LUND UNIVERSITY

Tobit Maximum-likelihood estimation of Censored Pathloss Data

Gustafson, Carl; Abbas, Taimoor; Bolin, David; Tufvesson, Fredrik

2015

[Link to publication](#)

Tobit Maximum-likelihood estimation of Censored Pathloss Data

Carl Gustafson, Taimoor Abbas, David Bolin and Fredrik Tufvesson

Abstract—**Pathloss is typically modeled using a log-distance power law with a large-scale fading term that is log-normal. However, the received signal is affected by the dynamic range and noise floor of the measurement system used to sound the channel, which can cause measurement samples to be truncated or censored. If the information about the censored samples are not included in the estimation method, as in ordinary least squares estimation, it can result in biased estimation of both the pathloss exponent and the large scale fading. This is solved by applying a Tobit maximum-likelihood estimator, which provides consistent estimates for the pathloss parameters. This technical report provides information about the Tobit maximum-likelihood estimator estimator and its asymptotic variance under certain conditions.**

Keywords—*Pathloss, maximum-likelihood estimation, ordinary least squares, censored data, truncated data, vehicular communication.*

I. Introduction

Pathloss describes the expected loss in received power as a function of the transmitter (Tx) and receiver (Rx) separation distance and the effects of random large scale fading. It includes losses due to the expansion of the radio wave front in space as well as losses due to reflection, scattering, diffraction and penetration. A number of pathloss models have been developed for a variety of wireless communication systems, e.g., cellular systems, Bluetooth, Wi-Fi, vehicle-to-vehicle communications, and, mm-wave point-to-point communications, operating 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. Pathloss models are often developed based on channel measurements in realistic user scenarios. The model parameters estimated from measurement data are thus typically valid only for a particular frequency range, antenna arrangement, and environment for the targeted user scenario.

However, 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. In many vehicle-to-vehicle measurements, this limitation due to the system noise is often present at longer distances [5]–[7]. Also, in mm-wave measurements, the pathloss

values are in general larger than at lower frequencies, which effectively can reduce the range in which the data is unaffected by the noise floor. Due to the limited dynamic range of the measurement system, sample data might be *truncated*, whereby all data above or below a certain range are immeasurable, or *censored*, meaning that all data above or below a certain range are counted, but not measured. Estimation of the cluster decay and cluster fading based on truncated data has previously been addressed in [8]. For clusters, the data is modeled as truncated, since it is generally impossible to measure or count clusters that are below the noise floor. However, in pathloss measurements, since the distances where the received power falls below the noise floor are known, it is possible to model the data as being censored. Estimating statistical parameters without considering the effects of censored or truncated data samples, can lead to erroneous results. The fact that this can be a problem for pathloss data is acknowledged in [6], however, the authors do not give any detailed information on how to solve this issue. In this technical report, we discuss the use of a Tobit model [9] for censored pathloss data and a maximum-likelihood (ML) method for the estimation of pathloss parameters [10].

II. Pathloss Modeling

Path loss is often modeled by a log-distance power law plus a large scale fading term [11]. In units of dB this can be written as

$$PL(d) = PL(d_0) + 10n\log_{10}\left(\frac{d}{d_0}\right) + \Psi_\sigma, \quad d \geq d_0, \quad (1)$$

where d is the distance, n is the pathloss exponent, $PL(d_0)$ is the pathloss at a reference distance of d_0 and Ψ_σ is a random variable that describes the large-scale fading about the distance-dependent mean pathloss. For measurement data, it is here assumed that the effects of small scale fading is removed from the data set. It is also assumed that the peak value of the aggregated antenna gain is removed from the measurement data [12]. Ideally, the variation of the aggregated antenna gain should be small, so that it does not affect the measured large-scale fading too much. The large-scale fading term is usually modeled by a log-normal distribution, which in the dB-domain corresponds to a zero-mean Gaussian distribution with standard deviation σ, i.e., $\Psi_\sigma \sim \mathcal{N}(0, \sigma^2)$. Hence, the pathloss is normally distributed with a distance dependent expected value, $PL(d) \sim \mathcal{N}(\mu(d), \sigma^2)$, where

$$\mu(d) = PL(d_0) + 10n\log_{10}\left(\frac{d}{d_0}\right). \quad (2)$$

The reference value $PL(d_0)$ can be estimated based on measurement data, or based on reference measurements at this distance. For line-of-sight (LOS) scenarios, it is sometimes deterministically modeled based on the free-space pathloss, as

$$PL(d_0) = 20\log_{10}\left(\frac{4\pi d_0}{\lambda}\right). \tag{3}$$

Here λ is the wavelength at the given frequency. Here, it is worth noting that the approach of using the deterministic reference value of Eq. (3) only provides theoretically correct results if the pathloss exponent is equal to 2. If the pathloss exponent is *not* equal to 2, but Eq. (3) is used to determine the reference value, the data model of Eq. 1 is inconsistent, as it depends on the choice of the reference distance d_0. For non-line-of-sight (NLOS) scenarios, it is clear that the free-space equation (3) does not hold, which means that the reference value in this case must be determined in another fashion. Due to the above, it is preferable to use actual measurements of the reference level, or, to estimate it based on the measurement data. In some cases, it might be difficult to produce reliable measurements of the reference value scenarios due to practical reasons, especially considering that it might be hard to produce a large number of uncorrelated measurement samples exactly at d_0.

III. Estimation by Ordinary Least Squares

To completely model the pathloss and large-scale fading for a given data set, we wish to estimate the three parameters of (1), i.e., n, $PL(d_0)$ and σ^2. The data under consideration is implicitly assumed to be Gaussian since Ψ_σ is Gaussian in the dB domain. Using (1) the data set for L path loss measurements can be written as,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \tag{4}$$

where

$$\begin{aligned}
\mathbf{y} &= [PL(\mathbf{d}/d_0)]_{L\times 1}, \\
\mathbf{X} &= [\mathbf{1} \quad 10\log_{10}(\mathbf{d}/d_0)]_{L\times 2}, \\
\boldsymbol{\alpha} &= [PL(d_0) \ n]^T.
\end{aligned} \tag{5}$$

The term $\boldsymbol{\epsilon} = [\boldsymbol{\Psi}_\sigma]_{L\times 1}$ is a row vector describing the large-scale fading term for each of the L different pathloss samples.

When there are no censored samples, the parameters of the log-distance power law can be estimated by applying ordinary least squares (OLS). The parameter $\boldsymbol{\alpha}$ is then estimated as[1]

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}. \tag{6}$$

The variance of the large-scale fading, σ^2, can then be estimated as

$$\hat{\sigma}^2 = \frac{1}{L-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}). \tag{7}$$

---

[1] As the variance σ^2 is assumed to be independent of delay, weighted least squares (WLS) are not applied. However, we note that WLS could be of use for cases when σ^2 is being modeled with a distance dependence.

The estimate $\hat{\boldsymbol{\alpha}}$ is Gaussian,

$$\hat{\alpha}_j \sim \mathcal{N}\left(\alpha_j, \sigma^2(\mathbf{X}^T\mathbf{X})_{jj}^{-1}\right), \ j = 1, 2, \tag{8}$$

which alternatively can be expressed as

$$\begin{aligned}
\hat{PL}(d_0) &= \hat{\alpha}_1 \sim \mathcal{N}\left(PL(d_0), \sigma^2\left(L^{-1} + \bar{\mathbf{x}}^2 S_{xx}^{-1}\right)\right), \\
\hat{n} &= \hat{\alpha}_2 \sim \mathcal{N}\left(n, \sigma^2 S_{xx}^{-1}\right),
\end{aligned} \tag{9}$$

where

$$\begin{aligned}
\bar{\mathbf{x}} &= \frac{1}{L}\sum_{l=1}^{L} 10\log_{10}(d_l/d_0), \\
S_{xx} &= \sum_{l=1}^{L}\left(10\log_{10}(d_l/d_0) - \bar{\mathbf{x}}\right)^2.
\end{aligned} \tag{10}$$

Using Eq. (9), standard errors[2] for $\hat{n}$ and $\hat{PL}(d_0)$ can be found by replacing the unknown standard deviation of the large scale fading, σ, by its estimate, $\hat{\sigma}$, which gives

$$\begin{aligned}
\hat{\text{SE}}(\hat{n}) &= \hat{\sigma}\sqrt{S_{xx}^{-1}}, \\
\hat{\text{SE}}(\hat{PL}(d_0)) &= \hat{\sigma}\sqrt{L^{-1} + \bar{\mathbf{x}}^2 S_{xx}^{-1}}.
\end{aligned} \tag{11}$$

The standard errors are useful for evaluating the accuracy of the estimated parameters. However, it should be stressed that these standard errors only applies when the data actually follows the log-distance power law model with a large-scale fading variance that is independent of delay. For this reason, it is often necessary to validate the measurement data against the presumed model. This could be done by investigating the residuals between the measured data and the regression, to make sure that the residuals do not exhibit any sort of distance dependence. If the data seems to be described by a different model, then a different pathloss model would have to be considered. The standard error of the parameters estimated with OLS depend on the number of samples and the exact pathloss sample distances that are used in the measurement. However, if the data is being censored, OLS would provide biased results, which means that Eq. (11) no longer applies.

IV. Estimation of Censored Pathloss Data

In order to estimate the pathloss exponent and fading variance of censored data, with a *known* number of missing samples where only the distance is known, it is possible to base the estimation on a censored normal distribution. Under this assumption, the observations follow a censored normal distribution [9]. The censoring occurs for data samples where the pathloss is greater than or equal to c. The value $-c$ is a channel gain that corresponds to the noise floor of the channel sounder or measurement device. In practice, c is chosen with some margin with respect to the noise floor, so that a limited number of samples dominated by noise are included as measurement data. Using the data set model in (4), the data is assumed to be censored so that observations with values at or above c are set to c, i.e.,

---

[2] The standard error is the standard deviation of the sampling distribution of a statistic.

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* < c \\ c & \text{if } y_i^* \geq c \end{cases} \quad (12)$$

where

$$y_i^* \sim \mathcal{N}(\mathbf{x}_i\boldsymbol{\alpha}, \sigma^2). \quad (13)$$

The probability of observing a censored observation at a distance d is given by

$$P(y \geq c) = 1 - \Phi\left(\frac{c - \mathbf{x_i}\boldsymbol{\alpha}}{\sigma}\right), \quad (14)$$

where Φ is the cumulative distribution function (CDF) of the standard normal distribution. Now, by using I as an indicator function that is set to 1 if the observation is uncensored and is set to 0 if the observation is censored, it is possible to write down the likelihood function as [9]

$$l(\sigma, \boldsymbol{\alpha}) = \prod_{i=1}^{N} \left[\frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x_i}\boldsymbol{\alpha}}{\sigma}\right)\right]^{I_i} \left[1 - \Phi\left(\frac{c - \mathbf{x_i}\boldsymbol{\alpha}}{\sigma}\right)\right]^{1-I_i},$$

where ϕ is the standard normal probability density function (PDF). The log-likelihood $L(\sigma, \boldsymbol{\alpha}) = \ln[l(\sigma, \boldsymbol{\alpha})]$ can now be written as

$$
\begin{aligned}
L(\sigma, \boldsymbol{\alpha}) =& \sum_{i=1}^{N} I_i \left[-\ln\sigma + \ln\phi\left(\frac{y_i - \mathbf{x_i}\boldsymbol{\alpha}}{\sigma}\right)\right] \\
&+ \sum_{i=1}^{N} (1-I_i)\ln\left[1 - \Phi\left(\frac{c - \mathbf{x_i}\boldsymbol{\alpha}}{\sigma}\right)\right].
\end{aligned} \quad (15)
$$

Using the log-likelihood, the parameters σ and $\boldsymbol{\alpha}$ are estimated using

$$[\hat{\sigma}, \hat{\boldsymbol{\alpha}}] = \arg\min_{\sigma, \boldsymbol{\alpha}}\{-L(\sigma, \boldsymbol{\alpha})\}, \quad (16)$$

which is easily solved by numerical optimization of $\boldsymbol{\alpha}$ and σ, using for instance the method of Newton [10]. In this work, we have solved this by using the fminsearch function in Matlab, which is based on a Nelder-Mead search method. The estimates obtained from OLS were used as initial values for the minimization. The presented method approach can easily be further extended, so that it supports other pathloss models.

A. Asymptotic Variance of the ML estimator

The asymptotic variance of the ML estimator has been derived in [10] for the problem with censoring of samples where $y_i \leq 0$. We therefore transform the data in Eq. (4), by letting

$$\mathbf{y}_t = -\mathbf{y} + c = -\mathbf{X}\boldsymbol{\alpha} - \boldsymbol{\epsilon} + c = \mathbf{X}\boldsymbol{\alpha}_t - \boldsymbol{\epsilon}, \quad (17)$$

where

$$\boldsymbol{\alpha}_t = [-PL(d_0) + c \quad -n]^T. \quad (18)$$

The parameters to be estimated for the transformed data are

$$\boldsymbol{\theta}_t = [\boldsymbol{\alpha}_t^T \quad \sigma^2]^T. \quad (19)$$

The asymptotic variance for the ML estimates of the original parameters, $\boldsymbol{\theta}$, are however the same as for the parameters of the transformed data, $\boldsymbol{\theta}_t$. Therefore, we can directly use the equations found in [10] to calculate the asymptotic variance as

$$\text{Avar}(\boldsymbol{\theta}) = \text{Avar}(\boldsymbol{\theta}_t) = \text{diag}\left\{\left(\sum_{i=1}^{N} \mathbf{A}_i(\mathbf{x}_i, \boldsymbol{\theta}_t)\right)^{-1}\right\}, \quad (20)$$

where

$$\mathbf{A}_i(\mathbf{X}_i, \boldsymbol{\theta}_t) = \begin{pmatrix} a_i\mathbf{x}_i^T\mathbf{x}_i & b_i\mathbf{x}_i^T \\ b_i\mathbf{x}_i & c_i \end{pmatrix}, \quad (21)$$

with coefficients

$$
\begin{aligned}
a_i &= -\sigma^{-2}\left[z_i\phi_i - \phi_i^2/(1-\Phi_i) - \Phi_i\right], \\
b_i &= \sigma^{-3}\left[z_i^2\phi_i + \phi_i - z_i\phi_i^2/(1-\Phi_i)\right]/2, \\
c_i &= -\sigma^{-4}\left[z_i^3\phi_i + z_i\phi_i - z_i^2\phi_i^2/(1-\Phi_i) - 2\Phi_i\right]/4.
\end{aligned} \quad (22)
$$

Here, $\phi_i = \phi_i(z_i)$ and $\Phi_i = \Phi_i(z_i)$ and $z_i = \mathbf{x}_i\boldsymbol{\alpha}_t/\sigma$. In order to avoid numerical issues when calculating the coefficients in Eq. 22, it is worthwhile to rewrite the ratio $\phi_i/(1-\Phi_i)$ as

$$\frac{\phi_i(z_i)}{1 - \Phi_i(z_i)} = \frac{\frac{1}{\sqrt{2\pi}}\exp(-z_i^2/2)}{1 - \frac{1}{2}\text{erfc}(-z_i/\sqrt{2})} = \frac{2}{\sqrt{2\pi}\text{erfcx}(z_i/\sqrt{2})}, \quad (23)$$

where $\text{erfc}(\cdot)$ is the complementary error function and $\text{erfcx}(\cdot)$ is the scaled complementary error function.

As stated previously, the asymptotic variances of the parameters $\boldsymbol{\theta}_t$ are the same as for $\boldsymbol{\theta}$. Therefore, the asymptotic variance of the parameters $PL(d_0)$, n and σ^2 are given by the three main diagonal elements of the matrix in Eq. (20). For measurement data, an estimate of the asymptotic variance can be found by replacing the true parameter values with their estimates, $\hat{PL}(d_0)$, $\hat{n}$ and $\hat{\sigma}^2$. Estimates of the standard errors can then be obtained simply by taking the square root of the asymptotic variance.

The standard errors of the estimated parameters depend on many different things, such as the pathloss sample distances, the level of the censoring, the number of samples as well as the exact values of $PL(d_0)$, n and σ^2. Therefore, it is often necessary to evaluate the standard errors for each individual measurement case. In order to give an illustrative example, Fig. 1 shows the standard errors calculated for synthetic pathloss data with known parameters, that is sampled at equally spaced distances on a linear scale, starting from 1 m. The thick lines are the standard errors for pathloss data that is censored at a level of 90 dB, whereas the thin lines are for data that is completely uncensored. The dashed line indicates the distance at which the expected value of the pathloss is equal to the censoring level. As the distance is increased, the standard error for the case with censoring starts to deviate from the case with no censoring, just as expected. This deviation seems to be especially severe for the large scale fading variance. However, for n and $PL(d_0)$, the standard errors are fairly similar for the two cases even for distances beyond the point where the pathloss starts to fall below the censoring level. Fig. 2 shows virtually the same thing as Fig. 1, except that the distance is sampled using equally spaced distances on a *logarithmic* scale.
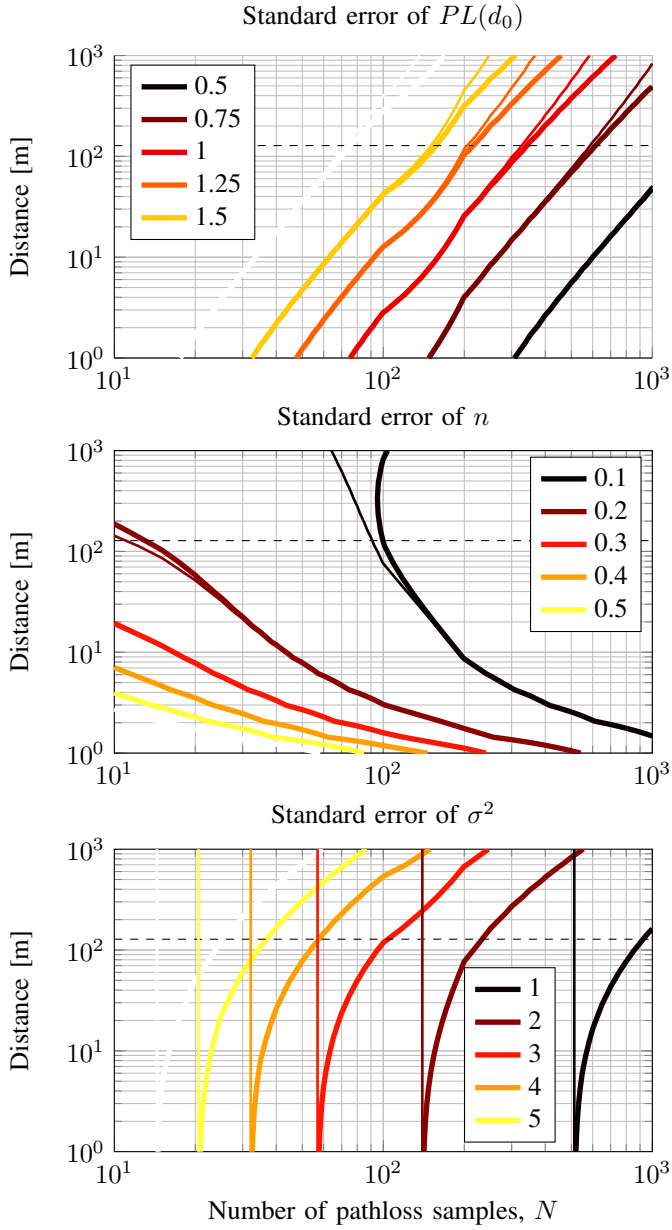
Fig. 1. Standard errors for pathloss parameters estimated with the ML estimator. The true values of the parameters are $PL(d_0) = 47.9$ dB, $n = 2$ and $\sigma = 4$. The thick lines are standard errors for data that is censored at a level corresponding to pathloss values of 90 dB, whereas the thin lines are for data that is completely uncensored (which corresponds to the standard errors of OLS). The dashed lines indicate the distance at which the expected value of the pathloss is equal to the censoring level. The data is sampled at equally spaced distances on a linear scale, starting from 1 m.

This shows that logarithmically spaced measurement distances requires fewer samples, compared to linearly spaced distances, in order to reach the same standard error for the estimated parameters.
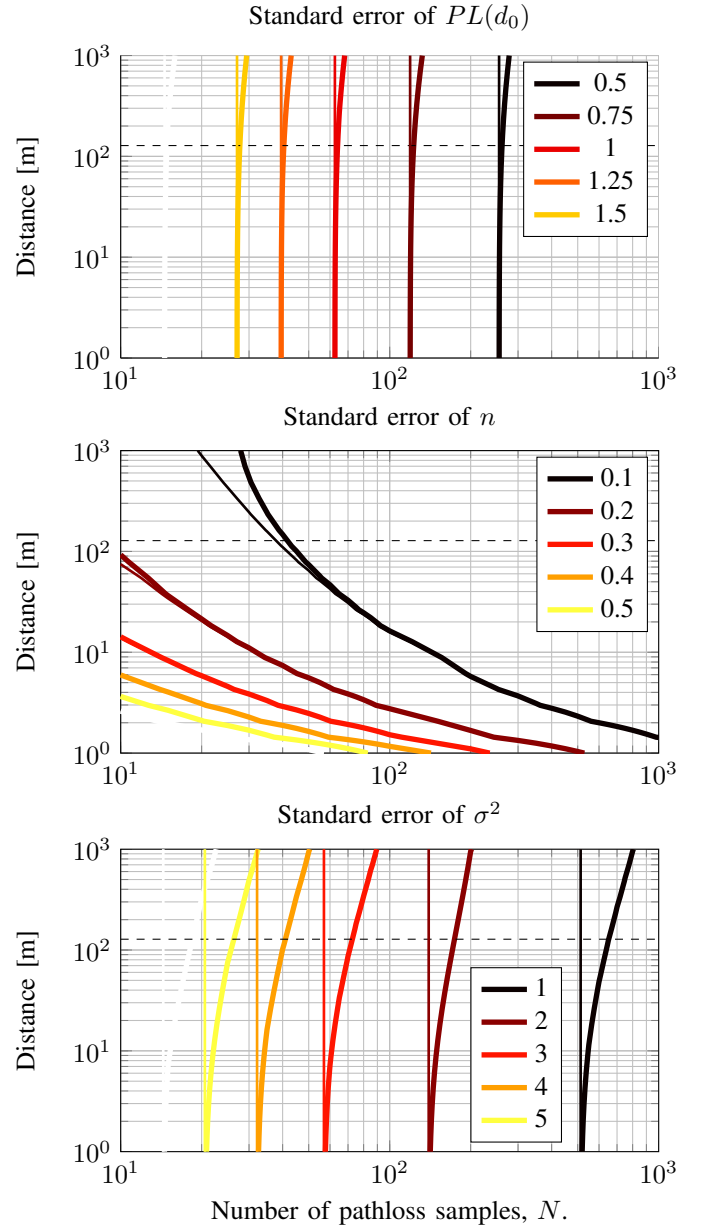

Fig. 2. Standard errors for pathloss parameters estimated with the ML estimator, with the same conditions as in Fig. 2, except that the data is sampled at equally spaced distances on a logarithmic scale, starting from 1 m.

V. RESULTS

As an example, synthetic data at 5.6 GHz was generated according to Eq. (1) with known parameters ($n = 2$ and $\sigma = 4$) and a synthetic censoring level at c. The parameters were estimated using OLS and the ML method described above. The result is shown in Fig. 3. The OLS method clearly underestimates both the pathloss exponent, $\hat{n}$, as well as the standard deviation of the large scale fading, $\hat{\sigma}$. The ML method on the other hand, is able to correctly estimate both parameters in this example. Fig. 4 shows the same thing as Fig. 3, but is for
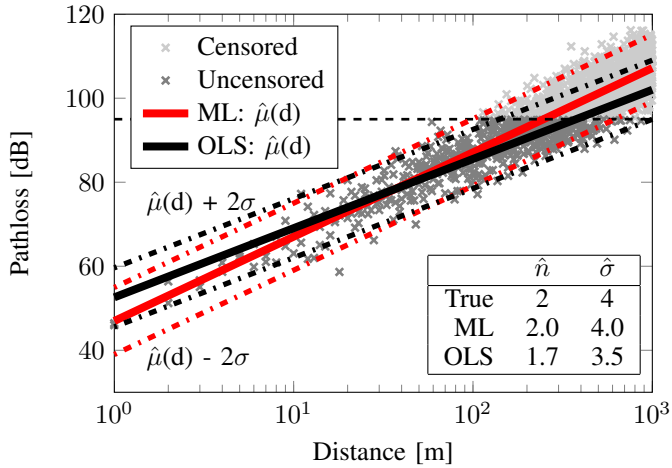
Fig. 3. Pathloss estimation based on censored synthetic data using the ML estimation method that considers censoring and using OLS without considering censoring. The ML method produces accurate estimates, whereas the OLS method underestimates n and σ.
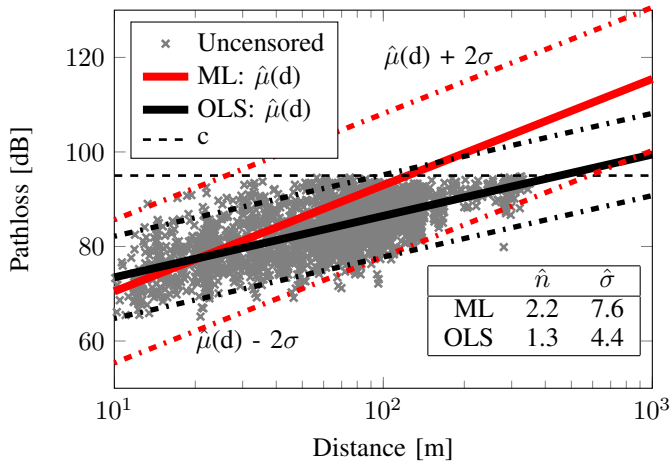

Fig. 4. Pathloss estimation of censored measurement data, using ML and OLS estimation. The estimated standard error for the ML estimates are $\mathrm{SE}(\hat{PL}(d_0)) = 0.72$ dB, $\mathrm{SE}(\hat{n}) = 0.04$ and $\mathrm{SE}(\hat{\sigma}^2) = 1.6$.

measured data from a vehicle-to-vehicle (V2V) measurement campaign for NLOS scenarios at 5.6 GHz [13]. In this case, the parameter estimates obtained using OLS show significantly smaller values compared to the parameter estimates for the ML method. This large discrepancy is due to the large number of censored samples in this data set; about 45 % of the measurement data points are censored. As a result, the OLS, which does not consider the censored samples, greatly underestimates the pathloss exponent and large scale fading. This shows the importance of taking censored samples into account when estimating the pathloss parameters.

VI. Conclusions

In this technical report, we suggest the use of a Tobit ML method [9] for the estimation of pathloss parameters based on censored data. When the data is censored, the standard approach of OLS, which has been widely used in the literature, is inconsistent, and yields biased estimates. The suggested ML estimator solves this problem by jointly estimating the parameters based on a censored normal distribution. Equations for the standard errors of this estimator are also provided. Using these equations, we show that the sampling distribution of the measurement samples can have a significant effect on the standard error in typical pathloss measurements. Using synthetic pathloss data that is censored, we also show that the ML method is able to correctly estimate the pathloss parameters, whereas OLS is biased and underestimates the pathloss exponent and the large-scale fading variance. Lastly, by using measured pathloss data from a V2V measurement campaign, we see that the ML method yields drastically different and more realistic estimates compared to the OLS method. In the appendix, we provide Matlab codes for the ML estimator and the asymptotic variance of the parameter estimates as well as an example on how to use these codes.

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. Kunisch and J. Pamp, "Wideband car-to-car radio channel measurements and model at 5.9 GHz," in *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*, Sept 2008, pp. 1–5.

[6] P. Paschalidis, K. Mahler, A. Kortke, M. Peter, and W. Keusgen, "Pathloss and multipath power decay of the wideband car-to-car channel at 5.7 GHz," in *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, May 2011, pp. 1–5.

[7] O. Onubogu, K. Ziri-Castro, D. Jayalath, K. Ansari, and H. Suzuki, "Empirical vehicle-to-vehicle pathloss modeling in highway, suburban and urban environments at 5.8 GHz," in *Signal Processing and Communication Systems (ICSPCS), 2014 8th International Conference on*, Dec 2014, pp. 1–6.

[8] C. Gustafson, D. Bolin, and F. Tufvesson, "Modeling the cluster decay in mm-wave channels," in *Antennas and Propagation (EuCAP), 2014 8th European Conference on*, April 2014, pp. 804–808.

[9] J. Tobin, "Estimation of relationships for limited dependent variables," in *Econometrica*, vol. 26, no. 1, Jan 1958, pp. 24–36.

[10] T. Amemiya, "Regression analysis when the dependent variable is truncated normal," in *Econometrica*, November 1973, pp. 997–1016.

[11] A. Molisch, *Wireless Communications*. Wiley-IEEE Press, 2005.

[12] J. Karedal, N. Czink, A. Paier, F. Tufvesson, and A. Molisch, "Path loss modeling for vehicle-to-vehicle communications," *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 1, pp. 323–328, Jan 2011.

[13] T. Abbas, K. Sjöberg, J. Karedal, and F. Tufvesson, "A measurement based shadow fading model for vehicle-to-vehicle network simulations," *International Journal of Antennas and Propagation*, 2015.

APPENDIX

A. Maximum-likelihood estimator

```
% CENSOREDML Numerical maximum-likelihood estimation of a censored normal
% distribtuion.
%
%   CENSOREDML(X,Y,C,T,A,S2) performs a numerical optimization of the parameters
%   for the log-likelihood of a censored normal distribution.
%   The data is assumed to be of the form Y=alpha*X+e, where alpha are regression
%   coefficients and e is r.v. that is Gaussian with zero mean.
%
%   C is a constant describing the censoring level and T is a vector indicating if
%   the samples are censored or uncensored (0 indicates that the sample is
%   censored and 1 indicates that it is uncensored). A and S are initial values for
%   the regression parameters and the variance of the error, respectively.
%   For pathloss estimation, A contains the initial values of [PL(d0); n] and S is
%   the initial value of the large scale fading variance sigma2. If You use this
%   code, or parts of it, please cite the following paper:
%   C. Gustafson, T. Abbas, D. Bolin and F. Tufvesson,
%   "Statistical Modeling and Estimation of Censored Pathloss Data",
%   IEEE Wireless Communications Letters, 2015.
%
%   Copyright (C) 2015, David Bolin, Carl Gustafson.
%
%   This program is free software: you can redistribute it and/or modify
%   it under the terms of the GNU General Public License as published by
%   the Free Software Foundation, either version 3 of the License, or
%   (at your option) any later version.
%
%   This program is distributed in the hope that it will be useful,
%   but WITHOUT ANY WARRANTY; without even the implied warranty of
%   MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
%   GNU General Public License for more details.
%
%   You should have received a copy of the GNU General Public License
%   along with this program.  If not, see <http://www.gnu.org/licenses/>.

function est=censoredml(x,y,c,t,a_est,s2e)

  opts = optimset('GradObj','off', 'Largescale','off','MaxFunEvals',20000);
  pars = fminsearch(@(pars) censoredllh(pars),[a_est;log(s2e)],opts);
  est=[pars(1:2); sqrt(exp(pars(3)))];

  function l = censoredllh(p)
    L = zeros(size(y));
    L(t==1)=-0.5*(y(t==1)-x(t==1,:)*p(1:2)).^2/exp(p(3))-log(sqrt(2*pi))-p(3)/2;
    L(t==0)=log(1-normcdf((c-x(t==0,:)*p(1:2))/exp(p(3)/2)));
    l = -sum(L);
  end
end
```

B. Asymptotic Variance

```
% CENSOREDVAR Computes the theoretical asymptotic variance for
% the estimates of the censored ML estimator.
%
%   CENSOREDVAR(X,C,PLd0,N,SIGMA) uses the equations from T. Amemiya,
%   "Regression Analysis when the dependent variable is truncated normal",
%   Econometrica, 1987. C is the censoring level, PLd0, N and SIGMA are
%   either true or estimated pathloss parameter values. If You use this
%   code, or parts of it, please cite the following paper:
%   C. Gustafson, T. Abbas, D. Bolin and F. Tufvesson,
%   "Statistical Modeling and Estimation of Censored Pathloss Data",
%   IEEE Wireless Communications Letters, 2015.
%
%   Copyright (C) 2015, David Bolin, Carl Gustafson.
%
%   This program is free software: you can redistribute it and/or modify
```

```
function censvar=censoredvar(x,c,PLd0,n,sigma)

z=x*[-PLd0+c -n]'/sigma;
phi=normpdf(z);
Phi=normcdf(z);

zp = z.*phi;
r=2./(sqrt(2*pi)*erfcx(z/(sqrt(2))));
ai=-(zp-phi.*r-Phi)/sigma^2;
bi=(z.*zp+phi-zp.*r)/(2*sigma^3);
ci=-(z.^2.*zp+zp-z.*zp.*r-2*Phi)/(4*sigma^4);
bx  =sum(bsxfun(@times,bi,x));
censvar=(diag(inv([bsxfun(@times,ai,x)'*x bx'; bx sum(ci)])));
```

C. Example on how to use the code

```
% EXAMPLE An example on how to use the Matlab codes CENSOREDML and
% CENSOREDVAR.
%
%   EXAMPLE defines synthetic pathloss data with given pathloss parameters
%   and a given censoring level, c. Then, estimates for the pathloss
%   parameters of censored data are calculated using ordinary least squares
%   (OLS) and the maximum-likelihood (ML) estimator found in CENSOREDML.
%   Standard errors for the parameter estimates of the ML estimator are
%   calculated using CENSOREDVAR. If You use this code, or parts of it,
%   please cite the following paper: C. Gustafson, T. Abbas, D. Bolin
%   and F. Tufvesson, "Statistical Modeling and Estimation of
%   Censored Pathloss Data", IEEE Wireless Communications Letters, 2015.
%
%   Copyright (C) 2015, David Bolin, Carl Gustafson.
%
%   This program is free software: you can redistribute it and/or modify
%   it under the terms of the GNU General Public License as published by
%   the Free Software Foundation, either version 3 of the License, or
%   (at your option) any later version.
%
%   This program is distributed in the hope that it will be useful,
%   but WITHOUT ANY WARRANTY; without even the implied warranty of
%   MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
%   GNU General Public License for more details.
%
%   You should have received a copy of the GNU General Public License
%   along with this program.  If not, see <http://www.gnu.org/licenses/>.

% Define synthetic parameters
d=linspace(10,200,200)'; %Distances [m]
d0=1;                    % Reference distance [m]
PLd0=20*log10(4*pi*d0/(3e8/5.9e9)); %Reference pathloss value at d0.
n=2;                     %Pathloss exponent.
sigma=4;                 %Large scale standard deviation [dB].
c=90;                    %Censoring level [dB]

% Generate synthetic data
y=PLd0+10*n*log10(d/d0)+sigma*randn(length(d),1);
x=[ones(length(y),1) 10*log10(d/d0)];

%Define indicator for censore and uncensored samples
```

```
t=zeros(1,length(d));
t(find(y<c))=1;
t(find(y>=c))=0;

%Censoring of the synthetic data
yt=zeros(length(d),1);
yt(find(y<c))=y(find(y<c));
yt(find(y>=c))=c;
xt=x(find(y<c),:);

%Ordinary Least Squares (OLS) estimates
a_est = inv(xt'*xt)*xt'*yt(find(t));
sigma2_est = var(yt(find(t))-xt*a_est);

%Maximum likelihood (ML) estimates
thetahat=censoredml(x,yt,c,t,a_est,sigma2_est)

%Asymptotic variance
Avarhat=censoredvar(x,c,thetahat(1),thetahat(2),thetahat(3));
Avar=censoredvar(x,c,PLd0,n,sigma);

%S=['PL(d0)';'n';'sigma'];
disp('   Estimate    OLS     ML    | True value');
disp(sprintf('   PL(d0)      %.2f   %.2f | %.2f', a_est(1), thetahat(1),PLd0))
disp(sprintf('   n           %.3f   %.3f | %.3f', a_est(2), thetahat(2),n))
disp(sprintf('   sigma       %.3f   %.3f | %.3f', sqrt(sigma2_est), thetahat(3),sigma))
disp('   --------------------------------');
disp('   Asymptotic standard errors (SE) of ML estimates');
disp('             SE(true value) SE(estimated value)');
disp(sprintf('   SE(PL(d0))   %.2f             %.2f', sqrt(Avar(1)), sqrt(Avarhat(1))))
disp(sprintf('   SE(n)        %.3f             %.3f', sqrt(Avar(2)), sqrt(Avarhat(2))))
disp(sprintf('   SE(sigma^2)  %.3f             %.3f', sqrt(Avar(3)), sqrt(Avarhat(3))))
```