aspects of a forecast, such as the abundance of events or the probability of a particularly large earthquake, can then be derived from statistics of the set of catalog continuations. Here, we consider 2-week earthquake forecasts with the Southern California earthquake catalog (Hutton et al., 2010) with an additional set of results for the San Jacinto fault catalog (Figures S7 and S8 in Supporting Information S1). Each 2-week forecast comprises 50,000 simulated catalog continuations during the testing period. We note that individual continuations exhibit realistic features, including aftershock decay, secondary bursts of aftershocks, and stochasticity (Figure 4a). To quantitatively compare forecasts, we compute the fraction of simulated continuations, r , that exactly reproduce the total number of events in the interval (Figure 4b and Figure S7B in Supporting Information S1). Forecast accuracy is measured by r and larger values represent more accurate forecasts (see Text S5 in Supporting Information S1 for further details and rationale).

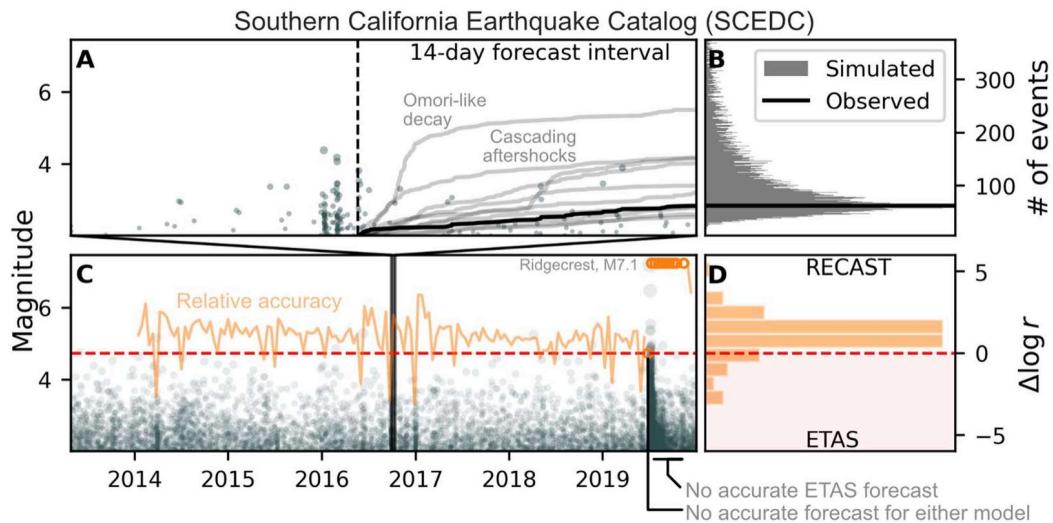


Figure 4. Evaluation of forecasts with RECAST and ETAS models for the Southern California earthquake catalog. (a) Example 2-week earthquake RECAST forecast issued at the vertical dashed line. The observed cumulative number of earthquakes is shown in black with sample RECAST simulations in gray. Punctuated increases in the cumulative number of events in these samples indicate spontaneous event clusters which gradually decay in time. (b) Full distribution of the cumulative number of events from the sampled trajectories compared to the observation (black). (c) Test catalog with the evolution of the log-score for the tested 14-day forecast intervals. Empty orange markers and the corresponding annotation indicate the intervals where both model forecasts (1 instances), or just the ETAS model forecast (9 instances) yielded no accurate forecast ($r = 0$ and $\log r$ is undefined). These all occur during the Ridgecrest earthquake sequence. (d) Comparison of the relative accuracy (forecast log-likelihood) of RECAST and ETAS models. A positive log-score indicates a more accurate RECAST forecast. In most intervals, the RECAST is more accurate.

The majority of 2-week intervals in the test set were best forecast by RECAST (138 out of 155 14-day intervals, Figure 4c-d, Figures S7C and S7D in Supporting Information S1). Best fit periods tend to be during background activity. Nonetheless, RECAST also performs better following the largest earthquakes. Another metric of success is the ability to accurately capture the full range of observations. 14-day windows with outcomes outside the 95% confidence interval of the forecasts are far less frequent for RECAST than for the ETAS model (14% vs. 33%, respectively, Figures S6 and S8 in Supporting Information S1). Outcomes not featured in the entire set of 50,000 simulations occurred only once for the RECAST forecasts and 10 times for the ETAS forecasts (Figure 4c). These full misses for ETAS occur during the Ridgecrest earthquake aftershock sequence. While RECAST does not foresee the Ridgecrest sequence, it quickly recovers relatively accurate forecasts after the onset of the sequence.

4. Discussion

Our analysis shows both the advantages and limitations of the RECAST model. Some are inherent to deep learning, while others stem from the choices in the model architecture.

RECAST is fundamentally different from ETAS. The exact functional form that relates catalog data to the likelihood of the next event time is not required. This aspect is well illustrated in our tests on synthetic catalogs. Prior to training, RECAST only assumes that inter-event times will be strictly positive. The RECAST architecture can be trained on a modest-sized data set to learn the relationship between the inputs, in this case, timing and magnitude, and the output, event likelihood. This flexibility to incorporate new data and relationships is a key advantage. In contrast, ETAS is a parametric model and requires known functional relationships. The benefit of the latter approach is that fewer parameters are required. The drawback is that the forecast is limited by the assumptions of the model which are often broken either due to physical changes in the system or because of artifacts in the preparation of the catalog data. RECAST may have the potential to outperform ETAS for real earthquake catalogs.

The new model requires a sufficiently large earthquake catalog for training. As exemplified in the benchmarks from the San Jacinto Fault zone (Figure 3b), for small data sets, the temporal ETAS model outperforms RECAST.

In this experiment, the model needs a larger catalog (10^4 events) before RECAST outperforms the benchmark. Much like ETAS, we anticipate poor performance if the training set is different from the period over which the model is evaluated. The improvements across every other test we conduct would suggest that the result is robust for such dense earthquake catalogs. Intriguingly, the relative gap in performance between RECAST and ETAS grows with catalog size, suggesting persistent improvements that may scale with increasingly available data. Unlike ETAS, RECAST's model architecture ensures that the volume of data currently available does not pose computational limits on this trend. Historically there was little reason for a method to use information for more than a few thousand events. Now catalogs offer more and demand the new technology.

It is difficult to diagnose the exact cause of improvements. We can, however, highlight some diagnostics. First, we note that the more pronounced improvements are in smaller subregions (Table S2, Figures S7 and S8 in Supporting Information S1). Second, it appears that the temporal ETAS model underfits event sequences and does not account for long-term trends in seismicity. In the case of the San Jacinto data set, ETAS under-predicts during periods of high seismicity and over-predicts during periods of low seismicity rates (e.g., Figure S8 in Supporting Information S1). In the case of the SCEDC catalog, ETAS forecasts reflect a high base rate of seismicity of the training set and are biased accordingly, consistently over-predicting the base rate in the testing interval (Figure S6 in Supporting Information S1). Although RECAST uses the same training periods in all cases, the model's memory (the hidden state in Equation 1) allows RECAST to account for time-varying trends in the validation and test sets and, as a result, may better track a constantly evolving seismicity rate. Altering the ETAS training procedure to retrain on a fixed window directly before each forecast might address this model limitation.

Improvements could stem from uncovering a physical process that governs the predictability of the system. Another interpretation is that deep learning models, such as RECAST, jointly address failure modes associated with standard parametric models, such as ETAS. Current practice entails many moving pieces. It is challenging to introduce and weigh the contribution of new assumptions about the data regarding non-stationarity (Kumazawa & Ogata, 2014; Llenos & Michael, 2019; Muir & Ross, 2023), anisotropy (Ogata, 2011), catalog incompleteness (Mizrahi et al., 2021; Page et al., 2016), magnitude errors (Werner & Sornette, 2008), changes in instrumental coverage (Woessner & Wiemer, 2005), pathologies, for example, orphaned aftershocks (van der Elst, 2017), criticality (Helmstetter & Sornette, 2001), and missing links (Wang et al., 2010). A subset of these factors, as implemented in state-of-the-art models, may close the performance gap between ETAS and RECAST. In our experiments, it seems most important to address long-term trends in seismicity either with a non-stationary ETAS implementation (Kumazawa & Ogata, 2014; Llenos & Michael, 2019) or with a sliding window for recalibration. Such prioritization is clear after the fact and may not robustly hold in further experiments. To the best of our knowledge, no model addresses the suite of limitations that ETAS models face concurrently, mainly due to the complexity involved in coordinating the competing effects while preserving the benefits of the original model. The advantages of deep learning as an approach may be that modeling builds on well-maintained and rapidly evolving libraries and provides a path forward relying less heavily on the idealizations of the data.

5. Conclusion

Dense earthquake catalogs reveal a richer window into the seismic cycle. Yet it has been a challenge to translate finer detail and increased data volume into improved forecasts. A deep-learning approach such as RECAST is sufficiently general to recover well-known statistical patterns. Our experiments on synthetic data sets suggest performance at least as capable as a temporal ETAS model provided a modest number of events and that RECAST may also capture processes that are not explicitly parameterized in an ETAS model on real-world catalogs.

This study is a proof of concept for the application of neural temporal point process models for earthquake forecasting. More generally, neural temporal point process models and deep learning provide important advantages. They do not require knowledge of the relationship between earthquake features to forecasting probabilities. With increasingly large earthquake data sets, their design enables increasingly complex relationships to be derived directly from the data. It remains to be seen whether the increase in performance is due to more accurately capturing the observational limitations of catalogs or reflects a physical process that controls long-term seismicity evolution. In either case, forecasters can capitalize on the production of enhanced catalogs by leveraging the flexibility and scalability of deep learning.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

All code used to produce figures, along with guidelines, data sets, and example workflows, are available for download here: <https://doi.org/10.5281/zenodo.8161777>. The following specifies the earthquake catalogs and links for data access: the Southern California Seismic Network Catalog (Hutton et al., 2010), <https://service.scedc.caltech.edu/ftp/catalogs/SCSN/>, the San Jacinto Catalog (White et al., 2019), <https://doi.org/10.17632/7y-wkdx7c62.1>, the Quake Template Matching catalog (Z. E. Ross et al., 2019), https://service.scedc.caltech.edu/ftp/QTMcatalog/qtm_final_12dev.hypo.

Acknowledgments

We thank the members of the UCSC seismology lab, Alba Rodriguez Padilla, Colleen Murphy, Coby Abrahams, and Nicholas Van Der Elst, for providing thoughtful insight and discussion. Funding: Supported by the Alexander Graham Bell postgraduate NSERC fellowship, NSF Grant EAR-1761987 and SCEC Grant 22114.

References

- Aalen, O., Borgan, O., & Gjessing, H. (2008). *Survival and event history analysis: A process point of view*. Springer.
- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). TensorFlow: A system for large-scale machine learning (pp. 265–283). Retrieved from <https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi>
- Beroza, G. C., Segou, M., & Mostafa Mousavi, S. (2021). Machine learning and earthquake forecasting—Next steps. *Nature Communications*, 12(1), 1. <https://doi.org/10.1038/s41467-021-24952-6>
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. ArXiv:1406.1078 [Cs, Stat]. Retrieved from <http://arxiv.org/abs/1406.1078>
- Du, N., Dai, H., Trivedi, R., Upadhyay, U., Gomez-Rodriguez, M., & Song, L. (2016). Recurrent marked temporal point processes: Embedding event history to vector. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1555–1564). <https://doi.org/10.1145/2939672.2939875>
- Field, E. H., Milner, K. R., Hardebeck, J. L., Page, M. T., van der Elst, N., Jordan, T. H., et al. (2017). A spatiotemporal clustering model for the third uniform California earthquake rupture forecast (UCERF3-ETAS): Toward an operational earthquake forecast. *Bulletin of the Seismological Society of America*, 107(3), 1049–1081. <https://doi.org/10.1785/0120160173>
- Gareth, J., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning (world; Second)*. Springer. <https://doi.org/10.5555/2517747>
- Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks (arXiv:1607.00653). arXiv Retrieved from <http://arxiv.org/abs/1607.00653>
- Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, 34(4), 4–188. <https://doi.org/10.1785/bssa0340040185>
- Helmstetter, A., & Sornette, D. (2001). Sub-critical and super-critical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research*, 107(B10), 10–11. <https://doi.org/10.1029/2001JB001580>
- Hutton, K., Woessner, J., & Hauksson, E. (2010). Earthquake monitoring in southern California for seventy-seven years (1932–2008). *Bulletin of the Seismological Society of America*, 100(2), 423–446. <https://doi.org/10.1785/0120090130>
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., et al. (2020). Scaling laws for neural language models (arXiv:2001.08361). arXiv. Retrieved from <http://arxiv.org/abs/2001.08361>
- Kumazawa, T., & Ogata, Y. (2014). Nonstationary ETAS models for nonstandard earthquakes. *Annals of Applied Statistics*, 8(3), 1825–1852. <https://doi.org/10.1214/14-AOAS759>
- Llenos, A. L., & Michael, A. J. (2017). Forecasting the (un)productivity of the 2014 M 6.0 South Napa aftershock sequence. *Seismological Research Letters*, 88(5), 1241–1251. <https://doi.org/10.1785/0220170050>
- Llenos, A. L., & Michael, A. J. (2019). Ensembles of ETAS models provide optimal operational earthquake forecasting during swarms: Insights from the 2015 San Ramon, California Swarm. *Bulletin of the Seismological Society of America*, 109(6), 2145–2158. <https://doi.org/10.1785/0120190020>
- Mai, P. M., Schorlemmer, D., Page, M., Ampuero, J., Asano, K., Causse, M., et al. (2016). The earthquake-source inversion validation (SIV) project. *Seismological Research Letters*, 87(3), 690–708. <https://doi.org/10.1785/0220150231>
- Mancini, S., Segou, M., Werner, M. J., Parsons, T., Beroza, G., & Chiaraluce, L. (2022). On the use of high-resolution and deep-learning seismic catalogs for short-term earthquake forecasts: Potential benefits and current limitations. *Journal of Geophysical Research: Solid Earth*, 127(11), e2022JB025202. <https://doi.org/10.1029/2022JB025202>
- Mei, H., & Eisner, J. (2017). The neural hawkes process: A neurally self-modulating multivariate point process. ArXiv:1612.09328 [Cs, Stat]. Retrieved from <http://arxiv.org/abs/1612.09328>
- Mignan, A., & Broccardo, M. (2020). Neural network applications in earthquake prediction (1994–2019): Meta-analytic and statistical insights on their limitations. *Seismological Research Letters*, 91(4), 2330–2342. <https://doi.org/10.1785/0220200021>
- Mizrahi, L., Nandan, S., & Wiemer, S. (2021). Embracing data incompleteness for better earthquake forecasting. *Journal of Geophysical Research: Solid Earth*, 126(12), e2021JB022379. <https://doi.org/10.1029/2021JB022379>
- Mousavi, S. M., & Beroza, G. C. (2023). Machine learning in earthquake seismology. *Annual Review of Earth and Planetary Sciences*, 51(1), 105–129. <https://doi.org/10.1146/annurev-earth-071822-100323>
- Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11(1), 3952. <https://doi.org/10.1038/s41467-020-17591-w>
- Muir, J. B., & Ross, Z. E. (2023). A deep Gaussian process model for seismicity background rates. *Geophysical Journal International*, 234(1), 427–438. <https://doi.org/10.1093/gji/ggad074>
- Obara, K. (2003). Hi-net: High sensitivity seismograph network, Japan. In T. Takanami & G. Kitagawa (Eds.), *Methods and applications of signal processing in seismic network operations* (pp. 79–88). Springer. <https://doi.org/10.1007/BFb0117698>
- Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401), 9–27. <https://doi.org/10.1080/01621459.1988.10478560>

- Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2), 379–402. <https://doi.org/10.1023/A:1003403601725>
- Ogata, Y. (2011). Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. *Earth, Planets and Space*, 63(3), 217–229. <https://doi.org/10.5047/eps.2010.09.001>
- Ogata, Y. (2017). Statistics of earthquake activity: Models and methods for earthquake predictability studies. *Annual Review of Earth and Planetary Sciences*, 45(1), 497–527. <https://doi.org/10.1146/annurev-earth-063016-015918>
- Omi, T., Ueda, N., & Aihara, K. (2019). Fully neural network based model for general temporal point processes. *Advances in Neural Information Processing Systems*, 32. <https://github.com/omitakahiro/NeuralNetworkPointProcess>
- Page, M. T., van der Elst, N. J., Hardebeck, J. L., Felzer, K. R., & Michael, A. J. (2016). Three ingredients for improved global aftershock forecasts: Tectonic region, time-dependent catalog incompleteness, and intersequence variability. *Bulletin of the Seismological Society of America*, 106(5), 2290–2301. <https://doi.org/10.1785/0120160073>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32. <https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bf9f7012727740-Abstract.html>
- Ross, G. J. (2021). Bayesian estimation of the ETAS model for earthquake occurrences. *Bulletin of the Seismological Society of America*, 111(3), 1473–1480. <https://doi.org/10.1785/0120200198>
- Ross, Z. E., Trugman, D. T., Hauksson, E., & Shearer, P. M. (2019). Searching for hidden earthquakes in Southern California. *Science*, 364(6442), 767–771. <https://doi.org/10.1126/science.aaw6888>
- Shchur, O., Biloš, M., & Günemann, S. (2020). Intensity-free learning of temporal point processes. *International conference on learning representations*. Retrieved from <http://arxiv.org/abs/1909.12127>
- Shchur, O., Türkmen, A. C., Januschowski, T., & Günemann, S. (2021). Neural temporal point processes: A review. International joint conference on artificial intelligence. <http://arxiv.org/abs/2104.03528>
- Shelly, D. R. (2017). A 15 year catalog of more than 1 million low-frequency earthquakes: Tracking tremor and slip along the deep San Andreas Fault. *Journal of Geophysical Research: Solid Earth*, 122(5), 3739–3753. <https://doi.org/10.1002/2017JB014047>
- Tan, Y. J., Waldhauser, F., Ellsworth, W. L., Zhang, M., Zhu, W., Michele, M., et al. (2021). Machine-learning-based high-resolution earthquake catalog reveals how complex fault structures were activated during the 2016–2017 central Italy sequence. *The Seismic Record*, 1(1), 11–19. <https://doi.org/10.1785/0320210001>
- Upadhyay, U., De, A., & Gomez Rodriguez, M. (2018). Deep reinforcement learning of marked temporal point processes. *Advances in Neural Information Processing Systems*, 31. <https://proceedings.neurips.cc/paper/2018/hash/71a58e8cb75904f24cde464161c3e766-Abstract.html>
- van der Elst, N. J. (2017). Accounting for orphaned aftershocks in the earthquake background rate. *Geophysical Journal International*, 211(2), 1108–1118. <https://doi.org/10.1093/gji/ggx329>
- van der Elst, N. J., Hardebeck, J. L., Michael, A. J., McBride, S. K., & Vanacore, E. (2022). Prospective and retrospective evaluation of the U.S. geological survey public aftershock forecast for the 2019–2021 southwest Puerto Rico earthquake and aftershocks. *Seismological Research Letters*, 93(2A), 620–640. <https://doi.org/10.1785/0220210222>
- Wang, Q., Jackson, D. D., & Zhuang, J. (2010). Missing links in earthquake clustering models. *Geophysical Research Letters*, 37(21). <https://doi.org/10.1029/2010GL044858>
- Werner, M. J., & Sornette, D. (2008). Magnitude uncertainties impact seismic rate estimates, forecasts, and predictability experiments. *Journal of Geophysical Research*, 113(B8). <https://doi.org/10.1029/2007JB005427>
- White, M. C. A., Ben-Zion, Y., & Vernon, F. L. (2019). A detailed earthquake catalog for the san Jacinto Fault-zone region in southern California. *Journal of Geophysical Research: Solid Earth*, 124(7), 6908–6930. <https://doi.org/10.1029/2019JB017641>
- Woessner, J., & Wiemer, S. (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty. *Bulletin of the Seismological Society of America*, 95(2), 684–698. <https://doi.org/10.1785/0120040007>
- Zhou, Z., Yang, X., Rossi, R., Zhao, H., & Yu, R. (2021). Neural point process for learning spatiotemporal event dynamics. ArXiv:2112.06351 [Cs] Retrieved from <http://arxiv.org/abs/2112.06351>
- Zhu, S., Li, S., Peng, Z., & Xie, Y. (2021). Imitation learning of neural spatio-temporal point processes. *IEEE Transactions on Knowledge and Data Engineering*, 34(11), 1–5402. <https://doi.org/10.1109/TKDE.2021.3054787>
- Zhuang, J. (2011). Next-day earthquake forecasts for the Japan region generated by the ETAS model. *Earth, Planets and Space*, 63(3), 3–216. <https://doi.org/10.5047/eps.2010.12.010>

References From the Supporting Information

- Felzer, K. R., Abercrombie, R. E., & Ekström, G. (2004). A common origin for aftershocks, foreshocks, and multiplets. *Bulletin of the Seismological Society of America*, 94(1), 88–98. <https://doi.org/10.1785/0120030069>
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378. <https://doi.org/10.1198/016214506000001437>
- Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. ArXiv:1412.6980 [Cs]. Retrieved from <http://arxiv.org/abs/1412.6980>
- Ogata, Y., & Katsura, K. (1993). Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues. *Geophysical Journal International*, 113(3), 727–738. <https://doi.org/10.1111/j.1365-246x.1993.tb04663.x>