LUND UNIVERSITY

Pathloss Estimation Techniques for Incomplete Channel Measurement Data

Abbas, Taimoor; Gustafson, Carl; Tufvesson, Fredrik

2014

Citation for published version (APA):
Abbas, T., Gustafson, C., & Tufvesson, F. (2014). *Pathloss Estimation Techniques for Incomplete Channel Measurement Data*. Paper presented at COST IC1004 10th Management Committee and Scientific Meeting, Aalborg, Denmark.

Total number of authors:
3

SOURCE: Dept. of Electrical and Information Technology, Lund University, Lund, Sweden

Pathloss Estimation Techniques for Incomplete Channel

Measurement Data

Taimoor Abbas
Department of Electrical and Information Technology
Lund University
Lund, Sweden
Phone: + 46 46 2224902
Fax: + 46 46 129948
Email: Taimoor.abbas@eit.lth.se

Pathloss Estimation Techniques for Incomplete Channel Measurement Data

Taimoor Abbas, Carl Gustafson, and Fredrik Tufvesson
Dept. of Electrical and Information Technology, Lund University, Lund, Sweden
firstname.lastname@eit.lth.se

Abstract—The pathloss exponent and the variance of the large-scale fading are two parameters that are of great importance when modeling or characterizing wireless propagation channels. The pathloss is typically modeled using a single-slope log-distance power law model, whereas the large-scale fading is modeled using a log-normal distribution. In practice, the received signal is affected by noise and it might also be corrupted by interference from other active transmitters that are transmitting in the same frequency band. Estimating the pathloss exponent and large scale fading without considering the effects of the noise and interference, can lead to erroneous results. In this paper, we show that the path loss and large scale fading estimates can be improved if the effects of samples located below the noise floor are taken into account in the estimation step. When the number of such samples is known, then the pathloss exponent and standard deviation of the large scale fading can be iteratively computed using maximum likelihood estimation from incomplete data via the expectation maximization (EM) algorithm. Alternatively, if the number of samples below the noise floor is unknown, we show that the pathloss and large scale fading parameters can be estimated based on a likelihood expression for a truncated normal distribution.

I. Introduction

Pathloss is the expected (mean) loss at a certain distance between the transmitter (TX) and the receiver (RX). A number of pathloss models have been developed for a variety of wireless communication systems, e.g., cellular systems, bluetooth, wifi vehicle-to-vehicle dedicted short range communications and mm-wave point-to-point communications, working over different frequency bands ranging from hundreds of MHz to tens of GHz [1]–[4]. These models have widely been used for the prediction and simulation of signal strengths for given TX-RX separation distances. Typically, pathloss models are developed with the help of channel measurements in realistic scenarios. The model parameters estimated from the measurement data are thus typically valid only for a particular frequency range, antenna arrangement, and environment for the target scenario.

Measurement data often has limitations in one way or another and thus there are a number of associated challenges, which often makes the estimation and modeling of the pathloss exponent from the channel measurement data non-trivial. Therefore, special considerations must be taken into account when modeling the pathloss exponent by measurement data. The examples of associated challenges are;

1) Typically samples are recorded at equi-spaced time or distance intervals. This makes the distribution of data samples to be non-uniform along *logarithmic* distance, i.e., the data samples have higher concentration at larger *logarithmic* distances, implying that the standard least-square (LS) or mean-square-error (MSE) estimation approaches will have a high weight for large distances and low weight for shorter distances. This problem can be solved by using weighted samples for improved prediction where weights are calculated w.r.t. the logarithmic sampling density [5]. It is very important to explicitly mention that if the *linear* or *logarithmic* distance sampling is used for modeling, because sample distribution is different for both cases and may lead to different estimates for the same data set.

2) The pathloss exponent and the standard deviation of the large scale fading are random variates that vary from location to location as well for different antenna heights and frequencies. It is important to understand these variations and their physical motivation.

3) In practice, the observation of the received signal power at the receiver is limited by the system noise, i.e., the signals with power below the noise floor can not be measured properly. Moreover, the received signal power might also be corrupted by interference from other active transmitters that are transmitting in the same frequency band. Estimating the pathloss exponent and large scale fading based on such data set, without considering the effects of the noise floor and interference, can lead to erroneous results. In this paper it is shown that the path loss and large scale fading estimates can be improved if the missing samples, which are located below the noise floor, are taken into account in the estimation step. Two different methods are provided that incorporate missing samples to improve the parameter estimation, first when the number of missing samples is unknown and second when the number of missing samples is known.

In section II we describe the pathloss and basic modeling assumptions. Section III explains the parameter estimation methods. Section IV presents results for the synthetic as well as for the channel measurement data. Finally, Section V summarizes the discussion.

II. Pathloss modelling

A simple log-distance power law [6] is often used to model the path loss to predict the reliable communication range between the TX and the RX. The generic form of this log-distance power law path loss model is given by,

$$PL(\boldsymbol{d}) = PL_0 + 10n \log_{10}\left(\frac{\boldsymbol{d}}{d_0}\right) + X_\sigma; d \geq d_0 \quad (1)$$

where PL_0 is an estimated (measured) reference level or $PL_0 = 20log_{10}(4\pi d_0/\lambda)$ the theoretical path loss at a reference distance d_0 in dB calculated using free-space propagation model. Furthermore λ is the wavelength in meters, $\boldsymbol{d}$ is the vector of distances between the TX and the RX, n is the pathloss exponent, and X_σ is a random variable to represent large scale fading about the distance dependent pathloss, respectively. If the effect of small scale fading is removed from the data set by averaging the data over the time samples corresponding to a wide-sense-stationary region, then the variations of the large-scale fading are typically modeled using a zero-mean Gaussian distribution with standard deviation σ, i.e., $X_\sigma \sim \mathcal{N}(0, \sigma^2)$. Hence, $PL(d) \sim \mathcal{N}(\mu(d), \sigma^2)$ with distance dependent deterministic expected value,

$$\mu(\boldsymbol{d}) = PL_0 + 10n \log_{10}\left(\frac{\boldsymbol{d}}{d_0}\right); \qquad (2)$$

The pathloss exponent n is an environment dependent parameter commonly provided in modeling papers, which is determined by field measurements. Usually, n is estimated by simple linear regression of $10 \log_{10}(d)$ to the measured power values in dB such that the mean square error (MSE) between the measured and modeled points is minimized.

For peer-to-peer communication in line-of-sight (LOS) propagation conditions it is often observed that a dual-slope model based on two-ray ground model, as stated in [7], can represent the measurement data more accurately. We thus characterize a dual-slope model as a piecewise-linear model with the assumption that the power decays with path loss exponent n_1 until the breakpoint distance (d_b) and from there it decays with path loss exponent n_2. The dual-slope model is given by,

$$PL(d) = \begin{cases} PL_0 + 10n_1 \log_{10}\left(\frac{d}{d_0}\right) + X_{\sigma 1}, & \text{if } d_0 \leq d \leq d_b \\ PL_0 + 10n_1 \log_{10}\left(\frac{d_b}{d_0}\right) + & \text{if } d > d_b \\ 10n_2 \log_{10}\left(\frac{d}{d_b}\right) + X_{\sigma 2}. \end{cases}$$
$$(3)$$

The typical flat earth model consider d_b as the distance at which the first Fresnel zone touches the ground or the first ground reflection has traveled $d_b + \lambda/4$ to reach RX. The d_b can be calculated as, $d_b = \frac{4h_{TX}h_{RX} - \lambda^2/4}{\lambda}$, where λ is the wavelength at carrier frequency f_c, and h_{TX} and h_{RX} are the height of the TX and RX antennas, respectively. $X_{\sigma 1}$ and $X_{\sigma 2}$ represent large scale fading before and after the break point. For $d < d_b$, this model is the same as the log-distance power law in (1).

III. ESTIMATION

To completely model the pathloss for a given data set, we wish to estimate three main parameters of (1) or (3), i.e., n, PL_0 and σ^2. The data under consideration is implicitly assumed to be Gaussian distributed because the large-scale or large scale fading on top of distance dependent deterministic pathloss is Gaussian, as mentioned above. Here, we will consider parameter estimation only for (1) where the result

can easily be extended for (3). Using (1) the data set can be modeled as,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \qquad (4)$$

where
$$\mathbf{y} = \begin{pmatrix} PL(d_1) \\ PL(d_2) \\ \vdots \\ PL(d_L) \end{pmatrix}, \ \mathbf{X} = \begin{pmatrix} 1 & 10log_{10}(d_1) \\ 1 & 10log_{10}d_2) \\ \vdots & \vdots \\ 1 & 10log_{10}(d_L) \end{pmatrix}, \ \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_L \end{pmatrix}$$
and $\boldsymbol{\alpha} = \begin{pmatrix} PL_0 \\ n \end{pmatrix}$

By applying ordinary least squares, the parameter $\boldsymbol{\alpha}$ can be estimated as

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}. \qquad (5)$$

Using the estimates contained in $\hat{\boldsymbol{\alpha}}$, the variance σ^2 can be estimated as

$$\hat{\sigma}^2 = \frac{1}{L-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}). \qquad (6)$$

A. Least-Square (LS) Estimation for Truncated Data

In order to estimate the pathloss exponent and fading variance of data truncated by a noise floor, with an *unknown* number of missing samples, it is possible to base the estimation on a truncated normal distribution. Under this assumption, each data observation follows a truncated normal distribution,

$$y_i \sim N_c(\mathbf{x}_i\boldsymbol{\alpha}, \sigma^2), \qquad (7)$$

where c denotes the noise floor level. The likelihood expression for this distribution is given by

$$l(\sigma, \boldsymbol{\alpha}) = \prod_{i=1}^{n} \frac{\frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\boldsymbol{\alpha}}{\sigma}\right)}{1 - \Phi\left(\frac{c - \mathbf{x}_i\boldsymbol{\alpha}}{\sigma}\right)}, \qquad (8)$$

where $\phi(\cdot)$ is the probability density function for the standard normal distribution and $\Phi(\cdot)$ is its cumulative distribution function, and hence the likelihood can be written as

$$l(\sigma, \boldsymbol{\alpha}) = \prod_{i=1}^{n} \frac{\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{1}{2\sigma^2}(y_i - \mathbf{x}_i\boldsymbol{\alpha})^2)}{\frac{1}{2}(1 - \text{erf}(\frac{c - \mathbf{x}_i\boldsymbol{\alpha}}{\sqrt{2}\sigma}))}, \qquad (9)$$

where $\text{erf}(\cdot)$ is the error function. Using the log-likelihood $L(\sigma, \boldsymbol{\alpha}) = \ln[l(\sigma, \boldsymbol{\alpha})]$, the parameters σ and $\boldsymbol{\alpha}$ are estimated using

$$\underset{\sigma, \boldsymbol{\alpha}}{\arg\min}\{-L(\sigma, \boldsymbol{\alpha})\}. \qquad (10)$$

which is easily solved by numerical optimization of $\boldsymbol{\alpha}$ and σ.

B. Expectation-Maximization (EM) Algorithm for Truncated Data

To estimate the pathloss exponent and variance of the incomplete data with a *known* number of missing samples, we make use of the expectation maximization (EM) algorithm by Dempster *et al.* in [8]. The iterative EM algorithm is developed to estimate the mean and standard deviation of a truncated

data set by maximizing the likelihood function based upon observed and missing samples.

The measured channel gains have a distance dependent mean and a certain standard deviation along the distance. Therefore, we divide the measured distance dependent channel gains into log-spaced distance bins so that the mean and variance for each bin can be estimated independently.

The data $\mathbf{y}' \sim \mathcal{N}(\mu, \sigma^2)$ associated to each distance bin is assumed to be Gaussian distributed, whereas the true mean μ and standard deviation σ are unknown. Let $y_1', y_2', ..., y_{k-l}'$ be the detected samples of $\mathbf{y}'$ and $y_{(k-l)+1}', y_{(k-l)+2}', ..., y_k'$ be the l undetected samples, if exist, lying below the noise floor c. The mean, μ_0, and standard deviation σ_0 of observed $(k-l)$ truncated data values, without considering missing samples, can then be estimated as,

$$\mu_0 = \sum_{i=1}^{k-l} \frac{y_i'}{(k-l)}, \tag{11}$$

and

$$\sigma_0 = \sum_{i=1}^{k-l} \frac{(y_i' - \mu_0^2)}{(k-l)}. \tag{12}$$

The likelihood expression for this truncated distribution is given by

$$\ln L(\mathbf{y}', \mu, \sigma) = l \ln \Phi(L) - (k-l) \ln \sigma - \sum_{i=1}^{k-l} (x_i - \mu)^2 / 2\sigma^2, \tag{13}$$

where $\Phi(L)$ is the cumulative distribution function of $L = c - \mu/\sigma$ representing the probability that an observation is less than c.

For such a left-truncated data set with a known threshold and known number of missing samples, The EM-algorithm of [8] makes use of these observed $(k-l)$ samples to calculate initial estimates of μ and σ. In each iteration the expectation of the conditional likelihood function of the complete data is maximized based on the type of truncation such as left-truncation in our case. At $(j+1)^{th}$ iteration, the estimates of μ and σ are given by [9], as follows

$$\hat{\mu}_{j+1} = \frac{\left[\sum_{i=1}^{k-l} x_i + \sum_{i=k-l}^{k} E_j(x_i | x_i \leq c) \right]}{k}, \tag{14}$$

$$\hat{\sigma}_{j+1}^2 = \frac{\left[\sum_{i=1}^{k-l} x_i - \hat{\mu_j}^2 + \sum_{i=1}^{l} E_j((x_i - \mu_j)^2 | x_i \leq c) \right]}{k-1}, \tag{15}$$

where

$$E_j(x_i | x_i \leq c) = \hat{\mu_j} - \hat{\sigma_j}[\phi(L_j)/\Phi(L_j)], \tag{16}$$

$$E_j((x_i - \mu_j)^2 | x_i \leq N_0) = \hat{\sigma_j}^2 (1 - L_j[\phi(L_j)/\Phi(L_j)]), \tag{17}$$

with

$$L_j = (c - \hat{\mu}_j)/\hat{\sigma}_j. \tag{18}$$

Hence, the iterative EM-algorithm improves the estimation of μ and σ in each iteration, where all of the non-detected
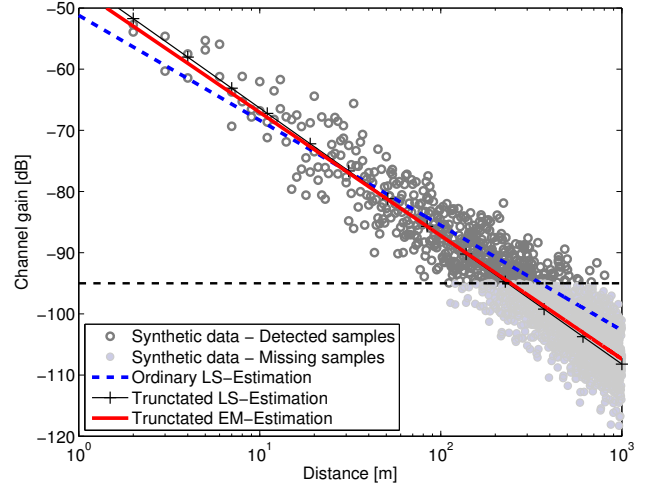

Fig. 1. Path loss estimation based on synthetic data with a noise threshold $c = lnN_0 = -95$ dB.

samples are replaced by the same conditional expected value as given by (16).

IV. ESTIMATION EXAMPLES

The significance of these estimation methods is shown by the help of examples where the path loss parameter estimation is performed for synthetic as well as for measured data.

A. Synthetic data

To make sure if both the estimation methods work fine we first consider an example using synthetic data, where data is generated according to (1) with known channel parameters.

The path loss parameters used to generate 2000 samples of the synthetic data distributed uniformly along the distance between $0 - 1000$ m are; $f_c = 5.6$ GHz, $n = 2$, $\sigma = 4$ dB, and $d_0 = 1$ m. The noise threshold $c = 10log_{10}(N_0) = -95$ dB is used to truncate the data, where N_0 is the non-dB value of noise floor. It is assumed that the samples lying below c are missing or un-detected, though the exact count of those samples is available.

The parameter estimation is done in two steps: 1) The parameters of truncated data are estimated without considering the effect of the noise threshold c by ordinary least-square (LS) estimation as given in (5) and (6). 2) The parameters of the truncated data are estimated by considering the effect of noise threshold. Both of the above described estimation methods, LS-estimation for truncated data and EM-Algorithm for truncated data, are tested while taking into account the unknown number of missing samples and known number of missing samples, respectively.

The synthetic data is shown in Fig. 1 with the estimated path loss slopes. It can be observed that the estimates are less accurate when using an ordinary LS-estimation for truncated data without considering the effect of missing samples. However, when the missing samples are taken into account the estimated values are very close to the true value. This means that, in the prior case, the estimated values of the received power will

be larger than the true values. The path loss model based on incorrect parameter estimates can thus lead to erroneous results when performing network simulations.

The true and estimated values are listed in Table I for each of the above described parameter estimation methods. The table shows that, for this specific example, when estimation is performed without considering the missing samples, n is over-estimated whereas the standard deviation is underestimated. However, when the effects of the noise floor or the number of missing samples are considered, then the estimation results are very similar to the true values. For the synthetic data in this example, both the truncated LS-estimation and the truncated EM-estimation methods have similar performance.

TABLE I
PATH LOSS PARAMETER ESTIMATION FOR SYNTHETIC DATA

	n	PL_0 [dB]	σ [dB]
True values	2	47.4	4
Ordinary LS-estimation	1.61	-53.5	3.46
Truncated LS-estimation	1.91	-48.8	3.86
Truncted EM-estimation	2.0	-46.6	3.84

B. Measurement data

The measurement data used here is collected for vehicle-to-vehicle (V2V) communication channel characterization at $5.6\,$GHz, for details see [10]. The path loss model parameter for the measurement data are estimated in the same way as it is done in Section IV-A. The measurement data for two different situation line-of-sight (LOS), when both the TX and RX has visual sight between them, and obstructed-LOS (OLOS), when a large object such as building, or another vehicle partly or completely obstruct the LOS between the TX and RX. Both the situations are significantly different from each other and receiver power in LOS is typically better than that in OLOS. These differences in the received power result in different number of missing samples is both scenarios for a fixed noise threshold c, i.e., more samples will be missing in OLOS situation.

The measurement data and the estimated path loss slopes for all three methods are shown in Fig. 2, where as the estimated parameters are listed in Table II. Figure. 2(a) shows that all three estimation methods have nearly similar parameter estimates for the LOS measurement data, it is due to the fact that only a few samples are missing. However, in Fig. 2(b) this is not the case, where the ordinary LS estimation clearly overestimates the pathloss exponent n and underestimates σ. The results for truncated LS-estimation are not so different from the ordinary LS-estimation because in OLOS case the number of undetected samples, 2187, is almost equal to the amount to detected samples 2722. Since, truncated LS-estimation does not take the number of missing samples into account it leads to unrealistic parameter estimates. The second method, truncated EM-estimation, however takes the number of missing samples into account and thus give more realistic estimates of the statistics.
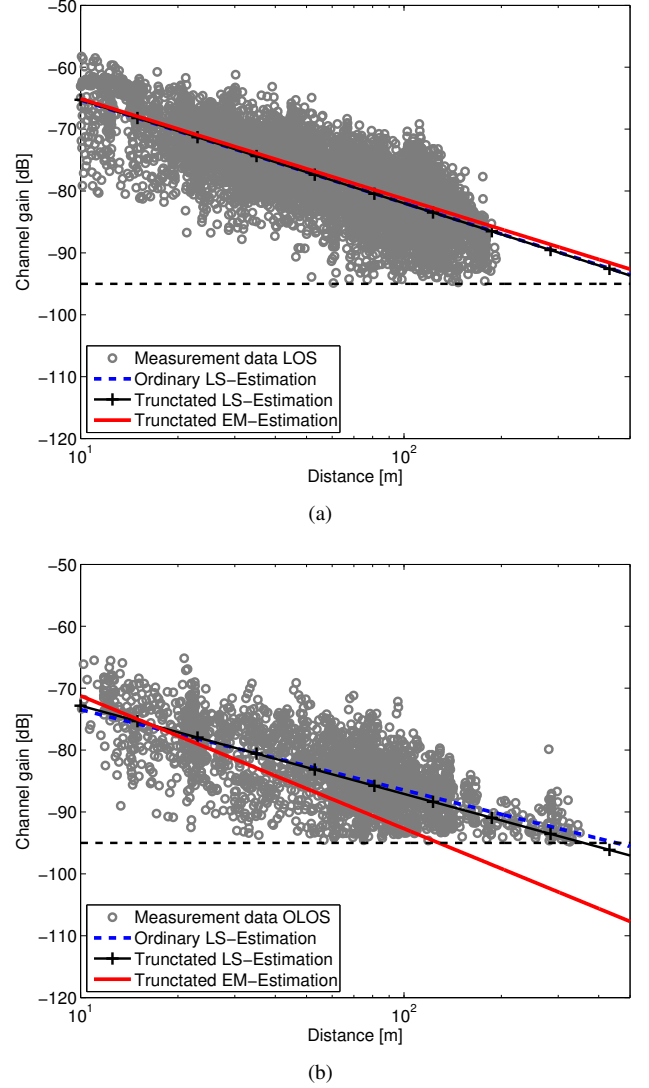

(a)

(b)

Fig. 2. Path loss estimation based on measurement data, recorded for vehicle-to-vehicle (V2V) communication in line-of-sight (LOS) and obstructed-line-of-sight (OLOS) situation, with a noise threshold $c = 10log_{10}(N_0) = -95\,$dB.

TABLE II
PATH LOSS PARAMETER ESTIMATION FOR MEASUREMENT DATA

		n	PL_0 [dB]	σ [dB]
LOS	Ordinary LS-estimation	1.66	-65.3	4.3
	Truncated LS-estimation	1.67	-65.2	4.3
	Truncted EM-estimation	1.63	-65.1	4.4
OLOS	Ordinary LS-estimation	1.3	-73.5	4.4
	Truncated LS-estimation	1.43	-72.8	4.6
	Truncted EM-estimation	2.1	-71.2	7.2

V. Summary and Conclusions

In this paper, we present two novel ways of estimating the pathloss model parameters for the measurement data set, which is incomplete or truncated due to the influence of the noise floor. We show that the estimates can be improved if the effects of the noise floor are taken into account in the estimation steps. In the truncated least-square (LS) estimation based method, when the number of samples below the noise floor is unknown, the noise floor is modeled with the help of a truncated normal distribution and the parameters are estimated based on a likelihood expression for this distribution. In the second method, when the number of samples above and below the noise floor is known, then both the noise floor information as well as the number of samples below noise floor are used to estimate the model parameters using the expectation maximization (EM) algorithm for truncated data. It is found that the EM-algorithm based method gives better estimates compared to LS-estimation based method, given that a large number of samples are below the noise floor and that number is known. When the number of samples below the noise floor is unknown, the trucated LS-estimation gives better parameter estimates than the ordinary LS method. All three methods give similar estimates when there are no samples below the noise floor.

References

[1] M. Hatay, "Empirical formula for propagation loss in land mobile radio services," *Vehicular Technology, IEEE Transactions on*, vol. 29, no. 3, pp. 317–325, Aug 1980.

[2] V. Erceg, L. Greenstein, S. Tjandra, S. Parkoff, A. Gupta, B. Kulic, A. Julius, and R. Bianchi, "An empirically based path loss model for wireless channels in suburban environments," *Selected Areas in Communications, IEEE Journal on*, vol. 17, no. 7, pp. 1205–1211, Jul 1999.

[3] G. Mao, B. D. O. Anderson, and B. Fidan, "Wsn06-4: Online calibration of path loss exponent in wireless sensor networks," in *Global Telecommunications Conference, 2006. GLOBECOM '06. IEEE*, Nov 2006, pp. 1–6.

[4] A. F. Molisch, F. Tufvesson, J. Karedal, and C. F. Mecklenbräuker, "A survey on vehicle-to-vehicle propagation channels," in *IEEE Wireless Commun. Mag.*, vol. 16, no. 6, 2009, pp. 12–22.

[5] J. Karedal, "Measurement-based modeling of wireless propagation channels - MIMO and UWB," Ph.D. dissertation, Department of Electrical and information technology, Lund University, Sweden, Feb. 2009.

[6] A. Molisch, *Wireless Communications*. Chichester, West Sussex, UK: IEEE Press-Wiley, 2005.

[7] L. Cheng, B. Henty, D. Stancil, F. Bai, and P. Mudalige, "Mobile vehicle-to-vehicle narrow-band channel measurement and characterization of the 5.9 GHz dedicated short range communication (DSRC) frequency band," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 8, pp. 1501 –1516, Oct. 2007.

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 39, no. 1, pp. 1–38, 1977.

[9] R. H. Shumway, A. S. Azari, and P. Johnson, "Estimating mean concentrations under transformation for environmental data with detection limits," *Technometrics*, vol. 31, no. 3, pp. pp. 347–356, 1989. [Online]. Available: http://www.jstor.org/stable/3556144

[10] T. Abbas, F. Tufvesson, and J. Karedal, "Measurement based shadow fading model for vehicle-to-vehicle network simulations," *ArXiv e-prints*, Mar. 2012.