CBMS Lecture 9

Alan E. Gelfand
Duke University

Spatio-temporal Models

- Point-referenced vs. areal unit data
- Continuous time vs. discretized time
- Association in space versus association in time!
  For geostatistical setting, t continuous, Gaussian data,

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + w(\mathbf{s}, t) + \epsilon(\mathbf{s}, t)$$

- For nonGaussian data, instead use appropriate likelihood with link $g(E(Y(\mathbf{s}, t))) = \mu(\mathbf{s}, t) + w(\mathbf{s}, t)$
- Don't treat time as a third coordinate – scale issue!

$$\text{sensible:}\ \ Cov(Y(\mathbf{s}, t), Y(\mathbf{s}', t')) = C(\mathbf{s} - \mathbf{s}', t - t')$$

$$\text{NOT sensible:}\ \ Cov(Y(\mathbf{s}, t), Y(\mathbf{s}', t')) = C((\mathbf{s}, t) - (\mathbf{s}', t'))$$

Spatio-temporal Models

- Separable form:

$$C(\mathbf{s}-\mathbf{s}', t-t') = \sigma^2 \rho_1(\mathbf{s}-\mathbf{s}'; \phi_1)\rho_2(t-t'; \phi_2)$$

- Limitation
- To create nonseparable versions:
  - Sum of independent separable processes
  - More generally, mixing of separable covariance functions
  - Spectral domain approaches - space-time spectral density and inverse Fourier transform

Spatio-temporal Models

- An EDA idea for space-time dependence
- Now suppose time is discretized, i.e. data are $Y_t(\mathbf{s}),\ t = 1, \ldots, T$
- Type of data: time series versus cross-sectional (e.g., real estate sales). Latter is problematic
- For time series data, exploratory analysis:
  - Arrange into an $n \times T$ matrix Y with entries $Y_t(\mathbf{s}_i)$
  - Center by row averages of Y yields Y_{rows}
  - Center by column averages of Y yields Y_{cols}
  - Sample spatial covariance matrix: $\frac{1}{T} Y_{rows} Y_{rows}^T$
  - Sample autocorrelation matrix: $\frac{1}{n} Y_{cols}^T Y_{cols}$
  - E, residuals matrix after a regression fitting

Empirical Orthogonal Functions

- Can learn about the structure of Y, Y_{rows}, Y_{cols}, E using empirical orthogonal functions
- Say for E, use singular value decomposition

$$E = UDV^T = \sum_{j=1}^{T} d_j \mathbf{u}_j \mathbf{v}_j^T$$

  where U is $n \times n$ orthogonal, V is $T \times T$ orthogonal and D is "almost diagonal"

- If we arrange the d_j in decreasing order then $\mathbf{u}_j \mathbf{v}_j^T$ is the jth empirical orthogonal function
- Typically, we only need a few terms in the sum to well approximate EE^T. With just the first term, suggests approximating $E(\mathbf{s}, t)$ by $u_1(\mathbf{s})v_1(t)$ - "separable".

Spatio-temporal Models

- If t discrete, a time series of spatial process realizations
- Modeling: $Y_t(\mathbf{s}) = \mu_t(\mathbf{s}) + w_t(\mathbf{s}) + \epsilon_t(\mathbf{s})$,
  or perhaps $g(E(Y_t(\mathbf{s})) = \mu_t(\mathbf{s}) + w_t(\mathbf{s})$
- For $\epsilon_t(\mathbf{s})$, independent $N(0, \tau_t^2)$
- For $w_t(\mathbf{s})$
  - $w_t(\mathbf{s}) = \alpha_t + w(\mathbf{s})$, $\alpha_t \sim ?$ Additive is *better* than multiplicative.
  - $w_t(\mathbf{s})$ independent for each t
  - $w_t(\mathbf{s}) = w_{t-1}(\mathbf{s}) + \eta_t(\mathbf{s})$, independent spatial process innovations - random walk (perhaps autoregressive)
  - Inference? Interpolate for new locations at an observed time; predict for a current location at a new time; new location and new time

Multivariate spatio-temporal models

- Again, connecting space and time, extending versions above
- $\mathbf{Y}(s,t) = \boldsymbol{\mu}(s,t) + \mathbf{w}(s,t) + \epsilon(s,t)$ with $\boldsymbol{\mu}(s,t) = \mathbf{X}(s,t)^T \boldsymbol{\beta}$
- For $\mathbf{w}(s,t)$, can supply an additive form in random effects
- Model for $\boldsymbol{\alpha}_t$? for $\mathbf{w}(s)$?
- Perhaps, most convenient is a coregionalization specification

cont.

- Coregionalization approach
- In the LMC, consider $\mathbf{w}(s,t) = A\mathbf{v}(s,t)$ where the components of $\mathbf{v}(\mathbf{s},t)$ are independent univariate space-time processes
- Can extend to A_t, to $A(s)$ or even $A(s,t)$
- Can make the evolution dynamic, e.g., $\mathbf{w}_t(s) = A\mathbf{v}_t(s)$ where $v_{lt}(s) = \phi_l v_{l,t-1}(s) + \epsilon_{lt}(s)$
- Further variations
- Could also try cross-convolution using valid $C_l(s,t)$

Dynamic Space Time models

- Again, t is discrete. More general two-stage specification. Dynamics in the mean
- Stage 1: Measurement equation

$$Y(s,t) = \mu(s,t) + \epsilon(s,t); \;\; \epsilon(s,t) \overset{ind}{\sim} N\left(0, \sigma_\epsilon^2\right).$$
$$\mu(s,t) = \mathbf{x}^T(s,t)\tilde{\boldsymbol{\beta}}(s,t).$$
$$\tilde{\boldsymbol{\beta}}(s,t) = \boldsymbol{\beta}_t + \boldsymbol{\beta}(s,t)$$

- Stage 2: Transition equation

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \overset{ind}{\sim} N_p\left(\mathbf{0}, \Sigma_{\boldsymbol{\eta}}\right).$$
$$\boldsymbol{\beta}(s,t) = \boldsymbol{\beta}(s,t-1) + \mathbf{w}(s,t).$$

- $\mathbf{w}(s,t)$ is a multivariate space-time process

cont.

- So, as above $\mathbf{w}(s,t) = A\mathbf{v}(s,t)$, with $\mathbf{v}(s,t) = (v_1(s,t), ..., v_p(s,t))^T$.
- The $v_l(s,t)$ are replications of a Gaussian processes with unit variance and correlation function $\rho_l(\phi_l, \cdot)$
- Can connect to linear Kalman filter

Dynamic Space Time models

- More general two-stage specification.
- Stage 1: Measurement equation

$$Y_t(s) = f(Z_t(s); \boldsymbol{\theta}) + \epsilon_t(s)$$

- $\epsilon_t(s)$ is a time-dependent white noise process
- Stage 2: Transition equation

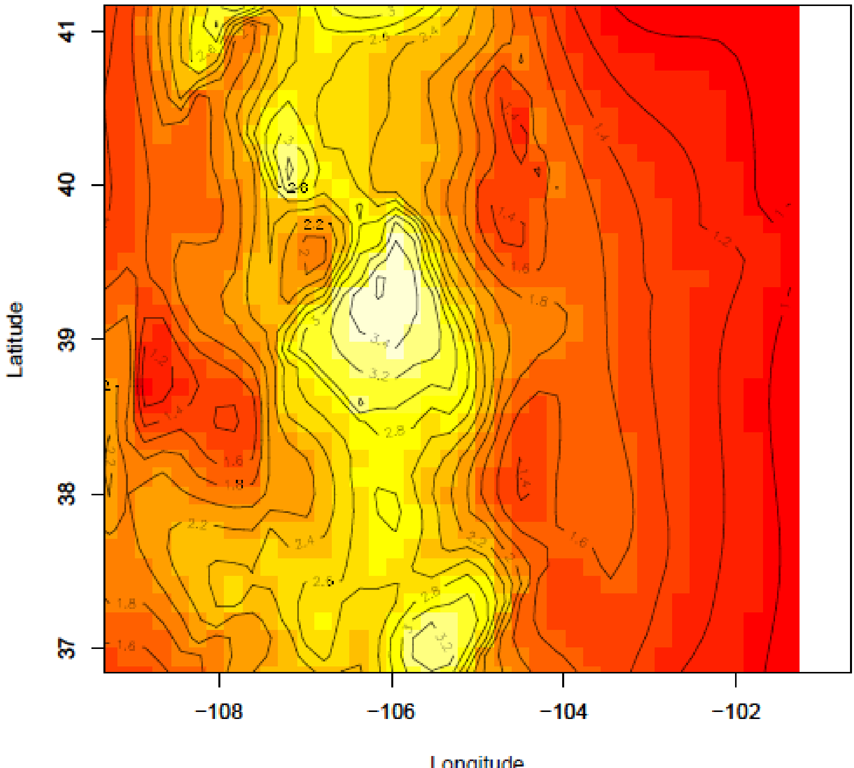
$$Z_t(s) = g(Z_{t-1}(s); \boldsymbol{\gamma}) + \eta_t(s)$$

- $\eta_t(s)$ is a discrete-time spatial process

An example:

- Modelling Temperature given precipitation.
- Sampled temperature data (maximum monthly temperature) and sampled precipitation data (maximum monthly precipitation), 50 sites across Colorado, January through December in 1997 (12 months).
- $$\text{temp}(s, t) = \tilde{\beta}_0(s, t) + \tilde{\beta}_1(s, t)\text{precip}(s, t) + \epsilon(s, t)$$
- Independent Gaussians for $\boldsymbol{\beta}_t$
- Space-time varying intercept process and slope process. Coregionalization for $\mathbf{w}(s, t)$

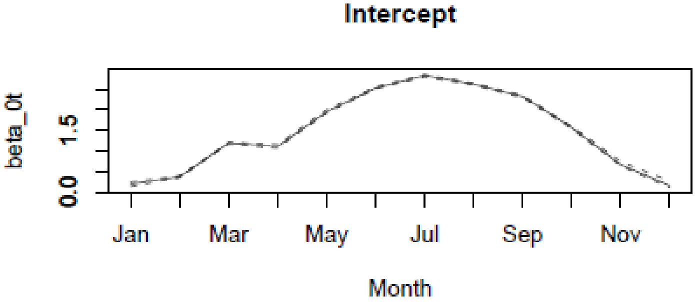
Spatial domain (with elevation)

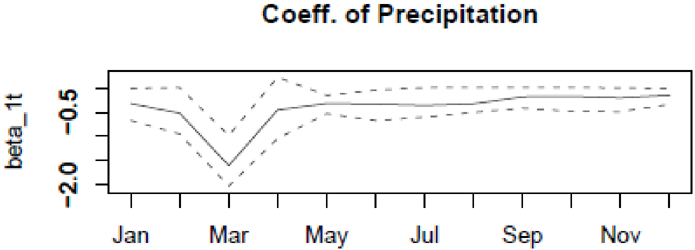


Parameter estimates

Parameters	median (2.5%, 97.5%)
$\Sigma_\eta[1,1]$	0.296 (0.130, 0.621)
$\Sigma_\eta[2,2]$	0.786 (0.198, 1.952)
$\Sigma_\eta[1,2]\sqrt{\Sigma_\eta[1,1]\Sigma_\eta[2,2]}$	-0.562 (-0.807, -0.137)
$\Sigma_{\mathbf{w}}[1,1]$	0.017 (0.016, 0.019)
$\Sigma_{\mathbf{w}}[2,2]$	0.026 (0.0065, 0.108)
$\Sigma_{\mathbf{w}}[1,2]\sqrt{\Sigma_{\mathbf{w}}[1,1]\Sigma_{\mathbf{w}}[2,2]}$	-0.704 (-0.843, -0.545)
σ_ϵ^2	0.134 (0.106, 0.185)
ϕ_1	1.09 (0.58, 2.04)
ϕ_2	0.58 (0.37, 1.97)
Range for intercept process	2.75 (1.47, 5.17)
Range for slope process	4.68 (1.60, 6.21)

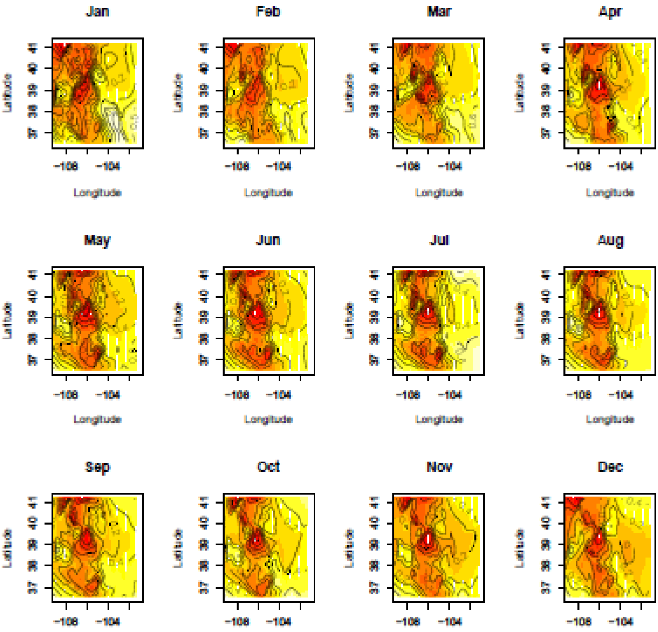
Time varying parameters





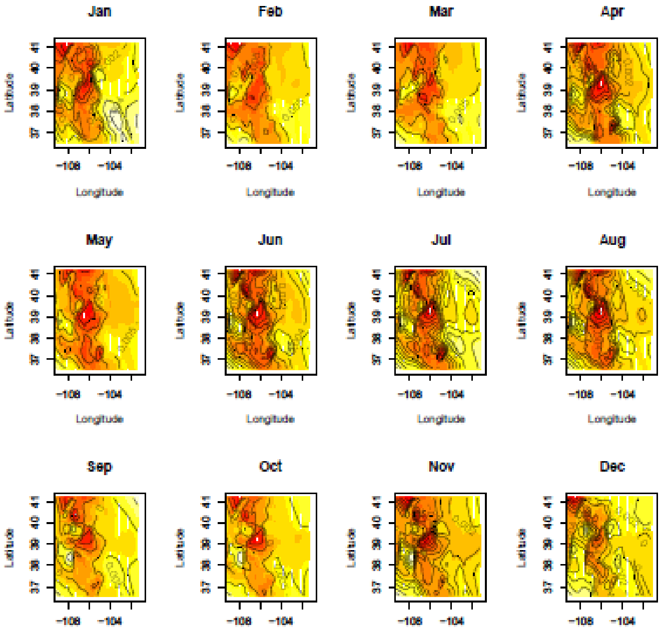
Intercept process

Time-sliced image-contour plots displaying the posterior mean surface of the spatial residuals



Slope process

Time-sliced image-contour plots displaying the posterior mean surface of the spatial residuals



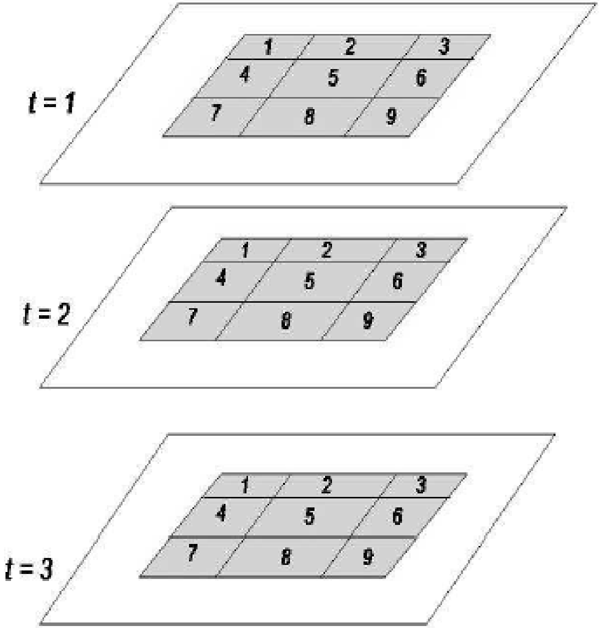
Areal unit data

- $Y_i(t)$, temporal process for each unit (rare, but, e.g., continuous time functional level across regions of the brain) Y_{it}, a time series for each unit (and occasionally, Y_{ijt}), is more common
- Again, $Y_{it} = \mu_{it} + \phi_{it} + \epsilon_{it}$,
  or for non-Gaussian data, $g(E(Y_{it})) = \mu_{it} + \phi_{it}$
- Again, $\epsilon_{it} \sim N(0, \tau_t^2)$
- Modeling for ϕ_{it}?? CAR in space and time!
  - Again, additive form, $\phi_i + \alpha_t$
  - For space nested within time, model $\phi_i^{(t)} \sim CAR(\lambda_t)$, with say $\lambda_t \overset{iid}{\sim} Gamma(a, b)$
  - $\phi_{it} | \boldsymbol{\phi}_{-(it)}$, space, time neighbors, weight for space, weight for time
  - MCAR, $\boldsymbol{\phi}_i = (\phi_{i1}, \phi_{i2}, ...\phi_{iT})$, if short series

A bit richer

- Suppose the “disease mapping” setting over space and time
- Data is a count Y_{ilt} for ith unit, lth subgroup, tth year
- $Y_{ilt} \sim Po(E_{ilt} r_{ilt})$. (In fact, n_{ilt}, p_{ilt} specification is preferred as before)
- $\log r_{ilt} = \mathbf{X}_l^T \boldsymbol{\beta} + \delta_t + \mathbf{Z}_t^T \gamma + \theta_{it} + \phi_{it}$
- θ's are pure error, ϕ's are CARs
- Again, should we include θ's?
- Again, modeling options for ϕ_{it}

Neighbors in time and space



Space-time autoregressive models

- STAR models (popular in the econometrics literature)
- Suppose we have n areal units observed over T time points
- We can imagine an $n \times n$ spatial proximity matrix, W_i and a $T \times T$ temporal proximity matrix
- If we concatenate the space-time random effects by time, i.e., a $Tn \times 1$ vector, ϕ, we can write

$$\phi = (W_t \otimes I_n + I_T \otimes W_i + W_t \otimes W_i)\phi + \epsilon$$

- Space-time model with clear interpretation

Diffusions in time over space

- Many interesting ecological diffusions
- Emerging diseases - avian flu, H1N1 flu
- Exotic organisms - invasive plants, gypsy moths
- Size and age distributions
- Transformation of landscape, deforestation, land use classifications, urban growth

The objective

- Forecast likely spread in space and time with associated uncertainty
- Nonlinear, nonhomogeneous in space and time
- Explanatory covariates
- Start with deterministic integro-differential equations or with partial differential equations
- How to add uncertainty?

cont.

- Theoretical models, “varying” around them
- Analytical solutions only available for models too simple to be what we need
- So, discretization to fit models
- Do we care of the deterministic equation? Should we just work with the discrete time version we want? Dynamic spatial models?

Hierarchical modeling

- Again, the hierarchical paradigm:

$$[\mathit{data}|\mathit{process}, \mathit{parameters}][\mathit{process}|\mathit{parameters}][\mathit{parameters}]$$

- A paradigm shift - designed experiments to observational studies; controlled experiments to integrated (big picture) investigation
- Prior information from: empirical studies, mechanistic knowledge, ecological theory, etc.
- Multiple information sources
- Conditional uncertainty in components (a natural way to specify models)
- Different resolutions in space and time
- Structured dependence in space and time
- Complex dependence structure through latent variables

Eurasian collared dove data

- An example from Wikle et al., using data from the Breeding Bird Survey (BBS)
- Escaped to U.S. from Bahamas, introduced in Florida, expanding dramatically across North America
- 4000+ routes in the survey (some sampled more than once per year, others not sampled in a given year), length of route is $\approx$ 40 kms, 50 stops per route, count birds by sight for 3 minutes
- 18 years: 1986-2003
- Response at a point is a count
- Aggregate to grid boxes
- Z_{it} is count in box i in year t, n_{it} is number of visits to cell i in year t.
- λ_{it} is *intensity* for box i in year t

Modeling specifics

- $Z_{it} \sim \text{Po}(n_{it}\lambda_{it})$
- $\log\lambda_{it} = w_{it} + \epsilon_{it}$
- ϵ_{it} are i.i.d. (pure error or micro-scale variation)
- The focus is on the $\mathbf{w}_t$. They tell the diffusion story, i.e., $\mathbf{w}_t = H(\boldsymbol{\delta})\mathbf{w}_{t-1} + \boldsymbol{\eta}_t$
- So, first order Markov and H is the *propagator* matrix
- Model for $\boldsymbol{\eta}_t$?
- $\mathbf{w}_0 \sim N(0, 10I)$
- $\boldsymbol{\delta}$ is the vector of local diffusion coefficients, one for each grid cell
- A dimension reduction for $\boldsymbol{\delta}$; many possibilities here - basis functions, EOF's, predictive processes

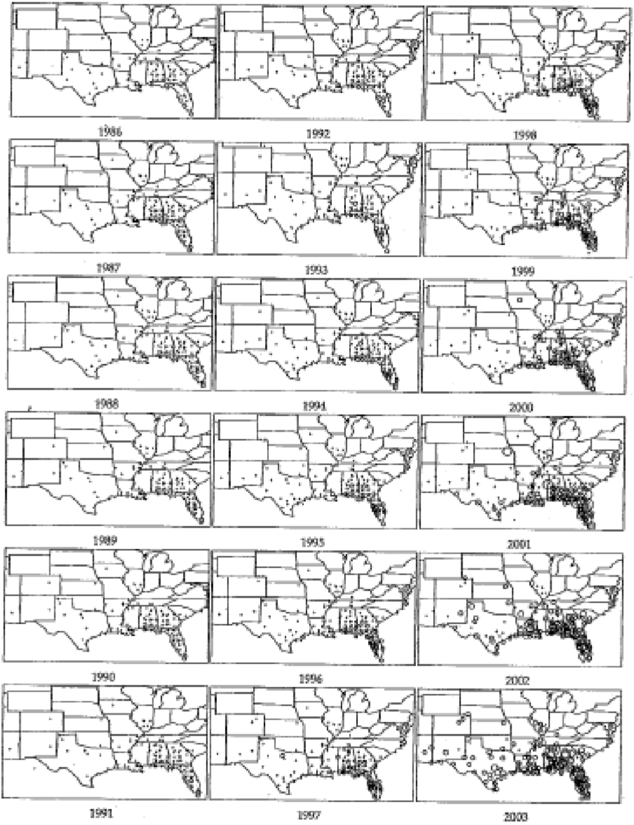


Figure 8.1. Location of BBS survey route (+) and observed Eurasian Collared-Dove count for years 1986–2003. The radius of the circles are proportional to the observed count.

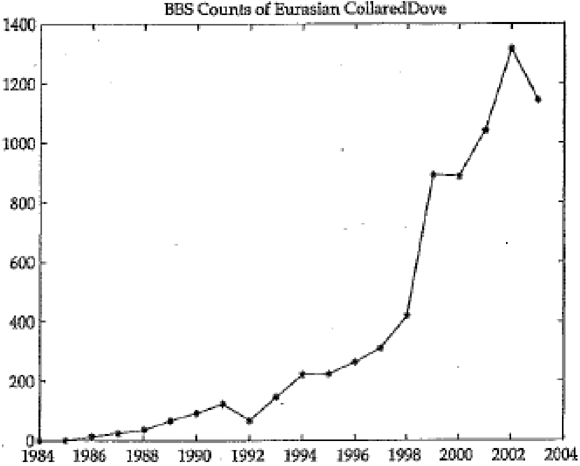


Figure 8.2. Sum of BBS Eurasian Collared-Dove counts over space for years 1986–2003.

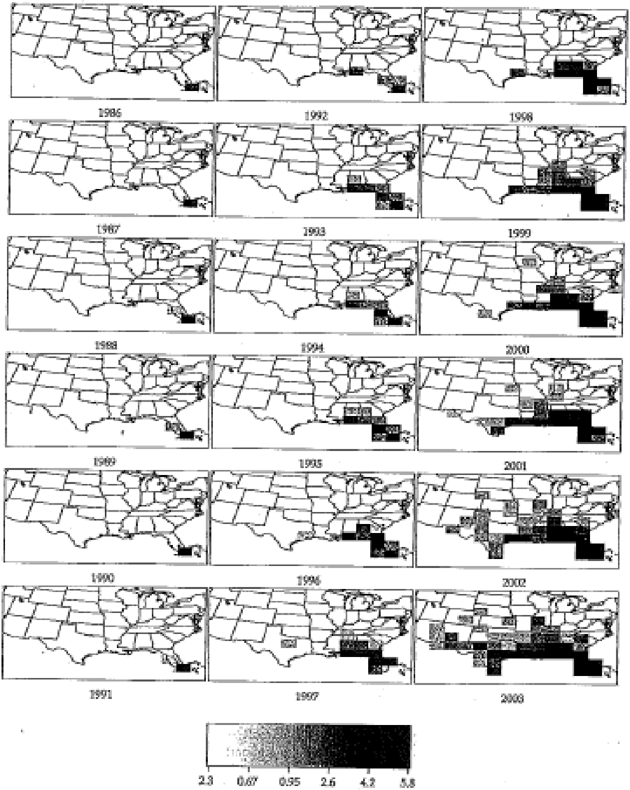


Figure 8.3. Log of Eurasian Collared-Dove BBS counts aggregated to a grid for years 1986–2003.

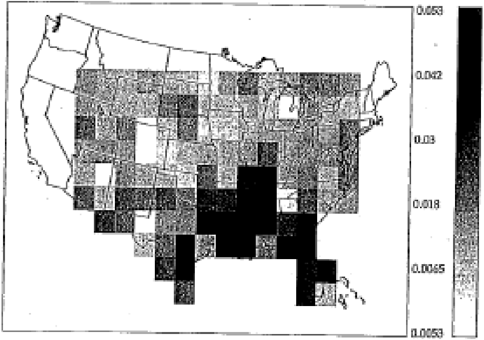


Figure 8.7. Posterior mean of δ, the diffusion coefficients.

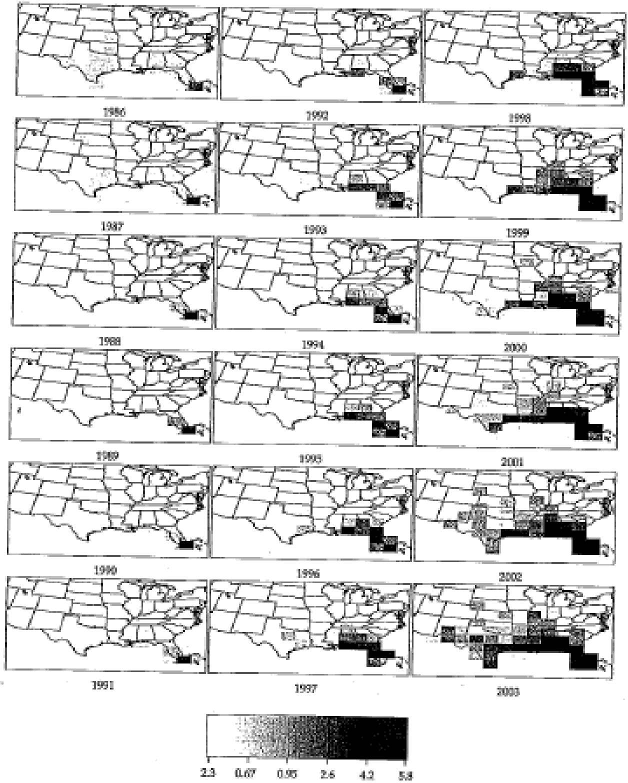


Figure 8.9. Posterior mean of $\log(M(s;t)/A(s;t))$ for years 1986–2003.

Dynamics

- With continuous space, discrete time, i.e., $w_t(s)$
- Without loss of generality $\mathbf{t} = (1, 2, ..., T)$
- Envision $w_t(s)$ as a *dynamical* process
- Again, simplify to first order Markov, i.e., for locations $s_1, s_2, ..., s_n$, let $\mathbf{w}_t = (w_t(s_1), w_t(s_2), ..., w_t(s_n))^T$. Then

$$[\mathbf{w}_t | \mathbf{w}_0, \mathbf{w}_1, ...\mathbf{w}_{t-1}] = [\mathbf{w}_t | \mathbf{w}_{t-1}]$$

- For example, $\mathbf{w}_t = H\mathbf{w}_{t-1} + \boldsymbol{\eta}_t$ where $\eta_t(s)$ incorporates spatial structure
- A vector AR(1) model and, again, H is called the *propagator* matrix
- ??Specifying H??

Specifying H

- $H = I$ - not stationary (explosive), no interaction across space and time, not realistic for most dynamic processes of interest
- $H = \text{Diag}(h)$ where $\text{Diag}(h)$ has diagonal elements $0 < h_i < 1$ - Still no space-time interactions
- Integro-difference equation (IDE) dynamics:

$$w_t(s) = \int h(s, r; \phi) w_{t-1}(r) dr + \eta_t(s)$$

- h is a "redistribution kernel" that determines the rate of diffusion and the advection
- h stationary, i.e., $h(s - r; \phi)$? h time dependent, i.e., $h_t(s, r; \phi)$? h spatially dependent, i.e., $h(s, r : \phi(r))$?

cont

- If require $w > 0$, work with

$$\log w_t(s) = \log\left(\int h(s, r; \phi) w_{t-1}(r) dr\right) + \eta_t(s)$$

- Alternatively,

$$v_t(s) = \int h(s, r; \phi) v_{t-1}(r) dr$$

and

$$\log w_t(s) = \log v_t(s) + \eta_t(s)$$

- Discretization to obtain H

cont

- Recall linear PDE, $\frac{dw(s,t)}{dt} = h(s)w(s,t)$
- Finite differencing yields $w(s, t + \Delta t) - w(s, t) = h(s)w(s, t)\Delta t$, i.e., $w(s, t + 1) \approx \tilde{h}(s)w(s, t)$. Same limitations as above.
- Need more general PDE's
- PDE can motivate IDE, can clarify H
- "forward" vs. "backward" perspective
- IDE's can be specified directly without using PDE's, e.g., $h(s, r)$ can be a sum of a survival/spread term + a birth/replenishment term

Diffusion PDE's

- Diffusion in one dimension - Fick's Law: diffusive flux from *high* concentration to *low* is $-\delta \frac{\partial w(x,t)}{\partial x}$ with δ, the diffusion coefficient. Location varying diffusion $\delta(x)$
- And, diffusion equation is $\partial w / \partial t = -\partial \text{flux}/\partial x$, i.e., $\frac{\partial w(x,t)}{\partial t} = \frac{\partial}{\partial x}(\delta(x)\frac{\partial w(x,t)}{\partial x})$
- That is, the 1-dim diffusion equation is

$$\frac{\partial w(x,t)}{\partial t} = \delta^{'}(x)\frac{\partial w(x,t)}{\partial x} + \delta(x)\frac{\partial^2 w(x,t)}{\partial x^2}$$

- In 2-dim, diffusive flux is $-\delta(x,y)\nabla w(x,y,t)$ ($\nabla w(x,y,t)$ is the concentration gradient at time t)
- The resulting diffusion PDE is

$$\frac{\partial w(x,y,t)}{\partial t} = \frac{\partial}{\partial x}(\delta(x,y)\frac{\partial w(x,y,t)}{\partial x}) + \frac{\partial}{\partial y}(\delta(x,y)\frac{\partial w(x,y,t)}{\partial y})$$

Discretizing the diffusion equation

- Complete chain rule differentiation of the diffusion equation
- Yields second order partial derivatives, $\frac{\partial^2 w}{\partial x^2}$ and $\frac{\partial^2 w}{\partial y^2}$
- Introduce Δt, Δx, Δy
- Replace ∂'s with finite differences (first forward and second order centered) - careful detail, ugly expression!
- After the smoke clears, we obtain $\mathbf{w}_{t+\Delta t} = H\mathbf{w}_t$
- Again, add $\boldsymbol{\eta}_t$
- We are back to our earlier redistribution form

Add growth rate

- Previous dynamics simply redistribute existing population spatially over time
- In many situations, there is also population growth/decline
- Options here - scale $h_t(s, r : \phi)$ in the IDE; overall temporal trend added to PDE; an exponential differential equation
- As an example, suppose population growth is captured by a logistic differential equation

$$\frac{\partial w(s,t)}{\partial t} = \gamma w(s,t)(1 - w(s,t)/K)$$

- γ is the growth rate, K is the carrying capacity
- $\gamma(s)$?, $\gamma(t)$, $\gamma(s,t)$ and even $K(s)$?

Enriching the modeling

- Again, we focus on $w(s,t)$
- $w(s,t)$ can arise as a mean model for a geostatistical model or in a space-time GLM (as in Wikle) or as a *cumulative* intensity $\Lambda(s,t)$ for a space-time point pattern (which drives the cumulative diffusion)
- A general diffusion PDE (nonstochastic) looks like $\frac{\partial w(s,t)}{\partial t} = a(w(s,t), z(s,t), \theta)$ where $z(s,t)$ are other potential variables ($z(s,t) = t$ for example)
- Discretization as proposed above
- $\theta(s)$?, $\theta(s,t)$?
- **How to make the PDE stochastic?**

First a differential equation

- Ignoring location s for the moment, we have:

$$dw(t) = a(w(t), t, \theta)dt \quad \text{with} \quad w(0) = w_0$$

  Simplest way to add stochasticity is to make θ random.

- Instead:

$$dw(t) = a(w(t), t, \theta)dt + b(w(t), t, \theta)dZ(t)$$

  where $Z(t)$ is Brownian motion over R^1 with a and b the "drift" and "volatility" respectively. Now a *stochastic* differential equation (SDE)

- θ would still be random
- Richer: $dw(t) = a(w(t), t, \theta(t))dt$ where $d\theta(t) = g(\theta(t), t, \beta)dt + h(\theta(t), \sigma)dZ(t)$ and $Z(t)$ is variance 1 Brownian motion.

Add space

- Now, we add space. First,

$$dw(s,t) = a(w(s,t), t, \theta(s))dt \ \text{ with } \ w(s,0)) = w_0(s),$$

  a PDE. Randomness through $\theta(s)$, a process realization, so $\theta(s)$ provides the spatial dependence. Hence, a stochastic process of differential equations.

- Richer,

$$dw(s,t) = a(w(s,t), t, \theta(s))dt + b(w(s,t), t, \theta(s))dZ(s,t)$$

- Modeling $Z(s,t)$? For a fixed finite set of spatial locations assume independent Brownian motion at each location.
- Or a discrete space approximation to spatial Brownian motion employing a Gaussian process (GP)on R^2