The authors propose to apply a method called Sequential Tempered MCMC combined with a FSP-based multifidelity approach to do Bayesian inference for the Chemical Master Equation. Their method is tested on three examples, one based on experimental data, and yields speedups on the order of 2x to 5x compared to standard MCMC. The approach is of interest to the community, being easily amenable to further improvement and expansion where required. Overall I think the paper is promising, but there are some issues that I think should be addressed before publication.

Most importantly the authors do not provide code, announcing their intent to release implementation details elsewhere (p. 23, lines 3-4). In the interest of reproducibility I would hesitate to recommend publication before the code for the experiments is made available.

Excluding this issue please see the points below for further feedback.

**Major comments:**

- The first half of Section 4 would benefit from some elaboration. The authors point out that different parameter settings usually require different numbers of states to be considered in the FSP and that this cannot easily be determined a priori, but I do not understand how the proposed surrogate CME solves these problems, or even what the difference is to using the classical FSP.

- The paper requires some proofreading before publication as there are typos and dubious formulations spread throughout.

- The authors mention that "traditional computational methods for solving inference problems such as Markov Chain Monte Carlo methods based on classical Metropolis-Hastings algorithm involve numerous serial evaluations of the likelihood function, which in turn requires expensive forward solutions of the chemical master equation (CME)." While this is true for many methods, its not for all. for e.g. in the paper "Accuracy of parameter estimation for auto-regulatory transcriptional feedback loops from noisy data." Journal of The Royal Society Interface 16.153 (2019): 20180967. a Bayesian approach using a likelihood built from moments and informed by moment closure is shown to lead to a fast and accurate way to obtain the posterior of parameters provided the moment closure is carefully chosen. Th authors should expand their discussion to discuss such Bayesian approaches and also other recent likelihood free approaches for the CME such as the paper "Parameter estimation for biochemical reaction networks using Wasserstein distances." Journal of Physics A: Mathematical and Theoretical (2019). This is important to be able to frame the relevance of their method in the context of already published methods.

**Minor comments:**

- While the authors use a relatively new method (ST-MCMC) based on recent previous publications by the first author the proposed benefits, except speed, are not very clear. Tempering is commonly used in a MCMC context to better explore multimodal posteriors; do the authors believe that this, or any other aspect of the problem, might be relevant here? Otherwise, why should we include tempering at all?

- Section 4.2: Line 12 claims that the estimate of the log-likelihood converges from below, line 5 in the proposition claims that a subsequence converges from below and lines 15-17 show that the whole sequence converges and is asymptotically increasing. These non-equivalent statements are confusing and should be replaced by "asymptotic convergence from below" (for example).

- In Eq. 8 (p. 8) the expression for the likelihood does not include dependencies between observations of individual cells at different times as would be expected in a Markov model.

- Sec. 3.4, how reliable is the estimate of the model evidence using the empirically estimated factors c\_j?

- On p. 18, lines 16-17 calling certain parameter combinations 'nonphysical' is potentially misleading

- I initially understood the sentence as claiming that all reaction systems requiring a large state space in the FSP are nonphysical, which is sadly not the case.

- In Figure 2 (p. 26) the y-axis for the right subfigure only has one tick which makes it impossible to deduce the scale of the plot. The same holds for the x-axes in Figure 4.