4.6 Health-related risk modelling

Authors

Holly C.Y. Lam and **Zhe Huang**, CCOUC, Faculty of Medicine, CUHK, Hong Kong SAR, China.

Emily Y.Y. Chan, CCOUC, Faculty of Medicine, CUHK, Hong Kong SAR, China; GX Foundation, Hong Kong SAR, China.

4.6.1 Learning objectives

To have a basic understanding of some modelling methods that might be applied in research studies relevant to the following issues for health emergency and disaster risk management (Health EDRM):

1. Short-term environmental health associations.
2. Factors associated with the uptake of protection behaviours.
3. Trends of influenza.
4. Health-related vulnerability index.

4.6.2 Introduction

Health EDRM is an important approach for reducing the numerous public health impacts of disasters and emergencies (Chapter 1.2). Other chapters in this book describe research methods that require the collection of new data in prospective studies; this chapter complements these by discussing the use of statistical modelling to establish mathematical associations between variables. The chapter focuses on health-related risk models that are applicable to Health EDRM and discusses models for four particular topics: short-term environmental health associations; factors associated with the uptake of protection behaviours; trends in influenza; and health-related vulnerability index.

4.6

4.6.3 Models for evaluating short-term environmental health associations

Hydrometeorological hazards (that is, hazards related to water and weather-related events) are common triggers of disasters and account for about 95% of the people affected by disasters caused by natural hazards in the past 50 years *(1)*. Climate change is causing these extreme events to become more common and severe, leading to further impacts on human health. Improved weather forecasting and better understanding of the health risks of extreme environmental events is allowing for the implementation of effective health protection plans and improvements in resource allocation. These are supported by modelling methods for evaluating short-term associations between environmental exposures and health outcomes, and this section uses ambient temperature as an example to illustrate this. Extreme temperatures are a silent killer, due to people's lower awareness compared to other hazards *(2)*, and have caused substantial public health problems *(3-7)*.

Similar to other environmental exposures (air-pollutants, storms, for example), ambient temperature usually has a short-term association with health outcomes, ranging from hours *(8)* to weeks *(9)*, depending on the degree of exposure and the health outcome considered. A delayed effect is commonly reported for the temperature-health association, but it is not always linear. For instance, since both extreme high and low temperature may cause adverse effects on human health, there may be a v-shaped association between ambient temperature and the risk of adverse health outcomes. Combined with a seasonal effect and some other confounding effects (such as air-pollutants and population-level demographic factors), the estimation of a temperature-health association is complicated. A time series design is the most common method to reveal these short-term temperature-health associations *(10)*.

Time series data are a series of sequential records in equal time units, such as the number of deaths and the average daily or weekly temperature within a specific time period. Bhaskaran and colleagues discussed and compared time series designs used in environmental epidemiology, identifying three main types of time series study: time stratified model, periodic functions and flexible spline functions *(11)*.

For the time stratified model, exposure and outcome are associated in stratified time units. Time intervals are indicated by indicator variables (such as time period: 1, 2, up to "n") instead of the true date record. This type of model is relatively easy to understand but many parameters are included in the model and it cannot facilitate the calculation of the continuous effect from one time unit to another *(11)*.

Periodic functions (Fourier terms) model exposure and outcome by using periodic functions such as sine and cosine function to represent the periodic characteristics (such as calendar months). This model type creates smooth predictions but the period of the pattern is fixed, and this might not be appropriate for representing trends that are mathematically complicated and do not have a fixed pattern *(11)*.

Flexible spline function is a modelling approach combining different polynomial curves *(11–12)*. This design is most commonly applied in assessing short-term associations between temperature and health

outcomes *(13–14)*. It allows the health outcome to be linked to a number of exposures with different non-linear associations at the same time. This is an important feature, because most temperature-outcome associations and long-term trends are non-linear and non-periodical. Another reason for using a flexible spline to model long-term trends is that it helps control the long-term demographic factors at a population level. For example, smoking is a potential confounder of the risk of admissions to hospitals for breathing problems when studying the association with temperature but, because the proportion of people in a population who smoke does not change significantly from day to day, it does not affect the daily association between temperature and these admissions. Therefore, overall changes in the proportion of smokers can be captured by fitting a spline function for the long-term trend.

To incorporate the non-linear delayed effects of ambient temperature on health outcomes into the spline model, Armstrong *(15)* and Gasparrini *(16)* introduced the Distributed Lagged Non-linear Model (DLNM) and the corresponding R package dlnm, respectively. This modelling approach is a three-dimensional data analysis. It considers the exposure, health outcome and the delay (time) dimension at the same time. In DLNM, spline functions can also be applied on the time dimensions, thereby addressing the need to model the non-linear delayed effect in exposure-outcome associations. The combination of flexible spline approach and DLNM tackles most of the concerns when evaluating short-term environmental health associations but is complicated because it involves one more dimension than other traditional time series designs. Similar to other time series approaches, the problem of effect modification by other factors (such as age and gender) still exists and needs to be handled separately (for example, by conducting subgroup analysis). More details about the method and some examples are available elsewhere *(17)*.

4.6.4 Identifying factors associated with the uptake of protective behaviours during extreme events

Applying appropriate protective behaviours during extreme events can lower exposure to hazards and hence reduce health risk. Sociodemographic factors *(19)* and knowledge, attitude and practice (KAP) *(20-21)* are common examples of determinants of health behaviours. Identifying associated sociodemographic factors and understanding KAP for a protective behaviour provides evidence to support health promotion policies. This section introduces a statistical method for identifying factors associated with the uptake of protective behaviours, using data collected from a cross-sectional KAP survey.

Logistic regression is a regression model characterized by one binary dependent variable (outcome) and multiple independent variables (explanatory variables) *(22)*. It allows users to investigate the association between the outcome variable and an explanatory variable with adjustment for other confounders. It is used widely for identifying factors (such as knowledge and gender) that might be associated with the likelihood of a group of people acting in a certain way (taking or not taking action, for example) and comparing this to a reference group of other people.

In Health EDRM, there are usually several explanatory variables to consider but including too many explanatory variables in the model compromises its power to reveal the real associations. A general guide is that there should be at least ten cases for each explanatory variable in each outcome group *(22)* and the power increases with increasing numbers of cases. To reduce the number of explanatory variables in a regression model, univariate analysis, such as the chi-square test (for categorical variables) and t-test (for continuous variables), can be used to provide a quick assessment of the potential associating factors. Explanatory variables showing potential association with the outcome in the univariate analysis, together with some core explanatory variables (supported by literature or hypothesis) are then entered into the logistic regression model. Model selection (the process of selecting explanatory variables for a model) can also be done by removing non-significant variables from a full model or adding variables and keeping those that are significant (see Case Study 4.6.1).

Case Study 4.6.1
Data collection by telephone survey

For a community with a high level of landline telephone penetration, data collection through a telephone survey might be an appropriate way to examine knowledge, attitude and practice (KAP) in community behaviour patterns. A population-based telephone survey among the Hong Kong population investigated their weather information acquisition pattern during an intense cold spell *(23)*. The Chi-square test and a logistic regression model were used to identify independent associated factors in a two-stage analysis. Univariate analyses were used to identify potential associated factors with the outcome and factors with a p-value from the chi-square test of less than 0.20 were entered to the second stage of the analysis, the multiple logistic regression analysis, to assess their independent association with the outcome. In the univariate analyses, educational attainment, age and marital status were significantly associated with current use of smartphone apps to acquire weather information. In multiple logistic regressions, only older age and lower education level remained significantly associated with lower smartphone app usage.

4.6.5 Prediction and forecasting of influenza trend

Influenza is a global public health burden, usually associated with cold-like symptoms but leading to serious illnesses in vulnerable groups (for example, young children and the elderly) *(24)*. Influenza causes health and economic burdens, with loss of work or school hours for patients and caretakers, large numbers of emergency room visits, hospitalizations and deaths *(25–27)*. Influenza viruses gradually mutate and when a new contagious strain emerges in a community without immunity, this may lead to an epidemic. To reduce the risk of disease outbreak and disease burden, accurate prediction of strain types and the number of cases is important for primary prevention strategies. Accurate prediction facilitates effective vaccine strain selection and resource planning for the healthcare sector, and various prediction models have been developed to meet different purposes and region-specific environmental conditions. This section

introduces predictive models for vaccine selection and the forecast of influenza activity *(28–30)*.

Vaccine selection is conducted annually, in general, and is a year-long process because of the long production time for the vaccines (approximately 6 to 8 months). The process is managed collaboratively between WHO and professionals around the world, supported by global surveillance data related to influenza virus circulating in humans *(29)*. Employing present and past data, predictive models are used to identify and predict emerging influenza clades (that is, groups of virus strains that are believed to comprise of evolutionary descendants of a common virus ancestor) that may be dominant in the following year. Most of the predictive models focus on the biological determinants of the evolution of influenza, with scale from molecular, within-host, population, regional to global level. Some models infer phenotypic properties of the current population *(29)*.

Antigenicity-stability fitness model *(31)*, Epitope Clade Growth *(32)* and Local Tree Shape *(33)* are probabilistic evolutionary focused models for predicting future viral populations *(29)*. Antigenicity-stability fitness model is a validated model estimating expected growth rate (fitness) of viral clades by input of a few years of genetic and antigenic data and is able to predict frequency of trajectory of clades for about one year ahead *(31)*. Epitope Clade Growth, a model based on genealogical tree, estimates antigenic differences by extrapolating recent growth hemagglutinin clades seeded by epitope mutation *(32)*. Local Tree Shape is another genealogical tree-based model. It estimates recent clade growth from information stored in the local shape of a hemagglutinin genealogical tree *(33)*.

Linking antigenic properties and genetic data, and identification of proposed vaccine strains are two ways of inferring phenotypic properties *(29)*. They estimate the effectiveness of current vaccines for the emerging influenza strains and identify new antigenic variants at an early stage of expansion *(29)*. Strain selection involves complex decisions that require the integration of the results from different models at different scales. Integration and interpretation of data for decisions are key challenges *(29)*.

Forecasts of influenza activity have been conducted worldwide to support preparedness activities *(28, 30)*. These forecasts can be based on single or multiple measures. Typical measures are peak periods (time), peak and outbreak magnitude and case counts by day or week *(30)*.

There are two main modelling approaches: (i) statistical models without consideration of the epidemiology process and (ii) epidemiological models *(28)*. The common statistical models are time series models, generalized linear models, Bayesian network and classification methods *(28)*. The susceptible-infections-removed (SIR) models and agent-based models (AMBs), which include exposure, infection, transmission and behaviours in the calculations, are the common epidemiological approaches for forecasting influenza activity *(28)*. Agent-based models can be operated by simulation algorithm to estimate key epidemiological parameters and then to forecast future activity (see Case Study 4.6.2). While time series models can capture the temporal dependence of health outcomes, epidemiological approaches are able to account for health-related human behaviours and address questions related to the impact of prevention measures on health. Dynamic virological data and syndromic influenza-like

illness are common input data for surveillance data forecast models *(28)*. Real-time forecast models, making use of retrospective forecast information have been developed for temperate regions, with seasonal winter epidemics such as the USA *(34–35)*. However, these real time models performed less well in subtropical regions, such as Hong Kong SAR, with a two peak or year-round pattern *(36)*.

Case Study 4.6.2
Forecast Model - Simulation Optimization (SIMOP)

Nsoesie and colleagues *(37)* introduced a **sim**ulation **op**timization (SIMOP) approach for forecasting influenza epidemic infection curves. This combines the individual-based epidemiology model and the optimization technique for model parameters estimation (Nelder-Mead simplex method). The three model parameters estimated were the disease transmissibility, incubation and infectious period distribution. The individual-based model consisted of a dynamic social contact network (representing Montgomery County in Virginia, Miami, Seattle and surrounding metropolitan regions of the USA) and a disease model with the several assumptions.

There were three main steps for the SIMOP: (i) initialize the individual-based model and the Nelder-Mead simplex method, (ii) run the Nelder-Mead algorithm to find new parameter sets, and (iii) simulate an epidemic using the proposed parameter set and evaluate the objective function. Steps 2 and 3 were repeated for convergency. The input measures were the sequential daily or weekly number of cases during the period of epidemic, which were simulated by the estimated disease transmissibility, incubation and infectious period distribution. The model was used to forecast the epidemic peak timing, counts of infected individuals and cumulative infected individuals.

The model predicted the peak time at seven weeks before the actual peak. Forecasting the peak count of infected and cumulative infected individual was more challenging because of the possibilities of the epidemic curve trajectories, but the forecast was found to be accurate for Montgomery County.

4.6.6 Compositing indicators/index to measure vulnerability

Climate change is set to increase the frequency and intensity of disasters due to natural hazards *(38)*. Risk assessment tools are important for saving lives and reducing losses in disasters. During disasters, the number of deaths, the number of people affected and economic loss are not only determined by the hazard itself, but also by the proportion of population exposed and the vulnerability of the community (Chapter 1.3). Understanding risk in all its dimensions is essential for effective Health EDRM, and as such, the collection of large volumes of data is a major focus of research and public interest, because it presents opportunities to describe reality accurately (Chapter 2.4). However, although large amounts of data provide information from many perspectives,there may be too many variables for a clear understanding. This problem is sometimes known as the “curse of dimensionality”.

If there are a large number of variables in a dataset, a dimension reduction method can be applied. This maps the numerous original variables into fewer independent dimensions, based on their correlation to each other. It is therefore more meaningful to summarize data as a few independent dimensions, while preserving as much of the original information as possible *(39)*.

On some occasions it is easier to interpret one composite index resulting from dimension reduction, rather than indicators from multiple perspectives, despite the simplification of the original data. A composite index can allow multi-country comparisons for complex issues, such as society development, vulnerability to environmental hazards and urban heat islands. A good quality composite index is based on careful variable selection and appropriate use of the dimension reduction method, and can facilitate communication and policy making.

Principal components analysis (PCA) and factor analysis (FA) are two examples of linear dimension reduction methods. They attempt to explain a multivariate dataset by reducing them into a smaller number of dimensions. PCA is one of the oldest multivariate techniques and is useful for displaying multivariate data as a set of dimensions (called 'principal components'). It simplifies the complexity by transforming correlated variables into a set of uncorrelated principal components *(40)*. Each principal component is rated according to the extent to which it represents the original dataset, and most of the information from the original variables is captured by the principal components rated the highest (see Case Study 4.6.3). In summary, PCA provides a concise summary of the original variables, with no probabilistic or statistical assumptions.

Case Study 4.6.3
Principal components analysis (PCA) to develop a Heat Vulnerability Index

PCA was used to combine socioeconomic indicators into a Heat Vulnerability Index in London, United Kingdom *(41)*. Nine variables were identified: households in rented tenure, households in a flat, population density (persons/hectare), households without central heating, population above 65 years old, population with self-reported health status, receiving any kind of social benefit, single pensioner households and ethnic group. These were included in the principal components analysis. Four principal components were then identified, which could be interpreted as high-density housing, poor health and welfare dependency, being elderly and isolated, and poor housing quality. Principal component loadings are weighted according to the variance they explain and summed to form the Heat Vulnerability Index. In this way, the number of independent factors (dimensions) associated with the outcomes could be decreased and interpretation of the findings was simplified.

If statistical assumptions are added into principal components analysis, the principal components analysis becomes a factor analysis (*42)*. The results from principal components analysis and factor analysis would not differ dramatically if the specific variances added are small. Like principal components analysis, factor analysis is a classical technique used to

derive fewer dimensions from a large set of variables. However, unlike principal components analysis, factor analysis can allow for further statistical inference and support assertions about a population (see Case Study 4.6.4). Although the use of factor analysis draws considerable criticism (due to the lack of uniqueness of the factor loadings, for example), it is a useful approximation for the truth and a suitable starting point for further investigation.

Case Study 4.6.4
Factor analysis to develop a Health Vulnerability Index

By using FA to create a linear combination of indicators, a Health Vulnerability Index for disaster risk reduction along the Belt and Road Initiative was developed *(17)*. The index is based on three latent factors: population status, disease prevention and coping capacity. These were derived from nine indicators: proportion of the population below 15 and above 65 years, under-five mortality ratio, maternal mortality ratio, tuberculosis prevalence, age-standardized raised blood pressure, physician ratio, hospital bed ratio, and coverage of the measles-containing-vaccine first-dose (MCV1) and diphtheria tetanus toxoid and pertussis (DTP3) vaccines.

Non-linear dimension reduction methods are an extension of the linear methods and are useful if Euclidean distances (that is, straight-line distance between two points) fail to capture the dissimilarity between the observations. These methods reduce the volume of data by simplifying it into a set of low-dimensional coordinates that preserve distances in the high-dimensional space as much as possible, but involves non-linear transformations of the data.

4.6.7 Conclusions

Risk modelling is well established and can be used in helping resource allocation in Health EDRM. In recent years, it has been applied to a wide range of temperature-related studies, but consistent associations were not often found for other climate-related topics such as rainfall or sea level rise *(17)*. Risk modelling in other contexts (such as complex emergencies) or between varying contexts (such as rural versus urban) is also needed to understand health-related impact of hazards and disasters.

4.6.8 Key messages

- **Time series analysis is widely used for establishing short-term associations between exposures and health outcomes.**
- **Factors associated with protective or preparedness behaviours can be identified by applying the multiple logistic regression method.**
- **Linking Antigenic Properties and Genetic Data, and Identification of Proposed Vaccine Strains are two ways of inference of phenotypic properties for influenza vaccine selection. They estimate the effectiveness of current vaccine strains for the emerging strains and identify new antigenic variants at an early stage of expansion.**
- **In predicting influenza trends, epidemiological approaches, such as the susceptible-infections-removed models and agent-based models, consider human behaviours and address questions related to the impact of prevention measures.**
- **In constructing a health-related risk index, dimension reduction approaches such as principle component analysis (PCA) and factor analysis are widely used to simplify the display of multivariate data.**

4.6.9 Further reading

Jackson JE. A user's guide to principal components. New York, NY: Wiley. 1991.

Wood SN. Generalized additive models: An introduction with R. Chapman and Hall/CRC. 2006.

Gasparrini A. Distributed lag linear and non-linear models in R: the package dlnm. Journal of Statistical Software. 2011: 43(8): 1.

Vynnycky E, White R. An introduction to infectious diseases modelling. Oxford, UK: Oxford University Press. 2010.

McSharry P. Parsimonious risk assessment and the role of transparent diverse models. In Risk modeling for hazards and disasters Elsevier. 2018. pp. 263-9.

4.6

4.6.10 References

1. Centre for Research on the Epidemiology of Disasters (CRED). EM-DAT: The Emergency Events Database. 2020. www.emdat.be (accessed 9 March 2020).
2. National Oceanic and Atmospheric Administration (NOAA). Excessive heat, a 'silent killer'. 2017. www.noaa.gov/stories/excessive-heat-silent-killer (accessed 9 March 2020).
3. Johnson H, Kovats RS, McGregor G, Stedman J, Gibbs M, Walton H, et al. The impact of the 2003 heat wave on mortality and hospital admissions in England. Health Statistics Quarterly. 2005: (25): 6-11.
4. Rey G, Jougla E, Fouillet A, Pavillon G, Bessemoulin P, Frayssinet P et al. The impact of major heat waves on all-cause and cause-specific mortality in France from 1971 to 2003. International Archives of Occupational and Environmental Health. 2007: 80(7): 615-26.
5. Yang J, Liu H, Ou C, Lin G, Ding Y, Zhou Q, et al. Impact of heat wave in 2005 on mortality in Guangzhou, China. Biomedical and Environmental Sciences. 2013: 26(8): 647-54.
6. Hajat S, Haines A. Associations of cold temperatures with GP consultations for respiratory and cardiovascular disease amongst the elderly in London. International Journal of Epidemiology. 2002: 31(4): 825-30.
7. Ryti NR, Guo Y, Jaakkola JJ. Global association of cold spells and adverse health effects: a systematic review and meta-analysis. Environmental Health Perspectives. 2015: 124(1): 12-22.
8. Bhaskaran K, Armstrong B, Hajat S, Haines A, Wilkinson P, Smeeth L. Heat and risk of myocardial infarction: hourly level case-crossover analysis of MINAP database. BMJ. 2012: 345: e8050.
9. Lam HCY, Chan EYY, Goggins WB. Comparison of short-term associations with meteorological variables between COPD and pneumonia hospitalization among the elderly in Hong Kong—a time-series study. International Journal of Biometeorology. 2018: 62(8): 1447-60.
10. Shrestha MS, Khan MR, Wagle N, Babar ZA, Khadgi VR, Sultan S. Chapter 13 - Review of Hydrometeorological Monitoring and Forecasting System for Floods in the Indus Basin in Pakistan. In: Indus River Basin. Elsevier. 2019: pp. 309-33.
11. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. International Journal of Epidemiology. 2013: 42(4): 1187-95.
12. Wood SN. Generalized additive models: An introduction with R. Chapman and Hall/CRC. 2006.
13. Chan EYY. Climate Change and Urban Health: The Case of Hong Kong as a Subtropical City. Routledge. 2019.
14. Gasparrini A, Guo Y, Hashizume M, Lavigne E, Zanobetti A, Schwartz J, et al. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. Lancet. 2015: 386(9991): 369-75.

15. Armstrong B. Models for the relationship between ambient temperature and daily mortality. Epidemiology 2006: 17: 624-31.

16. Gasparrini A . Distributed lag linear and non-linear models in R: the package dlnm. Journal of Statistical Software. 2011: 43(8): 1.

17. Chan EYY, Huang Z, Lam HCY, Wong CKP, Zou Q. Health vulnerability index for disaster risk reduction: application in belt and road initiative (BRI) region. International Journal of Environmental Research and Public Health 2019: 16(3): 380.

18. Chan EYY, Ho JY, Hung HH, Liu S, Lam HC. Health impact of climate change in cities of middle-income countries: the case of China. British Medical Bulletin. 2019: 130(1): 5-24.

19. Ulla Díez SM, Perez-Fortis A . Socio-demographic predictors of health behaviors in Mexican college students. Health Promotion International. 2009: 25(1): 85-93.

20. Health education: theoretical concepts, effective strategies and core competencies. WHO. 2012. http://applications.emro.who.int/dsaf/emrpub_2012_en_1362.pdf (accessed 9 March 2020).

21. Launiala A. How much can a KAP survey tell us about people's knowledge, attitudes and practices? Some observations from medical anthropology research on malaria in pregnancy in Malawi. 2009. www.anthropologymatters.com/index.php/anth_matters/article/view/31/53 (accessed 9 March 2020).

22. Sperandei S. Understanding logistic regression analysis. Biochemia Medica. 2014: 24(1): 12-8.

23. Chan EYY. Public health humanitarian responses to natural disasters. Routledge. 2017.

24. Influenza: are we ready? WHO. 2019. https://www.who.int/influenza/spotlight (accessed 9 March 2020).

25. Descalzo MA, Clara W, Guzmán G, Mena R, Armero J, Lara B, et al. Estimating the burden of influenza-associated hospitalizations and deaths in Central America. Influenza and Other Respiratory Viruses. 2016: 10(4): 340-5.

26. Vestergaard LS, Nielsen J, Krause TG, Espenhain L, Tersago K, Sierra NB, et al. Excess all-cause and influenza-attributable mortality in Europe, December 2016 to February 2017. Eurosurveillance. 2017: 22(14): 30506.

27. Young-Xu Y, van Aalst R, Russo E, Lee JK, Chit A. The annual burden of seasonal influenza in the US Veterans Affairs population. PloS ONE. 2017: 12(1): e0169344.

28. Chretien JP, George D, Shaman J, Chitale RA, McKenzie FE. Influenza forecasting in human populations: a scoping review. PloS ONE. 2014: 9(4): e94130.

29. Morris DH, Gostic KM, Pompei S, Bedford T, Łuksza M, Neher RA, et al. Predictive modeling of influenza shows the promise of applied evolutionary biology. Trends in Microbiology. 2018: 26(2): 102-18.

30. Nsoesie EO, Brownstein JS, Ramakrishnan N, Marathe MV. A systematic review of studies on forecasting the dynamics of influenza outbreaks. Influenza and Other Respiratory Viruses. 2014: 8(3): 309-16.

31. Łuksza M, Lässig M. A predictive fitness model for influenza. Nature. 2014: 507(7490): 57.

32. Steinbrück L, Klingen TR, McHardy AC. Computational prediction of vaccine strains for human influenza A (H3N2) viruses. Journal of Virology. 2014: 88(20): 12123-32.

33. Neher RA, Russell CA, Shraiman BI. Predicting evolution from the shape of genealogical trees. Elife. 2014: 3: e03568.

34. Hu H, Wang H, Wang F, Langley D, Avram A, Liu M. Prediction of influenza-like illness based on the improved artificial tree algorithm and artificial neural network. Scientific Reports. 2018: 8(1): 4895.

35. Reich NG, McGowan CJ, Yamana TK, Tushar A, Ray EL, Osthus D, et al. A Collaborative Multi-Model Ensemble for Real-Time Influenza Season Forecasting in the US. 2019. bioRxiv 566604.

36. Yang W, Cowling BJ, Lau EH, Shaman J. Forecasting influenza epidemics in Hong Kong. PLoS Computational Biology. 2015: 11(7): e1004383.

37. Nsoesie EO, Beckman RJ, Shashaani S, Nagaraj KS, Marathe MV. A simulation optimization approach to epidemic forecasting. PloS ONE. 2013: 8(6): e67164.

38. Intergovernmental Panel on Climate Change (IPCC) (2014). Climate Change 2014: Impacts, Adaptation, and Vulnerability. 2014. http://www.ipcc.ch/report/ar5/wg2 (accessed 9 March 2020).

39. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. New York, NY: Springer. 2013.

40. Everitt B, Hothorn T. An introduction to applied multivariate analysis with R. Springer Science & Business Media. 2011.

41. Wolf T, McGregor G. The development of a heat wave vulnerability index for London, United Kingdom. Weather and Climate Extremes. 2013: 1: 59-68.

41. Cosma S. Advanced Data Analysis from an Elementary Point of View. (in press). https://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/ (accessed 9 March 2020).