A Poisson cokriging method for bivariate count data

David Payares-Garcia [a],*, Frank Osei [a], Jorge Mateu [b], Alfred Stein [a]

[a] *ITC Faculty Geo-Information Science and Earth Observation, University of Twente, Enschede, The Netherlands*
[b] *Department of Mathematics, University Jaume I, Castellón, Spain*

ARTICLE INFO

ABSTRACT

Bivariate spatially correlated count data appear naturally in several domains such as ecology, economy and epidemiology. Current methods for analysing such data lack simplicity, interpretability and computational awareness. This paper introduces Poisson cokriging, a bivariate geostatistical technique to model and predict spatially correlated count variables. Our method exploits classical geostatistical theory and the bivariate Poisson distribution to propose an adaptation of cokriging when the underlying process follows a bivariate Poisson structure. A simulation study and a real data application using counts from two mosquito-borne diseases in Colombia showed that our model successfully performs spatial predictions at unobserved locations under different settings. We demonstrate the competitive convenience of Poisson cokriging in filtering rates and modelling highly variant population sizes against traditional geostatistical methods. We conclude that Poisson cokriging improves prediction accuracy and reduces variance prediction errors in comparison with ordinary cokriging.

1. Introduction

Bivariate spatially correlated count data occur in several disciplines, such as ecology, demography, economy and epidemiology. For example, public health institutions routinely collect

---

* Corresponding author.
E-mail address: d.e.payaresgarcia@utwente.nl (D. Payares-Garcia).

incidence counts of multiple diseases to identify risk factors, estimate health outcomes and implement integrated disease surveillance systems. In particular, epidemiology often exploits bivariate spatially-indexed count data to determine spatially varying disease-related factors and predict disease risk at unobserved locations (Holford, 2002). The joint modelling of spatial count data is particularly relevant in regression and prediction applications where a count variable is estimated using another spatially correlated count variable known throughout the study region.

Models for spatially dependent count data assume that the observed magnitude is the realisation of a random variable with a specific probability distribution, the Poisson distribution being the most frequent choice. There are three widely used frameworks for modelling counts: generalised linear mixed models (GLMM), model-based geostatistics and (classical, regionalised) geostatistics (named here as simply geostatistics). GLMMs are flexible in incorporating auxiliary information and spatial mixed effects and in explicitly modelling most sources of uncertainty. These models, however, are highly parametrised and imply complex Bayesian estimation of parameters (Goovaerts, 2017; Monestiez et al., 2006; Smith et al., 2020; Cressie, 1993). These characteristics make GLMM computationally demanding. Model-based geostatistics, as a relative of GLMM, exploits the convenience of geostatistical properties such as continuous modelling of spatial dependence, change of support and uncertainty quantification, but, unfortunately, it also inherits the Bayesian modelling limitations. Geostatistical applications to analyse count data are somewhat questionable as they assume that the underlying data distribution is typically second-order stationary and based on a Gaussian process. Methods to manage these limitations have been developed in traditional geostatistics (Berke, 2004; Goovaerts and Jacquez, 2004), but they directly induce undesirable characteristics in the semivariogram estimator and kriging predictions (Goovaerts, 2005). Nonetheless, the geostatistical framework is a simple, robust, scalable alternative to predict spatially correlated variables. Alternative methods for modelling spatially dependent count data have been proposed, including the Poisson random field (Morales-Navarrete et al., 2022) and Gaussian copula spatial regression (Masarotto and Varin, 2012). The Poisson random field offers interpretability advantages over traditional approaches such as the Poisson Log-Gaussian random field and Poisson Gaussian copula, but full likelihood inference is not feasible and an optimal predictor closed form is unavailable. Gaussian copula spatial regression also has strong interpretability, but simpler correlation models and estimation procedures are required to compute the likelihood.

A geostatistical model for count data, Poisson kriging, was introduced by Monestiez et al. (2006) to map the relative abundance of whales in the Mediterranean Sea. The model accommodates classical geostatistical theory for semivariogram estimation and optimal linear unbiased prediction (i.e. kriging) to an unobserved and non-stationary latent process that potentially drives the observed count data. Since Poisson kriging specifically deals with counts, its methodology and application have been mainly adopted to analyse rates that result from diving the observed counts by a sampling effort — for instance, dividing the number of deaths in a county by its total population. Poisson kriging accommodates the heteroscedasticity of rates derived from spatial counts, namely, the effect of the population size spatial variations on the process variance. Over the years, methodological extensions have been conducted to increase its adaptability to sampling settings, the change of support paradigm, and non-constant mean definitions (Goovaerts, 2005, 2006a, 2017; Bellier et al., 2013, 2010; Krivoruchko et al., 2011; Oliveira, 2014). Notably, Poisson kriging has become pivotal in disease risk applications as it adjusts for variance instability and spurious spatial predictions induced by population size variability (Osei and Stein, 2017).

While Poisson kriging is an ideal alternative to classical statistical methods for prediction based on count data (Goovaerts, 2017; Oliveira, 2014), it is restricted to univariate settings. Attempts to extend Poisson kriging into a multivariate scenario include binomial cokriging, an adaptation of kriging to a binomially distributed response variable (Oliver et al., 1998) and Poisson cokriging as a GLMM (Smith et al., 2020). Oliver et al. (1998) found the method better suited to model spatial disease proportions as it takes full advantage of integrating ancillary variables in the geostatistical modelling. Strictly speaking, however, the model does not directly incorporate auxiliary information in the prediction process. Similarly, Smith et al. (2020) proved that their model is an effective prediction method for spatially correlated count data with an auxiliary variable. Nonetheless, the method is purely spatial and is restricted to small sample sizes to overcome the computational burden of a Bayesian GMML.

This study aims to expand Poisson kriging into Poisson cokriging to predict spatially correlated counts in the presence of an ancillary spatially correlated auxiliary variable sharing the former's probability distribution. We build upon multivariate geostatistics theory for bivariate data (Wackernagel, 2003a; Cressie, 1993) and the bivariate Poisson distribution proposed by Kawamura (1973) to accommodate cokriging to spatially correlated count variables. Our ambition comprises assessing and adjusting for the effect population sizes have on the raw rates. First, a simulation study is conducted to examine the sampling properties of our proposed method, Poisson cokriging. Finally, a real data application is illustrated on Zika and Chikungunya incidence counts for two regions of Colombia in 2016.

2. Materials and methods

The definition of Poisson cokriging is based on traditional geostatistics. We followed the work by Monestiez et al. (2006), later extended by Bellier et al. (2010) and adapted into disease rates analysis by Goovaerts (2005). We also adopted the multivariate geostatistical definitions and procedures focusing on bivariate interactions from Stein and Corsten (1991), Wackernagel (2003a) and Cressie (1993).

2.1. Model

Let $\{r_\alpha(\mathbf{s}), r_\beta(\mathbf{s}), r_{\alpha\beta}(\mathbf{s}) : \mathbf{s} \in \mathcal{D} \subset \mathbb{R}^2\}$ be three dependent positive random fields describing three distinct spatially varying disease risks at location $\mathbf{s}$. As the values of the latent random fields are not observable, inference occurs through the mean of three random Poisson variables $X_{\alpha i}$, $X_{\beta i}$, and $X_{\alpha\beta i}$, representing the number of disease cases and $n_{\alpha i}$, $n_{\beta i}$ and $n_{\alpha\beta i}$ referring to the population sizes at sampling locations $\mathbf{s}_1, \ldots, s_k$.

The random variables $X_{\alpha i}$, $X_{\beta i}$, and $X_{\alpha\beta i}$ can be combined to generate two dependent Poisson random variables (Kawamura, 1973): $Y_{\alpha i} = X_{\alpha i} + X_{\alpha\beta i}$ and $Y_{\beta i} = X_{\beta i} + X_{\alpha\beta i}$. It follows that:

$$
\begin{aligned}
Y_{\alpha i} \mid r_\alpha(\mathbf{s}_i), r_{\alpha\beta}(\mathbf{s}_i) &\sim \text{Poisson}(n_{\alpha i} \cdot r_\alpha(\mathbf{s}_i) + n_{\alpha\beta i} \cdot r_{\alpha\beta}(\mathbf{s}_i)) \\
Y_{\beta i} \mid r_\beta(\mathbf{s}_i), r_{\alpha\beta}(\mathbf{s}_i) &\sim \text{Poisson}(n_{\beta i} \cdot r_\beta(\mathbf{s}_i) + n_{\alpha\beta i} \cdot r_{\alpha\beta}(\mathbf{s}_i))
\end{aligned}
\tag{1}
$$

where $r_\alpha(\mathbf{s}_i)$ and $r_\beta(\mathbf{s}_i)$ denote the marginal disease risks, and $r_{\alpha\beta}(\mathbf{s}_i)$ stands for the joint risk between the observed counts $Y_{\alpha i}$ and $Y_{\beta i}$ for any set of sampling locations $\mathbf{s}_1, \ldots, s_k \in \mathcal{D}$. The counts $Y_{\alpha i}$ and $Y_{\beta i}$ are conditionally independent given the risks $r_\alpha(\mathbf{s}_i)$ and $r_\beta(\mathbf{s}_i)$ and $r_{\alpha\beta}(\mathbf{s}_i)$. Eq. (1) can be further simplified by assuming that population sizes $n_{\alpha i} = n_{\alpha\beta i}$ and $n_{\beta i} = n_{\alpha\beta i}$. Reducing the independent marginal and joint risks can be reduced into a single term $\mathcal{R}_\alpha(\mathbf{s}_i) = r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)$ and $\mathcal{R}_\beta(\mathbf{s}_i) = r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)$, we obtain,

$$
\begin{aligned}
Y_{\alpha i} \mid \mathcal{R}_\alpha(\mathbf{s}_i) &\sim \text{Poisson}(n_{\alpha i} \cdot \mathcal{R}_\alpha(\mathbf{s}_i)), \qquad \mathcal{R}_\alpha(\mathbf{s}_i) = r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i) \\
Y_{\beta i} \mid \mathcal{R}_\beta(\mathbf{s}_i) &\sim \text{Poisson}(n_{\beta i} \cdot \mathcal{R}_\beta(\mathbf{s}_i)), \qquad \mathcal{R}_\beta(\mathbf{s}_i) = r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)
\end{aligned}
\tag{2}
$$

Here, the counts $Y_{\alpha i}$ and $Y_{\beta i}$ are Poisson distributed whose parameters are the product between the population sizes $n_{\alpha i}$ and $n_{\beta i}$ and the total risk $\mathcal{R}_\alpha(\mathbf{s}_i)$ and $\mathcal{R}_\beta(\mathbf{s}_i)$ at sampling sites $\mathbf{s}_i$, respectively. The total risks $\mathcal{R}_\alpha(\mathbf{s})$ and $\mathcal{R}_\beta(\mathbf{s})$ are positive random fields honouring order two stationarity, with means $m_\alpha + m_{\alpha\beta}$ and $m_\beta + m_{\alpha\beta}$, variances $\sigma_\alpha^2 + \sigma_{\alpha\beta}^2$ and $\sigma_\beta^2 + \sigma_{\alpha\beta}^2$ and covariance functions $C_\alpha^\mathcal{R}(|\mathbf{s}_i - \mathbf{s}_j|)$ and $C_\beta^\mathcal{R}(|\mathbf{s}_i - \mathbf{s}_j|)$, correspondingly. Note that the covariance functions are isotropic, that is, they depend only upon the distances between sites $\mathbf{s}_i$ and $\mathbf{s}_j$. Note also that the population sizes $n_{\alpha i}$, $n_{\beta i}$ and $n_{\alpha\beta i}$ are identical. For clarity, we will use $n_{\alpha i}$ and $n_{\beta i}$ independently but interchangeably to develop the model's definitions.

As stated in Monestiez et al. (2006) and Bellier et al. (2010), no further distributional assumptions are imposed on $\mathcal{R}_\alpha(\mathbf{s}_i)$ and $\mathcal{R}_\beta(\mathbf{s}_i)$, except the constraints $\mathcal{R}_\alpha(\mathbf{s}_i) \geq 0$ and $\mathcal{R}_\beta(\mathbf{s}_i) \geq 0$.

2.2. Expectation and variances of Y_α and Y_β

Several properties can be derived directly from (2). It follows that:

$$
\begin{aligned}
\mathrm{E}\left[Y_{\alpha i} \mid \mathcal{R}_\alpha(\mathbf{s}_i)\right] = \mathrm{Var}\left[Y_{\alpha i} \mid \mathcal{R}_\alpha(\mathbf{s}_i)\right] &= n_{\alpha i} \cdot \mathcal{R}_\alpha(\mathbf{s}_i) \\
&= n_{\alpha i} \cdot (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) \\
\mathrm{E}\left[(Y_{\alpha i})^2 \mid \mathcal{R}_\alpha(\mathbf{s}_i)\right] &= n_{\alpha i}\mathcal{R}_\alpha(\mathbf{s}_i) + n_{\alpha i}^2 \mathcal{R}_\alpha(\mathbf{s}_i)^2 \\
&= n_{\alpha i} \cdot (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) + n_{\alpha i}^2 \cdot (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))^2
\end{aligned}
\tag{3}
$$

and when marginalising:

$$
\begin{aligned}
\mathrm{E}[Y_{\alpha i}] &= n_{\alpha i} \cdot (m_\alpha + m_{\alpha\beta}) \\
\mathrm{Var}[Y_{\alpha i}] &= n_{\alpha i} \cdot (m_\alpha + m_{\alpha\beta}) + n_{\alpha i}^2 \cdot (\sigma_\alpha + \sigma_{\alpha\beta})
\end{aligned}
\tag{4}
$$

It is tacit that (3) and (4) also hold for $Y_{\beta i}$. The conditional independence of observations at different sites introduces two covariance expressions: a covariance within and between the processes $Y_{\alpha i} \mid \mathcal{R}_\alpha(\mathbf{s}_i)$ and $Y_{\beta i} \mid \mathcal{R}_\beta(\mathbf{s}_i)$. Deriving the former is trivial, and we obtain,

$$
\begin{aligned}
\mathrm{E}\left[Y_{\alpha i} Y_{\alpha j} \mid \mathcal{R}_\alpha\right] &= \mathrm{Cov}[Y_{\alpha i}, Y_{\alpha j} \mid \mathcal{R}_\alpha] + \mathrm{E}\left[Y_{\alpha i} \mid \mathcal{R}_\alpha(\mathbf{s}_i)\right] \mathrm{E}\left[Y_{\alpha j} \mid \mathcal{R}_\alpha(\mathbf{s}_j)\right] \\
&= \delta_{ij}(n_{\alpha i} \cdot (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))) + n_{\alpha i} \cdot (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))\, n_{\alpha j} \cdot (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))
\end{aligned}
\tag{5}
$$

where δ_{ij} is the Kronecker delta $\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$ indicating the conditional independence between locations.

Eq. (5) displays the covariance expression within $Y_\alpha \mid \mathcal{R}_\alpha$, and naturally, it also holds for $Y_\beta \mid \mathcal{R}_\beta$. The covariance between $Y_{\alpha i} \mid \mathcal{R}_\alpha(\mathbf{s}_i)$ and $Y_{\beta i} \mid \mathcal{R}_\beta(\mathbf{s}_i)$ is the shared parameter $n_{\alpha i} \cdot r_{\alpha\beta}(\mathbf{s}_i)$ which controls for the dependence between the two random fields (Kawamura, 1973). This expression exists owing to the joint marginal risk $r_{\alpha\beta}(\mathbf{s})$. It results that:

$$
\begin{aligned}
\mathrm{E}\left[Y_{\alpha i} Y_{\beta j} \mid \mathcal{R}\right] &= \mathrm{Cov}[Y_{\alpha i}, Y_{\beta j} \mid \mathcal{R}] + \mathrm{E}\left[Y_{\alpha i} \mid \mathcal{R}_\alpha(\mathbf{s}_i)\right] \mathrm{E}\left[Y_{\beta j} \mid \mathcal{R}_\beta(\mathbf{s}_j)\right] \\
&= \delta_{ij}[n_{\alpha i} r_{\alpha\beta}(\mathbf{s}_i)] + n_{\alpha i} \cdot (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))\, n_{\beta j} \cdot (r_\beta(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))
\end{aligned}
\tag{6}
$$

Eq. (6) is critical to estimate the semivariograms and Poisson cokriging equations.

2.3. Estimation of the semivariograms

2.3.1. Direct semivariogram

For any pair of locations $\mathbf{s}_i, \mathbf{s}_j \in \mathcal{D}$ for the random variable $Y_\alpha(\mathbf{s}_i)$, the direct semivariogram $\gamma_\alpha^Y(\mathbf{h})$ is defined in the context of the intrinsic hypothesis (intrinsic stationary of order two). We have that

$$
\begin{aligned}
\mathrm{E}\left[Y_{\alpha i} - Y_{\alpha j}\right] &= 0 \\
\mathrm{Var}\left[Y_{\alpha i} - Y_{\alpha j}\right] &= 2\gamma_\alpha^Y(\mathbf{h})
\end{aligned}
\tag{7}
$$

Under these two properties, the direct theoretical semivariogram can be written as:

$$
\gamma_\alpha^Y(\mathbf{h}) = \frac{1}{2}\,\mathrm{E}\left[(Y_{\alpha i} - Y_{\alpha j})^2\right]
\tag{8}
$$

However, (7) cannot be directly adopted for the counts $Y_\alpha(\mathbf{s}_i)$ due to the heteroscedasticity and non-stationarity induced by the population sizes n_α and the counts themselves. Likewise, we are primarily concerned with estimating the total risk $\mathcal{R}_\alpha(\mathbf{s}_i)$ rather than the counts $Y_\alpha(\mathbf{s}_i)$. Consequently, we need to infer the relationship between the semivariograms of the counts $Y_\alpha(\mathbf{s}_i)$ and the total risk $\mathcal{R}_\alpha(\mathbf{s}_i)$. Using the sample rates $\frac{Y_{\alpha i}}{n_{\alpha i}}$ and $\frac{Y_{\alpha j}}{n_{\alpha j}}$ and (7) and (8) we see that:

$$
\mathrm{E}\left[\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right] = 0
\tag{9}
$$

and

$$\frac{1}{2}\,\mathrm{E}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right)^2\right] = \frac{1}{2}(m_\alpha + m_{\alpha\beta})\left(\frac{1}{n_{\alpha i}} + \frac{1}{n_{\alpha j}}\right) - \delta_{ij}\frac{m_\alpha + m_{\alpha\beta}}{n_{\alpha i}}$$
$$+ \frac{1}{2}\,\mathrm{E}\left[(\mathcal{R}_\alpha(\mathbf{s}_i) - \mathcal{R}_\alpha(\mathbf{s}_j))^2\right]. \tag{10}$$

Let $\gamma_\alpha^Y(\mathbf{h})$ be the non-stationary theoretical direct semivariogram of the random field $\frac{Y_{\alpha i}}{n_{\alpha i}}$ as defined in (8) and $\mathbf{h} = \|\mathbf{s}_i - \mathbf{s}_j\|$. When $i \neq j$ we obtain:

$$\gamma_\alpha^{\mathcal{R}}(\mathbf{h}) = \gamma_\alpha^Y(\mathbf{h}) - \frac{1}{2}(m_\alpha + m_{\alpha\beta})\left(\frac{n_{\alpha i} + n_{\alpha j}}{n_{\alpha i}n_{\alpha j}}\right) \tag{11}$$

When $i = j$ (10), reduces to:

$$\gamma_\alpha^{\mathcal{R}}(0) = \gamma_\alpha^Y(0) - \frac{1}{2}(m_\alpha + m_{\alpha\beta})\left(\frac{1}{n_{\alpha i}} + \frac{1}{n_{\alpha i}}\right) + \delta_{ii}\frac{m_\alpha + m_{\alpha\beta}}{n_{\alpha i}}$$
$$\gamma_\alpha^{\mathcal{R}}(0) = \gamma_\alpha^Y(0) - (m_\alpha + m_{\alpha\beta})\left(\frac{1}{n_{\alpha i}}\right) + \frac{m_\alpha + m_{\alpha\beta}}{n_{\alpha i}} \tag{12}$$
$$\gamma_\alpha^{\mathcal{R}}(0) = \gamma_\alpha^Y(0) = 0$$

For $i \neq j$ we also have:

$$\mathrm{Var}\left[\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}} \mid \mathcal{R}_\alpha\right] = \frac{(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))}{n_{\alpha i}} + \frac{(r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))}{n_{\alpha j}}$$
$$\Rightarrow \mathrm{E}\left[\mathrm{Var}\left[\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}} \mid \mathcal{R}_\alpha\right]\right] = (m_\alpha + m_{\alpha\beta})\left(\frac{n_{\alpha i} + n_{\alpha j}}{n_{\alpha i}n_{\alpha j}}\right) \tag{13}$$

Details on how to obtain Eqs. (9), (10) and (13) can be found in Appendix A.

The direct semivariogram estimator of $\gamma_\alpha^{\mathcal{R}}(\mathbf{h})$ can be inferred from (11) and (13). Let $Y_{\alpha i}$ be the sample counts and $n_{\alpha i}$ the population sizes for the set of spatial locations $\mathbf{s}_i = 1, \ldots, k$. The expression of the direct semivariogram is defined as:

$$\gamma_\alpha^{\mathcal{R}}(\mathbf{h}) = \frac{1}{2\sum_{(i,j)\in N(\mathbf{h})} w_{ij}} \sum_{i,j}\left(w_{ij}\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right)^2 - (m_\alpha^* + m_{\alpha\beta}^*)\right) \tag{14}$$

where $N(\mathbf{h})$ is the number of observation pairs separated by a distance $\mathbf{h}$ between i and j, $w_{ij} = \frac{n_{\alpha i}n_{\alpha j}}{n_{\alpha i}+n_{\alpha j}}$, and $m_\alpha^* = \frac{\sum n_{\alpha i}r_\alpha(\mathbf{s}_i)}{\sum n_{\alpha i}}$ and $m_{\alpha\beta}^* = \frac{\sum n_{\alpha\beta i}r_{\alpha\beta}(\mathbf{s}_i)}{\sum n_{\alpha\beta i}}$ are the mean estimates of the marginal risks $r_\alpha(\mathbf{s}_i)$ and $r_{\alpha\beta}(\mathbf{s}_i)$, respectively. If the joint risk $r_{\alpha\beta}(\mathbf{s}_i) = 0$, the direct semivariogram estimator in Eq. (14) is identical to the semivariogram expression defined by Monestiez et al. (2006).

The definitions of the direct modified semivariogram $\gamma_\alpha^{\mathcal{R}}(\mathbf{h})$ unmistakably apply to the random variable $Y_\beta(\mathbf{s}_i)$ as well.

2.3.2. Cross-semivariogram

The covariance-based cross-semivariogram $\gamma_{\alpha\beta}(\mathbf{h})$ is defined in the context of a joint intrinsic hypothesis for two random variables $Y_\alpha(\mathbf{s}_i)$ and $Y_\beta(\mathbf{s}_i)$, when evaluated for any site $\mathbf{s}_i, \mathbf{s}_j \in \mathcal{D}$. It follows that:

$$\mathrm{E}\left[Y_{\alpha i} - Y_{\alpha j}\right] = 0$$
$$\mathrm{E}\left[Y_{\beta i} - Y_{\beta j}\right] = 0$$
$$\mathrm{Cov}\left[(Y_{\alpha i} - Y_{\alpha j})(Y_{\beta i} - Y_{\beta j})\right] = 2\gamma_{\alpha\beta}^Y(\mathbf{h}) \tag{15}$$

The cross-semivariogram is thus defined as half the expectation of the product of the increments of the two variables (Matheron, 1965; Wackernagel, 2003b):

$$\gamma_{\alpha\beta}^Y(\mathbf{h}) = \frac{1}{2}\mathrm{E}\left[\left(Y_{\alpha i} - Y_{\alpha j}\right) \cdot \left(Y_{\beta i} - Y_{\beta j}\right)\right] \tag{16}$$

Since we are interested in the sample rates $\frac{Y_{\alpha i}}{n_{\alpha i}}$ and $\frac{Y_{\beta j}}{n_{\beta j}}$, (15) solves exactly as (7). Then,

$$\mathrm{E}\left[\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right] = 0, \quad \mathrm{E}\left[\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}}\right] = 0 \tag{17}$$

It also results that:

$$\frac{1}{2}\mathrm{E}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right)\left(\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}}\right)\right] = \frac{1}{2}\left(\frac{m_{\alpha\beta}}{n_{\alpha i}} + \frac{m_{\alpha\beta}}{n_{\beta j}}\right) - \frac{1}{2}\left(\frac{m_{\alpha\beta}}{n_{\alpha i}}\delta_{\alpha\beta ij} + \frac{m_{\alpha\beta}}{n_{\beta j}}\delta_{\alpha\beta ij}\right)$$
$$+ \frac{1}{2}\mathrm{E}\left[\left(\mathcal{R}_\alpha(\mathbf{s}_i) - \mathcal{R}_\alpha(\mathbf{s}_j)\right)\left(\mathcal{R}_\beta(\mathbf{s}_i) - \mathcal{R}_\beta(\mathbf{s}_i)\right)\right]. \tag{18}$$

Let $\gamma_{\alpha\beta}^Y(\mathbf{h})$ be the non-stationary theoretical cross-semivariogram of the random fields $\frac{Y_{\alpha i}}{n_{\alpha i}}$ and $\frac{Y_{\beta j}}{n_{\beta j}}$ and $\mathbf{h} = \|\mathbf{s}_i - \mathbf{s}_j\|$. When $i \neq j$ we obtain:

$$\gamma_{\alpha\beta}^\mathcal{R}(\mathbf{h}) = \gamma_{\alpha\beta}^Y(\mathbf{h}) - \frac{1}{2}m_{\alpha\beta}\left(\frac{n_{\alpha i} + n_{\beta j}}{n_{\alpha i}n_{\beta j}}\right) \tag{19}$$

For $i = j$, (19) reduces to $\gamma_{\alpha\beta}^Y(0) = \gamma_{\alpha\beta}^\mathcal{R}(0) = 0$.

The expectation of the conditional covariance of the difference in the sample rates for $i \neq j$ equals:

$$\mathrm{Cov}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right), \left(\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}}\right) \mid \mathcal{R}\right] = \frac{r_{\alpha\beta}(\mathbf{s}_i)}{n_{\alpha i}} + \frac{r_{\alpha\beta}(\mathbf{s}_j)}{n_{\beta j}}$$
$$\mathrm{E}\left[\mathrm{Cov}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right), \left(\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}}\right) \mid \mathcal{R}\right]\right] = m_{\alpha\beta}\left(\frac{n_{\alpha i} + n_{\beta j}}{n_{\alpha i}n_{\beta j}}\right) \tag{20}$$

Appendix A contains details on how to solve Eqs. (17), (18) and (20).

Finally, the cross-semivariogram estimator of $\gamma_{\alpha\beta}^\mathcal{R}(\mathbf{h})$ can be derived using Eqs. (18) and (20). Let $Y_{\alpha i}$ be the sample counts and $n_{\alpha i}$ the observed population sizes for the set of spatial locations $\mathbf{s}_i = 1, \ldots, k$ and $Y_{\beta j}$ be the sample counts and $n_{\beta j}$ the observed population sizes for the set of spatial locations $\mathbf{s}_j = 1, \ldots, k$. The estimation of the cross-semivariogram for the sample rates can be written as:

$$\gamma_{\alpha\beta}^\mathcal{R}(\mathbf{h}) = \frac{1}{2\sum_{(i,j)\in N(\mathbf{h})} w_{\alpha\beta}} \sum_{i,j}\left(w_{\alpha\beta}\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right)\left(\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}}\right) - m_{\alpha\beta}^*\right) \tag{21}$$

where $w_{\alpha\beta} = \frac{n_{\alpha i}n_{\beta j}}{n_{\alpha i}+n_{\beta j}}$ is the number of observation pairs separated by a distance $\mathbf{h}$ between i and j and $m_{\alpha\beta}^* = \frac{\sum n_{\alpha\beta i}r_{\alpha\beta}(\mathbf{s}_i)}{\sum n_{\alpha\beta i}}$ is the mean estimate of the joint risk $r_{\alpha\beta}(\mathbf{s}_i)$.

Wackernagel (2003b) argued that cross-semivariogram, namely the covariance-based cross-variogram (Myers, 1982) is a better choice than the variance-based cross-variogram (pseudo-cross-variogram) in terms of interpretability, consistency and stationarity. Built on his premise, we chose the covariance-based cross-variogram to estimate the cross-variogram between the sample rates $\frac{Y_{\alpha i}}{n_{\alpha i}}$ and $\frac{Y_{\beta j}}{n_{\beta j}}$.

2.4. Prediction of the latent process

The spatial interpolation of the total risk $\mathcal{R}_\alpha(\mathbf{s}_0)$ at any site $s_0 \in \mathcal{D}$ is a linear predictor combining the observed data $Y_{\alpha i}$ weighted by population sizes $n_{\alpha i}$ and the observed data $Y_{\beta i}$ weighted by population sizes $n_{\beta i}$ located at observed points in the neighbourhood of site s_0. The sample rates $\frac{Y_{\alpha i}}{n_{\alpha i}}$ and $\frac{Y_{\beta j}}{n_{\beta j}}$ are defined on a set of locations k_α and k_β, respectively, in order to entail the possibility

of (partially) heterotopic data. The Poisson cokriging estimator equals:

$$\mathcal{R}_\alpha(\mathbf{s}_0)^* = \sum_{i=1}^{k_\alpha} \lambda_{\alpha i} \frac{Y_{\alpha i}}{n_{\alpha i}} + \sum_{i=1}^{k_\beta} \lambda_{\beta i} \frac{Y_{\beta i}}{n_{\beta i}} \tag{22}$$

The unbiasedness constraint leads to the standard weights as in classical (bivariate) ordinary cokriging:

$$\sum_{i=1}^{k_\alpha} \lambda_{\alpha i} = 1, \quad \sum_{i=1}^{k_\beta} \lambda_{\beta i} = 0 \tag{23}$$

Similarly to ordinary cokriging, we obtain the mean square error of prediction (MSEP) (see Appendix B). Its expression is:

$$
\begin{aligned}
\text{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)] = & \sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}^2}{n_{\alpha i}}(m_\alpha + m_{\alpha\beta}) + \sum_{i=1}^{k_\alpha} \frac{\lambda_{\beta i}^2}{n_{\beta i}}(m_\beta + m_{\alpha\beta}) + 2\sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \frac{\lambda_{\alpha i}\lambda_{\beta j}}{n_{\beta j}} m_{\alpha\beta} \\
& + \sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\alpha} \lambda_{\alpha i}\lambda_{\alpha j}C^{\mathcal{R}}_{\alpha\alpha ij} + \sum_{i=1}^{k_\beta}\sum_{j=1}^{k_\beta} \lambda_{\beta i}\lambda_{\beta j}C^{\mathcal{R}}_{\beta\beta ij} + 2\sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \lambda_{\alpha i}\lambda_{\beta j}C^{\mathcal{R}}_{\alpha\beta ij} \\
& - 2\sum_{i=1}^{k_\alpha} \lambda_{\alpha i}C^{\mathcal{R}}_{\alpha\alpha i0} - 2\sum_{i=1}^{k_\beta} \lambda_{\beta i}C^{\mathcal{R}}_{\beta\alpha i0} + (\sigma_\alpha^2 + \sigma_{\alpha\beta}^2)
\end{aligned}
\tag{24}
$$

The unbiasedness condition in (23) implies that $\text{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)] = \text{E}[(\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0))^2]$.

Minimising (24) on the weights λ_α and λ_β generates two parameters of Lagrange μ_α and μ_β and $(k_\alpha + k_\beta + 2)$ equations. We then obtain the Poisson cokriging system:

$$
\begin{cases}
\dfrac{\lambda_{\alpha i}}{n_{\alpha i}}(m_\alpha + m_{\alpha\beta}) + \displaystyle\sum_{i=1}^{k_\beta} \frac{\lambda_{\beta i}}{n_{\beta i}} m_{\alpha\beta} + \sum_{i=1}^{k_\alpha} \lambda_{\alpha i}C^{\mathcal{R}}_{\alpha\alpha ij} \\
\quad + \sum_{i=1}^{k_\beta} \lambda_{\beta i}C^{\mathcal{R}}_{\alpha\beta ij} + \mu_\alpha = C^{\mathcal{R}}_{\alpha\alpha i0} \quad \text{for } j = 1, 2, \ldots, k_\alpha \\[2mm]
\dfrac{\lambda_{\beta i}}{n_{\beta i}}(m_\beta + m_{\alpha\beta}) + \displaystyle\sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}}{n_{\alpha i}} m_{\alpha\beta} + \sum_{i=1}^{k_\beta} \lambda_{\alpha i}C^{\mathcal{R}}_{\beta\beta ij} \\
\quad + \sum_{i=1}^{k_\alpha} \lambda_{\alpha i}C^{\mathcal{R}}_{\alpha\beta ij} + \mu_\beta = C^{\mathcal{R}}_{\beta\alpha i0} \quad \text{for } j = 1, 2, \ldots, k_\beta \\[2mm]
\displaystyle\sum_{i=1}^{k_\alpha} \lambda_{\alpha i} = 1 \\[2mm]
\displaystyle\sum_{i=1}^{k_\alpha} \lambda_{\beta i} = 0
\end{cases}
\tag{25}
$$

Note that the system is expressed in terms of the covariance rather than the semivariogram. The system can be easily rewritten by using the relationship $C_{ij} = \sigma^2 - \gamma_{ij}$.

Eq. (24) can be further simplified by using the Poisson cokriging equation system (details in Appendix B). We obtain:

$$\text{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)] = (\sigma_\alpha^2 + \sigma_{\alpha\beta}^2) - \sum_{i=1}^{k_\alpha} \lambda_{\alpha i}C^{\mathcal{R}}_{\alpha\alpha i0} - \sum_{i=1}^{k_\beta} \lambda_{\beta i}C^{\mathcal{R}}_{\beta\alpha i0} - \mu_\alpha \tag{26}$$

Eq. (26) is the usual expression for the ordinary cokriging variance. Nonetheless, the solution of the weights $\lambda_{\alpha i}$ and $\lambda_{\beta i}$ and the variance value $\sigma^2_{\mathcal{R}_\alpha} = \sigma_\alpha^2 + \sigma_{\alpha\beta}^2$ lead to rather different kriging variances.

2.4.1. Matrix formulation

The Poisson cokriging equations can be as well written in matrix notation. This notation is computationally convenient to solve the equations system in (25). The Poisson cokriging equation system becomes:

$$\mathbf{C}\boldsymbol{\lambda} = \mathbf{c} \tag{27}$$

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{\alpha\alpha} & \mathbf{C}_{\alpha\beta} & \mathbf{1} & \mathbf{0} \\ \mathbf{C}_{\beta\alpha} & \mathbf{C}_{\beta\beta} & \mathbf{0} & \mathbf{1} \\ \mathbf{1}^T & \mathbf{0}^T & 0 & 0 \\ \mathbf{0}^T & \mathbf{1}^T & 0 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{C}^T & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0}^* \end{pmatrix}, \quad \boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{\lambda}_\alpha \\ \boldsymbol{\lambda}_\beta \\ \mu_\alpha \\ \mu_\beta \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} \mathbf{c}_{\alpha 0} \\ \mathbf{c}_{\beta 0} \\ 1 \\ 0 \end{pmatrix},$$

where

$$\begin{aligned}
\mathbf{c}_{\ell 0}^T &= (C_{\ell\alpha 10}, \ldots, C_{\ell\alpha n0}) & \text{for } \ell \in \alpha, \beta \\
\boldsymbol{\lambda}_\ell^T &= (\lambda_{\ell 1}, \ldots, \lambda_{\ell n}) & \text{for } \ell \in \alpha, \beta \\
\mathbf{C}_{\ell k} &= \begin{pmatrix} C_{\ell k 11} + \frac{W_{\ell k}}{n_{k1}} & \cdots & C_{\ell k 1n} \\ \vdots & \ddots & \vdots \\ C_{\ell k n1} & \cdots & C_{\ell k nn} + \frac{W_{\ell k}}{n_{kn}} \end{pmatrix} & \text{for } \ell, k \in \alpha, \beta
\end{aligned} \tag{28}$$

Here $W_{\ell k} = \begin{cases} m_\ell + m_{\ell k} & \text{if } \ell = k \\ m_{\ell k} & \text{if } \ell \neq k \end{cases}$ is a term added to the diagonal of the coefficient matrices.

Although (27) is analogous to the ordinary cokriging system, the coefficient matrix $\mathbf{C}_{\ell k}$ is more complex due to the added term $W_{\ell k}$.

Finally, the weights of the Poisson cokriging model are found by solving the matrix equation

$$\boldsymbol{\lambda} = \mathbf{C}^{-1}\mathbf{c} \tag{29}$$

The system of equations is to be solved for each prediction location individually.

3. Simulation study

A simulation study was conducted to assess and compare the performances of Poisson cokriging against ordinary cokriging (Wackernagel, 2003a). This simulation evaluates two main aspects in terms of semivariogram estimation and prediction and errors: (i) the effect of the variability of the population sizes and, (ii) the increasing correlation between the target risk variable and the auxiliary variable.

3.1. Simulation design

Multiple datasets following the specifications of the Poisson cokriging model were simulated at $n = 400$ sampling locations uniformly distributed over the regular grid $\mathcal{D} = [-10, 10] \times [-10, 10]$ following the linear model of coregionalisation (LMC) (Goulard and Voltz, 1992). We considered sampling locations distributed over a random spatial configuration. We assumed that the latent random fields $\mathcal{R}_\alpha(\mathbf{s})$ and $\mathcal{R}_\beta(\mathbf{s})$ are log-Gaussian and that, conditional on the latent process, the counts $Y_{\alpha i}$ and $Y_{\beta i}$ follow a Poisson distribution.

One thousand cross-correlated spatial random fields were generated using the LMC. Both total risk variables $\mathcal{R}_\alpha(\mathbf{s})$ and $\mathcal{R}_\beta(\mathbf{s})$ were simulated as bivariate log-Gaussian random fields with zero mean to ensure positive risk. The covariance matrix combines the isotropic exponential covariance function with range 2 and sill 1 and the spherical structure with a range 5 and sill 1. No nugget effect was added.

Based on (Kawamura, 1973) definition of the correlation between the two independent Poisson processes $\rho_Y = \frac{n_{\alpha\beta}r_{\alpha\beta}}{\sqrt{n_\alpha \mathcal{R}_\alpha \cdot n_\beta \mathcal{R}_\beta}}$, we produced six sets of the risks $\mathcal{R}_\alpha(\mathbf{s})$, $\mathcal{R}_\beta(\mathbf{s})$ and $r_{\alpha\beta}(\mathbf{s})$ by varying $\rho_Y = 0.001, 0.2, 0.4, 0.6, 0.8, 1.0$. Fig. 1 shows the simulated risks when $\rho_Y = 0.6$.

(a) Total risk $\log \mathcal{R}_\alpha(\mathbf{s})$

(b) Total risk $\log \mathcal{R}_\beta(\mathbf{s})$
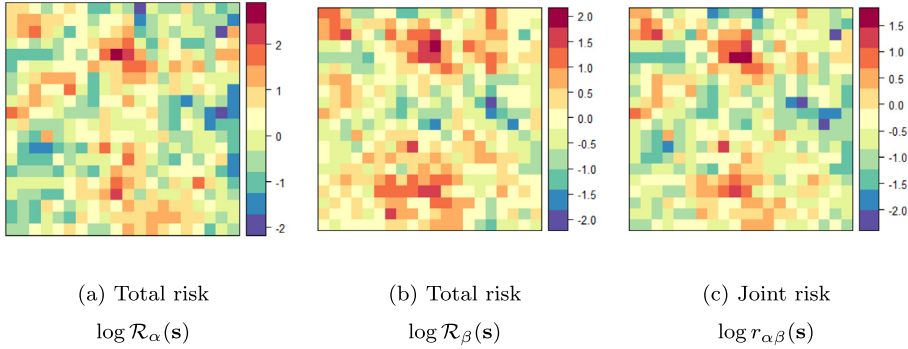
(c) Joint risk $\log r_{\alpha\beta}(\mathbf{s})$

Fig. 1. Simulation of the total risks (underlying log-Gaussian process) at each location; (a) $\log \mathcal{R}_\alpha(\mathbf{s})$ and (b) $\log \mathcal{R}_\beta(\mathbf{s})$ and, (c) joint marginal risk $\log r_{\alpha\beta}(\mathbf{s})$ when $\rho_Y = 0.6$.

(a) Uniform

(b) Poisson

(c) Beta-binomial
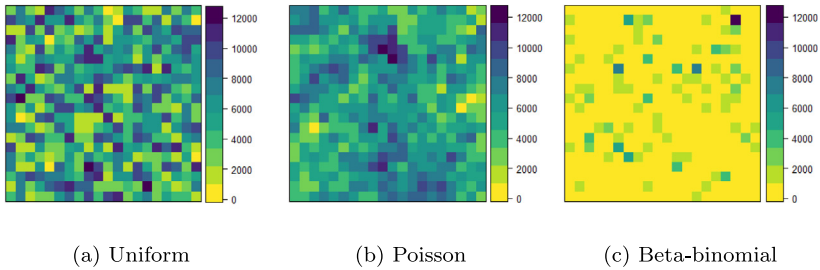
Fig. 2. Simulation of the counts Y_α derived from the definition of the risks $\mathcal{R}_\alpha$, $\mathcal{R}_\beta$ and $r_{\alpha\beta}$ using the different distributions of the population sizes n.
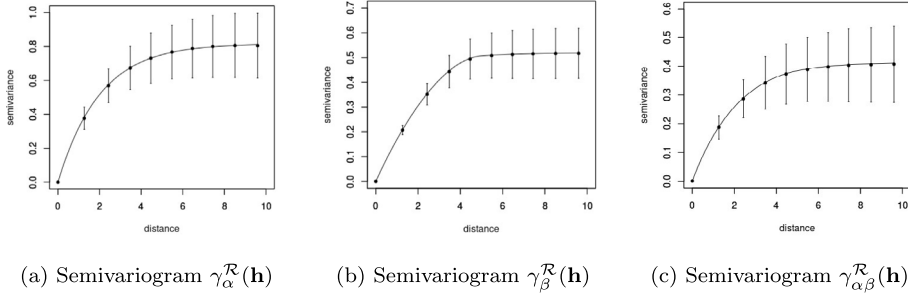
The two dependent Poisson variables $Y_{\alpha i}$ and $Y_{\beta i}$ were simulated (Fig. 2) through (2). We used the total risks $\mathcal{R}_\alpha(\mathbf{s})$ and $\mathcal{R}_\beta(\mathbf{s})$ displayed in Fig. 1, and population sizes $n_{\alpha i}$ and $n_{\beta i}$ drawn from the discreet uniform, Poisson and, beta-binomial distributions since we are interested in evaluating the effect the distribution and variability of population sizes have on the semivariogram estimation and the kriging of the risk. Our concern is to assess the robustness of the Poisson cokriging and direct and cross-semivariogram estimators under homoscedasticity violations. Poisson cokriging, just as Poisson kriging, is a method that accounts for the heteroscedasticity of disease rates, namely, the effect of the population size spatial variations on the variance (Goovaerts, 2005). This advantage translates into a robust estimation of the spatial dependency and the kriged values regardless of the underlying distribution of the population (Goovaerts, 2006b).

To generate *partially* heterotopic data, we randomly preserved 30% of the target variable and 100% of the auxiliary variable for every simulation. The shared locations between $\mathcal{R}_\alpha$ and $\mathcal{R}_\beta$, i.e. the isotopic subset of the data, allow inferring the cross-semivariogram in (21) as well as establishing the level of correlation between the count processes.

3.2. Simulation results

3.2.1. Direct and cross-semivariograms estimation

The direct and cross-semivariogram were estimated using (14) and (21) assuming isotropy, and the LMC was fitted using a weighted least-squares. We adjusted the bivariate sample semivariogram to the theoretical structures described in Section 3.1. We estimated and adjusted the direct and cross-semivariograms for the six correlation scenarios and fitted four population size distributions.

(a) Semivariogram $\gamma_\alpha^\mathcal{R}(\mathbf{h})$ (b) Semivariogram $\gamma_\beta^\mathcal{R}(\mathbf{h})$ (c) Semivariogram $\gamma_{\alpha\beta}^\mathcal{R}(\mathbf{h})$

Fig. 3. (a–b) Direct-semivariogram and (c) cross semivariogram estimation summaries when $\rho_Y = 0.6$ and n_α follows a Poisson distribution. Average semivariograms estimate ($\cdot$) and an 'error bar' with one standard deviation of the semivariogram estimates; the graph of the true semivariogram is overlaid.

We display the results for 1000 simulations when $\rho_Y = 0.6$ and n_α follows a Poisson distribution (Fig. 3). Results for remaining correlation and population size distribution scenarios were similar.

The direct semivariogram estimators are approximately unbiased across distances; the average semivariogram estimates overlap with the true semivariogram values. Contrarily, the cross-semivariogram average estimates are downward biased; however, these biases are almost negligible. The sampling variability of both the direct and cross-semivariogram estimators increases as the strength of spatial association decreases: at large distances, the estimates are more dispersed than at short distances. These results also hold for other correlation and population size distribution scenarios. Surprisingly, the correlation between the Poisson variables does not substantially affect the averaged parameter estimates as the bias slightly increases downwards when the correlation drops. The population sizes seem to not influence the semivariogram estimates. However, the skewed beta-binomial-distributed population sizes biased the estimates upwards and downwards. This behaviour likely occurs due to the high proportion of zero counts produced by generating the conditional counts using the beta-binomial distribution.

3.2.2. Prediction of the latent process

We predicted the total risk $\mathcal{R}_\alpha(\mathbf{s}_0)$ at every prediction location $s_0 \in \mathcal{D} = [-10, 10] \times [-10, 10]$ using the sample counts of the target variable $\frac{Y_{\alpha i}}{n_{\alpha i}}$ and auxiliary variable $\frac{Y_{\beta j}}{n_{\beta j}}$. Poisson cokriging predicted the total risk $\mathcal{R}_\alpha(\mathbf{s}_i)$ at 280 locations for 1000 datasets for each simulation scenario. The performance of our model was assessed with four different metrics: the average error, the mean squared prediction error (MSPE), the coverage probability of prediction intervals and Pearson's correlation ($\rho_\mathcal{R}$) between the predicted and actual values.

$$\text{Average error} = \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{k_\alpha} (\hat{\mathcal{R}}_{\alpha i}^m - \mathcal{R}_{\alpha i}^m), \qquad \text{MSPE} = \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{k_\alpha} (\hat{\mathcal{R}}_{\alpha i}^m - \mathcal{R}_{\alpha i}^m)^2$$

$$\text{Coverage (95\%)} = \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{k_\alpha} \mathbf{1} \left\{ \left| \frac{\hat{\mathcal{R}}_{\alpha i}^m - \mathcal{R}_{\alpha i}^m}{MSPE} \right| < 1.96 \right\}, \qquad \rho_\mathcal{R} = \frac{1}{M} \sum_{m=1}^{M} \rho(\hat{\mathcal{R}}_\alpha^m, \mathcal{R}_\alpha^m)$$

where $\hat{\mathcal{R}}_{\alpha i}^m$ is the predicted value of $\mathcal{R}_{\alpha i}^m$ at prediction location $\mathbf{s}_i$ in the mth simulation. The metrics were computed for each simulation scenario.

Table 1 shows the assessment metrics of the Poisson cokriging predictor for every simulated scenario. The assessment metrics improve when the correlation between the target and auxiliary variables increases: the average errors tend to be 0, the MSPE exhibits small values, the coverage probabilities are closer to 100%, and a strong positive correlation between the predicted and true values is appreciated. As in traditional cokriging, the correlation between the target and auxiliary variables determines the gain in predicting the former (Cressie, 1993). The distribution

Table 1
Summary assessment metrics for the simulated datasets.

ρ_Y	n_α distribution	Average error	MSPE	Coverage	$\rho_\mathcal{R}$
$\to 0$	Uniform	0.045	1.342	0.847	0.559
	Poisson	0.091	1.849	0.860	0.583
	Beta-binomial	−0.073	1.352	0.860	0.549
0.2	Uniform	0.010	0.803	0.946	0.685
	Poisson	0.050	0.607	0.989	0.691
	Beta-binomial	−0.071	0.640	0.952	0.688
0.4	Uniform	0.126	0.527	0.983	0.730
	Poisson	0.123	0.451	0.951	0.704
	Beta-binomial	−0.165	0.928	0.700	0.810
0.6	Uniform	0.042	0.465	0.966	0.788
	Poisson	0.044	**0.417**	0.917	0.791
	Beta-binomial	−0.031	0.641	0.900	0.793
0.8	Uniform	0.029	0.364	0.971	0.880
	Poisson	0.032	**0.320**	0.964	0.873
	Beta-binomial	−0.087	0.412	0.971	0.837
1.0	Uniform	0.057	0.2404	0.964	0.951
	Poisson	0.051	**0.210**	0.971	0.979
	Beta-binomial	−0.026	0.295	0.953	0.920

of the population sizes changes noticeably in the assessment metrics. Overall, the average error indicates that Poisson cokriging over-predicts when the population sizes follow a uniform and Poisson distribution. Surprisingly, under-predictions are present in the simulated data with highly skewed population sizes, i.e., the beta-binomial distribution as the average error adopts negative values. The uniform distribution is associated with the lowest average error in most correlation scenarios. The Poisson cokriging estimator is thus sensitive to the population sizes distribution, mainly if the correlation ρ_Y is high. The coverage probability of the distribution scenarios is reasonably consistent; no gross differences are observed. The Pearson correlation between the predicted and actual risks unequivocally shows that highly skewed population sizes worsen the predictions of Poisson cokriging. In all scenarios, the beta-binomial distribution has the lowest ρ_Y. Large differences in correlation are only found between the beta-binomial distribution and its counterparts.
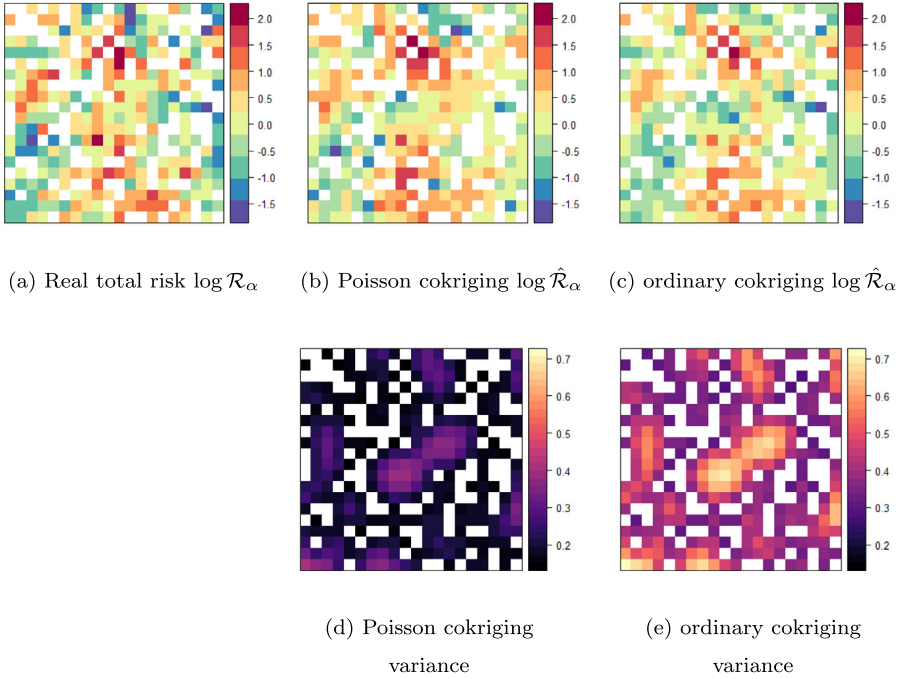
3.2.3. Poisson cokriging evaluation

In order to evaluate the contribution of Poisson cokriging to the geostatistical analysis of spatial count data, we compared its prediction (and error variance) to those of ordinary cokriging (Matheron, 1965; Goovaerts, 1998). We will consider a specific scenario when the correlation $\rho_Y = 0.6$ (moderate correlation) and the population size follows a beta-binomial distribution (highly spatially varying populations).

We assessed the methods using the assessment metrics described in the previous section, except the coverage probability. The results are displayed in Table 2. Our model outperforms ordinary cokriging. The average error of Poisson cokriging is close to zero. The models seem to overestimate the total risk; this is a result of the population sizes. Using Poisson cokriging reduces the MSPE by 46% as compared to ordinary kriging. This gain occurs since traditional geostatistical interpolators poorly deal with the problem of non-stationarity induced by the population sizes (Oliveira, 2014; Bellier et al., 2010; Goovaerts and Gebreab, 2008; Goovaerts, 2017). The linear correlation between the true total risk and the predicted values is the highest in our proposed method.

A visual inspection further corroborates Poisson cokriging's convenience against ordinary cokriging. Fig. 4 presents the real total risk map in the prediction locations in contrast with the prediction and variance errors obtained by Poisson cokriging and ordinary cokriging. Global patterns are similar in the two interpolations, but our method better captures the spatial risk patterns regardless of the smoothing effect produced. Poisson cokriging, as Poisson kriging, is a noise-filtered algorithm (Goovaerts and Gebreab, 2008); the model cleans the effect of non-systematics

Table 2

Summary assessment metrics to constrain Poisson cokriging and ordinary for the selected scenario and the 1000 simulated dataset.

Model	Metric		
	Average error	MSPE	$\rho_{\mathcal{R}}$
Poisson cokriging	−0.0423	0.648	0.780
Ordinary cokriging	−0.1670	1.203	0.595

(a) Real total risk log $\mathcal{R}_\alpha$ (b) Poisson cokriging log $\hat{\mathcal{R}}_\alpha$ (c) ordinary cokriging log $\hat{\mathcal{R}}_\alpha$

(d) Poisson cokriging
variance

(e) ordinary cokriging
variance

Fig. 4. Maps of the (a) true total risk log $\mathcal{R}_\alpha$, (b–c) predicted total risk and (d–e) variance of the prediction error using ordinary cokriging and Poisson kriging when $\rho_Y = 0.6$ and beta-binomial-distributed population sizes (average of 1000 simulations).

measurement errors from the observed rates, which translates into smoothed values. While the ordinary cokriging prediction map suggests adequate spatial risk variability modelling, the predicted values seem noisy and artificial, particularly in areas with blended high and low risks. The kriging variance maps show similar spatial distributions of the prediction error for both techniques. High kriging variances are generally found in the simulated grid regions where the observed risk is the most heterogeneous. The variances in ordinary cokriging are appreciably higher (24%) than in our model. The highest variances in Fig. 4 are proportional to the heterogeneous population sizes in such areas.

4. Application to disease risk data

In this section, we apply Poisson cokriging to disease risk data of two mosquito-borne diseases (Zika and Chikungunya) in two regions of Colombia (Antioquia and Cordoba) for 2016. For the analysis, public health information provided by the Public health surveillance system SIVIGILA (acronym in Spanish) was employed.

Mosquito-borne diseases are prevalent in tropical and subtropical regions of the world. Colombia has the highest incidence of diseases spread by mosquitoes as it possesses climatic, environmental, and socio-geographic conditions favouring their growth and development (Parselia et al., 2019). By 2018, Colombia reported more than 8 million cases of mosquito-borne diseases such as Dengue, Malaria, Zika, and Chikungunya (Borchering et al., 2019), the last two introduced in the country between 2014 and 2015. Studies have demonstrated that Chikungunya and Zika spatial distribution overlap in that they share common epidemiological characteristics such as vectors (i.e. the *Aedes* species), the population at risk and environmental risk factors (Mercado-Reyes et al., 2019a; Riou et al., 2017; Magalhaes et al., 2018). Furthermore, Zika and Chikungunya predominately co-circulate within the same region and populations, and the infection from one virus can trigger or coexist with the infection of the other virus (Mercado-Reyes et al., 2019b).

The first outbreak of the Zika virus disease in Colombia occurred in early October 2015. In 2016 Zika cases rose dramatically, especially in the country's Northwestern region, where alarming mosquito reservoir densities and poor socio-economical conditions prevail. Chikungunya circulates in Colombia since 2014. Rates were of public concern in the first three months following its introduction in Colombia, but the figures rapidly flattened in the subsequent months. Still, the rising cases of the new emerging disease Zika, encouraged the re-occurrence of Chikungunya, particularly in the regions with the most susceptible populations (Wichit et al., 2021; Mercado-Reyes et al., 2019b).

Poisson cokriging was adopted to predict and produce smooth maps of Zika incidence rates in the municipalities of Antioquia and Cordoba for 2016, using the Chikungunya incidence rates at the same period as auxiliary information. Given that our primary ambition is to estimate the spatial structure of the underlying Zika incidence rates using auxiliary information, Poisson cokriging is a suitable method. To convert the areal data into geostatistical data for this application, we have collapsed the municipalities from areal-level to their centroid at the point-level. Although we recognise the theoretical limitations of this method, it facilitates our geostatistical analysis in a similar fashion to classical block kriging (Olea, 1999)

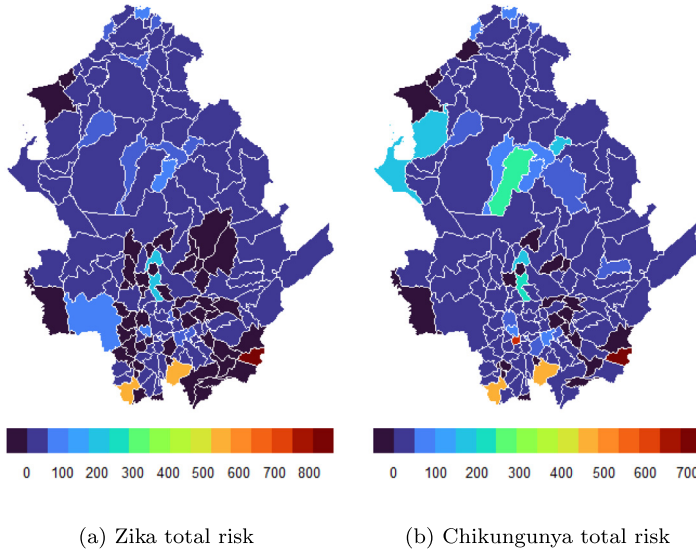## 4.1. Zika and Chikungunya incidence rates

The target variable refers to the Zika incidence rates and the auxiliary variable to the Chikungunya incidence rates. The sample rates were calculated by dividing the total number of cases per disease, namely the sum of the marginal and joint cases, over the population at risk at the period analysed. Fig. 5 shows the total incidence risk for the two mosquito-borne diseases.

The total risk for Zika and Chikungunya varies within the 155 municipalities that compose the Antioquia and Cordoba districts. In most of the region, the incidence risk for both viruses is relatively low; it oscillates between 0 and 200 cases per 10,000 population. There are, however, some municipalities in the northern region with higher risk values that range from the 600 to 700 cases per 10,000 population. Exceptionally, the three municipalities of Antioquia: Andes, Abejorral and Puerto Triunfo (orange and red polygons) present the highest incidence for both arboviruses. These areas tend to represent small municipalities with high Zika and Chikungunya infections.

Overall, the spatial patterns of both infectious diseases are similar. High total risks locate primarily in the Southern region of the study area. Moderate risk is found in the central region, and low-risk values replicate in most municipalities.

The municipalities' population sizes are presented in Fig. 6. The log map demonstrates the highly variable character of the population in the study area. At least 90% of the municipalities have a population inferior to 100,000 inhabitants corresponding to log values between 8 and 13. Solely 11 municipalities have large populations extending from 100,000 to 2,000,000 inhabitants (log values above 13). Poisson cokriging is a sensible alternative to model the incidence rates computed from the vastly different population sizes of Antioquia and Cordoba.
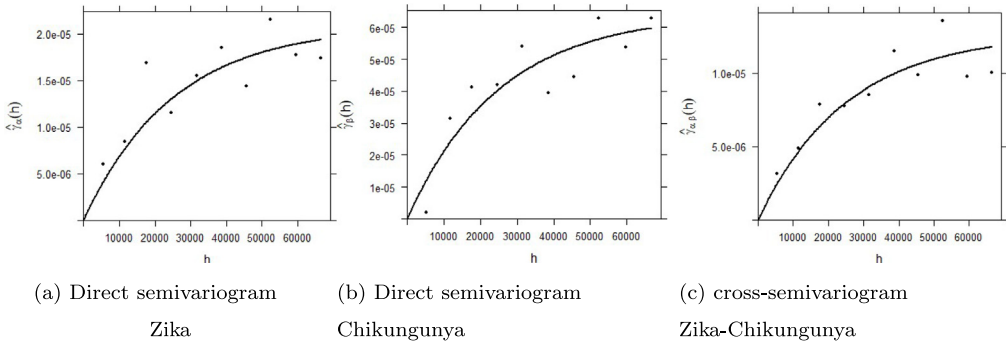
We calculated the correlation coefficient defined by Kawamura (1973) to assess the level of association between the counts of the two diseases. We found an average $\rho_Y = 0.49$. This number indicates that Poisson cokriging will exploit the Chikungunya incidence rates to predict Zika incidence rates.

(a) Zika total risk   (b) Chikungunya total risk

Fig. 5. Maps of the total risk of (a) Zika and (b) Chikungunya incidence in Antioquia and Cordoba (Colombia) for 2016. The legend displays the incidence rate per 10,000 population.
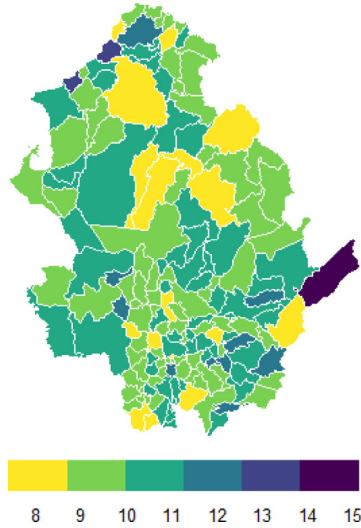
Fig. 6. Logarithmic map of the total Population at risk in Antioquia and Cordoba, Colombia for 2016.

(a) Direct semivariogram   (b) Direct semivariogram   (c) cross-semivariogram

Zika   Chikungunya   Zika-Chikungunya

Fig. 7. Estimation and fitting of the Poisson cokriging (a–b) direct-semivariograms and (c) cross-semivariograms under the LMC: Zika and Chikungunya incidence risks.

To estimate the semivariograms and krige at unsampled locations, the Zika incidence risks dataset was randomly split into prediction (70%) and sampling locations (30%). We use the complete information (100%) of the Chikungunya rates as the auxiliary variable.

4.2. Direct and cross-semivariogram estimation

The direct and cross-experimental semivariograms defined in Eqs. (14) and (21) were obtained under the assumption of isotropy using the sampling locations of the target and auxiliary variable. An exponential semivariogram model was fitted using the linear model of coregionalisation. The direct semivariogram and cross-semivariogram showed no nugget effect. Fig. 7 displays the experimental direct and cross-semivariograms and their corresponding fitting using weighted least-squares.

Based on the semivariograms, the spatial dependency within and between the total risks weakens at distances greater than 70 km. A significant amount of variability is appreciable in the data set, notably when the distances increase. Nonetheless, most point estimates suffer from biases due to the overdispersion in the data. Nonetheless, the fit of the semivariograms using the LMC shows good results.

The estimated semivariogram parameters are $b_1 = 2.130 \times 10^{-5}$, $b_2 = 6.42 \times 10^{-5}$, $b_{12} = 1.246 \times 10^{-5}$ as the sills of the direct semivariograms of Zika and Chikungunya and the cross-semivariogram, respectively. The three semivariograms shared an effective range of $a = 25$ km, which means that the spatial auto and cross-correlation are rather on a local-level scale and primarily involve the first-order neighbouring municipalities.

4.3. Poisson cokriging

Fig. 8 displays the spatial distribution of the true and predicted total risk for Zika for the predicted locations (70% data). The spatial patterns, as well as the range of the risk values, are generally preserved. It is evident the effect the Chikungunya rates have on the predicted Zika rates: predicted values are inflated in those municipalities where high Chikungunya risk occurs. The predicted risk is smoothed with excessively low and high figures filtered out. Nonetheless, the municipalities with extreme risks are still identifiable. Municipalities with sparse neighbours or scarce surrounding observations trend towards the mean.

The kriging variances associated with the risk map in Fig. 8.b are mapped in Fig. 9. The prediction variance ranges from 10 to 50 cases per 10,000 population approximately. Low variances mainly

(a) Zika observed total risk   (b) Zika predicted total risk
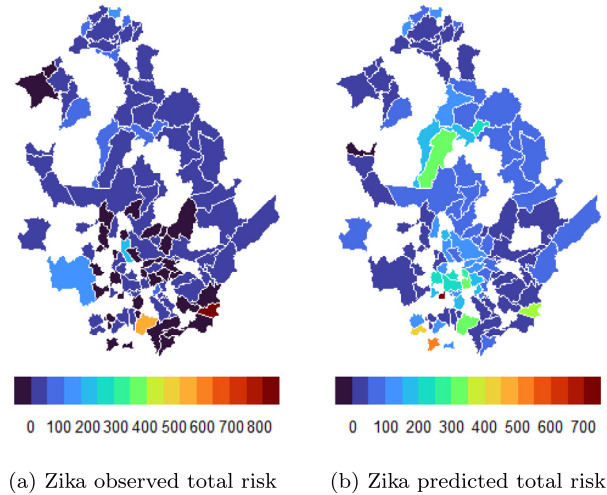
Fig. 8. Maps of the (a) observed and (b) predicted (Poisson cokriging) total risk of Zika incidence in Antioquia and Cordoba (Colombia) for 2016. The legend displays the incidence rate per 10,000 population.

occur along the southern region of the study area, where sampling locations abound and the distances between centroids are short. In these cases, the information provided by neighbouring municipalities is rich, leading to smaller kriging variances. High variances mainly affect municipalities surrounded by small population sizes and low-risk values. The effect of the population size is negligible in the kriging variance maps as it was incorporated into the Poisson cokriging system. The variance of the observed sample rates is filtered out from the variance of the total risk distributions (Monestiez et al., 2006).

5. Discussion and conclusion

We have introduced a novel geostatistical method for bivariate spatially correlated count data. From a methodological perspective, we have adapted classical multivariate geostatistics, appealing for their simplicity in dealing with spatially correlated variables, and the bivariate Poisson distribution, a model to deal with bivariate correlated count data, to develop an interpretable and transparent method to predict spatial counts in the presence of auxiliary information. The driving motivation emerged from disease risk applications where closely related diseases can explain their mutual occurrence. In particular, it is interesting to understand disease rate co-dependence using spatial counts. By conducting a simulation study, we have shown the robustness of Poisson cokriging under multiple scenarios and its superiority against alternative methods. An application to real disease count data demonstrated our method's convenience to interpolate disease rates at unobserved locations when ancillary data are available.

The simulation results show a large increase in prediction accuracy in terms of the assessment metrics considered in this study (Table 2). Such an increase is a direct consequence of introducing a secondary variable in the analysis. The simulation also shows how robust Poisson cokriging is under high variation in spatially varying population sizes. The method successfully denoises the measurement errors and stabilises the variances by adding weighting and correction bias terms in both the semivariogram estimators and the cokriging equations system. In a real data application to Zika and Chikungunya incidence counts data in Colombia, we observed our model's strength in capturing and modelling the spatial dependence between counts. Moreover, the high correspondence
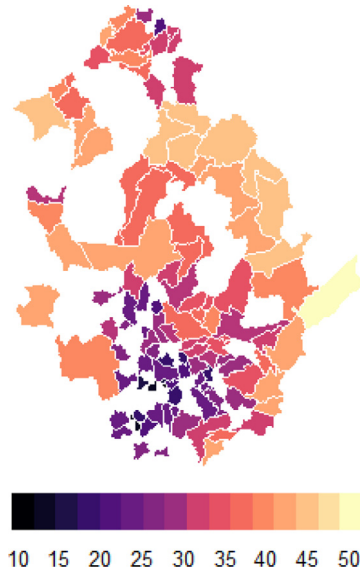
Fig. 9. Map of the kriging variance of the total risk of Zika incidence in Antioquia and Cordoba (Colombia) for 2016. The legend displays the incidence rate per 10,000 population.

between the real and predicted values and low prediction kriging variances suggests that Poisson cokriging is a suitable method for bivariate spatial prediction.

Despite the advantages of Poisson cokriging as a bivariate geostatistical estimator, it unavoidably inherits limitations from its founding theory, the bivariate Poisson distribution proposed by Kawamura (1973). First, the model cannot account for a negative correlation between the target and auxiliary variable; that is, the correlation coefficient ρ_Y must be strictly positive. This drawback restrains the applicability of Poisson cokriging to some specific applications. For example, in ecology, predator–prey interactions are often negatively correlated (Cortez and Weitz, 2014): the scarcer predators habit in an ecosystem, the higher the prey population becomes. Our model will not be able to model and predict the prey/predator population counts of such interaction. Nonetheless, this limitation has little impact on our disease mapping scope. Our primary concern is the positive (or zero) correlation between two diseases, i.e., the occurrence of one disease is associated with an increased likelihood of the other disease occurring. Specifically, we aim to address the question of how an increase in cases of disease A in a geographical unit influences the rise of both disease A and disease B cases in the same and surrounding units. The answer to this question has important implications for outbreak management, pandemic response, and disease control. We believe that our model's assumption of positive correlation can potentially guide the implementation of combined control measures, early detection and rapid response strategies, and minimise the impact of co-occurring diseases, as it might serve as an indication of causality between the two conditions. Another latent limitation is that the correlation ρ_Y cannot encompass every possible positive correlation. The coefficient ρ_Y merely comprises the correlations included in the squared root of the ratio between the total risks, i.e. $min\left[\sqrt{\mathcal{R}_\alpha/\mathcal{R}_\beta}, \sqrt{\mathcal{R}_\beta/\mathcal{R}_\alpha}\right]$ (Holgate, 1966). Nonetheless, this restriction does not present challenges for analysing spatial bivariate health count data, as we defined the correlation based on the dependence between the observed data. On another note, advancing the model to incorporate further auxiliary variables is problematic. While (Kawamura,

1973)'s theory allows for pairwise covariance and naturally adjusts to the cokriging covariance structure, introducing additional covariates makes the cross-semivariogram weights and Poisson cokriging equations system nearly intractable. Finally, Poisson cokriging cannot operate with count variables whose denominators (e.g. population sizes, sampling effort) differ. This latent disadvantage confines our method's analysis to rates whose independent quantities are identical.

Other limitations, common in geostatistics, also challenge Poisson cokriging's adoption in practical settings. For instance, the change of support problem. In our real data application, we employed Poisson cokriging to make municipality-level inferences using a centroid-based approach, transforming the area-level data to point-level data. This assumption oversimplifies the data latent spatial configuration as it implies that the distribution of both the population and the cases within a municipality are homogeneous. Additionally, the distance computed between such centroids might be artificial, leading to incorrect estimation of the semivariogram, over-smoothed maps, and higher prediction errors (Goovaerts, 2006b). Under this scenario, our model should explicitly handle the data change of support and account for distances that mirror the actual proximity between locations (i.e. population-weighed distances). Potential avenues for future research that we are considering to mitigate the challenges arising from transitioning from areal to point data include alternative geostatistical methods such as Area-to-Area kriging, Area-to-Point kriging (Goovaerts, 2006b; Goovaerts and Gebreab, 2008), weighted centroids, and incorporating point-based ancillary data.

As future work, Poisson cokriging can be improved both methodologically and practically. One natural methodological extension is to deal with more than one auxiliary variable. The cokriging method improves predictions by exploiting multiple sources of information, that is, several covariates (Wackernagel, 2003a). A potential approach includes (Mahamunulu, 1967; Kawamura, 1979; Karlis and Meligkotsidou, 2007) definitions of the multivariate Poisson distribution. Nevertheless, attention is needed as these lead to complex covariance structures making multivariate probability mass computationally and theoretically unmanageable. Accounting for negative correlation is another future improvement to our model. Although in epidemiological applications, one is often concerned with the positive stimuli diseases exert on each other for mitigation and resource allocation purposes, inverse relationships, while rare, might also arise. For instance, the inverse relationship between Alzheimer's disease and cancer (Li et al., 2014). Other research paths for Poisson cokriging include introducing the temporal dimension to comprehensive spatiotemporal analyses and external drift to account for shared co-founding effects between the variables. It is worth noting that we are currently working on extending Poisson cokriging to include covariates, in a similar fashion to kriging with external drift and kriging regression. We also believe a collocated version of our method would benefit if data are completely heterotopic.

Poisson cokriging, as a frequentist method for bivariate spatial count data, is often easier to interpret, implement and scale than its Bayesian counterparts. Our method avoids defining prior distributions and using iterative parameter estimation procedures, such as Markov Chain Monte Carlo (MCMC) or integrated nested Laplace approximation (INLA), which are computer intensive and hard to understand for most practitioners outside statistics. Poisson cokriging, similar to Poisson kriging, can be used for both interpolation and smoothing purposes (Goovaerts, 2017). In contrast to the prevalent univariate and multivariate techniques employed for disease risk mapping, such as conditional autoregressive (CAR) and multivariate autoregressive (MCAR) models, Poisson cokriging permits the estimation of risk at unobserved locations, provides a flexible modelling of the spatial (co)dependency structure, and optimal smoothing effect (Goovaerts and Gebreab, 2008). Moreover, the simplicity of the technique allows for easily inferring the underlying spatial structure between variables and predicting count variables with limited observations. Our study's findings are expected to be of significant interest to epidemiologists, ecologists, and experts from various fields where count data thrives.

In conclusion, this paper introduces a bivariate geostatistical method to predict a spatially correlated count variable in the presence of another variable of the exact nature. This model is particularly beneficial in domains where the variable of interest is under-sampled relative to a correlated generously available covariate and when the variables involved are rates computed from highly varying population sizes. Furthermore, as Poisson cokriging is a geostatistical method, its implementation and interpretation are unambiguous, and the computational burden is modest compared to alternative methods. The Poisson cokriging implementation can be found at https://github.com/DavidPayares/PCK.

Appendix A. Direct-semivariogram and cross-semivariogram calculations

A.1. Direct semivariogram

The relationships between the sample rates at two different spatial locations $\mathbf{s}_i$ and $\mathbf{s}_j$ are needed to infer the theoretical semivariogram. Using (3) and (7) we get the first order moment of $(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}})$:

$$
\begin{aligned}
\mathrm{E}\left[\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}} \mid \mathcal{R}_\alpha\right] &= \frac{1}{n_{\alpha i}}\,\mathrm{E}\left[Y_{\alpha i} \mid \mathcal{R}_\alpha(\mathbf{s}_i)\right] - \frac{1}{n_{\alpha j}}\,\mathrm{E}\left[Y_{\alpha j} \mid \mathcal{R}_\alpha(\mathbf{s}_j)\right] \\
&= \frac{1}{n_{\alpha i}}(n_{\alpha i} \cdot (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))) - \frac{1}{n_{\alpha j}}(n_{\alpha j} \cdot (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))) \\
&= (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) - (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j)) \\
\Rightarrow \mathrm{E}\left[\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right] &= \mathrm{E}\left[\mathrm{E}\left[\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}} \mid \mathcal{R}_\alpha\right]\right] \\
&= \mathrm{E}\left[(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) - (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\right] \\
&= (m_\alpha + m_{\alpha\beta}) - (m_\alpha + m_{\alpha\beta}) \\
&= 0
\end{aligned}
$$

and the second order moment,

$$
\begin{aligned}
\mathrm{E}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right)^2 \mid \mathcal{R}_\alpha\right] &= \mathrm{E}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}}\right)^2 + \left(\frac{Y_{\alpha j}}{n_{\alpha j}}\right)^2 - 2\frac{Y_{\alpha i}Y_{\alpha j}}{n_{\alpha i}n_{\alpha j}} \mid \mathcal{R}_\alpha\right] \\
&= \frac{1}{n_{\alpha i}^2}\,\mathrm{E}\left[Y_{\alpha i}^2 \mid \mathcal{R}_\alpha\right] + \frac{1}{n_{\alpha j}^2}\,\mathrm{E}\left[Y_{\alpha j}^2 \mid \mathcal{R}_\alpha\right] - \frac{2}{n_{\alpha i}n_{\alpha j}}\,\mathrm{E}\left[Y_{\alpha i}Y_{\alpha j} \mid \mathcal{R}_\alpha\right] \\
&= \frac{1}{n_{\alpha i}^2}\,\mathrm{E}\left[Y_{\alpha i}^2 \mid \mathcal{R}_\alpha\right] + \frac{1}{n_{\alpha j}^2}\,\mathrm{E}\left[Y_{\alpha j}^2 \mid \mathcal{R}_\alpha\right] - \frac{2}{n_{\alpha i}n_{\alpha j}}(\mathrm{Cov}[Y_{\alpha i}, Y_{\alpha j} \mid \mathcal{R}_\alpha] \\
&\quad + \mathrm{E}\left[Y_{\alpha i} \mid \mathcal{R}_\alpha\right]\mathrm{E}\left[Y_{\alpha j} \mid \mathcal{R}_\alpha\right]) \\
&= \frac{1}{n_{\alpha i}^2}\left(n_{\alpha i} \cdot (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) + n_{\alpha i}^2\left(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)\right)^2\right) \\
&\quad + \frac{1}{n_{\alpha j}^2}\left(n_{\alpha j} \cdot (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j)) + n_{\alpha j}^2\left(r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j)\right)^2\right) \\
&\quad - \frac{2}{n_{\alpha i}n_{\alpha j}}\left(\delta_{ij}(n_{\alpha i} \cdot (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))) + n_{\alpha i} \cdot (r_\alpha(\mathbf{s}_i)\right. \\
&\quad \left. + r_{\alpha\beta}(\mathbf{s}_i))\,n_{\alpha j} \cdot (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\right)
\end{aligned}
$$

$$
= \frac{(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))}{n_{\alpha i}} + (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))^2 + \frac{(r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))}{n_{\alpha j}}
$$
$$
+ (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))^2 - 2\delta_{ij}\frac{(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))}{n_{\alpha i}}
$$
$$
- 2(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))(r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))
$$
$$
= \frac{(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))}{n_{\alpha i}} + \frac{(r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))}{n_{\alpha j}} - 2\delta_{ij}\frac{(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))}{n_{\alpha i}}
$$
$$
+ \left((r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) - (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\right)^2
$$
$$
\Rightarrow \mathrm{E}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right)^2\right] = \mathrm{E}\left[\frac{(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))}{n_{\alpha i}} + \frac{(r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))}{n_{\alpha j}} - 2\delta_{ij}\frac{(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))}{n_{\alpha i}}\right.
$$
$$
\left. + \left((r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) - (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\right)^2\right]
$$
$$
= \frac{m_\alpha + m_{\alpha\beta}}{n_{\alpha i}} + \frac{m_\alpha + m_{\alpha\beta}}{n_{\alpha j}} - 2\delta_{ij}\frac{m_\alpha + m_{\alpha\beta}}{n_{\alpha i}}
$$
$$
+ \mathrm{E}\left[\left((r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) - (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\right)^2\right]
$$
$$
\frac{1}{2}\mathrm{E}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right)^2\right] = \frac{1}{2}(m_\alpha + m_{\alpha\beta})\left(\frac{1}{n_{\alpha i}} + \frac{1}{n_{\alpha j}}\right) - \delta_{ij}\frac{m_\alpha + m_{\alpha\beta}}{n_{\alpha i}} + \frac{1}{2}
$$
$$
\mathrm{E}\left[\left((r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) - (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\right)^2\right]
$$

We can re-write the equation in terms of $\mathcal{R}_\alpha(\mathbf{s})$ as the total risk $\mathcal{R}_\alpha(\mathbf{s}_i) = r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)$:

$$
\frac{1}{2}\mathrm{E}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right)^2\right] = \frac{1}{2}(m_\alpha + m_{\alpha\beta})\left(\frac{1}{n_{\alpha i}} + \frac{1}{n_{\alpha j}}\right) - \delta ij\frac{m_\alpha + m_{\alpha\beta}}{n_{\alpha i}}
$$
$$
+ \frac{1}{2}\mathrm{E}\left[(\mathcal{R}_\alpha(\mathbf{s}_i) - \mathcal{R}_\alpha(\mathbf{s}_j))^2\right]
$$

The expectation of the conditional variance of the difference in the squared rates when $i \neq j$ is,

$$
\mathrm{Var}\left[\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}} \mid \mathcal{R}_\alpha\right] = \mathrm{E}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right)^2 \mid \mathcal{R}_\alpha\right] + \mathrm{E}^2\left[\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}} \mid \mathcal{R}_\alpha\right]
$$
$$
= \frac{(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))}{n_{\alpha i}} + \frac{(r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))}{n_{\alpha j}}
$$
$$
+ \left((r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) - (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\right)^2
$$
$$
- \left((r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) - (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\right)^2
$$
$$
= \frac{(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))}{n_{\alpha i}} + \frac{(r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))}{n_{\alpha j}}
$$
$$
\Rightarrow \mathrm{E}\left[\mathrm{Var}\left[\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}} \mid \mathcal{R}_\alpha\right]\right] = \mathrm{E}\left[\frac{(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))}{n_{\alpha i}} + \frac{(r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))}{n_{\alpha j}}\right]
$$
$$
= (m_\alpha + m_{\alpha\beta})\left(\frac{n_{\alpha i} + n_{\alpha j}}{n_{\alpha i}n_{\alpha j}}\right)
$$

A.2. Cross-semivariogram

Through (15) and (3) we get the first order moment of $(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\beta j}}{n_{\beta j}})$:

$$\mathrm{E}\left[\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}} \mid \mathcal{R}\right] = (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) - (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))$$

$$\Rightarrow \mathrm{E}\left[\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right] = (m_\alpha + m_{\alpha\beta}) - (m_\alpha + m_{\alpha\beta})$$

$$= 0$$

And identically for $\mathrm{E}\left[\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}} \mid \mathcal{R}\right]$.

We also get that:

$$\mathrm{Cov}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right)\left(\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}}\right) \mid \mathcal{R}\right] = \mathrm{E}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right), \left(\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}}\right) \mid \mathcal{R}\right]$$

$$= \mathrm{E}\left[\frac{Y_{\alpha i}Y_{\beta i}}{n_{\alpha i}n_{\beta i}} - \frac{Y_{\alpha i}Y_{\beta j}}{n_{\alpha i}n_{\beta j}} - \frac{Y_{\alpha j}Y_{\beta i}}{n_{\alpha j}n_{\beta i}} + \frac{Y_{\alpha j}Y_{\beta j}}{n_{\alpha j}n_{\beta j}} \mid \mathcal{R}\right]$$

$$= \frac{1}{n_{\alpha i}n_{\beta i}}\mathrm{E}\left[Y_{\alpha i}Y_{\beta i} \mid \mathcal{R}\right] - \frac{1}{n_{\alpha i}n_{\beta j}}\mathrm{E}\left[Y_{\alpha i}Y_{\beta j} \mid \mathcal{R}\right]$$

$$- \frac{1}{n_{\alpha j}n_{\beta i}}\mathrm{E}\left[Y_{\alpha j}Y_{\beta j} \mid \mathcal{R}\right] + \frac{1}{n_{\alpha j}n_{\beta j}}\mathrm{E}\left[Y_{\alpha j}Y_{\beta j} \mid \mathcal{R}\right]$$

$$= \frac{1}{n_{\alpha i}n_{\beta i}}\Big[n_{\alpha i}r_{\alpha\beta}(\mathbf{s}_i) + n_{\alpha i}\cdot(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))\cdot n_{\beta i}\cdot(r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))\Big]$$

$$- \frac{1}{n_{\alpha i}n_{\beta j}}\Big[\delta_{\alpha\beta ij}\big(n_{\alpha i}r_{\alpha\beta}(\mathbf{s}_i)\big) + n_{\alpha i}\cdot(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))\cdot n_{\beta j}\cdot(r_\beta(\mathbf{s}_j)$$
$$+ r_{\alpha\beta}(\mathbf{s}_j))\Big]$$

$$- \frac{1}{n_{\alpha j}n_{\beta i}}\Big[\delta_{\alpha\beta ji}\big(n_{\alpha j}r_{\alpha\beta}(\mathbf{s}_j)\big) + n_{\alpha j}\cdot(r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\cdot n_{\beta i}\cdot(r_\beta(\mathbf{s}_i)$$
$$+ r_{\alpha\beta}(\mathbf{s}_i))\Big]$$

$$+ \frac{1}{n_{\alpha j}n_{\beta j}}\Big[n_{\alpha j}r_{\alpha\beta}(\mathbf{s}_j) + n_{\alpha j}\cdot(r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\cdot n_{\beta j}\cdot(r_\beta(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\Big]$$

$$= \frac{r_{\alpha\beta}(\mathbf{s}_i)}{n_{\alpha i}} + (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))(r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))$$

$$- \frac{r_{\alpha\beta}(\mathbf{s}_i)}{n_{\beta j}}\delta_{\alpha\beta ij} - (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))(r_\beta(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))$$

$$- \frac{r_{\alpha\beta}(\mathbf{s}_j)}{n_{\beta i}}\delta_{\alpha\beta ij} - (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))(r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))$$

$$+ \frac{r_{\alpha\beta}(\mathbf{s}_j)}{n_{\beta j}} + (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))(r_\beta(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))$$

$$= \frac{r_{\alpha\beta}(\mathbf{s}_i)}{n_{\alpha i}} - \frac{r_{\alpha\beta}(\mathbf{s}_i)}{n_{\beta j}}\delta_{\alpha\beta ij} - \frac{r_{\alpha\beta}(\mathbf{s}_j)}{n_{\beta i}}\delta_{\alpha\beta ij} + \frac{r_{\alpha\beta}(\mathbf{s}_j)}{n_{\beta j}}$$

$$+ \Big[\big((r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) - (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\big)\big((r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) - (r_\beta(\mathbf{s}_j)$$
$$+ r_{\alpha\beta}(\mathbf{s}_j))\big)\Big]$$

$$\Rightarrow \mathrm{Cov}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right), \left(\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}}\right)\right] = \mathrm{E}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right)\left(\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}}\right)\right]$$

$$= \mathrm{E}\left[\frac{r_{\alpha\beta}(\mathbf{s}_i)}{n_{\alpha i}} - \frac{r_{\alpha\beta}(\mathbf{s}_i)}{n_{\beta j}}\delta_{\alpha\beta ij} - \frac{r_{\alpha\beta}(\mathbf{s}_j)}{n_{\beta i}}\delta_{\alpha\beta ij} + \frac{r_{\alpha\beta}(\mathbf{s}_j)}{n_{\beta j}}\right.$$
$$+ \left[\left((r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) - (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\right)\right.$$
$$\times \left(\,(r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))\right.$$
$$\left.\left.-(r_\beta(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\,\right)\right]\Bigg]$$

$$= \frac{m_{\alpha\beta}}{n_{\alpha i}} - \frac{m_{\alpha\beta}}{n_{\beta j}}\delta_{\alpha\beta ij} - \frac{m_{\alpha\beta}}{n_{\beta i}}\delta_{\alpha\beta ij} + \frac{m_{\alpha\beta}}{n_{\alpha j}}$$
$$+ \mathrm{E}\left[\,\left((r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) - (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\right)\left(\,(r_\beta(\mathbf{s}_i)\right.\right.$$
$$+ r_{\alpha\beta}(\mathbf{s}_i))$$
$$\left.\left.- (r_\beta(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\,\right)\right]$$

$$\frac{1}{2}\mathrm{E}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right)\left(\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}}\right)\right] = \frac{1}{2}\left(\frac{m_{\alpha\beta}}{n_{\alpha i}} + \frac{m_{\alpha\beta}}{n_{\beta j}}\right) - \frac{1}{2}\left(\frac{m_{\alpha\beta}}{n_{\beta j}}\delta_{\alpha\beta ij} + \frac{m_{\alpha\beta}}{n_{\alpha i}}\delta_{\alpha\beta ij}\right)$$
$$+ \frac{1}{2}\mathrm{E}\left[\,\left((r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) - (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\right)\right.$$
$$\times \left(\,(r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))\right.$$
$$\left.\left.- (r_\beta(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\,\right)\right]$$

We can re-write the equation in terms of $\mathcal{R}_\alpha$ and $\mathcal{R}_\beta$ as the total risks $\mathcal{R}_\alpha(\mathbf{s}_i) = r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)$ and $\mathcal{R}_\beta(\mathbf{s}_i) = r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)$, correspondingly.

$$\frac{1}{2}\mathrm{E}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right)\left(\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}}\right)\right] = \frac{1}{2}\left(\frac{m_{\alpha\beta}}{n_{\alpha i}} + \frac{m_{\alpha\beta}}{n_{\beta j}}\right) - \frac{1}{2}\left(\frac{m_{\alpha\beta}}{n_{\beta j}}\delta_{\alpha\beta ij} + \frac{m_{\alpha\beta}}{n_{\alpha i}}\delta_{\alpha\beta ij}\right)$$
$$+ \frac{1}{2}\mathrm{E}\left[\left(\mathcal{R}_\alpha(\mathbf{s}_i) - \mathcal{R}_\alpha(\mathbf{s}_j)\right)\left(\mathcal{R}_\beta(\mathbf{s}_i) - \mathcal{R}_\beta(\mathbf{s}_i)\right)\right]$$

The expectation of the conditional variance of the difference in the squared rates when $i \neq j$ is,

$$\mathrm{Cov}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right), \left(\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}}\right) \mid \mathcal{R}\right] = \mathrm{E}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right)\left(\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}}\right) \mid \mathcal{R}\right]$$
$$- \mathrm{E}\left[\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha i}}{n_{\alpha j}} \mid \mathcal{R}\right]\mathrm{E}\left[\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta i}}{n_{\beta j}} \mid \mathcal{R}\right]$$
$$= \frac{r_{\alpha\beta}(\mathbf{s}_i)}{n_{\alpha i}} + \frac{r_{\alpha\beta}(\mathbf{s}_j)}{n_{\beta j}} + \left(\,(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))\right.$$
$$- (r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\,\big)\big(\,(r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))$$
$$- (r_\beta(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))\,\big)$$
$$= \frac{r_{\alpha\beta}(\mathbf{s}_i)}{n_{\alpha i}} + \frac{r_{\alpha\beta}(\mathbf{s}_j)}{n_{\beta j}}$$

$$\mathrm{E}\left[\mathrm{Cov}\left[\left(\frac{Y_{\alpha i}}{n_{\alpha i}} - \frac{Y_{\alpha j}}{n_{\alpha j}}\right), \left(\frac{Y_{\beta i}}{n_{\beta i}} - \frac{Y_{\beta j}}{n_{\beta j}}\right) \mid \mathcal{R}\right]\right] = \mathrm{E}\left[\frac{r_{\alpha\beta}(\mathbf{s}_i)}{n_{\alpha i}} + \frac{r_{\alpha\beta}(\mathbf{s}_j)}{n_{\beta j}}\right]$$
$$= m_{\alpha\beta}\left(\frac{n_{\alpha i} + n_{\beta j}}{n_{\alpha i}n_{\beta j}}\right)$$

Appendix B. Poisson cokriging variance

The best set of weights $\lambda_{\alpha i}$ and $\lambda_{\beta i}$ should be chosen such that the predicted values at a new location are unbiased and also minimise the prediction variance $\text{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)]$. To compute the prediction variance, we calculate first the conditional expectation

$$
\text{E}\left[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0) \mid \mathcal{R}\right] = \text{E}\left[\sum_{i=1}^{k_\alpha} \lambda_{\alpha i} \frac{Y_{\alpha i}}{n_{\alpha i}} + \sum_{i=1}^{k_\beta} \lambda_{\beta i} \frac{Y_{\beta i}}{n_{\beta i}} - \mathcal{R}_\alpha(\mathbf{s}_0) \mid \mathcal{R}\right]
$$

$$
= \sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}}{n_{\alpha i}} \text{E}\left[Y_{\alpha i} \mid \mathcal{R}\right] + \sum_{i=1}^{k_\beta} \frac{\lambda_{\beta i}}{n_{\beta i}} \text{E}\left[Y_{\beta i} \mid \mathcal{R}\right] - \text{E}\left[\mathcal{R}_\alpha(\mathbf{s}_0) \mid \mathcal{R}\right]
$$

$$
= \sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}}{n_{\alpha i}} (n_{\alpha i}\mathcal{R}_\alpha(\mathbf{s}_i)) + \sum_{i=1}^{k_\beta} \frac{\lambda_{\beta i}}{n_{\beta i}} (n_{\beta i}\mathcal{R}_\beta(\mathbf{s}_i)) - \mathcal{R}_\alpha(\mathbf{s}_0)
$$

$$
= \sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}}{n_{\alpha i}} \left[n_{\alpha i} \cdot (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))\right] + \sum_{i=1}^{k_\beta} \frac{\lambda_{\beta i}}{n_{\beta i}} \left[n_{\beta i} \cdot (r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))\right]
$$
$$
- (r_{\alpha 0} + r_{\alpha\beta 0})
$$

$$
\Rightarrow \text{E}\left[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)\right] = \text{E}\left[\sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}}{n_{\alpha i}} \left[n_{\alpha i} \cdot (r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))\right]\right.
$$

$$
+ \sum_{i=1}^{k_\beta} \frac{\lambda_{\beta i}}{n_{\beta i}} \left[n_{\beta i} \cdot (r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))\right]
$$

$$
\left. - (r_{\alpha 0} + r_{\alpha\beta 0})\right]
$$

$$
= \sum_{i=1}^{k_\alpha} \lambda_{\alpha i} \text{E}[r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)] + \sum_{i=1}^{k_\beta} \lambda_{\beta i} \text{E}[r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)]
$$

$$
- \text{E}[r_{\alpha 0} + r_{\alpha\beta 0}]
$$

$$
= (m_\alpha + m_{\alpha\beta}) \underbrace{\sum_{i=1}^{k_\alpha} \lambda_{\alpha i}}_{1} + (m_\beta + m_{\alpha\beta}) \underbrace{\sum_{i=1}^{k_\beta} \lambda_{\beta i}}_{0} - (m_\alpha + m_{\alpha\beta})
$$

$$
= (m_\alpha + m_{\alpha\beta}) - (m_\alpha + m_{\alpha\beta})
$$

$$
= 0
$$

Then we compute the conditional squared expectation,

$$
\text{E}\left[(\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0))^2 \mid \mathcal{R}\right] = \text{E}\left[\left(\sum_{i=1}^{k_\alpha} \lambda_{\alpha i} \frac{Y_{\alpha i}}{n_{\alpha i}} + \sum_{i=1}^{k_\beta} \lambda_{\beta i} \frac{Y_{\beta i}}{n_{\beta i}} - \mathcal{R}_\alpha(\mathbf{s}_0)\right)^2 \mid \mathcal{R}\right]
$$

$$
= \mathrm{E}\left[\left(\sum_{i=1}^{k_\alpha} \lambda_{\alpha i} \frac{Y_{\alpha i}}{n_{\alpha i}}\right)^2 + 2\left(\sum_{i=1}^{k_\alpha} \lambda_{\alpha i} \frac{Y_{\alpha i}}{n_{\alpha i}}\right)\left(\sum_{i=1}^{k_\beta} \lambda_{\beta i} \frac{Y_{\beta i}}{n_{\beta i}}\right) - 2\left(\sum_{i=1}^{k_\alpha} \lambda_{\alpha i} \frac{Y_{\alpha i}}{n_{\alpha i}}\right)\mathcal{R}_\alpha(\mathbf{s}_0) \right.
$$

$$
+ \left(\sum_{i=1}^{k_\beta} \lambda_{\beta i} \frac{Y_{\beta i}}{n_{\beta i}}\right)^2
$$

$$
\left. - 2\left(\sum_{i=1}^{k_\beta} \lambda_{\beta i} \frac{Y_{\beta i}}{n_{\beta i}}\right)\mathcal{R}_\alpha(\mathbf{s}_0) + \mathcal{R}_{\alpha 0}^2 \mid \mathcal{R} \right]
$$

$$
= \mathrm{E}\left[\sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\alpha} \lambda_{\alpha i}\lambda_{\alpha j} \frac{Y_{\alpha i}Y_{\alpha j}}{n_{\alpha i}n_{\alpha j}} + 2\sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \lambda_{\alpha i}\lambda_{\beta j} \frac{Y_{\alpha i}Y_{\beta j}}{n_{\alpha i}n_{\beta j}} - 2\mathcal{R}_\alpha(\mathbf{s}_0)\sum_{i=1}^{k_\alpha} \lambda_{\alpha i} \frac{Y_{\alpha i}}{n_{\alpha i}} \right.
$$

$$
\left. + \sum_{i=1}^{k_\beta}\sum_{j=1}^{k_\beta} \lambda_{\beta i}\lambda_{\beta j} \frac{Y_{\beta i}Y_{\beta j}}{n_{\beta i}n_{\beta j}} - 2\mathcal{R}_\alpha(\mathbf{s}_0)\sum_{i=1}^{k_\beta} \lambda_{\beta i} \frac{Y_{\beta i}}{n_{\beta i}} + \mathcal{R}_{\alpha 0}^2 \mid \mathcal{R} \right]
$$

$$
= \sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\alpha} \frac{\lambda_{\alpha i}\lambda_{\alpha j}}{n_{\alpha i}n_{\alpha j}} \mathrm{E}[Y_{\alpha i}Y_{\alpha j} \mid \mathcal{R}] + 2\sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \frac{\lambda_{\alpha i}\lambda_{\beta j}}{n_{\alpha i}n_{\beta j}} \mathrm{E}[Y_{\alpha i}Y_{\beta j} \mid \mathcal{R}] - 2\mathcal{R}_\alpha(\mathbf{s}_0)\sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}}{n_{\alpha i}} \mathrm{E}[Y_{\alpha i} \mid \mathcal{R}]
$$

$$
+ \sum_{i=1}^{k_\beta}\sum_{j=1}^{k_\beta} \frac{\lambda_{\beta i}\lambda_{\beta j}}{n_{\beta i}n_{\beta j}} \mathrm{E}[Y_{\beta i}Y_{\beta j} \mid r] - 2\mathcal{R}_\alpha(\mathbf{s}_0)\sum_{i=1}^{k_\beta} \frac{\lambda_{\beta i}}{n_{\beta i}} \mathrm{E}[Y_{\beta i} \mid r] + \mathcal{R}_{\alpha 0}^2
$$

$$
= \sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\alpha} \frac{\lambda_{\alpha i}\lambda_{\alpha j}}{n_{\alpha i}n_{\alpha j}}\left(\underbrace{\delta_{ij}}_{1}\left(n_{\alpha i}\cdot(r_\alpha(\mathbf{s}_i)+r_{\alpha\beta}(\mathbf{s}_i))\right) + n_{\alpha i}\cdot(r_\alpha(\mathbf{s}_i)+r_{\alpha\beta}(\mathbf{s}_i))\, n_{\alpha j}\cdot(r_\alpha(\mathbf{s}_j)+r_{\alpha\beta}(\mathbf{s}_j)) \right)
$$

$$
+ 2\sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \frac{\lambda_{\alpha i}\lambda_{\beta j}}{n_{\alpha i}n_{\beta j}}\left(\underbrace{\delta_{\alpha\beta ij}}_{1}\left(n_{\alpha i}r_{\alpha\beta}(\mathbf{s}_i)\right) + n_{\alpha i}\cdot(r_\alpha(\mathbf{s}_i)+r_{\alpha\beta}(\mathbf{s}_i))\, n_{\beta j}\cdot(r_\beta(\mathbf{s}_j)+r_{\alpha\beta}(\mathbf{s}_j)) \right)
$$

$$
+ \sum_{i=1}^{k_\beta}\sum_{j=1}^{k_\beta} \frac{\lambda_{\beta i}\lambda_{\beta j}}{n_{\beta i}n_{\beta j}}\left(\underbrace{\delta_{ij}}_{1}\left(n_{\beta i}\cdot(r_\beta(\mathbf{s}_i)+r_{\alpha\beta}(\mathbf{s}_i))\right) + n_{\beta i}\cdot(r_\beta(\mathbf{s}_i)+r_{\alpha\beta}(\mathbf{s}_i))\, n_{\beta j}\cdot(r_\beta(\mathbf{s}_j) \right.
$$

$$
\left. + r_{\alpha\beta}(\mathbf{s}_j)) \right)
$$

$$
- 2\mathcal{R}_\alpha(\mathbf{s}_0)\sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}}{n_{\alpha i}}\left(n_{\alpha i}\cdot(r_\alpha(\mathbf{s}_i)+r_{\alpha\beta}(\mathbf{s}_i))\right) - 2\mathcal{R}_\alpha(\mathbf{s}_0)\sum_{i=1}^{k_\beta} \frac{\lambda_{\beta i}}{n_{\beta i}}\left(n_{\beta i}\cdot(r_\beta(\mathbf{s}_i)+r_{\alpha\beta}(\mathbf{s}_i))\right)
$$

$$
+ \mathcal{R}_\alpha(\mathbf{s}_0)^2
$$

$$
= \sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}^2}{n_{\alpha i}}(r_\alpha(\mathbf{s}_i)+r_{\alpha\beta}(\mathbf{s}_i)) + \sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\alpha} \lambda_{\alpha i}\lambda_{\alpha j}(r_\alpha(\mathbf{s}_i)+r_{\alpha\beta}(\mathbf{s}_i))(r_\alpha(\mathbf{s}_j)+r_{\alpha\beta}(\mathbf{s}_j))
$$

$$
+ 2\sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \frac{\lambda_{\alpha i}\lambda_{\beta j}}{n_{\beta j}} r_{\alpha\beta}(\mathbf{s}_i)
$$

$$
+ 2\sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \lambda_{\alpha i}\lambda_{\beta j}(r_\alpha(\mathbf{s}_i)+r_{\alpha\beta}(\mathbf{s}_i))(r_\beta(\mathbf{s}_j)+r_{\alpha\beta}(\mathbf{s}_j)) + \sum_{i=1}^{k_\beta} \frac{\lambda_{\beta i}^2}{n_{\beta i}}(r_\beta(\mathbf{s}_i)+r_{\alpha\beta}(\mathbf{s}_i))
$$

$$+ \sum_{i=1}^{k_\alpha} \sum_{j=1}^{k_\beta} \lambda_{\beta i} \lambda_{\beta j} (r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))(r_\beta(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j)) - 2 \sum_{i=1}^{k_\alpha} \lambda_{\alpha i}(r_{\alpha 0} + r_{\alpha\beta 0})(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))$$

$$- 2 \sum_{i=1}^{k_\beta} \lambda_{\beta i}(r_{\alpha 0} + r_{\alpha\beta 0})(r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) + (r_{\alpha 0} + r_{\alpha\beta 0})^2$$

$$\Rightarrow \mathrm{E}\left[(\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0))^2\right] = \mathrm{E}\left[\sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}^2}{n_{\alpha i}}(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) + \sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\alpha} \lambda_{\alpha i}\lambda_{\alpha j}(r_\alpha(\mathbf{s}_i) \right.$$

$$+ r_{\alpha\beta}(\mathbf{s}_i))(r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j)) + 2 \sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \frac{\lambda_{\alpha i}\lambda_{\beta j}}{n_{\beta j}} r_{\alpha\beta}(\mathbf{s}_i)$$

$$+ 2 \sum_{i=1}^{k_\beta}\sum_{j=1}^{k_\beta} \lambda_{\alpha i}\lambda_{\beta j}(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))(r_\beta(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))$$

$$+ \sum_{i=1}^{k_\beta} \frac{\lambda_{\beta i}^2}{n_{\beta i}}(r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) + \sum_{i=1}^{k_\beta}\sum_{j=1}^{k_\beta} \lambda_{\beta i}\lambda_{\beta j}(r_\beta(\mathbf{s}_i)$$

$$+ r_{\alpha\beta}(\mathbf{s}_i))(r_\beta(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))$$

$$- 2 \sum_{i=1}^{k_\alpha} \lambda_{\alpha i}(r_{\alpha 0} + r_{\alpha\beta 0})(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))$$

$$\left. - 2 \sum_{i=1}^{k_\beta} \lambda_{\beta i}(r_{\alpha 0} + r_{\alpha\beta 0})(r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i)) + (r_{\alpha 0} + r_{\alpha\beta 0})^2 \right]$$

$$= \sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}^2}{n_{\alpha i}} \mathrm{E}[(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))] + \sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\alpha} \lambda_{\alpha i}\lambda_{\alpha j} \mathrm{E}[(r_\alpha(\mathbf{s}_i)$$

$$+ r_{\alpha\beta}(\mathbf{s}_i))(r_\alpha(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))] + 2 \sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \frac{\lambda_{\alpha i}\lambda_{\beta j}}{n_{\beta j}} \mathrm{E}[r_{\alpha\beta}(\mathbf{s}_i)]$$

$$+ 2 \sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \lambda_{\alpha i}\lambda_{\beta j} \mathrm{E}[(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))(r_\beta(\mathbf{s}_j) + r_{\alpha\beta}(\mathbf{s}_j))]$$

$$+ r_{\alpha\beta}(\mathbf{s}_i))(r_\beta(\mathbf{s}_j)$$

$$+ r_{\alpha\beta}(\mathbf{s}_j))] - 2 \sum_{i=1}^{k_\alpha} \lambda_{\alpha i} \mathrm{E}[(r_{\alpha 0} + r_{\alpha\beta 0})(r_\alpha(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))]$$

$$- 2 \sum_{i=1}^{k_\beta} \lambda_{\beta i} \mathrm{E}[(r_{\alpha 0} + r_{\alpha\beta 0})(r_\beta(\mathbf{s}_i) + r_{\alpha\beta}(\mathbf{s}_i))] + \mathrm{E}[(r_{\alpha 0} + r_{\alpha\beta 0})^2]$$

$$\times \left[C_{\alpha\alpha ij}^{\mathcal{R}} + (m_\alpha + m_{\alpha\beta})(m_\alpha + m_{\alpha\beta}) \right]$$

$$+ 2 \sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \frac{\lambda_{\alpha i}\lambda_{\beta j}}{n_{\beta j}} m_{\alpha\beta} + 2 \sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \lambda_{\alpha i}\lambda_{\beta j}$$

$$\times \left[C_{\alpha\beta ij}^{\mathcal{R}} + (m_\alpha + m_{\alpha\beta})(m_\beta + m_{\alpha\beta}) \right]$$

$$+ \sum_{i=1}^{k_\beta} \frac{\lambda_{\beta i}^2}{n_{\beta i}} (m_\beta + m_{\alpha\beta}) + \sum_{i=1}^{k_\beta} \sum_{j=1}^{k_\beta} \lambda_{\beta i} \lambda_{\beta j}$$

$$\times \left[C_{\beta\beta ij}^{\mathcal{R}} + (m_\beta + m_{\alpha\beta})(m_\beta + m_{\alpha\beta}) \right]$$

$$\mathrm{E}\left[(\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0))^2 \right] \quad - 2\sum_{i=1}^{k_\alpha} \lambda_{\alpha i} \left[C_{\alpha\alpha i0}^{\mathcal{R}} + (m_\alpha + m_{\alpha\beta})(m_\alpha + m_{\alpha\beta}) \right]$$

$$- 2\sum_{i=1}^{k_\beta} \lambda_{\beta i} \left[C_{\beta\alpha i0}^{\mathcal{R}} + (m_\alpha + m_{\alpha\beta})(m_\beta + r_{\alpha\beta}) \right] + (m_\alpha + m_{\alpha\beta})^2$$

$$+ (\sigma_\alpha^2 + \sigma_{\alpha\beta}^2)$$

$$= \sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}^2}{n_{\alpha i}} (m_\alpha + m_{\alpha\beta}) + \sum_{i=1}^{k_\alpha} \sum_{j=1}^{k_\alpha} \lambda_{\alpha i} \lambda_{\alpha j} C_{\alpha\alpha ij}^{\mathcal{R}} + (m_\alpha + m_{\alpha\beta})^2$$

$$+ \sum_{i=1}^{k_\beta} \frac{\lambda_{\beta i}^2}{n_{\beta i}} (m_\beta + m_{\alpha\beta})$$

$$+ \sum_{i=1}^{k_\beta} \sum_{j=1}^{k_\beta} \lambda_{\beta i} \lambda_{\beta j} C_{\beta\beta ij}^{\mathcal{R}} + 2\sum_{i=1}^{k_\alpha} \sum_{j=1}^{k_\beta} \frac{\lambda_{\alpha i} \lambda_{\beta j}}{n_{\beta j}} m_{\alpha\beta} + 2\sum_{i=1}^{k_\alpha} \sum_{j=1}^{k_\beta} \lambda_{\alpha i} \lambda_{\beta j} C_{\alpha\beta ij}^{\mathcal{R}}$$

$$- 2\sum_{i=1}^{k_\alpha} \lambda_{\alpha i} C_{\alpha\alpha i0}^{\mathcal{R}} - 2(m_\alpha + m_{\alpha\beta})^2 - 2\sum_{i=1}^{k_\beta} \lambda_{\beta i} C_{\beta\alpha i0}^{\mathcal{R}} + (m_\alpha + m_{\alpha\beta})^2$$

$$+ (\sigma_\alpha^2 \sigma_{\alpha\beta}^2)$$

$$= \sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}^2}{n_{\alpha i}} (m_\alpha + m_{\alpha\beta}) + \sum_{i=1}^{k_\alpha} \frac{\lambda_{\beta i}^2}{n_{\beta i}} (m_\beta + m_{\alpha\beta}) + 2\sum_{i=1}^{k_\alpha} \sum_{j=1}^{k_\beta} \frac{\lambda_{\alpha i} \lambda_{\beta j}}{n_{\beta j}} m_{\alpha\beta}$$

$$+ \sum_{i=1}^{k_\alpha} \sum_{j=1}^{k_\alpha} \lambda_{\alpha i} \lambda_{\alpha j} C_{\alpha\alpha ij}^{\mathcal{R}} + \sum_{i=1}^{k_\beta} \sum_{j=1}^{k_\beta} \lambda_{\beta i} \lambda_{\beta j} C_{\beta\beta ij}^{\mathcal{R}} + 2\sum_{i=1}^{k_\alpha} \sum_{j=1}^{k_\beta} \lambda_{\alpha i} \lambda_{\beta j} C_{\alpha\beta ij}^{\mathcal{R}}$$

$$- 2\sum_{i=1}^{k_\alpha} \lambda_{\alpha i} C_{\alpha\alpha i0}^{\mathcal{R}} - 2\sum_{i=1}^{k_\beta} \lambda_{\beta i} C_{\beta\alpha i0}^{\mathcal{R}} + (\sigma_\alpha^2 + \sigma_{\alpha\beta}^2)$$

The variance can be written in terms of expectations as

$$\mathrm{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)] = \mathrm{E}\left[(\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0))^2 \right] - \mathrm{E}^2\left[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0) \right]$$

$$= \sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}^2}{n_{\alpha i}} (m_\alpha + m_{\alpha\beta}) + \sum_{i=1}^{k_\alpha} \frac{\lambda_{\beta i}^2}{n_{\beta i}} (m_\beta + m_{\alpha\beta}) + 2\sum_{i=1}^{k_\alpha} \sum_{j=1}^{k_\beta} \frac{\lambda_{\alpha i} \lambda_{\beta j}}{n_{\beta j}} m_{\alpha\beta}$$

$$+ \sum_{i=1}^{k_\alpha} \sum_{j=1}^{k_\alpha} \lambda_{\alpha i} \lambda_{\alpha j} C_{\alpha\alpha ij}^{\mathcal{R}} + \sum_{i=1}^{k_\beta} \sum_{j=1}^{k_\beta} \lambda_{\beta i} \lambda_{\beta j} C_{\beta\beta ij}^{\mathcal{R}} + 2\sum_{i=1}^{k_\alpha} \sum_{j=1}^{k_\beta} \lambda_{\alpha i} \lambda_{\beta j} C_{\alpha\beta ij}^{\mathcal{R}}$$

$$- 2\sum_{i=1}^{k_\alpha} \lambda_{\alpha i} C_{\alpha\alpha i0}^{\mathcal{R}} - 2\sum_{i=1}^{k_\beta} \lambda_{\beta i} C_{\beta\alpha i0}^{\mathcal{R}} + (\sigma_\alpha^2 + \sigma_{\alpha\beta}^2)$$

where $C_{\alpha\beta ij}^{\mathcal{R}} = C_{\alpha\beta}^{\mathcal{R}}(|\mathbf{s}_i - \mathbf{s}_j|)$ and $C_{\alpha\beta i0}^{\mathcal{R}} = C_{\alpha\beta}(|s_0 - \mathbf{s}_i|)$. The variance of the process $\mathcal{R}_\alpha(\mathbf{s}_0)$ is $\sigma_\alpha^2 + \sigma_{\alpha\beta}^2$.

The final step is to set up a system of equations and minimise the prediction error variance. By minimising the above equation, the constraints on the weights generate two parameters of Lagrange μ_α and μ_β. Then

$$\text{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)] = \sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}^2}{n_{\alpha i}}(m_\alpha + m_{\alpha\beta}) + \sum_{i=1}^{k_\alpha} \frac{\lambda_{\beta i}^2}{n_{\beta i}}(m_\beta + m_{\alpha\beta}) + 2\sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \frac{\lambda_{\alpha i}\lambda_{\beta j}}{n_{\beta j}}m_{\alpha\beta}$$

$$+ \sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\alpha}\lambda_{\alpha i}\lambda_{\alpha j}C_{\alpha\alpha ij}^{\mathcal{R}} + \sum_{i=1}^{k_\beta}\sum_{j=1}^{k_\beta}\lambda_{\beta i}\lambda_{\beta j}C_{\beta\beta ij}^{\mathcal{R}} + 2\sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta}\lambda_{\alpha i}\lambda_{\beta j}C_{\alpha\beta ij}^{\mathcal{R}}$$

$$- 2\sum_{i=1}^{k_\alpha}\lambda_{\alpha i}C_{\alpha\alpha i0}^{\mathcal{R}} - 2\sum_{i=1}^{k_\beta}\lambda_{\beta i}C_{\beta\alpha i0}^{\mathcal{R}} + (\sigma_\alpha^2 + \sigma_{\alpha\beta}^2) + 2\mu_\alpha\left[\sum_{i=1}^{k_\alpha}\lambda_{\alpha i} - 1\right]$$

$$+ 2\mu_\beta\left[\sum_{i=1}^{k_\beta}\lambda_{\beta i}\right]$$

Note that the two additional terms are both equal to 0 and do not contribute to the error variance $\text{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)]$.

To minimise $\text{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)]$ we compute the partial derivatives of $\text{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)]$ with respect to the $k_\alpha + k_\beta$ weights and the two Lagrange multipliers:

$$\frac{\partial}{\partial \lambda_{\alpha j}} \text{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)] = 2\frac{\lambda_{\alpha i}}{n_{\alpha i}}(m_\alpha + m_{\alpha\beta}) + 2\sum_{i=1}^{k_\beta}\frac{\lambda_{\beta i}}{n_{\beta i}}m_{\alpha\beta} + 2\sum_{i=1}^{k_\alpha}\lambda_{\alpha i}C_{\alpha\alpha ij}^{\mathcal{R}}$$

$$+ 2\sum_{i=1}^{k_\beta}\lambda_{\beta i}C_{\alpha\beta ij}^{\mathcal{R}}$$

$$- 2C_{\alpha\alpha i0}^{\mathcal{R}} + 2\mu_\alpha$$

$$\frac{\partial}{\partial \lambda_{\beta j}} \text{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)] = 2\frac{\lambda_{\beta i}}{n_{\alpha i}}(m_\beta + m_{\alpha\beta}) + 2\sum_{i=1}^{k_\alpha}\frac{\lambda_{\alpha i}}{n_{\alpha i}}m_{\alpha\beta} + 2\sum_{i=1}^{k_\beta}\lambda_{\beta i}C_{\beta\beta ij}^{\mathcal{R}}$$

$$+ 2\sum_{i=1}^{k_\alpha}\lambda_{\alpha i}C_{\alpha\beta ij}^{\mathcal{R}}$$

$$- 2C_{\beta\alpha i0}^{\mathcal{R}} + 2\mu_\beta$$

$$\frac{\partial}{\partial \mu_\alpha} \text{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)] = 2\sum_{i=1}^{k_\alpha}\lambda_{\alpha i} - 1$$

$$\frac{\partial}{\partial \mu_\beta} \text{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)] = 2\sum_{i=1}^{k_\beta}\lambda_{\beta i}$$

The Poisson cokriging system in Eq. (25) is obtained by equating each of these $k_\alpha + k_\beta + 2$ equations to 0 and rearranging the individual terms.

The Poisson cokriging variance in Eq. (24) can be simplified by making substitutions using the Lagrange multipliers

$$
\begin{aligned}
\mathrm{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)] =& \sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}^2}{n_{\alpha i}}(m_\alpha + m_{\alpha\beta}) + \sum_{i=1}^{k_\beta} \frac{\lambda_{\beta i}^2}{n_{\beta i}}(m_\beta + m_{\alpha\beta}) + 2\sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \frac{\lambda_{\alpha i}\lambda_{\beta j}}{n_{\beta j}} m_{\alpha\beta} \\
& + \sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\alpha} \lambda_{\alpha i}\lambda_{\alpha j} C_{\alpha\alpha ij}^{\mathcal{R}} + \sum_{i=1}^{k_\beta}\sum_{j=1}^{k_\beta} \lambda_{\beta i}\lambda_{\beta j} C_{\beta\beta ij}^{\mathcal{R}} + 2\sum_{i=1}^{k_\alpha}\sum_{j=1}^{k_\beta} \lambda_{\alpha i}\lambda_{\beta j} C_{\alpha\beta ij}^{\mathcal{R}} \\
& - 2\sum_{i=1}^{k_\alpha} \lambda_{\alpha i} C_{\alpha\alpha i0}^{\mathcal{R}} - 2\sum_{i=1}^{k_\beta} \lambda_{\beta i} C_{\beta\beta i0}^{\mathcal{R}} + (\sigma_\alpha^2 + \sigma_{\alpha\beta}^2) \\
=& \sum_{i=1}^{k_\alpha} \lambda_{\alpha i} \underbrace{\left(\frac{\lambda_{\alpha i}}{n_{\alpha i}}(m_\alpha + m_{\alpha\beta}) + \sum_{i=1}^{k_\beta} \frac{\lambda_{\beta i}}{n_{\beta i}} m_{\alpha\beta} + \sum_{i=1}^{k_\alpha} \lambda_{\alpha i} C_{\alpha\alpha ij}^{\mathcal{R}} + \sum_{i=1}^{k_\beta} \lambda_{\beta i} C_{\alpha\beta ij}^{\mathcal{R}} - C_{\alpha\alpha i0}^{\mathcal{R}} \right)}_{-\mu_\alpha} \\
& + \sum_{i=1}^{k_\beta} \lambda_{\beta i} \underbrace{\left(\frac{\lambda_{\beta i}}{n_{\beta i}}(m_\beta + m_{\alpha\beta}) + \sum_{i=1}^{k_\alpha} \frac{\lambda_{\alpha i}}{n_{\alpha i}} m_{\alpha\beta} + \sum_{i=1}^{k_\beta} \lambda_{\beta i} C_{\beta\beta ij}^{\mathcal{R}} + \sum_{i=1}^{k_\alpha} \lambda_{\alpha i} C_{\alpha\beta ij}^{\mathcal{R}} - C_{\beta\alpha i0}^{\mathcal{R}} \right)}_{-\mu_\beta} \\
& - \sum_{i=1}^{k_\alpha} \lambda_{\alpha i} C_{\alpha\alpha i0}^{\mathcal{R}} - \sum_{i=1}^{k_\beta} \lambda_{\beta i} C_{\beta\alpha i0}^{\mathcal{R}} + (\sigma_\alpha^2 + \sigma_{\alpha\beta}^2) \\
=& (\sigma_\alpha^2 + \sigma_{\alpha\beta}^2) - \sum_{i=1}^{k_\alpha} \lambda_{\alpha i} C_{\alpha\alpha i0}^{\mathcal{R}} - \sum_{i=1}^{k_\beta} \lambda_{\beta i} C_{\beta\alpha i0}^{\mathcal{R}} - \mu_\alpha \underbrace{\sum_{i=1}^{k_\alpha} \lambda_{\alpha i}}_{1} - \mu_\beta \underbrace{\sum_{i=1}^{k_\beta} \lambda_{\beta i}}_{0}
\end{aligned}
$$

$$
\mathrm{Var}[\mathcal{R}_\alpha(\mathbf{s}_0)^* - \mathcal{R}_\alpha(\mathbf{s}_0)] = (\sigma_\alpha^2 + \sigma_{\alpha\beta}^2) - \sum_{i=1}^{k_\alpha} \lambda_{\alpha i} C_{\alpha\alpha i0}^{\mathcal{R}} - \sum_{i=1}^{k_\beta} \lambda_{\beta i} C_{\beta\alpha i0}^{\mathcal{R}} - \mu_\alpha
$$

References

Bellier, E., Monestiez, P., Certain, G., Chadœuf, J., Bretagnolle, V., 2013. Reducing the uncertainty of wildlife population abundance: Model-based versus design-based estimates. Environmetrics 24.

Bellier, E., Monestiez, P., Guinet, C., 2010. Geostatistical modelling of wildlife populations: A non-stationary hierarchical model for count data. pp. 1–12.

Berke, O., 2004. Exploratory disease mapping: Kriging the spatial risk function from regional count data. Int. J. Health Geogr. 3.

Borchering, R.K., Huang, A.T., y Teran-Romero, L.M., Rojas, D.P., Rodriguez-Barraquer, I., Katzelnick, L.C., Martinez, S.D., King, G.D., Cinkovich, S.C., Lessler, J., Cummings, D.A., 2019. Impacts of Zika emergence in Latin America on endemic Dengue transmission. Nature Commun. 10.

Cortez, M.H., Weitz, J.S., 2014. Coevolution can reverse predator-prey cycles. Proc. Natl. Acad. Sci. USA 111.

Cressie, N., 1993. Statistics for Spatial Data. In: Wiley Series in Probability and Statistics, Wiley.

Goovaerts, P., 1998. Ordinary cokriging revisited. Math. Geol. 30.

Goovaerts, P., 2005. Geostatistical analysis of disease data: Estimation of Cancer mortality risk from empirical frequencies using Poisson kriging. Int. J. Health Geogr. 4.

Goovaerts, P., 2006a. Geostatistical analysis of disease data: Accounting for spatial support and population density in the isopleth mapping of Cancer mortality risk using area-to-point Poisson kriging. Int. J. Health Geogr. 5.

Goovaerts, P., 2006b. Geostatistical analysis of disease data: Accounting for spatial support and population density in the isopleth mapping of Cancer mortality risk using area-to-point Poisson kriging. Int. J. Health Geogr. 5.

Goovaerts, P., 2017. Geostatistical Interpolation of Rate Data Using Poisson Kriging. Springer International Publishing, pp. 767–774.

Goovaerts, P., Gebreab, S., 2008. How does Poisson kriging compare to the popular BYM model for mapping disease risks? Int. J. Health Geogr. 7.

Goovaerts, P., Jacquez, G.M., 2004. Accounting for regional background and population size in the detection of spatial clusters and outliers using geostatistical filtering and spatial neutral models: The case of lung Cancer in Long Island, New York. Int. J. Health Geogr. 3.

Goulard, M., Voltz, M., 1992. Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. Math. Geol. 24.

Holford, T.R., 2002. Multivariate Methods in Epidemiology. Oxford University Press.

Holgate, P., 1966. Bivariate generalizations of Neyman's type A distribution. Biometrika 53.

Karlis, D., Meligkotsidou, L., 2007. Finite mixtures of multivariate Poisson distributions with application. J. Statist. Plann. Inference 137.

Kawamura, K., 1973. The structure of bivariate Poisson distribution. In: Kodai Mathematical Seminar Reports, Vol. 25. pp. 246–256.

Kawamura, K., 1979. The structure of multivariate Poisson distribution. Kodai Math. J. 2.

Krivoruchko, K., Gribov, A., Krause, E., 2011. Multivariate areal interpolation for continuous and count data. Procedia Environ. Sci. 3, 14–19.

Li, J.M., Liu, C., Hu, X., Cai, Y., Ma, C., Luo, X.G., Yan, X.X., 2014. Inverse correlation between Alzheimer's disease and cancer: implication for a strong impact of regenerative propensity on neurodegeneration? BMC Neurol. 14.

Magalhaes, T., Robison, A., Young, M.C., Black, W.C., Foy, B.D., Ebel, G.D., Rückert, C., 2018. Sequential infection of Aedes aegypti mosquitoes with Chikungunya virus and Zika virus enhances early Zika virus transmission. Insects 9.

Mahamunulu, D.M., 1967. A note on regression in the multivariate Poisson distribution. J. Amer. Statist. Assoc. 62.

Masarotto, G., Varin, C., 2012. Gaussian copula marginal regression. Electron. J. Stat. 6, 1517–1549. http://dx.doi.org/10.1214/12-EJS721.

Matheron, G., 1965. Les Variables Regionalisees Et Leur Estimation. Masson.

Mercado-Reyes, M., Acosta-Reyes, J., Navarro-Lechuga, E., Corchuelo, S., Rico, A., Parra, E., Tolosa, N., Pardo, L., González, M., Martìn-Rodriguez-Hernández, J., Karime-Osorio, L., Ospina-Martinez, M., Rodriguez-Perea, H., Rio-Pertuz, G.D., Viasus, D., 2019a. Dengue, Chikungunya and Zika virus coinfection: Results of the national surveillance during the Zika epidemic in colombia. Epidemiol. Infect. 147.

Mercado-Reyes, M., Acosta-Reyes, J., Navarro-Lechuga, E., Corchuelo, S., Rico, A., Parra, E., Tolosa, N., Pardo, L., González, M., Martìn-Rodriguez-Hernández, J., Karime-Osorio, L., Ospina-Martinez, M., Rodriguez-Perea, H., Rio-Pertuz, G.D., Viasus, D., 2019b. Dengue, Chikungunya and Zika virus coinfection: Results of the national surveillance during the Zika epidemic in colombia. Epidemiol. Infect. 147.

Monestiez, P., Dubroca, L., Bonnin, E., Durbec, J.P., Guinet, C., 2006. Geostatistical modelling of spatial distribution of Balaenoptera physalus in the Northwestern Mediterranean Sea from sparse count data and heterogeneous observation efforts. Ecol. Model. 193.

Morales-Navarrete, D., Bevilacqua, M., Caamaño-Carrillo, C., Castro, L.M., 2022. Modeling point referenced spatial count data: A Poisson process approach. J. Amer. Statist. Assoc. 1–14.

Myers, D.E., 1982. Matrix formulation of co-kriging. J. Int. Assoc. Math. Geol. 14, 249–257.

Olea, R.A., 1999. Block kriging. In: Geostatistics for Engineers and Earth Scientists. Springer.

Oliveira, V.D., 2014. Poisson kriging: A closer investigation. In: Spatial Statistics, Vol. 7.

Oliver, M.A., Webster, R., Lajaunie, C., Muir, K.R., Parkes, S.E., Cameron, A.H., Stevens, M.C., Mann, J.R., 1998. Binomial cokriging for estimating and mapping the risk of childhood Cancer. IMA J. Math. Appl. Med. Biol. 15.

Osei, F.B., Stein, A., 2017. Spatio-temporal analysis of small-area intestinal parasites infections in Ghana. Sci. Rep. 7.

Parselia, E., Kontoes, C., Tsouni, A., Hadjichristodoulou, C., Kioutsioukis, I., Magiorkinis, G., Stilianakis, N.I., 2019. Satellite Earth Observation data in epidemiological modeling of Malaria, dengue and West Nile Virus: A scoping review. Remote Sens. 11.

Riou, J., Poletto, C., Boëlle, P.Y., 2017. A comparative analysis of Chikungunya and Zika transmission. Epidemics 19.

Smith, L.M., Stroup, W.W., Marx, D.B., 2020. Poisson cokriging as a generalized linear mixed model. In: Spatial Statistics, Vol. 35.

Stein, A., Corsten, L.C.A., 1991. Universal kriging and cokriging as a regression procedure. Biometrics 47.

Wackernagel, H., 2003a. Multivariate Geostatistics: An Introduction with Applications. Springer Berlin Heidelberg.

Wackernagel, H., 2003b. Direct and cross covariances. In: Multivariate Geostatistics: An Introduction with Applications. Springer Berlin Heidelberg, pp. 145–150.

Wichit, S., Gumpangseth, N., Hamel, R., Yainoy, S., Arikit, S., Punsawad, C., Missé, D., 2021. Chikungunya and zika viruses: Co-circulation and the interplay between viral proteins and host factors. Pathogens 10 (4).