Spatial Modeling of Infectious Diseases: Environment and Health

Mahmoud Torabi, University of Manitoba, Lead Investigator

Abstract

Identifying disease clusters and spatial patterns of disease (from human/animal/plant) are important to inform policy makers, programs and interventions at both local and global scales. For instance, in the context of population health, Canadian health authorities depend on alerts provided by front-line clinicians or by members of the public when there is an increase in disease or illness (disease cluster). Health authorities need to respond to cluster inquiries to inform the public that: a) no clustering exists; or b) to warn the public and investigate the cause of the cluster. The recent emergence of the Zika virus as a global pandemic is one example of a critical public health threat that challenged management systems. The rapid spread of Zika across much of the Americas is not well understood. Space-time patterns of spread span multiple scales due to complex disease ecological processes and biases from surveillance data generated from multi-jurisdictions with varying sampling protocols are real challenges. These issues, which are also common to high priority diseases in Canada (e.g., Lyme disease), can be difficult to accommodate in quantitative frameworks, and hamper the ability to use data and modeling products to accurately monitor disease and identify vulnerable populations either spatially or over space-time. We will spearhead innovation in disease modeling by addressing several practical problems related to infectious diseases in environment and health by advancing statistical modeling techniques. *Our research goal is to better integrate population and environmental data for infectious diseases using spatial modeling techniques.* In particular, our study aims are to: 1) advance a spatial area-level statistical model (ALM) to address measurement errors in covariates that are related to an infectious disease outcome; 2) develop an area-level spatial model to relax the assumption of having the same distribution for all the areas of the population study by introducing a mixture model approach for an infectious disease outcome; 3) simultaneously study multiple infectious disease outcomes (e.g., influenza and meningococcal disease) by introducing multivariate area-level spatial models; 4) extend individual-level statistical models (ILMs) to a new class of geo-dependent ILMs to also account for the spatial location of the individuals in addition to the distance between susceptible and infectious individuals; and 5) develop a joint spatial survival model for modeling successive times to multiple events through the stages (susceptible, infectious, removed) of an infectious disease. Using our proposed models, which offer a better reflection of the true infectious disease dynamics and imperfect data, researchers and authorities will have improved models for understanding disease ecology and advising population health management to ultimately improve the health of Canadians.

6 Research aims

6.1 Background and recent progress: Monitoring, analyzing, and predicting the impact of infectious diseases on the environment and health of a society are cornerstones for identifying effective ways to prevent, control, and manage disease spread. A basic assumption is that infectious diseases are transmitted through space and time when susceptible individuals come into contact with pathological agents from direct contact with infected individuals or indirectly through environmental contamination by infected individuals. In theory, every transmission event can be traced along a specific spread network, and these networks are often influenced by behavioral and environmental factors. Furthermore, our ability to detect patterns in spread dynamics can depend on space-time scales of observation and measurement error. It is then essential to use appropriate models to study the spread of disease, and to understand the impacts of interventions to control and prevent further outbreaks, and how such interventions limit the devastating effects on a population.

A variety of analyses and modeling techniques have been developed assuming that there is a fundamental spatial structure of disease spread based on which the human and physical geographical world is formed. The classical method of statistical modeling considered a study population to be divided into distinct units, and each individual interacts with other individuals based on its immediate neighborhood. The simplest of these population-based models is the SIR (susceptible, infectious, removed) model, initially described by Kermack and McKendrick (1927) for a closed population in the context of a human transmitted disease. However, possible spatial-temporal spread and resulting implications of a disease outbreak in communities with unique behaviors and risk factors often affect disease dynamics and the types of public health interventions that would be effective. Traditional statistical models that represent the dynamics of infectious diseases usually use a non-spatial and population-based framework to view disease spread assuming homogeneity in disease transmission. Although these models are useful in estimating the size of the affected population, they do not explicitly address the casual factors in the development of epidemics.

The area-level statistical models (ALMs) have been studied to model the spatial and temporal patterns of (infectious) disease in area level analyses. In geographical epidemiology, diseases measured at the same spatial location may be also correlated so that the spatial structures of such diseases across the area under considerations are very similar, and may be influenced by the same set of spatially unobserved distribution or unmeasured risk factors where the presence of one disease might lead to the presence of another disease over an area. For example, there are studies assessing trends in tuberculosis (TB) and human immunodeficiency virus (HIV) comorbidity, which identify that HIV infection contributes to the progression from a recently acquired or latent TB infection to the active form of the disease (Lawn et al., 2009).

The individual-level statistical models (ILMs) have been developed to model the transmission between disease states (e.g., SIR) on the level of the individual instead of population groups at aggregated units (Longini et al., 2005; Deardon et al., 2010; Pokharel and Deardon, 2016). The general form of ILMs considers the transmission from a susceptible state to an infected state based on a time-dependent Poisson process for an individual, and then adds spatial structure of potential risk factors to explain the underlying disease dynamics. The key feature of the ILMs is that they can take into account covariate information on susceptible and infectious individuals (e.g., age, genetics, lifestyle factors) as well as shared covariate information such as geography or contact measures (e.g., sexual partnerships for a human sexually transmitted disease or shared household or workplaces). Reviews of infectious disease models are given by Daley and Gani (1999), Anderson and Britton (2000), Lawson (2006), Chen (2015), Smith et al. (2015) and references therein.

6.2 Program of research aims: Our research aims include the development of area-level and individual-level statistical models that incorporate heterogeneous populations and spatial spread of infectious diseases.

6.2.1 Program of research aims in the area-level statistical models (ALMs): To address both the incorporation of heterogeneous populations and the modeling of spatial spread of an infectious disease, we use the idea of temporal models for infectious disease which are often built around infective behavior and related changes to the susceptible population. These models offer fundamental frameworks for developing spatial models in the context of infectious diseases. Here, the disease outcome, case event or count, becomes the infective status or count of infectives, respectively. In this case, we need to consider at any given time, the number of susceptible, the number of infectives, and the number removed from the population, corresponding to SIR models. Assuming T discrete time periods and m number of small areas within a study region, we can define the *true* count of new infectives within time period $t\ (=1,...,T)$ and small area $i\ (=1,...,m)$ as I_{it} and the reported count as y_{it} by assuming that $y_{it} \sim Bin(I_{it}, p_{it})$, which allows for under-ascertainment via the reporting rate p_{it}. We also define S_{it} as the susceptible population at small area i and time t, and R_{it} as the count of removed cases during the time period t at small area i. Hence, the transmission, susceptible, and removal equations would take the form: $y_{it} \sim Bin(I_{it}, p_{it})$, $I_{it} \sim Pois(S_{it}\lambda_{it})$, $S_{i,t+1} \sim N(\mu_{i,t+1}, \sigma_s^2)$, $\mu_{i,t+1} = S_{it} - I_{it} - R_{it}$, $R_{it} \sim N(\delta I_{it}, \sigma_r^2)$, where λ_{it} takes different forms which depend on the structure of the data. For instance, if we believe that neighborhoods play a role in transmission then we could assume a spatial structure through λ_{it} as

$$\lambda_{it} = I_{i,t-1} \exp(\mathbf{x}_{it}'\boldsymbol{\beta} + u_i + v_t), \tag{1}$$

where $\mathbf{x}_{it}'$ is the vector of covariates (risk factors) with the corresponding regression coefficients $\boldsymbol{\beta}$; u_i is a spatial random effect which may be modeled as for example a proper conditional auto-regressive (CAR) model (Stern and Cressie, 1999; Torabi and Rosychuk, 2010) with parameters (τ, σ_u^2); and v_t is the temporal random effect which may be modeled as for example a random walk model with parameter σ_v^2 (Torabi, 2012). One can also define alternative forms for modeling λ_{it} and in particular for spatial and temporal random effects. Note that in the models above, it is assumed that the evolution of susceptibles and removed will have some associated noise and so normal distributions could be assigned to these components, although consideration of their discrete nature may also suggest a Poisson model.

Objective 1: In model (1), it is assumed that the covariates are error-free, but this may not be a valid assumption in many applications. For instance, it is well-known that data products derived from satellite imagery, which are increasingly used as covariates in infectious disease models, are subject to measurement error (Kotchi et al., 2016). Another example is to measure usual nutrient intake, which is an unobserved covariate (we observe a surrogate covariate instead), as a risk factor for TB (Cegielski et al., 2012). Our plan is to assess model (1) incorporating error-free covariates and covariates with measurement error. In particular, let us assume that $\mathbf{x}_{it}'\boldsymbol{\beta} = \mathbf{x}_{1it}'\boldsymbol{\beta}_1 + \mathbf{x}_{2it}'\boldsymbol{\beta}_2$ where $\mathbf{x}_{1it}$ are observed covariates but not $\mathbf{x}_{2it}$. However, we may observe $\mathbf{w}_{it}$ as surrogate to $\mathbf{x}_{2it}$ which is $\mathbf{w}_{it} = \mathbf{x}_{2it} + \boldsymbol{\eta}_{it}$, called classical measurement error, where $\boldsymbol{\eta}_{it}$ are measurement errors and possibly normally distributed with appropriate means and variance-covariance matrix. Depending on the nature of $\mathbf{x}_{2it}$, we will work on functional (assuming $\mathbf{x}_{2it}$ are fixed but unknown) or structural (assuming $\mathbf{x}_{2it}$ are random variables) measurement error models (Carroll et al., 2006; Torabi, 2013). Depending on the nature of study and covariates with measurement errors, other scenarios will also be investigated such as Berkson measurement error, multiplicative rather than additive measurement error, different distribution assumptions on the random variables rather than normal.

Note that the above well-established measurement error models have been extensively used in non-spatial models. Recently, there have been some work on spatial models in the context of non-infectious diseases with incorporating measurement error models (e.g., Huque et al., 2016); noting that the structure

of spatial models with infectious diseases is very different than the spatial models with non-infectious diseases. The novelty of this objective is to rigorously study the impact of the covariates with measurement error in the context of spatial models with infectious diseases. With the help of our Manitoba partner (see section 2), we will undertake to test the proposed measurement error model on the TB dataset in the province of Manitoba. This objective will be completed by a post-doc through collaboration with health scientists and our partners in Manitoba and Public Health Agency of Canada (PHAC). We will also explore other applications/datasets of this proposed approach using other rich datasets from our partners.

Objective 2: In model (1), the same distribution for the spatial random effects is assumed through the entire population density. However, in many circumstances, this is a strong assumption as some areas may have different levels of risk. For example, transmission risk of TB has been associated with poverty levels, although the direction and intensity of this association varies regionally (Olson et al., 2012). To address this issue, we propose a hierarchical mixture generalized linear model (Torabi, 2016) to analyze spatial outcome with flexible distribution for the spatial random effects. The basic idea is that the population under scrutiny may be decomposed into subareas (L) with different levels of risk. The parameters describing these levels of risk stem from a discrete parameter distribution $(\pi_1, ..., \pi_L)$. The nonparametric mixture model assumes that these area-specific random effects have a discrete probability distribution taking L values $\theta_{it1}, ..., \theta_{itL}$ with probabilities $\pi_{it1}, ..., \pi_{itL}$, respectively. Note that $\theta_{itl} = ln(\lambda_{itl})$, and λ_{it} in (1) can be expanded for the L subareas to obtain $\lambda_{itl}, (l = 1, ..., L)$. Each of these L components of the mixture represents a cluster containing a proportion π_{itl} (assigning a binary variable τ_{itl} to each cluster l with corresponding probability π_{itl}) from the population with θ_{itl} and with the constraint that $\sum_{l=1}^{L} \pi_{itl} = 1$.

With the help of our partners in Alberta (see section 2), we will undertake to test the proposed mixture model on the TB dataset in the province of Alberta. Other applications of the proposed approach are to model mosquito borne diseases like Zika, Dengue, and Chikungunya for estimating risk of transmission from these diseases to advise Canadian travelers to infected countries through the PHAC. To the best of our knowledge, the mixture models have not been explored in the context of ALMs, although in a closely related field, Ferguson et al. (2001) used a mathematical model as a mixture model to study foot-and-mouth disease (FMD) outbreak in the UK in 2001. However, they did not specifically study the mixture model rigorously and only briefly mentioned that they used the idea of mixture model for the FMD. The novelty of this objective is to rigorously study the mixture models, as detailed above, in the context of ALMs and also evaluate its impacts in different real data applications including the TB in the province of Alberta. This objective will be completed by a PhD student through collaboration with health scientists and our partners in Alberta and PHAC.

Objective 3: There are many applications in infectious and/or chronic diseases with multiple health outcomes (e.g., influenza and meningococcal disease; diabetes and influenza) (Badawi and Ryoo, 2016). Simultaneous modeling of different diseases can be a valuable tool both from the epidemiological and from the statistical point of view. In particular, when we have several measurements recorded at each spatial location, we need to consider multivariate models in order to handle the dependence among the multivariate components as well as the spatial dependence between locations. Note that the multivariate models for multiple health outcomes have been extensively studied in the context of spatial models with non-infectious diseases but not with infectious diseases as modeling for the latter is more complicated and complex. For this objective, we propose to extend the ALM (1) to study multiple disease outcomes jointly. In particular, we will use multivariate CAR, generalized version of the CAR, models for hierarchical modeling of multiple disease outcomes (Jin et al., 2005; Torabi, 2014). Alternatively, we will use a shared component spatial model to study multiple diseases that are spatially correlated (Feng and Dean, 2012).

In particular, we can extend (1) to $\lambda_{itj} = I_{ij,t-1} \exp(\mathbf{x}'_{itj}\boldsymbol{\beta}_j + u_{ij} + v_t)$ where $I_{ij,t-1}$ is the true count of

infectives within time period $t-1$ and small area i for the outcome (e.g., disease) $j(=1,...,J)$; $\mathbf{x}'_{itj}$ is the vector of covariates at small area i and time period t for the outcome j, with the corresponding regression coefficients $\boldsymbol{\beta}_j$; v_t is the same temporal random effect; and u_{ij} are spatial random effects of area i for outcome j, which can be defined as shared spatial component or multivariate CAR models. In particular, we can define $u_{ij} = \gamma_j u_i$, where γ_j is the factor loading for the shared spatial component on outcome j, with $\gamma_1 = 1$, and u_i is the same spatial random effect assumed to follow a proper CAR model. Another way to define spatial random effects is through multivariate CAR models. Let $\boldsymbol{u}_i = (u_{i1},...,u_{iJ})', (i = 1,...,m)$, be a $J-$dimensional vector with, for example, u_{i1} as a spatially random variable of the first outcome (e.g., disease) at the ith small area. Similar to the univariate CAR model, the unique joint distribution $\boldsymbol{u} = (\boldsymbol{u}_1,...,\boldsymbol{u}_m)'$ can be then constructed. We will investigate the performance of these two approaches in the multiple disease outcomes. With the help of our partner PHAC (see section 2), we will undertake to test the proposed multivariate model on the influenza and meningococcal datasets in the province of Ontario. This objective will be completed by a post-doc through collaboration with health scientists and our partner in PHAC. Other applications of the proposed approach are for co-circulating diseases of interest to PHAC such as Zika, Dengue, and Chikungunya which share the same mosquito vector.

6.2.2 Program of research aims in the individual-level statistical models (ILMs): When individual-level data are available, different types of models may be also applied. We initially formulate this in the context of the progression of an epidemic, for a disease where there is removal, no re-infection, and no latent period before an exposed individual becomes infectious, (i.e., a SIR framework). We may employ similar concepts as for a survival experiment. Our viewpoint is then a finite population out of which individuals contract the disease and are removed. This situation holds when there is a large pool of susceptibles, and removal rate is relatively low, so that the transmission process is the main concern. In this research proposal, we assume that a realization of n disease events occurs within a fixed spatial and temporal window. In particular, the disease events are cases of infection as $\{(z_i, t_i), i = 1,...,n\}$, which represent the locations and infection times of all the cases. We then define $I(t^*)$ as the set of infectives at t^* for the window $\{(z_{I_j}, t_{I_j}); j = 1,...,n_{t^*}\}$. To form the corresponding susceptible population, we consider a three-dimensional field representation $S(z,t)$, which represents the degree of local susceptibility in the population at (z,t). This specification of $S(z,t)$ can be viewed as a general approach to cover discrete susceptible locations (e.g., houses) where a series of spikes will occur or, perhaps, more continuous background fields (e.g., urban area effects). We assume that the first-order intensity of cases can capture the model structure adequately, and hence the incidence of cases, conditional on the current $I(\cdot)$, can be modeled through a heterogeneous Poisson process with first-order intensity as:

$$\zeta(z,t) = \rho S(z,t) \sum_{j=1}^{n(t)} h(z - z_{I_j}) g(t - t_{I_j}), \tag{2}$$

where $S(z,t)$ is the local density of susceptibles at (z,t); ρ is the overall density (space-time units); $h(\cdot)$ is a spatial cluster function, which relates the location of a susceptible to any current infective location; $g(\cdot)$ is a cluster function which relates the temporal position t to the time of infectivity of the known infectives t_{I_j}; and $n(t)$ is the current number of infectives (just before time t). This form of intensity definition considers the epidemic to be described by a space-time interaction term $h(\cdot)g(\cdot)$; however, it is also possible to define the intensity specification to include separate spatial and temporal components which purely specify spatial or temporal effects. This kind of modeling will also allow us to study spatial transmission between selected social groups.

Objective 4: In the standard ILM (2), it is assumed that the spatial function $h(\cdot)$ depends only upon the

distance between susceptible and infectious individuals, but not the spatial location of the individuals (e.g., Zinszer et al., 2017). For example, often a simple power law spatial function is used (e.g., Deardon et al., 2010) of the form $h(z - z_{I_j}) = (z - z_{I_j})^{-\delta}, (\delta > 0)$. However, spatial locations are often important for understanding the spread of emerging infectious diseases and identifying their causes, with spatially varying demographic and environmental factors could influence the disease transmission (e.g., Rees et al., 2015). For instance, the transmission of vector-borne infectious diseases are predominantly controlled by environmental variables such as temperature, air quality, rainfall, and humidity. Our plan is to generalize the ILMs to a new class of geo-dependent ILMs (GD-ILMs) in which $h(\cdot)$ depends upon spatial (e.g., areal) covariates and unobserved spatial structure of emerging infectious diseases. In particular, the GD-ILMs will incorporate spatial effects salient to infectious disease spread for formulating etiological hypotheses and identifying areas of unusually high risk to formulate preventive action.

For instance, consider a set of areas/neighborhoods indexed by i. Then, a generalized spatial function form can be given by $h(z - z_{I_j}) = \exp(\alpha + \mathbf{x}'_{ij}\boldsymbol{\beta}_1 + \mathbf{x}'_i\boldsymbol{\beta}_2 + u_i)(z - z_{I_j})^{-\delta}$ where α is the intercept, $\mathbf{x}_{ij}$ are covariates (e.g., lifestyle factors) associated with infectious individual j in area i with associated parameters $\boldsymbol{\beta}_1$, $\mathbf{x}_i$ are covariates (e.g., environmental) associated with area i with associated parameters $\boldsymbol{\beta}_2$, and u_i are the area level spatial random effects. These spatial random effects could be modelled with a CAR model. Alternatively, the spatial random effects can be included within the power law function, giving a generalized form of $h(z-z_{I_j}) = \exp(\alpha+\mathbf{x}'_{ij}\boldsymbol{\beta}_1+\mathbf{x}'_i\boldsymbol{\beta}_2)(z-z_{I_j})^{-\delta+u_i}$. Typically, the temporal function $g(\cdot)$ is assumed to be exponential or gamma. However, since times of infection and removal (e.g., recovery or death) are not normally observed, inference is usually carried out using Bayesian data augmented Markov chain Monte Carlo (MCMC) treating the infection and removal times as latent variables to be inferred and integrated out when considering the marginal posterior distributions of the model parameters. This, along with the complexity of the likelihood, tends to lead to computationally taxing analyses. Here, this problem will be exacerbated by the need to estimate spatial random effects. Thus, a substantial part of this project may involve the development of computationally efficient methods (e.g., approximate Bayesian computation or Gaussian process emulation) to carry out inference for large populations.

Also, the modelling framework will be extended to allow for more complicated disease life histories than that of the SIR framework. For example, an SEIR framework will be considered in which individuals enter a latent period before becoming infectious after *exposure* to the disease. The proposed approaches will be developed using simulated data, and then applied to data on H1N1 influenza data, and vector-borne infectious diseases such as Zika virus in the province of Alberta with the help of our partner at the PHAC. This objective will be completed by a PhD student through collaboration with health scientists and our partners in Alberta and PHAC. Other applications of the proposed approach will be also explored using other rich datasets from our partners.

Objective 5: Infectious diseases typically progress over time undergoing multiple stages. For example, in AIDS studies, the outcomes of interests are often hazard of patients being diagnosed with HIV progressing to AIDS, and the hazard of patients with AIDS progressing to death due to AIDS. To address this very important health issue which has impacts on the individuals, health care providers and policy makers for possible intervention, we propose a Bayesian framework to jointly model (Feng and Dean, 2015) the multiple successive events times to reflect the infectious disease dynamic.

For instance, in the above-mentioned motivating example, there are three successive states. Let T_{1ij} denote the time of HIV diagnosis, T_{2ij} denote the survival time from HIV diagnosis to AIDS onset, and T_{3ij} denote the survival time from AIDS onset to death for the jth patient from the ith geographical area. The joint modeling of T_{1ij}, T_{2ij}, and T_{3ij} is realized as the product of the conditional distribution of T_{3ij} given T_{2ij}, T_{2ij} given T_{1ij}, and the marginal distribution of T_{1ij} as:

$$f_{T_{1ij},T_{2ij},T_{3ij}}(t_{1ij}, t_{2ij}, t_{3ij}) = f_{T_{1ij}}(t_{1ij}).f_{T_{2ij}|T_{1ij}}(t_{2ij}|t_{1ij}).f_{T_{3ij}|T_{2ij}}(t_{3ij}|t_{2ij})$$

We also assume a Weibull distributional form for $f_{T_{1ij}}$, $f_{T_{2ij}|T_{1ij}}$, and $f_{T_{3ij}|T_{2ij}}$ through $f_{T_{1ij}} \sim Weibull(\omega_1, \zeta_{1ij})$, $f_{T_{2ij}|T_{1ij}} \sim Weibull(\omega_2, \zeta_{2ij}|t_{1ij})$, and $f_{T_{3ij}|T_{2ij}} \sim Weibull(\omega_3, \zeta_{3ij}|t_{2ij})$. The covariates effects are also modeled through the scale parameters as: $ln(\zeta_{kij}) = \alpha_k + \mathbf{x}'_{kij}\boldsymbol{\beta}_k + u_{ki}, (k = 1, 2, 3)$, where u_{ki} can be defined using a shared component spatial model or multivariate CAR model as explained in objective 3. With the help of our Saskatchewan partner (see section 2), we will undertake to test the proposed model on the HIV dataset in the province of Saskatchewan. This objective will be completed by a PhD student through collaboration with health scientists and our partners in Saskatchewan and PHAC. Other applications of this proposed approach will be also explored using other rich datasets from our partners.

In the all five objectives above, Bayesian and likelihood-based methods (e.g., data cloning: Lele et al., 2010; Torabi, Lele, and Prasad, 2015) will be used for the inference. As the Bayesian and likelihood methods have different paradigms, we will evaluate the impact of different inferences on results as well as feasibility of the approach to be used for specific models and data applications. Proposed models will be also evaluated through simulation studies and also by real datasets in the context of infectious diseases. In particular, our partners (see section 3) will provide health-related applications of our proposed objectives. We will provide the R software package for each proposed method/model for a broad spectrum of researchers and end-users to implement our proposed objectives.

6.3 Anticipated outcomes and impact: Complexities of infectious disease ecology and quality of data used to characterize disease processes pose important challenges to authorities who need to monitor and respond to emerging and existing diseases in humans, animals, and plants. Statistical methods can have crucial roles for investigating disease-host systems and informing appropriate population management strategies. By addressing the objectives proposed in this program of research, our team will provide new statistical techniques that solve prevalent problems stated above in the analysis of infectious disease data. An immediate outcome of this research program is helping our partner organizations across Canada to implement novel modeling products that will improve current technologies that are used to inform population health. Our proposed models will better reflect the true infectious disease dynamics and account for data imperfections, and therefore, researchers, end-users at various agencies, and decision-makers will have better tools for drawing the appropriate conclusions of disease ecology and devising effective disease management strategies to ultimately improve the health of Canadians.

6.4 References

Anderson H., and Britton T. (2000). *Stochastic Epidemic Models and their Statistical Analysis*, New York: Springer.

Badawi A., and Ryoo S. G. (2016). Prevalence of diabetes in the 2009 influenza A (H1N1) and the Middle East respiratory syndrome coronavirus?: a systematic review and meta-analysis, *Journal of Public Health Research*, 5:733.

Carroll R.J., Ruppert D., Stefanski L.A., and Crainiceanu C.M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, Chapman & Hall /CRC press, 2006.

Chen D. (2015). Modeling the spread of infectious diseases: a review. In *Analyzing and Modelling Spatial and Temporal Dynamics of Infectious Diseases*, edited by D. Chen, B. Moulin, J. Wu,19–42,Wiley.

Cegielski J.P., Arab L., and Cornoni-Huntly J. (2012). Nutritional risk factors for tuberculosis among adults in the United States, 1971–1992, *American Journal of Epidemiology*, 176, 409–422.

Daley D., and Gani J. (1999). *Epidemic Modeling: An Introduction*, New York: Cambridge University.

Deardon R., et al. (2010). Inference for individual-level models of infectious diseases in large populations, *Statistica Sinica*, 20, 239–261.

Feng C. X., and Dean C. B. (2012). Joint analysis of multivariate spatial count and zero-heavy count

outcomes using common spatial factor models, *Environmetrics*, 23, 493–508.

Feng C.X., and Dean C.B. (2015). Spatial pattern analysis of multivariate disease data. In *Analyzing and Modelling Spatial and Temporal Dynamics of Infectious Diseases*, edited by D. Chen, B. Moulin, and J. Wu, 283–296, Wiley.

Ferguson N.M., Donnelly C.A., and Anderson R.M. (2001). The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions, *Science*, 292, 1155–1160.

Huque Md. H., Bondell H. D., Carroll R. J., and Ryan, L. M. (2016). Spatial regression with covariate measurement error: A semiparametric approach, *Biometrics*, 72, 678–686.

Jin X., Carlin B.P., and Banerjee S. (2005). Generalized hierarchical multivariate CAR models for areal data, *Biometrics*, 61, 950–961.

Kermack W.O., and McKendrick A.G. (1927). A contribution to the mathematical theory of epidemics, *Proceedings of the Royal Society of London, Series A*, 115 (772), 700–721.

Kotchi S.O., et al. (2016). Estimation and uncertainty assessment of surface microclimate indicators at local scale using airborne infrared thermography and multispectral imagery. In: *Geospatial Technology - Environmental and Social Applications*, edited by P. Imperatore, InTech, DOI: 10.5772/64527.

Lawn S.D., and Churchyard G. (2009). Epidemiology of HIV-associated tuberculosis Running Head: Epidemiology of TB /HIV, *Current Opinion in HIV and AIDS*, 4, 325–333.

Lawson A.B. (2006). *Statistical Methods in Spatial Epidemiology*, 2nd ed., Wiley.

Lele, S.R., Nadeem, K., and Schmuland, B. (2010). Estimability and likelihood inference for generalized linear mixed models using data cloning, *J. Am. Statist. Assoc.*,105, 1617–1625.

Longini I.M., et al. (2005). Containing pandemic influenza at the source, *Science*, 309, 1083–1087.

Olson N.A., Davidow A.L., Winston C.A., Chen M.P., Gazmararian J.A., and Katz D.J. (2012). A national study of socioeconomic status and tuberculosis rates by country of birth, United States, 1996-2005, *BMC Public Health*, 12: 365, DOI: 10.1186/1471-2458-12-365.

Pokharel G., and Deardon R. (2016). Gaussian process emulators for spatial individual-level models for infectious disease, *The Canadian Journal of Statistics*, 44, 480–501.

Rees E.E., et al. (2015). Spatial patterns of sea lice infection among wild and captive salmon in western Canada, *Landscape Ecology*, 30, 989–1004.

Smith C.M., et al. (2015). Spatial methods for infectious disease outbreak investigations: systematic literature review, *EuroSurveillance*, 20(39), DOI: http://dx.doi.org/10.2807/1560-7917.ES.2015.20.39.30026.

Stern H.S., and Cressie N.A. (1999). Inference for extremes in disease mapping, In *Disease Mapping and Risk Assessment for Public Health*, A. Lawson, A. Biggeri, D. Bohning, E. Lesaffre, J-F. Viel and R. Bertollini (Eds.), Chichester: Wiley, 63–84.

Torabi M. (2012). Hierarchical Bayes estimation of spatial statistics for rates, *Journal of Statistical Planning and Inference*, 142, 358–365.

Torabi M. (2013). Likelihood inference in generalized linear mixed measurement error models, *Computational Statistics & Data Analysis*, 57, 549–557.

Torabi M. (2014). Likelihood inference in spatial generalized linear mixed models with multivariate CAR models for areal data, *Journal of Spatial Statistics*, 10, 12–26.

Torabi M. (2016). Hierarchical multivariate mixture generalized linear models for the analysis of spatial data: An application to disease mapping, *Biometrical Journal*, 58, 1138–1150.

Torabi M., Lele S., and Prasad N.G.N. (2015). Likelihood inference for small area estimation using data cloning, *Computational Statistics & Data Analysis*, 89, 158–171.

Torabi M., and Rosychuk R.J. (2010). Spatio-temporal modelling of disease mapping of rates, *Canadian Journal of Statistics*, 38, 698–715.

Zinszer K., Morrison K., Brownstein J.S., Marinho F., Santos A.F., and Nsoesie E.O. (2017). Reconstruction of Zika virus introduction in Brazil, *Emerging Infectious Diseases Journal*, 23, 1–7.