The Annals of Applied Statistics
2016, Vol. 10, No. 2, 667–689
DOI: 10.1214/15-AOAS851

NONPARAMETRIC MULTI-LEVEL CLUSTERING OF HUMAN EPILEPSY SEIZURES[1]

BY DRAUSIN F. WULSIN, SHANE T. JENSEN AND BRIAN LITT

University of Pennsylvania

Understanding neuronal activity in the human brain is an extremely difficult problem both in terms of measurement and statistical modeling. We address a particular research question in this area: the analysis of human intracranial electroencephalogram (iEEG) recordings of epileptic seizures from a collection of patients. In these data, each seizure of each patient is defined by the activities of many individual recording channels. The modeling of epileptic seizures is challenging due the large amount of heterogeneity in iEEG signal between channels within a particular seizure, between seizures within an individual, and across individuals. We develop a new nonparametric hierarchical Bayesian model that simultaneously addresses these multiple levels of heterogeneity in our epilepsy data. Our approach, which we call a multi-level clustering hierarchical Dirichlet process (MLC-HDP), clusters over channel activities within a seizure, over seizures of a patient and over patients. We demonstrate the advantages of our methodology over alternative approaches in human EEG seizure data and show that its seizure clustering is close to manual clustering by a physician expert. We also address important clinical questions like "to which seizures of other patients is this seizure similar?"

1. Introduction. Over 50 million people worldwide suffer from epilepsy [Saraceno, Avanzini and Lee (2005)], and 20–40% of those cases are not effectively treated by medications [French (2007)]. Resective brain surgery is currently the only alternative to pharmacological treatment, but determining what area(s) of brain tissue to remove is a difficult task. To this end, patients usually undergo continuous intracranial electroencephalogram (iEEG) monitoring for up to several weeks in a hospital epilepsy monitoring unit after implantation of up to several hundred intracranial electrodes. The goal of monitoring is to record seizures so that neurologists specializing in epilepsy can observe the seizure dynamics and use them in their clinical decision-making.

The iEEG data come from individual implanted electrodes (also known as channels) penetrating into the brain or under the dura on the brain's surface to record

Received February 2013; revised January 2015.

[1]Supported by grants from the US National Institutes of Health (National Institute of Neurological Disorders and Stroke RO1-NS041811, RO1-NS48598 and U24-NS063930-03), the Julie's Hope Award from the Citizens for Research in Epilepsy, and the Dr. Michel and Mrs. Anna Mirowski Discovery Fund for Epilepsy Research.

Key words and phrases. Epilepsy, seizures, intracranial electroencephalogram (iEEG), Dirichlet process, nonparametric Bayes, clustering.

electrophysiologic potentials relative to a reference channel. These recordings are sampled at a frequency of at least 200 Hz and collected continually for a period of up to a few weeks. A patient often has more than one hundred individual electrodes implanted at once. These electrodes are uniquely placed for each patient. The number of seizures recorded during continuous monitoring can vary from only one to over fifty per patient.

The clinical decision process for epilepsy surgery is still quite inexact due to the heterogeneity of the disease, the many modalities of patient data (including clinical history, MRIs, CTs, scalp EEG and intracranial EEG), and the numerous physicians and subspecialties involved. These challenges most likely contribute to the low seizure-freedom rates following resective surgery, particularly for seizures arising outside of the temporal lobes [de Tisi et al. (2011)].

In reviewing a patient's iEEG, an epileptologist must determine which areas of the brain should be removed to reduce and hopefully eliminate a patient's future seizures [Engel and Pedley (2008), Chapters 166–169]. Many factors go into this decision. Epileptologists pay particular attention to which channels seem involved with the onset of the seizure and the way in which other channels later join the seizure activity. Simplistically, the epileptologists manually cluster the channels into different groups that include onset channels, delayed onset channels and non-involved channels, to name but a few different types. On a higher level, epileptologists look at the collection of seizures a patient displays to understand the different seizure types a patient tends to display, including seizures with focal onset (a few very specific channels initiate the seizure), diffuse onset (many or all of the channels involved from the start) and many gradations in between. Understanding the full range of seizure types a patient can display is important in determining how removing particular brain tissue may affect a patient's prognosis.

Finally, from a very high level, the epileptologist considers how likely the patient is to benefit from the extremely invasive resective surgery given the outcomes of previous patients with a similar pattern of seizures. For example, patients with exclusively very focal onset seizures tend to have much higher seizure freedom rates than those with more of a variety of seizure etiologies [Engel and Pedley (2008), Chapter 167].

In current clinical practice, this analysis process is almost entirely manual and varies greatly depending on the center of treatment and individual training of the epileptologists. We believe statistical models can help reduce some of the uncertainty in this process and provide objective clinical decision support. On some level, one can think of the clinical analysis process described above as a clustering procedure done on a number of levels: the channel level, the seizure level and the patient level.

Most existing quantitative approaches to seizure modeling focus on individual seizures since generalizing across seizures and especially across patients is so difficult. Many models aim to understand relationships between the channel activities [cf., Bartolomei et al. (2010), Chaovalitwongse (2008), Hegde et al. (2007),

Ossadtchi et al. (2010), Paramanathan and Uthayakumar (2008)] in a seizure. Other models focus on the problem of seizure onset detection [cf., Chan et al. (2008), Klatchko et al. (1998), Quyen et al. (2005)], a well-studied but still very difficult problem in the epilepsy domain.

Models for a single seizure from one patient are not at all analogous to a physician's representation of the same seizure. That physician has—over the course of his or her training—seen the iEEGs of hundreds or thousands of seizures from many patients. This experience informs the physician's interpretation of the current seizure of interest. We believe that any model that hopes to reasonably represent the iEEG of seizures must also integrate information from many seizures over a diverse patient population. A model that clusters a multi-patient dataset of iEEGs from seizures on the channel level, the seizure level and the patient level would achieve our desired information sharing across seizures and patients.

Hierarchical Bayesian approaches provide a ready solution for the information-sharing problem in that each level in the hierarchy is a blend of local information (e.g., the channel activities in a particular seizure) and also global information (e.g., the other seizures of that patient and even the other seizures of other patients). Nonparametric Bayesian methods are also attractive since they reduce the amount of necessary model selection. Nonparametric Bayesian methods often build off the Dirichlet process (DP) [Ferguson (1973)], a discrete probability distribution over distributions. In particular, the hierarchical Dirichlet process (HDP) of Teh et al. (2006) and the nested Dirichlet process (NDP) of Rodríguez, Dunson and Gelfand (2008) provide ways of sharing information across a hierarchy and clustering on multiple levels of a dataset, respectively, and thus help inspire the model developed in this work. All of these approaches are agnostic to the observational model used at the bottom level, allowing the practitioner to choose the observational model most suitable for the specific application.

In this work, we develop a new hierarchical Bayesian model inspired by the problem of sharing information between multiple, diverse patients to effectively model seizures and the iEEG channel activities that comprise those seizures. Our model builds off nonparametric models like the hierarchical and nested Dirichlet processes, in some sense combining the efficient information sharing across data groups of HDP with the NDP's ability to perform multiple levels of clustering. Since the model uses an HDP for this clustering on each level (e.g., the iEEG channels, the seizure and the patient level) of the data, we call it the multi-level clustering hierarchical Dirichlet process (MLC-HDP).

In Section 2, we outline our dataset of intracranial electroencephalogram (iEEG) recordings from human epileptic seizures and the preprocessing that we apply to this data. We develop our multi-level clustering hierarchical Dirichlet process (MLC-HDP) model and its implementation in Section 3. The results from our dataset of intracranial electroencephalogram (iEEG) recordings are presented in Section 4 along with comparisons of our inferred seizure clusters to manual clustering by physician experts. We also examine our model inferences in the context

of several relevant clinical questions in Section 4 before concluding with a brief discussion in Section 5.

2. Data and preliminary processing. Our available data is 193 intracranial EEG (iEEG) seizure records from 10 patients from the Children's Hospital of Philadelphia. These patients display attributes common to most epileptic seizure iEEG recordings: (1) unique electrode placement for each patient, (2) large differences in the number of seizures per patient, and (3) differences in the number of usable channels across the seizures for each patient. Table 1 describes the number of seizures per patient as well as whether a patient's seizures contain the same number of active electrodes.

We chose to work with iEEG clips of all channels -30 seconds before to $+90$ seconds after the start of the seizure, with start of the seizure being manually defined by an epileptologist. An example of the raw iEEG recording with 128 channels is given in the top left plot of Figure 1. Rather than modeling these raw iEEG channel voltage traces directly, we chose to extract a set of simple features from each channel. Specifically, we extracted the $\log_{10}$ power in four clinically relevant frequency bands (4–8, 8–13, 13–30, 30–100 Hz) for each channel over the 120 seconds, using a sliding window of 500 ms with 50% overlap.

While rich literature exists on EEG features [cf., Adeli, Zhou and Dadmehr (2003), Ghosh-Dastidar, Adeli and Dadmehr (2008), Reijneveld et al. (2007), Srinivasan, Eswaran and Sriraam (2007), Stam (2005)], we chose these features because they closely resemble what we believe actual epileptologists consider when reading iEEG recordings. As an example, the middle right plot of Figure 1 gives one of these frequency features (13–30 Hz) for the 128 channels of the same seizure shown in the top left plot of Figure 1.

TABLE 1
The number of recorded seizures for each patient and whether all the seizures for that patient contained the same number of active channels

Patient	# Seizures	Same # channels
A	1	yes
B	9	yes
C	4	yes
D	18	no
E	61	no
F	50	no
G	1	yes
H	22	yes
I	13	no
J	14	yes

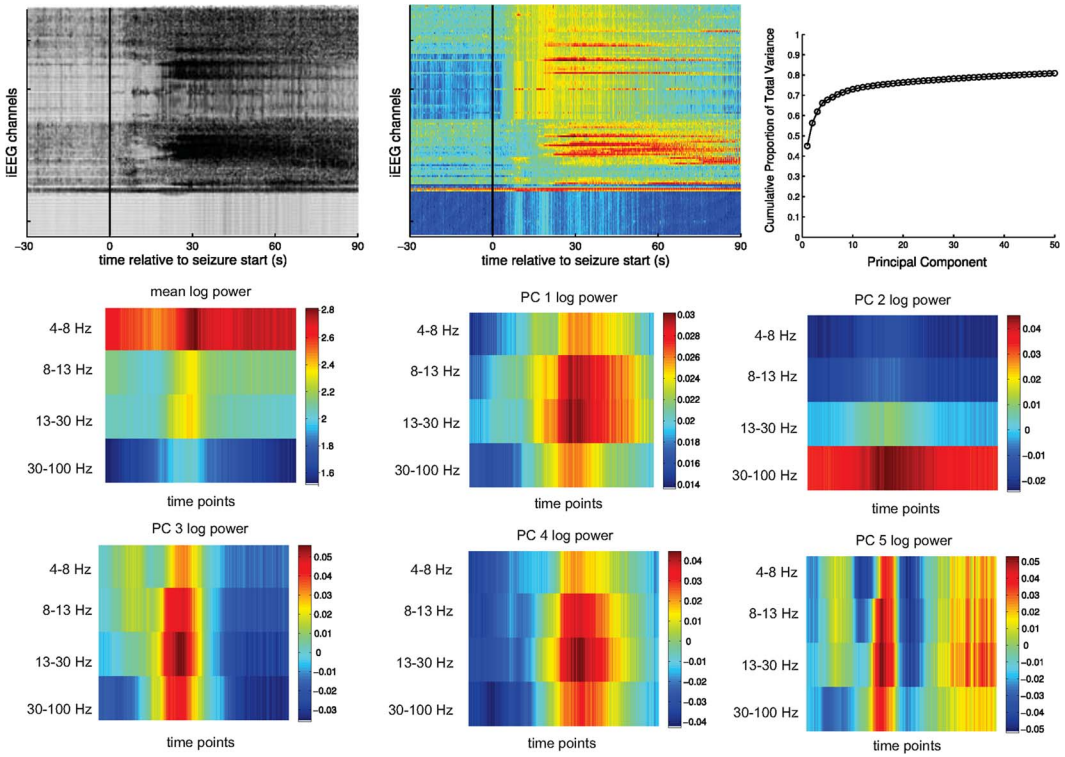

FIG. 1. Top left: *An example of* 128 *iEEG voltage traces over time for a seizure, where the start of the seizure is indicated with a vertical black line.* Top middle: *The* $\log_{10}$ *power in the* 13–30 *Hz frequency band for each channel of that same seizure, where red corresponds to larger values and blue to smaller values.* Top right: *A scree plot showing the cumulative proportion of variance of the first* 50 *principal components of the channel activities.* Middle and bottom rows: *Two-dimensional* (*each of four frequency bands at each time point*) *representations of the mean and first five principal components of the channel features.*

For each channel, the four frequency features at each of the 479 time points were concatenated into a 1916-dimensional feature vector representing that channel's activity during the seizure. Principal Components Analysis (PCA) was then applied to these feature vectors over all the seizures in our data in order to reduce the dimensionality of the channel features. The top right plot in Figure 1 gives the scree plot for the first 50 principal components of the channel features.

Based on this scree plot, we decided to focus our modeling on the first 5 principal components, which retained 68% of the original variance in the channel feature vectors.[2] In the middle and bottom rows of Figure 1, we represent these 5 principal components by plotting their value for the four frequency bands at each time point, along with the mean value for the four frequency bands at each time point. These two-dimensional representations show how different time points and frequency band powers are emphasized in different principal components.

Thus, our raw iEEG data has been reduced to a five-dimensional vector of principal component values that are the *channel observations* $\mathbf{x}_{tji}$ for channel i in seizure j of patient t. In Section 3, we develop a hierarchical Bayesian nonparametric model that shares information across channels within a seizure, across seizures within a patient and across patients.

3. Model and implementation. Given the complexity of epilepsy as a disease and the unique electrode placement for each patient, any model of seizure activity across patients will necessarily involve a number of simplifying assumptions to make it tractable. First and foremost, we assume that the channels of each seizure and the seizures of each patient are exchangeable. In reality, the spatial relationships between the channels and the temporal relationships between the seizures add considerable nuance to this already complex data. Nevertheless, we believe the most important information about a patient's seizures lies in the channel iEEG itself, so it should be the basis for our modeling approach.

As detailed in Section 2, our processed intracranial EEG (iEEG) seizure data consists of a five-dimensional vector of channel observations $\mathbf{x}_{tji} \in \mathbf{R}^5$, where $i = 1, \ldots, N_{tj}$ indexes each channel within a particular seizure, $j = 1, \ldots, J_t$ indexes each seizure with a particular patient, and $t = 1, \ldots, T$ indexes each patient. Specifically, we have $T = 13$ patients, each with a different number of seizures J_t as outlined in Table 1.

We use a multivariate normal likelihood with diagonal covariance to model these channel features,

$$\mathbf{x}_{tji} \sim \mathcal{N}(\boldsymbol{\mu}_{tji}, \boldsymbol{\sigma}^2_{tji}), \tag{1}$$

[2]We found that our results presented in Section 4 are reasonably similar when focusing our modeling on either 5, 10 or 20 principal components, which is not surprising given the small marginal increase in cumulative variance going from 5 to 20 principal components.

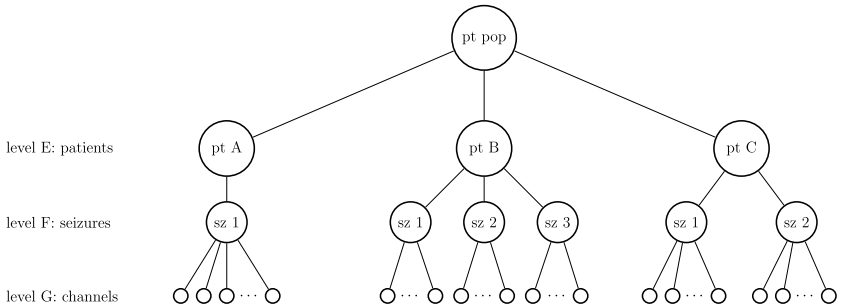

FIG. 2. *A schematic of the structure of our epilepsy iEEG data, with a set of patients that each contain a set of seizures that each contain a set of channel observations. The clusters at each level are shared horizontally across the hierarchy.*

where we collect the parameters of the mean vector $\boldsymbol{\mu}_{tji}$ and diagonal covariance matrix $\boldsymbol{\sigma}^2_{tji}$ into a single parameter vector $\boldsymbol{\phi}_k = (\boldsymbol{\mu}_{tji}, \boldsymbol{\sigma}^2_{tji})$. For more details about the likelihood, refer to our supplementary materials [Wulsin, Jensen and Litt (2016)].

We need to model the parameters $\boldsymbol{\phi}_{tji}$ of each channel observation $\mathbf{x}_{tji}$ so that we share information across channels within a seizure and across seizures within a patient. In Section 3.1, we outline our model for sharing information across seizures within a single patient. In Section 3.2, we extend our model to also share information across multiple patients. In Figure 2, we present an illustrative schematic of the three levels of information sharing in our epilepsy iEEG data: the channel observation level, the seizure level and the patient level.

Although every patient's seizures are unique, enough similarities exist on the channel and seizure level that we would like to share information between the models for each patient. However, the clinical aspects of a particular patient's seizures on the iEEG are much more similar to some patients than others, so we would like some way to more *selectively* share information across patients. For example, some patients display diffuse onset seizures, where most of the channels become hyperactive almost simultaneously. In other patients, activity on a small core of channels initiates the seizure and is joined later into the seizure by higher activity in other channels. In still other patients, the seizure is well localized only to a specific brain region, so channels located far from that region have barely any heightened activity during the seizure. We would ideally like to share information mostly between seizures within patients and between patients with a preference for finding common onset patterns.

Our approach to this desiderata is the development of a Bayesian hierarchical model that shares information at the multiple levels of channels within seizures, seizures within patients and across patients. As much as possible, we will avoid strong assumptions about the functional form of the distributions that underly

our parameters $\boldsymbol{\phi}_k$ but instead build off of recent advancements in nonparametric Bayesian modeling [e.g., Rodríguez, Dunson and Gelfand (2008), Teh et al. (2006)].

3.1. *Sharing information across seizures within a single patient.* We begin our development by first considering only a single patient t. We model the underlying parameters $\boldsymbol{\phi}_{tji}$ of each channel observation $\mathbf{x}_{tji}$ as sharing a common distribution

$$\boldsymbol{\phi}_{tji} \sim G$$

across channels i within a seizure and across seizures j within that patient t. The conventional nonparametric Bayesian approach for this common distribution G would be the Dirichlet process [Ferguson (1973)], where G is given a Dirichlet process prior, $G \sim \mathrm{DP}(\gamma, H)$, with base measure H for the parameters $\boldsymbol{\phi}_{tji}$ and scalar concentration parameter γ.

The stick-breaking construction of Sethuraman (1994) defines a realization $G \sim \mathrm{DP}(\gamma, H)$ from a Dirichlet process as

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\boldsymbol{\phi}_k}, \tag{2}$$

where $\delta_{\boldsymbol{\phi}_k}$ represents a unit measure concentrated at each sample $\boldsymbol{\phi}_k$ from the base measure $\boldsymbol{\phi}_k \sim H$. In our iEEG application, we use a normal inverse-χ^2 base measure for the parameters $\boldsymbol{\phi}_k$ that has a mean and diagonal covariance set using the method of moments to be equal to those across all the channel feature observations of all seizures and patients. This empirical prior setting provides a reasonable prior location while ensuring that it is sufficiently vague thanks to the large diversity in channel observations.

The $\boldsymbol{\pi}$ are weights for each sample $\boldsymbol{\phi}_k$ derived from the probabilistic breaking of a stick,

$$\pi_k = b_k \prod_{l=1}^{k-1} (1 - b_l) \qquad \text{where } b_k \sim \mathrm{Beta}(1, \gamma). \tag{3}$$

This generative scheme for the weights $\boldsymbol{\pi}$ is denoted $\boldsymbol{\pi} \sim \mathrm{GEM}(\gamma)$ after Griffiths, Engen and McCloskey [Pitman (2002)].

This stick-breaking construction clearly illustrates that under a Dirichlet process model, the distribution G is discrete: the parameters $\boldsymbol{\phi}_{tji}$ for each channel observation are generated from a set of atoms $\boldsymbol{\phi}_k \sim H$ with π_k specifying the probability of atom $\boldsymbol{\phi}_k$. The discrete nature of our model permits information sharing across channels i within a seizure and across seizures j by clustering our channel observations into groups that share the same underlying parameter atoms $\boldsymbol{\phi}_k$. We introduce the additional notation $z_{tji} = k$ to specify which of the unique parameter atoms $\boldsymbol{\phi}_k$ underlies the channel observation $\mathbf{x}_{tji}$.

However, a key disadvantage of the generic DP model is that it contains only one level of clustering: it would consider channel observations of all the seizures at once, without distinguishing between channel observations belonging to one seizure or another. While clustering channel activity alone is indeed a relevant enterprise, it is not as well suited for the more high-level clinical analysis of determining similarities (and differences) between seizures, especially for patients with many recorded seizures. This disadvantageous aspect of the DP model is illustrated with an example in the top left section of Figure 3.

We need to add more structure to our model in order to address the heterogeneity at multiple levels of our seizure data. The hierarchical Dirichlet process of Teh

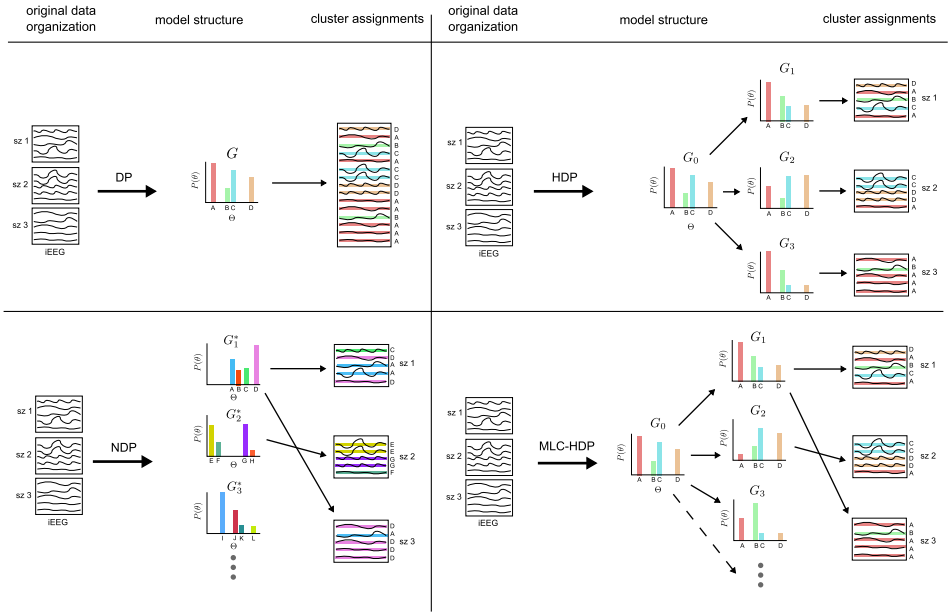

FIG. 3. *A comparison of how four different models—the Dirichlet process* (*DP*), *the hierarchical Dirichlet process* (*HDP*), *the nested Dirichlet process* (*NDP*) *and our multi-level clustering HDP* (*MLC-HDP*)*—would be applied to three seizures of a single patient*. Top left: *The DP yields a single discrete measure G over the channel observations*, *producing a clustering over the channels* (*shown by the color under each channel*) *that is indiscriminate of the seizure to which each channel belongs*. Top right: *The HDP yields a global discrete measure* G_0 *and measures* $\{G_j\}_{j=1}^3$ *for each seizure that shares atoms with* G_0 *which produces a clustering over channel observations with unique weights for each seizure. However, there is still only a single clustering at the level of channel observations and no direct clustering of seizures*. Bottom left: *The NDP yields a collection of measures* $(G_\ell^\star)$ *that each contain a collection of unique atoms and weights*, *resulting in a clustering over the seizure types and the channel types but without any sharing of channel atoms between seizures. Our MLC-HDP model also yields a collection of measures* (G_ℓ), *but they share the same collection of atoms while each having unique weights. Like the NDP, the MLC-HDP produces a clustering over the seizure types and the channel types*.

et al. (2006) is a natural extension of the DP model to grouped data, where the groups in this context would be the seizures within a patient. In the HDP, the channels of each seizure j would be grouped together and modeled by an unknown distribution that is unique to each seizure,

$$\boldsymbol{\phi}_{tji} \sim G_j,$$

where we employ a Dirichlet process prior $G_j \sim \mathrm{DP}(\alpha, G_0)$ to each seizure-specific distribution G_j. Information is shared between seizures via the common base measure G_0 that is itself modeled using a Dirichlet process prior $G_0 \sim \mathrm{DP}(\gamma, H)$.

The stick-breaking representation defines a realization from the HDP as

$$G_j = \sum_{k=1}^{\infty} \pi_{jk} \delta_{\boldsymbol{\phi}_k} \quad \text{and} \quad G_0 = \sum_{k=1}^{\infty} \beta_k \delta_{\boldsymbol{\phi}_k}, \tag{4}$$

where $\delta_{\boldsymbol{\phi}_k}$ again is a unit measure concentrated at each sample $\boldsymbol{\phi}_k$ from the base measure $\boldsymbol{\phi}_k \sim H$. The weights $\boldsymbol{\beta}$ for the global measure G_0 are generated from the same stick-breaking process as in the DP model, $\boldsymbol{\beta} \sim \mathrm{GEM}(\gamma)$. The weights $\boldsymbol{\pi}_j = (\pi_{jk})$ for each seizure-specific distribution G_j are themselves a sample from a Dirichlet process, $\boldsymbol{\pi}_j \sim \mathrm{DP}(\alpha, \boldsymbol{\beta})$.

A key aspect of the HDP model is that the same set of parameter atoms $\delta_{\boldsymbol{\phi}_k}$ are shared by both the global measure G_0 and the seizure-specific distributions G_j, though each seizure can have a unique set of weights $\boldsymbol{\pi}_j$ for those parameter atoms. Thus, the HDP model is still only clustering at the level of the channel observations, just as in the DP model.

If we also desire to cluster at the level of entire seizures, we could only do this with the HDP model by downstream analysis of the clustered channel observations. The lack of an explicit clustering of seizures is a potential disadvantage of the HDP model that we illustrate with an example in the top right section of Figure 3.

The nested Dirichlet process (NDP) of Rodríguez, Dunson and Gelfand (2008) is an alternative approach that allows for explicit clustering at both the level of channel observations and the level of seizures. The stick-breaking construction for the NDP is

$$G_\ell^\star = \sum_{k=1}^{\infty} \pi_{\ell k}^\star \delta_{\boldsymbol{\phi}_{\ell k}^\star} \quad \text{and} \quad G_0^\star = \sum_{k=1}^{\infty} \beta_k^\star \delta_{G_\ell^\star}. \tag{5}$$

In the NDP, the parameter atoms $\boldsymbol{\phi}_{\ell k}^\star$ are unique for each measure $G_\ell^\star$ but still sampled from the base measure $\boldsymbol{\phi}_{\ell k}^\star \sim H$. The associated weights $\pi_{\ell k}^\star \sim \mathrm{GEM}(\alpha)$ for each parameter atom are also unique for each group. The global measure G_0 is a mixture of these measures $G_\ell^\star$ with weights $\boldsymbol{\beta}^\star \sim \mathrm{GEM}(\gamma)$.

There are two key distinctions between the NDP and HDP approaches. First, the measures $(G_\ell^\star)_{\ell=1}^\infty$ of the NDP contain both unique parameter atoms and unique

weights, whereas the measures $\{G_j\}_{j=1}^{J}$ of the HDP contain only unique weights, with their parameter atoms all defined by the global measure G_0. Second, multiple seizures j can be assigned by the NDP to a particular measure $G_\ell^\star$, which allows for the clustering of seizures. In contrast, each seizure j has its own unique measure G_j in the HDP. Unlike the DP and the HDP, the NDP is indeed a two-level clustering model, where the channel observations are clustered on one level, and the seizures are clustered on another level. The NDP model is thus more clinically useful because the seizure level clusters can be used to organize the seizures directly.

Unfortunately, the assumption of the NDP that each seizure cluster $G_\ell^\star$ contains its own unique parameter atoms $\boldsymbol{\phi}_{\ell k}$ is not realistic for our seizure application. A clinician may deem two seizures to be fundamentally different even if a subset of their channels are behaving quite similarly. We would like to share channel-type atoms across the seizure-type clusters, which is not possible with the NDP. Furthermore, the unique channel-type atoms for each seizure-type mixture can lead to a large number of total channel-type atoms, making practical computational issues a potential barrier to scaling the model up to a realistic number of seizures. We illustrate this disadvantage of the NDP model with an example in the bottom left section of Figure 3.

To address the disadvantages of both the HDP and NDP approaches, we propose a new hierarchical nonparametric Bayesian model. Recall that we need to model the underlying parameters $\boldsymbol{\phi}_{tji}$ of each channel observation $\mathbf{x}_{tji}$ for a single patient t. We model the collection of underlying channel parameters $\boldsymbol{\phi}_{tji}$ as a set of atoms from a *seizure-type* measure G_ℓ,

$$G_\ell = \sum_{k=1}^{\infty} \pi_{\ell k}^{G} \delta_{\boldsymbol{\phi}_k}, \tag{6}$$

where each parameter atom $\boldsymbol{\phi}_k$ is generated from the base measure $\boldsymbol{\phi}_k \sim H$ and represents the different types of channel observations. Unlike the NDP, these channel parameter atoms $\boldsymbol{\phi}_k$ can be shared across different seizure-type measures G_ℓ. The weights $\boldsymbol{\pi}_\ell^G$ associated with measure G_ℓ are a sample from a Dirichlet process, $\boldsymbol{\pi}_\ell^G \sim \mathrm{DP}(\alpha^G, \boldsymbol{\beta}^G)$ with $\boldsymbol{\beta}^G \sim \mathrm{GEM}(\gamma^G)$.

Next, we model the collection of seizure-type measures G_ℓ as a set of atoms from an overall measure F shared across seizures,

$$F = \sum_{\ell=1}^{\infty} \pi_\ell^F \delta_{G_\ell}, \tag{7}$$

where the weights $\boldsymbol{\pi}^F$ are generated from a Dirichlet process $\boldsymbol{\pi}^F \sim \mathrm{DP}(\alpha^F, \boldsymbol{\beta}^F)$ with $\boldsymbol{\beta}^F \sim \mathrm{GEM}(\gamma^F)$. This formulation is similar to the NDP in that it implies a mixture of mixtures, but our weights $\boldsymbol{\pi}^F$ are sampled from a DP, whereas the analogous weights in the NDP are sampled from a GEM stick-breaking process.

Also, our model shares channel parameter atoms $\boldsymbol{\phi}_k$ across different seizure-type measures G_ℓ unlike the NDP.

We can think about our model's generative process as each seizure j selecting a G_ℓ from the mixture F. The mixture F shares information across seizures within a patient (with parameters $\boldsymbol{\pi}^F, \alpha^F$ and $\boldsymbol{\beta}^F$). The selected G_ℓ can be thought of as the seizure-type cluster to which seizure j is assigned.

Within that seizure type, particular parameter atoms $\boldsymbol{\phi}_k$ (which we may think of as describing the channel type) are chosen from G_ℓ as the underlying parameters for each channel observation $\mathbf{x}_{tji}$ in seizure j. The mixture G_ℓ shares information across channels within a seizure (with parameters $\boldsymbol{\pi}^G, \alpha^G$ and $\boldsymbol{\beta}^G$). In Section 3.2, we will add an additional level to our model where information is shared across patients.

We call our model the multi-level clustering hierarchical Dirichlet process (MLC-HDP). Our MLC-HDP blends the desirable aspects of the HDP and NDP for our seizure iEEG application. It incorporates the multiple levels of clustering introduced by the NDP, but—as in the HDP—the different seizure type mixtures G_ℓ share a common set of channel-type atoms $\boldsymbol{\phi}_k$, allowing for both across seizure-type sharing of channel information and efficient scaling as the number of seizure types grows.

Our two-level MLC-HDP model is actually equivalent to the hybrid NDP-HDP model described by James' comment to the NDP paper of Rodríguez, Dunson and Gelfand (2008). The final panel (bottom right) of Figure 3 depicts our MLC-HDP model applied to the same example dataset as the DP, HDP and NDP. We have explored the differences between our two MLC-HDP and the NDP with a simulation study in our supplementary materials. Under the same simulation settings as Rodríguez, Dunson and Gelfand (2008), our MLC-HDP model is able to better estimate subtle mixture densities.

3.2. *Sharing information across patients.* In Section 3.1 above, we developed a multi-level clustering hierarchical Dirichlet process model for sharing information across seizures within a single patient. We now extend this model to share information across the ten different patients in our iEEG data. Specifically, we add another hierarchical Dirichlet process layer to our model that contains a mixture of patient types.

Instead of a single patient mixture F over the seizure-types $(\delta_{G_\ell})_{\ell=1}^{\infty}$, as we had in equation (7), we define a mixture F_t,

$$F_t = \sum_{\ell=1}^{\infty} \pi_{t\ell}^F \delta_{G_\ell}, \tag{8}$$

with weights $\boldsymbol{\pi}_t^F \sim \mathrm{DP}(\alpha^F, \boldsymbol{\beta}^F)$ where $\boldsymbol{\beta}^F \sim \mathrm{GEM}(\gamma^F)$. These mixtures F_t represent a *patient-type*, which we link together with a mixture E over our population

of patients,

$$E = \sum_{t=1}^{\infty} \pi_t^E \delta_{F_t}, \tag{9}$$

with weights $\boldsymbol{\pi}^E \sim \mathrm{DP}(\alpha^E, \boldsymbol{\beta}^E)$ where $\boldsymbol{\beta}^E \sim \mathrm{GEM}(\gamma^E)$. Note that in equation (8) each patient-type t uses different weights $\boldsymbol{\pi}_t^F$ over the same seizure-types (δ_{G_ℓ}), a representation that parallels the fact that each seizure-type ℓ uses different weights $\boldsymbol{\pi}_\ell^G$ over the same channel-parameter atoms ($\delta_{\boldsymbol{\phi}_k}$) in equation (6).

In total, our MLC-HDP model induces clustering at three different levels of the iEEG data: channel observations are clustered within a particular seizure and seizures are clustered both within a particular patient as well as across our population of ten patients. Referring back to Figure 2, we have used different colors to illustrate the clustering induced by our MLC-HDP model at the levels of channels, seizures and patients. While the DP, HDP and NDP models can also be applied to a multi-patient dataset of seizures, those approaches suffer from the same problems discussed previously for those models in Section 3.1.

3.3. *Model implementation*. Our model implementation is based on the Gibbs sampler, a Markov chain Monte Carlo sampling strategy that iteratively samples from the conditional distributions of each parameter given current values of the other parameters [Geman and Geman (1984)]. We divide the parameters that need to be estimated into several sets: (1) the channel observation parameters $\boldsymbol{\phi}_k$, (2) the mixture weights $\boldsymbol{\pi}$ and their shared parameters $\boldsymbol{\beta}$ at each of the three model levels, (3) the concentration parameters γ and α at each of the three model levels, and (4) a set of indicator variables z_t^E, z_{tj}^F and z_{tji}^G that indicate the current clusters of patient-types, seizure-types and channel-types, respectively.

Following Teh et al. (2006), we use a Rao–Blackwellized direct assignment sampler (also known as a *collapsed* Gibbs sampler) at each level that marginalizes out the mixture weights $\boldsymbol{\pi}$ at each of the three model levels, as well as the channel observation parameter atoms $\boldsymbol{\phi}_k$. Such marginalization has been shown to improve efficiency of Gibbs samplers [Casella and Robert (1996), Gelfand and Smith (1990), Liu, Wong and Kong (1994)]. Below, we briefly summarize the three steps of our collapsed Gibbs sampler, with full details given in our supplementary materials. Matlab code implementing our model is also available from the first author's website.[3]

Sampling cluster indicators at each level. At the level of the channel observations, we sample the cluster indicator z_{tji}^G for channel observation i in seizure j of patient t from a multinomial where the probability of selecting cluster k is

$$p(z_{tji}^G = k | \cdots) \propto d_k(\mathbf{x}_{tji})(\alpha^G \beta_k^G + n_{\ell k}^G),$$

[3] www.seas.upenn.edu/~wulsin

where ℓ is the current cluster indicator for the seizure j, that is, $z_{tj}^F = \ell$ from the seizure level of the model. The count variable $n_{\ell k}^G$ represents the number of channel observations assigned to channel type k across all the seizures in seizure type ℓ. The value $d_k(\mathbf{x})$ is the posterior predictive likelihood which, under our Rao–Blackwellized scheme, is a multivariate Student-t [Gelman et al. (2004), page 88]. The conditional distributions for the seizure-level indicators z_{tj}^F and patient-level indicators z_t^E are similar, with details given in our supplementary materials.

Sampling global parameters at each level. As mentioned above, the mixture weights $\boldsymbol{\pi}$ are marginalized out of our sampling scheme, but we must still sample the global parameters $\boldsymbol{\beta}$ for those weights at each level of the model. For the channel-type parameters $\boldsymbol{\beta}^G$, we first generate values

$$m_{\ell k}^G = \sum_{s=1}^{n_{\ell k}^G} \theta_s \qquad \text{where } \theta_s | \cdots \sim \text{Ber}\Big(\frac{\alpha^G \beta_k^G}{\alpha^G \beta_k^G + s}\Big),$$

and then use those values to sample the channel-type parameters $\boldsymbol{\beta}^G$,

$$\boldsymbol{\beta}^G \sim \text{Dir}(m_{\cdot 1}^G, \ldots, m_{\cdot K}^G, \gamma^G),$$

where $m_{\cdot k}^G = \sum_\ell m_{\ell k}^G$. The sampling of the seizure-type global parameters $\boldsymbol{\beta}^F$ and patient-type global parameters $\boldsymbol{\beta}^E$ proceeds similarly.

Sampling concentration parameters at each level. If we set Gamma prior distributions for each of our concentration parameters (α^E, α^F, α^G) and (γ^E, γ^F, γ^G), then the conditional posterior distribution of these parameters is also Gamma. Details are provided in our supplementary materials.

In our EEG seizure application, we used Gamma(1, 1) priors for α^E, α^F, γ^E and γ^F and Gamma(5, 1) priors for α^G and γ^G. In our supplementary materials, we evaluate the sensitivity of our results to different prior distributions for our six concentration parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$. We find very little difference in our posterior clustering results when we employ priors for $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ that differ substantially in their prior predictive characteristics.

In a standard (i.e., noncollapsed) Gibbs sampler, we would also need to obtain posterior samples of the channel observation parameter atoms $\boldsymbol{\phi}_k$. We outline an alternative sampler that does not marginalize out the level weights and observation model parameters $\boldsymbol{\pi}$ and $\boldsymbol{\phi}_k$ in our supplementary materials, and we show that our collapsed Rao–Blackwellized (RB) sampler is superior to several alternative sampling schemes in terms of convergence and autocorrelation of the sampled values.

4. Application to intracranial EEG of human epileptic seizures. Our available data is 193 intracranial EEG (iEEG) seizure records from 10 patients from the Children's Hospital of Philadelphia. As outlined in Section 2, the raw iEEG data

was reduced to a five-dimensional vector of channel observations for each channel i ($i = 1, \ldots, N_{tj}$) in seizure j ($j = 1, \ldots, J_t$) of patient t ($t = 1, \ldots, 10$). Our MLC-HDP model was fit to this data using the collapsed Gibbs sampler that we briefly outlined in Section 3.3. We ran 25 chains, each yielding 200 samples after a burn-in of 500 iterations and thinning every 20 iterations.

4.1. *The advantages of a hierarchical model.* Since we were aware of no other hierarchical models that share information across seizures and patients in the epilepsy modeling literature, we explored the extent to which this information sharing improves the model for a subset of patient seizures that are held out of the model estimation. In particular, we are interested in comparing our MLC-HDP model to a nonhierarchical alternative, the Dirichlet process (DP), that does not address the grouped structure in our data, similar to the comparison given in Teh et al. (2006).

We compared the MLC-HDP trained on the full hierarchy over all patients and seizures to the DP trained with two different datasets: one with only the channel activities from the seizures of a single patient t and another with channel activities from all patients. We denote the MLC-HDP's modeling scenario as M_3 and the two DP modeling scenarios as M_1 and M_2, respectively. For a given patient t, we created $J_t - 1$ training and testing sets for each of these three modeling scenarios. We summarize these three scenarios below:

M_1: The channel observations from seizures $1, \ldots, J_t$ of patient t are used to train a standard DP mixture model,

M_2: The channel observations from seizures $1, \ldots, J_t$ of patient t and all the seizures $j' \in \{1, \ldots, J_{t'}\}$ of all the other patients $t' \neq t$ are used to train a standard DP mixture model,

M_3: The same data as M_2 is used but our full MLC-HDP model is implemented on the full data.

To evaluate these three models, similarly to Teh et al. (2006), we use the conditional perplexity—the log of the Shannon entropy [MacKay (2003), Shannon (1948)]—of the future held out seizures $\{j+1, \ldots, J_t\}$ given the cluster index for each channel observation

$$\mathrm{PP}(S_{tj+1}, \ldots, S_{tJ_t} | \cdots) = \exp\left(-\frac{1}{J_t - j} \sum_{j'=j+1}^{J_t} \log p\big(S_{tj'} | z^G_{tj'1}, \ldots, z^G_{tj'N_{tj'}}\big) \right),$$

where

$$p\big(S_{tj'} | z^G_{tj'1}, \ldots, z^G_{tj'N_{tj'}}\big) = \prod_{i=1}^{N_{tj'}} f_{(z^G_{tj'i})}(\mathbf{x}_{tj'i}),$$

with the notation $S_{tj} = \{\mathbf{x}_{tji}\}_{i=1}^{N_{tj}}$ denoting the set of channel observations occurring in seizure j of patient t. Lower perplexity values indicate a better model.

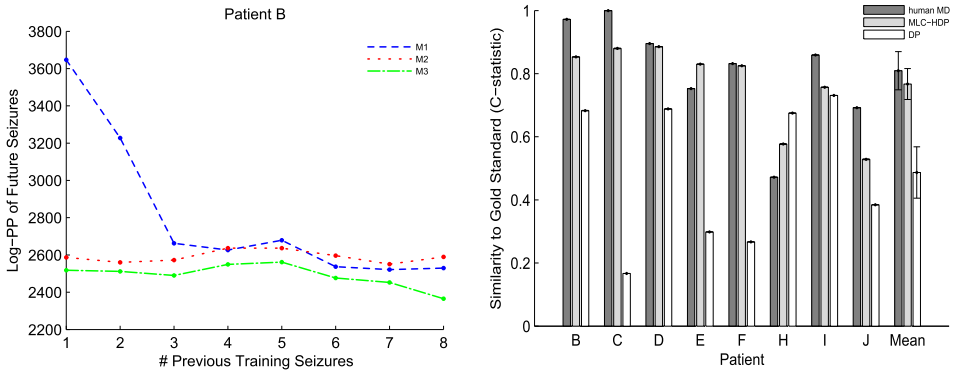

FIG. 4. Left: *The mean log-perplexity of patient's B future seizures as a function of the number of previous training seizures. The results for three models are displayed*: M_1, *a DP with training data from previous seizures from the patient*; M_2, *a DP using training data from previous seizures from the patient as well as all the seizures from all the other patients*; M_3, *a MLC-HDP model with its full patient-seizure-channel clustering. The supplementary materials contain image representations of each of patient's B seizures*. Right: *The seizure clustering similarity between an expert's clustering and the other expert clustering, as well as the clusterings found by the DP and MLC-HDP models. Patients A and G are excluded because they only had one seizure. Error bars bound the region of one standard error.*

The left side of Figure 4 shows the results for a single patient (patient B), with the corresponding plots for other patients being similar. For low amounts of training data, the DP model with only patient's B seizures (M_1) suffers relative to the other models. As more and more seizures are added to the training set, the DP model with only patient's B seizures (M_1) improves considerably. The DP model incorporating training data from all other patients (M_2) demonstrates much better performance than M_1 when the amount of training data from patient B is small but becomes slightly worse than M_1 after the first five seizures of patient B are included in the training set.

Though it has the same training data as M_2, the hierarchical MLC-HDP (M_3) model has consistently lower perplexity (better performance) than both M_1 and M_2 models across every number of patient B training seizures. These results show the value of a hierarchical model for our iEEG seizure application. The multiple levels of clustering allow the model to intelligently blend local data (e.g., the seizures for a particular patient) with global data (e.g., the seizures from all other patients). This hierarchical sharing of information is particularly appropriate for this application where the number of seizures varies widely between patients.

4.2. *Seizure clustering performance.* To assess how well the MLC-HDP clusters the 193 seizures from 10 patients, we had the seizures for each patient manually clustered by two board-certified epileptologists. This task is inherently subjective and uncertain, so we used two physicians to quantitatively assess the natural

uncertainty in this task. These two physicians have trained and worked with each other for over ten years, so their markings should be as close as two separate human markers can be. For our subsequent analysis, we arbitrarily chose one of the physicians as the "gold standard." Our results do not change substantially when the markings of the other physician are used as the gold standard instead.

In addition to the MLC-HDP and physician-seizure clusterings, we also desired a baseline-seizure clustering from another model. Of the related models we have previously considered (DP, HDP, NDP), the NDP is the only other one that naturally yields an explicit seizure clustering when channel activities are used as the observations. Unfortunately, the NDP involves assumptions and computational burdens that are impractical for this problem, as we discuss in Section 3.1. For the baseline-seizure clustering, we thus decided to work with a parameterization of a seizure that is agnostic to the number of channel activities in that seizure. Such a setting allows us to straightforwardly compare seizures with a single patient and across multiple patients. It also allows us to cluster seizures using a standard DP. We believe this method for producing baseline-seizure clusterings is the closest reasonable alternative to those produced by our MLC-HDP.

For the seizure parameterization, we worked with the six features of Schiff et al. (2005) since we believe they capture the most important dynamics of a seizure, namely, the synchronization of different areas of the brain and their frequency characteristics. These features were calculated using the same sliding window as our frequency-band features for individual channels and are described in the supplementary materials. As with the channel features, these seizure features were concatenated across time windows and reduced to the top 20 principal components, retaining 72.3% of the variance.

We use the c-statistic of Rand (1971) to determine the similarity between each seizure clustering and the gold standard because it elegantly handles different numbers of clusters and labels between two differ clusterings. The c-statistic between two clusterings Y and Y' of N objects is given by

$$c(Y, Y') = 1 - \frac{1}{\binom{N}{2}}\left(\frac{1}{2}\sum_i\left(\sum_j n_{ij}\right)^2 + \frac{1}{2}\sum_j\left(\sum_i n_{ij}\right)^2 - \sum_i\sum_j n_{ij}^2\right), \tag{10}$$

where n_{ij} is the number of points simultaneously in the ith cluster of Y and the jth cluster of Y'.

We estimated both the MLC-HDP and DP models using the seizures from each patient separately as well as using all patients together, and both models yield superior clustering performance relative to the gold standard manual clustering when all the seizures across patients are clustered together. We interpret this result by considering the performance of human experts who, in their clustering of each patient's seizures separately, do not forget about the many thousands of seizures they have seen before this task, and so the MLC-HDP and DP models that combine across patients are more effective. Clusterings over all patients are used for both the MLC-HDP and DP models for our subsequent results.

The right side of Figure 4 compares the MLC-HDP and DP clustering performance to the gold standard clustering from one physician, as well as the manual cluster from the other physician. The physicians tend to agree best with each other, as one would expect, but the MLC-HDP's clusterings are often close to those of the physicians. The DP model does not perform nearly as well as our MLC-HDP model in terms of similarity to the physician clusterings.

Patient H is an interesting exception to the general trend. The two physicians disagreed most on this patient. The seizures of patient H are extremely similar, and it seems that the models and the experts disagreed on the best way to split them up (we give the seizure images for patient H in our supplementary materials). The gold-standard expert and the DP had fewer clusters in the seizures of patient H, whereas the MLC-HDP and the other expert split them more.

We believe the main difference between the MLC-HDP and DP clusterings comes from the differences in how each method represents a seizure. In the seizure features used by the DP, the activity of a few channels can be washed out by the majority of the channels. Since the MLC-HDP explicitly models the activity of each individual channel, a few important channels can more easily sway the entire model of the seizure. Both physicians indicated that they followed the clinical practice of defining a seizure in large part by the activity of a few "leading" channels. These results incline us to believe that any attempt to model seizures must begin with modeling the activity of individual channels and build from there. We also believe that the absence of such methods until now explains the limited usage of seizure clustering within and between patients in the epilepsy literature.

4.3. *Finding similarities between seizures of different patients.* We can use the seizure clusterings over the 5000 samples (200 samples from each of the 25 chains) to derive a similarity metric between seizures. Similarity metrics that generalize between patients have received scant attention in the epilepsy literature. This metric is simply the posterior probability that two seizures occur in the same cluster. We use the posterior probability of the two seizures *not* clustering together as a distance metric. We then use least squares multidimensional scaling [Hastie, Tibshirani and Friedman (2001), page 502] to find a 2-dimensional projection of the seizures, where seizures closer together in the 2D space are more similar.

The seizure images corresponding to the 193 seizures in our dataset are plotted in Figure 5. We note that the seizures of the same patients (which share the same color outline) are often situated near each other, though they can also vary greatly. This 2D representation also allows us to find seizures of different patients that are similar to each other. For example, patients A, B, C, D and H all have seizures that are close in the 2D representation (top right area of Figure 5). This clustering task would be considerably more difficult if one had to manually wade though all the 193 seizures individually or manually examine them all simultaneously. We hope that our MLC-HDP model will be helpful in the future of organizing and mining

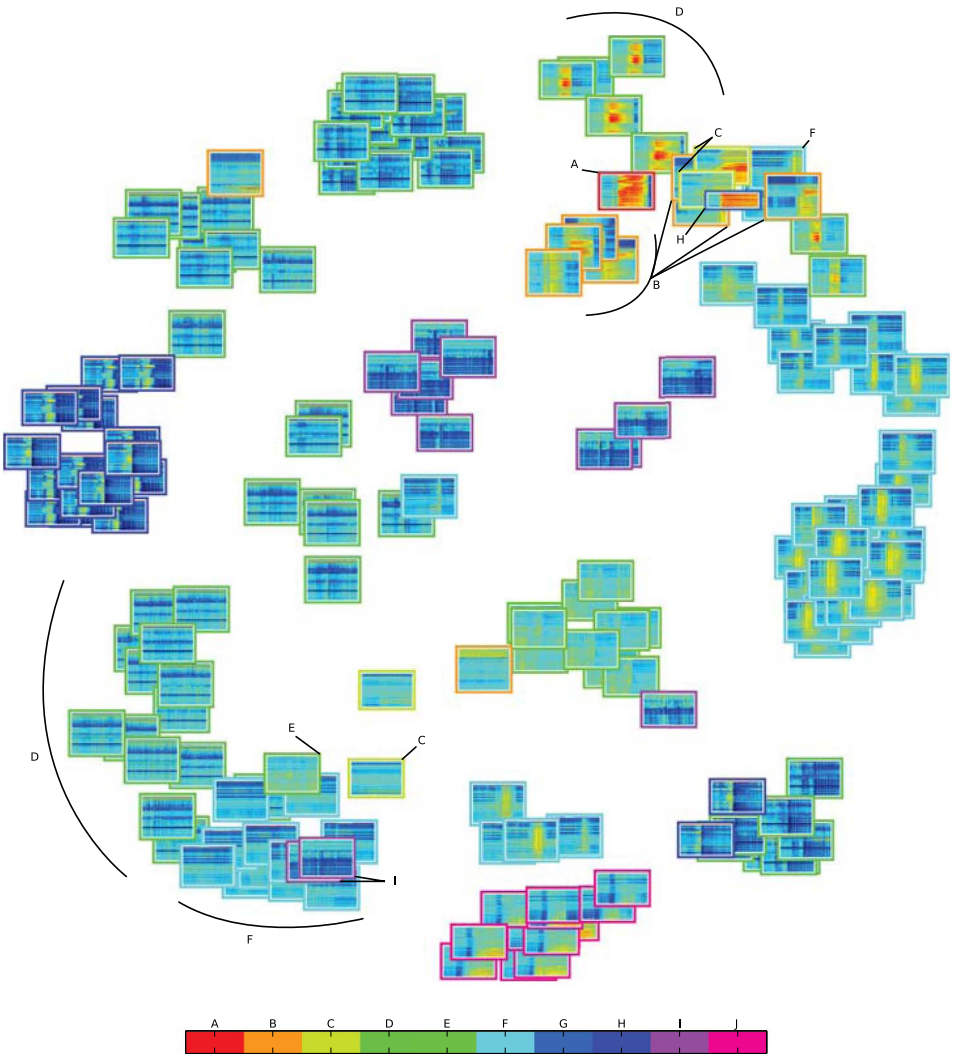

FIG. 5. *A 2D representation of the similarities between seizures across all patients in the MLC-HDP model, where seizure images closer together indicate more similarity. Image positions are slightly jittered to make more seizures visible. Each seizure image depicts the* 13–30 *Hz feature* (*blue*: *low intensity*, *red*: *high intensity*) *across all the channels* (*rows*) *at each time point* (*columns*). *The image border colors denote the patient for each seizure. The channel order of the seizures is consistent within a patient but not so between patients. Note that seizures from different patients can have quite similar dynamics. A few examples of these similar seizures are explicitly marked with the corresponding patient letter for each seizure.*

much larger intracranial EEG datasets with hundreds of patients and thousands of seizures.

In this work we have introduced a model meant for describing the data of many epilepsy patients, but when the number of patients is relatively small and cluster-

ing on them does not necessarily make sense (or is not useful), a slightly simpler model could also be appropriate. In this case, we would replace the clustering at the patient level with instead a single HDP on the seizure-type clusters, where each patients has its own weights over those seizure-type clusters. We plan to explore this alternative in future work.

In our supplementary materials, we investigate the influence of individual patients on our results by fitting our model after removing the seizures from a single individual from the data (and then repeating this procedure for each individual). We find that our MLC-HDP results are less sensitive to the influence of individual patients than the DP model that we also considered in this section.

5. Discussion. In this paper we introduce a new Bayesian nonparametric model: the multi-level clustering hierarchical Dirichlet process (MLC-HDP), which simultaneously shares information across multiple levels of a dataset via clustering. We applied our MLC-HDP model on an iEEG seizure dataset of 193 seizures from 10 patients. We find that the hierarchical aspect of the model has advantages over less structured alternatives such as the DP model, and we also show how the model clusters seizures reasonably well, compared to manual clustering by physician experts which is the standard method used in assessing patients undergoing iEEG monitoring.

Our model-based seizure clusterings can be used to produce visualizations helpful in answering questions like "what seizures from other patients is this seizure similar to?" and "what are the different types of seizures present in this dataset?" These questions have important implications for clinical care and decision-making during evaluation for epilepsy surgery. We also show in our supplementary materials that the MLC-HDP model has desirable properties such as low dependence on individual patients, low prior sensitivity and low autocorrelation. We believe these results show the model's use in organizing and understanding large, multi-level datasets like iEEG seizures from a number of patients.

Our MLC-HDP approach does not explicitly model the spatial relationships of channels since physical location of each channel is unique for each patient and difficult to determine exactly. Nevertheless, we plan to use the channel clusterings produced by MLC-HDP inference in conjunction with a 3D brain visualization tool developed by our research group to see whether any spatial patterns become apparent, as we would expect. We expect this analysis to aid practicing epileptologists when they determine the physiologic extent of seizure onset and spread regions for individual patients. We can also use the MLC-HDP's seizure clusterings to explore whether seizures that occur close to each other in time tend to also be of the same seizure type. Anecdotal evidence suggests that this is the case for

some patients. Such an analysis would aid epileptologists in visualizing ways in which a patient's seizures change over longer periods of time (days or weeks).

Acknowledgments. We thank Eric Marsh and Brenda Porter, of the Children's Hospital of Philadelphia, for the continuous iEEG records from which the seizures dataset derived as well as their manual-seizure clustering and helpful discussion. We also thank Emily Fox for her helpful ideas and discussion.

SUPPLEMENTARY MATERIAL

Supplement to "Nonparametric multi-level clustering of human epilepsy seizures" (DOI: 10.1214/15-AOAS851SUPP; .pdf). We present visual summaries of the seizures for each of our patients. We provide details of our Normal model and conjugate prior. We explore the sensitivity of our results to different priors for the concentration parameters as well as the influence of individual patients. We outline detailed algorithms of our MCMC model implementation as well as a comparison to alternative sampling schemes. We gives mathematical expressions for the 6 Schiff features mentioned in our results section. We provide a synthetic data comparison to the NDP model.

REFERENCES

ADELI, H., ZHOU, Z. and DADMEHR, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *J. Neurosci. Methods* **123** 69–87.

BARTOLOMEI, F., COSANDIER-RIMELE, D., MCGONIGAL, A., AUBERT, S., RÉGIS, J., GAVARET, M., WENDLING, F. and CHAUVEL, P. (2010). From mesial temporal lobe to temporoperisylvian seizures: A quantified study of temporal lobe seizure networks. *Epilepsia* **51** 2147–2158.

CASELLA, G. and ROBERT, C. P. (1996). Rao–Blackwellisation of sampling schemes. *Biometrika* **83** 81–94. MR1399157

CHAN, A. M., SUN, F. T., BOTO, E. H. and WINGEIER, B. M. (2008). Automated seizure onset detection for accurate onset time determination in intracranial EEG. *Clin. Neurophysiol.* **119** 2687–2696.

CHAOVALITWONGSE, W. A. (2008). Novel quadratic programming approach for time series clustering with biomedical application. *J. Comb. Optim.* **15** 225–241. MR2383309

DE TISI, J., BELL, G. S., PEACOCK, J. L., MCEVOY, A. W., HARKNESS, W. F. J., SANDER, J. W. and DUNCAN, J. S. (2011). The long-term outcome of adult epilepsy surgery, patterns of seizure remission, and relapse: A cohort study. *Lancet* **378** 1388–1395.

ENGEL, J. JR. and PEDLEY, T. A., eds. (2008). *Epilepsy*: *A Comprehensive Textbook*, 2nd ed. Lippincot Williams & Wilkins, Philadelphia, PA.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230. MR0350949

FRENCH, J. A. (2007). Refractory epilepsy: Clinical overview. *Epilepsia* **48** 3–7.

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409. MR1141740

GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (2004). *Bayesian Data Analysis*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL. MR2027492

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–741.

GHOSH-DASTIDAR, S., ADELI, H. and DADMEHR, N. (2008). Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Trans. Biomed. Eng.* **55** 512–518.

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2001). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*. Springer, New York. MR1851606

HEGDE, A., ERDOGMUS, D., SHIAU, D. S., PRINCIPE, J. C. and SACKELLARES, C. J. (2007). Clustering approach to quantify long-term spatio-temporal interactions in epileptic intracranial electroencephalography. *Comput. Intell. Neurosci.* **2007** 83416.

KLATCHKO, A., RAVIV, G., WEBBER, W. R. and LESSER, R. P. (1998). Enhancing the detection of seizures with a clustering algorithm. *Electroencephalogr. Clin. Neurophysiol.* **106** 52–63.

LIU, J. S., WONG, W. H. and KONG, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* **81** 27–40. MR1279653

MACKAY, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge Univ. Press, New York. MR2012999

OSSADTCHI, A., GREENBLATT, R. E., TOWLE, V. L., KOHRMAN, M. H. and KAMADA, K. (2010). Inferring spatiotemporal network patterns from intracranial EEG data. *Clin. Neurophysiol.* **121** 823–835.

PARAMANATHAN, P. and UTHAYAKUMAR, R. (2008). Application of fractal theory in analysis of human electroencephalographic signals. *Comput. Biol. Med.* **38** 372–378.

PITMAN, J. (2002). Poisson–Dirichlet and GEM invariant distributions for split-and-merge transformation of an interval partition. *Combin. Probab. Comput.* **11** 501–514. MR1930355

QUYEN, M. L. V., SOSS, J., NAVARRO, V., ROBERTSON, R., CHAVEZ, M., BAULAC, M. and MARTINERIE, J. (2005). Preictal state identification by synchronization changes in long-term intracranial EEG recordings. *Clin. Neurophysiol.* **116** 559–568.

RAND, M. (1971). Objective criteria for the evaluation of methods clustering. *J. Amer. Statist. Assoc.* **66** 846–850.

REIJNEVELD, J. C., PONTEN, S. C., BERENDSE, H. W. and STAM, C. J. (2007). The application of graph theoretical analysis to complex networks in the brain. *Clinical Neurophysiology* **118** 2317–2331.

RODRÍGUEZ, A., DUNSON, D. B. and GELFAND, A. E. (2008). The nested Dirichlet process. *J. Amer. Statist. Assoc.* **103** 1131–1144. MR2528831

SARACENO, B., AVANZINI, G. and LEE, P. (2005). Atlas: Epilepsy care in the world. Technical report, World Health Organization, Geneva.

SCHIFF, S. J., SAUER, T., KUMAR, R. and WEINSTEIN, S. L. (2005). Neuronal spatiotemporal pattern discrimination: The dynamical evolution of seizures. *NeuroImage* **28** 1043–1055.

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. MR1309433

SHANNON, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* **27** 379–423, 623–656. MR0026286

SRINIVASAN, V., ESWARAN, C. and SRIRAAM, N. (2007). Approximate entropy-based epileptic EEG detection using artificial neural networks. *IEEE Trans. Inf. Technol. Biomed.* **11** 288–295.

STAM, C. J. (2005). Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. *Clin. Neurophysiol.* **116** 2266–2301.

TEH, Y. W., JORDAN, M. I., BEAL, M. J. and BLEI, D. M. (2006). Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.* **101** 1566–1581. MR2279480

WULSIN, D. F., JENSEN, S. T. and LITT, B. (2016). Supplement to "Nonparametric multi-level clustering of human epilepsy seizures." DOI:10.1214/15-AOAS851SUPP.

D. F. WULSIN
DEPARTMENT OF BIOENGINEERING
UNIVERSITY OF PENNSYLVANIA
301 HAYDEN HALL
3320 SMITH WALK
PHILADELPHIA, PENNSYLVANIA 19104
USA
E-MAIL: wulsin@seas.upenn.edu

S. T. JENSEN
DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
463 HUNTSMAN HALL
3730 WALNUT STREET
PHILADELPHIA, PENNSYLVANIA 19104
USA
E-MAIL: stjensen@wharton.upenn.edu

B. LITT
DEPARTMENTS OF BIOENGINEERING AND NEUROLOGY
UNIVERSITY OF PENNSYLVANIA
301 HAYDEN HALL
3320 SMITH WALK
PHILADELPHIA, PENNSYLVANIA 19104
USA
E-MAIL: littb@mail.med.upenn.edu