Impacts of the day-of-the-week effect in epidemic trend inference

Phan Minh Quan Nguyen supervised by Oliver Eales and Freya Shearer

Infectious Disease Dynamics Unit, The University of Melbourne

Introduction

Inferring the trends of an epidemic time series (e.g. daily number of cases) is crucial for informing effective public health responses. There can be substantial noise in case time series due to variations in daily case reporting rates. Additionally, case reporting and testing behaviors may change depending on the day of the week, introducing further noise in case counts. **Aims:**

- **Retrospectively assess the performance of statistical models incorporating day-of-the-week (DOW) effects compared to models not incorporating DOW effects.**
- **Evaluate the influence of DOW effects on real-time inference across different phases of the epidemic.**

Methods

The Bayesian P-spline model [1] is used to fit smoothed trends to case time series. The model fits equally spaced (knots every 5 days) basis-splines (piecewise cubic polynomial functions) to approximate the underlying trend. A log link function transforms the smoothed trend as a linear combination of N B-splines into the expected daily number of cases:

$$\log \mu_t = \sum_{i=1}^{N} b_i B_{i,n}(t)$$

where μ_t is the expected number of cases at time t, b_i are the coefficients that weight each of the B-spline basis functions at time t.

The model assumes a Negative Binomial likelihood for the daily number of cases, with additional over-dispersion parameter (ϕ) to account for overdispersion. The expected number of cases, scaled by ϕ, serves as the mean in the likelihood function.

DOW effects are incorporated by adjusting the expected case count with a multiplicative factor θ_{DOW_t}, which accounts for systematic variations in case reporting across different weekdays: [2]

$$\log \mu_t = \sum_{i=1}^{N} b_i B_{i,n}(t) + \log \theta_{\text{DOW}_t}$$

where θ_{DOW_t} represents DOW effect for data at time t, and $\sum_{d=1}^{7} \theta_d = 1$.

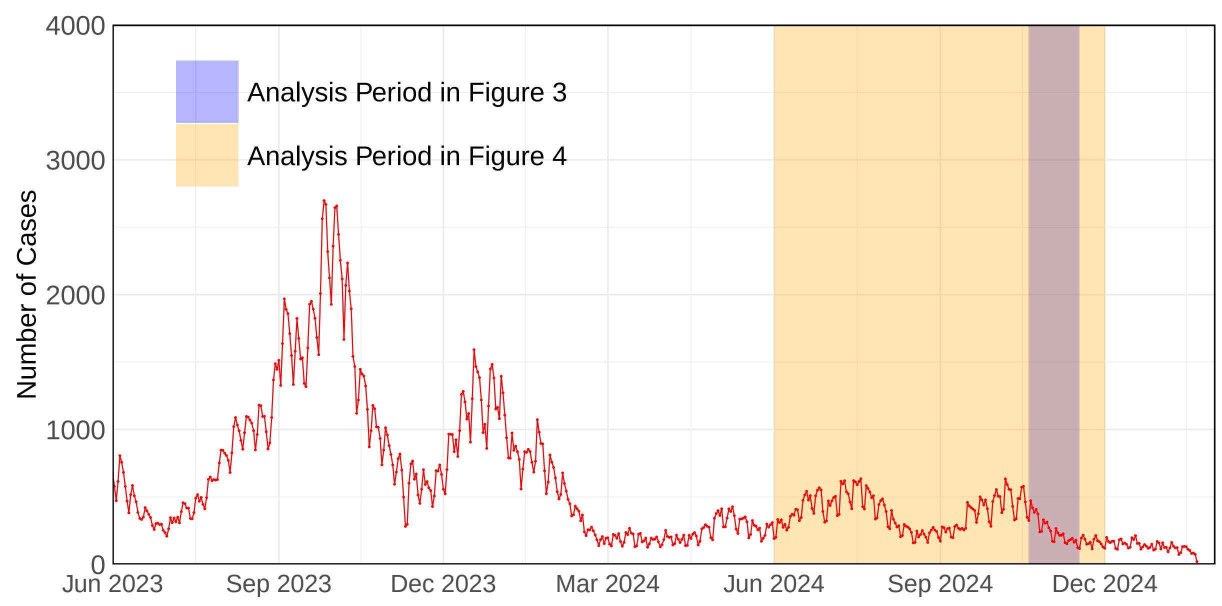

Figure 1. SARS-CoV-2 case time-series (June 2023 to January 2025) in the United Kingdom' (red).

Kullback-Leibler Divergence

Kullback-Leibler (KL) Divergence is a statistical measure that quantifies the differences between two probability distributions. It is computed as:

$$D_{KL}(P\|Q) = \int p(x) \log \frac{p(x)}{q(x)}\, dx$$

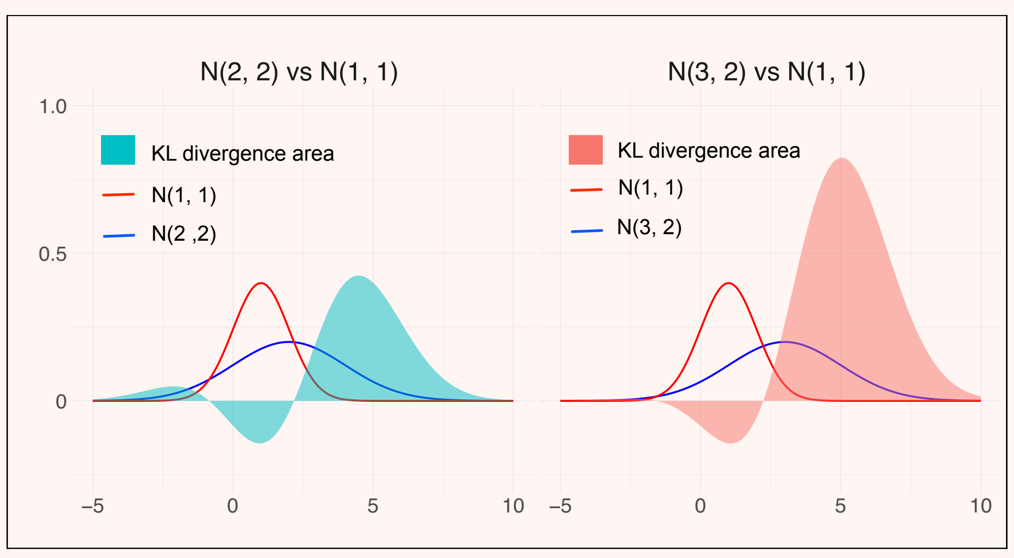

Figure 2. KL Divergence Visualization for Normal Distributions

We use the KL divergence to compare the posterior distributions of two models (Figure 4). We compare "real-time" models, fitted to the previous 365 days of data up to specific dates (each day of the analysis period), and "overall models", which are fit to the entire time series.

Results

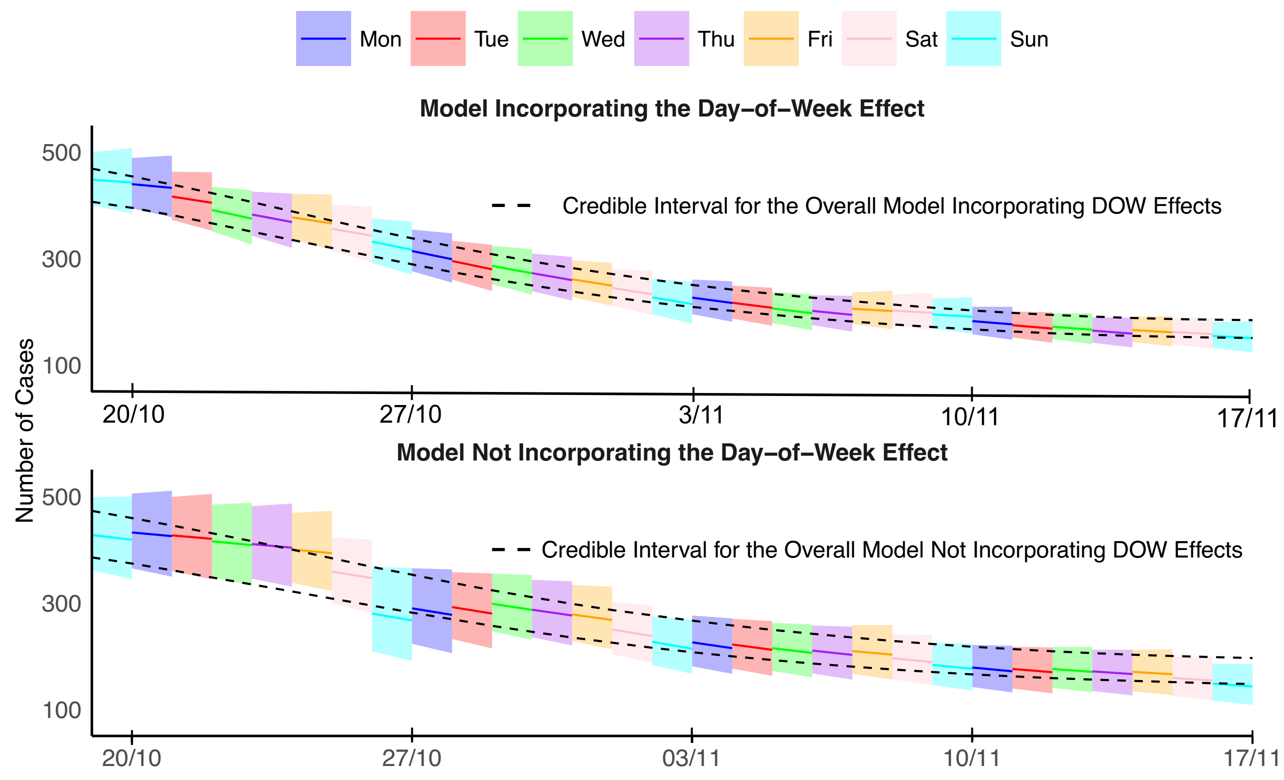

Figure 3. Comparing Posterior Estimates of Case Numbers for Real-Time and Overall Models, Incorporating and Not Incorporating DOW Effects.

- The model incorporating DOW effects has narrower credible intervals compared to the model not incorporating DOW effects.
- While both models show similar overall trends, the model incorporating DOW effects provides a more refined estimate by modeling the weekly noise due to DOW effects explicitly.

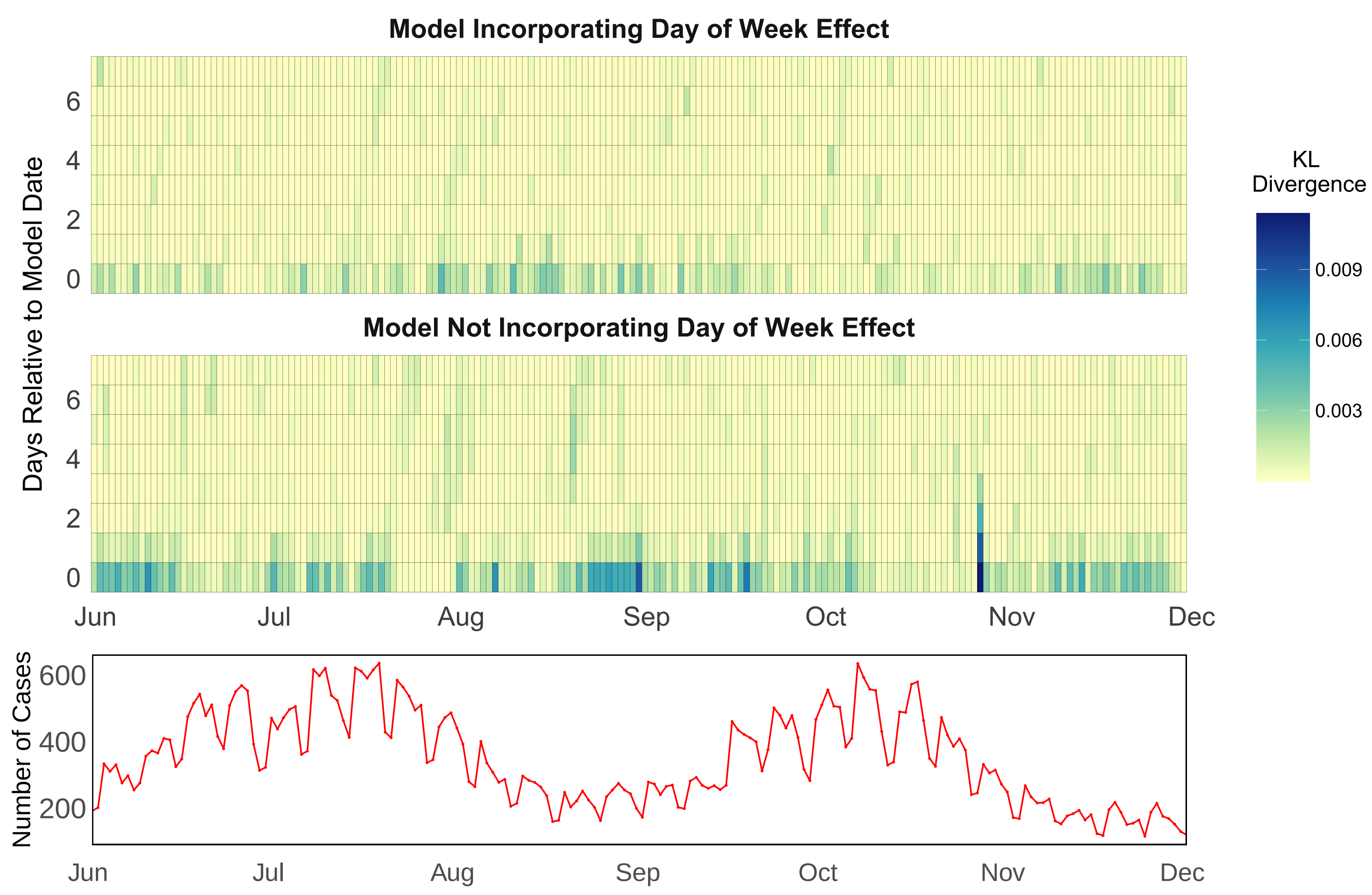

Figure 4. KL Divergence heatmap: real-time model inferences (most recent 8 days) vs. retrospective overall model.

- The inference for the most recent day of data (day 0) diverges most from the overall model for models both incorporating/not incorporating DOW effect, with a less pronounced difference in the model incorporating DOW effect.
- Generally, the model that does not incorporate DOW effects exhibits higher KL divergence, with the real-time model diverging more strongly from the overall model throughout the analysis period compared to the model incorporating DOW effects.

Discussion and Conclusions

- The inclusion of DOW effects significantly enhances the precision of the model's estimates, as evidenced by narrower credible intervals. By accounting for weekly patterns in noise, the model reduces uncertainty, providing more precise estimates of temporal dynamics, enabling better-informed public health responses to emerging trends.
- By incorporating DOW effects, the model captures and explains weekly fluctuations in the data, which, if unaccounted for, could distort the true epidemic trends and potentially result in misguided public health policies. This improves the interpretability of short-term patterns, especially in cases where day-to-day variations are critical.
- Further research is needed to examine how the magnitude and stability of DOW effects vary across different regions, diseases, and surveillance systems over time. Understanding these variations can help assess whether the model presented here remains robust across different settings.

References

[1] Eales, Oliver, K. E. C. Ainslie, C. E. Walters, et al. 2022. "Appropriately Smoothing Prevalence Data to Inform Estimates of Growth Rate and Reproduction Number." Epidemics 40: 100604. https://doi.org/10.1016/j.epidem.2022.100604.

[2] Eales, Oliver, Saras M. Windecker, James M. McCaw, and Freya M. Shearer. 2024. "Inferring Temporal Trends of Multiple Pathogens, Variants, Subtypes or Serotypes from Routine Surveillance Data." medRxiv. https://doi.org/10.1101/2024.11.03.24316681.