Fine-Grid Spatial Interaction Matrices for Surveillance Models with Application to Influenza in Germany

Manuel Stapper[1*] and Sebastian Funk[1]

[1]Department of Infectious Disease Epidemiology and Dynamics, London School of Hygiene and Tropical Medicine, London, UK.

*Corresponding author

Abstract

Background Spatial models of infectious disease transmission often rely on administrative boundaries and simple measure of proximity, such as the neighbourhood order. While these methods effectively capture key aspects of disease spread, they may not fully account for nuances in transmission risk. This study introduces a new approach that defines transmission risk at the individual level and aggregates it to the district level. This approach aims to improve both forecast accuracy and the spatial resolution of risk mapping.

Method We propose the Fine-Grid Spatial Interaction Matrix (FGSIM), which models transmission risk between districts based on distances between individuals. Five distance measures are transformed into contact intensity using a power-law function, forming a weight matrix applicable in the endemic-epidemic framework. We evaluate FGSIM using influenza data from Germany (2001–2020) and compare its performance to established methods and a simplified FGSIM variant. Forecasts for one to eight weeks ahead are assessed across four study regions using the weighted interval score (WIS) and ranked probability score (RPS).

Results FGSIM outperforms its simplified variants and models without spatial dependence in most cases when evaluated in-sample. For one-week-ahead forecasts, a centroid-based model performs best in three of four regions (two when evaluated on the logarithmic scale). For longer-term forecasts (four or eight weeks ahead), one FGSIM model consistently outperforms most others. Risk maps at 100m resolution demonstrate the ability of FGSIM to identify high-risk areas not aligned with administrative boundaries.

Conclusion FGSIM provides a flexible and computationally feasible approach to incorporate individual-level spatial structure into district-level infectious disease

models. While established methods perform well in short-term forecasts, FGSIM demonstrates competitive performance for longer forecasting horizons. In addition, it enables fine-scale risk mapping, making it a valuable tool for public health planning beyond administrative districts.

Keywords: Spatio-Temporal Modeling, Epidemiology, Forecasting

1 Introduction

Infectious diseases are transmitted through complex networks of interactions between individuals. While many factors influence contact patterns, such as social networks, workplace environments, and surrounding infrastructure, spatial proximity between an infected and a susceptible individual is especially important. The underlying assumption is that diseases tend to spread more easily between geographically close regions or individuals due to the higher frequency of interactions that can lead to transmission.

One of the aims of spatial modelling in epidemiology is quantifying and understanding the intensity of contact and thus transmission between geographical units. Capturing these dependencies allows for more accurate predictions of disease spread and provides insights into its underlying dynamics. However, modelling spatial dependence often requires a trade-off between simplicity and complexity. Simple models make strong assumptions about human behaviour, which makes them easy to implement and maintain and computationally efficient but limits their ability to capture critical factors influencing disease transmission. In contrast, complex models can incorporate detailed human movement patterns but demand extensive data and computational resources. An important challenge lies in making realistic assumptions about human movement while ensuring computational feasibility of fitting the model.

A straightforward measure of spatial proximity is to define the closeness of relevant areas (e.g. administrative districts) as the distances between their centres. From this starting point, the intensity of interaction between districts can be modeled as a function of those distances. For instance, one could assume a constant interaction intensity for the k nearest districts or use a decay function to model how contact intensity decreases with distance. Another commonly used approach is defining closeness based on the order of neighborhood, i.e., the minimum number of district boundaries an individual must cross to move between the districts. Contact intensity can then be specified using this measure; for example by considering only first-order neighbors or by assigning equal interaction intensities to districts within the same order of neighborhood. A widely used method, rooted in the work of Brockmann et al (2006), applies a power law to the order of neighborhood, assuming that contact intensity decays geometrically with increasing distance, see Meyer and Held (2014). While practical and effective, these models depend on area boundaries and are sensitive to where they are. See section A for a discussion on the sensitivity of the order of neighbourhood.

For modelling the probability of contact between two individuals, the distance, whether expressed as Euclidean distance or neighborhood order is not the only determining factor. Individuals living in densely populated districts are less likely to travel

long distances than those in remote areas. Similar to Newton's law of gravitation, Zipf (1946) proposed that human movement follows a "gravity model", where the probability of an individual traveling to a destination depends on the populations at the origin and destination and decreases with increasing distance between them, see for example Haynes and Fotheringham (1985).

Other approaches incorporate factors beyond distance to model human movement. Stouffer's law of intervening opportunities states that migration is proportional to the opportunities at the destination and inversely proportional to those between origin and destination, see Stouffer (1940). This principle motivates the radiation model (see for example Simini et al (2012) and Alis et al (2021)), which considers populations at the origin and destination while accounting for the population in the circle around the origin that touches the destination. For a detailed overview of human mobility models, see Barbosa et al (2018) and Tizzoni et al (2014).

Most of the models discussed above operate at the level of districts, aggregating population movement and interaction patterns. However, this level of generalization can be relaxed to increase model flexibility. Agent-based models, for instance, define movement and interactions at the individual level rather than the district level.

The remainder of this article introduces a novel approach to spatial dependence modelling that defines interactions at the individual level and then aggregates them to the district level. In this framework, individual interactions are treated as random variables, where stochasticity arises from the randomly selecting individuals from districts. Aggregation to the district level is performed by taking expectations. This method combines the advantages of individual-level modelling with the simplicity of district-level modelling. By focusing on individual locations rather than predefined district boundaries, this approach avoids restrictive assumptions about human movement patterns, while maintaining computational feasibility.

The following section introduces the proposed methodology. Section 3 applies the method to Influenza cases in Germany, evaluating in-sample model fit and out-of-sample forecast performance. Finally, Section 4 concludes with a discussion of advantages, limitations and extensions of this work.

2 Methodology

In this study, we apply the endemic-epidemic framework introduced by Held et al (2005) to model the number of newly infected individuals in spatial units. Its implementation in the surveillance package (see Meyer et al (2017)) is widely used for spatio-temporal epidemic data. The versatile framework partitions the expected number of new infections into endemic and epidemic components.

Let Y_{it} be the number of new cases in district i at time t and Y_t the vector of observations of all areas $i = 1, ..., N$. Given the information set available at time $t-1$, i.e. the sigma algebra $\mathcal{F}_{t-1}$ generated by $\{Y_{t-1}, Y_{t-2}, ...\}$, the expectation $\mu_{it} = \mathrm{E}(Y_{it}|\mathcal{F}_{t-1})$ is defined as

$$\mu_{it} = \nu_{it} e_{it} + \lambda_{it} \sum_{j=1}^{N} W_{ji} Y_{j,t-1} \ .$$

The first term captures the baseline infection counts, explained by usual circulation. It depends on e_{it}, the number of (susceptible) individuals in district i at time t. The second term, the epidemic component, captures the transmission of cases, with W denoting the weight matrix. Each element W_{ji} represents the intensity of contacts between units j and i that potentially lead to a transmission of the disease from j to i. The weight matrix is typically row-normalised allowing each row $W_{j\cdot}$ to be interpreted as the distribution of contacts from an infected individual in j across all districts.

Both components are scaled by log-linear predictors ν_{it} and λ_{it}, defined as

$$\log(\nu_{it}) = \alpha^{(\nu)} + \beta_i^{(\nu)} + \left(\gamma^{(\nu)}\right)^T z_t^{(\nu)}$$
$$\log(\lambda_{it}) = \alpha^{(\lambda)} + \beta_i^{(\lambda)} + \left(\gamma^{(\lambda)}\right)^T z_t^{(\nu)}$$

Here, $\alpha^{(\nu)}$ and $\alpha^{(\lambda)}$ are intercept parameters, β_i are district-specific effects, which can be either random or fixed effects. Other covariates z_t may include seasonality components and γ^{ν} and $\gamma^{(\lambda)}$ are the corresponding parameter vectors. To ensure identification, the intercept parameters $\alpha^{(\nu)}$ and $\alpha^{(\lambda)}$ are often dropped in presence of district effects or the average district effect is set to zero.

The number of new infections is then assumed to follow a Negative Binomial distribution:

$$Y_{it}|\mathcal{F}_{t-1} \sim \text{NegBin}(\mu_{it}, \psi)$$

where the Negative Binomial distribution can be replaced by a Poisson distribution or a Negative Binomial with district-specific overdispersion parameters ψ_i. For additional flexibility in modelling intra-district transmission, the epidemic term is often decomposed into an intra-district (autoregressive) component and an inter-district (neighbour-driven) component.

In the remainder of this article, we focus on the weights W_{ji}, which capture the dynamics of transmission. A well-established approach is to define weights using the neighbourhood order between districts O_{ji} and applying the power law such that unnormalised weights are given by $W_{ji} = O_{ji}^{-d}$, with d as a decay parameter.

The proposed alternative for defining weights is based on a similar principle, but uses a distance measure at the individual level. For instance, consider a pair of districts i and j. When selecting an individual from each district at random, we measure the distance D_{ji} between them (e.g. the beeline distance). Assuming that the probability of contact decays with distance according to a power law, the contact intensity is proportional to D_{ji}^{-d}. However, since D_{ji} is a random variable due to the random selection of individuals, the individual transmission risk is a random variable as well. Hence, we aggregate the individual transmission risk by taking the expectation $\mathrm{E}(D_{ji}^{-d})$, where the expectation is taken with respect to the distribution of individuals living in the two districts. Computing $\mathrm{E}(D_{ji}^{-d})$ exactly is computationally infeasible when considering all pairs of individuals. Instead, assumptions are made about the distribution of D_{ji}.

This process involves four steps. First, locations of individuals from each district are repeatedly sampled according to the population density. Second, distances between

the sampled pairs are calculated, yielding a sample of distances. Third, these distances are used to fit a parametric distribution. Finally, the fitted distribution is used to translate distances into weights by taking expectations.

The Lognormal distribution is a natural distributional assumption for modelling D_{ji}, as it takes only positive values and has desirable properties. Specifically, if D_{ji} is Lognormal, D_{ji}^{-d} also follows a Lognormal distribution, allowing for a closed-form expectation. However, the Lognormal distribution may not adequately account for bounded distances due to district boundaries and may be overly restrictive. To address this, we assume that D_{ji} follows a mixture of truncated Lognormal distributions. This approach preserves the closed-form expectation $\mathrm{E}(D_{ji}^{-d})$ while providing greater flexibility and ensuring boundedness. The mixture is fitted using an Expectation-Maximisation (EM) algorithm, where information criteria are used for the selection of the number of mixing components.

In practical applications, truncating the power law can be advantageous to prevent extreme values for small distances. Further, it is a reasonable assumption that individuals have a small radius around their home with uniform movement. To account for this, we generalise the power law by introducing a truncation threshold, $\delta > 0$ such that the weights are defined as the minimum of $\mathrm{E}(D_{ji}^{-d})$ and δ^{-d}. The truncation parameter δ is estimated alongside the other model parameters. For a full derivation of the closed-form expectations, see B. Derivatives of these expectations, required for numerical optimisation, are provided in C.

In this study, five distance measures are applied in the new approach. All measures begin by sampling a large number of individuals per district according to population density. Data on population distributions within districts is available from WorldPop (2018) on a 100m by 100m grid.

Beeline Distance

The simplest and most intuitive measure of distance between individuals is the geodesic, or beeline, distance. This measure is easy to compute and provides a baseline for comparison with more complex alternatives. It is calculated using the Haversine formula, as implemented in the geosphere package Hijmans et al (2023).

Travel Time

While the beeline distance serves as a straightforward starting point, it may not reflect real-world travel patterns. For example, two locations separated by a river may have a short beeline distance but require a significantly longer journey via a bridge. Additionally, regions with high volumes of traffic often have well-developed infrastructure, such as highways, which influence travel behaviour. These factors make travel time a more realistic measure of the likelihood of interaction between individuals.

Travel times are calculated using the osrm package Giraud (2022), which computes road travel times based on OpenStreetMap data. This package supports various modes of travel; here, car travel times are used. Although travel time provides valuable additional information, it is computationally more demanding than the beeline

distance due to the need for route calculations.

Gravity Model

Individuals are more likely to travel to areas with higher population densities, reflecting the gravitational pull of urban centres. The gravity model incorporates this principle by combining the beeline distance between two locations, D_{ji}, with population densities at the origin (P_j) and destination (P_i):

$$D_{ji}^{-d_1} P_j^{-d_2} P_i^{d_3}$$

Assuming the triple $(D_{ji}, P_j, P_i)'$ is jointly Lognormal, this formula has a closed-form expectation. However, assuming (mixtures of) truncated Lognormal distributions for at least one element of the triple jeopardises the closed-form expectation. Further, preliminary analyses suggest that P_i and P_j are not well-approximated by a mixture of bivariate Lognormals. To simplify the model, we assume equality of d_2 and d_3. Hence, we include the population ratio $P_{ji} = P_j/P_i$ which fits the Lognormal assumption more closely. Instead of the triple above, we measure the distance as $(D_{ji}, R_{ji})'$ and translate it to a contact intensity $D_{ji}^{-d_1} P_{ji}^{-d2}$.

Circle Distance

The circle distance expands on the gravity model by considering the population within a circular area around the origin. For a given destination, the measure accounts for all inhabitants within a circle centred at the origin and touching the destination. This approach captures the attractiveness of a destination relative to alternative locations within the same range.

Using high-resolution population data on a 100m grid enables precise calculation of circle populations, but the process is computationally intensive. For efficiency, we recommend different resolutions depending on the proximity of locations: a 100m grid for intra-district distances and a coarser 1km grid for inter-district distances. Adjustments to population counts are made to account for the discrete nature of grid-based computations, ensuring accurate results.

Radiation Model

The radiation model builds on the circle distance by incorporating the populations at both the origin and destination. It evaluates transmission risk using the following formula:

$$D_{ji} = \frac{(P_j + C_{ji})(P_i + P_j + C_{ji})}{P_i P_j}$$

where C_{ji} is the population within a circle around the origin that touches the destination. Traditionally used to estimate flux between locations in non-parametric settings, the radiation model here serves as a distance measure. It balances simplicity with the ability to account for population distributions, making it a natural competitor to the gravity model.

Simplifications

In addition to above weight matrix formulation, we add two levels of simplification. First, we add a version where transformation and expectation are inverted. For the beeline distance measure D_{ji}, the weights are $\mathrm{E}(D_{ji})^{-d}$ instead of $\mathrm{E}(D_{ji}^{-d})$. The resulting weights lose their interpretation as an average transmission risk between individuals. Instead, they can be seen as a transmission risk between average individuals of districts. The advantages are lower computational complexity as well as less memory needed to store the expectations $\mathrm{E}(D_{ji})$ compared to storing parameters of a mixture distribution. Although this simplification is feasible for all distance measures, we only include it for the beeline distance and the gravity distance. For the latter, the reformulation is similar and transforms expected distance and the expected population ratio to a transmission risk.

This simplification still uses detailed population density data. An additional step of simplification is ignoring population density data and base weights on the centroids of districts. Then, we can apply the power law to the distance between centroids to quantify transmission risk. We include weights based on the beeline distance between centroids as well as a centroid gravity model that accounts for the total population in districts. Similar to above fine grid approach, we define those centroid gravity weights as product of power laws applied to the beeline distance and the population ratio. To maintain the ability to express intra-district transmission risk, we introduce a new parameter for this approach such that the raw weights are $W_{ji} = \omega > 0$ for $j = i$. The approach is equivalent to extending the endemic epidemic model to three components where the autoregressive and neighbourhood epidemic component have equal loglinear predictors.

3 Application

The new approach is applied to influenza infections in Germany from 2001 to the 45th week of 2020. The data, sourced from Robert Koch Institute (2024), contains the number of weekly infections reported for the 401 administrative districts. The observation period overlaps with two relevant epidemic outbreaks: the 2009 swine-flu outbreak and COVID-19.

Data from the period of the H1N1 pandemic, the 52 weeks starting from the 17th week of 2009, are excluded from the analysis, as these cases likely do not follow with the transmission mechanisms captured by the model. This starting point is chosen based on the lowest infection count between seasonal peaks. The overlap with COVID-19 infection is ignored in the study, as the first minor outbreaks in Germany started at the beginning of 2020 where influenza infections had already declined to endemic levels.

The 401 districts are grouped into four regions (north, east south and west) and served two purposes:

- Reducing computational complexity, as the numbner of transmission paths grows quadratically with the number of districts.
- Exploring how model performance varies across regions with distinct spatial and demographic characteristics.

The northern region includes Schleswig-Holstein, Hamburg, Bremen, and Lower Saxony. The east is defined as Berlin, Brandenburg, Mecklenburg-Western Pomerania, Saxony, Saxony-Anhalt and Thuringia. The south comprises Bavaria and Baden-Württemberg, and all remaining states form the west. Figure 1 provides an overview. These regions differ significantly in characteristics such as population size, urbanisation, and spatial distribution. For example, the northern region includes sparsely populated rural areas as well as isolated urban centres like Hamburg and Bremen. The west exhibits a cluster of densely populated areas in the Ruhr region. The largest districts by area are sparsely populated and located in eastern Germany alongside the largest district by population: Berlin. In the south, we observe many nested districts, where urban districts are surrounded by their corresponding rural district.

For each of the four regions, we compare eleven ways to define spatial dependence. As a baseline, the first model assumed no spatial dependence, i.e. a diagonal weight matrix. The second is the power law approach applied to the order of neighbourhood. These approaches will be abbreviated as None and NGB in the following. Further, we include the five distance measures of the fine grid approach introduced in the previous section based on the beeline distance (FG-B), the travel time (FG-T), a gravity attraction (FG-G), the circle population (FG-C) and the radiation formula (FG-R) respectively. Simplified versions of the beeline and gravity approach will be denoted as FGs-B and FGs-G for the simplified fine grid approach and Cent-B and Cent-G for the centroid based weights.

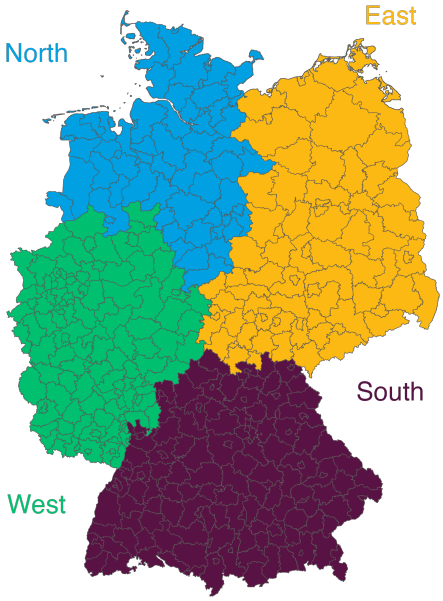


Fig. 1: Grouping of Germany's 401 districts into four regions

These approaches are evaluated in terms of their in-sample performance as well as in a forecast study. The out-of-sample performance is assessed for a rolling window of around 18 years of data for estimation and forecasting up to 8 weeks into the future before shifting the window by one week. This results in around 100 time windows, or two years.

To be able to use the FG approach, we need to fit a mixture of (truncated) LogNormal distributions for each distance measure and each pair of districts that is in the same region. The following steps are carried out:

- Sampling 2000 pairs of individuals, one from each district, weighted by the population density.
- Compute the distance measure for each pair.
- Fit a mixture of truncated Lognormal distributions to the sample of 2000 distances using the EM algorithm, where the number of mixture components is determined by information criteria and its upper bound set to five components.
- Saving the estimated parameters and mixing weights.

Sampling 2000 individuals per districts results in 802,000 individuals in total, or around 1% of the total population. We found the number of samples sufficient to obtain stable estimates of LogNormal parameter estimates while maintaining computational feasibility. Figure F2 in the appendix summarises the number of mixing components. We see that the upper bound of five components suffices. The radiation model tends to have few components. For all four regions, the distribution of radiation distances is well approximated using only one component for most pairs of districts with an average of 1.5 components. The bivariate gravity distance measure needs most mixing components with 2.7 on average.

The Endemic-Epidemic model is chosen to have a similar architecture across all four regions and for each distance measure. Fixed effects for the districts are included in both the endemic and epidemic component. In the endemic component, we further assume constant population sizes, such that the component reduces to ν_{it}. Harmonic waves with one year periodicity are included to capture seasonality. The number of harmonic waves for each distance measure is determined in a forward selection according to the BIC in the application to the training data that is not predicted in the forecast performance assessment. Starting with a model that omits seasonality terms, the order of the harmonic waves is increased by one, separately in the endemic and epidemic component and the largest decrease in the BIC determines which order increase is selected. The stepwise selection stops if an increase of the seasonality orders does not improve the BIC. Overdispersion parameters are assumed to be the same for all districts in a region.

For each distance measure, four variations are explored. With and without population scaling, i.e. accounting for heterogeneity in population sizes of districts at risk, as well as with or withour truncating the power law for small distances.

After fitting a model, the prediction generates a total of 1000 chains of length 8 into the future following the fitted model. From this approximation of the predictive distributions, we save quantiles (2.5%, 5%, 10%, 50%, 90%, 95% and 97.5%) and score

the prediction using the weighted interval score (WIS) and the ranked probability score (RPS).

The interval score for a level $\alpha \in (0, 1/2)$ is defined as

$$\mathrm{IS}(y, \alpha) = (U_\alpha - L_\alpha) + \frac{2}{\alpha}(L_\alpha - y)\mathbb{1}\left\{y < \hat{Y}_{\alpha/2}\right\} + \frac{2}{\alpha}(y - U_\alpha)\mathbb{1}\left\{y > U_\alpha\right\}$$

where $U_\alpha = \hat{Y}_{1-\alpha/2}$ and $L_\alpha = \hat{Y}_{\alpha/2}$ are bounds of the prediction interval, $\hat{Y}_\kappa$ is the κ quantile of the predictive distribution and y is the observation. For a set of multiple levels $\alpha_1, ..., \alpha_k$ and the median prediction $\hat{Y}_{1/2}$, the weighted interval score is

$$\mathrm{WIS}(y, \alpha_1, ..., \alpha_K) = \frac{1}{K+1}\left(\frac{1}{2}|y - \hat{Y}_{1/2}| + \sum_{k=1}^{K} \frac{\alpha_k}{2}\mathrm{IS}(\alpha_k, y)\right) .$$

The RPS is computed as

$$\mathrm{RPS}(y, \hat{y}_1, ..., \hat{y}_N) = \frac{1}{N}\sum_{i=1}^{K} |\hat{y}_i - y| - \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} |\hat{y}_i - \hat{y}_j|$$

where $\hat{y}_i$ are the generated samples of the chains. Both prediction scores are computed for forecasts of different horizons. Additionally, the log-transformed forecast is evaluated with the WIS to account for sharp increases and differences in the level of the time series. Regardless of the scoring rule forecasts are evaluated with, we can translate average scores for models to relative skills. For that purpose, we compute θ_{ij} as the ratio of average scores of models i and j and define the skill of model i as $\theta_i = (\prod_j \theta_{ij})^{1/M-1}$ where M is the number of models. Choosing a reference model i^*, we can then rescale θ_i/θ_{i^*} to obtain the skill of model i compared to the reference model.

We aim to assess the usefulness of the new method in three areas: In terms of additional insight into the dynamics of disease transmission, their forecast performance, and a more detailed spatial risk assessment.

3.1 In-Sample Performance

The novel method provides flexibility in defining weight matrices. To narrow down the selection of models included in the rolling window forecast evaluation, we begin by inspecting the in-sample performance. In a preliminary analysis, we observed that scaling raw weights by population sizes results in a lower BIC for all distance measures and across all four regions. On the other hand, truncating the power law - and thus introducing an additional parameter - does not consistently reduce the BIC. In cases where truncating the power law does reduce the BIC, the improvement is small. Therefore, for the following analysis, we focus on models with population scaling and an untruncated power law.

Table 1 summarises the selected seasonality orders and the BIC from fitting to the complete data set. The seasonality orders differ slightly between regions and distance measures, but is two or three except for Northern Germany. According to the BIC, the centroid gravity model has the best in-sample fit for Northern Germany, in the other three regions, the radiation approach outperforms the other approaches. In terms of levels of simplicity, we see that no simplification dominates the two other approaches when using the beeline distance. With a gravity distance measure, both simplifications have a lower BIC compared to the Fine Grid approach. In all other regions, the Fine Grid approach has a lower BIC that the simpler versions.

In addition to information criteria, Table 2 summarises the estimated decay parameters along with their 95% confidence intervals. Across regions, the estimates for each distance measure are comparable when not using the centroid approach. The similarity in decay parameter estimates and the selected seasonality orders indicates that an application of those models to all 401 spatial units jointly is an interesting alternative. For the two distance measures based on the geodesic distance between individuals - the beeline and gravity model - the decay estimates are between 1.4 and 1.6, which aligns with the findings of Brockmann et al (2006).

For the gravity models, we see that some estimates for the second decay parameter are negative, implying that a transmission from an infected person living in a less densely populated districts is less likely to a district with high population density. This counterintuitive result emphasises a limitation of the gravity model and other models with asymmetric weights. The transmission risks are based on the distance between an infected person and a destination the person might travel to. However, transmission can also occur when a susceptible person travels to a destination of an infected person. Accounting for those transmission paths is an interesting and necessary future extension.

Table 1: Selected seasonality order (Endemic, Epidemic) and the corresponding BIC from fitting to complete data

	North		East		South		West	
Model	Order	BIC	Order	BIC	Order	BIC	Order	BIC
None	(2,2)	85,972	(3,3)	143,807	(2,3)	220,688	(3,3)	162,591
NGB	(1,3)	80,144	(3,2)	135,720	(3,3)	204,397	(3,2)	150,741
FG-B	(1,2)	79,965	(3,2)	135,539	(3,3)	204,174	(3,2)	150,306
FGs-B	(1,3)	80,197	(3,2)	136,008	(3,3)	204,833	(3,2)	150,701
Cent-B	(1,2)	80,003	(3,2)	135,915	(3,3)	204,522	(3,2)	150,850
FG-T	(1,2)	79,908	(3,2)	136,644	(3,3)	205,756	(3,2)	151,914
FG-G	(1,2)	79,943	(3,2)	135,513	(3,3)	204,168	(3,2)	150,297
FGs-G	(1,3)	80,199	(3,2)	136,013	(3,3)	204,849	(3,2)	150,676
Cent-G	(1,2)	79,805	(3,2)	135,926	(3,3)	204,512	(3,2)	150,583
FG-C	(1,2)	80,172	(3,2)	135,793	(3,3)	204,944	(2,2)	150,839
FG-R	(1,2)	80,047	(3,2)	135,378	(3,3)	203,946	(3,2)	150,200

Table 2: Estimated decay parameters and 95% confidence intervals from fitting to complete data

	North		East		South		West	
Model	Est.	CI	Est.	CI	Est.	CI	Est.	CI
NGB	2.789	± 0.068	2.914	± 0.060	2.820	± 0.042	2.986	± 0.052
FG-B	1.435	± 0.025	1.448	± 0.021	1.472	± 0.016	1.573	± 0.020
FGs-B	2.062	± 0.060	2.012	± 0.056	1.981	± 0.038	2.208	± 0.046
Cent-B	1.093	± 0.104	0.452	± 0.102	0.514	± 0.080	0.348	± 0.116
FG-T	1.927	± 0.032	1.917	± 0.032	1.965	± 0.025	1.972	± 0.030
FG-G	1.428	± 0.025	1.435	± 0.020	1.466	± 0.017	1.566	± 0.021
	0.463	± 0.110	0.388	± 0.107	0.047	± 0.155	-0.130	± 0.207
FGs-G	2.079	± 0.061	2.026	± 0.056	1.993	± 0.038	2.211	± 0.046
	-0.281	± 0.158	-0.246	± 0.152	-0.160	± 0.148	-0.779	± 0.237
Cent-G	1.228	± 0.107	0.451	± 0.103	0.513	± 0.081	0.368	± 0.111
	0.867	± 0.115	0.047	± 0.148	0.398	± 0.153	1.113	± 0.119
FG-C	0.860	± 0.017	0.848	± 0.014	0.809	± 0.011	0.829	± 0.012
FG-R	0.424	± 0.007	0.421	± 0.006	0.411	± 0.004	0.429	± 0.005

3.2 Forecast Performance

The out-of-sample forecast performance is assessed in different ways. We compare the performance across methods, regions, and three different scores: the Weighted Interval Score (WIS) applied to the original forecasts, the WIS applied to log-transformed forecasts (lWIS), and the Ranked Probability Score (RPS). The forecast horizons vary between one and eight weeks.

We begin by examining the performance for one-week-ahead forecast. In total, we have 100 forecasts to be evaluated for each district. Forecast performance is aggregated over forecast windows or regions by averaging the performance measures of individual forecasts. Table 3 presents the aggregated performance measures as relative skills, with the model exhibiting the lowest (i.e. best) skill is highlighted in bold. The results show that all models perform better in one-week-ahead forecasts than a model with no spatial correlation. The centroid gravity approach outperforms fine-grid model approaches in the west regardless of the scoring rule. In the south and east, the circle population approach performs best when a scoring rule on the original scale is used. In the northern region, the centroid gravity approach performs best for scoring rules on the original scale and the travel time approach when applying the WIS on the log scale.

Next, we evaluate the performance for longer forecast horizons. Table 4 presents the weighted interval score on a log scale for forecast horizons of 1, 4, and 8 weeks. The methods are ranked according to their performance. In the first block - for the one-week-ahead forecast - we observe the same trends as previously. For the four-week-ahead forecast, the best-performing models are either the circle population measure or the centroid gravity approach. The former is consistently in the best two performing models whereas the centroid gravity model is ranked last for the eastern region. Except for the northern region, most models with spatial correlation do not outperform

the reference model without spatial correlation. Only the circle population measure performs better in all four regions.

For eight-week-ahead forecasts, we see that the best performing model in three regions is the approach without spatial correlation. The circle population model is in the top three performing models in all four regions. The poor performance of spatial models could be due to the complex nature of epidemic behaviour, where predicting infection counts two months in advance proves to be highly challenging. For longer forecast horizons, it seems that forecasting a typical infection count for the region is the most reliable approach.

Table 3: One week ahead forecast performance - Relative skill (in %) compared to all other models, where the model with no spatial correlation serves as reference model

	North			East			South			West		
Model	WIS	lWIS	RPS	WIS	lWIS	RPS	WIS	lWIS	RPS	WIS	lWIS	RPS
NGB	91.1	83.1	91.4	93.3	85.5	92.7	88.8	81.6	91.0	94.3	81.4	94.1
FG-B	90.4	81.7	91.1	92.5	84.6	92.2	88.4	81.4	90.9	95.4	81.4	94.6
FGs-B	91.4	83.2	91.4	93.5	85.4	93.4	88.9	82.2	91.5	95.6	81.9	95.0
Cent-B	90.0	82.0	91.2	92.9	86.6	92.9	89.0	80.9	90.8	97.9	82.7	96.8
FG-T	90.4	**81.5**	91.1	94.6	87.7	95.0	90.3	82.2	91.9	97.2	84.4	96.7
FG-G	89.9	81.7	90.1	92.5	**84.5**	92.3	88.5	81.4	90.9	95.3	81.2	94.6
FGs-G	91.5	83.2	91.4	93.1	85.3	93.1	88.9	82.6	91.4	95.2	81.7	94.5
Cent-G	**88.7**	81.7	**89.4**	92.7	86.6	92.9	88.4	**80.7**	90.1	**91.3**	**80.1**	**91.3**
FG-C	90.4	83.4	90.7	**92.2**	85.5	**91.9**	**87.1**	81.6	**89.3**	93.3	88.4	93.4
FG-R	89.6	81.7	90.8	92.3	84.6	92.0	88.4	81.2	90.7	95.1	81.0	94.6

3.3 Risk Mapping

So far, we have evaluated the in-sample and out-of-sample forecast performance, focusing on predicting the number of new cases at the district level. All fine-grid approaches began by defining distances at the individual level, subsequently aggregating transmission risks to the district level. Once a model is fitted, this process can be reversed, allowing us to derive the expected number of new infections at a finer resolution, instead of districts, to give a detailed risk map.

The expected number of new infections for a grid cell is determined by separately calculating the expected number of new endemic and epidemic infections. For endemic infections in a district i, the expected number is given by ν_{it}, which is then distributed among the district's grid cells in proportion to their population density. The calculation of expected weekly epidemic infections is carried out by computing the expected number of secondary infections in every grid cell caused by one infected individual residing in that district. To do so, iterate over all grid cells within the district of origin and calculate distances to all other grid cells within the study area. Translate these distances into contact intensities, e.g. apply the power law. Compute a weighted average where the weights are proportional to the population in the grid cell of origin. This yields the contact intensity from the district of origin to all grid cells. Normalise the contact intensities to represent the probability distribution of expected contacts

Table 4: Ranking for forecast horizons

	North		East		South		West	
Rank	Model	lWIS	Model	lWIS	Model	lWIS	Model	lWIS
				h = 1				
1	FG-T	12.06	FG-G	14.19	Cent-G	13.01	Cent-G	13.34
2	FG-G	12.09	FG-B	14.19	Cent-B	13.04	FG-R	13.47
3	FG-R	12.09	FG-R	14.20	FG-R	13.09	FG-G	13.51
4	FG-B	12.09	FGs-G	14.30	FG-B	13.11	FG-B	13.53
5	Cent-G	12.10	FGs-B	14.32	FG-G	13.12	NGB	13.53
6	Cent-B	12.13	FG-C	14.33	FG-C	13.14	FGs-G	13.58
7	NGB	12.28	NGB	14.34	NGB	13.15	FGs-B	13.61
8	FGs-G	12.29	Cent-G	14.50	FG-T	13.23	Cent-B	13.73
9	FGs-B	12.29	Cent-B	14.50	FGs-B	13.23	FG-T	13.98
10	FG-C	12.32	FG-T	14.67	FGs-G	13.29	FG-C	14.59
11	None	14.53	None	16.53	None	15.81	None	16.32
				h = 4				
1	Cent-G	26.71	FG-C	32.77	FG-C	31.97	Cent-G	34.58
2	FG-C	27.48	None	33.61	Cent-G	32.97	FG-C	36.94
3	FG-R	28.18	FGs-G	33.95	None	33.39	None	37.37
4	FG-G	28.64	NGB	33.98	Cent-B	33.41	FG-B	38.30
5	FG-B	28.79	FG-R	33.99	NGB	33.50	NGB	38.31
6	FG-T	28.82	FGs-B	34.02	FG-R	33.51	FG-R	38.41
7	FGs-G	28.87	FG-B	34.02	FG-G	33.66	Cent-B	38.47
8	FGs-B	29.08	FG-G	34.06	FG-B	33.68	FGs-G	38.59
9	NGB	29.43	FG-T	34.63	FG-T	33.78	FG-G	38.74
10	Cent-B	29.69	Cent-B	34.87	FGs-G	33.87	FGs-B	38.79
11	None	32.25	Cent-G	34.97	FGs-B	33.90	FG-T	39.02
				h = 8				
1	Cent-G	37.51	None	50.20	None	48.09	None	51.32
2	FG-C	39.45	FG-C	51.33	FG-C	50.26	Cent-G	53.81
3	None	41.11	NGB	52.53	Cent-G	53.18	FG-C	54.89
4	FG-G	42.32	FGs-B	52.75	NGB	53.82	FG-B	56.48
5	FG-R	43.07	FG-B	52.90	FG-R	53.87	FGs-G	57.81
6	FGs-G	43.50	FG-G	53.02	FG-G	54.20	NGB	57.97
7	FG-T	43.51	FGs-G	53.04	FGs-B	54.25	FGs-B	58.00
8	FG-B	43.87	FG-R	53.96	FGs-G	54.26	FG-R	58.13
9	FGs-B	43.89	FG-T	54.36	FG-B	54.31	FG-G	58.14
10	NGB	44.17	Cent-B	55.05	Cent-B	54.39	Cent-B	58.17
11	Cent-B	45.27	Cent-G	55.21	FG-T	54.47	FG-T	58.89

from one person in the district of origin to all other grid cells. Scale these probabilities by the observed number of cases in the district during the current week and sum them across all districts. The resulting value represents the aggregated risk from all observed cases across districts, which is then scaled by the corresponding epidemic log-linear predictor λ_{it}.

By summing the expected endemic and epidemic infections, we obtain the total risk for each grid cell. This risk measure highlights areas with a high expected number of cases in absolute terms, correlating with population density. To derive an individual

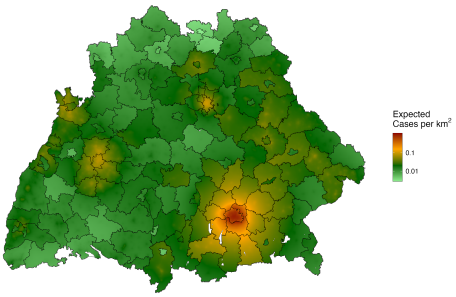


Fig. 2: Number of expected cases in Southern Germany in week 9 of 2019 per square kilometer based on beeline distance approach

Fig. 3: Individual risk of infection in Southern Germany in week 9 of 2019 based on the beeline distance approach

risk map, which reflects the per-person risk of infection, the total risk in a grid cell is divided by the number of individuals residing there.

As an example, we assess the risk for Southern Germany in week 9 of 2019, based on a fitted model that uses the beeline distance between individuals. This week was selected because it is the first week after the peak of 2019, and 2019 was chosen as the most recent year with complete observations. The beeline distance was selected because it is easy to compute. Figure 2 presents the overall risk for grid cells, revealing that areas around Munich exhibit the highest risk due to both the large number of infections in the district and its high population density.

Figure 3 illustrates the individual risk, showing variation across grid cells within single districts. This detailed risk mapping approach cannot currently be assessed without information on the location of infected individuals within districts. However, evaluating its performance using synthetic data in a simulation study is an interesting future research direction.

Detailed risk mapping provides a valuable tool for public health planning and intervention. Unlike district-level predictions, which are constrained by administrative boundaries, forecasts on grid cells can be adapted to the specific applications. For example, this approach enables the generation of forecasts for hospital catchment areas, which often do not align with district boundaries. Such tailored forecasts could then help hospital administrators allocating resources more effectively.

4 Conclusion

This study has assessed the performance of novel spatial approaches for forecasting infection dynamics. By incorporating individual-level distance measures and aggregating to district-level risks, these methods have shown robust predictive capabilities. They models often outperform the model based on neighbourhood orders but they do not consistently outperform the centroid gravity approach in short term forecasting.

Especially the circle population distance measure appeared to have reasonable forecast performance. Extending this general idea of intervening opportunities not only to include the population structure but other driving factors of human migration is an interesting research direction.

In addition to their forecasting strength, detailed risk mapping offers a significant enhancement, allowing for a nuanced understanding of spatial heterogeneity and highlighting areas at higher risk both in absolute terms and on a per-capita basis. These insights extend the usefulness of the models beyond administrative boundaries.

Despite these advances, this article also highlights that there is still much to explore. The application considered only one disease and one country. Assessing the robustness of the new method when applied to other diseases or countries can and should be addressed in future research.

The proposed methods are not without their challenges. The estimation procedures are sensitive to the choice of starting values, which can affect convergence and stability. Additionally, the preparation required for estimation is computationally intensive, particularly when dealing with large datasets or high-resolution spatial grids. While the new approaches generally perform well, their forecast performance is comparable to the established neighbourhood order-based method in certain scenarios.

4.1 Outlook

There are numerous directions for future research to address these limitations and expand the utility of the proposed methods:

Incorporating alternative distance measures into radiation models could provide insights into how spatial interactions vary under different assumptions. The validity of current assumptions should be investigated in more detail, such as the choice of the Lognormal distribution or translating distances to contact intensity with a power law. More flexible approaches for the latter, such as splines, could be an interesting alternative.

Similar to the detailed risk map, it would be an advantage for retrospective analysis and forecasting to identify clusters or hotspots invariant to district boundaries. Generally, the detachment from the district boundaries in modelling provides more flexibility for applications where district boundaries changed during the observation period. Whether the new approach is robust to such boundary changes can be assessed in a simulation study.

Incorporating additional covariates could offer a more comprehensive understanding of transmission dynamics and risk factors. For example, the age and gender structure of districts could be incorporated in the log-linear predictors or the definition of the weight matrix.

By addressing these areas, future work can build on the foundation laid in this study to develop even more accurate, flexible, and interpretable models for forecasting and risk assessment in epidemic contexts.

Declarations

Funding: SF and MS were supported by Wellcome (210758/Z/18/Z).

Conflict of interest: The authors declare no conflict of interest.

Data availability: All data used are publicly available.

Code availability: General functions to apply the methods are available in a toolbox at https://github.com/ManuelStapper/FGSIM. Application-specific code is available on request.

Appendix A Motivating Examples

Using the order of neighbourhood as a measure of closeness between district is simple, but entails pitfalls practitioners need to be aware of. To highlight some, consider the examples shown in Figure A1.

For some applications, there are reforms in the district boundaries during the observation period. In our application, the districts Göttingen and Osterode am Harz were aggregated in 2016. Figure A1a shows in black outline the current district and in red the former boundary. For the two spatial units highlighted in yellow, the order of neighbourhood changed due to the aggregation and thus how we quantify the transmission risk between the two districts. For us, district reforms have not been problematic, as the number of new infections is reported ob the basis of the current districts. However, that is not always the case for other applications. Simple district aggregations can be accounted for by summing up the number of cases in retrospect, but it is not straightforward how to deal with a district being split when only the total number of cases is available for times before the split.

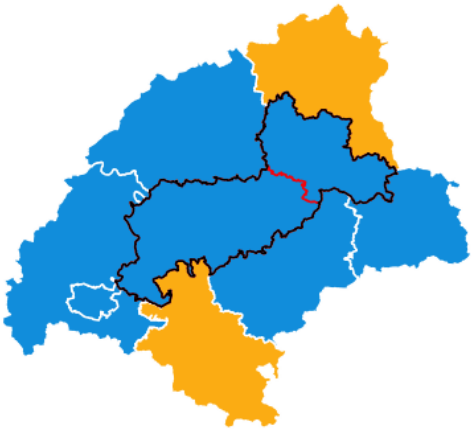
Another example for the sensitivity of neighbourhood orders as distance measure is how we decide for an area of application. Figure A1b shows districts in West Germany (green) and districts in Nothern Germany (blue). In the application we analyse the data separately. For the pair of spatial units highlighted in black outline, Steinfurt and Minden-Lübbecke, the order or neighbourhood is four. If we extend to study area and also include districts in Northern Germany, the order is only two, while most other orders remain the same. Thus, the transmission risk between the two districts is evaluated differently compared to other pairs of districts.

Figure A1c shows two pairs of two districts each both with a neighbourhood order of one. Hence, we would assume similar transmission risk between them. However, the pair on the right is larger in area compared to the pair on the left, where one districts is fully surrounded by the other. It is plausible that the transmission risk between the left pair is more likely than for the right pair. While on the left, individuals live at most 32 kilometers apart with an average of around 13, two individuals from the right pair may live up to 198 kilometers apart with an average of 106.

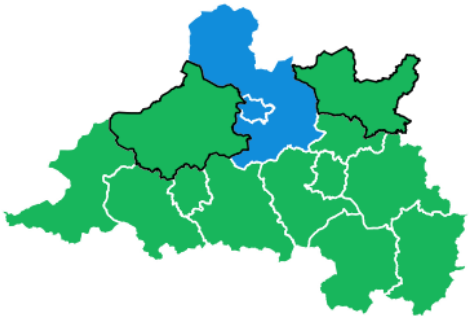
Example 4, shown in Figure A1d, displays four districts in Eastern Germany. The two districts at the bottom left share only a short border, while the other two districts'

border is substantially longer. Transmission risk is again assumed similar when using neighbourhood orders.

Finally, example 5 in Figure A1e shows a comparison of two pairs where the colours give the population density on a log scale. On the left, we see Mainz and Wiesbaden, to urban districts where the population of both is clustered near their border. On the left, we see the rural districts Uckermark and Vorpommern-Greifswald. The population is more spread out with the most densely populated are in the more northern district is location far away from the border. Such differences are not accounted for when basing transmission risk on the order of neighbourhood.



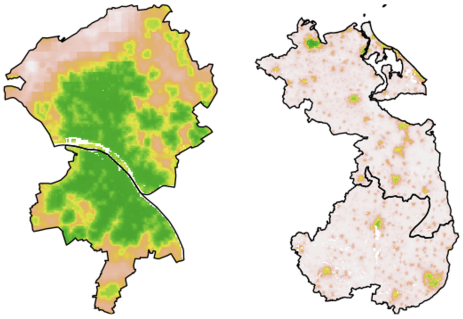
(a) Example 1 - Aggregation of districts



(b) Example 2 - Effect of study region

(c) Example 3 - Heterogeneity in area

(d) Example 4 - Heterogeneity in joint boundary length



(e) Example 5 - Heterogeneity in population density

Fig. A1: Comparison of different heterogeneity effects

Appendix B Derivation of Weights

In the following, the derivatives of weight matrices are summarised. Given a distance measure D_{ij}, we define unnormalised weights as the expectation $\mathrm{E}\left[D_{ij}^{-d}\right]$ regardless of what D_{ij} measures, the beeline distance, route time or radiation measure between individuals. If the power law shall be truncated, raw weights are the expectation $\mathrm{E}\left[D_{ij}^{-d} \wedge \exp\{-d\delta\}\right]$, where "$\wedge$" is the minimum operator. The distribution of D_{ij} is assumed to be either a Lognormal distribution, a truncated Lognormal distribution or a mixture of multiple (truncated) Lognormal distributions. Since truncation boundaries account for the minimum and maximum distance, they are assumed to be the same for all mixing components, if considered. All distributions considered are special cases of a mixture of truncated Lognormal distributions. Further, the regular power law is a limiting case of the truncated power law for $\delta \to 0$. Thus, we summarise derivatives of raw weights for the most general case, however, we ignore mixture distributions to simplify notation.

Let W be a weight such that

$$W = \mathrm{E}\left[D^{-d} \wedge \exp\{-d\delta\}\right]$$

where D is the distance measure, d and δ are parameters. D follows a truncated Lognormal distribution with parameters μ, σ^2 that takes values between $\exp(l)$ and $\exp(u)$. If $l < \delta < u$, we split above expectation

$$W = P\left(\log(D) \in [l, \delta]\right) \exp\{-d\delta\} + P\left(\log(D) \in [\delta, u]\right) \mathrm{E}\left[D^{-d} \middle| \log(D) > \delta\right] \ .$$

Then, the probabilities are

$$P\left(\log(D) \in [l, \delta]\right) = \frac{\Phi\left(\frac{\delta-\mu}{\sigma}\right) - \Phi\left(\frac{l-\mu}{\sigma}\right)}{\Phi\left(\frac{u-\mu}{\sigma}\right) - \Phi\left(\frac{l-\mu}{\sigma}\right)}$$

$$P\left(\log(D) \in [\delta, u]\right) = \frac{\Phi\left(\frac{u-\mu}{\sigma}\right) - \Phi\left(\frac{\delta-\mu}{\sigma}\right)}{\Phi\left(\frac{u-\mu}{\sigma}\right) - \Phi\left(\frac{l-\mu}{\sigma}\right)}$$

To find the conditional expectation, we see that $D^{-d} | \log(D) > \delta$ follows a truncated LogNormal distribution with parameters $-d\mu$ and $d^2\sigma^2$ truncated between $-du$ and $-d\delta$. Thus, according to Wang et al (2012), the expectation is

$$\mathrm{E}\left[D^{-d} \middle| \log(D) > \delta\right] = \exp\left\{-d\mu + \frac{d^2\sigma^2}{2}\right\} \frac{\Phi\left(\frac{-d\delta+d\mu}{d\sigma} - d\sigma\right) - \Phi\left(\frac{-du+d\mu}{d\sigma} - d\sigma\right)}{\Phi\left(\frac{-d\delta+d\mu}{d\sigma}\right) - \Phi\left(\frac{-du+d\mu}{d\sigma}\right)}$$

$$= \exp\left\{-d\mu + \frac{d^2\sigma^2}{2}\right\} \frac{\Phi\left(\frac{-\delta+\mu-d\sigma^2}{\sigma}\right) - \Phi\left(\frac{-u+\mu-d\sigma^2}{d\sigma}\right)}{\Phi\left(\frac{-\delta+\mu}{\sigma}\right) - \Phi\left(\frac{-u+\mu}{\sigma}\right)}$$

$$= \exp\left\{-d\mu + \frac{d^2\sigma^2}{2}\right\} \frac{\Phi\left(\frac{u-\mu+d\sigma^2}{d\sigma}\right) - \Phi\left(\frac{\delta-\mu+d\sigma^2}{\sigma}\right)}{\Phi\left(\frac{u-\mu}{\sigma}\right) - \Phi\left(\frac{\delta-\mu}{\sigma}\right)}$$

To sum up, we have

$$\begin{aligned} h \cdot W = &\left(\Phi\left(\frac{\delta-\mu}{\sigma}\right) - \Phi\left(\frac{l-\mu}{\sigma}\right)\right) \exp\{-d\delta\} \\ &+ \left(\Phi\left(\frac{\delta-\mu+d\sigma^2}{\sigma}\right) - \Phi\left(\frac{l-\mu+d\sigma^2}{\sigma}\right)\right) \exp\left\{-d\mu + \frac{d^2\sigma^2}{2}\right\} \end{aligned}$$

where $h := \Phi\left(\frac{u-\mu}{\sigma}\right) - \Phi\left(\frac{l-\mu}{\sigma}\right)$.

So far, we have restricted to the case $l \leq \delta \leq u$. Above findings can be generalised by defining

$$m(\delta) = \begin{cases} \delta & & l \leq \delta \leq u \\ l & \text{if} & \delta < l \\ u & & \delta > u \end{cases}$$

Finally we have weights for the general case in dependence of parameters δ, and d

$$\begin{aligned} W(\delta, d) &= \frac{1}{h}\left[p(\delta)f(\delta, d) + q(\delta, d)g(d)\right] \\ h &= \Phi\left(\frac{u-\mu}{\sigma}\right) - \Phi\left(\frac{l-\mu}{\sigma}\right) \\ p(\delta) &= \Phi\left(\frac{m(\delta)-\mu}{\sigma}\right) - \Phi\left(\frac{l-\mu}{\sigma}\right) \\ f(\delta, d) &= \exp\{-d\delta\} \\ q(\delta, d) &= \Phi\left(\frac{m(\delta)-\mu+d\sigma^2}{\sigma}\right) - \Phi\left(\frac{l-\mu+d\sigma^2}{\sigma}\right) \\ g(d) &= \exp\left\{-d\mu + \frac{d^2\sigma^2}{2}\right\} \end{aligned}$$

Appendix C Derivatives of Weights

Above weights are implemented to be used in the *surveillance* package. For numerical optimisation, it is beneficial to use its first two derivatives. To simplify notation, let $f^{(d)}(\delta, d)$ and $f^{(dd)}(\delta, d)$ be the first and second partial derivatives of $f(\delta, d)$ with respect to d and likewise for functions $p(\delta), q(\delta, d)$ and $g(d)$ as well as for derivatives

with respect to δ. Then, we have outer derivatives of $W(\delta, d)$

$$
\begin{aligned}
\frac{\partial}{\partial d} W(\delta, d) &= p(\delta) f^{(d)}(\delta, d) + q^{(d)}(\delta, d) g(d) + q(\delta, d) g^{(d)}(d) \\
\frac{\partial}{\partial d} W(\delta, d) &= p^{(\delta)}(\delta) f(\delta, d) + p^{(\delta)}(\delta) f(\delta, d) + q^{(\delta)}(\delta, d) g(d) \\
\frac{\partial^2}{\partial d^2} W(\delta, d) &= p(\delta) f^{(dd)}(\delta, d) + q^{(dd)}(\delta, d) g(d) 2 q^{(d)}(\delta, d) g^{(d)}(d) + q(\delta, d) g^{(dd)}(d) \\
\frac{\partial^2}{\partial \delta^2} W(\delta, d) &= p^{(\delta\delta)}(\delta) f(\delta, d) + 2 p^{(\delta)}(\delta, d) f^{(\delta)}(\delta, d) + p(\delta) f^{(\delta\delta)}(\delta, d) + q^{(\delta\delta)}(\delta, d) g(d) \\
\frac{\partial^2}{\partial \delta \partial d} W(\delta, d) &= p^{(\delta)}(\delta) f^{(d)}(\delta, d) + p(\delta) f^{(\delta d)}(\delta, d) + q^{(\delta d)}(\delta, d) g(d) + q^{(\delta)}(\delta, d) g^{(d)}(d)
\end{aligned}
$$

$$
\begin{aligned}
f^{(d)}(\delta, d) &= -\delta f(\delta, d) & f^{(\delta d)}(\delta, d) &= (\delta d - 1) f(\delta, d) \\
f^{(\delta)}(\delta, d) &= -d f(\delta, d) & & \\
f^{(dd)}(\delta, d) &= \delta^2 f(\delta, d) & g^{(d)}(d) &= g(d)(-\mu + d\sigma^2) \\
f^{(\delta\delta)}(\delta, d) &= d^2 f(\delta, d) & g^{(dd)}(d) &= g(d) \left[(-\mu + d\sigma^2)^2 + \sigma^2 \right]
\end{aligned}
$$

$$
\begin{aligned}
m^{(\delta)}(\delta) &= \mathbb{1}\{l \leq \delta \leq u\} \\
p^{(\delta)}(\delta) &= \phi\left(\frac{m(\delta) - \mu}{\sigma}\right) \frac{m^{(\delta)}(\delta)}{\sigma} \\
p^{(\delta\delta)}(\delta) &= \phi\left(\frac{m(\delta) - \mu}{\sigma}\right) \frac{(\mu - m(\delta)) m^{(\delta)}(\delta)}{\sigma^3}
\end{aligned}
$$

$$
\begin{aligned}
q^{(d)}(\delta, d) &= \left[\phi\left(\frac{u - \mu + d\sigma^2}{\sigma}\right) - \phi\left(\frac{m(\delta) - \mu + d\sigma^2}{\sigma}\right) \right] \sigma \\
q^{(\delta)}(\delta, d) &= -\phi\left(\frac{m(\delta) - \mu + d\sigma^2}{\sigma}\right) \frac{m^{(\delta)}(\delta)}{\sigma} \\
q^{(dd)}(\delta, d) &= \left[\phi\left(\frac{m(\delta) - \mu + d\sigma^2}{\sigma}\right) \left(m(\delta) - \mu + d\sigma^2\right) - \phi\left(\frac{u - \mu + d\sigma^2}{\sigma}\right) \left(u - \mu + d\sigma^2\right) \right] \sigma \\
q^{(\delta\delta)}(\delta, d) &= \phi\left(\frac{m(\delta) - \mu + d\sigma^2}{\sigma}\right) \frac{(m(\delta) - \mu + d\sigma^2) m(\delta)}{\sigma^2} \\
q^{(\delta d)}(\delta, d) &= \phi\left(\frac{m(\delta) - \mu + d\sigma^2}{\sigma}\right) \frac{(m(\delta) - \mu + d\sigma^2) m^{(\delta)}(\delta)}{\sigma}
\end{aligned}
$$

Appendix D Gravity Model Extension

In case we include gravity effects in the weights, we use that $(D, P)'$ is assumed to follow a mixture of bivariate Lognormal distributions where D is a distance measure and P is the ratio between population at the origin and destination. Again, for the sake of simplicity, weights are derived ignoring a potential mixture distribution. In that case, we have

$$\log \begin{pmatrix} D \\ P \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_D \\ \mu_P \end{pmatrix}, \begin{pmatrix} \sigma_D^2 & \sigma_{DP} \\ \sigma_{DP} & \sigma_P^2 \end{pmatrix} \right)$$

Raw weights are defined as $G := \mathrm{E}\left[D^{-d_1} P^{d_2}\right]$. By assumptions made, $D^{-d_1} P^{d_2}$ follows a Lognormal distribution with expectation

$$G(d_1, d_2) = \exp\left\{ -d_1\mu_D + d_2\mu_P + \frac{d_1^2\sigma_D^2}{2} - d_1 d_2 \sigma_{DP} + \frac{d_2^2\sigma_P^2}{2} \right\} .$$

It derivatives with respect to d_1 and d_2 are summarised as

$$\begin{aligned}
\frac{\partial}{\partial d_1} G(d_1, d_2) &= G(d_1, d_2)(-\mu_D - d_2\sigma_{DP} + d_1\sigma_D^2) \\
\frac{\partial}{\partial d_2} G(d_1, d_2) &= G(d_1, d_2)(\mu_P - d_1\sigma_{DP} + d_2\sigma_P^2) \\
\frac{\partial^2}{\partial d_1^2} G(d_1, d_2) &= G(d_1, d_2)\left[(-\mu_D - d_2\sigma_{DP} + d_1\sigma_D^2)^2 + \sigma_D^2\right] \\
\frac{\partial^2}{\partial d_2^2} G(d_1, d_2) &= G(d_1, d_2)\left[(\mu_P - d_1\sigma_{DP} + \alpha\sigma_P^2)^2 + \sigma_P^2\right] \\
\frac{\partial^2}{\partial d_1 \partial d_2} G(d_1, d_2) &= G(d_1, d_2)\left[(-\mu_D - d_2\sigma_{DP} + d_1\sigma_D^2)(\mu_P - d_1\sigma_{DP} + d_2\sigma_P^2) - \sigma_{DP}\right]
\end{aligned}$$

Appendix E Weight Transformations

From raw weights, whether they are from a distance measure or including gravity effects, we consider three different transformations of weights: 1) normalising the rows of the weight matrix 2) using the parameters on a log scale and 3) scaling the columns. When applying these transformations, the first and second derivatives need to be transformed as well. In the following let W be the matrix of weights before applying a transformation and $W^{(d)}$ as well as $W^{(dd)}$ its first and second derivatives with respect to d respectively. Likewise derivatives are defined for other parameters δ or d_2 and cross-derivatives.

<u>Row normalisation:</u> When the matrix of raw weights is row normalised, we first compute the row sums as $S = W1_{n\times n}$, where $1_{n\times n}$ is an $n \times n$ matrix of ones. This results in a matrix that has the same dimension as W and contains the corresponding row

sums. Thus, the normalised weight matrix can be expressed as $W \oslash S$, where $\oslash$ is the Hadamard division operator. Its derivative with respect to d is simply

$$\frac{\partial}{\partial d}(W \oslash S) = W^{(d)} \oslash S - W \odot S^{(d)} \oslash S^2 \ ,$$

where $\odot$ is the Hadamard product. Derivatives with respect to δ or d_2 are found analogously. The second derivative is

$$\frac{\partial^2}{\partial d^2}(W \oslash S) = W^{(dd)} \oslash S - 2W^{(d)} \odot S^{(d)} \oslash S^2 - W \odot S^{(dd)} \oslash S^2 + 2W \odot \left(S^{(d)}\right)^2 \oslash S^3$$

The cross derivatives with respect to d_1 and d_2 are

$$\begin{aligned}\frac{\partial^2}{\partial d_1 \partial d_2}(W \oslash S) =& W^{(d_2 d_1)} \oslash S - W^{(d_2)} \odot S^{(d_1)} \oslash S^2 - W^{(d_1)} \odot S^{(d_2)} \oslash S^2 \\ & - W \odot S^{(d_2 d_1)} \oslash S^2 + 2W \odot S^{(d_2)} \odot S^{(d_1)} \oslash S^2\end{aligned}$$

<u>Log Transformation</u> If weight matrices are computed using parameters on a log scale, i.e. parameters is transformed by the exponential function before inserted into the weight function, derivatives are simply

$$\frac{\partial}{\partial d} W(\exp(d)) = W^{(d)}(\exp(d)) \cdot \exp(d) \ ,$$

second derivatives

$$\frac{\partial^2}{\partial d^2} W(\exp(d)) = W^{(dd)}(\exp(d)) \odot \exp(2d) + W^{(d)}(\exp(d)) \cdot \exp(d)$$

and cross-derivatives

$$\frac{\partial^2}{\partial d_1 \partial d_2} W(\exp(d_1), \exp(d_2)) = W^{(d_2 d_1)}(\exp(d_1), \exp(d_2)) \cdot \exp(d_1) \exp(d_2) \ .$$

<u>Scaling</u>
If a weight matrix is scaled column-wise, i.e. $W_{ij} \mapsto W_{i,j} \cdot c_j$ for constants c_j, we can simply define C as the diagonal matrix containing c_j. Then, W is transformed to $W \cdot C$ and likewise all derivatives are simply multiplied by C.

Appendix F Application

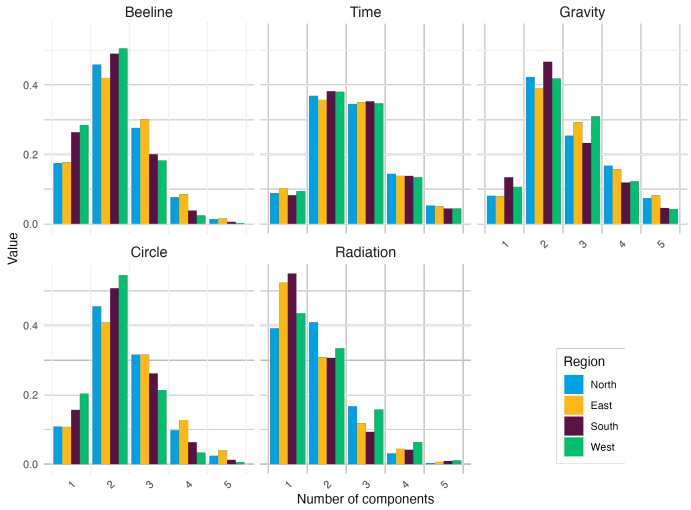


Fig. F2: Number of mixture components split by distance measure and region

References

Alis C, Legara E, Monterola C (2021) Generalized radiation model for human migration. Scientific reports 11(22707). https://doi.org/10.1038/s41598-021-02109-1

Barbosa H, Barthelemy M, Ghoshal G, et al (2018) Human mobility: Models and applications. Physics Reports 734:1–74. https://doi.org/10.1016/j.physrep.2018.01.001

Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. Nature 439(7075):462–465

Giraud T (2022) osrm: Interface Between R and the OpenStreetMap-Based Routing Service OSRM. Journal of Open Source Software 7(78):4574. https://doi.org/10.21105/joss.04574

Haynes KE, Fotheringham AS (1985) Gravity and Spatial Interaction Models. 07, Regional Research Institute, West Virginia University, URL https://ideas.repec.org/b/rri/wholbk/07.html

Held L, Höhle M, Hofmann M (2005) A statistical framework for the analysis of multivariate infectious disease surveillance counts. Statistical modelling 5(3):187–199. https://doi.org/10.1191/1471082X05st098oa

Hijmans RJ, Karney C, Eilliams E, et al (2023) geosphere: Spherical Trigonometry. https://doi.org/10.32614/CRAN.package.geosphere, URL https://CRAN.R-project.org/package=geosphere, r package version 1.5-20

Meyer S, Held L (2014) Power-law models for infectious disease spread. The Annals of Applied Statistics 8(3):1612 – 1639. https://doi.org/10.1214/14-AOAS743

Meyer S, Held L, Höhle M (2017) Spatio-temporal analysis of epidemic phenomena using the R package surveillance. Journal of Statistical Software 77(11):1–55. https://doi.org/10.18637/jss.v077.i11

Robert Koch Institute (2024) SurvStat@RKI 2.0: Online Analysis Tool for Infectious Disease Data. https://survstat.rki.de, accessed: 2024-10-02

Simini F, González MC, Maritan A, et al (2012) A universal model for mobility and migration patterns. Nature 484(7392):96–100. https://doi.org/10.1038/nature10856

Stouffer SA (1940) Intervening opportunities: A theory relating mobility and distance. American Sociological Review 5(6):845–867. https://doi.org/10.2307/2084520

Tizzoni M, Bajardi P, Decuyper A, et al (2014) On the use of human mobility proxies for modeling epidemics. PLOS Computational Biology 10(7):1–15. https://doi.org/10.1371/journal.pcbi.1003716

Wang Y, Dong W, Zhang L, et al (2012) Speed modeling and travel time estimation based on truncated normal and lognormal distributions. Transportation Research Record 2315(1):66–72. https://doi.org/10.3141/2315-07

WorldPop (2018) Global high resolution population denominators project - funded by the bill and melinda gates foundation (opp1134076). https://dx.doi.org/10.5258/SOTON/WP00645, www.worldpop.org - School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur and Center for International Earth Science Information Network (CIESIN), Columbia University

Zipf GK (1946) The p1 p2/d hypothesis: On the intercity movement of persons. American Sociological Review 11(6):677–686. URL http://www.jstor.org/stable/2087063