BMC Medical Research Methodology

RESEARCH

Open Access

A stabilized spatiotemporal kriging method for disease mapping and application to male oral cancer and female breast cancer in Taiwan

Dai-Rong Tsai[1†], Jing-Rong Jhuang[1,2†], Shih-Yung Su[1], Chun-Ju Chiang[1,2], Ya-Wen Yang[2] and Wen-Chung Lee[1,2*]

Abstract

Mapping spacetime disease rates can provide a more in-depth understanding of their distribution and trends. Traditional spatiotemporal kriging methods can break the constraints of geopolitical boundaries and time intervals. Still, disease rates in densely and sparsely populated areas are stabilized to the same degree, resulting in a map that is oversmoothed in some places but undersmoothed in others. The stabilized spatiotemporal kriging method proposed in this study overcomes this problem by allowing for nonconstant variances over space and time. A spatiotemporal map of the standardized incidence ratio for oral cancer in men in Taiwan between 1997 and 2017 reveals that the high-risk areas for oral cancer are in the midwestern and southeastern regions of Taiwan, spreading toward the center and north, with persistent cold spots in the northern and southwestern urban regions. However, the corresponding map for breast cancer in women in Taiwan reveals that the high-risk areas for breast cancer are concentrated in densely populated urban regions in the west. Spatiotemporal maps facilitate our understanding of disease risk dynamics. We recommend using the proposed stabilized spatiotemporal kriging method for mapping disease rates across space and time.

Keywords: Stabilized spatiotemporal kriging, Disease map, Incidence, Oral cancer, Breast cancer

Introduction

Disease mapping has long been a part of public health. Epidemiologists often construct disease maps to investigate the geographical variation in diseases. The choropleth map is the widely used map type for mapping disease rates, using local administrative areas (LAAs) such as counties, cities, or towns as geographic units. The isarithmic map was another common map type, using latitude and longitude as geographic units and breaking the constraints of geopolitical boundaries. Tracking the disease rates of each geographic unit over several years constitutes spatiotemporal data. The three-dimensional spatiotemporal map visualizes the distribution of disease risks over space and time, which can help detect possible risk factors and form the basis of health policies.

Mapping disease rates with raw data and visualizing with the choropleth map may incur too much spatial variability; in LAAs with smaller populations, the disease rates may become highly unstable. Empirical Bayes methods stabilize the rates in LAAs adaptively according to their population size [1]. Conditional autoregressive models further consider spatial correlations between LAAs [2]. The spatial random effects model involves the benefits of the abovementioned methods and can deal with confounding factors [3, 4]. These methods can also incorporate "time" as a covariate or as a random effect to deal with the variations in disease risks in the time domain [3, 4].

†Dai-Rong Tsai and Jing-Rong Jhuang contributed equally to this work.

*Correspondence: wenchung@ntu.edu.tw

[1] Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Rm. 536, No. 17, Xuzhou Rd., Taipei 100, Taiwan
Full list of author information is available at the end of the article

Mapping disease rates as a continuous surface and visualizing with the isarithmic map can more legibly depict the spatiotemporal patterns of disease risks. Trend surface analysis and kernel smoothing created a continuous surface of disease rates [5–7]. The spatiotemporal kriging method has been widely used in environmental and earth sciences [8–10]. Epidemiologists also applied the kriging method to investigate the spread of infectious diseases [11–13], the distribution of environmental risk factors [14, 15], and the spatiotemporal variations of disease rates [16, 17]. The kriging method estimates the spatiotemporal correlations based on a nonparametric model and extrapolates the disease rates at any given point in space and time. The method breaks the constraints of geopolitical boundaries and time intervals [18–20]. However, because of the homoscedasticity assumption, the kriging method stabilizes the disease rates in LAAs with different population sizes to the same degree, resulting in a map that is oversmoothed in some places but undersmoothed in others.

This study proposed a stabilized spatiotemporal kriging method for disease mapping. This method accounts for the heteroscedasticity in the spatiotemporal data, permitting the disease rates to be shrunk toward the mean to different degrees, thus avoiding oversmoothing and undersmoothing. Monte Carlo simulations were conducted to assess the properties of the proposed method compared with the traditional kriging method. Real data applications for oral cancer in men and breast cancer in women in Taiwan were also performed for demonstration.

Methods

Stabilized spatiotemporal kriging

Let the observed and expected number of cases for a particular disease in the i th LAA of the j th year be O_{ij} and E_{ij}, respectively $(i = 1, \ldots, N;\ j = 1, \ldots, M)$. The expected number of cases can be calculated as

$$E_{ij} = \sum_{k=1}^{K} p_{ijk} \times R_k,$$

where p_{ijk} is the population size in the i th LAA of the j th year for the k th age group and R_k is the incidence rate for the k th age group of a standard population $(k = 1, \ldots, K)$. The standardized incidence ratio, SIR, and its natural logarithm, logSIR, are then estimated as $\widehat{r}_{ij} = \frac{O_{ij}}{E_{ij}}$ and $\widehat{\theta}_{ij} = log_e\left(\widehat{r}_{ij}\right)$, respectively. We assume a linear mixed model as follows:

$$\widehat{\theta}_{ij} = \theta_{ij} + \varepsilon_{ij} = \mu + \delta_{ij} + \varepsilon_{ij}, \tag{1}$$

where θ_{ij} is the natural logarithm of the true standardized incidence ratio (logSIR) and ε_{ij} is the measurement error, which is unbiased and independent of one another; thus, $E\left(\varepsilon_{ij}\right) = 0$ $(i = 1, \ldots, n;\ j = 1, \ldots, m)$ and $Cov\left(\varepsilon_{ij}, \varepsilon_{kl}\right) = 0$ $(ij \neq kl)$. The θ_{ij} is the sum of the spatiotemporal mean (μ, a fixed effect) and the spatiotemporal variations (δ_{ij}, random effects with a mean of zero). The δ_{ij} is second-order stationary and spatially isotropic; accordingly, the variance of θ_{ij} is a constant τ^2 and the covariance between θ_{ij} and θ_{kl} ($C_{ij,kl}$, $ij \neq kl$) depends only on the spatial distance (h_{ik}; unit: km) between the centroids of the i th and k th LAAs and the temporal distance (u_{jl}; units: year) between the j th and l th years. Commonly used distance metric in health research involved straight line (Euclidean) distance and Manhattan distance, and, in this study, we used the Euclidean distance.

Data consists of $n \times m$ couples of $\left(\widehat{\theta}_{ij}, \widehat{v}_{ij}\right)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. The $\widehat{v}_{ij}$, the estimation of $Var\left(\varepsilon_{ij}\right)$, is $\frac{1}{O_{ij}}$ (Additional file 1: Appendix 1). Under the formula (1), a best linear unbiased predictor for the logSIR in a particular locality (subscript s; this can be an LAA centroid or any noncentroid geographical coordinate) at one specific time point (subscript t; this can be the middle or any time point in any year) is,

$$\widehat{\theta}_{st} = \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} \widehat{\theta}_{ij},$$

and the weights w_{ij} are

$$\begin{bmatrix} w_{11} \\ w_{12} \\ \vdots \\ w_{nm} \\ \lambda \end{bmatrix} = \begin{bmatrix} \tau^2 + \widehat{v}_{11} & C_{11,12} & \cdots & C_{11,nm} & 1 \\ C_{12,11} & \tau^2 + \widehat{v}_{12} & \cdots & C_{12,nm} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{nm,11} & C_{nm,12} & \cdots & \tau^2 + \widehat{v}_{nm} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} C_{11,st} \\ C_{12,st} \\ \vdots \\ C_{nm,st} \\ 1 \end{bmatrix}, \tag{2}$$

where λ is the Lagrange multiplier to ensure $\sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} = 1$.

The parameters τ^2 and $C_{ij,kl}$ $(C_{ij,kl} = C_{kl,ij};\ i, k = 1, \ldots, n;\ j, l = 1, \ldots, m;\ ij \neq kl)$ in formula (2) can be estimated from a semivariogram. First, we calculate a total of $\frac{1}{2} \times n \times m \times (n \times m - 1)$ paired sample semivariances: $\frac{1}{2}\left(\widehat{\theta}_{ij} - \widehat{\theta}_{kl}\right)^2$ for $i, k = 1, \ldots, n;\ j, l = 1, \ldots, m;$ $ij \neq kl$. We then produce the semivariogram by plotting these semivariances against the spatial distance h_{ik} and temporal distance u_{jl} between any pair of distinct spatiotemporal data points. We then fit a spatiotemporal semivariance function $\gamma(h, u)$, which monotonically increases as h and u increase, to the semivariogram based on generalized least squares. Herein, we use the "sum-metric" spatiotemporal semivariance structure detailed in [21]: $\gamma(h,u) = \gamma_S(h) + \gamma_T(u) + \gamma_{ST}\left[\sqrt{h^2 + (\alpha \times u)^2}\right]$, where $\gamma_S(\cdot)$, $\gamma_T(\cdot)$, and $\gamma_{ST}(\cdot)$ are the semivariance functions for space, time, and spacetime interactions, respectively, and α is a parameter that integrates the distances

in space and time into a single spacetime distance metric. Note that the functions $\gamma_S(\cdot)$, $\gamma_T(\cdot)$, and $\gamma_{ST}(\cdot)$ must increase monotonically and have upper limits so that the spatiotemporal semivariance function $\gamma(h,u)$ can approach a maximum (referred to as the "sill") when the distances in space and time become extremely large. Commonly used semivariance functions include the spherical, exponential, and Gaussian functions.

We derive the covariance function $C(h,u)$ by subtracting the semivariance function from the sill. This function is then a monotonically decreasing function of distances in both space and time. The parameters in formula (2) are then estimated as $\widehat{\tau}^2 = C(0,0)$, $\widehat{C}_{ij,kl} = C\left(h_{ik}, u_{jl}\right)$, and $\widehat{C}_{ij,st} = C\left(h_{is}, u_{jt}\right)$, respectively, for $i,k = 1,\ldots,n$; $j,l = 1,\ldots,m$; $ij \neq kl$.

We refer to this method as stabilized spatiotemporal kriging. The key to stabilization is that with the error variances, $v_{11},\ldots,v_{nm}$, added to the diagonal of the square matrix in formula (2), smaller weights are given to data with larger variances and larger weights to those with smaller variances. The approach extends the stabilized kriging method in a purely spatial setting, as proposed by Hsu et al. [22], to deal with spacetime data.

The stabilized spatiotemporal kriging method can be implemented through the following steps:

1. Produce a semivariogram and fit the semivariance function. This can be achieved using the same software package used for traditional spatiotemporal kriging.
2. Subtract the semivariance function from the sill (the maximum of the semivariance function) and derive the covariance function.
3. Estimate the parameters τ^2, $C_{ij,kl}$, and $C_{ij,st}$ ($C_{ij,kl} = C_{kl,ij}$; $i,k = 1,\ldots,n$; $j,l = 1,\ldots,m$; $ij \neq kl$) from a direct readout of the values from the covariance function.
4. Calculate the kriging weight for any spatiotemporal point to be estimated. This can be done using any statistical package that supports formula (2) matrix calculations.
5. Calculate the weighted average and then use exponentiation to obtain the kriging estimate of SIR at that spatiotemporal point.

We further use the annual percent change (APC) to measure the trend of SIR over time. We recommend using Cole's definition of the symmetric percentage change [23]. The APC at a certain spatiotemporal point, st, can be obtained as follows:

1. Calculate the kriging estimates of logSIR at the spatial point s within a 5-year time window centered at the temporal point t.
2. Perform a simple linear regression of the kriged logSIR against time (in years) to obtain the estimate of the slope $\widehat{\beta}$.
3. The APC at the spatiotemporal point st is $\widehat{\beta} \times 100(\%)$.

A positive APC indicates that the value of SIR is increasing, a negative APC that the value is declining, and an APC near zero that the value remains stable. Note that the conventional definition of APC leads to overt asymmetry; the value comparing two consecutive years is $\left[\exp\left(\widehat{\beta}\right) - 1\right] \times 100\%$ (with a lower limit of -100% and without an upper limit) with the first year as the reference, but is $\left[1 - \exp\left(-\widehat{\beta}\right)\right] \times 100\%$ (with an upper limit of 100% and without a lower limit) with the second year as the reference. However, Cole's definition of APC is always $\widehat{\beta} \times 100\%$.

Monte Carlo simulation

We generated 4417×21 spatiotemporal points (with longitudes between 120.00° and 121.95° and latitudes between 21.92° and 25.27° from 2000 to 2020) to encompass the entire Taiwan island (Additional file 1: Appendix 2). We considered a single-hotspot scenario at coordinate point h and a double-hotspots scenario at coordinate points h and h*, respectively (Additional file 1: Appendix 3). We set the true value of the logSIR for each spatiotemporal point according to the two-dimensional Gaussian function, which diminishes in value with distance and time, as follows:

1. A single hotspot at coordinate point h

$$\theta_{st} = \log(1.0) + \exp\left(-\frac{1}{2}\left(\frac{D_{s,h}}{50 - 9t}\right)^2\right),$$

2. Double hotspots at coordinate points h and h*, respectively,

$$\theta_{st} = \log(1.0) + \exp\left(-\frac{1}{2}\left(\frac{D_{s,h}}{50 - 9t}\right)^2\right) + \exp\left(-\frac{1}{2}\left(\frac{D_{s,h^*}}{50 - 9t}\right)^2\right),$$

where $D_{a,b}$ denotes the distance between two coordinate points a and b (unit: km).

We used a log-normal distribution to generate the population size of the 349 LAAs in Taiwan. We considered the expected values of the population size were 50,000 and 25,000; the coefficients of variation of the population size were 0.5 and 1.0. For an LAA, we multiplied its simulated population size by the simulated disease rate (assumed to follow an exponential distribution) at the centroid of that LAA to obtain the expected number of cases (E_{ij}). We then used a Poisson distribution (with a mean of $E_{ij} \times e^{\theta_{ij}}$) to generate the observed number of cases (O_{ij}) for each LAA. The estimated logSIR ($\widehat{\theta}_{ij}$) of the LAA is the logarithm of the ratio between the observed and expected numbers of cases.

We used the stabilized spatiotemporal kriging method and the traditional spatiotemporal kriging method to estimate the logSIR of all coordinate points within the geopolitical boundary of Taiwan based on the simulated data. We calculated the index of symmetric mean absolute percentage error (SMAPE):

$$\text{SMAPE} = \frac{100\%}{21 \times 4417} \sum_{t=1}^{21} \sum_{s=1}^{4417} \frac{|e^{\widehat{\theta}_{st}} - e^{\theta_{st}}|}{e^{\widehat{\theta}_{st}} + e^{\theta_{st}}}.$$

The smaller the SMAPE index, the closer the estimated value is to the true value. We performed 1,000 simulations for each scenario.

All analyses were performed using SAS version 9.4 and R version 3.5.2. R packages, including gstat, sp, and spacetime, were used to perform the spatiotemporal kriging method.

Results

Simulation results

The simulation results are presented in Table 1. The SMAPEs increased with decreasing population size and increasing coefficient of variation for both methods. For the single-hotspot scenario, the stabilized spatiotemporal kriging method and the traditional kriging method had similar performance. For the double-hotspots scenario, the stabilized spatiotemporal kriging method performed better (having smaller SMAPE) than the traditional spatiotemporal kriging method in all combinations of population size and coefficient of variation.

Real data applications

We used oral cancer in men and breast cancer in women in Taiwan as examples to demonstrate the methodology. The data of male patients with oral cancer and female patients with breast cancer from 1997 to 2017, including their ages and the LAAs they resided in, were extracted from the Taiwan Cancer Registry [24, 25]. The age-specific mid-year population numbers (the averages of the end-year population numbers in two consecutive years) for men and women in every LAA in Taiwan from 1997 to 2017 were extracted from an online database provided by the Department of Household Registration in Taiwan's Ministry of the Interior. We focused our analysis on the 349 LAAs on the main island of Taiwan (the 19 LAAs in the outlying islands were excluded). The patients' ages were divided into 18 age groups: 0–4, 5–9, ..., 80–84, and 85+. The average incidence rate for each age group on the main island over 21 years (1997–2017) was taken as the age-specific rate of the standard population. A total of $2 \times 349 \times 21 = 14,658$ SIRs (2 cancer types, 349 LAAs, 21 years) were then calculated as described previously. In this study, the representative point of an LAA was taken to be its geometric center and the middle of each year, July 1.

We generated a square mesh coordinate point system with a 1-km spacing (69,084 coordinate points, with longitudes between 120.00° and 122.02° and latitudes between 21.89° and 25.31°) to encompass the entire main island of Taiwan. We then used the proposed stabilized spatiotemporal kriging method to estimate the SIRs of all the coordinate points on the main island (a total of 32,263 coordinate points) for a total of $21 \times 12 = 252$ months (each month represented by its first day) based on the available data of $349 \times 21 = 7,329$ couples of $\left(\widehat{\theta}_{ij}, \widehat{\nu}_{ij}\right)$ for $i = 1, \ldots, 349$ and $j = 1, \ldots, 21$. We applied the sum-metric structure and used the spherical spatiotemporal semivariance

Table 1 The symmetric mean absolute percentage errors (%) of the traditional and stabilized spatiotemporal kriging in the scenarios of a single hotspot and double hotspots

		A single hotspot		Double hotspots	
Population size	**Coefficient of variation**	**Traditional spatiotemporal kriging**	**Stabilized spatiotemporal kriging**	**Traditional spatiotemporal kriging**	**Stabilized spatiotemporal kriging**
25,000	0.5	21.18	21.02	26.11	20.75
25,000	1.0	21.56	21.36	33.31	28.87
50,000	0.5	15.26	15.12	24.53	20.69
50,000	1.0	18.23	18.08	24.99	20.14

function. We also calculated the APC at each spatiotemporal point using kriged SIRs. Finally, we produced the spatiotemporal SIR and APC maps for oral cancer in men (Figs. 1 and 2; Additional file 1: Appendix 4) and breast cancer in women in Taiwan (Figs. 3 and 4; Additional file 1: Appendix 5).

The SIR map for oral cancer in men (Fig. 1) reveals that before 2000, people in most parts of Taiwan, except for the midwestern and southeastern regions, were at low risk of oral cancer, with cold spots in the north and southwest. After 2000, high-risk areas began to emerge in the Central and Southern mountainous region and the Eastern coastal area, connecting with the midwestern and southeastern high-risk regions to form a hot zone. The hot zone intensified over time and spread to the center and north. The risk in the Central and Northern regions changed from low to high after 2008. The risk of oral cancer in the northern and southwestern urban regions remained low over time.

The corresponding APC map for oral cancer in men (Fig. 2) indicates that the risk of oral cancer in Taiwan rose rapidly before 2005. The Central and Southern mountainous region and the Eastern coastal area had APC hot spots between 2001 and 2004, coinciding with the SIR hot zone formation mentioned above. After 2005, the APC in the whole of Taiwan began to decline, indicating that the increase in the risk of oral cancer had slowed down. The APC in the Central and Northern regions first decreased and then increased, indicating that the high-risk area of oral cancer had spread to the north. After 2015, the APC in Central and Southern Taiwan dropped below zero, indicating that oral cancer risk there had peaked and then declined.

The SIR map for breast cancer in women (Fig. 3) reveals that before 2005, the whole population of Taiwan was at low risk of breast cancer, but the cold spots were gradually shrinking. In 2005, an area of slightly elevated risk first appeared in the north, and later, hot spots emerged one after another in the Western, northeastern, and southeastern regions. These hot spots intensified and

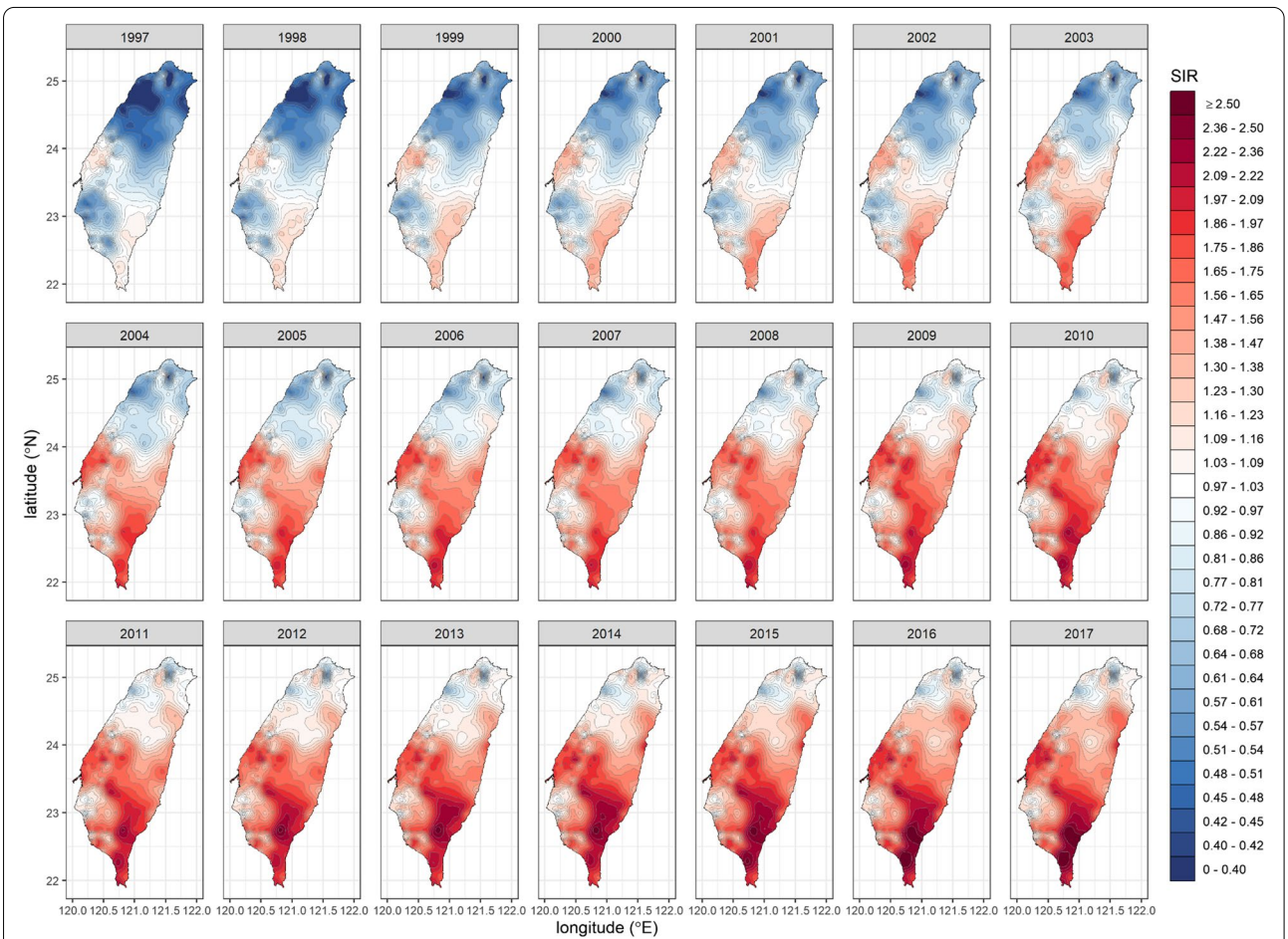

Fig. 1 The stabilized spatiotemporal kriging map for the standardized incidence ratios (SIRs) of oral cancer in men in Taiwan, 1997 to 2017

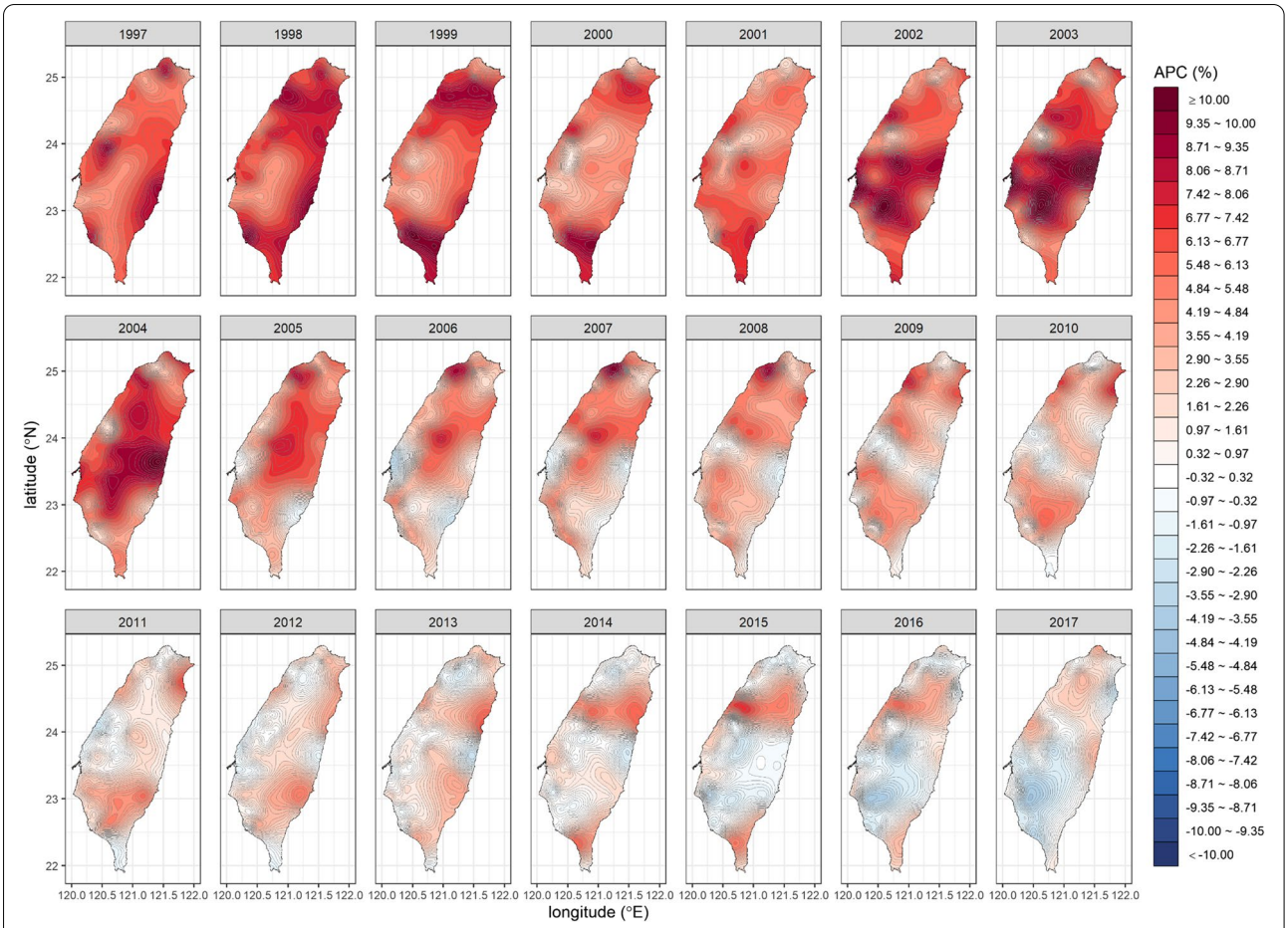

Fig. 2 The stabilized spatiotemporal kriging map for the annual percent changes (APCs) in the standardized incidence ratios of oral cancer in men in Taiwan, 1997 to 2017

spread outward over time. Note that the high-risk areas of breast cancer tend to be concentrated in densely populated urban regions in the west. In contrast, the risks in the Western and Eastern coastal areas and the Central mountainous region remain low.

The APC map for breast cancer in women (Fig. 4) indicates that the risk of breast cancer in Taiwan has been rising over the last 21 years. The APC hot spots in the Northern region between 2005 and 2008 corresponded to the emergence and growth of the SIR hot spots in that region. The APC varied greatly over space and time, with the APC hot spots mostly occurring in the Western and Eastern urban areas and the Southern mountainous region. After 2011, the APC stabilized, although an APC hot spot appeared briefly in the southeastern region.

The spatiotemporal dynamic maps (Additional file 1: Appendices 4 and 5) clarify our understanding of the nuances and trends of disease risk dynamics.

Discussion

The stabilized spatiotemporal kriging proposed in this study breaks the constraints of geopolitical boundaries and time intervals and stabilizes the disease rates properly. Hsu et al. [22] proposed the stabilized kriging method for spatial mapping. Their simulation results showed the benefits of the stabilized kriging method for spatial mapping of disease rates over the empirical Bayes and the traditional kriging methods. The spatiotemporal mapping (proposed in this study) is similar to the stabilized kriging method by Hsu et al.; however, in our proposed method, the additional variations in the time domain and those arising from spacetime interactions must also be considered. The simulation results of this study showed that the stabilized spatiotemporal kriging method improves on the traditional spatiotemporal kriging method when the spacetime interaction exists.

Disease rates are unstable in areas with small populations and appear extremely high or low. Therefore,

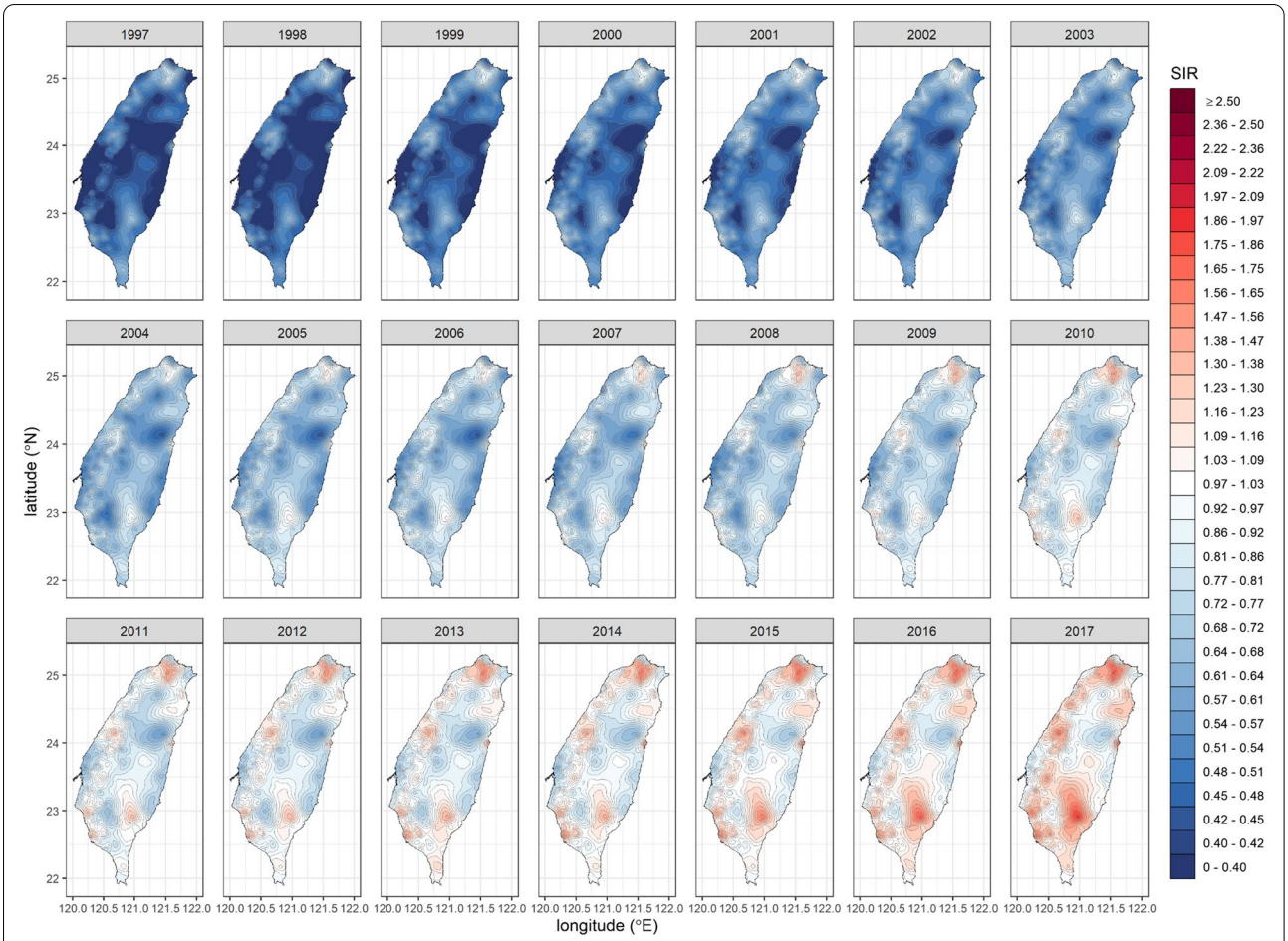

Fig. 3 The stabilized spatiotemporal kriging map for the standardized incidence ratios (SIRs) of breast cancer in women in Taiwan, 1997 to 2017

the LAA-based spatiotemporal map of the original disease rates exhibits large jumps in values, even in adjacent LAAs or consecutive years (Additional file 1: Appendices 6 and 7). The traditional spatiotemporal kriging without adjusting for the nugget effect (assuming no measurement errors) can produce a smooth curved surface of the cross-sectional disease rates at any time point. However, the kriging estimate for a spatiotemporal centroid point (the centroid of an LAA in the middle of a year) is the original rate of the LAA in that year without a stabilizing effect (Additional file 1 :Appendices 8 and 9). At the other extreme, the nugget-adjusted spatiotemporal kriging (with homoscedasticity assumption) stabilizes the rates indiscriminately; it oversmooths the rates in populous regions but undersmooths the rates in sparsely populated ones (Additional file 1: Appendices 10 and 11). The stabilized spatiotemporal kriging adjusts the smoothing of a data point according to its variance, resulting in a desirable effect (Figs. 1 and 3). In this study, the stabilized spatiotemporal kriging is anchored to the centroids. The method could be extended to an area-to-area or area-to-point kriging [26, 27] in future research to consider the shapes, sizes, and population distributions of LAAs. In addition, the proposed method assumed a constant effect (μ) across space and time, which was only suitable for a simple spatiotemporal process scenario. Kriging methods with the spatiotemporal trends fitted by straight lines, polynomials, or splines are also worthy of study.

The SIR map for oral cancer in men (Fig. 1) created by our proposed method reveals that the high-risk areas of oral cancer originated in the midwestern and southeastern regions and spread to the middle and the north, with persistent cold spots in the northern and southwestern urban regions. The spatiotemporal variations of oral cancer risk may be related to the Taiwanese habit of chewing betel nuts and concentrations of heavy metals in the soils [22, 28]. However, the SIR map for breast cancer in women (Fig. 3) reveals that the high-risk areas for breast cancer are concentrated in densely populated urban regions in the west, where

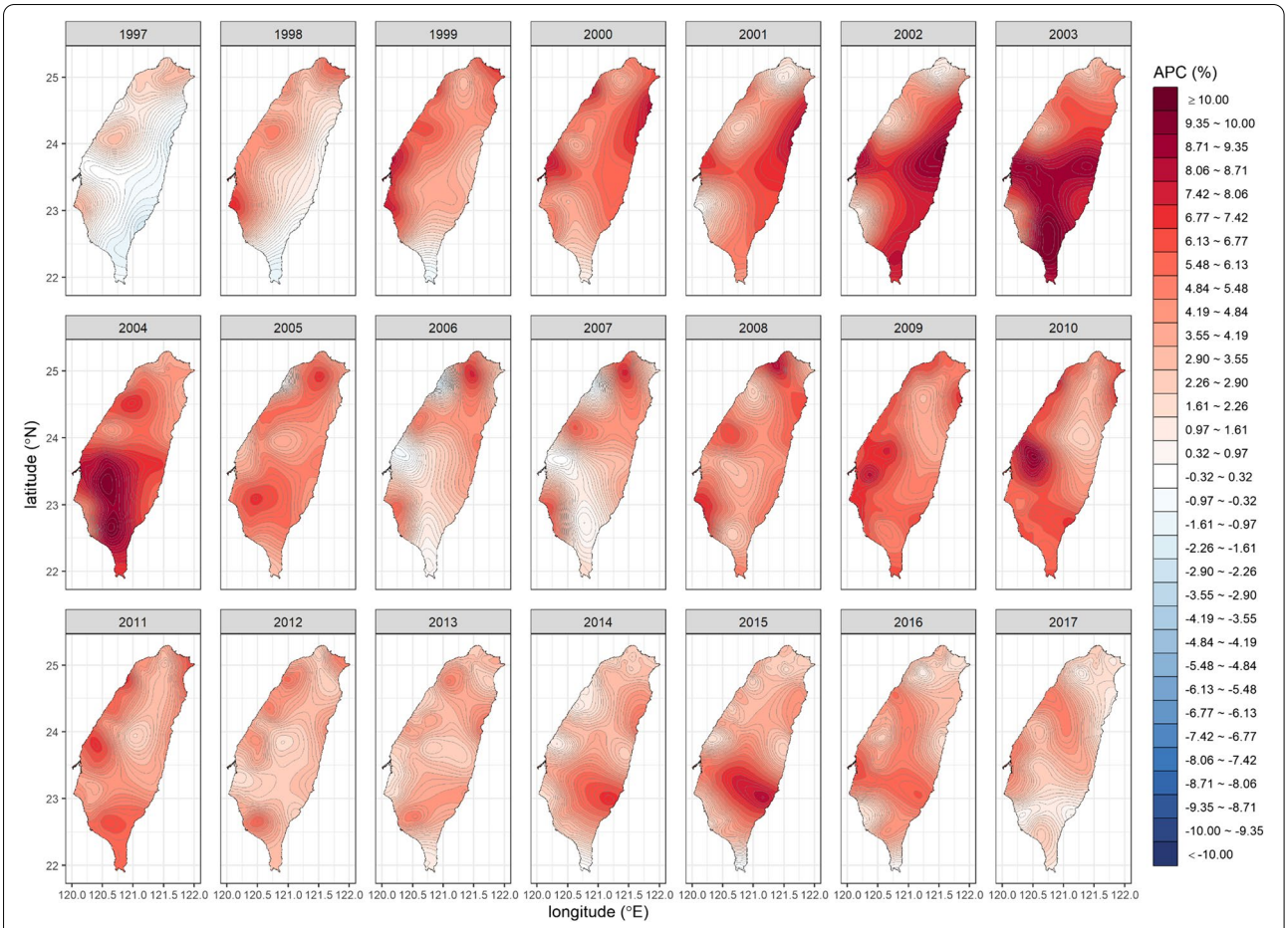

Fig. 4 The stabilized spatiotemporal kriging map for the annual percent changes (APCs) in the standardized incidence ratios of breast cancer in women in Taiwan, 1997 to 2017

people primarily have higher socioeconomic status and have adopted westernized lifestyles and eating habits [29]. Further studies can extend the methodology to the setting of the universal kriging (regression kriging) [9] to incorporate covariate information, if available, into spatiotemporal maps.

The stabilized spatiotemporal kriging has greater uncertainty in areas with smaller populations. To demonstrate the effect, we used the SIR hot spot for breast cancer in women in the mountainous, sparsely populated southeast region as an example (Fig. 3). We produced a SIR time-series plot for the hot spot centroid (Yanping Township in Taitung County). We did the same for a hot spot centroid in the densely populated Taipei Metropolis (Zhongzheng District in Taipei City). In Additional file 1: Appendix 12, the comparison reveals that the SIR trends are similar in the two LAAs, but the 95% confidence interval for the Yanping Township is much wider than that in the Zhongzheng District. Including 1.00 in the 95% SIR confidence interval for the Yanping Township raises whether it is genuinely a SIR hot spot.

Spatiotemporal maps facilitate the understanding of disease risk dynamics. We recommend using the stabilized spatiotemporal kriging method for mapping disease rates across space and time.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-022-01749-9.

Additional file 1: Appendix 1. Derivation of the variance of measurement error. **Appendix 2.** Taiwan island used in the simulation. **Appendix 3.** Simulation scenarios. **Appendix 4.** The spatiotemporal dynamic maps for the standardized incidence ratios (SIRs) and the annual percent changes (APCs) in the SIRs of oral cancer in men in Taiwan, 1997 to 2017. **Appendix 5.** The spatiotemporal dynamic maps for the standardized incidence ratios (SIRs) and the annual percent changes (APCs) in the SIRs of breast cancer in women in Taiwan, 1997 to 2017. **Appendix 6.** The local-administrative-area-based spatiotemporal map of the original disease rates of oral cancer in men in Taiwan, 1997 to 2017. **Appendix 7.** The local-administrative-area-based spatiotemporal map of the original disease rates of breast cancer in women in Taiwan, 1997 to

2017. **Appendix 8.** The traditional nugget-unadjusted spatiotemporal kriging map for the standardized incidence ratios (SIRs) of oral cancer in men in Taiwan, 1997 to 2017. **Appendix 9.** The traditional nugget-unadjusted spatiotemporal kriging map for the standardized incidence ratios (SIRs) of breast cancer in women in Taiwan, 1997 to 2017. **Appendix 10.** The traditional nugget-adjusted spatiotemporal kriging map for the standardized incidence ratios (SIRs) of oral cancer in men in Taiwan, 1997 to 2017. **Appendix 11.** The traditional nugget-adjusted spatiotemporal kriging map for the standardized incidence ratios (SIRs) of breast cancer in women in Taiwan, 1997 to 2017. **Appendix 12.** The time-series plot of the standardized incidence ratios (SIRs) of breast cancer in women in Taiwan, 1997 to 2017, for two hot spot centroids (Left: Zhongzheng District in Taipei City; Right: Yanping Township in Taitung County).

Acknowledgements
The content of this research may not represent the opinion of the Health Promotion Administration, Ministry of Health and Welfare in Taiwan. This study was based on the master thesis of Dai-Rong Tsai (the first author of this paper) completed at National Taiwan University under the supervision of Wen-Chung Lee (the corresponding author of this paper).

Authors' contributions
Dai-Rong Tsai and Wen-Chung Lee developed the models, Dai-Rong Tsai, Jing-Rong Jhuang, and Shih-Yung Su analyzed the data, Jing-Rong Jhuang, Chun-Ju Chiang, Ya-Wen Yang, and Wen-Chung Lee collected the data, Dai-Rong Tsai, Jing-Rong Jhuang, and Wen-Chung Lee drafted the manuscript, and all authors contributed substantially to its revision. Wen-Chung Lee has full access to all aspects of the research and writing process and takes final responsibility for the whole paper. The author(s) read and approved the final manuscript.

Funding
This paper was supported by grants from the Health Promotion Administration, Ministry of Health and Welfare (A1101009: Tobacco Health and Welfare Taxation) in Taiwan, and the Ministry of Science and Technology in Taiwan (MOST 111–2314-B-002–089-MY3). No additional external funding was received for this study. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials
Data are available from the National Taiwan Cancer Registry Database published by the Health Promotion Administration, Ministry of Health and Welfare of Taiwan. Due to legal restrictions imposed by the government of Taiwan in relation to the "Personal Information Protection Act", data cannot be made publicly available. Requests for data can be sent as a formal proposal to the Health Promotion Administration, Ministry of Health and Welfare of Taiwan.

Declarations

Ethics approval and consent to participate
This study protocol was approved by the National Taiwan University Research Ethics Committee (202101HM030) and the Data Release Review Board of the Health Promotion Administration, Ministry of Health and Welfare in Taiwan. All methods were performed in accordance with the relevant guidelines and regulations. In addition, the National Taiwan University Research Ethics Committee waived the requirement for informed consent due to the lack of personal information and secondary data in the study.

Consent for publication
Not applicable.

Competing interests
The authors declare no potential conflict of interest.

Author details
[1]Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Rm. 536, No. 17, Xuzhou Rd., Taipei 100, Taiwan. [2]Taiwan Cancer Registry, Taipei, Taiwan.

Received: 18 May 2022 Accepted: 6 October 2022
Published online: 13 October 2022

References

1. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics. 1987;43(3):671–81.
2. Gelfand AE, Vounatsou P. Proper multivariate conditional autoregressive models for spatial data analysis. Biostatistics. 2003;4(1):11–25.
3. Wakefield J. Disease mapping and spatial regression with count data. Biostatistics. 2007;8(2):158–83.
4. Lawson AB. Bayesian disease mapping: hierarchical modeling in spatial epidemiology. 3rd ed. Boca Raton: CRC Press; 2018.
5. Moore DA, Carpenter TE. Spatial Analytical methods and geographic information systems: use in health research and epidemiology. Epidemiol Rev. 1999;21(2):143–61.
6. Bithell JF. An application of density estimation to geographical epidemiology. Stat Med. 1990;9(6):691–701.
7. Nakaya T, Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted poisson regression for disease association mapping. Stat Med. 2005;24(17):2695–717.
8. Mateu J, Müller WG. Spatio-temporal design: advances in efficient data acquisition. UK: Wiley; 2013.
9. Heuvelink GBM, Griffith DA. Spacetime geostatistics for geography: a case study of radiation monitoring across parts of Germany. Geogr Anal. 2010;42(2):161–79.
10. Kilibarda M, Hengl T, Heuvelink GBM, Gräler B, Pebesma E, Perčec Tadić M, Bajat B. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. J Geophys Res Atmos. 2014;119(5):2294–313.
11. Carrat F, Valleron AJ. Epidemiologic mapping using the "kriging" method: application to an influenza-like illness epidemic in France. Am J Epidemiol. 1992;135(11):1293–300.
12. Forna A, Dorigatti I, Nouvellet P, Donnelly CA. Spatiotemporal variability in case fatality ratios for the 2013–2016 Ebola epidemic in West Africa. Int J Infect Dis. 2020;93:48–55.
13. Souty C, Guerrisi C, Masse S, et al. Impact of the lockdown on the burden of COVID-19 in outpatient care in France, spring 2020. Infect Dis (Lond). 2021;53(5):376–81.
14. Liang CP, Chen JS, Chien YC, Chen CF. Spatial analysis of the risk to human health from exposure to arsenic contaminated groundwater: a kriging approach. Sci Total Environ. 2018;627:1048–57.
15. Wu CD, Zeng YT, Lung SC. A hybrid kriging/land-use regression model to assess PM2.5 spatial-temporal variability. Sci Total Environ. 2018;645:1456–64.
16. Kou K, Guo X, Baade P, et al. Spatial analysis of esophageal cancer mortality in a high-risk population in china: consistent clustering pattern in 1970–74 and 2011–13. Asian Pac J Cancer Prev. 2018;19(11):3161–6.
17. Liao WB, Ju K, Zhou Q, Gao YM, Pan J. Forecasting PM2.5-induced lung cancer mortality and morbidity at county level in China using satellite-derived PM2.5 data from 1998 to 2016: a modeling study. Environ Sci Pollut Res Int. 2020;27(18):22946–55.
18. Huang HC, Martinez F, Mateu J, Montes F. Model comparison and selection for stationary spacetime models. Comput Stat Data Anal. 2007;51(9):4577–96.
19. Ma C. Recent developments on the construction of spatio-temporal covariance models. Stoch Env Res Risk Assess. 2007;22(S1):39–47.
20. Mateu J, Porcu E, Gregori P. Recent advances to model anisotropic space-time data. Stat Methods Appl. 2007;17(2):209–23.
21. Snepvangers JJJC, Heuvelink GBM, Huisman JA. Soil water content interpolation using spatio-temporal kriging with external drift. Geoderma. 2003;112(3):253–71.
22. Hsu CC, Tsai DR, Su SY, Jhuang JR, Chiang CJ, Yang YW, Lee WC. A stabilized kriging method for mapping disease rates. J Epidemiol. https://doi.org/10.2188/jea.JE20210276.
23. Cole TJ. Sympercents: symmetric percentage differences on the 100 loge scale simplify the presentation of log transformed data. Stat Med. 2000;19(22):3109–25.
24. Chiang CJ, Wang YW, Lee WC. Taiwan's Nationwide Cancer registry system of 40 years: past, present, and future. J Formos Med Assoc. 2019;118(5):856–8.

25. Kao CW, Chiang CJ, Lin LJ, Huang CW, Lee WC, Lee MY, Taiwan Society of Cancer Registry Expert Group. Accuracy of long-form data in the Taiwan cancer registry. J Formos Med Assoc. 2021;120(11):2037–41.
26. Goovaerts P. Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging. Int J Health Geogr. 2006;5(1):52.
27. Goovaerts P. Kriging and semivariogram deconvolution in the presence of irregular geographical units. Math Geosci. 2007;40(1):101–28.
28. Lin WC, Lin YP, Wang YC, Chang TK, Chiang LC. Assessing and mapping spatial associations among oral cancer mortality rates, concentrations of heavy metals in soil, and land use types based on multiple scale data. Int J Environ Res Public Health. 2014;11(2):2148–68.
29. Liaw YP, Lee WC, Chen CJ. Trend surface analysis for cancer mapping: application to breast cancer mortality in Taiwan. Taiwan Journal of Public Health. 1998;17(6):474–84.

Publisher's Note