Spatial discrete hazards using Hierarchical Bayesian Modeling

Mathias Graf
ETH Zurich, Institute for Structural Engineering,
Group Risk & Safety

Papers

-Maes, M.A., Dann M., Sarkar S., and Midtgaard, A.K., (2007) Fatality rate modeling within a spatial network using hierarchical Bayes methods, Web-published in the Proceedings International Forum on Engineering Decision Making, IFED2007, Port Stephens, December.

- Ng, K., Hung, W. and Wong, W. (2002) Algorithm for assessing the risk of traffic accident. Journal of Safety Research, 33, pp. 387-410.

Content

Risk analysis of a road network

- Establish a model using a classical approach
- Establish a model using a hierarchical Bayesian modeling approach
- Using WinBUGS to establish a model and estimate model parameters
- Using the model for risk analysis and decision making

Risk analysis of a road network

- Establishing an accident event estimation model by means of potential causal factors

- Identifying significant causal factors

- Identifying high risk zones

Classical approach

Example Hong Kong (Ng, K. et all)

-Data merging

-Data clustering

-Model calibration

-Black site identification

Data merging

Combining accident data with:

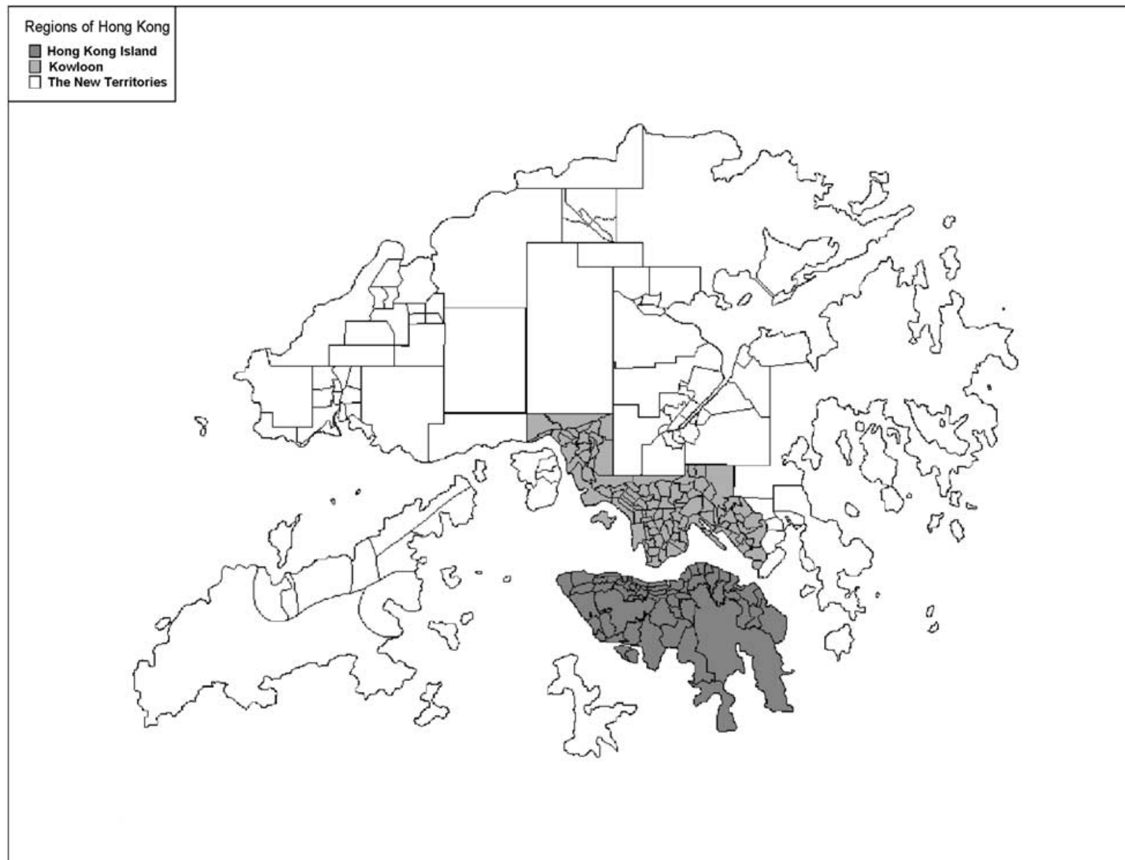
-Digital map
    -Zones

-Potential causal factors
    -Land use information
    -Road information

Data merging

-274 traffic analysis zone

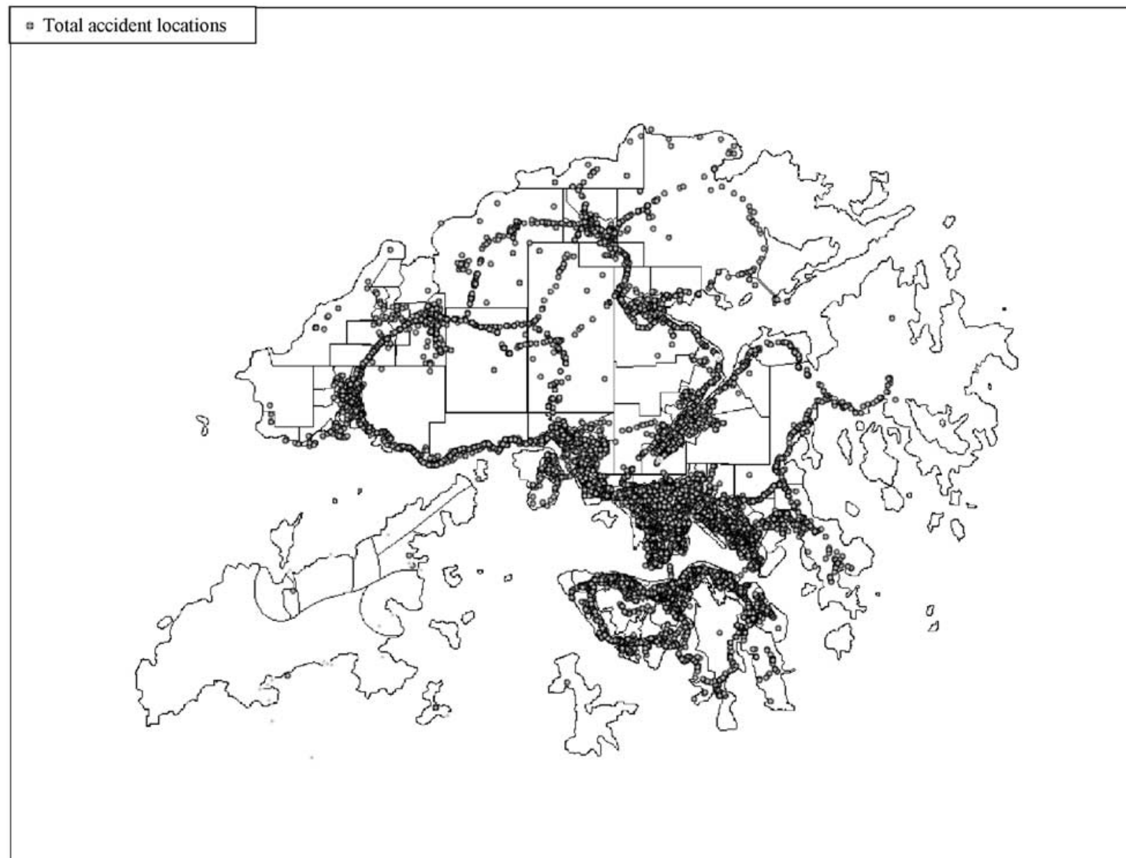
-Land use data



Data merging

-274 traffic analysis zone

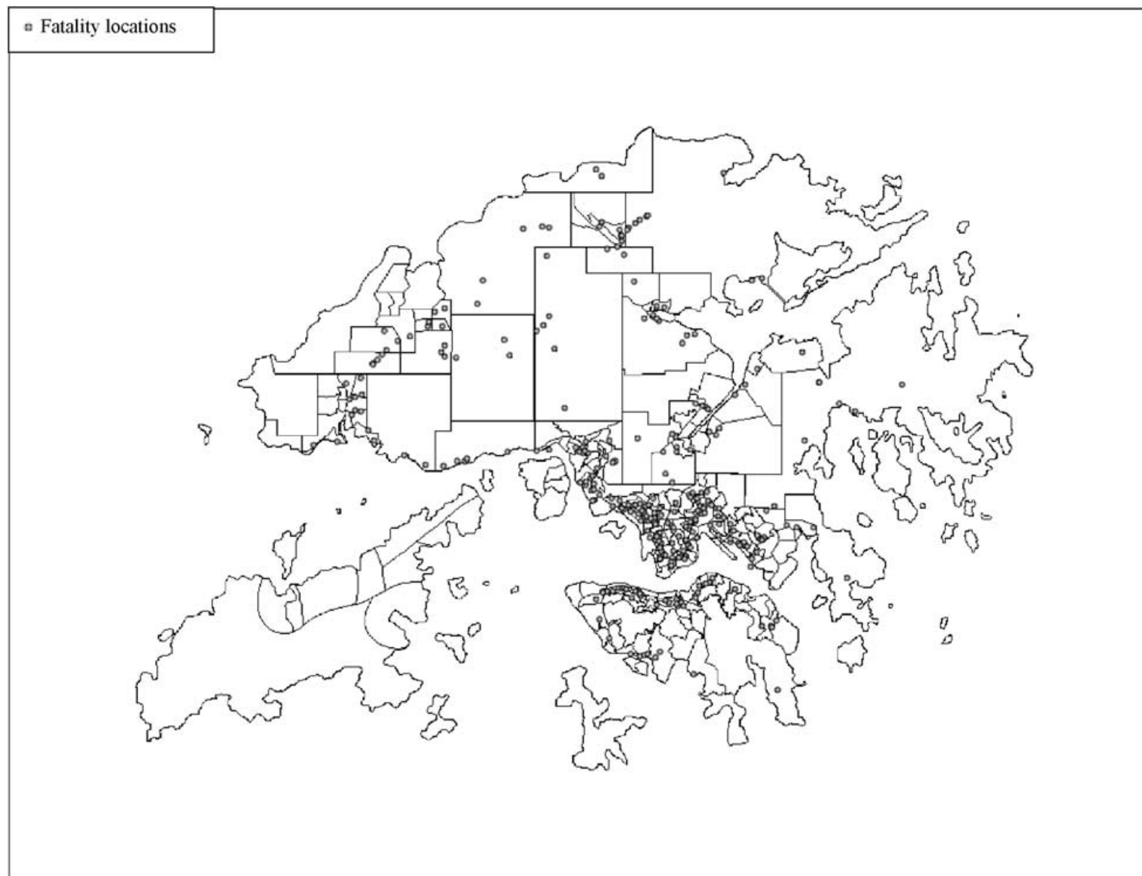
-Accident locations



Data merging

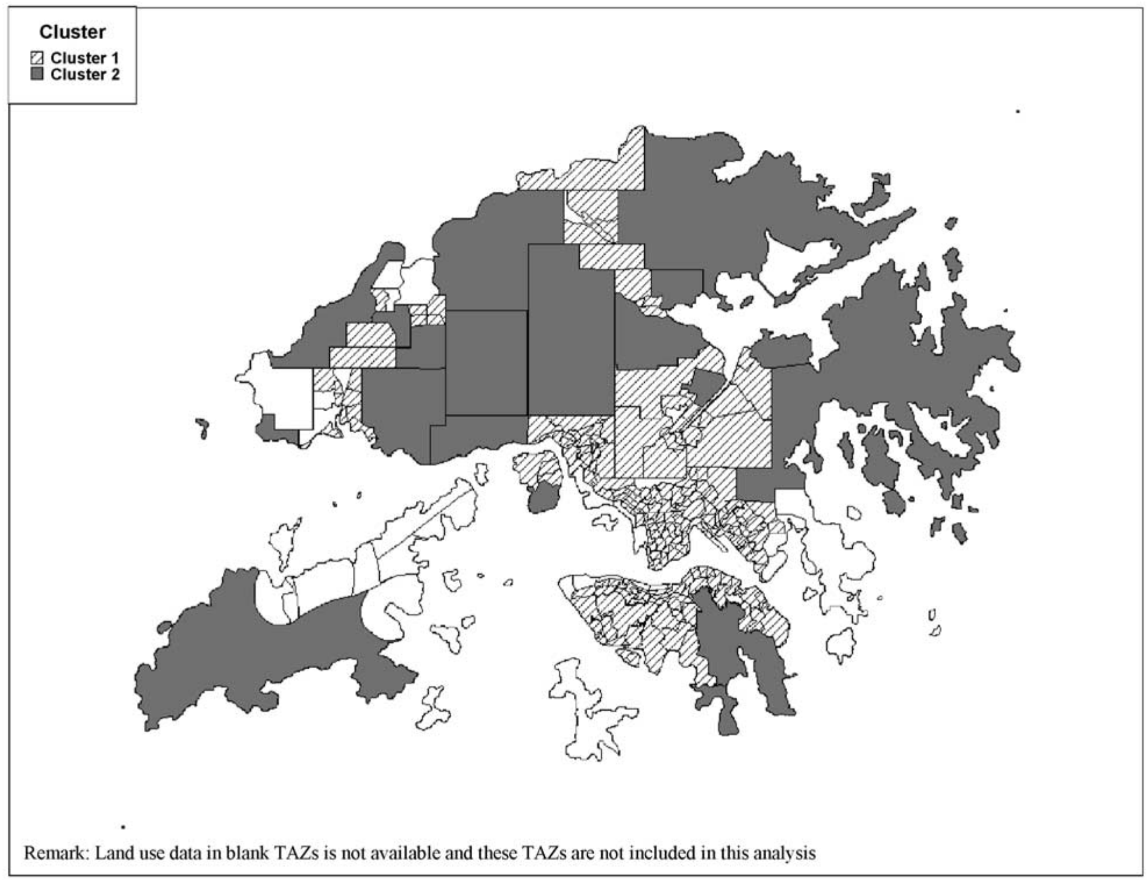
- 274 traffic analysis zone

- Fatality locations



Data clustering

-Internal homogenous & external heterogeneous



Defining model

- Literature search for existing models for similar problems.

- Appling physical models.

- Establishing a statistical model using availible data.

Model calibration

- Choosing model
- Estimating model parameters for each zone separately

$$P(Y=y)=\frac{\Gamma\left(y+\alpha^{-1}\right)}{y!\Gamma\left(\alpha^{-1}\right)}\left(\frac{\alpha\lambda}{1+\alpha\lambda}\right)^{y}\left(\frac{1}{1+\alpha\lambda}\right)^{\alpha-1}$$

α = overdispersion parameter $\alpha \geq 0$
y = number of accident events
λ = expected mean number of accident events.

Model calibration

- Choosing model
- Estimating model parameters for each zone separately

$$P(Y=y)=\frac{\Gamma\left(y+\alpha^{-1}\right)}{y!\Gamma\left(\alpha^{-1}\right)}\left(\frac{\alpha\lambda}{1+\alpha\lambda}\right)^{y}\left(\frac{1}{1+\alpha\lambda}\right)^{\alpha-1}$$

λ = expected mean number of accident events

$$\lambda=\exp(x\beta)$$

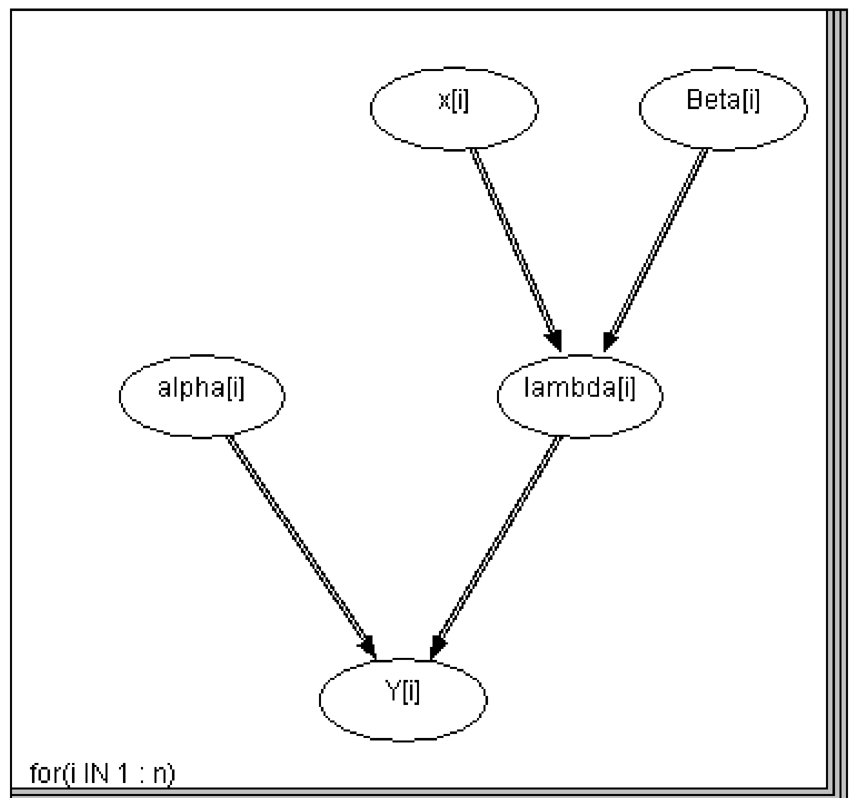
x = vector with indicators

β = vector with regression coefficients

Model calibration

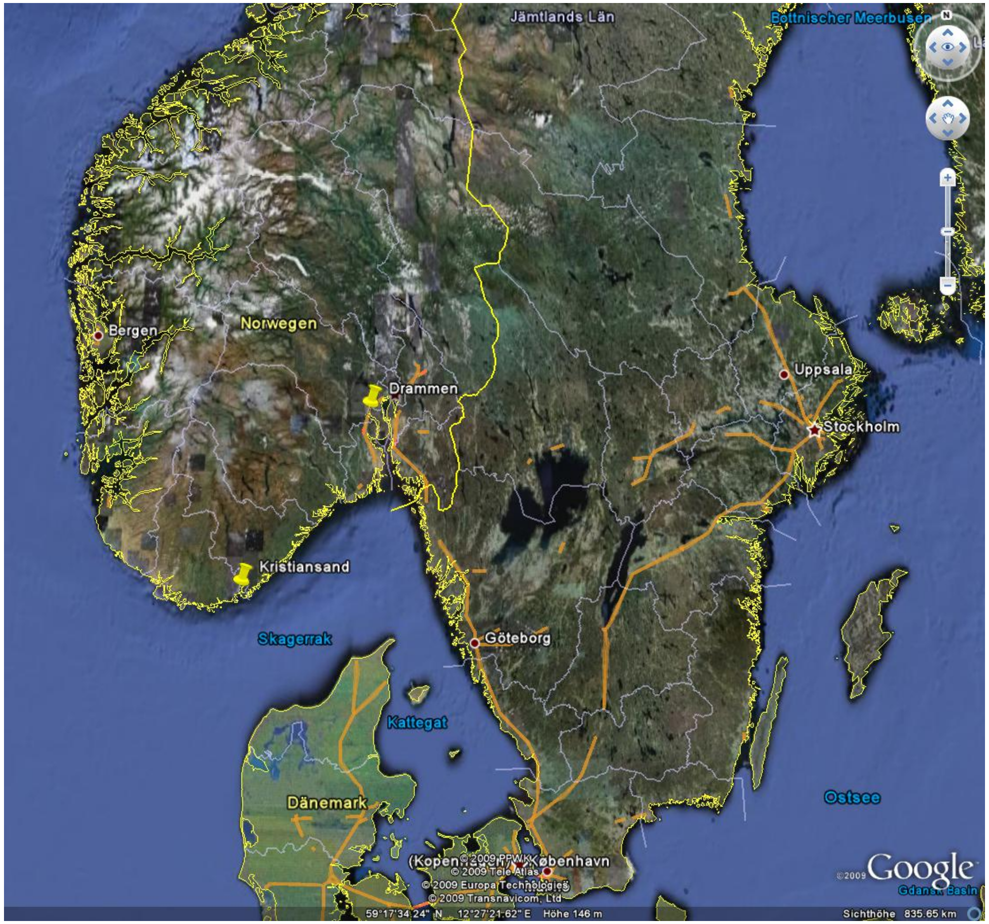
$$P(Y=y)=\frac{\Gamma\left(y+\alpha^{-1}\right)}{y!\,\Gamma\left(\alpha^{-1}\right)}\left(\frac{\alpha\lambda}{1+\alpha\lambda}\right)^{y}\left(\frac{1}{1+\alpha\lambda}\right)^{\alpha-1}$$

$$\lambda=\exp\left(x\beta\right)$$



Establishing a model using a hierarchical Bayesian modeling approach

Example Norway costal road E18 (Maes, M.A et all)



Segmentation of the road

- Different road characteristics

- Accident data

Accident data

Accident outcomes

Outcome		Definition
F	Fatality	Dead occurring within 30 days after the accident
D	Disability	Degree of disability more than 30%
S	Serious injury	Degree of disability less than 30% and at least 24h stay in hospital
I	Injury	No hospital stay required and light injury

Random variables describing the accident outcome

	Random Variable	Description
Events	Y_{AF}	Number of accident events involving fatalities F
	Y_{AFDS}	Number of accident events involving (F + D + S in Table 1)
People	Y_F	Number of fatalities F
	Y_{FDS}	Number of (F + D + S in Table 1)

Adding spatial reference to accident data

Number of events Y observed in 5 years and empirical estimates of annual accident and fatality rates θ

Segment	Y_{AF}	Y_{AFDS}	Y_F	Y_{FDS}	θ_{AF}	θ_{AFDS}	θ_F	θ_{FDS} 1/(km*year)
1	0	2	0	2	0.000	0.058	0.000	0.058
2	1	5	1	6	0.015	0.073	0.015	0.087
3	2	6	2	6	0.047	0.141	0.047	0.141
4	2	8	3	12	0.025	0.101	0.038	0.152
5	3	9	3	10	0.027	0.082	0.027	0.091
6	0	3	0	3	0.000	0.067	0.000	0.067
7	3	10	3	15	0.024	0.080	0.024	0.121
8	0	1	0	1	0.000	0.017	0.000	0.017
9	4	19	5	29	0.017	0.082	0.022	0.126
10	0	1	0	1	0.000	0.022	0.000	0.022
11	3	14	4	17	0.030	0.140	0.040	0.170
12	6	22	7	28	0.056	0.205	0.065	0.261
13	2	4	2	6	0.024	0.048	0.024	0.071
14	7	9	8	10	0.074	0.095	0.084	0.105
15	1	3	1	3	0.005	0.014	0.005	0.014
16	0	1	0	1	0.000	0.039	0.000	0.039
17	1	1	1	1	0.087	0.087	0.087	0.087
All segments:	35	118	40	151	0.024	0.080	0.027	0.103

Available Indicators

- Annual average daily traffic (AADT)
- Speed limit
- Number of crossings
- Number of lanes
- Curviness
- Bridges
- Tunnels

Available Indicators
For each segment

Nr	ROAD SEGMENT from	to	Length L_i km	AADT x_1 10^4 cars per day	Speed limit x_2 50 km / h	# x_3 1 / km	# Lanes x_4	Curviness x_5 1 / km	Bridges x_6 10%	Tunnels x_7 30%
1	Hellemyr	Timene	6.92	1.56	1.40	3.47	2	1.59	2.20	0.17
2	Timene	Kleivsmoen	13.75	0.84	1.40	1.09	2	3.43	0.19	0.18
3	Kleivsmoen	Helldal	8.50	0.93	1.40	0.94	2	2.60	0.19	0.00
4	Helldal	Morholt	15.83	0.88	1.40	0.57	2	2.19	0.19	0.00
5	Morholt	Stoa	22.01	1.15	1.60	0.73	2	1.26	0.73	0.11
6	Stoa	Brekke	9.01	0.99	1.20	0.56	2	3.70	0.15	0.00
7	Brekke	Kvienga	24.87	0.75	1.40	0.48	2	2.02	0.36	0.04
8	Kvienga	Osterhold	11.68	0.60	1.60	0.34	3	0.81	0.64	0.00
9	Osterhold	Bamble	46.09	0.78	1.40	0.30	2	1.28	0.49	0.00
10	Bamble	Skjelsvik	9.00	1.15	1.60	0.78	3	1.20	1.06	1.57
11	Skjelsvik	Langestrand	20.00	1.18	1.60	0.60	2	2.60	0.65	0.05
12	Langestrand	Kullerod	21.44	1.82	1.40	0.61	2	1.21	0.78	0.02
13	Kullerod	As	16.84	2.13	1.60	0.36	2	0.63	0.37	0.00
14	As	Nykirke	19.00	1.57	1.40	0.32	2	1.46	0.08	0.00
15	Nykirke	Skoger	42.00	1.08	2.00	0.12	4	0.56	0.24	1.12
16	Skoger	Fjell	5.13	2.76	1.80	0.20	4	0.96	0.66	0.00
17	Fjell	Drammen	2.31	2.77	1.40	0.87	2	1.34	0.81	3.14

Defining model

- The occurrence of Y_i within a segment with the length L_i during a period with a time t is modeled by

$$Y_i \sim Poisson\left(\theta_i L_i t\right)$$

Defining model

- The occurrence of Y_i within a segment with the length L_i during a period with a time t is modeled by

$$Y_i \sim Poisson(\theta_i L_i t)$$

- Whereby $$\theta_i = \exp(\mu_i + \varepsilon_i + \phi_i)$$

Defining model

- The occurrence of Y_i within a segment with the length L_i during a period with a time t is modeled by

$$Y_i \sim Poisson\left(\theta_i L_i t\right)$$

- Whereby $$\theta_i = \exp\left(\mu_i + \varepsilon_i + \phi_i\right)$$

$\mu_i = \text{meanlog-risk} \quad \mu_i = x_i \beta$

$x_i = \left\{1, x_{1i}, \ldots, x_{mi}\right\} = \text{spatial covariates characterizing the segment } i$

$\beta_i = \left\{\beta_0, \beta_1, \ldots, \beta_m\right\} = \text{regression coefficients}$

$m = \text{number of indicators}$

Defining model

- The occurrence of Y_i within a segment with the length L_i during a period with a time t is modeled by

$$Y_i \sim Poisson\left(\theta_i L_i t\right)$$

- Whereby $$\theta_i = \exp\left(\mu_i + \varepsilon_i + \phi_i\right)$$

$$\varepsilon_i \sim N\left(0, \sigma_\varepsilon^2\right) \quad \sigma_\varepsilon^2 \simeq \text{system wide variability (nugget effect)}$$

Defining model

- The occurrence of Y_i within a segment with the length L_i during a period with a time t is modeled by

$$Y_i \sim Poisson\left(\theta_i L_i t\right)$$

- Whereby $\theta_i = \exp\left(\mu_i + \varepsilon_i + \phi_i\right)$

ϕ_i = Zero-mean random field capturing any cluster effects

$$\{\phi_1, \ldots, \phi_n\} \mid \sigma_\phi^2, r_\phi \sim \text{Multivariate normal } \left(0, \Sigma \mid \sigma_\phi^2, r_\phi\right)$$

$$\left(\Sigma\left(\sigma_\phi^2, r_\phi\right)\right)_{ij} = \sigma_\phi^2 \exp\left(-\frac{d_{ij}}{r_\phi}\right)$$

r_ϕ = correlation lenght

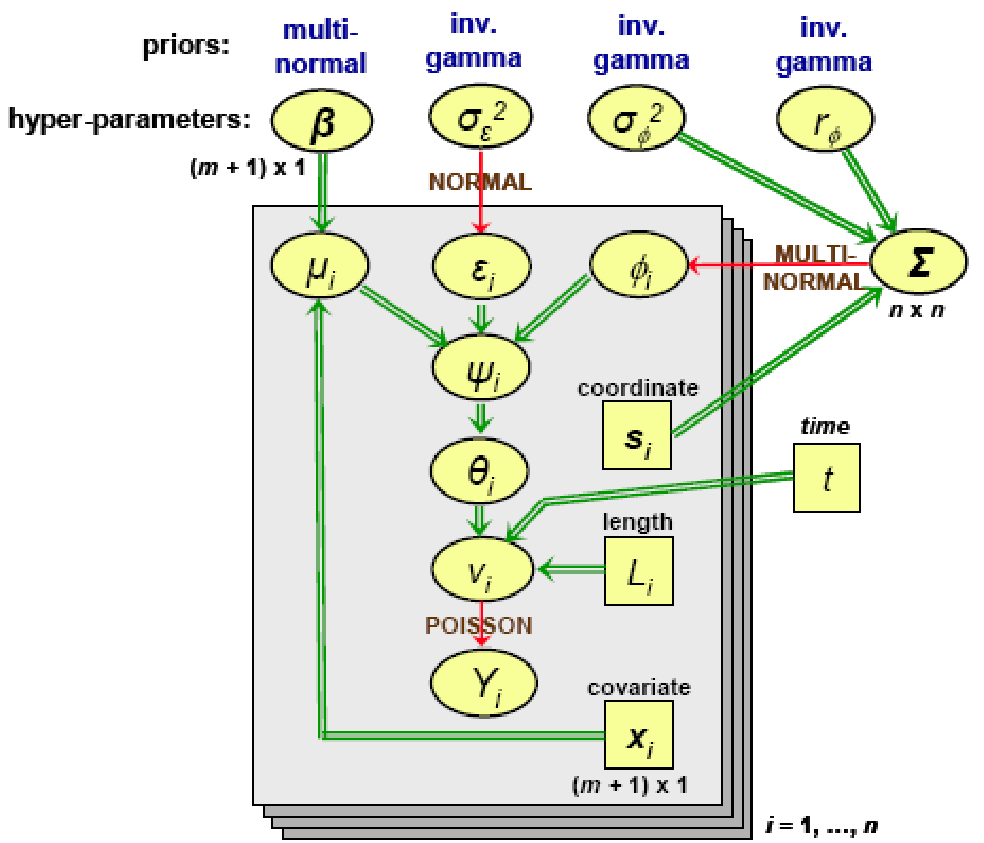
d_{ij} = distance between centroids of the segment i and segment j

Using WinBUGS for parameter estimation

- Using doodle to scratch model
- Generating code and defining model
- Preparing input data
- Updating the model to estimate parameters
- Using WinBUGS scripts to easily perform several parameter estimations in a batch mode

Using WinBUGS for parameter estimation

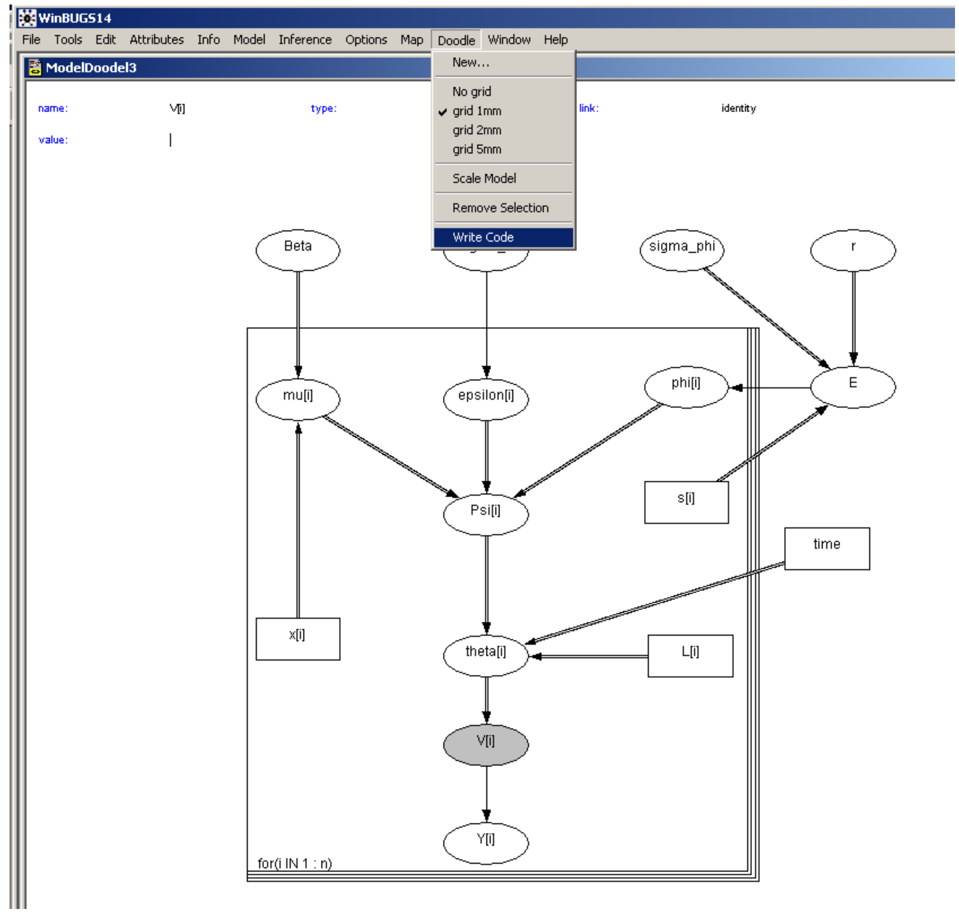
- Using doodle to scratch model



Using WinBUGS for parameter estimation

- Generate code and defining model

- Check definition of parameters in WinBUGS



Using WinBUGS for parameter estimation

- Preparing input data and initial values

WinBUGS format:

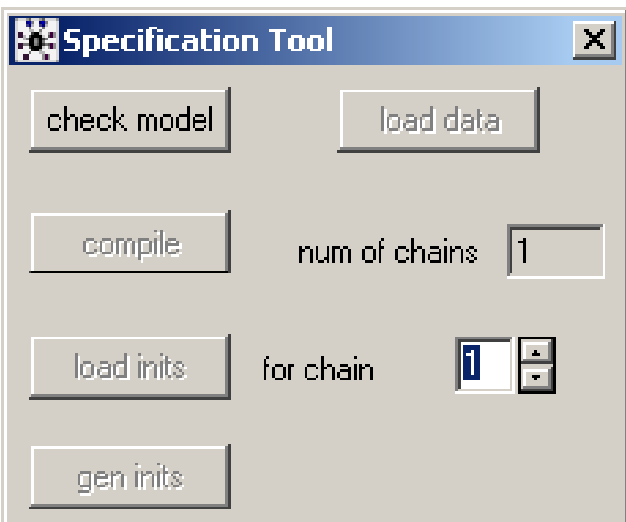
Start with list( )

- Single values: n = 17
- Vectors: Y = c(0,1,2,3,3,0,3,0,5,0,4,7,2,8,1,0,1)
- Matrix:
- x=structure(.Data=c(6.92,1.56, …,3.14),.Dim = c(17,8))

Using matlab to generate input files

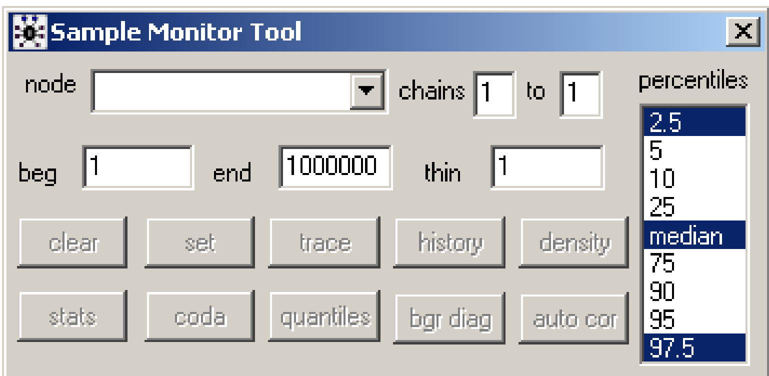
Using WinBUGS for parameter estimation

- Checking model and loading data



Using WinBUGS for parameter estimation

- Defining which variables will be monitored



- Updating: using Markov chain Monte Carlo (MCMC) methods to sample posterior distribution.

Using WinBUGS for parameter estimation

- Using WinBUGS scripts to easily perform several parameter estimations in a batch mode

- Sample script code:

```
display('log')          #progress is show in a log window
check('Model.odc')      #load and check the model
data('Data.txt')        #load data file
compile(1)              #compile model
gen.inits()             #generate initial values
set(lambda)             #define which variable will be monitored
set(theta)
```

Using WinBUGS for parameter estimation

- Using WinBUGS scripts to easily perform several parameter estimations in a batch mode

```
beg(1000)            # select a subset of the stored sample for
end(1000000)         analysis.
update(10000)        #set the number of MCMC updates
refresh(1000)        #start updates and refresh screen after
                     1000 updates

stats(*)             #show statistics of all monitored variables
save('Stats.txt')    #save the statistics in a file
```

Estimated Parameters

Acc. outcome	events Y_{AF}			events Y_{AFDS}			people Y_F			people Y_{FDS}		
Parameter	E[]	$[]_{2.5}$	$[]_{97.5}$	E[]	$[]_{2.5}$	$[]_{97.5}$	E[]	$[]_{2.5}$	$[]_{97.5}$	E[]	$[]_{2.5}$	$[]_{97.5}$
β_1 (AADT)	**0.34**	-4.6	18.0	**0.23**	-1.1	1.0	**0.28**	-1.7	1.8	**0.19**	-0.6	0.9
β_2 (speed limit)	**2.78**	-4.5	6.0	**-0.25**	-4.0	2.4	**1.68**	-4.8	10.5	**-0.40**	-3.8	2.5
β_3 (# crossings)	**-0.78**	-7.9	-0.3	**-0.75**	-4.1	0.5	**-1.00**	-4.8	1.9	**-0.69**	-1.8	0.4
β_4 (# lanes)	**-2.68**	-3.6	1.2	**-0.86**	-1.8	0.0	**-2.55**	-5.8	-0.4	**-0.99**	-2.0	0.0
β_5 (curviness)	**-0.14**	-6.4	1.5	**0.31**	-0.3	1.0	**-0.16**	-1.7	1.1	**0.19**	-0.3	0.8
β_6 (bridges)	**-1.40**	-1.9	2.2	**0.97**	-0.9	6.1	**-1.11**	-5.5	1.9	**0.71**	-1.0	2.2
β_7 (tunnels)	**0.36**	-14.6	6.8	**-0.30**	-1.1	0.4	**0.31**	-0.9	1.8	**-0.33**	-1.1	0.3
r in km	**9.63**	3.8	34.2	**11.62**	3.8	50.9	**9.96**	3.8	37.2	**11.51**	3.8	47.7
σ_ϕ	**0.52**	0.0	2.3	**0.34**	0.0	2.0	**0.53**	0.0	1.9	**0.29**	0.0	1.0
σ_ε	**0.45**	0.0	1.9	**0.21**	0.0	0.7	**0.44**	0.0	1.6	**0.24**	0.0	0.8
factor α (7)	**0.52**	0.0	1.0	**0.54**	0.1	0.9	**0.54**	0.1	0.9	**0.52**	0.1	0.9

- Check influence of indicators
- Use estimated parameters to estimate risk for road segments

Checking model

Cross validation

Establish model using 16 segments and test it on the 17^{th} segment.

Estimate the prediction error of the model.

Difference between hierarchical and non-hierarchical approach.

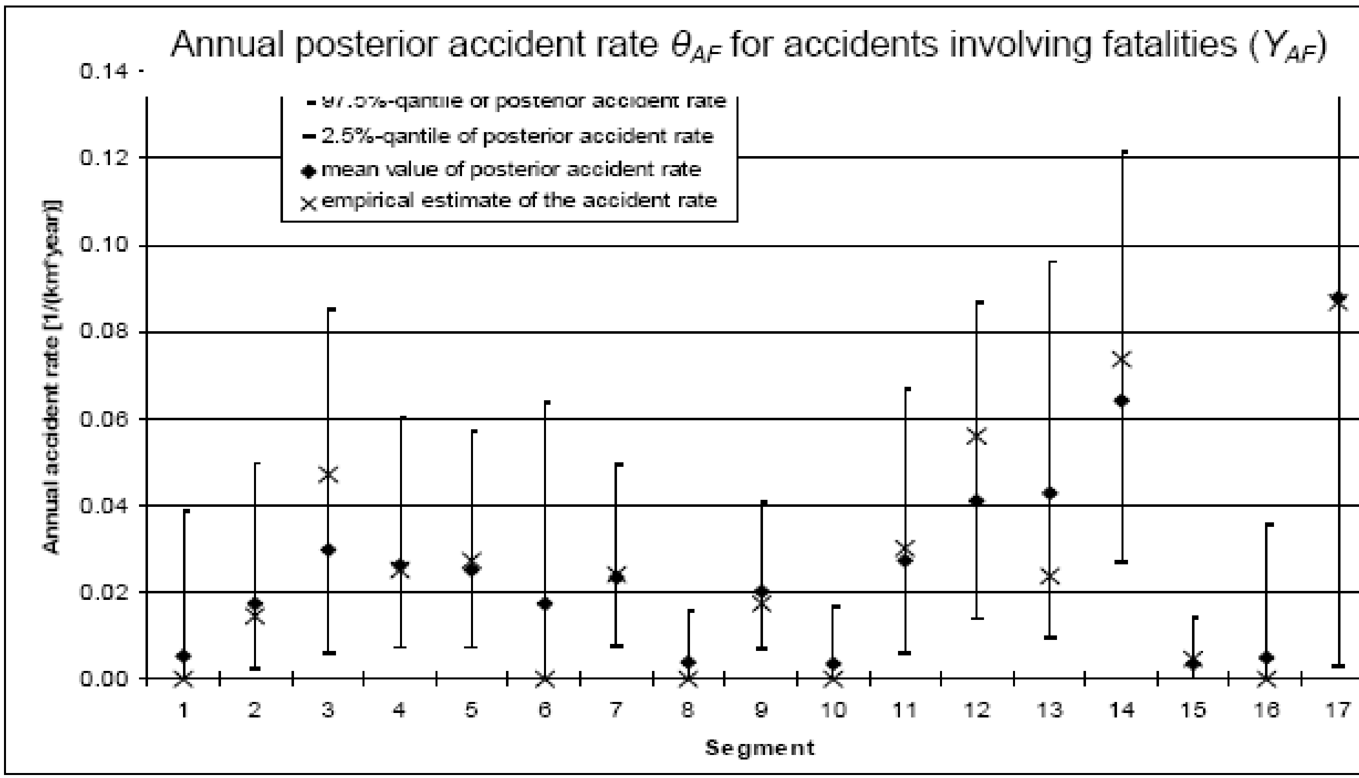


Figure 3: Annual prior and posterior accident rate θ_{AF}

Decision making

- Estimating risk for the selected road segments

- Estimating risk for new road segments

- Using the risk analysis to allocate resources to improve high risk areas

Real-time risk analysis

- Conditioning network on road conditions
  - Lanes closed, construction site…

- Including other indicators
  - Traffic news
  - Day time
  - Weather (forecast)

- Real time monitoring system