Stochastic Modelling of the Spatial Spread of Influenza in Germany (Supporting Material)

Christiane Dargatz, Vera Georgescu, Leonhard Held

From the Standard Deterministic to a Global Stochastic SIR Model

Global SIR Model – Justification

This is a more detailed description of how we came to set up the differential equations in the global SIR model:

$$\frac{ds_i}{dt} = \underbrace{-\alpha s_i j_i}_{\substack{\text{individuals}\\\text{contaminate:}\\ S_i + I_i \xrightarrow{\alpha} 2I_i}} \underbrace{-\sum_k \gamma_{ik} s_i}_{\substack{\text{susceptibles}\\\text{leave region } i:\\ S_i \xrightarrow{\gamma_{ik}} S_k}} + \underbrace{\sum_k \gamma_{ki} s_k}_{\substack{\text{susceptibles}\\\text{enter region } i:\\ S_k \xrightarrow{\gamma_{ki}} S_i}} + \frac{1}{\sqrt{N_i}} \sqrt{\alpha s_i j_i}\, \xi_1^{(i)}(t)$$
$$+ \frac{1}{\sqrt{N_i}} \sqrt{\sum_k \gamma_{ik} s_i}\, \xi_4^{(i)}(t) - \frac{1}{\sqrt{N_i}} \sqrt{\sum_k \gamma_{ki} s_k}\, \xi_5^{(i)}(t)$$

$$\frac{dj_i}{dt} = \underbrace{\alpha s_i j_i}_{\substack{\text{individuals}\\\text{contaminate}\\ S_i + I_i \xrightarrow{\alpha} 2I_i}} \underbrace{-\beta j_i}_{\substack{\text{infecteds}\\\text{recover:}\\ I_i \xrightarrow{\beta} R_i}} \underbrace{-\sum_k \gamma_{ik} j_i}_{\substack{\text{infecteds}\\\text{leave region } i:\\ I_i \xrightarrow{\gamma_{ik}} I_k}} + \underbrace{\sum_k \gamma_{ki} j_k}_{\substack{\text{infecteds}\\\text{enter region } i:\\ I_k \xrightarrow{\gamma_{ki}} I_i}} - \frac{1}{\sqrt{N_i}} \sqrt{\alpha s_i j_i}\, \xi_1^{(i)}(t) + \frac{1}{\sqrt{N_i}} \sqrt{\beta j_i}\, \xi_2^{(i)}(t)$$
$$+ \frac{1}{\sqrt{N_i}} \sqrt{\sum_k \gamma_{ik} j_i}\, \xi_4^{(i)}(t) - \frac{1}{\sqrt{N_i}} \sqrt{\sum_k \gamma_{ki} j_k}\, \xi_5^{(i)}(t)$$

$$\frac{dr_i}{dt} = \underbrace{\beta j_i}_{\substack{\text{infecteds}\\\text{recover:}\\ I_i \xrightarrow{\beta} R_i}} - \frac{1}{\sqrt{N_i}} \sqrt{\beta j_i}\, \xi_2^{(i)}(t)$$

Global SLIR Model

The global extension of the SLIR model reads as follows:

$$S_i + I_i \xrightarrow{\alpha} L_i + I_i, \quad L_i \xrightarrow{\varepsilon} I_i, \quad I_i \xrightarrow{\beta} R \qquad \text{and} \qquad S_i \xrightarrow{\gamma_{ij}} S_j, \quad L_i \xrightarrow{\gamma_{ij}} L_j, \quad I_i \xrightarrow{\gamma_{ij}} I_j,$$

leading to

$$
\begin{aligned}
\frac{dl_i}{dt} &= \alpha s_i j_i - \varepsilon l_i - \sum_k \gamma_{ik} l_i + \sum_k \gamma_{ki} l_k - \frac{1}{\sqrt{N_i}}\sqrt{\alpha s_i j_i}\,\xi_1^{(i)}(t) + \frac{1}{\sqrt{N_i}}\sqrt{\varepsilon l_i}\,\xi_3^{(i)}(t) \\
&\quad + \frac{1}{\sqrt{N_i}}\sqrt{\sum_k \gamma_{ik} l_i}\,\xi_4^{(i)}(t) - \frac{1}{\sqrt{N_i}}\sqrt{\sum_k \gamma_{ki} l_k}\,\xi_5^{(i)}(t) \\
\frac{dj_i}{dt} &= \varepsilon l_i - \beta j_i - \sum_k \gamma_{ik} j_i + \sum_k \gamma_{ki} j_k - \frac{1}{\sqrt{N_i}}\sqrt{\varepsilon l_i}\,\xi_3^{(i)}(t) + \frac{1}{\sqrt{N_i}}\sqrt{\beta j_i}\,\xi_2^{(i)}(t) \\
&\quad + \frac{1}{\sqrt{N_i}}\sqrt{\sum_k \gamma_{ik} j_i}\,\xi_4^{(i)}(t) - \frac{1}{\sqrt{N_i}}\sqrt{\sum_k \gamma_{ki} j_k}\,\xi_5^{(i)}(t),
\end{aligned}
$$

$i = 1, \ldots, n$ (the derivatives of s_i and r_i remain unchanged).

Implementation

Keeping the System Closed

In Section 3.1, we claimed

$$
\text{var}\left(\sum_{i=1}^{n} x_i(t)\,\xi_4^{(i)}(t)\right) = 0.
$$

Justification:

$$
\begin{aligned}
\text{var}\left(\sum_{i=1}^{n} x_i(t)\,\xi_4^{(i)}(t)\right) &= \sum_{i=1}^{n} \text{var}\big(x_i(t)\,\xi_4^{(i)}(t)\big) + \sum_{1\le i,j\le n, i\ne j} \text{cov}\left(x_i(t)\,\xi_4^{(i)}(t), x_j(t)\,\xi_4^{(j)}(t)\right) \\
&= \sum_{i=1}^{n} q_{ii}(t) + \sum_{1\le i,j\le n, i\ne j} q_{ij}(t) = \mathbf{1_n'}\mathbf{Q}(t)\mathbf{1_n} \overset{*}{=} 0,
\end{aligned}
$$

where equality $(*)$ holds due to

$$
\begin{aligned}
\sum_{i=1}^{n} q_{ii}(t) + \sum_{1\le i,j\le n, i\ne j} q_{ij}(t) &= \frac{1}{n^2}\sum_i \sum_{k\ne i} x_k^2(t) + \Big(\frac{n-1}{n}\Big)^2 \sum_i x_i^2(t) \\
&\qquad + \frac{1}{n^2}\sum_{(i,j), i\ne j}\ \sum_{k\ne i,j} x_k^2(t) - \frac{n-1}{n^2}\sum_{(i,j), i\ne j}\big(x_i^2(t) + x_j^2(t)\big) \\
&= \frac{n-1}{n^2}\sum_{i=1}^{n} x_i^2(t) + \Big(\frac{n-1}{n}\Big)^2 \sum_{i=1}^{n} x_i^2(t) \\
&\qquad + \frac{(n-1)(n-2)}{n^2}\sum_{i=1}^{n} x_i^2(t) + 2\Big(\frac{n-1}{n}\Big)^2 \sum_{i=1}^{n} x_i^2(t) \\
&= \frac{n-1}{n^2}\left[1 + (n-1) + (n-2) - 2(n-1)\right]\sum_{i=1}^{n} x_i^2(t) \\
&= 0.
\end{aligned}
$$

Algorithm

This presents the algorithm from Section 3.3, using the SLIR model. For $m = 0, \ldots, \lfloor t_{\max}/\delta \rfloor - 1$, repeat the following steps:

1. For $i = 1, \ldots, n$, calculate

$$\mu_i := \alpha\, s_i(t_m)\, j_i(t_m), \quad \nu_i := \beta\, j_i(t_m), \quad \varphi_i := \varepsilon\, l_i(t_m),$$

   and

$$\eta_i := \sum_{k=1}^{n} \gamma_{ik}\, s_i(t_m), \qquad \kappa_i := \sum_{k=1}^{n} \gamma_{ik}\, l_i(t_m), \qquad \rho_i := \sum_{k=1}^{n} \gamma_{ik}\, j_i(t_m),$$
$$\zeta_i := \sum_{k=1}^{n} \gamma_{ki}\, s_k(t_m), \qquad \lambda_i := \sum_{k=1}^{n} \gamma_{ki}\, l_k(t_m), \qquad \tau_i := \sum_{k=1}^{n} \gamma_{ki}\, j_k(t_m),$$

   where $t_m = m\delta$.

2. For $i = 1, \ldots, n$, compute $x_i = m_i(\sqrt{\eta_i} + \sqrt{\kappa_i} + \sqrt{\rho_i}\,)$ and $y_i = m_i(\sqrt{\zeta_i} + \sqrt{\lambda_i} + \sqrt{\tau_i}\,)$, where $m_i := \sqrt{N_i}^{-1}$.

3. Set $\mathbf{P_4} = \operatorname{diag}\big(x_1^2, \ldots, x_{n-1}^2\big) + x_n^2 \mathbf{1_{n-1}} \mathbf{1'_{n-1}}$ and $\mathbf{P_5} = \operatorname{diag}\big(y_1^2, \ldots, y_{n-1}^2\big) + y_n^2 \mathbf{1_{n-1}} \mathbf{1'_{n-1}}$.

4. Generate $\boldsymbol{\pi}^{\mathbf{(j)}} = \big(\pi_1^{(j)}, \ldots, \pi_n^{(j)}\big)'$, $j = 4, 5$, with $\big(\pi_1^{(j)}, \ldots, \pi_{n-1}^{(j)}\big)' \sim N(\mathbf{0}, \mathbf{P_j})$ and $\pi_n^{(j)} = 0$.

5. Compute $\mathbf{u} = (u_1, \ldots, u_n)' = \mathbf{M}\boldsymbol{\pi}^{\mathbf{(4)}}$ and $\mathbf{v} = (v_1, \ldots, v_n)' = \mathbf{M}\boldsymbol{\pi}^{\mathbf{(5)}}$ with $\mathbf{M} = \mathbf{I_n} - n^{-1}\mathbf{1_n}\mathbf{1'_n}$.

6. Evaluate $\boldsymbol{\xi_1}, \boldsymbol{\xi_2} \sim N(\mathbf{0}, \mathbf{I_n})$ and $\boldsymbol{\xi_4}, \boldsymbol{\xi_5}$ with $\xi_4^{(i)} = u_i/x_i$, $\xi_5^{(i)} = v_i/y_i$, $i = 1, \ldots, n$.

7. For $i = 1, \ldots, n$, calculate

$$\begin{array}{llll}
a_s(i) = -\mu_i - \eta_i + \zeta_i & a_l(i) = \mu_i - \varphi_i - \kappa_i + \lambda_i & a_j(i) = \varphi_i - \nu_i - \rho_i + \tau_i & a_r(i) = \nu_i \\
b_{s1}(i) = m_i\sqrt{\mu_i} & b_{l1}(i) = -m_i\sqrt{\mu_i} & b_{j1}(i) = 0 & b_{r1}(i) = 0 \\
b_{s2}(i) = 0 & b_{l2}(i) = 0 & b_{j2}(i) = m_i\sqrt{\nu_i} & b_{r2}(i) = -m_i\sqrt{\nu_i} \\
b_{s3}(i) = 0 & b_{l3}(i) = m_i\sqrt{\varphi_i} & b_{j3}(i) = -m_i\sqrt{\varphi_i} & b_{r3}(i) = 0 \\
b_{s4}(i) = m_i\sqrt{\eta_i} & b_{l4}(i) = m_i\sqrt{\kappa_i} & b_{j4}(i) = m_i\sqrt{\rho_i} & b_{r4}(i) = 0 \\
b_{s5}(i) = -m_i\sqrt{\zeta_i} & b_{l5}(i) = -m_i\sqrt{\lambda_i} & b_{j5}(i) = -m_i\sqrt{\tau_i} & b_{r5}(i) = 0.
\end{array}$$

8. Approximate $s_i(t_{m+1})$, $l_i(t_{m+1})$, $j_i(t_{m+1})$, and $r_i(t_{m+1})$, $i = 1, \ldots, n$, with the Euler-Maruyama formula.

9. For $i = 1, \ldots, n$, correct approximation errors by setting negative values of s_i, l_i, j_i, and r_i equal to zero.

10. (Optional step.) Rescale s_i, l_i, j_i, and r_i, $i = 1, \ldots, n$, via

$$s_i(t_{m+1}) \leftarrow s_i(t_{m+1}) \cdot \big(s_i(t_{m+1}) + l_i(t_{m+1}) + j_i(t_{m+1}) + r_i(t_{m+1})\big)^{-1}$$
$$l_i(t_{m+1}) \leftarrow l_i(t_{m+1}) \cdot \big(s_i(t_{m+1}) + l_i(t_{m+1}) + j_i(t_{m+1}) + r_i(t_{m+1})\big)^{-1}$$
$$j_i(t_{m+1}) \leftarrow j_i(t_{m+1}) \cdot \big(s_i(t_{m+1}) + l_i(t_{m+1}) + j_i(t_{m+1}) + r_i(t_{m+1})\big)^{-1}$$
$$r_i(t_{m+1}) \leftarrow r_i(t_{m+1}) \cdot \big(s_i(t_{m+1}) + l_i(t_{m+1}) + j_i(t_{m+1}) + r_i(t_{m+1})\big)^{-1}.$$

Initialization

Connectivity Matrix

In this subsection we want to fix the parameter γ, whose entries stand for the amount of traffic between the parts of Germany. Its installment certainly depends on the considered resolution (districts, counties, or states). In any case, the matrix is composed of the three components "neighbours", "trains", and "flights", standing for the exchange between adjacent regions, by long distance trains, and by airplane, respectively.

Let us first focus on the traffic between districts: When considering the traffic between adjacent regions, we assume the attraction of a district to be proportional to its population density because it is usually the big cities which appeal commuters, customers, and tourists. We set the strength of the connection between two districts with population densities d_1 and d_2 and n_1 and n_2 surrounding districts to $d_1/n_1 + d_2/n_2$. In the train network model, we include 57 cities which are served by ICE trains. Here, the importance of a connection depends on whether two stations are located in the same, in neighbouring or in more distant states (weights 4, 2, and 1). Like the commuter traffic, the train network is hence presumed to be symmetric. The data for the flight network was collected from OAGflights (`http://www.oagflights.com`). Like Hufnagel et al., we counted the daily amount M_{ij} of travellers from airport i to j via the available capacity of seats. The weight for the connection from i to j is then calculated as $M_{ij}/\sum_i M_{ij}$.

On the county and state level, the migration between regions are more homogeneous since we do not distinguish between cities and suburbs. For both counties and states, we simply attach all neighbourhoods with uniform weights and increase the influence of the rail and flight networks to $1/4$ and $1/16$ in comparison to the exchange between adjacent regions.

In the paper the overall influence of γ was estimated to be of magnitude 0.16 when considering the subdivision of Germany into 438 districts (without the island Rügen), i.e. the components of γ are scaled such that the average row total amounts 0.16. In case of counties and states, we provide γ with the weights 0.014 and 0.006, according to their lower numbers, which are 40 and 16, respectively. (Actually, today there are less than 40 counties since some of them were merged; but since the data is available for the old structure, we still use this division.)