Teodor Fredriksson
August 2, 2024

Supplementary Information

This section outlines how the Bayesian Data Analysis was performed. We follow the recommendations and guidlines of [1, 2, 3] amd asses the Markov Chain Monte Carlo (MCMC) chains and Posterior Predictive Checks.

1. Assesing the MCMC chains

All analysis were conducted using the `R` programming language and the `cmdstanr` library. Each model was computed using four parallel chains using 2000 iterations and an additional 200 warmup iterations. None of the iteration diverged as can be seen in the traceplots Figures 1, 2, 3, 4, 5, 6, 7, and 8 and therefore the MCMC simulation is considered valid.

In addition we investigate diagnostics of the posterior estimates, such as *Gelman-Rubin Potential Scale Reduction*($\widehat{R}$) [4] the number of *efficient samples* (n_{eff}). Tables 1, 2, 3, 4, 5, 6, 7, 8 contains three columns. The first column contains $\widehat{R}$ and the other two contains two different estiamtes of the number of effective samples. As a rule of thumb we should have $\widehat{R} < 1.01$ which is satisfied in all scenarios. The number of efficient samples should be atleast 200, which is satisfied in all cases.

Table 1. Diagnostics for the posterior ranks (aggregated)

Variable	Rhat	ess_bulk	ess_tail
a_alg[1]	1.000	11549.08	5515.995
a_alg[2]	1.000	14966.93	4698.992
a_alg[3]	1.001	11387.76	5792.326
a_alg[4]	1.000	13475.97	5838.628
a_alg[5]	1.000	13906.34	5240.604
a_alg[6]	1.000	13484.73	5755.981
a_alg[7]	1.000	13642.63	5674.657
a_alg[8]	1.002	13613.50	6151.843
a_alg[9]	1.001	11709.12	5973.553
a_alg[10]	1.000	13512.71	6059.267
a_alg[11]	1.001	12780.16	5426.069
a_alg[12]	1.000	13816.25	5731.842
a_alg[13]	1.000	13065.28	5764.492

2. Posterior Predictive Cheks

For posterior predictive checks, we investigate how well the mean error rate is captured by the posterior distribution. Figures 9, 10, 11, 12, 13, 14, 15, 16 shows the posterior distribution of the mean for each algorithm and the Figures show that the mean is well captured by the posterior distribution.

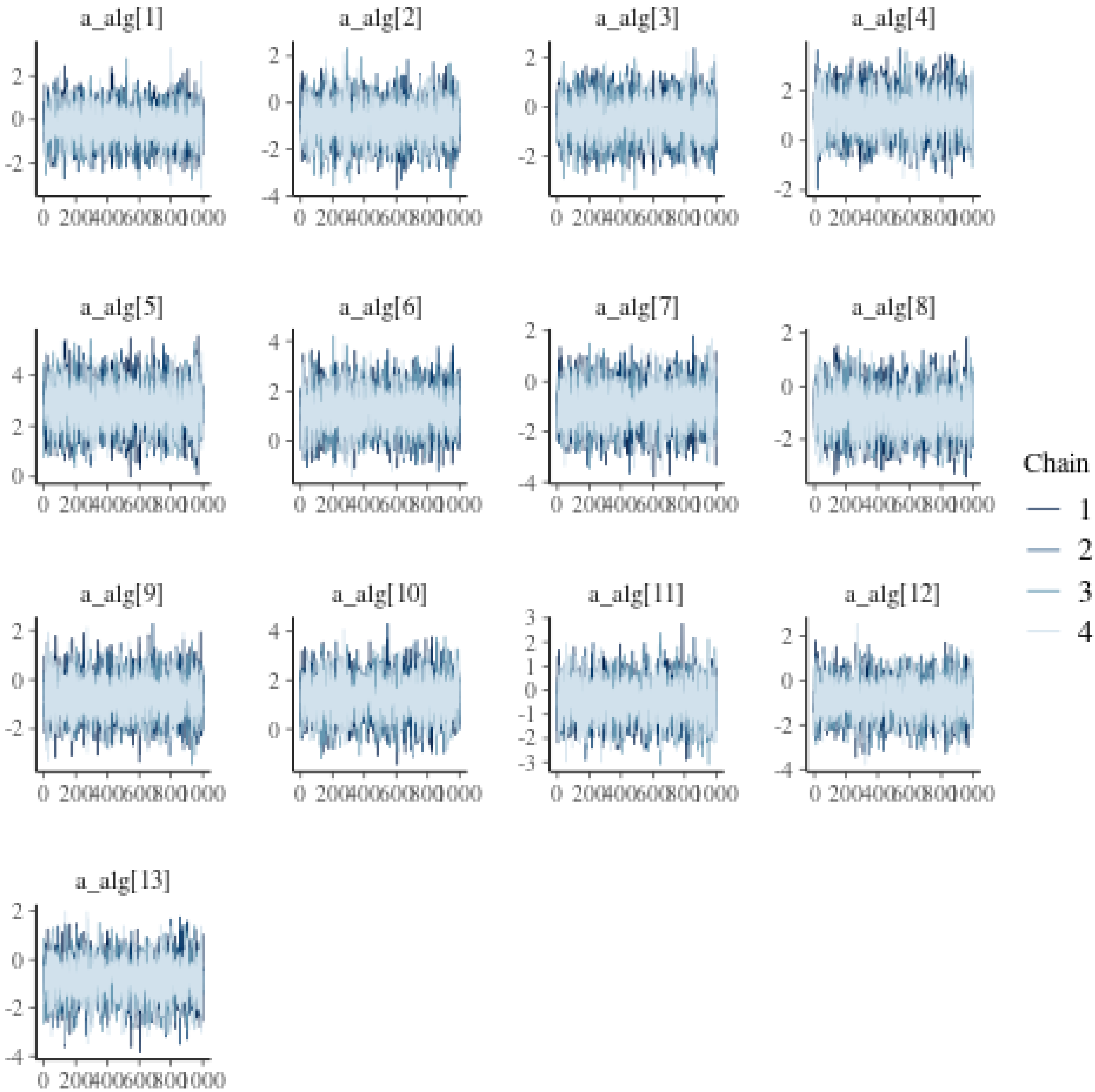

FIGURE 1. Simulated MCMC chains for the strength parameters (aggregated data).

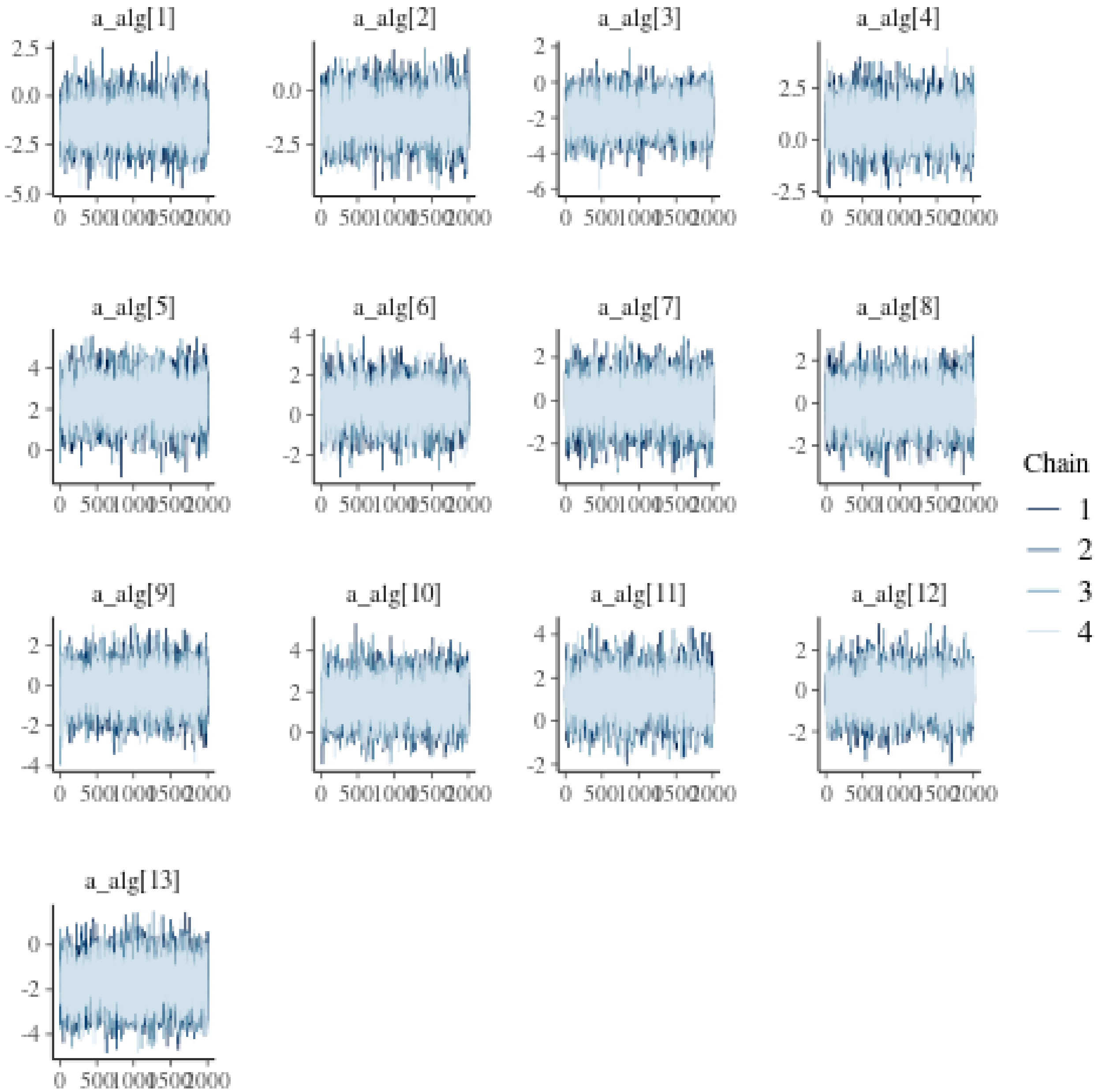

FIGURE 2. Simulated MCMC chains for the strength parameters (aggregated data).

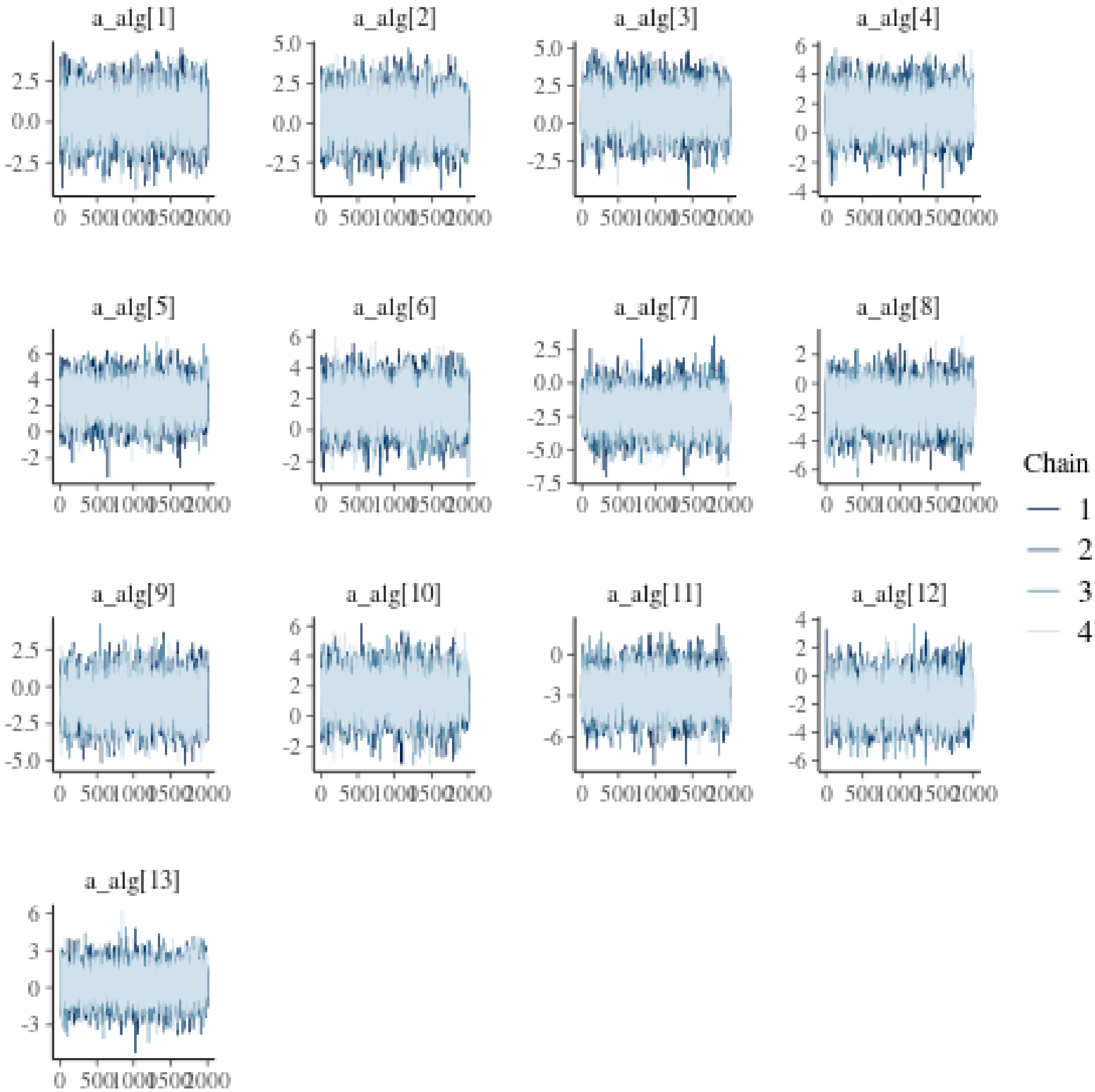

Figure 3. Simulated MCMC chains for the strength parameters (aggregated data).

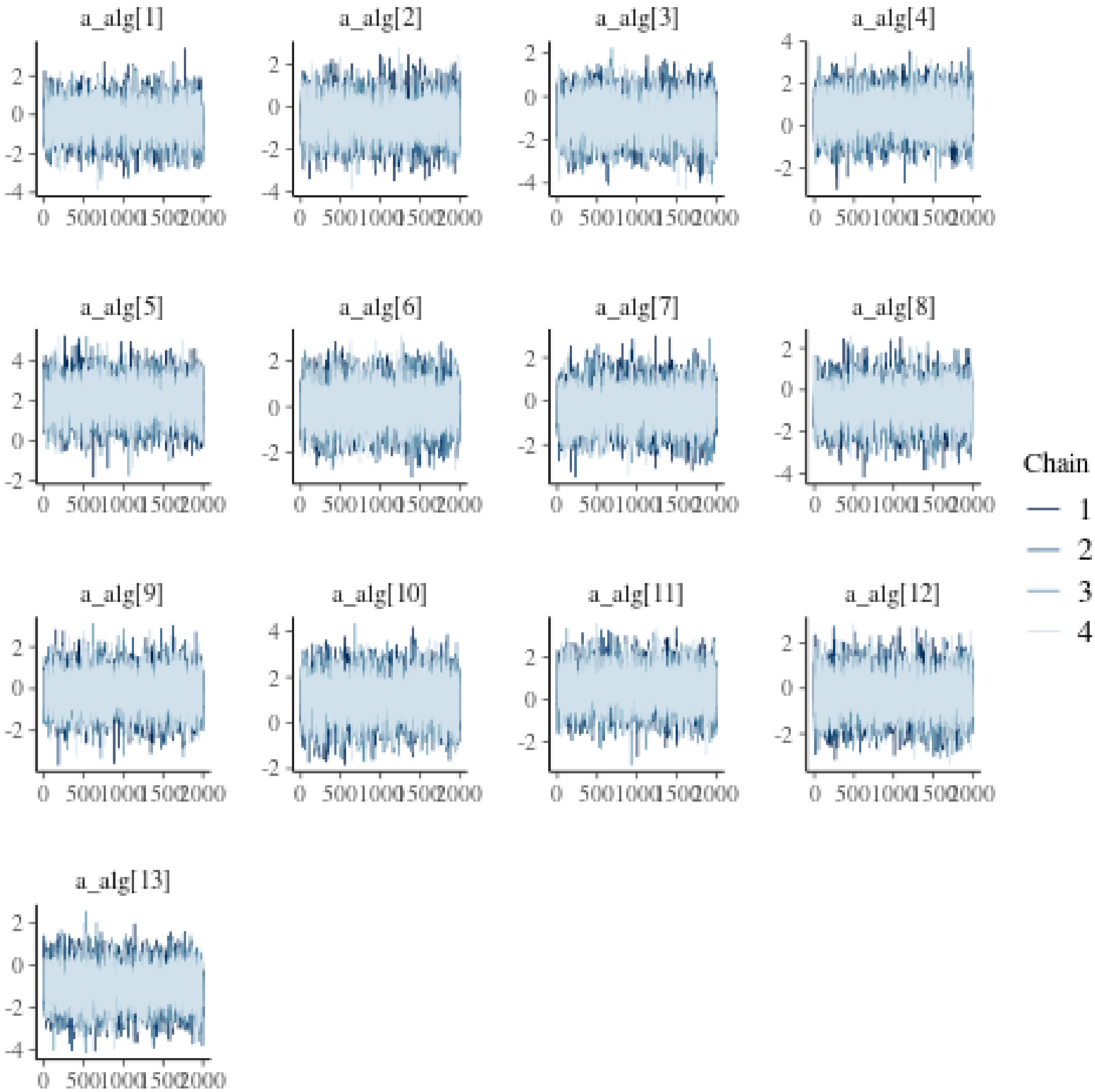

FIGURE 4. Simulated MCMC chains for the strength parameters (aggregated data).

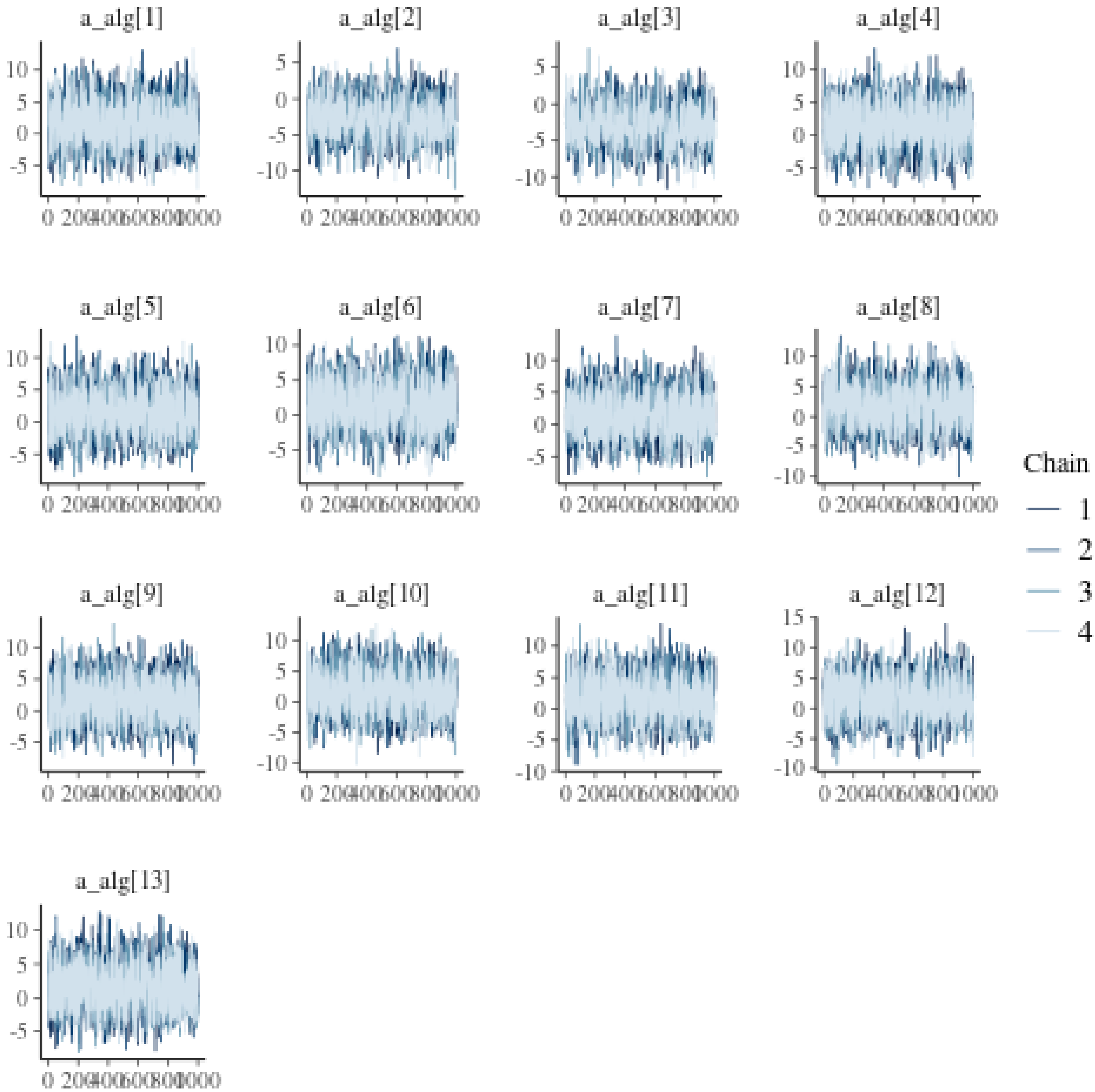

FIGURE 5. Simulated MCMC chains for the effect parameters (aggregated data).

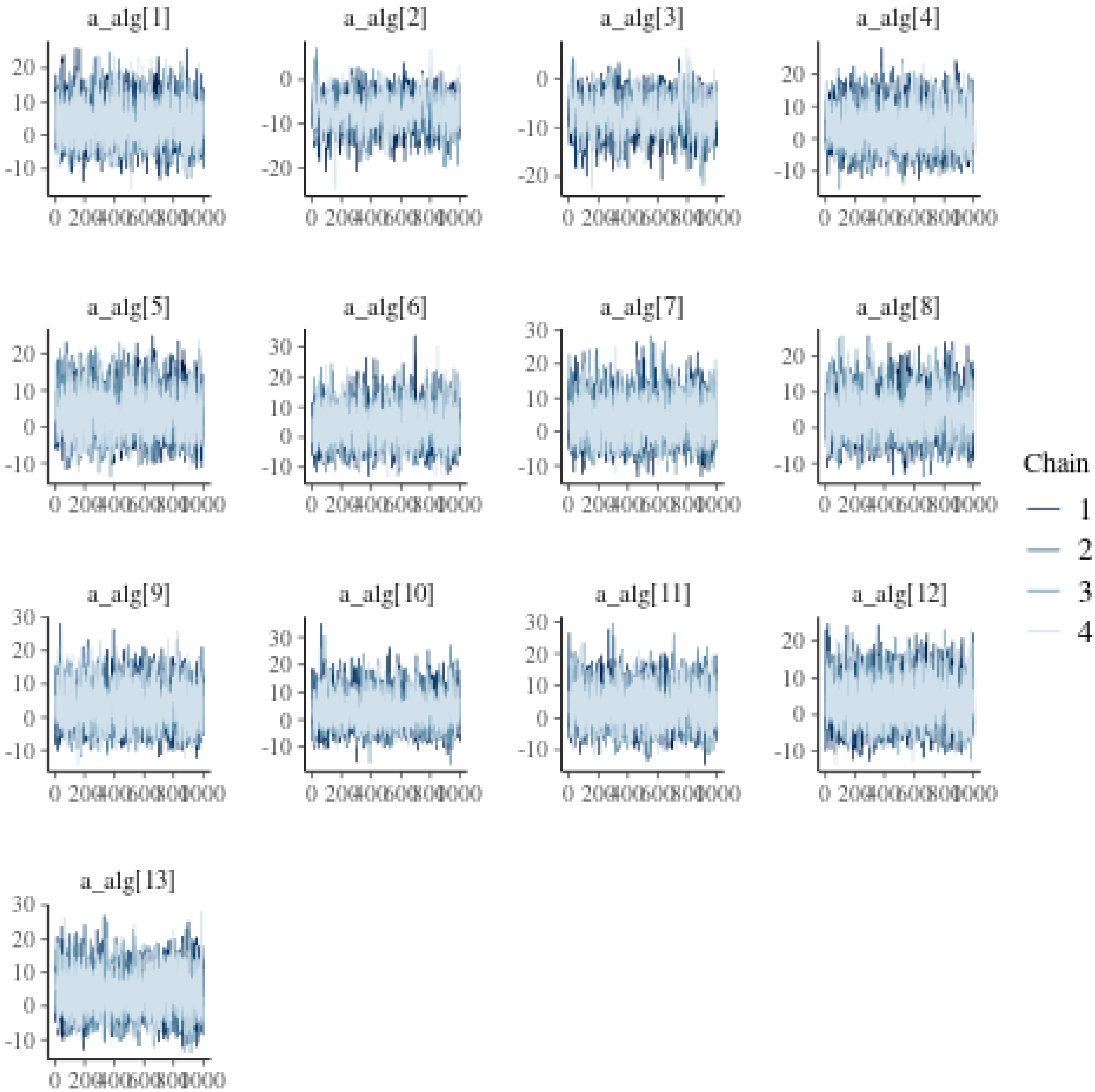

FIGURE 6. Simulated MCMC chains for the effect parameters (image data).

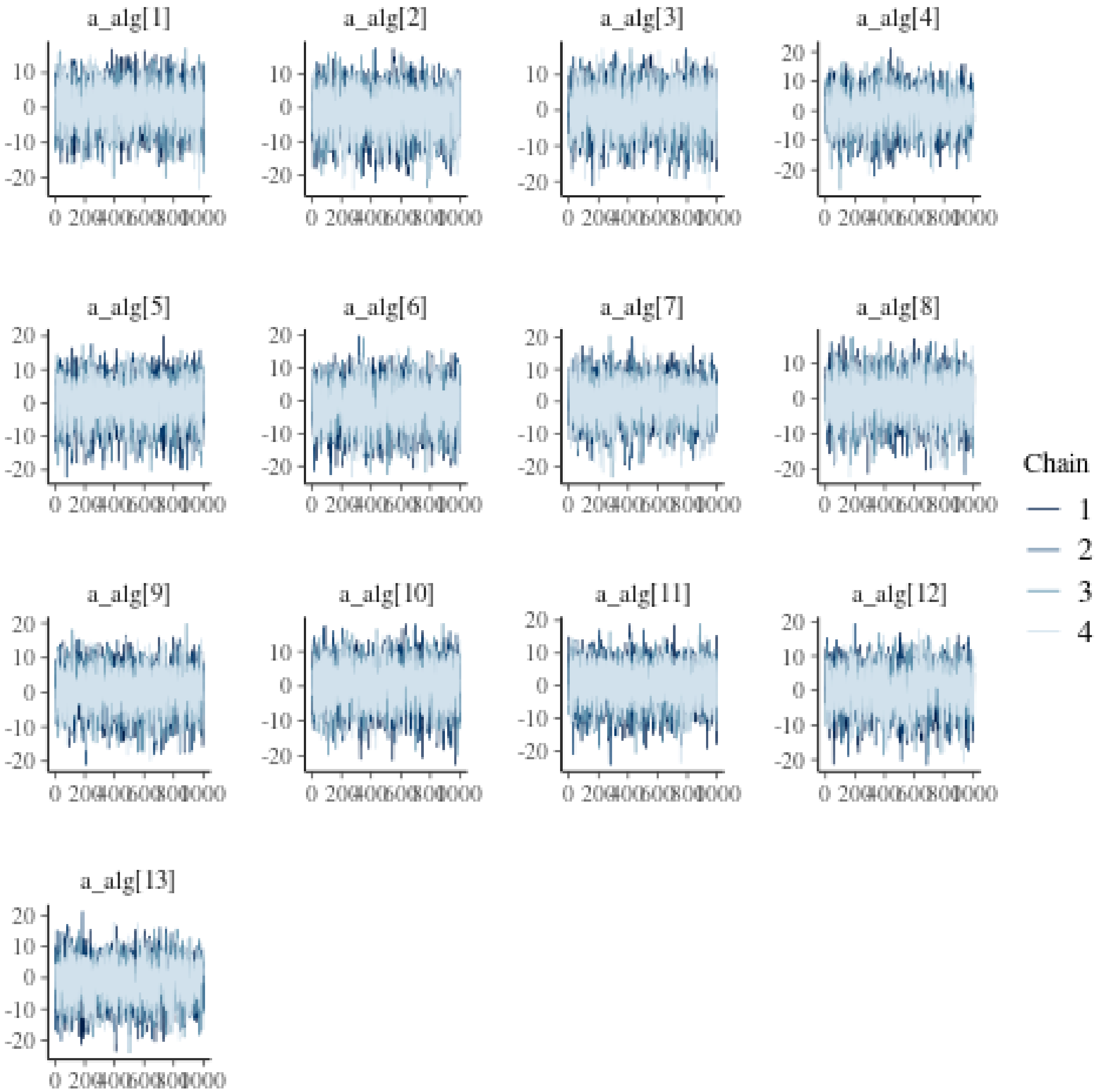

FIGURE 7. Simulated MCMC chains for the effect parameters (text data).

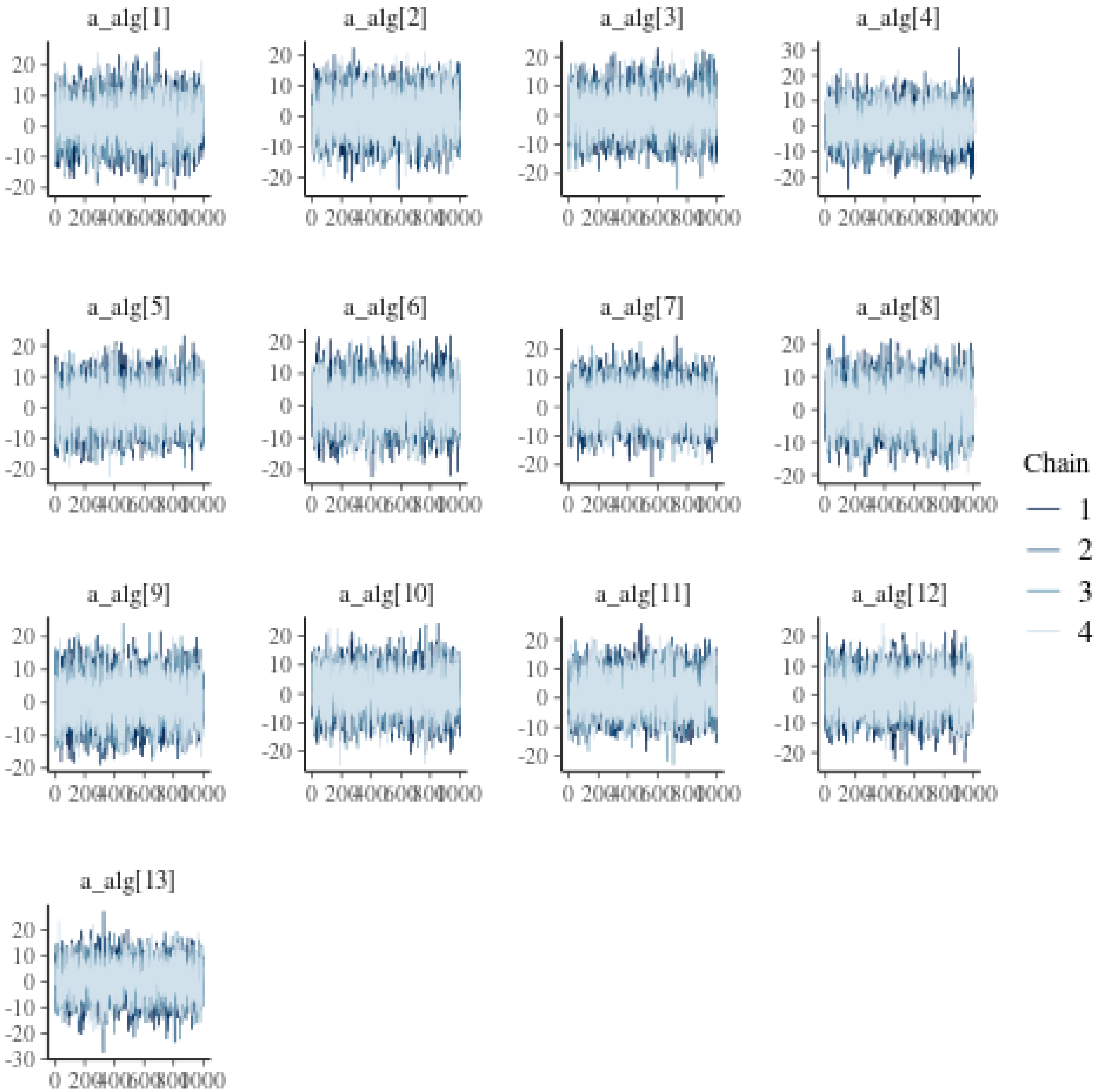

FIGURE 8. Simulated MCMC chains for the effect parameters (text data).

Table 2. Diagnostics for the posterior ranks (image data)

Variable	Rhat	ess_bulk	ess_tail
a_alg[1]	1.001	6843.458	5841.228
a_alg[2]	1.001	8010.731	6203.063
a_alg[3]	1.000	6983.924	6518.645
a_alg[4]	1.000	7532.715	6204.486
a_alg[5]	1.000	8215.520	5196.772
a_alg[6]	1.000	8507.300	5739.370
a_alg[7]	1.000	8227.085	5822.522
a_alg[8]	1.001	8274.565	5946.208
a_alg[9]	1.001	7045.012	6258.158
a_alg[10]	1.000	6562.233	5746.464
a_alg[11]	1.001	6863.201	5678.795
a_alg[12]	1.000	7220.158	5791.735
a_alg[13]	1.001	7059.938	5687.501

Table 3. Diagnostics for the posterior ranks (text data)

Variable	Rhat	ess_bulk	ess_tail
a_alg[1]	1.001	9219.006	5696.938
a_alg[2]	1.000	8159.844	5501.918
a_alg[3]	1.001	8194.701	5731.082
a_alg[4]	1.001	9698.246	5792.446
a_alg[5]	1.000	9523.302	5751.300
a_alg[6]	1.000	9379.086	5835.681
a_alg[7]	1.000	8112.099	5414.614
a_alg[8]	1.001	10008.326	5965.950
a_alg[9]	1.000	8248.915	5321.357
a_alg[10]	1.000	7389.070	5735.235
a_alg[11]	1.001	8273.447	5734.861
a_alg[12]	1.000	8682.404	6117.667
a_alg[13]	1.001	7615.126	5528.505

TABLE 4. Diagnostics for the posterior ranks (numeric data)

Variable	Rhat	ess_bulk	ess_tail
a_alg[1]	1	10066.931	6341.406
a_alg[2]	1	9986.580	5733.858
a_alg[3]	1	9972.198	5585.470
a_alg[4]	1	9539.656	5750.775
a_alg[5]	1	9804.196	5886.244
a_alg[6]	1	9090.924	5964.757
a_alg[7]	1	9255.918	6154.528
a_alg[8]	1	9996.392	6090.676
a_alg[9]	1	9648.361	6529.156
a_alg[10]	1	8911.369	6347.509
a_alg[11]	1	9361.509	6190.749
a_alg[12]	1	9680.378	6071.385
a_alg[13]	1	9482.599	5616.573

TABLE 5. Diagnostics for the posterior effects (aggregated data)

variable	rhat	ess_bulk	ess_tail
a_alg[1]	1.000	3875.538	5106.532
a_alg[2]	1.000	2316.687	3800.909
a_alg[3]	1.000	2387.592	4104.560
a_alg[4]	1.000	3815.961	4953.275
a_alg[5]	1.001	3602.684	4895.245
a_alg[6]	1.001	3892.622	5240.429
a_alg[7]	1.001	3380.358	4652.745
a_alg[8]	1.000	3391.268	4957.767
a_alg[9]	1.000	3685.936	4975.255
a_alg[10]	1.000	3556.614	5042.657
a_alg[11]	1.001	3434.150	4687.095
a_alg[12]	1.000	3460.173	4839.767
a_alg[13]	1.001	3255.699	4956.926

Table 6. Diagnostics for the posterior effects (image data)

variable	rhat	ess_bulk	ess_tail
a_alg[1]	1.000	5726.116	4957.001
a_alg[2]	1.002	1508.296	1981.993
a_alg[3]	1.002	1456.598	1921.736
a_alg[4]	1.000	6636.772	5545.676
a_alg[5]	1.001	7268.332	5666.927
a_alg[6]	1.000	6687.499	5804.222
a_alg[7]	1.000	6876.669	5432.274
a_alg[8]	1.001	5526.927	5044.321
a_alg[9]	1.001	6553.413	5195.253
a_alg[10]	1.001	7505.705	4898.517
a_alg[11]	1.000	5621.637	5269.483
a_alg[12]	1.000	6701.225	5162.609
a_alg[13]	1.000	7089.734	6008.329

Table 7. Diagnostics for the posterior effects (text data)

variable	rhat	ess_bulk	ess_tail
a_alg[1]	1.001	6805.015	5088.715
a_alg[2]	1.001	10406.872	6048.554
a_alg[3]	1.000	7776.333	5465.982
a_alg[4]	1.000	7981.526	5584.848
a_alg[5]	1.002	7361.233	5793.209
a_alg[6]	1.000	10203.453	5950.313
a_alg[7]	1.001	7854.116	6110.782
a_alg[8]	1.001	8100.850	5459.704
a_alg[9]	1.000	5826.955	5488.954
a_alg[10]	1.000	7574.068	5277.421
a_alg[11]	1.000	7137.205	5607.689
a_alg[12]	1.000	5937.118	5193.251
a_alg[13]	1.001	8334.488	5386.165

Table 8. Diagnostics for the posterior effects (numeric data)

variable	rhat	ess_bulk	ess_tail
a_alg[1]	1.000	7050.378	6189.147
a_alg[2]	1.000	7428.251	6821.047
a_alg[3]	1.000	7241.839	6043.851
a_alg[4]	1.000	7679.707	6553.354
a_alg[5]	1.001	7987.069	6358.697
a_alg[6]	1.000	8513.448	6423.963
a_alg[7]	1.001	8644.972	6599.283
a_alg[8]	1.000	7961.322	6731.745
a_alg[9]	1.000	8236.145	6686.385
a_alg[10]	1.000	7693.744	6464.176
a_alg[11]	1.001	6196.071	6122.471
a_alg[12]	1.000	6893.882	5953.944
a_alg[13]	1.000	7000.893	6042.653

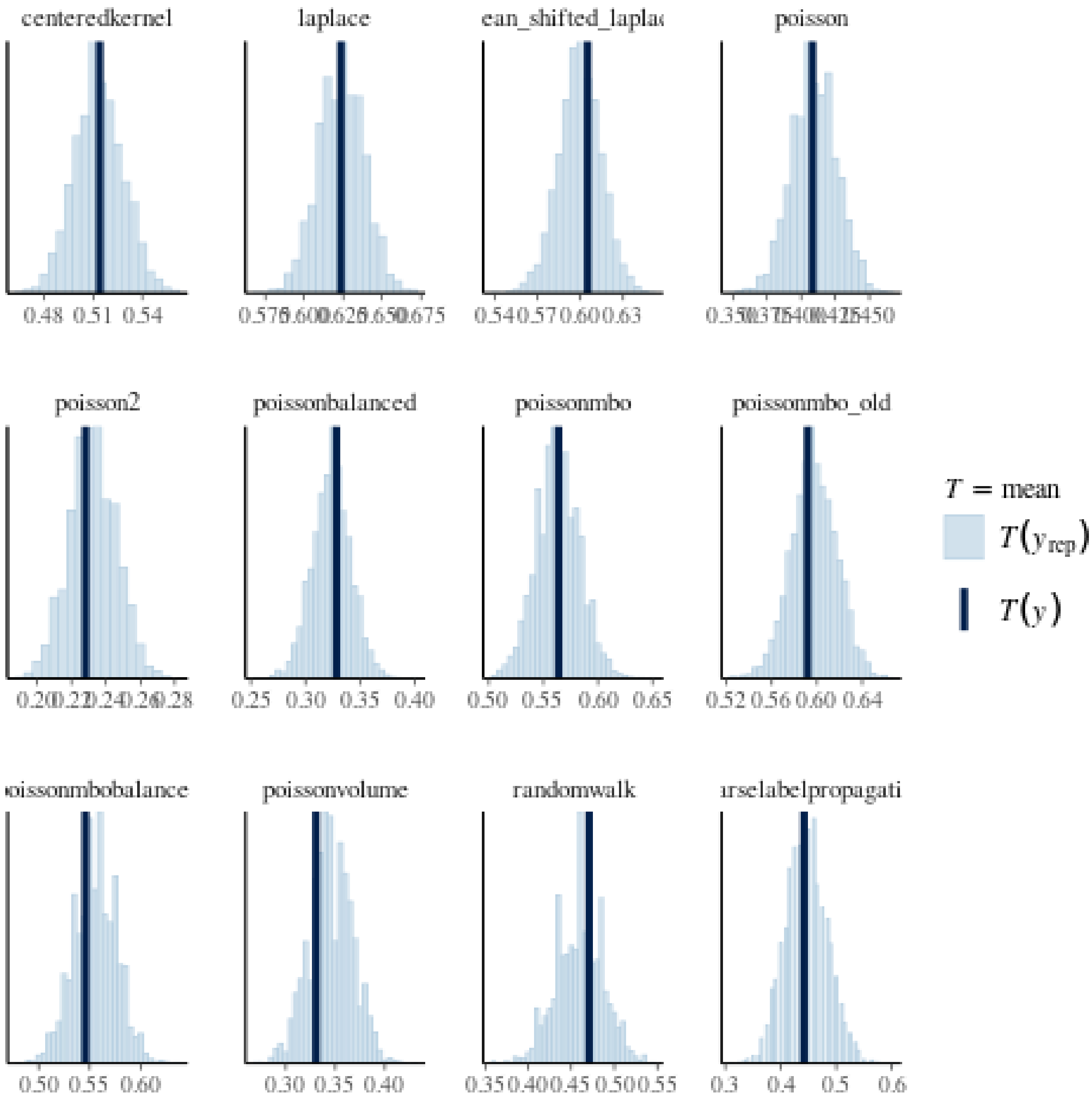

Figure 9. Posteririor predictice check of the ranks for each algorithm (aggregated)

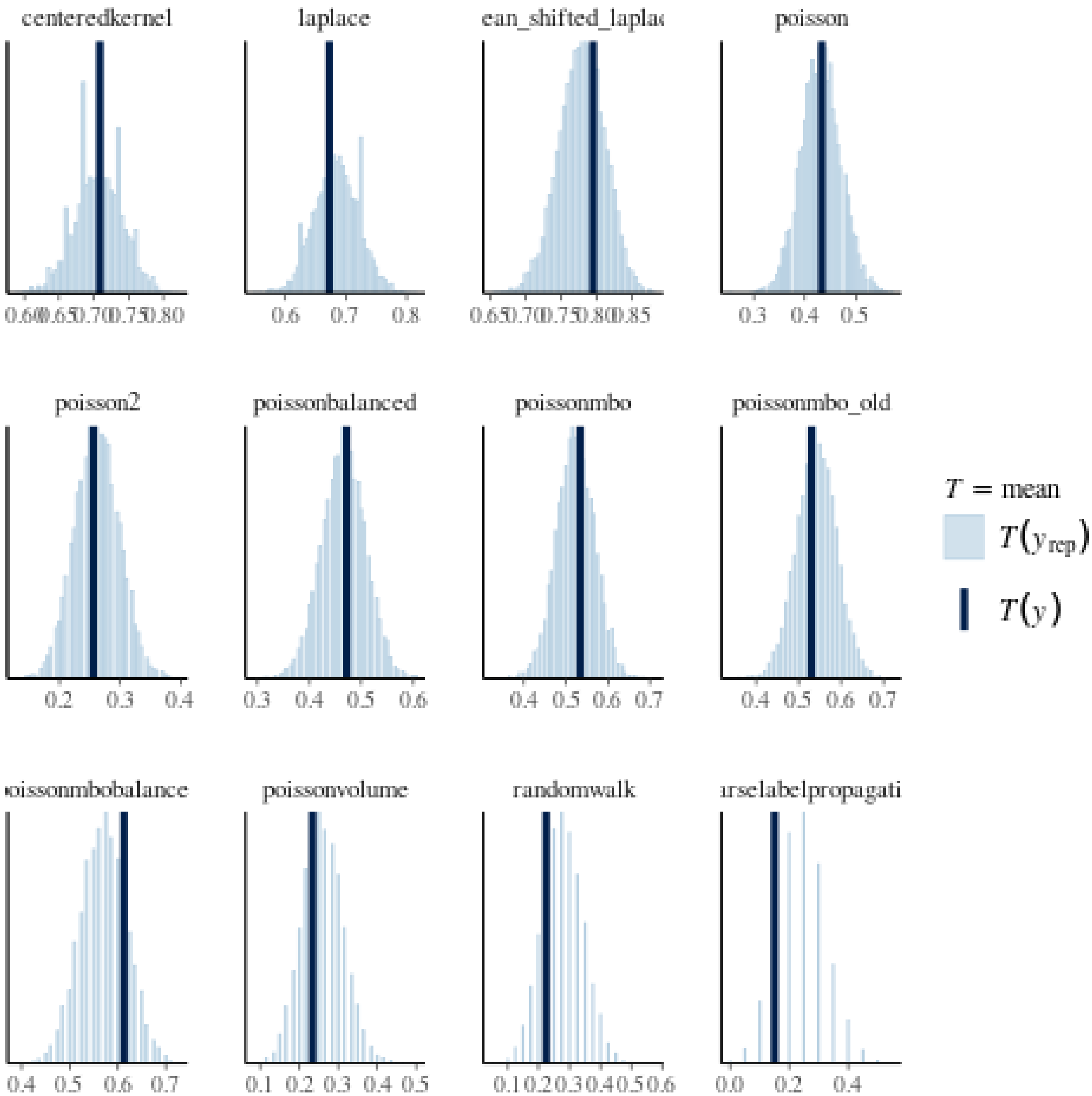

FIGURE 10. Posteririor predictice check of the ranks for each algorithm (image)

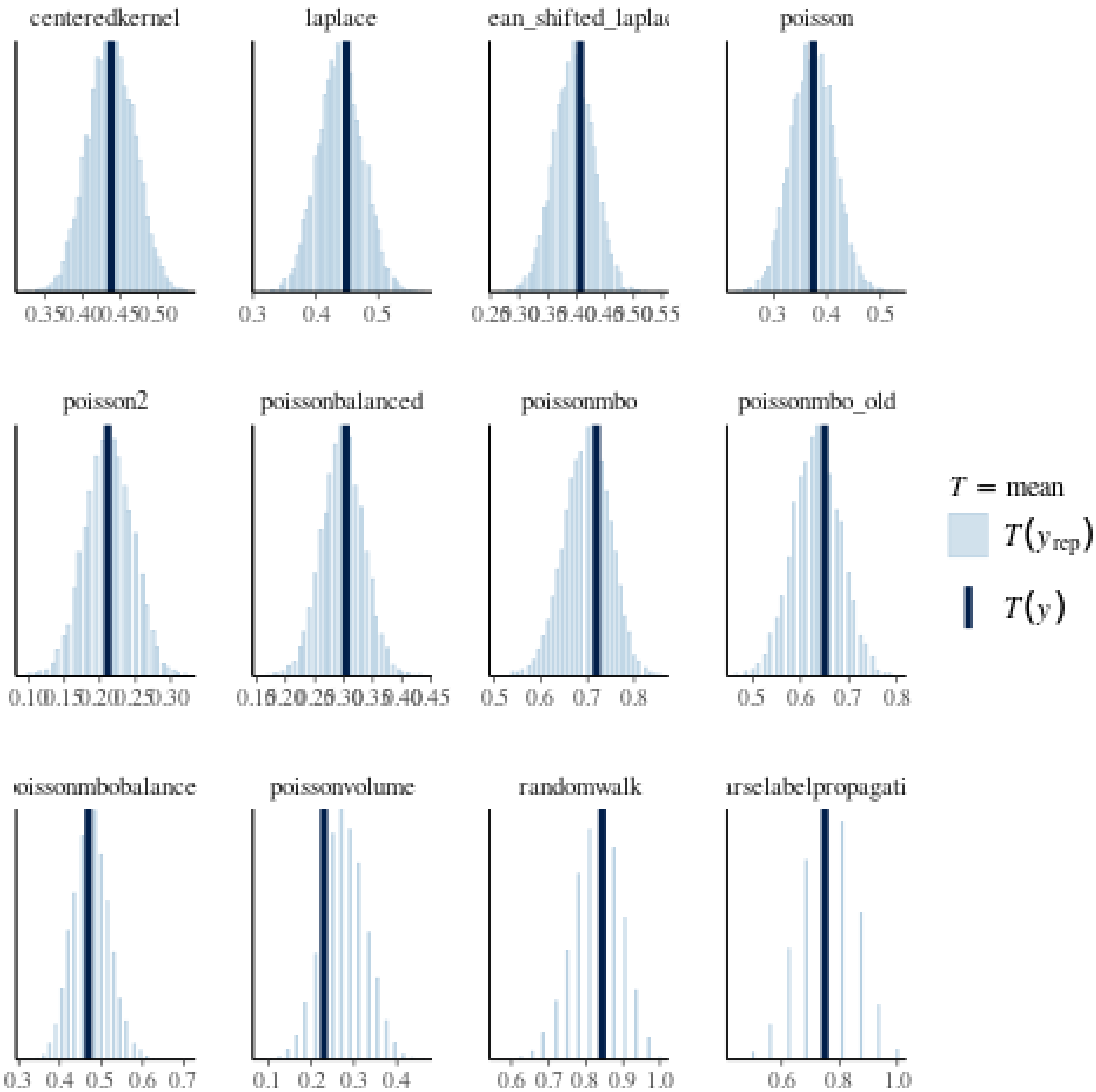

FIGURE 11. Posteririor predictice check of the ranks for each algorithm (text)

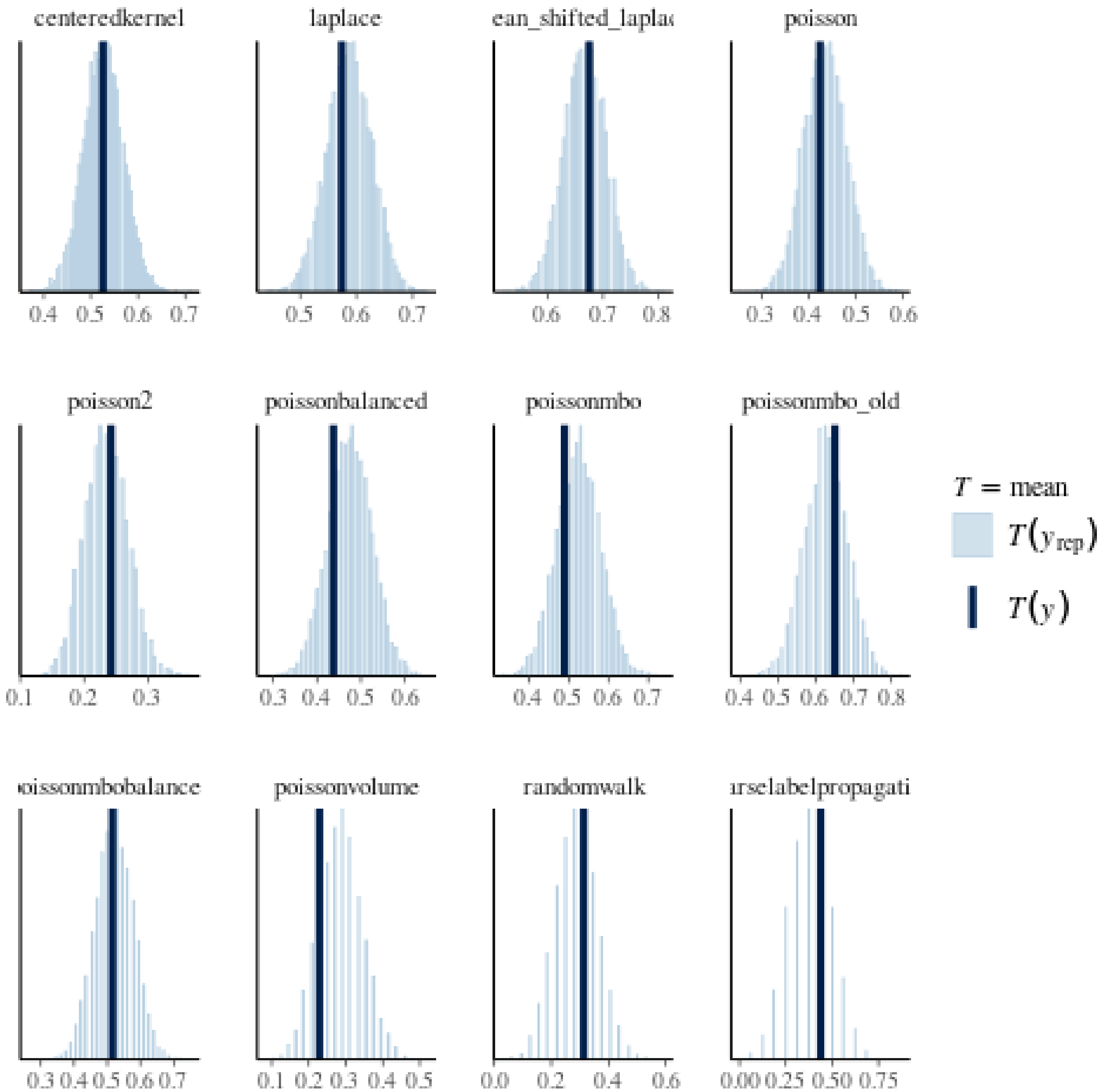

FIGURE 12. Posteririor predictice check of the ranks for each algorithm (numeric)

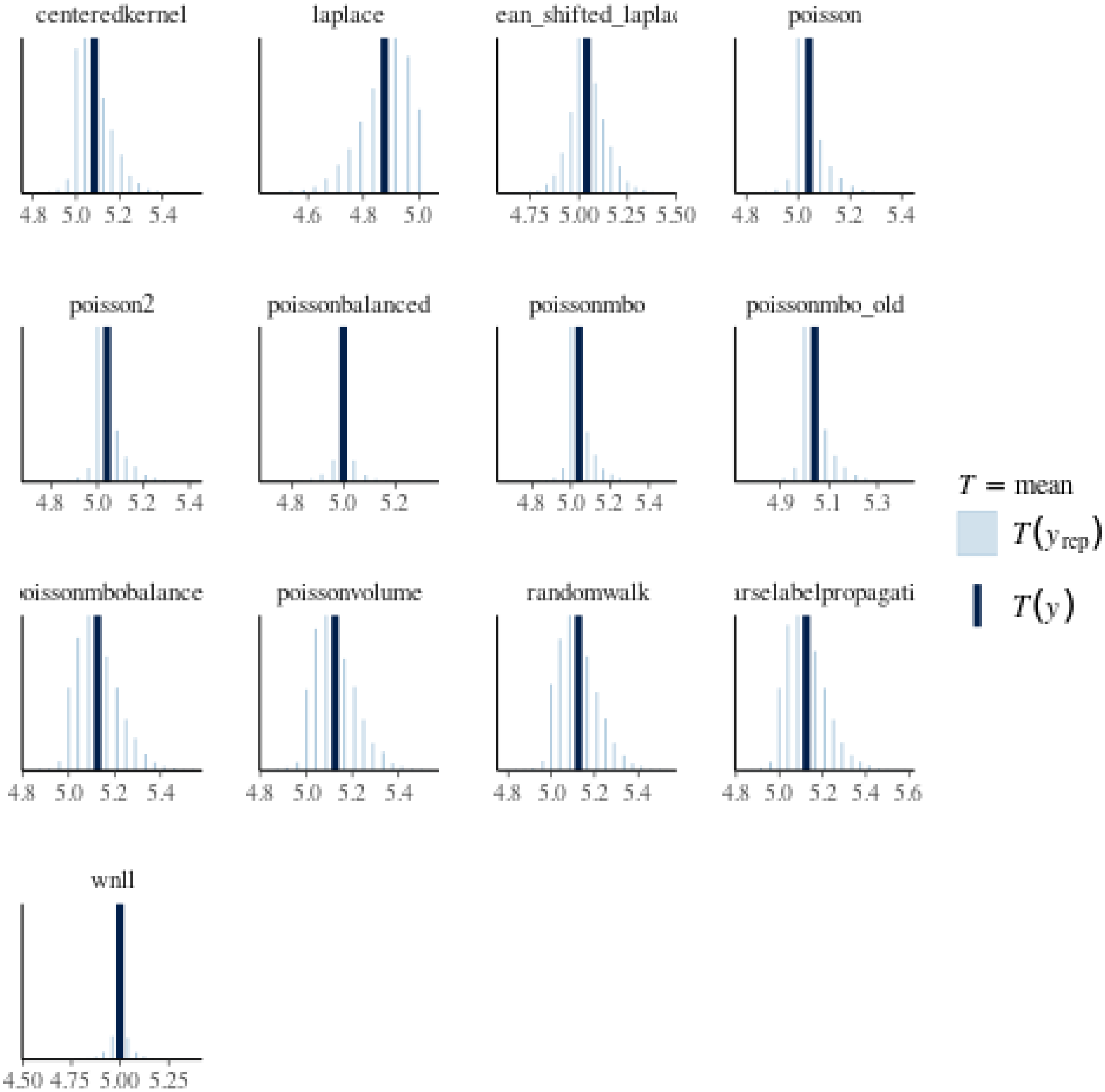

FIGURE 13. Posteririor predictice check of the probability of success for each algorithm (aggregated)

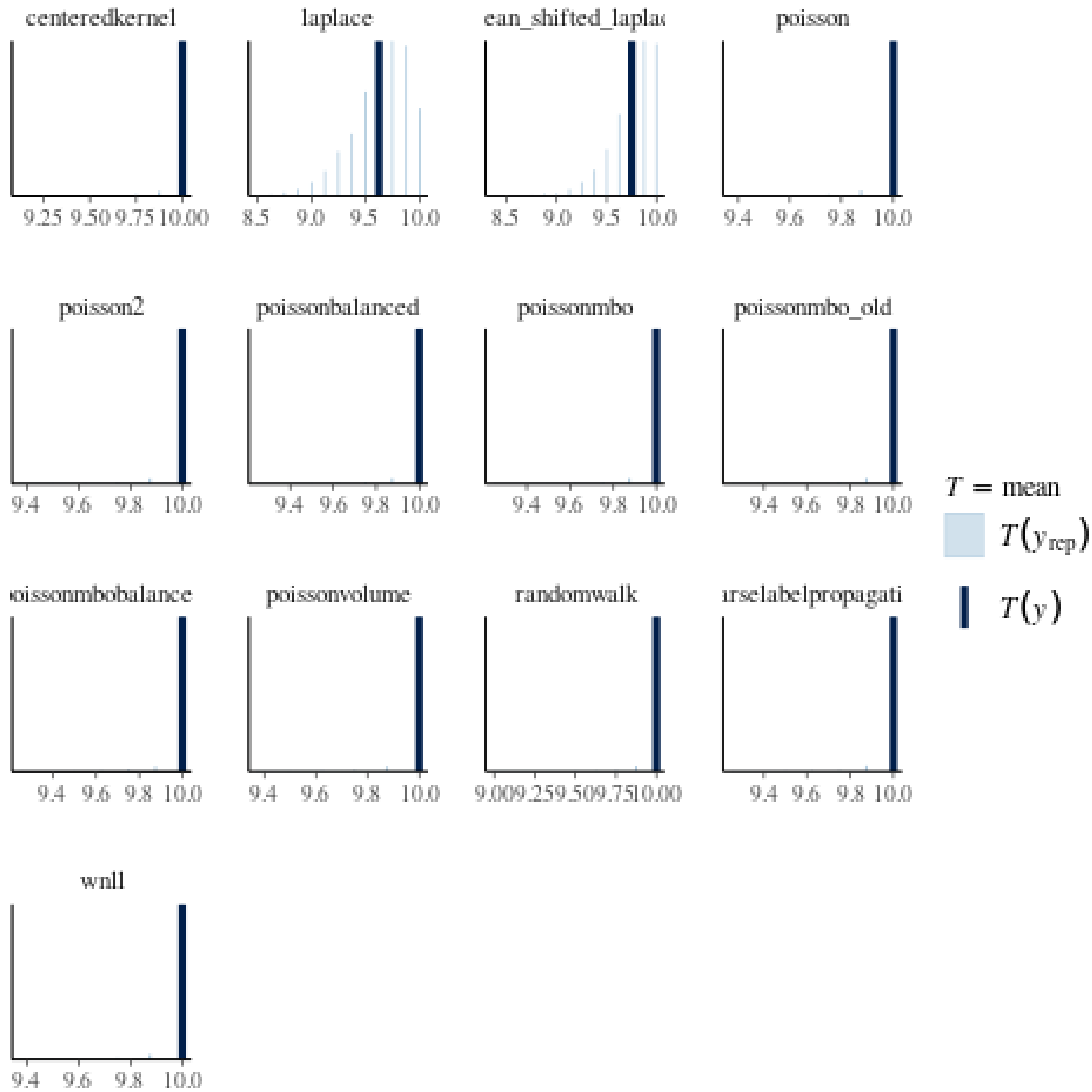

FIGURE 14. Posteririor predictice check of the probability of success for each algorithm (image)

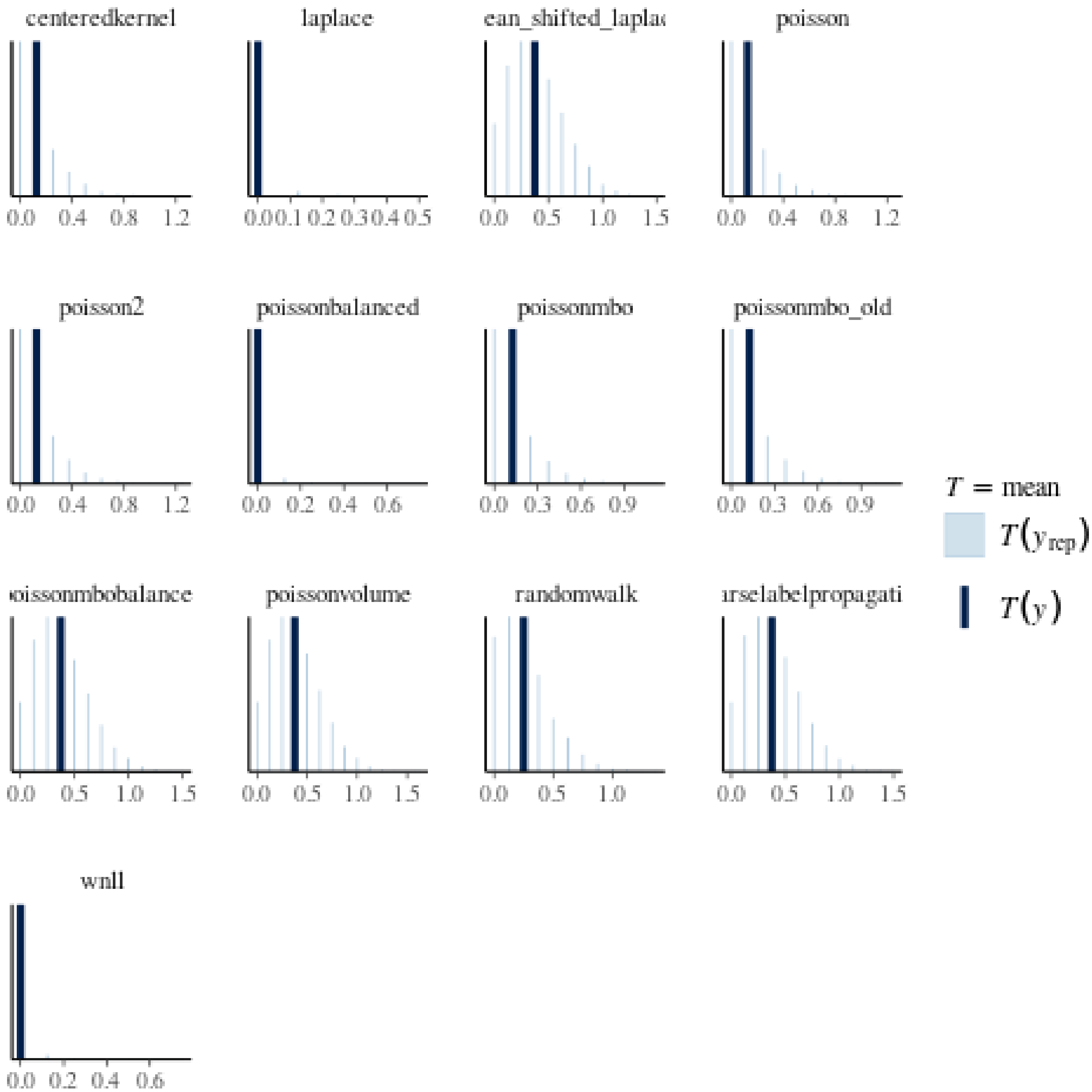

FIGURE 15. Posteririor predictice check of the probability of success for each algorithm (text)

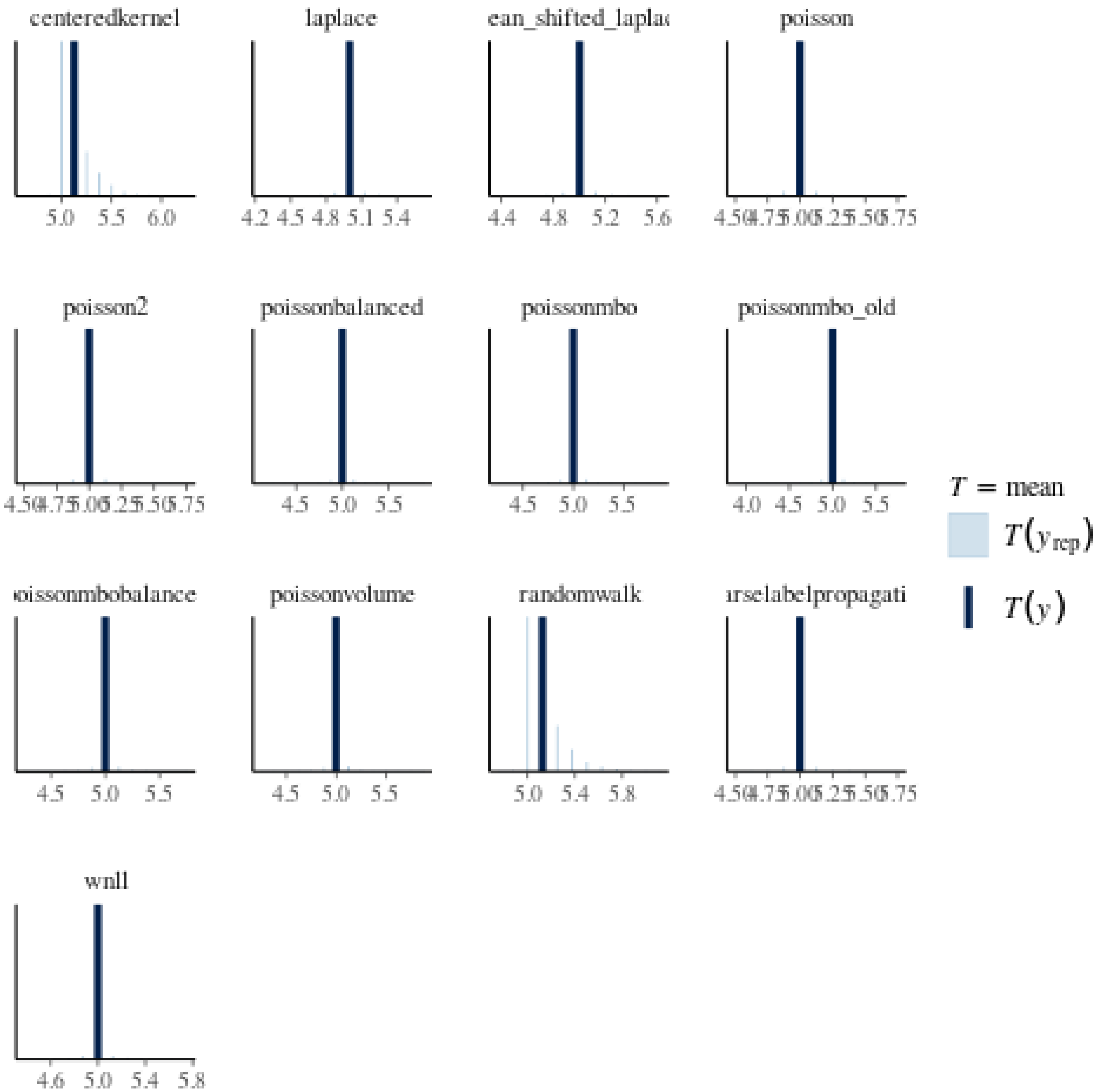

FIGURE 16. Posteririor predictice check of the probability of success for each algorithm (numerical)

References

[1] Carlo A Furia, Robert Feldt, and Richard Torkar. Bayesian data analysis in empirical software engineering research. *IEEE Transactions on Software Engineering*, 47(9):1786–1810, 2019.

[2] Jonah Gabry, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. Visualization in bayesian workflow. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 182(2):389–402, 2019.

[3] David Issa Mattos, Jan Bosch, and Helena Holmström Olsson. Statistical models for the analysis of optimization algorithms with benchmark functions. *IEEE Transactions on Evolutionary Computation*, 25(6):1163–1177, 2021.

[4] Richard McElreath. *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press, 2020.