Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology

Demetris Koutsoyiannis

Department of Water Resources, Faculty of Civil Engineering,

National Technical University, Athens

Heroon Polytechniou 5, GR-157 80 Zografou, Greece

(dk@hydro.ntua.gr)

Paper submitted to Water Resources Research

March 1998 / October 1998

Abstract. A new method is proposed for decomposing covariance matrices that appear in the parameter estimation phase of all multivariate stochastic models in hydrology. This method applies not only to positive definite covariance matrices (as do the typical methods of the literature) but to indefinite matrices, too, that often appear in stochastic hydrology. It is also appropriate for preserving the skewness coefficients of the model variables as it accounts for the resulting coefficients of skewness of the auxiliary (noise) variables used by the stochastic model, given that the latter coefficients are controlled by the decomposed matrix. The method is formulated in an optimization framework with the objective function being composed of three components aiming at (a) complete preservation of the variances of variables (b) optimal approximation of the covariances of variables, in case that complete preservation is not feasible due to inconsistent (i.e., not positive definite) structure of the covariance matrix, and (c) preservation of the skewness coefficients of the model variables by keeping the skewness of the auxiliary variables as low as possible. Analytical expressions of the derivatives of this objective function are derived, which allow the development of an effective nonlinear optimization algorithm using the Steepest Descent or the Conjugate Gradient methods. The method is illustrated and explored through a real-world application, which indicates a very satisfactory performance of the method.

1. Introduction

In the parameter estimation phase of all multivariate models of stochastic hydrology we confront the problem of decomposing a covariance matrix **c** into another matrix **b** such as $\mathbf{c} = \mathbf{b}\,\mathbf{b}^T$ (also known as "taking the square root" of **c**). It is well known that this problem has an infinite number of solutions when **c** is positive definite and no (real) solution otherwise. In the era of the first development of the multivariate stochastic models in hydrology, *Matalas and Wallis* [1971] (see also *Matalas and Wallis* [1976]) first pointed out that multivariate stochastic models in hydrology may be inconsistent in the sense that their covariance matrix **c**, estimated by the historical hydrological data, may not be positive definite; in that case the "square root matrix" **b**, which is necessary to express the model itself, does not exist. As it will be explained in section 2 below, such inconsistent matrices are encountered either when only a subset of the covariances among related variables are explicitly modeled, or when missing data affect the parameter estimation. In real world applications, such situations are not infrequent (see *Grygier and Stedinger*, p. 31, and section 5 below). Interestingly, *Slack* [1973] in his article with the emphatic title "I Would If I Could", showed that multivariate synthetic hydrological series may lead to inconsistent (i.e., indefinite) covariance matrix **c**, even if those series were the output of a consistent and simple stochastic model (such as the bivariate Markovian model).

Hydrologists have not hesitated to provide approximate solutions in cases of inconsistent matrices. Various approximate techniques were presented in several works which could be titled "I Could If I Should". Among them are those proposed by *Mejía and Millán* [1974] (also quoted by *Bras and Rodriguez-Iturbe* [1985, p. 98]), *Grygier and Stedinger* [1990, p. 31-33], *Koutsoyiannis* [1992], and *Koutsoyiannis and Manetas* [1996]. These techniques are rather empirical and result in more or less significant alteration of the covariance matrix **c**. A more theoretical approach was devised by *Rasmussem et al.* [1996],

who suggested determination of **b** by numerical constrained optimization, where the objective function is the squared difference between the observed covariance of the data and that produced by the estimated model, and the constraint is the requirement for a positive definite matrix **c**. This method was formulated for a two-site case (i.e. for a matrix **c** with size 2×2); in this case it is convenient to express the positive-definiteness constraint analytically, but it may be difficult to expand to higher dimensional problems.

Seeking for a more generalized theoretical basis for remedying inconsistent covariance matrices, we will assume that, whatever the dimensionality of the problem is, there exists an optimal matrix **b**, that results in the least significant alteration of **c**, or the best approximation of the original **c**. Then, the questions arise: (a) How we can objectively quantify the degree of approximation, and (b) How we can search systematically to find the optimal solution.

In case of a consistent (i.e., positive definite) covariance matrix **c** there exist two well known algorithms for deriving two different solutions **b** (see. e.g., *Bras and Rodriguez-Iturbe* [1985, p. 96]). The first and simpler algorithm, known as triangular or Cholesky decomposition, results in a lower triangular **b**. The second, known as singular value decomposition, results in a full **b** using the eigenvalues and eigenvectors of **c**. However, since it is known that there exists an infinite number of solutions **b**, the question arises whether there exists an optimal solution, possibly different from these two. The answer to this question would be negative if we had no other concern apart from the determination of **b**. In that case, the computationally simpler lower triangular **b** is the most preferable. However, as we will see, there are cases where other concerns must be considered and the answer to this question becomes positive. Then, the subsequent questions are (a) How we can quantify this optimality, and (b) How we can search systematically to find the optimal solution. Again, we have about the same questions for consistent matrices as in the case of inconsistent matrices. This enables a unique treatment of the decomposition problem for consistent or inconsistent matrices.

Another frequent problem in multivariate stochastic models is encountered when we attempt to preserve the coefficients of skewness of the model variables. The auxiliary variables associated with the stochastic model, also known as noise variables or innovation variables, may potentially have very high coefficients of skewness that are practically unachievable. This was first reported by *Todini* [1980], who encountered a coefficient of skewness greater than 30 and was not in position to preserve it. *Koutsoyiannis and Manetas* [1996] related the problem of high skewness with that of the determination of the matrix **b**, since the skewness coefficients are proportional to the inverse of matrix $\mathbf{b}^{(3)}$, that is the matrix whose elements are the cubes of **b**. The close relation of these two problems can contribute to the quantification of the optimality of matrix **b** in case of either consistent or inconsistent **c**. That is, we can set the requirement that the matrix **b** must result in as small coefficient of skewness of the auxiliary variables as possible.

The purpose of this study is the development of a systematic method to remedy all the above described parameter estimation problems for both consistent and inconsistent covariance matrices, also answering the questions set above. All problems are resolved in a unique manner at the grounds of an optimization framework. A single objective function incorporating all concerns about the matrix **b** is proposed and a procedure is developed for finding the optimal **b**. This procedure is based on the nonlinear optimization theory and utilizes both the objective function and its partial derivatives with respect to **b**, which are analytically derived in this paper.

The paper is organized in six sections. Section 2 is devoted to the clarification of notation and some introductory aspects of a generalized stochastic model that constitutes the basis for further analysis. In section 3 we formulate the conditions that determine an optimal matrix **b** and develop the objective function. In section 4 we develop the numerical procedure for determining the optimal **b**. In section 5 we present a case study to illustrate the method and investigate some practical issues. Section 6 is devoted to conclusions and discussion. In

addition, there is an appendix where we have placed an essential part of the paper, i.e., the analytical derivation of the derivatives of the objective function.

2. The stochastic model

In the following analysis we will consider a general type of linear model which is typical in stochastic hydrology, that is

$$\mathbf{Y} = \mathbf{a}\,\mathbf{Z} + \mathbf{b}\,\mathbf{V} \tag{1}$$

where $\mathbf{Y}$ is a vector of n stochastic variables to be generated, $\mathbf{Z}$ is a vector of m stochastic variables with known values (n and m may be equal or not), $\mathbf{V}$ is a vector of n random variates with unit variance, mutually independent, and also independent with $\mathbf{Z}$ (often called noise, innovation, or auxiliary variables), and $\mathbf{a}$ and $\mathbf{b}$ are matrices of coefficients with sizes $n \times m$ and $m \times m$, respectively. (In this paper we use uppercase letters for random variables and lowercase letters for values of variables or coefficients; also we use bold letters for matrices or vectors and regular letters for scalars). Generally, the elements of $\mathbf{Y}$ represent specific hydrologic processes (rainfall, runoff, etc.) at some locations specified by an index $l = 1, \ldots, n$, at a specific time period, whereas the elements of $\mathbf{Z}$ represent the same or other related hydrologic processes at the same or other locations, generally at a previous time period. The variables $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{V}$ are not necessarily standardized to have zero mean and unit variance although this is the case in most common models; however, $\mathbf{V}$ have by definition unit variance. Also, the variables $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{V}$ are not necessarily Gaussian.

For example, in the case of the stationary AR(1) model we set $\mathbf{Y} \equiv \mathbf{X}^t$ and $\mathbf{Z} \equiv \mathbf{X}^{t-1}$, where $\mathbf{X}^t$ represents hydrologic variables at n sites at the time period (typically year) t and (1) writes

$$\mathbf{X}^t = \mathbf{a}\,\mathbf{X}^{t-1} + \mathbf{b}\,\mathbf{V}^t \tag{2}$$

Similarly, in the case of the seasonal AR(1) (or PAR(1)) model (1) writes

$$\mathbf{X}^s = \mathbf{a}^s \, \mathbf{X}^{s-1} + \mathbf{b}^s \, \mathbf{V}^s \tag{3}$$

where now the matrices of coefficients depend on the season (typically month) s. In both these examples the vectors $\mathbf{Y}$ and $\mathbf{Z}$ have the same dimension $n = m$. In case of the AR(2) we have $\mathbf{Y} \equiv \mathbf{X}^t$ and $\mathbf{Z} \equiv [(\mathbf{X}^{t-1})^T, (\mathbf{X}^{t-2})^T]^T$ (where the exponent T denotes the transpose of a matrix or vector), so that $m = 2\,n$. Similar is the situation for the PAR(2) model. In another example, the *Valencia and Schaake*'s [1972, 1973] disaggregation model, $\mathbf{Y}$ represents the $n = 12\,m$ monthly hydrologic variables at m sites whereas $\mathbf{Z}$ represents the m annual values at the same locations. Some other disaggregation models can be also reduced in the form (1) after appropriate assignments of the variables $\mathbf{Y}$ and $\mathbf{Z}$.

The model parameters $\mathbf{a}$ and $\mathbf{b}$ are typically determined by the moment estimators that are

$$\mathbf{a} = \mathrm{Cov}[\mathbf{Y}, \mathbf{Z}] \; \{\mathrm{Cov}[\mathbf{Z}, \mathbf{Z}]\}^{-1} \tag{4}$$

$$\mathbf{b}\,\mathbf{b}^T = \mathrm{Cov}[\mathbf{Y}, \mathbf{Y}] - \mathbf{a}\,\mathrm{Cov}[\mathbf{Z}, \mathbf{Z}]\,\mathbf{a}^T \tag{5}$$

where $\mathrm{Cov}[\boldsymbol{\Xi}, \boldsymbol{\Psi}]$ denotes the covariance matrix of any two random vectors $\boldsymbol{\Xi}$ and $\boldsymbol{\Psi}$, i.e., $\mathrm{Cov}[\boldsymbol{\Xi}, \boldsymbol{\Psi}] := E\{(\boldsymbol{\Xi} - E[\boldsymbol{\Xi}])\,(\boldsymbol{\Psi}^T - E[\boldsymbol{\Psi}]^T)\}$ with $E[\,]$ denoting expected value (the symbol := stands for equality by definition). These equations are direct generalizations for the model (1) of the equations of the AR and PAR models given by *Matalas and Wallis* [1976, p. 63], *Salas et al.* [1988, p. 381], *Salas* [1993, p. 19.31], and *Koutsoyiannis and Manetas* [1996], among others.

Another group of model parameters are the moments of the auxiliary variables $\mathbf{V}$. The first moments (means) are directly obtained from (1) by taking expected values, i.e.,

$$E[\mathbf{V}] = \mathbf{b}^{-1} \; \{E[\mathbf{Y}] - \mathbf{a}\,E[\mathbf{Z}]\} \tag{6}$$

The variances are by definition 1, i.e.,

$$\mathrm{Var}[\mathbf{V}] = [1, \ldots, 1]^T \tag{7}$$

The third moments are obtained by cubing both sides of (1), where previously we have subtracted the means, and then taking expected values. Observing that, because of the independence of **Z** and **V**, joint third order terms of $(Z_k - E[Z_k])$ and $(V_l - E[V_l])$ have zero expected values, and similarly, because of the mutual independence of **V**, joint third order terms of $(V_k - E[V_k])$ and $(V_l - E[V_l])$ also have zero expected values, we get

$$\mu_3[\mathbf{V}] = \left(\mathbf{b}^{(3)}\right)^{-1} \{\mu_3[\mathbf{Y}] - \mu_3[\mathbf{a}\,\mathbf{Z}]\} \tag{8}$$

where $\mu_3[\boldsymbol{\Xi}]$ denotes the third central moments of any random vector $\boldsymbol{\Xi}$, i.e., $\mu_3[\boldsymbol{\Xi}] := E\{(\boldsymbol{\Xi} - E[\boldsymbol{\Xi}])^3\}$ and $\mathbf{b}^{(3)}$ denotes the matrix whose elements are the cubes of **b**. (Throughout this paper we will extend this notation, i.e., $\mathbf{u}^{(k)}$ for any matrix or vector **u** and any power *k*). Equation (8) is a generalization of those given by *Matalas and Wallis* [1976, p. 64], *Todini* [1980], and *Koutsoyiannis and Manetas* [1996]. Moments of order greater than three are not used nor they can be estimated in a similar manner.

The set of equations (4) to (8) determine completely the model parameters with an exception for (5), which does not estimate the parameter matrix **b** but the product $\mathbf{b}\,\mathbf{b}^T$. This has to be decomposed, as it will be discussed in the next section. Generally, all parameter estimation equations involve only moments of the original variables **Y** and **Z**, either marginal of order 1 to 3, or joint of order 2. There is an exception in (8), which involves third moments of a linear combination of Z_k (i.e., **a Z**) which cannot be estimated in terms of the marginal third moments of Z_k (in fact, third order joint moments of Z_k are involved, which are impractical to use). Thus, the solution is to estimate $\mu_3[\mathbf{a}\,\mathbf{Z}]$ from the available data for **Z** after estimating **a** and performing the linear transformation **a Z**. However, this is an inconvenient

situation that can be avoided only if each row of **a** contains only one non-zero element (for $m \leq n$), in which case (8) reduces to

$$\mu_3[\mathbf{V}] = \left(\mathbf{b}^{(3)}\right)^{-1} \{\mu_3[\mathbf{Y}] - \mathbf{a}^{(3)}\, \mu_3[\mathbf{Z}]\} \tag{9}$$

Apparently, (4) is not appropriate to construct such an **a** (i.e., with one non-zero element in each row), and therefore it must be replaced by a simplified form so that, if a_{ij} is the only non-zero element for row i, then

$$a_{ik} = \begin{cases} \text{Cov}[Y_i, Z_j] \,/\, \text{Var}[Z_j] & k = j \\ 0 & k \neq j \end{cases} \tag{10}$$

As a consequence, covariances among **Y** and **Z** apart from Cov[Y_i, Z_j] are not preserved in this case. All other parameter estimation equations are still valid in their form written above with (8) replaced by its simplified form (9). This special case is met, for example, in the so called contemporaneous AR(1) and contemporaneous PAR(1) model [*Matalas and Wallis*, 1976, p. 63; *Salas*, 1993, p. 19.31]. In both these models **a** is diagonal; for example in the latter model

$$\mathbf{a} = \text{diag}\left(\text{Cov}[X_1^s, X_1^{s-1}] \,/\, \text{Var}[X_1^{s-1}], \ldots, \text{Cov}[X_n^s, X_n^{s-1}] \,/\, \text{Var}[X_n^{s-1}]\right) \tag{11}$$

Having expressed all basic equations for parameter estimation, we can return to the discussion of the introduction (section 1) for the situations leading to inconsistent matrices **c** = **b** $\mathbf{b}^T$. We must emphasize that the model (1) with parameter estimators (4) and (5) and complete data sets for estimation, always results in positive definite, i.e. consistent, **c**. The only cases in which inconsistencies may appear are (a) the trivial case with one series of equal values (resulting in a covariance matrix column with all zeroes); (b) when different items of the covariance matrices are estimated using records of different lengths due to missing data; and (c) when a simplified form of the matrix **a** is adopted, such as in (10) (instead of the complete form (4)). To the case (b) we must incorporate a rather usual practice followed in

AR models (e.g., when $\mathbf{Y} \equiv \mathbf{X}^t$ and $\mathbf{Z} \equiv \mathbf{X}^{t-1}$), which may cause inconsistencies even in case of using the complete form of **a** and having no missing data: This occurs when the contemporaneous covariances matrices (e.g. Cov $[\mathbf{X}^t, \mathbf{X}^t]$) are estimated using the complete series of length k whereas lagged covariance matrices (e.g. Cov $[\mathbf{X}^t, \mathbf{X}^{t-1}]$) are estimated using (unavoidably) smaller lengths (e.g., $k - 1$).

Given all above equations for the model parameters we will focus on the estimation of **b** which also affects both $E[\mathbf{V}]$ and $\mu_3[\mathbf{V}]$, as it follows from (6)-(9).

3. Formulation of the conditions determining an optimal matrix b

As stated in the introduction, the purpose of this paper is the quantification of the performance of the parameter matrix **b** through an appropriate objective function that incorporates all concerns about that matrix. Given that objective function we can then develop an algorithm to find an optimal **b**.

As an initial step for a more convenient formulation of the method for determining an optimal matrix **b** we perform a standardization of **b** and the other matrices and vectors associated with it. We call **c** the known right hand term of (5), i.e.,

$$\mathbf{c} := \mathrm{Cov}[\mathbf{Y}, \mathbf{Y}] - \mathbf{a}\, \mathrm{Cov}[\mathbf{Z}, \mathbf{Z}]\, \mathbf{a}^T \tag{12}$$

The matrix **c** is in fact the variance-covariance matrix of the vector $\mathbf{Y} - \mathbf{a}\,\mathbf{Z}$, and thus all its diagonal elements are positive (they represents the variances of the vector components). Thus, we can standardize **c** using the diagonal matrix

$$\mathbf{h} := \mathrm{diag}\left(1 / \sqrt{c_{11}}, \ldots, 1 / \sqrt{c_{nn}}\right) \tag{13}$$

so that

$$\mathbf{c}' := \mathbf{h}\, \mathbf{c}\, \mathbf{h} \tag{14}$$

has all its diagonal elements equal to 1 and the off-diagonal elements between –1 and 1 (The off-diagonal elements may slightly violate this rule if **a** is constructed by (10) or (11) instead of (4)). If we define

$$\mathbf{b}' := \mathbf{h}\,\mathbf{b} \tag{15}$$

then it is easily shown that (5) becomes

$$\mathbf{b}'\,\mathbf{b}'^{T} = \mathbf{c}' \tag{16}$$

and if we also define the vectors

$$\boldsymbol{\varphi} := \mathbf{h}^{(3)}\,\{\mu_3[\mathbf{Y}] - \mu_3[\mathbf{a}\,\mathbf{Z}]\} \tag{17}$$

$$\boldsymbol{\xi} := \mu_3[\mathbf{V}] \tag{18}$$

then (8) becomes

$$\boldsymbol{\xi} = (\mathbf{b}'^{(3)})^{-1}\,\boldsymbol{\varphi} \tag{19}$$

The matrices **h** and **c**′ and the vector **φ** are known whereas **b**′ and **ξ** have to be determined; specifically **ξ** depends on **b**′. Equations (16) and (19) constitute the basis for the proposed method. Since (16) does not always have a real solution we set

$$\mathbf{d} := \mathbf{b}'\,\mathbf{b}'^{T} - \mathbf{c}' \tag{20}$$

and demand that all elements of **d** be as close to zero as possible. This requirement can be expressed mathematically as

$$\text{minimize} \quad \|\mathbf{d}\|^2 := \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 \tag{21}$$

Here we have used the notation ||**d**|| for the norm, as if **d** were a vector rather than a matrix. (If it were a vector then ||**d**|| would be its Euclidean or standard norm; see, e.g., *Marlow* [1993, p.

59]). We have used the square of this norm $||\mathbf{d}||^2$, instead of $||\mathbf{d}||$, for two reasons: first because the computations are simpler, as we will see in the next section, and second, because it was found that the convergence of the optimization procedure is faster for $||\mathbf{d}||^2$ rather than for $||\mathbf{d}||$.

In addition, we must set a restriction that all diagonal elements of **d** should be exactly zero. To justify this requirement we observe that if the diagonal elements of **d** are zero, then all diagonal elements of $\mathbf{b}\,\mathbf{b}^T$ will equal those of **c**. In turn, this will result in preservation of the diagonal elements of Cov[**Y**, **Y**] (as implied from (5) and (12)), i.e., in preservation of the variances of all elements of **Y**. The preservation of the variances must have priority against that of covariances because the former is related to the preservation of the marginal distribution functions of the components of **Y**. To express mathematically this requirement we introduce the diagonal matrix

$$\mathbf{d}^* := \mathrm{diag}(d_{11}, \ldots, d_{nn}) \tag{22}$$

and demand that

$$||\mathbf{d}^*||^2 = 0 \tag{23}$$

Another restriction arises when we consider the preservation of the coefficients of skewness of **Y**. The coefficients of skewness of **V** (i.e., $\boldsymbol{\xi}$) that preserve the coefficients of skewness of **Y** are those obtained by (19). However, (19) may result in arbitrary high elements of $\boldsymbol{\xi}$ if no relevant provision for **b** is sought. Nevertheless, in the model application phase (i.e. the generation of synthetic data), an arbitrary high coefficient of skewness is hardly achievable. Specifically, it is well known that in a finite sample of size k the coefficient of skewness is bounded [*Wallis et al.*, 1974; *Kirby*, 1974; *Todini*, 1980] between $-\xi_{\mathrm{ub}}$ and ξ_{ub}, where

$$\xi_{\mathrm{ub}} = \frac{k-2}{\sqrt{k-1}} \approx \sqrt{k} \tag{24}$$

In particular, a series of generated values v_r ($r = 1, \ldots, k$) will have skewness $\pm\xi_{ub}$ only if all but one values of the series are equal. Apparently, this is not an accepted arrangement of the generated series, and thus an acceptable coefficient of skewness ξ_{acc} must be much less than ξ_{ub} (e.g., $\xi_{acc} = 0.5\ \xi_{ub}$; for $k = 1000$, $\xi_{acc} = 0.5 \times 1000^{0.5} \approx 16$). Consequently, we must set the restriction that all n auxiliary variables V_i ($i = 1, \ldots, n$) have coefficients of skewness less than ξ_{acc}, i.e.,

$$\max\ \{|\xi_i|,\ i = 1, \ldots, n\} \le \xi_{acc} \tag{25}$$

The handling of (25) in an optimization problem is not mathematically convenient, especially when we wish to calculate derivatives, as is the case in the procedure that we will develop. However, we recall that if

$$\|\boldsymbol{\xi}\|_p := \left(\sum_{i=1}^{n} |\xi_i|^p \right)^{1/p} \tag{26}$$

is the pth norm of $\boldsymbol{\xi}$, then $\|\boldsymbol{\xi}\|_p \to \max\ \{|\xi_i|,\ i = 1, \ldots, n\}$ when $p \to \infty$ (see, e.g., Marlow [1993, p. 59]). Therefore, (25) can be replaced by

$$\|\boldsymbol{\xi}\|_p^2 \le \xi_{acc}^2 \tag{27}$$

where we have squared the norm as we already did in (21) and (23). To avoid the absolute values within the right-hand side of (26) it suffices to use an even p. Specifically, we have adopted the value $p = 8$, which it was numerically proved to be adequately high for $\|\boldsymbol{\xi}\|_p$ to approach $\max\ \{|\xi_i|,\ i = 1, \ldots, n\}$.

In this formulation, the problem of determining $\mathbf{b}'$ becomes a constrained optimization problem with three elements: (a) the objective function (21) to be minimized, (b) the equality constraint (23) and (c) the inequality constraint (27). Note that (27) is related to $\mathbf{b}'$ through (19). In terms of the Lagrangian method, the objective function and the constraints can be combined into the unconstrained minimization problem

$$\text{minimize} \quad \theta^2(\mathbf{b}', \lambda_2, \lambda_3) := \|\mathbf{d}(\mathbf{b}')\|^2 + \lambda_2 \|\mathbf{d}^*(\mathbf{b}')\|^2 + \lambda_3 \left(\|\boldsymbol{\xi}(\mathbf{b}')\|_p^2 - \xi_{\text{acc}}^2 \right) \tag{28}$$

where λ_2 and λ_3 are Lagrangian multipliers for the equality constraint (23) and the inequality constraint (27), respectively. A general theoretical framework for solving this problem might consist of two steps: In the first step we assume that $\lambda_3 = 0$ (i.e., ignore the inequality constraint) and determine $\mathbf{b}'$ by minimizing $\theta^2(\mathbf{b}', \lambda_2, 0)$. (This step may be avoided if $\mathbf{c}'$ is positive definite, in which case $\mathbf{b}'$ can be determined by the typical decomposition methods.) If the solution $\mathbf{b}'$ of this step satisfies (27), then this solution is the optimal one. Otherwise, we proceed to the second step, where we perform the optimization of the complete function (28) (i.e., with $\lambda_3 \neq 0$) to simultaneously determine $\mathbf{b}'$, λ_2 and λ_3. For this optimization the Kuhn-Tucker optimality conditions (see, e.g., *Marlow* [1993, p. 271]), must be satisfied.

Apparently, due to the complexity of the problem equations, an analytical solution of equations is not attainable and a numerical optimization scheme must be established. Initial attempt to establish a numerical procedure for (28) (using typical available solvers for multivariate constrained optimization) indicated that it is not easy to optimize it simultaneously for the parameters $\mathbf{b}'$ and the Lagrange multipliers. Particularly, it was observed that the initially assumed values of the Lagrange multipliers do not advance towards an optimal solution but remain almost constant through the consecutive steps of numerical optimization. At the same time, the objective function itself has a decreasing trajectory through the consecutive steps, owing to the appropriate modification of parameters $\mathbf{b}'$. This indicates that the formulation of the problem in terms of (28) is not effective in practice.

A direct alternative is the introduction of penalties for the constraints [e.g., *Pierre*, 1986, pp. 21, 333-340]. This alternative is similar with the Lagrangian formulation as it combines the objective function and the constraints into a simple unconstrained objective function; as we will see, this function can be very similar with (28). The essential difference between the two formulations is the fact that in the second method the coefficients λ_2 and λ_3 are preset weighting factors of the penalizing terms, no more determined by the optimization

procedure. However, given the preliminary numerical results discussed in the previous paragraph (i.e., the constancy of coefficients λ through consecutive steps) this difference does not have any practical meaning for the problem examined.

To formulate the problem in terms of penalty functions, we observe that the equality constraint (23) is easily incorporated into the objective function (21) by the addition of the penalty term $\lambda_2 \, ||\mathbf{d}^*||^2$, where λ_2 is assigned a large value so that even a slight departure of $||\mathbf{d}^*||^2$ from zero results in a significant "penalty". The inequality constraint (27) can be treated in several ways, of which the simplest is the addition of the penalty term $\lambda_3 \, ||\boldsymbol{\xi}||_p^2$ into the objective function (21), where λ_3 is a weighting factor appropriately chosen so that the penalty term be small enough when $||\boldsymbol{\xi}||_p \le \xi_{\text{acc}}$. The resulting objective function in this case is

$$\text{minimize} \quad \theta^2(\mathbf{b}') := \frac{\lambda_1}{n^2} ||\mathbf{d}(\mathbf{b}')||^2 + \frac{\lambda_2}{n} ||\mathbf{d}^*(\mathbf{b}')||^2 + \lambda_3 \, ||\boldsymbol{\xi}(\mathbf{b}')||_p^2 \tag{29}$$

Here, we have divided $||\mathbf{d}||^2$ by n^2, the number of elements of $\mathbf{d}$, to convert the total "square error" $||\mathbf{d}||^2$ into an average square error for each element of $\mathbf{d}$; similarly we have divided $||\mathbf{d}^*||^2$ by n because $\mathbf{d}^*$ is diagonal so that the number of nonzero elements is n. The coefficient λ_1 in the first term of the right-hand side of (29) may be considered equal to 1 unless specified different (see the case study in section 5). We observe that (29) is similar with (28) from a practical point of view: As discussed before, the theoretical difference that the parameters λ are preset constants in (29) has not any practical implication; also the omission in (29) of ξ_{acc}^2, which is a constant, apparently does not affect the minimization process.

A reasonable choice for λ_2 to ensure that $||\mathbf{d}^*||^2$ will be very close to zero (so that constraint (23) will hold), is $\lambda_2 = 10^3$; this value assigns a weight of the average square error of the diagonal elements three orders of magnitude higher than that of the off-diagonal ones. For the choice of an appropriate value of λ_3 let us assume that an acceptable value of $||\mathbf{d}||^2 / n^2$ (the first term of the right-hand side of (29)) will be of the order of 10^{-3} (recall that $\mathbf{c}$ is standardized so that its off-diagonal elements are less than 1) whereas an acceptable value of

$\lambda_3 \, \|\boldsymbol{\xi}\|_p^2$, as it results from (24) and the subsequent discussion, will be ($\lambda_3 \, k$ / 4), where k is the size of the synthetic record to be generated. Assuming that the acceptable values of the first and the third term are of the same order of magnitude, so that both terms have about the same significance in the objective function (29), we conclude that $\lambda_3 \, k / 4 \approx 10^{-3}$ or $\lambda_3 \approx 4 \times 10^{-3} / k$. We may assume that in typical generation problems of stochastic hydrology k does exceed 40, so that a maximum value of λ_3 must be about 10^{-4}. We may use a smaller value of λ_3 (e.g., 10^{-5} - 10^{-6}) either if the sample size is of higher order of magnitude or if we wish to give a greater importance to the preservation of covariances rather than to that of the coefficients of skewness.

The establishment of a numerical procedure for minimizing the objective function (29) is given in the next section.

4. Optimization procedure

To establish an effective optimization algorithm for (29) we need first to determine the partial derivatives of all components of θ^2 with respect to the unknown parameters b'_{ij}. This may seem a difficult task to accomplish analytically. However, it is not at all intractable. For the convenience of the reader we have placed all calculations of the derivatives in the Appendix and we present here only the final results of the calculations, which are very simple. In these results, given any scalar a we use the notation

$$\frac{d\alpha}{d\mathbf{b}'} := \begin{bmatrix} \frac{\partial \alpha}{\partial b'_{11}} & \frac{\partial \alpha}{\partial b'_{12}} & \cdots & \frac{\partial \alpha}{\partial b'_{1n}} \\ \frac{\partial \alpha}{\partial b'_{21}} & \frac{\partial \alpha}{\partial b'_{22}} & \cdots & \frac{\partial \alpha}{\partial b'_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \alpha}{\partial b'_{n1}} & \frac{\partial \alpha}{\partial b'_{n2}} & \cdots & \frac{\partial \alpha}{\partial b'_{nn}} \end{bmatrix} \tag{30}$$

for the matrix of its partial derivatives with respect to all b'_{ij}; this is an extension of the notation used for vectors [e.g., *Marlow*, 1993, p. 208].

As shown in the Appendix, the derivatives of all components of θ^2 are

$$\frac{d\,\|\mathbf{d}\|^2}{d\,\mathbf{b}'} = 4\,\mathbf{d}\,\mathbf{b}' \tag{31}$$

$$\frac{d\,\|\mathbf{d}^*\|^2}{d\,\mathbf{b}'} = 4\,\mathbf{d}^*\,\mathbf{b}' \tag{32}$$

$$\frac{d\,\|\boldsymbol{\xi}\|_p^2}{d\,\mathbf{b}'} = -6\,\|\boldsymbol{\xi}\|_p^{2-p}\,\mathbf{w} \tag{33}$$

where $\mathbf{w}$ is a matrix with elements

$$w_{ij} := b'^2_{ij}\,\xi_j\,\psi_i \tag{34}$$

and $\boldsymbol{\psi}$ is a vector defined by

$$\boldsymbol{\psi} := \left[\left(\mathbf{b}'^{(3)}\right)^{-1}\right]^T \boldsymbol{\xi}^{(p-1)} \tag{35}$$

Thus, the required matrix of derivatives of θ^2 with respect to the unknown parameters b'_{ij} is

$$\frac{d\theta^2}{d\mathbf{b}'} = \frac{4\,\lambda_1}{n^2}\,\mathbf{d}\,\mathbf{b}' + \frac{4\,\lambda_2}{n}\,\mathbf{d}^*\,\mathbf{b}' - 6\,\lambda_3\,\|\boldsymbol{\xi}\|_p^{2-p}\,\mathbf{w} \tag{36}$$

It is apparent from the objective function (29) and its derivatives (36) that we have a typical nonlinear optimization problem, whose solution can be achieved by iterations, starting with an initial matrix $\mathbf{b}'^{[0]}$. In the lth iteration we start with a known $\mathbf{b}'^{[l]}$ and we find an "improved" matrix $\mathbf{b}'^{[l+1]}$; we repeat this procedure until the solution converges. Several algorithms are known in the literature for advancing from $\mathbf{b}'^{[l]}$ to $\mathbf{b}'^{[l+1]}$ (see *Mays and Tung* [1996, p. 6.12], among others, who summarize the most common ones). Among them, the most suitable for our case are those of the Steepest Descend and Fletcher-Reeves Conjugate

Gradient methods. These are chosen for their mathematical simplicity and convenience, and their low memory requirements. Specifically, the mathematical formulation of both methods, which typically assume a vector arrangement of the unknown parameters, can be directly adapted for our case, which uses a matrix arrangement of the unknown parameters. Furthermore, the memory requirements are of great importance because in the problem examined the number of unknown parameters is large, i.e., n^2; the amount of memory locations for both chosen methods is of the order of n^2 whereas for other methods such as quasi-Newton methods would be n^4.

Both these methods are described by the following common expression, adapted from the typical expression given in the literature for vector-arranged variables [e.g., *Mays and Tung*, 1996, p. 6.12; *Press et al.*, 1992, p. 422],

$$\mathbf{b}'^{[l+1]} = \mathbf{b}'^{[l]} - \beta^{[l+1]} \left[\left(\frac{d\theta^2}{d\mathbf{b}'} \right)^{[l]} + \gamma^{[l]} \left(\frac{d\theta^2}{d\mathbf{b}'} \right)^{[l-1]} \right] \tag{37}$$

where

$$\gamma^{[l]} = \begin{cases} 0 & \text{for Steepest Descent method} \\ \left\| \left(\frac{d\theta^2}{d\mathbf{b}'} \right)^{[l]} \right\|^2 / \left\| \left(\frac{d\theta^2}{d\mathbf{b}'} \right)^{[l-1]} \right\|^2 & \text{for Fletcher-Reeves method} \end{cases} \tag{38}$$

and $\beta^{[l+1]}$ is a scalar whose value is obtained by a line search algorithm for each iteration l. For $l = 0$ (first step) the Fletcher-Reeves method cannot be used because $(d\theta^2 / d\mathbf{b}')^{[-1]}$ is not defined and, thus, we must use the Steepest Descent method (i.e., $\gamma^{[0]} = 0$). For the other steps, the numerical applications have indicated that the Fletcher-Reeves method is faster and thus preferable. However, it was observed that in some instances during the iteration procedure the Fletcher-Reeves method may become too slow; in such instances the procedure is accelerated if we perform one iteration with the Steepest Descent method and then continue again with the Fletcher-Reeves method.

What it remains is a procedure to construct an initial matrix $\mathbf{b}'^{[0]}$. This is not quite important because, as demonstrated in the case study below (section 5), the method is very fast in the first phase of the iteration procedure. Even if we start with a $\mathbf{b}'^{[0]}$ leading to an unrealistically high value of the objective function, this value is dramatically reduced in the first few iterations. The general idea for the construction of the initial matrix $\mathbf{b}'^{[0]}$ is to start decomposing the matrix $\mathbf{c}'$ by triangular decomposition, which is the simplest of all procedures [*Press et al.*, 1992, p. 97], and occasionally performing some corrections when this procedure fails. This will be demonstrated more clearly in the case study below (section 5).

5. Case study

To illustrate the developed method we present a part of a real-world case study. This case study aims at the generation of simultaneous monthly rainfall and runoff at three basins namely Evinos, Mornos and Yliki, supplying water to Athens, Greece [*Nalbantis and Koutsoyiannis*, 1997]. This constitutes a multivariate generation problem with 6 locations (2 variables × 3 basins). As a general framework for the generation, the Simple Disaggregation Model [*Koutsoyiannis and Manetas*, 1996] was adopted. Specifically, at a first step the annual variables for all locations are generated using a multivariate AR(1) model. Then, the annual quantities are disaggregated into monthly ones, using a combination of a multivariate contemporaneous seasonal AR(1) (or PAR(1)) model and an accurate adjusting procedure, as described by *Koutsoyiannis and Manetas* [1996]. Here we focus on the PAR(1) model, described by (3), which generates the vector of random variables $\mathbf{X}^s$ at the six locations for each month s, given those of the previous month $\mathbf{X}^{s-1}$. The model parameters are given by (11), (5), (6), (7), and (9), in which $\mathbf{Y}$ and $\mathbf{Z}$ must be replaced with $\mathbf{X}^s$ and $\mathbf{X}^{s-1}$, respectively.

In Table 1 we display the sample statistics of the monthly rainfall and runoff data that are necessary for estimating the contemporaneous PAR(1) model parameters for subperiod

(month) s = 8 (May; note that hydrologic year starts in October). These are the first three marginal moments at all locations (represented by the means, standard deviations, and coefficients of skewness) and the matrices of cross correlations for all locations for both subperiods s = 8 and $s - 1 = 7$; the autocorrelations among subperiods s = 8 and $s - 1 = 7$ for each location are also needed and given in Table 1. Note that the autocorrelation of rainfall was found statistically insignificant and thus the autocorrelation coefficient were set zero; on the contrary, the autocorrelation of runoff is significant for all three basins. To increase readability and understandability of the data, the form of the statistics given in Table 1 differs from that displayed in the equations (e.g., coefficients of skewness instead of third moments; matrices of correlation coefficients instead of covariance matrices); the transformation between the two forms is direct and given in textbooks.

There seems to be nothing strange about the statistics given in Table 1 as the cross-correlations and autocorrelations are not too high; the skewnesses are not zero and also are not too high (they do not exceed 1.75). However, the matrix $\mathbf{c}$ of (12) (and $\mathbf{c}'$ of (14)) is not positive definite, and thus there does not exist an exact solution for $\mathbf{b}$ (or $\mathbf{b}'$). Notably, the absence of positive definiteness is met in 10 out of 12 cases (months) for the monthly PAR(1) model as well as in the annual AR(1) model, in the case study examined. This indicates that the problem studied in this paper might be quite frequent in multivariate stochastic simulation problems.

To start remedying the problem for s = 8 we need an initial solution $\mathbf{b}'^{[0]}$. To this aim we apply the *Lane*'s [1979] procedure, which is appropriate for positive semidefinite matrices (see also *Salas et al.* [1988, p. 87]; *Lane and Frevert* [1990, p. V-15]). This procedure derives a lower triangular $\mathbf{b}'$ given the matrix $\mathbf{c}'$, first calculating the first matrix column, then the second, third, etc. When the matrix $\mathbf{c}'$ is not positive definite (as in our case), at some column (in our case at column 5) the procedure assigns all matrix elements equal to zero and, besides, the estimated matrix $\mathbf{b}'$ no longer obeys $\mathbf{b}'\ \mathbf{b}'^{T} = \mathbf{c}'$. Furthermore, a zero column within $\mathbf{b}'$ is

also transferred in $\mathbf{b}'^{(3)}$ so that $(\mathbf{b}'^{(3)})^{-1}$ that appears in (19) does not exist. Thus, the vector of the skewness coefficients of the auxiliary variables, $\boldsymbol{\xi}$, becomes infinite. To avoid this, it suffices to set the diagonal element of the column with zero elements equal to some arbitrary value b'_{min} (so that $0 < b'_{min} < 1$). There is no reason to investigate thoroughly what would be the most appropriate value of b'_{min} because this is used to establish an initial solution only; this initial solution is then modified by the optimization procedure. Here we have used $b'_{min} = 0.05$ and the derived solution is thereafter referred to as initial solution 0. The elements of $\mathbf{b}'$ for that solution are shown graphically in Figure 1(a). The resulting values of the objective function and its three components are shown in Table 2 (where for the sake of readability we display the square roots of the components). We observe that for the initial solution 0 the maximum coefficient of skewness of the auxiliary variables is extremely high, i.e., $\max_j(\xi_j) \approx ||\boldsymbol{\xi}|| = 29089.50$. Apparently, such skewness cannot be reached in any generated sample. Also, we observe that $||\mathbf{d}||$ is not zero and it could not be so because $\mathbf{c}'$ is not positive definite; also $||\mathbf{d}^*||$ is not zero.

For comparison we have also determined another initial solution referred to as initial solution 0a, and also shown graphically in Figure 1(b). For this solution we have used a more complicated procedure outlined by *Koutsoyiannis* [1992]. This procedure imposes a lower limit in each diagonal element of $\mathbf{b}'$ to avoid extremely high skewness coefficients, assures the preservation of the diagonal elements of $\mathbf{c}'$ (so that $||\mathbf{d}^*|| \approx 0$), and also attempts to preserve the highest cross-correlation coefficient in each matrix row simultaneously reducing the other cross-correlation coefficients. As shown in Table 2, the maximum coefficient of skewness of the initial solution 0a has been reduced to 18.31, a value quite less than 29089.50 of the initial solution 0; $||\mathbf{d}^*||$ is almost zero in the initial solution 0a whereas $||\mathbf{d}||$ unavoidably remains positive.

Given the initial solutions 0 and 0a we can directly proceed to the optimization procedure described in section 4. To this aim we use the objective function (29) and its

derivatives (36) with weights $\lambda_1 = 1$, $\lambda_2 = 10^3$, $\lambda_3 = 10^{-5}$ as explained in section 3. The evolution of the objective function and its three components for the first forty iterations are shown in Figure 2 for both initial solutions 0 and 0a. We observe in Figure 2 that the unreasonable value of $\|\boldsymbol{\xi}\| = 29089.50$ of the initial solution 0 reduces rapidly in the first 5-10 iterations towards more reasonable values. Also the positive value of $\|\mathbf{d}^*\|$ of the initial solution 0 reduces rapidly in the first 4 iterations to a value close to zero. As expected, the performance of the evolution of the initial solution 0a seems better than that of the initial solution 0 at the same iteration number. However, the finally obtained values of the objective function and its components are the same for both initial solutions. This indicates that there is no need to use a complicated procedure to obtain an initial solution $\mathbf{b}'^{[0]}$. It suffices to use the simple *Lane*'s [1979] procedure with the modification that the diagonal elements of $\mathbf{b}'$ are not allowed to take values smaller than b'_{min}. In the initial solution 0 of this example we assumed that $b'_{min} = 0.05$; however, we suggest a much higher value, e.g., $b'_{min} = 0.5$, to avoid unreasonably high initial skewness coefficients and accelerate convergence.

The final solutions $\mathbf{b}'$, referred to as final solutions 1 and 1a, which were obtained by the optimization procedure starting with the initial solutions 0 and 0a, respectively, are shown graphically in Figure 1(c, d). They are almost indistinguishable but they are not exactly the same. The resulting values of the objective function and its components, shown in Table 2, are the same for both final solutions 1 and 1a (in a normal computer precision). This may mean that there are (at least) two close local minima of the objective function. It may also be interpreted differently, that is, both obtained solutions are approximations of a single global minimum that cannot be located exactly due to computer precision limitations. Nevertheless, the exact theoretical meaning of the nature of the two solutions is not important; what is important is that we have one or more reasonable matrices $\mathbf{b}'$ and other related parameters that can be used directly for stochastic simulation. This simulation will preserve the variances of the variables because $\|\mathbf{d}^*\| = 0$, as well as their coefficients of skewness, because the

obtained $\max_j(\xi_j) = 5.37$ is low enough to be achieved for the auxiliary variables (assuming that the synthetic sample size will have length greater than about 50). However, the simulation will not preserve exactly the cross-correlation coefficients (since $||\mathbf{d}|| = 0.1404 > 0$), but this is an unavoidable consequence of the absence of positive definiteness of $\mathbf{c}'$. To have an indication of how large the errors in preserving these cross-correlation coefficients are, we have performed the inverse calculations, i.e., given $\mathbf{b}'$ we solved (12) for Cov[$\mathbf{Y}$, $\mathbf{Y}$] (where $\mathbf{Y} \equiv \mathbf{X}^s$) also replacing $\mathbf{c}$ with $(\mathbf{h}^{-1}\,\mathbf{b}'\,\mathbf{b}'^{T}\,\mathbf{h}^{-1})$. The resulting cross-correlation coefficients are shown in Table 3 among with their differences from the "correct" values of Table 1, as the latter were estimated from the historical data. We observe that errors are almost negligible (i.e. within ±0.03).

To acquire an indication of how low could ultimately be the error in preserving the cross-correlations if we ignore completely the preservation of skewness, we performed another optimization setting $\lambda_3 = 0$ in the objective function. The resulting final solution 2 is shown graphically in Figure 1(e) and the relevant values of the objective function and its components are shown in Table 2. We observe that the further reduction in $||\mathbf{d}||$ is not impressive ($||\mathbf{d}|| = 0.1386$ against $||\mathbf{d}|| = 0.1404$ of final solution 1).

As a further investigation, we also accessed the ultimate lowest value of the coefficient of skewness $\max_j(\xi_j)$ by ignoring the error in preserving cross-covariances. Thus, we performed another minimization of the objective function setting $\lambda_1 = 0$, $\lambda_2 = 10^3$, $\lambda_3 = 10^{-5}$. The resulting final solution 3 is shown graphically in Figure 1(f) and the relevant values of the objective function and its components are shown in Table 2. Again, the further reduction in $\max_j(\xi_j)$ is not impressive ($\max_j(\xi_j) = 4.57$ against $\max_j(\xi_j) = 5.37$ of final solution 1).

The results of the above investigations indicate that there is no strong conflict between the objectives of preserving the coefficient of skewness and the cross-correlation coefficients. The same indication is also acquired from Figure 1, where we observe that all final solutions

1, 1a, 2, and 3 are similar with each other although they differ significantly from the initial solutions 0 and 0a.

6. Summary, conclusions and discussion

A new method is proposed for decomposing covariance matrices that appear in the parameter estimation phase of all multivariate stochastic models in hydrology. This method applies not only to positive definite covariance matrices (as do the typical methods of the literature) but to indefinite matrices, too, that often appear in stochastic hydrology. It is also appropriate for preserving the skewness coefficients of the model variables as it accounts for the resulting coefficients of skewness of the auxiliary (noise) variables used by the stochastic model. The method is formulated in an optimization framework with the objective function being composed of three components aiming at (a) complete preservation of the variances of variables (b) optimal approximation of the covariances of variables, in case that complete preservation is not feasible due to inconsistent (i.e., not positive definite) structure of the covariance matrix, and (c) preservation of the skewness coefficients of the model variables by keeping the skewness of the auxiliary variables as low as possible. Analytical expressions for the derivatives of this objective function are derived, which allow the development of an effective nonlinear optimization algorithm using the Steepest Descent or the Conjugate Gradient methods.

An advantage of the method is its unique formulation, applicable on both positive definite or indefinite matrices, and symmetric or skewed distributions of variables. Besides, the weighting factors incorporated in the objective function allow for giving more or less emphasis to each one of its three components. In case of a consistent (positive definite) covariance matrix of variables having symmetric (e.g., Gaussian) distributions, the method will result in the triangular decomposition of the matrix, which is the simplest among all solutions. If the covariance matrix is indefinite whereas the variables have still symmetric

distributions, the method will result in a decomposed (square root) matrix, which corresponds to least significant alteration, or the best approximation, of the original covariance matrix. If the distributions of variables are skewed, then the method solution (either for consistent or inconsistent covariance matrices) will be appropriately modified to simultaneously account for avoiding too high coefficients of skewness of the auxiliary variables.

The real-world application examined for the sake of illustration and numerical exploration of the method indicates its very satisfactory performance both in approaching the covariances of an inconsistent matrix and in yielding low coefficients of skewness of the auxiliary variables, although the initial coefficients of skewness were extremely and unreasonably high. Moreover, it reveals that there is no strong conflict between the objectives of preserving the covariances and the coefficients of skewness. Finally, it indicates a stable behavior of the optimizing algorithm.

The stochastic model that was used as a ground for the development of the method (section 2) is generalized so as to represent many of the typical models of the literature. However, it does not cover explicitly all possible multivariate models, for example, the ARMA and PARMA models [*Stedinger et al.*, 1985; *Salas*, 1993; among others]. In these types of models the problem of decomposing covariance matrices still appears [*Salas*, 1993, pp. 19.29-19.30], and the proposed method may also be used. Some adaptations of the equations are needed, mainly those regarding the coefficients of skewness. Generally, the method allows of adaptations not only in cases of different types of models but also in situations where different conditions arise for the solution matrix. As a simplified example for a conditional rearrangement of the method, let us consider the case where the model variable of one location has known value (e.g., given as an output of another model). Obviously, this implies a linear constraint for the elements of a single row of the decomposed matrix. This constrained can be easily incorporated in the optimization framework, the simplest way being the appending of a penalty term into the objective function.

Appendix: Proof of equations (31)-(33)

From (20) we have

$$d_{kl} = \sum_{r=1}^{n} b'_{kr}\, b'_{lr} - c'_{kl} \tag{A1}$$

so that

$$\frac{\partial d_{kl}}{\partial b'_{ij}} = \sum_{r=1}^{n} b'_{kr} \frac{\partial b'_{lr}}{\partial b'_{ij}} + \sum_{r=1}^{n} \frac{\partial b'_{kr}}{\partial b'_{ij}} b'_{lr} \tag{A2}$$

Clearly,

$$\frac{\partial b'_{lr}}{\partial b'_{ij}} = \delta_{li}\, \delta_{rj} \tag{A3}$$

where

$$\delta_{ij} := \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \tag{A4}$$

Therefore,

$$\frac{\partial d_{kl}}{\partial b'_{ij}} = \sum_{r=1}^{n} b'_{kr}\, \delta_{li}\, \delta_{rj} + \sum_{r=1}^{n} b'_{lr}\, \delta_{ki}\, \delta_{rj} \tag{A5}$$

which results in

$$\frac{\partial d_{kl}}{\partial b'_{ij}} = b'_{kj}\, \delta_{li} + b'_{lj}\, \delta_{ki} \tag{A6}$$

The partial derivative of $||\mathbf{d}||^2$ with respect to b'_{ij} will be

$$\frac{\partial ||\mathbf{d}||^2}{\partial b'_{ij}} = \sum_{k=1}^{n} \sum_{l=1}^{n} 2\, d_{kl} \frac{\partial d_{kl}}{\partial b'_{ij}} = 2 \sum_{k=1}^{n} \sum_{l=1}^{n} d_{kl}\, (b'_{kj}\, \delta_{li} + b'_{lj}\, \delta_{ki}) \tag{A7}$$

or

$$\frac{\partial \|\mathbf{d}\|^2}{\partial b'_{ij}} = 2 \sum_{k=1}^{n} d_{ki}\, b'_{kj} + 2 \sum_{l=1}^{n} d_{il}\, b'_{lj} \tag{A8}$$

and, because **d** is symmetric,

$$\frac{\partial \|\mathbf{d}\|^2}{\partial b'_{ij}} = 4 \sum_{k=1}^{n} d_{ik}\, b'_{kj} \tag{A9}$$

We observe that the sum in the right-hand side of (A9) is the (i, j)th element of the matrix **d b** and, thus, this proves (31).

In a similar manner (and also considering (22)), the partial derivative of $\|\mathbf{d}^*\|^2$ with respect to b'_{ij} is

$$\frac{\partial \|\mathbf{d}^*\|^2}{\partial b'_{ij}} = \sum_{k=1}^{n} 2\, d_{kk} \frac{\partial d_{kk}}{\partial b'_{ij}} = 2 \sum_{k=1}^{n} d_{kk}\, (b'_{kj}\, \delta_{ki} + b'_{kj}\, \delta_{ki}) = 4 \sum_{k=1}^{n} d_{kk}\, b'_{kj}\, \delta_{ki} = 4\, d_{ii}\, b'_{ij} \tag{A10}$$

which proves (32).

To determine the partial derivatives of $\|\xi\|_p^2$ we must first find the partial derivatives of ξ. We denote

$$\mathbf{g} := (\mathbf{b}'^{(3)})^{-1} \tag{A11}$$

so that (19) becomes

$$\xi = \mathbf{g}\, \boldsymbol{\varphi} \tag{A12}$$

Since $\boldsymbol{\varphi}$ is a vector of constants we have

$$\frac{\partial \xi}{\partial b'_{ij}} = \frac{\partial \mathbf{g}}{\partial b'_{ij}}\, \boldsymbol{\varphi} \tag{A13}$$

(A11) can be written as

$$\mathbf{g}\, \mathbf{b}'^{(3)} = \mathbf{I} \tag{A14}$$

where **I** is the identity matrix. Therefore,

$$\frac{\partial \mathbf{g}}{\partial b'_{ij}} \mathbf{b}'^{(3)} + \mathbf{g} \frac{\partial \mathbf{b}'^{(3)}}{\partial b'_{ij}} = \mathbf{O} \tag{A15}$$

where **O** is the zero matrix (i.e., with all its elements zero), and thus

$$\frac{\partial \mathbf{g}}{\partial b'_{ij}} = -\mathbf{g} \frac{\partial \mathbf{b}'^{(3)}}{\partial b'_{ij}} (\mathbf{b}'^{(3)})^{-1} = -\mathbf{g} \frac{\partial \mathbf{b}'^{(3)}}{\partial b'_{ij}} \mathbf{g} \tag{A16}$$

Combining (A13), (A16), and (A12) we get

$$\frac{\partial \boldsymbol{\xi}}{\partial b'_{ij}} = -\mathbf{g} \frac{\partial \mathbf{b}'^{(3)}}{\partial b'_{ij}} \mathbf{g}\, \boldsymbol{\varphi} = -\mathbf{g} \frac{\partial \mathbf{b}'^{(3)}}{\partial b'_{ij}} \boldsymbol{\xi} \tag{A17}$$

Consequently,

$$\frac{\partial \xi_k}{\partial b'_{ij}} = -\sum_{r=1}^{n} \sum_{s=1}^{n} g_{ks} \frac{\partial b'^{3}_{sr}}{\partial b'_{ij}} \xi_r \tag{A18}$$

and since

$$\frac{\partial b'^{3}_{sr}}{\partial b'_{ij}} = 3\, b'^{2}_{sr}\, \delta_{si}\, \delta_{rj} \tag{A19}$$

we get

$$\frac{\partial \xi_k}{\partial b'_{ij}} = -3 \sum_{r=1}^{n} \sum_{s=1}^{n} g_{ks}\, b'^{2}_{sr}\, \delta_{si}\, \delta_{rj}\, \xi_r \tag{A20}$$

or

$$\frac{\partial \xi_k}{\partial b'_{ij}} = -3 \sum_{r=1}^{n} g_{ki}\, b'^{2}_{ir}\, \delta_{rj}\, \xi_r \tag{A21}$$

and finally,

$$\frac{\partial \xi_k}{\partial b'_{ij}} = -3\, g_{ki}\, b'^{2}_{ij}\, \xi_j \tag{A22}$$

Now we can proceed to the calculation of the derivatives of $\|\boldsymbol{\xi}\|_p^2$. Assuming that p is even (26) writes

$$\|\boldsymbol{\xi}\|_p^2 = \left(\sum_{k=1}^{n} \xi_k^p \right)^{2/p} \tag{A23}$$

so that

$$\frac{\partial \|\boldsymbol{\xi}\|_p^2}{\partial b'_{ij}} = \frac{2}{p} \left(\sum_{k=1}^{n} \xi_k^p \right)^{2/p-1} \frac{\partial}{\partial b'_{ij}} \sum_{k=1}^{n} \xi_k^p = \frac{2}{p} \|\boldsymbol{\xi}\|_p^{2-p} \; p \sum_{k=1}^{n} \xi_k^{p-1} \frac{\partial \xi_k}{\partial b'_{ij}} \tag{A24}$$

Combining (A22) and (A24) we get

$$\frac{\partial \|\boldsymbol{\xi}\|_p^2}{\partial b'_{ij}} = -6 \|\boldsymbol{\xi}\|_p^{2-p} \; b'^{\,2}_{ij} \, \xi_j \sum_{k=1}^{n} \xi_k^{p-1} g_{ki} \tag{A25}$$

whereas combining (A11) and (35) we have

$$\sum_{k=1}^{n} \xi_k^{p-1} g_{ki} = \psi_i \tag{A26}$$

Also considering (34) we get

$$b'^{\,2}_{ij} \, \xi_j \sum_{k=1}^{n} \xi_k^{p-1} g_{ki} = b'^{\,2}_{ij} \, \xi_j \, \psi_i = w_{ij} \tag{A27}$$

so that, finally,

$$\frac{\partial \|\boldsymbol{\xi}\|_p^2}{\partial b'_{ij}} = -6 \|\boldsymbol{\xi}\|_p^{2-p} \; w_{ij} \tag{A28}$$

which proves (33).

Acknowledgments. The research leading to this paper was performed within the framework of the project *Evaluation and Management of the Water Resources of Sterea Hellas*, Phase 3, project no 9476702, funded by the Greek Ministry of Environment, Regional Planning and Public Works, Directorate of Water Supply and Sewage. The author wishes to thank the director Th. Bakopoulos, the staff of this Directorate, and E. Tiligadas of the Directorate D7 for the support of this research. Thanks are also due to I. Kioustelidis for his helpful suggestions and I. Nalbantis for his comments. Finally, the author is grateful to the Associate Editor P. Rasmussen and the reviewer E. Todini for their constructive comments that were very helpful for an improved, more complete and more accurate presentation.

References

Bras, R. L. and I. Rodriguez-Iturbe, *Random functions in hydrology*, Addison-Wesley, 1985.

Grygier, J. C. and J. R. Stedinger, SPIGOT, A synthetic streamflow generation software package, Technical description, School of Civil and Environmental Engineering, Cornell University, Ithaca, NY., Version 2.5, 1990.

Kirby, W., Algebraic boundness of sample statistics, *Water Resour. Res.*, 10(2) 220-222, 1974.

Koutsoyiannis, D., A nonlinear disaggregation model with a reduced parameter set for simulation of hydrologic series, *Water Resour. Res.*, 28(12), 3175-3191, 1992.

Koutsoyiannis, D., and A. Manetas, Simple disaggregation by accurate adjusting procedures, *Water Resources Research*, 32(7), 2105-2117, 1996

Lane, W. L., Applied Stochastic Techniques, User's Manual, Bureau of Reclamation, Engineering and Research Center, Denver, Co., 1979.

Lane, W. L. and D. K. Frevert, Applied Stochastic Techniques, User's Manual, Bureau of Reclamation, Engineering and Research Center, Denver, Co., Personal Computer Version, 1990.

Marlow, W. H., *Mathematics for Operations Research*, Dover Publications, New York, 1993.

Matalas, N. C. and J. R. Wallis, Correlation constraints for generating processes, *Proceedings of the International Symposium on Mathematical Models in Hydrology*, pp. 697-707, Warsaw, Poland, July 1971.

Matalas, N. C. and J. R. Wallis, Generation of synthetic flow sequences, in *Systems Approach to Water Management*, edited by A. K. Biswas, McGraw-Hill, New York, 1976.

Mays, L. W., and Y.-K. Tung, Systems analysis, in *Water Resources Handbook*, edited by L. W. Mays, McGraw-Hill, New York, 1996.

Mejía, J. M., and J. Millán, Una Metodología para tratar el problema de matrices inconsistentes en la generatión multivariada de series hidrológicas, *VI Congreso Latino Americano de Hydraúlica*, Bogotá, Colombia, 1974.

Nalbantis, I., and D. Koutsoyiannis, A parametric rule for planning and management of multiple-reservoir systems, *Water Resour. Res.*, 33(9), 2165-2177, 1997.

Pierre, D. P., *Optimization Theory With Applications*, Dover, New York, 1986.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, Cambridge University Press, Cambridge, 1992.

Rasmussen, P. F., J. D Salas, L. Fagherazzi, J.-C. Rassam, and B. Bobée, Estimation and validation of contemporaneous PARMA models for streamflow simulation, Water Resour. Res., 32(1), 3151-3160, 1996.

Salas, J. D., Analysis and modeling of hydrologic time series, Chapter 19, *Handbook of hydrology*, edited by D. Maidment, McGraw-Hill, New York, 1993.

Salas, J. D., J. W. Delleur, V. Yevjevich, and W. L. Lane, *Applied Modeling of Hydrologic Time Series*, Water Resources Publications, Littleton, Colo. 1988.

Slack, J. R., I would if I could (Self-denial by conditional models), *Water Resour. Res.*, 9(1), 247-249, 1973.

Stedinger, J. R., D. P. Lettenmaier, and R. M. Vogel, Multisite ARMA(1,1) and disaggregation models for annual streamflow generation, *Water Resour. Res.*, 21(4) 497-509, 1985.

Todini, E., The preservation of skewness in linear disaggregation schemes, *J. Hydrol.*, 47, 199-214, 1980.

Valencia, D. and J. C. Schaake, A disaggregation model for time series analysis and synthesis, Report no. 149, Ralph M. Parsons Laboratory for Water Resources and Hydrodynamics, Massachusetts Institute of Technology, Cambridge, Mass., 1972.

Valencia, D. and J. C. Schaake, Disaggregation processes in Stochastic Hydrology, *Water Resour. Res.*, 9(3) 211-219, 1973.

Wallis, J. R., N. Matalas, and J. R. Slack, Just a moment!, *Water Resour. Res.*, 10(2) 211-219, 1974.

List of Figures

Figure 1 Plots of the elements of matrices **b**′ for the various solutions examined in the case study: (a) initial solution 0; (b) initial solution 0a; (c) final solution 1; (d) final solution 1a; (e) final solution 2; (f) final solution 3 (see text).

Figure 2 Evolution of the value of the square root of the objective function θ and its three components $\|\mathbf{d}\|$, $\|\mathbf{d}^*\|$, and $\|\boldsymbol{\xi}\|$ for the first forty iterations of the optimization procedure starting with initial solutions 0 and 0a (see text).

List of tables

Table 1 Sample statistics of the monthly rainfall and runoff data used for the case study for subperiods $s = 8$ (May) and $s - 1 = 7$ (April)

Table 2 Values of the square root of the objective function θ and its three components $||\mathbf{d}||$, $||\mathbf{d}^*||$, and $||\boldsymbol{\xi}||$ (also, in comparison with $\max_j(\xi_j)$) for the initial and final solutions of the case study examined

Table 3 Cross-correlation coefficients Corr[X, X] resulting from $\mathbf{b}'$ of the final solution 1a. The values in parentheses are the differences from the corresponding values of Table 1.

Table 1 Sample statistics of the monthly rainfall and runoff data used for the case study for subperiods $s = 8$ (May) and $s - 1 = 7$ (April)

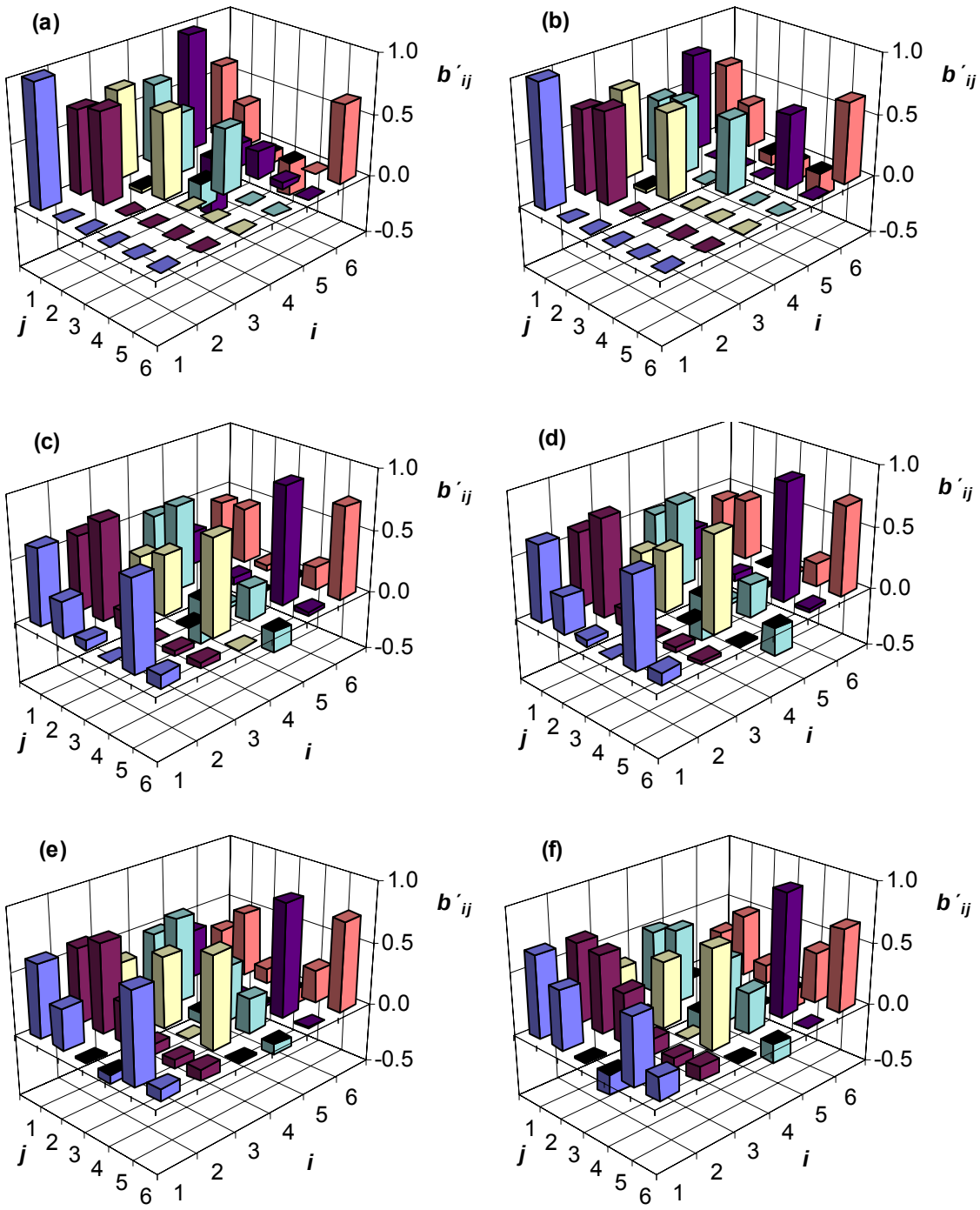
Location		$i = 1$ (Evinos runoff)	2 (Evinos rainfall)	3 (Mornos runoff)	4 (Mornos rainfall)	5 (Yliki runoff)	6 (Yliki rainfall)
Means							
$E[X_i^{s-1}]$		97.3	111.2	59.8	100.4	20.1	31.5
$E[X_i^s]$		53.1	69.5	43.2	65.3	9.4	19.3
Standard deviations							
StDev$[X_i^{s-1}]$		35.0	57.6	17.4	53.6	10.7	27.1
StDev$[X_i^s]$		20.2	30.8	18.8	26.5	7.2	14.4
Coefficients of skewness							
CS$[X_i^{s-1}]$		0.72	1.04	0.58	1.07	1.06	1.72
CS$[X_i^s]$		0.76	0.89	0.87	0.81	1.53	1.49
Cross-correlation coefficients							
Corr$[X_i^{s-1}, X_j^{s-1}]$	$j = 1$	1.00	0.76	0.85	0.56	0.22	0.16
	2	0.76	1.00	0.67	0.90	0.30	0.43
	3	0.85	0.67	1.00	0.58	0.69	0.14
	4	0.56	0.90	0.58	1.00	0.37	0.50
	5	0.22	0.30	0.69	0.37	1.00	0.83
	6	0.16	0.43	0.14	0.50	0.83	1.00
Corr$[X_i^s, X_j^s]$	$j = 1$	1.00	0.54	0.76	0.53	0.57	0.49
	2	0.54	1.00	0.29	0.81	0.15	0.67
	3	0.76	0.29	1.00	0.20	0.75	0.23
	4	0.53	0.81	0.20	1.00	0.27	0.45
	5	0.57	0.15	0.75	0.27	1.00	0.22
	6	0.49	0.67	0.23	0.45	0.22	1.00
Autocorrelation coefficients							
Corr$[X_i^s, X_i^{s-1}]$		0.60	0	0.78	0	0.80	0

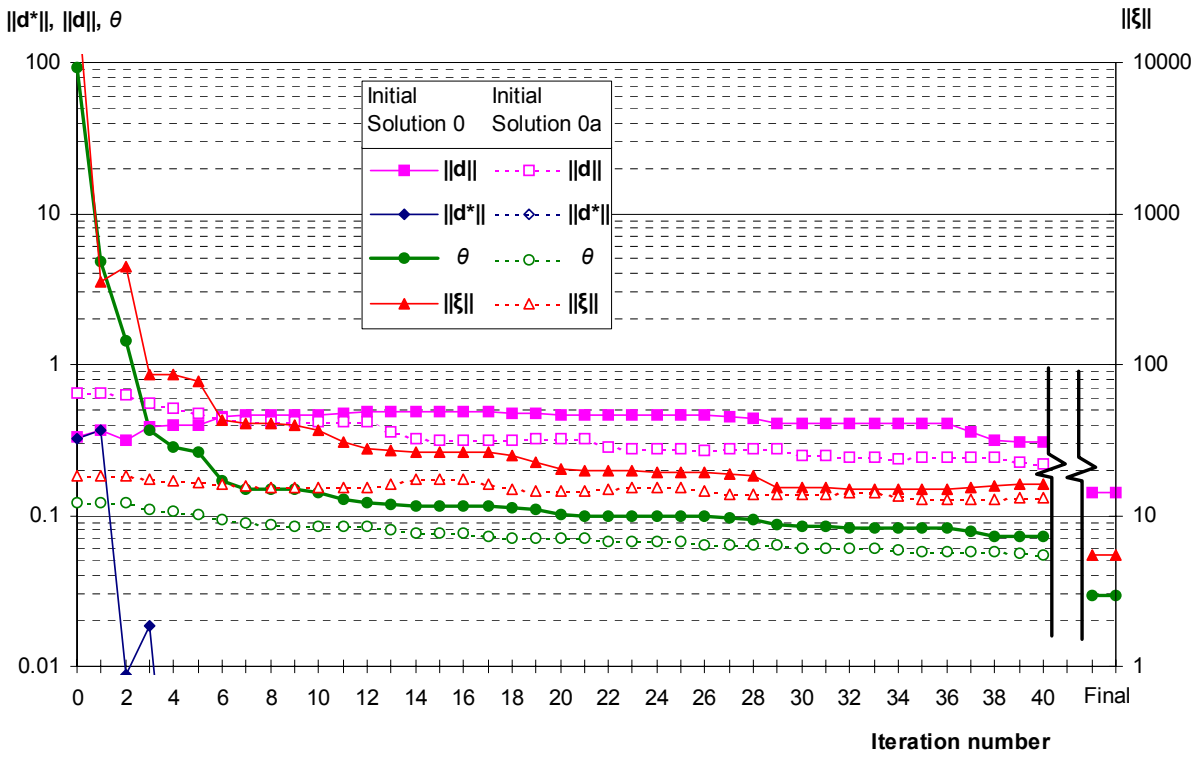
Table 2 Values of the square root of the objective function θ and its three components $\|\mathbf{d}\|$, $\|\mathbf{d}^*\|$, and $\|\boldsymbol{\xi}\|$ (also, in comparison with $\max_j(\xi_j)$) for the initial and final solutions of the case study examined.

	Initial solution 0	Initial solution 0a	Final solutions 1 and 1a – Optimum for the combination of $\|\mathbf{d}\|$ and $\|\boldsymbol{\xi}\|_p$ ($\lambda_1 = 1$, $\lambda_2 = 10^3$, $\lambda_3 = 10^{-5}$)	Final solution 2 Optimum for $\|\mathbf{d}\|$ ($\lambda_1 = 1$, $\lambda_2 = 10^3$, $\lambda_3 = 0$)	Final solution 3 Optimum for $\|\boldsymbol{\xi}\|_p$ ($\lambda_1 = 0$, $\lambda_2 = 10^3$, $\lambda_3 = 10^{-5}$)
$\|\mathbf{d}^*\|$	0.3238	0.0012	0.0000	0.0000	0.0000
$\|\mathbf{d}\|$	0.3297	0.6394	0.1404	0.1386	0.8136
$\|\boldsymbol{\xi}\|$	29089.50	18.31	5.50	7.94	4.58
$\max_j(\xi_j)$	29089.50	18.31	5.37	7.19	4.57
θ	4.1806	0.1076	0.0292	0.0231	0.0145

Table 3 Cross-correlation coefficients Corr[X_i^s, X_j^s] resulting from **b**′ of the final solution 1a. The values in parentheses are the differences from the corresponding values of Table 1.

	$i = 1$	2	3	4	5	6
$j = 1$	1.00 (0.00)	0.51(–0.03)	0.77 (0.01)	0.54 (0.01)	0.54(–0.03)	0.49 (0.00)
2	0.51(–0.03)	1.00 (0.00)	0.28(–0.01)	0.80(–0.01)	0.17 (0.02)	0.67 (0.00)
3	0.77 (0.01)	0.28(–0.01)	1.00 (0.00)	0.20 (0.00)	0.74(–0.01)	0.23 (0.00)
4	0.54 (0.01)	0.80(–0.01)	0.20 (0.00)	1.00 (0.00)	0.26(–0.01)	0.45 (0.00)
5	0.54(–0.03)	0.17 (0.02)	0.74(–0.01)	0.26(–0.01)	1.00 (0.00)	0.22 (0.00)
6	0.49 (0.00)	0.67 (0.00)	0.23 (0.00)	0.45 (0.00)	0.22 (0.00)	1.00 (0.00)


(a)
(b)
(c)
(d)
(e)
(f)
b´ij
1.0
0.5
0.0
-0.5
i
j
1
2
3
4
5
6


||d*||, ||d||, θ
||ξ||
100
10
1
0.1
0.01
10000
1000
100
10
1
Initial Solution 0
Initial Solution 0a
||d||
||d*||
θ
	ξ	
	d	
	d*	
θ
	ξ	
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
Final
Iteration number