Earthquake forecast evaluation using space-time point process residual analysis

Robert Alan Clements[1], Frederic Paik Schoenberg[1], and Danijel Schorlemmer[2]

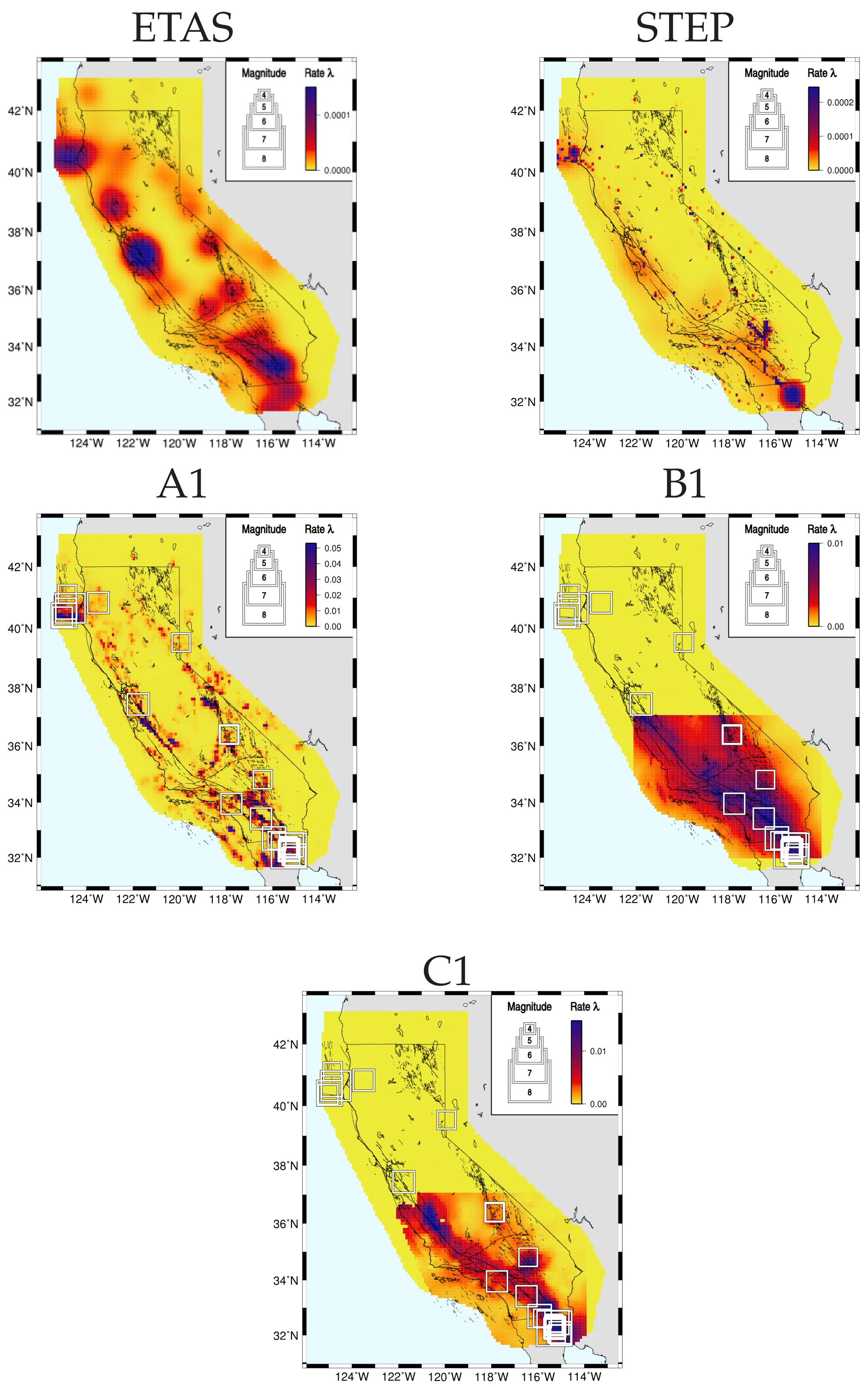
[1]UCLA Department of Statistics, [2]USC Southern California Earthquake Center

Introduction

A collection of earthquake (eq) forecast models from the Collaboratory for the Study of EQ Predictability have traditionally been tested using numerical summaries and error diagrams. Here, we propose using residual analysis methods for space-time point processes to evaluate and compare a subset of these models. Residuals such as these can help reveal specific locations where the forecast is failing.

EQ Forecast Models

Each model forecasts a rate of eqs in space-time-magnitude bins in California. Models are divided into 1-day models (ETAS and STEP) and 5-year models (A1 [3], B1 [4], and C1 [5]) with forecast period 2006-2011. 1-day models forecast eqs of M ≥ 3.95. 5-year models are extrapolated to forecast eqs of M ≥ 3.95.



Conditional Intensity

EQ occurrences can be modeled as a space-time point process, N, defined on a compact set, $\mathcal{S} \subset \mathbb{R}^d$. To model N, one typically models the conditional intensity $\lambda(\mathbf{x}|\mathcal{H}_t) < \infty$, defined as the rate at which eqs are expected to occur around a location $\mathbf{x}$, given the prior history $\mathcal{H}_t$.

Pixel-based Residuals

The testing region, $\mathcal{S}$, is divided into bins, B_i. Residuals can then be computed within each B_i using the estimated model for the conditional intensity, $\hat{\lambda}$.

Pearson residuals [1]:

$$R_P(B_i) = \sum_{(\mathbf{x}_i)\in B_i} \frac{1}{\sqrt{\hat{\lambda}(\mathbf{x}_i|\mathcal{H}_t)}} - \int_{B_i} \sqrt{\hat{\lambda}(\mathbf{x}|\mathcal{H}_t)}d\mathbf{x}$$

Deviance residuals of Model 1 vs Model 2 [5] (a model comparison tool):

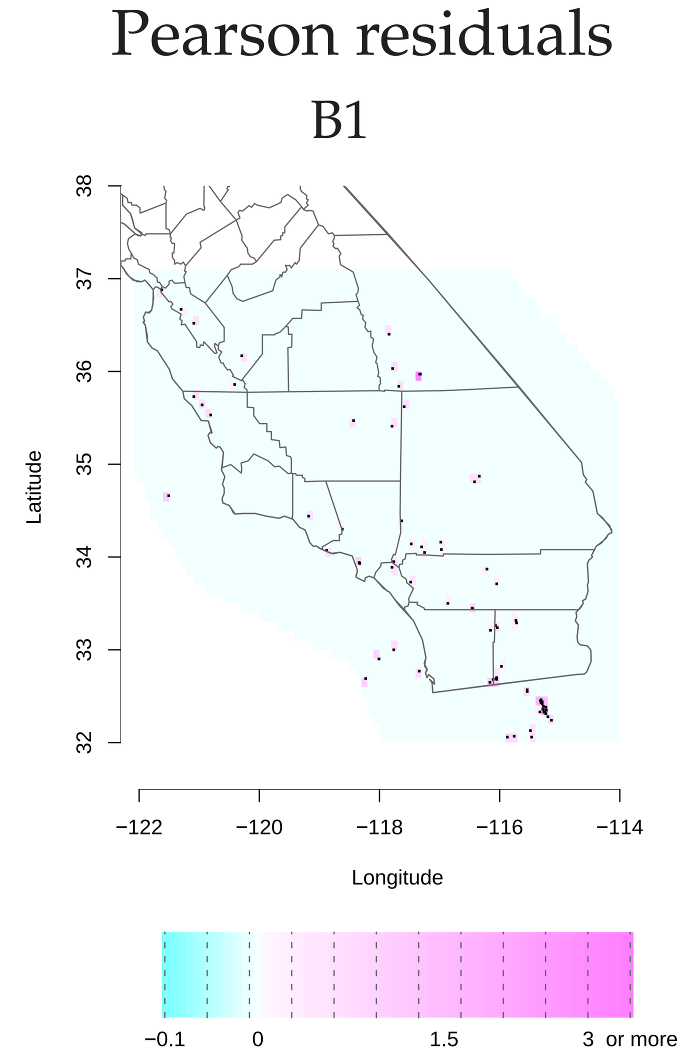
$$R_D(B_i) = \sum_{(\mathbf{x}_i)\in B_i} \log\hat{\lambda}_1 - \int_{B_i} \hat{\lambda}_1 d\mathbf{x} - \left(\sum_{(\mathbf{x}_i)\in B_i} \log\hat{\lambda}_2 - \int_{B_i} \hat{\lambda}_2 d\mathbf{x} \right)$$

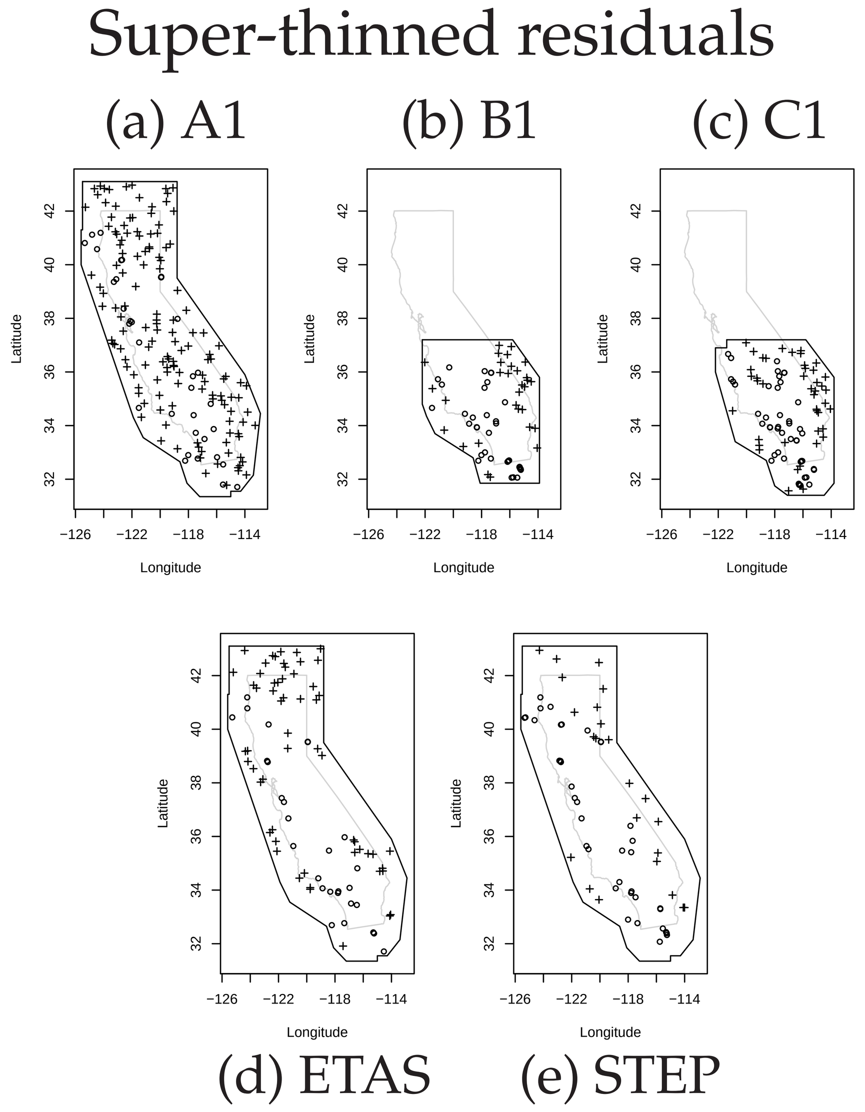
Super-thinned residuals

Using the super-thinning method, N can be transformed into Z, a homogeneous Poisson process, iff $\hat{\lambda} = \lambda$. Any clustering, inhibition, or spatial trend in Z is indicative of a lack-of-fit in those locations.

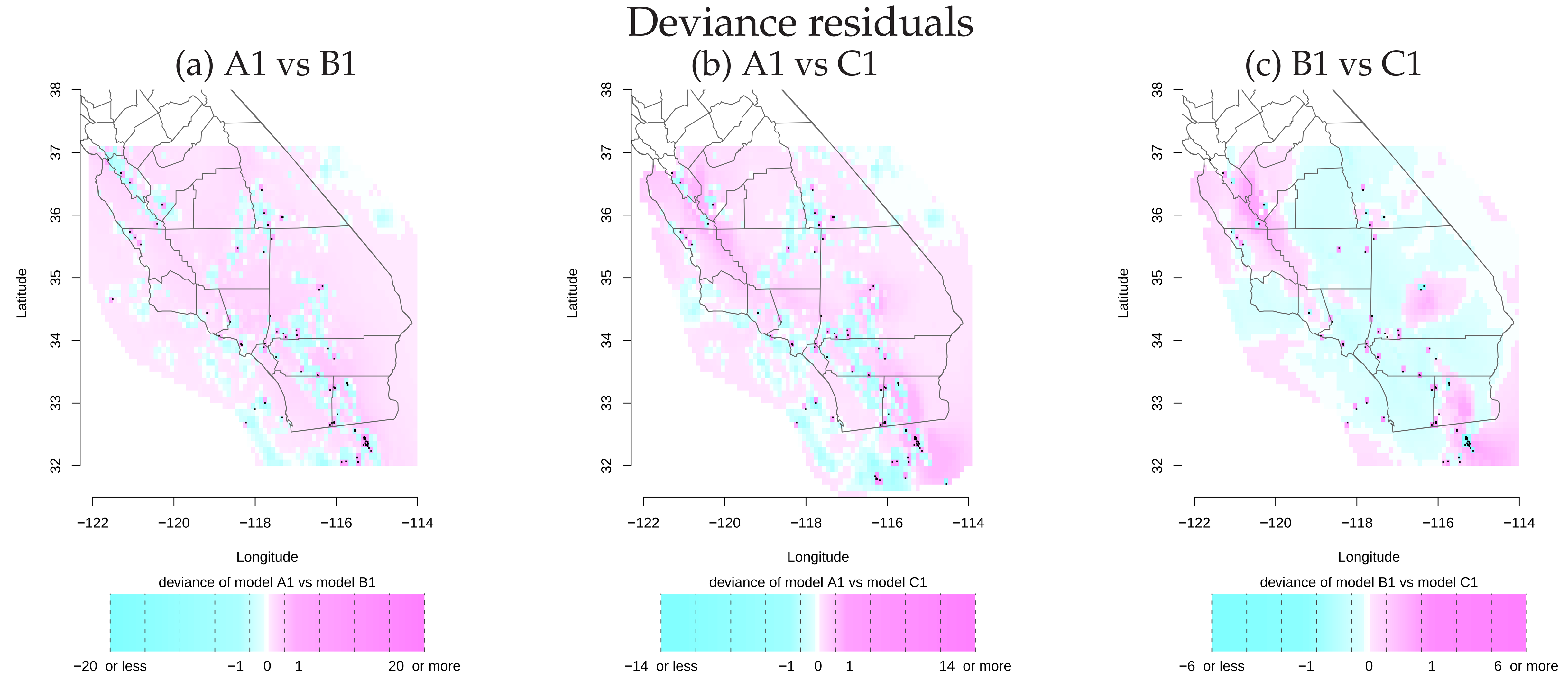
- **superpose** points if $\hat{\lambda}(\mathbf{x}|\mathcal{H}_t) < k$, $\inf\{\hat{\lambda}\} \leq k \leq \sup\{\hat{\lambda}\}$
  - simulate points with rate $k - \hat{\lambda}(\mathbf{x}|\mathcal{H}_t)$
- **thin** points if $\hat{\lambda}(\mathbf{x}_i|\mathcal{H}_t) \geq k$
  - keep each point with probability $k/\hat{\lambda}(\mathbf{x}_i|\mathcal{H}_t)$
- the result, called ***super-thinned*** residuals, is homogeneous Poisson, with rate k, iff $\hat{\lambda} = \lambda$ [2]

Results





The Pearson residuals typically reveal only the largest residuals which tend to occur in bins containing eqs. Clustering leftover in the Imperial fault zone (directly south of the California-Mexico border), and large areas of empty space in the super-thinned residuals imply under-prediction and over-prediction, respectively.



pink $\Longrightarrow$ Model 1 is preferred; blue $\Longrightarrow$ Model 2 is preferred. Overall, Model A1 > C1 > B1; ETAS > STEP (not shown). The results indicate over-prediction of seismicity in inter-fault zones for A1; under-prediction in B1 and C1 around the Imperial, Laguna Salada, Baja, and Panamint clusters.

References

[1] A. Baddeley, R. Turner, J. Moller, and M. Hazelton (2005). Residual analysis for spatial point processes. *JRSS* **67** 617-666.
[2] R. Clements, F. Schoenberg, and A. Veen (2010). Evaluation of space-time point process models using super-thinning. *UCLA Statistics Preprints* **579**.
[3] A. Helmstetter, Y. Kagan, and D. Jackson (2007). High-resolution time-independent grid-based forecast M$\geq$5 earthquakes in California. *SRL* **78** 78-86.
[4] Y. Kagan, D. Jackson, and Y. Rong (2007). A testable five-year forecast of moderate and large earthquakes in southern California based on smoothed seismicity. *SRL* **78** 94-98.
[5] Z. Shen, D. Jackson, and Y. Kagan (2007). Implications of geodetic strain rate for future earthquakes, with a five-year forecast of M5 earthquakes in southern California. *SRL* **78** 116-120.
[6] K. Wong and F. Schoenberg (2010). On mainshock focal mechanisms and the spatial distribution of aftershocks. *In review*.