A Statistical Modeling Approach to Location Estimation

Teemu Roos, Petri Myllymäki, and Henry Tirri, *Member*, *IEEE*

Abstract—Some location estimation methods, such as the GPS satellite navigation system, require nonstandard features either in the mobile terminal or the network. Solutions based on generic technologies not intended for location estimation purposes, such as the cell-ID method in GSM/GPRS cellular networks, are usually problematic due to their inadequate location estimation accuracy. In order to enable accurate location estimation when only inaccurate measurements are available, we present an approach to location estimation that is different from the prevailing geometric one. We call our approach the *statistical modeling approach*. As an example application of the proposed statistical modeling framework, we present a location estimation method based on a statistical signal power model. We also present encouraging empirical results from simulated experiments supported by real-world field tests.

Index Terms—Location estimation, mobile terminals, signal propagation, statistical modeling.

1 Introduction

Location-aware computing is a recent interesting research area that exploits the possibilities of modern communication technology [6], [9], [14], [17]. Location-aware devices can be located or can locate themselves, while, by location-aware services, we mean services based upon such location technologies. Location-aware computing has great potential in areas such as personal security, navigation, tourism, and entertainment. The most obvious location-based service is the one answering questions like "Where am I?" and "Where is the nearest shop/bus-stop/hospital?" Now that graphical and interactive applications are technically feasible, it would even be easy to implement an application that presents a map labeled with a mark pointing "You are here." On the other hand, location can be regarded as a filter for the ever-increasing amount of information available to us every day. For instance, people probably do not want to know about daily offerings of supermarkets located hundreds of kilometers away, but information about the nearby supermarket might be of interest.

Location information can also be useful for other people than the user of the location-aware device. For instance, people want to know where their friends are, companies want to know where their delivery vehicles are, rescue officials want to know where injured people are, etc. In the United States, location-based services and, in particular, location of the origin of emergency calls have been considered so important that the service is becoming obligatory the local network operators. This so-called Enhanced-911 requirement was scheduled to become effective in October 2001. Similar actions have been considered in the European Union as well.

• *The authors are with the Complex Systems Computation Group, Helsinki Institute for Information Technology, PO Box 9800, FIN-02015 HUT, Finland. E-mail: {teemu.roos, petri.myllymaki, henry.tirri}@hiit.fi.*

Manuscript received 5 Dec. 2001; revised 2 May 2002; accepted 3 May, 2002. For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number 13-122001.

The location of a mobile terminal can be estimated using radio signals transmitted or received by the terminal [2], [7], [21], [26]. The problem is called by various names: location estimation, geolocation, location identification, localization, positioning, etc. Some location estimation methods, such as GPS, are based on signals transmitted from satellites, while others rely on terrestrial communication. Additional costs to the service provider are minimal in systems based on existing network infrastructure. However, with devices that are not designed for location estimation purposes, the measurements that can be exploited are often scarce. In many networks, the only thing that is available is the received signal strength indication (RSSI) value. Consequently, the location estimation accuracy of such a system is often inadequate for many location services.

Improving the accuracy and applicability of location estimation systems based on the existing network infrastructure would be very useful and it is the main motivation of this work. We focus primarily on cellular—especially GSM/GPRS—networks, but most of the ideas and concepts are applicable to many other networks where received signal power or other suitable location-dependent quantities are available.

One of the most severe problems facing cellular telephone systems is the complex propagation of radio waves in environments with obstructions, reflecting objects, and interference from adjacent cells. In order to ensure good coverage in their cellular networks, operators use so-called cell planning tools that are based on radio wave propagation models [19]. Such models use information about the environment and combine it with knowledge about phenomena such as signal attenuation, reflection, diffraction, and interference. The dependency between the location of the receiver and observable signal properties is important for location estimation as well. Despite this fact, the fusion of propagation models and location estimation is rarely mentioned in the literature.

The common, geometric approach to location estimation is based on angle and distance estimates from which a location estimate is deduced using standard geometry. We will discuss location estimation from a point of view which is different from the geometric one. In particular, a location estimation method based on a statistical propagation model will be proposed.[1] The basic idea is to construct a statistical propagation model that describes the distribution of received signal power at any given location and to use the model for estimating the mobile unit's location when the received power is observed. The proposed approach renders the determination of the user's location a statistical estimation problem. This enables the use of a wealth of statistical machinery designed to handle problems caused by uncertainty and errors in measurements and missing data.

The paper is organized as follows: In Section 2, we describe the statistical modeling approach and contrast it to the geometric approach. A detailed description of a suitable propagation model is given in Section 3. In Section 4, we show how the model's parameters can be estimated from empirical data with the expectation-maximization (EM) algorithm [12], [18]. Section 5 discusses estimation of location using the model after the parameter values have been fixed and some encouraging empirical results are reported in Section 6. Section 7 concludes and suggests directions for future research.

2 Statistical Modeling Approach

The conceptual development of location estimation methods has been modest since the ancient Egyptians and Greeks invented the art of triangulation. The problem has been mostly considered by engineers familiar with geomatics and, consequently, a majority of proposed solutions are geometric in nature. For instance, the Angle of Arrival method is nothing more than triangulation. In addition to triangulation-based methods, there are several geometric methods, such as Time of Arrival and Time Difference of Arrival—which is used in the GPS system—that are based on distance measurements rather than angle measurements. The geometric solutions work very well in ideal conditions. However, if the signal propagation environment differs significantly from ideal conditions, the distance or angle measurements are unreliable. In such cases, serious problems occur because the various measurements are inaccurate at best, incompatible at worst. Special ad hoc heuristics have to be applied in order to compensate for these errors.

Here, we take an alternative approach to the location estimation problem. In this approach, which we call the statistical modeling approach, signal properties, such as received power, angle of arrival, and/or propagation delay, are treated as random variables which are statistically dependent on the locations of the transmitter, the receiver, and the propagation environment. Because of this dependency, an observation of the signal properties allows inferences about the location. A similar idea has been sketched independently in [16], [30].

The conceptual difference between the statistical modeling approach and the geometric approach is clear in the following sense: In the geometric approach, all the interest is in mapping measured signal properties to the location. The same is true in some nongeometric methods also [24]. In contrast to this, the statistical modeling approach emphasizes propagation modeling, which describes the dependency of the measured signal properties on the location variable, i.e., the reasoning proceeds from the location to the signal properties. The location estimation problem is then solved as an inverse—or, rather, inference—problem, which is the kind of reasoning that is typical of statistics, in general. In statistical terms, the propagation model is a sampling distribution whose parameters—in the first phase, the propagation parameters and, in the second phase, the location variable—we wish to estimate.

The problem of incompatible measurements is not present in the statistical modeling approach, unlike the geometric one, because, no matter how unlikely the obtained measurement results are, they are always possible. Of course, if the propagation model does not fitl the actual propagation phenomena and the environment well, the propagation prediction accuracy and, accordingly, the location estimation accuracy is poor. However, whereas the only possibility of enhancing the accuracy of the geometric location estimation methods is to increase the accuracy of the angle and distance measurements, this is not the case with methods based on the statistical modeling approach. Their accuracy can be enhanced also by switching to another propagation model that is better suited for predicting the relevant signal properties in the environment in question. One can even use very sophisticated propagation modeling techniques, such as ray tracing or neural networks (see [31]).

Having said this, we note that it is also possible to combine some of the good aspects of both the statistical modeling approach and the geometric approach. Namely, one can use a propagation model for extracting accurate range or angle estimates and then use these estimates for geometric location determination. As a prime example, consider the received signal power: It is impossible to obtain any range estimates whatsoever without some kind of a propagation model. A method where ranging is based on propagation modeling and location determination is based on geometry has been applied in ad hoc sensor networks [25]. Such hybrid systems may lead to sufficient accuracy with very small time consumption.

So-called empirical location estimation methods require on-site calibration, but their accuracy is usually better than the one of other methods. There are several empirical methods for location estimation, such as RADAR [2], Nibble [8], RadioCamera [15], and the ones developed in the robotics community (see [28] and the references therein). In a forthcoming paper [23], we discuss empirical location estimation and present real-world results from a wireless local area network (WLAN) trial. In the present paper, we focus on an instance of the statistical modeling approach where offsite calibration is necessary to tune the propagation parameters for the *type* of environment (rural, urban, etc.) in question, but no onsite calibration is required.

1. This paper extends the work described in the thesis [29] by T. Roos (former name Tonteri).

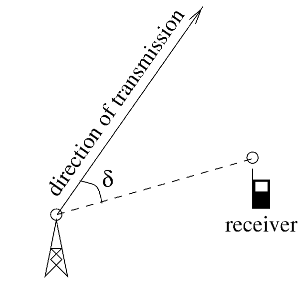

Fig. 1. The deviation, δ, between the direction of transmission and the direction of the receiver as measured from the transmitter.

3 Signal Propagation Model

A propagation model is a mathematical model that predicts some properties of a radio signal at a given location [1], [10], [13], [22], [27], [31]. If the "output" of the model is a probability distribution of the signal's properties, the model is *statistical*, as opposed to a *deterministic* model that gives a point estimate for each of the predicted properties. Received signal power, denoted by r, will be used throughout the paper as the variable quantifying the observed radio signal, although the approach is applicable to any observable property or properties of the signal.

3.1 Single Transmitter Model

The so-called log-loss model (termed log-distance in [20]) can be used as a statistical propagation model as long as an error term with a specified distribution is added. If a zero-mean Gaussian distribution with a constant variance is used for the error term, denoted by e, the model is a linear regression model with three parameters: two regression coefficients, β_0 and β_1, which define the mean value of the received power at a given distance, and the variance of e, denoted by σ^2. The mean value of the received power is given by[2]

$$\mu(d, p, \theta) = p + \beta_0 + \beta_1 \ln d, \tag{1}$$

where d is the transmitter-receiver distance, p is the transmitted power in decibels, and θ denotes the set of parameters.

The transmitters of cellular networks are often directed to some *direction of transmission* to which the transmitted power is higher than to other directions. Therefore, the log-loss model can be improved by adding a term which depends on the deviation between the direction of the receiver and the direction of transmission. Let the deviation be denoted by δ. The values of δ are clearly between zero and 180 degrees (see Fig. 1).

In addition to the parameters of the log-loss model, the improved log-loss model has an additional parameter, β_2, that is associated with δ. The mean value of r is given by

$$\begin{aligned}\mu(d, \delta, p, \theta) &= p + \beta_0 + \beta_1 \ln d + \beta_2 \delta \ln d \\ &= p + \beta_0 + (\beta_1 + \beta_2 \delta) \ln d.\end{aligned} \tag{2}$$

It can be seen on the second line of (2) that, if the deviation, δ, is constant, the improved model is identical to the normal log-loss model with β_1 replaced by $\beta_1 + \beta_2\delta$. In other words, attenuation obeys the log-loss model along any straight line originating from the transmitter. Fig. 2 shows values of μ evaluated using (2).

2. We denote the natural (base e) logarithm by $\ln$.

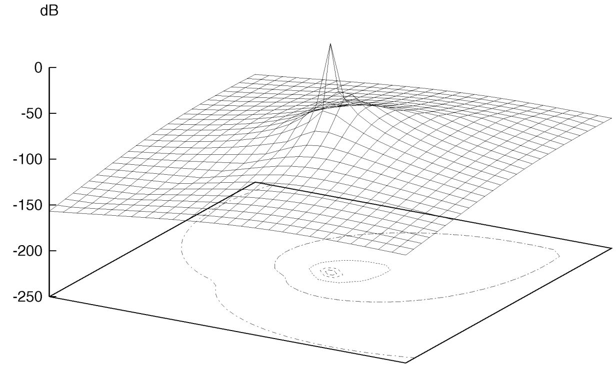

Fig. 2. An illustration of the average attenuation evaluated using (2). The transmitter is located at the center of the area and its direction of transmission is toward the upper right corner of the area.

The distribution of r is Gaussian with the following p.d.f.:

$$f(r|d, \delta, p, \theta) = \frac{1}{\sqrt{2\pi}\,\sigma}\,\xi\left(\frac{r - \mu(d, \delta, p, \theta)}{\sigma}\right), \tag{3}$$

where $\xi(x)$ is just a short-hand notation defined as

$$\xi(x) \overset{\text{df.}}{=} \exp\left(-\frac{1}{2}x^2\right). \tag{4}$$

3.2 Multiple Transmitters Model

We have now described how the distribution of the received signal power is evaluated with respect to one transmitter. Let us now extend the model to several transmitters. First, because, in cellular networks, many *channels*, each operating on a separate frequency range, are used simultaneously, there are actually as many received signal power variables as there are channels. Let r_j denote the received power of channel j and c_i denote the channel of transmitter i. Second, transmitters are classified depending on their transmission properties and location with respect to buildings. For instance, the signal received from an indoor transmitter is usually weaker than the signal received from an outdoor transmitter at the same distance because of the attenuation caused by buildings. In order to take these differences into account, we could use different parameters for each transmitter type. This modification would be straightforward, but, for the sake of simplicity, we do not introduce it here. The details can be found in [29].

If there are two transmitters on the same channel, they cause interference and it is difficult to predict the resulting field strength. However, the situations in which two transmitters on the same channel are close to each other are intentionally avoided while planning the layout of the network and, hence, the power received from no more than one transmitter is likely to be significant. In most cases, a good approximation is obtained by ignoring all transmitters except the one whose mean power according to (2) is the highest at the location in question.

Thus, each transmitter i has location, denoted by l_i, direction of transmission, denoted by α_i, and transmitted power, denoted by p_i. Let g_j denote the p.d.f. of the received power on channel j, given that the measurement is performed at location l. It is given by the equation

$$g_j(r|l,\theta) \stackrel{\text{df.}}{=} f\big(r|d(l,l_i),\ \delta(l,l_i,\alpha_i),\ p_i,\ \theta\big), \tag{5}$$

where $d(l,l_i)$ is the distance between locations l and l_i, $\delta(l,l_i,\alpha_i)$ is the deviation at location l with respect to a transmitter located at l_i and directed to α_i. The index i is chosen so that it maximizes the mean received power:

$$i = \text{argmax}_{\{\,i\,:\,c_i=j\,\}}\mu\big(d(l,l_i),\ \delta(l,l_i,\alpha_i),\ p_i,\ \theta\big), \tag{6}$$

where function μ is given by (2). Thus, when the propagation parameters and the location, channel, direction of transmission, and transmitted power of the transmitters are fixed, an estimate of the distribution of the signal strength r_j, for each channel j, is available for every location. We shall next consider how to deal with the unknown propagation parameters.

4 Estimation of Propagation Parameters

In most propagation models, there are some parameters whose values cannot be derived from the underlying theory. These parameters are typically somehow related to the environment and, hence, there are no universally good values for them. In such cases, it is obligatory to use empirical data to obtain information about the parameter values. Note, however, that it is generally unjustified to assume the existence of some *true* parameter values, which are referred to in the following quote:

> In many statistics problems, the probability distribution that generated the experimental data is completely known except for the values of one or more parameters. [11]

When modeling phenomena as complex as radio wave propagation, the assumption is certainly incorrect. Instead of trying to find the "true" parameter values, a more realistic goal would be to maximize the predictive accuracy. A reasonable solution, frequently used in statistics, is to use the *maximum likelihood* parameters. In the following, we will show how to obtain the maximum likelihood parameters for the presented propagation model.

In our case, the propagation parameters are estimated from data consisting of n measurements of received signal power, each labeled with the corresponding channel and location of the receiver. The transmitter information consists of the already mentioned properties, namely the location, channel, direction of transmission, and transmitted power of each transmitter. Based on the data, we need to estimate the parameters $\beta_0, \beta_1, \beta_2$, and σ of (2) and (3).

As a preprocessing step, the transmitter information is combined with the received power measurements in order to produce a table consisting of the following columns:

1. received signal powers, $\mathbf{r} = (r^{(1)}, \dots, r^{(n)})$,
2. distances between the transmitter and the receiver, $\mathbf{d} = (d^{(1)}, \dots, d^{(n)})$,
3. deviations between the direction of transmission and the direction of the receiver, $\delta = (\delta^{(1)}, \dots, \delta^{(n)})$, and
4. transmitted powers, $\mathbf{p} = (p^{(1)}, \dots, p^{(n)})$.

Filling in fields 2-4 requires that the source of each measured signal is identified, also in cases where there are several transmitters on the same channel. In such cases, we assume that the signal is coming from the transmitter that is nearest to the receiver, although, in principle, (6) should be used, and these two criteria do not always agree. This is a deliberate pragmatic choice: Using (6) would require treating the ambiguous cases as missing data because θ, whose value is unknown, appears in the equation.

We shall next describe how to obtain maximum likelihood estimates (MLEs) of the parameters, or approximations thereof, from empirical data. First, the simple case where none of the data is missing is discussed, after which a solution to the realistic missing data case is presented.

4.1 Maximum Likelihood from Complete Data

Evaluating the MLEs from complete data can be performed easily by exploiting the fact that what we have is, in effect, a linear regression model and the standard methods for solving MLEs for linear regression models can be applied.

Given n fully observed data vectors, $\mathbf{r}$, $\mathbf{d}$, δ, and $\mathbf{p}$, the likelihood, $\mathcal{L}(\theta)$, is a product of the conditional p.d.f.s of the individual observations:

$$\begin{aligned}\mathcal{L}(\theta) &= \prod_{i=1}^{n} f\big(r^{(i)}|d^{(i)}, \delta^{(i)}, p^{(i)}, \theta\big) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma}\ \xi\Big(\frac{r^{(i)} - \mu^{(i)}}{\sigma}\Big),\end{aligned} \tag{7}$$

where $\mu^{(i)}$ is given by

$$\mu^{(i)} \stackrel{\text{df.}}{=} \mu\big(d^{(i)}, \delta^{(i)}, p^{(i)}, \theta\big). \tag{8}$$

The factorization of $\mathcal{L}(\theta)$ is based on the assumption that the variables $r^{(1)}, \dots, r^{(n)}$ are independent.

The likelihood function can be rewritten as

$$\mathcal{L}(\theta) = \Big(\frac{1}{\sqrt{2\pi}\,\sigma}\Big)^n \exp\Big(-\frac{\text{SSE}}{2\,\sigma^2}\Big), \tag{9}$$

where the sum of squared errors, SSE, is given by

$$\text{SSE} \stackrel{\text{df.}}{=} \sum_{i=1}^{n}\big(r^{(i)} - \mu^{(i)}\big)^2. \tag{10}$$

Without proof—one can be found in [11]—we state that the maximum likelihood estimates[3] $\widehat{\beta}_0$, $\widehat{\beta}_1$, and $\widehat{\beta}_2$ are independent of $\widehat{\sigma}$ and that they can be obtained by minimizing SSE. Using matrix notation,[4] the solution is given by

$$\widehat{\beta} = \big(\mathbf{Z}^T\,\mathbf{Z}\big)^{-1}\,\mathbf{Z}^T\,\mathbf{Y}, \tag{11}$$

where $\widehat{\beta}$, $\mathbf{Z}$, and $\mathbf{Y}$ are defined as

$$\widehat{\beta} \stackrel{\text{df.}}{=} \begin{bmatrix}\widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2\end{bmatrix}, \tag{12}$$

3. We denote the MLE of variable X by $\widehat{X}$.
4. The notion $\mathbf{A}^T$ denotes the transpose of matrix $\mathbf{A}$ and the inverse of matrix $\mathbf{A}$ is denoted by $\mathbf{A}^{-1}$.

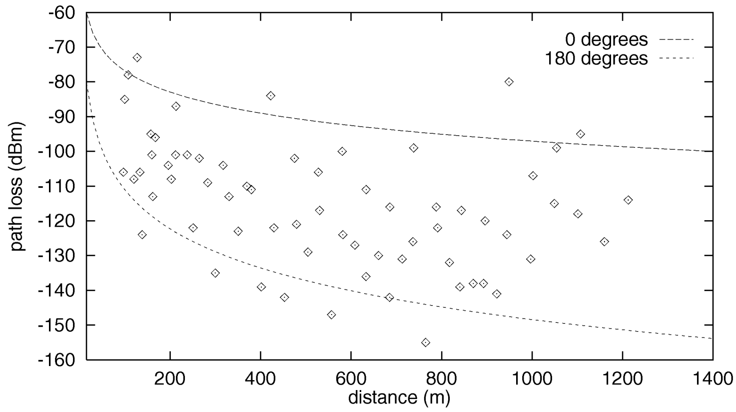

Fig. 3. Mean path loss curves obtained from sample data. Small blots represent observed path loss values at varying distances from the transmitter. The two curves show the mean path loss to the direction of transmissions ($\delta = 0^\circ$) and to the opposite direction ($\delta = 180^\circ$).

$$\mathbf{Y} \overset{\text{df.}}{=} \begin{bmatrix} r^{(1)} - p^{(1)} \\ r^{(2)} - p^{(2)} \\ \vdots \\ r^{(n)} - p^{(n)} \end{bmatrix}, \tag{13}$$

$$\mathbf{Z} \overset{\text{df.}}{=} \begin{bmatrix} 1 & \ln d^{(1)} & \delta^{(1)} \ln d^{(1)} \\ 1 & \ln d^{(2)} & \delta^{(2)} \ln d^{(2)} \\ \vdots & \vdots & \vdots \\ 1 & \ln d^{(n)} & \delta^{(n)} \ln d^{(n)} \end{bmatrix}. \tag{14}$$

Finally, the MLE of σ can be obtained from

$$\widehat{\sigma} = \sqrt{\frac{\text{SSE}}{n}}. \tag{15}$$

The value of SSE is obtained by fixing the values of the β-parameters to their MLEs given by (11). Equations (11) and (15) give us the MLEs of the parameter in closed form when the data is complete. The somewhat more complicated missing data case is discussed in the next section.

Example 1. Fig. 3 shows an artificial data set containing 66 observations. The path loss values plotted on the vertical axis are the same values that are contained in matrix $\mathbf{Y}$. The data was generated by sampling from the propagation model presented in this section. Table 1 shows the parameters used for generating the data and the MLEs evaluated using (11) and (15).

4.2 Maximum Likelihood from Incomplete Data

In general, the received power values cannot be directly observed because of physical restrictions. First, the received power values have to be *binned*, i.e., rounded to finite accuracy. Second, because of sensitivity limitations, the received power on only some channels—those with the strongest signal—is reported. Moreover, in the case of GSM/GPRS telephones, the received power values of only six to eight channels are available. The only information about the other channels is that their received power does not exceed the power on any of the reported channels. In such cases, we say that the received power variable is *truncated* at a point given by the smallest of the known values. We will now present a method for handling binned and truncated variables.

Let the random vector $\mathbf{o} = o^{(1)}, \ldots, o^{(n)}$ denote the observations. For simplicity, we assume that the observations are labeled in such a way that the first m variables correspond to binned observations and the $n - m$ other ones correspond to truncated observations. Thus, the relationship between $\mathbf{o}$ and $\mathbf{r}$ is defined by

$$\begin{aligned} o^{(i)} - \tfrac{\epsilon}{2} \leq\ & r^{(i)} < o^{(i)} + \tfrac{\epsilon}{2} && \text{for } i \in \{1, \ldots, m\} \\ & r^{(i)} \leq o^{(i)} + \tfrac{\epsilon}{2} && \text{for } i \in \{m+1, \ldots, n\}, \end{aligned} \tag{16}$$

where the accuracy is determined by ϵ, whose value can be, for instance, 1.0 dBm.

The likelihood function for incomplete data, $\mathcal{L}_{\mathcal{I}}$ (the $\mathcal{I}$ stands for *incomplete*), for an observation vector $\mathbf{o}$ is given by

$$\begin{aligned} \mathcal{L}_{\mathcal{I}}(\theta) = & \prod_{i=1}^{m} \int_{o^{(i)} - \frac{\epsilon}{2}}^{o^{(i)} + \frac{\epsilon}{2}} f\big(r | d^{(i)}, \delta^{(i)}, p^{(i)}, \theta\big)\, dr \\ & \prod_{i=m+1}^{n} \int_{-\infty}^{o^{(i)} + \frac{\epsilon}{2}} f\big(r | d^{(i)}, \delta^{(i)}, p^{(i)}, \theta\big)\, dr. \end{aligned} \tag{17}$$

The equation is analogous to the likelihood function for complete data given by (7). However, it is not straightforward to derive a closed form solution analogous to the complete-data solution. Instead, there is a method which can be used to approximate a local maximum of the likelihood function from incomplete data, namely the Expectation–Maximization (EM) algorithm [12], [18].

The EM algorithm can be applied whenever it is possible to evaluate the expected value of the logarithm of the complete data likelihood (log-likelihood). In order to evaluate the expected log-likelihood, we need a probability distribution for the missing received power values. In the EM algorithm, the distribution is obtained by fixing the parameters to some hypothetical values, say θ_t. The expectation of the log-likelihood function, denoted by $Q(\theta, \theta_t)$, is then evaluated in the *expectation step* using the equation:[5]

TABLE 1
The Actual Parameter Values Used when Generating the Data Set of Example 1 and the Corresponding MLEs

	β_0	β_1	β_2	σ
MLE	**-36.43**	**-8.77**	**-0.0413**	**9.8**
actual	-30.00	-10.00	-0.0400	10.00

5. The notation E_{θ_t} denotes the conditional expectation given θ_t.

$$Q(\theta, \theta_t) \overset{\text{df.}}{=} E_{\theta_t} \ln \mathcal{L}(\theta), \tag{18}$$

where $\mathcal{L}(\theta)$ is the complete-data likelihood, given by (7). In the *maximization step*, the parameter values are replaced by ones which maximize the expected log-likelihood, thus giving

$$\theta_{t+1} = \text{argmax}_\theta Q(\theta, \theta_t), \tag{19}$$

where θ_t denotes the parameters on step t. The algorithm consists of repeating these two steps, one after the other. It can be shown that the likelihood of the parameter values never decreases during an iteration. Thus, if the algorithm converges, it converges to a local maximum of the likelihood function.

It now remains to be shown how to obtain a set of parameter values from (19). By taking the logarithm of $\mathcal{L}(\theta)$, given by (7), and substituting it into (18), we get

$$\begin{aligned} Q(\theta, \theta_t) = E_{\theta_t} \sum_{i=1}^{n} \Bigg[& -\frac{1}{2}\ln 2\pi - \ln\sigma \\ & -\frac{1}{2}\left(\frac{r^{(i)} - \mu^{(i)}}{\sigma}\right)^2 \Bigg]. \end{aligned} \tag{20}$$

By switching the order of the expectation and sum operators and taking terms that are independent of $\mathbf{r}$ outside of the expectation, the equation becomes

$$\begin{aligned} Q(\theta, \theta_t) = \sum_{i=1}^{n} \Big[& -\frac{1}{2}\ln 2\pi - \ln\sigma \\ & -\frac{1}{2\,\sigma^2} E_{\theta_t}\big(r^{(i)} - \mu^{(i)}\big)^2 \Big]. \end{aligned} \tag{21}$$

Further rewriting yields

$$\begin{aligned} Q(\theta, \theta_t) = & \, n\left(-\frac{1}{2}\ln 2\pi - \ln\sigma\right) \\ & -\frac{1}{2\,\sigma^2}\ \text{SESE}, \end{aligned} \tag{22}$$

where SESE is the sum of expected squared errors given by

$$\text{SESE} \overset{\text{df.}}{=} \sum_{i=1}^{n} E_{\theta_t}\big(r^{(i)} - \mu^{(i)}\big)^2. \tag{23}$$

The value of β maximizing (22) is given by (see [29] for a proof)

$$\widehat{\beta} = \left(\mathbf{Z}^T\, \mathbf{Z}\right)^{-1} \mathbf{Z}^T\, \mathbf{Y}, \tag{24}$$

where $\widehat{\beta}$,[6] and $\mathbf{Z}$ are defined by (12) and (14), and $\mathbf{Y}$ is defined as

$$\mathbf{Y} \overset{\text{df.}}{=} \begin{bmatrix} E_{\theta_t} r^{(1)} - p^{(1)} \\ E_{\theta_t} r^{(1)} - p^{(2)} \\ \vdots \\ E_{\theta_t} r^{(n)} - p^{(n)} \end{bmatrix}. \tag{25}$$

Thus, in order to obtain estimates of the β-parameters, we need to evaluate the expected value of $r^{(i)}$ for each $i \in \{1, \ldots, n\}$. For binned observations, the expected value of $r^{(i)}$ is (see [29] for a proof)

$$E_{\theta_t} r^{(i)} = \frac{\big(\xi(a^{(i)}) - \xi(b^{(i)})\,\big)\sigma_t}{\sqrt{2\pi}\,\left(\Phi(b^{(i)}) - \Phi(a^{(i)})\right)} + \mu_t^{(i)}, \tag{26}$$

where Φ is the cumulative distribution function of a Gaussian distribution with zero mean and unity variance; $\mu_t^{(i)}$ is the mean received power value according to the log-loss model with parameters θ_t:

$$\mu_t^{(i)} \overset{\text{df.}}{=} \mu\big(d^{(i)}, \delta^{(i)}, p^{(i)}, \theta_t\big), \tag{27}$$

and $a^{(i)}$ and $b^{(i)}$ are given by

$$a^{(i)} \overset{\text{df.}}{=} \frac{o^{(i)} - \frac{\epsilon}{2} - \mu_t^{(i)}}{\sigma_t}, \tag{28}$$

$$b^{(i)} \overset{\text{df.}}{=} \frac{o^{(i)} + \frac{\epsilon}{2} - \mu_t^{(i)}}{\sigma_t}. \tag{29}$$

Because the value of $r^{(i)}$ is known to be within the range $o^{(i)} \pm \frac{\epsilon}{2}$, its expected value must also be within the same range. The difference between the exact solution and $o^{(i)}$ is bound by the equation

$$\left|E_{\theta_t} r^{(i)} - o^{(i)}\right| \leq \frac{\epsilon}{2}. \tag{30}$$

Thus, the expectation can be approximated by $o^{(i)}$.

For truncated observations, the expectation of $r^{(i)}$ is given by (see [29] for a proof)

$$E_{\theta_t} r^{(i)} = -\frac{\xi(b^{(i)})\ \sigma_t}{\sqrt{2\pi}\ \Phi(b^{(i)}) + \mu_t^{(i)}}, \tag{31}$$

where $b^{(i)}$ is given by (29).

The value of σ maximizing (22) is given by (see [29] for a proof)

$$\widehat{\sigma} = \sqrt{\frac{\text{SESE}}{n}}. \tag{32}$$

In order to evaluate SESE that appears in (32), we need a closed form solution for the expected squared error $E_{\theta_t}(r^{(i)} - \mu^{(i)})^2$. For binned observations, it is given by (see [29]for a proof)

$$\begin{aligned} E_{\theta_t}\big(r^{(i)} - \mu^{(i)}\big)^2 = & \\ & \frac{\sigma_t^2\ \left(a^{(i)}\xi(a^{(i)}) - b^{(i)}\xi(b^{(i)})\right)}{\sqrt{2\pi}\ (\Phi(b^{(i)}) - \Phi(a^{(i)}))} + \sigma_t^2 \\ & + \frac{2\ \sigma_t\big(\mu_t^{(i)} - \mu^{(i)}\big)\ \left(\xi(a^{(i)}) - \xi(b^{(i)})\right)}{\sqrt{2\pi}\ (\Phi(b^{(i)}) - \Phi(a^{(i)}))} \\ & + \big(\mu_t^{(i)} - \mu^{(i)}\big)^2, \end{aligned} \tag{33}$$

where $a^{(i)}$ and $b^{(i)}$ are given by (28) and (29) and $\mu^{(i)}$ is obtained by using the estimates of the β-parameters given by (24). A reasonable approximation to (33) is given by $(o^{(i)} - \mu^{(i)})^2$ because $r^{(i)}$ is known to be within the range $o^{(i)} \pm \frac{\epsilon}{2}$.[7]

For truncated observations, the expected squared error is given by (see [29] for a proof):

6. The notation $\widehat{\beta}$ is used, although strictly speaking, the solution is the maximizer of $Q(\theta, \theta_t)$, not the (incomplete-data) likelihood.

7. Such an approximation is, in fact, implicity used every time finite-precision values are treated as real values. This was the case in the complete-data case of the previous section.

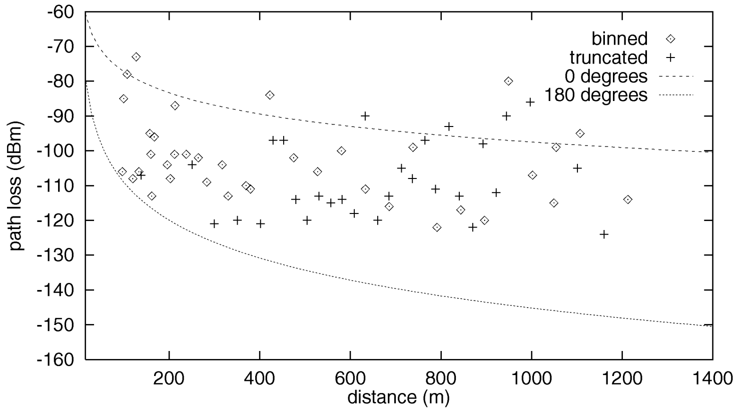

Fig. 4. Mean path loss curves obtained from sample data. Small symbols represent binned (△) and truncated (+) path loss values at varying distances from the transmitter. The two curves show the mean path loss to the direction of transmission ($\delta = 0°$), and to the opposite direction

$$E_{\theta_t}\left(r^{(i)} - \mu^{(i)}\right)^2 = - \frac{\sigma_t^2\ b^{(i)}\xi(b^{(i)})}{\sqrt{2\pi}\ \Phi(b^{(i)})} + \sigma_t^2 - \frac{2\ \sigma_t\left(\mu_t^{(i)} - \mu^{(i)}\right)\ \xi(b^{(i)})}{\sqrt{2\pi}\ \Phi(b^{(i)})} + \left(\mu_t^{(i)} - \mu^{(i)}\right)^2. \tag{34}$$

By looking at (34), one can see that the last two terms can be ignored if we assume that the difference $|\mu_t^{(i)} - \mu^{(i)}|$, i.e., the difference between two consecutive estimates of the mean received power is very small. Unless the EM-algorithm does not converge at all, this is guaranteed to be the case in the long run.

We now have closed form solutions for the parameters maximizing (22): (24) for β and (32) for σ. By using them, we obtain the parameters θ maximizing (18). This is all that is needed to solve (19) and, in fact, all that is needed to perform an iteration of the EM algorithm. To run the algorithm, in practice, we still need to determine the initial parameter values θ_0 and, as EM converges only to a local optimum, it is obvious that this choice may have a strong effect on the final result. This issue is not addressed in this paper, but, for the remainder of the paper, we assume that the initial parameter values are computed from the nontruncated observations by using (11) and (15)—a choice we have found to work quite well in practice.

Example 2. Fig. 4 shows an artificial data set containing 66 observations, 37 of which are binned, while the 29 other ones are truncated. For truncated observations, the figure shows the truncation point which is known to be higher than the unknown path loss value. Table 2 shows the parameters used for generating the data and estimates obtained with the EM-algorithm. The data set of Example 2 is the same as the one used in Example 1 with the exception that, in Example 2, some of the observations are truncated.

The EM algorithm scales efficiently to large data sets. We generated a data set with 40,320 observations (5,040 binned, 35,280 truncated) with the same parameter values as in Examples 1 and 2. The algorithm converged in 573 iterations producing parameter estimates (-32.03, -9.73, -0.0398, 10.00). Thus, the differences between the actual values and the ones obtained in Examples 1 and 2 were mainly caused by the relatively small sample size, not the estimation method.

5 Location Estimation

Given the estimates of the propagation parameters $\hat{\theta}$, the p.d.f. of the received power on channel j at location l is given by $g_j(r_j|l, \hat{\theta})$, where g_j is defined by (5). The posterior p.d.f. of the location variable l is given by the Bayes rule:[8]

$$p(l|\mathbf{r}, \hat{\theta}) = \frac{g(\mathbf{r}|l, \hat{\theta})\ \pi(l)}{\int\ g(\mathbf{r}|l', \hat{\theta})\pi(l')dl'}, \tag{35}$$

where $\mathbf{r}$ is a vector consisting of the received power values r_j for each channel j and $g(\mathbf{r}|l, \hat{\theta})$ is the likelihood function given by

$$g(\mathbf{r}|l, \hat{\theta}) = \prod_j g_j(r_j|l, \hat{\theta}) \tag{36}$$

and π is the prior p.d.f. of the location variable.

However, (36) is not directly applicable for practical location estimation purposes if some of the received power observations are truncated.[9] It is not the actual received power vector, $\mathbf{r}$, that is observed, but the observation vector, $\mathbf{o}$, whose relation to $\mathbf{r}$ is the following:

$$\begin{aligned} o_j - \tfrac{\epsilon}{2} \le r_j < o_j + \tfrac{\epsilon}{2} \quad & \text{if } j \in \mathcal{B} \\ r_j \le o_j + \tfrac{\epsilon}{2} \quad & \text{if } j \in \mathcal{T}, \end{aligned} \tag{37}$$

where $\mathcal{B}$ is the set of binned channels and $\mathcal{T}$ is the set of truncated channels and the accuracy of the measurements is determined by ϵ.

8. The application of the Bayes rule might be opposed by some people who prefer the frequentist statistical theory over its Bayesian correspondent [4], [5]. The primary concern of the opponent is usually related to the concept of prior distributions. However, in this case, the results obtained with frequentist statistical methods would be similar to the ones presented here, as we will note later.

9. If there are no truncated observations, i.e., all the observations are binned, (36) is applicable because one can use the center points of the bins as approximations to the actual values of the received power variables unless the bins are very wide.

TABLE 2
The Values of the Parameter Estimates for EM-Iterations 1-25 with the Data Set of Example 2

iter.	β_0	β_1	β_2	σ
1	-51.84	-6.38	-0.0280	9.4
2	-48.57	-6.98	-0.0313	9.4
3	-44.58	-7.59	-0.0337	9.5
4	-42.01	-7.98	-0.0352	9.5
5	-40.36	-8.23	-0.0362	9.6
6	-39.27	-8.40	-0.0368	9.6
7	-38.52	-8.51	-0.0373	9.7
8	-38.01	-8.59	-0.0376	9.7
9	-37.65	-8.65	-0.0378	9.7
10	-37.40	-8.69	-0.0379	9.7
11	-37.22	-8.71	-0.0380	9.8
12	-37.09	-8.73	-0.0381	9.8
13	-37.00	-8.75	-0.0382	9.8
14	-36.94	-8.76	-0.0382	9.8
15	-36.89	-8.76	-0.0383	9.8
16	-36.85	-8.77	-0.0383	9.8
17	-36.83	-8.77	-0.0383	9.8
18	-36.81	-8.77	-0.0383	9.8
19	-36.80	-8.78	-0.0383	9.8
20	-36.79	-8.78	-0.0383	9.8
21	-36.79	-8.78	-0.0383	9.8
22	-36.78	-8.78	-0.0383	9.8
23	-36.78	-8.78	-0.0383	9.8
24	-36.78	-8.78	-0.0383	9.8
25	**-36.77**	**-8.78**	**-0.0383**	**9.8**
actual	-30.00	-10.00	-0.0400	10.00

The algorithm has converged with the precision used in the table by iteration 25. The actual parameter values used when generating the data set are shown at the bottom of the table.

Now that the propagation parameters, $\hat{\theta}$, are fixed, the likelihood function is defined with respect to the location variable, l, and, thus, the likelihood function is given by

$$g(\mathbf{o}|l,\hat{\theta}) = \prod_{j\in\mathcal{B}} \int_{o_j-\frac{\epsilon}{2}}^{o_j+\frac{\epsilon}{2}} g_j(r|l,\hat{\theta})\ dr \prod_{j\in\mathcal{T}} \int_{-\infty}^{o_j+\frac{\epsilon}{2}} g_j(r|l,\hat{\theta})\ dr. \tag{38}$$

The corresponding posterior p.d.f. of the location variable is then

$$p(l|\mathbf{o},\hat{\theta}) = \frac{g(\mathbf{o}|l,\hat{\theta})\ \pi(l)}{\int g(\mathbf{o}|l',\hat{\theta})\ \pi(l')\ dl'}. \tag{39}$$

The denominator of the right-hand side of (39) is constant with respect to l and, thus, the posterior p.d.f. of the location variable is proportional to the numerator:

$$p(l|\mathbf{o},\hat{\theta}) \propto g(\mathbf{o}|l,\hat{\theta})\ \pi(l). \tag{40}$$

In theory, the location variable might be continuous in $\mathbb{R}^2$. In that case, no proper uniform prior π would exist.[10] In practice, however, the location variable is always restricted to some area and, thus, a uniform prior can be used. Of course, if an informative prior is available, it should be used instead.

A location estimate is chosen depending on the penalty function, which defines how different errors are penalized. Two reasonable estimates are the maximum a posteriori location, i.e., the location maximizing (40), and the expected value of the location variable. The latter minimizes the expected value of the squared error of the location estimate. If a uniform prior is used, the maximum a posteriori location is the same as the maximum likelihood estimate of l, which would probably be the solution preferred by advocates of the frequentist statistical theory.

No closed form solution for either the maximum a posteriori value or the expected value is available. Therefore, one has to resort to some numerical method in order to obtain an approximate solution. One can, for instance, discretize the location variable into squares of fixed size, say 50×50 meters, and use the center point of each square to evaluate the distribution of the received power in that particular square. After discretization, the maximum a posteriori value can be obtained simply by going through each of the squares and choosing the value that maximizes (40). The expected value of the location variable can be obtained by calculating an average of the location variable weighted by (40).

6 Simulation Results

We have evaluated the empirical performance of the presented method in a proprietary real-life test case with two industrial partners. As we are, however, not at liberty to present the details of that trial, we will, in this paper, present illustrative examples using an artificial network layout, shown in the background of Fig. 6. The simulation environment was designed to correspond to our real-life test case as closely as possible so that the simulation results are practically equivalent to the real-world results.

Example 3. Assuming a hypothetical network layout, Fig. 5 shows four examples of the posterior p.d.f. of the location variable, the resulting maximum a posteriori location estimate, and the expected value of the location variable with artificial received power measurement results. In Fig. 5a, the received power on channel 14 is known to be -65 dBm and information concerning the other channels is nonexistent. In Fig. 5b, in addition to channel 14, the received power on channel 13 is observed to be -65 dBm. The posterior density is much more concentrated in Fig. 5b than in Fig. 5a.

Figs. 5c and 5d illustrate the effect of truncated observations. The measurement data is the same as in Figs. 5a and 5b, respectively. By comparing, in particular, Fig. 5a to Fig. 5c, one can see that the truncated observations can be useful when estimating location; more posterior probability is concentrated near transmitter 14 because other transmitters should be observed elsewhere. The location estimates in Figs. 5b, 5c, and 5d are intuitively sensible because the estimated location is within the region where the transmitters on channels 13 and 14 are located.

10. A prior $\pi(x)$ is proper if it is nonnegative, $\pi(x) \leq 0$, for all x, and it integrates to one, $\int \pi(x)dx = 1$. A uniform prior $\pi(x) = c$, where c is constant, violates the latter condition unless the range of x is finite.

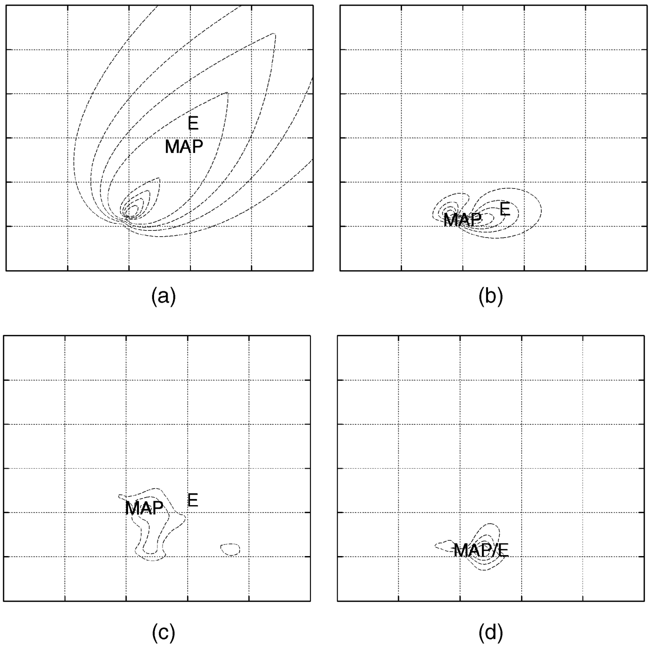

Fig. 5. Examples of the posterior p.d.f. of the location variable with sets of received power observations described in Example 3. Labels indicate the maximum a posteriori estimate ("MAP"), and the expected value ("E") of the location varaible. In (d), they are practically identical.

Example 3 gives an indication of the feasibility of the statistical modeling approach. However, the most important point was still missing because we actually had no means to evaluate the location estimation accuracy. This was because there was no correct location to which the estimate could have been compared. We now present results from a more complete simulation.

We constructed an imaginative trajectory of a mobile terminal. In order to obtain received power measurements, we sampled the improved log-loss model presented above. Parameter values used for sampling were $(-30.00,\ -10.00,\ -0.0400,\ 10.0)$ (see Table 2). The presented location estimation method was then used for location estimation and the location estimates were compared to the trajectory. The parameter values given to the location estimation method were the ones actually obtained with the EM algorithm, i.e., $(-32.03,\ -9.73,\ -0.0398,\ 10.0)$ (see Section 4.2).

Example 4. Using the same network layout as in Example 3 and an imaginary trajectory, a location estimation simulation was performed. As the mobile terminal moved along the trajectory, a sample of received power measurements, each measurement labeled with the correct location, was drawn from the improved log-loss model. The simulated measurements were fed to the location estimation method described in this work. Finally, the location estimates (in particular, the expected value of the location variable, denoted by E in Fig. 5) were compared to the correct coordinates. For comparison purposes, the same data was also used to measure the performance of the standard cell-ID location method in which the location of the transmitter of the channel on which the highest signal power is received is used as a location estimate.[11] Figs. 6 and 7 and Table 3 present a summary of the results.

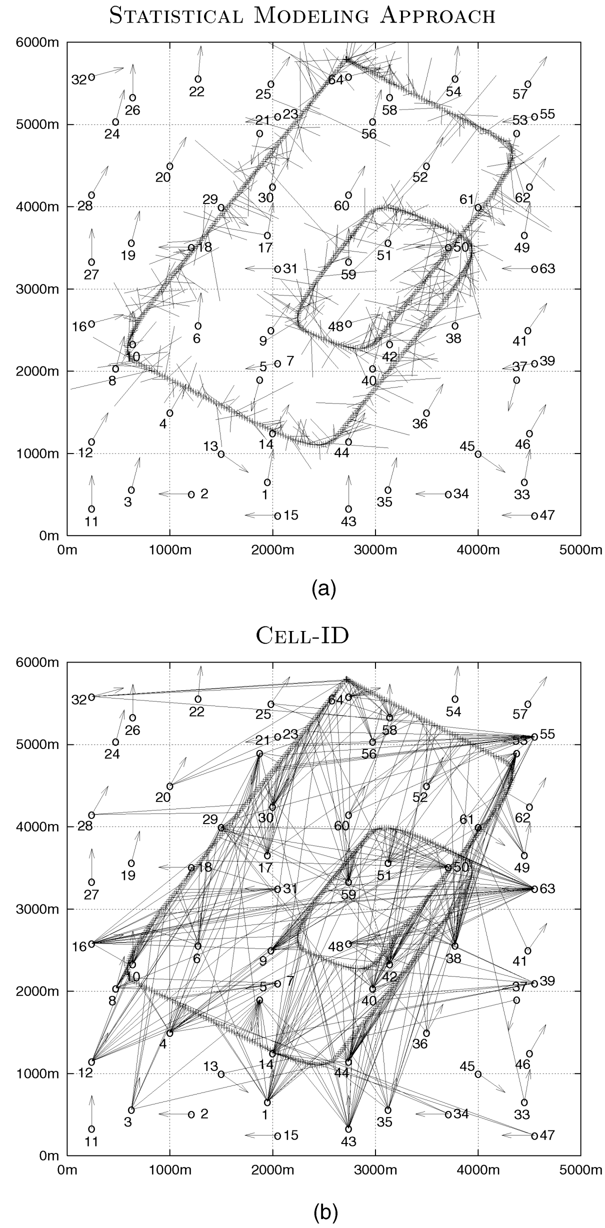

Fig. 6. A hypothetical network layout consisting of 64 transmitters. (The same layout is in the background of both images.) Arrows indicate direction of transmission and labels indicate channels. The superimposed lines correspond to error vectors on an imaginative trajectory of Example 4. The endpoints of each line are located at the correct and the estimated coordinates. (a) Location estimated using the statistical modeling approach. (b) Location estimated using the standard cell-ID method.

In the simulated test case with the presented location estimation method, 67 percent of the location errors were less than 320 meters, the average error being 279 meters. In our real-life trial, we obtained similar (even somewhat

11. To be precise, in the cell-ID method, the location of the *serving cell* is used as a location estimate. The transmitter with the strongest signal might not be chosen as the serving cell if the received power it signals varies much. However, in most cases, the serving call is associated with the transmitter with the strongest signal.

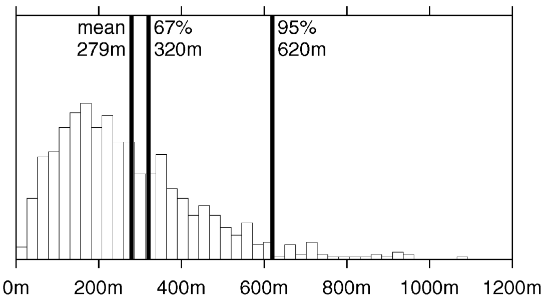

Fig. 7. Error distribution of Example 4 using the statistical modeling approach. Vertical lines denote the mean value and the 67th and 95th percentile.

better) accuracy. In the same setting, the corresponding values for the cell-ID method were 1,262 and 1,092 meters, respectively. It is possible to further enhance the accuracy by Kalman filtering or other tracking methods (see, e.g., [3]).

7 Conclusions

We have presented a statistical location estimation method based on a propagation prediction model. To conclude, we shall now repeat the central aspects of our work.

The advantages of the statistical modeling approach include certain types of flexibility that presented itself in the present work. In some mobile networks (e.g., GSM/GPRS), the observations made by a mobile unit in order to be located are associated with one set of channels whose received power is known and another set of channels whose received power is only bounded from above. We called the latter kind of partial observations *truncated*. The geometric approach provides no principled way of exploiting the information contained in the truncated observations. However, as we showed, the statistical modeling approach lends itself easily to exploiting any kind of observations, partial or complete.

The empirical performance of the presented method has been evaluated in a real-life environment provided by two industrial partners. However, as we are not at liberty to present the details of the proprietary field tests in this paper, we presented illustrative examples using an artificial network layout. The simulation environment was designed to correspond to the real-life environment as closely as possible so that the simulation results are practically equivalent to the real-world results. The resulting location estimation error was approximately 70-75 percent lower than the location estimation error obtained by the cell-ID method.

Although the empirical results are already encouraging, the proposed approach provides a theoretical framework for developing more sophisticated techniques with even better accuracy. For instance, the approach is by no means restricted to the use of received signal power measurements: One could also use angle or timing measurements as long as the used propagation model is capable of handling them. The flexibility of the approach also allows the fusion of different types of measurement results, for instance, received power and timing information.

TABLE 3
Summary Statistics for the Simulation of Example 4

	Statistical Modeling Approach	Cell-ID
mean	279 m	1092 m
median	237 m	773 m
67 %	320 m	1262 m
95 %	620 m	3108 m
maximum	1930 m	5015 m

Rows 67 percent and 95 percent denote the 67th and 95th percentile, respectively.

The propagation model considered in this work does not take into account the effect of the heterogeneity of the propagation environment. Location estimation based on empirical propagation prediction methods that avoid this weakness through onsite calibration is another interesting line of investigation for future research. Our preliminary empirical results [23] suggest that such solutions produce very accurate location estimation at the cost of onsite calibration, whereas the method presented in this paper is based on a more general-purpose model that is less accurate, but can be calibrated offsite with very little effort.

Acknowledgments

This research has been supported by the National Technology Agency and the Academy of Finland.

References

[1] J.B. Andersen, T.S. Rappaport, and S. Yoshida, "Propagation Measurements and Models for Wireless Communications Channels," *IEEE Comm. Magazine,* vol. 33, no. 1, pp. 42–49, 1995.
[2] P. Bahl, V.N. Padmanabhan, and A. Balachandran, "Enhancements to the RADAR User Location and Tracking System," Technical Report MSR-TR-00-12, Microsoft Research, Feb. 2000.
[3] Y. Bar-Shalom and X.R. Li, *Estimation and Tracking: Principles, Techniques and Software.* Boston: Artech House, 1993.
[4] J. Berger, *Statistical Decision Theory—Foundations, Concepts, and Methods.* New York: Springer-Verlag, 1980.
[5] G.E.P. Box and G.C. Tiao, *Bayesian Inference in Statistical Analysis.* Reading, Mass.: Addison-Wesley, 1973.
[6] P.J. Brown, J.D. Bovey, and X. Chen, "Context-Aware Applications: From the Laboratory to the Marketplace," *IEEE Personal Comm.,* vol. 4, no. 5, pp. 58–64, 1997.
[7] N. Bulusu, J. Heidemann, and D. Estrin, "GPS-Less Low Cost Outdoor Localization for Very Small Devices," *IEEE Personal Comm.,* no. 7, no. 5, pp. 28–34, 2000.
[8] P. Castro, P. Chiu, T. Kremenek, and R. Muntz, "A Probabilistic Room Location Service for Wireless Networked Environments," *Proc. Third Int'l Conf. Ubiquitous Computing (Ubicomp 2001),* Sept./Oct. 2001.
[9] G. Chen and D. Kotz, "A Survey of Context-Aware Mobile Computing Research," Technical Report TR2000-381, Dept. of Computer Science, Dartmouth College, Nov. 2000.
[10] *Digital Mobile Radio towards Future Generation Systems: COST-231 Final Report,* COST-231, E. Damosso and L.M. Correia, eds., 1998.
[11] M.H. DeGroot, *Probability and Statistics,* second ed. Reading, Mass.: Addison-Wesley, 1986.
[12] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm (with Discussion)," *J. Royal Statistical Soc., Series B,* pp. 1–38, 1977.
[13] B.H. Fleury and P.E. Leuthold, "Radiowave Propagation in Mobile Communications: An Overview of European Research," *IEEE Comm. Magazine,* vol. 34, no. 2, pp. 70–81, 1996.

[14] J. Hightower and G. Borriello, "Location Systems for Ubiquitous-Computing," *Computer,* special issue on location-aware computing, vol. 34, no. 8, pp. 57–66, 2001.
[15] O. Hilsenrath and M. Wax, "Radio Transmitter Location Finding for Wireless Communication Network Services and Management," US Patent 6,026,304, U.S. Wireless Corp., Jan. 1997.
[16] J.-M. Latapy, "GSM Mobile Station Locating," master's thesis, Norwegian Univ. Science and Technology, Trondheim, 1996.
[17] U. Leonhardt, "Supporting Location-Awareness in Open Distributed Systems," PhD thesis, Dept. of Computing, Imperial College, London, May 1998.
[18] G.J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions.* New York: John Wiley & Sons, 1997.
[19] R. Pattuelli and V. Zingarelli, "Precision of the Estimation of Area Coverage by Planning Tools in Cellular Systems," *IEEE Personal Comm.,* vol. 7, no. 3, pp. 50–53, 2000.
[20] T.S. Rappaport, *Wireless Communications: Principles & Practice.* Prentice Hall, 1996.
[21] T.S. Rappaport, J.H. Reed, and B.D. Woerner, "Position Location Using Wireless Communications on Highways of the Future," *IEEE Comm. Magazine,* vol. 34, pp. 33–41, 1996.
[22] K. Rizk, "Propagation in Microcellular and Small Cell Urban Environment," PhD thesis, Swiss Federal Inst. of Technology of Lausanne, 1997.
[23] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen, "A Probabilistic Approach to WLAN User Location Estimation," *Int'l J. Wireless Information Networks,* to appear.
[24] Z. Salcic, "AGPCS—An Automatic GSM-Based Positioning and Communication System," *Proc. Second Ann. Conf. GeoComputation and SIRC,* pp. 15–22, 1997.
[25] A. Savvides, C.C. Han, and M.B. Srivastava, "Dynamic Fine-Grained Localization in Ad-Hoc Wireless Sensor Networks," *Proc. Seventh Int'l Conf. Mobile Computing and Networking (MobiCom 2001),* July 2001.
[26] J. Syrjärinne, "Studies of Modern Techniques for Personal Positioning," D.Sc. thesis, Tampere Univ. of Technology, 2001.
[27] C. Tepedelenlioglu, A. Abdi, G.B. Giannakis, and M. Kaveh, "Estimation of Doppler Spread and Signal Strength in Mobile Communications with Applications to Handoff and Adaptive Transmission," *Wireless Comm. and Mobile Computing,* vol. 1, no. 2, pp. 221–242, 2001.
[28] S. Thrun, D. Fox, W. Burgard, and F. Dellaert, "Robust Monte Carlo Localization for Mobile Robots," *Artificial Intelligence,* vol. 128, nos. 1-2, pp. 99–141, 2001.
[29] T. Tonteri, "A Statistical Modeling Approach to Location Estimation," master's thesis, Dept. Computer Science, Univ. of Helsinki, May 2001. Available at http://cosco.hiit.fi/Articles/ttthesis.ps.gz.
[30] S.Y. Willassen, "A Method for Implementing Mobile Station Location in GSM," diploma thesis, Norwegian Univ. of Science and Technology, Trondheim, Dec. 1998.
[31] G. Wölfle and F.M. Landstorfer, "Prediction of the Field Strength inside Buildings with Empirical, Neural, and Ray-Optical Prediction Models," *Proc. Seventh COST-259 MCM-Meeting,* 1999.

Teemu T. Roos (formerly Tonteri) received the MSc degree from the University of Helsinki in 2001. He is currently pursuing the PhD degree in computer science. In 1999, he spent a period of three months at the Centro Studi e Laboratori Telecomunicazioni in Turin, Italy. Since 1999, he has been with the Complex Systems Computation Group (Helsinki Institute for Information Technology). His current research interests are primarily in Bayesian and information-theoretic data analysis and machine learning.

Petri J. Myllymäki received the MSc degree in computer science from the University of Helsinki in 1991 and the PhD degree in 1995. He is one of the cofounders of the Complex Systems Computation (CoSCo) Research Group and his current special research interests are in Bayesian and information-theoretic modeling, in particular, with models such as Bayesian networks or finite mixture models and in stochastic optimization methods. He has been an editorial board member, program committee member, and reviewer for several international scientific journals and conferences and he has published more than 50 scientific articles in his research area. He has also worked as a project manager in numerous applied research projects with companies, like Nokia, Kone, StoraEnso, ABB, and AlmaMedia, and the cooperation has lead to a number of fielded applications and patent applications. Dr. Myllymäki is currently working as a research fellow for the Academy of Finland.

Henry R. Tirri received the BSc, MSc, and PhD degrees in computer science from the University of Helsinki. Dr. Tirri's academic experience includes both research and teaching positions at the University of Helsinki, University of Texas at Austin, Microelectronics and Computer Technology Corporation (MCC), AT&T Bell Laboratories, Purdue University, NASA Ames Research Center, and Stanford University, where he is currently a visiting professor for 2001-2002. Since 1998, he has been a professor of computer science at the University of Helsinki, and an adjunct professor of computational engineering at the Helsinki University of Technology. He is currently engaged in research on various aspects of learning and adaptation in artificial systems. In particular, he is interested in methods for building predictive models from data using Bayesian and information-theoretic approach. Dr. Tirri was a member of the editorial board of the *Computer Journal* (Oxford University Press) for the period 1995-1998. He was a member of the IT-development Board of the Safety Technology Authority for the period 2000-2001, and is currently a member of the Ministry of Education Board for Learning Environments as part of the National Information Technology Strategy in Education 2000-2004 and a member of the Scientific Advisory Board of WSOY Publishing Corporation. He is a member of various professional societies such as ACM, IEEE, the IEEE Computer Society, International Neural Network Society, AAAI, International Society for Bayesian Analysis, and the American Educational Research Association.