**Hydrology and
Earth System
Sciences**

Calibration of the modified Bartlett-Lewis model using global optimization techniques and alternative objective functions

W. J. Vanhaute[1], S. Vandenberghe[1], K. Scheerlinck[2], B. De Baets[2], and N. E. C. Verhoest[1]

[1]Laboratory of Hydrology and Water Management, Ghent University, Coupure links 653, 9000 Ghent, Belgium
[2]Department of Mathematical modelling, Statistics and Bioinformatics, Ghent University, Coupure links 653, 9000 Ghent, Belgium

Correspondence to: W. J. Vanhaute (willem.vanhaute@ugent.be)

Abstract. The calibration of stochastic point process rainfall models, such as of the Bartlett-Lewis type, suffers from the presence of multiple local minima which local search algorithms usually fail to avoid. To meet this shortcoming, four relatively new global optimization methods are presented and tested for their ability to calibrate the Modified Bartlett-Lewis Model. The list of tested methods consists of: the Downhill Simplex Method, Simplex-Simulated Annealing, Particle Swarm Optimization and Shuffled Complex Evolution. The parameters of these algorithms are first optimized to ensure optimal performance, after which they are used for calibration of the Modified Bartlett-Lewis model. Furthermore, this paper addresses the choice of weights in the objective function. Three alternative weighing methods are compared to determine whether or not simulation results (obtained after calibration with the best optimization method) are influenced by the choice of weights.

1 Introduction

Rainfall is an important input for many models in various branches of applied sciences. Generally, observed time series of rainfall can be used. However, certain applications (such as design studies) require very long time series which are not available from observations (Wheater et al., 2006). To circumvent this problem, one can make use of rainfall models (Boughton and Droop, 2003). When using such models, it is of paramount importance to ensure the modelled rainfall adequately reflects meteorological conditions in the area of interest.

The use of stochastic rainfall models dates back several decades (Waymire and Gupta, 1981), and has since received much attention in literature. Both single-site and spatial-temporal models have been researched extensively. However, the scope of this article is limited to the use of single-site models. For more information about spatial-temporal models, one is referred to Wheater et al. (2005).

An important branch of rainfall models is based on the generation of rectangular pulses. Within these rectangular pulses models, one may discern the Bartlett-Lewis (BL) (Rodriguez-Iturbe et al., 1987a) and Neyman-Scott (NS) (Kavvas and Delleur, 1981) type rainfall models. The distinction between the NS and BL models can be made in the third order moment and proportion dry but not in the second order properties. Hence, both are virtually interchangeable, with the possible exception that the NS models may generate marginally more extreme values. Since empirical analysis revealed that for data observed at Uccle (near Brussels, Belgium), the BL model is preferable (Verhoest et al., 1997), and since this paper makes use of the same data (albeit a more extended dataset), the NS models will not be discussed further. Nonetheless, obtained results may also be applicable to the NS models, due to their strong similarity with the BL models.

Since the formulation of the original BL model by Rodriguez-Iturbe et al. (1987a), this model has been subjected to a number of modifications and extensions. Introducing a jitter, for example, results in more realistically irregular cell intensities (Onof and Wheater, 1994a; Gyasi-Agyei and Willgoose, 1999). Allowing for different cell types to exist, introduces certain variations between storms, which is in accordance with the existence of different types of rainfall

(such as frontal and convective rainfall). To achieve the latter, the mean cell duration can be randomized (Rodriguez-Iturbe et al., 1988), multiple cell types can be defined (Cowpertwait, 1994), or one can make use of multiple superposed processes (Cowpertwait, 2004; Cowpertwait et al., 2007). Finally, to improve extreme value behaviour, the probability distribution of cell intensities can be adjusted to a distribution with a heavier tail (Onof and Wheater, 1994b).

To fit the model to a series of observations, the generalized method of moments is used. In this method, the model is fitted to observed sample properties of rainfall intensity at different aggregation levels. For this purpose, analytical expressions of the expected value of the modelled properties were derived as a function of the model parameters (Rodriguez-Iturbe et al., 1987a).

The calibration of the BL models has proven to be a cumbersome task because of the presence of multiple local minima (Verhoest et al., 1997). Traditional local search techniques sometimes fail to avoid these local minima, resulting in a suboptimal solution to the optimization problem.

Furthermore, the calibration result is influenced by the choice of weights in the objective function. Different approaches exist, but it is not clear which of them leads to better simulation results.

To address these issues, this paper proposes to use relatively new optimization methods as they are expected to be more robust than more traditional local search methods. Furthermore, three different approaches to the weighing of the objective function are compared in order to shed some light on their advantages and disadvantages in terms of model performance and practicality. For these purposes, data recorded at the Uccle-site of the Royal Meteorological Institute (RMI) in Brussels (Belgium), are used. The data set consists of 105 yr of recorded rainfall at an aggregation level of 10 min (De Jongh et al., 2006).

2   The modified Bartlett-Lewis model

Aside from the potential adjustments to the original BL modelling structure, mentioned in Sect. 1, the basic principles of the model have remained intact: storm arrivals occur in a Poisson process with parameter λ, and each storm arrival is followed by a number of cell origins, which also occur in a Poisson process, characterized by parameter β. The duration of the interval in which cell origins are generated is exponentially distributed with parameter γ. Each cell origin is coupled with a rainfall cell, having a random depth and duration, both drawn from exponential distributions characterized by parameters $1/\mu_x$ and η respectively. The superposition of the rainfall cells eventually leads to a continuous rainfall time series.

Extensive analysis of the model by Rodriguez-Iturbe et al. (1987a,b) revealed that the original BL model was well capable of reproducing general rainfall statistics. Conversely, the

wet-dry properties, commonly expressed by the zero depth probability (ZDP), an important feature of the rainfall time series, were not adequately reproduced. These findings led to an adjustment of the model. In the Modified BL (MBL) model, the average cell duration is allowed to vary between storms. This is achieved by letting the parameter of the exponentially distributed cell duration η follow a Gamma distribution with shape and rate parameters α and ν, respectively (Rodriguez-Iturbe et al., 1988). This results in $E[\eta] = \alpha/\nu$ and $Var[\eta] = \alpha/\nu^2$. For the expected duration of a cell to be finite, it is assumed that $\alpha > 1$. Furthermore, Rodriguez-Iturbe et al. (1987a) introduced dimensionless parameters $\kappa = \beta/\eta$ and $\phi = \gamma/\eta$, which are used in the calibration. This way, it can be seen that by keeping κ and ϕ constant, and varying η, storms which exhibit similar structures but consist of different types of cells (i.e. with different average duration and variance) are generated. In total, the MBL model contains 6 parameters that are to be calibrated.

Analyses of the MBL model show that the model is an improvement in comparison with the original BL model (Rodriguez-Iturbe et al., 1988; Entekhabi et al., 1989). Better reproduction of the ZDP, the autocorrelation structure of the rainfall, and the manifestation of extreme rainfall events is observed (Velghe et al., 1994). Nonetheless, the model generates excessive values for the autocorrelation with a lag larger than 12 h (Onof and Wheater, 1993) and the fit of the extremes is still not completely satisfactory.

To improve extreme value behaviour of the model, third order statistics could be included in the objective function. This approach has already proven to lead to good results for the Neyman-Scott models (Cowpertwait, 1998; Burton et al., 2008). Expansion of this concept to the Bartlett-Lewis models can thus be expected to yield similar satisfying results. For the Bartlett-Lewis model at hand, analytical expressions for the third order statistics have not yet been published. However, research on this topic is ongoing and the implementation of third order statistics into the Bartlett-Lewis modelling framework will be addressed in the near future.

3   Calibration procedure

The calibration of Bartlett-Lewis models, and stochastic rainfall models in general, is usually based on the generalized method of moments (GMM). The GMM seeks to minimize the difference between observed sample properties of rainfall intensity and those generated by rainfall models. Alternative methods, such as likelihood approximation (Cameron et al., 2001) or Bayesian inference (Hartig et al., 2011), could be considered for calibration, but these methods tend to be computationally very expensive (Obeysekera et al., 1987). It is not clear whether or not this would have any merit in practical applications (Rodriguez-Iturbe et al., 1988), as from a practical point of view the models are usually treated as being fully identifiable, i.e. only one parameter set is used for

simulation. Moreover, it is not possible to obtain a likelihood function in a closed form, so maximum likelihood approximation is not available as a parameter estimation method.

The use of the GMM for the calibration of stochastic rainfall models is widespread. An array of empirical studies have been performed in which Bartlett-Lewis models were fitted to rainfall time series in Great Britain (Onof and Wheater, 1993, 1994a; Cameron et al., 2000), Ireland (Khaliq and Cunnane, 1996), Belgium (Verhoest et al., 1997; Vandenberghe et al., 2011), the United States (Rodriguez-Iturbe et al., 1987b; Velghe et al., 1994), New Zealand (Cowpertwait et al., 2007), Australia (Gyasi-Agyei and Willgoose, 1999; Heneker et al., 2001), South Africa (Smithers et al., 2002), etc.

In general, the objective function f, which is to be minimized, can be written as:

$$f(\boldsymbol{x}) = (\boldsymbol{M}' - \boldsymbol{M}(\boldsymbol{x}))^T \mathbf{W} (\boldsymbol{M}' - \boldsymbol{M}(\boldsymbol{x})) \tag{1}$$

where $\boldsymbol{x}$ is the parameter vector, $\boldsymbol{M}'$ is the vector of observed values for a set of k properties, $\boldsymbol{M}(\boldsymbol{x})$ is the vector of their expected values under the model (calculated through analytical expressions), and $\mathbf{W}$ is a $k \times k$ positive definite matrix of weights. The objective function value f for a given set of parameters $\boldsymbol{x}$ is also referred to as the fitness of the proposed parameter vector, as it reflects the quality of this potential solution to the optimization problem.

The chosen fitting properties in the current work include the mean (Avg), variance (Var), lag-1 autocovariance (Cov), and the proportion of dry intervals or zero depth probability (ZDP). Each of these are evaluated at aggregation levels of 10 h, 1 h, 6 h and 24 h. This is similar to the fitting properties chosen by Cowpertwait et al. (2007).

As discussed in Sect. 1, the parameters of the MBL model all follow a different probability distribution function. The support for each of these parameters is the interval $[0, +\infty]$, except for α, for which a lower boundary of 1 is assumed (see Rodriguez-Iturbe et al., 1988). The used algorithms should obey these boundaries, otherwise, numerical instabilities might emerge when calculating the analytical expressions using negative values, which would trouble the calibration or lead to erroneous results. Another implication that arises is the impracticality of working with $+\infty$ as an upper boundary, as this might impede the convergence of the optimization method. Therefore, the theoretical upper boundaries are tightened so that the parameters can still take a wide range of feasible values, and in addition, contain the previously calibrated values of the MBL model (Verhoest et al., 1997). Table 1 shows the set of boundaries which is assumed to constitute the feasible parameter space of the model.

The choice of $\mathbf{W}$ is rather subjective. Many different approaches have been explored in literature. The theory of Hansen (1982) suggests that the inverse of the covariance matrix of the observed properties should be used as $\mathbf{W}$. In terms of parameter identifiability, this would be the theoretically optimal starting point (Kaczmarska, 2011). However,

Table 1. Pre-defined boundaries for the parameters of the MBL model during calibration, applicable to all months.

Parameter	λ	κ	ϕ	μ_x	α	ν
Lower boundary	0	0	0	0	1	0
Upper boundary	0.1	20	1	15	20	20

for simplicity, in most cases $\mathbf{W}$ is chosen to be a diagonal matrix. In that case the objective function is reduced to:

$$f(\boldsymbol{x}) = \sum_{i=1}^{k} w_i \left(M_i' - M_i(\boldsymbol{x}) \right)^2 \tag{2}$$

Frequently, w_i is set equal to $a_i / M_i'^2$, where a_i is a user defined value (Entekhabi et al., 1989; Cowpertwait, 1991; Velghe et al., 1994; Verhoest et al., 1997; Smithers et al., 2002; Cowpertwait, 2004). Division of the squared model error by the sample estimate ensures that large values do not dominate the minimization procedure (Cowpertwait et al., 2007). The variable a is usually chosen arbitrarily, to ensure a good reproduction of certain fitting properties. Cowpertwait (1991), for example, chooses $a = 100$ for the mean and $a = 1$ for the other fitting properties. Velghe et al. (1994) and Verhoest et al. (1997), on the other hand, choose $a = 1$ for all fitting properties. In this paper we will follow the approach of Velghe et al. (1994) and Verhoest et al. (1997) and call this objective function OF1.

An alternative configuration of the objective function is suggested by Cowpertwait et al. (2007):

$$f(\boldsymbol{x}) = \sum_{i=1}^{k} \left[\left(\frac{M_i(\boldsymbol{x})}{M_i'} - 1 \right)^2 + \left(\frac{M_i'}{M_i(\boldsymbol{x})} - 1 \right)^2 \right] \tag{3}$$

The use of an additional term which contains the reciprocal value of the division present in Eq. (2) helps to ensure that no bias is present in the optimal solution, in case an exact fit is not obtained. This objective function will be referred to as OF2.

Finally, a simplification of the theory of Hansen (1982) can be used to weigh the objective function. Here, $w_i = 1/\text{Var}[M_i']$ is used. This makes sense because "in least squares problems with unequal variances, observations should be weighed according to the inverse of their variances" (Chandler, 2004). The empirical variance of the observed sample properties is obtained by calculating these properties for each year separately, which results in a series of 105 repetitions of that particular property. The variance of these repetitions is calculated and used in the objective function. This approach will further be referred to as OF3. A more clarifying overview of the used objective functions is given in Table 2.

Finally, it should be mentioned that the MBL model is fitted on a monthly basis, i.e. 12 different parameter sets have

Table 2. Expressions of used objective functions for the calibration of the Modified Bartlett-Lewis model.

Name	Expression
OF1 :	$f(\boldsymbol{x}) = \sum_{i=1}^{k} \dfrac{\left(M_i' - M_i(\boldsymbol{x})\right)^2}{M_i'^2}$
OF2 :	$f(\boldsymbol{x}) = \sum_{i=1}^{k} \left[\left(\dfrac{M_i(\boldsymbol{x})}{M_i'} - 1 \right)^2 + \left(\dfrac{M_i'}{M_i(\boldsymbol{x})} - 1 \right)^2 \right]$
OF3 :	$f(\boldsymbol{x}) = \sum_{i=1}^{k} \dfrac{\left(M_i' - M_i(\boldsymbol{x})\right)^2}{\mathrm{Var}[M_i']}$

to be calibrated, i.e. one for each month. This approach is upheld to cancel out any seasonal effects present in the rainfall time series. This is necessary to ensure temporal homogeneity (Obeysekera et al., 1987; Verhoest et al., 1997).

The objective of this paper is to determine whether the choice of the objective function has a significant impact on the estimated parameters. If this is the case, the impact of the choice of the objective function can be assessed by considering properties that were not used during the fitting, but are of hydrological importance. If no significant impact can be observed, a distinction can be made based on the efficiency of the calibration.

4 Optimization methods

The calibration of the Bartlett-Lewis models has been reported as being a cumbersome task (Verhoest et al., 1997), as the optimization is troubled by the presence of multiple local minima, in which local optimization techniques tend to get trapped. In the past, such techniques have been used in the majority of the cases. For example, Velghe et al. (1994); Verhoest et al. (1997); Onof and Wheater (1993) use Powell's method (Press et al., 1986) for calibration. This gradient-based method is prone to get stuck in local optima. Furthermore, the user has to supply the algorithm with an initial guess of the solution, which may lead to a bias in the results (Khaliq and Cunnane, 1996). More recently, most authors opt to calibrate using the Simplex method by Nelder and Mead (1965) with multiple starting points. Occasionally the outcome is further minimized using a gradient-based method (Wheater et al., 2006; Cowpertwait et al., 2007; Kaczmarska, 2011).

In recent years, an array of global optimization methods has been developed. In this work, we test four of those algorithms with respect to the calibration of the BL models. The following sections discuss the theoretical background of the used optimization methods which, ultimately, stem from fundamentally different conceptual backgrounds. The current paper does not aim at discriminating against or rallying for a certain method, merely an objective comparison, highlighting advantages and disadvantages of the presented methods is aspired.

4.1 Downhill simplex method

The Downhill Simplex Method (DSM) is based on an idea by Spendley et al. (1962) for tracking ideal operating conditions by evaluating the output of a system at a set of points, forming a simplex, in the parameter space, and the continuous formation of new simplices by reflecting a point in the hyperspace of the other points. Nelder and Mead (1965) acknowledged this concept's merit in the optimization of mathematical formulas. The simplex moves autonomously through the parameter space by a sequence of intermittent reflections, contractions and expansions.

Suppose an objective function contains D variables and is subjected to a minimization procedure without posing any restrictions on the values of the variables. Then, suppose $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_D$ are $(D+1)$ points in the D-dimensional space which form the current simplex. The objective function value of each point $\boldsymbol{x}_i$ is written as y_i.

The point with the lowest objective function value receives the subscript l ($y_l = \min_i y_i$), the point with the highest objective function value receives subscript h ($y_h = \max_i y_i$). Point $\bar{\boldsymbol{x}}$ is defined as the centroid of those points for which $i \neq h$. The distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is expressed by $\mathrm{d}(\boldsymbol{x}_i, \boldsymbol{x}_j)$. During each step of the process, $\boldsymbol{x}_h$ is replaced by a new point by a reflection, contraction or expansion of the simplex. The reflection of $\boldsymbol{x}_h$ can be written as $\boldsymbol{x}_h^*$, whose coordinates can be found by the following relationship :

$$\boldsymbol{x}^* = (1+\alpha)\bar{\boldsymbol{x}} - \alpha \boldsymbol{x}_h \tag{4}$$

with α a positive constant, known as the *reflection coefficient*. In other words, $\boldsymbol{x}^*$ lies on a straight line, connecting $\boldsymbol{x}_h$ and $\bar{\boldsymbol{x}}$, opposite to $\bar{\boldsymbol{x}}$, with $\mathrm{d}(\boldsymbol{x}^*, \bar{\boldsymbol{x}}) = \alpha \mathrm{d}(\boldsymbol{x}_h, \bar{\boldsymbol{x}})$. $\boldsymbol{x}_h$ is replaced by $\boldsymbol{x}_h^*$ and the process starts over with the new simplex if $y_h > y^* > y_l$.

If, on the other hand, the reflection has created a new minimum ($y^* < y_l$), the simplex is expanded from $\boldsymbol{x}^*$ towards $\boldsymbol{x}^{**}$ according to :

$$\boldsymbol{x}^{**} = \gamma \boldsymbol{x}^* + (1-\gamma)\bar{\boldsymbol{x}} \tag{5}$$

The *expansion coefficient* γ, which is larger than 1, equals $\mathrm{d}(\boldsymbol{x}^{**}, \bar{\boldsymbol{x}})$ divided by $\mathrm{d}(\boldsymbol{x}^*, \bar{\boldsymbol{x}})$. If $y^{**} < y_l$, then $\boldsymbol{x}_h$ is replaced by $\boldsymbol{x}^{**}$ and the process starts over. If, on the other hand, $y^{**} > y_l$, then the expansion is considered to have failed and $\boldsymbol{x}_h$ is replaced by $\boldsymbol{x}^*$ before re-initiating the process.

If the reflection of $\boldsymbol{x}$ to $\boldsymbol{x}^*$ results in a situation where $y^* > y_i$ for all $i \neq h$, i.e. replacing $\boldsymbol{x}$ by $\boldsymbol{x}^*$ results in the creation of a new maximum y^*, then $\boldsymbol{x}_h$ becomes either $\boldsymbol{x}_h$ or $\boldsymbol{x}^*$, whichever has the smallest function value, and $\boldsymbol{x}^{**}$ is calculated as follows :

$$\boldsymbol{x}^{**} = \beta \boldsymbol{x}_h + (1-\beta)\bar{\boldsymbol{x}}. \tag{6}$$

The *contraction coefficient* $\beta \in [0,1]$ is the ratio of $d(\boldsymbol{x}^{**}, \bar{\boldsymbol{x}})$ to $d(\boldsymbol{x}, \bar{\boldsymbol{x}})$. $\boldsymbol{x}^{**}$ replaces $\boldsymbol{x}_h$ and the algorithm proceeds to the next iteration, unless $y^{**} > \min(y_h, y^*)$. In the latter case the contraction resulted in a point that has a higher function value than both $\boldsymbol{x}_h$ and $\boldsymbol{x}^*$. In case of such a failed contraction all $\boldsymbol{x}_i$'s are replaced by $(\boldsymbol{x}_i + \boldsymbol{x}_l)/2$ before continuing to the next iteration. The propagation of the simplex continues until a suitable result (expressed by the stopping criteria) is obtained.

As can be deduced from the description above, the DSM is a local, rather than a global search method. The outcome of the optimization strongly depends on the initial position, provided by the user. To enable a more global search, and to avoid a biased outcome resulting from a user provided initial position, multiple starting points are chosen randomly within the boundaries of the search space. In this specific case 30 initial positions are generated. The DSM is applied to each of these initial positions, after which the single best result is selected as the outcome of the DSM with multiple starting points. This method is used throughout the article, any further mentioning of the DSM thus refers to the methodology using 30 initial positions.

4.2 Simplex-simulated annealing

Simplex-Simulated Annealing (SIMPSA) is a hybridization between the DSM by Nelder and Mead (1965) and Simulated Annealing (Kirkpatrick et al., 1983; Kirkpatrick, 1984). The latter method, which is based on the metallurgic process of Annealing, enforces the DSM through its ability to escape local minima and thus avoid premature convergence. The annealing process was first simulated by Metropolis et al. (1953), and was later picked up by Kirkpatrick et al. (1983), to be used as an optimization algorithm. In the original Metropolis algorithm, an equilibrium composition of molecules, which yields minimum energy at a given temperature, is sought after through successive random displacements. Because a thermic balance is characterized by a Boltzmann distribution of energy levels, transitions towards a lower, as well as towards a higher energy level are possible. This feature is thought to be the main reason why a minimum energy level can be reached (Aarts and Van Laarhoven, 1987).

This concept can easily be translated towards an optimization algorithm. By replacing the energy level of the system with the value of an objective function, the search for a minimum energy level is converted to a search for the minimum of the objective function. As stated before, this leads to an optimization scheme which is fairly consistent and, above all, less prone to get stuck in local minima, in comparison with gradient-based methods. An issue, however, is that Simulated Annealing relies on a random walk through the parameter space. This shortcoming can be mitigated by the use of the DSM to guide the search of the optimum, which allows for a more structured search. This is what constitutes

SIMPSA (Press and Teukolsky, 1991; Cardoso et al., 1996). From the point of view of the DSM, the incorporation of the Simulated annealing framework enforces the DSM in its global search. By allowing occasional missteps, the algorithm can be steered away from local minima, increasing the robustness of the outcome.

The probability of a misstep is controlled by the temperature T of the system. A new configuration corresponding to a lower energy level (i.e. $\Delta E < 0$) or lower function value, is unconditionally accepted. When, on the other hand, a solution is found which cannot be accepted as an improvement (i.e. $\Delta E \geq 0$), there still is a chance $P(\Delta E) = \exp(-\Delta E/k_b T)$ that it will nevertheless be accepted. Thus, the probability of making a move in the "wrong" direction is very high at the beginning of the cooling process, i.e. a global search of the parameter space in conducted. As the temperature decreases, the chances of making a wrong move decrease accordingly, approaching zero and thus the algorithm ultimately converges towards the original DSM.

It is clear that, in order to fully exploit the potential of SIMPSA, the initial temperature has to be chosen carefully, satisfying the condition that almost all wrong steps should be accepted at the start of the iteration process. Then, as the optimization progresses, the chances of making erroneous moves should decrease, which means the temperature will steadily decrease according to a predefined cooling schedule. The cooling schedule proposed by Aarts and Van Laarhoven (1985) is used :

$$T^{j+1} = \frac{T^j}{1 + \frac{T^j \cdot \ln(1+\delta)}{3\sigma}} \tag{7}$$

with δ and σ trajectory parameters, and j the current iteration. δ is the cooling rate, which controls the speed of the cooling. Small values (<1) will result in slow convergence while larger values (>1) result in convergence to inferior local minima. Finally, σ is the standard deviation of the objective function value of all configurations at a certain temperature T^j at iteration step j.

In order to initiate the optimization scheme an initial guess of the parameter vector has to be supplied by the user. This initial guess is used by SIMPSA to construct the additional D points of the simplex needed in a D-dimensional problem. Therefore, the initial guess is perturbed in one of its D dimensions to create the initial simplex. Once the initial simplex is formed, the iteration process can commence and new simplex configurations are formed like in the original DSM.

The way in which SIMPSA incorporates the Simulated Annealing methodology of making the occasional faulty move is as follows: when propagating the simplex following the aforementioned rules (Sect. 4.1), a positive, log-distributed variable, proportional to the control temperature T, is added to the function value of each of the points of the simplex. In a similar fashion, the function value of the newly found point is diminished by a randomly chosen value. In

this manner, a new point (e.g. created by a reflection) with a higher objective function value than the other points of the simplex (which would be rejected by the DSM's rules), still has a chance (proportional to T) of being accepted.

4.3 Particle swarm optimization

The accomplishment of complex objectives through the use of simple individual interactions has been an important source of influence for a certain type of artificial intelligence, collectively termed *Swarm Intelligence*. One of those techniques is Particle Swarm Optimization (PSO), initially introduced to simulate social behaviour and later adapted to be used as an optimization method (Kennedy and Eberhart, 1995). PSO is based on the behaviour of herd animals, characterized by the absence of a leader in the herd. Notwithstanding the absence of such a leader, the herd is able to act as a collective, mainly due to local interactions between the individuals. This allows them to attain certain goals, such as the gathering of food and the evasion of predators. The simulation of this behaviour and the replacement of the aspired goal by some sort of objective function, makes this method particularly useful for solving an optimization problem (Engelbrecht, 2006).

The PSO algorithm consists of a swarm of N particles, each particle representing a possible solution to the problem at hand. The particles travel the multi-dimensional search space, in search of the global optimum. The search is led by a combination of information gathered by the particle itself, and by the community as a whole. In a D-dimensional search space, the position and velocity of a certain particle i, with $i = 1, ..., N$, can be represented by a D-dimensional vector $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{iD})$ and $\boldsymbol{v}_i = (v_{i1}, v_{i2}, ..., v_{iD})$, respectively. The position $\boldsymbol{x}_i$ of a particle can be adjusted by adding its speed vector $\boldsymbol{v}_i$ to the current position. This can be expressed by the following equation:

$$\boldsymbol{x}_i(t+1) = \boldsymbol{x}_i(t) + \boldsymbol{v}_i(t+1) \tag{8}$$

for which t and $t+1$ express the current and subsequent iteration step.

The velocity drives the optimization. It determines the speed and direction in which the particles move, thus orchestrating the collective search of, and convergence towards, the global optimum. For this purpose, the aforementioned vector is equipped with two components, each containing specific information about the objective. (1) The *cognitive component* reflects the personal experience of a given particle, while (2) the *social component* bears information gathered by the particle's neighbourhood. Many different approaches exist for defining the size and shape of this neighbourhood, more detailed information can be found in Engelbrecht (2006). For the sake of simplicity, a global neighbourhood is selected for the current application. The social component thus consists of information gathered by the swarm as

a whole during all preceding iterations and the current, and is represented by the best found position $\boldsymbol{p}_g$ by the swarm. Accordingly, the cognitive component consists of information about the objective function, obtained by a certain particle i. It is expressed by the best previously visited position $\boldsymbol{p}_i$ of particle i in the search space. Both components are combined to update the velocity:

$$\boldsymbol{v}_i(t+1) = w \cdot \boldsymbol{v}_i(t) + c_1 \cdot r_1(t) \cdot [\boldsymbol{p}_i(t) - \boldsymbol{x}_i(t)]$$
$$+ c_2 \cdot r_2(t) \cdot [\boldsymbol{p}_g(t) - \boldsymbol{x}_i(t)], \tag{9}$$

with $\boldsymbol{v}_i(t)$ the velocity of particle i at iteration step t; $\boldsymbol{x}_i(t)$ the position of the i-th particle at iteration step t; c_1 and c_2 positive acceleration constants, used to scale the influence of the cognitive and social components. The variables $r_1(t)$ and $r_2(t)$ are uniformly distributed between 0 and 1, and introduce a stochastic component in the optimization. Finally, w is the inertia weight. These parameters all have an important influence on the performance of the PSO algorithm, a more detailed discussion of their significance and a method for their selection are reported in Sect. 5.2.

4.4 Shuffled complex evolution

The Shuffled Complex Evolution algorithm (SCE-UA), originally developed for the calibration of a watershed model (Duan et al., 1994), is based on a synthesis of four concepts: (1) a combination of deterministic and probabilistic approaches; (2) systematic evolution of a "complex" of points spanning the parameter space, in the direction of global improvement; (3) competitive evolution; (4) complex shuffling. The combination of these concepts, most of which have already proven their merit in global optimization problems, makes the SCE-UA method robust, effective, flexible and efficient (Duan et al., 1994). The SCE-UA method (following the description by Duan et al., 1994) can be summarized by the following steps:

1. at the start of the optimization procedure a random sample of s points is generated in the feasible parameter space (defined by the user). Since no prior information about the approximate location of the global optimum is available, a uniform distribution is used to generate this initial sample. In each of the s points, one can calculate the corresponding objective function value.

2. Once the objective function values of the generated samples are known, they can be ranked according to their objective function values, the first having the lowest function value and the last having the highest function value (or the other way around for a maximization problem).

3. The s points are then partitioned into p complexes, each containing m points. The complexes are partitioned in such a manner that complex i, where $i = 1, 2, ..., p$,

contains every $p(k - 1) + i$ ranked point, with $k = 1, 2, \ldots, m$.

4. The constructed complexes are allowed to evolve according to the competitive complex evolution (CCE) algorithm (which will be discussed further below).

5. After the complexes have evolved they are shuffled. This is accomplished by combining them into a single sample population: sorting the population in order of increasing objective function value and ultimately shuffling the sample population into p complexes following the procedure outlined in step 3.

6. Before continuing the iterative process convergence criteria are checked. If none of them are met the process continues, otherwise the process is aborted and the optimum is assumed to have been found.

7. In a final step the reduction of the number of complexes is checked. If the minimum number of complexes required in the population, $p_{\min}$, is less than p, then the complex with the lowest ranked points is removed, p is replaced by $p - 1$ and $s = pm$, after which the process restarts at step 4. If, however, $p_{\min} = p$, the algorithm returns to step 4.

The effectiveness of the SCE-UA method can be attributed to several factors. First of all the use of a population avoids biases resulting from the use of a single user-defined initial point. On the other hand, the partitioning into different complexes allows for an extensive exploitation of the parameter space, while the shuffling of the complexes is a way of sharing knowledge on a larger scale, representing the explorative character of the algorithm.

A key component of the SCE-UA method is the CCE algorithm. The CCE algorithm controls the evolution of the points within a certain complex. Within each complex, a subcomplex is formed. A fixed number of points is drawn from a trapezoidal probability distribution (constructed so that the point with the best function value has the best odds of being selected) and are assigned to the subcomplex. The members of the subcomplex can be regarded as parents which are about to generate offspring. The idea of competitiveness, introduced in the formation of the subcomplexes (not all points of the complex are allowed to procreate), expedites the search towards promising regions. Offspring is generated via the use of the DSM by Nelder and Mead (1965). The simplex is formed by the points of the subcomplex, and is allowed to progress for a fixed number of inner loop iterations (nI), resulting in the offspring. The generation of offspring is repeated a number of times before the complexes are shuffled and the process starts over. As the optimization progresses, the entire population is expected to converge towards the neighbourhood of the global optimum, provided that the initial population size and the number of complexes is sufficiently large.

5 Implementation of the optimization methods

For each of the optimization algorithms described, several parameters have to be selected in order to fully exploit their potential. The selection of those parameters will have an influence on the effectiveness of the algorithms, so a careful consideration of the available options will contribute to the objectiveness of the overall comparison. Therefore, the conducted selection procedure for the parameters of the optimization algorithms is described below. Furthermore, the implementation entails the specification of measures to be taken against infeasible parameter combinations, i.e. points which lie outside the delineated boundaries. These measures will also be described in the following sections.

All algorithms that are based on a simplex design are stopped by the same stopping criteria. The iterative process is called to a halt when the differences in objective function values between the points of the simplex are smaller than a certain threshold, or when the positions of the simplex points differ less than a given threshold. For PSO, other criteria need to be specified because of the different conceptual approach. This is further elaborated in Sect. 5.2.

5.1 Simplex-simulated annealing

The performance of SIMPSA can be enhanced for a specific problem by fine-tuning some of the parameters of the algorithm. Most of the parameters have been set to their recommended value (Cardoso et al., 1996), except for the cooling rate δ (see Eq. 7) and the final temperature. To avoid the algorithm getting stuck in local optima, it is opted to choose the cooling rate $\delta < 1$, which will lead to slower but steady convergence instead of fast convergence towards (possibly) inferior optima. Therefore, different cooling rates are tested, δ is varied between 0.2 and 0.9 with an increment of 0.1.

The second parameter that can be adjusted is the final temperature. The final temperature is the temperature at which SIMPSA is reduced to the original DSM. Put differently, if the final temperature is reached, the temperature is set to zero and the chances of making movements in the wrong direction become equal to zero. Obviously, the higher the final temperature, the faster the algorithm will come to a halt. Care must be taken however, a final temperature that is too high will nullify the expected advantages of using SIMPSA instead of the DSM. Different values for the final temperature are tested to investigate its influence on the overall calibration result. The final temperature is tested at {1, 0.1, 0.01, 0.001, 0.0001, 0.00001}.

Figure 1 shows the dependence of the objective function value on the cooling rate and the final temperature. For each month, calibrations were performed using these different combinations for cooling rate and final temperature. Figure 1 displays the average result obtained over the different months, so as to provide an image on how well the combination of cooling rate and final temperature performs for the
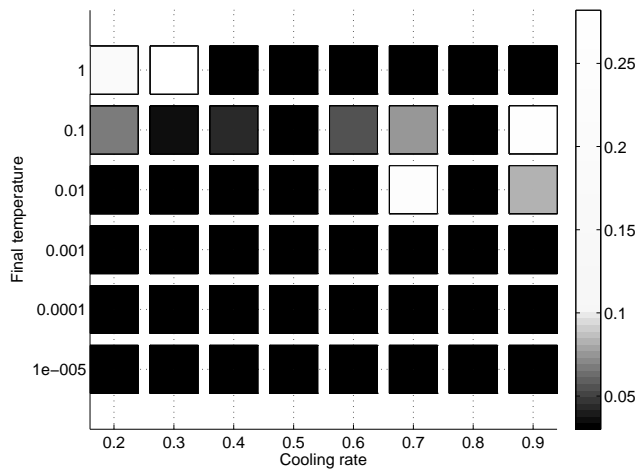
Fig. 1. Dependence of the fitness on the cooling rate δ and the final temperature of SIMPSA.

calibration of all months. It should be noted that for different combinations of cooling rate and final temperature, the same minima are obtained. To highlight this, and to increase interpretability of the plots, the color scale of Fig. 1 (and of plots in the following sections) is adjusted so that only parameter combinations with the lowest obtained objective function values are attributed gray shades. Parameter combinations that resulted in the minima are attributed a black colour.

Results confirm that for lower final temperatures better results are obtained, this is because the parameter space is explored more intensively. For the cooling rate δ, it seems that the best results are obtained when it is set to 0.5 or 0.8. Figure 1 shows that for $\delta = 0.5$ and $\delta = 0.8$ equally good results are obtained for different final temperatures. We choose to set δ equal to 0.5, to favour a slower convergence, minimizing the risk of getting stuck in a local optimum. The final temperature is set to 1, as this will lead to a faster calibration than would be the case if the final temperature were to be set lower.

Besides the choice of the above-mentioned parameters, the user has to set up rules about what to do with generated points which are outside the parameter boundaries. One possible way is to just place boundary violating points at the boundary (Box, 1965). This, however, may lead to considerable instabilities in the simplex which may cause it to collapse (Cardoso et al., 1996). Another solution is to reset them at a random position in the feasible parameter space. In this paper, we opted for accepting infeasible points, however, the objective function is adjusted so that infeasible points automatically receive a very high objective function value, which will force the simplex back into the feasible parameter space (Nelder and Mead, 1965). Note that the same approach is used for the DSM.

5.2 Particle swarm optimization

As outlined in Sect. 4.3, the performance of PSO is influenced by several parameters. To ensure efficient convergence to the global optimum, care must be taken in the selection of these parameters. For example, the population size N must be chosen large enough so that the parameter space is sufficiently explored, but not too large because of the obvious increase in computational burden accompanied with such an increase in population size. Furthermore, the cognitive parameter c_1, social parameter c_2 and the inertia weight w will also have an impact on the speed and efficiency of the algorithm. The relative magnitude of c_1 and c_2 determine the exploration/exploitation trade-off made by PSO. Exploration means the particles will be able to explore the whole parameter space and identify promising parameter regions. They will, however, lack in accuracy to find a satisfactory optimum in this promising region. Exploitation, on the other hand, means each particle will be pre-occupied with its own local search, not interacting much with the other particles. This way, the parameter space will not be explored sufficiently, so inferior results might be obtained. For the PSO algorithm to work properly, the balance of the exploration/exploitation trade-off is of key importance. A larger social parameter c_2, for example, means that more importance is given to the global best position, i.e. exploration is favoured. If, on the other hand, the cognitive parameter c_1 outweighs the social parameter, much more care is given to the local exploitative search by the particles. Finally, the inertia weight w ($0 < w < 1$) will slow down the velocity of the particle at a previous iteration step. Large values of w facilitate exploration of the parameter space (higher velocities will lead to more extensive coverage of the parameter space), whilst a small w facilitates exploitation (Engelbrecht, 2006).

To select the parameters which will lead to the best calibration results for the model under study, an exhaustive parameter search is conducted (Scheerlinck et al., 2009). Parameters c_1, c_2 and w are varied and monthly calibrations are conducted for the different combinations. The population size N is fixed at 30 particles (Engelbrecht, 2006). The other parameters are varied in their convergence domain (Trelea, 2003; Jiang et al., 2007). Parameters c_1 and c_2 range from 0.5 to 2.5 and w ranges from 0.2 to 1 with an increment of 0.2.

Figure 2 displays the results of the exhaustive parameter search. This figure shows the mean obtained objective function value for the monthly calibrations. For the three parameters, a value can be chosen such that, on average, the algorithm performs well for all months. Figure 2a shows the dependence of the fitness with regard to the social and the cognitive parameters, c_1 and c_2, with a fixed inertia weight ($w = 0.4$). The results show that, in order to obtain good calibration results, a trade-off has to be made between c_1 and c_2. High values of c_1 in combination with low values of c_2 yield good results and vice versa. A choice thus has to be made by the user in favour of either a more explorative, or a more
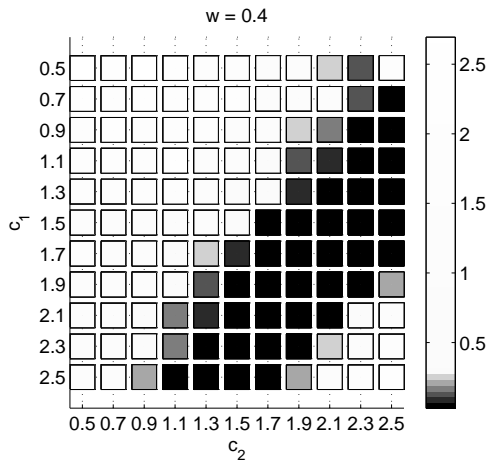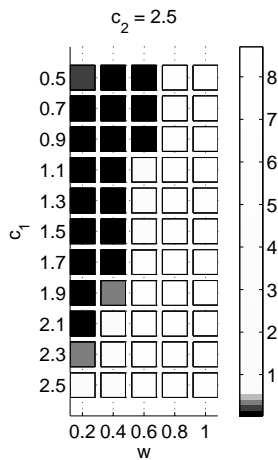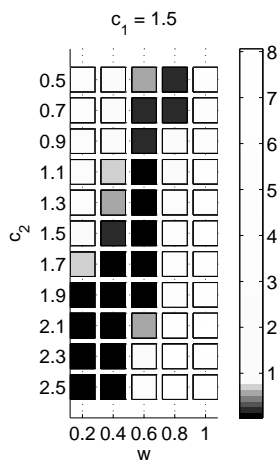
Fig. 2. Dependence of the objective function value on the number of complexes p and the number of inner loop iterations nI of SCE-UA.
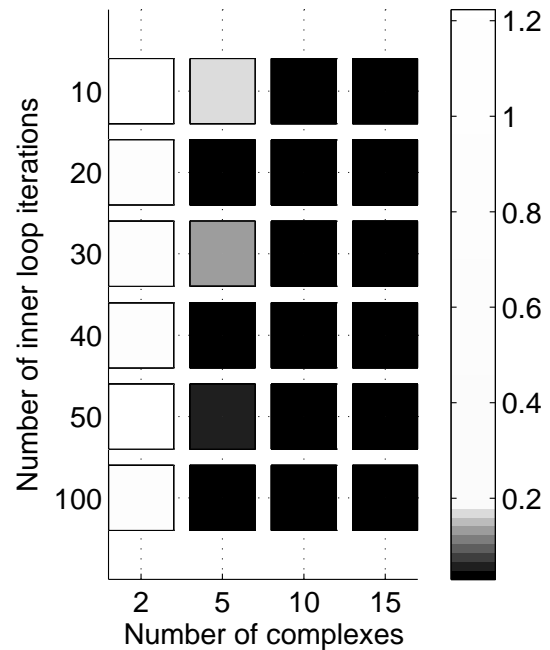

Fig. 3. Distributions of estimation error for each parameter under different optimization methods and objective functions.

exploitative search. In this case, c_2 is set at 2.5, a search with an explorative character is preferred.

Figure 2b shows the results of the calibrations for different combinations of w and c_1 when c_2 is fixed at 2.5. Results show that good results are obtained when the inertia weight w is kept relatively small, whilst the choice of c_1 does not seem to have a profound effect on the result when using small values for w. A minimum could, however, be detected for $c_1 = 1.5$ and consequently this value is chosen for c_1.

Finally, Fig. 2c shows the results for different combinations of w and c_2 with c_1 fixed at 1.5. This figure shows that there is a negative correlation between both parameters. Higher values of c_2 in combination with lower values of w lead to promising results, and vice versa. Because the value c_2 was previously set at 2.5, the value of w is set at 0.4. The figure shows that for this combination good results are obtained.

In addition to the selection of the PSO parameters, several other settings have to be specified. When a population member attempts to cross parameter boundaries, that boundary acts as a perfect reflector. In other words, the direction of displacement of the particle is inverted in order to keep it inside the parameter boundaries.

The choice of a stopping criterion will also affect the obtained calibration results. In addition to the already mentioned general stopping criteria several other approaches can be utilized for PSO. In the current paper the search procedure is stopped if the global best solution $\boldsymbol{p}_g$ does not change during 30 subsequent iterations. This indicates that no better
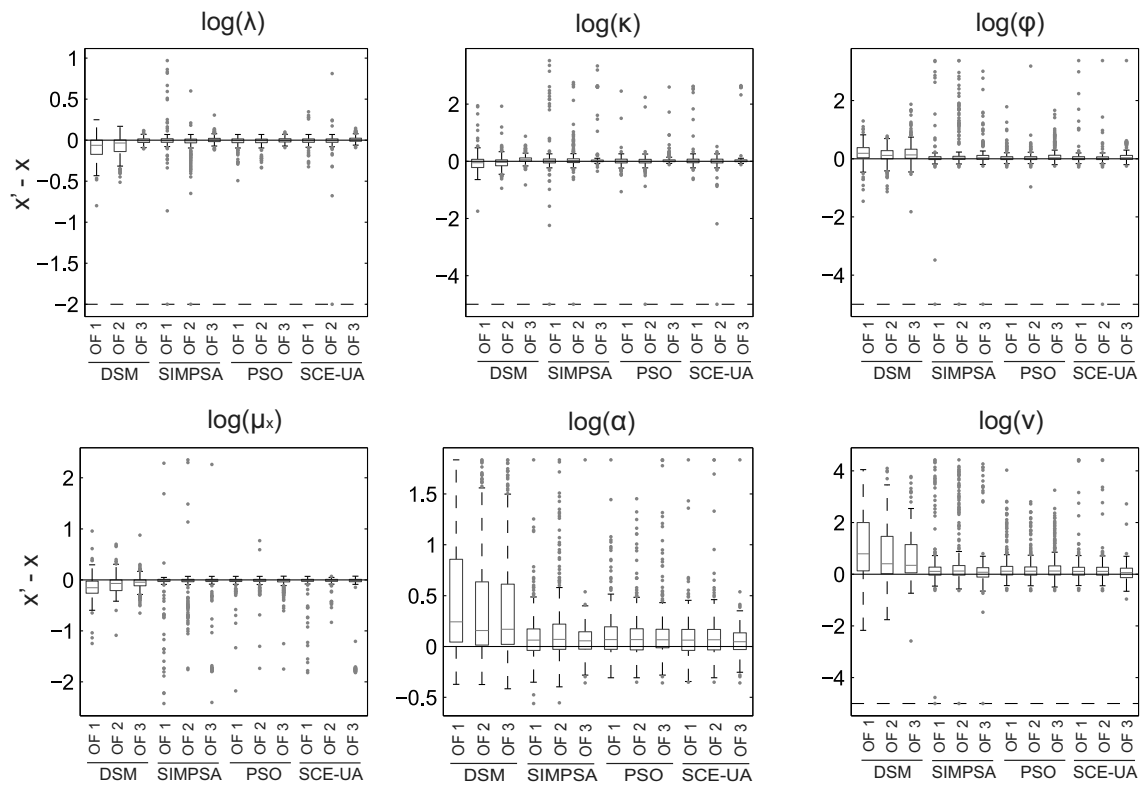
Fig. 4. Distributions of estimation error for each parameter, grouped according to the used objective function.

solutions are being found. This criterion is preferred against a convergence criterion, i.e. a certain fraction of the population has to converge to the same solution in order for the algorithm to stop, because, from personal experience, it has become clear that the latter is quite time consuming.

5.3 Shuffled complex evolution

The parameters of the SCE-UA algorithm have been set to their recommended values (Duan et al., 1994), except for the number of complexes p and the number of inner loop iterations nI which are evaluated more closely. The value of p is recommended to lie between 2 and 20. Therefore p is evaluated at different values within this interval. It is expected that the use of more complexes will lead to better results, but at a higher computational cost. The number of inner loop iterations will also have an influence on the results. The higher the number of inner loop iterations, the better the offspring that is being generated (the simplex will be able to progress further towards an optimum), again, bearing in mind the augmented computational burden. The parameter search thus seeks to find a good balance between results and efficiency. The procedure is the same as with PSO and SIMPSA. For every combination of SCE-UA parameters calibrations are conducted for each month. Those results are averaged and displayed in Fig. 3. These results confirm that a higher number of complexes leads to better results, as does a higher num-

ber of inner loop iterations. The number of complexes p, however, is best set at 5, since for this value good results are obtained throughout and this is less computationally expensive than 10 or 20 complexes. The value of nI is chosen as 20. It can be seen in Fig. 3 that for this value better results are obtained than is the case when nI would be equal to 10 or 30, and the results are equal to those obtained when $nI > 30$, however $nI = 20$ is computationally more efficient.

6 Comparison of optimization methods

After successfully implementing the presented optimization methods, the calibration of the MBL model can be performed. The performance of the optimization methods is evaluated in two steps. First, a known parameter set is used to perform several simulations to which the MBL model will be fitted using the different optimization methods and objective functions. This way, the optimization methods and the objective functions' ability to retrieve a known parameter set will be assessed. Second, the MBL model is fitted to the observations at Uccle for each optimization method and objective function.
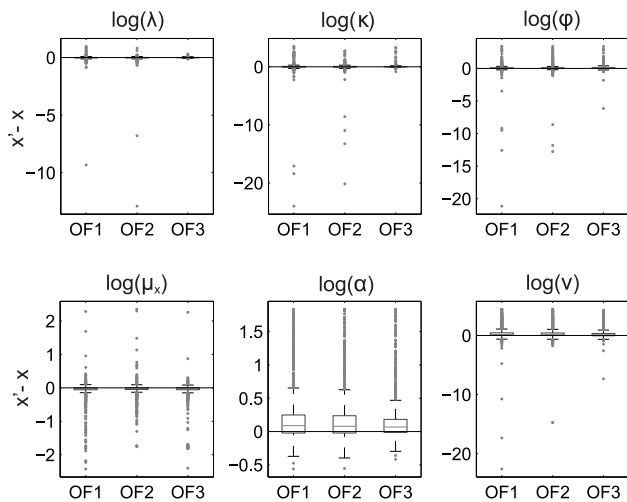
Fig. 5. Distributions of estimation error for each parameter, grouped according to the used optimization method.

Table 3. Descriptive statistics of the performance of different optimization methods for the calibration of the Modified Bartlett-Lewis Rectangular Pulses model.

		minimum	median	StDev	duration (min)
DSM	OF1	0.0300	0.0467	0.0247	9
	OF2	0.0594	0.0922	0.0580	15
	OF3	0.1908	0.2063	0.0524	9
SIMPSA	OF1	0.0300	0.0300	0.4696	18
	OF2	0.0594	0.0594	18.2159	100
	OF3	0.1908	0.1908	0.6206	4
PSO	OF1	0.0283	0.0311	0.1531	13
	OF2	0.0594	0.0624	0.8168	31
	OF3	0.1908	0.1908	0.2359	14
SCE-UA	OF1	0.0300	0.0300	0.1150	14
	OF2	0.0594	0.0594	0.3552	24
	OF3	0.1908	0.1908	0.3433	11

6.1 Retrieval of known parameters

To assess the ability of the optimization methods the objective functions to retrieve a known parameter set, a parameter set was taken from Verhoest et al. (1997) to perform a total of 400 simulations with the MBL model. The length of each simulation is equal to 105 yr. Afterwards, the MBL model was fitted to these simulations. For each objective function, and for each optimization method, 400 calibrations were carried out, i.e. only one repetition of the calibration for each simulated data set. Ideally, each of these calibrations should result in the retrieval of the known parameter set. However, since this is very unlikely, the distribution of estimation errors will shed light on the algorithms and objective functions' performance in their attempt to retrieve known parameters.

Figure 4 visualises the distribution of the estimation errors on the six MBL model parameters for the month of January. It can be seen that SIMPSA, PSO, and SCE-UA perform better in identifying the true parameters in comparison with the DSM with multiple starting points. Significant differences between the former optimization methods, or between the objective functions, are not clearly visible. To facilitate this, the calibration results are grouped according to the objective function with which they were obtained, regardless of the used optimization method (see Fig. 5). Similarly, Fig. 6 displays the distribution of the estimation error in function of the used optimization method, regardless of the used objective function.

Figure 5 indicates that the use of OF3 might lead to better identifiability (this is especially visible for α), however differences are very small.

As for the ability of the optimization methods to identify the true parameters, Fig. 6 confirms the DSM's inability to do so. SIMPSA seems to lead to very large estimation errors

on several occasions. PSO seems to be the most consistent in identifying the true parameter, however, its results are comparable to those of SCE-UA, apart from a few outliers.

6.2 Fit to Uccle data

To evaluate the performance of the different optimization methods and objective functions in a realistic situation, 30 repetitions of the calibration are performed for each optimization method, each objective function, and each month. The 30 repetitions vary in that for each of them, the algorithm starts from a different initial situation. For the DSM, PSO, and SCE-UA, the initial population is chosen randomly within the preset parameter boundaries according to a uniform distribution. Similarly, the initial simplex for SIMPSA is created around a randomly chosen point in parameter space, also sampled from a uniform distribution.

In order to compare the performance of the used optimization methods, several approaches can be adopted. Here, we will first summarize the obtained results by a set of descriptive statistics. This will give a first indication of how well certain methods perform compared with the others. Table 3 displays the aforementioned descriptive statistics for the different objective functions, fitted by the different optimization methods. The minimum and median values and the standard deviation (StDev) of the objective function values obtained after 30 repetitions, displayed in Table 3 are calculated by, first, determining those statistics for each month separately and, second, taking the mean over 12 months. The results are displayed in such a way to enhance interpretation without loss of generality. The duration of the calibration is the mean duration of the total calibration procedure, i.e. taking into account all of the 12 months, on a PC with an Intel®Core™i7-2600 CPU at 3.40 GHZ. All software is implemented in Matlab®.
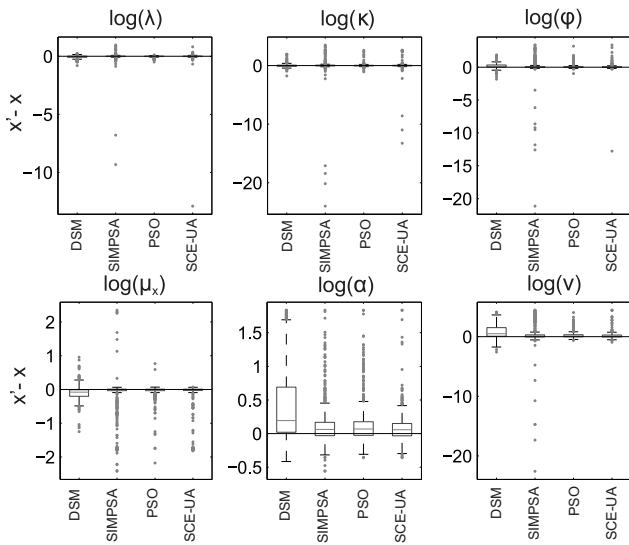
Fig. 6. Box plots comparing calibration results of DSM, SIMPSA, PSO and SCE-UA. All months and repetitions are lumped together, no averages have been taken.

Several observations can be made on the basis of Table 3. First, it seems that the obtained minima for the different objective functions are largely the same. This means that, when the optimization is repeated at least 30 times, each of the optimization methods is able to find the same minima. Further inspection of these minima indicates they result from the same minimizers, i.e. the same parameter combinations are being found by the different optimization methods. This suggests that the parameters are highly identifiable, seeing that the same minima are being found by independent optimization methods. This evidence is corroborated by the fact that the median values, at least for SIMPSA and SCE-UA, are equal to the minima, i.e. the same points are being withheld as the minimum in the majority of the calibration runs. As for DSM and PSO, the median values are fairly close to the minimum. Thus, when it comes to finding a suitable minimum, DSM, SIMPSA, PSO and SCE-UA are almost interchangeable. All four are able to locate the same minima on multiple occasions. Note that PSO is able to find a minimum with a lower objective function value for OF1, in comparison with SIMPSA and SCE-UA. Yet the value reported in Table 3 is the mean value of the minima of the 12 different months. Further investigation of the data uncovers that for the month of April, PSO attained a better solution, however, the differences in the parameter values are rather small, so it is doubtful that it will lead to significant differences in the simulations.
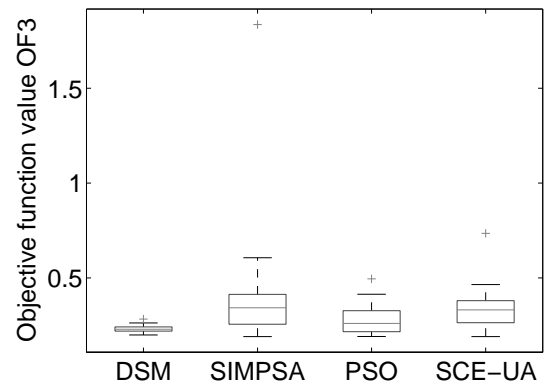
Second, the robustness of each optimization method can be judged in several ways. A first indicator is the standard deviation of the objective function values of the found minima after 30 repetitions. Clearly, DSM is far more robust (i.e. by an order of magnitude) than the other optimization
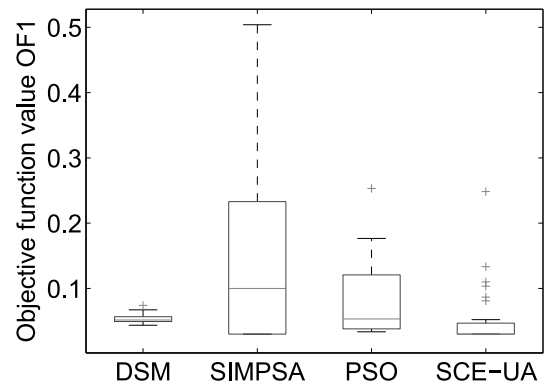
(a) OF1
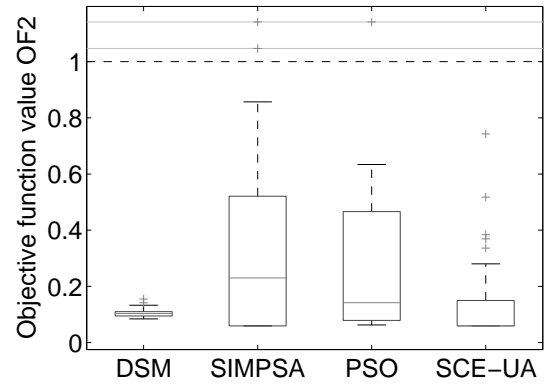
(b) OF2

(c) OF3

Fig. 7. Empirical Cumulative distribution function of the mean 10 min rainfall depth in January as simulated with OF1 vs. observed value.

methods. When comparing the remaining three methods, it can be seen that PSO and SCE-UA have resulted in roughly the same standard deviations, whereas SIMPSA clearly had problems in minimizing OF2. It shows a much higher standard deviation for OF2, and moreover, the average duration

Table 4. Mean ranks of the different optimizaton method's performances.

	DSM	SIMPSA	PSO	SCE-UA
Mean rank	2391	2085	2236	1929

of the calibration is 3 to 4 times higher than for DSM, PSO and SCE-UA, of which DSM proves to be slightly faster in obtaining results. This may lead to the conclusion that SIMPSA is not as flexible as PSO or SCE-UA.

DSM, PSO and SCE-UA also display longer calibration runtimes and a slightly elevated standard deviation for OF2, but not of the same magnitude as SIMPSA. As the former are more flexible, they are more adaptable to changes in the objective function, which of course is an advantage because the cumbersome optimization of the parameters does not have to be repeated in order to obtain satisfactory results.

To further assess the spread of the results obtained by the different optimization methods, a box plot is created. Figure 7a, b and c show comparative box plots for OF1, OF2 and OF3, respectively. To enhance the interpretability of these plots, box plot 7b is clipped. A dotted line marks the limit if any points are outside it. The points outside the limit are plotted in a compression region delineated by two solid lines. The density of the point in the compression region gives an indication of the number of points outside the limit.

For OF1 and OF2, it is clear that DSM and SCE-UA exhibit the best performances, followed by PSO and SIMPSA. Figure 7a and b clearly show that the outcome of a calibration with the DSM or SCE-UA is more robust than with PSO and SIMPSA. However, the DSM is slightly less accurate because the mean of the obtained calibration results does not coincide with the minimum, which is the case for SCE-UA. Conversely, SCE-UA has more outliers, suggesting that it is slightly less robust. Thus, DSM is able to locate near-optimal solutions in a very robust manner, whereas SCE-UA is able to locate more accurate results in a slightly less robust manner. Finally, the performances of PSO and SIMPSA combined with both OF1 and OF2 seem to be more alike, however, objective function values obtained by SIMPSA are spread more than those obtained by PSO, indicating a less robust result.

For OF3 (see Fig. 7c), DSM clearly outcompetes the other optimization methods. It is more accurate, more robust, faster (see Table 3) and more user friendly. Differences between SIMPA, PSO and SCE-UA are less apparent. Subtle differences exist, however. PSO seems to be more robust than SCE-UA and SIMPSA respectively, and again, SIMPSA exhibits the least desirable results (albeit the differences are minute, and therefore may be negligable). Since, based on this last plot, no obvious distinction can be made between PSO and SCE-UA, a more objective measure is needed to determine whether or not there are significant differences in the results.

Table 5. p-values for pairwise Wilcoxon rank sum tests between different optimization methods.

	DSM	SIMPSA	PSO	SCE-UA
DSM	1	1.14×10^{-08}	0.0043	9.20×10^{-18}
SIMPSA		1	0.0059	0.0031
PSO			1	1.08×10^{-08}
SCE-UA				1

In order to compare the performance of the different optimization methods objectively, a Kruskal-Wallis test is performed (Kruskal and Wallis, 1952). Under the null hypothesis, the populations from which the samples are generated have the same median value. If the null hypothesis is rejected it can be concluded that the populations show significant differences. Post-hoc analysis has to be performed to determine which of the methods differ significantly. For this, a pairwise Wilcoxon rank sum test (Gibbons, 1985) is performed. The null hypothesis of the Wilcoxon rank sum test states that the compared samples are independent samples from identical continuous distributions with identical medians (Gibbons, 1985). These statistical tests are performed for all the obtained data, i.e. the results for the different months and the different objective functions are gathered and then tested. This approach is chosen because it allows to define which of the optimization methods displays the overall best performance.

It can be expected that the Kruskal-Wallis test will find that the different optimization methods differ significantly. This assumption is confirmed. A p-value of 1.3×10^{-17} is obtained when conducting the Kruskal-Wallis test. The mean ranks are shown in Table 4. At a 5 % significance level, there is a significant difference between the populations' medians. The mean ranks lead to believe that SCE-UA performs best, followed by SIMPSA, PSO and the DSM. The fact that SIMPSA performs second best and that DSM has the lowest mean rank is quite surprising. Table 3, as well as the box plots lead to believe that both DSM and PSO performed better than SIMPSA. This shows the importance of using multiple evaluation criteria. The standard deviatons (listed in Table 3) and the box plots (Fig. 7) are both seemingly sensitive to outliers. The use of a more robust statistic (mean ranks) unveils this sensitivity.

To determine which of these differences are significant, pairwise Wilcoxon rank sum tests are performed. The resulting p-values are shown in Table 5. Note that, for the interpretation of these p-values, a Bonferroni correction must be applied, i.e. results are significant at an α% significance level if the p-value of the tested hypothesis is smaller than α/n% (with n the number of tested hypotheses). In this case, the null hypothesis of the pairwise Wilcoxon rank sum test can be rejected at a 5 % significance level if the p-value of the hypothesis is smaller than $0.05/6 = 0.0083$. Consequently,

Table 6. Final parameters resulting from calibration with OF1 found by each of the optimization methods, except for the month of April for which the displayed minimum was only found by PSO.

	λ	κ	ϕ	μ_x	α	ν	fitness
Jan	0.0324	0.3810	0.0593	0.9609	3.2957	0.9053	0.0345
Feb	0.0272	0.2973	0.0485	0.9483	3.4148	1.0634	0.0389
Mar	0.0327	0.5920	0.0327	1.4127	2.7201	0.1541	0.0356
Apr	0.0297	0.2918	0.0333	1.6022	3.0000	0.3353	0.0281
May	0.0257	0.1122	0.0296	3.3370	3.1148	0.4208	0.0393
Jun	0.0284	0.2003	0.0326	4.9128	3.0996	0.1899	0.0146
Jul	0.0275	0.0896	0.0268	7.4132	3.3296	0.2781	0.0134
Aug	0.0288	0.1727	0.0294	6.8543	2.8370	0.1282	0.0170
Sep	0.0266	0.1208	0.0226	3.9448	2.7106	0.2233	0.0279
Oct	0.0299	0.3523	0.0323	2.1634	2.4490	0.1775	0.0208
Nov	0.0306	0.2994	0.0490	1.1666	2.8800	0.7642	0.0365
Dec	0.0286	0.2266	0.0448	1.0846	3.1943	1.1887	0.0336

Table 7. Final parameters resulting from calibration with OF2 found by each of the optimization methods.

	λ	κ	ϕ	μ_x	α	ν	fitness
Jan	0.0327	0.3883	0.0582	0.9842	3.2394	0.8253	0.0674
Feb	0.0276	0.2990	0.0476	0.9755	3.3447	0.9702	0.0772
Mar	0.0323	0.6014	0.0324	1.4060	2.7365	0.1547	0.0695
Apr	0.0295	0.2955	0.0336	1.6068	3.0217	0.3402	0.0932
May	0.0257	0.1109	0.0294	3.3918	3.1143	0.4169	0.0792
Jun	0.0282	0.2013	0.0326	4.9003	3.1173	0.1927	0.0288
Jul	0.0273	0.0895	0.0269	7.3857	3.3546	0.2849	0.0264
Aug	0.0286	0.1749	0.0295	6.8458	2.8455	0.1286	0.0339
Sep	0.0265	0.1204	0.0229	3.9312	2.7200	0.2294	0.0562
Oct	0.0297	0.3551	0.0329	2.1488	2.4548	0.1820	0.0411
Nov	0.0309	0.3021	0.0483	1.1964	2.8494	0.7105	0.0724
Dec	0.0289	0.2280	0.0444	1.1090	3.1501	1.1111	0.0670

it can be concluded that, at a 5 % significance level, all the compared optimization methods are found to differ significantly in their median values. So, it is appropriate to conclude that SCE-UA is the best method for the calibration of the MBL model, compared with the other tested methods. The fact that the duration of the calibration is reasonable and quite robust contributes to the validity of this conclusion. According to these statistical tests SIMPSA takes second place and is followed by PSO and DSM. However, from a practical point of view, these results can be disputed. The DSM method clearly shows to be more robust, faster and more user friendly. Besides, it is not clear whether the subtle differences in the calibration results between the different methods would result in significant different modelling outcomes. So, pros and cons must be weighed before making a choice of optimization method, bearing in mind the aforementioned results.

7   Comparison of objective functions

In order to compare the performance of the models, fitted with the respective objective functions, it seems appropriate to incorporate several different performance measures to assess the impact of the configuration of the objective function on various aspects of the fitted model. The parameter sets that provided the best fits in Sect. 6 are used here to enable the comparison. These parameter sets are given in Tables 6 to 8.

The objective function itself, focuses on the expected moments of the rainfall time series and its wet-dry properties. However, certain properties of the generated rainfall time series cannot be expressed by analytical expressions, extreme values, for example, and have to be evaluated through simulation. As a consequence, for each of the three fitted models, an ensemble of 50 simulations is carried out, each of which is 105 yr long. For each moment or property of the rainfall time series, a list of 50 values is obtained, which can each be inter-

preted as the probability distributions of the respective moments or properties when rainfall time series are simulated. To exemplify this, Fig. 8 shows the Empirical Cumulative Distribution Function (ECDF) of the mean 10 min rainfall depth in the month of January. It would be preferable that the observed value at Uccle coincides with the obtained median value, which is obviously not the case. Figure 9 shows the ECDF of the lag-1 autocovariance of the data at an aggregation level of 12 h, for the month of March. This figure shows a much more satisfying fit to the observations. Note that these ECDFs were obtained through simulation using the best obtained parameter set after 30 repetitions of the calibration with OF1. For each of the used objective functions, similar figures for different moments and wet-dry properties could be made, this at different aggregation levels. This would, however, lead to a vast number of figures, rendering comparison impossible. Therefore, the median values of the simulated moments and wet-dry properties are compared with the observed rainfall time series.

Figure 10, for example, shows the mean 10 min rainfall generated by the different fitted models with regard to the observations. It can be seen that none of them provides a perfect fit to the observations. However, OF1 and OF2 seem to largely coincide whereas OF3 tends to deviate more severely form the observed mean values. Similar plots can be made for the other moments of the observed and simulated rainfall time series. These plots can provide a general idea of the orders of magnitude of the deviations from the observations, but they will not allow to distinguish between the suggested objective functions, for their interpretation is far too subjective.

More objective goodness of fit measures of the rainfall time series' moments are given in Table 9. Thus, Table 9 displays the mean percentage error (MPE), mean absolute percentage error (MAPE) and maximum percentage error (MAXPE) of the simulated moments and ZDP in comparison with the observed values. The MPE is particularly useful in uncovering the presence of a significant bias in the model
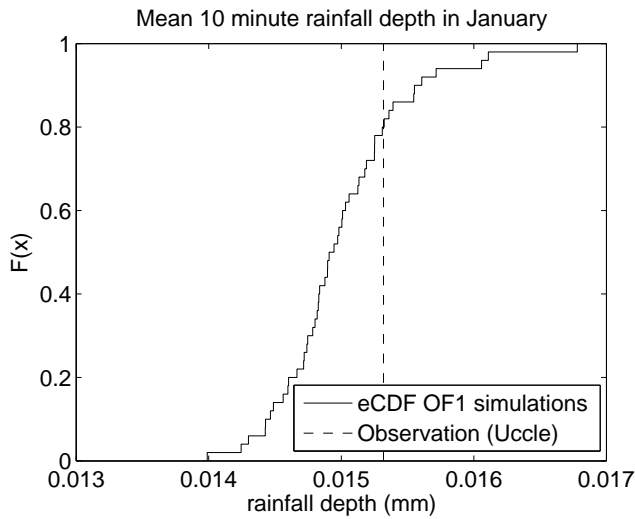
Fig. 8. Empirical Cumulative distribution function of the lag-1 autocovariance of the 12 hourly-rainfall depth in March as simulated with OF1 vs. observed value.
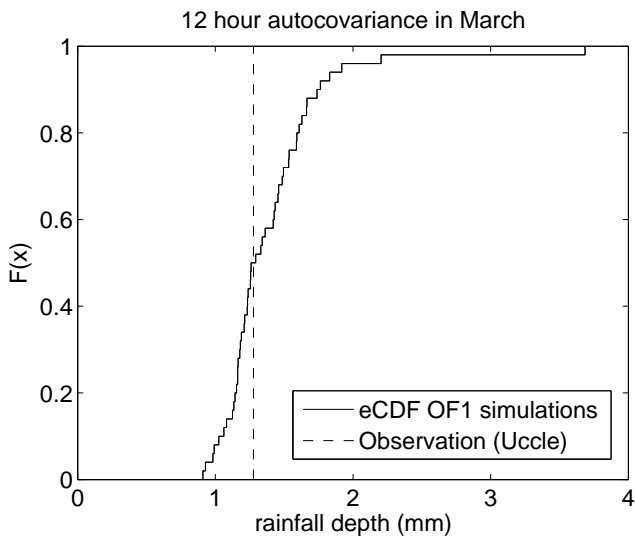

Fig. 9. Mean 10 min rainfall, average simulation result vs. observations.

Table 8. Final parameters resulting from calibration with OF3 found by each of the optimization methods.

	λ	κ	ϕ	μ_x	α	ν	fitness
Jan	0.0314	0.2385	0.0463	1.2099	2.8871	0.6913	0.2979
Feb	0.0284	0.2422	0.0448	1.2191	2.8032	0.6100	0.2523
Mar	0.0271	0.2707	0.0474	1.3024	2.9791	0.6322	0.2471
Apr	0.0267	0.2139	0.0338	1.6379	3.0981	0.4922	0.2482
May	0.0241	0.1455	0.0336	2.7151	3.6000	0.6365	0.0989
Jun	0.0242	0.1495	0.0340	3.0595	4.2021	0.7554	0.1277
Jul	0.0248	0.1133	0.0294	4.3469	4.2751	0.7048	0.3039
Aug	0.0260	0.1292	0.0235	5.1469	3.0247	0.2313	0.1574
Sep	0.0240	0.1356	0.0276	2.9538	3.1375	0.4571	0.0637
Oct	0.0259	0.2061	0.0329	1.9844	2.5324	0.3799	0.0706
Nov	0.0300	0.2140	0.0406	1.4124	2.6273	0.5974	0.2119
Dec	0.0315	0.2254	0.0463	1.3147	2.7612	0.6977	0.2100

Table 9. Model performance measures for the MBL model fitted with different objective functions.

	MPE	MAPE	MAXPE
OF1	−4 %	12 %	−94 %
OF2	−5 %	12 %	−98 %
OF3	−2 %	14 %	−83 %

observed values. The only valid conclusion would be that OF3 shows relatively little bias in comparison with the fitting by the other two objective functions. In order to assess the quality of the overall goodness of fit of the different fits, the MAPE is best used. This shows that both OF1 and OF2, although more biased, show an overall better fit to the observed moments than OF3. The MAXPE, finally, reveals that OF3 results in a lower maximum percentual error, when compared with OF1 and OF2. So, judging by these overall model performance measures, it seems that the use of OF1 and OF2 leads to, more or less, the same results, albeit that OF2 generates a slightly higher bias than the generated rainfall time series fitted with OF1. Taking this information into account, and, looking back at Table 3, which shows that the calibration process tends to be a bit more time consuming when OF2 is used, it is reasonable to conclude that OF1 could be a good objective function for calibrating the MBL model.

Based on these criteria, the results with OF3 are promising, a smaller bias is achieved and the maximum error is smaller. However, the MAPE suggests an overall larger deviation from the observations than when the model is fitted with OF1 or OF2. To further investigate this, Figs. 11 and 12 show the MAPE and MPE, respectively, for a set of moments and the ZDP. Figure 11 clearly reveals the fact that the simulation results obtained by fitting with OF3 almost systematically display a larger deviation from the observations than when the model is fitted with the other configurations of the objective function, which was already concluded on the basis of Table 9. Interestingly however, OF3 seems to be
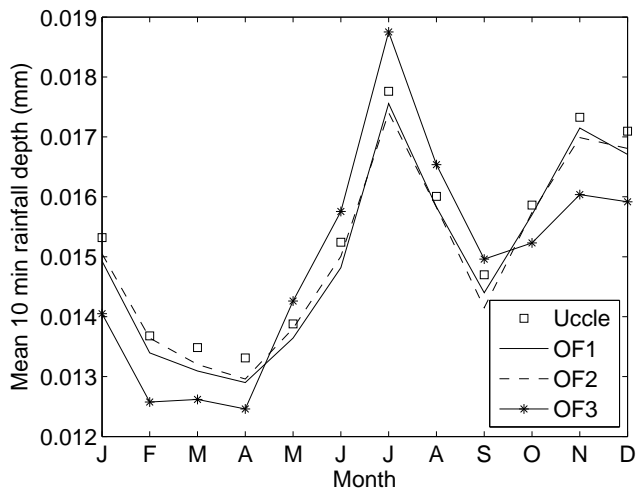
fit. As outlined in Sect. 3, OF2 was especially designed to reduce the overall bias in the fitting of the BL models. Judging from Table 9, this attempt to reduce overall bias in the model fit had an adverse effect. The fitting of the MBL model with OF2 seems to lead to a bigger underestimation of the moments of the rainfall time series (a negative value corresponds to underestimation by model), however it can be argued that the difference is rather negligable.

Judging solely by the MPE, one might find it reasonable to conclude that calibration using OF3 leads to the best overall model fit. The MPE, however, reveals very little about the distribution nor the magnitude of the deviations of the

Fig. 10. Mean Absolute Percentage Error of different fits, averaged over different months and aggregation levels (10 min, 30 min, 1 h, 6 h, 12 h and 24 h).

Fig. 11. Mean Percentage Error of different fits, averaged over different months and aggregation levels (10 min, 30 min, 1 h, 6 h, 12 h and 24 h).

more capable in producing more satisfying wet-dry probabilities (ZDP) in comparison with its contenders. The reason why the overall bias is smaller with OF3 is explained by combining Figs. 11 and 12. It seems the overall deviation of the OF3 is larger, but, more balanced between under- and overestimation, especially for lag-2 and lag-3 autocorrelation and skewness of the rainfall time series. These results make it abundantly clear that a straightforward conclusion in favour of one of three objective function configurations is very hard to make. From a practical point of view, however, it seems reasonable to prefer OF1 over OF2, as was already suggested earlier.

The impact of the used objective functions on the reproduction of extreme rainfall events will not be discussed here. The MBL model suffers from several flaws which need to be resolved first. The MBL model systematically underestimates extreme rainfall events (Verhoest et al., 1997) and occasionally creates unrealistic rainfall cells (Verhoest et al., 2010). The first issue might be resolved by introducing the third order moment into the objective function (Cowpertwait, 1998). The second issue might be resolved by truncating the distribution from which average cell durations are drawn (i.e. a truncated MBL model) (Verhoest et al., 2010). However, these approaches need further investigation and are out of the scope of this paper.

8 Conclusions

In this paper, further refinement of the calibration procedure of Bartlett-Lewis type stochastic point rainfall models was attempted. For this purpose, the Modified Bartlett Lewis (MBL) model was used. These results may however be applicable to other variants of the Bartlett-Lewis model
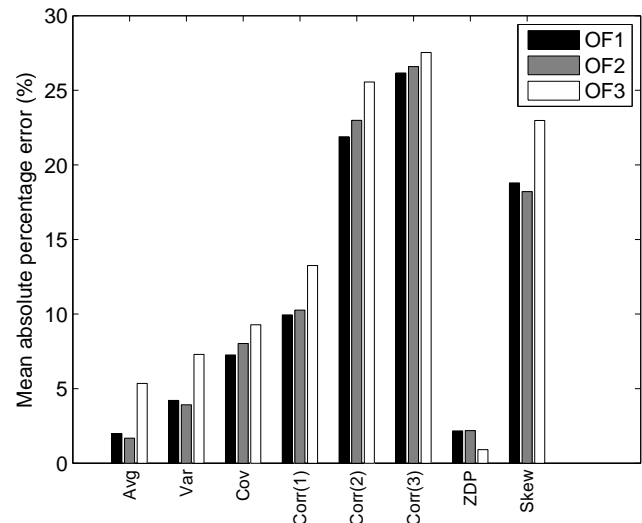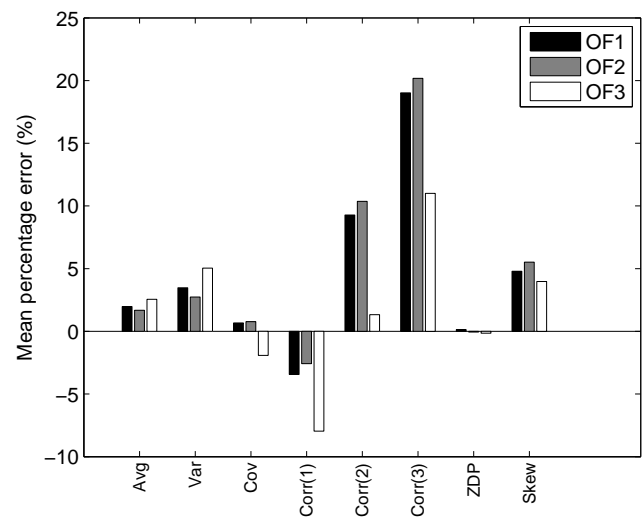

Fig. 12. Mean Percentage Error of different fits, averaged over different months and aggregation levels (10 min, 30 min, 1 h, 6 h, 12 h and 24 h).

and the Neyman-scott type models, due to strong similarities between them.

A first issue that was addressed is the fitting procedure itself. The fitting procedure is characterized by the presence of multiple local minima, making it hard for conventional "local search" methods to reach satisfactory results in a robust manner. Therefore, four different optimization methods were tested and compared with each other. To ensure a fair comparison, the parameters of the respective methods were first fine-tuned for this specific minimization problem, after which the actual calibration of the MBL model

was performed. The performance was then judged by the accuracy, robustness and time consumption of the optimization methods. The following algorithms were used: the DSM by Nelder and Mead (1965), Simplex-Simulated Annealing (SIMPSA) (Cardoso et al., 1996), Particle Swarm Optimization (PSO) (Shi and Eberhart, 1998), and Shuffled Complex Evolution (SCE-UA) (Duan et al., 1994).

Secondly, the choice of weights in the objective function is addressed. In many empirical studies, the choice of the weights of the used properties in the objective function is made rather subjectively. As different approaches exist, the impact of the approaches on the results after calibration is therefore investigated. Three objective functions were used, one classically used objective function (OF1), according to Verhoest et al. (1997), an implicitly weighed objective function (OF2) (Cowpertwait et al., 2007) and one weighted by the empirical variances of the included properties (OF3) (Chandler, 2004). Simulated moments obtained after fitting with each of these three objective functions were compared with each other.

The fine-tuning of the parameters of the optimization methods generally revealed that different parameter combinations lead to the same results. Therefore, in future research, when calibrating BL type models, one can rely on the feasible parameter ranges presented in this paper. When however, one would attempt to repeat this search for a different optimization problem, it should be noted that the search for the optimal parameters of the optimization methods requires most efforts for PSO. For SCE-UA and SIMPSA, the exhaustive parameter search is less time consuming and more straightforward.

Analysis of the optimization methods' abilities to retrieve known parameters reveals that the estimation errors obtained by PSO, SIMPSA, and SCE-UA are comparable. Estimation errors obtained by DSM are, on the other hand, systematically larger. Similarly, the analysis is used to determine whether the different objective functions lead to higher identifiability of the parameters. However, differences in the objective functions' abilities to identify a known parameter set are negligable and thus not convincing to make a discrimination between them.

Descriptive statistics of repeated calibration runs show that all of the tested optimization methods were able to find the same minima. The final result could thus not be used as a basis for discrimination. Judging by the median objective function value and the standard deviation, the robustness of the methods was assessed. SIMPSA lacked in robustness for OF2, which is also reflected in the poor computational efficiency. This leads to the conclusion that SIMPSA might not be the best method to use for the calibration of the MBL model. On the other hand, DSM exhibited much more robust behaviour in comparison with SCE-UA and PSO. To distinguish between DSM, PSO and SCE-UA, according to the accuracy of the performance, a pairwise Wilcoxon rank sum test demonstrated that the calibration results obtained by each of the three methods differ significantly. SCE-UA is most accurate, followed by PSO and DSM, however, DSM is faster, more robust and more practical in use. Thus no straightforward conclusion can be made based on these results. However, the choice between them will depend on the application's desired accuracy, expertise of the user, and the timespan in which results have to be obtained.

The results of the comparison of the three tested alternative objective function configurations are less clear-cut. In summary, it can be concluded that the use of OF3 is not to be encouraged, as it leads to an overall larger deviation from the observations, except for the ZDP, than when OF1 or OF2 are used. Both OF1 and OF2 perform equally well, however, it could be noticed that the calibration with OF2 tended to be more computationally intensive, which is of course an argument in favour of the use of OF1.

References

Aarts, E. H. L. and Van Laarhoven, P. J. M.: Statistical cooling – A general-approach to combinatorial optimization problems, Philips J. Res., 40, 193–226, 1985.

Aarts, E. H. L. and Van Laarhoven, P. J. M.: Simulated Annealing: Theory and Applications, D. Reidel, Dordrecht, 1987.

Boughton, W. and Droop, O.: Continuous simulation for design flood estimation – a review, Environ. Modell. Softw., 18, 309–318, 2003.

Box, M. J.: A new method of constrained optimization and a comparison with other methods, Comput. J., 8, 42–52, 1965.

Burton, A., Kilsby, C., Fowler, H., Cowpertwait, P., and O'Connell, P.: RainSim: A spatialtemporal stochastic rainfall modelling system, Environ. Modell. Softw., 23, 1356–1369, 2008.

Cameron, D., Beven, K., and Tawn, J.: An evaluation of three stochastic rainfall models, J. Hydrol., 228, 130–149, 2000.

Cameron, D., Beven, K., and Tawn, J.: Modelling extreme rainfall using a modified random pulse Bartlett-Lewis stochastic rainfall model (with uncertainty), Adv. Water Resour., 24, 203–211, 2001.

Cardoso, M. F., Salcedo, R. L., and De Azevedo, S. F.: The Simplex-Simulated Annealing approach to continuous nonlinear optimization, Comput. Chem. Eng., 20, 1065–1080, 1996.

Chandler, R.: Moment-based inference for stochastic-mechanistic models, internal report no. 7, DEFRA project: Improved methods for national spatial-temporal rainfall and evaporation modelling for BSM, 2004.

Cowpertwait, P. S. P.: Further developments of the Neyman-Scott clustered point process for modeling rainfall, Water Resour. Res., 27, 1431–1438, 1991.

Cowpertwait, P. S. P.: A generalized point process model for rainfall, P. Roy. Soc. Lond. A Mat., 447, 23–37, 1994.

Cowpertwait, P. S. P.: A Poisson-cluster model of rainfall : high-order moments and extreme values, P. Roy. Soc. Lond. A Mat., 454, 885–898, 1998.

Cowpertwait, P. S. P.: Mixed rectangular pulses models of rainfall, Hydrol. Earth Syst. Sci., 8, 993–1000, doi:10.5194/hess-8-993-2004, 2004.

Cowpertwait, P. S. P., Isham, V., and Onof, C.: Point process models of rainfall: developments for fine-scale structure, P. Roy. Soc. Lond. A Mat., 463, 2569–2587, 2007.

De Jongh, I. L. M., Verhoest, N. E. C., and De Troch, F. P.: Analysis of a 105-year time series of precipitation observed at Uccle, Belgium, Int. J. Climatol., 26, 2023–2039, 2006.

Duan, Q., Sorooshian, S., and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, J. Hydrol., 158, 265–284, 1994.

Engelbrecht, A.: Fundamentals of Computational Swarm Intelligence, John Wiley, London, 2006.

Entekhabi, D., Rodriguez-Iturbe, I., and Eagleson, P. S.: Probabilistic representation of the temporal rainfall process by a Modified Neyman-Scott Rectangular Pulses Model: Parameter estimation and validation, Water Resour. Res., 25, 295–302, 1989.

Gibbons, J. D.: Nonparametric statistical inference, Marcel Dekker, New York, 1985.

Gyasi-Agyei, Y. and Willgoose, G.: Generalisation of a hybrid model for point rainfall, J. Hydrol., 219, 218–224, 1999.

Hansen, L. P.: Large properties of generalized method of moments estimators, Econometrica, 50, 1029–1054, as cited by Kaczmarska (2011), 1982.

Hartig, F., Calabrese, J. M., Reineking, B., Wiegand, T., and Huth, A.: Statistical inference for stochastic simulation models - theory and application, Ecol. Lett., 14, 816–827, 2011.

Heneker, T. M., Lambert, M. F., and Kuczera, G.: A point rainfall model for risk-based design, J. Hydrol., 247, 54–71, 2001.

Jiang, M., Luo, Y. P., and Yang, S. Y.: Stochastic convergence analysis and parameter selection of the standard Particle Swarm Optimization algorithm, Inform. Process. Lett., 102, 8–16, doi:10.1016/j.ipl.2006.10.005, 2007.

Kaczmarska, J.: Further development of Bartlett-Lewis models for fine-resolution rainfall, Tech. Rep., Department of Statistical Science, University College London, 2011.

Kavvas, M. L. and Delleur, J. W.: A stochastic cluster model of daily rainfall sequences, Water Resour. Res., 17, 1151–1160, 1981.

Kennedy, J. and Eberhart, R.: Particle swarm optimization, in: Proceedings of the IEEE International conference on neural networks, IV, 1942–1948, Piscataway, NJ: IEEE Service Center, 1995.

Khaliq, M. N. and Cunnane, C.: Modelling point rainfall occurences with the Modified Bartlett-Lewis Rectangular Pulses Model, J. Hydrol., 180, 109–138, 1996.

Kirkpatrick, S.: Optimization by simulated annealing – Quantitative studies, J. Stat. Phys., 34, 975–986, 1984.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P.: Optimization by Simulated Annealing, Science, 220, 671–680, 1983.

Kruskal, W. H. and Wallis, W. A.: Use of ranks in one-criterion variance analysis, J. Am. Stat. Assoc., 47, 584–621, 1952.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E.: Equation of state calculations by fast computing machines, J. Chem. Phys., 21, 1087–1092, 1953.

Nelder, J. A. and Mead, R.: A simplex-method for function minimization, Comput. J., 7, 308–313, 1965.

Obeysekera, J. T. B., Tabios III, G. Q., and Salas, J. D.: On parameter estimation of temporal rainfall models, Water Resour. Res., 23, 1837–1850, 1987.

Onof, C. and Wheater, H. S.: Modelling of British rainfall using a random parameter Bartlett-Lewis Rectangular Pulse Model, J. Hydrol., 149, 67–95, 1993.

Onof, C. and Wheater, H. S.: Improved fitting of the Bartlett-Lewis Rectangular Pulse Model for hourly rainfall, Hydrolog. Sci., 39, 663–680, 1994a.

Onof, C. and Wheater, H. S.: Improvements to the modeling of British rainfall using a modified random parameter Bartlett-Lewis rectangular pulse model, J. Hydrol., 157, 177–195, 1994b.

Press, W. H. and Teukolsky, S. A.: Simulated Annealing over continuous spaces, Computers in Physics, 5, 426, 1991.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical Recipes - The Art of Scientific Computing, Cambridge University Press, New York, 1986.

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V.: Some models for rainfall based on stochastic point processes, P. Roy. Soc. Lond. A Mat., 410, 269–288, 1987a.

Rodriguez-Iturbe, I., Febres de Power, B., and Valdés, J. B.: Rectangular pulses point process models for rainfall: analysis of empirical data, J. Geophys. Res., 92, 9645–9656, 1987b.

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V.: A point process model for rainfall: further developments, P. Roy. Soc. Lond. A Mat., 417, 283–298, 1988.

Scheerlinck, K., Pauwels, V. R. N., Vernieuwe, H., and De Baets, B.: Calibration of a water and energy balance model: Recursive parameter estimation versus particle swarm optimization, Water Resour. Res., 45, W10422, doi:10.1029/2009WR008051, 2009.

Shi, Y. and Eberhart, R.: A modified particle swarm optimizer, in: Proceedings of the IEEE International Conference on Evolutionary Computation, 69–73, Piscataway, NJ: IEEE Press, 1998.

Smithers, J. C., Pegram, G. G. S., and Schulze, R. E.: Design rainfall in South Africa using Bartlett-Lewis rectangular pulse rainfall models, J. Hydrol., 258, 83–99, 2002.

Spendley, W., Hext, G. R., and Himsworth, F. R.: Sequential Application of Simplex Designs in Optimisation and Evolutionary Operation, Technometrics, 4, 441–461, 1962.

Trelea, I.: The Particle Swarm Optimization algorithm: convergence analysis and parameter selection, Inform. Process. Lett., 85, 317–325, 2003.

Vandenberghe, S., Verhoest, N. E. C., Onof, C., and De Baets, B.: A comparative copula-based bivariate frequency analysis of observed and simulated storm events: A case study on Bartlett-Lewis modeled rainfall, Water Resour. Res., 47, W07529, doi:10.1029/2009WR008388, 2011.

Velghe, P. A., Troch, P. A., De Troch, F. P., and Van de Velde, J.: Evaluation of cluster-based rectangular pulses point process models for rainfall, Water Resour. Res., 30, 2847–2857, 1994.

Verhoest, N. E. C., Troch, P. A., and De Troch, F. P.: On the applicability of Bartlett-Lewis rectangular pulses models in the modeling of design storms at a point, J. Hydrol., 202, 108–120, 1997.

Verhoest, N. E. C., Vandenberghe, S., Cabus, P., Onof, C., Meca-Figueras, T., and Jameleddine, S.: Are stochastic point rainfall models able to preserve extreme flood statistics?, Hydrol. Process., 24, 3439–3445, 2010.

Waymire, E. and Gupta, V.: The mathematical structure of rainfall representations. 1. A review of stochastic rainfall models, Water Resour. Res., 17, 1261–1272, 1981.

Wheater, H. S., Isham, V. S., Chandler, R. E., Onof, C. J., and Stewart, E. J.: Improved methods for national spatial-temporal rainfall and evaporation modelling for BSM, Tech. Rep. F2105/TR, Department for Environment, Food and Rural Affairs, London, 2006.

Wheater, H. S., Chandler, R. E., J., O. C., Isham, V. S., Bellone, E., Yang, C., Lekkas, D., Lourmas, G., and Segond, M.-L.: Spatial-temporal rainfall modelling for flood risk estimation, Stoch. Env. Res. Risk A., 19, 403–416, 2005.