

Chain Graphs for Spatial Dependence in Ecological Data

Alix I Gitelman¹ & Alan Herlihy²

¹Statistics Department

²Department of Fisheries & Wildlife

Oregon State University

STARMAP, Colorado State University

September, 2004

STAR Meetings, Ft Collins CO

The Disclaimer

This presentation was developed under STAR (Science to Achieve Results) Research Assistance Agreements CR-829095 and CR-829095 awarded by the US Environmental Protection Agency (EPA) to Colorado State University and Oregon State University, respectively. The presentation has not been formally reviewed by the EPA. The views expressed here are solely those the authors and respective programs under these two agreements. The EPA does not endorse any products or commercial services mentioned in this presentation.

A Detour

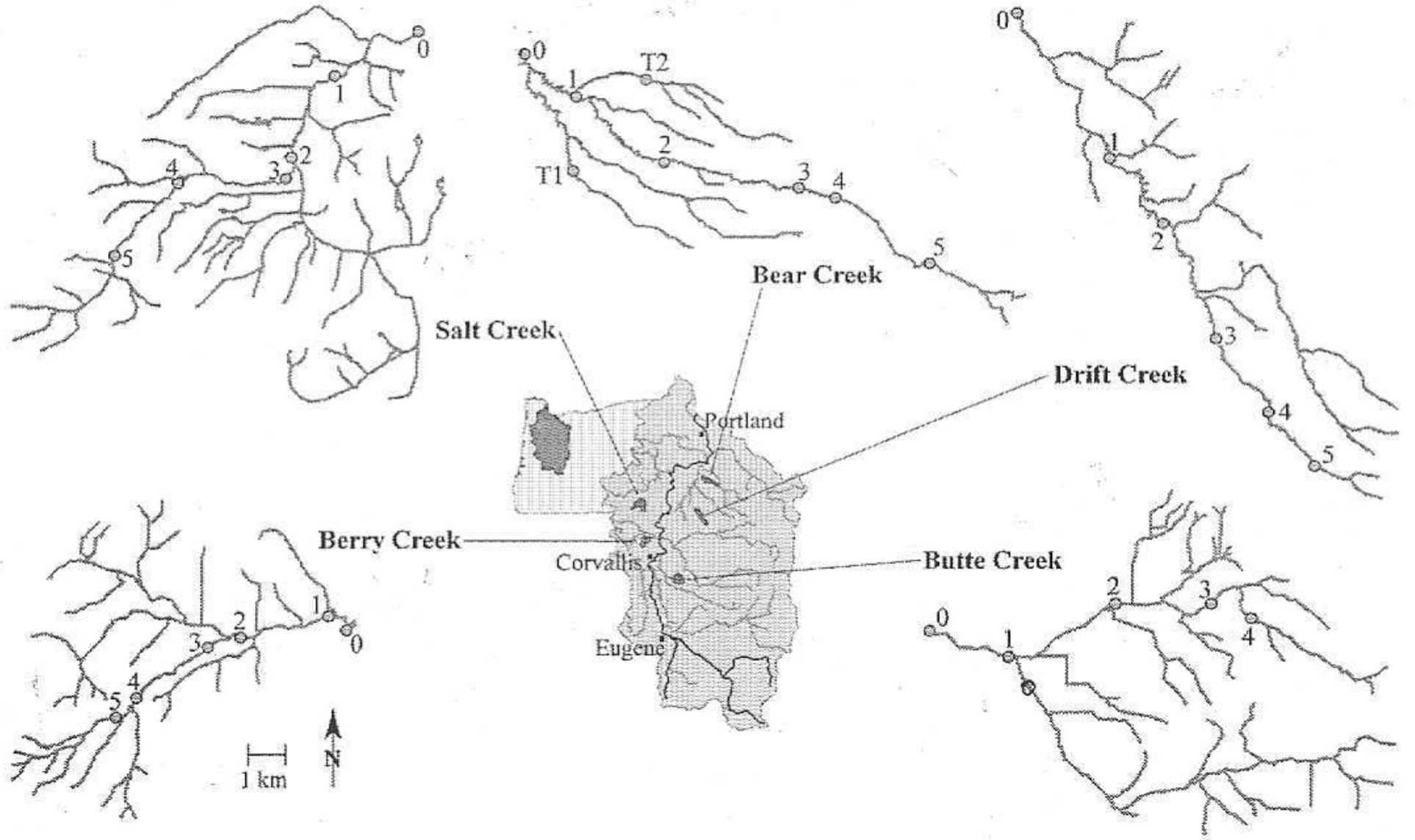
1. Collaboration with Ken Reckhow's group at Duke
2. ICG paper
3. Steve Jensen, MS; CleverSet, Inc.

Collaboration with Duke

- Linking Bayesian SPARROW (Qian et al. 2004) with a Bayes network model for chlorophyll (Neu-BERN; Borsuk et al. 2003).
- SPARROW is a “process” model connecting land-use to nitrogen load.
- Neu-BERN is a Bayes network model implemented on the Neuse River Estuary in NC; it connects chlorophyll concentration to nitrogen loads.

ICG Models for Spatial Dependence

- Willamette Valley Data
- Some schematics
- Isomorphic chain graphs
- Back to the Willamette Valley
- Computing issues



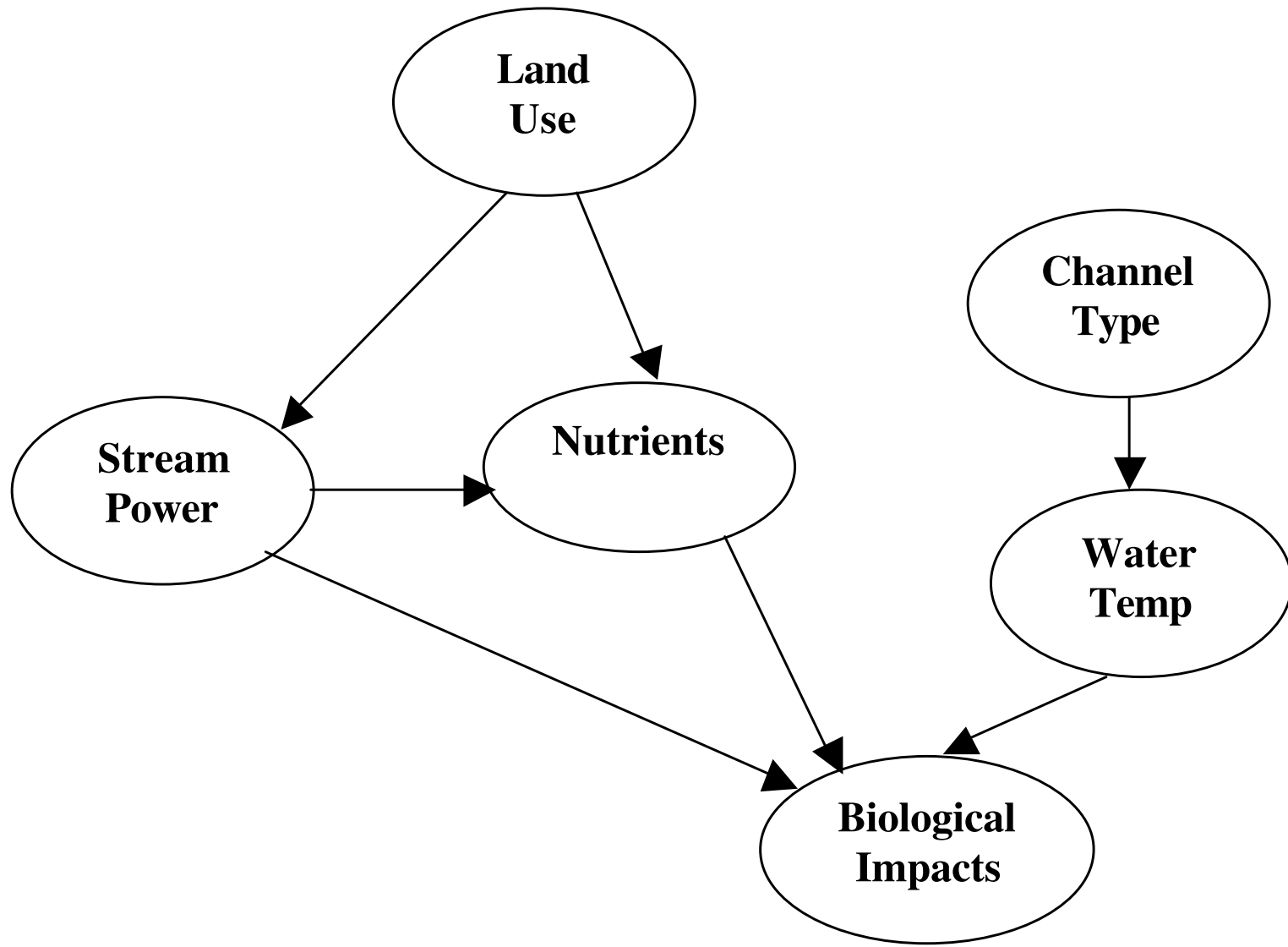
Willamette Valley Data

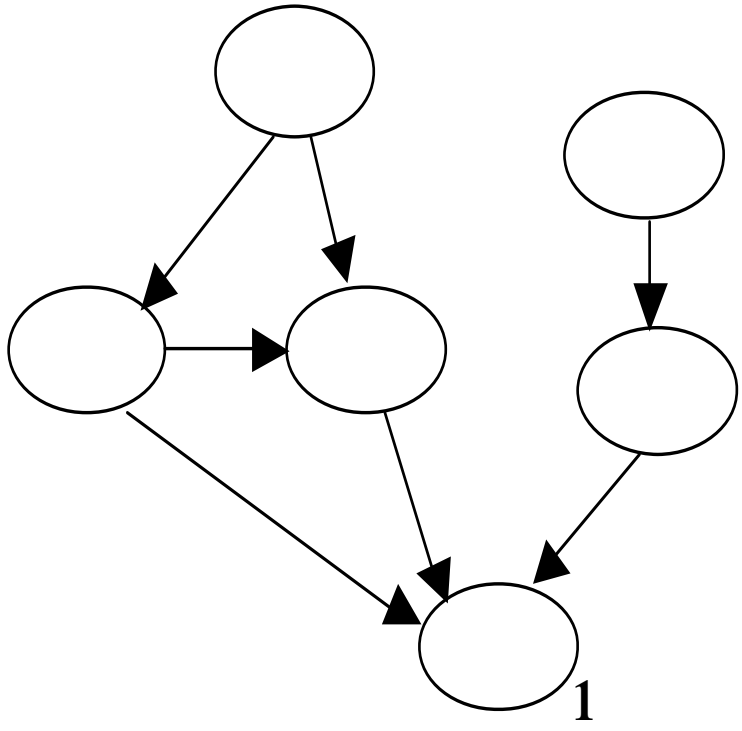
land use	stream power	temperature
nutrients	sediment	channel complexity
riparian condition	biological impacts	

n = 76

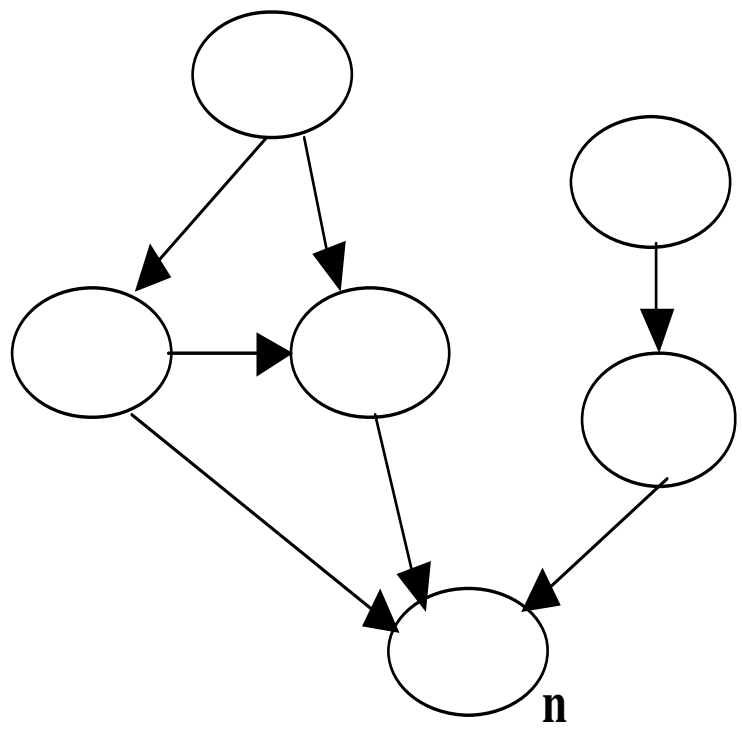
sampled: 1997, 1999

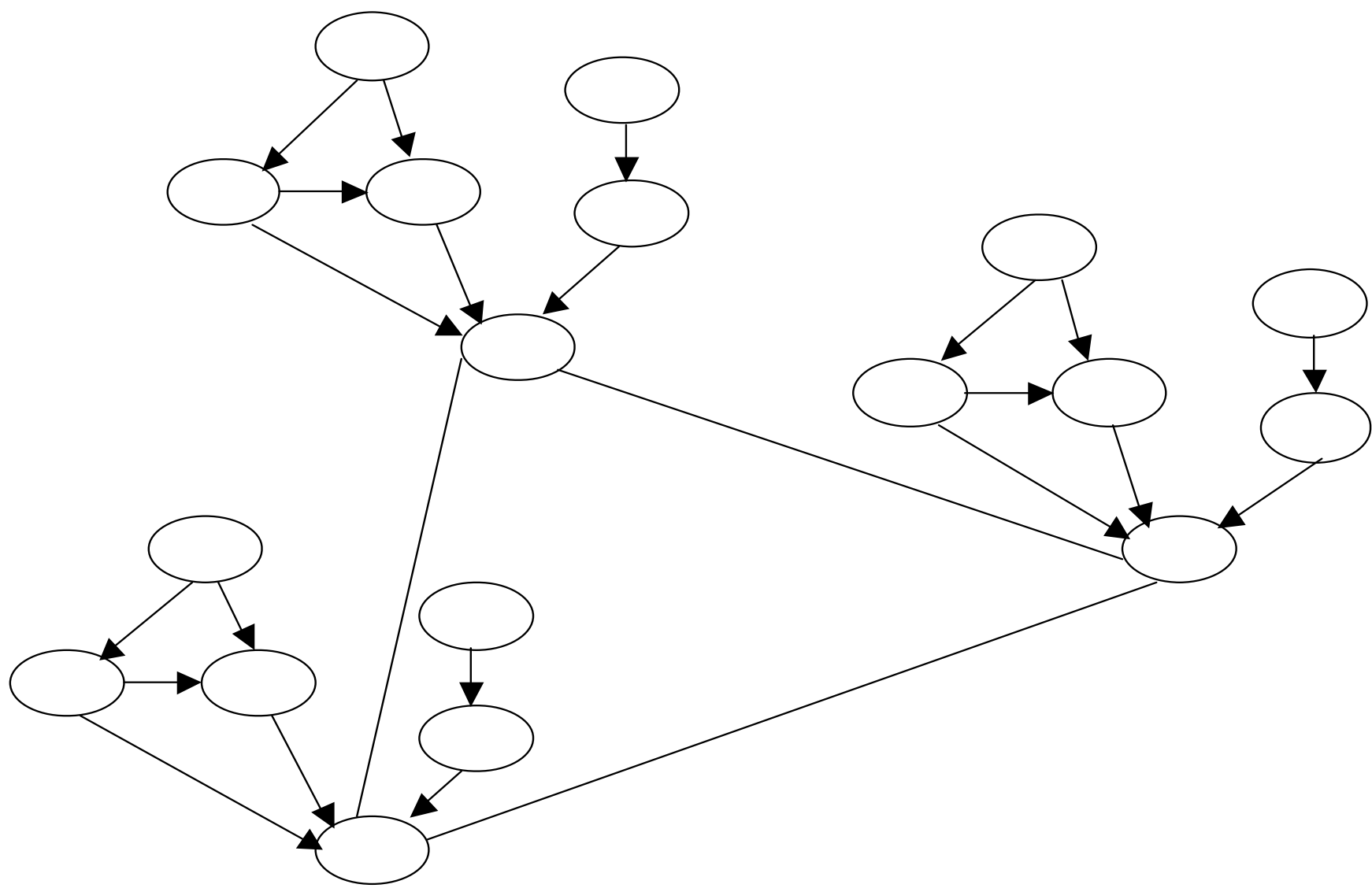
Details in Van Sickle et al. (2004).

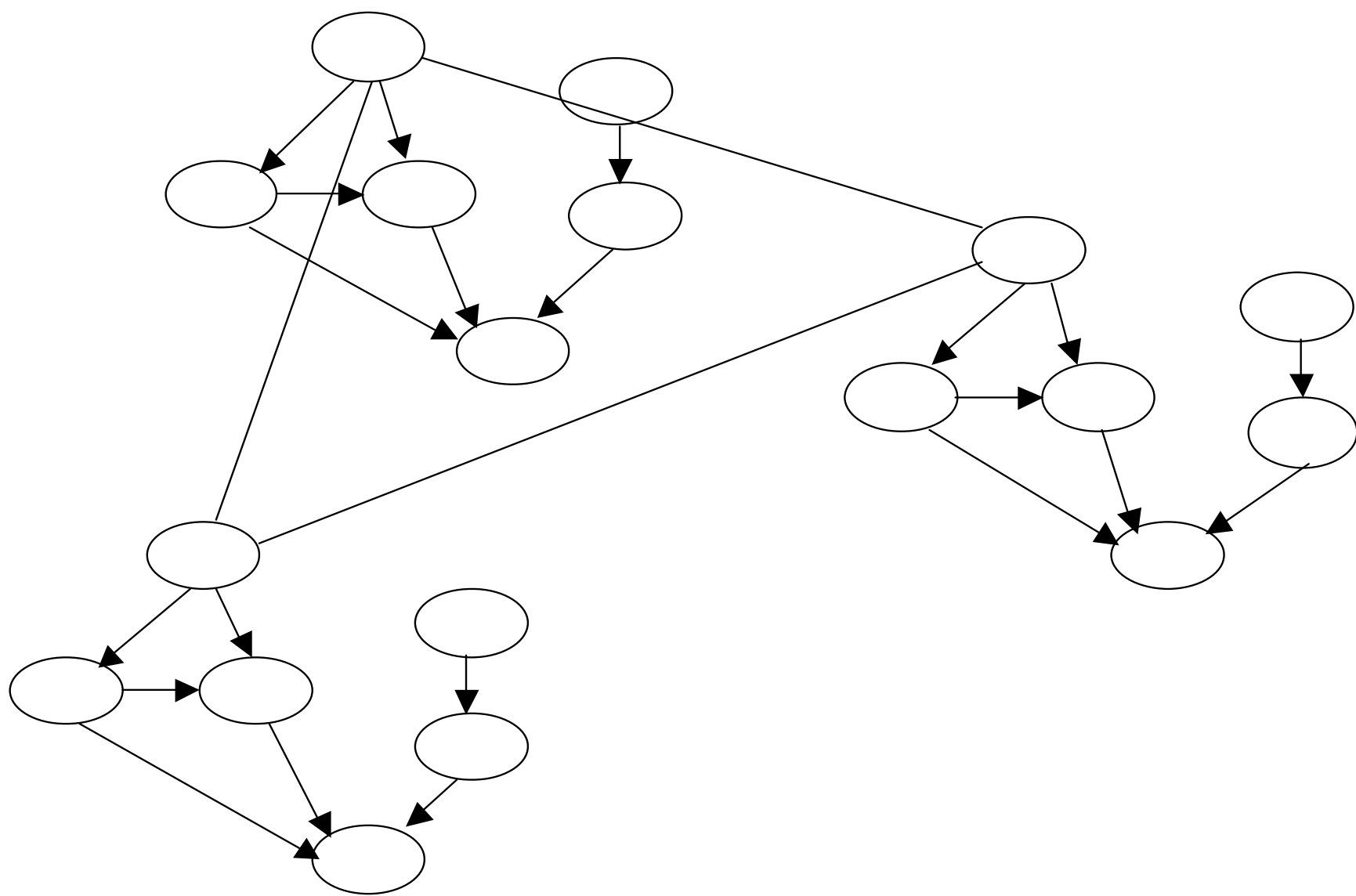




...







Summary of Results 1 & 2

(from Gitelman & Herlihy 2004).

1. The conditional independencies of the marginal Bayes networks are retained in the isomorphic chain graph.
2. Conditional independence holds across marginal Bayes networks, conditioned on the isomorphic nodes.

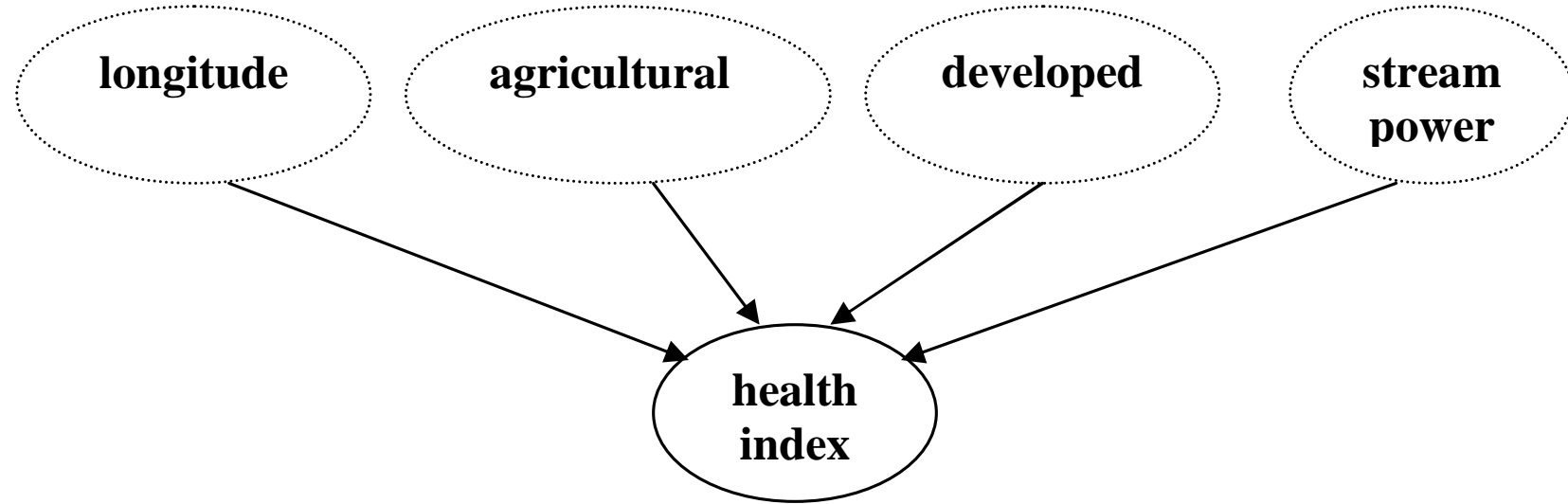
This means that the joint probability distribution associated with an ICG model has a Markovian factorization.

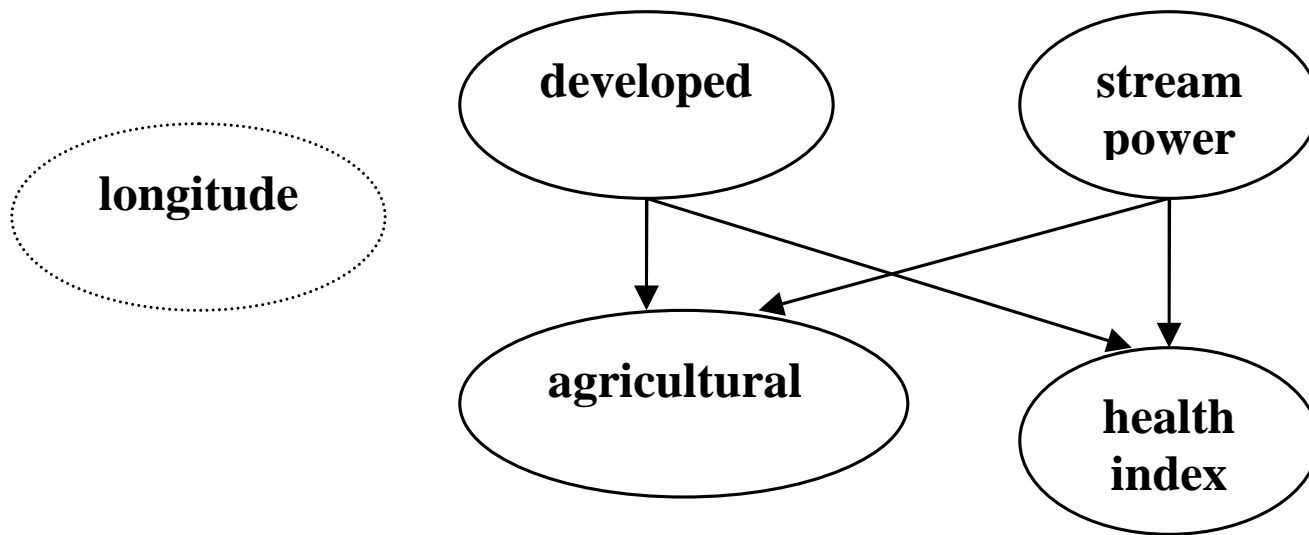
Willamette Valley Data

Van Sickle et al. (2004) find longitude, percent agricultural land cover, percent developed land cover and stream power to be useful for understanding the variation in WINOE (Willamette invertebrate observed/expected index)

We use DOE (dispersers observed/expected index) instead ($\hat{\rho} = 0.93$).

For comparison, we build an ICG model using the same “explanatory” variables.





Multiple Linear Regression Model

The multiple linear regression model, parameterized as a Bayes network model:

$$U_i \stackrel{iid}{\sim} \text{Exp}(\theta)$$

$$P_i \stackrel{iid}{\sim} N(\mu_p, \sigma_p^2)$$

$$A_i \stackrel{ind}{\sim} N(\mu_a, \sigma_a^2)$$

$$H_i \stackrel{ind}{\sim} N(\gamma_i, \sigma^2)$$

where

$$\gamma_i = \beta_0 + \beta_1 L_i + \beta_2 A_i + \beta_3 U_i + \beta_4 P_i$$

The Bayes Network Model

$$U_i \stackrel{iid}{\sim} \text{Exp}(\theta)$$

$$P_i \stackrel{iid}{\sim} N(\mu_p, \sigma_p^2)$$

$$A_i \stackrel{ind}{\sim} N(\delta_i, \sigma_a^2)$$

$$H_i \stackrel{ind}{\sim} N(\gamma_i, \sigma^2)$$

where

$$\delta_i = \phi_0 + \phi_1 P_i + \phi_2 U_i$$

$$\gamma_i = \beta_0 + \beta_1 P_i + \beta_2 U_i$$

ICG Models

- In each case, the isomorphic node distributions are modeled using a spatial autoregressive model (Ord, 1975):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\epsilon}$$

where \mathbf{W} is a weight-matrix based on sample location adjacencies. In particular, $W[i, j] = 0$ if samples i and j are from a different stream.

- ICG₁ DOE is the isomorphic node
- ICG₂ Agriculture is the isomorphic node

parameter	MLR	BN	ICG ₁	ICG ₂
Stream power				
μ_p	1.14 (.03)	1.14 (0.03)	1.14 (0.04)	1.14 (0.04)
σ_p	0.13 (0.01)	0.14 (0.01)	0.13 (0.01)	0.14 (0.01)
Agriculture				
μ_a	42.7 (2.8)	NA	NA	NA
σ_a	24.3 (2.0)	20.54 (1.72)	20.58 (1.75)	20.79 (1.78)
intercept	NA	87.58 (9.6)	87.63 (9.6)	85.74 (10.0)
strm power	NA	-0.46 (0.11)	-0.46 (0.11)	-0.46 (0.12)
developed	NA	-35.04 (8.0)	-35.06 (8.0)	-34.12 (8.3)
Health metric				
intercept	31.58 (7.9)	0.14 (0.07)	0.13 (0.06)	0.14 (0.07)
longitude	-0.25 (0.06)	NA	NA	NA
agriculture	-0.0023 (0.0007)	NA	NA	NA
developed	-0.0032 (0.0008)	-0.0019 (0.0008)	-0.0014 (0.0008)	-0.0019 (0.0008)
strm power	0.30 (0.05)	0.34 (0.07)	0.31 (0.05)	0.34 (0.06)
σ	0.30 (0.02)	0.30 (0.02)	0.30 (0.02)	0.30 (0.02)
ρ	NA	NA	0.42 (0.16)	0.02 (0.02)
Developed				
θ	0.09 (0.01)	0.09 (0.01)	0.09 (0.01)	0.09 (0.01)

Model Comparison

Because all 4 models have similar likelihood components, it seems reasonable to compare them using BIC.

Model	BIC	number of parameters
MLR	1184	11
BN	1179	11
ICG ₁	1174	12
ICG ₂	1184	12

Computing Issues

1. Need to incorporate spatial dependence in the model selection process
2. Reversible Jump MCMC
3. Here's Steve....