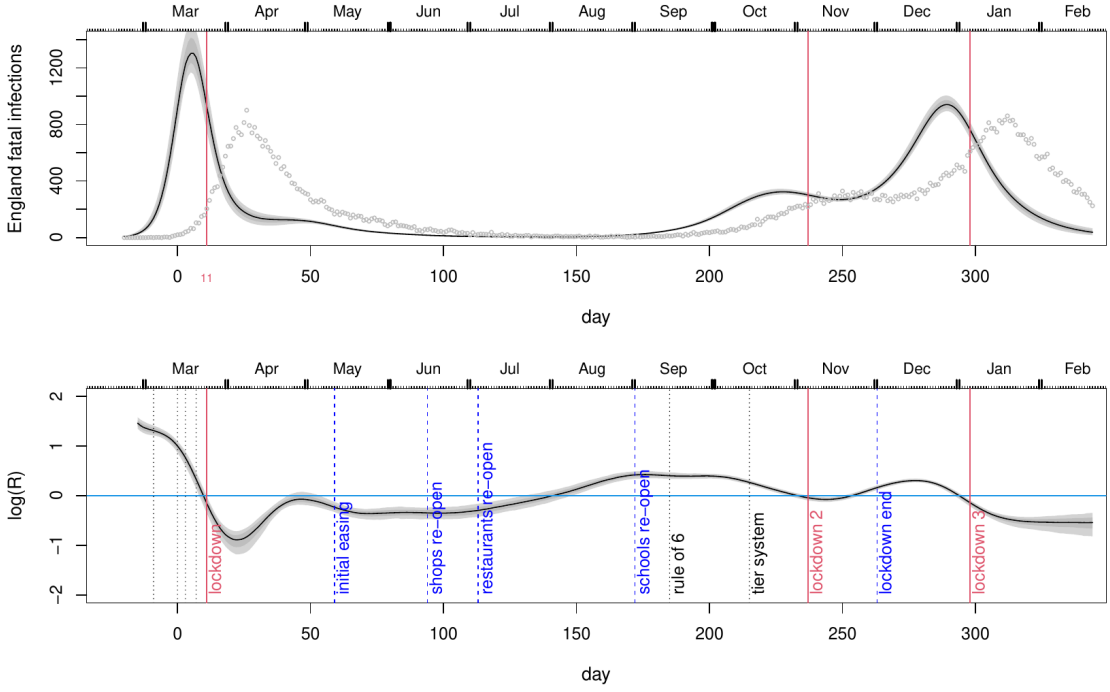
Simon N Wood

***Chair of Computational Statistics*, School of Mathematics, James Clerk Maxwell Building, University of Edinburgh, EH9 3FD**

`simon.wood ed.ac.uk` I'm not on Twitter: it promotes thinking fast in a world that needs to think slow (in the Daniel Kahneman sense), and it's hard to think of a better way to encourage availability bias.

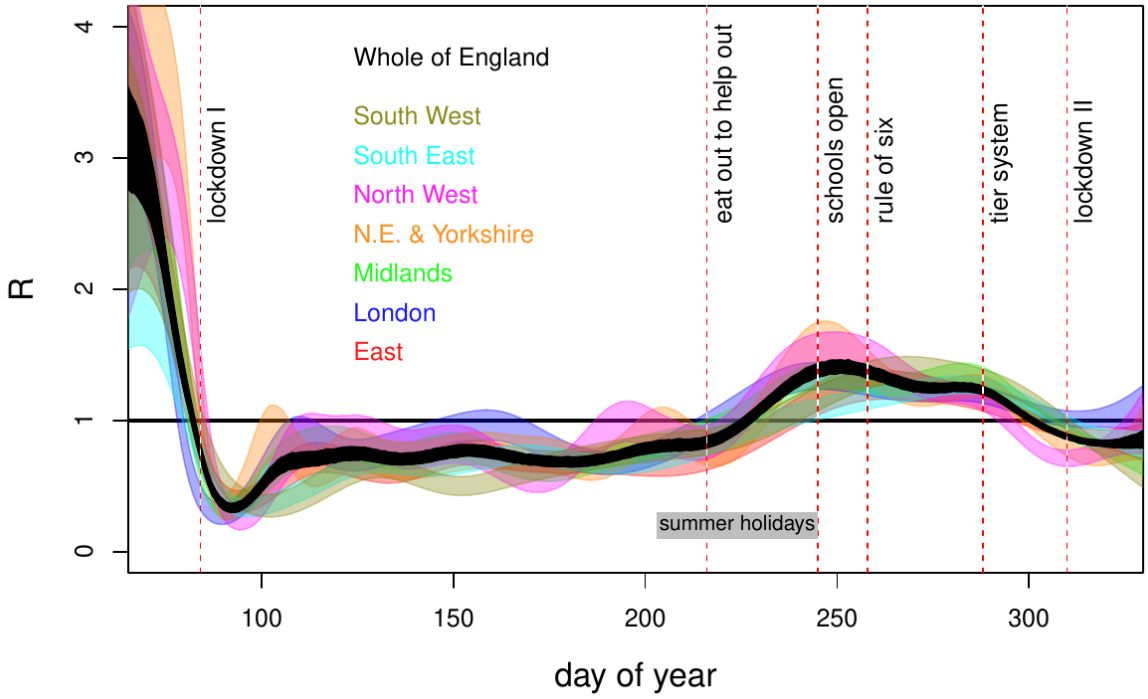
I work as a professor of statistics. Statistics is about the honest interpretation of data. It's not always a popular subject: honest interpretation of data is difficult, and much less appealing than less honest interpretation. Disraeli is often supposed to have said that there are lies, damned lies and statistics. Gladstone was president of the Royal Statistical Society. I tend to think both were right.

Here is the time course of fatal infections from Covid-19 estimated from daily hospital fatalities in England (grey circles), along with the corresponding R estimates (log scale, so R<1 when log(R)<0). The estimated infections are already in decline and R<1 before each lockdown is called, but the timing of lockdown appears to roughly coincide with when the rate of increase of daily deaths reaches its maximum each time. The dip and recovery in R after the first lockdown is expected (unless lockdown did nothing).



The method is described in this paper in press in *Biometrics*, which also shows that the approach of Flaxman et al. (2020, Nature 584, 257-261) gives similar results if we estimate *how and when* R changed, in addition to how much it changed (Flaxman et al. assumed they knew the how and when part). Furthermore Sweden, which did not lockdown, saw declining infections only a day or two after the UK (Swedish GDP dropped about 3% in 2020 against the UK's 10% drop). The paper also briefly reviews the UK statistical evidence on life loss from economic shocks, which suggests that lockdowns may not in fact be net life savers (see e.g. this graphic). Even if life lost to Covid-19 is the only life loss that counts, the cost per life year saved by these non-pharmaceutical-interventions (NPIs) appears to be an order of magnitude higher than the threshold for funding pharmaceutical interventions. See this article.

The results in the above picture also broadly agree with the results obtained from repeating the analysis in the Imperial Covid response team's Report 41, as described in this pre-print, which again replaces some unusual model assumptions in the original study with more defensible ones. The R reconstruction, by English region, and for the whole of England, for most of 2020 then looks like this.



One question is whether our results are somehow caused by smoothness assumptions, rather than the data. The Biometrics paper investigates this and concludes not, but there is also much simpler evidence. The results are rather insensitive to how heavily we smooth, but if we increase the amount of smoothing enough, the R<1 point moves *closer* to lockdown (monotonically). If our results were driven by the smoothness assumptions then the opposite should happen.

So a minimal assumption analysis, as in my Biometics paper, or a replication of the Flaxman et al. approach, but removing the assumption that we know how and when R changed, or a re-doing of Imperial's Report 41 analysis with more defensible assumptions, all imply R<1 and incidence in decline before the lockdowns. This is not proof, but it is strongly suggestive.

Something that the Covid-19 crisis has emphasised is the fact that statisticians have not managed to adequately communicate how fundamental random sampling is to proper measurement of things like infection rates. That data *somehow related* to the thing we want to measure are not the same as data that actually measure it. Here is an attempt to explain random sampling and why it means that you should trust plots like the ones from the ONS Covid survey in preference to plots of cases, like the first one here.

A lockdown reading list

- Collapse (Jared Diamond) on how societies are destroyed, not by external forces, but by their failure to adapt their cultural norms to those forces.
- Thinking Fast and Slow (Daniel Kahneman) on the pitfalls of our intuitive reasoning, especially about risk and uncertainty.
- Mistakes were made, but not by me (Carol Tarvis and Elliot Aronson) on the psychology of sticking with bad decisions.
- The Parable of the Old Man and the Young by Wilfred Owen, on consequences of the above.
- Economics The User's Guide (Ha-Joon Chang) on what you really need to know about economics, and how it is not mainly about money, nor just a scaled up version of household accounting.
- The Great Crash 1929 (Galbraith) a delightful disection of economic hubris (and the need for stabilizing controls that we long since did away with).
- The Rise and Fall of the Third Reich (William Shirer) detailing exactly how things went wrong in Germany after the Great Depression.
- Witch hunting in Scotland (Brian Levack) on the Scottish experience of the great European witchcraft panic (James I/VI wrote a treatise on Witchcraft).
- Wood and Thomas paper on the problems of prediction with disease models in the absense of direct validation data (the least impressive item here).

(from March 2020, but in retrospect p142 of 'Thinking Fast and Slow', on Availability Cascades is quite interesting.)

Notes on debugging R code and C code from R

Apologies if I have not replied to you on an mgcv related query: I've got rather behind on mgcv email, especially on the interesting stuff that requires thought.

Boring career stuff about me: state comprehensive school educated; BSc in Physics from Manchester; PhD on biological modelling (Dept Applied Physics Strathclyde); short stint (1989-90) as a civil service bioeconomic modeller MAFF; 4 years postdoc at Imperial (biology) on biological dynamic models. Lecturer - Reader in statistical ecology, St Andrews (maths), also RSS graduate diploma in statistics; Reader - Prof in statistics (Glasgow, Bath, Bristol).

Books

Core Statistics (2015) is a short textbook in the CUP IMS textbook series. The idea is to offer a concise coverage of the essentials that anyone starting a statistics PhD ought to know, in the form of a brief introduction to statistics for the numerate. A pdf version is here (A5 format - ok for e-reading). Try this version for less wasteful printing on A4. Comments (including typo and error reports) very welcome. Here is the errata list and the algae and urchin datasets. (e.g. alg <- read.table("http://www.maths.bris.ac.uk/~sw15190/data/algae.txt") to read directly into R.). If you find the free download useful please consider buying the book (click top right to change location).

Generalized Additive Models: An Introduction with R (2nd ed) (2017) provides an introduction to linear (mixed) models, generalized linear (mixed) models, generalized additive models and their mixed model extensions. The second edition has a completely revised structure, with greater emphasis on mixed models and the equivalence of smooths and Gaussian random fields. A greatly enhanced range of smoothers is covered, along side a thorough upgrading of the chapter on GAM theory, and many new examples including functional data analysis, survival analysis, location-scale modelling and more.

I work as a professor in the statistics group at the university of Edinburgh. I think Brexit was probably a rather poor decision. I kind of liked being in a block that produces roughly 100% of its own food, as opposed to an island that produces roughly 60%, and it seemed like a sensible thing to be part of a group large enough to stand up to China and the USA. But then I also tend to think of the British Empire as a rather racist enterprise responsible for the death of millions and the subjugation of far more, rather than, say, an indication of the country's former greatness. I'm currently joint editor of JRSSB and have two main research interests.

- **Smoothing.** In particular methods for generalized additive modelling and applications of generalized additive models (GAMs). I am especially interested in smoothness selection, and low rank spline smoothing, and have written an R package called mgcv which implements GAMs. Some recent example smoothing papers are
  - Wood, SN, N Pya and B Saefken (2016) *Smoothing parameter and model selection for general smooth models* (with discussion). Journal of the American Statistical Association.
  - Wood, SN, Z Li, G Shaddick and NH Augustin (2017) *Generalized additive models for gigadata: modelling the UK black smoke network daily data* Journal of the American Statistical Association. Here are the data set black_smoke.RData and its description.
  - Wood, SN and M Fasiolo (2017) *A Generalized Fellner-Schall Method for Smoothing Parameter Optimization with Application to Tweedie Location, Scale and Shape Models* Biometrics.
  - Wood, SN (2019) *Simplified Integrated Nested Laplace Approximation* in press Biometrika.
  - Wood, SN (2020) *Inference and computation with generalized additive models and their extensions (with discussion from Greven, Scheipl, Kneib and Eilers)* TEST
- **Statistical Ecology.** In particular using ecological dynamic models as statistical models to help understand ecological mechanisms, and ecological applications of nonlinear random effects models and smooth models, as part of NCSE. Some recent example statistical ecology papers are
  - Wood SN (2010) *Statistical inference for noisy nonlinear ecological dynamic systems.* Nature 466(26): 1102-1104. (or short web version with links to supplementary material and methods).
  - Augustin, NH, V Trenkel, SN Wood and P Lorance (2013) *Space-time modelling of blue ling for fisheries stock management* Environmetrics 24(2), 109-119.
  - M Fasiolo, N Pya and SN Wood (2016) *A Comparison of Inferential Methods for Highly Nonlinear State Space Models in Ecology and Epidemiology* Statistical Science 31 (1), 96-118

Fuller lists of papers are at researcherid and google scholar . Here is a 2014 BIRS talk on inference for ecological dynamic models.

I am interested in taking on PhD students working on any area related to my research interests. Here are a couple of example projects: GAMs for big data and GAMs for multivariate data. The department has funding for strong students.

Advisees:

- Emiko Dupont, works on spatial modelling.
- Bertrand Nortier, works on quantile GAM methodology.

Here is a selection of talks. It's not exhaustive, but hopefully gives some idea of what I work on.

- Simple statistical methods for complex ecological dynamics (ISEC 2014)
- Smooth modelling with regular likelihoods
- Modelling with smooth functions
- REML estimation of GAMs
- GAMs and mgcv
- An introduction to GAMs
- Implementing mgcv in R
- Soap film smoothing how to smooth over domains were the boundary matters.
- Inaugural lecture with slides. [12/3/2009. A university of Bath podcast]
- Slides from a couple of short GAM course .

This year I'm teaching Theory of Inference. Here are a couple of examples of previous courses.

- A first course in statistics
- A short course on likelihood based inference (with R).

- mgcv - generalized additive (mixed) models and other generalized ridge regression for R.
- Software
- C.V. (not up to date).
- Bristol maps