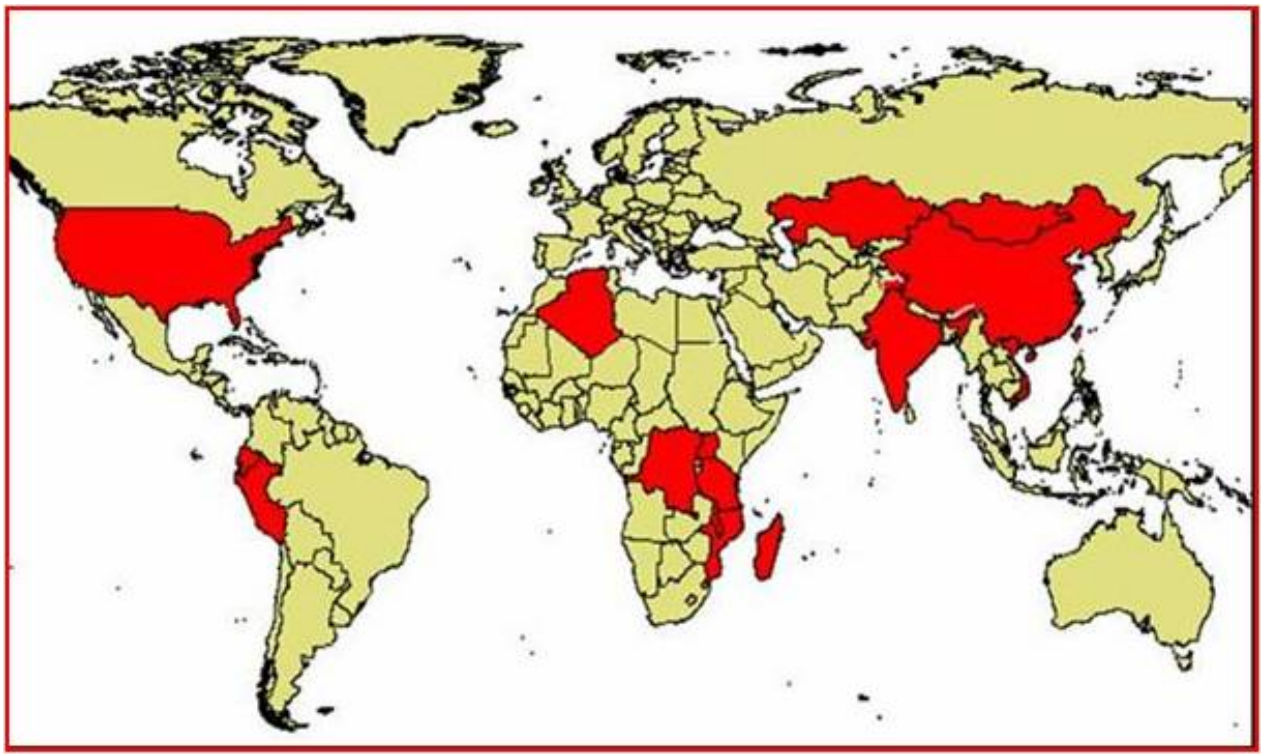
Bayesian analysis of plague in Kazakhstan

Geir Storvik
Dept. of Mathematics, University of Oslo
SFI^2, Statistics for inovation
Norwegian Computing Center
SFF-CEES, Ecology and Evolution synthesis

Smögen, August 16-20, 2010
Lise Heier, Hildegunn Viljugrein, Nils Chr Stenseth, Bård Ø. Kvaal

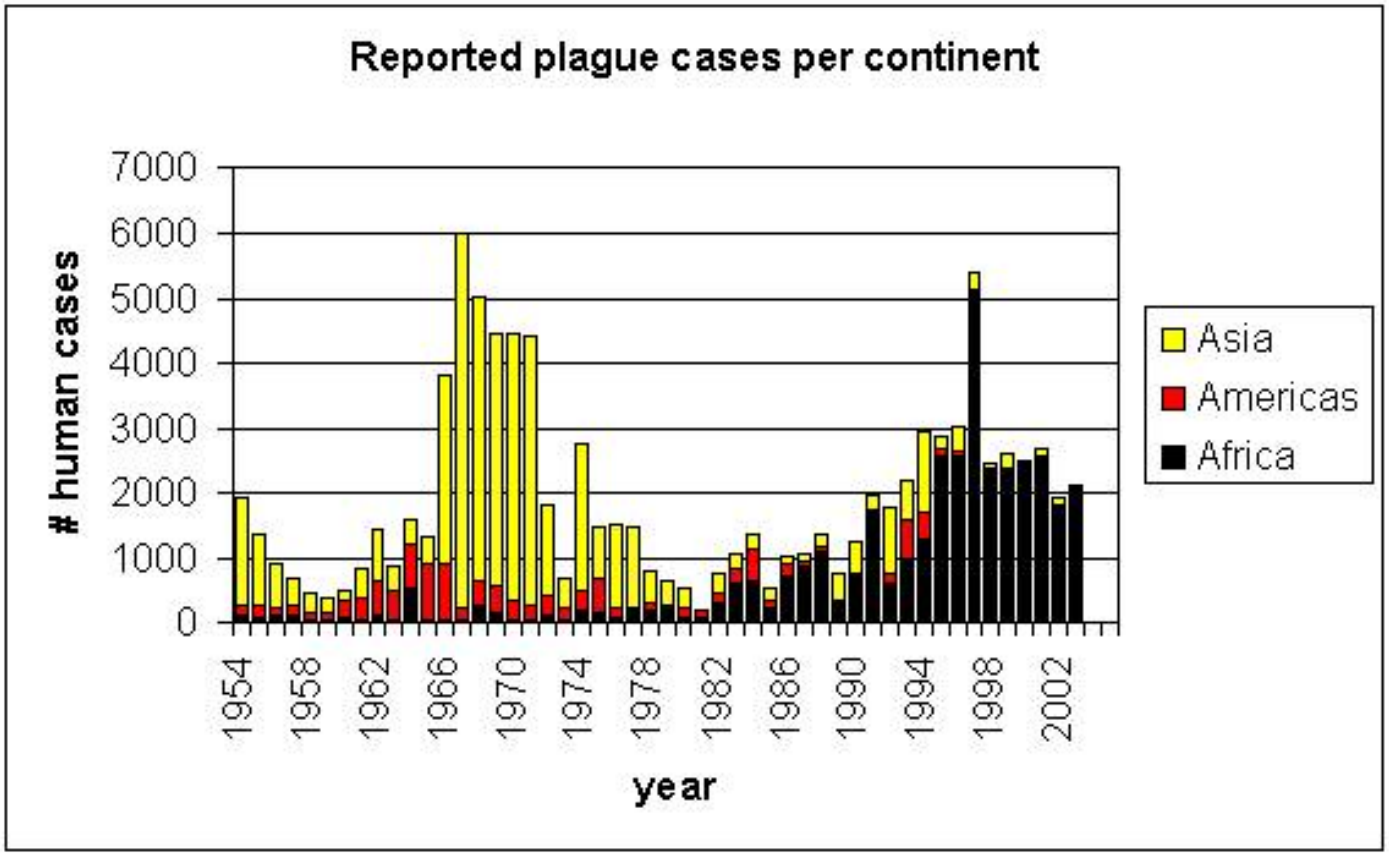
Outline

Human cases of Plague



Stenseth and Kausrud [2009]

Cases per continent



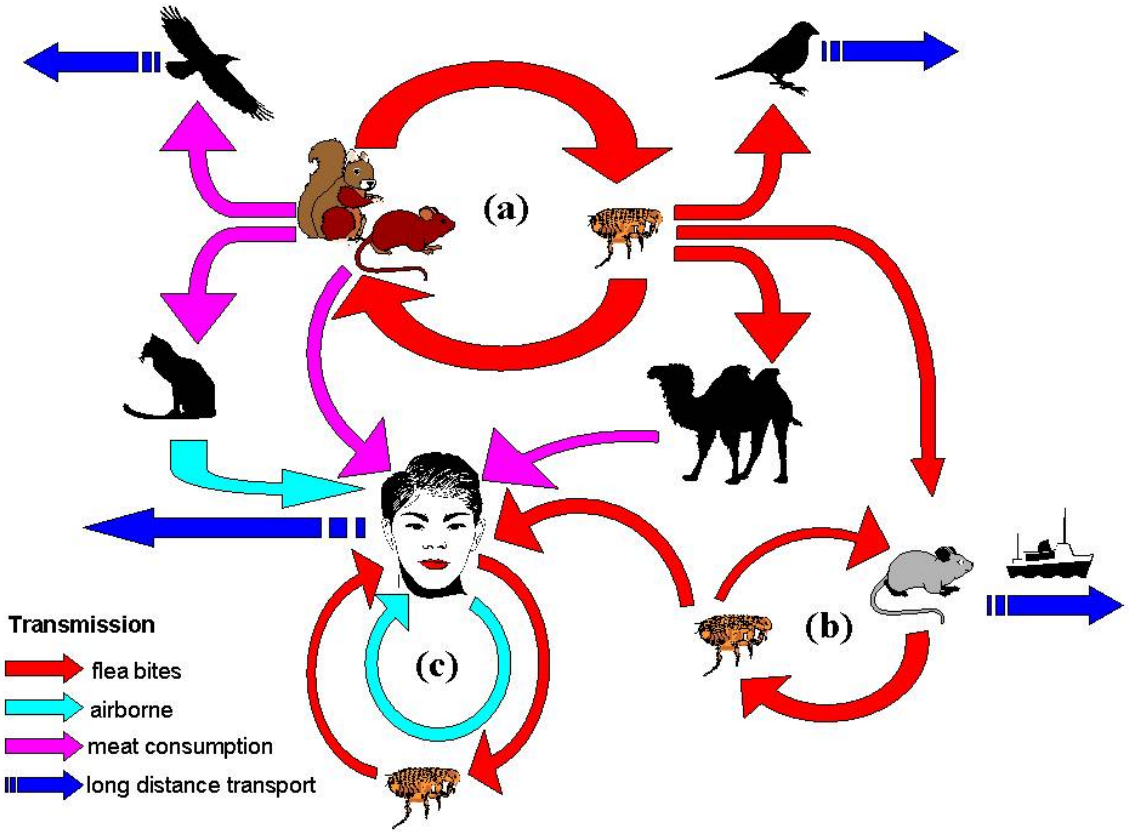
Stenseth and Kausrud [2009]

Plague is a highly variable disease

- Given rise to at least three pandemics, with smaller endemics in between.

- A variety of host species: Great Gerbils
- A variety of vector species (transmitting the bacteria (Yersinia pestis))
- Complex temporal patterns
- Complex spatial patterns

Plague system



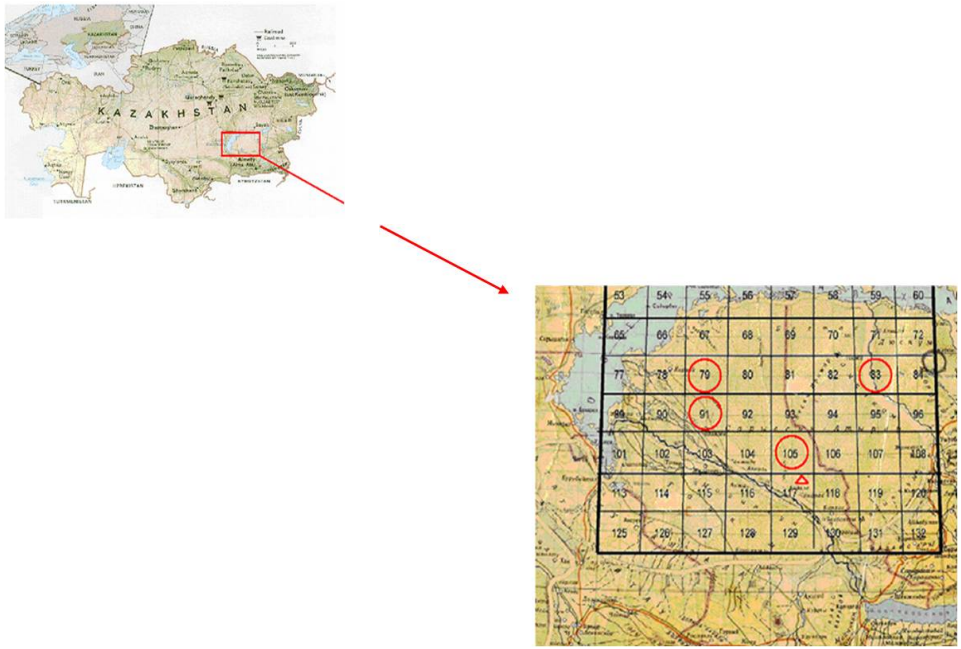
Stenseth and Kausrud [2009]

Questions of interest

- Biology: Understand the dynamics of plague
  - Why do plague “disappear” in periods?
  - How do population sizes influence occurence of plague
  - How do climatic variables influence plague and/or population sizes?
  - Are there spatial structure in where plague appear?
- Here: Mainly modelling absence/presence of plague

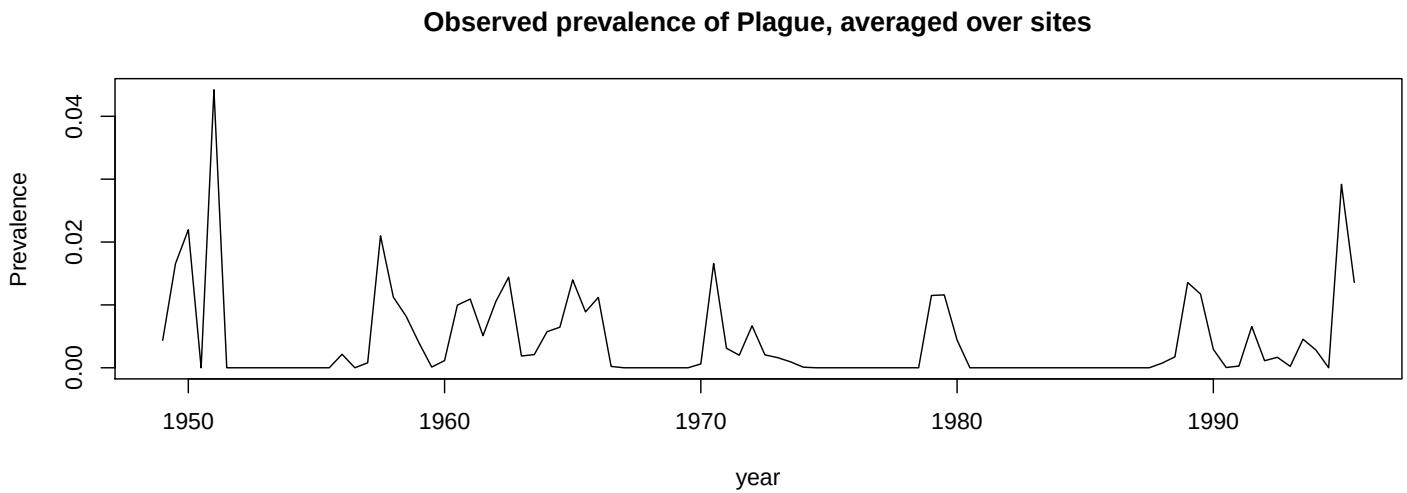
Kazakhstan data

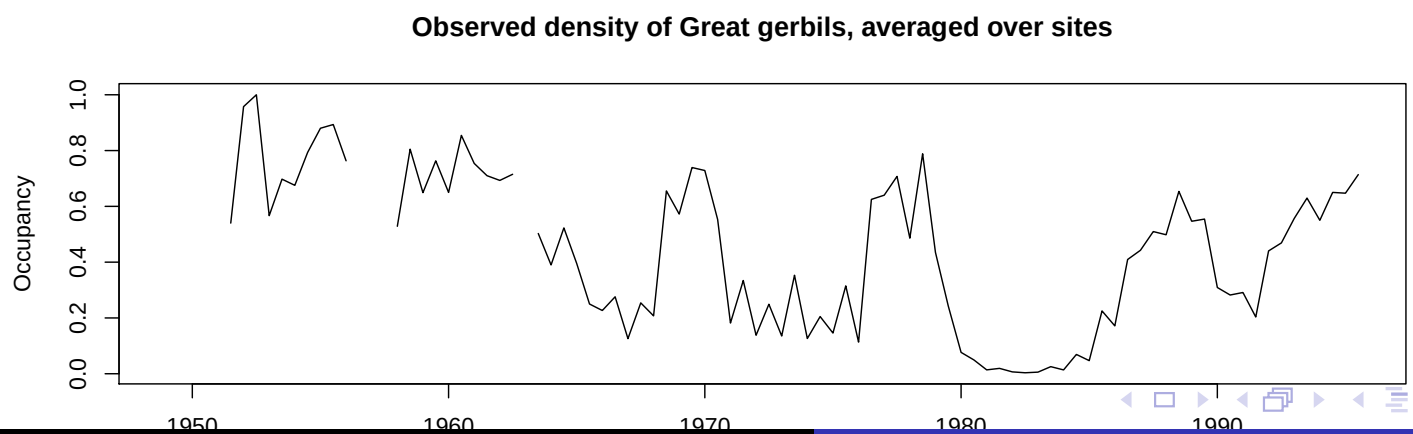
- Data collected within sectors (squares), spring and autumn



- Gerbils:
  - Observed ratio of occupied burrows, $R_{t,i}$ ("density").
  - Number of tested $N_{t,i}$ and positive $Y_{t,i}$ animals
- Similar data for fleas.
- Much missing data

Data aggregated over space





Previous analyses

- Time series approach, Spatially aggregated/single sites [Davis et al., 2004, Park et al., 2006a,b]
- Bayesian analysis, only Gerbils [Frigessi et al., 2005]
- GLMM approach with spatial random effects [Heier, L. and Storvik, G. and Davis, S. A. and Viljugrein, H. and Ageyev, V, A and Klassovskaya, E. and Stenseth NC, 2010]
  - Can use standard software
  - Many missing variables of plague or gerbil density (multiple imputation possible)
  - Difficult to take into account measurement errors (both in responses and covariates)
- Main interest been in presence/absence of plague

Overall model structure

- Process models
  - Model for plague given host density
    - So far only considering presence/absence of plague
  - Model for host “density”
    - Density actually fraction of occupied collonies
    - Model needed because of missing observations
- Measurement models
  - Plague data
  - Density data
- Bayesian approach
  - Prior models

Model for plague

- $z_{t,i} = 1$ if plague in site i at time t
- Logistic type models with random effects for all cases.

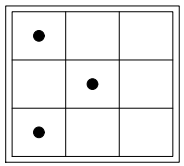
$$\text{logit}(\Pr(z_{t,i} = 1|\text{past})) = \alpha_{t,i}(\text{past}) + \varepsilon_{t,i}$$

- $\{z_{t,i}, i = 1, ..., N\}$ independent given $\{\alpha_{t,i}\}$ and $\{\varepsilon_{t,i}\}$.
- Spatial correlation through spatial correlation in $\{\varepsilon_{t,i}\}$.
- Temporal correlation through $\alpha_{t,i}(\text{past})$.
- $\varepsilon_t \sim GCAR(\rho_\varepsilon, \tau_\varepsilon)$, independent in time.

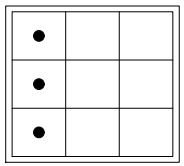
Model for α's

- Three "cases" of plague, depending on past:

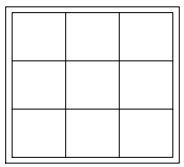
Persistence: $z_{t-1,i} = 1$



Spread: $z_{t-1,i} = 0, \sum_{j \sim i} z_{t-1,j} > 0$



Invasion: $z_{t-1,i} = 0, \sum_{j \sim i} z_{t-1,j} = 0$



Models for α's (cont)

- $\phi_{t,i}$ is occupancy of gerbils at logit-scale.
- Persistence

$$\alpha_{t,i} = b_{p,P}\phi_{t,i} + c_{p,P}\sum_{j\in\delta_i} z_{t-1,j} + \mathbf{x}_{t,i}^T\boldsymbol{\alpha}_P$$

- Spread

$$\alpha_{t,i} = b_{p,S}\phi_{t,i} + c_{p,S}\sum_{j\in\delta_i} z_{t-1,j} + \mathbf{x}_{t,i}^T\boldsymbol{\alpha}_S$$

- Invasion

$$\alpha_{t,i} = b_{p,I}\phi_{t,i} + \mathbf{x}_{t,i}^T\boldsymbol{\alpha}_I + s_{I,i}$$

- $\{s_{I,i}\}$ spatial process $\sim GCAR(\rho_S, \tau_S)$.

Measurement model

- Observed $(N_{t,i}, Y_{t,i})$.
- Assume no error in test
- Assume $Y_{i,t} \sim \text{Binom}(N_{t,i}, p_{t,i})$, $p_{t,i}$ prevalence of plague
- $Z_{t,i} = 1$ iff $p_{t,i} > 0$
- If $Y_{t,i} > 0$, then $Z_{t,i} = 1$
- If $Y_{t,i} = 0$, then $Z_{t,i} =$?
- Assume $p_{t,i} \sim \text{Beta}(\alpha, \beta)$ if $Z_{t,i} = 1$.
- Gives

$$\Pr(Y_{t,i} = 0 | Z_{t,i} = 0) = 1$$

$$\Pr(Y_{t,i} = 0 | Z_{t,i} = 1) = \frac{\Gamma(\alpha + \beta)\Gamma(n + \beta)}{\Gamma(\beta)\Gamma(n + \alpha + \beta)}$$

Model for occupancy

$$\phi_{t,i} = \mathbf{x}_{t,i}^T \boldsymbol{\beta} + \rho_t(\phi_{t-1,i} - \mathbf{x}_{t-1,i}^T \boldsymbol{\beta}) + u_{t,i}$$

- $\mathbf{u}_t \sim GCAR(\rho_\phi, \tau_\phi)$, independent in time.

Spatial models

GCAR(ρ, τ) Gaussian conditional autoregressive model

$$\begin{aligned}\mathbf{U} \sim & \text{MVN}(\mathbf{0}, (\tau \mathbf{Q}(\rho))^{-1}) \\ \mathbf{Q}(\rho) = & \rho \mathbf{P} + (1-\rho)\mathbf{I}\end{aligned}$$

$$P_{ij} = \begin{cases} m_i & j = i \\ -1 & j \in \delta_i \\ 0 & \text{otherwise} \end{cases}$$

MacNab [2003]: avoiding potential identifiability problems.
$\rho = 1$ corresponds to the intrinsic autregression model
$\rho = 0$ corresponds to an independence model.
Note: $\mathbf{Q}$ is sparse!

"Direct" approach

- θ_a parameters for abundance model
- θ_p parameters for plague model

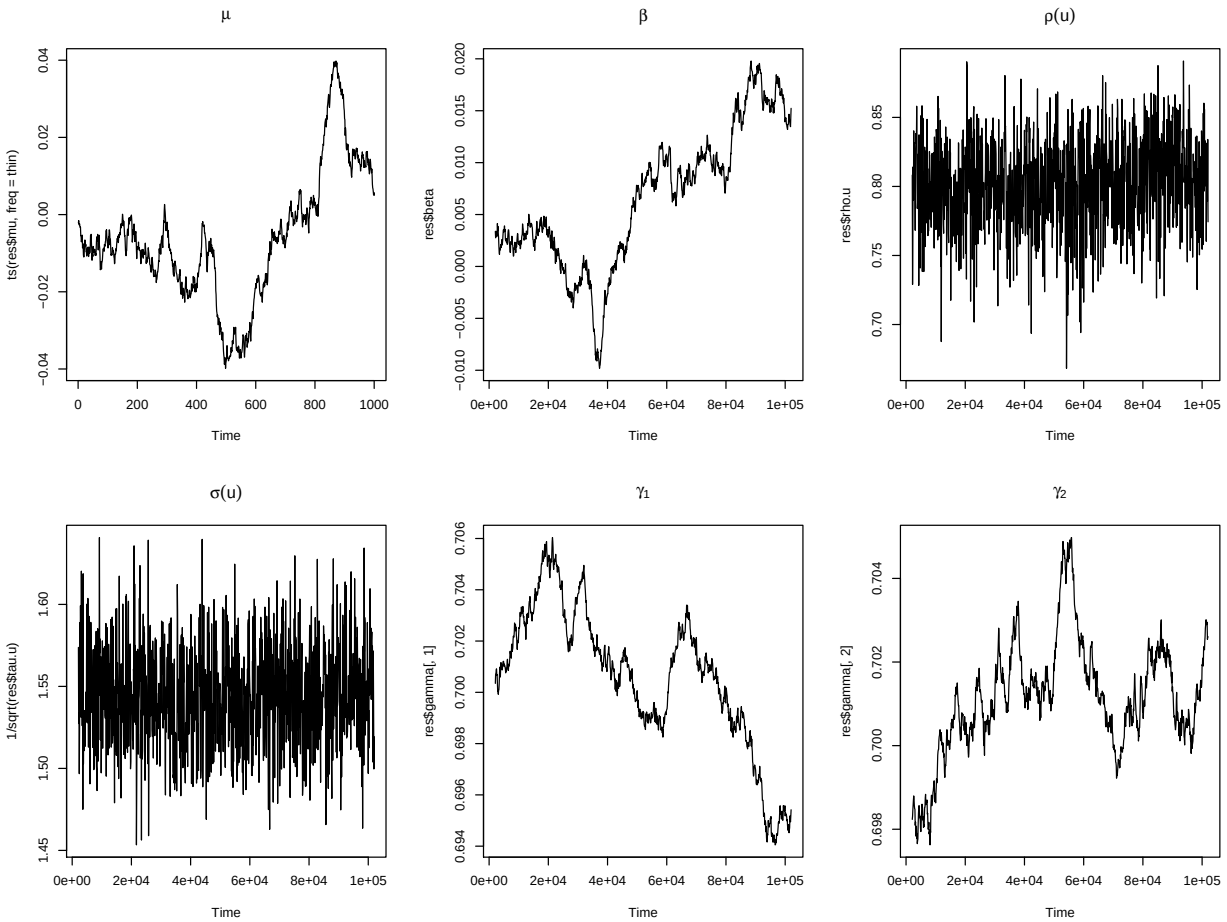
$$
\begin{aligned}
f(\phi, \mathbf{Z}|\theta_a, \theta_p) &= \prod_t f(\phi_t|\phi_{t-1}; \theta_a) f(\mathbf{Z}_t|\mathbf{Z}_{t-1}, \phi_t; \theta_p) \\
&= \prod_t f(\phi_t|\phi_{t-1}; \theta_a) \\
&\quad \prod_t f(\mathbf{Z}_t|\mathbf{Z}_{t-1}, \phi_t; \theta_p)
\end{aligned}
$$

- Each part GLMM type
- Missing/incomplete data: Multiple imputation
- Works, but complicated, not utilizing data fully

Bayesian analysis

- Denote $\boldsymbol{\theta}$ hyperparameters involved
  - Regression parameters
  - Precisions
  - Correlation parameters
- Prior $\pi(\boldsymbol{\theta})$ mainly noninformative
- Bayesian inference: $\pi(\boldsymbol{\theta}, \phi, \mathbf{z}|\mathbf{y})$
- Computation through Markov Chain Monte Carlo
- Possible to use Winbugs/Jags?

Using existing software - JAGS



Markov chain Monte Carlo

- Many variables to be updated
- Plague: $\{z_{t,i}\}$, $N \times T$, binary
- Random effects plague: $\{\varepsilon_{t,i}\}$, $N \times T$, Gaussian prior
- Invasion process: $\{s_{I,i}\}$, N, Gaussian prior
- Occupancy: $\{\phi_{t,i}\}$, $N \times T$, Gaussian prior
- Regression parameters, $b_{p,P/S/I}, c_{p,P/S}, \alpha_P, \alpha_S, \alpha_I, \beta$
- Precision parameters, $\tau_\varepsilon, \tau_S, \tau_\phi$
- Correlation parameters: $\rho_\varepsilon, \rho_S, \rho_\phi, \rho_t$
- Fast convergence: Update (large) blocks simultaneously (GMRFLib)

Block-updating continuous processes

- Knorr-Held and Rue [2002], Rue and Held [2005]
- Continuous processes of the form

$$\pi(\mathbf{x}|\theta,\mathbf{y}) \propto \exp\{-\tfrac{1}{2}(\mathbf{x}-\mu(\theta))^T\mathbf{Q}(\theta)(\mathbf{x}-\mu(\theta))+ \sum_i d_i \log g(x_i;y_i,\theta)\}$$

- Regression parameters included in $\mathbf{x}$.
- Metropolis Hastings: $\theta^* \sim q_1(\cdot|\theta)$, $\mathbf{x}^* \sim q_2(\cdot|\theta^*,\mathbf{x})$
- q_2 Gaussian approx. to $\pi(\mathbf{x}|\theta^*)$, independent of $\mathbf{x}$!
- Acceptance ratio

$$R = \frac{\pi(\theta^*)\pi(\mathbf{x}^*|\theta^*)f(\mathbf{y}|\mathbf{x}^*,\theta^*)}{\pi(\theta)\pi(\mathbf{x}|\theta^*)f(\mathbf{y}|\mathbf{x},\theta)} \times \frac{q_1(\theta|\theta^*)q_2(\mathbf{x}|\theta,\mathbf{x}^*)}{q_1(\theta^*|\theta)q_2(\mathbf{x}^*|\theta^*,\mathbf{x})}$$

Sparse matrix operations

- **Q** sparse, both in time and space.
- GMRFLib [Rue and Held, 2005], a C-library for automatic simulation from

$$\pi(\theta, \mathbf{x}|\mathbf{y}) \propto \pi(\theta) \exp\{-\tfrac{1}{2}(\mathbf{x} - \mu(\theta))^T \mathbf{Q}(\theta)(\mathbf{x} - \mu(\theta)) + \sum_i d_i \log g(x_i; y_i, \theta)\}$$

- Sparseness in time could have been handled through Kalman filter approximation.
- GMRFLib automatially handles spareness both in space and time.
- Example: Q 3480 $\times$ 3480, : Acceptance rate 0.29!

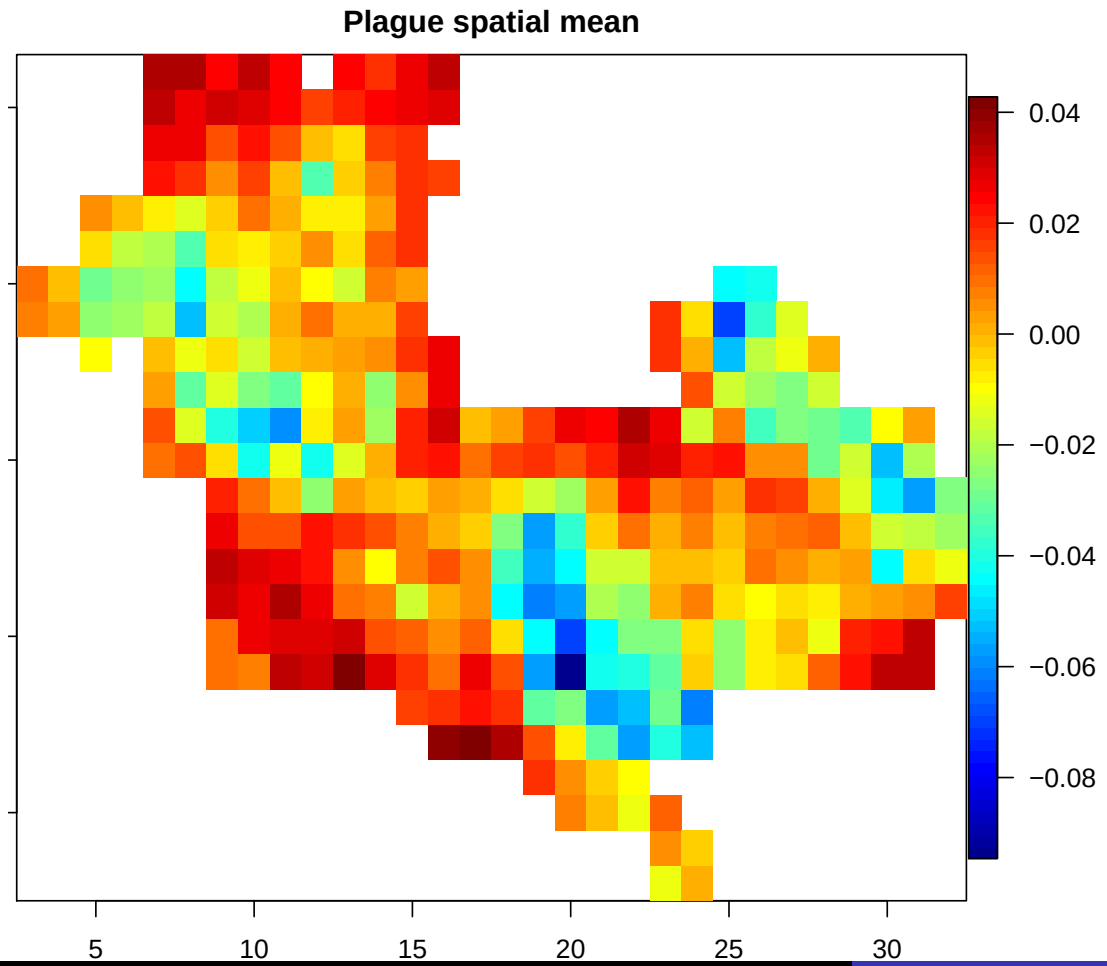
Preliminary results

- Years 1986-1995
- Only gerbils data
- Only data from fall
- Covariate: Absolute minimum of air temperature in one month.
- MCMC: 1000 burnin, 50000 additional iterations

Results

- Density highly significant for plague
- Past presence of plague slightly significant
- Clear spatial structure both in density and in plague
- Absolute minimum of air temperature in one month slightly significant on density but not on plague
- Spatial structure in invasion but significant?

Plague spatial mean (logit-scale)



Residuals

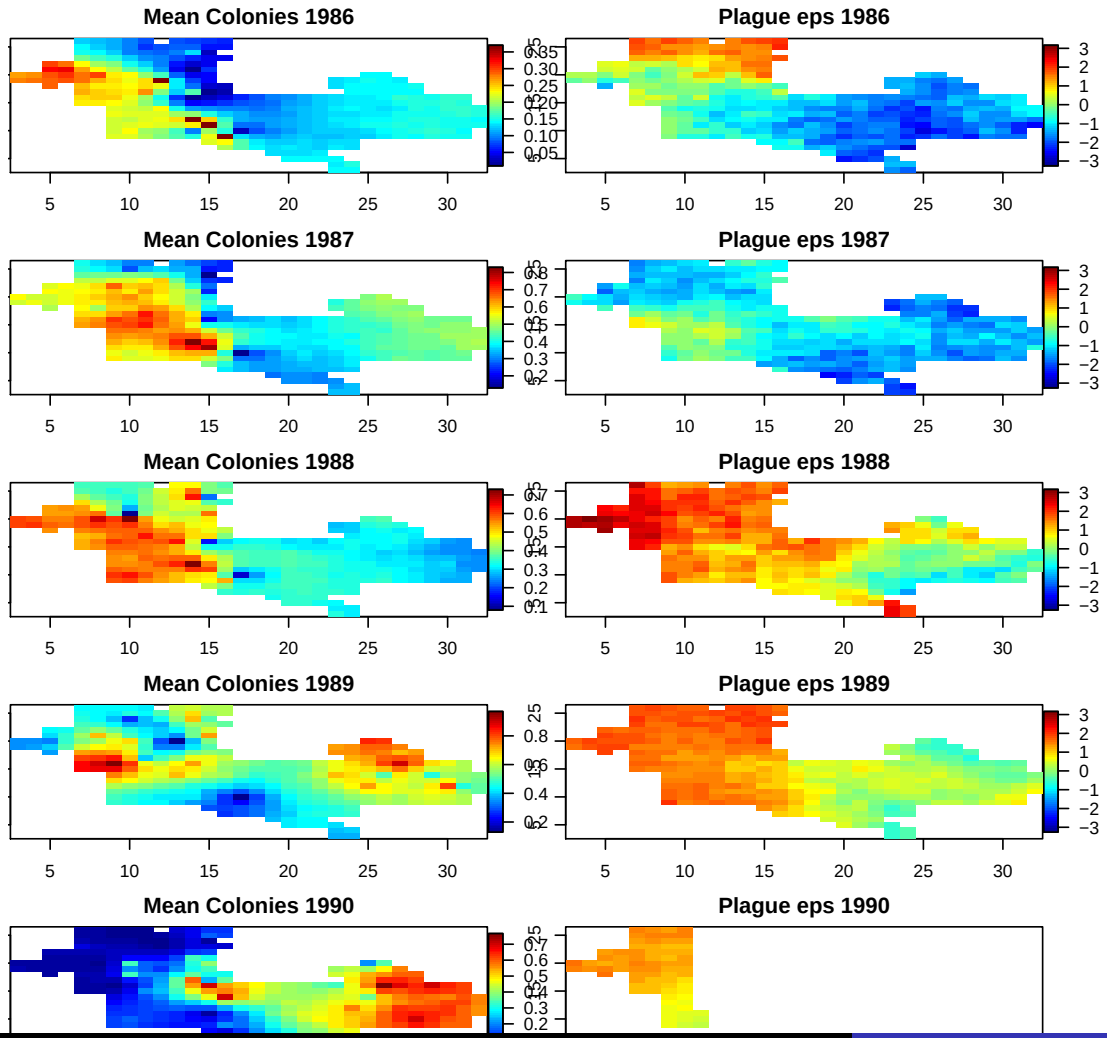
Model

$$\phi_{t,i} = \rho_t \phi_{t-1,i} + \mathbf{x}_{t,i}^T \boldsymbol{\beta} + u_{t,i}$$

$$\text{logit}(\Pr(z_{t,i} = 1 | \text{past})) = \alpha_{t,i}(\text{past}) + \varepsilon_{t,i}$$

Structure in $\{\varepsilon_{t,i}\}$, $\{u_{t,i}\}$?

Residuals


Mean Colonies 1986
Plague eps 1986
Mean Colonies 1987
Plague eps 1987
Mean Colonies 1988
Plague eps 1988
Mean Colonies 1989
Plague eps 1989
Mean Colonies 1990
Plague eps 1990

Extensions

- Better observation models
- Including fleas data
- Nonlinear relation to gerbil density
- Better proposals in MCMC
- Model critisism/evaluation!

Summary/Discussion

- Preliminary version of full Gerbil-plague analysis
- Seems to confirm previous analyses
- Efficient MCMC implementation
- Further improvements/extensions needed

N. C. Stenseth and K. L. Kausrud. Plague - historical perspectives to modern infection trends. Talk on Society for Applied Microbiology, Manchester, July 7 2009.

S. Davis, M. Begon, L. De Bruyn, V. S. Ageyev, N. L. Klassovskly, S. B. Poel, H. Viljugrein, N. C. Stenseth, and H. Leirs. Predictive Thresholds for Plague in Kazakhstan. *Science*, 304:736–738, 2004.

S. Park, K. Chan, H. Viljugrein, L. Nekrassova, B. Suleimenov, V. S. Ageyev, N. L. Klassovskiy, S. B. Pole, and B. C. Stenseth. Statistical analysis of the dynamics of antibody loss to a disease-causing agent: plague in natural populations of great gerbils as an example. *J. R. Soc. Interface*, ??:??–??, 2006a.

S. Park, K. Chan, H. Viljugrein, L. Nekrassova, B. Suleimenov, V. S. Ageyev, N. L. Klassovskiy, S. B. Pole, and B. C. Stenseth. . *J. R. Soc. Interface*, ??: ??–??, 2006b.

A Frigessi, M. Holden, C. Marshall, H. Viljugrein, N. C. Stenseth, L. Holden, V. S. Ageyev, and N. L. Klassovskly. Bayesian population dynamics of interacting species: Great gerbils and fleas in Kazakhstan. *Biometrics*, 61: 230–238, 2005.

Heier, L. and Storvik, G. and Davis, S. A. and Viljugrein, H. and Ageyev, V, A and Klassovskaya, E. and Stenseth NC. Emergence, Spread, Persistence and Fade-out of Sylvatic Plague in Kazakhstan. Submitted, 2010.

Y.C. MacNab. Hierarchical Bayesian modeling of spatially correlated health service outcome and utilization rates. *Biometrics*, 59(2):305–316, 2003.

L. Knorr-Held and H. Rue. On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics*, 29(4):597–614, 2002.

H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*. Chapman & Hall, 2005.