The endemic-epidemic model as a semi-mechanistic spatio-temporal model of disease spread

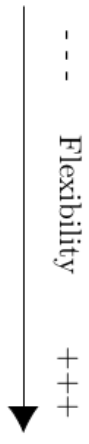
Johannes Bracher

Karlsruhe Institute of Technology
Heidelberg Institute for Theoretical Studies

September 6, 2024

A continuum of modelling approaches

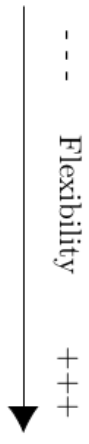
- *Agent-based modelling:* detailed simulation of epidemics at the individual level
- *Compartmental models:* mechanistic description of infection processes at the population level via compartments
- *Statistical/empirical models:* (interpretable) statistical description of observable patterns
- *Machine learning approaches:* mostly black-box approaches to capture observable patterns



Source: my PhD thesis, so highly authoritative.

A continuum of modelling approaches

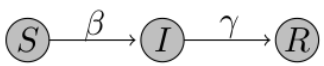
- *Agent-based modelling:* detailed simulation of epidemics at the individual level
- *Compartmental models:* mechanistic description of infection processes at the population level via compartments
- *Endemic-epidemic model.*
- *Statistical/empirical models:* (interpretable) statistical description of observable patterns
- *Machine learning approaches:* mostly black-box approaches to capture observable patterns

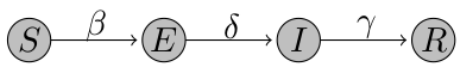


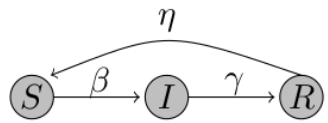
Source: my PhD thesis, so highly authoritative.

EE and compartmental epidemic models

- Epidemics are often modelled using compartmental models.







  - "mechanistic" reflection of disease spread.
  - traditionally continuous-time and deterministic (ODEs).

$$\frac{\mathrm{d}S(t)}{\mathrm{d}t} = -\frac{\beta}{N}S(t)I(t), \quad \frac{\mathrm{d}I(t)}{\mathrm{d}t} = \frac{\beta}{N}S(t)I(t) - \gamma I(t), \quad \frac{\mathrm{d}R(t)}{\mathrm{d}t} = \gamma I(t).$$

  - susceptible dynamics are key for model behaviour.

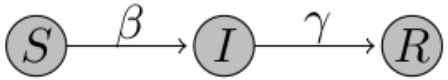
EE and compartmental epidemic models (II)

- The endemic-epidemic model can be seen as a strongly simplified discrete-time stochastic SIR model.
  - see Bauer and Wakefield[1] for detailed derivations.
- But ultimately the endemic-epidemic model is **not** a fully mechanistic model.
  - susceptible dynamics are ignored.
  - many model elements are pragmatic rather than derived from first principles (e.g., negative binomial distribution).

[1] C Bauer and J Wakefield (2018): Stratified space-time infectious disease modelling, with an application to hand, foot and mouth disease in China. JRSSA.

EE vs TSIR

- If you are looking for a (univariate) model with susceptible dynamics, **T**ime Series **SIR** may be the right choice for you.



$$I_t \mid I_{t-1}, S_{t-1} \sim \text{NegBin}(\lambda_t, 1/I_{t-1})$$

$$\lambda_t = \frac{\beta}{N} S_{t-1} I_{t-1}^{\alpha}$$

$$S_t = S_{t-1} - I_t.$$

- R package: Becker and Grenfell (2017): `tsiR`: An R package for time-series Susceptible-Infected-Recovered models of epidemics. PLOS One.

EE and count time series models

- Technically, the EE model is a multivariate **I**nteger-valued **G**eneralized **A**utoregressive **C**onditional **H**eteroscedasticity (INGARCH) model.
- If you care about ergodicity, stationarity etc, there is a vast literature on INGARCH models.
- Several R packages exist:

tscount: An R Package for Analysis of Count Time Series Following Generalized Linear Models

Tobias Liboschik
TU Dortmund University

Konstantinos Fokianos
University of Cyprus

Roland Fried
TU Dortmund University

Inference for Network Count Time Series with the R Package PNAR

The R Journal

by Mirko Armillotta, Michail Tsagris, and Konstantinos Fokianos

What the EE model offers

- EE is **more pragmatic than full mechanistic models.**
  - simple base model facilitates multivariate extension.
  - latent susceptible dynamics are ignored.
  - simple maximum likelihood inference can be used.
- EE is **more tailored than generic count time series models.**
  - identifiability ensured by "semi-mechanistic" parameterizations.
  - complexity "spent" on epidemiologically relevant aspects.
- The EE model has a **robust and longstanding implementation** in the R package `surveillance`.

Reminder: Multivariate model structure

- The multivariate endemic-epidemic model is defined as

$$Y_{rt} \mid \text{past} \sim \text{NegBin}(\mu_{rt}, \psi_r) \tag{1}$$

$$\mu_{rt} = \nu_{rt} + \phi_{rt} \times \sum_{r'=1}^{N} w_{r'r} \times Y_{r',t-1} \tag{2}$$

- As in `surveillance` within-region dynamics are given extra flexibility we often also write

$$\mu_{rt} = \underbrace{\nu_{rt}}_{\texttt{end}} + \underbrace{\lambda_{rt} \times Y_{r,t-1}}_{\texttt{ar}} + \underbrace{\phi_{rt} \times \sum_{r' \neq r} w_{r'r} \times Y_{r',t-1}}_{\texttt{ne}}. \tag{3}$$

- How do we handle all these parameters smartly?

Semi-mechanistic elements

- The EE framework accommodates the following epidemiologically meaningful mechanisms:
  - seasonality (and other external drivers).
  - simple but well-motivated mechanisms for spatial spread.
  - integration of social contact data.
  - encoding of generation times.
  - spatially smooth efects (random effects).

Seasonality and covariates

- Reminder: the parameters $\nu_{r,t}$, $\lambda_{r,t}$ and $\phi_{r,t}$ are modelled in a log-linear fashion to account for seasonality or other covariates, e.g.,

$$\log(\nu_{r,t}) = \alpha_i + \gamma \sin(2\pi t/52) + \delta \cos(2\pi t/52).$$

- Often it makes sense to share some parameters (γ, δ) across units $r = 1, \ldots, N$, while others are unit-specific (α_i).

```
formula_end <- addSeason2formula(
                    ~0 + fe(1, unitSpecific = TRUE),
                    S = 1)
```

- Intuition: Seasonality and other covariates modify disease import and transmission ($\approx$ reproductive numbers).

The power law

- A simple spatial coupling (and the default in `surveillance`) is to set

$$w_{r'r} = \begin{cases} 1 & \text{if } r, r' \text{ are neighbours} \\ 0 & \text{else.} \end{cases}$$

- A smart way to allow dependences between indirect neighbours is a power law,

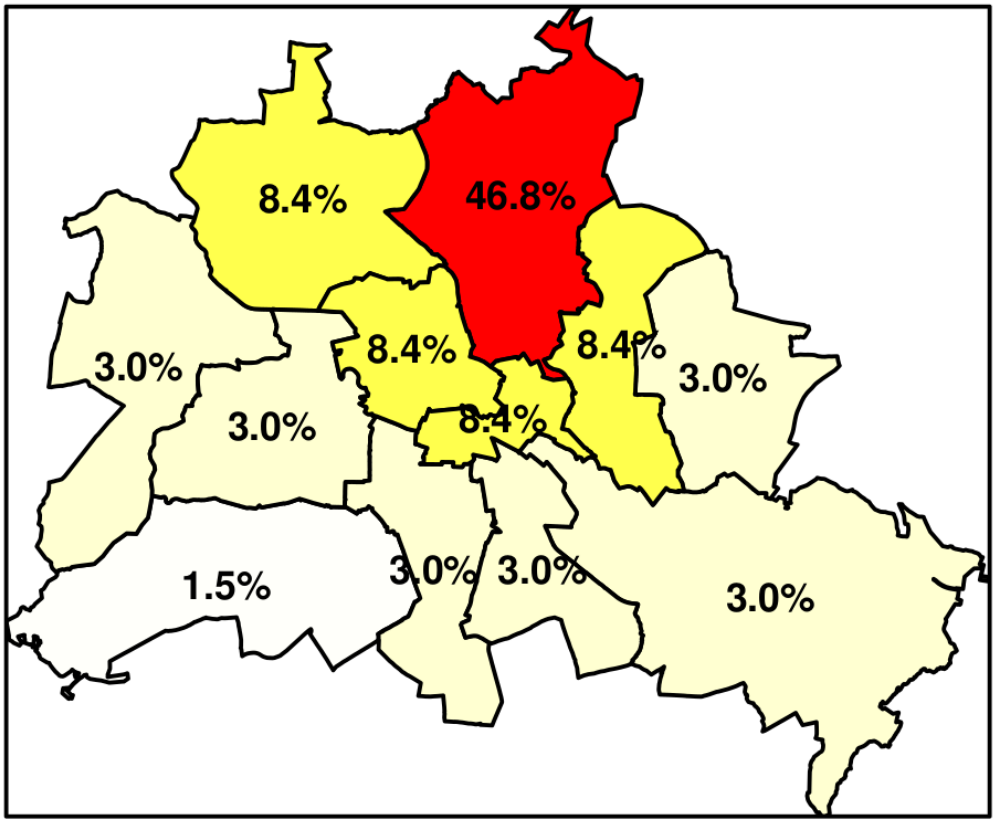
$$w_{r'r} \propto (o_{r'r} + 1)^{-\rho}.$$

- Weights are typically normalized such that $\sum_{r'=1}^{N} w_{rr'} = 1$.

```
formula_ne <- list(f = ~0 + fe(1, unitSpecific = TRUE),
                           weights = W_powerlaw(maxlag = 5,
                                                normalize = TRUE,
                                                log = TRUE))
```

Power law (II)

- Example: How does one district "distribute" its infectious pressure under the power law ($\rho = 2.5$)?

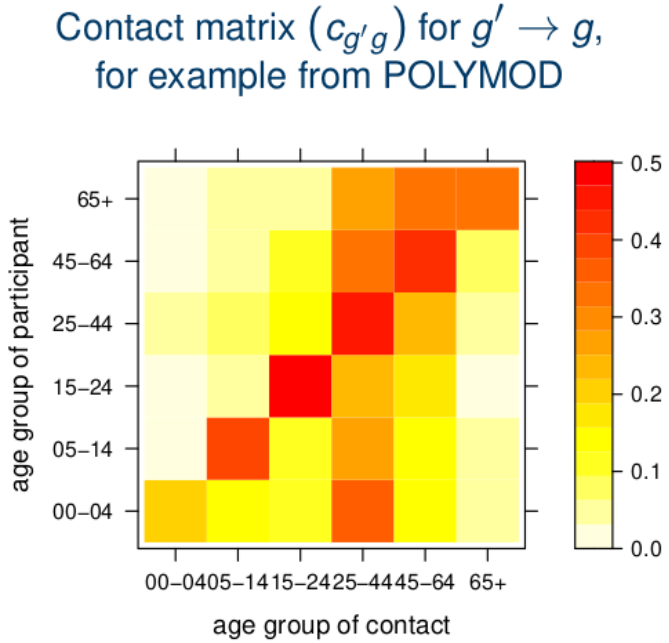


Why a power law?

- Empirical evidence indicates that "the distribution of travelling distances decays as a power law."
  - D Brockmann, L Hufnagel, T Geisel (2006): The Scaling Laws of Human Travel. Nature.
- In the EE framework, power laws have been found to outperform other (more complex) specifications.
  - S Meyer and L Held (2014): Power-law models for infectious disease spread. AOAS.
  - Geilhufe et al (2014): Power law approximations of movement network data for modeling infectious disease spread. Biometrical Journal.

Social contact matrices

- When modelling spread across age groups rather than space, social contact data can be used to parameterize the $w_{r',r}$.[2]



- Side note: Pioneering work[3] came from U Hasselt!

[2] S Meyer and L Held (2017): Incorporating social contact data in spatio-temporal models for infectious disease spread. Biostatistics.

[3] Hens et al (2009): Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries. BMC Infectious Diseases.

Generation times

- Using the `hhh4addon` package, the EE model can be extended to[4]

$$Y_{rt} \mid \text{past} \sim \text{NegBin}(\mu_{rt}, \psi_r) \tag{4}$$

$$\mu_{rt} = \nu_{rt} + \phi_{rt} \times \sum_{r'=1}^{N} \sum_{d=1}^{D} w_{r'r} \times u_d \times Y_{r',t-d}, \tag{5}$$

where $u_1, \ldots, u_D$ is the generation time / serial interval distribution.

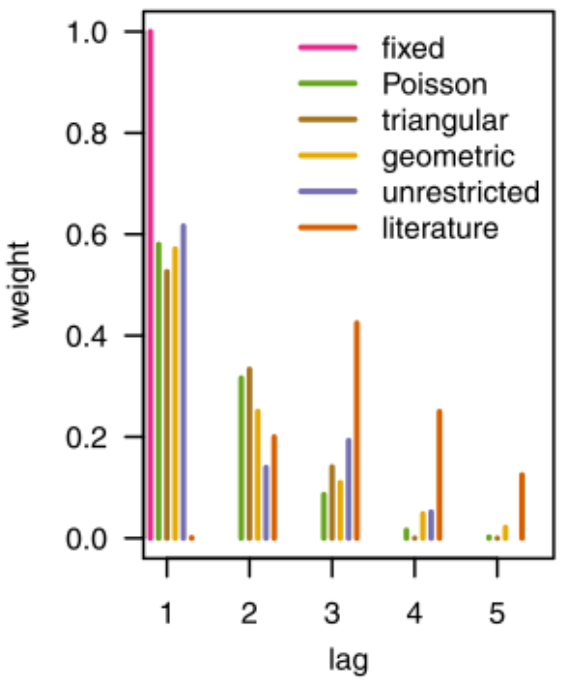
[4] Bracher and Held (2020). Endemic-epidemic models with discrete-time serial interval distributions for infectious disease prediction. IJF.

Generation times (II)

- Generation time distributions can be fixed based on literature estimates or estimated parametrically (`?profile_par_lag`), e.g.,

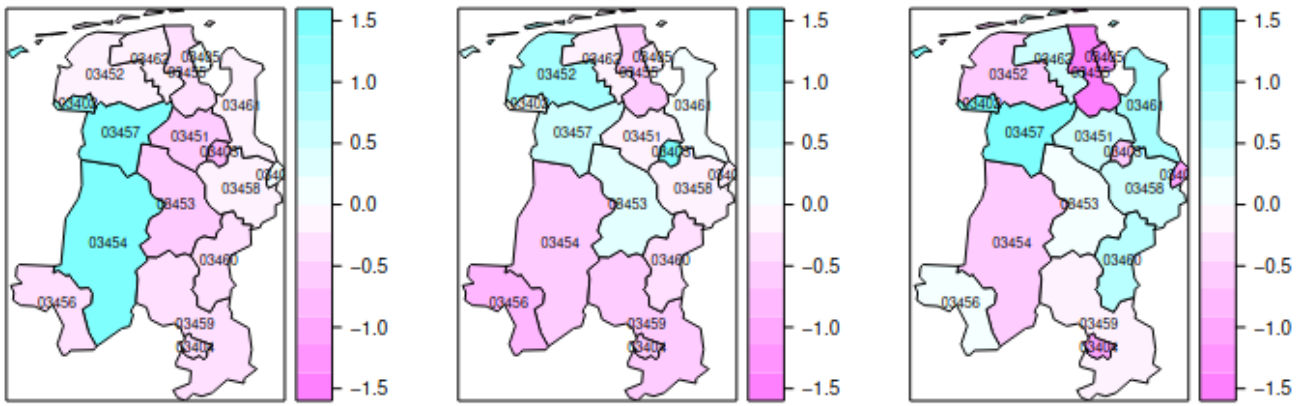
$$u_d = (1 - \pi)^{k-1}\pi$$

- Example: Dengue in Puerto Rico (Bracher and Held 2020).



Spatial random effects

- For models with many strata and many parameters, spatially structured (CAR) random effects can be used.
- Example from Meyer et al[5]:



(a) Autoregressive $\alpha_i^{(\lambda)}$ (b) Spatio-temporal $\alpha_i^{(\phi)}$ (c) Endemic $\alpha_i^{(\nu)}$

Figure 18: Maps of the estimated random intercepts.

[5] S Meyer et al (2017): Spatio-Temporal Analysis of Epidemic Phenomena Using the R Package surveillance. JSS.

What is the EE model used for?

- The EE model was conceived as a generic tool to *"provide an adequate fit, reliable one-step-ahead prediction intervals"* and *"capture space–time dependence caused by the spatial spread of a disease over time"* (Held, Höhle, Hoffmann 2005).
- Over time it has been used for a variety of purposes (some anticipated, some not).

Forecasting

- Robert et al[6] use the EE framework to generate national and subnational-level forecasts of COVID-19 cases and deaths.

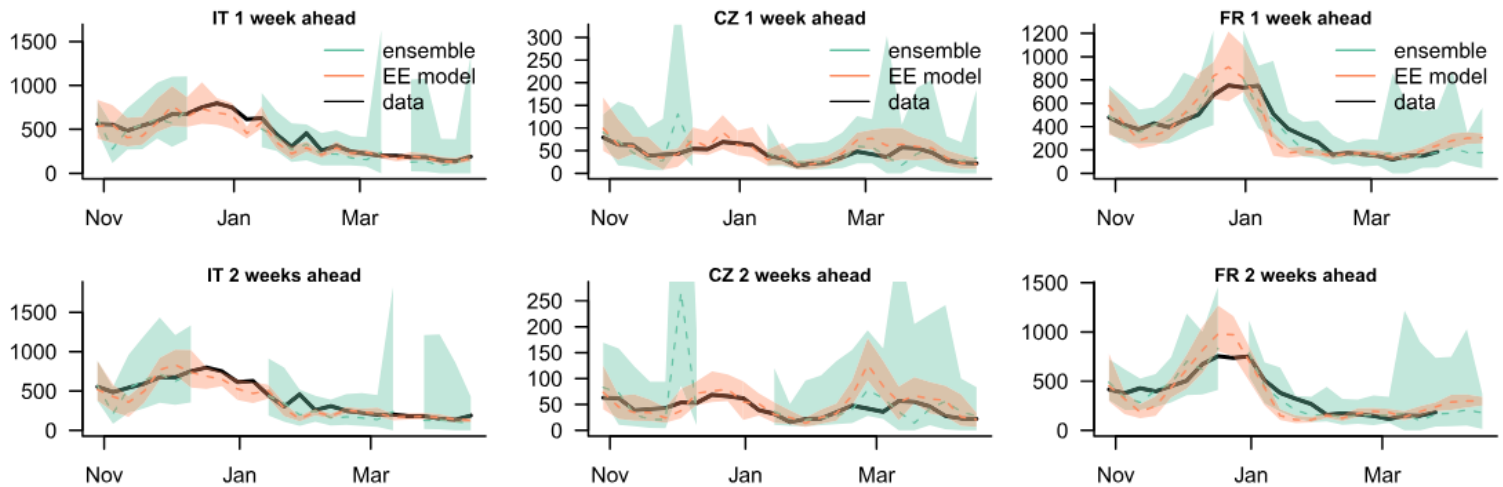
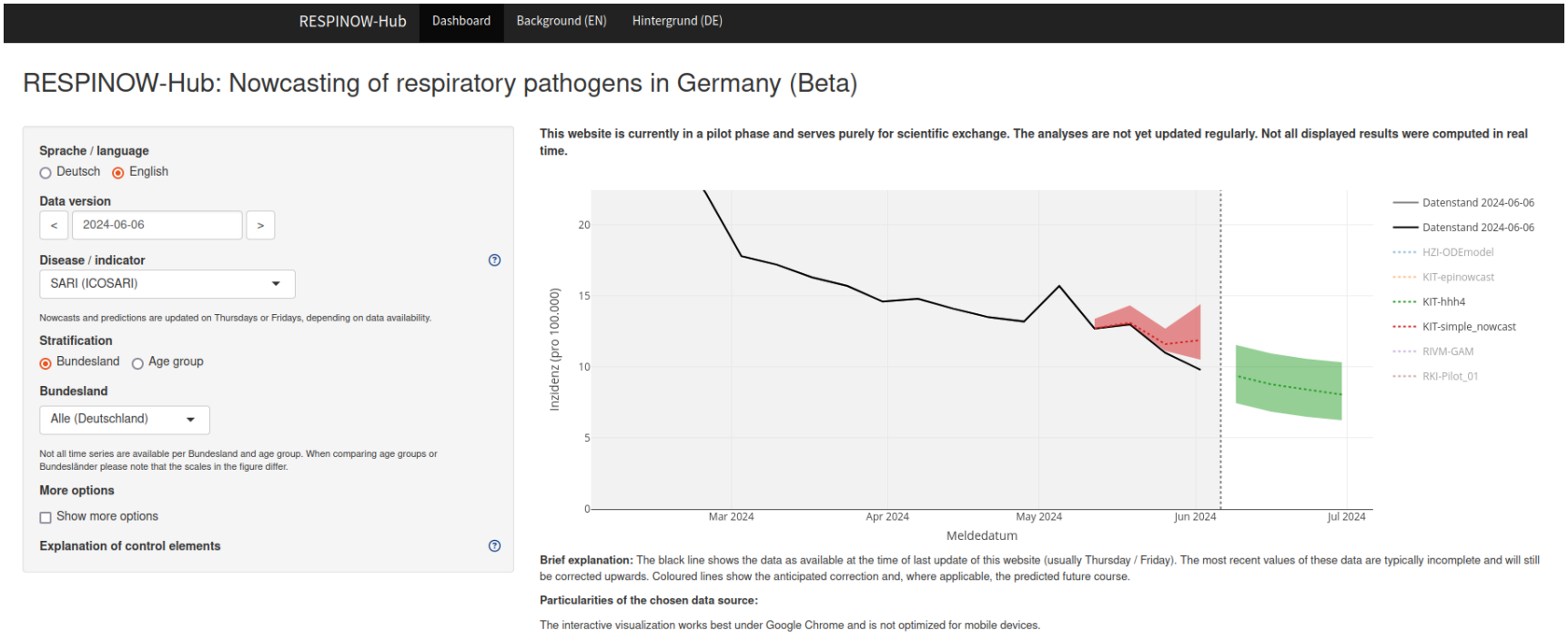


Image license: `http://creativecommons.org/licenses/by/4.0/`

[6] A Robert et al (2024): Predicting subnational incidence of COVID-19 cases and deaths in EU countries. BMC Infetious Diseases.

Forecasting (II)

- Within the RespiNow Consortium, we use the EE model e.g., to predict weekly SARI hospitalizations in Germany:



Vaccination

- Herzog et al[7] study the impact of measles vaccination coverage on the occurrence of measles.
- Model for bi-weekly measles counts Y_{rt} in 16 German states:

$$
\begin{aligned}
Y_{rt} \mid \text{ past} &\sim \text{NegBin}(\mu_{rt}, \psi_r) \\
\mu_{rt} &= \nu_{rt} + \lambda_r \times X_{t-1} \\
\log(\nu_{r,t}) &= \alpha_i + \gamma \sin(2\pi t/26) + \delta \cos(2\pi t/26). \\
\lambda_r &= \beta_0 + \beta_1 \times \log(\text{proportion unvaccinated school starters in } r)
\end{aligned}
$$

- Result: "... a significant association between estimated vaccination coverage at school entry and the overall incidence of measles'.'
- Data are available in `surveillance`.

[7] S Herzog et al (2011): Heterogeneity in vaccination coverage explains the size and occurrence of measles epidemics in German surveillance data. Epi&Inf.

NPIs / counterfactuals

- Grimée et al[8] study the impact of border closures between Switzerland and Italy by producing conterfactual scenarios.

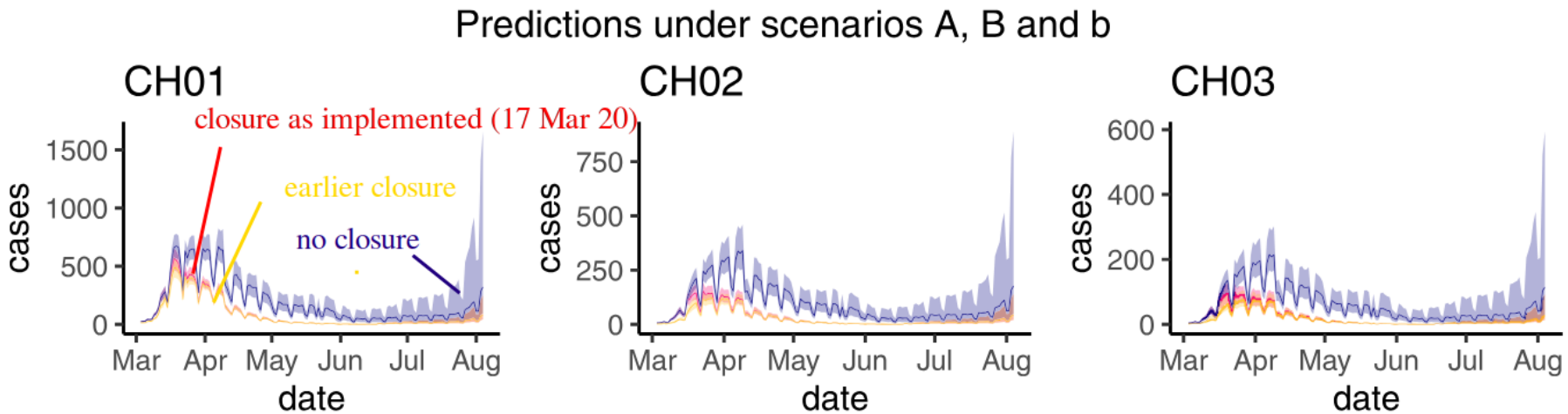


Image license: `https://creativecommons.org/licenses/by/4.0/`

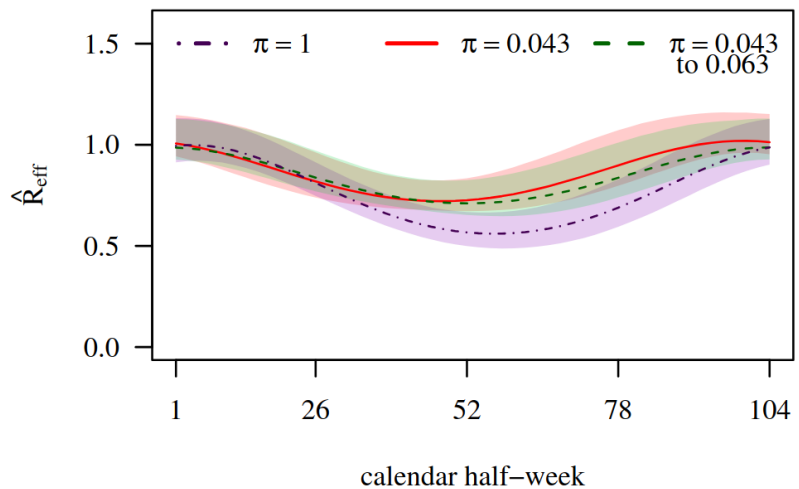
[8] M Grimée et al (2022): Modelling the effect of a border closure between Switzerland and Italy on the spatiotemporal spread of COVID-19 in Switzerland. Spatial Statistics.

Estimation of local reproductive numbers

- Bauer and Wakefield (HMF disease) and Bracher and Held[9] (rotavirus) estimate local effective reproductive numbers R_t.
- In multivariate models (vector notation),

$$\mathbb{E}(\mathrm{Y}_t) \mid \text{past} = \boldsymbol{\nu}_t + \boldsymbol{\Phi}_t \mathrm{Y}_{t-1},$$

  the largest eigenvalue of $\boldsymbol{\Phi}_t$ corresponds to R_t.
- Example: R_t of rotavirus in Berlin:

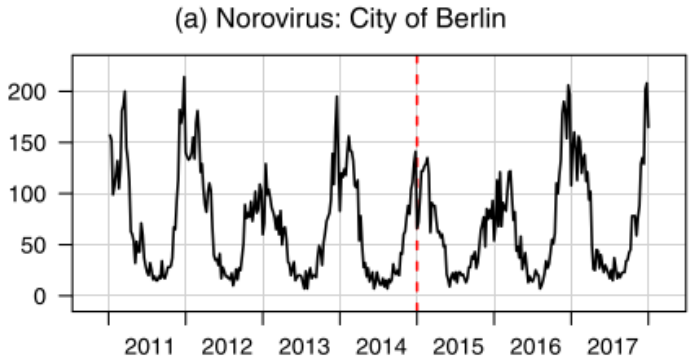


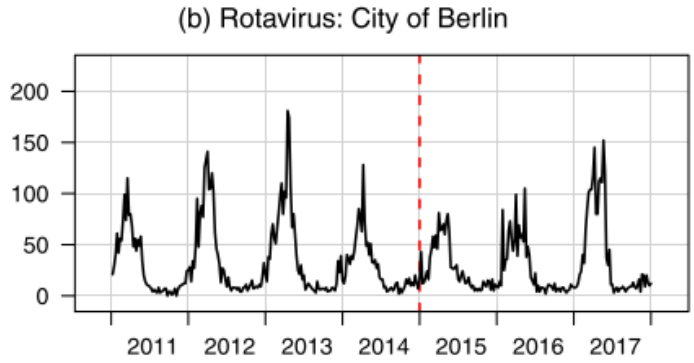
[9] Bracher and Held (2020): A Marginal Moment Matching Approach for Fitting Endemic-Epidemic Models to Underreported Disease Surveillance Counts. Biometrics.

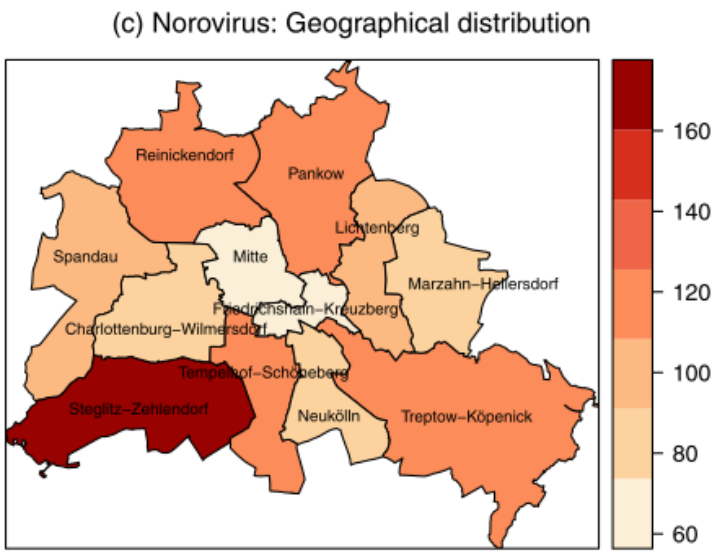
Tutorial

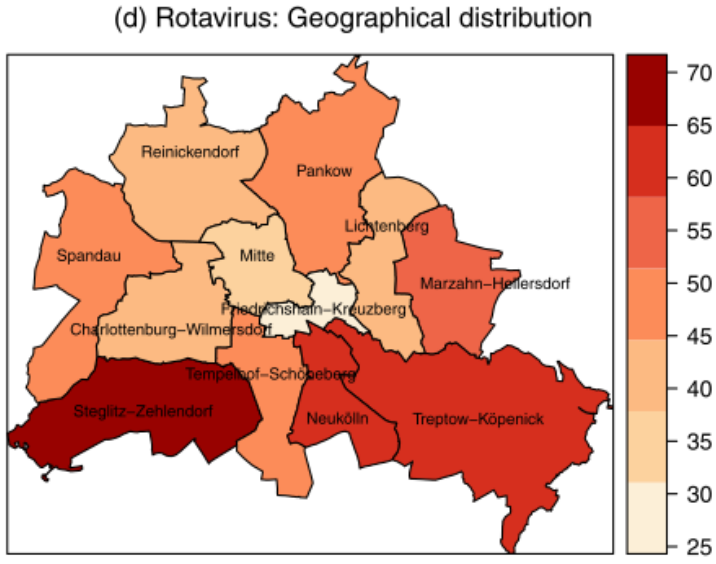
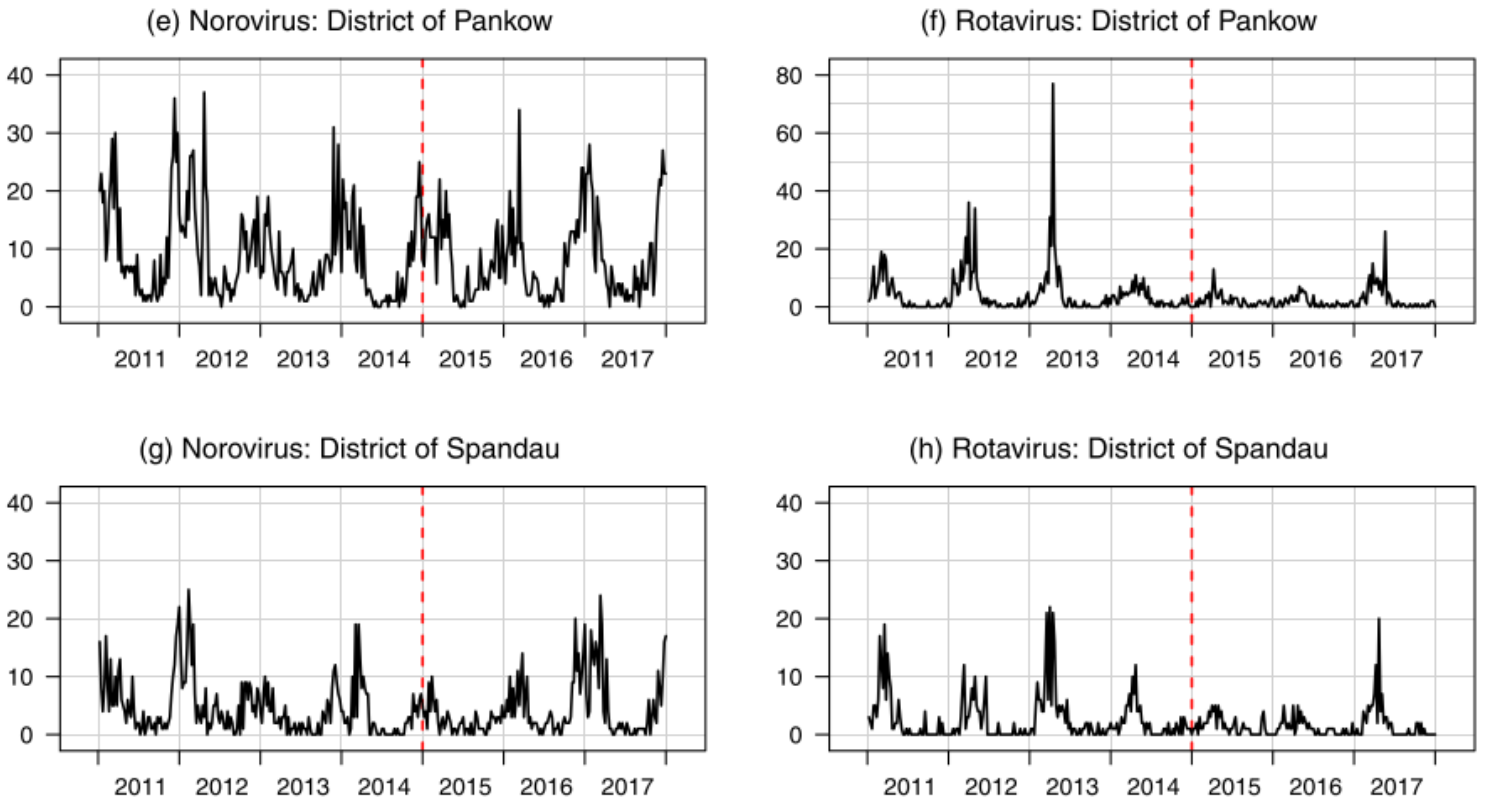
- We will now run through the development of a simple multivariate model.
- Head over to `https://codeberg.org/smeyer/hhh4geomed`.

Case study: Norovirus (and rotavirus) in Berlin









Case study: Norovirus (and rotavirus) in Berlin