

# A QUASI-LIKELIHOOD APPROACH TO PARAMETER ESTIMATION FOR SIMULATABLE STATISTICAL MODELS

MARKUS BAASKE<sup>✉</sup>,<sup>1</sup> FELIX BALLANI<sup>1</sup> AND KARL GERALD VAN DEN BOOGAART<sup>1,2</sup>

<sup>1</sup>Institute of Stochastics, Faculty of Mathematics and Computer Science, TU Bergakademie Freiberg, D-09596 Freiberg, Germany; <sup>2</sup>Helmholtz Institute Freiberg for Resource Technology, Halsbrücker Str. 34, D-09599 Freiberg, Germany

e-mail: markus.baaske@math.tu-freiberg.de, ballani@math.tu-freiberg.de, boogaart@hzdr.de

(Received November 13, 2013; revised March 21, 2014; accepted March 24, 2014)

## ABSTRACT

This paper introduces a parameter estimation method for a general class of statistical models. The method exclusively relies on the possibility to conduct simulations for the construction of interpolation-based meta-models of informative empirical characteristics and some subjectively chosen correlation structure of the underlying spatial random process. In the absence of likelihood functions for such statistical models, which is often the case in stochastic geometric modelling, the idea is to follow a quasi-likelihood (QL) approach to construct an optimal estimating function surrogate based on a set of interpolated summary statistics. Solving these estimating equations one can account for both the random errors due to simulations and the uncertainty about the meta-models. Thus, putting the QL approach to parameter estimation into a stochastic simulation setting the proposed method essentially consists of finding roots to a sequence of approximating quasi-score functions. As a simple demonstrating example, the proposed method is applied to a special parameter estimation problem of a planar Boolean model with discs. Here, the quasi-score function has a half-analytical, numerically tractable representation and allows for the comparison of the model parameter estimates found by the simulation-based method and obtained from solving the exact quasi-score equations.

Keywords: kriging meta-modelling, parameter estimation, quasi-likelihood, simulation-based optimization.

## INTRODUCTION

Our initial aim motivating this research is fitting parameters of statistical models from stochastic geometry for modelling porous media. In many of these models no closed forms or direct computation algorithms for likelihood functions or distribution characteristics as function of the model parameters are known. This precludes many standard techniques of parameter estimation, such as maximum likelihood (ML), typical Bayesian algorithms (including Markov-chain-Monte-Carlo-type algorithms), or, least squares (including method of moments) based on exact model characteristics. It is, however, relatively easy – though still computationally demanding and afflicted with substantial random error – to approximately compute expected values of empirical characteristics through Monte Carlo simulations by doing a random simulation of the process and computing characteristics from the simulation result. Such empirical characteristics can be any type of descriptive statistic, for instance, for a random closed set, the densities of the Minkowski functionals like volume fraction or specific surface, the covariance and certain contact distribution functions (Schneider and Weil, 2008; Ohser and Schladitz, 2009; Chiu *et al.*, 2013), or, for a random point pattern, the intensity

(function), the pair correlation function and the nearest neighbour distance distribution function (Møller and Waagepetersen, 2003; Illian *et al.*, 2008), or any other sort of summary statistic maybe tailored specifically for the given estimation problem. In particular, this includes those cases where a spatial random structure is practically observable only via planar sections or projections and, hence, only statistics based on sections or projections are available.

If we could compute these characteristics precisely and fast enough, a general approach for estimation would be to find those model parameters producing characteristics most similar to the observed characteristics, following a least squares, minimum contrast or estimation equation approach. We would formulate the problem as a minimization problem and solve it using a standard optimization code. This is exactly the way how model parameters are commonly estimated in such a situation (see, *e.g.*, Redenbach, 2009).

However, on the one hand optimization codes expect their objective functions to be evaluated precisely. On the other hand, both the Monte Carlo simulations of the model as well as the computation of the empirical characteristics are often very time-consuming. In particular, this applies for

stochastic geometric models of complex random structures such as fiber systems (Altendorf and Jeulin, 2011; Gaiselmann *et al.*, 2013) or foam structures (Lautensack, 2008; Redenbach, 2009).

That leaves us with the task of solving a global non-convex optimization or adjustment problem where the objective function can be evaluated only slowly and with a substantial random error. The aim of this paper is to give a fast algorithm for solving this sort of problems. Although it is in principle possible to apply the ideas of this algorithm also to a least squares approach, we will concentrate on a certain estimation equation approach coming from the *quasi-likelihood (QL) theory* (Godambe and Heyde, 1987; Heyde, 1997) which proposes to use the root of the so-called *quasi-score function* as the parameter estimate.

The paper is organized as follows. At first, the necessary ideas from the QL theory are introduced. Then, in a step-by-step manner, a detailed description of our ideas to solve the problem of finding a root of the quasi-score function in practice is given. Then, for the purpose of comparison, we apply our method to the example of a planar Boolean model with discs where also other ways of parameter estimation are available. Finally, we end up with some discussion.

## QUASI-LIKELIHOOD METHOD

Let  $X$  be a random variable on the sample space  $\mathcal{X}$  whose distribution depends on an unknown parameter  $\theta$  taking values in an open subset  $\Theta$  of the  $q$ -dimensional Euclidean space  $\mathbb{R}^q$ . The possible probability measures  $\{\mathbb{P}_\theta\}$  for  $X$  form a parametric family indexed by  $\theta$ . We define a mapping  $Y = T(X) \in \mathbb{R}^p$  denoting a transformation of  $X$  to a set of possible summary statistics. The objective is the optimal estimation of the parameter  $\theta$  in the sense of (asymptotic) minimum variance unbiased estimation based on the specification of the first two moments of  $Y = T(X)$  without imposing any distributional assumptions in the situation when likelihood functions are unavailable. Let  $Z(\theta) = \mathbb{E}_\theta[Y] \in \mathbb{R}^p$  be a function of the parameter to the expected value of the summary statistics and denote the variance of  $Y$  as a function  $V(\theta) = \text{Var}_\theta(Y) \in \mathbb{R}^{p \times p}$  of the parameter under  $\mathbb{P}_\theta$ . An unbiased *estimating function* for  $\theta$  is defined as a function of the data  $Y = y$  and the parameter of interest, such that  $\mathbb{E}_\theta[G(\theta, Y)] = 0$  for each  $\mathbb{P}_\theta$ . For the class  $\mathcal{G} = \{A(\theta)(y - Z(\theta))\}$  of all linear unbiased estimating functions, where  $A(\theta)$  is any non-singular matrix,  $G(\theta, y) = y - Z(\theta)$  is an optimal estimating

function for which the “information criterion”

$$\mathcal{E}(G) = \left( \mathbb{E}_\theta \left[ \frac{\partial G}{\partial \theta} \right] \right)' \left( \mathbb{E}_\theta [GG'] \right)^{-1} \left( \mathbb{E}_\theta \left[ \frac{\partial G}{\partial \theta} \right] \right) \quad (1)$$

is maximized in the partial order of non-negative definite matrices (Heyde, 1997). Then the QL theory states that

$$Q(\theta, y) = (Z'(\theta))' V(\theta)^{-1} (y - Z(\theta)), \quad (2)$$

where  $Z'$  is the Jacobian of  $Z$ .  $Q$  is called the *quasi-score function*, which is an optimal estimating function among all functions of  $\mathcal{G}$ . The “information criterion” (Eq. 1) can be seen as a generalization of the well-known Fisher information and coincides with it in case a likelihood is available and  $G$  is the usual score function, *i.e.*, the derivative of the log-likelihood w.r.t. the parameter. Thus, in analogy to ML estimation,  $\mathcal{E}(G)^{-1}$  has a direct interpretation as the asymptotic variance of the estimator  $\hat{\theta}$ . Moreover,  $\mathcal{E}(G)$  might serve as a measure of how much the characteristics  $Y = T(X)$  contribute to the precision of the estimator derived from the quasi-score function. Under rather minor regularity conditions, the estimator  $\hat{\theta}$  obtained from solving the estimating equation  $Q(\hat{\theta}, y) = 0$  has minimum asymptotic variance among all functions  $G \in \mathcal{G}$  whereas consistency is mainly established due to the unbiasedness assumption of the estimating equation (Liang and Zeger, 1995) which yields, in terms of its root, a consistent estimator  $\hat{\theta}$  even if the covariance structure  $V(\theta)$  is incorrectly specified. Note that this is in general not the case if we applied weighted least squares with  $\theta$  dependent weighting to our setting.

In what follows we always assume that the family of models  $\{\mathbb{P}_\theta\}$  is rich enough to find a root of the quasi-score equation. This could fail like for instance in classical ML when the observed characteristics are outside the range of the expected model characteristics or if the mapping from the parameter space to the expected characteristics is not continuous.

It is worth noting that *quasi-likelihood* is a concept differing from *pseudo-likelihood*, which assumes the existence of a likelihood and estimates parameters by maximizing an approximation of the full likelihood rather than the full likelihood itself, *e.g.*, because the latter is difficult to evaluate in case of complex dependencies. The most important variants are *composite marginal likelihoods* (see Varin (2008) for a comprehensive overview) first of all *pairwise likelihoods*, which have been used, *e.g.*, for parameter estimation in random set models (Nott and Rydén, 1999; Wilson and Nott, 2001).

## QUASI-LIKELIHOOD-BASED ESTIMATION APPROACH

### SIMULATION AND INTERPOLATION OF THE QUASI-SCORE

Since closed form expressions of expectations are generally unavailable for complex stochastic models we assume there is no other way of inferring  $Z(\theta)$  or  $V(\theta)$  than over extensive simulations of  $X$ . We assume that we can simulate realizations from the distributions  $\mathbb{P}_\theta$  for any  $\theta \in \Theta$ . An iterative algorithm solving the estimating equations based on the quasi-score function has thus to conduct simulations of  $X$  for various  $\theta$  inferring  $Z(\theta)$  and  $V(\theta)$  from the simulation output. Depending on the complexity of the simulation models this can be very time consuming and typical solvers will be confused by the remaining random fluctuation. We solve the problem by estimating an approximating model for the dependency of the quasi-score function on the parameters from individual simulations and then solve the quasi-score equation on an iteratively improved version of this approximation.

We propose the use of an interpolation-based meta-model of  $Z(\theta)$  and later  $V(\theta)$  for the construction of an approximating quasi-score function, since unlike the quasi-score function itself, these can be estimated unbiasedly from simulations. We use *kriging* (Cressie, 1993; Chiles and Delfiner, 1999) for this interpolation since it allows automatic adaption of the prediction rule to the properties of the unknown function to be interpolated and provides an indication of precision of the interpolation. Kriging is typically defined in a probabilistic context but it can also be understood as a method to approximate or interpolate data such as splines do (Wahba, 1990). It is primarily based on estimated covariance functions which quantify the change rates of the interpolated functions and their derivatives.

Besides its usage in truly geostatistical problems, kriging has been also widely applied in the context of *simulation optimization* (Kleijnen and van Beers, 2009) and independently studied within the optimization literature of *Design and Analysis of Computer Experiments* (DACE) (Sacks *et al.*, 1989) for expensive to evaluate, scalar valued objective functions. Unlike these approaches which use stationary random functions we propose the use of intrinsic random functions because quasi-scores typically diverge.

#### Prediction model

Let  $Z(\theta)$  denote one of the functions to be interpolated, *e.g.*,  $Z(\theta) = \mathbb{E}_\theta[Y]$ . The problem

considered now is the prediction of the value  $Z(\theta_0)$  for some new sampling point  $\theta_0$  in the parameter space given a finite set  $S_n = \{(\theta_i, z_i), \theta_i \in \mathbb{R}^q, i = 1, \dots, n\}$  of pairs of a *sampling point*  $\theta_i$  and an “observed”  $z_i \approx Z(\theta_i)$ . The observations  $z_i = Z(\theta_i) + \varepsilon_i$  are affected by the variance from the Monte Carlo simulations  $\varepsilon_i$ . The  $\varepsilon_i$  are assumed to be stochastically independent of each other. Its variance is estimated in the sampling procedure. In our setting, given  $S_n$ , we model the functional dependence between the parameter  $\theta$  and the value  $Z(\theta)$  by

$$Z(\theta) = m_\beta(\theta) + W(\theta), \quad (3)$$

where  $m_\beta(\theta) = \sum_{l=0}^r \beta_l f^l(\theta)$  denotes a linear model for the expected value of the “random” function  $Z$ . The functions  $f^l$  are known basis functions, typically monomials for a polynomial approximation of the mean of  $Z$ . The index  $l$  has the meaning of a power, *e.g.*, in 1D,  $f^l(\theta) = \theta^l$ . The case  $l = 0$  is included for the constant mean, *i.e.*,  $f^0 = 1$ . In the  $q$ -dimensional space there are  $k_n = \binom{q+k}{k}$  monomials of degree  $\leq k$  so that the term  $m_\beta$  consists of  $k_n + 1$  functions  $f^l$ . The unknown coefficients  $\beta_l$  have to be estimated from the data.  $W$  is a zero-mean *intrinsic random function* (see next section). Though the form of  $W$  is in general unknown the covariance structure of  $W$  is subjectively chosen so that it relates to the data phenomenon and its smoothness properties as usually done in geostatistics.

#### Intrinsic kriging

Intrinsic Kriging extends the scope of Kriging to the case of intrinsic random functions (IRF) (Matheron, 1973) with unknown mean whose increments are second-order stationary. The main idea is to filter out the mean so to consider a zero-mean process again. Here, we shortly recall the basic notion of IRF. For a set of weights  $\lambda_i \in \mathbb{R}$  and points  $\theta_i \in \Theta \subset \mathbb{R}^q, i = 1, \dots, n$ , we define a discrete measure  $\lambda$ , say  $\lambda = \sum_{i=1}^n \lambda_i \delta_{\theta_i}$ . Let  $\mathcal{F}$  be the finite-dimensional vector space generated by  $f^l$ . For a function  $f \in \mathcal{F}$  we define the linear application of  $\lambda$  on the function  $f$  by  $f(\lambda) = \sum_{i=1}^n \lambda_i f(\theta_i)$ . For  $k \geq 0$ ,  $\Lambda_k$  denotes the set of all measures  $\lambda$  for which  $f(\lambda) = 0$ , that is,  $\lambda$  annihilates separately all monomials of degree  $\leq k$ , *i.e.*,  $f^l(\theta) = \theta_1^{l_1} \dots \theta_q^{l_q}, \theta \in \Theta$ , with the vector-valued index  $l = (l_1, \dots, l_q), l_j = 0, 1, \dots, k$  and  $|l| = \sum_{j=1}^q l_j \leq k$ . Such measure  $\lambda \in \Lambda_k$  is called an *allowable measure*, see, *e.g.*, Chiles and Delfiner (1999). For any translation vector  $h \in \mathbb{R}^q$  and  $\lambda$ , we define the translate of  $\lambda$  by  $\tau_h \lambda = \sum_{i=1}^n \lambda_i \delta_{\theta_i+h}$  with the same weights as  $\lambda$ . A second-order random function  $Z(\theta)$  is said to be an *intrinsic random function of order  $k$*  (IRF- $k$ ), if for any allowable measure  $\lambda \in \Lambda_k$  the random function  $Z(\tau_h \lambda)$  is second-order stationary. Since the mean of  $Z(\theta)$  is in  $\mathcal{F}$ ,  $Z(\lambda), \lambda \in \Lambda_k$ , is a zero-mean random function

with covariance  $\text{Cov}(Z(\tau_{h_1}\lambda), Z(\tau_{h_2}\lambda))$  as a function of  $h_1 - h_2$  only. From Matheron (1973) we know that any continuous IRF- $k$  has a continuous *generalized covariance function*  $K(h)$  such that for any pair of measures  $\lambda = \sum_{i=1}^{n_1} \lambda_i \delta_{\theta_i} \in \Lambda_k$ ,  $\mu = \sum_{j=1}^{n_2} \mu_j \delta_{\theta_j^*} \in \Lambda_k$  and sequences of points  $\theta = (\theta_i)$ ,  $\theta^* = (\theta_j^*)$ ,

$$\text{Cov}(Z(\lambda), Z(\mu)) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \lambda_i \mu_j K(\theta_i - \theta_j^*),$$

where  $K$  is unique up to an even polynomial of degree  $2k$  for  $k \geq 0$ . For further details the reader is referred to the general geostatistical literature (Cressie, 1993; Chiles and Delfiner, 1999; Wackernagel, 2003). In our examples we use the generalized covariance functions proposed by Mardia (1996). We need this advanced covariance functions to adequately describe the typical divergence properties of the moments as functions of parameters.

### Kriging the statistics

Let the sample mean at some evaluation location  $\theta$  be

$$\bar{Y}(\theta) = \frac{1}{N} \sum_{j=1}^N Y(\theta, j),$$

where  $Y(\theta, j) \in \mathbb{R}$  denotes the  $j$ th simulation,  $j = 1, \dots, N$ , of some real-valued characteristic. The value  $\bar{Y}$  is regarded as a realization of an IRF- $k$  of known order  $k$  and known generalized covariance function  $K(h)$ . We are concerned with the estimation of the value  $\bar{Y}(\theta_0)$  for some evaluation location  $\theta_0$ . The *kriging predictor* (best linear unbiased estimator) is given by

$$\hat{Y}(\theta_0) = \sum_{i=1}^n \hat{\lambda}_i(\theta_0) \bar{Y}(\theta_i), \quad (4)$$

where the weights  $\hat{\lambda}_i$ , depending on the prediction location, are the solutions to the so-called *intrinsic kriging equations* (Chiles and Delfiner, 1999, sect. 4.6.). The *prediction variance* of this estimator is then given by

$$\hat{\sigma}^2(\theta_0) = K(0) - \sum_{i=1}^n \hat{\lambda}_i K(\theta_0 - \theta_i) - \sum_{|l|=0}^k \mu_l \theta_0^l, \quad (5)$$

where  $\mu$  is the vector of Lagrange multipliers, introduced in order to ensure the unbiasedness of the kriging predictor.

### Modelling the variance of statistics

The quasi-score function (Eq. 2) involves the variance-covariance matrix  $V(\theta)$  of the data  $Y = y$  which in general depends on the unknown parameter.

Since we only assume that we can infer  $V(\theta)$  from simulations the idea is to construct an interpolated version of the sample variance-covariance matrix based on a Cholesky decomposition.

Let the sample covariance matrix be  $\bar{V}(\theta) = [v]_{lk}$ , where, for  $l, k = 1, \dots, p$ ,

$$v_{lk} = \frac{1}{N-1} \sum_{j=1}^N [Y_l(\theta, j) - \bar{Y}_l(\theta)] [Y_k(\theta, j) - \bar{Y}_k(\theta)].$$

The Cholesky decomposition of  $\bar{V}$  is of the form  $\bar{V} = LL^t$  where  $L$  is a unique lower triangular matrix with real and positive diagonal entries since  $\bar{V}$  is (assumed to be) positive definite. Suppose we have previously computed a series of such matrices  $\bar{V}(\theta_1), \dots, \bar{V}(\theta_n)$  for various sample points. In order to apply kriging we rewrite each Cholesky decomposition  $\bar{V}(\theta_k) = L^{(k)}(L^{(k)})^t$  as a row vector which allows us to treat the entries  $L_{ij}^{(k)}$ ,  $1 \leq j \leq i \leq p$ ,  $k = 1, \dots, n$ , as the data to essentially the same type of kriging model as noted in (3). We define the following data matrix

$$\begin{bmatrix} L_{11}^{(1)} & L_{21}^{(1)} & \dots & L_{pp}^{(1)} \\ \vdots & \vdots & & \vdots \\ L_{11}^{(n)} & L_{21}^{(n)} & \dots & L_{pp}^{(n)} \end{bmatrix} \in \mathbb{R}^{n \times m}, \quad (6)$$

where  $m = p(p+1)/2$ . Now, each column represents the kriging data for a single kriging interpolation model which amounts in overall  $m$  kriging models. For some  $\theta_0$  the kriging prediction for each model results in the kriging estimates  $\hat{L}^{(0)} = (\hat{L}_{11}^{(0)}, \dots, \hat{L}_{pp}^{(0)})^t$ . Finally, as a unique kriging approximation to  $\bar{V}(\theta_0)$ , we obtain

$$\hat{V}(\theta_0) = \hat{L}^{(0)}(\hat{L}^{(0)})^t. \quad (7)$$

As done before in the case of kriging the statistics, one could also incorporate a measurement error of the sample variances. However, as this quantity as a fourth order moment is difficult to estimate, we use a global nugget effect in the covariance function model to capture it.

### Modeling the covariance

We use the restricted maximum likelihood method (REML) (Mardia and Marshall, 1984; Zimmermann, 1989) to estimate the vector  $\phi \in \mathbb{R}^r$  of covariance parameters. The variable under study is still  $Z(\theta)$  which is treated as non-stationary in its mean  $\mathbb{E}[Z(\theta)] = m_\beta(\theta)$  and has a generalized covariance model  $K(\cdot; \phi)$ . Clearly the response  $\bar{Y}(\theta)$  cannot be observed without noise. The variance of the measurement noise  $\varepsilon(\theta)$ , which is assumed Gaussian with zero mean, is estimated by

$$\hat{\sigma}_\varepsilon^2(\theta) = \frac{1}{N(N-1)} \sum_{j=1}^N (Y(\theta, j) - \bar{Y}(\theta))^2.$$

In the application of kriging in geostatistics the variance of the measurement error is due to the so-called “nugget effect”. To account for the noise due to simulation replications we take

$$\hat{\Sigma}_\varepsilon(\theta) = \text{Diag}\{\hat{\sigma}_\varepsilon^2(\theta_1), \dots, \hat{\sigma}_\varepsilon^2(\theta_n)\}$$

as the diagonal matrix of estimated variances. Then the REML estimation is based on the overall covariance matrix

$$\tilde{K}(\cdot; \phi) = K(\cdot; \phi) + \hat{\Sigma}_\varepsilon(\cdot), \quad (8)$$

which yields an estimate  $\hat{\phi}$  of the covariance parameter. The kriging estimator and its prediction variance are further derived from Eq. 4 and Eq. 5 with  $K$  replaced by  $\tilde{K}(\cdot; \hat{\phi})$ . Other approaches would be possible, but we found this approach to be reasonable stable and effective.

### Quasi-score approximation

We are now in the position to construct the approximated quasi-score based on the interpolated quantities which we derived in the previous sections.

Suppose that for a given set  $S_n$  of sampled points and corresponding values we have constructed the kriging predictors  $\hat{Z} = (\hat{Y}_1, \dots, \hat{Y}_p)^t$  for each statistic together with the approximating variance-covariance matrix  $\hat{V}$ . Note that all predictors depend on the given sampled data set  $S_n$ . Let

$$\hat{\Sigma}_K(\theta) = \text{Diag}\{\hat{\sigma}_1^2(\theta), \dots, \hat{\sigma}_p^2(\theta)\} \quad (9)$$

denote the diagonal matrix of kriging variances (Eq. 5) for each statistic. Based on the optimal estimating function (Eq. 2) we construct the approximated quasi-score function  $\hat{Q}(\theta, y) \in \mathbb{R}^q$  by replacing the involved expectations and variances with the corresponding kriging approximations

$$\hat{Q}(\theta, y) = (\hat{Z}'(\theta))^t (\hat{V}(\theta) + \hat{\Sigma}_K)^{-1} (y - \hat{Z}(\theta)), \quad (10)$$

such that the approximated “information criterion” becomes

$$\hat{I}(\theta) = (\hat{Z}'(\theta))^t (\hat{V}(\theta) + \hat{\Sigma}_K)^{-1} (\hat{Z}'(\theta)). \quad (11)$$

The term  $\hat{\Sigma}_K$  explicitly accounts for the kriging prediction error of the variances of the statistics. This modification of the variances leads to a systematic reduction of the prediction error induced by the kriging model when new solutions are added to the set  $S_n$  of all previously sampled points and values. For a consistent estimation recall that we do not need to correctly specify the (unknown) covariance structure  $V$ .

The Jacobian of  $\hat{Z}$ , *i.e.*, the matrix of partial derivatives in Eq. 10, is numerically calculated based on finite differences of the kriging-interpolated statistics though one could also obtain estimates of  $\hat{Z}'(\theta)$  by a cokriging approach (Chiles and Delfiner, 1999, sect. 5.5.1).

## SOLVING THE QUASI-SCORE EQUATION

In analogy to the ML method one can regard the quasi-score function as some kind of a gradient specification of a “quantity” similar to a likelihood which, however, is not considered in the QL theory. Likewise the “information criterion” (Eq. 1) is a generalization of the well-known Fisher information defined in the ML theory. Therefore, the same Newton-Raphson-type iteration used to find the maximum likelihood estimate as a root of the score function, which is usually called *Fisher scoring*, can be applied to the QL setting (Osborne, 1992). This leads to the *Fisher quasi-scoring* iteration

$$\theta_{k+1} = \theta_k + t_k d_k, \quad d_k = -\hat{I}(\theta_k)^{-1} \hat{Q}(\theta_k). \quad (12)$$

The step length parameter  $t_k$  in Eq. 12 has to be chosen so that the iteration does not become unstable. Therefore  $t_k$  is chosen as a minimizer of the auxiliary optimization problem

$$\min_{t \in (0,1)} \|\hat{Q}(\theta_k + t d_k)\|_2^2, \quad (13)$$

(see Osborne, 1992, sect. 4 for details). Particularly Osborne (1992) shows that the quasi-scoring iteration converges a.s. to the QL estimator provided that the effective sample size is large enough and the chosen starting point  $\theta_0$  is sufficiently close to a root of Eq. 10. The aspect of the necessary proximity of the starting point is dealt with by a simple restart. If the iteration fails a direct search method (Spall, 2003) is used to find a better starting value. This is possible since the kriging interpolations are relatively cheap to evaluate. The increasing sample size corresponds to an increasing window size in a spatial setting. Hence, ergodicity or mixing properties are required. Note that the iteration is based on the approximated quasi-score function (Eq. 10) and therefore will find a root of Eq. 10 rather than of the theoretical quasi-score function (Eq. 2).

## UPDATING THE QUASI-SCORE INTERPOLATION

Since the overall aim is to sequentially improve the QL estimate based on Eq. 10 towards the solution of Eq. 2, the approximate quasi-score is updated as follows: To the current set  $S_n$  of all previous sample points and values the root  $\hat{\theta}$  of the quasi-score equation (Eq. 10) as well as further points sampled from  $\mathcal{N}(\hat{\theta}, \hat{I}(\hat{\theta})^{-1})$  are added. This involves

new simulations at all new locations followed by refining the approximation  $\hat{Q}$  through an update of the kriging-interpolated statistics and variance-covariance matrix (Eq. 7) along the lines of the previous subsections. The additional locations are used to improve not only the estimate of the value but also the estimate of the Jacobian  $\hat{Z}'(\theta)$  and, hence, also that of the quasi-score function. The chosen sampling distribution for the additional sample points is motivated by the idea that the estimation error is (under certain conditions ensuring asymptotic normality of the statistics  $T$ , like ergodicity and sufficient regularity) asymptotically normal with this variance (see the subsequent subsection) and thus the true parameter value could be any of these (Godambe and Heyde, 1987, sect. 4.3).

## ESTIMATION ERROR

The precision of the QL estimator is under certain conditions asymptotically equivalent to the inverse of the “information criterion” (Eq. 1). The variance of the quasi-score function  $Q(\theta, y)$  is given by Eq. 1 (Heyde, 1997) and reads

$$I(\theta) = Z'(\theta)'V(\theta)^{-1}Z'(\theta), \quad (14)$$

which we call the *quasi-information matrix*. We can use a first-order Taylor approximation of  $Q(\hat{\theta}, y)$  in  $\hat{\theta} = \theta$  and  $y = Z(\theta)$ ,

$$\begin{aligned} Q(\hat{\theta}, y) &\approx Q(\theta, Z(\theta)) + \frac{\partial}{\partial \theta} Q(\theta, Z(\theta))(\hat{\theta} - \theta) \\ &\quad + Z'(\theta)'V(\theta)^{-1}(y - Z(\theta)) \\ &= 0 - I(\theta)(\hat{\theta} - \theta) + Q(\theta, y), \end{aligned}$$

which leads to

$$\hat{\theta} \approx \theta + I(\theta)^{-1}Q(\theta, y).$$

Using the law of error propagation we finally get

$$\text{Var}(\hat{\theta}) \approx I(\theta)^{-1}I(\theta)I(\theta)^{-1} = I(\theta)^{-1} \quad (15)$$

for sufficiently large datasets such that  $y$  is near  $\mathbb{E}[Y]$  just like in ML estimation.

## STOPPING CRITERIA

The estimation algorithm has a twofold iteration. The inner loop searches for a root of the approximated quasi-score by means of the stabilized Fisher quasi-scoring iteration. This inner iteration terminates when the “Mahalanobis distance” of the quasi-score (Heyde, 1997, sect. 4.4) is numerically negligible, *i.e.*,

$$C = Q'I^{-1}Q \leq \delta \ll 1, \quad (16)$$

*e.g.*,  $\delta = 10^{-7}$  in our simulation study. If the algorithm converges numerically (*i.e.*, the step size in  $\theta$  drops below some limit which is small relative to the extent of the parameter space, in our simulation study set to  $10^{-6}$ ) we distinguish two cases. If  $C$  is significantly distinguishable from 0 according to its approximate  $\chi_q^2$  distribution, the Fisher quasi-scoring iteration is restarted from multiple (*e.g.*, Eq. 15 in our simulation study) random locations in the parameter space following the reasoning that this is not an approximate root. The root with the smallest sum of eigenvalues of the inverse quasi-information matrix (Eq. 14) is selected according to the idea that if really multiple substantially different parameters occur which would explain the data we should select one with a small local estimation error. Obviously this idea can be improved towards some more sophisticated method to select the best root according to additional evidence in the data, *e.g.*, based on a difference quotient quasi-score

$$\left( \frac{Z_j(\theta_2) - Z_j(\theta_1)}{\theta_{2i} - \theta_{1i}} \right)_{ij} \left( \frac{V(\theta_2) + V(\theta_1)}{2} \right)^{-1} (Y - Z(\theta)),$$

for evaluation locations  $\theta_1, \theta_2 \in \Theta$ . If it could be a root statistically according to Eq. 16, the inner iteration is interrupted to improve the approximated quasi-score locally by new simulations in this area of the parameter space.

The stopping rule for the outer iteration, which samples further locations, is based on the comparison of the interpolation error of the quasi-score approximation to the quasi-information matrix (Eq. 14). This interpolation error is measured by the kriging error of  $Z(\theta)$  and transformed into an error of the quasi-score by the law of error propagation assuming  $Z'(\theta)V(\theta)^{-1}$  and the expected statistical error of the quasi-score to be known. The approximation error of the quasi-score is calculated by

$$B(\theta) = Z'(\theta)V(\theta)^{-1}\hat{\Sigma}_K V(\theta)^{-1}Z'(\theta)'$$

The outer iteration is stopped as soon as the first is negligible compared to the second in all directions, *i.e.*, if the largest generalized eigenvalue (Golub and Loan, 1996) of the two matrices drops below some limit  $b \ll 1$  (*e.g.*,  $b = 10^{-4}$  in our simulation study),  $B \leq_L b \cdot I(\theta)$ , in the Loewner half order of non-negative definite matrices.

## SUMMARY OF QUASI-SCORE ESTIMATION

This section is intended to give a short outline of the intermediate steps during the overall estimation procedure based on the approximated quasi-score.

The parameter estimation method requires the definition of a stochastic simulation model which – to some extent – describes the data-generating process. For such model, we assume that a fixed set of chosen statistics can be efficiently simulated for various values of the model parameters. For the initialization of the method one has to select a start sample of parameter values  $S_{n_0}$  of sufficiently large size  $n_0$  where to simulate the model and to compute the values of these statistics needed to construct the initial kriging interpolation models. It is not important to select  $S_{n_0}$  according to a specific design methodology as long as it “reasonably” covers the parameter space such as space filling designs do, e.g., a *Latin hypercube design* (Santner *et al.*, 2003, sect. 5). One could also account for some “user” knowledge of most likely regions of possible roots of the approximated quasi-score and generate a sample of initial model parameters from a pre-specified a-priori distribution.

The construction and calculation of the approximate quasi-score involves the following steps: Suppose that in the  $k$ th iteration of the algorithm a root  $\hat{\theta}_k$  of Eq. 10 was found. Given a set of sampling locations  $S_{n_k}$  together with the corresponding simulation results the mean values and variances of the statistics are computed and the kriging interpolation models are individually updated. This step involves the REML estimation of the covariance parameters based on the modified covariance matrix (Eq. 8) once for each kriging predictor of the statistics and all entries of the Cholesky decomposition of the variance-covariance matrix of the statistics. The kriging interpolation of the Cholesky decomposition of the variance-covariance matrix of the statistics is based on the same kriging model (Eq. 3) as used for the interpolation of the statistics but with separately fitted parameters. Besides the calculation of the Jacobian of the statistics, as a final step to calculate the approximation to the quasi-score, the kriging prediction variances (Eq. 9) of the statistics are added to the kriging interpolation  $\hat{V}$  in Eq. 10. Note that this only requires the calculation of the kriging prediction variances by Eq. 5 once for each statistic since at this point the kriging equations have already been solved for the prediction of the statistics at some specific parameter value.

Based on the current estimated root  $\hat{\theta}_k$  and kriging interpolation models a stabilized version of

the Fisher quasi-scoring iteration is applied to solve Eq. 10 including the step-length control problem (Eq. 13) combined with a multi-start strategy to ensure convergence. This leads to an improved estimate  $\hat{\theta}_{k+1}$  followed by the above described update procedure of the quasi-score interpolation involving further simulations.

## AN ILLUSTRATIVE EXAMPLE

By way of example we shall demonstrate the power of the proposed general parameter estimation method in what follows.

### 2D BOOLEAN MODEL WITH DISCS

Let  $X$  be a stationary Boolean model in  $\mathbb{R}^2$  with intensity  $\lambda$  and with “typical” grain equal to a disc of some fixed radius  $R$ , that is,  $X$  is a random closed set of the form  $X = \bigcup_i B(x_i, R)$  where the centres  $x_i$  of the discs  $B(x_i, R)$  of radius  $R$  are the points of a homogeneous Poisson point process in  $\mathbb{R}^2$  with intensity  $\lambda$  (Schneider and Weil, 2008).

Since Boolean models in general have been paid attention in many respects there exist many results on the statistics of Boolean models in the literature, in particular several special methods of parameter estimation in Boolean models have been established and investigated (Heinrich, 1993; Molchanov, 1997; Chiu *et al.*, 2013).

A certain method, which is known as the *method of intensities* (Molchanov, 1997) or *method of densities* (Chiu *et al.*, 2013), relies on the fact that the area fraction  $A_A$ , the specific boundary length  $L_A$  and the specific Euler number  $\chi_A$  of  $X$  as the intensities/densities of the three Minkowski functionals in  $\mathbb{R}^2$  (Schneider and Weil, 2008) are given as closed form expressions of, in the present example,  $\lambda$  and  $R$ :

$$A_A = 1 - \exp(-\lambda \pi R^2), \quad (17)$$

$$L_A = 2\lambda \pi R \exp(-\lambda \pi R^2), \quad (18)$$

$$\chi_A = (\lambda - \lambda^2 \pi R^2) \exp(-\lambda \pi R^2). \quad (19)$$

Solving for  $\lambda$  and  $R$  leads to

$$\lambda = \frac{\chi_A}{1 - A_A} + \frac{1}{4\pi} \frac{L_A^2}{(1 - A_A)^2}, \quad (20)$$

$$R = \frac{2L_A(1 - A_A)}{4\pi(1 - A_A)\chi_A + L_A^2}, \quad (21)$$

cf. Molchanov (1997, p. 82), or, using only  $A_A$  and  $L_A$ , to

$$\lambda = \frac{L_A^2}{-4\pi(1 - A_A)^2 \cdot \ln(1 - A_A)}, \quad (22)$$

$$R = \frac{-2(1-A_A) \cdot \ln(1-A_A)}{L_A}. \quad (23)$$

Replacing the model characteristics  $A_A$ ,  $L_A$  and  $\chi_A$  by certain corresponding empirical counterparts will lead to estimators of  $\lambda$  and  $R$ .

In the following we always assume that we observe our model  $X$  in some rectangular window  $W$ . However, we distinguish the following two cases. On the one hand, we use the fact that  $X$  restricted to  $W$  is essentially a finite union of discs from which the area  $A_X(W)$ , the boundary length  $L_X(W)$  and the Euler number  $\chi_X(W)$  of  $X$  within  $W$  can be computed exactly (Tscheschel, 2005). This results in the (unbiased) estimators

$$\widehat{A}_A = \frac{A_X(W)}{A(W)}, \quad \widehat{L}_A = \frac{L_X(W)}{A(W)}, \quad \widehat{\chi}_A = \frac{\chi_X(W)}{A(W)}, \quad (24)$$

where  $A(W)$  is the area of  $W$ . Inserting these estimators into Eqs. 20 and 21 gives the estimator  $(\widehat{\lambda}_{M1a}, \widehat{R}_{M1a})$  for  $(\lambda, R)$ , inserting instead into Eqs. 22 and 23 gives  $(\widehat{\lambda}_{M1b}, \widehat{R}_{M1b})$ .

On the other hand, we consider the case that  $X$  restricted to  $W$  is only observable in a discretized form, *i.e.*, on a rectangular lattice of some lattice constant  $a$ . In this case we use estimators  $\widetilde{A}_A$ ,  $\widetilde{L}_A$ , and  $\widetilde{\chi}_A$  for  $A_A$ ,  $L_A$  and  $\chi_A$ , respectively, along the lines given in Ohser and Mücklich (2000, sect. 4.2, filter mask  $F_1$ ), finally resulting in  $(\widetilde{\lambda}_{M1a}, \widetilde{R}_{M1a})$  when inserting in Eqs. 20 and 21, and  $(\widetilde{\lambda}_{M1b}, \widetilde{R}_{M1b})$  when inserting in Eqs. 22 and 23.

Subsequently, in the application of quasi-score estimation, either the triple  $(\widehat{A}_A, \widehat{L}_A, \widehat{\chi}_A)$  or the triple  $(\widetilde{A}_A, \widetilde{L}_A, \widetilde{\chi}_A)$  will play the role of the set of statistics  $Y = T(X)$ .

## EXACT QUASI-SCORE ESTIMATION

Due to unbiasedness the expected values of  $(\widehat{A}_A, \widehat{L}_A, \widehat{\chi}_A)$  are  $Z(\lambda, R) = (A_A, L_A, \chi_A)$ , explicitly given by Eqs. 17–19. This in turn leads to

$$\frac{\partial Z(\lambda, R)}{\partial(\lambda, R)} = \exp(-\lambda \pi R^2) \times \begin{pmatrix} \pi R^2 & 2\lambda \pi R \\ 2\pi R(1 - \lambda \pi R^2) & 2\pi \lambda(1 - 2\lambda \pi R^2) \\ 1 - 3\lambda \pi R^2 + \lambda^2 \pi^2 R^4 & 2\lambda^2 \pi R(\lambda \pi R^2 - 2) \end{pmatrix}.$$

Furthermore, also the matrix  $V(\lambda, R)$  of variances and covariances of  $(\widehat{A}_A, \widehat{L}_A, \widehat{\chi}_A)$  can be calculated, at least numerically.

The simplest case is the variance  $V_{AA} = \mathbb{E}[A_X(W)^2/A(W)^2] - A_A^2$  of  $\widehat{A}_A$ . We have

$$V_{AA} = \frac{1}{A(W)^2} \int C_{AA}(h) \gamma_W(h) dh - A_A^2,$$

where  $C_{AA}(h) = 2A_A - 1 + (1 - A_A)^2 \exp(\lambda \gamma_{B(o,R)}(h))$  is the covariance of  $X$  and  $\gamma_W(h) = A(W \cap (W - h))$  is the set covariance of  $W$  (Chiu *et al.*, 2013).

$V_{AA}$  is directly connected to the second-order moment measure  $M_{AA}$  of the random measure  $A_X$ , *i.e.*,  $M_{AA}(W_1, W_2) := \mathbb{E}[A_X(W_1)A_X(W_2)]$ , and  $V_{AA} = M_{AA}(W, W)/A(W)^2 - A_A^2$ . Note that  $C_{AA}$  is a density of  $M_{AA}$  with respect to the 4-dimensional Lebesgue measure. In order to determine the variances and covariances  $V_{ij}$ ,  $i, j \in \{A, L, \chi\}$ , it is thus necessary to know the (mixed) second-order moment measures  $M_{ij}$ ,  $i, j \in \{A, L, \chi\}$ , and, if existing, their densities  $C_{ij}$  with respect to Lebesgue measure. For exactly the model under consideration these densities have been derived in the literature (Mecke, 2001; Torquato, 2002; Ballani, 2007). For instance, we have

$$V_{LL} = \frac{1}{A(W)^2} \int C_{LL}(h) \gamma_W(h) dh - L_A^2,$$

$$C_{LL}(h) = \left( \lambda^2 \left[ 2\pi R - \mathbf{1}_{[0,2R)}(\|h\|) 2R \arccos\left(\frac{\|h\|}{2R}\right) \right]^2 + \lambda \mathbf{1}_{[0,2R)}(\|h\|) \frac{4R^2}{r\sqrt{4R^2 - r^2}} \right) \bar{C}_{AA}(h),$$

and

$$V_{AL} = V_{LA} = \frac{1}{A(W)^2} \int C_{AL}(h) \gamma_W(h) dh - A_A L_A,$$

$$C_{AL}(h) = L_A - \lambda \bar{C}_{AA}(h) \left[ 2\pi R - \mathbf{1}_{[0,2R)}(\|h\|) 2R \arccos\left(\frac{\|h\|}{2R}\right) \right],$$

where  $\bar{C}_{AA}(h) = (1 - A_A)^2 \exp(\lambda \gamma_{B(o,R)}(h))$ . It turns out that all measures  $M_{ij}$  but  $M_{\chi\chi}$  have a density with respect to Lebesgue measure. The “density” of  $M_{\chi\chi}$  contains a certain delta component (Mecke, 2001) which can be treated separately without any problem.

All in all, for the present example of a Boolean model the quasi-score function  $Q$  is known exactly in the sense that it can be evaluated numerically. Hence, here we are in the position to find the estimates  $\widehat{\lambda}_{EQS}$  and  $\widehat{R}_{EQS}$  as the root of the exact quasi-score (EQS) equation  $Q((\lambda, R), y) = 0$  by the type of iteration shown in Eq. 12.



### SIMULATED QUASI-SCORE ESTIMATION

For the simulation-based quasi-score (SQS) estimation we consider the two cases that both the observation  $y$  and each of the realizations of the statistics of  $\bar{Y}(\lambda, R)$  of the Boolean model used during the estimation procedure is either given by the exact estimators  $(\widehat{A}_A, \widehat{L}_A, \widehat{\chi}_A)$  or by the estimators  $(\widetilde{A}_A, \widetilde{L}_A, \widetilde{\chi}_A)$  coming from the discretization. Thus we obtain again two sets of estimators  $(\widehat{\lambda}_{SQS}, \widehat{R}_{SQS})$  and  $(\widetilde{\lambda}_{SQS}, \widetilde{R}_{SQS})$ , respectively, for  $(\lambda, R)$ .

### SIMULATION STUDY

We conducted a simulation study by applying the above described estimation methods SQS, EQS and MIa, MIb to the same set of model realizations. The first part of this study investigates the mean squared errors between re-estimated model parameters and true ones over randomly sampled locations of the parameter space. The second part addresses the estimation precision of the proposed method based on the SQS.

Throughout the whole simulation study the observation window  $W$  was chosen as  $[0, 5]^2$ . Intentionally the lattice constant  $a$  was 0.05 leading to a relatively coarse discretization of  $100 \times 100$  pixels. Further, the following process parameters of the SQS estimation method were fixed. The amount of sampling locations  $n_0$  used to construct the initial quasi-score approximation was set to 12. The number of simulation replications spent for each sample location was fixed to 100 throughout the whole estimation procedure. The parameter space for the feasible values of  $\theta = (\lambda, R)$  was set to the intervals  $[5, 30]$  and  $[0.01, 0.2]$  for  $\lambda$  and  $R$ , respectively. The maximum number of admissible new sampling locations added during the iterations of the method was restricted to 50.

Within the first part of the study  $N = 200$  parameter values of  $\lambda$  and  $R$  were sampled from a uniform distribution, each parametrized by the interval  $[8, 25]$  and  $[0.05, 0.18]$ , respectively. Simulating the model  $X$  once for each  $\theta_i = (\lambda_i, R_i)$ ,  $i = 1, \dots, N$  yielded the observations  $x_i$ . The exact statistics  $(\widehat{A}_A, \widehat{L}_A, \widehat{\chi}_A)$  and their counterparts  $(\widetilde{A}_A, \widetilde{L}_A, \widetilde{\chi}_A)$  for the discretizations were both applied to these observations in order to compare the performance of the estimation methods. Fig. 1 presents the true parameters plotted against their estimated values obtained from a single-run SQS and EQS estimation for each  $\theta_i$ . The mean number of newly added locations in case of running the SQS estimation method based on the exact statistics

(Eq. 24) amounted to 20 whereas using the discretized versions required 35 additional sampling locations on average.

The *Bayesian root mean square error* (BRMSE)

$$\text{BRMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\theta_{ik} - \hat{\theta}_{ik})^2}, \quad k = 1, 2$$

of  $N$  re-estimated parameters was separately calculated over  $\hat{\theta}_i = (\hat{\lambda}_i, \hat{R}_i)$  for each estimation method.

For 13 out of 200 single-runs of the SQS estimation method the current implementation caused the method to fail to converge towards the “correct” root. This algorithmic behaviour due to the current implementation of the SQS estimation method is induced by the non-convexity properties of the quasi-score in general, thus making it difficult to distinguish between multiple roots (Heyde, 1997, sect. 13.3).

First of all the BRMSE for the estimators based on the method of intensities, see Table 1, show a well-known behaviour, namely, that MIb outperforms MIa due to the fact that the estimators for the specific Euler number  $\chi_A$  have much higher variability than that for  $A_A$  and  $L_A$  (Chiu *et al.*, 2013). For both MIa and MIb the relatively high BRMSE of  $\widehat{\lambda}$  and  $\widehat{R}$  compared to that of  $\widetilde{\lambda}$  and  $\widetilde{R}$  is due to the fact that the estimators  $\widetilde{L}_A$  and  $\widetilde{\chi}_A$  based on the discretized realizations are not unbiased for  $L_A$  and  $\chi_A$ , respectively, see Serra (1982); Ohser and Mücklich (2000). This also leads to a bias for  $\widetilde{\lambda}$  and  $\widetilde{R}$ , which is here quite pronounced because of the very coarse applied discretization.

Table 1. *Bayesian root mean square error (BRMSE) for the estimators of  $\lambda$  and  $R$  according to the method of intensities with  $(A_A, L_A, \chi_A)$  (MIa), with  $(A_A, L_A)$  (MIb), exact quasi-score (EQS), and simulated quasi-score (SQS) estimation. Values are rounded to two significant digits. The simulation study standard errors are given in parentheses. The exact quasi-score is not applicable for the discretized data.*

Method	BRMSE( $\widehat{\lambda}$ )	BRMSE( $\widetilde{\lambda}$ )
MIa	1.8 (0.21)	9.5 (0.36)
MIb	1.3 (0.10)	7.0 (0.21)
EQS	1.1 (0.07)	–
SQS	2.5 (0.32)	2.5 (0.34)

Method	BRMSE( $\widehat{R}$ )	BRMSE( $\widetilde{R}$ )
MIa	0.011 (0.0029)	0.299 (0.0674)
MIb	0.003 (0.0003)	0.044 (0.0017)
EQS	0.003 (0.0002)	–
SQS	0.005 (0.0009)	0.007 (0.0011)

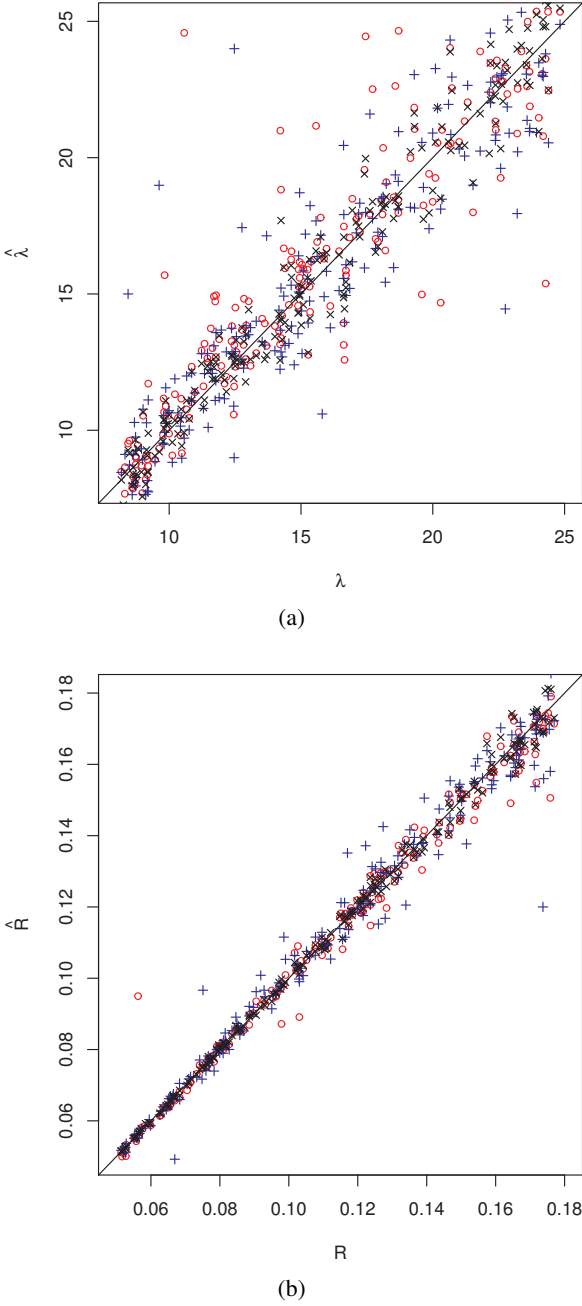


Fig. 1. Plot of 200 re-estimated parameters vs. true parameters for (a)  $\lambda$  and (b)  $R$  using the estimators  $(\hat{\lambda}_{SQS}, \hat{R}_{SQS})$  (red),  $(\tilde{\lambda}_{SQS}, \tilde{R}_{SQS})$  (blue) and  $(\hat{\lambda}_{EQS}, \hat{R}_{EQS})$  (black).

Furthermore, although the BRMSE of the new SQS estimation method using exact statistics is higher than that of MIa and MIb, the BRMSE of the EQS estimation method is smaller than that of MIb, *i.e.*, EQS even outperforms MIb. On the one hand this demonstrates that the quasi-score estimation method can use the available statistical information in an optimal way, thus yielding more precise parameter

estimates. On the other hand it shows that the applied simulation-based implementation of the quasi-score method (SQS) cannot fully exhaust the theoretically possible potential. The reasons for the latter are that Monte Carlo simulation is involved and that in a few cases “wrong” roots are found (see the above discussion), leading to a higher BRMSE.

However, Table 1 also shows that the BRMSE of the new SQS estimation method using the exact or discretized version of the statistics are of the same order of magnitude (cf. also Fig. 1) and that for the case of discretized data SQS clearly outperforms MIa and MIb. This is because in the SQS estimation method the observed and simulation-based statistics are equally treated with respect to the way of how the data is observed, namely, in both cases the estimation of the statistics is based on the discretized model realizations. In particular it is not important whether or not the characteristics  $(A_A, L_A, \chi_A)$  are estimated unbiasedly. More generally, it is even irrelevant for the SQS estimation procedure whether the used statistics possess any direct interpretation, *e.g.*, whether they are estimators of any certain model characteristics. This is one of the fundamental strengths of the proposed SQS estimation method.

In the remainder of this section we assess the precision of the SQS estimates and compare it to the error predicted by the inverse quasi-information matrix. We generated  $N = 100$  observations of the model at the fixed parameter value  $\theta = (15, 0.15)$  and applied the above two kinds of statistics to these observations as we did in the first part. The SQS estimation method was applied, the resultant estimates are shown in Fig. 2 including the MIa and MIb estimates. Fig. 2a shows the comparable performance of the three estimators (MIa, MIb and SQS) in case of exact statistics. Fig. 2b shows again the problem generated by the bias of the estimators of  $L_A$  and  $\chi_A$  in case of discretized data leading to a substantial bias for the method-of-intensities estimators.

The comparison of the empirical estimation error with the predicted error is based on the empirical mean squared error matrix  $MSEM = \frac{1}{N} \sum_{i=1}^N (\theta - \hat{\theta}_i)(\theta - \hat{\theta}_i)^t$  either in the case of the exact statistics, denoted as  $\widehat{MSEM}$ , or for their discretized counterparts,  $\widetilde{MSEM}$ . They read

$$\widehat{MSEM} = \begin{pmatrix} 2.03 & -3.05 \cdot 10^{-3} \\ -3.05 \cdot 10^{-3} & 1.56 \cdot 10^{-5} \end{pmatrix},$$

$$\widetilde{MSEM} = \begin{pmatrix} 2.68 & -5.00 \cdot 10^{-3} \\ -5.00 \cdot 10^{-3} & 2.80 \cdot 10^{-5} \end{pmatrix}.$$

Eq. 15 suggests that the inverse of the quasi-information matrix (Eq. 14) should be very similar to

the mean squared error matrix of  $\hat{\theta}$ . We thus compare this to the average

$$\Sigma = \frac{1}{N} \sum_{i=1}^N \hat{I}(\hat{\theta}_i)^{-1} \in \mathbb{R}^{q \times q} \quad (25)$$

of the approximated quasi-information matrices in the estimated parameter value for each of the 100 simulations, which read, respectively,

$$\hat{\Sigma} = \begin{pmatrix} 2.03 & -4.27 \cdot 10^{-3} \\ -4.27 \cdot 10^{-3} & 2.99 \cdot 10^{-5} \end{pmatrix},$$

$$\tilde{\Sigma} = \begin{pmatrix} 3.19 & -2.86 \cdot 10^{-3} \\ -2.86 \cdot 10^{-3} & 2.61 \cdot 10^{-5} \end{pmatrix}.$$

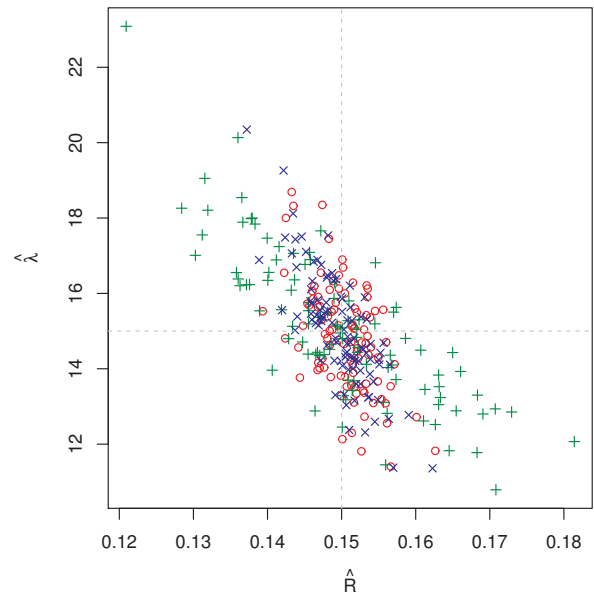
Both compare well to the corresponding observed mean squared error matrix. Fig. 3 compares them graphically. The exact quasi-score method provides a different quasi-information matrix,

$$I(\theta)^{-1} = \begin{pmatrix} 1.30 & -1.89 \cdot 10^{-3} \\ -1.89 \cdot 10^{-3} & 5.07 \cdot 10^{-6} \end{pmatrix},$$

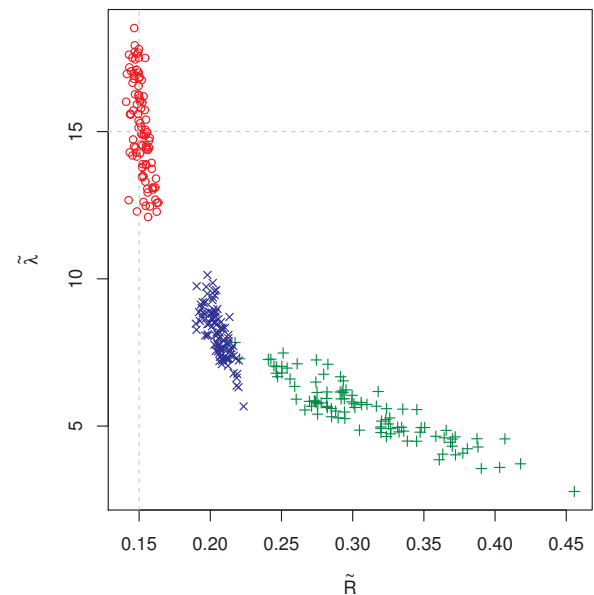
which would suggest a lower estimation error. This corresponds to the fact that this method is more precise. It can be expected that using more simulations would allow to increase the performance of the SQS method towards EQS. The theoretical and empirical evidence thus suggests that the inverse quasi-information might serve as an estimate for the precision of the method. More studies are, however, required to support generalizability of this result.

## DISCUSSION

The new simulation-based quasi-likelihood method allows to estimate parameters for statistical models where neither the likelihood nor moments or characteristics can be computed directly. The user is only required to provide informative statistics and a method to simulate realizations of these statistics for different model parameters. Based on an iterative simulation procedure providing successively improved approximations to the quasi-score function, the estimation algorithm provides a numerical approximation to the efficient quasi-likelihood estimation method. The inverse of the quasi-information matrix computable on the approximation seems to provide a good measure of the estimation error. This allows estimation of model parameters in very general settings.



(a)



(b)

Fig. 2. Re-estimated parameters by SQS (red), MIa (green), MIb (blue) for 100 realizations of the data at  $(\lambda, R) = (15, 0.15)$  using the statistics (a)  $(\hat{A}_A, \hat{L}_A, \hat{\chi}_A)$  or (b)  $(\tilde{A}_A, \tilde{L}_A, \tilde{\chi}_A)$ .

Obviously the procedure is still in its infancy. The global convergence is not guaranteed by the current algorithm and led to substantial errors in about 7% of the runs. The error is still substantially above the theoretical limit given by the exact quasi-likelihood method. The exact quasi-likelihood method can typically not be applied since the necessary ingredients (moments of statistics and their derivatives) often

can not be computed. However, if they could be computed and the statistics are well chosen, one would typically expect better or similar performance to most known estimation procedures. ML estimators

in exponential families are special cases of quasi-likelihood estimators. Quasi-likelihood estimators use moments more efficiently than typical method-of-moment estimators and can use more information.

### ACKNOWLEDGMENTS

This research has been supported by the German Federal Ministry of Education and Research under grant 03MS603B/05M10OFA. We are grateful to André Tscheschel for the opportunity to use his Morph2D software.

### REFERENCES

Altendorf H, Jeulin D (2011). Random-walk-based stochastic modeling of three-dimensional fiber systems. *Phys Rev E* 83:041804.

Ballani F (2007). The surface pair correlation function for stationary Boolean models. *Adv Appl Probab* SGSA 39:1–15.

Chiles JP, Delfiner P (1999). *Geostatistics: Modelling spatial uncertainty*. New York: J. Wiley & Sons.

Chiu SN, Stoyan D, Kendall WS, Mecke J (2013). *Stochastic geometry and its applications*. Chichester: J. Wiley & Sons, 3rd ed.

Cressie NAC (1993). *Statistics for spatial data*. New York: J. Wiley & Sons.

Gaiselmann G, Froning D, Toetzke C, Quick C, Manke I, Lehnert W, *et al.* (2013). Stochastic 3D modeling of non-woven materials with wet-proofing agent. *Int J Hydrogen Energ* 38:8448–60.

Godambe VP, Heyde CC (1987). Quasi-likelihood and optimal estimation. *Int Stat Rev* 55:231–44.

Golub GH, Loan CFV (1996). *Matrix Computations*. Baltimore, MD, USA: JHU Press.

Heinrich L (1993). Asymptotic properties of minimum contrast estimators for parameters of Boolean models. *Metrika* 31:349–60.

Heyde CC (1997). *Quasi-likelihood and its applications: A general approach to optimal parameter estimation*. New York: Springer.

Illian J, Penttinen A, Stoyan H, Stoyan D (2008). *Statistical analysis and modelling of spatial point patterns*. Chichester: J. Wiley & Sons.

Kleijnen JPC, van Beers WCM (2009). Kriging metamodelling in simulation: A review. *Eur J Oper Res* 192:707–16.

Lautensack C (2008). Fitting three-dimensional Laguerre tessellations to foam structures. *J Appl Statist* 35:985–95.

Liang KY, Zeger SL (1995). Inference based on estimating

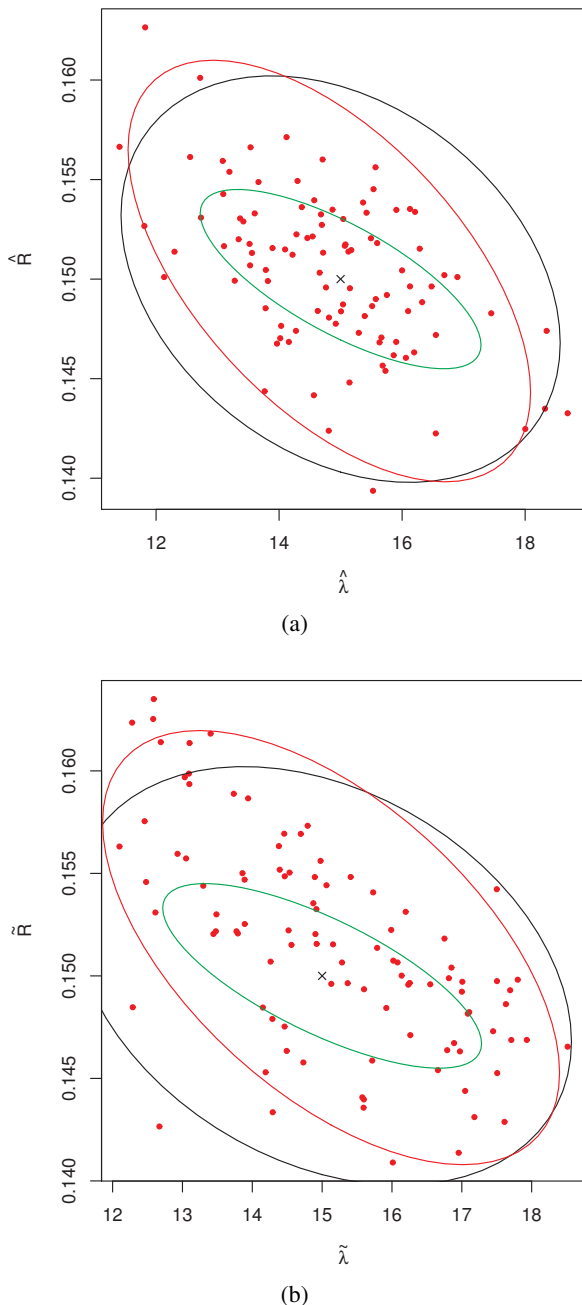


Fig. 3. Re-estimated parameters by the SQS for 100 realizations of the data at  $(\lambda, R) = (15, 0.15)$  using the statistics (a)  $(\widehat{A}_A, \widehat{L}_A, \widehat{\chi}_A)$  or (b)  $(\widetilde{A}_A, \widetilde{L}_A, \widetilde{\chi}_A)$ . The  $2\sigma$ -ellipses are given for the averaged inverse quasi-information (black), the empirical mean squared error matrix of the residuals (red), and the exact inverse quasi-information matrix of the exact quasi-likelihood method (green). The figure illustrates the matching of the first two.

- functions in the presence of nuisance parameters. *Statist Sci* 10:158–73.
- Mardia KV (1996). Kriging and splines with derivative information. *Biometrika* 83:207–21.
- Mardia KV, Marshall RJ (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* 71:135–46.
- Matheron G (1973). The intrinsic random functions and their applications. *Adv Appl Probab* 5:439–68.
- Mecke KR (2001). Exact moments of curvature measures in the Boolean model. *J Stat Phys* 102:1343–81.
- Molchanov IS (1997). *Statistics of the Boolean model for practitioners and mathematicians*. New York: J. Wiley & Sons.
- Møller J, Waagepetersen RP (2003). *Statistical inference and simulation for spatial point processes*. Boca Raton: Chapman & Hall/CRC.
- Nott DJ, Rydén T (1999). Pairwise likelihood methods for inference in image models. *Biometrika* 86:661–76.
- Ohser J, Mücklich F (2000). *Statistical analysis of microstructures in materials science*. Chichester: J. Wiley & Sons.
- Ohser J, Schladitz K (2009). *3D images of materials structures*. Weinheim: Wiley-VCH.
- Osborne MR (1992). Fisher's method of scoring. *Int Stat Rev* 60:99–117.
- Redenbach C (2009). Microstructure models for cellular materials. *Comp Mater Sci* 44:1397–407.
- Sacks J, Welch WJ, Mitchel TJ, Wynn HP (1989). Design and analysis of computer experiments. *Statist Sci* 4:409–23.
- Santner T, Williams B, Notz W (2003). *The design and analysis of computer experiments*. New York: Springer.
- Schneider R, Weil W (2008). *Stochastic and integral geometry*. Berlin: Springer.
- Serra J (1982). *Image analysis and mathematical morphology*. London: Academic Press.
- Spall JC (2003). *Introduction to stochastic search and optimization: estimation, simulation, and control*. Hoboken, New Jersey: J. Wiley & Sons.
- Torquato S (2002). *Random heterogeneous materials: microstructure and macroscopic properties..* New York: Springer.
- Tscheschel A (2005). *Räumliche Statistik zur Charakterisierung gefüllter Elastomere*. Ph.D. thesis, TU Bergakademie Freiberg, Germany.
- Varin C (2008). On composite marginal likelihoods. *Adv Statist Anal* 92:1–28.
- Wackernagel H (2003). *Multivariate geostatistics*. Berlin: Springer.
- Wahba G (1990). *Spline models for observational data*. Philadelphia, PA: SIAM.
- Wilson RJ, Nott DJ (2001). Review of recent results on excursion set models. *Image Anal Stereol* 20:71–8.
- Zimmermann DL (1989). Computationally efficient restricted maximum likelihood estimation of generalized covariance functions. *Math Geol* 21:655–72.