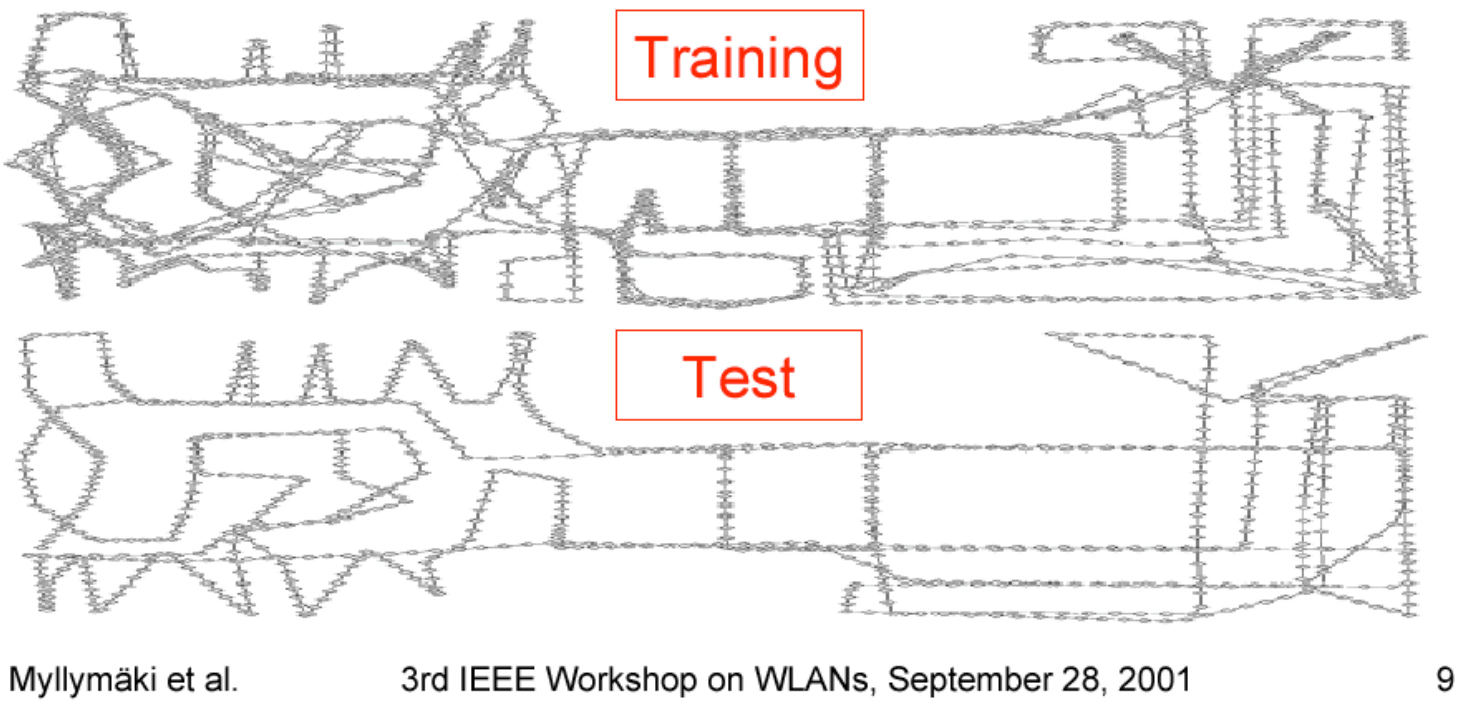
Bayesian Graphical Models for Location Determination

David Madigan
Columbia University

joint work with Wen-Hua Ju, P. Krishnan, and A.S. Krishnakumar at Avaya Labs Research and Richard P. Martin and Eiman Elnahrawy at Rutgers CS

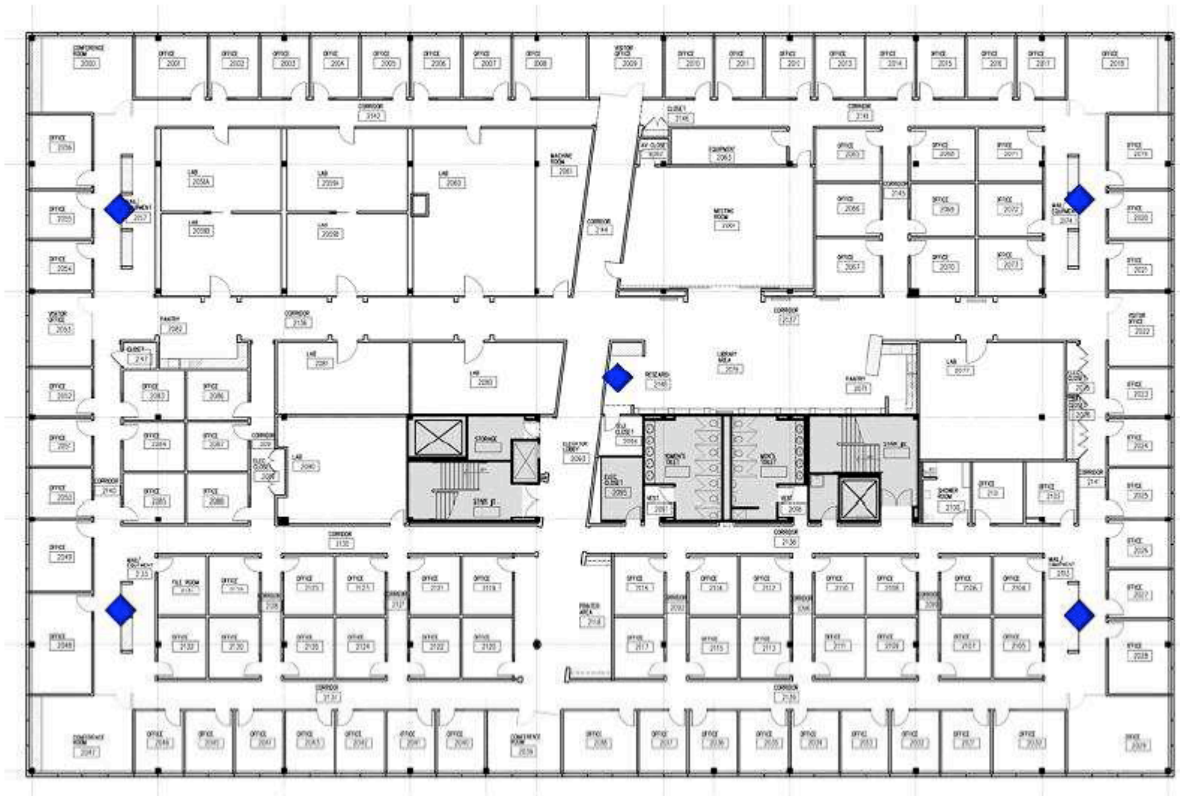
Location Determination via Statistical Modeling



- Data collection is slow, expensive ("profiling")
- "Productization"
- Either the access points or the wireless devices can gather the data
- Focus on predictive accuracy

The Problem

- **Estimate the physical location of a wireless terminal/user in an enterprise**
  - Radio wireless communication network, specifically, 802.11-based



Example Applications

- Use the closest resource, e.g., printing to the closest printer
- Security: in/out of a building
- Emergency 911 services
- Privileges based on security regions (e.g., in a manufacturing plant)
- Equipment location (e.g., in a hospital)
- Mobile robotics
- Museum information systems

WhereNet

WhereNet, wireless solutions for tracking and managing assets.

HUMMER FACTORY IMPLEMENTS WHERENET'S WIRELESS SOLUTIONS TO ENHANCE ASSEMBLY LINE OPERATIONS AND EXPEDITE PRODUCTION

Wireless Supply Chain Technology for Military Transportation, Logistics & Security

ekahau

Ekahau Positioning Engine™ 2.1

AeroScout®
Enterprise Visibility Solutions

LEGOLAND

AeroScout and Kidspotter launch world's largest Wi-FI location network for child tracking at LegoLand Denmark.

KidSpotter
Location-Finding System

NEWBURY NETWORKS

Locate.
Manage.
Secure.
with
WiFi Workplace

Automotive News®

SEPTEMBER 8, 2003

Automakers go wireless to find cars in plants

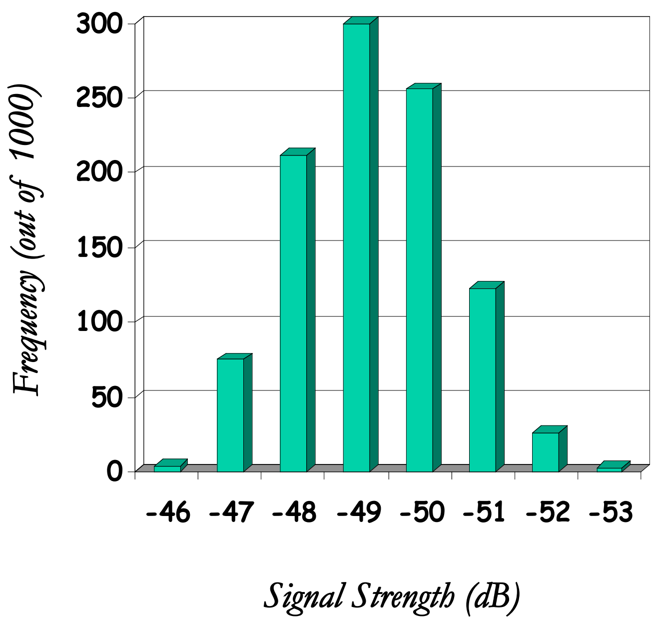
Physical Features Available for Use

- **Received Signal Strength (RSS) from multiple access points**
- **Angles of arrival**
- **Time deltas of arrival**
- **Which access point (AP) you are associated with**

- **We use RSS and AP association**
  - RSS is the only reasonable estimate with current commercial hardware

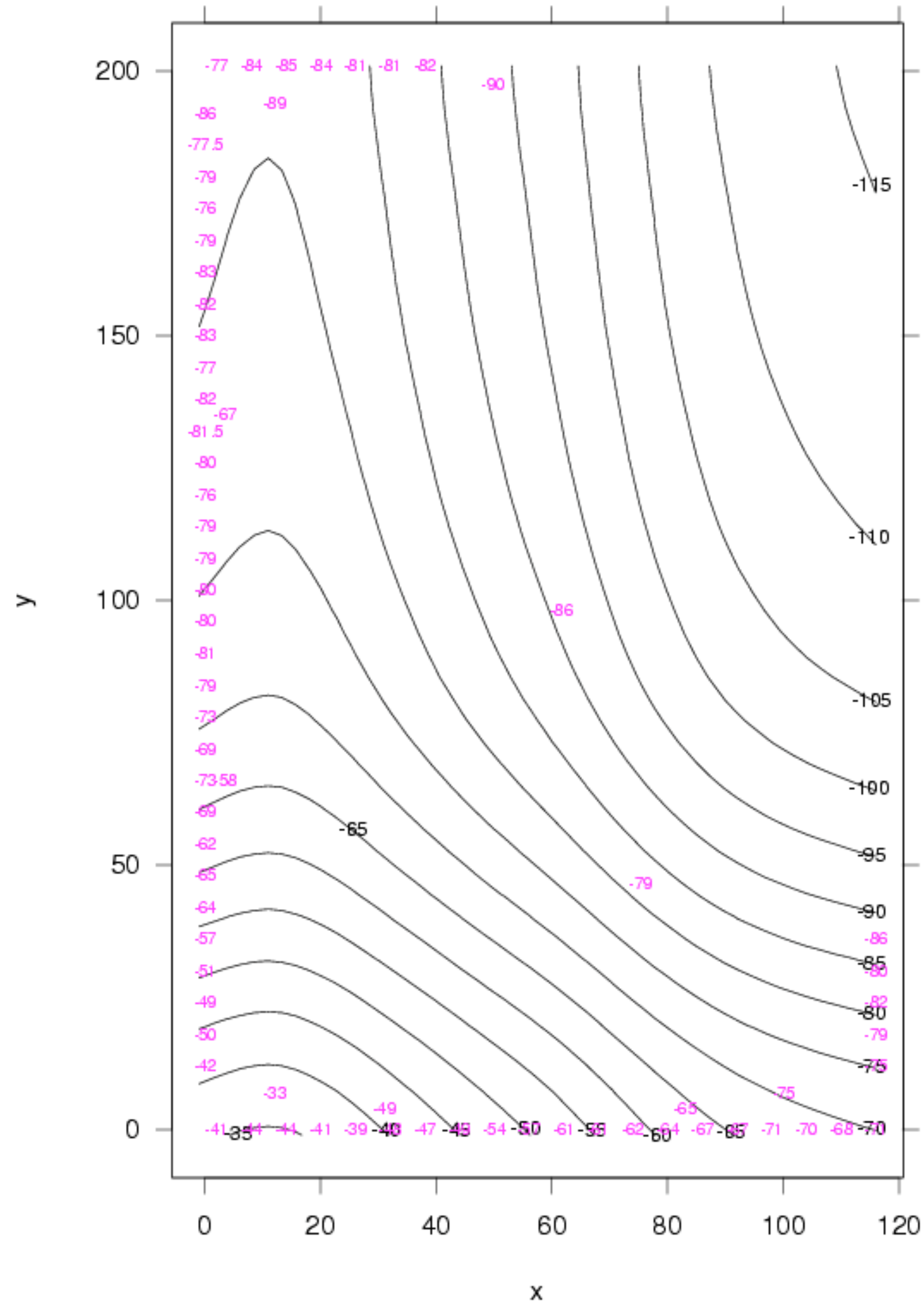
Known Properties of Signal Strength

- Signal strength at a location is known to vary as a log-normal distribution with some environment-dependent σ
- Variation caused by people, appliances, climate, etc.

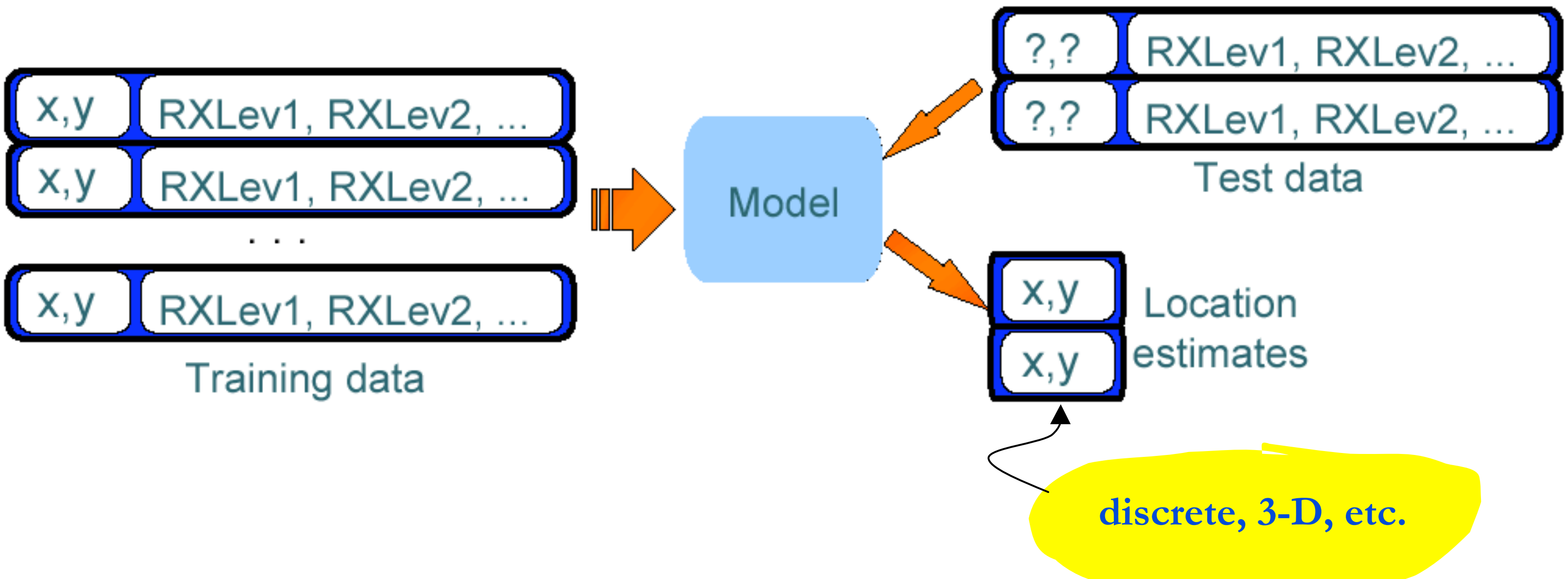


- The Physics: signal strength (SS; in dB) is known to decay with distance (d) as $SS = k_1 + k_2 \log d$

Access point: 1



Prior Work



- Take signal strength measures at many points in the site and do a closest match to these points in signal strength vector space. [e.g. Microsoft's RADAR system]
- Take signal strength measures at many points in the site and build a multivariate regression model to predict location (e.g., Tirri's group in Finland)
- Some work has utilized wall thickness and materials

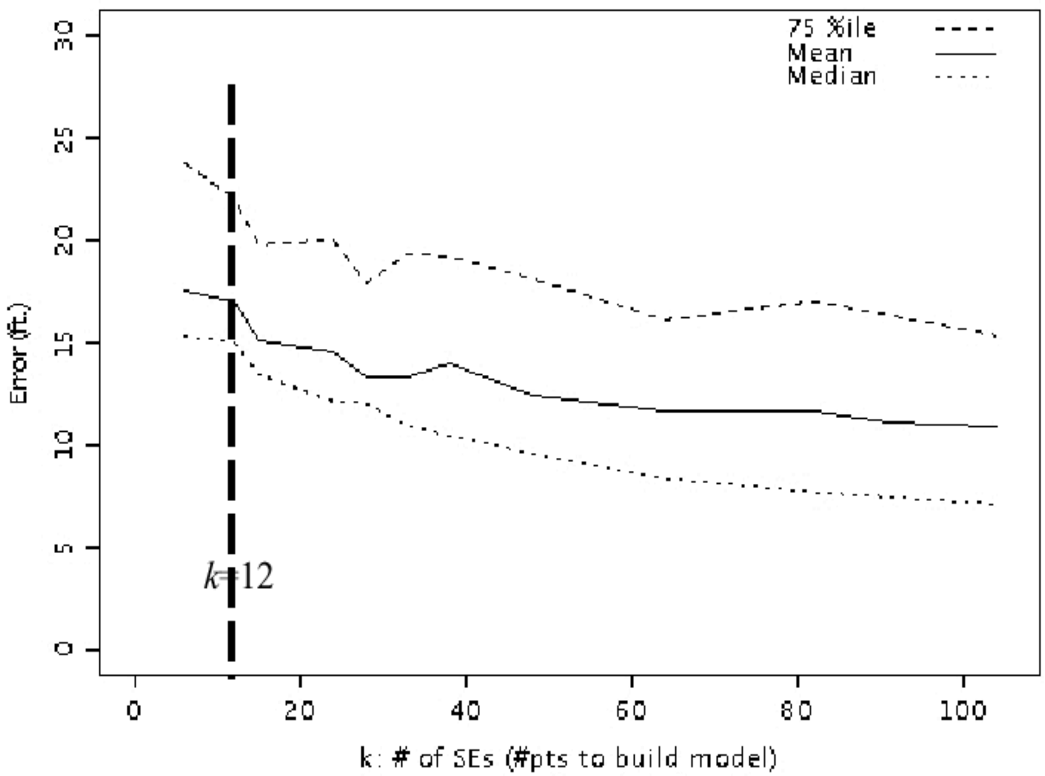
Prior Work

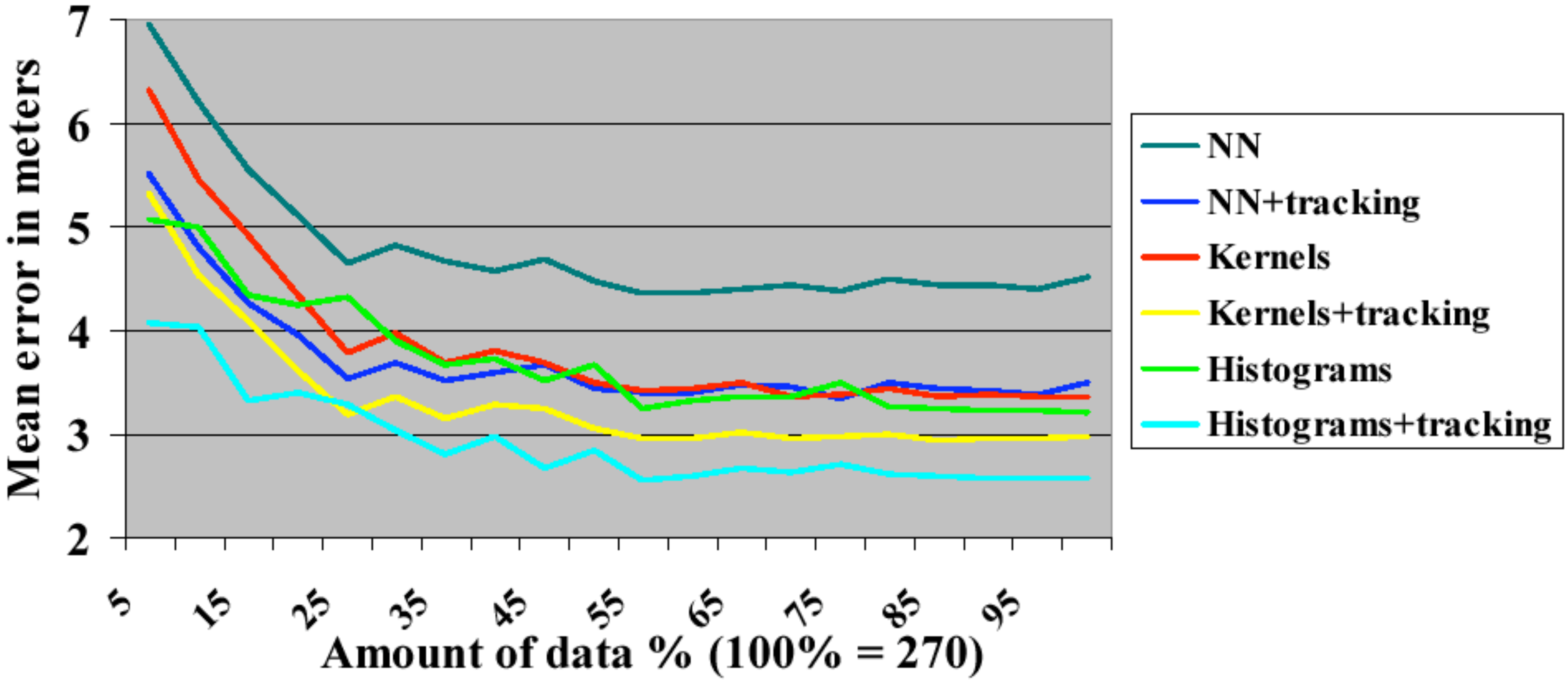
- **Use physical characteristics of signal strength propagation and build a model augmented with a wall attenuation factor**
- **Needs detailed (wall) map of the building; model portability needs to be determined**
  - [RADAR; INFOCOM 2000] based on [Rappaport 1992]

Krishnan et al. Results

Infocom 2004

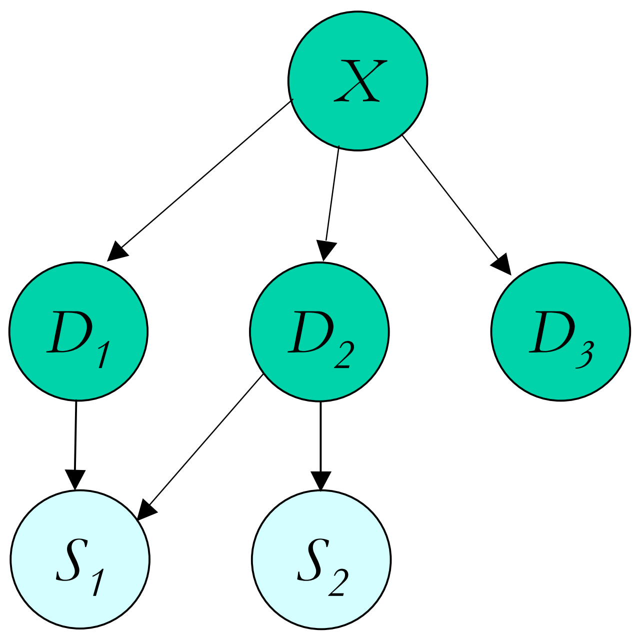
– Smoothed signal map per access point + nearest neighbor





- Best result: mean error 2.57 meters (90% below 4.52 meters) obtained with the probabilistic histogram method with tracking.
- Surprisingly robust with respect to the amount of training data.

Probabilistic Graphical Models



- Graphical model = picture of some conditional independence assumptions
- For example, D_1 is conditionally independent of D_3 given X

Markov Properties for Acyclic Directed Graphs

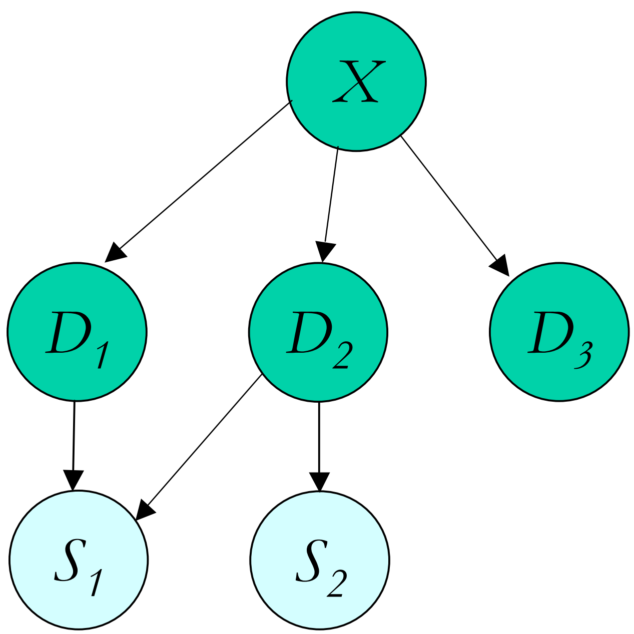
(Bayesian Networks)

(**Global**) S separates A from B in $G_{an(A,B,S)}{}^{m} \Rightarrow A \perp\!\!\!\perp B \mid S$

(**Local**) $\alpha \perp\!\!\!\perp nd(\alpha) \backslash pa(\alpha) \mid pa(\alpha)$

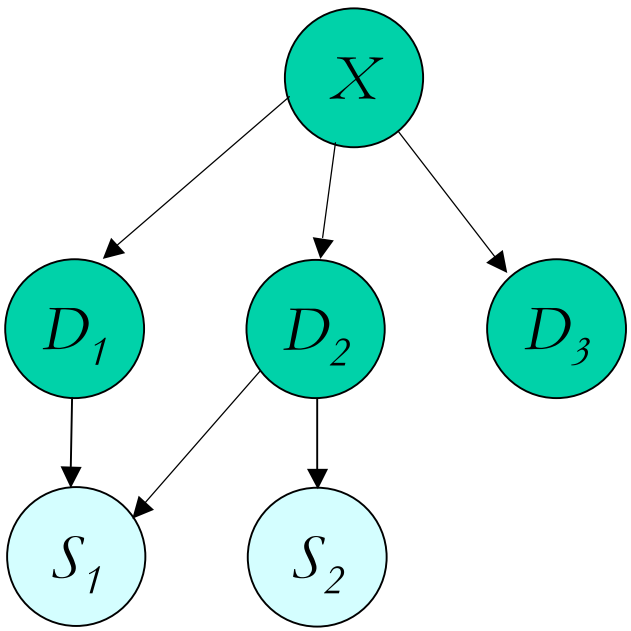
(**Factorization**) $f(x) = \prod f(x_v \mid x_{pa(v)})$

equivalent



$$\Leftrightarrow \quad p(X,D_1,D_2,D_3,S_1,S_2) = p(X)\, p(D_1 | X)\, p(D_2 | X)\, p(D_3 | X)\, p(S_1 | D_1,D_2)\, p(S_2 | D_2)$$

Monte Carlo Methods and Graphical Models



Simple Monte Carlo: Sample in turn from

$$p(X),\ p(D_1 \mid X),\ p(D_2 \mid X),\ p(D_3 \mid X),$$
$$p(S_1 \mid D_1, D_2), \text{ and } p(S_2 \mid D_2)$$

Gibbs Sampling: Sample in turn from

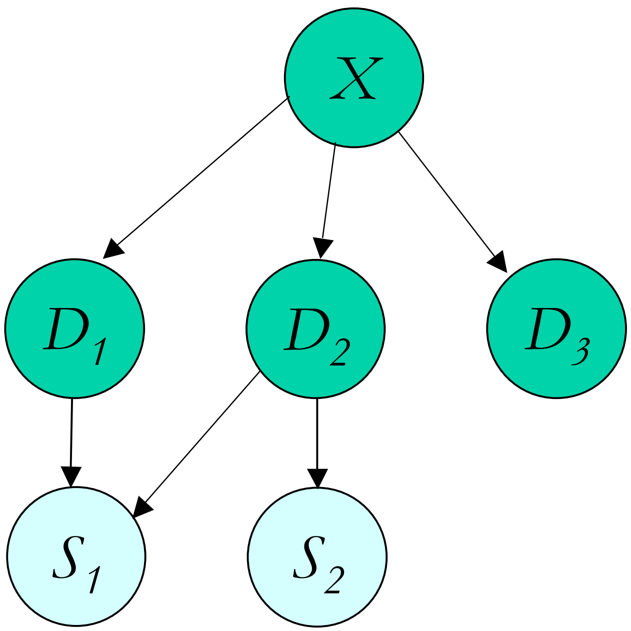
$$p(X \mid D_1, D_2, D_3, S_1, S_2)$$

$$p(D_1 \mid X, D_2, D_3, S_1, S_2)$$

$$\ldots$$

$$p(S_2 \mid X, D_1, D_2, D_3, S_1)$$

Full Conditionals from the Graphical Model



$$p(D_1 \mid X, D_2, D_3, S_1, S_2)$$

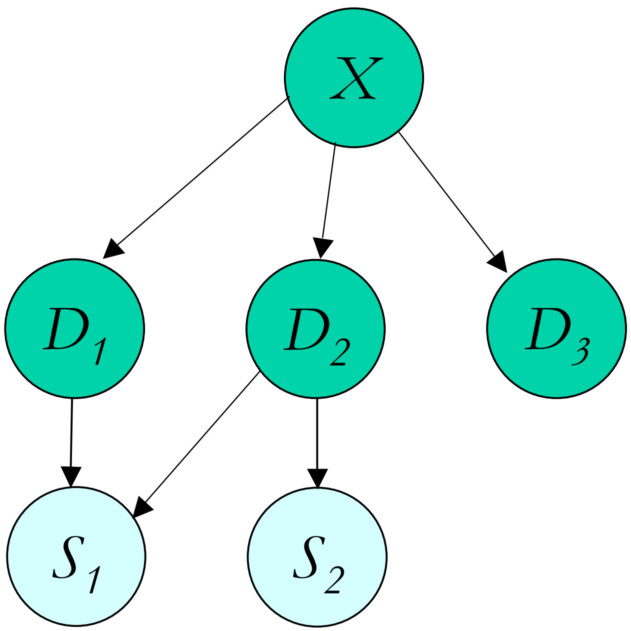
$$\propto \; p(X, D_1, D_2, D_3, S_1, S_2)$$

$$= \; p(X)\, p(D_1 | X)\, p(D_2 | X)\, p(D_3 | X)$$
$$p(S_1 | D_1, D_2)\, p(S_2 | D_2)$$

$$\propto \; p(D_1 | X)\, p(S_1 | D_1, D_2)$$

BUGS/WinBUGS automates this via adaptive rejection sampling and slice sampling

Full Conditionals from the Graphical Model



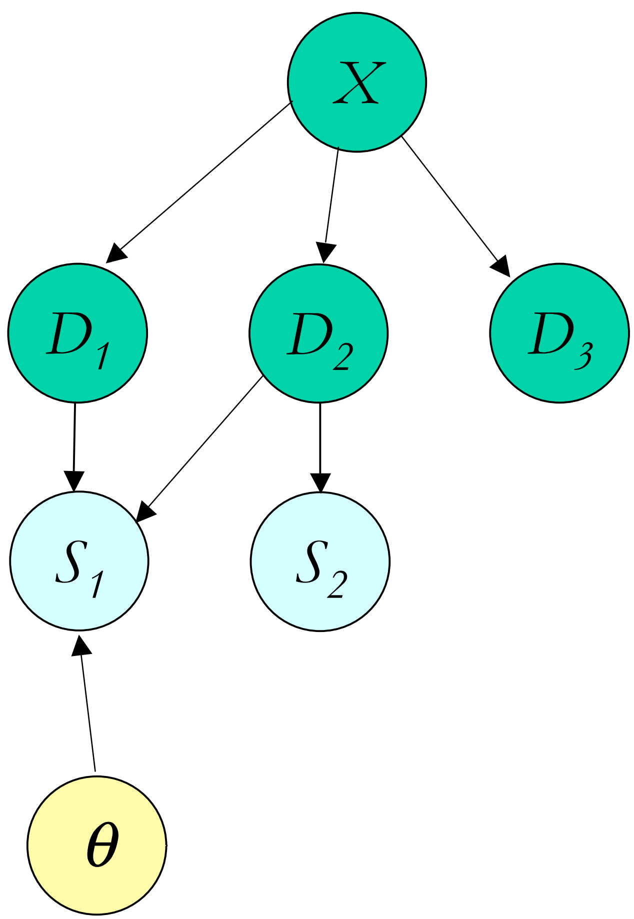
Incorporating Data, etc. Suppose the *D*'s were observed. Then sample from:

$$p(X \mid D_1, D_2, D_3, S_1, S_2)$$

$$p(S_1 \mid X, D_1, D_2, D_3, S_2)$$

$$p(S_2 \mid X, D_1, D_2, D_3, S_1)$$

Full Conditionals from the Graphical Model



Incorporating Data, etc. Suppose the *D*'s were observed. Then sample from:

$$p(X \mid D_1, D_2, D_3, S_1, S_2)$$

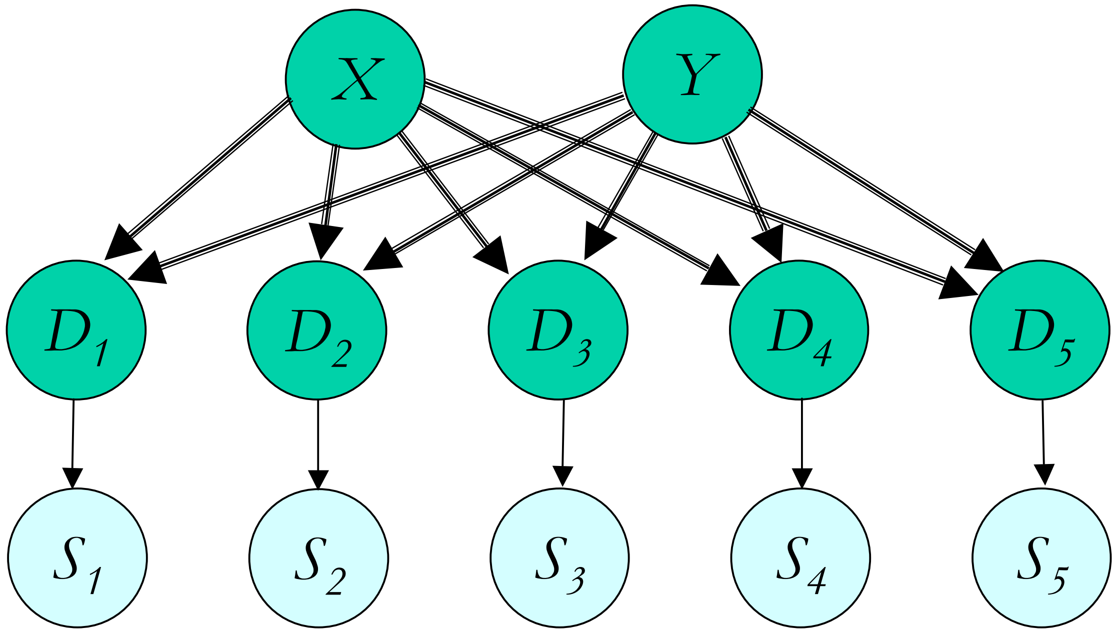
$$p(S_1 \mid X, D_1, D_2, D_3, S_2)$$

$$p(S_2 \mid X, D_1, D_2, D_3, S_1)$$

Bayesian Analysis. Treat "parameters" the same as everything else.

- BR has 5 APs, site dimension: 225 ft X 144 ft
  - 259 blue (corridor) data points taken earlier

Bayesian Graphical Model Approach

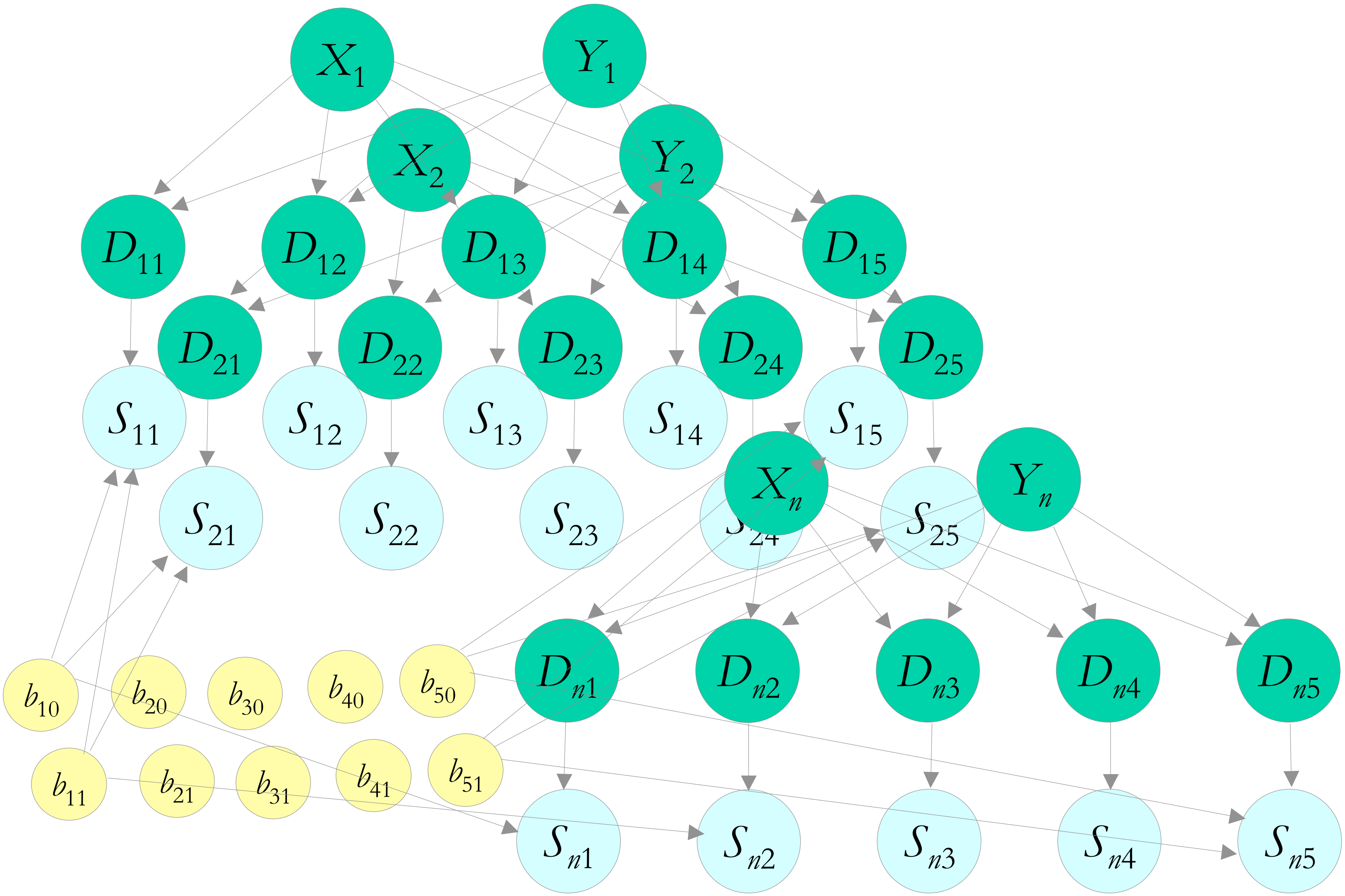


$$X, Y \sim unif$$

$$D_i(X, Y) = \text{distance to the ith access point}$$

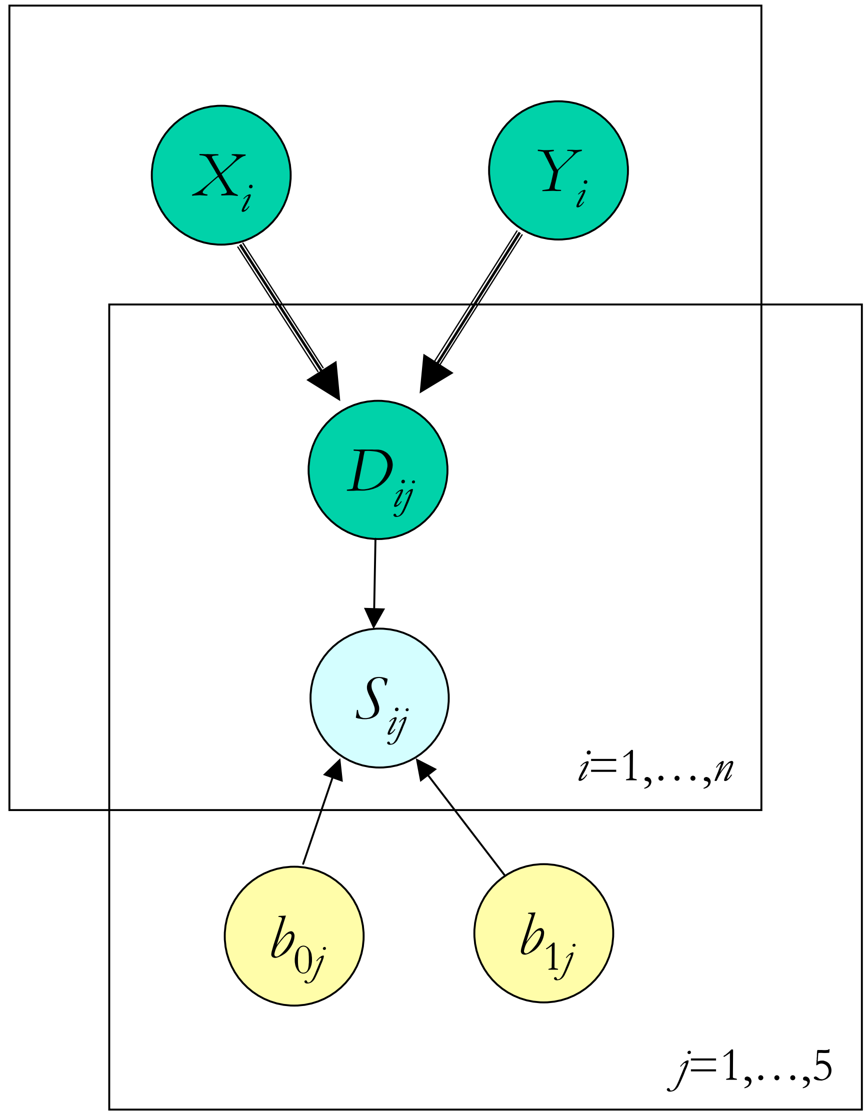
$$S_i \sim N(b_{i0} + b_{i1} \log D_i, \sigma_i^2),\ i = 1,\ldots,5$$

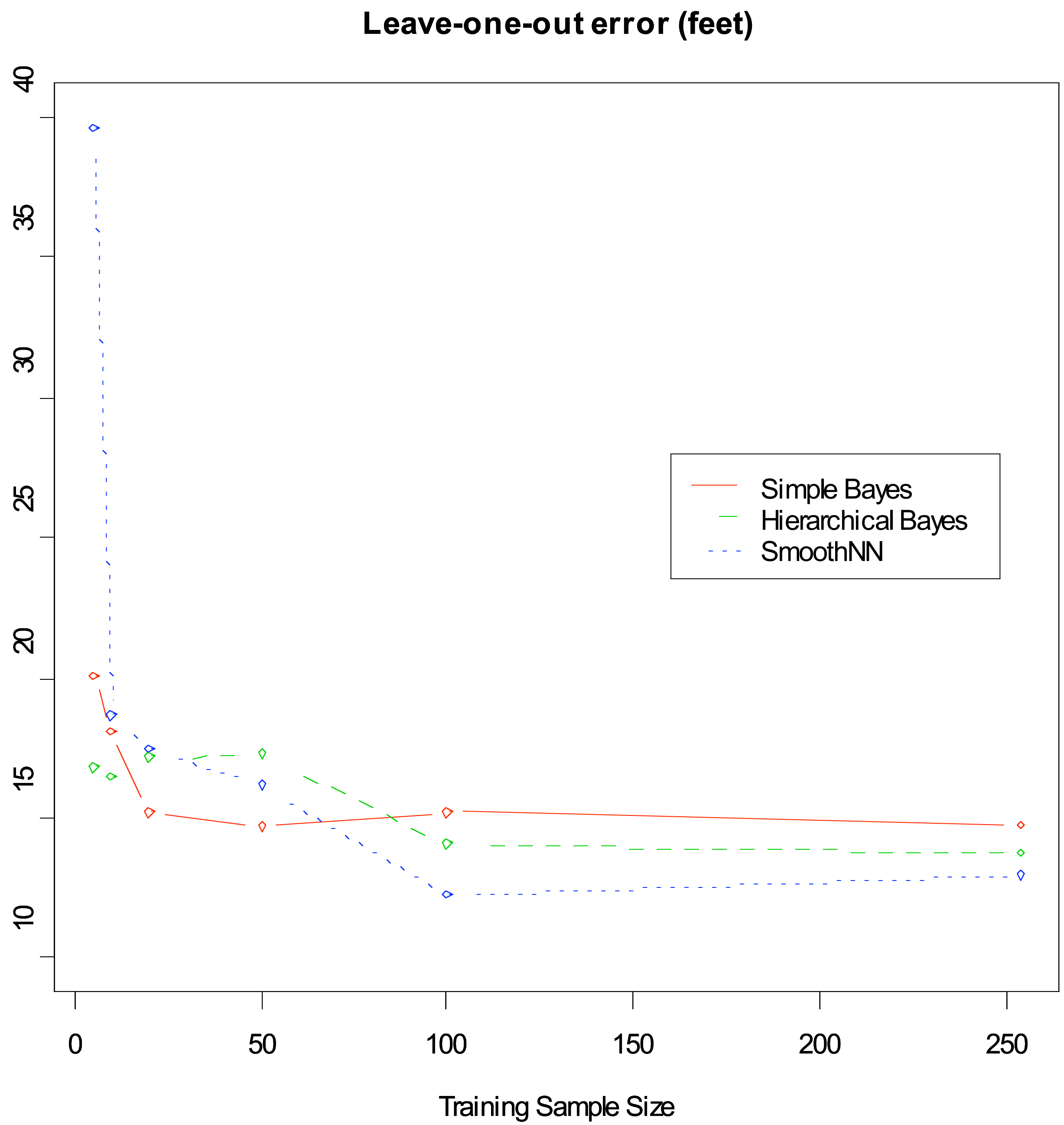
average

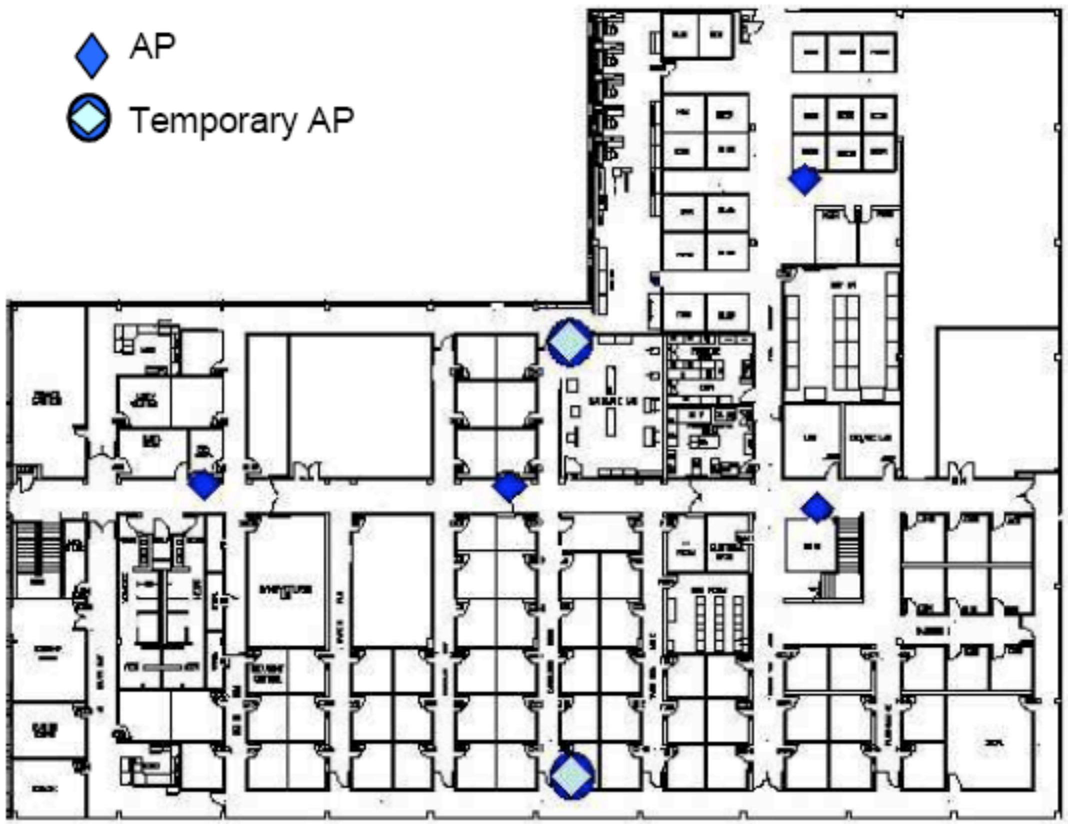


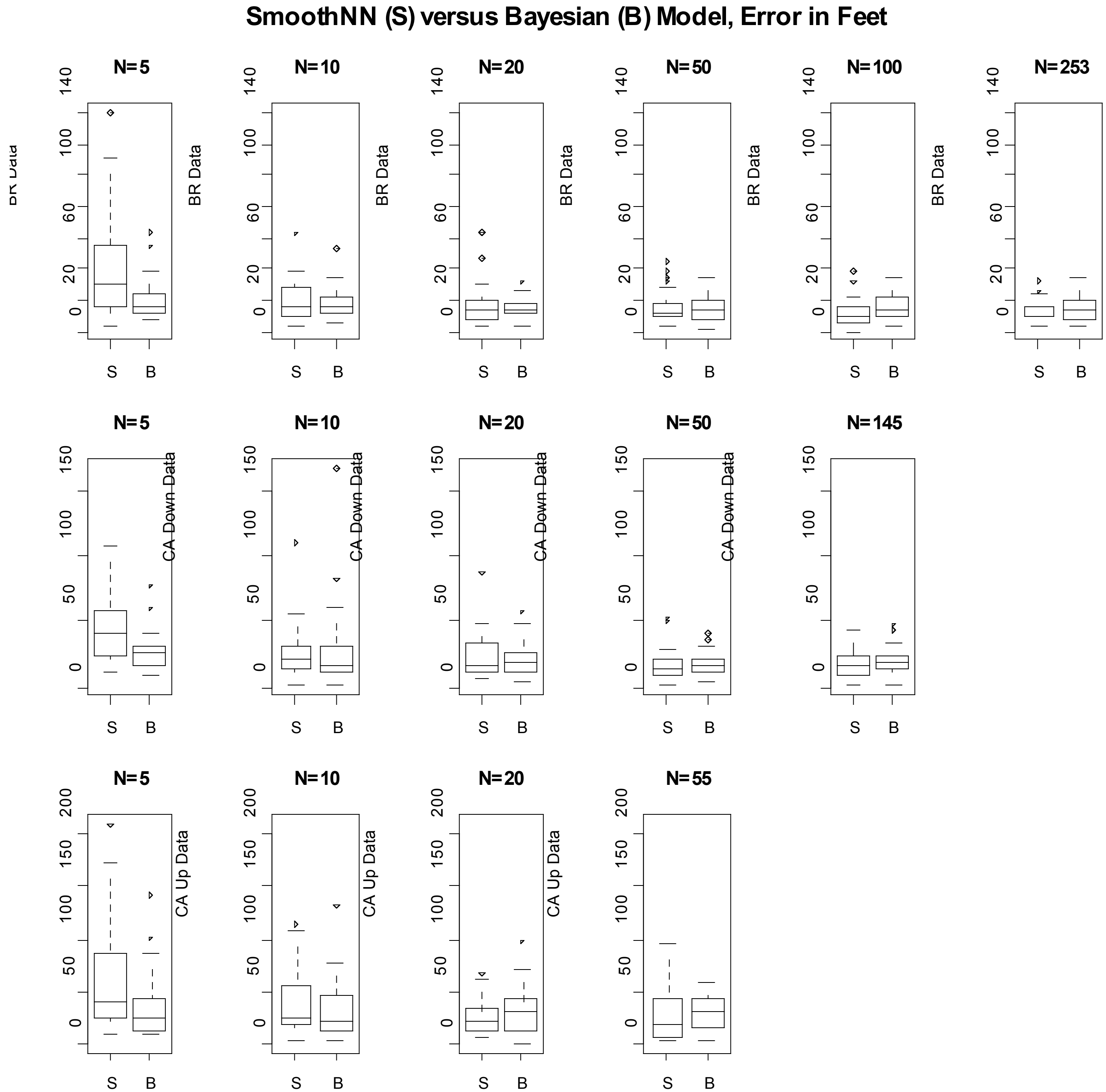
X_1
Y_1
X_2
Y_2
D_{11}
D_{12}
D_{13}
D_{14}
D_{15}
D_{21}
D_{22}
D_{23}
D_{24}
D_{25}
S_{11}
S_{12}
S_{13}
S_{14}
S_{15}
S_{21}
S_{22}
S_{23}
S_{24}
S_{25}
X_n
Y_n
b_{10}
b_{20}
b_{30}
b_{40}
b_{50}
b_{11}
b_{21}
b_{31}
b_{41}
b_{51}
D_{n1}
D_{n2}
D_{n3}
D_{n4}
D_{n5}
S_{n1}
S_{n2}
S_{n3}
S_{n4}
S_{n5}

Plate Notation

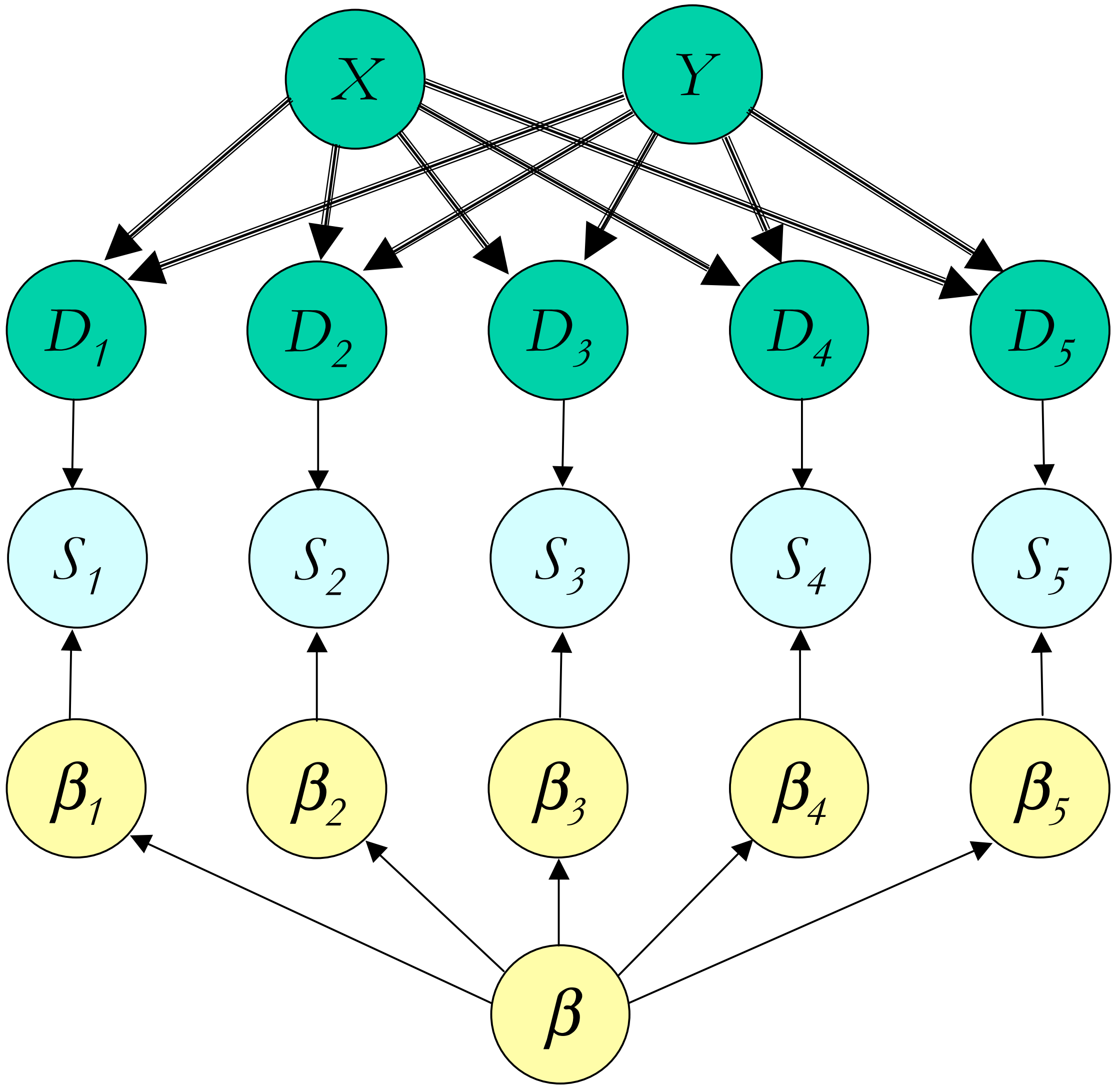



Leave-one-out error (feet)
40
35
30
25
20
15
10
0
50
100
150
200
250
Training Sample Size
Simple Bayes
Hierarchical Bayes
SmoothNN

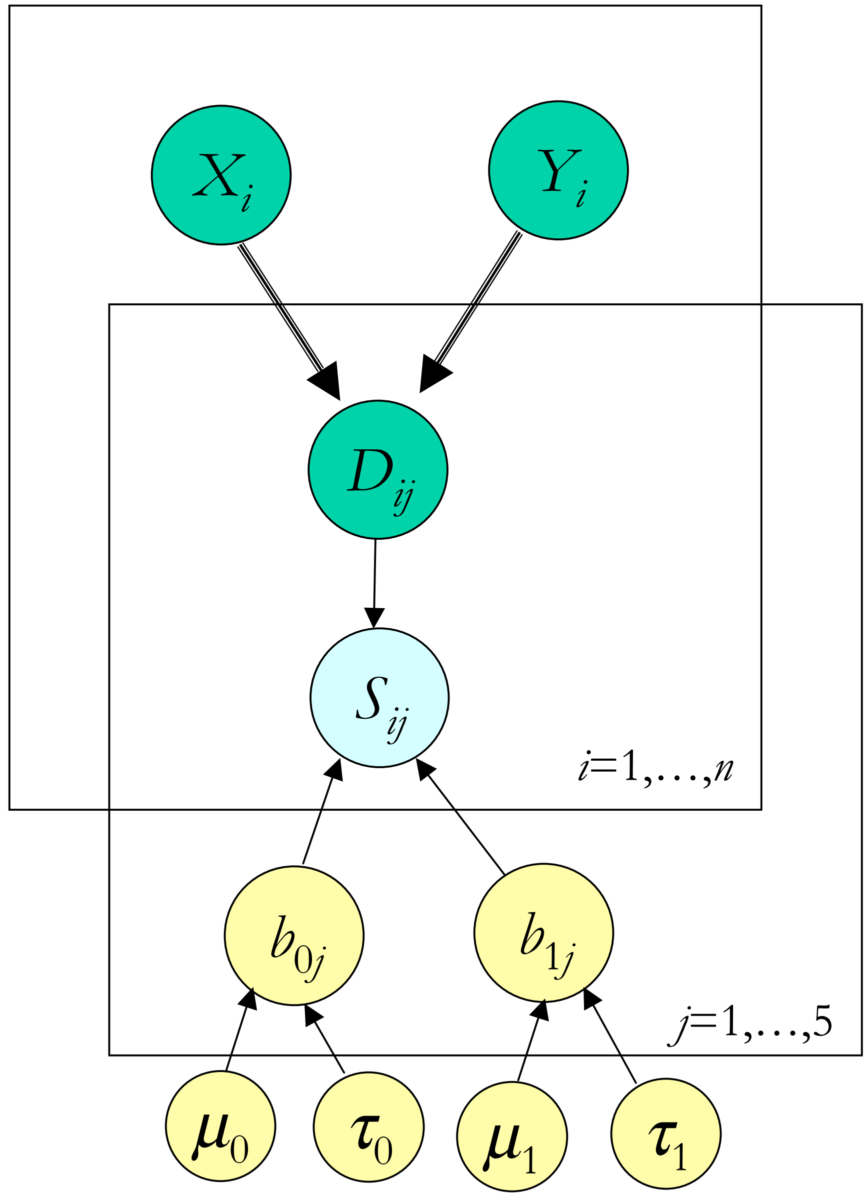

AP
Temporary AP


SmoothNN (S) versus Bayesian (B) Model, Error in Feet
N=5
N=10
N=20
N=50
N=100
N=253
N=145
N=55
BR Data
CA Down Data
CA Up Data
S
B

Hierarchical Model



Hierarchical Model

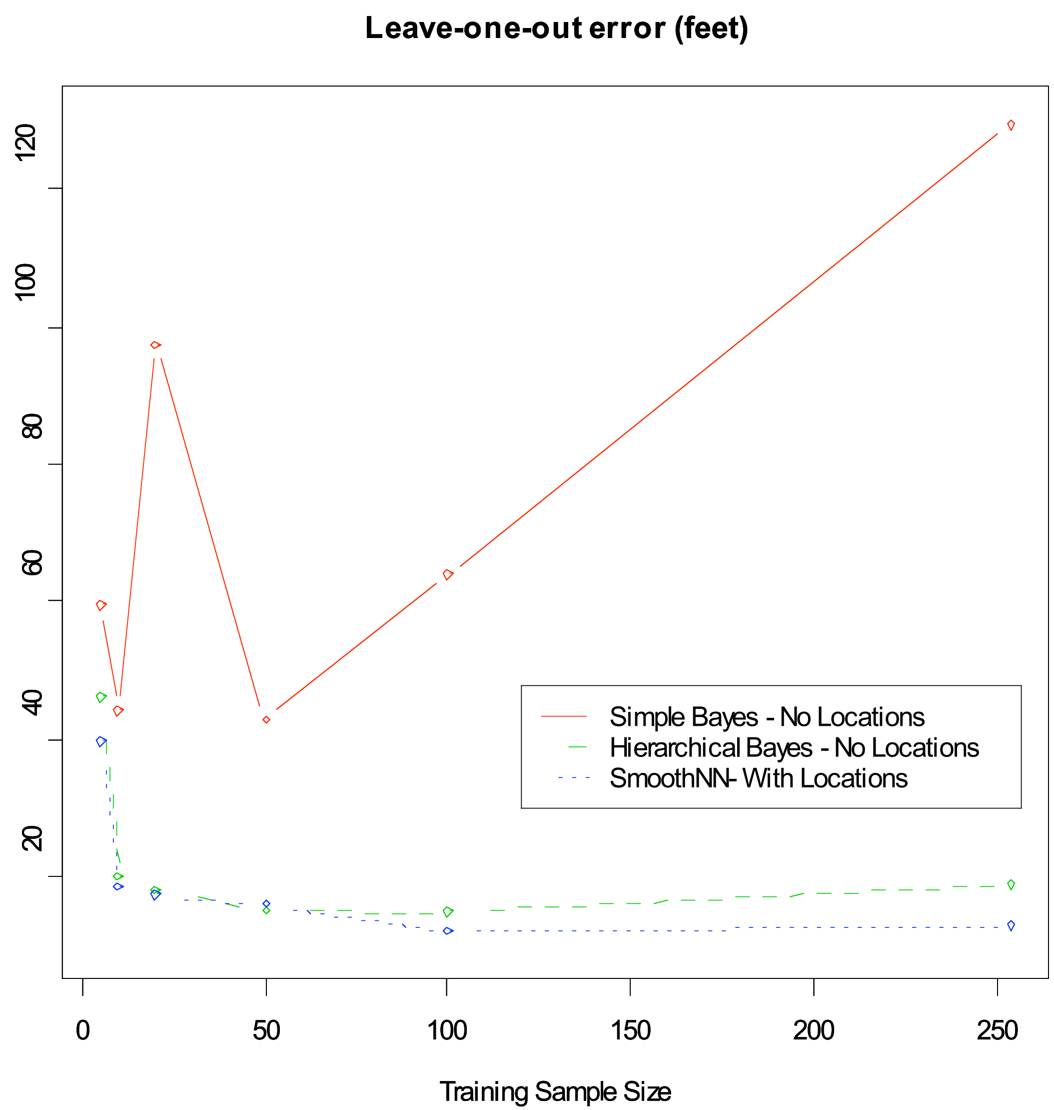


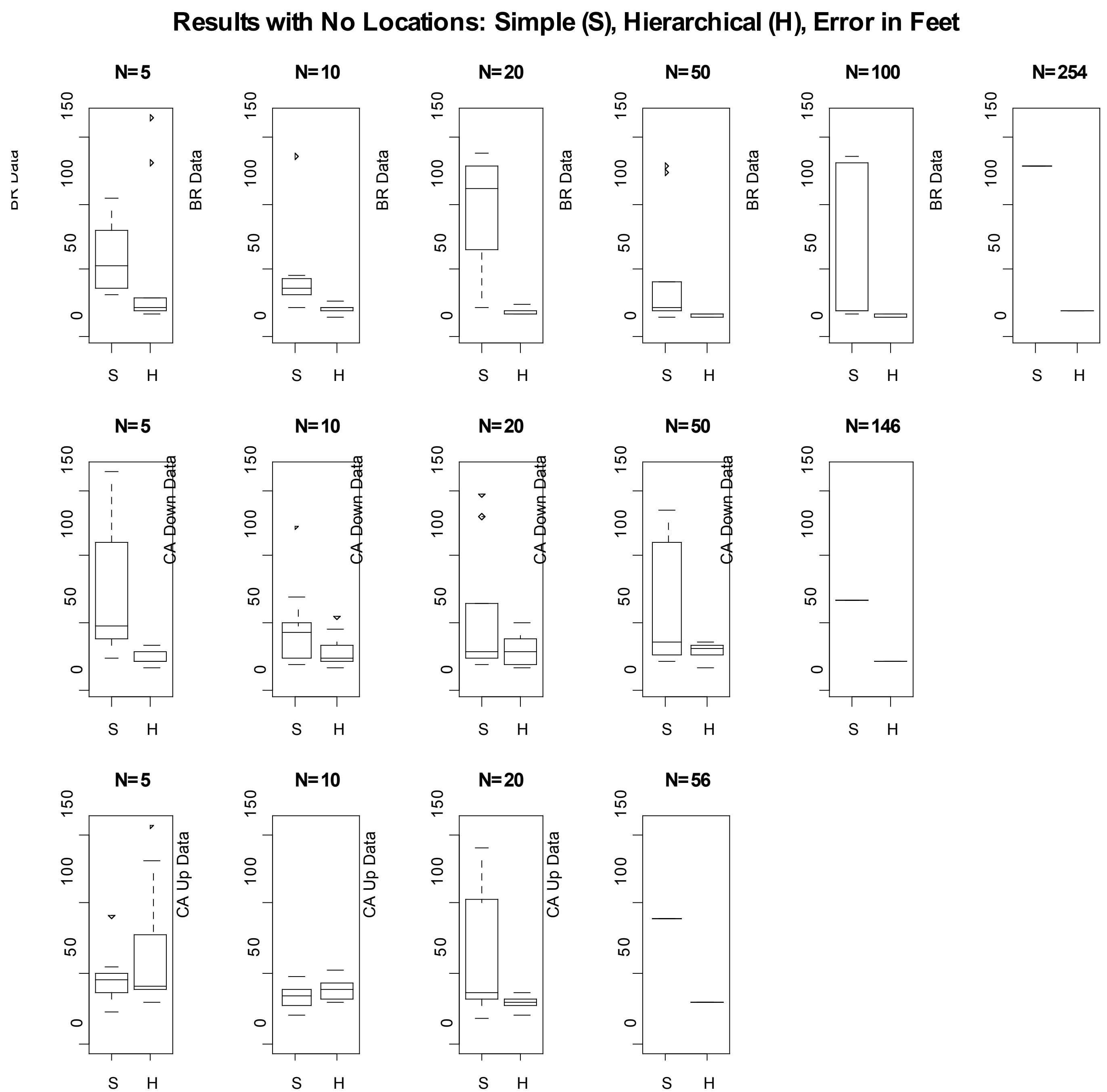
Simple Bayesian (B) versus Hierarchical Bayesian (H) Model, Error in Feet

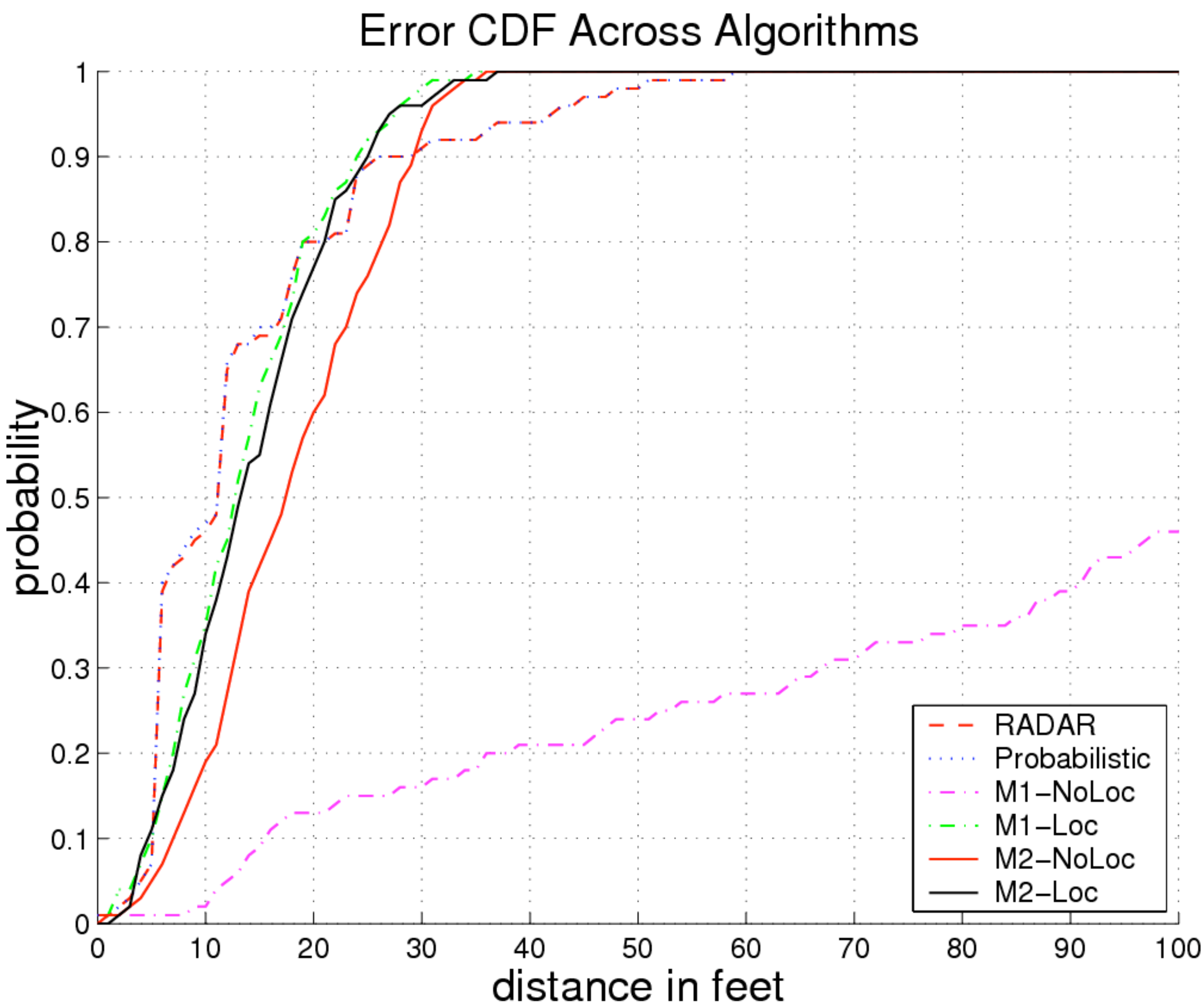
Pros and Cons

- **Bayesian model produces a predictive distribution for location**
- **MCMC can be slow**
- **Difficult to automate MCMC (convergence issues)**
- **Perl-WinBUGS (perl selects training and test data, writes the WinBUGS code, calls WinBUGS, parses the output file)**

What if we had no locations in the training data?




Results with No Locations: Simple (S), Hierarchical (H), Error in Feet
N=5
N=10
N=20
N=50
N=100
N=254
N=146
N=56
BR Data
CA Down Data
CA Up Data
S
H
0
50
100
150


Error CDF Across Algorithms
probability
distance in feet
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80
90
100
RADAR
Probabilistic
M1−NoLoc
M1−Loc
M2−NoLoc
M2−Loc

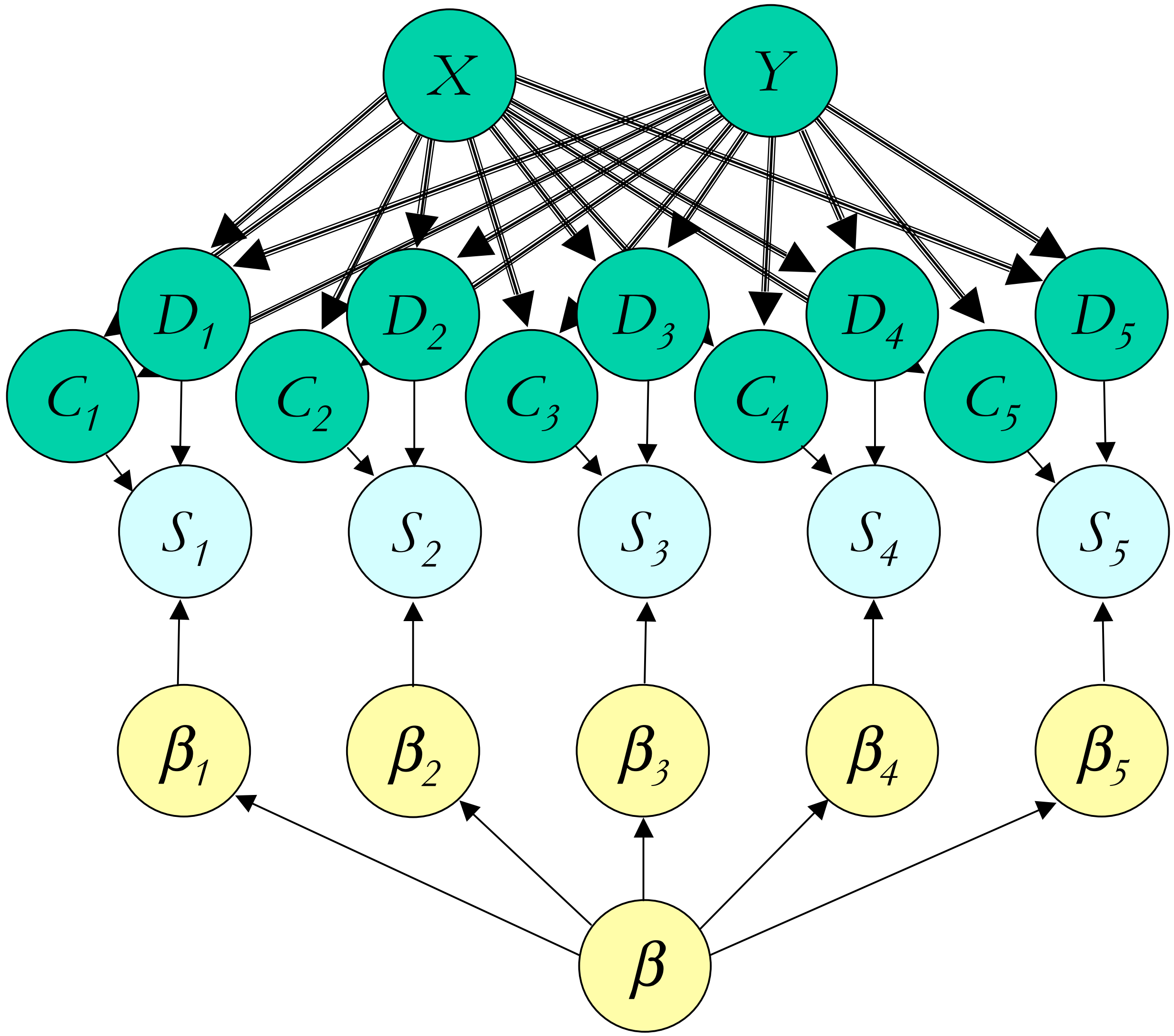
Zero Profiling?

- Simple sniffing devices can gather signal strength vectors from available WiFi devices
- Can do this repeatedly
- Locations of the Access Points

Why does this work?

- **Prior knowledge about distance-signal strength**
- **Prior knowledge that access points behave similarly**
- **Estimating several locations simultaneously**

Corridor Effects



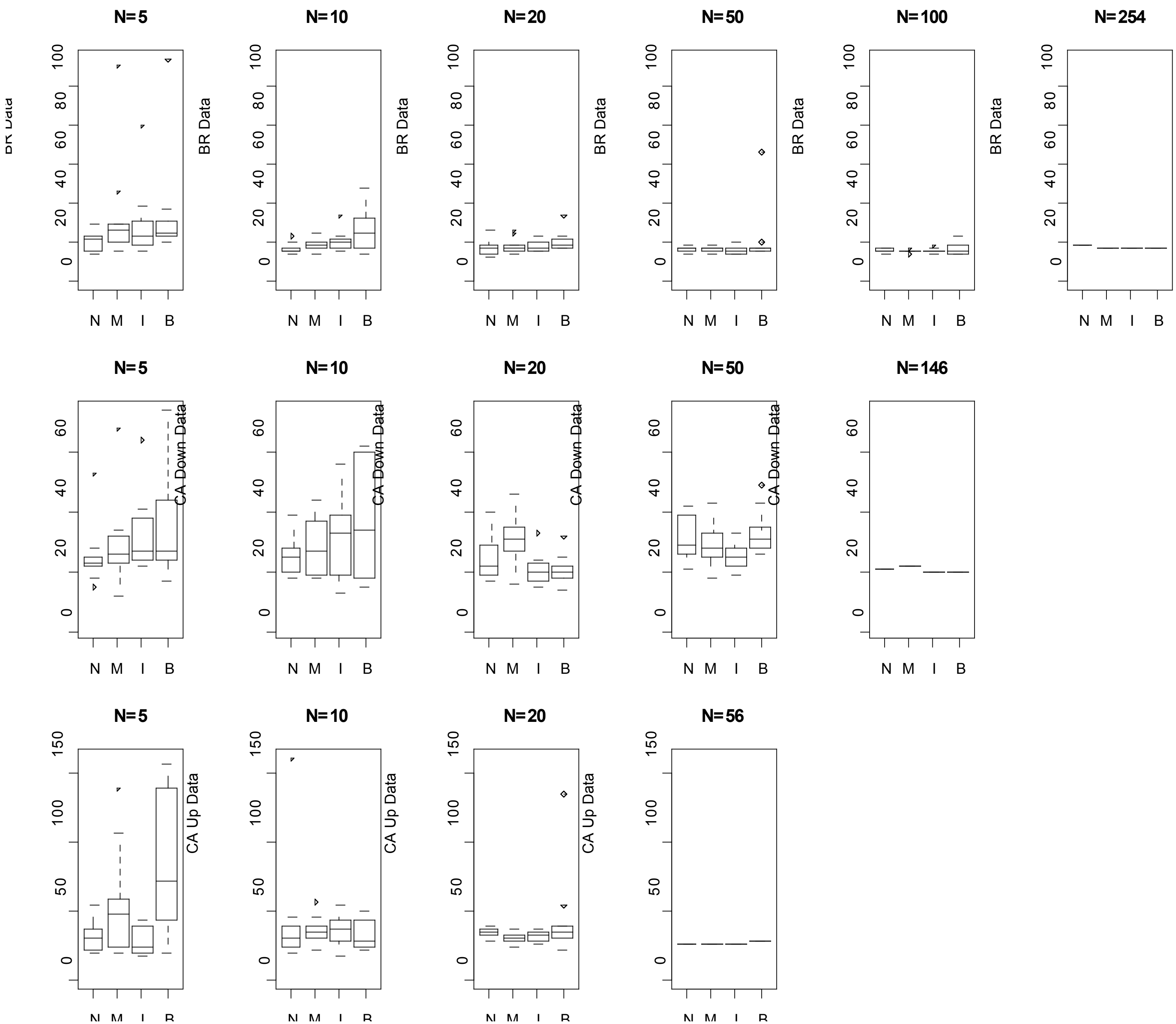
Results for N=20, no locations

corridor main effect	corridor -distance interaction	average error
0	0	20.8
0	1	16.7
1	0	17.8
1	1	17.3

with mildly informative prior on the distance main effect…

corridor main effect	corridor -distance interaction	average error
0	0	16.3
0	1	14.7
1	0	15.8
1	1	15.9

Corridor Effect: None (N), Main (M), Interaction (I), Both (B), Error in Feet



Discussion

- Informative priors
- Convenience and flexibility of the graphical modeling framework
- Censoring (30% of the signal strength measurements)
- Repeated measurements & normal error model
- Tracking

- Machine learning-style experimentation is clumsy with perl-WinBUGS

- BR has 5 APs, site dimension: 225 ft X 144 ft
  - 259 blue (corridor) data points taken earlier