Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
International Journal of Rock Mechanics & Mining Sciences 43 (2006) 877–893
International Journal of Rock Mechanics and Mining Sciences
www.elsevier.com/locate/ijrmms

Inference of discontinuity trace length distributions using statistical graphical models

R. Jimenez-Rodriguez*, N. Sitar

Department of Civil and Environmental Engineering, University of California at Berkeley, CA-94720-1710 USA

Accepted 16 December 2005
Available online 17 February 2006

Abstract

The characterization of discontinuities within rock masses is often accomplished using stochastic discontinuity network models, in which the stochastic nature of the discontinuity network is represented by means of statistical distributions. We present a flexible methodology for maximum likelihood inference of the distribution of discontinuity trace lengths based on observations at rock outcrops. The inference problem is formulated using statistical graphical models and target distributions with several Gaussian mixture components. We use the Expectation–Maximization algorithm to exploit the relations of conditional independence between variables in the maximum likelihood estimation problem. Initial results using artificially generated discontinuity traces show that the method has good inference capabilities, and inferred trace length distributions closely reproduce those used for generation. In addition, the convergence of the algorithm is shown to be fast.

Keywords: EM-algorithm; Fracture network; Discontinuity size; Maximum likelihood

1. Introduction

Discontinuities have a significant impact on the deformability, strength, and permeability of rock masses [1,2]; consequently, their characterization is an important element of rock mass characterization [1–5]. However, deterministic characterization of individual discontinuities in the rock mass is usually an insurmountable site characterization challenge and, in general, it is only feasible for major features. The characterization of other (i.e., not major) discontinuities can be, on the other hand, accomplished using stochastic discontinuity network models. In these models, the rock mass is represented as an assemblage of discontinuities intersecting a volume of intact rock and the stochastic nature of the discontinuity network is represented using statistical distributions [6,7]; in some cases, additional non-geometrical aspects are included to imitate geological processes leading to the formation of discontinuities within the rock mass [5,8].

To be able to use stochastic discontinuity network models in engineering applications, the problem of calibration of network parameters remains, and we need methods for the characterization of discontinuity networks based on information available at the design stage. In particular, one of the main difficulties in estimating discontinuity sizes is the fact that direct observation of their complete three dimensional extent is not possible. As a result, the distribution of discontinuity dimensions is commonly inferred using information about the distribution of trace lengths at rock exposures by means of, for example, stereological or fractal considerations [9–15].

Hence, proper characterization of the distribution of trace lengths is an essential step in the characterization of the distribution of discontinuity dimensions. There are two additional difficulties in the solution of the problem of estimating the trace length distribution: The first is that the observations of discontinuity traces are biased [15–20]; and the second is due to the complexity of the geological processes leading to the development of rock discontinuities

*Corresponding author. Present address: ETS Ing. de Caminos, Canales y Puertos. Universidad Politécnica de Madrid, Spain.
Tel.: +34 91 336 6710; fax: +34 91 336 6774.
E-mail address: rjimenez@caminos.upm.es (R. Jimenez-Rodriguez).

1365-1609/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijrmms.2005.12.008

[11]. Such complexity is responsible for uncertainties about the most adequate type of distribution to be used in real applications. Exponential and lognormal distributions are most commonly employed (see e.g., [21,22]), but there are cases in which bi-modal types of distributions—which cannot be properly described by the distributions mentioned above—seem to be more adequate, as suggested by case histories or geomechanical modeling results (see e.g., [18,23,24]).

To deal with these problems, the measured (i.e., biased) trace length distribution is commonly estimated first; then, it is assumed that the real (i.e., unbiased) trace length distribution has the same distribution form, so that only the parameters of the distribution (usually the mean and variance) need to be obtained [7,21]. Several methods for the estimation of the mean of the real distribution of trace lengths have been proposed [17,19,25,26]; the variance can be estimated by assuming that the values of the variances of the observed and real trace length distributions are equal [7] or, better yet, by assuming that the coefficients of variation of both distributions are equal [21]. Additional statistical approaches have been used to estimate the distribution of discontinuity trace lengths: Song and Lee [20] estimated trace lengths using areal sampling and probabilistic relations derived for the Poisson disk model; and maximum likelihood estimation methods for scanline or areal sampling have been proposed as well [18,23,27–29].

In this work, we present a novel approach based on the use of statistical graphical models for maximum likelihood inference of the real (i.e., corrected for biases) distribution of discontinuity trace lengths [30,31]. The identification of the type or structure of the trace length distribution is not our main interest; rather, we are interested in working with a model that provides reasonable estimates of the probability distribution without making strong assumptions about its type. Accordingly, we avoid the assumption that the measured and real trace lengths distributions are of the same type, and we approach the inference problem by considering a broad family of target mixture distribution models. That is, we work with a *target* distribution that is flexible enough to mimic the main features of the *real* (and unknown) distribution of trace lengths, allowing the observed data to "select" the most adequate distribution for each case.

We believe that this type of model based on a statistical analysis of observed discontinuity trace data will gain even more significance in the years to come, as traditional methods for discontinuity surveying in rock engineering [32] are being replaced by automated techniques, which allow more efficient and detailed acquisition of discontinuity data [33–35].

2. Generation and sampling of discontinuities

We assume that the rock outcrop is a planar surface and that the *sampling domain* (where traces are observed) is a rectangular region within the rock outcrop of dimensions $W_o \times H_o$. We further assume that the sampling domain is contained within the *generation domain*, where we use Monte Carlo simulation to generate populations of discontinuity traces. We consider the size of the generation domain to be "much larger" than the size of the sampling domain and "much larger" than the length of generated traces, so that the consequences of biases, edge effects, or both are negligible in the generation process. (For a detailed analysis of the influence of stochastic network parameters in the occurrence of edge effects, see [36].) We also assume that discontinuities are parallel and flat circular disks of negligible thickness, with their centers uniformly distributed in space (i.e., the Poisson disk model [37]); accordingly, discontinuity traces are parallel straight lines of negligible width, with centers uniformly located within the generation domain.

The Poisson disk model has been extensively used in rock mechanics applications (see e.g., [5,6,9,20,21,29,36–41]). Here, we use the Poisson disk model because it has been found to generate fractures and fracture traces that are similar to natural fracture patterns in many cases and because it has been recognized that some fracture systems are best described by this type of models [5,11]. Additional advantage of the Poisson disk model is that it is simple and easy to program [5]; it is also mathematically convenient [14], allowing simple derivations of analytical expressions.

In other cases, however, characteristics of fracture systems in rock masses are best described by power laws and fractal geometry [11,12], and fractals have been widely used to describe fracture geometry (see e.g., [42–46]). (For further arguments in favor of fractals, see [5,11].) Therefore, the method proposed herein should not be applied to rock masses with a demonstrable fractal system without due consideration to the errors that may be introduced.

Fig. 1 illustrates the types trace maps, sampling domains, and generation domains used in this work.

3. Existing biases

3.1. Description of types of bias

Observations of discontinuity traces at rock exposures are subjected to orientation, truncation, censoring and size bias (e.g., [16–21]). The terms *curtailment* and *trimming* have also been used to refer to censoring and truncation [15], but we chose to use the former terminology for overall consistency with the rock mechanics literature.

In this work, we consider the commonly used model of a single set of parallel traces (see e.g., [15,20,25,26,29,47,48]) and, therefore, orientation bias does not affect the results discussed herein. Methods for correction of orientation bias—including the use of circular domains—are quite common [16,21,49–51], and extensions to consider trace data with variable orientation are also available [23,52]. Similarly, truncation bias is not significant in the context of block formation, since the truncation threshold may be

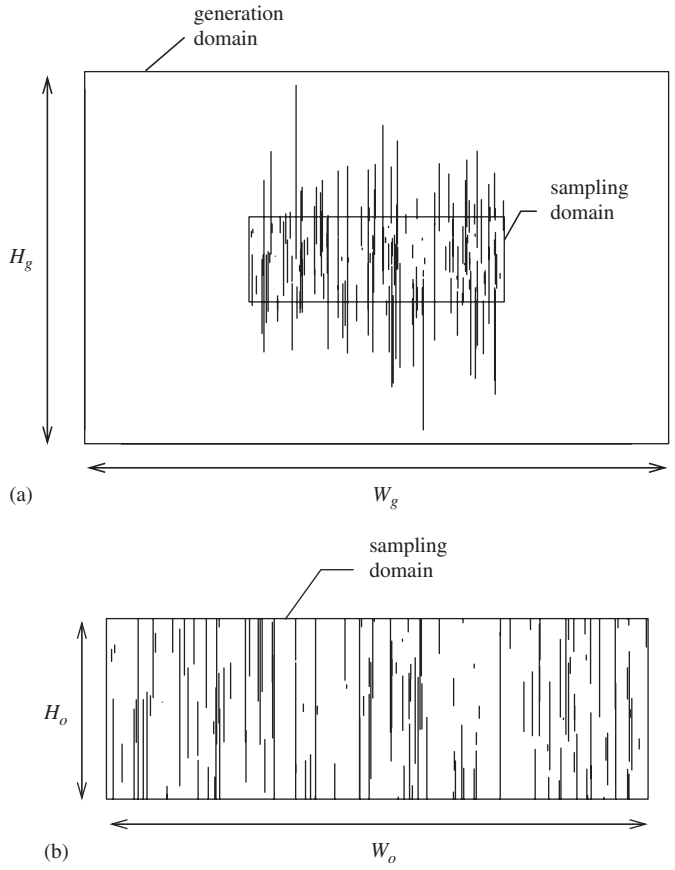


Fig. 1. Generation and sampling of discontinuities, with typical examples of observed trace maps. (a) Size-biased sample of observed traces (observations not censored), (b) observed traces (censored) for the sample above.

easily decreased so that it has negligible influence on the formation of medium to large size blocks, which are of main interest in engineering design. Truncation bias could be significant, however, in cases in which flow through discontinuities in the rock mass is important. Care should be taken in such cases so that the truncation threshold is decreased to an adequate low value.

The problem of censoring bias is, on the other hand, a more significant issue, since it prevents traces from being completely observed. That is, censoring bias causes observed trace lengths to be shorter than the corresponding real traces. This has negative safety consequences in the analysis, as larger discontinuities are more likely to produce larger blocks, hence, in the event of failure, greater consequences; similarly, large discontinuities can also serve as preferential paths for underground flow. Highly censored data sets obtained when discontinuities are much larger than the size of the outcrop impose some additional challenges (see e.g., [17]) that cannot be easily overcome without additional assumptions. In such cases, geophysical methods provide a promising tool to characterize discontinuity sizes but, despite progress, further research is needed before they can be widely applied to rock engineering projects [11,24]. Empirical relations between size and other discontinuity parameters (e.g., aperture) provide another alternative to infer sizes of discontinuities much larger than the size of the outcrop [11,53].

Finally, size bias occurs because longer traces are more likely to be sampled than shorter ones. One alternative would be to ignore size bias, with the argument that this simplification is on the side of safety. We believe, however, that it is preferable to develop as accurate methods of estimation as possible, so that more efficient engineering solutions can be achieved. Thus, a solution to the problem of size bias is discussed next.

3.2. Correction for size bias

Fig. 1(a) shows an example of one size biased sample of discontinuity traces. To correct for size bias, we need to estimate the probability density function (PDF) of the real (i.e., unbiased) trace length distribution, $f(l)$, from estimates of the distribution of total lengths of observed traces (i.e., not censored, but size biased), $f'(l)$. As trace centers are assumed to be uniformly distributed within the generation domain and traces are oriented perpendicular to the horizontal sides of the rectangular sampling domain, the probability of a trace of length $L = l$ intersecting the sampling domain is proportional to $l + H_o$, where H_o is the vertical dimension of the sampling domain [29] (see Fig. 1(b)).[1] That is,

$$f'(l) \propto K_1(l + H_o)f(l), \tag{1}$$

or equivalently,

$$f(l) \propto K_2 \frac{f'(l)}{l + H_o}. \tag{2}$$

Imposing the condition that $f(l)$ integrates to one, we obtain the proportionality constant to be $K_2 = (\int_0^\infty (f'(l)/l + H_o)\,dl)^{-1}$. Therefore, the PDF of the distribution of inferred "real" (i.e., unbiased) discontinuity trace lengths is given by:

$$f(l) = \frac{f'(l)}{(l + H_o)\int_0^\infty \frac{f'(\xi)}{\xi + H_o}\,d\xi}, \tag{3}$$

where ξ is an integration variable along values of l.

4. Statistical graphical model

We use *target* mixture distributions to overcome uncertainties about the most adequate type of trace length distribution to be used in real applications. Within this context, mixture distributions are used as a "mathematical artifact" that provides increased inference flexibility and

[1] Facing a similar problem—in the context of length estimation of textile fibers—Cox [54] proposed a correction for size bias, in which he considered the probability of sampling a certain fiber with a sampling line to be proportional to the fiber length. Such correction has later been extensively employed in the context of scanline sampling of rock discontinuities (e.g., [15,18,47]).

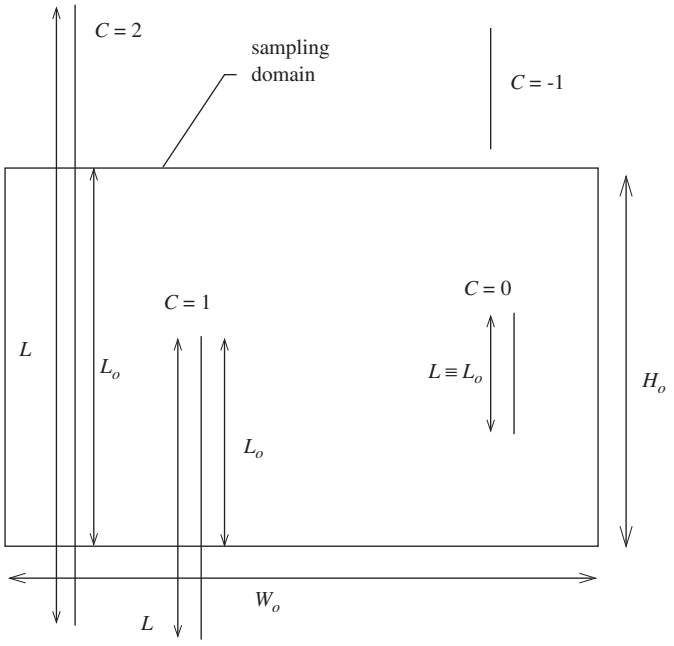


Fig. 2. Random variables for different types of discontinuity traces.

broadens the class of distributions that may be adequately reproduced by the model. For a mixture of K components, $\pi \equiv (\pi_1, \ldots, \pi_K)$ is the set of mixture proportions (with $\pi_i \geqslant 0$ and $\sum_{i=1}^{K} \pi_i = 1$),[2] and $\theta \equiv (\theta_1, \ldots, \theta_K)$ is the set of component distribution parameters. (i.e., $\Theta \equiv (\pi, \theta)$ is the *complete set* of parameters of the mixture.) Then, if $p_i(\cdot|\theta_i)$ is the PDF of the distribution of component i, the PDF of the mixture distribution is given by [55]:

$$p(\cdot|\Theta) = \sum_{i=1}^{K} \pi_i p_i(\cdot|\theta_i). \tag{4}$$

We characterize each trace observed (up to a total of N_o), by means of a number of random variables (see Fig. 2). L represents the *total* trace length of observed discontinuities; due to censoring bias (see Section 3), this random variable is not observed in general. Alternatively, we have two other random variables that are observed: L_o represents the *observed* trace length (it is therefore a lower bound for L); and C represents the corresponding censoring conditions. Depending on the number and the type of intersections (if any) between the discontinuity trace and the boundaries of the sampling domain, we consider the censoring conditions described in Table 1. Finally, we use Z (a multinomial random variable with K possible values), to represent the mixture component to which each observed trace is assigned.

Based on the definitions above, we propose the statistical graphical model presented in Fig. 3. Edges joining random variables denote statistical dependence between them, and shaded nodes are used to represent observed random variables. In addition, the plate representation in our graphical model denotes statistical independence between observations; that is, our sample is composed of a set of N_o observations of statistically independent and identically distributed random variables.

[2]There may be additional constraints on the domain of the parameters θ_i that form an acceptable solution set for the distribution of each mixture component. For instance, if we use normal random variables as mixture components, an additional constraint is that they should all have positive variance.

Table 1
Possible censoring conditions for a discontinuity trace

Value	Description
$C = -1$	*Not observed*: The discontinuity trace is located *outside* of the sampling domain.
$C = 0$	*No censoring*: The discontinuity trace is located *inside* of the sampling domain and it does not intersect its boundaries.
$C = 1$	*Censoring on one side*: Only one of the extremes of the discontinuity trace is observed; i.e., the other one is censored.
$C = 2$	*Censoring on both sides*: None of the extremes of the discontinuity trace are observed; i.e., both of them are censored.

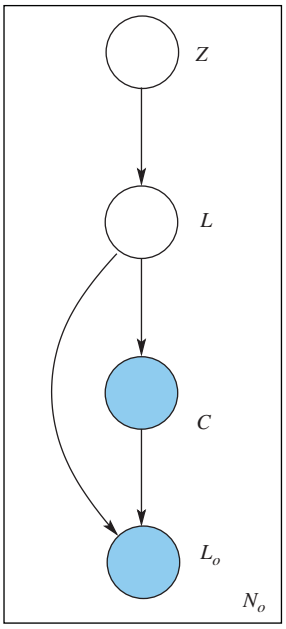


Fig. 3. Proposed graphical model.

5. Estimation of trace length distribution parameters

5.1. Introduction

To illustrate the maximum likelihood inference methodology, we use X to denote the set of observed variables, and we use Y to denote the set of unobserved variables. The *complete* probability model is then given by $p(x, y|\Theta)$, where Θ is the set of parameters of the model. If random variables Y were observed, the log likelihood of the model (which we refer to as the *complete log likelihood*) would be

computed as:

$$l_c(\Theta; x, y) \equiv \log p(x, y|\Theta). \tag{5}$$

Random variables Y are, however, not observed by definition, and the likelihood of the model for observed data $X = x$ has to be computed by marginalizing over the unobserved variables in Eq. (5). The *incomplete log likelihood* is then given by:

$$l(\Theta; x) \equiv \log p(x|\Theta) = \log \sum_y p(x, y|\Theta), \tag{6}$$

where we use the summation sign to indicate marginalization—the derivation would be equivalent (with the use of integration) for continuous random variables.

Unlike in Eq. (5), the logarithm in Eq. (6) is separated from the probability term by the marginalization (i.e., summation or integration) expression. This has the undesirable consequence that we cannot directly use the factorizations (i.e., conditional independence relations) of $p(x, y|\Theta)$ to simplify the maximization of the logarithmic term in Eq. (6); that is, the model does not "decouple".

The expectation–maximization (EM) algorithm provides a general method for maximum likelihood parameter estimation in graphical models with unobserved variables. An alternative approach would be to use a nonlinear optimization algorithm, such as conjugate gradient or Newton–Raphson [56,57]. The EM algorithm is applicable to arbitrary graphical models with unobserved variables, with the advantage that it allows us to exploit the independence structure of the probability model to its full extent [56,57]. This provides significant computational advantages, particularly for mixture models with several Gaussian components, which we use as target distributions.

Here, we present a brief introduction to the EM algorithm based on Jordan [56]. For a more in-depth discussion, the seminal reference is Dempster et al. [58]; Redner and Walker [55] discuss the problem of maximum likelihood estimation in the context of mixture densities; additional references of interest are Cowell et al. [59] and Xu and Jordan [57].

5.2. The EM algorithm

The EM algorithm uses an averaging distribution, $q(y|x)$, to obtain the expectation of the complete log likelihood with respect to the unobserved variables in the model. The *expected complete log likelihood* is defined as:

$$\langle l_c(\Theta; x, y)\rangle_q \equiv \sum_y q(y|x, \Theta) l_c(\Theta; x, y) \tag{7}$$

$$= \sum_y q(y|x, \Theta) \log p(x, y|\Theta). \tag{8}$$

In that way, the uncertainty introduced by the unknown values of Y is removed and, for a set of observations $X = x$, the expected complete log likelihood is a deterministic quantity that only depends on Θ. It also inherits the good computational properties of the complete log likelihood, since the log expression in Eq. (8) is directly applied to the complete probability model, allowing us to exploit statistical independence relations in the graphical model.

For an arbitrary distribution $q(y|x)$, we define an *auxiliary function* $\mathscr{L}(q, \Theta)$ which is a lower bound to the log likelihood:

$$\mathscr{L}(q, \Theta) \equiv \sum_y q(y|x) \log \frac{p(x, y|\Theta)}{q(y|x)} \leqslant l(\Theta; x). \tag{9}$$

The EM algorithm provides a coordinate ascent algorithm on $\mathscr{L}(q, \Theta)$, with the following steps:

E-step:

$$q^{(t+1)} = \arg\max_q \mathscr{L}(q, \Theta^{(t)}), \tag{10}$$

M-step:

$$\Theta^{(t+1)} = \arg\max_\Theta \mathscr{L}(q^{(t+1)}, \Theta). \tag{11}$$

The M step may be shown to maximize the expected complete log likelihood with respect to Θ. At each step of the algorithm, the maximization in the E step is achieved by the election of a distribution of the form $q^{(t+1)}(y|x) = p(y|x, \Theta^{(t)})$. Such averaging distribution not only maximizes the auxiliary function, but also assures that the auxiliary function and the log likelihood are *equal* at each EM iteration. Then, as $\mathscr{L}(q, \Theta^{(t)})$ is a lower bound for $l(\Theta^{(t)}; x)$, finding a local maximum of the auxiliary function is equivalent to finding a local maximum of the log likelihood, and the EM algorithm therefore provides a local optimum to the maximum likelihood estimation problem.

5.3. Likelihood functions

Let us use $\mathscr{D}_c$ to denote the *complete* set of data that we would have if we were able to observe the unobserved variables of the model, and use $\mathscr{D}_o$ to refer to the *observed* data that we actually have; that is, $\mathscr{D}_o$ is given by the set of observed trace lengths and censoring conditions of the N_o traces observed in our sample. Based on $\mathscr{D}_o$, we compute the (incomplete) log likelihood as:[3]

$$l(\Theta; \mathscr{D}_o) = \log p(\mathscr{D}_o|\Theta) \tag{12}$$

$$= \sum_{n=1}^{N_o} \log p(c, l_o|\Theta). \tag{13}$$

[3]In the derivations below, l, l_o, c, and z refer to the total length, observed length, censoring condition, and mixture component corresponding to the n-th trace of the N_o observed traces. To lighten the notation, however, we avoid making explicit reference to n when using these variables.

If we had observations of all variables in $\mathscr{D}_c$, the complete log likelihood would be computed as:

$$l_c(\Theta;\mathscr{D}_c) = \log p(\mathscr{D}_c|\Theta) \tag{14}$$

$$= \sum_{n=1}^{N_o} \log p(z,l,c,l_o|\Theta). \tag{15}$$

Using the averaging distribution $q(z,l|c,l_o)$, we can also compute the expected complete log likelihood, as:

$$\langle l_c(\Theta;\mathscr{D}_o)\rangle_q = \sum_{n=1}^{N_o}\sum_{z}\int_{l} q(z,l|c,l_o) \times \log p(z,l,c,l_o|\Theta)\,\mathrm{d}l, \tag{16}$$

and the auxiliary function results in:

$$\mathscr{L}(q,\Theta) = \sum_{n=1}^{N_o}\sum_{z}\int_{l} q(z,l|c,l_o)\log\frac{p(z,l,c,l_o|\Theta)}{q(z,l|c,l_o)}\,\mathrm{d}l. \tag{17}$$

6. Complete probability model

The expected complete log likelihood (see Eq. (16)) is computed as a function of the complete probability model $p(z,l,c,l_o|\Theta)$. This distribution may be factorized considering the relations of conditional independence in the graphical model, as:

$$p(z,l,c,l_o|\Theta) = p(z,l|\Theta)p(c|l,\Theta)\,p(l_o|c,l,\Theta) \tag{18}$$

$$= \left(\prod_{i=1}^{K}\left[\pi_i f'_i(l|\theta_i)\right]^{z^i}\right)p(c|l,\Theta) \times p(l_o|c,l,\Theta), \tag{19}$$

where $f'_i(l|\theta_i)$ is the PDF of the ith component of the *target* mixture distribution of total lengths for sampled (i.e., size biased) traces, $f'(l)$.

To complete the probabilistic description of the factorized model, however, we need expressions for the conditional distributions $p(c|l,\Theta)$ and $p(l_o|c,l,\Theta)$ in Eq. (19). To that end, we identify regions where centers of traces of length $L=l$ need to be located to have the censoring conditions presented in Table 1 (See Figs. 4 and 5).[4] The relative sizes of such regions (with respect to the size of the region for which traces of that length are sampled) are used to derive $p(c|l)$ and $p(l_o|c,l)$ for each censoring condition.

6.1. Traces with no censoring

In this case traces are completely contained within the sampling domain and their centers are located within the *No censoring* region in Fig. 4. As trace centers are uniformly located within the generation domain (see Section 2), the distribution $p(C=0|l)$ can be computed considering the size of the *No censoring* region with respect to the total size of the region for which traces with any censoring condition are sampled. We obtain:

$$p(C=0|l) = \begin{cases} \dfrac{H_o - l}{H_o + l} & \text{if } 0 \leqslant l < H_o, \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

The observed trace length is assured to be equal to the real length (i.e., $l \equiv l_o$); accordingly, the conditional distribution of observed trace lengths is given by:

$$p(l_o|C=0,l) = \begin{cases} \delta(l-l_o) & \text{if } 0 \leqslant l < H_o, \\ 0 & \text{otherwise,} \end{cases} \tag{21}$$

where $\delta(l-l_o)$ is the (continuous) Dirac delta function.

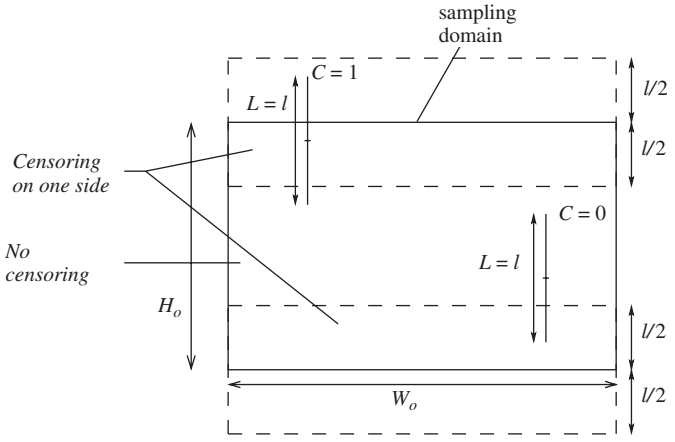


Fig. 4. Region where centers of traces with length $l<H_o$ need to be located to have $C=0$ or $C=1$ censoring conditions.

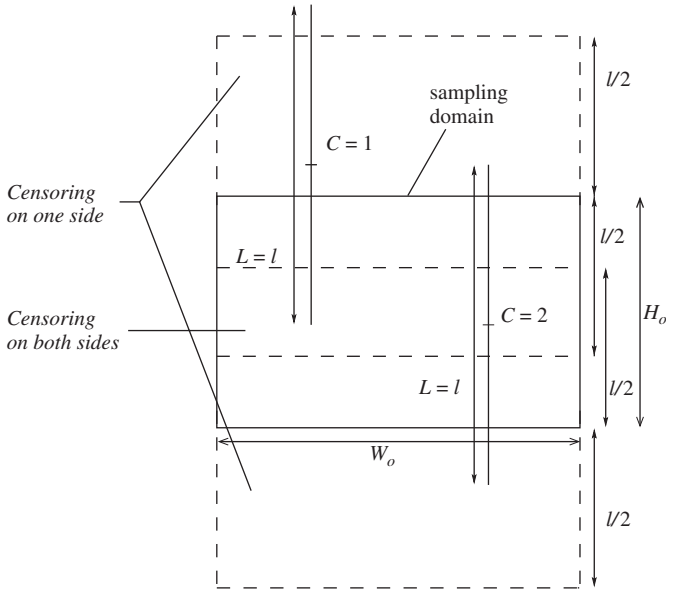


Fig. 5. Region where centers of traces with length $l>H_o$ need to be located to have $C=1$ or $C=2$ censoring conditions.

[4]Note that such regions are *completely defined* (i.e., deterministic) once the trace length is specified. Also, as expected given the proposed graphical model, the conditional distributions $p(c,l,\Theta)$ and $p(l_o|c,l,\Theta)$ are independent of the parameters of the target distribution. (i.e., $p(c|l,\Theta) \equiv p(c|l)$ and $p(l_o|c,l,\Theta) \equiv p(l_o|c,l)$.) Accordingly, to lighten the notation, we avoid further reference to these parameters in the derivations of $p(c|l)$ and $p(l_o|c,l)$.

6.2. Traces censored on one side

The conditional distribution $p(C=1|l)$ is also computed using the assumption of uniform location of trace centers within the generation domain, and considering the size of the *Censoring on one side* regions in Figs. 4 and 5 with respect to the total size of the region for which traces are sampled. We obtain:

$$p(C=1|l)=\begin{cases}\dfrac{2l}{H_o+l} & \text{if } 0\leqslant l<H_o,\\ \dfrac{2H_o}{H_o+l} & \text{if } l\geqslant H_o,\\ 0 & \text{otherwise.}\end{cases} \tag{22}$$

Since we assume that the distribution of trace center locations is uniform, the conditional distribution of observed trace lengths $p(l_o|C=1,l)$ is given by:

$$p(l_o|C=1,l)=\begin{cases}\dfrac{1}{l} & \text{if } 0\leqslant l_o\leqslant l; l<H_o,\\ \dfrac{1}{H_o} & \text{if } 0\leqslant l_o<H_o; l\geqslant H_o,\\ 0 & \text{otherwise.}\end{cases} \tag{23}$$

6.3. Traces censored on both sides

In this case the trace lengths are longer than the sampling domain dimension (i.e., $l\geqslant H_o$). Using the assumption of uniformly distributed trace centers, the distribution $p(C=2|l)$ can be computed considering the relative size of the *Censoring on both sides* region in Fig. 5 with respect to the size of the region for which traces with any censoring condition are sampled. We obtain:

$$p(C=2|l)=\begin{cases}\dfrac{l-H_o}{l+H_o} & \text{if } l\geqslant H_o,\\ 0 & \text{otherwise.}\end{cases} \tag{24}$$

The observed trace length is assured to be equal to the sampling domain dimension (i.e., $l_o\equiv H_o$); accordingly, the conditional distribution of observed trace lengths may be expressed in terms of the (continuous) Dirac delta function, as follows:

$$p(l_o|C=2,l)=\begin{cases}\delta(l_o-H_o) & \text{if } l\geqslant H_o,\\ 0 & \text{otherwise.}\end{cases} \tag{25}$$

7. Averaging distribution

As discussed in Section 5.2, the distribution $q(z,l|c,l_o,\Theta)=p(z,l|c,l_o,\Theta)$ assures that the solution obtained with the EM algorithm provides a local optimum of the log likelihood. Having taken into account the conditional independence of Z with respect to C and L_o that is implied by the graphical model, we may express $p(z,l|c,l_o,\Theta)$ as:

$$p(z,l|c,l_o,\Theta)=p(z|l,\Theta)p(l|c,l_o,\Theta). \tag{26}$$

We start with the distribution $p(z|l,\Theta)$. We assume that $Z^i=1$ if the observed trace is assigned to the ith component of the mixture and $Z^i=0$ otherwise, so that $\sum_{i=1}^{K}z^i=1$. Using Bayes theorem, we obtain:

$$p(Z^i=1|l,\Theta)=\frac{\pi_i f_i'(l|\theta_i)}{f'(l|\Theta)}. \tag{27}$$

The conditional distribution $p(l|c,l_o,\Theta)$ is then obtained as:

$$p(l|c,l_o,\Theta)=\frac{p(l,c,l_o|\Theta)}{\int_{l_o}^{\infty}p(\xi,c,l_o|\Theta)\,d\xi}, \tag{28}$$

where

$$p(l,c,l_o|\Theta)=f'(l|\Theta)p(c|l)p(l_o|c,l), \tag{29}$$

and $p(c|l)$ and $p(l_o|c,l)$ are obtained as in Section 6. The distributions for each censoring case are derived below.

7.1. Traces with no censoring

Total trace lengths are identical to observed lengths in this case. The conditional distribution may be expressed as:

$$p(l|C=0,l_o,\Theta)=\begin{cases}\delta(l-l_o) & \text{if } 0<l_o<H_o,\\ 0 & \text{otherwise.}\end{cases} \tag{30}$$

7.2. Traces censored on one side only

We start from Eq. (28), and we marginalize with respect to the (unobserved) trace length. Substituting Eqs. (22) and (23) back into Eq. (29), we obtain the following distribution (defined for $0<l_o<H_o$ and $l\geqslant 0$, with $p(l|C=1,l_o,\Theta)=0$ otherwise):

$$p(l|C=1,l_o,\Theta)=\frac{(1/(l+H_o))f'(l|\Theta)}{\int_{l_o}^{\infty}(1/(\xi+H_o))f'(\xi|\Theta)\,d\xi}. \tag{31}$$

Table 2
Statistical distributions and intensity values used for generation of discontinuity traces

(a) *Distribution types and parameters*

Exponential	Lognormal	
μ [m]	μ [m]	σ [m]
5.0	5.0	2.0
10.0	10.0	4.0
15.0	20.0	8.0
20.0		
30.0		

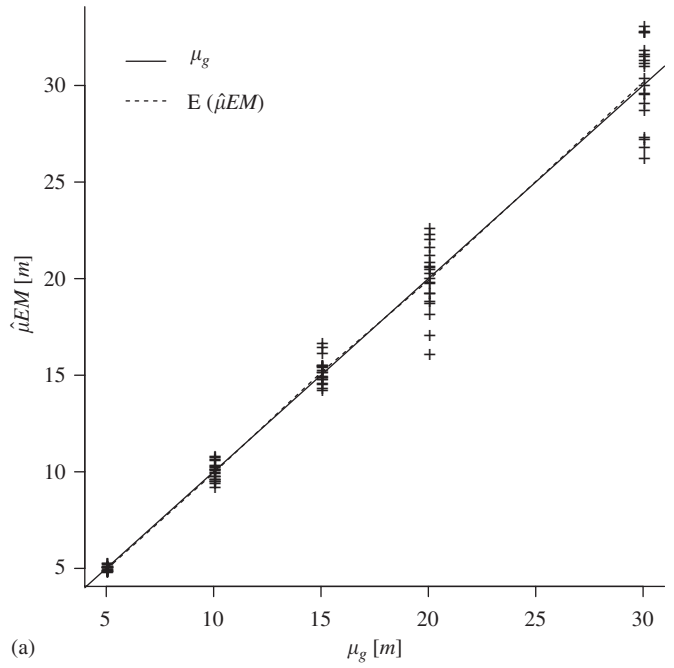
(b) *Intensity values*

P_{22} [m/m^2]
0.5
1.0
2.0
5.0

7.3. Trace censored on both sides

We marginalize with respect to the (unobserved) trace length in Eq. (28), and we substitute (24) and (25) back into the probability model in Eq. (29), obtaining the following distribution (for $l_o = H_o$ and $l \geqslant H_o$; with $p(l|C=2, l_o, \Theta) = 0$ otherwise):

$$p(l|C=2, l_o, \Theta) = \frac{((l-H_o)/(l+H_o))f'(l|\Theta)}{\int_{l_o}^{\infty}((\xi-H_o)/(\xi+H_o))f'(\xi|\Theta)\,d\xi}. \tag{32}$$



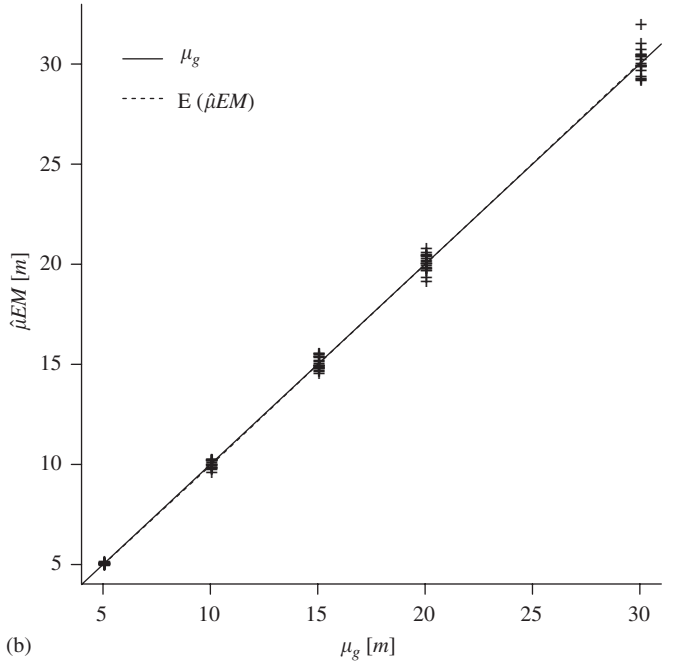


Fig. 6. Computed parameter estimates when an exponential distribution is used for generation and target. (a) $P_{22} = 0.5\,\mathrm{m}^{-1}$, (b) $P_{22} = 5.0\,\mathrm{m}^{-1}$.

8. M step optimization

The M step of the EM algorithm is equivalent to maximizing the expected complete log likelihood with respect to the parameters of the target trace length distribution. Substituting $p(z, l, c, l_o|\Theta)$, and $q(z, l|c, l_o, \Theta)$ (see Eqs. (19) and (26)), back into the expected complete log likelihood in Eq. (16), we obtain:

$$\langle l_c(\Theta; \mathscr{D}_o)\rangle_q = \sum_{n=1}^{N_o}\sum_{z}\int_l p(z|l, \Theta^{(t)})p(l|c, l_o, \Theta^{(t)}) \times \left(\sum_{i=1}^{K} z^i \log(\pi_i^{(t+1)} f_i'(l|\theta_i^{(t+1)})) + \log p(c|l) + \log p(l_o|c, l)\right) dl, \tag{33}$$

where the $^{(t)}$ superscript indicates *known* estimates of the target distribution parameters at the previous step, and the $^{(t+1)}$ superscript indicates *unknown* parameters (which we aim to compute) that maximize the expected complete log likelihood at the current M step. The $^{(t+1)}$ parameters only appear in the first term of the expression in parenthesis in Eq. (33). Using that term, we obtain:

$$\sum_{n=1}^{N_o}\int_l p(l|c, l_o, \Theta^{(t)}) \sum_{i=1}^{K} p(Z^i = 1|l, \Theta^{(t)}) \times \log(\pi_i^{(t+1)} f_i'(l|\theta_i^{(t+1)}))\,dl. \tag{34}$$

Table 3
Computed parameter estimates, and their variability, when an exponential distribution is used for generation and target

μ_g [m]	$E(\hat{\mu}_{EM})$ [m]	$\mathrm{Var}[\hat{\mu}_{EM}]$ [m^2]	$\delta(\hat{\mu}_{EM})$
(a) $P_{22} = 0.5\,\mathrm{m}^{-1}$			
5.00	4.94	0.02	0.06
10.00	9.94	0.22	0.15
15.00	15.13	0.43	0.17
20.00	19.90	2.82	0.38
30.00	30.15	4.55	0.39
(b) $P_{22} = 1.0\,\mathrm{m}^{-1}$			
5.00	5.02	0.01	0.05
10.00	9.87	0.10	0.10
15.00	15.12	0.19	0.11
20.00	20.09	0.85	0.21
30.00	30.15	2.28	0.27
(c) $P_{22} = 2.0\,\mathrm{m}^{-1}$			
5.00	5.02	0.02	0.06
10.00	9.99	0.07	0.08
15.00	14.81	0.17	0.11
20.00	20.06	0.67	0.18
30.00	29.70	1.36	0.21
(d) $P_{22} = 5.0\,\mathrm{m}^{-1}$			
5.00	4.99	0.00	0.02
10.00	9.94	0.04	0.06
15.00	14.99	0.09	0.08
20.00	19.99	0.17	0.09
30.00	30.05	0.44	0.12

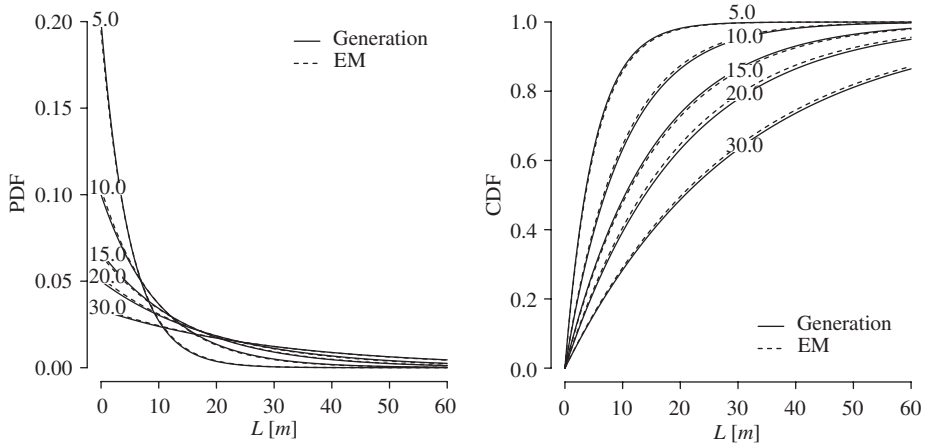


Fig. 7. Comparison between the exponential distribution used for generation (with $P_{22} = 1.0\,\mathrm{m}^{-1}$) and that inferred using the EM algorithm.

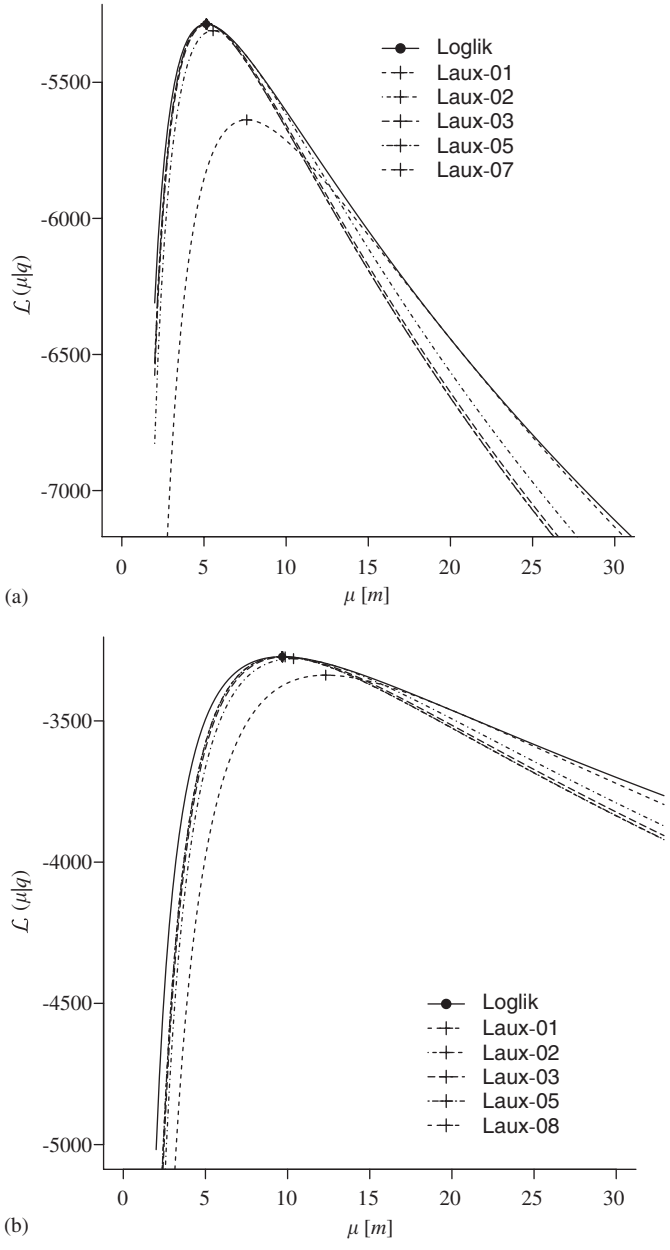


Fig. 8. Evolution of the auxiliary function for successive iterations of the EM algorithm in the exponential distribution case, and its relation to the log likelihood ($P_{22} = 1.0\,\mathrm{m}^{-1}$). (a) $\mu_g = 5.0\,\mathrm{m}$, (b) $\mu_g = 10.0\,\mathrm{m}$.

The $^{(t+1)}$ parameters in Eq. (34) are "decoupled"; that is, they appear in different terms within the summation expression and they can be computed independently, hence reducing the dimensionality of the optimization problem. The reduction of the dimensionality of the optimization problem is not, however, the only advantage of the EM algorithm. A closed-form solution of the optimal mixture proportions at each M step, $\pi^{(t+1)}$, may be obtained in general [60]. This means that, instead of an optimization problem with $K(m+1)-1$ unknowns, we only have to solve K numerical optimization sub-problems, each one of dimension m (m is the number of parameters in the component distributions—e.g., $m = 2$ in the case of Gaussian components). Furthermore, closed-form solutions for the m parameters of each mixture component can be obtained when certain types of probability distributions are used. This completely eliminates the need to perform numerical optimization in the M step, greatly increasing the computational efficiency of the optimization process. (For the solution of the Gaussian mixture component case, see [60].)

9. Example application

9.1. Introduction

To illustrate and validate the proposed methodology, we use discontinuity traces generated using the model in Section 2. The sampling domain is rectangular, with dimensions $W_o = 150\,\mathrm{m}$ and $H_o = 50\,\mathrm{m}$, and the generation domain has dimensions ten times larger than the sampling domain, so that biases and edge effects are negligible during trace generation.

We present two sets of examples: one for models with a single target distribution (it is therefore a simplified version of the model in Fig. 3 in which $K = 1$), and another set of

examples for models using mixture target distributions with Gaussian components. In particular, we test the inference capabilities and the convergence of the EM algorithm for several distribution parameters and trace intensity values [61].

9.2. Single distribution case

Given the simplicity of this model, the size bias correction can be included in the derivations and the parameters of the *real* trace length distribution can be inferred directly—thus, they can be compared with the generation parameters; see [60]. Table 2 presents the distribution parameters and the intensity values used for generation of traces; twenty simulations were performed for each possible combination of distribution parameters and intensity values. (We use exponential and lognormal target distributions because they are commonly used to model discontinuity trace lengths [21].)

9.2.1. Exponential distribution results

Fig. 6 shows examples of computed ML parameter estimates, $\hat{\mu}_{EM}$; Table 3 presents the mean parameter estimates obtained for each intensity value, together with

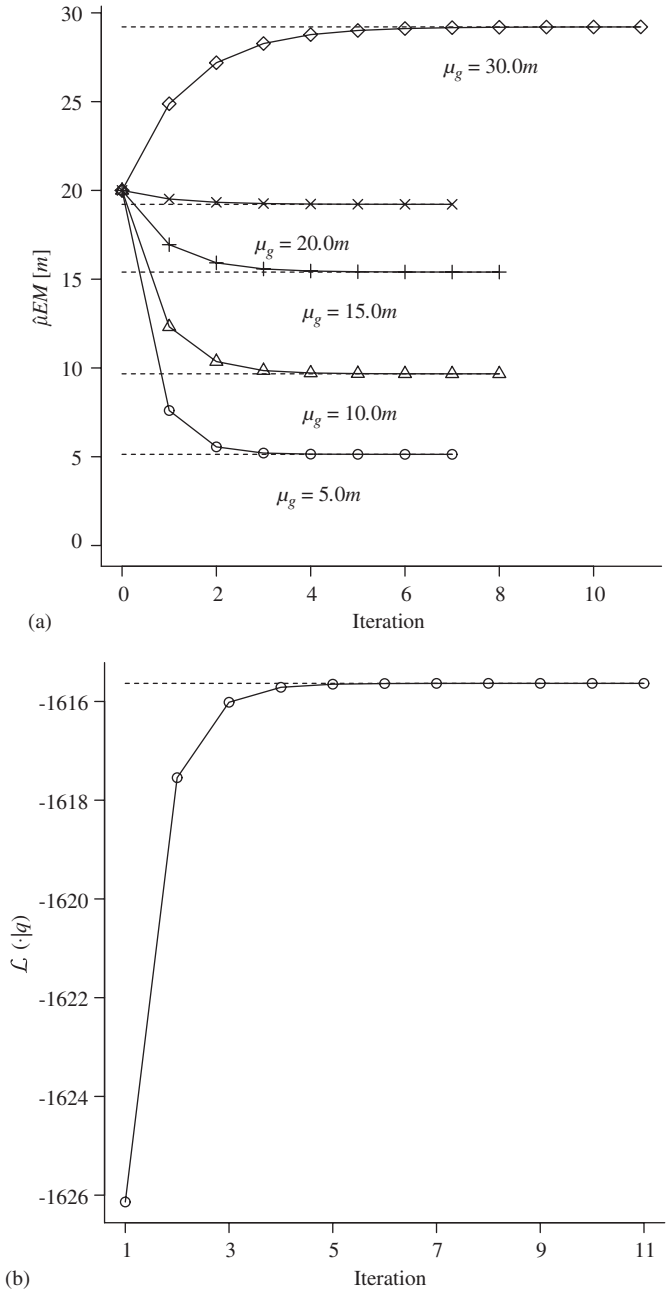


Fig. 9. Convergence performance in the exponential distribution case ($P_{22} = 1.0\,\mathrm{m}^{-1}$). (a) Inferred parameters, (b) log likelihood ($\mu_g = 30.0\,\mathrm{m}$).

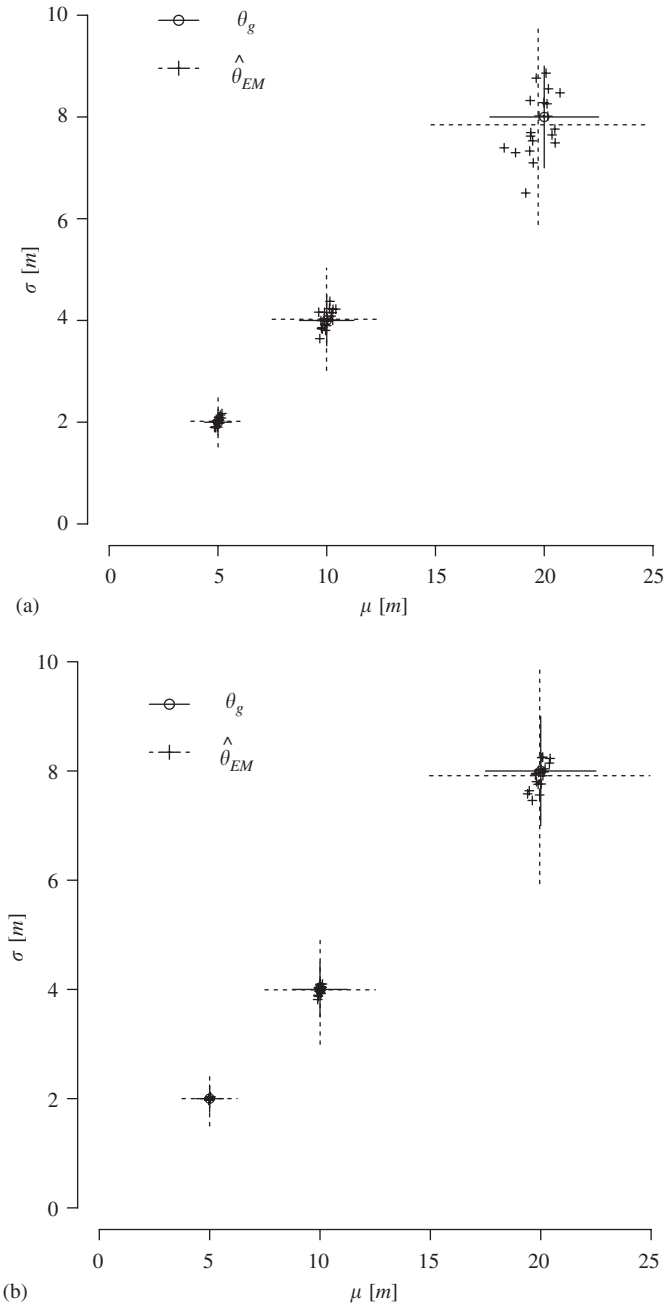


Fig. 10. Computed parameter estimates when a lognormal distribution is used for generation and target. (a) $P_{22} = 0.5\,\mathrm{m}^{-1}$, (b) $P_{22} = 5.0\,\mathrm{m}^{-1}$.

their variances and coefficients of variation. The variability increases as traces are larger with respect to the sampling domain size, whereas it decreases as the number of observations increases. In Fig. 7, we compare the generation distribution and the distribution inferred using the EM algorithm. These results show the good inference capabilities of the method: i.e., the means of the inferred parameters match the parameters of the original distributions; and the inferred distributions closely match the original distributions.

Fig. 8 is a plot of the log likelihood function and the auxiliary functions at several iterations (until convergence) of the algorithm. As can be seen, the auxiliary function is a lower bound to the log likelihood function, whose maximum monotonically approaches that of the log likelihood. Another observation is that—due to the problem of censoring bias [27,28]—the log likelihood becomes "flatter" as the relative length of discontinuities increases with respect to the size of the sampling domain. We observe, however, that the EM algorithm behaves well in all cases, converging to the asymptotic solution of the problem (marked using solid circles in Fig. 8).

In Fig. 9(a) we present the convergence performance of the computed parameter estimates; Fig. 9(b) illustrates the convergence with respect to log likelihood values. A dashed line indicates the asymptotic maximum likelihood solution in both cases. The algorithm is shown to converge quickly and values very close to the asymptotic solution are achieved after only a few iterations.

9.2.2. Lognormal distribution results

Fig. 10 shows the computed parameter estimates for each simulation; the generation parameters (solid lines) and the mean values of the computed parameter estimates (dashed lines) are indicated as well. As before, the variability of the computed estimates increases as the size of traces increases (i.e., as censoring bias increases), and it decreases as the intensity value increases (i.e., as we observe more traces). The variances and coefficients of variation of the lognormal distribution parameters used as target are summarized in Table 4. Fig. 11 shows comparisons between the generation distribution (solid lines) and the inferred distribution (dashed lines). Again, the inference capabilities are good, and the inferred distribution closely matches the original one.

Fig. 12 shows the evolution of the auxiliary function at several iterations of the EM algorithm. The log likelihood function is represented using a dotted surface with dashed boundaries, whereas the auxiliary functions are represented with solid surfaces. (The maximum values of the log likelihood and of the auxiliary function are plotted as well, using upward and downward-pointing triangles.) The auxiliary function is a lower bound of the log likelihood so that, at convergence, the maximum of the auxiliary function is equal to the maximum of the log likelihood, hence providing a local solution of the maximum likelihood estimation problem.

Next, we present the convergence performance of the algorithm, both with respect to the computed parameter estimates and with respect to the auxiliary function values. In Fig. 13(a) we show an example of the evolution of the parameter estimates at several iterations of the EM algorithm. (The initial parameter set is marked using a solid-line cross, and computed parameters are marked using dashed-line crosses.) Finally, Fig. 13(b) illustrates the convergence in log likelihood. As before, the convergence is shown to be fast in both cases, and a value very similar to the asymptotic solution is usually achieved in less than ten iterations.

9.3. Mixture distribution case

We explore the performance of the methodology when the mixture model in Fig. 3 is used for inference of the

Table 4
Computed parameter estimates, and their variability, when a lognormal distribution is used for generation and target

μ_g [m]	$E(\hat{\mu}_{EM})$ [m]	$\mathrm{Var}(\hat{\mu}_{EM})$ [m^2]	$\delta(\hat{\mu}_{EM})$	σ_g [m]	$E(\hat{\sigma}_{EM})$ [m]	$\mathrm{Var}(\hat{\mu}_{EM})$ [m^2]	$\delta(\hat{\mu}_{EM})$
(a) $P_{22} = 0.5\,\mathrm{m}^{-1}$							
5.00	5.01	0.01	0.04	2.00	2.02	0.01	0.06
10.00	9.99	0.05	0.07	4.00	4.02	0.03	0.09
20.00	19.72	0.42	0.15	8.00	7.85	0.36	0.21
(b) $P_{22} = 1.0\,\mathrm{m}^{-1}$							
5.00	4.99	0.00	0.02	2.00	1.99	0.00	0.03
10.00	9.98	0.01	0.04	4.00	4.01	0.01	0.04
20.00	20.01	0.17	0.09	8.00	7.99	0.24	0.17
(c) $P_{22} = 2.0\,\mathrm{m}^{-1}$							
5.00	4.99	0.00	0.02	2.00	2.00	0.00	0.02
10.00	9.99	0.01	0.04	4.00	3.99	0.01	0.05
20.00	19.93	0.11	0.07	8.00	7.96	0.12	0.12
(d) $P_{22} = 5.0\,\mathrm{m}^{-1}$							
5.00	5.00	0.00	0.01	2.00	2.00	0.00	0.02
10.00	10.01	0.00	0.02	4.00	3.99	0.01	0.04
20.00	19.95	0.07	0.06	8.00	7.91	0.05	0.08

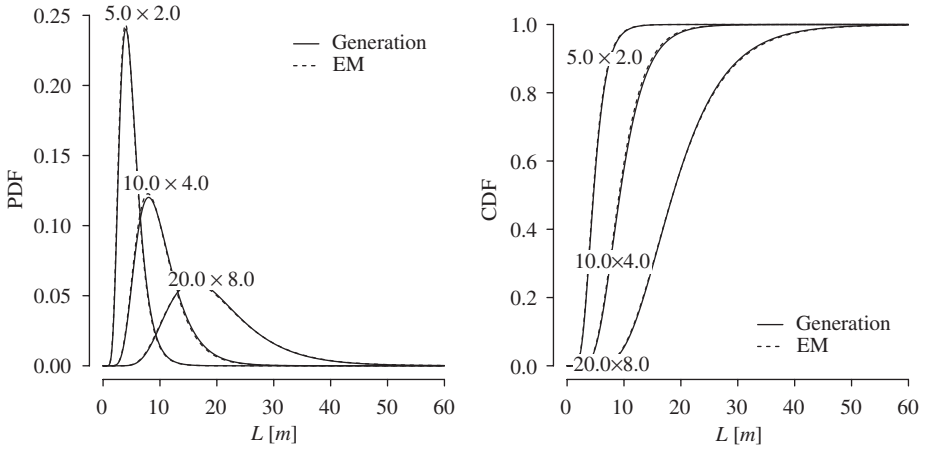


Fig. 11. Comparison between generation and inferred distributions in the lognormal distribution case ($P_{22} = 2.0\,\mathrm{m}^{-1}$).

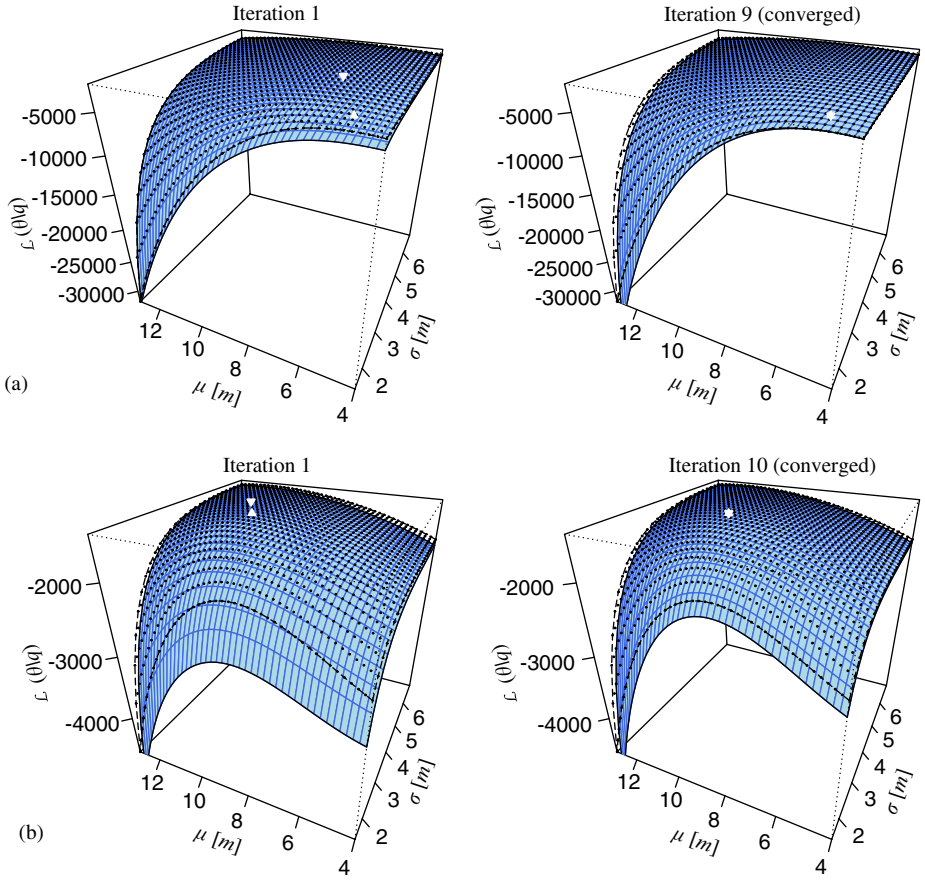


Fig. 12. Examples of evolution of auxiliary functions for different iterations of the EM algorithm, and their relation to the log likelihood ($P_{22} = 0.5\,\mathrm{m}^{-1}$). (a) $\theta_g = 5.0 \times 2.0\,\mathrm{m}$, (b) $\theta_g = 10.0 \times 4.0\,\mathrm{m}$.

distribution of discontinuity trace lengths. The distribution that we use for generation of traces is composed of a mixture of four lognormally distributed components, whose parameters are listed in Table 5(a). Such generation distribution provides a challenging test for the methodology because it is multi-modal, and because it has a significant probability of producing traces longer than about half the size of the sampling window—hence increasing the influence of biases (see Section 9.2). Table 5(b) presents intensity values used for generation of discontinuities and the corresponding mean (after 100 simulations) of the number of observed traces, N_o.

We use target mixtures with $K = 3$, 5, 8, and 12 Gaussian components for trace length distribution inference. This allow us to avoid making strong a-priori assumptions about the most adequate type of distribution;

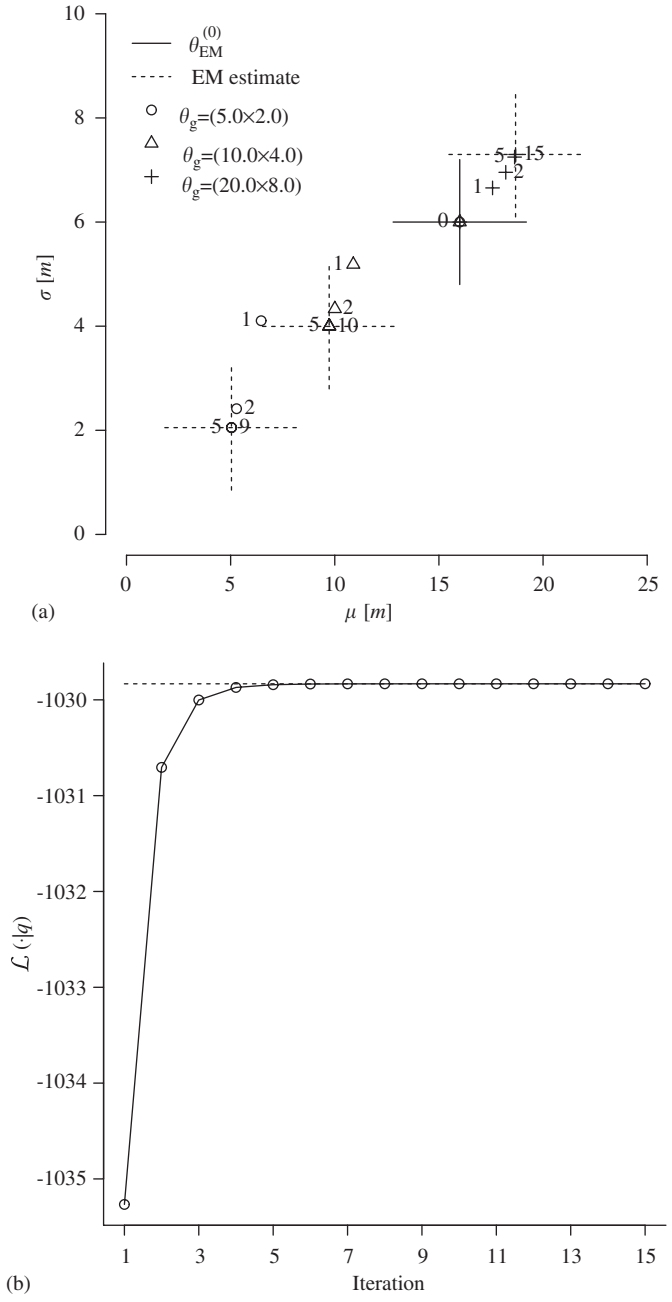


Fig. 13. Convergence performance in the lognormal distribution case ($P_{22} = 0.5\,\mathrm{m}^{-1}$). (a) Inferred parameters, (b) log likelihood ($\theta_g = 20.0 \times 8.0\,\mathrm{m}$).

computational efficiency is another advantage, since closed-form expressions can be obtained for the maximization performed in the M step, thus avoiding the need for numerical optimization and reducing the computational cost. To check the similarity between the original distribution, $f_g(l)$, and the inferred distribution, $f(l)$, we use the Kullback–Leibler (KL) divergence [62]:

$$D(f_g(\cdot)||f(\cdot)) = \int_l f_g(l) \log \frac{f_g(l)}{f(l)}\,\mathrm{d}l. \quad (35)$$

Table 6 presents the influence of the number of target components and the number of observed traces on the computed values of KL divergence. (Twenty simulations were performed for each combination of K and P_{22} values.) The predictive capabilities of the method are observed to improve (i.e., the KL divergence decreases) as the intensity value increases (i.e., as we have more observed traces). We also observe that increasing the number of target mixture components does not necessarily lead to improved inference capabilities; this is due to over-fitting, an effect that occurs when one or more of the target mixture components

Table 5
Mixture distribution used for generation of trace lengths, intensity values, and expected number of observed traces

(a) *Types of distributions and parameter values used as mixture components of the generation distribution*

Component	Type	Parameters		
i		π_i	μ_i [m]	σ_i [m]
1	Lognormal	0.25	9.0	4.0
2	Lognormal	0.15	20.0	4.0
3	Lognormal	0.30	31.0	3.0
4	Lognormal	0.30	40.0	4.0

(b) *Mean number of traces observed for each intensity value considered (after 100 simulations)*

P_{22} [m/m^2]	$E(N_o)$
0.5	217
1.0	435
2.0	865
5.0	2160

Table 6
Influence of number of target distribution components and number of observed traces on computed KL divergence values

	P_{22} [m^{-1}]	KL divergence			
		Mean	Std. dev.	Min	Max
$K = 3$	0.5	7.71×10^{-02}	3.56×10^{-02}	3.92×10^{-02}	1.83×10^{-01}
	1.0	5.81×10^{-02}	1.87×10^{-02}	3.72×10^{-02}	1.08×10^{-01}
	2.0	4.12×10^{-02}	1.05×10^{-02}	2.40×10^{-02}	6.49×10^{-02}
	5.0	3.37×10^{-02}	5.11×10^{-03}	2.63×10^{-02}	4.42×10^{-02}
$K = 5$	0.5	1.33×10^{-01}	6.96×10^{-02}	2.88×10^{-02}	2.96×10^{-01}
	1.0	5.84×10^{-02}	1.88×10^{-02}	2.62×10^{-02}	9.11×10^{-02}
	2.0	3.05×10^{-02}	1.33×10^{-02}	5.35×10^{-03}	5.68×10^{-02}
	5.0	1.58×10^{-02}	6.01×10^{-03}	6.09×10^{-03}	2.97×10^{-02}
$K = 8$	0.5	2.37×10^{-01}	1.47×10^{-01}	8.48×10^{-02}	5.35×10^{-01}
	1.0	8.66×10^{-02}	4.01×10^{-02}	3.11×10^{-02}	1.74×10^{-01}
	2.0	4.60×10^{-02}	1.71×10^{-02}	2.25×10^{-02}	7.67×10^{-02}
	5.0	2.66×10^{-02}	8.45×10^{-03}	1.58×10^{-02}	4.88×10^{-02}
$K = 12$	0.5	3.60×10^{-01}	2.44×10^{-01}	1.10×10^{-01}	9.49×10^{-01}
	1.0	8.86×10^{-02}	3.44×10^{-02}	2.75×10^{-02}	1.69×10^{-01}
	2.0	5.24×10^{-02}	2.01×10^{-02}	2.38×10^{-02}	1.08×10^{-01}
	5.0	2.40×10^{-02}	6.04×10^{-03}	1.52×10^{-02}	3.57×10^{-02}

are assigned a small variance [55,63]. The occurrence of over-fitting problems may be identified when spikes (corresponding to mixture components with low variance) are observed in the inferred PDF; in this research we found that over-fitting problems are likely to appear when the ratio between the number of observed discontinuities and the number of mixture components is smaller than approximately 150 (i.e., $N_o/K < 150$).

Several methods have been proposed to reduce the occurrence of over-fitting. The introduction of penalty terms in the log likelihood is sometimes employed [64], whereas other authors have proposed to solve a constrained maximization problem in the M step [65–67]. With this approach, restrictions are imposed, for instance, on the ratio between the maximum and the minimum standard deviations of the mixture components; minimum values for the mixture proportions may be specified as well. This method has been shown to be well posed, with a consistent, global solution [66,67]; for an illustration of the improvements obtained, see [60].

Fig. 14 compares the original distribution with inferred distributions (after bias correction) for examples cases with different number of target components. In all cases, we generate traces using an intensity value of $P_{22} = 5.0\,\mathrm{m}^{-1}$, observing a number of traces in the order of $N_o \approx 2160$ (see Table 5(b)). The inference capability of the methodology is generally quite good. However, only a maximum of three modes may be reproduced with a mixture with three components and, accordingly, we also see that the results improve when more than three mixture components are employed. In that sense, the results are particularly good when five to eight mixture components are used (see Fig. 14 and Table 6). Based on these observations, we suggest using target distributions with about two mixture components for each mode of the *real* trace length distribution, as long as there are at least 250 observed traces for each component in the target distribution.

Fig. 15 shows an example of the evolution of the log likelihood function values computed at each iteration of the EM algorithm. The results show that the convergence in log likelihood is fast, and values very similar to the asymptotic value are usually obtained after fifteen or twenty iterations.

10. Conclusions

We present a method for the estimation of the distribution of discontinuity trace lengths based on observations at rock outcrops. The methodology is based in the use of a statistical graphical model approach in

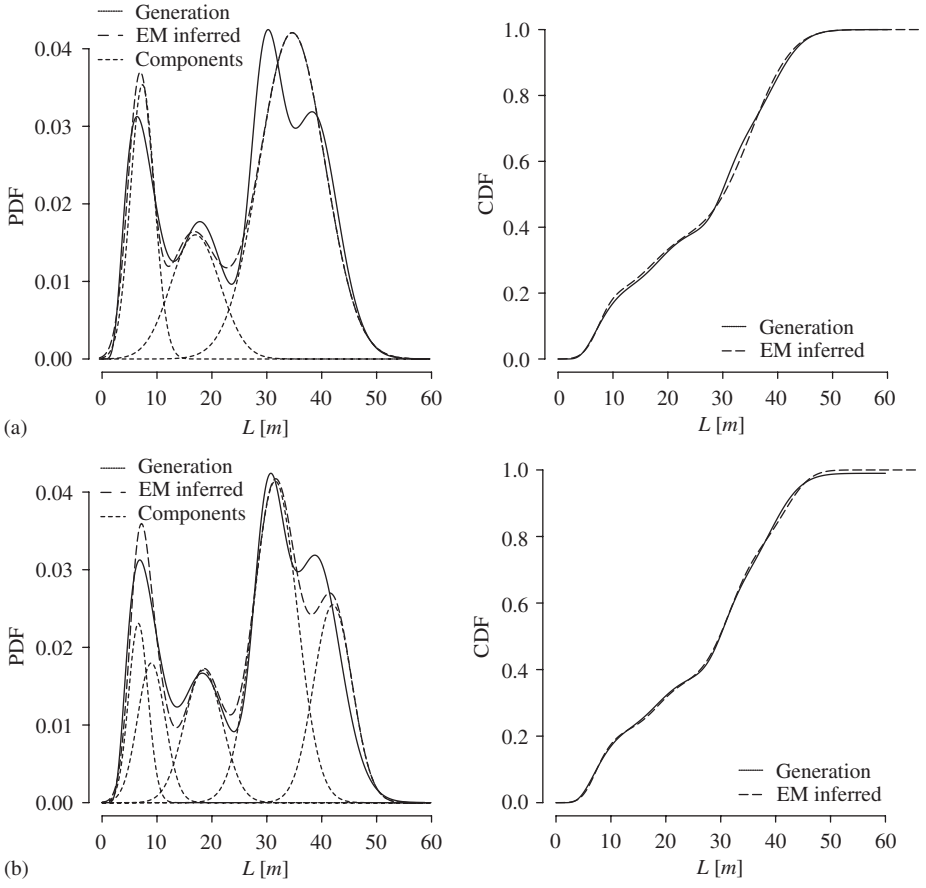


Fig. 14. Examples of inferred distributions (after bias correction) for different number of target mixture components ($P_{22} = 5.0\,\mathrm{m}^{-1}$). (a) $K = 3$ ($N_o/K \approx 720$), (b) $K = 5$ ($N_o/K \approx 432$), (c) $K = 8$ ($N_o/K \approx 270$), (d) $K = 12$ ($N_o/K \approx 180$).

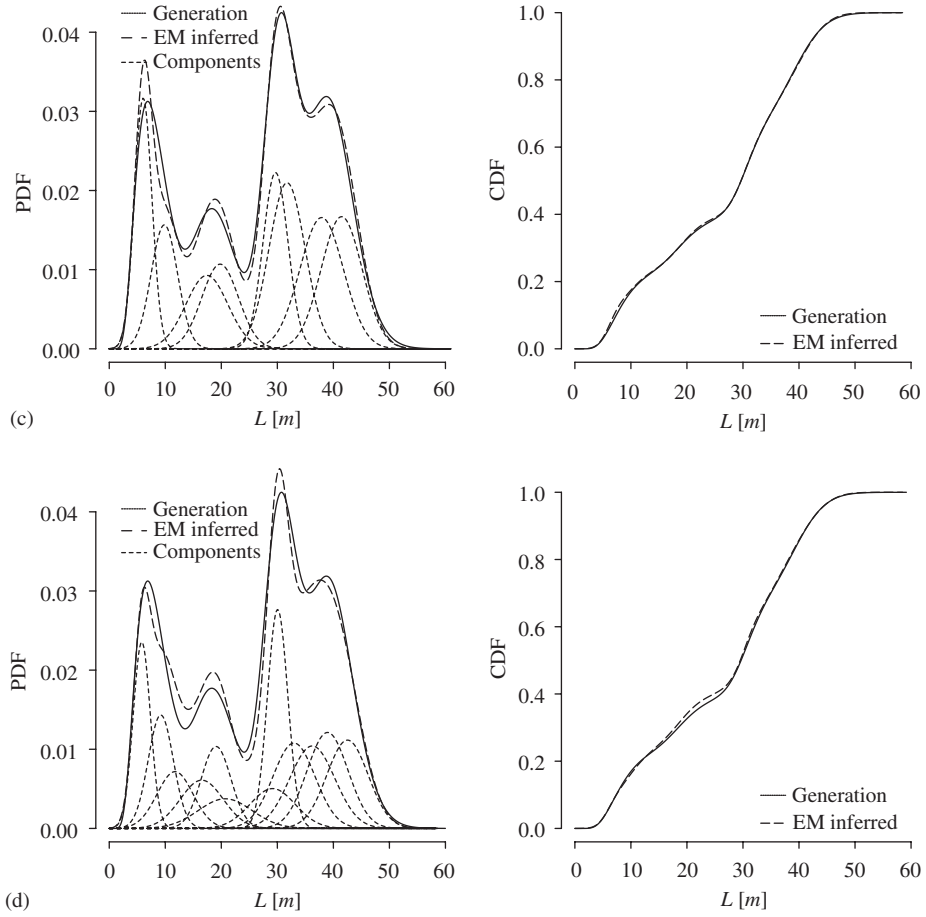


Fig. 14. (*Continued*)

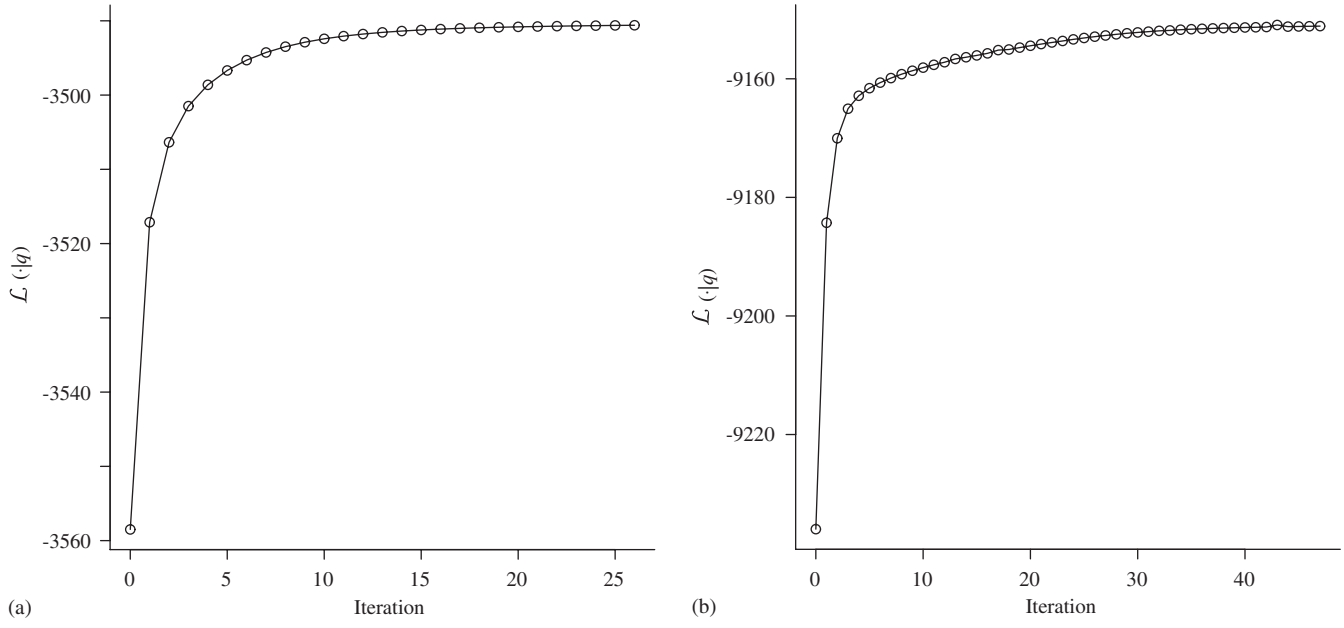


Fig. 15. Examples of evolution of the log likelihood values with respect to the iteration number of the EM algorithm. (a) $K = 5$; $P_{22} = 2.0\,\mathrm{m}^{-1}$, (b) $K = 8$; $P_{22} = 5.0\,\mathrm{m}^{-1}$.

which we consider a wide, flexible class of target probability distributions that allow us to obtain reasonable estimates without making strong assumptions about its type. Statistical graphical models further allow us to take advantage of the relations of conditional independence between the variables in the model. In particular, we use

the expectation–maximization (EM) algorithm to solve the maximum likelihood parameter estimation problem, as it provides a natural way of exploiting the conditional independence relations in models with unobserved variables.

The EM algorithm is an iterative optimization algorithm, in which a lower bound to the log likelihood to be maximized is used as an auxiliary function in the optimization process. The E step needs to be solved only once in each particular problem, and the computed averaging function assures that the EM solution is a local maximum of the log likelihood. Similarly, the maximization at the M step may be performed by maximizing the expected log likelihood at each iteration; the algorithm "decouples" the parameters of interest—hence reducing the dimensionality of the optimization problem—and closed-form analytical expressions can be developed when target mixtures with Gaussian components are employed—thus avoiding the need for numerical optimization and reducing the computational cost.

Artificial trace data generated using Monte Carlo methods are used to illustrate and validate the proposed methodology. As expected, the performance improves as the length of traces decreases, and it also improves as the number of observed traces increases. Results show that adequate estimates of trace length distribution parameters are obtained when single distributions are used as target; the inferred distributions have also been shown to compare well with the original distributions. Similarly, the overall inference capabilities are good in the Gaussian mixture case, and they improve as we increase the number of target components, as long as over fitting effects are avoided. Finally, the convergence of the algorithm is shown to be fast, particularly with respect to the log likelihood values at each iteration, and the results show that computed log likelihoods very similar to the asymptotic value are usually obtained after fifteen or twenty iterations.

Acknowledgements

The authors are indebted to Professor Peter L. Bartlett for his advice during the development of this work. Financial support for this research was provided by the Jane Lewis Fellowship of the University of California and the National Science Foundation (grant CMS-0355079). The support of *Grupo de Investigaciones Medioambientales: Riesgos Geológicos e Ingeniería del Terreno* (RNM 221, PAI) of the University of Granada is gratefully acknowledged as well.

References

[1] Hudson JA, Harrison JP. Engineering rock mechanics: an introduction to the principles, 1st ed. Tarrytown, NY: Pergamon; 1997.
[2] Goodman RE. Methods of geological engineering in discontinuous rocks. St. Paul: West Publishing Co.; 1976.
[3] ISRM. Suggested methods for the quantitative description of discontinuities in rock masses. Int J Rock Mech Mining Sci Geomech Abstracts 1978;15(6):319–68.
[4] Priest SD. The collection and analysis of discontinuity orientation data for engineering design, with examples. In: Hudson JA, editor. Comprehensive rock engineering; principles, practice & projects: rock testing and site characterization. Oxford: Pergamon Press; 1993. p. 167–92.
[5] La Pointe PR. Pattern analysis and simulation of joints for rock engineering. In: Hudson JA, editor. Comprehensive rock engineering. volume 3—rock testing and site characterization. Oxford: Pergamon Press; 1993. p. 215–39.
[6] Dershowitz WS, Einstein HH. Characterizing rock joint geometry with joint system models. Rock Mech Rock Eng 1988;21(1):21–51.
[7] Kulatilake PHSW, Wathugala DN, Stephansson O. Stochastic three dimensional joint size, intensity and system modelling and a validation to an area in Stripa Mine, Sweden. Soils Foundations 1993;33(1):55–70.
[8] Meyer T, Einstein HH. Geologic stochastic modeling and connectivity assessment of fracture systems in the Boston area. Rock Mech Rock Eng 2002;35(1):23–44.
[9] Warburton PM. A stereological interpretation of joint trace data. Int J Rock Mech Mining Sci Geomech Abstracts 1980;17:181–90.
[10] Warburton PM. Stereological interpretation of joint trace data: influence of joint shape and implications for geological surveys. Int J Rock Mech Mining Sci Geomech Abstracts 1980;17(6):305–16.
[11] Bonnet E, Bour O, Odling N, Davy P, Main I, Cowie P, Berkowitz B. Scaling of fracture systems in geological media. Rev Geophys 2001;39(3):347–83.
[12] La Pointe PR. Derivation of parent fracture population statistics from trace length measurements of fractal fracture populations. Int J Rock Mech Mining Sci 2002;39:381–8.
[13] Lyman GJ. Stereological and other methods applied to rock joint size estimation—does Crofton's theorem apply? Math Geol 2003;35(1):9–23.
[14] Zhang L, Einstein HH, Dershowitz WS. Stereological relationship between trace length and size distribution of elliptical discontinuities. Géotechnique 2002;52(6):419–33.
[15] Priest SD. Determination of discontinuity size distributions from scanline data. Rock Mech Rock Eng 2004;37(5):347–68.
[16] Terzaghi RD. Sources of errors in joint surveys. Geotechnique 1965;15:287–304.
[17] Zhang L, Einstein HH. Estimating the mean trace length of rock discontinuities. Rock Mech Rock Eng 1998;31(4):217–35.
[18] Villaescusa E, Brown ET. Maximum likelihood estimation of joint size from trace length measurements. Rock Mech Rock Eng 1992;25(2):67–87.
[19] Kulatilake PHSW, Wu TH. Estimation of mean trace length of discontinuities. Rock Mech Rock Eng 1984;17(4):243–53.
[20] Song JJ, Lee CI. Estimation of joint length distribution using window sampling. Int J Rock Mech Mining Sci 2001;38(4):519–28.
[21] Zhang LY, Einstein HH. Estimating the intensity of rock discontinuities. Int J Rock Mech Mining Sci 2000;37(5):819–37.
[22] Wu F, Wang S. Statistical model for structure of jointed rock mass. Géotechnique 2002;52(2):137–40.
[23] Laslett GM. Censoring and edge effects in areal and line transect sampling of rock joint traces. J Int Assoc Math Geol 1982;14(7):125–40.
[24] Laubach S, Marrett R, Olson J. New directions in fracture characterization. The leading edge 2000. p. 704–11. [Also available at http://www.beg.texas.edu/indassoc/fraccity/public/documents/TLE%Fractures.pdf].
[25] Pahl PJ. Estimating the mean length of discontinuity traces. Int J Rock Mech Mining Sci Geomech Abstracts 1981;18(3):221–8.
[26] Mauldon M. Estimating mean fracture trace length and density from observations in convex windows. Rock Mech Rock Eng 1998;31(4):201–16.
[27] Lee JS, Veneziano D, Einstein HH. Hierarchical fracture trace model. In: Hustrulid WA, Johnson GA, editors. Rock mechanics; contributions and challenges; proceedings of the 31st U.S. symposium. Rotterdam: A.A. Balkema; 1990. p. 261–8.

[28] Einstein HH. Modern developments in discontinuity analysis—the persistence-discontinuity problem. In: Hudson JA, editor. Comprehensive rock engineering, volume 3—rock testing and site characterization. New York: Pergamon Press; 1993. p. 215–39.
[29] Lyman GJ. Rock fracture mean trace length estimation and confidence interval calculation using maximum likelihood methods. Int J Rock Mech Mining Sci 2003;40:825–32.
[30] Jimenez-Rodriguez R, Sitar N, Bartlett PL. Maximum likelihood estimation of trace length distribution parameters using the EM algorithm. In: Barla G, Barla M, editors. Prediction, analysis and design in geomechanical applications: proceedings of the eleventh international conference on computer methods and advances in geomechanics (IACMAG-2005), vol. 1. Bologna: Pàtron Editore; 2005. p. 619–26.
[31] Jimenez-Rodriguez R, Sitar N, Chacón J. Caracterización de discontinuidades en macizos rocosos mediante modelos gráficos probabilísticos [Characterization of discontinuities in rock masses by means of probabilistic graphical models]. In: Pérez Aparicio J, Rodríguez Ferrán A, Martins J, Gallego R, César de Sá J, editors. Métodos Numéricos en Ingeniería 2005 [Numerical Methods in Eng 2005]. Barcelona: SEMNI, Sociedad Española de Métodos Numéricos en Ingeniería; 2005. p. 222. Extended abstract in Proceedings; full paper in CD-Rom.
[32] Priest SD. Discontinuity analysis for rock engineering, 1st ed. London, New York: Chapman & Hall; 1993.
[33] Fasching A, Gaich A, Schubert W. Data acquisition in engineering geology. An improvement of acquisition methods for geotechnical rock mass parameters. Felsbau 2001;19(5):93–101.
[34] Lemy F, Hadjigeorgiou J. Discontinuity trace map construction using photographs of rock exposures. Int J Rock Mech Mining Sci 2003;40(6):903–17.
[35] Hadjigeorgiou J, Lemy F, Côté P, Maldague X. An evaluation of image analysis algorithms for constructing discontinuity trace maps. Rock Mech Rock Eng 2003;36(2):163–79.
[36] Chan LY. Application of block theory and simulation techniques to optimum design of rock excavations. Ph.D. thesis, University of California, Berkeley; 1987.
[37] Baecher GB, Einstein HH, Lanney NA. Statistical description of rock properties and sampling. In: Wang FD, Clark GB, editors. Energy resources and excavation technology; proceedings, 18th U.S. symposium on rock mechanics. Golden: Colo. Sch. Mines Press; 1977. p. 5C1.1–8.
[38] Chan LY, Goodman RE. Predicting the number of dimensions of key blocks of an excavation using block theory and joint statistics. In: Farmer IW, Daemen JJK, Desai CS, Glass CE, Neuman SP, editors. Rock mechanics; proceedings of the 28th U.S. symposium. Rotterdam: A.A. Balkema; 1987. p. 81–7.
[39] Hoerger SF, Young DS. Probabilistic prediction of keyblock occurrences. In: Hustrulid WA, Johnson GA, editors. Rock mechanics; contributions and challenges; proceedings of the 31st U.S. symposium. Rotterdam: A.A. Balkema; 1990. p. 229–36.
[40] Hoerger SF, Young DS. Probabilistic analysis of keyblock failures. In: Rossmanith HP, editor. Mechanics of jointed and faulted rock; proceedings of the international conference. Rotterdam: A.A. Balkema; 1990. p. 503–8.
[41] Song JJ, Lee CI, Seto M. Stability analysis of rock blocks around a tunnel using a statistical joint modeling technique. Tunnelling Underground Space Technol 2001;16(4):341–51.
[42] Barton C, Larsen E. Fractal geometry of two-dimensional fracture networks at Yucca Mountain, Southwestern Nevada. In: Stephansson O, editor. Proceedings of the international symposium on fundamentals of rock joints. Lulea, Sweden: Centek; 1985. p. 77–84.
[43] Boadu FK, Long LT. The fractal character of fracture spacing and rqd. Int J Rock Mech Mining Sci Geomech Abstracts 1994;31(2):127–34.
[44] Kulatilake P, Fiedler R, Panda BB. Box fractal dimension as a measure of statistical homogeneity of jointed rock masses. Engineering Geology 1997;48(3–4):217–29.
[45] Odling N. Scaling and connectivity of joint systems in sandstones from western Norway. J Struct Geol 1997;19(10):1257–71.
[46] Ehlen J. Fractal analysis of joint patterns in granite. Int J Rock Mech Mining Sci 2000;37(6):909–22.
[47] Priest SD, Hudson JA. Estimation of discontinuity spacing and trace length using scanline surveys. Int J Rock Mech Mining Sci Geomech Abstracts 1981;18(3):183–97.
[48] Mauldon M, Dunne WM, Rohrbaugh MB. Circular scanlines and circular windows: new tools for characterizing the geometry of fracture traces. J Struct Geol 2001;23(2–3):247–58.
[49] Kulatilake PHSW, Wu TH. Sampling bias on orientation of discontinuities. Rock Mech Rock Eng 1984;17(4):215–32.
[50] Wathugala DN, Kulatilake PHSW, Wathugala GW, Stephansson O. A general procedure to correct sampling bias on joint orientation using a vector approach. Comput Geotech 1990;10:1–31.
[51] Mauldon M, Mauldon JG. Fracture sampling on a cylinder: from scanlines to boreholes and tunnels. Rock Mech Rock Eng 1997;30(3):129–44.
[52] Laslett GM. The survival curve under monotone density constraints with applications to two-dimensional line segment processes. Biometrika 1982;69(1):153–60.
[53] Starzec P. Characterisation and modeling of discontinuous rock mass: geostatistical interpretation of seismic response to fracture occurrence, discrete fracture network modeling and stability predictions for underground excavations. Ph.D. thesis, Chalmers University of Technology; 2001.
[54] Cox DR. Some sampling problems in technology. In: Johnson NL, Smith Jr. H, editors. New developments in survey sampling. New York: Wiley; 1969. p. 506–27.
[55] Redner RA, Walker HF. Mixture densities, maximum likelihood and the EM algorithm. SIAM Rev 1984;26(2):195–239.
[56] Jordan MI. An introduction to probabilistic graphical models, 2003. (unpublished manuscript). Department of Statistics. University of California, Berkeley.
[57] Xu L, Jordan MI. On convergence properties of the EM algorithm for Gaussian mixtures. Neural Comput 1996;8(1):129–51.
[58] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc (B) 1977;39(1):1–38.
[59] Cowell RG, Dawid AP, Lauritzen SL, Spiegelhalter DJ. Probabilistic networks and expert systems. New York: Springer; 1999.
[60] Jimenez-Rodriguez R, Sitar N. Maximum likelihood inference of discontinuity trace lengths based on observations at rock outcrops. GeoEngineering Report, Department of Civil and Environmental Engineering, University of California, Berkeley; 2005, in preparation.
[61] Dershowitz WS, Herda HH. Interpretation of fracture spacing and intensity. In: Tillerson JR, Wawersik WR, editors. Rock mechanics; Proceedings of the 33rd U.S. symposium Rock mechanics. Rotterdam: A.A. Balkema; 1992. p. 757–66.
[62] Kullback S, Leibler RA. On information and sufficiency. Ann Math Stat 1951;22(1):79–86.
[63] Day NE. Estimating the components of a mixture of normal distributions. Biometrika 1969;56(3):463–74.
[64] Ridolfi A, Idier J. Penalized maximum likelihood estimation for univariate normal mixture distributions. In: Actes du 17^{e} colloque GRETSI, (Vannes, France); 1999. p. 259–62.
[65] Tong B, Viele K. Smooth estimates of normal mixtures. Cana J Stat 2000;28(3):551–61.
[66] Hathaway RJ. A constrained formulation of maximum-likelihood estimation for normal mixture distributions. Ann Stat 1985;13(2):795–800.
[67] Davenport JW, Bezdek JC, Hathaway RJ. Parameter estimation for finite mixture distributions. Comput Math Appl 1988;15(10):819–28.