A BAYESIAN MODEL OF DOSE-RESPONSE FOR CANCER DRUG STUDIES

BY WESLEY TANSEY[1,a], CHRISTOPHER TOSH[2,b] AND DAVID M. BLEI[2,c]

[1]*Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center,* [a]*tanseyw@mskcc.org*

[2]*Data Science Institute, Columbia University,* [b]*ct2915@columbia.edu,* [c]*david.blei@columbia.edu*

Exploratory cancer drug studies test multiple tumor cell lines against multiple candidate drugs. The goal in each paired (cell line, drug) experiment is to map out the dose-response curve of the cell line as the dose level of the drug increases. We propose Bayesian tensor filtering (BTF), a hierarchical Bayesian model for dose-response modeling in multisample, multitreatment cancer drug studies. BTF uses low-dimensional embeddings to share statistical strength between similar drugs and similar cell lines. Structured shrinkage priors in BTF encourage smoothness in the dose-response curves while remaining adaptive to sharp jumps when the data call for it. We focus on a pair of cancer drug studies exhibiting a particular pathology in their experimental design, leading us to a nonconjugate monotone mixture-of-gammas likelihood. To perform posterior inference, we develop a variant of the elliptical slice sampling algorithm for sampling from linearly-constrained multivariate normal priors with nonconjugate likelihoods. In benchmarks, BTF outperforms state-of-the-art methods for covariance regression and dynamic Poisson matrix factorization. On the two cancer drug studies, BTF outperforms the current standard approach in biology and reveals potential new biomarkers of drug sensitivity in cancer. Code is available at https://github.com/tansey/functionalmf.

1. Introduction. To search for new therapeutics, biologists carry out exploratory studies of drugs. They test multiple drugs at different doses and against multiple biological samples. The goal is to trace the dose-response curves and to understand the efficacy of each drug.

This article concerns dose-response modeling in exploratory drug studies. In particular, we are concerned with studies where the experimental design makes it difficult to perform statistical inference on the resulting data. We consider two such studies, both involving anticancer drugs being tested in vitro on models of human tumors known as organoids (Drost and Clevers (2018)). A dose-response curve in the studies represents the expected cell survival rate (response) for a specific organoid as a function of the concentration (dose) of an anticancer drug. The studies differ primarily in their size. The first study is a small-scale pilot study conducted internally at Columbia University Medical Center with 35 drugs and 28 organoids; the second is a large-scale, "landscape" study conducted at Samsung Medical Center with 67 drugs and 284 organoids (Lee et al. (2018)). The experiments in each study are costly; each one can take weeks or months to conduct in the lab. Consequently, the exhaustive set of all (organoid, drug) combinations is not available. This leaves missing data, dose-response curves for which no data is available that must be imputed.

Figure 1 shows data from the landscape study. Each panel illustrates the interaction of one type of drug with one organoid sample. The gray points are the results of a set of experiments, each set with two replicates measured at seven different doses. The goal is to use the observations (gray points) to infer the true dose-response curves. The predictions—the orange lines and uncertainty bands—come from the Bayesian dose-response model we propose in this paper; each of the nine panels in Figure 1 were held out from the model at fitting time.
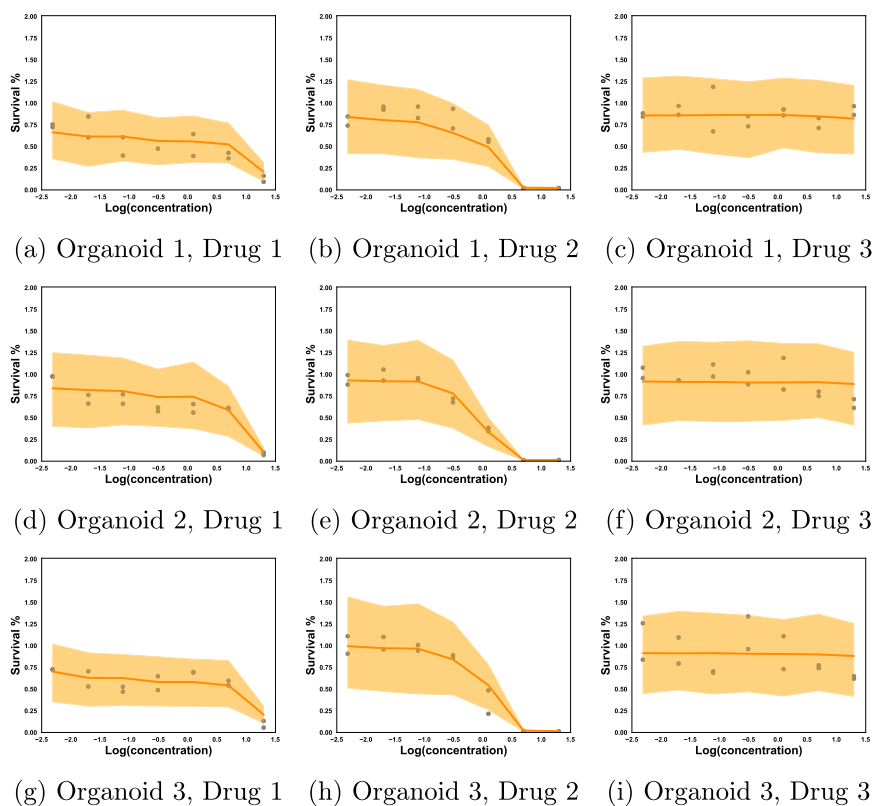
FIG. 1. *Sample of data from an organoid cancer drug experiment. Gray dots are observed outcomes*; *the orange line is the mean predicted response*; *bands represent* 50% *posterior predictive credible intervals for observations. All nine experiments were held out from the model at training time.*

Notice there is structure in the outcomes: each drug has similar effects on each organoid. Thus, we treat modeling of dose-response as a factorization problem. The structure in the data arises because organoids share latent molecular attributes, such as genomic mutations, and drugs share latent pharmaceutical attributes, such as chemical structures. In each experiment, organoid and drug attributes interact, creating the shared patterns of dose-response.

While traditional factorization considers a matrix of scalars, the entries of this matrix are latent dose-response curves subsampled at different doses. To model such curves, we model drug attributes as evolving with the dose level. While the effects usually vary smoothly between dose levels, there are occasional sharp jumps, such as between the final two dose levels of drug 1. Capturing latent structure in dose-response curves requires handling this type of nonstationarity.

The observation model for these experiments is also nonstandard. The measured outcome is a positive, real-valued measurement of cell survival relative to a noisy baseline. The model we propose uses a nonconjugate mixture-of-gamma-shapes likelihood with the latent dose-response entering through the scale parameter. This is reflected in Figure 1, where the uncertainty intervals shrink as the predicted survival rate drops.

As a final wrinkle, the drugs in these cancer studies are all cytotoxic, meaning that they will only kill cells, not facilitate growth. This biological prior knowledge implies the expected survival rate can only decrease as the dose increases. Cytotoxicity adds a shape constraint, requiring the dose-response curves are all monotonic. Further, since these drugs will not facilitate growth, effects in each curve are also upper bounded.

The dose-response model in this paper addresses these requirements. First, we propose Bayesian tensor filtering (BTF), a probabilistic method for smoothed tensor factorization.

BTF uses structured shrinkage priors that encourage smoothness between successive dose levels while simultaneously enabling sharp jumps when the data calls for it. Second, we develop generalized analytic slice sampling (GASS), a new MCMC inference algorithm that, when combined with BTF, enables the proposed dose-response model to support arbitrary likelihoods and linear inequality constraints on dose-response curves. The ability of GASS to handle linear constraints enables inference of monotone dose-response curves with upper and lower bounds.

The remainder of this paper is structured as follows. We first review related work on Bayesian dose-response modeling in Section 2. Section 3 then provides an overview of the two cancer drug studies and a motivation for the mixture-of-gamma-shapes likelihood. Section 4.2 presents Bayesian tensor filtering, a flexible model for smoothed tensor factorization. In Section 5 we detail generalized analytic slice sampling, a procedure for sampling from posteriors with constrained multivariate normal priors and nonconjugate likelihoods. Section 6 presents quantitative performance benchmarks for BTF, GASS, and the proposed dose-response model. Finally, in Section 7 the dose-response model is extended to handle side information in the form of molecular features about organoids. An analysis of the landscape study dataset with 115 features reveals potential new biomarkers of drug sensitivity in a subset of organoids.

2. Relevant literature. We survey a collection of the most relevant work to the proposed dose-response model. In each category of work, we focus on methods that enable uncertainty quantification, primarily through Bayesian inference.

Bayesian factor modeling. Many models have been developed for Bayesian factor analysis with smooth structure. Zhang and Paisley (2018) apply a group lasso penalty to the rows and columns of a matrix then derive a variational expectation maximization (EM) algorithm (Bishop (2006)) for inference. Hahn, He and Lopes (2018) use horseshoe priors (Carvalho, Polson and Scott (2010)) for sparse Bayesian factor analysis in causal inference scenarios with many instrumental variables. Kowal, Matteson and Ruppert (2019) develop a time series factor model using a Bayesian trend filtering prior (Faulkner and Minin (2018)) on top of a linear dynamical system with Pólya-gamma augmentation (Polson, Scott and Windle (2013)) for binomial observations. Schein, Wallach and Zhou (2016) develop Poisson-gamma dynamical systems (PGDS), a dynamic matrix factorization model specifically for Poisson-distributed observations; we compare BTF with a tensor extension of PGDS in Section 6. Unlike the above models, BTF is likelihood-agnostic through GASS inference and enables modeling of independently-evolving columns rather than a common time dimension.

Independent dose-response curve estimation. A number of authors have investigated Bayesian methods for modeling monotone dose-response curves. These are typically done through a mixture of monotone functions. Perron and Mengersen (2001) use a mixture of triangular distributions. Neelon and Dunson (2004) use an autoregressive mixture prior of truncated normals in a piecewise linear spline model. Bornkamp and Ickstadt (2009) propose a Bayesian nonparametric (BNP) model with a potentially-infinite mixture of two-sided power distributions. Shively, Sager and Walker (2009) also propose a BNP model which improves upon the model of Neelon and Dunson (2004) with the key idea being to model the mean of the monotone curve as the integral of a positive function. Ghebretinsae et al. (2013) present a Bayesian hierarchical model for nonnegative, real-valued comet assays with a gamma outcome model on the shape. Lin and Dunson (2014) propose a Gaussian process model with a posterior projection approach for shape-constrained curves. These methods all focus on the case of individual curve estimation. The datasets we consider here differ in that there are multiple samples and multiple drugs, with the goal to share statistical strength between samples and drugs both to denoise the existing curves and to predict drug effects on samples without that specific (sample, drug) pair yet tested.

Joint dose-response curve estimation.   A smaller body of work considers joint modeling of a set of dose-response curves. Both Vis et al. (2016) and Abbas-Aghababazadeh, Lu and Fridley (2019) use nonlinear mixed effect models for cancer drug response in a regression setting. Patel et al. (2012) take a Bayesian B-spline approach to dose-response surface modeling. Cai and Dunson (2007) propose a Bayesian spline model for multiple dependent isotonic response curves where latent factors account for per-sample response offsets. Fridley et al. (2009) propose a hierarchical Bayesian log-linear model for dose-response in cytotoxicity studies. Fox and Dunson (2015) consider the similar setting of covariance regression for influenza infection rates, imposing a sparse factorization on the covariance matrix that evolves over time in a Bayesian nonparametric setting. Wilson, Reif and Reich (2014) use monotone piecewise-linear splines in a hierarchical model of chemical toxicity assays, imposing a hierarchical Bayesian model that shrinks across similar molecules.

These models all have the common property that they shrink together samples via hierarchical priors, sharing statistical strength among rows to improve dose-response curve estimation. However, the hierarchical priors do not model any relational structure to shrink across samples and assays and do not provide any way to infer missing curves. Modeling the relational structure in multisample, multidrug studies is crucial for cancer drug studies, as often only a subset of samples have been tested for any given drug. Predictions about the missing curves can inform which experiments, among the many possible (sample, drug) combinations, show promise and should be carried out next. We discuss the deeper connections between BTF and both Fox and Dunson (2015) and Wilson, Reif and Reich (2014) after presenting the details of BTF; we also compare against Fox and Dunson (2015) with a Gaussian likelihood version of BTF in the benchmarks.

Predictive dose-response modeling.   In many experiments, descriptive features are gathered representing useful side information about the samples or assays. A natural approach in these scenarios is to build predictive models that map from features to dose-response curves, enabling out-of-sample prediction for untested (sample, drug) pairs. Low-Kam et al. (2015) proposed a Bayesian regression tree model with spline leaf nodes, enabling prediction of entire dose-response curves from chemical descriptors in nanoparticle experiments. Wheeler (2019) modeled dose-response with molecular descriptors via additive Gaussian process tensor products over a real-valued feature space. In the case of binary or discrete data, Wheeler (2019) first takes a principal components decomposition to project features to a continuous space, making feature interpretation difficult. Both methods also assume a Gaussian noise model with no shape constraints; dealing with nonconjugate likelihoods and shape constraints would require a novel inference scheme similar to the proposed GASS algorithm. Tansey et al. (2021) use deep neural networks to predict monotone dose-response curves from molecular features in an approximate Bayesian model but require a large dataset of experiments and features to train the neural network. More generally, predictive methods generally assume features to be available and complete. One notable exception is Moran et al. (2019) which focuses on a Bayesian multiview factorization approach for dose-response modeling in the single-assay setting. In the pilot dataset no features are present; in the landscape dataset many samples are missing feature information. In Section 7 we extend the dose-response model to include potentially-missing features via a multiview factorization approach.

3. Study design and dataset details.   We detail the specific protocol used for the internal pilot study at Columbia University Medical Center. The landscape study uses a different number of plates, wells per plate, concentrations, drugs per plate, and replications. These distinctions are changes to the dimensions of the resulting tensor of observations, but the fundamental statistical inference challenges remain the same.
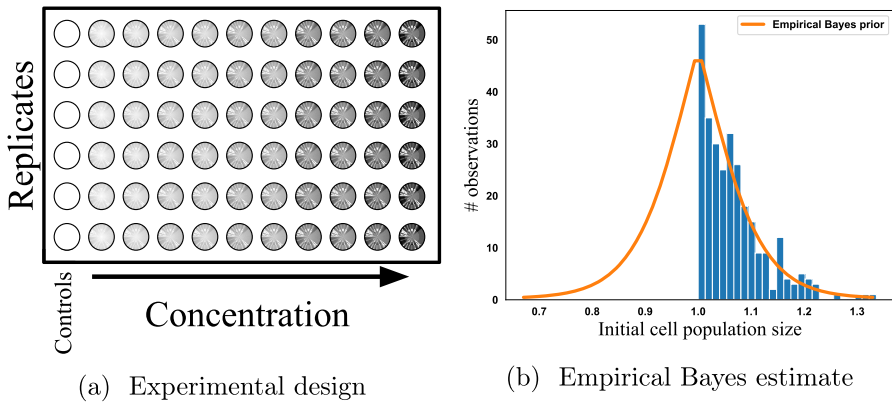
(a) Experimental design

(b) Empirical Bayes estimate

FIG. 2. *Left*: *The layout of each microwell plate experiment used to generate a single dose-response experiment. Cells are pipetted one column at a time, leading to correlated errors. Right*: *Estimate of the prior distribution of mean cell counts in each column relative to the control column mean. The prior is estimated empirically, assuming the lowest concentration had no effect if it had a higher mean.*

Each dose-response experiment is conducted on a microwell plate where a single drug is tested against a single biological sample. Each experiment measures a proxy for cell abundance 72 hours after applying the drug. Cell abundance is reported relative to a baseline control population where no drug was applied. For the control and each concentration level, six replicates are tested. Figure 2a shows the design of each 60-well plate experiment in the pilot study. The proxy measurement is a fluorescence assay; a fluorescent protein is added after 72 hours that binds to a molecule kept at near-constant levels in living cells. The degree of brightness of each microwell measures the relative abundance of cells alive but does not correspond to an exact cell count. All observations in the final dataset are normalized by dividing the brightness measurements in each microwell by the mean brightness for the control wells.

The details of how these organoid experiments were carried out in the wet lab are important, as they induce a particular form of correlated errors in the observations. The first step in each experiment is to pipette an initial population of cells into each of the 60 microwells on the plate. This is a time-consuming process for the biologist, often taking hours to pipette a single plate. To speed up the plating process, biologists use a multiheaded pipette that enables them to simultaneously fill an entire column of each plate. This reduces the burden on the biologist but comes at a cost—correlated errors.

When a biologist fills a microwell, they first draw a pool of cells into the pipette. Given the small volumes involved in laboratory experiments, the actual number of cells drawn can vary substantially on a relative basis. Using a multiheaded pipette transforms this variation into a hierarchical model: first, a pool of cells is drawn into the pipette, then it is split among all the heads. The majority of the variation comes in the initial sampling, with small noise added in the splitting process. This has the unintended side effect of creating correlated errors between all microwells in a single column. The drug concentrations are replicated along the same column, leading to a fundamentally unanswerable question: was the drug particularly effective at this dose, or was the initial sample of cells particularly small by chance?

In other cancer drug experiments, such as high-throughput cancer cell line experiments (Garnett et al. (2012), Ghandi et al. (2019), Shoemaker (2006)), wells are filled using designs that avoid such correlated errors. In particular, these experiments typically use an orthogonal design, pipetting the drug replicates along rows and cells along columns. High-throughput screens exhibit their own correlated error pathologies, often referred to as batch

effects. A number of techniques have been developed to preprocess high-throughput screening data and remove or limit batch effects (Johnson, Li and Rabinovic (2007), Lachmann et al. (2016), Leek et al. (2010), Mazoure, Nadon and Makarenkov (2017), Tansey et al. (2021)). These techniques account for the microenvironment similarities (e.g., temperature, humidity) that create spatial and temporal correlation between errors in wells nearby on the same plate or run in the lab on the same day, in high-throughput screens.

Unfortunately, batch effect correction techniques are not applicable here. The correlation between wells is due to multiheaded pipetting, not similarities in the plate microenvironment between temporally or spatially-related wells. Unlike smoothly-varying spatial batch effects, the experimental bias in the organoid datasets creates correlations between wells in the same column but does not vary smoothly between columns. Drug concentrations are also varied across columns but constant across rows. This makes it impossible to disentangle the drug effect from the pipetting error without any assumptions.

4. Generative dose-response model for cancer organoid studies.   To model the latent dose-response curves in the pilot and landscape studies, we make two high-level design decisions. First, the unidentifiable experimental design necessitates making some assumption about the dose-response curves. We take an empirical Bayes approach, leveraging the fact that most drugs are ineffective at the smallest dose level. After specifying the likelihood, a hierarchical structure is imposed to share statistical strength between similar organoids and similar drugs. We propose a Bayesian hierarchical model that imposes data-adaptive smoothness between successive doses and shares statistical strength via latent, low-dimensional embeddings.

4.1. *Smallest-dose assumption and empirical Bayes likelihood estimation.*   The correlated errors in the columns render the exact effects unidentifiable. Each column has two latent variables affecting the final population size of cells: a dose-level effect from the drug and an initial population size from the pipetting. Since both of these variables affect all replicates in a column, disentangling them precisely is impossible.

We take an empirical Bayes approach to disentangling the variation in drug effects from the technical error in pipetting. In most experiments the lowest concentration tested is too small to have any affect on cell survival. We, therefore, make the assumption that any experiment where the mean of the control replicates is lower than the mean of the replicates treated at the lowest concentration has effectively two sets of control columns. This enables estimation of the variation between means and an empirical Bayes prior for the pipetting error.

Specifically, we form a histogram of all lowest-concentration means greater than the control mean on the same plate. We then fit a Poisson GLM with three degrees of freedom to the histogram to estimate the prior probability that the mean of the initial population of cells was higher than the control mean. We assume the true distribution is symmetric and obtain an empirical Bayes prior on the means. Figure 2b shows the histogram and empirical Bayes prior estimate for the pilot study dataset. The within-column variance is identifiable and estimated, using the control replicates.

We integrate out the uncertainty in the initial population mean. For each organoid sample $i = 1, \ldots, N$ treated with drug $j = 1, \ldots, M$ at dose level $t = 1, \ldots, T$, with replicates $r = 1, \ldots, R$, this yields a gamma mixture model likelihood,

$$(1) \qquad P(y_{ijtr} \mid \mu_{ijt}) = \prod_{r=1}^{R} \left(\sum_{k=1}^{K} \hat{m}_k \, \mathrm{Ga}(y_{ijtr}; \hat{a}_k, \hat{b}_k \mu_{ijt}) \right) \mathbb{1}[0 \leq \mu_{ijt} \leq 1],$$

where $(\hat{m}, \hat{a}, \hat{b})_k$ are derived from the empirical Bayes procedure. The latent drug effect $\mu_{ijt} \in [0, 1]$ is the probability of a cell in organoid i surviving treatment with drug j at

dose level t. The drug effect enters in the gamma scale as the initial cell population size sampled from $\text{Ga}(y_{ijtr}; \hat{a}_k, \hat{b}_k)$ is being multiplied by μ_{ijt}. The drug effect is constrained to be a proportion, as the drugs are known not to help any cells grow (i.e., the proportion can be at most 1) and a drug cannot kill more than all of the cells.

4.2. *Bayesian tensor filtering for organoid dose-response modeling.* To capture shared structure between dose-response curves of similar organoid samples and similar drugs, we place smoothed factor model priors on the dose-response curve in a hierarchical Bayesian model we term Bayesian tensor filtering (BTF),

$$
\begin{aligned}
y_{ijtr} &\sim P(y_{ijtr} \mid \mu_{ijt})\mathbb{1}[\mu_{ijt} \leq \mu_{ij(t-1)}], \\
\mu_{ijt} &= w_i^\top v_{jt}, \\
w_i &\sim \mathcal{N}_D(\mathbf{0}, \sigma^2 I_D), \\
(\Delta^{(k)} V_j)_\ell &\sim \mathcal{N}_D(\mathbf{0}, \rho^2 \tau_{j\ell}^2 I_D), \\
\tau_{j\ell} &\sim \mathrm{C}^+(0, \phi_{j\ell}), \\
\phi_{j\ell} &\sim \mathrm{C}^+(0, 1), \\
\sigma^{-2} &\sim \mathrm{gamma}(0.1, 0.1), \\
\rho &\sim P(\rho).
\end{aligned}
$$

(2)

In the above model, $w_i, v_{jt} \in \mathbb{R}^D$ are the latent factors and loadings for each organoid i and drug-dose (j, t), respectively. The $T \times D$ drug loadings matrix V_j contains all $(v_{j1}, \ldots, v_{jT})$ loading vectors for drug j. The choice of the number of latent factors, D, is a hyperparameter. We will occasionally refer to factors and loadings as embeddings or attributes, and D as the embedding dimension.

The generative model in equation (2) incorporates a number of design decisions motivated by prior knowledge about biology and the nature of the experiments conducted. We explain the rest of the model in the context of these design decisions and the properties they induce in the resulting dose-response curves.

Monotonicity in the dose-response curve. All drugs in both organoid studies are cytotoxic, only inducing a higher rate of cell death as the dose increases. This monotonic dose-response relationship is encoded by the hard constraint at the top of equation (2). Each dose-response effect μ_{ijt} is required to be, at least, as toxic as the previous $\mu_{ij(t-1)}$ effect; we assume $\mu_{ij0} = 1$.

Latent attributes for biological samples. Organoid samples share molecular attributes. In cancer, different tumor samples contain similar patterns of genomic mutations, copy number alterations, and gene expression (Weinstein et al. (2013)). In mixed tissue experiments, cells that have differentiated into the same type will often respond similarly (e.g., Huang, Wu and Xing (2011)). These attributes are captured in BTF with a latent vector, $w_i \in \mathbb{R}^D$ for the ith sample, as in standard matrix factorization. For identifiability (West (2003)) we assume a lower-triangular structure on the factors matrix $W = (w_1, \ldots, w_N)$, though interpretable factors are not our primary goal here.

Independent dose-specific latent attributes for drugs. For each drug in the dataset, BTF models each dose level $t = 1, \ldots, T$ with its own embedding. For a single drug j, there are T embeddings forming a drug embedding matrix $V_j \in \mathbb{R}^{T \times D}$. In BTF, columns (drug effects) are evolving independently, though potentially with similar latent attributes. This column independence distinguishes BTF from time-series tensor factorization models (Gauvin,

Panisson and Cattuto (2014), Spiegel et al. (2011), Takeuchi, Kashima and Ueda (2017), Xiong et al. (2010)) where all columns are progressing through time together. Independent column evolution captures the notion that two drugs treated at the same concentration may have totally different effects due to the molecular size of the drug, its targeting receptor, and its chemical structure. Different drugs may also be treated at entirely different concentration levels, as they have different molecular properties that require higher or lower concentration ranges to map out the dose-response relationship. For ease of notation, we assume all drugs are treated at T concentrations; however, conceptually, drugs could be treated at different numbers T_j of concentrations.

Nonstationary group smoothness priors on drug attributes. We assume drug effects typically vary smoothly with dose, with occasional sharp jumps. To encode this, we place hierarchical smoothness priors on the differences between dose-specific drug embeddings. Specifically, we place a Bayesian group trend filtering prior on drug embeddings.

Trend filtering is an adaptive smoothing technique originally developed in the penalized regression case (Kim et al. (2009), Tibshirani (2014)). The penalized regression formulation places ℓ_1 penalties on the kth-order differences of neighboring points on a 1D grid. For instance, in the $k = 1$ case this is the total variation norm penalty,

$$(3) \qquad \underset{\beta \in \mathbb{R}^n}{\text{minimize}} \|Y - \beta\|_2^2 + \lambda \|\beta_{1:(n-1)} - \beta_{2:n}\|_1.$$

The solution to the convex optimization problem in equation (3) leads to piecewise constant plateaus in β due to the lasso penalty driving first differences to zero; higher order differences lead to piecewise-polynomial solutions (Tibshirani (2014)).

Faulkner and Minin (2018) extended trend filtering to the Bayesian context and considered a number of different priors on the kth-order differences. They consider three different priors on the differences: normal priors, equivalent to ℓ_2 Laplacian smoothing in the regression case; Laplace priors, the direct Bayesian analog of ℓ_1 lasso priors, and horseshoe priors (Carvalho, Polson and Scott (2010)), a heavier tailed distribution that does not suffer from the nondiminishing bias of the lasso (van der Pas, Kleijn and van der Vaart (2014)). The horseshoe priors are shown to perform best across a range of problems. We adapt the horseshoe prior to the group trend filtering case in BTF.

We call $\Delta^{(k)} \in \mathbb{R}^{L \times T}$ the composite trend filtering matrix; it contains all linear operators needed to encode the $(1, \ldots, k)$th-order differences. The ordinary trend filtering matrix encodes only the kth-order differences, implicitly assuming all lower-order differences are not smooth. For example, the $k = 2$ case yields a prior on the first and second order differences,

$$(4) \qquad \Delta^{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & \ldots & 0 & 0 & 0 \\ & & & & \ldots & & & \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 & -1 \\ 1 & -2 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \ldots & 0 & 0 & 0 \\ & & & & \ldots & & & \\ 0 & 0 & 0 & 0 & \ldots & 1 & -2 & 1 \end{bmatrix}.$$

The first line of equation (4) places an independent prior on the embedding vector for the first dose level in each drug, v_{j1}, making the matrix $(\Delta^\top \mathcal{T} \Delta)$ nonsingular, where $\mathcal{T} = \text{diag}(1/(\rho^2 \tau_j^2))$. This ensures the resulting prior on V_j is proper; see the Supplementary Material (Tansey, Tosh and Blei (2022)) for details.

We impose a group trend-filtering prior on the drug effects by placing a group horseshoe prior on the ℓth row of the $(\Delta^{(k)} V_j)$ differences matrix. To do this, in equation (2) we

adapt the Bayesian formulation of the group lasso (Kyung et al. (2010)) to the global-local shrinkage view of the horseshoe prior (Polson and Scott (2010)). Each row $(\Delta^{(k)} V_j)_\ell$ in the differences matrix has both a local $\tau_{j\ell}^2$ variance and a global ρ^2 variance term. Small values of ρ^2 and $\tau_{j\ell}^2$ will shrink the ℓth difference vector to nearly zero, resulting in the curve being smoother; larger values enable the curve to jump in response to the data.

Following Bhadra et al. (2017), we place a half-Cauchy prior on $\phi_{j\ell}$, the scale term in the local horseshoe shrinkage prior; this is referred to as the horseshoe+ prior. A full Bayesian specification could choose a reasonable prior for ρ, such as a standard Cauchy or Uniform$(0, 1)$. If an estimate of the number of nonzero entries is available, van der Pas, Kleijn and van der Vaart (2014) make an asymptotic argument for setting $\hat{\rho}$ to the expected number of nonzeros. We find BTF is robust to the choice of global shrinkage parameter; we default to a half-Cauchy prior on ρ in our implementation and also support performing a grid search over a range of discrete ρ values via deviance information criteria (Celeux et al. (2006)). The value k in $\Delta^{(k)}$ is left as a hyperparameter; we suggest $k = 2$ as a reasonable default choice for most datasets.

4.3. *Deeper connections to related work.* The BTF model is closely related to two works: the Bayesian nonparametric covariance regression model of Fox and Dunson (2015) and the zero-inflated piecewise log-logistic dose-response model of Wilson, Reif and Reich (2014). Before proceeding to inference in BTF, we first provide a detailed discussion of the deeper connections to these methods.

Bayesian covariance regression. Fox and Dunson (2015) introduced Bayesian nonparametric covariance regression (BNP-CovReg). The BNP-CovReg model poses a Gaussian noise model for a set of p curves observed at n points,

$$(5) \qquad \mathbf{y}_i = \boldsymbol{\mu}(\mathbf{x}_i) + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}_p(0, \Sigma), i = 1, \ldots, n,$$

where $\mathbf{y}_i = \log \mathbf{r}_i$, the vector of observations in the p curves at point x_i, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$. A factor model is imposed on the latent mean,

$$(6) \qquad \mathbf{y}_i = \boldsymbol{\Lambda}(\mathbf{x}_i)\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\eta}_i \sim \mathcal{N}_k\big(\boldsymbol{\psi}(\mathbf{x}_i), I_k\big),$$

where $\boldsymbol{\Lambda}(\mathbf{x}_i)$ are the factor loadings at point $\mathbf{x}_i$ and $\boldsymbol{\eta}_i$ the latent factors associated with observation $\mathbf{y}_i$. Here, $k \ll p$ imposes a low-rank assumption on the factor model, and independent Gaussian process priors are placed on each ψ_h, for $h = 1, \ldots, k$ with squared exponential kernel. Independent conjugate inverse-gamma priors are placed on each σ_i.

For computational feasibility the factor loadings matrix $\boldsymbol{\Lambda}(\mathbf{x})$ is expressed as a weighted combination of a smaller set of L basis functions,

$$(7) \qquad \boldsymbol{\Lambda}(\mathbf{x}) = \boldsymbol{\Theta}\boldsymbol{\xi}(\mathbf{x}),$$

where $\boldsymbol{\Theta}$ is a $p \times L$ matrix of coefficients and $\boldsymbol{\xi}(\mathbf{x})$ an $L \times k$ array of basis functions. A global-local shrinkage prior is placed on the elements of $\boldsymbol{\Theta}$ to effectively reduce the dimension of the basis to much smaller than L or k.

A number of follow-up works have investigated similar models. Kunihama, Halpern and Herring (2019) extend BNP-CovReg to longitudinal data with covariate information. Li et al. (2019) use fixed factors, sacrificing the flexibility of the nonparametric approach of Fox and Dunson (2015) for increased scalability. Heaukulani and van der Wilk (2019) derive a variational inference approach for inverse Wishart processes that leads to a scalable approximate inference scheme for BNP-CovReg.

In the Gaussian likelihood case, BTF also uses a low-dimensional factor model for the response mean. However, rather than assuming independent sparsity and smoothness priors

on a set of basis coefficients and latent factors, BTF imposes smoothness directly on the curves. This translates to a *group* smoothness assumption on the latent factors, enforcing that the kth-order differences in successive latent means be shrunk to zero. As we show in Section 6.1, BTF outperforms the BNP-CovReg model on the same dataset that motivated the design of BNP-CovReg. It also enjoys substantially faster computational times (around $10\times$ faster on a 2018 MacBook Pro) and trivial parallelization in the Gibbs sampler if further scalability is needed. Furthermore, BTF is extensible to a number of other likelihoods, including nonconjugate models with linear constraints like monotonicity, through the GASS inference algorithm.

Piecewise log-logistic, monotone dose-response modeling. Wilson, Reif and Reich (2014) introduced the zero-inflated piecewise log-logistic (ZIPLL) model for estimating dose-response curves in chemical toxicity assays. Similar to the cancer drug study scenario, the ZIPLL model is explicitly designed for multisample, multiassay studies, where observations form a tensor with concentration as a third dimension. The ZIPLL model assumes a Gaussian noise distribution on observations y_{ijt} of sample i, assay j, concentration x_{ijt}, respectively,

$$(8) \qquad y_{ijt} = f_{ij}(x_{ijt}) + \epsilon_{ijt}, \quad \epsilon_{ijt} \sim \mathcal{N}(0, \sigma^2).$$

In the toxicity study considered, the minimum and maximum values corresponding to no-effect and total toxicity are unknown and modeled as latent variables,

$$(9) \qquad f(x_{ijt}) = \begin{cases} u_{ij} - (u_{ij} - l_{ij}) \times \text{Logistic}(g(x_{ijt}; a_{ij}, \mathbf{w}_{ij})) & \text{if } Z_{ij} = 1, \\ l_{ij} & \text{if } Z_{ij} = 0, \end{cases}$$

where $u_{i}j$ is the maximum response, $l_{i}j$ is the minimum response, $a_{i}j$ is the location parameter of the curve (the 50% survival point), and $\mathbf{w}_{ij}$ are the shape parameters. The Z_{ij} variable captures the zero-inflated property of the chemicals in the toxicity dataset, where sparsity in effects is expected. The ZIPLL model uses a monotonic log-linear spline basis to model $g(x_{ijt}; a_{ij}, \mathbf{w}_{ij})$ using a fixed symmetric grid of internal knots. Multivariate normal priors are placed on $\boldsymbol{\theta}_{ij} = \log(l_{ij}, u_{ij}, a_{ij})$ and a Gaussian autoregressive prior is placed on $\log \mathbf{w}_{ij}$ to encourage smoothness in the dose-response curve. Shrinkage in ZIPLL is performed across samples in the same assay via a hierarchical prior,

$$(10) \qquad \boldsymbol{\theta}_{ij} \mid \boldsymbol{\mu}_j, \Sigma_j \sim \mathcal{N}_3(\boldsymbol{\mu}_j, \Sigma_j), \quad \boldsymbol{\mu}_j \sim \mathcal{N}_3(\boldsymbol{\mu}, \Sigma),$$

where hard-coded hyperparameters are used for conjugate priors on $\boldsymbol{\mu}$, Σ_j, and Σ as well as the other hyperparameters in the hierarchical model.

The ZIPLL model is closely related to the BTF dose-response model in both modeling constraints and target application domain. The use of a monotone basis with Gaussian priors on log-transformed latent parameters ensures the posterior curves satisfy the monotonicity constraints in cancer dose-response modeling and the logistic transform maps to the $[0, 1]$ interval as well. However, the Gaussian noise model is also misaligned with the cancer drug studies likelihood, and the stationary smoothness assumptions of the Gaussian autoregressive prior may not handle to sharp jumps found in the data.

Moreover, ZIPLL only pools statistical strength across samples (rows) in the dose-response tensor but not assays (columns); this prevents imputing assays that are missing entirely. Extending the comparatively simple scalar basis coefficients $\mathbf{w}_{ijl}$ with autoregressive priors to a factor model with equivalent smoothness would be nontrivial, as posterior inference is nonconjugate and may require a novel approximate inference scheme. These are similar challenges to the inference in the BTF dose-response model, all of which led us to the development of the GASS algorithm in Section 5. The ability to impute out of sample experiments is critical for the cancer organoid drug studies, where only a subset of samples are tested for each drug.

5. Posterior inference. Posterior inference in BTF is performed through Gibbs sampling. In its most basic form, Gibbs sampling requires us to sample from the conditional distributions for each of the parameters. The updates for the latent attributes W and V depend on the form of likelihood, $P(y_{ijtr}; w_i^\top v_{jt})$. The derivations for Gaussian and binomial likelihoods as well as the horseshoe parameter updates are provided in the Supplementary Material (Tansey, Tosh and Blei (2022)). Here, we focus on the crux of the posterior inference challenge in BTF—the constrained, nonconjugate likelihood in equation (1).

To sample from the complete conditional for each of the latent attributes, we run MCMC-within-Gibbs, running a separate Markov chain whenever a Gibbs step requires a sample from the conditional distributions of the latent attributes w_i and v_{jt}. The challenge to this step is that the likelihood imposes hard constraints on the values of entries. The inner product must be a probability, requiring $w_i^\top v_{jt} \in [0, 1]$. Each inner product must also be no greater than the previous inner product, requiring $w_i^\top (v_{jt} \leq v_{j(t-1)}) \leq 0$. Running a naive MCMC algorithm such as Metropolis–Hastings within the Gibbs sampler is likely to have a high rejection rate and lead to poor mixing. Instead, we develop an MCMC algorithm that is capable of directly handling generic likelihoods and arbitrary linear constraints.

5.1. *Generalized analytic slice sampling.* Sampling from the conditional distributions of the latent attributes can be reduced to the problem of sampling from the posterior of a vector x with a multivariate normal prior constrained by a set of linear inequalities,

$$(11) \qquad x \sim P(y; x) \, \mathrm{MVN}(x; \mu, \Sigma) \mathbb{I}[Dx \geq \gamma].$$

Note in equation (11) we are describing a generic problem that is distinct from equation (2). The variables $(y, x, \mu, \Sigma, D, \gamma)$ in equation (11) are correspondingly generic variables that are distinct from those used in equation (2).

A natural candidate for sampling from such a distribution is elliptical slice sampling (Murray, Adams and MacKay (2010)), an empirically successful exact MCMC method for sampling from distributions with arbitrary likelihoods and multivariate normal priors. Elliptical slice sampling builds on the idea of slice sampling (Neal (2003)), an MCMC algorithm that, given a point x, samples a new point x' by first drawing a value u uniformly over the range 0 up to the likehood of x, and then by drawing x' uniformly from the set of points whose likelihood is, at least, u. Computing the set of points whose likelihood is above a certain threshold is infeasible in general, and thus some form of rejection sampling is required. Unfortunately, in high dimensions these rejection rates will be very large.

However, when we have a multivariate normal prior, we may utilize the fact that the contours of equal probability on the multivariate normal distribution are elliptical regions. Elliptical slice sampling exploits this observation by sampling a point v from the prior distribution and computing the ellipse $\{x \cos(\theta) + v \sin(\theta) : \theta \in [-\pi, \pi]\}$ containing both x and v. Then, as in slice sampling, it samples a likelihood u and then performs a form of rejection sampling to sample x' uniformly from the set of points *on the ellipse* whose likelihood is at least u. Note that when the likelihood is reasonably smooth, there will always be a reasonably large interval of points around x on the interval above this likelihood, and we will not have too many rejections. However, when the likelihood has hard constraints, these regions can be very small and lead to high rejection rates.

To address this, we extend elliptical slice sampling to directly handle constrained multivariate normal priors. The approach, which we call generalized analytic slice sampling (GASS), is a natural extension of the analytic slice sampling procedure of Fagan, Bhandari and Cunningham (2016) for truncated multivariate normals. The key difference is that the original analytic slice sampler only considered centered truncated multivariate normals with no likelihood component. Generalizing this procedure to handle the more general case in equation (11) requires handling several edge cases.

Algorithm 1: Generalized analytic slice sampling (GASS) for constrained MVN priors

Data: Valid current point x, mean μ, covariance Σ, log-likelihood $\mathcal{L}$, constraints (D, γ)
Result: MCMC sample from $P(x') \propto \exp(\mathcal{L}(x')) \text{MVN}(x'; \mu, \Sigma) \mathbb{I}[Dx' \geq \gamma]$
$t = \mathcal{L}(x) + \log \epsilon, \epsilon \sim U(0, 1)$;
Sample proposal $v \sim \text{MVN}(v; \mathbf{0}, \Sigma)$;
Grid approximation $\mathcal{G} = \text{grid}(-\pi, \pi)$;
foreach *constraint* $(d_i, \gamma_i) \in (D, \gamma)$ **do**
    $a = d_i^\top (x - \mu), b = d_i^\top v, c = \gamma_i - d_i^\top \mu$;
    if $a^2 + b^2 - c^2 \geq 0$ *and* $a \neq -c$ **then**
        Get θ_1, θ_2 as in equation (12);
        if $a^2 > c^2$ **then**
            $\mathcal{G} = \mathcal{G} \cap [\theta_1, \theta_2]$;
        else
            $\mathcal{G} = \mathcal{G} \cap ([-\pi, \theta_1] \cup [\theta_2, \pi])$;
        end
    end
end
Generate candidate samples $\mathcal{X} = \{x': x \cos(\theta_g) + v \sin(\theta_g) + \mu, \theta_g \in \mathcal{G}\}$;
Select uniformly from sufficiently likely candidates $\{x': \mathcal{L}(x') \geq t, x' \in \mathcal{X}\}$.

5.2. *Algorithm.* The full GASS procedure is presented in Algorithm 1. The idea of GASS is to note that the constraints can be pushed inside the proposal update. Namely, given an ellipse and a set of linear constraints, it is relatively straightforward to exactly compute their intersection and thus to restrict proposals to that region.

To see why, consider the case where we have a single linear constraint requiring that the output point satisfies $d^\top x' \geq \gamma$. Then, a valid angle θ must satisfy $a \cos \theta + b \sin \theta - c \geq 0$, where $a = d^\top (x - \mu)$, $b = d^\top (v - \mu)$, and $c = \gamma - d^\top \mu$. Basic trigonometry implies that the feasible range of θ is a subset of $[-\pi, \pi]$ whose boundary points are given by

$$(12) \qquad \theta_1, \theta_2 = 2 \arctan\left(\frac{b \pm \sqrt{a^2 + b^2 - c^2}}{a + c}\right).$$

There are two edge cases where the entire ellipse is valid: (i) $(a^2 + b^2 - c^2) < 0$ and (ii) $a = -c$. In the first case, we trivially have $a^2 + b^2 < c^2$ and, therefore, $a \cos \theta + b \sin \theta > c$ for all θ. In the second case, the only place the constraint boundary intersects the ellipse exactly at a single point of the ellipse, and thus its selection has probability zero. For all other cases the subset is determined based on the sign of $a^2 - c^2$. A positive sign indicates the quadratic in the inequality is concave and equation (12) defines the boundaries of a contiguous region; a negative sign indicates convexity and thus the complement of the interval.

When there are many linear constraints, we can solve for the valid regions of each of the individual constraints separately and then take their intersection. Finally, after computing the region of valid θ's, we approximate it with a fine-grained 1D grid. We draw a likelihood value u and filter out all the grid points with likelihood smaller than u. Finally, we draw a sample uniformly over the remaining grid points. An illustration of the algorithm is given in Figure 3.

It is not difficult to show that GASS is a valid Markov chain which will converge to the distribution given in equation (11). For completeness a proof of this convergence is provided in the Supplementary Material (Tansey, Tosh and Blei (2022)).
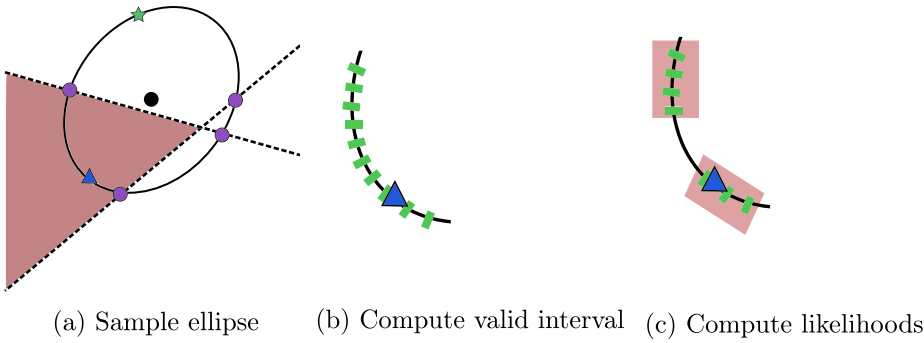
(a) Sample ellipse   (b) Compute valid interval (c) Compute likelihoods

FIG. 3. *An illustration of one step of the GASS chain starting from the blue triangle. (a) First, the green star is sampled form the multivariate normal centered at the black circle. The green star and the blue triangle determine an ellipse around the the black circle. Here, the dashed lines denote the linear constraints, and the red region is the feasible region. (b) Then, the intersection of the feasible region with the ellipse is computed and gridded. (c) Finally, a likelihood u is sampled, and those points that have likelihood at least u are retained, denoted by the red region. The next state is randomly selected from the remaining grid points.*

5.3. *Conditioning heuristic.* Elliptical slice sampling schemes like GASS can suffer from poor mixing when the likelihood overwhelms the multivariate normal prior. In such settings the sampled ellipses, which are generated with respect the prior and not the likelihood, may only have a small region centered around the current sample with likelihood comparable with the current likelihood, causing the chain to take only very small steps. Motivated by this observation, Fagan, Bhandari and Cunningham (2016) suggest performing expectation propagation (Minka (2001)) to better align the prior with likelihood. Unfortunately, in the context of BTF this is impractical as the prior parameters for W are a function of V, and vice versa, which would require us to perform expectation propagation every iteration of the Gibbs sampler.

We instead approximate the entire tensor once at the start by a nonnegative, monotone tensor factorization. To do this, we find an approximate solution to the optimization problem,

$$\hat{W}, \hat{V} = \underset{W, V}{\text{minimize}} \sum_{ijtr} (y_{ijtr} - w_i^\top v_{jt})^2$$

(13)

$$\text{subject to } 0 \le w_i^\top v_{jt} \le 1,$$

$$w_i^\top (v_{jt} - v_{j(t-1)}) \le 0.$$

The approximation solution to equation (13) is found via alternating constrained minimization for the rows and columns; we run the alternating minimization procedure until convergence.

After fitting the rows and columns, we calculate an overestimate of the variance, analogous to an EP approximation, as a multiple of the empirical squared error in the estimate for each column and row,

$$\hat{\mu}_{ijt} = \hat{w}_i^\top \hat{v}_{jt},$$

(14)

$$\hat{s} = \frac{\sum_{ijtr} (y_{ijtr} - \hat{\mu}_{ijt})^2}{N \times M \times T \times R},$$

$$\hat{\Sigma} = c\hat{s}I,$$

where $c \ge 1$ is a hyperparameter. This overestimates the empirical variance, accounting for a wider range of possible samples to correct for the error in the NMF procedure. The main

sensitivity of BTF inference is initialization and conditioning with an accurate mean; the co-variance is less important. A reasonably accurate $\hat{\mu}$ yields the majority of the gains, whereas the method is insensitive to reasonable choices of c. As we discuss in the benchmarks, the NMF model produces a reasonably accurate $\hat{\mu}$ in terms of RMSE and serves as a good starting point for the BTF dose-response model; for all BTF benchmarks we set $c = 3$.

The BTF dose-response model uses the conditioning heuristic in the W and V steps in the Gibbs sampler to calculate an adjusted prior. The log-likelihood used in the GASS procedure is then the original log-likelihood minus the log-conditioning likelihood, leaving the resulting distribution equivalent but better aligning the prior and likelihood.

6. Benchmarks and performance comparisons. We study the performance of the proposed dose-response model and its components, BTF and GASS. We first benchmark GASS against different alternative methods for nonconjugate inference, where GASS mixes faster and has lower error. Then, we study BTF on a dynamic matrix factorization problem with nonconjugate Poisson observations; BTF outperforms a recent Bayesian tensor decomposition approach designed for time-evolving count matrices. Finally, we apply the dose-response model to a real cancer drug study. We run five independent trials, holding out a different subset of entire dose-response curves and report averages over all trials; the BTF-based dose-response model outperforms all baselines in terms of log-probability on held out data.

6.1. *Gaussian BTF benchmarks.* We benchmark the conjugate Gaussian-likelihood BTF model on the Google Flu Trends dataset[1] modeled in Fox and Dunson (2015). This dataset contains weekly influenza-like infection (ILI) counts from 183 different regions in the United States from 1996 to 2014. The set of regions contains nested information, including both major cities and entire states. For benchmarking purposes we focus only on the 50 states to ensure that held out data is not leaked in through other nested regions; this makes the inference task strictly more difficult.

Fox and Dunson (2015) model the weekly log-count of infections with the Gaussian noise model in equations (5)–(6), where $\mathbf{y}_i = \log \mathbf{r}_i$ is the vector of log Google-estimated ILI rates in the 50 states at time x_i.

We compare against the performance of the Bayesian nonparametric covariance regression (BNP-CovReg) model of Fox and Dunson (2015), designed specifically for the Google Flu Trends data. For BNP-CovReg we keep the same hyperparameter settings with truncation parameters $\bar{L} = 10$ and $\bar{k} = 20$; we use the reference implementation provided by the authors. For Gaussian BTF we initialize the global shrinkage parameter ρ^2 to 0.1 and sample it with a HS+ prior; we also place an weakly-informative inverse-gamma(0.1, 0.1) prior on the likelihood variance. To measure performance, we hold out 10% of all years, selected uniformly at random across all available state-years. Model performance is measured in both root mean squared error (RMSE) and mean absolute error (MAE) on held-out data.

Table 1 shows the results for both the BNP-CovReg model and Gaussian BTF with $d = 2$, 5, and 10 latent factors. The BTF method outperforms BNP-CovReg in each case. The model performs best out of sample with $d = 5$ latent factors, suggesting it overfits with $d = 10$ factors. The BTF model also has good coverage, with 95% credible intervals having 95.83% coverage on in-sample data and 92.82% coverage for out-of-sample data with $d = 5$ factors.

Figure 4 shows an example of a single state (Alabama) comparing BNP-CovReg, and $d = 5$ BTF. The BNP-CovReg model over-smooths, leading it to underestimate large peaks both in- and out-of-sample. The BTF model closely tracks the data, even in out-of-sample predicted weeks, suggesting it has learned latent structure between the states.

---

[1]http://www.google.org/flutrends/

*Posterior mean results on the Google Flu Trends dataset. The Gaussian BTF model
outperforms the BNP-CovReg model of Fox and Dunson (2015) with as few as two
latent factors. MAE: Mean absolute deviation from data; RMSE: Root
mean-squared error from data*

| | Google Flu Trends | | | |
| | In-sample | | Out-sample | |
Model	MAE	RMSE	MAE	RMSE
BNP-CovReg	0.31	0.41	0.31	0.42
Gaussian BTF ($d = 2$)	0.15	0.22	0.15	0.21
Gaussian BTF ($d = 5$)	0.10	0.14	**0.12**	**0.16**
Gaussian BTF ($d = 10$)	**0.07**	**0.10**	0.13	0.21

6.2. *GASS benchmarks.* We benchmark GASS on a simulation study with a constrained multivariate normal prior with a nonconjugate gamma scale likelihood,

$$y_i^{(r)} \sim \text{gamma}(y_i; a, \theta_i),$$

$$\boldsymbol{\theta} \sim \text{MVN}(\boldsymbol{\theta}; \boldsymbol{\mu}, \Sigma)\mathbb{1}[0 \leq \boldsymbol{\theta} \leq 1] \prod_{i=1}^{n-1} \mathbb{1}[\theta_i \geq \theta_{i+1}],$$

(15)

$$\boldsymbol{\mu} = [0.95, 0.8, 0.75, 0.5, 0.29, 0.2, 0.17, 0.15, 0.01, 0.0001],$$

$$\Sigma_{ij} = \tau \exp\left(-\frac{1}{2b}(i - j)^2\right).$$

The covariance matrix in the unconstrained prior corresponds to a squared exponential kernel. We set the hyperparameters $a = 100$, $\tau = 0.1$, and $b = 3$; all hyperparameters are assumed known. We use $R = 3$ replicates for **y**. We compare GASS against four different variants of elliptical slice sampling (ESS):

- *Rejection sampling (RS).* Samples are drawn using the unconstrained ESS model, with the constraints pushed into the likelihood. Any violated constraint generates a zero probability and corresponds to a rejection sampler.

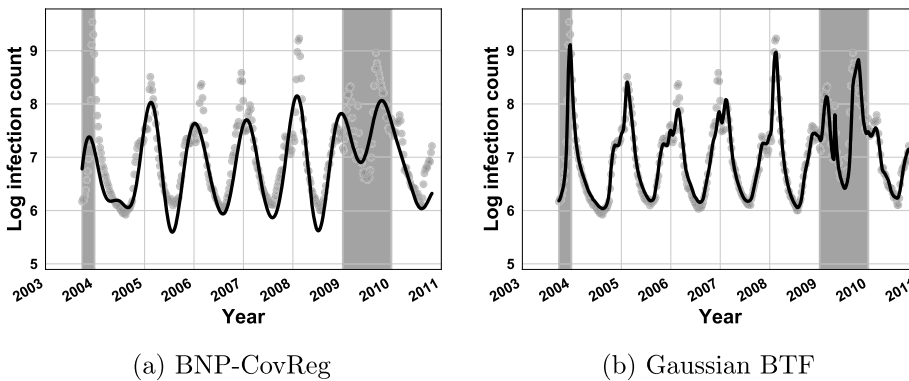

(a) BNP-CovReg   (b) Gaussian BTF

FIG. 4. *Example fits for BNP-CovReg and Gaussian BTF ($d = 5$) on a single state (Alabama) in the Google Flu Trends dataset. The gray shaded regions represent held out periods. BNP-CovReg oversmooths, failing to capture large peaks and short movements. BTF fits the data tightly both on in-sample examples and imputed held out periods, suggesting it has captured latent structure between states.*
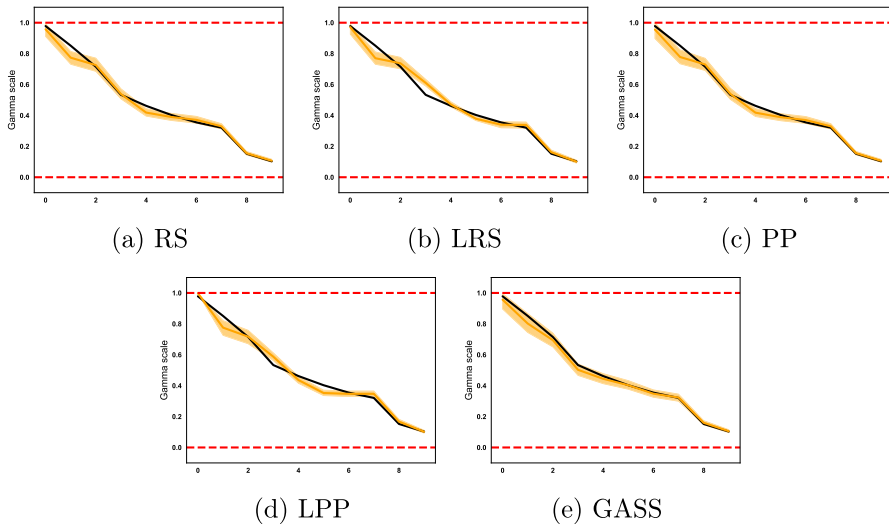
FIG. 5. *Sample fits for different methods on the gamma scale estimation benchmark; black is the true scale, orange is the estimated scale, bands are* 90% *credible intervals, dashed lines are constraint boundaries. GASS captures the shape of the curve and has good coverage after* 10K *Gibbs iterations.*

- *Logistic rejection sampling (LRS).* The ESS is used to model logits which are then passed through the logistic transform to satisfy the [0, 1] constraint; rejection sampling again handles the monotonicity constraint.
- *Posterior projections (PP).* No constraints are imposed on the model during posterior inference. Instead, we use the posterior projection approach of Lin and Dunson (2014) to post hoc enforce the constraints.
- *Logistic posterior projections (LPP).* A hybrid of the previous two approaches combined: modeling the logits for [0, 1] constraints but projecting the posterior samples to the monotone surface.

All models use the true prior mean and covariance; for the logistic models we empirically estimate the covariance of the logit-transformed θ. We compare performance with $2m$ MCMC steps, where the first m are a burn-in phase and the last m are used for posterior approximation; we consider $m = [100, 500, 1000, 5000, 10{,}000]$. Performance is measured in terms of mean squared error (MSE) and coverage rate of the 90% credible intervals for every θ_i point. Results are averaged over 100 independent trials with $(\boldsymbol{\theta}, \mathbf{y})$ resampled from equation (15) at the start of each trial.

Figure 5 shows examples of the fits for each method with 90% credible intervals. GASS is the only procedure that results in good coverage of the true mean and captures the shape of the overall curve after 20K MCMC steps; the other methods tend to oversmooth the curve and underestimate the uncertainty.

Table 2 shows the aggregate results of the benchmarks. GASS outperforms all four comparison methods in terms of both error and coverage. After $m = 100$ samples, the MSE for GASS is lower and coverage is higher than any of the other strategies after $m = 10{,}000$ samples. Further, the model appears to have almost fully mixed after 5000 samples with the coverage rate close to 90%.

6.3. *Nonstationary Poisson dynamical systems.* We benchmark BTF on a synthetic Poisson tensor dataset where the observations are Poisson distributed with a latent rate curve for each function. The rate at every point in the curve is the inner product of two gamma random

Benchmark performance for GASS vs. alternative nonconjugate elliptical sampling approaches. Results are averages over 100 *independent trials* $\pm$ *standard error*

	MSE $(\times 10^3)$				
Sampler	$m = 100$	$m = 500$	$m = 1000$	$m = 5000$	$m = 10000$
RS	1.54 ± 0.09	1.43 ± 0.09	1.43 ± 0.09	1.41 ± 0.09	1.41 ± 0.09
LRS	1.76 ± 0.14	1.75 ± 0.14	1.73 ± 0.14	1.66 ± 0.13	1.63 ± 0.13
PP	1.35 ± 0.09	1.31 ± 0.08	1.29 ± 0.09	1.29 ± 0.09	1.28 ± 0.08
LPP	1.66 ± 0.12	1.62 ± 0.12	1.56 ± 0.12	1.52 ± 0.13	1.50 ± 0.12
GASS	$\mathbf{0.74 \pm 0.05}$	$\mathbf{0.66 \pm 0.04}$	$\mathbf{0.63 \pm 0.04}$	$\mathbf{0.52 \pm 0.03}$	$\mathbf{0.49 \pm 0.03}$

	90% Credible Interval Coverage				
Sampler	$m = 100$	$m = 500$	$m = 1000$	$m = 5000$	$m = 10000$
RS	0.37 ± 0.02	0.47 ± 0.02	0.49 ± 0.02	0.50 ± 0.02	0.52 ± 0.02
LRS	0.23 ± 0.01	0.34 ± 0.02	0.36 ± 0.02	0.46 ± 0.02	0.48 ± 0.02
PP	0.44 ± 0.02	0.54 ± 0.02	0.56 ± 0.02	0.57 ± 0.02	0.58 ± 0.02
LPP	0.30 ± 0.01	0.40 ± 0.02	0.44 ± 0.02	0.53 ± 0.01	0.53 ± 0.02
GASS	$\mathbf{0.58 \pm 0.02}$	$\mathbf{0.73 \pm 0.02}$	$\mathbf{0.77 \pm 0.02}$	$\mathbf{0.86 \pm 0.01}$	$\mathbf{0.87 \pm 0.01}$

vectors,

$$h_{j\ell} \sim \text{Bern}(0.2), \qquad u_{j\ell d} \sim (1 - h_{j\ell})\delta_0 + h_{j\ell}\,\text{Ga}(1, 1), \qquad v_{jtd} = \sum_{\ell=1}^{t} u_{j\ell d},$$

$$w_{id} \sim \text{Ga}(1, 1), \qquad y_{ijt} \sim \text{Pois}(\langle w_i, v_{jt} \rangle).$$

The resulting true rates form a monotonic curve of constant plateaus with occasional jumps. As in the dose-response data, the columns evolve independently of each other, rather than through a common time parameter. We set the latent factor dimension to 3.

We compare a Poisson likelihood version of BTF with GASS inference (Poisson BTF) to nonnegative matrix factorization (NMF), the Poisson-gamma dynamical system (PGDS) model of Schein, Wallach and Zhou (2016), and a negative binomial likelihood version of BTF (NBinom BTF) with Pólya-gamma augmentation (Polson, Scott and Windle (2013)). For PGDS we contacted the authors who suggested we try three different values of the hyperparameter $\tau = (0.25, 0.5, 1)$. For NBinom BTF we use MCMC-within-Gibbs and sample the latent rate parameter with 30 steps of random walk Metropolis–Hastings for every Gibbs step. For Poisson BTF we initialize with NMF and did not use any conditioning heuristic.

We run all models for 5000 burn-in iterations and collect 5000 samples on an $11 \times 12 \times 20$ tensor with the upper left $3 \times 3 \times 20$ corner held out. We conduct five independent trials, regenerating new data each time and evaluating the models on the held out data and true latent rate. We measure performance in three categories of metrics: (i) mean absolute error (MAE), root mean squared error (RMSE), and negative log-likelihood (NLL) on held out observations; (ii) MAE and RMSE on the true latent rate, and (iii) posterior credible interval coverage of the true rate at 50%, 75%, and 90% targets. We evaluate all models at embedding dimensions $d = (2, 3, 5, 10)$ to compare sensitivity to the common hyperparameter.

Table 3 presents the results. The Poisson BTF model performs similarly across the range of dimension embeddings, whereas the other models are more sensitive. The NMF model generally performs better with a smaller embedding dimension while the PGDS model performs better with larger dimensions. In the case of NMF, this is due to overfitting without

TABLE 3
Mean results on the Poisson dynamical system benchmark; smaller is better for all metrics. NMF: Nonnegative matrix factorization; PGDS(τ): Poisson-gamma dynamical system with hyperparameter τ; NBinom-BTF: Bayesian tensor filtering with conditionally-conjugate negative binomial likelihood; Poisson-BTF: Bayesian tensor filtering with constrained, nonconjugate Poisson likelihood via GASS inference; NLL: Negative log-likelihood; MAE: Mean absolute deviation from truth; RMSE: Root mean-squared error from truth. Bold indicates the best performance at each embedding dimension setting

$d = 2$								
Model	Observations			True Rate		Coverage		
	MAE	RMSE	NLL	MAE	RMSE	50%	75%	95%
NMF	1.70	2.47	364.29	1.12	1.72	N/A	N/A	N/A
PGDS(0.25)	1.76	2.47	360.76	1.29	2.00	15.32	25.35	40.59
PGDS(0.5)	1.74	2.43	359.96	1.31	2.03	14.43	23.96	38.94
PGDS(1)	1.74	2.43	360.29	1.30	2.02	14.58	23.93	39.09
NBinom BTF	$> 10^2$	$> 10^3$	$> 10^5$	48.10	$> 10^2$	30.17	47.80	69.67
Poisson BTF	**1.66**	**2.34**	**354.52**	**0.91**	**1.43**	**34.16**	**53.40**	**74.72**

$d = 3$								
Model	Observations			True Rate		Coverage		
	MAE	RMSE	NLL	MAE	RMSE	50%	75%	95%
NMF	2.26	3.48	552.10	1.15	1.88	N/A	N/A	N/A
PGDS(0.25)	1.72	2.42	365.26	1.06	1.68	21.58	35.83	56.24
PGDS(0.5)	**1.71**	2.42	364.39	1.06	1.68	21.51	35.58	55.69
PGDS(1)	**1.71**	**2.41**	363.86	1.05	1.68	21.27	35.00	54.81
NBinom BTF	$> 10^3$	$> 10^4$	$> 10^6$	$> 10^2$	$> 10^3$	37.14	59.20	82.60
Poisson BTF	1.78	2.47	**358.95**	**0.79**	**1.29**	**40.45**	**62.11**	**82.87**

$d = 5$								
Model	Observations			True Rate		Coverage		
	MAE	RMSE	NLL	MAE	RMSE	50%	75%	95%
NMF	2.65	4.21	596.37	1.43	2.31	N/A	N/A	N/A
PGDS(0.25)	1.74	2.44	368.10	0.86	1.37	31.46	50.20	73.55
PGDS(0.5)	1.76	2.43	369.59	0.87	1.37	31.35	50.46	73.83
PGDS(1)	1.80	2.56	371.89	0.88	1.40	29.80	49.05	71.91
NBinom BTF	$> 10^2$	$> 10^2$	$> 10^4$	10.77	94.94	**45.31**	**67.85**	**89.66**
Poisson BTF	**1.71**	**2.26**	**355.75**	0.81	**1.25**	41.64	63.55	83.98

$d = 10$								
Model	Observations			True Rate		Coverage		
	MAE	RMSE	NLL	MAE	RMSE	50%	75%	95%
NMF	4.53	7.45	$> 10^3$	1.89	3.12	N/A	N/A	N/A
PGDS(0.25)	1.86	2.74	391.60	0.82	1.33	38.25	60.21	83.84
PGDS(0.5)	1.89	2.80	402.63	**0.79**	1.29	39.20	61.48	84.40
PGDS(1)	1.79	2.52	376.62	**0.79**	**1.26**	39.21	61.52	84.61
NBinom BTF	31.87	42.16	$> 10^3$	3.13	11.14	43.33	66.17	**87.55**
Poisson BTF	**1.72**	**2.43**	**359.12**	0.83	1.32	**44.05**	**67.03**	86.59

any smoothness prior built in. PGDS uses a canonical tensor decomposition, where the third tensor dimension is modeled with an embedding that is shared between rows and columns. This requires an inflation of the embedding dimension when columns are all evolving independently in order to sufficiently capture all of the latent structure in the data.

The negative binomial BTF model performs poorly on held out observations and true-rate estimation. This is due to the instability of the model on held-out data. The Pólya-gamma approach to negative binomial likelihoods uses the $NB(r, \sigma(\beta))$ parameterization, where r is an unknown dispersion parameter, σ is the logistic function, and β are log-odds. The mean of the distribution is

$$\frac{n\sigma(\beta)}{1 - \sigma(\beta)}.$$

When the failure rate $\sigma(\beta)$ is near zero or one, small changes in β lead to large changes in the mean. Thus, small errors in $\beta_{ijt} = w_i^\top v_{jt}$ on the held-out data lead to large errors in the observational NLL and the MAE and RMSE metrics we consider.

For each embedding dimension the Poisson BTF model performs competitively or better than the other models in each category. The negative log-likelihood on held-out observations is always lowest for the Poisson BTF model. For other metrics, Poisson BTF is either the best performing model or within 10% of the best performing model. By contrast, the other models are off by more substantial amounts in certain categories at certain dimension embedding settings. The credible intervals for Poisson BTF also consistently have better coverage than PGDS for the same embedding dimension. This includes the true embedding dimension $d = 3$, where Poisson BTF shows the best performance in terms of MAE to the true rate, and competitive performance with other choices of d. By contrast, both NMF and PGDS perform substantially worse in the $d = 3$ regime than in other choices. Finally, unlike the negative binomial model, the Poisson BTF model maintains a stable prediction on held out entries.

6.4. *Cancer drug study.* We evaluate the proposed empirical Bayes dose-response model built on top of BTF on two cancer drug studies. First, we use a small internal pilot study conducted at Columbia University Medical Center. The pilot study tested 35 drugs against 28 tumor organoids, each at nine different concentrations with six replicates. Second, we run on a large-scale, "landscape" study (Lee et al. (2018)) that tested 67 drugs against 284 tumor organoids, each at seven different concentrations with two replicates. For the pilot study we run five independent trials, holding out 30 curves at random, subject to the constraint that no column or row is left without any observations in the training set. For the landscape study we hold out a single test set of 1000 curves ($\approx 5\%$ of the total entries). Since this is real data, MAE and RMSE from the truth are not available; we measure performance solely in terms of negative log-likelihood on the held out data.

The standard dose-response modeling approach in cancer datasets is a log-linear logistic model (Vis et al. (2016)). For a baseline, we extend that model to a logistic factor model (LFM), using the same preprocessing strategy. We also compare to NMF as a second baseline. To ensure the monotonicity, we project the NMF results to be monotone curves using the PAV algorithm, as in Lin and Dunson (2014). We choose the factor size in both models by five-fold cross-validation on the training set.

For BTF we perform a grid search over hyperparameters: $\rho^2 = \{0.001, 0.01, 0.1\}$, factor size $D = \{1, 3, 5, 8\}$, and the order of the trend filtering matrix $k = \{0, 1\}$; we select the best model using the deviance information criterion (Celeux et al. (2006)). We evaluate the BTF model using the average of the posterior draws, rather than the full Bayes estimate; this enables us to fairly compare with the NMF and LFM point estimates of the latent mean. We run 10,000 Gibbs sampling steps in both studies, discarding the first 5000 as burn-in.

TABLE 4
Left: *Mean results* $\pm$ *standard error on held out data for the pilot
cancer drug studies. Right*: *Results on a single test set of* 1000
curves for the landscape study. NMF: *Nonnegative matrix
factorization*; *LFM*: *Logistic factor model*; *BTF*: *Bayesian tensor
filtering*; *NLL*: *Negative log-likelihood*

| | Cancer Drug Studies | |
| | Pilot Study | Landscape Study |
Model	NLL	NLL
NMF	262.75 ± 308.12	25573.14
LMF	589.17 ± 582.29	$> 10^6$
BTF	$\mathbf{-80.22 \pm 9.67}$	$\mathbf{-3268.11}$

Table 4 present the results. The BTF dose-response model outperforms both baselines in terms of negative log-likelihood of the held-out data in both studies. Results in terms of RMSE and MAE (not shown) on the raw observations were similar for all three models in both studies (e.g., RMSE 0.14 ± 0.01, MAE 0.20 ± 0.01 in the pilot study). The BTF procedure is also more stable in the pilot study cross-validation with a much lower reconstruction variance than either baseline. This suggests BTF not only forms a more accurate basis for a dose-response model but is also more reliable.

Qualitative results on the held out predictions are in Figure 1 (orange). All nine plots are for real data from the landscape study, with the gray observations held out. The orange line shows the posterior mean of the predicted curves. The curves have all of the desired properties: monotonicity with dose, bounded between zero and one, mostly smooth, locally adaptive to sharp jumps in the data, and highly predictive of the outcomes of the experiments. The orange bands show the 50% approximate posterior credible intervals using the empirical Bayes likelihood model. The credible intervals are conservative, estimating a larger variance than is actually observed in the outcomes. Even still, the NMF and LMF models far exceed these bands in certain points in the curve. This is due to the heteroskedastic nature of the likelihood and the misspecification of the NMF and LMF loss functions. Both competing models optimize for squared error, effectively making a heteroskedastic assumption on the model. In RMSE terms, all three models perform nearly identically, within ± 0.01 of each other on both datasets. Judging the models by RMSE would be misleading, since the high degree of noise in the first three dose levels dominates the overall loss and obscures the real fit of the model.

7. Landscape study analysis. In many cancer drug studies, molecular features such as gene expression, genomic mutations, or copy number alterations are gathered. These features represent useful side information that can help further denoise the dose-response data. Features also enable one to address the "cold-start" problem, enabling predictions for samples that have no dose-response data available. Features that are predictive of sensitivity or resistance in the dose-response experiments may represent "biomarkers"— diagnostic indicators of drug response. Biomarkers are candidate targets for future experimental investigation, such as targeted drug development. The landscape study of Lee et al. (2018) considered 115 molecular features, all binary, with a subset of the features gathered on a subset of organoid cell lines.

To incorporate potentially-missing features into the BTF dose-response model, we take a multiview factorization approach. For each feature x_{im}, $m = 1, \ldots, M$, for organoid cell

line i, we assume a latent factor model between the cell line embeddings and feature embeddings,

$$x_{im} \mid w_i, u_m \sim \text{Bern}(w_i^\top u_m)\mathbb{1}[0 \le w_i^\top u_m \le 1],$$

(16)

$$u_m \sim \text{MVN}(\mathbf{0}, \sigma_u^2 I).$$

Rather than a logistic link function with an unconstrained likelihood model, we use an identity link with a [0, 1] constraint. A logistic link would put the feature likelihood and dose-response likelihoods on different scales. By using a constrained identity link function in equation (16), a change in w_i has the same effect on the probability of dose-specific survival as in the probability of a feature being positive.

For the landscape study we set $\sigma_u = 1$ and use a latent embedding dimension of 10. All other hyperparameters and settings are the same, as in the benchmarks in Section 6.4.

7.1. *Site of origin structure captured by organoid embeddings.* We first check whether the model has uncovered latent structure in the data. Tumor molecular profiles and drug response are strongly associated with the site of origin of the tumor. The BTF dose-response model was not supplied any direct site-of-origin information. Nonetheless, we expect that some clustering of organoids into site of origin should emerge.

Visualizing structure in the 10-dimensional organoid embeddings is challenging, as with any data of more than four dimensions. Figure 6 shows a two-dimensional principal components analysis (PCA) of the posterior mean of the 10-dimensional organoid embeddings. The 2D projection of the embeddings confirms that site of origin is associated with the first two principal components. Ovarian cancers are predominantly clustered in the bottom-center of the 2D space, liver cancers skew left, and breast and gastric cancers skew right. Brain cancers, cover the entire space but also represent the largest and most diverse set of original tumors in the landscape dataset.
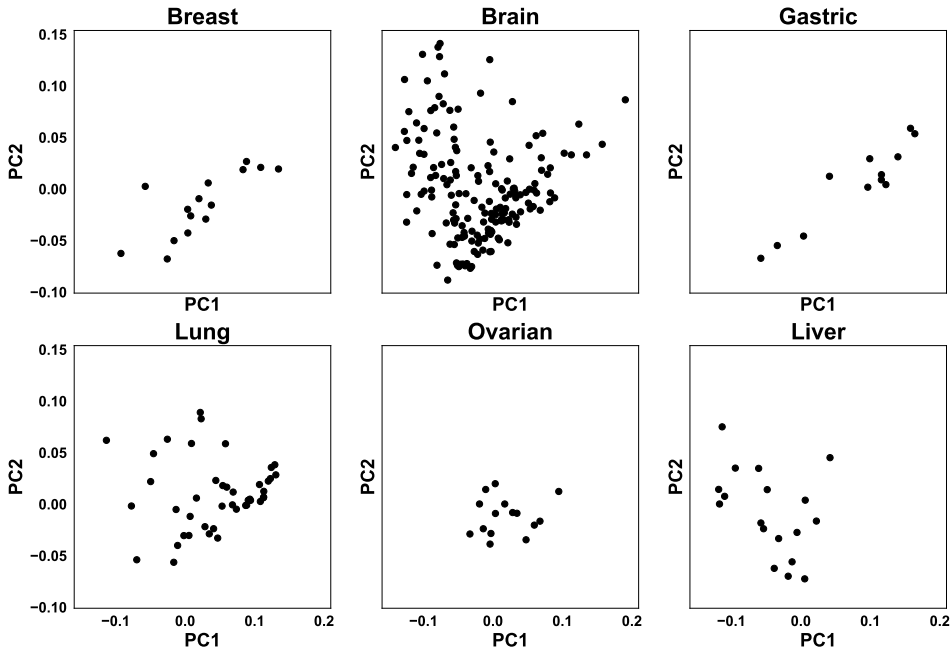

FIG. 6. *Two-dimensional PCA projection of the learned organoid embeddings, stratified by tumor site of origin. The embeddings reveal learned structure relating to site of origin which was not an explicit input to the model.*
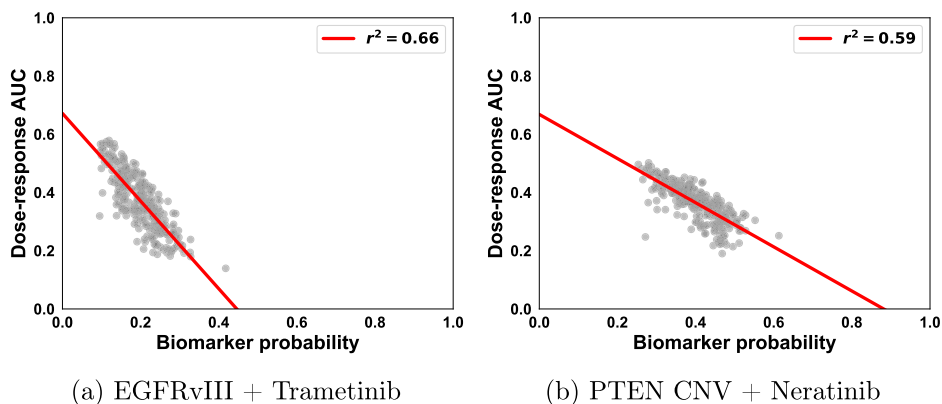
(a) EGFRvIII + Trametinib (b) PTEN CNV + Neratinib

Fig. 7. *Top two biomarker results correlating drug sensitivity with biomarker presence. Left: The vIII rearrangement in the EGFR gene is associated with sensitivity to treatment with the drug Trametinib. Right: Copy number variation, typically in the form of a copy number loss, in the tumor suppressor gene PTEN is associated with sensitivity to treatment with the drug Neratinib.*

7.2. *Uncovering biomarkers associated with drug sensitivity.* To discover potential biomarkers, we correlate the area under the dose-response curve (AUC) with the feature probability. To calculate the AUC, we approximate the curve with a piecewise-linear fit between successive dose intervals,

$$(17) \qquad AUC = \frac{1}{T-1} \sum_{t=1}^{T-1} w_i^\top (v_{jt} - v_{j(t+1)}).$$

We fit independent linear models to predict the posterior mean *AUC* values as a function of the biomarker probability. Features are ranked by r^2 value and stratified into *sensitivity* or *resistance*, based on the directionality of their slope. We filter out spurious results by removing features and drugs with standard deviation between samples is less than 0.05.

Figure 7 shows the top two results. Both features are flagged as potential biomarkers of drug sensitivity, indicating that as the probability of the feature increases, the AUC decreases. The top result (Figure 7a) associates the vIII rearrangement of the epidermal growth factor receptor (EGFR) gene with sensitivity to treatment with the drug Trametinib. EGFR is involved in many proliferation-inducing signaling pathways, including the important RAS/REF/MEK/ERK pathway (Kolch et al. (2015)). The vIII rearrangement of EGFR (EGFRvIII) leads to continual activation of EGFR, promoting oncogenesis (Guo et al. (2015)). Current approaches to targeting EFGRvIII have not seen major success (An et al. (2018)). The top result flags organoids with the EGFRvIII biomarker as associated with sensitivity to Trametinib, a MEK1/2 inhibitor. This may be due to Trametinib silencing the RAS/REF/MEK/ERK pathway being constitutively activated by the EGFRvIII rearrangement, suggesting a subpopulation of patients with EGFRvIII would benefit from Trametinib.

The second top result (Figure 7b) is copy number alterations in the phosphatase and tensin homolog (PTEN) gene being associated with sensitivity to treatment with Neratinib. PTEN is a tumor suppressor gene frequently lost in cancer, in particular glioblastoma (Koul (2008)) (GBM). The landscape study includes a plurality of GBM organoids and correspondingly the vast majority of the PTEN copy number variations in the dataset are due to PTEN loss. PTEN loss predicts resistance in Trastuzumab (Nagata et al. (2004)) and Neratinib has been shown to overcome Trastuzumab resistance in a subset of breast cancer patients (Canonici et al. (2013)). The association between PTEN loss and Neratinib sensitivity suggests a possible mechanistic explanation for these effects as well as a biomarker for other treatment-resistant patients.

8. Discussion. Multisample, multidrug cancer studies are time and resource intensive. The outcomes from these studies are noisy, often incomplete, observations of biological responses to candidate therapies. Denoising observations and imputing missing experiments is an important step in the scientific analysis and drug discovery pipeline. The Bayesian tensor filtering model we presented in this paper enables scientists to flexibly model dose-response curves with consideration for measurement error and biological constraints on the shape of the curve. While the BTF model is an improvement over the state of the art, we believe there are several improvements that could be made.

The BTF model assumes a regularly-spaced grid. Extensions to irregular grids of dose levels may be possible via a Bayesian group trend filtering extension to an irregular grid approach for scalar trend filtering. The approach taken in Faulkner and Minin (2018) was based on integrated Wiener processes. In the case of horseshoe priors, the method requires an approximation (Lindgren and Rue (2008)) that was only tractable up to second-order differences. We have found higher-order priors to not be necessary in our experience, suggesting irregular grid extensions here hold promise. Other methods from the trend-filtering literature, such as the falling factorial basis (Wang, Smola and Tibshirani (2014)), may also be adaptable to the Bayesian group trend filtering case.

The AUC values computed in Section 7.2 are likely to be slightly biased due to the piecewise-linear approximation of the true curve. For more precise estimation of the AUC curves, one could adopt the Nadaraya–Watson kernel smoothing of Piegorsch et al. (2012) by introducing pseudo-concentrations for each posterior sample. An alternative approach would be to introduce the pseduo-concentrations as missing data and have the model impute the values directly during posterior sampling with the composite trend filtering enforcing nonstationary smoothness. One could think of the latter approach as the more Bayesian approach, whereas the post hoc smoothing will be computationally faster. The modeling choice is analogous to how we enforce monotonicity directly in the posterior, as opposed to post hoc merging via the PAV approach of Lin and Dunson (2014).

The current BTF model is computationally intensive. For small scale studies like the pilot study, the model runs in a few hours on a laptop. The landscape study required several days on a compute cluster to perform the hyperparameter search. Relative to the years required for the landscape experiments, the run time is negligible. Nevertheless, offering an alternative inference approach that can scale more efficiently, such as variational inference, may make the BTF model useful for a broader group of scientists.

Finally, BTF does not support adding chemical features about the drugs. In organoid studies, scientists generally know the class of approved and potentially-translatable drugs. Even in high-throughput screening studies for cancer cell lines (i.e., not organoids), only known drugs—mostly chemotherapy agents—are typically tested. The goal in these studies is to find biological markers of resistance or sensitivity to the set of available compounds. The molecular feature analysis in Section 7 addresses this. However, if we were trying to discover entirely new drugs, such as might be done at a pharmaceutical company, extending the model to include drug features would be useful. This may be possible by including a second side-information matrix and adding a drug-specific embedding in a manner similar to canonical tensor decomposition.

SUPPLEMENTARY MATERIAL

Supplemental material A: Gibbs sampler derivations (DOI: 10.1214/21-AOAS1485SUPPA; .pdf). Full derivations for the conditional updates in the Gibbs sampler for BTF are available in the supplemental material.

Supplementary material B: Source code (DOI: 10.1214/21-AOAS1485SUPPB; .zip). All source code for BTF. The latest code release can be found at the GitHub repository, https://github.com/tansey/functionalmf.

REFERENCES

ABBAS-AGHABABAZADEH, F., LU, P. and FRIDLEY, B. L. (2019). Nonlinear mixed-effects models for modeling *in vitro* drug response data to determine problematic cancer cell lines. *Sci. Rep.* **9** 1–9.

AN, Z., AKSOY, O., ZHENG, T., FAN, Q.-W. and WEISS, W. A. (2018). Epidermal growth factor receptor and EGFRvIII in glioblastoma: Signaling pathways and targeted therapies. *Oncogene* **37** 1561–1575.

BHADRA, A., DATTA, J., POLSON, N. G. and WILLARD, B. (2017). The horseshoe+ estimator of ultra-sparse signals. *Bayesian Anal.* **12** 1105–1131. MR3724980 https://doi.org/10.1214/16-BA1028

BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning. Information Science and Statistics*. Springer, New York. MR2247587 https://doi.org/10.1007/978-0-387-45528-0

BORNKAMP, B. and ICKSTADT, K. (2009). Bayesian nonparametric estimation of continuous monotone functions with applications to dose-response analysis. *Biometrics* **65** 198–205. MR2665861 https://doi.org/10.1111/j.1541-0420.2008.01060.x

CAI, B. and DUNSON, D. B. (2007). Bayesian multivariate isotonic regression splines: Applications to carcinogenicity studies. *J. Amer. Statist. Assoc.* **102** 1158–1171. MR2412540 https://doi.org/10.1198/016214506000000942

CANONICI, A., GIJSEN, M., MULLOOLY, M., BENNETT, R., BOUGUERN, N., PEDERSEN, K., O'BRIEN, N. A., ROXANIS, I., LI, J.-L. et al. (2013). Neratinib overcomes trastuzumab resistance in HER2 amplified breast cancer. *Oncotarget* **4** 1592.

CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97** 465–480. MR2650751 https://doi.org/10.1093/biomet/asq017

CELEUX, G., FORBES, F., ROBERT, C. P. and TITTERINGTON, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Anal.* **1** 651–673. MR2282197 https://doi.org/10.1214/06-BA122

DROST, J. and CLEVERS, H. (2018). Organoids in cancer research. *Nat. Rev. Cancer* **18** 407–418. https://doi.org/10.1038/s41568-018-0007-6

FAGAN, F., BHANDARI, J. and CUNNINGHAM, J. (2016). Elliptical slice sampling with expectation propagation. In *Uncertainty in Artificial Intelligence*.

FAULKNER, J. R. and MININ, V. N. (2018). Locally adaptive smoothing with Markov random fields and shrinkage priors. *Bayesian Anal.* **13** 225–252. MR3737950 https://doi.org/10.1214/17-BA1050

FOX, E. B. and DUNSON, D. B. (2015). Bayesian nonparametric covariance regression. *J. Mach. Learn. Res.* **16** 2501–2542. MR3450515

FRIDLEY, B. L., JENKINS, G., SCHAID, D. J. and WANG, L. (2009). A Bayesian hierarchical nonlinear model for assessing the association between genetic variation and drug cytotoxicity. *Stat. Med.* **28** 2709–2722. MR2751045 https://doi.org/10.1002/sim.3649

GARNETT, M. J., EDELMAN, E. J., HEIDORN, S. J., GREENMAN, C. D., DASTUR, A., LAU, K. W., GRENINGER, P., THOMPSON, I. R., LUO, X. et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483** 570.

GAUVIN, L., PANISSON, A. and CATTUTO, C. (2014). Detecting the community structure and activity patterns of temporal networks: A non-negative tensor factorization approach. *PLoS ONE* **9** e86028. https://doi.org/10.1371/journal.pone.0086028

GHANDI, M., HUANG, F. W., JANÉ-VALBUENA, J., KRYUKOV, G. V., LO, C. C., MCDONALD, E. R., BARRETINA, J., GELFAND, E. T., BIELSKI, C. M. et al. (2019). Next-generation characterization of the cancer cell line encyclopedia. *Nature* **569** 503–508.

GHEBRETINSAE, A. H., FAES, C., MOLENBERGHS, G., DE BOECK, M. and GEYS, H. (2013). A Bayesian, generalized frailty model for comet assays. *J. Biopharm. Statist.* **23** 618–636. MR3049131 https://doi.org/10.1080/10543406.2012.756499

GUO, G., GONG, K., WOHLFELD, B., HATANPAA, K. J., ZHAO, D. and HABIB, A. A. (2015). Ligand-independent EGFR signaling. *Cancer Res.* **75** 3436–3441.

HAHN, P. R., HE, J. and LOPES, H. (2018). Bayesian factor model shrinkage for linear IV regression with many instruments. *J. Bus. Econom. Statist.* **36** 278–287. MR3790214 https://doi.org/10.1080/07350015.2016.1172968

HEAUKULANI, C. and VAN DER WILK, M. (2019). Scalable Bayesian dynamic covariance modeling with variational Wishart and inverse Wishart processes. In *Advances in Neural Information Processing Systems* 4582–4592.

HUANG, L., WU, S. and XING, D. (2011). High fluence low-power laser irradiation induces apoptosis via inactivation of akt/GSK3β signaling pathway. *J. Cell. Physiol.* **226** 588–601.

JOHNSON, W. E., LI, C. and RABINOVIC, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8** 118–127.

KIM, S.-J., KOH, K., BOYD, S. and GORINEVSKY, D. (2009). l_1 trend filtering. *SIAM Rev.* **51** 339–360. MR2505584 https://doi.org/10.1137/070690274

KOLCH, W., HALASZ, M., GRANOVSKAYA, M. and KHOLODENKO, B. N. (2015). The dynamic control of signal transduction networks in cancer cells. *Nat. Rev. Cancer* **15** 515–527.

KOUL, D. (2008). PTEN signaling pathways in glioblastoma. *Cancer Biol. Ther.* **7** 1321–1325.

KOWAL, D. R., MATTESON, D. S. and RUPPERT, D. (2019). Dynamic shrinkage processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 781–804. MR3997101 https://doi.org/10.1111/rssb.12325

KUNIHAMA, T., HALPERN, C. T. and HERRING, A. H. (2019). Non-parametric Bayes models for mixed scale longitudinal surveys. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **68** 1091–1109. MR4002385 https://doi.org/10.1111/rssc.12348

KYUNG, M., GILL, J., GHOSH, M. and CASELLA, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Anal.* **5** 369–411. MR2719657 https://doi.org/10.1214/10-BA607

LACHMANN, A., GIORGI, F. M., ALVAREZ, M. J. and CALIFANO, A. (2016). Detection and removal of spatial bias in multiwell assays. *Bioinformatics* **32** 1959–1965.

LEE, J.-K., LIU, Z., SA, J. K., SHIN, S., WANG, J., BORDYUH, M., CHO, H. J., ELLIOTT, O., CHU, T. et al. (2018). Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy. *Nat. Genet.* **50** 1399–1411.

LEEK, J. T., SCHARPF, R. B., BRAVO, H. C., SIMCHA, D., LANGMEAD, B., JOHNSON, W. E., GEMAN, D., BAGGERLY, K. and IRIZARRY, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **11** 733–739.

LI, L., PLUTA, D., SHAHBABA, B., FORTIN, N., OMBAO, H. and BALDI, P. (2019). Modeling dynamic functional connectivity with latent factor Gaussian processes. In *Advances in Neural Information Processing Systems* 8263–8273.

LIN, L. and DUNSON, D. B. (2014). Bayesian monotone regression using Gaussian process projection. *Biometrika* **101** 303–317. MR3215349 https://doi.org/10.1093/biomet/ast063

LINDGREN, F. and RUE, H. (2008). On the second-order random walk model for irregular locations. *Scand. J. Stat.* **35** 691–700. MR2468870 https://doi.org/10.1111/j.1467-9469.2008.00610.x

LOW-KAM, C., TELESCA, D., JI, Z., ZHANG, H., XIA, T., ZINK, J. I. and NEL, A. E. (2015). A Bayesian regression tree approach to identify the effect of nanoparticles' properties on toxicity profiles. *Ann. Appl. Stat.* **9** 383–401. MR3341120 https://doi.org/10.1214/14-AOAS797

MAZOURE, B., NADON, R. and MAKARENKOV, V. (2017). Identification and correction of spatial bias are essential for obtaining quality data in high-throughput screening technologies. *Sci. Rep.* **7** 11921. https://doi.org/10.1038/s41598-017-11940-4

MINKA, T. P. (2001). Expectation propagation for approximate Bayesian inference. In *Uncertainty in Artificial Intelligence* 362–369.

MORAN, K. R., DUNSON, D., WHEELER, M. W. and HERRING, A. H. (2019). Bayesian joint modeling of chemical structure and dose response curves. arXiv preprint arXiv:1912.12228.

MURRAY, I., ADAMS, R. and MACKAY, D. (2010). Elliptical slice sampling. In *Artificial Intelligence and Statistics*.

NAGATA, Y., LAN, K.-H., ZHOU, X., TAN, M., ESTEVA, F. J., SAHIN, A. A., KLOS, K. S., LI, P., MONIA, B. P. et al. (2004). PTEN activation contributes to tumor inhibition by trastuzumab, and loss of PTEN predicts trastuzumab resistance in patients. *Cancer Cell* **6** 117–127.

NEAL, R. M. (2003). Slice sampling. *Ann. Statist.* **31** 705–767. With discussions and a rejoinder by the author. MR1994729 https://doi.org/10.1214/aos/1056562461

NEELON, B. and DUNSON, D. B. (2004). Bayesian isotonic regression and trend analysis. *Biometrics* **60** 398–406. MR2066274 https://doi.org/10.1111/j.0006-341X.2004.00184.x

PATEL, T., TELESCA, D., GEORGE, S. and NEL, A. E. (2012). Toxicity profiling of engineered nanomaterials via multivariate dose-response surface modeling. *Ann. Appl. Stat.* **6** 1707–1729. MR3058681 https://doi.org/10.1214/12-AOAS563

PERRON, F. and MENGERSEN, K. (2001). Bayesian nonparametric modeling using mixtures of triangular distributions. *Biometrics* **57** 518–528. MR1855686 https://doi.org/10.1111/j.0006-341X.2001.00518.x

PIEGORSCH, W. W., XIONG, H., BHATTACHARYA, R. N. and LIN, L. (2012). Nonparametric estimation of benchmark doses in environmental risk assessment. *Environmetrics* **23** 717–728. MR3019063 https://doi.org/10.1002/env.2175

POLSON, N. G. and SCOTT, J. G. (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction. *Bayesian Stat.* **9** 501–538. MR3204017 https://doi.org/10.1093/acprof:oso/9780199694587.003.0017

POLSON, N. G., SCOTT, J. G. and WINDLE, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *J. Amer. Statist. Assoc.* **108** 1339–1349. MR3174712 https://doi.org/10.1080/01621459.2013.829001

SCHEIN, A., WALLACH, H. and ZHOU, M. (2016). Poisson–Gamma dynamical systems. In *Advances in Neural Information Processing Systems*.

SHIVELY, T. S., SAGER, T. W. and WALKER, S. G. (2009). A Bayesian approach to non-parametric monotone function estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 159–175. MR2655528 https://doi.org/10.1111/j.1467-9868.2008.00677.x

SHOEMAKER, R. H. (2006). The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer* **6** 813–823.

SPIEGEL, S., CLAUSEN, J., ALBAYRAK, S. and KUNEGIS, J. (2011). Link prediction on evolving data using tensor factorization. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*.

TAKEUCHI, K., KASHIMA, H. and UEDA, N. (2017). Autoregressive tensor factorization for spatio-temporal predictions. In *International Conference on Data Mining*.

TANSEY, W., TOSH, C. and BLEI, D. M. (2022). Supplement to "A Bayesian model of dose-response for cancer drug studies." https://doi.org/10.1214/21-AOAS1485SUPPA, https://doi.org/10.1214/21-AOAS1485SUPPB

TANSEY, W., LI, K., ZHANG, H., LINDERMAN, S. W., RABADAN, R., BLEI, D. M. and WIGGINS, C. H. (2021). Dose–response modeling in high-throughput cancer drug screenings: An end-to-end approach. *Biostatistics*.

TIBSHIRANI, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Ann. Statist.* **42** 285–323. MR3189487 https://doi.org/10.1214/13-AOS1189

VAN DER PAS, S. L., KLEIJN, B. J. K. and VAN DER VAART, A. W. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electron. J. Stat.* **8** 2585–2618. MR3285877 https://doi.org/10.1214/14-EJS962

VIS, D. J., BOMBARDELLI, L., LIGHTFOOT, H., IORIO, F., GARNETT, M. J. and WESSELS, L. F. (2016). Multilevel models improve precision and speed of IC50 estimates. *Pharmacogenomics J.* **17** 691–700. https://doi.org/10.2217/pgs.16.15

WANG, Y.-X., SMOLA, A. and TIBSHIRANI, R. (2014). The falling factorial basis and its statistical applications. In *International Conference on Machine Learning* 730–738.

WEINSTEIN, J. N., COLLISSON, E. A., MILLS, G. B., SHAW, K. R. M., OZENBERGER, B. A., ELLROTT, K., SHMULEVICH, I., SANDER, C., STUART, J. M. et al. (2013). The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **45** 1113.

WEST, M. (2003). Bayesian factor regression models in the "large *p*, small *n*" paradigm. In *Bayesian Statistics*, 7 (*Tenerife*, 2002) 733–742. Oxford Univ. Press, New York. MR2003537

WHEELER, M. W. (2019). Bayesian additive adaptive basis tensor product models for modeling high dimensional surfaces: An application to high-throughput toxicity testing. *Biometrics* **75** 193–201. MR3953720

WILSON, A., REIF, D. M. and REICH, B. J. (2014). Hierarchical dose-response modeling for high-throughput toxicity screening of environmental chemicals. *Biometrics* **70** 237–246. MR3251684 https://doi.org/10.1111/biom.12114

XIONG, L., CHEN, X., HUANG, T.-K., SCHNEIDER, J. and CARBONELL, J. G. (2010). Temporal collaborative filtering with Bayesian probabilistic tensor factorization. In *International Conference on Data Mining*.

ZHANG, A. and PAISLEY, J. (2018). Deep Bayesian nonparametric tracking. In *International Conference on Machine Learning* 5828–5836.