

REVISTA
DE LA
ACADEMIA
DE
CIENCIAS

Exactas
Físicas
Químicas y
Naturales

DE
ZARAGOZA



Serie 2.^a
Volumen 50

1995

INDICE DE MATERIAS

	<i>Págs.</i>
I. K. Argyros and D. Chen. — «On a two-point Newton method in Banach spaces of order four and applications»	5
I. K. Argyros. — «Improved error bounds for fast two-point Newton methods of order three»	15
B. K. Sen and A. K. Chongdar. — «Lie theory and some generating functions of extended Jacobi polynomials»	21
A. K. Chongdar. — «On the extension of trilateral generating functions of extended Jacobi polynomials»	27
D. Chen, I. K. Argyros and Q. Qian. — «A Study on the Order of Convergence of a Rational Iteration for Solving Quadratic Operator Equations in a Banach Space»	35
P. Stanimirović. — «General determinantal representation of pseudoinverses and its computation»	41
— M. Ruiz Espejo. — «Una relación entre cuasicovarianzas muestral y poblacional»	51
— M. Ruiz Espejo, J. Santos Peñas y M. Rueda García. — «Optimum linear integration of independent sample means from finite populations»	55
— M. Ruiz Espejo. — «Un estimador más preciso que la media muestral para estudios analíticos»	61
— M. Ruiz Espejo. — «Sesgo de intencionalidad en la selección de la estrategia muestral»	65
— L. Agud y O. R. García. — «Energía de una señal usando el teorema de Shannon»	69
— R. C. Díaz Ara y A. Novales Claraco. — «Determinación de fluoruros en baños de galvanotecnia por electrodos selectivos de iones»	85
— O. Artieda, J. Mandado y J. Herrero. — «Las "areniscas" dolomíticas del área de Quinto (Zaragoza). Modelo petrogenético»	93
— L. F. Auqué, J. Mandado, M. J. Gimeno, P. L. López y V. Vallés. — «Cálculos de simulación de la alteración de microclina. I. Análisis comparativo de códigos de modelización geoquímica»	103
— L. F. Auqué, P. L. López, M. J. Gimeno, J. Mandado, V. Vallés, A. Zamora y J. Gómez. — «Análisis comparativo de códigos de modelización geoquímica y bases de datos termodinámicos. I. Simulación de la alteración de microclina»	115
— L. F. Auqué, P. L. López, M. J. Gimeno, J. Mandado, V. Vallés, A. Zamora y J. Gómez. — «Análisis comparativo de códigos de modelización geoquímica y bases de datos termodinámicos. II. Simulación de procesos de dedolomitización»	129
— L. F. Auqué, P. López, M. J. Gimeno, J. Mandado, J. Gómez y A. Zamora. — «Pautas de especiación de las Tierras Raras en las soluciones ácidas de Osamu-Utsuni, Poços de Caldas (Brasil)»	145

ON A TWO-POINT NEWTON METHOD IN BANACH SPACES OF ORDER FOUR AND APPLICATIONS *

Ioannis K. Argyros[†] and Dong Chen^{††}

[†] Cameron University. Department of Mathematics. LAWTON, OK 73505 U.S.A.

^{††} University of Arkansas. Department of Mathematics. FAYETTEVILLE, AR 72701, U.S.A

Abstract

Sufficient conditions are given for the convergence of a two-point Newton method to a zero of a linear operator equation in a Banach space. The order of convergence is four.

1. Introduction

In this study we are concerned with the problem of approximating a locally unique zero x^* of the equation

$$F(x) = 0 \quad (1)$$

in a Banach space E_1 , where F is a nonlinear operator defined on some convex subset D of E_1 with values in another Banach space E_2 .

The convergence of single-step methods, like Newton's method as well as Newton-like methods to a zero x^* of equation (1) has been studied extensively in a Banach space setting [1]-[19]. But the convergence analysis for multipoint methods is less developed, although the fundamental theory was developed several years ago, [15],[16] and [18]. The reason is that the expression $F(x)$ cannot easily be dominated by a real scalar function. It is well known, from the efficiency index point view [15] that multipoint methods are faster than single-step methods.

Here, in particular we consider a Newton two-step method of the form

$$y_n = x_n - F'(x_n)^{-1} F(x_n) \quad (2)$$

$$x_{n+1} = y_n - F'(y_n)^{-1} F(y_n) \quad (3)$$

*AMS (MOS) subject classification: 47H17, 65H10, 65J15

for some arbitrary $x_0 \in D$ and for all $n \geq 0$. The linear operator $F'(x_n)$ is the Fréchet-derivative of F evaluated at $x = x_n$.

Using standard Kantorovich assumptions we will show that the two-point method (2)-(3) converges to a zero x^* of equation (1) with order 4.

2. Convergence Analysis

Let $x_0 \in D$ be arbitrary and for $R > 0$ such that $U(x_0, R) = \{x \in E_1 / \|x - x_0\| \leq R\} \subseteq D$. We assume that

$$\|F''(x)\| \leq M \quad (4), \quad \|F'(x) - F'(y)\| \leq K \|x - y\| \quad (5),$$

for all $x, y \in D$. It is convenient to introduce the constants

$$\eta \geq \|y_0 - x_0\|, \quad \beta \geq \|F'(x_0)^{-1}\|, \quad t_0 = 0, \quad h = M\eta\beta \quad (6),$$

$$r_1 = \frac{1 - \sqrt{1 - 2h}}{h} \eta \quad (7), \quad r_2 = \frac{1 + \sqrt{1 - 2h}}{h} \eta \quad (8), \quad \theta = \frac{r_1}{r_2} \quad (9),$$

and the scalar iterations

$$s_n = t_n - \frac{g(t_n)}{g'(t_n)} \quad (10), \quad t_{n+1} = s_n - \frac{g(s_n)}{g'(s_n)} \quad (11),$$

for all $n \geq 0$, where

$$g(t) = \frac{M}{2}t^2 - \frac{1}{\beta}t + \frac{\eta}{\beta} \quad (12).$$

Note that r_1 is the smallest zero of the equation

$$g(t) = 0 \quad (\text{if } 2h \leq 1) \quad (13)$$

We can now prove the main result:

Theorem. Let $F : D \subseteq E_1 \rightarrow E_2$ be a nonlinear operator defined on some convex subset D of a Banach space E_1 with values in another Banach space E_2 . Assume:

(a) F is twice-Fréchet differentiable on $U(x_0, r_1) \subseteq D$ for some $x_0 \in D$, and satisfies (4)-(5).

(b) The inverse of the linear operator $F'(x_0)$ for $x_0 \in D$, exists and is bounded.

The following estimates are true

$$h \leq \frac{1}{2} \quad (14), \quad \beta K(3r_1 + r_2) < 2. \quad (15)$$

Then the two-point Newton method generated by (2)-(3) is well defined, remains in $U(x_0, r_1)$ for all $n \geq 0$ and converges to a unique zero x^* of the equation $F(x) = 0$ in $U(x^*, r_2)$.

Moreover, the following estimates are true:

$$\|x_n - x^*\| \leq r_1 - t_n \quad (16), \quad \|y_n - x^*\| \leq r_1 - s_n \quad (17),$$

$$r_1 - t_n = \frac{(1 - \theta^2)\eta\theta^{4^n-1}}{1 - \theta^{4^n-1}}, \quad (18)$$

for all $n \geq 0$.

Proof. We will show that if

$$\|y_n - x_n\| \leq s_n - t_n, \quad \|F(x_n)\| \leq g(t_n), \quad \|F(y_n)\| \leq g(s_n), \quad (19)$$

$$\|F'(x_n)^{-1}\| \leq -g'(t_n)^{-1} \quad (20), \quad \|F'(y_n)^{-1}\| \leq -g'(s_n)^{-1} \quad (21),$$

then

$$\|x_{n+1} - y_n\| \leq t_{n+1} - s_n \quad (22), \quad \|y_{n+1} - x_{n+1}\| \leq s_{n+1} - t_{n+1} \quad (23),$$

$$\|F(x_{n+1})\| \leq g(t_{n+1}) \quad (24), \quad \|F(y_{n+1})\| \leq g(s_{n+1}) \quad (25),$$

for all $n \geq 0$.

Using (3), (19) and (21), we get

$$\|x_{n+1} - y_n\| \leq \|F'(y_n)^{-1}\| \|F(y_n)\| \leq -g'(s_n)^{-1} g(s_n) = t_{n+1} - s_n.$$

Hence, (22) is true.

From (2)-(4), (22) and the approximation

$$\begin{aligned} F(x_{n+1}) &= F(x_{n+1}) - F(y_n) - F'(y_n)(x_{n+1} - y_n) \\ &= \int_0^1 F''[y_n + t(x_{n+1} - y_n)](1-t)dt(x_{n+1} - y_n)^2 \end{aligned}$$

we obtain

$$\|F(x_{n+1})\| \leq \frac{1}{2}M \|x_{n+1} - y_n\|^2 \leq \frac{1}{2}M(t_{n+1} - s_n)^2 = g(t_{n+1}).$$

Hence, (24) is true.

Also by (2), (20) and (25) we get

$$\|y_{n+1} - x_{n+1}\| \leq \|F'(x_{n+1})^{-1}\| \|F(x_{n+1})\| \leq -g'(t_{n+1})^{-1} g(t_{n+1}) = s_{n+1} - t_{n+1}.$$

Hence, (23) is also true.

Similarly, from (2), (4), (23) and the approximation

$$\begin{aligned} F(y_{n+1}) &= F(y_{n+1}) - F(x_{n+1})(y_{n+1} - x_{n+1}) \\ &= \int_0^1 F''[x_{n+1} + t(y_{n+1} - x_{n+1})](1-t)dt(y_{n+1} - x_{n+1})^2 \end{aligned}$$

we get

$$\|F(y_{n+1})\| \leq \frac{1}{2}M \|x_{n+1} - y_{n+1}\|^2 \leq \frac{1}{2}M(s_{n+1} - t_{n+1})^2 = g(s_{n+1}).$$

Hence, (25) is also true.

Putting $\|Y_n\| = \|y_n - y_0\| + \|y_0 - x_0\|$, we have also the estimates

$$\begin{aligned} \|x_{n+1} - x_0\| &\leq \|x_{n+1} - y_0\| + \|y_0 - x_0\| \leq \|x_{n+1} - y_n\| + \|Y_n\| \\ &\leq \dots \leq (t_{n+1} - s_n) + (s_n - s_0) + s_0 \leq t_{n+1} \leq r_1. \end{aligned} \quad (26)$$

and

$$\begin{aligned} \|y_{n+1} - x_0\| &\leq \|y_{n+1} - y_0\| + \|y_0 - x_0\| \leq \|y_{n+1} - x_{n+1}\| + \|x_{n+1} - y_n\| + \|Y_n\| \\ &\leq \dots \leq (s_{n+1} - t_{n+1}) + (t_{n+1} - s_n) + (s_{n+1} - s_n) + (s_n - s_0) + s_0 \leq s_{n+1} \leq r_1. \end{aligned} \quad (27)$$

Moreover, from (4), (6), (26) and the estimate

$$\begin{aligned} \|F'(x_0)^{-1}\| \|F'(x_n) - F'(x_0)\| &\leq \|F'(x_0)^{-1}\| \int_0^1 \|F''[x_0 + t(x_n - x_0)]\| (x_n - x_0) dt \| \\ &\leq \beta M \|x_n - x_0\| \leq \beta M(t_n - t_0) \leq \beta M r_1 < 1, \end{aligned} \quad (28)$$

it follows from the Banach lemma on invertible operators that $F'(x_n)^{-1}$ exists and for all $n \geq 1$

$$\|F'(x_n)^{-1}\| \leq \frac{\|F'(x_n)^{-1}\|}{1 - \|F'(x_0)^{-1}\| \|F'(x_n) - F'(x_0)\|} \leq \frac{\beta}{1 - \beta M t_n} = -g'(t_n)^{-1} \quad (29)$$

Similarly, we obtain

$$\|F'(y_n)^{-1}\| \leq -g'(s_n)^{-1} \quad (30)$$

for all $n \geq 0$.

Hence, the iterates (2)-(3) are well defined for all $n \geq 0$.

It now follows that the sequence $\{x_n\}$ is Cauchy in a Banach space and as such it converges to some $x^* \in U(x_0, r_1)$, which by taking the limit as $n \rightarrow \infty$ in (2) becomes a zero of F since $F(x^*) = 0$. Moreover by (26) and (27) $x_n, y_n \in U(x_0, r_1)$. The estimates (16) and (17) follow easily from

$$\|x_{n+1} - x_n\| \leq \|x_{n+1} - y_n\| + \|y_n - x_n\| \leq t_{n+1} - s_n + s_n - t_n = t_{n+1} - t_n$$

and

$$\|y_{n+1} - y_n\| \leq \|y_{n+1} - x_{n+1}\| + \|x_{n+1} - y_n\| \leq s_{n+1} - t_{n+1} + t_{n+1} - s_n = s_{n+1} - s_n$$

for all $n \geq 0$.

Moreover, using (7), (8) and (11) we get

$$r_1 - s_n = \frac{K\beta}{2(1 - K\beta t_n)}(r_1 - t_n)^2, \quad r_1 - t_{n+1} = \frac{K\beta}{2(1 - K\beta t_n)}(r_1 - s_n)^2$$

and

$$r_2 - t_{n+1} = \frac{K\beta}{2(1 - K\beta s_n)}(r_2 - s_n)^2$$

Hence, we get

$$\frac{r_1 - t_{n+1}}{r_2 - t_{n+1}} = \left(\frac{r_1 - s_n}{r_2 - s_n} \right)^2 = \left(\frac{r_1 - t_n}{r_2 - t_n} \right)^4 = \dots = \left(\frac{r_1 - t_0}{r_2 - t_0} \right)^{4^{n+1}} = \theta^{4^{n+1}}. \quad (31)$$

Since,

$$r_2 - t_n = r_1 - t_n + (1 - \theta^2) \frac{\eta}{\theta} \quad (32)$$

the result (18) follows from (31) and (32).

To show uniqueness, let us assume that there exists another zero y^* of equation (1) in $U(x^*, r_2)$. Then from (5) and (24) we obtain

$$\begin{aligned} & \|F'(x^*)^{-1}\| \int_0^1 \|F'[x^* + t(y^* - x^*)] - F'(x^*)\| dt \\ & \leq \frac{\beta K}{1 - \beta K r_1} \|x^* - y^*\| \int_0^1 t dt \leq \frac{\beta K}{2(1 - \beta K r_1)}(r_1 + r_2) < 1 \end{aligned}$$

since $\|x^* - y^*\| \leq \|x^* - x_0\| + \|y^* - x_0\| \leq r_1 + r_2$.

It now follows from the above inequality that the linear operator

$$\int_0^1 F'[x^* + t(y^* - x^*)] dt$$

is invertible. From this and the approximation

$$F(y^*) - F(x^*) = \int_0^1 F'[x^* + t(y^* - x^*)] dt (y^* - x^*)$$

it follows $x^* = y^*$, that completes the proof of the theorem.

Remarks

(a) From the estimates

$$\|x_n - y_0\| \leq \|x_n - y_n\| + \|y_n - y_0\| \leq t_n - s_n + s_n - s_0 \leq t_n - \eta \leq r_1 - \eta$$

and

$$\begin{aligned} & \|y_{n+1} - y_0\| \leq \|y_{n+1} - x_{n+1}\| + \|x_{n+1} - y_n\| + \|y_n - y_0\| \\ & \leq s_{n+1} - t_{n+1} + t_{n+1} - s_n + s_n - s_0 \leq s_{n+1} - \eta \leq r_1 - \eta \end{aligned}$$

it follows that $x_n, y_n \in U(y_0, r_1 - \eta)$ for all $n \geq 0$.

(b) We can use the two-point method to approximate nonlinear equations with non-differentiable operators. Indeed consider the equation

$$F_1(x) = 0, \quad (33)$$

where $F_1(x) = F(x) + Q(x)$ with F as before and Q satisfying an estimate of the form

$$\|Q(x) - Q(y)\| \leq K_1 \|x - y\| \quad (34)$$

for all $x \in D$.

Note that the differentiability of Q is not assumed here. Replace F in (2) and (3) by F_1 and leave the Fréchet-derivatives as they are. Define the sequences $\{\bar{t}_n\}$ and $\{\bar{s}_n\}$ as the corresponding $\{t_n\}$ and $\{s_n\}$ given by (10) and (11) respectively. The only change will be an extra term of the form $K_1(s_n - t_n)$ and $K_1(t_{n+1} - s_n)$ added at the right hand sides of (10) and (11) respectively and multiplied by the corresponding fractions. Define also g_1 as g in (12) but add the term $K_1 t$ at the numerator. Then following the proof of the above theorem step by step we can show a similar theorem with identical hypotheses and conclusions, but holding for equation (27). (See, also [4]).

(c) Similar theorems can be proved if relation (5) is replaced by a weaker Hölder estimate of the form

$$\|F'(x - F'(y))\| \leq K \|x - y\|^p$$

for all $x, y \in D$ and some $p \in [0, 1]$. (See, also [5]).

(d) Note that using the approximation

$$\int_0^1 \{F'[x^* + t(y^* - x^*)] - F'(x^*)\} dt = \int_0^1 \int_{x^*}^{x^* + t(y^* - x^*)} F''(z) dz$$

we can show that (15) can be replaced by $M\beta(3r_1 + r_2) < 2$, which may be useful, especially when $M \leq K$.

3. Applications

In this section we use the theorem to suggest new approaches to the solution of quadratic integral equations of the form

$$x(s) = y(s) + \lambda x(s) \int_0^1 q(s, t) x(t) dt \quad (35)$$

in the space $E_1 = C[0, 1]$ of all functions continuous on the interval $[0, 1]$, with norm

$$\|x\| = \max_{0 \leq s \leq 1} |x(s)|$$

Here we assume that λ is a real number called *albedo* for scattering and the kernel $q(s, t)$ is a continuous function of two variables s, t with $0 < s, t < 1$ and satisfying the conditions

- (i) $0 < q(s, t) < 1, 0 \leq s, t \leq 1,$
- (ii) $q(s, t) + q(t, s) = 1, 0 \leq s, t \leq 1.$

The function $y(s)$ is a given continuous function defined on $[0, 1]$, and finally $x(s)$ is the unknown function sought in $[0, 1]$.

Equations of this type are closely related with the work of S. Chandrasekhar [7], (Nobel prize of physics 1983), and arise in the theories of radiative transfer, neutron transport and in the kinetic theory of gasses, [1], [2], [7].

There exists an extensive literature on equations like (35) under various assumptions on the kernel $q(s, t)$ and λ is a real or complex number. One can refer to the recent work in [1], [2] and the references there. Here we demonstrate that the theorem via the two-point Newton method (2)-(3) provides existence results for (35). Moreover the two-point Newton iterative method (2)-(3) converges faster to the solution than all the previous known ones. Furthermore a better information on the location of the solutions is given. Note that the computational cost is not higher than the corresponding one of previous methods.

For simplicity, without loss of generality, we will assume that

$$q(s, t) = \frac{s}{s+t}$$

for all $0 \leq s, t \leq 1$.

Note that q so defined satisfies (i) and (ii) above.

Let us now choose $\lambda = .25, y(s) = 1$ for all $s \in [0, 1]$; and define the operator F on E_1 by

$$F(x) = \lambda x(s) \int_0^1 \frac{s}{s+t} x(t) dt - x(s) + 1.$$

Note that every zero of the equation $F(x) = 0$ satisfies the equation (35).

Set $x_0(s) = 1$, use the definition of the first and second Fréchet-derivatives of the operator F to obtain using and the theorem.

$$K = M = 2|\lambda| \max_{0 \leq s \leq 1} \left| \int_0^1 \frac{s}{s+t} dt \right| = 2|\lambda| \ln 2 = .34657359,$$

$$\beta = \|F'(1)^{-1}\| = 1.53039421,$$

$$\eta \geq \|F'(1)^{-1}F(1)\| \geq \beta|\lambda| \ln 2 = .265197107,$$

$$h = .140659011 < \frac{1}{2}, \quad r_1 = .28704852, \quad r_2 = 3.4837317,$$

$$\theta = .08239685, \quad 2\beta K r_1 = .304497749 < 1,$$

wich shows that x^* is unique in $U(x^*, r_1)$ and not in $U(x^*, r_2)$; since (15) is violated. (For detailed computations, see also [1], [2]).

Therefore according to the theorem equation (35) has a solution x^* and the two-point Newton method (2)-(3) converges to x^* faster than any other method used so far according to (14) and (16). (See also, [1], [2], [7]). Moreover the information on the location of the solution given here is better than the ones given before.

Referencias

- [1] Argyros, I.K. (1985): Quadratic equations and applications to Chandrasekhar's and related equations, *Bull. Austral. Math. Soc.*, **32**, 275-292.
- [2] Argyros, I.K. (1988): On a class of nonlinear integral equations arising in neutron transport. *Aequationes Mathematicae*, **36**, 99-111.
- [3] Argyros, I.K. (1990): Improved error bounds for a certain class of Newton-like methods. *ATA*, **6**, 1-19.
- [4] Argyros, I.K. (1990): On the solution of equations with nondifferentiable and Ptak error estimates. *BIT*, **30**, 752-754.
- [5] Argyros, I.K. (1990): A mesh-independence principle for nonlinear operator equations and their discretizations under mild differentiability conditions. *Computing*, **45**, 265-268.
- [6] Argyros, I.K. (1990): The secant method in generalized Banach spaces. *Applied Math. Comput.*, **39**, 111-121.
- [7] Chandrasekhar, S.: Radiative transfer. Dover Publ. New York, 1960.
- [8] Chen Dong (1989): On the convergence and optimal error estimates of King's iteration procedures for solving nonlinear equations. *Intern. J. Computer Math.*, **26**:(3+4), 229-237.
- [9] Chen Dong (1990): Kantorovich-Ostrowski convergence theorems and optimal error bounds for Jarratt's iterative methods. *Intern. J. Computer Math.*, **31**:(3+4), 221-235.
- [10] Chen Dong (1990): On the convergence of a class of generalized Steffensen's iterative procedures and error analysis. *Intern. J. Computer Math.*, **31**:(3+4), 195-203.

- [11] Dennis, J.E.: On the convergence of Newton-like methods, In Numerical Methods for Nonlinear Algebraic Equations. Edited by P. Rabinowitz, Gordon and Breach, New York, 1970.
- [12] Dennis, J.E.: Toward a unified convergence theory for Newton-like methods, In Nonlinear Functional Analysis and Applications. edited by L.B. Rall, Academic Press, New York, 1971.
- [13] Gragg, W.B. and Tapia, R.A. (1974): Optimal error bounds for the Newton-Kantorovich theorem. *SIAM J. Numer. Anal.*, **11**, 10-13.
- [14] Kantorovich, L.V. and Akilov, G.P.: Functional Analysis in Normed Spaces. Pergamon Press, New York, 1964.
- [15] Ortega, J.M. and Rheinboldt, W.C.: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press., New York, 1970.
- [16] Ostrowski, A.M.: Solution of Equations in Euclidean and Banach Spaces. Academic Press, New York, 3rd ed., 1973.
- [17] Potra, F.A. and Ptak, V. (1980): Sharp error bounds for Newton's process. *Numer. Math.*, **34**, 63-72.
- [18] Traub, J.F.: Iterative Methods for the Solutions of Equations. Prentice Hall, Englewood Cliffs, 1964.
- [19] Taylor, A.E.: Introduction to Functional Analysis. Wiley, New York, 1957.

IMPROVED ERROR BOUNDS FOR FAST TWO-POINT NEWTON METHODS OF ORDER THREE *

Ioannis K. Argyros

Cameron University. Department of Mathematics. LAWTON, OK 73505 U.S.A.

Abstract

Using standard Newton-Kantorovich assumptions as well as the method of *continuous induction* we provide an error analysis for two point Newton methods.

1. Introduction

In this note we are concerned with the problem of approximating a locally unique zero x^* of the equation

$$F(x) = 0 \quad (1)$$

where F is a nonlinear operator defined on some convex subset D of a Banach space E_1 with values in a Banach space E_2 . We use two-step Newton method of the form

$$y_n = x_n - F'(x_n)^{-1}F(x_n) \quad (2)$$

$$x_{n+1} = y_n - F'(x_n)^{-1}F(y_n) \quad (3)$$

for some arbitrary $x_0 \in D$ and for all $n \geq 0$. The linear operator $F'(x_n)$ is the Frechet derivative of F evaluated at $x = x_n$. Let $U(x_0, R)$ denote the closed ball centered at x_0 and of radius $R > 0$, and assume $U(x_0, R) \subseteq D$, and $\|F'(x_0)^{-1}\| \leq \beta$. Moreover, assume that the following conditions are satisfied.

$$\|F'(x) - F'(y)\| \leq K \|x - y\| \quad (4)$$

for all $x, y \in U(x_0, R)$, and some $K > 0$,

$$\|F'(x) - F'(y)\| \leq q(r) \|x - y\| \quad (5)$$

for all $x \in U(x_0, r)$, $0 \leq r \leq R$, $0 \leq \|h\| \leq R - r$,

$$\|F'(x+h) - F'(x)\| \leq D(r, \|h\|) \quad (6)$$

*AMS (MOS) subject classification: 47H17, 65H10, 65J15

for all $x \in U(x_0, r)$, $0 \leq r \leq R$, $0 \leq \|h\| \leq R - r$.

Conditions (4), (5) and (6) have been used in [1], [3], [6], respectively for single step methods. Here, q is a nondecreasing function on $[0, R]$. Moreover, D is a nonnegative and continuous function of two variables such that if one of the variables is fixed then D is a nondecreasing function of the other on the interval $[0, R]$. Also, the function $\frac{\partial D(0, t)}{\partial t}$ is positive, continuous and nondecreasing on $[0, R-r]$ with $D(0, 0) = 0$.

It is convenient to introduce the constants

$$\begin{aligned} \eta &\geq \|y_0 - x_0\| > 0, \quad t_0 = t_0^1 = t_0^2 = 0, \quad s_0, s_0^1, s_0^2 \geq \eta, \\ t_1^1 &\geq s_0^1 + p_1, \quad s_1^1 \geq t_1^1 + p_2, \quad t_1^2 \geq s_0^2 + p_3, \quad s_1^2 \geq t_1^2 + p_4, \\ p_1 &= \frac{\beta}{1 - \beta w(t_0^1)} B(w, t_0^1, s_0^1), \quad p_2 = \frac{\beta}{1 - \beta w(t_1^1)} A(w, t_0^1, s_0^1, t_1^1), \\ p_3 &= \frac{\beta}{1 - \beta D(0, t_0^2)} \int_{t_0^2}^{s_0^2} D(t_0^2, t) dt, \quad p_4 = \frac{\beta}{1 - \beta D(0, t_1^2)} \int_{s_0^2}^{t_1^2} D(s_0^2, t) dt, \end{aligned}$$

the functions

$$\begin{aligned} g(t) &= \frac{K}{2} t^2 - \frac{1}{\beta} t + \frac{\eta}{\beta}, \quad w(r) = \int_0^r q(t) dt, \\ T_1(r) &= s_0^1 + \frac{\beta}{1 - \beta w(r)} \int_0^r w(t) dt, \quad T_2(r) = s_0^2 + \frac{\beta}{1 - \beta D(0, r)} \int_0^r D(r, t) dt, \end{aligned}$$

and for all $n \geq 0$ the scalar iterations

$$\begin{aligned} s_n &= t_n - \frac{g(t_n)}{g'(t_n)}, \quad s_{n+1}^1 = t_{n+1}^1 + \frac{\beta}{1 - \beta w(t_{n+1}^1)} A(w, t_n^1, s_n^1, t_{n+1}^1), \\ t_{n+1} &= s_n - \frac{g(s_n)}{g'(s_n)}, \quad t_{n+1}^1 = s_n^1 + \frac{\beta}{1 - \beta w(t_n^1)} B(w, t_n^1, s_n^1,), \\ A(w, t_n^1, s_n^1, t_{n+1}^1) &= \int_{s_n^1}^{t_{n+1}^1} w(t) dt - w(s_n^1)(t_{n+1}^1 - s_n^1), \\ B(w, t_n, s_n) &= \int_{t_n^1}^{s_n^1} w(t) dt - w(t_n^1)(s_n^1 - t_n^1), \\ s_{n+1}^2 &= t_{n+1}^2 + \frac{\beta}{1 - \beta D(0, t_{n+1}^2)} \int_{s_n^2}^{t_{n+1}^2} D(s_n^2, t) dt, \\ t_{n+1}^2 &= s_n^2 + \frac{\beta}{1 - \beta D(0, t_n^2)} \int_{t_n^2}^{s_n^2} D(t_n^2, t) dt. \end{aligned}$$

Moreover, assume:

(i) The condition

$$h = K\eta\beta < c = \frac{2d}{(d+1)^2},$$

where

$$d = \sqrt[3]{2}, \quad r_1 = \frac{1 + \sqrt{1 - 2h}}{h} \eta, \quad r_2 = \frac{1 - \sqrt{1 - 2h}}{h} \eta,$$

is satisfied.

(ii) There exists a minimum nonnegative number r_1^1 satisfying $T_1(r_1^1) \leq r_1^1$.

Moreover, the following estimates are true:

$$r_1^1 \leq R, \quad w(R) < \frac{1}{\beta}.$$

(iii) There exists a minimum nonnegative number r_1^2 satisfying $T_2(r_1^2) \leq r_1^2$.

Furthermore the following estimates are true

$$r_1^2 \leq R, \quad \beta D(0, R) < 1.$$

(iv) The following conditions

$$\frac{\beta}{1 - \beta w(r)} \left[\int_{s_0}^r w(t) dt + w(r)r \right] \leq s_1 - t_1, \quad \text{or } \leq -\frac{g(r)}{g'(r)},$$

$$\frac{\beta}{1 - \beta w(r)} \left[\int_0^r w(t) dt + w(r)r \right] \leq t_1 - s_0, \quad \text{or } \leq -\frac{g(r)}{g'(r)},$$

are satisfied for all $r \in [0, R]$, provided that $s_0 = s_0^1$, for all $r \in [0, R]$.

(v) The following conditions:

$$D(0, t) \leq w(t), \quad D(r_2^2, t) \leq w(t),$$

are satisfied for all $t \in [0, R]$, provided that $s_0^1 = s_0^2$. Then we can prove the following result (see also [2], [6] and the references there).

Theorem: Let $F : D \subseteq E_1 \rightarrow E_2$ be a nonlinear operator defined on some convex subset D of a Banach space E_1 with values in a Banach space E_2 . Assume (4)-(6) and (i)-(v) above are true. Then:

(a) The sequences $\{t_n\}$, $\{t_n^1\}$, $\{t_n^2\}$, (similarly for the sequences $\{s_n\}$, $\{s_n^1\}$, $\{s_n^2\}$) are monotonically increasing and bounded above by their limits r_2 , r_2^1 and r_2^2 , respectively.

(b) The two point step Newton method generated by (2)-(3) is well defined, remains in $U(x_0, v)$ ($v = r_2$, or r_2^1 , or r_2^2) for all $n \geq 0$, and converges to a unique solution x^* of equation $F(x) = 0$ in $U(x_0, v_1)$ ($v_1 = r_1$, or r_1^1 , or r_1^2).

Moreover the following estimates are true:

$$\begin{aligned} \|y_n - x_n\| &\leq s_n - t_n, \quad \|x_{n+1} - y_n\| \leq t_{n+1} - s_n, \\ \|x_n - x^*\| &\leq r_2 - t_n, \quad \|y_n - x^*\| \leq r_2 - s_n \end{aligned} \quad (7)$$

with

$$r_2 - t_n = \frac{(1 - \theta^2)\eta(d\theta)^{3^n-1}}{1 - (d\theta)^{3^n}}, \quad \theta = \frac{r_2}{r_1}$$

$$\begin{aligned} \|y_n - x_n\| &\leq s_n^1 - t_n^1, \quad \|x_{n+1} - y_n\| \leq t_{n+1}^1 - s_n^1, \\ \|x_n - x^*\| &\leq r_2^1 - t_n^1, \quad \|y_n - x^*\| \leq r_2^1 - s_n^1 \end{aligned} \quad (8)$$

$$\|y_n - x_n\| \leq s_n^2 - t_n^2, \quad \|x_{n+1} - y_n\| \leq t_{n+1}^2 - s_n^2, \quad (9)$$

$$\|x_n - x^*\| \leq r_2^2 - t_n^2, \quad \|y_n - x^*\| \leq r_2^2 - s_n^2$$

$$s_n^1 - t_n^1 \leq s_n - t_n, \quad t_{n+1} - s_n \leq t_{n+1}^1 - s_n^1, \quad r_2^1 - t_n^1 \leq r_2 - t_n, \quad r_2^1 - s_n^1 \leq r_2 - s_n \quad (10)$$

and

$$s_n^2 - t_n^2 \leq s_n^1 - t_n^1, \quad t_{n+1}^2 - s_n^2 \leq t_{n+1}^1 - s_n^1, \quad r_2^2 - t_n^2 \leq r_2^1 - t_n^1, \quad r_2^2 - s_n^2 \leq r_2^1 - s_n^1 \quad (11)$$

for all $n \geq 0$.

Note that conditions (i), or (ii), or (iii), or (i), (ii) and (iv), or (ii), (iii) and (v) imply the results (7), or (8), or (9), or (7), (8) and (10), or (8), (9) and (11).

Remarks:

1. The function D can be chosen to be

$$D(r, \|h\|) = \int_r^{r+\|h\|} q(t) dt$$

Another choice for D can be

$$D(r, \|h\|) = \sup_{x, y \in U(x_0, r), \|h\| \leq R-r} \|F'(x+h) - F'(x)\|$$

We will then have

$$D(r, \|h\|) \leq K \|h\|$$

and

$$D(r, \|h\|) = \int_r^{r+\|h\|} q(t) dt$$

for all $0 \leq r \leq R, 0 \leq \|h\| \leq R-r$

All of the above improve the corresponding results in [1]-[6]. Other choices are also possible. One can refer to [2] for applications of these ideas to the solution of integral equations.

2. Under the hypotheses of the theorem, conditions (v) can be replaced by the weaker

$$\frac{1}{1 - \beta D(0, r)} \int_0^r D(r, t) dt \leq \frac{1}{1 - \beta w(r)} \int_0^r w(t) dt$$

for all $r \in [0, R]$.

3. A choice for the function q can be given by

$$q(r) = \sup_{x, y \in U(x_0, r)} \frac{\|F'(x) - F'(y)\|}{\|x - y\|}$$

in which case $q(t) \leq K$ for all $x, y \in U(x_0, r)$

4. If the right hand side of condition (6) changes to $D(r, r + \|h\|)$, then a new result involving this condition can easily follow. Moreover the particular choice of D given by remark 1 above (first choice), will then reduce these results to the corresponding ones involving condition (5). The details are left to the motivated reader.

Referencias

- [1] Argyros, I.K. (1990): On the solution of equations with nondifferentiable operators and the Ptak error estimates. *BIT*, **30**, 752-754.
- [2] Argyros, I.K. (1994): On a multistep Newton method in Banach spaces and the Ptak error estimates. *CAM 94, University of Central Oklahoma, Edmond, OK U.S.A.*, 1-15.
- [3] Argyros, I.K. and Szidarovszky, F (1993): The theory and applications of iteration methods *C.R.C. Press, Inc., Boca Raton, Florida*.
- [4] Argyros, I.K. and Chen Dong. (1992): On the midpoint method for solving nonlinear operator equations in Banach spaces, *Appl. Math. Letters*, Vol. **5**, No **4**, 7-9.
- [5] Potra, F.A. and Ptak, V. (1980): Sharp error bounds for Newton's process, *Numer. Math.*, **34**, 63-72.
- [6] Zabrejko, P.P. and Nguen, D.F. (1987): The majorant method in the theory of Newton-Kantorovich approximations and the Ptak error estimates. *Numer. Funct. Anal. and Optimiz.* **9**, 671-684.

Weisner's group-theoretic method of obtaining generating functions is applied in the study of extended Jacobi polynomials. The main result of this paper is derived in section 3. Some particular cases of interest are also discussed.

1. Introduction

The extended Jacobi polynomial $F_n(\alpha, \beta; x)$ is defined by [1]:

$$F_n(\alpha, \beta; x) = \frac{(-1)^n}{n!} (x-a)^{-\alpha} (b-x)^{-\beta} \left(\frac{x}{b-a} \right) D^n [(x-a)^{\alpha+\alpha} (b-x)^{\beta+\beta}] \quad (1.1)$$

where $D = \frac{d}{dx}$.

The object of the present paper is to derive some novel generating functions of $F_n(\alpha, \beta; x)$, a modification of $F_n(\alpha, \beta; x)$, which satisfies the following ordinary differential equation [2]:

$$[(x-a)(b-x)] D^2 + [(\alpha+1)(b-x) - (\beta-\alpha+1)(x-a)] D + n(1+\beta+\alpha) v = 0. \quad (1.2)$$

by using Weisner's group-theoretic method [3] which is vividly presented in the monograph of E.B. McBride [4].

LIE THEORY AND SOME GENERATING FUNCTIONS OF EXTENDED JACOBI POLYNOMIALS.

B.K. Sen (*) and A.K. Chongdar (**)

(*) Department of Mathematics
Vidyanagar College,

P.O. Charashyambdas, South 24-Parganas,
West Bengal, INDIA.

(**) Department of Mathematics
Bancabasi Evening College
19, R.K. Chakraborty Sarani,
Calcuta-700009, INDIA.

Abstract.

Weisner's group-theoretic method of obtaining generating functions is applied in the study of extended Jacobi polynomials. The main result of this paper is derived in section 3. Some particular cases of interest are also discussed.

1. Introduction.

The extended Jacobi polynomial $F_n(\alpha, \beta; x)$ is defined by [1] :

$$F_n(\alpha, \beta; x) = \frac{(-1)^n}{n!} (z-a)^{-\alpha} (b-x)^{-\beta} \left(\frac{\lambda}{b-a} \right) D^n [(x-a)^{n+\alpha} (b-x)^{n+\beta}] \quad (1.1)$$

where $D \equiv \frac{d}{dx}$.

The object of the present paper is to derive some novel generating functions of $F_n(\alpha, \beta; x)$, a modification of $F_n(\alpha, \beta; x)$, which satisfies the following ordinary differential equation [2] :

$$[(x-a)(b-x) D^2 + \{(\alpha+1)(b-x) - (\beta-\alpha+1)(x-a)\} D + n(1+\beta+n)]v = 0, \quad (1.2)$$

by using Weisner's group-theoretic method [3] which is vividly presented in the monograph of E.B. McBride [4].

The main result of our investigation is given in (3.3). For previous works see [2,5,6].

2. Group-theoretic discussion.

Replacing $\frac{d}{dx}$ by $\frac{\partial}{\partial x}$, α by $y\frac{\partial}{\partial y}$ and v by $u(x,y)$ in (1.2), we obtain the following partial differential equation:

$$[(x-a)(b-x)\frac{\partial^2}{\partial x^2} + (b-a)y\frac{\partial^2}{\partial x\partial y} + \{a+b-2x-\beta(x-a)\}\frac{\partial}{\partial x} + n(1+\beta+n)]u = 0 \quad (1.2)$$

We now define the following first order linear partial differential operators:

$$\begin{aligned} A_1 &= y\frac{\partial}{\partial y} \\ A_2 &= (x-b)y\frac{\partial}{\partial x} - y^2\frac{\partial}{\partial y} + y\beta \\ A_3 &= (x-a)y^{-1}\frac{\partial}{\partial x} + \frac{\partial}{\partial y} \end{aligned} \quad (2.2)$$

such that

$$\begin{aligned} A_1[F_n(\alpha, \beta-\alpha; x)y^\alpha] &= \alpha F_n(\alpha, \beta-\alpha; x)y^\alpha \\ A_2[F_n(\alpha, \beta-\alpha; x)y^\alpha] &= (\beta-\alpha+n)F_n(\alpha+1, \beta-\alpha+1; x)y^{\alpha+1} \\ A_3[F_n(\alpha, \beta-\alpha; x)y^\alpha] &= (n+\alpha)F_n(\alpha-1, \beta-\alpha-1; x)y^{\alpha-1} \end{aligned} \quad (2.3)$$

Now we have the following commutator relations:

$$\begin{aligned} [A_1, A_2] &= A_2 \\ [A_1, A_3] &= -A_3 \\ [A_1, A_2] &= 2A_1 - \beta \end{aligned} \quad (2.4)$$

where

$$[A, B]u = (AB - BA)u$$

So from (2.4), we arrive at the following theorem:

Theorem : The set of operators $\{1, A_i : i = 1, 2, 3\}$, where 1 stands for the identity operator, generates a Lie algebra Z .

The partial differential operator L , given by

$$L = (x-a)(b-x) \frac{\partial^2}{\partial x^2} + (b-a)y \frac{\partial^2}{\partial x \partial y} + \{a+b-2x-\beta(x-a)\} \frac{\partial}{\partial x} + n(1+\beta+n)$$

can be expressed as follows:

$$L = -A_2 A_3 - A_1^2 + (1+\beta)A_1 \quad (2.5)$$

one can easily verify that L commutes with each A_i ($i = 1, 2, 3$), i.e.

$$[L, A_i] = 0 \quad (2.6)$$

The extended form of the group, generated by A_i ($i = 1, 2, 3$), are given below

$$e^{a_1 A_1} f(x, y) = f(x, e^{a_1} y)$$

$$e^{a_2 A_2} f(x, y) = (1+a_2 y)^\beta f\left[x + (x-b)a_2 y, \frac{y}{1+a_2 y}\right]$$

$$e^{a_3 A_3} f(x, y) = f\left[x + (x-a)\frac{a_3}{y}, y + a_3\right],$$

where all a_i ($i = 1, 2, 3$) are arbitrary constants.

Thus we have

$$\begin{aligned} e^{a_1 A_1} e^{a_2 A_2} e^{a_3 A_3} f(x, y) &= \{1 + a_2(y + a_3)\}^\beta f\left[x + (x-a)\frac{a_3}{y} + \right. \\ &\quad \left. + \{x + (x-a)\frac{a_3}{y} - b\} a_2(y + a_3), \frac{e^{a_1}(y + a_3)}{1 + a_2(y + a_3)}\right]. \end{aligned} \quad (2.8)$$

3. Generating functions.

From the above discussion, we see that $u(x,y) = F_n(\alpha, \beta-\alpha; x)y^\alpha$ is a solution, of the system:

$$L u = 0, \quad (A_1 - \alpha) u = 0$$

Also from (2.6), we observe that

$$SL[F_n(\alpha, \beta-\alpha; x)y^\alpha] = LS[F_n(\alpha, \beta-\alpha; x)y^\alpha] = 0,$$

where

$$S = e^{a_3 A_3} e^{a_2 A_2} e^{a_1 A_1}$$

So the transformation $S[F_n(\alpha, \beta-\alpha; x)y^\alpha]$ is annihilated by L .

Now, substituting $a_1 = 0$ and writing $f(x,y) = F_n(\alpha, \beta-\alpha; x)y^\alpha$ in (2.8), we get

$$\begin{aligned} e^{a_3 A_3} e^{a_2 A_2} [F_n(\alpha, \beta-\alpha; x)y^\alpha] &= [1 + a_2(y + a_3)]^{\beta-\alpha} F_n[\alpha, \beta-\alpha; x + \\ &+ (x-a)\frac{a_3}{y} + \{x + (x-a)\frac{a_3}{y} - b\} a_2(y+a_3)](y+a_3)^\alpha \end{aligned} \quad (3.1)$$

But

$$\begin{aligned} e^{a_3 A_3} e^{a_2 A_2} [F_n(\alpha, \beta-\alpha; x)y^\alpha] &= \\ &= \sum_{p=0}^{\infty} \sum_{k=0}^{\infty} \frac{(-a_2 y)^k}{k!} \frac{(-a_3/y)^p}{p!} (\alpha-\beta-n)_k (-\alpha-k-n)_p F_n(\alpha+k-p, \beta-k+p; x)y^\alpha \end{aligned} \quad (3.2)$$

Equating (3.1) and (3.2), we get

$$\begin{aligned} [1+a_2(y+a_3)]^{\beta-\alpha} F_n[\alpha, \beta-\alpha; x+(x-a)\frac{a_3}{y} + \{x+(x-a)\frac{a_3}{y} - b\} a_2(y+a_3)](y+a_3)^\alpha &= \\ &= \sum_{p=0}^{\infty} \sum_{k=0}^{\infty} \frac{(-a_2 y)^k}{k!} \frac{(-a_3/y)^p}{p!} (\alpha-\beta-n)_k (-\alpha-k-n)_p F_n(\alpha+k-p, \beta-k+p; x)y^\alpha \end{aligned} \quad (3.3)$$

We now consider the following cases:

Case 1 : Putting $a_3 = 0$ and $-a_2 y = t$, we get

$$(1-t)^{-\alpha+\beta} F_n[\alpha, \beta-\alpha; x-(x-b)t] = \sum_{k=0}^{\infty} \frac{(\alpha-\beta-n)_k}{k!} F_n(\alpha+k, \beta-\alpha-k; x) t^k \quad (3.4)$$

Special Case 1 : If we put $b = -a = 1$, $\lambda = 1$, we get

$$(1-t)^{\beta} P_n^{(\beta, \alpha)}(x - x-1 t) = \sum_{k=0}^{\infty} \frac{(-\beta-n)_k}{k!} P_n^{(\beta-k, \alpha+k)}(x) t^k \quad (3.5)$$

which is found derived in [6].

Subcase : Putting $\alpha = 0$, we get

$$(1-t)^{\beta} P_n^{(\beta, 0)}(x - x-1 t) = \sum_{k=0}^{\infty} \frac{(-\beta-n)_k}{k!} P_n^{(\beta-k, k)}(x) t^k$$

Case 2 : Putting $a_2 = 0$ and replacing $-a_3/y$ by t , we get

$$F_n[\alpha, \beta-\alpha; x-(x-a)t(1-t)^{\alpha}] = \sum_{k=0}^{\infty} \frac{(\alpha-p)_p}{p!} F_n(\alpha-p, \beta-\alpha+p; x) t^p \quad (3.6)$$

Case 3 : Putting $a_2 = 1/w$, $a_3 = 1$, $1/y = t$, we get

$$\begin{aligned} (1+t)^{\alpha} \left\{ 1 + \frac{1+t}{wt} \right\}^{\beta-\alpha} F_n[\alpha, \beta-\alpha; x+(x-a)t + \{x+(x-a)t-b\} \frac{1+t}{wt}] = \\ = \sum_{p=0}^{\infty} \sum_{k=0}^{\infty} \frac{(-1/wt)^k}{k!} \frac{(-t)^p}{p!} (\alpha-\beta-n)_k (-\alpha-k-n)_p F_n(\alpha+k-p, \beta-\alpha-k+p; x) \end{aligned} \quad (3.7)$$

Special Case : Putting $-a = b = 1$, $\lambda = 1$, we get

$$\begin{aligned} (1+t)^{\alpha} \left\{ 1 + \frac{1+t}{wt} \right\}^{\beta-\alpha} P_n^{(\beta-\alpha, \alpha)} \left[x+(x+1)t + \{x+(x+1)t-1\} \frac{1+t}{wt} \right] = \\ = \sum_{p=0}^{\infty} \sum_{k=0}^{\infty} \frac{(-1/wt)^k}{k!} \frac{(-t)^p}{p!} (\alpha-\beta-n)_k (-\alpha-k-n)_p P_n^{(\beta-\alpha-k+p, \alpha+k-p)}(x) \end{aligned} \quad (3.8)$$

Writing $\beta+\alpha$ in place of β and then interchanging α and β , we get

$$(1+t)^{\beta} \left\{ 1 + \frac{1+t}{w} \right\}^{\alpha} \left[\{x+(x+1)t\} \left\{ 1 + \frac{1+t}{w} \right\} - \frac{1+t}{w} \right] =$$

$$= \sum_{p=0}^{\infty} \sum_{k=0}^{\infty} \frac{(-1/w)^k}{k!} \frac{(-1)^p}{p!} (-\alpha-n)_k (-\beta-k-n)_p P_n^{(\alpha-k+p, \beta+k-p)}(x) t^{p-k}$$

which is the correct version of the result (1.4) of B. Ghosh [7].

References

- [1] K.R. Patil and N.K. Thakare: *Operational formulas and generating functions in the unified form for the classical orthogonal polynomials*. The Math. Student, 45(1) (1997), 41-51.
- [2] P.N. Srivastava and S.S. Dhillon: *Lie operators and classical orthogonal polynomials - II*. Pure Math-Manuscript, 7(1988), 129-136.
- [3] L. Weisner: *Group-theoretic origin of certain generating functions*. Pacific Jour. Maths., 5 (1955), 1033-1039.
- [4] E.B. McBride: *Obtaining generating functions*. Springer - Verlag (1971), Berlin.
- [5] A.K. Chongdar and S.K. Pan: *Quelques fonctions génératrices de certain polynomes orthogonaux du point de vue du groupe Lie*. Communicated.
- [6] A.K. Chongdar and Majumdar: *Some generating functions of extended Jacobi polynomials by group-theoretic method*. Communicated.
- [7] B. Ghosh: *Group-theoretic origin of certain generating functions of Jacobi polynomials - III*. Pure Math. Manuscript, 4 (1985), 5-10.

ON THE EXTENSION OF TRILATERAL GENERATING FUNCTIONS OF EXTENDED JACOBI POLYNOMIALS.

A. K. Chongdar

Department of Mathematics,
Bangabasi Evening College,
19, R.K. Chakraborti Sarani,
Calcutta - 700009, INDIA.

Abstract

In this article, the author obtains an extension of his result on mixed trilateral generating functions with the help of group-theoretic method. Some special cases of interest are also discussed.

1. Introduction.

The extended Jacobi polynomial as designated by N.K. Thakare [1] is given by

$$F_n(\alpha, \beta; n) = \frac{(-1)^n}{n} \left(\frac{\lambda}{b-a} \right)^n (x-a)^{-\alpha} (b-x)^{-\beta} D^n \left[(x-a)^{\alpha+n} (b-x)^{\beta+n} \right] \quad (1.1)$$

where $D = \frac{d}{dx}$.

In a recent paper [2], the present author has proved the following theorem in connection with mixed trilateral generating functions of modified extended Jacobi polynomials $F_n(\alpha-n, \beta; x)$.

Putting $\mu = \frac{x-a}{b-a}$ we have

Theorem 1 : If

$$G(x,u,\omega) = \sum_{n=0}^{\infty} a_n F_n(\alpha-n, \beta; x) g_n(u) \omega^n \quad (1.2)$$

where $g_n(u)$ is an arbitrary polynomial of degree n , then

$$(1-\lambda t)^{\alpha} (1-\lambda \mu t)^{-\alpha-\beta-1} G\left(\frac{x-\lambda \mu b t}{1-\lambda \mu t}, u, \frac{z t}{1-\lambda t}\right) = \sum_{n=0}^{\infty} F_n(\alpha-n, \beta; x) \sigma_n(z, u) t^n \quad (1.3)$$

where

$$\sigma_n(z, u) = \sum_{k=0}^n a_k \binom{n}{k} g_k(u) z^k \quad (1.4)$$

Here we have obtained an extension of the above result which do not seem to have appeared in the earlier works. The main result is as follows:

Theorem 2. If

$$G(x,u,\omega) = \sum_{n=0}^{\infty} a_n F_{n+m}(\alpha-n, \beta; x) g_n(u) \omega^n \quad (1.5)$$

then

$$(1-\lambda t)^{\alpha} (1-\lambda \mu t)^{-\alpha-\beta-1} G\left(\frac{x-\lambda \mu b t}{1-\lambda \mu t}, u, \frac{z t}{1-\lambda t}\right) = \sum_{n=0}^{\infty} F_{n+m}(\alpha-n, \beta; x) \sigma_n(z, u) t^n \quad (1.6)$$

where

$$\sigma_n(z, u) = \sum_{k=0}^n a_k \binom{n+m}{k+m} g_k(u) z^k \quad (1.7)$$

It is of interest to mention that theorem 2, in special cases, yields some interesting results (Theorem 3 to Theorem 6 of this paper) on mixed trilateral generating functions involving Hermite, modified Laguerre, modified Bessel and modified Jacobi polynomials.

2. Proof of the Theorem.

We first define the following linear partial differential operator R :

$$R = \lambda \mu (b-x) y \frac{\partial}{\partial x} - \lambda y^2 \frac{\partial}{\partial x} + [(b-x)\alpha - (\beta+m+1)(x-a)] \frac{\lambda y}{b-a}$$

such that

$$R(F_{n+m}(\alpha-n, \beta; x) y^n) = -(n+m+1) F_{n+m+1}(\alpha-n-1, \beta; x) y^{n+1} \quad (2.1)$$

and

$$\exp(\omega R)f(x,y) = (1+\lambda\omega y)^\alpha [1 + \lambda\omega\mu y]^{-\alpha-\beta-m-1} f\left(\frac{x+\lambda\omega\mu by}{1+\lambda\omega\mu y}, \frac{y}{1+\lambda\omega y}\right) \quad (2.2)$$

We now consider the following generating relation:

$$G(x,u,\omega) = \sum_{n=0}^{\infty} F_{n+m}(\alpha-n,\beta;x) g_n(u) \omega^n \quad (2.3)$$

Now replacing ω by ωyz in (2.3) and then operating $\exp(\omega R)$ on both sides, we get

$$\exp(\omega R) G(x,u,\omega yz) = \exp(\omega R) \sum_{n=0}^{\infty} a_n(\omega z)^n [F_{n+m}(\alpha-n,\beta;x) y^n] g_n(u) \quad (2.4)$$

The left member of (2.4) is

$$(1+\lambda\omega y)^\alpha [1 + \lambda\omega\mu y]^{-\alpha-\beta-m-1} f\left(\frac{x+\lambda\omega\mu by}{1+\lambda\omega\mu y}, u, \frac{\omega yz}{1+\lambda\omega y}\right) \quad (2.5)$$

The right member of (2.4) is

$$\sum_{n=0}^{\infty} \sum_{k=0}^{\infty} a_n(\omega z)^n \frac{\omega^k}{k!} R^k [F_{n+m}(\alpha-n,\beta;x) y^n] g_n(u) =$$

$$\sum_{n=0}^{\infty} \sum_{k=0}^{\infty} a_n(\omega z)^n \frac{\omega^k}{k!} (-1)^k (n+m+1)_k F_{n+m+k}(\alpha-n-k,\beta;x) y^{n+k} g_n(u) =$$

$$\sum_{n=0}^{\infty} \sum_{k=0}^{\infty} a_n(-z)^n \frac{(-\omega y)^{n+k}}{k!} (n+m+1)_k F_{n+m+k}(\alpha-n-k,\beta;x) g_n(u) = \quad (2.6)$$

$$\sum_{n=0}^{\infty} \sum_{k=0}^{\infty} a_{n-k}(-z)^{n-k} \frac{(-\omega y)^n}{k!} (n-k+m+1)_k F_{n+m}(\alpha-n,\beta;x) g_{n-k}(u) =$$

$$\sum_{n=0}^{\infty} \left(\sum_{k=0}^n a_{n-k} \frac{(n-k+m+1)_k}{k!} g_{n-k}(u) (-z)^{n-k} \right) F_{n+m}(\alpha-n,\beta;x) (-\omega y)^n$$

Equating the above two members (2.5) and (2.6) we get

$$(1+\lambda\omega y)^\alpha [1 + \lambda\omega\mu y]^{-\alpha-\beta-m-1} G\left(\frac{x+\lambda\omega\mu by}{1+\lambda\omega\mu y}, u, \frac{\omega yz}{1+\lambda\omega y}\right) =$$

$$= \sum_{n=0}^{\infty} \left(\sum_{k=0}^n a_{n-k} \frac{(n-k+m+1)_k}{k!} g_{n-k}(u) (-z)^{n-k} \right) F_{n+m}(\alpha-n, \beta; x) (-\omega y)^n \quad (2.7)$$

Replacing ωy by $-t$ and z by $-z$ in (2.7), we get

$$(1-\lambda t)^{\alpha} (1-\lambda \mu t)^{-\alpha-\beta-1} G\left(\frac{x-\lambda \mu b t}{1-\lambda \mu t}, u, \frac{z t}{1-\lambda t}\right) = \sum_{n=0}^{\infty} F_{n+m}(\alpha-n, \beta; x) \sigma_n(z, u) t^n \quad (2.8)$$

where $\sigma_n(z, u)$ is defined in (1.7) and this completes the proof of the theorem.

Corollary 1 : Now if put $m = 0$ in theorem 2, we get the theorem 1.

3. Some Special Cases.

We now discuss some special cases of our result.

Special Case 1 : Putting $\alpha = \beta$, $-a = b = \sqrt{\alpha}$, recalling $\lambda \rightarrow \frac{2}{\sqrt{\alpha}}$ and finally taking limit when $\alpha \rightarrow \infty$, we get the following extension of mixed trilateral generating relation involving Hermite polynomial:

Result 1 : If

$$G(x, u, \omega) = \sum_{n=0}^{\infty} a_n H_{n+m}(x) g_n(u) \frac{\omega^n}{n!} \quad (3.1)$$

then

$$\exp(2xt - t^2) G(x-t, u, tz) = \sum_{n=0}^{\infty} H_{n+m}(x) \sigma_n(u, z) \frac{t^n}{n!} \quad (3.2)$$

where $\sigma_n(z, u)$ is defined in (1.7) and which is note worthy.

Now putting $m = 0$ in the above theorem, we get

Corollary 2 : If

$$G(x, u, \omega) = \sum_{n=0}^{\infty} a_n H_n(x) g_n(u) \frac{\omega^n}{n!} \quad (3.3)$$

then

$$\exp(2xt - t^2) G(x-t, u, tz) = \sum_{n=0}^{\infty} H_n(x) \sigma_n(u, z) \frac{t^n}{n!} \quad (3.4)$$

where $\sigma_n(z, u)$ is defined in (1.4) and which is found derived in [3,4].

Special Case 2 : Putting $a = 0$, $\beta = b$, $\lambda = 1$, and then taking limit when $b \rightarrow \infty$, we get the following theorem on the extension of mixed trilateral generating relation involving Laguerre polynomials:

Result 2 : If

$$G(x, u, \omega) = \sum_{n=0}^{\infty} a_n L_{n+m}^{(\alpha-n)}(x) g_n(u) \omega^n \quad (3.5)$$

then

$$(1+t)^\alpha \exp(-xt) G(x(1+t), u, \frac{tz}{1+t}) = \sum_{n=0}^{\infty} L_{n+m}^{(\alpha-n)}(x) \sigma_n(u, z) t^n \quad (3.6)$$

where $\sigma_n(z, u)$ is defined in (1.7) and which is noteworthy.

Now putting $m = 0$ in theorem 4, we get

Corollary 3 : If

$$G(x, u, \omega) = \sum_{n=0}^{\infty} a_n L_n^{(\alpha-n)}(x) g_n(u) \omega^n \quad (3.7)$$

then

$$(1+t)^\alpha \exp(-xt) G(x(1+t), u, \frac{tz}{1+t}) = \sum_{n=0}^{\infty} L_n^{(\alpha-n)}(x) \sigma_n(u, z) t^n \quad (3.8)$$

where $\sigma_n(z, u)$ is defined in (1.4) and which is found derived in [5].

Special Case 3 : At first we write t in place of ω in the relation (1.5). Then putting $-a = b = \lambda = 1$, $\alpha = v - \epsilon - 1$, $\beta = \epsilon - 1$, and replacing x by $1 + \frac{2x\epsilon}{s}$, t by $\frac{s\omega}{\epsilon}$, and finally taking limit as $\epsilon \rightarrow \infty$, we get the following result on Bessel polynomials:

Result 3 : If

$$G(x, u, \omega) = \sum_{n=0}^{\infty} a_n Y_{n+m}(x, v-n, \epsilon) g_n(u) \frac{(\epsilon\omega)^n}{n!} \quad (3.9)$$

then

[3] A.K. Chongdar: *Some generating functions of $H_n^r(x, a, p)$ - the generalization of Hermite polynomials.* Tamkang Jour. Math., 19(2), (1988), 49-51.

[4] S.K. Chatterjea: *An extension of a class of bilateral generating functions for certain special functions.* Bull. Inst. Math. Acad. Sinica, 5(2), (1977), 323-331.

[5] M.C. Mukherjee and A.K. Chongdar: *On the unification of trilateral generating functions for certain special functions.* Tamkang Jour. Math., 19(3), (1988), 41-48.

S. Das: *On partial differential operator for Jacobi polynomials.* Pure Math. Manuscript, 3, (1984), 95-104.

A Study on the Order of Convergence of a Rational Iteration for Solving Quadratic Operator Equations in a Banach Space

D. Chen * I.K. Argyros † and Q. Qian ‡

Abstract

We introduce a new rational iteration, called the super-Halley method to solve nonlinear operator equations in a Banach space. The order of convergence of this method is four for quadratic operator equations. The order found by others (including the famous conjecture by Traub) for the methods that compare with ours is only three.

Key words: Banach space, quadratic operator, Super-Halley method.

AMS(MOS) Classification: 65H10, 65J15, 47H17

1 Introduction

In this study, we are concerned with approximating a solution x^* of the operator equation

$$F(x) = 0, \quad (1)$$

where $F : D \subset X \rightarrow Y$ is a nonlinear operator defined on some convex subset D of a Banach space X with values in a Banach space Y . Problems of this type appear frequently in many modern fields (see Ortega and Rheinboldt [1970], Dennis and Schnabel [1983], Cuyt and Rall [1985], Laub [1991], Argyros and Szidarovszky [1993], etc.) One of the well-known iterations for solving equation (1) is the third-order Halley. Given an approximation x_k , compute x_{k+1} by

*Department of Mathematical Sciences, University of Arkansas, Fayetteville, Arkansas 72701, USA

†Department of Mathematics, Cameron University, Lawton, Oklahoma 73505, USA

‡Department of Mathematics, University of Kentucky, Lexington, Kentucky 40506, USA

$$x_{k+1} = x_k - [F'(x_k) - \frac{1}{2}F''(x_k)F'(x_k)^{-1}F(x_k)]^{-1}F(x_k), \quad (2)$$

for all $k \geq 0$. or Here $F'(x_k)$, $F''(x_k)$ denote the first and second Frechet derivatives of F evaluated at $x = x_k$. Several authors have already studied the convergence of iteration (2) in a Banach space setting, Yamamoto [1988], Candela and Marquina [1990], Kanno [1992], Argyros and Szidarovszky [1993]. Recent applications of this method in various fields can be found in Davis [1983], Cuyt and Rall [1985], Petkovic [1989] and Kenney and Laub [1992]. By introducing a Newton step, we define a new iteration, called the super-Halley method:

$$y_k = x_k - F'(x_k)^{-1}F(x_k)$$

$$H_k = F'(x_k)^{-1}F''(x_k)(y_k - x_k)$$

$$x_{k+1} = y_k - \frac{1}{2}[I + H_k]^{-1}H_k(y_k - x_k), \quad (3)$$

for all $k \geq 0$, and some initial guess $x_0 \in D$. The famous conjecture by Traub for general iterations states that their maximum order of convergence is three, when the function is of one variable (see Traub [1974]). Here we restrict ourselves to quadratic operator equations of the form

$$P(x) = 0, \quad (4)$$

where

$$P(x) = Q(x, x) + L(x) + z_0, \quad (5)$$

where Q is a bounded symmetric bilinear operator on X , L is a bounded linear operator on X and $z_0 \in Y$ is a fixed. We show that the super-Halley method is of order four, whereas all previously mentioned authors have showed that it is only of order three

2 A Local Convergence

We can now prove the following result:

Theorem 2.1 *Let P be a quadratic operator defined on an open convex domain D of in a Banach space X with values in Banach space Y . Assume:*

a.- there exists a simple solution x^* of the equation $P(x) = 0$;

b.- the inverse of the Frechet-derivative of $P'(x^*)$ exists, and $\|P'(x^*)^{-1}\| \leq \beta^*$;

c.- there exists a $\delta > 0$ such that $\|Q\| \beta^* \delta \leq \frac{1}{6}$ and $C\delta^3 \leq q(\delta) < 1$ with $C = \frac{1}{\alpha}(1 + \beta^* \delta \|Q\|)^4 \beta^{*2} \frac{\|Q\|^3}{[1 - 4\beta^* \|Q\|(1 + 2\beta^* \delta \|Q\|)\delta]^2}$, where $x_0 \in D$ with $\|x_0 - x^*\| \leq \delta$ and for some fixed $\alpha > 0$.

Then the iteration $\{x_k\}$ generated by (3) is well defined, remains in $U(x^*, \delta)$ for all $k \geq 0$, and converges to x^* with

$$\|x_{k+1} - x^*\| \leq C \|x_k - x^*\|^4, k \geq 0, \quad (6)$$

Proof: Using assumption (c) we deduce that $x_0 \in U(x^*, \delta)$. From (5), we can show that

$$\|P'(x_k) - P'(x^*)\| \leq 2 \|Q\| \|x_k - x^*\| \leq 2 \|Q\| \delta,$$

and

$$\|P'(x^*)^{-1}\| \|P'(x_k) - P'(x^*)\| \leq 2\beta^* \delta \|Q\| < 1$$

by (c). Hence by the Banach lemma an invertible operator $P'(x_k)^{-1}$ exists and

$$\|P'(x_k)^{-1}\| \leq \frac{\|P'(x^*)^{-1}\|}{1 - \|P'(x^*)^{-1}\| \|P'(x_k) - P'(x^*)\|} \quad (7)$$

$$\leq \frac{\beta^*}{1 - 2\beta^* \delta \|Q\|} \quad (8)$$

$$\leq 2\beta^*. \quad (9)$$

From the approximation (see, Dennis and Schnabel [1983])

$$x^* - y_k = -P'(x_k)^{-1} \int_0^1 P''[x^* + t(x_k - x^*)] t dt (x^* - x_k)^2, \quad (10)$$

we obtain in turn

$$\|y_k - x^*\| \leq 2\beta^* \|Q\| \|x_k - x^*\|^2 \leq 2\beta^* \delta \|Q\| \|x_k - x^*\| < \delta. \quad (11)$$

Hence we get $y_k \in U(x^*, \delta)$. We now assume that $x_{k+1} \in U(x^*, \delta)$, and consequently (11) can now be proved again but holding for $k+1$. Moreover, from (3) and (7)-(11), we obtain in turn

$$\|H_k\| \leq 4\beta^* \|Q\| \|y_k - x_k\| \quad (12)$$

$$\leq 4\beta^* \|Q\| (\|y_k - x^*\| + \|x_k - x^*\|) \quad (13)$$

$$\leq 4\beta^* \|Q\| (2\beta^* \delta \|Q\| + 1) \delta < 1, \quad (14)$$

Therefore, $(I - H_k)^{-1}$ exists, and

$$\|(I - H_k)^{-1}\| \leq \frac{1}{1 - 4\beta^* \|Q\| (1 + 2\beta^* \|Q\| \delta) \delta}. \quad (15)$$

By the continuity of $P'(x^*)^{-1}$, there is an $\alpha > 0$ such that

$$\|P(x_{k+1}) - P(x^*)\| \geq \alpha \|x_{k+1} - x^*\|,$$

which can yield to the following estimate

$$\begin{aligned} \frac{\|P(x_{k+1})\|}{\|y_k - x_k\|} &\geq \frac{\|P(x_{k+1}) - P(x^*)\|}{\|y_k - x_k\|^4} \\ &\geq \frac{\alpha \|x_{k+1} - x^*\|}{\|y_k - x_k\|^4} \\ &\geq \frac{\alpha \|x_{k+1} - x^*\|}{(\|x_k - x^*\| + \|y_k - x^*\|)^4}, \end{aligned}$$

and

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^4} \leq \frac{1}{\alpha} \left[1 + \frac{\|y_k - x^*\|}{\|x_k - x^*\|}\right]^4 \frac{\|P(x_{k+1})\|}{\|y_k - x_k\|^4}. \quad (16)$$

Using (3) and (5) we can obtain the following approximation

$$\begin{aligned} F(x_{k+1}) &= \int_0^1 F''[y_k + t(x_{k+1} - y_k)](1-t) dt (x_{k+1} - y_k)^2 \\ &\quad - \frac{1}{2} \int_0^1 F''[x_k + t(y_k - x_k)](2t-1) dt (y_k - x_k)(I + H_k)^{-1} H_k (y_k - x_k) \\ &\quad + \int_0^1 [F''(x_k + t(y_k - x_k))(1-t) - \frac{1}{2} F''(x_k)](y_k - x_k)(I + H_k)^{-1} dt (y_k - x_k). \end{aligned}$$

We apply this approximation for $F = P$ to obtain

$$P(x_{k+1}) = Q(x_{k+1} - y_k)^2,$$

and then we deduce in turn, using also (3) and (5) that

$$\begin{aligned} \|P(x_{k+1})\| &\leq \|Q\| \left[\frac{\|Q\|}{1 - \|H_k\|} \|P'(x_k)^{-1}\| \|y_k - x_k\|^2 \right]^2 \\ &\leq 4\beta^2 \frac{\|Q\|^3}{[1 - 4\beta^* \|Q\| (1 + 2\beta^* \delta \|Q\| \delta)^2]} \|y_k - x_k\|^4, \end{aligned}$$

which estimate combined with (14) and (16) shows (6). Furthermore from (c) and (6) we have that $\|x_{k+1} - x^*\| \leq q \|x_k - x^*\|$ and $\|x_k - x^*\| \leq q^k \|x_0 - x^*\|$ for $k \geq 0$ which show that $x_k \in U(x^*, \delta)$ for all $k \geq 0$ and $\lim_{k \rightarrow \infty} x_k = x^*$. That completes the proof of the theorem.

Applications (1) Our method can be applied to solve quadratic integral equations arising in radiative transfer [4]. See Argyros [1] and Argyros and Szidarovszky [2] for details.

(2) Our method can be used to solve quadratic operator equations in a Banach space [1].

(3) Our method can be used to solve autonomous differential equations [2] and [10]. This is equations satisfying (5), and operator equations of the form

$$P'(x) = R(P(x)),$$

where R is a known continuous operator. Since $P'(x^*) = R(P(x^*)) = R(0)$, and $R(0)$ is known, it follows that $P'(x^*)$ is known, without actually knowing x^* .

(4) Furthermore, our method can be used to solve quadratic algebraic systems [5], [6] and [9].

(5) The basic hypothesis in most convergence theorems developed so far (see, Kanno [8], Yamamoto [12], and the reference there) has been that

$$\|F''(x) - F''(y)\| \leq N \|x - y\|,$$

for all $x, y \in D$, and for some constant $N > 0$. But for $F = P$, we obtain $N = 0$. Consequently the convergence theorems mentioned above cannot apply to solve quadratic operator equations of the form (5) at least directly. In the limiting case if the proofs of those theorems are reworked then the case $N = 0$ can be included. But those even then will guarantee only quadratic order of convergence. However the results obtained here require $N \geq 0$ only.

References

- [1] I.K. Argyros, On a Class of Nonlinear Integral Equations Arising in neutron Transport, *Aequations Mathematicae*, 36(1988),99-111.
- [2] I.K. Argyros and F. Szidarovszky, *The Theory and Application of Iteration Methods*, C.R.C. Press, Inc., Boca Raton, Florida, 1993.
- [3] V. Candela and A. Marquina, Recurrence Relations for Rational Cubic Methods I: The Halley Method, *Computing*, 44(1990), 169-184.
- [4] S. Chandrasekhar, *Radiative Transfer*, Dover Pub., New York, 1960.
- [5] A.M. Cuyt and L.B. Rall, Computational Implementation of the Multivariate Halley Method for Solving Nonlinear Systems of Equations, *ACM Transaction on Math. Software*, 11(1985), 20-36.
- [6] G. Davis, An Algorithm to Compute Solvents of the Matrix Equation $AX^2 + BX + C = 0$, *ACM Transaction on Math. and Software*, 9(1983), 246-254.
- [7] J.E. Dennis and R.B. Schnabel, *Numerical Methods for Unconstrained optimization and Nonlinear Equations*, Prentice-Hall, Inc., Englewood, New Jersey, 1983.
- [8] S. Kanno, Convergence Theorems for the Method of Tangent Hyperbolas, *Mathematica Japonica*, 37:4(1992),711-722.
- [9] C. Kenney and A.J. Laub, Rational Iterative Methods for the Matrix Sign Function, *SIAM J. Matrix Anal. Appl.*, 12(1991), 273-291.
- [10] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations In Several Variables*, Academic Press, New York, 1970.
- [11] J.F. Traub, *Theory of Optimal Algorithms, Software for Numerical Mathematics*, D.J. Evans, ed., Academic Press, New York, 1974, 1-13.
- [12] T. Yamamoto, On the Method of Tangent Hyperbolas in Banach Spaces, *J. Computational and Applied Math.*, 21(1988), 75-88.

GENERAL DETERMINANTAL REPRESENTATION OF PSEUDOINVERSES AND ITS COMPUTATION

Predrag Stanimirović

University of Niš

Faculty of Philosophy, Department of mathematics

Čirila i Metodija 2, 18000 Niš, Yugoslavia

Abstract. In this paper we establish general determinantal representation of generalized inverses and general form of different definitions of rectangular determinants and induced generalized inverses, in terms of minors of a matrix, satisfying certain conditions. Using this representation we obtain a general algorithm for exact computation of different classes of pseudoinverses: Moore-Penrose inverse, weighted Moore-Penrose inverse, group inverse, $\{1, 2, 3\}$, $\{1, 2, 4\}$, $\{1, 2\}$ inverses, left/right inverses, Radić's and Stojaković's generalized inverses. We also give some examples which illustrate our results.

1. Introduction

Let $\mathbb{C}_r^{m \times n}$ be the set of $m \times n$ complex matrices whose rank is r . Conjugate, transposed and conjugate-transposed matrix of A will be denoted by \bar{A} , A^T and A^* respectively. Submatrix and minor of A containing rows $\alpha_1, \dots, \alpha_t$ and columns β_1, \dots, β_t will be denoted by $A \begin{bmatrix} \alpha_1 \dots \alpha_t \\ \beta_1 \dots \beta_t \end{bmatrix}$ and $A \begin{pmatrix} \alpha_1 \dots \alpha_t \\ \beta_1 \dots \beta_t \end{pmatrix}$ respectively, and their algebraic complement corresponding to the element a_{ij} is $A_{ij} \begin{pmatrix} \alpha_1 \dots \alpha_{p-1} & i & \alpha_{p+1} \dots \alpha_t \\ \beta_1 \dots \beta_{q-1} & j & \beta_{q+1} \dots \beta_t \end{pmatrix} = (-1)^{p+q} A \begin{pmatrix} \alpha_1 \dots \alpha_{p-1} & \alpha_{p+1} \dots \alpha_t \\ \beta_1 \dots \beta_{q-1} & \beta_{q+1} \dots \beta_t \end{pmatrix}$.

For any matrix $A \in \mathbb{C}^{m \times n}$ consider the following equations in X [13]:

$$(1) \quad AXA = A, \quad (2) \quad XAX = X, \quad (3) \quad (AX)^* = AX, \quad (4) \quad (XA)^* = XA$$

and if $m = n$, also

$$(5) \quad AX = XA.$$

For a subset S of $\{1, 2, 3, 4, 5\}$ the set of matrices G obeying the conditions represented in S will be denoted by $A\{S\}$. A matrix $G \in A\{S\}$ is called an S -inverse of A and denoted by $A^{(S)}$. In particular for any $A \in \mathbb{C}^{m \times n}$ the set $A\{1, 2, 3, 4\}$ consists of a single element, the Moore-Penrose inverse of A , denoted by A^\dagger [13]. In the case $m = n$, the group inverse, denoted by $A^\#$, of A is the unique $\{1, 2, 5\}$ inverse, and exists if and only if $\text{ind}(A) = \min\{k : k > 0 \text{ and } \text{rank}(A^{k+1}) = \text{rank}(A^k)\} = 1$. [3]

The starting point of the investigations of this paper is the determinantal representation of Moore-Penrose inverse, studied in [1], [4], [6], [7], [8], [11]. The main result of these papers is:

Theorem 1.1. Element a_{ij}^\dagger , lying on the i -row and j -column of the Moore-Penrose pseudoinverse of a given matrix $A \in \mathbb{C}_r^{m \times n}$ is given by

$$a_{ij}^\dagger = \frac{\sum_{\substack{1 \leq \beta_1 < \dots < \beta_r \leq n \\ 1 \leq \alpha_1 < \dots < \alpha_r \leq m}} \bar{A} \begin{pmatrix} \alpha_1 \dots j \dots \alpha_r \\ \beta_1 \dots i \dots \beta_r \end{pmatrix} A_{ji} \begin{pmatrix} \alpha_1 \dots j \dots \alpha_r \\ \beta_1 \dots i \dots \beta_r \end{pmatrix}}{\sum_{\substack{1 \leq \delta_1 < \dots < \delta_r \leq n \\ 1 \leq \gamma_1 < \dots < \gamma_r \leq m}} \bar{A} \begin{pmatrix} \gamma_1 \dots \gamma_r \\ \delta_1 \dots \delta_r \end{pmatrix} A \begin{pmatrix} \gamma_1 \dots \gamma_r \\ \delta_1 \dots \delta_r \end{pmatrix}}, \quad \begin{pmatrix} 1 \leq i \leq n \\ 1 \leq j \leq m \end{pmatrix}.$$

For the sake of completeness in the following definition we unite definitions of generalized inverses introduced by M. Radić [16], [17], M. Stojaković [21] and V.N. Joshi [9].

Definition 1.1. Let i, j be integers, $1 \leq i \leq n$, $1 \leq j \leq m$. Then the (i, j) -th entry of Radić's, Stojaković's and Joshi's generalized inverse $A \in \mathbb{C}^{m \times n}$ is defined by

$$\sum_{\substack{1 \leq j_1 < \dots < j_r \leq n \\ 1 \leq i_1 < \dots < i_r \leq m}} \epsilon^{(i_1 + \dots + i_r) + (j_1 + \dots + j_r)} A_{ji} \begin{pmatrix} i_1 & \dots & j & \dots & i_r \\ j_1 & \dots & i & \dots & j_r \end{pmatrix} \\ \sum_{\substack{1 \leq \alpha_1 < \dots < \alpha_r \leq m \\ 1 \leq \beta_1 < \dots < \beta_r \leq n}} \epsilon^{(\alpha_1 + \dots + \alpha_r) + (\beta_1 + \dots + \beta_r)} A \begin{pmatrix} \alpha_1 & \dots & \alpha_r \\ \beta_1 & \dots & \beta_r \end{pmatrix}, \quad \text{where } \epsilon \in \{-1, 1\}.$$

For $\epsilon = 1$ we get Stojaković's definition, and for $\epsilon = -1$, we get Radić's definition.

In the following theorem is described an useful representation of generalized inverses [15]:

Theorem 1.2. If $A \in \mathbb{C}^{m \times n}$ has a full-rank factorization $A = PQ$, $P \in \mathbb{C}^{m \times r}$, $Q \in \mathbb{C}^{r \times n}$, $W_1 \in \mathbb{C}^{n \times r}$ and $W_2 \in \mathbb{C}^{r \times m}$ are some matrices such that $\text{rank}(QW_1) = \text{rank}(W_2P) = \text{rank}(A)$, then:

- $A^\dagger = Q^\dagger P^\dagger = Q^*(QQ^*)^{-1}(P^*P)^{-1}P^*$;
- the general solution of the equations (1), (2) is given by $W_1(QW_1)^{-1}(W_2P)^{-1}W_2$;
- the general solution of the equations (1), (2), (3) is given by $W_1(QW_1)^{-1}(P^*P)^{-1}P^*$;
- the general solution of the equations (1), (2), (4) is given by $Q^*(QQ^*)^{-1}(W_2P)^{-1}W_2$.

In [20] we develop determinantal representation of class of $\{1, 2\}$ inverses and the weighted Moore-Penrose inverse, as follows in the following two theorems:

Theorem 1.3. If $A = PQ$ is a full rank factorization of $A \in \mathbb{C}^{m \times n}$ and $W_1 \in \mathbb{C}^{n \times r}$, $W_2 \in \mathbb{C}^{r \times m}$ are some matrices such that $\text{rank}(QW_1) = \text{rank}(W_2P) = \text{rank}(W_1W_2) = \text{rank}(A)$, then an element $a_{ij}^{(1,2)} \in A^{(1,2)}$ is given by

$$a_{ij}^{(1,2)} = \frac{\sum_{\substack{1 \leq \beta_1 < \dots < \beta_r \leq n \\ 1 \leq \alpha_1 < \dots < \alpha_r \leq m}} (W_1W_2)^T \begin{pmatrix} \alpha_1 & \dots & i & \dots & \alpha_r \\ \beta_1 & \dots & j & \dots & \beta_r \end{pmatrix} A_{ji} \begin{pmatrix} \alpha_1 & \dots & j & \dots & \alpha_r \\ \beta_1 & \dots & i & \dots & \beta_r \end{pmatrix}}{\sum_{\substack{1 \leq \delta_1 < \dots < \delta_r \leq n \\ 1 \leq \gamma_1 < \dots < \gamma_r \leq m}} A \begin{pmatrix} \gamma_1 & \dots & \gamma_r \\ \delta_1 & \dots & \delta_r \end{pmatrix} (W_1W_2)^T \begin{pmatrix} \gamma_1 & \dots & \gamma_r \\ \delta_1 & \dots & \delta_r \end{pmatrix}}.$$

Theorem 1.4. Let $M \in \mathbb{C}^{m \times m}$ and $N \in \mathbb{C}^{n \times n}$ be positive definite, and $A = PQ$ be a full rank factorization of A , such that $\text{rank}(P^*MP) = \text{rank}(QNQ^*) = \text{rank}(MAN) = r$.

Element of the weighted Moore-Penrose inverse

$A_{M \bullet, \bullet N}^\dagger = (QN)^*(Q(QN)^*)^{-1}((MP)^*P)^{-1}(MP)^*$ lying on the i th row and j th column, can be represented in terms of square minors as follows:

$$(a_{M \bullet, \bullet N}^\dagger)_{ij} = \frac{\sum_{\substack{1 \leq \beta_1 < \dots < \beta_r \leq n \\ 1 \leq \alpha_1 < \dots < \alpha_r \leq m}} (MAN) \begin{pmatrix} \alpha_1 & \dots & j & \dots & \alpha_r \\ \beta_1 & \dots & i & \dots & \beta_r \end{pmatrix} A_{ji} \begin{pmatrix} \alpha_1 & \dots & j & \dots & \alpha_r \\ \beta_1 & \dots & i & \dots & \beta_r \end{pmatrix}}{\sum_{\substack{1 \leq \delta_1 < \dots < \delta_r \leq n \\ 1 \leq \gamma_1 < \dots < \gamma_r \leq m}} A \begin{pmatrix} \gamma_1 & \dots & \gamma_r \\ \delta_1 & \dots & \delta_r \end{pmatrix} (MAN) \begin{pmatrix} \gamma_1 & \dots & \gamma_r \\ \delta_1 & \dots & \delta_r \end{pmatrix}}.$$

Now we describe the main results of the paper. Firstly we obtain general representation of $\{i, j, k\}$ generalized inverses and the group inverse. Then we show that determinantal representation of the Moore-Penrose inverse and the weighted Moore-Penrose, determinantal representation of the Group inverse, $\{i, j, k\}$ inverses are only partial cases of the common general determinantal representation. Also, all introduced definitions of rectangular determinants and induced generalized inverses possesses the same general form. We also show that the class of left/right generalized inverses can be obtained from this general determinantal representation.

Later we describe algorithms for exact computation of generalized inverses based on the introduced determinantal representation. In this way, this paper is a generalization of an earlier paper [19], where the algorithms for computing of determinantal representation of the Moore-Penrose inverse and for computing the Radić's and Stojaković's inverse are described.

Finally, we give several examples which illustrate presented theory and algorithms.

2. General determinantal representation

Applying method used in [20, Theorem 2.1] and using general form of $\{i, j, k\}$ inverses (Theorem 1.2), the following lemma can be proved:

Lemma 2.1. If $A \in \mathbb{C}^{m \times n}$ and $A = PQ$ is a full rank factorization of A , and if $W_1 \in \mathbb{C}^{m \times r}$ and $W_2 \in \mathbb{C}^{r \times n}$ are some matrices such that $\text{rank}(QW_1) = \text{rank}(W_2P) = \text{rank}(A)$, then (i, j) -elements $a_{ij}^{(1,2,3)} \in A^{(1,2,3)}$ and $a_{ij}^{(1,2,4)} \in A^{(1,2,4)}$

$$a_{ij}^{(1,2,3)} = \frac{\sum_{\substack{1 \leq \beta_1 < \dots < \beta_r \leq n \\ 1 \leq \alpha_1 < \dots < \alpha_r \leq m}} (W_1 P^*)^T \begin{pmatrix} \alpha_1 & \dots & j & \dots & \alpha_r \\ \beta_1 & \dots & i & \dots & \beta_r \end{pmatrix} A_{ji} \begin{pmatrix} \alpha_1 & \dots & j & \dots & \alpha_r \\ \beta_1 & \dots & i & \dots & \beta_r \end{pmatrix}}{\sum_{\substack{1 \leq \delta_1 < \dots < \delta_r \leq n \\ 1 \leq \gamma_1 < \dots < \gamma_r \leq m}} A \begin{pmatrix} \gamma_1 & \dots & \gamma_r \\ \delta_1 & \dots & \delta_r \end{pmatrix} (W_1 P^*)^T \begin{pmatrix} \gamma_1 & \dots & \gamma_r \\ \delta_1 & \dots & \delta_r \end{pmatrix}}$$

$$a_{ij}^{(1,2,4)} = \frac{\sum_{\substack{1 \leq \beta_1 < \dots < \beta_r \leq n \\ 1 \leq \alpha_1 < \dots < \alpha_r \leq m}} (Q^* W_2)^T \begin{pmatrix} \alpha_1 & \dots & j & \dots & \alpha_r \\ \beta_1 & \dots & i & \dots & \beta_r \end{pmatrix} A_{ji} \begin{pmatrix} \alpha_1 & \dots & j & \dots & \alpha_r \\ \beta_1 & \dots & i & \dots & \beta_r \end{pmatrix}}{\sum_{\substack{1 \leq \delta_1 < \dots < \delta_r \leq n \\ 1 \leq \gamma_1 < \dots < \gamma_r \leq m}} A \begin{pmatrix} \gamma_1 & \dots & \gamma_r \\ \delta_1 & \dots & \delta_r \end{pmatrix} (Q^* W_2)^T \begin{pmatrix} \gamma_1 & \dots & \gamma_r \\ \delta_1 & \dots & \delta_r \end{pmatrix}}.$$

Note that conditions $\text{rank}(QW_1) = \text{rank}(W_2P) = \text{rank}(A)$ are satisfied if and only if $\text{rank}(W_1) = \text{rank}(W_2) = \text{rank}(A)$.

Determinantal representation of the Group inverse of a singular n by n matrix can be obtained from [2] and [10]:

Lemma 2.2. The group inverse $A^\# = (a_{ij}^\#)$ of $A \in \mathbb{C}^{n \times n}$ exists if

$$u = \sum_{1 \leq \mu_1 < \dots < \mu_r \leq n} A \begin{pmatrix} \mu_1 & \dots & \mu_r \\ \mu_1 & \dots & \mu_r \end{pmatrix} \neq 0.$$

Then $A^\#$ has the following determinantal representation:

$$a_{ij}^\# = \frac{\sum_{\substack{1 \leq \alpha_1 < \dots < \alpha_r \leq n \\ 1 \leq \beta_1 < \dots < \beta_r \leq n}} A^T \begin{pmatrix} \alpha_1 & \dots & j & \dots & \alpha_r \\ \beta_1 & \dots & i & \dots & \beta_r \end{pmatrix} A_{ji} \begin{pmatrix} \alpha_1 & \dots & j & \dots & \alpha_r \\ \beta_1 & \dots & i & \dots & \beta_r \end{pmatrix}}{\sum_{\substack{1 \leq \gamma_1 < \dots < \gamma_r \leq n \\ 1 \leq \delta_1 < \dots < \delta_r \leq n}} A^T \begin{pmatrix} \gamma_1 & \dots & \gamma_r \\ \delta_1 & \dots & \delta_r \end{pmatrix} A \begin{pmatrix} \gamma_1 & \dots & \gamma_r \\ \delta_1 & \dots & \delta_r \end{pmatrix}}$$

Proof. In view of the supposition $u \neq 0$ from [10, Theorem 8.] we get the following determinantal representation for the (i, j) -th entry of $A^\#$:

$$a_{ij}^\# = \frac{\sum_{\substack{1 \leq \alpha_1 < \dots < \alpha_r \leq n \\ 1 \leq \beta_1 < \dots < \beta_r \leq n}} A \begin{pmatrix} \beta_1 & \dots & i & \dots & \beta_r \\ \alpha_1 & \dots & j & \dots & \alpha_r \end{pmatrix} A_{ji} \begin{pmatrix} \alpha_1 & \dots & j & \dots & \alpha_r \\ \beta_1 & \dots & i & \dots & \beta_r \end{pmatrix}}{u^2}$$

Now the proof of the lemma can be completed using the following relation [10, Theorem 8.]

$$u^{-2} \cdot A \begin{pmatrix} \alpha_1 & \dots & \alpha_r \\ \beta_1 & \dots & \beta_r \end{pmatrix} A \begin{pmatrix} \beta_1 & \dots & \beta_r \\ \alpha_1 & \dots & \alpha_r \end{pmatrix} = 1$$

and the following, evident relation: $A \begin{pmatrix} \alpha_1 & \dots & \alpha_r \\ \beta_1 & \dots & \beta_r \end{pmatrix} = A^T \begin{pmatrix} \beta_1 & \dots & \beta_r \\ \alpha_1 & \dots & \alpha_r \end{pmatrix}$. ■

Now we introduce a general determinantal representation which includes the determinantal representations of all of presented generalized inverses, generalized inverses introduced by M. Stojaković, M. Radić and V.N. Joshi and the well known concept of the inversion of a square regular matrices.

Theorem 2.1. Let $A \in \mathbb{C}^{m \times n}$ and matrix R satisfies condition

$$(U_1) \quad \text{rank}(AR^*) = \text{rank}(R^*A) = \text{rank}(A).$$

Determinantal representation of an (i, j) -element of an arbitrary $\{1, 2\}$, $\{1, 2, 3\}$, $\{1, 2, 4\}$ inverse, left, right inverses, A^\dagger and $A^\#$ is

$$(2.1) \quad g_{ij} = \frac{A_{ij}^{(R,t)}}{\det_{(R,t)}(A)} = \frac{\sum_{\substack{1 \leq \beta_1 < \dots < \beta_t \leq n \\ 1 \leq \alpha_1 < \dots < \alpha_t \leq n}} \overline{R} \begin{pmatrix} \alpha_1 & \dots & j & \dots & \alpha_t \\ \beta_1 & \dots & i & \dots & \beta_t \end{pmatrix} A_{ji} \begin{pmatrix} \alpha_1 & \dots & j & \dots & \alpha_t \\ \beta_1 & \dots & i & \dots & \beta_t \end{pmatrix}}{\sum_{\substack{1 \leq \delta_1 < \dots < \delta_t \leq n \\ 1 \leq \gamma_1 < \dots < \gamma_t \leq m}} A \begin{pmatrix} \gamma_1 & \dots & \gamma_t \\ \delta_1 & \dots & \delta_t \end{pmatrix} \overline{R} \begin{pmatrix} \gamma_1 & \dots & \gamma_t \\ \delta_1 & \dots & \delta_t \end{pmatrix}}.$$

The numerator in (2.1) we simply denote by $A_{ij}^{(R,t)}$ and denominator by $\det_{(R,t)}(A)$. The expression $A_{ij}^{(R,t)}$ will be called generalized algebraic complement corresponding to element a_{ji} , and $\det_{(R,t)}(A)$ will be called generalized determinant.

Notation $r_c(A) = t \leq r \leq \min\{m, n\}$ denotes the greatest integer which ensures $\det_{(R,t)}(A) \neq 0$.

Proof. Consider the following cases:

1. Suppose that $t = m \leq n$. Using Laplace's development for the square minors of A , we get

$$\begin{aligned} \det_{(R,m)}(A) &= \det(AR^*) = \sum_{j_1 < \dots < j_m} \overline{R} \begin{pmatrix} 1 & \dots & m \\ j_1 & \dots & j_m \end{pmatrix} \left[\sum_{k=1}^r a_{ijk} A_{ik} \begin{pmatrix} 1 & \dots & m \\ j_1 & \dots & j_m \end{pmatrix} \right] = \\ &= \sum_{l=1}^n a_{il} \left[\sum_{j_1 < \dots < j_m} \overline{R} \begin{pmatrix} 1 & \dots & m \\ j_1 & \dots & l & \dots & j_m \end{pmatrix} A_{il} \begin{pmatrix} 1 & \dots & m \\ j_1 & \dots & l & \dots & j_m \end{pmatrix} \right] = \sum_{l=1}^n a_{il} A_{li}^{(R,m)}. \end{aligned}$$

For two integers $p \neq q$, $1 \leq p, q \leq m$, substituting in the minors of A in the expression

$$\det_{(R,m)}(A) = \sum_{j_1 < \dots < j_m} \overline{R} \begin{pmatrix} 1 & \dots & m \\ j_1 & \dots & j_m \end{pmatrix} A \begin{pmatrix} 1 & \dots & m \\ j_1 & \dots & j_m \end{pmatrix}$$

the q -row by the p -row, starting from $\det_{(R,m)}(A) = 0$, in a similar way can be proved

$$\sum_{l=1}^n a_{pl} A_{lq}^{(R,m)} = 0.$$

In this way, $g_{ij} = \delta_{ij} \det_{(R,m)}(A)$, and consequently $A \cdot A_{(R,m)}^{-1} = I_m$, which means that $A_{(R,m)}^{-1}$ represents a right inverse of a full-rank matrix A .

In a similar way can be proved that $A_{(R,n)}^{-1}$ represents a left inverse in the case $t = n \leq m$. Now is evident that (2.1) represents general determinantal representation of right/left inverses of a full rank matrix A .

2. For $R = A$ we obtain determinantal representation of A^\dagger , presented in Theorem 1.3. Then is trivial to conclude $r_c(A) = r$, which represents known result in [8].

3. If $m = n$, $\text{ind}(A) = 1$ and $R = A^*$ the determinantal representation of the group inverse is obtained (Lemma 2.2).

4. If $r_c(A) = r$ and matrix R satisfies condition

$$(U_2) \quad \overline{R} \begin{pmatrix} i_1 & \dots & i_r \\ j_1 & \dots & j_r \end{pmatrix} = K \cdot \epsilon^{(i_1 + \dots + i_r) + (j_1 + \dots + j_r)}, \quad \text{where } K \in \mathbb{C} \text{ and } \epsilon \in \{-1, 1\},$$

for all combinations $1 \leq i_1 < \dots < i_r \leq m$; $1 \leq j_1 < \dots < j_r \leq n$,

then, in the case $\epsilon = 1$, $A_{(R,r)}^{-1}$ is equal to Stojaković's inverse and the Radić's inverse in the case $\epsilon = -1$ (Definition 1.1).

5. If $A = PQ$ is a full-rank factorization of $A \in \mathbb{C}_r^{m \times n}$ and if $W_1 \in \mathbb{C}^{n \times r}$ and $W_2 \in \mathbb{C}^{r \times m}$ are some matrices such that $\text{rank}(QW_1) = \text{rank}(W_2P) = \text{rank}(A) = r$, then:

- for $R = PW_1^*$ we obtain determinantal representation of $A\{1, 2, 3\}$ (Lemma 2.1);

- for $R = W_2^*Q$ we obtain determinantal representation of $A\{1, 2, 4\}$ (Lemma 2.1);

- for $R = (W_1W_2)^*$ we obtain determinantal representation of $A\{1, 2\}$ (Theorem 1.3);

6. For $R = MAN$, where M and N are positive definite matrices of the appropriate sizes, $A_{(R,r)}^{-1}$ reduces to the weighted Moore-Penrose inverse (Theorem 1.4).

7. For a regular matrix A formulae (2.1) is transformed into the well known inversion of regular square matrices, for an arbitrary n by n matrix R .

8. If A is of full-rank matrix and each minor of R is the degree of the corresponding minor of A with the exponent $\frac{1}{2k-1}$, for some positive integer k , we obtain Radić's generalization of the Arghiriade-Dragomir representation of the Moore-Penrose inverse [18]. ■

3. Algorithms

Addition, subtraction, multiplication and division of complex or rational numbers in the internal forms we denote by \oplus , \ominus , \otimes , \oslash , respectively (makrooperations). The determinant of a square matrix A represented in the internal form is written as $\det(A)$.

Global parameters, for all of here presented algorithms are:

- ◊ S : the actual value of the norm $DET_{(R,i)}(A)$.
- ◊ $p(1:n)$, $q(1:n)$: vectors representing combinations of rows and columns of A respectively.

In the following algorithm D we describe algorithm for computing value of *generalized determinant*.

Procedure $D(u, m, n, x, y)$ { Computing generalized determinant of A . }

- ◊ $u = \text{rank}(A)$,
- ◊ m, n : number of rows and number of columns of A , respectively.
- ◊ x, y : the internal form of A and R respectively.

begin

$p(1:u) \leftarrow (1:u)$; $pl(1:u) \leftarrow (1:u)$; $q(1:u) \leftarrow (1:u)$; $ql(1:u) \leftarrow (1:u)$

$S \leftarrow 0$

$j \leftarrow m$

while $j \geq 1$ { terminates when all combinations $[p]$ are formed. }

$j_1 \leftarrow n$

while $j_1 \geq 1$ { terminates when all combinations $[q]$ are formed. }

Step 1.

Form the internal form of $\bar{R} \begin{bmatrix} p_1 & \dots & p_u \\ q_1 & \dots & q_u \end{bmatrix}$ and $A \begin{bmatrix} p_1 & \dots & p_u \\ q_1 & \dots & q_u \end{bmatrix}$, using y and x , respectively.

Step 2.

$p(1:u) \leftarrow (1:u)$; $q(1:u) \leftarrow (1:u)$

$S \leftarrow S \oplus \det(\bar{R} \begin{bmatrix} p_1 & \dots & p_u \\ q_1 & \dots & q_u \end{bmatrix}) \otimes \det(A \begin{bmatrix} p_1 & \dots & p_u \\ q_1 & \dots & q_u \end{bmatrix})$

$p(1:u) \leftarrow p_1(1:u)$; $q(1:u) \leftarrow q_1(1:u)$

if $q[u] = n$ **then** $j_1 \leftarrow j_1 - 1$ { form a new combination $[q]$ }

else $j_1 \leftarrow u$

if $j_1 \geq 1$ **then** **for** $i = u$ **downto** j_1

$q[i] \leftarrow q[j_1] + i - j_1 + 1$; $ql[i] \leftarrow ql[j_1]$ { new combination $[q]$ is formed }

end while

if $p[u] = m$ **then** $j \leftarrow j - 1$ { form a new combination $[p]$ }

else $j \leftarrow u$

if $j \geq 1$ **then** **for** $i = u$ **downto** j

$p[i] \leftarrow p[j] + i - j + 1$; $pl[i] \leftarrow pl[j]$ { new combination $[p]$ is formed }

end while

end { D }.

Now we describe Algorithm I for exact computation of different classes of generalized inverses.

Procedure $I(m, n, x, y, T)$ { Computing generalized inverses of A . }

- ◊ m, n : number of rows and number of columns of A , respectively.
- ◊ x, y : the internal form of A and R .
- ◊ $T = (t_{ij})$: the internal form of the computed generalized inverse of A .


```

begin
  u ← rank(A) + 1
  repeat
    u ← u - 1;    norma ← D(u, m, n, x, y)
  until norma ≠ 0
  p(1 : u) ← (1 : u);  p1(1 : u) ← (1 : u);  q(1 : u) ← (1 : u);  q1(1 : u) ← (1 : u)
  for w = 1 : n
    for v = 1 : m
      suma ← 0
      j ← m
      while j ≥ 1    { terminates when all combinations [p] are formed. }
        j1 ← n
        while j1 ≥ 1  { terminates when all combinations [q] are formed. }
          Step 1.
          Form the internal form of  $\overline{R} \begin{bmatrix} p_1 & \dots & p_u \\ q_1 & \dots & q_u \end{bmatrix}$  and  $A_{vw} \begin{bmatrix} p_1 & \dots & p_u \\ q_1 & \dots & q_u \end{bmatrix}$ ,
          using y and x, respectively.
          Step 2.
          p(1 : u) ← (1 : u);  q(1 : u) ← (1 : u)
          suma ← suma ⊕ det( $\overline{R} \begin{bmatrix} p_1 & \dots & p_u \\ q_1 & \dots & q_u \end{bmatrix}$ ) ⊗ det( $A_{vw} \begin{bmatrix} p_1 & \dots & p_u \\ q_1 & \dots & q_u \end{bmatrix}$ )
          p(1 : u) ← p1(1 : u);  q(1 : u) ← q1(1 : u)
          if q[u] = n then  j1 ← j1 - 1    { form a new combination [q] }
          else  j1 ← u
          if j1 ≥ 1 then  for i = u downto j1
            q[i] ← q[j1] + i - j1 + 1;  q1[i] ← q[i] { new combination [q] is formed }
          end while
          if p[u] = m then  { form a new combination [p] }  j ← j - 1
          else  j ← u
          if j ≥ 1 then  for i = u downto j
            p[i] ← p[j] + i - j + 1;  p1[i] ← p[i] { new combination [p] is formed }
          end while
          twv ← suma ⊙ norma
        end for
      end for
    end for
  end { I } .

```

4. Numerical examples

If a matrix R runs over the set of m by n matrices, in (2.1) we get various definitions of generalized inverses.

1. If $r = r_c(A)$ and a matrix R satisfies condition (U_1) , then $A_{(R,r)}^{-1}$ is equal to Stojaković's inverse, i.e. equivalent Joshi's inverse, in the case $\epsilon = 1$ and the Radi's inverse, in the case $\epsilon = -1$.

For example, consider matrix $A = \begin{pmatrix} \frac{11}{2} & \frac{23}{15} & 1 \\ \frac{3}{20} & -2 & \frac{234}{233} \end{pmatrix}$. Using $R = \begin{pmatrix} 2 & 0 & -2 \\ 1 & 1 & 0 \end{pmatrix}$ we get the Stojaković's inverse of A : $A_{(R,2)}^{-1} = \begin{pmatrix} \frac{58600}{440131} & \frac{619780}{1320573} \\ \frac{132135}{880182} & \frac{366975}{440131} \\ \frac{22135}{880382} & \frac{1720705}{1320573} \end{pmatrix}$. Now, using fixed point representation for the elements in A and the same matrix R we get

$$A = \begin{pmatrix} 5.50000000000000000000 & 1.53333333333333344000 & 1.00000000000000000000 \\ 0.14999999999999999999 & -0.28571428571428569800 & 1.00429184549356232000 \end{pmatrix},$$

and the following Stojaković's inverse of A :

$$A_{(R,2)}^{-1} = \begin{pmatrix} 0.13312403025050492700 & -0.46932657263172883300 \\ 0.15826652521291895100 & 0.83367220138530784300 \\ 0.02514249496241404220 & 1.30299877401703657000 \end{pmatrix}.$$

2. Furthermore, if $R = A$ satisfies (U_1) , then $A_{(R,r)}^{-1} = A^\dagger$, and both generalized inverses are identical to the Stojaković's or Radić's generalized inverse.

Concretly, for $R = A = \begin{pmatrix} 5729 & 5729 & 0 \\ 0 & 5729 & 5729 \\ -5729 & 0 & -5729 \end{pmatrix}$ we get the following Moore-Penrose inverse of A :

$$A_{(R,2)}^{-1} = A^\dagger = \begin{pmatrix} \frac{2008044837}{256295929} & 0 & -\frac{2008044837}{256295929} \\ \frac{2008044837}{256295929} & \frac{2008044837}{256295929} & 0 \\ 0 & \frac{2008044837}{256295929} & \frac{2008044837}{256295929} \end{pmatrix},$$

which is identical to the Stojaković's generalized inverse.

3. If $A \in \mathbb{C}_r^{m \times n}$ and $R = A$ we get $A_{(R,r)}^{-1} = A^\dagger$.

a) For example, if we use $R = A = \begin{pmatrix} 175 & 0 & 175 \\ 0 & 13 & 175 \\ 175 & 13 & 525 \\ 0 & 13 & 175 \end{pmatrix}$, then is obtained

$$A_{(R,2)}^{-1} = A^\dagger = \begin{pmatrix} \frac{192878339}{497627891} & -\frac{201395239}{995255782} & -\frac{4258450}{497627891} & -\frac{201395239}{995255782} \\ \frac{497627891}{1684865000} & -\frac{1263648750}{497627891} & -\frac{421216250}{497627891} & -\frac{2263648750}{497627891} \\ \frac{1684865000}{497627891} & \frac{497627891}{1075979571} & \frac{497627891}{1281633260} & \frac{497627891}{1075979571} \\ \frac{497627891}{995255782} & \frac{995255782}{995255782} & \frac{995255782}{995255782} & \frac{995255782}{995255782} \end{pmatrix}.$$

b) Using test matrix $Z_4 = \begin{pmatrix} a+1 & a & a & a \\ a & a-1 & a & a \\ a & a & a+1 & a \\ a & a & a & a-1 \end{pmatrix}$ from [22], for $a = 1$, we get

$$Z_4^\dagger = Z_4^{-1} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & -2 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & -2 \end{pmatrix}$$

4. For a square matrix A , such that $\text{ind}(A) \leq 1$ and $R = A^*$ we get $A_{(R,m)}^{-1} = A^\#$. For example, let given a matrix $A = \begin{pmatrix} 21.93-3i & 4. & \frac{275}{35917} & 9.13570+2950.84725i \\ 11.35 & 35.75-2i & 0 & 1257420 \\ 257384 & \frac{91584}{23} & 12+15i & 213574/5762403 \\ 159384-135i & 109825.23 & 183294/7359 & 0.000579 \end{pmatrix}$, where $i = \sqrt{-1}$. Using $R = A^*$, we get

$$A^\# = \begin{pmatrix} -0.005992-0.010995i & -0.000026+0.000014i & 0.000004+0.000001i & 0 \\ -0.029221+0.010681i & 0.000025+0.000068i & -0.000004+0.000001i & 0.000010 \\ 167.244553-23.230652i & 0.053301-0.392650i & -0.0109+0.000438i & -0.005703-0.000733i \\ 0.000001 & 0.000001 & 0 & 0 \end{pmatrix}.$$

5. Consider the following matrix of rank 2: $A = \begin{pmatrix} 2 & 0 & -5 & 4 \\ 7 & -4 & -9 & \frac{3}{2} \\ 3 & -4 & 1 & -\frac{13}{2} \end{pmatrix}$. Matrices $P = \begin{pmatrix} 2 & 0 \\ 7 & -4 \\ 3 & -4 \end{pmatrix}$ and $Q = \begin{pmatrix} 1 & 0 & -\frac{5}{8} & \frac{2}{8} \\ 0 & 1 & -\frac{17}{8} & \frac{25}{8} \end{pmatrix}$ form a full-rank factorization of A . If we use matrix $W_1 = \begin{pmatrix} 0 & 11 \\ -2 & 2 \\ 4 & 3 \\ 2 & 7 \end{pmatrix}$ satisfying condition $\text{rank}(QW_1) = \text{rank}(A)$ and $R = PW_1^T$, then the following $\{1, 2, 3\}$ inverse of A can be obtained: $A_{(R,2)}^{-1} = \begin{pmatrix} \frac{143}{735} & -\frac{11}{147} & -\frac{341}{735} \\ -\frac{3}{245} & \frac{20}{245} & \frac{26}{245} \\ \frac{109}{735} & -\frac{1}{147} & -\frac{313}{735} \\ \frac{6}{35} & -\frac{1}{7} & -\frac{17}{35} \end{pmatrix}$.

6. In this example we show necessity of the condition $\text{rank}(QW_1) = \text{rank}(A)$. Given matrix $A = \begin{pmatrix} 2 & 0 & -5 & 4 \\ 7 & -4 & -9 & \frac{3}{2} \\ 3 & -4 & 7 & -\frac{13}{2} \end{pmatrix}$ of rank 3. Full-rank factorization of A can be obtained from matrices $P = \begin{pmatrix} 2 & 0 & -5 \\ 7 & -4 & -9 \\ 3 & -4 & 7 \end{pmatrix}$ and $Q = \begin{pmatrix} 1 & 0 & \frac{21}{8} \\ 0 & 1 & \frac{117}{32} \\ 0 & 0 & \frac{1}{4} \end{pmatrix}$. Now we use the matrix $W_1 = \begin{pmatrix} 0 & 11 & 0 \\ -2 & 2 & 4 \\ 4 & 3 & -8 \\ 2 & 7 & -4 \end{pmatrix}$ such that $\text{rank}(QW_1) = 2$. For $R = PW_1^T$ we get: $A_{(R,2)}^{-1} = \begin{pmatrix} \frac{11}{5717} & \frac{990}{5717} & \frac{957}{5717} \\ \frac{129}{5717} & \frac{176}{5717} & -\frac{211}{5717} \\ -\frac{251}{5717} & \frac{278}{5717} & \frac{1031}{5717} \\ -\frac{5717}{5717} & \frac{5717}{5717} & \frac{5717}{5717} \end{pmatrix}$, and it is trivial to verify that the equations (1), (2) and (3) are not satisfied.

7. For $A \in \mathbb{C}^{m \times n}$ using $R = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$, $R \in \mathbb{C}^{m \times n}$, we obtain

$$DET_{(R,r)}(A) = A \begin{pmatrix} 1 \dots r \\ 1 \dots r \end{pmatrix}$$

and the following algebraic complement of the element a_{ij}

$$A_{ij}^{(R,r)} = \begin{cases} 0, & \text{for } j > r \text{ or } i > r \\ A_{ji} \begin{pmatrix} 1 \dots i \dots r \\ 1 \dots j \dots r \end{pmatrix}, & \text{for } j, i \leq r. \end{cases}$$

Generalized inverse of $A \in \mathbb{C}^{m \times n}$, is equal to

$$A_{(R,r)}^{-1} = \frac{1}{A \begin{pmatrix} 1 \dots r \\ 1 \dots r \end{pmatrix}} \begin{pmatrix} A_{11} \begin{pmatrix} 1 \dots r \\ 1 \dots r \end{pmatrix} \dots A_{r1} \begin{pmatrix} 1 \dots r \\ 1 \dots r \end{pmatrix} & 0 \dots 0 \\ \dots & \dots \\ A_{1r} \begin{pmatrix} 1 \dots r \\ 1 \dots r \end{pmatrix} \dots A_{rr} \begin{pmatrix} 1 \dots r \\ 1 \dots r \end{pmatrix} & 0 \dots 0 \\ 0 & \dots & 0 \dots 0 \\ \dots & \dots & \dots \\ 0 & \dots & 0 \dots 0 \end{pmatrix}.$$

Concretely, for $A = \begin{pmatrix} 13 & 115 & 476 \\ 56 & 13 & 23 \\ 3 & -372 & 26 \\ -3 & 14 & 21 \\ 12 & 1 & 0 \end{pmatrix}$ and $R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ can be obtained the following right

$$\text{generalized inverse of } A: A_{(R,2)}^{-1} = \begin{pmatrix} 10552600 & -8558144 & -1364947612 \\ 6188751 & 80453763 & -26817921 \\ 35980 & -3615752 & -7448669 \\ 2062917 & 8939307 & 160907526 \\ 76388480 & 7857808 & -1097248 \\ 18566253 & 6188751 & 6188751 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

8. Consider matrix $A = \begin{pmatrix} -\frac{2}{3} & 0 & -\frac{2}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \\ 5 & 5 & 15 \end{pmatrix}$ and positive-definite matrices $M = \begin{pmatrix} 5 & -1 & 3 \\ -1 & 2 & -2 \\ 3 & -2 & 3 \end{pmatrix}$ and $N = \begin{pmatrix} 4 & -3 & 3 \\ -3 & 4 & 1 \\ 3 & 1 & 5 \end{pmatrix}$. Using $R = MAN = \begin{pmatrix} 135 & 62 & 796/3 \\ -476 & -184 & -2692 \\ 698 & 292 & 4021 \end{pmatrix}$ we get the following weighted

$$\text{Moore-Penrose inverse of } A: A_{(R,2)}^{-1} = A_{M \bullet, \bullet N}^{\dagger} = \begin{pmatrix} -3516685 & -69873 & 12501 \\ 2227904 & 79568 & 1113952 \\ 2229475 & 41919 & 19557 \\ 1113952 & 39784 & 556976 \\ 205075 & 9327 & 51419 \\ 2227904 & -79568 & 1113952 \end{pmatrix}.$$

5. Conclusions

1. The memory requirements of the above presented procedures for $A \in \mathbb{C}_r^{m \times n}$ are two $r_c(A) \times r_c(A)$ matrices.

2. The advantage of the presented algorithms is their generality induced by theoretical weight of Theorem 2.1.

3. The efficiency of these algorithms is identical to the efficiency of the algorithms presented in [19]. The number of operations in most of the methods for computing the Moore-Penrose depends on the rank of the matrix. But high rank does not always lead to more operations. There presented method possesses this property.

references

- [1] Arghiriade, E. et Dragomir, A., *Une nouvelle definition de l'inverse generalisee d'une matrice*, Lincei - Rend. Sc. fis. mat e nat. **35** (1963), 158-165.
- [2] Bapat, R.B.; Bhaskara, K.P.S. and Manjunatha Prasad, K., *Generalized inverses over integral domains*, Linear Algebra Appl. **140** (1990), 181-196.
- [3] Ben-Israel, A. and Greville, T.N.E., *Generalized Inverses: Theory and applications*, Wiley-Interscience, New York, 1974..
- [4] Ben-Israel, A., *Generalized inverses of matrices: a perspective of the work of Penrose*, Math. Proc. Camb. Phil. Soc. **100** (1986), 407-425.
- [5] Duane Pyle, L., *The weighted generalized inverse in nonlinear programming-active set selection using a variable-metric generalization of the simplex algorithm*, Lecture Notes in Economics and Mathematical Systems **174** (1977), 197-231.

- [6] Gabriel, R., *Extinderea complementilor Algebrici generalizati la gatrici oarecare*, Studii si cercetari matematice 17 -Nr. 10 (1965), 1566-1581.
- [7] Gabriel, R., *Das verallgemeinerte inverse einer matrix, deren Elemente einem beliebigen Körper angehören*, J. Rewie ansew Math. 234 (1967), 107-122.
- [8] Gabriel, R., *Das verallgemeinerte inverse einer Matrix, über einem beliebigen Körper - analytisch betrachtet*, J. Rewie ansew Math. 244(V) (1970), 83-93.
- [9] Joshi, V.N., *A Determinant for rectangular matrices*, Bull. Australl. Math. Soc. 21 (1980), 137-146.
- [10] Manjunatha, P.; Bhaskara, K.P.S. and Bapat, R.B., *Generalized inverses over integral domains. II. Group inverses and Drazin inverses*, Linear Algebra Appl. 146 (1991), 31-47.
- [11] Moore, E.H., *General Analysis, Part I. The Algebra of Matrices*, (compiled and edited by R.W. Barnard), (The Amer. Philos. Soc. 1935).
- [12] Moore, E.H., *On the reciprocal of the general algebraic matrix (Abstract)*, Bull. Amer. Math. Soc. 26 (1920), 394-395.
- [13] Penrose, R., *A generalized inverse for matrices*, Proc. Cambridge Philos. Soc. 51 (1955), 406-413.
- [14] Radhakrishna R., C. and Kumar, M., *Generalized Inverse of Matrices and its Applications*, John Wiley & Sons, Inc, New York, London, Sydney, Toronto, 1971.
- [15] Radić, M., *Some contributions to the inversions of rectangular matrices*, Glasnik Matematički 1 (21) -No. 1 (1966), 23-37.
- [16] Radić, M., *Inverzija Pravokutnih Matrica*, Doktorska disertacija, 1964.
- [17] Radić, M., *A definition of the determinant of a rectangular matrix*, Glasnik matematički 1(21)-No. 1 (1966), 17-22.
- [18] Radić, M., *On a generalization of the Arghiriade-Dragomir representation of the Moore-Penrose inverse*, Lincei-Rend. Sc. fis. mat. e nat. 44 (1968), 333-336.
- [19] Stanimirović, P. and Stanković, M., *Computing pseudoinverses of rectangular matrices in terms of square submatrices*, VIII Conference on Applied Mathematics, Tivat (1993), 207-216.
- [20] Stanimirović, P. and Stanković, M., *Determinantal representation of weighted Moore-Penrose inverse*, Matematički Vesnik 46 (1994), 41-50.
- [21] Stojaković, M., *Determinante Nkvadratnih Matrica*, Vesnik, D.M.N.R.S., Beograd 1-2 (1952), 9-21.
- [22] G. Zielke, *Report on test matrices for generalized inverses*, Computing 36 (1986), 105-162.

Una relación entre cuasicovarianzas muestral y poblacional.

M. Ruiz Espejo

Departamento de Economía Aplicada Cuantitativa
Facultad de Ciencias Económicas y Empresariales (U.N.E.D.)
Apartado 60.142. 28080 MADRID.

Summary. Under simple random sampling without replacement of fixed sample size n , we give an original proof of a result known, that sample quasicovariance is an unbiased estimator of population quasicovariance. As a corollary, the sample quasivariance is unbiased for population quasivariance.

1. Introducción

Un resultado importante en teoría de muestras es que la cuasivarianza muestral es insesgada para estimar la cuasivarianza poblacional, bajo muestreo aleatorio simple sin reemplazamiento -mas- de tamaño muestral fijo n . Este resultado es aplicado a la estimación de varianzas de estimadores en poblaciones finitas.

Sea $U = \{1, 2, \dots, N\}$ la población finita, y $S(n)$ el conjunto de muestras aleatorias simples sin reemplazamiento de tamaño n , $1 < n < N$; es decir,

$$S(n) = \{s \subset U : \text{card}(s) = n\},$$

donde $p(s) = 1/\binom{N}{n}$ para toda muestra $s \in S(n)$.

Si disponemos de dos variables x e y definidas sobre la población finita, vamos a justificar que bajo diseño mas, la cuasicovarianza muestral

$$s_{xy} = \frac{1}{n-1} \sum_{i \in s} (x_i - \bar{x}_s)(y_i - \bar{y}_s) = \frac{1}{n-1} \left(\sum_{i \in s} x_i y_i - n \bar{x}_s \bar{y}_s \right),$$

es un estimador insesgado de la cuasivarianza poblacional

$$S_{xy} = \frac{1}{N-1} \sum_{i \in U} (x_i - \bar{X}) (y_i - \bar{Y}) = \frac{1}{N-1} \left(\sum_{i \in U} x_i y_i - N \bar{X} \bar{Y} \right),$$

donde \bar{z}_s y \bar{Z} son las medias muestral (asociada a la muestra s) y poblacional (asociada a la población finita U) para la variable $z (= x \text{ ó } y)$.

2. Resultado

Teorema. Bajo diseño mas-n, $E(s_{xy}) = S_{xy}$ □

Demostración.

$$\begin{aligned} E(s_{xy}) &= \frac{1}{n-1} E \left[\sum_{i \in s} x_i y_i - \frac{1}{n} \left(\sum_{i \in s} x_i y_i + \sum_{i \neq j \in s} x_i y_j \right) \right] = \\ &= \frac{1}{n-1} E \left[(n-1) \frac{1}{n} \sum_{i \in s} x_i y_i - \frac{1}{n} \sum_{i \neq j \in s} x_i y_j \right] = \\ &= E \left(\frac{1}{n} \sum_{i \in s} x_i y_i \right) - \frac{1}{n(n-1)} E \left(\sum_{i \neq j \in s} x_i y_j \right) \end{aligned} \quad (1)$$

Pero como la media muestral \bar{z}_s (de $z_i = x_i y_i$) tiene por esperanza a la media poblacional \bar{Z} ,

$$E \left(\frac{1}{n} \sum_{i \in s} x_i y_i \right) = \frac{1}{N} \sum_{i \in U} x_i y_i. \quad (2)$$

Además, razonando de modo similar a Ruiz (3) podemos calcular la esperanza matemática siguiente

$$\begin{aligned} E \left(\sum_{i \neq j \in s} x_i y_j \right) &= \sum_{s \in S(n)} \left(\sum_{i \neq j \in s} x_i y_j \right) p(s) = \\ &= \sum_{i \neq j \in U} x_i y_j \text{card}\{s \in S(n) : i \neq j \in s\} \frac{1}{\binom{N}{n}} = \\ &= \frac{\binom{N-2}{n-2}}{\binom{N}{n}} \sum_{i \neq j \in U} x_i y_j = \frac{n(n-1)}{N(N-1)} \sum_{i \neq j \in U} x_i y_j \end{aligned} \quad (3)$$

pues

y

$$p(s) = \frac{1}{\binom{N}{n}}$$

$$\text{card}\{s \in S(n) : i \neq j \in s\} = \binom{N-2}{n-2}.$$

Así, sustituyendo (2) y (3) en (1), tenemos:

$$\begin{aligned} E(s_{xy}) &= \frac{1}{N} \sum_{i \in U} x_i y_i - \frac{1}{N(N-1)} \sum_{i \neq j \in U} x_i y_j = \\ &= \frac{1}{N} \left(\frac{N \sum_{i \in U} x_i y_i - \sum_{i \in U} x_i y_i - \sum_{i \neq j \in U} x_i y_j}{N-1} \right) = \\ &= \frac{1}{N-1} \left(\sum_{i \in U} x_i y_i - N \bar{X} \bar{Y} \right) = S_{xy} \quad \square \end{aligned}$$

Denotando $s_x^2 = s_{xx}$ y $S_x^2 = S_{xx}$, tenemos que:

Corolario. Bajo diseño mas-n, $E(s_x^2) = S_x^2 \quad \square$

Demostración. Inmediato del teorema anterior, si la variable $y = x \quad \square$

3. Comentarios finales

Estos resultados han sido citados sin demostración por Foreman (2, p. 58), y probados con una notación algo más tradicional y menos avanzada por De Vries (1, pp. 341-344). El corolario es un resultado clásico y ampliamente difundido desde los años 50 del actual siglo XX.

Referencias

- (1) De Vries, P. G. (1986). *Sampling Theory for Forest Inventory*. Springer-Verlag. Nueva York.
- (2) Foreman, E. K. (1991). *Survey Sampling Principles*. Marcel Dekker. Nueva York.
- (3) Ruiz, M. (1992-93). Problema 46. *Qüestiió* 16, 176; 17, 171-172.

OPTIMUM LINEAR INTEGRATION OF INDEPENDENT SAMPLE MEANS FROM FINITE POPULATIONS.

M. Ruiz Espejo (*), J. Santos Peñas y M. Rueda García (**)

(*) Deptº de Economía Aplicada Cuantitativa
Fac. de Cienc. Económicas y Empresariales (U.N.E.D.)
Apartado 60142 - 28080 - MADRID

(**) Deptº de Estadística e Investigación Operativa
Fac. de Cienc. - Universidad de Granada
Campus de Fuente Nueva - 18014 - GRANADA

Summary.

We obtain the optimum linear integrated estimators from a number of independent unbiased strategies with common simple random sampling with or without replacement (srsr or srs) design respectively, from a same finite population. Moreover, we deduce the optimum linear unbiased estimators asymptotically to estimate the variance of both integrated estimators

1. Introduction.

The problem of optimum integration that we are going to resolve consists of carrying out - from a same finite population of size N and on an interest variable "y" defined in its N units - simple random samples with or without replacement (srsr or srs, respectively) of distinct sample sizes $n_i \geq 1$ ($i = 1, 2, \dots, k$), k being the number of independent samples, and

$$n = \sum_{i=1}^k n_i$$

If the strategies used are $e_i \equiv (srsr-n_i, \bar{y}_i)$, with \bar{y}_i the sample mean of the n_i units observed by "srsr", then we can integrate all the unbiased strategies " e_i " in one single estimator (more precise than each strategy e_i) which will efficiently exploit all the information,

$$t = \sum_{i=1}^k t_i \bar{y}_i$$

where t_i ($i = 1, 2, \dots, k$) are positive relative weights such as

$$\sum_{i=1}^k t_i = 1$$

From Raj (1968) or Ruiz (1992) it follows that the optimum weights of t_i, \bar{t}_i (say), are precisely

$$\bar{t} = \frac{n_i}{n}$$

That is to say, the sample mean estimator of the n observations,

$$\bar{t} = \sum_{i=1}^k \bar{t}_i \bar{y}_i = \sum_{i=1}^k \bar{y} \left(\frac{n_i}{n} \right) = \bar{y}_{(n)},$$

is optimum in the sense of being unbiased and of minimum variance among the possible t estimators. Furthermore, it can be deduced that

the variance of \bar{t} is

$$V(\bar{t}) = \frac{\sigma^2}{n}$$

σ^2 being the finite population variance.

We studying two cases of integration and their variance estimation in practice.

2. Variance estimation of \bar{t}

The unbiased estimator of $V(\bar{t})$ of minimum variance (we reasoning in a similar way) is obtained from the linear estimator

$$v = \sum_{i=1}^k v_i s_i^2$$

with $v_i \geq 0$ and

and σ^2 its the sample quasivariance from the i 'th independent study of size n_i , which is unbiased for σ^2 . Its variance is (Hansen et al., 1953)

$$V(s_i^2) = \frac{\mu_4}{n_i} - \frac{(n_i-3)\sigma^4}{n_i(n_i-1)} = a_i \text{ (say)}$$

which, if $n_i \rightarrow \infty$,

$a_i \leftrightarrow \frac{\mu_4 - \sigma^4}{n_i}$ are asymptotically equivalent, μ_4 being the central moment of order 4 of the finite population. The optimum weights of integration are

$$\bar{v}_i = (1/a_i) / \left(\sum_{j=1}^k (1/a_j) \right)$$

that are asymptotically equivalent, if $n_i \rightarrow \infty$ ($i = 1, 2, \dots, k$), to

$$\bar{v}_i \leftrightarrow \frac{n_i}{n}.$$

As such the optimum linear integrated unbiased estimator of σ^2 will result

$$\bar{v} = \sum_{i=1}^k \bar{v}_i s_i^2$$

and asymptotically will be equivalent, if $n_i \rightarrow \infty$ ($i = 1, 2, \dots, k$), to

$$\bar{v} \leftrightarrow \sum_{i=1}^k \frac{n_i}{n} s_i^2 \quad (2.1)$$

as such the optimum linear integrated unbiased estimator of $V(\bar{t})$ (and asymptotically equivalent, respectively) will be

$$\hat{V}(\bar{t}) = \frac{\bar{v}}{n} \left(\leftrightarrow \frac{1}{n^2} \sum_{i=1}^k n_i s_i^2 \right)$$

an estimator that requires (and not respectively) the knowledge of a_i ($i = 1, 2, \dots, k$), a value which is not accessible usually in practice, but its asymptotic expression based on (2.1) is, if the sample sizes n_i (in each independent strategy e_i) are sufficiently large. Moreover, the variance $V(\hat{V}(\bar{t}))$ is (and if $n_i \rightarrow \infty$ ($i = 1, 2, \dots, k$), asymptotically equivalent respectively)

$$V(\hat{V}(\bar{t})) = \frac{1}{n^2} \hat{V}(\bar{v}) = \frac{1}{n^2} \sum_{i=1}^k a_i \bar{v}_i^2 \left(\leftrightarrow \frac{1}{n^2} \sum_{i=1}^k \frac{n_i}{n^2} (\mu_4 - \sigma^4) = \frac{\mu_4 - \sigma^4}{n^3} \right).$$

3. Strategies $e' \equiv (srs - n_i, \bar{y}_i)$

In this case, if strategies e_i' ($i = 1, 2, \dots, k$) are independent of each other, the estimator of population mean, which is unbiased and integrates linearly all the strategies will be

$$t' = \sum_{i=1}^k t_i' \bar{y}_i$$

with $t_i' \geq 0$ ($i = 1, 2, \dots, k$) and $\sum_{i=1}^k t_i' = 1$

Now,

$$V(\bar{y}_i) = \left(\frac{1}{n_i} - \frac{1}{N} \right) S^2 = b_i \text{ (say)}$$

being $S^2 = N\sigma^2/(N-1)$ the population quasivariance. Then the optimum weights are

$$\bar{t}_i' = (1/b_i) / \left(\sum_{j=1}^k (1/b_j) \right) = \frac{n_i}{N - n_i} / \left(\sum_{j=1}^k \frac{n_j}{N - n_j} \right)$$

which only depends of n_i and N , but not of S^2 . As such the optimum linear integrated estimator for the population mean, from the independent e_i' strategies is exactly

$$\bar{t}' = \sum_{i=1}^k t_i' \bar{y}_i = \left(\sum_{i=1}^k \frac{n_i \bar{y}_i}{N - n_i} \right) / \left(\sum_{j=1}^k \frac{n_j}{N - n_j} \right),$$

a simple estimator that does not use unknown parameters. The variance of \bar{t}' is exactly

$$V(\bar{t}') = \frac{S^2}{N \left(\sum_{i=1}^k \frac{n_i}{N-n_i} \right)} = \frac{\sigma^2}{(N-1) \left(\sum_{i=1}^k \frac{n_i}{N-n_i} \right)}.$$

It is not very complicated to prove that for a fixed sample size n ,

$$V(\text{srs-n}, \bar{y}) = \frac{N-n}{n} \frac{S^2}{n} \leq V(\bar{t}') \leq \frac{\sigma^2}{n} = V(\text{srsr-n}, \bar{y}),$$

as such \bar{t}' constitutes an unbiased and intermediate in precision strategy for the already classic $(\text{srs-n}, \bar{y})$ and $(\text{srsr-n}, \bar{y})$ in the sense studied by Ruiz et al. (1989).

4. Variance estimation of \bar{t}'

We can unbiasedly estimate S^2 from the integrated and linear way

$$v' = \sum_{i=1}^k v_i' s_i^2$$

with $v_i' \geq 0$ and

$$\sum_{i=1}^k v_i' = 1$$

The optimum linear integrated estimator assigns (and if $n_i \rightarrow \infty$, $i = 1, 2, \dots, k$, are asymptotically equivalent to)

$$\bar{v}_i' = (1/a_i) / \left(\sum_{j=1}^k (1/a_j) \right) \quad (\leftrightarrow n_i/n)$$

for \bar{v}_i' , being

$$a_i \equiv V(\text{srsr-n}_i, s_i^2)$$

given in Hansen et al. (1953). From which (and asymptotically equivalent, if $n_i \rightarrow \infty$, $i = 1, 2, \dots, k$, to)

$$V(\hat{V}(\bar{v}')) = \sum_{i=1}^k (1/a_i) \quad \left(\leftrightarrow \frac{n}{\mu_4 - \sigma^4} \right).$$

As such the integrated estimator of the variance of \bar{t} is

$$\hat{V}(\bar{t}) = \frac{\sum_{i=1}^k (s_i^2/a_i)}{N \left(\sum_{i=1}^k \frac{n_i}{N-n_i} \right) \left(\sum_{j=1}^k (1/a_j) \right)}$$

which depends of unknown parameters such as μ_4 and σ^4 , but asymptotically, if $n_i \rightarrow \infty$ ($i = 1, 2, \dots, k$), it will be equivalent to

$$\hat{V}(\bar{t}) \leftrightarrow \frac{\sum_{i=1}^k \frac{n_i}{n} s_i^2}{N \left(\sum_{j=1}^k \frac{n_j}{N-n_j} \right)}$$

which does not depend of any unknown parameter. Furthermore, from Hansen et al. (1953), $n_i \rightarrow \infty$ ($i = 1, 2, \dots, k$), we have equivalently

$$V(\hat{V}(\bar{t})) \leftrightarrow \frac{\mu_4 - \sigma^4}{N^2 n \left(\sum_{i=1}^k \frac{n_i}{N-n_i} \right)^2}$$

which is asymptotically less than $V(\hat{V}(\bar{t}))$ given in the previous section 2.

References

- Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953) : *Sample Survey Methods and Theory (Volume II)*. John Wiley, New York.
- Raj, D. (1968) : *Sampling Theory*. McGraw-Hill. New York.
- Ruiz, M. (1992) *Optimality of sample mean*. Rev. Acad. Cienc. Zaragoza **47**, 125-128.
- Ruiz, M. and Santos, J. (1989) : *Intermediate sampling strategies*. Estadística Española, **31**, 227-235.

Un estimador más preciso que la media muestral para estudios analíticos.

M. Ruiz Espejo

Departamento de Economía Aplicada Cuantitativa
Facultad de Ciencias Económicas y Empresariales (U.N.E.D.)
Apartado 60.142. 28080 - MADRID.

Summary.

In a paper by Ruiz [2] it is proved that for certain conditions, the sample mean is not UMV (uniformly of minimum variance), counterexample to Koop [1] result when a wider kind of unbiased estimators of the population mean of the infinite population, used for analytic studies, is considered. In this paper we prove the existence of a such estimator, and we give its explicit mathematical expression.

1. Introducción.

El estudio de la caracterización de estimadores insesgados UMV en la inferencia para estudios analíticos fue iniciado por Koop [1]. En este trabajo se demostró que la media muestral, basada en el tamaño muestral efectivo, es insesgado UMV para la media de la población infinita, μ , en estudios analíticos para cierta clase de estimadores.

Posteriormente, Ruiz [2] justificó que ampliando ligeramente la clase de estimadores, que admitan incorporar en el estimador el conocimiento de dos cuantiles de la población infinita, la media muestral, aun siendo insesgada, deja de ser UMV.

Sin embargo en el trabajo no se proporciona ninguna expresión matemática explícita de un estimador que mejore en algún caso a la media muestral y_s .

En la demostración de Ruiz [2] se utiliza la existencia de una variable aleatoria conocida, z , de esperanza nula y cuya covarianza con y_s es $\text{Cov}(y_s, z) > 0$.

2. Expresión explícita del estimador

Llamamos \bar{t} al estimador insesgado, para la media de la población infinita, μ , definido por la igualdad

$$\bar{t} = \bar{y}_s + kz, \quad (2.1)$$

donde k es una constante real por determinar. La varianza del estimador \bar{t} resulta ser

$$V(\bar{t}) = V(\bar{y}_s) + k^2V(z) + 2k\text{Cov}(\bar{y}_s, z). \quad (2.2)$$

Si

$$k = -\text{Cov}(\bar{y}_s, z)/V(z), \quad (2.3)$$

que para una población finita, procedente de un modelo infinito, es una constante real, entonces

$$V(\bar{t}) = V(\bar{y}_s) - \frac{\text{Cov}^2(\bar{y}_s, z)}{V(z)} = V(\bar{y}_s) (1-\rho^2) < V(\bar{y}_s), \quad (2.4)$$

siendo $\rho > 0$ el coeficiente de correlación lineal de \bar{y}_s y z . Es decir, \bar{t} es insesgado y, uniformemente, más preciso que \bar{y}_s . Con lo cual se deduce que \bar{y}_s deja de ser UMV e incluso deja de ser admisible, puesto que es uniformemente superado en precisión por el estimador \bar{t} , con la constante $k = -\text{Cov}(\bar{y}_s, z)/V(z)$, en cualquier punto o vector y del espacio real N -dimensional, siendo N el tamaño de la población finita observable.

La constante k existe, puesto que $V(z) > 0$, al ser $\text{Cov}(\bar{y}_s, z) > 0$, como se probó en Ruiz [2].

La expresión matemática explícita del estimador \bar{t} , según (2.1) y (2.3), es

$$\bar{t} = -\frac{\text{Cov}(\bar{y}_s, z)}{V(z)} z \quad (2.5)$$

que siempre existe para cualquier diseño, no "por conglomerados unietápico con muestras de un solo conglomerado".

Desde un punto de vista práctico, la varianza $V(z)$ es calculable (Ruiz [2]), en tanto que la covarianza $Cov(y_s, z)$ no es calculable aunque sí estimable.

Por otra parte, el estimador t es más preciso que \bar{y}_s para cualquier constante k en el intervalo abierto $(-2Cov(y_s, z)/V(z), 0)$, alcanzando $V(t)$ el mínimo para $k = -Cov(y_s, z)/V(z)$, por lo que una decisión prudente consiste en tomar un valor de k negativo, suficientemente próximo a 0, que haga muy probable o casi segura la mejora de t , en precisión, sobre la media muestral.

Referencias

- [1] Koop, J.C. (1986): *Some problems of statistical inference from sample survey data for analytic studies*. Statistics **17**, 237-247.
- [2] Ruiz, M. (1993): *Sobre estimadores UMV para estudios analíticos*. Estadística **45**, 107-116.

1. Introducción.

El problema que vamos a tratar consiste en estudiar para un diseño muestral con probabilidades iguales de selección p de una muestra cualquiera, si existe sesgo al considerar una estrategia basada en un estimador t insesgado para una función paramétrica T cuando en la práctica se dispone de dos estimaciones independientes t_1 y t_2 con el mismo diseño, y se estima T por el valor de la estimación mayor t_j .

La respuesta es que t_j sobrestima en términos esperados a la función paramétrica T , o en otras palabras tiene sesgo positivo si y sólo si el estimador t admite al menos dos valores distintos. Sólo será t_j insesgado para T cuando y sólo cuando t sea constante para todas las muestras posibles con probabilidad positiva igual a " p ".

Aparentemente podría pensarse que no habría sesgo al considerar que tanto la estimación t_1 como t_2 se han obtenido con el mismo diseño, y que las muestras respectivas s_1 y s_2 son igualmente probables; sin embargo, el hecho de comparar t_1 y t_2 e intencionalmente seleccionar la estimación mayor de ellas, o equivalentemente la muestra correspondiente tras la comparación, de modo que se prefiere la estimación mayor, reduce en un sesgo positivo o sobrestimación de la función paramétrica T . La justificación del

Sesgo de intencionalidad en la selección de la estrategia muestral

M. Ruiz Espejo

Departamento de Economía Aplicada Cuantitativa
Facultad de Ciencias Económicas y Empresariales (U.N.E.D.)
Apartado 60.142. 28080 MADRID.

Summary.

In this paper, we prove that the systematical selection derivelately of a strategy, between two independent estimations with common design of equal probabilities of each possible sample, implies a bias for estimating the parametric function T , when the original strategy was unbiased for T . Some consequences and applications are suggested.

1. Introducción.

El problema que vamos a tratar consiste en estudiar para un diseño muestral con probabilidades iguales de selección, p , de una muestra cualquiera, s , si existe sesgo al considerar una estrategia basada en un estimador t insesgado para una función paramétrica T cuando en la práctica se dispone de dos estimaciones independientes $t_i \leq t_j$ con el mismo diseño, y se estima T por el valor de la estimación mayor t_j .

La respuesta es que t_j sobreestima en términos esperados a la función paramétrica T , o en otras palabras tiene sesgo positivo sí y sólo sí el estimador t admite al menos dos valores distintos. Sólo será t_j insesgado para T cuando y sólo cuando t sea constante para todas las muestras posibles con probabilidad positiva igual a " p ".

Aparentemente podría pensarse que no habría sesgo al considerar que tanto la estimación t_i como t_j se han obtenido con el mismo diseño, y que las muestras respectivas s_i, s_j , son igualmente probables; sin embargo, el hecho de comparar t_i y t_j e intencionalmente seleccionar la estimación mayor de ellas, o equivalentemente la muestra correspondiente tras la comparación, de modo que se prefiere la estimación mayor, redundará en un sesgo positivo o sobreestimación de la función paramétrica T . La justificación del

resultado esbozado hasta aquí es el objeto de los siguientes párrafos, así como de sus consecuencias inmediatas.

2. Nuevo diseño muestral.

Si disponemos de un número finito de muestras, s_i , con probabilidad común "p", las posibles muestras pueden ordenarse por el subíndice

$$s_1 \lesssim s_2 \lesssim \dots \lesssim s_i \lesssim \dots \lesssim s_{(1/p)}$$

si y sólo si

$$t_1 \leq t_2 \leq \dots \leq t_i \leq \dots \leq t_{(1/p)}$$

siendo $t_i = t(s_i)$ el i-ésimo estadístico ordenado.

Admitiendo la intencionalidad explicada, el diseño muestral nuevo resultante, será

$$p(s_1) = p(s_1, s_1) = p^2$$

$$p(s_2) = p(s_1, s_2) + p(s_2, s_1) + p(s_2, s_2) = 3p^2$$

$$p(s_3) = 2p(s_1, s_3) + 2p(s_2, s_3) + p(s_3, s_3) = 5p^2$$

$$p(s_i) = 2 \sum_{j=1}^{i-1} p(s_j, s_i) + p(s_i, s_i) = [2(i-1) + 1]p^2 = (2i-1)p^2$$

$$p(s_{(1/p)}) = (2/p - 1)p^2 = 2p - p^2 = p(2-p)$$

Efectivamente $p(s_i)$ es un diseño muestral para $i = 1, 2, \dots, 1/p$. En primer lugar, $p(s_i) > 0$ para todo $i = 1, 2, \dots, 1/p$; y además:

$$\sum_{i=1}^{1/p} p(s_i) = \sum_{i=1}^{1/p} (2i-1)p^2 = p^2 \sum_{i=1}^{1/p} (2i-1) = p^2 \left(2 \sum_{i=1}^{1/p} i - \frac{1}{p} \right) = p^2 \frac{1}{p^2} = 1$$

Hemos supuesto que t_i es insesgado para T , es decir,

$$T = \sum_{i=1}^{1/p} p t_i = E(t_i)$$

Al introducir la intencionalidad, podemos considerar la variable aleatoria bidimensional (t_i, i) como auxiliar de trabajo, con $p(t_i, i) = p$. Entonces, la nueva esperanza matemática de t_i , resulta ser:

$$\sum_{i=1}^{1/p} t_i p(s_i) = \sum_{i=1}^{1/p} t_i (2i-1) p^2 = p \sum_{i=1}^{1/p} t_i (2i-1) p = p(2\alpha_{11} - \alpha_{10})$$

de modo que será mayor o igual a $T = \alpha_{10}$, si y sólo si se tiene

$$2\alpha_{11} \geq \left(1 + \frac{1}{p}\right) \alpha_{10} = 2\alpha_{10} \alpha_{01} \quad (2.1)$$

y de la última igualdad se deduce que

$$\alpha_{01} = E(i) = \sum_{i=1}^{1/p} i p = p \sum_{i=1}^{1/p} i = \frac{1}{2} \left(1 + \frac{1}{p}\right)$$

Por tanto, la inecuación (2.1) se dará si y sólo si la covarianza es positiva, es decir

$$\mu_{11} = \text{Cov}(t_i, i) = \alpha_{11} - \alpha_{10} \alpha_{01} \geq 0,$$

lo que es siempre cierto, pues el coeficiente de correlación lineal $\rho(t_i, i)$ verifica la condición $\rho(t_i, i) \geq 0$, de tal modo que la igualdad sólo se cumple si t_i es constante para todo $i = 1, 2, \dots, 1/p$.

En consecuencia, la nueva estrategia es tanto más sesgada positivamente cuanto mayor sea el valor positivo de la covarianza μ_{11} .

3. Consecuencias.

De modo similar, el diseño intencional en el que se toma el estadístico menor entre los dos independientes disponibles, procedentes de un mismo diseño con probabilidad constante para las distintas muestras, produce siempre un sesgo negativo tanto menor cuanto mayor sea μ_{11} (que es siempre positiva, o cero para t_i siempre constante).

Entre los diseños iniciales con probabilidades constantes para las diferentes muestras, podemos destacar como más usados (Véase Fernández-Abascal y otros [2]), los siguientes:

a) $p = 1/\binom{N}{n}$,

procedente de muestras de combinaciones de N unidades de la población tomadas de n en n ; (muestreo aleatorio simple con reemplazamiento).

b) $p = 1/N^n$,

procedente de muestras de variaciones con repetición de N unidades de la población tomadas de n en n ; (muestreo aleatorio simple con reemplazamiento).

c) $p = 1/\binom{N+n-1}{n}$,

procedente de muestras de combinaciones con repetición de N unidades poblacionales tomadas de n en n .

d) $p = 1 / \left[\binom{N}{n} n! \right]$,

procedente de muestras de variaciones de N unidades de la población tomadas de n en n .

Entre las funciones paramétricas T a las que pueden aplicarse estos resultados son, por ejemplo, la media poblacional y la varianza poblacional en los casos a) y b), y la media poblacional en c) y d), con adecuados estimadores insesgados para tales diseños, aunque estos son casos particulares de las implicaciones más amplias en el modelo de muestreo de poblaciones finitas fijadas (Cassel et al...[1]).

Referencias.

- [1] Cassel, C.M. Särndal, C.E. and Wretman, J.H. (1977): *Foundations of Inference in Survey Sampling*. John Wiley. New York.
- [2] Fernández-Abascal, H. Guijarro, M.M., Rojo, J.L. y Sanz, J.A. (1994): *Cálculo de Probabilidades y Estadística*. Ariel. Barcelona.

ENERGIA DE UNA SEÑAL USANDO EL TEOREMA DE SHANNON

L. Agud y O.R. García

Departamento de Matemáticas. Area de Análisis Matemático.

Facultad de Ciencias.

50009 - Universidad de Zaragoza.

Abstract

In this article, the fisical concept of signal energy is treated, working with no-analogic signals. In that way we have attained different expressions of the energy using evaluations of the signal and its derivatives until the choosen order. We have also studied the way errors in the evaluations of the samples disturb the value of the energy obtained using an expression of the described ones.

§1.- INTRODUCCION.

Una palabra clave en el presente artículo es el concepto de señal. Matemáticamente, entenderemos por señal una función cualquiera. Para dar más uniformidad al estudio admitiremos que una señal $x(t)$ pueda tomar valores complejos. Físicamente exigiremos que sea una función temporal o espacial que transporte información. Esto último hace que la función constante no resulte interesante.

Tomaremos el par de transformadas de Fourier [1][4]:

$$X(\nu) = \int_{-\infty}^{+\infty} x(t) e^{-2\pi j \nu t} dt \quad (1.1)$$

$$x(t) = \int_{-\infty}^{+\infty} X(\nu) e^{2\pi j \nu t} d\nu \quad (1.2)$$

A la transformada de Fourier de la señal $x(t)$ la llamaremos espectro de la señal y lo denotaremos con $X(\nu)$, debido a la analogía existente con el espectro de la luz, ya que la expresión (1.1) no es más que un análisis de la proporción de frecuencia ν que se encuentra presente en la señal $x(t)$.

Diremos que una señal es de banda limitada con anchura de banda B si su espectro es de soporte acotado contenido en el intervalo $[-B, B]$.

A lo largo del artículo utilizaremos frecuentemente la señal seno compacto y su espectro, que denotaremos respectivamente $\text{sinc}(t)$ y $\Pi(\nu)$ y que vendrán definidas como en [2] por:

$$\text{sinc}(t) = \begin{cases} \frac{\text{sen}(\pi t)}{\pi t} & t \neq 0 \\ 1 & t = 0 \end{cases} \quad (1.3)$$

$$\Pi(\nu) = \begin{cases} 1 & -1/2 \leq \nu \leq 1/2 \\ 0 & \text{otro caso.} \end{cases} \quad (1.4)$$

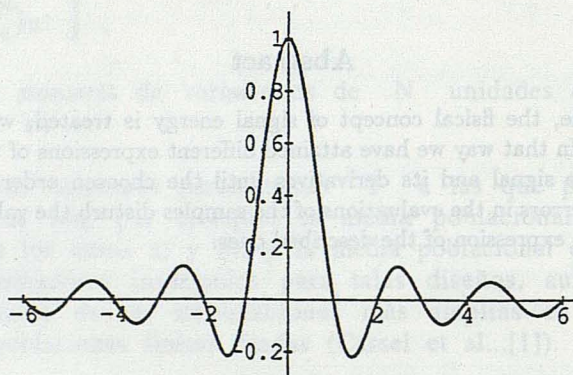


Figura 1.- Señal $\text{sinc}(t)$

El presente artículo está dedicado a un aspecto físico tan importante en una señal como es su energía. Definimos la energía de una señal $x(t)$ como el cuadrado de su norma-2 y la denotaremos por $K[x]$ [3][4].

$$K[x(t)] = \int_{-\infty}^{\infty} |x(t)|^2 dt \quad (1.5)$$

Esta definición se ve justificada por el siguiente hecho físico: Una de las leyes más importantes en el campo de la electricidad es la ley de Ohm, que afirma que la diferencia de potencial entre los extremos de un circuito que tiene una resistencia, es igual al producto de la intensidad que circula por el valor de dicha resistencia. Esto se traduce en la siguiente fórmula:

$$v = iR$$

Por otra parte, de todos es conocida la fórmula de la potencia (derivada de la energía con respecto al tiempo) en la que al hacer uso de la ley de Ohm obtenemos la expresión siguiente:

$$W = v i = \frac{v^2}{R}$$

con lo cual obtenemos la expresión siguiente para la energía:

$$K[v] = \int_{-\infty}^{\infty} W(t) dt = \frac{1}{R} \int_{-\infty}^{\infty} v^2(t) dt$$

que se ajusta perfectamente a la definición introducida.

Utilizaremos la herramienta que nos proporciona el teorema de Shannon sobre el conocimiento de una señal a partir de sus muestras, de forma que llegaremos a una expresión de la energía de la señal a partir de ellas.

Posteriormente veremos cómo errores en la obtención o evaluación de la función en distintos puntos, que darán origen a las muestras de las que partamos, influirán en la obtención de la energía de la señal y calcularemos la esperanza matemática de la variable aleatoria que mide la diferencia de energía entre la señal de partida y la señal modificada por los errores.

Debido a que estaremos usando iniciales para denotar errores, energías y esperanzas matemáticas, hemos tenido que hacer uso de la siguiente notación:

$E \rightarrow$ para denotar el operador esperanza matemática.

$K \rightarrow$ para denotar el operador energía.

$F \rightarrow$ para denotar el error cometido.

Estas dos últimas iniciales han sido tomadas de las palabras alemanas Kraft (energía) y Fehler (error), puesto que tanto en castellano como en inglés o francés los tres términos utilizados tienen la misma inicial.

El pilar en el que cimentaremos nuestro trabajo es el Teorema de Shannon [5][6], también conocido con el nombre de teorema del muestreo, que asegura:

Para toda señal de energía finita y de banda limitada, con anchura de banda B , se tiene:

$$x(t) = \sum_{k=-\infty}^{\infty} x(k\tau) \operatorname{sinc}\left(\frac{t}{\tau} - k\right) \quad (1.6)$$

$\forall \tau \leq \tau_0$, donde $\tau_0 = \frac{1}{2B}$.

La fórmula (1.6) es cierta en $L^2(\mathbb{R})$ en el sentido de valor propio de Cauchy. Y la convergencia es uniforme si:

$$\sum_{n=-\infty}^{\infty} |x(n\tau)| < +\infty$$

lo cual es cierto para el tipo de señales que estamos considerando.

Más adelante utilizaremos la noción de filtro [7][8][9]. Entenderemos por filtro cualquier transformación T verificando $T(x(t)) = y(t)$ con un determinado sentido físico. Por ello le impondremos que cumpla las siguientes dos propiedades de linealidad e invarianza en el tiempo:

$$\left. \begin{aligned} T(x(t)) &= y(t) \\ T(z(t)) &= \omega(t) \end{aligned} \right\} \implies T(\lambda x(t) + \mu z(t)) = \lambda y(t) + \mu \omega(t) \quad (1.7)$$

$$T(x(t - t_0)) = y(t - t_0) \quad (1.8)$$

La salida del filtro $y(t)$ es, en función de la entrada $x(t)$, el producto de convolución siguiente: $y(t) = (x * h)(t)$ donde $h(t)$ es la respuesta del filtro al impulso, es decir, $h(t) = T(\delta(t))$ y por tanto es fija para cada filtro.

Si consideramos la actuación del filtro observando el comportamiento frecuencial en lugar del temporal y por tanto estudiamos las salidas cuando la entrada es el espectro de una señal, observamos una acción del tipo:

$$Y(\nu) = X(\nu) H(\nu) \quad (1.9)$$

donde $H(\nu)$ es la transformada de Fourier de $h(t)$. La función $h(t)$ recibe el nombre de *respuesta al impulso* y $H(\nu)$ se llama *función de transferencia del filtro*.

§2.- INTERPRETACION DEL TEOREMA DE SHANNON.

Veamos en este párrafo cómo actúa el teorema del muestreo para recomponer la señal $x(t)$. Además, esto servirá principalmente para destacar la importancia de tomar frecuencias en el muestreo $\tau \leq \tau_0$, donde τ_0 denota la frecuencia de Nyquist.

Para realizar el muestreo de la señal $x(t)$, basta con multiplicar la señal en cuestión por una sucesión de impulsos de altura la unidad separados la anchura con que deseamos muestrear. Esta sucesión de impulsos se denomina Peine de Dirac.

La nueva señal, que llamamos \bar{x} está formada por impulsos separados una cierta anchura τ de altura igual al valor de la señal $x(t)$ en esos puntos.

Esto queda reflejado en la figura 2.

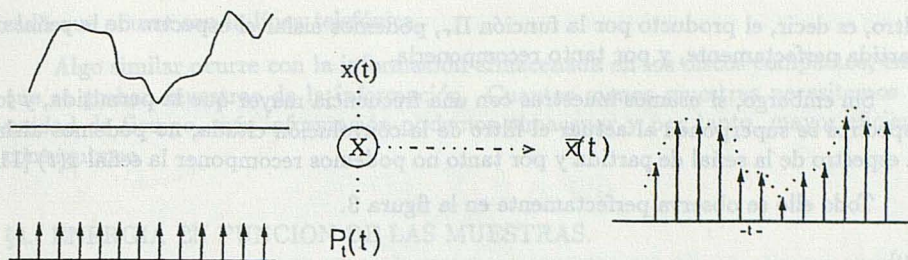


Figura 2.- Muestreo de la señal $x(t)$.

Estudiemos cómo funciona el teorema de Shannon observando el plano frecuencial, es decir, la transformada de Fourier de todo cuanto hagamos. Así, el producto anterior de la señal con el peine de Dirac [10], se convierte en el producto de convolución de sus respectivos espectros.

Denotamos con X el espectro de la señal $x(t)$ y con P_1 el espectro del peine de Dirac de frecuencia τ . Observamos que es otro peine de frecuencia inversa a la del primero y altura también la unidad.

Debido a que muestreamos con una frecuencia menor o igual a la frecuencia de Nyquist, al realizar el producto de convolución señalado, podemos asegurar que los espectros no se van a superponer.

Si la frecuencia con la que muestreamos es mayor que la frecuencia de Nyquist, entonces los espectros se superponen al realizar la convolución. A continuación veremos en qué se traduce esto.

Estos dos casos se encuentran reflejados en la figura 3.

El siguiente paso para la obtención de la expresión:

$$\sum_{n=-\infty}^{\infty} x(n\tau) \operatorname{sinc}\left(\frac{t}{\tau} - n\right)$$

es realizar la convolución $(\bar{x}(t)) * (\operatorname{sinc}(\frac{t}{\tau}))(\nu)$.

En función de sus espectros, lo que tenemos es el producto siguiente:

$$X(\nu) \Pi_{\tau}(\nu)$$

Donde $\Pi_{\tau}(\nu)$ es el espectro de la señal sinc y es la función que vale 1 para las frecuencias comprendidas entre $\frac{-1}{2\tau}$ y $\frac{1}{2\tau}$, y se anula en el resto, [10].

Si la frecuencia con la que hemos muestreado es menor o igual a la de Nyquist, hemos visto que los espectros no se superponen. De esta forma al hacer el paso del

filtro, es decir, el producto por la función Π_T , podemos aislar el espectro de la señal de partida perfectamente, y por tanto recomponerla.

Sin embargo, si usamos muestras con una frecuencia mayor que la permitida, y los espectros se superponen al actuar el filtro de la convolución citada, no podemos aislar el espectro de la señal de partida y por tanto no podemos recomponer la señal $x(t)$ [11].

Todo ello se observa perfectamente en la figura 3.

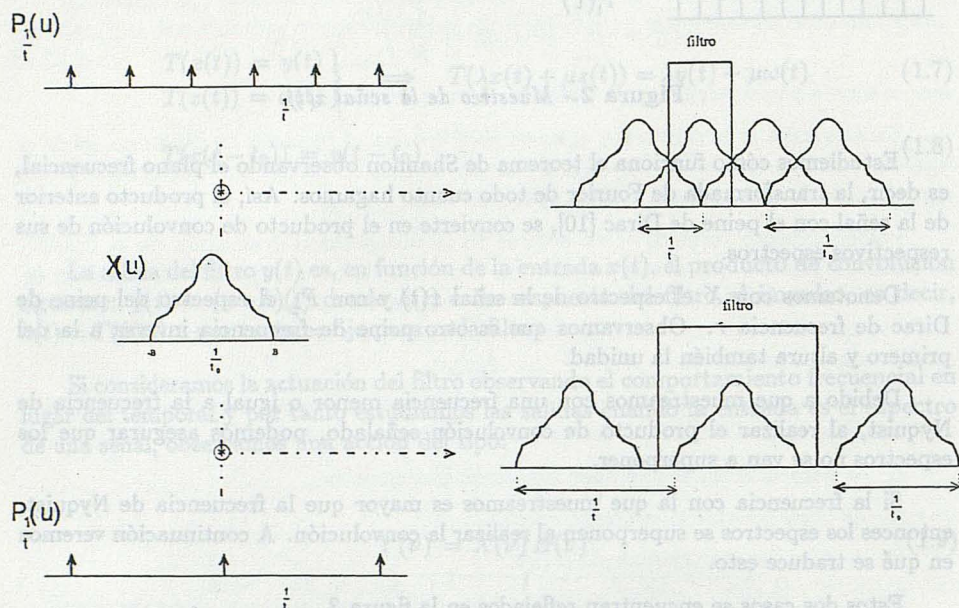


Figura 3.- Filtrado de $\tilde{X}(\nu)$.

Físicamente la realización de un peine de Dirac perfecto no es posible. Por ello en la realidad, en lugar de trabajar con una sucesión de impulsos, se trabaja con una sucesión de pulsos.

El resultado al convolucionar espectros es que el lóbulo central sigue siendo el espectro de la señal de partida y los anteriores y posteriores se van desvirtuando. Sin embargo como es el lóbulo central el que se utiliza para el paso del filtro, la recomposición de la señal sigue siendo perfecta.

§3.- IMPORTANCIA TECNICA.

La importancia técnica del teorema radica en la utilización de canales de transmisión de señal. Si enviamos información de forma analógica, ie $x(t) \forall t \in \mathbb{R}$, ocupamos el canal de transmisión continuamente, mientras que si enviamos sólo muestras de la señal, entre dos muestras, el canal queda libre y podemos enviar otras señales. Esto

es lo que ocurre con la línea telefónica.

Algo similar ocurre con la información almacenada en los discos compactos, en los que se graban muestras de la información. Cuantas menos muestras necesitemos por unidad de tiempo, más información podemos almacenar y por tanto, mayor eficiencia conseguimos.

§4.- ENERGIA EN FUNCION DE LAS MUESTRAS.

En este apartado vamos a calcular la energía de una señal partiendo únicamente de un muestreo de ésta.

Para la obtención de dicha expresión, usaremos de forma importante la recomposición que proporciona el Teorema de Shannon, así como la ortogonalidad de las señales $\{\text{sinc}(nt_0)\}_{n \in \mathbb{Z}}$ en $L^2(\mathbb{R})$.

De esta forma por (1.6), $\forall \tau \leq \tau_0$:

$$K[x(t)] = K\left[\sum_{k=-\infty}^{\infty} x(k\tau) \text{sinc}\left(\frac{t}{\tau} - k\right)\right] \quad (4.1)$$

Debido a que la energía de la suma de señales ortogonales es igual a la suma de energías de cada una de dichas señales, podemos escribir:

$$\begin{aligned} K[x(t)] &= \sum_{k=-\infty}^{\infty} K[x(k\tau) \text{sinc}\left(\frac{t}{\tau} - k\right)] \\ &= \sum_{k=-\infty}^{\infty} |x(k\tau)|^2 K[\text{sinc}\left(\frac{t}{\tau} - k\right)] \\ &= \sum_{k=-\infty}^{\infty} |x(k\tau)|^2 \int_{-\infty}^{\infty} |\text{sinc}\left(\frac{t}{\tau} - k\right)|^2 dt \end{aligned} \quad (4.2)$$

Utilizando la invarianza por traslaciones de la norma euclídea y el Teorema de Parseval:

$$\langle x(t), y(t) \rangle = \langle X(\nu), Y(\nu) \rangle$$

se llega a:

$$K[x] = \tau \sum_{k=-\infty}^{\infty} |x(k\tau)|^2 \quad (4.3)$$

§5.- EXPRESION DE LA ENERGIA EN FUNCION DE LA PRIMERA DERIVADA.

En lo que sigue supondremos $x(t)$ una señal real.

Consideramos la expresión (4.3), que puede verse como una función de τ :

$$H(\tau) = \tau \sum_{k=-\infty}^{\infty} [x(k\tau)]^2 \quad (5.1)$$

Esta función resulta ser constante respecto de τ en el intervalo $(0, \tau_0)$, debido a que la expresión representa la energía de una señal, la cual no depende realmente de la separación que se tome entre las muestras, como prueba su obtención en el apartado anterior. Por ello, $H(\tau) \in C^\infty(\mathbb{R})$.

Luego derivando respecto dicha variable la expresión (5.1), se obtiene:

$$0 = H'(\tau) = \sum_{k=-\infty}^{+\infty} [x(k\tau)]^2 + 2\tau \sum_{k=-\infty}^{+\infty} [k x(k\tau) x'(k\tau)]$$

Por ello:

$$2\tau \sum_{k=-\infty}^{+\infty} [k x(k\tau) x'(k\tau)] = - \sum_{k=-\infty}^{+\infty} [x(k\tau)]^2 \Rightarrow$$

$$2\tau^2 \sum_{k=-\infty}^{+\infty} [k x(k\tau) x'(k\tau)] = -\tau \sum_{k=-\infty}^{+\infty} [x(k\tau)]^2$$

Así, se llega a la fórmula que nos da otra expresión de la energía en la que también interviene la primera derivada de nuestra señal :

$$K[x] = -2\tau^2 \sum_{k=-\infty}^{+\infty} [k x(k\tau) x'(k\tau)] \quad (5.2)$$

§6.- EXPRESION DE LA ENERGIA EN FUNCION DE SUCEVAS DERIVADAS.

En este párrafo vamos a generalizar el resultado para obtener una nueva fórmula que nos permita expresar la energía de toda señal de banda limitada en función de las derivadas de cualquier orden de derivación superior deseado:

$$K[x] = \frac{(-1)^k \tau^{k+1}}{k!} \sum_{n=-\infty}^{\infty} n^k \left[\sum_{i=0}^k \binom{k}{i} x^{(i)}(n\tau) x^{(k-i)}(n\tau) \right] \quad (6.1)$$

Esta generalización se obtiene por el proceso de inducción sobre k , siendo k el orden de derivación con el que nos interesa trabajar. Por lo que el párrafo anterior sería con $k = 1$, paso previo para nuestra inducción. Así:

* Para $k = 1$ es cierto, como hemos visto en el párrafo anterior.

* Supuesto cierto para k veamos para $k + 1$:

Para la demostración vamos a reiterar los pasos necesarios para calcular $K[x]$, partiendo de:

$$K[x] = \frac{(-1)^k \tau^{k+1}}{k!} \sum_{n=-\infty}^{\infty} n^k \left[\sum_{i=0}^k \binom{k}{i} x^{(i)(n\tau)} x^{(k-i)(n\tau)} \right] \quad (6.1)$$

Derivando respecto τ dicha expresión y pasando al otro miembro, se llega a:

$$\begin{aligned} & \frac{(-1)^{k+1} (k+1) \tau^k}{k!} \sum_{n=-\infty}^{\infty} n^k \left[\sum_{i=0}^k \binom{k}{i} x^{(i)(n\tau)} x^{(k-i)(n\tau)} \right] \\ &= \frac{(-1)^k \tau^{k+1}}{k!} \sum_{n=-\infty}^{\infty} n^k \left[\sum_{i=0}^k n \binom{k}{i} \left(x^{(i+1)(n\tau)} x^{(k-i)(n\tau)} \right. \right. \\ & \quad \left. \left. + x^{(i)(n\tau)} x^{(k-i+1)(n\tau)} \right) \right] \end{aligned}$$

Multiplicando en ambos miembros por $\frac{-\tau}{k+1}$ para obtener (2) en el primer miembro:

$$\begin{aligned} K[x] &= \frac{(-1)^k \tau^{k+1}}{k!} \sum_{n=-\infty}^{\infty} n^k \left[\sum_{i=0}^k \binom{k}{i} x^{(i)(n\tau)} x^{(k-i)(n\tau)} \right] \\ &= \frac{(-1)^{k+1} \tau^{k+2}}{(k+1)!} \sum_{n=-\infty}^{\infty} n^{k+1} \left[\sum_{i=0}^k \binom{k}{i} x^{(i+1)(n\tau)} x^{(k-i)(n\tau)} \right. \\ & \quad \left. + x^{(i)(n\tau)} x^{(k-i+1)(n\tau)} \right] \quad (6.2) \end{aligned}$$

Para que se verifique la inducción falta comprobar si es cierta la siguiente igualdad:

$$\begin{aligned} & \sum_{i=0}^k \left(\binom{k}{i} x^{(i+1)(n\tau)} x^{(k-i)(n\tau)} \right. \\ & \quad \left. + x^{(i)(n\tau)} x^{(k-i+1)(n\tau)} \right) \\ &= \sum_{i=0}^{k+1} \binom{k+1}{i} x^{(i)(n\tau)} x^{(k-i+1)(n\tau)} \quad (6.3) \end{aligned}$$

Gracias a la propiedad de números combinatorios que nos dice:

$$\binom{k+1}{j} = \binom{k}{j-1} + \binom{k}{j} \quad j \neq 0, k+1 \quad (6.4)$$

puede observarse que tanto el número de sumandos se corresponde, como que al desarrollar las sumas se tiene que los sumandos son los mismos. Por lo que la expresión (3) es cierta. Con lo cual hemos probado que la inducción se verifica y así se tiene una nueva expresión para la energía en función de derivadas superiores de la señal.

$$K[x] = \frac{(-1)^k \tau^{k+1}}{k!} \sum_{n=-\infty}^{\infty} n^k \left[\sum_{i=0}^k \binom{k}{i} x^{(i)(n\tau)} x^{(k-i)(n\tau)} \right] \quad (6.5)$$

Aparentemente esta fórmula presenta cierta complejidad debido tanto a la presencia de dos sumatorios, como a la utilización de algunas derivadas sucesivas.

Ahora bien, si tenemos en cuenta que tanto la señal como sus derivadas son prácticamente nulas fuera de un intervalo simétrico suficientemente grande, el primer sumatorio de (6.5) se reduce a una suma finita.

El segundo inconveniente parece ser más serio, pues estamos hablando de derivadas que hemos obtenido a partir de muestras de nuestra señal. Sin embargo, esta dificultad se puede resolver si recordamos cual es el espectro de la derivada y utilizamos el algoritmo de la transformada rápida de Fourier (F.F.T.).

Así, para obtener la primera derivada $x'(t)$ seguiríamos este esquema:

$$\begin{array}{ccc} x(t) & & x'(t) \\ \text{FFT} \downarrow & & \uparrow (\text{FFT})^{-1} \\ X(\nu) & \rightarrow & 2\pi j\nu X(\nu) \end{array}$$

Otra objeción que se puede presentar a la fórmula (6.5) es su poca utilidad por cuanto que se necesitan los mismos datos que en (4.3) (muestras de $x(t)$) y además muestras de las derivadas sucesivas.

A esto hemos de contestar que a veces las derivadas de una señal informan de aspectos interesantes de la señal que no se aprecian con los valores puntuales $x(t)$. Así recientemente Balcou [12][13] determina algunos parámetros de transitorias sinusoidales derivando respecto de la frecuencia.

La importancia de utilizar no sólo evaluaciones de la señal, sino también de sus muestras, nuevamente se hace patente en [14], donde tenemos un teorema del muestreo usando evaluaciones de la señal a recomponer y muestras de sus derivadas.

§7.- TRANSMISION DE ERRORES.

En todo lo que antecede hemos estado trabajando con muestras de la señal que es objeto de estudio. En este último párrafo, hemos tratado el tema de la energía de una señal y hemos visto una fórmula que permite calcularla a partir de tan sólo evaluaciones

de la propia señal en determinados puntos. Estos puntos vienen determinados por la frecuencia de Nyquist de la señal en cuestión.

Sin embargo, hasta el momento hemos olvidado un aspecto importante en el mundo real, al que intentamos adaptarnos. Este aspecto es la existencia de errores. Errores que pueden aparecer tanto por el aparato que utilicemos para evaluar la función en los distintos puntos, como por ruidos en la transmisión de datos, o por errores en la aproximación de los datos con los que operamos, etc...

El presente apartado pretende ver cómo errores debidos a mediciones afectan al valor de la energía de una señal, calculado con la fórmula obtenida en el párrafo cuarto.

Sea una señal $x(t)$ de banda limitada a la que le aplicamos el teorema del muestreo, obteniendo la representación (1.6).

Si cometemos errores de este tipo, que se traducen en alteraciones del valor de la evaluación de la señal en los distintos puntos, y que denotaremos por $\varepsilon(n\tau)$, entonces, al intentar recomponer la señal con estos errores, obtenemos otras señales, que llamaremos $\tilde{x}_\tau(t)$, y que están definidas para cada τ mediante la siguiente serie:

$$\tilde{x}_\tau(t) = \sum_{n=-\infty}^{\infty} [x(n\tau) + \varepsilon(n\tau)] \text{sinc}(2Bt - n) \quad (7.1)$$

Observemos que, en principio, para cada τ obtenemos una señal diferente. De esto trataremos más detenida y extensamente en el siguiente apartado.

La diferencia entre ambas señales es :

$$\tilde{x}_\tau(t) - x(t) = \sum_{n=-\infty}^{\infty} \varepsilon(n\tau) \text{sinc}(2Bt - n) \quad (7.2)$$

Con esta última igualdad, podemos conocer la energía de la diferencia entre la señal original y la modificada por los errores en las evaluaciones de la función, en los puntos que entran en juego en la aplicación del teorema del muestreo para una frecuencia determinada. La energía de una señal venía dada, según el resultado obtenido en el apartado cuarto, por la expresión (4.3).

Aplicando este resultado a la señal que nos ocupa en este caso, obtenemos:

$$K[\tilde{x}_\tau(t) - x(t)] = \tau \sum_{n=-\infty}^{\infty} |\varepsilon(n\tau)|^2 \quad (7.3)$$

Por otra parte, si calculamos la energía de la señal original, sin errores, obtenemos la siguiente expresión:

$$K[x(t)] = \tau \sum_{n=-\infty}^{\infty} |x(n\tau)|^2 \quad (4.3)$$

Y la energía de la señal modificada por los errores de evaluación en los distintos puntos quedará:

$$K[\tilde{x}_\tau(t)] = \tau \sum_{n=-\infty}^{\infty} |\varepsilon(n\tau) + x(n\tau)|^2 \quad (7.4)$$

Consideramos la diferencia entre la expresión obtenida para la energía de la señal cuando tenemos en cuenta los posibles errores cometidos en la evaluación de la señal y la expresión de la energía si suponemos que estos errores no tienen lugar.

$$\begin{aligned} K[\tilde{x}_\tau(t)] - K[x(t)] &= \tau \left\{ \sum_{n=-\infty}^{\infty} |x(n\tau) + \varepsilon(n\tau)|^2 - \sum_{n=-\infty}^{\infty} |x(n\tau)|^2 \right\} \\ &= \tau \left\{ \sum_{n=-\infty}^{\infty} |\varepsilon(n\tau)|^2 + 2 \sum_{n=-\infty}^{\infty} x(n\tau) \varepsilon(n\tau) \right\} \quad (7.5) \end{aligned}$$

§8.- TRANSMISION DE ERRORES RELATIVOS.

Comenzamos nuevamente con una señal $x(t)$ de banda limitada.

En el apartado anterior, estamos suponiendo que en la transmisión de las muestras o en la obtención de las mismas, podemos cometer unos errores absolutos que hemos llamado $\varepsilon(n\tau)$.

En realidad, como tanto τ como n son variables, el producto $n\tau$ puede tomar cualquier valor real. Así resulta que el error en la evaluación del valor de la señal puede ser cometido en cualquier punto. Por ello consideramos el proceso aleatorio continuo $\varepsilon(t)$, que señala para cada t_0 fijo una variable aleatoria que representa el error que se puede cometer en la obtención de $x(t_0)$. De este proceso continuo, para cada frecuencia fija τ , obtenemos un proceso aleatorio discreto, que es el que viene dado por la expresión $\varepsilon(n\tau)$.

Consideramos la señal modificada por los posibles errores que cometamos al evaluar la señal en cada punto. Sea esta la señal $\tilde{x}(t)$. Por tanto tenemos:

$$\tilde{x}(t) = x(t) + \varepsilon(t) \quad (8.1)$$

Aunque partimos de una señal $x(t)$ de banda limitada, es posible que la señal modificada por los errores ya no tenga espectro con soporte acotado. Por tanto no tenemos seguridad de que aplicando el teorema del muestreo para cada valor de τ , válido para la recomposición de $x(t)$, a la señal que contempla los errores, podamos recomponerla a partir de sus muestras. Así, para cada τ , tendremos, en general, una señal, que denotaremos $\tilde{x}_\tau(t)$, definida de la forma:

(8.2)

$$\tilde{x}_\tau(t) = \sum_{n=-\infty}^{\infty} \tilde{x}(n\tau) \operatorname{sinc}\left(\frac{t}{\tau} - n\right) \quad (8.2)$$

Hasta este momento hemos estado tratando el error cometido al evaluar la función en cada punto, como un error absoluto. Sin embargo, esto carece de interés en la práctica. Lo lógico es que el error sea mayor cuanto mayor sea la cantidad que estemos midiendo.

De aquí que resulte más conveniente considerar errores relativos en las mediciones. Definimos, pues, el error relativo cometido al evaluar la señal $x(t)$ en un punto como:

$$\varepsilon_r(t) = \begin{cases} \frac{\varepsilon(t)}{x(t)} & \text{si } x(t) \neq 0 \\ 0 & \text{en otro caso} \end{cases} \quad (8.3)$$

Notemos que aquí hablamos solamente de errores debidos a mediciones en las muestras y por tanto la anulación de $x(t)$ supone la no existencia de señal y por consiguiente la no existencia de error de medición. Naturalmente sí que puede haber ruido del canal, pero este no lo consideramos en este momento.

En el apartado anterior habíamos obtenido expresiones para la energía de la diferencia de las señales $x(t)$ y $\tilde{x}_\tau(t)$ y para la diferencia entre las energías de cada una de ellas, en función de muestras de la señal de partida y del proceso aleatorio discreto $(\varepsilon(n\tau))_n$, de la forma:

$$K[\tilde{x}_\tau - x] = \tau \sum_{n=-\infty}^{\infty} |\varepsilon(n\tau)|^2 \quad (1.22)$$

$$K[\tilde{x}_\tau] - K[x] = \tau \sum_{n=-\infty}^{\infty} [|\varepsilon(n\tau)|^2 + 2\varepsilon(n\tau)x(n\tau)] \quad (1.24)$$

Los primeros miembros de las dos igualdades anteriores, claramente, por lo comentado previamente sobre la aleatoriedad de $\varepsilon(n\tau)$, son variables aleatorias. Cada una de estas variables está definiendo un error. Denotaremos estos errores con las expresiones:

$$F_1(\tau) = K[\tilde{x}_\tau - x] \quad (8.4)$$

$$F_2(\tau) = K[\tilde{x}_\tau] - K[x] \quad (8.5)$$

Tenemos así expresadas dos variables aleatorias función de los errores absolutos que estamos cometiendo al evaluar la señal $x(t)$ en distintos puntos. En primer lugar, vamos a expresar estos dos errores en el cálculo de la energía en función de los errores relativos, pues así nos ajustamos mucho más a la realidad. Teniendo en cuenta que

$$\varepsilon(n\tau) = \varepsilon_r(n\tau) x(n\tau) \quad (8.6)$$

llegamos a la obtención de las expresiones:

$$F_1(\tau) = \tau \sum_{n=-\infty}^{\infty} |x(n\tau)|^2 |\varepsilon_r(n\tau)|^2 \quad (8.7)$$

$$F_2(\tau) = \tau \left\{ \sum_{n=-\infty}^{\infty} |x(n\tau)|^2 |\varepsilon_r(n\tau)|^2 + 2 \sum_{n=-\infty}^{\infty} |x(n\tau)|^2 |\varepsilon_r(n\tau)| \right\} \quad (8.8)$$

Estableceremos, a continuación, algunas hipótesis sobre estos errores relativos, como:

a) Los errores relativos cometidos en un punto son una variable aleatoria centrada, es decir, con media nula:

$$E[\varepsilon_r(t)] = 0 \quad \forall t \quad (8.9)$$

b) El proceso $(\varepsilon_r(n\tau))$ es un proceso discreto, estacionario de orden dos en sentido amplio, es decir, no hay zonas en el tiempo con mayor probabilidad de cometer más errores que en otras zonas. Esto se traduce en que la esperanza cuadrática de la variable aleatoria sea constante. Sea V esta constante:

$$E[(\varepsilon_r(n\tau))^2] = V \quad \forall n \quad \forall \tau \quad (8.10)$$

Como el proceso es centrado, esta esperanza coincide con la varianza.

Una vez hechas estas hipótesis de partida, calculamos las esperanzas de las dos variables aleatorias que tenemos. Así, tomamos la variable que tiene información sobre la energía de la diferencia entre la señal de partida y la modificada por los errores relativos y calculamos su media:

$$E[F_1(\tau)] = E\left[\tau \sum_{n=-\infty}^{\infty} |x(n\tau)|^2 |\varepsilon_r(n\tau)|^2\right]$$

usando la linealidad de la esperanza matemática obtenemos:

$$E[F_1(\tau)] = \tau \sum_{n=-\infty}^{\infty} |x(n\tau)|^2 E[|\varepsilon_r(n\tau)|^2]$$

y haciendo uso de la segunda hipótesis que hemos formulado, llegamos a las siguientes igualdades:

$$\begin{aligned}
 E[F_1(\tau)] &= \tau \sum_{n=-\infty}^{\infty} |x(n\tau)|^2 V \\
 &= V K[x]
 \end{aligned}
 \tag{8.11}$$

sin más que hacer uso de la expresión que obtuvimos para la energía de una señal en función de sus muestras en el apartado cuarto del presente artículo, y llamando $K[x]$ a la energía de la señal $x(t)$.

Si hacemos lo mismo con la diferencia entre las energías de las dos señales que estamos considerando, obtenemos:

$$\begin{aligned}
 E[F_2(\tau)] &= E \left[\tau \left\{ \sum_{n=-\infty}^{\infty} |x(n\tau)|^2 |\varepsilon_r(n\tau)|^2 + 2 \sum_{n=-\infty}^{\infty} |x(n\tau)|^2 |\varepsilon_r(n\tau)| \right\} \right] \\
 &= \tau \left\{ \sum_{n=-\infty}^{\infty} |x(n\tau)|^2 E[|\varepsilon_r(n\tau)|^2] + 2 \sum_{n=-\infty}^{\infty} |x(n\tau)|^2 E[|\varepsilon_r(n\tau)|] \right\} \\
 &= \tau \sum_{n=-\infty}^{\infty} |x(n\tau)|^2 V = V K[x]
 \end{aligned}
 \tag{8.12}$$

usando en este caso, además, que $E[\varepsilon_r(n\tau)] = 0$ por la primera hipótesis que hemos hecho sobre los errores relativos.

De estos resultados podemos sacar diferentes conclusiones. La primera de ellas es observar que ambas variables aleatorias F_1 y F_2 tienen la misma media.

De las fórmulas se observa además la lógica dependencia, en el error cometido, de la energía de la primera señal y de la varianza de los errores que cometamos. Claramente cabía esperar que a mayor varianza de los errores que cometamos, mayor separación de los errores de su media, mayor sería el error cometido al estimar la energía de la señal $x(t)$ por la energía de $\tilde{x}(t)$.

Observamos también que la esperanza de las variables aleatorias marcando el error cometido en el cálculo de la energía al tomar una señal u otra, no depende de la frecuencia, es decir, de τ , con la que tomemos las muestras. Esto se traduce en que, a efectos de valoración y estimación de la energía de la señal $x(t)$, no influye que estemos tomando más muestras en la recomposición de ésta, a través del teorema del muestreo. Por ello no resulta más eficiente tomar una frecuencia $\tau \leq \tau_0$, siendo τ_0 la frecuencia de Nyquist de la señal $x(t)$.

REFERENCIAS BIBLIOGRAFICAS:

- [1] R. Bracewell. 1978. *The Fourier Transform and its Applications*. Ed. McGraw-Hill, U.S.A.
- [2] A. Bruce Carlson. 1975. *Communications Systems*. Ed. McGraw-Hill, U.S.A.
- [3] H.J. Landau - H.O. Pollack. 1961. *The Bell System Technical Journal*. Vol.40, págs. 65-84.
- [4] D. Slepian, H.O. Pollack. 1961. *The Bell System Technical Journal*. Vol.40, págs. 43-63.
- [5] B. Picinbono. 1977. *Eléments de Théorie du signal*. Ed. Dunod, París.
- [6] I. Daubechies. 1992. *Ten Lectures on Wavelets*. Ed. SIAM, Vermont.
- [7] C. Gasquet - P. Witomski. 1990. *Analyse de Fourier et Applications*. Ed. Masson, París.
- [8] Athanasios Papoulis. 1987. *Signal Analysis*. Ed. McGraw-Hill, Singapur.
- [9] Charles S. Williams. 1986. *Designing Digital Filters*. Ed. Prentice-Hall International Editions, U.S.A.
- [10] B. Picinbono. 1989. *Théorie des signaux et des systèmes*. Ed. Dunod, París.
- [11] J. Auvray. 1979. *Electronique des signaux échantillonnés et numériques*. Ed. Dunod Université, París.
- [12] Y. Balcou. 1994. *Rapid Communications in mass spectrometry*. Vol.8, págs. 942-944.
- [13] Y. Balcou, J.P. Guegari, S. Sinbandhit. 1994. *Journal Chim Phys*. Vol.91, págs. 821-827.
- [14] D.A. Linden. 1959. *Proceedings of the IRE*. Vol.July, págs. 1219-1226.

Determinación de fluoruros en baños de galvanotecnia po electrodos selectivos de iones.

R.C. Díaz Ara , A. Novales Claraco

Departamento de Química Analítica

Escuela Universitaria Politécnica de Huesca

Crtra. Zaragoza, s/n. 22071 - HUESCA.

Abstract:

We have used a commercial fluoride ion - selective electrode for the determination of fluoride anion in galvanized baths. The proposed method use a calibration graph and the addition standard. They have been applied to the determination of fluoride in two types of Spanish galvanized baths. Fluoride was found in all samples in concentrations ranging between 0,0011 and 97,00 % (w/w). Some interfering ions have also been studied.

1.- Introducción.

La presencia de fluoruros en algunos baños de galvanotecnia es necesaria para el recubrimiento metálico de las superficies tratadas. Es necesario conocer el rango de concentraciones de fluoruro para obtener un buen recubrimiento metálico. Existe una amplia bibliografía sobre la determinación de fluoruros en muestras diversas (1-10), pero no hemos encontrado referencia alguna acerca de su determinación en baños de galvanotecnia de las características de los que analizamos. El uso de un electrodo selectivo de fluoruros nos permite determinar rápidamente la concentración de fluoruros. Este método propuesto puede usarse para el control de calidad de este tipo de baños de galvanotecnia.

2.- Parte experimental.

2.1. Aparatos.

Se ha utilizado un medidor digital pH/mV Orion 920 A, un electrodo selectivo de fluoruros Orion 94-01 y un electrodo de referencia de unión simple de Ag/AgCl, KCl saturado. Las determinaciones se han

llevado a cabo en un baño termostático Haake, a $20,0 \pm 0,2$ °C.

2.2. Reactivos.

Todos los reactivos son de grado analítico. Las muestras de NR-10 y NR-90 fueron facilitadas por industrias de galvanotecnia. La composición cualitativa de las muestras aparece en la Tabla 1.

TABLA 1

NR - 10	NR - 90
Fe^{2+}	F^-
Cr^{6+}	Na^+
SO_4^{2-}	K^+
SiF_6^{2-}	Ca^{2+}

Tabla 1: Composición cualitativa de dos muestras usadas en baños de galvanotecnia.

2.3. Procedimientos.

2.3.1. Estudio de las características del electrodo de fluoruros.

Se han estudiado las características analíticas del electrodo de fluoruros. El pH del medio, la fuerza iónica y la temperatura de la disolución pueden afectar a la respuesta del electrodo. Las condiciones óptimas para su uso son: el pH ha de estar comprendido entre 5,0 y 8,0 ; la fuerza iónica entre 0,5 y 1,6 M y la temperatura de $20,0 \pm 0,2$ °C , regulada con la ayuda de un baño termostático.

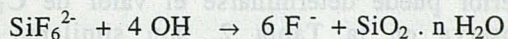
El ion hidróxido interfiere (11). Ajustando el pH entre 5,0-8,0 con una disolución de acetato sódico se elimina el efecto de los iones OH^- .

Se ha preparado una curva de calibrado con concentraciones patrón de fluoruro. La sensibilidad S, determinada por la pendiente de la curva de calibrado, es de $S = 55,10 \pm 0,25$ mV. década⁻¹. El rango de respuesta lineal es de 10^{-1} - 10^{-5} M. y el límite aproximado de detección es

de 10^{-5} M., calculado por la intersección de los dos tramos rectos de la curva de calibrado (12,13).

2.3.2. Tratamiento de las muestras.

Para las muestras de NR-10, se pesan 5 g., se trituran con un mortero y se disuelven con agua, añadiendo una disolución de carbonato sódico. La reacción que se produce es la siguiente (14) :



Esta reacción se produce a una temperatura de $50^\circ\text{--}60^\circ \text{C}$. A continuación, se deja enfriar y se filtra. El líquido filtrado se recoge en un matraz aforado de 50 ml. Se añade una disolución de acetato sódico en relación 10:1 (v/v) con respecto al volumen recogido en el filtrado (11) y se enrasa hasta 50 ml con agua destilada.

Para las muestras de NR-90, se pesan 5 g., se trituran y se disuelven en agua destilada, añadiendo la disolución de NaAc en relación 10:1, enrasando hasta 50 ml con agua destilada.

3.- Resultados y discusión.

3.1. Determinación por potenciometría directa.

Se toman 20 ml. de la disolución de la muestra, a la que se añade NaAc, que se coloca en un vaso de precipitados, junto al electrodo indicador de fluoruro y al de referencia, manteniendo la temperatura a 20°C . Se determina la concentración de fluoruro en la muestra por interpolación en la curva de calibrado, preparada a partir de disoluciones patrón de fluoruro.

Igualmente se ha determinado la concentración de fluoruro en las muestras (C_1) por la técnica de adición estandar. Para ello se toman 20 ml. de la disolución de la muestra, añadiendo NaAc, y midiendo a continuación el potencial de la muestra (E_1). Se añade 2 ml. de una disolución 10^{-3} M. de fluoruro (C_2), midiendo de nuevo el potencial resultante (E_2).

Haciendo uso de la ecuación de Nernst:

$$E_1 = K - S \lg C_1$$

siendo S la pendiente de la curva de calibrado.

$$E_2 = K - S \lg [(C_1 V_1 + C_2 V_2) / (V_1 + V_2)]$$

Restando $E_2 - E_1$:

$$\Delta E = E_2 - E_1 = -S \lg [(C_1 V_1 + C_2 V_2) / C_1 (V_1 + V_2)]$$

Y, finalmente:

$$C_1 = [C_2 V_2 / 10^{\Delta E/S} (V_1 + V_2) - V_1]$$

Usando la ecuación anterior puede determinarse el valor de C_1 . Los resultados, tal como se observa en la Tabla 2, son similares a los obtenidos por interpolación en la curva de calibrado. Como método de referencia se ha utilizado la valoración con nitrato de torio. Los valores son los resultados medios de 5 determinaciones.

TABLA 2

Muestra	Curva de calibrado	Adición estandar	Método de referencia
NR-10	0.0011	0.0010	0.001
NR-90	97.00	97.08	97.11

Tabla 2: Resultados obtenidos en la determinación de fluoruro en baños de galvanotecnia.

3.2. Estudios de reproducibilidad.

Se han realizado 10 determinaciones para una misma muestra. Los resultados que se obtienen conducen a un valor del 3,5% del coeficiente de variación.

3.3. Estudio de interferencias.

Los electrodos selectivos de iones (ESI) responden preferentemente al ion a determinar, aunque al potencial final contribuyen también el resto de los iones presentes en la disolución. La ecuación que rige la relación entre el potencial y la concentración de los iones en disolución es la de Eisenman - Nicolsky (15):

$$E = K + 2,303 \frac{RT}{z_i F} \lg [a_i + S K_{i,j}^{\text{POT}} \cdot a_j^{z_i/z_j}]$$

El estudio de interferencias se realiza calculando los valores de los coeficientes de selectividad potenciométricos ($K_{i,j}^{\text{POT}}$). Existen varias formas de determinación, dependiendo los valores encontrados del

método seguido y de la relación de actividades (o concentraciones si se mantiene constante la fuerza iónica de la disolución) entre el ion interferente y el ion fluoruro. En este caso se han determinado por el método de las disoluciones mezcladas, midiendo el potencial de disoluciones que contienen un volumen fijo de fluoruro (10^{-3} M.) y variando la concentración del ion interferente, de acuerdo a la ecuación (16):

$$K_{ij}^{\text{POT}} = -C_i V_0 / C_j V_x \quad \text{si } z_i = z_j$$

o bien:

$$K_{ij}^{\text{POT}} = -C_i (V_0 / C_j V_x)^{z_i / z_j} \quad \text{si } z_i < z_j$$

siendo:

$C_i = 10^{-3}$ M. la concentración inicial de fluoruro

$C_j = 1$ M. la concentración inicial del ion interferente

$V_0 = 20$ ml. el volumen inicial de la disolución de fluoruro

$z_i =$ carga del ion fluoruro

$z_j =$ carga del ion interferente

V_x (en ml.) se ha calculado experimentalmente por extrapolación lineal para $\Phi = 0$, siendo Φ la función:

$$\Phi = (V_0 + V)10^{E/S}$$

con:

$V =$ volumen añadido de disolución de ion interferente

$E =$ potencial determinado experimentalmente

$S =$ pendiente de la curva de calibrado, calculada previamente

Usando un programa de ajuste de rectas por mínimos cuadrados y, en base a las consideraciones anteriores, los valores encontrados para los coeficientes de selectividad son los que aparecen en la Tabla 3. No se observa ninguna interferencia apreciable ($K_{ij}^{\text{POT}} \ll 1$).

TABLA 3

Ion interferente	K_{ij}^{POT}
SiO_3^{2-}	$2,3 \cdot 10^{-3}$

Cl^-	$8,0 \cdot 10^{-3}$
$\text{H}_2\text{PO}_4^{2-}$	$6,4 \cdot 10^{-3}$
SO_4^{2-}	$2,7 \cdot 10^{-3}$
Ac^-	$4,5 \cdot 10^{-3}$

Tabla 3: Coeficientes de selectividad para el estudio de interferencias.

4. Conclusiones.

Se ha puesto a punto un procedimiento para determinar fluoruro en muestras procedentes de baños de galvanotecnia, basado en el uso de un electrodo convencional de fluoruro. Las determinaciones, realizadas por interpolación en la curva de calibrado y por adición estandar simple, conducen a valores comparables a los obtenidos con el método de referencia. No se observa ninguna interferencia entre los posibles iones presentes en las muestras que se han estudiado. El método del electrodo selectivo de fluoruro es rápido, de bajo coste y elevada precisión, pudiéndose utilizar incluso en muestras coloreadas, para las que la valoración con nitrato de torio resulta imposible.

5. Bibliografía.

- (1) Van Loon, J.C., 1968, Anal. Letters, 1(6), 393.
- (2) Shane, N., Miele, D., 1968, J. Pharm. Sci., 57(7), 1260.
- (3) Mc Cann, H.G., 1968, Arch. Oral. Biol., 13, 475.
- (4) Tusl, J., 1970, J. Clin. Chim. Acta, 27, 216.
- (5) Louw, C.W., Richards, J.F., 1972, Analyst, 97(5), 334.
- (6) Chang, F., Tsai, U., Wu, S., 1974, Anal. Chim. Acta, 71, 477.
- (7) Ferraro, T.A., 1975, Army. Mat. and Mech. Res. Center, PTR 75-1.
- (8) Singer L., Ophawg R.H., 1977, Anal. Chem., 49(1), 38.
- (9) Gallego, R., Bernal, J.L., Pardo, R., 1978, Afinidad, 4, 333.
- (10) Efstathiou, C.E., Hadjiioannou, T.P., 1979, Anal. Chim. Acta, 109, 319.

- (11) Instruction manual fluoride electrodes, Orion Research, 1983.
- (12) IUPAC, 1976, Pure and Appl. Chem., 48, 127.
- (13) Guilbault, G.C., 1981, Pure and Appl. Chem., 53, 1908.
- (14) Burriel, F., Lucena, F., Arribas, S., Hernández, J., 1991, Química Analítica Cualitativa, Editorial Paraninfo.
- (15) Koryta, J., Stulik, S., 1983, Ion - Selective Electrodes, 2ª Edición, Cambridge University Press.
- (16) Macca, C., Cakrt, M., 1983, Anal. Chim. Acta, 154, 54.

Abstract

Paleochannels formed by dolomite of replacement can be found in the Quinsá area (Province of Zaragoza), located on the eastern area of the Central Sector of the Ebro Basin. The homogeneous size of the crystals gives to the rock the appearance of a "sedimentologic sand". The presence in these rocks of gypsum crystals containing dolomite crystals with petrographic features, typical of diagenesis, points out that the dolomitization process is not epidiagenetic.

The study by X-ray diffraction techniques of the limestones within which the paleochannels are found proves the existence of anhydrite, interbedded, illite-smectite and dolomite that can be considered as the sources of magnesium for the dolomitization process.

1. Introducción

1.1. Situación geográfica y geológica

La zona de estudio se encuentra situada al SE de la Provincia de Zaragoza. Corresponde a la totalidad del término municipal de Quinsá. El resto de Molinos y parte de los términos de Ciesra y Villa de Ebro (Figura 1).

Desde el punto de vista geológico se sitúa en el extremo oriental del Sector Central de la Cuenca del Ebro. Los materiales presentes pertenecen a las UTS 24 (Oligoceno-Mioceno) y N1 (Mioceno inferior) (Pérez et al., 1988) y son fundamentalmente lutitas, calizas y yesos, con intercalaciones de arenas, dispuestos en estratos horizontales o ligeramente basculados.

La descripción de 38 columnas litológicas permitió establecer siete Unidades Litológicas, fundamentalmente lutíticas, que sirvieron para la realización de una cartografía a escala 1:25.000 (Artieda, 1993). En tres de estas Unidades aparecen

Las "areniscas" dolomíticas del área de Quinto (Zaragoza). Modelo petrogenético.

O. Artieda ⁽¹⁾, J. Mandado ⁽²⁾ y J. Herrero ⁽¹⁾

(1) Unidad de Suelos y Riegos, Servicio de Investigación Agraria (DGA);
Laboratorio Asociado de Agronomía y Medio Ambiente (CSIC-DGA).
Apdo. 727, 50080-Zaragoza (España).

(2) Área de Petrología y Geoquímica. Departamento de Geología. Facultad de
Ciencias. Univ. Zaragoza. 50009. Zaragoza. (España).

Abstract

Paleochannels formed by dolomite of replacement can be found in the Quinto area (Province of Zaragoza), located on the eastern area of the Central Sector of the Ebro Basin. The homogeneous size of the crystals gives to the rock the appearance of a "sedimentologic sand". The presence in these rocks of gypsum crystals containing dolomite crystals with petrographic features, typical of diagenesis, points out that the dolomitization process is not epidiagenetic.

The study by X-ray diffraction techniques of the lutites within which the paleochannels are found proves the existence of smectites, interlayered illite-smectite and dolomite that can be considered as the sources of magnesium for the dolomitization process.

1. Introducción

1.1. Situación geográfica y geológica

La zona de estudio se encuentra situada al SE de la Provincia de Zaragoza. Corresponde a la totalidad del término municipal de Quinto, El Real de Belchite y parte de los términos de Gelsa y Velilla de Ebro (Figura 1).

Desde el punto de vista geológico se sitúa en el extremo oriental del Sector Central de la Cuenca del Ebro. Los materiales presentes pertenecen a las UTS T4 (Oligoceno-Mioceno) y N1 (Mioceno inferior) (Pérez *et al.*, 1988) y son fundamentalmente lutitas, calizas y yesos, con intercalaciones de arenitas, dispuestos en estratos horizontales o ligeramente basculados.

La descripción de 38 columnas litológicas permitió establecer siete Unidades Litológicas, fundamentalmente lutíticas, que sirvieron para la realización de una cartografía a escala 1:25.000 (Artieda, 1993). En tres de estas Unidades aparecen

paleocanales constituidos por rocas detríticas reemplazadas por dolomita llegando a constituir actualmente una "arenisca" dolomítica. También aparecen materiales limo-arenosos, cuyas características petrográficas indican su origen detrítico y arcillitas de modo generalizado en todas las Unidades.

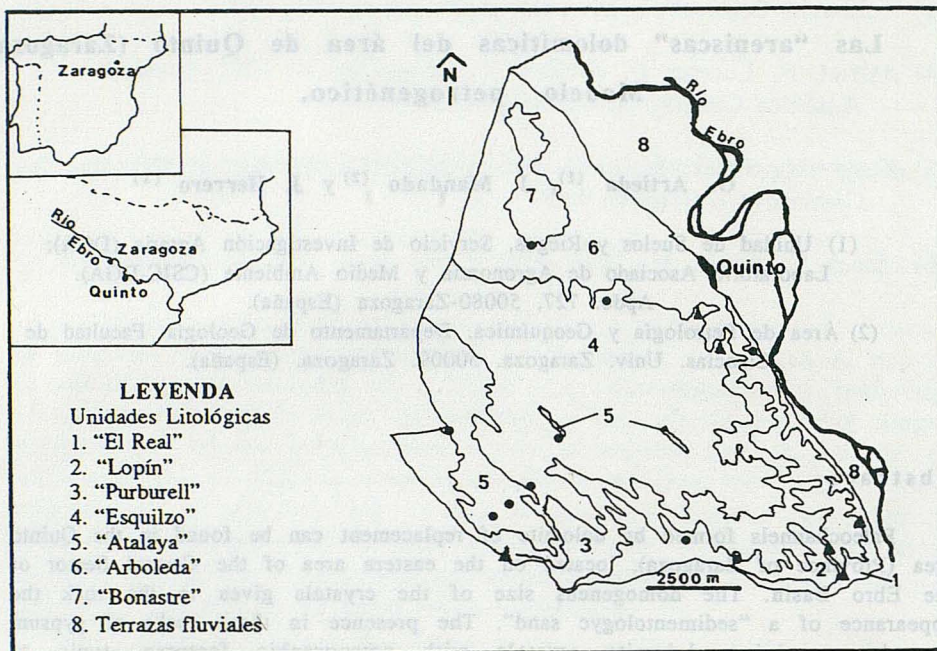


Figura 1. Mapa geológico de la zona estudiada. Las Unidades Litológicas corresponden a las definidas por Artieda (1993). En el mapa se sitúan los puntos de muestreo de: ▲ arenitas y • lutitas.

Parte de los materiales estudiados en este trabajo corresponden a las "Calizas de Sástago", definidas por Quirantes (1978). En esta formación el autor menciona la existencia de areniscas y arenitas yesíferas, definiendo las arenitas yesíferas como areniscas de grano fino a muy fino, con yeso y carbonatos como cementos y con clastos de cuarzo, calcita, yeso y otros accesorios. Entre los granos cita la existencia de hasta un 60 % de "dolomita secundaria", resultando confusa su consideración como granos si se entiende por dolomita secundaria aquella que procede de un proceso de dolomitización de calcita.

Algunos trabajos realizados en zonas próximas a la de estudio (Riba *et al.*, 1967; Quirantes y Martínez, 1967-68 y Birnbaum, 1976) mencionan la existencia de rocas detríticas, dispuestas en paleocanales o capas horizontales, constituidas por un armazón de granos de composición variada (cuarzo, feldespato, calcita, dolomita) y cementadas por calcita o yeso.

En cuanto a las lutitas, la zona de estudio se incluiría en la serie Sur (S-1) de Pinilla y Alonso (1969), caracterizada por abundancia de mica e illita, siendo escasa la haloisita. La característica de esta serie es el aumento de sepiolita, que

indicaría, en general, que se trata de medios distales dentro del dispositivo sedimentario. En este sentido, Fernández-Nieto *et al.* (1989), en el perfil de Vizcuerno, próximo espacialmente a la zona de estudio y cuyos materiales pertenecen a las UTS T4 y T5 (Oligoceno superior-Mioceno inferior), indican la presencia de esmectitas. En cuanto al origen de los minerales arcillosos, Mandado (1987) mantiene que en el Sector Aragonés de la Cuenca del Ebro, las facies arcillosas son esencialmente de carácter detrítico en los bordes y detríticas y neoformadas en las zonas centrales.

2. METODOLOGÍA

Para el estudio petrográfico de estos materiales se eligieron 8 muestras, correspondientes a niveles de arenitas, y 5 a niveles lutíticos aparecidos en las columnas descritas. En algunos casos fue necesaria la impregnación al vacío de la muestra con resina. De cada una de ellas se confeccionó una lámina delgada. Las secciones delgadas se realizaron utilizando aceite como refrigerante en el proceso de rectificación, de modo que se evitaron los artefactos de disolución. La temperatura se mantuvo durante todo el proceso por debajo de los 40°C para evitar posibles deshidrataciones. El tamaño de las láminas fue variable: 45 x 58 y 58 x 135 mm.

Para la diferenciación de los carbonatos se utilizó el método de tinción de Lindholm y Fikelman (1971) con ligeras modificaciones en los tiempos de ataque.

El estudio mineralógico se efectuó mediante difracción de R.X., por el método del polvo en agregados no orientados, y se eligieron 12 muestras correspondientes a lutitas. Estos análisis se realizaron en el Instituto "Jaume Almera", del CSIC, en Barcelona, con un difractómetro SIEMENS D-500. La radiación utilizada fue la $K\alpha_1$ del cobre, con $\lambda_1 = 1.5405 \text{ \AA}$. Las condiciones de excitación fueron 40 KV Y 20 mA y la constante de tiempo 2 segundos.

3. RESULTADOS

3.1. Arenitas

3.1.1. Características macromorfológicas.

En la zona estudiada aparecen cuerpos rocosos de geometría canaliforme, más o menos laxa, con estructuras sedimentarias (laminación cruzada, paralela, etc). Estos niveles están intercalados entre lutitas y pueden ser considerados como depósitos detríticos. El aspecto de campo es el de una arenita calcárea de grano fino-medio, con tonalidades de ocre a grises, compacta en corte fresco, pero deleznable en superficies expuestas. El desarrollo longitudinal es hectométrico y la potencia llega a superar el metro (Figura 2).

En algunos afloramientos se observan yesos nodulares incluidos en las "arenitas". Se trata de nódulos alargados en la vertical y con constricciones que les dan un aspecto característico. Se han denominado "turrítelados" al recordar la sección de estos fósiles (turrítelas). El diámetro de los nódulos no supera los 3 cm y su longitud está en torno a los 10 cm. Se trata de yeso nodular, de textura sacaroidea y de color blanco (Figura 2).

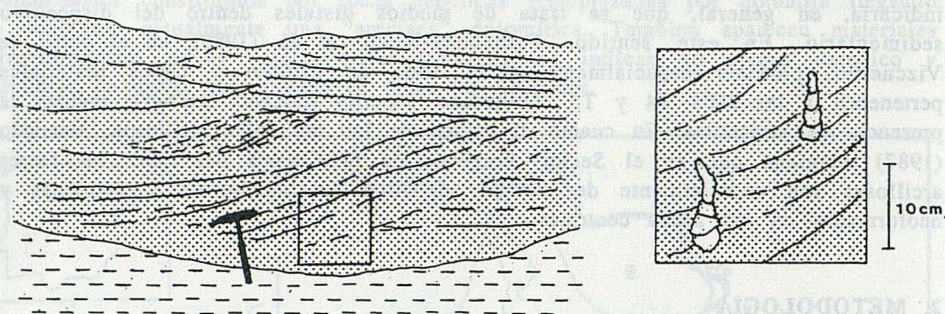


Figura 2. Aspecto de campo que presentan los paleocanales y yesos turrilados incluidos.

3.1.2. Características petrográficas.

El estudio petrográfico no evidencia estructuras que puedan considerarse sedimentarias, aun cuando en campo parecían insinuarse. La mineralogía de estos materiales queda resumida en el Cuadro 1. Están constituidos fundamentalmente por dolomita de reemplazamiento. Se trata de un mosaico idiomórfico constituido por romboedros de dolomita no férrica (como demuestra la tinción) junto a granos de cuarzo, feldespato, turmalina, micas blancas, opacos y fragmentos de caliza. Los romboedros de dolomita aparecen generalmente con el área interna más oscura, apareciendo en algunos casos cristales zonados, con dolomita férrica en el centro y dolomita (no férrica) en el exterior. El tamaño de estos cristales es variable de una muestra a otra, apareciendo desde cristales cuyo eje mayor es de 300 μm , a otros con menos de 30 μm .

La muestra PU-I/54 presentaba en campo el aspecto de una marga, estando constituida petrográficamente por romboedros de dolomita cuyo tamaño cristalino corresponde al limo.

Los granos de cuarzo no superan el 20 %. Se trata de granos de formas subangulosas, tabulares o esferoidales, con diámetro medio entre 150 y 200 μm .

Los fragmentos de rocas calcáreas llegan a representar hasta el 20 % (Muestra VA-I/1 y LO-IX/1)). Se trata de granos de geometría irregular, subredondeados y de tamaño variable entre 100 y 300 μm .

En cuanto a las micas, las más frecuentes, aunque en porcentajes inferiores al 2 %, son las blancas.

Los feldespatos, turmalina y opacos, son por lo general minoritarios.

La matriz resulta difícil de determinar, especialmente si se tiene en cuenta el grado de transformación que han sufrido estos materiales.

En cuanto a los nódulos de yeso incluidos, puede constatar que se trata de yeso alabastrino, constituido por cristales con extinción no uniforme (c.e.n.u.) con

impurezas de anhidrita. Estos c.e.n.u. pueden englobar romboedros de dolomita. En otros casos, el yeso aparece como cemento, interponiéndose en los espacios intergranulares, no siendo rara la existencia de yeso megacristalino.

Como resumen se puede decir que los materiales estudiados no corresponden a "arenas sedimentológicas", sino a rocas en las cuales el mayor porcentaje corresponde a dolomita de reemplazamiento, con ausencia de matriz, y una textura de apariencia arenosa o limosa favorecida por el tamaño homogéneo de los cristales de dolomita

Cuadro 1. Características macromorfológicas y petrográficas de las "arenitas". Q: cuarzo; Fc: fragmentos de rocas calcáreas; M: micas blancas; Ot: otros (turmalinas, feldespatos y opacos).

Lámina	Estructuras macromorfológicas	Mineralogía					
		Granos (%)				Cemento	
		Q	Fc	M	Ot	%	Naturaleza
ER-IV/2	Laminación cruzada en surco	10	-	-	<1	40	Yeso
BA-I/3	Masivas	10	-	2	-	10	Yeso
ER-II/A-3	Masiva	15	10	2	<1	<1	Cuarzo
LO-IX/1	Laminación paralela. Yeso turritelado	20	20	2	1	15	Yeso
LO-IX/2	Laminación paralela	5	4	1	-	-	-
PU-I/54	Laminación paralela	5	10	<1	-	-	-
PU-I/55	Masivas Yeso turritelados	5	<1	<1	-	4	Cuarzo-yeso
VA-I/1	Laminación paralela	18	20	1	-	-	-

3. 2. Lutitas

3.2.1. Descripción macromorfológica.

Dentro de estas rocas se incluirán tanto limolitas como arcillitas. Aparecen formando paquetes de espesor variable, de centimétrico a métrico, intercalados entre yesos y calizas.

Los colores varían entre rojos, violetas, verdosos, etc. Las coloraciones verdosas, grises o pardas suelen aparecer en la base de niveles calizos o de yesos, aunque sea constituyendo niveles centimétricos. Suelen aflorar en niveles masivos o como niveles laminados, distinguiéndose ésta por variaciones de color. En los términos más finos, arcillitas, es frecuente la presencia de fractura

concoidea y disyunción bolar, que llega a individualizar nódulos centimétricos, de caras curvas y en cuya superficie pueden observarse débiles estrías radiales. Las limolitas se caracterizan por presentar laminación paralela irregular.

3.2.2 Caracterización mineralógica. R.X.

Con el fin de poder conocer cualitativamente la composición mineralógica de las lutitas se presentan en el Cuadro 2 los minerales identificados mediante el estudio de R.X.

Cuadro 2. Resultados obtenidos en el análisis de las lutitas por difracción de RX por el método del polvo en agregados no orientados. Sm: esmectitas; I/Sm: interestratificados illita-esmectita; I: illita; Cl: clorita; Q: cuarzo; Ct: calcita; Dl: dolomita; Ys: yeso.

Muestra	Mineralogía							
	M.A.				Q	Ct.	Dl	Ys
	Sm	I/Sm	I	Cl				
LO-II/C3			SI	SI	SI		SI	
LO-III/2			SI	SI	SI	SI	SI	
PU-I/2	SI		SI	SI	SI		SI	
PU-I/10	SI		SI	SI	SI		SI	
ES-Z/1-3			SI	SI	SI	SI		SI
VA-II/1			SI	SI	SI	SI	SI	SI
ES/1			SI	SI	SI	SI	SI	SI
CA-I/1	SI		SI	SI	SI	SI	SI	SI
CA-III/1	SI		SI	SI	SI	SI	SI	SI
AT/1		SI	SI	SI	SI	SI	SI	
AT-3		SI	SI	SI	SI		SI	SI
TO-I/7	SI		SI	SI	SI	SI	SI	SI

Del Cuadro 2 se deduce que en 7 muestras (58 %) aparecen esmectitas o interestratificados illita-esmectita, apareciendo illita y clorita como minerales omnipresentes. La dolomita está presente en la práctica totalidad de las muestras (90 %).

3.2.3. Descripción petrográfica.

La mayor parte de los trabajos dedicados al estudio de lutitas hacen referencia a la composición mineralógica, a través de estudios por difracción de RX,

composición química o aspectos estructurales mediante el uso del microscopio electrónico; sin embargo, las referencias bibliográficas encontradas dedicadas a la descripción petrográfica de las lutitas son escasas. Zimmerle (1991) utiliza esta técnica, en muestras impregnadas con resina, como complemento en el estudio sedimentológico, indicando su utilidad en la observación de estructuras sedimentarias, evidenciadas por variaciones de la pigmentación y composición mineral.

En las muestras estudiadas la laminación paralela, horizontal o convolucionada, es un rasgo prácticamente común, evidenciándose por variaciones en el tamaño de grano o por variaciones del color. Los procesos de bioturbación se evidencian por la existencia de galerías rellenas que rompen la laminación. Es frecuente encontrar en las muestras de tamaño de grano menor, arcillitas, dominios de birrefringencia, así como zonas con "b-fábrica estriada paralela total" (Bullock *et al.*, 1985).

En el Cuadro 3 aparecen resumidos los rasgos petrográficos de cada una de las muestras estudiadas.

Cuadro 3. Composición mineralógica de las lutitas, a partir del estudio petrográfico. M.A.: minerales de la arcilla; Q.: cuarzo; F.C.: fragmentos de rocas calcáreas; M.: micas; T.: turmalinas; O.: Opacos; Ca.: calcita; Dol.: dolomita; Ys.: yeso.

Muestra	Clasificación visu	Mineralogía (%)								
		M.A	Q.	F.C	M.	T.	O.	Ca.	Dol	Ys.
AT-3	Limolita	30	35	3	10	-	15	1	3	3
CE-I/2	Limolita	10	50	25	5	-	5	1	1	3
TO-I/1	Limolita	35	30	-	20	-	10	3	2	-
CA-I/1	Arcillita	95	5	<1	-	-	-	-	-	-
CA-III/1	Arcillita	75	30	<1	5	-	10	-	-	-

Los granos detríticos de cuarzo llegan a representar el 50% de la roca con tamaños por debajo de los 200 μm , y de formas angulosas. En algunas muestras el cuarzo aparece formando el cemento, como mineral autigénico en forma de lutecita, de hasta 150 μm de diámetro, o como cristales euhedrales a anhedrales, de 80 x 500 μm de tamaño máximo (muestra AT-3).

Los fragmentos de rocas calcáreas, de formas generalmente redondeadas y de hasta 400 μm de diámetro no superan el 25 % de la roca (Muestra CE-I/2)

En las limolitas aparecen nódulos irregulares de diámetros inferiores a los 200 μm , constituidos por carbonatos y materia orgánica y dispuestos en bandas.

Las micas son muy frecuentes en todas las muestras, apareciendo tanto blancas como verdes. Su tamaño es variable pero es raro que superen los 150 μm , presentándose generalmente orientadas de modo paralelo a la laminación.

Los opacos llegan a ser muy abundantes en algunas muestras (hasta 15%), con formas redondeadas y diámetros no superiores a los 100 μm ; mientras que las turmalinas son minoritarias.

El yeso, aunque en proporciones inferiores al 5%, es frecuente en las muestras de mayor tamaño de grano, limolitas. Se trata de yeso con texturas claramente diagenéticas, fundamentalmente c.e.n.u. de hasta 7 mm de diámetro.

Los cristales anhedrales a subhedrales de calcita y dolomita, de diámetros inferiores a los 250 μm , aparecen igualmente en las limolitas, constituyendo nódulos de hasta 1 mm de diámetro, si bien en proporciones que no superan el 3%.

4. DISCUSIÓN Y CONCLUSIONES

El origen detrítico del sedimento original que da lugar a las "areniscas" viene confirmado por las características estructurales, morfología de canal y estructuras sedimentarias, y no por los aspectos texturales, al tratarse de una roca con importantes procesos diagenéticos modificadores de la textura primaria. La existencia de una textura "arenosa" se debe, en este caso, a la presencia de importantes cantidades de dolomita que ha reemplazado a unos hipotéticos granos carbonatados previos.

El sedimento original podría ser una calcilita o litoarenita calcárea (Folk, 1968) que diagenéticamente ha visto modificada su textura y composición original por un proceso de dolomitización.

El medio sedimentario es el de una cuenca de playa-lago, en su zona de llanura de barros, y los depósitos canaliformes de arenitas corresponderían a los materiales depositados por flujos jerarquizados, esporádicos, correspondientes a periodos de tormentas con avenidas más importantes. Este esquema paleogeográfico se confirma tanto por la disposición de facies sedimentológicas, asociación con lutitas, presencia de yesos nodulares dispersos, indicadores de la proximidad al nivel de base, como por la mineralogía de las arcillas, con evidencias de una importante proporción de minerales neoformados, característicos de la zona de transición entre las facies detríticas finas y las evaporíticas (Huertas *et al.*, 1974).

El proceso de dolomitización se produjo en condiciones de cuasiequilibrio, como lo evidencia la uniformidad granulométrica de los cristales, su morfología idiomórfica romboédrica y estructura zonada. Estas características permiten eliminar como momento de la dolomitización la etapa epidiagenética, que generaría mosaicos más irregulares y una marcada inhomogeneidad del frente de dolomitización; pero además, la evidencia petrográfica, con romboedros de dolomita incluidos en los c.e.n.u., confirma que el proceso de dolomitización es, cuanto menos, anterior al de hidratación epidiagenética de anhidrita a yeso secundario.

El mecanismo concreto de dolomitización puede ser tanto sinsedimentario, como diagenético temprano.

En un proceso sinsedimentario, la fuente de magnesio serían las aguas freáticas circulantes, procedentes del área fuente, ricas en magnesio y con un

posible enriquecimiento de la relación Mg/Ca como consecuencia de la precipitación de sulfato cálcico (yeso o anhidrita).

En condiciones de mayor enterramiento, durante las primeras etapas epidiagenéticas, las aguas expulsadas por la compactación de los materiales subyacentes pueden aportar importantes cantidades de magnesio procedente de los minerales magnésicos existentes en las lutitas, entre las que se intercalan los paleocanales. En este sentido cabe recordar que el 90 % de las muestras estudiadas en el Apartado 3.4, presentaban dolomita y más del 50 % de ellas contenían esmectitas o interestratificados illita-esmectita, que han podido ser la fuente de magnesio necesaria para la dolomitización de los sedimentos originales de los paleocanales, cuya mayor permeabilidad los convierte en áreas de circulación preferencial.

AGRADECIMIENTOS

Durante la realización de este trabajo O. Artieda ha disfrutado de una beca de investigación del Consejo Asesor de Investigación de la Diputación General de Aragón (CONAI BCA-3391) Así mismo el trabajo ha sido financiado por los proyectos INIA 8525 y 93/056.

BIBLIOGRAFÍA

Artieda, O. 1993. *Factores geológicos que inciden en el desarrollo de los suelos en un medio semiárido. El caso de Quinto (Zaragoza)*. Tesis de Licenciatura. Univ. Zaragoza. 305 pp. + Apéndices y mapas. Inédita.

Birnbaum, S.J. 1976. *Non-marine evaporite and carbonate deposition, Ebro Basin*. Thesis, Univ. Cambridge. 148 pp. Inédita

Bullock, P.; Fedoroff, N.; Jongerius, A.; Stoops, G. & Tursina, T. 1985. *Handbook for the description of thin sections of soils*. Waine Res. Publ. Wolverhampton, u.k. 150 pp.

Fernández-Nieto, C.; González, A.; González López, J.M.; González Martínez, J. y López-Aguayo, F. 1989. El análisis mineralógico aplicado a la evolución paleogeográfica del Terciario en el sector SE de la Depresión del Ebro. *Bol. Soc. Esp. Min.*, 12: 221-228.

Folk, R. 1968. *Petrology of sedimentary rocks*. Ed. Hemphill's, Austin. 170 pp.

Huertas, F.; Linares, J. y Martín-Vivaldi, J.L. 1974. Minerales fibrosos de la arcilla en cuencas sedimentarias españolas. II. Cuencas del Guadalquivir, Ebro y Depresión de Granada. III. Consideraciones genéticas. *Est. geol.*, 30: 359-366.

Lindholm, R.C. & Fikelman, R.B. 1971. Calcite staining: semiquantitative determination of ferrous iron. *Jour. Sed. Petrol.*, 41: 239-242.

Mandado, J. 1987. *Litofacies yesíferas del Sector Aragónés de la Cuenca Terciaria del Ebro*. Petrogénesis y Geoquímica. Tesis Doct. Univ. Zaragoza. 443 pp. + lam.

Pérez, A.; Muñoz, A. y Sanchez, J.A. 1988. *Sistemas lacustres neógenos del Margen Ibérico de la Cuenca del Ebro*. Guía de campo de la III Reunión del G.E.T. Zaragoza. 104 pp.

Pinilla, A. y Alonso, J.J. 1969. Estudio sedimentológico de la zona aragonesa de la Cuenca Terciaria del Valle del Ebro. V Análisis de las arcillas con microscopio electrónico. *An. Edafol. Agrob. T. XXVIII*. Nº 11-12: 895-906.

Quirantes, J. 1978. *Estudio sedimentológico y estratigráfico del Terciario continental de los Monegros*. Publicación Nº 178 de la Institución Fernando el Católico.(C.S.I.C) D.P.Z. 207 pp.

Quirantes, J. y Martínez, E. 1967-68. Un tipo de roca poco conocido: las arenitas yesíferas de los Monegros. *Brev. Geol. Astúrica*. Vol XI: 113-124.

Zimmerle, W. 1991. Thin-section petrography of pelites, a promising approach in sedimentology. *Geologie en Mijnbouw*, 70: 163-174.

Cálculos de simulación de la alteración de microclina.

I. Análisis comparativo de códigos de modelización geoquímica.

Auqué, L.F.⁽¹⁾; Mandado, J.⁽¹⁾; Gimeno, M.J.⁽²⁾; López, P.L.⁽¹⁾ y Vallés, V.⁽³⁾

(1) Área de Petrología y Geoquímica. Depto. Ciencias de la Tierra. Fac. Ciencias.
Universidad de Zaragoza. 50009 ZARAGOZA

(2) Unidad de Seguridad de Emplazamientos. Edificio 20 T. CIEMAT.
Avda. Complutense 22. 28006 MADRID.

(3) Laboratoire du Science du Sol d'Avignon, Institut National de la Recherche Agronomique.
Domaine St. Paul, B.P. 91, 84143 Montfavet CEDEX (France).

Abstract

Simulations of microcline weathering reactions (in closed system and reacting with a HCl solution at 25 °C and pH=4) are formulated with the more recent versions of the computer codes EQ3/6, CHILLER and PHREEQE. The calculated reaction paths are strongly dependent of the thermodynamic constants for mineral equilibria but almost insensitive to the particular chemical model and speciation reactions enclosed in these codes for the conditions of simulation.

Changes in the thermodynamic values for the involved mineral equilibria lead to important numerical differences in the predicted mass transfers, solution compositional parameters and also affect the number of mineral fields crossed by the solution on a log-log activity diagram. So, differences in the thermodynamic data can induce variations in reaction path calculations as important as those produced by changes in the initial conditions of simulation.

1. Introducción

Desde que Helgeson (1968) y Helgeson *et al.* (1969) plantearon y aplicaron el primer código de pautas de reacción (PATH) a la simulación de un proceso de hidrólisis de feldespato potásico, este ejemplo ha pasado a constituir un problema-tipo en la testificación y presentación del funcionamiento de los más modernos códigos de este tipo. Y como tal es frecuente que aparezca en sus propios manuales de usuario (p. ej. Wolery & Daveler, 1992; Parkhurst *et al.*, 1980, 1990).

Pero a la vez, la modelización de este proceso ha pasado a utilizarse como esquema básico sobre el que analizar, metodológica o conceptualmente, tanto la incidencia de las condiciones iniciales de simulación (composición inicial de la solución interactuante; Steinmann *et al.*, 1994), como del número y tipo de los productos minerales de alteración considerados (Wolery, 1979; o indirectamente, Drever, 1988). E, incluso, ha sido planteada desde supuestos distintos a los utilizados en los cálculos de pautas de reacción (evolución a través de sucesivos

estado de equilibrio parcial) usando formulaciones cinéticas (Helgeson & Murphy, 1983), de acoplamiento flujo-reacción, o conjugando ambas (Lichtner, 1988; Steefel & Van Capellen, 1990).

La diversidad de planteamientos con los que ha sido abordada la simulación de este proceso ha permitido una muy amplia discusión de los distintos factores que controlan los resultados obtenidos. Sin embargo, y volviendo a esquemas puramente termodinámicos (pautas de reacción), son bastante escasos los estudios que han incidido en la influencia de los propios planteamientos del código y de los datos termodinámicos utilizados sobre el resultado de los cálculos. Únicamente INTERA (1983) utiliza este caso-tipo para comparar los resultados suministrados por las versiones entonces existentes de los códigos EQ3/6 y PHREEQE.

En este trabajo se plantean dos objetivos estrechamente relacionados: por un lado, un estudio comparativo de las últimas versiones de los códigos de pautas de reacción actualmente disponibles; y por otro un análisis sobre la incidencia de los conjuntos de datos termodinámicos elegidos para las reacciones de equilibrio mineral en los resultados suministrados por estos códigos, utilizando como ejemplo de referencia una simulación de alteración de microclina en ambos casos.

2. Códigos seleccionados y metodología de simulación.

Los códigos seleccionados para el análisis comparativo han sido el EQ3/6 (versión 7.2a; Wolery & Daveler, 1992), CHILLER (Reed, 1982; Spycher & Reed, 1989, 1992) y PHREEQE (revisión de 1990; Parkhurst *et al.*, 1980, 1990). Se trata de códigos de pautas de reacción pertenecientes al grupo de resolución mediante la constante de equilibrio y basados todos ellos en modelos de Asociación Iónica (distintas variantes de la ecuación de Debye-Hückel: tipo "B-dot" para EQ3/6 y CHILLER, y tipo WATEQ y ecuación de Debye-Hückel ampliada en PHREEQE; Tena *et al.*, 1990).

Todos ellos pueden plantear la simulación de un proceso irreversible de alteración de microclina a través de sucesivos estados de equilibrio parcial, aunque desde distintas filosofías de cálculo. La diferencia fundamental se refiere a la presencia (EQ3/6 y CHILLER) o ausencia (PHREEQE) de un algoritmo de definición de las fases minerales estables en cada punto discreto del avance de la reacción (Tena *et al.*, 1990). Circunstancia que se traduce en una mayor necesidad de datos de entrada al plantear la simulación con el PHREEQE ya que, por sí mismo, es incapaz de determinar las fases minerales estables (que precipitan) a medida que avanza el proceso irreversible.

Asimismo, los supuestos y planteamientos de las ecuaciones de gobierno en cada código son también diferentes, especialmente en el caso de las relacionadas con el control de hidrógeno, oxígeno y/o agua. Entre los tres códigos abarcan la casi totalidad de posibles planteamientos actualmente existentes. Una descripción más detallada de las diferencias existentes en las ecuaciones de gobierno de estos códigos puede verse en Reed (1982) o Tena *et al.* (1990).

La base de datos termodinámicos utilizada por el EQ3/6 en el sistema composicional involucrado procede casi totalmente del SUPCRT 92 (Johnson *et al.*, 1991); la del CHILLER utiliza la base de datos SOLTHERM (Reed & Spycher, 1987, 1989) en gran parte derivada de la versión de SUPCRT 84 (Helgeson, 1984) y la del PHREEQE procede fundamentalmente de la revisión de Nordstrom *et al.* (1990).

La simulación del proceso de alteración de microclina se realizará inicialmente en condiciones isoterma a 25 °C, considerando una solución de HCl diluido (pH=4) y reduciendo el conjunto de fases minerales susceptibles de precipitar durante el proceso a gibbsita, caolinita y moscovita (los valores de las constantes de equilibrio utilizadas por los distintos códigos para las reacciones de disolución de estos minerales puede verse en la tabla 1).

El planteamiento de la simulación se realizará en condiciones de sistema cerrado; es decir, manteniendo dentro del sistema los productos minerales precipitados durante la disolución irreversible de la microclina (y que, por tanto, podrán disolverse posteriormente conforme cambian las condiciones del sistema; p. ej. Wolery, 1979). La variable de progreso de reacción queda definida por el número de moles disueltos de microclina y el proceso se detendrá cuando la fase acuosa alcance el equilibrio con ese mineral. Estas condiciones de simulación son las mismas que aparecen en los manuales de EQ3/6 (Wolery & Daveler, 1992) y PHREEQE (Parkhurst *et al.*, 1980, 1990), así como en el análisis comparativo de anteriores versiones de EQ3/6 y PHREEQE realizado por INTERA (1983).

Operativamente, la simulación puede contemplarse como el cálculo de los puntos de intersección de la solución con los límites de las fases minerales consideradas dentro de un diagrama de campos de estabilidad. Y desde esta perspectiva, la comparación de las pautas de reacción obtenidas será realizada sobre un diagrama de estabilidad mineral $\log aK^+/aH^+ - \log aSiO_2$, construido a partir de los datos termodinámicos considerados por cada código para las fases minerales involucradas (microclina, gibbsita, moscovita y caolinita). Sobre este diagrama se proyectarán las características composicionales de la solución correspondientes a las situaciones de equilibrio (puntos de intersección con los límites de fase) calculadas en cada caso. Los datos numéricos presentados en las tablas de resultados corresponderán a estos puntos de corte de las pautas calculadas.

Finalmente, los tres códigos citados necesitan la especificación de una solución inicial que satisfaga las condiciones definidas anteriormente. Para las características de pH = 4 (HCl diluido) se considerará agua pura, fijando un pH = 4 y añadiendo Cl⁻ hasta satisfacer totalmente la condición de neutralidad eléctrica de la solución. Esta operación supone la adición de 0.10116 mmoles de Cl⁻ en la versión examinada de EQ3/6 y de 0.1012 en las de CHILLER y PHREEQE. Esta coincidencia de valores constata el similar funcionamiento del algoritmo relacionado con la condición de electroneutralidad en los códigos.

Tabla 1: Valores de las constantes de equilibrio (a 25 °C) incluidas en las bases de datos de los distintos códigos para las reacciones de disolución de los minerales involucrados en la simulación.

	EQ3/6 (SUPCRT 92)	CHILLER (SOLTHERM)	PHREEQE (Versión 1990)	Robie <i>et al.</i> (1978)
Gibbsita	6.796	7.960	8.11 ⁽¹⁾	8.049
Caolinita	5.101	7.420	7.435	5.708
Moscovita	11.022	14.56	12.97	12.97
Microclina	-1.130	0.080	0.875	0.875

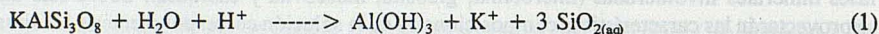
⁽¹⁾ Valor propuesto para la gibbsita cristalina por Nordstrom *et al.* (1990).

3. Comparación de los resultados obtenidos por PHREEQE, CHILLER y EQ3/6.

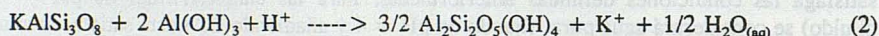
Los resultados obtenidos por los distintos códigos utilizados se resumen en las tablas 2 y 3, en las que se indican los caracteres más destacables del quimismo de la solución y de las transferencias de masa relacionadas con los límites de fase involucrados en el proceso.

En la tabla 2 se muestran los valores calculados por EQ3/6 y CHILLER para el proceso de alteración de microclina. Ambos códigos discretizan la pauta de reacción calculada en el mismo número de etapas, con caracteres cualitativamente similares pero con algunas diferencias cuantitativas en las proporciones de progreso de reacción, en las características composicionales de la solución y en los valores de transferencia de masa calculados. En la figura 1 A se ha representado la evolución obtenida por cada uno de estos códigos sobre el diagrama de estabilidad: los puntos indicados 1 a 5 corresponden a los límites de fase calculados por el EQ3/6; y los señalados de A a E son los obtenidos por el código CHILLER. Los campos de estabilidad definidos a partir de los datos termodinámicos considerados por estos códigos son prácticamente coincidentes pese a las diferencias en las constantes de equilibrio de las reacciones involucradas (tabla 1).

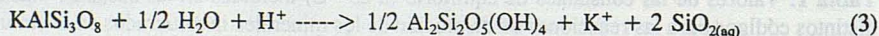
La pauta obtenida por EQ3/6 y CHILLER predice que el sistema alcanza la saturación respecto a la gibbsita (puntos 1 y A; figura 1A) tras disolver congruentemente $4.458 \cdot 10^{-3}$ y $13.9 \cdot 10^{-3}$ mmoles de microclina, respectivamente. A partir de este momento la disolución del feldespato es incongruente según la reacción:



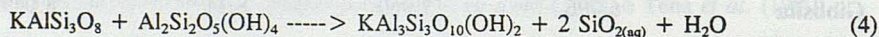
es decir, precipitando gibbsita hasta que el sistema alcanza la saturación con caolinita (punto 2 y B) después de haber disuelto casi $19 \cdot 10^{-3}$ mmoles de microclina y precipitado $15.84 \cdot 10^{-3}$ mmoles de gibbsita en el caso de EQ3/6 y $7 \cdot 10^{-5}$ mmoles en el caso de CHILLER. A partir de este punto, la disolución incongruente de microclina continúa según la reacción:



precipitando caolinita y disolviendo gibbsita hasta que, cuando se han disuelto $22.7 \cdot 10^{-3}$ mmoles o $27.19 \cdot 10^{-3}$ mmoles de microclina (según CHILLER y EQ3/6 respectivamente), se han consumido las proporciones de gibbsita precipitadas anteriormente y el sistema deja de estar saturado en este mineral a partir de los puntos 3 y C. En esos puntos las proporciones de caolinita precipitada calculadas son de $12.41 \cdot 10^{-3}$ mmoles (EQ3/6) y $5.46 \cdot 10^{-3}$ mmoles (CHILLER). A partir de ese momento la reacción pasa a ser incongruente respecto a la caolinita:



y después de $102.5 \cdot 10^{-3}$ mmoles de microclina disuelta el sistema llega al equilibrio con la moscovita, (punto 4 y D), con precipitación de casi $51.3 \cdot 10^{-3}$ mmoles de caolinita para ambos códigos. La reacción continúa según:



hasta que, cuando se han disuelto $152.66 \cdot 10^{-3}$ mmoles (EQ3/6) o $148.5 \cdot 10^{-3}$ mmoles (CHILLER) de feldespato, el sistema alcanza el equilibrio con la microclina (también lo está con caolinita y moscovita), precipitando $45 \cdot 10^{-3}$ - $47 \cdot 10^{-3}$ mmoles de moscovita y disolviendo gran parte de la caolinita anteriormente precipitada, deteniéndose el proceso en el punto 5 o E, cuando quedan sólo $5 \cdot 10^{-3}$ - $6 \cdot 10^{-3}$ mmoles de caolinita.

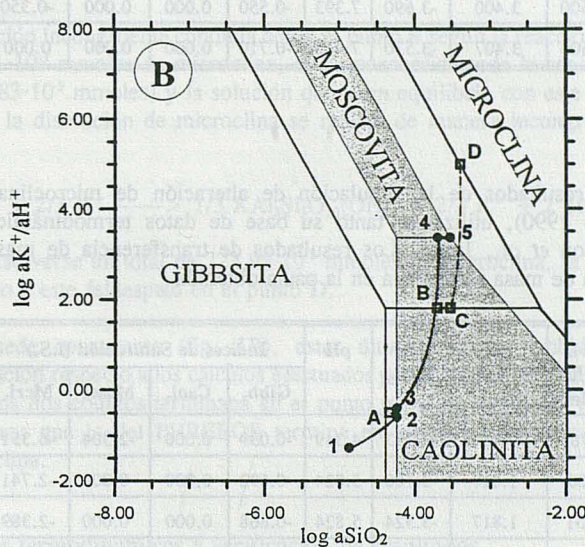
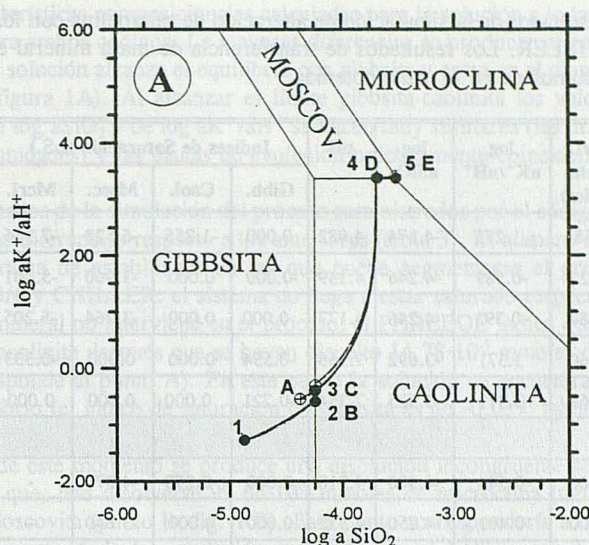


Figura 1.- A. Pautas de evolución calculadas por el código EQ3/6 (puntos 1-5) y CHILLER (puntos A-E) para la simulación de alteración de microclina sobre un diagrama de estabilidad mineral $\log aK^+/aH^+ - \log aSiO_2$. El espacio de fases obtenido con los datos termodinámicos de ambos códigos es prácticamente coincidente. B. Pauta de evolución calculada por el código PHREEQE (puntos A-D) sobre el diagrama de estabilidad obtenido a partir de los valores de su base de datos. A efectos comparativos se mantiene la pauta definida por el código EQ3/6 (puntos 1-5) y las áreas sombreadas corresponden a los campos de estabilidad de caolinita y moscovita definidos con los datos termodinámicos de este código.

Tabla 2: Algunos resultados de la simulación de alteración de microclina con los códigos EQ3/6 (versión 7.2a) y CHILLER. Los resultados de transferencia de masa mineral están expresados en forma de masa acumulada en la pauta obtenida.

EQ3/6 (vers.7.2A)	Micrl. disuelta (μ moles)	log aK ⁺ /aH ⁺	log aSiO ₂	pH	Indices de Saturación (I.S.)				μ moles precipitados		
					Gibb.	Caol.	Mosc.	Micrl.	Gib.	Caol.	Mosc.
1) Gibbsita	4.458	-1.275	-4.874	4.082	0.000	-1.255	-6.528	-7.966	0.00	0.00	0.00
2) Caol.- gib.	18.923	-0.589	-4.246	4.139	0.000	0.000	-3.960	-5.401	15.84	0.00	0.00
3) Caol.- gib.	27.187	-0.393	-4.246	4.177	0.000	0.000	-3.764	-5.205	0.00	12.41	0.00
4) Moscovita	102.60	3.371	-3.692	7.365	-0.554	0.000	0.000	-0.333	0.00	51.29	0.00
5) Microclina	152.66	3.371	-3.525	7.354	-0.721	0.000	0.000	0.000	0.00	5.25	47.37
CHILLER											
A) Gibbsita	13.900	-0.544	-4.380	4.318	0.000	-0.270	-4.370	-	0.06	0.00	0.00
B) Caol.-gib.	18.800	-0.401	-4.250	4.332	0.000	0.000	-3.840	-	0.07	0.00	0.00
C) Caol.-gib.	22.700	-0.307	-4.250	4.344	0.000	0.000	-3.730	-	0.00	5.46	0.00
D) Moscovita	102.500	3.400	-3.690	7.393	-0.550	0.000	0.000	-0.350	0.00	51.23	0.00
E) Microclina	148.500	3.407	-3.530	7.413	-0.710	0.000	0.000	0.000	0.00	6.04	45.46

Tabla 3: Algunos resultados de la simulación de alteración de microclina con el código PHREEQE (versión 1990), utilizando tanto su base de datos termodinámicos como la de SUPCRT 92 (Johnson *et al.*, 1991). Los resultados de transferencia de masa mineral están expresados en forma de masa acumulada en la pauta obtenida.

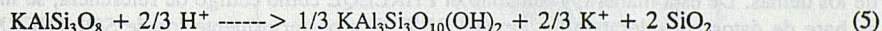
PHREEQE (1990)	Micrl. disuelta (μ moles)	log aK ⁺ /aH ⁺	log aSiO ₂	pH	Indices de Saturación (I.S.)				μ moles precipitados		
					Gibb.	Caol.	Mosc.	Micrl.	Gib.	Caol.	Mosc.
A) Caolinita	14.78	-0.487	-4.353	4.349	-0.039	0.000	-2.304	-6.351	0.00	0.00	0.00
B) Mos.- caol.	99.66	1.817	-3.700	5.824	-0.692	0.000	0.000	-2.741	0.00	49.83	0.00
C) Mos.-caol.	149.51	1.817	-3.524	5.824	-0.868	0.000	0.000	-2.389	0.00	0.00	49.83
D) Microclina	223.99	5.000	-3.391	8.831	-2.062	-2.121	0.000	0.000	0.00	0.00	74.66
PHREEQE (SUPCRT 92)											
* Gibbsita	4.455	-1.275	-4.874	4.082	0.000	-1.257	-6.530	-7.971	0.00	0.00	0.00
* Caol.- gib.	18.94	-0.589	-4.246	4.139	0.000	0.000	-3.957	-5.400	15.89	0.00	0.00
* Caol.- gib.	27.23	-0.392	-4.246	4.178	0.000	0.000	-3.761	-5.203	0.00	12.44	0.00
* Moscovita	102.05	3.369	-3.692	7.365	-0.554	0.000	0.000	-0.333	0.00	51.02	0.00
* Microclina	149.95	3.368	-3.525	7.363	-0.721	0.000	0.000	0.000	0.00	3.61	47.57

Las características composicionales calculadas para la solución a lo largo del proceso son muy similares para ambos códigos. Las mayores diferencias se producen al comienzo de la simulación, cuando la solución alcanza el equilibrio con gibbsita y entra en el diagrama de estabilidad (puntos 1 y A; figura 1A). Al alcanzar el límite gibbsita-caolinita los valores calculados por ambos códigos de $\log a\text{SiO}_2$ y de $\log a\text{K}^+/\text{aH}^+$ se hacen muy similares (las diferencias de pH son inferiores a 0.2 unidades) y las pautas de evolución prácticamente coincidentes.

Los resultados de la simulación del proceso suministrados por el código PHREEQE sí que presentan notables diferencias respecto a los anteriores (tabla 3). El número de etapas (puntos de corte en el diagrama de estabilidad) en las que puede segmentarse el proceso es distinto al definido por EQ3/6 y CHILLER: el sistema no llega a estar saturado respecto a la gibbsita, por lo que esta fase mineral no interviene en el proceso; el PHREEQE indica que el sistema alcanza la saturación en caolinita después que se hayan disuelto $14.78 \cdot 10^{-3}$ mmoles de microclina (en la figura 1 B corresponde al punto A). En este punto la solución se encuentra, no obstante, muy próxima al equilibrio (el índice de saturación de gibbsita es de -0.039; tabla 3).

A partir de este momento se produce una disolución incongruente según la reacción (3) y continúa hasta que, tras disolverse $99.66 \cdot 10^{-3}$ mmoles de microclina (tabla 3), se alcanza la saturación con moscovita (punto B; fig. 1 B). Este punto correspondería a un estado análogo al indicado para el punto 4 de la simulación realizada por el EQ3/6, aunque evidentemente la composición de la solución resultante es distinta en los dos casos.

La disolución incongruente continúa desde el punto B según la reacción (4) hasta que, tras disolverse $149.51 \cdot 10^{-3}$ mmoles de microclina, se ha consumido toda la caolinita precipitada en el paso A-B ($49.83 \cdot 10^{-3}$ mmoles) y la solución queda en equilibrio con este mineral (punto C). A partir de aquí la disolución de microclina se realiza de manera incongruente respecto a la moscovita según:



hasta que, tras disolverse un total de $223.99 \cdot 10^{-3}$ mmoles de microclina, la solución alcanza el equilibrio respecto a este feldespato en el punto D.

Como puede constatarse (fig. 1B), estas últimas etapas también constituyen una significativa variación respecto a los cálculos efectuados por el EQ3/6 y CHILLER. La evolución calculada por estos dos códigos terminaba en el punto triple de equilibrio caolinita-moscovita-microclina mientras que la del PHREEQE termina en un punto sobre la recta de equilibrio moscovita-microclina.

4. Bases de datos termodinámicos y verificación de resultados.

En el epígrafe anterior se han comentado los resultados definidos por tres códigos con sus respectivas bases de datos. Sin embargo, en el manual de referencia del PHREEQE (Parkhurst *et al.*, 1980, 1990) se utilizan los datos propuestos para las fases minerales por Robie *et al.* (1978) en la resolución de esta simulación geoquímica, omitiendo los de su base de datos y sin verificar las diferencias que ello supone.

Por ello, se ha repetido la simulación de alteración de microclina incorporando al código PHREEQE los datos de Robie *et al.* (1978); y los resultados obtenidos se han representado en la figura 2A manteniendo, a efectos comparativos, los proporcionados por este código con su

base de datos original. El número de etapas (puntos de corte en el espacio de fases) es en ambos casos similar y cualitativamente el proceso puede describirse en idénticos términos. Para los dos conjuntos de datos termodinámicos la solución entra en el diagrama de fases al alcanzar el equilibrio con caolinita sin que aparezca, por tanto, gibbsita en el proceso de simulación.

Sin embargo, las variaciones en las transferencias de masa y en las características químicas de la solución provocadas por el cambio en los datos termodinámicos son notables. Por ejemplo, la magnitud del proceso irreversible para alcanzar el punto B' es de $111 \cdot 10^{-3}$ mmoles de microclina disuelta mientras que con los datos originales suponía la disolución de $99.66 \cdot 10^{-3}$ mmoles. Para llegar al punto C' hace falta que se disuelvan $172 \cdot 10^{-3}$ mmoles de microclina mientras que para alcanzar su homólogo, el C, hacen falta $149.51 \cdot 10^{-3}$ mmoles (tabla 3).

Las diferencias en las características composicionales de la solución pueden apreciarse fácilmente si observamos los valores de $\log aK^+/aH^+$ que presenta para las situaciones de equilibrio caolinita - moscovita - solución (puntos B-C y B'-C', figura 2A). El valor de pH para los puntos B y C es de 5.8 mientras que para los puntos B' y C' es de 8.3.

Resultan evidentes, por tanto, las importantes modificaciones en las pautas de reacción obtenidas al variar los datos termodinámicos utilizados por un determinado código. Sin embargo, en los distintos resultados proporcionados por PHREEQE, CHILLER y EQ3/6 han podido contribuir, aparte de los distintos datos termodinámicos para las reacciones de equilibrio mineral, las variantes del modelo químico de Asociación Iónica considerado por cada código, su diferente estructura matemática y, sobre todo, las diferencias en los datos termodinámicos correspondientes a las reacciones de especiación.

Para verificar la incidencia de estos factores se procedió a repetir la simulación con un único código, considerando los valores termodinámicos para las reacciones de equilibrio mineral de los demás. De esta manera, y tomando el PHREEQE como código de referencia, se modificó su base de datos considerando los de SUPCRT 92 y posteriormente los de SOLTHERM. Los resultados obtenidos de esta manera reproducían casi con total exactitud los obtenidos por cada uno de los códigos que incorporaban originalmente los datos seleccionados. En la tabla 3 se presentan los resultados numéricos de la simulación realizada considerando la base de datos de SUPCRT 92, prácticamente idénticos a los obtenidos por EQ3/6 (tabla 2).

Los distintos planteamientos de cálculo de coeficientes de actividad (modelo químico) utilizados en los códigos analizados proporcionan similares resultados para las fuerzas iónicas de la solución resultante de la simulación (fuerza iónica $< 10^{-4}$ molal; p. ej. Drever, 1988). Por otro lado, incluyendo también los correspondientes valores de las reacciones de especiación (y no sólo los de las reacciones de equilibrio mineral) en los conjuntos de datos termodinámicos manejados por el PHREEQE en esta verificación no se consiguen aproximaciones mucho más exactas a los resultados ofrecidos por cada código; lo que indica que las diferencias en los algoritmos de resolución (véase también INTERA, 1983) y/o en las reacciones de especiación de cada código no son responsables de variaciones importantes en los resultados. Y, por tanto, las marcadas diferencias en las pautas de reacción definidas por cada código son resultado, fundamentalmente, de los datos termodinámicos considerados para las reacciones de equilibrio mineral-solución.

Desde esta perspectiva, la selección de los datos termodinámicos para el planteamiento del cálculo de pautas de reacción adquiere casi más importancia que la del propio código, especialmente en aproximaciones de método inverso (Plummer, 1984) en sistemas naturales. Los resultados reflejados en este análisis evidencian que el conjunto de datos seleccionado puede condicionar no sólo los valores numéricos de las transferencias de masa o características

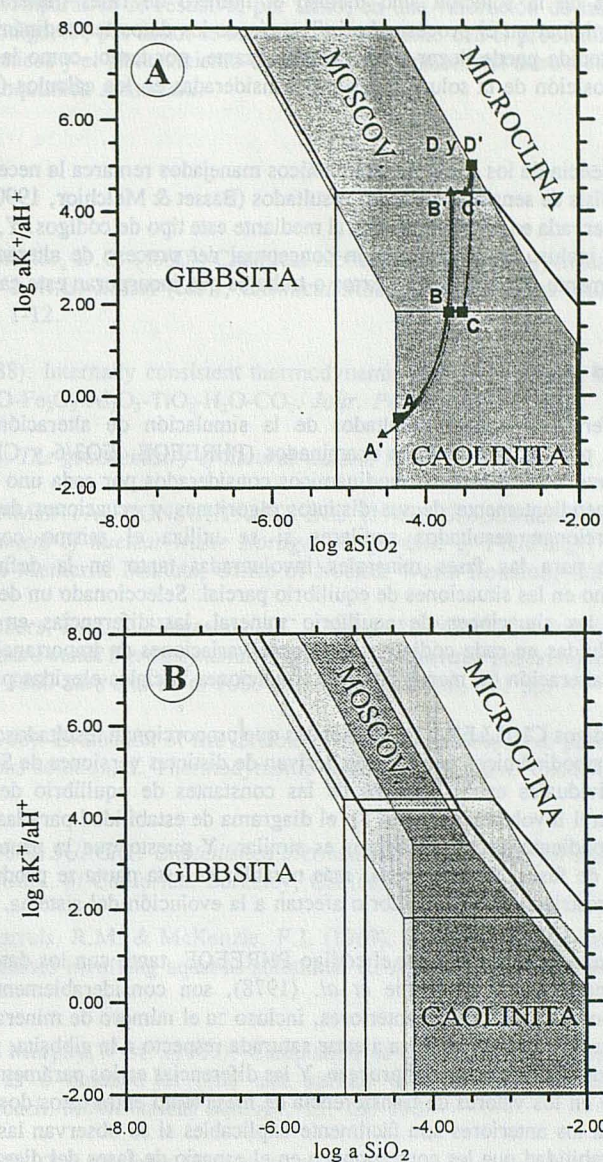


Figura 2.- A.- Pautas de evolución calculadas por el código PHREEQE tanto con su base de datos original (puntos A-D) como con los datos termodinámicos de Robie *et al.* (1978) (puntos A'-D') sobre sus respectivos diagramas de estabilidad mineral, $\log aK^+/aH^+ - \log aSiO_2$. Las áreas sombreadas corresponden a los campos de estabilidad de caolinita y moscovita definidos con los valores termodinámicos incluidos en la base de datos de este código. B.- Comparación de los diagramas de estabilidad definidos a partir de los datos de SUPCRT 92 y SOLTHERM (campos de estabilidad de caolinita y moscovita con sombreado intenso), PHREEQE (1990) (campos de estabilidad de caolinita y moscovita con sombreado menos intenso), Robie *et al.* (1978) (campos de estabilidad con trazo grueso) y Berman (1988) (campos de estabilidad con trazo muy grueso).

composicionales de la solución sino incluso el número de fases minerales susceptibles de intervenir y participar en el proceso. La influencia de los datos termodinámicos sobre la pauta de reacción obtenida puede llegar a ser tan importante, por tanto, como la de las condiciones iniciales (composición de la solución acuosa) consideradas en los cálculos (p. ej. Steinmann *et al.*, 1994).

La influencia de los datos termodinámicos manejados remarca la necesidad metodológica de plantear análisis de sensibilidad de los resultados (Basset & Melchior, 1990) al ajustar la pauta de reacción observada en un sistema natural mediante este tipo de códigos. Y, a la vez, evidencia su importancia incluso en la descripción conceptual del proceso de alteración de microclina, aspecto normalmente obviado en los textos o trabajos que incorporan este caso-tipo para tal fin.

5. Conclusiones

Las diferencias en los resultados de la simulación de alteración de la microclina proporcionados por los tres códigos examinados (PHREEQE, EQ3/6 y CHILLER) se deben fundamentalmente a los valores termodinámicos considerados por cada uno de ellos en su base de datos. Independientemente de sus distintos algoritmos y ecuaciones de gobierno, los tres códigos proporcionan resultados similares si se utiliza el mismo conjunto de valores termodinámicos para las fases minerales involucradas tanto en la definición del proceso irreversible como en las situaciones de equilibrio parcial. Seleccionado un determinado conjunto de datos para las situaciones de equilibrio mineral, las diferencias en las reacciones de especiación incluidas en cada código no provocan variaciones de importancia en los resultados del proceso de alteración (al menos para las condiciones iniciales elegidas para la simulación).

Los códigos CHILLER y EQ3/6 son los que proporcionan resultados más parecidos dado que los datos termodinámicos que utilizan derivan de distintas versiones de SUPCRT. Pese a las diferencias individuales en los valores de las constantes de equilibrio de las reacciones de disolución mineral involucradas (tabla 1), el diagrama de estabilidad para las fases en equilibrio parcial correspondiente a ambos códigos es similar. Y puesto que la pauta evoluciona en un mismo espacio de fases, las diferencias más notables en esta pauta se producirán al principio, cuando menos restricciones de equilibrio afectan a la evolución del sistema.

Los resultados definidos por el código PHREEQE, tanto con los datos de Nordstrom *et al.* (1990) como con los de Robie *et al.* (1978), son considerablemente distintos de los suministrados por los dos códigos anteriores, incluso en el número de minerales que intervienen en la simulación: la solución no llega a estar saturada respecto a la gibbsita, por lo que esta fase queda excluida de la descripción del proceso. Y las diferencias en los parámetros composicionales de la solución y en los valores de transferencia de masa tanto entre estos dos conjuntos de datos como respecto a los anteriores son fácilmente explicables si se observan las variaciones en los dominios de estabilidad que les corresponden en el espacio de fases del diagrama de estabilidad mineral (fig. 2 A y B).

Resulta, por tanto, evidente que la selección de un determinado conjunto de datos termodinámicos puede condicionar no sólo las magnitudes asociadas a las etapas que describen el proceso sino el propio número de etapas o de fases minerales involucradas en el proceso de alteración de microclina. La sistemática en la descripción de este proceso utilizada en la bibliografía corresponde, básicamente, a la de los datos termodinámicos de SUPCRT o EQ3/6. Y las variaciones inducidas por modificaciones en los valores termodinámicos pueden ser tan importantes como un cambio en las condiciones iniciales del cálculo.

El análisis comparativo se resuelve, por tanto, en una valoración de las bases de datos incluidas en cada código. Y, por tanto, se entronca directamente con el problema de la selección de datos termodinámicos y el planteamiento de análisis de sensibilidad dentro de la metodología de modelización geoquímica *sensu lato*.

Bibliografía

Basset, R.L. & Melchior, D.C. (1990): Chemical modeling of aqueous systems. An overview. In: D.C. Melchior & R.L. Basset (eds.), *Chemical Modeling of Aqueous Systems II*. A.C.S. Symp. Series, 416, 1-12.

Berman, R.G. (1988). Internally consistent thermodynamic data for minerals in system $\text{Na}_2\text{O}-\text{K}_2\text{O}-\text{CaO}-\text{MgO}-\text{FeO}-\text{Fe}_2\text{O}_3-\text{Al}_2\text{O}_3-\text{TiO}_2-\text{H}_2\text{O}-\text{CO}_2$. *Jour. Petrol.*, 29, 445-522.

Drever, J.J. (1988): *The geochemistry of natural waters*. Prentice-Hall Inc., N.J., 437 pp.

INTERA ENVIRONMENTAL CONSULTANTS INC. (1983): *Geochemical models suitable for performance assessment of nuclear waste storage: comparison of PHREEQE and EQ3/EQ6*. ONWI-473. Battelle Memorial Institute, Office of Nuclear Waste Isolation, Columbus, Ohio.

Johnson, J.W., Oelkers, E.H. & Helgeson, H.C. (1991): *SUPCRT92, a software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species and reactions from 1 to 5000 bars and 0° to 1000 °C*. User's manual, 101 pp.

Helgeson, H.C. (1968): Evaluation of irreversible reactions in geochemical processes involving minerals and aqueous solution. I. Thermodynamic relations. *Geochim. Cosmochim. Acta*, 32, 853-877.

Helgeson, H.C. (1984): *SUPCRT. Unpublished thermodynamic data base for minerals, aqueous species and gases*. Univ. of California, Berkeley, California.

Helgeson, H.C., Garrels, R.M. & McKenzie, F.I. (1969): Evaluation of irreversible reactions in geochemical processes involving aqueous solutions. II. Applications. *Geochim. Cosmochim. Acta*, 33, 455-481.

Helgeson, H.C. & Murphy, W.M. (1983). Calculation of mass transfer among minerals and aqueous solutions as a function of time and surface area in geochemical processes. I. Computational approach. *Mathematical Geology*, 15, 109-130.

Lichtner, P.C. (1988). The quasistationary state approximation to coupled mass transport and fluid rock interaction in a porous medium. *Geochim. Cosmochim. Acta*, 52, 143-165.

Nordstrom, D.K., Plummer, L.N., Langmuir, L., Busenberg, E., May, H.M., Jones, B.F. & Parkhurst, D.L. (1990). Revised chemical equilibrium data for major water-mineral reactions and their limitations. In: D.C. Melchior & R.L. Basset (eds.), *Chemical Modeling of Aqueous Systems II*. A.C.S. Symp. Series, 416, 398-413.

Parkhurst, D.L., Thorstenson, D.C. & Plummer, L.N. (1980): PHREEQE, a computer program for geochemical calculations. *U.S. Geol. Surv. Water Res. Inv.*, 80-96, 193pp.

Parkhurst, D.L., Thorstenson, D.C. & Plummer, L.N. (1990): PHREEQE, a computer program for geochemical calculations. Revised by J.V. Tisarranni & P.D. Glynn. *U.S. Geol. Surv. Water Res. Inv.*, 80-96, 193pp.

Plummer, L.N. (1984): Geochemical modeling: a comparison of forward and inverse methods. In: Hitchon, B. & Wallick, E.I. (eds.), *Practical Applications of Ground Water Geochemistry*. Proceedings of First Canadian/American Conference on Hydrogeology. Banff, Alberta, Canada. pp. 149-177.

Reed, M.H. (1982): Calculation of multicomponent chemical equilibria and reaction processes in systems involving minerals, gases and aqueous phase. *Geochim. Cosmochim. Acta*, 46, 513-528.

Reed, M.H. & Spycher, N.F. (1987): *SOLTHERM data base. A compilation of thermodynamic data from 25 °C to 300 °C for aqueous species, minerals and gases*. Eugene, Oregon, Dept. Geol. Sci., Univ. Oregon, 30 pp.

Reed, M.H. & Spycher, N.F. (1989): *SOLTHERM: Data base of equilibrium constants for aqueous-mineral-gas equilibria*. Dept. Geol. Sci., Univ. Oregon. Eugene, Oregon. 45 pp.

Robie, R.A., Hemingway, B.S. & Fisher, J.R. (1978): Thermodynamic properties of minerals and related substances at 298.15 °K and 1 bar (10^5 pascals) pressure and at higher temperatures. *Geol. Surv. Bull.*, 1452, 456 pp.

Spycher, N.F. & Reed, N.H. (1989): *CHILLER: A program for computing water-rock reactions, boiling, mixing and other reaction processes in aqueous-mineral-gas systems*. Dept. of Geological Sciences. University of Oregon. Eugene, Oregon, 63 pp.

Spycher, N.F. & Reed, N.H. (1992): Microcomputer-based modeling of speciation and water-mineral-gas reactions using programs SOLVEQ and CHILLER. In: Y.K. Kharaka and A.M. Maest (eds.), *Water-Rock Interaction*, 1087-1090. Balkema, Rotterdam.

Steefel, C.I. & Van Capellen, P. (1990): A new kinetic approach to modeling water-rock interaction: the role of nucleation, precursors and Ostwald ripening. *Geochim. Cosmochim. Acta*, 54, 2657-2677.

Steinmann, P.; Lichtner, P.C. & Shotyk, W. (1994). Reaction path approach to mineral weathering reactions. *Clays and Clay Minerals*, 42, 197-206.

Tena, J.M., Auqué, L.F., Gimeno, M.J., Fernández, J. y Mandado, J. (1990). *Estudio comparativo de software para solución de equilibrios fisicoquímicos*. CIEMAT, Madrid (no publicado), 4 vols.

Wolery, T.J. (1979): *Calculations of Chemical Equilibrium between aqueous solution and minerals: the EQ3/6 software Package*. NTIS UCRL-52658.

Wolery, T.J. & Daveler, S.A. (1992). *EQ6, a computer program for reaction path modeling of aqueous geochemical systems: theoretical manual, user's guide and related documentation (version 7.0)*. Lawrence Livermore National Laboratory. UCRL-MA-110662 PT IV.

Análisis comparativo de códigos de modelización geoquímica y bases de datos termodinámicos. I. Simulación de la alteración de microclina.

Auqué, L.F.⁽¹⁾; López, P.L.⁽¹⁾; Gimeno, M.J.⁽²⁾; Mandado, J.⁽¹⁾; Vallés, V.⁽³⁾; Zamora, A.⁽¹⁾ y Gómez, J.⁽¹⁾

(1) Área de Petrología y Geoquímica. Depto. Ciencias de la Tierra. Fac. Ciencias. Universidad de Zaragoza. 50009 ZARAGOZA

(2) Unidad de Seguridad de Emplazamientos. Edificio 20 T. CIEMAT. Avda. Complutense 22. 28006 MADRID.

(3) Laboratoire du Science du Sol d'Avignon, Institut National de la Recherche Agronomique. Domaine St. Paul, B.P. 91, 84143 Montfavet CEDEX (France).

Abstract

The more recent versions of the computer codes EQ3/6, CHILLER, PHREEQE and HARPHRQ are used to simulate a microcline weathering process (in closed system and reacting with a HCl solution of pH = 4, at 25 °C) with comparative purposes.

The calculated reaction paths are almost insensitive to the particular chemical model, governing equations or numerical algorithm used in those programs. But sensitivity of modeling results are strongly dependent of thermodynamic values enclosed in the code data-base. Differences in the equilibrium constants for the mineral reactions lead to important numerical changes in the predicted mass transfers, solution compositional parameters and even in the number of secondary minerals involved in the simulation. Comparatively, differences in the speciation reactions for the dissolved components (e.g. aluminum) lead to minor discrepancies in the results of the simulated weathering process.

In spite of efforts on database updating, model sensitivity to mineral equilibrium constants must be performed in reaction path modeling related with aluminous minerals.

1. Introducción

El desarrollo de las técnicas informáticas ha hecho posible que los cálculos de modelización geoquímica para el estudio de procesos de interacción agua-roca sean cada vez más frecuentes; y asimismo ha provocado una considerable proliferación de códigos capaces de realizar este tipo de cálculos desde aproximaciones termodinámicas (cálculos de pautas de reacción), mediante distintos planteamientos y con diferentes opciones de cálculo (p.ej. Basset y Melchior, 1990; Tena *et al.*, 1990). Esta circunstancia ha provocado que el análisis comparativo de códigos, desde múltiples perspectivas, se haya convertido en una de las prioridades en distintos proyectos o centros de investigación para delimitar el más adecuado y/o fiable en cada caso (Jenne, 1981; INTERA, 1983; Tena *et al.*, 1990; etc).

Como conclusión de muchos de esos trabajos se señalaba, implícita o explícitamente, que de los cuatro ejes básicos en la comparación de códigos, modelo químico considerado, planteamiento de las ecuaciones de gobierno, métodos de resolución matemática y bases de datos termodinámicos, era este último el principal responsable de las diferencias suministradas por los programas. Conclusión que, no obstante, se solía exponer genéricamente sin entrar a valorar el peso específico de los distintos tipos de reacciones incluidas en esas bases de datos y sin más precisión sobre sus consecuencias cuantitativas en la modelización de procesos geoquímicos.

El número de recopilaciones y bases de datos termodinámicos actualmente disponibles es considerable y los esfuerzos realizados por distintos organismos y grupos de trabajo en su revisión crítica, mejora y homogeneización han sido importantes. No obstante, muchos de estos esfuerzos han estado dirigidos a asegurar la consistencia interna (Basset y Melchior, 1990) de esas recopilaciones y/o valorar las diferencias existentes entre ellas (p. ej. Engi, 1992), pero también casi siempre de forma ajena a sus implicaciones y consecuencias en la modelización de procesos concretos.

El ritmo de aparición de nuevos códigos capaces de realizar cálculos de pautas de reacción y/o de nuevas versiones de programas ya existentes (con distintas mejoras en sus algoritmos y capacidades de cálculo pero, además, con bases de datos termodinámicos progresivamente más refinadas y revisadas) es elevado. Y, por ello, resulta aconsejable actualizar periódicamente los análisis comparativos sobre estas nuevas versiones y programas. Con este artículo se inicia una serie de trabajos cuyo propósito es el análisis comparado de algunos de los códigos más ampliamente utilizados en el campo de la modelización termodinámica de procesos de interacción agua-roca; análisis que se centrará específicamente en la incidencia de sus diferentes bases de datos termodinámicos sobre los resultados obtenidos en la simulación de determinados procesos.

Los códigos comparados han sido diseñados para tratar procesos muy variados y en ambientes con caracteres fisicoquímicos diversos, razón por la que sus bases de datos incluyen un amplio número de elementos químicos con sus correspondientes sistemas de especiación y reacciones de equilibrio mineral (además de incluir los datos necesarios para trabajar en rangos más o menos amplios de temperatura y, a veces, de presión). Este análisis comparativo se realizará mediante simulaciones que se desarrollen preponderantemente en determinados sistemas composicionales para poder analizar separadamente las discrepancias asociadas específicamente a alguno de los más importantes. En este primer trabajo se utilizará la simulación de un proceso de hidrólisis de microclina, a 25 °C como referencia de la comparación. Este proceso evolucionará de forma simplificada sobre el sistema Si-Al-K-H₂O y permitirá restringir el análisis sobre las tabulaciones termodinámicas existentes para las reacciones de especiación y equilibrio mineral del aluminio, uno de los elementos básicos en la modelización de procesos naturales.

2. Antecedentes y planteamiento del estudio.

La simulación del proceso de alteración de microclina constituye uno de los ejemplos normalmente incluido en los manuales de usuario de los programas termodinámicos con capacidad de plantear este tipo de cálculo (como por ejemplo EQ3/6 y PHREEQE; Wolery y Daveler, 1992; Parkhurst *et al.*, 1980, 1990). Y, además, este proceso se ha convertido en el ejemplo por excelencia de un cálculo de pautas de reacción en distintos textos docentes (p. ej. Richardson y McSween, 1989) o en el proceso sobre el que analizar didácticamente las variaciones producidas por algunos de los condicionamientos de un fenómeno de alteración (Drever, 1988; Faure, 1992).

Sin embargo, ha sido pocas veces utilizado para realizar explícitamente análisis comparativos entre distintos códigos y únicamente los trabajos de INTERA (1983) y Tena *et al.* (1990) emplearon esta simulación para comparar los resultados de las versiones entonces existentes de EQ3/6 y PHREEQE. El planteamiento de este análisis comparativo seguirá, a grandes rasgos, el esquema utilizado por Tena *et al.* (1990): inicialmente se planteará la simulación con cada uno de los códigos comparados para luego pasar a repetirla con un único código pero cambiando, parcial o totalmente, su base de datos por la de los demás.

Esta metodología de comparación permite, por un lado, una valoración directa de la influencia de las distintas bases de datos sobre los resultados del proceso y, por otro, una estimación indirecta de la incidencia del resto de factores condicionantes: modelo químico, algoritmo numérico, etc. La separación específica de los efectos producidos por variaciones en las reacciones de especiación y por los debidos a las de equilibrio mineral se realizará también de forma indirecta, cambiando los datos termodinámicos para las reacciones minerales involucradas en la simulación de un código a otro. Evidentemente, esta aproximación no plantea una comparación precisa de las diferencias cuantitativas entre los datos termodinámicos manejados pero realza, de forma explícita, las consecuencias de esas diferencias en los resultados de un cálculo de simulación.

Verificado que el resto de posibles condicionantes de los resultados de la simulación no influyen de forma apreciable, este planteamiento permite reinterpretar el análisis comparativo en términos de un "análisis de sensibilidad" (*model sensitivity analysis*; Nordstrom y Ball, 1989; Basset y Melchior, 1990) respecto a los datos termodinámicos manejados en la modelización: tomando como referencia el proceso simulado en unas determinadas condiciones puede evaluarse la incidencia de los datos termodinámicos sobre el modelo de alteración de microclina definido.

3. Códigos seleccionados y metodología de simulación.

Los códigos seleccionados para el análisis comparativo han sido el EQ3/6 (versión 7.2a; Wolery y Daveler, 1992), CHILLER (Reed, 1982; Spycher y Reed, 1989, 1992), PHREEQE (revisión de 1990; Parkhurst *et al.*, 1980, 1985, 1990) y HARPHRQ (Brown *et al.*, 1991). Se trata de códigos de pautas de reacción pertenecientes al grupo de resolución mediante la constante de equilibrio y basados todos ellos en modelos de Asociación Iónica (distintas variantes de la ecuación de Debye-Hückel: tipo "B-dot" para EQ3/6 y CHILLER, y tipo WATEQ, ecuación de Debye-Hückel ampliada o ecuación de Davies en PHREEQE y HARPHRQ; Tena *et al.*, 1990).

Todos ellos pueden plantear la simulación de un proceso irreversible de alteración de microclina a través de sucesivos estados de equilibrio parcial, aunque desde distintas filosofías de cálculo. La diferencia fundamental se refiere a la presencia (EQ3/6, CHILLER y HARPHRQ) o ausencia (PHREEQE) de un algoritmo de definición de las fases minerales estables en cada punto discreto del avance de la reacción (Tena *et al.*, 1990). Circunstancia que se traduce en una mayor necesidad de datos de entrada al plantear la simulación con el PHREEQE ya que, por sí mismo, es incapaz de determinar las fases minerales estables a medida que avanza el proceso irreversible. Asimismo, los supuestos y planteamientos de las ecuaciones de gobierno en cada código son también diferentes, especialmente en el caso de las relacionadas con el control de hidrógeno, oxígeno y/o agua. Entre los cuatro códigos abarcan la casi totalidad de posibles planteamientos actualmente existentes. Una descripción más detallada de las diferencias de planteamiento en las ecuaciones de gobierno de estos códigos puede verse en Reed (1982) o Tena *et al.* (1990).

La versión del código EQ3/6 utilizada en este trabajo dispone de cinco bases de datos distintas, seleccionables de forma independiente según las preferencias del usuario. Para la realización de este trabajo se han seleccionado dos de ellas, la DATA.COM y la derivada de SUPCRT92 (Johnson *et al.*, 1991), como bases más adecuadas a las características composicionales del proceso simulado; el CHILLER utiliza la base de datos SOLTHERM (Reed y Spycher, 1987, 1989) en gran parte derivada de la versión de SUPCRT 84 (Helgeson, 1984); la base de datos del PHREEQE procede fundamentalmente de las revisiones de Nordstrom *et al.* (1990) y de Tisaranani y Glynn (1990; en Parkhurst *et al.*, 1990); y el HARPHRQ hace uso de la base de datos HATCHES (Bond *et al.*, 1992) del Laboratorio Harwell (Gran Bretaña).

Con cada código se planteará la simulación del proceso de alteración de microclina considerando una solución de HCl diluido (pH=4), a 25 °C. El conjunto de fases minerales susceptibles de precipitar durante el proceso se reducirá a gibbsita, caolinita y moscovita (los valores de las constantes de equilibrio utilizadas por los distintos códigos para las reacciones de disolución de estos minerales puede verse en la tabla 1).

El planteamiento de las simulaciones se realizará en condiciones de sistema cerrado; es decir, manteniendo dentro del sistema los productos minerales precipitados durante la disolución irreversible de la microclina (y que, por tanto, podrán disolverse posteriormente conforme cambian las condiciones del sistema; p. ej. Wolery, 1979). La variable de progreso de reacción queda definida por el número de moles disueltos de microclina y el proceso se detendrá cuando la fase acuosa alcance el equilibrio con ese mineral. Estas condiciones de simulación son las mismas que aparecen en los manuales de EQ3/6 (Wolery y Daveler, 1992) y PHREEQE (Parkhurst *et al.*, 1980, 1990), así como en el análisis comparativo de anteriores versiones de EQ3/6 y PHREEQE realizado por INTERA (1983) o por Tena *et al.* (1990).

Operativamente, las simulaciones pueden contemplarse como el cálculo de los puntos de intersección de la solución con los límites de las fases minerales consideradas dentro de un diagrama de campos de estabilidad. Y así, la comparación de las pautas de reacción obtenidas será realizada sobre un diagrama de estabilidad mineral $\log aK^+/aH^+ - \log aSiO_2$, construido a partir de los datos termodinámicos considerados por cada código para las fases minerales involucradas (microclina, gibbsita, moscovita y caolinita). Sobre este diagrama se proyectarán las características composicionales de la solución correspondientes a las situaciones de equilibrio (puntos de intersección con los límites de fase) calculadas en cada caso. Los datos presentados en las tablas corresponderán a estos puntos de corte de las pautas calculadas.

Tabla 1: Valores de las constantes de equilibrio (log K, a 25 °C) incluidas en las bases de datos de los distintos códigos para las reacciones de disolución de los minerales involucrados en la simulación.

	EQ3/6 SUPCRT92	EQ3/6 DATA.COM	CHILLER SOLTHERM	PHREEQE Versión 1990	HARPHRQ HATCHES	Robie <i>et al.</i> (1978)
Gibbsita	6.796	6.796	7.960	8.110	8.770	8.049
Caolinita	5.101	5.101	7.420	7.435	8.480	5.708
Moscovita	11.022	11.022	14.560	12.970	14.560	12.970
Microclina	-1.130	-1.130	0.080	0.875	0.080	0.875

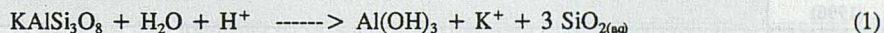
Finalmente, los cuatro códigos usados necesitan la especificación de una solución inicial que satisfaga las condiciones definidas anteriormente. Para las características de $\text{pH} = 4$ (HCl diluido) se considerará agua pura, fijando un $\text{pH} = 4$ y añadiendo Cl^- hasta satisfacer totalmente la condición de neutralidad eléctrica de la solución. Esta operación supone la adición de 0.10116 mmoles de Cl^- /kg de agua en la versión examinada de EQ3/6 y de 0.1012 mmoles/kg de agua en las de CHILLER, PHREEQE y HARPHRQ. Esta coincidencia de valores entre los códigos constata el similar funcionamiento del algoritmo relacionado con la condición de electro-neutralidad.

4. Resultados de la simulación. Análisis comparativo.

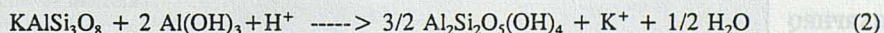
Los valores calculados por EQ3/6, CHILLER, PHREEQE y HARPHRQ para el proceso de alteración de microclina a 25°C se resumen en la tabla 2, indicándose los caracteres más destacables del quimismo de la solución y de las transferencias de masa relacionadas con los límites de fase involucrados en el proceso. El EQ3/6 (con las dos bases de datos manejadas) y el CHILLER discretizan la pauta de reacción en el mismo número de etapas (puntos de corte en el espacio de fases), con caracteres cualitativamente similares pero con algunas diferencias cuantitativas en las proporciones de progreso de reacción, en las características composicionales de la solución y en los valores de transferencia de masa calculados.

En la figura 1A se ha representado la evolución obtenida por cada uno de estos códigos sobre el diagrama actividad-actividad: los puntos indicados 1 a 5 corresponden a los límites de fase calculados por el EQ3/6 (para los parámetros involucrados en el diagrama los resultados proporcionados por las dos bases de datos manejadas son iguales); y los señalados de A a E son los obtenidos por el código CHILLER. Los campos de estabilidad definidos a partir de los datos termodinámicos considerados en las dos bases de datos de EQ3/6 son iguales (las constantes de equilibrio para las fases representadas también lo son; tabla 1). Y, a su vez, estos campos resultan prácticamente coincidentes con los definidos a partir de los datos del código CHILLER pese a las diferentes constantes de equilibrio que presenta este código (tabla 1).

La pauta obtenida por EQ3/6 y CHILLER predice que el sistema alcanza la saturación respecto a la gibbsita (puntos 1 y A; figura 1A) tras disolver congruentemente $4.422 \cdot 10^{-3}$ y $13.9 \cdot 10^{-3}$ mmoles de microclina, respectivamente. A partir de este momento la disolución del feldespato es incongruente según la reacción:



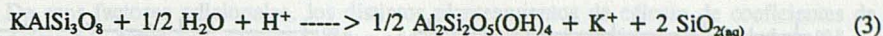
es decir, precipitando gibbsita hasta que el sistema alcanza la saturación con caolinita (punto 2 y B) después de haber disuelto casi $19 \cdot 10^{-3}$ mmoles de microclina y precipitado $15.87 \cdot 10^{-3}$ mmoles de gibbsita en el caso de EQ3/6 y $7 \cdot 10^{-5}$ mmoles en el caso de CHILLER. A partir de este punto, la disolución incongruente de microclina continúa según la reacción:



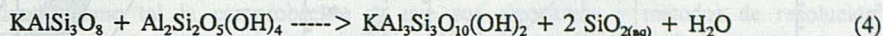
precipitando caolinita y disolviendo gibbsita hasta que, cuando se han disuelto $22.7 \cdot 10^{-3}$ mmoles o $27.2 \cdot 10^{-3}$ mmoles de microclina (según CHILLER y EQ3/6 respectivamente), se han consumido las proporciones de gibbsita precipitadas anteriormente y el sistema deja de estar saturado en este mineral a partir de los puntos 3 y C. En esos puntos, las proporciones de caolinita precipitada calculadas son de $12.4 \cdot 10^{-3}$ mmoles (EQ3/6, con ambas bases de datos) y $5.46 \cdot 10^{-3}$ mmoles (CHILLER). A partir de ese momento la reacción pasa a ser incongruente respecto a la caolinita:

Tabla 2: Resultados de la simulación de alteración de microclina a 25 °C con los códigos EQ3/6 (versión 7.2a, con las bases de datos SUPCRT92 y DATA.COM), CHILLER, PHREEQE (1990) y HARPHRQ. Los resultados de transferencia de masa mineral están expresados en forma de masa acumulada en la pauta obtenida.

EQ3/6 SUPCRT92	Micrl. disuelta (μ moles)	log aK ⁺ /aH ⁺	log aSiO ₂	pH	Indices de Saturación (I.S.)				μ moles precipitados		
					Gibb.	Caol.	Mosc.	Micrl.	Gib.	Caol.	Mosc.
1) Gibbsita	4.422	-1.278	-4.877	4.081	0.000	-1.262	-6.540	-7.980	0.00	0.00	0.00
2) Caol.- gib.	18.923	-0.589	-4.246	4.139	0.000	0.000	-3.960	-5.402	15.87	0.02	0.00
3) Caol.- gib.	27.204	-0.393	-4.246	4.177	0.000	0.000	-3.764	-5.205	0.00	12.44	0.00
4) Moscovita	101.90	3.371	-3.692	7.367	-0.554	0.000	0.000	-0.333	0.00	50.95	0.00
5) Microclina	149.59	3.371	-3.525	7.366	-0.721	0.000	0.000	0.000	0.00	3.72	47.34
EQ3/6 DATA.COM											
* Gibbsita	4.458	-1.275	-4.874	4.082	0.000	-1.255	-6.528	-7.966	0.00	0.00	0.00
* Caol.-gib.	18.923	-0.589	-4.246	4.139	0.000	0.000	-3.960	-5.401	15.84	0.00	0.00
* Caol.-gib.	27.187	-0.393	-4.246	4.177	0.000	0.000	-3.764	-5.205	0.00	12.41	0.00
* Moscovita	102.60	3.371	-3.692	7.365	-0.554	0.000	0.000	-0.333	0.00	51.29	0.00
* Microclina	152.66	3.371	-3.525	7.354	-0.721	0.000	0.000	0.000	0.00	5.25	47.37
CHILLER											
A) Gibbsita	13.900	-0.544	-4.380	4.318	0.000	-0.270	-4.370	-	0.00	0.00	0.00
B) Caol.-gib.	18.800	-0.401	-4.250	4.332	0.000	0.000	-3.840	-	0.07	0.00	0.00
C) Caol.-gib.	22.700	-0.307	-4.250	4.344	0.000	0.000	-3.730	-	0.00	5.46	0.00
D) Moscovita	102.500	3.400	-3.690	7.393	-0.550	0.000	0.000	-0.350	0.00	51.23	0.00
E) Microclina	148.500	3.407	-3.530	7.413	-0.710	0.000	0.000	0.000	0.00	6.04	45.46
PHREEQE (1990)											
A) Caolinita	14.780	-0.487	-4.353	4.349	-0.039	0.000	-2.304	-6.351	0.00	0.00	0.00
B) Mos.-caol.	99.660	1.817	-3.700	5.824	-0.692	0.000	0.000	-2.741	0.00	49.83	0.0
C) Mos.-caol.	149.51	1.817	-3.524	5.824	-0.868	0.000	0.000	-2.389	0.00	0.00	49.83
D) Microclina	223.99	5.000	-3.391	8.831	-2.052	-2.121	0.000	0.000	0.00	0.00	74.66
HARPHRQ											
* Caolinita	18.182	-0.271	-4.263	4.475	-0.267	0.000	-2.111	-4.637	0.00	0.00	0.00
* Mos.-caol.	99.718	1.840	-3.700	5.846	-0.830	0.000	0.000	-1.400	0.00	49.85	0.00
* Mos.-caol.	149.59	1.840	-3.524	5.846	-1.006	0.000	0.000	-1.048	0.00	0.00	49.85
* Microclina	152.99	3.387	-3.516	7.383	-1.530	-1.031	0.000	0.000	0.00	0.00	50.99



y después de disolverse en torno a los $102 \cdot 10^{-3}$ mmoles de microclina, el sistema llega al equilibrio con la moscovita, (punto 4 y D), precipitando alrededor de $51.0 \cdot 10^{-3}$ mmoles de caolinita según ambos códigos. La reacción continúa según:



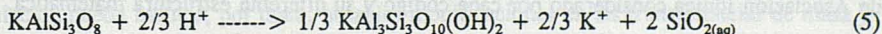
hasta que, cuando se han disuelto $149.59 \cdot 10^{-3}$ mmoles (EQ3/6-SUPCRT92), $152.66 \cdot 10^{-3}$ mmoles (EQ3/6-DATA.COM) o $148.5 \cdot 10^{-3}$ mmoles (CHILLER) de feldespato, el sistema alcanza el equilibrio con la microclina (también lo está con caolinita y moscovita), precipitando $45 \cdot 10^{-3}$ - $47 \cdot 10^{-3}$ mmoles de moscovita y disolviendo gran parte de la caolinita anteriormente precipitada, deteniéndose el proceso en el punto 5 o E, cuando quedan sólo $3.72 \cdot 10^{-3}$, $5.25 \cdot 10^{-3}$ o $6 \cdot 10^{-3}$ mmoles de caolinita según EQ3/6-SUPCRT92, EQ3/6-DATA.COM y CHILLER, respectivamente.

Las características composicionales calculadas para la solución a lo largo del proceso son muy similares para ambos códigos. La mayores diferencias entre EQ3/6 (con cualquiera de las bases de datos manejadas) y CHILLER se producen al comienzo de la simulación, cuando la solución alcanza el equilibrio con gibbsita y entra en el diagrama de estabilidad (puntos 1 y A; figura 1A). Al alcanzar el límite gibbsita-caolinita los valores calculados por ambos códigos de $\log a\text{SiO}_2$ y de $\log (a\text{K}^+/a\text{H}^+)$ se hacen muy similares (las diferencias de pH son inferiores a 0.2 unidades) y las pautas de evolución prácticamente coincidentes.

Los resultados de la simulación del proceso suministrados por los códigos PHREEQE y HARPHRQ sí que presentan notables diferencias respecto a los anteriores (tabla 2). El número de etapas (puntos de corte en el diagrama de estabilidad) en las que puede segmentarse el proceso es distinto al definido por EQ3/6 y CHILLER y el sistema no llega a estar saturado respecto a la gibbsita, por lo que esta fase mineral no interviene en el proceso. El PHREEQE indica que el sistema alcanza la saturación en caolinita después que se hayan disuelto $14.78 \cdot 10^{-3}$ mmoles de microclina (en la figura 1B corresponde al punto A), aunque en este punto la solución se encuentra muy próxima al equilibrio con gibbsita (el índice de saturación respecto a esa fase es de -0.039; tabla 2). El HARPHRQ alcanza esta situación de equilibrio con caolinita cuando se han disuelto $18.18 \cdot 10^{-3}$ mmoles de microclina, estando la solución algo más subsaturada (índice de saturación = -0.267; tabla 2) respecto a la gibbsita que con el PHREEQE.

A partir de este momento se produce una disolución incongruente según la reacción (3) y continúa hasta que, tras disolverse $99.7 \cdot 10^{-3}$ mmoles de microclina (tabla 2) según ambos códigos, se alcanza la saturación con moscovita (punto B de la figura 1B en el caso del PHREEQE; la situación de este punto calculada por el HARPHRQ sería prácticamente coincidente). Este punto correspondería a un estado análogo al indicado para el punto 4 de la simulación realizada por el EQ3/6 aunque, evidentemente, la composición de la solución resultante es distinta.

La disolución incongruente continúa desde el punto B según la reacción (4) hasta que, tras disolverse $149.5 \cdot 10^{-3}$ mmoles de microclina, se ha consumido toda la caolinita precipitada en el paso A-B ($49.8 \cdot 10^{-3}$ mmoles para ambos códigos) y la solución queda en equilibrio con este mineral (punto C; figura 1B). A partir de aquí la disolución de microclina se realiza de manera incongruente respecto a la moscovita según:



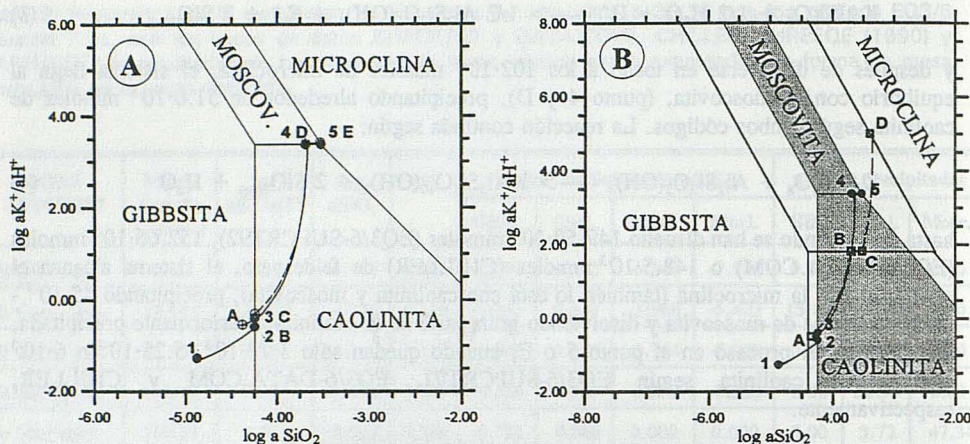


Figura 1.- A. Pautas de evolución calculadas por el código EQ3/6 (puntos 1-5) y CHILLER (puntos A-E) para la simulación de alteración de microclina representadas sobre un diagrama de estabilidad mineral $\log aK^+/aH^+ - \log aSiO_2$. El espacio de fases obtenido con los datos termodinámicos de ambos códigos es prácticamente coincidente. B. Pauta de evolución calculada por el código PHREEQE (puntos A-D) sobre el diagrama de estabilidad obtenido a partir de los valores de su base de datos. A efectos comparativos se mantiene la pauta definida por el código EQ3/6 (puntos 1-5) y las áreas sombreadas corresponden a los campos de estabilidad de caolinita y moscovita definidos con los datos termodinámicos de este último código.

hasta que, tras disolverse un total de $223.99 \cdot 10^3$ mmoles de microclina según el PHREEQE (punto D; fig. 1B) o $152.99 \cdot 10^3$ mmoles según el HARPHRQ, la solución alcanza el equilibrio respecto a este feldespato.

Como puede constatarse (fig. 1B), esta última etapa también constituye una significativa variación respecto a los cálculos efectuados por el EQ3/6 y CHILLER. La evolución calculada por estos dos códigos terminaba en el punto triple de equilibrio caolinita-moscovita-microclina mientras que la del PHREEQE y HARPHRQ termina en un punto sobre la recta de equilibrio moscovita-microclina. Por otro lado, en esta última etapa es en la que se aprecian también mayores diferencias entre PHREEQE y HARPHRQ: los resultados de transferencia de masa y los caracteres composicionales de la solución calculados por ese último código tienden a aproximarse a los obtenidos por EQ3/6 y CHILLER. Resultado justificable *a priori* si consideramos que los valores de la constante de equilibrio para microclina y moscovita utilizados por HARPHRQ son los mismos que los de EQ3/6 y CHILLER (tabla 1).

5. Bases de datos termodinámicos y verificación de resultados.

En el epígrafe anterior se han comparado, fundamentalmente, las pautas definidas por cuatro códigos con sus respectivas bases de datos. Y, por ello, a las diferencias en los resultados proporcionadas por PHREEQE, CHILLER, EQ3/6 y HARPHRQ han podido contribuir, aparte de los distintos datos termodinámicos para las reacciones de equilibrio mineral, las diferencias en los datos correspondientes a las reacciones de especiación, las variantes del modelo químico de Asociación Iónica considerado por cada código y su diferente estructura matemática.

De esos factores adicionales, los distintos planteamientos de cálculo de coeficientes de actividad (modelo químico) utilizados en los códigos analizados no deben tener transcendencia ya que todos ellos proporcionan similares resultados para las fuerzas iónicas ($< 10^{-4}$ molal) de la solución resultante de la simulación (INTERA, 1983; Drever, 1988; Tena *et al.*, 1990; etc.). Por otro lado, los distintos planteamientos matemáticos incluidos en cada código tampoco serán responsables *a priori* de estas diferencias, ya que al menos la verificación de los programas (entendiendo como tal la comprobación de que sus algoritmos y métodos de resolución matemática funcionan correctamente; p. ej. Silling, 1983; Tsang, 1991) puede darse por supuesta dada la amplia utilización de estos códigos.

Para verificar la incidencia de los condicionamientos asociados a los datos termodinámicos se procedió a repetir la simulación con un único código, considerando las constantes para las reacciones de equilibrio mineral de los demás. De esta manera, y tomando el PHREEQE como código de referencia, se modificó su base de datos considerando los de SUPCRT92, SOLTHERM y HARPHRQ. Los resultados obtenidos por PHREEQE de esta manera reproducían casi con total exactitud los obtenidos por cada uno de los códigos que incorporaban originalmente los datos seleccionados.

Adicionalmente, esta incidencia preponderante de los datos termodinámicos considerados para las reacciones de equilibrio mineral puede comprobarse analizando y ampliando el planteamiento de esta simulación que aparece en el propio manual de referencia del código PHREEQE (Parkhurst *et al.*, 1980, 1985, 1990). En el ejemplo de simulación del proceso de alteración de microclina que se desarrolla en este manual se utilizan los datos propuestos para las fases minerales (tabla 1) por Robie *et al.* (1978), omitiendo los de la propia base de datos del código y sin verificar las diferencias que ello supone. Cuestión de notable interés puesto que, si comparamos los valores de las constantes de equilibrio incluidas en la base de datos de este código (versión de 1990) y los propuestos por Robie *et al.* (1978) puede apreciarse que únicamente hay diferencias significativas en la constante de la caolinita (tabla 1).

Si comparamos los resultados de la simulación de alteración de microclina, a 25 °C, considerando los datos de Robie *et al.* (1978; tabla 3, versión 1990), con los proporcionados por este código con su base de datos (tabla 2) puede constatar (figura 2A) que el número de etapas es en ambos casos similar y cualitativamente el proceso puede describirse en idénticos términos. Con los datos de Robie *et al.* (1978) la solución entra en el diagrama de fases dentro del campo de la caolinita (punto A'; fig. 2A), en un punto más alejado del límite gibbsita-caolinita que cuando lo hacía con los de su base de datos (punto A; fig. 2A). Con los dos conjuntos de datos termodinámicos la gibbsita queda excluida del proceso.

Sin embargo, las variaciones en las transferencias de masa y en las características químicas de la solución son notables, especialmente en aquellos en los que se ve involucrada la caolinita. Por ejemplo, la magnitud del proceso irreversible para alcanzar el punto B' es de $111 \cdot 10^{-3}$ mmoles de microclina disuelta mientras que con los datos originales suponía la disolución de $99.66 \cdot 10^{-3}$ mmoles. Para alcanzar el punto C' hace falta que se disuelvan $172 \cdot 10^{-3}$ mmoles de microclina mientras que para alcanzar su homólogo, el C, hacen falta $149.51 \cdot 10^{-3}$ mmoles. Las diferencias en las características composicionales de la solución pueden apreciarse fácilmente si observamos los valores de $\log aK^+/aH^+$ que presenta para las situaciones de equilibrio caolinita - moscovita - solución (puntos B-C y B'-C', figura 2A). El valor de pH para los puntos B y C es de 5.8 mientras que para los puntos B' y C' es de 8.3.

La variación de una sola de las constantes de los equilibrios minerales involucrados en la simulación puede inducir, por tanto, notables modificaciones en las transferencias de masa y

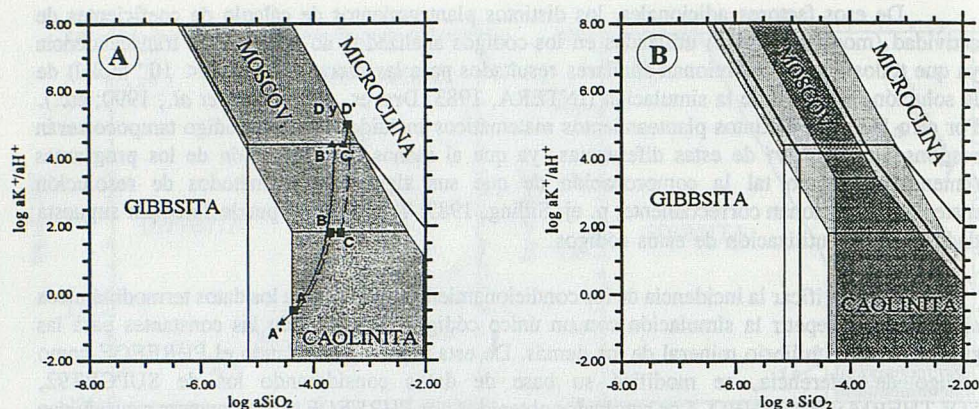


Figura 2.- A.- Pautas de evolución calculadas por el código PHREEQE tanto con su base de datos original (puntos A-D; tabla 2) como con los datos termodinámicos de Robie *et al.* (1978) (puntos A'-D'; tabla 3, versión 1990) sobre sus respectivos diagramas de estabilidad mineral, $\log aK^+/aH^+ - \log aSiO_2$. Las áreas sombreadas corresponden a los campos de estabilidad de caolinita y moscovita definidos con los valores termodinámicos incluidos en la base de datos de este código. B.- Comparación de los diagramas de estabilidad definidos a partir de los datos de SUPCRT92 y SOLTHERM (campos de estabilidad de caolinita y moscovita con sombreado intenso), PHREEQE (1990) (campos de estabilidad de caolinita y moscovita con sombreado menos intenso), Robie *et al.* (1978) (campos de estabilidad con trazo grueso) y Berman (1988) (campos de estabilidad con trazo muy grueso).

caracteres composicionales de la solución; modificaciones que, cualitativamente y para determinadas etapas de la evolución del proceso, pueden ser de magnitudes similares a las provocadas por el manejo de dos códigos distintos.

Una estimación precisa de la incidencia que tienen las diferencias en las reacciones de especiación sobre los resultados de la simulación puede obtenerse manejando dos bases de datos, diferentes pero con iguales valores para las constantes de equilibrio mineral, en un mismo código (con ello se consigue que todo el resto de posibles condicionantes de los resultados sean iguales, salvo los referidos a las reacciones de especiación manejadas). Este planteamiento puede verificarse considerando dos versiones distintas del código PHREEQE, la actual de 1990 y la inmediatamente anterior, de 1985 (Parkhurst *et al.*, 1985), considerando un mismo conjunto de datos para las reacciones de equilibrio involucradas. Las variaciones entre una y otra versión no afectan al número de reacciones de especiación incluidas pero sí a los valores termodinámicos considerados para las de aluminio y sílice, elementos de comportamiento fundamental en el proceso simulado.

Los resultados de la simulación con ambas versiones, considerando los datos termodinámicos de Robie *et al.* (1978), se resumen en la tabla 3. Y, como puede apreciarse, las diferencias existentes se reducen a escasas variaciones numéricas en los valores de transferencia de masa y quimismo de la solución durante la evolución del proceso. Variaciones que, numéricamente y salvando las diferencias en la discretización de la pauta, son del mismo rango que las definidas por EQ3/6 y CHILLER (tabla 1).

La utilización del código EQ3/6 con dos bases de datos distintas también verifica las premisas planteadas para el análisis de la incidencia de las reacciones de especiación. Como puede apreciarse (tabla 1), las diferencias en los resultados proporcionados por el uso de estas dos bases es muy pequeña. Los caracteres descriptivos del quimismo de la solución son iguales, existiendo pequeñas variaciones en las transferencias de masa calculadas únicamente apreciables en los estadios más avanzados de la simulación. De esta forma, con la base de datos SUPCRT92 la simulación termina tras haber disuelto $149.59 \cdot 10^3$ mmoles de microclina, mientras que con la base de datos DATA.COM lo hace tras disolver $152.66 \cdot 10^3$ mmoles (tabla 1). La causa de esta diferencia se encuentra en que la base de datos DATA.COM incluye una especie de la sílice, $H_6(H_2SiO_4)_4^{2-}$, inexistente en el resto de bases de datos y cuya concentración comienza a ser importante únicamente cuando el pH de la solución alcanza valores superiores a 7; es decir, en los últimos estadios de la simulación. Eliminando esta especie de DATA.COM los resultados proporcionados son totalmente iguales a los obtenidos con SUPCRT92.

Los resultados de la simulación son, por tanto, preponderantemente sensibles a los valores considerados para las reacciones heterogéneas intervinientes en el proceso. Y en las bases de datos utilizadas existen importantes diferencias en los valores de las constantes de equilibrio para esas reacciones. Diferencias que, en definitiva, se traducen en la definición de campos de estabilidad mineral considerablemente distintos para el espacio de fases en el que debe evolucionar el proceso. En la figura 2B se han representado de forma comparada los campos de estabilidad mineral correspondientes a los distintos conjuntos de datos manejados (a los que se ha añadido los de otra de las bases de datos más frecuentemente utilizada en la bibliografía sobre el tema, la de Berman, 1988) y las diferencias existentes entre algunos de los límites de fase superan los dos órdenes de magnitud. Circunstancia que, evidentemente, justifica las importantes variaciones delimitadas en los resultados del proceso simulado.

Tabla 3: Resultados de la simulación de alteración de microclina a 25 °C, considerando los datos de Robie *et al.* (1978), con dos versiones distintas del código PHREEQE (versiones de 1985 y 1990). Los resultados de transferencia de masa mineral están expresados en forma de masa acumulada en la pauta obtenida.

PHREEQE (1985)	Micrl. disuelta (μ moles)	log aK ⁺ /aH ⁺	log aSiO ₂	pH	Índices de Saturación (I.S.)				μ moles precipitados		
					Gibb.	Caol.	Mosc.	Micrl.	Gib.	Caol.	Mosc.
* Caolinita	9.222	-0.853	-4.558	4.188	-0.637	0.000	-5.261	-7.990	0.00	0.00	0.00
* Mos.-caol.	109.58	4.408	-3.671	8.374	-1.524	0.000	0.000	-0.955	0.00	54.78	0.00
* Mos.-caol.	168.78	4.408	-3.483	8.362	-1.712	0.000	0.000	-0.578	0.00	0.00	56.26
* Microclina	216.74	5.028	-3.400	8.874	-2.001	-0.413	0.000	0.000	0.00	0.00	72.24
PHREEQE (1990)											
A' Caolinita	9.206	-0.854	-4.559	4.188	-0.636	0.000	-5.262	-7.990	0.00	0.00	0.00
B' Mos.-caol.	111.00	4.408	-3.668	8.368	-1.527	0.000	0.000	-0.950	0.00	55.49	0.00
C' Mos.-caol.	172.01	4.408	-3.478	8.356	1.717	0.000	0.000	-0.569	0.00	0.00	57.33
D' Microclina	223.99	5.000	-3.391	8.831	-2.001	-0.394	0.000	0.000	0.00	0.00	74.66

6. Consideraciones

Las diferencias en los resultados de la simulación de alteración de la microclina proporcionados por los cuatro códigos examinados (PHREEQE, EQ3/6, CHILLER y HARPHRQ) se deben fundamentalmente a los valores termodinámicos considerados por cada uno de ellos en su base de datos y, específicamente, a los correspondientes a las reacciones de equilibrio mineral intervinientes. Comparativamente, las diferencias existentes en las reacciones de especiación de los elementos involucrados (en especial para el aluminio) no provocan cambios de magnitud significativa en los resultados de la simulación (al menos para las condiciones iniciales elegidas en el cálculo).

Los códigos CHILLER y EQ3/6 son los que proporcionan resultados más parecidos dado que los datos termodinámicos que utilizan derivan de distintas versiones de SUPCRT. Pese a las diferencias individuales en los valores de las constantes de equilibrio de las reacciones de disolución mineral involucradas (tabla 1), el diagrama de estabilidad para las fases en equilibrio parcial correspondiente a ambos códigos es similar. Y puesto que la pauta evoluciona en un mismo espacio de fases, las diferencias más notables en esta pauta se producirán al principio, cuando menos restricciones de equilibrio afectan a la evolución del sistema.

Los resultados definidos por el código PHREEQE, tanto con la base de datos de la versión de 1990 como con los datos de Robie *et al.* (1978), son considerablemente distintos de los suministrados por los dos códigos anteriores, incluso en el número de minerales que intervienen en la simulación: la solución no llega a estar saturada respecto a la gibbsita, por lo que esta fase queda excluida de la descripción del proceso. Y las diferencias en los parámetros composicionales de la solución o en los valores de transferencia de masa tanto entre estos dos conjuntos de datos como respecto a los anteriores son fácilmente explicables si se observan las variaciones en los dominios de estabilidad que les corresponden en el espacio de fases del diagrama de estabilidad mineral (figura 2A).

La selección de un determinado conjunto de datos termodinámicos para las fases minerales de aparición factible en el proceso puede condicionar no sólo las magnitudes asociadas a las etapas que describen el proceso sino el propio número de etapas o de fases minerales realmente involucradas en la alteración de microclina. La variación de una de las constantes de equilibrio de los minerales intervinientes en el proceso es suficiente para producir notables cambios en los resultados ya que se propaga, de forma más o menos importante, al resto de los campos de estabilidad mineral sobre los que debe evolucionar la simulación.

Desde esta perspectiva, la selección de los datos termodinámicos para el planteamiento de cálculos de modelización adquiere más importancia que la del propio código cuando se trabaja sobre sistemas naturales. Las condiciones impuestas en la simulación constituyen una simplificación de la mineralogía relacionada con la evolución real del proceso; e incluso para la específicamente considerada (gibbsita, caolinita y moscovita) se ha hecho abstracción de problemas tales como la variación de solubilidad de gibbsita y caolinita con el grado de cristalinidad de su estructura y se ha utilizado la moscovita como polo composicional extremo representativo de las arcillas illíticas. Circunstancias que han restringido el rango de variabilidad de los datos termodinámicos considerados en la simulación pero que evidentemente pueden llegar a ser todavía más importantes en aproximaciones "más realistas" del proceso.

Las discrepancias existentes en las tabulaciones termodinámicas que afectan al sistema aluminosilicatado y las constatadas consecuencias que tienen sobre la modelización de un proceso tan idealizado como el empleado en este trabajo aconsejan la realización de análisis de

sensibilidad de los resultados frente a los parámetros termodinámicos utilizados en la modelización de procesos naturales desde aproximaciones de problema inverso.

Bibliografía

Basset, R.L. & Melchior, D.C. (1990): Chemical modeling of aqueous systems. An overview. In: D.C. Melchior & R.L. Basset (eds.), *Chemical Modeling of Aqueous Systems II*. A.C.S. Symp. Series, 416, 1-12.

Berman, R.G. (1988). Internally consistent thermodynamic data for minerals in system $\text{Na}_2\text{O}-\text{K}_2\text{O}-\text{CaO}-\text{MgO}-\text{FeO}-\text{Fe}_2\text{O}_3-\text{Al}_2\text{O}_3-\text{TiO}_2-\text{H}_2\text{O}-\text{CO}_2$. *Jour. Petrol.*, 29, 445-522.

Bond, K.A.; Moreton, A.D. & Heath, T.G. (1992): *The HATCHES User Manual*. Chemical Studies Department. Radwaste Disposal Division. Harwell Laboratory. Oxon. OX11 0RA, 20 pp.

Brown, P.L.; Haworth, A.; Sharland, S.M. & Tweed, C.J. (1991): *HARPHRO. A geochemical program based on PHREEQE. User's Guide*. Theoretical Studies Department. Radwaste Disposal Division. Harwell Laboratory. Didcot. Oxon. OX11 0RA. 18 pp.

Drever, J.J. (1988): *The geochemistry of natural waters*. Prentice-Hall Inc., N.J., 437 pp.

Engi, M. (1992): Thermodynamic data for minerals: a critical assesment. In: *The Stability of Minerals* (G.D. Price & N.L. Ross, ed.). Chapman & Hall, London. Chapter Eight, 267-328.

Faure, G. (1992): *Principles and Applications of Inorganic Geochemistry*. MacMillan Pub. Co., 626 pp.

INTERA Environmental Consultants Inc. (1983): *Geochemical models suitable for performance assessment of nuclear waste storage: comparison of PHREEQE and EQ3/EQ6*. ONWI-473. Battelle Memorial Institute, Office of Nuclear Waste Isolation, Columbus, Ohio.

Jenne, E.A. (1981): *Geochemical modeling: A review*. Waste/rock Interactions Technology Program. Pacific Northwest Laboratory, PNL-3574, UC-70, 47 pp.

Johnson, J.W., Oelkers, E.H. & Helgeson, H.C. (1991): *SUPCRT92, a software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species and reactions from 1 to 5000 bars and 0° to 1000 °C*. User's manual, 101 pp.

Helgeson, H.C. (1984): *SUPCRT. Unpublished thermodynamic data base for minerals, aqueous species and gases*. Univ. of California, Berkeley, California.

Nordstrom D.K. & Ball, J.W. (1989): Mineral saturation states in natural waters and their sensitivity to thermodynamic and analytic errors. *Sci. Geol. Bull.*, 42, 269-280.

Nordstrom, D.K., Plummer, L.N., Langmuir, L., Busenberg, E., May, H.M., Jones, B.F. & Parkhurst, D.L. (1990). Revised chemical equilibrium data for major water-mineral reactions and their limitations. In: D.C. Melchior & R.L. Basset (eds.), *Chemical Modeling of Aqueous Systems II*. A.C.S. Symp. Series, 416, 398-413.

Parkhurst, D.L., Thorstenson, D.C. & Plummer, L.N. (1980): PHREEQE, a computer program for geochemical calculations. *U.S. Geol. Surv. Water Res. Inv.*, 80-96, 193pp. November, 1980.

Parkhurst, D.L., Thorstenson, D.C. & Plummer, L.N. (1985): PHREEQE, a computer program for geochemical calculations. *U.S. Geol. Surv. Water Res. Inv.*, 80-96, 193pp. January, 1985.

Parkhurst, D.L., Thorstenson, D.C. & Plummer, L.N. (1990): PHREEQE, a computer program for geochemical calculations. *U.S. Geol. Surv. Water Res. Inv.*, 80-96, 193pp. August, 1990.

Reed, M.H. (1982): Calculation of multicomponent chemical equilibria and reaction processes in systems involving minerals, gases and aqueous phase. *Geochim. Cosmochim. Acta*, 46, 513-528.

Reed, M.H. & Spycher, N.F. (1987): *SOLTHERM data base. A compilation of thermodynamic data from 25 °C to 300 °C for aqueous species, minerals and gases*. Eugene, Oregon, Dept. Geol. Sci., Univ. Oregon, 30 pp.

Reed, M.H. & Spycher, N.F. (1989): *SOLTHERM: Data base of equilibrium constants for aqueous-mineral-gas equilibria*. Dept. Geol. Sci., Univ. Oregon. Eugene, Oregon. 45 pp.

Richardson, S.M. & McSweeney Jr., H.Y. (1989): *Geochemistry. Pathways and Processes*. Prentice Hall, Inc., 488 pp.

Robie, R.A., Hemingway, B.S. & Fisher, J.R. (1978): Thermodynamic properties of minerals and related substances at 298.15 °K and 1 bar (10^5 pascals) pressure and at higher temperatures. *Geol. Surv. Bull.*, 1452, 456 pp.

Silling, S.A. (1983): *Final technical position on documentation of computer codes for high-level waste management*. U.S. Nuclear Regulatory, Washington D.C., NUREG-0856, 12 pp.

Spycher, N.F. & Reed, N.H. (1989): *CHILLER: A program for computing water-rock reactions, boiling, mixing and other reaction processes in aqueous-mineral-gas systems*. Dept. of Geological Sciences. University of Oregon. Eugene, Oregon, 63 pp.

Spycher, N.F. & Reed, N.H. (1992): Microcomputer-based modeling of speciation and water-mineral-gas reactions using programs SOLVEQ and CHILLER. In: Y.K. Kharaka and A.M. Maest (eds.), *Water-Rock Interaction*, 1087-1090. Balkema, Rotterdam.

Tena, J.M., Auqué, L.F., Gimeno, M.J., Fernández, J. y Mandado, J. (1990). *Estudio comparativo de software para solución de equilibrios fisicoquímicos*. CIEMAT, Madrid (no publicado), 4 vols.

Tsang, C.F. (1991): The modeling process and model validation. *Ground Water*, 29, 825-831.

Wolery, T.J. (1979): *Calculations of Chemical Equilibrium between aqueous solution and minerals: the EQ3/6 software Package*. NTIS UCRL-52658.

Wolery, T.J. & Daveler, S.A. (1992). *EQ6, a computer program for reaction path modeling of aqueous geochemical systems: theoretical manual, user's guide and related documentation (version 7.0)*. Lawrence Livermore National Laboratory. UCRL-MA-110662 PT IV.

Análisis comparativo de códigos de modelización geoquímica y bases de datos termodinámicos. II. Simulación de procesos de dedolomitización.

Auqué, L.F.⁽¹⁾; López, P.L.⁽¹⁾; Gimeno, M.J.⁽²⁾; Mandado, J.⁽¹⁾; Vallès, V.⁽³⁾; Zamora, A.⁽¹⁾ y Gómez, J.⁽¹⁾

(1) Área de Petrología y Geoquímica. Depto. Ciencias de la Tierra. Fac. Ciencias. Universidad de Zaragoza. 50009 ZARAGOZA (España).

(2) Unidad de Seguridad de Emplazamientos. Edificio 20 T. CIEMAT. Avda. Complutense 22. 28006 MADRID (España).

(3) Laboratoire du Science du Sol d'Avignon, Institut National de la Recherche Agronomique. Domaine St. Paul, Site AGROPARC, 84914-AVIGNON. CEDEX 9 (France).

Abstract

Simulation of a water-rock interaction process, mainly constrained by dedolomitization conditions (dolomite dissolution and calcite precipitation driven irreversibly by anhydrite dissolution), is performed with the more recent versions of EQ3/6, PHREEQE, CHILLER and HARPHRQ geochemical codes.

Results are almost insensitive to differences in the involved speciation reactions. But order-disorder in dolomite has a very significant effect on the calculated reaction-path. Differences in the equilibrium constant of ordered-disordered dolomite lead to important discrepancies even in the resulting conceptual model: opposite paths (dolomitization or dedolomitization) can be defined depending on the ordering of the selected dolomite for calculations.

Effects of structural order-disorder in dolomite can be considered analogous to those related with degree of crystallinity for other minerals; and they must be taken into account in sensitivity analysis of reaction path modeling.

1. Introducción

En el análisis comparativo de los códigos de modelización termodinámica PHREEQE, EQ3/6, CHILLER y HARPHRQ realizado por Auqué *et al.* (1995) se señalaba que las distintas constantes de equilibrio para las reacciones de disolución o precipitación de fases aluminicas o aluminosilicatadas eran las casi únicas responsables de las diferencias suministradas en la simulación de un proceso de alteración de microclina. Por tanto, el análisis comparativo se reducía a una verificación de la sensibilidad del modelo geoquímico obtenido en la simulación frente a los datos termodinámicos específicamente manejados en el sistema Si-Al-K-H₂O.

Otro sistema composicional básico en la modelización de procesos de interacción agua-roca es el Ca-Mg-SO₄-CO₃-HCO₃-H₂O, responsable en muchas ocasiones del control de un parámetro tan importante como el pH y para el que, lógicamente, se incluyen tabulaciones

termodinámicas más o menos amplias en todos los códigos ya utilizados en la simulación del proceso de alteración de microclina.

Para completar el estudio comparativo de los códigos anteriormente mencionados, en este artículo se analizan los resultados que proporcionan en la simulación de un proceso de interacción agua-roca desarrollado en ese segundo sistema composicional. Análisis que, en definitiva, podrá reducirse a una valoración de los efectos producidos por las diferencias en los datos termodinámicos considerados por los distintos programas y, por tanto, en un análisis de sensibilidad (Nordstrom y Ball, 1989) del modelo geoquímico obtenido en la simulación. Sin embargo, los resultados de este nuevo análisis pondrán de manifiesto un aspecto adicional no explícitamente evidenciado en el anterior trabajo (Auqué *et al.*, 1995): la importancia del grado de ordenación de la estructura de una de las fases involucradas (y su reflejo en el valor de su constante de equilibrio) sobre los resultados de la simulación.

2. Antecedentes y planteamiento del estudio

Normalmente, en los manuales de usuario de los códigos de modelización suelen incluirse distintos ejemplos de cálculo desarrollados específicamente en el sistema carbonatado (p. ej. Parkhurst *et al.*, 1990; Wolery y Daveler, 1992) como demostración de la capacidad de trabajo del programa en este sistema; y en los estudios de comparación de códigos termodinámicos se ha considerado casi siempre la simulación de un proceso de interacción agua-roca desarrollado específicamente en el sistema $\text{Ca-Mg-SO}_4\text{-CO}_3\text{-HCO}_3\text{-H}_2\text{O}$. Pese a presentar distintos matices en las condiciones concretas, casi todas las simulaciones se desarrollan dentro de los esquemas normalmente adscritos a la modelización de procesos de dedolomitización.

El fenómeno de dedolomitización, efectivo en acuíferos carbonatados, consiste en un proceso conjugado de disolución de dolomita y precipitación de calcita desencadenado por la disolución de anhidrita. Y como tal ha sido ampliamente estudiado en distintos sistemas naturales (Back *et al.*, 1983; Busby *et al.*, 1983, 1990; Plummer *et al.*, 1990; Bischoff *et al.*, 1994; etc). Metodológicamente, y tanto desde aproximaciones de problema inverso como directo (Plummer, 1984), la modelización de este proceso puede ser termodinámicamente descrita considerando la disolución de anhidrita como un proceso irreversible limitado por sucesivos estadios de equilibrio parcial de la solución respecto a calcita y dolomita.

Se trata de un proceso típicamente caracterizable mediante un cálculo de pautas de reacción y, por tanto, resulta explicable que haya sido tomado preferentemente como simulación de referencia en análisis comparativos de códigos. De esta forma, INTERA (1983) y Tena *et al.* (1990) analizan los resultados de EQ3/6 y PHREEQE en la simulación explícita de un proceso de dedolomitización desarrollado en campo variable de temperatura; y Broyd *et al.* (1985) consideran la simulación a 25 °C de una pauta de evolución algo más compleja pero delimitada, implícitamente, por las condiciones adscritas a un proceso de este tipo.

Estos trabajos previos han constatado genéricamente la importancia de los datos termodinámicos manejados por los distintos códigos en los resultados del proceso simulado. Sin embargo, una valoración más atenta de esos estudios comparativos pone de manifiesto que las diferencias en las bases de datos de los códigos pueden llegar a provocar que la simulación se resuelva en un proceso opuesto al de dedolomitización. En el estudio de Broyd *et al.* (1985), y pese a considerarse implícitamente la simulación de un proceso de dedolomitización como referencia de comparación, los resultados obtenidos con el código EQ3/6 corresponden formalmente a un proceso simple de precipitación de dolomita (dolomitización). Sin embargo,

en el trabajo de INTERA (1983), las pautas obtenidas por ese mismo código responden a las esperables de un proceso de dedolomitización, si bien la simulación en este caso tiene lugar en condiciones no isotermas. Ambas situaciones fueron verificadas por Tena *et al.* (1990).

Estos resultados parecen sugerir la existencia de unas muy importantes discrepancias en la tabulación termodinámica del sistema en cuestión; discrepancias cualitativamente más importantes incluso que las asignadas al sistema aluminosilicatado (Drever, 1988; Auqué *et al.*, 1995). Y que, por tanto, hacen necesario un estudio más detallado sobre la incidencia de las bases de datos en la simulación de este tipo de procesos.

El planteamiento del análisis comparativo que sigue a continuación responde a este propósito, y utilizó un esquema similar al empleado por Auqué *et al.* (1995). Inicialmente, se planteó la simulación del proceso de interacción agua-roca considerado por Broyd *et al.* (1985) con cada uno de los códigos comparados; y posteriormente se volvió a repetir esta simulación con un único código pero cambiando parcialmente su base de datos, sustituyendo las constantes de los equilibrios heterogéneos involucrados por las existentes en las del resto de códigos.

Para el caso simulado, este mínimo cambio en la base de datos resulta suficiente para conseguir una valoración precisa sobre la incidencia específica tanto de las distintas tabulaciones en las reacciones de especiación como en las de equilibrio mineral. Y además permite una adecuada delimitación del aspecto más destacado observado en este análisis comparativo: el grado de ordenación de la estructura mineral de la dolomita (una de las fases involucradas en la simulación) como principal fuente de incertidumbre en los cálculos.

3. Metodología de simulación: códigos seleccionados y bases de datos termodinámicos.

El ejemplo escogido para la realización de este trabajo corresponde a la simulación definida por Broyd *et al.* (1985) en su estudio comparativo de códigos de modelización. Estos autores consideran un proceso teórico de interacción agua-roca en un medio carbonatado-sulfatado, a temperatura constante (25 °C) y desarrollado en sistema cerrado respecto al CO₂, que puede esquematizarse en los siguientes pasos:

(a) Se parte de una solución acuosa inicial con las siguientes características composicionales (concentraciones expresadas en ppm): Ca²⁺ = 82.2; Mg²⁺ = 55.4; HCO₃⁻ = 165.0; K⁺ = 129.0; Na⁺ = 3470.0; SO₄²⁻ = 282.0; Cl⁻ = 5360.0; F⁻ = 1.6; SiO₂ = 1.09; pH = 7.05; T = 25 °C.

(b) La solución reacciona con dolomita, disolviéndola hasta llegar al equilibrio.

(c) Esta solución en equilibrio con dolomita disuelve posteriormente CaSO₄ (anhidrita), en cantidades acumulativas de 1 mmol hasta un total de 6 mmoles. Durante las adiciones progresivas de CaSO₄ se impone que la solución esté en equilibrio con dolomita, precipitando calcita cuando se sobrepase su producto de solubilidad.

Se trata, por tanto, de una simulación con dos partes diferenciadas: en la primera el proceso irreversible está representado por la disolución de dolomita, y la variable de avance de reacción estará delimitada por la propia transferencia de masa de esa fase mineral (moles de dolomita disuelta) hasta que se alcance una situación de equilibrio; en la segunda parte, el proceso irreversible y la variable de progreso de reacción estarán definidas por la adición de CaSO₄ a la solución hasta el total indicado (6 mmoles), manteniéndose de forma permanente

Tabla 1: Valores de las constantes de equilibrio (log K, a 25 °C) para las reacciones de disolución de calcita y dolomita, incluidas en los códigos comparados en este trabajo, y estado de saturación de la solución inicial respecto a esas fases carbonatada calculados por cada uno de ellos.

	PHREEQE	HARPHRQ HATCHES	CHILLER SOLThERM	EQ3/6 SUPCRT92	EQ3/6 DATA.COM
Calcita	-8.48	-8.47	-8.60	-8.48	-8.48
Dolomita	-17.09	-17.02	-18.14	-18.14	-18.14
Dolomita desord.	-16.54*	-16.60	-16.59	-16.60	-16.60
Dolomita ordenada	---	-18.14	-18.14	-18.14	-18.14
Estados de saturación de la solución inicial (I.S.)					
I.S. calcita	-0.694	-0.723	-0.670	-0.684	-0.719
I.S. dolomita media	-1.192	-1.319	---	---	---
I.S. dolomita desor.	-1.742	-1.731	-1.940	-1.621	-1.722
I.S. dolomita ord.	---	-0.200	-0.400	-0.077	-0.177

* Valor no incluido en la base de datos del código, pero tomado de la revisión de datos termodinámicos (Nordstrom *et al.*, 1990) que sirvió para construirla.

Tabla 2: Resultados de la simulación obtenidos mediante el código HARPHRQ, con los distintos valores termodinámicos para la constante de equilibrio de la dolomita incluida en su base de datos (HATCHES). Los valores de Ca_T , Mg_T y alcalinidad están expresados en mmoles/kg de agua.

		Disolución de dolomita	Adición de CaSO ₄ (mmoles)						
			0.7	1.0	2.0	3.0	4.0	5.0	6.0
HARPHRQ (dolomita media)	pH	7.615	7.563	7.544	7.487	7.439	7.398	7.362	7.330
	Ca _T	2.208	2.917	3.078	3.617	4.151	4.700	5.240	5.787
	Mg _T	2.440	2.415	2.547	2.989	3.430	3.872	4.314	4.756
	Alc. _T	3.282	3.251	3.238	3.198	3.162	3.130	3.100	3.073
	Dolomita*	-0.138	-0.114	-0.246	-0.687	-1.129	-1.571	-2.013	-2.455
	Calcita*	(-)™	(-)	0.238	1.141	2.042	2.942	3.840	4.738
HARPHRQ (dolomita desord.)	pH	7.835	7.785	7.765	7.707	7.658	7.616	7.579	7.546
	Ca _T	1.484	1.701	1.794	2.106	2.419	2.733	3.048	3.364
	Mg _T	3.234	3.703	3.904	4.576	5.249	5.922	6.595	7.269
	Alc. _T	3.422	3.394	3.383	3.351	3.323	3.297	3.274	3.253
	Dolomita*	-0.932	-1.402	-1.603	-2.275	-2.947	-3.621	-4.294	-4.968
	Calcita*	1.520	2.471	2.879	4.239	5.599	6.958	8.317	9.675
HARPHRQ (dolomita ord.)	pH	7.129	7.083	7.067	7.025	6.994	6.969	6.949	6.932
	Ca _T	2.096	2.780	3.077	4.060	5.050	6.044	7.038	8.032
	Mg _T	2.327	2.310	2.308	2.294	2.284	2.275	2.268	2.262
	Alc. _T	2.834	2.778	2.758	2.703	2.660	2.626	2.598	2.575
	Dolomita*	-0.026	-0.012	-0.007	0.007	0.017	0.026	0.033	0.039
	Calcita*	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)

* Masa mineral acumulada durante la simulación, expresada en mmoles de mineral disuelto (con valores negativos) o precipitado (valores positivos) por kg de agua.

** (-) significa que el mineral se encuentra subsaturado en la correspondiente etapa de la simulación.

situaciones de equilibrio parcial con dolomita (lo que implica que esta fase se disuelva o precipite) y con calcita sólo cuando la solución alcance la saturación respecto a esa fase (lo que condiciona que las transferencias de masa sean sólo de precipitación). La situación teórica planteada en la simulación podría asimilarse a la pauta de evolución de una solución en un acuífero de naturaleza exclusivamente dolomítico-sulfatada.

Los códigos seleccionados para este análisis comparativo han sido los mismos que ya fueron utilizados en el tratamiento de la alteración de la microclina (Auqué *et al.*, 1995): el EQ3/6 (versión 7.2a; Wolery y Daveler, 1992), CHILLER (Reed, 1982; Spycher y Reed, 1989, 1992), PHREEQE (revisión de 1990; Parkhurst *et al.*, 1990) y HARPHRQ (Brown *et al.*, 1991). Todos ellos son capaces de plantear la simulación del proceso descrito a través de sucesivos estadios de Equilibrio Parcial (cálculos de pautas de reacción). Un resumen de sus características más destacadas puede verse en Auqué *et al.* (1995) y con más detalle en Tena *et al.* (1990).

Operativamente, merece la pena destacar dos circunstancias en el planteamiento de la simulación con estos códigos. Por un lado, la estructura de los códigos PHREEQE y HARPHRQ permite enlazar ambas partes de la simulación en una única entrada de datos. Sin embargo, existen problemas con el algoritmo utilizado por HARPHRQ en esta modalidad de entrada (se producen algunas pequeñas modificaciones en los valores de alcalinidad y balance de cargas de la solución al pasar de una parte a otra), motivo por el que se prefirió trabajar planteando dos entradas sucesivas e independientes, una para cada parte de la simulación. Por otro lado, las diferencias en el tratamiento de la alcalinidad que presenta el código EQ3/6 ha hecho necesario introducir el dato de carbono total calculado por los otros códigos (muy similar en todos los casos) en vez del valor de HCO_3^- considerado en el análisis de la solución inicial.

Las bases de datos termodinámicos utilizadas con EQ3/6 han sido SUPCRT92 (Johnson *et al.*, 1991) y DATA.COM (Wolery y Daveler, 1992); el CHILLER utiliza la base de datos SOLTHERM (Reed y Spycher, 1987, 1989) en gran parte derivada de la versión de SUPCRT84 (Helgeson, 1984); la del PHREEQE procede fundamentalmente de las revisiones de Nordstrom *et al.* (1990) y de Tisarrani y Glynn (1990; en Parkhurst *et al.*, 1990); y la de HARPHRQ hace uso de la base de datos HATCHES (Bond *et al.*, 1992) del Laboratorio Harwell (Gran Bretaña).

Los valores de las constantes de equilibrio para calcita y dolomita a 25 °C incluidas en estas bases de datos se indican en la tabla 1. Y como puede apreciarse, los correspondientes a la calcita son prácticamente iguales para todos los códigos y bases de datos manejadas. El valor dominante de $\log K = -8.48$ procede de los resultados experimentales de Plummer y Busenberg (1982). Y el valor de $\log K = -8.60$, incluido en SOLTHERM, de las propuestas de Helgeson *et al.* (1978) y SUPCRT84 (Helgeson, 1984).

Mayores diferencias existen en los valores incluidos para la dolomita en los distintos códigos, e incluso en la denominación específica de la fase en cuestión. Así en PHREEQE (tabla 1) se incluyen dos valores, uno para "dolomita" y otro para dolomita desordenada; HARPHRQ presenta tres valores distintos, para "dolomita", dolomita ordenada y dolomita desordenada; SOLTHERM, SUPCRT92 y DATA.COM incluyen también esas tres denominaciones pero sólo dos valores de constante de equilibrio diferentes (los valores considerados para la "dolomita" son iguales a los de la dolomita ordenada; tabla 1). En las diferencias del valor de $\log K$ para la dolomita se incluye, por tanto, un aspecto adicional que conviene tener en cuenta: el grado de ordenación de la estructura al que se asocia ese valor.

Helgeson *et al.* (1978) demostraron que la solubilidad de la dolomita varía significativamente en función del grado de ordenación de las posiciones estructurales en la red

mineral; y tabularon las constantes de equilibrio para dos dolomitas estequiométricas, una totalmente ordenada y otra totalmente desordenada. Desde entonces, estos calibrados se incluyeron en las bases de datos de distintos códigos, en los que con el término "dolomita" se volvían a considerar los datos de la dolomita estequiométrica ordenada. Sin embargo, en otros códigos se incluyen bajo el mismo término de "dolomita" (sin referencia explícita al grado de ordenación de su estructura) los datos experimentales propuestos por Robie *et al.* (1978) con valores en torno a $\log K = -17.00$, a 25 °C.

Normalmente, en los ejercicios de comparación de códigos se seleccionaba la fase con una denominación común en sus bases de datos, la "dolomita". Y esta circunstancia implicaba considerar variaciones de más de un orden de magnitud en la constante de equilibrio de esa fase mineral, por lo que en general se aludía siempre a una importante sensibilidad de los resultados respecto a la base de datos utilizada sin considerar que, en realidad, se podían estar comparando fases de caracteres distintos y, por tanto, no estrictamente comparables.

En este estudio se analizará en detalle la incidencia de la ordenación de la dolomita considerada en la simulación, manteniéndose los términos de ordenada y desordenada, tal como fueron definidos originalmente por Helgeson *et al.* (1978). Los valores experimentales para la dolomita propuestos por Robie *et al.* (1978) corresponden a una dolomita estequiométrica pero para la que no se determinó el grado de ordenación de su estructura. Sin embargo, por el valor de su constante de equilibrio puede deducirse que se trataba de una dolomita con un cierto grado de desorden (Carpenter, 1980) y será denominada como dolomita media.

4. Análisis de resultados: incidencia de la ordenación de la dolomita.

4.1. Estados de saturación de la solución inicial y transferencias de masa.

Los caracteres fisicoquímicos y estados de saturación definidos para la solución problema por los distintos códigos condicionarán el posterior desarrollo de los procesos heterogéneos involucrados en la simulación. En la tabla 1 se señalan los estados de saturación (expresados como índices de saturación, $I.S. = \log (PAI/K(T))$, donde PAI es el producto de actividad iónica y $K(T)$ la constante de equilibrio a la temperatura de interés) definidos respecto a la calcita y a las distintas dolomitas en el cálculo previo de especiación-solubilidad realizado por los programas comparados. Como puede apreciarse, la solución de partida se encuentra subsaturada respecto a calcita sea cual sea el código considerado y con valores de I.S. muy similares ($I.S. = -0.6985 \pm 0.0245$). Y también se encuentra siempre subsaturada respecto a cualquiera de las dolomitas consideradas, aunque en distinto grado según la ordenación de esa fase mineral: los mayores estados de subsaturación se alcanzan respecto a la dolomita desordenada ($I.S. = -1.815 \pm 0.125$), siendo importantes respecto a la dolomita de orden medio ($I.S. = -1.2555 \pm 0.063$) y mínimos (prácticamente caracterizando situaciones próximas al equilibrio) en el caso de la dolomita ordenada ($I.S. = -0.238 \pm 0.16$).

Estos resultados ponen de manifiesto varios hechos de interés. Por un lado, avanzan que las posibles diferencias en las reacciones de especiación involucradas en el cálculo y/o en el modelo químico utilizado por cada código son poco importantes: los valores de índice de saturación para las distintas fases consideradas son prácticamente coincidentes, dado que los valores de sus constantes de equilibrio también lo son. Y, por otro, que es el tipo de dolomita considerada el responsable de las mayores diferencias en la caracterización inicial de la solución, independientemente del código manejado.

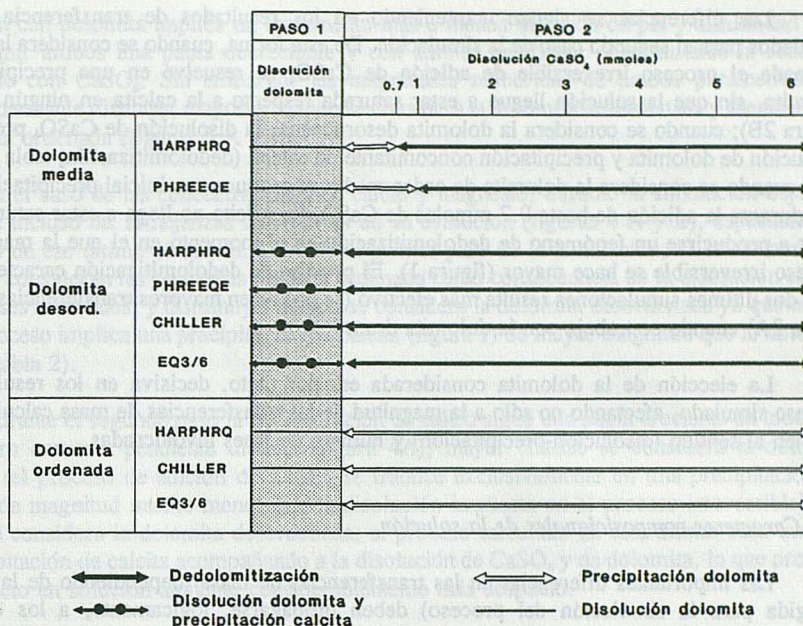


Figura 1.- Comparación de los procesos efectivos deducidos para las dos partes de la simulación por los códigos EQ3/6, PHREEQE, CHILLER y HARPHRQ. Los resultados se agrupan en función del grado de ordenación de la dolomita utilizada en los cálculos.

En todos los casos, la solución inicial podrá disolver dolomita y verificar el primer paso de la simulación. Sin embargo, resulta evidente que el tipo de dolomita escogida condicionará el valor de transferencia de masa (dolomita disuelta) obtenido en ese primer paso de la simulación y, por tanto, también las condiciones iniciales de desarrollo del segundo paso. De hecho, y en una primera valoración cualitativa (figura 1), los resultados proporcionados por los distintos códigos para todo el conjunto de la simulación responden a pautas idénticas siempre que se consideren dolomitas de similar ordenación.

Esta circunstancia evidencia ya *a priori* la particular importancia de esta fase en el proceso simulado. Se puede obtener una valoración general de los efectos producidos por la utilización de los distintos tipos de dolomita analizando los resultados proporcionados por HARPHRQ, código que incluye en su base de datos los tres tipos más frecuentes de dolomita. En la tabla 2 se presentan los resultados numéricos de las simulaciones obtenidos con cada una de las dolomitas de ese código.

Al considerar la dolomita media o la ordenada, los resultados de transferencia de masa involucrados en el primer paso de la simulación hacen referencia exclusivamente a la disolución de esa fase mineral (mayor en el caso de la dolomita media que en el de la ordenada; 0.138 y 0.026 mmoles/kg de solución, respectivamente; tabla 2); sin embargo, al considerar la dolomita desordenada, la solución alcanza la saturación respecto a la calcita durante el desarrollo del proceso irreversible (disolución de dolomita) y la simulación del primer paso se resuelve en la disolución de dolomita desordenada y subsiguiente precipitación de calcita (figura 1).

Las diferencias se siguen manteniendo en los resultados de transferencia de masa calculados para el segundo paso de la simulación. De esta forma, cuando se considera la dolomita ordenada el proceso irreversible de adición de CaSO_4 se resuelve en una precipitación de dolomita, sin que la solución llegue a estar saturada respecto a la calcita en ningún momento (figura 2B); cuando se considera la dolomita desordenada, la disolución de CaSO_4 provoca una disolución de dolomita y precipitación concomitante de calcita (dedolomitización; tabla 2 y figura 1); y cuando se considera la dolomita de orden medio se produce una inicial precipitación de esa fase durante la adición de hasta 0.7 mmoles de CaSO_4 (la calcita no llega a estar saturada) para pasar a producirse un fenómeno de dedolomitización en el momento en el que la magnitud del proceso irreversible se hace mayor (figura 1). El proceso de dedolomitización caracterizado en estas dos últimas simulaciones resulta más efectivo (se producen mayores transferencias de masa; figura 2A) cuando se trabaja con la dolomita desordenada.

La elección de la dolomita considerada es, por tanto, decisiva en los resultados del proceso simulado, afectando no sólo a la magnitud de las transferencias de masa calculadas sino también al sentido (disolución-precipitación) y número de fases involucradas.

4.2. Caracteres composicionales de la solución.

Las importantes diferencias en las transferencias de masa (dependiendo de la dolomita escogida para la simulación del proceso) deben propagarse, lógicamente, a los caracteres composicionales de la solución resultante en cada punto discreto de avance del progreso de la reacción. Los resultados obtenidos con HARPHRQ indican que, cualitativamente, las pautas de pH y alcalinidad son similares, independientemente de la dolomita considerada (figuras 3 A y B):

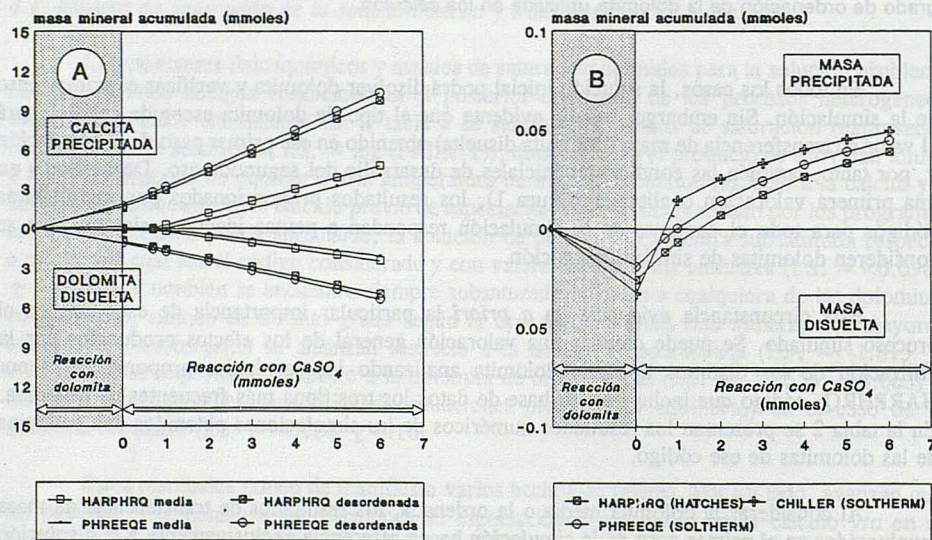


Figura 2.- Comparación de las transferencias de masa definidas respecto a las fases carbonatadas involucradas la simulación. A.- Masas de calcita precipitada y de dolomita disuelta calculadas por distintos códigos, considerando la dolomita desordenada y la de ordenación media de sus bases de datos. B.- Masa de dolomita precipitada o disuelta calculada por distintos códigos pero considerando la intervención de una dolomita ordenada. Obsérvese que en este caso no existen transferencias de masa respecto a la calcita en ninguna de las dos partes de la simulación.

la reacción con dolomita implica un incremento más o menos marcado de pH y alcalinidad para luego definir ambos una pauta decreciente y con similar pendiente al producirse la reacción irreversible con CaSO_4 . Sin embargo, las magnitudes numéricas de ambos parámetros son considerablemente distintas, mayores cuando se considera la dolomita desordenada y menores con la dolomita ordenada (figuras 3 A y B).

En el caso de las concentraciones de calcio y magnesio, durante la simulación dejan de apreciarse incluso las semejanzas cualitativas en su evolución (figuras 4 A y B), especialmente en el caso de ese último elemento. Durante el primer paso de la simulación, el calcio aumenta cuando se consideran las dolomitas media y ordenada como consecuencia de la disolución simple de esas fases minerales; y disminuye cuando se considera la dolomita desordenada ya que en ese caso el proceso implica una precipitación de calcita (figura 1) de mayor magnitud que la dolomita disuelta (tabla 2).

Durante el segundo paso la concentración de calcio sigue una pauta creciente en todos los casos, pero con una pendiente distinta (figura 4A), mayor cuando se considera la dolomita ordenada (el proceso de adición de CaSO_4 se traduce exclusivamente en una precipitación de dolomita de magnitud mucho menor que la disolución impuesta en el proceso irreversible) que cuando se considera la dolomita desordenada; el proceso calculado en este último caso implica una precipitación de calcita acompañando a la disolución de CaSO_4 y de dolomita, lo que provoca que el calcio en solución aumente considerablemente más despacio.

Estos efectos de las diferencias en las transferencias de masa son los responsables de las importantes variaciones definidas en la evolución del magnesio (figura 4B). Durante el primer paso de la simulación, sólo se produce un incremento notable en la concentración de magnesio

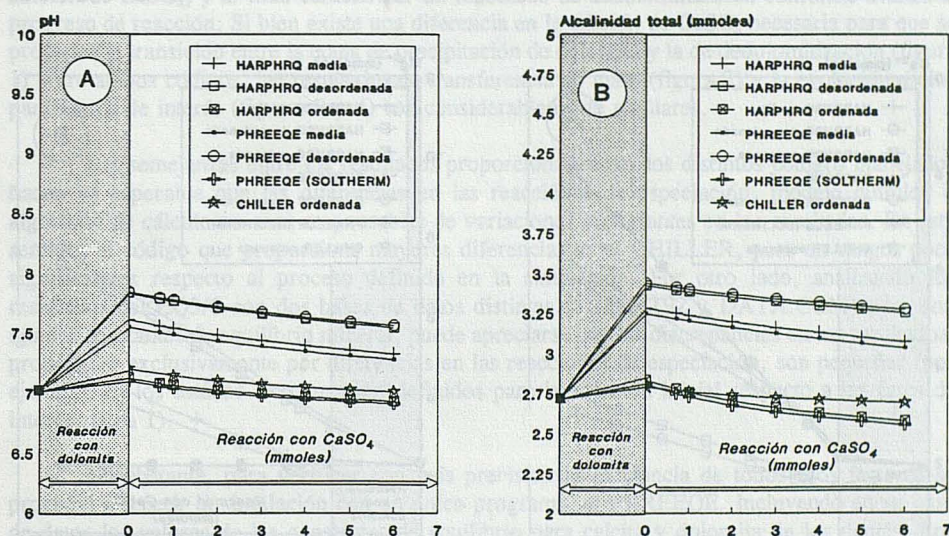


Figura 3.- Comparación de la evolución del pH (A) y de la alcalinidad total (B) de la solución a lo largo del proceso simulado. Se incluyen los resultados proporcionados por HARPHRQ (considerando los tres grados de ordenación para la estructura de la dolomita que aparecen en su base de datos), CHILLER (considerando la dolomita ordenada de su base de datos) y PHREEQE (considerando, por un lado, la dolomita desordenada y de orden medio de su base de datos y, por otro, las constantes de equilibrio de calcita y dolomita ordenada de la base de datos SOLTHERM-CHILLER).

cuando se utiliza la dolomita desordenada; mientras que con la dolomita ordenada o con la media, el aumento de magnesio es muy pequeño. En el primer caso, el notable incremento de magnesio está relacionado tanto con la importante subsaturación de la solución inicial respecto a la dolomita desordenada (tabla 1), como con la existencia de un proceso adicional de precipitación de calcita (figura 1; tabla 2); proceso que, a su vez, provoca una mayor disolución de dolomita. En los otros dos casos, el aumento en la concentración de magnesio es proporcional al estado de subsaturación de la solución inicial respecto a cada una de las dolomitas (mayor en el caso de la dolomita media ya que respecto a la dolomita ordenada la solución se encuentra casi en equilibrio).

Las diferencias más notables en la evolución de este elemento tienen lugar en la segunda parte de la simulación, llegándose a producir pautas de pendiente de signo opuesto: con la dolomita desordenada y media se produce un incremento en la concentración de magnesio, de mayor pendiente para la primera que para la segunda; pero al considerar la dolomita ordenada la pauta de este elemento viene definida por una progresiva disminución conforme avanza (figura 4B) el proceso irreversible de disolución de CaSO_4 , puesto que, en este caso, la simulación predice un proceso de precipitación (y no de disolución como en los anteriores) de esa dolomita.

5. Bases de datos termodinámicos y verificación de resultados.

Al valorar los resultados referidos en el epígrafe anterior se han utilizado los obtenidos con un determinado código (HARPHRQ) y, por tanto, las importantes diferencias definidas estaban únicamente condicionadas por los valores de $\log K$ de la dolomita escogida (el modelo químico y los valores de las constantes de equilibrio del resto de reacciones involucradas eran las mismas en las tres simulaciones consideradas dado que se empleaba un mismo código). Al

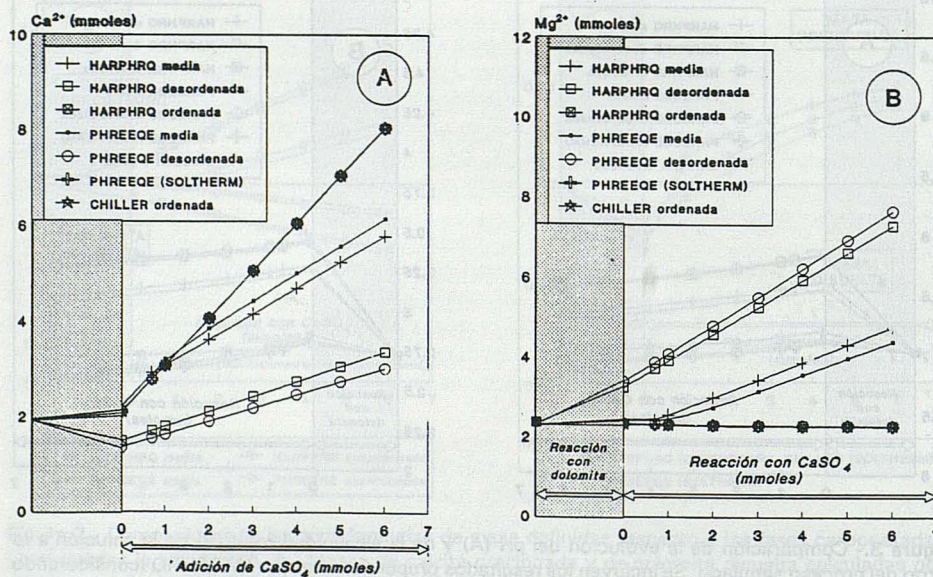


Figura 4.- Comparación de las concentraciones de calcio (A) y magnesio (B) de la solución a lo largo del proceso simulado. Las pautas de evolución indicadas corresponden a los mismos códigos y datos termodinámicos empleados en la figura 3.

utilizar distintos códigos podrían producirse variaciones adicionales debidas a sus particulares algoritmos, modelos químicos o valores de las reacciones de especiación involucradas.

Pero como ya se ha comentado, de la comparación directa de los resultados obtenidos por los distintos códigos manejados puede evidenciarse que los procesos globalmente definidos por PHREEQE, HARPHRQ, CHILLER y EQ3/6 son similares (figura 1) siempre que en las simulaciones comparadas utilicen dolomitas de ordenación análoga. Cuando se considera una dolomita ordenada, todos los códigos que incluyen una fase de este tipo predicen su disolución inicial en el primer paso y una precipitación en el segundo. Las transferencias de masa puestas en juego son muy similares: las de HARPHRQ y EQ3/6 (con la base de datos SUPCRT92 y DATA.COM) son prácticamente coincidentes (por ello en la figura 2B sólo se incluyen las de HARPHRQ) y algo más diferenciada la de CHILLER. E idéntica situación puede señalarse para la evolución de los parámetros de interés (figuras 3 y 4), siempre con variaciones ligeramente mayores para CHILLER.

Los resultados vuelven a ser similares si comparamos los que utilizan una dolomita desordenada (figura 1). Todos los códigos resuelven ambos pasos de la simulación en procesos de precipitación de calcita y disolución de dolomita. Los valores de transferencia de masa son muy similares para todos los códigos, lo mismo que la evolución de pH, alcalinidad, calcio y magnesio (figuras 2A, 3 y 4; en estas figuras sólo se comparan los resultados obtenidos por PHREEQE y HARPHRQ).

Y, por último, para los códigos que incluyen una dolomita media en su base de datos (PHREEQE y HARPHRQ) los resultados proporcionados definen un proceso de disolución inicial de dolomita para seguir con un proceso de precipitación de esa misma fase (cuando comienza la adición de CaSO_4) y al final caracterizar un fenómeno de dedolomitización conforme avanza el progreso de reacción. Si bien existe una diferencia en la adición de CaSO_4 necesaria para que se produzca la transición entre la etapa de precipitación de dolomita y la de dedolomitización (figura 1) entre ambos códigos, los resultados de transferencia de masa (figura 2) y la evolución de los parámetros de interés (figuras 3 y 4) son considerablemente similares.

Las semejanzas entre los resultados proporcionados por los distintos códigos manejados hacen ya esperable que las diferencias en las reacciones de especiación, modelo químico o algoritmo de cálculo no sean responsables de variaciones importantes en los resultados. En este sentido, el código que proporciona mayores diferencias es el CHILLER, pero en rangos poco significativos respecto al proceso definido en la simulación. Por otro lado, analizando los resultados del EQ3/6 con dos bases de datos distintas (SUPCRT92 y DATA.COM) pero con iguales constantes de equilibrio mineral, puede apreciarse que las discrepancias en los resultados, propagadas exclusivamente por diferencias en las reacciones de especiación, son pequeñas (por ejemplo, en los estados de saturación definidos para la solución inicial respecto a las fases de interés; tabla 1).

No obstante, para verificar con más precisión la incidencia de todos esos factores se procedió a repetir la simulación con un único programa, el PHREEQE, incluyendo en su base de datos los valores de las constantes de equilibrio para calcita y dolomita de los demás. Los resultados proporcionados por el PHREEQE reproducían con escasas variaciones los obtenidos por los programas de los que procedían los valores termodinámicos. En la figura 5 se señalan cualitativamente algunos de los resultados proporcionados por el PHREEQE, en concreto los correspondientes a la utilización de las constantes de equilibrio de calcita y dolomita incluidas en CHILLER-SOLTHERM y a la de la dolomita media de HARPHRQ-HATCHES (la constante de la calcita de esta base de datos es muy parecida a la existente en la de PHREEQE; tabla 1).

Como puede observarse, los resultados cualitativos definidos por PHREEQE reproducen idénticas pautas a las obtenidas originalmente por CHILLER y HARPHRQ (figura 1).

Los valores numéricos de transferencia de masa y de los parámetros de interés son asimismo considerablemente parecidos: como ejemplo, en las figuras 2B, 3 y 4 se han representado los valores correspondientes a la simulación realizada por PHREEQE con los valores de SOLTHERM para calcita y dolomita ordenada; y como se aprecia, los valores de los parámetros considerados en esas figuras son muy similares a los definidos por el propio código CHILLER en idénticas condiciones de simulación.

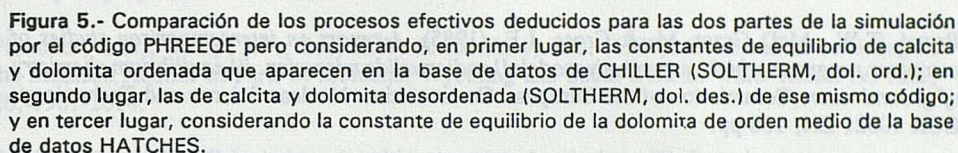
Por tanto, puede considerarse que las diferencias asignables a las distintas reacciones de especiación utilizadas por los códigos tienen escasa incidencia en los resultados del proceso simulado, incluso en aquél que mayores discrepancias parecía mostrar en este sentido (CHILLER). Por otro lado, si consideramos los resultados definidos por el PHREEQE con la dolomita media de HATCHES puede valorarse la marcada dependencia que, por el contrario, presentan los resultados de los valores de la constante de equilibrio de la dolomita: una modificación del valor original incluido en PHREEQE ($\log K = -17.09$) al de HATCHES ($\log K = -17.02$) es suficiente para variar en la segunda parte de la simulación el punto de transición entre la precipitación de dolomita y el fenómeno de dedolomitización (compárense los resultados presentados en las figuras 1 y 5).

6. Consideraciones.

El análisis realizado ha puesto de manifiesto que los valores de la constante de equilibrio de la dolomita constituyen la principal fuente de variación de los resultados cuando interviene en un proceso heterogéneo de interacción agua-roca desarrollado en un sistema carbonatado. Las diferencias en la constante de equilibrio de calcita, en las reacciones de especiación, en el modelo químico o en el algoritmo numérico de los códigos comparados no introducen, comparativamente, cambios significativos en el modelo de reacción obtenido.

La selección en la simulación del tipo de dolomita (o de la constante de equilibrio de la reacción de disolución de este mineral) puede condicionar una variación, no ya en los valores de transferencia de masa, sino incluso en el sentido en el que se producen estas transferencias o en el número de minerales involucrados en las mismas, llegando a proporcionar pautas de evolución totalmente opuestas (dedolomitización vs. dolomitización) en los casos más extremos. Circunstancia justificable si consideramos que las diferencias en las constantes de equilibrio para las dos dolomitas con grados de ordenación extremos varían en más de un orden de magnitud.

En planteamientos teóricos de modelización (aproximaciones abordadas desde métodos directos; Plummer, 1984) como los usados normalmente en análisis comparativos de códigos, es frecuente trabajar con las dolomitas incluidas en las bases de datos de los códigos sin matizar o tener en cuenta su grado de ordenación. Lo cierto es que todavía existen imprecisiones y heterogeneidades en las propias bases de datos sobre el valor considerado para su constante de equilibrio y muchas veces aparece una fase denominada específicamente como "dolomita" que, según el código o base de datos manejada, en unos casos corresponde a una dolomita de orden medio y otras a una dolomita ordenada. Ello conduce a asignar genéricamente para los resultados obtenidos una importante dependencia de la base de datos o código manejado. Dependencia que, una vez filtrado el problema del grado de ordenación de la dolomita, resulta ser considerablemente menor, al menos para las últimas versiones de los códigos comparados en este trabajo.



La modelización de procesos naturales en los que intervenga la dolomita como fase primaria o secundaria se verá además condicionada por la normal existencia de dolomitas no estequiométricas, característica que se superpondrá a los distintos grados de ordenación de su estructura y que provocará variaciones adicionales en la constante de equilibrio todavía no perfectamente delimitadas. Teniendo presente que variaciones en esa constante de rango bastante inferior al definido para las dolomitas de ordenación extrema son capaces de matizar cuando menos la pauta de evolución obtenida, es aconsejable el planteamiento de análisis de sensibilidad respecto a ese valor en la modelización de procesos en los que intervenga esta fase mineral.

Los efectos producidos por el grado de ordenación de la estructura de la dolomita en cálculos de modelización pueden ser considerados como ejemplo o análogo simplificado de los generados por el grado de cristalinidad mineral (término en el que a veces se engloban factores como tamaño de partícula, defectos cristalinos o el propio grado de ordenación estructural; p. ej. Nordstrom *et al.*, 1990), y que afecta a otros muchos minerales (p. ej. siderita, rodocrosita, gibsita, etc). Por ello, resulta aconsejable tener presente este aspecto particular en la valoración de los resultados alcanzados en la modelización de procesos en los que intervengan estas fases.

Bibliografía

- Auqué, L.F., López, P.L., Gimeno, M.J., Mandado, J. y Vallès, V. (1995). Análisis comparativo de códigos de modelización geoquímica y bases de datos termodinámicos. I. Simulación de la alteración de microclina. *Rev. Acad. Ciencias Zaragoza* (en este volumen).
- Back, W., Hanshaw, B.B., Plummer, L.N., Rahn, P.H., Rightmire, C.T. & Rubin, M. (1983): Process and rate of dedolomitization: mass transfer and ^{14}C dating in a regional carbonate aquifer. *Geol. Soc. Am. Bull.*, 94, 1415-1429.
- Bischoff, J.L., Juliá, R., Shanks III, W.C. & Rosenbauer, R.J. (1994). Karstification without carbonic acid: bedrock dissolution by gypsum-driven dedolomitization. *Geology*, 22, 995-998.
- Bond, K.A.; Moreton, A.D. & Heath, T.G. (1992): *The HATCHES User Manual*. Chemical Studies Department. Radwaste Disposal Division. Harwell Lab. Oxon. OX11 0RA, 20 pp.
- Brown, P.L.; Haworth, A.; Sharland, S.M. & Tweed, C.J. (1991): *HARPHRO. A geochemical program based on PHREEQE. User's Guide*. Theoretical Studies Department. Radwaste Disposal Division. Harwell Laboratory. Didcot. Oxon. OX11 0RA. 18 pp.
- Broyd, T.W., McD Grant, M. & Cross, J.E. (1985). *A report on intercomparison studies of computer programs which respectively model: i) radionuclide migration. ii) equilibrium chemistry of groundwater*. Commission of the European Communities. Nuclear Science and Technology. EUR 10321 EN, 188 pp.
- Busby, J. F., Lee, R. W. & Hanshaw, B. B. (1983): Major geochemical processes related to the hydrology of the Madison Aquifer System and associated rocks in parts of Montana, South Dakota, Wyoming. *U.S. Geol. Surv. Water Resour. Invest. Rep.*, 83, 180 pp.
- Busby, J.F., Plummer, L.N., Lee, R.W. & Hanshaw, B.B. (1990): Geochemical evolution of water in the Madison aquifer in parts of Montana, South Dakota and Wyoming. *U.S. Geol. Surv. Prof. Pap.*, 1273-F, 89 pp.
- Carpenter, A.B. (1980). The chemistry of dolomite formation I: the stability of dolomite. *SEPM Special Publication*, 28, 111-121.
- Drever, J.J. (1988): *The geochemistry of natural waters*. Prentice-Hall Inc., N.J., 437 pp.
- Helgeson, H.C. (1984): *SUPCRT. Unpublished thermodynamic data base for minerals, aqueous species and gases*. Univ. of California, Berkeley, California.
- Helgeson, H.C., Delany, J.M., Nesbitt, H.W. & Bird, D.K. (1978): Summary and critique of the thermodynamic properties of rock-forming minerals. *Am. J. Sci.*, 278, 229 pp.
- INTERA ENVIRONMENTAL CONSULTANTS INC. (1983): *Geochemical models suitable for performance assessment of nuclear waste storage: comparison of PHREEQE and EQ3/EQ6*. ONWI-473. Battelle Memorial Institute, Office of Nuclear Waste Isolation, Columbus, Ohio.
- Johnson, J.W., Oelkers, E.H. & Helgeson, H.C. (1991): *SUPCRT92, a software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species and reactions from 1 to 5000 bars and 0° to 1000 °C*. User's manual, 101 pp.

Nordstrom D.K. & Ball, J.W. (1989): Mineral saturation states in natural waters and their sensitivity to thermodynamic and analytic errors. *Sci. Geol. Bull.*, 42, 269-280.

Nordstrom, D.K., Plummer, L.N., Langmuir, L., Busenberg, E., May, H.M., Jones, B.F. & Parkhurst, D.L. (1990). Revised chemical equilibrium data for major water-mineral reactions and their limitations. In: D.C. Melchior & R.L. Basset (eds.), *Chemical Modeling of Aqueous Systems II*. A.C.S. Symp. Series, 416, 398-413.

Parkhurst, D.L., Thorstenson, D.C. & Plummer, L.N. (1990): PHREEQE, a computer program for geochemical calculations. *U.S. Geol. Surv. Water Res. Inv.*, 80-96, 193pp.

Plummer, L.N. (1984): Geochemical modeling: a comparison of forward and inverse methods. In: Hitchon, B. & Wallick, E.I. (eds.), *Practical Applications of Ground Water Geochemistry*. Proceedings of First Canadian/American Conference on Hydrogeology. Banff, Alberta, Canada. pp. 149-177.

Plummer, L.N. & Busenberg, E. (1982): The solubilities of calcite, aragonite and vaterite in CO_2 - H_2O solutions between 0 and 90 °C, and an evaluation of the aqueous model for the system CaCO_3 - CO_2 - H_2O . *Geochim. Cosmochim. Acta*, 46, 1011-1040.

Plummer, L.N., Busby, J.F., Lee, R.W. & Hanshaw, B.B. (1990). Geochemical modeling of the Madison Aquifer in Parts of Montana, Wyoming and South Dakota. *Water Resources Res.*, 26, 1981-2014.

Reed, M.H. (1982): Calculation of multicomponent chemical equilibria and reaction processes in systems involving minerals, gases and aqueous phase. *Geochim. Cosmochim. Acta*, 46, 513-528.

Reed, M.H. & Spycher, N.F. (1987): *SOLTHERM data base. A compilation of thermodynamic data from 25 °C to 300 °C for aqueous species, minerals and gases*. Eugene, Oregon, Dept. Geol. Sci., Univ. Oregon, 30 pp.

Reed, M.H. & Spycher, N.F. (1989): *SOLTECHM: Data base of equilibrium constants for aqueous-mineral-gas equilibria*. Dept. Geol. Sci., Univ. Oregon. Eugene, Oregon. 45 pp.

Robie, R.A., Hemingway, B.S. & Fisher, J.R. (1978): Thermodynamic properties of minerals and related substances at 298.15 °K and 1 bar (10^5 pascals) pressure and at higher temperatures. *Geol. Surv. Bull.*, 1452, 456 pp.

Spycher, N.F. & Reed, N.H. (1989a): *SOLVEQ: A computer program for computing aqueous-mineral-gas equilibria*. Dept. of Geological Sciences. University of Oregon. Eugene, Oregon, 37 pp.

Spycher, N.F. & Reed, N.H. (1989b): *CHILLER: A program for computing water-rock reactions, boiling, mixing and other reaction processes in aqueous-mineral-gas systems*. Dept. of Geological Sciences. University of Oregon. Eugene, Oregon, 63 pp.

Spycher, N.F. & Reed, N.H. (1992): Microcomputer-based modeling of speciation and water-mineral-gas reactions using programs SOLVEQ and CHILLER. In: Y.K. Kharaka and A.M. Maest (eds.), *Water-Rock Interaction*, 1087-1090. Balkema, Rotterdam.

Tena, J.M., Auqué, L.F., Gimeno, M.J., Fernández, J. y Mandado, J. (1990). *Estudio comparativo de software para solución de equilibrios fisicoquímicos*. CIEMAT, Madrid (no publicado), 4 vols.

Wolery, T.J. & Daveler, S.A. (1992). *EQ6, a computer program for reaction path modeling of aqueous geochemical systems: theoretical manual, user's guide and related documentation (version 7.0)*. Lawrence Livermore National Laboratory. UCRL-MA-110662 PT IV.

Pautas de especiación de las Tierras Raras en las soluciones ácidas de Osamu-Utsumi, Poços de Caldas (Brasil).

Auqué, L.⁽¹⁾; López, P.⁽¹⁾; Gimeno, M.J.⁽²⁾; Mandado, J.⁽¹⁾; Gómez, J.⁽¹⁾ y Zamora, A.⁽¹⁾

(1) Área de Petrología y Geoquímica. Depto. Ciencias de la Tierra. Fac. Ciencias. Universidad de Zaragoza. 50009 ZARAGOZA (España).

(2) Unidad de Seguridad de Emplazamientos. Edificio 20 T. CIEMAT. Avda. Complutense 22. 28006 MADRID (España).

Abstract

Ground waters from the Osamu Utsumi uranium mine are of K-Fe-SO₄⁻-F type, a highly unusual composition related to weathering of hydrothermally altered complex of phonolites. The highest REE concentrations ever published for natural solutions were measured in the near-surface waters of low pH and high sulfate concentrations.

The speciation of the REEs was performed on several samples of those acidic waters. Results indicate that REEs exist as either free metal species (Ln⁺³) or as sulfate complexes (mainly LnSO₄⁺), depending on sulfate concentrations. REE fluoride complexing is absent in spite of the high concentrations of this ligand because of the competition of Al for fluoride speciation. This result indicates that REE fluoride complexing will be of minor importance in acidic waters where general high Al concentrations are frequent.

1. Introducción.

El estudio del comportamiento de las Tierras Raras (REE) en distintos sistemas acuosos se ha visto recientemente favorecido tanto por una generalización de las técnicas analíticas adecuadas para su determinación (ICP-MS, NAA, etc.) como por la progresiva tabulación de sus reacciones de formación de complejos en solución. Los trabajos sobre la distribución de REE en soluciones naturales abarcan ya una amplia variedad de ambientes (Michard, 1989; Elderfield *et al.*, 1990; Fee *et al.*, 1992; etc.) y los cálculos de especiación han permitido caracterizar, de forma incipiente, algunas de las pautas de comportamiento de estos elementos especialmente en aguas de mar y/o de caracteres neutro-básicos de baja temperatura (p. ej. Möller y Bau, 1993; Johannesson *et al.*, 1994).

Dentro de este marco, los estudios sobre el comportamiento de las REE en aguas de carácter ácido han recibido una atención considerablemente menor. Aunque, sin embargo, los pocos estudios que se han centrado en este tipo de soluciones han destacado la extraña e intrigante pauta convexa de distribución de estos elementos (con máximos centrados en las Tierras Raras Intermedias, IREE) y su aparente relación con procesos específicos de interacción agua-

mineral (Auqué *et al.*, 1993, 1994a; Johannesson y Lyons, 1995), así como sus particularidades en los procesos de fraccionamiento respecto a las fases coloidales relacionadas (Auqué *et al.*, 1994b). Todo ello remarca la importancia que pueden tener estos sistemas de aguas ácidas en la interpretación y comprensión del comportamiento geoquímico de las REE.

Los trabajos recientemente realizados dentro del Proyecto Internacional de Poços de Caldas en Brasil (Chapman *et al.*, 1992) han suministrado interesantes datos sobre la geoquímica de las soluciones ácidas relacionadas con uno de los análogos naturales estudiados, el de la mina de uranio de Osamu Utsumi. Pero además, los análisis realizados en algunas de las muestras ácidas de este sistema presentan los mayores contenidos de REE determinados hasta la fecha en aguas naturales y su pauta de distribución muestra aspectos distintivos de los hasta ahora definidos en este tipo de soluciones; caracteres que hacen especialmente interesante este sistema para el estudio del comportamiento de las Tierras Raras.

La caracterización de las pautas de distribución de REE realizadas dentro del mencionado Proyecto ha sido considerablemente detallada, abarcando aspectos sobre los fenómenos de fraccionamiento entre la carga en suspensión y las soluciones, los relativos a los procesos de interacción agua-roca e incluso la caracterización de los contenidos asociados a la fracción coloidal. Sin embargo, quedan todavía muchos puntos que delimitar de forma más detallada, y uno de ellos se refiere a la distribución de especies (especiación) de estos elementos en la solución.

En este trabajo se aborda la realización de un análisis específico de distribución de especies de las Tierras Raras para algunas de las muestras más destacadas de este sistema, aprovechando las determinaciones más recientes sobre las constantes de estabilidad de las reacciones de formación de complejos de estos elementos. Las especiales características composicionales de estas soluciones con determinados ligandos en concentraciones anormalmente elevadas (caso del F⁻) hacen particularmente interesante el análisis detallado de la distribución de especies de las REE en este sistema.

2. Caracteres generales del sistema y distribución de REE.

Las soluciones asociadas a la mina de uranio de Osamu Utsumi son de tipo K-Fe-SO₄⁼, con altos contenidos en flúor y de caracteres ácidos y oxidantes (Nordstrom *et al.*, 1990, 1992; Miekeley *et al.*, 1992a; tabla 1), resultado de un intenso lavado del complejo fonolítico (muy alterado hidrotermalmente) con el que se relacionan. Las aguas más ácidas, sulfatadas y con mayores contenidos en elementos metálicos disueltos son las de carácter más superficial (el muestreo del sistema fue realizado a partir de sondeos a distintas profundidades). La oxidación de la pirita y esfalerita dispersas en el complejo alterado confiere a las soluciones las características típicas de un agua ácida de mina (Nordstrom *et al.*, 1992). También son las que presentan mayores concentraciones de F⁻ procedente de la disolución de fluorita, mineral abundante en la ganga de la mina.

Estas soluciones superficiales han sido las elegidas para realizar el estudio, puesto que son las que presentan las mayores concentraciones de REE analizadas hasta el momento a nivel mundial. Y de ellas se han seleccionado la GW-47 (Σ REE = 29 ppm) y la GW-58 (Σ REE = 1ppm), como muestras de caracteres composicionales extremos (tanto en lo que se refiere a elementos mayores, traza y REE; tabla 1) de todo el conjunto.

Tabla 1: Análisis químicos de las aguas ácidas de Osamu Utsumi (datos de Nordstrom *et al.*, 1990, 1992) y contenidos en REE (datos de Miekeley *et al.*, 1992a). Concentraciones expresadas en ppm.

	GW-47	GW-58		GW-47	GW-58
Temp (°C)	20.0	20.0	La	13.5	0.567
pH	3.57	4.43	Ce	6.270	0.124
Eh (mV)	751.0	614.0	Pr	---	---
HCO ₃ ⁻	---	---	Nd	6.490	0.250
Cl ⁻	<2.0	1.00	Sm	0.764	0.0285
SO ₄ ²⁻	445.00	8.80	Eu	0.206	0.0078
F ⁻	10.5	0.46	Gd	0.585	0.0228
PO ₄ ³⁻	<0.05	0.30	Tb	---	---
Na ⁺	0.62	1.56	Dy	0.595	0.0233
K ⁺	24.5	13.3	Ho	0.130	0.0053
Ca ²⁺	24.7	2.65	Er	0.309	0.0122
Mg ²⁺	3.6	0.85	Tm	---	---
Al _T	23.0	1.18	Yb	0.135	0.0055
Fe _T	0.21	0.07	Lu	0.014	0.00059

En la figura 1, y a efectos únicamente comparativos, se ha representado la distribución de REE normalizada frente a NASC (Haskin *et al.*, 1968; Gromet *et al.*, 1984) de esas dos muestras junto con la determinada para otros dos medios de aguas ácidas (las soluciones del Arroyo del Val y de Colour Lake; Auqué *et al.*, 1994b y Johannesson y Lyons, 1995, respectivamente). Asimismo, se incluye la distribución correspondiente a un agua de río (Amazonas, Elderfield *et al.*, 1990) y la característica de un agua marina (Elderfield y Greaves, 1992). Como puede apreciarse, las concentraciones determinadas para las aguas de Osamu Utsumi pueden llegar a ser hasta cinco o seis órdenes de magnitud superiores a los contenidos característicos de aguas superficiales de carácter neutro-básico (aguas de río y de mar). El único sistema con unas concentraciones de REE cercanas a las de Osamu Utsumi corresponden a las aguas, también ácidas, del Arroyo del Val. No obstante, los contenidos de la muestra GW-47 son cuando menos dos órdenes de magnitud superiores a cualquiera de las de ese arroyo.

Para el conjunto de aguas seleccionado, la pauta de distribución de las REE es muy similar, tanto considerando la normalización frente a NASC (figura 1) como la utilizada por Miekeley *et al.* (1992a) respecto a condrito. La obtenida frente a la roca fonolítica en contacto con las soluciones presenta pequeñas variaciones aunque no modifica los caracteres básicos observables a partir de las otras normalizaciones. Así, la pauta de estas aguas ácidas está caracterizada por la existencia de enriquecimientos netos de las Tierras Raras pesadas (HREE) respecto a las ligeras (LREE) de forma que las relaciones de los valores normalizados frente a

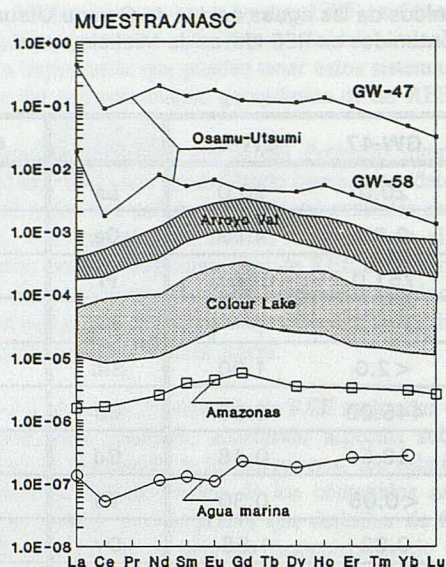


Figura 1. Pautas de distribución de las REE en las muestras GW-47 y GW-58 de las aguas ácidas de Osamu Utsumi (Brasil). A efectos comparativos se incluyen las pautas de otros sistemas de aguas ácidas estudiados en la bibliografía (Arroyo del Val y Colour Lake) así como las correspondientes a un agua de río y al agua de mar (ver texto).

NASC de La y Lu (La_N/Lu_N) son considerablemente superiores a la unidad, en torno a 14.4 para ambas muestras. Dentro de esta pauta general destaca la existencia de una clara anomalía negativa para el Ce, asociada a la oxidación del Ce (III) en las soluciones y a la retención preferencial de Ce (IV) en las partículas en suspensión (Miekeley *et al.*, 1992 a y b).

En conjunto, la pauta de distribución de REE de las aguas ácidas superficiales de Osamu Utsumi es considerablemente distinta de la obtenida en otros sistemas de aguas ácidas (figura 1) en los que se encuentra caracterizada por enriquecimientos netos de las HREE respecto a las LREE y, especialmente, por el desarrollo de una marcada convexidad definida por los valores normalizados para las Tierras Raras Intermedias (IREE; Auqué *et al.*, 1993, 1994 a y b; Johannesson y Lyons, 1995), sin que tampoco sean apreciables anomalías relacionadas con Ce o Eu. Diferencias que justifican un análisis más detallado de las pautas distribución de REE de estas aguas, comenzando por la determinación de su distribución de especies en disolución.

3. Metodología.

La descripción detallada de la metodología de muestreo y análisis para la determinación de elementos mayores y traza de las soluciones utilizadas en este trabajo puede verse en Nordstrom *et al.* (1990, 1992). Y la correspondiente a la de las Tierras Raras en Miekeley *et al.* (1992a).

Para la realización de los cálculos de especiación en las soluciones estudiadas se ha utilizado el código PHREEQE (Parkhurst *et al.*, 1990) incluyendo en su base de datos termodinámicos los valores de las constantes de equilibrio propuestas por Wood (1990) para las reacciones de formación de complejos de las REE con sulfatos, cloruros, fluoruros e hidroxilos; y los recientemente propuestos por Lee y Byrne (1992) para los complejos carbonatados y fosfatados. La consistencia de los resultados proporcionados por PHREEQE con esta base de datos ha sido verificada en soluciones de distintos caracteres composicionales, incluyendo aguas ácidas (Tena *et al.*, 1993; Auqué *et al.*, 1994a; Gimeno *et al.*, 1995).

Al valorar los resultados hay que tener presente, inicialmente, que la inexistencia de datos termodinámicos precisos sobre las reacciones de óxido-reducción que afectan al Ce en este caso específico ha obligado a simplificar el cálculo de especiación, asumiéndose que todos los elementos se comportan con igual valencia (+3). Y, por otro lado, que tanto la concentración como las pautas de REE en solución están condicionadas directamente por la metodología de filtración seguida en el muestreo.

Los resultados manejados en este trabajo corresponden a muestras filtradas a través de mallas de 0.45 micras, para las que protocolos de filtración más detallados han puesto de manifiesto la presencia de fases coloidales capaces de atravesar ese tamaño de poro. Miekeley *et al.* (1992b) indican que, pese a que la fracción coloidal (coloides orgánicos y de hierro) en las aguas de Osamu Utsumi no es cuantitativamente importante, el 50 % de la concentración de las REE que pasan a través del filtro de 0.45 micras puede encontrarse asociada a esa fracción coloidal. Desgraciadamente los resultados presentados en los mencionados trabajos sobre contenidos de REE en solución verdadera son muy incompletos y hacen referencia al Ce, elemento afectado además por procesos de óxido-reducción en este sistema de aguas ácidas.

Por ello, los resultados analíticos manejados para las REE son los correspondientes a las muestras filtradas a través de mallas de 0.45 micras y, estrictamente, representan la suma de dos fracciones distintas, la coloidal y la fracción en disolución verdadera. Su utilización en un cálculo de especiación supone evidentemente una incorrección formal y falsea las concentraciones reales asignables a las distintas especies. No obstante, la distribución porcentual de las especies de las REE, mediante la que se expresarán los resultados de este análisis, resulta independiente de la concentración de elemento total utilizada en el cálculo y permite delimitar tanto las pautas de especiación generales de estos elementos como la posible efectividad de las reacciones de formación de complejos en los fenómenos de fraccionamiento de las REE.

4. Distribución de especies de las Tierras Raras.

La distribución de especies calculada para toda la serie de las REE es similar en las dos muestras (GW-47 y GW-58) consideradas. Por ello, en la figura 2 se indica únicamente la distribución porcentual de las especies de La y Lu, como elementos extremos de la serie representativos de la especiación de todo el conjunto de las REE en estas soluciones. Y, asimismo, al hablar de los resultados de especiación se hará referencia genéricamente a la de los lantánidos (Ln).

En la muestra de carácter más ácido (GW-47; tabla 1) los complejos sulfatados representan el 82 % del total de las especies de REE con concentraciones apreciables (la especie dominante es la LnSO_4^+ con un 72 % y resulta también apreciable la de $\text{Ln}(\text{SO}_4)_2^-$ con un 10 %), estando el resto ($\approx 18\%$) representado por las formas libres. En el caso de la muestra GW-58 (figura 2) la distribución porcentual de especies se invierte: las especies predominantes están

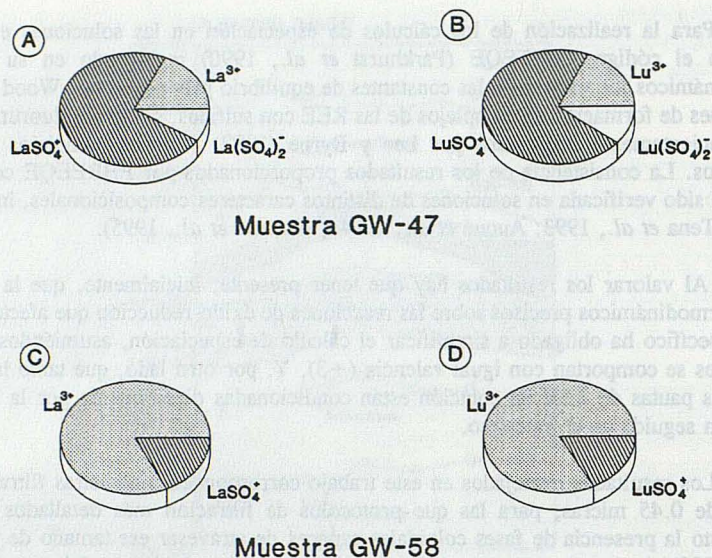


Figura 2. Resultados de los cálculos de especiación para el lantano (A y C) y lutecio (B y D) de las muestras GW-47 y GW-58, respectivamente.

constituidas por las formas libres (Ln^{+3}) con un 82 % y los complejos sulfatados (únicamente representados en este caso por la especie LnSO_4^+) constituyen el 18 % restante.

Esta inversión en los porcentajes de formas libres y complejos sulfatados se encuentra condicionada por los contenidos considerablemente distintos de SO_4^{2-} que presentan ambas muestras (tabla 1): la importante concentración de este ligando en la muestra GW-47 explica su preponderancia en la especiación de las REE; preponderancia que pierde en detrimento de las formas libres para la muestra GW-58 (figura 2), en la que incluso se minimiza el número de complejos sulfatados con concentraciones apreciables.

En cualquier caso, y aunque el número de especies y sus proporciones relativas varían de una muestra a otra, puede concluirse que la especiación de las REE en las aguas ácidas de Osamu Utsumi se encuentra controlada por la presencia de complejos sulfatados y formas libres (Ln^{+3}). Lo que implica, además, que el resto de ligandos (CO_3^{2-} , PO_4^{3-} y F^-) capaces de competir con el SO_4^{2-} en las reacciones de especiación de los lantánidos no tienen una incidencia apreciable en el cuadro de especiación obtenido, resultado especialmente evidente en el caso de los carbonatos, inexistentes en soluciones de caracteres tan ácidos como las de Osamu Utsumi.

Las concentraciones de PO_4^{3-} en estas aguas son lo suficientemente bajas como para influir en la distribución de especies de las REE; y ni siquiera en la muestra GW-58, en la que las mínimas concentraciones de SO_4^{2-} podrían favorecer la intervención de aquel ligando en la especiación, existe una presencia efectiva de complejos fosfatados: la baja concentración de SO_4^{2-} se traduce simplemente en el predominio de las formas libres (Ln^{+3}), tal como se ha indicado

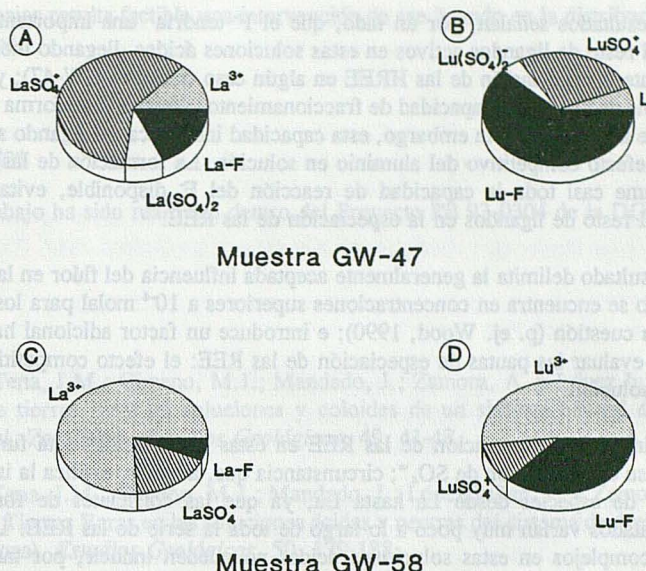


Figura 3. Resultados de los cálculos de especiación realizados sin considerar el dato de aluminio para lantano (A y C) y lutecio (B y D) de las muestras GW-47 y GW-58, respectivamente. Con el símbolo Ln-F se hace referencia al total de especies fluoruradas ($\text{LnF}^{2+} + \text{LnF}_2^+$) en cada caso.

anteriormente. Estos resultados son consistentes con el hecho de que la incidencia del PO_4^{3-} en la especiación de las REE en soluciones de pH inferior a 6 no resulta apreciable a menos que las concentraciones de este ligando sean considerablemente elevadas (Wood, 1990; Auqué *et al.*, 1994b; Gimeno *et al.*, 1995).

La ausencia de complejos fluorurados, sin embargo, merece una valoración más detallada, especialmente en el caso de la muestra GW-47 que presenta unos contenidos de flúor anormalmente altos respecto a los existentes en cualquier sistema hídrico natural ($0.5 \cdot 10^{-3}$ molal; tabla 1). Sin embargo, este ligando ve saturada su capacidad de reacción por el aluminio, lo que impide que pueda ejercer un control efectivo sobre las REE. Una manera de visualizar este efecto consiste en realizar el análisis de especiación sin incluir el aluminio en los cálculos. Los resultados obtenidos para las dos muestras consideradas pueden verse en la figura 3.

En el caso de la muestra GW-47, las proporciones de complejos fluorurados alcanzan casi el 19% del total en el caso del lantano (fundamentalmente como LaF^{2+} y, en menor proporción, como LaF_2^+ y LaF_3^0) aumentando progresivamente a lo largo de la serie de las REE hasta alcanzar el 62% en el caso del lutecio (con un 42% de LuF^{2+} y un 14% de LuF_2^+). Para la muestra GW-58, y pese a las considerablemente menores proporciones de F, la proporción de especies fluoruradas va desde el 7 a casi el 37% desde lantano a lutecio (representadas exclusivamente por la especie LnF^{2+} para ambos elementos).

Estos resultados señalan, por un lado, que el F⁻ tendría una importante capacidad de competir con el resto de ligandos activos en estas soluciones ácidas, llegando incluso a controlar mayoritariamente la especiación de las HREE en algún caso (muestra GW-47); y, por otro, que sería responsable de una cierta capacidad de fraccionamiento, afectando de forma más importante a las HREE que a las LREE. Sin embargo, esta capacidad intrínseca del ligando se ve totalmente anulada por el efecto competitivo del aluminio en solución. La formación de las especies AlF^{2+} y AlF_2^+ consume casi toda la capacidad de reacción del F⁻ disponible, evitando que pueda competir con el resto de ligandos en la especiación de las REE.

Este resultado delimita la generalmente aceptada influencia del flúor en la especiación de las REE cuando se encuentra en concentraciones superiores a 10^{-4} molal para los valores de pH de las aguas en cuestión (p. ej. Wood, 1990); e introduce un factor adicional hasta ahora poco considerado al evaluar las pautas de especiación de las REE: el efecto competitivo del resto de cationes de la solución.

En definitiva, la especiación de las REE en estas aguas ácidas está fundamentalmente delimitada por su concentración de SO_4^{2-} ; circunstancia que, además, explica la invariabilidad en la distribución de especies desde La hasta Lu, ya que las constantes de formación de los complejos sulfatados varían muy poco a lo largo de toda la serie de las REE. Los procesos de formación de complejos en estas soluciones ácidas no pueden inducir, por tanto, fenómenos importantes de fraccionamiento en los procesos de interacción agua-roca, hecho coincidente con los resultados alcanzados en otros sistemas de aguas ácidas (p. ej. Auqué *et al.*, 1994b; Johannesson y Lyons, 1995) pero opuestos a los definidos en aguas de carácter alcalino (Möller y Bau, 1993; Johannesson y Lyons, 1994).

5. Conclusiones

Los cálculos de especiación de las REE realizados con las soluciones ácidas de Osamu Utsumi han puesto de manifiesto que la distribución de especies es similar para toda la serie de estos elementos, encontrándose delimitada fundamentalmente por las concentraciones de SO_4^{2-} . En función de los contenidos de este ligando, la especiación de las REE se encuentra controlada por complejos sulfatados (LnSO_4^+ y $\text{Ln}(\text{SO}_4)_2^+$) o por sus formas libres (Ln^{3+}).

La escasa variación en las constantes de equilibrio de las reacciones de formación de complejos sulfatados a lo largo de toda la serie de los elementos de las Tierras Raras justifica un cuadro de especiación similar para todos ellos. Circunstancia que, a su vez, expresa la escasa capacidad de fraccionamiento de los procesos de especiación de las REE en soluciones de estas características, ácidas y sulfatadas.

Estos resultados, similares a los obtenidos en otros sistemas de aguas ácidas (p. ej. Auqué *et al.*, 1994), están matizados por las especiales características composicionales de las aguas de Osamu Utsumi, en las que las elevadas concentraciones de flúor hacían esperable la intervención de este ligando en la especiación de las REE. No obstante, los cálculos realizados han evidenciado que los efectos competitivos del aluminio sobre este ligando anulan su capacidad de reacción con los elementos de las Tierras Raras.

Teniendo en cuenta que las concentraciones de aluminio en soluciones ácidas suelen ser considerablemente elevadas resulta esperable que la incidencia del flúor en la especiación de las REE sea menor de la hasta ahora señalada. Resultado que, por otro lado, permite predecir que únicamente en sistemas ácidos en los que, por cualquier circunstancia, las concentraciones de

aluminio sean bajas resulta factible una intervención de ese ligando en la distribución de especies de los lantánidos.

Agradecimientos

Este trabajo ha sido realizado dentro del Proyecto PB 93-0304 de la DGICYT.

Bibliografía

- Auqué, L.F.; Tena, J.M.; Gimeno, M.J.; Mandado, J.; Zamora, A. y López Julián, P. (1993). Distribución de tierras raras en soluciones y coloides de un sistema natural de aguas ácidas (Arroyo del Val, Zaragoza). *Estudios Geológicos*, 49, 41-48.
- Auqué, L.F.; Tena, J.M.; Gimeno, M.J.; Mandado, J.; López Julián, P. y Zamora, A. (1994a). Especiación de Tierras Raras en las soluciones ácidas y neutras del sistema de drenaje del Arroyo del Val (Zaragoza). *Estudios Geológicos*, 50, 179-188.
- Auqué, L.F.; Gimeno, M.J.; López, P.L. y Mandado, J.M. (1994b). Pautas de evolución en la distribución de Tierras Raras a lo largo de un curso natural de aguas ácidas (Arroyo del Val, Zaragoza). *Rev. Acad. Ciencias Zaragoza*, 49, 165-176.
- Chapman, N.A.; McKinley, I.G.; Penna Franca, E.; Shea, M.S. & Smellie, J.A.T. (1992). The Poços de Caldas Project: an introduction and summary of its implications for radioactive waste disposal. *Jour. Geochem. Explor.*, 45, 1-24.
- Elderfield, H. & Greaves, M.J. (1982). The rare earth elements in seawater. *Nature*, 296, 214-219.
- Elderfield, H.; Upstill-Goddard, R. & Sholkovitz, E.R. (1990). The rare earth elements in rivers, estuaries and coastal seas and their significance to the composition of ocean waters. *Geochim. Cosmochim. Acta*, 54, 971-991.
- Fee, J.A.; Gaudette, H.E.; Berry Lyons, W. & Long, D.T. (1992). Rare-earth element distribution in Lake Tyrrell groundwaters, Victoria, Australia. *Chem. Geol.*, 96, 67-93.
- Gimeno, M.J.; Auqué, L.F.; López, P. & Mandado, J. (1995). Rare earth element aqueous speciation in the acidic and neutral solutions of the "Arroyo del Val" system, Iberian Chain, Spain. *MIGRATION'95*, Abstracts, pág. 97. Saint-Malo, France. September, 10-15 1995.
- Gromet, L.P.; Dymek, R.F.; Haskin, L.A. & Korotev, R.L. (1984). The "North American Shale Composite": its compilation, major and trace element characteristics. *Geochim. Cosmochim. Acta*, 48, 2469-2482.
- Haskin, L.A.; Haskin, M.A.; Frey, F.A. & Wildeman, T.R. (1968). Relative and absolute terrestrial abundances of the rare earths. In: *Origin and Distribution of the Elements* (I.H. Ahrens), pp. 889-911. Pergamon Press.

Johanneson, K.H. & Berry Lyons, W. (1995). Rare earth element geochemistry of Colour Lake, an acidic freshwater lake on Axel Heiberg Island, Northwest Territories, Canada. *Chem. Geol.*, 119, 209-223.

Johannesson, K.H.; Berry Lyons, W. & Bird, D.A. (1994). Rare earth element concentrations and speciation in alkaline lakes from the western USA. *Geophys. Res. Lett.*, 21, 773-776.

Lee, J.H. & Byrne, R.H. (1992). Examination of comparative rare earth element complexation behavior using linear free-energy relationships. *Geochim. Cosmochim. Acta*, 56, 1127-1137.

Michard, A. (1989). Rare earth element systematics in hydrothermal fluids. *Geochim. Cosmochim. Acta*, 53, 745-750.

Miekeley, N.; Couthino de Jesus, H.; Porto da Silveira, C.L.; Linsalata, P. & Morse, R. (1992a). Rare-earth elements in groundwaters from the Osamu Utsumi mine and Morro do Ferro analogue study sites, Poços de Caldas, Brazil. *Jour. Geochem. Explor.*, 45, 365-387.

Miekeley, N.; Couthino de Jesus, H.; Porto da Silveira, C.L. & Degueldre, C. (1992b). Chemical and physical characterization of suspended particles and colloids in waters from the Osamu Utsumi mine and Morro do Ferro analogue study sites, Poços de Caldas, Brazil. *Jour. Geochem. Explor.*, 45, 409-437.

Möller, P. & Bau, M. (1993). Rare-earth patterns with positive cerium anomaly in alkaline waters from Lake Van, Turkey. *Earth Planet. Sci. Lett.*, 117, 671-676.

Nordstrom, D.K.; Smellie, J.A. & Wolf, M. (1990). *Chemical and isotopic composition of groundwaters and their seasonal variability at the Osamu Utsumi mine and Morro do Ferro analogue study sites, Poços de Caldas, Brazil*. SKB Technical Report, 90-15, 111 pp.

Nordstrom, D.K.; McNutt, R.H.; Puigdomènech, I.; Smellie, J.A.T. & Wolf, M. (1992). Ground water chemistry and geochemical modelling of water-rock interactions at the Osamu Utsumi mine and Morro do Ferro analogue study sites, Poços de Caldas, Brazil. *Jour. Geochem. Explor.*, 45, 249-287.

Parkhurst, D.L.; Thorstenson, D.C. & Plummer, L.N. (1990). *PHREEQE. A computer program for geochemical calculations*. U.S. Geological Survey, Water Resources Investigations 80-96.

Tena, J. M.; Auqué, L.F.; Gimeno, M.J. y Mandado, J. (1993). Verificación de las pautas de especiación de Ce y Eu mediante aplicación del código PHREEQE. *V Congreso de Geoquímica de España*. Sección Primera: Geoquímica de los materiales y procesos geológicos. pp. 72-77.

Wood, S.A. (1990). The aqueous geochemistry of the rare-earth elements and yttrium. 1. Review of available low-temperature data for inorganic complexes and the inorganic REE speciation of natural waters. *Chem. Geol.*, 82, 159-186.