

# Efficient real-time monitoring of an emerging influenza epidemic: how feasible?

## Web Appendix

Birrell PJ, Wernisch L, Tom BDM, Held L, Roberts GO,  
Pebody RP, De Angelis D.

### Appendix A Model Parameterisation

Table A1 presents the model parameters together with the values used in the simulation of the data in the two study scenarios.

Further detail on the parameterisation of the transmission model component can be found in the Supplementary Information to Birrell et al. (2011). Here, some further detail is given on the parameterisation chosen here for the background consultation model in Scenario 2.

#### A.1 The Background Consultation Model

The model for the rates of background primary care consultations used in both the generation of data in Section 5.2 and in the model fitted in Sections 6.2 and 8 is piecewise linear, with a step change between  $t_k = 83$  and  $t_k = 84$  days, reflecting the introduction of a public health intervention. Formally, the log-consultation rate is (at time  $t$ , in age group  $a$ , is:

$$\log B_{ta} = \begin{cases} \mu^0 + \beta_{A(a)}^0 + s^0(t) & t < 84 \\ \mu^1 + \beta_{A(a)}^1 + s^1(t) & t \geq 84 \end{cases} \quad (1)$$

where  $A(a) = \mathbb{1}_{\{a > 15\}}$  is an indicator function and the  $s^i(t)$ ,  $i = 0, 1$  are linear B-splines with general form

$$s^i(t) = \sum_{j=1}^{J^i} \alpha_j^i \Lambda_j^i(t)$$

where  $\Lambda_j^i(t)$  is the triangular function

$$\Lambda_j^i(t) = \begin{cases} (t - t_{j-1}^i)/(t_j^i - t_{j-1}^i) & t_{j-1}^i \leq t \leq t_j^i \\ (t_{j+1}^i - t)/(t_{j+1}^i - t_j^i) & t_j^i < t < t_{j+1}^i \end{cases}, \quad j = 1, \dots, J_i; \quad i = 0, 1.$$

Where  $t_j^i$  is the  $j^{\text{th}}$  knot of B-spline  $i$ .

For identifiability it is imposed that  $\sum_{a=0}^1 \beta_a = \sum_{j=1}^{J^i} \alpha_j^i = 0$ . The two splines have vectors of knot points  $\mathbf{t}^0 = (1, 42, 83)$  and  $\mathbf{t}^1 = (84, 129, 177, 245)$ . Overall, this gives a total of 9 free parameters governing the background model of (1). The prior for these nine parameters is multivariate normal with a block diagonal matrix, designed such that the  $B_{t_j^i a}$  are all, *a priori*,

Table A1:  
Parameters used in the simulation of (confirmed case) epidemic data

<i>Parameter</i>	<i>Description (dimension)</i>	<i>Value/Prior</i>
$\eta$	Dispersion parameters, split either side of a public health intervention at $t_k = 83$ , denoted $(\eta_1, \eta_2)$ (2).	(3.00, 2.15)
$d_I$	The mean infectious period. (1)	3.47
$\phi$	The proportion of symptomatic infections. (1)	0.278
$m$	Multipliers applied to the contact matrices (e.g. to describe the school-holiday effects). (5)	(0.403, 0.495, 0.0588, 0.301, 0.421)
$\psi$	Exponential growth rate. (1)	0.133
$\nu$	A reparameterisation of the initial number of infectives, a function of $I_0$ . (1)	-13.9
$\mathbf{p}^{\text{conf}}$ or $\mathbf{p}^{\text{doc}}$	Parameters governing the population propensity of individuals with ILI symptoms to appear in the data. Split either side of the public health intervention at $t_k = 83$ , with different rates for adults and children. (4)	$p_{t_k, a}^e = \begin{cases} p_1 & t_k \leq 83, a < 4 \\ p_2 & t_k \leq 83, a \geq 4 \\ p_3 p_1 & t_k > 83, a < 4 \\ p_4 p_2 & t_k > 83, a \geq 4 \end{cases}$ $p = (0.278, 0.162, 0.137, 0.441)$
$\beta^B$	Parameters of the piecewise log-linear model for the background ILI consultation rates, with change-points at $t_k = 83, 129, 177$ .	See Web Appendix B.

identically distributed. This is done by considering an unconstrained (and hence unidentifiable) version of (1), evaluated below at the knot points:

$$\log B_{t_j^i a} = m^i + b_{A(a)}^i + a_j^i; \quad j = 1, \dots, J_i; \quad i = 0, 1.$$

Assume that the vectorised parameter  $(m^0, \mathbf{b}^0, \mathbf{a}^0, m^1, \mathbf{b}^1, \mathbf{a}^1) \sim N(0, \mathbf{D})$  for some diagonal covariance matrix  $\mathbf{D}$  with diagonal elements corresponding to prior variances of the  $m^i$  being equal, and similarly for the  $b_a^i$  variances and the  $a_j^i$  variances. Consider a linear transformation

$$\begin{aligned} \mu^i &= m^i + \bar{b}^i + \bar{a}^i \\ \beta_a^i &= b_a^i - \bar{b}^i \\ \alpha_j^i &= a_j^i - \bar{a}^i. \end{aligned}$$

Then if this transformation is represented by a matrix  $\mathbf{E}$ , then the vector  $(\mu^0, \beta^0, \alpha^0, \mu^1, \beta^1, \alpha^1)$  has distribution  $N(0, \mathbf{EDE}^T)$ . The matrix  $\mathbf{D}' = \mathbf{EDE}^T$  is block-diagonal and degenerate. Rows are removed corresponding to parameters  $\beta_1^0, \alpha_{j_0}^0, \beta_1^1, \alpha_{j_1}^1$  and setting these to:

$$\begin{aligned} \beta_1^i &= -\beta_0^i \\ \alpha_{j_i}^i &= -\sum_{j=1}^{j_i-1} \alpha_j^i \end{aligned}$$

for  $i = 0, 1$ .

Table B2: KL statistics and likelihood evaluations per day (‘Run Time’) for each resample-move algorithm used to analyse the ‘contaminated’ primary care consultation and serology dataset over each of the time periods studied.

	<i>Proposal Method</i>	<i>Correlated Random-Walk</i>	<i>Component-wise approx. Gibbs</i>	<i>Block approx. Gibbs</i>
<i>Intervals</i>				
51-70	KL	3.54	1.43	3.98
	Run Time	33000	27000	9500
71-83	KL	1.54	7.74	10.1
	Run Time	23100	18500	6920
84-120	KL	222	26400	17700
	Run Time	29200	30800	9460
121-164	KL	2.63	0.477	2.17
	Run Time	15000	15000	9320
165-245	KL	9.63	2.24	1.90
	Run Time	12600	12600	9750

## Appendix B Supplementary Results to Section 4.1

Table B2 contain the Kullback-Leibler (KL) divergences of the posterior distributions derived using the naive SMC algorithm from the MCMC-obtained posteriors in Scenario 2, analogous to Table 1 for Scenario 1. As can be seen from the arbitrarily high values for the KL divergence, the increased dimensionality of the particles makes it even harder for the algorithm to find suitable areas of high posterior density with only a limited number of MH-steps.

## Appendix C Calculation of the Intra-class Correlation Coefficient (ICC)

In general, suppose a partitioning of data  $\mathbf{G}$  into  $I$  clusters, each of size  $d_i, i = 1, \dots, I$ , is proposed together with a hierarchical model:

$$g_{ij} = \mu + a_i + \epsilon_{ij}; \quad i = 1, \dots, I, j = 1, \dots, d_i,$$

where it is assumed that the cluster effects  $a_i$  are distributed as iid  $N(0, \sigma_a^2)$  and the residual errors,  $\epsilon_{ij}$  are distributed as iid  $N(0, \sigma_\epsilon^2)$ . Then the intra-class correlation coefficient,  $\rho$ , is defined (Donner and Koval, 1980):

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_\epsilon^2}. \quad (2)$$

The absence of any clustering in the data is indicated by  $\rho = 0$ .

The variance terms in (2) are, however, unknown. A common estimate used for  $\rho$  is the analysis of variance intra-class correlation coefficient, given by

$$r_A = \frac{(MS_a - MS_w)/d_0}{(MS_a - MS_w)/d_0 + MS_w},$$

where  $MS_a = \frac{1}{I-1} \sum_{i=1}^I d_i (\bar{g}_{i\cdot} - \bar{g}_{\cdot\cdot})^2$ ,  $MS_w = \frac{1}{d-1} \sum_{i=1}^I \sum_{j=1}^{d_i} (g_{ij} - \bar{g}_{i\cdot})^2$  and  $d_0 = \bar{d} - \frac{1}{d(I-1)} \sum_{i=1}^I (d_i - \bar{d})^2$ , represent a between-class mean sum-of-squares, a within-class mean sum-of-squares and an average class size respectively, with  $d = \sum_{i=1}^I d_i$ ,  $\bar{g}_{i\cdot} = \frac{1}{d_i} \sum_{j=1}^{d_i} g_{ij}$  and  $\bar{g}_{\cdot\cdot} = \frac{1}{d} \sum_{i=1}^I \sum_{j=1}^{d_i} g_{ij}$ . Clusters with  $d_i = 1$  are initially removed as they have zero within cluster variation (Donner and Koval, 1980).

## Appendix D Supplementary Results to Section 6

### D.1 ESS decay and the timing of rejuvenations

Figure D1 shows that, for both the discrete and continuous SMC, the number and timing of rejuvenations appears independent of the choice of  $r_A^*$ . This suggests that the decline in the ESS is independent of the quality of the initial sample.

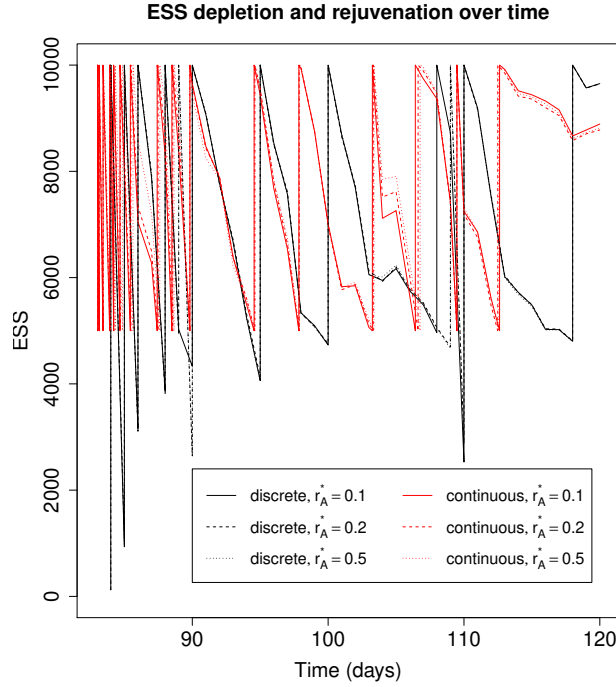


Figure D1: The declines in ESS over time and the timing of rejuvenations.

### D.2 KL calculations including background parameters

The KL calculations behind Table 3 have been based on approximations to the posterior distribution for  $\theta$  excluding the parameters of the background consultation model,  $\beta^B$ . Table D3 presents this table using KL calculations based on the full parameter vector  $\theta$ .

Table D3: Performance of the information-adjusted SMC algorithm over the interval 83-120 days, using the continuous filter and the continuous filter alternative with the negative binomial dispersion parameters removed from the block proposals.

<i>ICC threshold</i>	<i>0.5</i>	<i>0.2</i>	<i>0.1</i>	<i>ICC threshold</i>	<i>0.5</i>	<i>0.2</i>	<i>0.1</i>
<b>84 Days</b> (KL target = 6.54)				<b>90 Days</b> (KL target = 0.684)			
Continuous	<b>3.03</b>	<b>2.96</b>	<b>2.91</b>	Continuous	1.96	<b>0.528</b>	<b>0.232</b>
Cts. Reduced	<b>3.09</b>	<b>2.93</b>	<b>2.95</b>	Cts. Reduced	2.28	<b>0.268</b>	1.58
<b>85 Days</b> (KL target = 2.14)				<b>100 Days</b> (KL target = 1.42)			
Continuous	3.24	3.17	3.16	Continuous	<b>0.476</b>	<b>0.412</b>	<b>0.376</b>
Cts. Reduced	3.25	3.13	3.15	Cts. Reduced	<b>0.438</b>	<b>0.389</b>	<b>0.369</b>
<b>86 Days</b> (KL target = 2.23)				<b>110 Days</b> (KL target = 1.08)			
Continuous	3.47	3.43	3.43	Continuous	<b>0.360</b>	<b>0.259</b>	<b>0.283</b>
Cts. Reduced	3.48	3.39	3.46	Cts. Reduced	<b>0.404</b>	<b>0.194</b>	<b>0.203</b>
<b>87 Days</b> (KL target = 5.94)				<b>120 Days</b> (KL target = 1.50)			
Continuous	<b>2.67</b>	<b>2.58</b>	<b>2.55</b>	Continuous	<b>0.273</b>	<b>0.141</b>	<b>0.165</b>
Cts. Reduced	<b>2.68</b>	<b>2.62</b>	<b>2.56</b>	Cts. Reduced	<b>0.211</b>	<b>0.143</b>	<b>0.151</b>

## Appendix E Assessing one-step ahead forecasts

The Discussion of Section 7 of the main text introduces methods for assessing model adequacy based on a series of one-step ahead forecasts based on distributions  $F_k \equiv F_k(y_k | \mathbf{y}_{1:(k-1)})$ , the distribution function with corresponding (posterior) density  $\pi(y_k | \mathbf{y}_{1:(k-1)})$ ,  $k = 1, \dots, K - 1$ . For integer data  $y_k$ , the probability integral transform (PIT) as defined by Czado et al. (2009) can be used to informally assess the quality of these predictions. Suppose at time  $t_k$ , the predictive distribution function is  $F_k$  with subsequently realized value  $Y_k = y_k$ , the PIT for this prediction is defined to be:

$$G_k(u | y_k) = \begin{cases} 0 & u \leq F_k(y_k - 1) \\ (u - F_k(y_k - 1)) / (F_k(y_k) - F_k(y_k - 1)) & F_k(y_k - 1) < u \leq F_k(y_k) \\ 1 & u \geq F_k(y_k) \end{cases}.$$

Each of the datasets considered in the scenarios of Section 4 of the main text have a denominator associated with each observation. For the serological and virological data, these denominators are the respective sample sizes. For primary care consultations, these will be the size of the observable population on each day. Therefore, over a collection of such probabilistic forecasts  $F_k$ ,  $k = k_0, \dots, K$ , for which observations  $(y_{k_0}, \dots, y_K)$  are subsequently made based on denominators  $(n_{k_0}, \dots, n_K)$ , one can compile an overall PIT:

$$\bar{G}(u) = \frac{1}{\sum_{j=k_0}^K n_j} \sum_{i=k_0}^K n_i G_i(u | y_i).$$

Under a null hypothesis that the predictive distributions are those that generate the data, this overall PIT should resemble the distribution function of a uniform (0, 1) random variable, i.e.  $\bar{G}(u) = u$ . Alternatively, one can transform to a PIT histogram to check that it resembles the flat uniform distribution (as seen in Figure 7A).

Section 7 of the main text introduces the use of the logarithmic score function to formally assess the quality of probabilistic epidemic forecasts. As demonstrated by Seillier-Moiseiwitsch

and Dawid (1993), scoring rules can be used to test a hypothesis of prediction adequacy as illustrated by Held et al. (2017) who use a ranked probability scoring rule for discrete data,

$$s_{\text{rps}}(P, y) = \sum_{l=0}^{\infty} \{P(l) - \mathbb{1}(y \leq l)\}^2.$$

Over a collection of forecasts as described above a two-sided test of predictive adequacy can be calculated based on the  $z$ -statistic

$$z_{\text{rps}} = \frac{\bar{s}_{\text{rps}} - \mathbb{E}\bar{s}_{\text{rps}}}{\text{Var}(\bar{s}_{\text{rps}})}.$$

where

$$\bar{s}_{\text{rps}} = \frac{1}{\sum_{j=1}^K n_j} \sum_{k=1}^K n_k s_{\text{rps}}(F_k, \mathbf{y}_k).$$

quantities whose calculation may not be intuitively obvious. To calculate the  $z$  statistic (assuming for easy of representation that  $k_0 = 1$ ), we need:

$$\begin{aligned} \mathbb{E}(\bar{s}_{\text{rps}}) &= \frac{1}{\sum_{j=1}^K n_j} \sum_{k=1}^K n_k \mathbb{E}(s_{\text{rps}}(F_k, \mathbf{y}_k)) \\ &= \frac{1}{\sum_{j=1}^K n_j} \sum_{k=1}^K n_k \sum_{l=1}^{\infty} \mathbb{E}(F_k(l) - \mathbb{1}(\mathbf{y}_k \leq l))^2 \\ &= \frac{1}{\sum_{j=1}^K n_j} \sum_{k=1}^K n_k \sum_{l=1}^{\infty} \text{Var}(\mathbb{1}(\mathbf{y}_k \leq l)) \\ &= \frac{1}{\sum_{j=1}^K n_j} \sum_{k=1}^K n_k \sum_{l=1}^{\infty} F_k(l) (1 - F_k(l)). \end{aligned}$$

and denoting  $\mathbb{1}(\mathbf{y}_i \leq l)$  by  $\mathbb{1}_{il}$  and  $F_k(l) = p_{kl}$ :

$$\begin{aligned} \text{Var}(\bar{s}_{\text{rps}}) &= \frac{1}{\left(\sum_{j=1}^K n_j\right)^2} \sum_{k=1}^K n_k^2 \text{Var}(s_{\text{rps}}(P_k, \mathbf{y}_k)) \\ &= \frac{1}{\left(\sum_{j=1}^K n_j\right)^2} \sum_{k=1}^K n_k^2 \text{Var}(s_{\text{rps}}(F_k, \mathbf{y}_k) - \mathbb{E}(s_{\text{rps}}(F_k, \mathbf{y}_k))) \\ &= \frac{1}{\left(\sum_{j=1}^K n_j\right)^2} \sum_{k=1}^K n_k^2 \text{Var}\left(\sum_{l=0}^{\infty} (p_{kl} - \mathbb{1}_{kl})^2 - \mathbb{E}\left(\sum_{l=0}^{\infty} (p_{kl} - \mathbb{1}_{kl})^2\right)\right) \\ &= \frac{1}{\left(\sum_{j=1}^K n_j\right)^2} \sum_{k=1}^K n_k^2 \text{Var}\left(\sum_{l=0}^{\infty} (2p_{kl} - 1)(p_{kl} - \mathbb{1}_{kl})\right) \\ &= \frac{1}{\left(\sum_{j=1}^K n_j\right)^2} \sum_{k=1}^K n_k^2 \sum_{l=0}^{\infty} (2p_{kl} - 1)(1 - p_{kl}) \left(p_{kl}(2p_{kl} - 1) + 2 \sum_{m=0}^{l-1} p_{km}(2p_{km} - 1)\right). \end{aligned}$$

Calculating these statistics for each of the data types in scenario 2 gives the test statistics in Table E4, showing some evidence against the hypothesis of prediction adequacy in the doctor consultations.

Table E4:  $z$ -statistics and corresponding two-sided  $p$ -values testing the adequacy of one-step ahead forecasts of each type of data in Scenario 2

<i>Data</i>	<i>z</i>	<i>p-value</i>
Primary Care Consultations	2.10	0.036
Virological Sampling	-1.27	0.202
Serological Sampling	1.19	0.234

## Appendix F Code

Packaged code is in development to implement the algorithms in this paper. Unpolished R code used in all the analyses presented here can be found at <https://gitlab.com/pjbirrell/Rmpi-smc>.

## References

- Birrell, P. J., G. Ketsetzis, N. G. Gay, B. S. Cooper, A. M. Presanis, R. J. Harris, A. Charlett, X.-S. Zhang, P. White, R. G. Pebody, and D. De Angelis (2011, November). Bayesian modelling to unmask and predict the influenza A/H1N1pdm dynamics in London. *Proc. Natn. Acad. Sci. USA* 108(45), 18238–18243.
- Czado, C., T. Gneiting, and L. Held (2009). Predictive Model Assessment for Count Data. *Biometrics* 65(4), 1254–1261.
- Donner, A. and J. J. Koval (1980). The estimation of intraclass correlation in the analysis of family data. *Biometrics* 36(1), 19–25.
- Held, L., S. Meyer, and J. Bracher (2017). Probabilistic forecasting in infectious disease epidemiology: the 13th armitage lecture. *Statistics in Medicine* 36(22), 3443–3460. sim.7363.
- Seillier-Moiseiwitsch, F. and A. P. Dawid (1993). On testing the validity of sequential probability forecasts. *J. Am. Statist. Ass.* 88(421), 355–359.