Details on Expectation-Maximization (EM) Algorithm Estimation Procedure

Let $(x^{(1)}, \dots, x^{(K)})$ denote the epidemic source locations for each $k = 1, \dots, K$ with $x^{(k)} \in \mathbb{R}^2$. Now an arbitrary location $x \in \mathbb{R}^2$ is associated with K sets of polar coordinates $(r^{(1)}, \phi^{(1)}), \dots, (r^{(K)}, \phi^{(K)})$ where the k-th polar coordinate pair indicates the distance $r^{(k)}$ and angle $\phi^{(k)}$ to the kth source point $x^{(k)}$. Applying the model framework given in the main paper to each set of coordinates yields the collection of velocity models

$$\log\left(1 + \frac{r^{(k)}}{g_k(\phi^{(k)})}\right) = -M_k t + h_k(\phi^{(k)}) \ , \qquad k = 1, \dots, K \tag{1}$$

Now, if multiple sources are present, any given location could be subject to disease exposure from as many as K wavefronts moving simultaneously. Yet, depending on conditions, the movement patterns of the wavefronts, and relative distances to each epicenter, an infection event at any particular time and location is attributable to the different sources with varying probability. In other words, disease at particular locations is more likely due to certain sources rather than others. To accommodate this intuition, a latent process Z is introduced that indicates the relative probabilities of disease associated with each of the K sources, and the collection of models given in Equation (1) describe (r, ϕ, t) conditional on the possible values of Z.

$$\mathbf{Z}_{K\times 1} \sim \text{Multinomial}\left(p^{(1)}, \dots, p^{(K)}\right) \tag{2}$$

$$\log\left(1 + \frac{r^{(k)}}{g_k\left(\phi^{(k)}\right)}\right) = -M_k t + h_k\left(\phi^{(k)}\right) \ , \qquad \text{if } Z_k = 1 \tag{3}$$

We propose an estimation procedure wherein velocity models are fit using regression methods conditional on known g_k. The functions g_k introduce anisotropy in the model. In many applications, known variables drive anistropy, so it is plausible to estimate g_k from covariate information or secondary data sources.

The velocity models (Equation (1)) are fitted conditional on g_k to disease occurrence data (presence or absence) of the form $\mathcal{Y} = \left\{(r_i^{(1)}, \phi_i^{(1)}), \dots, (r_i^{(K)}, \phi_i^{(K)}), t_i\right\}_{i=1}^{n}$ indicating the locations and times of the first observed disease case. For the purpose of exposition, suppose one is fitting only the kth model: consider just the data $(r_i^{(k)}, \phi_i^{(k)}, t_i)$ and assume $P(Z_i = k) = 1$. Now, adding an offset c_k and Gaussian error term $\epsilon_i^{(k)}$ to Equation (1) yields the statistical model

$$\log\left(1 + \frac{r_i^{(k)}}{g_k\left(\phi_i^{(k)}\right)}\right) = c_k - M_k t_i + h_k\left(\phi_i^{(k)}\right) + \epsilon_i^{(k)} \qquad \begin{cases} \epsilon_i^{(k)} \overset{iid}{\sim} N\left(0, \sigma_k^2\right) \\ i = 1, \dots, n \end{cases} \tag{4}$$

Under this multiple source situation, the *complete data* can be given as $\mathcal{X} = \left\{\mathbf{Z}_i, (r_i^{(1)}, \phi_i^{(1)}), \dots, (r_i^{(K)}, \phi_i^{(K)}), t_i\right\}_{i=1}^{n}$, where $\mathbf{Z}_i \in \{0,1\}^K$ for all $i \in \{1, \dots, n\}$ with only one 1 rest all 0's in each $\mathbf{Z}_i$. $\{\mathbf{Z}_i\}_{i=1}^{n}$ are considered as the *unobserved data*. The source with which the i-th location is principally associated with is given by $\mathbf{Z}_i$ by taking the value 1. The unknown parameter set is given by $\Theta = \left\{c_k, M_k, h_k, \sigma_k^2, p^{(k)}\right\}_{k=1}^{K}$.

The likelihood of the parameters Θ given the complete data $\mathcal{X}$ is given by

$$(5)\qquad \mathcal{L}\left(\Theta|\mathcal{X}\right) = \sum_{k=1}^{K}\prod_{i=1}^{n}\left[\varphi\left(\log\left(1+\frac{r_i^{(k)}}{g_k\left(\phi_i^{(k)}\right)}\right) - c_k + M_k t_i - h_k\left(\phi_i^{(k)}\right), \sigma_k^2\right)\right]^{\mathbf{1}(Z_{ik}=1)}$$

where $\varphi(x,\sigma^2)$ is the probability density function of a Gaussian random variable with mean zero and variance σ^2 evaluated at value x. The log-likelihood of the parameters Θ given the complete data $\mathcal{X}$ is given by

$$(6)\qquad \ell\left(\Theta|\mathcal{X}\right) = \sum_{k=1}^{K}\mathbf{1}(Z_{ik}=1)\sum_{i=1}^{n}\left[-\frac{\left(\log\left(1+\frac{r_i^{(k)}}{g_k\left(\phi_i^{(k)}\right)}\right) - c_k + M_k t_i - h_k\left(\phi_i^{(k)}\right)\right)^2}{2\sigma_k^2}\right]$$

The expected log-likelihood of the parameters Θ given the complete data $\mathcal{X}$ is given by

$$(7)\qquad \mathbb{E}\left[\ell\left(\Theta|\mathcal{X}\right)\right] = -\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{p_i^{(k)}}{\sigma_k^2}\left(\log\left(1+\frac{r_i^{(k)}}{g_k\left(\phi_i^{(k)}\right)}\right) - c_k + M_k t_i - h_k\left(\phi_i^{(k)}\right)\right)^2$$

where, $p_i^{(k)} = \mathbb{E}(Z_{ik}=1|\mathcal{Y})$, the probability of i-th location being principally associated with the k-th source ($i \in \{1,\ldots,n\}$, $k \in \{1,\ldots,K\}$). The Expectation-Maximization (EM) algorithm is an iterative algorithm which iterates between the expected log-likelihood and maximizing the expected log-likelihood.
The maximization of the complete log-likelihood in Equation (7) can be broken down into K minimization problems involving weighted least squares problems -

$$(8)\qquad \ell_k(c_k, M_k, h_k, \sigma_k^2) = -\sum_{i=1}^{n}\frac{p_i^{(k)}}{\sigma_k^2}\left(\log\left(1+\frac{r_i^{(k)}}{g_k\left(\phi_i^{(k)}\right)}\right) - c_k + M_k t_i - h_k\left(\phi_i^{(k)}\right)\right)^2$$

The minimization of weighted least squares loss function in Equation (8) leads to estimates of c_k, M_k and h_k given prior estimates of σ_k^2 and $\left\{p_i^{(k)}\right\}_{i=1}^{n}$. Estimates of c_k, M_k and h_k are easily computed using semiparametric regression. Let $s_1(\cdot),\ldots,s_B(\cdot)$ denote a set of B basis functions. Now, rewriting Equation (4) we obtain

$$(9)\qquad \log\left(1+\frac{r_i^{(k)}}{g_k\left(\phi_i^{(k)}\right)}\right) = c_k + (-M_k)t_i + \beta_1^{(k)}s_1\left(\phi_i^{(k)}\right) + \cdots + \beta_B^{(k)}s_B\left(\phi_i^{(k)}\right) + \epsilon_i^{(k)}$$

The weighted least squares (WLS) solution to Equation (9) with weights given by $\left\{p_i^{(k)}/\sigma_k^2\right\}_{i=1}^{n}$, subsequently yields estimates of $\hat{c}_k$, $\hat{M}_k$ and $\hat{h}_k = \sum_b \hat{\beta}_b^{(k)} s_b$ for each $k = 1,\ldots,K$. The maximum likelihood estimate of σ_k^2 becomes

$$(10)\qquad \hat{\sigma}_k^2 = \sum_{i=1}^{n}\hat{p}_i^{(k)}\left(\log\left(1+\frac{r_i^{(k)}}{g_k\left(\phi_i^{(k)}\right)}\right) - \hat{c}_k + \hat{M}_k t_i - \hat{h}_k\left(\phi_i^{(k)}\right)\right)^2.$$

The estimate of $\left\{p_i^{(k)}\right\}_{i=1}^n$ given the estimates $\hat{c}_k, \hat{M}_k, \hat{h}_k, \hat{\sigma}_k^2$, and $\left\{p_i^{(k)}\right\}_{i=1}^n$, the maximum likelihood estimate of $\left\{p_i^{(k)}\right\}_{i=1}^n$ becomes

$$\hat{p}_i^{(k)} = \frac{\varphi\left(\log\left(1 + \frac{r_i^{(k)}}{g_k\left(\phi_i^{(k)}\right)}\right) - \hat{c}_k + \hat{M}_k t_i - \hat{h}_k\left(\phi_i^{(k)}\right), \hat{\sigma}_k^2\right) p_i^{(k)}}{\sum_{k=1}^K \varphi\left(\log\left(1 + \frac{r_i^{(k)}}{g_k\left(\phi_i^{(k)}\right)}\right) - \hat{c}_k + \hat{M}_k t_i - \hat{h}_k\left(\phi_i^{(k)}\right), \hat{\sigma}_k^2\right) p_i^{(k)}}. \tag{11}$$

Finally, this estimation strategy is extended to the full collection of K models by accounting for the latent variables Z_i that attribute each of the i-th data points to one of the K sources. Formally, the joint likelihood of the data arising from Equations (2) and (3) is maximized with respect to the parameters $p^{(k)} \in \mathbb{R}^N$, $\beta_k \in \mathbb{R}^{B+2}$, and σ_k^2 for $k = 1, \ldots, K$. The EM algorithm is used to iteratively update estimated multinomial probabilities $\hat{p}_i^{(1)}, \ldots, \hat{p}_i^{(K)}$ for each data point in alternation with fitting the regression models in Equation (9) using the estimate $\hat{p}_i^{(k)}$ as a regression weight for the i-th data point in fitting the k-th model. In detail, the iterations are given by:

1. Initiate $\hat{p}_i^{(k)}$ as the weight of ith data-point to be associated with kth source, where $\sum_{k=1}^K \hat{p}_i^{(k)} = 1$.
2. Compute/update the estimates $\left(\hat{c}_k, \hat{M}_k, \hat{h}_k, \hat{\sigma}_k^2\right)_{k=1}^K$ by fitting each of the models in Equation (9) with weights $\hat{p}_i^{(k)}$ for the ith data point and the kth model.
3. Update $\hat{p}_i^{(k)}$ by

$$\hat{p}_i^{(k)} = \frac{\varphi\left(\log\left(1 + \frac{r_i^{(k)}}{g_k\left(\phi_i^{(k)}\right)}\right) - \hat{c}_k + \hat{M}_k t_i - \hat{h}_k\left(\phi_i^{(k)}\right), \hat{\sigma}_k^2\right) \hat{p}_i^{(k)}}{\sum_{k=1}^K \varphi\left(\log\left(1 + \frac{r_i^{(k)}}{g_k\left(\phi_i^{(k)}\right)}\right) - \hat{c}_k + \hat{M}_k t_i - \hat{h}_k\left(\phi_i^{(k)}\right), \hat{\sigma}_k^2\right) \hat{p}_i^{(k)}} \tag{12}$$

   where $\varphi(x, \sigma^2)$ is the probability density function of a Gaussian random variable with mean zero and variance σ^2 evaluated at value x.
4. Repeat steps 2-3 until convergence.

A simple heuristic for the initialization step is to use as $\hat{p}_i^{(k)}$ the estimated probabilities obtained by logistic regression of an indicator of whether the kth source is closest on the variables $r^{(1)}/\hat{g}_1(\phi^{(1)}), \ldots, r^{(K)}/g_k(\phi^{(K)})$. We note that an isotropic model with one or many sources can be recovered within this framework as a special case by fixing $g_k(x) = 1/2\pi$ for $x \in [0, 2\pi]$, with the consequence that $h_k \equiv 0$.

S2 Supplementary Table 1. The mean parameter estimates and standard deviation for two-source models fit to simulated data. Two sets of estimates are reported corresponding to a (0,0) placement for a first source and a (2000,2000) placement for a second source.

			Source 1				Source 2			
Start time	n	σ^2	Intercept	Time	Basis 1	Basis 2	Intercept	Time	Basis 1	Basis 2
True values			4.850	0.030	0.500	0.000	3.750	0.020	0.000	-1.000
Synchronous	100	0.5	4.862	0.030	0.507	-0.007	3.797	0.019	0.011	-0.985
			(0.216)	(0.002)	(0.130)	(0.124)	(0.348)	(0.003)	(0.200)	(0.202)
		1	5.291	0.026	0.493	-0.119	4.009	0.018	0.026	-0.596
			(0.585)	(0.005)	(0.252)	(0.253)	(1.193)	(0.011)	(0.626)	(0.914)
	500	0.5	4.869	0.030	0.504	-0.006	3.787	0.019	0.001	-0.986
			(0.103)	(0.001)	(0.054)	(0.055)	(0.155)	(0.002)	(0.096)	(0.096)
		1	5.691	0.023	0.499	-0.219	3.985	0.017	0.011	-0.699
			(0.204)	(0.002)	(0.118)	(0.104)	(0.523)	(0.005)	(0.277)	(0.330)
Asynchronous	100	0.5	4.889	0.030	0.498	0.003	3.943	0.018	0.002	-0.949
			(0.221)	(0.002)	(0.123)	(0.121)	(1.052)	(0.008)	(0.221)	(0.245)
		1	5.227	0.027	0.498	0.001	6.771	-0.005	-0.006	0.167
			(0.516)	(0.004)	(0.263)	(0.233)	(13.537)	(0.021)	(0.934)	(13.342)
	500	0.5	4.987	0.029	0.501	0.004	4.027	0.017	-0.002	-0.931
			(0.127)	(0.001)	(0.054)	(0.050)	(0.550)	(0.004)	(0.126)	(0.126)
		1	5.201	0.027	0.496	0.001	9.534	-0.018	-0.010	-0.888
			(0.207)	(0.002)	(0.109)	(0.094)	(12.000)	(0.014)	(0.485)	(11.952)