

POWER-LAW MODELS FOR INFECTIOUS DISEASE SPREAD*

BY SEBASTIAN MEYER AND LEONHARD HELD

University of Zurich

Short-time human travel behaviour can be described by a power law with respect to distance. We incorporate this information in space-time models for infectious disease surveillance data to better capture the dynamics of disease spread. Two previously established model classes are extended, which both decompose disease risk additively into endemic and epidemic components: a spatio-temporal point process model for individual-level data, and a multivariate time series model for aggregated count data. In both frameworks, a power-law decay of spatial interaction is embedded into the epidemic component and estimated jointly with all other unknown parameters using (penalized) likelihood inference. Whereas the power law can be based on Euclidean distance in the point process model, a novel formulation is proposed for count data where the power law depends on the order of adjacency of discrete spatial units. The performance of the new approach is investigated by a re-analysis of individual cases of invasive meningococcal disease in Germany (2002–2008), and count data on influenza in 140 administrative districts of Southern Germany (2001–2008). In both applications, the power law substantially improves model fit and predictions, and is reasonably close to alternative qualitative formulations, where distance and order of neighbourhood, respectively, is treated as a factor. Implementation in the R package `surveillance` allows the approach to be applied in other settings.

1. Introduction. The surveillance of infectious diseases constitutes a key issue of public health and modelling their spread is basic to the prevention and control of epidemics. An important task is the timely detection of disease outbreaks, for which popular methods are the Farrington algorithm (Farrington et al., 1996; Noufaily et al., 2013) and cumulative sum (CUSUM) likelihood ratio detectors inspired by statistical process control (Höhle and Paul, 2008; Höhle, Paul and Held, 2009). As opposed to such prospective surveillance, retrospective surveillance is concerned with explaining the spread of an epidemic through statistical modelling, thereby

*Funded by the Swiss National Science Foundation (project #137919).

Keywords and phrases: power law, spatial interaction function, infectious disease surveillance, stochastic epidemic modelling, branching process with immigration, multivariate time series of counts, spatio-temporal point process.

assessing the role of environmental and socio-demographic factors or contact networks in shaping the evolution of an epidemic. The spatio-temporal data for such modelling primarily originate from routine public health surveillance of the occurrence of infectious diseases, and is ideally accompanied by additional data on influential factors to be accounted for. Surveillance data are available in different spatio-temporal resolutions each type requiring an appropriate model framework.

This paper covers both a spatio-temporal point process model for individual-level data (proposed by Meyer, Elias and Höhle, 2012, and motivated by the work of Höhle, 2009), and a multivariate time-series model for aggregated count data (established by Held and Paul, 2012, and earlier work). Although these two models are designed for different types of spatio-temporal surveillance data, both are inspired by the approach of Held, Höhle and Hofmann (2005) decomposing disease risk additively into “endemic” and “epidemic” components. The endemic component captures exogenous factors such as population, socio-demographic variables, long-term trends, seasonality, climate, or concurrent incidence of related diseases (all varying in time and/or space). Explicit dependence between observed cases, i.e., infectiousness, is then introduced through epidemic components driven by the observed past.

To describe disease spread in space, both models account for spatial interaction between units or individuals, respectively, but up to now, this has been incorporated rather crudely. The point process model used a Gaussian kernel to capture spatial interaction, and the multivariate time-series model restricted epidemic spread from time t to $t + 1$ to adjacent regions. However, a simple form of dispersal can be motivated by the findings of Brockmann, Hufnagel and Geisel (2006): They inferred from the dispersal of bank notes in the United States that (short-time) human travel behaviour can be well described by a decreasing power law of the distance x , i.e. $f(x) \propto x^{-d}$ with positive decay parameter d . An important characteristic of this power law is its slow convergence to zero (“heavy tail”), which in our application enables occasional long-range transmissions of infectious agents in addition to principal short-range infections. In the words of Brockmann, Hufnagel and Geisel (2006), their results “can serve as a starting point for the development of a new class of models for the spread of human infectious diseases”.

Power laws are well-known from the work by Pareto (1896) for the distribution of income and Zipf (1949) for city sizes and word frequencies in texts. They describe the distribution of earthquake magnitudes (Gutenberg and Richter, 1944) and many other natural phenomena (see Newman, 2005; Pinto, Lopes and Machado, 2012, for a review of power laws). Liljeros et al. (2001) reported on a power-law distribution of the number of sexual part-

ners, and [Albert and Barabási \(2002\)](#) review recent advances in network theory including scale-free networks where the number of edges is distributed according to a power law. Interestingly, a power law was also used as the distance decay function in geographic profiling for serial violent crime investigation ([Rossmo, 2000](#)) as well as in an application of this technique to infectious disease control ([Le Comber et al., 2011](#)). Examples of power-law transmission kernels to model the spatial dynamics of infectious diseases can be found in plant epidemiology ([Gibson, 1997](#); [Soubeyrand et al., 2008](#)) and in models for the 2001 UK foot-and-mouth disease epidemic ([Chis Ster and Ferguson, 2007](#)). Recently, [Geilhufe et al. \(2014\)](#) found that using (fixed) power-law weights between regions performed better than real traffic data in predicting influenza counts in Northern Norway. In both models for spatio-temporal surveillance data presented in the following sections, the power law will be estimated jointly with all other unknown parameters. Since the choice of a power law is a strong (yet well motivated) assumption, a comparison with alternative qualitative formulations is provided.

This paper is organized as follows: In Sections 2 and 3, respectively, the two model frameworks are reviewed and extended with power-law formulations for the spatial interaction of units. In Section 4, surveillance data on invasive meningococcal disease (IMD) and influenza are re-analysed using power laws and alternative qualitative approaches to be evaluated against previously used models for these data. We close with some discussion in Section 5 and a software overview in Appendix A.

2. Individual-level model.

2.1. *Introduction.* The spatio-temporal point process model proposed by [Meyer, Elias and Höhle \(2012\)](#) is designed for time-space-mark data $\{(t_i, \mathbf{s}_i, \mathbf{m}_i) : i = 1, \dots, n\}$ of individual case reports to describe the occurrence of infections (“events”) and their potential to trigger secondary cases. Formally, the model characterizes a point process in a region \mathbf{W} observed during a period $(0, T]$ through the conditional intensity function

$$(1) \quad \lambda(t, \mathbf{s}) = \nu_{[t][s]} \rho_{[t][s]} + \sum_{j:t_j < t} \eta_j \cdot g(t - t_j) \cdot f(\|\mathbf{s} - \mathbf{s}_j\|).$$

Related models are the purely temporal, “self-exciting” process proposed by [Hawkes \(1971\)](#), the spatio-temporal epidemic-type aftershock-sequences (ETAS) model from earthquake research ([Ogata, 1998](#)), the point process models discussed by [Diggle \(2007\)](#) and an additive-multiplicative point process model for discrete-space surveillance data proposed by [Höhle \(2009\)](#).

The first, endemic component in model (1) consists of a log-linear predictor $\log(\nu_{[t][s]}) = \beta_0 + \beta^\top \mathbf{z}_{[t][s]}$ proportional to an offset $\rho_{[t][s]}$, typically the population density. Both the offset and the exogenous covariates are given piecewise constant on a spatio-temporal grid (e.g. week \times district), hence the notation $[t][s]$ for the period which contains t in the region covering \mathbf{s} . In the IMD application in Section 4.1, $\mathbf{z}_{[t][s]} = ([t], \sin(\omega \cdot [t]), \cos(\omega \cdot [t]))^\top$ incorporates a time trend with one sinusoidal wave of frequency $\omega = 2\pi/365$.

A purely endemic, piecewise constant intensity model without the observation-driven epidemic component is equivalent to a Poisson regression model for the aggregated number of cases on the chosen spatio-temporal grid. However, with an epidemic component the intensity process depends on previously infected individuals and becomes “self-exciting”. Specifically, the epidemic force of infection at (t, \mathbf{s}) is the superposition of the infection pressures caused by each previously infected individual j . The individual infection pressure is weighted by the log-linear predictor $\log(\eta_j) = \gamma_0 + \gamma^\top \mathbf{m}_j$, which models the effect of individual/infection-specific characteristics \mathbf{m}_j such as the age of the infective. Regional-level covariates could also be included in \mathbf{m}_j , e.g. to model ecological effects on infectivity. Note, however, since the epidemic is modelled through a point process, the susceptible “population” consists of the continuous observation region $\mathbf{W} \subset \mathbb{R}^2$ and is thus infinite. Consequently, the model cannot include information on susceptibles, nor an autoregressive term as in time-series models.

Decreasing infection pressure of individual j over space and time is described by $f(x)$ and $g(t)$, parametric functions of the spatial distance x and of the elapsed time t since individual j became infectious, respectively. The spatial interaction could also be described more generally by a non-isotropic function $f_2(\mathbf{s})$ of the vector \mathbf{s} to the host, e.g. to incorporate the dominant wind direction in vector-borne diseases. However, in our application, f essentially reflects people’s movements and we assume that $f_2(\mathbf{s}) = f(\|\mathbf{s}\|)$ only depends on the distance to the host. For instance, Meyer, Elias and Höhle (2012) used an isotropic Gaussian kernel

$$(2) \quad f(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

with scale parameter σ . In what follows, we propose an alternative spatial interaction function, which allows for occasional long-range transmission of infections: a power law.

2.2. *Power-law extension.* The basic power law $f(x) = x^{-d}$, $d > 0$, is not a suitable choice for the distance decay of infectivity since it has a pole

at $x = 0$. For $x \geq \sigma > 0$, x^{-d} is the kernel of a Pareto density, but a shifted version to the domain \mathbb{R}_0^+ , known as Pareto type II and sometimes named after [Lomax \(1954\)](#), has density kernel

$$(3) \quad f(x) = (x + \sigma)^{-d} \propto \left(1 + \frac{x}{\sigma}\right)^{-d}$$

(see [Figure 1a](#)). Note that there is no need for the spatial interaction function to be normalized to a density. It is actually more closely related to correlation functions known from stationary random field models for geostatistical data ([Chilès and Delfiner, 2012](#)). For instance, the rescaled version $(1 + x/\sigma)^{-d}$ is a member of the Cauchy class introduced by [Gneiting and Schlather \(2004\)](#), which provides asymptotic power-law correlation as $x \rightarrow \infty$.

For short-range travel within 10 km, [Brockmann, Hufnagel and Geisel \(2006\)](#) found a uniform distribution instead of power-law behaviour, which suggests an alternative formulation with a “lagged” power law:

$$(4) \quad f(x) = \begin{cases} 1, & \text{for } x \leq \sigma, \\ \left(\frac{x}{\sigma}\right)^{-d}, & \text{otherwise.} \end{cases}$$

Spatial interaction is now constant up to the change point $\sigma > 0$, followed by a power-law decay for larger distances (see [Figure 1b](#)). A similar kernel was used by [Deardon et al. \(2010\)](#), therein called “geometric”) for the 2001 UK foot-and-mouth disease epidemic, additionally limiting spatial interaction to a pre-specified upper-bound distance.

A change-point-free kernel also unifying intended short-range and long-range characteristics is the Student kernel

$$(5) \quad f(x) = (x^2 + \sigma^2)^{-d} \propto \left(1 + \left(\frac{x}{\sigma}\right)^2\right)^{-d}$$

with scale parameter σ and shape (decay) parameter d (see [Figure 1c](#)). This kernel implements a power law of the squared distance and is known as the “Cauchy model” in geostatistics ([Chilès and Delfiner, 2012](#)). For $d > 0.5$, it describes a Student distribution with $2d - 1$ degrees of freedom.

To investigate the appropriateness of the assumed power-law decay, we also estimate an unconstrained step function

$$(6) \quad f(x) = \sum_{k=0}^K \alpha_k \mathbb{1}(x \in I_k),$$

which corresponds to treating the distance x – categorized into consecutive intervals I_k – as a qualitative variable.

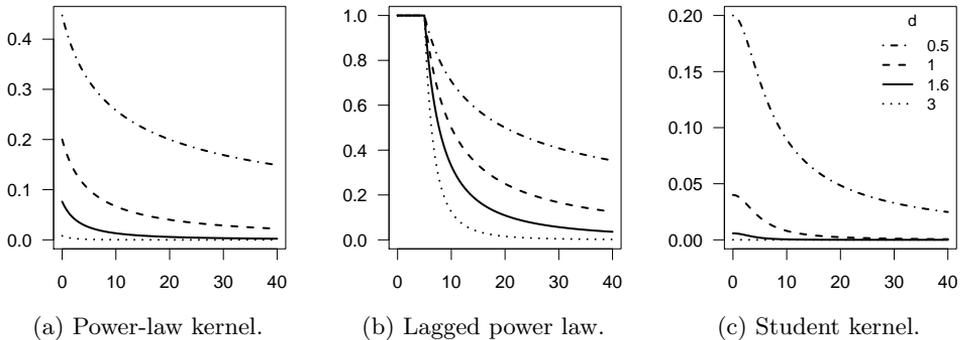


FIG 1. Power-law kernels as a function of the distance x for various choices of the decay parameter $d > 0$ and $\sigma = 5$.

2.3. *Inference.* Model parameters are estimated via maximization of the full (log-)likelihood, applying a quasi-Newton algorithm with analytical gradient and Hessian (cf. [Supplement B](#)). We estimate kernel parameters on the log-scale to avoid constrained optimization. For the step function, $\alpha_0 = 1$ is fixed to ensure identifiability.

The point process likelihood incorporates the integral of $f_2(\mathbf{s})$ over shifted versions of the observation region \mathbf{W} , which is represented by polygons. The same is true for the score function and approximate Fisher information with f_2 replaced by its partial derivatives. Except for the step function kernel (6), this requires a method of numerical integration such as the two-dimensional midpoint rule with an adaptive bandwidth, which was found to be best suited for the Gaussian kernel ([Meyer, Elias and Höhle, 2012](#)). For the other kernels we use a more sophisticated approach inspired by product Gauss cubature over polygons ([Sommariva and Vianello, 2007](#)). This cubature rule is based on Green's theorem, which relates the double integral over the polygon to a line integral along the polygon boundary. Its efficiency can be greatly improved in our specific case by taking analytical advantage of the isotropy of f_2 , after which numerical integration remains in only one dimension (see [Supplement B](#)). Regardless of any sophisticated cubature rule, the required integration of f_2 over n polygons in the log-likelihood is the part that makes model fitting cumbersome: it introduces numerical errors which have to be controlled such that they do not corrupt numerical likelihood maximization, and it increases computational cost by several orders of magnitude. For instance, in our IMD application in [Section 4.1](#) a single likelihood evaluation would only take 0.02 seconds if we used a constant spatial interaction function $f(x) \equiv 1$, where the integral does not depend on parameters be-

ing optimized and simply equals the area of the polygonal domain. For the Gaussian kernel, a single evaluation takes about 5 seconds, the step function takes 7 seconds, and the power law and Student kernel take about 20 seconds. The above and all following runtime statements refer to total CPU time at 2.80GHz (real elapsed time is shorter since some computations run in parallel on multiple CPUs).

3. Count data model.

3.1. *Introduction.* The multivariate time-series model established by [Held and Paul \(2012\)](#) (see also [Paul and Held, 2011](#); [Paul, Held and Toschke, 2008](#); [Held, Höhle and Hofmann, 2005](#)) is designed for spatially and temporally aggregated surveillance data, i.e. disease counts Y_{it} in regions $i = 1, \dots, I$ and periods $t = 1, \dots, T$. Formally, the counts Y_{it} are assumed to follow a negative binomial distribution

$$Y_{it} | \mathbf{Y}_{\cdot, t-1} \sim \text{NegBin}(\mu_{it}, \psi) \quad i = 1, \dots, I, t = 1, \dots, T$$

with additively decomposed mean

$$(7) \quad \mu_{it} = \nu_{it} e_{it} + \lambda_{it} Y_{i, t-1} + \phi_{it} \sum_{j \neq i} w_{ji} Y_{j, t-1},$$

and overdispersion parameter ψ such that the conditional variance of Y_{it} is $\mu_{it}(1 + \psi\mu_{it})$. The Poisson distribution results as a special case if $\psi = 0$. In (7), the first component represents the endemic component similar to the point process model (1). The endemic mean is proportional to an offset of known expected counts e_{it} typically reflecting the population at risk. The other two components are observation-driven epidemic components: The second component models an autoregression on the number of cases at the previous time point, whereas the third, “spatio-temporal” component captures transmission from other units. Note that without these epidemic components, the model would reduce to a negative binomial regression model for independent observations.

Each of the parameters ν_{it} , λ_{it} , and ϕ_{it} is a log-linear predictor of the form

$$\log(\cdot)_{it} = \alpha^{(\cdot)} + b_i^{(\cdot)} + \boldsymbol{\beta}^{(\cdot)\top} \mathbf{z}_{it}^{(\cdot)}$$

(where “.” is one of ν , λ , ϕ), containing fixed and region-specific intercepts as well as effects of exogenous covariates $\mathbf{z}_{it}^{(\cdot)}$ including time effects. For example, in the influenza application in Section 4.2,

$$\mathbf{z}_{it}^{(\nu)} = (t, \sin(1 \cdot \omega t), \cos(1 \cdot \omega t), \dots, \sin(S \cdot \omega t), \cos(S \cdot \omega t))^\top$$

describes an endemic time trend with a superposition of S harmonic waves of fundamental frequency $\omega = 2\pi/52$ (Held and Paul, 2012). The random effects $\mathbf{b}_i := (b_i^{(\lambda)}, b_i^{(\phi)}, b_i^{(\nu)})^\top$ account for heterogeneity between regions, and are assumed to follow independently a trivariate normal distribution with mean zero and covariance matrix Σ . Accounting for correlation of random effects across regions is possible by adopting a conditional autoregressive (CAR) model (Paul and Held, 2011).

The weights w_{ji} of the spatio-temporal component in (7) describe the strength of transmission from region j to region i , collected into an $I \times I$ weight matrix (w_{ji}) . In contrast to the individual-level model, all of the $Y_{j,t-1}$ cases of the neighbour j by aggregation contribute with the same weight w_{ji} to infections in region i . In previous work, these weights were assumed to be known and restricted to first-order neighbours:

$$(8) \quad w_{ji} = \begin{cases} 1/n_j, & \text{for } i \sim j, \\ 0, & \text{otherwise,} \end{cases}$$

where the symbol “ \sim ” denotes “is adjacent to” and n_j is the number of direct (first-order) neighbours of region j . This is a normalized version of the “raw” adjacency indicator matrix $\mathbf{A} = (\mathbb{1}(i \sim j))_{j,i=1,\dots,I}$, which is binary and symmetric. The idea behind normalization is that each region j distributes its cases uniformly to its n_j neighbours (Paul, Held and Toschke, 2008). Accordingly, the weight matrix is normalized to proportions such that all rows sum to 1. A simple alternative weight matrix considering only first-order neighbours would result from the definition $w_{ji} = 1/n_i$ for $i \sim j$ (i.e. columns sum to 1), meaning that the number of cases in a region i at time t is promoted by the mean of the neighbours at time $t - 1$. However, the first definition seems more natural in the framework of branching processes, where the point of view is from the infective source. Furthermore, the factor $1/n_i$ would be confounded with the region-specific effects $b_i^{(\phi)}$.

In either case, with the above weight matrix, the epidemic can only spread to first-order neighbours during the period $t \rightarrow t + 1$, except for independently imported cases via the endemic component. This ignores the ability of humans to travel further. In what follows, we propose a parametric generalization of the neighbourhood weights: a power law.

3.2. Power-law extension. To implement the power-law principle in the network of geographical regions, we first need to define a distance measure on which the power law acts. There are two natural choices: Euclidean distance between the regions’ centroid coordinates, and the order of the neighbourhoods between regions. The first one conforms to a continuous power

law whereas the second one is discrete. However, using centroid coordinates interferes with the area and shape of the regions. Specifically, a tiny neighbouring region would be attributed a stronger link than a large neighbour with centroid further apart, even if the latter shares more boundary than the tiny region. Using the common boundary length as a measure of “coupling” (Keeling and Rohani, 2002) would only cover adjacent regions. We thus opt for the discrete measure of neighbourhood order.

Formally, a region j is a k th-order neighbour of another region i , denoted $o_{ji} = o_{ij} = k$, if it is adjacent to a $(k - 1)$ th-order neighbour of i and if it is not itself a neighbour of order $k - 1$ of region i . In other words, two regions are k th-order neighbours, if the shortest route between them has k steps across distinct regions. The network of regions thus features a symmetric $I \times I$ matrix of neighbourhood orders with zeroes on the diagonal by convention.

Given this discrete distance measure, we generalize the previously used first-order weight matrix to higher-order neighbours assuming a power law with decay parameter $d > 0$:

$$(9) \quad w_{ji} = o_{ji}^{-d}$$

for $j \neq i$, and $w_{jj} = 0$. This may also be recognized as the kernel of the Zipf (1949) probability distribution. The raw power-law weights (9) can be normalized to

$$(10) \quad w_{ji} = \frac{o_{ji}^{-d}}{\sum_{k=1}^I o_{jk}^{-d}}$$

such that $\sum_{k=1}^I w_{jk} = 1$ for all rows j of the weight matrix. The higher the decay parameter d , the less important are higher-order neighbours. The limit $d \rightarrow \infty$ corresponds to the previously used first-order dependency, whereas $d = 0$ would assign equal weight to all regions.

Similarly to the point process modelling in Section 2.2, we also estimate the weights in a qualitative way by treating the order of neighbourhood as a factor:

$$(11) \quad w_{ji} = \sum_{o=1}^{M-1} \omega_o \cdot \mathbb{1}(o_{ji} = o) + \omega_M \cdot \mathbb{1}(o_{ji} \geq M).$$

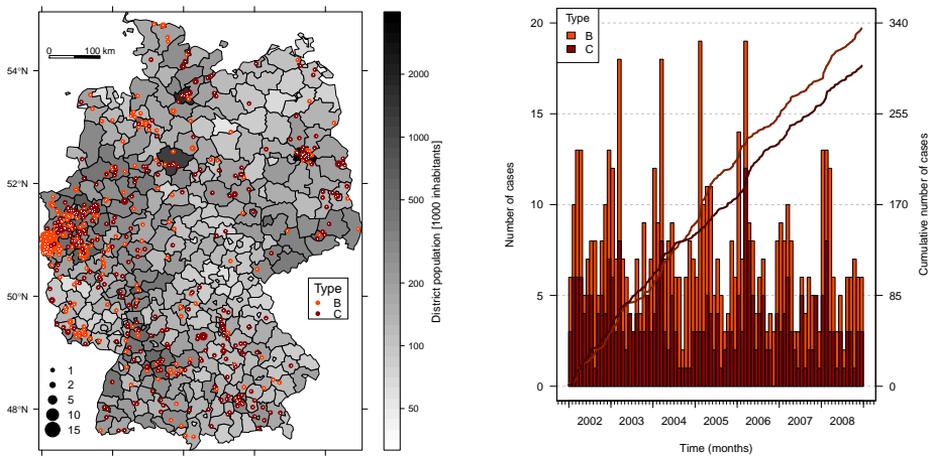
Aggregation of higher orders ($o_{ji} \geq M$) is necessary since the available information becomes increasingly sparse. As before, the unconstrained weights (11) can be normalized to $w_{ji} / \sum_{k=1}^I w_{jk}$.

3.3. Inference. We set $\omega_1 = 1$ for identifiability and estimate the decay parameter d and the unconstrained weights $\omega_2, \dots, \omega_M$ on the log-scale to enforce positivity. Supplied with the enhanced score function and Fisher information matrix, estimation of parametric weights is still possible within the penalized likelihood framework established by [Paul and Held \(2011\)](#) (cf. [Supplement B](#)). The authors argue, however, that classical model choice criteria such as Akaike's Information Criterion (AIC) can not be used straightforwardly for models with random effects. Therefore, performance of the power-law models and the previous first-order formulations is compared based on the assessment of one-step-ahead forecasts with strictly proper scoring rules: the logarithmic score (logS) and the ranked probability score (RPS) advocated by [Czado, Gneiting and Held \(2009\)](#) for count data:

$$\begin{aligned} \log S(P, y) &= -\log P(Y = y), \\ \text{RPS}(P, y) &= \sum_{k=0}^{\infty} [P(Y \leq k) - \mathbb{1}(y \leq k)]^2. \end{aligned}$$

These scores evaluate the discrepancy between the predictive distribution P from a fitted model and the later observed value y . Thus, lower scores correspond to better predictions. Note that the infinite sum in the RPS can be approximated by truncation at some large k in a way such that a pre-specified absolute approximation error is maintained ([Wei and Held, 2014](#)). Such scoring rules have already been used for previous analyses of the influenza surveillance data ([Held and Paul, 2012](#)). Along these lines, one-step-ahead predictions and associated scores are computed and statistical significance of the difference in mean scores is assessed using a Monte-Carlo permutation test for paired data.

4. Applications. We now apply the power-law formulations of both model frameworks to previously analysed surveillance data and investigate potential improvements with respect to predictive performance. We investigate the appropriateness of the power-law shape by alternative qualitative estimates of spatial interaction. In [Section 4.1](#), 635 individual case reports of IMD caused by the two most common bacterial finetypes of meningococci in Germany from 2002 to 2008 are analysed with the point process model (1). In [Section 4.2](#), the multivariate time-series model (7) is applied to weekly numbers of reported cases of influenza in the 140 administrative districts of the federal states Bavaria and Baden-Württemberg in Southern Germany from 2001 to 2008. In [Section 4.3](#), we evaluate a simulation-based long-term forecast of the 2008 influenza wave. Space-time animations of both surveillance data sets are provided in the web-based [Supplement A](#).



(a) Spatial point pattern with dot size proportional to the number of cases at the respective location (postcode level).

(b) Monthly aggregated time series and evolution of the cumulative number of cases (original resolution: date of specimen sampling).

FIG 2. Distribution of the 635 IMD cases in Germany, 2002-2008, caused by the two most common meningococcal finetypes B:P1.7-2,4:F1-5 (335 cases) and C:P1.5,2:F3-3 (300 cases), as reported to and typed by the German Reference Centre for Meningococci.

4.1. *Cases of invasive meningococcal disease in Germany, 2002–2008* (see Figure 2). In the original analysis of the IMD data (Meyer, Elias and Höhle, 2012), comprehensive AIC-based model selection yielded a linear time trend, a sinusoidal time-of-year effect ($S = 1$), and no effect of the (lagged) number of local influenza cases in the endemic component. The epidemic component included an effect of the meningococcal finetype (C:P1.5,2:F3-3 being less infectious than B:P1.7-2,4:F1-5, abbreviated by C and B in the following), a small age effect (3-18 year old patients tending to be more infectious), and supported an isotropic Gaussian spatial interaction function f compared to homogeneous spatial spread ($f(x) \equiv 1$). The analysis assumed constant infectivity over time until 30 days after infection when infectivity vanishes to zero, i.e. $g(t) = \mathbb{1}_{[0,30]}(t)$. In this paper, we replace the Gaussian kernel in the selected model by the proposed power-law distance decay (3) to investigate if it better captures the dynamics of IMD spread.

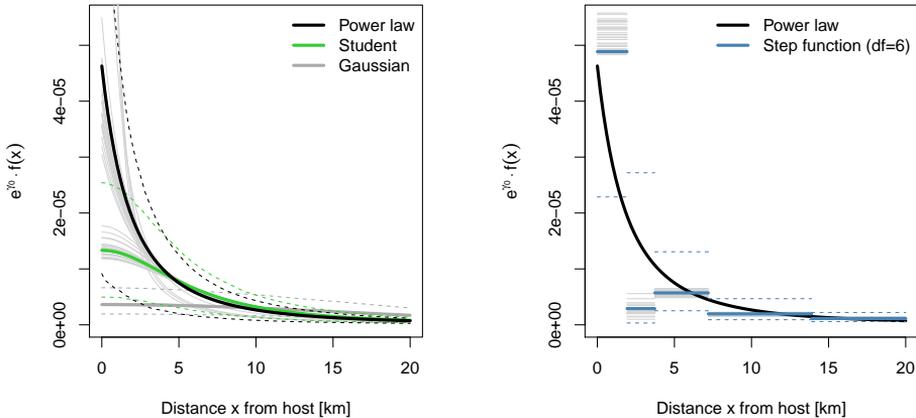
Note that the distinction between two finetypes in this application actually corresponds to a marked version of the point process model (1). It is described by an intensity function $\lambda(t, \mathbf{s}, k)$, where the sum in (1) is restricted to previously infected individuals of one specific bacterial finetype k , since we assume that infections of different finetypes are not associated

via transmission (Meyer, Elias and Höhle, 2012). For convenience, we kept notation simple and comparable to the multivariate time-series model of Section 3.

Prior to fitting point process models to the IMD data, the interval-censored nature of the data caused by a restricted resolution in space and time has to be taken into account: we only observed dates and residence postcodes of the cases (implicitly assuming that infections effectively happened within the residential neighbourhood). This makes the data interval-censored yielding tied observations. However, ties are not compatible with our (continuous-time, continuous-space) point process model since observing two events at the exact same time point or location has zero probability. In the original analyses with a Gaussian kernel f , events were untied in time by subtracting a $U(0, 1)$ -distributed random number from all observed time points (Meyer, Elias and Höhle, 2012), i.e., random sampling within each day, which is also the preferred method used by Diggle, Kaimi and Abellana (2010). To identify the two-parameter power law $(x + \sigma)^{-d}$, it was additionally necessary to break ties in space since otherwise $\log \sigma$ diverged to $-\infty$ yielding a pole at $x = 0$. A possible solution is to shift all locations randomly in space within their round-off intervals similar to the tie-breaking in time. Lacking a shapefile of the postcode regions, we shifted locations by a vector uniformly drawn from the disc with radius $\epsilon/2$, where ϵ is the minimum observed spatial separation of distinct points, here $\epsilon = 1.17\text{km}$. Accordingly, a sensitivity analysis was conducted by applying the random tie-breaking in time and space 30 times and fitting the models to all replicates.

Figure 3a displays estimated spatial interaction functions – appropriately scaled by $\exp(\hat{\gamma}_0)$ – together with confidence intervals and estimates from the sensitivity analysis (see Table 1 for values of $\hat{\gamma}_0$, $\hat{\sigma}$, and \hat{d}). The power law puts much more weight on localized transmissions with an initially faster distance decay of infectivity. Furthermore, it features a heavier tail than the Gaussian kernel, which facilitates the geographical spread of IMD by occasional long-range transmissions. Maps of the accumulated epidemic intensity in Supplement B visualize the impact of the power law on the modelled infectivity. Sensitivity analysis shows that AIC clearly prefers the new power-law kernel against the Gaussian kernel (mean $\Delta\text{AIC} = -27.6$, SD = 1.5). The Student kernel represents a compromise between the other two parametric kernels with short-range properties similar to the Gaussian kernel but with a heavy tail. However, AIC improvement is not as large as for the above power law (mean $\Delta\text{AIC} = -15.5$, SD = 0.9).

For these three kernels, sensitivity analysis of the random tie-breaking procedure in space and time generally confirmed the results. The Gaussian



(a) Power laws vs. the Gaussian kernel.

(b) Power law vs. a step function.

FIG 3. Estimated spatial interaction functions – appropriately scaled by the epidemic intercept $\exp(\gamma_0)$. The dashed lines represent 95% bootstrap confidence intervals, i.e., the pointwise 2.5% and 97.5% quantiles of the functions evaluated for 999 samples from the asymptotic multivariate normal distribution of the affected parameters. The light grey lines are estimates obtained from a sensitivity analysis with repeated random tie-breaking.

kernel was least affected by the small-scale perturbation of event times and locations. Some replicates for the power-law model yielded a slightly steeper shape, which is due to closely located points after random tie-breaking. Such an artifact would have been avoided if we had used constrained sampling in that the randomly shifted points obey a minimum separation of say 0.1 km.

The estimated lagged version of the power law (4) is shown in [Supplement B](#). It has a uniform short-range dispersal radius of $\hat{\sigma} = 0.40$ (95% CI: 0.18 to 0.86) kilometres. However, such a small σ is not interpretable since it is actually not covered by the spatial resolution of the data. Accordingly, the 30 estimates of the sensitivity analysis are more dispersed, as is the goodness-of-fit compared to the Gaussian kernel (mean $\Delta\text{AIC} = -21.1$, SD = 3.8).

Figure 3b shows a comparison of the estimated power law with a step function (6) for spatial interaction. An upper boundary knot had to be specified, which we set at 100 kilometres, where the step function drops to 0. We chose six knots to be equidistant on the log-scale within $[0, \log(100)]$, i.e., steps at 1.9, 3.7, 7.2, 13.9, 26.8 and 51.8 kilometres. Estimation took only 72 seconds due to the analytical implementation of the integration of f_2 over polygonal domains, whereas the power-law model took 42 minutes. The power law is well confirmed by the step function; it is almost completely enclosed by its

TABLE 1

Parameter estimates and 95% Wald confidence intervals for the Gaussian and the power-law model. Results for the Gaussian kernel are slightly different from those reported by Meyer, Elias and Höhle (2012) due to improved numerical integration. Note that we use the symbol σ for the scale parameter and d for the decay parameter in all spatial interaction functions, but these parameters as well as γ_0 are not directly comparable (instead see Figure 3).

	Gaussian kernel (2)		Power-law kernel (3)	
	Estimate	95% CI	Estimate	95% CI
β_0	-20.53	-20.62 to -20.44	-20.58	-20.68 to -20.47
β_{trend}	-0.05	-0.09 to -0.00	-0.05	-0.09 to 0.00
β_{sin}	0.26	0.14 to 0.39	0.26	0.12 to 0.39
β_{cos}	0.26	0.14 to 0.39	0.27	0.14 to 0.40
γ_0	-12.53	-13.15 to -11.91	-6.21	-9.32 to -3.10
γ_C	-0.91	-1.44 to -0.39	-0.80	-1.31 to -0.29
γ_{3-18}	0.67	0.04 to 1.31	0.78	0.11 to 1.45
$\gamma_{\geq 19}$	-0.29	-1.19 to 0.61	-0.18	-1.11 to 0.75
σ	16.37	13.95 to 19.21	4.60	1.80 to 11.71
d			2.47	1.80 to 3.39

95% confidence interval. The step function suggests an even steeper initial decay and has a slightly better fit in terms of AIC (mean $\Delta\text{AIC} = -6.9$, SD = 4.0 compared to the power law). However, it depends on the choice of knots, it is sensitive for artifacts of the data and forfeits monotonicity, which contradicts Tobler's *first law of geography* (Tobler, 1970).

Parameter estimates and confidence intervals for the Gaussian and the power-law model are presented in Table 1 (see Supplement B for parameter estimates of the other models). The parameters of the endemic component characterising time trend and seasonality were not affected by the change of the shape of spatial interaction, and also the epidemic coefficients of finetype and age group do not differ much between the models retaining their signs and orders of magnitude. For instance, also with the power-law kernel, the C-type is approximately half as infectious as the B-type, which is estimated by the multiplicative type-effect $\exp(\hat{\gamma}_C) = 0.45$ (95% CI: 0.27 to 0.75) on the force of infection (type B is the reference category here).

An important quantity in epidemic modelling is the expected number R of offspring (secondary infections) each case generates. This reproduction number can be derived from the fitted models for each event by integrating its triggering function $\eta_j g(t - t_j) f(\|\mathbf{s} - \mathbf{s}_j\|)$ over the observation region \mathbf{W} and period $[t_j, T]$ (Meyer, Elias and Höhle, 2012). Type-specific estimates of R are then obtained by averaging over the individual estimates by finetype.

TABLE 2

Estimated type-specific reproduction numbers with 95% confidence intervals based on parametric bootstrap (199 samples from the asymptotic multivariate normal distribution of the parameters).

	Gaussian kernel (2)		Power-law kernel (3)	
	Estimate	95% CI	Estimate	95% CI
B	0.22	0.17 to 0.31	0.26	0.10 to 0.35
C	0.10	0.06 to 0.15	0.13	0.05 to 0.19

Table 2 shows that the reproduction numbers become slightly larger, which is related to the heavier tail of the power law enabling additional interaction between events at far distances.

We close this application with two additional ideas for improvement of the model. First, it might be worth considering a population effect also in the *epidemic* component to reflect higher contact rates and thus infectivity in regions with a denser population. Using the log-population density of the infective's district, $\log \rho_{[t_j][s_j]}$, the corresponding parameter is estimated to be $\hat{\gamma}_{\log(\rho)} = 0.21$ (95% CI: -0.07 to 0.48), i.e. individual infectivity scales with $\rho^{0.21}$, where ρ ranges from 39 to 4225 km². Although the positive point estimate supports this idea, the wide confidence interval does not reflect strong evidence for such a population effect in the IMD data.

However, it is helpful to allow for spatial heterogeneity in the *endemic* component. For instance, an indicator for districts at the border or the distance of the district's centroids from the border could serve as proxies for simple edge effects. The idea is that as we get closer to the edge of the observation window (Germany) more infections will originate from external sources not directly linked to the observed history of the epidemic within Germany. We thus model a spatially varying risk of importing cases through the endemic component. For the Greater Aachen Region in the central-west part of Germany, where a spatial disease cluster is apparent in Figure 2a such a cross-border effect with the Netherlands was indeed identified by Elias et al. (2010) for the serogroup B finetype during our observation period using molecular sequence typing of bacterial strains in infected patients from both countries. Inclusion of an edge indicator in the endemic covariates $z_{[t][s]}$ improves AIC by 5 with an estimated rate ratio of 1.37 (95% CI: 1.10 to 1.70) for districts at the border versus inner districts. If we instead use the distance to the border, AIC improves by 20 with an estimated risk reduction of 5.0% (95% CI: 3.0% to 7.0%) per 10 km increase in distance to the border.

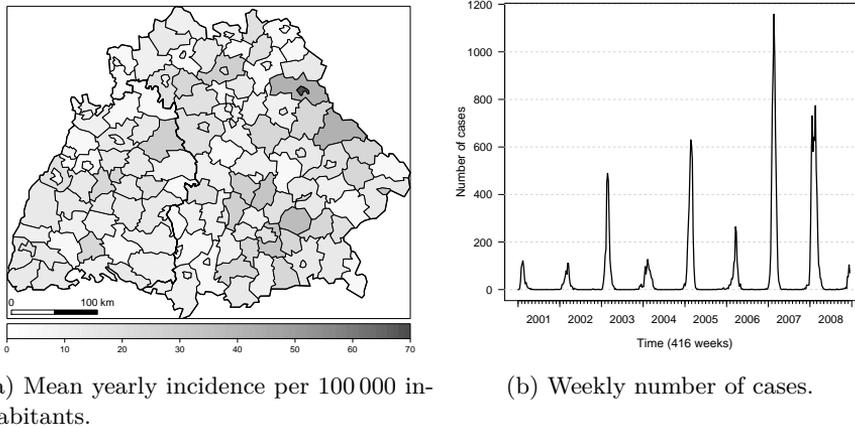
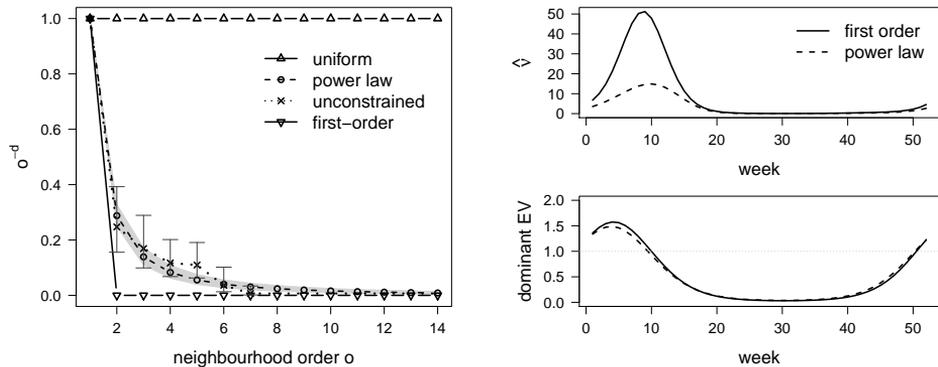


FIG 4. *Spatial and temporal distribution of reported influenza cases in the 140 districts of Bavaria and Baden-Württemberg during the years 2001 to 2008.*

4.2. *Influenza surveillance data from Southern Germany, 2001–2008* (see *Figure 4*). The best model (with respect to logS and RPS) for the influenza surveillance data found by [Held and Paul \(2012\)](#) using normalized first-order weights included $S = 1$ sinusoidal wave in each of the autoregressive (λ_{it}) and spatio-temporal (ϕ_{it}) components and $S = 3$ harmonic waves with a linear trend in the endemic component ν_{it} with the population fraction e_i in region i as offset. We now fit an extended model by estimating (raw or normalized) power-law neighbourhood weights (9) or (10) as described in Section 3.2, which replace the previously used fixed adjacency indicator.

Figure 5a shows the estimated normalized power law with $\hat{d} = 1.80$ (95% CI: 1.61 to 2.01). This decay is remarkably close to the power-law exponent 1.59 estimated by [Brockmann, Hufnagel and Geisel \(2006\)](#) for short-time travel in the USA with respect to distance (in kilometres), even though neighbourhood order is a discretized measure with no one-to-one correspondence to Euclidean distances, and travel behaviour in the USA is potentially different from that in Southern Germany. The plot also shows the estimated unconstrained weights for comparison with the power law. The sixth order of neighbourhood was the highest for which we could estimate an individual weight; higher orders had to be aggregated corresponding to $M = 7$ in (11). The unconstrained weights decrease monotonically and resemble nicely the estimated power law, which is enclosed by the corresponding 95% confidence intervals (except for order 5, which has a slightly higher weight). The results with raw weights are very similar and shown in [Supplement B](#).



(a) Power-law (10) and unconstrained weights (11) with 95% confidence intervals. (b) Seasonal variation of the endemic (top) and epidemic (bottom) components.

FIG 5. Estimated power-law and unconstrained weights (a), and seasonal variation (b) using normalized weights.

Figure 5b shows the estimated seasonal variation in the endemic component, and the course of the dominant eigenvalue (Held and Paul, 2012) for the normalized weight models. The dominant eigenvalue is a combination of the two epidemic components: if it is smaller than 1, it can be interpreted as the epidemic proportion of total disease incidence, otherwise it indicates an outbreak period. Whereas the course of this combined measure is more or less unchanged upon accounting for higher-order neighbours with a power law, the weight of the endemic component decreases remarkably. This goes hand in hand with an increased importance of the spatio-temporal component since in the power-law formulation much more information can be borrowed from the number of cases in other regions. Jumps of the epidemic to non-adjacent regions within one week are no longer dedicated to the endemic component only.

Concerning the remaining coefficients, there is less overdispersion in the power-law models (see ψ in Table 3), which indicates reduced residual heterogeneity. For the variance and correlation estimates of the random effects, there is no substantial difference between first-order and power-law models and even less between raw and normalized formulations.

To assess if the power-law formulation actually improves the previous first-order model, their predictive performance is compared based on one-week-ahead predictions for all 140 regions and the 104 weeks of the last two years. Computing these predictions for one model takes about 3 hours, since it needs to be refitted for every time point. Table 4 shows the resulting

TABLE 3

Estimated model parameters (with standard errors) excluding intercepts and trend/seasonal coefficients. The parameter $\beta_{\log(\text{pop})}^{(\phi)}$ in the first row belongs to a further extended power-law (PL) model, which accounts for population in the spatio-temporal component (last column). The σ^2 and ρ . parameters are the variances and correlations of the random effects (from Σ). The last row shows the final values of the penalized and marginal log-likelihoods.

	raw weights		normalized weights		
	first order	power law	first order	power law	PL + pop.
$\beta_{\log(\text{pop})}^{(\phi)}$	—	—	—	—	0.76 (0.13)
d	—	1.72 (0.10)	—	1.80 (0.10)	1.65 (0.10)
ψ	0.93 (0.03)	0.86 (0.03)	0.92 (0.03)	0.86 (0.03)	0.86 (0.03)
σ_{λ}^2	0.14	0.17	0.13	0.17	0.16
σ_{λ}^2	0.94	0.92	0.98	0.89	0.71
σ_{ν}^2	0.5	0.67	0.51	0.67	0.66
$\rho_{\lambda\phi}$	0.02	0.2	0.03	0.21	0.13
$\rho_{\lambda\nu}$	0.11	0.31	0.12	0.31	0.27
$\rho_{\phi\nu}$	0.56	0.29	0.55	0.3	0.39
$l_{\text{pen}}(l_{\text{mar}})$	-18400 (-433)	-18129 (-456)	-18387 (-436)	-18124 (-453)	-18124 (-439)

TABLE 4

Mean scores of 104×140 one-week-ahead predictions over the last two years, accompanied with p -values for comparing power-law and first-order weights obtained via permutation tests with 19999 random permutations. Note that the values obtained for normalized first-order weights are slightly different from the ones published by [Held and Paul \(2012\)](#) due to a correction of a recording error in the last week of the influenza data.

	raw weights		normalized weights	
	logS	RPS	logS	RPS
first order	0.5522	0.4205	0.5511	0.4194
power law	0.5453	0.4174	0.5448	0.4168
p -value	0.00005	0.11	0.0001	0.19

mean scores with associated p -values. Both logS and RPS improve when accounting for higher-order neighbours with a power law, while the difference is only significant for the logarithmic score. Furthermore, the normalized formulation performs slightly better than the raw weights. For instance, the mean difference in the logarithmic scores of the power-law models has an associated p -value of 0.0009. In the following, we therefore only consider the normalized versions. For additional comparison, the simple uniform weight model ($w_{ji} \equiv 1$), which takes into account higher-order neighbours but with equal weight, has mean logS = 0.5484 and mean RPS = 0.4215, and thus performs worse than a power-law decay and, according to the RPS, even worse than first-order weights.

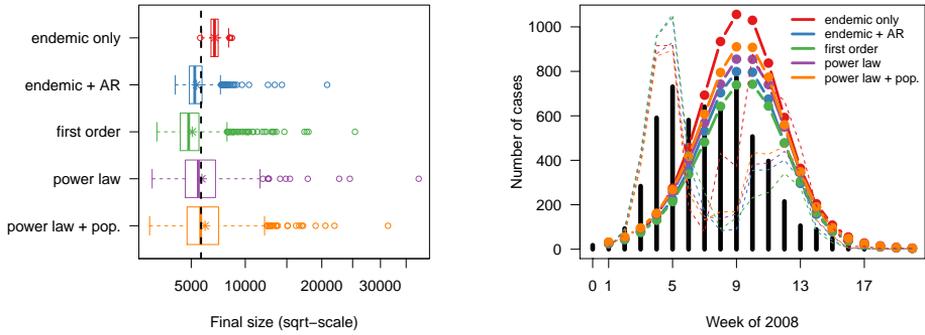
Similarly to the IMD analysis, further improvement of the model's descrip-

tion of human mobility can be achieved by accounting for the district-specific population also in the spatio-temporal component. The idea is that there tends to be more traffic to regional conurbations, i.e., districts with larger population, which are thus expected to import a bigger amount of cases from neighbouring regions. Note that inclusion of the log-population in $z_{it}^{(\phi)}$ affects susceptibility rather than infectivity, which is inverse to modelling the force of infection in the individual-based framework. The influenza data yield an estimated coefficient of $\hat{\beta}_{\log(\text{pop})}^{(\phi)} = 0.76$ (95% CI: 0.50 to 1.01), which provides strong evidence for such an agglomeration effect. This result agrees quite well with the finding of [Bartlett \(1957\)](#), that the number of infectious imports scales with the square root of the population size. The variance of the random effect $b_i^{(\phi)}$ of the spatio-temporal component is slightly reduced from 0.89 to 0.71 reflecting a decrease in residual heterogeneity between districts. The decay parameter is estimated to be slightly smaller in the extended model ($\hat{d} = 1.65$ (95% CI: 1.45 to 1.86)) and all other effects remain approximately unchanged (see [Table 3](#)). However, the predictive performance improves only minimally, e.g., the logarithmic score decreases from 0.5448 to 0.5447 ($p = 0.66$). This small change could be related to the random effects $b_i^{(\phi)}$, which replace parts of the population effect if it is not included as a covariate. Indeed, there is substantial correlation ($r_{\text{Pearson}} = 0.41$) between $\log(\text{pop}_i)$ and $b_i^{(\phi)}$ in the model without an explicit population effect in ϕ_{it} (see [Supplement B](#) for a scatterplot).

4.3. Long-term forecast of the 2008 influenza wave. For further evaluation of the power-law models described in [Section 4.2](#), we carry out a long-term forecast of the wave of influenza in 2008. Specifically, we simulate the evolution of the epidemic during the first 20 weeks in 2008 for each model trained by the previous years and initialized by the 18 cases of the last week of 2007 (see the animation in [Supplement A](#) for their spatial distribution). Predictive performance is then evaluated by the final size distributions and by proper scoring rules assessing the empirical distributions induced by the simulated counts both in the temporal and spatial domains. Since the logarithmic score is infinite in the case of zero predictive probability for the observed count, we instead use the [Dawid and Sebastiani \(1999\)](#) score

$$\text{DSS}(P, y) = \frac{(y - \mu_P)^2}{\sigma_P^2} + \log \sigma_P^2,$$

where μ_P and σ_P^2 denote the mean and the variance of P (see also [Gneiting and Raftery, 2007](#)).



(a) Final size distributions ($\sqrt{\cdot}$ -scale). The star in each box represents the mean, and the vertical dashed line marks the observed final size of 5781 cases.

(b) Time series of observed (bars) and mean simulated (dots) counts aggregated over all districts. Week 0 corresponds to the initial condition (2007-W52). The dashed lines show the (scaled) RPS (see also Table 5).

FIG 6. *Summary statistics of 1000 simulations of the wave of influenza during the first 20 weeks of 2008 for five competing models.*

Figure 6a shows the final size distributions of the simulated waves of influenza during the first 20 weeks of 2008. Note that model complexity increases from top to bottom and that we also considered the naive endemic model, i.e. independent counts, and the model without a spatio-temporal component as additional benchmarks. The endemic-only model, which decomposes disease incidence into spatial variation across districts, a seasonal and a log-linear time trend, overestimates the reported size of 5781 cases. It also does not allow for much variability in the size of the outbreak as opposed to the models with epidemic potential. The power-law models show the greatest amount of variation but best meet the reported final size: the power-law model without the population effect yields a simulated mean of 6022 (95% CI: 3126 to 10808). The huge uncertainty seems plausible with regard to the long forecast horizon over a whole epidemic wave.

Figure 6b shows the time series of observed and mean simulated counts aggregated over all districts. In 2008, the wave grew two or more weeks earlier than in previous years trained by the sinusoidal terms in the three components. This phenomenon cannot be captured by the simulations, which are solely based on the observed pattern during 2001-2007 and the distribution of the cases from the last week of 2007. Furthermore, instead of two peaks as observed specifically in 2008, the simulations yield a single, larger peak where the power-law models on average induce the best amplitudes with respect

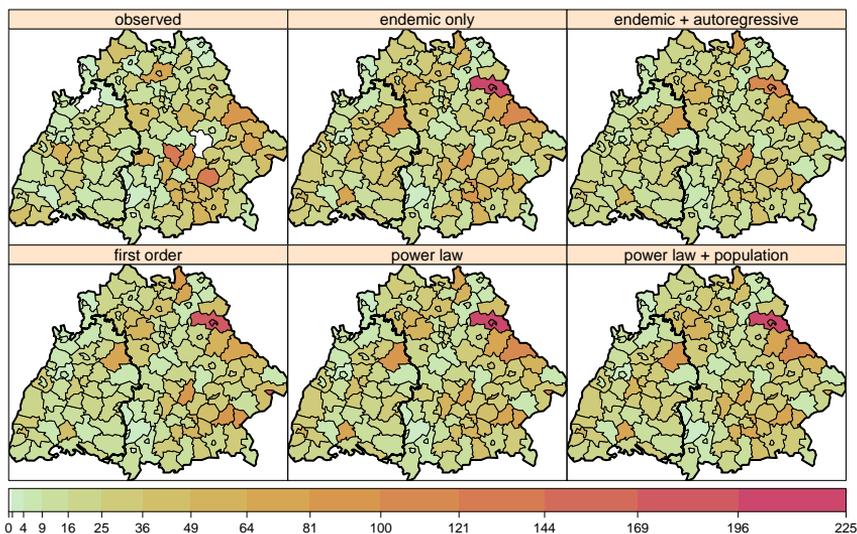


FIG 7. Observed and mean simulated incidence (cases per 100 000 inhabitants) aggregated over the 20 weeks forecast horizon (see [Supplement B](#) for scatterplots).

to final size. The simulated spatial distribution of the cases (see Figure 7) is very similar among the various models and agrees quite well with the observed pattern. Animations of the observed and mean simulated epidemics provide more insight about the epidemic spread and are available as [Supplement A](#). It is difficult to see a clear-cut traveling-wave of influenza in the reported data, which suggests that both an endemic component capturing immigration as well as scale-free jumps via the spatio-temporal component, i.e. power-law weights w_{ji} , are important. [Supplement A](#) also includes an animated series of weekly probability integral transform (PIT) histograms ([Gneiting, Balabdaoui and Raftery, 2007](#)) using the non-randomized version for count data proposed by [Czado, Gneiting and Held \(2009\)](#). These sequential PIT histograms mainly reflect the above time shift of the predictions. More clearly than the plots, the mean scores in [Table 5](#) show that predictive performance generally improves with increasing model complexity and use of a power-law decay.

5. Discussion. Motivated by the finding of [Brockmann, Hufnagel and Geisel \(2006\)](#) that short-time human travel roughly follows a power law with respect to distance, we investigated a power-law decay of spatial dependence between infections in two modelling frameworks for spatio-temporal surveil-

TABLE 5

Long-term predictive performance of 5 competing models in the temporal and spatial dimensions measured by mean DSS and RPS for the 2008 wave of influenza.

Model	Time		Space		Space-Time	
	DSS	RPS	DSS	RPS	DSS	RPS
endemic only	27.03	149.77	7.85	15.39	2.91	1.31
endemic + autoregressive	31.36	112.15	7.59	15.04	2.58	1.26
first order	26.46	108.61	7.51	15.63	2.50	1.26
power law	16.41	110.20	7.36	14.75	2.29	1.25
power law + population	15.49	111.86	7.24	14.30	2.29	1.24

lance data. A spatio-temporal point process model was applied to case reports of invasive meningococcal disease, and a multivariate time-series model was applied to counts of influenza aggregated by week and district. Since human mobility is an important driver of epidemic spread, the aim was to improve the predictive performance of these models using a power-law transmission kernel with respect to distance or neighbourhood order, respectively, where the decay is estimated jointly with all other model parameters.

In both applications considered, the power-law formulations performed better than previously used naive Gaussian or first-order interaction models, respectively. Furthermore, alternative piecewise constant, but otherwise unrestricted interaction models were in line with the estimated power laws. This confirms that the power-law distribution of short-time human travel translates to the modelling of infectious disease spread. We note that the qualitative interaction models could be replaced by (cubic) smoothing spline formulations, either in a continuous (Eubank, 2000) or in discrete fashion (Fahrmeir and Knorr-Held, 2000). In order to penalize deviations from the power law, this should be done on a log-log scale, where the power law is a simple linear relationship. However, data-driven estimation of the smoothing parameter may become difficult.

The heavy tail of the power law allows for long-range dependence between cases, which accordingly increased the importance of the epidemic component in both models. An alternative formulation of spatial interaction with occasional long-range transmission was used by Diggle (2006), who added a small distance-independent value to a powered exponential term of the scaled distance. However, this offset and the power parameter were poorly identified so that the power was fixed at 0.5 for the 2001 UK foot-and-mouth disease epidemic to correspond to the transmission kernel estimated by Keeling et al. (2001).

Regions at the edge of the observation window are missing potential sources of infection from the unobserved side of the border. To capture

unobserved heterogeneity due to immigration/edge effects, the count data model includes region-specific random effects $b_i^{(\nu)}$ in the endemic component. However, there was no clear pattern in their estimates with respect to regions being close to the border or not (Supplement B). In contrast, the IMD data supported edge effects, specifically concerning the border to the Netherlands. The spatial occurrence of cases met our simplistic approach of including the distance to the border as a covariate in the endemic component. This ignores that immigration might be more important in large metropolitan areas attracting people from abroad regardless of the location within Germany. A better way of accounting for edge effects would thus be to explicitly incorporate immigration data. For instance, Geilhufe et al. (2014) used incoming road or air traffic from outside North Norway as a proxy for the risk of importing cases of influenza, which led to improved predictive performance while also accounting for population in the spatio-temporal component.

Scaling regional susceptibility by population size proved very informative also for influenza in Southern Germany: More populated regions seem to attract more infections from neighbours than smaller regions, which reflects commuter-type imports (see Viboud et al., 2006, and Keeling and Rohani, 2008, Section 6.3.3.1). An exception of such a population effect in the spatio-temporal component might be seasonal accumulations in low-populated touristic regions. In the point process model for the IMD cases, the effect of population density on infectivity was less evident, which might be related to the very limited size of the point pattern with less than 100 cases per year over all of Germany.

We considered power laws as a description of spatial dispersal of infectious diseases as motivated by human travelling behaviour. Concerning temporal dispersal, power laws are usually not an appropriate description of the evolution of infectivity over time. Infectious diseases typically feature a very limited period of infectivity after the incubation period, since an infected individual will receive treatment and typically restrict its interaction radius upon the appearance of symptoms. Due to the small number of cases in the IMD data, we could not estimate a parametric temporal interaction function $g(t)$ and simply assumed constant infectivity during 30 days. More generally, $g(t)$ could represent an increasing level of infectivity beginning from exposure, followed by a plateau and then decreasing and eventually vanishing infectivity (Lawson and Leimich, 2000, Section 5.3). Even simpler, the counts in the multivariate time-series model were restricted to only explicitly depend on the previous week. This is reasonable if the generation time, the time consumed by an infective to cause a secondary case, is not larger

than the aggregation time in the surveillance data. For human influenza, [Cowling et al. \(2009\)](#) report a mean generation time of 3.6 days (95% CI: 2.9 days to 4.3 days).

Long-term simulated forecast of the 2008 influenza wave confirmed that the power-law model yields better predictions. However, the model was not able to describe the onset in 2008, which was two weeks earlier than in the years 2001-2007. For this to work, it would be necessary to further enrich the model by external processes such as vaccination coverage (as in [Herzog, Paul and Held, 2011](#)) or climate conditions ([Fuhrmann, 2010](#); [Willem et al., 2012](#)) entering as covariates in the endemic and/or epidemic components. An alternative approach has been suggested by [Fanshawe et al. \(2008\)](#) where seasonality parameters are allowed to change from year to year according to a random walk model. Implementation would then require Markov chain Monte Carlo or other more demanding techniques for inference. Despite the open issue of dynamic seasonality, the simulated final size and spatial distribution matched the reported epidemic quite well.

This success also indicates that under-reporting of influenza seems to be roughly constant over time. For instance, the 4 districts which did not report any cases during the 2008 forecast period (SK Kempten, SK Memmingen, LK Kelheim and SK Aschaffenburg) only reported 1, 0, 20 and 4 cases in total during 2001-2007. However, we can only model the effectively reported number of cases, which may be affected by time-varying attention drawn to influenza in the media. Syndromic surveillance systems aim to unify various routinely collected data sources, e.g., web searches for outbreak detection and monitoring ([Josseran et al., 2006](#); [Hulth, Rydevik and Linde, 2009](#)), and may thereby provide a more realistic picture of influenza.

Prospective detection of outbreaks is also possible based on the count data model presented here. A statistic could be based on quantiles of the distribution of $Y_{i,t+1} | \mathbf{Y}_{\cdot,t}$, e.g., an alarm could be triggered if the actual observed counts at $t + 1$ are above the 99% quantile, say ([Held et al., 2006](#)). Note that by including seasonality in the model, a yearly wave at the beginning of the year would be “planned” and not necessarily considered a deviation from default behaviour.

Our power-law approach is very useful in the absence of movement network data, (e.g., plane and train traffic). However, if such data were available ([Lazer et al., 2009](#)), neighbourhood weights w_{ji} in the count data model could instead be based on the connectivity between regions, which was investigated by [Schrödle, Held and Rue \(2012\)](#) for the spread of Coxiellosis in Swiss cows and by [Geilhufe et al. \(2014\)](#) for the spread of influenza in Northern Norway. In recent work, [Brockmann and Helbing \(2013\)](#) introduce the

“effective distance” to describe the 2009 H1N1 influenza pandemic. Their approach relates to what has already been termed “functional distance” by [Brown and Horton \(1970\)](#), i.e., a function of (inter-)regional properties like population and commuter or travel flows such that it “reflects the net effect of entity properties upon the propensity of the entities to interact” ([Brown and Holmes, 1971](#)). A recent example of using telephone call data as a measure of human interaction can be found in [Ratti et al. \(2010\)](#). Another fruitful area of future research is the statistical analysis of age-stratified surveillance data. Contact patterns vary across age ([Mosson et al., 2008](#); [Truscott et al., 2012](#)) calling for a unified analysis across age groups and regions.

APPENDIX A: SOFTWARE

All calculations have been carried out in the statistical software environment R 3.0.2 ([R Core Team, 2013](#)). Both model frameworks and their power-law extensions presented in this paper are implemented in the R package **surveillance** ([Höhle, Meyer and Paul, 2014](#)) as of version 1.6-0 available from the Comprehensive R Archive Network ([CRAN.R-project.org](#)). The two analysed datasets are included therein as `data("imdepi")` (courtesy of the German Reference Centre for Meningococci) and `data("fluBYBW")` (raw data obtained from the German national surveillance system operated by the [Robert Koch Institute, 2009](#)). The point process model (1) for individual point-referenced data can be fitted by the function `twinstim()`, and the multivariate time-series model (7) for count data is estimated by `hhh4()`. The implementations are flexible enough to allow for other specifications of the spatial interaction function f and the weights w_{ji} , respectively. A related two-component epidemic model ([Höhle, 2009](#)), which is designed for time-continuous individual surveillance data of a closed population with a fixed set of locations, e.g. for farm- or household-based epidemics, is also included as function `twinSIR()`.

Spatial integrals in the point process likelihood have been evaluated using cubature methods implemented in the R package **polyCub** 0.4-2 ([Meyer, 2014](#)). Maps have been produced using **sp** 1.0-14 ([Bivand, Pebesma and Gómez-Rubio, 2013](#)), animations using **animation** 2.2 ([Xie, 2013](#)).

ACKNOWLEDGEMENTS

We thank Michaela Paul for technical support on the original count data model, as well as Johannes Elias and Ulrich Vogel from the German Reference Centre for Meningococci, University of Würzburg, for providing us with the IMD data. This work was presented at the Summer School on Topics in Space-Time Modeling and Inference at Aalborg University, May 2013,

which enabled fruitful discussions with its participants. These also gave rise to using the more efficient cubature rule for isotropic functions over polygonal domains elaborated in [Supplement B](#) with valuable support by Emil Hedevang and Christian Reiher. We appreciate helpful comments on earlier versions of this manuscript by Julia Meyer, the editor Tilman Gneiting, and two anonymous referees. The research is financially supported by the Swiss National Science Foundation (project 137919: *Statistical methods for spatio-temporal modelling and prediction of infectious diseases*).

SUPPLEMENTARY MATERIAL

Supplement A: Animations of the IMD and influenza epidemics
(<http://www.biostat.uzh.ch/research/manuscripts/powerlaw.html>).

- Observed evolution of the IMD and influenza epidemics
- Simulated counts from various models for the 2008 influenza wave
- Weekly mean PIT histograms for these predictions

Supplement B: Inference details, an isotropic cubature method over polygons, and additional figures and tables
(doi: **TO BE COMPLETED BY THE TYPESETTER**).

- Details on likelihood inference for both models
- Integration of radially symmetric functions over polygonal domains
- Additional figures and tables of the power-law models for invasive meningococcal disease and influenza

REFERENCES

- ALBERT, R. and BARABÁSI, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics* **74** 47–97.
- BARTLETT, M. S. (1957). Measles periodicity and community size. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **120** 48–70.
- BIVAND, R. S., PEBESMA, E. and GÓMEZ-RUBIO, V. (2013). *Applied Spatial Data Analysis with R*, 2nd ed. *Use R!* **10**. Springer.
- BROCKMANN, D. and HELBING, D. (2013). The hidden geometry of complex, network-driven contagion phenomena. *Science* **342** 1337–1342.
- BROCKMANN, D., HUFNAGEL, L. and GEISEL, T. (2006). The scaling laws of human travel. *Nature* **439** 462–465.
- BROWN, L. A. and HOLMES, J. (1971). The delimitation of functional regions, nodal regions, and hierarchies by functional distance approaches. *Journal of Regional Science* **11** 57–72.
- BROWN, L. A. and HORTON, F. E. (1970). Functional distance: an operational approach. *Geographical Analysis* **2** 76–83.
- CHILÈS, J.-P. and DELFINER, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*, 2nd ed. *Wiley Series in Probability and Statistics* **713**. John Wiley & Sons.

- CHIS STER, I. and FERGUSON, N. M. (2007). Transmission parameters of the 2001 foot and mouth epidemic in Great Britain. *PLoS ONE* **2** e502.
- COWLING, B. J., FANG, V. J., RILEY, S., PEIRIS, J. M. S. and LEUNG, G. M. (2009). Estimation of the serial interval of influenza. *Epidemiology* **20** 344–347.
- CZADO, C., GNEITING, T. and HELD, L. (2009). Predictive model assessment for count data. *Biometrics* **65** 1254–1261.
- DAWID, A. P. and SEBASTIANI, P. (1999). Coherent dispersion criteria for optimal experimental design. *Annals of Statistics* **27** 65–81.
- DEARDON, R., BROOKS, S. P., GRENFELL, B. T., KEELING, M. J., TILDESLEY, M. J., SAVILL, N. J., SHAW, D. J. and WOOLHOUSE, M. E. J. (2010). Inference for individual-level models of infectious diseases in large populations. *Statistica Sinica* **20** 239–261.
- DIGGLE, P. J. (2006). Spatio-temporal point processes, partial likelihood, foot and mouth disease. *Statistical Methods in Medical Research* **15** 325–336.
- DIGGLE, P. J. (2007). Spatio-Temporal Point Processes: Methods and Applications. In *Statistical Methods for Spatio-Temporal Systems* (B. Finkenstädt, L. Held and V. Isham, eds.) 1, 1–45. Chapman & Hall/CRC, Boca Raton.
- DIGGLE, P. J., KAIMI, I. and ABELLANA, R. (2010). Partial-likelihood analysis of spatio-temporal point-process data. *Biometrics* **66** 347–354.
- ELIAS, J., SCHOULS, L. M., VAN DE POL, I., KEIJZERS, W. C., MARTIN, D. R., GLENNE, A., OSTER, P., FROSCH, M., VOGEL, U. and VAN DER ENDE, A. (2010). Vaccine preventability of meningococcal clone, Greater Aachen Region, Germany. *Emerging Infectious Diseases* **16** 465–472.
- EUBANK, R. L. (2000). Spline Regression. In *Smoothing and Regression: Approaches, Computation, and Application*, (M. G. Schimek, ed.). *Wiley Series in Probability and Statistics* Ch. 1, 1–18. John Wiley & Sons.
- FAHRMEIR, L. and KNORR-HELD, L. (2000). Dynamic and Semiparametric Models. In *Smoothing and Regression: Approaches, Computation, and Application*, (M. G. Schimek, ed.). *Wiley Series in Probability and Statistics* Ch. 18, 513–544. John Wiley & Sons.
- FANSHAW, T. R., DIGGLE, P. J., RUSHTON, S., SANDERSON, R., LURZ, P. W. W., GLINIANAIA, S. V., PEARCE, M. S., PARKER, L., CHARLTON, M. and PLESS-MULLOLI, T. (2008). Modelling spatio-temporal variation in exposure to particulate matter: a two-stage approach. *Environmetrics* **19** 549–566.
- FARRINGTON, C. P., ANDREWS, N. J., BEALE, A. D. and CATCHPOLE, M. A. (1996). A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **159** 547–563.
- FUHRMANN, C. (2010). The effects of weather and climate on the seasonality of influenza: What we know and what we need to know. *Geography Compass* **4** 718–730.
- GEILHUF, M., HELD, L., SKRØVSETH, S. O., SIMONSEN, G. S. and GODTLIEBSEN, F. (2014). Power law approximations of movement network data for modeling infectious disease spread. *Biometrical Journal*. In press.
- GIBSON, G. J. (1997). Markov chain Monte Carlo methods for fitting spatiotemporal stochastic models in plant epidemiology. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **46** 215–233.
- GNEITING, T., BALABDAOUI, F. and RAFTERY, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **69** 243–268.
- GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102** 359–378.
- GNEITING, T. and SCHLATHER, M. (2004). Stochastic models that separate fractal dimension and the hurst effect. *SIAM Review* **46** 269–282.

- GUTENBERG, B. and RICHTER, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America* **34** 185–188.
- HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** 83–90.
- HELD, L., HÖHLE, M. and HOFMANN, M. (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling* **5** 187–199.
- HELD, L. and PAUL, M. (2012). Modeling seasonality in space-time infectious disease surveillance data. *Biometrical Journal* **54** 824–843.
- HELD, L., HOFMANN, M., HÖHLE, M. and SCHMID, V. (2006). A two-component model for counts of infectious diseases. *Biostatistics* **7** 422–437.
- HERZOG, S. A., PAUL, M. and HELD, L. (2011). Heterogeneity in vaccination coverage explains the size and occurrence of measles epidemics in German surveillance data. *Epidemiology and Infection* **139** 505–515.
- HULTH, A., RYDEVIK, G. and LINDE, A. (2009). Web queries as a source for syndromic surveillance. *PLoS ONE* **4** e4378.
- HÖHLE, M. (2009). Additive-multiplicative regression models for spatio-temporal epidemics. *Biometrical Journal* **51** 961–978.
- HÖHLE, M., MEYER, S. and PAUL, M. (2014). **surveillance**: Temporal and Spatio-Temporal Modeling and Monitoring of Epidemic Phenomena. R package version 1.8-0.
- HÖHLE, M. and PAUL, M. (2008). Count data regression charts for the monitoring of surveillance time series. *Computational Statistics and Data Analysis* **52** 4357–4368.
- HÖHLE, M., PAUL, M. and HELD, L. (2009). Statistical approaches to the monitoring and surveillance of infectious diseases for veterinary public health. *Preventive Veterinary Medicine* **91** 2–10. Special Issue: GisVet 2007.
- ROBERT KOCH INSTITUTE (2009). SurvStat@RKI. <http://www3.rki.de/SurvStat>. Queried on March 6, 2009.
- JOSSERAN, L., NICOLAU, J., CAILLÈRE, N., ASTAGNEAU, P. and BRÜCKER, G. (2006). Syndromic surveillance based on emergency department activity and crude mortality: two examples. *Eurosurveillance* **11** 225–229.
- KEELING, M. J. and ROHANI, P. (2002). Estimating spatial coupling in epidemiological systems: a mechanistic approach. *Ecology Letters* **5** 20–29.
- KEELING, M. J. and ROHANI, P. (2008). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.
- KEELING, M. J., WOOLHOUSE, M. E. J., SHAW, D. J., MATTHEWS, L., CHASE-TOPPING, M., HAYDON, D. T., CORNELL, S. J., KAPPEY, J., WILESMITH, J. and GRENFELL, B. T. (2001). Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* **294** 813–817.
- LAWSON, A. B. and LEIMICH, P. (2000). Approaches to the space-time modelling of infectious disease behaviour. *IMA Journal of Mathematics Applied in Medicine and Biology* **17** 1–13.
- LAZER, D., PENTLAND, A., ADAMIC, L., ARAL, S., BARABÁSI, A.-L., BREWER, D., CHRISTAKIS, N., CONTRACTOR, N., FOWLER, J., GUTMANN, M., JEBARA, T., KING, G., MACY, M., ROY, D. and VAN ALSTYNE, M. (2009). Computational Social Science. *Science* **323** 721–723.
- LE COMBER, S., ROSSMO, D. K., HASSAN, A., FULLER, D. and BEIER, J. (2011). Geographic profiling as a novel spatial tool for targeting infectious disease control. *International Journal of Health Geographics* **10** 35.
- LILJEROS, F., EDLING, C. R., AMARAL, L. A. N., STANLEY, H. E. and ABERG, Y. (2001). The web of human sexual contacts. *Nature* **411** 907–908.
- LOMAX, K. S. (1954). Business failures: another example of the analysis of failure data.

- Journal of the American Statistical Association* **49** 847–852.
- MEYER, S. (2014). **polyCub**: Cubature over Polygonal Domains. R package version 0.4-2.
- MEYER, S., ELIAS, J. and HÖHLE, M. (2012). A space-time conditional intensity model for invasive meningococcal disease occurrence. *Biometrics* **68** 607–616.
- MOSSONG, J., HENS, N., JIT, M., BEUTELS, P., AURANEN, K., MIKOLAJCZYK, R., MAS-SARI, M., SALMASO, S., TOMBA, G. S., WALLINGA, J., HELJNE, J., SADKOWSKA-TODYS, M., ROSINSKA, M. and EDMUNDS, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine* **5** e74.
- NEWMAN, M. E. J. (2005). Power laws, Pareto distributions and Zipf’s law. *Contemporary Physics* **46** 323–351.
- NOUFAILY, A., ENKI, D. G., FARRINGTON, P., GARTHWAITE, P., ANDREWS, N. and CHARLETT, A. (2013). An improved algorithm for outbreak detection in multiple surveillance systems. *Statistics in Medicine* **32** 1206–1222.
- OGATA, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics* **50** 379–402.
- PARETO, V. (1896). *Cours d’Économie Politique* **1**. F. Rouge, Lausanne.
- PAUL, M., HELD, L. and TOSCHKE, A. (2008). Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine* **27** 6250–6267.
- PAUL, M. and HELD, L. (2011). Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. *Statistics in Medicine* **30** 1118–1136.
- PINTO, C. M. A., LOPES, A. M. and MACHADO, J. A. T. (2012). A review of power laws in real life phenomena. *Communications in Nonlinear Science and Numerical Simulation* **17** 3558–3578.
- RATTI, C., SOBOLEVSKY, S., CALABRESE, F., ANDRIS, C., READES, J., MARTINO, M., CLAXTON, R. and STROGATZ, S. H. (2010). Redrawing the map of Great Britain from a network of human interactions. *PLoS ONE* **5** e14248.
- ROSSMO, D. K. (2000). *Geographic Profiling*. CRC Press, Boca Raton.
- SCHRÖDLE, B., HELD, L. and RUE, H. (2012). Assessing the impact of a movement network on the spatiotemporal spread of infectious diseases. *Biometrics* **68** 736–744.
- SOMMARIVA, A. and VIANELLO, M. (2007). Product Gauss cubature over polygons based on Green’s integration formula. *Bit Numerical Mathematics* **47** 441–453.
- SOUBEYRAND, S., HELD, L., HÖHLE, M. and SACHE, I. (2008). Modelling the spread in space and time of an airborne plant disease. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **57** 253–272.
- R CORE TEAM (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- TOBLER, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography* **46** 234–240.
- TRUSCOTT, J., FRASER, C., CAUCHEMEZ, S., MEEYAI, A., HINSLEY, W., DONNELLY, C. A., GHANI, A. and FERGUSON, N. (2012). Essential epidemiological mechanisms underpinning the transmission dynamics of seasonal influenza. *Journal of the Royal Society Interface* **9** 304–312.
- VIBOUD, C., BJØRNSTAD, O. N., SMITH, D. L., SIMONSEN, L., MILLER, M. A. and GRENFELL, B. T. (2006). Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312** 447–451.
- WEI, W. and HELD, L. (2014). Calibration tests for count data. *TEST*. In revision.
- WILLEM, L., VAN KERCKHOVE, K., CHAO, D. L., HENS, N. and BEUTELS, P. (2012). A nice day for an infection? Weather conditions and social contact patterns relevant to influenza transmission. *PLoS ONE* **7** e48695.

- XIE, Y. (2013). **animation**: An R package for creating animations and demonstrating statistical methods. *Journal of Statistical Software* **53** 1–27.
- ZIPF, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley Press, Cambridge, Massachusetts, USA.

UNIVERSITY OF ZÜRICH
INSTITUTE OF SOCIAL AND PREVENTIVE MEDICINE
DIVISION OF BIOSTATISTICS
HIRSCHENGRABEN 84
CH-8001 ZÜRICH
SWITZERLAND
E-MAIL: Sebastian.Meyer@ifspm.uzh.ch
Leonhard.Held@ifspm.uzh.ch
URL: www.biostat.uzh.ch