

Abstract

In previous research, childhood cancer incidence rates have been positively linked to county-level crop intensity after controlling for the effects of other socio-demographic factors, e.g., population density or median household income. Although the findings are intriguing, the data analysis has important methodological limitations, including: a) the cancer incidence rates were aggregated by relatively large geographic units, e.g., within a county or census tract, and b) the analysis was conducted at one level, e.g., either at county or census tract level. Since these studies are ecologic in nature, ecologic bias cannot be ruled out, no definitive causal pathways can be established and risk estimates at the aggregate level may not reflect risk estimates at the individual level. We propose to adopt an integrated approach to overcome these pitfalls. The method is based on a *spatial point process* modeling approach and importantly, this spatial point process modeling method has the potential to generate causal hypothesis on an individual-level.

We proposed a hierarchical spatial point process model which adjust the estimates for misaligned covariates and incorporate the uncertainty that is inherent to different data sources. By using the Ohio childhood cancer incidences, we compared these estimates with the estimates obtained from spatial point process model and area level model where data are aggregated at county level. In addition, we compared the county level relative risk estimates when the area level model is accounted for the modifiable areal unit problem (MAUP).

Introduction

The etiology of childhood cancer is largely unknown. Research into the cause of increasing childhood cancer incidence has suggested links to air pollution and pesticide use among other factors. Increased pesticide levels have been reported in communities neighboring agricultural land, potentially increasing exposure among pregnant women and young children living close to the farmland. In recent studies that have employed ecological models, childhood cancer incidence rates have been positively linked to county-level crop intensity after controlling for the effects of other socio-demographic factors, e.g., population density or median household income. Since ecological bias cannot be ruled out when applying these models, no definitive causal relation can be established. The risks that are estimated at the area level may not reflect risk estimates at the individual level.

When events of interest are observed at geographical locations (e.g., the residence address where a childhood cancer event was registered), the data can be viewed as spatial point processes, especially when this location depends on our prior understanding of the exposure process and disease etiology. In practice, assuming an inhomogeneous point process is more appropriate, especially when one of the objectives of the study is to find the risk factors that influence the process.

The problem of spatial misalignment is faced when some of the covariates in the intensity models are not observed at the locations where the intensities should be estimated. Moreover, the misaligned data generally comes from various sources. When employed in specifying the main model for the response variable, this variability in data generating mechanism adds another level of uncertainty in estimation, known as *data uncertainty*. To overcome the problem of misalignment and data uncertainty, we propose hierarchical spatial point process models which build up separate models to each type of misaligned data hierarchically, in addition to the main model for the spatial point process.

The modifiable areal unit problem (MAUP) occurs when the aggregation of data at one level changes the correlation structure that is observed at another level with the same dataset. We propose an MAUP adjusted area level model as similar to the research in ecological modeling.

Aim

The study focused on developing spatial point process models for the location of childhood cancer incidence appropriately incorporating the multilevel and cross-classified structure of the covariates, and employ the model to identify potential risk factors for childhood cancer. Model development employs a hierarchical spatial point process approach, adjusts for misaligned data, and incorporates the uncertainty that is inherent to different data sources. We apply these models to the childhood cancer cases reported to the Ohio Cancer Incidence Surveillance System (OCISS) to estimate intensity (i.e., incidence rate) maps, and to find significant risk factors.

In addition, we focused on developing area level models accounting for the MAUP to minimize the effect of ecological fallacy, and the results are compared with a model without accounting for MAUP.

Methods

Spatial point process models: Assume that the data is presented by a set of geographic location $s_i \in B, i = 1, \dots, n$, and B is the study region. We consider the models for the locations as inhomogeneous Poisson processes, with the intensities at the data locations as

$$s = (s_1, \dots, s_n)^T \sim PP(\lambda(s|\theta))$$

$$\log \lambda(s|\theta) = \log \lambda_0(s) + Z(s)^T \beta + \nu(s) + u(s)$$

where $\lambda_0(s)$ represent the background information, e.g., the

population density

$Z(s)$ is the set of covariates observed at individual level (e.g., mother's ethnicity, child age, genetic markers); at census tract level (e.g., median household income or population density); at county level (e.g. percentage of land used in agriculture); at EPA air pollutants monitoring level

$\nu(s)$ is spatially correlated random effect

$u(s)$ is spatially uncorrelated random effect.

Hierarchical spatial point process models: Commonly, the estimation of background information and the covariates at the data points are done in advance (e.g., by kernel method or kriging predictions), and the estimation procedure follows the profile likelihood approach. The estimates of intensity surface is sensitive to the plug-in estimates of background information and covariates. In hierarchical spatial point process modeling, the spatial point process model is combined with the background information and the covariates to obtain the joint uncertainties:

$$[s, \lambda_0(s), Z(s) | P_s, P_{\lambda_0}, P_Z] = [s | \lambda_0(s), Z(s) P_s, P_{\lambda_0}, P_Z] [\lambda_0(s) | P_{\lambda_0}] \times [Z(s) | P_Z]$$

where P_D is the set of parameters that are used to define the model for D .

Area level models: When the incidences of childhood cancer are aggregated at county level, the data represent a count variable and a hierarchical generalized spatial model will be used at county level to find the significant risk factors. The adjustment for MAUP can be made by considering the county-specific spatial distributions of background information and covariates in the linear predictor model.

Methods, cont'd

Adjacency matrix: Spatially adjacent points will be determined by the Dirichlet tessellation method which divide an area into contiguous non-overlapping tiles, one per data point, with no gaps in between them. A map of these tiles yields an adjacency matrix of neighbors. To determine the adjacency matrix for the area level data, we will use the first-order neighborhood structure if the counties share common administrative boundaries.

Data Sources

Incidents of cancer cases with Ohio residential address for the years 1996-2009 and childhood age group (0-24 years) are obtained from the Ohio Cancer Incidence Surveillance System (OCISS). The residential addresses are geocoded into longitude and latitude coordinates. County specific percentage of land used in agriculture by crop types was derived from the 1997 Agriculture Census conducted by the US Department of Agriculture (USDA). Population characteristics at census tract and county levels for the year 2000 are acquired from the US Census Bureau.

Results

The current results are limited to all incidents of Ohio cancer cases in year 2000. Two software, R and WinBUGS, are used for analysis. The convergence of all model parameters are ensured by following the standard approach.

Figure 1: Percentage of Ohio land used in agriculture in 1997: county specific data from USDA (left); and smooth map from hierarchical spatial point process models (right).

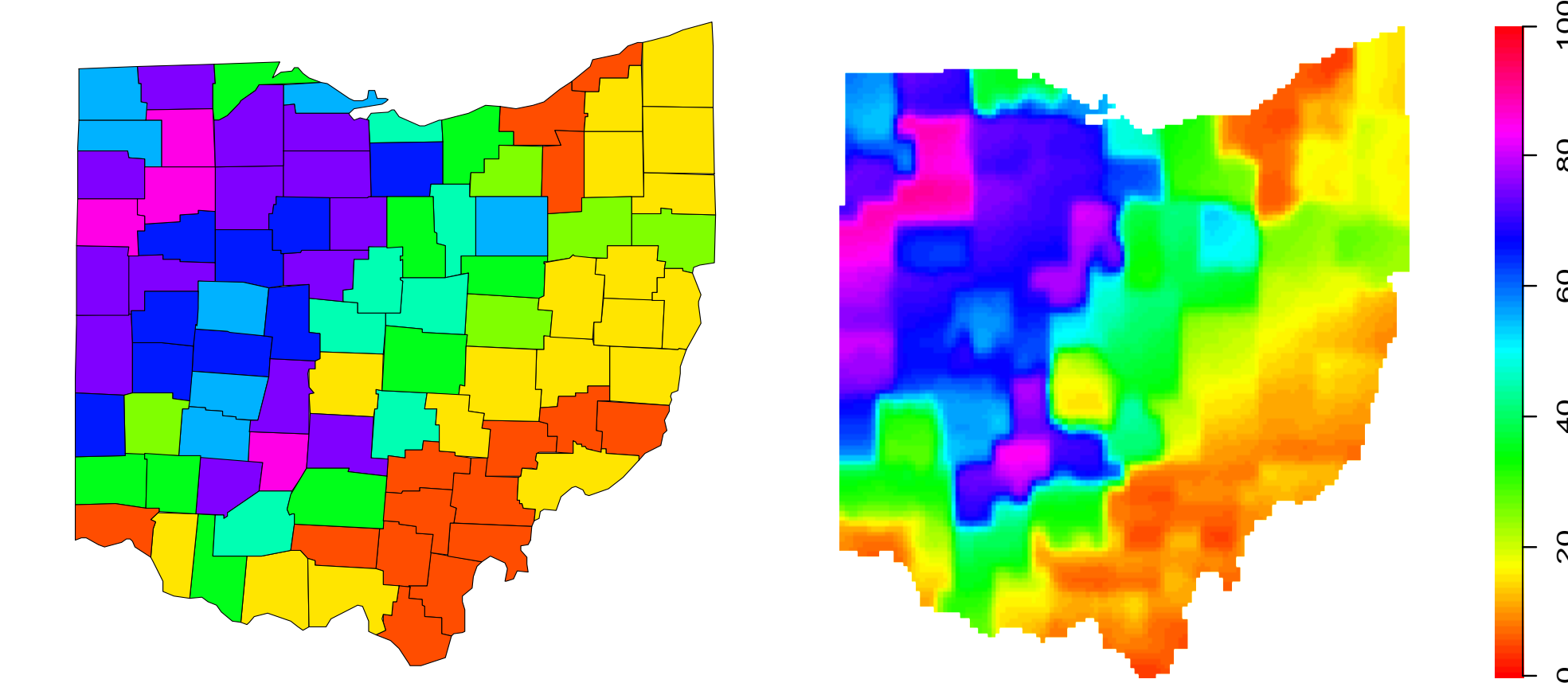
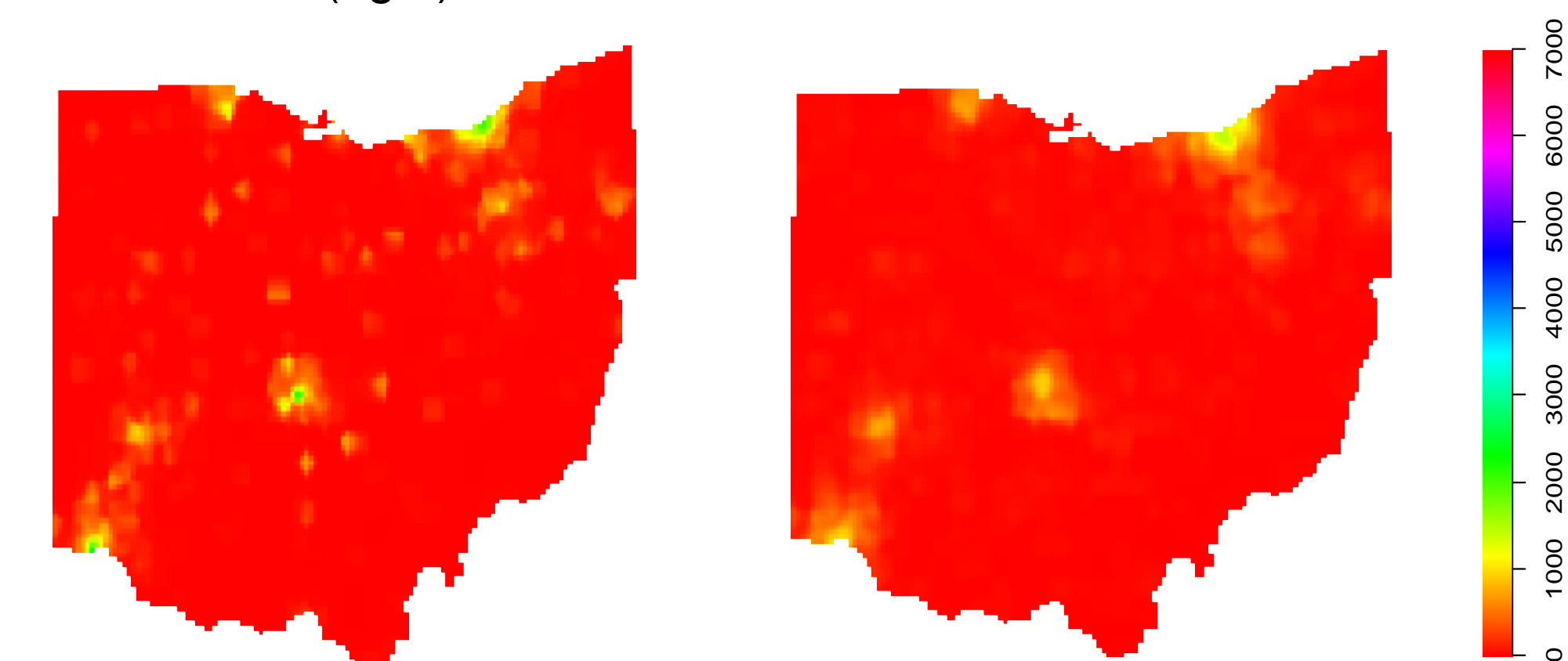


Figure 2: Intensity estimates of Ohio childhood cancer incidences in year 2000: spatial point process models (left); and hierarchical spatial point process models (right).

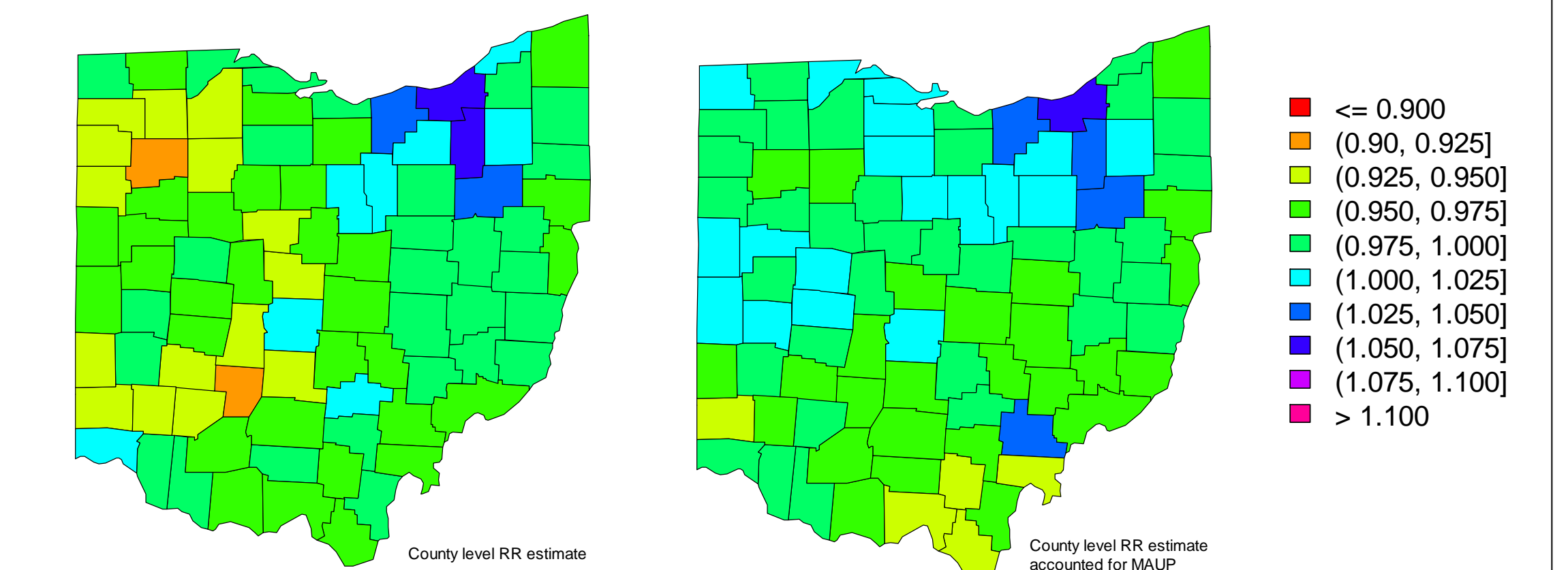


Results, cont'd

Table 1: Parameter estimates (95% CI) from from the spatial models fitted to Ohio childhood cancer incidences in year 2000.

	Spatial point process models	Hierarchical spatial point process models	County level models
Intercept	5.558 (5.445, 5.671)	4.046 (3.660, 4.360)	3.802 (3.600, 3.984)
Medium exposure	-0.076 (-0.256, 0.101)	-0.101 (-0.611, 0.482)	0.065 (-0.188, 0.328)
High exposure	-0.051 (-0.310, 0.197)	-1.249 (-2.037, -0.371)	0.095 (-0.235, 0.419)
Sigma.v	0.180 (0.066, 0.369)	2.066 (1.803, 2.384)	0.209 (0.027, 0.539)
Sigma.u	0.138 (0.062, 0.251)	0.106 (0.034, 0.188)	0.188 (0.026, 0.348)
-2Log-likelihood	-5613.0	-5446.0	---

Figure 3: County specific relative risk estimates of Ohio childhood cancer incidences in year 2000: area level model (left); and area level model accounted for MAUP (right).



Conclusions

We proposed a hierarchical spatial point process model. In modeling the intensity of incidents of childhood cancer cases of Ohio in year 2000, we have used the population density as a background information and the percentage of land used in agriculture as an exposure variable.

Viewing figures 1 and 2, clearly hierarchical spatial point process models generates much smoother estimates over the maps, and we also observed better goodness-of-fit estimate (-2 log-likelihood value in Table 1) in comparing to spatial point process model.

Results (Table 1) indicate that the estimated effects of medium level of crop intensity from all these models, and effect of high level of crop intensity from spatial point process model and area level model did not have any significant effect on cancer incidences comparing to the low level of crop intensity. The significant effect of high level crop intensity from hierarchical spatial point process model needs to be validated by designing a number of simulation studies for various spatial scenarios.

Adjustment for the spatial distribution of exposure (i.e., MAUP) (Figure 2, right panel) also contribute to the reduction in incidence ratio variations, comparing to without MAUP adjustment (Figure 2, left panel).

The reported results are limited to one year incidents of cancer cases.