Reply to Comment on "Sloppy models, parameter uncertainty, and the role of experimental design"

David R. Hagen[a,b,†], **Joshua F. Apgar**[a,b,†,‡], **David K. Witmer**[b,c,†], **Forest M. White**[a,d], and **Bruce Tidor**[a,b,c,*]

[a]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[b]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[c]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[d]David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

We welcome the commentary from Chachra, Transtrum, and Sethna[1] regarding our paper "Sloppy models, parameter uncertainty, and the role of experimental design,"[2] as their intriguing work shaped our thinking in this area.[3] Sethna and colleagues introduced the notion of sloppy models, in which the uncertainty in the values of some combinations of parameters is many orders of magnitude greater than others.[4] In our work we explored the extent to which large parameter uncertainties are an intrinsic characteristic of systems biology network models, or whether uncertainties are instead closely related to the collection of experiments used for model estimation. We were gratified to find the latter result — that parameters are in principle knowable, which is important for the field of systems biology. The work also showed that small parameter uncertainties can be achieved and that the process can be greatly accelerated by using computational experimental design approaches[5–9] deployed to select sets of experiments that effectively exercise the system in complementary directions.[2]

The comment by Chachra *et al.* does not disagree with any of these points, but rather emphasizes two quantitative issues.[1] Firstly, even when all parameter combinations have small uncertainties, the fit model can still be sloppy in that some parameter combinations are known orders of magnitude better than others (in our paper this ratio of uncertainties was around 300).[1,2] This is certainly correct, although to truly ask whether sloppiness is inherent in the model or is due to the experiments used for fitting, one should apply optimal experimental design to the objective of minimizing sloppiness. We have done an initial trial and were able to establish all parameter directions to near 10% or less uncertainty while reducing the ratio to 55, and we expect that with more effort further reductions could be achieved.

Secondly, Chachra *et al.* commented that the quantity of data required to achieve small parameter uncertainties could be large.[1] We certainly agree. In our paper we effectively used

---

[*]Correspondence tidor@mit.edu.
[†]These authors contributed equally
[‡]Current address JF Apgar, Boehringer Ingelheim Pharmaceuticals, 900 Ridgebury Road, PO Box 368, Ridgefield, Connecticut 06877-0368, USA

3,000 individual measurements spread across five experimental perturbations (600 data points per experiment), each measurement with the relatively high precision of 10%, to fit just 48 parameters. A greater number of less precise experimental measurements would be required, but the number could be decreased if less precision in the fit parameters were required. As but one example of how this tradeoff plays out in the example used in our paper,[2] if the total number of measurements were reduced from 600 data points per experiment to just 68, then 13 experimental perturbations are required. If the experimental uncertainty were then doubled from 10% to 20%, then the required number of perturbations would increase further to 33, but if the desired parameter uncertainty then were to similarly double from 10% to 20%, the number of experimental perturbations would return to 13 (and is, in fact, a mathematically equivalent problem with an identical set of solutions). It should be noted that we did not optimize the selection of species or time points to measure, although it is known that not all contribute equally,[7–9] and our techniques applied to species and time point selection could presumably lead to significant data reductions. This consideration, coupled with dramatic increases in capacities of new technologies for making large-scale measurements in systems biology, makes it is less likely that data limitations will be determining. Moreover, the application of optimal experimental design computations to plan experimental campaigns should then be increasingly useful to strategically plan experiments. This example emphasizes the tradeoff between the number of measurements per experimental perturbation and the number of experimental perturbations. Depending on the relative effort of producing one or the other, an appropriately customized campaign could be developed.

Finally, it remains an unanswered question just how accurately parameters need to be known to achieve accurate predictions. One of the notions arising from the concept of model sloppiness is that *some* predictions can be made quite accurately with very inaccurate parameters,[3] but this is of little use without a method for knowing when one is in this situation. Propagation of parameter uncertainty is one approach to estimating prediction accuracy. By clarifying the link between parameter uncertainty and experimental conditions, our work points to another approach.[2] Because the link between parameter uncertainty and experimental conditions extends to experiments that have yet to be done (namely, predictions), combinations of experimental perturbations and measurements that would not further reduce parameter uncertainty significantly are expected to be well represented by the current model and should be relatively high confidence predictions. We are investigating this relationship in more detail.

References

1. Chachra R, Transtrum MK, Sethna JP. Mol. Biosyst. 2011 **XX**, XXXX-XXXX.

2. Apgar JF, Witmer DK, White FM, Tidor B. Mol. Biosyst. 2010; 6:1890–1900. [PubMed: 20556289]

3. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. PLoS Comput. Biol. 2007; 3:1871–1878. [PubMed: 17922568]

4. Brown KS, Hill CC, Calero GA, Myers CR, Lee KH, Sethna JP, Cerione RA. Phys. Biol. 2004; 1:184–195. [PubMed: 16204838]

5. Casey FP, Baird D, Feng Q, Gutenkunst RN, Waterfall JJ, Myers CR, Brown KS, Cerione RA, Sethna JP. IET Syst Biol. 2007; 1:190–202. [PubMed: 17591178]

6. Cho KH, Shin SY, Kolch W, Wolkenhauer O. Simulation. 2003; 79:726–739.

7. Donckels BMR, De Pauw DJW, Vanrolleghem PA, De Baets B. J. Comput. Chem. 2009; 30:2064–2077. [PubMed: 19165773]

8. Faller D, Klingmuller U, Timmer J. Simulation. 2003; 79:717–725.

9. Kutalik Z, Cho KH, Wolkenhauer O. Biosystems. 2004; 75:43–55. [PubMed: 15245803]