# Stochastic Modelling of the Spatial Spread of Influenza in Germany

Christiane Dargatz, Leonhard Held

Department of Statistics
Ludwig-Maximilians-University Munich

Financial support by the German Research Foundation (DFG), SFB 386

Hannover, November 2005

# Influenza

- Worldwide one of the most common and severe infectious diseases.
- Major epidemics and pandemics of the 20th century: Spanish Flu (1918-20), Asian Flu (1957-58), Hong Kong Flu (1969)
- Annual number of deaths caused by influenza in Germany is twice as high as those caused by road accidents, nevertheless low vaccination rates.
- Steadily new antigen mutants of the influenza virus coming up.
- In the opinion of experts the next pandemic is just a question of time (caused e.g. by avian flu).

⇒ Emergency plans are essential.

# Motivation

- Mathematical modelling of the spread of epidemics dates back to 1760 (Daniel Bernoulli) and has been well developed since then.
- Globalization has induced "new era" of epidemiology: People travel more frequently, faster and further than in former times.
- Hufnagel et al. (2004) have successfully simulated the world-wide spatial spread of the SARS epidemic in 2003 using a spatial SIR model based on global air traffic data.

# Outline

# Standard SIR Model

- Divide population of size $N$ into susceptible ($S$), infected ($I$), and removed ($R$) individuals.
- Transitions:

$$S + I \xrightarrow{\alpha} 2I, \quad I \xrightarrow{\beta} R$$

$\alpha$: contact rate of an infectious individual sufficient to spread the disease,
$\beta$: reciprocal average infectious period.

- Infection dynamics:

$$ds/dt = -\alpha sj, \quad dj/dt = \alpha sj - \beta j, \qquad (1)$$

$s = S/N$, $j = I/N$, $r = R/N = 1 - s - j$.

# Standard SIR Model (cont.)

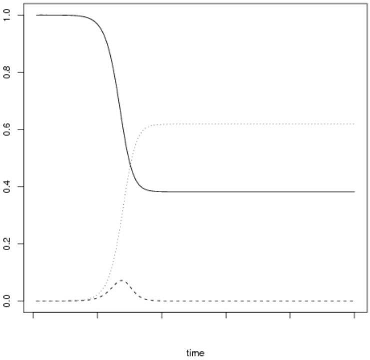
Crucial parameter: the basic reproduction number $\rho = \alpha/\beta$, *the average number of persons directly infected by an infectious case during its entire infectious period after entering a totally susceptible population.*

$\rho^{-1} > s(0)$: no epidemic will occur
$\rho^{-1} > s(t)$: epidemic decays,

i.e. major epidemic will occur when, in the early stages of an outbreak, each infective on average produces more than one further infective.

# Standard SIR Model (cont.)

Evolution of proportions of susceptible (solid), infected (dashed), and removed (dotted) individuals in the standard SIR model; $\rho = 1.5$.



# Stochastic SIR Model

Infection and recovery processes are of rather stochastic than deterministic character.

⇒ Write (1) in terms of Langevin equations:

$$\begin{aligned}
\frac{ds}{dt} &= -\alpha sj + \frac{1}{\sqrt{N}}\sqrt{\alpha sj}\,\xi_1(t) \\
\frac{dj}{dt} &= \alpha sj - \beta j - \frac{1}{\sqrt{N}}\sqrt{\alpha sj}\,\xi_1(t) + \frac{1}{\sqrt{N}}\sqrt{\beta j}\,\xi_2(t),
\end{aligned}$$

where $\xi_1(t)$ and $\xi_2(t)$ are independent Gaussian white noise forces, modelling fluctuations in disease transmission and recovery (Hufnagel et al., 2004).

# Spatial SIR Model

**Problem:** Assumption of homogeneous mixing is not given in our fully connected world anymore!

**Idea:** Introduce network of subregions $i = 1, \dots, n$ of sizes $N_i$.

- Local infection dynamics within a subregion is given by stochastic SIR model as before:

$$S_i + I_i \xrightarrow{\alpha} 2I_i, \quad I_i \xrightarrow{\beta} R_i.$$

- Global dispersal between nodes of network is rated in a connectivity matrix $\gamma = (\gamma_{ij})_{ij}$:

$$S_i \xrightarrow{\gamma_{ij}} S_j, \quad I_i \xrightarrow{\gamma_{ij}} I_j.$$

# Spatial SIR model (cont.)

The system of stochastic differential equations now changes to

$$
\begin{aligned}
\frac{ds_i}{dt} &= -\alpha s_i j_i - \sum_k \gamma_{ik} s_i + \sum_k \gamma_{ki} s_k + \frac{1}{\sqrt{N_i}} \sqrt{\alpha s_i j_i}\ \xi_1^{(i)}(t) \\
&\quad + \frac{1}{\sqrt{N_i}} \sqrt{\sum_k \gamma_{ik} s_i}\ \xi_4^{(i)}(t) - \frac{1}{\sqrt{N_i}} \sqrt{\sum_k \gamma_{ki} s_k}\ \xi_5^{(i)}(t) \\
\frac{dj_i}{dt} &= \alpha s_i j_i - \beta j_i - \sum_k \gamma_{ik} j_i + \sum_k \gamma_{ki} j_k - \frac{1}{\sqrt{N_i}} \sqrt{\alpha s_i j_i}\ \xi_1^{(i)}(t) + \frac{1}{\sqrt{N_i}} \sqrt{\beta j_i}\ \xi_2^{(i)}(t) \\
&\quad + \frac{1}{\sqrt{N_i}} \sqrt{\sum_k \gamma_{ik} j_i}\ \xi_4^{(i)}(t) - \frac{1}{\sqrt{N_i}} \sqrt{\sum_k \gamma_{ki} j_k}\ \xi_5^{(i)}(t) \\
\frac{dr_i}{dt} &= \beta j_i - \frac{1}{\sqrt{N_i}} \sqrt{\beta j_i}\ \xi_2^{(i)}(t).
\end{aligned}
$$

for $i = 1, \ldots, n$, where $\boldsymbol{\xi_1(t)}$, $\boldsymbol{\xi_2(t)}$, $\boldsymbol{\xi_4(t)}$, and $\boldsymbol{\xi_5(t)}$ denote independent vector-valued white noise forces standing for fluctuations in transmission, recovery, and outbound and inbound traffic, respectively.

# Keeping the System Closed

Area of $n$ regions is assumed to be closed. i.e. we have to require

$$\sum_{i=1}^{n} \left( \frac{ds_i}{dt} + \frac{dj_i}{dt} + \frac{dr_i}{dt} \right) = 0 \ .$$

$\Rightarrow$ Introduce a weak form of dependence to the white noise forces such that the above equality holds almost surely.

# Numerical Scheme

Define functions $a_j$ and $b_{jk}$, $1 \leq j \leq 3$, $1 \leq k \leq 5$, such that

$$ds_i(t) = a_1(t, s_i(t))\,dt + \sum_{k=1}^{5} b_{1k}(t, s_i(t))\,dW_k^{(i)}(t)$$

$$dj_i(t) = a_2(t, j_i(t))\,dt + \sum_{k=1}^{5} b_{2k}(t, j_i(t))\,dW_k^{(i)}(t)$$

$$dr_i(t) = a_3(t, r_i(t))\,dt + \sum_{k=1}^{5} b_{3k}(t, r_i(t))\,dW_k^{(i)}(t).$$

# Numerical Scheme (cont.)

For example, for $i = 1, \dots, n$:

$$ds_i = \underbrace{\left(-\alpha s_i j_i - \sum_k \gamma_{ik} s_i + \sum_k \gamma_{ki} s_k\right)}_{:=a_1(t,s_i(t))} dt + \underbrace{\frac{1}{\sqrt{N_i}} \sqrt{\alpha s_i j_i}}_{:=b_{11}(t,s_i(t))} \xi_1^{(i)}(t)dt$$

$$+ \underbrace{\frac{1}{\sqrt{N_i}} \sqrt{\sum_k \gamma_{ik} s_i}}_{:=b_{14}(t,s_i(t))} \xi_4^{(i)}(t)dt - \underbrace{\frac{1}{\sqrt{N_i}} \sqrt{\sum_k \gamma_{ki} s_k}}_{:=b_{15}(t,s_i(t))} \xi_5^{(i)}(t)dt,$$

$$b_{12}(t, s_i(t)) = b_{13}(t, s_i(t)) = 0, \ \xi_k^{(i)}(t)dt = dW_k^{(i)}(t).$$

# Numerical Scheme (cont.)

Apply Euler-Maruyama approximation scheme to numerically solve the system of SDEs at discrete, equidistant instants $0, \delta, 2\delta, \ldots$ in the time domain:

$$s_i(m\delta) = s_i((m-1)\delta) + a_1((m-1)\delta, s_i((m-1)\delta))\,\delta + \sum_{k=1}^{5} b_{1k}((m-1)\delta, s_i((m-1)\delta))\,\triangle W_k^{(i)}(m)$$

$$j_i(m\delta) = j_i((m-1)\delta) + a_2((m-1)\delta, j_i((m-1)\delta))\,\delta + \sum_{k=1}^{5} b_{2k}((m-1)\delta, j_i((m-1)\delta))\,\triangle W_k^{(i)}(m)$$

$$r_i(m\delta) = r_i((m-1)\delta) + a_3((m-1)\delta, r_i((m-1)\delta))\,\delta + \sum_{k=1}^{5} b_{3k}((m-1)\delta, r_i((m-1)\delta))\,\triangle W_k^{(i)}(m)$$

for $m \geq 1$, $i = 1, \ldots, n$ and $\triangle W_k^{(i)}(m) := W_k^{(i)}(m\delta) - W_k^{(i)}((m-1)\delta)$.

# Application: Influenza in Germany

- Consider the 438 districts of Germany.
- Data about incidences of influenza taken from Robert Koch Institute.
- Build up connectivity matrix to describe the strength between parts of Germany, considering
  - dispersal between adjacent regions, caused e.g. by commuters (info from Federal Statistical Office Germany),
  - domestic train traffic (ICE),
  - domestic air traffic (www.oagflights.com),

  where each of these components is provided with a weight regulating its influence.

# Parameter choice

- We assume that the basic reproduction number $\rho = \alpha/\beta$ depends on population density:

$$\rho(d_i) = 1.0179 + 10^{-5} \cdot d_i,$$

where $d_i$ is the population density of region $i$. (Disease is more likely to spread in areas with high population densities.)

- Infectious period of influenza: 4-5 days, hence we choose $\beta = 2/9$.
- We obtain the contact rates $\alpha_i$ via $\beta \cdot \rho(d_i)$, $i = 1, \dots, n$.

# Simulation

1. Simulation 1:
    - Starting values based on surveillance data from week 5/2005.
    - Animation shows proportion of infectives and time trend.
    - Surprisingly good agreement with actual course of the influenza epidemic 2005.
2. Simulation 2:
    - Artificial starting values in three districts.
    - Simulation based on three different connectivity matrices:
        1. local, train and air,
        2. local and train,
        3. only local.

# Conclusion

- Spatial extension of classical SIR model.
- Implementation of certain sum-to-zero contraints and numerical approximation.
- Parameter choice based on external knowledge.
- Simulation of the spread of an influenza epidemic in Germany.

# Construction Sites

- Prevalence data of influenza is highly underreported.
    - Simulate underreporting.
    - Other data sources (Sentinella).
    - Consider different diseases.
- Model does not take into account important factors like
    - weather/temperature,
    - seasonal holiday travel,
    - measures undertaken to lower transmission rate.

# Future Work

Main purposes will be

- finding surveillance strategies in case of a sudden outbreak of an epidemic (isolation, vaccination, observation of migration),
- more formal statistical inference on model parameters based on available data from surveillance databases.

# References


- Dargatz, C., Georgescu, V. and Held, L. (2005) Stochastic Modelling of the Spatial Spread of Influenza in Germany. Technical Report, Munich University.
- Hufnagel, L., Brockmann, D. and Geisel, T. (2004). Forecast and Control of Epidemics in a Globalized World. *Proceedings of the National Academy of Sciences*, **101**, 15124-15129.