



HAL
open science

Spatial-temporal patterns and risk factors for human Leptospirosis in Thailand, 2012-2018

Sudarat Chadsuthi, Karine Chalvet-Monfray, Suchada Geawduanglek,
Phrutsamon Wongnak, Julien Cappelle

► **To cite this version:**

Sudarat Chadsuthi, Karine Chalvet-Monfray, Suchada Geawduanglek, Phrutsamon Wongnak, Julien Cappelle. Spatial-temporal patterns and risk factors for human Leptospirosis in Thailand, 2012-2018. Scientific Reports, 2022, 12 (1), pp.1-11. 10.1038/s41598-022-09079-y . hal-03541478

HAL Id: hal-03541478

<https://hal.inrae.fr/hal-03541478>

Submitted on 13 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



OPEN

Spatial–temporal patterns and risk factors for human leptospirosis in Thailand, 2012–2018

Sudarat Chadsuthi^{1✉}, Karine Chalvet-Monfray^{2,3}, Suchada Geawduanglek⁴, Phrutsamon Wongnak^{2,3} & Julien Cappelle^{2,5,6}

Leptospirosis is a globally important zoonotic disease. The disease is particularly important in tropical and subtropical countries. Infections in humans can be caused by exposure to infected animals or contaminated soil or water, which are suitable for *Leptospira*. To explore the cluster area, the Global Moran's I index was calculated for incidences per 100,000 population at the province level during 2012–2018, using the monthly and annual data. The high-risk and low-risk provinces were identified using the local indicators of spatial association (LISA). The risk factors for leptospirosis were evaluated using a generalized linear mixed model (GLMM) with zero-inflation. We also added spatial and temporal correlation terms to take into account the spatial and temporal structures. The Global Moran's I index showed significant positive values. It did not demonstrate a random distribution throughout the period of study. The high-risk provinces were almost all in the lower north-east and south parts of Thailand. For yearly reported cases, the significant risk factors from the final best-fitted model were population density, elevation, and primary rice crop arable areas. Interestingly, our study showed that leptospirosis cases were associated with large areas of rice production but were less prevalent in areas of high rice productivity. For monthly reported cases, the model using temperature range was found to be a better fit than using percentage of flooded area. The significant risk factors from the model using temperature range were temporal correlation, average soil moisture, normalized difference vegetation index, and temperature range. Temperature range, which has strongly negative correlation to percentage of flooded area was a significant risk factor for monthly data. Flood exposure controls should be used to reduce the risk of leptospirosis infection. These results could be used to develop a leptospirosis warning system to support public health organizations in Thailand.

Leptospirosis is a globally important zoonotic disease and causes approximately 1.03 million cases and 58,900 deaths worldwide each year¹. The disease is particularly prevalent in tropical and subtropical climates. Infections in humans and animals can be caused by exposure to infected animals or soil and water contaminated, which are suitable for *Leptospira*² and with urine from infected animals^{3,4}.

Studies have found that the pathogenic *Leptospira* can survive in water and soil for months^{5–7}. Climatic variables could potentially be associated with leptospirosis transmission, such as rainfall, temperature, and floods. The effects of rainfall and temperature on leptospirosis incidences were investigated in Thailand^{8,9}, and at different setting and climatic zones such as in Sri Lanka¹⁰, China¹¹, and Philippines¹². Flooding is one of the important drivers for leptospirosis outbreaks. Leptospirosis outbreaks have been reported to be associated with flooding or flooded areas (e.g., Indonesia¹³, Pakistan¹⁴, Brazil¹⁵, and Laos¹⁶). As such, the risk of leptospirosis infection could be identified using remotely sensed data^{11,17,18}.

In Thailand, the annual human leptospirosis cases are reported as 4.25 cases per 100,000 people, from 2012 to 2018¹⁹. Most reported cases are in agricultural workers, who are the most exposed to contaminated environments. A previous study in Nakhon Ratchasima province of Thailand found that rice cultivators were at risk for leptospirosis²⁰. Rice paddies have been found to be associated with increased leptospirosis incidences^{21,22}. A study in Nan province of Thailand found that the open habitat near rivers or in rice fields that were prone to

¹Department of Physics, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand. ²Université de Lyon, INRAE, VetAgro Sup, UMR EPIA, 69280 Marcy l'Etoile, France. ³Université Clermont Auvergne, INRAE, VetAgro Sup, UMR EPIA, 63122 Saint-Genès-Champanelle, France. ⁴Medical and Graduate Education Division, Faculty of Science, Mahidol University, Bangkok 10400, Thailand. ⁵UMR ASTRE, CIRAD, INRAE, 34398 Montpellier, France. ⁶CIRAD, UMR ASTRE, 34398 Montpellier, France. ✉email: sudaratc@nu.ac.th

flooding, were related to human leptospirosis infections²³. The activities of rice farming are also associated with an increase of leptospirosis infections in other countries, for example, in Iran²⁴ or Tanzania²⁵. Thus far, to our knowledge, there have been few studies on the impact of rice production activities on leptospirosis infections.

Spatial–temporal variation of human leptospirosis incidences have also been investigated to identify high-risk areas, where preventive measure should be implemented^{9,26–28}. The study of leptospirosis spatial–temporal patterns may help to prioritize the prevention and surveillance of leptospirosis in high-risk areas. Also, risk factors should be identified in order to control human leptospirosis.

In this work, we established a large-scale dataset of human leptospirosis in Thailand over seven years from 2012 to 2018. The objectives of this study are (1) to investigate the spatial–temporal patterns of leptospirosis incidence, (2) to determine the high-risk and low-risk provinces across Thailand, and (3) to examine the effect of rice-production-related factors, remotely sensed environmental factors, and climatic factors on leptospirosis risk in humans using a generalized linear mixed model (GLMM).

Materials and Methods

Data collections. Human leptospirosis case reports were retrieved from the database of national disease surveillance (report 506), Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health, Thailand¹⁹. Most positive cases were suspected leptospirosis cases, based on the clinical diagnosis made by attending physicians and were mainly reported from public hospitals. In this research, we analysed all reported cases from 2012 to 2018 at the province level.

The percentage of flooded area was obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) of the Terra satellite (Surface Reflectance 8-Day L3 Global 500 m SIN Grid V005 (MOD09A1)). Using the surface spectral reflectance, band 4 (green) and band 7 (infrared), which measured at ground level in the absence of atmospheric scattering or absorption, the modified normalized difference water index (MNDWI) was calculated^{17,29}. To calculate the percentage of flooded area, the number of pixels for which the MNDWI value was greater than or equal to zero, in each province, for each month and for each year, were counted.

To reflect the vegetation intensity, the normalized difference vegetation index (NDVI) was calculated from the value of red band (wavelength: 620–670 nm) and near-infrared (NIR) band (841–876 nm)³⁰. The vegetation area was classified into two types: low vegetation (such as shrub and grassland) and high vegetation (such as temperate and tropical urban forest). The vegetation type was calculated from the percentage of NDVI pixels that had a value within 0.2–0.5 and 0.5–1, respectively³⁰.

The monthly values of rainfall, minimum and maximum temperature, and soil moisture were obtained from TerraClimate³¹, calculated over a grid of approximately 4 km², and aggregated at the province level. The annual average values were calculated from monthly data. We also added the temperature range (i.e., the average difference between maximum and minimum temperature) to test the model. Due to the relation of temperature range with cloud cover³², we hypothesize that this parameter may be used as a proxy for rainfall and flooding.

Elevation data was derived from the NASA Shuttle Radar Topographic Mission (SRTM) 90 m Digital Elevation Data³³ as it may drive leptospirosis infection²⁷. The average elevation at the province level was used in this analysis. Slope data was calculated using QGIS version 3.6.0 from the elevation data in units of degree.

The human population data was obtained from the WorldPop database (<http://www.worldpop.org>) and summarised as human population density per year for each province. Livestock population density (buffalo, cattle, and pigs) were obtained from the Information and Communication Technology Center (ICT), Department of Livestock Development of Thailand at the district level³⁴.

In this work, we also incorporated rice production data such as rice crop arable area, rice cultivated area and rice yield, which may associate with leptospirosis occurrences. In Thailand, most cases were identified in rice farmers³⁵. Rice farmers typically plant rice in two seasons, i.e., the in-season rice is called the primary rice crop and the off-season rice is called the secondary rice crop. The primary and secondary rice season data was obtained from the Office of Agricultural Economics, Thailand (<http://www.oae.go.th/>). The descriptions of the variables are presented in Table 1.

Spatial–temporal autocorrelation. To determine the spatial autocorrelation, the annual incidence cases and monthly incidence cases per 100,000 population were used to calculate the Global Moran's I index³⁶ at the province level. The Moran's I value, which ranges from -1 to $+1$, is used to classify the spatial clustering. A Moran's I value of more than zero indicates positive autocorrelation, and a value of less than zero indicates negative autocorrelation. When the Moran's I value is close to zero, it indicates a random distribution. The value is calculated from the spatial weight matrix, which is assigned from the nearest neighbour province. To test the significant level, we used a Monte-Carlo simulation method with 999 permutations. A significance of less than 0.05 is used to consider the cluster or no autocorrelation. We used R package 'spdep' to calculate the Moran's I value³⁷.

We also determined the local indicators of spatial association (LISA) at the province level. LISA was used to describe leptospirosis clusters with hot spots (High–High), cold spots (Low–Low) and spatial outliers (High–Low and Low–High)³⁸. Hot spots or High–High provinces indicate provinces that had a high leptospirosis incidence rate and were surrounded by provinces with high incidence rates. Cold spots represent provinces with low incidences surrounded by other low incidence provinces. Spatial outliers represent high incidence provinces surrounded by low incidence provinces (High–Low) and low incidence provinces surrounded by high incidence provinces (Low–High). LISA was calculated using GeoDA version 1.18 software (<https://geodacenter.github.io/>)³⁹. The spatial weights file was created based on an inverse distance function with a distance band of 150 km. We considered a p-value less than 0.05 as a statistically significant result using 999 permutations.

Variable	Description, (unit)	Reference
Temporal correlation	The current number of case reports per population of the province compared with the previous year or month	http://www.boe.moph.go.th/boedb/surdata/disease.php?ds=43
Spatial correlation	Correlation of exponential decay of distance between centroids	
Population density	Annual human population per area at each province (population/km ²) Average human population per area at each province across 7 years (population/km ²)	http://www.worldpop.org
Cattle density	Annual cattle population per area at each province (population/km ²) Average cattle population per area at each province across 7 years (population/km ²)	http://ict.dld.go.th
Pig density	Annual pig population per area at each province (population/km ²) Average pig population per area at each province across 7 years (population/km ²)	http://ict.dld.go.th
Buffalo density	Annual buffalo population per area at each province (population/km ²) Average buffalo population per area at each province across 7 years (population/km ²)	http://ict.dld.go.th
Soil moisture	Annual average of monthly soil moisture at each province Monthly soil moisture at each province, averaged across 7 years (mm)	http://www.climatologylab.org/terraclimate.html
Precipitation	Annual average of monthly precipitation at each province Monthly precipitation at each province, averaged across 7 years (mm)	http://www.climatologylab.org/terraclimate.html
Minimum temperature	Annual average of monthly minimum temperature at each province Monthly minimum temperature at each province, averaged across 7 years (°C)	http://www.climatologylab.org/terraclimate.html
Maximum temperature	Annual average of monthly maximum temperature at each province Monthly average maximum temperature at each province, averaged across 7 years (°C)	http://www.climatologylab.org/terraclimate.html
Temperature range	Difference between maximum annual average temperature and minimum annual average temperature at each province Difference between maximum monthly average temperature and minimum monthly average temperature at each province (°C)	http://www.climatologylab.org/terraclimate.html
Percentage of flooded area	The percentage of pixels with an MNDWI value greater than or equal to zero (250 m, 8 days) for each month and for each year	MOD09A1
NDVI-1	The percentage of NDVI pixels that are within 0.2–0.5 (250 m, 8 days) for each month and for each year	MOD09A1
NDVI-2	The percentage of NDVI pixels that are within 0.5–1.0 (250 m, 8 days) for each month and for each year	MOD09A1
Elevation	Average elevation of each province (90 m spatial resolution)	https://srtm.csi.cgiar.org/
Slope	Average angle of inclination at each province (90 m spatial resolution)	Calculated from elevation
Primary rice crop arable area	Annual arable area for primary rice crop at each province (km ²)	http://www.oae.go.th/
Primary rice cultivated area	Annual cultivated area for primary rice crop at each province (km ²)	http://www.oae.go.th/
Primary rice yield	Annual primary rice yield of each province (1000 kg/km ²)	http://www.oae.go.th/
Secondary rice crop arable area	Annual arable area for secondary rice crop at each province (km ²)	http://www.oae.go.th/
Second rice cultivated area	Annual cultivated area for secondary rice crop at the province level (km ²)	http://www.oae.go.th/
Secondary rice yield	Annual secondary rice yield of each province (1000 kg/km ²)	http://www.oae.go.th/

Table 1. Variables used to identify risk of infection from GLMM and their descriptions.

Statistical model. To study the spatial and temporal distribution of leptospirosis, we used a generalized linear mixed model (GLMM). Mixed effect models were used to account for the spatial dependence using province ID and for the temporal dependence, using time as random effect variables. The response variable of this model was the number of reported cases. Due to the data containing many zeros, the model would underpredict the cases. In this work, we used a zero-inflated negative binomial distribution (ZINB), which is an extension of the Poisson regression and captures the excess number of zeros^{40,41}. To fit ZINB, we used the R package ‘glm-mTMB’⁴². We also added spatial and temporal correlation terms to consider the spatial and temporal structures (Table 1).

All variables were scaled by subtracting their mean and dividing by the standard deviation before investigating the effect of factors. We first selected the variables by studying the cross-correlation. For pairs of variables that have highly correlated, (more than |0.5|), only one variable was selected based on the minimum AIC value of the univariate model. Then, multivariate models, including all selected variables, were examined using the dredge function from the R package ‘MuMIn’⁴³. This function generates the 2^N model table with subsets of fixed effect terms, where N is the number of variables. The best models are sorted based on the criterion (AIC). The final model was the model with the least number of variables in the set of models with a $\Delta AIC < 2$ in comparison with the model of the lowest AIC.

Ethics statement. All information of leptospirosis surveillance data were collected from the Thai Ministry of Public Health. This study was approved by the Institutional Review Board of Naresuan University (P10003/64). The need of informed consent was waived by the Institutional Review Board of Naresuan University as all data of our study are deidentified. The analyses were performed at aggregate level (province level) and no confidential

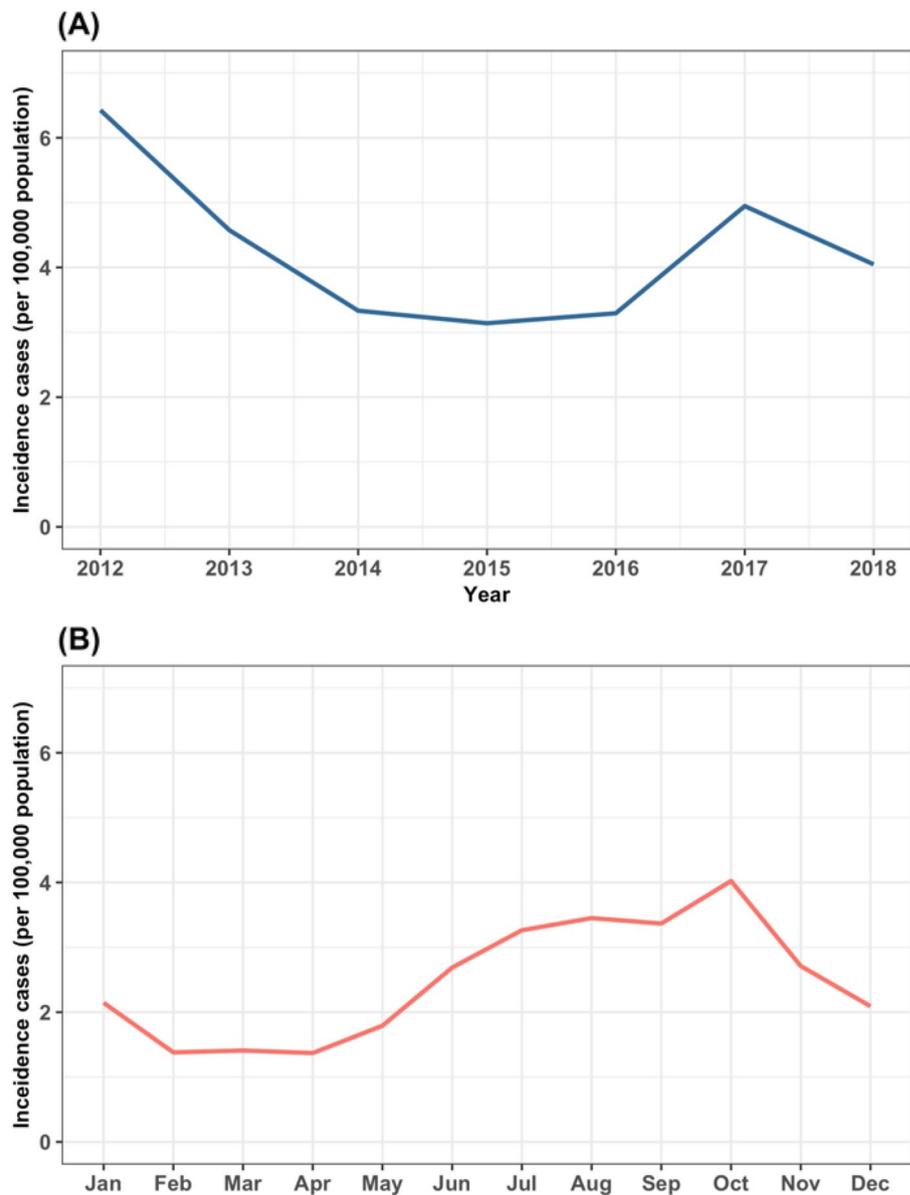


Figure 1. Annual (A) and monthly (B) leptospirosis incidence rates in Thailand (per 100,000 population).

information was involved. All methods were performed in accordance with the relevant guidelines and regulations. The location and date of onset of illness were also gathered. The maps presented in this paper do not identify patients' addresses.

Software. For the maps and for data analysis, we used R program version 4.1.0⁴⁴ with package tidyverse 1.3.0⁴⁵, stringr 1.4.0⁴⁶, spdep 1.1–5⁴⁷, sf 0.9–6⁴⁸, and ggpubr 0.4.0⁴⁹. The cross-correlation plots were created using R package ggcorrplot 0.1.3⁵⁰. For the heat map created using Microsoft Excel version 16.57 (<https://www.microsoft.com/th-th/microsoft-365/excel>).

Results

Descriptive results. From 2012 to 2018, there were 20,459 reported leptospirosis cases in Thailand. The annual incidence per 100,000 population (incidence rate) is plotted in Fig. 1. The incidence rate ranged from 3.1 in 2015 to 6.43 in 2012. After 2012, the incidence rate declined till 2014, where it levelled off, then slightly increased after 2016. In addition, the monthly incidence rate was also analysed (Fig. 1). We found that the incidence rates were higher, than the average of 2.47, between June and November, with the highest value in October. Spatial variation was observed in the northern and southern parts Thailand, while the incidence rate in the north-eastern part was observed all year round (Figs. S1 and S2). The incidence rate in the central part was found to be lower than other regions.

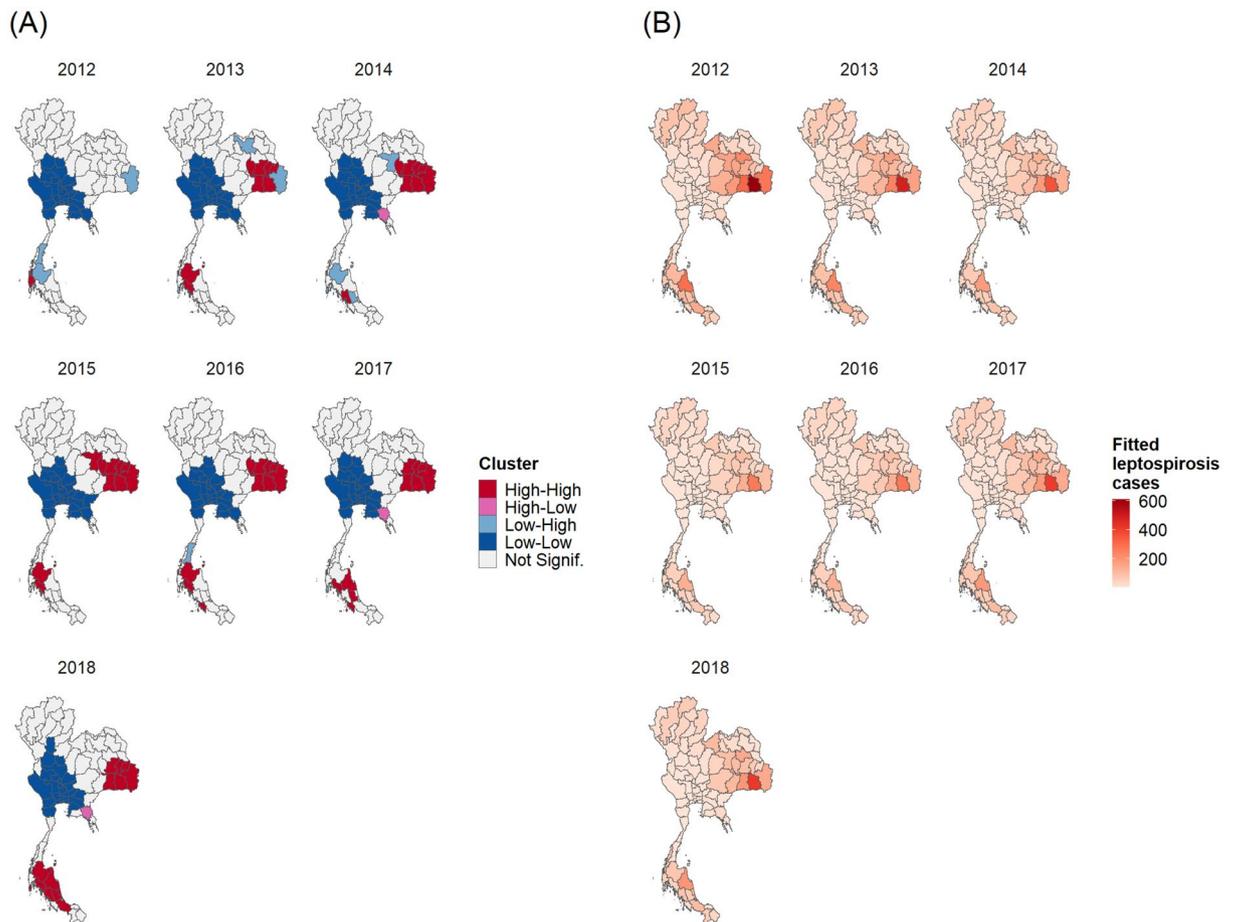


Figure 2. The plots of annual spatial clustering of incidence rate as determined by LISA (A) and annual leptospirosis cases estimated by the final GLMM (B). Maps created using R Program version 4.0.3 (<https://www.r-project.org/>).

Spatial-temporal analysis. To investigate the spatial autocorrelation, the Global Moran's I values were calculated (Tables S1 and S2). The results showed significant positive values, demonstrating a non-random distribution throughout the period studied, for both annual and monthly incidence rates. To identify high-risk provinces, results of the analysis of local indicators of spatial association were mapped in Figs. 2A and 3A. For the annual data, the LISA maps show that high-high clusters appeared in the lower north-east, except in 2012, the year with the lowest Moran's I value. In the south, the high-risk provinces display different patterns. For monthly data, we found high-risk provinces in north-east during rainy season correspond to Fig. 4. The high-risk areas were observed in the south-west correspond to high rainfall (Fig. 4). However, for south-east, the high-risk areas were observed all months except in September and October. Low-risk provinces were mainly detected in central Thailand for both annual and monthly data. Whereas the LISA results of the north showed that there were no significant clusters.

Risk factors. To select explanatory variables in the model for annual reported data, we calculated the cross-correlation of the scaled annual variables (Fig. S3). Because precipitation, percentage of flooded area, temperature range, and elevation were highly correlated, they could not be selected in the same model. Elevation, the variable with the lowest univariate model AIC (best model) was kept for model selection. The set of best models (with a $\Delta AIC < 2$ to the model with the lowest AIC) is shown in Table S3 and, following the parsimony principle, we selected the model with the lowest number of variables within this set of best models. Thus, for annual reported data, the final model included annual population density, elevation, and primary rice crop arable areas. These three variables were positively associated with the annual reported cases (Table 2). The set of these explanatory variables is mapped and shown in the supplementary data (Figs. S7–S15).

Higher population density was generally found to increase the risk of infection. Elevation was also a predictor for leptospirosis. Specifically in the north-east region, we found that the primary rice crop arable areas were associated with a high probability of leptospirosis (Fig. S13). The estimated annual leptospirosis cases in all parts of Thailand except for the Central region, where there were relatively few cases. Figure 5 showed that our model can fit to the reported cases. Note that the weather variables were not included in the final model because of the selection criteria.

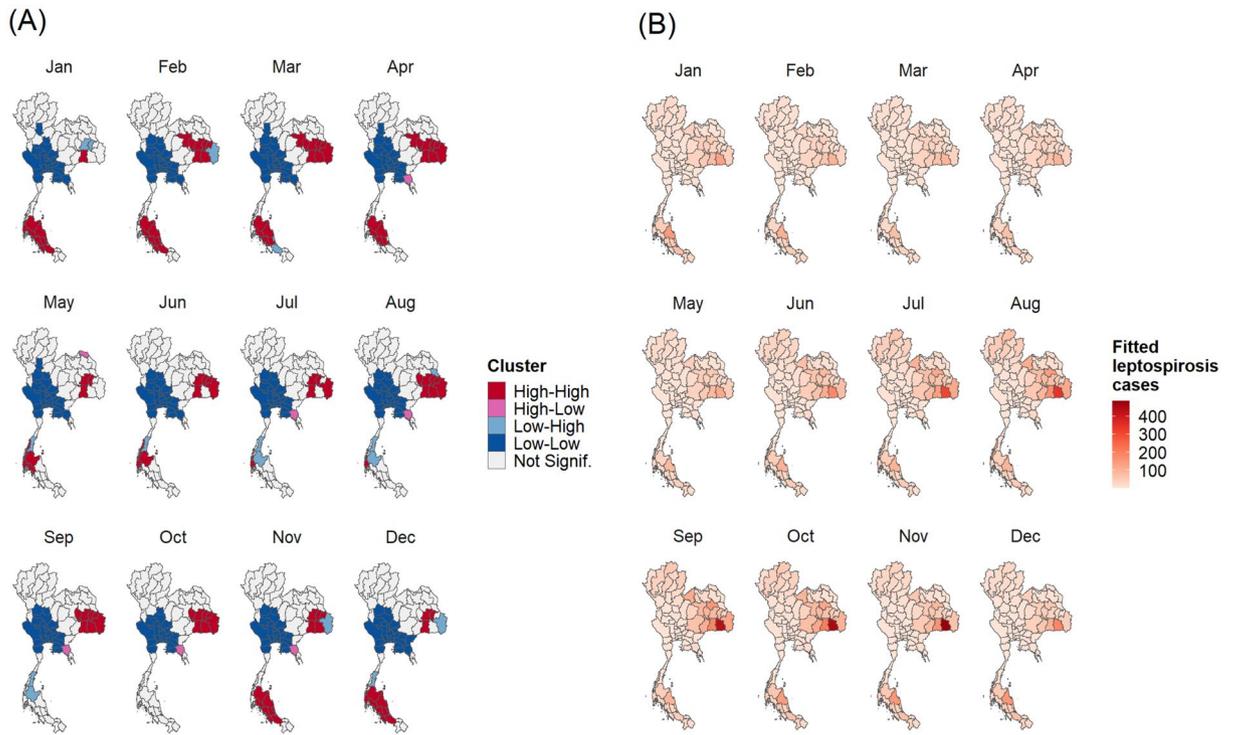


Figure 3. Plots of monthly spatial clustering of incidence rate as determined by LISA (A) and monthly leptospirosis cases estimated by the final GLMM (B). Maps created using R Program version 4.0.3 (<https://www.r-project.org/>).

Region	Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
North-eastern 	Precipitation	Blue	Blue	Light Blue	Light Blue	Light Pink	Light Pink	Red	Red	Red	Light Pink	Light Blue	Blue
	Percentage of flooded area	Blue	Blue	Blue	Blue	Light Pink	Light Pink	Red	Red	Red	Light Pink	Light Blue	Blue
	Temperature range	Red	Red	Red	Light Pink	Light Blue	Blue	Blue	Blue	Blue	Light Blue	Light Pink	Light Pink
	Incidence rate	Blue	Blue	Blue	Blue	Light Pink	Light Pink	Red	Red	Red	Red	Light Pink	Light Blue
South-eastern 	Precipitation	Light Blue	Blue	Blue	Light Blue	Light Pink	Red	Red					
	Percentage of flooded area	Light Blue	Blue	Blue	Blue	Light Pink	Red	Red					
	Temperature range	Light Blue	Red	Red	Red	Light Pink	Light Blue	Blue	Blue				
	Incidence rate	Red	Blue	Blue	Blue	Light Blue	Light Pink	Light Pink	Light Pink	Light Blue	Light Pink	Light Pink	Red
South-western 	Precipitation	Blue	Blue	Blue	Light Blue	Light Pink	Light Pink	Light Pink	Red	Red	Light Pink	Light Pink	Light Blue
	Percentage of flooded area	Blue	Blue	Blue	Light Blue	Light Pink	Light Pink	Light Pink	Red	Red	Light Pink	Light Pink	Light Blue
	Temperature range	Light Pink	Red	Red	Red	Light Pink	Light Blue	Blue	Light Blue	Blue	Blue	Blue	Light Pink
	Incidence rate	Light Blue	Blue	Blue	Blue	Light Blue	Red	Red	Red	Light Pink	Light Pink	Light Pink	Light Pink

Figure 4. Heatmaps of monthly precipitation, monthly percentage of flooded area, monthly temperature range, and average of incidence rate for 3 regions. Red highlights the high values. Blue highlights the low value. The 3 regions were mapped in dark grey. The heat map created using Microsoft Excel version 16.57 (<https://www.microsoft.com/th-th/microsoft-365/excel>). Maps created using R Program version 4.0.3 (<https://www.r-project.org/>).

Variables	OR (95% confidence interval)	P-value
Primary rice crop arable area	1.9576 (1.4398–2.6617)	<0.0001
Elevation	1.9259 (1.3125–2.8259)	0.0008
Annual population density	1.2708 (1.0213–1.5812)	0.0316

Table 2. Results of the final generalized linear mixed model for annual reported cases. *OR* odds ratio.

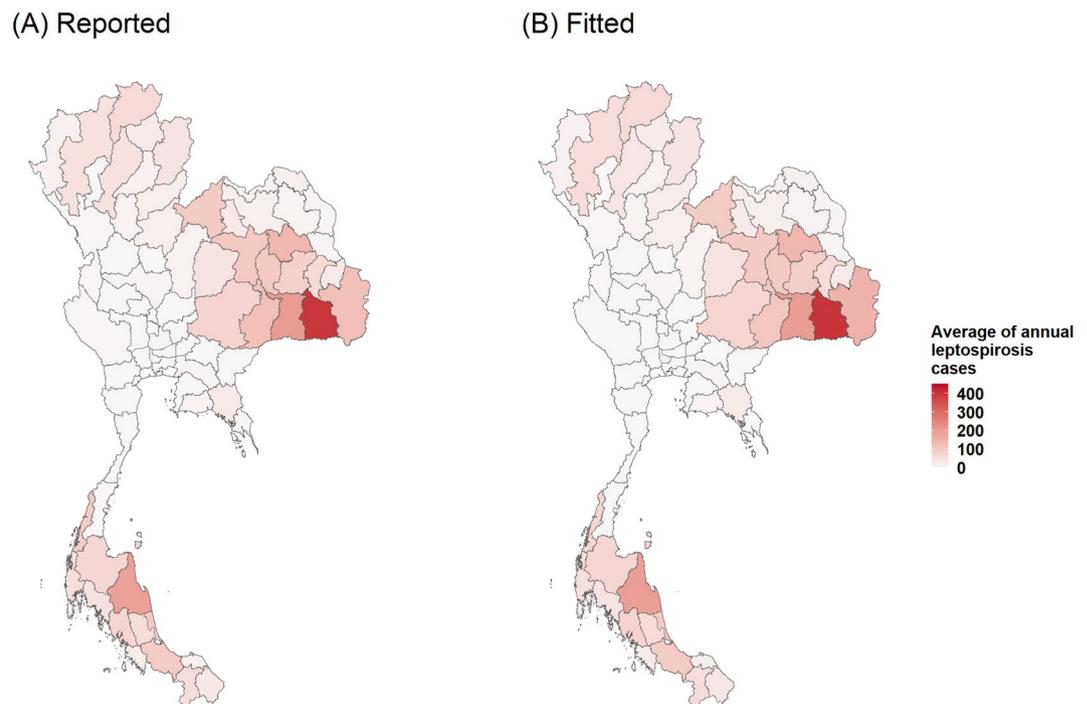


Figure 5. Plots of comparison between average annual reported leptospirosis cases (A) and average annual fitted leptospirosis cases estimated by the final GLMM (B). Maps created using R Program version 4.0.3 (<https://www.r-project.org/>).

Variables	OR (95% confidence interval)	P-value
Monthly temporal correlation	1.1253 (1.0843–1.1678)	<0.0001
Monthly soil moisture	1.3206 (1.2257–1.4230)	<0.0001
Monthly temperature range	0.7869 (0.7462–0.8297)	<0.0001
Monthly percentage of NDVI-2	0.9533 (0.9186–0.9893)	0.0115

Table 3. Results of the final generalized linear mixed model for monthly reported cases. *OR* odds ratio.

The same procedure was applied to the monthly data. For monthly reported data, the variables for selecting the final model were temporal correlation, pig density, buffalo density, monthly average soil moisture, monthly temperature range, and monthly NDVI-2 percentage (Table S4). The set of selected variables is also mapped in the supplementary data (Figs. S15–S21). The set of models with a $\Delta AIC < 2$ is shown in Tables S4 and S5. We found that the model using temperature range showed a lower AIC compared to the one using the percentage of flooded area. However, the model using flooded area used less variables, as NDVI-2 was not included. The results from both final models are presented in Table 3 and Table S6. For both models, the temporal autocorrelation parameter was included, indicating the positive correlation of the previous month. The monthly average soil moisture was found to be a positive correlation to the cases, implying the greater the soil moisture the more risk of infection. Temperature range was found to be strongly negative correlation to percentage of flooded area ($r \leq -0.7$). Also, the monthly NDVI-2 percentage were negatively with the monthly reported cases. With the alternative final models (Table S6), we found that the percentage of flooded area is positively correlated with monthly reported cases.

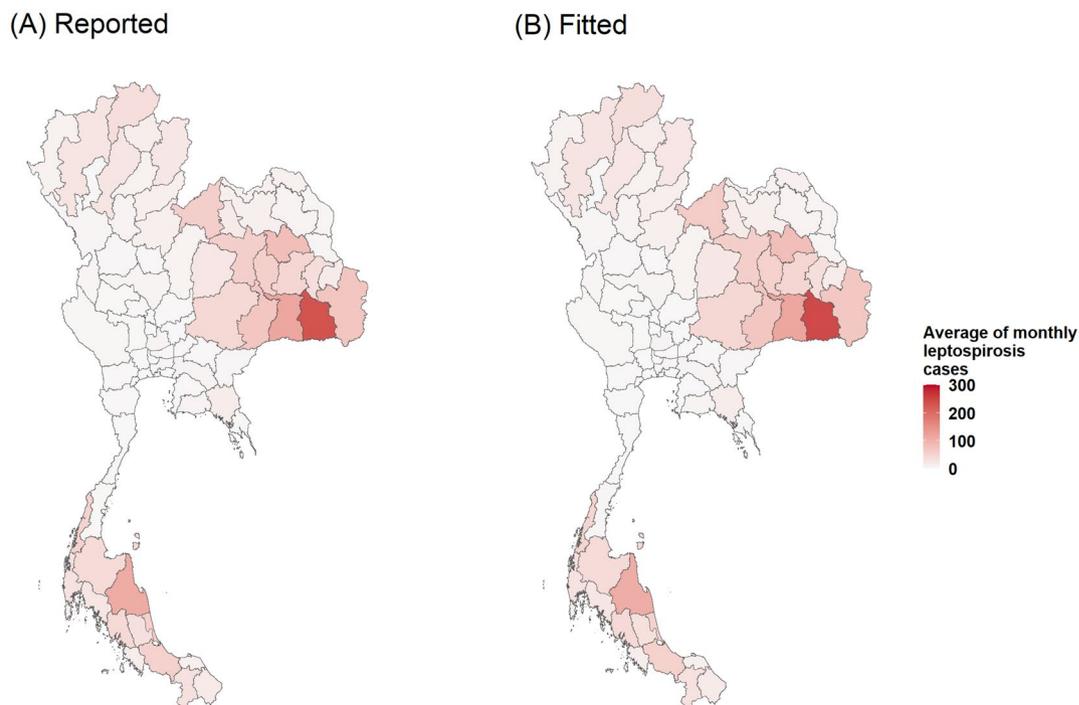


Figure 6. Plots of comparison between average monthly reported leptospirosis cases (A) and average monthly fitted leptospirosis cases estimated by the final GLMM for temperature range parameter (B). Maps created using R Program version 4.0.3 (<https://www.r-project.org/>).

The results indicated that low temperatures together with a high percentage of flooded area could increase infection risk. The estimated monthly leptospirosis cases are mapped in Fig. 3B. Both models can be used to illustrate the average monthly leptospirosis cases as shown in Fig. 6, Figs. S5, S6.

Discussion

The spatial–temporal analysis of all human leptospirosis cases reported in Thailand from 2012 to 2018 allowed us to highlight the main spatial and temporal patterns and to identify key risk factors associated with leptospirosis infection in Thailand. There are few studies of the spatial–temporal patterns of leptospirosis in the country and none to our knowledge using such a large data set. Although the association of climatic and environmental factors on leptospirosis has been studied in many previous works^{51,52}, few studies used remotely sensed data to analyze directly the impact of flooding and took into account primary and secondary rice crop arable land, cultivated areas and yields.

Overall, our analysis showed high incidence rates for specific provinces and seasons in Thailand, suggesting that leptospirosis is still a major public health concern as it included in the prevention research program to control disease by Department of Disease Control, Ministry of Public Health. The spatial autocorrelation analysis highlighted the significance of the annual and monthly spatial clustering of the leptospirosis cases. For the annual data, there are limited differences between the years, as the same provinces (provinces from north-eastern Thailand) have higher risk almost every year. Whereas for the monthly data, the high-risk provinces were different between the months for different parts of the country, especially for the north-eastern and southern parts. It is interesting that in October, which has the highest incidence rate, the high-risk provinces were observed only in north-eastern parts. This may be explained by the process of rice cultivation, which involves many activities such as preparing the land, sowing the wet fields and weeding as discussed below. The hot spots in the north-eastern are almost exclusively rural areas. On the other hand, the hot spots in the south may not be related only to rice cultivation but rather to flooding, as an outbreak occurred 2 weeks after flooding in Nakhon Si Thammarat in January 2017⁵³. It also may have a different seasonality component as compared to the north-eastern part⁵⁴ and a different landscape. The low-risk provinces were clustered in the central region of Thailand, which has the highest rice productivity in the country⁵⁵.

Our analysis highlights that the areas of different rice cultivated types (in-season and off-season) and rice yield should be considered separately to describe the risk of leptospirosis infection. Consistent with our results, rice farming activities have been identified to be an important risk factor^{20,24,25}. In Thailand, rice farming is the predominant occupation and rice farmers usually cultivate rice in two seasons. Our results demonstrated that the primary rice crop arable area (rice grown using rainwater) was positively associated with the annual leptospirosis incidence. The primary rice crop (in-season) is cultivated during the rainy season (May–October), thus increasing activities in the field, such as fertilizing and ploughing rice in wet fields, leading to a higher risk of exposure to *Leptospira*. In contrast, the secondary rice crop arable area (off-season) was not considered to be a risk factor for leptospirosis as it corresponded to rice growing activities during the non-rainy season. However,

rice yield, mainly in the central region, was found to be negatively correlated to primary rice crop arable area. The central region has soil conditions more suitable for increased rice production per area and higher water resources allowing for several rice harvest per year. Our results may suggest that the contamination of *Leptospira* may depend on the soil characteristics⁵⁶. It may also be due to the different practices for rice cultivation in the central region, which has more mechanized processes than farmers in north-eastern region. Another reason may be the highly effective healthcare system in the central region compared to other parts of the country. The leptospirosis infection could be explained by using land use such as rice growing data and landscape such as elevation. Elevation may be a proxy for increased flooding because lower elevations store rainwater. However, elevation was calculated at province level, the finer spatial resolution should further analysis to be more specific calculation. Due to the limitation of the data (annual data for rice yield), further analysis for the monthly data should be evaluated to calculate the risk factors associated with different seasons.

For the monthly data, we found that using temperature data leads to a minor improvement in the model compared to using the percentage of flooded area. The results showed both parameters can be predictor variables for leptospirosis. Our analyses revealed that temperature has a strongly negative correlation to percentage of flooded area. Usually, in tropical climates, low temperatures correlate to cloudy weather, where there is high rainfall and therefore prone to flood³². The temperature range could be a predictor variable, when the percentage of flooded area is not available. However, using temperature range instead of the percentage of flooding area should be only used in similar climatic areas as Thailand where the model was built and where the correlation between these two factors is likely to hold true. More generally, any extrapolation of this model to other climatic regions should be implemented carefully.

Our finding allows the potential risk of leptospirosis infection to be estimated almost in real-time. *Leptospira* can survive in freshwater² where flooding events could increase the number of pathogenic *Leptospira*⁵⁷. The average soil moisture, or the soil water capacity³¹, was identified as a risk factor for monthly data in this work. *Leptospira* can survive in soils with a moisture content of $\geq 20\%$ ⁵⁸. Increased soil moisture may increase the survival rate in contaminated soil and water²⁶. Note that our results were based on the case reports in the surveillance system. This may not be accurate because some mild symptoms or asymptomatic cases are not going to the hospital, resulting in underreporting of leptospirosis cases. A finer spatial and temporal scale should also be conducted, when well-represented data is available. The analysis could only be performed using the data set by the month and year due to the low number of cases. Regardless of these limitations, our study has provided important knowledge on leptospirosis occurrences by characterizing the hot-spots and key risk factors.

Conclusion

In summary, our study highlighted that when studied over a long period, leptospirosis cases demonstrated a spatial-temporal distribution and showed hot-spots and clustered areas. Regression analysis was used to find the risk factors associated with leptospirosis cases. Working in rice field, which have high soil moisture, could increase risk of infection due to occupational behaviour of working in wet fields facilitating the survival of *Leptospira* bacteria. The temperature range, which is negatively correlated to the percentage of flooded area was found to be negatively associated with cases. Therefore, flood exposure controls should be used to reduce the risk of leptospirosis infection. These results could be used to improve prevention measures and control actions for public health organizations. The investigation of the associated risk factors can contribute to the prevention and early warning of leptospirosis in Thailand.

Data availability

The leptospirosis dataset used in the current study are available from Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health, Thailand upon reasonable request. The data supporting the findings can be found in the main paper and in Supplementary Information file.

Received: 24 August 2021; Accepted: 16 March 2022

Published online: 24 March 2022

References

- Costa, F. *et al.* Global morbidity and mortality of leptospirosis: A systematic review. *PLoS Negl. Trop. Dis.* **9**, e0003898. <https://doi.org/10.1371/journal.pntd.0003898> (2015).
- Bierque, E., Thibeaux, R., Girault, D., Soupe-Gilbert, M. E. & Goarant, C. A systematic review of *Leptospira* in water and soil environments. *PLoS ONE* **15**, e0227055. <https://doi.org/10.1371/journal.pone.0227055> (2020).
- Adler, B. & de la Pena Moctezuma, A. *Leptospira* and leptospirosis. *Vet. Microbiol.* **140**, 287–296. <https://doi.org/10.1016/j.vetmic.2009.03.012> (2010).
- Haake, D. A. & Levett, P. N. Leptospirosis in humans. *Curr. Top. Microbiol. Immunol.* **387**, 65–97. https://doi.org/10.1007/978-3-662-45059-8_5 (2015).
- Andre-Fontaine, G., Aviat, F. & Thorin, C. Waterborne Leptospirosis: Survival and preservation of the virulence of pathogenic *Leptospira* spp. in fresh water. *Curr. Microbiol.* **71**, 136–142. <https://doi.org/10.1007/s00284-015-0836-4> (2015).
- Casanovas-Massana, A. *et al.* Quantification of *Leptospira* interorgans survival in soil and water microcosms. *Appl. Environ. Microbiol.* **84**, 1. <https://doi.org/10.1128/AEM.00507-18> (2018).
- Thibeaux, R. *et al.* Seeking the environmental source of Leptospirosis reveals durable bacterial viability in river soils. *PLoS Negl. Trop. Dis.* **11**, e0005414. <https://doi.org/10.1371/journal.pntd.0005414> (2017).
- Chadsuthi, S., Modchang, C., Lenbury, Y., Iamsirithaworn, S. & Triampo, W. Modeling seasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using time-series and ARIMAX analyses. *Asian Pac. J. Trop. Med.* **5**, 539–546. [https://doi.org/10.1016/S1995-7645\(12\)60095-9](https://doi.org/10.1016/S1995-7645(12)60095-9) (2012).
- Luenam, A. & Puttanapong, N. Modelling and analyzing spatial clusters of leptospirosis based on satellite-generated measurements of environmental factors in Thailand during 2013–2015. *Geospat. Health* **15**. <https://doi.org/10.4081/gh.2020.856> (2020).

10. Ehelepola, N. D. B., Ariyaratne, K. & Dissanayake, D. S. The interrelationship between meteorological parameters and leptospirosis incidence in Hambantota district, Sri Lanka 2008–2017 and practical implications. *PLoS ONE* **16**, e0245366. <https://doi.org/10.1371/journal.pone.0245366> (2021).
11. Dhewantara, P. W. *et al.* Climate variability, satellite-derived physical environmental data and human leptospirosis: A retrospective ecological study in China. *Environ. Res.* **176**, 108523. <https://doi.org/10.1016/j.envres.2019.06.004> (2019).
12. Sumi, A. *et al.* Effect of temperature, relative humidity and rainfall on dengue fever and leptospirosis infections in Manila, the Philippines. *Epidemiol. Infect.* **145**, 78–86. <https://doi.org/10.1017/S095026881600203X> (2017).
13. Syamsuar, A. D., Maria, I. L., Hatta, M. & Mallongi, A. Determinant Factors of Leptospirosis in Indonesia: Flood Prone Area Setting in Wajo District in ICER-PH 2018: Proceedings of the 3rd International Conference on Environmental Risks and Public Health, ICER-PH 2018, 26–27, October 2018, Makassar, Indonesia. 254 (European Alliance for Innovation).
14. Ijaz, M. *et al.* Sero-epidemiology and hemato-biochemical study of bovine leptospirosis in flood affected zone of Pakistan. *Acta Trop.* **177**, 51–57. <https://doi.org/10.1016/j.actatropica.2017.09.032> (2018).
15. de Resende Londe, L., Da Conceição, R. S., Bernardes, T. & de Assis Dias, M. C. Flood-related leptospirosis outbreaks in Brazil: perspectives for a joint monitoring by health services and disaster monitoring centers. *Nat. Hazards* **84**, 1419–1435 (2016).
16. Kawaguchi, L. *et al.* Seroprevalence of leptospirosis and risk factor analysis in flood-prone rural areas in Lao PDR. *Am. J. Trop. Med. Hyg.* **78**, 957–961 (2008).
17. Ledien, J. *et al.* Assessing the performance of remotely-sensed flooding indicators and their potential contribution to early warning for leptospirosis in Cambodia. *PLoS ONE* **12**, e0181044. <https://doi.org/10.1371/journal.pone.0181044> (2017).
18. Raghavan, R. K., Brenner, K. M., Harrington, J. A. Jr., Higgins, J. J. & Harkin, K. R. Spatial scale effects in environmental risk-factor modelling for diseases. *Geospat. Health* **7**, 169–182. <https://doi.org/10.4081/gh.2013.78> (2013).
19. Bureau of Epidemiology, Department of Disease Control & Ministry of Public Health. *Leptospirosis [in Thai]*. <http://doe.moph.go.th/surdata/disease.php?ds=43> (2021).
20. Tangkanakul, W. *et al.* Risk factors associated with leptospirosis in northeastern Thailand, 1998. *Am. J. Trop. Med. Hyg.* **63**, 204–208. <https://doi.org/10.4269/ajtmh.2000.63.204> (2000).
21. Hinjoy, S. *et al.* Environmental and behavioral risk factors for severe leptospirosis in Thailand. *Trop. Med. Infect. Dis.* **4**. <https://doi.org/10.3390/tropicalmed4020079> (2019).
22. Owers, K. A. *et al.* Timing and spatial heterogeneity of leptospirosis transmission in Northeast Thailand. *Am. J. Trop. Med. Hyg.* **97**, 1–674 (2017).
23. Della Rossa, P. *et al.* Environmental factors and public health policy associated with human and rodent infection by leptospirosis: a land cover-based study in Nan province, Thailand. *Epidemiol. Infect.*, 1–13 (2015).
24. Sahneh, E., Delpisheh, A., Sayehmiri, K., Khodabakhshi, B. & Moafi-Madani, M. Investigation of risk factors associated with leptospirosis in the North of Iran (2011–2017). *J. Health Sci. Res.* **19**, e00449 (2019).
25. Maze, M. J. *et al.* Risk factors for human acute leptospirosis in northern Tanzania. *PLoS Negl. Trop. Dis.* **12**, e0006372. <https://doi.org/10.1371/journal.pntd.0006372> (2018).
26. Baquero, O. S. & Machado, G. Spatiotemporal dynamics and risk factors for human Leptospirosis in Brazil. *Sci. Rep.* **8**, 15170. <https://doi.org/10.1038/s41598-018-33381-3> (2018).
27. Dhewantara, P. W. *et al.* Geographical and temporal distribution of the residual clusters of human leptospirosis in China, 2005–2016. *Sci. Rep.* **8**, 16650. <https://doi.org/10.1038/s41598-018-35074-3> (2018).
28. Nunez-Gonzalez, S., Gault, C., Granja, P. & Simancas-Racines, D. Spatial patterns of leptospirosis in Ecuador, 2013–2018. *Trans. R. Soc. Trop. Med. Hyg.* **114**, 545–548. <https://doi.org/10.1093/trstmh/trz124> (2020).
29. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **27**, 3025–3033 (2006).
30. Hashim, H., Abd Latif, Z. & Adnan, N. A. Urban vegetation classification with NDVI threshold value method with very high resolution (VHR) PLEIADES Imagery. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, 237–240 (2019).
31. Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A. & Hegewisch, K. C. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* **5**, 170191. <https://doi.org/10.1038/sdata.2017.191> (2018).
32. Yong, W. & Zablon, S. W. Variability of diurnal temperature range in east Africa during 1921–2010. *J. Trop. Meteorol.* **23**, 345–356 (2017).
33. Consortium for Spatial Information (CGIAR-CSI). *SRTM 90m DEM Digital Elevation Database*. <https://srtm.csi.cgiar.org/srtmdata/> (2020).
34. Department of Livestock Development (DLD) & Ministry of Agriculture and Cooperatives of Thailand. *Number of farmer households and livestock classified by species in Thailand [in Thai]*. <https://ict.dld.go.th/webnew/index.php/th/service-ict/report> (2020).
35. Hinjoy, S. Epidemiology of leptospirosis from Thai national disease surveillance system, 2003–2012. *OSIR J.* **7**, 1–5 (2016).
36. Moran, P. A. The interpretation of statistical maps. *J. R. Stat. Soc., B: Stat. Methodol.* **10**, 243–251 (1948).
37. Bivand, R. *et al.* *spdep: Spatial dependence: Weighting schemes, statistics*. <https://cran.r-project.org/web/packages/spdep/index.html> (2015).
38. Anselin, L. Local indicators of spatial association—LISA. *Geogr. Anal.* **27**, 93–115 (1995).
39. Anselin, L., Syabri, I. & Kho, Y. GeoDa: An introduction to spatial data analysis. *Geogr. Anal.* **38**, 5–22. <https://doi.org/10.1111/j.0016-7363.2005.00671.x> (2006).
40. Brooks, M. E. *et al.* glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J.* **9**, 378–400 (2017).
41. Brooks, M. E. *et al.* Modeling zero-inflated count data with glmmTMB. Preprint at <https://doi.org/10.1101/132753v1.full> (2017).
42. Magnusson, A. *et al.* *glmmTMB: Generalized linear mixed models using template model builder*. <https://cran.r-project.org/web/packages/glmmTMB/index.html> (2017).
43. Barton, K. & Barton, M. K. *MuMIn: Multi-model inference*. <https://cran.r-project.org/web/packages/MuMIn/index.html> (2015).
44. R Core Team. *R: A language and environment for statistical computing*. <https://www.R-project.org> (2019).
45. Hadley, W. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
46. Hadley, W. stringr: Simple, consistent wrappers for common string operations. <http://stringr.tidyverse.org>. <https://github.com/tidyverse/stringr> (2021).
47. Bivand, R. S. & Wong, D. W. Comparing implementations of global and local indicators of spatial association. *TEST* **27**, 716–748 (2018).
48. Pebesma, E. J. Simple features for R: standardized support for spatial vector data. *R J.* **10**, 439 (2018).
49. Alboukadel, K. *ggpubr: 'ggplot2' Based Publication Ready Plots*. <https://CRAN.R-project.org/package=ggpubr> (2020).
50. Alboukadel, K. *ggcorrplot: Visualization of a Correlation Matrix using 'ggplot2'*. <https://CRAN.R-project.org/package=ggcorrplot> (2019).
51. Chadsuthi, S., Chalvet-Monfray, K., Wiratsudakul, A. & Modchang, C. The effects of flooding and weather conditions on leptospirosis transmission in Thailand. *Sci. Rep.* **11**, 1486. <https://doi.org/10.1038/s41598-020-79546-x> (2021).
52. Chadsuthi, S., Chalvet-Monfray, K., Wiratsudakul, A., Suwancharoen, D. & Cappelle, J. A remotely sensed flooding indicator associated with cattle and buffalo leptospirosis cases in Thailand 2011–2013. *BMC Infect. Dis.* **18**, 602. <https://doi.org/10.1186/s12879-018-3537-3> (2018).

53. Chutinantakul, A., Chanakitjanukit, P. & Buathong, R. Outbreaks of leptospirosis after a flood in Thung Song District, Nakhon Si Thammarat, January 2017. *Dis. Control J.* **45**, 317–329 (2019).
54. Climatological Group, Meteorological Development Bureau & Meteorological Department. *The Climate of Thailand*. https://www.tmd.go.th/en/archive/thailand_climate.pdf (2015).
55. Suebpongsang, P., Ekasingh, B. & Cramb, R. *Commercialisation of rice farming in northeast thailand in White Gold: The Commercialisation of Rice Farming in the Lower Mekong Basin* 39–68 (Palgrave Macmillan, 2020).
56. Gamage, C. D., Sato, Y., Kimura, R., Yamashiro, T. & Toma, C. Understanding leptospirosis eco-epidemiology by environmental DNA metabarcoding of irrigation water from two agro-ecological regions of Sri Lanka. *PLoS Negl. Trop. Dis.* **14**, e0008437 (2020).
57. Naing, C., Reid, S. A., Aye, S. N., Htet, N. H. & Ambu, S. Risk factors for human leptospirosis following flooding: A meta-analysis of observational studies. *PLoS ONE* **14**, e0217643. <https://doi.org/10.1371/journal.pone.0217643> (2019).
58. Saito, M. *et al.* Comparative analysis of *Leptospira* strains isolated from environmental soil and water in the Philippines and Japan. *Appl. Environ. Microbiol.* **79**, 601–609. <https://doi.org/10.1128/AEM.02728-12> (2013).

Acknowledgements

We thank Faculty of Science, Naresuan University for their support.

Author contributions

S.C.: conceptualized, participated in its design, performed the analysis, wrote the first draft and wrote the manuscript. J.C.: conceptualized, participated in its design, analysis and wrote the manuscript. K.C.M.: conceptualized, participated in its design and analysis. S.G.: participated in data extraction, interpretation, and analysis. P.W.: participated in its design and analysis. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-09079-y>.

Correspondence and requests for materials should be addressed to S.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022