An Additive Mixed-effects Model (AMM) with Kernel Smoothers and a Permutation Test on Temporal Heterogeneity of Geospatial Risk Patterns

Yannan Tang
Advisor: Professor Daniel Gillen

Department of Statistics
University of California, Irvine

January 7, 2020

Motivating example: MA birth defects study

- Girguis et al. (2016) conducted a study of birth defects in MA.
- A retrospective case-control study was conducted using all live and still births from the Massachusetts state birth registry between 2001-2009.
- Cases: all birth defects. Controls: random selected 1000 infants each year among all live births without defects.
- Among the defects, Patent Ductus Arteriosus (PDA) is one of the most common.

PDA

Before birth, the two major arteries—the aorta and the pulmonary artery—are connected by a blood vessel called the ductus arteriosus. This vessel is supposed to close after birth.
In some babies the ductus arteriosus remains open (patent). In PDA, abnormal blood flow occurs and can put strain on the heart and increase blood pressure in the lung arteries.

Spatial effects

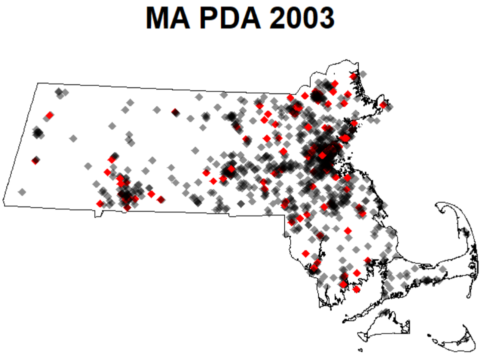

Figure 1: Red: PDA cases; Black: non-PDA cases.

- Spatial disparities in risk - a risk surface
- Surrogate effects

Notations

- $s()$ stands for smoothing function in general.
- $lo()$ stands for LOESS smoother.
- i is observation index, $i = 1, 2, \ldots, N$.
- j is time index, $j = 1, 2, \ldots, J$.
- $X\beta$ is the linear term.
- $Z\gamma$ is the random effect.
- (u, v) stands for longitude and latitude, respectively.

Model spatial effects in generalized additive model

- Spatial effects

Model spatial effects in generalized additive model

- Spatial effects
- $lo(u, v)$ (LOESS smoother)

Model spatial effects in generalized additive model

- Spatial effects
- $lo(u, v)$ (LOESS smoother)
- Confounding variables (maternal age, adequacy of prenatal care, etc)

Model spatial effects in generalized additive model

- Spatial effects
- $lo(u, v)$ (LOESS smoother)
- Confounding variables (maternal age, adequacy of prenatal care, etc)
- $X\beta$

Model spatial effects in generalized additive model

- Spatial effects
- $lo(u, v)$ (LOESS smoother)
- Confounding variables (maternal age, adequacy of prenatal care, etc)
- $X\beta$
- $g(\mu) = X\beta + lo(u, v)$ (generalized additive model, Hastie and Tibshirani (1990))

Model spatial effects in generalized additive model

- Spatial effects
- $lo(u, v)$ (LOESS smoother)
- Confounding variables (maternal age, adequacy of prenatal care, etc)
- $X\beta$
- $g(\mu) = X\beta + lo(u, v)$ (generalized additive model, Hastie and Tibshirani (1990))

- Looks good for a cross-sectional study

Extension to longitudinal studies

- Longitudinal studies with repeated measurements
- linear model (LM) $\xrightarrow{\text{mixed effects}}$ linear mixed model (LMM)
- additive model (AM) $\xrightarrow{\text{mixed effects}}$ additive mixed model (AMM)
- Lin and Zhang (1999) proposed generalized additive mixed models using smoothing splines
- No current work has been done using kernel smoothers (such as LOESS)

Smoothers comparison: kernel vs splines

	Kernel	Spline
Parametrizable	No	Yes
Efficient calculation	No	Yes
Understandable	Yes	Depends
Use of distance	Yes	No
Fit to irregular map	Yes	Depends

Table 1: A brief comparison between kernel smoother and spline smoother

- Researchers have varied preference in smoother selection
- It is meaningful to provide sufficient statistical tools using kernel smoothers

Fit AM: backfitting algorithm

Model

$y = X\beta + lo(u, v) + \epsilon$

Fitting

1. Fit $y - \hat{lo}(u, v) = X\beta + \epsilon$
2. Fit $y - X\hat{\beta} = lo(u, v) + \epsilon$
3. Repeat Step 1 and 2 until convergence

Fit LMM: Maximum likelihood

Model

$y = X\beta + Z\gamma + \epsilon$

Fitting

1. $\hat{\beta} = \hat{\beta}(\gamma, \sigma^2)$
2. $\ell(\beta, \gamma, \sigma^2) = \ell'(\gamma, \sigma^2)$
3. Maximum ℓ'

Fit AMM: A combined algorithm

Model

$y = X\beta + lo(u, v) + Z\gamma + \epsilon$

Fitting

1. Fit $y - \hat{lo}(u, v) = X\beta + Z\gamma + \epsilon$ using ML
2. Fit $y - X\hat{\beta} = lo(u, v) + \epsilon$ given $(\hat{\gamma}, \hat{\sigma^2})$
3. Repeat Step 1 and 2 until convergence

Empirical performance of AMM

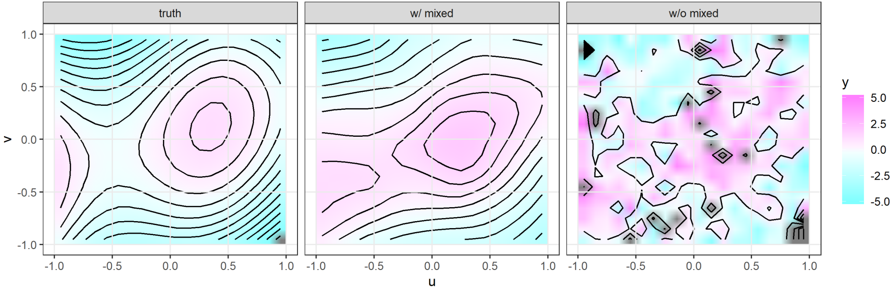

Figure 2: True pattern and estimated patterns with AMM and AM.
$J = 5, \sigma^2 = 1, V_0 = 3, V_1 = 0.09$

PMSD test for time-varying geospatial pattern

One question

- AMM: $y = X\beta + lo(u, v) + Z\gamma + \epsilon$
- The model use one marginal smoother $lo(u, v)$ for multiple time points
- How can we decide if the assumption holds?

Idea

- Fitted model $y = X\hat{\beta} + \hat{lo}(u, v) + Z\hat{\gamma} + \hat{\epsilon}$
- If $lo(u, v)$ is not sufficient, temporal heterogeneity remains in $\hat{\epsilon}$
- Check $y' = \hat{lo} + \hat{\epsilon}$ for heterogeneity

PMSD test for time-varying geospatial pattern

MSD statistic

- Fit $y' = lo_j(u, v) + \epsilon$ (time-specific smoothers)
- Fit $y' = lo_0(u, v) + \epsilon$ (marginal smoother)
- Define MSD $= \frac{1}{JN_g}(\sum_{j=1}^{J}\sum_{g=1}^{N_g}(\hat{lo}_j(u^{(g)}, v^{(g)}) - \hat{lo}_0(u^{(g)}, v^{(g)}))^2)$, where a set of locations $\{(u^{(g)}, v^{(g)}), g = 1, \ldots, N_g\}$ covers the area of interest to be tested

PMSD test

- MSD would be large if one smoother does not suffice
- To get a reference distribution, permute time label and recalculate MSD
- Reject H_0 if $\frac{1}{N_{perm}}\sum_{p=1}^{N_{perm}} I\{\text{PMSD}_p > \text{OMSD}\} < \alpha$

Empirical performance of PMSD test

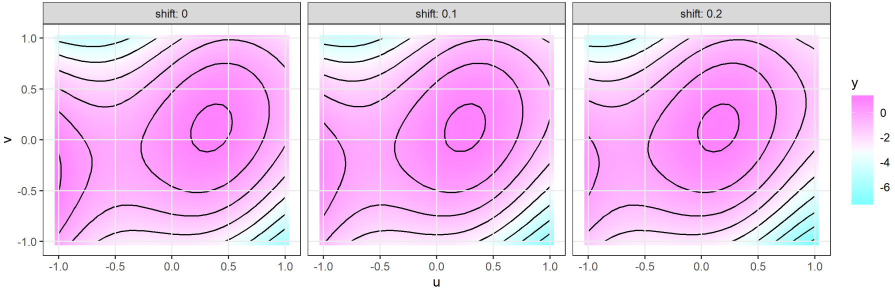

Figure 3: Examples of shifted patterns

Shift	0.00	0.04	0.08	0.12	0.16	0.20
Power	0.06	0.13	0.30	0.53	0.74	0.94

Table 2: Maximum shift of the geospatial pattern and corresponding powers.

To summarize

This presented work

- Extended additive models to additive mixed models with kernel smoothers
- Developed a permutation test for temporal homogeneity of geospatial risk surface

Ongoing works

- Complete the framework to generalized additive mixed models
- Detection of varying areas

References

Mariam S Girguis, Matthew J Strickland, Xuefei Hu, Yang Liu, Scott M Bartell, and Verónica M Vieira. Maternal exposure to traffic-related air pollution and birth defects in massachusetts. *Environmental research*, 146:1–9, 2016.

Trevor Hastie and Robert Tibshirani. *Generalized additive models*. Wiley Online Library, 1990.

Xihong Lin and Daowen Zhang. Inference in generalized additive mixed models by using smoothing splines. *Journal of the royal statistical society: Series b (statistical methodology)*, 61(2):381–400, 1999.