

# Bayesian zero-inflated spatio-temporal modelling of scrub typhus data in Korea, 2010-2014

Dayun Kang,<sup>1</sup> Jungsoon Choi<sup>2,3</sup>

<sup>1</sup>Department of Applied Statistics; <sup>2</sup>Department of Mathematics; <sup>3</sup>Research Institute for Natural Sciences, Hanyang University, Seoul, South Korea

## Abstract

Scrub typhus, a bacterial, febrile disease commonly occurring in the autumn, can easily be cured if diagnosed early. However, it can develop serious complications and even lead to death. For this reason, it is an important issue to find the risk factors and thus be able to prevent outbreaks. We analyzed the monthly scrub typhus data over the entire areas of South Korea from 2010 through 2014. A 2-stage hierarchical framework was considered since weather data are covariates and the scrub typhus data have different spatial resolutions. At the first stage, we obtained the administrative-level estimates for weather data using a spatial model; in the second, we applied a Bayesian zero-inflated spatio-temporal model since the scrub typhus data include excess zero counts. We found that the

zero-inflated model considering the spatio-temporal interaction terms improves fitting and prediction performance. This study found that low humidity and a high proportion of elderly people are significantly associated with scrub typhus incidence.

## Introduction

Scrub typhus is an acute febrile disease spread by the bites of the larvae of trombiculid mites infected with *Orientia tsutsugamushi*, a bacterium similar to *Rickettsia* (Ogawa *et al.*, 2002). The infection usually occurs in autumn when there is a high chance of contact with these chigger larvae (Cracco *et al.*, 2000). It is commonly distributed in the Asia-Pacific area (Figure 1) within the so-called tsutsugamushi triangle region (McCrum *et al.*, 1957). In South Korea within this triangle, an average of 8,329 patients per year were diagnosed with scrub typhus from 2010 to 2016.

Patients with scrub typhus have symptoms such as fever, headache, fatigue, swollen lymph nodes and muscle pain. They are easily cured by antibiotics (tetracycline or chloramphenicol) when administered in the early stage; however, patients who are not treated appropriately can develop complications that can lead to death, such as pneumonia, encephalitis, and multi-organ failure. Finding the risk factors for scrub typhus is important as this would contribute to prevention of outbreaks of the disease.

Previous studies suggest that meteorological factors and the proportion of elderly people influence the number of scrub typhus cases (Ogawa *et al.*, 2002; Kuo *et al.*, 2011; Tsai *et al.*, 2013; Li *et al.*, 2014). Kuo *et al.* (2011) focused the spatial distribution of scrub typhus in 350 administrative districts of Taiwan, showing by application of the Spearman rank correlation coefficient. They showed that the scrub typhus occurs more often if the temperature increases and rainfall, normalized difference vegetation index, the proportion of farmers and dry land decrease. Li *et al.* (2014) investigated the association between meteorological factors and the monthly scrub typhus incidence in Guangzhou, China for the period 2006-2012 through negative binomial regression. They found that temperature had a positive association and humidity a negative one, *i.e.* less infections with lower temperature and higher humidity. They only considered the temporal resolution and their results might be applicable only in countries with similar weather patterns. Ogawa *et al.* (2002) analyzed the clinical characteristics of scrub typhus in Japan by use of a questionnaire approach involving healthcare workers in 1998, showing that females and people over 51 years old had a higher chance of acquiring scrub typhus.

In recent years, a few studies have examined the spatial or spatio-temporal distribution of scrub typhus. Kuo *et al.* (2011) conducted spatial clustering of scrub typhus incidence using Moran's *I* (Li *et al.*, 2007) and LISA (Anselin, 1995). Wardrop *et al.* (2013)

Correspondence: Jungsoon Choi, Department of Mathematics, Hanyang University, Seongdong-gu, Seoul 133-791, South Korea.  
Tel.: +82.2.2220.2621 - Fax: +82.2.2281.0019.  
E-mail: jungsoonchoi@hanyang.ac.kr

Key words: Scrub typhus; Spatial Kriging; Spatio-temporal model; Zero-inflated Poisson model; Bayesian hierarchical modelling.

See online Appendix for an additional Table.

Contributions: the authors contributed equally.  
This paper is an extended version of master's thesis of Dayun Kang.

Conflict of interest: the authors declare no potential conflict of interest.

Funding: this work was supported by the research fund of the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, and Future Planning (NRF-2015R1C1A1A02037495) and the Ministry of Education (NRF-2018R1D1A1B07047712).

Received for publication: 27 December 2017.  
Revision received: 14 August 2018.  
Accepted for publication: 14 August 2018.

©Copyright D. Kang and J. Choi, 2018  
Licensee PAGEPress, Italy  
Geospatial Health 2018; 13:665  
doi:10.4081/gh.2018.665

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

conducted a spatial analysis using a Poisson regression model with weather covariates. Wu *et al.* (2016) explored the spatio-temporal patterns of scrub typhus incidence to detect hotspots using clustering methods, while Noh *et al.* (2013) analyzed the scrub typhus incidence dataset in Korea by considering the spatio-temporal dependency structures within a Bayesian framework. However, neither Wu *et al.* (2016) nor Noh *et al.* (2013) considered the possible risk factors, which are particularly important with reference to policy decision-making. In addition, although Noh *et al.* (2013) considered space and time, they only took into account a single spatial dependency structure and a single temporal dependency structure over the entire domain investigated. A more versatile approach is needed as the spatial and temporal patterns of scrub typhus incidence could vary across space and time. For example, the spatial distribution of scrub typhus incidence this year might not be the same as the one in the past. In such cases, it is important that the interaction of space and time be considered in modelling in the statistical analysis in order to avoid distorted, even wrong results.

In this paper, we discuss the analysis of monthly scrub typhus incidence data for all administrative districts of South Korea, while also considering the complicated spatio-temporal dependency structures. To the best of our knowledge, this is the first study to adopt a spatio-temporal zero-inflated model for scrub typhus data.

## Materials and Methods

We used meteorological and socioeconomic factors as covariates and propose a Bayesian hierarchical model for the building of flexible spatio-temporal structures by combining prior knowledge

with the data at hand. We examined whether such a space-time interaction structure should be adopted in analyzing the data along with the overall spatial and temporal dependency structures. In South Korea, most of the scrub typhus incidence is concentrated in the south-western regions of the country and in the autumn season because of harvest and increased outdoor activities. Taking the whole country into account, most of the monthly incidence data had zero counts, which leads to over-dispersion. Therefore, we used a zero-inflated Poisson (ZIP) distribution (Lambert, 1992) to account for such data distribution characteristics. Since the meteorological data as covariates are gathered from monitoring stations and the scrub typhus incidence data are collected based on administrative area, they have different spatial data resolutions, which is often called spatial misalignment (Gotway and Young, 2002). To overcome this problem, we applied a 2-stage framework (Choi *et al.*, 2009). At the first stage, we obtained weather estimates for all administrative districts through a spatial weather model, to be used as inputs for the next stage. At the second stage, we applied the spatio-temporal ZIP model to the scrub typhus data to investigate the associations between the meteorological factors and scrub typhus incidence. Finally, the performance of the proposed model is compared to competing models.

## Study region and data

We used monthly datasets in South Korea from 2010 to 2014 covering 251 administrative districts and 60 months. The basic characteristics of all variables are shown in Table 1. The monthly scrub typhus dataset obtained from the Korea Centers for Disease Control and Prevention (<http://is.cdc.go.kr/dstat>) contains the number of patients diagnosed with scrub typhus per month in each administrative area. The zero count of this dataset is about 73%,

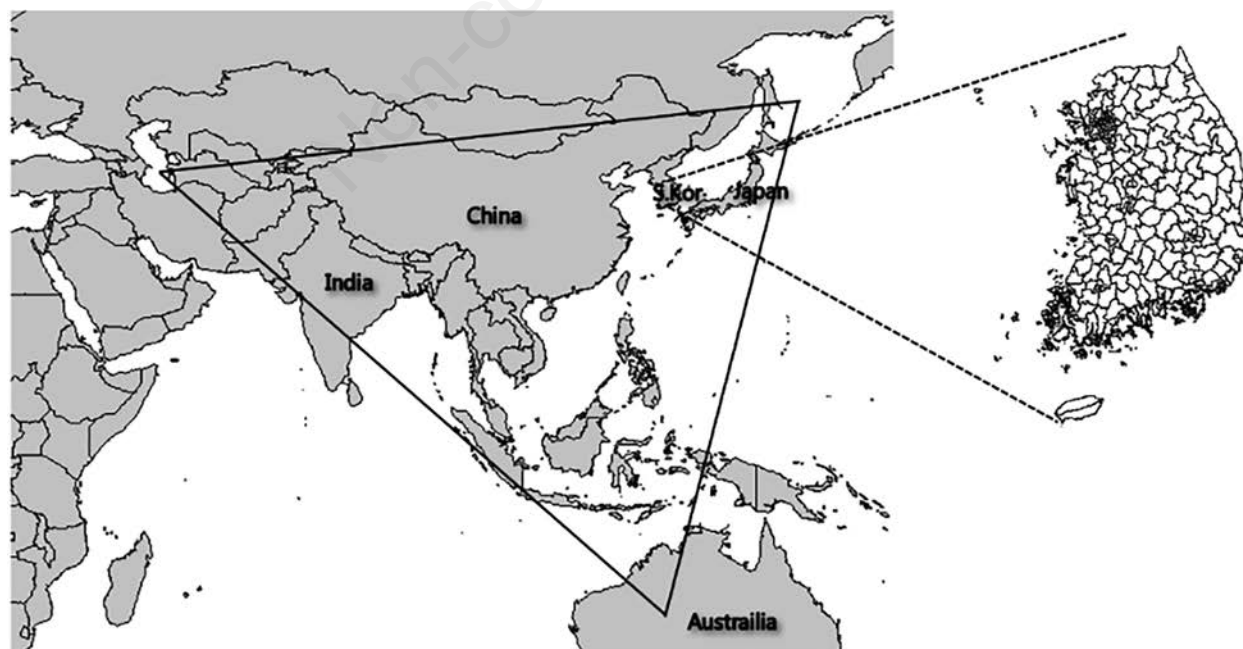


Figure 1. Map of Tsutsugamushi triangle.

showing that our data are highly zero-inflated. Due to this, we summarized the incidence data with and without zero counts in Table 1. The daily precipitation, temperature, and humidity datasets were obtained from the Korea Meteorological Administration (<http://data.kma.go.kr>). Precipitation and temperature data were collected from 487 monitoring stations (Figure 2A), and humidity from 95 monitoring stations (Figure 2B). Monthly averaged values were used for the analysis. Because the number of scrub typhus cases is related to the population and the proportion of elderly people (age 65 and over), we also considered these factors as an offset and a covariate. These datasets were obtained from the Korean Statistical Information Service (<http://kosis.kr>). In the Korean Government system, the total population dataset is collected monthly, while the elder population dataset is collected on an annual basis. Thus, the monthly variation of the proportion of elderly people could be inferred.

### Statistical modelling

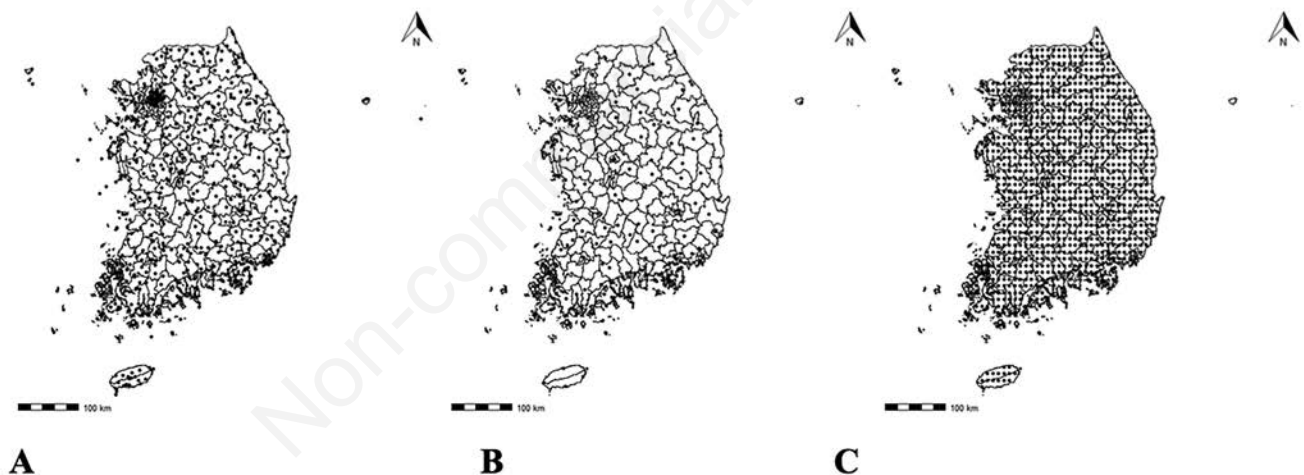
We proposed a 2-stage hierarchical framework to overcome the different spatial data resolutions. At the first stage, we predicted weather values for all administrative districts using a spatial model in which projected coordinates of longitude, latitude and weather data are covariates. At the second stage, we fitted a

Bayesian spatio-temporal zero-inflated model to the incidence data, using the predicted weather values and the proportion of elderly people as covariates. The detailed framework is shown in Figure 3.

Stage one: spatial modelling for meteorological data. We assumed that the spatial model for each weather data is as follows:

$$W(s,t) = Z(s,t)\gamma + \Psi(s,t), \quad \Psi \sim N(0, \Sigma), \quad \text{Eq. 1}$$

where  $W(s,t)$  is the observed weather value at monitoring station  $s$  and time  $t$ , and  $Z(s,t)$  the vector of covariates with the corresponding coefficient vector  $\gamma$ . The vector  $\Psi$  with the element  $\Psi(s,t)$  explains the spatial effects and measurement error with the covariance matrix  $\Sigma$ , in which a Matern spatial covariance structure (Banerjee *et al.*, 2014) provides the best prediction performance. Based on an exploratory data analysis, we used projected coordinates as covariates for precipitation data. Additional covariates included precipitation for temperature data, and temperature as well as precipitation for humidity data. The parameters were assumed to follow non-informative prior distributions to make the most use of the data at hand and to avoid bias. We obtained the predicted weather values at each time and location of interest, *i.e.* a Kriging (Banerjee *et al.*, 2014) approach.



**Figure 2. Map of monitoring stations and kriging locations for the weather model. (A) Map of temperature and precipitation stations; (B) Map of humidity stations; (C) Map of kriging locations.**

**Table 1. Summary of variables.**

Variables	Description	Mean	Standard deviation	Q1	Q3	IQR
Scrub typhus incidence ( $y_{it}$ )	Monthly incidence in each administrative area	2.5	8.7	0	1	1
Non-zero incidence ( $y_{it}$ )	Monthly non-zero incidence in each administrative area	9.3	14.8	1	11	10
Precipitation ( $W_{1it}$ )	Monthly average precipitation (mm)	3.4	3.7	1.0	4.2	3.2
Temperature ( $W_{2it}$ )	Monthly average temperature (°C)	12.3	9.7	3.8	21.5	17.7
Humidity ( $W_{3it}$ ) relative humidity (%)	Monthly average	67.2	9.2	60.0	74.7	14.7
Proportion of elderly people ( $X_{it}$ )	Proportion of people older than 65 years (%)	16.0	7.5	9.9	21.8	11.9

Q1, first quartile; Q3, third quartile; IQR, Q3-Q1.



Here, we estimated the true weather values at about 1,000 locations for each time point in Figure 2C. The estimated weather value at administrative area  $i$  in month  $t$  was averaged by the estimates of true weather values within area  $i$  and used as inputs for the second stage. Stage two: zero-inflated spatio-temporal modelling for scrub typhus incidence data. The incidence of scrub typhus for administrative area  $i$  and month  $t$ ,  $y_{it}$  follows a zero-inflated Poisson distribution:

$$y_{it} = \begin{cases} 0 & \text{with probability } p_{it} \\ \text{Poisson}(\mu_{it} = \theta_{it} N_{it}) & \text{with probability } (1 - p_{it}) \end{cases}, \quad t = 1, \dots, I, \quad t = 1, \dots, T$$

$$\text{logit}(p_{it}) = \alpha_0 + \alpha_1 W_{1it} + \alpha_2 W_{2it} + \alpha_3 W_{3it} + \alpha_4 X_{it} + \log(N_{it}),$$

Eq. 2

where  $p_{it}$  is the probability of structural zeros and  $N_{it}$  the population. The  $\text{logit}(p_{it})$  is the linear combination of precipitation  $W_{1it}$ , temperature  $W_{2it}$ , humidity  $W_{3it}$ , and the proportion of elderly people  $X_{it}$ . The corresponding coefficients  $\alpha_j, j = 0, 1, \dots, 4$  indicate the effects of the covariates. The log relative risk  $\log(\theta_{it})$  was modelled with fixed effects and space-time random effects:

$$\log(\theta_{it}) = \beta_0 + \beta_1 W_{1it} + \beta_2 W_{2it} + \beta_3 W_{3it} + \beta_4 X_{it} + u_i + l_t + v_i + k_t + \phi_{it},$$

Eq. 3

where the random effects  $u_i \sim N(0, \sigma_u^2)$  and  $l_t \sim N(0, \sigma_l^2)$  are the spatially and temporally unstructured terms, respectively. The spatially correlated random effect  $v_i$  follows a conditional autoregressive (CAR) model (Besag, 1974),  $v = (v_1, v_2, \dots, v_i)^T \sim \text{CAR}(\sigma_v^2)$ . Generally speaking, a CAR model is constructed based on neighbourhood information. The mean of a specific area is defined as the weighted average of its neighbours, and the variance is inversely proportion-

al to the number of neighbours. The temporally correlated random term  $k_t$  follows a first-order autoregressive AR(1) process (Yule, 1921). Knorr-Held (2000) proposed four different types of the spatio-temporal interaction term  $\phi_{it}$  and its covariance can be expressed as  $\Sigma_\phi = \Sigma_s \otimes \Sigma_T$  using the Kronecker product, where the matrices  $\Sigma_s$  and  $\Sigma_T$  indicate covariance matrices of space and time, respectively. In our data, we found that the temporal pattern for each area was not identical to that of other areas and the spatial pattern varied from year to year, supporting the fact that considering a space-time interaction term is reasonable. We used non-informative priors for the parameters: *Normal*  $(0, 10^5)$  for the coefficients  $\beta_j$  and  $\alpha_j, j = 0, 1, \dots, 4$ , and *U*  $(0, 100)$  for the standard deviations  $\sigma_u, \sigma_l, \sigma_v, \sigma_k$ , and  $\sigma_\phi$ .

The WinBUGS statistical package (<http://www.mrc-bsu.cam.ac.uk/software/bugs>) was used. Two chains with different initial values were used to check the sample convergence. Every 50th sample was extracted as a posterior sample. After the burn-in, 2,500 samples for each chain, in total of 5,000 samples, were used for parameter estimation. We checked the convergence using trace plots, the Gelman-Rubin statistic (Gelman *et al.*, 1992), and autocorrelation plots. All outcome Figures in this paper were produced with open source program R (<https://www.r-project.org>).

We additionally considered seven competing models. All models (models 1-8) are listed in the Appendix. Model 1 to Model 4 are Poisson models, and Model 5 to Model 8 are ZIP models. Models 1 and 5 only consider covariates. Models 2 and 6 additionally contain spatially and temporally uncorrelated terms. Spatially and temporally correlated random terms were added in models 3 and 7. Finally, in models 4 and 8, spatio-temporal interaction term was considered additionally. We investigated the performance of the proposed model (model 8) and other competing models (models 1-

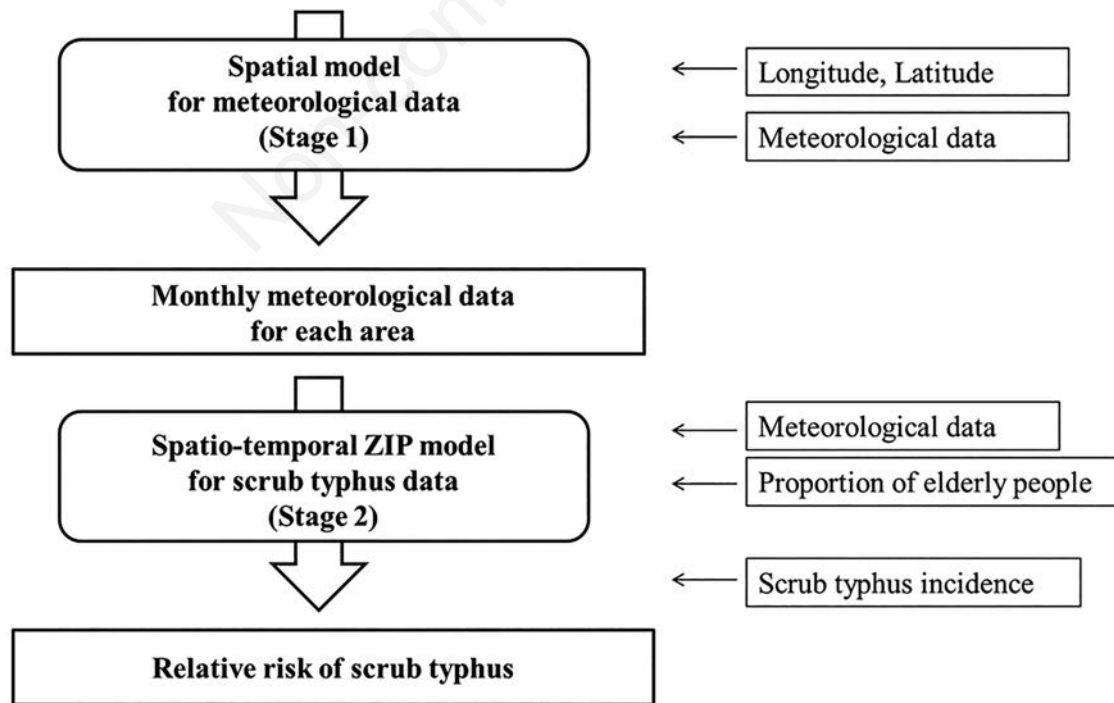


Figure 3. Flowchart of the two-stage model. ZIP, zero-inflated Poisson.

7) in terms of deviance information criterion (DIC) and mean squared prediction error (MSPE). A model with a smaller MSPE and DIC value has better performance.

## Results

### Weather data results

To examine the prediction performance of the proposed spatial

model, we compared the values observed at the monitoring stations and the predicted values for the administrative district in which each station is located. We chose three administrative areas that contain weather monitoring stations: Inje-gun in Gangwon Province, Youngdong-gun in Chungcheong Province and Mungyeong-si in Gyeongsangbuk Province. Figure 4 shows that most points are located close to the line  $y = x$ , indicating that the predicted weather values are similar to the observations from the stations. We also found that the observed and predicted values for the other areas were similar. Thus, the proposed spatial weather model fits the data well.

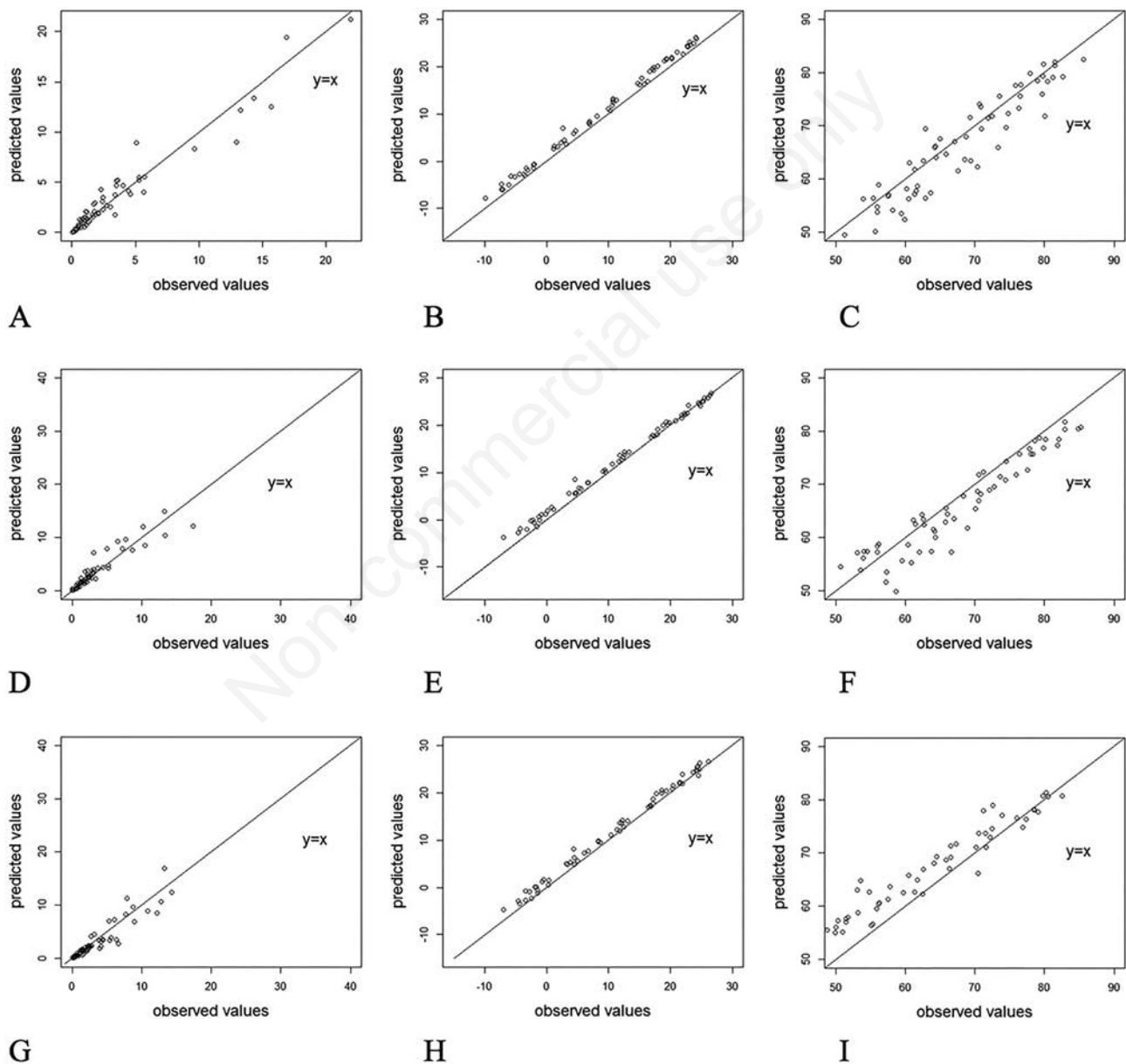


Figure 4. Calibration plots of the weather data model for Inje-gun in Gangwon Province (first row), Youngdong-gun in Chungcheong Province (second row), and Mungyeong-si in Gyeongsangbuk Province (third row). (A, D, G) precipitation; (B, E, H) temperature; (C, F, I) humidity.



### Scrub typhus data results

Table 2 summarizes the model performances. The MSPE values of models 4 and 8 decreased to about one-sixtieth and one-eightieth of those of models 1 and 5, respectively. The DIC values also decreased dramatically for the spatio-temporal models compared with the non-spatio-temporal ones. Moreover, the space-time interaction terms not only provide smaller DIC values but also smaller MSPE values. Overall, the Poisson models (models 1-4) have larger MSPE and DIC values than the ZIP models (models 5-8). As more complicated spatio-temporal structures are included in a Poisson model, its performance becomes more similar to the performance of the ZIP model with the corresponding spatio-temporal dependency structure. Therefore, space-time random components explain the over-dispersion in Poisson regression models. Since model 8 works the best of the eight models in terms of MSPE and DIC values, the ZIP model with the space-time interaction term was deemed more suitable for our data than the other models.

We also compared the empirical probability of zero counts from the real data with the estimated probability from the models. Around 73% of the incidence had zero values. In Table 2, the estimate of the probability of zero counts was 0.157 in the simple Poisson model (model 1), but 0.739 in model 8, which is almost the same as the observed probability of zero. Therefore, using a spatio-temporal ZIP model significantly improves the ability to capture zero-inflation. The other models (models 2-7) had similar values because the space-time random terms explain most of the zero-inflation.

**Table 2. Model performance.**

Distribution	Model	Mean square prediction error	Deviance	pD	Deviance information criterion	Est.Pr (Y=0)
Real data						0.730
Poisson	Model 1	75.07	160938	5.02	160943	0.157
	Model 2	10.65	29633	306.58	29939	0.739
	Model 3	10.65	29640	295.16	29935	0.740
	Model 4	0.96	22640	2122.03	24762	0.739
Zero-inflated Poisson	Model 5	61.24	76150	4.38	76154	0.731
	Model 6	10.07	28330	357.53	28688	0.740
	Model 7	10.15	28410	366.04	28776	0.710
	Model 8	0.97	22320	2025.35	24345	0.739

Est.Pr (Y=0): the estimated probability of zero counts.

**Table 3. Posterior summaries for the spatio-temporal zero-inflated Poisson model.**

	Est.	Standard deviation	Monte Carlo error	2.50%	Median	97.50%	Relative risk
Intercept ( $\beta_0$ )	-6.883	0.432	0.043	-7.673	-6.714	-6.285	0.001
Precipitation ( $\beta_1$ )	0.023	0.017	0.001	-0.008	0.022	0.055	1.023
Temperature ( $\beta_2$ )	-0.016	0.011	0.001	-0.041	-0.017	0.007	0.984
Humidity ( $\beta_3$ )	-0.098	0.007	0.001	-0.110	-0.099	-0.085	0.907
Elderly people proportion ( $\beta_4$ )	0.066	0.006	0.001	0.054	0.066	0.075	1.068
$\sigma_u$	0.418	0.030	0.002	0.365	0.416	0.482	
$\sigma_1$	0.263	0.163	0.016	0.091	0.211	0.726	
$\sigma_v$	0.770	0.058	0.004	0.666	0.767	0.892	
$\sigma_k$	1.525	0.149	0.005	1.262	1.514	1.852	
$\sigma_\phi$	0.835	0.020	0.002	0.800	0.833	0.871	

Est.: posterior mean; 2.50% and 97.50%: lower limit and upper limit of 95% credible interval, respectively.

The parameter estimates of the best model, model 8, are shown in Table 3. Only the coefficients of humidity and proportion of elderly people are statistically significant since the 95% credible intervals did not contain zero. The regression coefficient of humidity was negative and that of proportion of elderly people positive. The estimated coefficients of precipitation and temperature were positive and negative, respectively, but not statistically significant.

We compared the observed values with the predicted values from model 8. In Figure 5, the observed incidence of scrub typhus and the predicted values are located with the regression line  $y = 0.94x + 0.15$ . Figure 6 presents the time series plots of the observed data and predicted values for two selected areas. Here, Gwanak-gu in Seoul City, and Ulju-gun in Ulsan City have the highest incidence within Seoul and South Korea, respectively. Figure 7 presents the observed and predicted maps of the incidence in October 2013 and October 2014. These comparison methods show that the predicted values were similar to the observed data.

### Discussion

Investigating the relationship between weather factors and scrub typhus has led to the result that humidity is a significant risk factor, but a negative one. We found that the number of the scrub typhus cases increases as humidity decreases. This negative association can be explained by the fact that the autumn season is rel-

atively dry and the incidence is mostly centred at that time. This result is parallel to the negative correlation of relative humidity and scrub typhus incidence shown by Li *et al.* (2014) and Wu *et al.* (2016). Since both these studies and our own made use of the scrub typhus incidence during all seasons, the effect of humidity on the disease might be different if the incidence data were restricted to the autumn season.

In addition, we showed that the higher the proportion of elderly people is, the more scrub typhus occurs, which is supported by Ogawa *et al.* (2002). Since scrub typhus commonly occurs in farmland and farm workers are mostly aged over 60 in South Korea, this result seems to be reasonable. Also, there is a high chance that older people have a less vigorous immune system and therefore are more at risk of scrub typhus infection than young people.

A negative binomial zero-inflated spatio-temporal model as an alternative for our data can be considered, but as it has larger DIC and MSPE values (DIC = 25553 and MSPE = 6.44) than the proposed Poisson zero-inflated spatio-temporal model, the latter would then be better in terms of model performance.

Since most of the hotspots are in rural areas, interventions specified for those areas can effectively prevent scrub typhus. A high proportion of the residents in rural areas are senior citizens who are likely to lack information on scrub typhus. Therefore, a key approach would be to provide education to all residents in the endemic areas before peak season. As an example, Koryung County, South Korea, effectively prevented the disease by educating its residents, especially the elderly. People who had experienced scrub typhus were invited as guest speakers and as soon as the first case of the disease occurred, information went out. In addition, the government of Koryung County distributed tick repellent and protective clothing to the residents. In doing so, the incidence of scrub typhus in Koryung County decreased compared to previous years. The prevention policies should especially be focused on the autumn season due to ensuing harvest and increased outdoor activities.

All models in this study used adopted Bayesian methods. In spite of a high computational cost, they have advantages over frequentist methods. Unlike the difficult interpretation of confidence

intervals in frequentist inference, credible intervals in Bayesian inference are more straightforward and easy to interpret. Especially in spatial modelling, the Bayesian framework enables understanding based on hierarchical models highly intuitive. Combining prior knowledge with real-world data is another benefit of Bayesian inference. Here, careful selection of appropriate priors is required and we used non-informative priors. To understand how the prior distributions influence the results, we conducted a sensitivity analysis using inverse gamma distributions for variances. These prior distributions provided almost similar results.

We had a minor support problem in using the weather data as covariates in this study. As a solution, we used a two-stage model which can offer location datasets without monitoring stations using a relatively small number of observed data. Owing to this strength, statistical analysis can be conducted with the complete covariates and find the significant risk factors. Based on these results, we were able to prevent and deal with the disease effectively.

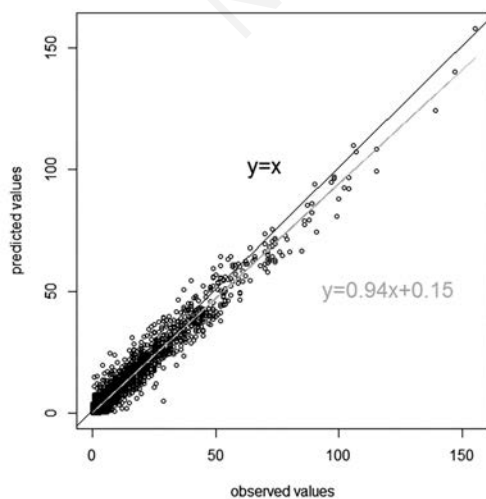


Figure 5. Calibration plot of the scrub typhus data model (Model 8).

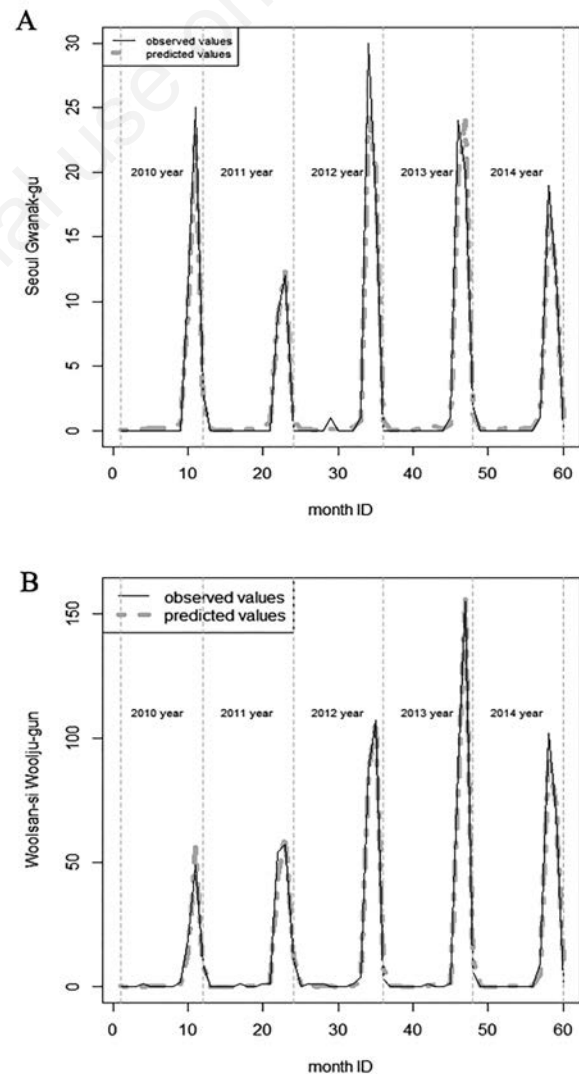
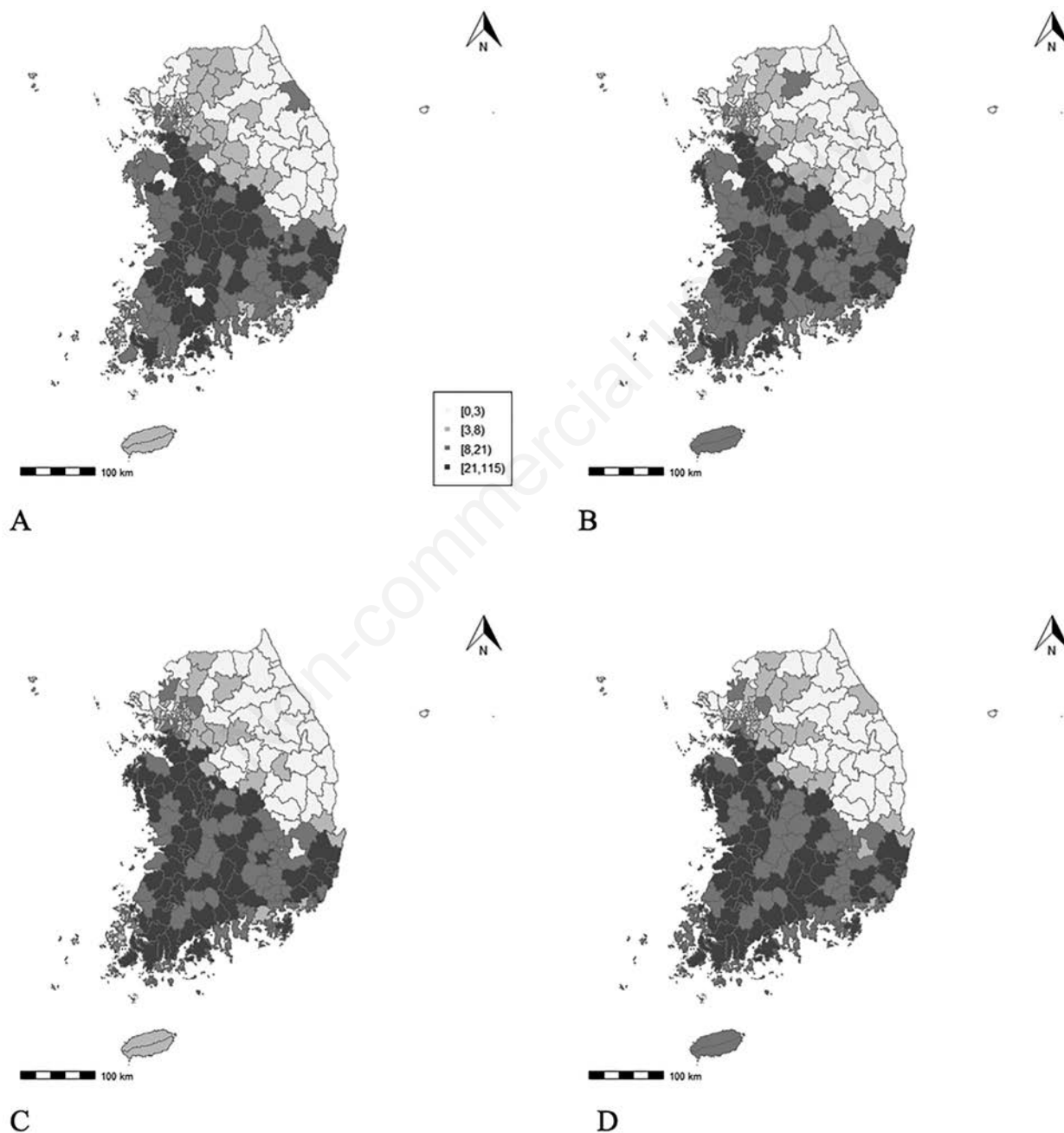


Figure 6. Comparison of scrub typhus incidence and the predicted values for two representative regions. (A) Gwanak-gu in Seoul City; (B) Ulju-gun in Ulsan City.

Several further tasks remain to be done. First, the model for weather data in the first stage is limited to the spatial model in this study. Adopting a spatio-temporal model for meteorological data might improve the predictive performance. Also, combining weather observation values and predicted values from numerical models might enhance the predictive performance. Second, we expect to be able to analyze the data using sex- and age-adjusted individual patient data in the future, but we were unable to obtain

this information with reference to the people diagnosed with scrub typhus in this study. Third, because scrub typhus occurs mostly in the autumn, analyzing only autumn data but on a daily basis might help locate detailed trends. In addition, conducting a spatio-temporal clustering, might be helpful in deriving interventions for each season and could lead to a simulation study to investigate the effects of the interventions.



**Figure 7.** Predicted and observed maps of scrub typhus incidence. (A) The incidence of scrub typhus, October 2013; (B) the predicted values, October 2013; (C) the incidence of scrub typhus, October 2014; (D) the predicted values, October 2014.



## Conclusions

This study is the first attempt to use a Bayesian spatio-temporal ZIP model for the association between the incidence of scrub typhus in Korea and the weather and proportion of people older than 65 years. Our spatio-temporal model dramatically increased the performance. This supports that spatio-temporal models should be applied for the data with spatio-temporal dynamics. Given that many epidemiological data contain spatial and temporal dependencies, our model could be a template for the use of spatio-temporal models with epidemiological data.

## References

- Anselin L, 1995. Local Indicators of Spatial Association – LISA. *Geogr Anal* 2:93-115.
- Banerjee S, Carlin BP, Gelfand AE, 2014. Hierarchical modelling and analysis for spatial data. Boca Raton, FL: CRC Press.
- Besag J, 1974. Spatial interaction and the statistical analysis of lattice systems. *J Roy Stat Soc B Met* 36:192-236.
- Choi J, Fuentes M, Reich BJ, 2009. Spatial-temporal association between fine particulate matter and daily mortality. *Comput. Statist. Data Anal* 53:2989-3000.
- Cracco C, Delafosse C, Baril L, Lefort Y, Morelot C, Derenne JP, Bricaire F, Similowski T, 2000. Multiple organ failure complicating probable scrub typhus. *Clin Infect Dis* 31:191-2.
- Gotway C, Young L, 2000. Combining incompatible spatial data. *J Am Statist Assoc* 97:632-48.
- Knorr-Held L, 2000. Bayesian modelling of inseparable space-time variation in disease risk. Collaborative Research Center 386 Discussion Paper 147. *Stat Med* 19:2555-67.
- Kuo CC, Huang JL, Ko CY, Lee PF, Wang HC, 2011. Spatial analysis of scrub typhus infection and its association with environmental and socioeconomic factors in Taiwan. *Acta Tropica* 120:52-8.
- Lambert D, 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34:1-14.
- Li H, Calder CA, Cressie N, 2007. Beyond Moran's I: Testing for spatial dependence based on the spatial autoregressive model. *Geogr Anal* 39:357-75.
- Li T, Yang Z, Dong Z, Wang M, 2014. Meteorological factors and risk of scrub typhus in Guangzhou, southern China, 2006-2012. *BMC Infect Dis* 14:139.
- McCrumb FR, Stockard JL, Robinson CR, Turner LH, Levis DG, Maisey CW, Kellecher MF, Gleiser CA, Smadel JE, 1957. Leptospirosis in Malaya. *Am J Trop Med Hyg* 6:238-56.
- Noh M, Lee Y, Chu C, Gwack J, Youn SK, Huh S, 2013. Are there spatial and temporal correlations in the incidence distribution of scrub typhus in Korea? *Osong Public Health Res Perspect* 4:39-44.
- Ogawa M, Hagiwara T, Kishimoto T, Shiga S, Yoshida Y, Furuya Y, Kaiho I, Ito T, Nemoto H, Yamamoto N, Masukawa K, 2002. Scrub typhus in Japan: Epidemiology and clinical features of cases reported in 1998. *Am J Trop Med Hyg* 67:162-5.
- Tsai PJ, Yeh HC, 2013. Scrub typhus islands in the Taiwan area and the association between scrub typhus disease and forest land use and farmer population density: geographically weighted regression. *BMC Infect Dis* 13:191.
- Wardrop NA, Kuo CC, Wang HC, Clements AC, Lee PF, Atkinson PM, 2013. Bayesian spatial modelling and the significance of agricultural land use to scrub typhus infection in Taiwan. *Geospat Health* 8:229-39.
- Wu YC, Qian Q, Magalhaes RJS, Han ZH, Hu WB, Haque U, Weppelmann TA, Wang Y, Lui YX, Li HL, Sun HL, Sun YS, Clements AC, Li SH, Zhang WY, 2016. Spatiotemporal dynamics of scrub typhus transmission in mainland China, 2006-2014. *PLoS Negl Trop Dis* 10:e0004875.
- Yule GU, 1921. On the time-correlation problem, with especial reference to the variate-difference. *J Roy Stat Soc* 84:497-537.