Evaluating a simplified forecast model in comparison to the ECDC forecasting hub ensemble

Epiforecasts meeting (30th June 2022)

Sam Abbott
@seabbs
samabbott.co.uk

cmmid
centre for the mathematical modelling of infectious diseases

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

An 80% right paper before a policy decision is made is worth ten 95% right papers afterwards, provided the methodological limitations imposed by doing it fast are made clear.

Whitty et al. BMC Medicine, 2015, DOI: 10.1186/s12916-015-0544-8

Talk overview

So what are we talking about?

Model

Implementation

Results

Summary

Slides: samabbott.co.uk/presentations/2022/epiforecasts-surrogate-ecdc-ensemble-model.pdf

Evaluating a simplified forecast model in comparison to the ECDC forecasting hub ensemble

Epiforecasts meeting (30th June 2022)

So what are we talking about?

Sam Abbott
@seabbs
samabbott.co.uk

centre for the mathematical modelling of infectious diseases
LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

So 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.

- Ensemble forecasts from these projects have a range of downsides including the considerable organisational and resource cost required to produce them.

- Emulation approaches are widely used in climate models as a solution to high compute requirements.

- We also want to understand our models.

- Here we discuss the use of surrogate model which tries to replicate observed model behaviour based on a series of assumptions around how the target ensemble model performs.

- Here we motivate the assumptions behind this model, discuss its implementation, and evaluate its performance against the ECDC forecast ensemble.

- We aim for this work to provide a low resource option for forecasters, and provide inspiration for forecasters looking to improve.

Evaluating a simplified forecast model in comparison to the ECDC forecasting hub ensemble

Epiforecasts meeting (30th June 2022)

Model

Sam Abbott
@seabbs
samabbott.co.uk

centre *for the* mathematical modelling *of* infectious diseases

LONDON SCHOOL *of* HYGIENE &TROPICAL MEDICINE

Insights from submitted models

What can we learn from other models

- Ensemble is an unweighted quantile median of independent forecasts
- These forecasts come from a range of models based on mechanistic, statistical, hybrid, and human judgement frameworks.
- Most models use daily data with some adjustment for day of the week effects and outlier observations.
- Our submissions have focussed on trying to minimally represent the infection process whilst retaining statistical flexibility.
- Our submissions have generally captured changes in trend quite well but were very susceptible to outliers, and typically overpredicted during periods of growth.
- Very few models tried to capture NPIs

Insights from the ECDC ensemble

What can we learn from the combined ensemble forecast

- Ensemble is an unweighted quantile median of independent forecasts
- 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.

Simplified model assumptions

What assumptions should we make

- Simplified epidemiological process that still assumes exponential growth but focuses only on observations.

- Weekly data.

- The growth rate should be non-stationary over time.

- The growth rate should be modelled to decay towards zero over the forecast horizon in order to represent external changes in transmission.

Model definition

We model the expectation (λ_t) of reported cases (C_t) as an order 1 autoregressive (AR(P)) process by epidemiological week (t). The model is initialised by assuming that the initial reported cases are representative with a small amount of error (2.5%). We assume a negative binomial observation model with overdispersion ϕ for reported cases (C_t).

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

Where r_t can be interpreted as the growth rate. r_t is then modelled as a piecewise constant differenced AR(1) process modified such that the dependence of r_{t-1} is multiplied by a decay factor ($\xi_{+,-}$) that varies dynamically according to the sign of r_{t-1}. The assumptions of this modelling approach are that the growth rate is non-stationary with a trend that is independent of the current growth rate (the differenced AR(1) process), the additional decay factor encodes the belief that larger absolute growth rates will tend more quickly towards no growth and that this process may work differently for positive or negative growth rates.

Model definition

This process can be defined as follows,

$$
\begin{aligned}
r_0 &\sim \text{Normal}(0, 0.25) \\
r_t &= \left(\mathscr{H}(r_{t-1} > 0)\xi_+ + \mathscr{H}(r_{t-1} \leq 0)\xi_-\right) r_{t-1} + \varepsilon_t \\
\varepsilon_t &= \mathscr{H}(t > 0)\beta\varepsilon_{t-1} + \eta_t
\end{aligned}
$$

Where $\mathscr{H}$ is the Heaviside step function and is defined such that it attains the value of 1 if the argument is true and the value of 0 otherwise. The following priors are used,

$$
\begin{aligned}
\xi_+ &\sim \text{Beta}(3, 1) \\
\xi_- &\sim \text{Beta}(3, 1) \\
\beta &\sim \text{Normal}(0, 0.25) \\
\eta_t &\sim \text{Half-Normal}(0, \sigma) \\
\sigma &\sim \text{Normal}(0, 0.2) \\
\frac{1}{\sqrt{\phi}} &\sim \text{Half-Normal}(0, 1)
\end{aligned}
$$

Evaluating a simplified forecast model in comparison to the ECDC forecasting hub ensemble

Epiforecasts meeting (30th June 2022)

Implementation

Sam Abbott
@seabbs
samabbott.co.uk

centre *for the* mathematical modelling *of* infectious diseases

LONDON SCHOOL *of* HYGIENE & TROPICAL MEDICINE

Implementation

What did we do?

- Submitted to the ECDC 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 a simplified forecast model in comparison to the ECDC forecasting hub ensemble

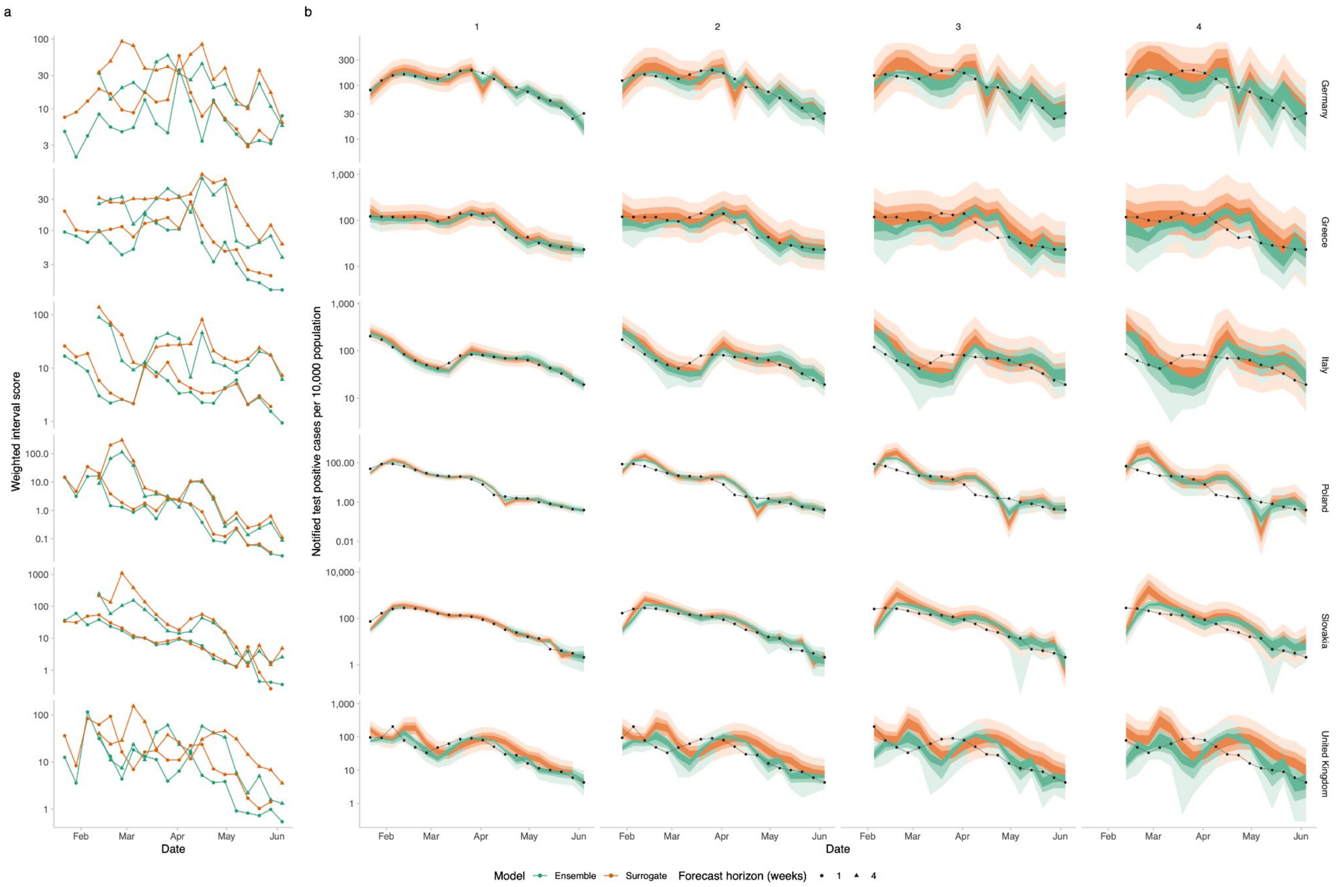
Epiforecasts meeting (30th June 2022)

Results

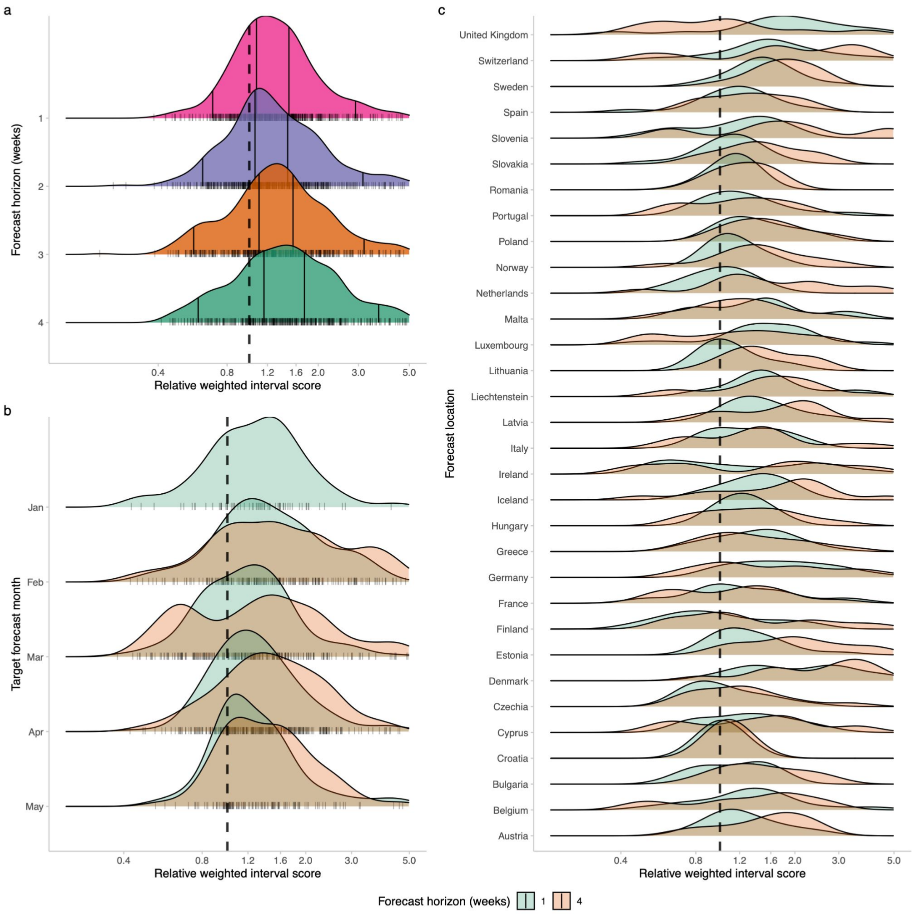
Sam Abbott
@seabbs
samabbott.co.uk

cmmid centre for the mathematical modelling of infectious diseases
LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

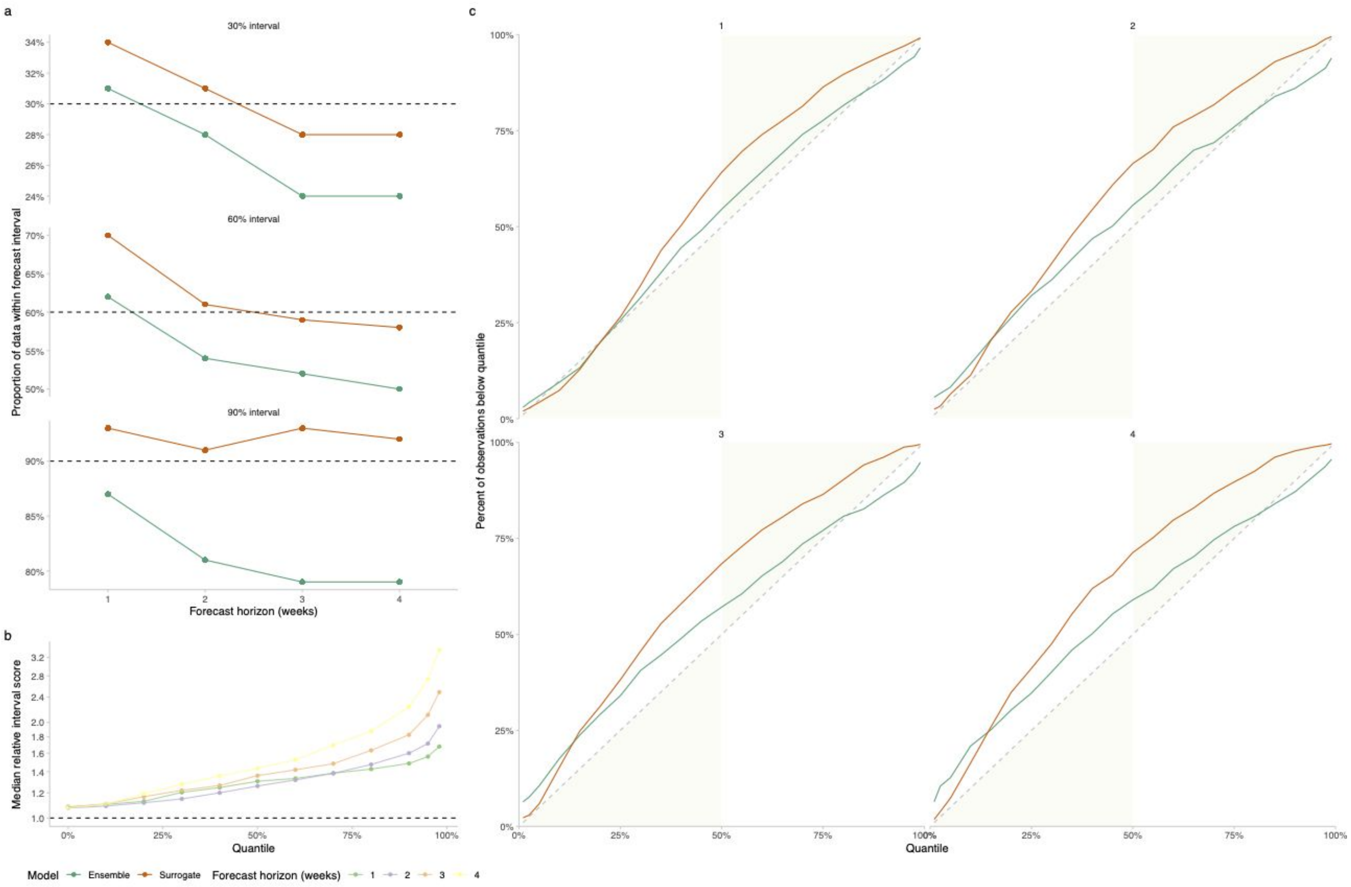
Forecasts in example locations



Relative performance of the surrogate model



Relative coverage of the surrogate model



Evaluating a simplified forecast model in comparison to the ECDC forecasting hub ensemble

Epiforecasts meeting (30th June 2022)

Summary

Sam Abbott
@seabbs
samabbott.co.uk

cmmid centre for the mathematical modelling of infectious diseases
LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

Summary

So what are we talking about?

Model

Implementation

Results

What is left to do?

Slides: samabbott.co.uk/presentations/2022/epiforecasts-surrogate-ecdc-ensemble-model.pdf

Thanks to the epiforecast.io group and my collaborators.