Institution: Imperial College London

Unit of Assessment: 10 – Mathematical Sciences

Title of case study: B10-3 General-purpose Bayesian inference software has revolutionized data science globally

Period when the underpinning research was undertaken: 2000-2012

Details of staff conducting the underpinning research from the submitting unit:

Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
Nicky Best David Lunn Andrew Thomas	NB: Professor of Statistics and Epidemiology; DL: Research Fellow; AT: Senior Computing Officer	NB: 1996 - 2013 DL: 2000-2007 AT: 2000-2004

Period when the claimed impact occurred: 2013-2020

Is this case study continued from a case study submitted in 2014? Y

1. Summary of the impact

The global proliferation of data science would be unimaginable without access to freely available software that automate the computations required for statistical learning. The first such general-purpose Bayesian software was WinBUGS. This widely used software brought flexible Bayesian data analysis to non-statisticians, stimulated the development of other currently popular Bayesian inference engines, and thereby created significant societal benefits and economic value. In terms of significance and reach, Bayesian inference engines are now used worldwide, making Bayesian data science more accessible to companies, government and NGOs. Monetary revenue is in the millions of USD, especially in the pharmaceutical industry. Wildlife from frogs to elephants benefitted through better conservation management. Hundreds of thousands of UK patients benefitted from improved clinical care based on cost-effectiveness analyses conducted with WinBUGS.

2. Underpinning research

Since the development of Markov Chain Monte Carlo methods in the 1990s, Bayesian methods and modelling have revolutionized statistical data science. Bayesian models are used in a variety of contexts – for example, they are used to forecast the weather, predict the spread of infectious diseases and optimize investment portfolios – however in all these real-world applications, sophisticated computational inference routines are essential. Implementing such algorithms typically requires expert skills at PhD level, and is time consuming. Therefore, the development of software for automated computational inference of Bayesian models has been a global major milestone in making modern statistical analyses broadly accessible. The first such software with global reach, WinBUGS **[1, 2]**, was developed at Imperial College London (Imperial) and the MRC Biostatistics Unit in Cambridge (MRC-BSU). It is freely available online **[1].** The contributions from researchers at both institutions were approximately equal **[2]**. Key collaborators were David Spiegelhalter (MRC-BSU and Statistical Laboratory Cambridge), Andrew Thomas (MRC-BSU after 2004), and Chris Jackson (MRC-BSU, 2007 onwards).

Prior to 2000, the first version of BUGS/WinBUGS was developed at MRC-BSU and could be used to fit Bayesian models described by log-concave full conditional distributions, for example linear regression models. In 1996, senior developers and the project moved from MRC-BSU to Imperial, and subsequent work, largely after 2000, turned early versions of the software into the first general-purpose Bayesian inference engine.

In 2000, the WinBUGS paper **[3]** was published in “Statistics and Computing”, authored by Lunn, Thomas, Best (all Imperial) and Spiegelhalter (MRC-BSU). The paper, which has >5,000 citations, presented the first general-purpose software for fitting complex Bayesian models through an interactive graphical interface, using flowcharts or simple commands to design models. This innovation made advanced statistical modelling and inference accessible to non-statisticians. The paper also showed that the principles of the inference engine were flexible and extensible. This prompted many further extensions that made WinBUGS a broadly applicable data science tool and inspired the development of better inference engines.

Since 2000, major extensions were spearheaded at Imperial, most of which are described in detail in the WinBUGS book **[4]**. Self-tuning Metropolis-Hastings samplers, self-tuning slice samplers, and reversible-jump samplers for fitting variable-dimension models were implemented **[5]**, which greatly broadened applicability to models with full conditional distributions that are neither available in closed form nor log-concave. This allowed, for example, statistical inference with dynamic and non-linear pharmacokinetic/dynamic (PKPD) models **[6]** and led to widespread use of WinBUGS in pharma. The entire software architecture was also overhauled, especially so that sets of correlated random variables could be updated jointly. This enabled the routine inference of geospatial models, disease mapping, biodiversity mapping, and spatio-temporal regression, all of which taken together had a transformative impact on quantitative ecology, conservation biology, and public health.

Industry applications were further supported by the development of specialized interfaces at Imperial. Most notable among these were PKBUGS for fitting PKPD models as routinely used in pharma **[6]**, and GeoBUGS for statistical inference with geospatial models **[7]** as routinely done for public health disease mapping. Building on the basic WinBUGS framework, further additions to the BUGs family of software were made outside of Imperial, such as the open-source version OpenBugs, interfaces to R, and parallel MCMC samplers for faster inference on multi-core computers.

3. References to the research

[1] The BUGS project, https://www.mrc-bsu.cam.ac.uk/software/bugs/
[2] Lunn DJ, Spiegelhalter D, **Thomas A**, **Best N**, *The BUGS project: Evolution, critique and future directions*. Statistics in Medicine, 28, (25) 3049-3067 (2009), doi:10.1002/sim.3680.
[3] Lunn DJ, **Thomas A**, **Best N** and Spiegelhalter D, *WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility*, Statistics and Computing, 10, 325-337 (2000), doi:10.1023/A:1008929526011.
[4] Lunn D, Jackson C, **Best N**, **Thomas A**, Spiegelhalter D, *The BUGS book: A practical introduction to Bayesian analysis*. (2012). https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-the-bugs-book/
[5] Lunn DJ, **Best N** and **Whittaker JC**, *Generic reversible jump MCMC using graphical models*, Statistics and Computing, 19, 395-408 (2009), doi:10.1007/s11222-008-9100-0.
[6] Lunn DJ, **Best N**, **Thomas A**, **Wakefield J** and Spiegelhalter D, *Bayesian Analysis of Population PK/PD Models: General Concepts and Software*, Journal of pharmacokinetics and pharmacodynamics, 29, 271-307 (2002), doi:10.1023/A:1020206907668.
[7] Best N, **Richardson S**, and **Thomson A**, *A comparison of Bayesian spatial models for disease mapping*, Stat Methods Med Res, 14(1), 35-59 (2005), doi:10.1191/0962280205sm388oa.

4. Details of the impact

WinBUGS has had wide-reaching and very significant impact. Here we describe the continuing impact of WinBUGS since August 2013: how WinBUGS underpins the most popular tools for Bayesian data science, enabled the wide-spread dissemination of Bayesian methods, and how this had led to increasing impact to the economy, public health and conservation, in tandem with increasing use of Bayesian methods since 2013.

WinBUGS underpins globally used Bayesian inference software tools. Each day, data scientists across the globe utilize general-purpose software to approximate posterior distributions numerically and perform Bayesian inference. The most widespread software tools during the impact period have been WinBUGS, Stan, JAGS, NIMBLE and OpenBUGS, which are freely available including for commercial purposes. It was WinBUGS that first stimulated the development of this ecosystem of software solutions, which all started their development after WinBUGS (Stan was first released in 2012, NIMBLE in 2015, JAGS in 2007 and OpenBUGS in 2004 **[A])**, and continue to evolve (Stan last updated in 2021, NIMBLE: 2020, JAGS: 2015, OpenBUGS: 2013 **[A]**).

From the Stan Development Team **[B]**:

> "*Stan is based on similar principles as BUGS. […] We could not have done any of this without the groundbreaking work by the BUGS developers, both in terms of technology and community. … WinBUGS was way ahead of everything else in providing (1) a simple language to describe Bayesian models in shareable model files, and (2) technology to derive a generalized […] sampler for automated inference for many Bayesian models."*

From the co-lead developer of NIMBLE **[B]**:

> *"BUGS has inspired many of the software packages that followed over time, including the very popular JAGS and Stan packages, as well as the NIMBLE package […]. It is safe to say that NIMBLE would not exist without the great success of WinBUGS […].*"

From the JAGS user manual **[B]**:

> "*Many thanks to the BUGS development team, without whom JAGS would not exist."*

It is difficult to track how many data science developers and analysists rely on any one of the freely available software solutions that WinBUGS inspired, though for Stan alone the number of users was estimated above 100,000 globally in 2020 **[C]**, and the *rstan*, *nimble*, *rjags*, *R2OpenBUGS* and *R2WinBUGS* packages have been downloaded from R's package distribution network CRAN more than 3.6 million times between 2013 and 2020 **[C]**.

WinBUGS led to the global acceptance of Bayesian data analysis. WinBUGS not only met demand, but also created a data science market. The software helped professionals without a statistics degree to become well-versed in formulating and applying advanced statistical models in their domains, which in turn drove methodological development and better software. Indeed, WinBUGS has been directly linked to the explosive adoption of Bayesian methods across the world, and has become the de-facto standard for data analysis in some industries:

Prof. Lawson, wrote in this book "Bayesian Disease Mapping" in 2018 **[D]**:

> *"the development of the BUGS package and its Windows incarnation WinBUGS have had a huge effect on the dissemination and acceptance of [Bayesian] methods. A brief search for recently published papers referencing WinBUGS turned up applications in food safety, forestry, mental health policy, AIDS clinical trials, population genetics, pharmacokinetics, paediatric neurology and other diverse fields, indicating that Bayesian methods with WinBUGS indeed are finding widespread use*".

Professor McElreath, Director of the Max Planck Institute for Evolutionary Anthropology, wrote in November 2017 **[E]**:

> "*BUGS started a revolution in efficient, desktop Bayesian computation…It has done more than any other initiative to promote and advance applied Bayesian data analysis."*

The Director and Principal Statistician for systematic reviews and meta-analysis at Evidera PPD states in on 26 October 2020 **[F]**:

> *"The majority of our work is conducting Bayesian network meta-analyses (NMAs), generally to support health economic modeling. As you may know, the methodological leadership in the area comes from NICE, and their guidance exemplar codes are in WinBUGS.*"

Impact in the Economy: WinBUGS and the many data science software solutions that it stimulated are providing substantial economic value, estimated to be in millions of US dollars. For example:

The Director and Principal Statistician for systematic reviews and meta-analysis projects at Evidera PPD, writes **[F]**:

> *"WinBUGS and its descendants has greatly facilitated our application of the most advanced statistical methods to the challenging problem of indirect comparisons between treatments [of pharmaceuticals]. Providing such inferences and evidence is mandatory in more and more settings, and the commercial value to Evidera PPD for those projects has been in the millions of dollars, leading to many submissions (and approvals) for our clients."*

The Head of Advanced Methodology and Data Science (AMDS) group, Novartis Global Drug Development, writes **[G]**:

> "*Especially over the last five years, we have streamlined and now routinely use robust and automated statistical inference software in drug development, and WinBUGS played a pivot role in this development. Thus, WinBUGS and related projects have undoubtedly had an enormous impact on the efficiency, cost-effectiveness and economic value of drug development benefiting the pharmaceutical R&D process at Novartis but more importantly benefiting patients across the world*."

Impact in public health: The recommendations of the National Institute for Health and Care Excellence (NICE) aim to optimally allocate resources and maximize the quality of life of people in the UK based on cost-effectiveness and are typically taken up in care **[H]**. WinBUGS or JAGS were used in 26% (n=37) of the clinical guidelines that were published between 2013/08 and 2017/12 and had the full guidelines text available online **[H]**, illustrating how these tools have helped drive effective and affordable healthcare in the UK. For example, the NICE impact mental health report describes the early intervention pathways that were implemented across NHS England as a result of guideline CG178, benefiting approximately 550,000 people diagnosed with severe mental illness in 2017/18 in the UK **[H].** CG178 was published in 2014; WinBUGS was the main tool used for the statistical evaluation of the evidence **[H]**.

Impact in Conservation: The ability of WinBUGS to fit geospatial models, and the same ability of the later software tools that BUGS inspired, have been frequently used in quantitative ecology, and have contributed to conservation management and policy.

One exemplar is the illegal ivory trade, for which JAGS was used annually between 2013-2018 to characterize global trends for the United Nations Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) **[I]**. The estimates underpinned situation reports, informed CITES policy, numerous press releases, and national ivory action plans in Central Africa **[I]**.

There are hundreds of similar examples, for example on wildfire management or the growling grassfrog, an endangered amphibian mainly found on the outskirts of Melbourne, for which national conservation management plans were created based on OpenBUGS. Co-author Professor McCarthy of the University of Melbourne wrote to us **[J]**:

> *"This model is now being used by the Victorian government to plan management actions that are worth tens of millions of dollars, helping to ensure that this money is spent as efficiently as possible."*

5. Sources to corroborate the impact

[A] First release dates of software building on WinBUGS (Archived here)

- Stan: https://github.com/stan-dev/stan/releases?after=v1.0.1
- JAGS: version 4.3.0 user manual (page 5) https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf, https://sourceforge.net/projects/mcmc-jags/files/JAGS/1.0/
- NIBMLE: https://github.com/nimble-dev/nimble/releases

- OpenBUGS: http://openbugs.net/w/Overview

[B] Evidence that software tools build on WinBUGS

- Letter of support from the Stan Development team.
- Letter of support from the NIMBLE Development team.
- JAGS manual https://sourceforge.net/projects/mcmc-jags/files/Manuals/4.x/jags_user_manual.pdf/download; quote is on page 5 (Archived here) .

[C] Evidence that software tools are widely used (Archived here)

- Stan user forum, post https://discourse.mc-stan.org/t/how-many-stan-users-are-there/12196/22, published 2020/12/22. Last accessed 2021/01/25.
- CRAN download statistics, R command sum(cranlogs::cran_downloads(package = c("rstan","rjags","nimble","R2WinBUGS","R2OpenBUGS", "BRugs"), from = "2013-08-01", to = "2020-12-13")$count). Last accessed 2021/01/25.

[D] Lawson AB, Bayesian Disease Mapping, 2018, ISBN-10: 1584888407 p3 (See https://www.amazon.co.uk/Bayesian-Disease-Mapping-Hierarchical-Interdisciplinary/dp/1584888407 - look inside preview (Archived here)).

[E] Online post published 2017/11/08, https://elevanth.org/blog/2017/11/28/build-a-better-markov-chain/. (Archived here)

[F] Letter of support by the Director and Principal Statistician for systematic reviews and meta-analysis projects at Evidera PPD.

[G] Letter of support from the Head of Advanced Methodology and Data Science (AMDS) group, Novartis Global Drug Development.

[H] Impact on NICE evaluations (Archived here):

- Take-up of NICE recommendations in care https://www.nice.org.uk/about/what-we-do/into-practice/measuring-the-uptake-of-nice-guidance, last accessed 2021/01/25;
- NICE Guidelines using WinBUGS, 2013/07 to 2018/12 search results, pdf;
- specific evidence on impact in mental health clinical practice https://www.nice.org.uk/Media/Default/About/what-we-do/Into-practice/measuring-uptake/NICEimpact-mental-health.pdf, last accessed 2021/01/25.
- Use of WinBUGS in CG178 (e.g. 391 WinBUGS cite in methods) https://www.nice.org.uk/guidance/cg178/evidence/full-guideline-pdf-490503565

[I] The original paper using OpenBUGS is doi:10.1371/ journal.pone.0076539, published 2013/10 (Archived here). Situation reports and CITES policy guidance documents are available online at https://www.cites.org/eng/prog/mike/index.php. Press releases are online available at https://www.cites.org/eng/news/pr/index.php. Last accessed 2021/01/25.

[J] Letter of support from the Deputy Director of the Australian Research Council Centre for Excellence for Environmental Decisions.