CompAMa
www.compama.co.usb.ve
Bull. Comput. Appl. Math., Vol.3, No.2, 2015
ISSN 2244-8659

A deterministic optimization approach for solving the rainfall disaggregation problem

Debora Cores *, *Lelys Guenni* †, *Lisbeth Torres* ‡

CompAMa Vol.3, No.2, pp.7-29, 2015 - Accepted October 3, 2015

Abstract

One of the main problems in hydrology is the time scale of the historical rainfall data, available from many meteorological data bases. Most of the rainfall data is given at a time scale coarser than the one needed for many applications in hydrology and environmental sciences, as the estimation of spatially continuous rainfall at finer time scales, for drainage systems design and extreme rainfall analysis. A method to disaggregate monthly rainfall to daily or finer temporal scale is very important in many applications. Many authors have addressed this problem by using some stochastic methods including several stochastic rainfall models. The lowering resolution methods must be low-cost and low-storage since the amount of rainfall data is large. The purpose of this work is to formulate this problem as a constrained optimization problem and solve it with a low-cost and low-storage deterministic optimization method. We modify the objective function proposed by Guenni and Bárdossy for solving the disaggregation rainfall problem and we use the low-cost spectral projected gradient (SPG) method.

*Departamento de Cómputo Científico y Estadística, and Centro de Estadística y Software Matemático (CEsMA), Universidad Simón Bolívar, Apartado 89000, Caracas 1080-A, Venezuela (cores@usb.ve).

†Departamento de Cómputo Científico y Estadística, and Centro de Estadística y Software Matemático (CEsMA), Universidad Simón Bolívar, Apartado 89000, Caracas 1080-A, Venezuela (lbravo@cesma.usb.ve).

‡Departamento de Cómputo Científico y Estadística, and Centro de Estadística y Software Matemático (CEsMA), Universidad Simón Bolívar, Apartado 89000, Caracas 1080-A, Venezuela (ltorres@gmail.com).

In contrast with the stochastic method, a deterministic approach will take into account important information, as for example the gradient of the objective function. The proposed method was applied to a data set from a rainfall network of the central plains of Venezuela, in which rainfall is highly seasonal and data availability at a daily time scale or even higher temporal resolution is very limited. The numerical results show that the SPG method for solving the disaggregation rainfall problem avoids daily precipitations outliers that might occur as an artifact of the simulation procedure and accurately reproduces the probability distribution. Also, the proposed model and methodology outperforms the one proposed by Guenni and Bárdossy (2002) in the sense that it reduces the absolute error value for the statistical properties from the observed data.

Keywords: Rainfall disaggregation, spectral projected gradient, simulated annealing.

1 Introduction

In hydrology, there are many physical processes that should be monitored at time scales of days, hours or even minutes. Rainfall-runoff processes, landslides movements caused by extreme rainfall intensities, flash floods occurring in urban areas due to rainfall episodes of short time duration, are just some examples. The key study variable for all these processes is rainfall, which needs to be collected at high temporal and spatial resolutions in order to understand the physical dynamics of such processes.

A large number of studies are available to investigate the possibility of transforming rainfall data from one temporal scale to another. In particular to go from coarse time scales to more detailed ones there are the pioneer works of Valencia and Schaake ([1], [2]). An extensive review of methodologies for rainfall disaggregation are presented by Koutsoyiannis in [3]. In this review disaggregation methods based on stochastic rainfall modeling are discussed going from methods that assume normality of the high resolution data to generalizations for the non-normal case. Methods to represent an aggregate amount into individual storms using point process models are discussed by Glasbey et al. in [4], Cowperwait et al. in [5] and Koutsoyanis and Onof in [6]. Another approach based on artificial neural networks has been proposed by Burian et al. [7] to disaggregate hourly rainfall into shorter time intervals. Guenni and Bárdossy in [8] proposed a simulated annealing approach applied to a synthetically generated daily rainfall series from a truncated normal model. This method gets daily rainfall samples by minimizing an objective function preserving different statistical characteristics of the high resolution data including scaling properties at different time scales.

All these approaches have in common the stochastic nature of the analysis in the sense that high resolution data are considered as realizations of a stochastic process. Sivakumar et al. in [9] and [10] introduced a non-stochastic approach based on the notion of deterministic chaos. They use a chaotic model to disaggregate rainfall by using six successively time resolutions. Their study suggests that a chaotic framework might be more suitable to model rainfall dynamics than a purely stochastic process.

Following the idea that a non stochastic approach might also be adequate for representing rainfall dynamics at high temporal resolutions, we proposed a methodology based on an extension of the Spectral Gradient (SG) method [11], that allows to restrict an optimization problem over any convex set, let say Ω. This technique developed by Birgin et al. in [12], is known as the Spectral Projected Gradient (SPG) method and utilizes the projection over Ω of the spectral gradient direction. The spectral choice of the step length makes this approach a non-monotone descent scheme. This methodology is a low-cost and low-storage deterministic approach that only requires first order information and converges to a stationary point of the problem from any initial guess, even if it is not close to the solution. For this particular application, we propose to solve an optimization problem where the objective function is a slight perturbation of the one proposed by Guenni and Bárdossy [8] in order to force some desirable statistical properties in the objective function. We also restrict the problem in such a way that daily precipitations will be bounded by a minimum and maximum value, that could be obtained from the historical daily rainfall or any prior information. The numerical results obtained with the proposed methodology illustrate that the resulting daily series follow the behaviour of the historical series keeping the extreme outliers on their bounds. Moreover, the statistical measures for the computed daily series were better approximated using the proposed optimization problem and technique than using the Guenni and Bárdossy model and strategy [8].

On the other hand, the SPG method requires fewer function evaluations to get a solution when compared to the stochastic approaches (particulary with the stochastic methodology proposed in [8]).

2 Formulation of the Disaggregation Problem

In this section we formulate the optimization problem to disaggregate monthly rainfall to daily precipitation. Suppose that C is the accumulated rainfall value for a particular month of a year for which the daily rainfall series is unknown. Also assume that daily rainfall data corresponding to the same month and different years (historic or observed rainfall series) are given. The disaggregation problem consists in estimating the unknown daily rainfall for this particular month of a year such that some desired statistical properties are satisfied. In particular, we want to preserve the autocorrelation function, the indicator autocorrelation function, some scaling properties, the dry days proportion and the accumulated monthly rainfall C, between the historic daily rainfall (HDR) series and the estimated daily rainfall (EDR) series.

In order to write explicitly the problem we need to introduce some notation. For a fixed given month j, let n_j to denote the number of days for month j, and $z^*_{i,k}$ denote the amount of precipitation on day k and year i, for the fixed j month. So, the HDR sequence for the j month containing s years of information, can be written as the vector:

$$\begin{aligned} Z^* &= (z^*_{1,1}, z^*_{1,2}, \dots, z^*_{1,n_j}, z^*_{2,1}, \dots, z^*_{2,n_j}, \dots, z^*_{s,1}, \dots, z^*_{s,n_j})^T \\ &:= (z^*_1, z^*_2, \dots, z^*_p)^T \in \mathbb{R}^p \end{aligned}$$

where $p = s * n_j$ is the total number of days of the HDR sequence. We want to estimate the daily precipitation for month j, corresponding to a year that it is not in the HDR series. For the month j, of each year $j = 1, \dots, s$, we let z_i be the estimated amount of precipitation on day i, then the EDR series for the month j is given by

$$Z = (z_1, z_2, \dots, z_{n_j})^T \in \mathbb{R}^{n_j}.$$

The complete EDR series contains the daily rainfall for the j month of each year $j = 1, \dots, s$. Therefore, we solve s restricted optimization problems, one for each year. For each year, $j = 1, \dots, s$, we propose to preserve some statistical properties that define the objective function. In order to preserve the autocorrelation function between the HDR series and the EDR series, we define:

$$\Theta_1(Z) = \sum_{k=1}^{K} (\rho^*_k(Z^*) - \rho_k(Z))^2, \tag{1}$$

where

$$\begin{aligned}
\rho_k^*(Z^*) &= \frac{\sum_{t=1}^{p-k}(z_t^* - \bar{z}^*)(z_{t+k}^* - \bar{z}^*)}{\sum_{t=1}^{p-k}(z_t^* - \bar{z}^*)^2}, \quad \text{(observed autocorrelation function)},\\
\rho_k(Z) &= \frac{\sum_{t=1}^{n_j-k}(z_t - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=1}^{n_j-k}(z_t - \bar{z})^2}, \quad \text{(estimated autocorrelation function)},\\
\bar{z}^*(Z^*) &= \frac{\sum_{t=1}^{p} z_t^*}{p},\\
\bar{z}(Z) &= \frac{\sum_{t=1}^{n_j} z_t}{n_j},
\end{aligned}$$

and k is the lag number.

On the other hand, we also want to minimize the squared difference between the indicator autocorrelation function for the HDR series and the EDR series. So, let

$$\Theta_2(Z) = \sum_{k=1}^{K}(\rho_{I_k}^*(Z^*) - \rho_{I_k}(Z))^2, \tag{2}$$

K is the total number of lags being considered. where

$$\begin{aligned}
\rho_{I_k}^*(Z^*) &= \frac{\sum_{t=1}^{p-K}(I(z_t^*) - I(\bar{z}^*))(I(z_{t+K}^*) - I(\bar{z}^*))}{\sum_{t=1}^{p-K}(I(z_t^*) - I(\bar{z}^*))^2},\\
&\quad \text{(observed indicator autocorrelation function)}\\
\rho_{I_k}(Z) &= \frac{\sum_{t=1}^{n_j-K}(I(z_t) - I(\bar{z}))(I(z_{t+K}) - I(\bar{z}))}{\sum_{t=1}^{n_j-K}(I(z_t) - I(\bar{z}))^2}, \quad \text{and}\\
&\quad \text{(estimated indicator autocorrelation function)}
\end{aligned}$$

$$I(z_t) = \begin{cases} 1 & if \;\; z_t > \delta \\ 0 & if \;\; z_t \leq \delta. \end{cases} \tag{3}$$

In equation (3), the parameter δ means the rainfall threshold representing a small rainfall amount.

To maintain some scaling properties of the HDR series we consider

$$\Theta_3(Z) = \sum_{l=1}^{L}(\Psi_l^*(Z^*) - \Psi_l(Z))^2, \tag{4}$$

where $\Psi_l(Z)$ is the exponent of the Burlando and Rosso multiple scaling relationship [13], which is given by:

$$E(\hat{z}^l(\lambda h)) = \lambda^{\Psi_l} E(\hat{z}^l(h)), \tag{5}$$

where $\hat{z}(\lambda h)$ is the accumulated rainfall at duration λh of the series Z, E is the expected value and l is the order of the moment.

It is important to stress out that in the work by Guenni and Bárdossy in [8], the authors consider as the objective function of the disaggregated problem, a linear combination of the functions Θ_1, Θ_2 and Θ_3. That is,

$$\begin{array}{ll} \text{Minimize} & \gamma_1\Theta_1(Z) + \gamma_2\Theta_2(Z) + \gamma_3\Theta_3(Z) \\ Z \in \mathbb{R}^{n_j} & \end{array} \tag{6}$$

where the $\gamma_i' s$ are arbitrary positive weights. The authors solved problem (6) using simulating annealing and the value of $\frac{1}{3}$ for the weights γ_i for $i = 1, 2, 3$. The initial iterate in Guenni and Bárdossy in [8] is generated by simulating from a truncated normal model. Therefore, it is a plausible realization of the rainfall process. Simulated series from this model would preserve some of the statistical properties of the observed rainfall series as mean, variance and probability of dry days. The simulating annealing algorithm would enhance the simulation process in the sense that it would reshuffle the simulated rainfall sequences in order to preserve additional statistical characteristics which are included in the objective function. However, a deterministic optimization method for solving problem (6) starting with the same initial iterate would not have to preserve the same desired statistic properties of the initial guess during the process. The only way to guarantee it, is by considering those desirable properties as constraints of the optimization problem or by including them in the objective function. Since the purpose of this work is to solve the disaggregation problem by a deterministic low-cost optimization method, we need to redefine the Guenni and Bárdossy optimization problem by adding up additional constraints.

To preserve the dry days proportion we consider:

$$\Theta_4(Z) = (P^*(Z^*)) - P(Z))^2, \tag{7}$$

where $P^*(Z^*) = \sum_{t=1}^{p} \hat{I}(z_t^*)$ is the amount of dry days of the HDR series,

$P(Z) = \sum_{t=1}^{n_j} \hat{I}(z_t)$ is the amount of dry days of the EDR series where,

$$\hat{I}(z_t) = \begin{cases} 1 & if \;\; z_t = 0 \\ 0 & if \;\; z_t > 0 \end{cases} . \tag{8}$$

To maintain the given accumulated monthly rainfall C in the EDR series, we define:

$$\Theta_5(Z) = (C - \sum_{t=1}^{n_j} z_t)^2. \tag{9}$$

To adjust the mean ($\bar{z}^*$) of the observed series (Z^*) we define:

$$\Theta_6(Z) = (\bar{z}^* - \bar{z})^2. \tag{10}$$

Notice that functions Θ_1 and Θ_2 are discontinuous at the points Z such that $z_t = \bar{z}$ or $I(z_t) = I(\bar{z})$, for all $t = 1, \ldots, n_j - k$ respectively. These points do not belong to the solution space, they do not represent the real behavior of the rainfall and they are not samples from its probability distribution. In order to avoid those discontinuities, different lower and upper bounds for each z_t are considered as restrictions to our problem. The lower bounds are slightly different for each z_t but close to lowest observed precipitation. Similar considerations are taken for the upper bounds.

Finally, a disaggregated optimization problem that preserve the autocorrelation function, the indicator autocorrelation function, the Burlando and Rosso scaling properties [13], the dry days proportion, the accumulated monthly rainfall C and the mean between the HDR series and the EDR series is written as

$$\begin{array}{ll} \text{Minimize} & \Theta(Z) = \sum_{m=1}^{6} \Theta_m(Z) \\ \text{s.t.} & Z \in \Omega = \{Z \in \mathbb{R}^{n_j} : l_t \leq z_t \leq u_t \quad t = 1, \ldots, n_j\}, \end{array} \tag{11}$$

where the values of l_t and u_t, for $t = 1, \ldots, n_j$ will bound the extreme values of the EDR series and will also avoid discontinuities of the objective function. The values of l_t and u_t, for $t = 1, \ldots, n_j$ could be given by the user or they can be obtained from the HDR series or other reference values.

3 Application of the Spectral Projected Gradient Method

Let Ω be a convex set. The Spectral Projected Gradient (SPG) method applied to any convex constrained optimization problem of the form:

$$\begin{array}{cc} \text{Minimize} & \Theta(Z) \\ \text{s.t.} & Z \in \Omega \end{array} \tag{12}$$

is a developed iterative technique, based on the Spectral Gradient (SG) method that allows to minimize a nonlinear function over a convex set, [12]. The SG method utilizes the negative gradient direction (minimum descent direction) with a particular choice of the steplength, named spectral length (for more details on this see [11]). Therefore, the SPG method finds a solution of problem (12) by using the projection of the spectral gradient step over the convex set as any projection gradient method. That is:

$$Z_{n+1} = Z_n + \lambda d_n \tag{13}$$

where $d_n = P_\Omega(Z_n - \alpha_n \nabla\Theta(Z_n)) - Z_n$ and P_Ω is the projection on the convex set Ω. The parameter α_n is the spectral steplength, which is given by

$$\alpha_n = \frac{s_{n-1}^T s_{n-1}}{s_{n-1}^T y_{n-1}}, \tag{14}$$

where $s_{n-1} = Z_n - Z_{n-1}$, $y_{n-1} = \nabla\Theta(Z_n) - \nabla\Theta(Z_{n-1})$ and λ is chosen to guarantee convergence from any initial guess, by satisfying the following non-monotone condition:

$$\Theta(Z_{n+1}) \leq (\max_{0 \leq j \leq \min\{n, M-1\}} \Theta(Z_{n-j})) + \gamma\lambda\nabla\Theta(Z_n)^T d_n, \tag{15}$$

where the parameter M used in the globalization strategy of the SPG method guarantees descent of the objective function at least every M iterations and convergence to a stationary solution from any initial guess. The SPG method does not guarantee descent of the objective function at every iteration in contrast with the classic Projected Gradient method because of the spectral steplength (14) and the non-monotone condition (15) (see [14]). In [12], the authors show that for many examples or applications, this non-monotone behavior of the objective function makes the method very competitive and sometimes preferable than other optimization techniques. Moreover, in [11], they compare the results obtained for many problems with the (SG) method, with the ones obtained with the Polak-Ribiere implementation of the conjugate gradient method (PR^+), by Gilbert and Nocedal [15] and CONMIN by Shanno and Phua [16]. The line search strategy for the PR^+ method is based on the algorithm of Moré and Thuente [17]. They conclude that the SG method out performs CONMIN and PR^+ in number of gradient evaluations and CPU time for most of the problems tested. Also, in [12], the authors claim that in the SPG method, the spectral choice of the steplength is the essential feature that puts the efficiency in the projected gradient methodology. The SPG method is a low-cost and low-storage technique since few floating point operations and only first order information are required. On the other hand, the non-monotone condition (15) allows the objective function to increase at some iterations, which implies less function evaluations than the Armijo type condition used frequently in the classical projected gradient methods or gradient type methods (see [18]).

The set of restrictions for problem (11) is

$$\Omega = \{Z \in \mathbb{R}^{n_j} \text{ such that } l_t \le z_t \le u_t \text{ for } t = 1, \ldots, n_j\}.$$

It is clear that Ω is a convex set. Moreover, the projection over Ω (P_Ω) requires a marginal additional computational cost, and is given by:

$$P_\Omega = \begin{cases} l_t & \text{if } z_t < l_t \\ u_t & \text{if } z_t > u_t \\ z_t & \text{if } l_t \le z_t \le u_t \end{cases} \tag{16}$$

The lower and upper bounds l_t and u_t, for $t = 1, \ldots, n_j$, can be obtained from the HDR series or as a priory information based on physical considerations. We expect that the EDR sequence obtained by solving problem (11) will not have extreme peaks.

In this work, the gradient vector of the objective function is:

$$\nabla\Theta(Z) = \nabla\Theta_1(Z)+\nabla\Theta_2(Z)+\nabla\Theta_3(Z)+\nabla\Theta_4(Z)+\nabla\Theta_5(Z)+\nabla\Theta_6(Z) \tag{17}$$

and

$$
\begin{aligned}
\nabla\Theta_1(Z) &= -\sum_{k=1}^{K} 2(\rho_k^*(Z^*) - \rho_k(Z))\nabla\rho_k(Z),\\
\nabla\Theta_2(Z) &= -\sum_{k=1}^{K} 2(\rho_{I_k}^*(Z^*) - \rho_{I_k}(Z))\nabla\rho_{I_k}(Z),\\
\nabla\Theta_3(Z) &= -\sum_{l=1}^{L} 2(\Psi_l^*(Z^*) - \Psi_l(Z))\nabla\Psi_l(Z),\\
\nabla\Theta_4(Z) &= -2(P^*(Z^*) - P(Z))\nabla P(Z),\\
\nabla\Theta_5(Z) &= -2(C - \sum_{t=1}^{n_j} z_t)e,\\
\nabla\Theta_6(Z) &= -2(\bar{z}^* - \bar{z})\nabla\bar{z}(Z),
\end{aligned}
\tag{18}
$$

where $e \in \mathbb{R}^{n_j}$ is a vector of all ones.

The partial derivatives in the gradient vectors that appear in equations (18) are computed via finite central difference as follows:

$$
\begin{aligned}
\frac{\partial\rho_k}{\partial z_i}(Z) &= \frac{\rho_k(Z + he_i) - \rho_k(Z - he_i)}{2h},\\
\frac{\partial\rho_{I_k}}{\partial z_i}(Z) &= \frac{\rho_{I_k}(Z + he_i) - \rho_{I_k}(Z - he_i)}{2h},\\
\frac{\partial\Psi_k}{\partial z_i}(Z) &= \frac{\Psi_k(Z + he_i) - \Psi_k(Z - he_i)}{2h},\\
\frac{\partial P}{\partial z_i}(Z) &= \frac{P(Z + he_i) - P(Z - he_i)}{2h},\\
\frac{\partial\bar{z}}{\partial z_i}(Z) &= \frac{\bar{z}(Z + he_i) - \bar{z}(Z - he_i)}{2h},
\end{aligned}
$$

where $k = 1, 2, 3$ are the lags, $h = \sqrt{\epsilon_{machine}} \times \max|z_i|$, and e_i, for $i = 1, \ldots, n_j$, are the canonical vectors in $\mathbb{R}^{n_j}$.

It is important to know that any solution $\hat{Z} \in \mathbb{R}^{n_j}$ of problem (12) satisfies that

$$P_\Omega(\hat{Z} - \nabla\Theta(\hat{Z})) - \hat{Z} = 0. \tag{19}$$

This condition is also a sufficient condition for problem (12) if the objective function is convex.

4 Numerical Results

We use daily rainfall data from twelve stations of Guárico State, located at the central plains of Venezuela. The stations are located in an area of 250×250 km^2 approximately and the southern limits are roughly in the same direction as the Orinoco River. The data were registered from 1967 to 1991. Rainfall at these locations is strongly seasonal with a dry season from November to April and a rainy season from May to October. We consider one of the locations the rainfall data consisting of all 24 months of August between 1967 and 1991. For this particular station there exists daily rainfall data for the month of August. We denoted the rainfall data for each different month of August as "Year j" where $j = 1, \ldots, 24$. Then, we solve the daily disaggregation problem for August, which has a small variance since it is a month in the wet season. So, we use the results obtained by Guenni and Bárdossy in [8] when the SA algorithm is applied to problem (6) for this location and we apply the SPG method to disaggregate monthly data from this month at the same location by solving problem (11). The purpose of this work is to establish the advantages of considering the optimization model (11) and the deterministic SPG method for disaggregation of rainfall data. The results obtained with the proposed optimization model and methodology are compared with the results obtained with model (6) and the simulating annealing scheme presented in [8]. The availability of rainfall at a daily time scale allows us to estimate the statistical properties of the daily information we wish to preserve. On the other hand, in order to approximate the daily rainfall sequence for each August month, we consider the same HDR data but without the August month corresponding to the one we want to estimate. So, a fair comparison between the two strategies versus the real daily rainfall series could be attained from these results for August.

All computations were performed in MATLAB, which has unit roundoff $\mu \approx 1.1 \times 10^{-16}$. An initial iterate of a daily sequence, for both methods, was obtained by simulating a truncated normal model, the same used Guenni and Bárdossy in [8]. The stopping criterion for the SPG method is

$$\|P_\Omega(Z_n - \nabla\Theta(Z_n)) - Z_n\|_2^2 \leq 1.0e - 06$$

and the input parameter $M = 5$. The value of parameter M must be different from 1, since in that case the method is the low computational classical gradient method. On the other hand, any number between 5 and 12 could be considered, and for those numbers there is not much difference in number of iterations to converge.

We estimate the daily precipitation using the results obtained when problem (6) is solved with the SA method, and problem (11) is solved with the SPG method. Since each month of August has 31 days and the rainfall data consist of 24 months, the dimension of the vector Z is $n_j = 31$ and the dimension of the vector Z^* is $s * n_j = 24 * 31 = 744$. Each EDR series tries to preserve the autocorrelation function, the indicator autocorrelation function, the Burlando and Rosso scaling parameter, the dry days proportion, the accumulated monthly rainfall C and the mean of the historic daily rainfall series. The value of the threshold δ that represents the smallest rainfall is 5 mm. The lower bounds for each z_t are different but close to $1.0e-15$ and the upper bounds for z_t are distinct but approximately 150. The values $1.0e-15$ and 150 correspond to the minimum and maximum values of the observed rainfall respectively. The values of the weights γ_i in model (11) are all one, and in the Guenni and Bárdossy model (6) are $\frac{1}{3}$. The parameter λ in the Burlando and Rosso multiple scaling relation (5) takes values between 1,2 and 3.

	HDR	SPG_1	SA	$e(SPG_1)$	$e(SA)$
Mean	5.9122	5.1232	5.9057	0.7890	0.0065
Autoc.					
lag 1	0.0519	0.0538	0.0765	0.0019	0.0246
lag 2	0.0161	0.0146	-0.522	0.0015	0.5381
lag 3	0.0218	0.0219	0.1046	0.0001	0.0828
I. Autoc.					
lag 1	0.1354	0.1347	-0.0465	0.0007	0.1819
lag 2	0.5556	0.1305	-0.0664	0.4251	0.6220
lag 3	-0.0026	0.1639	0.1317	0.1665	0.1343
Scale	0.8537	0.9557	1	0.1020	0.1463
Dry days	266	576	395	310	129
Acc. Rain	4398.7	3811.7	4393.9	587	4.8

Tab. 1: Statistics values for EDR series obtained with SPG method and SA method vs. statistics values for the HDR series

Tables 1 and 2 show the values of the desired statistics properties for the

	HDR	SPG_1	SPG_2	$e(SPG_1)$	$e(SPG_2)$
Mean	5.9122	5.1232	5.9133	0.7890	0.0011
Autoc.					
lag 1	0.0519	0.0538	0.0774	0.0019	0.0255
lag 2	0.0161	0.0146	-0.0050	0.0015	0.0211
lag 3	0.0218	0.0219	0.0816	0.0001	0.0598
I. Autoc.					
lag 1	0.1354	0.1347	0.1196	0.0007	0.0158
lag 2	0.5556	0.1305	0.1101	0.4251	0.4455
lag 3	-0.0026	0.1639	0.1485	0.1665	0.1511
Scale	0.8537	0.9557	1	0.1020	0.1463
Dry days	266	576	225	310	41
Acc. Rain	4398.7	3811.7	4399.5	587	0.8

Tab. 2: Statistics values for EDR series obtained with SPG method and SA method vs. statistics values for the HDR series

EDR sequence obtained by the SA method and by the SPG method using 24 years of data. In the rows of Tables 1 and 2 we denote by *"Autoc."* the autocorrelation coefficient function, by *"I. Autoc."* the indicator autocorrelation coefficient function, by *"Scale"* the scaling exponent, by *"Dry days"* the number of dry days and by *"Acc. Rain"* the Accumulated rain. The values of statistical properties for the EDR series obtained by the SPG method and by SA method respectively are shown from the second through the fourth column of Tables 1 and 2. However, in the tables we can distinguish two different SPG methods. The only difference between SPG_1 and SPG_2 is that SPG_1 denotes the results obtained with the SPG method applied to the problem (6) which corresponds to Guenni and Bárdossy model, and SPG_2 denotes the results attained by applying the SPG method to problem (11). The reason of considering SPG_1 and SPG_2 is to show that a deterministic optimization method will adjust the parameters that appears in the objective function or as restrictions. The second column of Tables 1 and 2 presents the computed statistical values for the HDR series. The last two columns of Tables 1 and 2 show the absolute value difference (error) between the SPG method and the HDR series (denoted $e(SPG_1)$ and $e(SPG_2)$, depending of the objective function used for desegregating the rainfall), and between the

SA method and the HDR series (denoted $E(SA)$).

From Table 1, is clear that model (6) and the deterministic methodology SPG (denoted by SPG_1) gets an EDR series with statistical properties closer to the statistical values of the HDR series except for values of the mean, of the dry day proportion and of the accumulated rainfall. This behavior responds to the fact that SPG_1 uses the SPG method for solving problem (6), which does not consider optimization of the mean, of the dry day proportion and of the accumulated rainfall with respect to the HDR series. Indeed, the estimated values of the statistical properties that were included on the objective function given by (6) are better approximated with SPG_1 than the other two methods, since the SPG method tries to adjust a function with less parameters. In Table 2 it is clear that the mean, the dry day proportion and the accumulated rainfall are better estimated with SPG_2. Moreover, in a global sense, the SPG_2 gets better estimates for the statistics properties. In order to make fair comparisons, from now on, we only consider the SA and SPG_2 methods, and we will refer to the last one as SPG. Therefore, the numerical method that better adjust all the desired statistical properties of the HDR series is the SPG, since those statistics properties are all included in the objective function. Especially there is a big difference in the number of dry days for the EDR series when using the SA method, 48-49% more dry days than in the observed daily rainfall sequence. In contrast the EDR series for the SPG method has 15-16% less dry days than the HDR series.

In order to verify which method could get an EDR sequence closer to a global or local solution of the problem (11), we present Table 3. So, Table 3 shows the value of the necessary optimality condition for both strategies, which is given by

$$\|P_\Omega(Z_n - \nabla\Theta(Z_n)) - Z_n\|_2^2,$$

where Z_n is the EDR series, at iteration n, for the considered method. Observe that the value of the necessary condition when solving with the proposed machinery is always less than $1.0e{-}06$, since it is the stopping criterium of the SPG method. However, we can not assure that the obtained solution, is the global one, unless the objective function is convex. But we guarantee that the EDR series for the SPG method for each month is an approximation to a stationary solution of the problem (11). On the other hand, the results presented in Table 3 seem to indicate that the EDR sequence obtained for the SA method for some years is far from a stationary solution of problem (11). However, more experimentation is required for comparing both strategies.

Month	SA method	SPG method	Month	SA method	SPG method
1	8.4564	9.4112E-08	13	0.7774	2.1674E-09
2	4.3839E-02	4.0221E-07	14	7.9978E-02	3.6480E-08
3	7.9774E-05	3.5501E-08	15	1.8378E-02	0
4	3.5760E-05	6.7244E-10	16	0.6168	3.2446E-08
5	1.4488E-02	6.6479E-09	17	3.1125	2.0108E-08
6	11.8078	9.5493E-07	18	5.3111E-03	1.4790E-08
7	1.7624E-04	1.6848E-07	19	0.1663	8.2377E-10
8	9.1821E-03	1.5037E-08	20	1.0135	1.8744E-09
9	6.4237E-05	2.6747E-07	21	1.8400E-02	7.1469E-10
10	1.7320E-02	1.4192E-08	22	5.9611E-03	1.6070E-08
11	5.3593	2.6557E-08	23	1.0362E-02	2.3340E-07
12	8.6781E-04	2.9402E-09	24	0.3919	3.3866E-09

Tab. 3: Necessary condition $\|P_\Omega(Z_n - \nabla\Theta(Z_n)) - Z_n\|$ where Z_n is the EDR series for SA method and SPG method

The difference in the necessary optimality condition for both strategies came from the fact that the SA algorithm obtains an EDR series with a poor approximation on the dry day proportion, the accumulated monthly rainfall and the mean. Moreover, the SA scheme does not handle upper and lower bounds on the rain. These results indicate that the proposed optimization model and optimization strategy allows us to incorporate additional statistical properties to the objective function without changing the optimization strategy and the initial iterate. Moreover, the projection of the rainfall iterate on the constrained set Ω can be easily computed and has low computational cost. On the other hand, the dry day proportion, the accumulated monthly rainfall and the mean values could be considered in the objective function without increasing the complexity of the SPG method. It is well known that the SA method is a global technique, however there is no guarantee that the attained solution, when the SA algorithm stops, is in a neighborhood of the global solution. This will depend on the cooling or temperature function.

Figures 1 and 2 show the quantile-quantile plot for the obtained EDR series when solving the disaggregation problem (6) with SA method and when solving problem (11) with the SPG method respectively. It was observed from both figures that the estimated data for the SA method and for the

SPG method follow almost the same disaggregated probability distribution than the HDR series. We also observed from Figure 2 that at the lower tail of the series there are more quantiles where the probability distribution for the EDR series with SPG is closer to the observed probability distribution than in the upper tail. Moreover, it is clear from both figures that the estimated data it is always on the straight line or above it, except for one quantile at the upper tail of the estimated SA series.

On the other hand, using a different graphical tool, the boxplot for each estimated series and the time series corresponding to the observed and the estimated series give a better idea of the behavior of both models and numerical strategies (variability, extreme values) for solving the disaggregation problem.

Figure 3 illustrates the time series plot for the three data series. This figure shows that the SA strategy generates extreme values on the rainfall series that do not match the extreme values of the observed series. There is not synchronization on peak time of occurrence between the observed series and the SA series. In contrast, the SPG estimated series preserves the shape of the observed series and maintains bounded the extreme values of the rainfall, which implies less variability on the estimated data. These comments are also supported by Figure 4, which presents the boxplot corresponding to the three rainfall series. In fact, in Figure 4, we observed more atypical data for the estimated SA series than in the other two series. We think that the behavior of SPG estimated series is related to the constraints imposed as upper and lower bounds on rainfall extreme values, and the good approximations obtained for the values of the media, accumulated rainfall and dry day proportion. These extreme values are input parameters and could be obtained from a priori information taking into account physical considerations of the historical information.

5 Conclusions

The proposed optimization model and the deterministic low-cost optimization methodology has certain advantages and limitations as any stochastic method. Indeed, it was possible to obtain a disaggregated daily series by using the SPG method, requiring less function evaluations and less CPU time than the SA method. The results indicate that the obtained EDR series will depend on the bounds of the extreme maximum and minimum values of the

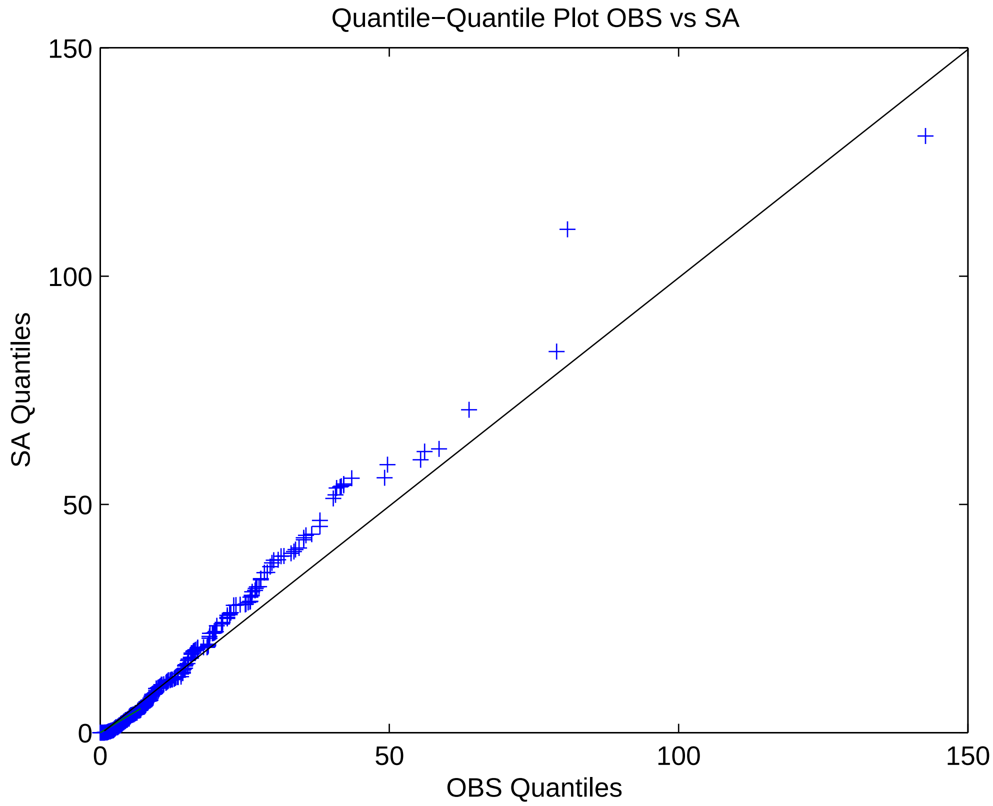

Fig. 1: Quantile-Quantile plot of the EDR series obtained with the SA method

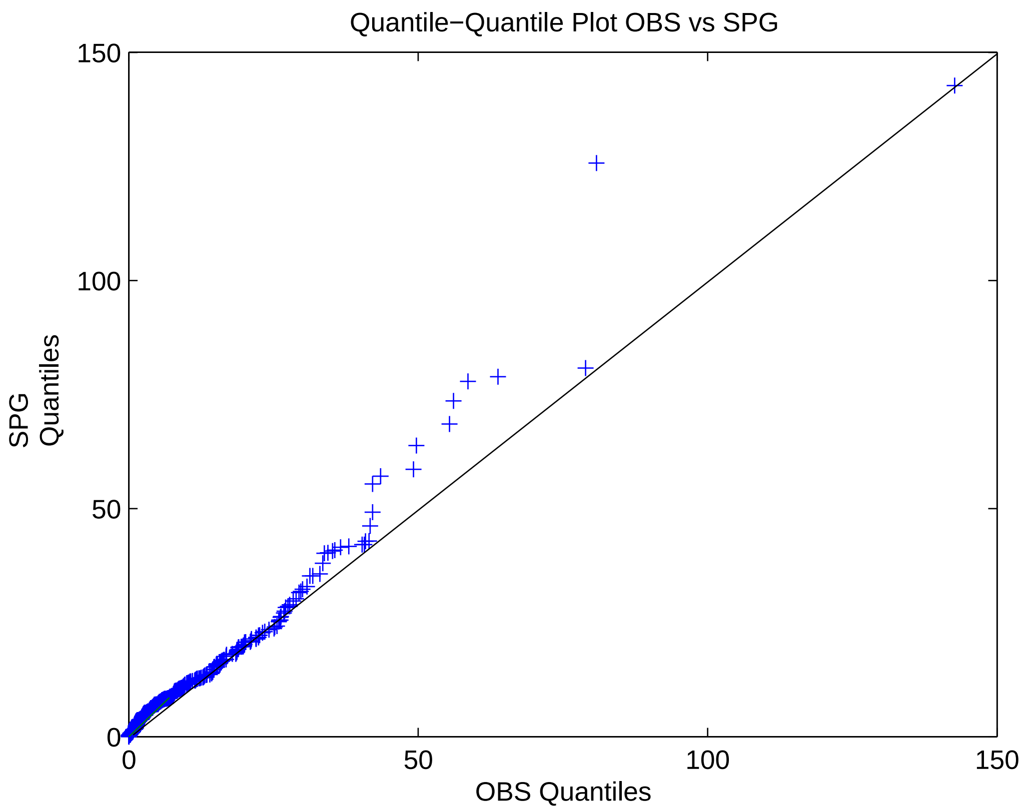

Fig. 2: Quantile-Quantile plot of the EDR series obtained with the SPG method

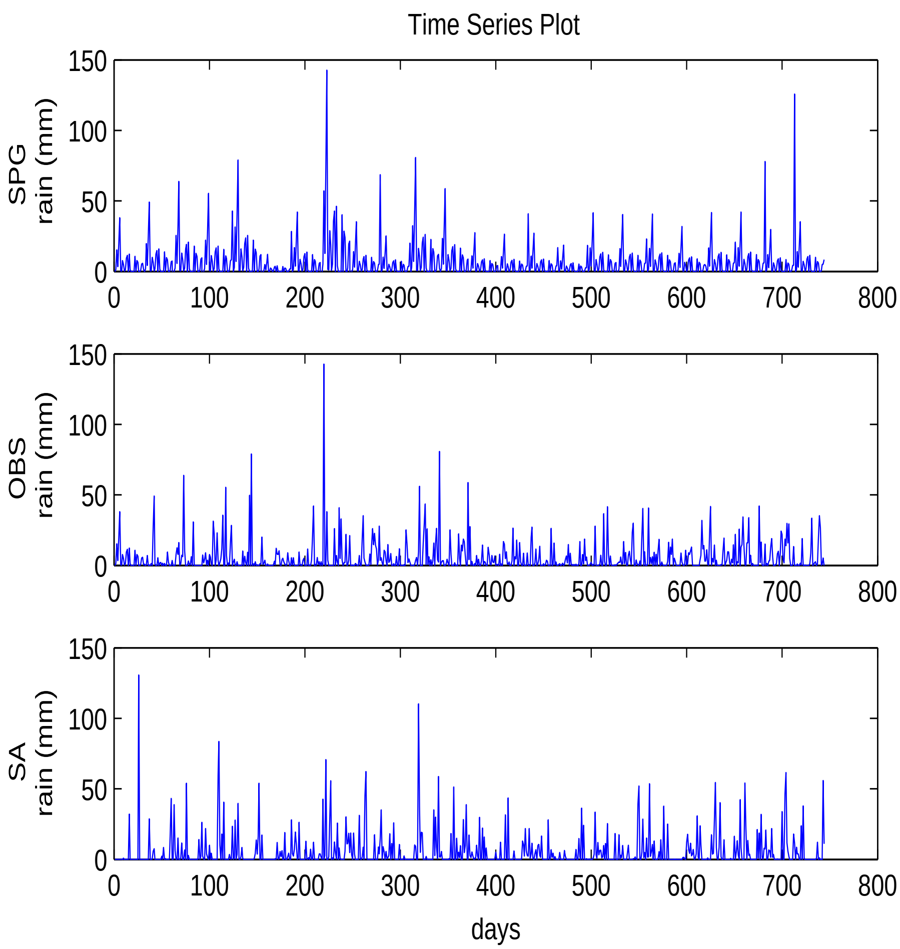

Fig. 3: Time series

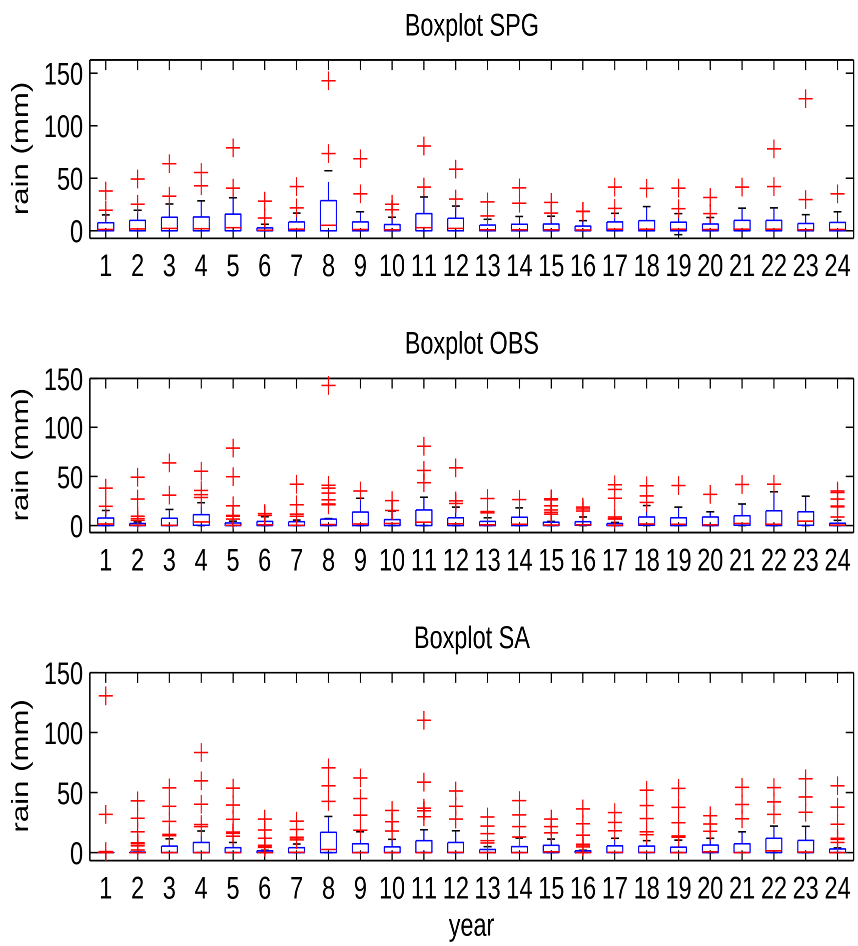

Fig. 4: Boxplots for the EDR series obtained with SA and SPG methods vs. the observed disaggregated sequences for the month of August

daily precipitation. The SPG method is very simple since it only requires the gradient of the objective function whose gradient can be approximated by finite differences, which only require function evaluations as in the SA strategy. The proposed methodology could be used at a shorter time scale, as for example, hours. The dimension of the problem in this case will increase, but since the proposed strategy is an iterative low-cost and low-storage scheme, then it is convenient for large size problems (shorter time scales).

From the numerical results we can infer that the SPG method and the proposed optimization model, problem (11), are appropriate for disaggregating monthly to daily precipitations preserving some basic statistical properties as the ones proposed in this work and in the work by Guenni and Bárdossy in [8].

The EDR series obtained using SA method does not accurately reproduce the dry days probability of the observed series. In particulary, in the paper by Guenni and Bárdossy in [8], the authors suggest to include the probability distribution of consecutive dry days as part of the objective function in order to overcome this problem. However, they also claim that this idea will imply more computational complexity and a larger number of parameters. However, the proposed model, problem (11), allows to attain a better approximation of the dry days proportion in the estimated daily series without adding major complexity to the problem.

In general, daily rainfall information is not available for many locations and most of the months and years. But, it is possible to infer them from the observed data or from any other information. For example, the maximum and minimum values of the precipitation can be inferred from the observed data. So, this priori information can be used to adjust the maximum and minimum extreme values in the estimated daily precipitation sequence as we proposed in this work. The projection over the set containing the bounds of the extreme values of the precipitation, does not require additional computation (i.e., floating point operations) as we explained in Section 2. The inclusion of these constraints in the SA methodology will require to restrict the iterates during the SA process.

The major difference between the SPG method and the SA strategy for disaggregating daily precipitation is that the initial iterate in the SA algorithm must be generated each time the disaggregation process starts, taking into account the observed accumulated precipitation, the probability distribution and the dry days proportion. In contrast the SPG method only requires that the probability distribution of the initial guess be the same as the observed one.

Despite the proposed model adds three new components to the objective function, the increase in the floating point operations when evaluating the objective function, does not make much difference in the performance of the SPG method with respect to the CPU time but it improves the estimated precipitation series.

References

[1] R. D. Valencia and J. C. Schaake. Dissagregation processes in stochastic hydrology. *Water Resources Research*, 9(3):580–585, 1973. doi:10.1029/WR009i003p00580.

[2] R. D. Valencia and J. C. Schaake. A disaggregation model for tiem series analysis and synthesis. Technical Report Rep. 149, Ralph M. Parsons Lab. for Water Resour. and Hydrodyn., Mass. Inst. of Technology, Cambridge-MA, 1973.

[3] D. Koutsoyiannis, C. Onof, and H.S. Wheater. Multivariate rainfall disaggregation at a fine timescale. *Water Resources Research*, 39(7):1173–1190, 2003. doi:10.1029/2002WR001600.

[4] C.A. Glasbey, G. Cooper, and M.B. Mc.Gechan. Disaggregation of daily rainfall by conditional simulation from a point-process model. *Journal of Hydrology*, 165(1):1–9, 1995. doi:10.1016/0022-1694(94)02598-6.

[5] P.S.P. Cowperwait, P.E. O'Connell, A.V. Metcalfe, and J.A. Mawdsley. Stochastic point process modelling of rainfall. II regionalisation and disaggregation. *Journal of Hydrology*, 175:47–65, 1996. doi:10.1016/S0022-1694(96)80005-9.

[6] D. Koutsoyiannis and C. Onof. Rainfall disaggregation using adjusting procedures on a Poisson cluster model. *Journal of Hydrology*, 246(1):109–122, 2001. doi:10.1016/S0022-1694(01)00363-8.

[7] S.J. Burian, S.R. Durrans, S.J. Nix, and R.E. Pitt. Training artificial neural networks to perform rainfall disaggregation. *J. Hydrol. Eng.-ASCE*, 6(1):43–51, 2001. doi:10.1061/(ASCE)1084-0699(2001)6:1(43).

[8] L. Guenni and A. Bárdossy. A two steps disaggregation method for highly seasonal monthly rainfall. *Stochastic Environmental Research and Risk Assessment*, 16(3):188–206, 2002. doi:10.1007/s00477-002-0094-4.

[9] B. Sivakumar, R. Berndtsson, J. Olsson, and K. Jinno. Evidence of chaos in the rainfall-runoff process. *Hydrol. Sci. Journal*, 46(1):131–145, 2001. doi:10.1080/02626660109492805.

[10] B. Sivakumar. Rainfall dynamics at different temporal scales : a chaotic perspective. *Hydrol. Earth System Sci.*, 5(4):645–651, 2001. doi:10.5194/hess-5-645-2001.

[11] M. Raydan. The Barzilai and Borwein gradient method for the large-scale unconstrained minimization problem. *SIAM J. Opt.*, 7(1):26–33, 1997. doi:10.1137/S1052623494266365.

[12] E. G. Birgin, J.M. Martínez, and M. Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, 10(4):1196–1211, 2000. doi:10.1137/S1052623497330963.

[13] P. Burlando and R. Rosso. Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. *J. Hydrol. Eng.*, 187:45–64, 1996. doi:10.1016/S0022-1694(96)03086-7.

[14] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer, New York, 1999.

[15] G.C. Gilbert and J. Nocedal. Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal of Optimization*, 2:21–42, 1992. doi:10.1137/0802003.

[16] D.F. Shanno and K.H. Phua. Remark on algorithm 500: Minimization of unconstrained multivariate functions. *ACM Trans. Math. Software*, 6(4):618–622, 1980. doi:10.1145/355921.355933.

[17] J.J. Moré and D. J. Thuente. On line search algorithms with guaranteed sufficient decrease. *Mathematics and Computer Science Division, Argone National Laboratory*, 1990. preprint MCS-P153-0590.

[18] J.E. Dennis Jr. and R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs, NJ, 1983.