# Bayesian Spatial Health Surveillance

Allan Clark and Andrew Lawson

University of South Carolina

# Two important problems

Clustering of disease:                    PART 1
- Development of Space-time models
- Modelling vs Testing
- Hidden process models
- Example from Scotland
- Future directions

Relative risk change detection:      PART 2
- What is meant by surveillance
- Statistical aspects of surveillance
- Bayesian models space-time risk estimation
- discussion

# Clustering of disease

What is a cluster ?

- No universally accepted definition

- A working (spatial) definition was given by Knox (1989)

    'A geographically bounded group of occurrences of sufficient size and concentration to be unlikely to have occurred by chance'

- In space-time slightly more complicated since we can have 3 different types of clusters.

    - Temporal cluster: occurs in the whole study region, but for a limited time.
    - Spatial cluster: occurs during the whole study period, but in a small area.
    - Spatio-temporal cluster: should exist in a small area and for a limited time.

# Testing or modelling

Large number of tests for clustering exist
- Diggle-Chetwynd, Knox, Scan, Besag-Newell, etc

All have different properties, however some common concerns are
- adjustment for multiple testing results in loss of power
- how to deal with covariates, stratification is wasteful.

Better to fit a model
- no adjustment needed for multiple testing
- deals with covariates in 'traditional' way
- flexibility

# Hidden Process Model development

- From definition of clustering we expect clustering to be a localized phenomenon and occur around some 'centre'.

- 'Centre' can be any shape – most commonly a point or a line
    - Point may reflect the location of a factory or a waste processing site
    - Line may reflect the location of a highway or a river

- Centres not directly observed, they form a hidden process

- Observed cases form a realisation of a hetrogeneous Poisson process with intensity given by

$$\lambda(x,t) = g(x,t)\exp(\beta^T z)m(c_1, c_2, c_3, x, t)$$

  - $x, t$ = space and time coordinates
  - $g(x,t)$ = background rate due to 'population at risk'
  - $\beta^T z$ = is the linear predictor of covariates
  - $c_1$ = spatial cluster centres
  - $c_2$ = temporal cluster centres
  - $c_3$ = spatio-temporal cluster centres

- Different choices of the function $m$ result in different models. We normally use

$$m(c_1, c_2, c_3, x, t) = 1 + \alpha_1 \sum_{i=1}^{n^s} K(x - c_{1i})$$

$$+ \alpha_2 \sum_{i=1}^{n^t} K(t - c_{2i})$$

$$+ \alpha_3 \sum_{i=1}^{n^{st}} K((x, t) - c_{3i})$$

- The $K$ functions are called cluster distribution functions

- The $K$ functions describes how the cases are distributed around the cluster centre. It makes sense to assume that the density of cases decreases with distance from the centre.

- For points we use a radial symmetric Gaussian distribution

$$K(x - c) = \frac{1}{2\pi\kappa} \exp\left[-\frac{1}{2\kappa} \|x - c\|^2\right]$$

# Model Estimation

## Background:

- $g(x, t)$ is a nuisance function which must be properly accounted for.

- It seems natural to consider a nonparametric estimate as we don not wish to make inferences regarding $g(x, t)$

- If we have a control disease (one with similar age-sex distribution) we can construct an estimate based on kernel smoothing

$$\hat{g}(x, t) = \sum w_1 \left( \frac{x - v_j}{h_1} \right) w_2 \left( \frac{t - \tau_j}{h_2} \right)$$

- $w_1$ and $w_2$ are kernel functions, $h_1$ and $h_2$ are smoothing parameters, $\{v_j, \tau_j\}$ are the space-time coordinates of a realisation of the control disease.

# Prior distribution for cluster centres

- Assume cluster centres are points
- We need to have a prior point process model for the **location** and **number**
- The Strauss prior is suitable for this since it is natural to assume that cluster centres are not very close to each other.

$$f(c) \propto b^k \gamma^{R(c,r)},$$

- where $b$ is a rate parameter, $r$ an inhibition distance, $k$ number of points, $R(c,r)$ is the number of r-close pairs, $\gamma$ is an inhibition parameter.
- These parameters are usually fixed before fitting the model to given weak inhibition.

# Other prior distributions

- Weights

$$p(\alpha_1, \alpha_2, \alpha_3) = \exp(-\alpha_1 - \alpha_2 - \alpha_3)$$

  i.e. each weight is exponentially distributed with mean 1.0
- Regression parameter vector has a uniform prior

# Likelihood

- Recall, $\lambda(x,t)$ is the intensity.

$$L = \frac{\prod_{i=1}^{n} \lambda(x_i, t_i)}{\left( \int_A \int_T \lambda(u,v) du dv \right)^n}$$

# Count Data

- Above model is for individual level (point) data. Often we observe data at an aggregated level (e.g. ZIP code, Postcode sector).

- Possible to adapt a model at the individual level to the aggregated level

- For count data we replace the intensity by the integrated intensity over the region.

$$m_{it} = \int_{A_i} \int_{T_t} \lambda(u,v)\,du\,dv$$

$$\simeq e_{it} \exp(\beta^T z) \int_{A_i} \int_{T_t} 1 + \alpha_1 \sum_{i=1}^{n^s} K(s - c_{1i})$$

$$+ \alpha_2 \sum_{i=1}^{n^t} K(t - c_{2i})$$

$$+ \alpha_3 \sum_{i=1}^{n^{st}} K((s,t) - c_{3i})\,dt\,ds$$

# Birth/Death/Shift Algorithm

- Unknown parameters:
  - Cluster variances: $\kappa_1, \kappa_2, \kappa_3$
  - Cluster centres: $c_1, c_2, c_3$

    (note both number and location unknown)
  - 
    - weights: $\alpha_1, \alpha_2, \alpha_3$
    - covariate parameter vector: $\beta_1, \beta_2, \beta_3$

- Cluster variances, weights, covariate parameter can be estimated using standard MCMC, centres need to be estimated via BDS

- BDS consists of doing one of the following at each iteration
  - Death: remove a centre;
  - Birth: add a centre;
  - Shift: move a centre

# Example: Scottish Birth Abnormalities

- Complete set of all singleton Birth abnormalities in Scotland during the period Jan 1991 to Dec 1995 with postcode sectors DD or PH. 92 postcode sectors and 60 months
- Complete set of all live singleton births during same period.
- Carstairs deprivation index computed from 1991 census.
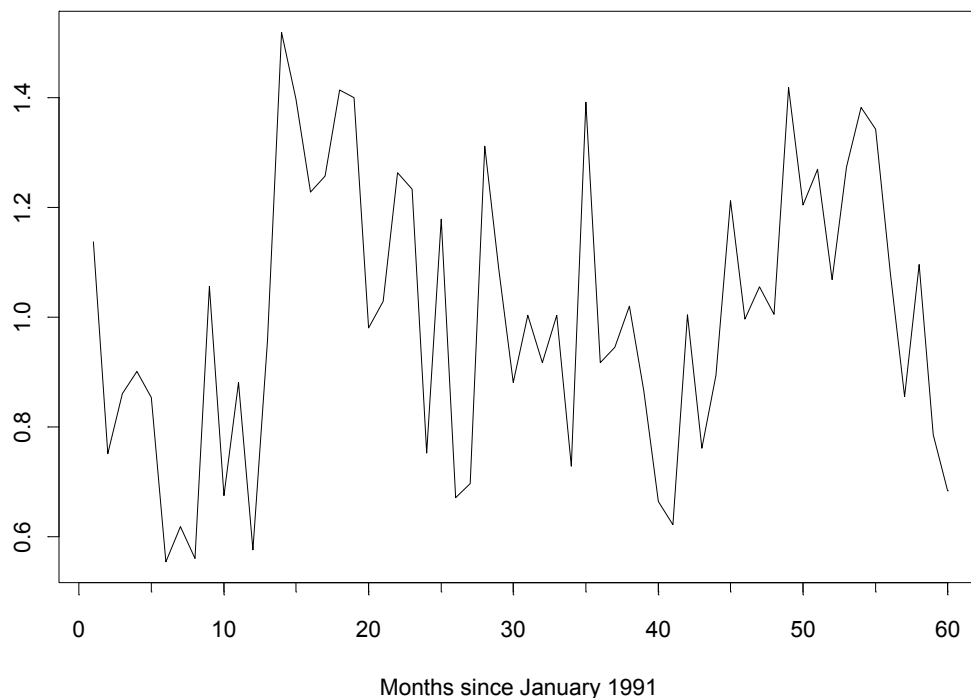- Question: do the birth abnormalities cluster in space-time

- Observed count: number of singleton birth abnormalities in postcode sector i and month t ($m_{it}$)
- Expected count: is
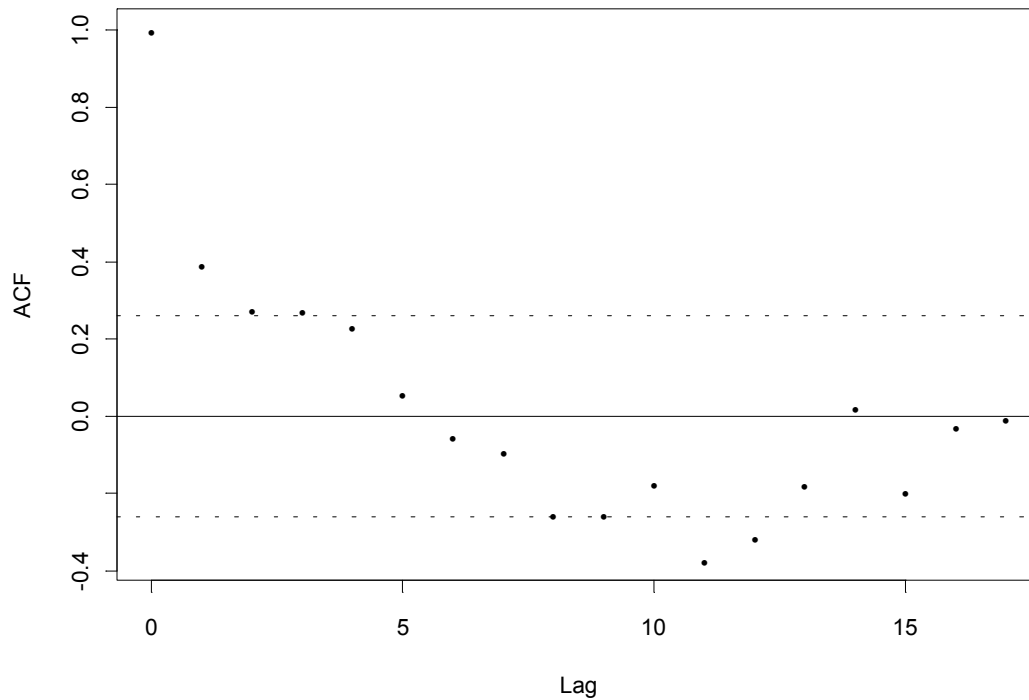
$$e_{it} = n_{it} \cdot \frac{\sum m_{it}}{\sum n_{it}}$$

# Exploratory analysis

- Time series plot of the standardised mortality ratio (obs/exd) is
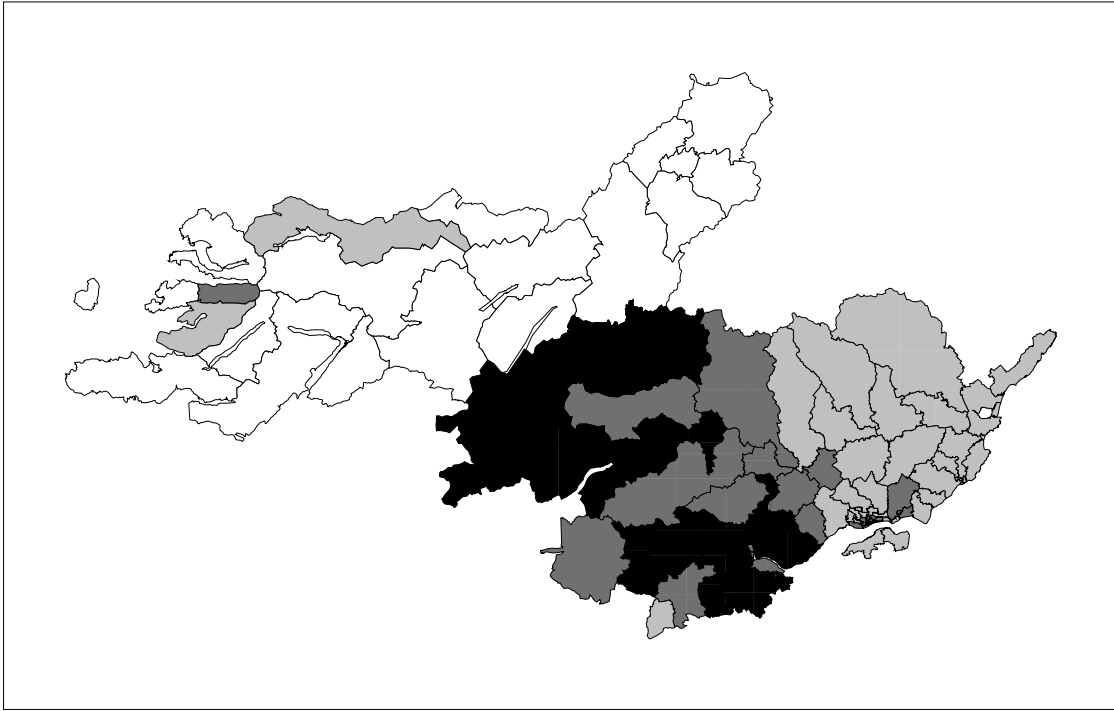


Months since January 1991

- No evidence of any trend in time
- Some evidence of clustering at 20 and 50 months
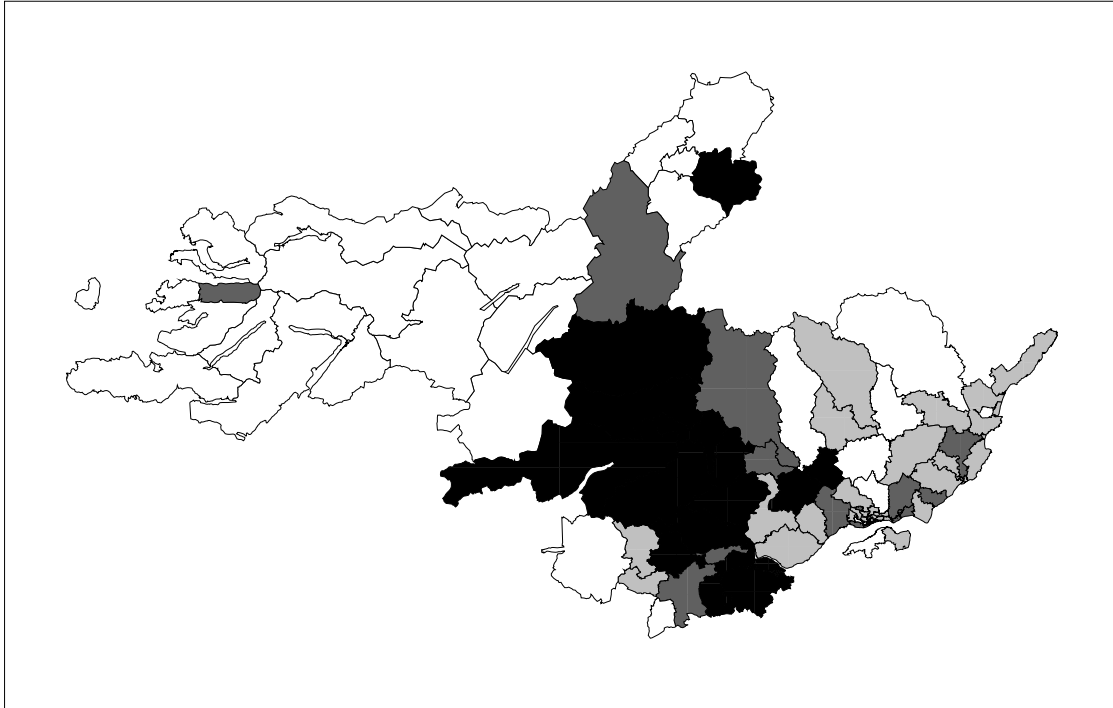
- Autocorrelation plot



- Evidence of short-term correlation suggests clustering.
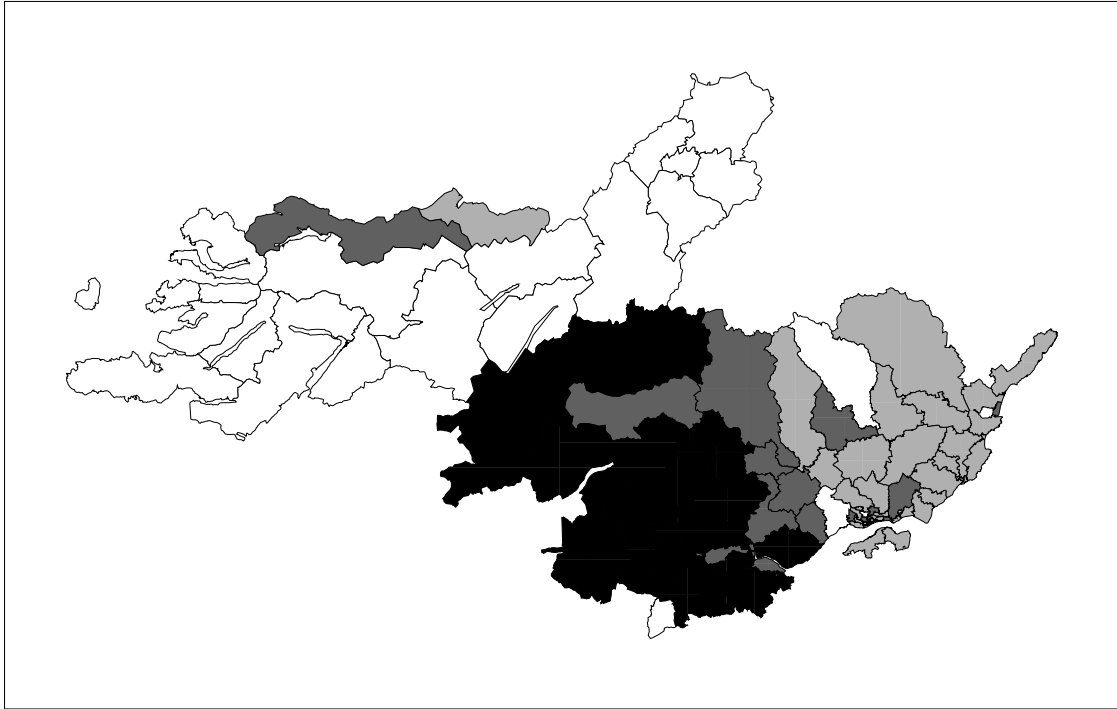
- A plot of the spatial SMR is



- small SMR is rural (north west) area, and high SMR in city areas (Dundee and Perth)
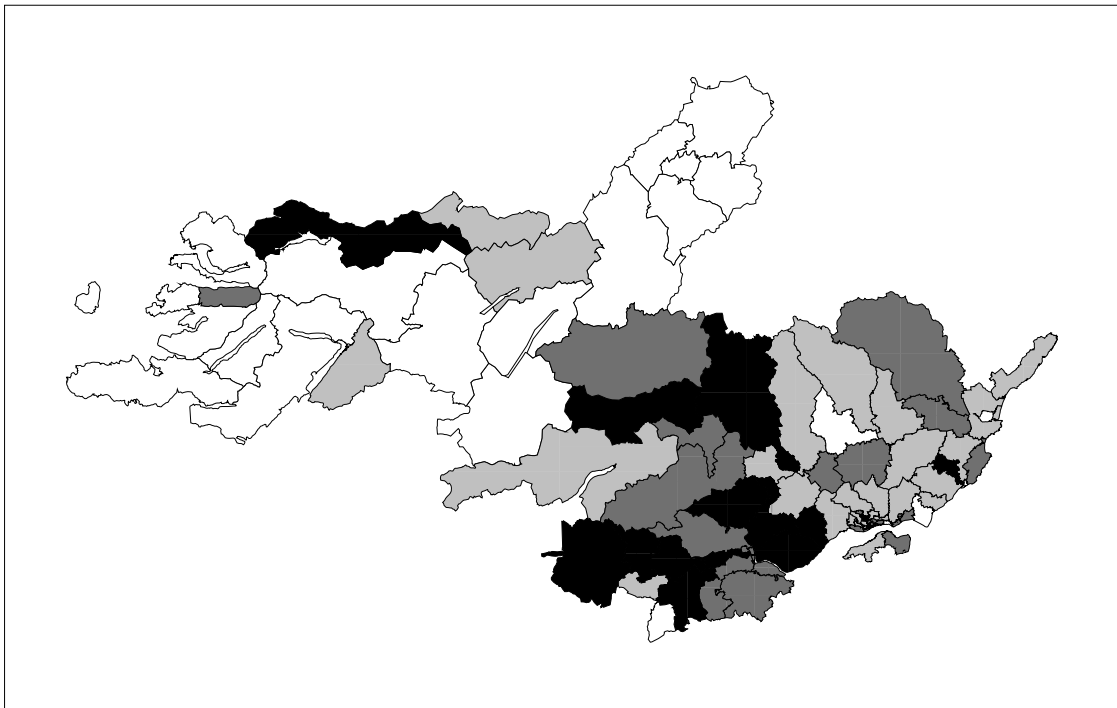- elevated SMR tend to cluster in middle of map

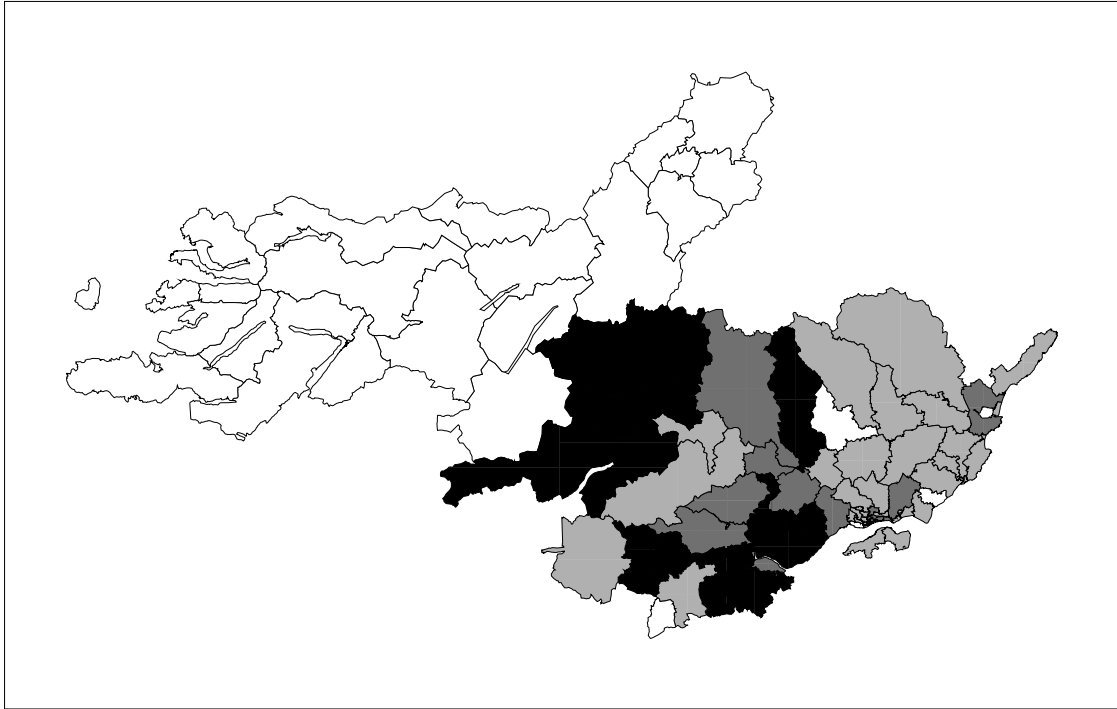- Yearly SMRs can be used to reveal patterns in space-time
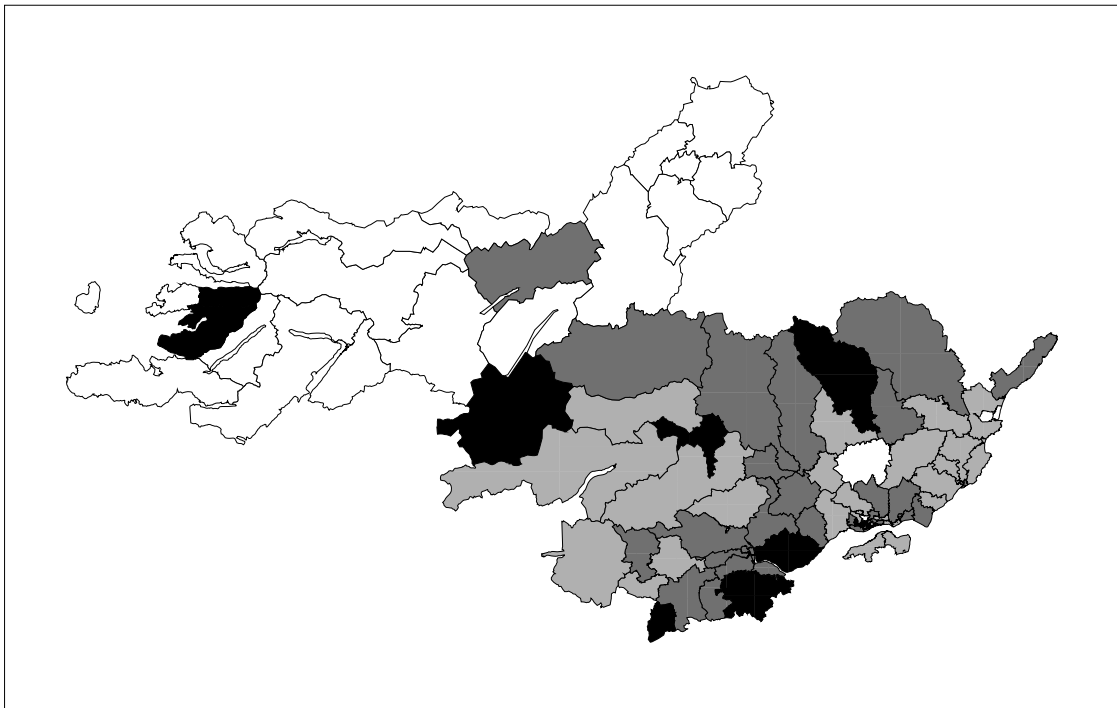


1991

1992



1993

22

1994



1995

23

# Model-based results

- fitted count model for month-postcode sector data
- covariates: x-direction, y-direction, deprivation index
- Strauss parameters set at 0.2 for the radius parameter, 3 for the rate parameters, 0.1 for the inhibition parameter (measured on unit square).
- Ran MCMC algorithm for 20,000 iterations, based inference on next 500 iterations.
- Fitted full model and reduced models by removing one-term only

| model | 2*loglik | ns | nt | nst | BIC |
|---|---|---|---|---|---|
| full | -27740.36 | 6.8 | 4.0 | 9.0 | -55712.7 |
| dep | -27723.70 | 6.7 | 5.0 | 14.0 | -55718.8 |
| x-coord | -27712.79 | 8.6 | 2.1 | 10.0 | -55656.8 |
| y-coord | -27785.42 | 7.8 | 2.1 | 6.4 | -55766.6 |
| spatial | -28007.71 | - | 5.0 | 6.4 | -56163.6 |
| time | -27759.02 | 6.7 | - | 9.0 | -55700.9 |
| s-t | -27781.62 | 6.0 | 2.1 | - | -55684.9 |

- The number of spatial and temporal clusters change only slightly, whereas space-time clusters change dramatically.
- Lack of relationship with deprivation,
- East/west gradient
- Temporal clustering

# Location of Clusters
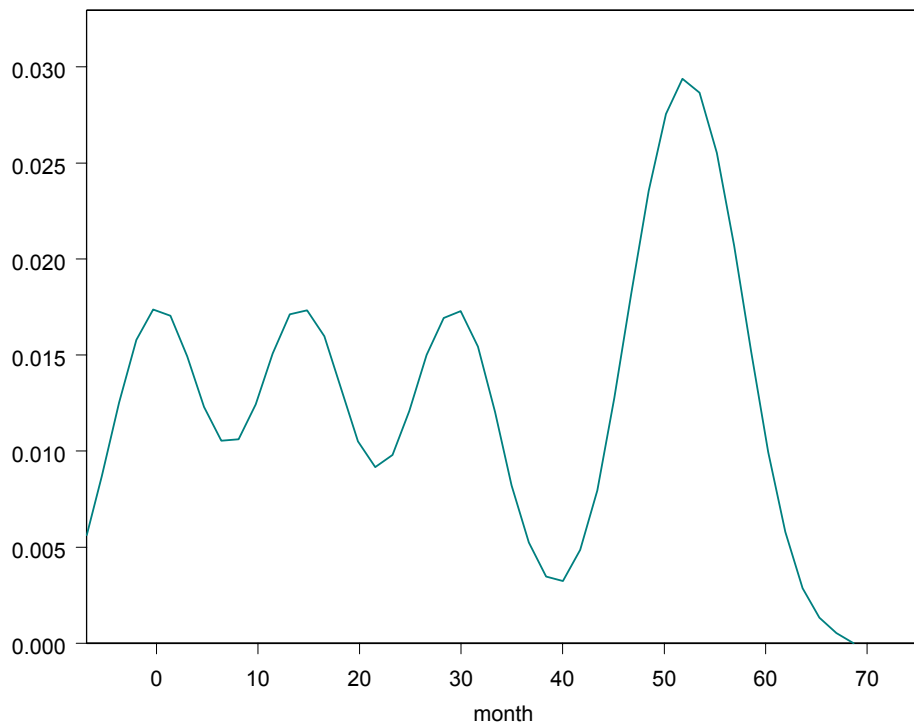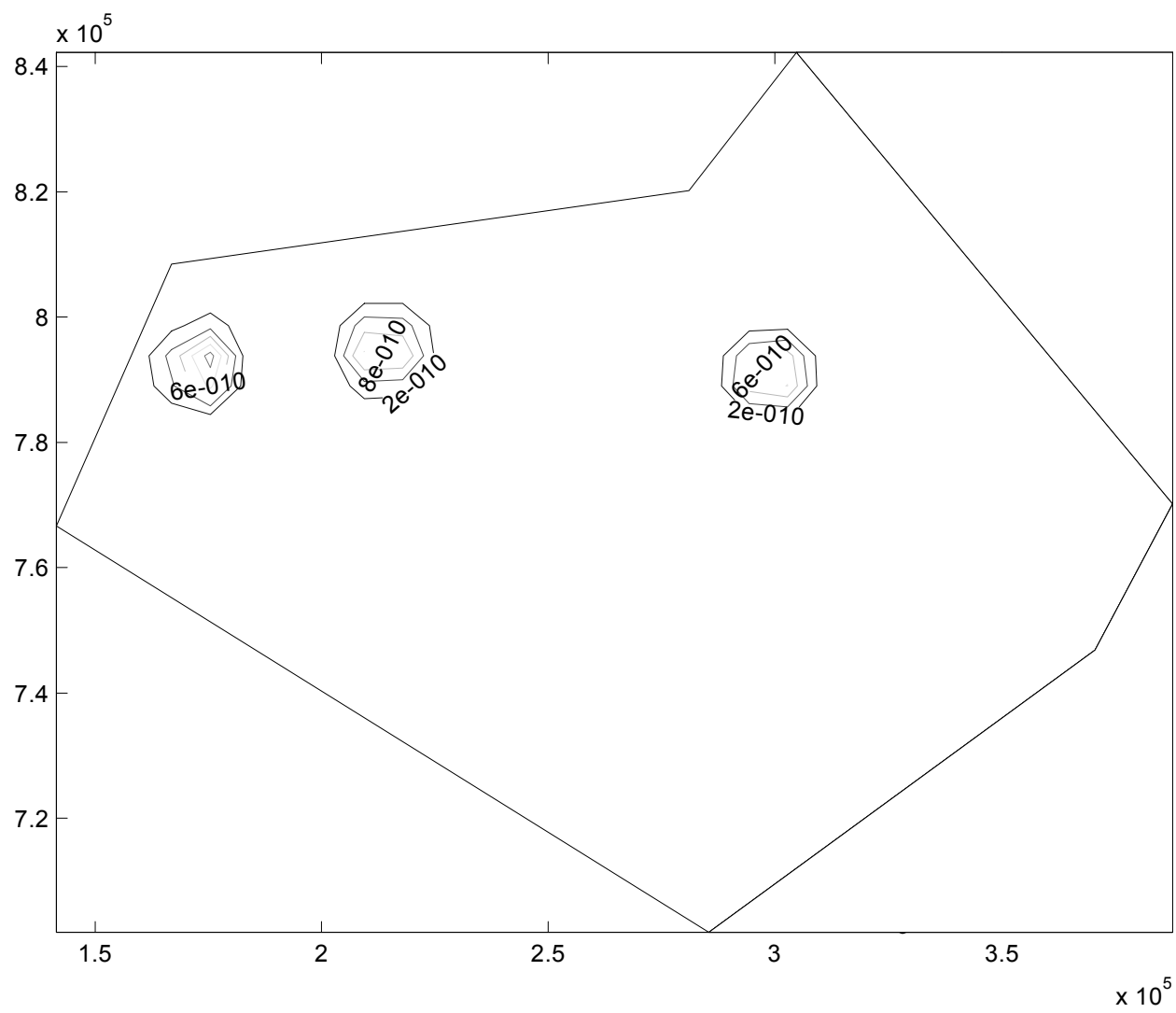
# PART 2

# Relative risk change detection

# Background

The Centers for Disease Control (CDC) defines public health surveillance as: *the ongoing, systematic collection, analysis, and interpretation of health data essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those who need to know. The final link of the surveillance chain is the application of these data to prevention and control. A surveillance system includes a functional capacity for data collection, analysis, and dissemination linked to public health programs.* Thacker and Berkelman (1992)

- ■ Broad definition of surveillance implied

- ■ How can statistical methods can be developed or employed to aid the task of surveillance of populations?

- ■ Clearly *spatial* statistics may be useful but a *temporal* element must also be included in the analysis, where changes in disease distribution are possible

- ■ Spatio-temporal (ST) methods will be important

- ■ However good ST models arent necesarily the best models for surveillance

# Building Disease Surveillance Models

● ■ In a model-based approach to surveillance it is natural to consider monitoring of parameters.

■ In spatio-temporal surveillance, it is clearly important that models be chosen which

1.  are capable of describing the overall behaviour of disease in space and time, and
2. will be sensitive to *changes* in the spatio-temporal structure in time,
3. be reasonably flexible so that *multiple foci* can be included.

■ The first criteria means that the model should be able to capture spatial and temporal effects and spatio-temporal interaction effects.

- The second criteria suggest that the parameterisation should allow for *changes* in time, as well as modelling time.
- This may be achieved by maintaining a relatively simple model, from which important deviations can be detected.
- The third criterion implies that the model should be flexible enough to encompass a variety of behaviours.

# A Simple Count Surveillance Model

- Define a possible simple model for disease count data surveillance

- We want to examine a change within a fixed time period $t$ and spatial unit $j,$ then we could assume the following model for the count of disease in the $j, t$ th unit:

$$y_{jt} \sim Poisson(e_{jt}.\theta_{jt})$$

where $y_{jt}$ is the observed number of cases in the $j$ th region in time period $t,$ $e_{jt}$ is the expected number of cases in the $j$ th region in time period $t$ and $\theta_{jt}$ is the relative risk in the $j$ th region in time period $t$.

- A model could be assumed for the relative risk of the form:

$$\ln \theta_{jt} = \lambda_{jt} = \rho + \varphi_t + \phi_j + \tau_j + \gamma_{jt},$$

- where $\rho$ is an intercept term defining the overall level of the relative risk, $\varphi_t$ is a component describing the temporal variation, $\phi_j, \tau_j$ are components describing the spatial extra-variation

- The component $\phi_j$ is the correlated component and $\tau_j$ the uncorrelated component.

- The final term $\gamma_{jt}$ represents the interaction between spatial and temporal effects in the maps.

- In this form, independent temporal and spatial terms are used and the relation between these is assumed to be included in the $\gamma_{jt}$ term.

- This model is a relatively simple one

compared to the possibly large number of effects which could be conceived. For example, in some published studies, $\phi_j$ and $\tau_j$ are regarded as time-dependent also and so terms such as $\phi_{jt}$ and $\tau_{jt}$ appears. Other studies include simple time trends, and it is also possible to include spatial trends.

● For this model we can specifiy prior distributions for the parameters. These distributions help to define the role of the parameters within the relative risk and make them easier to estimate.

- The prior distributional specifications assumed are as follows. The temporal effect distribution is defined as:

$$\varphi_t | \varphi_{t-1} \sim N(v\varphi_{t-1}, K_1 . \sigma_t^2)$$

  where $v$ is an autoregressive parameter and $\sigma_t^2$ is the variance. This allows there to be a smooth time-on-time variation in risk at any spatial site.

- The spatial components are specified as

$$\phi | \phi_{-j} \sim N(\overline{\phi}_{\delta_j}, K_2 . \frac{\sigma_{ss}^2}{m_j})$$

  where $m_j$ is the number of spatial neighbours of the $j$ th region, $\delta_j$ is the set of neighbours of the $j$ th region and $\phi_{-j}$ is the parameter set excluding $\phi_j$, $\sigma_{ss}^2$ is the correlated spatial component variance and

$$\tau_j \sim N(0, K_3 . \sigma_{us}^2).$$

  where $\sigma_{us}^2$ is the uncorellated spatial component variance.

- These components are not dependent on time and are estimated for the complete data available at any given time. The spatio-temporal effect is defined as

$$\gamma_{jt} \sim N(0, K_4 . \sigma_{st}^2).$$

  where $\sigma_{st}^2$ is the spatio-temporal component variance. This component is estimated for each site at each period.
- The $K_*$ parameters are scaling constants which we will use in the surveillance exercise.

- In surveillance in space-time we want to monitor changes to a process via changes in parameter values. We can monitor a variety of changes by examining changes in $K_1, K_2, K_3,\ K_4$. If the process is in control then $K_1 = K_2 = K_3 = K_4 = 1$. If $K_1 > 1$ then a sharp jump in the risk occurs in time, $K_2 > 1$ is a change in the global spatial correlation structure, $K_3 > 1$ suggests a change in variability across the map, while $K_4 > 1$ is a change in the risk at a particular space-time location. Hence the basic procedure examined here is the examination of changes to global model parameters via the sequential fitting of a global model.

- Variants of this model have been examined by Knorr-Held (2000), who fitted the model variants to *complete* space-time sequences of lung cancer in Ohio.
- He found that a variant of the above model fits the complete 21 year sequence well. The variant has the interaction prior distribution specified as:

$$\gamma_{jt} \sim N(\gamma_{jt-1}, K_4.\sigma_{st}^2)$$

where there is a random walk dependence in the interaction, as opposed to a zero-mean prior distribution.
- That author did not examine the surveillance of the sequence of maps of lung cancer as they arose. Here I examine the difference in these two interaction models applied to the *surveillance* of the Ohio lung cancer example
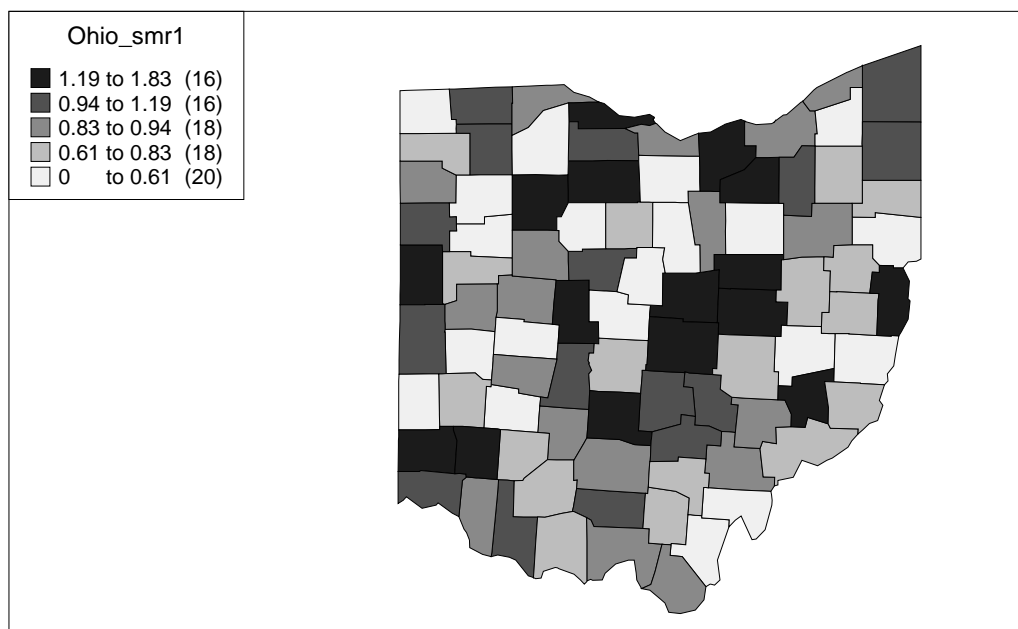
# Ohio Lung Cancer Example

● The Ohio lung cancer data set described by Carlin and Louis (1996) (amongst others) is considered here.


● This data set consists of lung cancer counts for 88 counties of Ohio for 21 year time periods (1968-1988).

● The surveillance of the 21 years of the map sequence is examined here.

● MCMC algorithms can be used to posterior sample these models

● These algorithms are not readily available for use by non-specialists, although the package WinBugs does provide an environment in which spatio-temporal models can be fitted to *complete* sequences of maps.

● It appears that sequential analysis of maps is not possible within WinBugs currently.

● For the models presented here the posterior sampling was custom-programmed in Visual Fortran.
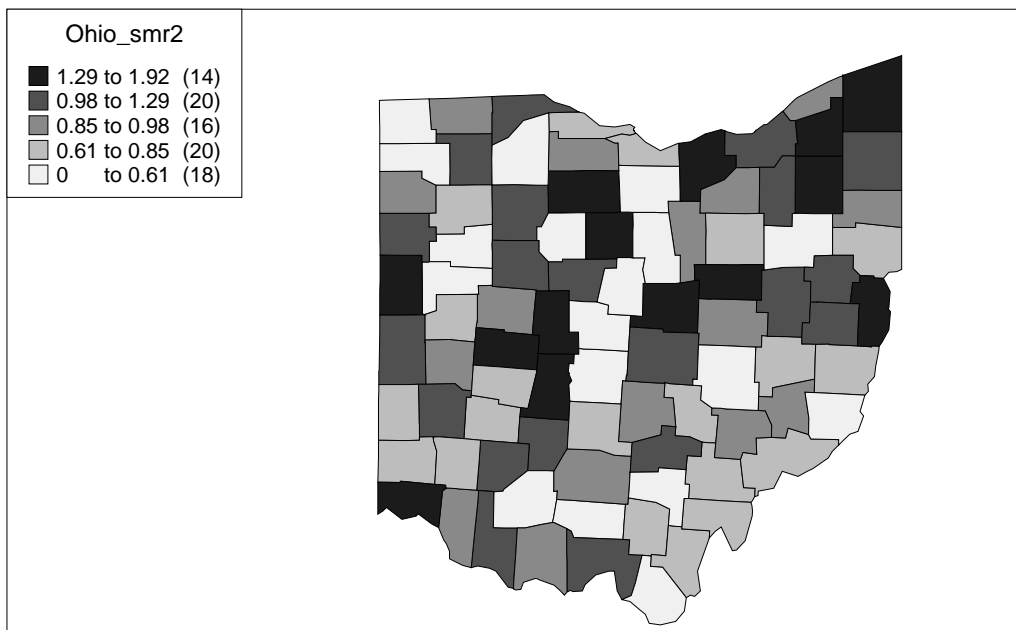
- The models fitted were sampled using Metropolis-Hastings steps. These steps are straightforward to implement, however sequential re-fitting of these models presents a problem as at each new time period a new set of data is added and a new set of parameters are included.
- A simple approach to the problem is to adopt a sliding window of time units within which the effects are estimated.
- This is clearly an approximate procedure as any temporal effects longer than the window will not be properly estimated.
- Here we have used such moving window to reduce the computational time (see also particle filters and SIR methodology e.g. Doucet et al (2001)).
- Each model fit was monitored for convergence using a range of diagnostics including posterior convergence plotting and variance comparisons.
- For simplicity we have estimated the $K_*$ parameters as ratios of current to lag-one variances.
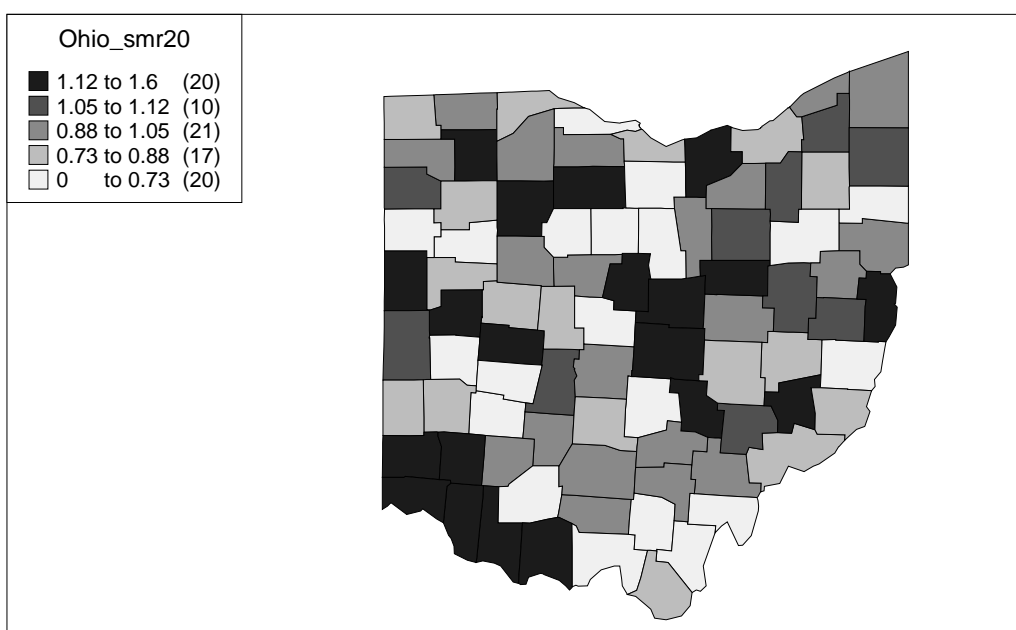
- Here, the multiple time series of $K_1, K_2, K_3,$ $K_4$ and the corresponding variances are examined.
- The SMRs for lung cancer are presented here for time periods: 1, 2, 20, 21, based on a Ohio state yearly rate for each year period.
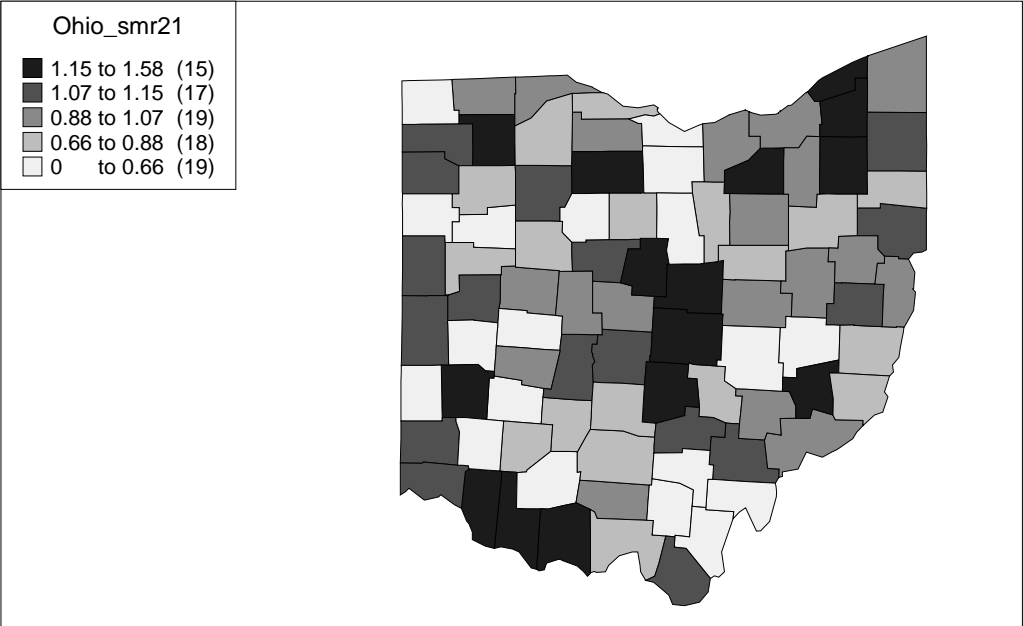- Two basic models were examined: the models with independent space-time interaction and the dependence model.
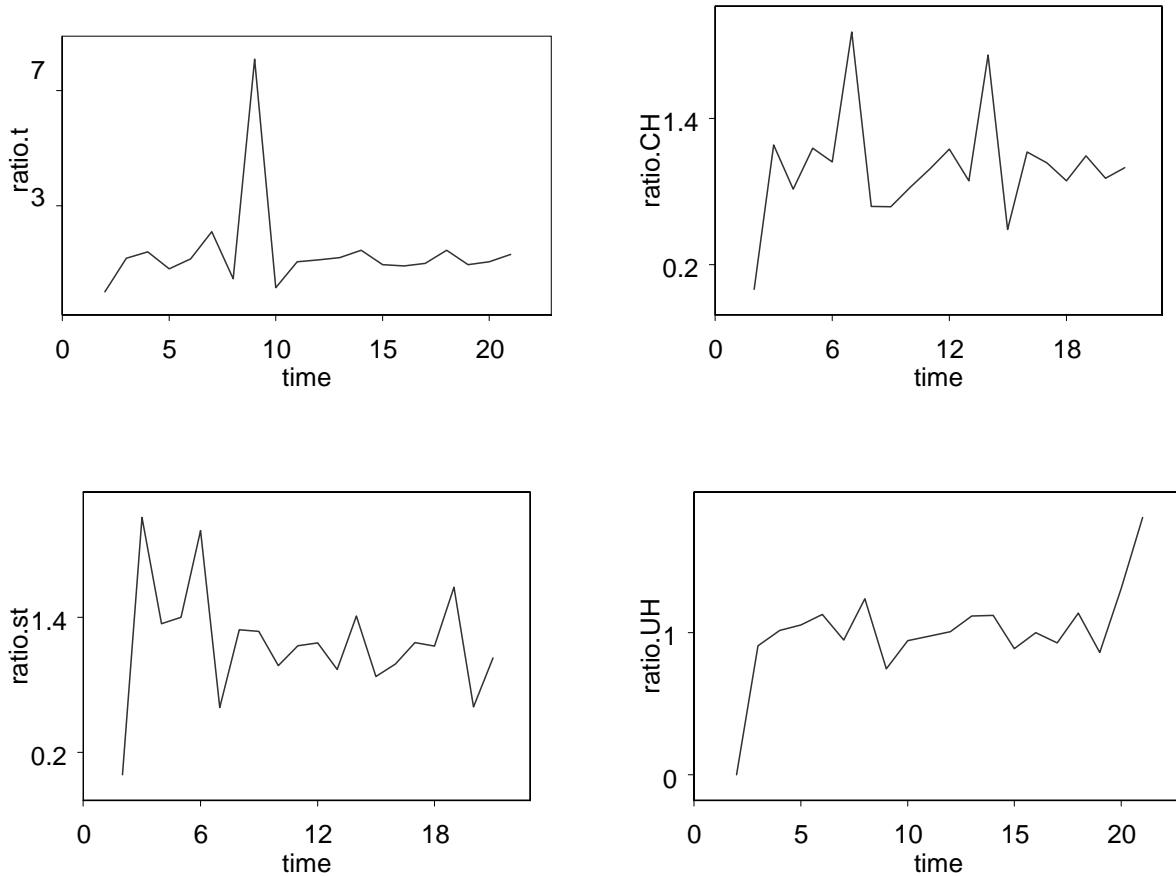


Ohio_smr1

■ 1.19 to 1.83 (16)
■ 0.94 to 1.19 (16)
■ 0.83 to 0.94 (18)
■ 0.61 to 0.83 (18)
□ 0 to 0.61 (20)

1968

Ohio_smr2

- ■ 1.29 to 1.92 (14)
- ▨ 0.98 to 1.29 (20)
- ▨ 0.85 to 0.98 (16)
- ▨ 0.61 to 0.85 (20)
- □ 0 to 0.61 (18)

1969



Ohio_smr20

- ■ 1.12 to 1.6 (20)
- ▨ 1.05 to 1.12 (10)
- ▨ 0.88 to 1.05 (21)
- ▨ 0.73 to 0.88 (17)
- □ 0 to 0.73 (20)

1987

Ohio_smr21

■ 1.15 to 1.58 (15)
■ 1.07 to 1.15 (17)
■ 0.88 to 1.07 (19)
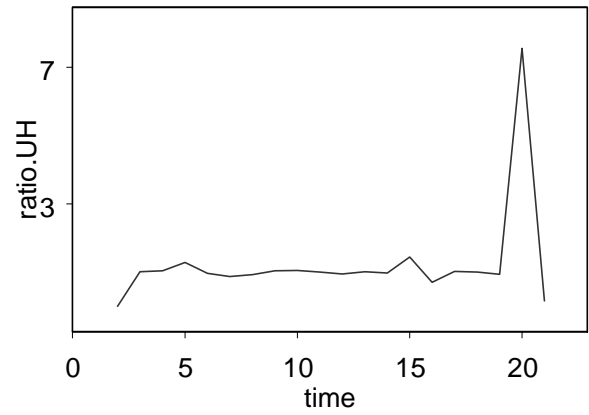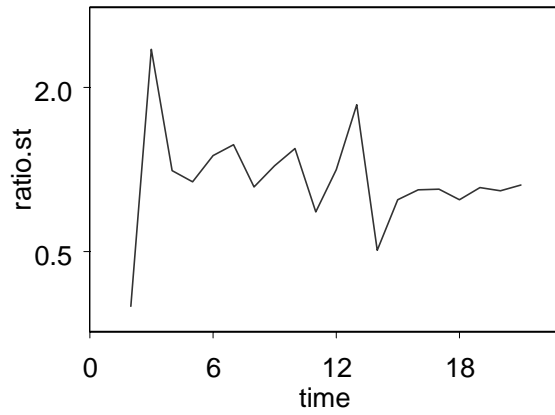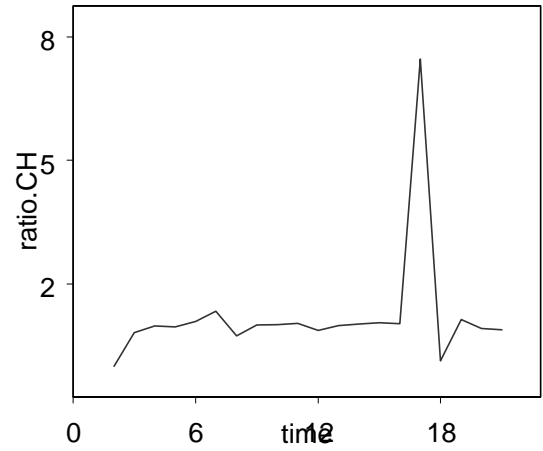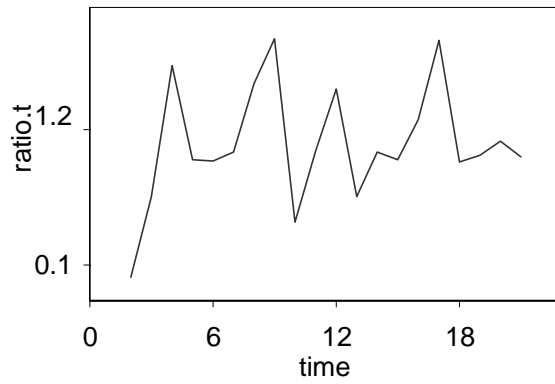□ 0.66 to 0.88 (18)
□ 0 to 0.66 (19)

1988

# Results



Empirical lag-one ratios of variance parameters for the independence S-T interaction model (top left: time; bottom left: space-time; right panels: spatial correlated and uncorrelated ratios)

Empirical lag-one ratios of variance parameters for the random walk S-T interaction model.

# Discussion

- Empirical ratios of variances as crude estimates of the $K_*$ parameters only reported here

- Further hyperpriors could be used to distinguish the parameters but our need for parsimony deters such steps.

- The variance parameters display a variety of differences: The spatial parameters are similar between models and show a few isolated peaks particularly near the end of the period and ratios close to one except near the period end. This suggests short term changes in variability of the counties and also possibly some clustering.

- The time variances show increasing trend in both models. The spatio-temporal increase suggests that as time proceeds there are changes in localised incidence.

- There is some evidence of interaction between the unstructured spatial component and time parameters, and this may also be related to the spatio-temporal interaction effects.

- In other studies of the Ohio data set it has been found that there is considerable upward trend overall in the study region, while there are localised differences in pattern.

- In Knorr-Held(2000) rural increases in trend in interaction were found. Here we report the existence of jumps or changepoints in the temporal increases and in the interactions suggesting clustering changes.

- Carlin and Louis (1996) found marked positive temporal trends for the study area and also marked increases in the south west counties (Hamilton, Clermont, Butler and Warren). Although in the case of Hamilton there is the presence of a large urban area (Cincinnati).

- Neither previous analysis could provide the temporal measure of parameter change which is provided here.

- While we don't examine in depth the *spatial* distribution of the parameter change results, we note the importance of being able to flag, in real time, any changes to pattern which arise and to be able to examine quickly the corresponding spatial distribution of changes.

- The sequential approach could be used to isolate forms of variation in trend, s-t clustering, and global changes in the spatial structure. Bayesian alarm monitoring could form the next extension to this study.

- Another feature of this approach is the examination of overall goodness-of-fit (GOF) over time. As time proceeds we can assess whether our model fits the data well or if it diverges globally in its goodness-of-fit. As we are not in a 'control' system and therefore can't adjust the behaviour of the system immediately, we would want to alter our model if the global goodness-of-fit suggested a serious lack of fit.

- To this end we here display the sequential Bayesian Information Criterion (SBIC) for the interaction dependence model. Note that as time proceeds then the number of parameters, and the number of data points

also increase and so the  balance between the likelihood and penalty could change because of this.

● Displayed is the successive BIC values for this model.
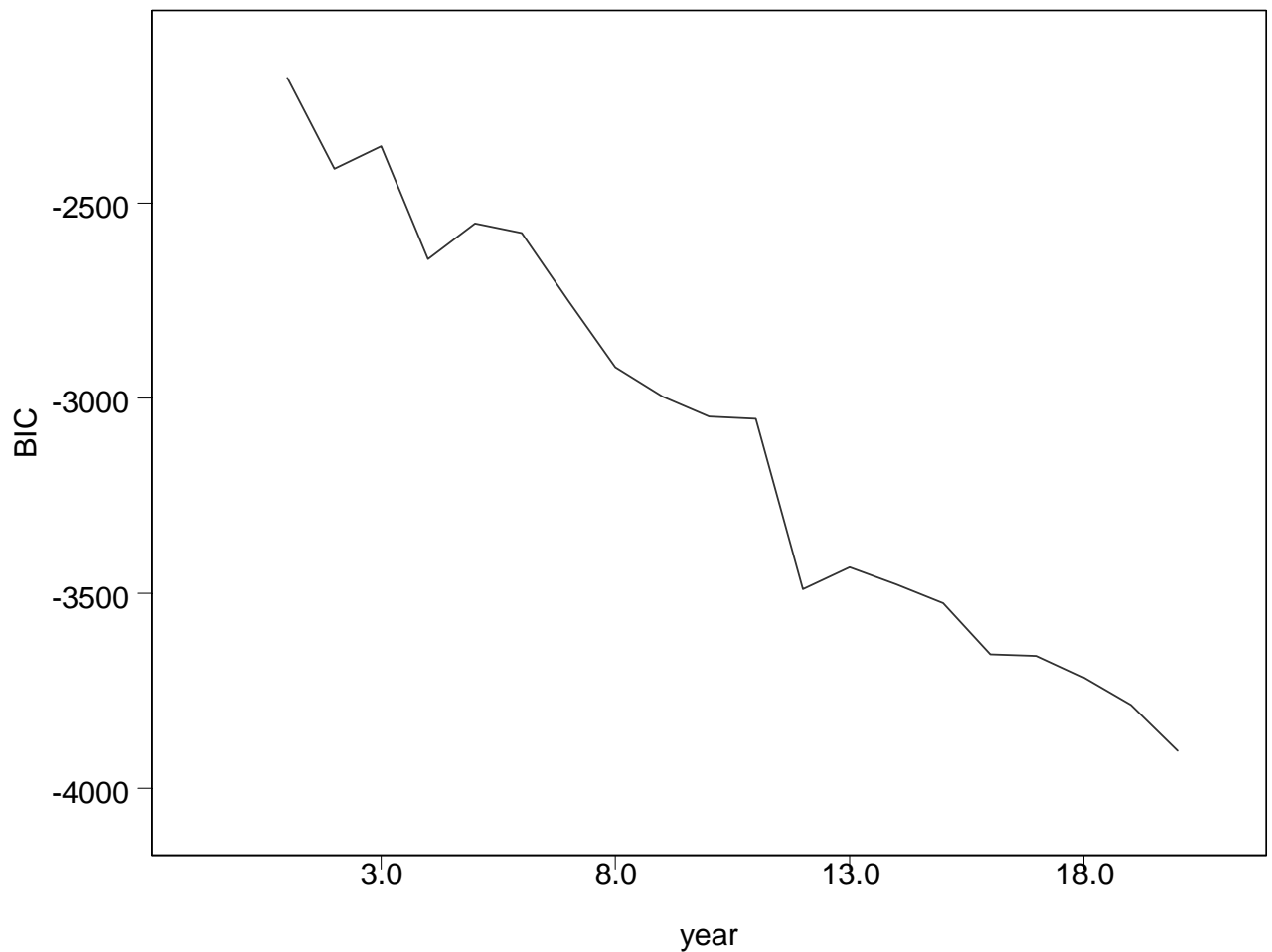


Successive BIC values for the interaction dependence model

- As time proceeds there is an overall reduction in validity of the model. Of course this may suggest that a more flexible modelling strategy should be envisaged. One possibility would be to include a general AR1 prior specification for the interaction component, thereby allowing a range of models within one prior:

$$\gamma_{jt} \sim N(\alpha \gamma_{jt-1}, K_4.\sigma_{st}^2).$$

- Of course care would have to be exercised in separation of non-stationary components, and this also raises the question of how to determine which components of the model are lacking in GOF, and how to allow for this within an active sequential surveillance system.

# Conclusions

- There is considerable scope for development of new methods within the general area of surveillance of disease maps.

- There is a need to develop spatial methods which are sensitive to the sequential nature of the surveillance task. This could be via updating algorithms or through the sequential methods discussed.

- Ultimately it would be useful to develop methods which could be employed easily or routinely within a public health surveillance context.

- This development may require both methods development, dissemination and the incorporation of methods into a suitable surveillance system as tools which can be used by public health analysts.