Bayesian Indoor Positioning Systems

David Madigan[1,2], Eiman Elnahrawy[1], Richard P. Martin[1],
Wen-Hua Ju[2], P. Krishnan[2], and A.S. Krishnakumar[2]

Rutgers University[1], Piscataway, NJ 08840
Avaya Labs[2] Basking Ridge, NJ 07920

dmadigan@rutgers.edu,{rmartin,eiman}@cs.rutgers.edu,{whju,pk,ask}@avaya.com

***Abstract*— In this paper, we introduce a new approach to location estimation where, instead of locating a single client, we *simultaneously* locate a set of wireless clients. We present a Bayesian hierarchical model for indoor location estimation in wireless networks. We demonstrate that our model achieves accuracy that is similar to other published models and algorithms. By harnessing prior knowledge, our model eliminates the requirement for training data as compared with existing approaches, thereby introducing the notion of a fully adaptive *zero profiling* approach to location estimation.**

***Index Terms*— Experimentation with real networks/Testbed, Statistics, WLAN, localization, RSS/fingerprinting, Bayesian graphical models**

I. Introduction

The growth of wireless networking has generated commercial and research interest in statistical methods to track people and things. Inside stores, hospitals, warehouses, and factories, where Global Positioning System devices generally do not work, Indoor Positioning Systems (IPS) aim to provide location estimates for wireless devices such as laptop computers, handheld devices, and electronic badges. The proliferation of "Wi-Fi" (IEEE 802.11b) wireless Internet access in cafes, college campuses, airports, hotels, and homes has generated particular interest in indoor positioning systems that utilize physical attributes of Wi-Fi signals. Typical applications include tracking equipment and personnel in hospitals, providing location-specific information in supermarkets, museums, and libraries, and location-based access control.

In a standard Wi-Fi setup, one or more access points serve end-users. In what follows we focus on networks with multiple access points (typical of networks in office buildings or large public spaces). Wi-Fi location estimation can employ one or more of several physical attributes of the medium. Typical features include received signal strength (RSS) from the access points, the angle of arrival of the signal, and the time difference of arrival. Among these, RSS is the only feature that reasonably priced off-the-shelf hardware can currently measure. There exists a substantial literature on using RSS for location estimation in wireless networks - see, for example, [1], [13], [20], [29], [12]. Related websites include `www.ekahau.com`, `www.bluesoft-inc.com`, and `www.newburynetworks.com`. In a laboratory setting, RSS decays linearly with log distance and a simple triangulation using RSS from three access points could uniquely identify a location in a two-dimensional space. In practice, physical characteristics of a building such as walls, elevators, and furniture, as well as human activity add significant noise to RSS measurements. Consequently statistical approaches to location estimation prevail.

The standard approach uses supervised learning techniques. The training data comprise vectors of signal strengths, one for each of a collection of known locations. The dimension of each vector equals the number of access points. The corresponding location could be one-dimensional (e.g., location on a long airport corridor), two-dimensional (e.g., location on one floor of a museum), or three-dimensional (e.g. location within a multi-story office building). Collection of the location data is labor intensive requiring physical distance measurements with respect to a reference object such as a wall. The model building phase then learns a predictive model that maps signal strength vectors to locations. In the online phase, a wireless terminal is located by matching the terminal's RSS measurements against the model. Researchers have applied many supervised learning methods to this problem including nearest neighbor methods, support vector machines, and assorted probabilistic techniques.

Gathering extensive training data and the requisite physical measures of location ("profiling") involves a steep upfront cost and deployment effort ([22]). Furthermore, even in normal office environments, changing environmental, building, and occupancy conditions can affect signal propagation and require repeated data gathering to maintain predictive accuracy ([2]). Consequently, minimizing the number of training observations needed to adequately profile a particular site and having the model be adaptable are important objectives. Similarly we seek to minimize data requirements concerning internal wall materials, flooring, occupancy, etc. In [12] this issue was addressed by deploying additional hardware called stationary emitters at known locations to help rebuild the model periodically. Minimizing the number of stationary emitters was also shown to be equivalent to minimizing the amount of profiling needed. However, all techniques in the literature so far either require profiling or detailed maps of the building to build the signal strength model.

In this paper we address the location estimation problem using a novel approach. Instead of trying to locate a single terminal, our model tries to *simultaneously* locate a set of terminals. By appropriately exploiting signal strength information from a collection of terminals, we show that the location estimation for the entire set can be improved. The model is particularly relevant as the number of wireless terminals are increasing. We also propose and use for the first time hierarchical Bayesian graphical models [23], [5] for wireless location estimation. Our key finding is that a hierarchical Bayesian approach, incorporating physical knowledge about the nature of Wi-Fi signals, can provide accurate location estimates without *any* location information in the training data, leading to a truly adaptive, zero-profiling technique for location estimation.

Section II provides some additional background. Our approach uses probabilistic graphical models and Section III describes the Bayesian statistical framework that we use for location prediction. Sections IV and V describe the datasets we used for experimentation as well as various results. Section VI describes some potential future work.

II. Related Work

Location estimation techniques in wireless networks can be broadly classified based on the methods used to build models and methods used to search the models in the online phase. For building models, most techniques profile the entire site and collect one or more signal strength samples from all visible access points at each sample point. Each point is mapped to either a signal strength vector ([1], [13], [17], [29], [21] or a signal strength probability distribution ([3], [20], [26], [29], [30]). Such profiling techniques require considerable investment in data gathering. Alternatively, a parametric model that uses signal propagation physics and calculates signal degradation based on a detailed map of the building, the walls, obstructions and their construction material, has been proposed ([1]). Obtaining detailed maps of the building and its changes over time is, however, a hurdle that needs to be overcome for the use of this method. [12] propose an architecture that allows the model to be updated periodically and automatically, while also reducing (but not eliminating) the profiling needed.

Custom sensors have been used for location estimation in other interesting ways ([18], [27], [28]). In [27] and similar systems, infra-red (IR) wireless technology is used; IR technology has limited range and hence has not become very popular. In [18] a decentralized (client-based) approach using time difference of arrival between ultrasound and RF signals from custom sensors is used for location estimation. The system in [28] uses expensive custom RF-based hardware for location estimation. In contrast, our approach is easier to bootstrap, is based on RSS and can be built with off the shelf components, thus providing a cost-effective solution. Recent advances in sensor technology ([10]) and projected decreases in the manufacturing cost allow us to provide a cost-effective solution in our system.

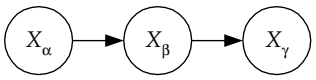

Fig. 1. *A simple acyclic directed graphical model.*

III. Bayesian Graphical Models

A *graphical model* is a multivariate statistical model embodying a set of conditional independence relationships. A graph displays the independence relationships. The vertices of the graph correspond to random variables and the edges encode the relationships. To date, most graphical models research has focused on acyclic digraphs, chordal undirected graphs, and chain graphs that allow both directed and undirected edges, but have no partially directed cycles ([14]).

Here we focus on acyclic digraphs (ADGs) with both continuous and categorical random variables. In an ADG, *all* the edges are directed and the graph represents them with arrows (see Figure 1). Each vertex in the graph corresponds to a random variable $X_v, v \in V$ taking values in a sample space $\mathcal{X}_v$. To simplify notation, we use v in place of X_v in what follows. In an ADG, the *parents* of a vertex v, pa(v), are those vertices from which vertices point into v. The *descendants* of a vertex v are the vertices which are reachable from v along a directed path. A vertex w is a *child* of v if there is an edge from v to w. The parents of v are taken to be the only direct influences on v, so that v is independent of its non-descendants given its parents. This property implies a factorization of the joint density of $X_v, v \in V$, which we denote by $p(V)$, given by

$$p(V) = \prod_{v \in V} p(v|\mathrm{pa}(v)). \tag{1}$$

Figure 1 shows a simple example. This directed graph represents the assumption that X_γ and X_α are conditionally independent given X_β. The joint density of the three variables factors accordingly,

$$p(X_\alpha, X_\beta, X_\gamma) = p(X_\alpha)p(X_\beta|X_\alpha)p(X_\gamma|X_\beta).$$

For graphical models where all the variables are discrete, [24] presented a Bayesian analysis and showed how independent Dirichlet prior distributions can be updated locally to form posterior distributions as data arrive. [8] provided corresponding closed-form expressions for complete-data likelihoods and posterior model probabilities. [15] described corresponding Bayesian model averaging procedures. In the Bayesian framework, model parameters are random variables and appear as vertices in the graph.

When some variables are discrete and others continuous, or when some of the variables are latent or have missing values, a closed-form Bayesian analysis generally does not exist. Analysis then requires either analytic approximations of some kind or simulation methods. Here we consider a Markov chain Monte Carlo (MCMC) simulation method. [23] provides a brief introduction to a particular MCMC algorithm, the univariate Gibbs sampler, for Bayesian graphical models as follows.

The Gibbs sampler starts with some initial values for each unknown quantity (that is, model parameters, missing values, and latent variables), and then cycles through the graph simulating each variable v in turn from its conditional probability distribution, given all the other quantities, denoted $V \backslash v$, fixed at their current values (known as the "full conditional"). The simulated v replaces the old value and the simulation shifts to the next quantity. After sufficient iterations of the procedure one assumes that the Markov chain has reached its stationary distribution, and then future simulated values for vertices of interest are monitored. Inferences concerning unknown quantities are then based on data analytic summaries of these monitored values, such as empirical medians and 95% intervals. Some delicate issues do arise with the Gibbs sampler such as assessment of convergence, sampling routines, etc. [7] provide a full discussion.

The crucial connection between directed graphical models and Gibbs sampling lies in expression (1). The full conditional distribution for any vertex v is equal to:

$$\begin{aligned} p(v|V\backslash v) &\propto p(v, V\backslash v) \\ &\propto \text{terms in } p(V) \text{ containing } v \\ &= p(v|\mathrm{pa}(v)) \prod_{w \in \mathrm{child}(v)} p(w|\mathrm{pa}(w)), \end{aligned}$$

i.e., a prior term and a set of likelihood terms, one for each child of v. Thus, when sampling from the full conditional for v, we need only consider vertices which are parents, children, or parents of children of v, and we can perform local computations. The BUGS language and software [25] implements a version of the Gibbs sampler for Bayesian graphical models. We utilized BUGS for the experiments we report below.

IV. Datasets

We collected RSS data from three floors at two sites, referred to in this paper as BR, CA Up and CA Down. Both the BR and CA sites are office buildings and have deployed 802.11b wireless networks. Figure 2 shows the floor plans for the two sites.

To make our RSS measurements, we used a Linux IPAQ with a modified driver updated to scan for access points. The IPAQ had a custom client and a standard Konqueror web browser. The user making RSS measurements clicked on their current location in an image of the floor as displayed on the browser. The posting of this information triggered an RSS measurement request at the client from the web server on a separate TCP channel. The web server then recorded the coordinate and RSS vector information at that location. We did not specifically orient the IPAQ in any way while taking measurements.

The BR site has 5 access points and measures 225 ft X 144 ft. We made 254 RSS measurements at this site. The measurements were made over different sessions spanning several days.

The CA Down floor has 4 access points, three of which are collinear, and measures 250 ft X 175 ft, with a "slice" removed. Due to the collinearity of the three access points,

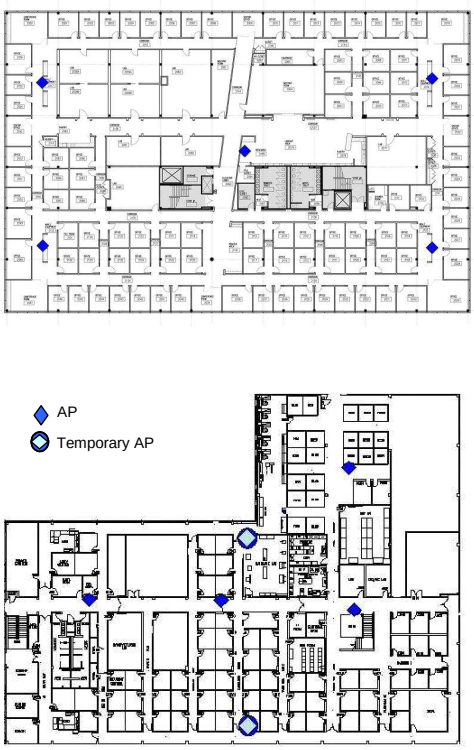

Fig. 2. *Floor plans for the BR and CA sites showing the access points (APs).*

we installed two temporary access points. The CA Up floor has 4 access points. At the CA site, a colleague took 146 measurements on the "Down" floor and 56 measurements on the "Up" floor.

V. Models and Experiments

Our goal is to construct a model that embodies extant knowledge about Wi-Fi signals as well as physical constraints implied the target building. We present a series of models of increasing complexity, in each case showing results with varying training dataset sizes. We focus throughout on predictive accuracy.

A. A Non-Hierarchical Bayesian Graphical Model

Figure 3 shows a particular graphical model for a two-dimensional location estimation problem in a building with four access points. In what follows we refer to this model as M_1 (although the number of access points varies).

The vertices X and Y represent location. The vertex D_1 (respectively D_2, D_3, and D_4) represents the Euclidean distance between the location specified by X and Y and the first (respectively second, third, and fourth) access point. Since we assume the locations of the access points are known, the D_i's are deterministic functions of X and Y. The vertex S_i represents the signal strength measured at (X, Y) with respect to the ith access point, $i = 1, \ldots, 4$. The model assumes that X and Y are marginally independent.

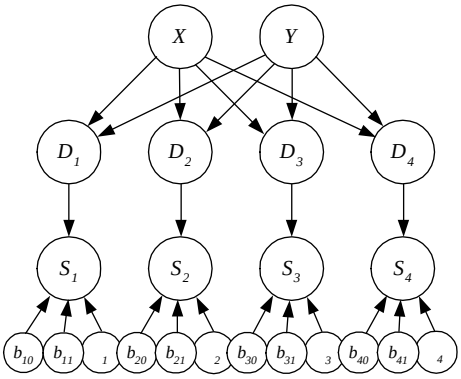

Fig. 3. *A Bayesian graphical model for location estimation. This is model* M_1.

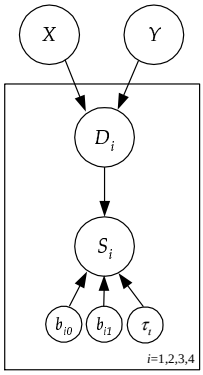

Fig. 4. *A Bayesian graphical model using plate notation. This is also model* M_1.

Specification of the model requires a conditional density for each vertex given its parents as follows:

$$\begin{aligned} X &\sim \text{uniform}(0, L), \\ Y &\sim \text{uniform}(0, B), \\ S_i &\sim N(b_{i0} + b_{i1} \log D_i, \tau_i), i = 1, 2, 3, 4, \\ b_{i0} &\sim N(0, 0.001), i = 1, 2, 3, 4, \\ b_{i1} &\sim N(0, 0.001), i = 1, 2, 3, 4. \end{aligned}$$

Here L and B denote the length and breadth of the building respectively. The distributions for X and Y reflect the physical constraints of the building. The model for S_i reflects the fact that signal strength, decays approximately linearly with log distance. Note that we use $N(\mu, \tau)$ to denote a Gaussian distribution with mean μ and precision τ so that the prior distributions for b_{i0} and b_{i1} have large variance.

Figure 4 shows a more compact representation for M_1 using the BUGS plate notation for replicated sub-models and with d denoting the number of access points.

Figure 5 shows the predictive performance of model M_1 on the BR data, as a function of training set size. Specifically, for each training set size N, we plot the average performance for 30 replications of a random test-training split, using N observations for training and

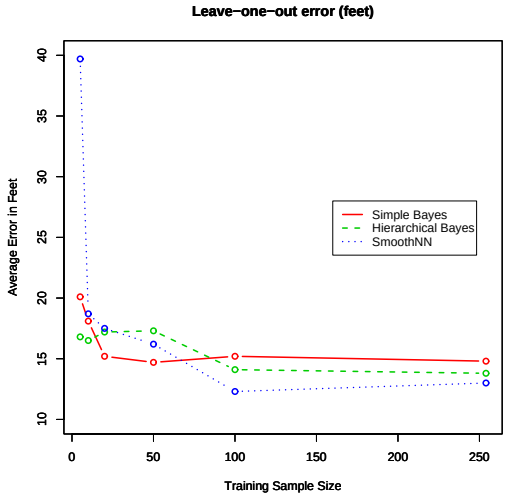

Fig. 5. *Average predictive accuracy of the non-hierarchical Bayesian graphical model* M_1*, the hierarchical model* M_2*, and the SmoothNN model on the BR data.*

one observation for testing. The solid curve shows the results for M_1 (i.e., the model of Figure 4). In each case, and throughout the paper, the estimates resulted from 110,000 MCMC iterations, discarding the first 10,000. This seemed to provide adequate convergence in most cases, according to standard BUGS diagnostics. We return to this issue at the end of the paper. For comparison purposes, the dotted curve shows the equivalent results for the smoothed nearest-neighbor "SmoothNN" model of [12]. The SmoothNN model proved highly competitive in comparison with two other benchmark systems and hence we use it for comparison purposes in this paper. The Figure 5 shows that M_1 outperforms the SmoothNN model with smaller training sample sizes but underperforms the SmoothNN model at the larger sample sizes.

Figure 6 provides more detail and also shows results for the other two datasets. The results for the three different datasets are qualitatively similar. Table I provides corresponding summary statistics. Note that predictive accuracy does tend to improve with training sample size, although not in every case.

B. A Hierarchical Bayesian Graphical Model

Next we seek to incorporate the knowledge that the coefficients of the linear regression models corresponding to each of the access points should be similar since the similar physical processes are in play at each access point. Physical differences between locations of the different access points will tend to mitigate the similarity but nonetheless, borrowing strength across the different regression models might provide some predictive benefits.

Figure 7 shows the hierarchical model M_2. The conditional densities for this model are:

$$\begin{aligned} X &\sim \text{uniform}(0, L), \\ Y &\sim \text{uniform}(0, B), \end{aligned}$$

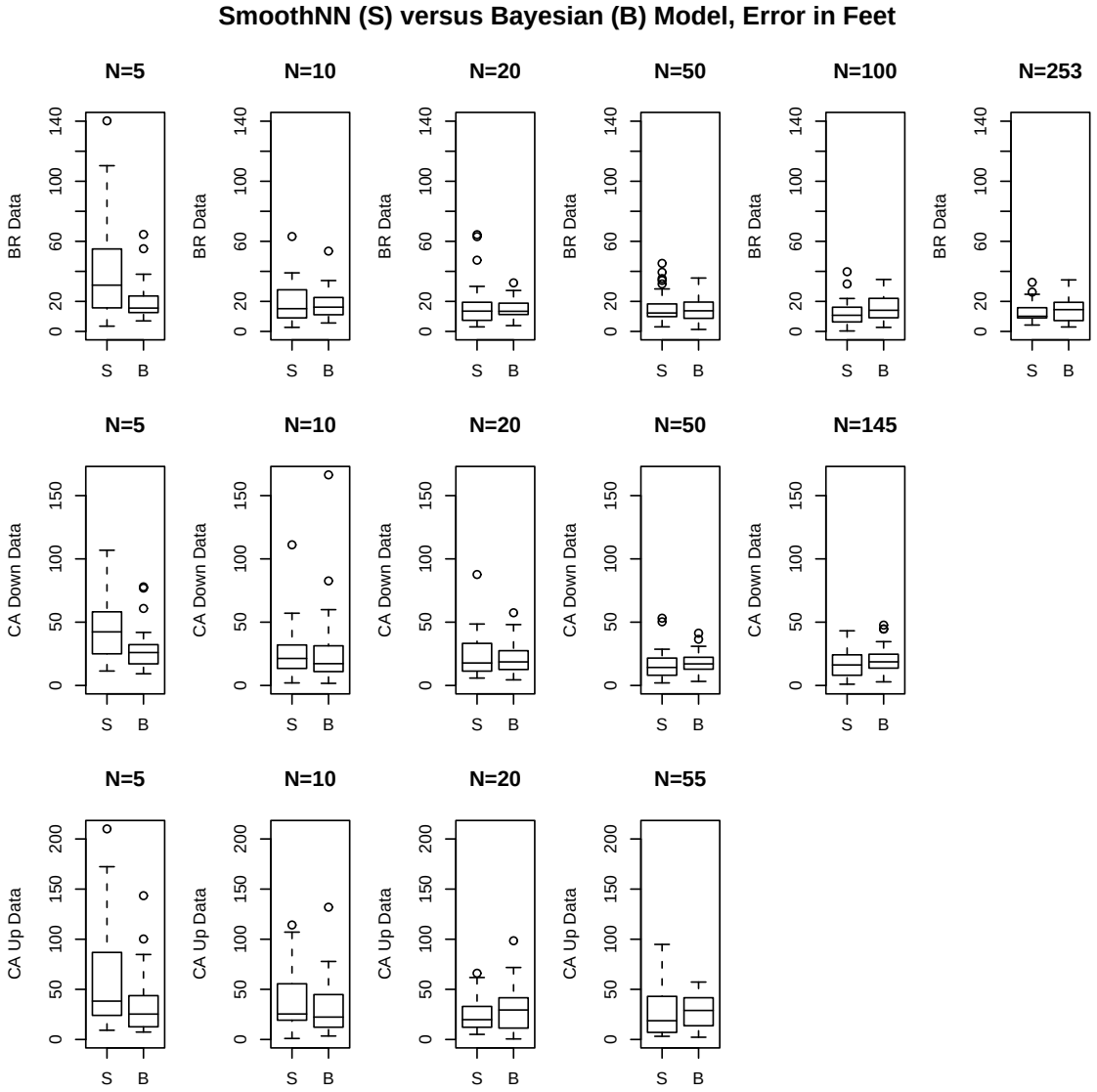

Fig. 6. *Predictive accuracy of the SmoothNN model versus the non-hierarchical Bayesian graphical model (M_1) for the BR data, CA Down data, and CA Up Data. The boxplots show 75th, 50th, and 25th percentiles and tails that show either the maximum and the minimum or shorter tails and individual extreme points. The vertical axes show error in feet.*

$$
\begin{aligned}
S_i &\sim N(b_{i0} + b_{i1} \log D_i, \tau_i), i = 1, \ldots, d, \\
b_{i0} &\sim N(b_0, \tau_{b0}), i = 1, \ldots, d, \\
b_{i1} &\sim N(b_1, \tau_{b1}), i = 1, \ldots, d, \\
b_0 &\sim N(0, 0.001), \\
b_1 &\sim N(0, 0.001), \\
\tau_{b0} &\sim \text{Gamma}(0.001, 0.001) \\
\tau_{b1} &\sim \text{Gamma}(0.001, 0.001)
\end{aligned}
$$

The dashed curve in Figure 5 shows the predictive accuracy of M_2 on the BR data. A comparison of M_1 and M_2 shows that the hierarchical model performs similarly to its non-hierarchical counterpart, although M_2 does provide improvement in average error for the smallest training sample size of 5.

Table II provides summary statistics for all three datasets. Again, the results for the different datasets are qualitatively similar. In general, the results show small differences between the non-hierarchical model M_1 and the hierarchical model M_2.

C. Training Data With No Location Information

Model M_2 incorporates two sources of prior knowledge. First, M_2 embodies the knowledge that signal strength decays approximately linearly with log distance. Second, the hierarchical portion of M_2 reflects prior knowledge that the different access points behave similarly. Here we pursue the idea that perhaps this prior knowledge provides sufficient constraints to obviate the need to know the actual

BR Data						
	Training Sample Size					
Model	5	10	20	50	100	253
M_1	20.1	18.1	15.2	14.7	15.2	14.8
SmoothNN	39.7	18.7	17.5	16.2	12.3	13.0
CA Down Data						
	Training Sample Size					
Model	5	10	20	50	145	
M_1	28.8	27.4	21.6	18.2	19.9	
SmoothNN	46.3	26.7	24.3	17.1	17.4	
CA Up Data						
	Training Sample Size					
Model	5	10	20	55		
M_1	35.4	31.7	30.5	28.5		
SmoothNN	59.9	36.3	25.2	28.2		

TABLE I
LEAVE-ONE-OUT AVERAGE ACCURACY IN FEET FOR THE BR, CA DOWN, AND CA UP DATA. RESULTS ARE AVERAGED OVER 30 REPLICATIONS. THE CORRESPONDING STANDARD ERRORS RANGE FROM ABOUT 1.5 TO 5.7.

BR Data						
	Training Sample Size					
Model	5	10	20	50	100	253
M_1	20.1	18.1	15.2	14.7	15.2	14.8
M_2	16.8	16.5	17.2	17.3	14.1	13.8
CA Down Data						
	Training Sample Size					
Model	5	10	20	50	145	
M_1	28.8	27.4	21.6	18.2	19.9	
M_2	21.3	26.3	25.0	20.3	18.7	
CA Up Data						
	Training Sample Size					
Model	5	10	20	55		
M_1	35.4	31.7	30.5	28.5		
M_2	30.6	37.9	33.0	33.5		

TABLE II
LEAVE-ONE-OUT AVERAGE ACCURACY IN FEET FOR THE BR, CA DOWN, AND CA UP DATA. RESULTS ARE AVERAGED OVER 30 REPLICATIONS. THE CORRESPONDING STANDARD ERRORS RANGE FROM ABOUT 1.3 TO 7.5.

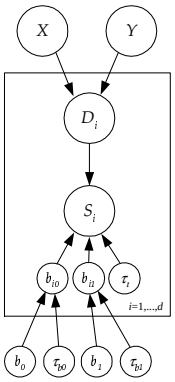

Fig. 7. *A Bayesian hierarchical graphical model using plate notation. This is model M_2.*

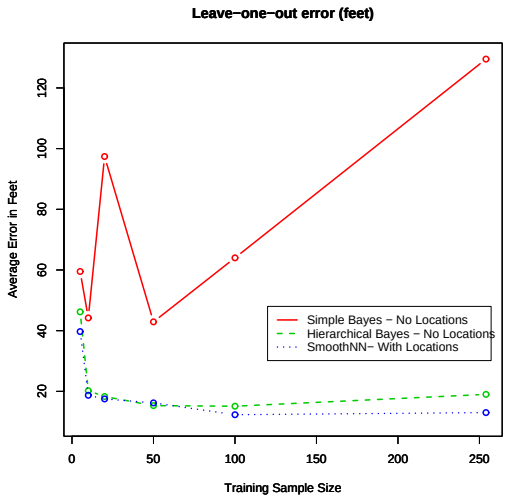

Fig. 8. *Average predictive accuracy of the non-hierarchical Bayesian graphical model on the BR data with no location data.*

locations of the training data observations. Specifically, the training data now comprise vectors of signal strengths with *unknown* locations; X and Y in M_1 and M_2 become latent variables.

Figure 8 shows the average predictive performance for the BR data with different numbers of randomly signal strength vectors, simultaneously estimating the locations of these signal strength vectors. In each case the results shows averages over 30 replications, except for the maximal case (254 for BR, 146 for CA Down, 56 for CA Up) which uses all the signal strength vectors in the training data. The solid curve corresponds to M_1 and the dashed curve to M_2. The results for the SmoothNN model are reproduced from Figure 5 and reflect training data *with known locations*. These results show some striking features. With no location information, M_1 performs poorly and shows no improvement with increasing numbers of signal strength vectors. Model M_2, however, from about 10 training vectors onwards, performs almost as well as the SmoothNN model trained on data with complete location information for each signal strength vector!

Figure 9 provides shows more detail and also shows results for the other two datasets. Once again, the results for the three different datasets are qualitatively similar. Table III provides corresponding summary statistics. In each case the hierarchical model, even with no location information, provides predictive performance that is close to, although not as good as, the state-of-the-art SmoothNN model.

Dropping the location data requirement affords significant practical benefits. As discussed in Section I, the location measurement process is slow and human-intensive. By contrast, gathering signal strengths vectors without the corresponding locations does not require human intervention; suitably instrumented access points or sniffing devices

BR Data						
	Training Sample Size					
Model	5	10	20	50	100	253
M_1, No Loc.	59.5	44.2	97.4	42.9	64.0	129.5
M_2, No Loc.	46.2	20.2	18.3	15.3	15.1	19.0
SmoothNN,w/Loc.	39.7	18.7	17.5	16.2	12.3	13.0

CA Down Data						
	Training Sample Size					
Model	5	10	20	50	145	
M_1, No Loc.	66.6	46.9	52.5	58.1	67.5	
M_2, No Loc.	23.9	29.4	29.2	29.8	21.9	
SmoothNN,w/Loc.	46.3	26.7	24.3	17.1	17.4	

CA Up Data						
	Training Sample Size					
Model	5	10	20	55		
M_1, No Loc.	46.4	33.2	59.1	89.0		
M_2, No Loc.	63.6	38.1	28.9	30.6		
SmoothNN,w/Loc.	59.9	36.3	25.2	28.2		

TABLE III
AVERAGE ACCURACY IN FEET FOR THE BR, CA DOWN, AND CA UP DATA. NO LOCATION INFORMATION IN THE TRAINING DATA. RESULTS ARE AVERAGED OVER 10 REPLICATIONS.

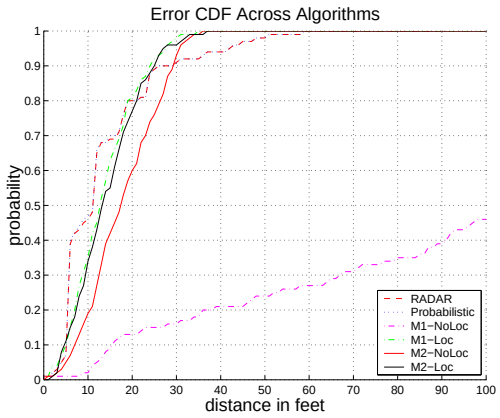

Fig. 10. *Error CDF across the various algorithms.*

can solicit signal strength measurements from existing Wi-Fi devices and can do this repeatedly at essentially no cost. We conjecture that typical Wi-Fi environments will feature a substantial number of scattered wireless devices. Repeated polling of these devices coupled with the Bayesian hierarchical model will enable fully adaptive location estimation that responds to changes in the Wi-Fi environment.

We note the existing location estimation algorithms that we are aware of all require location information in the training data to produce *any* estimates. The approach we have outlined does require the location of the actual access points (although recent experimentation shows that accurate location estimate does not require the location of all the access points).

D. Comparison To Previous Work

Figure 10 shows the error CDF on the BR data set for M_1 and M_2, both with and without location information, plotted along with CDFs of two widely known previous works: RADAR[1] and a probabilistic approach used in [29]. A size of 85 training points is shown, since this is close to where algorithms' performance converges.

The results show that only the M_1 with no location information is not competitive with the other algorithms. Indeed, the figure shows that the performance of all of the remaining Bayesian models presented in this paper compare favorably to existing approaches. Given that RADAR and SmoothNN were shown to have comparable performance[12], we can surmise that SmoothNN would have comparable performance to the above CDFs as well. Figure 10 thus shows that we can obtain performance close to a wide range of existing approaches with a fully adaptive, zero profiling approach without excessive sampling.

E. Incorporating Corridor Effects and Other Prior Knowledge

The graphical modeling framework coupled with MCMC provides a very flexible tool for multivariate modeling. Here we pursue two ideas that demonstrate this flexibility and aim to improve predictive accuracy, especially when the training data contain no location information.

Corridor Model.

All three datasets show striking corridor effects. That is, when an access point is located in a corridor, the signal strength tends to be substantially stronger along the entire corridor. In the three office building floors we have examined, corridors are mostly parallel to the walls. Hence, a location that shares either an $x-$coordinate or a $y-$coordinate with an access point (at least approximately) tends to be in the same corridor as that access point.

Figure 11 shows a model (M_3) with a corridor effect, C_i. The variable C_i takes the value 1 if the location (X, Y) shares a corridor with access point i and 0 otherwise. We define "sharing a corridor" as having an $x-$ or $y-$coordinate within three feet of the corresponding access point coordinate. Since corridor width varies from building to building, this definition should vary accordingly, although we do not pursue this here.

The conditional densities for model M_3 are:

$$
\begin{aligned}
X &\sim \text{uniform}(0, L), \\
Y &\sim \text{uniform}(0, B), \\
S_i &\sim N(b_{i0} + b_{i1} \log D_i + b_{i2} C_i + b_{i3} C_i D_i, \tau_i), \\
&\quad i = 1, \ldots, d \\
b_{ij} &\sim N(b_j, \tau_{bj}), i = 1, \ldots, d, j = 0, 1, 2, 3, \\
b_j &\sim N(0, 0.001), j = 0, 1, 2, 3, \\
\tau_{bj} &\sim \text{Gamma}(0.001, 0.001), j = 0, 1, 2, 3.
\end{aligned}
$$

Note we have included a corridor main effect as well as a corridor-distance interaction term. Figure 12 shows the results for various training sample sizes, all with no location information.

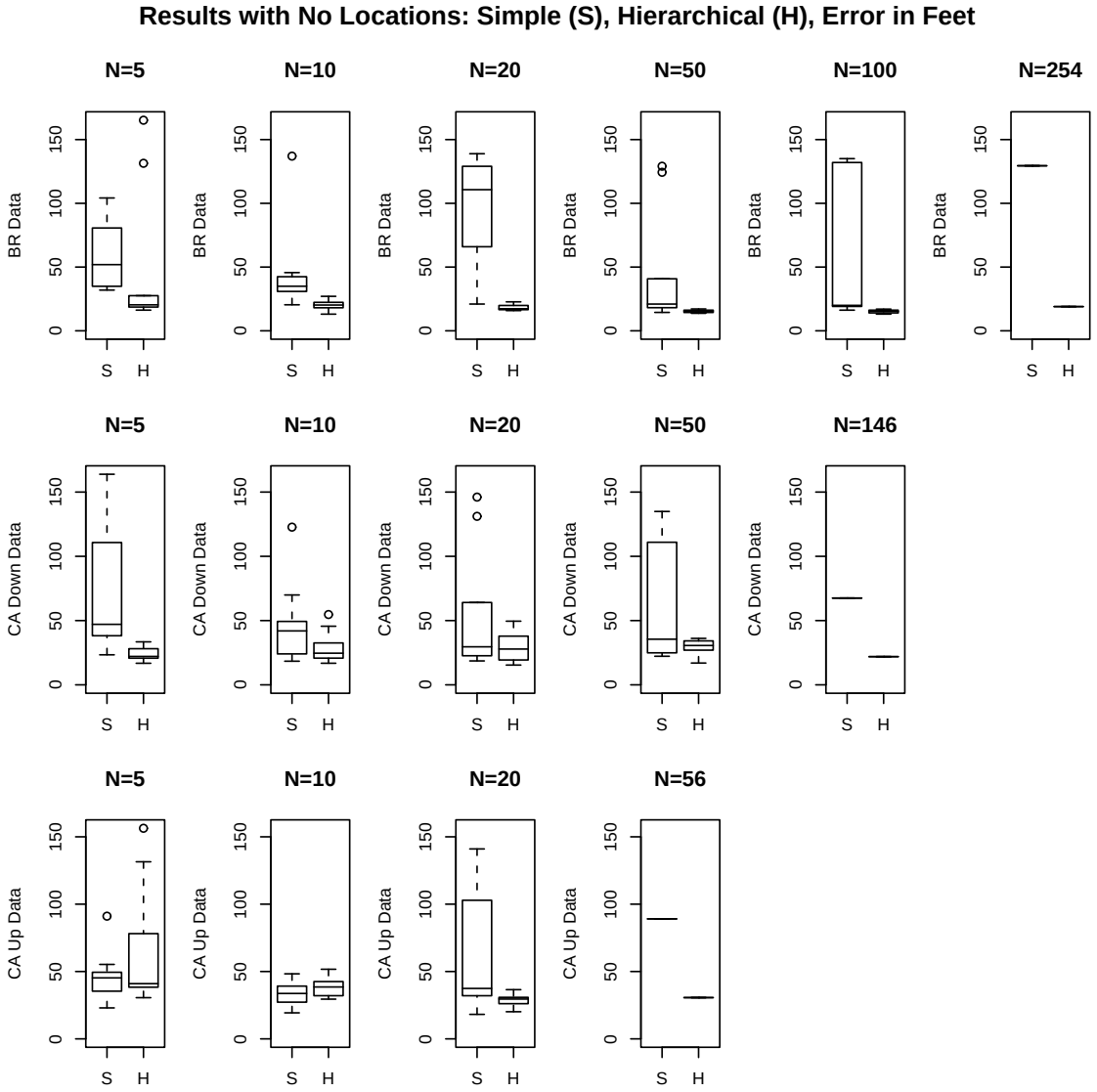

Fig. 9. *Predictive accuracy of the Bayesian graphical model with no location information. Non-hierarchical model (M_1) versus the hierarchical Bayesian graphical model (M_2) for the BR data, CA Down data, and CA Up Data. The verticcal axes show error in feet.*

The results for the case of training data with location information are similar. These analyses suggest that our particular approach to modeling a corridor effect does not improve predictive performance.

Informative Priors for the Regression Co-Efficients.

A second direction we considered was the incorporation of mildly informative prior distributions for the regression co-efficients. Specifically, we used a $N(10, 0.1)$ prior for b_0 and a $N(-19.5, 0.1)$ prior for b_1 in Model M_2. The means of these priors correspond to the average intercept and slope from a maximum likelihood analysis of the combined data over all access points from all three locations. The precisions of 0.1 permit considerable posterior variability around these values.

Figure 13 shows the results using training data with no location information. The informative priors do provide some improved predictive performance, especially for the experiments with small numbers of signal strength vectors. For training data with location information, the informative priors did not improve predictive performance.

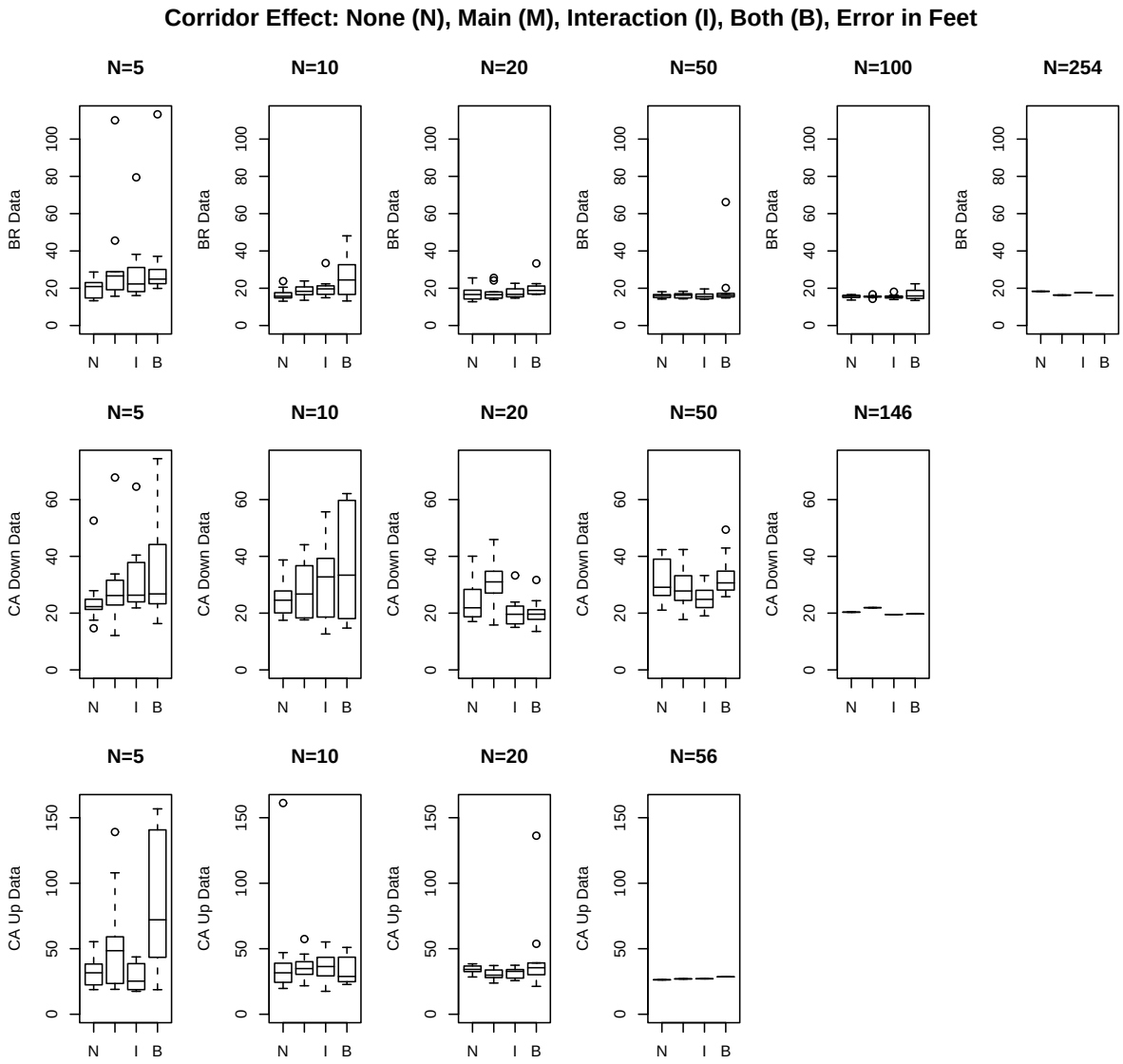

Fig. 12. *Predictive accuracy of the hierarchical Bayesian graphical model (M_2) versus the hierarchical Bayesian graphical model with corridor effects (M_3) for the BR data, CA Down data, and CA Up Data. No location information. "N" corresponds to no corridor effect and is the same as M_2. "M," "I," and "B" correspond to model M_3 with main effect only, interaction only, and both main effect and interaction, respectively. The vertical axes show error in feet.*

VI. Conclusions and Future Work

Our conclusion is the Bayesian hierarchical modeling approach shows considerable promise for localization. The ability to incorporate specific types of prior knowledge proves useful and greatly reduces the requisite profiling effort.

Several directions for future work suggest themselves. In the first instance, we will explore several generalizations of the current model including:

- Piecewise linear or spline-based models for the core signal strength-log distance relationship. The data show some evidence of non-linearity, especially at shorter distances. In particular we will explore the transformation selection algorithm of [9].
- Models that can incorporate approximate location information. For example, when sensors are attached to wireline telephones, the room location may be available but not the location of the sensor within the room.
- Extensions of the corridor effects we discussed above to include more detailed information concerning wall locations as well as locations of potentially interfering objects such as elevators, kitchens, printers, etc.
- Incorporation of other data pertaining to the signal such as angle of arrival.

Currently Markov chain Monte Carlo algorithm estimate the parameters and produce location estimates. For real-time or for larger-scale applications such simulation-based

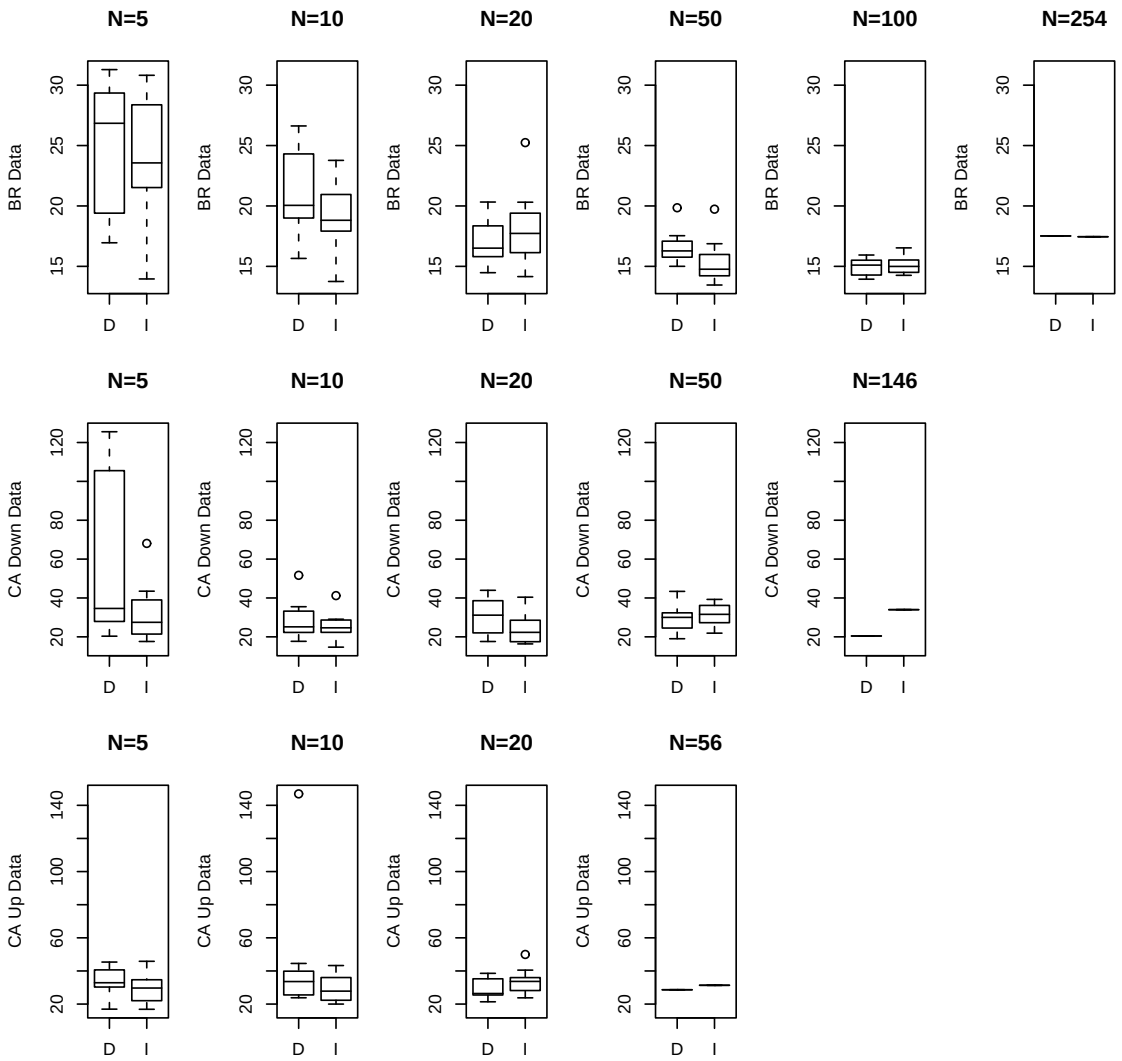

Fig. 13. *Predictive accuracy of the hierarchical Bayesian graphical model (M_2) versus the hierarchical Bayesian graphical model with informative priors and No Location Information, BR data, CA Down data, and CA Up data.*

approaches may prove impractical and we will explore alternatives. In particular, so-called "variational approximations" (see, for example, [11]) may prove useful.

Since our experiments involve multiple test-training splits, manual MCMC convergence checking is not possible. We carried out several runs of 1,000,000 iterations for a few of the models and observed that predictive performance did not improve. Nonetheless, some more systematic, automated approach to convergence diagnostics would be more satisfactory.

A major future thrust of our work will be to extend the current model to dynamic tracking applications. Such applications may begin with a known location (e.g., a location where a user takes a wireless device off a power rack) or not. In either case, the model will assume that the true location varies smoothly over time according to a low-order hidden stochastic process. Robotics has stimulated prior work in this direction and Monte Carlo algorithms for such applications exist. Work on so-called "particle filters is relevant - see, for example, [26], [6], [19], [4]. Again, alternatives to simulation such as online EM algorithms or quasi-Bayes ([16]) procedures may prove necessary.

Acknowledgments

The U.S. National Science Foundation supports Madigan's and Martin's work. Perl and BUGS code for the various models this paper describes are available from the first author. We are grateful to the anonymous reviewers for helpful comments.

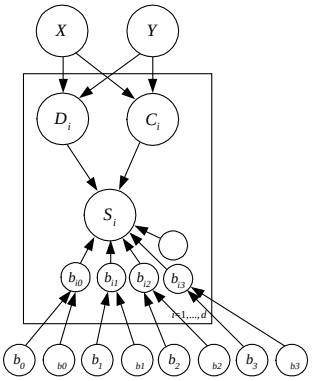

Fig. 11. *An extension of M_2 to include a corridor main effect. This is model M_3.*

REFERENCES

[1] P. Bahl and V.N. Padmanabhan. RADAR: An in-building RF-based user location and tracking system. In *Proceedings of IEEE Infocom 2000*, 2000.

[2] P. Bahl, V.N. Padmanabhan, and A. Balachandran. Enhancements to the RADAR user location and tracking system. Technical report, Microsoft Research Technical Report, February 2000, 2000.

[3] R. Battiti, M. Brunato, and A. Villani. Statistical learning theory for location fingerprinting in wireless LANs. Technical Report DIT-020086, Dipartimento di Informaticae Telecomunicazioni, Universita di Trento, 2002.

[4] M. Berna, B. Lisien, B. Sellner, G. Gordon, F. Pfenning, and S. Thrun. A learning algorithm for localizing people based on wireless signal strength that uses labeled and unlabeled data. In *Proceedings of IJCAI 03*, pages 1427–1428, 2003.

[5] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis (second edition)*. Chapman and Hall, London, 2003.

[6] W.R. Gilks and C. Berzuini. Following a moving target. *Journal of the Royal Statistical Society (Series B)*, 63:127–146, 2001.

[7] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov chain Monte Carlo in Practice*. Chapman and Hall, London, 1996.

[8] David Heckerman, Dan Geiger, and David M. Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20:197, 1995.

[9] J. Hoeting, A.E. Raftery, and D. Madigan. A method for simultaneous variable and transformation selection in linear regression. *Journal of Computational and Graphical Statistics*, 11:485–507, 2002.

[10] G.T. Huang. Casting the wireless sensor net. *Technology Review, MIT's Magazine of Innovation, July/August 2003*, pages 51–56, 2003.

[11] T. Jaakola and M.I. Jordan. Bayesian parameter estimation via variational methods. *Statistics and Computing*, 10:25–37, 2000.

[12] P. Krishnan, A.S. Krishnakumar, W. Ju, C. Mallows, and S. Ganu. A system for LEASE: System location estimation assisted by stationary emitters for indoor rf wireless networks. In *Proceedings of IEEE Infocom 2004*, 2004.

[13] Andrew M. Ladd, Kostas E. Bekris, Algis Rudys, Lydia E. Kavraki, Dan S. Wallach, and Guillaume Marceau. Robotics-based location sensing using wireless ethernet. In *Proceedings of the eighth Annual International Conference on Mobile Computing and Networking (MOBICOM-02)*, pages 227–238, New York, September 23–28 2002. ACM Press.

[14] Steffen L. Lauritzen. *Graphical Models*. Cambridge University Press, 1996.

[15] D. Madigan and J. York. Bayesian graphical models for discrete data. *International Statistical Review*, 63:215–232, 1995.

[16] M. Opper. *A Bayesian approach to online learning*. Cambridge University Press, 1998.

[17] Prasithsangaree, P. Krishnamurthy, and P.K. Chrysanthis. On indoor position location with wireless LANs. In *The 13th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC 2002)*, 2002.

[18] N.B. Priyantha, A. Chakraborty, and H. Balakrishnan. The cricket location support system. In *Proceedings of the Sixth Annual ACM International Conference on Mobile Computing and Networking*, 2000.

[19] G. Ridgeway and D. Madigan. A sequential monte carlo method for bayesian analysis of massive datasets. *Journal of Knowledge Discovery and Data Mining*, 7:301–319, 2003.

[20] T. Roos, P. Myllymaki, and H. Tirri. A statistical modeling approach to location estimation. *IEEE Transactions on Mobile Computing*, 1:59–69, 2002.

[21] S. Saha, K. Chaudhuri, D. Sanghi, and P. Bhagwat. Location determination of a mobile device using IEEE 802.11 access point signals. In *IEEE Wireless Communications and Networking Conference (WCNC)*, 2003.

[22] A. Smailagic, D.P. Siewiorek, J. Anhalt, D. Kogan, and Y. Wang. Location sensing and privacy in a context aware computing environment. *Pervasive Computing 2001*, 2001.

[23] David J. Spiegelhalter. Bayesian graphical modelling: A case-study in monitoring health outcomes. *Applied Statistics*, 47(1):115–133, 1998.

[24] D.J. Spiegelhalter and S.L. Lauritzen. Sequential updating of conditional probabilities on directed graphical stuctures. *Networks*, 20:579–605, 1990.

[25] D.J. Spiegelhalter, A. Thomas, and N.G. Best. WinBUGS version 1.2 user manual. Technical report, MRC Biostatistics Unit, 1999.

[26] S. Thrun. Probabilistic algorithms in robotics. Technical Report CMU-CS-00-126, Computer Science Department, Carnegie Mellon University, 2000.

[27] Roy Want, Andy Hopper, Veronica Falcao, and Jonathan Gibbons. The active badge location system. *ACM Transactions on Information Systems*, 10(1):91–102, January 1992. ftp://ftp.orl.co.uk/pub/docs/ORL/tr.92.1.ps.Z.

[28] J. Werb and C. Lanzl. Designing a positioning system for finding things and people indoors. *IEEE Spectrum*, pages 71–78, 1998.

[29] M. Youssef, A. Agrawala, and A. Udaya Shankar. WLAN location determination via clustering and probability distributions. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2003.

[30] M. Youssef and A.K. Agrawala. Handling samples correlation in the horus system. In *IEEE Infocom*, 2004.