8th October 2019

|  |  |
| --- | --- |
|  |  |

Tim Holt   
Senior Publishing Editor   
Journal of The Royal Society Interface

Dear Dr Holt,

First and foremost, we want to thank you for the reviews we received (and, of course, the reviewers for their insightful comments). Following the reviewers’ comments, we have adapted our manuscript and we outline these changes below, where we respond to specific suggestions or concerns.

Second, we were also pleased that the reviewers largely agreed that the subject matter and approach taken were worthwhile. Reviewer 1 said, “I enjoyed reading this excellent piece of work”, and indicated, “This framework has very wide potential interest and I think that it will be adopted by the modelling community because the presentation is lucid and high-quality codes are made available by the authors.” Reviewer 2 agreed that the subject matter was worthwhile and that, “Estimation of parameter distributions from single cell data is an extremely challenging problem, so new methods to tackle it are always valuable contributions” and that, “The method the authors present has multiple advantages (e.g., better computational scalability with number of single cells due to approximation of density)…”. Reviewer 3 agreed that the subject matter was worthwhile although, admittedly, was more circumspect with their praise.

We do recognise, however, that there were omissions on our original submission which, following the reviewers’ comments, we are able to address. To summarise the reviewers’ main concerns, they fall broadly under the following headings:

1. The lack of discussion of nonlinear fixed effects models and how these methods have been used to model single cell data.
2. That our method concerns only underdetermined models.

We are grateful that the reviewers pointed out each of these concerns. In response to 1., we have added a paragraph to the introduction. Within this we do, however, indicate that the nonlinear fixed effect model literature (including all the references the reviewers included) concerns problems of a different nature to ours: the data for fitting nonlinear fixed effects models has trajectories of individual cells through time; our data is a snapshot of one cell’s properties at a single timepoint. To reliably estimate the cell-specific effects of mixed effect models, typically requires more data than a single observation per cell – otherwise, it is hard to differentiate between parameter variability (i.e. random effects) and measurement uncertainty. Additionally, mixed effects methods do not scale well with data size because each cell requires its own individual simulation. Our method does not suffer from this scalability issue because as inputs it uses distributions fit to raw data opposed to the raw data itself.

In summarising point 2., reviewer 2 is concerned of the, “poor performance of the method under poor parameter identifiability”. We feel that because we did not previously make clear from the outset that our method concerns only underdetermined models, it was possible to mistake the lack of identifiability of parameters for a shortcoming of the method. Of course, in underdetermined models, there are more than a single parameter set consistent with data; by their nature, these models – which represent the majority of models used in computational biology – are not fully identified. That our method does not resolve parameter values is then correct and, we should like to highlight, that any method that does estimate parameter values with high precision is, at best, overconfident; at worse, disingenuous. In response to this review, and also from similar comments from reviewer 3, we have made changes to the abstract that indicate from the outset that our method concerns only underdetermined models. To address reviewers 3’s related concern that, “Where parameters are simulated, only the negative effect of bad priors on the inference of the unknown parameters is discussed”, we have added an additional result to Section X.

In light of the above, we kindly request that you reconsider the previous decision to reject the manuscript. We recognise the hard demands that are placed on editors’ time, however, and appreciate the time taken to read our arguments.

Please find attached a reworked version of our manuscript to reflect the reviewers’ comments (there are actually two versions: one with the edits marked up in blue and, another, unmarked version).

We look forward to hearing from you in the near future.

With best wishes

Ben Lambert

**Referee: 1**  
  
Comments to the Author  
I enjoyed reading this excellent piece of work. Lambert et al. are absolutely right to point to the fact that variability is a major challenge in interpreting biological and biochemical experiments and applications.  Personally I am always surprised that more theoretical tools have not been developed to address these kinds of concerns because these questions are wide open for theoreticians to tackle them.

In this work the authors present a very general framework, based on ODEs, nearly called HODEs to incorporate heterogeneity, and use Bayesian methodologies to work with sparse time snap shots to resolve the posterior densities of the heterogeneity.  This framework has very wide potential interest and I think that it will be adopted by the modelling community because the presentation is lucid and high-quality codes are made available by the authors.  I note that the figures are high quality and the manuscript reads well without any significant typos.

* We are pleased to hear that the reviewer liked the method and presentation and also agree that this area of application (fitting to snapshots of cell properties) is underrepresented.

I suggest the authors make a minor revision of their work to address the following questions before publication:  
  
1. The CMC is presented in terms of working with a series of time snapshots and working with a system of deterministic ODEs where the parameters vary from cell-to-cell.  This is entirely reasonable, but I see the framework as being very flexible with potential extensions that might include many other features such as measurement error in the data, generalising the deterministic ODEs to working with stochastic ODEs and even working with spatial problems in terms of PDEs.  The CMC framework, in my view, extends to all of these cases and some brief discussion about how such extensions would be achieved would be useful.

* We agree that more extensions than those discussed previously are possible and thank the reviewer for mentioning these. We have now added the following paragraph to the discussion to include these, “Whilst we have illustrated our approach by fitting ODE models to data, we recognise that our approach is applicable to deterministic forward models in general. These include a large swathe of models used in computational biology, such as partial differential equations and difference equations. Similarly, whilst we have illustrated our approach by fitting to models with time-invariant parameters, it could also be used to determine how parameters vary throughout the course of an experiment - so long as the dynamic evolution of parameter values is itself parameterised.”
* Unfortunately, we have not yet determined how to tackle the issue of stochasticity in the forward map – be it due to inherent stochasticity (as in an SDE) or measurement noise.

We hope that others, like the reviewer, recognise the worth of this method and, by publishing the article in JRSI, that others may also be motivated to take up the task of extending the approach.

2. I wonder how well the method is able to resolve problems where the parameters evolve as a function of time?  This kind of temporal behaviour is observed in experiments of cell migration where some kind of adaptation takes place, or some kind of nutrient resource is exhausted during an experiment.  Can the authors comment on the ability of the approach to resolve temporal variations in parameter values?

* This is a good suggestion and we have now added, as previously noted, the following sentence to the discussion, “Similarly, whilst we have illustrated our approach by fitting to models with time-invariant parameters, it could also be used to determine how parameters vary throughout the course of an experiment - so long as the dynamic evolution of parameter values is itself parameterised.”

3. A big question in my mind is hinted at in the discussion, and this is the question of model selection.  I agree with the authors that a nice feature of the CMC is that given some posterior distribution you could run forward simulations and compare with data.  Of course, when we work with a mathematical model we are always forced to include certain features and to exclude other features that are thought to be less important.  If the forward simulations do not compare well with the data then perhaps the model is incorrect (or incongruent as the authors state, very politely).  Since the CMC is Bayesian, it should be able to be coupled with Bayesian selection tools (e.g. Bayes' factors or some other flavour of information criteria AIC, BIC etc) and I think explicitly pointing this out would be useful.

* This is an astute point that had eluded us, thank you. We have now included a description of how CMC could be used for model simplification in our discussion with the following paragraph, “In Figure 4, we present the workflow for our approach, which includes as its last step comparing output samples with the target distribution. As discussed above, if output samples do not correspond with the target, this may indicate that a model isn't fit for purpose. Conversely, if there is correspondence with the target distribution, it is possible that a simplified model -- with (say) one or more fewer parameters -- could also recapitulate the same results. Thus, a process of repeated rounds of model simplification then CMC could be pursued until output samples no longer correspond with the target. The most parsimonious model would then be the simplest case where the output samples still match the target. We note however, that may be dangerous when using the most parsimonious model to predict other functionals”

**Referee: 2**  
Comments to the Author  
The authors present a novel method to infer parameter distributions from single cell data from heterogenous distributions, which is a very relevant topic in the life sciences at the moment. The authors showcase their method on a large number of different examples, which includes models with non-identifiable parameters.  
  
Estimation of parameter distributions from single cell data is an extremely challenging problem, so new methods to tackle it are always valuable contributions.

* We are glad that the referee recognises the lengths we go to in order to showcase our method. We also agree with them – inference from single cell data *is* very challenging and we believe merits more attention in the literature.

Despite their careful evaluation of the method, a numerical comparison (convergence with sample size, robustness to noise, computation time) with previously developed methods is missing. Furthermore, it seems that the authors are unaware of the development of nonlinear mixed effect models, which are exactly what they have now denoted as "HODES" (see e.g.(Dharmarajan, L. et al. A simple and flexible computational framework for inferring sources of heterogeneity from single-cell dynamics. Cell Systems), (Fröhlich, F et al. Multi-experiment nonlinear mixed effect modeling of single-cell translation kinetics after transfection. NPJ systems biology and applications), (Almquist, J. et al. A Nonlinear Mixed Effects Approach for Modeling the Cell-To-Cell Variability of Mig1 Dynamics in Yeast. PLoS One), (Karlsson, M. et al. Nonlinear mixed-effects modelling for single cell estimation: when, why, and how to use it. BMC Syst. Biol).

* We do believe, however, that the nonlinear mixed effects literature tackles a different problem to ours. To estimate models with mixed effects (with random effects for each individual cell), it typically requires more than one observation per cell. Otherwise, it is hard to differentiate between true cell-to-cell variations versus measurement noise. We believe that this distinction is important and now indicate this from the outset, with the following sentence in the abstract, *“*Nor are they [the data] amenable to standard nonlinear mixed effect methods, since a single observation per cell is typically too few to estimate parameter variability.”
* We thank the reviewer for pointing out this literature to us. We have now included a discussion of these approaches in our introduction, “Since HODEs assume the state of each cell evolves continuously over time, experimental data tracing individual cell trajectories through time constitutes a richer data resource. Fluorescent Recovery After Photo-bleaching (FRAP) is one such method, which follows the time-dependent response of cells after an initial bleaching (Karlsson et al., 2015). Methods exists, broadly under the banner of ``nonlinear mixed effects models'', which uses cell trajectories - individual time series of cellular quantities - to estimate both cellular variation and qualities of measurement noise. See, for example, Karlsson et al., 2015, Zechner et al., 2014, Dharmarajan et al., 2019. The demands of obtaining such data are, however, higher and typically involve either tracking individual cells through imaging methods (Hilsenbeck et al., 2016), or trapping cells in a spatial position where they can be monitored over time (Fritzsch et al., 2012). These techniques impose severe restrictions on experimental practices meaning they cannot be used in many circumstances, including for online monitoring of biotechnological processes or analysis of in vivo studies. For this reason, ``snapshot'' data continues to play an important role for determining cell level variability in many applications and in this paper we restrict analysis to only such data.”

The method the authors present has multiple advantages (e.g., better computational scalability with number of single cells due to approximation of density) but probably also disadvantages (e.g. robustness to noise/non-identifiability).

* We are pleased that the reviewer recognises the advantages of our approach but do take on board the comment that our method has disadvantages – namely, that our approach does not currently allow for measurement noise. We recognise this explicitly in the discussion, with the following paragraph, “Failure to reproduce a given output distribution can also indicate that the generating model (the priors and the forward model) are incongruent with experimental results. This may either be due to misspecification of the ODE system or because the assumption of a deterministic forward model is inappropriate. Our approach currently assumes that output variation is dominated by cellular variation in the parameter values of the underlying ODE, with measurement noise making a negligible contribution. Whether this is a reasonable assumption depends on the system under investigation and, more importantly, on experimental details. We recognise that neglecting measurement noise when it is, in fact, important in determining observed data means CMC will overstate cellular variation. It may also mean that some output distributions cannot be obtained by our model system (i.e. HODEs without noise). Future work incorporating a stochastic noise process or, more generally, including stochastic cellular mechanisms is thus likely to be worthwhile.”

Major Issues:  
  
A) The poor performance of the method under poor parameter identifiability is a crucial shortcoming of the method as the authors claim that their method allows the identification of sources of variability. If the reconstructed parameter distributions are not accurate, the identification of sources of variability cannot be trusted. Accordingly the authors should more carefully evaluate how well their method can actually reconstruct parameter distributions. In particular the finding that reconstructed parameter ranges do not even contain the true parameters for the TNF example suggests that the method is not really doing anything meaningful. If this is really only due to the imposed prior, weaken the prior.

* We apologise for our previous presentation where we did not make it explicit that our approach applies only to underdetermined models – those where all parameter values cannot explicitly be determined from data. To highlight this distinction, we now include the following sentence in our abstract, “Our method is appropriate for underdetermined systems where observed variation is mostly due to variability in cellular processes rather than experimental measurement error, which may be the case for many systems due to continued improvements in resolution of laboratory techniques.”

Overall the authors seems to have a couple of misconceptions around the issue of non-identifiability that should be clarified (and accordingly corrected in the paper):  
  
1) The presence of non-identifiability in a non-linear ODE model does not necessarily depend on the number of QOIs/datapoints to parameters, but requires more advanced methods such as profile likelihood or bayesian sampling to be rigorously evaluated

* This is true and we thank the reviewer for highlighting this. We have not clarified our use of the word underdetermined as follows, “Additionally, our approach is suitable only for underdetermined models which we define as the case where there are fewer parameters than output quantities of interest.”

2) Bayesian priors should not be used to render problem identifiable just for the sake of rendering the problem identifiable. Priors should encode prior beliefs/information about the system that were drawn from previous experiments or literature.

* We agree with the distinction in theory. In practice, however, the construction of priors and data analysis is increasingly being considered in tandem. Often the model fitting indicates that parameters aren’t identified and this can force a modeller to re-evaluate their prior prejudices – moving towards more informative priors that more accurately represent their viewpoint. See, for example, Schad et al., 2019, "Toward a principled Bayesian workflow in cognitive science." *arXiv preprint arXiv:1904.12765*.

B) The difference between snapshot vs timecourse data is not sufficiently stressed in the manuscript and possibly also incorrectly accounted for in the method. In timecourse data, there is correlation between QOIs at individual timepoints, whereas the sets of QOIs are assumed to be statisticially independent in snapshot data.

* We thank the reviewer for this suggestion. As noted above, we have now added the following discussion in our introduction that highlights this distinction, “Since HODEs assume the state of each cell evolves continuously over time, experimental data tracing individual cell trajectories through time constitutes a richer data resource. Fluorescent Recovery After Photo-bleaching (FRAP) is one such method, which follows the time-dependent response of cells after an initial bleaching (Karlsson et al., 2015). Methods exists, broadly under the banner of ``nonlinear mixed effects models'', which uses cell trajectories - individual time series of cellular quantities - to estimate both cellular variation and qualities of measurement noise. See, for example, Karlsson et al., 2015, Zechner et al., 2014, Dharmarajan et al., 2019. The demands of obtaining such data are, however, higher and typically involve either tracking individual cells through imaging methods (Hilsenbeck et al., 2016), or trapping cells in a spatial position where they can be monitored over time (Fritzsch et al., 2012). These techniques impose severe restrictions on experimental practices meaning they cannot be used in many circumstances, including for online monitoring of biotechnological processes or analysis of in vivo studies. For this reason, ``snapshot'' data continues to play an important role for determining cell level variability in many applications and in this paper we restrict analysis to only such data.”

To avoid correlation between simulated QOIs across timepoints, it is necessary to generate an independent set of parameter samples and simulations for every considered timepoints at which QOIs are collected. This introduces a dependence of the computation time on the number of timepoints, which is not a huge problem as the number of timepoints in snapshot data is usually small, but should be mentioned in the discussion.

* Our method does not resolve individual cell trajectories since it is fit to ``marginal’’ snapshot distributions. As such, we do not need to generate independent simulations for each cell/time point. This is in stark contrast with existing approaches that must simulate once per each cell. This is a benefit of our approach that we believe is worth highlighting as we now do with the following sentence in our abstract, “The CMC algorithm fits to snapshot probability distributions rather than raw data, which means its computational burden does not, like existing approaches, increase with the number of cells observed.”

C) How realistic is the assumption that the conditional distribution of parameters (line 277) is independent of data? I would think that if the distribution of parameters is independent of data, any kind of inference is a silly endeavor.

* We thank the reviewer for their comment here. We do believe that the proof, however, is somewhat in the pudding – our method seems to generate sensible results when we treat this object as independent of data; that is, as a prior.

Minor Issues:  
I would recommend the authors to perform their sampling in logarithmic parameter coordinates as parameter distributions are usually log-normal  
  
Why suddenly use a histogram to compare densities in figure 11B? Is the approximation of the target QOI distribution with a gaussian density really necessary? Why not use a vine KDE approximation?  
  
  
**Referee: 3**  
Comments to the Author  
The paper addresses estimation of parameter heterogeneity in single-cell response models based on an MCMC approach.  
  
The topic is relevant for cellular biology, and the paper is reasonably well-written. However I see several fundamental problems with this contribution:  
  
1. The authors do not seem aware of relevant literature, which starts being rather rich (in the following, I only give a partial account). Inference of heterogeneous single-cell response models has been already treated under the name of Mixed-Effects modelling/inference eg in:  
  
Llamosi A, Gonzalez-Vargas AM, Versari C, et al. What Population Reveals about Individual Cell Identity: Single-Cell Parameter Estimation of Models of Gene Expression in Yeast. PLoS Comput Biol. 2016;12(2):e1004706.  
  
In fact, Mixed-Effects modelling is a well-known framework from Pharmacokinetics that can well represent the name of the paradigm that the authors are looking for (authors call it "HODE"), precisely aiming at estimating variability of individual response parameters. Though in the above reference inference is from single-cell profiles, the modelling approach may well apply to snapshot data.  
  
Bayesian inference of individual response variability with MCMC approaches has been treated eg in:  
Zechner C, Unger M, Pelet S, Peter M, Koeppl H., Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. Nat Methods 2014, 11(2):197-202  
  
From population snapshot data, inference of individual-cell parameter variability is also part of the work:  
Zechner C, Ruess J, Krenn P, et al. Moment-based inference predicts bimodality in transient gene expression. Proc Natl Acad Sci U S A. 2012;109(21):8340–8345  
  
Relevant work has also been carried out (in this and other works) by the authors of:  
Gregor Neuert, Brian Munsky, Rui Zhen Tan, Leonid Teytelman, Mustafa Khammash and Alexander van Oudenaarden, Systematic Identification of Signal-Activated Stochastic Gene Regulation , Science, vol. 339: no. 6119, pp. 584-587  
  
Additional work where simulation-based approaches are used to fit snapshot data is scattered through the computational biology literature.

* We thank the reviewer for these detailed references. As noted previously, we have now included the below paragraph in our introduction to address the applicability of our methods to our specific problem, “Since HODEs assume the state of each cell evolves continuously over time, experimental data tracing individual cell trajectories through time constitutes a richer data resource. Fluorescent Recovery After Photo-bleaching (FRAP) is one such method, which follows the time-dependent response of cells after an initial bleaching (Karlsson et al., 2015). Methods exists, broadly under the banner of ``nonlinear mixed effects models'', which uses cell trajectories - individual time series of cellular quantities - to estimate both cellular variation and qualities of measurement noise. See, for example, Karlsson et al., 2015, Zechner et al., 2014, Dharmarajan et al., 2019. The demands of obtaining such data are, however, higher and typically involve either tracking individual cells through imaging methods (Hilsenbeck et al., 2016), or trapping cells in a spatial position where they can be monitored over time (Fritzsch et al., 2012). These techniques impose severe restrictions on experimental practices meaning they cannot be used in many circumstances, including for online monitoring of biotechnological processes or analysis of in vivo studies. For this reason, ``snapshot'' data continues to play an important role for determining cell level variability in many applications and in this paper we restrict analysis to only such data.”

2. The originality of this work relative to approaches in the computational biology literature is the focus on underdetermined systems, where unknown individual parameters are more than measurements. Though the analysis proposed is interesting in itself, it does not cite any of the identifiability studies in the literature (as eg as part of some of the works above, and more works by the same authors and others).

* We have now cited the above studies in our introduction.

More ackwardly, it focuses on showing that the wrong Bayesian priors drive parameter estimates far from their actual values (or regions of variability), although the observed snapshots are well fitted. In the interest of the application (which seems to be the main goal of the paper, see also point 3 below), I don't see the utility of this. Little is said in the paper on how well variability of parameters can be estimated even in presence of unidentifiability (for instance, by a - Bayesian ? - inference of all equivalent parameters compatible with the observations, which should ideally include the true parameter distributions). Surprisingly, in most of the examples given, parameters do not even seem to be simulated, while the authors directly generate arbitrary snapshot measurements. Where parameters are simulated, only the negative effect of bad priors on the inference of the unknown parameters is discussed (Section 4.3.1). Finally, no word is spent on the relevance of these underdetermined problems: In practice, people try to avoid them either by collection of more data (as is most often the case in practice), or by model reduction (notably to address structural identifiability issues).

* We agree that we may have gone too far to emphasise how underdetermined models are, by their nature, typically un-identified.

3. It is not clear what the contribution of this paper is relative to the so-referred companion paper [22]. The CMC method is already presented in [22], while here only a breif - not very transparent - account of the mathematical principles of [22] is provided (section 3.2). One then concludes that the originality of the submission under review here is the application and not the method. As such, though, the contribution is rather limited (three simulated examples showing more the limitations inherent in underdetermined problems than the practical utility of the method for parameter reconstruction).

* We thank the reviewer for this comment. We do believe, however, that we have made significant attempts to highlight that this (unpublished) ``companion’’ paper deals with the theory behind the method in detail. For example, in our methods, we say, “See our companion paper [22] for a more comprehensive discussion”, and, “we defer to our companion piece for detailed explanation of eqs. (10) and (11).
* We also believe that the comment is perhaps, a little uncharitable: we believe this additional paper – whilst adding little to our research output, being unpublished – does help a reader to understand the theoretical underpinnings of our approach. There is currently a movement within funding bodies such as the Wellcome trust, to advocate for additional resources, be they tutorial papers, videos and dynamic code help to make research more generally useful for readers. Alongside this paper, we essentially provide an accompanying tutorial paper and, additionally, fully-commented Jupyter Python notebooks with code blocks that readers can run to recapitulate our results. As such, we ask that these additions be taken as complementary rather than detracting from this publication.

Additional (technical) comments follow:  
  
- The focus on underdetermined problems is not clear from title abstract and intro.

* We have sought to address this deficiency.

- The lack of account for measurement error is practically a severe limitation. This is generally not the case in computational biology and, given the typically non-negligible strength of measurement noise, most methods in the literature above consider it in a way or another.

* We agree this is a limiting case, but argue it is a useful one to consider. Butler et al provide an a priori bound on the recovered distributions

- Eq. (4)-(5) and discussion around it: Definitions are a bit confusing, they are initially given for generic QOIs but then eventually reduced to same functionals measured at several times (eg snapshots)

- First paragraph of page 8: It seems to me that these examples are only correct if m=p-1 (only one dimension unidentifiable), not for generic m<p as claimed .

* Extension to the method can be found in Butler, Estep and Tavener….

- Eq.(8) - (9) etc: The derivations using dirac delta functions are not very clear, a more proper derivation would be desirable.

* ???

- Page 15: These discussions on sensitivity and experiment design are all relevant and legitimate but fail to account for literature treating these points (and for exp. design, the choice of the best experiment also depends on the true unknown parameters, which makes the application of your method to this purpose not obvious)

* ???

- Results section: In none of the examples parameter posteriors are visually compared with their actual distributions to be recovered (unfortunately, these distributions are in fact not even defined in most examples)

* ???

- Section 4.1.2: First paragraph is redundant and should anyhow come in the section before

* Agreed. This has been moved earlier

- Section 4.2.2: Here you increase the number of unknowns to make the problem underdetermined. This is odd, no-one would do this in practice, but rather prefer that a method that works for underdetermined problems also works for well- or over-determined ones. A comment on this is necessary.

* Methods for under and over-determined systems are necessarily very different. We have commented.

- Figure 9: It is not obvious to me that if your priors cover some marginals they also cover the whole joint distribution

* ???

- Section 4.3.1:; Again, first paragraph is redundant

* We disagree

- Figure 11: Are units of the different parameters compatible, thus justifying a single plot for all of them?

* Figure 11 is a complicated figure and we have extended our discussion

- Discussion: The mentioned approaches for stochastic dynamics that you claim being necessary already exist (though of course not in the context of your approach, see literature above).

* We have added references to some of this extensive literature, while continuing to claim the differences and advantages of our approach.