



A Probabilistic Spatial Dengue Fever Risk Assessment by a Threshold-Based-Quantile Regression Method

Chuan-Hung Chiu¹, Tzai-Hung Wen², Lung-Chang Chien³, Hwa-Lung Yu^{1*}

1 Department of Bioenvironmental systems engineering, National Taiwan University, Taipei, Taiwan, **2** Department of Geography, National Taiwan University, Taipei, Taiwan, **3** Division of Biostatistics, University of Texas School of Public Health at San Antonio Regional Campus, San Antonio, Texas, United States of America; Research to Advance Community Health Center, University of Texas Health Science Center at San Antonio Regional Campus, San Antonio, Texas, United States of America

Abstract

Understanding the spatial characteristics of dengue fever (DF) incidences is crucial for governmental agencies to implement effective disease control strategies. We investigated the associations between environmental and socioeconomic factors and DF geographic distribution, and proposed a probabilistic risk assessment approach that uses threshold-based quantile regression to identify the significant risk factors for DF transmission and estimate the spatial distribution of DF risk regarding full probability distributions. To interpret risk, return period was also included to characterize the frequency pattern of DF geographic occurrences. The study area included old Kaohsiung City and Fongshan District, two areas in Taiwan that have been affected by severe DF infections in recent decades. Results indicated that water-related facilities, including canals and ditches, and various types of residential area, as well as the interactions between them, were significant factors that elevated DF risk. By contrast, the increase of per capita income and its associated interactions with residential areas mitigated the DF risk in the study area. Nonlinear associations between these factors and DF risk were present in various quantiles, implying that water-related factors characterized the underlying spatial patterns of DF, and high-density residential areas indicated the potential for high DF incidence (e.g., clustered infections). The spatial distributions of DF risks were assessed in terms of three distinct map presentations: expected incidence rates, incidence rates in various return periods, and return periods at distinct incidence rates. These probability-based spatial risk maps exhibited distinct DF risks associated with environmental factors, expressed as various DF magnitudes and occurrence probabilities across Kaohsiung, and can serve as a reference for local governmental agencies.

Citation: Chiu C-H, Wen T-H, Chien L-C, Yu H-L (2014) A Probabilistic Spatial Dengue Fever Risk Assessment by a Threshold-Based-Quantile Regression Method. PLoS ONE 9(10): e106334. doi:10.1371/journal.pone.0106334

Editor: Osman Alimamy Sankoh, INDEPTH Network, Ghana

Received: February 5, 2014; **Accepted:** August 4, 2014; **Published:** October 10, 2014

Copyright: © 2014 Chiu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by funds from Taiwan Ministry of Science and Technology (MOST 101-2628-E-002-017-MY3 and MOST 102-2221-E-002-140-MY3), and Centers for Disease Control in Taiwan (DOH102-DC-1203). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: hlyu@ntu.edu.tw

Introduction

Dengue fever (DF) is among the most severe vector-borne infectious diseases spread by mosquitoes in tropical and subtropical regions; 2.5 billion people in over 100 countries are at risk of contracting the dengue virus [1–5]. More than 50 million infections occur annually, particularly in Southeast Asian and western Pacific regions [5,6]. Several species of mosquito (e.g., *Aedes aegypti* and *Aedes albopictus*) transmit five distinct dengue virus serotypes, including DENV-1, DENV-2, DENV-3, and DENV-4 [7,8]. No commercially available vaccine exists to mitigate the disease spread, and therefore, determining the spatiotemporal distribution of DF incidence is among the top priorities for DF prevention and control [9–13]. Although numerous DF studies have focused on determining the DF etiology to facilitate space–time prediction, a risk assessment framework to account for DF risk factors and provide risk measures across space (e.g., the probability and magnitude of DF incidences) must be established to control and manage the disease.

Identifying the major surrogate factors for DF incidence is key to assessing DF risk. Ideal surrogate factors are risk factors that not only are closely associated with DF etiology, but also should be

readily available [13], i.e., easier to be accessed or observed. For example, climatic variables are a critical factor characterizing the temporal patterns of DF incidence [14]. Regarding the spatial characteristics of DF incidences, previous research has explored land-use indicators and socioeconomic status to approximate the local magnitudes of complex interactions between infected and susceptible human hosts and DF vectors [15–18]. High-risk DF areas are closely associated with the size and location of agricultural, forest, and residential areas, which are the preferred habitats of DF vectors. Populated areas close to vector-preferred areas exhibit increased DF occurrence risk [15,19–21]. In addition, previous studies have indicated that high population densities, residential areas, and areas with low income family exhibit increased risk of DF transmission [22–25]. The statistical relationships among DF incidence, land use, and socioeconomic factors frequently change across the study area because of the distinct environmental and climatic conditions. Therefore, determining the associations between DF outbreaks and the environmental factors of each study area is essential. We used quantile regression to characterize the relationships between risk factors and DF incidence across the quantiles of the DF incidence probability distribution.

Probabilistic risk assessment is a systematic approach to evaluating the risk of an event in terms of probabilities and their associated magnitudes [26–28]. Although a probabilistic approach provides a means to characterize the uncertainties of the events (e.g., a DF epidemic), interpreting the probability associated with risks is difficult for the general population and decision makers [29]. By contrast, frequency analysis is used to estimate the return period a prespecified event (e.g., extreme rainfall, floods, tornados, earthquakes) [30]. The return period is the average number of years between occurrences of the annual event with magnitude larger than a specified level. The selected and analyzed annual events should be homogeneous and independent [31]. In other words, interpreting the probabilities of risks is possible by estimating their associated return periods, using frequency analysis.

This paper proposes a threshold-based quantile regression approach to investigate the functional relationships between the spatial distributions of environmental risk factors, including land use and socioeconomic factors, and DF incidence. The functional relationship was characterized using estimations at prespecified quantiles of DF incidence (i.e., 0.1, 0.2, ..., 0.9) that comprise the spatially varying nonparametric probability distribution of DF risk associated with the spatial factors. The spatial distributions of DF risk were estimated and expressed as return periods for ease of interpretation.

Materials

Study area

For decades, the DF annual incidence rate has been among the highest for infectious diseases in Taiwan [32] and is particularly high in old Kaohsiung City (Fig. 1), the second largest city in Taiwan with a population of 2.9 million distributed over 11 districts and an area of 154 km². Fongshan District, located to the southeast of Kaohsiung City and covering an area of 26 km², was the former capital of Kaohsiung County and has a population of approximately 36 000. Figure 1 depicts the area covered in this study: old Kaohsiung City and Fongshan District, which both include major commercial, residential, and political areas of Southern Taiwan. This area comprises 535 *li* (the smallest administrative unit in Taiwan) with population sizes ranging from 160 to 37 000 and an average population density of approximately 950 people per square kilometer (Fig. 1). The study area is located in a tropical region that has a clear distinction between dry and wet seasons. The wet season typically lasts for half a year, from mid-May to mid-October, during which time the daily rainfall can exceed 500 mm during typhoon periods. The dry season comprises the rest of the year, during which time there is little to no rain. The average annual rainfall of this area is approximately 1884.9 mm. The daily temperature exceeds 20°C for most of the year, with minimum and maximum average monthly temperatures of 19.3°C in January and 29.2°C in July, respectively. Because of the limited area (i.e., 180 km²) and topographical changes of the study area, the variability of climatic variables (e.g., temperature) across the study area was relatively small.

Data

We investigated DF cases based on surveillance data obtained from the Centers for Disease Control (CDC) in Taiwan. Data was obtained by following the standard data request procedure on the CDC website [33]. The dataset included total DF cases in the 353-*li* area, including both local and imported cases, from 2004 to 2011. Only 2% of the total reported DF cases were considered

imported in the dataset, and no significant distribution patterns of imported cases across space and time were observed. Figure 2 illustrates the spatial distribution of the average annual incidence of DF. The DF risk factors considered in this study included a various land uses and socioeconomic attributes. Land-use data were obtained from the Taiwan National Land Surveying and Mapping Center (NLSC) [34]. This analysis included 103 land-use factors, including agriculture, forest, public utility, and residential areas. The NLSC website lists a detailed description and definition of every land-use factor [35,36]. Figure 3 shows various land-use factors (i.e., canals, ditches, and residential areas). The land-use factors for each *li* are expressed according to area ratio and size, i.e., the percentage and areal size of a land-use factor in the *li*'s. Socioeconomic data were collected from the Fiscal Information Agency of the Ministry of Finance in Taiwan in 2009. This dataset comprises the spatial distributions of 12 socioeconomic attributes, such as per capita income, and average population density across the 535 *li*.

Method

Koenker and Bassett [37] first introduced quantile regression specifically to estimate the relationship between covariates and the quantile of response variable distribution. The regression is flexible, allowing covariates to exert various effects at distinct points of the distribution, and the estimation is robust to nonnormality and skewed tails of the data distribution [38]. We adopted quantile regression to investigate the underlying relationships between average annual DF incidence rate *Y* (i.e., average cases per 10 000 people) and potential risk factors *x* (i.e., land use and socioeconomic attributes). Unlike conventional regression approaches, quantile regression not only considers the central tendency of the response variable regarding covariate changes, but also determines the covariate effect conditional on specific quantiles of the response variable.

Let τ within the range of [0,1] be a quantile level of the *Y* distribution. The relationship between τ and *Y* is denoted as $q_Y(\tau) = F_Y^{-1}(\tau)$, where $F_Y^{-1}(\tau)$ is an inverse of the cumulative distribution function (CDF) of *Y*. Given *li* location $s_i, i = 1, \dots, n$, *Y* and *x* are rewritten as $Y(s_i)$ and $x(s_i)$, respectively. In addition, $Y(s_i) = N(s_i)/Z(s_i)$, where $N(s_i)$ and $Z(s_i)$ are average dengue cases and population size s_i , respectively. Thus, the relationship between $Y(s_i)$ and $x(s_i)$ in selected quantile τ can be written as

$$q_{Y(s_i)}[\tau|x(s_i)] = \sum_{j=0}^m \tilde{\beta}_j(\tau) x_j(s_i), \tag{1}$$

where $\tilde{\beta}_0(\tau), \tilde{\beta}_1(\tau), \dots, \tilde{\beta}_m(\tau)$ denotes the estimated intercept and coefficients in a prespecified quantile level τ , and *m* is the number of $x(s_i)$. The DF dataset indicated the absence of DF incidence in some *li* during the study period, implying that the relationship between the covariate changes and dengue incidence quantiles were not linear; specifically, thresholds existed for the risk factors above which the factors significantly affected the DF risk change. Based on this assumption, the general quantile regression formula shown in Eq. (1) was transformed into a threshold-based quantile regression to account for the threshold values for each identified covariate. To obtain the model thresholds, the original dataset was classified into two sets, $[Y^0(s_i), x_j^0(s_i)]$ and $[Y^1(s_i), x_j^1(s_i)]$, where $Y^0(s_i)$ and $Y^1(s_i)$ represent the observed average incident rate equal to zero and larger than zero, respectively, and $x_j^0(s_i)$ and $x_j^1(s_i)$ are their corresponding risk factors at location s_i . The

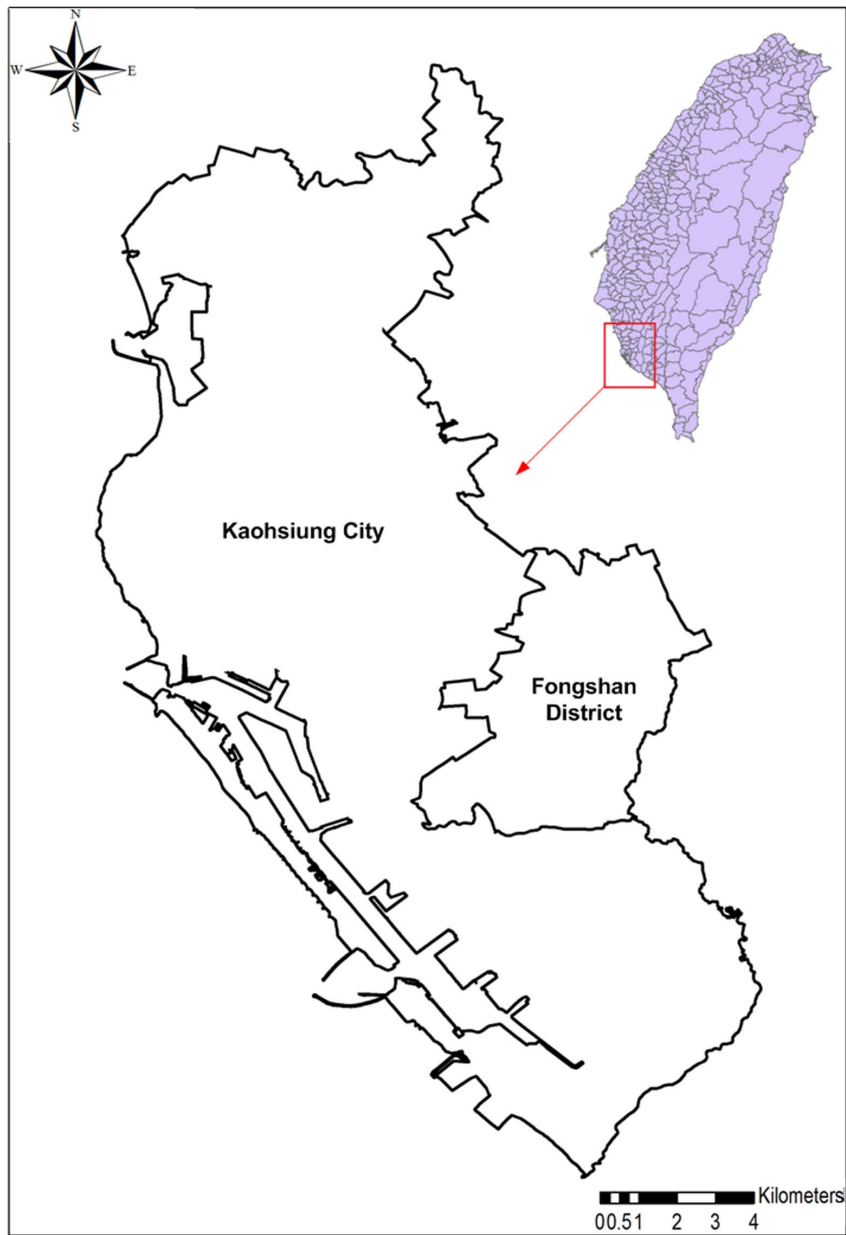


Figure 1. Map of Kaohsiung city and Fongshan district.

doi:10.1371/journal.pone.0106334.g001

proposed threshold-based quantile regression model can be expressed as

$$q_{Y^1(s_i)}[\tau|u(s_i)] = \sum_{j=0}^m \hat{\beta}_j(\tau) u_j(s_i) \quad (2)$$

where $u_j(s_i) = \begin{cases} x_j^1(s_i) - \bar{x}_j^0(s_i) & \text{while } x_j^1(s_i) > \bar{x}_j^0(s_i); \\ 0 & \text{while } x_j^1(s_i) \leq \bar{x}_j^0(s_i); \end{cases}$ $\hat{\beta}_j(\tau)$, $j = 1, \dots, m$ are the estimated coefficients for threshold-adjusted covariates $u_j(s_i)$; and $\hat{\beta}_0(\tau)$ is the intercept. To determine the influential variables, a two-stage variable selection process (i.e., variable screening and stepwise variable selection) was employed. In the first stage, variables were removed

as long as they are closely linear correlated with other variables and, based upon literature review, not evident to be a risk factor of DF incidence. In the variable screening process, before performing a quantile regression, a Pearson correlation, ρ , was implemented to evaluate collinearity among the risk factors, where $x_j(s_i)$, $j = 1, \dots, m$, with absolute correlation coefficients larger than 0.35 with other covariates were removed before further analysis. For example, the variables of road areal ratio and population density were highly correlated ($\rho > 0.35$). Road areal ratio was removed because population density exerted a greater influence on DF in studies [22,24,39]. In the variable selection process, the covariate selection in Eq. (2) followed the stepwise selection technique [40], based on Akaike's information criterion (AIC) [41] (i.e., the lower the AIC, the better the performance of the model is). The optimal thresholds $\bar{x}_j^0(s_i)$ for each covariate

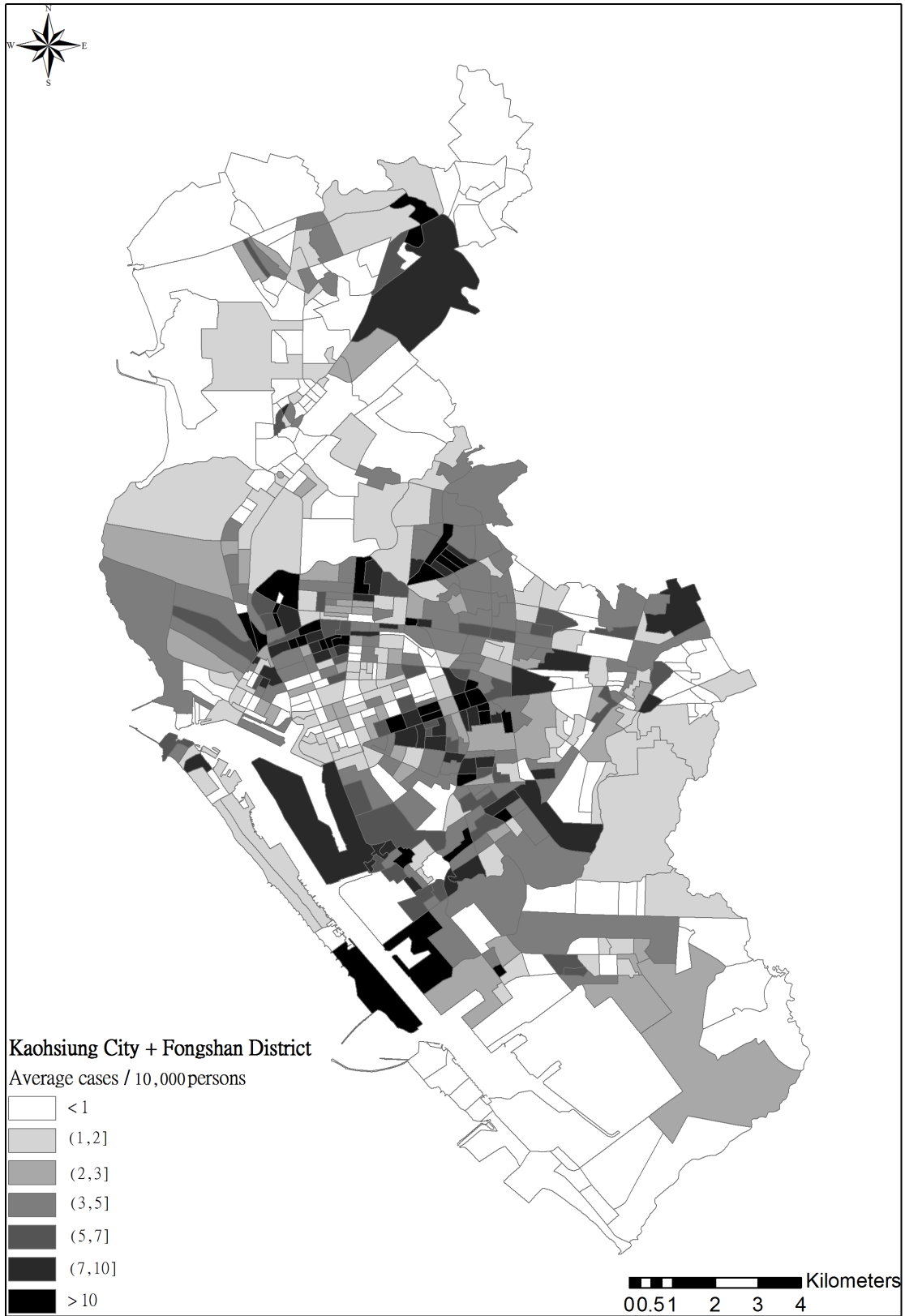


Figure 2. Spatial distribution of average DF incidence rate in 535 Li's during 2004–2011.
doi:10.1371/journal.pone.0106334.g002

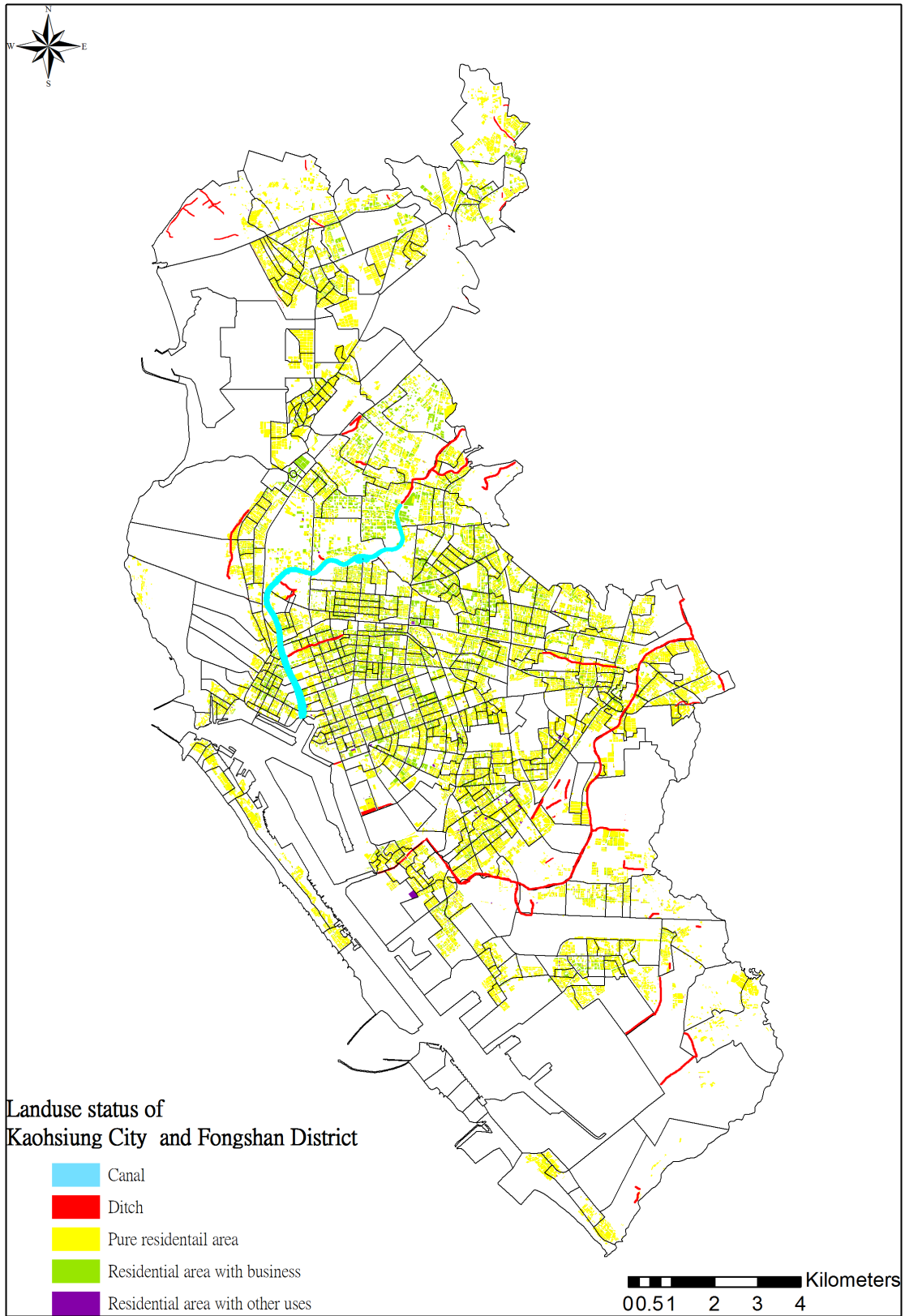


Figure 3. Spatial distribution of landuse factors important to DF across Kaohsiung-Fongshan area.
 doi:10.1371/journal.pone.0106334.g003

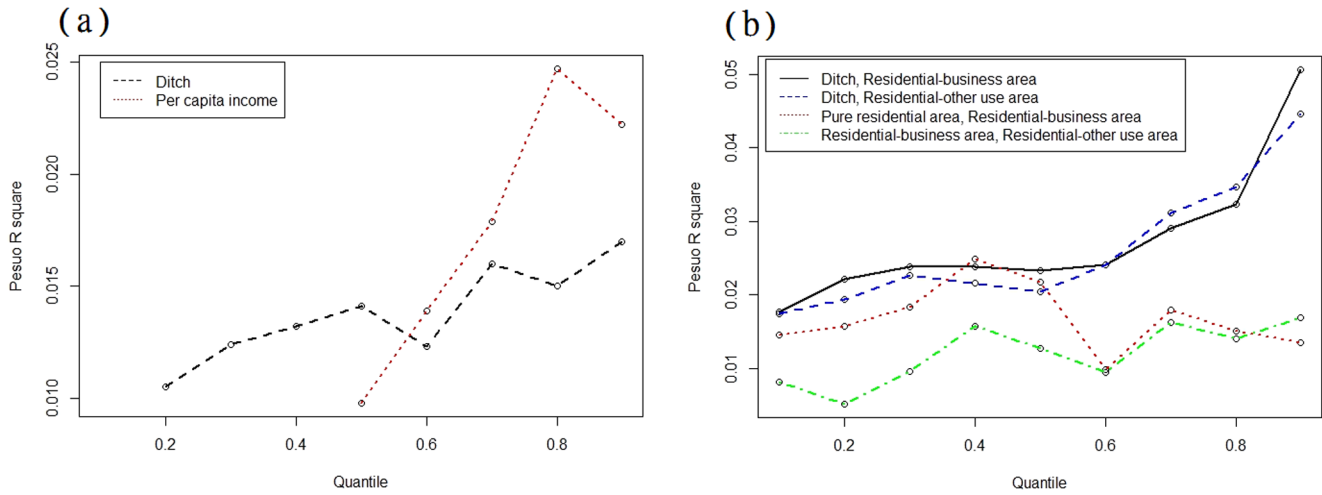


Figure 4. Variations of Pseudo R-square across quantile levels of (a) the two most significant risk factors, and (b) the four most significant interaction effects.

doi:10.1371/journal.pone.0106334.g004

were estimated by maximizing the pseudo R^2 (pR^2). The pR^2 of the optimal models at various quantile levels were calculated separately to assess their performances. The analyses of the quantile regression models were performed using the quantreg package in R [42].

The semiparametric CDF for $Y(s_i)$ was estimated with respect to τ and $u_j(s_i)$ by using Eq. (2). The quantile-based CDF can be used to estimate the exceedance probability (EP) of $Y(s_i)$, $p = Pr(Y(s_i) > y)$, where y is a selected annual DF incidence rate. In frequency analysis, a distinct perspective of EP can be obtained by using return period due to the close relationship between the probability distribution and return period of an event. Assuming the yearly DF cases during the study period were independent, the return period of a specified annual incidence rate

would be equal to the reciprocal of its EP (i.e., $1/p$) based on frequency analysis. For example, a 10-year DF outbreak has a $1/10 = 0.1$ or 10% chance of being exceeded in one year. This approach facilitated a probabilistic risk assessment of DF epidemics by estimating the full CDFs of DF risk across the study area without distributional assumptions. DF risk spatial distributions can be expressed as DF incidence statistical characteristics (e.g., mean and median) or EP and return periods regarding the prespecified EP and incidence rates. Based on risk maps, areas of dense DF incidence (e.g., locations with high average incidence rates or short return periods) can be identified. This approach can encourage public health authorities to prioritize DF prevention.

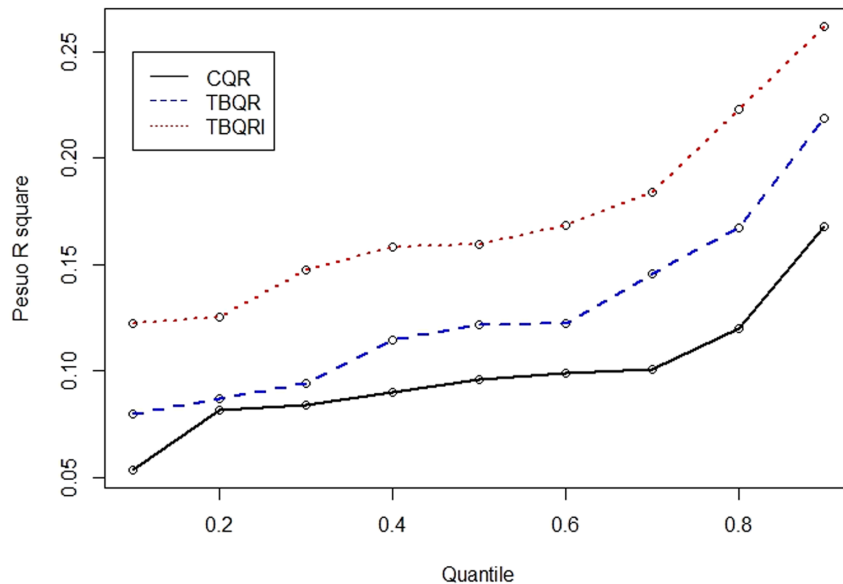


Figure 5. Comparison of pseudo R-squares across quantile levels among 1) conventional quantile regression model (CQR), 2) threshold-based quantile regression model (TBQR), and 3) threshold-based quantile regression considering interaction effects (TBQRI).

doi:10.1371/journal.pone.0106334.g005

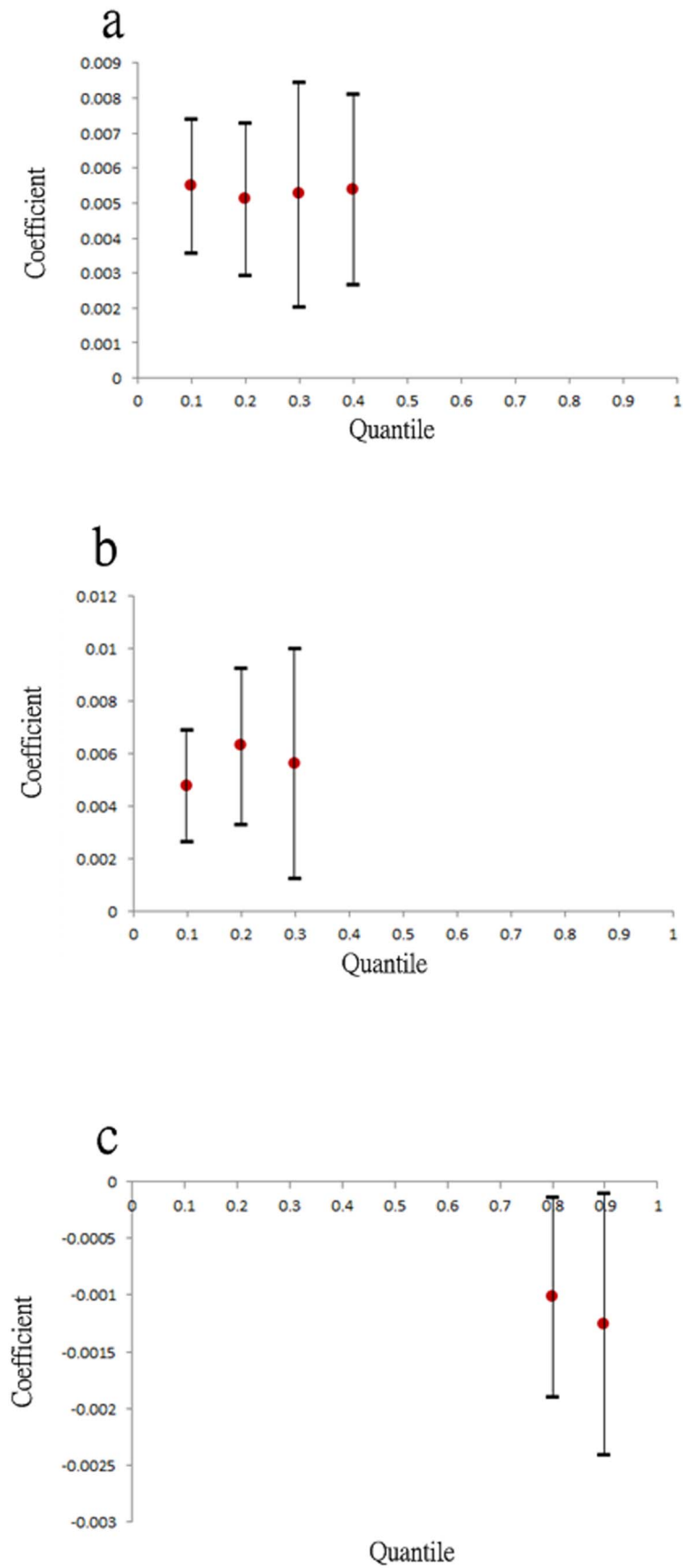


Figure 6. The 95% confident interval at the significant quantile levels for (a) canal (b) ditch, and (c) per capita income, respectively.
 doi:10.1371/journal.pone.0106334.g006

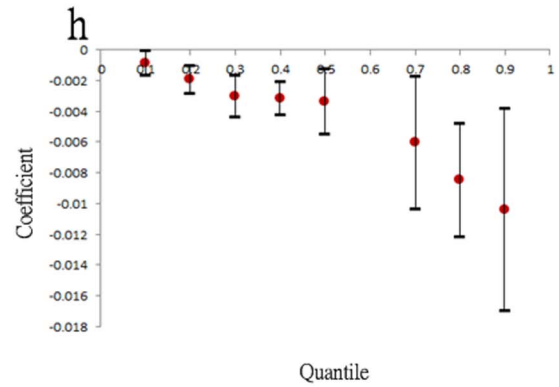
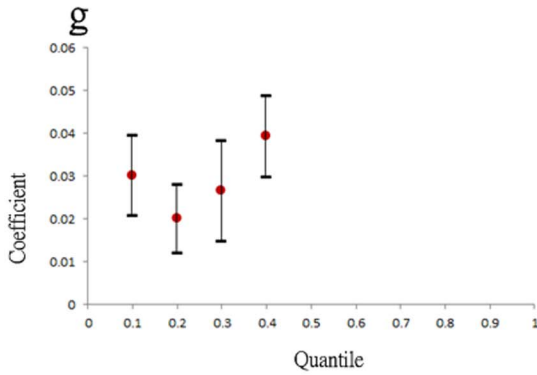
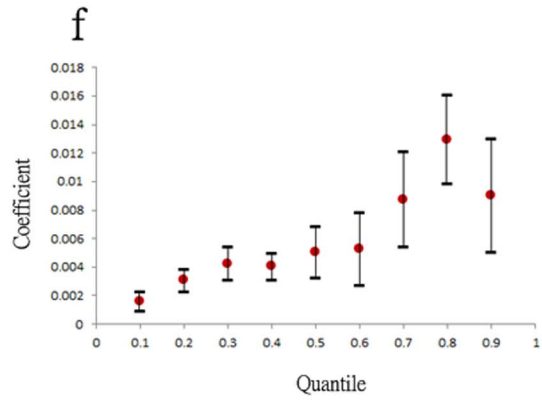
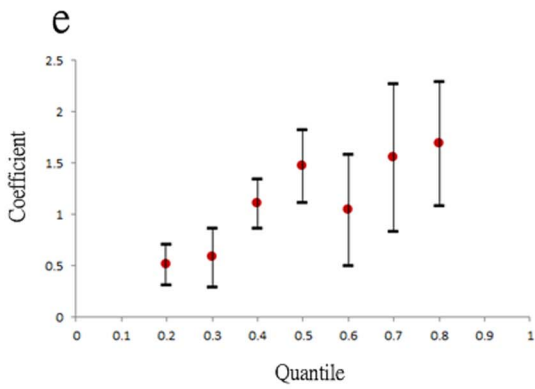
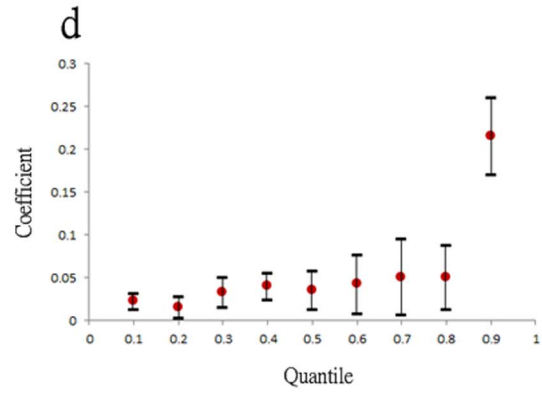
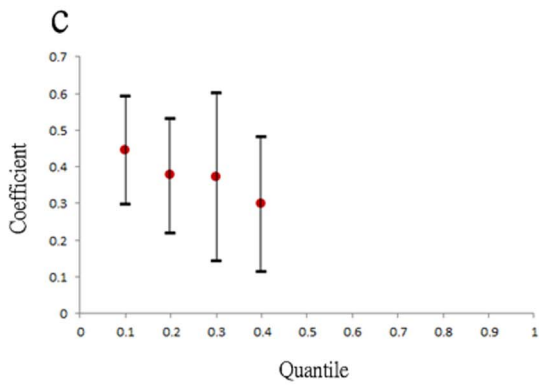
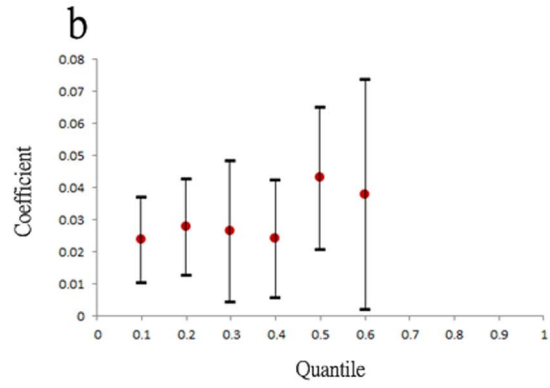
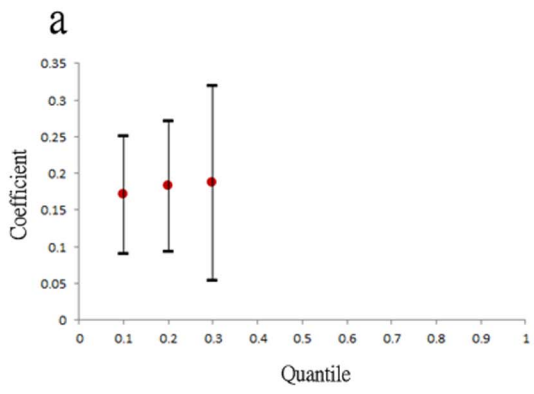


Figure 7. The 95% confident intervals at the significant quantile levels for interaction effects between (a) canal and ditches, (b) canal and residential-business area, (c) canal and residential-other use area, (d) ditch and residential-business area, (e) ditch and residential-other use area, (f) pure residential area and residential-business area, (g) residential-business area and residential-other use area, and (h) residential-business area and per capita income, respectively.

doi:10.1371/journal.pone.0106334.g007

Results

Associations between the spatial distributions of DF incidence rates and risk factors (i.e., land use and socioeconomic attributes) were estimated using quantile regression for every selected quantile level, $\tau = 0.1, 0.2, \dots, 0.9$. Before analysis, Pearson's rho was used to evaluate collinearity instead of the PCA because the pR^2 was below 10%. Following the covariate selection procedure, major DF risk factors for all quantiles were identified, including area ratios of canals, ditches, purely residential areas, residential areas with businesses, residential areas with other uses, warehouses, government agencies, elementary schools, high schools, and the socioeconomic factor, per capita income. Among these risk factors, the most significant were ditches and per capita income, indicating that the effect on DF risk can change among various quantiles (Fig. 4a). Spatial distributions of ditches and per capita income explained approximately 1% to 1.5% and 1% to 2% of DF variations, respectively.

Furthermore, we considered interactions between the risk factors. Figure 4b shows the four most significant interaction factors, among which the interactions between ditches and residential areas contributed to changes in DF risk. Specifically, interactions between ditches and residential-business mixed areas and ditches and residential-other purpose mixed areas accounted for approximately 2% to 5% and 2% to 4%, respectively, of the DF spatial variability among quantile levels. Figure 5 depicts the pR^2 values in the selected quantile levels for the optimal models with three distinct settings (i.e., conventional quantile regression model, threshold-based quantile regression model, and threshold-based quantile regression that considers interaction effects). Consideration of the covariate threshold and interactions among risk factors can significantly improve the explanatory power of the model. In addition, all three models performed more effectively at higher quantile levels (i.e., areas of higher DF risk).

Figure 6 depicts the estimated coefficients and their corresponding 95% confidence intervals for the factors of canals, ditches, and per capita income, which significantly affected DF occurrence at various quantile levels. Among them, canals and ditches were critical to DF incidence in areas with lower incident rates (i.e., $0.1 \leq \tau \leq 0.4$). These land-use effects at the identified quantile levels were approximately constant, that is, a 1% increase in land-use ratios of canals or ditches elevated DF incidence rates by approximately 0.005%. By contrast, per capita income was negatively associated with DF, particularly at relatively high incidence rate quantile levels ($\tau = 0.8, 0.9$). Figure 7 illustrates the identified critical interactions for DF incidence. Generally, the interactions between water-related utilities and residential areas were vital to the DF incidence rate spatial distribution and nonlinearly affected DF occurrence across the entire range of quantile levels, implying that the relationships between these interaction factors and DF risk can change significantly with the DF incidence rate level. Among the interactions, canal-related interactions with ditches, residential-business mixed areas, and residential-other-purposes mixed areas exerted significant effects at low-median quantile levels, (i.e., $0.1 \leq \tau \leq 0.6$). The ditch-related interactions with residential-business mixed areas and residential-other-purposes mixed areas exerted increasing effects when the DF incidence rate quantile levels increased. When

residential-business areas were combined with purely residential areas, higher DF risk in relatively high quantile levels occurred compared with when they were combined with residential-other-purpose areas. Conversely, areas with high residential area per capita income reduced the effects of residential-business areas on DF risk, particularly at the high quantile levels.

The semiparametric CDF of DF incidence rates in the studied 535 *li* were estimated using the threshold-based quantile regression results. Figure 8 depicts the expected DF incidence rates across the area, based on the CDFs, similar to the observed average incident rates in Figure 2. In addition, Figures 9a–c show the incidence rate spatial distributions when the three selected exceedance probabilities were 0.9, 0.5, and 0.1, regarding the return periods (i.e., 1.11, 2, and 10 years). When the return period was 1.11 years, higher DF incidence rates occurred in areas close to canals or ditches. When return periods increased, the high risk areas shifted to areas with a mix of residential areas, canals, and ditches. Spatial distributions of the return periods at the three selected DF incidence cases are shown in Figures 10a–c. Approximately 80% of the *li* had return periods shorter than 2 years when the annual DF incidence rate was one. As the selected annual DF cases increased, the *li* return periods also increased. The *li* with shorter return periods (i.e., 2–5 years) were located in areas with a mix of residential areas, canals, and ditches.

Discussion

This paper proposed a novel probabilistic risk assessment approach to determine the DF risk spatial distributions, using quantile regression. Although quantile regression has been widely applied to analyze nonlinear functional relationships between quantities of concerns and covariates, according to an extensive review of the literature, this study is the first to integrate quantile regression and frequency analysis in probabilistic disease risk assessment. We further accounted for the potential threshold effects of DF occurrence risk factors and expanded conventional quantile regression into a threshold-based model. Using the threshold approach allows for the assumption that risk factors only take effect when they are larger than preselected values [43,44] and is useful in environmental management and policy making [45,46]. The proposed threshold-based quantile regression approach allowed for nonlinearities in both the relationships and predictor values. We not only provided probabilistic-based DF risk maps, but also identified the major risk factors affecting the spatial distribution of DF incidence rates, which were expressed as various return periods that presented the occurrence frequency of prescribed magnitudes. This presentation of DF risk can be interpreted and used to inform the general public and decision makers [29].

We considered land use and socioeconomic factors as proxies for DF disease modeling, because both of these factors are closely associated with DF vector habitats [47–49]. Land use patterns can indicate whether the environmental conditions of surrounding areas are favorable to vector breeding [50], and socioeconomic factors can signal human behavior and activities in the areas; both of these factors have been widely used for DF modeling [19,20,51,52]. Unlike conventional approaches, linear models have been used extensively to identify significant risk factors

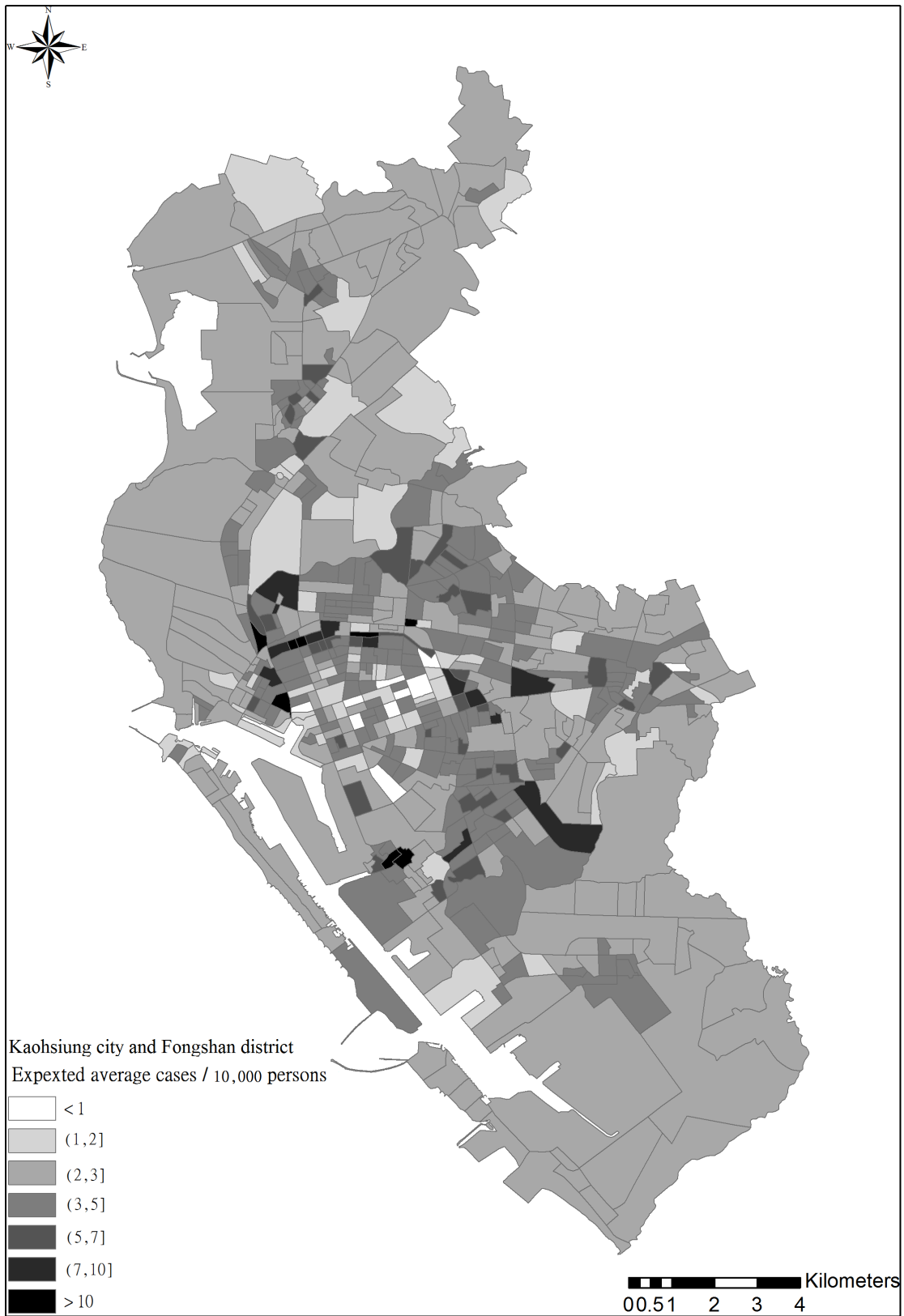


Figure 8. Spatial distribution of expected average incidence rates in the study area.
 doi:10.1371/journal.pone.0106334.g008

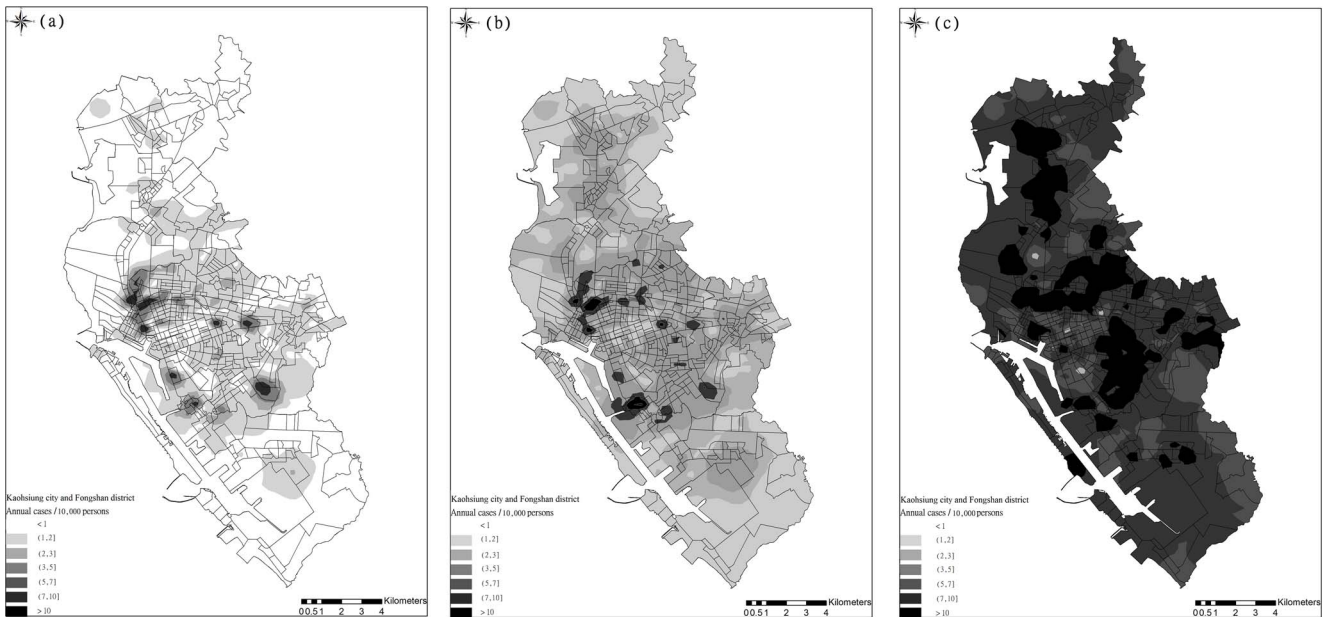


Figure 9. Spatial distribution of DF risks in terms of estimated incidence rates at return period of (a) 1.111 years, (b) 2 years, and (c) 10 years.
doi:10.1371/journal.pone.0106334.g009

[15,16,21,53]. However, the associations between DF incidence and risk factors are not necessarily linear; the magnitudes of these associations depend on the levels of risk factors. We determined the varying associations between significant risk factors and DF incidence in areas with various epidemic conditions, using quantile regression. This study is the first to use quantile regression to analyze the functional relations between DF risk factors and incidence. A two-stage variable selection process (i.e., variable screening and stepwise variable selection) was used to reduce collinearity issues and identify the most influential and explainable

factors in each DF quantile. Although other variable selection or reduction methods are available (e.g., principle component analysis), the proposed variable selection process was used to provide direct relationships between landuse and socioeconomic patterns and DF risk. Therefore, these findings can serve as a reference for governmental agencies when implementing disease control practices in the study area.

Areas including both residential areas and open water channels (e.g., canals and ditches) exhibited elevated DF risk, which is consistent with previous findings [14,16,22,24,39]. The findings in

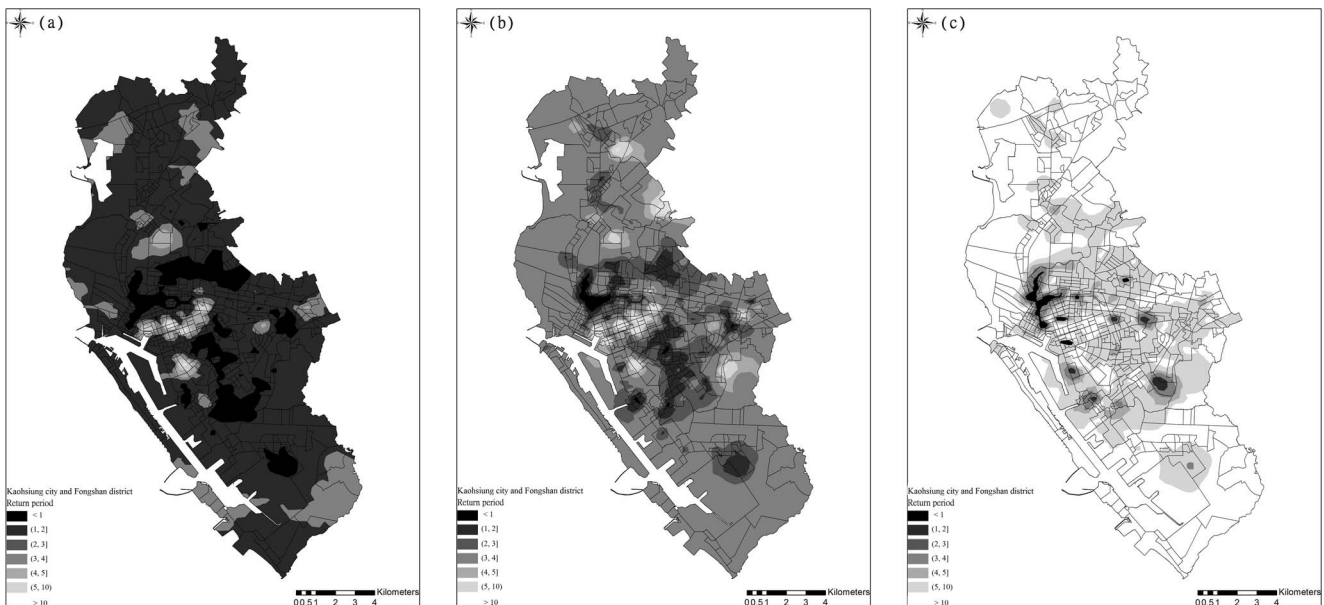


Figure 10. Spatial distribution of DF risks in terms of the estimated return periods that incidence rate is greater than (a) 2 cases/10,000 persons, (b) 1 case/10,000 persons, and (c) 8 case/10,000 persons.
doi:10.1371/journal.pone.0106334.g010

this study indicated that *Aedes aegypti* is the primary vector responsible for DF epidemics in Taiwan, even though both *Aedes aegypti* and *Aedes albopictus* coexist in the country [54,55]. Studies have revealed that water-related facilities are closely associated with an abundance of *Aedes aegypti* [56,57], particularly in areas with numerous water-holding containers [55,58]. In addition, urban areas are generally the preferred habitat of *Aedes aegypti* [59,60], an observation that supports the finding in this study that urbanization level was a highly influential factor that elevated DF risk [24]. The results further determined that residential areas with small businesses had higher DF risk than other types of residential area (e.g., purely residential areas; Fig. 7), possibly because of the relatively high water storage requirements for business activities (e.g., open-air food and grocery markets) [35,36]. DF risk was further elevated by the existence of ditches in the area (Fig. 7). By contrast, high per capita income mitigated DF risk, particularly in residential–business areas in which banks or business offices were located on the lower floors of buildings. The approach used in this study emphasized the nonlinear contributions of environmental factors to DF occurrence that can be overlooked when using conventional regression methods. For example, the significance of water-related facilities to DF occurrence in the study area was only observed in lower quantiles (Figs. 6 and 7), a phenomenon that is easily overlooked when using conventional regression.

Distinct spatial patterns of DF incidence rates regarding various return periods indicated that areas close to ditches and downstream of major canals, (i.e., the Ai river) are generally DF epidemic foci. Figure 9 illustrates the areas with high DF incidence rates in the three selected return periods (i.e., 1.11, 2, and 10 years). These areas were exposed to high DF risk nearly every year, whereas downtown areas, such as Cianjhen District, have longer return periods (i.e., a high DF incidence in this area was observed only once every several years). These results suggested that open water channels in the city increase DF transmission, particularly in the areas downstream of the Ai river where both high-density residential and business areas are mixed. Clustered infections can easily occur in the high-density downtown residential areas once a DF epidemic has occurred. These findings provide insight into previous results that determined that urbanization in Kaohsiung City can increase the DF risk [24]. Another presentation shows the spatial distributions of return periods according to selected DF incidence rates. Most parts of the study area exhibited at least mild DF transmission (i.e., morbidity rate of 1/10000) every 3 years, except for Sinsing District. Areas with short return periods (e.g., 1–2 years), were restricted to areas downstream of the Ai river and those in Cianjhen District once the

DF risk level increased, (e.g., 2/10000). The return period map can provide an alternative view for governmental agencies when estimating the frequency of and contributions to DF occurrence in the area. Canals and ditches are critical factors in the lower quantiles of DF occurrence in Kaohsiung City (Figs. 6 and 7), and therefore, are major risk factors for annual DF outbreaks. By contrast, residential areas were identified as the major factors in extreme DF outbreaks.

Risk analysis based on full probability densities of DF across the entire city can provide a multifaceted view of the associations between spatial distributions of environmental conditions and DF risk. This analysis can be used to emphasize the importance of considering nonlinear functional associations between environmental conditions and DF incidence at various quantile levels. Nonlinearity is crucial to the interpretation of the DF risk spatial distribution. However, the environmental conditions considered in this study (i.e., land use and socioeconomic factors) only partially explained the spatial characteristics of DF occurrence. More specifically, the factors considered in this study were only proxies to characterize the vector habitats and the interactions between humans and vectors. DF incidence can also depend on other direct or indirect risk factors, such as climatic variables, virus serotypes, imported cases, clustered transmissions, and disease control interventions [58,61]. Consequently, the pR^2 of quantile regression models only accounted for approximately 25% of the space-time variations in the disease data. The spatial variation of climatic variables within the study area should have been limited, because of the small size and topographical changes. Other limitations resulting from data uncertainty should also be considered (i.e., the space-time observation scales varied between DF data and proxy factors). Land-use data for Taiwan were only available during 2007, and therefore, no land-use change was considered in this analysis during the study period. In addition, only country-scale socioeconomic data were available; therefore, this study did not include smaller-scale data (i.e., *li*). According to these data uncertainties, the probability and return period maps presented in this study can only provide spatial distribution patterns of DF risk rather than estimations of DF epidemic magnitude.

Author Contributions

Conceived and designed the experiments: CHC HLY. Performed the experiments: CHC HLY. Analyzed the data: CHC HLY. Contributed reagents/materials/analysis tools: CHC THW LCC HLY. Wrote the paper: CHC THW LCC HLY.

References

- Aldstadt J (2007) An incremental Knox test for the determination of the serial interval between successive cases of an infectious disease. *Stochastic Environmental Research and Risk Assessment* 21: 487–500.
- WHO (1997) Dengue haemorrhagic fever: diagnosis, treatment, prevention, and control: World Health Organization.
- Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, et al. (2013) The global distribution and burden of dengue. *Nature*.
- Whitehorn J, Farrar J (2010) Dengue. *British medical bulletin* 95: 161–173.
- Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, et al. (2012) Refining the Global Spatial Limits of Dengue Virus Transmission by Evidence-Based Consensus. *Plos Neglected Tropical Diseases* 6.
- WHO (2009) Dengue and dengue haemorrhagic fever. Fact sheet no 117.
- Normile D (2013) Surprising New Dengue Virus Throws a Spanner in Disease Control Efforts. *Science* 342: 415–415.
- Rodenhuis-Zybert IA, Wilschut J, Smit JM (2010) Dengue virus life cycle: viral and host factors modulating infectivity. *Cellular and molecular life sciences* 67: 2773–2786.
- Maidana NA, Yang HM (2008) Describing the geographic spread of dengue disease by traveling waves. *Mathematical Biosciences* 215: 64–77.
- Rotela C, Fouque F, Lamfri M, Sabatier P, Introini V, et al. (2007) Space-time analysis of the dengue spreading dynamics in the 2004 Tartagal outbreak, Northern Argentina. *Acta Tropica* 103: 1–13.
- Wen T-H, Lin NH, Lin C-H, King C-C, Su M-D (2006) Spatial Mapping of Temporal Risk Characteristics to Improve Environmental Health Risk Identification: A Case Study of a Dengue Epidemic in Taiwan. *The Science of the Total Environment* 367: 631–640.
- Yu HL, Yang SJ, Yen HJ, Christakos G (2011) A spatio-temporal climate-based model of early dengue fever warning in southern Taiwan. *Stochastic Environmental Research and Risk Assessment* 25: 485–494.
- Stewart-Ibarra AM, Lowe R (2013) Climate and Non-Climate Drivers of Dengue Epidemics in Southern Coastal Ecuador. *American Journal of Tropical Medicine and Hygiene* 88: 971–981.
- Lifson AR (1996) Mosquitoes, models, and dengue. *Lancet* 347: 1201–1202.
- Vanwambeke SO, Lambin EF, Eichhorn MP, Flasse SP, Harbach RE, et al. (2007) Impact of land-use change on dengue and malaria in northern Thailand. *Ecohealth* 4: 37–51.
- Mondini A, Chiaravalloti-Neto F (2008) Spatial correlation of incidence of dengue with socioeconomic, demographic and environmental variables in a Brazilian city. *Science of the Total Environment* 393: 241–248.

17. Gubler DJ, Clark GG (1995) Dengue Hemorrhagic-Fever - the Emergence of a Global Health Problem. *Emerging Infectious Diseases* 1: 55–57.
18. Dowling Z, Armbruster P, LaDeau SL, DeCotiis M, Mottley J, et al. (2013) Linking Mosquito Infestation to Resident Socioeconomic Status, Knowledge, and Source Reduction Practices in Suburban Washington, DC. *Ecohealth* 10: 36–47.
19. Van Benthem BHB, Vanwambeke SO, Khantikul N, Burghoorn-Maas C, Panart K, et al. (2005) Spatial patterns of and risk factors for seropositivity for dengue infection. *American Journal of Tropical Medicine and Hygiene* 72: 201–208.
20. Raju K, Sokhi B (2008) Application of GIS modeling for dengue fever prone area based on socio-cultural and environmental factors—a case study of Delhi city zone. *Int Arch Photogramm Remote Sens Spat Inf Sci* 37: 165–170.
21. Braga C, Luna CF, Martelli CM, Souza WVd, Cordeiro MT, et al. (2010) Seroprevalence and risk factors for dengue infection in socio-economically distinct areas of Recife, Brazil. *Acta tropica* 113: 234–240.
22. Gharbi M, Quenel P, Gustave J, Cassadou S, La Ruche G, et al. (2011) Time series analysis of dengue incidence in Guadeloupe, French West Indies: Forecasting models using climate variables as predictors. *Bmc Infectious Diseases* 11.
23. Gubler D (1998) The global pandemic of dengue/dengue haemorrhagic fever: current status and prospects for the future. *Annals of the Academy of Medicine, Singapore* 27: 227–234.
24. Wu PC, Lay JG, Guo HR, Lin CY, Lung SC, et al. (2009) Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical Taiwan. *Science of the Total Environment* 407: 2224–2233.
25. Hales S, Weinstein P, Soares Y, Woodward A (1999) El Niño and the dynamics of vectorborne disease transmission. *Environmental Health Perspectives* 107: 99.
26. Kumamoto H, Henley EJ, Henley EJ (1996) Probabilistic risk assessment and management for engineers and scientists. New York: IEEE Press. xvii, 597 p. p.
27. Bari RA (2003) Probabilistic risk assessment: Springer.
28. Vesely WE (2011) Probabilistic Risk Assessment. *System Health Management: With Aerospace Applications*: 253–263.
29. O'Hagan A (2006) Uncertain judgements : eliciting experts' probabilities. London ; Hoboken, NJ: Wiley. xiii, 321 p. p.
30. Mays LW (2010) Water resources engineering: Wiley. com.
31. Benjamin JR (1970) Probability, statistics, and decision for civil engineers.
32. Wu PC, Guo HR, Lung SC, Lin CY, Su HJ (2007) Weather as an effective predictor for occurrence of dengue fever in Taiwan. *Acta Tropica* 103: 50–57.
33. Taiwan CDC (2013) Procedures of data application. In: Welfare MoHa, editor. Governmental Data Taipei, Taiwan: Centers of Disease Control, Taiwan.
34. NLSA (2013) Land Use Investigation of Taiwan.
35. National Land Surveying and Mapping Center (2007) Construction use land taxonomy table. In: Center NLSaM, editor. Landuse Investigation of Taiwan. Taichung, Taiwan: Ministry of Interior.
36. National Land Surveying and Mapping Center (2012) Land use classification system doubt casebook. In: National Land Surveying and Mapping Center, editor. Taichung, Taiwan: Ministry of Interior.
37. Koehler R, Bassett G (1978) Regression Quantiles. *Econometrica* 46: 33–50.
38. Mata J, Machado JA (1996) Firm start-up size: A conditional quantile approach. *European Economic Review* 40: 1305–1323.
39. Hales S, Weinstein P, Soares Y, Woodward A (1999) El Niño and the dynamics of vectorborne disease transmission. *Environmental Health Perspectives* 107: 99–102.
40. Montgomery DC, Peck EA, Vining GG (2012) Introduction to linear regression analysis: Wiley.
41. Akaike H (1987) Factor analysis and AIC. *Psychometrika* 52: 317–332.
42. Koehler R, Koehler MR (2013) Package 'quantreg'.
43. Kozlowski J (1985) Threshold approach in environmental planning. *Ekistics* 52: 146–153.
44. Kozlowski J, Rosier J, Hill G (1988) Ultimate environmental threshold (UET) method in a marine environment (Great Barrier Reef Marine Park in Australia). *Landscape and urban planning* 15: 327–336.
45. Kim SY, Lee JT, Hong YC, Ahn KJ, Kim H (2004) Determining the threshold effect of ozone on daily mortality: an analysis of ozone and mortality in Seoul, Korea, 1995–1999. *Environmental Research* 94: 113–119.
46. Daniels MJ, Dominici F, Samet JM, Zeger SL (2000) Estimating particulate matter-mortality dose-response curves and threshold levels: An analysis of daily time-series for the 20 largest US cities. *American Journal of Epidemiology* 152: 397–406.
47. Lei H-Y, Huang J-H, Huang K-J, Chang C (2002) Status of dengue control programme in Taiwan-2001. *Dengue Bulletin* 26: 14–23.
48. Vanwambeke SO, Bennett SN, Kapan DD (2011) Spatially disaggregated disease transmission risk: land cover, land use and risk of dengue transmission on the island of Oahu. *Tropical Medicine & International Health* 16: 174–185.
49. Sarfraz MS, Tripathi NK, Tipdecho T, Thongbu T, Kerdtong P, et al. (2012) Analyzing the spatio-temporal relationship between dengue vector larval density and land-use using factor analysis and spatial ring mapping. *BMC public health* 12: 853.
50. Vanwambeke SO, Somboon P, Harbach RE, Isenstadt M, Lambin EF, et al. (2007) Landscape and land cover factors influence the presence of Aedes and Anopheles larvae. *Journal of Medical Entomology* 44: 133–144.
51. Kienberger S, Hagenlocher M, Delmelle E, Casas I (2013) A WebGIS tool for visualizing and exploring socioeconomic vulnerability to dengue fever in Cali, Colombia. *Geospatial health* 8: 313–316.
52. Hagenlocher M, Delmelle E, Casas I, Kienberger S (2013) Assessing socioeconomic vulnerability to dengue fever in Cali, Colombia: statistical vs expert-based modeling. *International journal of health geographics* 12: 36.
53. Almeida MCD, Caiaffa WT, Assuncao RM, Proietti FA (2007) Spatial vulnerability to dengue in a Brazilian urban area during a 7-year surveillance. *Journal of Urban Health-Bulletin of the New York Academy of Medicine* 84: 334–345.
54. Chang SF, Huang JH, Shu PY (2012) Characteristics of dengue epidemics in Taiwan. *Journal of the Formosan Medical Association* 111: 297–299.
55. Tuan Y-C, Hung M-N, Lin L-J, Shin W-Y, Huang C-C, et al. (2009) Analysis on Dengue Vector Density Survey in Kaohsiung and Pingtung Areas of Southern Taiwan, 2004–2008. *Taiwan Epidemiology Bulletin* 25: 462–485.
56. Saifur RG, Hassan AA, Dieng H, Salmah MRC, Saad AR, et al. (2013) Temporal and Spatial Distribution of Dengue Vector Mosquitoes and their Habitat Patterns in Penang Island, Malaysia. *Journal of the American Mosquito Control Association* 29: 33–43.
57. Hiscox A, Kaye A, Vongphayloth K, Banks I, Piffer M, et al. (2013) Risk Factors for the Presence of Aedes aegypti and Aedes albopictus in Domestic Water-Holding Containers in Areas Impacted by the Nam Theun 2 Hydroelectric Project, Laos. *The American journal of tropical medicine and hygiene* 88: 1070–1078.
58. Shang CS, Fang CT, Liu CM, Wen TH, Tsai KH, et al. (2010) The Role of Imported Cases and Favorable Meteorological Conditions in the Onset of Dengue Epidemics. *Plos Neglected Tropical Diseases* 4.
59. Tsuda Y, Suwonkerd W, Chawprom S, Prajakwong S, Takagi M (2006) Different spatial distribution of Aedes aegypti and Aedes albopictus along an urban-rural gradient and the relating environmental factors examined in three villages in northern Thailand. *Journal of the American Mosquito Control Association* 22: 222–228.
60. Hiscox A, Kaye A, Vongphayloth K, Banks I, Piffer M, et al. (2013) Risk Factors for the Presence of Aedes aegypti and Aedes albopictus in Domestic Water-Holding Containers in Areas Impacted by the Nam Theun 2 Hydroelectric Project, Laos. *American Journal of Tropical Medicine and Hygiene* 88: 1070–1078.
61. Knerer G, Currie CS, Brailsford SC (2013) Impact of combined vector-control and vaccination strategies on transmission dynamics of dengue fever: a model-based analysis. *Health care management science*: 1–13.