

THE ISIF PRESIDENT LOOKS IN THE MIRROR

CONTEMPLATING HOW TO MAKE A DIFFERENCE

Sitting in Liverpool in early 2020, it was clear that COVID-19 was going to present a challenge to global health, and also to global science. The deluge of COVID-19 papers and the apparent inability for that corpus to be rallied to influence reality had already begun to be seen. The question that quickly arose was then: “How could one team in Liverpool make a difference?”

Prior to COVID-19, Liverpool researchers had begun to focus effort on “Big Hypotheses” [1]¹, a five year project begun in 2018 to develop a game-changing ability to use large computers to make statistical inferences from noisy data. This activity was initially motivated by the observation that, while the computational resources used to apply Deep Learning are doubling every four months [2], the resources used in the context of numerical Bayesian inference (i.e., Markov Chain Monte Carlo (MCMC)) have stagnated to be those on researchers' desktops. While Bayesian inference can operate effectively in the “data-starved” and “understanding-rich” contexts where Deep Learning can struggle, Bayesian inference needs to evolve to be able to compete. Big Hypotheses' initial aim is to make it possible to use N computers to make numerical Bayesian inference run N times faster, where N might be large (e.g., 86,400: the number of seconds in a day). One might think that this can be achieved by simply implementing MCMC using languages that are amenable to distributed implementation on High Performance Computing (HPC) or in the cloud. Unfortunately, that is necessary but not sufficient: MCMC has an initial burn-in phase which, in general, is very challenging to parallelize since it involves an inherently sequential process of taking a sequence of steps, each of which brings the algorithm a tiny bit closer to convergence.¹ The solution that Big Hypotheses adopts is to replace MCMC with an alternative algorithmic work-horse, the Sequential Monte Carlo (SMC) sampler. The Liverpool team believe that SMC samplers can be configured to implement parallelized numerical Bayesian inference. Indeed, because SMC samplers have different constraints to MCMC, the ultimate (overtly ambitious but, it is believed, achievable) vision for Big Hypotheses is energy-efficiency, whereby one computer can perform numerical Bayesian inference N times faster than MCMC.

Strong scaling of an existing algorithm, i.e., making it run faster is, perhaps counter-intuitively, rarely significantly op-

¹ Of course, you can run multiple short MCMC chains in parallel. If you stop each of them before burn-in has completed, this typically (but not always) degrades estimation performance very significantly.

erationally useful; if the algorithm is already being used, it is unlikely that freeing up resources will be game-changing. Strong scaling's utility is more likely to be associated with transforming problems that are assumed to be too time-consuming to be practically useful into operational systems. What strong scaling delivers is an ability to be substantially more ambitious in terms of the complexity or size of problems that can be considered. It is that ambition that is potentially game-changing.

For any game-changing ability to apply numerical Bayesian inference to experience widespread adoption, it would need not only to be readily applied to arbitrary problems but also to deliver benefit relative to existing state-of-the-art solutions that are accessible to the people who need to solve those problems. Probabilistic programming languages (PPL) are widespread across those end-users since they provide a flexible way for end-users to succinctly articulate their probabilistic model in a form that allows a state-of-the-art MCMC algorithm to be applied: the No-U-Turn-Sampler (NUTS) [3] underpins many PPLs and exploits local gradient information to efficiently explore the parameter space (even when it is high dimensional). Big Hypotheses therefore focuses on interfacing to a specific PPL, Stan [4] (named after Stanislaw Ulam), and on articulating benefit in the context of a portfolio of models that have been collated and curated to provide benchmarks for performance comparison [5].

Returning to early 2020, Liverpool researchers were, somewhat fortuitously, already working on combining data from each of multiple sources to detect outbreaks of infectious disease [6]. This work made apparent that there was already a pressing need to calibrate sophisticated models for the spread of infectious diseases. COVID-19 made clear that the world needed Big Hypotheses.

Unfortunately, Big Hypotheses wasn't ready.

INTRIGUE LED TO ENGAGEMENT

The UK government's response to the need to monitor the spread of COVID-19 was an emergency response; epidemiologists were rallied to inform government decisions. These scientists, and others already working with government who had relevant skills and experience, sensibly made use of tools and techniques that they had to hand or could rapidly produce.

Simon Maskell
University of Liverpool
Brownlow Hill
Liverpool, UK
s.maskell@liverpool.ac.uk



Figure 1
The University of Liverpool's Signal Processing Group in mid-2019 (just before COVID-19).

The modelling that was undertaken then helped inform decision making in the UK: the UK entered lockdown #1.

Various governmental bodies stood forward to try to combine the confusing mass of data that was being collated (e.g., from various applications (apps) that were each monitoring intersecting sets of facts about differently biased subpopulations of the UK) and the equally confusing mass of scientific literature and understanding that was beginning to emerge. Meanwhile, the Liverpool research team engaged with some of these government bodies and some of the aforementioned scientists that were helping inform government decisions.

At about this point in time, the vaccines started to emerge and the UK government app, which uses an unscented Kalman smoother to infer close contacts from Bluetooth signal strength [7] and that went on to make a significant impact on COVID-19 [8], was released. Both events were poignant for the Liverpool team: the team had tried to win funding to work on using Twitter to monitor the side effects of vaccines but had been unsuccessful because the reviewers took the view that there was a low chance that vaccines would exist; the team knew some of the developers of the app but were unsure if they should have diverted more of the Liverpool work towards supporting the app developers.

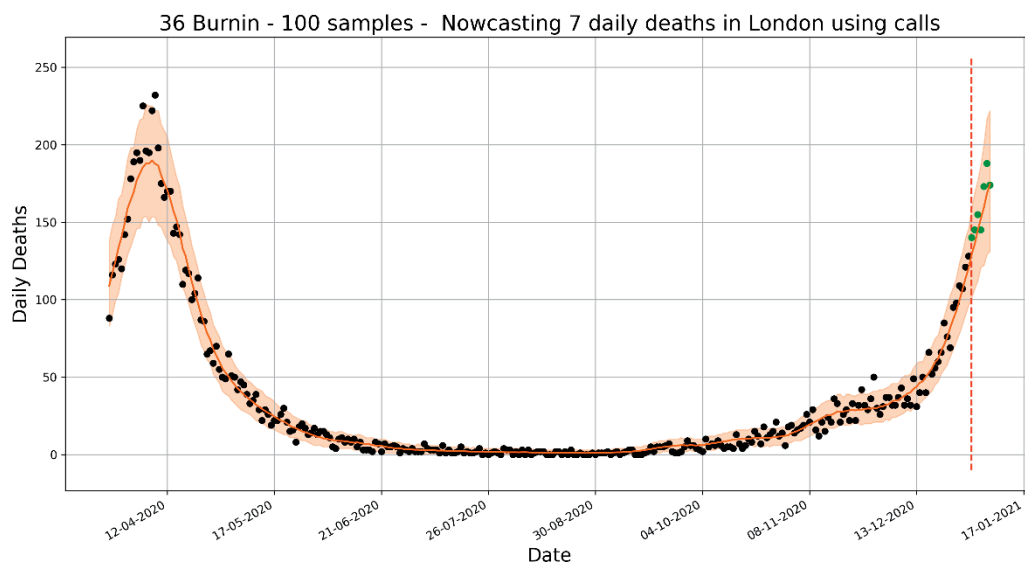


Figure 2
Combining low latency and high accuracy feeds can help inform difficult decisions: black dots are observed COVID-19 deaths in London, red line is prediction (with shaded area indicating associated confidence interval), and green dots are (retrospective) actual deaths.

The team had developed a processing chain for monitoring symptoms of COVID-19 reported in Twitter [9] in the hope that the biases present in those data would vary less with time compared to the COVID-19 tests. However, in the UK, the counts of geolocated Tweets that were indicative of symptoms were low. This motivated a focus on combining the high accuracy records of deaths with another low-latency feed: 111 calls and 111 online interactions.² By fusing the 111 data and the deaths, it was possible to calibrate the 111 calls as a low-latency forecast (or predictor in statistical language) of future death counts. This use of data fusion combined with Big Hypotheses' vision helped solidify the Liverpool team's engagement with the UK's Joint Biosecurity Centre (JBC): the JBC is charged with transitioning the emergency response into crisis-as-usual and adopting an engineering approach to using the noisy data to inform the difficult decisions that the UK government has to make.

While some epidemiologists were intrigued by the concept of Big Hypotheses, researchers at Imperial College London were arguably the most interested. Imperial had written a paper [10] on simultaneously analyzing multiple geographies to disentangle the impact of different nonpharmaceutical interventions (e.g., shutting schools, closing shops, lockdowns, etc.). They used Stan, but the scale at which they could apply the model was constrained by the inability to use HPC or cloud resources.

Imperial was ambitious. Big Hypotheses was not ready. Then it started to feel like the tide began to turn.

PROGRESS!

Stan is developed by an international team of researchers but its genesis and the center of gravity for its development is Columbia University in New York. Columbia won National Science Foundation funding to work with the Liverpool team to build COvid DAta MOdels (CODATMO) [11]. CODATMO was a response to different epidemiologists using different programming languages, making it hard to synergize ideas and approaches: CODATMO collates articulations (in Stan) of several Epidemiologists' models, the data they use when assessing such models, as well as frameworks for evaluating the models (including mechanisms for simulating epidemics, such that assessment can exploit known ground-truth). CODATMO aims to make it easier to extend, synergize, develop, and deploy such models. It has been picked up by researchers in Brazil [12] and it has intrigued Stan developers, who have contributed insights that have directly influenced the direction taken by the Liverpool team in their interactions with the JBC. Openness has delivered.

The Liverpool team have also been involved in the analysis (using Stan) of data from wider activities at Liverpool related to UK pilots of mass testing and of large scale events (a business

² 111 is the UK telephone number for a public call center that people call for urgent healthcare assistance that does not qualify as an emergency involving an immediate threat to life. The numbers of calls (and interactions with an associated website) that relate to COVID-19 are published by the UK government.

event, a rave, and a festival) with no social distancing. Interest also started to grow in Streaming-Stan [13], a variant of Stan that the Liverpool team had developed, as an off-shoot from Big Hypotheses, as a PPL for tracking problems: the team had several “knocks on the door” of people requesting to be beta-testers.

The sustained hard work of the Liverpool team then started to deliver. Big Hypotheses began to show its potential to achieve strong scaling; promising preliminary results in low dimensions and other promising results in the context of discrete variables emerged. The initial attempt to achieve strong scaling in high dimensions failed: NUTS can generate good samples in high dimensions, but the team couldn't shake off the need for burn-in. The second attempt did successfully avoid burn-in in a few high dimensional examples, but didn't provide the general-purpose Bayesian blunderbuss that the Liverpool team believe they can produce.

Today, the team sense they are close. Big Hypotheses oscillates between being close and feeling far away: it is not yet a reality.

REFLECTING ON THE PAST AND LOOKING TO THE FUTURE

The reality I now see is that decision makers need to learn how to balance advice from multiple scientists from diverse disciplines: epidemiologists, who understand the impacts of the disease on physical health; psychologists, who understand the impact of interventions on mental health; and economists, who understand the financial ramifications of these interventions. This advice derives from data from each of multiple disparate sources. A common reference frame needs to be defined to triangulate the data and models developed that articulate the scientists' uncertain, imprecise, conflicting, and ambiguous understanding in that reference frame. The parameters of the models need to be estimated from historical data, used to make predictions as time evolves and communicated to decision makers in such a way that they, and the public, can understand. This is Data Fusion for 2021.

So, as I look in the mirror, I ask myself:

1. Was the world lucky that scientists developed vaccines that were more effective than we could have hoped, and that the delta variant only arrived relatively late in the day?
2. Does the Fusion community need to embrace HPC and cloud computing environments, standardized datasets like those associated with CODATMO, and probabilistic programming languages like Stan?
3. Was it ambitious and yet sufficiently realistic to think that a hard-working and purposeful team in Liverpool could mature Big Hypotheses to produce a game-changing ability to perform numerical Bayesian inference for COVID-19?

4. Is Big Hypotheses tantalizingly close to demonstrating a revolutionary advance?
5. Is there still lots of interesting, important, and inspiring work left to do?
6. Was COVID-19 the warning shot for the Fusion community?

Yes!

With apologies to JFK: we choose to make Big Hypotheses a reality, not because it is easy, but because it is hard. Indeed, while I wish COVID-19 had not happened, it's fantastic to work with a team that continues to be spurred on by the belief that we are close to making a significant advance and is relentlessly driven forwards by the opportunity to make a difference. I hope the wider Fusion community can learn from our experiences in Liverpool and thereby make a significant contribution to the fight against a common enemy that measures less than a micron across and yet is having an impact that spans the planet. We must coordinate and collaborate if we are to help understand the spread of the pathogen and the utility of different interventions. Only then will we have helped win this fight and positioned the Fusion community to play the pivotal role it should when the next pandemic strikes.

REFERENCES

1. Engineering and Physical Sciences Research Council. Grant application. Available: <https://gow.epsrc.ukri.org/NGBOViewGrant.aspx?GrantRef=EP/R018537/1>.
2. Synced. AI doubling its compute every 3.5 months, <https://medium.com/@Synced/ai-doubling-its-compute-every-3-5-months-596b1b60fab>.
3. Hoffman, M., and Gelman, A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, Vol. 15, 1 (2014), 1593–1623.
4. Gelman, A., Lee, D., and Guo, J. Stan: a probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, Vol. 40, 5 (2015), 530–543.
5. GitHub. Stan dev, <https://github.com/stan-dev/posteriorodb>.
6. Jenkins, S., Bull, M., Abbs, S., Maskell, S., and Horridge, P. Biosurveillance and data fusion for early outbreak detection and classification. Presented at DTRA CBD S&T 2019, Cincinnati, OH, Nov. 18–21, 2019.
7. Lovett, T., Briers, M., Charalambides, M., Jersakova, R., Lomax, J., and Holmes, C. Inferring proximity from bluetooth low energy RSSI with unscented Kalman smoothers. arXiv preprint, arXiv:2007.05057.
8. Wymant, C., Ferretti, L., Tsallis, D., Charalambides, M., Abeler-Dörner, L., Bonsall, D., et al. The epidemiological impact of the NHS COVID-19 app. *Nature*, Vol. 594, 7863 (2021), 408–412.
9. Symptom Watch UK. Daily counts of UK COVID-19 symptom tweets, <https://pgb.liv.ac.uk/~johnheap/>.
10. Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, Vol. 584, 7820 (2020), 257–261.
11. GitHub. CODMATO, <https://github.com/codatmo>.
12. Storopoli, J., Ferreira dos Santos, A. L. M., Guedes Pellini, A. C., and Baldwin, B. Simulation-driven COVID-19 epidemiological modeling with social media. <https://arxiv.org/abs/2106.11686>.
13. Maskell, S., and Varsi, A. Towards an interface for streaming Stan. Presented at StanCon 2020. <https://www.youtube.com/watch?v=8dGhMYmyHnU>.

Simon Maskell is a Professor of Autonomous Systems at the University of Liverpool. At Liverpool, Simon leads the wider Signal Processing group and is also principal investigator for a number of research projects, mostly funded by EPSRC and Dstl with a focus on synergistic use of next-generation algorithms and parallel computing resources. Simon is also director of the EPSRC Centre for Doctoral Training in Distributed Algorithms, a prestigious grant to train the next generation of data scientists through 60 PhD studentships. Simon has editorial roles for two IEEE journals and one Wiley journal and was general chair for Fusion 2010, co-chair for Fusion 2018, and is now President of the International Society of Information Fusion for 2020–2022. Simon was recently named as a “hero” by the IET for his work to support the UK government’s response to COVID-19.

