Spatio-Temporal Analysis of Epidemic Phenomena Using the R Package \textit{surveillance}

Sebastian Meyer

\begin{verbatim}
> sessionInfo()
[1] "June 30 - July 3, 2015"
[2] "Aalborg, Denmark"
\end{verbatim}
Epidemic phenomena

Examples:
- Earth quakes
- Riots / crimes
- Infectious diseases

Data: Surveillance systems routinely collect
- time-stamped
- geo-referenced

case reports
Case study I: Invasive meningococcal disease

library("surveillance"); data("imdepi")

plot(imdepi, "space")

plot(imdepi, "time")

Dot size proportional to the number of cases (residence postcode)

Monthly and cumulative number of cases (by date of specimen sampling)
 animate(subset(imdepi, type="B"),
   time.spacing = 7)

 animate(subset(imdepi, type="C"),
   time.spacing = 7)

Does the force of infection depend on the bacterial finetype?
Case study II: Measles

Publically available surveillance data:

time series of counts of newly reported infections by district

```r
plot(measlesWeserEms, type = observed ~ unit)
plot(measlesWeserEms, type = observed ~ time)
```
Is local vaccination coverage related to disease dynamics?
Characteristics of epidemic-type data

- Low number of cases
- Seasonality
- Occassional outbreaks ("self-exciting" process)
- Dependence between areas, age groups, etc.
- Underreporting, reporting delays
Characteristics of epidemic-type data

- Low number of cases
- Seasonality
- Occasional outbreaks (“self-exciting” process)
- Dependence between areas, age groups, etc.
- Underreporting, reporting delays

Aims of surveillance

Monitoring (prospective): Outbreak prediction and detection
(→ “Zombie Preparedness” talk by Michael Höhle)

Modelling (retrospective): Quantify epidemicity and effects of external covariates on disease dynamics
**Place in the world of R packages**

`surveillance` is the first and only software package dedicated to the **space-time modelling and monitoring of epidemic phenomena**

**Related packages:**

- **spacetime**: Basic classes and methods for spatio-temporal data
- **spatstat**: THE package for **purely spatial** point patterns
- **tscout**, **EpiEstim**, **outbreaker**, **amei**: Several packages dealing with **purely temporal** epidemic data
- **stpp**: Simulation & visualization of space-time point patterns

For a more complete picture:

→ CRAN task view “Handling and Analyzing Spatio-Temporal Data”
## Three modelling frameworks in surveillance

<table>
<thead>
<tr>
<th>Data Resolution</th>
<th>Example</th>
<th>Model</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>individual events in continuous space-time</td>
<td>cases of invasive meningococcal disease (IMD) Meyer et al., 2012</td>
<td>spatio-temporal point process</td>
<td>twinstim()</td>
</tr>
<tr>
<td>event counts aggregated in space &amp; time</td>
<td>week × district counts of measles Meyer et al., 2014</td>
<td>multivariate NegBin time series</td>
<td>hhh4()</td>
</tr>
<tr>
<td>individual SIR event history of a fixed population</td>
<td>spread of classical swine fever among domestic pig farms Höhle, 2009</td>
<td>multivariate temporal point process</td>
<td>twinSIR()</td>
</tr>
</tbody>
</table>
Basic modelling concept
Stochastic branching process with immigration

Decomposed disease risk:
- **Endemic**: seasonality, population, socio-demography, ...
- **Epidemic**: force of previously infected individuals

- Ebola: $R_0$ of about 1.5 – 2.5
- Force of infection may depend on age and spatial/temporal distance to infective
Spatial interaction

Tobler’s First Law of Geography:

*Everything is related to everything else, but near things are more related than distant things.*

Brockmann et al., 2006 (dollar bill tracking):

*The distribution of travelling distances decays as a power law.*

\[ f(x) = x^{-1.6} \]
\[ \log(f(x)) = -1.6 \cdot \log(x) \]
Case study I: Invasive meningococcal disease
Regression framework for the conditional intensity function

\[ \lambda(\mathbf{s}, t) = \rho[\mathbf{s}][t] \nu[\mathbf{s}][t] \]

Endemic component

- Piecewise constant on a suitable space-time grid
- Explanatory variables in a log-linear predictor \( \nu[\mathbf{s}][t] \)
- Equivalent to Poisson-GLM for aggregated counts
Case study I: Invasive meningococcal disease
Regression framework for the conditional intensity function

\[
\lambda(s, t) = \rho[s][t] \nu[s][t] + \sum_{j: t_j < t} \eta_j f(\|s - s_j\|) g(t - t_j)
\]

**Endemic component**
- Piecewise constant on a suitable space-time grid
- Explanatory variables in a log-linear predictor \( \nu[s][t] \)
- Equivalent to Poisson-GLM for aggregated counts

**Force of infection**
- Depends on event-specific characteristics \( m_j \) via
  \[
  \log(\eta_j) = \gamma_0 + \gamma^T m_j
  \]
- Decays over space/time according to parametric interaction function \( f(\cdot)/g(\cdot) \)

Likelihood inference
- \( \text{nlminb() with analytical score function} \)
- \( \text{R package polyCub for cubature of } f(\|s\|) \) over polygons
Case study I: Invasive meningococcal disease

Regression framework for the conditional intensity function

\[ \lambda(s, t) = \rho[s][t] \nu[s][t] + \sum_{j: t_j < t} \eta_j f(\|s - s_j\|) g(t - t_j) \]

**Endemic component**

- Piecewise constant on a suitable space-time grid
- Explanatory variables in a log-linear predictor \( \nu[s][t] \)
- Equivalent to Poisson-GLM for aggregated counts

**Force of infection**

- Depends on event-specific characteristics \( m_j \) via
  \[ \log(\eta_j) = \gamma_0 + \gamma^\top m_j \]
- Decays over space/time according to parametric interaction function \( f(\cdot)/g(\cdot) \)

**Likelihood inference**

- `nlminb()` with analytical score function and Fisher info
- R package `polyCub` for cubature of \( f(\|s\|) \) over polygons
Case study I: Invasive meningococcal disease

Model estimation

```r
imdfit <- twinstim(
  endemic = ~ offset(log(popdensity)) + I(start/365 - 3.5) +
    sin(2 * pi * start/365) + cos(2 * pi * start/365),
  epidemic = ~ type + agegrp,
  siaf = siaf.powerlaw(), tiaf = tiaf.constant(),
  data = imdepi, subset = !is.na(agegrp),
  start = c("e.(Intercept")=-6.5, "e.siaf.1"=1.5, "e.siaf.2"=0.9),
  optim.args = list(fixed = "e.siaf.1"), model = TRUE, cores = 4)
```

<table>
<thead>
<tr>
<th>RR</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>h.I(start/365 - 3.5)</td>
<td>0.959</td>
<td>0.92–1.00</td>
</tr>
<tr>
<td>h.sin(2 * pi * start/365)</td>
<td>1.231</td>
<td>1.08–1.41</td>
</tr>
<tr>
<td>h.cos(2 * pi * start/365)</td>
<td>1.379</td>
<td>1.21–1.57</td>
</tr>
<tr>
<td>e.typeC</td>
<td>0.450</td>
<td>0.27–0.74</td>
</tr>
<tr>
<td>e.agegrp[3,19)</td>
<td>2.133</td>
<td>1.10–4.12</td>
</tr>
<tr>
<td>e.agegrp[19,Inf)</td>
<td>0.824</td>
<td>0.33–2.05</td>
</tr>
</tbody>
</table>

### Case study I: Invasive meningococcal disease

#### Model estimation

```r
imdfit <- twinstim(
  endemic = ~offset(log(popdensity)) + I(start/365 - 3.5) +
  sin(2 * pi * start/365) + cos(2 * pi * start/365),
  epidemic = ~type + agegrp,
  siaf = siaf.powerlaw(), tiaf = tiaf.constant(),
  data = imdepi, subset = !is.na(agegrp),
  start = c("e.(Intercept)"=-6.5, "e.siaf.1"=1.5, "e.siaf.2"=0.9),
  optim.args = list(fixed = "e.siaf.1"), model = TRUE, cores = 4)

xtable(imdfit)
```

<table>
<thead>
<tr>
<th></th>
<th>RR</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>h.I(start/365 - 3.5)</td>
<td>0.959</td>
<td>0.92–1.00</td>
<td>0.071</td>
</tr>
<tr>
<td>h.sin(2 * pi * start/365)</td>
<td>1.231</td>
<td>1.08–1.41</td>
<td>0.0022</td>
</tr>
<tr>
<td>h.cos(2 * pi * start/365)</td>
<td>1.379</td>
<td>1.21–1.57</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>e.typeC</td>
<td>0.450</td>
<td>0.27–0.74</td>
<td>0.0017</td>
</tr>
<tr>
<td>e.agegrp[3,19)</td>
<td>2.133</td>
<td>1.10–4.12</td>
<td>0.024</td>
</tr>
<tr>
<td>e.agegrp[19,Inf)</td>
<td>0.824</td>
<td>0.33–2.05</td>
<td>0.68</td>
</tr>
</tbody>
</table>
Case study I: Invasive meningococcal disease

Estimated spatial interaction

```r
plot(imdfit, which = "siaf",
     xlim = c(0, 50))

imdfit_fstep <- update(imdfit,
     siaf = siaf.step(
         knots = exp((1:4)*log(100)/5),
         maxRange = 100),
     optim.args = list(fixed = NULL))
plot(imdfit_fstep, which = "siaf",
     add = TRUE, col.estimate = 1)
```
Case study I: Invasive meningococcal disease

Estimated spatial interaction

```r
plot(imdfit, which = "siaf",
xlim = c(0, 50))

imdfit_fstep <- update(imdfit,
siaf = siaf.step(
knots = exp((1:4)*log(100)/5),
maxRange = 100),
optim.args = list(fixed = NULL))
plot(imdfit_fstep, which = "siaf",
add = TRUE, col.estimate = 1)
```

Predefined interaction functions:

<table>
<thead>
<tr>
<th>Spatial (siaf.*)</th>
<th>Temporal (tiaf.*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>constant</td>
</tr>
<tr>
<td>gaussian</td>
<td>exponential</td>
</tr>
<tr>
<td>powerlaw</td>
<td>step</td>
</tr>
<tr>
<td>powerlawL</td>
<td></td>
</tr>
<tr>
<td>step</td>
<td></td>
</tr>
<tr>
<td>student</td>
<td></td>
</tr>
</tbody>
</table>

Distance $x$ from host
$e^{-\gamma 0 \cdot f(x)}$

Power law
Step (df=4)
Case study I: Invasive meningococcal disease

Fitted ground intensity \( \int \hat{\lambda}(s, t) \, ds \)

\[
\text{plot(imdfit, which = "total intensity", aggregate = "time", types = 1, ylim = c(0,0.3), tgrid = 2500)}
\]
### Case study I: Invasive meningococcal disease

#### Methods for "twinstim"

<table>
<thead>
<tr>
<th>Display</th>
<th>Extract</th>
<th>Modify</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>print</td>
<td>nobs</td>
<td>update</td>
<td>simulate</td>
</tr>
<tr>
<td>summary</td>
<td>vcov</td>
<td>add1</td>
<td>coeflist</td>
</tr>
<tr>
<td>xtable</td>
<td>logLik</td>
<td>drop1</td>
<td></td>
</tr>
<tr>
<td>plot</td>
<td>extractAIC</td>
<td>stepComponent</td>
<td></td>
</tr>
<tr>
<td>intensityplot</td>
<td>profile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iafplot</td>
<td>residuals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>checkResidualProcess</td>
<td>terms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Case study II: Measles (areal count time series)

Regression framework

Number of cases in region $r$ at time $t$

$$Y_{rt} \mid Y_{.,t-1} \sim \text{NegBin}(\mu_{rt}, \psi)$$

Endemic-Epidemic decomposition of disease risk:

$$\mu_{rt} = e_{rt} \nu_{rt} + \lambda_{rt} Y_{r,t-1} + \phi_{rt} \sum_{s \neq r} w_{sr} Y_{s,t-1}$$

- $e_{rt}$: population offset
- $\nu_{rt}, \lambda_{rt}, \phi_{rt}$: log-linear predictors, e.g., vaccination coverage
- $w_{sr}$: weight for $s$ to $r$ transmission, e.g., $w_{sr} = o_{sr}^{-d}$
Case study II: Measles (areal count time series)
Model estimation

```r
( endemic <- addSeason2formula(~log(pSusceptible) + t))

## ~log(pSusceptible) + t + sin(2 * pi * t/52) + cos(2 * pi * t/52)

measlesModel <- list(
  end = list(f = endemic, offset = population(measlesWeserEms)),
  ar = list(f = ~1),
  ne = list(f = ~1, weights = W_powerlaw(maxlag = 5)),
  family = "NegBin1", data = list(pSusceptible = 1 - pVacc))

measlesFit <- hhh4(measlesWeserEms, control = measlesModel)
```
Case study II: Measles (areal count time series)
Fitted mean components

plot(measlesFit, type = "fitted", units = c(7,12), hide0s = TRUE)
Case study II: Measles (areal count time series)
Association with vaccination coverage $v_r$

Endemic incidence is proportional to $(1 - v_r)^{\beta_v}$:

$$\text{summary(measlesFit)$fixef["end.log(pSusceptible)" ,]}$$

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.054</td>
<td>0.379</td>
</tr>
</tbody>
</table>
Case study II: Measles (areal count time series)
Association with vaccination coverage $v_r$

Endemic incidence is proportional to $(1 - v_r)^{\beta v}$:

```r
summary(measlesFit)$fixef["end.log(pSusceptible)", ]
```

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.054</td>
<td>0.379</td>
</tr>
</tbody>
</table>

Other methods:
`update()`, `simulate()`, `oneStepAhead()`, ...
Conclusion

**surveillance** offers a comprehensive framework for the **spatio-temporal analysis of epidemic phenomena**, including visualisation, modelling, inference and simulation of:

- (multivariate) surveillance time series
- spatio-temporal point patterns
- geo-referenced SIR event histories

**Key references:**

- [http://surveillance.r-forge.r-project.org/](http://surveillance.r-forge.r-project.org/)
- arXiv:1411.0416 (Meyer et al., 2014): a guide to the three presented endemic-epidemic model classes
Acknowledgments

Joint work with:
  – Leonhard Held (University of Zurich)
  – Michael Höhle (University of Stockholm)

Funding:
  – Munich Center of Health Sciences (2007–2010)
  – Swiss National Science Foundation (2012–2015)
References


Feedback?
☞ sebastian.meyer@uzh.ch