Submitted to the Annals of Applied Statistics

A COUPLED ETAS-I^{2}GMM POINT PROCESS WITH APPLICATIONS TO SEISMIC FAULT DETECTION

By Yicheng Cheng Murat Dundar George Mohler

Indiana University Purdue University Indianapolis

Abstract Epidemic-type aftershock sequence (ETAS) point processes are a common model for the occurrence of earthquake events. ETAS models consist of a stationary background Poisson process modeling spontaneous earthquakes and a triggering kernel representing the space-time-magnitude distribution of aftershocks. Two popular non-parametric methods for estimation of the background intensity include histograms and kernel density estimators. While these methods are able to capture local spatial heterogeneity in the intensity of spontaneous events, they do not capture patterns resulting from fault-line structure over larger spatial scales. Here we propose a two-layer infinite Gaussian mixture model for clustering of earthquake events into fault-like groups over intermediate spatial scales. We introduce a Monte-Carlo expectation-maximization (EM) algorithm for joint inference of the ETAS-I^{2}GMM model and then apply the model to the Southern California earthquake Catalog. We illustrate the advantages of the ETAS-I^{2}GMM model in terms of both goodness of fit of the intensity and recovery of fault line clusters in the Community Fault Model 3.0 from earthquake occurrence data.

1. Introduction.

1.1. *Background on point-process models of seismicity.* The Epidemic-Type Aftershock Sequence (ETAS) model of earthquake occurrence [Oga88a, Oga98] is a self-exciting point-process model where the conditional intensity $\lambda(t, x, y|H_t)$ of events is determined by a stationary Poisson intensity generating spontaneous earthquake events along with a dynamic term representing a branching process of aftershocks:

$$\lambda(t, x, y|H_t) = \mu(x, y) + \sum_{i:t_i<t} g(t - t_i, x - x_i, y - y_i, m_i). \tag{1}$$

Here (x, y) is the epicenter of an earthquake event described by longitude and latitude in decimal degrees of the WGS 84 coordinate system, m is its magnitude on the Richter scale computed using a body-wave magnitude

Keywords and phrases: Infinite Gaussian Mixture Model, Epidemic-Type Aftershock Sequence, Point Process

formula [SSC89], $H_t = \{(t_i, x_i, y_i, m_i) : t_i < t\}$ is the history of all earthquake events up to time t in a catalog, and $\mu(x, y)$ is the background intensity reflecting spatial heterogeneity of spontaneous earthquakes and the fact that earthquake catalogs with aftershocks removed are approximately Poisson in time [GK74].

The space-time-magnitude distribution of parent-offspring events in the branching process given by the function $g(t, x, y, m)$ is called the triggering kernel, typically following Omori's law [Uts61],

(2)
$$g(t - t_i, x - x_i, y - y_i, m_i) = \frac{K_0 e^{a(m_i - m_0)}}{(t - t_i + c)^{(1+w)}((x - x_i)^2 + (y - y_i)^2 + d)^{(1+\rho)}}$$

where m_0 is the cutoff magnitude of the considering dataset following [Oga88b], $K_0, a, c, \omega, d, \rho > 0$ are the parameters under estimation. Estimation of Equation 1 typically consists of constructing a non-parametric estimate for $\mu(x, y)$ along with finding maximum likelihood estimators for the parameters of the triggering kernel in Equation 2. Methods for maximizing the likelihood include quasi-Newton [Oga88a] and expectation-maximization (EM) [VS08], and the most common estimators for $\mu(x, y)$ are spatial histograms [ML08, VS08] or isotropic kernel density estimators [ZOVJ02, AC15].

1.2. *A New Model: Coupled ETAS-I^{2}GMM.* Earthquakes cluster at multiple scales, as earthquakes cluster locally through aftershock activity but also over larger scales along fault lines (see Figure 1). While there is research on the reconstruction of aftershock clusters from event data [ZOVJ02, ZGKBW08], existing point-process models of earthquake activity fail to capture clustering patterns at the larger scale of fault lines. In particular, histograms and kernel density estimators are able to capture spatial heterogeneity in the risk of spontaneous earthquakes, but the methods capture variation over only one scale. To our knowledge, our work here is the first to attempt to reconstruct the community fault model [PSB+07] with a statistical model based on earthquake event data.

In this paper we introduce a new type of ETAS model that can capture multiscale clustering in earthquake patterns. In particular, we propose using an infinite mixture of infinite Gaussian mixtures (I^{2}GMM) [YRD14] to estimate the background rate of earthquakes $\mu(x, y)$. The I^{2}GMM uses a different Dirichlet process mixture of Gaussians (DPMG) for each cluster that simultaneously predicts the number of clusters along with performing model inference. While I^{2}GMM has been introduced for high-dimensional clustering and ETAS is well known in seismology, what is brand new in this paper is the use of I^{2}GMM for modeling the intensity of a point process and

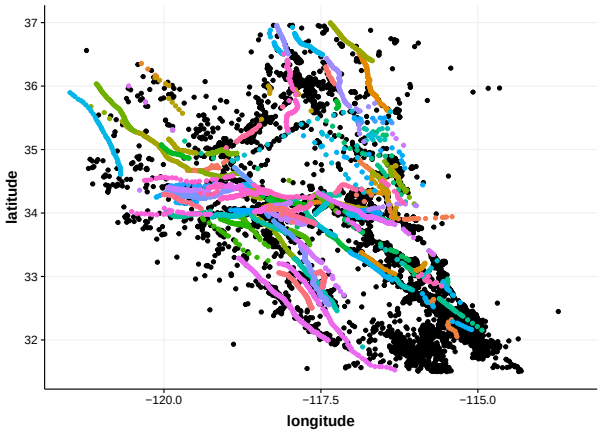

FIGURE 1. *Southern California earthquakes magnitude 2.5 and greater (black) and faults corresponding to the Community Fault Model 3.0 (marked by lines).*

the coupling of these two techniques for multiscale modeling of space-time event patterns. Through the use of an expectation-maximization algorithm, the benefit of our approach is that earthquakes are assigned membership to aftershock clusters in addition to a larger-scale fault-line cluster.

Another advantage of our approach is that multi-modal and skewed clusters are more accurately captured. In the case of spatial earthquake patterns, each fault may be considered as a separate cluster with multi-modality and skewness that the I^2GMM can handle better than histograms and KDE estimators. One additional advantage of the I^2GMM model is that earthquakes are assigned membership to clusters inferred by the model. In this research, we explore the relevance of cluster membership to automatic detection of fault lines within the ETAS-I^2GMM framework.

1.3. *Outline of the paper.* In Section 2 we describe our methodology, including an overview of the I^2GMM model and details on a Monte-Carlo EM algorithm for joint inference of the ETAS-I^2GMM model. In Section 3 we present results for several experiments where the ETAS-I^2GMM is applied to a Southern California earthquake catalog [sce]. We compared the goodness of fit of the estimated intensity of the model to a baseline approach. In Section 3, we also use the Community Fault Model 3.0 [PSB+07] to explore the ETAS-I^2GMM model's ability to detect fault locations and event-fault linkage from space-time event data.

2. Methods.

2.1. *Infinite mixture of infinite Gaussian mixtures.* The finite Gaussian mixture model (GMM) uses a single Gaussian for each cluster and requires the number of clusters to be specified. In the infinite version of GMM (IGMM) [Fer73], the number of components is estimated along with the component mean vectors and covariances. Both GMM and IGMM are used for clustering problems, albeit with limited success, as these techniques often overestimate the number of mixture components so as to more accurately estimate the density of the underlying dataset. However, a more accurate estimation of the density does not necessarily translate into a more accurate estimation of cluster distributions, as density estimation does not readily solve the problem of many-to-one mappings between components and clusters. Two different approaches are considered in the literature to overcome this limitation.

The first approach replaces Gaussian mixture components by student-t [PM00, SB05, AV07, AM12] or skewed-t [LM14] distributions and their Pearson-type extensions [FW14, SKG10] in an effort to better model cluster distributions with heavy tails. Although closed-form solutions for maximum-likelihood estimations of parameters do not in general exist under these settings, extensions of the EM algorithm can still be derived for this family of mixture models by placing certain restrictions on the original model. This line of models has proved quite effective in clustering datasets with skewed distributions but are less ideal for clusters with multi-mode distributions.

The second approach generates a large number of Gaussian components and merges them according to various metrics in an effort to recover true cluster distributions. The study in [FJ02] initializes the model with a large number of components and uses the concept of minimum message length to merge components. Another technique uses the Bayesian information criterion (BIC) to choose the initial number of components and merges components to minimize entropy [BRC+10]. Other options for assigning components to clusters include clustering modes of components [GS12] and ridgeline analysis and Bhattacharyya dissimilarity [Hen10]. Compared to mixtures of student-t or skewed-t distributions, this line of models is more flexible in terms of the type of distribution they can model. However, the main limitation of these techniques is the independence assumption made during component estimation that makes EM derivations possible. These techniques assume that all components are generated independently, which is not a very realistic assumption in a setting where some clusters are known to be multi-mode, and components originating from the same cluster are more likely to share certain latent parameters than two random components. Another limitation of these techniques is the computational complexity that increases with the square of the number of Gaussian components, as the

decision to merge two components requires evaluating the metric for every possible pair of components. This puts a constraint on the maximum number of components that can be used to model datasets.

When clustering datasets with skewed/multi-mode cluster distributions two dependent subproblems, namely density estimation and component clustering, need to be addressed jointly. The two-layer non-parametric GMM (I^2GMM) model, which can grow arbitrarily large in the number of components and clusters generated, was introduced earlier to more accurately cluster datasets with multi-mode and skewed cluster distributions [YRD14].

In I^2GMM the lower layer estimates the density of the overall dataset by clustering individual data points to components, while the upper layer associates components with clusters to allow for cluster recovery. More specifically, the generative model is a two-layer hierarchical Dirichlet process mixture (DPM) model where the lower layer uses one DPM for each cluster and the upper layer uses a global DPM for modeling cluster shapes and sizes. The dependency between the two layers is achieved by centering the base distributions of DPMs in the lower layer on a unique parameter distributed according to the global DPM. Inference, which involves sampling component indicator variables for individual data points and sampling cluster indicator variables for components, is performed by a collapsed Gibbs sampler, enabling optimization of two subproblems simultaneously.

We believe that I^2GMM has three unique features that would make it very suitable for the estimation of background intensity $\mu(x, y)$ in the ETAS model.

- As a two-layer non-parametric model, I^2GMM allows the number of clusters and the number of mixture components in each cluster to grow arbitrarily large, offering great flexibility in modeling clusters with multi-mode/skewed distributions. This is the main feature that distinguishes I^2GMM from other model-based clustering techniques that use one component for each cluster.
- As a Bayesian model, I^2GMM has hyper-parameters that can be tuned to recover clusters with varying shapes and different levels of rarity without facing singularities during model estimation. This distinguishes I^2GMM from purely data-driven techniques such as finite mixture of Gaussians and t-distributions that rely on EM and its extensions during model learning.
- As a hierarchical model, I^2GMM can share parameters not only across different clusters but also across different components of the same cluster. In other words, I^2GMM assumes that components are generated independently only when conditioned on the unique parameter defining

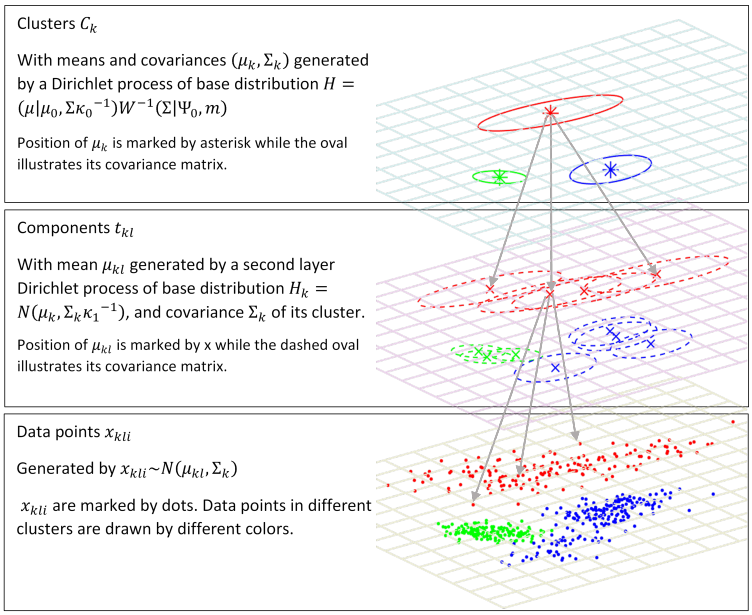

FIGURE 2. *The hierarchy of I^2GMM model illustrated on a synthetic dataset.*

their clusters of origin. This differentiates the proposed work from other techniques that estimate a large number of Gaussian components and merges them sequentially to recover clusters, thus violating component dependence.

In Figure 2 we provide an illustration of the generative model for I^{2}GMM: where t_{kl} indicates the l^{th} component in the k^{th} cluster C_k; x_{kl_i} indicates the ith data point in the l^{th} component in the k^{th} cluster. In the generative process, t_{kl} is a Gaussian distribution and C_k is a Gaussian mixture defined by its components. We will use the top-level label (i.e., C_k) to identify different clusters of spontaneous earthquakes.

The generative model for I^{2}GMM is given by

$$
\begin{aligned}
H &= NIW(\mu, \Sigma|\mu_0, \Psi_0, \kappa_0, m) = N(\mu|\mu_0, \Sigma\kappa_0^{-1})W^{-1}(\Sigma|\Psi_0, m) \\
G &\sim DP(\gamma H) \\
(\mu_k, \Sigma_k) &= \theta_k \sim G \\
H_k &= N(\mu_k, \kappa_1^{-1}\Sigma_k) \\
G_k &= DP(\alpha H_k) \\
\mu_{kl} &\sim G_k \\
x_{kl_i} &\sim N(\mu_{kl}, \Sigma_k)
\end{aligned}
\tag{3}
$$

In this generative model $DP(\gamma H)$ is a global Dirichlet process with normal-inverse-Wishart base distribution H and a concentration parameter γ. G is a discrete mixing measure sampled from the global DP. Center μ_k and covariance Σ_k of the cluster k are drawn from G. For each cluster generated by the global DPM, a local DPM is defined with base distribution H_k and concentration parameter α. All H_k are Gaussian distributions with centers μ_k and covariances $\kappa_1^{-1}\Sigma_k$. G_k is the cluster-specific discrete mixing measure drawn from the local DP. The components in cluster k are generated with mean vectors μ_{kl} drawn from G_k. Data points x_{kl_i} are generated from the Gaussian components with mean vectors μ_{kl} and covariance matrices Σ_k. User-specified hyper-parameters $(\mu_0, \Psi_0, \kappa_0, \kappa_1, \alpha, \gamma, m)$ are listed in Table 1 along with their descriptions.

To perform inference with the I^{2}GMM model using spatial event data, we first initialize the cluster and component indicators for each event to some arbitrary values (for example put all data in the same component of a cluster) and then use a collapsed Gibbs sampler to infer values for indicator variables one at a time, given all other indicator variables [YRD14]. Conditioned on the indicator variables, the location vectors and scale matrices are determined by maximizing the complete data log-likelihood and have closed form solutions. One sweep of the Gibbs sampler will go over all events in the dataset; convergence typically requires several hundred to thousand Gibbs sweeps.

2.2. *EM inference for ETAS.* The ETAS model given in Equation 1 can be viewed as a branching process where first-generation events occur according to a Poisson process with intensity $\mu(x, y)$. Events (from all generations) each give birth to direct offspring events determined by the triggering kernel $g(t - t_i, x - x_i, y - y_i, m_i)$.

Given an initial guess for the parameters of the triggering kernel in Equation 2 and the background rate $\mu(x, y)$, the branching structure along with

μ_0	Expected mean vector for each cluster. This is usually set to the mean of the overall dataset.
Ψ_0	$\frac{\Psi_0}{m-d-1}$ is the expected covariance matrix for clusters. This is usually set to identity.
κ_0	A positive scaling constant that adjusts the separation among clusters. The smaller the κ_0, the more separated the clusters will be from each other.
κ_1	A positive scaling constant that adjusts the separation among components of a given cluster. The smaller the κ_1, the more separated the components will be from each other, and clusters will tend to emerge with multi-modal distributions.
α	The concentration parameter for the local DPMs that controls the expected number of components and their sizes within a cluster.
γ	The concentration parameters for the global DPM that controls the expected number of clusters and their sizes.
m	Degree of freedom for the inverse Wishart that controls the degree of deviation of actual component covariances from the expected covariance. The higher the m, the less the deviation and the more similar component shapes will be.

TABLE 1
Hyper-parameters for I^2GMM

the model parameters of the triggering kernel can be estimated using an EM algorithm [VS08, MSB+11]. In the E-step of the EM algorithm the probability p_{ij} that event i is a direct offspring of event j is estimated, along with the probability p_i^b that the event was generated by the Poisson process μ.

$$p_{ij} = \frac{g(t_i - t_j, x_i - x_j, y_i - y_j)}{\lambda(t_i, x_i, y_i)}, \tag{4}$$

$$p_i^b = \frac{\mu(x_i, y_i)}{\lambda(t_i, x_i, y_i)}, \tag{5}$$

Given the probabilistic estimate of the branching structure, the complete data log-likelihood is then maximized in the M-step (using standard methods for estimating a Pareto distribution) [VS08], providing an estimate of the model parameters.

2.3. *Joint inference of the ETAS-I^2GMM .* We propose three variants for inferring the joint ETAS-I^{2}GMM model.

ETAS-I^{2}GMM 1. In the first variant we start by clustering all events spatially using I^{2}GMM. We then evaluate the convex hull of each cluster and enforce $\mu(x, y)$ to be constant for each cluster in the convex hull. With

the support of $\mu(x, y)$ fixed, we then proceed to estimate ETAS using the EM algorithm in Section 2.2.

ETAS-I^{2}GMM 2. In the next variant we perform joint inference using a Monte-Carlo EM algorithm. In particular, at each EM iteration we perform the following steps:

i. (I^{2}GMM-step) Sample background events from probabilistic branching structure p_{ij}. Estimate $\mu(x, y)$ based on the clusters generated by I^2GMM on the sampled background events only.

ii. (E-step) Estimate probabilistic branching structure and model parameters of triggering kernel as in Section 2.2.

ETAS-I^{2}GMM 3. In the last variant we use a weighted I^{2}GMM algorithm in place of the i step in variant two above. Instead of estimating $\mu(x, y)$ based on sampled background events, we estimate $\mu(x, y)$ using weighted I^{2}GMM on all events whose weights are estimated by (5). In the first EM iteration, since p_i^b does not exist we initialized the weight to 1.

2.4. *Baseline ETAS model with histogram estimator.* We use the histogram estimator proposed in [VS08] as a baseline model for comparison. In particular, we let the background rate $\mu(x, y)$ be a constant

$$\mu(x, y) = \mu_k \text{ if (x,y) is in cell } k, k \in 1, \ldots, K \tag{6}$$

over each rectangular cell of a regular grid. There are then $K + 6$ parameters $\theta = (\mu_1, \ldots, \mu_K, a, c, d, w, \rho, K_0)$ that we need to estimate (assuming there are K cells in the grid) and for that purpose we use the EM algorithm in [VS08].

3. Experiments and Results.

3.1. *Experiment 1: goodness of fit of ETAS-I^{2}GMM applied to CA earthquakes 3.5 and greater since 2000.* We apply our models to the California earthquake event data filtered by year (greater than 2000) and magnitude (greater than 3.5). The geographic bounds range from $46.116 >$ latitude > 29.615 and $-113.581 >$ longitude > -130.427. The dataset is divided into training and testing using time point 2010-01-01 00:00:00 for cutoff. All events before this timestamp are placed in the training dataset while all events after it are placed in the testing dataset. We performed experiments

with the following six models to analyze how the performance varies by adopting different modeling strategies:

1. ETAS-I^{2}GMM 1.
2. ETAS-I^{2}GMM 2.
3. ETAS-I^{2}GMM 3.
4. 4×4 grid baseline model.
5. 3×4 grid baseline model.
6. 3×3 grid baseline model.

Note that experiments 1 to 3 are repeated 10 times and the means of the likelihood are recorded. For I^{2}GMM we run 400 Gibbs sweeps; the hyper-parameters are set as follows: $\mu_0 = [36.4603 - 119.3265]$ the mean of the data set; $\Sigma_0 = [21.4972 \quad 0; 0 \quad 23.1351)]$ the diagonal of the covariance of the data set; $m = 22$; $\kappa_0 = 0.1$; $\kappa_1 = 0.5$. The values of m, κ_0, κ_1 are tuned to let I^{2}GMM generate equivalent number of clusters as the 4×4 grid baseline model.

We use the log-likelihood function

$$\text{logL} = \sum_{i=1}^{N} \log(\lambda(t_i, x_i, y_i)) - \int_0^T \int_S \lambda(t, x, y) dx dy dt \tag{7}$$

to evaluate the competing models for the background intensity. The results are shown in Table 2.

Model	logL	$\sum_i \log(\lambda_i)$	$\int \lambda$
ETAS-I^{2}GMM 1	**-4619**	-1206	3413
ETAS-I^{2}GMM 2	-4686	-1283	3403
ETAS-I^{2}GMM 3	-4716	-1319	3397
4×4 Grid	-4980	-1590	3390
3×4 Grid	-4937	-1607	3330
3×3 Grid	-5023	-1582	3440

TABLE 2
Log-likelihood model comparison.

In Table 2 we find that all three ETAS-I^{2}GMM models outperform ETAS with a histogram estimator. Between the three ETAS-I^{2}GMM variants, the best-performing model is variant 1, where I^{2}GMM is first estimated and the EM algorithm is run separately to estimate the triggering kernel parameters. It is worthwhile to note that finer clusters do not necessarily yield better results. Even though the 4×4 grid model generates more clusters it produces lower likelihood than the 3×4 grid model.

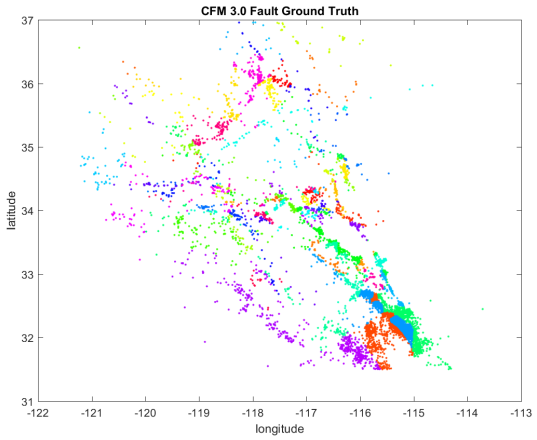

FIGURE 3. *Cluster membership using the nearest CFM 3.0 fault to each earthquake (ground truth).*

3.2. *Experiment 2: ETAS-I^{2}GMM for event-fault linkage from space-time event data.* Next, we investigate the extent to which the ETAS-I^{2}GMM model can learn fault structure from space-time event data. For this purpose we use the Community Fault Model 3.0 [PSB+07], which is a three dimensional representation (latitude, longitude, and elevation) of faults in Southern California. The CFM is a collaborative project undertaken by scientists of the Southern California Earthquake Center (SCEC) for studying active faults and earthquake phenomena and to improve regional earthquake hazard assessments. Our goal here is to assess how well ETAS-I^{2}GMM recovers a 2D projection of the CFM 3.0 using only space-time-magnitude earthquake incident data as input. In particular, we generate a fault label for each event in the dataset by assigning fault membership as the nearest fault in CFM 3.0 (see Figure 3).

The ETAS-I^{2}GMM-predicted label is taken from the first layer of the I^{2}GMM model clusters, where offspring events are assigned to the cluster of their nearest neighbor among background events. To allow for comparison to the CFM 3.0, we restrict the geographic bounds of the CA earthquake event data to $36.958 > \text{latitude} > 31.518$ and $-113.719 > \text{longitude} > -121.176$, but we expand the magnitude threshold down to 2.5.

Since there are 145 actual fault lines in CFM 3.0 but all the six models we used in previous experiments generate at most 26 clusters, we added two additional models in the experiments for fault recovery:

- A I^{2}GMM with parameters tuned to generate approximately 145 clus-

ters on average, and name this version as ETAS-I^{2}GMM 145 in our experiments.

- A 16×15 grid model that contains 143 non-empty clusters.

	Accuracy	$\overline{Acc}_{10}$
ETAS-I2GMM 145	0.46	0.52
ETAS-I2GMM 1	**0.50**	**0.67**
ETAS-I2GMM 2	0.45	0.46
ETAS-I2GMM 3	0.41	0.33
Grid 16x15	0.45	0.47
Grid 4x4	0.37	0.37
Grid 3x4	0.36	0.28
Grid 3x3	0.35	0.32

TABLE 3
Accuracy comparison of fault classification.

Given that the number of clusters estimated by I^{2}GMM may be different from the number of faults in the CFM 3.0, we evaluate the success of fault-cluster recovery by considering the percentage of correctly classified data points. In addition to the overall accuracy, we evaluate the mean accuracy for the 10 largest faults, which contains 67% of data points across 145 faults. In Table 3 we present the accuracy for the six models listed in 3.1 as well as the two additional models. To calculate the accuracy, we first align the generated clusters with the ground-truth classes using the Hungarian algorithm [Kuh55, Ste00], and then calculate the percentage of the data points that fall into their classes of origin. We adopt accuracy for its simplicity and its invariance to potential switching between ground truth and predicted class labels. Mean accuracy is calculated as below:

$$\overline{Acc} = \frac{1}{|C^*|} \sum_{C_k^* \in C^*} \frac{|C_k^* \cap C_k|}{|C_k^*|} \tag{8}$$

where C^* contains all fault clusters under consideration; C_k is the predicted cluster corresponding to C_k^* after alignment; $C_k^* \cap C_k$ indicates data points in both C_k^* and C_k; $|S|$ denote the cardinality of the set S. To compute the mean accuracy for the ten largest faults $\overline{Acc}_{10}$ we set C^* to contain the ten largest true fault clusters in the above equation.

Here again we see that ETAS-I^{2}GMM 1 performs best both in terms of accuracy and $\overline{Acc}_{10}$. In Figure 4 we plot the clusters recovered corresponding to ETAS-I^{2}GMM 145, ETAS-I^{2}GMM 1, 16×15 grid, 4×4 grid, and the true clusters for a better understanding of this outcome. Despite the fact that

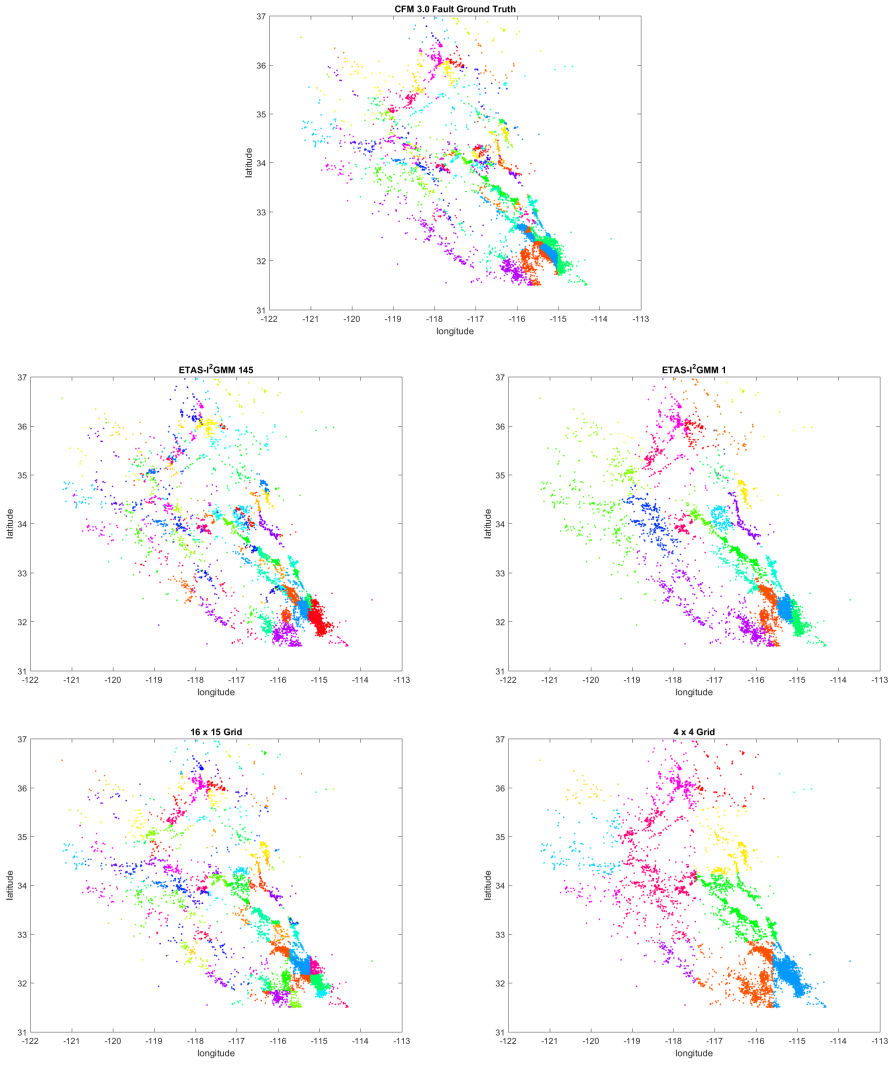

FIGURE 4. *True and predicted CFM fault groupings. Events with same labels are shown by the same color.*

I^2GMM generated only 26 unique clusters on average compared to 145 actual fault lines in CFM 3.0, a meaningful accuracy of 0.5 was achieved. Results suggest that a majority of events in fault clusters that tend to have elongated, skewed, and in some cases multi-mode shapes are clustered correctly by I^2GMM. In contrast, the clusters in the two histogram models have abrupt boundaries formed from the grid irrespective of the shape of the underlying faults, as shown in Figure 4. Moreover, when we adjust the parameters of the I^2GMM to get approximately the same number of clusters as the true number of fault clusters, we observe that the accuracy does not improve owing to erroneous splitting of events belonging to larger fault lines into multiple clusters. This is also true for the 16×15 grid model that generates 143 non-empty clusters. Although the accuracy improves with this model compared to the grid model with a smaller number of clusters, overall accuracy achieved by this model is still less than that achieved by I^2GMM with 26 clusters (0.45 vs 0.50). The difference in accuracy between the two models increases in favor of I^2GMM when we take into account only the largest ten fault clusters (0.47 vs 0.67). This is a natural result of the grid model arbitrarily splitting fault clusters compared to the more effective handling of elongated fault cluster shapes by I^2GMM.

We illustrate this over-splitting problem in Figure 5 by plotting the clustering results of the ten largest faults. From the $\overline{Acc}_{10}$ results in Table 3 and Figure 5 we can see that ETAS-I^2GMM 1 did the best by achieving a mean accuracy of 0.67 across ten faults while recovering several of them by an accuracy of over 0.9. On the other hand, for the grid models the $\overline{Acc}_{10}$ values are consistent with their corresponding overall accuracies.

4. Discussion. We introduced a coupled ETAS-I^2GMM model for jointly estimating multi-scale clustering in earthquake data with parameters governing earthquake productivity and self-excitation. We also introduced what we believe is a novel machine-learning task for statistical seismology, namely estimating CFM fault clusters using unlabeled space-time-magnitude event data. Improving upon algorithms aimed at solving this task could aid in the development of future versions of CFM, as well as fault models in other regions of the world.

We also believe that the I^2GMM model may have applications to point processes beyond those arising in seismology. Space-time self-exciting point processes arise in the study of crime [MSB+11, Moh14, MSM+15], conflict [LM11], and terrorism [PW+12, M+13, WPM13, WP14], as well as in social-network event dynamics, for example in social media [LMY+14, SJ12, ZEH+15]. In the case of crime, clusters arise naturally from the su-

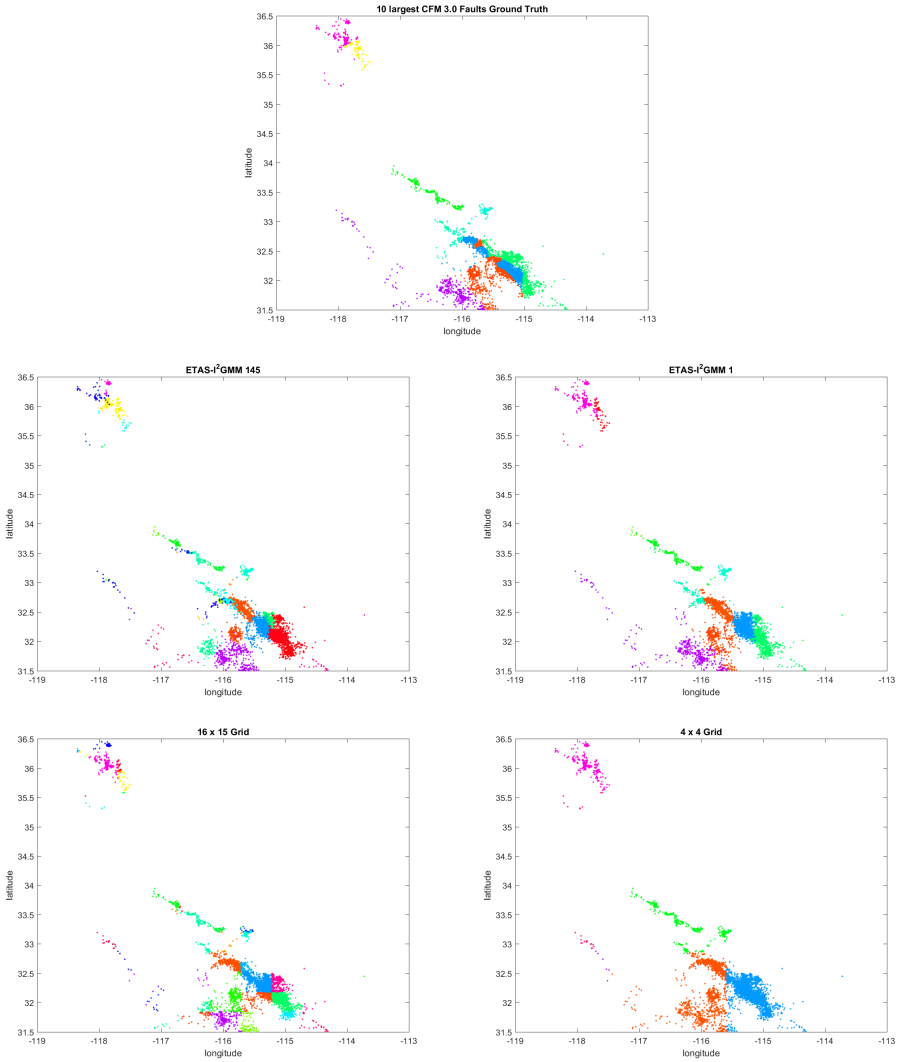

FIGURE 5. *Ten largest CFM faults and recovered clusters. Events with same labels are shown by the same color.*

perposition of events committed by different offenders with different modus operandi. Similar clusters may arise from the operations of different terrorist groups within a geographic region. I^2GMM is a flexible model for capturing this type of clustering in the intensity of events of a point process.

Acknowledgment. This research was sponsored by the National Science Foundation (NSF) under Grant Numbers IIS-1252648 (CAREER) and SES-1343123. The content is solely the responsibility of the authors and does not necessarily represent the official view of NSF.

REFERENCES

[AC15] Giada Adelfio and Marcello Chiodi. Alternated estimation in semi-parametric space-time branching-type point processes with application to seismic catalogs. *Stochastic Environmental Research and Risk Assessment*, 29(2):443–450, 2015.

[AM12] Jeffrey L Andrews and Paul D Mcnicholas. Model-based clustering, classification, and discriminant analysis via mixtures of multivariate t-distributions. *Statistics and Computing*, 22(5):1021–1029, 2012.

[AV07] Cédric Archambeau and Michel Verleysen. Robust bayesian clustering. *Neural Networks*, 20(1):129–138, 2007.

[BRC+10] Jean-Patrick Baudry, Adrian E Raftery, Gilles Celeux, Kenneth Lo, and Raphael Gottardo. Combining mixture components for clustering. *Journal of Computational and Graphical Statistics*, 19(2), 2010.

[Fer73] Thomas S Ferguson. A bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230, 1973.

[FJ02] Mario AT Figueiredo and Anil K Jain. Unsupervised learning of finite mixture models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(3):381–396, 2002.

[FW14] Florence Forbes and Darren Wraith. A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweight: application to robust clustering. *Statistics and Computing*, 24(6):971–984, 2014.

[GK74] JK Gardner and Leon Knopoff. Is the sequence of earthquakes in southern california, with aftershocks removed, poissonian? *Bulletin of the Seismological Society of America*, 64(5):1363–1367, 1974.

[GS12] Yongchao Ge and Stuart C. Sealfon. Flowpeaks: A fast unsupervised clustering for flow cytometry data via k-means and density peak finding. *Bioinformatics*, 28(15):2052–2058, aug 2012.

[Hen10] Christian Hennig. Methods for merging gaussian mixture components. *Advances in data analysis and classification*, 4(1):3–34, 2010.

[Kuh55] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[LM11] Erik Lewis and George Mohler. A nonparametric em algorithm for multiscale hawkes processes. *preprint*, pages 1–16, 2011.

[LM14] Sharon Lee and Geoffrey J McLachlan. Finite mixtures of multivariate skew t-distributions: some recent and new results. *Statistics and Computing*, 24(2):181–202, 2014.

[LMY+14] Eric Lai, Daniel Moyer, Baichuan Yuan, Eric Fox, Blake Hunter, Andrea L Bertozzi, and Jeffrey Brantingham. Topic time series analysis of microblogs. Technical report, DTIC Document, 2014.

[M+13] George Mohler et al. Modeling and estimation of multi-source clustering in crime and security data. *The Annals of Applied Statistics*, 7(3):1525–1539, 2013.

[ML08] David Marsan and Olivier Lengline. Extending earthquakes' reach through cascading. *Science*, 319(5866):1076–1079, 2008.

[Moh14] George Mohler. Marked point process hotspot maps for homicide and gun crime prediction in chicago. *International Journal of Forecasting*, 30(3):491–497, 2014.

[MSB+11] George O Mohler, Martin B Short, P Jeffrey Brantingham, Frederic Paik Schoenberg, and George E Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108, 2011.

[MSM+15] George O Mohler, Martin B Short, Sean Malinowski, Mark Johnson, George E Tita, Andrea L Bertozzi, and P Jeffrey Brantingham. Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*, 110(512):1399–1411, 2015.

[Oga88a] Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401):9–27, 1988.

[Oga88b] Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.

[Oga98] Yosihiko Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998.

[PM00] David Peel and Geoffrey J McLachlan. Robust mixture modelling using the t distribution. *Statistics and computing*, 10(4):339–348, 2000.

[PSB+07] Andreas Plesch, John H Shaw, Christine Benson, William A Bryant, Sara Carena, Michele Cooke, James Dolan, Gary Fuis, Eldon Gath, Lisa Grant, et al. Community fault model (cfm) for southern california. *Bulletin of the Seismological Society of America*, 97(6):1793–1802, 2007.

[PW+12] Michael D Porter, Gentry White, et al. Self-exciting hurdle models for terrorist activity. *The Annals of Applied Statistics*, 6(1):106–124, 2012.

[SB05] Markus Svensén and Christopher M Bishop. Robust bayesian mixture modelling. *Neurocomputing*, 64:235–252, 2005.

[sce] https://service.scedc.caltech.edu/eq-catalogs/.

[SJ12] Aleksandr Simma and Michael I Jordan. Modeling events with cascades of poisson processes. *arXiv preprint arXiv:1203.3516*, 2012.

[SKG10] Jianyong Sun, Ata Kaban, and Jonathan M Garibaldi. Robust mixture modeling using the pearson type vii distribution. In *Neural Networks (IJCNN), The 2010 International Joint Conference on*, pages 1–7. IEEE, 2010.

[SSC89] William Spence, Stuart A Sipkin, and George L Choy. Measuring the size of an earthquake. *Earthquake Information Bulletin (USGS)*, 21(1):58–63, 1989.

[Ste00] Matthew Stephens. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):795–809, 2000.

[Uts61] T Utsu. A statistical study on the occurrence of aftershocks. *Geophysical Magazine*, 30(4), 1961.

[VS08] Alejandro Veen and Frederic P Schoenberg. Estimation of space–time branching process models in seismology using an em–type algorithm. *Journal of the American Statistical Association*, 103(482):614–624, 2008.

[WP14] G. White and M.D. Porter. GPU accelerated MCMC for modeling terrorist activity. *Computational Statistics & Data Analysis*, 71:643–651, 2014.

[WPM13] G. White, M.D. Porter, and L. Mazerolle. Terrorism risk, resilience, and volatility: A comparison of terrorism in three Southeast Asian countries. *Journal of Quantitative Criminology*, 29(2):295–320, 2013.

[YRD14] Halid Z Yerebakan, Bartek Rajwa, and Murat Dundar. The infinite mixture of infinite gaussian mixtures. In *Advances in neural information processing systems*, pages 28–36, 2014.

[ZEH+15] Qingyuan Zhao, Murat A Erdogdu, Hera Y He, Anand Rajaraman, and Jure Leskovec. Seismic: A self-exciting point process model for predicting tweet popularity. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1513–1522. ACM, 2015.

[ZGKBW08] Ilya Zaliapin, Andrei Gabrielov, Vladimir Keilis-Borok, and Henry Wong. Clustering analysis of seismicity and aftershock identification. *Physical Review Letters*, 101(1):018501, 2008.

[ZOVJ02] Jiancang Zhuang, Yosihiko Ogata, and David Vere-Jones. Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97(458):369–380, 2002.

Indiana University Purdue University Indianapolis
E-mail: chengyic@umail.iu.edu