Identifying Under-Reported Events in Networks with Spatial Latent Variable Models

Gabriel Agostini [1] **Emma Pierson** [*1] **Nikhil Garg** [*1]

Abstract

Decision-makers often observe the occurrence of events through a reporting process. City governments, for example, rely on resident reports to register and then resolve urban infrastructural problems such as fallen street trees, over-flooding sewers, or rat infestations. In the absence of additional assumptions, events that occur but are not reported cannot be distinguished from events that truly did not occur, leading to systematic neglect in addressing problems in neighborhoods that comparatively under-report events. In this paper, we leverage a Bayesian model to describe this setting in the presence of *network correlations* in the event occurrence process. We present a sampling routine to estimate the report rates and the event occurrence incidence, as well as infer the ground truth of discrete latent states. We apply the model to flooding reports in New York City, publicly available via the 311 data portal.

1. Introduction

Training data in real-world classification problems is often not fully labeled. For example, although interactions in social media tell whether a user enjoys a particular type of content, many platforms do not have an explicit way for expressing "dislike" so that not observing a positive interaction between the user and some content could either mean that the user dislikes the content or has not seen it. The data is split into two classes: positively labeled datapoints and unlabeled datapoints. Points in the latter class could be unlabeled either because they were not classified at all or because they were classified as negative. In machine learning, *positive-unlabeled (PU) learning* methods attempt to solve this group of problems (Liu et al., 2003; Shanmugam & Pierson, 2021).

Without further assumptions, the proportion of true positive points — the *prevalence* — is unidentifiable because a positive-class, unlabeled datapoint is indistinguishable from a negative-class datapoint. Hence, many PU learning methods either require prevalence as an input or make further assumptions (Bekker & Davis, 2020). For example, a common assumption is that each true positive point has the same uniform probability of being labeled positive (Elkan & Noto, 2008). Even this strong assumption, which often does not hold in real-world settings where the labeling probability is non-uniform, is alone not sufficient to identify the model.

One such problematic setting of PU learning, which we study in this paper, occurs in urban crowdsourcing systems. City governments rely on residents to report issues they experience, as ubiquitous inspections would be too costly. Examples of issues that decision-makers mostly get to know through reports are power outages, pest infestations, or street floods. These reports, however, are scarce and heterogeneous: many authors have studied inequity in 311 report rates, often concluding that factors such as race and home ownership contribute significantly towards reporting (Liu & Garg, 2023; Kontokosta et al., 2017; Minkoff, 2016; O'Brien et al., 2017). Without further assumptions on the prevalence of such issues or on their heterogeneous report rates, these parameters are unidentifiable.

As it is common to urban phenomena, most of these issues are spatially correlated: neighboring regions are likely to suffer from the same problems. We focus on *street floods*, events that are correlated across space due to the occurrence of an exogenous incident like a hurricane. Therefore, our PU learning setting differs from the case typically studied in PU learning, where the datapoints are assumed to be drawn independently. In this paper, we leverage a Bayesian model to describe the reporting rate and the probability of a flood event at multiple locations under the assumption that events are correlated across adjacent areas.

2. Related Work

Some Bayesian models for spatial latent variable settings have been developed in the context of crowdsourcing. Some

[*] Equal contribution [1] Cornell Tech, New York, United States. Correspondence to: Gabriel Agostini <gsagostini@infosci.cornell.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

works use weighted regression models to allocate "importance" for citizen expertise levels on their seemingly independent classifications (Peterson et al., 2020). Specific to 311 data, Liu & Garg (2023) show that time-stamped, duplicate reports about the same event can be used to identify the parameters of the reporting process, also without access to ground truth information; in contrast we leverage reports that are not duplicates but rather spatially correlated.

Applied ecologists have looked at the problem of estimating under-reported ground-truth event indicators in the context of species distributions (Heikkinen & Hogmander, 1994; Sicacha-Parada et al., 2021; Della Rocca & Milanesi, 2022). Humans, upon surveying a spatial area, report whether or not they noticed an animal of that species in the region. These reports are often not a ground truth, as the animals could have been away or hiding. Assuming reports that are spatially correlated, Santos-Fernandez et al. (2021) fit a Bayesian model to correct for misreporting errors in coral detection, where the latent variables are continuous proportions rather than discrete event indicators as the class labels in the PU setting.

Most relevant to our work is that of Spezia et al. (2018). They present a model for animal species presence inspired by statistical mechanics—the Ising Model—and a Bernoulli reporting process. The current version of our work applies their model to our setting.

Finally, our work relates to a much broader literature on methods to quantify and compensate for the effects of missing and imperfect data in inequality-related contexts, including healthcare, policing, education, and government inspections (Coston et al., 2021; Rambachan et al., 2021; Movva et al., 2023; Franchi et al., 2023; Laufer et al., 2023; Guerdan et al., 2023; Zink et al., 2023; Cai et al., 2020; Pierson, 2020; Obermeyer et al., 2019; Kleinberg et al., 2018; Zanger-Tishler et al., 2023; Jung et al., 2018; Garg et al., 2021; Lakkaraju et al., 2017; Arnold et al., 2022). This broader literature considers many types of missingness besides the PU-missingness we study here, and many types of identification approaches besides the spatial correlations leveraged here.

3. Model

Consider a network G whose vertices are indexed by integers 1 through N and whose adjacency matrix is given by E. There are two binary random variables describing the state of each node: $A_i \in \{-1, +1\}$, which describes the node's *latent*, *ground-truth* state, and $T_i \in \{0, 1\}$, which describes the node's *observed*, *reporting* state. In the flood setting, $A_i = 1$ if a flood occurred in that node and -1 if not, while $T_i = 1$ if there was at least one 311 report in that node and 0 if there was none. The key challenge is that we do not observe ground truth A_i, only reports T_i.

We make three assumptions about the nature of our problem:

(A1) **Markovian property of incidents:** the conditional distribution of each ground-truth state A_i given all other A_k is dependent only on the ground-truth states of node i's neighbors (j such that $j \sim i$).

$$\Pr(A_i \mid A_k \; \forall\, k \neq i) = \Pr(A_i \mid A_j \; \forall\, j \sim i)$$

(A2) **Conditional independence of reports:** the reporting states of two nodes are independent given the ground-truth states.

$$T_i \perp T_j \mid A_i \;\; \forall i \neq j$$

(A3) **PU property:** there are no false positive reports.

$$\Pr\left(T_i = 1 \mid A_i = -1\right) = 0$$

The occurrence of incidents follows an Ising Model after Spezia et al. (2018). There are two real-valued parameters, θ_0 and θ_1, controlling respectively the event incidence rate and the event correlation. The probability distribution of the vector $\vec{A} \in \{\pm 1\}^N$ is:

$$\begin{aligned}\Pr(\vec{A}) &= \frac{q(\vec{A} \mid \theta_0, \theta_1)}{\mathcal{Z}(\theta_0, \theta_1)} \\ &= \frac{\exp\left(\theta_0 \sum_i A_i + \theta_1 \sum_{i,j} A_i A_j \cdot E_{ij}\right)}{\mathcal{Z}(\theta_0, \theta_1)} \end{aligned} \tag{1}$$

In Equation (1), the denominator $\mathcal{Z}(\theta_0, \theta_1)$ is the distribution's normalization constant (i.e., the *partition function*). This constant is intractable, as it must be evaluated for 2^N values of the ground-truth state vector $\vec{A}$:

$$\mathcal{Z}(\theta_0, \theta_1) = \sum_{A_1 \in \{\pm 1\}} \cdots \sum_{A_N \in \{\pm 1\}} q(\vec{A} \mid \theta_0, \theta_1) \tag{2}$$

The report states T_i follow, due to (A2) and (A3), a Bernoulli distribution conditioned on their ground-truth state. The reporting process is controlled by a parameter $\psi \in (0, 1)$.

$$T_i \mid A_i \;\sim\; \text{Ber}\left(\psi \cdot \frac{A_i + 1}{2}\right) \tag{3}$$

The parameter vector of our model is then $\vec{\Theta} = (\psi, \theta_0, \theta_1)$. We impose priors on the parameters that are coherent with the 311 flood reporting setting. We assume that the spatial correlation θ_1 is slightly positive, but are more liberal with our treatment of the event occurrence parameter θ_0 and center it at zero. The priors we use are:

$$\psi \sim \text{Beta}(1.2, 0.8)$$

$$\theta_0 \sim \mathcal{N}(0, 0.2) \quad \theta_1 \sim \mathcal{N}(0.3, 0.1)$$

3.1. Sampling Routine

We use a Gibbs sampling MCMC to conduct posterior inference: namely, at each iteration, we draw each latent variable from its conditional distribution given the current values of all the other variables. All variables are initialized at random.

Sampling A_i: We sample each of the A_i in a random order (re-sorted at every iteration). As proven by Besag (1974), the conditional distributions have the form:

$$\Pr(A_i = a \mid \vec{A}_k \;\forall\, k \neq i) = \frac{1}{1 + \exp(2 \cdot a \cdot \tau_i)} \tag{4}$$

The exponent τ_i aggregates the states only of node i's neighbors, in accordance with (A1):

$$\tau_i = \theta_0 + \theta_1 \sum_{j \sim i} A_j \tag{5}$$

The conditional distribution of A_i given all the other variables is dependent only on $\vec{\Theta}$, T_i, and A_j for all $j \sim i$. If the corresponding $T_i = 1$, the assumption (A3) forces that $A_i = 1$. If $T_i = 0$, then we sample A_i with:

$$\Pr(A_i \mid T_i, \vec{\Theta}, A_j \forall\, j \sim i) \propto \begin{cases} \frac{1-\psi}{1+\exp(-2\cdot\tau_i)} & A_i = 1 \\ \frac{1}{1+\exp(2\cdot\tau_i)} & A_i = -1 \end{cases} \tag{6}$$

Sampling ψ: The conditional distribution of ψ given all other variables depends only on the number of incidents which occurred but not reported n_0 and the number of occurred events that was reported n_1. We sample ψ from a Beta posterior distribution, updating our prior parameters:

$$\psi \sim \text{Beta}\,(1.2 + n_1, 0.8 + n_0) \tag{7}$$

Sampling θ_0 and θ_1: The conditional distribution of θ_0 and θ_1 given all other variables depends only on $\vec{A}$. We cannot directly compute it due to the presence of the partition function $\mathcal{Z}$ (Murray et al., 2006). We use the Single-Variable Exchange Algorithm (SVEA) to circumvent this difficulty (Møller et al., 2006).

The SVEA is a Metropolis-Hasting type sampling algorithm that introduces an auxiliary variable $\vec{w}$ to cancel two terms with the partition function when computing the acceptance ratio. To do so, $\vec{w}$ must be sampled from the same distribution family as $\vec{A}$. We generate auxiliary variables from the distribution in eq. 1 using the Swendsen-Wang algorithm (Swendsen & Wang, 1987; Wolff, 1989) when possible. This is an efficient method to sample from an Ising Model with positive spatial correlation θ_1 (Park et al., 2017; Cooper et al., 2000). We sample $\vec{w}$ using Gibbs sampling if at the current step $\theta_1 \leq 0$.

4. Results

We present our current results in two sections. First, we evaluate our modeling approach and fitting procedure on synthetic data, showing that we can indeed recover ground truth parameters. Second, we apply the model to real data from resident reports of street floods in New York City during the passage of Hurricane Ida.

4.1. Model Calibration on Synthetic Data

We generate synthetic data from the model with parameters sampled from our prior distributions for $n = 25$ experiments. In each experiment, we run two independent chains. The convergence diagnostics for one such experiment along with hyper-parameters is shown in Figure 3.

Across these experiments, we calculate the proportion of times that the true parameter value falls within a given confidence interval of the inferred posterior distribution, for each of $5, 10, \ldots 95\%$ confidence intervals. As Figure 1 shows, we verify that the model is *calibrated*: at any given confidence level the corresponding confidence interval contain the true parameters at the expected rate.

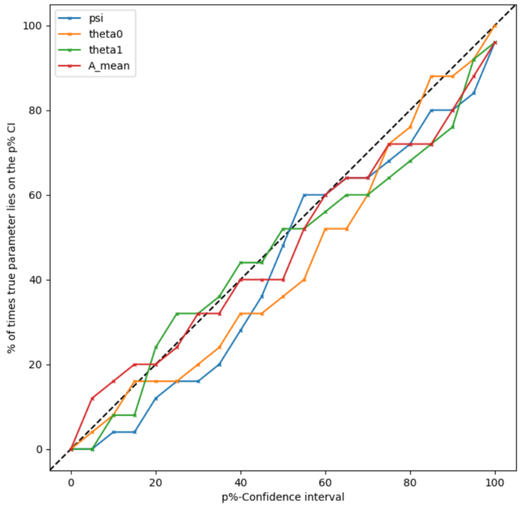
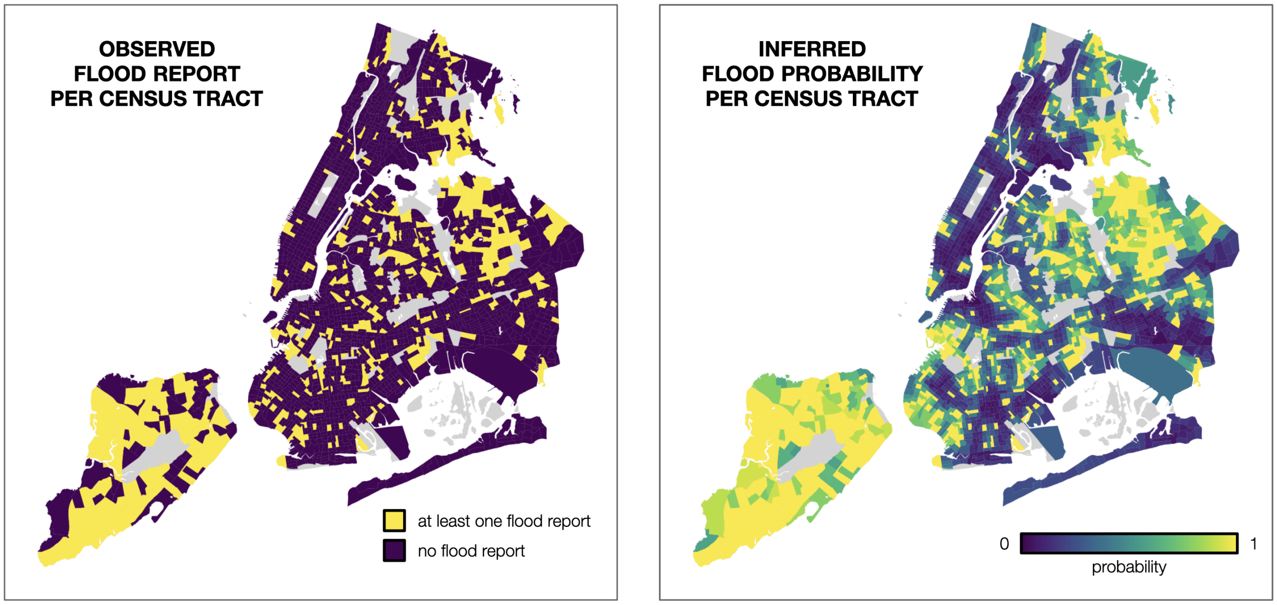


Figure 1. Calibration of the model. A point (x, y) is marked if in $y\%$ of the experiments ($n = 25$) the true parameter value was contained in the $x\%$ percentile of the inferred posterior distribution. Perfect calibration is given by the identity line, which our model approximates very well.

We verify that we can recover the true parameters in the majority of the experiments, as shown in Figure 4.

4.2. Performance on 311 Data

We use data on 311 reports during the first week of September 2021 (NYC Open Data, 2023), following the passage of Hurricane Ida in New York City. The reports are mapped



(a) Reports per census tract i.e. T_i. The reports are scarce (about 20% of tracts have a positive report) and show signal of positive spatial correlation.

(b) Flood probability per census tract i.e. $\Pr(A_i = 1)$. Tracts where the reports happened will by construction have probability of flood equal to 1.

Figure 2. Map of census tracts in New York City with their (a) observed report status and (b) inferred probability of flooding by our model between September 1st and 8th, 2021. Adjacent tracts to those with positive reports seem to have high probability of flooding, as the model pools together information from neighboring areas. Census tracts in grey were ignored.

per census tract (our unit of analysis) in Figure 2. We note that this section is a work in progress – as we discuss, the graph structure substantially affects model inference, in a manner that requires further work.

The spatial structure is given by the adjacency of census tracts in New York City. Two tracts are neighbors if they share a boundary. We remove nodes with degree greater than 10 to account for outlier census tracts—mostly parks with no population, which span a significant area of the city and border many tracts. We found that the graph structure resulting from this process requires further considerations when choosing priors to adequately explore the entire parameter space. As shown in Figure 5, the average mean of nodes with incidents is more sensitive to spatial correlation parameter θ_1 in the real graph than it is on a regular grid graph with comparable number of nodes. In other words, for the same (positive) value of the spatial correlation θ_1, real-world graphs tend to be more homogeneous (all nodes are either 1 or -1) than synthetic graphs for every value of the prevalence θ_0.

With our model we estimate the true parameters. The convergence diagnostics and the full posteriors for these results are shown in Figure 6. The mean and standard deviations estimated for the three parameters are shown in Table 1.

	mean	standard deviation
ψ	0.43	0.04
θ_0	-0.01	0.01
θ_1	0.26	0.02

Table 1. Estimated means and standard deviations for the parameters (pooling inferred posteriors from all chains).

Our inferred parameter estimates are plausible for our application. The spatial correlation θ_1 is high, which agrees with our assumption that flooding is correlated between adjacent areas. The event prevalence parameter θ_0 is around zero, meaning that census tract is as likely to get flood as it is not to (ignoring correlations). The reporting rate ψ is estimated to be around 0.4, which incurs scarcity in the observed reporting numbers. However, given the results in Figure 5 about the effect of the graph structure, it is unclear whether these results are a consequence of the priors and graph structure – in future work, we will validate and improve these inferences.

Finally, we estimate the probability that a given census tract is flooded, that is $\Pr(A_i = 1)$. The probability is computed as the proportion of post burn-in samples in which $A_i = 1$ for node i. This map is also shown in Figure 2.

5. Future Work

This work shows the promise of using *spatial correlation* of incidents (such as flooding events and rats) to diagnose (1) under-reporting of events in municipal resident crowdsourcing systems; and (2) in turn, learn incident *ground truth* in the presence of such under-reporting. However, substantial work remains to be done to realize this promise.

Most immediately, our current results require further validation for robustness and we hypothesize that choosing the correct graph structure will improve model fit. The impact of θ_1 joint with the census graph structure in the average proportion of positive class nodes suggests that a more polished, grid-like graph structure could capture fluctuations in the true parameters more precisely. In particular, to attain a fraction of more than 0.2 nodes with ground truth $A_i = 0$ (resp. $A_i = 1$) in a regime where the spatial correlation parameter θ_1 is around 0.25, the propensity parameter θ_0 will likely be inferred close to zero—the result we obtained. As we move forward with this work, we are working on replicating our experiments using geohashes and other grid-like graph structures.

Second, we are working to model the heterogeneity in reporting rates through a vector $\vec{\psi}$ which varies across nodes as opposed to a constant. For each node we observe a vector $\vec{X}_i \in \mathbb{R}^K$ of K demographic features (predominant race, income, education level, home-ownership status). Then, at every step of our MCMC, we estimate a vector $\vec{\beta} \in \mathbb{R}^K$ with a Bayesian logistic regression of the form:

$$\psi_i = \text{logit}^{-1}\left(\alpha + \sum_{m=1}^{K} \beta_m X_{mi}\right).$$

References

Arnold, D., Dobbie, W., and Hull, P. Measuring racial discrimination in bail decisions. *American Economic Review*, 112(9):2992–3038, 2022.

Bekker, J. and Davis, J. Learning from positive and unlabeled data: a survey. *Machine Learning*, 109(4):719–760, April 2020. ISSN 1573-0565. doi: 10.1007/s10994-020-05877-5. URL https://doi.org/10.1007/s10994-020-05877-5.

Besag, J. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225, January 1974. ISSN 00359246. doi: 10.1111/j.2517-6161.1974.tb00999.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1974.tb00999.x.

Cai, W., Gaebler, J., Garg, N., and Goel, S. Fair allocation through selective information acquisition. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 22–28, 2020.

Cooper, C., Dyer, M. E., Frieze, A. M., and Rue, R. Mixing properties of the Swendsen–Wang process on the complete graph and narrow grids. *Journal of Mathematical Physics*, 41(3):1499–1527, March 2000. ISSN 0022-2488, 1089-7658. doi: 10.1063/1.533194. URL https://pubs.aip.org/aip/jmp/article/41/3/1499-1527/462339.

Coston, A., Rambachan, A., and Chouldechova, A. Characterizing fairness over the set of good models under selective labels. In *International Conference on Machine Learning*, pp. 2144–2155. PMLR, 2021.

Della Rocca, F. and Milanesi, P. The New Dominator of the World: Modeling the Global Distribution of the Japanese Beetle under Land Use and Climate Change Scenarios. *Land*, 11(4):567, April 2022. ISSN 2073-445X. doi: 10.3390/land11040567. URL https://www.mdpi.com/2073-445X/11/4/567. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

Elkan, C. and Noto, K. Learning classifiers from only positive and unlabeled data. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, pp. 213–220, New York, NY, USA, August 2008. Association for Computing Machinery. ISBN 978-1-60558-193-4. doi: 10.1145/1401890.1401920. URL https://doi.org/10.1145/1401890.1401920.

Franchi, M., Zamfirescu-Pereira, J., Ju, W., and Pierson, E. Detecting disparities in police deployments using dashcam data. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pp. 534–544, 2023.

Garg, N., Li, H., and Monachou, F. Standardized tests and affirmative action: The role of bias and variance. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 261–261, 2021.

Guerdan, L., Coston, A., Holstein, K., and Wu, Z. S. Counterfactual prediction under outcome measurement error. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pp. 1584–1598, 2023.

Heikkinen, J. and Hogmander, H. Fully Bayesian Approach to Image Restoration with an Application in Biogeography. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43(4):569–582, 1994. ISSN 0035-9254. doi: 10.2307/2986258. URL https://www.jstor.org/stable/2986258. Publisher: [Wiley, Royal Statistical Society].

Jung, J., Corbett-Davies, S., Shroff, R., and Goel, S. Omitted and included variable bias in tests for disparate impact. *arXiv preprint arXiv:1809.05651*, 2018.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. Human decisions and machine predictions. *The quarterly journal of economics*, 133(1):237–293, 2018.

Kontokosta, C., Hong, B., and Korsberg, K. Equity in 311 Reporting: Understanding Socio-Spatial Differentials in the Propensity to Complain, October 2017. URL http://arxiv.org/abs/1710.02452. arXiv:1710.02452 [cs].

Lakkaraju, H., Kleinberg, J., Leskovec, J., Ludwig, J., and Mullainathan, S. The selective labels problem: Evaluating algorithmic predictions in the presence of unobservables. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 275–284, 2017.

Laufer, B., Pierson, E., and Garg, N. Detecting Disparities in Capacity-Constrained Service Allocations. 2023.

Liu, B., Dai, Y., Li, X., Lee, W., and Yu, P. Building text classifiers using positive and unlabeled examples. In *Third IEEE International Conference on Data Mining*, pp. 179–186, November 2003. doi: 10.1109/ICDM.2003.1250918.

Liu, Z. and Garg, N. Quantifying Spatial Underreporting Disparities in Resident Crowdsourcing, March 2023. URL http://arxiv.org/abs/2204.08620. arXiv:2204.08620 [cs, stat].

Minkoff, S. L. NYC 311: A Tract-Level Analysis of Citizen–Government Contacting in New York City. *Urban Affairs Review*, 52(2):211–246, March 2016. ISSN 1078-0874. doi: 10.1177/1078087415577796. URL https://doi.org/10.1177/1078087415577796. Publisher: SAGE Publications Inc.

Movva, R., Shanmugam, D., Hou, K., Pathak, P., Guttag, J., Garg, N., and Pierson, E. Coarse race data conceals disparities in clinical risk score performance. *arXiv preprint arXiv:2304.09270*, 2023.

Murray, I., Ghahramani, Z., and MacKay, D. J. C. Mcmc for doubly-intractable distributions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'06, pp. 359–366, Arlington, Virginia, USA, 2006. AUAI Press. ISBN 0974903922.

Møller, J., Pettitt, A. N., Reeves, R., and Berthelsen, K. K. An Efficient Markov Chain Monte Carlo Method for Distributions with Intractable Normalising Constants. *Biometrika*, 93(2):451–458, 2006. ISSN 0006-3444. URL https://www.jstor.org/stable/20441294. Publisher: [Oxford University Press, Biometrika Trust].

NYC Open Data. 311 Service Requests from 2010 to Present, 2023. URL https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9.

Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, 2019.

O'Brien, D. T., Offenhuber, D., Baldwin-Philippi, J., Sands, M., and Gordon, E. Uncharted Territoriality in Coproduction: The Motivations for 311 Reporting. *Journal of Public Administration Research and Theory*, 27 (2):320–335, April 2017. ISSN 1053-1858. doi: 10.1093/jopart/muw046. URL https://doi.org/10.1093/jopart/muw046.

Park, S., Jang, Y., Galanis, A., Shin, J., Stefankovic, D., and Vigoda, E. Rapid Mixing Swendsen-Wang Sampler for Stochastic Partitioned Attractive Models, April 2017. URL http://arxiv.org/abs/1704.02232. arXiv:1704.02232 [cs, stat].

Peterson, E. E., Santos-Fernández, E., Chen, C., Clifford, S., Vercelloni, J., Pearse, A., Brown, R., Christensen, B., James, A., Anthony, K., Loder, J., González-Rivero, M., Roelfsema, C., Caley, M. J., Mellin, C., Bednarz, T., and Mengersen, K. Monitoring through many eyes: Integrating disparate datasets to improve monitoring of the Great Barrier Reef. *Environmental Modelling & Software*, 124:104557, February 2020. ISSN 1364-8152. doi: 10.1016/j.envsoft.2019.104557. URL https://www.sciencedirect.com/science/article/pii/S1364815219309582.

Pierson, E. Assessing racial inequality in covid-19 testing with bayesian threshold tests. *NeurIPS ML4H Workshop*, 2020.

Rambachan, A. et al. Identifying prediction mistakes in observational data. *Harvard University*, 2021.

Santos-Fernandez, E., Peterson, E. E., Vercelloni, J., Rushworth, E., and Mengersen, K. Correcting Misclassification Errors in Crowdsourced Ecological Data: A Bayesian Perspective. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 70(1):147–173, January 2021. ISSN 0035-9254. doi: 10.1111/rssc.12453. URL https://doi.org/10.1111/rssc.12453.

Shanmugam, D. and Pierson, E. Quantifying inequality in underreported medical conditions. *arXiv preprint arXiv:2110.04133*, 2021.

Sicacha-Parada, J., Steinsland, I., Cretois, B., and Borgelt, J. Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. *Spatial Statistics*, 42:100446, April 2021. ISSN 2211-6753. doi: 10.1016/j.spasta.2020.100446. URL https://www.sciencedirect.com/science/article/pii/S2211675320300403.

Spezia, L., Friel, N., and Gimona, A. Spatial hidden Markov models and species distributions. *Journal of Applied Statistics*, 45(9):1595–1615, July 2018. ISSN 0266-4763, 1360-0532. doi: 10.1080/02664763.2017.1386771. URL https://www.tandfonline.com/doi/full/10.1080/02664763.2017.1386771.

Swendsen, R. H. and Wang, J.-S. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, 58(2):86–88, January 1987. doi: 10.1103/PhysRevLett.58.86. URL https://link.aps.org/doi/10.1103/PhysRevLett.58.86. Publisher: American Physical Society.

Wolff, U. Collective Monte Carlo Updating for Spin Systems. *Physical Review Letters*, 62(4):361–364, January 1989. doi: 10.1103/PhysRevLett.62.361. URL https://link.aps.org/doi/10.1103/PhysRevLett.62.361. Publisher: American Physical Society.

Zanger-Tishler, M., Nyarko, J., and Goel, S. Risk scores, label bias, and everything but the kitchen sink. *arXiv preprint arXiv:2305.12638*, 2023.

Zink, A., Obermeyer, Z., and Pierson, E. Race corrections in clinical models: Examining family history and cancer risk. *medRxiv*, pp. 2023–03, 2023.

A. Model Convergence Diagnostics

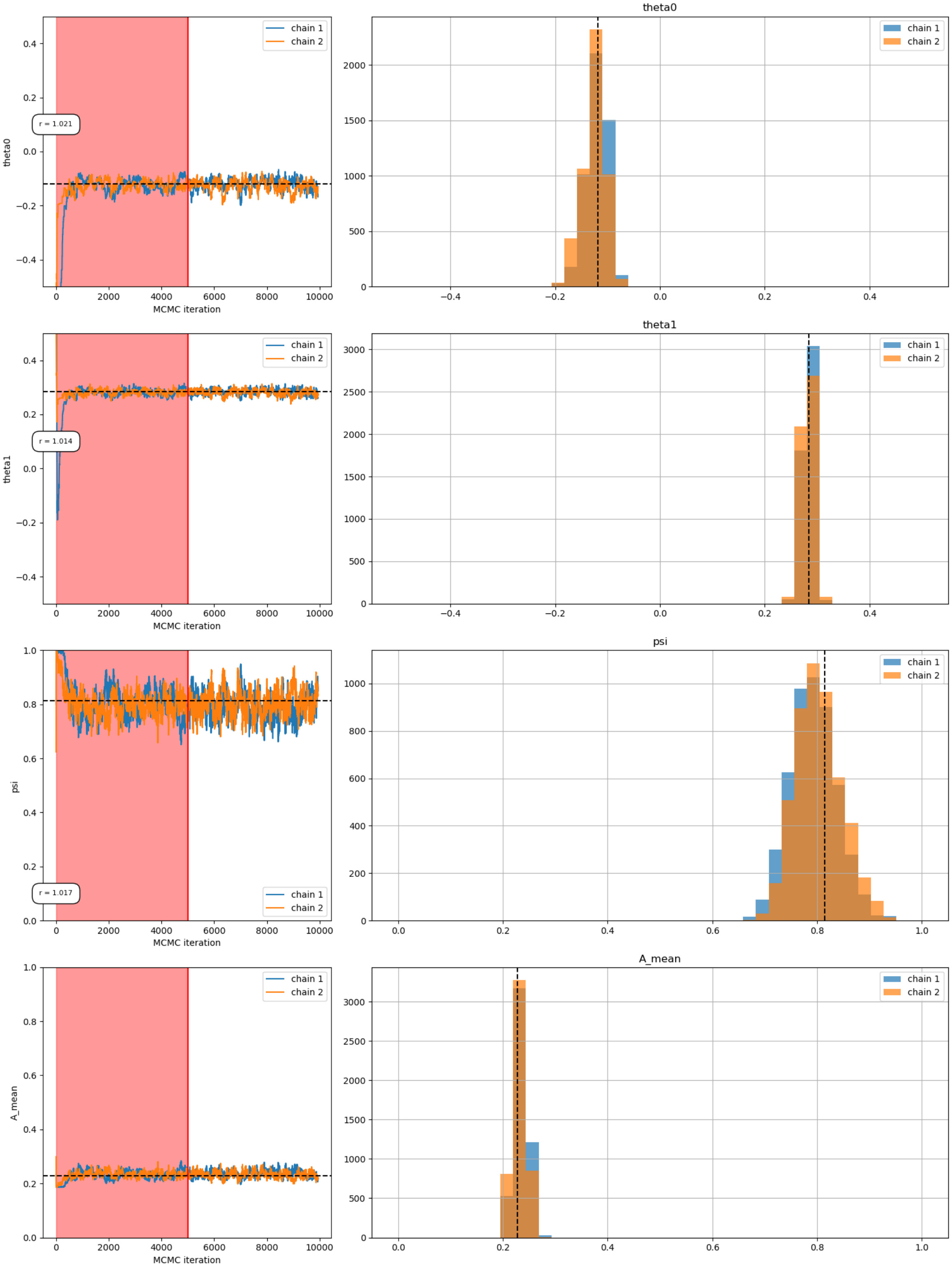


Figure 3. Trace plots and Inferred Posteriors for a synthetic data setting where $\psi = 0.81$, $\theta_0 = -0.12$, and $\theta_1 = 0.28$. Two independent chains were sampled for 15000 iterations with a thinning fraction of 0.5, and $\hat{r}$ convergence statistics are given in the figure. The red region in the trace plots delineates the burn-in period (5000 iterations), during which samples were discarded and the hyperparameters were tuned. The means of the estimated posteriors for ψ, $theta_0$, and θ_1 were respectively 0.80, -0.12, and 0.28—exactly the true parameters—with standard deviations 0.04, 0.02, and 0.01.

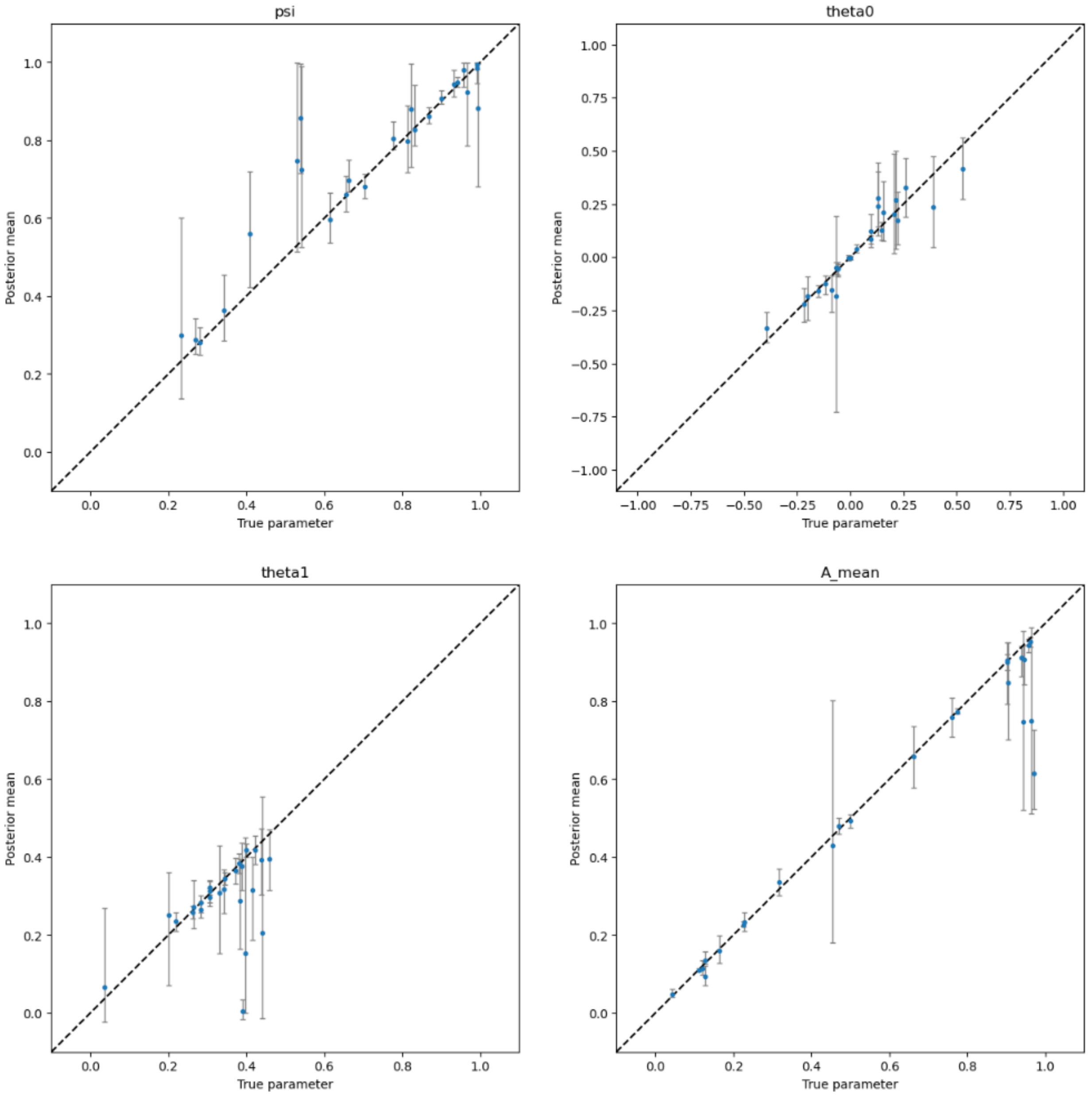


Figure 4. We recover the true parameters in the majority of the $n = 25$ synthetic data experiments. The y-axis shows the inferred means of the posteriors for ψ, θ_0, θ_1, and (the mean of) $\vec{A}$, while the x-axis shows the true value of those variables. Each experiment is then mapped to one point (posteriors aggregated across both chains), and the error bars represent 95% confidence intervals. The proximity of most points to the identity line show the model is recovering true parameters at the same rate through the parameter space.

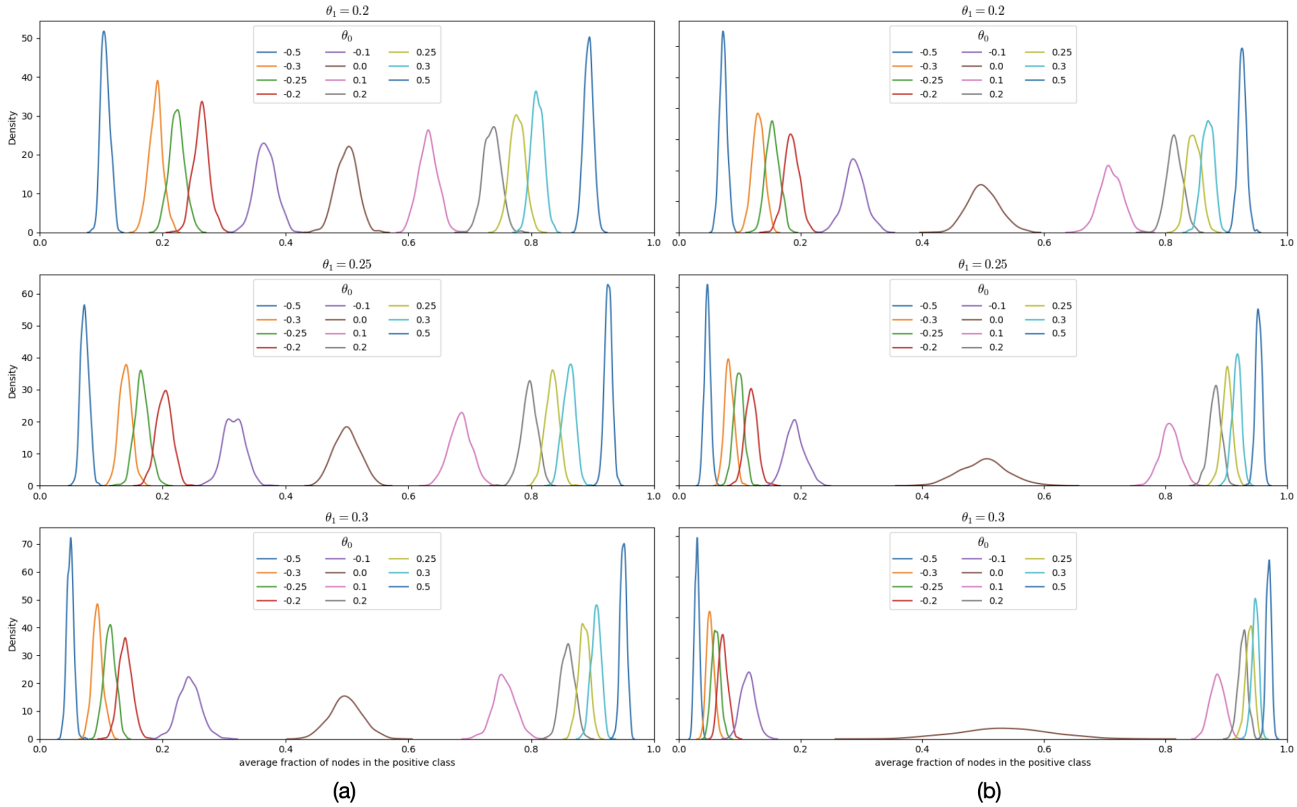


Figure 5. Average fraction of nodes with positive A_i for a given set of parameters θ_0 and θ_1 on **(a)** a grid graph with $N = 2500$ nodes, **(b)** the census graph with high-degree nodes removed ($N = 2274$ nodes). The distribution of $\vec{A}$ is highly sensitive to the values of θ_1 on the real world regime: as θ_1 increases, the vector $\vec{A}$ approaches a constant vector at either -1 or $+1$ depending on the signal of θ_0 (except when $\theta_0 = 0$). Averages were computed across 500 trials.

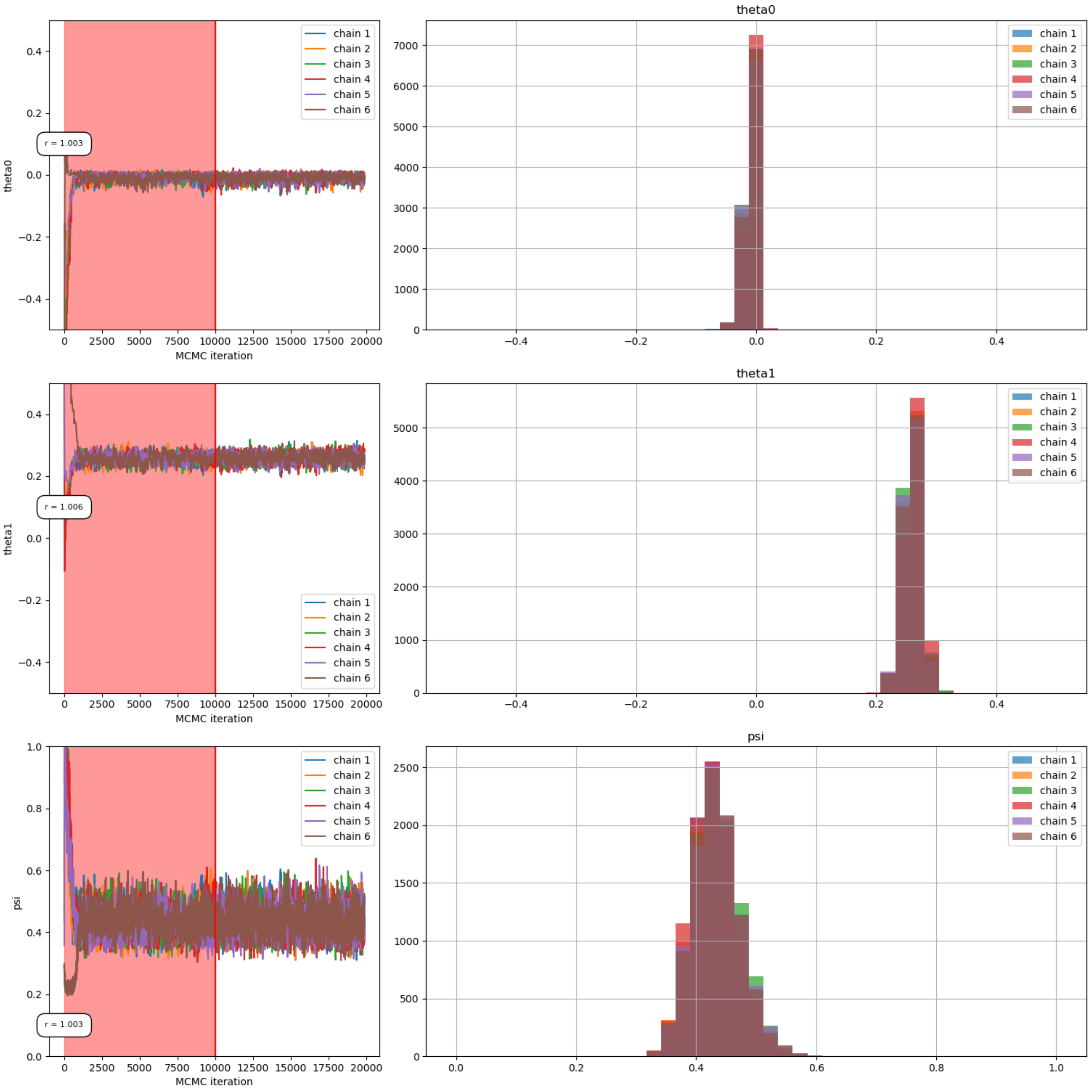


Figure 6. Trace plots and Inferred Posteriors for ψ, θ_0, and θ_1 in the real 311 data. Six independent chains were sampled for 25000 iterations with a thinning fraction of 0.5. The red region in the trace plots delineates the burn-in period (10000 iterations), during which samples were discarded and the hyperparameters were tuned.