Evaluating an epidemiologically motivated surrogate model of a multi-model ensemble

IISA 2022 - Pandemic forecasting: Lessons learnt from COVID-19

Sam Abbott
fosstodon.org/@seabbs
samabbott.co.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES
CSH Cold Spring Harbor Laboratory BMJ Yale

Evaluating an epidemiologically motivated surrogate model of a multi-model ensemble

Sam Abbott, Katharine Sherratt, Nikos Bosse, Hugo Gruson, Johannes Bracher, Sebastian Funk

doi: https://doi.org/10.1101/2022.10.12.22280917

DOI: 10.1101/2022.10.12.22280917

Evaluating an epidemiologically motivated surrogate model of a multi-model ensemble

Multi-model and multi-team ensemble forecasts have become widely used to generate reliable short-term forecasts. Whilst there is robust and consistent evidence that multi-team ensemble forecasts provide reliable and performant forecasts across domains they also have a range of downsides. The most significant is the difficulty in interpreting them. Here we develop a parsimonious forecast model based on observations of ensemble behaviour and evaluate its predictive performance.

AUTHOR
Sam Abbott

PUBLISHED
Oct. 13, 2022

CITATION
Abbott, 2022

samabbott.co.uk/posts/2022-10-13-surrogate-ensemble-forecast/

Talk overview

Who am I

What are we talking about?

The surrogate model

Implementation

Results

Summary

Slides: samabbott.co.uk/presentations/2023/iisa-surrogate-ecdc-ensemble-model.pdf

Evaluating an epidemiologically motivated surrogate model of a multi-model ensemble ensemble

IISA 2022 - Pandemic forecasting: Lessons learnt from COVID-19

Who am I?

Sam Abbott
fosstodon.org/@seabbs
samabbott.co.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Sam Abbott

Twitter | Posts | Notes | Research | Ideas | Code | Papers

Hi there 👋

I'm an infectious disease researcher interested in real-time analysis, forecasting, semi-mechanistic modelling, and open source tool development. More on my research interests here.

NOW

- Working at the London School of Hygiene and Tropical Medicine in the Epiforecasts group;
- ⭐ Crafting extensions to epinowcast 📦;
- 🌓 Currently working on:
  - Evaluating a simplified forecast model in comparison to the ECDC forecasting hub ensemble
  - Evaluating the use of real-time sequences for short-term forecasting
  - Evaluating a new method for nowcasting right truncated count data.

BIO

- 🏢 I'm currently working at London School of Hygiene and Tropical Medicine
- 🎓 I did my PhD at the University of Bristol
- ⚙️ I use daily: `R`, `stan`
- 🏗️ I like to perform analysis using novel models on interesting data and generalise those approaches into software 📦s
- 🌍 I'm mostly active within the R Community
- 🌱 (Very slowly) Learning all about Julia and Turing.jl
- 📘 Reading: Modeling Discrete Time-to-Event Data
- 💬 Ping me about statistical modelling of infectious diseases, real-time analysis of infectious diseases, estimating transmission dynamics in real-time, and team science opportunities
- 📫 Reach me: sam.abbott@lshtm.ac.uk

samabbott.co.uk

Who am I?

Submitted short-term forecasts, reproduction number estimates, and a range of real-time analyses since April 2020 to SAGE/SPIM-O (UK advisory committee).

Produced and submitted (i.e to the US and European Forecasting Hubs) similar forecasts and estimates for most countries.

These estimates were based on open source tools I developed and these are now commonly used by academics and in public health departments for real-time surveillance of infectious disease.

Currently working on the next generation of methods/tools (see epinowcast.org)

Evaluating an epidemiologically motivated surrogate model of a multi-model ensemble ensemble

IISA 2022 - Pandemic forecasting: Lessons learnt from COVID-19

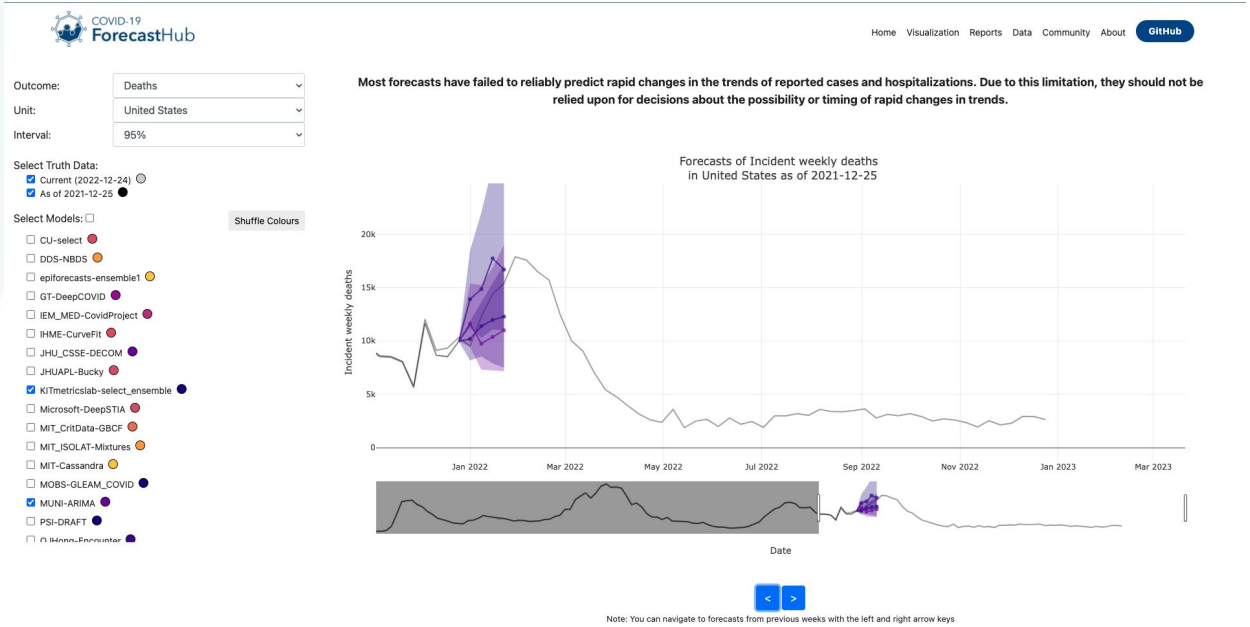
What are we talking about?

Sam Abbott
fosstodon.org/@seabbs
samabbott.co.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

What are we talking about?

- Multi-model, multi-team, ensembles have become increasingly popular as an approach to increase the robustness and performance of infectious disease forecasts. covid19forecasthub.org



What are we talking about?

- Multi-model, multi-team, ensembles have become increasingly popular as an approach to increase the robustness and performance of infectious disease forecasts. covid19forecasthub.eu

What are we talking about?

Ensemble forecasts from these projects have a range of downsides including:

- The considerable organisational and resource cost required to produce them

- Difficulty in interpreting the assumptions behind a given forecast

Emulation approaches are widely used in climate modelling as a solution to high compute requirements but are generally non-parametric. Can we do something similar but without losing interpretability?

What are we talking about? The surrogate model

1. We want to understand our models behaviour.
2. We want “good” forecast behaviour
3. We want forecasting to be relatively “cheap”

Can we create a model with these features?

What are we talking about? How

- Review the Hub ensemble and derive a set of observed behaviours.
- Review our previous forecast efforts and those of others to explore the kinds of assumptions we commonly make in our models.
- Make a set of assumptions we think will produce the behaviour we have observed.
- Define a model based on these assumptions
- Evaluate in real-time for 6 months in comparison to the Hub ensemble.

Evaluating an epidemiologically motivated surrogate model of a multi-model ensemble ensemble

IISA 2022 - Pandemic forecasting: Lessons learnt from COVID-19

The surrogate model

Sam Abbott
fosstodon.org/@seabbs
samabbott.co.uk

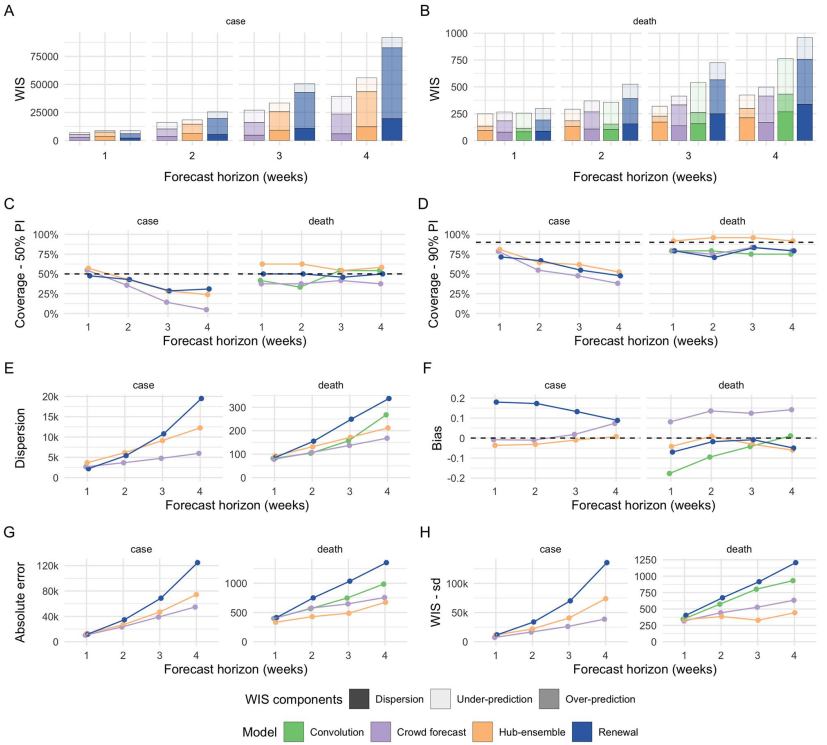
LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Insights from the European Forecast Hub

- Ensemble is an unweighted quantile median of independent forecasts
- Forecasts come from a range of models based on mechanistic, statistical, hybrid, and human judgement frameworks.
- Typically robust to daily reporting artefacts
- Some ability to forecast future changes in trend that were not clearly present in the observed data
- Highly auto-correlated over time and generally less reactive during periods of stability.
- Relatively sharp forecasts with a noticeable bias away from forecasts significantly larger than currently observed cases.

Insights from our submitted models

- Focussed on trying to minimally represent the infection process whilst retaining statistical flexibility
- Negative Binomial likelihood with a log link.
- Used daily data with adjustment for day of the week effects.
- Adjustment for changes in the susceptible population.
- Generally captured changes in trend quite well but were very susceptible to outliers, and typically overpredicted during periods of growth.



Citation: Bosse NI, Abbott S, Bracher J, Hain H, Quilty BJ, Jit M, et al. (2022) Comparing human and model-based forecasts of COVID-19 in Germany and Poland. PLoS Comput Biol 18(9): e1010405. https://doi.org/10.1371/journal.pcbi.1010405 DOI: 10.1371/journal.pcbi.1010405

Model assumptions

Table 1. Observations on the relative performance of the Forecast Hub ensemble compared to our forecast submissions.

No.	Observation
1	Robust to daily reporting artefacts
2	Some ability to forecast future trend changes
3	Less reactive to apparent observed changes in trend
4	Sharper forecasts
5	A tendency towards underprediction
6	Modelling the reporting process appears to have little impact

Table 2. Assumptions/simplifications based on observations of the relative performance of Forecast Hub ensembles compared to our forecast submissions.

Assumption	Observation
Reported cases can be modelled using weekly data and a generative process discretised by week	1, and 2
Reported cases can be modelled as if they represented infections	6
The growth rate of infections can be represented as an auto-regressive process with an order of 1 week	3 and 4
Unobserved interventions and more general changes in transmission towards a stable state can be represented using a multiplicative decay parameter	2, and 5

Model definition

A weekly AR(1) process on the growth rate with additional decay to a steady state.

$$\lambda_0 \sim \text{LogNormal}\left(\log C_0, 0.025 \times \log C_0\right)$$
$$\lambda_t = C_{t-1} e^{r_t},\ t > 0$$
$$C_t \mid \lambda_t \sim \text{NB}\left(\lambda_t, \phi\right)$$

where the mean and variance of the negative binomial are given by

$$\mathbb{E}[C_t \mid \lambda_t] = \lambda_t \quad \text{and} \quad \text{Var}[C_t \mid \lambda_t] = \lambda_t + \frac{\lambda_t^2}{\phi}.$$

$$r_0 \sim \text{Normal}\left(0, 0.25\right)$$
$$r_t = \left(\mathbf{1}_{r_{t-1}>0}\xi_+ + \mathbf{1}_{r_{t-1}\leq 0}\xi_-\right) r_{t-1} + \varepsilon_t$$
$$\varepsilon_t = \mathbf{1}_{t>0}\beta\,\varepsilon_{t-1} + \eta_t$$

Evaluating an epidemiologically motivated surrogate model of a multi-model ensemble ensemble

IISA 2022 - Pandemic forecasting: Lessons learnt from COVID-19

Implementation

Sam Abbott
fosstodon.org/@seabbs
samabbott.co.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Implementation

What did we do?

- Submitted to the European Forecast Hub from February 2022
- Model written using stan and as an extension of the forecast.vocs package
- Model deployed using GitHub Actions during development to ensure computational resources are appropriately restricted.
- Model submitted in an automated fashion using the ECDC hub GitHub Actions framework with no manual interventions during the submission period.
- Evaluation fully reproducible ensuring that the complete analysis can be replicated by others if interested in developing a model with similar resource constraints.

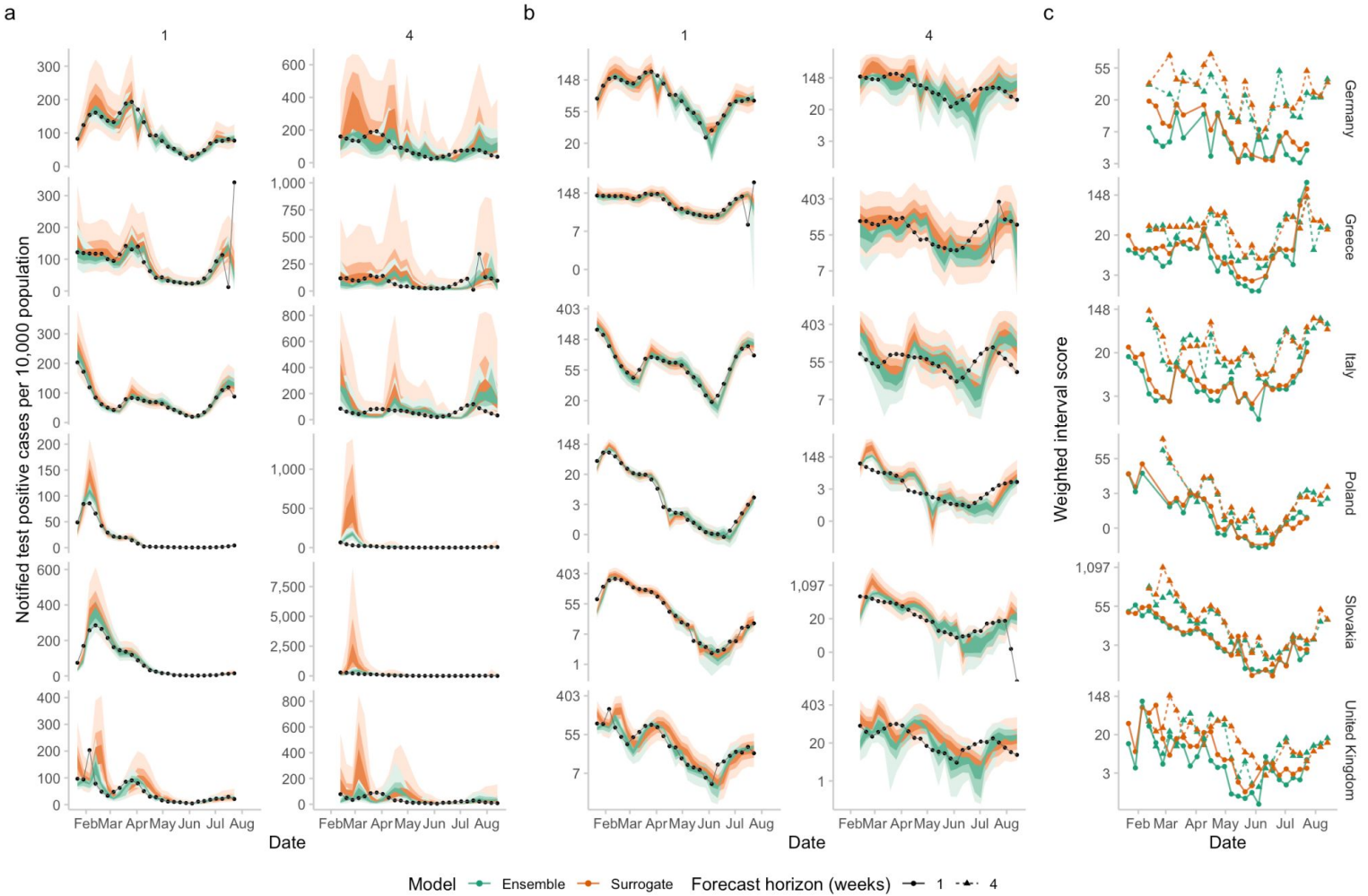
Evaluating an epidemiologically motivated surrogate model of a multi-model ensemble ensemble

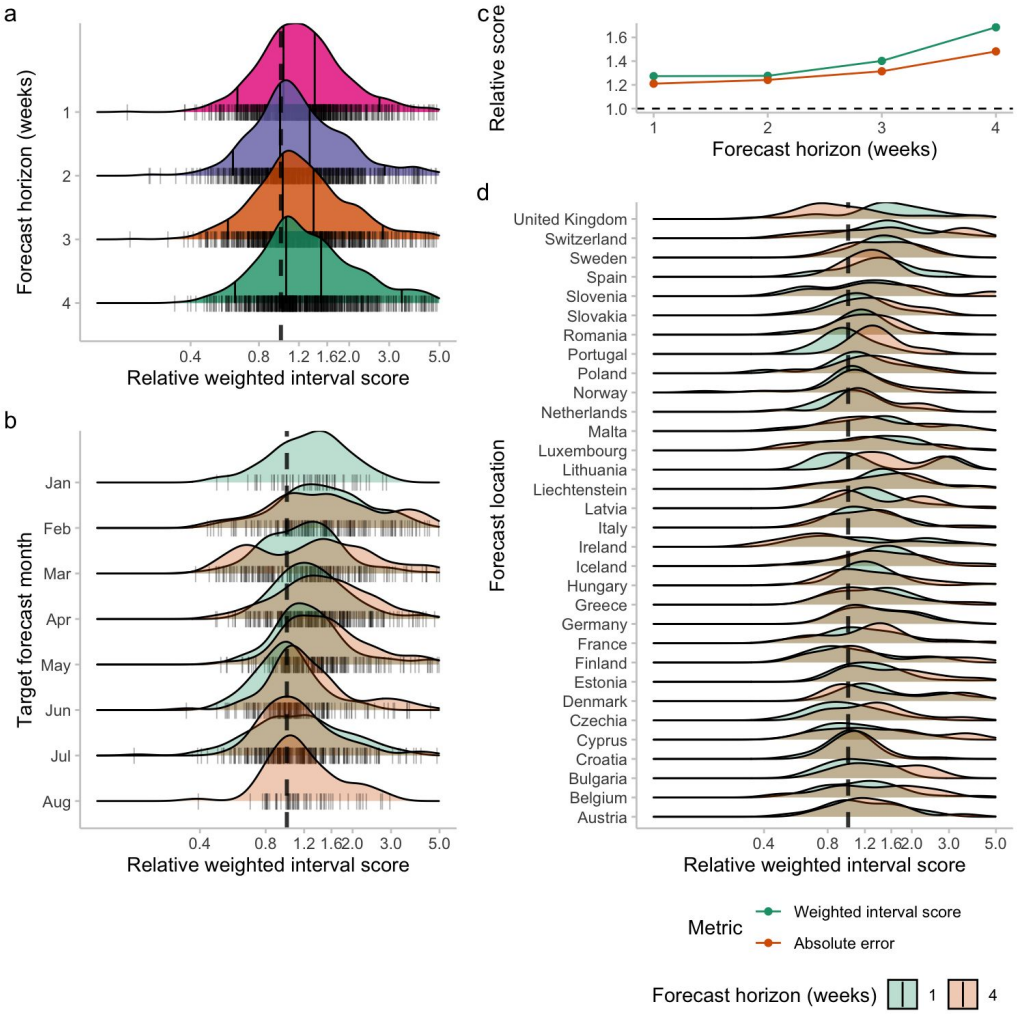
IISA 2022 - Pandemic forecasting: Lessons learnt from COVID-19

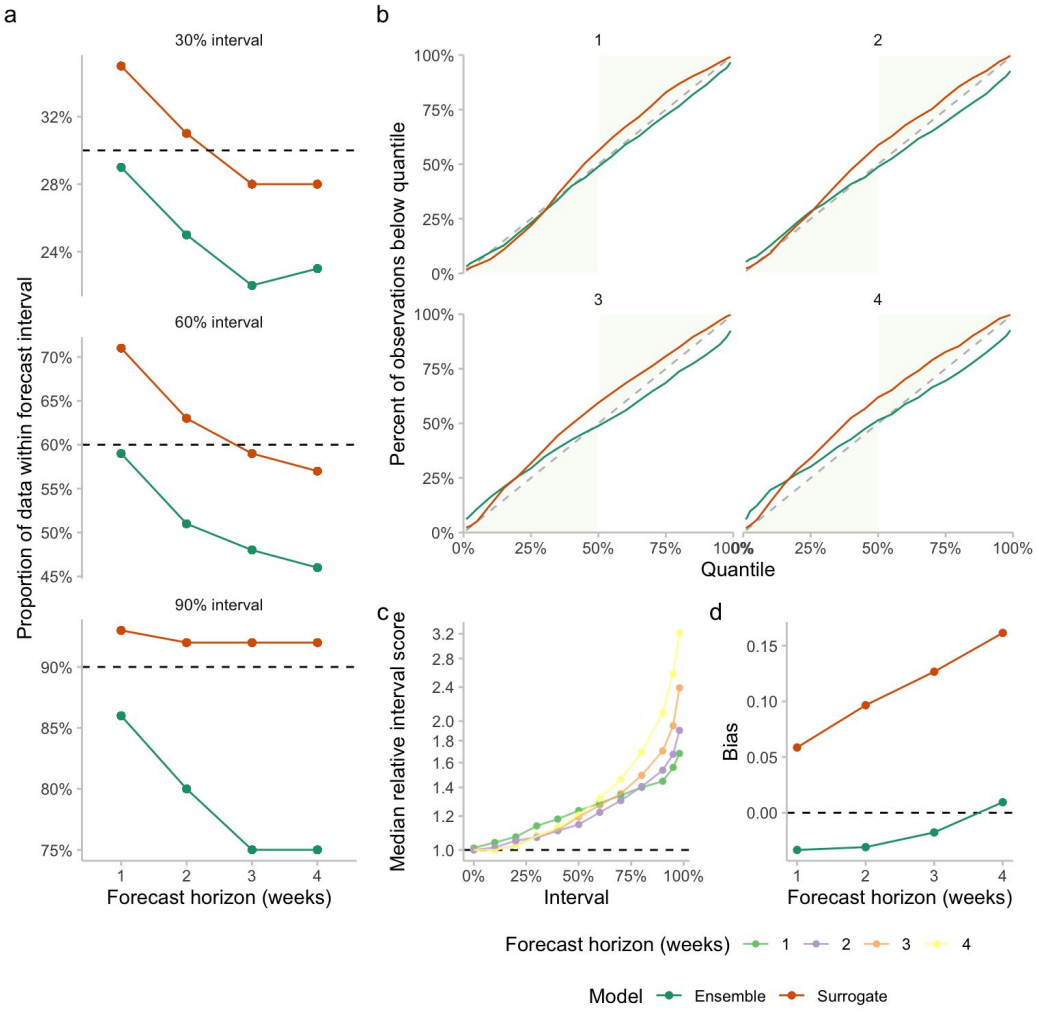
Results

Sam Abbott
fosstodon.org/@seabbs
samabbott.co.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE


a
b
c
1
4
Notified test positive cases per 10,000 population
Weighted interval score
Germany
Greece
Italy
Poland
Slovakia
United Kingdom
Feb Mar Apr May Jun Jul Aug
Date
Model
Ensemble
Surrogate
Forecast horizon (weeks)
1
4


a
Forecast horizon (weeks)
1
2
3
4
0.4
0.8
1.2
1.6
2.0
3.0
5.0
Relative weighted interval score
b
Target forecast month
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
0.4
0.8
1.2
1.6
2.0
3.0
5.0
Relative weighted interval score
c
Relative score
1.6
1.4
1.2
1.0
1
2
3
4
Forecast horizon (weeks)
d
Forecast location
United Kingdom
Switzerland
Sweden
Spain
Slovenia
Slovakia
Romania
Portugal
Poland
Norway
Netherlands
Malta
Luxembourg
Lithuania
Liechtenstein
Latvia
Italy
Ireland
Iceland
Hungary
Greece
Germany
France
Finland
Estonia
Denmark
Czechia
Cyprus
Croatia
Bulgaria
Belgium
Austria
0.4
0.8
1.2
1.6
2.0
3.0
5.0
Relative weighted interval score
Metric
Weighted interval score
Absolute error
Forecast horizon (weeks)
1
4


a
30% interval
60% interval
90% interval
Proportion of data within forecast interval
Forecast horizon (weeks)
b
Percent of observations below quantile
Quantile
c
Median relative interval score
Interval
d
Bias
Forecast horizon (weeks)
Forecast horizon (weeks) 1 2 3 4
Model Ensemble Surrogate

Evaluating an epidemiologically motivated surrogate model of a multi-model ensemble ensemble

IISA 2022 - Pandemic forecasting: Lessons learnt from COVID-19

Summary

Sam Abbott
fosstodon.org/@seabbs
samabbott.co.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Summary

So what are we talking about? Improving access to ensemble quality forecasts whilst also trying to understand how they work

What did we try and do. Develop an interpretable surrogate model based on our observations of the ensembles behaviour and our prior forecasting experience.

The surrogate model. A weekly AR(1) process on the growth rate with additional decay to a steady state.

Implementation. Fully automated and cloud based submission and a reproducible analysis framework.

Results. Some qualitative similarity but with increased uncertainty and generally quantitatively worse performance. Improved calibration vs the ensemble.

What is left to do? How can we do better? Should be trying to replicate ensemble based behaviour at all. Is it what we want?

Thanks to my co-authors for doing all the work, our reviewers, all those who submitted forecasts which informed the ensemble that this is built on, and to you all for listening to this talk.

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES
CSH Cold Spring Harbor Laboratory BMJ Yale

Evaluating an epidemiologically motivated surrogate model of a multi-model ensemble

Sam Abbott, Katharine Sherratt, Nikos Bosse, Hugo Gruson, Johannes Bracher, Sebastian Funk

doi: https://doi.org/10.1101/2022.10.12.22280917

DOI: 10.1101/2022.10.12.22280917