Top 10 most cited MCMC papers, chronologically

By Hedibert F. Lopes (www.hedibert.org) – May 2026

Foundations (1953–1990)

1. Metropolis, Rosenbluth, Rosenbluth, Teller & Teller (1953)
Equation of state calculations by fast computing machines.
Journal of Chemical Physics, 21(6), 1087–1092. ≈ 50,000+ citations.
The original Metropolis algorithm — symmetric proposal, accept/reject step on configurations of interacting particles.

2. Hastings (1970)
Monte Carlo sampling methods using Markov chains and their applications.
Biometrika, 57(1), 97–109. ≈ 25,000+ citations.
Generalises the 1953 algorithm to non-symmetric proposals — the modern Metropolis-Hastings acceptance ratio.

3. Geman & Geman (1984)
Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6(6), 721–741. ≈ 30,000+ citations. Introduces the Gibbs sampler in the context of image processing; also gives the first formal proof of geometric convergence for finite state spaces.

4. Tanner & Wong (1987)
The calculation of posterior distributions by data augmentation.
Journal of the American Statistical Association, 82(398), 528–540.
≈ 6,000+ citations. The data-augmentation algorithm that opened the door to Bayesian computation for hierarchical and missing-data models.

5. Gelfand & Smith (1990)
Sampling-based approaches to calculating marginal densities.
JASA, 85(410), 398–409. ≈ 15,000+ citations.
The paper that popularised the Gibbs sampler in mainstream statistics.

Diagnostics and model-space MCMC (1992–1995)

6. Gelman & Rubin (1992)
Inference from iterative simulation using multiple sequences.
Statistical Science, 7(4), 457–472. ≈ 15,000+ citations.
The R-hat convergence diagnostic.

7. Chib & Greenberg (1995)
Understanding the Metropolis-Hastings algorithm.
The American Statistician, 49(4), 327–335. ≈ 5,000+ citations.
The canonical tutorial — the paper most people are sent to when they want to learn MH from scratch.

8. Green (1995)
Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4), 711–732. ≈ 10,000+ citations.
RJMCMC — sampling across models of different dimensions.

Modern era (2003–2014)

9. Andrieu, de Freitas, Doucet & Jordan (2003)
An introduction to MCMC for machine learning.
Machine Learning, 50, 5–43. ≈ 6,000+ citations.
The bridge paper that brought MCMC into the ML mainstream.

10. Hoffman & Gelman (2014)
The No-U-Turn Sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1), 1593–1623.
≈ 5,000+ citations. NUTS — the algorithm at the heart of Stan and PyMC.

Honourable mentions

- Neal (2011) ***MCMC using Hamiltonian dynamics*** (Handbook of MCMC)
- Haario, Saksman & Tamminen (2001) ***An adaptive Metropolis algorithm***;
- Roberts & Rosenthal (2001) ***Optimal scaling for various MH algorithms***;
- Chib (1995) ***Marginal likelihood from the Gibbs output***;
- Neal (2003) ***Slice sampling***.