Investigating the UK measles outbreak between 1944 to 1962 using the hhh4 model

Kajal Dodhia[1], Jordan J Hood[2]

University of Glasgow[1], STOR-i, Lancaster University[2]

Background

- SARS-CoV-2 (Covid-19) virus has had a significant impact on our lives since January 2020. Vaccines do not provide complete immunity, hence SARS-CoV-2 will remain a part of our lives until we reach a resolution.
- World events and societies changing behaviours greatly influence what global infections look like. However, such a scenario can be determined from other infectious disease such as Influenza which exhibits local endemic and epidemic phases.
- The prevaccine UK measles data set from 1944-62 is analysed using the hhh4 model, starting with a basic model, adding covariates and taking into account random effects in order to explore the best fitting model and predicting future outbreaks.

Basic Model

An endemic-epidemic multivariate time-series model for infectious disease counts Y_{it} from units $i = 1, \cdots, 60$ during periods $t = 1, \cdots, 493$ where i denotes city and t denotes fortnightly time. The hhh4 model assumes that, $Y_{it}|\mathcal{F}_{t-1} \sim \mathcal{NB}(\mu_{it}, \psi)$ where

$$\mu_{it} = e_i\nu_t + \lambda Y_{i,t-1} + \phi \sum_{j \neq i} w_{ji} Y_{i,t-1},$$
$$\log(\nu_t) = \alpha^{(\nu)} + \beta_t t + \gamma \sin(\omega t) + \delta \cos(\omega t),$$

and overdispersion parameter $\psi_i > 0$ such that the conditional variance of Y_{it} is $\mu_{it}(1 + \psi_i \mu_{it})$. The link function is $\log(\mu_i)$.

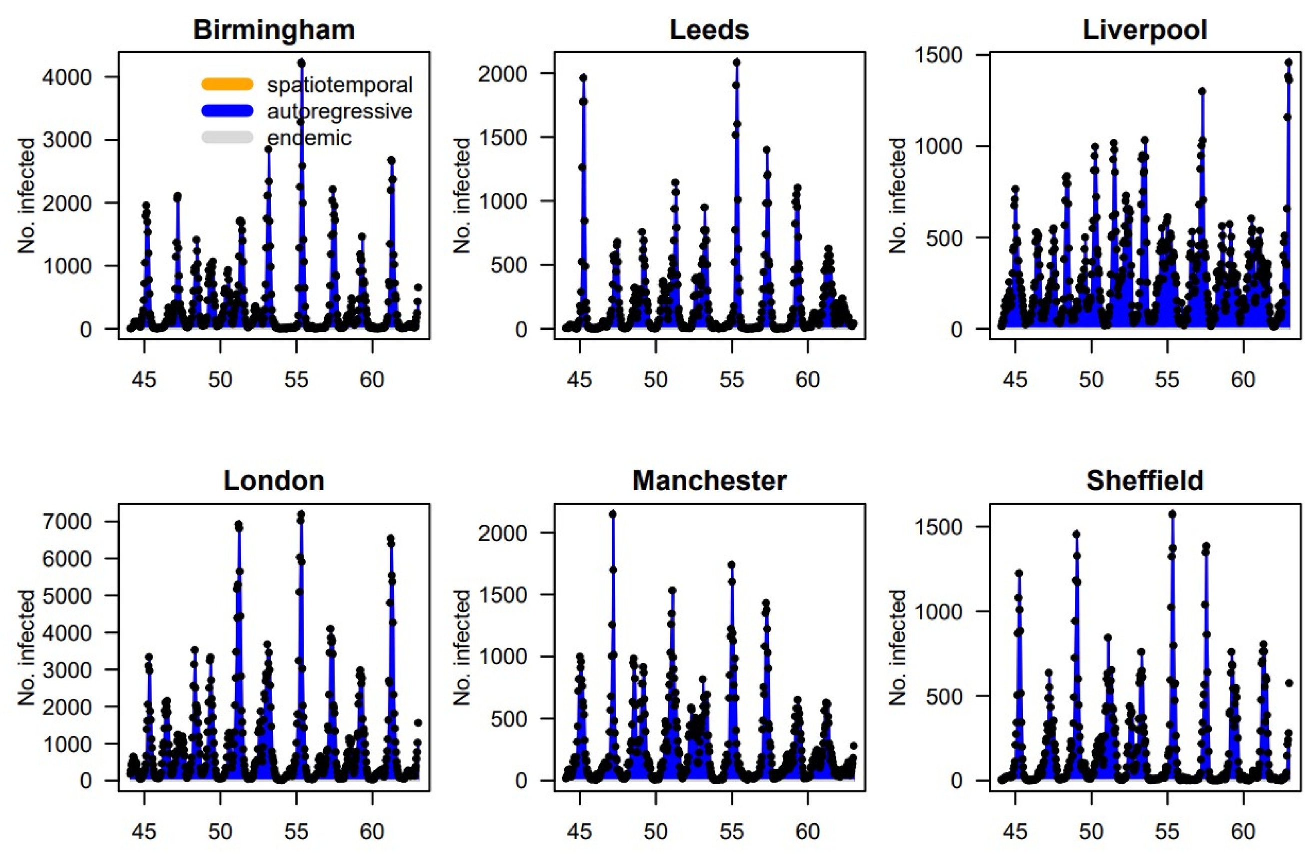

Figure 1: Fitted components in the initial model for the cities with more than 80,000 total infections. Dots are drawn for positive weekly counts. The basic model shows that the largest proportion of the fitted mean results from the within-city autoregressive component to the data and captures seasonality.

Model parameters explained below:

1. **Endemic log-linear predictor ν_t:**
   - Temporal variation of disease incidence incorporates an overall trend and a sinusoidal wave of frequency $\omega = \frac{2\pi}{26}$.
   - Population fraction as multiplicative offset e_i.
2. **Epidemic Component:**
   - Autoregressive parameter: $\lambda = \exp(\alpha)^{\lambda}$.
   - Spatio-temporal parameter: $\phi = \exp(\alpha)^{\phi}$.

- These are assumed homogeneous across cities and constant over time and in this model the epidemic can only arrive from directly adjacent districts.
- The Negative Binomial which accounts for overdispersion is a better fit compared to the Poisson model as it has a lower Akaike information criterion (AIC) value of 218770.4.

Adding Covariates Step by Step

1. Scaled infections against **average employment domain score (E)** and **average domain income score (I)** had a moderate positive correlation 0.5854129 and 0.4786521 respectively. By performing AIC-based model selection, the lowest AIC for both employment and income shows that they can be added to the endemic predictor as a covariate.
2. **Random effects** are added to the model as cities display heterogeneous incidence levels not explained by observed covariates. This could be caused by under reporting (GLMM).

Final Model

The final model with independent random effects in all three components: $\alpha_i^{(\nu)} \overset{\text{iid}}{\sim} \mathcal{N}(\alpha^{(\nu)}, \sigma_\nu^2)$, $\alpha_i^{(\lambda)} \overset{\text{iid}}{\sim} \mathcal{N}(\alpha^{(\lambda)}, \sigma_\lambda^2)$ and $\alpha_i^{(\phi)} \overset{\text{iid}}{\sim} \mathcal{N}(\alpha^{(\phi)}, \sigma_\phi^2)$,

$$\mu_{it} = e_i\nu_t + \lambda Y_{i,t-1} + \phi \sum_{j \neq i} w_{ji} Y_{i,t-1},$$
$$\log(\nu_t) = \alpha_i^{(\nu)} + \beta_t t + \gamma \sin(\omega t) + \delta \cos(\omega t) + \beta_E \log(E_i) + \beta_I \log(I_i)$$

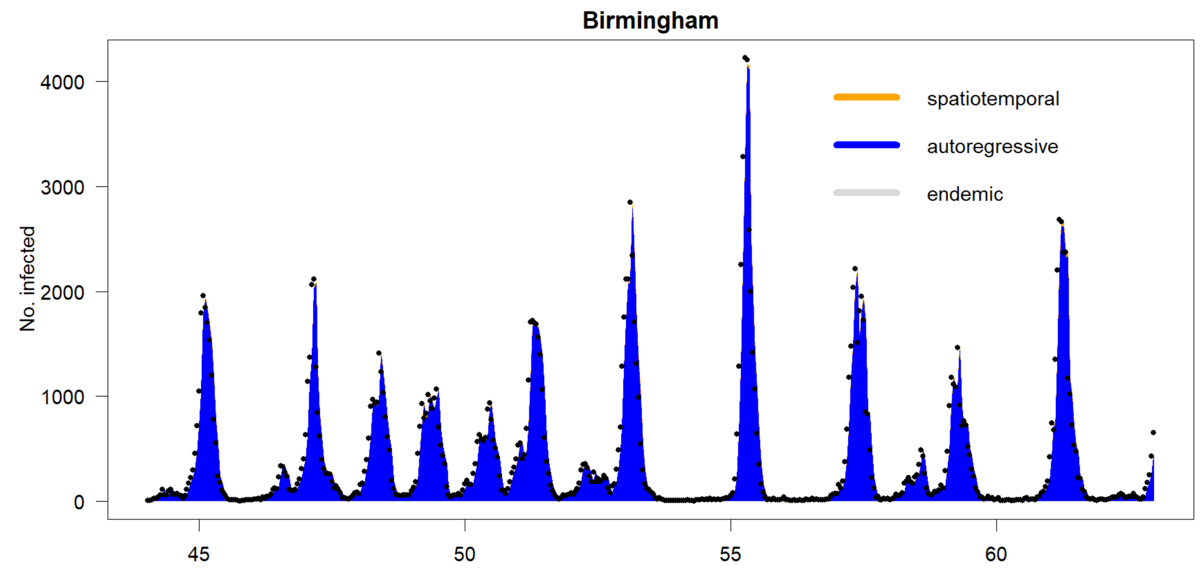

Figure 2: Fitted components in the final model for Birmingham.

Fitted components in the final model for most cities are almost identical to Figure 1 but there is a slight increase in the proportion of fitted mean captured by the spatio-temporal component for Birmingham, Figure 2. The lack of transmission between cities could be due to less ways and needs of travelling between cities, hence less physical contact to pass the disease amongst cities compared to today.

Predictive Model Assessment: Test Period: 1962

- **Goodness of fit assessment:** The final model gave the smallest mean score for all scoring methods, hence it is the best fitting model.
- **True two-week-ahead predictions:** The most parsimonious initial model is the final model which gives the best two-week-ahead predictions in terms of overall mean scores.
- **Paired t-test:** The p value (0.00052) is small, so there is significant difference between the mean scores of the basic and the final model with regard to predictive performance during the test period.
- **Calibration Test:** The p value (2.2e-16) is small, therefore null hypothesis that the model is well calibrated is rejected. Figure 3 shows miscalibration.

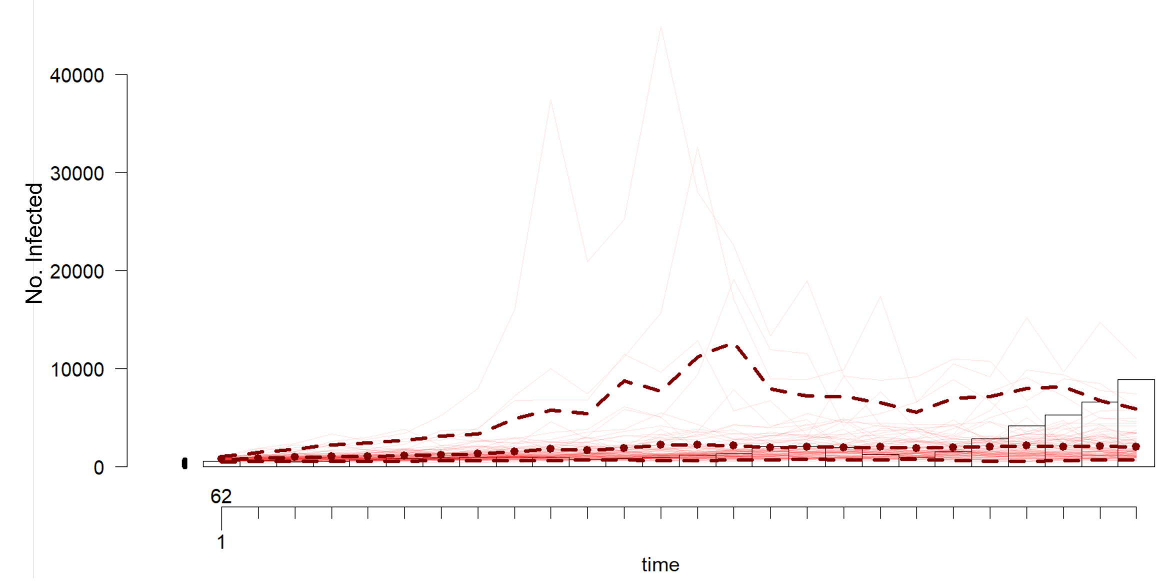

Figure 3: Simulation-based forecast of 1962 starting from the second last week in 1961 (vertical bar on the left), showing the counts aggregated over all districts. The fortnightly mean of the simulations is represented by dots, and the dashed lines correspond to the pointwise 2.5% and 97.5% quantiles. The actually observed counts are shown in the background.

Conclusions

The final model which incorporates the covariates for average of employment and income domain score and random effects is the best fit for the measles data set. However, in predicting, it does not capture the large number of cases for the year 1962. The largest portion of the fitted mean results from the within-district autoregressive component, a very small spatio-temporal and almost negligible endemic component to the data.

References

Damon Berridge, Robert Crouchley, and Daniel Grose. *Multivariate generalized linear mixed models using R*. CRC Press Boca Raton, FL, 2011.

Sebastian Meyer, Leonhard Held, and Michael Höhle. Spatio-temporal analysis of epidemic phenomena using the r package surveillance. *Journal of Statistical Software*, 77(11):1–55, 2017.

2366389D@student.gla.ac.uk