Strategies for Distributed Bayesian Inference with Independent and Spatio-temporally Correlated Data

Rajarshi Guhaniyogi

Associate Professor, Department of Statistics, Texas A & M University,

3143 TAMU, College Station, TX 77843-3143 , E-mail: rajguhaniyogi@tamu.edu

May 6, 2024

Abstract

Due to recent advancements in technology, the utilization of massive datasets, whether independent or correlated, has become prevalent across various fields such as forestry, ocean science, neuroimaging, and public health. Traditional Bayesian statistical approaches for analyzing such datasets operate on the assumption that all data from different units are stored and processed centrally. This centralized data processing paradigm necessitates significant computing and storage resources at the central server, while also raising valid privacy concerns for individual units contributing data. To address these challenges, a comprehensive Bayesian analysis framework has been devised, allowing for approximate Bayesian inference in a decentralized manner, with model computation distributed across multiple machines. This article presents a review of pertinent methods in distributed Bayesian inference developed in recent years, catering to both independent and spatio-temporally correlated data.

Keywords: Bayesian statistics; Distributed inference; spatio-temporal data.

1 Introduction

Due to advancements in technology across various scientific domains, the prevalence of massive structured datasets has become widespread in crucial applications, including climatology, neuroscience, bioinformatics, forestry, and environmental science. Flexible models for structured data, incorporating high-dimensional parameters and/or correlated random effects, are utilized to capture the intricate features of these complex datasets. Despite numerous successes, current studies

on such complex structured data models often assume a centralized data processing approach, where all data is stored and processed at a central location. However, this centralized paradigm poses several challenges in modern applications, particularly related to storage and model computation. These challenges are magnified when sample units exhibit spatial or spatio-temporal correlations (Heaton et al., 2019). By adopting a decentralized approach to model computation, various advantages emerge:

1. *Improved Privacy:* Decentralized inference allows for the sharing of focused updates from local computations instead of raw data, enhancing privacy and ensuring that sensitive information remains localized.

2. *Reduced Computation and Storage Needs:* De-centralized computation shifts the central server's role primarily to combining analyses from different processors, resulting in decreased computation and storage demands on the server.

3. *Reduced Latency and Communication Traffic:* De-centralized inference eliminates the need for massive raw data exchange and communication between the central server and processors during statistical analysis, leading to reduced latency and communication traffic.

4. *Faster Inference:* Efficient computation on smaller data shards at each processor leads to faster overall inference.

Distributed Bayesian inferential approaches have gained popularity by simultaneously addressing these advantages. These approaches typically operate in three stages. In the first stage, the sample of size n is divided into J subsets, with the jth subset having a sample size of n_j, where $\sum_{j=1}^{J} n_j = n$. In the second stage, a suitable model is fitted on the J subsets in parallel to obtain J subset posteriors. This stage allows for Markov Chain Monte Carlo (MCMC) approximations of the full data posterior without exchanging data between processors or communication with a central server. Each subset's likelihood is adjusted to account for the missing fraction of data. At the final stage, the subset posteriors are combined to obtain a "collaborative pseudo posterior," serving as a computationally efficient approximation to the full posterior. Section 2 provides a comprehensive review of existing literature on three-stage distributed Bayesian inferential techniques.

While distributed Bayesian approaches share some similarities with composite likelihood approaches (Varin et al., 2011), the two methodologies exhibit distinct differences. For instance, Chandler and Bate (2007) and Ribatet et al. (2012) create pseudo-likelihood as a substitute for the full data likelihood. This pseudo-likelihood aims to capture essential features of the full data likelihood while ensuring computational efficiency. In spatial or spatio-temporal data modeling with Gaussian Processes (GP) or their variants, pseudo-likelihood, based on the independence of data blocks to some extent, is employed for computational efficiency. To address the incorrect asymptotic distribution of the posterior resulting from the assumption of incorrect independence, various adjustments in the composite log likelihood, such as margin and curvature adjustments, have been proposed. Similar to these approaches, the likelihood adjustment in each subset during the second stage of the general distributed Bayesian approach is motivated by the need to scale the asymptotic variance of subset posteriors to the same order as the asymptotic variance of the full posterior. However, unlike composite likelihood approaches, the discussed distributed approaches do not assume any restrictive structure, such as block independence, in the data likelihood. There is no guarantee that the induced data likelihood, leading to the collaborative pseudo posterior for any distributed method, assumes a block independence form. Moreover, Srivastava et al. (2018) provides an example of embedding a composite likelihood in a distributed setup. Similarly, we believe that most of these "flexible" composite likelihoods can be employed in extensions of the distributed framework for subset sampling in models where the true likelihood is either unavailable or expensive to compute.

2 Distributed Bayesian Inference: A Three-Stage Strategy

2.1 Partitioning of Data

The initial step in distributed Bayesian inference involves the creation of subsets $D_1, ..., D_J$ of the full data D, such that $D_1 \cup \cdots \cup D_J = D$, $D_i \cap D_j = \phi$. While the partitioning of data samples is less critical in cases where the data is i.i.d., caution is warranted when dealing with spatio-temporally correlated data. Several approaches have been explored in the literature (Guhaniyogi and Banerjee, 2018; Guhaniyogi et al., 2022; Guhaniyogi and Banerjee, 2019; Guhaniyogi et al., 2023), outlined below,

Random Partitioning

This scheme involves the random allocation of location-time tuples into J subsets, which may or may not overlap. The objective is to ensure that each subset contains representative data samples from all subregions within the domain.

Stratified Partitioning

In this approach, the space-time domain is divided into subdomains, and each of the J subsets is constructed with representative samples from each subdomain.

Both random and stratified partitioning methods aim to guarantee that each subset comprises samples from the entire space-time domain. This inclusivity is crucial for achieving reliable inference in distributed Bayesian methods, as emphasized in Guhaniyogi et al. (2023); Guhaniyogi and Banerjee (2018). However, it is worth noting that even when each subset contains representative samples from the entire space-time domain, inference in distributed Bayesian learning may be somewhat sensitive to the choice of data subsets Guhaniyogi et al. (2023).

2.2 Fitting a Model on Data Subsets: Construction of Subset Posteriors

In the second stage, the distributed Bayesian approach involves fitting a suitable statistical model tailored to the data in each subset. Let $f(D_j|\eta)$ denote the likelihood of subset D_j corresponding to a suitable model characterized by a scalar- or vector-valued parameter η. We establish subset posterior distributions by modifying the likelihood in $f(D_j|\eta)$. More specifically, we define the jth subset posterior distribution of η as follows:

$$\pi_{n_j}(\eta \mid D_j) = \frac{f(D_j|\eta)^{n/n_j}\pi(\eta)}{\int f(D_j|\eta)^{n/n_j}\pi(\eta)d\eta}, \tag{1}$$

where we assume that $\int f(D_j|\eta)^{n/n_j}\pi(\eta)d\eta < \infty$, and the subscript '$n_j$' indicates that the density conditions on n_j data samples in the jth subset D_j. The modification of the likelihood to yield the subset posterior density in (1) is termed *stochastic approximation* (Minsker et al., 2014). Raising the likelihood to the power of n/n_j is equivalent to replicating every D_j, n/n_j times ($i = 1, \ldots, n_j$), so stochastic approximation takes into account that the jth subset posterior distribution conditions on an (n_j/n)-fraction of the full data, ensuring that its variance is of the same order (as a function of n) as that of the full data posterior distribution. Markov Chain Monte Carlo (MCMC) samples are drawn from each subset posterior distribution to obtain an empirical approximation of the

subset posteriors.

The existing body of literature on distributed Bayesian approaches extensively explores the fitting of different $f(D_j|\eta)$ originating from structured data models with independent errors, as well as those with spatially/spatio-temporally correlated data. In this regard, Minsker et al. (2014); Srivastava et al. (2018); Li et al. (2017) have delved into fitting ordinary linear regression with varying dimensions, finite mixtures of Gaussians, and probabilistic parafac models within the distributed Bayesian framework. Regarding the exploration of distributed Bayesian inference for spatial/spatio-temporal data, Guhaniyogi and Banerjee (2018); Guhaniyogi et al. (2023) have fitted spatially dependent Gaussian process (GP) models, Guhaniyogi and Banerjee (2019) have applied multivariate Gaussian process models, and Guhaniyogi et al. (2022) have employed multivariate spatially varying coefficient models.

2.3 Combining Subset Posteriors: Construction of Collaborative Pseudo Posterior

In this section, we explore different approaches for merging subset posteriors to create a "collaborative pseudo posterior," serving as a computationally efficient substitute for the full data posterior. The discussed combination strategies encompass several methods commonly employed in divide-and-conquer inference for both independent and correlated data scenarios. As part of our discourse in this article, we will provide a concise overview of the following combination schemes: (i) Consensus Monte Carlo (CMC); (ii) Double Parallel Monte Carlo (DPMC); (iii) combination through median posterior (iv) combination through Wasserstein barycenter; and (v) Aggregated Monte Carlo (AMC).

Consensus Monte Carlo (CMC)

For any parameter of interest, denoted as η and representing either a scalar or a vector, the Consensus Monte Carlo (CMC) strategy (Scott et al., 2022; Rendell et al., 2020; Rabinovich et al., 2015) offers a method to draw samples from an approximation of the full posterior. Let $\left\{\eta_j^{(1)}, ..., \eta_j^{(L)}\right\}$ represent the L posterior samples of η generated from subset j post-convergence. Leveraging the Bernstein-von Mises theorem, which asserts that the posterior converges to a normal distribution centered around the true parameter value with increasing observations, Scott et al. (2022) suggested utilizing the weighted average $\sum_{j=1}^{J} w_j \eta_j^{(l)}$, where $l = 1, ..., L$, to approximate L samples from the

full data posterior. Here, the weight w_j corresponds to the inverse of the empirical covariance matrix of $\left\{\eta_j^{(1)}, ..., \eta_j^{(L)}\right\}$. This algorithm provides an exact solution when each subset posterior follows a Gaussian distribution.

Double Parallel Monte Carlo (DPMC)

For any parameter of interest, denoted as η and representing either a scalar or a vector, let $\overline{\eta}_1, ...\overline{\eta}_J$ denote the empirically estimated mean of η from the J subset posteriors, and $\overline{\eta} = \frac{1}{J}\sum_{j=1}^{J}\overline{\eta}_j$ denote their average. The Double Parallel Monte Carlo (DPMC) method (Xue and Liang, 2019) re-centers each subset posterior to $\overline{\eta}$ and then utilizes the mixture of re-centered subset posteriors, expressed as $\frac{1}{J}\sum_{j=1}^{J}\pi_{n_j}(\eta - \overline{\eta} + \overline{\eta}_j|D_j)$, to approximate the full data posterior (Xue and Liang, 2019). The DPMC algorithm employs the population approximation stochastic Monte Carlo (pop-SAMC) technique to sample from the approximate posterior distribution for parameters.

Median posterior

This approach relies upon the unique Geometric Median (GM) of the subset posteriors (Minsker et al., 2014; Minsker, 2015; Guhaniyogi and Banerjee, 2018). Assume that the subset posterior densities $\pi_{n_j}(\cdot)$ reside on a Banach space $\mathcal{P}$ equipped with norm $||\cdot||_\rho$. The GM is defined as $\pi^* = arg\min_{\pi\in\mathcal{P}}\sum_{j=1}^{J}||\pi_{n_j} - \pi||_\rho$. The norm quantifies distance between any two densities $\pi_{n_j}, \pi_{n_{j'}}$ as $||\pi_{n_j} - \pi_{n_{j'}}||_\rho = |\int \rho(\eta, \cdot)d(\pi_{n_j} - \pi_{n_{j'}})(\eta)|$. When π_{n_j} and $\pi_{n_{j'}}$ are empirically approximated from the post burn-in iterates of η, $||\pi_{n_j} - \pi_{n_{j'}}||_\rho$ also assumes an empirical approximation. To elaborate, let the empirically approximations of π_{n_j} and $\pi_{n_{j'}}$ be given by $\pi_{n_j} = \sum_{m=1}^{M_j}\gamma_{mj}1_{\xi_{mj}}$ and $\pi_{n_{j'}} = \sum_{m=1}^{M_{j'}}\gamma_{mj'}1_{\xi_{mj'}}$, then $||\pi_{n_j} - \pi_{n_{j'}}||_\rho = \sum_{m=1}^{M_j}\sum_{m'=1}^{M_j}\gamma_{mj}\gamma_{m'j}\rho(\xi_{mj}, \xi_{m'j}) + \sum_{m=1}^{M_{j'}}\sum_{m'=1}^{M_{j'}}\gamma_{mj}\gamma_{m'j'}\rho(\xi_{mj'}, \xi_{m'j'}) - 2\sum_{m=1}^{M_j}\sum_{m'=1}^{M_{j'}}\gamma_{mj}\gamma_{m'j'}\rho(\xi_{mj}, \xi_{m'j'})$ is available in closed form. To compute GM, Weiszfeld's algorithm (Guhaniyogi and Banerjee, 2018) is employed which estimates weights $\alpha_{\rho,j}(D)$, such that $\pi^* = \sum_{j=1}^{J}\alpha_{\rho,j}(D)\pi_{n_j}$, with $\alpha_{\rho,j}(D) \geq 0$ and $\sum_{j=1}^{J}\alpha_{\rho,j}(D) = 1$. The approach finds the GM distribution in the convex hull of the posterior distributions.

Wasserstein barycenter

While GM relies on the "median" of the subset posterior distributions, this combination scheme relies on finding the "mean" of the subset posterior distribution. In the space of distributions, the notion of "mean" is defined through Wasserstein barycenter (Srivastava et al., 2018). We first provide some background on Wasserstein barycenter. Let $(\mathcal{H}, \zeta)$ be a complete separable metric space and $\mathcal{P}(\mathcal{H})$ be the space of all probability measures on $\mathcal{H}$. The Wasserstein space of order 2 is a set

of probability distributions defined as $\mathcal{P}_2(\mathcal{H}) = \{\mu \in \mathcal{P}(\mathcal{H}) : \int_{\mathcal{H}} \zeta^2(\eta, \eta_0)\mu(d\eta) < \infty\}$, where $\eta_0 \in \mathcal{H}$ is arbitrary and $\mathcal{P}_2(\mathcal{H})$ does not depend on the choice of η_0. The Wasserstein distance of order 2, denoted as W_2, is a metric on $\mathcal{P}_2(\mathcal{H})$. Let μ, ν be two probability measures in $\mathcal{P}_2(H)$ and $\Pi(\mu, \nu)$ be the set of all probability measures on $\mathcal{H} \times \mathcal{H}$ with marginals μ and ν, then W_2 distance between μ and ν is defined as $W_2(\mu, \nu) = \underset{\pi \in \Pi(\mu,\nu)}{Inf} \int_{\mathcal{H}\times\mathcal{H}} \zeta^2(x, y)\, d\pi(x, y)\}^{1/2}$. Let $\nu_1, \ldots, \nu_J \in \mathcal{P}_2(\mathcal{H})$, then the Wasserstein barycenter of $\nu_1, \ldots, \nu_J$ is defined as

$$\overline{\nu} = argmin_{\nu \in \mathcal{P}_2(\mathcal{H})} \frac{1}{J} \sum_{j=1}^{J} W_2^2(\nu, \nu_j). \tag{2}$$

It is known that $\overline{\nu}$ exists and is unique (Agueh and Carlier, 2011).

In this combination framework, for any parameter of interest η, whether scalar or vector, the term "collaborative pseudo posterior" refers to the Wasserstein barycenter of the J subset posterior distributions of η. To explain further, in Wasserstein barycenter-based approaches, the components $\nu_1, \ldots, \nu_J$ in (2) are identified as the J subset posterior distributions of η, denoted as $\nu_j = \pi_{n_j}(\cdot)$. Consequently, the collaborative pseudo posterior, obtained through the mathematical computation of the Wasserstein barycenter $\overline{\pi}$ in (2), represents a general concept for deriving the mean of J potentially dependent subset posterior distributions.

In the context of Bayesian inference, the exact subset posteriors of η (expressed as $\pi_{n_1}, \ldots, \pi_{n_J}$ in (2)) are often analytically intractable. However, effective approximations can be achieved using MCMC samples from the subset posterior distributions of η. The empirical version of the Wasserstein barycenter $\overline{\pi}$ is then estimated empirically by solving a sparse linear program efficiently, as outlined in Cuturi and Doucet (2014); Srivastava et al. (2015, 2018). This approach to combining subset posteriors leads to the collaborative pseudo posterior known as the Wasserstein posterior (WASP). Empirical evidence suggests that, particularly for independent data, WASP is a preferable choice over various other combination methods (Srivastava et al., 2018). For instance, directly averaging over numerous subset posterior densities with different means can often yield an undesirable multimodal collaborative pseudo posterior distribution. WASP does not encounter this issue and can recover a unimodal posterior, as illustrated in Srivastava et al. (2018). Moreover, it does not rely on the asymptotic normality of subset posterior distributions, distinguishing it from other methods like CMC, as discussed earlier (Scott et al., 2022).

While WASP relies upon iterative algorithms to compute $\overline{\pi}$ from subset posteriors $\pi_{n_1}, ..., \pi_{n_J}$, an approach free of implementing iterative algorithms can be employed if η represents a one-dimensional functional of interest. For a scalar parameter the Wasserstein barycenter of η can be easily obtained by averaging empirical subset posterior quantiles (Li et al., 2017; Guhaniyogi et al., 2023). We refer to this approach as the quantile aggregation (QA) and the collaborative pseudo posterior is called as the QA posterior. The simple combination formula exists as the W_2 distance between two univariate distributions is the same as the L_2 distance between their quantile functions (Lemma 8.2 of Bickel and Freedman 1981). In particular, let $\overline{\pi}$ be the Wasserstein barycenter of $\pi_{n_1}, \ldots, \pi_{n_J}$ as in (2). For any $q \in (0,1)$, let $\pi_{n_j}^{-q}$ be the qth empirical quantile of π_{n_j} based on the MCMC samples from π_{n_j}, and $\overline{\pi}^{-q}$ be the qth quantile of the empirical version of $\overline{\pi}$. Then, $\overline{\pi}^{-q}$ can be computed as

$$\overline{\pi}^{-q} = \frac{1}{J} \sum_{j=1}^{J} \pi_{n_j}^{-q}. \tag{3}$$

Li et al. (2017); Guhaniyogi et al. (2023) suggested computing the empirical quantiles over a values of q on a grid. If the grid is fine enough in (3), then the parameter MCMC samples from different marginals of the collaborative pseudo posterior are obtained by inverting the empirical distribution function supported on the quantile estimates.

The choice of the grid size is primarily influenced by the Monte Carlo approximation error associated with each subset posterior. In general, this approximation error for subset posteriors is measured in terms of the size of MCMC samples, denoted as L. It is crucial to note that this evaluation involves considering L approaching infinity, distinguishing it from the statistical error, which occurs when the sample size n tends to infinity. For the divide-and-conquer Bayes method applied to models with i.i.d. data, Theorem 3 in the supplementary material of Li et al. (2017) demonstrates that the Monte Carlo error typically follows some polynomial order of L, such as $O(L^{-1/2})$ and $O(L^{-1/4})$, depending on the chosen distance measure. Notably, this error is independent of the statistical error defined in terms of n. Following this insight, practical recommendations suggest drawing a minimum of 10^4 MCMC samples for each subset posterior and utilizing the entire set for constructing the desired quantiles.

Aggregated Monte Carlo

Notably, WASP offers Wasserstein barycenter for the subset posterior distribution for all parameters jointly, but it relies on an iterative algorithm to compute it. On the other hand, QA technique allows combination of separate marginal distributions of subset posteriors. In order to combine subset posterior distributions jointly, as well as to avoid iterative algorithms, Guhaniyogi et al. (2022) proposed the Aggregated Monte Carlo (AMC) algorithm. With $\eta_j^{(1)}, ..., \eta_j^{(L)}$ being the MCMC iterates from the jth subset posterior for parameter η, AMC proposes computing the empirical mean and empirical variance of η from the jth subset, given by

$$\mu_{j\eta} = \frac{1}{L}\sum_{l=1}^{L} \eta_j^{(l)}, \;\; \Sigma_{j\eta} = \frac{1}{L}\sum_{l=1}^{L} (\eta_j^{(l)} - \mu_{j\eta})(\eta_j^{(l)} - \mu_{j\eta})^T, \;\; j = 1, .., J. \tag{4}$$

We now summarize the AMC algorithm for obtaining draws from the collaborative pseudo posterior using the subset posterior draws. First, define the combined empirical mean and covariance matrix for η draws using the subset posterior empirical means and covariance matrices in (4) as

$$\mu_\eta = \frac{1}{J}\sum_{j=1}^{J} \mu_{j\eta}, \;\; \Sigma_\eta = \frac{1}{J}\sum_{j=1}^{J} \Sigma_{j\eta}$$

Second, center and scale the jth subset posterior draws of η as

$$q_{j\eta}^{(l)} = \Sigma_{j\eta}^{-1}(\eta_j^{(l)} - \mu_{j\eta})^T, \;\; j = 1, ..., J; \; l = 1, ..., L.$$

Third, rescale and recenter the η draws from all the subsets as

$$\eta_{l'} = \mu_\eta + \Sigma_\eta^{1/2} q_{j\eta}^{(l)}, \;\; l' = l + (j-1)L, \; j = 1, .., J$$

to obtain l'th draws from the collaborative pseudo posterior distribution of η.

While CMC, DPMC and WASP allow combining subset posteriors of a multi-variate parameter, quantile aggregation approach is based on combining marginals of subset posteriors separately. In practice, the primary interest often lies in the marginal distributions of model parameters and predicted values; that is, the posterior distribution of some one-dimensional functional η; therefore, the uni-variate Wasserstein barycenter obtained by averaging quantiles in (3) accomplishes this with great generality and convenient implementation. Guhaniyogi et al. (2022) show practically

indistinguishable performance from these approaches.

An essential aspect of the combination scheme in distributed approaches is its independence from the chosen model, given the subset posterior MCMC samples. Importantly, the computational efficiency of the combination steps is notable, with the primary computational burden residing in computing each subset posterior. Assuming ample computational resources, a sufficiently large J is selected to ensure the feasibility of subset posterior computations. However, the choice of J is constrained from growing arbitrarily with respect to n to maintain the desired inference accuracy from the collaborative pseudo posterior. Guhaniyogi et al. (2022, 2023) comprehensively explore theoretical properties of distributed Bayesian approaches for correlated spatial/spatio-temporal data, and offers a guideline to choose J as a function of n and properties of the fitted model in each subset.

3 Conclusion

This article provides a comprehensive overview of the literature concerning three-stage distributed Bayesian approaches. It is crucial to note that the inferences drawn from the distributed Bayesian framework are contingent upon the selection of data subsets. While both random and stratified partitioning yield similar predictive inferences, they may diverge in terms of parameter inference, depending on the characteristics of the true random fields. To address the sensitivity of inference arising from data partitioning, recent efforts have introduced a data sketching approach. Specifically, the partitioning of data has been substituted with multiple random sketches of the data. By implementing the third stage of distributed Bayesian inference over various random sketches, this approach helps mitigate the sensitivity of inference due to the selection of data sketches. The author envisions that the simplicity and generality of distributed Bayesian approaches will help solving scalability of many important inferential problems, such as to solve the serious scalability issues in Bayesian density estimation problems, spatio-temporal point process modeling.

References

Agueh, M. and G. Carlier (2011). Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis 43*(2), 904–924.

Bickel, P. J. and D. A. Freedman (1981). Some asymptotic theory for the bootstrap. *The annals*

of statistics 9(6), 1196–1217.

Chandler, R. E. and S. Bate (2007). Inference for clustered data using the independence loglikelihood. *Biometrika 94*(1), 167–183.

Cuturi, M. and A. Doucet (2014). Fast computation of wasserstein barycenters. In *International conference on machine learning*, pp. 685–693. PMLR.

Guhaniyogi, R. and S. Banerjee (2018). Meta-kriging: scalable Bayesian modeling and inference for massive spatial datasets. *Technometrics 60*(4), 430–444.

Guhaniyogi, R. and S. Banerjee (2019). Multivariate spatial meta kriging. *Statistics & probability letters 144*, 3–8.

Guhaniyogi, R., C. Li, T. Savitsky, and S. Srivastava (2023). Distributed bayesian inference in massive spatial data. *Statistical science 38*(2), 262–284.

Guhaniyogi, R., C. Li, T. D. Savitsky, and S. Srivastava (2022). Distributed bayesian varying coefficient modeling using a gaussian process prior. *The Journal of Machine Learning Research 23*(1), 3642–3700.

Heaton, M. J., A. Datta, A. O. Finley, R. Furrer, J. Guinness, R. Guhaniyogi, F. Gerber, R. B. Gramacy, D. Hammerling, M. Katzfuss, et al. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics 24*, 398–425.

Li, C., S. Srivastava, and D. B. Dunson (2017). Simple, scalable and accurate posterior interval estimation. *Biometrika 104*(3), 665–680.

Minsker, S. (2015). Geometric median and robust estimation in banach spaces.

Minsker, S., S. Srivastava, L. Lin, and D. Dunson (2014). Scalable and robust bayesian inference via the median posterior. In *International conference on machine learning*, pp. 1656–1664. PMLR.

Rabinovich, M., E. Angelino, and M. I. Jordan (2015). Variational consensus monte carlo. *Advances in Neural Information Processing Systems 28*.

Rendell, L. J., A. M. Johansen, A. Lee, and N. Whiteley (2020). Global consensus monte carlo. *Journal of Computational and Graphical Statistics 30*(2), 249–259.

Ribatet, M., D. Cooley, and A. C. Davison (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 813–845.

Scott, S. L., A. W. Blocker, F. V. Bonassi, H. A. Chipman, E. I. George, and R. E. McCulloch (2022). Bayes and big data: The consensus monte carlo algorithm. In *Big Data and Information Theory*, pp. 8–18. Routledge.

Srivastava, S., V. Cevher, Q. Dinh, and D. Dunson (2015). WASP: Scalable Bayes via barycenters of subset posteriors. In *Artificial Intelligence and Statistics*, pp. 912–920. PMLR.

Srivastava, S., C. Li, and D. B. Dunson (2018). Scalable bayes via barycenter in wasserstein space. *The Journal of Machine Learning Research 19*(1), 312–346.

Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica*, 5–42.

Xue, J. and F. Liang (2019). Double-parallel monte carlo for bayesian analysis of big data. *Statistics and computing 29*(1), 23–32.