.

Generalized Spatial Structural Equation Models

Melanie M. Wall

Division of Biostatistics

School of Public Health

University of Minnesota,

Co-authors: Xuan Liu, James S. Hodges

Structural equation modeling (SEM)

- Most generally, structural equation modeling combines the ideas of **factor analysis** with **regression**.
- researcher interested in possibly several regression-type relationships, but some or all of the variables of interest can not be measured directly (i.e., they are **latent**).
- a set of observable variables is assumed to represent imperfect measure of the underlying latent variable of interest
- Then a structural equation model assumes a factor analysis type model to "measure" the latent variables via the multiple imperfect measures, while simultaneously assuming a regression-type model for the relationship among the latent variables.

SEM - Traditionally

Restricted to:

- normally distributed observed variables
- linear relations among the latent variables
- independently observed individuals
- used in sociology and psychology

SEM - Developments

For example,

- Methods for allowing observed variables of mixed types from an exponential family (Muthén 1984; Sammel et al., 1997; Moustaki and Knott, 2000)
- Methods for including nonlinear relationships among latent variables (Wall and Amemiya, 2000, 2001; Lee and Zhu, 2002; Lee and Song, 2003)
- Methods for clustered individuals, i.e., "multi-level" sampling designs (McDonald and Goldstein, 1989; Muthén,1989; Dunson, 2000; Lee and Shi 2001, Rasbash et al., 2002)

In this paper/talk extend the method for use with geographically referenced population based public health data.

Multivariate geographically referenced data in public health

Multivariate geographically referenced (e.g. state, county, census tract) data are very common in population-based data sources used for assessing public health and socioeconomic research.

- Vital records - births and deaths - are geographically coded to county of residence and often coded to smaller regions.
- National Cancer Institute - SEER - County-level incidence rates for different cancers
- Behavioral Risk Factor Surveillance Survey - population phone survey contains county level information.
- Several different education and economic variables collected by the US Census summarized geographically

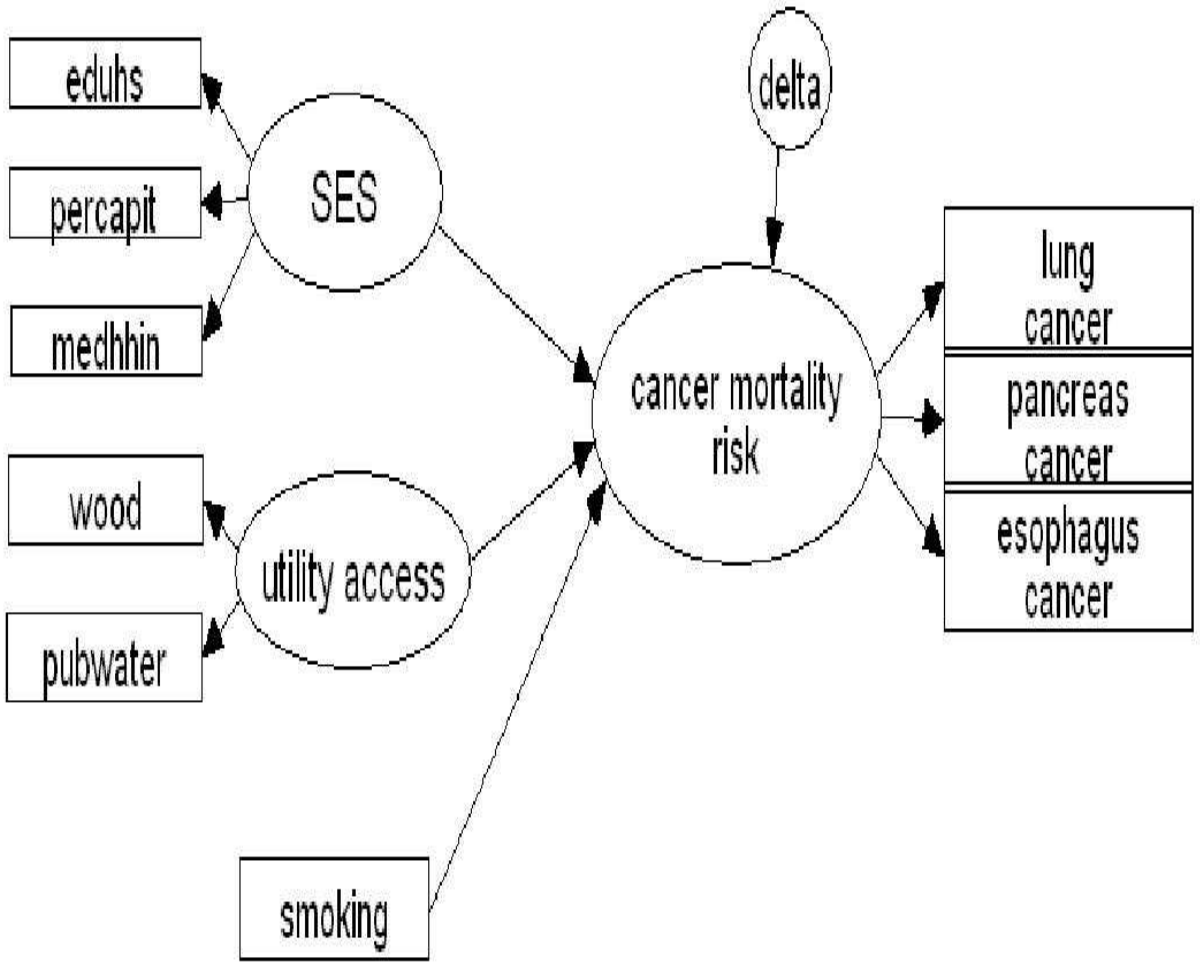
Multivariate geographically referenced data in public health

- Large number of variables available at each geographic unit
- Possibility of performing ecological-type regressions for investigating influences of risk factors on outcomes.
- Health Disparities Initiative - Quantify and assess relations between poverty, minorities, and health.

Example - Minnesota cancer mortality data

- Three groups of observed variables of Minnesota counties are used corresponding to three underlying factors.
  - Death counts due to *esophagus*, *pancreas*, and *lung* cancer, sharing underlying factor called common cancer risk factor
  - Three census variables, *high school education*, *median household income*, and *percapita income* measuring a underlying factor called social economic status (SES)
  - Two census variables, *public water*, and *home heat wood* measuring a underlying factor called access to public utilities
- County level smoking prevalence variable of Minnesota - a known covariate of interest (BRFSS)
- The interest of this study is the relationship between the shared common cancer factor, social economic status, access to public utilities, and smoking.

Example - Minnesota cancer mortality data




CANADA
MEXICO
Edmonton
Vancouver
Regina
Winnipeg
Seattle
Olympia
Washington
Montana
North Dakota
Minnesota
Québec
Ottawa
Montréal
Maine
Augusta
Salem
Helena
Bismarck
Oregon
Idaho
Boise
South Dakota
Pierre
St. Paul
Wisconsin
Michigan
Toronto
Montpelier
Albany
Concord
Boston
Providence
New York
Wyoming
Madison
Lansing
Detroit
Harrisburg
Philadelphia
Trenton
Dover
Annapolis
Washington, D.C.
Sacramento
Carson City
Salt Lake City
Cheyenne
Nebraska
Lincoln
Des Moines
Iowa
Chicago
Nevada
Utah
Denver
Colorado
Illinois
Springfield
Indiana
Indianapolis
Ohio
Columbus
San Francisco
Topeka
St. Louis
Jefferson City
Missouri
Kansas
Charleston
Richmond
Virginia
Frankfort
Kentucky
California
Las Vegas
Raleigh
Nashville
North Carolina
Tennessee
Los Angeles
Santa Fe
Oklahoma
Oklahoma City
Memphis
Columbia
Arizona
Phoenix
New Mexico
Arkansas
Little Rock
Atlanta
North Atlantic Ocean
North Pacific Ocean
Mexicali
Dallas
Mississippi
Alabama
Montgomery
Georgia
Jackson
Louisiana
Texas
Tallahassee
Austin
Baton Rouge
New Orleans
Florida
Hermosillo
Houston
Miami
THE BAHAMAS
Nassau
Chihuahua
Gulf of Mexico
Monterrey
State capital
Scale 1:27,000,000
Albers Equal-Area Projection
standard parallels 28°30'N and 45°30'N
0
500 Kilometers
0
500 Miles

Summary of Minnesota Census data

Population around 5 million

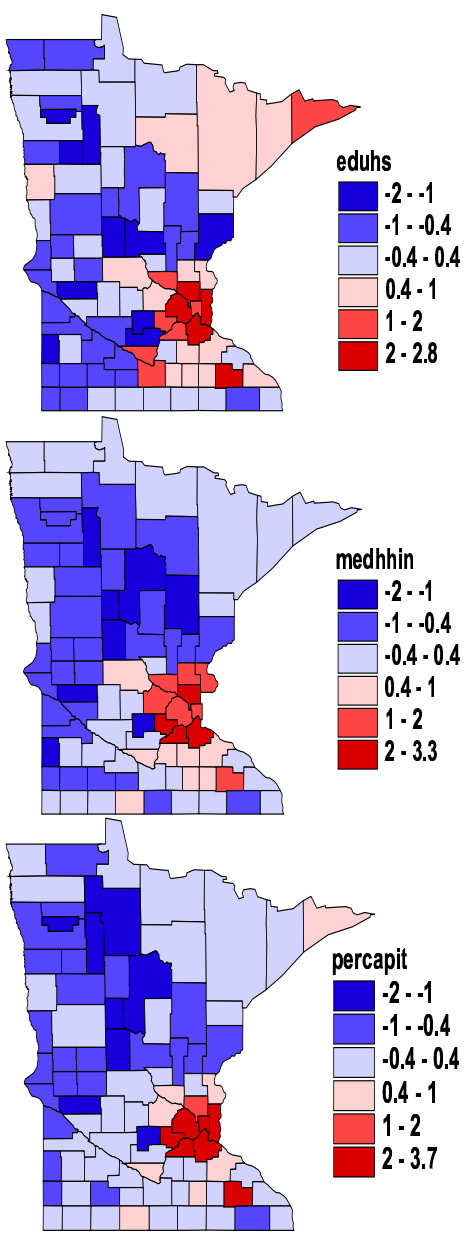
Selected 1990 census variables.

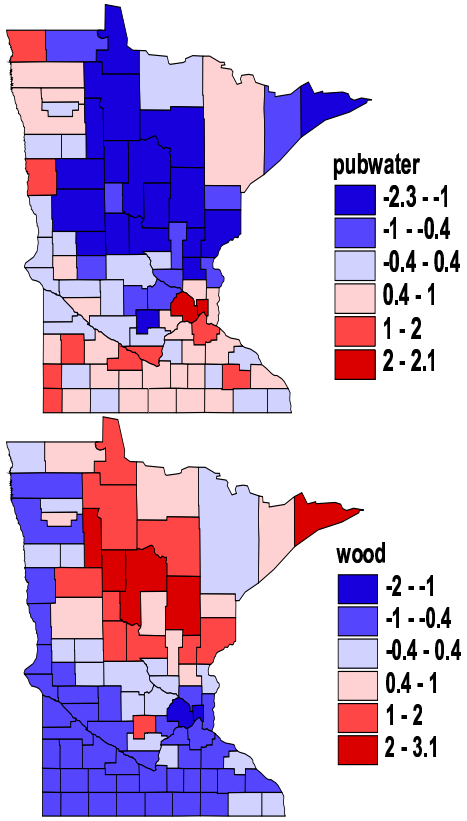
For the whole state: household median income $30,900,
per capita income $14,389, 73% access to public water,
4.9% use wood to heat home, 82.4% have at least high school education

Summary of the 87 individual counties...

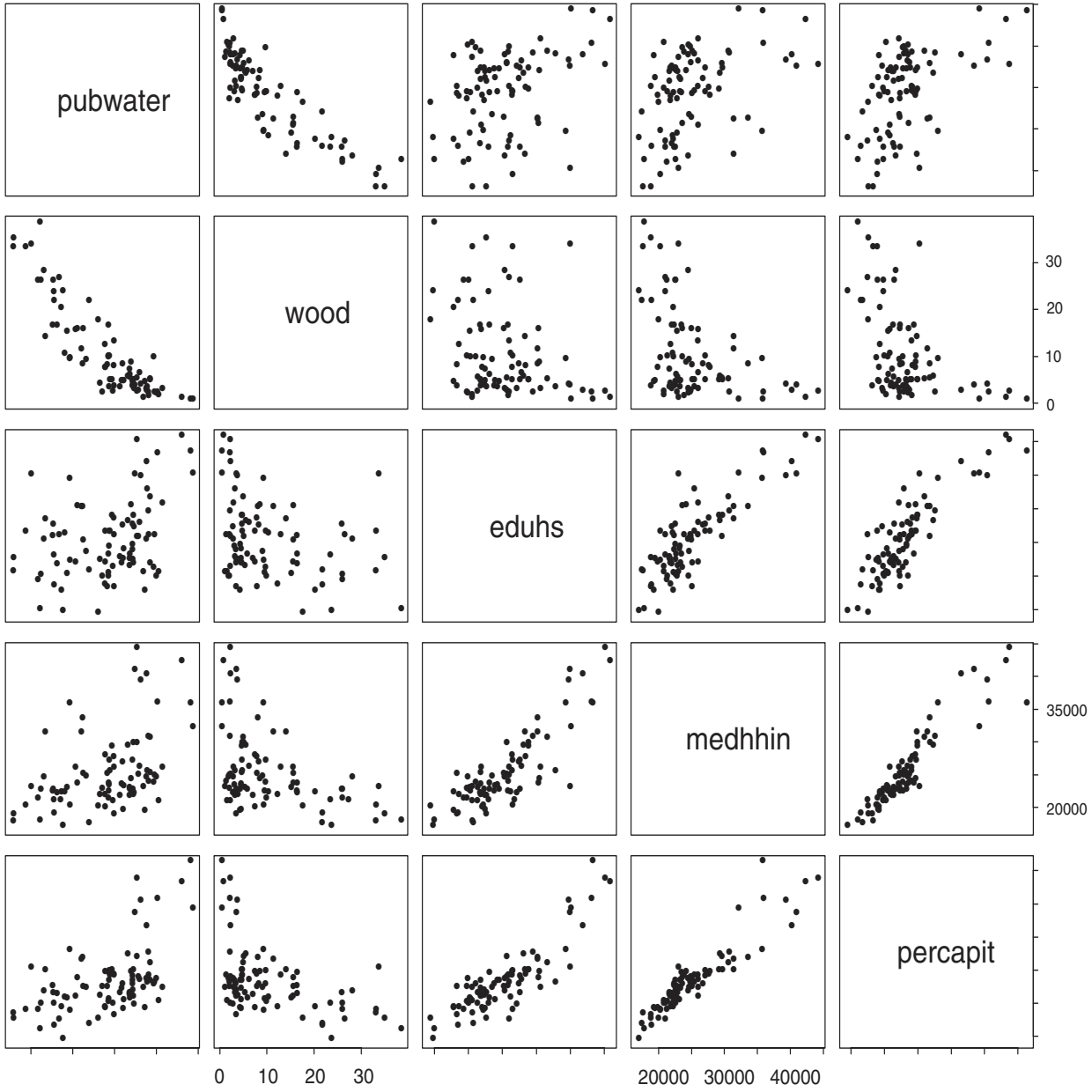
variable		mean	min	max
eduhs	Per cent with high-school education	75.2	64.3	90.7
medhhin	Median household income (in dollars)	25 052	16 924	44 122
percapit	Per capita income (in dollars)	11 227	7737	18 496
pubwater	Per cent of households with access to public water	56.4	11.2	97.1
wood	Per cent of households using wood to heat the home	10.0	0.3	38.4

Maps of five of the 5 census variables





Scatterplot of the 5 census variables by county



Summary of Minnesota County Cancer data

	lung cancer		pancreas cancer		esophagus cancer	
	Min	Max	Min	Max	Min	Max
Observed	15	3797	2	830	0	319
Expected	21	3528	5	791	2	298

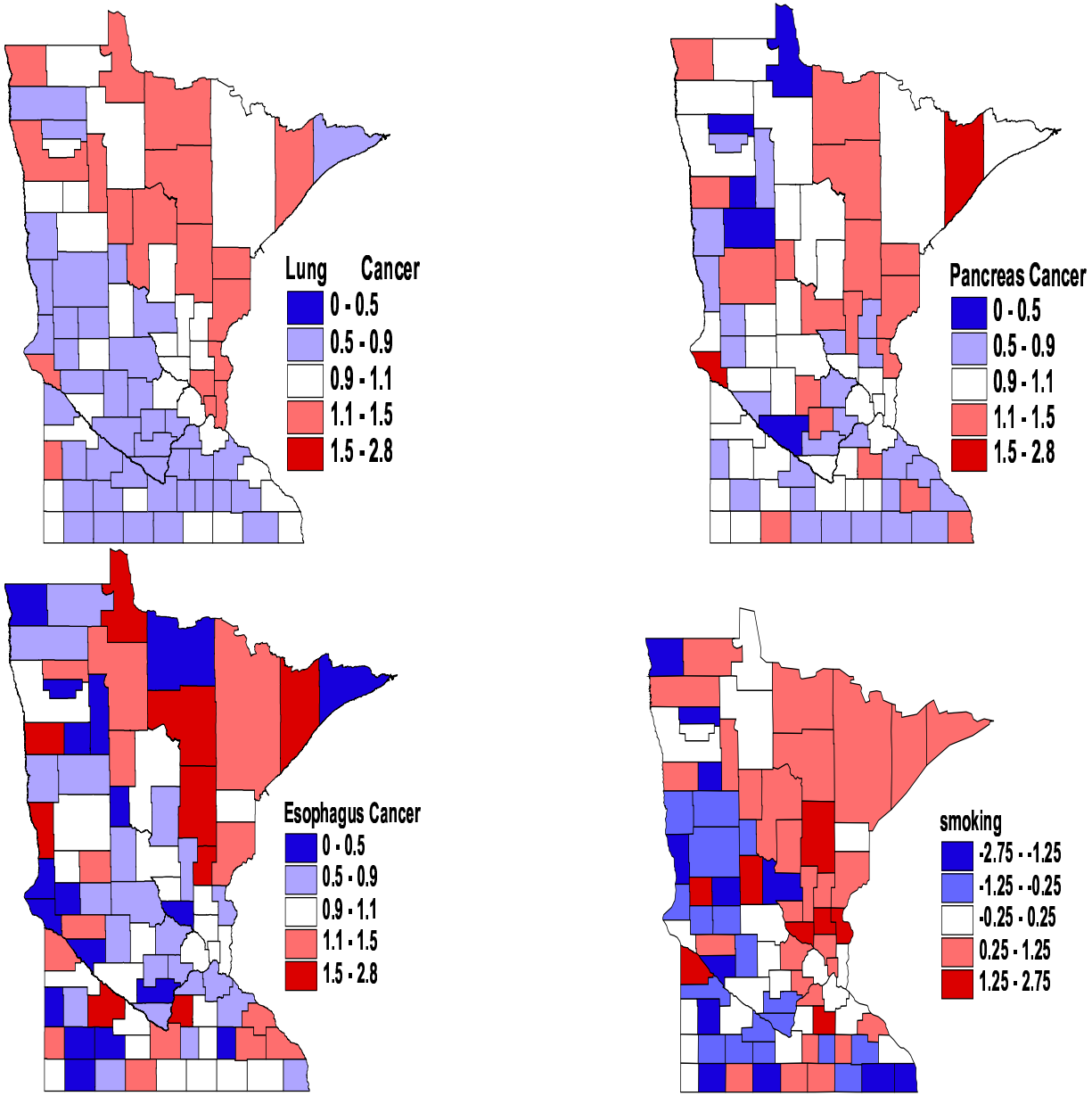
Expected counts

$E_{ij} = \sum_k \lambda_{jk} N_{ik}$

Where λ_{jk} is the statewide "age specific" mortality rate for cancer j in the kth age group, and N_{ik} is the population of people in county i who are in the kth age group.

This use of Expected counts (which will serve as the "offset" in the later Poisson model) prevents the results being confounded with age.

SMR cancer and smoking percents



Related methods for multivariate Spatial data

Methods for reducing dimensionality

- Related methods mainly use principle component methods on the variance covariance matrix of the data to generate components of different spatial scales (e.g., Switzer and Green, 1984, Grunsky and Agterberg, 1992, Grzebyk and Wackernagel 1994; Wackernagel 2003; or Banerjee, Carlin, and Gelfand 2004)
- Descriptive in the sense that they perform direct operation on the data instead of making explicit statistical modeling assumptions.

Model based spatial factor analysis

- Christensen and Amemiya (2002, 2003) developed exploratory factor analysis frameworks for multivariate spatial data aiming to explore the relationship between the underlying factors and the observed variables, applied their methods on agricultural data.

In our approach concentrate on full statistical modeling based methods. More confirmatory than exploratory.

Traditional factor analysis model

- Let $\mathbf{Z}_i$ be a $p \times 1$ observed random vector, then

$$\mathbf{Z}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_i + \mathbf{e}_i, \ (i = 1, \ldots, n)$$

  where $\mathbf{f}_i$ is a $Q \times 1$ vector of underlying factors, $\boldsymbol{\Lambda}$ is a $p \times Q$ matrix of unknown parameters, called factor loadings matrix and $Q << P$.

- The assumptions are:
  - $\mathbf{f}_i \overset{iid}{\sim} N_Q(\mathbf{0}, \boldsymbol{\Phi}), \ (i = 1, \ldots, n)$
  - $\mathbf{e}_i \overset{iid}{\sim} N_p(\mathbf{0}, \boldsymbol{\Psi}), \ (i = 1, \ldots, n)$ where $\boldsymbol{\Psi} = Diag(\psi_1, \ldots, \psi_p)$,
  - $\mathbf{e}_i$ and $\mathbf{f}_i$ are independent.

- These assumptions imply that any correlations found between the variables in $\mathbf{Z}_i$ are the result of their relationship with the shared underlying factors $\mathbf{f}_i$.

The generalized single common spatial factor model

(Wang and Wall, 2003)

Variable correlated within site because they share a common latent factor, variables correlated across sites because the common factor is spatially distributed.

- Let Z_{ij} be the jth random variable observed at location s_i (i=1,. . . ,n, j=1,. . . ,p),

$$Z_{ij} \mid \theta_{ij}, \sigma_j \overset{ind}{\sim} F(\theta_{ij}, \sigma_j^2),$$

  and for some function $g(\cdot)$,

$$g(\theta_{ij}) = \mu_j + \lambda_j f^{(s_i)},$$

- Assumptions:
  - There is 1 underlying common factor, i.e. $Q = 1$.
  - Z_{ij}'s given the underlying factors are independent.
  - Let $\mathbf{f} = (f^{(s_1)}, \ldots, f^{(s_n)})'$, then

$$\mathbf{f} \sim N(\mu_{\mathbf{f}} \mathbf{1}_n, \mathbf{C}(\boldsymbol{\alpha})),$$

    where $\mathbf{C}(\boldsymbol{\alpha})$ is a spatial covariance matrix, e.g. exponential covariance structure or CAR (conditional autoregressive).

Use of Generalized Common Spatial factor model

- Minnesota cancer mortality data - Wang and Wall (*Biostatistics*, 2003)

$$Z_{ij} \mid \theta_{ij} \overset{ind}{\sim} Poi(\theta_{ij}), \qquad (i = 1, \ldots, n, \quad j = 1, \ldots, p)$$
$$log(\theta_{ij}) = log(E_{ij}) + \lambda_j f^{(s_i)}, \qquad (i = 1, \ldots, n, \quad j = 1, \ldots, p)$$

  where $n = 87$ represents the 87 counties of Minnesota; $p = 4$ county-level death counts due to cancer of the lung, pancreas, esophagus, and stomach; E_{ij} known constant for standardized expected number of deaths in county s_i for cancer j. A conditional autoregressive (CAR) structure was used for the spatial covariance structure for $\mathbf{f}$.

- Material deprivation - Hogan and Tchernis (*JASA*, 2004)

$$Z_{ij} \mid \theta_{ij}, \sigma_j \overset{ind}{\sim} N(\theta_{ij}, \sigma_j^2), \qquad (i = 1, \ldots, n, \quad j = 1, \ldots, p)$$
$$\theta_{ij} = \alpha_{ij} + \lambda_j f^{(s_i)}, \qquad (i = 1, \ldots, n, \quad j = 1, \ldots, p).$$

  Used p census variables related to ownership of property and goods collected in census tracts of Rhode Island. Proposed several parametric covariance structures $\mathbf{C}(\boldsymbol{\alpha})$ for the underlying factor including a combination of geostatistical and areal spatial analysis methods where a CAR model is used with a distance-dependent neighbor structure. A posterior predictive criterion was used for selection.

Generalized spatial structural equation model (GSSEM): motivation

- Structural equation models (SEM) offer a unified method by combining factor analysis and regression analysis.
- A SEM incorporates:
  - A **measurement model** - relating observed data to latent variables.
  - A **structural model** - relating latent variables to other latent variables.
- The motivation of GSSEM is to extend the SEM methods to **spatial data**.

GSSEM : The model

- Measurement model:
  Suppose there are Q underlying factors measured by Q separate groups of observed variables:

$$Z_{ij}^{(q)} \mid \theta_{ij}^{(q)}, \sigma_j^{(q)} \overset{ind}{\sim} F(\theta_{ij}^{(q)}, \sigma_j^{2^{(q)}}), \qquad (i = 1, \ldots, n, \quad j = 1, \ldots, p_q)$$

$$g(\theta_{ij}^{(q)}) = \boldsymbol{\beta}_j^{(q)} \mathbf{x}_{ij}^{(q)} + \lambda_j^{(q)} f_q^{(s_i)}, \qquad (i = 1, \ldots, n, \quad j = 1, \ldots, p_q),$$

  where $\mathbf{x}_{ij}^{(q)}$ is a vector of possible observed covariates relating specifically to the jth variable measuring the qth factor.

GSSEM : The model (cont.)

- Structural model:

  Let $\mathbf{f}_i$ a vector of the Q underlying factors at location s_i. We partition $\mathbf{f}_i$ into two vectors of lengths Q_1 and Q_2, i.e. $\mathbf{f}_i^T = (\mathbf{f}_{1i}^T, \mathbf{f}_{2i}^T)$. Then,

$$\mathbf{f}_{1i} = \boldsymbol{\Xi}\mathbf{X}_i + \boldsymbol{\Gamma}\mathbf{f}_{2i} + \boldsymbol{\delta}_i,$$

  where

  $\mathbf{f}_{1i}$ is called *endogenous* (dependent) underlying factors
  $\mathbf{f}_{2i}$ is called *exogenous* (independent) underlying factors

  and the $\mathbf{X}_i$ is a matrix of possible observed covariates influencing the endogenous factor, $\boldsymbol{\Xi}$ and $\boldsymbol{\Gamma}$ are matrices of unknown constants, and $\boldsymbol{\delta}_i$ is a Q_1 vector of errors that is independent of $\mathbf{f}_{2i}$.

GSSEM : The model (cont.)

- Spatial distributions for $\mathbf{f}_i$ and $\boldsymbol{\delta}_i$: We specify those two multivariate spatial process using the **linear model of coregionalization** (LMC) idea, i.e.

$$\boldsymbol{\delta}_i = \mathbf{A}_1 \mathbf{w}_i, \qquad (i = 1, \ldots, n)$$

$$\mathbf{f}_{2i} = \mathbf{A}_2 \mathbf{v}_i, \qquad (i = 1, \ldots, n)$$

  where $\mathbf{A}_1$ and $\mathbf{A}_2$ are two upper triangular matrix of size Q_1 and Q_2; $\mathbf{w}_i$ and $\mathbf{v}_i$ are independent zero-mean and unit-variance spatial processes of dimension Q_1 and Q_2, where these independent spatial processes can be from different spatial parametric distributions, such as, isotropic exponential distributions, or conditional autoregressive distributions (CAR).

Linear model of coregionalization (LMC)

- LMC is a flexible method to jointly modeling multiple processes.
- Let $\mathbf{Y}(\mathbf{s}_i)$ $(p \times 1)$ be a realization of p variate spatial process at location $\mathbf{s}_i$. $(i = 1, \ldots, n)$. Then LMC model:

$$\mathbf{Y}(\mathbf{s}_i) = \mathbf{A}\mathbf{v}(\mathbf{s}_i),$$

  where $\mathbf{A}$ is a $p \times p$ full rank matrix, $\mathbf{v}(\mathbf{s}_i)$ is a vector of p independent spatial processes $v_j(s_i)$ $(j = 1, \ldots, p)$ at s_i.

  - If the p independent processes have the same covariance structure with covariance matrix $\mathbf{C}$, then the covariance matrix of $\mathbf{Y}^T = (\mathbf{Y}(\mathbf{s}_1), \ldots, \mathbf{Y}(\mathbf{s}_n))$, $\mathbf{\Sigma}$ is:

$$\mathbf{\Sigma} = \mathbf{C} \otimes \mathbf{T}, \quad \text{where } \mathbf{T} = \mathbf{A}\mathbf{A}^T.$$

  - If the p independent process $v_1, \ldots, v_p$ to have p different covariance structures with covariance matrix $\mathbf{C}_1, \ldots, \mathbf{C}_p$, then

$$\mathbf{\Sigma} = \sum_{j=1}^{p} \mathbf{C}_j \otimes \mathbf{T}_j,$$

    where $\mathbf{T}_j = \mathbf{A}_j \mathbf{A}_j^T$, and $\mathbf{A}_j$ is the jth column of $\mathbf{A}$.

- The covariance structures of $v_1, \ldots, v_p$ can take any parametric continuous spatial distributions with a proper covariance matrix, e.g. exponential, proper CAR structures.

Model fitting of the GSSEM

- A full Bayesian approach will be used to fit the the GSSEM.
- Prior specifications:
  - Vague normal priors for all the fixed constants in the mean structure of both the measurement and structural part
  - Vague Inverse Gamma priors for all the variance and spatial covariance components in the model
  - Vague Inverse Gamma priors for the diagonal elements of matrix $\mathbf{A}_1$ and $\mathbf{A}_2$ and vague normal priors for their off diagonal elements

Model fitting of the GSSEM (cont.)

- Posterior distributions: Let $\mathbf{Z}$ be the $(np) \times 1$ vector of all the observed variables, $\mathbf{f}_1$ be a $(Q_1 n) \times 1$ vector of endogenous factors, and $\mathbf{f}_2$ be a $(Q_2 n) \times 1$ vector of exogenous factors over n locations. Then the joint posterior of all the unknown parameters and the factors is:

$$P(\mathbf{\Phi}, \mathbf{f}_1, \mathbf{f}_2 \mid \mathbf{Z}, \mathbf{x}) \propto P(\mathbf{Z} \mid \mathbf{x}, \mathbf{f}_1, \mathbf{f}_2, \{\boldsymbol{\beta}^{(q)}, \boldsymbol{\sigma}^{2^{(q)}}, \boldsymbol{\lambda}^{(q)} : \ q = 1, \ldots, Q\})$$
$$P(\mathbf{f}_1 \mid \mathbf{X}, \mathbf{\Gamma}, \mathbf{f}_2, \mathbf{\Xi}, \boldsymbol{\alpha}_\delta) P(\mathbf{f}_2 \mid \mathbf{v}, \boldsymbol{\alpha}_v) P(\mathbf{\Phi})$$

- The marginal posterior distributions can be obtained through sampling based MCMC method, i.e. Gibbs sampler and Metropolis-Hasting algorithm.

- The model fitting can be implemented in Winbugs.

Measurement model

$Z_1^{(1)}, Z_2^{(1)}, Z_3^{(1)}$ are the *esophagus*, *pancreas*, and *lung* cancer death counts for each county i which manifest the first factor (cancer mortality risk)

$Z_1^{(2)}, Z_2^{(2)}, Z_3^{(2)}$ are *eduhs*, *medhhin* and *percapit*, measuring the second factor (SES)

$Z_1^{(3)}, Z_2^{(3)}$ are *wood* and *pubwater*, measuring the third factor (utility accessibility)

For simplicity, drop county index i. The **measurement model** for this example is:

$$
\begin{aligned}
Z_1^{(1)} \mid \theta_1^{(1)} &\sim Poi(\theta_1^{(1)}) \\
Z_2^{(1)} \mid \theta_2^{(1)} &\sim Poi(\theta_2^{(1)}) \\
Z_3^{(1)} \mid \theta_3^{(1)} &\sim Poi(\theta_3^{(1)}) \\
Z_1^{(2)} \mid \theta_1^{(2)}, \sigma_1^{2^{(2)}} &\sim N(\theta_1^{(2)}, \sigma_1^{2^{(2)}}) \\
Z_2^{(2)} \mid \theta_2^{(2)}, \sigma_2^{2^{(2)}} &\sim N(\theta_2^{(2)}, \sigma_2^{2^{(2)}}) \\
Z_3^{(2)} \mid \theta_3^{(2)}, \sigma_3^{2^{(2)}} &\sim N(\theta_3^{(2)}, \sigma_3^{2^{(2)}}) \\
Z_1^{(3)} \mid \theta_1^{(3)}, \sigma_1^{2^{(3)}} &\sim N(\theta_1^{(3)}, \sigma_1^{2^{(3)}}) \\
Z_2^{(3)} \mid \theta_2^{(3)}, \sigma_2^{2^{(3)}} &\sim N(\theta_2^{(3)}, \sigma_2^{2^{(3)}}).
\end{aligned}
$$

Measurement model continued

Let $\boldsymbol{\theta}^T = (\theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}, \theta_1^{(2)}, \theta_2^{(2)}, \theta_3^{(2)}, \theta_1^{(3)}, \theta_2^{(3)})$, then define the link and joint mean structure as

$$\mathbf{g}(\boldsymbol{\theta}) \quad = \quad \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}$$

$$\begin{pmatrix} log(\theta_1^{(1)}) \\ log(\theta_2^{(1)}) \\ log(\theta_3^{(1)}) \\ \theta_1^{(2)} \\ \theta_2^{(2)} \\ \theta_3^{(2)} \\ \theta_1^{(3)} \\ \theta_2^{(3)} \end{pmatrix} = \begin{pmatrix} log(E_1) \\ log(E_2) \\ log(E_3) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \lambda_1^{(1)} & 0 & 0 \\ \lambda_2^{(1)} & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \lambda_1^{(2)} & 0 \\ 0 & \lambda_2^{(2)} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda_1^{(3)} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}$$

where E_1, E_2, E_3 are age-adjusted expected number of deaths for the three cancers.

Structural model

The f_1, f_2 and f_3 are the underlying factors representing cancer mortality risk, SES, and utility accessibility, respectively. The relationships among the factors is then modeled in the structural equation model:

$$f_1 = \beta H + \gamma_1 f_2 + \gamma_2 f_3 + \delta,$$

where H is the fixed covariate, smoking prevalence and β, γ_1, and γ_2 are unknown constants.

Spatial model for factors and errors

Let $\boldsymbol{\delta}$ be the vector of δ over the 87 Minnesota counties. Since we are only considering one structural model, i.e., $Q_1 = 1$, it is natural to assume that $\boldsymbol{\delta}$ has a univariate CAR covariance structure with covariance parameters $\boldsymbol{\alpha}_\delta = (\tau_\delta, \rho_\delta)$, where τ_δ is the precision parameter and ρ_δ the spatial correlation parameter. We use LMC to model the joint distribution of $(f_2, f_3)^T$:

$$\begin{pmatrix} f_2 \\ f_3 \end{pmatrix} = \mathbf{A}\mathbf{v} = \begin{pmatrix} a_1 & a_2 \\ 0 & a_3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix},$$

where $\mathbf{v}_1$ and $\mathbf{v}_2$ are vectors of v_1 and v_2 over 87 Minnesota counties (the subscript i is again suppressed). We assume $\mathbf{v}_1$ and $\mathbf{v}_2$ are independent spatial processes with CAR covariance structures having overall scale parameter set to 1 and spatial correlation parameters ρ_{v_1} and ρ_{v_2} respectively.

Estimation

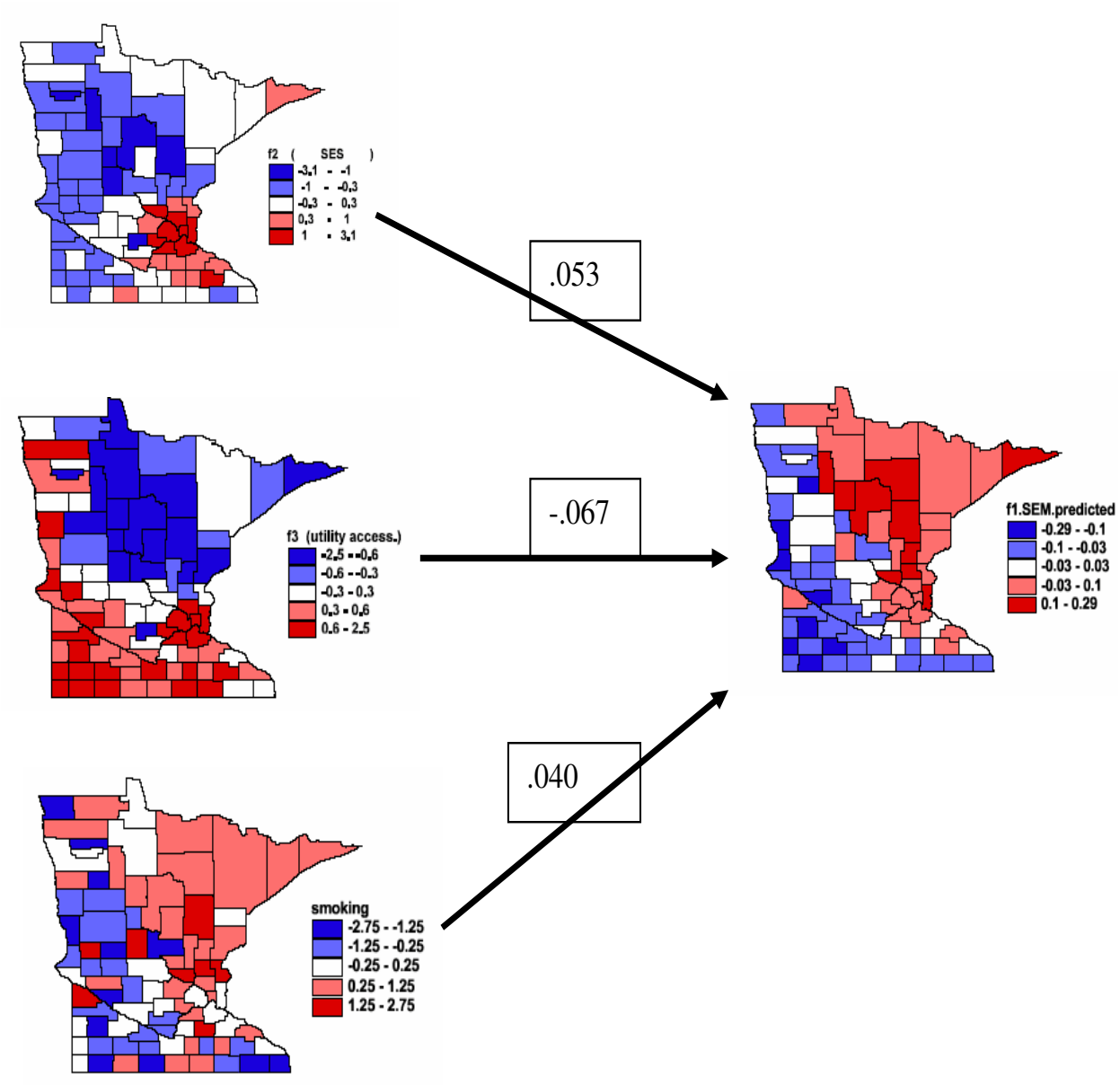
- Implemented in Winbugs software to obtain posterior summaries for the parameters.
- Ran three Markov Chains simultaneously started from different initial points.
- Monitor plots show that the three chains mixed well within 5000 iterations.
- The lag 1 posterior sample autocorrelations of most parameters (e.g., λs, β, γs, and ρs) are lower than 0.5, and the Gelman-Rubin diagnostic plot of the chains are mostly within the 0.8 to 1.2 band, which suggests satisfactory convergence.
- We also calculate the posterior summaries at different points of the Markov Chain, and the result summaries are almost identical after burn-in period. No thinning on the draws seems necessary.

Posterior Estimates

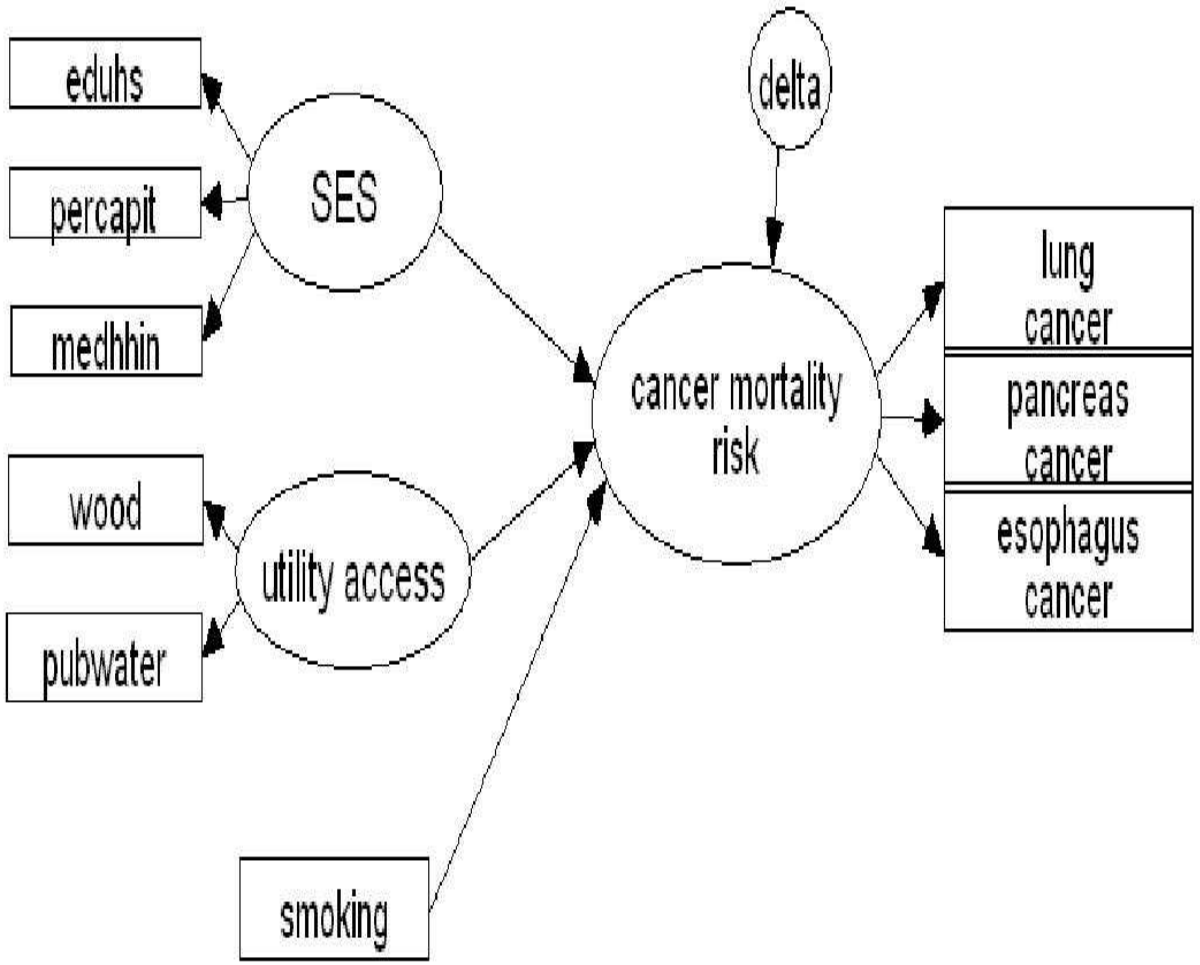
Table 2: Posterior summaries for Minnesota cancer data analyzed using GSSEM

	Posterior mean (Posterior .025 and .975 quantiles)					
	$\lambda_1^{(1)}$	0.82	(0.36, 1.28)	–	–	–
	$\lambda_2^{(1)}$	0.46	(0.19, 0.75)	–	–	–
Measurement	$\lambda_3^{(1)}$	1	–	–	–	–
Model	$\lambda_1^{(2)}$	0.90	(0.79, 1.02)	$\sigma_1^{2(2)}$	0.23	(0.16, 0.33)
Parameters	$\lambda_2^{(2)}$	0.97	(0.88, 1.07)	$\sigma_2^{2(2)}$	0.10	(0.06, 0.16)
	$\lambda_3^{(2)}$	1	–	$\sigma_3^{2(2)}$	0.047	(0.003, 0.094)
	$\lambda_1^{(3)}$	-1.15	(-1.29, -0.99)	$\sigma_1^{2(3)}$	0.016	(0.0005, 0.079)
	$\lambda_2^{(3)}$	1	–	$\sigma_2^{2(3)}$	0.23	(0.16, 0.32)
	β	0.04	(0.003, 0.073)	τ_δ	0.036	(0.004, 0.08)
Structural	γ_1	0.053	(0.008, 0.098)	ρ_δ	0.97	(0.88, 0.998)
Equation	γ_2	-0.067	(-0.13, -0.008)	ρ_{v_1}	0.96	(0.88, 0.997)
Parameters	a_1	1.49	(1.25, 1.76)	ρ_{v_2}	0.97	(0.91, 0.998)
	a_2	0.72	(0.45, 1.02)			
	a_3	1.06	(0.86, 1.3)			

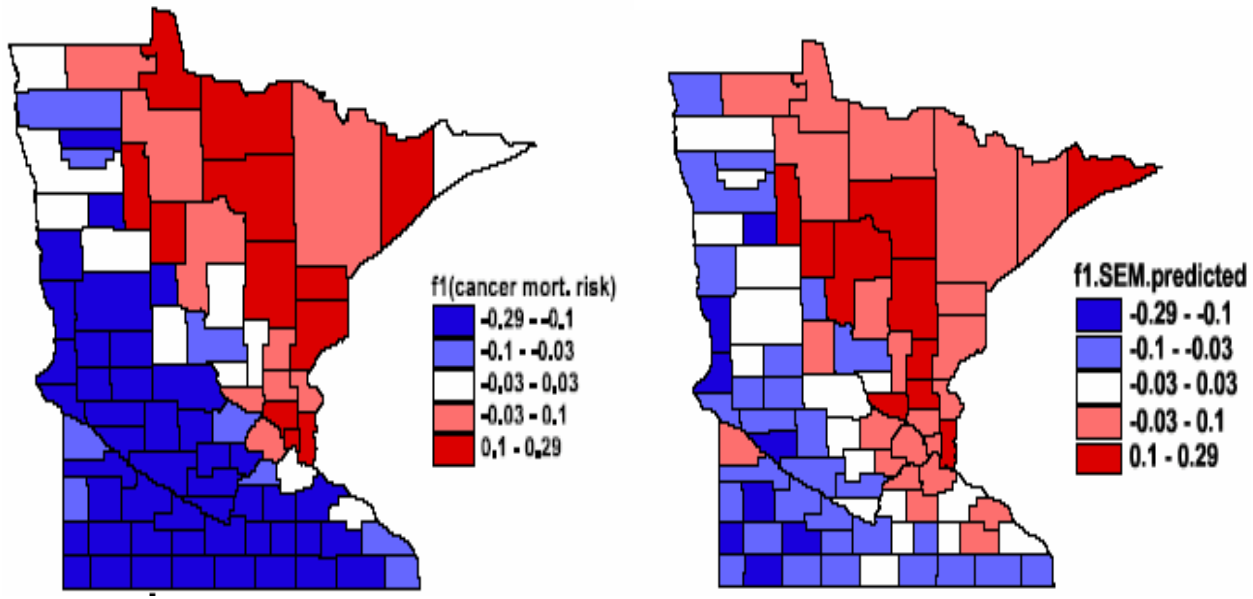
Posterior predicted values

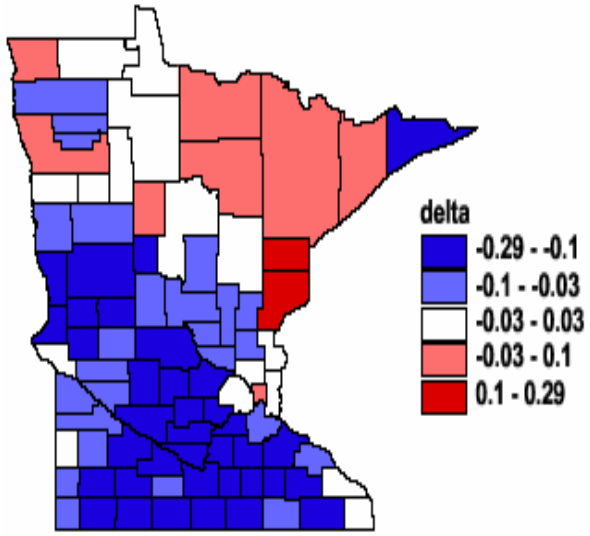


Example - Minnesota cancer mortality data



Residual - still a missing spatial factor?





Discussion

- Generalized spatial structural equation modeling provides a method for researchers to perform ecological regressions incorporating many correlated variables in a meaningful way while taking account of spatial structure.
- Model could be extended to incorporate both space and time.
- Model could be extended to allow underlying cluster process for underlying factors (factors could be categorical).

. .

C'est tout