



An Introductory Framework for Choosing Spatiotemporal Analytical Tools in Population-Level Eco-Epidemiological Research

Kaushi S. T. Kanankege^{1*}, Julio Alvarez², Lin Zhang³ and Andres M. Perez¹

¹ Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, Saint Paul, MN, United States, ² Departamento de Sanidad Animal, Centro de Vigilancia Sanitaria Veterinaria (VISAVET), Facultad de Veterinaria, Universidad Complutense, Madrid, Spain, ³ Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, United States

OPEN ACCESS

Edited by:

Flavie Vial,
Animal and Plant Health Agency,
United Kingdom

Reviewed by:

Kayoko Shioda,
Yale University, United States
Maysa Pellizzaro,
Universidade Federal da Bahia, Brazil
Alessandra Carioli,
University of Southampton,
United Kingdom

*Correspondence:

Kaushi S. T. Kanankege
kanan009@umn.edu

Specialty section:

This article was submitted to
Veterinary Epidemiology and
Economics,
a section of the journal
Frontiers in Veterinary Science

Received: 04 February 2020

Accepted: 15 May 2020

Published: 07 July 2020

Citation:

Kanankege KST, Alvarez J, Zhang L
and Perez AM (2020) An Introductory
Framework for Choosing
Spatiotemporal Analytical Tools in
Population-Level Eco-Epidemiological
Research. *Front. Vet. Sci.* 7:339.
doi: 10.3389/fvets.2020.00339

Spatiotemporal visualization and analytical tools (SATs) are increasingly being applied to risk-based surveillance/monitoring of adverse health events affecting humans, animals, and ecosystems. Different disciplines use diverse SATs to address similar research questions. The juxtaposition of these diverse techniques provides a list of options for researchers who are new to population-level spatial eco-epidemiology. Here, we are conducting a narrative review to provide an overview of the multiple available SATs, and introducing a framework for choosing among them when addressing common research questions across disciplines. The framework is comprised of three stages: (a) pre-hypothesis testing stage, in which hypotheses regarding the spatial dependence of events are generated; (b) primary hypothesis testing stage, in which the existence of spatial dependence and patterns are tested; and (c) secondary-hypothesis testing and spatial modeling stage, in which predictions and inferences were made based on the identified spatial dependences and associated covariates. In this step-wise process, six key research questions are formulated, and the answers to those questions should lead researchers to select one or more methods from four broad categories of SATs: (T1) visualization and descriptive analysis; (T2) spatial/spatiotemporal dependence and pattern recognition; (T3) spatial smoothing and interpolation; and (T4) geographic correlation studies (i.e., spatial modeling and regression). The SATs described here include both those used for decades and also other relatively new tools. Through this framework review, we intend to facilitate the choice among available SATs and promote their interdisciplinary use to support improving human, animal, and ecosystem health.

Keywords: geographical/spatial analysis, geostatistics, epidemiology, disease mapping, framework

SPATIAL EPIDEMIOLOGY

Spatial epidemiology is defined as “the description and analysis of geographic variations in disease with respect to demographic, environmental, behavioral, socioeconomic, and infectious risk factors” (1). The importance of understanding the interplay between genetic, population, and environmental factors, and temporal characteristics of diseases in relation to space (2–4) has provided a set of powerful reasons to further develop the field of spatial epidemiology. The integration of epidemiological concepts, spatial analysis, geographic information system (GIS), and

statistics leads to the accomplishment of the objectives of spatial epidemiology in understanding and modeling spatiotemporally explicit health risks (5–10). Essentially, geostatistics was originated in fields of geoscience, and the use of geostatistics on health data is synonymously referred to as “medical/health geography” or “spatial/geographical epidemiology” (11, 12).

The poster child of spatiotemporal epidemiological studies is Dr. John Snow’s map of cholera deaths in Soho, London, in 1854 (13, 14). Dr. Snow used the map to support his theory that disease was associated with contaminated water, contrary to the popular belief at the time that it was airborne (14). Dr. Snow’s classic work is an early example of how spatial epidemiological methods may support improving the quality of epidemiological investigations, eventually providing risk estimates in a timely manner to support decision and policy in preventive and control measures (15–17). Traditionally, spatial epidemiology focused on two major concepts: (a) mapping and spatial pattern analysis, such as cluster analysis, to determine visual and geographical relational cues (pre-hypothetical stages of research), and (b) using ecologic approaches to recognize etiologic clues of disease spread and explanatory factors (hypothesis-driven research) (18). However, the emergence of a large variety of tools and methods over the last decades has made the landscape of spatiotemporal epidemiological tools quite complex, challenging researchers ability to identify the analytical approaches most suitable for their needs.

SPATIOTEMPORAL VISUALIZATION AND ANALYTICAL TOOLS (SATs)

A plethora of SATs, especially geostatistical tools, have been published and used in the field of spatial epidemiology (15, 19). However, for a beginner in spatial eco-epidemiology, selecting an appropriate analytical tool is often a challenging decision. Different disciplines, including epidemiology, econometrics, and ecology, use different SATs to address similar research questions (20–23). Juxtaposing these diverse techniques may support an interdisciplinary approach of shared knowledge while providing a list of options for researchers. The choice of SATs depends on a variety of factors/criteria. The majority of the published reviews and books on SATs are focused on describing the features of the tools/methods and do not guide a beginner researcher through the options to consider when choosing a spatial eco-epidemiological analysis. The objective of the paper here was to suggest a framework that facilitates choosing SATs which enables the researchers to analyze existing epidemiological data, draw inferences, and plan future research in spatiotemporal epidemiology.

DATA USED IN SPATIOTEMPORAL ANALYSIS

The types of spatial data that can be used in epidemiology to represent the distribution of diseases and adverse events in space include (1) point-referenced data (presence and absence of the disease or number of animals at each farm location), (2) point-pattern data (presence of the disease: where the

disease occurrence itself is random giving rise to a “spatial point process”), and (3) areal data or “lattice data” (number of disease cases aggregated by an administrative division such as counties) (19, 24). The first case is often referred to as “geocoded” or “geostatistical” data (19). The point-referenced data and areal data may be of binary, count, or continuous in nature. The key difference between point-referenced and point-pattern data is that the former has a set of pre-known locations from which a certain value for a given variable was observed, whereas in the latter the events are assumed to have a stochastic or random nature (19). Therefore, in point-pattern data both the location and the observation of the disease themselves are random or stochastic. While the term “lattice data” may lead to the assumption that the areal units are regular shaped grids, in practice most areal data are summarized over irregular lattice such as administrative divisions. Reduced spatial explicitness may lead to aggregation of the events by administrative divisions and non-availability of the temporal details would limit the researcher to use purely spatial tools for the analysis.

While disease status data are the primary focus, epidemiological studies often look into association of the disease with underlying risk factors, such as human population density, air pollution parameters, temperature, precipitation, or soil pH among many other possible examples, which vary continuously over the space. These variables that are usable on GIS platforms are available from various data base sources in the form of point-referenced observations, polygon maps, or gridded i.e., “raster” maps. WorldClim [www.worldclim.org; (25, 26)] and LandScan Global Population Database (27) are examples of such data sources. The relevant value of these continuous variables, at each location where the disease status has been determined, can be extracted and used for further analysis, i.e., point-referenced data (19). The availability of exact location details and the time of the case supports more spatiotemporally explicit and reliable analysis. Unless specified as applicable to a particular type of data only, SATs described here are suitable to be used point-pattern, point-referenced, or areal data. It is important to notice that under certain circumstances the data types can be converted from one form to another. Point-referenced data can be summarized and represented by administrative divisions (i.e., polygon data). For example, point-referenced data representing 10 different farm locations recorded with a disease can be represented as 10 cases with in the county. Similarly, disaggregation of areal data with certain assumptions, such as density dependent disaggregation (28), is possible. Representing the area by the centroid of each polygon, thus, converting areal data into a point-referenced format, which, of course, is a simplification of the analysis that may be acceptable only under certain circumstances.

A FRAMEWORK FOR CHOOSING SPATIOTEMPORAL EPIDEMIOLOGICAL TOOLS

Here, we are suggesting a framework for choosing SATs (Figure 1). The framework is classified into three stages: (a) pre-hypothesis testing/hypothesis generating stage; (b) primary

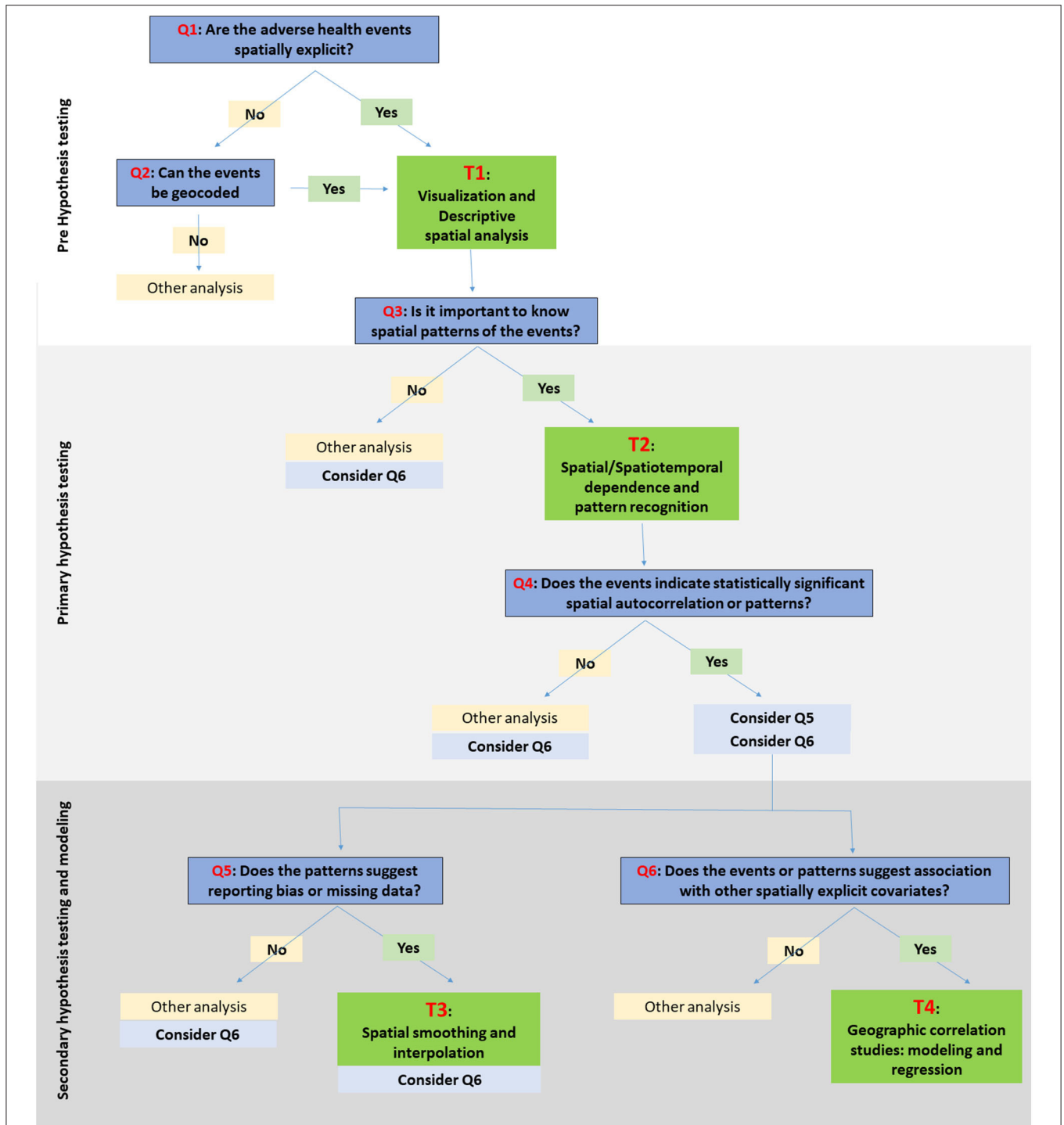


FIGURE 1 | Schematic illustration of a framework for choosing spatiotemporal visualization and analytical tools (SATs). The research questions/objectives are identified with Q1:Q6. The specific SATs under the relevant categories, i.e., T1:T4, are listed in **Table 1**.

hypothesis testing stage; and (c) secondary-hypothesis testing and spatial modeling stage where the predictions and inferences are made. The primary hypothesis refers to the existence of spatial dependence and spatial patterns in the distribution

of adverse health events, while the secondary hypotheses involve the association of the events with risk factors/covariates. The different types of SAT are broadly classified into four categories: (T1) visualization and descriptive analysis; (T2)

spatial/Spatiotemporal dependence and pattern recognition; (T3) spatial smoothing and interpolation; and (T4) spatial correlation studies: modeling and regression. The types of data primarily applicable with different SATs are listed under T1:T4. The framework seeks to suggest a suitable category of the SAT among the four, based on the stage of the research question. The types of SAT that are commonly used in epidemiological studies are listed under each category (T1:T4) in **Table 1** and discussed briefly below. The usage of tools are further discussed in relation to one example case study. It is important to note, however, that this is not a systematic review on the existing SATs, and that the classification used here is, somewhat, arbitrary, given the subjective nature of the problem. This contribution of a narrative review, while not an exhaustive description of SATs, intends to provide a short guide to introductory-level population and ecological scientists on commonly used tools and encourage the users to explore the diverse algorithms for more informed conclusions. Detailed reviews on SATs can be found elsewhere (6, 7, 10, 23, 138), as well as, a glossary of commonly used terms and their definitions in spatial epidemiology is found in Rezaeian et al. (11).

COMMONLY USED SPATIOTEMPORAL VISUALIZATION AND ANALYTICAL TOOLS (SATs)

T1 Tools for Visualization and Descriptive Analysis

Spatial data visualization is one of the key steps in understanding and generating hypotheses on the spatial distribution of events. Global Navigation Satellite Systems (GNSS), such as Global Positioning System (GPS); Global Navigation Satellite System (GLONASS); Galileo; Navigation Indian Constellation (NavIC); and BeiDou provide the ability to position the exact geospatial locations during the data collection phase. In the absence of GNSS based data, geocoding plays a major role to generate spatially explicit databases (29, 30). In addition to the visualization, description of the extent of spatial distribution by means of size, shape, and directionality of the spread supports understanding the extent of the adverse health/environmental effect. Descriptive analysis using T1 tools may support planning primary interventions including assigning vaccine or surveillance buffer zones and recognizing the distance to closest epidemiologically important features.

GIS is a system which enables capturing, storing, visualizing, and analyzing spatially explicit or “georeferenced” data to cartographic projections (31, 139). The true value of the ability to place data or measurements on a map, either as discrete events using its exact location (i.e., point-referenced data) or as continuous data by regular grids (i.e., raster data), is the ability to assess possible relationships within the data. GIS technology makes it technically feasible to integrate large amounts of data collected from different sources into a single georeferenced map/model for analysis. Therefore, GIS plays a major role in the spatial analysis as a platform which facilitates bringing data and

analytical techniques together. The key analytical tools are listed under T2:T4.

T2 Tools for Spatial/Spatiotemporal Dependence and Pattern Recognition Measures of Spatial Autocorrelation

According to Walter Tobler’s First Law of Geography, “everything is related to everything else, but near things are more related than distant things (140).” This phenomenon, otherwise known as spatial autocorrelation or spatial dependence, is a key component of spatial epidemiology. The majority of the T2 techniques are focused on determining the extent to which data are spatially autocorrelated and performing hypothesis tests after accounting for spatial autocorrelation (141). Assumptions involved in the analytics include the spatial stationarity, isotropic spatial autocorrelation, and spatial continuity (141). In simpler terms these assumptions imply that events (infectious diseases in animals for example) of the considered spatial process are homogeneously distributed across the region regardless of geographical directions or barriers. However, understanding the violations of these assumptions, i.e., detecting patterns of non-stationarity or anisotropy, is paired with the descriptive analytics (32). Moran’s I (37), Geary’s C (38), Mantel test (39), and Getis Ord (40, 41), which often referred to as “global spatial autocorrelation indices” (142) are the commonly used techniques to measure spatial autocorrelation.

Measurement of spatial heterogeneity, i.e., uneven distribution of the populations and risk factors across the geographical space, is another important component for understanding the disease process. Spatial heterogeneity measures could be either (1) local where we measure whether an attribute at one site is different from its surrounding or (2) stratified where the attributes are stratified within strata, such as Agro-ecological zones or land use categories in which the spatial variance between strata was measured. An example of local measures of spatial heterogeneity is Getis Ord G_i^* [i.e., hot-pot/cold spot analysis; (40, 41)]. Other techniques such as G-statistics are increasingly available facilitating the measurement of stratified spatial heterogeneity (51). The indices of spatial heterogeneity provide opportunity to quantitatively measure the differences and compare the landscape patterns of populations and risk factors.

Spatial Cluster Analysis

A spatial cluster is an excess of events or measurements in certain areas in geographic space, compared to the null expectation of complete spatial randomness (143). The cluster analysis is generally aimed at detecting if there is any clustering in the spatial data (i.e., Global cluster analysis), and detecting and locating the clusters (local cluster analysis and focused cluster analysis). In general, the cluster analysis provides information about the cluster morphology, including the magnitude of the excess/deficit feature, geographic size, shape, and the locations of spatial clusters.

Detecting first-order adjacencies such as Local Indicators of Spatial Autocorrelation (LISA) statistics (41, 50) and nearest-neighbors relationships such as used in Cuzick and Edward’s

TABLE 1 | A summary of types of common spatial analytical tools and their purpose.

| | Purpose | Measure | Commonly used techniques | D* | References |
|---|---|--|---|---------------------------|--|
| T1: Visualization and descriptive analysis | Transformation of locational information into geographic coordinates | Geocoding/georeferencing | GIS based geocoding of street address, postal code, or administrative divisions | pp, pr, ar | (29–31) |
| T2: Spatial/Spatiotemporal dependence and pattern recognition | Visualization and description of the size and shape of the spatial distribution | Exploratory spatial data analysis | Mean center | pp, pr, ar | (32) |
| | | | Median center | | (32) |
| | | | Convex hull | | (33) |
| | | | Standard deviation (weighted by attributes) | | (32) |
| | | | Directional mean and variance | | (34) |
| | | | Moran scatter plot | | (35) |
| | Characterize nearby features | Features with in a distance band/buffer zone | pr, ar | | (31, 36) |
| | | | | Distance to feature | (31) |
| | | | | Overlaying features | (31) |
| | Test whether there is spatial dependence in the event data | Spatial autocorrelation | pr, ar | Global Moran's I | (37) |
| | | | | Geary's C | (38) |
| | | | | Mantel test | (39) |
| | | | | Geti's ord | (40, 41) |
| | | Spatial autocorrelation among regression residuals | pr, ar | Moran's I test | (42, 43) |
| | | | | Kelejian–Robinson test | (44, 45) |
| | | Distance analysis | pr, ar | Nearest neighbor analysis | (46) |
| | | | | Ripley's K | (47, 48) |
| | Distance matrices | | | (31) | |
| | Measure the uneven distribution of the populations and risk factors | Local or stratified spatial heterogeneity | pr, ar | Getis Ord Gi* | (40, 41) |
| | | | | K-means clustering | (49) |
| Anselin's local Moran's I (L-Moran) | | | | (50) | |
| Spatial stratified heterogeneity test | | | | (51) | |
| Measure the spatial dependence while accounting for background population | | | Oden's Ipop | ar | [(52, 53); https://www.biomedware.com] |
| Test whether there is any spatial trends | Testing for first-order effects | | Trend analysis | pr, ar | (18, 54, 55) |
| Test whether there is any spatial clustering in the data | Global cluster detection | pp, pr, ar | Nearest neighbor test | (46) | |
| | | | Cuzick and Edward's test (case-control data) | (56) | |
| | | | Local indicators of spatial association (LISA) | (50) | |
| Locate the clusters and the statistical significance of the clustering | Purely spatial local cluster detection | ar | Spatial scan statistics | (57–59) | |
| | | | Flexscan | (60) | |
| | | | Turnbull's test | pr, ar | (61) |
| Test whether there is space and time clustering in the data | Spatiotemporal cluster detection | pp, pr, ar | Besag and Newell's test | (62) | |
| | | | Knox test | (63) | |
| | | | Mantel test | (39) | |
| | | | Barton's test | (64) | |
| | | | kth nearest neighbor test for time-space interaction | (65) | |
| Detect the direction of progression of an event over time | Spatiotemporal directionality | pr, ar | Space-time permutation scan statistic | (66, 67) | |
| | | | Edrer-Myers-Mantel test | (68, 69) | |
| | | | Spatiotemporal directionality test | pr, ar | [(53, 70); https://www.biomedware.com] |
| | | | Spatiotemporal anisotropy parameter | | |

(Continued)

TABLE 1 | Continued

| | Purpose | Measure | Commonly used techniques | D* | References | | |
|--|--|--|--|---|-------------------|----------------|------------|
| T3: Spatial smoothing and interpolation | Quantifying spatial variations in event intensity; spatial point pattern (SPP) intensity | Density based point pattern recognition | Univariate Kernel density estimation (KDE) | pr | (73–75) | | |
| | | | Multidimensional KDE | | (76, 77) | | |
| | | | Empirical Bayes smoothing (EBS) | ar | (78, 79) | | |
| | Smoothing and interpolation | Deterministic spatial interpolation | | Thiessen (Voronoi) polygons | pr | (80) | |
| | | | | Neighborhood matrices | | (31) | |
| | | | | Inverse Distance Estimation (IDW) | | (32, 81, 82) | |
| | | | | Triangulated Irregular Network (TIN) | | (83, 84) | |
| | | | | Headbang smoothing | | (85–87) | |
| | | Spatial modeling with stochastic partial differential equations (SPDE) | | pr | (88, 89) | | |
| | | Geostatistical interpolation and spatial regression | | Kriging | pr | (32, 90, 91) | |
| | Spline regression models | | | (92) | | | |
| | Trend Surface Interpolation | | | (93–96) | | | |
| | Multivariate spatial interpolation | | Co-kriging | pr | (32, 91, 97) | | |
| Regression kriging | | | (98–100) | | | | |
| Spatiotemporal interpolation | | Space-time kriging | pr | (101, 102) | | | |
| | | Autoregressive spatial smoothing and temporal Spline smoothing | | (103) | | | |
| T4: Geographic correlation studies: modeling and regression | Estimate the probability of disease spread using explanatory variables | Regression at spatial units | Ordinary least square regression and test for spatial autocorrelation of residuals | pp, pr, ar | (42, 43, 45) | | |
| | | | Spatial lag model with independent variable representing neighbors | | (104, 105) | | |
| | | Spatial and spatiotemporal error autoregression models for areal data (When regression residuals have spatial autocorrelation) | | Simultaneous autoregressive (SAR) models | pr, ar | (19, 24, 106) | |
| | | | | Geographically weighted regression (GWR) | | (107, 108) | |
| | | | | Purely spatial: Conditional autoregressive (CAR) models | | (19, 109, 110) | |
| | | | | Spatiotemporal CAR models | | (111, 112) | |
| | | | | Two-stage space-time mixture modeling | | (113) | |
| | | | | Latent structure models | | (113–115) | |
| | | Spatial and spatiotemporal models for point-level data | | Point process models with weighted sum approximation | pp | (116, 117) | |
| | | | | Conditional logistic model | | pp, pr | (118, 119) |
| | | | | Separable models for spatiotemporal data | | | (19) |
| | | | | Non-separable models for spatiotemporal | | | (19) |
| | | Measure the gravitation of adverse effects and the risk factors based on distance | Estimate most probable spatial interactions between entities | Gravity models | pr, ar | (120–123) | |
| Analysis of spatially explicit time-to-event data | Spatial survival models | Spatial cure rate model | pr | (124) | | | |
| | | Frailty models | | (124) | | | |
| Estimate the probability of disease when the disease occurrence is correlated with environmental variables | Environmental/Ecological niche modeling | Maximum Entropy Ecological Niche modeling (Maxent) | pr | (125–127) | | | |
| | | Genetic Algorithm for Rule Set Production (GARP) | | (128–130) | | | |
| | | Machine/statistical learning techniques | | pr | (131, 132) | | |
| | Generalized additive models (GAMs) | (133–135) | | | | | |
| | Artificial neural networks (ANN) | (136, 137) | | | | | |

D* Column represents the type of data primarily applicable on the set of tools, where, pp, point-pattern; pr, point-referenced; ar, areal data.

(56) test can be considered as global cluster detection techniques. Most local cluster-detection techniques employ circular scanning windows, such as the scan statistic (58), Turnbull's test (61), and Besag and Newell's (62) test. In scan statistics, a circular scanning window of varying sizes that moves across the study area is used to compare the observed-to-expected ratio of the cases compared to the expected spatial randomness was calculated, and the windows that maximize this likelihood ratio were recognized as the most likely clusters (58). Some of these local cluster analyses such as scan statistics have been incorporated into widely used software such as SaTScan that enable temporal, spatial, and spatiotemporal cluster analysis in a user-friendly manner. However, it is essential to realize that spatial variation and hence cluster morphology is complex, and may not be well-described by the circular cluster window approaches (143, 144). Therefore, alternative approaches that are flexible for the cluster shape such as Flex scan (60), Upper Level Set scan statistics (145), and B-statistics (146) have been introduced. A detailed description on the spatial pattern recognition and cluster analytical techniques are found elsewhere (143). The performance of SATs designed to detect clusters can be highly sensitive to the level of aggregation of the data (147). Therefore, while the clusters detected based on point-pattern or point-referenced data are intuitive to interpret, the clusters of data aggregated at large areal units requires caution. Distance based assignment of the neighbors instead of considering shared borders between areal units has been suggested (147). Morris and Munasinghe (148) have offered a solution through a user defined computer algorithm that combines existing areal units, such as administrative divisions, into regions with populations large enough to diminish spurious variability in disease rates while limiting the loss in resolution.

T3 Tools for Spatial Smoothing and Interpolation

Spatial Smoothing Techniques

Many research studies on adverse health/environmental events apply spatial smoothing and interpolation techniques to improve estimation and for exploratory mapping of risk (149). There is a variety of smoothing techniques and they can be broadly categorized as global (the same function is applied to all the data points and predictions are made using the entire dataset) and local (the same function is applied to sub-sets of data points based on the neighborhood) smoothing techniques. Kernel smoothing, one of the widely used techniques, facilitates visualization of the intensity of events (73) while accounting for background spatial distribution of the population at risk (150), and generate tolerance contours (i.e., confidence regions) for which the relative risk of a disease is significantly high (74, 75). Kernel smoothing can be used to describe and visualize the intensity or the spatial relative risk of health threats. Smoothing techniques are used to reduce noise by shrinking values toward the adjacent observations and estimate the spatial trend, which is applicable to both homogenous and heterogeneous point processes (75, 151). In a heterogeneous point process in which the intensity of the spatially varying event varies within the study area, smoothing is used to increase accuracy of the estimation of the event

intensity using either parametric or non-parametric methods (73–75). Spatial smoothing techniques use a moving weighted function to reduce the noise component, where the differences in the values on a surface are accentuated resulting in a spatially continuous map. Commonly used spatial smoothing techniques include kernel density estimation (KDE) [(73, 74, 152, 153)] and headbanging (85–87), which are considered as alternatives of detecting circumscribing clusters of varying shapes in lieu of circular clusters (74, 143). Empirical Bayes smoothing (EBS) is a specific case of spatial smoothing where the denominator i.e., varying population at risk over the map is used as a measure of the confidence in risk estimates. Therefore, the confidence of estimates are higher in highly populated areas, whereas, the estimates of relative risk would have high margins of error in the less populated areas (79). For example, if two counties have same the standardized incidence ratio (SIR) but have different population sizes, the confidence of EBS estimates would be higher for the county with a larger population size.

Spatial Interpolation Techniques

Spatial interpolation techniques are used to estimate or predict values at unknown locations using available/known data points (32). These tools can be broadly categorized as deterministic (they use the extent of similarity or distance to create the surface using measured points) and geostatistical (they use the statistical properties of the measured points to create the interpolated surface) interpolations. The resulting interpolated surfaces i.e., statistical surfaces are raster layers and often can be considered as risk maps in epidemiological analyses. There are multiple spatial interpolation techniques including Inverse distance estimation (IDW) (81), Triangulated Irregular Network (TIN) (5, 83), Kriging as well as its variations such as Co-kriging (32), and Trend Surface Interpolation (93–96) are among the commonly used techniques. TIN represents the surface by a set of contiguous and non-overlapping triangles connecting the original data points and allows construction of 3-dimensional surfaces based on a secondary variable of a researcher's choice, which, for example, the prevalence of a disease in a farm location. A review by Li and Heap (84) summarizes and compares several interpolation methods used in environmental sciences that are highly applicable in eco-epidemiological studies as well.

Geostatistical interpolation, such as kriging can be understood as a two-step process, where, step 1 is fitting the spatial variogram or likelihood for the data observed at the sampled points; and step 2 involves the interpolation of values for unsampled points or blocks using the weights derived from this covariance structure (32). In situations in which disease events are biased or undersampled, co-kriging can be used to enhance the accuracy of the estimation using a highly sampled auxiliary variable (154). For example, when invasive species detected at lakes are underreported, but the known invasions are highly correlated with the visitors/boater traffic in-and-out of the lakes and data are available for this variable, boater traffic network may use as an auxiliary variable to determine the lakes that are likely to be invaded (155). Trend surface interpolation facilitates mapping variables while allowing for the local fluctuations. Therefore, trend surface analysis may reflect the regional distribution,

trend, and the local variation of the mapped disease (156, 157). Interpolation techniques, their model assumptions, and usage are discussed extensively, elsewhere (32, 96).

Spatiotemporal interpolation techniques are used to predict variables in-between and beyond observation times (101, 102). In space-time kriging, the spatial, temporal, and spatiotemporal dependence structures are modeled using spatiotemporal variograms (102). Modeling the spatial and temporal components independently is one of the drawbacks in most of the spatiotemporal interpolation techniques (158). A detailed discussion on the spatiotemporal interpolation techniques used in the environmental modeling is found elsewhere (158). Recent developments including spatial modeling with stochastic partial differential equations (SPDE) have further improved spatial and spatiotemporal smoothing using Bayesian inference (88, 89).

T4 Tools for Geographical Correlation Studies: Modeling and Regression

Spatial Regression Models

In geographic correlation studies in epidemiology, spatial regression analysis is commonly used to examine the effects of certain risk factors/covariates on disease incidence while accounting for the spatial autocorrelation/dependence (19, 104, 159–161). Spatial dependence is incorporated into the model specifications typically using a spatial lag term or spatial error autoregression models [i.e., assigning autoregression terms for regression residuals; (104, 160)]. This is because the standard regression models assume that observations are independent, an assumption that is not met when spatially dependent data are analyzed. Fitting regression models while assigning a variable to represent the neighbor effect is one way of modeling the spatial dependence. For example in spatial lag model in which we assume that disease status in at one location is affected by the disease status at the nearby locations, a “lag” term, which is a specification of disease status at nearby locations, is included in the regression, and its coefficient and *p*-value are interpreted as for the independent variables (104). Both Frequentist and Bayesian spatial regression techniques have been extensively used in epidemiological analyses. Spatial regression models vary by their computational complexity, capacity of capturing spatial heterogeneity, and the quantification of uncertainty associated with parameter estimates (161).

Spatial error autoregressive models for discrete/areal data include: Simultaneous autoregressive (SAR) models (19, 24, 106, 162), Geographically weighted regression (163), and Conditional autoregressive models (CAR) with neighborhood structures defined based on Besag, York, and Mollie (BYM) model or Leurox (109, 110). Defining the neighbors for areal data is done based on contiguity including first-order contiguity (i.e., presence of shared borders between polygons such as adjacent counties); graph-based contiguity (i.e., based on defined algorithms such as nearest-neighbor graphs); or distance-based contiguity [i.e., neighbors within 10 km; (45)]. Due to sampling and reporting variabilities of disease incidences and risk factors, borrowing

strength from neighboring regions to get more reliable estimates is the motivation behind these spatially dependent regression models (e.g., closer neighbors might receive higher weights). This strategy of borrowing information from neighbors is applicable in autoregressive models, where the spatial or spatiotemporal structure is modeled via sets of autocorrelated random effects (19, 109, 164).

In addition to accounting for the spatial dependency, multiple spatiotemporal regression models have been used in epidemiological studies that enable the researchers to analyze the influence of spatial and temporal dependence of disease events and risk factors (19, 165). Detailed descriptions on spatial and spatiotemporal autoregressive models can be found elsewhere (19, 165). For example, latent structure models which accounts for the heterogeneity or the discontinuity in risk surface such that homogenous areas can be grouped together while discriminating for the risk levels (114).

When the events are recorded as point-referenced data from locations within a continuous spatial domain, such as by households or animal farms in a certain area, the binary outcome that the adverse event occurs in each location is assumed to have an underlying continuous spatial process. Spatial processes with binary outcomes are usually modeled by spatial logistic or probit regression models. Assigning the spatial dependence and neighbors in spatial process is complicated. This is because point-referenced spatial data often come as multivariate measurements at each location and we anticipate dependence between measurements both at a particular location as well as across locations. For example presence of a certain animal disease in a farm is correlated with the farms own characteristics including number of animals and management practices, as well as the presence of neighboring farms. Separable and non-separable spatiotemporal regression models are commonly used to model spatial point processes (19, 166).

Environmental/Ecological Models

Ecological niche modeling (ENM) approaches are widely used to characterize the complexity and heterogeneity of the landscapes in research related to epidemiologically relevant vector and parasite-reservoir distributions (167, 168). In addition to the characterization of the areas where disease is distributed, ENM is used to identify potential distributional areas in response to the likely geographic shifts in distributional areas of species or phenomena under scenarios of climate change or changing land use (169). Genetic Algorithm for Rule Set Production (GARP) (129, 130); Maximum Entropy Ecological Niche modeling (Maxent) (125, 126); and Machine/statistical Learning Techniques such as random forest (131, 132) and artificial neural networks (ANN) (136, 137) are the commonly used algorithms in epidemiology. Most ENM studies use presence-only data for the analyses. Further details regarding GARP, Maxent, and other ENM algorithms are found elsewhere [(125, 126, 128, 129)]. Additionally, hybrid methods that are bringing together multiple tools are being used in several disciplines to improve estimation and prediction abilities in spatial analysis.

EVALUATING THE PERFORMANCE OF SPATIOTEMPORAL ANALYTICAL TOOLS

Model Performance Indicators

Evaluating model performance is important when choosing between similar SATs (Especially those listed under T3 and T4). These measures include correct classification rate (CCR) (170), model sensitivity and specificity (i.e., the number of correctly classified cases) and area under the receiver operating characteristics (ROC) curve (170, 171). The sensitivity of a spatial model in disease mapping can be defined as the model's ability to correctly predict high-risk areas/locations, whereas, the specificity of the model would be its ability to correctly identify low-risk areas/locations. Error and accuracy measures, such as root mean squared error (RMSE), are also used to measure how wrong the resultant model estimates can be (138). Similarly, penalized-likelihood criteria for comparing models including Akaike information criterion (AIC) (172), Bayesian information criterion (BIC) (173), Deviance information criterion (DIC) (174, 175), and Watanabe-Akaike information criterion (WAIC) (176) are used in regression models as relative measures to compare between models and evaluate goodness of fit with penalty on model complexity. Further reading on the choice of model selection criterion is found elsewhere (177, 178).

Model Validation Techniques

The SATs, especially the predictive modeling and correlation models (listed under T3 and T4 of **Table 1**), are evaluated for their performance because the predictions would have no merit if the accuracy of the models cannot be assessed using independent data (138, 170, 179). A variety of techniques are available to validate the SATs (Listed under T3 and T4 of **Table 1**). Data partitioning techniques such as bootstrapping (180, 181), randomization (182), prospective sampling (182, 183), and k-fold partitioning (184, 185), leave-one-out cross-validation (138) are commonly used to determine training and testing datasets for model validations.

Cross validation, i.e., partitioning the data into several subsets and each fitting the model excluding one subset and validating the fitted model's ability to correctly predict the risk areas using the excluded subset of data, is one of the common practices in spatial model validation (138, 185). This includes dividing the data over space or time. For example, if the incident data are from 2000 through 2018, fitting model using early data/incidents and validation of the model predictions using recent events is considered an approach of temporal cross validation. Temporal cross validation is also achieved through the prospective sampling where new cases are evaluated against already built models from a different region or from a different time (170). A review by Anselin (179) discuss model validation techniques used in spatial econometrics in relation to the statistical validity of the models. The model fitting concerns related to theory, hypothesis testing, choice of criteria, and practical considerations are discussed under this criteria of model validations (179).

AVAILABLE SOFTWARE TOOLS FACILITATING SAT

Multiple free and proprietary software tools are available facilitating the spatiotemporal analytical studies. However, there is no quality control over to assess the accuracy, reliability, and sustainability of the majority of those non-proprietary software. Some software, such as SaTScanTM (<https://www.satscan.org>) and ArcGIS (<https://geocode.arcgis.com>), have become successful commercial products that are widely in use (7, 186), while others are underutilized due to less popularity and irregular maintenance. Sustainability and maintenance of these software is essential when incorporating these software based eco-epidemiological analyses into surveillance or intervention measures. An overview of the spatial data analytical software is found elsewhere (186).

Geocoding can be implemented using either commercial GIS software or online that are developed by governmental (Ex. USGS map locator: <https://store.usgs.gov/map-locator>), private (ArcGIS Online Geocoding Service by Esri (<https://geocode.arcgis.com/arcgis/>); QGIS Geocoding Plugins (<https://plugins.qgis.org/plugins/GeoCoding/>); Geocoding using Google maps (<https://cloud.google.com/maps-platform>), or through educational organizations (e.g., TAMU Geo coding Services of the University of Texas A&M: <http://geoservices.tamu.edu/>). Similarly, Python based geocoding using open or commercial spatial data repositories and spatial database management systems such as Google geocoding application programming interface (API) and improving the capacity of spatial computing is a field in developing (187). These software and tools enable both batch geocoding where multiple addresses are submitted at once for geocoding, and reverse geocoding, i.e., determining the nearest street address based on given coordinates.

The commonly used user-friendly software in the spatiotemporal analysis that are capable of performing the descriptive analysis, spatial pattern recognition, smoothing/interpolation, and/or spatial modeling are ArcGIS (188), QGIS (189), GRASS (190), GeoDa [(191); <http://geodacenter.github.io/index.html>], Clusterseer [(53); <https://www.biomedware.com/>], SaTScan (<http://www.satscan.org/version.9.6>), and CrimeStat (192). Similarly, there are multiple toolboxes relevant to spatiotemporal analysis that can be used through following software: R statistical software (193), SAS (194) (SAS/STAT[®] software), STATA (195), and Matlab (Matlab: <https://www.mathworks.com>)¹. platforms that are specifically developed for handling geospatial analysis. Some of the advanced statistical software packages enables performing both frequentist and Bayesian spatial analyses. For example, the R package "spatialreg" (196, 197) enables performing frequentist spatial error models including CAR models (listed under T4), while R packages "CARBayes" (198), "CARBayesST" (165), and "R-INLA" [(88); www.r-inla.org; (199)] enables fitting Bayesian CAR models using Markov Chain Monte Carlo (MCMC)

¹MATLAB and Statistics Toolbox TM Release 2018a. Natick, MA: The MathWorks, Inc.

or Integrated Nested Laplace approximation (INLA) based estimation of the posterior distributions, respectively.

HOW TO USE THE FRAMEWORK TO CHOOSE SAT: AN EXAMPLE

While we have introduced a framework and a categorization of commonly used SATs, it is important to note that the choice of the SATs is entirely a researcher-driven decision. There are certain factors/criteria associated with the decision of choosing one method over the other. The factors include: (1) characteristics of the disease/adverse event; (2) study design; (3) spatial explicitness of data; (4) data quality and availability; (5) research question and hypothesis; (6) stakeholder involvement; and (7) existence of resources, policy, and regulations for the mitigation of events (200). These factors influence the six questions (Q1:Q6) illustrated in the framework (Figure 1).

For example, assume a researcher is interested in understanding epidemiological characteristics of natural Anthrax in animal populations and intends to use that information to plan a surveillance/vaccination program in an endemic area. Let us assume that the final output the researcher intends to have is a criteria to define zoning distances for ring vaccination or surveillance when at least one Anthrax case is reported. Firstly, understanding the extent of spread and duration of previous Anthrax outbreaks would play a major role when determining this surveillance/vaccination radii. Secondly, understanding the association between the epidemiological drivers of the disease and the characteristics of susceptible population would be of importance when planning an area-based surveillance/vaccination program.

At the pre-hypothesis stage of the framework (Supplementary Figure 1), answering questions Q1 and Q2 would guide the researcher to use T1 tools and obtain a spatially explicit data set that is ready for further spatial analysis. Anthrax, caused by a spore-forming bacterium *Bacillus anthracis*, is characterized by the prolonged survival of the spores on soil and wide range of hosts including wildlife, livestock, and human (201, 202). Therefore, the observational study designs on Anthrax are likely to be retrospective based on reported cases (203). Given Anthrax is reportable to the animal and public health authorities, most likely type of data available would be point-referenced in nature (i.e., presence of the disease at farm locations or grazing lands). Although in rare situations, data may be available aggregated at administrative divisions due to privacy policy. If the coordinates of case locations are not recorded along with the case report, geocoding the locations based on the descriptions or farm addresses would be the initiating step.

Once geocoded, answering the Q3 and the use of SATs listed under T2 would facilitate the recognition of spatiotemporal dependence between the reported cases (i.e., the primary hypothesis testing stage). Given the prolonged survival of Anthrax spores in contaminated soils/environment, in addition to the initial testing for spatial dependence, understanding the spatiotemporal dependence and spatiotemporal directionality is

the key to understand the extent of past spread of the disease. Testing whether there are space and time clustering in the data would facilitate determining any particular area/s with high relative risk for disease clusters at a specific time [i.e., disease hot-spots; (203)].

Once geocoded, the primary hypothesis testing stage of the framework and the T2 tools would facilitate the recognition of spatiotemporal dependence between the reported cases and determining any particular area with high relative risk for disease clusters [i.e., disease hot-spots; (203)]. Given the prolonged survival of the Anthrax spores conducting purely spatial and spatiotemporal dependence and directionality is the key to understand the extent of past spread of the disease. This spatiotemporal pattern detection may lead to the refinement of further research questions (Q4: Q6 of the framework) and secondary hypothesis testing using the SATs listed under T3 and T4 (Supplementary Figure 1).

Because the pathogen is invariably dependent upon the distribution of susceptible species and environmental characteristics such as soil pH, rain fall, and flood plains; the choice of predictive modeling using correlated environmental factors such as regression or ecological niche modeling (ENM) (204) is a suitable option to consider (i.e., tools under T4). However, it is important to recognize that the ideal analysis for a chronic disease like Anthrax would be spatiotemporal correlation models that enable incorporating temporal changes of both the disease and underlying environmental characteristics, in addition to space.

Once the range of cluster radii (T2 tools) and key epidemiologically important environmental factors by area (T4 tools) were identified, these two key pieces of information would facilitate informing the decisions of planning the ring vaccination/surveillance programs. For example, recognition of which areas are at high risk for Anthrax based on the models outputs from T4 tools, such as ENM (204), and the extent/cluster radii of past outbreaks using T2 tools would allow us to inform defining the minimum and maximum zoning distances for ring vaccination/surveillance.

ADVANTAGE, CHALLENGES, AND DRAWBACKS OF SATs

The framework provides an introductory guide for choosing SATs for eco-epidemiological studies. Use of SATs improves an eco-epidemiological investigation by adding precision, facilitating the comparison of distributions by means of quantitative criteria, and capturing risk factors and characteristics that are unlikely to be detected by visual inspection or analyzing data without the spatial component (6). Therefore, SAT outcomes, commonly represented as “risk maps,” may serve as estimates of the effects of “real” exposures to human, animal, and environmental health threats and facilitate recognizing the effect size at more vulnerable locations and time periods.

Common weaknesses associated with the spatial analysis and risk mapping are related to shortcomings in the accuracy of data,

choices of mapping and projections, choice of the analytical/modeling tools and relevant assumptions, and eventually the decisions related to the representation of the risk maps to the end users (205, 206). In relation to the data aggregated by administrative divisions, commonly discussed issues include “edge effect” i.e., problems posed by the presence of adjacent locations not included in the analysis but that can influence its outcome, such as an unknown disease status in a country adjacent to the study area [(207, 208)]; and the “modifiable areal unit problem (MAUP)” i.e., the existence of differences in the analytical results obtained through the analysis of the same input data after aggregation at different levels. Examples include aggregation of point data from dairy farms in to counties or data available at sub districts level into provinces. The MAUP pertains to scale and zoning effect of the divisions (209, 210). A variety of methods are discussed in the literature to quantify and account for the edge effect and MAUP issues (211, 212). When spatial analytics and models are conducted based on available and potentially biased data, the resulting risk maps are invariably subjected to the negative impact of the data quality. However, we emphasize the use of existing data, bringing several databases together, and the spatiotemporal analytical tools can support initiating the process of improving data quality.

The choice of SAT, as discussed, varies with multiple factors. Inevitably, all analytical tools and models involve certain assumptions on statistical properties of variables and often these assumptions are violated in natural environments. In other words, none of the SAT are perfect matches for any particular situation (158). For example, spatial continuity of risk is a common assumption in risk-mapping process while there can be natural (e.g., mountain range acting as a physical barrier) or infrastructural barriers (e.g., urban vs. rural neighborhoods) that violate the continuity assumption resulting in step changes of risk between adjacent areas (112). Therefore, clarity on the choice of SAT, underlying assumptions, and the seven factors/criteria is essential when choosing SAT to address eco-epidemiological problems.

FUTURE DIRECTIONS

Improving the quality of spatially explicit health and environmental data through systematic collection of high-resolution data and public participation GIS approaches such as “crowdsourcing” or “citizen science data” is increasingly popular in both public and environmental health monitoring efforts (213–215). Additionally, the use of existing databases as passive surveillance systems and improving systematic data collection are suggested as ways to generate spatially explicit animal health databases (203).

While the geostatistical techniques introduced here, especially those under T4, commonly are frequentist approaches. The hierarchical specification of geostatistical models (216), therefore the adoption of a Bayesian framework for inference and suitable

Gibbs sampling, MCMC, or INLA [(88); www.r-inla.org; (199)] for model fitting is being increasingly used. In addition to the geostatistical SATs discussed here, there are non-geostatistical spatial analytical tools such as Agent-based modeling (217–219) that are increasingly used by the researchers interested in spatial eco-epidemiological studies.

When modeling complex systems of adverse health and environmental effects, incorporation of several other analytical and modeling techniques in addition to SATs may support further exploring the phenomena including understanding the network effects (21). Spatial networks are another branch of the complex system approaches to spatial data. Because complex systems are often organized under the form of networks where nodes and edges are embedded in space, such as transportation networks of swine farms or water connectivity networks between salmon farms, the importance of connectivity in addition to the spatial proximity has a major role when determining disease transmission (220).

Predicting where the phenomenon would move/flow/spread next is an essential component in spatial modeling. SATs such as space-time kriging (T3 of **Table 1**) are capable of estimating such phenomena (221). Atmospheric dispersion models such as plume models (222) and Hybrid Single Particle Lagrangian Integrated Trajectory Model (HYSPLIT) (223) are examples of applications of spatial models that account for flow directions and cost surfaces used to predict wind-mediated transmission of arthropod-borne diseases. While these models can be considered as advanced spatiotemporal variations of SATs listed under T4 here, they can be computationally costly. Hence, for the researchers who are new to population-level spatial analysis and models, it is recommendable to start with the simpler and more established SATs to explore health or environmental threats prior to applying novel modeling techniques.

AUTHOR CONTRIBUTIONS

KK designed the framework, directed the review process, and wrote the article. JA and LZ provided expertise in methods and edited and reviewed the manuscript. AP contributed in design, expertise in methods, supervision, and revision of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded in part by the Minnesota Discovery, Research, and Innovation Economy (MnDRIVE) program and Office of the Vice President for Research (OVPR) of the University of Minnesota.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fvets.2020.00339/full#supplementary-material>

REFERENCES

1. Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. *Environ Health Pers.* (2004) 112:998–1006. doi: 10.1289/ehp.6735
2. Mayer JD. The role of spatial analysis and geographic data in the detection of disease causation. *Pergamon Press Ltd Soc Sci Med.* (1983) 17:1213–21. doi: 10.1016/0277-9536(83)90014-X
3. March D, Susser E. The eco- in eco-epidemiology. *Int. J. Epidemiol.* (2006) 35:1379–83. doi: 10.1093/ije/dyl249
4. Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* (2012) 13:97–109. doi: 10.1038/nrg3142
5. Sanson RL, Pfeiffer DU, Morris RS. Geographic information systems: their application in animal disease control. *Rev Sci Tech.* (1991) 10:179–95. doi: 10.20506/rst.10.1.541
6. Ward MP, Carpenter TE. Analysis of time-space clustering in veterinary epidemiology. *Prev Vet Med.* (2000) 43:225–37. doi: 10.1016/S0167-5877(99)00111-7
7. Beale L, Abellan JJ, Hodgson S, Jarup L. Methodologic issues and approaches to spatial epidemiology. *Environ Health Pers.* (2008) 116:1105–10. doi: 10.1289/ehp.10816
8. Clements ACA, Pfeiffer DU. Emerging viral zoonoses: frameworks for spatial and spatiotemporal risk assessment and resource planning. *Vet J.* (2009) 182:21–30. doi: 10.1016/j.tvjl.2008.05.010
9. Richardson DB, Volkow ND, Kwan M, Kaplan RM, Goodchild MF, Croyle, RT. Spatial turn in health research. *Science.* (2013) 339:1390–92. doi: 10.1126/science.1232257
10. Kirby RS, Delmelle E, Eberth JM. Advances in spatial epidemiology and geographic information systems. *Ann Epidemiol.* (2017) 27:1–9. doi: 10.1016/j.annepidem.2016.12.001
11. Rezaeian M, Dunn G, St Leger S, Appleby L. Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *J Epidemiol Commun Health.* (2007) 61:98–102. doi: 10.1136/jech.2005.043117
12. Goovaerts P. Geostatistics: a common link between medical geography, mathematical geology, and medical geology. *J South Afr Inst Min Metall.* (2014) 114:605–12.
13. Snow J. *Report of the Cholera Outbreak in the Parish of St. James, Westminster, During the Autumn of 1854.* London: John Churchill (1855). p. 97–120.
14. Koch T. *Disease Maps: Epidemics on the Ground.* Chicago, IL: University of Chicago Press. (2011). doi: 10.7208/chicago/9780226449401.001.0001
15. Cromley EK, McLafferty SL. *GIS and Public Health.* 2nd ed. New York, NY: The Guilford Publications Inc. (2011).
16. Wagner RS, Bauer SE, Vena JE. Integration of Different Epidemiologic Perspectives Applications to Spatial Epidemiology. In: Lawson AB, Banerjee S, Haining RP, Ugarte M, Chapman D, editors. *Handbook of Modern Statistical Methods. Handbook of Spatial Epidemiology.* Boca Raton, FL: Hall/CRC Press Taylor Francis Group (2016). p. 3–2.
17. Stevens KB, Pfeiffer DU. The role of spatial analysis in risk-based animal disease management. In: Lawson AB, Banerjee S, Haining RP, Ugarte, M, Chapman D, editors. *Handbook of Modern Statistical Methods. Handbook of Spatial Epidemiology.* Boca Raton, FL: Hall/CRC Press Taylor Francis Group (2016). p. 450–63.
18. Lawson AB. *Statistical Methods in Spatial Epidemiology.* 2nd Ed. Chichester: John Wiley & Sons Ltd. (2006) 3–20. doi: 10.1002/9780470035771
19. Banerjee S, Carlin BP, Gelfand, AE. *Hierarchical Modeling and Analysis for Spatial Data, Second Edition, Monographs on Statistics and Applied Probability.* 2nd ed. Boca Raton, FL: Chapman and Hall/CRC, CRC Press, Taylor and Francis Group (2014).
20. Arbia G. *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence.* Berlin: Springer-Verlag (2006).
21. Anselin L. Thirty years of spatial econometrics. *Papers Reg Sci.* (2010) 89:3–25. doi: 10.1111/j.1435-5957.2010.00279.x
22. Fortin MJ, James PMA, MacKenzie A, Melles SJ, Rayfield B. Spatial statistics, spatial regression, and graph theory in ecology. *Spat Stat.* (2012) 1:100–9. doi: 10.1016/j.spasta.2012.02.004
23. Carroll LN, Au AP, Detwiler LT, Fu TC, Painter IS, Abernethy NF. Visualization and analytics tools for infectious disease epidemiology: a systematic review. *J Biomed Inform.* (2014) 51:287–98. doi: 10.1016/j.jbi.2014.04.006
24. Cressie, NAC. *Statistics for Spatial Data. Wiley Series in Probability and Mathematical Statistics.* 2nd Revised ed. New York, NY: Wiley (1993) doi: 10.1002/9781119115151
25. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol.* (2005) 25:1965–78. doi: 10.1002/joc.1276
26. Fick SE, Hijmans, RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol.* (2017) 37:4302–15. doi: 10.1002/joc.5086
27. Bright EA, Rose AN, Urban ML. LandScan 2012™. *High Resolution Global Population Data.* UT-Battelle, LLC. Oak Ridge National Laboratory. U.S. Department of Energy (2013). Available online at: <https://landscan.ornl.gov/landscan-datasets> (accessed July 06, 2020).
28. Perez AM, Thurmond MC, Grant PW, Carpenter TE. Use of the scan statistic on disaggregated province-based data: foot-and-mouth disease in Iran. *Prev Vet Med.* (2005) 71:197–207. doi: 10.1016/j.prevetmed.2005.07.005
29. Zandbergen PA. A comparison of address point, parcel and street geocoding techniques. *Comput Environ Urban Syst.* (2008) 32:214–32. doi: 10.1016/j.compenvurbsys.2007.11.006
30. Hart TC, Zandbergen PA. Reference data and geocoding quality examining completeness and positional accuracy of street geocoded crime incidents. *Policing Int J Police Strat Manag.* (2013) 36:263–94. doi: 10.1108/13639511311329705
31. Wiecezorek AF, Delmerico AM. Geographic information systems. *Comput Stat.* (2009) 1:167–86. doi: 10.1002/wics.21
32. Isaaks EH, Srivastava, RM. *Applied Geostatistics.* New York, NY: Oxford University Press (1989).
33. Worton BJ. A convex hull-based estimator of home-range size. *Biometrics.* (1995) 51:1206–15. doi: 10.2307/2533254
34. Levefer DW. Measuring geographic concentration by means of the standard deviation ellipse. *Am J Sociol.* (1926) 32:88–94. doi: 10.1086/214027
35. Anselin L. The moran scatterplot as an ESDA tool to assess local instability in spatial association. In: Fischer M, Scholten H, Unwin D, editors. *Spatial Analytical Perspective on GIS.* London: Taylor and Francis (1996). p. 111–25. doi: 10.1201/9780203739051-8
36. Moore DA, Carpenter TE. Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiol Rev.* (1999) 21:143–6. doi: 10.1093/oxfordjournals.epirev.a017993
37. Moran PAP. Notes on continuous stochastic phenomena. *Biometrika.* (1950) 37:17–23. doi: 10.1093/biomet/37.1-2.17
38. Geary RC. The contiguity ratio and statistical mapping. *Incorporat Stat.* (1954) 5:115–45. doi: 10.2307/2986645
39. Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* (1967) 27:209–20.
40. Getis A, Ord JK. The analysis of spatial association by distance statistics. *Geogr Anal.* (1992) 24:189–206. doi: 10.1111/j.1538-4632.1992.tb00261.x
41. Ord JK, Getis A. Local spatial autocorrelation statistics: distributional issues and an application. *Geogr Anal.* (1995) 27:286–306. doi: 10.1111/j.1538-4632.1995.tb00912.x
42. Cliff A, Ord JK. Testing for spatial autocorrelation among regression residuals. *Geogr Anal.* (1972) 4:267–84. doi: 10.1111/j.1538-4632.1972.tb00475.x
43. Burridge P. On the cliff-ord test for spatial autocorrelation. *J R Stat Soc B.* (1980) 42:107–8. doi: 10.1111/j.2517-6161.1980.tb01108.x
44. Kelejian H, Robinson D. Spatial autocorrelation: a new computationally simple test with an application to per capita country police expenditures. *Reg Sci Urban Econ.* 22:317–31. doi: 10.1016/0166-0462(92)90032-V
45. Anselin L, Bera AK. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah A, Giles DE, editors. *Handbook of Applied Economic Statistics.* Marcel Dekker (1998). p. 237–89.
46. Clark PJ, Evans FC. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology.* (1954) 35:445–53. doi: 10.2307/1931034
47. Ripley BD. Modelling spatial patterns. *J. R. Stat. Soc. B Stat Methodol.* (1977) 39:172–92. doi: 10.1111/j.2517-6161.1977.tb01615.x
48. Dixon PM. Ripley's K function. *Encyclopedia Environ.* (2002) 3:1796–803.

49. MacQueen JB. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Symposium on Math, Statistics, and Probability*. Berkeley, CA: University of California Press (1967). p. 281–97.
50. Anselin L. Local indicators of spatial association – LISA. *Geogr Anal.* (1995) 27:93–115. doi: 10.1111/j.1538-4632.1995.tb00338.x
51. Chen X, Wang K. Geographic area-based rate as a novel indicator to enhance research and precision intervention for more effective HIV/AIDS control. *Prev Med Rep.* (2017) 5:301–7. doi: 10.1016/j.pmedr.2017.01.009
52. Oden N. Adjusting Moran's I for population density. *Stat Med.* (1995) 14:17–26. doi: 10.1002/sim.4780140104
53. Jacquez GM, Estberg L, Greiling D, Durbeck H, Do E, Long A, Rommel B. ClusterSeer v.2.05. In: *User Manual book 2. Software for the Detection and Analysis of Event Clusters*. Ann Arbor, MI: BioMedware Inc. (2012)
54. Fotheringham AS. Trends in quantitative methods I: stressing the local. *Progr Hum Geogr.* (1997) 21:88–96. doi: 10.1191/030913297676693207
55. Fotheringham AS. Trends in quantitative methods III: stressing the visual. *Progr Hum Geogr.* (1999) 23:597–606 doi: 10.1191/030913299667756016
56. Cuzick J, Edwards R. Spatial clustering for inhomogeneous populations. *J R Stat Soc B.* (1990) 52:73–104. doi: 10.1111/j.2517-6161.1990.tb01773.x
57. Kulldorff M. A spatial scan statistic. *Commun Stat Theory Methods.* (1997) 26:1481–96. doi: 10.1080/03610929708831995
58. Kulldorff M, Nagarwalla N. Spatial disease clusters and inference. *Stat Med.* (1995) 14:799–810. doi: 10.1002/sim.4780140809
59. Kulldorff M. *SaTScanTM v9.4.1: Software for the Spatial and Space-Time Scan Statistics.* (2009). Available online at: <http://www.satscan.org/> (accessed July 06, 2020).
60. Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *Int J Health Geogr.* (2005) 4:11. doi: 10.1186/1476-072X-4-11
61. Turnbull BW, Iwano EJ, Burnett W, Howe HL, Clark LC. Monitoring for clusters of disease: applications to leukaemia incidence in upstate New York. *Am J Epidemiol.* (1990) 132:S136–43. doi: 10.1093/oxfordjournals.aje.a115775
62. Besag J, Newell J. The detection of clusters in rare diseases. *J R Stat Soc A.* (1991) 154:143–55. doi: 10.2307/2982708
63. Knox EG. The detection of space-time interactions. *J R Stat Soc C Appl Stat.* (1964) 13:25–30. doi: 10.2307/2985220
64. Barton DE, David FN, Merrington M. A criterion for testing contagion in time and space. *Ann Hum Genet.* (1965) 29:97–102. doi: 10.1111/j.1469-1809.1965.tb00504.x
65. Jacquez GM. A k nearest neighbour test for space-time interaction. *Stat Med.* (1996) 15:1935–1949. doi: 10.1002/(SICI)1097-0258(19960930)15:183.0.CO;2-I
66. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. *PLoS Med.* (2005) 2:216–24. doi: 10.1371/journal.pmed.0020059
67. Kulldorff M, Glaz J, Pozdnyakov V, Wallenstein S, Marcelo Azevedo C. Applications of spatial scan statistics: a review. In: Glaz J, Vladimir P, Sylvan W, editors. *Scan Statistics: Methods and Applications*. Birkhauser (2009). p. 129–152. doi: 10.1007/978-0-8176-4749-0
68. Ederer F, Myers MH, Mantel N. A statistical problem in space and time: do leukemia cases come in clusters? *Biometrics.* (1964) 20:626–38. doi: 10.2307/2528500
69. Stark CR, Mantel N. Lack of seasonal or temporal spatial clustering of Down's Syndrome births in Michigan. *Am J Epidemiology.* (1967) 86:199–213. doi: 10.1093/oxfordjournals.aje.a120725
70. Jacquez GM, Greiling D, Durbeck H, Estberg L, Do E, Long E, et al. ClusterSeer. In: *User Guide 2: Software for Identifying Disease Clusters*. Ann Arbor, MI: TerraSeer Press (2002). p. 316.
71. Bilonick RA. Monthly hydrogen ion deposition maps for the northeastern U.S. from July 1982 to September 1984. *Atmos Environ.* (1988) 22:1909–24. doi: 10.1016/0004-6981(88)90080-7
72. Snepvangers JJJ, Heuvelink GBM, Huisman JA. Soilwater content interpolation using spatio-temporal kriging with external drift. *Geoderma.* (2003). 112:253–71. doi: 10.1016/S0016-7061(02)00310-5
73. Bithell J. An application of density estimation to geographical epidemiology. *Stat Med.* (1990) 9:691–701. doi: 10.1002/sim.4780090616
74. Kelsall JE, Diggle PJ. Kernel estimation of relative risk. *Bernoulli.* (1995) 1:3–16. doi: 10.2307/3318678
75. Hazelton ML. Kernel smoothing methods. In: Lawson AB, Banerjee S, Haining RP, Ugarte MD, editors. *Handbooks of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Boca Raton, FL: Chapman and Hall/CRC Press (2016). p. 195–205.
76. Silverman BW. *Density Estimation for Statistics and Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC (1986).
77. Wand MP, Jones MC. *Kernel Smoothing*. Boca Raton, FL: Chapman Hall/CRC. (1995). doi: 10.1007/978-1-4899-4493-1
78. Cressie N, Read TRC. Spatial data analysis of regional counts. *Biometrical J.* (1989) 31:699–719. doi: 10.1002/bimj.4710310607
79. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics.* (1987) 43:671–81. doi: 10.2307/2532003
80. Aurenhammer F. Voronoi diagrams – a survey of a fundamental geometric data structure. *ACM Comput Surveys.* (1991) 23:345–405. doi: 10.1145/116873.116880
81. Tomczak M. Spatial interpolation and its uncertainty using automated anisotropic inverse distance weighting (IDW) - cross-validation/jackknife approach. *J Geogr Inform Decis Anal.* (1998) 2:18–30.
82. Lu GY, Wong DW. An adaptive inverse-distance weighting spatial interpolation technique. *Comput Geosci.* (2008) 34:1044–55. doi: 10.1016/j.cageo.2007.07.010
83. Burrough PA, McDonnell RA. *Principles of Geographical Information Systems*. Oxford: Oxford University Press (1998).
84. Li J, Heap AD. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecol Inform.* (2011) 6:228–41. doi: 10.1016/j.ecoinf.2010.12.003
85. Kafadar K. Smoothing geographical data, particularly rates of disease. *Stat Med.* (1996) 15:2539–60. doi: 10.1002/(SICI)1097-0258(19961215)15:23<2539::AID-SIM379>3.0.CO;2-B
86. Mungiole M, Pickle LW, Simonson KH. Application of a weighted headbanging algorithm to mortality data maps. *Stat Med.* (1999) 18:3201–9. doi: 10.1002/(sici)1097-0258(19991215)18:23<3201::aid-sim310>3.0.co;2-u
87. Gelman A, Price PN, Lin C. A method for quantifying artefacts in mapping methods illustrated by application to headbanging. *Stat Med.* (2000) 19:2309–20. doi: 10.1002/1097-0258(20000915/30)19:17/18<2309::aid-sim571>3.0.co;2-h
88. Rue H, Martino S, Chopin N. Approximate bayesian inference for latent Gaussian models using integrated nested laplace approximations (with discussion). *J R Stat Soc B.* (2009) 71:319–392. doi: 10.1111/j.1467-9868.2008.00700.x
89. Lindgren F, Rue H, Lindstrom J. An explicit link between gaussian fields and Gaussian Markov random fields: the SPDE approach (with discussion). *J R Stat Soc B.* (2011) 73:423–98. doi: 10.1111/j.1467-9868.2011.00777.x
90. Matheron G. *The Theory of the Regionalized Variables and its Applications*. Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau No. 5. Paris: École Nationale Supérieure des Mines de Paris (1971). p. 211.
91. Wackernagel H. *Multivariate Geostatistics: An Introduction with Applications*. 3rd ed. Berlin; Heidelberg: Springer-Verlag (2003). doi: 10.1007/978-3-662-05294-5_1
92. MacNab YC, Gustafson P. Regression B-spline smoothing in bayesian disease mapping: with an application to patient safety surveillance. *Stat Med.* (2007) 26:4455–74. doi: 10.1002/sim.2868
93. Lustig LK. Trend-surface analysis of the Basin Range Province, some geomorphic implications. *U.S. Geological Survey Professional Paper 500-D*. Washington DC: U.S. Government printing office (1969). p. 70. doi: 10.3133/pp500D
94. Davis JC. *Statistics and Data Analysis in Geology*. New York, NY: John Wiley and Sons (1973). p. 550.
95. Agterberg FP. Trend surface analysis. In: Gaile GL, Willmott CJ, editors. *Spatial Statistics Models*. Dordrecht: Riedel (1984). p. 147–71. doi: 10.1007/978-94-017-3048-8_8

96. Mitas L, Mitasova H. Spatial Interpolation. In: Longley P, Goodchild MF, Maguire DJ, Rhind DW, editors. *Geographical Information Systems: Principles, Techniques, Management and Applications*. Hoboken, NJ: Wiley (1999), p.481–92.
97. Stein A, Corsten LCA. Universal kriging and co-kriging as a regression procedure. *Biometrics*. (1991) 47:575–87. doi: 10.2307/2532147
98. Odeh IOA, McBratney AB, Chittleborough DJ. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma*. (1995) 67:215–26. doi: 10.1016/0016-7061(95)00007-B
99. Goovaerts P. *Geostatistics for Natural Resource Evaluation*. Applied Geostatistics. New York, NY: Oxford University Press (1997).
100. Eldeiry AA, Garcia LA. Comparison of ordinary kriging, regression kriging, and cokriging techniques to estimate soil salinity using LANDSAT images. *J Irrig Drain Eng*. (2010) 136:355–64. doi: 10.1061/(ASCE)IR.1943-4774.0000208
101. Cressie N, Wikle CK. *Statistics for Spatio-Temporal Data*. Wiley Series in Probability and Statistics John Wiley & sons Inc. (2011).
102. Biondi F. Space-time kriging extension of precipitation variability at 12km spacing from tree-ring chronologies and its implications for drought analysis. *Hydrol Earth Syst Sci Discuss*. (2013) 10:4301–35. doi: 10.5194/hessd-10-4301-2013
103. MacNab YC, Dean CB. Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics*. (2004) 57:949–56. doi: 10.1111/j.0006-341X.2001.00949.x
104. Anselin L. *Spatial Econometrics: Methods and Models*. Boston, MA; Dordrecht: Kluwer Academic Publishers (1988).
105. Anselin L. Spatial Econometrics. In: Baltagi BH editor. *A Companion to Theoretical Econometrics*. Blackwell Publishing Ltd. (2001) 310–30. doi: 10.1002/9780470996249.ch15
106. Haining R. *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press (2003). doi: 10.1017/CBO9780511754944
107. Brunson C, Fotheringham AS, Charlton ME. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geogr Anal*. (1996) 28:281–98. doi: 10.1111/j.1538-4632.1996.tb00936.x
108. Nakaya T, Fotheringham AS, Brunson C, Charlton M. Geographically weighted poisson regression for disease association mapping. *Stat Med*. (2005) 24:2695–717. doi: 10.1002/sim.2129
109. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math*. (1991) 43:1–20. doi: 10.1007/BF00116466
110. Leroux BG, Lei X, Breslow N. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Miller Jr. W, Editor. *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. Springer. (2000), p. 179–91. doi: 10.1007/978-1-4612-1284-3_4
111. Mariella L, Tarantino M. Spatial temporal conditional auto-regressive model: a new autoregressive matrix. *Austr J Stat*. (2010) 39:223–44. doi: 10.17713/ajs.v39i3.246
112. Rushworth A, Lee D, Sarran C. An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk. *J R Stat Soc C Appl Stat*. (2017) 66:141–57. doi: 10.1111/rssc.12155
113. Lawson AB, Choi J, Cai B, Hossain M, Kirby RS, Liu JH. Bayesian 2-stage space-time mixture modeling with spatial misalignment of the exposure in small area health data. *J Agric Biol Environ Stat*. (2012) 17:417–41. doi: 10.1007/s13253-012-0100-3
114. Hossain MM, Lawson AB. Space-time bayesian small area disease risk models: development and evaluation with a focus on cluster detection. *Environ Ecol Stat*. (2010) 17:73–95. doi: 10.1007/s10651-008-0102-z
115. Cai B, Lawson AB, Hossain MM, Choi J. Bayesian latent structure models with space-time dependent covariates. *Stat Modell*. (2012) 12:145–64. doi: 10.1177/1471082X1001200202
116. Berman M, Turner R. Approximating point process likelihoods with GLIM. *Appl Stat*. (1992) 41:31–8. doi: 10.2307/2347614
117. Hossain MM, Lawson AB. Approximate methods in bayesian point process spatial models. *Comput Stat Data Anal*. (2009) 53:2831–42. doi: 10.1016/j.csda.2008.05.017
118. Diggle P, Rowlingson B. A conditional approach to point process modeling of elevated risk. *J Royal Stat Soc Series A*. (1994) 157:433–40. doi: 10.2307/2983529
119. Diggle P. Spatio-temporal point processes. In: Finkenstadt B, Held L, Isham V, editors. *Methods Applications. Statistical Methods for Spatio-Temporal Systems* London: CRC Press (2007). p. 1–45. doi: 10.1201/9781420011050.ch1
120. Nijkamp P. Reflections on gravity and entropy models. *Reg Sci Urban Econ*. (1975) 5:203–25. doi: 10.1016/0166-0462(75)90004-6
121. Bailey T, Gatrell A. *Interactive Spatial Data Analysis*. Essex: Longman (1996). doi: 10.1016/0277-9536(95)00183-2
122. Barrios JM, Verstraeten WW, Maes P, Aerts JM, Farifteh J, Coppin P. Using the gravity model to estimate the spatial spread of vector-borne diseases. *Int J Environ Res Public Health*. (2012) 9:4346–64. doi: 10.3390/ijerph9124346
123. Truscott J, Ferguson NM. Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling. *PLoS Comput Biol*. (2012) 8:e1002699. doi: 10.1371/journal.pcbi.1002699
124. Banerjee S. Spatial survival models. In: Lawson AB, Banerjee S, Haining RP, Ugarte MD, editors. *Handbooks of Modern Statistical Methods. Handbook of Spatial Epidemiology*. Boca Raton, FL: Chapman Hall/CRC Press Taylor Francis Group (2016). p. 303–311.
125. Elith J, Phillips SJ, Hastie T, Dudik M, Chee YE, Yates CJ. A statistical explanation of MaxEnt for ecologists. *Divers Distrib*. (2011) 17:43–57. doi: 10.1111/j.1472-4642.2010.00725.x
126. Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. *Ecol Model*. (2006) 190:231–59. doi: 10.1016/j.ecolmodel.2005.03.026
127. Merow C, Smith MJ, Silander JA Jr. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*. (2013) 36:001–12. doi: 10.1111/j.1600-0587.2013.07872.x
128. Stockwell DRB, Peters DG. The GARP modelling system: problems and solutions to automated spatial prediction. *Int J Geogr Inform Syst*. (1999) 13:143–58. doi: 10.1080/136588199241391
129. Stockwell DRB, Peterson AT. Effects of sample size on accuracy of species distribution models. *Ecol Model*. (2002) 148:1–13. doi: 10.1016/S0304-3800(01)00388-X
130. Blackburn JK, McNyset KM, Curtis A, Hugh-Jones ME. Modeling the geographic distribution of bacillus anthracis, the causative agent of anthrax disease, for the contiguous United States using predictive ecologic niche modeling. *Am J Trop Med Hygiene*. (2007) 77:1103–10. doi: 10.4269/ajtmh.2007.77.1103
131. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324
132. Mi C, Huettmann F, Guo Y, Han X, Wen L. Why choose random forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ*. (2016) 5:e2849. doi: 10.7717/peerj.2849
133. Hastie T, Tibshirani R. *Generalized Additive Models*. London: Chapman and Hall (1990).
134. Leathwick JR, Elith J, Hastie T. Comparative performance of generalised additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol Model*. (2006) 199:188–96. doi: 10.1016/j.ecolmodel.2006.05.022
135. Wood SN. *Generalized Additive Models: An Introduction with R*. Boca Raton FL: Chapman & Hall/CRC (2006) doi: 10.1201/9781420010404
136. Civco DL. Artificial neural networks for land-cover classification and mapping. *Int J Geogr Inform Sci*. (1993) 7:173–86. doi: 10.1080/02693799308901949
137. Sordo M. Introduction to neural networks in healthcare. *Open Clin*. (2002). p. 1–7. Available online at: <http://www.openclinical.org/docs/int/neuralnetworks011.pdf> (accessed on April 14, 2020).
138. Li J. A critical review of spatial predictive modeling process in environmental sciences with reproducible examples in R. *Appl Sci*. (2019) 9:2048. doi: 10.3390/app9102048
139. Maliene V, Grigonis V, Palevicius V, Griffiths S. Geographic information system: old principles with new capabilities. *Urban Des Int*. (2011) 16:1–6. doi: 10.1057/udi.2010.25

140. Tobler WR. A computer movie simulating urban growth in the detroit region. *Econ Geogr.* (1970) 46:234–40. doi: 10.2307/143141
141. Dormann CF, McPherson JM, Araujo MB, Bivand R, Bolliger J, Carl G, et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography.* (2007) 30:609–28. doi: 10.1111/j.2007.0906-7590.05171.x
142. Salima BA, Bellefon MD. Spatial autocorrelation indices. In: Loonis V, Bellefon MD, editors. *Handbook of Spatial Analysis: Theory Application with R.* (2018). p. 51–68.
143. Jacquez GM. Spatial cluster analysis. In: Fotheringham S, Wilson J, editors. *The Handbook of Geographic Information Science.* Blackwell Publishing (2008). p. 395–416. doi: 10.1002/9780470690819.ch22
144. Costa MA, Assuncao RM. A fair comparison between the spatial scan and the besag–newell disease clustering tests. *Environ Ecol Stat.* (2005) 12:301–19. doi: 10.1007/s10651-005-1515-6
145. Patil GP, Taillie C. Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environ Ecol Stat.* (2004) 11:183–97. doi: 10.1023/B:EEST.0000027208.48919.7e
146. Jacquez GM, Kaufmann A, Goovaerts P. Boundaries, links and clusters: a new paradigm in spatial analysis? *Environ Ecol Stat.* (2008) 15:403–19. doi: 10.1007/s10651-007-0066-4
147. Waller LA, Turnbull BW. The Effects of scale on tests for disease clustering. *Stat Med.* (1993) 12:1869–84. doi: 10.1002/sim.4780121913
148. Morris RD, Munasinghe RL. Aggregation of existing geographic regions to diminish spurious variability of disease rates. *Stat Med.* (1993) 12:1915–29. doi: 10.1002/sim.4780121916
149. Berke O. Exploratory spatial relative risk mapping. *Prev Vet Med.* (2005) 71:173–82. doi: 10.1016/j.prevetmed.2005.07.003
150. Diggle P, Rowlingson B, Su T. Point process methodology for on-line spatiotemporal disease surveillance. *Environmetrics.* (2005) 16:423–34. doi: 10.1002/env.712
151. Brunson C, Fotheringham S, Charlton M. Geographically weighted summary statistics: a framework for localised exploratory data analysis. *Comput Environ Urban Syst.* (2002) 26:501–24. doi: 10.1016/S0198-9715(01)00009-6
152. Rushton G. Improving the geographic basis of health surveillance using GIS. In: Gatrell AC, Loytonen M, editors. *GIS and Health.* London: Taylor and Francis (1998). p. 63–79.
153. Rushton G, Peleg I, Banerjee A, Smith G, West M. Analyzing geographic patterns of disease incidence: rates of late-stage colorectal cancer in Iowa. *J Med Syst.* (2004) 28:223–36. doi: 10.1023/b:joms.0000032841.39701.36
154. Rogers DJ, Sedda L. Statistical models for spatially explicit biological data. *Parasitol.* (2012) 139:1852–69. doi: 10.1017/S0031182012001345
155. Kanankege KST, Alkhamis MA, Phelps NBD, Perez AM. A probability co-kriging model to account for reporting bias and recognize areas at high risk for zebra mussels and Eurasian watermilfoil invasions in Minnesota. *Front Vet Sci.* (2018) 4:231. doi: 10.3389/fvets.2017.00231
156. Huanxin W, Nan H. Trend-surface analysis was applied to the study of the geographic distribution for LBWI in China. *Chinese J Hlth Stat.* (1991) 8:12.
157. Watson GS. *Trend Surface Analysis and Spatial Correlation.* Technical Report 124. AD 699 163. Department of Statistics, The John Hopkins University, Baltimore, MD (1969). Available online at: <https://apps.dtic.mil/dtic/tr/fulltext/u2/699163.pdf>
158. Susanto F, de Souza P, He J. Spatiotemporal interpolation for environmental modelling. *Sensors.* (2016) 16:1245. doi: 10.3390/s16081245
159. Frome EL, Checkoway H. Use of poisson regression models in estimating incidence rates and ratios. *Am J Epidemiol.* (1985) 121:309–23. doi: 10.1093/oxfordjournals.aje.a114001
160. Anselin L. Under the hood: issues in the specification and interpretation of spatial regression models. *Agric Econ.* (2002) 27:247–67. doi: 10.1111/j.1574-0862.2002.tb00120.x
161. Auchincloss AH, Gebreab SY, Mair C, Roux AVD. A review of spatial methods in epidemiology, 2000–2010. *Annu Rev Public Health.* (2012) 33:107–22. doi: 10.1146/annurev-publhealth-031811-124655
162. Wall MM. A close look at the spatial structure implied by the CAR and SAR models. *J Stat Plan Inference.* (2004) 121:311–24. doi: 10.1016/S0378-3758(03)00111-3
163. Fotheringham AS, Charlton ME, Brunson C. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environ Plan A.* (1998) 30:1905–27. doi: 10.1068/a301905
164. Lawson AB. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology.* Boca Raton, FL: Chapman and Hall/CRC Press; Taylor & Francis Group (2013).
165. Lee D, Rushworth A, Napier G. CARBayesST: spatio-temporal areal unit modeling in R with conditional autoregressive priors using the CARBayesST package. *J Stat Softw.* (2018) 84, 1–39. doi: 10.18637/jss.v084.i09
166. Paciorek CJ. Computational techniques for spatial logistic regression with large datasets. *Comput Stat Data Anal.* (2007) 51:3631–53. doi: 10.1016/j.csda.2006.11.008
167. Peterson AT, Sanchez-Cordero V, Ben Beard C, Ramsey JM. Ecologic niche modeling and potential reservoirs for chagas disease, Mexico. *Emerg Infect Dis.* (2002) 8:662–7. doi: 10.3201/eid0807.010454
168. Escobar LE, Craft ME. Advances and limitations of disease biogeography using ecological niche modeling. *Front Microbiol.* (2016) 7:1174. doi: 10.3389/fmicb.2016.01174
169. Peterson AT, Tian H, Martínez-Meyer E, Soberón J, Sánchez-Cordero V, Huntley B. Modeling distributional shifts of individual species biomes. In: Lovejoy TE, Hannah L, editors. *Climate Change and Biodiversity.* New Haven, CT: Yale University Press (2005). p. 211–28.
170. Fielding AL, Bell JF. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv.* (1997) 24:38–49. doi: 10.1017/S0376892997000088
171. Lusted LB. ROC recollected. *Editorial Med Decisi Making.* (1984) 4:131–4. doi: 10.1177/0272989X8400400201
172. Akaike H. On entropy maximization principle. In: Krishnaiah PR editor. *Applications of Statistics.* Amsterdam: North-Holland Publishing Co. (1977). p. 27–41.
173. Schwarz GE. Estimating the dimension of a model. *Ann Stat.* (1978) 6:461–4. doi: 10.1214/aos/1176344136
174. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). *J R Stat Soc B.* (2002) 64:583–639. doi: 10.1111/1467-9868.00353
175. van der Linde A. (2005). DIC in variable selection. *Statistica. Neerlandica.* 1:45–56. doi: 10.1111/j.1467-9574.2005.00278.x
176. Watanabe S. Asymptotic equivalence of bayes crossvalidation and widely applicable information criterion in singular learning theory. *J Mach Learn Res.* (2010) 11:3571–94. Available online at: <http://www.jmlr.org/papers/volume11/watanabe10a/watanabe10a.pdf>
177. Lee H, Ghosh S. Performance of information criteria for spatial models. *J Stat Comput Simul.* (2009) 79:93–106. doi: 10.1080/00949650701611143
178. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Stat Comput.* (2014) 24:997–1016. doi: 10.1007/s11222-013-9416-2
179. Anselin L. Model validation in spatial econometrics: a review and evaluation of alternative approaches. *Int Reg Sci Rev.* (1988) 11:279–316. doi: 10.1177/016001768801100307
180. Buckland ST, Elston DA. Empirical models for the spatial distribution of wildlife. *J Appl Ecol.* (1993) 30:478–95. doi: 10.2307/2404188
181. Verbyla DL, Litaitis JA. Resampling methods for evaluating classification accuracy of wildlife habitat models. *Environ Manag.* (1989) 13:783–7. doi: 10.1007/BF01868317
182. Capen DE, Fenwick JW, Inkleby DB, Boynton AC. Multivariate models of songbird habitat in New England forests. In: Verner JA, Morrison ML, Ralph CJ, editors. *Wildlife 2000: Modelling Habitat Relationships of Terrestrial Vertebrates.* Madison WI: University of Wisconsin Press (1986). p. 171–75.
183. Fielding AH, Haworth PF. Testing the generality of bird-habitat models. *Conserv Biol.* (1995) 9:1466–81. doi: 10.1046/j.1523-1739.1995.09061466.x
184. Stockwell DRB. *Machine learning and the problem of prediction and explanation in ecological modeling* (Ph.D. thesis). Australian National University, Canberra, ACT, Australia (1992).
185. Kohavi R. A study of cross-validation bootstrap for accuracy estimation and model selection. In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI),* Montreal QC (1995). p. 1137–43.
186. Anselin L. From spacestat to CyberGIS: twenty years of spatial data analysis software. *Int Reg Sci Rev.* (2012) 35:131–57. doi: 10.1177/0160017612438615

187. Kuhn W, Ballatore A. Designing a language for spatial computing. In: Bacao F, Santos MY, Painho M, editors. *AGILE*. Springer International Publishing (2015). p. 309–326. doi: 10.1007/978-3-319-16787-9_18
188. *ESRI ArcMap Version 10.7*. Redlands CA: Environmental Research Institute Inc. (2019).
189. QGIS Development Team. *QGIS Geographic Information System*. Open Source Geospatial Foundation (2018). Available online at: <https://qgis.org> (accessed July 06, 2020).
190. GRASS Development Team. *Geographic Resources Analysis Support System (GRASS) Software, Version 7.2*. Open Source Geospatial Foundation. Electronic Document (2017). Available online at: <http://grass.osgeo.org>
191. Anselin L, Syabri I, Kho Y. GeoDa: an introduction to spatial data analysis. *Geogr Anal.* (2006) 38:5–22. doi: 10.1111/j.0016-7363.2005.00671.x
192. Levine N. *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.3)*. Houston, TX; Washington, DC: Ned Levine Associates; the National Institute of Justice (2010).
193. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna (2017). Available online at: <https://grass.osgeo.org> (accessed July 06, 2020).
194. SAS Institute Inc. *SAS/STAT® 14.1*. Cary, NC: SAS Institute Inc. (2015).
195. StataCorp. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp, LLC. (2019).
196. Bivand RS, Hauke J, Kossowski T. Computing the jacobian in gaussian spatial autoregressive models: an illustrated comparison of available methods. *Geogr Anal.* (2013) 45:150–79. doi: 10.1111/gean.12008
197. Bivand R, Piras G. Comparing implementations of estimation methods for spatial econometrics. *J Stat Softw.* (2015) 63:1–36. doi: 10.18637/jss.v063.i18
198. Lee D. CARBayes: an R package for bayesian spatial modeling with conditional autoregressive priors. *J Stat Softw.* (2013) 55:1–24. doi: 10.18637/jss.v055.i13
199. Schrodle B, Held L. A primer on disease mapping and ecological regression using INLA. *Comput Stat.* (2011) 26:241–58. doi: 10.1007/s00180-010-0208-2
200. Kanankege KST. *The use of spatiotemporal analytical tools to inform decisions and policy in One Health scenarios* (Ph.D. dissertaion) (2019). Available online at: <http://hdl.handle.net/11299/202418> (accessed July 06, 2020).
201. Sternbach G. The history of anthrax. *J Emerg Med.* (2003) 24:463–7. doi: 10.1016/S0736-4679(03)00079-9
202. Driks A. The bacillus anthracis spore. *Mol Aspects Med.* (2009) 30:368–73. doi: 10.1016/j.mam.2009.08.001
203. Kanankege KST, Abdrakhmanov SK, Alvarez J, Glaser L, Bender JB, Mukhanbetkaliyev YY, et al. Comparison of spatiotemporal patterns of historic animal Anthrax outbreaks in Minnesota and Kazakhstan. *PLoS ONE.* (2019) 14:e0217144. doi: 10.1371/journal.pone.0217144
204. Mullins JC, Garofolo G, Van Ert M, Fasanella A, Lukhnova L, Hugh-Jones ME, et al. Ecological niche modeling of *Bacillus anthracis* on three continents: evidence for genetic-ecological divergence? *Plos One.* (2013) 8:8. doi: 10.1371/journal.pone.0072451
205. Ocaña-Riola R. Common errors in disease mapping. *Geospat Health.* (2010) 4:139–54. doi: 10.4081/gh.2010.196
206. Loth L, Gilbert M, Wu J, Czarnnecki C, Hidayat M, Xiao X. Identifying risk factors of highly pathogenic avian influenza (H5N1 Subtype) in Indonesia. *Prev Vet Med.* (2011) 102:50–8. doi: 10.1016/j.prevetmed.2011.06.006
207. Griffith DA, Amrhein CG. An evaluation of correction techniques for boundary effects in spatial statistical analysis: traditional methods. *Geogr Anal.* (1983) 15:352–60. doi: 10.1111/j.1538-4632.1983.tb00794.x
208. Griffith DA. An evaluation of correction techniques for boundary effects in spatial statistical analysis: contemporary methods. *Geogr Anal.* (1985) 17:81–8. doi: 10.1111/j.1538-4632.1985.tb00828.x
209. Openshaw S. The modifiable areal unit problem. In: *Concepts and Techniques in Modern Geography*. Norwich: Geo Books (1984). Available online at: <https://www.uio.no/studier/emner/sv/iss/SGO9010/openshaw1983.pdf>
210. Fotheringham AS, Wong DWS. The modifiable areal unit problem in multivariate statistical analysis. *Environ Plan A.* (1991) 23:1025–44. doi: 10.1068/a231025
211. Laurance WF, Nascimento HEM, Laurance SG, Andrade A, Ewers RM, Harms KE, et al. Habitat fragmentation, variable edge effects, and the landscape-divergence hypothesis. *Plos One.* (2007) 2:e1017. doi: 10.1371/journal.pone.0001017
212. Mu L, Wang F. A scale-space clustering method: mitigating the effect of scale in the analysis zone-based data. *Ann Assoc Am Geogr.* (2008) 98:85–101. doi: 10.1080/00045600701734224
213. Linard C, Tatem AJ. Large-scale spatial population databases in infectious disease modeling and metric derivation. *Int J Health Geogr.* (2012) 11:7. doi: 10.1186/1476-072X-11-7
214. Tatem AJ, Adamo S, Bharti N, Burgert CR, Castro M, Dorelien A, et al. Mapping populations at risk: improving spatial demographic data for infectious disease modeling and metric derivation. *Popul Health Metr.* (2012) 10:8. doi: 10.1186/1478-7954-10-8
215. Sun CC, Royle JA, Fuller AK. Incorporating citizen science data in spatially explicit integrated population models. *Ecology.* (2019) 100:e02777. doi: 10.1002/ecy.2777
216. Gelfand AE, Banerjee S. Bayesian modeling and analysis of geostatistical data. In: *Annual Review of Statistics and Its Application, Vol 4*. Palo Alto: Annual Reviews (2017) p. 245–66.
217. Auchincloss AH, Diez Roux AV. A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health. *Am J Epidemiol.* (2008) 168:1–8. doi: 10.1093/aje/kwn118
218. Crooks AT, Heppenstall, AJ. Agent-based models of geographical systems. Heppenstall AJ, Crooks AT, See LM, Batty, M, editors. Heidelberg, NY: Springer Dordrecht. (2012) 219–52. doi: 10.1007/978-90-481-8927-4
219. Tracy M, Cerda M, Keyes KM. Agent-based modeling in public health: current applications and future directions. *Annu Rev Public Health.* (2018) 39:77–94. doi: 10.1146/annurev-publhealth-040617-014317
220. Craft ME. Infectious disease transmission and contact networks in wildlife and livestock. *Philos Trans R Soc B Biol Sci.* (2015) 370:20140107. doi: 10.1098/rstb.2014.0107
221. Iglesias I, Montes F, Martínez M, Pérez A, Gogin A, Kolbasov D, et al. Spatio-temporal kriging analysis to identify the role of wild boar in the spread of African swine fever in the Russian federation. *Spat Stat.* (2018) 28:226–35. doi: 10.1016/j.spasta.2018.07.002
222. Gloster J, Mellor PS, Manning AJ, Webster HN, Hort MC. Assessing the risk of windborne spread of bluetongue in the 2006 outbreak of disease in northern Europe. *Vet Rec.* (2007) 160:54–56. doi: 10.1136/vr.160.2.54
223. Durr PA, Graham K, van Klinken RD. Sellers' revisited: a big data reassessment of historical outbreaks of bluetongue and African horse sickness due to the long-distance wind dispersion of culicoides midges. *Front Vet Sci.* (2017) 4:98. doi: 10.3389/fvets.2017.00098

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kanankege, Alvarez, Zhang and Perez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.