molecular systems biology

Properties of cell death models calibrated and compared using Bayesian approaches

Hoda Eydgahi, William W. Chen, Jeremy L. Muhlich, Dennis Vitkup, John N. Tsitsiklis, Peter K. Sorger

Corresponding author: Peter K. Sorger, Harvard Medical School

Review timeline:

Submission date:	04 October 2011
Editorial Decision:	30 October 2011
Re- submission:	22 June 2012
Editorial Decision:	06 August 2012
Appeal:	20 August 2012
Editorial Decision:	14 September 2012
Revision received:	15 October 2012
Editorial Decision:	26 November 2012
Revision received:	14 December 2012
Accepted:	17 December 2012

Editor: Thomas Lemberger

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 30 October 2012

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from two of the three referees whom we asked to evaluate your manuscript. Given that they provide very similar recommendations, I prefer to make a decision now rather than delaying further the process. As you will see from the reports below, these referees raise substantial concerns on your work, which, I am afraid to say, preclude its publication in Molecular Systems Biology.

The reviewers acknowledge the extent of the parameter analysis performed in this study. However, they are not convinced that the conclusions reached by this analysis would represent a decisive conceptual advance over previous works. In addition, they feel that the study would be better suited for publication in a more specialized journal.

Under these circumstances, I am afraid the the level of enthusiasm expressed by the reviewers remains rather limited and I see no other choice than to return the manuscript with the message that we cannot offer to publish it.

In any case, thank you gain for the opportunity to examine your work. I hope that the points raised in the reports will prove useful to you and that you will not be discouraged from submitting future work to Molecular Systems Biology.

Reviewer #1 (Remarks to the Author):

The authors describe an approach to sample parameters of non-identifiable model using a combination of Bayesian inference and Monte-Carlo-Markov-chain. They apply it to a model of apoptosis. From their study, they derive three conclusions: First, distributions for different parameters exhibit large variability in their size. Second, using parameters from the multi-dimensional parameter distribution, predictions could be made. Third, there is strong non-linear co-variation between parameters.

While the authors claim that the last conclusion is surprising, this is an effect that has been studied intensively and en efficient algorithm has been developed to calculate these manifolds (e.g. Raue et al, Bioinformatics, 25(15), 1923-1929). The overall approach itself is not novel, but possibly it has not been applied in that specific combination to a model of this scale. While the results may be interesting for specialists in parameter estimation, they provide no insight into biology.

Smaller comments:
Typically, initial conditions would also need to be estimated, as measurements of these are noisy, thus they would go into the likelihood term.
The text is very verbose, maybe the text could benefit from making it more concise.

Reviewer #2 (Remarks to the Author):

The authors explore the 78-dimensional parameter space of a complex model for receptor-mediated apoptosis in single cells. They adopt a Bayesian approach, using Markov-Chain Monte Carlo to sample parameter vectors that are consistent with both the available data and biophysically-motivated priors. They find that efficient sampling requires a hessian-guided MCMC routine, with periodic updates of the hessian matrix (although this does not strictly satisfy detailed balance). Taking two-dimensional slices of the parameter landscape, they find ellipsoidal basins, canyons, and walls. They find that the parameter vectors they sample yield good agreement with the fit data and well-constrained predictions. By contrast, they find that independently sampling from the marginal distributions for each parameter yields very poor fits and wide predictions. Moreover, they find that sampling parameter vectors using the covariance matrix also yields poor fits and wide predictions, suggesting that nonlinearities in their parameter space are very important.

The message that modelers should carefully explore parameter space is an important one, and the field should certainly move that direction. Building upon work by Klinke, Sethna, Stumpf, and others, the authors have done a heroic job exploring the parameter space of a very complex model. Many of their observations about parameter spaces have been reported previously in similar investigations, but not on systems as large as that the authors consider and perhaps not thoroughly and coherently in a single publication.

Major points:

- Fig 2: Neither the legend nor the text mention the black "actual marginal" lines. Am I correct that this is the marginal distribution arising from sampling the full 78-dimensional parameter space, as opposed to the 2-dimensional slice of parameter space over which the red and green ensembles were generated?

- Page 14: The difference between the black and red/green distributions in Fig 2 emphasizes that 2-D slices of a high dimensional parameter space may be deceptive, in that the shape of the 2D basin may depend on the values of the other 76 parameters. This should be explicitly noted in the manuscript text, to avoid confusion on the readers' part.

- Page 15 and Fig 3A: The authors note that the 2D parameter space slice corresponding to k38 and k58 is ellipsoidal. They find this "intriguing", but it appears to be straightforward. The principle axes of the ellipsoid lie along the k38 and k58 axes, so the values of k38 and k58 that fit the data are uncorrelated with each other. This is *very* different from the Michaelis-Menten case they compare it to. In the Michaelis-Menten case, estimates of the forward and reverse binding constants are

highly correlated with each other, indicating that it is only the ratio that can be estimated. In the k38 and k58 case, the estimates are independent from each other, which is not surprising given that they are separated in the network. In short, the intriguing cases are when parameters are highly correlated, as in Michaelis-Menten, not when they are uncorrelated. (If we constructed two completely decoupled systems, pairs of parameters between systems would also be uncorrelated and yield ellipsoidal basins, but for a trivial reason.)
If the authors wish to find interesting relationships among parameter values, they should focus on pairs of parameters which show strong dependence between each other. A simple way to do this is to look at those parameter pairs with high covariance. The visualization methods from Erguler and Stumpf. Mol BioSystems 7:1593 (2011) may prove helpful: http://dx.doi.org/10.1039/c0mb00107d. A more sophisticated method that accounts for nonlinearity would be to examine mutation information between pairs of parameters.

- Page 16 and Fig 4AB: The distinction between the "best-fit" vectors and the manifold sampling is unclear to me. How were the "best-fit" vectors chosen? Are they simply those vectors from the manifold sampling that have highest posterior probability?

- Page 17: It is worth noting that author authors have also found that the parameter covariance matrix is not an ideal approximation to the posterior, particularly when making predictions. See, for example, Gutenkunst et al. Ann NY Acad Sci 1115:203 (2007): http://dx.doi.org/10.1196/annals.1407.003 , although the difference between covariance matrix and actual posterior there is not as dramatic as the present case.

- Page 26: The authors note that future modeling results should ideally include not just point estimates for parameters, or confidence intervals, or even covariance matrices, but rather samples from the posterior. They note, however, that a MCMC can generate an enormous number of samples from the posterior. (In their case 1.5e8/79 ~ 2e5.) However, one is most interested in *independent* parameter vectors from the posterior, and their 2e5 are almost certainly not independent from each other. Because the sample is generated by a random walk, and they've restricted step sizes in that walk, consecutive parameter vectors are highly correlated with each other. (In fact, if a move is not accepted, they are identical.)
To estimate the number of independent samples in their chains, a simple procedure is to calculate the autocorrelation time each of the parameters in their chains. The number of independent parameter vectors is then roughly the total length of the chain divided by that longest correlation time. I suspect the authors have only a few hundred truly independent parameter vectors, which is still daunting, but not unreasonable in computer-readable tables.

- Since the details of the MCMC are important to the paper, it would be helpful to report the acceptance ratio, i.e., the fraction of proposed steps that are accepted. This is an important quantity for tuning the MCMC. (The acceptance ratio is not simply related to the convergence time, but it can point computational speedups. For example, if the acceptance ratio is low, then on a parallel computer one can save time by trying many moves simultaneously.) The authors can calculate this from their saved Monte-Carlo chains by simply recording the fraction of times in which consecutive. parameter vectors differ.

Minor points:
- The abstract mentions "three key properties of this parameter distribution, one unexpected". It is not clear which result the authors found to be unexpected.
- The bibliography is messy, with numerous papers listed twice, based on small changes in the citation, such as journal abbreviation.
- The legends for Fig 5 and Fig 6 appear to have been swapped.
- Suppl Fig 4: The flowchart of the algorithm should include the hessian recalculation step.
- Suppl Fig 6: The large plots for k10 and k39 obscure several distributions. If the authors wish to highlight k10 and k39, they should do so without obscuring other data.

Re- submission 22 June 2012

Please find attached a manuscript entitled "Bayesian parameter estimation and model discrimination for complex biochemical networks" that represents a complete rework of our previous manuscript MSB-11-3289. At your suggestion, we are submitting this as a new paper but still including a detailed response to the review of the previous manuscript.

Our paper describes a Bayesian approach to parameter estimation and model discrimination that we test using a previously published ODE-based model of extrinsic apoptosis (about which we have recently published in MSB). The distinguishing feature of a Bayesian estimation scheme is that it returns distributions for parameters and this makes model-based predictions probabilistic.

The fundamental advance in this current paper relative to our earlier submission is that we now use Bayesian estimation to perform model discrimination. Discrimination involves computing the Bayes factor (the ratio of the evidence) for competing models in a way the accounts for experimental error, non-identifiability and different numbers of parameters. The latter is important because different representations of interactions among a fixed set of proteins invariably results in models with different numbers of free parameters. Computing the Bayes factor is fundamentally more rigorous way to score models than using maximum likelihood and the Bayesian Information Criterion (BIC), a simple metric in relatively wide-spread use. The Bayes factor reduces to the BIC under conditions of non-identifiability, which do not prevail in most realistic situations.
We apply our model discrimination approach to "direct" and "indirect" models that instantiate competing hypotheses about the regulation of mitochondrial outer membrane permeabilization. Both models have considerable support in the literature and are therefore of biological interest. The two models have equally good fits to experimental data but the evidence for the direct model is greater, reflecting its greater "robustness" to variation in otherwise unknown parameters. This conclusion is supported by recent data from Doug Green and colleagues.

With respect to the presentation of our results we have taken account of the previous critique that we were too wordy and have shortened things considerably by moving much of the established methodology into two text Boxes and other figures into the supplement.
The paper does not follow the style of an applied mathematics paper, however, in being equation rich and text poor - we simply do not believe this will serve an audience for which a more didactic approach is easier to follow.
It seems to us that reviewers 2 and 3 from the previous manuscript might usefully be contacted about this new paper. Reviewer 1 dismissed our paper without seeming to read it and claimed an equivalence with a previous paper that was simply wrong. Given our inclusion of thermodynamic integration as a means to evaluate Bayes factor, we also suggest Fabian Theis, Group leader and Professor for Mathematics in Systems Biology at the Helmoholtz Center in Munich as a third reviewer.
The software developed for this paper is an important part of the overall work and we are releasing it into the open source via Github. It was developed in MatLab but this is too slow for routine use (some calculations took us over a month on a cluster computer) and we are currently re-writing it in Python and C++. The estimation module is complete and the thermodynamic integration model nearly done. All will be in the open source and we are happy to follow any additional recommendations you have on data release. Models and all other artifacts of the work will be released as SBML and other prevailing standards.

Detailed Response to Review of MSB-11-3289:

Reviewer #1 Specific Comments:
- Third, there is strong non-linear co-variation between parameters. While the authors claim that the last conclusion is surprising, this is an effect that has been studied intensively and en efficient algorithm has been developed to calculate these manifolds (e.g. Raue et al, Bioinformatics, 25(15), 1923-1929). The overall approach itself is not novel, but possibly it has not been applied in that specific combination to a model of this scale.

RESPONSE: We no longer claim this is a surprising result and have deemphasized this section of the manuscript in favor of a focus on model discrimination. However, the profile likelihood method in Raue's study is very far removed from a thorough sampling of parameter space. It only samples along the deep valleys in the objective function's parameter space where parameter vectors are considered to be highly likely. It misses high likelihood points that aren't directly in the valley/profile. It is necessary to sample both regions of high and low likelihood in parameter space in order to draw statistically correct and biologically accurate conclusions.

- While the results may be interesting for specialists in parameter estimation, they provide no insight into biology.

RESPONSE: We have extended the Bayesian framework presented in the manuscript to a comparison of competing biological hypotheses. We believe that this is of substantially greater biological significance.

- Typically, initial conditions would also need to be estimated, as measurements of these are noisy, thus they would go into the likelihood term.

RESPONSE: This is true but many of our initial protein concentrations were measured and it seems reasonable to assume them. In addition, we are currently estimating 78 parameters for a highly unconstrained, large-scale model of apoptosis. As a proof of concept, we did not deem it necessary or practical to estimate all parameters for the models. Moreover, biologically significant results are not presented despite our only have calibrated the models for rate parameters.

- The text is very verbose, maybe the text could benefit from making it more concise.

RESPONSE: The text has been re-written; in addition, some of the background information can now been found in Boxes.

Reviewer #2 Specific Comments:
- Fig 2: Neither the legend nor the text mention the black "actual marginal" lines. Am I correct that this is the marginal distribution arising from sampling the full 78-dimensional parameter space, as opposed to the 2-dimensional slice of parameter space over which the red and green ensembles were generated?

RESPONSE: We have reworked this completely and black 'actual marginal' lines have been removed.

- Page 14: The difference between the black and red/green distributions in Fig 2 emphasizes that 2-D slices of a high dimensional parameter space may be deceptive, in that the shape of the 2D basin may depend on the values of the other 76 parameters. This should be explicitly noted in the manuscript text, to avoid confusion on the readers' part.

RESPONSE: We have completely reworked this text and the accompanying figures to clarify this point.

- Page 15 and Fig 3A: The authors note that the 2D parameter space slice corresponding to k38 and k58 is ellipsoidal. They find this "intriguing", but it appears to be straightforward. The principle axes of the ellipsoid lie along the k38 and k58 axes, so the values of k38 and k58 that fit the data are uncorrelated with each other. This is ****very**** *different from the Michaelis-Menten case they compare it to. In the Michaelis-Menten case, estimates of the forward and reverse binding constants are highly correlated with each other, indicating that it is only the ratio that can be estimated. In the k38 and k58 case, the estimates are independent from each other, which is not surprising given that they are separated in the network. In short, the intriguing cases are when parameters are highly correlated, as in Michaelis-Menten, not when they are uncorrelated. (If we constructed two completely decoupled systems, pairs of parameters between systems would also be uncorrelated and yield ellipsoidal basins, but for a trivial reason.) If the authors wish to find interesting relationships among parameter values, they should focus on pairs of parameters which show strong dependence between each other. A simple way to do this is to look at those parameter pairs with high covariance. The visualization methods from Erguler and Stumpf. Mol BioSystems 7:1593 (2011) may prove helpful: http://dx.doi.org/10.1039/c0mb00107d. A more sophisticated method that accounts for nonlinearity would be to examine mutation information between pairs of parameters.*

RESPONSE: The reviewer is correct here and we intended to make exactly the same point! This did not come across and we have now rewritten the results from scratch and included a clarifying example in Box 2.

- Page 16 and Fig 4AB: The distinction between the "best-fit" vectors and the manifold sampling is unclear to me. How were the "best-fit" vectors chosen? Are they simply those vectors from the manifold sampling that have highest posterior probability?

RESPONSE: Yes, the "best-fit" vectors are simply those from manifold sampling that have highest posterior probability. We have removed the "best-fit" figure from the manuscript to reduce redundancy and confusion. We have also tried to explain the sampling methods more thoroughly.

- Page 17: It is worth noting that authors have also found that the parameter covariance matrix is not an ideal approximation to the posterior, particularly when making predictions. See, for example, Gutenkunst et al. Ann NY Acad Sci 1115:203 (2007): http://dx.doi.org/10.1196/annals.1407.003 , although the difference between covariance matrix and actual posterior there is not as dramatic as the present case.

RESPONSE: We don't claim this observation is novel and in fact have also cited another of Gutenkunst's et. al's works that talks about this point (Gutenkunst et al. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS*

Comput Biol **3:** 1871-1878). Overall, the emphasis on this aspect of the work is greatly reduced given our new focus on model discrimination.

- Page 26: The authors note that future modeling results should ideally include not just point estimates for parameters, or confidence intervals, or even covariance matrices, but rather samples from the posterior. They note, however, that a MCMC can generate an enormous number of samples from the posterior. (In their case 1.5e8/79 ~ 2e5.) However, one is most interested in ****independent**** *parameter vectors from the posterior, and their 2e5 are almost certainly not independent from each other. Because the sample is generated by a random walk, and they've restricted step sizes in that walk, consecutive parameter vectors are highly correlated with each other. (In fact, if a move is not accepted, they are identical.) To estimate the number of independent samples in their chains, a simple procedure is to calculate the autocorrelation time each of the parameters in their chains.*

RESPONSE: This is an excellent point that we over-looked. The autocorrelation length for the 78 parameters ranged from 100 to 10,000; most of these were around 1000. Following the reviewers suggestion we thinned out the samples and rewrote the relevant portion of the text.

- Since the details of the MCMC are important to the paper, it would be helpful to report the acceptance ratio, i.e., the fraction of proposed steps that are accepted.

RESPONSE: The acceptance ratio for Hessian-guided MCMC walks were approximately 15-19%. We now report these rates and include a short discussion of optimal acceptance rates.

- The abstract mentions "three key properties of this parameter distribution, one unexpected". It is not clear which result the authors found to be unexpected.

RESPONSE: This text has been abandon and the entire premise of the paper reworked.

- The bibliography is messy, with numerous papers listed twice, based on small changes in the citation, such as journal abbreviation.

RESPONSE: The bibliography has been cleaned so that papers are only listed once. We appear to have had a problem with Endnote and apologize for the error.

- The legends for Fig 5 and Fig 6 appear to have been swapped.

RESPONSE: We thank the reviewer for noticing this inconsistency; the legends now correspond to the correct figures, although we have extensively reworked all of the figures

- Suppl Fig 4: The flowchart of the algorithm should include the hessian recalculation step.

RESPONSE: This is an excellent suggestion - the flowchart now includes the Hessian recalculation step.

- Suppl Fig 6: The large plots for k10 and k39 obscure several distributions. If the authors wish to highlight k10 and k39, they should do so without obscuring other data.

RESPONSE: This figure has been removed.

Reviewer #3 Specific Comments:
-In theory, constructing a MCMC algorithm is straightforward. In practice, one has to demonstrate that the algorithm is functioning correctly. One of the strategies used for posterior stimulation is to condition the proposal distribution such that the acceptance rates are near 20% (see Gelman Bayesian Data Analysis - Chapter 11). There is no discussion of this point. If the acceptance rates are too high, the MCMC chains reflect the prior - as is observed here in multiple cases.

RESPONSE: This is a good number that we shouldn't have omitted. We now report the acceptance rate in the Methods section. The acceptance rate over all chains was approximately 0.15-0.19. We added a discussion of optimality of the acceptance rate that refer to Gelman's work on optimality of jumping rules.

- Given that the available data informs the posterior distribution, it is unclear as to what data is actually being used to calculate the likelihood. Moreover, the authors state in the abstract that the parameters in the model are non-identifiable, it seems odd that posterior distributions in all of the parameters converged with largely symmetric distributions, as shown in Supplemental Figure 6. If the authors use less data and the same approach, do some of the parameters exhibit unbounded distributions?

RESPONSE: In our system, the prior distributions are fairly "narrow". Each parameter starts off with the assumption that they should span only 2 logs (base 10) on either side of the nominal. This is based on our reading of literature that reported rate constants. The result is that when the likelihood is uniform, the posteriors will look like the priors. This explains why many posteriors end up having very similar widths to the priors.

But the shape of the marginals only tell part of the story. We illustrate this in a new examination of top 10 parameters which show the least change from prior to marginal posterior. Sampling independently from only these 10 parameters shows a much worse fit, suggesting that even though they pass through the MCMC walk with the shape of the marginal posterior relatively unchanged, there are important correlations between these 10 parameters. We have added next text and a new Supplementary Figure 3 to illustrate this (compare to Figure 5).

- It is unclear here how the adaptive method proposed here has a theoretical basis for convergence. In addition, how exactly was the Hessian calculated? The authors claim (pg 23) that approximating the posterior landscape with a Hessian matrix is justified in the current case by the good match between paraboloids derived from the Hessian matrices and the actual local landscape. This seems like circular reasoning and doesn't demonstrate the independence of the posterior from the prior (or proposal). Previous concerns raise additional questions as to the dependence of the posterior on the prior.

RESPONSE: We clarify the stationarity and convergence properties of this type of adaptive walk in the text. We also cite references that use this type adaptation.

Use of the Hessian-guidance matrix doesn't guarantee correct sampling of posterior, but our empirical tests with and without the Hessian showed very little difference in the sampled posterior marginals. We have reworked the text extensively to make clear the tradeoffs involved in using adaptive MCMC methods.

- Establishing convergence should not be the end of the algorithm, as shown in Supplemental figure 4. Establishing convergence just means that the chains are marching around a similar likelihood space and no longer in the tails that reflect the overdisperse starting points. Once the chains are converged then subsequent samples can be used as representative samples derived from the posterior distribution. Are the histograms just based on the converged segments? Moreover, one does not need to report the whole chain but can thin the chain to report representative samples.

RESPONSE: We agree with the reviewer: establishing convergence should not, and is not, the end of the algorithm. We simply use it as a metric to 1) distinguish our methods from others that do not achieve convergence and 2) demonstrate the validity of our sampling algorithm. The histograms are based on the converged samples only. In addition, we agree that the entire chain does not need to be reported, only the independent samples. Independent samples were generated from the originals by selecting one out of every N samples, where N is the autocorrelation length of each parameter. This is now described in the text.

- There is some confusion in the paper as a prior distribution and a proposal distribution are used synonymously - they are not the same.

RESPONSE: We thank the reviewer for pointing this out. It was sloppy writing on our point that we have now corrected.

- p12 line 8-9 - was Hessian-guidance only applied to steps that are multiples of 25,000 and all other steps were unguided? This seems unclear.

RESPONSE: all the steps were Hessian-guided; the Hessian, however, was only calculated once every 25,000 steps. The text has been changed to clarify this point.

- The first order term in the Taylor expansion for a symmetric distribution evaluated at the mode is zero.

RESPONSE: If this is in reference to the page 29, our ln(post) isn't necessarily symmetric.

- In Simulated Annealing section first sentence - change high posterior values to high likelihood. As the samples are from unconverged chain segments, they are not samples from the posterior.

RESPONSE: Yes – this is true and to perform Simulated Annealing, we step through parameter space according to the posterior, which is the sum of the likelihood and the prior. The text has been changed to clarify this point.

- How exactly were the MCMC evaluations of 3-5 chains distributed across 70 cores?

RESPONSE: Each MCMC chain was run on a single node of the cluster; every 25,000 steps, the main node created 20 parallel jobs, each of which calculated a certain portion of the 78x78 Hessian matrix. We now describe this in the methods.

2nd Editorial Decision 06 August 2012

Thank you again for submitting your work to Molecular Systems Biology. We have now finally heard back from the three referees whom we asked to evaluate your manuscript. Given that this new submission was a strongly modified version of the previous manuscript, we sent it to one of the former reviewers (reviewer #2) and to two new referees. As you will see from the reports below, while referee #2 acknowledge that the manuscript has been significantly refocused and improved as compared to the previous submission, referee #3 is not convinced that the presented approach is conceptually novel and points to several prior related works. Referee #1 is overall positive but also raises the issue on whether "this work introduces any new methodological idea or concept".

We circulated the reports among the three referees to give them the opportunity to comment on each other's evaluation. In view of the comments provided by reviewer #3, Reviewer #2 noted "I am not as familiar with the applications of Bayes Factor analysis to systems biology as reviewer 3; certainly the analysis is very similar to that of Xu et al. and is not carried as far" and felt that the "paper is a nice verification and compendium of these techniques and general observations, and it does apply them to a larger system. However, I don't think there are any dramatic conceptual or methodological advances in the manuscript, and it probably doesn't have the novelty required for MSB." Reviewer #1 agreed that "the authors should better distinguish their work from other related work in the literature" but remained overall positive because "the authors have considered single-cell data rather than population-averaged data and the model is larger by ~2-fold, which is probably not insignificant given the computational expense of the techniques being demonstrated." Reviewer #3 remained very reserved and re-emphasized the overlap of the presented methodology with previous works, in particular Xu et al. Science Signaling 2010, and also mentioned additional related works (Calderhead & M. Girolami, 2011, Vyshemirsky & Girolami Bioinformatics 2008).

When considering the work by Xu et al and the other related papers, it appears indeed that Baysian parameter estimation, adaptive MCMC, Bayes factor model discrimination and thermodynamic integration were all applied previously in the context of the analysis of models of signaling pathways, with follow up experimental validation in the case of Xu et al. It seems thus that similar methods for the "Bayesian parameter estimation and model discrimination for complex biochemical networks" have been reported before. We do appreciate the elegance of the approach and the clear and very didactic presentation. We also recognize that the size of your model is larger and that you use single-cell data in the context of models of apoptosis. We are however not convinced that the combination of "rigorous Bayesian approaches to model calibration and discrimination" would represent major conceptual advance in view of the prior literature highlighted by the reviewers.

Under these circumstances, I see no other choice than to return the manuscript with the message that we cannot offer to publish it. I am very sorry not to be able to bring better news on this occasion. In any case, thank you again for the opportunity to examine this work and I hope that the points raised in the reports will prove useful to you.

Reviewer #1 (Remarks to the Author):

General Impression: The authors have done a commendable job of demonstrating a theoretically rigorous Bayesian workflow to estimate free parameters in ODE-based biochemical models. This approach enables discrimination between alternative models with different topologies and different numbers of parameters. This workflow has been demonstrated using a previously validated but non-identifiable model of receptor-mediated apoptosis in human cells as a starting point.

Major comments:

This manuscript presents a statistical framework for assessing complex biochemical network models. The framework provides a means to address parametric and topological uncertainties. The framework is based on a Bayesian approach and provides a way to address issues that are commonly ignored in modeling work. In their demonstration, the authors consider a published model of

TRAIL-mediated cell apoptosis that has a complex reaction network and 78 rate parameters. The authors assess the published model's predictive performance against a form of the model incorporating Bayesian estimates of the parameter values. Using the Bayes factor criterion, the authors further compare two variants of the model differing both in network topology and parameter number. Overall, the reported study leads to several interesting findings, which are enumerated below.

1) Bayesian inference of parameter values (posterior density estimation) can help rigorously deduce the confidence interval and prediction accuracy of complex biochemical network models despite the lack of knowledge of the parameters and network topology; 2) Bayesian parameter estimates can help achieve robustness in a model's predictive performance with respect to parameter variation; and 3) the Bayes factor can be efficiently used as a model selection criterion because it can compare candidate models of different structures and parametric dimensions. Furthermore, the Bayes factor can differentiate models based on robustness or parameter sensitivity, whereas maximum likelihood-based inferences may fail to make such distinctions.

The impact of the reported work is likely to be significant. The reported study can potentially serve as a useful guide for statistical validation of complex network models.

There are a few minor issues that the authors may wish to address. Apart from showing the merit of Bayesian approaches in modeling of biochemical networks, it is not clearly indicated whether this work introduces any new methodological idea or concept, or applied any existing theory in a unique manner. Bayesian concepts are widely used and their usefulness in model validation/comparison is known in other fields. The reported work certainly underscores the importance of these concepts, and provides a valuable domain-specific demonstration. It is not clearly stated whether the primary motivation for the work was to demonstrate the usefulness of Bayesian parameter estimates in the context of a large multivariate biochemical network model. Is the computational approach implemented to improve/accelerate the Markov Chain Monte Carlo (MCMC) walk very unique? It would be helpful to better highlight the distinctive features of this work.

The authors provide details about the computational approach used to accelerate the MCMC walk to convergence in the 78-dimensional parameter space of the model. Multiple MCMC chains were initiated at random points in the parameter space to ensure global convergence. Combined simulated annealing (at initial stage) and semi-adaptive Hessian-guided moves in the MCMC walk accelerated chain propagation to convergence. However, it has not been clearly stated to what extent the new computational approach contributes to efficiency. It would probably be useful to provide a comparison of CPU time (or number of MCMC steps) for the following three cases: 1) basic MCMC, 2) MCMC with simulated annealing, without Hessian-guidance, and 3) MCMC with both simulated annealing and Hessian-guidance (the approach recommended by the authors). Convergence rates in the three cases could also be compared, with the results given in supplemental material.

Are there more/less sophisticated approaches that have been invented to improve the MCMC search? It seems unlikely that no attempt has been made in the context considered by the authors. However, if alternative approaches are available, it would be useful to show how the new computational approach introduced here performs against those existing approaches in terms of convergence and computational efficiency.

The authors may wish to briefly discuss bootstrapping (a simple procedure described, for example, in Numerical Recipes by Press et al.), which is another way to obtain confidence limits on parameter estimates.

The authors may wish to discuss physical constraints on parameter values and how these can be handled within their recommended Bayesian approach. I am thinking of upper limits set by diffusion and constraints of detailed balance, which make it impossible to specify all rate constants independently. The authors cite Schoeberl et al. (2002) on p.3 and the model reported in this paper is one that violates constraints of detailed balance; this paper has been cited in some of the work about detailed balance constraints. The size of parameter space is reduced by these types of constraints, so it would seem like something worthy of a comment, and moreover, a model parameterization that violates the diffusion limit or a constraint of detailed balance, is then incorrect in a bad way. (It's OK

for models to be incorrect in good ways but not bad ways.)

The authors may wish to discuss philosophically whether or when AIC, BIC or Bayes factor are appropriate criteria for model selection, as these criteria basically evaluate the goodness of a model as a fitting function. If I have two data points, all of these criteria will scream straight line. However, if I understand something about the underlying physics, I might prefer a parabola. In the case of the authors' own work, they have developed models based on mechanistic considerations that would be considered by any model selection criterion to be poor relative to a set of polynomial fitting functions, I am guessing. It seems that an important point not forcefully made is that alternative models are being considered in the authors' work and the different models are all deemed mechanistic, or plausibly mechanistic. The authors, I do not believe, are advocating that we always favor models that are optimal only w.r.t. a model selection criterion, but some researchers may misunderstand and become over exuberant and ignore mechanistic considerations? I am actually not sure how statisticians evaluate models that have too many parameters for the data on hand but that have features that are aligned with mechanistic understanding vs. models with few parameters and non-mechanistic. I for one would appreciate some discussion of the issue. Perhaps others too.

Minor comments:

The citation of Danos et al. (2007) on p. 3 is an odd choice (vs. the other models cited), given that the model reported in this paper has all parameter values set to 1.

The terms "in vitro" and "in vivo" are not used in italics consistently throughout the manuscript.

Page 4: I thought a genetic algorithm is a type of evolutionary computing, but the authors make a distinction. ??

Page 9: Replace "can by recovered" with "can be recovered"

Page 9: It was mentioned that "the two parameters balance each other out in a subtle way." Please explain in more detail.

Page 11: Please provide a reference for the Komlogorov-Smirnov test

Page 20: "...with improvement in speed..." Speed of what?

Page 22: "appropriate as a means" or "appropriate as means"??

Page 22: In the sentence starting with "Under the assumption", instead of "then we are forced" consider using "we are forced."

Page 26: Instead of "but this a weak criterion" it should be "but this is a weak criterion."

Page 27: Instead of "but in some case it makes more sense" it should be "but in some cases it makes more sense"

Page 27: Replace "speed computation" with "speed of computation"

Reviewer #2 (Remarks to the Author):

This manuscript is heavily revised from a previous submission, downplaying the previous focus on parameter space exploration and adding an application of Bayesian model choice.

The authors first introduce ODE modeling and the difficulties of large parameter spaces, before focusing on their model, EARM1.3, and potential alternative mechanisms for MOMP control. They then explore parameter space via MCMC walkm pointing out in both their model and in the classic Robertson system (in Box 2) that combinations of parameters may be constrained rather than individual parameters. They note that an adaptive method with Hessian guidance was useful to ensure convergent sampling and does not seem to strongly bias the resulting parameter distributions,

even though it violates detailed balance. They also note that non-uniform priors are necessary for convergence and help avoid difficult-to-integrate regions of parameter space. Using their parameter distributions, they then demonstrate the ability to make prediction that agree with data. They then use a Bayes factor (calculated using thermodynamic integration) to discriminate between direct and indirect models for MOMP (both of which fit the available data well). Finally, they consider the properties of the posterior landscape, showing that marginal distributions and the covariance matrix are poor representations of the nonlinear landscape. In the discussion, the authors consider how modeling papers should report parameter values, contrast Bayes factors with AIC and BIC, and discuss computational costs of the approach.

The updated manuscript is much stronger than the previous version. The additional model discrimination analysis dramatically increases the novelty of the work, and will, I think, make the manuscript much more interesting to non-modelers. In addition, the errors and ambiguities from the prior manuscript have been corrected, and the present work appears to be all-around solid. I have only minor suggestions for improvement.

Minor points:
1) Page 1: "rules-based modeling" in the keywords seems inappropriate. The model presented is an ODE model, and whether or not it was generated from a set of rules in the Hlavacek/Fontana sense is immaterial to the analysis presented.

2) Page 9: The text claims that Figure 2 shows marginal posterior distributions, but figure does not show them.

3) Page 15: A confidence interval was estimated for the Bayes factor by "propagating estimates of the error on the curves". It is not clear to me what this entails, and it would be helpful to detail it in the Methods.

4) Page 17: It is not clear to me what parameters sampled "from the peaks of independent marginal distributions" means. Are these simply sampled from the marginals, in which case I find the "from the peaks" language confusing and suggest striking it.

5) Page 14-19: Currently the paper switches from properties of a single model, to comparing two models (Model discrimination... section), back to properties of a single model (Properties of the posterior... section). The discussion might be more clear if the order of those last two sections were switched.

6) Page 36 (Box 1): It is perhaps confusing that the update rule for \Theta_J -> \Theta_{test} shown does not use the Hessian guidance that is used in the actual application. It would be helpful to note than many proposal distributions are possible, and the one used in the application is more complex than the symmetric one shown.

Typos:
1) Page 4: "experimental *** using genetic algorithms"

2) Page 21: "applicable to all cellular *** rather than priors"

3) Reference duplicates: Gutenkunst (2009a, 2009b); Klinke (2009a, 2009b); Spencer (2009a, 2009b)

Reviewer #3 (Remarks to the Author):

Probabilistic analysis of mass-action models is nothing new in both the frequentist and Bayesian traditions. The work of Jens Timmer and his group has considered practical means of addressing non-identifiability in biochemical models of Jak-Stat signalling based on likelihood-based inference. The work of Girolami (as referenced by the authors) has employed Bayesian inference using MCMC on large scale biochemical models of ErK signalling.

There is nothing that is really novel in this manuscript that has not been published in the literature

already. The manuscript adds nothing to what has already in the following publications - and these are just a small sample of a large literature.

Melke, P. et al A rate equation approach to elucidate the kinetics and robustness of the TGF-beta pathway. Biophysics Journal, 91, 4368-4380, 2006

Xu, T.R, Vyshemirsky, V., Gormand, A., Girolami, M., Baillie, G.S., Ketley, D., Milligan, G., Dunlop, A.J., Houslay, M.D., and Kolch. W., Inferring Signalling Pathway Topologies from Single Species Multiple Perturbation Measurements, Science Signaling, Vol.3, Issue 113, p. ra20.

J. Bachmann, A. Raue, M. Schilling, V. Becker, J. Timmer, U. Klingm√1/4ller.
Predictive mathematical models of cancer signaling pathways. J. Internal Medicine 271, 2012, 155-165

A. Raue, C. Kreutz, T. Maiwald, U. Klingm√1/4ller, J. Timmer.
Addressing Parameter identifiability by model-based experimentation. IET Systems

Girolami, M., Calderhead, B., Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods (with discussion), Journal of the Royal Statistical Society - Series B, 73(2), 123 - 214. Biology 5, 2011, 120-130

Appeal 20 August 2012

Harvard Medical School
DEPARTMENT OF SYSTEMS BIOLOGY

200 Longwood Avenue
Warren Alpert 438
Boston, Massachusetts 02115-5730

Peter K. Sorger, Ph.D. Tel: 617.432.6901 peter_sorger@hms.harvard.edu
Professor of Systems Biology FAX: 617.432.6990 Christopher_Bird@hms.harvard.edu
Professor of Biological Engineering (Massachusetts Institute of Technology

Thomas Lemberger, PhD
Chief Editor
Molecular Systems Biology
re: Manuscript MSB-12-3864

August 15, 2012

Dear Thomas,

I am writing in connection with your recent decision to reject our paper MSB-12-3864 (*Bayesian parameter estimation and model discrimination for complex biochemical networks*). We agree that we could have worded the paper better to emphasize novel elements and that we should have included references suggested by reviewer 3 (in addition to other papers by the same authors). Some of the revisions in a previous round of review led us to overemphasize methodology as opposed to results as the key feature of the manuscript. However, we believe that reviewer 3 is incorrect in concluding that our paper repeats previous work. We ask that you therefore reconsider a revised paper in which we address the remaining technical issues from reviewers 1 and 2 and make textual changes that address the very brief comments from reviewer 3.

In the comments that follow, we have focused in particular on the issues summarized in your cover letter regarding the comparison with Xu et al. (2010) and the relationship between our methods and BioBayes and GNU MCSim. Reviewer 3 mentions the work of Jens Timmer as being overlapping, but having worked closely with Jens for several years and I believe that his approach is not Bayesian in the sense we use it here. He has made use of complexity metrics such as AIC (Akaike Information Criterion) whose applicability we question. We think that our work is complementary not duplicative of research from the Timmer lab.

We are also unable to confirm the claim that most of the publications listed by reviewer 3 as overlapping are "just a small sample of a large literature" except insofar as MCMC sampling or Bayesian methods are generally mentioned (a point addressed below). We would also like to point out that there is little discussion about the biological novelty of our work. Discriminating between direct and indirect models of extrinsic apoptosis is an important topic in cell death research and has not previously been addressed using a formal model-based approach. Reviewers 1 and 2 are positive about this but the highly critical reviewer 3 does not mention it.

We were clearly at fault for not citing Xu et al. (2010) in our manuscript and we are of course happy to correct this. We also admit that in several key places (the end of the

abstract, end of the introduction etc.) we did a poor job of explaining why our paper is new and interesting. We never claimed to have been the first to apply MCMC methods, Bayes factors or Bayesian inference to ODE models and we recognize that they have been in routine use in applied math for many decades. The key question is not the methods themselves (used in hundreds of papers in physical sciences and engineering) but whether their application is significant and new. We are responsible for this misreading of our work by Reviewer 3 and will make appropriate corrections.

The novelty of our paper arises from our explicit analysis of what it means to estimate parameters for a biochemical (kinetic reaction) model, the implications, strengths, and limitations of Bayesian inference and how this all relates to conventional understanding of mass action biochemistry (specific points are listed below). Reviewer 3 and many mathematicians of our acquaintance have a stylistic objection to this treatment of the topic. They favor a compact, formula-rich style in which algorithmic novelty is the key and the actual results are left for the reader to work out. We cannot comment on what is appropriate for applied mathematics but we find even the most relevant applied mathematics requires substantial translation before it can be effectively applied in a biological setting. I hope that you will appreciate that it is extraordinarily difficult to get the "tone" of interdisciplinary papers such as ours correct, and we certainly did not intend to cause offense or preach to the converted. The reviewer's negative comment that our paper is "didactic" stems from this difference in style. Our goal is research novelty not pedagogy but we think it is incorrect to assume clear and thorough explanations cannot go hand-in-hand with new science (MSB "boxes" help is this regard). Reviewers 1 and 2 would appear to agree with this position, and have pushed us towards clearer and more complete explanations.

By way of example, consider how parameters in mass-action models relate to each other. It is well established that convergent MCMC sampling of parameters returns their joint probabilities. It is also clear from previous literature that rate parameters can be highly correlated: ratios such as K_M are generally more identifiable than the underlying forward and reverse rate constants. However, it has not previously been shown (to our knowledge) that a very large fraction of the information captured by model calibration lies in parameter co-variation and that this co-variation is sufficiently non-linear to be missed by correlation matrixes. The general habit of describing parameters in terms of individual (marginal) parameter distributions explicitly ignores this. However, we show that ignoring co-variation (treating parameters as independent) largely destroys model predictability whereas using the co-variant joint distribution returns predictions whose uncertainties match the actual uncertainties obtained by single-cell measurement of dying cells. Thus, the near-universal practice of reporting parameters as tables of values (or their SBML equivalents) is invalid. Instead, we proposed that the correct description of *n* parameters requires an *n+1* dimensional matrix of values obtained by thinning MCMC walks.

The existence of parameter co-variation also means that biochemical models have "too many" parameters. A large applied math literature describes how methods akin to PCA

can be used to reduce parameter number through clustering and orthogonalization. However, this yields parameters with no relationship to mass action rate constants and we are among the first to point out that having to manage underdetermined co-variant parameters is the inevitable cost of working with physical systems based on reversible mass action or stochastic reaction kinetics. Again, the underlying math is well known, at least by experts, but the implications for a specific physical system (a biochemical network) have not previously been elucidated.

More generally, a critical point missed by reviewer 3 is that attempts to relate the results of Bayesian estimation to our common understanding of biochemistry is essential for ensuring that the approach is understood and adopted by others. Reasoning about models based on single "best" fits remains ubiquitous (even in the pages of MSB) but I think all three reviewers would concur that it is deeply flawed. In our opinion, this unsatisfactory situation exists precisely because the absence to date of any discussion of the *results* rather than the *methods* of parameter estimation. A countervailing example is the work from Sethna and colleagues who coined the term parameter sloppiness and analyzed models from this perspective. As you know, the "sloppy" model paper of Gutenkunst et al. (PLoS Comp Biol 2007 Vol 3:10, 1871-1878, Times Cited: 176) is one of the most highly cited papers in PLOS Computational Biology. By explaining our findings in the Sethna framework we aim to emulate this success, relate our findings to his and also cover related work by Arkin, Klinke and others who have worked on non-identifiablility. Xu et al. (2010) does not explicitly address these issues or non-identifiablility and sloppiness.

A secondary problem we address is the absence of general-purpose open-source code for Bayesian analysis of biochemical models. The topic is currently the domain of specialists whose methods and code not readily used on biochemical models; in some cases the source code not available to the general community. Although Xu et al 2010 includes several distinguished applied mathematicians, there is no analysis of the assumptions made in analyzing mass action models with Bayesian methods. Overall, the methodological treatment is brief with no more than a few sentences in the main body of the paper and some mathematical definitions in the supplementary methods. This brevity masks many conceptual and technical issues associated with Bayesian model discrimination and the methods are embedded in closed-source code.

In sum, the novel discoveries and observations in our paper include the following:

- We show that Bayesian estimation allows high-confidence predictions to be made from sloppy models with high structural and parametric nonidentifiability while accounting for the uncertainty of parameter values. Sloppiness in biochemical models is a property of a fully sampled landscape and the concept is therefore relevant beyond the specific local problems examined by Sethna et al. This rebuts claims from Tidor and others that clever design can overcome the sloppiness problem (which we suggest to be intrinsic to the mass-action approximation).

- We demonstrate the value of probabilistic prediction using single-cell data on extrinsic cell death space and show its power to make predictions about the onset and duration of apoptosis that match single-cell data collected from HeLa cells. Probabilistic prediction is not part of Xu et al but reviewers 1 and 2 recognize this as a valuable aspect of our work.

- Calibration of complex models generates sets of parameters with high, non-linear co-variation. It is essential for this information to be used for accurate model-based prediction (see above). To our knowledge, this feature of biochemical parameters has never previously been analyzed although, in retrospect, it can be understood as an extension of concepts formalized by Michaelis and Menten in which K_M is recognized to be the identifiable parameter in simple catalytic reactions.

- Because of the high information content in parameter co-variation, it is incorrect to assess the degree of identifiability of a model simply by looking at individual parameter distributions (a point implicitly captured by the Fischer information matrix but often overlooked; see above).

- The Bayes factor includes a natural "Occam's razor" that, while not immediately intuitive, is an essential aspect of its ability to compare models with different numbers of parameters. In the so-called "asymptotic" limit, the Bayes factor reduces to the AIC and BIC, widely used metrics for scoring model complexity, but the limit conditions are unlikely to be satisfied with most biochemical models and use of the AIC and BIC is therefore highly suspect. This is an important point because it is not generally appreciated that testing alternative mechanistic models involving the same sets of proteins usually involves a change in the number of parameters. Correctly accounting for this difference is essential for model discrimination. We attempts to give this feature of Bayes a physical explanation.

- There is an equivalence between robustness and nonidentifiability, and Bayesian methods rigorously connect the two. Robustness, nonidentifiability and Bayesian estimation are three popular topics well represented in literature that not refer to each other. Our major finding is that "robustness" as it is usually defined is indistinguishable from nonidentifiability, and that robust/nonidentifiable models are favored unless they provide no explanatory value (careful investigators like Arthur Lander and Thomas Hoffer use a more interesting definition of robustness). This balance can be quantified by the Bayes factor. The focus of previous papers by Girolami et al and, Xu et al is methodological and the discussion provides little insight into these topics.

- A more technical point is our finding that efficient movement of MCMC walks in the landscape of non-identifiable models requires adaptive approaches but that such approaches have a tradeoff in terms of assumptions about stationarity; we examine this with a careful study of convergence, correlation, and walk lengths for adaptive and non-adaptive approaches. This is an area that clearly needs for further development.

- A practical value is that we provide a roadmap for tacking uncertainty in parameter values and model topology within a cohesive, general-purpose set of tools and software. The software is open-source and extensible and is being integrated into our PySB suite of tools. All of this can already be downloaded from the open source repository Github.

It is not the purpose of this letter to re-review, Xu et al., but it is necessary to point out that it is impossible to determine precisely what was done in that paper with respect to model discrimination. The point of that paper is very different: it describes a very interesting piece of biology and only briefly covers methods and approaches. Thus, it does not supersede ours with respect to methods on the simple grounds of reproducibility: there is no easy way to apply the reasoning or methods in Xu et al to other models. Xu et al also fails address technical issues analyzed in our manuscript (several pointed out by the reviewers) including convergence, proposal distribution, stationarity etc. Finally, Xu et al does not actually examine the resulting parameter distributions in any detail.

In conclusion then, a fair number of papers describe MCMC methods for parameter estimation and a somewhat smaller set use Bayes factors. Our paper is new in describing how these tools can be understood in the context of mass action biological networks and how they reveal previously underappreciated features of parameter space. Given the widespread interest in Sethna's work and the urgency of approaching cellular biochemistry from a more rigorous perspective we believe that it is worth covering these topics in some detail. It is also obvious that previously developed software in this area is not being widely used. Whether our tools will change this remains to be seen, but it will certainly not be for lack of our making them freely available.

At your request, we have also examined potential antecedents to the methodological aspects of our paper. Vyshemirsky and Girolami (*BioBayes: A software package for Bayesian inference in systems biology, 2008*) is the technical foundation of Xu et al and describes the software package BioBayes, a GUI-based, SBML importer and MCMC sampler for ODE models. The paper confines itself to the technical features of the sampler, such as the choice of a proposal distribution, the use of multiple chains in tempered annealing (parallelizing an MCMC walk), and the statistic used to assess convergence. Aside from its SBML importer, BioBayes could very well be a general purpose package for Bayesian estimation in any field of natural science. There is no discussion of why specific values for adjustable parameters were chosen or whether information of actual or models motivated specific choices. In contrast, our paper explains why the adaptive proposal distribution is essential and the landscape is globally sloppy.

A clear limitation of BioBayes is that the source code is not explicitly open (available by request) and written in Java, making it harder to understand, debug and modify than MatLab. The fact that users interact with BioBayes via a GUI means that functionality is restricted to what the programmers intended. You need to ask the developers to make

changes for you. Even were the code made available, most systems biologist are proficient in the use of high level mathematical packages like MatLab and Mathematica or languages such as Python but not Java and C.

GNU MCSim is a text-based model building tool and MCMC sampler in which models are written in ANSI-C and then compiled to binary by the user. The choice of C is motivated by speed (historically, the fastest numerical packages have been developed in Fortran and C) but the cost of this approach is developmental complexity: coding in C has a much higher learning curve than coding in MATLAB or Python. Indeed, GNU MCSim's own documentation admits that "*Other programs have been created to the same end, the Matlab family of graphical interactive programs being some of the more general and easy to use.*" In contrast, the software in our paper is available as source code via a public version control repository in both MATLAB and Python versions. We chose these languages because of their complementary histories: MATLAB is closed source but by far the most popular environment in systems biology; Python is open-sourced and its scientific libraries Numpy/Scipy bear a strong resemblance to MATLAB. While MATLAB and Python are traditionally regarded as less numerically efficient, they are correspondingly easier to use and program and modern computers mitigate the inefficiency.

From a practical perspective we believe that increased understanding of Bayesian estimation coupled with open-source software in familiar languages will enable the spread of the method beyond its current use by a restricted set of applied mathematicians to a general community of systems biologists. Individuals involved in this field have deep expertise (greater than ours) but they have thus far failed to appreciate or examine the biological systems they are working on.

Based on these arguments we ask that you re-consider a final revision of our paper in which we address all remaining issues raised by reviewers one and two, cite the appropriate papers from reviewer 3 and make limited modifications to the text to emphasize the novelty of our work. We also agree to provide the source code and user manual for our estimation scheme (indeed, some of it is already available on Github). We believe this will prove to be the right decision some years down the road when rigorous Bayesian parameter estimation and model discrimination are routine features of kinetic modeling. Thank you for your consideration,

Yours sincerely,

Peter Sorger

3rd Editorial Decision 14 September 2012

Thank you again for your detailed reply to our previous decision with regard to your manuscript. We have now had the time to examine your arguments in detail and to ask our Advisory Editorial Board for advice. I am pleased to inform you that the outcome of the consultation was positive and that we can consider a revision of the manuscript.

While several of the methods used in this study have applied before in the context of biological systems (Xu et al 2010, in particular), including Bayesian parameter estimation, MCMC sampling, thermodynamical integration and Bayes factor model discrimination, we accept the argument that the focus and main novelty of your study is to demonstrate the implications of joint parameter distribution analysis rather than the development of new methods. We can also agree that providing open source software may contribute to a wider adoption of these approaches and concepts.

We would thus kindly invite you to revise the manuscript according to the reviewers comments and by taking into account the following points:

- the methods used in your study should be better placed in the context of previous applications to biological systems and the relevant literature appropriately cited and discussed.

- the text and abstract, and possibly the title, should be carefully revised to place a clearer emphasis on the main aspect of novelty of your work.

- as suggested in your reply letter, links to the Python source code would be particularly appreciated.

- the analysis of MCMC walks (page 10-11) and the discussion of the Bayes factor on page 23 is fairly technical for a broad audience and may contribute to dilute somewhat the main and more novel message of the paper.

Thank you for submitting this paper to Molecular Systems Biology.

1st Revision - authors' response 15 October 2012

Harvard Medical School
DEPARTMENT OF SYSTEMS BIOLOGY

200 Longwood Avenue
Warren Alpert 438
Boston, Massachusetts 02115-5730

Peter K. Sorger, Ph.D. Tel: 617.432.6901 peter_sorger@hms.harvard.edu
Professor of Systems Biology FAX: 617.432.6990 Christopher_Bird@hms.harvard.edu
Professor of Biological Engineering (Massachusetts Institute of Technology

Thomas Lemberger, PhD
Chief Editor
Molecular Systems Biology
re: Manuscript MSB-12-3864R-Q

October 14, 2012

We would like to thank you again for the care with which you reconsidered our paper and for the excellent suggestions for a final round of revision.

We attached a modified draft and a slightly revised version of the rebuttal letter, which addresses most of the issues raised in the last round of review of our paper, particularly those raised by reviewer 3. The current letter does not rehash this, instead addressing specific concerns raised by reviewers 1-2 and describing the changes we have made to the text.

Most important, we now correctly reference Xu et al (2010) and remove any inappropriate claims of novelty on our part. Methods are placed in the context of previous applications to biological systems and the relevant literature is cited insofar as we can find it. References to Bayesian estimation for biochemical models which were already included in our original draft include:

- Battogtokh D, Asch D, Case M, Arnold J, Schuttler HB (2002) An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of Neurospora crassa. *Proceedings of the National Academy of Sciences* **99:** 16904.

- Flaherty P, Radhakrishnan ML, Dinh T, Rebres RA, Roach TI, Jordan MI, Arkin AP (2008) A dual receptor crosstalk model of G-protein-coupled signal transduction. *PLoS Comput Biol* **4:** e1000185.

- Klinke DJ (2009a) An empirical Bayesian approach for model-based inference of cellular signaling networks. *BMC bioinformatics* **10:** 371.

Additional references suggested by reviewers as being relevant to Bayesian estimation of biochemical models have been included:

- Bois FY (2009) GNU MCSim: Bayesian statistical inference for SBML-coded systems biology models. *Bioinformatics* **25:** 1453-1454.

- Vyshemirsky V, Girolami M (2008) BioBayes: a software package for Bayesian inference in systems biology. *Bioinformatics* **24:** 1933-1934.

- Xu TR, Vyshemirsky V, Gormand A, von Kriegsheim A, Girolami M, Baillie GS, Ketley D, Dunlop AJ, Milligan G, Houslay MD, Kolch W (2010) Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species. *Sci Signal* **3:** ra20.

Second, the title, text and abstract of the paper have been revised to more correctly reflect the novelty of our work; in part this involves bringing back points made in an earlier draft.

Third, we have substantially advanced the Python source code and made it available for download at the open-source repository Github under the name of BayesSB (https://sorgerlab.github.com/bayessb). We continue to develop and extend the software and to integrate with our PySB package, and we will make this development path transparent to potential users.

Fourth, the discussion of MCMC walks on page 10-11 has been moved into a supplementary note to make the paper more accessible. We think the discussion of the Bayes factor on page 23 is important however, and we have tried again to write it in an accessible way. If you think we have nonetheless failed, we are happy to move this into the supplement as well.

Thank you again for your support,

Peter

Revisions made in response to review

Reviewer 1:

Apart from showing the merit of Bayesian approaches in modeling of biochemical networks, it is not clearly indicated whether this work introduces any new methodological idea or concept, or applied any existing theory in a unique manner.

R1.1 We have rewritten the abstract and text (as mentioned above) to make clear that our manuscript does not describe a new method but rather the surprising features of multivariate parameter landscapes.

It is not clearly stated whether the primary motivation for the work was to demonstrate the usefulness of Bayesian parameter estimates in the context of a large multivariate biochemical network model.

R1.2.We have addressed this point with extensive revision of the introduction and abstract.

Is the computational approach implemented to improve/accelerate the Markov Chain Monte Carlo (MCMC) walk very unique? It would be helpful to better highlight the distinctive features of this work.

R1.3. In response to the editorial cover letter, we have now moved the discussion of MCMC walks into the supplementary materials "Properties of MCMC walks." Our treatment is not unique but many implementation details such as convergence and detailed balance have been considered in detail. As a supplementary note we believe the discussion remains valuable for those seeking to use our methods. In addition, we now cite a fine summary of adaptive methods in the reference Gilks et al, 1996 (see p 10). This and related references describe some adaptive methods that violate stationarity and others that do not but are inefficient. As mentioned in the manuscript, we sought a middle ground between efficiency and stationarity.

It has not been clearly stated to what extent the new computational approach contributes to efficiency. It would probably be useful to provide a comparison of CPU time (or number of MCMC steps) for the following three cases: 1) basic MCMC, 2) MCMC with simulated annealing, without Hessian-guidance, and 3) MCMC with both simulated annealing and Hessian-guidance (the approach recommended by the authors). Convergence rates in the three cases could also be compared, with the results given in supplemental material. Are there more/less sophisticated approaches that have been invented to improve the MCMC search? It seems unlikely that no attempt has been made in the context considered by the authors. However, if alternative approaches are available, it would be useful to show how the new computational approach introduced here performs against those existing approaches in terms of convergence and computational efficiency.

R1.4 To address this concern, we have provided a comparison of the number of MCMC steps and the number of parameters that have converged under two scenarios, 1) MCMC with simulated annealing and without Hessian-guidance and 2) MCMC with both simulated annealing and Hessian-guidance. These results are summarized in Table 1. Implementing another adaptive approach could be interesting but does not appear to be congruent with editorial guidance to reduce discussion of this point. Thus, we do not

believe that it does not fit within the scope of this paper, and we take pains to point out this paper work is not a study in methods but rather a study of implications.

The authors may wish to briefly discuss bootstrapping (a simple procedure described, for example, in Numerical Recipes by Press et al.), which is another way to obtain confidence limits on parameter estimates.

R1. 5 A short discussion on bootstrapping has been added. We argue that similarity is minimal given the statistical basis of our method. Boot strapping generates a new distribution of data points but all of these would lie within some equivalent chi-squared bands, and the Bayes method already accounts for the possibility that the measured data is merely a sampling of the "true" data.

The authors may wish to discuss physical constraints on parameter values and how these can be handled within their recommended Bayesian approach. I am thinking of upper limits set by diffusion and constraints of detailed balance, which make it impossible to specify all rate constants independently. The authors cite Schoeberl et al. (2002) on p.3 and the model reported in this paper is one that violates constraints of detailed balance; this paper has been cited in some of the work about detailed balance constraints. The size of parameter space is reduced by these types of constraints, so it would seem like something worthy of a comment, and moreover, a model parameterization that violates the diffusion limit or a constraint of detailed balance, is then incorrect in a bad way. (It's OK for models to be incorrect in good ways but not bad ways.)

R.1 6 This is an important point and we have clarified the relationship between hard physical constraints and soft priors in the manuscript. As the reviewer will appreciate, a prior in the Bayesian framework serves to set physical constraints on parameter values. Please see the discussion in the section "Choosing Priors" on page 12.

The authors may wish to discuss philosophically whether or when AIC, BIC or Bayes factor are appropriate criteria for model selection, as these criteria basically evaluate the goodness of a model as a fitting function. If I have two data points, all of these criteria will scream straight line. However, if I understand something about the underlying physics, I might prefer a parabola. In the case of the authors' own work, they have developed models based on mechanistic considerations that would be considered by any model selection criterion to be poor relative to a set of polynomial fitting functions, I am guessing. It seems that an important point not forcefully made is that alternative models are being considered in the authors' work and the different models are all deemed mechanistic, or plausibly mechanistic. The authors, I do not believe, are advocating that we always favor models that are optimal only w.r.t. a model selection criterion, but some researchers may misunderstand and become over exuberant and ignore mechanistic considerations?

I am actually not sure how statisticians evaluate models that have too many parameters for the data on hand but that have features that are aligned with mechanistic understanding vs. models with few parameters and non-mechanistic. I for one would appreciate some discussion of the issue. Perhaps others too.

R1.7 The reviewer is correct in that if we chose to compare a model based on mechanistic considerations with model that consisted of only a polynomial function, AIC, BIC, and

Bayes Factor will all favor the more simple polynomial function. The purpose of our work here is not to simply to fit a function, but to be able to learn something about the biochemistry underlying the network as a whole. The parameter values that produce good fits to a polynomial function will not necessarily produce kinetics that would be in line with the remainder of the biological network. Thus, we think our approach is congruent with the viewpoint of the reviewer and we now discuss this in the text.

Minor comments:
The citation of Danos et al. (2007) on p. 3 is an odd choice (vs. the other models cited), given that the model reported in this paper has all parameter values set to 1.

R1.8 The citation has been removed.

The terms "in vitro" and "in vivo" are not used in italics consistently throughout the manuscript.

R1.9 The inconsistent italicization has been fixed.

Page 4: I thought a genetic algorithm is a type of evolutionary computing, but the authors make a distinction?

R1.10 The reviewer is correct – this was sloppy writing on our part. Genetic algorithms are a type of evolutionary computing, as is evolutionary strategies. The distinction has now been made References to genetic algorithms and evolutionary strategies have been grouped under "genetic algorithms".

Page 9: Replace "can by recovered" with "can be recovered"

R1.11 Done

Page 9: It was mentioned that "the two parameters balance each other out in a subtle way." Please explain in more detail.

R1.12 As it is now mentioned in the text, the two parameters balance each other out in a subtle way insofar as k_1 is the forward rate of the ligand-binding reaction (which promotes cell death) and k_2 is the forward rate of the FLIP-binding reaction (which inhibits cell death).

Page 11: Please provide a reference for the Komlogorov-Smirnov test

R1.13 Done

Page 20: "...with improvement in speed..." Speed of what?

R1. 14 We have removed the ambiguous phrase "improvement in speed" and replaced it with "faster convergence".

Page 22: "appropriate as a means" or "appropriate as means"??

R1. 15 "Means" as method or path or way forward. It is correct as written.

Page 22: In the sentence starting with "Under the assumption", instead of "then we are forced" consider using "we are forced."

R1.16 Fixed.

Page 26: Instead of "but this a weak criterion" it should be "but this is a weak criterion."

R1.17 Fixed.

Page 27: Instead of "but in some case it makes more sense" it should be "but in some cases it makes more sense"

R1. 18 Fixed

Page 27: Replace "speed computation" with "speed of computation"

R1.19 Fixed.

Reviewer #2:

Page 1: "rules-based modeling" in the keywords seems inappropriate. The model presented is an ODE model, and whether or not it was generated from a set of rules in the Hlavacek/Fontana sense is immaterial to the analysis presented.

R2.1 We removed the keyword which does indeed seem to have entered in error.

Page 9: The text claims that Figure 2 shows marginal posterior distributions, but figure does not show them.

R.2.2 The marginal posterior distributions were present in a previous draft but were removed. However, the textual reference was erroneously left in. We have fixed this.

Page 15: A confidence interval was estimated for the Bayes factor by "propagating estimates of the error on the curves". It is not clear to me what this entails, and it would be helpful to detail it in the Methods.

R2.3 We created a Gaussian distribution at each temperature and the mean of the distribution was set to the mean of the ln(likelihood) values we obtained from the MCMC chains. Variance was set to set to the variance on the means obtained. Using the Gaussian distributions we were able to generate many curves describing the behavior of ln(likelihood) as temperature was varied, which, in turn, led to confidence intervals on the Bayes Factor values calculated. A sentence was added in the text on page 15 to explain this.

Page 17: It is not clear to me what parameters sampled "from the peaks of independent marginal distributions" means. Are these simply sampled from the marginals, in which case I find the "from the peaks" language confusing and suggest striking it.

R2.4 Yes, the reviewer's assumption is correct and wording inadvertently confusing. The text has been modified to make the point more clear.

Page 14-19: Currently the paper switches from properties of a single model, to comparing two models (Model discrimination... section), back to properties of a single model (Properties of the posterior... section). The discussion might be more clear if the order of those last two sections were switched.

R2.5 We have carefully considered this suggestion and find that the other ordering has drawbacks as well. We have therefore retained the current order with some tweaks to the test.

Page 36 (Box 1): It is perhaps confusing that the update rule for \Theta_J -> \Theta_{test} shown does not use the Hessian guidance that is used in the actual application. It would be helpful to note than many proposal distributions are possible, and the one used in the application is more complex than the symmetric one shown.
We appreciate the reviewer's commentary, however, Hessian guidance is not used in the entirety of the walk; it is applied after simulation annealing is complete. Furthermore, the purpose of the Box, and specifically that particular equation, is to introduce the reader to an overall introduction to the Metropolis-Hastings, which does not necessarily use a Hessian matrix to generate update positions.

Typos:
Page 4: "experimental *** using genetic algorithms"

R2.6 Fixed. It is now "experimental data."

Page 21: "applicable to all cellular *** rather than priors"

R2.7 Fixed. It is now "cellular functions."

Reference duplicates: Gutenkunst (2009a, 2009b); Klinke (2009a, 2009b); Spencer (2009a, 2009b)

R2.8 Fixed.

4th Editorial Decision 26 November 2012

Thank you again for submitting your work to Molecular Systems Biology. We have now finally heard back from reviewer #1 whom we asked to evaluate your revised manuscript. I am pleased to say that this reviewer is now supportive and we will thus be able to accept your study for publication pending the following minor points:

- I could not find the code for Bayes factor calculations by thermodynamic integration; since this may represent a non-trivial aspect of the computational analysis, we would kindly ask you to include this to the BayesSB resource.

Reviewer #1 (Remarks to the Author):

I think the authors made an honest effort to address the concerns raised by all reviewers. The manuscript reports an interesting study. The novelty of the study is in the use of Bayesian methods to analyze a large model for a cell signaling system.

2nd Revision - authors' response 14 December 2012

We would like to thank you again for the care with which our paper "Properties of cell death models calibrated and compared using Bayesian approaches" has been edited and reviewed and for the excellent suggestions for a final round of revision. We hope the paper proves widely useful to the field and we are fully committed to further developing the methods and code.

In the current draft we have incorporated the following changes, as raised by the editors:

- A further modification of the title in which "second generation" is removed and antecedent work (the "first generation" described fully in the manuscript itself.

First, and most importantly, we now correctly reference Xu et al (2010) and have scoured our paper to remove any inappropriate claims of novelty. Methods are placed in the context of previous applications of Bayesian calibration to biological systems and the relevant literature is cited insofar as we can find it.

References to Bayesian estimation for biochemical models now include:
- Battogtokh D, Asch D, Case M, Arnold J, Schuttler HB (2002) An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of Neurospora crassa. Proceedings of the National Academy of Sciences 99: 16904.
idth=211 x height=157 pixels)
- Flaherty P, Radhakrishnan ML, Dinh T, Rebres RA, Roach TI, Jordan MI, Arkin AP (2008) A dual receptor crosstalk model of G-protein-coupled signal transduction. PLoS Comput Biol 4: e1000185.
- Klinke DJ (2009a) An empirical Bayesian approach for model-based inference of cellular signaling networks. BMC bioinformatics 10: 371.
- Bois FY (2009) GNU MCSim: Bayesian statistical inference for SBML-coded systems biology models. Bioinformatics 25: 1453-1454.
- Vyshemirsky V, Girolami M (2008) BioBayes: a software package for Bayesian inference in systems biology. Bioinformatics 24: 1933-1934.
- Xu TR, Vyshemirsky V, Gormand A, von Kriegsheim A, Girolami M, Baillie GS, Ketley D, Dunlop AJ, Milligan G, Houslay MD, Kolch W (2010)

Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species. Sci Signal 3: ra20.

Second, the revised title, text and abstract of the paper have been changed to more accurately reflect the actual contribution of our work relative to antecedences; in part this involves bringing back points made in an earlier draft.

Third, we have substantially advanced the Python source code and made it available for download at the open-source repository Github under the name of BayesSB. We continue to develop and extend the software and to integrate BayesSB with our PySB package, and we will make this development path transparent to potential users. We are also exploring ways to make Bayesian calibration available as a virtual machine on the cloud (probably via the Amazon elastic compute cluster; see sorgerlab.github.com/bayessb).

Fourth, the discussion of MCMC walks on page 10-11 has been moved into a supplementary note to make the paper more accessible. We think the discussion of Bayes factor on page 23 is important however, and rather than banish it to the supplement we have tried again to write it in an accessible way. If you think we have still failed, we are willing to cut the text.

Thank you again for your support.

