

SCEC Project Report 25227: Space-time neural point process for earthquake forecasting

Kelian Dascher-Cousineau (PI) and Daniel Wells, Utah State University

1. Objectives and summary

Neural temporal point processes (NTPPs) parameterize the probability distribution of the next earthquake directly from the event history and can be trained efficiently on very large catalogs. Our time-only Recurrent Earthquake foreCAST (RECAST) outperformed a temporal Epidemic-Type Aftershock Sequence (ETAS) benchmark on large southern California catalogs, with an advantage that grew with the size of the training set (Dascher-Cousineau et al., 2023). This project set out to extend RECAST to joint space-time forecasting and to test it rigorously against a space-time ETAS (ST-ETAS) model. The award supported a few months of postdoctoral effort; progress on the four proposed tasks is summarized in Table 1.

The main outcomes are: (i) a working space-time model, ST-RECAST, in which the location of the next event is explicitly conditioned on its timing; (ii) a pipeline to train ST-RECAST and compare it head-to-head with ST-ETAS; (iii) the finding that ST-RECAST does not yet match ST-ETAS: on synthetic catalogs it largely reproduces the temporal clustering but issues overly diffuse spatial forecasts, and its skill improves steadily, but slowly, with the number of training sequences; and (iv) a global, temporal version of RECAST, trained on sub-catalogs around roughly a thousand $M \geq 6.5$ mainshocks worldwide, that issues better-calibrated and more accurate aftershock forecasts than ETAS fitted to local data. Outcome (iv) grew out of (iii): our working hypothesis is that a space-time NTPP needs more training data than is available locally.

Table 1. Progress on the proposed tasks.

Proposed task	Status
1. Implementation. ST-RECAST with locations conditioned on event timing; sampling routines for simulation-based forecasts.	Completed: model implemented with locations conditioned on event timing; it generates space-time forecasts (Section 2.1; Figure 2D).
2. Synthetic tests. Iterative development on catalogs simulated with ST-ETAS.	Tests run with up to 5931 training and 1977 test catalogs: ST-RECAST improves with training-set size but does not yet reach the ST-ETAS benchmark (Section 2.3). Development is ongoing.
3. State-wide testing against benchmark models and datasets.	Partially completed: head-to-head tests against ST-ETAS on the southern California catalog, where ST-RECAST also underperforms. Other benchmark datasets (e.g., EarthquakeNPP; Stockman et al., 2026) and CSEP consistency tests are pending.
4. Larger earthquakes. Document performance for large events.	Not yet done for ST-RECAST. Related work: aftershock forecasts following $M \geq 6.5$ mainshocks with a globally trained temporal model (Section 3).

2. Space-time RECAST

2.1 Model. RECAST encodes the event history $H_i = \{(t_j, m_j): j < i\}$ with a gated recurrent unit into a latent state and decodes that state into a mixture of Weibull distributions for the waiting time $\tau_i = t_i - t_{i-1}$ (Figure 1). Because the density is parameterized directly rather than through the conditional intensity (Shchur et al., 2020), likelihood evaluation and sampling are exact and inexpensive. For the space-time task we factorize the density of the next event as

$$f(\tau_i, x_i, y_i | H_i) = f(\tau_i | H_i) \cdot f(x_i, y_i | \tau_i, H_i). \quad (1)$$

The second factor is the crux. Our preliminary models (shown in the proposal) treated location as independent of timing, $f(x_i, y_i | H_i)$. That shortcut discards information. Even in ST-ETAS, whose triggering kernel is separable in time and space for each parent event, the location of the next event depends on when it

occurs, because the relative weights of recent parents and of the background change with the waiting time. In an illustrative ST-ETAS calculation with an M6 mainshock as the latest event, more than 99% of the probability mass for the next location lies in the mainshock’s aftershock kernel if that event occurs one hour later, but only about a fifth if it occurs 30 days later; the rest comes from the background. ST-RECAST therefore re-encodes the waiting time: the decoder state is updated with τ_i before it parameterizes a mixture density over location (Figure 1), analogous to the time-conditioned spatial densities of neural spatio-temporal point processes (Chen et al., 2021). During training, τ_i is the observed waiting time. Forecasts are generated by simulation: τ_i is drawn first, then the location from its conditional density, and the event is appended to the history before the next draw.

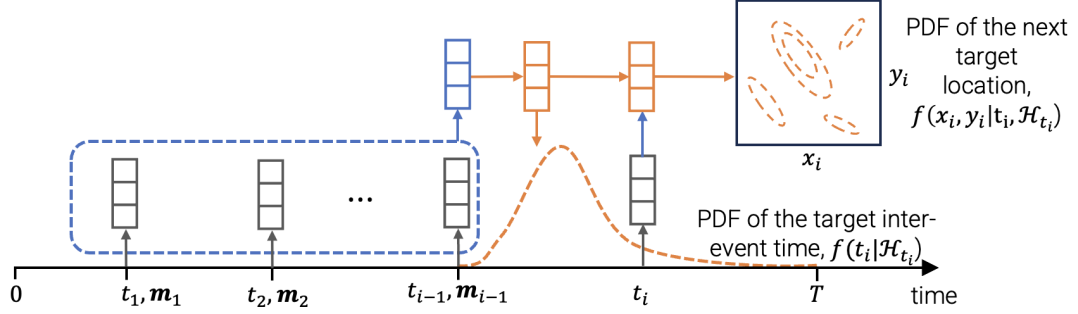


Figure 1. ST-RECAST architecture (adapted from the proposal). A recurrent encoder summarizes the event history (times t_j and marks m_j) into a latent state that is decoded into the density of the next waiting time; the waiting time is then re-encoded to condition the density of the next location.

2.2 Benchmark and testing machinery. The benchmark is the ST-ETAS model of Ogata (1998), with a spatially variable background rate and a magnitude-dependent power-law spatial kernel. Both models are fitted on training sequences and compared on held-out test sequences, by their negative log-likelihood (NLL) per day of catalog and by the space-time forecasts they generate.

2.3 Results. Figure 2 summarizes tests on catalogs simulated with ST-ETAS. Because these catalogs are generated by ST-ETAS, the benchmark has the correct functional form and is close to the best attainable forecaster, so the question is how close ST-RECAST gets for a given amount of training data. ST-RECAST captures the timing of the clustering: in the example shown, its conditional intensity rises and decays with that of ETAS through each burst, although its integrated intensity falls short of the observed count (Figure 2A–B). Its space-time forecasts are much more diffuse than those of ETAS, which concentrate rate around recent activity (Figure 2D–E). With 5931 training sequences, ETAS still has the lower NLL on the large majority of the 1977 test sequences, and the training and test distributions are similar, so the deficit is systematic rather than a symptom of overfitting (Figure 2C). In an earlier sweep with up to 600 training sequences, the test NLL of ST-RECAST improved steadily with training-set size, from ≈ 22.7 per day with 5–25 sequences to ≈ 20.5 with 100 and ≈ 18.5 with 600, against 14.2 per day for ETAS (Figure 2F). On the southern California catalog, ST-RECAST likewise trails ST-ETAS. We regard these results as provisional: model development is ongoing.

2.4 Interpretation. ST-ETAS encodes strong, physically grounded structure with a handful of parameters: power-law decay of triggering in time and with distance from each parent, magnitude-dependent productivity and spatial extent, and a spatially variable background rate. ST-RECAST must learn the equivalent structure, including where aftershocks fall relative to their parents, from the events themselves. Earthquake forecasting is label-poor in an important sense: a denser catalog adds events, but the number of independent sequences, and of large events, grows slowly. In the temporal problem, RECAST needed roughly 10^4 training events to overtake temporal ETAS on the San Jacinto catalog (Dascher-Cousineau et al., 2023). The synthetic tests show that the space-time problem is far more demanding: ST-RECAST improves steadily with training data, yet with about 6000 training sequences ETAS still performs better on most test sequences. A regional catalog spanning a few decades contains only a few hundred non-overlapping 30-day windows, an order of magnitude fewer than the training sets used here. If this reasoning

holds, progress depends on pooling training data across regions as much as on architecture, which motivated the global extension below.

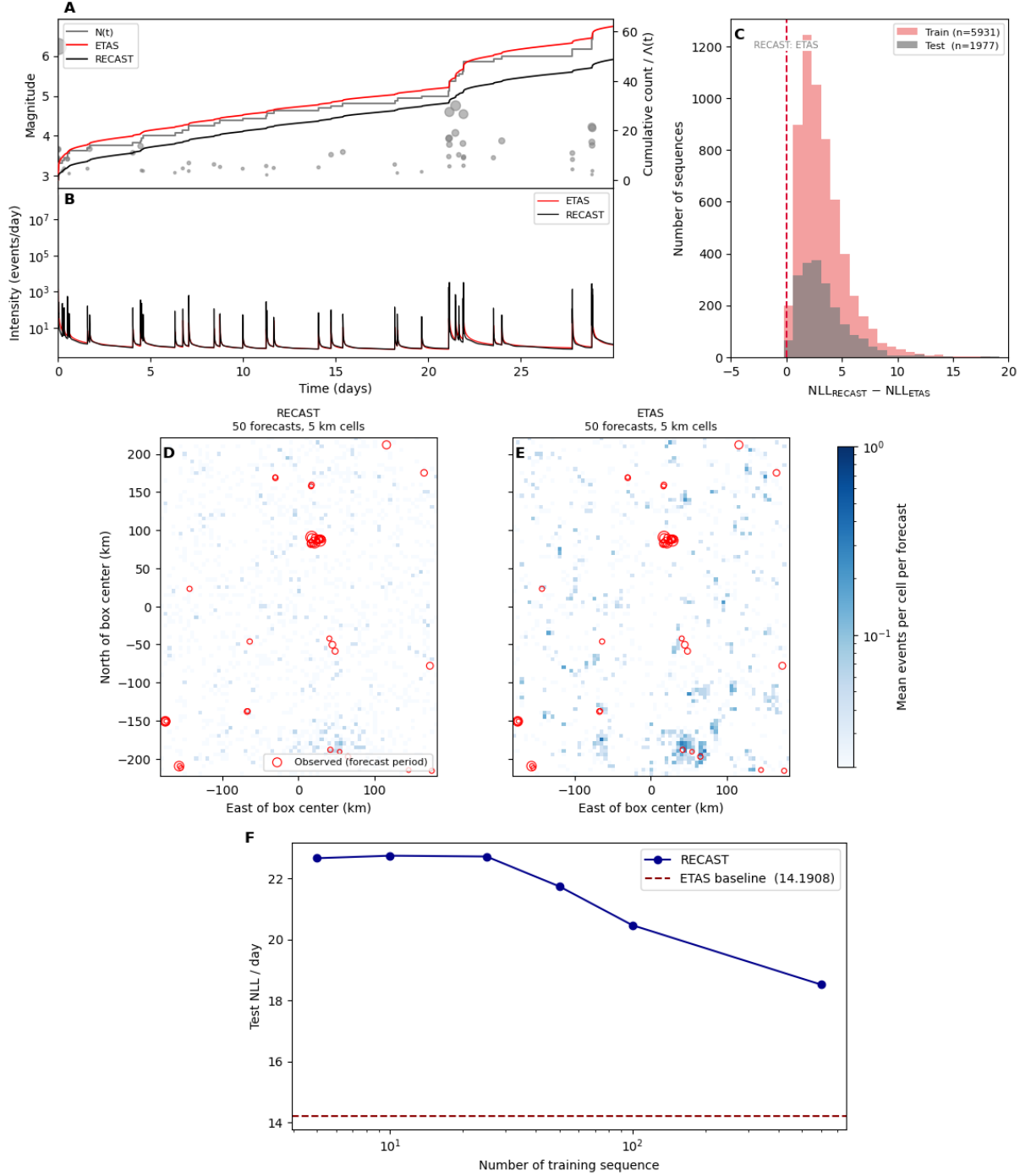


Figure 2. ST-RECAST versus ST-ETAS on synthetic ST-ETAS catalogs. (A) Example 30-day sequence: magnitudes (gray), observed cumulative count $N(t)$ and each model's integrated intensity $\Lambda(t)$. (B) Conditional intensities for the same sequence. (C) Per-sequence difference in time-averaged NLL (RECAST minus ETAS) for training ($n = 5931$) and test ($n = 1977$) sequences; positive values favor ETAS. (D, E) Mean number of events per 5 km cell over 50 forecasts from each model, with events observed during the forecast periods (red circles). (F) Test NLL per day versus number of training sequences, from an earlier sweep with up to 600 training sequences; dashed line, ETAS (14.19).

3. A global RECAST model for aftershock forecasting

3.1 Motivation. Operational aftershock forecasts are, to first order, temporal: they give the expected number of aftershocks above a magnitude threshold in a region over days to months after a mainshock. They also lean on information pooled across many sequences, for example generic parameters by tectonic regime that are updated with sequence-specific data (Page et al., 2016; van der Elst et al., 2022). We asked whether a single RECAST model, trained on aftershock sequences worldwide, can forecast a new sequence from its local history alone, without retraining on local data.

3.2 Data and training machinery. We build sub-catalogs from the ANSS Comprehensive Earthquake Catalog (ComCat; U.S. Geological Survey, 2017) at $M \geq 4.5$. Each sub-catalog contains all events within 250 km of a randomly drawn $M \geq 6.5$ mainshock; a randomly placed 1-year window that brackets the mainshock defines the interval over which the likelihood is optimized, and all events since 1990 that precede the window serve as conditioning history. Mainshocks from 1990–2010, 2010–2015 and 2015–2020 form the training (3000 sub-catalogs), validation (1000) and test (1000) sets. ComCat contains only about 1000 $M \geq 6.5$ events from 1990–2010 and about 270 from 2010–2015, so sub-catalogs reuse mainshocks with different windows and neighboring sequences overlap: the number of independent sequences is much smaller than the number of sub-catalogs. To learn from many catalogs of uneven quality, we built machinery in which each sub-catalog carries its own magnitude of completeness and the contributions of the sub-catalogs to the training loss are balanced. Because independent sequences are scarce, we also explored data augmentations that encode known properties of earthquake catalogs: temporal dilation and contraction; re-windowing in magnitude, in time (which orphans aftershocks from their parents) and in space; and perturbation of magnitudes, locations and times within their uncertainties. In preliminary tests, augmentation improved RECAST’s validation performance.

3.3 Benchmark and evaluation. For each sub-catalog, a temporal ETAS model is fitted to all events from 1990 up to the forecast issue time, 10 minutes after the mainshock. Both models are then conditioned on the same history and generate thousands of event-by-event continuations over the following 14 days, which define the forecast distribution of the number of $M \geq 4.5$ aftershocks. We measure calibration by the quantile of the observed count within each forecast distribution, which is uniform for a calibrated forecast, and accuracy by the ratio $r = (N_{\text{obs}} + 1)/(N_{\text{pred}} + 1)$ between the observed count and the median forecast count, summarized by the median relative error.

3.4 Results. Across the validation sub-catalogs (Figure 3), RECAST is better calibrated and more accurate than ETAS: the median relative error drops from 3.4 to 2.5 (–26%), and fewer outcomes fall in the extreme tails of the forecast distribution. Because the validation set was also used for early stopping, these numbers are somewhat optimistic; the test set (mainshocks from 2015–2020) is held out for final evaluation and is not reported here. Neither model is well calibrated: both place far more outcomes in the end bins than the roughly 10% expected for a calibrated forecast. In the validation results, RECAST’s gain comes mostly from fewer over-predictions, and it still under-predicts a substantial fraction of sequences.

3.5 Caveats. Two caveats temper these results. First, the benchmark ETAS is fitted to each sub-catalog’s local history, whereas operational systems combine generic, pooled parameters with sequence-specific updating (Page et al., 2016). A comparison against ETAS with globally or tectonically pooled priors is the necessary next test. Second, because RECAST learns from the catalog as recorded, part of its advantage may come from learning systematic catalog features, such as short-term aftershock incompleteness after large mainshocks, that the locally fitted ETAS does not model. This is operationally useful, but it should be distinguished from improved physical forecasting (Dascher-Cousineau et al., 2023).

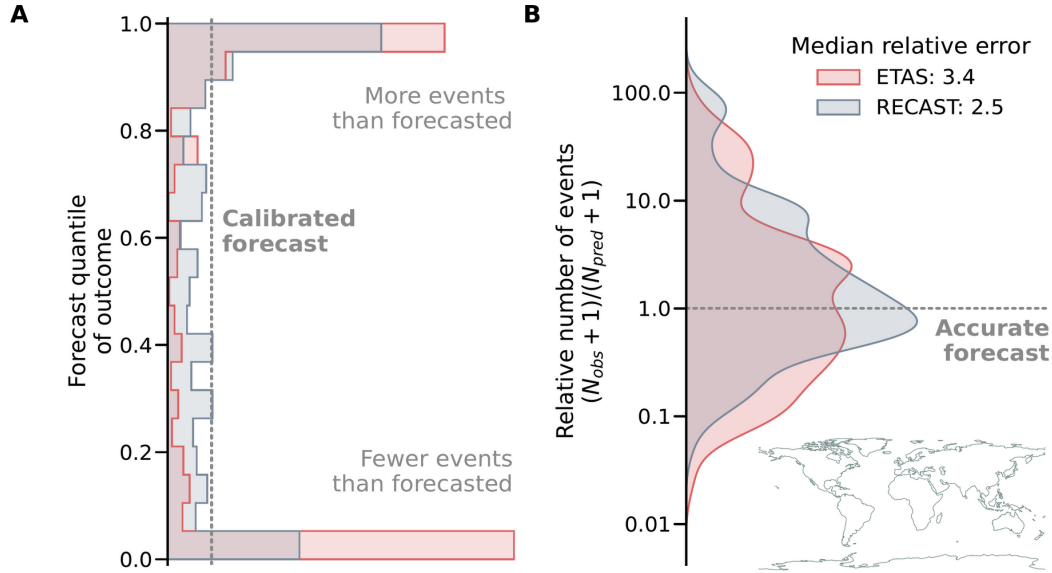


Figure 3. Performance of global RECAST and locally fitted ETAS on the validation sub-catalogs (also used for early stopping), for 14-day forecasts issued 10 minutes after the first $M \geq 6.5$ mainshock in each. (A) Calibration: histogram of the quantile of the observed count within each forecast distribution. A calibrated forecast is uniform (dotted line); the end bins collect outcomes in the tails of the forecast distribution (roughly below its 5th or above its 95th percentile). (B) Accuracy: distribution of the ratio of observed to median forecast counts, $(N_{\text{obs}} + 1)/(N_{\text{pred}} + 1)$; legend values are median relative errors.

4. Status and next steps

The space-time model and its testing framework exist; the space-time comparison is not yet favorable but improves with training data; and the global temporal model already improves on locally fitted ETAS in a configuration that mirrors operational aftershock forecasting. The next steps follow directly: (1) train ST-RECAST with the global sub-catalog machinery and data augmentation, which may supply the training volume the regional experiments lacked, then re-test it in California; (2) extend the head-to-head tests to other benchmark datasets (e.g., EarthquakeNPP) with CSEP consistency tests; and (3) benchmark the global temporal model against ETAS with pooled priors, on the held-out test set and pseudo-prospectively on recent sequences.

Data and code availability. Code developed under this award is publicly available at https://github.com/danedwells/recast_2, a fork of the RECAST repository (<https://github.com/keliankaz/recast>; Dascher-Cousineau et al., 2023). Earthquake data: ANSS ComCat (U.S. Geological Survey, 2017) and SCEDC (2013).

References

- Chen, R. T. Q., Amos, B., & Nickel, M. (2021). Neural spatio-temporal point processes. *International Conference on Learning Representations (ICLR)*.
- Dascher-Cousineau, K., Shchur, O., Brodsky, E. E., & Günnemann, S. (2023). Using deep learning for flexible and scalable earthquake forecasting. *Geophysical Research Letters*, 50(17), e2023GL103909. <https://doi.org/10.1029/2023GL103909>
- Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2), 379–402.
- Page, M. T., van der Elst, N., Hardebeck, J., Felzer, K., & 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.

- SCEDC (2013). Southern California Earthquake Data Center. Caltech. Dataset.
<https://doi.org/10.7909/C3WD3XH1>
- Shchur, O., Biloš, M., & Günnemann, S. (2020). Intensity-free learning of temporal point processes. *International Conference on Learning Representations (ICLR)*.
- Stockman, S., Lawson, D. J., & Werner, M. J. (2026). EarthquakeNPP: A benchmark for earthquake forecasting with neural point processes. *arXiv:2410.08226v3* (first version 2024).
- U.S. Geological Survey, Earthquake Hazards Program (2017). Advanced National Seismic System (ANSS) Comprehensive Catalog of Earthquake Events and Products.
<https://doi.org/10.5066/F7MS3QZH>
- 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.