Author’s response to reviews

Title: A spatial-temporal statistical analysis of health seasonality: explaining HFMD infections within a children population along the Vietnamese south central coast

Authors:

NGOC PHUONG TRUONG (truongngocphuong@gmail.com)
Thuong Nguyen (nvthuong@pasteurhcm.gov.vn)
Thao Nguyen (nttthao@pasteurhcm.gov.vn)
Alfred Stein (a.stein@utwente.nl)

Version: 1 Date: 13 Jun 2019

Author’s response to reviews:

Point-by-point responses to the comments of the reviewer.

Jinfeng Wang (Reviewer 1)

1. The authors partition some spatial-temporal strata for the stationarity of modelling. The strata themselves, however, are information as well, can be tested and attributed by spatial stratified heterogeneity q statistic; the interaction between the strata as proxies of driving forces can be investigated.

Response: We followed the recommendation to use the q-statistic from the Geodetector software to measure the spatiotemporal heterogeneity and identify the dominant risk factors. We added to the revised manuscript a new section, section 2.2, lines 98-105, page 6 and lines 229-234, page 13 to present the method and the results of this spatial-temporal heterogeneity analysis. The key references of the q-statistical method and the Geodetector software were also cited in section 2.2, line 105, page 6, lines 411-415, page 23.

2. S and T should be explained when they appear at the first time;

Response: We added the definition of S and T from the first use of them in lines 19-22, page 2.

3. For clarification, list the variables or their proxies used in the paper, and draw their maps or figures.
Response: The time plot of the two risk factors or their proxies was added in Fig 4, line 243-247, page 14.

Reviewer 2

General comments: Many thanks for asking me to review this manuscript. The authors report the analysis of temporal statistics of health seasonality of HFMD infections among a population of children in Vietnam from 2012 to 2016. Social contact and preschooling amongst children under five were observed as driving factors of HFMD.

However, there is an insufficient analysis of the spatial model. There is no particular spatial model used in this study, which makes the manuscript become a temporal analysis or time series analysis of HFMD. I strongly suggest that the authors include a spatial model to map the affected area.

Response: In this work, we used the spatial-temporal autoregressive model (STAR), as we clearly mentioned in the abstract (lines 22-23) and explained with key references in the methods, section 2.5 (lines 164-165, 172-173, page 9). This STAR model also takes into account the spatial spillover effects of the incidence proportions of HFMD between the strata of the districts in the study area. We apologize for the lack of clarity in the previous text and we have now better presented it by clarifying the model in full.

Additional requests/suggestions:

1. For this manuscript to be titled "spatial-temporal analysis", the authors should include a defined and specific spatial model.

Response: The spatial model we used is in the STAR model including the spatial autoregressive model, please see our responses to the same comment above.

2. The manuscript would benefit from a review for epidemiological definitions and language.

Response: To improve the language of the manuscript for the audience from both backgrounds of spatial and medical epidemiology of HFMD, we have carefully corrected the English; and the second and third co-authors from the Pasteur Institute in Viet Nam (Thuong Vu Nguyen and Thao Thi Thanh Nguyen), who are well-known experts in the epidemiology of hand-foot-mouth
disease, have again very critically read the manuscript and improved the language for the audience from epidemiology of hand-foot-mouth disease.

3. My concern about the Fourier series analysis is that the classical periodogram used in your Fourier series is not a good estimator ('inconsistent estimator') because the number of parameters grows with the number of data points, which is not applicable in this study. However, the Fourier transforms, and its probabilities also depend on several strong assumptions that are rarely achieved in real data: evenly spaced data of infinite duration with a high sampling rate, Gaussian noise, single frequency periodicity with sinusoidal shape and stationary behavior. Each of these constraints is violated in various problems. Shape may be non-sinusoidal (e.g., elliptical orbits, eclipses). Periods may not be constant. I strongly suggest that the authors look into this critically by improving the estimator and graphical representation using the "smoothing method".

Response: The Fourier series used in this study is one of the common smoothing methods based on spectral analysis to identify and formally describe the patterns of the reported timeseries of the incidences of HFMD. The systematic patterns are the trend and seasonality. They are the salient patterns of the space-time series. The trend is the pattern that does not repeat during the time frame of the study, while the seasonality repeats its pattern over regular time intervals. The random noises (i.e. the residuals or the errors) are spatiotemporally autocorrelated.

A Fourier analysis is a regression based method on the basic sine and cosine functions. The prominent advantage of using the Fourier analysis is that the uneven patterns of the seasonality (unequal amplitudes and periods) can be efficiently modeled. This method does not require the assumption of a single frequency periodicity and stationary behavior. The problem of increasing the parameters of the model when the number of observations increased cannot be avoidable when using other smoothing methods that are usually regression based methods. However, in our study, this problem has been controlled by minimizing the generalized cross-validation value that penalizes the increasing number of parameters (lines 152-155, page 9). Other advantages of the Fourier analysis have been discussed in the manuscript at lines 304-312, pages 17-18.

The time series are monthly data, i.e. evenly spaced data in a finite duration of five years and the normality of the residuals satisfies the Gaussian assumption according to the Shapiro–Wilk test (lines 179-180, page 10, lines 226-228, page 13). In our work, we also used the LOESS smoothing method, one of the efficient and consistent method, as the benchmark. The high correlation (correlation coefficient is larger than 0.85) between the results from both methods show that the results of the Fourier analysis are acceptable (lines 299-304, page 17).