Stoch Environ Res Risk Assess (2017) 31:829–834
DOI 10.1007/s00477-016-1374-8

EDITORIAL

Spatio-temporal data mining in ecological and veterinary epidemiology

Aristides Moustakas[1]

Published online: 9 January 2017

Abstract Understanding the spread of any disease is a highly complex and interdisciplinary exercise as biological, social, geographic, economic, and medical factors may shape the way a disease moves through a population and options for its eventual control or eradication. Disease spread poses a serious threat in animal and plant health and has implications for ecosystem functioning and species extinctions as well as implications in society through food security and potential disease spread in humans. Space–time epidemiology is based on the concept that various characteristics of the pathogenic agents and the environment interact in order to alter the probability of disease occurrence and form temporal or spatial patterns. Epidemiology aims to identify these patterns and factors, to assess the relevant uncertainty sources, and to describe disease in the population. Thus disease spread at the population level differs from the approach traditionally taken by veterinary practitioners that are principally concerned with the health status of the individual. Patterns of disease occurrence provide insights into which factors may be affecting the health of the population, through investigating which individuals are affected, where are these individuals located and when did they become infected. With the rapid development of smart sensors, social networks, as well as digital maps and remotely-sensed imagery spatio-temporal data are more ubiquitous and richer than ever before. The availability of such large datasets (big data) poses great challenges in data analysis. In addition, increased availability of computing power facilitates the use of computationally-intensive methods for the analysis of such data. Thus new methods as well as case studies are needed to understand veterinary and ecological epidemiology. A special issue aimed to address this topic.

Keywords Epidemiology · Data analytics · Spatial analysis · Temporal analysis · Networks · Computational modelling

1 Introduction

Understanding the spread of any disease is a highly complex and interdisciplinary exercise as biological, social, geographic, economic, and medical factors may shape the way a disease moves through a population and options for its eventual control or eradication (Moustakas and Evans 2016a; Oleś et al. 2012). Disease spread poses a serious threat in animal and plant health and has implications for ecosystem functioning and species extinctions (Fisher et al. 2012) as well as implications in society through food security and potential disease spread in humans (Graham et al. 2008; Tomley and Shirley 2009).

Space–time epidemiology (Knox and Bartlett 1964) is based on the concept that various characteristics of the pathogenic agents and the environment interact in order to alter the probability of disease occurrence and form temporal or spatial patterns (Snow 1855; Ward and Carpenter 2000). Epidemiology aims to identify these patterns and factors, to assess the relevant uncertainty sources, and to describe disease in the population. Thus disease spread at the population level differs from the approach traditionally taken by veterinary practitioners that are principally concerned with the health status of the individual (Arah 2009). Patterns of disease occurrence (Markatou and Ball 2014)

✉ Aristides Moustakas
arismoustakas@gmail.com

[1] School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK

provide insights into which factors may be affecting the health of the population, through investigating which individuals are affected, where are these individuals located and when did they become infected.

2 Technological advancements

With the rapid development of smart sensors (Aanensen et al. 2009), social networks, as well as digital maps and remotely-sensed imagery spatio-temporal data are more ubiquitous and richer than ever before (Gange and Golub 2016) epidemiology in the big data era needs to integrate novel methods (Mooney et al. 2015; Pfeiffer and Stevens 2015). The availability of such large datasets (big data) poses great challenges in data analysis (Fan et al. 2014; Najafabadi et al. 2015). In addition, increased availability of computing power facilitates the use of computationally-intensive methods for the analysis of such data (Moustakas and Evans 2015). Data mining—methods combining statistics and computer science—are increasingly employed (Lynch and Moore 2016) and may provide novel insights into epidemiological problems (McCormick et al. 2014; Nelson et al. 2014).

3 Let the data speak?

Can big data replace theory? It has been suggested that the availability of a large volume of data, data deluge will make the scientific method obsolete (Anderson 2008); hypothesis-driven, or equation-driven research will become irrelevant and data mining will be used instead (Anderson 2008). This thesis has generated a large scientific discussion—for some examples across scientific disciplines see (Benson 2016; Chiolero 2013; Levallois et al. 2013; Toh and Platt 2013), for online discussions see: https://www.edge.org/discourse/the_end_of_theory.html. Adding up to the discussion it has been suggested that experts will decline in importance in the big data sector (Mayer-Schönberger and Cukier 2013). There are cases where model-free forecasting (using machine learning methods) outperforms the correct mechanistic model for simulated and experimental data (Perretti et al. 2013). However if one simply relies on data-driven science several components of scientific methods will be made poorer: thought experiments (McAllister 1996), stochastic reasoning (Christakos 2010; Pearl 1987), or theoretically-derived predictions may open a new field and propose as a testable hypothesis (Gorelick 2011); something feasible in the mathematical universe is something that may happen in the biological/physical universe (regardless upon how likely is that to happen). A classic example derives from Einstein's general relativity theory. The theory was based on the observed difference for Mercury's precession between Newtonian theory and observation i.e. the deviance between observation and a model. The theory at the time that was developed lacked data but it was at later time steps verified by data. A data-driven science is welcome but we cannot afford to lose well established, tested through time scientific methods.

4 Are more data always better?

While the answer may look an obvious yes and that the only challenge is how to handle, visualise, and analyse large datasets, this is not always the case. Big datasets bring a lot of spurious correlations which appear to be simply relationships between things that are just random noise (Silver 2012). In addition big datasets allow easier 'cherry-peaking', people can choose which fractions of the data to use in order to show something that they already support or simply to produce a novel result, while a larger dataset may have simply falsified the reported result (Silver 2012), or simply verified something that was already known (Donoho and Jin 2015), therefore this would not merit a groundbreaking result/publication (Silver 2012). In addition, factor analysis in time series in econometrics showed that collating several datasets together may generate cross-correlated idiosyncratic errors, or a dominant factor in a smaller dataset may be a dominated factor in a larger dataset (Boivin and Ng 2006). In such cases smaller datasets have yielded results at least as satisfactory or in fact even better than larger datasets (Boivin and Ng 2006; Caggiano et al. 2011). Methods accounting for the effects of cross correlated errors have been proposed (Blair and Bar-Shalom 1996). While these examples are mentioned in order to highlight problematic issues related with big data, more often than not certainly more data are desirable than fewer.

5 Data availability and model complexity

A study in climate modelling has shown that as the models are becoming increasingly complex and realistic, they are also becoming less accurate because of cumulative uncertainties (Maslin and Austin 2012). In the case of climate modelling earlier models did not account for many important factors that are now being included (Maslin and Austin 2012). The simplicity of the models also prevented the uncertainties associated with these factors from being included in the modelling. The uncertainty remained hidden. More complex models that include more factors are also associated with higher uncertainties (Maslin and

Austin 2012). There is thus the paradox that as models are becoming more complex and more realistic (matching the real world better) they also become more uncertain. Ecological systems are quite complex with many small tapering effects, large heterogeneity, and interactions that are generally unknown. On an information-theoretic approach, 'information' about the biological system under study exists in the data and the goal is to express this information in a compact way (Evans et al. 2014; Lonergan 2014); the more data available the more information exists, i.e. a more complicated statistical model may approximate the data (Burnham and Anderson 2002) and more complex predictive models (process based models such as individual based models) may be calibrated (Evans and Moustakas 2016).

6 The importance of pubic data

While several new technologies providing a large volume of data exist (mentioned earlier in this paper), publicly available data from governmental organizations as well as data sharing among scientists (Michener 2015) having public data repositories are easier than ever due to large computer storage availabilities as well as fast network connections for downloading them. These public data promote transparency and accountability in the analysis, the potential for data expansion by merging several datasets together, as well as building up the impact of the work (Kenall et al. 2014; Piwowar and Vision 2013). In order to predict and mitigate disease spread informed decisions are needed. Often decisions involve conflicts between several stakeholders (Krebs et al. 1998; Moustakas 2016). These decisions need to be taken based on data analysis and predictive models calibrated with data. Making publicly available data will greatly facilitate their analysis and to informed decisions. For a review of publicly available veterinary epidemiological data with web sources links see (Pfeiffer and Stevens 2015).

7 Spatio-temporal data mining in veterinary and ecological epidemiology

There is thus a need for new methods as well as case studies to enhance our understanding in spatio-temporal data mining in veterinary and ecological epidemiology. A special issue in the journal Stochastic Environmental Research and Risk Assessment aimed to address this topic. Potential thematics included: spatiotemporal statistics (Biggeri et al. 2016; Picado et al. 2007), stochastic analysis (Heesterbeek 2000; Marx et al. 2015), Bayesian maximum entropy modeling (Biggeri et al. 2006; Juan et al. 2016), big data analytics (Andreu-Perez et al. 2015; Guernier et al. 2016), GIS and Remote Sensing (Ferrè et al. 2016; Norman et al. 2012), Trajectories and GPS tracking (Demšar et al. 2015; Zhang et al. 2011), Agent Based Modelling calibrated with data (Dion et al. 2011; Moustakas and Evans 2015; Smith et al. 2016), decision making and risk assessment (Fei et al. 2016; Lowe et al. 2015), network and connectivity analysis (Nobert et al. 2016; Ortiz-Pelaez et al. 2006) and co-occurrence and moving objects (Miller 2012; Webb 2005). Nine contributions were finally accepted after peer reviewing.

Bayesian analysis of spatial data often uses a conditionally autoregressive prior, expresses spatial dependence commonly present in underlying risks or rates. These conditionally autoregressive priors assume a normal density and uniform local smoothing for underlying risks often violated by heteroscedasticity or spatial outliers encountered in epidemiological data. Congdon (2016) proposes a spatial prior representing spatial heteroscedasticity within a model accommodating both spatial and non-spatial variation. The method is applied both in a simulation example based on US states, as well as in a real data application considers Tuberculosis incidence in England (Congdon 2016). The code used for generating simulations is also provided in R (R Development Core Team 2016).

An understanding of the factors that affect the spread of endemic bovine tuberculosis is critical for the control of the disease. Analyses of data need to account for spatial heterogeneity, or spatial autocorrelation may inflate the significance of explanatory covariates. Brunton et al. (2016) used three methods, least-squares linear regression with a spatial autocorrelation term, geographically weighted regression, and boosted regression tree analysis, to identify the factors that influence the spread of endemic bovine tuberculosis at a local level in England and Wales. The methods identified factors related to flooding, disease history and the presence of multiple genotypes of endemic bovine tuberculosis and these factors were consistent across two of the three methods (Brunton et al. 2016).

Early warning indicators are particularly useful for monitoring and control of any disease. Malesios et al. (2016) provide an early warning method of sheep pox epidemic applied in data from Evros region, Greece. To provide inference on the mechanisms governing the progress of sheep pox epidemic (Malesios et al. 2016) follow a two-stage procedure. At the first stage, a stochastic regression model is fitted to the complete epidemic data. The second stage uses an analogy of the fitted model with branching processes in order to obtain a system of estimating the probability of the epidemic going extinct at each of several time points during this epidemic. The end result is an evidence-based early warning system that could inform the authorities about the potential spread of the disease, in real-time.

Japanese encephalitis, a vector-borne disease transmitted by mosquitoes and maintained in birds and pigs. To examine the potential epidemiology of the disease in the USA, Riad et al. (2016) use an individual-level network model that explicitly considers the feral pig population and implicitly considers mosquitoes and birds in specific areas of Florida and Carolina. To model the virus transmission among feral pigs, two network topologies are considered: fully connected and random with a defined probability networks. Patterns of simulated outbreaks support the use of the random network similar to the peak incidence of the closely related West Nile virus, another virus in the Japanese encephalitis group (Riad et al. 2016). Simulation analysis suggested two important mitigation strategies.

Disease outbreaks are often followed by a large volume of data, usually in the form of movements, locations and tests. These data are a valuable resource in which data analysts and epidemiologists can reconstruct the transmission pathways and parameters and thus devise control strategies. However, the spatiotemporal data gathered can be both vast whilst at the same time incomplete or contain errors. Enright and O'Hare (2016) provide a user friendly introduction to the techniques used in dealing with the large datasets that exists in epidemiological and ecological science and the common pitfalls that are to be avoided as well as an introduction to Bayesian inference techniques for estimating parameter values for mathematical models from spatiotemporal datasets. The analysis is showcased with a large dataset from Scotland and the code and data used in this paper are also provided (Enright and O'Hare 2016).

Mechanistic epidemiological modelling has a role in predicting the spatial and temporal spread of emerging disease outbreaks and purposeful application of control treatment in animal populations. Lange and Thulke (2016) address the newly emerging epidemic of African swine fever spreading in Eurasian wild boar using an existing spatio-temporally explicit individual-based model of wild boar. Lange and Thulke (2016) propose a mechanistic quantitative procedure to optimise calibration of several uncertain parameters based on the spatio-temporal simulation model output and the spatio-temporal data of infectious disease notifications. The best agreement with the spatio-temporal spreading pattern was achieved by parameterisation that suggests ubiquitous accessibility to carcasses but with marginal chance of being contacted by conspecifics e.g., avoidance behaviour. The parameter estimation procedure is fully general and applicable to problems where spatio-temporal explicit data recording and spatial-explicit dynamic modelling is performed.

In the last two decades, two important avian influenza viruses infecting humans emerged in China, the highly pathogenic avian influenza H5N1 virus, and the low pathogenic avian influenza H7N9 virus. China is home to the largest population of chickens and ducks, with a significant part of poultry sold through live-poultry markets potentially contributing to the spread of avian influenza viruses. Artois et al. (2016) compiled and reprocessed a new set of poultry census data and used these to analyse H5N1 and H7N9 distributions with boosted regression trees models. Artois et al. (2016) found a positive and previously unreported association between H5N1 outbreaks and the density of live-poultry markets.

Transmitted infectious diseases, aggregate regional chronic diseases, and seasonal or transitory acute diseases can cause extensive morbidity, mortality and economic burden. Since the space–time distribution of a disease attribute is generally characterized by considerable uncertainty, the attribute distribution can be mathematically represented as a spatiotemporal random field model. Christakos et al. (2016) present a random field model of disease attribute that transfers the study of the attribute distribution from the original spatiotemporal domain onto a lower-dimensionality travelling domain that moves along the direction of disease velocity. The partial differential equations connecting the disease attribute covariances in the original and the travelling domain are derived with coefficients that are functions of the disease velocity. The theoretical model is illustrated and additional insight is gained by means of a numerical mortality simulation study, which shows that the proposed model is at least as accurate but computationally more efficient than mainstream mapping techniques of higher dimensionality (Christakos et al. 2016).

Moustakas and Evans (2016b) use a very large dataset generated by a calibrated agent based model to perform network analysis, spatial, and temporal analysis of bovine tuberculosis between cattle in farms and badgers. Infected network connectedness was lower in badgers than in cattle. The contribution of an infected individual to the mean distance of disease spread over time was considerably lower for badger than cattle. The majority of badger-induced infections occurred when individual badgers leave their home sett, and this was positively correlated with badger population growth rates. The spatial aggregation pattern of the disease in cattle and badgers is different across scales—in badgers, we find that the disease is found in clusters whereas in cattle the disease is much more random and dispersed. There is little geographical overlap between farms with infected cattle and setts with infected badgers, and cycles of infections between the two species are not synchronised. The findings reflect the movements of the animals—for example, cattle move greater distances within their grounds or they can be sold to farms further afield. Conversely, badgers are social animals that live in groups, and rarely leave their homes, meaning that the

presence of TB is more clustered (Christakos et al. 2016). The research suggests that an efficient way to vaccinate badgers might be to follow the spatial pattern of TB infections. This targeted approach would save labour and costs to control the spread of the disease.

Acknowledgements I wish to thank all contributing authors for their time and effort in preparing their manuscripts as well as the reviewers whose constructive comments considerably improved all manuscripts. I wish to thank the journal's editor-in-chief George Christakos for his encouragement and support as well as Helen James, from the Journals Editorial Office, for her professionalism in bookkeeping and tracking every manuscript and the associated deadlines with timely reminders.

References

Aanensen DM, Huntley DM, Feil EJ, Al-Own FA, Spratt BG (2009) EpiCollect: linking smartphones to web applications for epidemiology, ecology and community data collection. PLoS ONE 4:e6968

Anderson C (2008) The end of theory. Wired Mag 16:16-07

Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang GZ (2015) Big Data for Health. IEEE J Biomed Health Inform 19:1193–1208

Arah OA (2009) On the relationship between individual and population health. Med Health Care Philos 12:235–244

Artois J, Lai S, Feng L, Jiang H, Zhou H, Li X, Dhingra MS, Linard C, Nicolas G, Xiao X, Robinson TP, Yu H, Gilbert M (2016) H7N9 and H5N1 avian influenza suitability models for China: accounting for new poultry and live-poultry markets distribution data. Stoch Environ Res Risk Assess. doi:10.1007/s00477-016-1362-z

Benson ES (2016) Trackable life: data, sequence, and organism in movement ecology. Stud Hist Philos Sci Part C 57:137–147

Biggeri A, Dreassi E, Catelan D, Rinaldi L, Lagazio C, Cringoli G (2006) Disease mapping in veterinary epidemiology: a Bayesian geostatistical approach. Stat Methods Med Res 15:337–352

Biggeri A, Catelan D, Conesa D, Vounatsou P (2016) Spatio-temporal statistics: applications in epidemiology, veterinary medicine and ecology. Geospat Health 11:469. doi:10.4081/gh.2016.469

Blair W, Bar-Shalom T (1996) Tracking maneuvering targets with multiple sensors: does more data always mean better estimates? IEEE Trans Aerosp Electron Syst 32:450–456

Boivin J, Ng S (2006) Are more data always better for factor analysis? J Econom 132:169–194

Brunton LA, Alexander N, Wint W, Ashton A, Broughan JM (2016) Using geographically weighted regression to explore the spatially heterogeneous spread of bovine tuberculosis in England and Wales. Stoch Environ Res Risk Assess. doi:10.1007/s00477-016-1320-9

Burnham KP, Anderson DR (2002) Model selection and multimodel inference. Springer, New York

Caggiano G, Kapetanios G, Labhard V (2011) Are more data always better for factor analysis? Results for the euro area, the six largest euro area countries and the UK. J Forecast 30:736–752

Chiolero A (2013) Big data in epidemiology: too big to fail? Epidemiology 24:938–939

Christakos G (2010) Integrative problem-solving in a time of decadence. Springer, Dordrecht, pp 243–300

Christakos G, Zhang C, He J (2016) A traveling epidemic model of space–time disease spread. Stoch Environ Res Risk Assess. doi:10.1007/s00477-016-1298-3

Congdon P (2016) Representing spatial dependence and spatial discontinuity in ecological epidemiology: a scale mixture approach. Stoch Environ Res Risk Assess. doi:10.1007/s00477-016-1292-9

Demšar U, Buchin K, Cagnacci F, Safi K, Speckmann B, Van de Weghe N, Weiskopf D, Weibel R (2015) Analysis and visualisation of movement: an interdisciplinary review. Mov Ecol 3:5

Development Core Team R (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0

Dion E, VanSchalkwyk L, Lambin EF (2011) The landscape epidemiology of foot-and-mouth disease in South Africa: a spatially explicit multi-agent simulation. Ecol Model 222:2059–2072

Donoho D, Jin J (2015) Higher criticism for large-scale inference, especially for rare and weak effects. Stat Sci 30:1–25

Enright JA, O'Hare A (2016) Reconstructing disease transmission dynamics from animal movements and test data. Stoch Environ Res Risk Assess. doi:10.1007/s00477-016-1354-z

Evans MR, Moustakas A (2016) A comparison between data requirements and availability for calibrating predictive ecological models for lowland UK woodlands: learning new tricks from old trees. Ecol Evol 6:4812–4822

Evans MR, Benton TG, Grimm V, Lessells CM, O'Malley MA, Moustakas A, Weisberg M (2014) Data availability and model complexity, generality, and utility: a reply to Lonergan. Trends Ecol Evol 29:302–303

Fan J, Han F, Liu H (2014) Challenges of big data analysis. Natl Sci Rev 1:293–314

Fei X, Wu J, Liu Q, Ren Y, Lou Z (2016) Spatiotemporal analysis and risk assessment of thyroid cancer in Hangzhou, China. Stoch Environ Res Risk Assess 30:2155–2168

Ferrè N, Songyin Q, Mazzucato M, Ponzoni A, Mulatti P, Morini M, Fan J, Xiaofei L, Shulong D, Xiangmei L, Marangon S (2016) GIS applications to support entry-exit inspection and quarantine activities. In: Gervasi O, Murgante B, Misra S, Rocha AMAC, Torre CM, Taniar D, Apduhan BO, Stankova E, Wang S (eds.), Computational science and its applications—ICCSA 2016: 16th International Conference, Beijing, China, July 4–7, 2016. Proceedings, Part III. Springer International Publishing, Cham, pp. 85–97

Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ (2012) Emerging fungal threats to animal, plant and ecosystem health. Nature 484:186–194

Gange SJ, Golub ET (2016) From smallpox to big data: the next 100 years of epidemiologic methods. Am J Epidemiol 183:423–426

Gorelick R (2011) What is theory? Ideas Ecol Evol 4:1–10

Graham JP, Leibler JH, Price LB, Otte JM, Pfeiffer DU, Tiensin T, Silbergeld EK (2008) The animal-human interface and infectious disease in industrial food animal production: rethinking biosecurity and biocontainment. Public Health Rep 123:282–299

Guernier V, Milinovich GJ, Santos MAB, Haworth M, Coleman G, Magalhaes RJS (2016) Use of big data in the surveillance of veterinary diseases: early detection of tick paralysis in companion animals. Parasit Vectors 9:1

Heesterbeek J (2000) Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. Wiley, Hoboken

Juan P, Díaz-Avalos C, Mejía-Domínguez NR, Mateu J (2016) Hierarchical spatial modeling of the presence of Chagas disease insect vectors in Argentina. A comparative approach. Stoch Environ Res Risk Assess. doi:10.1007/s00477-016-1340-5

Kenall A, Harold S, Foote C (2014) An open future for ecological and evolutionary data? BMC Evol Biol 14:66

Knox E, Bartlett M (1964) The detection of space-time interactions. J R Stat Soc Ser C (Appl Stat) 13:25–30

Krebs JR, Anderson RM, Clutton-Brock T, Donnelly CA, Frost S, Morrison WI, Woodroffe R, Young D (1998) Badgers and

bovine TB: conflicts between conservation and health. Science 279:817–818
Lange M, Thulke H-H (2016) Elucidating transmission parameters of African swine fever through wild boar carcasses by combining spatio-temporal notification data and agent-based modelling. Stoch Environ Res Risk Assess. doi:10.1007/s00477-016-1358-8
Levallois C, Steinmetz S, Wouters P (2013) Sloppy data floods or precise social science methodologies? Dilemmas in the transition to data-intensive research in sociology and economics (Chapter 5). In: Beaulieu A, Scharnhorst A, Wyatt S, Wouters P (eds) Virtual knowledge. MIT Press, Cambridge
Lonergan M (2014) Data availability constrains model complexity, generality, and utility: a response to Evans et al. Trends Ecol Evol 29:301–302
Lowe R, Cazelles B, Paul R, Rodó X (2015) Quantifying the added value of climate information in a spatio-temporal dengue model. Stoch Environ Res Risk Assess. doi:10.1007/s00477-015-1053-1
Lynch SM, Moore JH (2016) A call for biological data mining approaches in epidemiology. BioData Min 9:1
Malesios C, Kostoulas P, Dadousis K, Demiris N (2016) An early warning indicator for monitoring infectious animal diseases and its application in the case of a sheep pox epidemic. Stoch Environ Res Risk Assess. doi:10.1007/s00477-016-1316-5
Markatou M, Ball R (2014) A pattern discovery framework for adverse event evaluation and inference in spontaneous reporting systems. Stat Anal Data Min 7:352–367
Marx C, Mühlbauer V, Krebs P, Kuehn V (2015) Species-related risk assessment of antibiotics using the probability distribution of long-term toxicity data as weighting function: a case study. Stoch Environ Res Risk Assess 29:2073–2085
Maslin M, Austin P (2012) Uncertainty: climate models at their limit? Nature 486:183–184
Mayer-Schönberger V, Cukier K (2013) Big data: a revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt, Boston
McAllister JW (1996) The evidential significance of thought experiment in science. Stud Hist Philos Sci Part A 27:233–250
McCormick TH, Ferrell R, Karr AF, Ryan PB (2014) Big data, big results: knowledge discovery in output from large-scale analytics. Stat Anal Data Min 7:404–412
Michener WK (2015) Ecological data sharing. Ecol Inform 29:33–44
Miller JA (2012) Using spatially explicit simulated data to analyze animal interactions: a case study with brown hyenas in northern Botswana. Trans GIS 16:271–291
Mooney SJ, Westreich DJ, El-Sayed AM (2015) Epidemiology in the era of big data. Epidemiology (Cambridge, Mass.) 26:390–394
Moustakas A (2016) The effects of marine protected areas over time and species' dispersal potential: a quantitative conservation conflict attempt. Web Ecol 16:113–122
Moustakas A, Evans M (2015) Coupling models of cattle and farms with models of badgers for predicting the dynamics of bovine tuberculosis (TB). Stoch Environ Res Risk Assess 29:623–635
Moustakas A, Evans MR (2016a) Regional and temporal characteristics of bovine tuberculosis of cattle in Great Britain. Stoch Environ Res Risk Assess 30:989–1003
Moustakas A, Evans MR (2016b) A big-data spatial, temporal and network analysis of bovine tuberculosis between wildlife (badgers) and cattle. Stoch Environ Res Risk Assess. doi:10.1007/s00477-016-1311-x
Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E (2015) Deep learning applications and challenges in big data analytics. J Big Data 2:1–21
Nelson JC, Shortreed SM, Yu O, Peterson D, Baxter R, Fireman B, Lewis N, McClure D, Weintraub E, Xu S, Jackson LA, On Behalf of the Vaccine Safety Datalink Project (2014) Integrating database knowledge and epidemiological design to improve the implementation of data mining methods that evaluate vaccine safety in large healthcare databases. Stat Anal Data Min 7:337–351
Nobert BR, Merrill EH, Pybus MJ, Bollinger TK, Hwang YT (2016) Landscape connectivity predicts chronic wasting disease risk in Canada. J Appl Ecol 53:1450–1459
Norman SA, Huggins J, Carpenter TE, Case JT, Lambourn DM, Rice J, Calambokidis J, Gaydos JK, Hanson MB, Duffield DA (2012) The application of GIS and spatiotemporal analyses to investigations of unusual marine mammal strandings and mortality events. Mar Mamm Sci 28:E251–E266
Oleś K, Gudowska-Nowak E, Kleczkowski A (2012) Understanding disease control: influence of epidemiological and economic factors. PLoS ONE 7:e36026
Ortiz-Pelaez A, Pfeiffer D, Soares-Magalhaes R, Guitian F (2006) Use of social network analysis to characterize the pattern of animal movements in the initial phases of the 2001 foot and mouth disease (FMD) epidemic in the UK. Prev Vet Med 76:40–55
Pearl J (1987) Evidential reasoning using stochastic simulation of causal models. Artif Intell 32:245–257
Perretti CT, Munch SB, Sugihara G (2013) Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. Proc Natl Acad Sci 110:5253–5257
Pfeiffer DU, Stevens KB (2015) Spatial and temporal epidemiological analysis in the Big Data era. Prev Vet Med 122:213–220
Picado A, Guitian F, Pfeiffer D (2007) Space–time interaction as an indicator of local spread during the 2001 FMD outbreak in the UK. Prev Vet Med 79:3–19
Piwowar HA, Vision TJ (2013) Data reuse and the open data citation advantage. PeerJ 1:e175
Riad MH, Scoglio CM, McVey DS, Cohnstaedt LW (2016) An individual-level network model for a hypothetical outbreak of Japanese encephalitis in the USA. Stoch Environ Res Risk Assess. doi:10.1007/s00477-016-1353-0
Silver N (2012) The signal and the noise: why so many predictions fail-but some don't. Penguin Books, London
Smith R, Lee BY, Moustakas A, Zeigler A, Prague L, Santos R, Chung M, Gras R, Forbes V, Borg S, Comans T, Ma Y, Punt N, Jusko W, Brotz L, Hyder A (2016) Population modelling by examples ii. In: Proceedings of the summer computer simulation conference. Society for computer simulation international, Montreal, Quebec, Canada, pp 1–8
Snow J (1855) On the mode of communication of cholera. John Churchill, Marlborough
Toh S, Platt R (2013) Big data in epidemiology: too big to fail? Epidemiology 24:939
Tomley FM, Shirley MW (2009) Livestock infectious diseases and zoonoses. Philos Trans R Soc B 364:2637–2642
Ward MP, Carpenter TE (2000) Techniques for analysis of disease clustering in space and in time in veterinary epidemiology. Prev Veterin Med 45:257–284
Webb CR (2005) Farm animal networks: unraveling the contact structure of the British sheep population. Prev Veterin Med 68:3–17
Zhang Z, Chen D, Liu W, Racine JS, Ong S, Chen Y, Zhao G, Jiang Q (2011) Nonparametric evaluation of dynamic disease risk: a spatio-temporal kernel approach. PLoS ONE 6:e17381