

REVISTA  
DE LA

ACADEMIA  
DE  
CIENCIAS

Exactas

Físicas

Químicas y

Naturales

DE

ZARAGOZA



Serie 2.<sup>a</sup>

Volumen 52

1997

## INDICE DE MATERIAS

	<i>Págs.</i>
L. J. Boya Balet. – «Escritos sobre física y filosofía» .....	5
A. El Kinani et M. Oudadess. – «Involution generalisee et structure de $C^*$ -Algebre» .....	15
I. K. Argyros. – «Extending the region of convergence for a certain class of modified iterative methods on banach spaces and applications» .....	17
I. K. Argyros. – «Results involving nondifferentiable equations on banach spaces with a convergence structure and newton methods» .....	23
I. K. Argyros. – «The halley method in banach spaces and the ptak error estimates» .....	31
G. Anastassiou. – «Shape-Preserving Bivariate Shift-Invariant Operators» .....	43
R. Barrio. – «Minimax Chebyshev coefficients of the real dilogarithm» .....	51
M. Ruiz Espejo. – «Sobre la cuasicovarianza muestral en el muestreo aleatorio simple» .....	55
M. Ruiz Espejo. – «Control de cuentas económicas por muestreo con número conocido de unidades grandes» .....	59
M. Ruiz Espejo. – «Nota sobre estrategias muestrales para estudios socioeconómicos en España» .....	63
M. Ruiz Espejo. – «Muestreo doble posestratificado especial para control de cuentas económicas» .....	67
M. Ruiz Espejo. – «El teorema de rao-blackwell en poblaciones finitas e implicaciones informativo-económicas» .....	71
M. Ruiz Espejo. – «Unbiased estimation of economic and social parameters with probable individual response» .....	75
M. Ruiz Espejo. – «Optimalidad insesgada de la media muestral» .....	81
A. A. Makinistian. – «Los comienzos del bipedismo en el proceso de hominización. (The beginning of bipedism in the hominization process)» .....	83
L. A. Inda y M. L. Peleato. – «Estudios en <i>Chlorella fusca</i> como modelo para el desarrollo de un método de valoración de deficiencia de hierro en ecosistemas acuáticos» .....	99
R. C. Díaz Ara y C. Albás Giral. – «Determinación de ácido úrico en orina por HPLC» .....	113
A. García-Olivares y E. Carrasco. – «Comparison of Biospheric Models of Radionuclides Transfer» .....	121
F. Sanz Morales, L. F. Auqué Sanz, J. Mandado Collado, R. B. Joven Barcelona y F. J. Torrijo Echarri. – «Precipitados carbonatados en el sistema termal de Alhama de Aragón-Jaraba» .....	145
F. J. Torrijo Echarri, J. Mandado Collado, R. B. Joven Barcelona y F. J. Sanz Morales. – «Nódulos y concreciones. Modelos genéticos» .....	159
R. B. Joven Barcelona, J. Mandado Collado, F. J. Sanz Morales y J. Torrijo Echarri. – «Rasgos edáficos en la periferia del lago terciario central de la cuenca de Calatayud» .....	171
A. Maestro González y A. L. Cortés Gracia. – «Análisis de la estabilidad de laderas en suelos utilizando los abacos de Hoek y Bray: un caso real en Camarena de la Sierra (Teruel)» .....	185
V. Sánchez Cela y A. Aparicio Yagüe. – «A Model for the Formation of Ultramafic Rocks in the Crust: Peridotites of Southern Spain and Northern Morocco» ....	197

**Comentarios sobre el libro de Wolfgang Pauli (\*)**

**ESCRITOS SOBRE FISICA Y FILOSOFIA**

por

Luis J. Boya Balet

Departamento de Física Teórica, Facultad de Ciencias  
Universidad de Zaragoza. - 50009 - ZARAGOZA, Spain

**Introducción.**

Se trata de 21 ensayos sobre física moderna y filosofía de la ciencia escritos por el gran físico teórico W. Pauli (1.900-1.958), publicados en la prensa diaria, revistas de filosofía, etc . Se dividen en tres grandes bloques: I) Escritos sobre la Teoría Cuántica, II) Escritos filosóficos generales y III) Celebraciones (*Festschriften* y obituarios) . La distinción entre los grupos es a veces algo artificial.

**I.- Escritos sobre la Teoría Cuántica.**

#1. MATERIA (1.954). Pauli nos invita a un paseo por los conceptos de materia, desde la inercia newtoniana, donde incluye la aguda observación de que la indefinición de fuerza se justifica con la tercera ley, pasando por las leyes de conservación en química, y entrando en el concepto de campo de Faraday y Maxwell, que aceptado como un punto de vista superior, condujo a Einstein a reinterpretar la gravitación como una teoría de campos. Interviene la atomística, y la equivalencia de masa y energía. Se introduce la mecánica cuántica, con énfasis en la complementariedad.

---

(\*) Editorial Debate. Colección Pensamiento. Madrid, XI-1.996.  
Traducción de "Writings on Physics and Philosophy",  
Springer, Heidelberg 1.994. Editado por C.P. Enz y K. von Meyenn

#10.- ESPACIO, TIEMPO Y CAUSALIDAD EN FISICA MODERNA (1.936). Considera aquí Pauli las tres constantes de la naturaleza ( $h$ ,  $c$ , y  $G$ ) y los tres dominios de la física clásica, en que  $h$  es aun despreciable, a saber: la física newtoniana ( $c = \infty$ ,  $G = 0$ ), la relatividad especial ( $c$  finita) y la teoría relativista de la gravitación ( $c$  y  $G$  finitas).

Epistemológicamente, la relatividad especial nos obliga a modificar el concepto de espacio-tiempo, y la relatividad general unifica el movimiento con gravedad con el movimiento libre (geodésico). La introducción de la mecánica cuántica produce un cambio aun más drástico; Pauli suscribe completamente el *Kobenhägenergeist* de la teoría cuántica, pero enfatiza más la complementariedad (Bohr) que la incertidumbre (Heisenberg). Es muy sabrosa la discusión de la emisión de luz por un átomo excitado, contraponiendo la onda esférica (caracter ondulatoria) con la *Nadelsthralung* (corpuscular), la misma comparación que había preocupado a Einstein. Es curiosa su conjetura que relaciona la constante de Fermi  $G_F$  con la gravitación, quizá el único intento de relacionar directamente estas dos interacciones.

# 18. PRINCIPIO DE EXCLUSION. (Conferencia Nobel (1.945), publicada en 1.948. Es un valioso documento histórico, que retrotrae el principio (¡de Pauli!) a los datos espectroscópicos de los años 20. Pero Pauli no dice porque no interpretó el cuarto número cuántico, como debido al spin del electrón, sino a una *unadualidad* [*Zweideutigkeit*] irreducible a la mecánica clásica; de hecho, Pauli (como Bohr, Heisenberg y otros) tardaron en aceptar la hipótesis de Uhlenbeck y Goudsmith del spin, hasta que L.H.Thomas demostró que explicaba correctamente el factor 1/2 necesario para el ajuste empírico de los dobletes alcalinos. Hace un repaso de la teoría cuántica de los años 30, destacando el principio de simetrización, la arbitrariedad en la estadística (¿porque sólo Bose-Einstein o Fermi-Dirac?), la energía del vacío, las subtracciones de Heisenberg y Dirac (que hoy día entendemos como renormalización de la carga y masa); etc. Muestra su insatisfacción por esa *Substraktionphysik* y por la *Limesakrobatik*. Y, muy en el espíritu de la época, (Heisenberg apuntó una idea parecida) atribuye al valor preciso de la constante de estructura fina, la corrección de las divergencias (esperanza vana, por cierto, y hoy día irrelevante: el concepto *derunning coupling constant* plantea actualmente el problema de modo muy distinto).

# 19. VIOLACION DE LA PARIDAD (1.958). La resistencia de Pauli a la hipótesis de Yang y Lee (1.956), sobre violación de la paridad en la desintegración del kaon cargado, es bien conocida. Fueron los decisivos experimentos de C.S. Wu con el cobalto 60 y los de Garwin *et al.* en la cadena pión  $\rightarrow$  muón, los que le convencieron. Resulta asombroso como un físico teórico de 58 años, a las puertas de la muerte, pudo seguir con

tanto detalle los complicados resultados experimentales, del otro lado del atlántico, que demostraron inequívocamente la violación de la paridad en (todos) los procesos débiles. Advierte, por ejemplo, que no puede ser el neutrino ("un fantasma de cuya existencia no estoy del todo libre de culpa", dice) la causa de todo, pues hay procesos débiles no leptónicos que tampoco conservan la paridad.

# 20. HISTORIA DEL NEUTRINO (1.958). Cuando hacia mitad de los años 20 se comprobó que el espectro de la radiactividad beta era genuinamente un espectro continuo, se adoptaron dos posiciones: Niels Bohr, siempre dispuesto a lo más radical, propuso (por segunda vez) la no conservación de la energía. Pauli, como Einstein, Heisenberg, etc., no pueden renunciar a un principio tan fundamental bajo ningún concepto. Pauli lanza la hipótesis más audaz de toda su carrera, que choca algo con su espíritu, en general conservador: existe una partícula neutra, fermiónica, de masa pequeña, que se emite con el electrón en los procesos beta. Para los sociólogos es de destacar que Pauli en la época (1.930) sufría una crisis personal, al haberle abandonado la bailarina con quien se había casado. Pespiczamente, advierte que la teoría de Fermi (1.934) tiene solo un valor fenomenológico (es una teoría "efectiva", diríamos hoy en día, no renormalizable). En cambio, no "olfatea" que el valor de la constante de Fermi apunta a una escala de energías del orden de 100-200 GeV. Es curioso el prejuicio de la época por la coincidencia del radio clásico del electrón, con el tamaño de los núcleos (que Heisenberg plasmó en su hipótesis de la longitud fundamental (1.938), del orden del fermi), lo que impidió una visión más profunda de los procesos débiles.

## II.- Escritos filisóficos.

#2.- SIGNIFICADO FILOSOFICO DE LA IDEA DE COMPLEMENTARIEDAD (1.950). Ante un auditorio de filósofos, Pauli se esfuerza por recabar de éstos el estudio de la física moderna, en especial las implicaciones filosóficas de la teoría cuántica. Dedicó especial énfasis a discutir el principio de complementariedad de Bohr. Presenta la complementariedad como incompletitud en la descripción del mundo físico: la física moderna ha restringido las ideas intuitivas sobre descripción completa del Universo (una idea también muy resaltada por Heisenberg: el progreso introduce necesariamente limitaciones). Pauli elabora el principio de complementariedad, para salvar la dualidad onda/corpusculo, lo que le lleva al principio de incertidumbre, al cuanto mínimo de acción, e incluso a ciertas excursiones sobre la conciencia que desarrolla mejor en otros ensayos.

#3.- PROBABILIDAD Y FISICA (1.952). Comienza con una carta de Pauli a Bohr (15-II-1.955), donde relata la oposición de Einstein al papel preponderante del observador, citando el famoso ejemplo de la Luna ("la Luna existe independiente de que la observemos; luego lo mismo debe ocurrir con el átomo o el electrón, puesto que no hay distinción esencial entre macro y micro-objetos. Pauli repasa luego el concepto de probabilidad en Boltzmann y Gibbs, es decir, en la mecánica estadística; esta noción es compatible con un determinismo estricto, pero Pauli advierte que la axiomática del cálculo de probabilidades puede considerarse como primaria, irreducible a otras cosas, y eso es lo que ha hecho la mecánica ondulatoria (en la interpretación de Max Born), donde la probabilidad se predica de sucesos individuales y es irreducible a otras cualidades (ausencia de variables ocultas). Naturalmente, esa concepción chocó con la mentalidad objetiva de físicos como Einstein o de Broglie, oposición que ya hemos relatado en otros comentarios. Pauli atribuye gran significación al hecho de que el concepto puro (axiomático) de probabilidad encaje también con la idea central de complementariedad, que es para Pauli la piedra angular de la filosofía de la mecánica cuántica.

#11.- LA TEORIA DE LA RELATIVIDAD Y LA CIENCIA (1.956). Filosóficamente es un ensayo importante. Establece Pauli los sucesivos avances epistemológicos, desde Aristóteles (las cualidades de las cosas), pasando por Galileo-Newton (esas cualidades están sujetas a leyes dinámicas que hay que descubrir) y Einstein (teoría especial y general de la relatividad): aquí el propio espacio adquiere independencia (eliminación del éter, primero, y sustitución de fuerza de atracción por curvatura, después). El progreso alcanza un nivel definitivo provisional con la mecánica cuántica, que restringe el concepto objetivo, primario, del campo, y lo somete a las incertidumbres cuánticas (naturalmente, Einstein no siguió por ahí, aferrándose siempre a la descripción objetiva, independiente del observador). Termina Pauli el ensayo aceptando, sin embargo, la posición de Einstein de que el científico debe exhibir varias personalidades al abordar los problemas: debe ser un poco realista, idealista en otros momentos, y hasta pitagórico a veces.

#14: TEORIA Y EXPERIMENTO (1.952).

#15: FENOMENO Y REALIDAD FISICA (1.957). En el primero de estos dos ensayos, abunda Pauli en la filosofía dualista de Goussier y Bernays, ampliándola a la relación entre lo externo (físico) e interno (psíquico). El punto de vista puramente empírico ("descripción" exhaustiva de los fenómenos) es incompleto, no tiene en cuenta que la conceptualización incluye siempre un elemento subjetivo. El segundo ensayo es más extenso y más interesante, por abordar el problema de la realidad subyacente a los fenómenos; Pauli declara de entrada no estar

asociado a ningún "-istmo" filosófico particular, tomando claramente una posición ecléctica. Insiste a continuación en que todo fenómeno tiene una carga de teoría, indudablemente los complejos fenómenos que observamos en el microcosmos. Comulga de nuevo con Bohr en admitir una definición muy laxa de objetividad, para que quepa en ella la axiomática cuántica. Observa luego que la lección aportada por la relatividad especial es destruir el concepto viejo de causalidad, en el sentido de que el movimiento uniforme no tiene causa, o mejor no más causa que el reposo. Pauli atribuye este perfilamiento del concepto a la existencia del grupo de transformaciones de Lorentz. Aprovechamos para señalar que Pauli fue el primero en utilizar ampliamente el concepto de grupo en física, siguiendo, aunque sin citarlo, el famoso programa de Erlangen de F. Klein. Se discute a continuación la epistemología cuántica, incidiendo una vez más en su carácter no determinista, aunque objetivable y racional. Y una aportación importante, expresada un tanto dialécticamente: coexisten en la ciencia cuántica actual un sentido de terminación, en cuanto la interpretación de Kopenhague es provisionalmente final, y un talante abierto, en el sentido de que nuestros enormes vacíos de conocimiento del mundo subnuclear dejan una puerta abierta a posibles e importantes descubrimientos epistemológicos futuros.

#16.- CIENCIA Y PENSAMIENTO OCCIDENTAL (1.955). ¿Es la ciencia un producto exclusivamente occidental? ¿Hay influencias místicas, orientales, en nuestra filosofía científica? Pauli parte de los griegos, analizando especialmente las elucubraciones de Demócrito, Platón y Pitágoras, más bien complementarias que opuestas, y cuyas consecuencias se filtran en toda la física, desde la edad moderna (Galileo y Newton), hasta la teoría cuántica y N. Bohr. Es muy interesante como descubre ingredientes místicos en la ciencia, y como los puede retrotraer al pensamiento hindú, etc. Piensa que el destino del mundo occidental es el de mantener una tensión continua, entre la actitud racional y crítica que intenta comprender, y la irracional y mística, que busca la salvación en la unidad; su antagonismo es un reflejo más de la complementariedad (me pregunto si Bohr hubiese llegado tan lejos). La tentación mística florece repetidamente en occidente, y puede verse hasta en la alquimia del siglo XVII; las concepciones científicas de un Göthe, por ejemplo, se explican mejor por sus antecedentes alquimistas. La conclusión de Pauli es para meditar: la concepción racionalista occidental ha sobrepasado su zenit, y se encuentra ahora (1.955) constreñida. Nosotros podemos agregar, desde la perspectivas de fin de siglo, que la componente irracional de la vida humana es cada vez más aparente.

#17. IDEAS DEL INCONSCIENTE. (en el 80 aniversario de C. Jung, 1.954). El filósofo Carl JUNG (1.875-1.961) sostuvo una interesante correspondencia con Pauli (véase Pauli y Jung, *Un intercambio epistolar*,

Alianza Editorial, Madrid 1.996). Para honrar a su amigo, Pauli se mete deliberadamente en el terreno del psicoanálisis y del inconsciente. Descubre analogías entre el inconsciente de Jung y la idea de campo físico, al desplazarse ambos desde el ámbito visual al paradójico. Al concepto de "arquetipo", central en Jung, se antepone el de leyes estadísticas primarias (es decir, irreducibles) de la naturaleza; ambas formulaciones tienden a ampliar la antigua idea de causalidad estricta. Pauli piensa que estas ideas sobre el inconsciente puedan extenderse más allá de la terapéutica, confluyendo con las corrientes científicas modernas. Hay otras analogías; por ejemplo, Pauli ve en el "sacrificio" de la física moderna (e.g. renuncia a una descripción microcausal, incertidumbre) un parecido con el "sacrificio" en psicología (pérdida del conocimiento).

Pauli quiere ver en los escritos de Jung un reconocimiento de lo no físico en relación al problema de la unidad psicofísica, cuyo primer paso pudo ser el encuentro de la psicología con la alquimia, hecho al que él atribuye un gran valor. Es notorio también la abierta actitud del Pauli frente a los fenómenos de percepción extrasensorial (ESP), ante experimentos indicativos y desde luego no fraudulentos. Termina el interesante estudio volviendo a las posibles aplicaciones de las ideas acerca del inconsciente a las ciencias cuantitativas, especialmente a la transferencia de los arquetipos al pensamiento pitagórico y kepleriano; esto le lleva a consideraciones muy interesantes sobre biología evolutiva, descalificando ciertos intentos lamackianos en favor del mecanismo clásico de mutación y selección de Darwin, aplicado ahora a la preservación de los patrones específicos de comportamiento.

#21.- LAS IDEAS DE LOS ARQUETIPOS EN KEPLER (1.952). Es éste quizá el artículo filosófico más famoso de Pauli, muy citado en la actualidad por físicos y filósofos. Se estudia aquí la ciencia en el s. XVII, y Pauli se interesa especialmente en la relación entre las percepciones sensoriales y los conceptos físicos, es decir, una aproximación no puramente empírica a la naturaleza; es el punto de vista platónico, de ideas innatas, que conjugadas con los duros hechos empíricos alumbran las teorías científicas. Llamamos "arquetipos" a estas imágenes primarias que el alma percibe, que actúan como un fondo o tapiz para escribir y describir nuestras sensaciones elaboradas; Pauli retrotrae esta concepción desde Platón a Jung, pasando por J. Kepler (1.571- 1.630), a quien estudia en detalle.

Para empezar, las leyes de Kepler sobre el movimiento planetario, que lo han inmortalizado, no era lo que él buscaba; intentaba más bien encontrar "la armonía de las esferas" en las órbitas de los cuerpos celestes. La estructura es jerárquica y fuertemente teológica, con Dios en la cumbre y la mente humana a su imagen. Matemáticas y

en particular Geometría parecen innatas al alma humana, pero Pauli denuncia enseguida que excesos en esta línea lleva a los juicios "a priori" de Kant, a sus categorías, que no admite, como la inevitabilidad de la geometría euclídea, solo una entre tantas, como se sabía desde Gauss en el s. XIX y sobre todo desde el descubrimiento de la relatividad general. El universo de Kepler sigue luego, en el mundo de lo corpóreo, con el Sol en el centro y los esféricos Planetas, con la Tierra por supuesto, en el siguiente nivel. Es curiosa la comparación de la Trinidad con la tridimensionalidad del espacio; Pauli ve también un antecedente a la idea de campo (de Faraday y Maxwell) en el influjo gravitatorio del Sol sobre los Planetas, lo que relaciona con la emisión de luz. Interesa menos, en mi opinión, la comparación de los actos humanos a la luz de la astrología, que Kepler indirectamente defiende.

Hay también una oposición maniquea entre luz (Sol) y oscuridad (materia): la ontología kepleriana distingue a los seres por la proporción relativa de una y otra, desde los ángeles a la materia inerte. La astrología racional de Kepler se opone a la fundamentalista de Fludd, autor coetáneo, cuya controversia con Kepler sigue Pauli en detalle; la ventaja esencial, a favor de Kepler, es que éste es cuantitativo. Pauli saca siempre a colación la física moderna, y destaca en particular que la dicotomía cualidad/cantidad se refleja de algún modo en la dualidad partícula/onda o en el principio de complementariedad. Termina el ensayo con ciertas divagaciones sobre la relación del alma con el cuerpo, problema que Pauli considera aun abierto.

### III.- Escritos Obituarios y *Festschrifts*.

# 4. EN EL SESENTA ANIVERSARIO DE N. BOHR (1.945). Es remarcable que todos los creadores de la mecánica cuántica profesasen una admiración sin límites por Niels Bohr, que va mucho más allá del valor intrínseco de las contribuciones científicas de éste. Se nota ello en Heisenberg, Dirac, Schrödinger, Pauli, el propio Einstein, etc. Pauli realiza en este trabajo onomástico un repaso a la obra de Bohr, destacando el carácter provisional de la mecánica cuántica antigua, que nadie vió como el propio Bohr, y pasó por alto algún desliz del gran danés, como la conservación sólo estadística de la energía en el célebre trabajo con Kramers y Slater (1.924) que, sin embargo, sentó las bases para que Heisenberg descubriese la verdadera mecánica cuántica al año siguiente.

# 5.- 80 ANIVERSARIO DE SOMMERFELD (1.948). Sommerfeld fue el físico matemático por excelencia. El "llevó la teoría cuántica a las masas", diríamos parodiando a Schwinger (en relación con Feynman). La aplicación de la relatividad especial a la estructura fina del hidrógeno es un trabajo magistral, que la naturaleza ha premiado, haciéndola válida

hoy día (fórmula de Dirac, 1.928) por una milagrosa compensación de dos defectos (ausencia de spin y reglas de cuantificación Bohr, no mecánica cuántica actual). Sin embargo, tanto Pauli como Heisenberg experimentaron la misma sensación al pasar de Sommerfeld a Bohr (a raíz del primer Bohr *Festspiele* en Göttingen, 1.922): él se dió más cuenta que nadie del carácter revolucionario y al mismo tiempo provisional de la teoría cuántica antigua.

# 6.- OBITUARIO POR SOMMERFELD (1.951). Repite, más brevemente, las consideraciones anteriores sobre el ilustre difunto. Añadamos el disgusto que tanto Bohr como Heisenberg expresaron varias veces en privado, de que Sommerfeld no recibiese el Premio Nobel, a pesar de haber sido propuesto varias veces.

#7.- RYDBERG Y EL SISTEMA PERIODICO (1.955). El espectroscopista sueco Rydberg (1.854-1.919) fue el primero que intentó entender las regularidades que Mendeleev había observado en el sistema periódico. Su trabajo es muy primitivo, y el problema no fue resuelto hasta las investigaciones de Bohr (1.922) y Stoner (1.924). Compárese el modelo espiral de Rydberg (1.914) con el primero de Bohr (1.922). El nombre de Rydberg ha quedado para siempre ligado a su constante, la unidad fundamental en los espectros atómicos, conocida experimentalmente hoy día con precisión superior al billón.

#8.- OBITUARIO POR P. EHRENFEST (1.933). La contribución esencial de Ehrenfest (1.880-1.933) fue la hipótesis adiabática, la manera más racional de introducir las reglas cuánticas en la teoría de Bohr. Su trabajo ha sido revalorizado en tiempos modernos por Guztwiller, en relación con la cuantificación de sistemas caóticos. Uno tiene la sensación que Ehrenfest no recibió el crédito suficiente por esta aportación; la incorporación de la termodinámica a la teoría cuántica es hoy rutinaria, pero en 1.914 supuso un punto de vista demasiado novedoso. Es trágico que la impotencia de Ehrenfest, ya mayor, por entender los nuevos desarrollos cuánticos (desde 1.927) fué lo que probablemente le llevó al suicidio.

# 9.- EINSTEIN Y LA TEORIA CUANTICA. Uno de los mejores ensayos de esta colección, para mi gusto. A la vista de la mecánica cuántica moderna (desde 1.925) y dado además la animadversión de Einstein a la interpretación convencional de la misma, es difícil apreciar la enorme labor de Einstein en el marco de la teoría cuántica antigua, desde la aparición de la ley del cuerpo negro de Planck; así como éste pretendió toda su vida deducir su ley sin forzar la rotura con la física clásica, Einstein se dio cuenta el primero que la hipótesis de Planck suponía una rotura total, no solo estadística o macroscópica, con la física clásica. Tanto las consideraciones termodinámicas de la ley de Planck

(fluctuaciones, etc.) como la teoría estadística de los calores específicos (1.907) o como la teoría de absorción y emisión (1.917), son trabajos magistrales que Pauli glosa acertadamente. Einstein se da cuenta, por ejemplo, del ineludible rasgo de aleatoriedad que tienen las transiciones cuánticas, y mantiene la esperanza (¿vana?) de que ello se arregle en una teoría futura. Por último, no es de desdeñar la contribución de Einstein en el descubrimiento de la mecánica ondulatoria, siendo una especie de vehículo entre de Broglie y Schrödinger; la relación pasa por la carta del indio Bose a Einstein, que dió lugar al nacimiento de la estadística que lleva sus nombres.

#12.- IMPRESIONES SOBRE EINSTEIN (Obituario, 1.955). En un cierto momento (1.945) Einstein había considerado a Pauli su sucesor; la admiración era, pues, recíproca; colaboraron científicamente. En este obituario, publicado en un periódico, rememora Pauli la época de Einstein en la oficina de patentes de Berna, cuando concibió la relatividad especial entre otras cosas. Pauli nos recuerda que "el rasgo característico del nuevo modo de pensar que Einstein introdujo en la física es el análisis riguroso de los conceptos tradicionales a la luz de principios más generales; en este proceso, determinadas hipótesis se revelan como restrictivas, y algunos conceptos son susceptibles de ser eliminados". Se nota cierta desilusión cuando Pauli narra la oposición de Einstein a la nueva mecánica cuántica, cuando él tanto había contribuido a que se originase; por cierto, esto fué también un *shock* enorme para Bohr, Heisenberg, Born, etc. Parece ser que, en las puertas de la muerte, Einstein reconoció a Pauli que sus intentos de unificación de electricidad y gravitación al margen de la teoría cuántica habían sido en balde.

#13.- A. EINSTEIN Y EL DESARROLLO DE LA FISICA (1.958). En una glosa de la obra de Einstein parecida a la anterior (#12), Pauli nos recuerda que el aferrarse Einstein al concepto de campo proviene de su gran hazaña de liberarse del éter, mejorando así la visión de Maxwell del electromagnetismo; naturalmente, ese mismo apegamiento le impidió, quizá, aceptar la visión menos realista y objetiva que la nueva mecánica cuántica proponía.

### **Comentario general.**

Estos *Escritos sobre Física y Filosofía* suponen un documento de incalculable valor para el estudio de la filosofía científica del siglo XX. Tienen también un valor biográfico innegable: las agudas observaciones de Pauli sobre Einstein, Bohr, Sommerfeld, etc., harán la delicia de los estudiosos. No nos pronunciamos sobre sus aventuras en psicología y en especial su énfasis en el carácter central del inconsciente, donde Pauli

solamente paga un tributo, creo yo, a la filosofía dominante, protagonizada por su arquetipo particular, Carl Jung.

Para terminar, queremos hacer notar que la edición y traducción española de la obra han sido esmeradas. Sólo nos hubiese gustado que los términos alemanes entre paréntesis se pusiesen inmediatamente después de la palabra castellana, y no al final de la frase. El original inglés, preparado por C. Enz (discípulo directo de Pauli) y von Meyenn (editor de su correspondencia) es una obra valiosa, y debemos decir que la traducción no desmerece. Recomendamos el libro especialmente a los estudiosos de las ideas aportadas por la teoría cuántica en la cultura humana en general.

Luis Joaquín BOYA BALET.

## INVOLUTION GENERALISEE ET STRUCTURE DE $C^*$ -ALGEBRE

A. El kinani et M. Oudadess

Ecole Normale Supérieure, B.P: 5118 Takaddoum  
10105 Rabat (Maroc)

**Abstract:** We examine Banach algebras  $A$  endowed with an antimorphism  $x \mapsto x^*$  i.e. a vector involution such that  $(xy)^* = x^*y^*$  for every  $x$  and  $y$  in  $A$ . We show that they are  $C^*$ -algebras if the condition  $\|xx^*\| = \|x\|^2$ , for every  $x \in A$ , is fulfilled.

### Introduction

Par définition, une involution généralisée sur une algèbre complexe  $A$  est une involution d'espace vectoriel  $x \mapsto x^*$  ([1]) telle que  $(xy)^* = y^*x^*$ , pour tous  $x, y$  dans  $A$  (auquel cas on dit que c'est une involution d'algèbre), ou  $(xy)^* = x^*y^*$ , pour tous  $x, y$  dans  $A$ . Nous montrons que toute algèbre de Banach (non nécessairement unitaire) munie d'une involution généralisée telle que  $\|xx^*\| = \|x\|^2$ , pour tout  $x$  dans  $A$ , est en fait une  $C^*$ -algèbre. Tout revient à montrer que l'algèbre est nécessairement commutative et donc que l'involution généralisée est en fait une involution d'algèbre.

**Théorème:** Soit  $(A, \|\cdot\|)$  une algèbre de Banach complexe, munie d'une involution généralisée  $x \mapsto x^*$ . Si  $\|xx^*\| = \|x\|^2$ , pour tout  $x$  dans  $A$ , alors  $A$  est une  $C^*$ -algèbre.

**Preuve:** Le cas à considérer est celui où il ne s'agit pas d'une involution d'algèbre. Posons  $Sym(A) = \{x \in A : x = x^*\}$ . Il suffit de montrer que  $hk = kh$ , pour tous  $h, k$  dans  $Sym(A)$ . Remarquons que  $Sym(A)$  est stable par produit et que

$$\rho(h) = \|h\|, \quad \forall h \in Sym(A). \quad (1)$$

Notons aussi que si  $A^1 = A \oplus C$  est l'algèbre obtenue par adjonction de l'unité à  $A$ , alors  $x \mapsto x^*$  se prolonge à  $A^1$  en une involution généralisée continue. Soient maintenant  $h, k$  dans  $Sym(A)$  et considérons la fonction

$$f(\lambda) = e^{\lambda h} h k h e^{-\lambda h}, \quad \lambda \in C.$$

Pour toute forme linéaire continue  $\varphi$  sur  $A$ , la fonction  $\varphi \circ f$  est holomorphe. Elle est de plus bornée car si  $\lambda = \alpha + i\beta$ , on a

$$\begin{aligned} |\varphi(f(\lambda))| &\leq \|\varphi\| \|f(\lambda)\| \\ &\leq \|\varphi\| \|e^{i\beta h} h\| \|e^{\alpha h} k e^{-\alpha h}\| \|h e^{-i\beta h}\| \\ &\leq \|\varphi\| \|h^2\| \rho(k), \end{aligned}$$

vu que  $\|e^{i\beta h} h\|^2 = \|e^{i\beta h} h e^{-i\beta h} h\| = \|h^2\|$ ,  $\|h e^{-i\beta h}\|^2 = \|h^2\|$  et  $\|e^{\alpha h} k e^{-\alpha h}\| = \rho(k)$ . D'après le théorème de Liouville et le théorème de Hahn-Banach, on a  $f(\lambda) = f(0)$  pour tout  $\lambda$ . En dérivant et en prenant  $\lambda = 0$ , on obtient  $h^2 k h = h k h^2$ . Alors, par (1), on a

$$(h k - k h) h^2 = 0 \quad \text{et} \quad h(h k - k h) h = 0. \quad (2)$$

Si  $[a, b]$  désigne le commutateur  $ab - ba$  de  $a$  et  $b$ , on vérifie que  $[h, [h, kh]] = 0$ . D'où

$$e^{\lambda h} k h e^{-\lambda h} = kh + \lambda [h, kh], \quad \text{pour tout } \lambda \in \mathbb{C}.$$

Et l'on a, pour tout  $t \in \mathbb{R}^*$ ,  $\rho(t kh + [h, kh]) = |t| \rho(kh)$ . Alors, par (1),

$$\|t kh + [h, kh]\| = |t| \rho(kh); \quad \text{pour tout } t \in \mathbb{R}^*.$$

Ce qui donne  $(hk - kh)h = 0$  et  $h(hk - kh) = 0$ . Mais alors  $[h, [h, k]] = 0$ . Et l'on obtient comme ci avant  $[h, k] = 0$  i.e.  $hk = kh$ . ♦

**Remarque 1:** La preuve peut être conduite autrement. En effet; on peut partir de (2) et appliquer le lemme de Kleinecke- Shirokov ([3], [4]) deux fois.

**Remarque 2:** On peut donner, dans le cas unitaire, une preuve plus courte. En effet, pour deux éléments  $h, k$  dans  $Sym(A)$ , on considère la fonction  $f: \lambda \mapsto e^{-\lambda h} k e^{\lambda h}$ .

#### Références:

- [1] Bonsal, F. F., Duncan, J. : "Complete normed algebras", *Ergebnisse der Mathematik, Band 80, Springer Verlag, (1973)*.
- [2] Dixmier, J. : "Les  $C^*$ - algèbres et leurs représentations ". *Paris: Gauthier-Villars, 2<sup>e</sup> édition, (1969)*.
- [3] Kleinecke, D. C. : "On operator commutators." *Proc. Amer. Math. Soc.* **8**, 535-536. (1957).
- [4] Shirokov, F.V. : "Proof of a conjecture of Kaplansky". *Uspehi. Mat. Nauk* **11**, 167-168 (1956).

EXTENDING THE REGION OF CONVERGENCE FOR  
A CERTAIN CLASS OF MODIFIED ITERATIVE METHODS  
ON BANACH SPACES AND APPLICATIONS

Ioannis K. Argyros  
Cameron University  
Department of Mathematics  
Lawton, OK 73505, U.S.A.

**Abstract.** We provide sufficient conditions for the convergence of a certain class of modified iterative methods to a locally unique solution of an equation on a Banach space. In particular we find a ball centered at a certain point that contains the solution and earlier convergent balls. This allows us to access the solution from new starting points. Our results apply to solve a nonlinear integral equation appearing in radiation transfer in connection with the problem of determination of the angular distribution of the radiant flux emerging from a plane radiation field. Relevant work can be found especially in [1], [4], [6], [7].

AMS (MOS) Subject Classification: 65J15, 47H17, 49D15.

Key Words and Phrases: Banach space, contraction mapping principle, region of convergence, modified iterative methods, fixed point.

I. Introduction

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

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

where  $F$  is a Fréchet-differentiable operator defined on a closed convex subset  $D$  of a Banach space  $E_1$  with values in a Banach space  $E_2$ .

We introduce the modified iterative methods of the form

$$x_{n+1} = x_n - A^{-1}F(x_n) \quad (n \geq 0) \quad (2)$$

to generate a sequence  $\{x_n\}$  ( $n \geq 0$ ) converging to  $x^*$ . Here  $A$  is a fixed invertible operator with  $A \in L(E_1, E_2)$  the space of bounded linear operators from  $E_1$  into  $E_2$ . For  $A = F'(x_0)$  we obtain the modified Newton-Kantorovich method [1], [2], [5], where as for  $A = [x_{-1}, x_0; F]$  we obtain the modified secant method [1], [2], [5]. The operator  $[x, y; F]$  is a divided difference of order one of  $F$  at the points  $x, y \in D$  [1], [2], [5]. Several other choices for  $A$  are also possible [2], [4], [6], [7]. The convergence of the last two methods has been extensively studied [1], [2], [5]. Several authors have found a ball  $U(x_0, r) = \{x \in E_1 \mid \|x - x_0\| \leq r\}$  containing the solution and such that for any starting point  $\bar{x}_0$  convergence to  $x^*$  is achieved.

We provide semilocal convergence results for iteration (2) and in the case of the modified Newton-Kantorovich method or the modified secant method we show that  $x^* \in U(x_0, r) \subseteq U(x_0, \bar{r})$ . That is we find a ring containing new starting points  $\bar{x}_0 \in U(x_0, \bar{r}) - U(x_0, r)$  from which  $x^*$  can be accessed via method (2).

Finally we apply our results to solve a nonlinear integral equation appearing in radiative transfer in connection with the problem of determination of the angular distribution of the radiant flux emerging from a plane radiation field [2], [3], [5].

## II. Convergence Analysis

We state and prove the following semilocal convergence result for method (2).

**Theorem 1.** *Let  $F$  be a Fréchet-differentiable operator defined on a closed convex subset  $D$  of a Banach space  $E_1$  with values in a Banach space  $E_2$ , and  $A \in L(E_1, E_2)$  be an invertible linear operator. Assume:*

(a) *for  $x_0 \in D$  there exist  $a \in (1, +\infty)$ ,  $b \in (0, +\infty)$ ,  $\eta \in [0, +\infty)$  such that*

$$\|A^{-1}F(x_0)\| \leq \eta, \quad (3)$$

$$\|A^{-1}(A - F'(x))\| \leq c(r) = a + br, \quad (4)$$

$$4b\eta \leq (1 - a)^2, \quad (5)$$

*for all  $x \in U(x_0, r)$ ,  $r \in [r_0, r_1)$ , where*

$$r_0 = \frac{1 - a - \sqrt{(1 - a)^2 - 4b\eta}}{2b} \quad \text{and} \quad r_1 = \frac{1 - a}{b}. \quad (6)$$

(b)  $U(x_0, r) \subseteq D$  for all  $r \in [r_0, r_1)$ .

Then the following hold: (i) Operator  $P: D \subseteq E_1 \rightarrow E_2$  given by

$$P(x) = x - A^{-1}F(x) \quad (7)$$

has a unique fixed point  $x^*$  in  $U(x_0, r)$ ,  $r \in [r_0, r_1)$ .

(ii) Modified iterative method (2) converges to  $x^*$ .

Moreover the following error bound holds for all  $n \geq 0$ .

$$\|x_n - x^*\| \leq c^n(r)r, \quad r \in [r_0, r_1). \quad (8)$$

**Proof.** By the choice of  $r \in [r_0, r_1)$ , (4) and (7) we obtain

$$P'(x) = I - A^{-1}F'(x) = A^{-1}(A - F'(x)),$$

$$\|P'(x)\| = \|A^{-1}(A - F'(x))\| \leq c(r) < 1, \quad (9)$$

and

$$r \geq \frac{\eta}{1 - c(r)}. \quad (10)$$

It follows from (2) (for  $n = 0$ ) and (10) that

$$\|x_1 - x_0\| \leq \eta \leq (1 - c(r))r \leq r$$

which shows  $x_1 \in U(x_0, r)$ .

Assume

$$x_k \in U(x_0, r) \quad \text{and} \quad \|x_k - x_0\| \leq (1 - c^k(r))r \leq r \quad (11)$$

for  $k = 0, 1, 2, \dots, n$ .

Using (2), (9) and the induction hypothesis (11) we obtain in turn

$$\begin{aligned} \|x_{k+1} - x_k\| &= \|P(x_k) - P(x_{k-1})\| \leq \sup_{y \in [x_{k-1}, x_k]} \|P'(y)\| \|x_k - x_{k-1}\| \\ &\leq c(r)\|x_k - x_{k-1}\| \leq \dots \leq c^k(r)\|x_1 - x_0\| = c^k(r)(1 - c(r))r. \end{aligned} \quad (12)$$

and

$$\begin{aligned}\|x_{k+1} - x_0\| &\leq \|x_{k+1} - x_k\| + \|x_k - x_0\| \leq c^k(t)(1 - c(r))r + (1 - c^k(r))r \\ &= (1 - c^{k+1}(r))r \leq r,\end{aligned}$$

which completes the induction for (11).

By (12) for  $m, n \in N$  we get

$$\|x_{n+m} - x_n\| \leq (1 - c(r))^m c^n(r)r. \quad (13)$$

Estimate (13) shows that  $\{x_n\}$  ( $n \geq 0$ ) is a Cauchy sequence in a Banach space  $E_1$  and as such it converges to some  $x^*$ . Letting  $n \rightarrow \infty$  in (2) and using the continuity of  $A$  and  $F$  we obtain  $P(x^*) = x^*$  and  $F(x^*) = 0$ .

To show uniqueness let us assume that  $y^* \in U(x_0, r)$  with  $P(y^*) = 0$ . Then by (9) we get

$$\begin{aligned}\|x^* - y^*\| &= \|P(x^*) - P(y^*)\| \leq \sup_{z \in [x^*, y^*]} \|P'(z)\| \|x^* - y^*\| \\ &\leq c(r) \|x^* - y^*\|,\end{aligned}$$

which shows  $x^* = y^*$ .

Finally, by letting  $m \rightarrow \infty$  in (13) we obtain estimate (8).

That completes the proof of Theorem 1.

The following result is a consequence of the contraction mapping principle [2], [5].

**Theorem 2.** Under the assumptions of Theorem 1, the modified iterative method

$$y_{n+1} = P(y_n) \quad (n \geq 0) \quad (14)$$

converges to the unique fixed point  $x^*$  of  $P$  in  $U(x_0, r)$   $r \in [r_0, r_1)$  provided that  $y_0 \in U(x_0, r)$ . Moreover the following error bounds hold for all  $n \geq 0$

$$\|y_n - x^*\| \leq \frac{c^n(r)}{1 - c(r)} \|y_1 - y_0\|$$

and

$$\|y_n - x^*\| \leq c^n(r) \|x^* - y_0\|.$$

**Proof.** For  $x \in U(x_0, r)$ ,  $r \in [r_0, r_1)$  we obtain

$$\|y_1 - P(x)\| = \|P(y_0) - P(x)\| \leq c(r) \|x - x_0\| \leq c(r)r.$$

Hence  $P: U(x_0, r) \rightarrow U(x_0, r)$  is a  $c(r) < 1$  contraction.

The result now follows from the contraction mapping principle [5, p. 146], [2].

That completes the proof of Theorem 2.

**Remark 1.** (Modified Newton-Kantorovich Method). For  $A = F'(x_0)$ , set  $a = 0$  in (4). Theorem 1 compares favorably to Theorem 12.2 in [5, p. 146], which guarantees convergence of method (14) in this case to  $x^*$  provided that  $y_0 \in U(x_0, r_0)$ . However our Theorem 1 guarantees convergence of method (14) to  $x^*$  provided that  $y_0 \in U(x_0, r)$  and  $U(x_0, r_0) \subseteq U(x_0, r)$ ,  $r \in [r_0, r_1)$ . That is we found infinitely many new starting points  $y_0 \in U(x_0, r) - U(x_0, r_0)$  from which the solution  $x^*$  can be accessed. Moreover the uniqueness of the solutions is guaranteed in a larger ball since

$$r_2 = \frac{1 + \sqrt{1 - 4b\eta}}{2b} < \frac{1}{b} = r_1.$$

Finally our conditions (9) are weaker than the corresponding conditions (12.7) in [5, p. 146], [4].

**Remark 2.** (Modified Secant Method). For  $A = [x_{-1}, x_0; F]$ , assume  $F'(x) = [x, x; F]$  ( $x \in D$ ),  $\|A^{-1}(A - [x, x; F])\| \leq K(\|x - x_{-1}\| + \|x - x_0\|)$  and  $\|x_0 - x_{-1}\| \leq d$ ,  $x_0 \neq x_{-1}$ . Then we can set  $a = Kd$  and  $b = 2K$ . A favorable comparison similar to the one given in Remark 1 can be provided between our Theorem 1 and the corresponding theorem on the secant method which can be found, e.g., in [5, p. 146], [2]. The computational details are left to the motivated reader.

### III. Applications

We complete this study with two examples:

**Example 1.** Let  $E_1 = E_2 = R$ ,  $D = [0, 2]$  and define  $F: D \rightarrow E_2$  by

$$F(x) = x^2 - \frac{1}{4}.$$

Take  $x_0 = .6$  to obtain  $a = 0$ ,  $b = 2$ ,  $\eta = .0916667$ ,  $r_0 = .1209005$ ,  $r_1 = .5$ ,  $U(x_0, r_0) = [.4790995, .7209005]$  and  $U(x_0, r)$  can be as large as  $(.1, 1.1)$ . That is we can choose as a starting point in (14)  $y_0 = 1 \notin U(x_0, r_0)$  and  $\{y_n\}$  ( $n \geq 0$ ) will converge to  $x^* = .5$  (according to Theorem 2). Indeed using (14) we obtain

$$\begin{aligned} y_1 &= .375, & y_2 &= .559375, & y_3 &= .506958, & y_4 &= .501193, \\ y_5 &= .5001855, & y_6 &= .5000309, & y_7 &= .5000052, & y_8 &= .5000009, \\ y_9 &= .5000002 & \text{and} & & y_{10} &= .5 = x^*. \end{aligned}$$

**Example 2.** Let us consider the quadratic integral equation of the form

$$x(s) = y(s) + \lambda x(s) \int_0^1 q(s, t)x(t)dt \quad (15)$$

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

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

We assume that  $\lambda$  is a given real number and the kernel  $q(s, t)$  is a continuous function of two variables  $s, t$  satisfying

$$0 < q(s, t) < 1, \quad s, t \in (0, 1), \quad q(0, 0) = 1 \quad (16)$$

and

$$q(s, t) + q(t, s) = 1, \quad s, t \in [0, 1]. \quad (17)$$

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

Equations of this type are closely related to the work of S. Chandrasekhar [3] (Nobel prize of Physics, 1983), and arise in the theories of radioactive transfer, neutron transport and in the kinetic theory of gases [2], [3]. In particular, these equations are used to solve the problem of determination of the angular distribution of the radiant flux emerging from a plane radiation field. This radiation field must be isotropic at a point, that is, the distribution is independent of direction at that point. The positive parameter  $\lambda$  is called the albedo for scattering. It represents the fraction of the radiation lost due to scattering on a plane radiation field.

For simplicity we assume that

$$q(s, t) = \frac{s}{s+t} \quad \text{for all } s, t \in (0, 1) \text{ and } q(0, 0) = 1.$$

The function  $q$  so defined satisfies (16) and (17). Define the operator  $F$  on  $E_1$  by

$$F(x) = \lambda x(s) \int_0^1 \frac{s}{s+t} x(t)dt - x(s) + y(s). \quad (18)$$

Note that every zero  $x^*$  of the equation  $F(x) = 0$  satisfies (15).

Set  $y(s) = x_0(s) = 1$ ,  $\lambda = .25$  and by using the definition of the first and second Fréchet-derivative of the operator  $F$  we obtain in turn

$$a = 0, \quad b = 2|\lambda| \|A^{-1}\| \cdot \max_{0 \leq s \leq 1} \left| \int_0^1 \frac{s}{s+t} dt \right| = 2|\lambda| \ln 2 \|F'(1)^{-1}\| = .530394201,$$

$$\eta = .265197107, \quad 4b\eta = .562636021 < 1,$$

$$r_0 = .319257803 \quad \text{and} \quad r_1 = 1.8853901.$$

The hypotheses of our Theorems 1 and 2 are satisfied. Hence the conclusions follow. In particular there exist infinitely many new starting points  $y_0 \in U(x_0, r)$ ,  $r \in [r_0, r_1]$  for which method (14) converges to  $x^*$ . However this is not guaranteed by the Newton-Kantorovich Theorem (see the comments made in Remark 1).

### Conclusion

In this study we gave sufficient conditions for the convergence of a certain class of modified iterative methods to a locally unique solution of an equation on a Banach space setting. We found a ring containing infinitely many new starting points for which our iterative method converges to the solution. However the Newton-Kantorovich Theorem for the modified Newton-Kantorovich method does not guarantee convergence. We also showed uniqueness of the solution in a ball larger than the ones in earlier results. We applied our results to solve a class of nonlinear integral equations appearing in radiative transfer in connection with the problem of determination of the angular distribution of the radiant flux emerging from a plane radiation field. Relevant work can be found especially in [1], [4], [5], [6], [7].

### References

- [1] Argyros, I.K. A convergence theorem for Newton-like methods under generalized Chen-Yamamoto-type assumptions, *Appl. Math. Comp.* 61, 1, (1994), 25-37.
- [2] Argyros, I.K. and Szidarovszky, F. *The Theory and Applications of Iteration Methods*, C.R.C. Press Inc., Boca Raton, Florida, U.S.A., 1993.
- [3] Chandrasekhar, S. *Radiative Transfer*, Dover Publ., New York, 1960.
- [4] Gutierrez, J.M., Hernandez, M.A., and Salanova, M.A. Accessibility of solutions by Newton's method, *Intern. J. Computer Math.* 57, (1995), 239-247.
- [5] Krasnosel'skii, M.A., Vainikko, G.M., and Zabrejko, P.P. *Approximate Solution of Operator Equations*, Walters-Noordhoff Publ. Groningen, Jerusalem, 1972.
- [6] Liu, D. and Szidarovszky, F. Global asymptotic stability of dynamic systems with modified contractions, *Appl. Math. Comp.* 43, 3, (1991), 237-240.
- [7] Szidarovszky, F. and Palusinski, O. A special matrix equation and its applications in microelectronics, *Appl. Math. Comp.* 64, (1994), 115-119.

RESULTS INVOLVING NONDIFFERENTIABLE EQUATIONS ON  
BANACH SPACES WITH A CONVERGENCE STRUCTURE  
AND NEWTON METHODS

Ioannis K. Argyros

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

**Abstract.** In this study we use Newton methods to find solutions of nonlinear, nondifferentiable operator equations on Banach spaces with a convergence structure. This technique involves the introduction of a generalized norm as an operator from a linear space into a partially ordered Banach space. In this way the metric properties of the examined problem can be analyzed more precisely. Moreover, this approach allows us to derive from the same theorem, on the one hand, semi-local results of Kantorovich-type, and on the other hand, global results based on monotonicity considerations. Furthermore, we show that special cases of our results reduce to corresponding ones already in the literature. Finally, our results are used to solve applied problems that cannot be solved with existing methods.

AMS (MOS) (1991) Subject Classification Codes: 65J15, 65B05, 47H17, 49D15.

Key Words and Phrases: Banach space, Newton methods, nondifferentiable operator.

I. INTRODUCTION

In this report we are concerned with approximating a solution of a nonlinear operator equation

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

with

$$P(x) = F(x) + Q(x), \quad (2)$$

where  $F$  is a Fréchet-differentiable nonlinear operator defined on a convex subset  $D$  of a Banach space  $X$  with values in  $X$ , and  $Q$  is a nondifferentiable nonlinear operator with the same domain and values in  $X$ .

We introduce the Newton method

$$x_{n+1} = x_n + F'(x_n)^{-1}[-P(x_n)], \quad x_0 = 0, \quad n \geq 0 \quad (3)$$

to approximate a solution  $x^*$  of equation (1). Here  $F'(x_n)^{-1}$  ( $n \geq 0$ ) denotes a linear operator which is an approximation for  $F'(x_n)^{-1}$  ( $n \geq 0$ ). The notion of a Banach space with a convergence structure was used in the elegant paper in [5] (see also [1], [2], [3], [7], [8]) to solve (1), when  $Q(x) = 0$  in (1), (2) and (3) for all  $x \in D$ . However, there are many interesting real life applications already in the literature, where  $P$  contains a nondifferentiable term. That is why we are motivated to report on this case [3], [9].

As in [6], we provide semi-local results of Kantorovich-type and global results based on monotonicity considerations from the same general theorem.

Moreover, we show that our results can be reduced to the ones obtained in [6] when  $Q(x) = 0$  for all  $x \in D$ , and furthermore, to the ones obtained in [1], [2], [3], [6], [7], [8] by further relaxing the requirements on  $X$ .

Finally, our results are used to solve problems that cannot be approached with existing methods.

## II. PRELIMINARIES

We will need the definitions:

**Definition 1.** The triple  $(X, V, E)$  is a Banach space with a convergence structure if

- (C<sub>1</sub>)  $(X, \|\cdot\|)$  is a real Banach space.
- (C<sub>2</sub>)  $(V, C, \|\cdot\|_V)$  is a real Banach space which is partially ordered by the closed convex cone  $C$ ; the norm  $\|\cdot\|_V$  is assumed to be monotone on  $C$ .
- (C<sub>3</sub>)  $E$  is a closed convex cone in  $X \times V$  satisfying  $\{0\} \times C \subseteq E \subseteq X \times C$ .
- (C<sub>4</sub>) The operator  $\| : D_0 \rightarrow C$  is well defined:

$$|x| = \inf\{q \in C \mid (x, q) \in E\}$$

for

$$x \in D_0 = \{x \in X \mid \exists q \in C : (x, q) \in E\}.$$

and

- (C<sub>5</sub>) For all  $x \in D_0 : \|x\| \leq \|x\|_V$ .

The set

$$U(a) = \{x \in X \mid (x, a) \in E\}$$

defines a sort of generalized neighborhood of zero.

Let us give the following motivational examples for  $X := \mathbb{R}^m$  with the maximum-norm:

- (a)  $V := \mathbb{R}, E := \{(x, e) \in \mathbb{R}^m \times \mathbb{R} \mid \|x\|_\infty \leq e\}$ .
- (b)  $V := \mathbb{R}^m, E := \{(x, e) \in \mathbb{R}^m \times \mathbb{R}^m \mid |x| \leq e\}$   
(componentwise absolute value).
- (c)  $V := \mathbb{R}^n, E := \{(x, e) \in \mathbb{R}^m \times \mathbb{R}^m \mid 0 \leq x \leq e\}$ .

Case (a) involves classical convergence analysis in a Banach space, Case (b) concerns componentwise analysis and error estimates, and (c) is used for monotone convergence analysis.

The proofs of the results stated in Section III can be provided upon request.

**Definition 2.** Let  $C^1(V_1 \rightarrow V)$  denote the set of all continuously differentiable operators defined on an open convex subset  $V_1$  of an ordered Banach space  $V$  with values in  $V$ . An operator  $L \in C^1(V_1 \rightarrow V)$  is order convex on  $[a, b] \subseteq V_1$  if

$$c, d \in [a, b], c \leq d \Rightarrow L'(d) - L'(c) \in L_+(V),$$

where for  $m \geq 0$

$$L_+(V^m) = \{L \in L(V^m) \mid 0 \leq x_i \Rightarrow 0 \leq L(x_1, x_2, \dots, x_m)\}$$

and  $L(V^m)$  denotes the space of  $m$ -linear, symmetric, bounded operators on  $V$ .

**Definition 3.** The set of bounds for an operator  $H \in L(X^m)$  is defined to be

$$B(H) = \{L \in L_+(V^m) \mid (x_i, q_i) \in E \Rightarrow (H(x_1, \dots, x_m), L(q_1, \dots, q_m)) \in E\}.$$

**Definition 4.** Let  $H \in L(X)$  and  $y \in X$  be given, then

$$H^*(y) = z \Leftrightarrow z = T^\infty(0) = \lim_{n \rightarrow \infty} T^n(0),$$

$$T(x) = (I - H)(x) + y \Leftrightarrow z = \sum_{i=0}^{\infty} (I - H)^i y,$$

if this limit exists.

We will also need the Lemmas [2], [3], [5], [6]:

**Lemma 1.** Let  $L \in L_+(V)$  and  $a, q \in C$  be given such that:

$$L(q) + a \leq q \text{ and } L^n(q) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Then the operator

$$(I - L)^* : [0, a] \rightarrow [0, a]$$

is well defined and continuous.

The following is a generalization of Banach's lemma [3], [5], [6]:

**Lemma 2.** Let  $H \in L(X)$ ,  $L \in B(H)$ ,  $y \in D_0$  and  $q \in C$  be such that

$$L(q) + |y| \leq q \text{ and } L^n(q) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Then the points  $x = (I - H)^*(y)$  is well defined,  $x \in S$  and

$$|x| \leq (I - L)^* |y| \leq q.$$

Moreover, the sequence

$$b_{n+1} = L(b_n) + |y|, \quad b_0 = 0$$

is well defined and

$$b_{n+1} \leq q, \quad \lim_{n \rightarrow \infty} b_n = b = (I - L)^* |y| \leq q.$$

**Lemma 3.** Let  $H_1 : [0, 1] \rightarrow L(X^m)$  and  $H_2 : [0, 1] \rightarrow L_+(V^m)$  be continuous operators, then for all  $t \in [0, 1]$ :

$$H_2(t) \in B(H_1(t)) \Rightarrow \int_0^1 H_2(t) dt \in B \left( \int_0^1 H_1(t) dt \right)$$

which will be used for the remainder of Taylor's formula [5].

The convergence analysis will be based on monotonicity considerations in the space  $X \times V$ . Let  $(x_n, e_n)$  be an increasing sequence in  $E^N$ , then

$$(x_n, e_n) \leq (x_{n+k}, e_{n+k}) \Rightarrow 0 \leq (x_{n+k} - x_n, e_{n+k} - e_n).$$

If  $e_n \rightarrow e$ , we obtain:  $0 \leq (x_{n+k} - x_n, e - e_n)$  and hence by (C<sub>5</sub>)

$$\|x_{n+k} - x_n\| \leq \|e - e_n\|_V \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Hence  $\{x_n\}$  ( $n \geq 0$ ) is a Cauchy sequence. When deriving error estimates, we shall as well use sequences  $e_n = w_0 - w_n$  with a decreasing sequence  $\{w_n\}$  ( $n \geq 0$ ) in  $C^N$  to obtain the estimate

$$0 \leq (x_{n+k} - x_n, w_n - w_{n+k}) \leq (x_{n+k} - x_n, w_n).$$

If  $x_n \rightarrow x^*$  as  $n \rightarrow \infty$  this implies the estimate  $|x^* - x_n| \leq w_n$ . Note also that for  $(x, e) \in E$ ,  $x \in D_0$  and by (C<sub>4</sub>)  $|x| \leq e$ .

### III. CONVERGENCE ANALYSIS

We can now state the main result:

**Theorem 1.** Let  $X$  be a Banach space with convergence structure  $(X, V, E)$  with  $V = (V, C, \|\cdot\|_V)$ , an operator  $F \in C^1(D \rightarrow X)$  with  $D \subseteq X$ , an operator  $Q \in C(D \rightarrow X)$ , an operator  $L \in C^1(V_1 \rightarrow V)$  with  $V_1 \subseteq V$ , an operator  $M \in C(V_1 \rightarrow V)$ , and a point  $a \in C$  such that the following conditions are satisfied:

(C<sub>6</sub>) The inclusion  $U(a) \subseteq D$  and  $[0, a] \subseteq V_1$  are true;

(C<sub>7</sub>)  $L$  is order-convex on  $[0, a]$  and

$$L'(|x| + |y|) - L'(|x|) \in B(F'(x) - F'(x + y))$$

for all  $x, y \in U(a)$  with  $|x| + |y| \leq a$ ;

(C<sub>8</sub>)  $0 \leq (Q(x) - Q(x + y), M(|x| + |y|) - M(|x|)) \in E$  for all  $x, y \in U(a)$  with  $|x| + |y| \leq a$ , and

$$M(w_1 + w_2) - M(w_1) \leq M(w_3 + w_4) - M(w_3), \quad M(w) \geq 0$$

for all  $w, w_1, w_2, w_3, w_4 \in [0, a]$  with  $w_1 \leq w_3$  and  $w_2 \leq w_4$ .

(C<sub>9</sub>)  $L'(0) \in B(I - F'(0))$  and  $-(F'(0) + Q(0)), L(0) + M(0) \in E$ ;

(C<sub>10</sub>)  $(L + M)(a) \leq a$ ,  $0 \leq L + M$ ,  
and

(C<sub>11</sub>)  $(L'(a) + M)^n a \rightarrow 0$  as  $n \rightarrow \infty$ .

Then the sequence  $(x_n, d_n) \in (X \times V)^N$  is well defined, remains in  $E^N$ , is monotone and satisfies for all  $n \geq 0$

$$d_n \leq b,$$

where

$$d_{n+1} = (L + M)(d_n) + L'(|x_n|)c_n, \quad d_0 = 0, \quad c_n = |x_{n+1} - x_n|$$

and

$$b = (L + M)^\infty(0)$$

is the smallest fixed point of  $L + M$  in  $[0, a]$ .

Moreover, the sequence  $\{x_n\}$  ( $n \geq 0$ ) generated by (3) converges to a solution  $x^* \in U(b)$  of the equation  $P(x) = 0$ , which is unique in  $U(a)$ .

Furthermore, the following estimates are true for all  $n \geq 0$ :

$$b_n \leq d_n \leq b,$$

$$b_n \leq a_n,$$

$$|x_{n+1} - x_n| \leq d_{n+1} - d_n,$$

$$|x_n - x^*| \leq b - d_n,$$

and

$$|x_n - x^*| \leq a_n - b_n$$

where

$$b_n = (L + M)^n(0) \quad \text{and} \quad a_n = (L + M)^n(a).$$

We will now introduce results on a posteriori estimates. It is convenient to define the operator

$$R_n(q) = (I - L'(|x_n|))S_n(q) + c_n$$

where

$$S_n(q) = (L + M)(|x_n| + q) - (L + M)(|x_n|) - L'(|x_n|)(q)$$

and the interval

$$I_n = [0, a - |x_n|].$$

Then we can show:

**Proposition.** *The following implications are true:*

(i) *If  $q \in I_n$  satisfy  $R_n(q) \leq q$  then*

$$c_n \leq R_n(q) = p \leq q$$

and

$$R_{n+1}(p - c_n) \leq p - c_n \quad \text{for all } n \geq 0.$$

(ii) *Under the hypotheses of Theorem 1, let  $q_n \in I_n$  be a solution of  $R_n(q) \leq q$  then*

$$|x^* - x_m| \leq a_m \quad (m \geq n)$$

where

$$a_n = q_n \quad \text{and} \quad a_{m+1} = R_m(a_m) - c_m.$$

and

(iii) *Under the hypotheses of Theorem 1, any solution  $q \in I_n$  of  $R_n(q) \leq q$  is such that*

$$|x^* - x_n| \leq R_n^\infty(0) \leq q.$$

**Remarks 1.** (a) The results obtained in Theorem 1 and the proposition reduce immediately to the corresponding one in [6] (Theorem 5 and Lemmas 10-12) when  $Q(x) = 0$  for all  $x \in D$  ( $M$  will be 0 then).

(b) As in [6], we can also show that if conditions  $(C_6)$ - $(C_9)$  are satisfied and there exists  $t \in (0, 1)$  such that  $(L + M)(a) \leq ta$ , then there exists  $a_1 \in [0, ta]$  satisfying conditions  $(C_6)$ - $(C_{11})$ . The solution  $x^* \in U(a_1)$  is unique in  $U(a)$ .

#### IV. THE MONOTONE CASE

Let  $A \in L(X \rightarrow X)$  be a given operator. Define the operators  $P, T(D \rightarrow X)$  by  $P(x) = AT(x + u)$ ,  $T(x) = G(x) + R(x)$ ,  $P(x) = F(x) + Q(x)$ ,  $F(x) = AG(x + u)$ , and  $Q(x) = AR(x + u)$ , where  $G, R$  are as  $F, Q$  respectively. We deduce immediately that under the hypotheses of Theorem 1, the zero  $x^*$  of  $P$  is a zero of  $AT$  also if  $u = 0$ .

We will now provide a monotonicity result to find a zero  $x^*$  of  $AT$ . The space  $X$  is assumed to be partially ordered and satisfies the conditions for  $V$  given in  $(C_1)$ - $(C_5)$ . Moreover, we set  $X = V$ ,  $D = C^2$  so that  $|\cdot|$  turns out to be  $I$ .

We can now state the main result of this section:

**Theorem 2.** *Let  $V$  be a partially ordered Banach space satisfying conditions  $(C_1)$ - $(C_5)$ ,  $Y$  a Banach space,  $G \in C^1(D \rightarrow Y)$ ,  $R \in C(D \rightarrow Y)$  with  $D \subseteq X$ ,  $A \in L(X \rightarrow V)$ ,  $M \in C(D \rightarrow V)$  with  $M \in L_+(V)$  and  $u, v \in X$  such that:*

$$(C_{12}) \quad [u, v] \subseteq D;$$

$$(C_{13}) \quad I - AG'(u) + M \in L_+(V);$$

$$(C_{14}) \quad \text{For all } w_1, w_2 \in [u, v] : w_1 \leq w_2 \Rightarrow AG'(w_1) \geq AG'(w_2);$$

$$(C_{15}) \quad AT(u) \leq 0, -Q(0) \leq M(0) \quad \text{and} \quad AT(v) \geq 0;$$

(C<sub>16</sub>) Condition (C<sub>3</sub>) is satisfied and  $M(v-u) \leq -Q(v-u)$ ; and

(C<sub>17</sub>)  $(I - AG'(v) + M)^n(v-u) \rightarrow 0$  as  $n \rightarrow \infty$ .

Then the Newton sequence

$$z_0 = u, \quad z_{n+1} = z_n + (AG'(z_n))^{-1}[-AT(z_n)]$$

is well defined for all  $n \geq 0$ , monotone and converges to a unique zero  $x^*$  of  $AT$  in  $[u, v]$ .

*Proof.* The proof follows immediately from Theorem 1 by setting

$$P(x) = AT(x+u), \quad L(q) = q - F(q) \quad \text{and} \quad a = v - u.$$

**Remarks 2.** (a) If  $R$  is a linear operator then condition (C<sub>16</sub>) will be satisfied if we set  $M(q) = -Q(q)$ .

(b) If  $M$  is such that  $0 \leq M - AG'(v)$ , then condition (C<sub>17</sub>) can be replaced by the weaker

(C<sub>17</sub>)  $(M - AG'(v))^n(v-u) \rightarrow 0$  as  $n \rightarrow \infty$ .

(c) By setting  $R(x) = 0$  for all  $x \in D$  we immediately deduce that our results reduce to the ones obtained in [6, Thm. 13].

## V. APPLICATIONS

We will complete this report with two examples that show how to choose  $L, M$  in practical applications.

**Example 1.** We discuss the case of a real Banach space with norm  $\|\cdot\|$ . Assume that  $F'(0) = I$  and there exists a monotone operator

$$f : [0, a] \rightarrow \mathbf{R} \quad \text{such that} \quad \|F''(x)\| \leq f(\|x\|) \quad \text{for all } x \in U(a),$$

and a continuous, nondecreasing function  $g$  on  $[0, r]$ ,  $r \leq a$  such that

$$\|Q(x) - Q(y)\| \leq g(r)\|x - y\| \quad \text{for all } x, y \in U\left(\frac{r}{2}\right). \quad (4)$$

We showed in [3], (see also [9]), that (4) implies that

$$\|Q(x+t) - Q(x)\| \leq h(r + \|t\|) - h(r), \quad x \in U(a), \quad \|t\| \leq a - r, \quad (5)$$

where

$$h(r) = \int_0^r g(t) dt.$$

Conversely, it is not hard to see that we may assume, without loss of generality, that the function  $h$  and all functions  $h(r+t) - h(r)$  are monotone in  $r$ . Hence we may assume that  $h(r)$  is convex, and hence differentiable from the right. Then as in [9] we show that (5) implies (4) and  $g(r) = h'(r+0)$ .

Hence, we can now set

$$L(q) = \|F(0) + Q(0)\| + \int_0^q ds \int_0^s f(t) dt \quad (6)$$

and

$$M(q) = \int_0^q g(t) dt + \|Q(0)\|. \quad (7)$$

Then condition (C<sub>10</sub>) will be true if

$$\frac{1}{2}f(a)a^2 - (1-g(a))a + \|F(0)\| + \|Q(0)\| \leq 0. \quad (8)$$

If we set  $Q = 0$  and  $g = 0$ , (8) is true if  $\|F(0)\|f(a) \leq \frac{1}{2}$ , which is a well known condition due to Kantorovich [4, Ch. 18]. If  $Q \neq 0$ , condition (8) is the same condition with the one found in [9], [3] for the Zinchenko iteration.

In the example that follows, we show that our results can apply to solve nonlinear integral equations involving a nondifferentiable term, whereas the results obtained in [6] (or in [5], [7], [8]) cannot apply.

**Example 2.** Let  $X = V = C[0, 1]$ , and consider the integral equation

$$x(t) = \int_0^1 k(t, s, x(s)) ds \quad \text{on } X, \quad (9)$$

where the kernel  $k(t, s, x(s))$  with  $(t, s) \in [0, 1] \times [0, 1]$  is a nondifferentiable operator on  $X$ . Consider (9) in the form

$$P(x)(t) = 0, \quad (10)$$

where  $P, F, Q : X \rightarrow X$  are given by

$$P(x)(t) = x(t) - \int_0^1 k(t, s, x(s)) ds \quad (11)$$

$$F(x)(t) = Ix(t) \quad (12)$$

and

$$Q(x)(t) = - \int_0^1 k(t, s, x(s)) ds. \quad (13)$$

The operator  $\|$  is defined by considering the sup-norm. We assume that  $V$  is equipped with natural partial ordering, and there exists  $\alpha, a \in [0, +\infty)$  and a real function  $\alpha(t, s)$  such that

$$\|k(t, s, x) - k(t, s, y)\| \leq \alpha(t, s)\|x - y\| \quad (14)$$

for all  $t, s \in [0, 1]$ ,  $x, y \in U(\frac{a}{2})$ , and

$$\alpha \geq \sup_{t \in [0, 1]} \int_0^1 \alpha(t, s) ds.$$

Define the real functions  $h, f, g$  on  $[0, a]$  by

$$h(r) = \alpha r, \quad f(r) = 0 \quad \text{and} \quad g(r) = \alpha \quad \text{for all } r \in [0, a].$$

By choosing  $L, M$  as in (6) and (7) respectively and using (11), (12) and (14), we can easily see that all conditions of Theorem 1 are satisfied. In particular condition (8) reduces to

$$(1 - \alpha)a - \|Q(x)\| \geq 0, \quad (15)$$

which is true in the following cases: if  $0 \leq \alpha < 1$ , choose  $a \geq \beta = \frac{\|Q(0)\|}{1 - \alpha}$ ; if  $\alpha = 1$  and  $Q(0) = 0$ , choose  $a \geq 0$ ; if  $\alpha > 1$  and  $Q(0) = 0$ , choose  $a = 0$ . If in (15) the strict inequality is valid, then there exists a solution,  $a^*$  of inequality (15) satisfying condition (C<sub>11</sub>). Note that if we choose  $\alpha \in [0, 1)$  and  $a \in (\beta, +\infty)$ , condition (15) is valid as a strict inequality.

Finally we remark that the results obtained in [6] (or in [5], [7], [8]) cannot apply here to solve equation (10), since  $Q$  is nondifferentiable on  $X$ .

## CONCLUSION

In this study, we used Newton methods to find solutions of nonlinear, nondifferentiable operator equations on Banach spaces with a convergence structure. This technique involves the introduction of a generalized norm as an operator from a linear space into a partially ordered Banach space. This way the metric

properties of the examined problem have been shown to be analyzed more precisely. Convergence results and error estimates have been improved compared with the real norm theory. Moreover, this approach also allowed us to derive, on the one hand, semi-local results of Kantorovich-type and, on the other hand, global results based on monotonicity considerations, from the same general theorem. All our convergence theorems were derived from metric estimates. Furthermore, we show that special cases of our results reduce to the corresponding ones already in the literature [1], [2], [5], [6], [7]. Finally, we used our results to solve applied problems that cannot be solved with existing methods.

#### REFERENCES

- [1] Argyros, I.K. The secant method in generalized Banach spaces, *Appl. Math. Comp.* **39**, (1990), 111-121.
- [2] Argyros, I.K. Newton-like methods in generalized Banach spaces, *Functiones et Approximatio Comen-tarii Mathematici*, **XXII**, (1993), 107-114.
- [3] Argyros, I.K. and Szidarovszky, F. *The Theory and Applications of Iteration Methods*, C.R.C. Press, Inc. Boca Raton, Florida, U.S.A., 1993.
- [4] Kantorovich, L.V. and Akilov, G.P. *Functional Analysis in Normed Spaces*, Academic Press, New York, 1978.
- [5] Meyer, P.W. Newton's method in generalized Banach spaces, *Numer. Funct. Anal. Optimiz.* **9**, (3 and 4), (1987), 249-259.
- [6] Meyer, P.W. A unifying method for Newton's method, *Numer. Funct. Anal. Optimiz.* **13** (5 and 6), (1992), 463-473.
- [7] Pandian, M.C. A convergence test and component-wise error estimates for Newton-type methods, *SIAM J. Numer. Anal.* **22**, (1985), 779-799.
- [8] Vandergraft, J.A. Newton's method for convex operators in partially ordered spaces, *SIAM J. Numer. Anal.* **4**, (1967), 406-432.
- [9] Zabrejko, P.P. and Nguen, D.F. The majorant method in the theory of Newton-Kantorovich approxi-mations and the Ptak error estimates, *Numer. Funct. Anal. Optimiz.* **9** (5 and 6), (1987), 671-684.

## THE HALLEY METHOD IN BANACH SPACES AND THE PTAK ERROR ESTIMATES

Ioannis K. Argyros

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

**Abstract.** Ptak error estimates are obtained for the Halley method in Banach spaces using the majorant method.

AMS (MOS) (1991) Subject Classification: 47H17, 65H10, 65J15, 49D15.

Key Words and Phrases: Halley method, Banach space, closed graph theorem.

### I. INTRODUCTION

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

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

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

Ptak, Potra [16], Zabrejko, Nguen [22] and Argyros [4], [9] have used the method of "continuous induction" to find error estimates for Newton's method as well as Newton-like methods under various assumptions. This method is based on a special variant of Banach's closed graph theorem.

Let  $x_0 \in E_1$  be arbitrary and define the Halley method on  $E_1$  for all  $n \geq 0$  by

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

$$H(x_n, y_n) = -F'(x_n)^{-1}F''(x_n)(y_n - x_n) \quad (3)$$

and

$$x_{n+1} = y_n - \frac{1}{2}F'(x_n)^{-1} \left[ I - \frac{1}{2}H(x_n, y_n) \right]^{-1} F''(x_n)(y_n - x_n)^2. \quad (4)$$

Here  $F'(x_n)$  and  $F''(x_n)$  denote the first and second Fréchet-derivatives of  $F$  evaluated at  $x = x_n$  ( $n \geq 0$ ). Note that  $F'(x_n)$  is a linear operator whereas  $F''(x_n)$  is a bilinear operator for all  $n \geq 0$  [2], [3]. If the sequence  $\{x_n\}$  ( $n \geq 0$ ) generated by (2)-(4) converges to a limit  $x^*$ , then  $x^*$  is a zero of equation (1) (if  $F'(x)$  and  $F$  are continuous on  $D$ ). Halley's method has a very long history. One can refer to [5], [6], [9], [12], [14], [15], [18], [21] and the references there for some background.

Using the majorant method we will show that under very simple and workable Newton-Kantorovich assumptions on the pair  $(F, x_0)$  the Halley method converges to a locally unique zero  $x^*$  of equation (1). We also provide Ptak error estimates on the distances  $\|x_n - x^*\|$  for all  $n \geq 0$ .

Finally, we study the problem when  $F$  is not Fréchet-differentiable. Our results compare favorably with the ones in [11]-[22] and the references there.

## II. Convergence Analysis

Let  $F: D \subset E_1 \rightarrow E_2$ , where  $D$  is a closed convex domain and  $x_0 \in D$ . We assume that the Fréchet-derivatives  $F', F''$  of  $F$  satisfy

$$\|F'(x_1) - F'(x_2)\| \leq q_1(r)\|x_1 - x_2\|, \quad \|F''(x)\| \leq M, \quad (5)$$

and

$$\|F''(x_1) - F''(x_2)\| \leq q_2(r)\|x_1 - x_2\| \quad (6)$$

for all  $x_1, x_2 \in U(x_0, r) = \{x \in E_1 \mid \|x - x_0\| \leq r\}$ . The functions  $q_1$  and  $q_2$  are nondecreasing, and defined on  $[0, R]$  for some  $R \geq 0$ .

It is convenient to introduce the constants

$$\eta \geq \|y_0 - x_0\|, \quad \beta \geq \|F'(x_0)^{-1}\|, \quad t_0 = 0, \quad s_0 \geq \eta, \quad t_1 \geq s_0^* = \frac{\beta M \eta^2}{2 - \beta M \eta}, \quad (7)$$

the iterations

$$s_{n+1} = t_{n+1} + \frac{\beta}{1 - \beta \omega_1(t_{n+1})} A(\omega_1, \omega_2, t_n, s_n), \quad (8)$$

$$A(\omega_1, \omega_2, t_n, s_n) = \int_{s_n}^{t_{n+1}} \omega_1(t) dt - \omega_1(s_n)(t_{n+1} - s_n) + (\omega_1(s_n) - \omega_1(t_n))(t_{n+1} - s_n) \\ + \frac{1}{2} c(t_n) B(t_n) M (s_n - t_n) \left[ \int_{t_n}^{s_n} \omega_1(t) dt - \omega_1(t_n)(s_n - t_n) \right] + \frac{1}{2} c(t_n) \left[ \int_{t_n}^{s_n} \omega_2(t) dt - \omega_2(t_n)(s_n - t_n) \right], \quad (9)$$

$$B(t_n) = \frac{\beta}{1 - \beta \omega_1(t_n)}, \quad (10)$$

$$c(t_n) = \left[ 1 - \frac{1}{2} \frac{\beta M}{1 - \beta \omega_1(t_n)} (s_n - t_n) \right]^{-1}, \quad (11)$$

$$t_{n+2} = s_{n+1} + \frac{1}{2} \frac{\beta M}{1 - \beta \omega_1(t_{n+1})} c(t_{n+1})(s_{n+1} - t_{n+1})^2, \quad (12)$$

where

$$\omega_1(r) = \int_0^r q_1(t) dt \quad \text{and} \quad \omega_2(t) = \int_0^r q_2(t) dt. \quad (13)$$

Moreover, we define the function  $T$  on  $[0, R]$  by

$$T(r) = t_1 + \int_0^r \omega_1(t) dt + \omega_1(r)r + \frac{M}{2} c(r) B(r) r \int_0^r \omega_1(t) dt \\ + \frac{1}{2} c(r) \int_0^r \omega_2(t) dt + \frac{M}{2} B(r) c(r) r^2. \quad (14)$$

We can now prove the main result:

**Theorem 1.** Let  $F: D \subset E_1 \rightarrow E_2$  be a nonlinear operator defined on some closed convex subset  $D$  of a Banach space  $E_1$  with values in  $E_2$ . Assume:

- (a)  $F$  is twice Fréchet-differentiable on  $U(x_0, R) \subseteq D$  for some  $x_0 \in D$ ,  $R \geq 0$  and satisfies (4)-(5);  
 (b) the inverse of the linear operator  $F'(x_0)$  exists;  
 (c) there exists a minimum nonnegative number  $R_1$  with

$$T(R_1) \leq R_1; \quad (15)$$

- (d) the following estimates are also satisfied:

$$R_1 \leq R, \quad (16)$$

$$\omega_1(R) < \frac{1}{\beta} \quad (17)$$

and

$$\frac{1}{2} \frac{\beta M}{1 - \beta \omega_1(R_1)} R_1 < 1. \quad (18)$$

Then

- (i) The sequence  $\{t_n\}$  ( $n \geq 0$ ) defined by (8)-(12) is monotonically increasing and bounded above by its limit  $R_1$  for all  $n \geq 0$ .  
 (ii) The Halley method generated by (2)-(4) is well defined, remains in  $U(x_0, R_1)$  for all  $n \geq 0$ , and converges to a unique zero  $x^*$  of equation  $F(x) = 0$  in  $U(x_0, R)$ .

Moreover, the following estimates hold:

$$\|x_n - x^*\| \leq R_1 - t_n \quad (19)$$

and

$$\|y_n - x^*\| \leq R_1 - s_n \quad \text{for all } n \geq 0. \quad (20)$$

**Proof.** (i) We will show that the sequence  $\{t_n\}$  ( $n \geq 0$ ) is monotonically increasing and bounded above by  $R_1$ . From (6)-(13), (17) and (18) we get  $t_0 \leq s_0 \leq t_1 \leq s_1 \leq t_2$ . By assuming  $t_k \leq s_k \leq t_{k+1}$ ,  $k = 0, 1, 2, \dots, n$ , we obtain  $t_{k+1} \leq s_{k+1} \leq t_{k+2}$  from (8), (12), (17) and (18). Hence  $\{t_n\}$  ( $n \geq 0$ ) is monotonically increasing. From (7) and (16)  $t_0 \leq t_1 \leq R_1$  and from (12) for  $n = 0$ ,  $t_2 \leq T(R_1) \leq R_1$ . Let us assume that  $t_k \leq R_1$ , for  $k = 0, 1, 2, \dots, n + 1$ . Then from (8) and (12)

$$\begin{aligned} t_{n+2} \leq & t_1 + B(R_1) \left[ \sum_{k=0}^n \int_{s_k}^{t_{k+1}} \omega_1(\tau) d\tau + \omega_1(R_1) \sum_{k=0}^n (t_{k+1} - s_k) \right. \\ & \left. + \frac{M}{2} c(R_1) B(R_1) R_1 \left( \sum_{k=0}^n \int_{t_k}^{s_k} \omega_1(t) dt \right) + \frac{1}{2} c(R_1) \left( \sum_{k=0}^n \int_{t_k}^{s_k} \omega_2(t) dt \right) \right] \\ & + \frac{M}{2} B(R_1) c(R_1) \left( \sum_{k=0}^n (s_{k+1} - t_{k+1})^2 \right) \leq \dots \leq T(R_1) \leq R_1 \quad (\text{by (15)}). \end{aligned} \quad (21)$$

Hence,  $\{t_n\}$  ( $n \geq 0$ ) is bounded above by  $R_1$ . By hypothesis (c),  $\lim_{n \rightarrow \infty} t_n = \lim_{n \rightarrow \infty} s_n = R_1$ . Moreover  $t_k \leq s_k \leq t_{k+1} \leq R_1$  for all  $k \geq 0$ . That completes the proof of part (i).

(ii) We will show that if

$$\|y_n - x_n\| \leq s_n - t_n, \quad n \geq 0 \quad (22)$$

$$\|F(x_n)\| \leq A(\omega_1, \omega_2, t_{n-1}, s_{n-1}), \quad n \geq 1 \quad (23)$$

$$\|F'(x_{n+1})^{-1}\| \leq B(t_{n+1}), \quad n \geq -1 \quad (24)$$

and

$$\|H(x_n, y_n)\| \leq MB(t_n)(s_n - t_n) < 1, \quad (25)$$

then

$$\|x_{n+1} - y_n\| \leq t_{n+1} - s_n, \quad (26)$$

$$\|F(x_{n+1})\| \leq A(\omega_1, \omega_2, t_n, s_n) \quad (27)$$

and

$$\|y_{n+1} - x_{n+1}\| \leq s_{n+1} - t_{n+1} \quad \text{for all } n \geq 0. \quad (28)$$

From (4), (5), (17) and (24) we obtain

$$\begin{aligned} \|x_{n+1} - y_n\| &\leq \frac{1}{2} \|F'(x_n)^{-1}\| \left\| \left( I - \frac{1}{2} H(x_n, y_n) \right)^{-1} \|F''(x_n)\| \|y_n - x_n\|^2 \right. \\ &\leq \frac{M}{2} B(t_n) c(t_n) (s_n - t_n)^2 = t_{n+1} - s_n. \end{aligned}$$

Hence, (26) is true.

Using (2)-(6), (13), (22)-(25) we can obtain the approximation

$$\begin{aligned} F(x_{n+1}) &= \int_0^1 [F'(y_n + t(x_{n+1} - y_n)) - F'(y_n)](x_{n+1} - y_n) dt + (F'(y_n) - F'(x_n))(x_{n+1} - y_n) \\ &\quad + \frac{1}{2} \left[ I - \frac{1}{2} H(x_n, y_n) \right]^{-1} F'(x_n)^{-1} F''(x_n)(y_n - x_n) \int_0^1 [F'(x_n + t(y_n - x_n)) \\ &\quad - F'(x_n)](y_n - x_n) dt \\ &\quad + \left[ I - \frac{1}{2} H(x_n, y_n) \right]^{-1} \int_0^1 [F''(x_n + t(y_n - x_n)) - F''(x_n)](1-t) dt (y_n - x_n)^2. \quad (29) \end{aligned}$$

We get by using the triangle inequality that the norm of each quantity at the right-hand side of (28) is such that

$$\begin{aligned} \|F(x_{n+1})\| &\leq \int_0^1 [\omega_1((1-t)s_n + t_{n+1}) - \omega_1(s_n)] dt (t_{n+1} - s_n) + (\omega_1(s_n) - \omega_1(t_n))(t_{n+1} - s_n) \\ &\quad + \frac{1}{2} c(t_n) B(t_n) M (s_n - t_n) \int_0^1 [\omega_1((1-t)t_n + ts_n) - \omega_1(t_n)] (s_n - t_n) dt \\ &\quad + \frac{1}{2} c(t_n) \int_0^1 [\omega_2((1-t)t_n + ts_n) - \omega_2(t_n)] dt (s_n - t_n)^2 \\ &\leq \int_{s_n}^{t_{n+1}} \omega_1(t) dt - \omega_1(s_n)(t_{n+1} - s_n) + (\omega_1(s_n) - \omega_1(t_n))(t_{n+1} - s_n) \\ &\quad + \frac{M}{2} c(t_n) B(t_n) (s_n - t_n) \left[ \int_{t_n}^{s_n} \omega_1(t) dt - \omega_1(t_n)(s_n - t_n) \right] \\ &\quad + \frac{1}{2} c(t_n) \left[ \int_{t_n}^{s_n} \omega_2(t) dt - \omega_2(t_n)(s_n - t_n) \right] = A(\omega_1, \omega_2, t_n, s_n). \end{aligned}$$

We have also used the estimates

$$\begin{aligned} \|x_{n+1} - x_0\| &\leq \|x_{n+1} - y_0\| + \|y_0 - x_0\| \leq \|x_{n+1} - y_n\| + \|y_n - y_0\| + \|y_0 - x_0\| \\ &\leq \cdots \leq (t_{n+1} - s_n) + (s_n - s_0) + s_0 = t_{n+1} \leq R_1, \end{aligned} \quad (30)$$

$$\begin{aligned} \|y_{n+1} - x_0\| &\leq \|y_{n+1} - y_0\| + \|y_0 - x_0\| \leq \|y_{n+1} - x_{n+1}\| + \|x_{n+1} - y_n\| + \|y_n - y_0\| + \|y_0 - x_0\| \\ &\leq \cdots \leq (s_{n+1} - t_{n+1}) + (t_{n+1} - s_n) + (s_n - s_0) \leq s_{n+1} \leq R_1 \end{aligned} \quad (31)$$

and the facts that if (5) and (6) are true, then

$$\|F'(x+h) - F'(x)\| \leq \omega_1(r + \|h\|) - \omega_1(r) \quad (32)$$

and

$$\|F''(x+h) - F''(x)\| \leq \omega_2(r + \|h\|) - \omega_2(r) \quad (33)$$

for all  $x \in U(x_0, r)$ ,  $\|h\| \leq R_1 - r$  (see, [22, p. 674]). Hence, (27) is true.

From (2), (24) and (27)

$$\|y_{n+1} - x_{n+1}\| \leq \|F'(x_{n+1})^{-1}\| \cdot \|F(x_{n+1})\| \leq B(t_n)A(\omega_1, \omega_2, t_n, s_n) = s_{n+1} - t_{n+1}.$$

Hence, (28) is also true.

Moreover, from (5), (7), (17), (32) and the estimate

$$\|F'(x_0)^{-1}\| \cdot \|F'(x_n) - F'(x_0)\| \leq \beta\omega_1(t_n) \leq \beta\omega_1(R_1) < 1,$$

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

$$\|F'(x_n)^{-1}\| \leq \frac{\|F'(x_0)^{-1}\|}{1 - \|F'(x_0)^{-1}\| \cdot \|F'(x_n) - F'(x_0)\|} \leq B(t_n). \quad (34)$$

for all  $n \geq 1$ .

Furthermore, from (3), (5), (34) and the estimate

$$\begin{aligned} \frac{1}{2}\|H(x_n, y_n)\| &\leq \frac{1}{2}\|F'(x_n)^{-1}\| \cdot \|F''(x_n)\| \|x_n - y_n\| \\ &\leq \frac{M}{2}B(t_n)(s_n - t_n) \leq \frac{M}{2}B(R_1)R_1 < 1, \quad (\text{by (18)}) \end{aligned}$$

it follows that  $I - \frac{1}{2}H(x_n, y_n)$  is invertible, and

$$\left\| \left( I - \frac{1}{2}H(x_n, y_n) \right)^{-1} \right\| \leq c(t_n) \quad \text{for all } n \geq 0. \quad (35)$$

Hence, the iterates generated by (2)-(4) are well defined for all  $n \geq 0$ . Also, by (26) and (28)

$$\|y_{n+1} - y_n\| \leq s_{n+1} - s_n \quad \text{and} \quad \|x_{n+1} - x_n\| \leq t_{n+1} - t_n \quad \text{for all } n \geq 0. \quad (36)$$

It now follows from (36) and (i) that the sequence  $\{x_n\}$  ( $n \geq 0$ ) 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 (30) and (31)  $x_n, y_n \in U(x_0, R_1)$  for all  $n \geq 0$ . The estimates (20) and (21) now follow from (36).

Finally to show uniqueness we assume that there exists another zero  $y^*$  of equation (1) in  $U(x_0, R)$ . Then from (19), (32) and (34) we obtain

$$\begin{aligned} \|F'(x_0)^{-1}\| \int_0^1 \|F'(y^* + t(x^* - y^*)) - F'(x_0)\| dt &\leq \beta\omega_1((1-t)\|x_0 - y^*\| + t\|x_0 - x^*\|) \\ &\leq \beta\omega_1((1-t)R + tR_1) \leq \beta\omega_1(R) < 1, \quad (\text{by (17)}). \end{aligned} \quad (37)$$

It now follows from (37) that the linear operator  $\int_0^1 F'(y^* + t(x^* - y^*)) dt$  is invertible. From this fact and the approximation

$$F(x^*) - F(y^*) = \int_0^1 F'(y^* + t(x^* - y^*)) dt (x^* - y^*) \quad (38)$$

it follows that  $x^* = y^*$ .

That completes the proof of Theorem 1.

**Remarks.** (a) From the estimates

$$\|x_n - y_0\| \leq \|x_n - y_n\| + \|y_n - y_0\| \leq (t_n - s_n) + (s_n - t_0) \leq t_n - \eta \leq R_1 - \eta \quad (39)$$

and

$$\begin{aligned} \|y_{n+1} - y_0\| &\leq \|y_{n+1} - x_{n+1}\| + \|x_{n+1} - y_n\| + \|y_n - y_0\| \\ &\leq (s_{n+1} - t_{n+1}) + (t_{n+1} - s_n) + (s_n - t_n) \leq s_{n+1} - \eta \leq R_1 - \eta, \end{aligned} \quad (40)$$

it follows that  $x_n, y_n \in U(y_0, R_1 - \eta)$  for all  $n \geq 0$ . Note also that  $R_1$  is the unique positive zero of  $T(r) - r = 0$  in  $(0, R_1]$ .

(b) If we take  $y_n = x_{n+1}$  in (2) and ignore the approximation (4), then our method coincides with Newton's and our results can be reduced to the ones in [4], [9], [16], [22] under the corresponding hypotheses.

(c) We can use the Halley method to approximate nonlinear equations with nondifferentiable operators. Indeed, consider the equation

$$F_1(x) = 0 \quad (41)$$

where

$$F_1(x) = F(x) + Q(x) \quad (42)$$

with  $F$  as before, and  $Q$  satisfying an estimate of the form

$$\|Q(x_1) - Q(x_2)\| \leq q_3(r)\|x_1 - x_2\| \quad (43)$$

for all  $x_1, x_2 \in U(x_0, R_1)$ , where  $q_3$  is a nondecreasing function on  $[0, R_1]$  with  $q_3(0) = 0$ .

Note that the differentiability of  $Q$  is not assumed here. Let us define the function

$$\omega_3(r) = \int_0^r q_3(t) dt. \quad (44)$$

Replace  $F$  in (2) by  $F_1$  and leave  $F'(x_n)^{-1}$  and approximations (3) and (4) as they are. Define the sequences  $\{\bar{t}_n\}$  and  $\{\bar{s}_n\}$  ( $n \geq 0$ ) as the corresponding  $\{t_n\}$  and  $\{s_n\}$  ( $n \geq 0$ ) given by (8) and (12). The only change will be an extra term of the form  $\omega_3(s_n) - \omega_3(t_n)$  inserted in the definition of  $A(\omega_1, \omega_2, t_n, s_n)$ . Define  $T_1$  as  $T$  in (14) but insert an extra term of the form  $\omega_3(r)$  (given by

(44)). Then following the proof of the above theorem with identical hypotheses and conclusions, we can show a similar theorem, but holding for equations (41) (see, also [4], [9], [22]).

(d) Similar theorems can be proved if the terms  $\|x_1 - x_2\|$  in (6) and (42) are replaced by a Hölder estimate of the form  $\|x_1 - x_2\|^p$  for some  $p \in [0, 1)$ . The details are omitted (see, also [4], [9]).

(e) In [9] we showed the result:

**Theorem 2.** Let  $F: D \subset E_1 \rightarrow E_2$ ,  $E_1, E_2$  are real Banach spaces, and  $D$  is a closed convex subset of  $E_1$ . Assume that  $F$  has second order continuous Fréchet derivatives on  $D$  and that the following conditions are satisfied:

$$\|F'(x) - F'(y)\| \leq \ell \|x - y\|, \|F''(x)\| \leq M, \|F''(x) - F''(y)\| \leq N \|x - y\|, \quad \text{for all } x, y \in D$$

$$\|F'(x_0)^{-1}\| \leq \beta, \quad \|y_0 - x_0\| \leq \eta,$$

$$\left[ 3M^2 + \frac{2N}{3\beta} \right]^{1/2} \leq K,$$

$$h = K\beta\eta \leq \frac{1}{2}$$

and

$$U(y_0, r_1 - \eta) \subset D.$$

Moreover, we define

$$g(t) = \frac{1}{2}Kt^2 - \frac{1}{\beta}t + \frac{\eta}{\beta},$$

$$r_1 = \frac{1 - \sqrt{1 - 2h}}{h}\eta,$$

and

$$\theta = \frac{1 - \sqrt{1 - 2h}}{1 + \sqrt{1 - 2h}},$$

where  $r_1$  is the smallest zero of the equation  $g(t) = 0$ . Then the Halley method (2)-(4) is convergent. Also  $x_n, y_n \in U(y_0, r_1 - \eta)$ , for all  $n \in N_0$ . The limit  $x^*$  is the unique zero of the equation  $F(x) = 0$  in  $U(x_0, r_2^*)$ ,  $r_1 \leq r_2^* < r_2$  if  $\ell = K$  (or  $M = K$ ) and  $r_2^* = r_2$  if  $\ell < K$  (or  $M < K$ ). Moreover, we have the following error estimates and the optimal error constant:

$$\|x_n - x^*\| \leq r_1 - t_n^1,$$

$$\|y_n - x^*\| \leq r_1 - s_n^1,$$

and

$$r_1 - t_n^1 = \frac{(1 - \theta^2)\eta}{1 - \theta^{3n}} \theta^{3n-1} \quad \text{for all } n \geq 0,$$

where

$$s_n^1 = t_n^1 - \frac{g(t_n^1)}{g'(t_n^1)}, \quad t_0^1 = 0,$$

$$h(t_n^1, s_n^1) = -g'(t_n^1)^{-1} g''(t_n^1)(s_n^1 - t_n^1)$$

and

$$t_{n+1}^1 = s_n^1 - \frac{1}{2}(s_n^1 - t_n^1) \frac{2g'(t_n^1)^{-1} g''(t_n^1)}{1 - \frac{1}{2}h(t_n^1, s_n^1)} \quad \text{for all } n \geq 0.$$

**Remarks.** (f) Several sufficient conditions can be given to show for example that under the hypotheses of Theorems 1 and 2

$$s_n - t_n \leq s_n^1 - t_n^1 \quad \text{for all } n \geq 0.$$

One such condition can be

$$\frac{\beta}{1 - \beta\omega_1(r)} \left[ \int_0^r \omega_1(t) dt + \omega_1(r)r + \frac{M}{2}c(r)B(r)r \int_0^r \omega_1(t) dt + \frac{1}{2}c(r) \int_0^r \omega_2(t) dt \right] \leq s_1^1 - t_1^1, \quad \text{or } \leq -\frac{g(r)}{g'(r)}$$

for all  $r \in [0, R]$ . The details are left for the motivated reader.

(g) By Theorems 1 and 2, we conclude that under Newton-Kantorovich assumptions the order of convergence of the Halley method in three, whereas for Newton's method is only two [4], [9], [13], [16].

(h) The functions  $q_1$  and  $q_2$  can be chosen by

$$q_1(r) = \sup_{x,y \in U(x_0,r)} \frac{\|F''(x_1) - F''(x_2)\|}{\|x_1 - x_2\|} \quad (\text{when } E_1 = E_2 \neq \mathbb{R} \text{ for example})$$

and

$$q_2(r) = \sup_{x,y \in U(x_0,r)} \frac{\|F''(x_1) - F''(x_2)\|}{\|x_1 - x_2\|}.$$

We will then have

$$q_1(r) \leq M \quad \text{and} \quad q_2(r) \leq N \quad \text{for all } r \in [0, R].$$

One can refer to [9] for some applications of these ideas to the solution of integral equations.

(i) Using the estimate  $\|F''(x)\| \leq \|F''(x) - F''(x_0)\| + \|F''(x_0)\| \leq \omega_2(R_1) + \|F''(x_0)\| = M^*$ , we see that hypothesis (5) can be replaced by the weaker one  $\|F''(x)\| \leq M^*$ .

(j) The Lipschitz condition (6) can be dropped, but the order of convergence will be slower (see [5], [9]).

### III. Applications

In this section we will give an example for Theorem 2 (similarly we can work on Theorem 1). We first note that by eliminating  $y_n$  ( $n \geq 0$ ) from approximations (2)-(4) we can obtain the method of tangent hyperbolas (or Chebyshev-Halley) which has been extensively studied in [1], [5], [6], [9], [12], [14], [15], [17], [18], [20], [21]. In all but our references it is assumed that  $N > 0$ , which means that their results cannot apply to solve quadratic operator equations of the form

$$P(x) = B(x, x) + L(x) + z, \quad (45)$$

where  $B, L$  are bounded quadratic and linear operators respectively with  $z$  fixed in  $E_1$ . We then have that  $P'(x) = 2B(x) + L$  and  $P''(x) = 2Q$ . Hence we get  $M = 2\|B\|$  and  $N = 0$ . Integral equations that can be formulated in the form  $P(x) = 0$  have very important applications in radiative transfer [2], [3], [9], [10].

As a specific example, let us consider the solution of quadratic integral equations of the form

$$x(s) = y(s) + \lambda x(s) \int_0^1 q(s,t)x(t)dt \quad (46)$$

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  $\lambda$  is a real number called the "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

(i)  $0 < q(s,t) < 1, 0 \leq s,t \leq 1$  with  $q(0,0) = 1$ ;

(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 [10], (Nobel prize of physics, 1983), and arise in the theories of radiative transfer, neutron transport and in the kinetic theory of gasses, [2], [3], [9], [10].

There exists an extensive literature on equations like (46) under various assumptions on the kernel  $q(s,t)$  and  $\lambda$  is a real or complex number. One can refer to the recent work in [2], [3], [9] and the references there. Here we demonstrate that the theorem via the iterative procedure (2)-(4) provides existence results for (46).

For simplicity (without loss of generality) we will assume that

$$q(s,t) = \frac{s}{s+t} \quad \text{for all } 0 \leq s,t \leq 1 \quad \text{with } q(0,0) = 1.$$

Note that  $q(s,t)$  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  $P$  on  $E_1$  by

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

Note that every zero of the equation  $P(x) = 0$  satisfies the equation (46).

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

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

$$K = M\sqrt{3} = .600283066$$

$$\beta = \|P'(1)^{-1}\| = 1.53039421.$$

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

$$h = .243628554 < .5,$$

$$r_1 = .3090766, \quad r_2 = 1.867984353$$

and

$$\theta = .165459951.$$

(For detailed computations, see also [2], [9] and [10].)

Therefore according to Theorem 2, equation (48) has a solution  $x^*$  and the two-point method (2)–(4) converges to  $x^*$ . Note that the results obtained in [1], [12], [14], [15], [17], [18], [20], [21] cannot apply here since  $N = 0$ . For Theorem 1 we can take  $q_1(r) = \ell$  and  $q_2(r) = 0$  for all  $r \in [0, R]$ . The computational details for this case are left to the motivated reader.

#### References

- [1] Altman, M. Iterative methods of higher order, Bull. Acad. Polon. Sci. Ser. Math. Astr. Phys. 9, (1961), 63–68.
- [2] Argyros, I.K. Quadratic equations and applications to Chandrasekhar's and related equations, Bull. Austral. Math. Soc. 32, (1985), 275–292.
- [3] Argyros, I.K. On a class of nonlinear integral equations arising in neutron transport, Aequationes Mathematicae, 36, (1988), 99–111.
- [4] Argyros, I.K. On the solution of equations with nondifferentiable operators and the Ptak error estimates, BIT, 90, (1990), 752–754.
- [5] Argyros, I.K. On the convergence of a Chebyshev-Halley-type method under Newton-Kantorovich hypotheses, Appl. Math. Letters, Vol. 6, No. 5, (1993), 71–74.
- [6] Argyros, I.K. and Chen, D. A note on the Halley method in Banach spaces, Appl. Math. and Comp. 58, (1993), 215–224.
- [7] Argyros, I.K. and Chen, D. A fourth order method in Banach spaces. Appl. Math. Letters, Vol. 6, No. 4, (1993), 97–98.
- [8] Argyros, I.K. and Szidarovszky, F. On the monotone convergence of general Newton-like methods, Bull. Austral. Math. Soc., Vol. 45, (1992), 489–502.
- [9] Argyros, I.K. and Szidarovszky, F. The Theory and Applications of Iteration Methods, C.R.C. Press, Inc., Boca Raton, Florida, 1993.
- [10] Chandrasekhar, S. Radiative Transfer, Dover Publ., New York, 1960.
- [11] Chen, X. and Yamamoto, T. Convergence domains of certain iterative methods for solving nonlinear equations, Numer. Funct. Anal. and Optimiz. 10, (1 and 2), (1989), 37–48.
- [12] Kanno, S. Convergence theorems for the method of tangent hyperbolas. Math. Japonica, 37, No. 4, (1992), 711–722.
- [13] Kantorovich, L.V. and Akilov, G.P. Functional Analysis in Normed Spaces, Pergamon Press, New York, 1964.
- [14] Mertvecova, M.A. An analog of the process of tangent hyperbolas for general functional equations (Russian), Dokl. Akad. Nauk. SSSR, 88, (1953), 611–614.

- [15] Necepurenko, M.T. On Chebyshev's method for functional equations (Russian), *Usephi Mat. Nauk.* 9, (1954), 163-170.
- [16] Potra, F.A. and Ptak, V. Sharp error bounds for Newton's process, *Numer. Math.* 34, (1980), 63-72.
- [17] Potra, F.A. On an iterative algorithm of order 1.839... for solving nonlinear operator equations, *Numer. Funct. Anal. and Optimiz.* 7, (1), (1984-85), 75-106.
- [18] Safev, R.A. The method of tangent hyperbolas, *Sov. Math. Dokl.* 4, (1963), 482-485.
- [19] Taylor, A.E. *Introduction to Functional Analysis*, Wiley Publ., New York, 1957.
- [20] Ul'm, S. Iteration methods with divided differences of the second order (Russian), *Dokl. Akad. Nauk. SSSR*, 158, (1964), 55-58. *Soviet Math. Dokl.* 5, 1187-1190.
- [21] Yamamoto, T. On the method of tangent hyperbolas in Banach spaces, *J. Comput. Appl. Math.* 21, (1988), 75-86.
- [22] Zabrejko, P.P. and Nguen, D.F. The majorant method in the theory of Newton-Kantorovich approximations and the Ptak error estimates, *Numer. Funct. Anal. and Optimiz.* 9, (5 and 6), (1987), 671-684.

## Shape-Preserving Bivariate Shift-Invariant Operators

by

George Anastassiou

Department of Mathematical Sciences

The University of Memphis

Memphis, Tennessee 38152

U.S.A.

and

X.M. Yu

Department of Mathematics

Southwest Missouri State University

Springfield, Missouri 65803

U.S.A.

### Abstract

Shape-preserving properties of some naturally arising bivariate shift-invariant operators  $B$  are examined. Namely, let  $f \in C^k(\mathbb{R}^2)$ ,  $k > 0$ ,  $r, s \geq 0$  all integers such that  $r + s = k$ . If

$$\frac{\partial^{r+s} f}{\partial x^r \partial y^s}(x, y) \geq 0,$$

then it is proved, under mild conditions on  $B$ , that

$$\frac{\partial^{r+s}}{\partial x^r \partial y^s} B(f)(x, y) \geq 0.$$

Associated simultaneous shape-preserving results are also presented for special type of operators  $B$ .

### I. Introduction

Let  $f \in C(\mathbb{R}^2)$ , and  $L(f)$  be an operator mapping  $C(\mathbb{R}^2)$  into  $C(\mathbb{R}^2)$ . Denote  $f_\alpha := f(\cdot + \alpha)$  for  $\alpha \in \mathbb{R}^2$ . If for each  $f \in C(\mathbb{R}^2)$  and  $\alpha \in \mathbb{Z}^2$ , we have  $L(f_\alpha) = (L(f))_\alpha$ , then we say that the operator  $L(f)$  is shift-invariant.

Let  $\varphi \in C(\mathbb{R}^2)$  be compactly supported on  $\mathbb{R}^2$  with  $\text{supp } \varphi \subseteq [-a, a]^2$ ,  $0 < a < +\infty$ . For  $f \in C(\mathbb{R}^2)$  we define

$$B(f)(x, y) := \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} f(i, j)\varphi(x - i, y - j), \quad (x, y) \in \mathbb{R}^2. \quad (1)$$

Since  $\varphi$  is compactly supported, there are only finite non-zero terms involved in the summations of (1). So  $B(f)$  is well-defined on  $\mathbb{R}^2$ , maps  $C(\mathbb{R}^2)$  into  $C(\mathbb{R}^2)$  and is shift-invariant.

We are concerned with the shape-preserving properties of such a shift-invariant operator  $B(f)$ , and want to know when to hold

$$\frac{\partial^{r+s} B(f)}{\partial x^r \partial y^s}(x, y) \geq 0 \quad \text{if} \quad \frac{\partial^{r+s} f}{\partial x^r \partial y^s}(x, y) \geq 0, \quad (x, y) \in \mathbb{R}^2.$$

We also consider the simultaneous shape-preserving problem. That is, for each  $r^* = 0, 1, \dots, r$ ,  $s^* = 0, 1, \dots, s$ , if

$$\varepsilon_{r^*, s^*} \frac{\partial^{r^*+s^*} f}{\partial x^{r^*} \partial y^{s^*}}(x, y) \geq 0,$$

where  $\varepsilon_{r^*, s^*} = \pm 1$ ,  $(x, y) \in \mathbb{R}^2$ , then we hold the same inequality for  $B(f)$ . The following Theorem 1 and Theorem 2 give some answers to these problems.

**Theorem 1.** *Let  $r, s$  be non-negative integers,  $k$  be a positive integer and  $r+s = k$ . Suppose that  $\varphi \in C(\mathbb{R}^2)$  is compactly supported on  $\mathbb{R}^2$  with  $\text{supp } \varphi \subseteq [-a, a]^2$ ,  $0 < a < +\infty$ ,  $\frac{\partial^{r+s} \varphi}{\partial x^r \partial y^s} \in C(\mathbb{R}^2)$  and satisfies the following conditions:*

(i) *for any fixed  $j$  and  $y$ ,  $\sum_{i=-\infty}^{\infty} p(i, j)\varphi(x - i, y - j)$  is a polynomial of degree  $< r$  with respect to  $x$  whenever  $p(x, y)$  is a polynomial of degree  $< r$  with respect to  $x$ .*

(ii) *for any fixed  $i$  and  $x$ ,  $\sum_{j=-\infty}^{\infty} p(i, j)\varphi(x - i, y - j)$  is a polynomial of degree  $< s$  with respect to  $y$  whenever  $p(x, y)$  is a polynomial of degree  $< s$  with respect to  $y$ .*

(iii) There are  $k$  real numbers  $-\infty =: x_0 < x_1 \leq \dots \leq x_r < x_{r+1} := +\infty$  and  $-\infty =: y_0 < y_1 \leq \dots \leq y_s < y_{s+1} := +\infty$  such that

$$(-1)^{m+l} \frac{\partial^{r+s} \varphi}{\partial x^r \partial y^s}(x, y) \geq 0, \quad \begin{array}{l} x_m \leq x \leq x_{m+1} \\ y_l \leq y \leq y_{l+1} \\ m = 0, 1, \dots, r, \\ l = 0, 1, \dots, s. \end{array} \quad (2)$$

Then, if  $f \in C(\mathbb{R}^2)$ ,  $\frac{\partial^{r+s} f}{\partial x^r \partial y^s}(x, y) \in C(\mathbb{R}^2)$  and  $\frac{\partial^{r+s} f}{\partial x^r \partial y^s}(x, y) \geq 0$ ,  $(x, y) \in \mathbb{R}^2$ , we have

$$\frac{\partial^{r+s} B(f)}{\partial x^r \partial y^s}(x, y) \geq 0, \quad (x, y) \in \mathbb{R}^2. \quad (3)$$

**Theorem 2.** Let  $r, s$  be non-negative integers,  $k$  be a positive integer and  $r + s = k$ . Suppose that  $\varphi(x) \in C(\mathbb{R})$  is compactly supported with  $\text{supp } \varphi \subseteq [-a, a]$ ,  $0 < a < +\infty$ ,  $\varphi^{(k)} \in C(\mathbb{R})$  and satisfies the following conditions:

(i) for each  $k^* = 0, 1, \dots, k-1$ ,  $\sum_{i=-\infty}^{\infty} p(i)\varphi(x-i)$  is a polynomial of degree  $k^*$  whenever  $p(x)$  is a polynomial of degree  $k^*$ .

(ii) there are  $k$  real numbers  $-\infty =: x_0 < x_1 \leq \dots \leq x_k < x_{k+1} := +\infty$  such that

$$(-1)^m \varphi^{(k)}(x) \geq 0, \quad x_m \leq x \leq x_{m+1}, \quad m = 0, 1, \dots, k.$$

Then, if  $f \in C(\mathbb{R}^2)$ ,  $\frac{\partial^{r+s} f}{\partial x^r \partial y^s} \in C(\mathbb{R}^2)$  and

$$\varepsilon_{r^*, s^*} \cdot \frac{\partial^{r^*+s^*} f}{\partial x^{r^*} \partial y^{s^*}}(x, y) \geq 0, \quad (x, y) \in \mathbb{R}^2, \quad (4)$$

where  $r^* = 0, 1, \dots, r$ ,  $s^* = 0, 1, \dots, s$  and  $\varepsilon_{r^*, s^*} = \pm 1$ , for the  $B(f)$  defined by

(1) with  $\varphi(x, y) := \varphi(x)\varphi(y)$ , we have

$$\varepsilon_{r^*, s^*} \cdot \frac{\partial^{r^*+s^*} B(f)}{\partial x^{r^*} \partial y^{s^*}}(x, y) \geq 0, \quad (x, y) \in \mathbb{R}^2, \quad \begin{array}{l} r^* = 0, 1, \dots, r, \\ s^* = 0, 1, \dots, s. \end{array} \quad (5)$$

## II. Proofs

**Proof of Theorem 1.** Let  $(\bar{x}, \bar{y})$  be a fixed point on  $\mathbb{R}^2$ . We want to prove (3) for  $(\bar{x}, \bar{y})$ . First of all, we check how  $\frac{\partial^{r+s} \varphi}{\partial x^r \partial y^s}(\bar{x} - i, \bar{y} - j)$  changes signs as  $i, j$  change. Let  $i_m$  and  $j_l$  ( $m = 1, 2, \dots, r; l = 1, 2, \dots, s$ ) be the integers such that

$$\bar{x} - i_m \leq x_{r-m+1} < \bar{x} - i_m + 1, \quad m = 1, 2, \dots, r, \quad (6)$$

$$\bar{y} - j_l \leq y_{s-l+1} < \bar{y} - j_l + 1, \quad l = 1, 2, \dots, s, \quad (7)$$

where  $x_m$  and  $y_l$  ( $m = 1, \dots, r; l = 1, \dots, s$ ) are the numbers given by the condition (iii). Since  $x_m \leq x_{m+1}$  ( $m = 1, \dots, r-1$ ) and  $y_l \leq y_{l+1}$  ( $l = 1, \dots, s-1$ ), we have  $i_m \leq i_{m+1}$  ( $m = 1, \dots, r-1$ ) and  $j_l \leq j_{l+1}$  ( $l = 1, \dots, s-1$ ). If for some  $m$  we have  $i_m = i_{m+1}$ , we redefine  $i_{m+1} := i_m + 1$ . Hence, after the refinement, we have  $i_m < i_{m+1}$ . Indeed, the refinement is going on in the order of  $m$ . If  $m_0$  is the smallest positive integer such that  $i_{m_0} = i_{m_0+1}$ , we redefine  $i_{m_0+1} := i_{m_0} + 1$ . If  $i_{m_0+2} \leq$  the refinement of  $i_{m_0+1}$ , we redefine  $i_{m_0+2} := i_{m_0} + 2$ , and so on until we have some positive integer  $q$  such that  $i_{m_0+q} \geq i_{m_0} + q$ . Then we check for the next  $m_1$  such that  $i_{m_1} = i_{m_1+1}$ . Do the same refinement. In this way, we can modify  $i_m$  such that  $i_m < i_{m+1}$ ,  $m = 1, \dots, r-1$ . Similarly, we can redefine  $j_l$  such that  $j_l < j_{l+1}$ ,  $l = 1, \dots, s-1$ .

Set  $i_0 = j_0 := -\infty$ ,  $i_{r+1} = j_{s+1} := +\infty$ . Suppose that for the fixed  $m$  there is some integer  $i$  such that  $i_m < i < i_{m+1}$ . Then this  $i_{m+1}$  is not the refinement of the original  $i_{m+1}$ . Hence, in this case, the  $i_{m+1}$  satisfies (6). Meanwhile even if  $i_m$  is a refinement,  $i_m$  still satisfies the inequality on the left-hand side of (6), because the refinement is greater than the original one. Thus, for  $i_m < i < i_{m+1}$ , by (6), we have

$$\bar{x} - i < \bar{x} - i_m \leq x_{r-m+1},$$

and noticing  $i \leq i_{m+1} - 1$ , we have

$$\bar{x} - i \geq \bar{x} - i_{m+1} + 1 > x_{r-m}.$$

Hence

$$x_{r-m} < \bar{x} - i < x_{r-m+1}, \quad i_m < i < i_{m+1}, \quad m = 0, 1, \dots, r. \quad (8)$$

Similarly, we have

$$y_{s-l} < \bar{y} - j < y_{s-l+1}, \quad j_l < j < j_{l+1}, \quad l = 0, 1, \dots, s. \quad (9)$$

It follows from (8), (9) and (2) that

$$(-1)^{r+s-m-l} \frac{\partial^{r+s} \varphi}{\partial x^r \partial y^s}(\bar{x} - i, \bar{y} - j) \geq 0, \quad \begin{array}{l} i_m < i < i_{m+1}, \\ j_l < j < j_{l+1}, \\ m = 0, 1, \dots, r; \\ l = 0, 1, \dots, s. \end{array} \quad (10)$$

Now we are going to find a function  $F(x, y)$  such that

$$\frac{\partial^{r+s} B(f)}{\partial x^r \partial y^s}(\bar{x}, \bar{y}) = \frac{\partial^{r+s} B(F)}{\partial x^r \partial y^s}(\bar{x}, \bar{y}) \quad (11)$$

and

$$(-1)^{r+s+m+l} F(x, y) \geq 0, \quad \begin{array}{l} i_m < x < i_{m+1}, \\ j_l < y < j_{l+1}, \\ m = 0, 1, \dots, r; \\ l = 0, 1, \dots, s. \end{array} \quad (12)$$

For  $f(x, y)$  and each fixed  $y$ , we can construct a polynomial  $P_1(x, y)$  of degree  $r-1$  with respect to  $x$  such that  $P_1(i_m, y) = f(i_m, y)$ ,  $m = 1, 2, \dots, r$ . Then, for the function  $f(x, y) - P_1(x, y)$  and each fixed  $x$ , we can construct a polynomial  $P_2(x, y)$  of degree  $s-1$  with respect to  $y$  such that  $P_2(x, j_l) = f(x, j_l) - P_1(x, j_l)$ ,  $l = 1, 2, \dots, s$ . Since  $f(i_m, y) - P_1(i_m, y) \equiv 0$ , we have  $P_2(i_m, y) \equiv 0$ . Thus, we define

$$F(x, y) := f(x, y) - P_1(x, y) - P_2(x, y) \quad \text{on } \mathbb{R}^2,$$

we have

$$F(x, y) = 0, \text{ if } x = i_m \text{ (} m = 1, 2, \dots, r \text{) or } y = j_l \text{ (} l = 1, 2, \dots, s \text{)}. \quad (13)$$

We also have

$$\frac{\partial^{r+s} F}{\partial x^r \partial y^s} = \frac{\partial^{r+s} f}{\partial x^r \partial y^s} \geq 0 \text{ on } \mathbb{R}^2, \quad (14)$$

and

$$\frac{\partial^s F}{\partial y^s}(x, y) = 0, \text{ for } x = i_m \text{ (} m = 1, 2, \dots, r \text{)} \quad (15)$$

because of (13).

For fixed  $y$ , the function  $g(x) := \frac{\partial^s F}{\partial y^s}(x, y)$  has

$$g^{(r)}(x) \geq 0 \text{ on } \mathbb{R},$$

and

$$g(i_m) = 0 \text{ (} m = 1, 2, \dots, r \text{),}$$

by (14) and (15). Hence

$$\begin{aligned} (-1)^{r+m} g(x) &= (-1)^{r+m} \prod_{m=1}^r (x - i_m) \cdot [i_1, \dots, i_r, x]g \\ &= (-1)^{r+m} \prod_{m=1}^r (x - i_m) g^{(r)}(\xi) / r! \geq 0, \\ & \quad i_m < x < i_{m+1}, \quad m = 0, 1, \dots, r, \end{aligned} \quad (16)$$

where  $[i_1, \dots, i_r, x]g$  is the  $r$ -th divided difference at the points  $i_1, \dots, i_r$  and  $x$ ,  $\xi$  is the point between  $x$  and some  $i_m$  by the Mean Value Theorem.

Now let  $m$  be fixed and  $x$  be fixed with  $i_m < x < i_{m+1}$ . The function  $q(y) := F(x, y)$  has

$$(-1)^{r+m} q^{(s)}(y) \geq 0 \text{ on } \mathbb{R}$$

and

$$q(j_l) = 0 \quad (l = 1, 2, \dots, s),$$

by (16) and (13). Then, similar to (16), we have

$$(-1)^{r+m+s+l} q(y) \geq 0, \quad j_l < y < j_{l+1}, \quad l = 0, 1, \dots, s, \quad (17)$$

that gives (12). For the fixed  $(\bar{x}, \bar{y})$ , from the conditions (i) and (ii), we obtain

(11). Hence, from (11), (13), (10) and (12), we have

$$\begin{aligned} \frac{\partial^{r+s} B(f)}{\partial x^r \partial y^s}(\bar{x}, \bar{y}) &= \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} F(i, j) \frac{\partial^{r+s} \varphi}{\partial x^r \partial y^s}(\bar{x} - i, \bar{y} - j) \\ &= \sum_{l=0}^s \sum_{m=0}^r \sum_{j_l < j < j_{l+1}} \sum_{i_m < i < i_{m+1}} F(i, j) \frac{\partial^{r+s} \varphi}{\partial x^r \partial y^s}(\bar{x} - i, \bar{y} - j) \\ &\geq 0. \quad \square \end{aligned}$$

To prove Theorem 2, we only need to use Theorem 1 and the following fact:

Suppose that  $\varphi(x) \in C(\mathbb{R})$  is compactly supported on  $\mathbb{R}$  with  $\text{supp } \varphi \subseteq [-a, a]$ ,  $0 < a < +\infty$ . If for some positive integer  $k$ ,  $\varphi^{(k)} \in C(\mathbb{R})$  and there are  $k$  real numbers  $-\infty =: x_0 < x_1 \leq x_2 \leq \dots \leq x_k < x_{k+1} := +\infty$  such that

$$(-1)^i \varphi^{(k)}(x) \geq 0, \quad x \in (x_i, x_{i+1}), \quad i = 0, 1, \dots, k,$$

then for any positive integer  $s$ ,  $s < k$ , there are  $s$  real numbers  $-\infty =: x'_0 < x'_1 \leq x'_2 \leq \dots \leq x'_s < x'_{s+1} := +\infty$  such that

$$(-1)^i \varphi^{(s)}(x) \geq 0, \quad x \in (x'_i, x'_{i+1}), \quad i = 0, 1, \dots, s.$$

**Example.** Let  $\varphi(x)$  be the  $B$ -spline of order  $k+2$ ,  $B_{k+2}(x)$ , where  $B_n(x)$  is defined inductively as follows:

$$B_1(x) = \frac{1}{2} \left( \chi_{[-\frac{1}{2}, \frac{1}{2}]}(x) + \chi_{(-\frac{1}{2}, \frac{1}{2})}(x) \right),$$

$$B_n(x) = B_{n-1} * B_1(x), \quad n = 2, 3, \dots$$

Take  $\varphi(x, y) := \varphi(x) \cdot \varphi(y)$ . Such a  $\varphi(x, y)$  satisfies all the assumptions of Theorems 1 and 2.

**Remark.** We mentioned in [1] that the similar results hold for the univariate case and proved the  $r$ -th convexity preserving for  $r \leq 3$ . The results of this paper can be also generalized to the multivariate case.

### References

- [1] G. Anastassiou and X.M. Yu, "Convex and Coconvex Probabilistic Wavelet Approximation," *Stochastic Analysis & Appl.*, Vol. 10, No. 5, pp. 507-521 (1992).
- [2] G. Anastassiou and X.M. Yu, "Bivariate Probabilistic Wavelet Approximation," in Proceedings of 6th S.E.A. Intern. Conference on Approx. Theory held in Memphis, March 1991, pp. 79-92, Marcel Dekker, Inc., 1992, New York.

## Minimax Chebyshev coefficients of the real dilogarithm

Roberto Barrio

GME, Depto. de Matemática Aplicada, Centro Politécnico Superior,  
Universidad de Zaragoza, E-50015 Zaragoza. Spain

### Abstract

The Chebyshev coefficients of a near best uniform polynomial approximation of the dilogarithm function with real argument are presented up to an error of the order  $10^{-18}$ .

**Keywords:** Dilogarithm, Chebyshev polynomials, minimax approximation.

### 1. Introduction

The dilogarithm or polylogarithm of second order is a function of complex argument that frequently appears in several fields of mathematics, physics and engineering, such as quantum electrodynamics, thermodynamics, number theory, celestial mechanics or network analysis. This function is defined as the line integral

$$\text{Li}_2(z) = \int_0^z \frac{\ln(1-t)}{t} dt,$$

where  $z \in \mathbb{C}$ ; the integration path is a straight line joining the origin and the point  $z$ .

Several algorithms have been written to evaluate the dilogarithm of real argument ( $x \in \mathbb{R}$ ). For instance, Kölbig, Mignaco and Remidi [5] give the coefficients of its development in Chebyshev polynomials of the first kind,  $T_n(x)$ , on the interval  $[-1, 1/2]$ . They need 27 coefficients to obtain an error of the order  $10^{-14}$ . Other authors, as Ginsberg and Zaborowski [4], use the series formulation of the dilogarithm

$$\text{Li}_2(x) = \sum_{k=1}^{\infty} \frac{x^k}{k^2}, \quad \text{for } x \in (-1, 1).$$

From our part, we propose to employ a different approach that allows to obtain better precision: we calculate the polynomial that is the best uniform approximation (minimax) (see [8]) of the function on a discrete set of points. This type of approximations it has been used in the construction of evaluation algorithms of other special functions [7] and permits to reduce the number of coefficients. Besides, this approximation gives an oscillatory error behavior.

### 2. Minimax coefficients

Firstly, we enunciate some properties [1, 4, 6] of the dilogarithm function that permit us to reduce the size of the interval of approximation.

**Proposition 1** The dilogarithm of a real argument,  $\text{Li}_2(x)$  with  $x \in \mathbb{R}$ , verifies

(i)  $\text{Li}_2(x) \in \mathbb{R}$  for  $x \leq 1$ .

(ii)  $\text{Im}(\text{Li}_2(x)) = -\sqrt{-1} \pi \ln(x)$  for  $x > 1$ .

(iii) If  $x > 1$

$$\text{Li}_2\left(\frac{1}{x}\right) + \text{Li}_2(x) = \frac{\pi^2}{3} - \frac{1}{2} \ln^2(x) - \sqrt{-1} \pi \ln(x).$$

(iv) If  $x < -1$

$$\text{Li}_2\left(\frac{1}{x}\right) + \text{Li}_2(x) = -\frac{\pi^2}{6} - \frac{1}{2} \ln^2(-x).$$

(v) If  $-1 \leq x < 0$

$$\text{Li}_2(x) = -\frac{1}{2} \ln^2(1-x) - \text{Li}_2\left(\frac{x}{x-1}\right).$$

(vi) If  $1/2 < x < 1$

$$\text{Li}_2(x) + \text{Li}_2(1-x) = \frac{\pi^2}{6} - \ln(x) \ln(1-x).$$

(vii)  $\text{Li}_2(1) = \frac{\pi^2}{6}$ .

The properties (i) and (ii) give us the value of the complex component of the dilogarithm of a real argument. The properties (iii-vii) reduce the interval to  $[0, 1/2]$ .

Taking into account these properties, we have obtained, in the interval  $[0, 1/2]$ , the coefficients of a near best polynomial approximation in the  $l_\infty$  norm expressed as a finite series of Chebyshev polynomials of the first kind. The values of the function at a point  $x \in \mathbb{R} \notin [0, 1/2]$  may be calculated by using the different mappings of the properties (iii-vii). Once the polynomial is given, it can be efficiently evaluated at any point with the Clenshaw's algorithm [3]. Besides, we may calculate the derivatives of any order by using the Smith's algorithm [9].

In order to obtain the coefficients of the approximation polynomial we may use several methods. In general, they are given [5] by calculating the Fourier-Chebyshev coefficients of the development of the function. In our case, we employ another classical approach, we calculate the coefficients of the polynomial that gives the best uniform approximation (minimax) of the function on a discrete set of points. The advantages of such election are the oscillatory behavior of the error and the lower number of coefficients needed in relation with the above mentioned possibility.

The chosen discrete set of points on which we evaluate the dilogarithm function is

$$x_i = \frac{1}{4} \left( \cos \frac{i \pi}{1000} + 1 \right), \quad 0 \leq i \leq 1000.$$

that is to say, the zeroes of the Chebyshev polynomial  $T_{1000}(x)$  mapped to the interval of study  $[0, 1/2]$ . Once we have the values of the dilogarithm function at these points, we have to solve an overdetermined system of linear equations in the norm  $l_\infty$ ; in our case we solve it by means of the algorithm of Barrodale and Phillips [2].

In Table 1 we present the coefficients necessary to obtain an error of  $10^{-18}$ . Besides, we show the coefficients of the same polynomial expressed as a power series of the variable  $x$ . In this case, in contrast to the Chebyshev polynomials, the coefficients result to be of

Table 1.—Coefficients of the approximation.

$\sum_{i=0}^{20} c_i T_i(4x-1)$	$\sum_{i=0}^{20} a_i x^i$
$c_0 = 2.79304142558718841 \text{ E} - 01$	$a_0 = 2. \text{ E} - 18$
$c_1 = 2.90242984773879951 \text{ E} - 01$	$a_1 = 9.99999999999997728 \text{ E} - 01$
$c_2 = 1.1732757105440220 \text{ E} - 02$	$a_2 = 2.50000000000473760 \text{ E} - 01$
$c_3 = 8.68216018804862 \text{ E} - 04$	$a_3 = 1.11111111066912512 \text{ E} - 01$
$c_4 = 8.2291327634940 \text{ E} - 05$	$a_4 = 6.2500002306744320 \text{ E} - 02$
$c_5 = 8.927973216141 \text{ E} - 06$	$a_5 = 3.9999924011941888 \text{ E} - 02$
$c_6 = 1.054646394612 \text{ E} - 06$	$a_6 = 2.7779479238868992 \text{ E} - 02$
$c_7 = 1.32088830190 \text{ E} - 07$	$a_7 = 2.0380957773135872 \text{ E} - 02$
$c_8 = 1.7264694403 \text{ E} - 08$	$a_8 = 1.5946529068023808 \text{ E} - 02$
$c_9 = 2.331132198 \text{ E} - 09$	$a_9 = 9.467284472987648 \text{ E} - 03$
$c_{10} = 3.22910388 \text{ E} - 10$	$a_{10} = 2.9856026534608896 \text{ E} - 02$
$c_{11} = 4.5662984 \text{ E} - 11$	$a_{11} = 9.8478497676132352 \text{ E} - 02$
$c_{12} = 6.568064 \text{ E} - 12$	$a_{12} = 4.56892023878713344 \text{ E} - 01$
$c_{13} = 9.58311 \text{ E} - 13$	$a_{13} = -1.483109794230304768 \text{ E} + 00$
$c_{14} = 1.41529 \text{ E} - 13$	$a_{14} = 3.858377616889217024 \text{ E} + 00$
$c_{15} = 2.1121 \text{ E} - 14$	$a_{15} = -7.713153233685839873 \text{ E} + 00$
$c_{16} = 3.181 \text{ E} - 15$	$a_{16} = 1.174440266576376627 \text{ E} + 01$
$c_{17} = 4.83 \text{ E} - 16$	$a_{17} = -1.313474831322605158 \text{ E} + 01$
$c_{18} = 7.4 \text{ E} - 17$	$a_{18} = 1.023217835338576691 \text{ E} + 01$
$c_{19} = 1.1 \text{ E} - 17$	$a_{19} = -4.971973988617027584 \text{ E} + 00$
$c_{20} = 2. \text{ E} - 18$	$a_{20} = 1.152921504606846976 \text{ E} + 00$

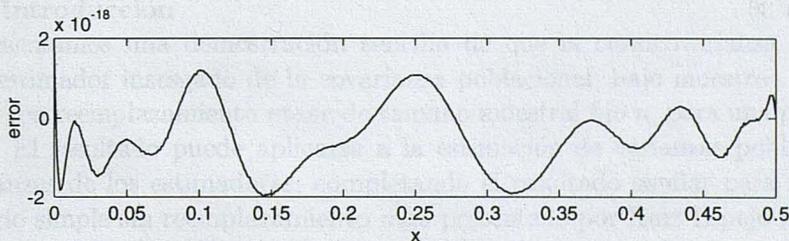


Figure 1.—Error in the approximation of the dilogarithm on the real interval  $[0, 1/2]$ .

a similar order, which makes impossible to know the quality of the approximation and giving a no-stable representation of the polynomial.

In Figure 1 we plot the error obtained when 21 coefficients are taken. We may remark the oscillatory behavior of such error, typical of the uniform approximations, and the estimation of the error given by the last coefficient.

**Acknowledgments:** The author is partially supported by the Centre National d'Etudes Spatiales (Toulouse, France) and by the Ministerio de Educación y Ciencia, project PB95-0807, (Spain)

### References

- [1] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover publications, Inc., New York, 1985.

- [2] I. Barrodale and C. Phillips, Solutions of an overdetermined system of linear equations in the Chebyshev norm, *CACM* algorithm 495, *ACM Trans. Math. Software* **1** (1975), pp. 264-278.
- [3] C.W. Clenshaw, A note on the summation of Chebyshev series, *Math. Tab. Wash.* **9** (1955), pp. 118-120.
- [4] E.S. Ginsberg and D. Zaborowski, The Dilogarithm Function of a Real Argument *Collected Algorithms from ACM.*, algorithm 490 (1973).
- [5] K.S. Kölbig, J.A. Mignaco and E. Remiddi, On Nielsen's generalized polylogarithms and their numerical calculation *BIT* **10** (1970), pp. 38-74.
- [6] L. Lewin, *Dilogarithms and Associated Functions*, North-Holland, New York, 1981.
- [7] C. W. Clenshaw, *Chebyshev series for mathematical functions*, National Physical Laboratory Mathematical Tables, 5, London H.M.S.O, 1962.
- [8] M.J.D. Powell, *Approximation theory and methods*, Cambridge University Press, 1981.
- [9] F.J. Smith, An algorithm for summing orthogonal polynomial series and their derivatives with applications to curve-fitting and interpolation, *Math. Comp.* **19** (1965), pp. 33-36.

SOBRE LA CUASICOVARIANZA MUESTRAL  
EN EL MUESTREO ALEATORIO SIMPLE

M. Ruiz Espejo

Deptº de Economía Aplicada Cuantitativa.

Fac. de Ciencias Económicas y Empresariales.

Univ. a Distancia. c/ Senda del Rey. s/n. 28040-Madrid.

**Summary.** Under simple random sampling with replacement of fixed sample size  $n$ , we give a simple proof of which we deduce that the sample quasivariance is an unbiased estimator of the population covariance. As a corollary, the sample quasivariance is unbiased for population variance (Hansen *et al.*, 1953, p. 98).

**1. Introducción**

Presentamos una demostración sencilla de que la cuasicovarianza muestral es un estimador insesgado de la covarianza poblacional, bajo muestreo aleatorio simple con reemplazamiento *masr* de tamaño muestral fijo  $n$ , para una población finita. El resultado puede aplicarse a la estimación de varianzas poblacionales y varianzas de los estimadores, completando el resultado similar para muestreo aleatorio simple sin reemplazamiento *mas* presentado por Ruiz Espejo (1995).

Sea  $U = \{1, 2, \dots, N\}$  la población finita, y la muestra aleatoria simple con reemplazamiento es de tamaño  $n$ ,  $1 < n$ . Denotamos por  $\bar{z}$  (y  $\bar{Z}$ ) a la media muestral (y poblacional) para la variable  $z$ . Si disponemos de dos variables  $x$  e  $y$  definidas para todas las unidades de la población finita, la cuasicovarianza muestral

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})$$

(donde  $i$  indica el lugar en la muestra ordenada) es un estimador insesgado, bajo diseño *masr*, para la covarianza poblacional

$$\sigma_{xy} = \frac{1}{N} \sum_{i \in U} (y_i - \bar{Y})(x_i - \bar{X}).$$

Para probarlo presentamos dos lemas previos.

## 2. Resultados

Lema 1. Bajo diseño *masr* de tamaño muestral  $n$ , si  $i$  es la  $i$ -ésima unidad extraída en la muestra ordenada, entonces:

$$E(\bar{y}_{x_i}) = E(y_i|\bar{x}) = \bar{X}\bar{Y} + \frac{\sigma_{xy}}{n} \square$$

*Demostración.*

$$\begin{aligned} E(\bar{y}_{x_i}) &= EE(\bar{y}_{x_i}|i) = EE\left(\frac{y_i x_i}{n} + \frac{x_i}{n} \sum_{j \neq i}^n y_j | i\right) = \\ &= E\left[\frac{x_i y_i}{n} + \frac{x_i}{n} E\left(\sum_{j \neq i}^n y_j | i\right)\right] = E\left[\frac{x_i y_i}{n} + \frac{x_i}{n} (n-1)\bar{Y}\right] = \\ &= \frac{\alpha_{11}}{n} + \frac{n-1}{n} \bar{X}\bar{Y} = \bar{X}\bar{Y} + \frac{1}{n} (\alpha_{11} - \bar{X}\bar{Y}) = \bar{X}\bar{Y} + \frac{\sigma_{xy}}{n}. \end{aligned}$$

El resultado simétrico para  $E(y_i|\bar{x})$  es análogo  $\square$

Lema 2. Bajo diseño *masr* de tamaño muestral  $n$ , entonces:

$$E(\bar{x}\bar{y}) = \bar{X}\bar{Y} + \frac{\sigma_{xy}}{n} \square$$

*Demostración.* Partimos de la relación

$$E(\bar{x}\bar{y}) = \bar{X}\bar{Y} + Cov(\bar{x}, \bar{y}),$$

y como

$$\begin{aligned} Cov(\bar{x}, \bar{y}) &= Cov\left(\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{j=1}^n y_j\right) = \frac{1}{n^2} \left[ \sum_{i=1}^n Cov(x_i, y_i) \right] = \\ &= \frac{1}{n^2} n \sigma_{xy} = \frac{\sigma_{xy}}{n}, \end{aligned}$$

resulta que

$$E(\bar{x}\bar{y}) = \bar{X}\bar{Y} + \frac{\sigma_{xy}}{n} \square$$

*Teorema.* Bajo diseño *masr* de tamaño muestral  $n$ :  $E(s_{xy}) = \sigma_{xy} \square$

*Demostración.*

$$\begin{aligned}
 E(s_{xy}) &= \frac{1}{n-1} \sum_{i=1}^n E[(y_i - \bar{y})(x_i - \bar{x})] = \\
 &= \frac{n}{n-1} E(y_i x_i - y_i \bar{x} - \bar{y} x_i + \bar{y} \bar{x}) = \\
 &= \frac{n}{n-1} [E(y_i x_i) - E(y_i \bar{x}) - E(\bar{y} x_i) + E(\bar{x} \bar{y})],
 \end{aligned}$$

de donde, como  $E(y_i x_i) = \alpha_{11}$  (momento no central de orden 1 y 1, en la población) y por los lemas 1 y 2,

$$\begin{aligned}
 E(s_{xy}) &= \frac{n}{n-1} \left[ \alpha_{11} - \left( \frac{\sigma_{xy}}{n} + \bar{X}\bar{Y} \right) \right] = \\
 &= \frac{n}{n-1} \left[ \left( \alpha_{11} - \bar{X}\bar{Y} \right) - \frac{\sigma_{xy}}{n} \right] = \\
 &= \frac{n}{n-1} \left( \sigma_{xy} - \frac{\sigma_{xy}}{n} \right) = \sigma_{xy} \quad \square
 \end{aligned}$$

Denotando  $s_x^2 = s_{xx}$  y  $\sigma_x^2 = \sigma_{xx}$ :

**Corolario.** Bajo diseño *masr*, de tamaño  $n$ :  $E(s_x^2) = \sigma_x^2 \quad \square$

**Demostración.** La cuasivarianza muestral es un estimador insesgado de la varianza poblacional como consecuencia del teorema anterior, cuando  $x = y \quad \square$

Este último corolario está enunciado en Hansen *et al.* (1953, p. 98) sin demostración.

## Referencias

- [1] Hansen, M.H., Hurwitz, W.N. y Madow, W.G. (1953). *Sample Survey Methods and Theory* (Volume II). John Wiley. Nueva York.
- [2] Ruiz Espejo, M. (1995). Una relación entre cuasivarianzas muestral y poblacional. *Rev. Acad. Cienc. Zaragoza* 50, 51-53.

CONTROL DE CUENTAS ECONÓMICAS POR MUESTREO  
CON NÚMERO CONOCIDO DE UNIDADES GRANDES

por

M. Ruiz Espejo

*Departamento de Economía Aplicada Cuantitativa*

*Facultad de Ciencias Económicas y Empresariales*

*Universidad a Distancia*

*C/ Senda del Rey, s/n*

*28040 Madrid*

**Summary.** In this note we provide two sampling methods to estimate without bias the true population mean, and the respective variance estimators are given. Our results are useful for the control of skewed economic populations, with a number known of large units.

### 1. Introducción

En este trabajo abordamos el control de cuentas económicas por muestreo cuando es conocido el número de unidades grandes o que más contribuyen al total económico, en poblaciones oblicuas (Glasser, 1962).

Proponemos dos estimadores insesgados del total económico verdadero, estimando insesgadamente la varianza de tales estimadores (Wolter, 1985).

### 2. Diseños muestrales y estimadores

Partimos de una población finita de tamaño  $N$ , que posestratificamos en  $L$  estratos de tamaños  $N_h$  ( $h = 1, 2, \dots, L$ ) de modo que el estrato  $L$ -ésimo o de unidades grandes tiene un tamaño conocido  $N_L$ . Si  $L = 2$ , Glasser (1962) estudió el estimador usual en muestreo estratificado. Si  $L > 2$ , proponemos dos métodos de estimación:

(a) El diseño usado es de "muestreo aleatorio simple con reemplazamiento", masr, de tamaño total  $m - N_L$  sobre los  $L - 1$  primeros estratos, donde se obtiene para  $h = 1, 2, \dots, L - 1$ , el número de unidades seleccionadas  $m_h$  en el estrato  $h$ -ésimo, de modo que

$$w_h = \frac{m_h}{m - N_L} \quad (h = 1, 2, \dots, L-1), \quad W_h = \frac{N_h}{N} \quad (h = 1, 2, \dots, L),$$

$$W_h^* = \frac{N_h}{N - N_L} \quad (h = 1, 2, \dots, L-1) \quad \text{y} \quad m = \sum_{h=1}^{L-1} m_h + N_L.$$

En el estrato  $h$ -ésimo se submuestra seleccionando  $n_h$  unidades donde se observa la variable  $x'$  verdadera por diseño masr, y con un censo en el estrato de unidades grandes, es decir,  $n_L = N_L$ .

El estimador propuesto es (para la media poblacional):

$$\bar{x}' = \sum_{h=1}^{L-1} w_h \bar{x}'_h + W_L \bar{X}'_L,$$

donde  $\bar{x}'_h$  es la media muestral de las  $n_h$  unidades seleccionadas en el estrato  $h$ -ésimo ( $h = 1, 2, \dots, L-1$ ) para la variable  $x'$ , y  $\bar{X}'_L$  es la media del estrato  $L$  para las  $N_L$  unidades donde observamos la variable verdadera  $x'$ .

Este estimador es insesgado para la media poblacional verdadera  $\bar{X}'$ , pues llamando  $\mathbf{m} = (m_1, m_2, \dots, m_{L-1})$ ,

$$\begin{aligned} E(\bar{x}') &= E[E(\bar{x}'|\mathbf{m})] = \sum_{h=1}^{L-1} E(w_h|\mathbf{m}) E(\bar{x}'_h) + W_L \bar{X}'_L = \\ &= \sum_{h=1}^{L-1} W_h \bar{X}'_h + W_L \bar{X}'_L = \bar{X}', \end{aligned}$$

donde  $\bar{X}'_h$  es la media verdadera del estrato  $h$ -ésimo. La varianza es

$$\begin{aligned} V(\bar{x}') &= V\left(\sum_{h=1}^{L-1} W_h \bar{x}'_h + W_L \bar{X}'_L\right) + E\left[\sum_{h=1}^{L-1} \frac{W_h^* (1 - W_h^*)}{m_h} \bar{x}'_h{}^2\right] = \\ &= \sum_{h=1}^{L-1} W_h^2 \frac{\sigma_h'^2}{n_h} + \sum_{h=1}^{L-1} W_h^* (1 - W_h^*) E\left(\frac{1}{m_h}\right) E(\bar{x}'_h{}^2), \end{aligned}$$

donde

$$\sigma_h'^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (x'_{hi} - \bar{X}'_h)^2$$

es la varianza verdadera en el estrato  $h$ -ésimo ( $h = 1, 2, \dots, L-1$ ) y  $x'_{hi}$  es la variable  $x'$  en la  $i$ -ésima unidad del estrato  $h$ . Un estimador insesgado de esta varianza es

$$\hat{V}(\bar{x}') = \sum_{h=1}^{L-1} \left[ w_h^2 - \hat{V}(w_h) \right] \frac{s_h'^2}{n_h} + \sum_{h=1}^{L-1} \frac{m - N_L}{m - N_L - 1} w_h (1 - w_h) \frac{\bar{x}'_h{}^2}{m_h},$$

donde

$$w_h = \frac{m_h}{m - N_L}, \quad s_h'^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x'_{hi} - \bar{x}'_h)^2, \quad (h = 1, 2, \dots, L - 1)$$

y

$$\begin{aligned} \hat{V}(w_h) &= \hat{V}\left(\frac{m_h}{m - N_L}\right) = \frac{1}{(m - N_L)^2} \hat{V}(m_h) = \frac{1}{m - N_L} [W_h^* (\widehat{1 - W_h^*})] = \\ &= \frac{w_h(1 - w_h)}{m - N_L - 1}, \quad \text{para } h = 1, 2, \dots, L - 1. \end{aligned}$$

Por ello, el estimador insesgado del total verdadero es  $N\bar{x}'$  y su estimador insesgado de la varianza es  $N^2\hat{V}(\bar{x}')$ .

Un caso particular es cuando para  $h = 1, 2, \dots, L - 1$ , tenemos  $n_h = km_h$  ( $0 < k \leq 1$ ) y  $n_L = N_L$ , y entonces

$$n = \sum_{h=1}^L n_h = \sum_{h=1}^{L-1} n_h + N_L = k \sum_{h=1}^{L-1} m_h + N_L = k(m - N_L) + N_L$$

de donde

$$k = \frac{n - N_L}{m - N_L},$$

y por ello,

$$n_h = \frac{n - N_L}{m - N_L} m_h \quad (h = 1, 2, \dots, L - 1) \quad \text{y} \quad n_L = N_L.$$

(b) Otro diseño muestral es considerar una "muestra aleatoria simple con reemplazamiento" de tamaño  $m$  sobre la totalidad de la población finita, y dentro de cada estrato  $h$  ( $h = 1, 2, \dots, L - 1$ ) se obtiene una muestra de tamaño  $m_h$  y se submuestra con diseño masr de tamaño  $n_h$  (sobre la muestra ordenada de tamaño  $m_h$  obtenida previamente) donde se observa la variable verdadera  $x'$ , y  $n_L = m_L$  es el correspondiente tamaño muestral para el estrato  $L$ .

El segundo estimador propuesto para la media verdadera es

$$\bar{x}'' = \sum_{h=1}^{L-1} w_h \bar{x}'_h + W_L \bar{x}'_L.$$

Obviamente es insesgado pues

$$E(\bar{x}'') = \bar{X}',$$

y la varianza resulta ser

$$V(\bar{x}'') = V\left(\sum_{h=1}^L W_h \bar{x}'_h\right) + E\left[\sum_{h=1}^{L-1} \frac{W_h^*(1-W_h^*)}{m_h} \bar{x}'_h{}^2\right] = \\ = \sum_{h=1}^L W_h^2 \frac{\sigma_h'^2}{n_h} + \sum_{h=1}^{L-1} W_h^*(1-W_h^*) E\left(\frac{1}{m_h}\right) E(\bar{x}'_h{}^2),$$

por lo que un estimador insesgado de esta varianza es

$$\hat{V}(\bar{x}'') = \sum_{h=1}^{L-1} [w_h^2 - \hat{V}(w_h)] \frac{s_h'^2}{n_h} + W_L^2 \frac{s_L'^2}{n_L} + \sum_{h=1}^{L-1} \frac{m}{m-1} w_h(1-w_h) \frac{\bar{x}'_h{}^2}{m_h},$$

con

$$\hat{V}(w_h) = \frac{1}{m^2} \hat{V}(m_h) = \frac{1}{m^2} m [W_h^*(1-W_h^*)] = \frac{w_h(1-w_h)}{m-1}$$

para  $h = 1, 2, \dots, L-1$ , y

$$s_h'^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x'_{hi} - \bar{x}'_h)^2$$

para  $h = 1, 2, \dots, L$ .

Aquí también, como en (a), el estimador insesgado del total verdadero es  $N\bar{x}''$  y su estimador insesgado de la varianza es  $N^2\hat{V}(\bar{x}'')$ .

Otro caso particular es cuando  $n_h = km_h$  ( $h = 1, 2, \dots, L-1$ ) con  $0 < k \leq 1$ , y  $n_L = m_L$ . Así pues

$$n = \sum_{h=1}^L n_h = k \sum_{h=1}^{L-1} m_h + m_L = k \sum_{h=1}^L m_h - km_L + m_L = km + (1-k)m_L,$$

donde

$$k = \frac{n - m_L}{L-1} = \frac{n - m_L}{\sum_{h=1}^{L-1} m_h},$$

y por ello

$$n_h = \frac{n - m_L}{m - m_L} m_h \quad (h = 1, 2, \dots, L-1) \quad \text{y} \quad n_L = m_L.$$

### Bibliografía

- [1] Glasser, G.J. (1962). On the complete coverage of large units in a statistical study. *Rev. Internat. Statist. Inst.* **30**, 28-32.
- [2] Wolter, K.M. (1985). *Introduction to Variance Estimation*. Springer-Verlag. Nueva York.

NOTA SOBRE ESTRATEGIAS MUESTRALES  
PARA ESTUDIOS SOCIOECONÓMICOS EN ESPAÑA

por

M. Ruiz Espejo

*Departamento de Economía Aplicada Cuantitativa*

*Facultad de Ciencias Económicas y Empresariales*

*Universidad a Distancia*

*C/ Senda del Rey, s/n*

*28040 Madrid*

**Summary.** For the socioeconomic studies by sampling in Spain, it is used a stratified sampling design of *clusters* (censal sections of unequal size) with subsampling of hearths, and usually the individual are subsampled. For this sampling design, we propose an unbiased estimator of the population mean and from here we propose an unbiased variance estimator. This is valid moreover with nonresponse, and response (or measure) errors.

### 1. Introducción

Los estudios socioeconómicos por muestreo a escala nacional en España, tal y como actualmente realiza el Instituto Nacional de Estadística y empresas colaboradoras, se basan en estratificar los hogares o los individuos personales por secciones censales; en cada sección censal seleccionada en un estrato, se submuestran los hogares (como unidades últimas en algunos casos) o bien se submuestran nuevamente cada hogar obteniendo una muestra final de individuos: este último será nuestro punto de partida en el contenido sucesivo del presente trabajo.

El procedimiento utilizado hasta ahora trata la no respuesta de una unidad con sustituciones de nuevas unidades finales en una lista ordenada de unidades alternativas, método que no garantiza la eliminación de sesgos como consecuencia de tales sustituciones y de no buscar la verdadera respuesta en una muestra del sector de no respondientes. Además no se inspeccionan las respuestas anotadas, dando por ciertos unos datos que no se corresponden con la realidad como consecuencia de respuestas deliberadamente falsas, errores de prestigio (consistentes en dar una respuesta más de apariencia y que pretendidamente dé mejor imagen, pero no real), etc. que contribuyen a nuevos sesgos de estimación.

La solución a estos problemas, es el objeto de esta nota de investigación estadística, que garantiza una estrategia muestral que afronta la complejidad del diseño muestral que aprovecha las secciones censales administrativas, proponiendo un estimador insesgado para la media poblacional de interés frente a la no respuesta y errores de respuesta independientemente de sus causas. También se estima sin sesgo la varianza de tal estrategia, lo que permite conocer su variabilidad estimada y como consecuencia, estimar su error de muestreo para construir intervalos de confianza aproximados.

## 2. Metodología de la solución

Sea  $L$  el número de estratos, y  $W_h$  el tamaño relativo del estrato  $h$  ( $= 1, 2, \dots, L$ ), es decir el número de individuos del estrato  $h$  entre el número de individuos total en los  $L$  estratos. El estimador insesgado de la media poblacional  $\bar{Y}$  será:

$$\hat{t} = \sum_{h=1}^L W_h t_h,$$

donde  $t_h$  es el estimador Hansen-Hurwitz en el estrato  $h$ ,

$$t_h = \sum_{k \in s_1} \frac{\tau_{hk}}{N_h n_h p_{hk}},$$

donde  $s_1$  es una muestra de  $n_h$  conglomerados de entre los  $N_h$  existentes (secciones censales del estrato  $h$ ) muestra obtenida con diseño de probabilidades proporcionales al tamaño ( $p_{hk}$ ) con reemplazamiento (*pptr*). Además si  $p'_{hk}$  es el número de hogares en el conglomerado  $k$  del estrato  $h$ , y  $p''_h$  es el número de hogares en el estrato  $h$ , entonces el tamaño normalizado es

$$p_{hk} = \frac{p'_{hk}}{p''_h} \begin{cases} h = 1, 2, \dots, L \text{ y} \\ k = 1, 2, \dots, N_h. \end{cases}$$

Además, a su vez  $\tau_{hk}$  es el estimador Hansen-Hurwitz en el conglomerado  $k$  del estrato  $h$ ,

$$\tau_{hk} = \sum_{i \in s_2} \frac{\tau_{hki}}{M_{hk} m_{hk} p_{hki}},$$

donde  $s_2$  es una muestra de  $m_{hk}$  hogares (entre los  $M_{hk}$  existentes) en el conglomerado  $k$  del estrato  $h$ , obtenida por diseño *pptr* con tamaño normalizado

$$p_{hki} = \frac{p'_{hki}}{p''_{hk}} \begin{cases} h = 1, 2, \dots, L, \\ k = 1, 2, \dots, N_h \text{ e} \\ i = 1, 2, \dots, M_{hk}, \end{cases}$$

donde  $p'_{hki}$  es el número de individuos en el hogar  $i$  del conglomerado  $k$  del estrato  $h$ , y  $p''_{hk}$  es el número de individuos en el conglomerado  $k$  del estrato  $h$ .

Finalmente,  $\tau_{hki}$  es la estrategia media muestral para muestreo aleatorio simple  $\bar{y}_{hki}$  (para  $s_3$ ) en el caso de no existir no respuesta y en ausencia de errores de respuesta; si existe no respuesta y errores de respuesta, un estimador insesgado es el presentado por Ruiz Espejo (1988), que incorpora la posibilidad de inspeccionar datos o disuadir a los individuos de que colaboren con respuestas veraces.

De este modo, el estimador  $t$  es insesgado para estimar el parámetro  $\bar{Y}$ . Esto puede demostrarse descomponiendo su esperanza del modo

$$E(t) = E(E\{E[E(t|s_3)|s_2]|s_1\}) = \bar{Y}.$$

Si tenemos varias estimaciones insesgadas  $t_1, t_2, \dots, t_r$  independientes con diseño muestral visto anteriormente aunque varíen sus tamaños muestrales, un estimador insesgado de  $\bar{Y}$ , es

$$t^* = \frac{1}{r} \sum_{i=1}^r t_i,$$

y un estimador insesgado de su varianza para  $r \geq 2$ , es (véase Wolter, 1985)

$$\hat{V}(t^*) = \frac{1}{r(r-1)} \sum_{i=1}^r (t_i - t^*)^2.$$

De este modo queda resuelto el problema metodológico para estudios socioeconómicos en España para el diseño muestral complejo usado habitualmente en la selección de la muestra de individuos, usando un diseño muestral probabilístico que permite estimar sin sesgo la media  $\bar{Y}$  mediante  $t^*$ , y estimar sin sesgo la varianza  $V(t^*)$ , que nos permite a su vez estimar el error de muestreo al construir intervalos de confianza para  $\bar{Y}$ , a un nivel de confianza acotado inferiormente por  $1 - \alpha$ , mediante la desigualdad de Chebychev.

### Referencias

- Ruiz Espejo, M. (1988). Estimación insesgada con observaciones erradas y no respuesta. *Trabajos de Estadística* 3, 71-80.
- Wolter, K.M. (1985). *Introduction to Variance Estimation*. Springer-Verlag. Nueva York.

MUESTREO DOBLE POSESTRATIFICADO ESPECIAL  
PARA CONTROL DE CUENTAS ECONÓMICAS

por

M. Ruiz Espejo

*Departamento de Economía Aplicada Cuantitativa*

*Facultad de Ciencias Económicas y Empresariales*

*Universidad a Distancia*

*C/ Senda del Rey, s/n*

*28040 Madrid*

**Summary.** We provide a new sampling method for the control of economic accounts, alternative to the proposed in the literature, of unbiased type and with estimable variance adapted to skewed populations.

### 1. Introducción

Existen diversas metodologías para el control por muestreo de cuentas económicas. Algunas de ellas están recogidas en las referencias citadas.

En este trabajo estudiamos otra metodología alternativa a las anteriores, en la que estimamos insesgadamente el *total económico verdadero*, el *total defraudado* y la *proporción de defraudadores*, y estimamos insesgadamente la *varianza* del estimador propuesto para el total verdadero.

### 2. Diseño muestral

Sea una población finita de  $N$  unidades o sumandos de un total económico que denotamos total verdadero

$$X' = \sum_{i=1}^N x'_i,$$

donde el valor  $x'_i$  es el valor verdadero en el sumando  $i$ -ésimo ( $i = 1, 2, \dots, N$ ). La unidad  $i$  tiene un valor declarado  $x_i$ , de modo que el total declarado es

$$X = \sum_{i=1}^N x_i.$$

Suponemos que el número de sumandos o unidades es conocido de antemano. Queremos estimar el total verdadero  $X'$ , el fraude total  $F = X' - X$  y la proporción de unidades defraudadoras  $\bar{D}$ . Para ello, y como  $N$  es conocido, el primer problema estadístico-económico a resolver es equivalente a estimar la media poblacional  $\bar{X}' = X'/N$ . El diseño muestral *doble estratificado especial* que aportamos consiste en dos fases:

**1ª fase.** Se selecciona una muestra aleatoria simple con reemplazamiento,  $s$ , de tamaño  $m$  (masr- $m$ ) de la población finita, donde se observa el valor declarado de cada unidad y con esta información se procede a estratificar la población y por ello a  $s$  en  $L$  estratos. Denotando por  $s_i$  a la submuestra ordenada de  $s$  contenida en el estrato  $i$ -ésimo ( $i = 1, 2, \dots, L$ ), llamamos  $m_i$  al tamaño muestral de  $s_i$ . Es evidente que

$$m = \sum_{i=1}^L m_i.$$

Usamos el vector aleatorio  $\mathbf{m} = (m_1, m_2, \dots, m_L)$  con distribución multinomial (véase Casas Sánchez *et al.*, 1996). Sea  $w_i = m_i/m$ .

**2ª fase.** Si el estrato  $i$  contiene  $m_i$  unidades de la muestra  $s_i$  contenida en  $s$ , se procede a extraer una muestra aleatoria simple con reemplazamiento de tamaño  $n_i$  (masr- $n_i$ ) del conjunto ordenado  $s_i$ . Usamos muestreo estratificado especial (para poblaciones oblicuas de tipo económico) pues

$$n_i = km_i \text{ para } i = 1, 2, \dots, L - 1 \quad (0 < k < 1)$$

$$\text{y } n_L = m_L \text{ (asignación especial en unidades grandes).}$$

En la submuestra de tamaño

$$n = \sum_{i=1}^L n_i,$$

se observa por inspección el valor verdadero  $x'$  en las unidades seleccionadas.

Los tamaños relativos de los estratos  $W_i = N_i/N$  ( $i = 1, 2, \dots, L$ ) son desconocidos; es necesario para la estimación insesgada de la varianza (en el sentido de Wolter, 1985) usar diseño masr- $n_i$  como veremos, en lugar de muestreo aleatorio simple sin reemplazamiento mas- $n_i$ . Aunque se pierde algo de precisión al usar diseño masr- $n_i$  en lugar de mas- $n_i$ , existen dos ventajas que nos inclinan a su uso:

a) El diseño masr es más económico que el mas, pues pueden aparecer unidades repetidas en la muestra y, en tal caso, basta observar solo una vez tal unidad, ahorrando el gasto de observación de las restantes repeticiones.

b) Además el diseño masr permite estimar sin sesgo la varianza de la media muestral en el estrato  $i$ -ésimo  $\bar{x}'_i$ , mediante

$$\hat{V}(\bar{x}'_i) = \frac{m_i}{m_i - 1} \frac{s_{i2}^2}{n_i} \quad (i = 1, 2, \dots, L - 1) \text{ y } \hat{V}(\bar{x}'_L) = \frac{s_{L2}^2}{n_L}$$

siendo  $s_{i2}^2$  la cuasivarianza muestral de la submuestra de tamaño  $n_i$ , del estrato  $i$ -ésimo. Si utilizáramos diseño mas- $n_i$  (siendo  $S_i'^2$  la cuasivarianza del estrato  $i$ ) la varianza sería

$$V(\bar{x}'_i) = \left( \frac{1}{n_i} - \frac{1}{N_i} \right) S_i'^2,$$

que contiene la constante desconocida  $N_i$  que hace difícilmente estimable insesgadamente a dicha varianza.

### 3. Estimadores insesgados

El estimador insesgado de  $X'$  es  $N\bar{x}'$ , donde

$$\bar{x}' = \sum_{i=1}^L w_i \bar{x}'_i,$$

siendo

$$\bar{x}'_i = \frac{1}{n_i} \sum_{j \in s_{i2}} x'_j$$

de la media de la muestra  $s_{i2}$  ( $\subset s_i$ ). En efecto,

$$\begin{aligned} E(\bar{x}') &= E[E(\bar{x}'|\mathbf{m})] = E\left[E\left(\sum_{i=1}^L w_i \bar{x}'_i | \mathbf{m}\right)\right] = E\left[\sum_{i=1}^L w_i E(\bar{x}'_i | \mathbf{m})\right] \\ &= E\left(\sum_{i=1}^L w_i \bar{X}'_i\right) = \sum_{i=1}^L E(w_i) \bar{X}'_i = \sum_{i=1}^L W_i \bar{X}'_i = \bar{X}', \end{aligned}$$

donde  $W_i$  y  $\bar{X}'_i$  son el "peso relativo" y la "media" del estrato  $i$ -ésimo.

El estimador insesgado de  $F = X' - X$  es

$$\hat{F} = N(\bar{x}' - \bar{x})$$

con  $\bar{x}$  la media muestral para la variable declarada  $x$  y para la muestra  $s$ .

El estimador insesgado de  $\bar{D}$  es  $\bar{x}'$ , donde ahora

$$x'_j = \begin{cases} 1 & \text{si la unidad } j \text{ sí defrauda} \\ 0 & \text{si la unidad } j \text{ no defrauda} \end{cases}$$

### 4. Estimación de la varianza

Aplicando el teorema de Madow,

$$\begin{aligned} V(\bar{x}') &= V[E(\bar{x}'|\mathbf{m})] + E[V(\bar{x}'|\mathbf{m})] = \\ &= V\left(\sum_{i=1}^L w_i \bar{X}'_i\right) + E\left(\sum_{i=1}^L w_i^2 \frac{\sigma_i'^2}{n_i}\right) = (*) \end{aligned}$$

donde  $\sigma_i'^2$  es la varianza de la variable  $x'$  en el estrato  $i$ .

$$(*) = \sum_{i=1}^L \bar{X}_i'^2 V(w_i) + \sum_{i=1}^{L-1} \sum_{j=i+1}^L \bar{X}_i' \bar{X}_j' Cov(w_i, w_j) + \sum_{i=1}^L E(w_i^2) \frac{\sigma_i'^2}{n_i}$$

siendo

$$V(w_i) = \frac{1}{m} W_i (1 - W_i)$$

y si  $i \neq j$ ,

$$Cov(w_i, w_j) = -\frac{1}{m} W_i W_j.$$

Luego un estimador insesgado para la varianza  $V(\bar{x}')$ , es:

$$\hat{V}(\bar{x}') = \sum_{i=1}^L \widehat{X}_i'^2 \hat{V}(w_i) + 2 \sum_{i=1}^{L-1} \sum_{j=i+1}^L \widehat{X}_i' \widehat{X}_j' \widehat{Cov}(w_i, w_j) + \sum_{i=1}^L w_i^2 \frac{\widehat{\sigma}_i'^2}{n_i}$$

donde

$$\hat{V}(w_i) = \frac{1}{m-1} w_i (1 - w_i),$$

$$\widehat{Cov}(w_i, w_j) = -\frac{1}{m-1} w_i w_j,$$

$$\widehat{X}_i'^2 = \bar{x}_i'^2 + \hat{V}(\bar{x}_i') = \bar{x}_i'^2 + \frac{\sigma_i'^2}{n_i},$$

$$\widehat{X}_i' = \bar{x}_i'$$

y si  $i = 1, 2, \dots, L-1$ ,

$$\widehat{\sigma}_i'^2 = \frac{m_i}{m_i - 1} s_{i2}^2$$

donde  $s_{i2}^2$  es la cuasivarianza de la muestra  $s_{i2}$  de tamaño  $n_i$ , para la variable  $x'$ , y

$$\widehat{\sigma}_L'^2 = s_{L2}^2$$

que es la cuasivarianza de la muestra  $s_L$  de tamaño  $m_L$ , para la variable  $x'$ .

### Referencias

Casas Sánchez, J.M. y Santos Peñas, J. (1996). *Introducción a la Estadística para Economía y Administración de Empresas*. CERA. Madrid.

Ruiz Espejo, M. (1997). Una metodología para el control por muestreo de cuentas económicas. *Estudios de Economía Aplicada*, en prensa.

——— (1997). Muestreo doble posestratificado proporcional para cuentas económicas. *Estudios de Economía Aplicada* (enviado).

Wolter, K.M. (1985). *Introduction to Variance Estimation*. Springer-Verlag. Nueva York.

EL TEOREMA DE RAO-BLACKWELL EN POBLACIONES  
FINITAS E IMPLICACIONES INFORMATIVO-ECONOMICAS

M. Ruiz Espejo

Deptº de Economía Aplicada Cuantitativa.

Fac. de Ciencias Económicas y Empresariales.

Univ. a Distancia. c/ Senda del Rey. s/n. 28040-Madrid.

**Summary.** We give a new version of the Rao-Blackwell theorem adapted to the sampling for finite populations. As a consequence we obtain an essentially complete class of estimators for any parametric function in the fixed population model. Some informative-economic implications completes the study.

1. Introducción

Siguiendo básicamente la notación de Cassel *et al.* (1977), el estadístico suficiente  $D = r(\mathbf{D}) = \{(k, y_k) : k \in \mathbf{S}\}$  puede ser utilizado con propósitos de estimación bajo diseño ordenado. Por el teorema de Rao-Blackwell, cualquier estimador  $t$  que depende de del estadístico ordenado  $\mathbf{D}$ , admite otro mejor o equivalente, en sentido de error cuadrático medio que depende de  $\mathbf{D}$  solo a través de  $D = r(\mathbf{D})$ . La no existencia de estimador uniformemente de mínima varianza se debe a la incompletitud del estadístico  $D$ . El teorema fundamental sobre completitud en teoría de muestras se debe a Liu (1983) que generaliza resultados anteriores de Watson (1964), Royall (1968) y de Bellhouse, Thompson y Godambe (1977).

*Definición 1.1.* Dado un diseño muestral  $p$ , un estimador  $t_1$  es mejor o equivalente al estimador  $t_2$  si  $R(\mathbf{y}; p, t_1) \leq R(\mathbf{y}; p, t_2)$  para todo  $\mathbf{y} \in \mathbb{R}^N$ , siendo  $R$  el riesgo asociado a un parámetro  $\mathbf{y}$  y una estrategia muestral  $(p, t)$ . El error cuadrático medio es un caso particular de riesgo en el que la función de pérdida es la cuadrática.

Dado un estimador  $t = t(\mathbf{D})$  no necesariamente insesgado de la función paramétrica  $f(\mathbf{y})$ , para cada dato  $d$  del estadístico suficiente minimal  $D = r(\mathbf{D})$ , definimos el estimador Rao-Blackwell-izado

$$t_{RB}(d) = E[t(\mathbf{D}) | D = d] = \frac{\sum_{\mathbf{s} \in r^{-1}(s)} t(\mathbf{d}) p(\mathbf{s})}{\sum_{\mathbf{s} \in r^{-1}(s)} p(\mathbf{s})} \quad (1)$$

siendo  $r$  la función de reducción de datos, que pierde la información relativa al orden y multiplicidad de unidades seleccionadas en la muestra ordenada  $s$ , obtenida según el diseño muestral  $p$ .

El teorema de Rao-Blackwell fue considerado en el ámbito de teoría de muestras por primera vez por Basu (1958).

Veamos el siguiente resultado que se basa en los teoremas proporcionados por Cassel *et al.* (1977) y por Stenger (1977).

## 2. Resultado

**Teorema 2.1.** Sea  $t = t(\mathbf{D})$  un estimador no necesariamente insesgado de la función paramétrica  $f(\mathbf{y})$ , y sea  $t_{RB} = t_{RB}(D)$  el estimador Rao-Blackwell-izado asociado a  $t$ , definido en (1). Sea  $L(\mathbf{y}, t)$  una función de pérdida convexa en  $t$  y la función de riesgo

$$R(\mathbf{y}; p, t) = E_p[L(\mathbf{y}, t)] = \sum_{s \in \mathcal{S}} L[\mathbf{y}, t(\mathbf{d})] p(s).$$

Entonces:

- a)  $E(t) = E(t_{RB})$ ,  $\mathbf{y}$   
 b)  $R(\mathbf{y}; p, t_{RB}) \leq R(\mathbf{y}; p, t)$  para todo  $\mathbf{y} \in \mathbb{R}^N$ .  $\square$

*Demostración*

a)

$$\begin{aligned} E(t) &= \sum_{s \in \mathcal{S}} t(\mathbf{d}) p(s) = \sum_{s \in \mathcal{S}} \sum_{s \in r^{-1}(s)} t(\mathbf{d}) p(s) = \\ &= \sum_{s \in \mathcal{S}} \left[ \sum_{s \in r^{-1}(s)} p(s) \right] \frac{\sum_{s \in r^{-1}(s)} t(\mathbf{d}) p(s)}{\sum_{s \in r^{-1}(s)} p(s)} = \sum_{s \in \mathcal{S}} p(s) t_{RB}(s) = E(t_{RB}). \end{aligned}$$

b)

$$\begin{aligned} R(\mathbf{y}; p, t) &= \sum_{s \in \mathcal{S}} L[\mathbf{y}; t(\mathbf{d})] p(s) = \sum_{s \in \mathcal{S}} \sum_{s \in r^{-1}(s)} L[\mathbf{y}; t(\mathbf{d})] p(s) = \\ &= \sum_{s \in \mathcal{S}} \left[ \sum_{s \in r^{-1}(s)} p(s) \right] \cdot \sum_{s \in r^{-1}(s)} L[\mathbf{y}, t(\mathbf{d})] \frac{p(s)}{\sum_{s \in r^{-1}(s)} p(s)} \geq \\ &\geq \sum_{s \in \mathcal{S}} \left[ \sum_{s \in r^{-1}(s)} p(s) \right] \cdot L \left[ \mathbf{y}, \frac{\sum_{s \in r^{-1}(s)} t(\mathbf{d}) p(s)}{\sum_{s \in r^{-1}(s)} p(s)} \right] = \\ &= \sum_{s \in \mathcal{S}} p(s) L[\mathbf{y}, t_{RB}(d)] = R(\mathbf{y}; p, t_{RB}), \end{aligned}$$

para todo  $\mathbf{y} \in \mathbb{R}^N$ , aplicando la desigualdad de Jensen.  $\square$

### 3. Consecuencias

*Definición 3.1.* Diremos que el conjunto  $C$  de estimadores es esencialmente completo si bajo un diseño muestral  $p$ , para todo estimador  $t' \notin C$ , existe  $t \in C$  que es mejor o equivalente a  $t'$ .

*Corolario 3.2.* Si la función de pérdida  $L(\mathbf{y}, t)$  es convexa en la componente  $t$ , en particular con la función de pérdida cuadrática  $L(\mathbf{y}, t) = [f(\mathbf{y}) - t]^2$ , bajo el diseño muestral  $p$ , el conjunto de estimadores  $C = \{t: t = t(D)\}$  es esencialmente completo.  $\square$

*Demostración.* Es consecuencia del teorema 2.1, siendo  $C$  el conjunto de estimadores que dependen solo de  $D$ , como ocurre en el caso de estimadores Rao-Blackwell-izados.  $\square$

Un resultado análogo puede obtenerse para estrategias muestrales. Otras clases completas en teoría de muestras han sido estudiadas por Stenger y Gabler (1981), Roy y Chakravarti (1960) y Ramakrishnan (1975).

### 4. Aplicaciones

El tamaño muestral fijo de un diseño muestral no ordenado equivale a tener un costo fijo, con sus correspondientes implicaciones informativo-económicas.

El estimador Hansen-Hurwitz (1943), más próximamente estudiado por Jessen (1978), y el estimador Sánchez-Crespo (1980, p. 154) coinciden en su expresión

$$\hat{Y} = \frac{X}{nN} \sum_{i=1}^N \frac{Y_i}{X_i} e_i,$$

donde para Hansen y Hurwitz  $e_i$  sigue una distribución multinomial, mientras que para Sánchez-Crespo  $e_i$  sigue una distribución hipergeométrica generalizada (Sukhatme y Sukhatme, 1970). En ambos casos, dichas estrategias admiten otras también insesgadas para estimar la media poblacional

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i,$$

que mejoran las precisiones respectivas (véase Ruiz Espejo, 1988).

Los métodos de remuestreo como las técnicas bootstrap no mejoran la precisión de estimadores Rao-Blackwell-izados pues aquéllos se basan en muestras ordenadas y posibles repeticiones de las mismas observaciones.

## Referencias

- Basu, D. (1958). On sampling with and without replacement. *Sankhyā* **20**, 287-294.
- Bellhouse, D.R., Thompson, M.E. y Godambe, V.P. (1977). Two-stage sampling with exchangeable prior distributions. *Biometrika* **64**, 97-103.
- Cassel, C.M., Särndal, C.E. y Wretman, J.H. (1977). *Foundations of Inference in Survey Sampling*. John Wiley. Nueva York.
- Hansen, M.H. y Hurwitz, W.N. (1943). On the theory of sampling from finite populations. *Ann. Math. Statist.* **14**, 333-362.
- Jessen, R.J. (1978). *Statistical Survey Techniques*. John Wiley. Nueva York.
- Liu, T.P. (1983). A general completeness theorem in sampling theory. *J. Roy. Statist. Soc. B* **45**, 369-372.
- Ramakrishnan, M.K. (1975). Choice of an optimum sampling strategy, I. *Ann. Statist.* **3**, 669-679.
- Roy, J. y Chakravarti, J.M. (1960). Estimating the mean of a finite population. *Ann. Math. Statist.* **31**, 392-398.
- Royall, R.M. (1968). An old approach to finite population sampling theory. *J. Amer. Statist. Assoc.* **63**, 1269-1279.
- Ruiz Espejo, M. (1988). Estimación insesgada generalizada en poblaciones finitas. *Qüestió* **12**, 315-321.
- Sánchez-Crespo, J.L. (1980). *Curso Intensivo de Muestreo en Poblaciones Finitas* (Segunda edición). Instituto Nacional de Estadística. Madrid.
- Stenger, H. (1977). Sequential sampling from finite populations. *Sankhyā Ser. C* **39**, 10-20.
- Stenger, H. y Gabler, S. (1981). On the completeness of the class of finite size sampling strategies. *Ann. Statist.* **9**, 229-232.
- Sukhatme, P.V. y Sukhatme, B.V. (1970). *Sampling Theory of Surveys with Applications*. Iowa State University Press. Ames.
- Watson, G.S. (1964). Estimation in finite population. Informe no publicado.

UNBIASED ESTIMATION OF ECONOMIC AND SOCIAL  
PARAMETERS WITH PROBABLE INDIVIDUAL RESPONSE

por

M. Ruiz Espejo

*Departamento de Economía Aplicada Cuantitativa*

*Facultad de Ciencias Económicas y Empresariales*

*Universidad a Distancia*

*C/ Senda del Rey, s/n*

*28040 Madrid*

**Summary.** We provide unbiased sampling strategies to estimate the population mean and variance of a certain economic or social interest variable, with probable individual response. This method is based on the knowledge of the individual response probability in the units of a population of respondents, selected in a simple random sample without replacement. We calculate the variance for the estimator of the mean, and we give a nonnegative and unbiased variance estimator.

### 1. Introduction

When with an unbiased strategy we estimate the economic or social mean of a finite population  $U = \{1, 2, \dots, N\}$  of size  $N$ , from a sample of fixed effective size  $n$  ( $1 \leq n \leq N$ ), it can occur that one or several of the selected units in the sample refuse to facilitate the required information to be able to realize the estimation as it was designed. The "individual response probability" has been estimated by Giommi (1987), and used among others such as Anderson (1979) and Eltinge (1992). If the sampling strategy is the classical (SRSWOR,  $\bar{y}$ ) characterized by "simple random sampling without replacement" design (SRSWOR), and  $\bar{y}$  is the sample mean estimator, Cochran (1977, section 13.2) has justified that the use of the sample mean of response  $\bar{y}_1$ , of the units selected in the initial sample  $s$  and moreover they respond (let us say  $s_1$ ), provides a bias to estimate the finite population mean  $\bar{Y}$ . Cochran (1977, chapter 13) gives some practical possible strategies to eliminate this bias. for example Politz and Simmons (1949, 1950). In this context, the general idea of weighting adjustment and imputation of data for nonresponse has been studied by Czajka et al. (1992), Little (1986), Oh and Scheuren (1983), Ruiz Espejo et al. (1995) and references cited therein, but usually these estimation methods are biased and not of easy unbiased error estimation.

In this paper we provide a new unbiased strategy to estimate the finite population mean  $\bar{Y}$ , when the individual response probabilities  $q_k$  ( $k \in s$ ) are known in all the units of the effective sample  $s$ . Moreover, we suppose that the response probability is positive for all  $k \in U$ . The main new contribution here involves the possibility of allowing that each individual unit to have a different response probability, in contrast with most "quasirandomization-based" nonresponse work, which assumes that response probabilities are equal within certain "response homogeneity groups" or "cells".

If the selected sample under SRSWOR design is  $s$ , of size  $n$ , and the respondents are a random number  $n_1$  of units of the subsample  $s_1$ ,  $\phi \neq s_1 \subset s$  with  $1 \leq n_1 \leq n$ , the nonrespondent are  $n_2 = n - n_1$ . We can consider the estimator

$$\bar{y}' = \frac{1}{n} \sum_{k \in s_1} \frac{y_k}{p_k}$$

where  $p_k = q_k/p$  ( $s_1 \neq \phi$ ) and where  $y_k$  is the observed interest variable in unit  $k \in s_1$ , being not necessary the values  $y_k$  when  $k \in s - s_1$ .

We introduce the indicator  $I_k(s_1)$  of the event  $\{k \in s_1\}$  conditioned by the event  $\{k \in s\}$ , that is to say, if  $k \in s$ :  $I_k(s_1) = 1$  if  $k \in s_1$ ; and  $I_k(s_1) = 0$  if  $k \notin s_1$ ; and then we can rewrite the estimator  $\bar{y}'$  in the equivalent expression

$$\bar{y}' = \frac{1}{n} \sum_{k \in s} \frac{y_k}{p_k} I_k(s_1),$$

being  $p_k = p(k \in s_1 | k \in s) = p[I_k(s_1) = 1 | k \in s]$ , which for the case of independent individual response probabilities  $q_k$  ( $k \in s$ ), is

$$p_k = \frac{q_k}{1 - \prod_{i \in s} (1 - q_i)} \quad \text{for all } k \in s,$$

where  $p_k$  is the probability of the "response of the unit  $k$ " conditioned by the event (" $s_1 \neq \phi$ " intersection with " $s$ "). The probability of this event is

$$p[(s_1 \neq \phi) \cap s] = 1 - \prod_{i \in s} (1 - q_i).$$

In section 2 we will see that this estimator,  $\bar{y}'$ , is unbiased for estimating the finite population mean  $\bar{Y}$  with probable individual response, in section 3 we calculate its variance  $V(\bar{y}')$ , and in section 4 we propose unbiased and nonnegative estimators  $\hat{\sigma}^2$  and  $\hat{V}(\bar{y}')$  for  $\sigma^2$  and  $V(\bar{y}')$  respectively.

From a formal viewpoint, the estimator  $\bar{y}'$  is subjected to a two-phase sampling design; in the first phase we use SRSWOR of size  $n$  to select the sample  $s$ , and in the second phase each unit  $k \in s$  is drawn in  $s_1 \neq \phi$  with response probability  $p_k$ . Thus, in the present paper we use a natural quasirandomization approach, distinct of the fixed-nonresponse approach emphasized by Cochran (1977, chapter 13).

## 2. Unbiased estimation of the mean

If  $k \in s$ ,  $I_k(s_1)$  follows the binomial distribution  $B(1, p_k)$  or the Bernoulli distribution  $B(p_k)$ . Then using the expression

$$\begin{aligned} E(\bar{y}') &= E_1[E_2(\bar{y}'|s)] = E_1\left\{E_2\left[\frac{1}{n}\sum_{k \in s} \frac{y_k}{p_k} I_k(s_1) | s\right]\right\} = \\ &= E_1\left\{\frac{1}{n}\sum_{k \in s} \frac{y_k}{p_k} E_2[I_k(s_1)|s]\right\} = E_1\left(\frac{1}{n}\sum_{k \in s} \frac{y_k}{p_k} p_k\right) = E_1(\bar{y}) = \bar{Y}, \end{aligned}$$

our strategy is unbiased for estimating the population mean  $\bar{Y}$  with probable individual response in the non empty sample  $s_1$ .

## 3. Variance of the strategy

Applying Madow theorem (see Cochran, 1977, section 10.2),

$$V(\bar{y}') = E_1 V_2(\bar{y}'|s) + V_1 E_2(\bar{y}'|s)$$

where

$$\begin{aligned} V_2(\bar{y}'|s) &= V_2\left[\frac{1}{n}\sum_{k \in s} \frac{y_k}{p_k} I_k(s_1) | s\right] = \frac{1}{n^2} \sum_{k \in s} \sum_{i \in s} \frac{y_k y_i}{p_k p_i} \text{Cov}_2[I_k(s_1), I_i(s_1) | s] = \\ &= \frac{1}{n^2} \sum_{k \in s} \sum_{i \in s} \frac{y_k y_i}{\sqrt{p_k p_i}} \rho_{ki}(s_1|s) \sqrt{(1-p_k)(1-p_i)} = \frac{1}{n^2} \sum_{k \in s} \sum_{i \in s} A_{ki}, \end{aligned}$$

in which  $\rho_{ki}(s_1|s)$  is the correlation coefficient of the random variables  $I_k(s_1)$  and  $I_i(s_1)$  conditioned by  $s$ . Thus  $\rho_{ki}(s_1|s)$  depends of all the units of  $s$ . If, in the following, we assume that  $\rho_{ki}(s_1|s) = \rho_{ki}$  (is independent of  $s$  and  $s_1$ ), being  $\rho_{ii} = 1$ ,

$$A_{ki} = y_k y_i \rho_{ki} \sqrt{\frac{(1-p_k)(1-p_i)}{p_k p_i}}$$

and

$$A_{ii} = A_i = y_i^2 \frac{1-p_i}{p_i}.$$

We have used that  $V_2[I_k(s_1) | s] = p_k(1-p_k)$  for all  $k \in s$ .

If we call  $S = \{s : s \subset U, \text{card}(s) = n\}$ , with SRSWOR,  $p(s) = 1/\binom{N}{n}$  for all sample  $s$  of  $S$ . Then,

$$E_1 V_2(\bar{y}'|s) = E_1\left(\frac{1}{n^2} \sum_{k \in s} \sum_{i \in s} A_{ki}\right) = \sum_{s \in S} \left(\frac{1}{n^2} \sum_{k \in s} \sum_{i \in s} A_{ki}\right) p(s) =$$

$$= \frac{1}{n^2} \frac{1}{\binom{N}{n}} \left( \sum_{\substack{k \in U \\ i \in U \\ i \neq k}} A_{ki} \text{card}\{s \in S : k, i \in s\} + \sum_{i \in U} A_i \text{card}\{s \in S : i \in s\} \right) =$$

$$= \frac{n-1}{nN(N-1)} \sum_{\substack{k \in U \\ i \in U \\ i \neq k}} A_{ki} + \frac{1}{nN} \sum_{i \in U} A_i,$$

because if  $k \neq i \in U$ ,  $\text{card}\{s \in S : k, i \in s\} = \binom{N-2}{n-2}$ , and  $\text{card}\{s \in S : i \in s\} = \binom{N-1}{n-1}$ .

On the other hand, as in section 2,  $E_2(\bar{y}'|s) = \bar{y}$  and consequently

$$V_1 E_2(\bar{y}'|s) = V_1(\bar{y}) = \frac{N-n}{N} \frac{S_y^2}{n}$$

(from Cochran, 1977, section 2.5) where

$$S_y^2 = \frac{1}{N-1} \sum_{k \in U} (y_k - \bar{Y})^2 = \frac{N}{N-1} \sigma^2$$

is the population quasivariance. From the previous results we have that

$$V(\bar{y}') = \frac{n-1}{nN(N-1)} \sum_{\substack{k \in U \\ i \in U \\ i \neq k}} A_{ki} + \frac{1}{nN} \sum_{i \in U} A_i + \frac{N-n}{N} \frac{S_y^2}{n}$$

is the variance of the strategy proposed in section 1.

#### 4. Unbiased variance estimation

An unbiased estimator of the variance  $V(\bar{y}')$ , is

$$\hat{V}(\bar{y}') = \frac{n-1}{nN(N-1)} \sum_{\substack{k \in s_1 \\ i \in s_1 \\ i \neq k}} \frac{A_{ki}}{E[I_k(s_1)I_i(s_1)]} + \frac{1}{nN} \sum_{i \in s_1} \frac{A_i}{p_i} + \frac{N-n}{N-1} \frac{\hat{\sigma}^2}{n}$$

where

$$E[I_k(s_1)I_i(s_1)] = \text{Cov}[I_k(s_1), I_i(s_1)] + E[I_k(s_1)]E[I_i(s_1)] =$$

$$= \rho_{ki} \sqrt{p_k(1-p_k)p_i(1-p_i)} + p_k p_i$$

and  $\rho_{ki}$  must be known, except when  $\rho_{ki} = 0$ , e.g. in case of independence between  $k$  and  $i$ ; from  $s_1$ , an unbiased and nonnegative estimator of the population variance  $\sigma^2$  is, according to Murthy (1963):

$$\hat{\sigma}^2 = \frac{1}{N^2} \sum_{k \in s_1} \sum_{\substack{i \in s_1 \\ i > k}} \frac{(y_k - y_i)^2}{E[I_k(s_1)I_i(s_1)]}.$$

If moreover  $\rho_{ki} = 0$  for all  $k \neq i \in s_1$ , a nonnegative and unbiased estimator of  $V(\bar{y}')$ , particular case of  $\hat{V}(\bar{y}')$ , is

$$\hat{V}^*(\bar{y}') = \frac{1}{nN} \sum_{i \in s_1} \frac{y_i^2(1-p_i)}{p_i^2} + \frac{N-n}{nN^2(N-1)} \sum_{k \in s_1} \sum_{\substack{i \in s_1 \\ i > k}} \frac{(y_k - y_i)^2}{p_k p_i}.$$

The estimators  $\hat{V}(\bar{y}')$  and  $\hat{V}^*(\bar{y}')$  are realizable for  $n_1 \geq 2$ . Also for  $n_1 \geq 2$ , the estimator  $\hat{\sigma}^2$  is nonnegative and unbiased for the population variance, in the case of independence between  $k \neq i \in s_1$ :  $\rho_{ki} = 0$ . An example of this situation is for  $q_i$  ( $i = 1, 2, \dots, N$ ) independent and  $\text{card}\{i \in U: q_i = 1\} \geq N - n + 2$ ; then,  $p_i$  ( $i \in U$ ) are independent and  $n_1 \geq 2$ .

### References

- Anderson, H. (1979). On nonresponse bias and response probabilities. *Scand. J. Statist.* **6**, 107-112.
- Cochran, W.G. (1977). *Sampling Techniques*. Third Edition. John Wiley. New York.
- Czajka, J.L., Hirabayashi, S.M., Little, R.J.A. and Rubin, D.B. (1992). Projecting from advance data using propensity modeling: An application to income and tax statistics. *J. Bus. Econ. Statist.* **10**, 117-131.
- Eltinge, J.L. (1992). Conditions for approximation of the bias and mean squared error of a sample mean under nonresponse. *Statist. Probab. Lett.* **15**, 267-276.
- Giommi, A. (1987). Nonparametric methods for estimating individual response probabilities. *Survey Methodology* **13**, 127-134.
- Hansen, M.H. and Hurwitz, M.N. (1946). The problem of the non-response in sample surveys. *J. Amer. Statist. Assoc.* **41**, 517-529.
- Little, R.J.A. (1986). Survey nonresponse adjustments for estimates of means. *Internat. Statist. Rev.* **54**, 139-157.
- Murthy, M.N. (1963). Generalised unbiased estimation in sampling from finite populations. *Sankhyā Ser. B* **25**, 245-262.
- Oh, H.L. and Scheuren, F.J. (1983). Weighting adjustment for unit nonresponse. *Incomplete Data in Sample Surveys 2* (W.G. Madow, I. Olkin and D.B. Rubin, eds.). Academic Press. New York, 143-184.
- Politz, A.N. and Simmons, W.R. (1949, 1950). An attempt to get the "not at homes" into the sample without callbacks. *J. Amer. Statist. Assoc.* **44**, 9-31; **45**, 136-137.
- Ruiz Espejo, M., Rueda García, M. and Arcos Cebrián, A. (1995). Finite population sampling and imputation of missing data. *Proceedings. Applied Stochastic Models and Data Analysis 7* (J. Janssen and S. McClean, eds.). University of Ulster Press. Coleraine, 551-558.

OPTIMALIDAD INSESGADA DE LA MEDIA MUESTRAL

M. Ruiz Espejo

Dept° de Economía Aplicada Cuantitativa.

Fac. de Ciencias Económicas y Empresariales.

Univ. a Distancia. c/ Senda del Rey. s/n. 28040-Madrid.

**Summary.** We provide an original proof of the unbiased optimality of the sample mean for the class of "all convex linear combination of the observations" estimators (with positive coefficients) for simple random sampling with replacement of size  $n$ , of a finite or infinite population.

1. Introducción y resultado

Consideremos la clase de estimadores

$$t = \frac{\sum_{i \in s} w_i y_i}{\sum_{i \in s} w_i} \text{ con } w_i > 0 \text{ (} i = 1, 2, \dots, n \text{),}$$

siendo  $s$  la muestra ordenada de tamaño  $n$ , de unidades de una población finita o no, seleccionada por muestreo aleatorio simple con reemplazamiento. Si a la clase de estimadores se le exige que sean insesgados para la media poblacional  $\bar{y}$ , entonces resulta que

$$\begin{aligned} E(t) &= E[E(t|s)] = E \left[ E \left( \frac{\sum_{i \in s} w_i y_i}{\sum_{i \in s} w_i} \middle| s \right) \right] = E \left[ \frac{\sum_{i \in s} w_i E(y_i|s)}{\sum_{i \in s} w_i} \right] = \\ &= E \left( \frac{\sum_{i \in s} w_i \bar{y}}{\sum_{i \in s} w_i} \right) = E(\bar{y}) = \bar{y}, \end{aligned}$$

es decir, todos los estimadores de la clase son insesgados para estimar  $\bar{y}$ . Luego la condición de insesgación no reduce en ningún estimador  $t$  a la clase introducida, o bien, la clase propuesta es de estimadores insesgados.

Si calculamos la varianza de un estimador  $t$  cualquiera de la clase, como  $E(t|s) = \bar{y}$ , entonces  $V[E(t|s)] = V(\bar{y}) = 0$ , y de aquí:

$$V(t) = E[V(t|s)] + V[E(t|s)] = E[V(t|s)] =$$

$$= E \left[ \frac{\sum_{i \in S} w_i^2 V(y_i | s)}{\left( \sum_{i \in S} w_i \right)^2} \right] = \sigma_y^2 E \left[ \frac{\sum_{i \in S} w_i^2}{\left( \sum_{i \in S} w_i \right)^2} \right], \quad (1)$$

pues

$$V(y_i | s) = \sigma_y^2$$

es la varianza poblacional.

Si tenemos en cuenta la desigualdad de Cauchy-Schwarz:

$$\left( \sum_{i=1}^n a_i b_i \right)^2 \leq \left( \sum_{i=1}^n a_i^2 \right) \left( \sum_{i=1}^n b_i^2 \right),$$

llamando  $a_i = w_i$  y  $b_i = 1$ , tendremos que

$$\left( \sum_{i \in S} w_i \right)^2 \leq \left( \sum_{i \in S} w_i^2 \right) n.$$

Entonces:

$$\frac{1}{n} \leq \frac{\sum_{i \in S} w_i^2}{\left( \sum_{i \in S} w_i \right)^2}, \quad \frac{1}{n} \leq E \left[ \frac{\sum_{i \in S} w_i^2}{\left( \sum_{i \in S} w_i \right)^2} \right],$$

tomando esperanzas matemáticas a ambos miembros, por lo que de aquí y de (1) deducimos que

$$V(\bar{y}_s) = \frac{\sigma_y^2}{n} \leq V(t),$$

para todo estimador  $t$  de la clase. Así, el estimador media muestral

$$\bar{y}_s = \frac{1}{n} \sum_{i \in S} y_i,$$

es un estimador de la clase, pues  $w_i = cte > 0$  (para  $i \in s$ ), es insesgada y además posee varianza mínima en la clase. En conclusión, la media muestral es el estimador insesgado óptimo de la clase introducida, bajo diseño de muestreo aleatorio simple con reemplazamiento de tamaño fijo  $n$ . Este resultado es válido indistintamente para poblaciones finitas o infinitas, y completa el resultado ya avanzado por Ruiz Espejo (1992). La demostración más conocida del resultado presentado se basa en minimizar la varianza de  $t$  sujeto a la restricción de su insesgación, mediante el método de los multiplicadores de Lagrange.

#### Referencia

Ruiz Espejo, M. (1992). Optimalidad de la media muestral. *Rev. Acad. Cienc. Zaragoza* 47, 125-128.

**LOS COMIENZOS DEL BIPEDISMO EN EL PROCESO DE  
HOMINIZACION**  
(The beginning of bipedism in the hominization process)

**Autor:** Alberto A. Makinistian

**Institución:** Departamento de Bioantropología y Evolución. Escuela de Antropología. Facultad de Humanidades y Artes. Universidad Nacional de Rosario. Entre Ríos 758. 2.000 Rosario (Pcia. de Santa Fe). República Argentina.

**Abstract.**

¿What events led to the adoption of a biped posture in the process of hominization? In fact, it is not exactly known, but it is probable that the modification of the geographic landscape of eastern Africa would have contributed to it.

Notwithstanding, the origins of bipedestation cannot be fully understood if his biped posture is not considered as a part of a general survival strategy, which was the result of an adaptative response to the new conditions of life.

It is also object of discussion if those individuals who lived more than 4 millions years ago were as efficient bipeds as modern human beings, or if on the contrary, they still preserved sufficient ability for tree climbing.

Finally, although we agree on the fact that bipedism could have played a significant role in the hominization process, we think that it not only this trait but the overall features of hominization what makes us different from other living creatures.

**NOTA:** El presente trabajo es el texto de la conferencia pronunciada por el autor en el Salón de Actos de la Caja de Ahorros de La Inmaculada, el día 6/3/97, respondiendo a una invitación de la Academia de Ciencias Exactas, Físicas, Químicas y Naturales de Zaragoza.

## 1.- Introducción.

Entre todos los mamíferos, la especie humana se caracteriza por disponer de una postura erecta y marcha bípeda permanentes. Si bien es cierto que ella nos ocasiona diversos inconvenientes, tales como hernias de disco, pies planos, prolapso uterino y várices (que no se presentan en los póngidos), la misma supone, entre otras ventajas, la de posibilitar que la traslación no dependa de todas las extremidades sino solo de las posteriores, por ese motivo más largas y fuertes y, consiguientemente, que los pies se conviertan en órganos de apoyo para sostener el peso del cuerpo, perdiendo así su capacidad para la prensión.

En este sentido, la postura erecta implica, entonces, un cambio significativo: ya no es la cría la que se aferra al cuerpo de la madre con manos y pies, sino que es la madre la que transporta a la cría sujetándola con sus miembros anteriores completamente libres de su función locomotriz (convirtiéndose así en órganos aptos para la aprehensión de los objetos, para cargarlos y transportarlos, para fabricarlos, para defensa o ataque, etc.)

¿Qué importancia tuvo la adopción del bipedismo en el proceso de hominización? Para algunos autores la adquisición de la postura erecta constituye un hecho de verdadera trascendencia. No solo por lo que significa en sí misma, sino, además, por los numerosos cambios correlativos que provoca.

En este sentido, C. Owen Lovejoy, de la Kent State University, de E.E.U.U., sostiene que: *En conjunto, la adopción del bipedismo es uno de los cambios anatómicos más sorprendentes que puedan verse en biología evolutiva* (1).

Del mismo modo, Stephen Jay Gould señala que el bipedismo es el paso crucial de la hominización y Richard Leakey, a su vez, afirma: *En mi opinión, la distinción fundamental entre nosotros y nuestros parientes más próximos no es el lenguaje, ni la cultura, ni la tecnología. Es el hecho de andar erguidos, el uso de nuestras extremidades inferiores para sostenernos y desplazarnos y liberar de esas funciones a nuestras extremidades superiores* (2).

Y más adelante: *para mí "humano" y "simio bípedo" son sinónimos* (3).

¿Qué fue lo que llevó a la adopción del bipedismo? No lo sabemos exactamente. Sin embargo, pensamos que las modificaciones del paisaje geográfico del este africano durante el Mioceno tuvieron mucho que ver con su aparición.

## **2.- Cambios en el paisaje geográfico de Africa oriental durante el Mioceno.**

A comienzos del Mioceno las regiones templadas africanas se caracterizaban por tener temperaturas más elevadas que en la actualidad. Como consecuencia, selvas y bosques cubrían particularmente el área comprendida entre los 10° de latitud norte y 10° de latitud sur. Pero desde mediados del Mioceno el clima fue cambiando lentamente, con una evidente tendencia al enfriamiento. Ello ocasionó la paulatina reducción de las vastas zonas selváticas y boscosas y el consiguiente incremento de estepas y sabanas arboladas.

El cambio es notable, porque mientras las selvas trópico-ecuatoriales se caracterizan por una temperatura elevada, poco variable, sin alternancia marcada de estaciones y alta pluviosidad, con una vegetación que persiste durante todo el año, en las sabanas las variaciones térmicas son más marcadas, existen escasas lluvias y clara alternancia de estaciones, con largos períodos de sequía. Las gramíneas y leguminosas, ricas en granos, constituyen el único alimento disponible, en estas zonas, solo durante parte del año.

La vida en regiones de densa vegetación brinda, a quienes allí viven, comida y protección. Refiriéndose al alimento disponible, Ashley Montagu señalaba que: *Un ambiente boscoso no presenta muchos inconvenientes, pues los animales que lo habitan encuentran allí abundantes alimentos que pueden comer directamente. Por así decirlo, la-mesa está puesta* (4).

En cambio, la vida en áreas más despejadas plantea nuevas y más duras exigencias.

Hacia fines del Mioceno, o comienzos del Plioceno, hace unos 8 millones de años, se produce, en el este africano, un acontecimiento geológico de

importancia: una falla o fractura de magnitud, conocida con el nombre de Valle del Rift, que se inicia en Asia Menor, mas precisamente en el sur de Turquía, descendiendo por Israel, cruza el Mar Rojo y penetra en Africa por Etiopía, recorriendo toda la zona oriental (los sitios de Hadar, Valle del Omo, Lago Turkana, Laetoli y Olduvai están dentro de la gran grieta) hasta llegar a la desembocadura del río Zambeze, en Mozambique, luego de recorrer unos 5.000 kms. y alcanzando en algunos lugares un ancho de 80 kms. y una profundidad superior a los 300 metros.

Ambos acontecimientos, el descenso de las temperaturas y la formación del Valle del Rift, mas la presencia de ríos, lagos y montañas, llevó a la transformación del este africano en una región rica en ecosistemas y barreras ecológicas convirtiéndose en un verdadero "mosaico de hábitats". Los cambios ambientales, como la mayoría de los investigadores considera, constituyen un importante estímulo del cambio evolutivo, como apunta Roger Lewin, dado la elevada presión selectiva existente. Así, es muy probable que ese "mosaico de hábitats" haya tenido particular relación con los comienzos del bipedismo.

Una respuesta eficaz a las exigencias del medio pudo haber sido, de este modo, la principal causa de que un grupo de primates iniciara el recorrido de un camino evolutivo diferente: el que conduciría a nosotros mismos. Gorilas y chimpancés, en cambio, quedaron reclusos, en su mayor parte, a las selvas y bosques densos.

### **3.- La documentación paleoantropológica**

¿Qué información proporcionan los hallazgos paleoantropológicos? Hasta 1972 no se consideraba una antigüedad superior a los 2,8 millones de años para la postura bípeda. Pero en 1973, durante la campaña realizada por Donald Johanson, Maurice Taieb e Yves Coppens a la zona de Hadar, al noreste de Etiopía, el primero de los autores mencionados descubrió, semienterrado, un fragmento superior de tibia y a unos metros un fragmento inferior de fémur que estaba partido por la mitad, de manera que solo conservaba uno de sus cóndilos (el otro cóndilo estaba tirado en la arena, muy cerca).

Cuando Johanson articuló las tres partes observó sorprendido que el fémur y la tibia se habían unido formando ángulo, disposición diferente a la

de los póngidos, en quienes se unen en línea recta. Dicho ángulo, denominado ángulo valgus, existe en el hombre moderno y es decisivo para la locomoción bípeda ya que el pie se sitúa por debajo del centro de gravedad mientras se avanza. En cambio, los póngidos "anadean" cuando se trasladan en posición bípeda debido a que su fémur cae en forma totalmente vertical. Esta comprobación permitía suponer que se trataba de un homínido de locomoción erguida, cuya antigüedad se calculaba entre 3 y 4 millones de años.

Con posterioridad, en 1974 y 1975, los esqueletos de "Lucy" y los de la "Primera Familia" corroboraron la existencia de la bipedestación en homínidos del noreste de Etiopía, hace más de tres millones de años.

Años después, en 1978, también se encontraron pisadas fósiles de homínidos en el sitio de Laetoli, a unos 40 kms. al sur de la Garganta de Olduvai, en Tanzania. Dichas pisadas, datadas en 3,7 millones de años, revelan el desplazamiento bípedo de tres homínidos, a juzgar por la fuerte pisada del talón, bóveda plantar bien marcada y dedo pulgar alineado junto a los otros dedos (no separado, como se observa en los póngidos). Se deduce claramente, por lo tanto, que el pie se había convertido en órgano de apoyo para sustentar el peso del cuerpo.

La mayoría de los autores ha aceptado la propuesta de Johanson y Tim White en cuanto a que las poblaciones de Hadar y Laetoli representan el mismo tipo de homínido, al que denominaron *Australopithecus afarensis*.

Restos de *afarensis*, con una antigüedad superior a los 4 millones de años, también fueron encontrados, en 1981, en Middle Awash, Etiopía, por una expedición conjunta de la Universidad de California y la de Addis Abeba.

A mediados de 1995, cuatro huesos del pie encontrados en las cuevas de Sterkfontein, cerca de Johannesburgo, en Sudáfrica, datados en 3,5 millones de años, demuestran postura bípeda.

Por entonces, la paleontóloga Meave Epps Leakey descubrió restos fósiles de una antigüedad de entre 3,9 y 4,1 millones de años, a los que denominó *Australopithecus anamensis*, en la zona del Lago Turkana. Los huesos tibiales del mismo presentan el típico borde delantero propio de individuos que se desplazan en forma bípeda.

En definitiva, la documentación paleoantropológica existente a la fecha asegura la presencia de la postura bípeda en homínidos que vivieron hace no menos de 4 millones de años.

#### 4.- Los australopitecos afarensis: ¿eran bípedos eficientes?

Los científicos no coinciden en sus apreciaciones comparativas entre la postura erecta y marcha bípeda de los afarensis y la nuestra. Algunos consideran que los mencionados homínidos, aunque bípedos, no eran demasiado eficientes en tierra y que alternaban con una vida parcialmente arborícola.

En este sentido, Peter Schmid, quien intentó una reconstrucción del esqueleto de Lucy, observó que la caja torácica no presentaba forma de tonel, como en el hombre moderno, sino forma de embudo, como en los póngidos. Además, y según esta reconstrucción, los huesos ilíacos de la pelvis tendrían un aspecto más simiesco que humano. Del mismo modo, los trabajos de Christine Tardieu demostrarían una locomoción no definitivamente terrestre en los australopitécidos, sino más indiferenciada, conservando cierta aptitud para la trepa arborícola.

Dos investigadores de la Universidad estatal de Nueva York, en Stony Brook, Randall L. Susman y Jack T. Stern, llegaron a la conclusión de que la mezcla de caracteres de la pelvis de Lucy es indicativa de una forma un tanto simiesca de bipedismo. Asimismo, interpretan que los huesos claramente curvados de los dedos de las manos y de los pies son adaptaciones a la aprehensión, en una vida en gran medida arborícola. Lo mismo indicaría el omóplato, en su articulación con el húmero y la articulación del codo es compatible con la suspensión arborícola.

Los estudios sobre el esqueleto de Lucy y de otros ejemplares de Hadar, parecen mostrar que afarensis tenía los miembros anteriores largos y los posteriores relativamente cortos; sin embargo, Milford Wolpoff, de la Universidad de Michigan, sostiene que las cortas piernas de Lucy tienen la longitud que cabría esperar en un ser humano de su diminuta estatura.

Según Lovejoy, los afarensis *parecen increíblemente primitivos del cuello para arriba e increíblemente modernos del cuello para abajo* (5), lo

que probaría que el fenómeno de adaptación fundamental en estos homínidos antiguos habría sido la marcha bípeda y no el desarrollo del cerebro.

Al respecto, considera que la pelvis de Lucy fue diseñada especialmente para la bipedia (6) y, en ciertos aspectos, ofrece un diseño mejor para la bipedestación que el aportado por la especie humana (7) debido a que el tamaño del cráneo todavía era reducido y no requería de un ensanchamiento de la abertura pélvica que pudiese afectar, de algún modo, la bipedestación. Lucy se habría anticipado a la drástica expansión del tamaño del cerebro. Por ello, sus partos debieron ser mucho más fáciles que entre los seres humanos (el cráneo de sus fetos no sería mayor que el del feto de un chimpancé).

Si bien Lovejoy admite la posibilidad de que Lucy no abandonara totalmente la trepa arborícola, señala que existen claros indicios de que la estructura interna del cuello del fémur de los afarensis es de tipo humano ya que presenta una delgada capa ósea en la parte superior, lo que permite soportar únicamente los esfuerzos derivados de la marcha o la carrera en posición erguida. De esta observación, Lovejoy concluye que en Lucy el cuello del fémur estaba diseñado exclusivamente para la locomoción bípeda. No se trata de que pudiera caminar erguida de vez en cuando: esa postura se había convertido en su única alternativa (8).

## 5.- Probables causas

¿Por qué nuestros antepasados adoptaron la postura bípeda? Tampoco lo sabemos con exactitud, pero es evidente que las causas que la originaron debieron ser lo suficientemente importantes como para justificar un cambio tan profundo. Por qué sucedió sigue siendo materia de conjeturas. No se trata de enumerar las ventajas que proporcionó, una vez establecida, sino qué fue lo que llevó a su aparición.

Entre las probables causas, suelen citarse las siguientes:

### 5.1. Fabricación de instrumentos.

La postura bípeda supone la liberación de los miembros anteriores y la consiguiente posibilidad de fabricación de herramientas. En 1974, Sherwood

Washburn escribió: *La locomoción y el uso de herramientas fueron simultáneamente causa y efecto uno del otro* (9).

Sin embargo, a pesar de intensos trabajos de rastreo en Laetoli, y en otros sitios de similar antigüedad, no fueron encontrados instrumentos líticos. Los más antiguos, hasta ahora, fueron encontrados hace poco tiempo en el Valle del Gona, a unos 20 kms. de Hadar, en Etiopía, y cuentan con 2,6 millones de años, de acuerdo con la aplicación del paleomagnetismo. Ello significaría la existencia de un período de más de un millón de años privado de instrumentos líticos. ¿Cómo se explica esta situación? Las posibles respuestas a este interrogante son:

- a) Los instrumentos existen, pero aún no han sido descubiertos.
- b) Los primeros instrumentos no serían de piedra sino de madera que, como sabemos, no se conserva a través de tanto tiempo (sin embargo, si tendrían que encontrarse los percutores de piedra utilizados.
- c) Estos homínidos no habían alcanzado todavía capacidad suficiente para la fabricación de instrumentos.

En lo que respecta a esta última alternativa, debemos recordar que sus cerebros no superaban los 500 c.c. y que no se aprecia un aumento significativo de la capacidad cerebral hasta la aparición de *Homo habilis*.

Ateniéndonos a la documentación paleoantropológica existente a la fecha, debemos considerar que la adopción de la postura erecta y marcha bípeda permanente sería claramente anterior al desarrollo del cerebro y no tendría relación directa con la fabricación de instrumentos como, entre otros autores, sostenía Washburn, sino que el motivo sería otro.

## 5.2. Vigilancia.

La vigilancia es otro aspecto del comportamiento antidepredador que tiende a favorecer una postura erguida.

Mientras el bosque y la selva ofrecen protección a los animales que allí viven, las zonas de sabana, abiertas, despejadas, dejan prácticamente desprotegidos del ataque de carnívoros a quienes dependen de ellas para su alimentación y, por lo tanto, para su subsistencia.

Por ello, una de las causas considerada como probable por algunos autores estaría relacionada con la alimentación de los homínidos en zonas abiertas. En esas circunstancias, una eficaz vigilancia practicada por algunos individuos en postura erecta (permitiendo dominar más y mejor el paisaje circundante), brindaría cierta seguridad al grupo ante el permanente riesgo de ataque por parte de carnívoros.

Este comportamiento, muy común entre los animales que habitan zonas abiertas (su supervivencia depende de la unión y de la cooperación), puede observarse, por ejemplo, en los papiones, que se adentran en la sabana y viven en el suelo. En un grupo de sesenta, habrá siempre 10 ó 12 individuos de pie, sobre sus patas traseras, olvidando momentáneamente la búsqueda de comida para desempeñar mejor su papel de vigías.

También el pequeño suricata, un tipo de mangosta que habita Africa meridional, en la zona del Kalahari, adopta la postura bípeda (con la cola como órgano de apoyo o bastón) en la tarea de vigilancia, en la que se turnan permanentemente.

### 5.3. Atemorizar a depredadores y congéneres.

En el caso de que el ataque de un carnívoro llegara a concretarse, la huida puede resultar absolutamente inútil por la distancia que separa a la víctima de algún árbol o sitio de protección. Por tanto, al no quedar otra opción que enfrentar la situación, la presa puede ensayar alguna conducta que quizá impacte o impresione al atacante: en suma, que sea más efectista que efectiva, pero que logre el propósito buscado: ahuyentar al enemigo.

Entonces, no es extraño que se adopte la postura erecta con el fin de atemorizar a los depredadores. Al erguirse sobre sus extremidades posteriores, un cuadrúpedo adquiere una dimensión distinta, desconocida para el atacante, y ello puede hacer que éste decida rehusar la confrontación.

Para amedrentar a los depredadores, y también a sus propios congéneres, los chimpancés recurren, frecuentemente, a las posturas bípedas acompañadas del manejo de grandes ramas. También arrojando piedras. Aunque estos comportamientos difícilmente constituyen una grave amenaza para el depredador que se aproxima, el mero hecho de adoptarlos produce cierto efecto.

En los últimos años, Nina Jablonski y George Chaplin, de la Universidad de Australia occidental, han considerado que el bipedismo es el resultado de la sistematización de un comportamiento común en los grandes monos: la parada de intimidación, que consiste en ponerse de pie para impresionar a un adversario.

#### 5.4. Transporte de alimentos

##### a) *hipótesis de Adrienne Zihlman.*

Como señala Roger Lewin, esta hipótesis, propuesta inicialmente por Adrienne Zihlman, de la Universidad de California en Santa Cruz, a principios de la década del setenta, fue un reto ideológico planteado frente al modelo del "hombre cazador" orientado al macho (10).

A través de la misma, se pone el acento en el hecho de que, como se sabe, en los grupos humanos de economía recolectora, la alimentación cotidiana, consistente en un 70 a 80 % de vegetales, es provista por las mujeres, mientras que el resto, carne, resulta aportado por los varones (despiece y traslado de animales grandes ya muertos y caza de animales pequeños). Entre los ¡Kung, por ejemplo, las nueces constituyen el alimento básico ya que cada individuo consume aproximadamente 300 nueces de mongongo por día (lo que equivale a un bistec de unos 400 gramos).

En este modelo, la bipedestación facilitaría a las mujeres la recolección de vegetales y frutas al mismo tiempo que el transporte de las crías. La antropóloga norteamericana Nancy Tanner sostiene similar opinión.

##### b) *hipótesis de Owen Lovejoy.*

A comienzos de los ochenta, C. Owen Lovejoy sostuvo que la locomoción no podía ser entendida sino como parte de una estrategia general de supervivencia.

En el modelo propuesto, no existe una primera causa desencadenante de las demás, sino que se trata de un circuito de realimentación donde cada factor ejerció efecto sobre los demás, produciéndose reforzamientos mutuos (fenómenos sinérgicos).

Lovejoy cree que el bipedismo habría permitido a los primeros homínidos procrear más, evitando el espaciamiento en los nacimientos típico de los póngidos. En este sentido, un aumento de la tasa de natalidad pudo haber sido una respuesta adaptativa a la vida en zonas abiertas.

Por otra parte, un cambio muy importante se habría producido cuando la atracción entre los sexos no se limitó al período de celo o estro de las hembras, ni estuvo restringida al momento de la ovulación, sino que fue continua; y esto constituyó la base de la pareja humana.

Pero también existen razones alimentarias que justificarían esta relación permanente. Según Lovejoy, si la hembra consigue una menor movilidad, podrá tener hijos con mayor frecuencia. Pero si el espaciamiento entre los nacimientos es menor, ello exige una mayor dedicación de la hembra al cuidado de las crías al mismo tiempo que aumentan las necesidades alimentarias (hay más bocas para alimentar). Como la menor movilidad de la hembra le resta posibilidades de obtener comida, esta tarea habría quedado asignada al macho, cuya postura bípeda le habría permitido acarrear alimentos al campamento.

Así, el macho habría provisionado a la hembra de alimentos de elevado nivel energético, aumentando la capacidad de la madre para nutrir y proteger a las crías y, además, le permitiría engendrar con mayor frecuencia. Pero, según Lovejoy, el macho no abastecería a la hembra a menos que esté seguro de que los hijos son suyos (aquí el autor sostiene ideas socio-biológicas).

Obviamente, la bipedia resultaría, en este caso, una adquisición sumamente ventajosa por la liberación de los miembros anteriores y la consiguiente posibilidad de recolectar y transportar el alimento a una morada base.

Lejos de compartir la postura de Lovejoy, Rebecca Cann considera que parece más bien una extrapolación al pasado del modelo occidental actual de la familia nuclear. Lo cierto es que la hipótesis de Lovejoy, que tuvo difusión y repercusión en la década del ochenta, fue luego muy criticada y actualmente ha perdido consenso.

### 5.5. Adaptación a una nueva distribución de los alimentos.

En los últimos años se ha destacado la propuesta de Peter Rodman y Henry Mc Henry, de la Universidad de California, en Davis, quienes consideran que el bipedismo no es producto de un cambio en la dieta alimenticia sino de un cambio en la distribución del alimento.

Como ya hemos visto, hacia fines del Mioceno muchas regiones de densa vegetación del este africano se fueron transformando en zonas abiertas, lo que llevó a una dispersión cada vez mayor de los alimentos, que, por este motivo, requerían de desplazamientos más largos para su recolección. Esta necesidad cotidiana de un desplazamiento mayor para poder subsistir pudo haber ejercido una significativa presión de selección para una locomoción más eficaz.

En tales circunstancias, la marcha bípeda habría resultado energéticamente más eficaz que el caminar sobre nudillos del chimpancé. De esta manera, también las manos, libres, habrían ido adquiriendo gran habilidad para la prensión de los pequeños bocados (granos y semillas, por ejemplo).

Según Richard Leakey, la hipótesis de Rodman y Mc Henry *es una de las más persuasivas con que contamos* (11).

### **6.- La abertura pélvica, el cerebro y la bipedestación: un delicado equilibrio.**

Una de las consecuencias de la bipedestación en las mujeres ha sido la reducción de su canal de parto. Contrariamente, el paulatino incremento del cerebro a lo largo del proceso de hominización condujo a un ensanchamiento cada vez mayor de la abertura pélvica para dar paso a un feto con mayor capacidad cerebral. Pero este ensanchamiento tuvo un límite: el de no llegar a afectar el normal desplazamiento bípedo de las mujeres.

Es probable que todo ello llevara a un acortamiento del período de gestación humana ya que el cerebro del feto crece aceleradamente en los últimos tramos de su vida intrauterina y, de continuar, no podría pasar luego por el canal de parto (esto explicaría el hecho de que, inmediatamente después de nacer, y hasta el año de vida, el cerebro del niño se incrementa en forma explosiva hasta alcanzar nada menos que el 60 % de su cerebro

adulto). Explicaría también el grado de inmadurez del recién nacido y, por tanto, su necesidad de recibir un mayor cuidado y protección.

La situación crítica en que se encuentra la pelvis de la mujer, por estar más cerca de los límites de su capacidad estructural y funcional, constituiría el motivo, según Lovejoy, de que su parto sea uno de los más traumáticos del reino animal y que, probablemente, a eso se deba la mayor frecuencia de las fracturas de la articulación de la cadera en las mujeres, respecto a los varones.

Afortunadamente, y tal como se observa, se ha llegado en las mujeres a una delicada armonía: la máxima abertura pélvica posible, capaz de dar paso a un voluminoso cerebro, y una normal bipedestación.

Dos consideraciones finales:

Por un lado, nos parece importante destacar que la bipedia no es un rasgo puramente biológico, sino un carácter bio-socio-cultural. No es innato, ya que los niños la imitan o aprenden en el seno de una sociedad. Sin dudas, nuestra conformación física la posibilita, y muy bien, pero todavía es preciso que se nos enseñe a ponernos de pie, a desplazarnos sobre nuestras dos piernas y no a cuatro patas. Cuando el niño tiene alrededor de un año intenta enderezarse, mantenerse en pie aferrado a algún soporte, pero cuando desea trasladarse lo hace en forma cuadrúpeda.

Dice Pierre Grassé al respecto: *Al no ser la posición bípeda innata en el hombre actual, es lógico deducir que tampoco lo era en los australopitecos y pitecántropos, quienes también tendrían que enseñarla a sus pequeños. Sólo pudo desarrollarse en el seno de una sociedad, o como mucho en el seno de familias que conservasen a sus crías largo tiempo entre ellas* (12).

La mejor prueba de que la postura bípeda no es un carácter exclusivamente biológico la tenemos en el hecho de que, en todos los casos en que niños fueron abandonados en zonas inhóspitas al poco tiempo de nacer, cuando se los encontró se desplazaban en forma cuadrúpeda, y, al parecer, este hecho debe ser relacionado directamente con la carencia del factor socio-cultural (es el caso de los llamados "niños lobos").

La segunda consideración tiene que ver con la importancia de la postura bípeda en el proceso de hominización.

Aunque para algunos autores, como Richard Leakey (según vimos al inicio de este texto) la bipedestación desempeñó un papel único en el proceso de hominización, y si bien estamos de acuerdo en que, efectivamente, el bipedismo pudo haber tenido un rol decisivo, creemos que la postura bípeda no constituyó el carácter exclusivo que nos diferencia de los demás seres vivos. Tampoco los otros caracteres de la hominización, considerados separadamente (tales como el desarrollo del cerebro, la fabricación de instrumentos o la posesión de un lenguaje articulado, por ejemplo).

En este sentido, concluíamos en un trabajo anterior: *Lo que nos diferencia es la posesión de todos, en conjunto. Estamos convencidos de que allí reside, precisamente, la originalidad de la línea filogenética que condujo a Homo sapiens* (13)

#### Citas bibliográficas

- (1). citado por Richard Leakey en **La formación de la humanidad**. Ediciones del Serbal. Barcelona, 1981. Pág. 52.
- (2) y (3). **Nuestros orígenes**. Ediciones Crítica. Barcelona, 1994. Pág. 81.
- (4). **La revolución del hombre**. Ed. Paidós. Biblioteca del hombre contemporáneo. Volumen 185. Buenos Aires, 1967. Pág. 21.
- (5). cita de Richard Leakey en **La formación de la humanidad**, pág. 69.
- (6) y (7). "Evolución de la marcha humana". En Revista Investigación y Ciencia (edición en español de Scientific American). Prensa Científica S.A. Barcelona. Nº148, del mes de enero de 1989, pág. 77.
- (8). Ibidem, pág. 80.
- (9). citado por Donald C. Johanson y Maitland Edey en **El primer antepasado del hombre**. Ed. Planeta. Colección Al filo del tiempo. Barcelona, 1982. Pág. 290.

(10). **Evolución humana.** Ed. Salvat. Biblioteca Científica Salvat N° 23. Barcelona, 1993. Pág. 181.

(11). **Nuestros orígenes,** pág. 88.

(12). **El Hombre, ese dios en miniatura.** H. Blume Ediciones. Madrid, 1977. Pág. III.

(13). "El ser del hombre desde la perspectiva paleoantropológica". En **Naturaleza y (o) Lenguaje.** Ed. Universidad Nacional de Rosario. Rosario, Argentina, 1996. Pág. 47.

## **Estudios en *Chlorella fusca* como modelo para el desarrollo de un método de valoración de deficiencia de hierro en ecosistemas acuáticos (\*).**

Luis A. Inda y M. Luisa Peleato.

Departamento de Bioquímica y Biología Molecular.

Facultad de Ciencias. Universidad de Zaragoza.

(\*) Premio de la Academia a la investigación (1996-1997)

### **Resumen**

Con objeto de conocer la respuesta a la deficiencia de hierro en el alga verde *Chlorella fusca*, se han purificado diversas proteínas que pueden ver alterada su síntesis en esas condiciones: ferredoxina, flavodoxina y citocromo *c<sub>6</sub>*. La obtención de anticuerpos policlonales específicos, y la valoración de la flavodoxina y la ferredoxina presente en las algas mediante inmuno-electrodifusión puede permitir una medida del estado nutricional de hierro del fitoplancton.

### **Abstract**

*Chlorella fusca* response to iron deficiency has been studied in order to develop methods for evaluating iron stress in phytoplankton, based in quantification of flavodoxin and ferredoxin present in the cells. The proteins have been purified and their correspondent antibodies raised. Preliminary studies in modelization of the response of *Chlorella* cells to iron stress have been performed.

## Introducción.

Durante muchos años, la comunidad científica ha prestado mucha atención al problema de las repercusiones de la deficiencia de hierro en los organismos fotosintéticos (los productores primarios), tanto en medios acuáticos como en terrestres. Una de las facetas más importantes de este problema se plantea en los océanos, y ha sido motivo de larga polémica el papel del hierro en el control de la productividad primaria de los océanos. Recientemente, se ha demostrado por primera vez de una forma concluyente, que los niveles de hierro controlan la productividad, mediante un experimento de fertilización "in situ" llamado IronEx II. Tras fertilizar con hierro aguas superficiales de una región ecuatorial del Océano Pacífico, se detectó un crecimiento masivo del fitoplancton, demostrando inequívocamente que la productividad estaba limitada por la disponibilidad de hierro (Coale et al. 1996). Las repercusiones de este hecho son innumerables, no sólo desde el punto de vista científico, sino también económico y social; por ejemplo, las llamadas mareas tóxicas que frecuentemente sufren muchas zonas costeras, podrían estar causadas por un aporte de hierro que permitiría la proliferación de los organismos responsables de ellas. Otro aspecto importante es la captación de CO<sub>2</sub> implícito con la mayor o menor cantidad de fitoplancton, que podría incluso influenciar el clima global de la tierra (Cooper et al, 1996).

Los oceanógrafos llevan mucho tiempo interesados en el desarrollo de un método que les permita valorar el estado nutricional respecto al hierro que presentan los organismos planctónicos, y este interés es también importante en aguas dulces, en el que pueden producirse proliferaciones indeseadas de algas y cianobacterias que pueden afectar a la salud humana (por ejemplo, la desmesurada proliferación de *Aphanizomenon ovalisporum* en el lago Kinneret, que es la mayor fuente de agua potable en Israel; este organismo produce la hepatotoxina cilindroespermopsina). La valoración directa del hierro puede no ser una medida adecuada, pues su disponibilidad para ser utilizado va a depender del organismo en cuestión y de las condiciones ambientales (pH, oxígeno disuelto, etc.) (Huntsman and Sunda, 1980), por lo que las medidas directas del hierro no van a ser buenos indicadores del estado nutricional en hierro de un determinado grupo de organismos. En los últimos años se ha buscado un indicador o marcador fisiológico que diera información acerca de esta cuestión. Lógicamente estos marcadores o indicadores de estrés de hierro han estado

relacionados con la valoración de respuestas adaptativas a la deficiencia de hierro, o diversos cambios observados como consecuencia de esta limitación. Inicialmente, se intentaron valorar cambios en gran cantidad de enzimas que contenían hierro (catalasa, lipooxigenasa, peroxidasa), directamente correlacionados con los niveles de hierro suministrados, así como razones entre citocromo *f* y clorofila *a* (Glober, 1977), o transportadores electrónicos como la ferredoxina (Entsch et al, 1983). Pero estos indicadores pueden estar influenciados también por otros factores, como toxicidad o limitación causada por otros elementos (Bouma, 1983). Evidentemente parecía más interesante el uso de un método de diagnóstico "positivo", que se desencadenara al manifestarse la limitación de hierro, que un método "negativo" es decir que se afectaba en mayor o menor grado en función de otros factores ambientales y fisiológicos. Entre los "marcadores" de deficiencia que se desencadenan como consecuencia de limitación podemos distinguir dos grupos: 1) en primer lugar, los polipéptidos relacionados con el transporte del hierro, y 2) proteínas que sustituyen en su función a ferroproteínas afectadas por la deficiencia. En este último grupo encontramos a la candidata por excelencia, la flavodoxina.

La flavodoxina es una pequeña flavoproteína (14.5-24 kDa), que contiene un mol de FMN (flavin mononucleotido) por molécula. Esta proteína está presente en algunos procariotas, tanto fotosintéticos como heterótrofos, y en algunas algas. No se ha descrito en plantas superiores. En algunos casos, esta proteína es constitutiva y en la mayoría de los organismos fotosintéticos se induce como respuesta a deficiencia de hierro en el medio. Su misión es sustituir a la ferredoxina, proteína sulfoférrica, en el transporte de electrones fotosintético asociado al fotosistema I, que dará lugar a la fotorreducción del NADP<sup>+</sup>. Puede sustituir también a la ferredoxina en otros procesos en los que esta última interviene, como el transporte de electrones a la nitrito reductasa y ferredoxina-glutamato sintasa (Vigara et al, 1997). Cuando las células son cultivadas en medios deficientes en hierro, se produce una disminución de la cantidad de ferredoxina (Pardo et al., 1990). En células capaces de inducir flavodoxina, los estudios en el laboratorio mostraron que la inducción de flavodoxina ocurría en tiempos muy cortos después de la exposición de las células a la deficiencia, y que las cantidades detectadas eran correlativas con la cantidad de hierro presente en el medio de cultivo, y con la disminución de la ferredoxina (Hutbert et al., 1977; Sandmann et al., 1990). El hecho de que además es una molécula extremadamente estable, y relativamente fácilmente

cuantificable llevaron a Entsch et al. (1983) a proponer el uso de la flavodoxina como indicador de estres nutricional de hierro en medios naturales.

Los métodos para cuantificar la flavodoxina son variados y sensibles. Mediante el uso de métodos inmunoquímicos, con anticuerpos específicos, se pueden valorar cantidades muy pequeñas de esta proteína en extractos crudos procedentes de muy pequeña cantidad de células. Las técnicas de cromatografía líquida de alta resolución (HPLC y FPLC) permiten purificar y valorar de una forma muy precisa el contenido en flavodoxina también de relativamente pequeñas cantidades de células. Mientras que en el laboratorio no se presentan grandes problemas en este sentido, en medios naturales la situación es mucho más compleja. El problema planteado es valorar el estatus de nutrición en hierro de una población muy heterogénea, con distintas respuestas a la limitación de hierro y la síntesis de flavodoxina, de una manera rápida y fiable. No solamente la respuesta en cuanto a la síntesis de flavodoxina va a poder ser diferente, sino que las propias moléculas de flavodoxina pueden ser especies moleculares bastante distintas en cuanto a características físico-químicas, debido a las distancias filogenéticas entre los organismos que integran las poblaciones.

En este trabajo se plantean estudios básicos que podrían llevar al desarrollo de un método de valoración de deficiencia biológica de hierro en aguas dulces, extrapolable a océanos utilizando otros organismos.

## **Material y Métodos**

### Cultivo de *Chlorella fusca*:

Las células se cultivaron a 28°C con el medio de cultivo descrito por Kessler y Czygan (1970). Las células, se iluminaron con un panel de luces fluorescentes ( $100 \mu\text{mol m}^{-2} \text{s}^{-1}$ ) y se burbujearon con aire-CO<sub>2</sub>. Los cultivos deficientes en hierro contenían 0.18  $\mu\text{M}$  de hierro, mientras los cultivos sin deficiencia eran 18  $\mu\text{M}$  en hierro.

### Purificación de proteínas:

La flavodoxina se purificó según el método previamente descrito por Peleato et al. (1994), utilizando células cultivadas en medios deficientes. La

ferredoxina se purificó utilizando células cultivadas con medio completo, utilizando un método basado en el descrito por Bes et al.(1995). El citocromo *c*<sub>6</sub> se purificó según el método previamente descrito (Inda et al., 1997), y la plastocianina por un método basado en el descrito por Katoh (1960). La concentración de las proteínas purificadas se determinó espectroscópicamente utilizando los siguientes coeficientes de extinción, determinados en este trabajo : flavodoxina 10 mM<sup>-1</sup> cm<sup>-1</sup> (462 nm), ferredoxina 10.4 mM<sup>-1</sup> cm<sup>-1</sup> (422 nm), citocromo *c*<sub>6</sub> 24.8 mM<sup>-1</sup> cm<sup>-1</sup> (553 nm) y plastocianina 3.7mM<sup>-1</sup> cm<sup>-1</sup> (596 nm).

#### Obtención de anticuerpos:

Los anticuerpos se prepararon utilizando proteínas puras, en conejos New Zealand White . Se utilizó una primera inmunización utilizando adyuvante completo de Freund e igual volumen de la solución de proteínas puras. (0.5 mg, excepto en el caso de la ferredoxina que fue necesario utilizar 0.1 mg). Al cabo de un mes, se llevó a cabo una segunda inmunización con adyuvante incompleto de Freund e igual volumen y concentración de proteínas. En el caso del citocromo *c*<sub>6</sub>, el anticuerpo se encargó a ZEU-Inmunotec.

#### Métodos inmunológicos:

La cuantificación de las proteínas se llevó a cabo utilizando inmuno-electrodifusión (Laurell, 1966).

#### Métodos analíticos:

Las proteínas totales fueron determinadas por el método de Lowry (1951). El contenido en clorofila *a* se determinó espectrofotométricamente según el procedimiento descrito por Mackinney (1941), utilizando acetona al 80% y la ecuación  $12.7(A_{663\text{nm}}) - 2.58(A_{645\text{nm}}) = \mu\text{g cl } a / \text{ml}$

### **Resultados**

El trabajo realizado constituye la base del desarrollo de un método inmunoenzimático para la valoración de la flavodoxina y la ferredoxina en el

alga verde *Chlorella fusca*. Se han purificado las proteínas flavodoxina y ferredoxina del alga verde *Chlorella fusca*, así como otra proteína que contiene hierro, el citocromo *c6*, y la plastocianina, proteína que contiene cobre. Estas proteínas puras se han utilizado para la obtención de anticuerpos policlonales, que permiten la cuantificación de las proteínas en extractos crudos, y en concentraciones extremadamente bajas. Los cuatro anticuerpos obtenidos son monoespecíficos, y se dispone de ellos en cantidad suficiente para permitir una purificación posterior.

La deficiencia de hierro de las condiciones del cultivo utilizadas afecta sólo ligeramente al crecimiento celular. El tiempo de duplicación de las células deficientes (0.18  $\mu\text{M}$ ) resulta ser de 12 horas, mientras que cuando las células crecen en medio completo (18  $\mu\text{M}$ ) es de 10 horas. Sin embargo, la composición en ferredoxina/flavodoxina en las distintas condiciones de cultivo resultó ser muy diferente. Cada 100 g de peso humedo de algas cultivadas en medio completo (18  $\mu\text{M}$  de hierro) se obtuvieron 11.5 mg de ferredoxina pura y no se detectó la presencia de flavodoxina. Cuando las células se cultivaron en medio deficiente (0.18  $\mu\text{M}$  de hierro), la cantidad de flavodoxina purificada de 100 g de algas fue de unos 20 mg, detectándose la presencia de unos 0.4 mg de ferredoxina. Estos datos, en que la sustitución de la ferredoxina por la flavodoxina ha ocurrido, y sin embargo no está demasiado afectado el crecimiento celular, están de acuerdo con los datos obtenidos "in vitro", en que se observa que la flavodoxina sustituye eficazmente a la ferredoxina como transportador de electrones en la fotosíntesis (Peleato et al, 1994) y en la asimilación de nitrógeno (Vigara et al. 1997). Probablemente, otros procesos dependientes de hierro, como la propia síntesis de clorofila son los que se ven más afectados (Figura 1).

El contenido en proteína total de las células cultivadas en las dos situaciones de disponibilidad de hierro, es muy similar (Figura 2), y esto era previsible dado que en nuestros cultivos la deficiencia no estaba afectando fuertemente el crecimiento. El hecho de que las proteínas totales no estuvieran condicionadas por la deficiencia, permite referir el resto de los datos obtenidos al contenido en proteínas.

En la Figura 3 se expresan los resultados de un experimento en que se provocó la inducción de la flavodoxina. Se cultivaron células en medio com-

pleto, y en su fase exponencial, se separaron en dos alícuotas. Estas alícuotas fueron lavadas con medios de cultivo esteriles, uno completo y otro deficiente con respecto al hierro y se transfirieron a medios respectivamente completo y deficiente. Se tomaron muestras de las células cada cierto tiempo y utilizando inmunolectrodifusión, se cuantificó la flavodoxina presente en los extractos de la células recolectadas. La flavodoxina se induce al tercer día de cultivo deficiente, aunque este factor podría estar condicionado por la reserva de hierro que pudieran tener las células en cuestión. A partir del séptimo día, los niveles de flavodoxina permanecen relativamente invariables, y no se aprecian cambios significativos. Las células utilizadas como control, en medio completo, no presentaron flavodoxina.

Se purificó también otra ferroproteína, el citocromo *c6*, con objeto de estudiar las variaciones observadas en los cultivos deficientes y obtener anticuerpos que nos permitieran valorarlo en extractos crudos. En este caso los rendimientos de la purificación han sido menores, 4 mg. por 100 g. de peso humedo de *Chlorella*, cultivada en medio completo, y 1.1 mg. cuando se utilizaron células deficientes.

## Discusión

El genero *Chlorella* está ampliamente distribuido en las aguas dulces, por lo que podría ser uno de los organismos de elección para ser utilizado como indicador de estrés biológico de hierro en limnología. Una de las ventajas determinadas para el uso de *Chlorella* sería el hecho de que la flavodoxina no es constitutiva (Peleato et al. 1994), con lo cual en situación de no deficiencia no va a existir flavodoxina que pueda llevar a interpretaciones equívocas.

El trabajo realizado al purificar las proteínas y obtener sus anticuerpos específicos es el primer paso para la puesta a punto del método que nos permitirá valorar en extractos crudos, y en cantidades extremadamente bajas, los niveles de estas proteínas, y estudiar la correlación de estos valores y del estrés biológico de hierro. Se pretende estudiar la pareja ferredoxina/flavodoxina, y citocromo *c6*/ plastocianina, aunque este último par es menos prometedor, ya que sus cantidades relativas no solo van a estar influenciadas por el hierro, sino también por el cobre.

En el laboratorio los niveles de flavodoxina medidos presentan una buena correlación con los valores de hierro presentes en el medio de cultivo, y su rápida inducción asegura una respuesta adecuada a cambios ambientales. La referencia para la comparación de datos de organismos fotosintéticos suele ser el contenido en clorofila, pero en el caso de la deficiencia de hierro, ésta no debe de ser utilizada como referencia, ya que su síntesis es uno de los sucesos afectados por la falta de hierro, y se produce clorosis. Las proteínas totales como valor de referencia puede resultar útil en una situación de laboratorio, con las condiciones controladas de forma que todos los parámetros van a resultar muy repetitivos, y siempre que estemos utilizando un cultivo axénico. Otra alternativa podría ser referir la cantidad de flavodoxina a la de ferredoxina-NADP<sup>+</sup> reductasa (FNR), enzima esencial de la misma cadena transportadora de electrones, que no se ve afectada por la deficiencia de hierro (Inda y Peleato, resultados no publicados), y que nos daría idea de la organización y funcionalidad de las membranas fotosintéticas. Sin embargo, la valoración de la cantidad de ferredoxina remanente puede ser el referente más adecuado (Doucette et al. 1996).

Un indicador biológico ideal debería de permitirnos valorar de una forma rápida y precisa la disponibilidad de hierro en medios poblados por un fitoplancton muy heterogéneo. Un método ideal debería de poder ser utilizado "in situ", es decir valorando los parámetros en una población de organismos muy heterogénea, sin cultivos ni manipulaciones posteriores, no sólo costosos en tiempo, sino que podrían desvirtuar completamente los resultados. La valoración de todas las flavodoxinas presentes en el conjunto de organismos podría ser una alternativa, pero presenta el problema de la distinta respuesta que pueden tener los organismos a la expresión de esta proteína, sin desdeñar la posible contribución de bacterias heterótrofas que podrían acrecentar los valores medidos (Tortell et al. 1996). Esta opción implicaría tener una buena referencia, puesto que en este caso las proteínas totales iban a estar determinadas por el conjunto de la población, no sólo por los productores de flavodoxina, por lo que, como proponen Doucette et al (1996), la medida más adecuada podría ser la razón flavodoxina/ferredoxina. La elección de un organismo indicador, bien estudiado en el laboratorio, tendría bastantes ventajas frente a la heterogeneidad de partida, aunque el problema sería en este caso separarlo, o diferenciar claramente sus proteínas. Esta elección de un organismo indicador, o por el contrario el trabajo con el conjunto de la población, determina el método

elegido para valorar la flavodoxina/ferredoxina. Los métodos inmunoquímicos, sea cual sea su índole, presentan cierta especificidad en el reconocimiento de las proteínas, de forma que por una parte hay reacciones cruzadas con proteínas similares, y no con la misma intensidad, y por otra, no reconocen la totalidad de, por ejemplo, las flavodoxinas presentes en la población. Por esta razón este tipo de métodos deben de ser usados con mucha precaución si se trata de poblaciones hetero- géneas. Los métodos en los que se cuantifican el total de las flavodoxinas mediante cromatografía de alta resolución (HPLC) (Doucette et al, 1996), pueden ser más adecuados para valoraciones de toda la flavodoxina presente en un filtrado de planctón, pero la interpretación de los datos puede llevar a conclusiones mucho más erróneas, dados los distintos comportamientos respecto a la síntesis de flavodoxina que pueden tener los organismos: ser constitutiva, ausencia total o inducida como respuesta a la deficiencia de hierro.

Con objeto de estudiar el comportamiento de otra proteína que contenía hierro, y también transportador electrónico fotosintético, se purificó el citocromo *c<sub>6</sub>*. Esta proteína es un donador de electrones al fotosistema I, que puede sustituir a la plastocianina (que desempeña este papel en otros organismos), cuando la deficiencia de cobre impide la síntesis de esta cuproproteína. En algas verdes coexisten ambas proteínas, y con objeto de encontrar respuestas cuantificables al estrés de hierro, se ha estudiado la presencia de esta proteína en *Chlorella*, siempre bajo la óptica del posible uso de este organismo como marcador. Sin embargo, la cantidad de este citocromo en las células va a ser también dependiente de la cantidad de cobre disponible, por lo que valorar el citocromo *c<sub>6</sub>* va a ser presumiblemente menos interesante que valorar la flavodoxina presente o ferredoxina remanente.

Nuestra propuesta es:

- 1) Elección de un organismo (o grupo de organismos) indicador, que deberá reunir una serie de requisitos como ubicuidad, distancia filogenética adecuada con los restantes, etc...
- 2) Que en este organismo esté perfectamente tipcada su respuesta a la deficiencia de hierro.
- 3) Que inmunológicamente pueda valorarse la flavodoxina, la ferredoxina, y presumiblemente la FNR, mediante técnicas de ELISA, inmunolectrodifusión, o incluso microscopía con inmunofluorescencia.

Alguna característica antigénica de ese organismo o grupo de organismos podría servir para distinguirlo de los demás.

La flavodoxina es un indicador de deficiencia biológica de hierro muy adecuado, aunque sigan pendientes gran número de problemas por resolver. Los estudios en el laboratorio sentarán las bases para proponer métodos que puedan ser cotejados en poblaciones naturales.

### **Bibliografía**

- Bes, M.T., Razquin, P. y Gómez-Moreno, C. (1995). *Preparative Biochemistry*, 25: 89-97.
- Bouma, D. (1983). En : *Inorganic Plant Nutrition*. Editado por Lauchi, A y Bielecki, R.L. Springer-Verlag, Berlin. pg.120-146.
- Coale, K. H., Johnson, K.S., Fitzwater, S.E., Gordon, R.M., Tanner, S. Chavez, F.P. Ferioli, L. Sakamoto, C. Rogers, P., Millero, F. Steinberg, P. Nightingale, P., Cooper, D. Cochlan, WP., Landry, M.R, Constantinou, J., Rollwagen, G., Trasvina, A. and Kudela, R. (1996). *Nature* 383, 495-501.
- Cooper, D.J., Watson, A.J. y Nightingale, P.D. (1996). *Nature* 383: 511-513.
- Entsch, B., Sim, R.G. y Hatcher, B.G. (1983). *Marine Biology* 73: 17-30.
- Doucette, G. J., Erdner, D.L., Peleato, M.L., Hartman, J.J. and Anderson, D.M. (1996) *Mar. Ecol. Progress Series*, 130, 269-276.
- Geider, R.J. y La Roche, J. (1994). *Photosynthesis Research* 39: 275-301.
- Glober, H.E. (1977). *J. Phycology* 13: 208-212.
- Huntsman, S.A. and Sunda, W.G. (1980). En: *The physiological ecology of phytoplankton*. pg 285-328. Editado por I.Morris. Blackwell Scientific, Oxford (Inglaterra).

- Hutbert, G.N., Huston, K.G y Rogers, L.J. (1977). FEMS Microbiol.Letters 1: 193-196.
- Inda, L.A., Medina, M., Saraiva, L.M., Gómez-Moreno, C., Teixeira, M. and Peleato, M.L. (1997). Photosynthesis Research : - .(En prensa)
- Katoh, S. (1960). Nature 4724:533-534.
- Kessler, E. and Czygan, F.C. (1970) Arch. Mikrobiol. 70, 211-216.
- Laurell, C.B. (1966) Anal. Biochem. 15, 45-52.
- Lowry, O.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J. (1951) J.Biol. Chem. 193, 265-275.
- MacKinney G. (1941) J.Biol. Chem. 140, 314-322.
- Medina, M., Louro, R., Peleato, M.L., Mendes, J., Gognon, J., Gómez-Moreno, C., Xavier, A. y Teixeira, M. (1997). Journal Biol. Inorganic. Chemistry 2, 225-234.
- Pardo, B. Gómez-Moreno, C y Peleato, M.L. (1990). Arch. Microbiol.153: 528-530.
- Peleato, M.L., Ayora, S., Inda, L.A. and Gómez-Moreno, C. (1994) Biochem. J. 302, 807-811
- Sandmann, G., Peleato, M.L., Fillat, M.F.Lázaro, M.C. y Gómez-Moreno, C. (1990). Photosynthesis Research 26: 119-125.
- Tortell, P.D., Maldonado, M.T. y Price, N.M.(1996) Nature 383: 330-332.
- Vígara, J., Inda, L.A., Vega, J.M., Gómez-Moreno, C. y Peleato, M.L. (1997) Enviado a Photochemistry and Photobiology.

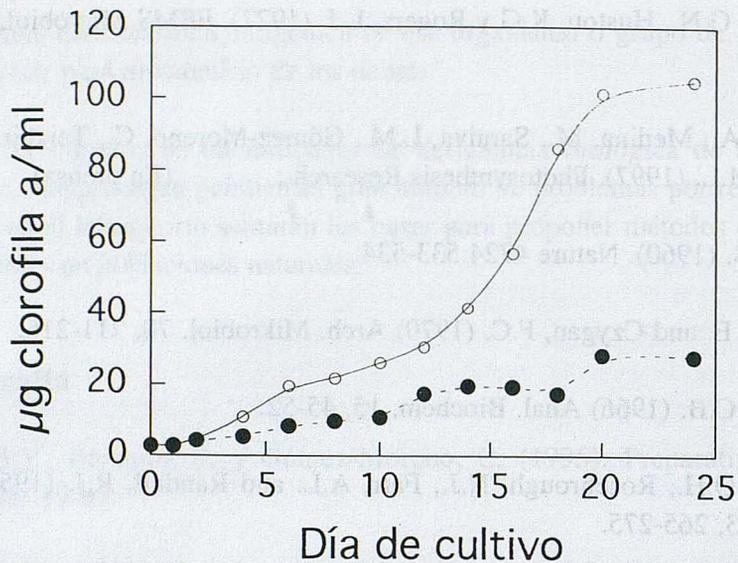


Figura 1: Contenido en clorofila *a* de células de *Chlorella fusca* cultivadas en medio 18  $\mu\text{M}$  de hierro (o), y cultivadas en 0.18  $\mu\text{M}$  de hierro (o).

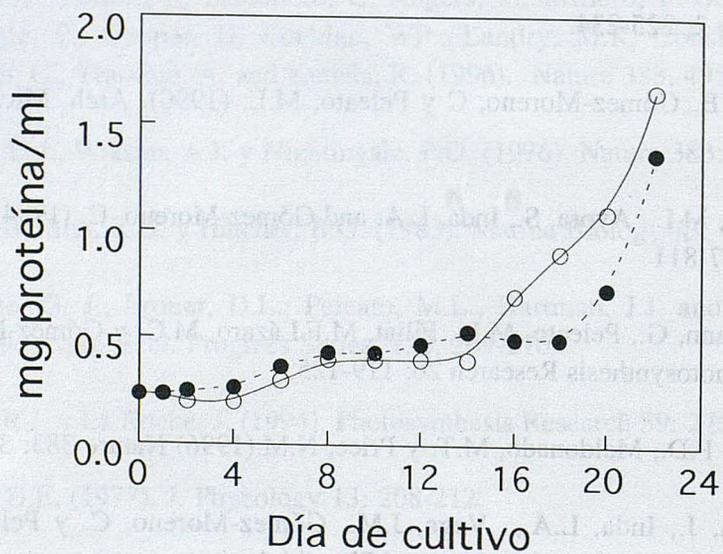


Figura 2: Determinación del contenido en proteínas de células crecidas en medio completo 18  $\mu\text{M}$  hierro (o), o medio deficiente 0.18  $\mu\text{M}$  hierro (o).

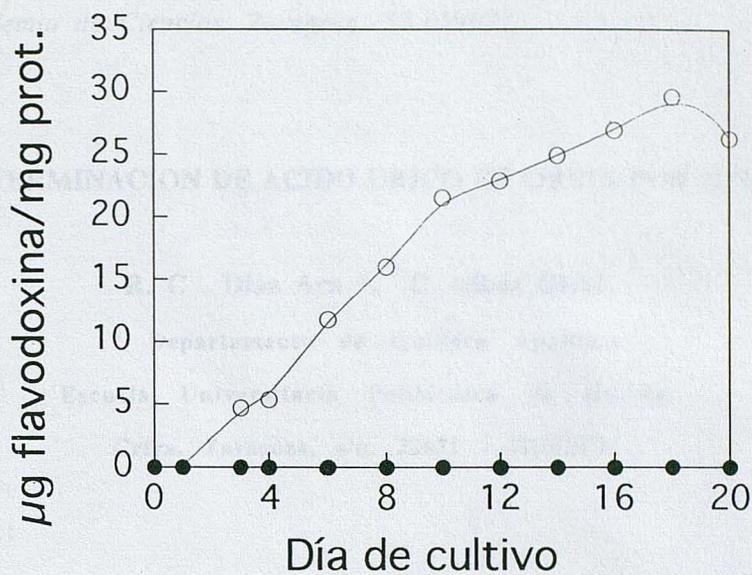


Figura 3: Contenido de flavodoxina en células mantenidas en medio completo 18  $\mu\text{M}$  de hierro (o), o transferidas a un medio deficiente 0.18  $\mu\text{M}$  de hierro (●).

## DETERMINACION DE ACIDO URICO EN ORINA POR HPLC

R. C . Díaz Ara \*, C. Albás Giral

\* Departamento de Química Analítica

Escuela Universitaria Politécnica de Huesca

Crtra. Zaragoza, s/n. 22071 - HUESCA

### Abstract:

A new HPLC method for the determination of uric acid in urine samples, using both external and internal standard techniques, in isocratic régime, has been proposed. The results are comparable to the obtained by the enzymatic method, used as reference, for 50 samples.

### 1.- Introducción.

La mayor parte de los métodos usados para la determinación de ácido úrico en orina suponen su oxidación a alantoína por medios químicos o enzimáticos. El método más específico consiste en la oxidación por la uricasa del ácido úrico (1). Se relaciona la diferencia de absorbancias antes y después del tratamiento de la muestra con la concentración de ácido úrico en la muestra de orina. En base a la reacción anterior, también puede conocerse indirectamente el contenido en ácido úrico, por determinación del agua oxigenada producida.

La mayoría de los métodos espectrofotométricos de determinación de ácido úrico se basan en su acción reductora sobre el fosfowolframato, en solución alcalina, dando lugar a la formación de un cromógeno azul (2). Cabe destacar el método clásico descrito por Folín (3), que ha sido modificado para aumentar la sensibilidad de la determinación (4-7).

Entre los métodos cinéticos merecen especial mención los basados en el empleo de sistemas enzimáticos, con seguimiento espectrofotométrico en la zona del ultravioleta (8-11), que son poco sensibles o selectivos. Los métodos cromatográficos también se han usado con

frecuencia. La cromatografía de gases permite la determinación simultánea de hipoxantina, xantina, alopurinol, oxipurinol y ácido úrico en fluidos fisiológicos (12-17). Por HPLC se ha determinado el ácido úrico proveniente de la transformación enzimática de la xantina (18) y simultáneamente con alantoína y oxipurinas en fluidos biológicos (19).

La mayoría de los métodos citados presentan algún inconveniente, como la excesiva duración del tratamiento previo de la muestra, baja sensibilidad y selectividad, etc. En este trabajo se presenta un método sencillo y rápido de determinación de ácido úrico en muestras de orina, por HPLC en régimen isocrático.

## **2.- Parte experimental.**

### **2.1. Aparatos.**

Se ha utilizado un cromatógrafo de líquidos Hewlett - Packard de la serie 1050. La columna es una Nucleosil 120 C - 18, de 25 cm de larga, 0,46 cm de diámetro interno y 5 mm de espesor de película. El integrador es de Hewlett - Packard, Modelo 3396 A.

### **2.2. Reactivos.**

Todos los reactivos son de grado analítico. El agua es de calidad ultrapura. El acetonitrilo usado como fase móvil es de calidad para HPLC.

### **2.3. Procedimientos.**

#### **2.3.1. Influencia del pH.**

El valor del pH de las muestras determina la aparición o no de dos picos correspondientes a la disociación del ácido úrico. Para evitarlo, se usan a menudo disoluciones tampón de fosfatos (12,19), que, sin embargo, provocan una inestabilidad en la presión, lo que no aconseja su uso. Ya que el pK del ácido úrico es de 5,75, se trabaja a un pH inferior, para asegurarse que el ácido úrico se encuentra en su forma no disociada. El pH óptimo resulta ser de 3 y se logra por adición de ácido acético. A  $\text{pH} < 2$  pueden producirse fenómenos de hidrólisis que originan precipitados que dañan la columna.

#### **2.3.2. Tratamiento de las muestras.**

Previamente a las determinaciones, las muestras se conservan en una nevera a 4°C. Para la determinación por el método del patrón externo, las muestras de orina, una vez filtradas, se diluyen 1:20 (v:v) con

dietilamina, añadiendo ácido acético para regular el pH. Se centrifuga a continuación durante 10 minutos. La curva de calibrado se realiza con patrones de ácido úrico, en el rango 0-6 mg.dl<sup>-1</sup>.

En el método del patrón interno, se diluye la muestra 1:20 citada anteriormente 1:2 (v:v) con alopurinol, usado como patrón interno. A los patrones de ácido úrico se les añade la misma cantidad de alopurinol (1 ml) que a las muestras.

### 2.3.3. Optimización de parámetros cromatográficos.

Tras la realización de los correspondientes ensayos previos, con diferentes fases móviles y variando los correspondientes flujos, se obtienen los parámetros cromatográficos que se observan en la Tabla 1.

**TABLA 1**

F.Móvil (% v:v)	Flujo F.M. (ml.min <sup>-1</sup> )	Compuesto	T.retención (min)
Tampon fosfatos	1,5	Acido úrico	6,35
Tampon fosfatos	1,5	Alopurinol	13,81
Agua	1,5	Acido úrico	6,79
Agua	1,5	Alopurinol	13,96
Agua:CH <sub>3</sub> CN, 99:1	1,5	Acido úrico	4,15
Agua:CH <sub>3</sub> CN, 99:1	1,5	Alopurinol	7,28
Agua	1,3	Acido úrico	8,13
Agua	1,3	Alopurinol	14,39
Agua:CH <sub>3</sub> CN, 99:1	1,3	Acido úrico	5,86
Agua:CH <sub>3</sub> CN, 99:1	1,3	Alopurinol	21,84
Agua:CH <sub>3</sub> CN, 98:2	1,3	Acido úrico	4,30
Agua:CH <sub>3</sub> CN, 98:2	1,3	Alopurinol	7,36

**Tabla I.- Datos de algunos parámetros cromatográficos obtenidos en ensayos previos.**

En base a los datos obtenidos, se realiza la determinación con agua:acetonitrilo como fase móvil, con un flujo de 1,3 ml.min<sup>-1</sup>, en régimen isocrático, a temperatura ambiente y con detección a 280 nm. El pH se mantiene en 3, para evitar la disociación del ácido úrico. Se inyectan 10 ml.

### 3.- Resultados y discusión.

#### 3.1. Determinación de ácido úrico en orina por patrón externo.

Se preparan patrones de ácido úrico en las condiciones citadas anteriormente, en el rango  $0-6 \text{ mg.dl}^{-1}$ . Se obtiene una recta cuya ecuación es  $y = 9191,41 + 957164,97x$ , siendo  $y$  el área, expresada en cuentas ( $1 \text{ cuenta} = 1,24 \cdot 10^{-4} \text{ mV}$ ), y  $x$  la concentración de ácido úrico. El coeficiente de regresión es  $R = 0,9999$ . A continuación se inyecta la muestra de orina, previamente preparada, y el resultado se interpola en la recta obtenida. Un cromatograma de la muestra inyectada se incluye en la siguiente Figura 1.

FIGURA 1

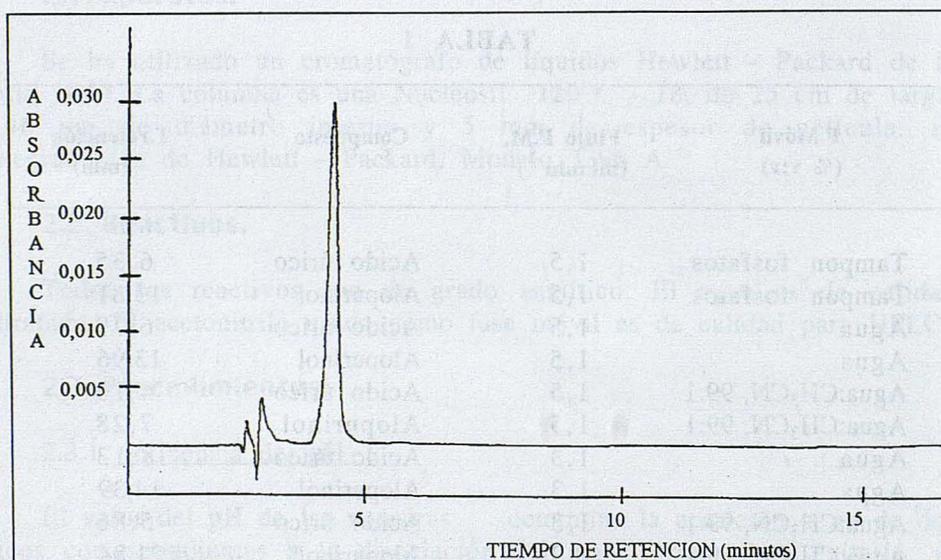


Figura 1: Cromatograma de una muestra de orina para determinar ácido úrico por patrón externo.

#### 3.2. Determinación de ácido úrico en orina por patrón interno.

Se añade a los patrones y muestras 1 ml de alopurinol 0,5 mM, como patrón interno. La curva de calibrado se obtiene representando  $A_{\text{ácido úrico}} / A_{\text{alopurinol}}$  frente a  $[\text{Acido úrico}] / [\text{Alopurinol}]$ . La respuesta es lineal en el rango  $0-0,5 \text{ mg.dl}^{-1}$ , siendo la ecuación de la recta  $A_{\text{ácido úrico}} / A_{\text{alopurinol}} = 0,043 + 11,126 \cdot [\text{Acido úrico}] / [\text{Alopurinol}]$ , con un

coeficiente de correlación,  $R = 0,9999$ . Un cromatograma de la muestra inyectada se muestra en la siguiente Figura 2.

FIGURA 2

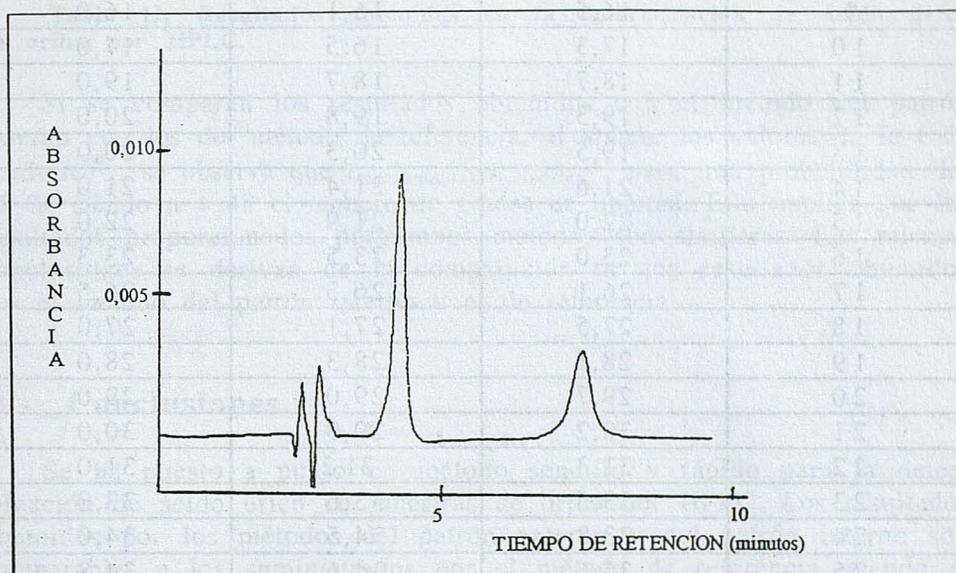


Figura 2: Cromatograma de una muestra de orina para determinar ácido úrico con patrón interno.

El orden de los picos que aparecen es el siguiente: ácido acético, dietilamina, ácido úrico y alopurinol.

Se ha realizado la determinación de ácido úrico en 50 muestras de orina por los dos métodos citados. Los valores medios para 3 determinaciones se muestran en la tabla II.

Los valores obtenidos con el método de referencia (1) han sido suministrados junto a las muestras correspondientes en el Hospital San Jorge de Huesca.

TABLA II

Número de muestra	Patrón externo	Patrón interno	Método referencia
1	6,5	6,6	6,7
2	9,3	9,7	9,5
3	9,4	9,7	9,6

4	10,7	10,3	10,0
5	11,2	11,7	12,0
6	13,5	13,2	13,0
7	12,7	13,6	13,0
8	13,5	13,8	14,0
9	16,5	16,1	16,0
10	17,3	16,5	17,0
11	18,7	18,7	19,0
12	19,3	19,8	20,0
13	19,5	20,3	20,0
14	21,6	21,4	21,0
15	22,0	21,9	22,0
16	23,0	23,5	23,3
17	26,1	26,3	26,5
18	27,5	27,1	27,0
19	28,8	28,3	28,0
20	29,7	29,0	29,0
21	30,2	29,9	30,0
22	31,3	31,2	31,0
23	33,5	33,6	33,0
24	33,8	34,5	34,0
25	35,6	34,6	34,8
26	35,7	35,4	35,0
27	35,8	36,5	36,0
28	39,1	39,4	39,7
29	40,2	41,1	40,6
30	42,1	41,8	41,6
31	43,4	43,5	43,7
32	43,9	44,6	44,5
33	45,1	45,0	45,6
34	46,2	46,5	46,0
35	47,8	47,4	48,0
36	48,5	48,5	49,0
37	50,9	49,8	50,0
38	51,3	51,9	52,0
39	52,7	53,7	53,2
40	53,8	54,1	54,0
41	54,9	55,3	55,7
42	57,2	57,3	57,0
43	58,2	58,2	58,0
44	58,4	58,5	59,0
45	59,1	59,1	59,5
46	59,2	59,8	59,8

47	60,6	60,9	61,0
48	63,3	62,8	63,1
49	64,8	63,8	64,2
50	67,5	67,9	68,0

**Tabla II: Resultados obtenidos en la determinación de ácido úrico en orina por HPLC.**

Si se comparan los resultados obtenidos con el método del patrón externo con los del método de referencia, al aplicar los valores de la  $t$  de student<sup>20</sup>, se observa que  $t_{\text{experimental}} < t_{\text{teórica}}$ , para una probabilidad del 95 %, siendo  $n = 48$  el número de grados de libertad. Ello implica que los resultados proporcionados por ambos métodos son similares. Las mismas conclusiones se derivan de la comparación de los resultados obtenidos por el método del patrón interno y el de referencia.

#### 4.- Conclusiones.

Se ha puesto a punto un método sencillo y rápido para la determinación de ácido úrico en muestras de orina por HPLC. Los resultados obtenidos por los métodos del patrón externo y del patrón interno son comparables a los suministrados por el método de referencia, siendo el coste aproximado de cada determinación por cromatografía líquida bastante inferior al del método oficial de los hospitales del Insalud.

#### 5.- Bibliografía.

- (1) Henry, J.B.(1988): *Diagnóstico y tratamiento clínicos por el laboratorio*, 8ª Edición. Editorial Salvat Barcelona.
- (2) Henry, R.J., Cannon, D.C., Winkelman, J.W., 1989, *Química Clínica. Principios y Técnicas*, Editorial Jims.
- (3) Folin, O. (1930): *Biol. Chem.*, 86, 179.
- (4) Patel, C.P. (1968): *Clin. Chem.*, 14, 764.
- (5) Wells, M.G. (1968): *Clin. Chim. Acta*, 22, 379.
- (6) Jung, D.H., Parekh, A.C. (1970): *Clin. Chem.*, 16, 247.
- (7) Pileggi, V.J., Di Giorgio, J., Wybenga, D.R. (1972): *Clin. Chim. Acta*, 37, 141.
- (8) Dilena, B.A., Peake, M.J., Pardue, H.L., Skoug, J.W. (1986): *Clin. Chem.*, 32, 486.
- (9) Hamilton, S.D., Pardue, H.L. (1982): *Clin. Chem.*, 28, 2359.
- (10) Frew, J.E., Jones, P. (1985): *Anal. Lett.*, 18, 1579.

- (11) Manuel de Villena, F.J., Pérez, R., Polo, L.M. (1990): Anales de Química, Serie B, 86, 271.
- (12) Putterman, G.J., Skaik, B., Hallmark, M.R., Sawyer, C.G., Hixson, C.V., Perini, F. (1979): Anal. Biochem., 98, 18.
- (13) Iwase, H., Kimura, T., Sugiyama, T., Murai, A. (1975): J. Chromatography, 106, 213.
- (14) Miller, V., Pacáková, V., Smolkova, E. (1976): J. Chromatography, 119, 355.
- (15) Sweetman, L., Nyhan, W.L. (1968): J. Chromatography, 32, 662.
- (16) Sweetman, L., Nyhan, W.L. (1971): J. Chromatography, 59, 349.
- (17) Brown, M., Bye, A. (1977): J. Chromatography, 143, 195.
- (18) Chen, X.B., Vrskov, E.R., Howell, DeB. (1990): J. Nutr., 63, 121.
- (19) Balcells, J., Guada, J.A., Peiró, J.M., Parker, D.S. (1992): J. Chromatography, 575, 153.
- (20) Miller, J.C., Miller, J.N. (1996): *Estadística para Química Analítica*. 2ª Edición. Addison - Wesley Iberoamericana.

## Comparison of Biospheric Models of Radionuclides Transfer

Antonio García-Olivares<sup>1</sup> & Elena Carrasco

### Abstract

After the Chernobyl accident, the international BIOMOVS A4 exercise made it possible that a set of biospheric transfer models could predict the daily radionuclide concentration in soils, forage and in some animal products (cow milk and beef) in areas affected by the Chernobyl accident. The aim was to compare these predictions with experimental results in 13 locations around the world. The data provided were essentially the daily air contamination and precipitation and some site-dependent parameters.

It was a blind test, the locations and experimental measures were not revealed in advance. Twenty-three models (quasi-steady state and time-dependent models) were involved in the study.

In this paper some explicit criteria have been used to select the models that better matched the experimental results. In nine selected locations a comparative analysis between these models has been carried out for obtaining the structural and parametric coincidences that could explain their relatively good performance.

The first evidence obtained has been that a wide set of models were able to predict the order of magnitude of the nuclides time-integrated concentrations in several biospheric compartments. But only a few models, all of them with a 'dynamical' structure, fitted the daily behavior with a reasonable agreement.

The dynamical structure of the most successful models at predicting for Caesium 137 (CIRCLE, ECOSYS, PRYMA, RAGTIME, FARMLAND and BILTH) shows some common patterns that may be relevant for modelling of nuclear accident scenarios.

### 1. Introduction.

Ten years after the Chernobyl accident, the scientific study of the radionuclides released by the accident has made it possible to quantify the rate of transfer of the two most important nuclides (Cs-137 and I-131) through some important biospheric paths. The effort was in part leadered by the BIOMOVS (Biospheric Models Validation Studies) and VAMP (Validation of Models Predictions) international programs. One of the most important projects in the BIOMOVS program was the "A4 exercise", which tried to compare the predictions of models designed in different institutes with experimental data obtained previously in 13 locations around the world.

The data provided were essentially the daily air contamination and precipitation and some site-dependent parameters. It was a blind test: the locations and experimental measures were not revealed in advance.

<sup>1</sup> Instituto LITEC. Consejo Superior de Investigaciones Científicas. C/ Marfa de Luna, 3. 50015 Zaragoza. Spain.

Twenty five models participated in the BIOMOVS A4 exercise, two of them included in a later stage. It is possible to classify these models either by their method of prediction or by means of their intended accuracy.

In the first case the classification criteria used are:

(1) *Steady* models: models using transfer ratios. A transfer ratio is defined as the concentration expected at the steady state, obtained when  $t \rightarrow \infty$  with a constant input of activity, per unit of input;

(2) *Dynamic* models: models using transfer rates and first order differential equations for the mass balance over the time.

In the second case the criteria used are: (a) models attempting a *conservative* prediction of the nuclides' concentration, with radiological protection purposes, and (b) models attempting a *best estimate* prediction.

Using the two criteria four groups result: Steady/conservative (St/c, 3 cases), Steady/best (St/b, 7 cases), Dynamic/conservative (Dy/c, 3 cases) and Dynamic/best (Dy/b, 12 cases).

In order to analyse the performance of the different models it is necessary to compare their predictions with the observed concentrations in vegetation, and milk/beef of cattle. When trying to compare the predictions with the observations it is possible to use the daily values (predicted and observed) or the time-integrated values.

## 2. Reliability of the models in the prediction of integrated values.

In order to quantify this performance in a way as objectively as possible, the number of "hits" of each model was quantified as follows: a "hit" was defined as a prediction/observation ratio in the range 0.5 to 2. All the available observations (50 data on 9 selected locations) of integrated deposition, vegetation concentration and milk concentration of  $^{137}\text{Cs}$  and  $^{131}\text{I}$  were used as geometric means, as were the geometric means of the transfer factors calculated for air to deposition, air to milk, deposition to vegetation and vegetation to milk for iodine and caesium. The beef concentration was not used due to the scarcity of experimental data available [Kölher 1991].

Table 1 shows the ranking of models obtained.

A steady state model appears in the ranking, showing that some steady state models are able to predict integrated quantities with roughly the same accuracy as dynamic models.

Figures 1 to 6 show the variability obtained in the prediction/ observation ratios (P/O) of some integrated magnitudes [Kölher 1991]. In these figures, the symbol "St/c", "Dy/b", etc, was added to identify the type of each model in our classification.

As can be observed in these figures, the bias and variability of the P/O ratios in integrated variables is not very different in steady state models in relation to best-estimate models. Since for some purposes, i.e. regulatory, an estimation of the integrated concentration is enough, the use of good steady state models seems to be acceptable.

The best models according to these criteria are: RAGTIME, PRYMA, SCHRAADLO, ECOSYS, FARMLAND2, BILTH and CIRCLE.

The authors and/or users of these models are introduced in the following list.

**RAGTIME** (Pleasant et al, 1980) (Killough & Hoffman, 1987). Users: G.G. Killough, F.O. Hoffman, Hendecagon Corporation ORNL, Oak Ridge, TN, USA.

**PRYMA** (García-Olivares 1992)(García-Olivares et al, 1993). Users: A. García-Olivares, E. Carrasco, A. Suañez, CIEMAT-IMA, Madrid, Spain.

**SCHRAADLO**. Users: J. Horyna, Nuclear Research Institute, Rez, Czechoslovakia.

**ECOSYS**. Users: H. Müller, G. Pröhl, GSF, Neuherberg, Germany.

**FARMLAND**. Users: J. Brown, A. Jones, NRPB, Chilton, U.K.

**BILTH**. Users: R. Blaauboer, RIVM, Bilthoven, The Netherlands.

**CIRCLE**. Users: T. Homma, O. Togawa, JAERI, Tokai, Japan.

The intended accuracy is best estimate and in five of them the type of prediction is deterministic (RAGTIME, ECOSYS, PRYMA, FARMLAND and SCHRAADLO). In two of them the prediction is a distribution (BILTH and CIRCLE). Except for SCHRAADLO, the remaining models are dynamic.

The models ECOSYS and BILTH were calibrated to Chernobyl data at their own site, and the model RAGTIME was calibrated after knowing the A4 data.

The FARMLAND code was used to make two sets of predictions for iodine. In FARMLAND2 the iodine speciation was taken into account and in FARMLAND1 it was not. FARMLAND2 corresponds to reality much better than FARMLAND1 and it will be called "FARMLAND" in the following.

In the following pages the cited models except SCHRAADLO will be compared with each other in order to detect some patterns in their structure and the parameters that they use.

### 3. Comparison of the six models.

The six models analysed are best-estimate dynamical models using first order differential equations for the mass balance in time. Two of them use transfer ratios also (BILTH and ECOSYS).

Figures 7 to 12 illustrate the kind of predictions obtained from dynamic models. In these figures the predictions are plotted together with the observed values in two out of 13 locations.

#### 3.1 Deposition.

In three of the models (ECOSYS, PRYMA and BILTH) the wet deposition  $F_w$  ( $Bq/m^2d$ ) is calculated through an expression involving the cloud height  $H$  (m) and a washout constant  $\alpha$  ( $h\ mm^{-1}s^{-1}$ ):

$$F_w = 3600 \alpha H I c_{air}(t)$$

where  $I$  is the daily precipitation rate (mm/d) and  $c_{air}$  the air concentration ( $Bq\ m^{-3}$ ). FARMLAND uses a washout constant for a fixed  $H$ , in units of  $s^{-1}$ .

In the other two models (RAGTIME and CIRCLE) it is calculated through an expression involving the distribution coefficient  $W$  [ $(Bq/l)/(Bq/m^3)$ ]

representing the nuclide concentration in rain droplets per unit air concentration:

$$F_w = W I_{\text{air}}(t)$$

Both expressions can be related, however, since  $W = 3600 \alpha H$ .

Table 2 summarizes the values of the effective  $W$  used by the models.

The assumptions made about the iodine speciation are different in different models. Some models (FARMLAND2, ECOSYS and BILTH) assume a gaseous  $^{131}\text{I}$  fraction equal to that measured at Neuharberg (Germany), one of the two locations where the iodine speciation was specified.

Two models (RAGTIME and CIRCLE) assume that the speciation in Tokai (Japan) was representative of all the other locations. PRYMA assumes two different speciation kinds which depend on the distance from the location to Chernobyl: the fractions measured in Neuharberg are applied in European locations and the fractions measured in Tokai are applied in the remaining locations.

The dry deposition of contamination ( $\text{Bq/m}^2\text{d}$ ) onto soil and plants is modeled through a deposition velocity  $v_g$  (m/s) by all the models (Table 3).

The models use a different deposition velocity for different physicochemical forms of iodine (elemental, organic and particulate), except CIRCLE which does not consider iodine speciation.

Table 3 shows that the deposition velocity values used range between a minimum value of  $4.E-4$  m/s and a maximum value of  $3.E-3$  m/s for caesium; and between  $1.E-6$  m/s and  $1.5E-2$  m/s for iodine.

For iodine the observed range was  $4.3E-3$  (Loviisa) to  $9.E-3$  m/s (Neuharberg) with mean value  $6.6E-3$  m/s. For Caesium the observed range was  $2.8E-3$  (Anguillara) to  $1.8E-2$  m/s (Neuharberg) with a mean value of  $1.1E-2$  m/s [Köhler et al 1991].

Therefore some of the best models have used a central value for  $v_g$  which is out of the observed range. However, this had no visible effect on their performance due to the fact that the wet deposition was dominant over the dry deposition in most of the locations. This may be related to the depth of the contaminated cloud of Chernobyl [Fulkner 1987].

### 3.2 Interception.

The dry interception fraction by the vegetation ( $R$ ) is simulated by the Chamberlain expression in ECOSYS, RAGTIME, FARMLAND and CIRCLE :

$$R = 1 - \exp(-\mu Y)$$

where  $R$  is the fraction of deposition intercepted by the vegetation (forage crops in this case),  $\mu$  is a coefficient typically between 2.3 and  $3.3 \text{ m}^2/\text{Kg}$  according to [Till 1983] and  $Y$  is the yield factor in dry weight units ( $\text{Kg/m}^2$ ).

Three models (RAGTIME, CIRCLE and FARMLAND) use a median  $\mu$  value of  $2.8 \text{ m}^2/\text{Kg}$ . ECOSYS uses a median value of  $1.0 \text{ m}^2/\text{Kg}$ .

BILTH and PRYMA use mass interception approach with an R value of 0.25 in BILTH and 0.2 for iodine and 0.25 for caesium in PRYMA.

The wet interception fraction was modeled as being equal to the dry interception fraction by three models (RAGTIME, PRYMA and BILTH). The model ECOSYS uses a specific wet interception dependent on the rain rate I (mm/d):

$$R_{wet} = r_1 [1 - \exp(-I \ln 2 / r_2)] [1 - \exp(-r_3 Y)] / I$$

where  $r_1$  is 0.7 mm for iodine and 1.4 mm for caesium;  $r_2$  is 0.3 mm for iodine and 0.6 mm for caesium;  $r_3$  is  $1 \text{ m}^2/\text{Kg}$ .

The model CIRCLE uses a wet interception value of 0.15 for caesium and 0. for iodine.

RAGTIME and PRYMA consider the rain everyday and this approach allows the stormy events to be simulated better. The R values used are listed in Table 4.

### 3.3 Reduction rates.

The main rate of loss of contamination out of the plant is shown in TABLE 5 for the six models.

Three models consider only the losses due to weathering (BILTH, CIRCLE and FARMLAND) and the other three consider growth dilution in addition to weathering. PRYMA and RAGTIME take into account cattle consumption.

As Table 5 shows, the fraction of contamination lost per day by the vegetation is between: 0.05 to 0.13 for iodine and 0.02 to 0.12 for caesium. This is equivalent to residence times between 7 and 14 days for iodine, and 6 and 35 days for caesium.

### 3.4 Vegetation contamination.

Table 6 shows the processes of vegetation contamination and removal that are considered by the models studied.

A way to compare the performance of the different models independently of the particular addressed processes is by means of the concentration factor air to vegetation ( $\text{m}^3/\text{kg d.w.}$ ) (Table 7).

Figure 13 shows the "effective" concentration factors as obtained from the models' predictions in the 13 locations and it compares them with the values experimentally estimated for this parameter.

If the whole set of participant models is used [Köhler et al, 1991], the geometric mean for the air to vegetation ratios of Iodine concentration ranges from  $2110 \text{ m}^3/\text{kg}$  to  $34256 \text{ m}^3/\text{kg}$  with mean value in  $7522 \text{ m}^3/\text{kg}$ . Therefore, the range of the best models is in the lower extreme of this range. For caesium, the range is from  $3392 \text{ m}^3/\text{kg}$  to  $64323 \text{ m}^3/\text{kg}$ , with a mean of  $21034 \text{ m}^3/\text{kg}$ .

If the set of the non-selected models is used the geometric mean ranges from  $2255 \text{ m}^3/\text{kg}$  to  $34256 \text{ m}^3/\text{kg}$  with a mean  $9167 \text{ m}^3/\text{kg}$  for iodine, and for caesium the range is from  $3392 \text{ m}^3/\text{kg}$  to  $64323 \text{ m}^3/\text{kg}$  with a mean of  $25723 \text{ m}^3/\text{kg}$ .

It can be observed that the variability generated by the best models is lower or similar to the observed variability for iodine as well as for caesium. However, this is a feature observed also in the non-selected models.

Figures 13 and 14 show the predicted versus the observed concentration ratio for  $^{131}\text{I}$  and  $^{137}\text{Cs}$  for the seven selected models. Observed concentration ratios were defined as the quotient of the integrated vegetation concentration to integrated air concentration, both observed in the whole period. Theoretical concentrations ratios were defined as a similar quotient but using the values predicted by the models. It can be observed that the scatterplots of the six dynamic best-estimate models are closer to the diagonal of the figure, when compared with the steady conservative model selected.

The relatively good concentration ratio emerging from the selected models does not imply, however, an identical choice of compartments and parameters. In fact, the effective concentration ratio from air to vegetation is controlled by the values of the deposition parameters, by the values of the loss-rates, and by the number of intermediate compartments delaying the loss-rates in relation to the deposition flow.

If the air concentration and the rain rate (I) were steady, this last feature would not be relevant, and the effective concentration ratio would coincide with the theoretical concentration ratio  $CF_{th}$  obtained through the following expression:

$$CF_{th} = \frac{R v_g / \lambda Y + R I W}{(\lambda_d + \lambda_w) Y} = \frac{R v_g / \lambda Y}{(\lambda_d + \lambda_w) Y} + \frac{R I W}{(\lambda_d + \lambda_w) Y} = CF_{dry} + CF_{wet}$$

where R is the fraction of deposition intercepted by the crop,  $v_g$  is the dry deposition velocity,  $\lambda$  is the radioactive decay constant, Y is the yield factor of the crop, I is the time precipitation rate, W is the effective distribution coefficient,  $\lambda_d$  is the loss rate due to dry weathering (wind drag and sublimation),  $\lambda_w$  is the loss rate due to wet weathering (rain drag), and  $CF_{dry}$  and  $CF_{wet}$  are the theoretical dry and wet concentration factors.

Table 8 shows the  $CF_{dry}$  and  $CF_{wet}$  obtained for the six selected models for a rain rate equal to 1.37 mm/d.

The values obtained in table 8 are in the range of the effective concentration factor observed in the locations (fig. 13 and 14) for almost all the models but they are not equal. The difference must be related to: (i) the differences in the input and output parameter values and to differences in the number of compartments; (ii) the fact that the Chernobyl cloud and deposition was not stationary but sharply varying and intermittent.

### 3.5 Milk contamination.

The contamination of animal products (milk and beef) were calculated through a time-dependent approach in three cases (ECOSYS, RAGTIME and CIRCLE) and the other three cases (PRYMA, FARMLAND and BILTH) with a multicompartamental model approach. In the time-dependent approach the equilibrium transfer factor (Fm) was converted into a time-dependent form, calculating radionuclide concentrations in milk and beef on the time-dependence of radionuclide concentration in forage.

Only one of the models considers cow inhalation of air contaminated by resuspended radionuclides (FARMLAND, 130 m<sup>3</sup> per day). In three cases (PRYMA, FARMLAND and BILTH) soil consumption by the cow was considered (4% of dry matter intake in the two first models and 0.75 kg per day in the third one).

Table 9 shows the transfer factor Fm (d/l) derived from the six models. Fm is the ratio of concentration that appears in the long term in the milk (Bq/l) per unit of the radionuclide ingestion (Bq/d) if a steady contamination is given to the animal.

As a reference to compare these values, the observed values for iodine in cattle of different locations ranged between 1.E-3 d/l at Belgium and 5.E-3 d/l in Hungary, with a mean value of 2.6E-3 d/l. The mean value for caesium was 4.3E-3 d/l, a value very similar to that observed from weapons fallout studies [Köhler et al 1991].

The value most commonly used for iodine by the whole set of participants models was 1.E-2 d/l, i.e. a factor of four over the observed value. The six dynamical models we analysed show a more accurate Fm for iodine, with central values ranging between 1.E-3 d/l and 9.9E-3 d/l with a mean value of 3.E-3 d/l, which is roughly compatible with the observed variability. For caesium the mean value is 4.4E-3 d/l, almost the same as the observed mean value.

Figures 15 and 16 show the predicted versus the observed concentration ratio for <sup>131</sup>I and <sup>137</sup>Cs transfer from vegetation to milk in one steady conservative model as well as in the six selected dynamic best-estimate models. As can be observed from these figures, for air to milk transfer the performance of some conservative models is quite similar to the performance of the six selected models. However the accumulated errors in the air-vegetation subsystem make those models to overpredict systematically.

#### 4. Conclusions.

The fact that a steady state model ranks in second position in its performance in the prediction of integrated concentrations shows that complexity is not a necessity for many assessment purposes.

However, when a good fit of the daily concentrations is intended, dynamic models show a better performance than quasi steady models do.

The model SPADE 2 is quite similar in structure to one of the 6 selected models: the PRYMA model. However, a different choice of parameter values made the predictions of the latter model somewhat more accurate.

The volumetric distribution parameter W (ratio between the concentration of contamination in rain droplets and the concentration in air) that the 6 best models used is roughly compatible with the range reported by Coughtrey for Europe [Coughtrey 1990]: W from 200 to 800 (Bq/l)/(Bq/m<sup>3</sup>).

A general feature observed in most of the participant models in the A4 exercise which is also observed in three of the six selected models (CIRCLE for Iodine and RAGTIME and PRYMA for Caesium) is that they give good predictions for the total deposition and not so good for wet and dry deposition separately.

This feature illustrates the occurrence of compensatory errors in different subsystems of the model.

Some of the best models used a central value for  $v_g$  which is out of the observed range. However, this had no visible effect on their performance. This is due to the fact that the wet deposition was dominant over the dry deposition in most of the locations. This can be explained by the depth of the contaminated cloud of Chernobyl [Fulkner 1987].

The selected models use the Chamberlain formula or some similar approach that correlates the interception  $R$  and the yield factor  $Y$ .

The selected models use central values for the weathering decay time in the interval of the experimental variability ( $1d < \tau_{exp} < 35 d$  for caesium and  $7 d < \tau_{exp} < 14 d$  for iodine).

The selected models use dynamical expressions for daily deposition and weather washout and dynamical metabolic models for caesium and for iodine.

A good summary of the performance of the air to vegetation subsystem of each model is the comparison between the observed and obtained concentration ratios for the transfer from air to vegetation. The range of the air to vegetation transfer ratios obtained by the best models for iodine is from 0.61 to 1.8 the observed mean value. For caesium the range is from 0.4 to 1.1 the observed mean value.

However this transfer ratio is obtained through very different mechanisms in different models. For example, for  $^{137}\text{Cs}$ , some models are very sensitive to precipitation events, e.g. RAGTIME (see Table 2), and others are less sensitive (e.g. CIRCLE). However, in practice these structural differences only generate an effective transfer ratio which presents a relatively good covariation with the observed one.

The pasture to milk transfer factors ( $F_m$ ) derived from the six best models are roughly in the range of the observed variability for caesium and for iodine. This contrasts with the mean of the  $F_m$  used by the rest of models, which tends to overpredict. For iodine, this overprediction is as large as a factor of four.

### Acknowledgements

This work would not have been possible without the previous analysis and documentation of results carried out by Harry Köhler, Sylvie-Ring Peterson and F. Owen Hoffman during the A4 exercise of BIOMOVS. This study was partially realized in the frame of a contract between CIEMAT-IMA and ENRESA, in Spain.

### References.

- Coughtrey, P.J. and Thorne, M.C. 1983, "Radionuclide distribution and transport in terrestrial and aquatic ecosystems. A critical review of data", A.A. Balkema, Rotterdam 1983.

- Coughtrey, P.J. et al., 1990, "Evaluation of food-chain transfer data for use in accident consequence assessment", Proceedings of the Seminar and methods and codes for assessing the off-site consequences of nuclear accidents", Athens, pag. 175-208, Commission of the European Communities, DG XII-D-3 and DG XI-A-1.

- Fulkner, M.J. 1987, "Aspects of Environmental monitoring by British Nuclear Fuels PLC Following the Chernobyl Reactor Accident", J. Environmental Radioactivity 5, 235-244, 1987.

- García-Olivares, A. 1992, "Modelling Uncertainties in a Dynamical Environmental Model", Modelling Geo-Biosphere Processes, No. 1, Catena-Verlag, pp. 41-81.

- García-Olivares, A., Carrasco, E. and Suañez, A., 1993, "PRYMA-T0, A model of radionuclides transfer from air into foodstuff. Test with data from the Chernobyl accident", Editora CIEMAT, Madrid, 1993.

- Killough G.G., and Hoffman, F.O., "Validation of RAGTIME87 dynamic food-chain model against fallout data from the Chernobyl accident", J. Tenn. Acad. Sci. 63: 46, 1991.

- Kölher, H., Peterson, S.-R. & Hoffman, O. (eds.), 1991, "BIOMOVS Scenario A4, Multiple Model Testing using Chernobyl Fallout Data of I-131 in Forage and Milk and Cs-137 in Forage, Milk, Beef and Grain", Technical Report 13, Stockholm, March 1991.

- Pleasant J.C., McDowell-Boyer L.M., Killough G.G., "RAGTIME: A FORTRAN IV implementation of a time-dependent model for radionuclides in agricultural systems", ORNL/NUREG/TM-371, Union Carbide Corp., Nuclear Division, ORNL, TN, USA, 1980.

- Simmonds, J.R., "The influence of the year on the transfer of radionuclides to terrestrial foods following an accidental release to atmosphere". Chilton, NRPB-M121, 1985.

- Till, J.E. and Meyer, H.R. 1983, "Radiological Assessment", NUREG/CR--3332, US NRC, Washington.

**TABLE 1**  
**Ranking of model reliability**

MODEL	Integrated predictions "hits"	Geometric means "hits"	Transfer factors "hits"	Type of model
RAGTIME	32	8	10	Dynamic/best
PRYMA	35	7	7	Dynamic/best
SCHRAADLO	32	6	6	Steady/best
ECOSYS	23	9	9	Dynamic/best
FARMLAND	22	8	6	Dynamic/best
BILTH	19	8	8	Dynamic/best
CIRCLE	18	6	6	Dynamic/best

**TABLE 2**  
**Effective Distribution Coefficient (W)**  
**values used by the 6 selected models**  
**for caesium and iodine [(Bq/l)/(Bq/m<sup>3</sup>)]**

MODELS	W(Iele)	W(Iorg)	W(Ipar)	W(Cs)
ECOSYS	475	4.75	475	475
RAGTIME	80	0.4	600	600
CIRCLE	200	2.0	200	200
PRYMA	396	0	576	576
BILTH	396	0	576	576
FARMLAND	360	3.6	360	360

Iele, Iorg and Ipar designate the chemical forms of the iodine: elemental, organic and particulate

TABLE 3

Values of the deposition velocity (m/s)

MODEL	ECOSYS	RAGTIME	CIRCLE	PRYMA	FARMLAND	BILTH
Iele	1.0E-2	1.5E-2	3.5E-3	1.0E-2	1.0E-2	5.0E-3
Iorg	2.0E-4	3.0E-6	-	1.0E-4	1.0E-6	0
Ipar	2.0E-3	4.0E-4	-	1.0E-3	1.0E-3	2.0E-3
Cs	2.0E-3	4.0E-4	7.5E-4	2.4E-3	1.0E-3	3.0E-3

TABLE 4

R (interception factor) values assuming  
 $Y = 0.2 \text{ Kg d.v./m}^2$  and  $I = 500 \text{ mm/year}$

MODEL	R I WET	R I DRY	R Cs WET	R Cs DRY
ECOSYS	0.09	1.00	0.08	1.00
PRYMA	0.20	0.20	0.25	0.25
RAGTIME	0.43	0.43	0.43	0.43
CIRCLE	0.00	0.43	0.15	0.43
FARMLAND	0.43	0.43	0.43	0.43
BILTH	0.25	0.25	0.25	0.25

TABLE 5

Main rate of loss without radioactive decay (d-1)

MODEL	ECOSYS	RAGTIME	CIRCLE	PRYMA	BILTH	FARMLAND
Cs	5.8E-2	7.53E-2	4.95E-2	1.23E-1	2.3E-2	4.19E-2
I	8.7E-2	7.53E-2	1.26E-1	1.03E-1	1.0E-1	4.95E-2

TABLE 6

The models' first order differential equations take into account the following phenomena

PHENOMENA	ECOSYS	RAGTIME	PRYMA	FARMLAND	BILTH	CIRCLE
Daily deposition	X	X	X	X	X	X
Weathering washout	X	X	X	X	X	X
Growth dilution	X		X			X
Cattle consumption		X	X			
Resuspension	X		X	X		X
Leave absorption		X				
Radicular absorpt.	X	X	X	X	X	X

TABLE 7

Concentration ratio air to vegetation (m<sup>3</sup>/kg d.w.) arising from the models studied and concentration factor observed

	G.Mean (I)	G.STD (I)	G.Mean(Cs)	G.STD (Cs)
Observed Value	3448	1.7	14526	2.2
ECOSYS	4010	1.4	13082	1.4
PRYMA	3237	1.7	9331	1.7
RAGTIME	2282	1.5	10201	2.0
CIRCLE	2602	1.3	6847	2.4
BILTH	2513	1.7	5886	1.9
FARMLAND	6292	1.7	15351	1.9

G.Mean: Geometric Mean. G.STD: Geometric Standard Deviation

TABLE 8

Concentration ratio air to vegetation (m<sup>3</sup>/kg d.w.)  
calculated for the models studied assuming steady  
air concentration and a rain rate of 1.37 mm/d.

MODEL	CFwet I	CFdry I	CFwet Cs	CFdry Cs
ECOSYS	442 n	6752 n	4481	6405
PRYMA	2230 n	9015 n	7994	3410
	2000 t	1331 t		
RAGTIME	254 t	206 t	636	987
CIRCLE	0 t	5160 nu	6227	2841
BILTH	4021 n	1901 n	42750	9391
FARMLAND *	99485 n	10662 n	443340	4433

n: Neuherberg iodine speciation, t: Tokai iodine speciation  
nu: iodine speciation not used

\* FARMLAND uses a precipitation rate of 24 mm/d

TABLE 9

Ingestion-milk concentration factor F<sub>m</sub> (d/l)

MODEL	ECOSYS	RAGTIME	CIRCLE	PRYMA	BILTH*	FARMLAND
Cs	3.0E-3	4.0E-3	7.1E-3	4.0E-3	1.0E-3	7.1E-3
I	3.0E-3	2.0E-3	1.0E-3	1.0E-3	1.1E-3	9.9E-3

\* Values calculated with time-integrated results.

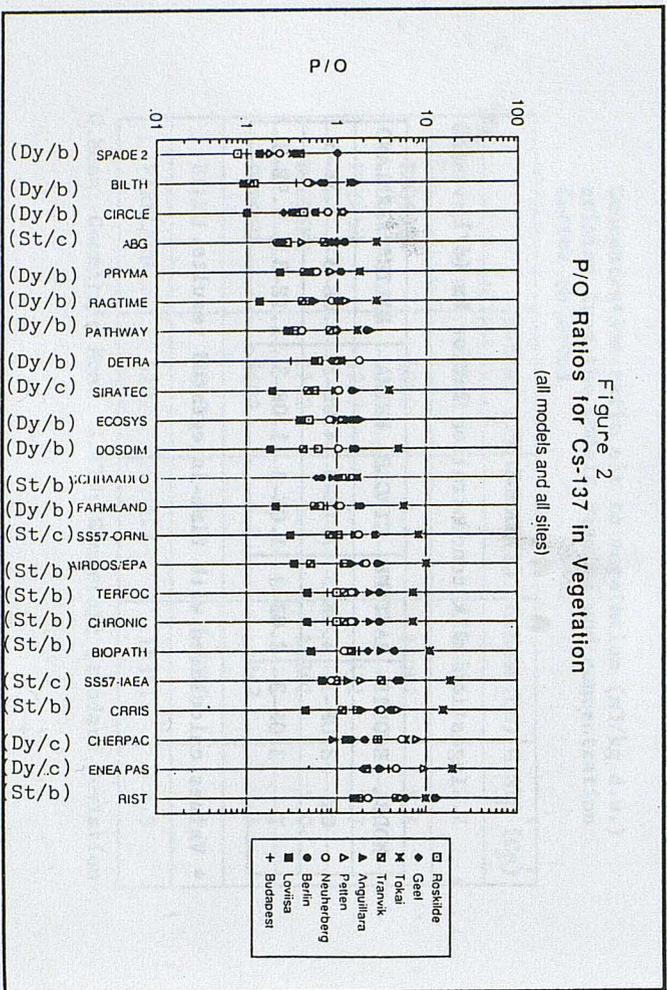
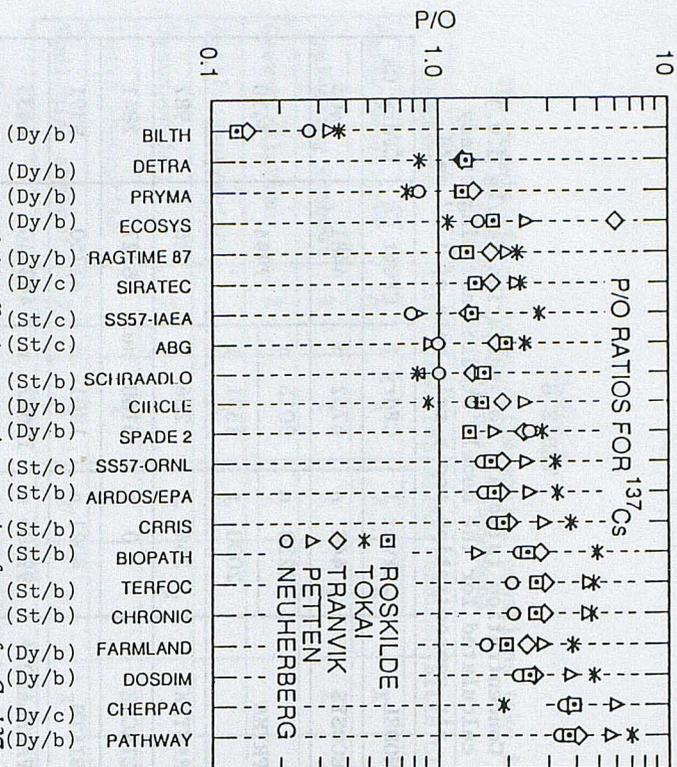
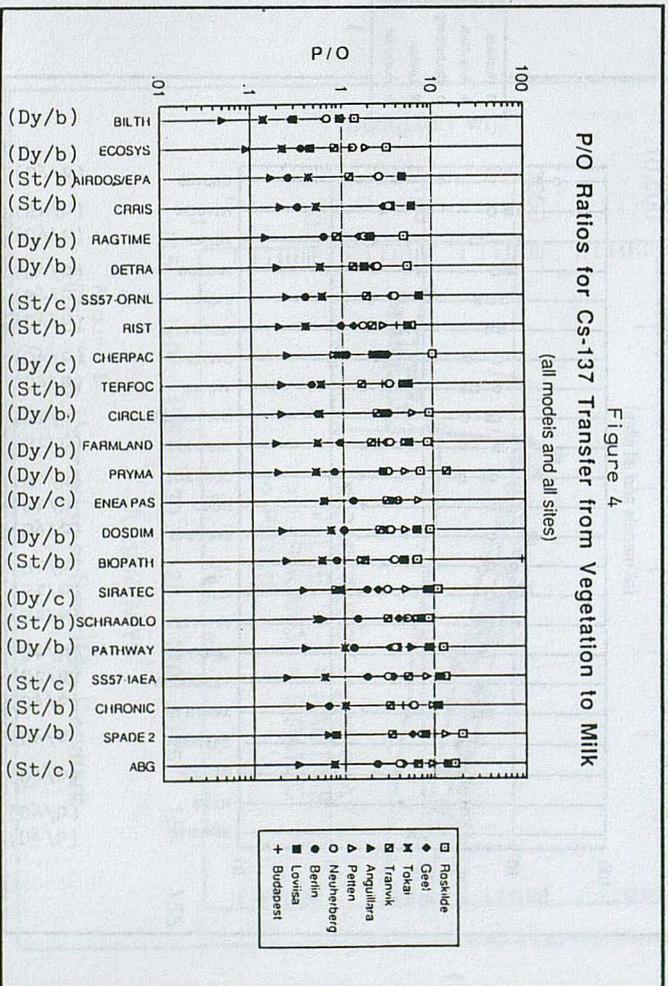
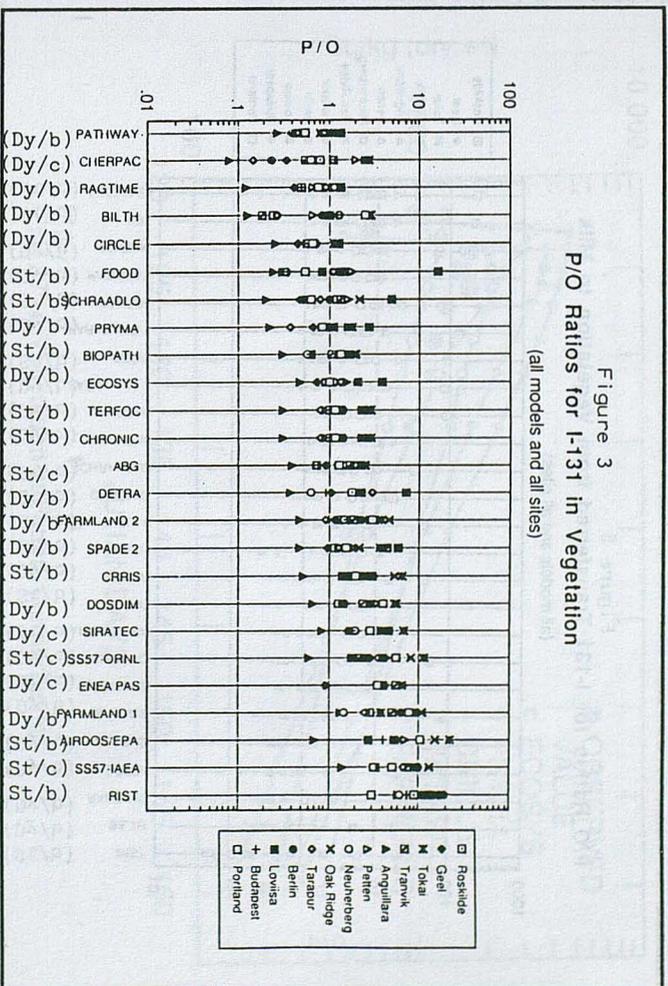


Figure 1 P/O ratios of the effective mass interception factors for Cs-137 calculated from model predictions.







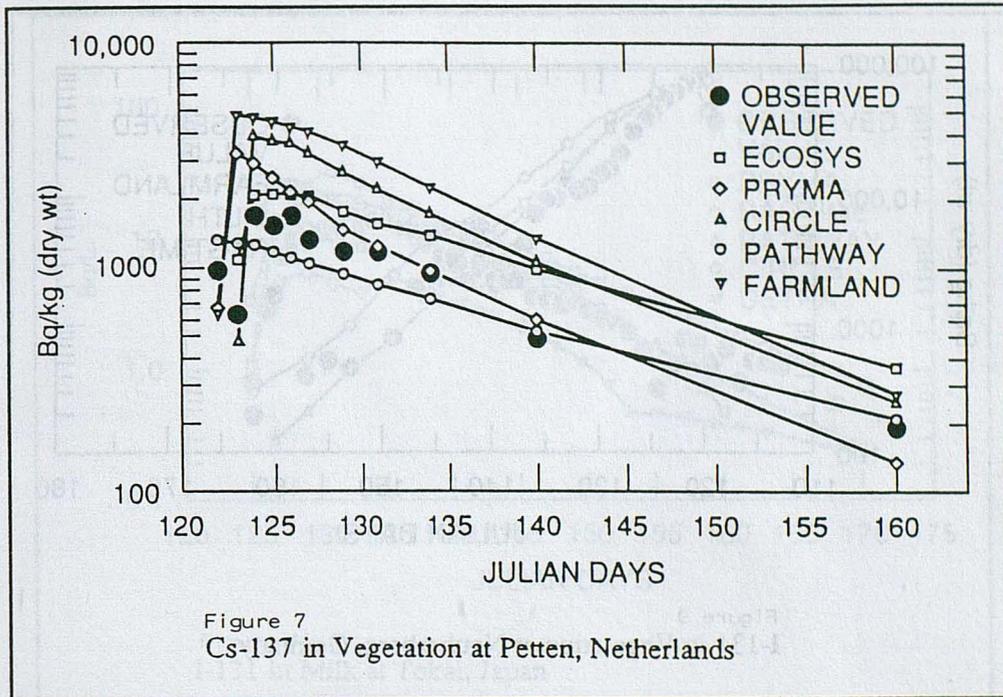


Figure 7  
Cs-137 in Vegetation at Petten, Netherlands

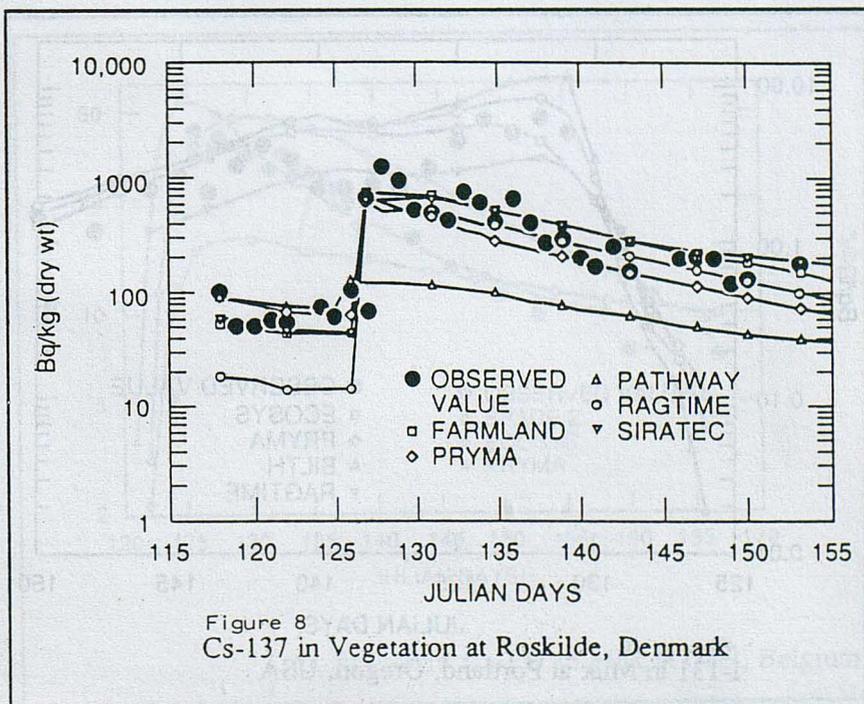
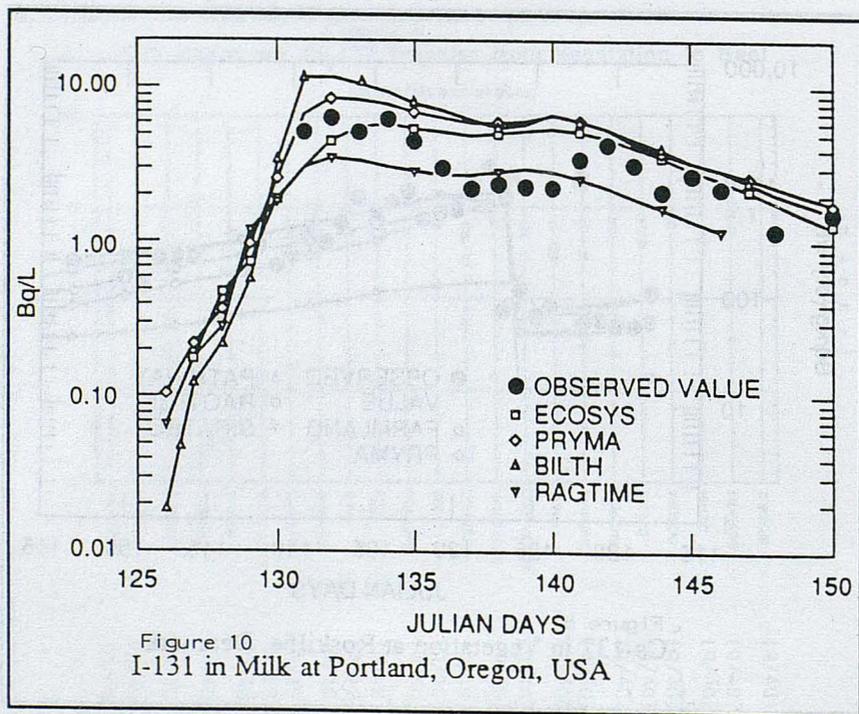
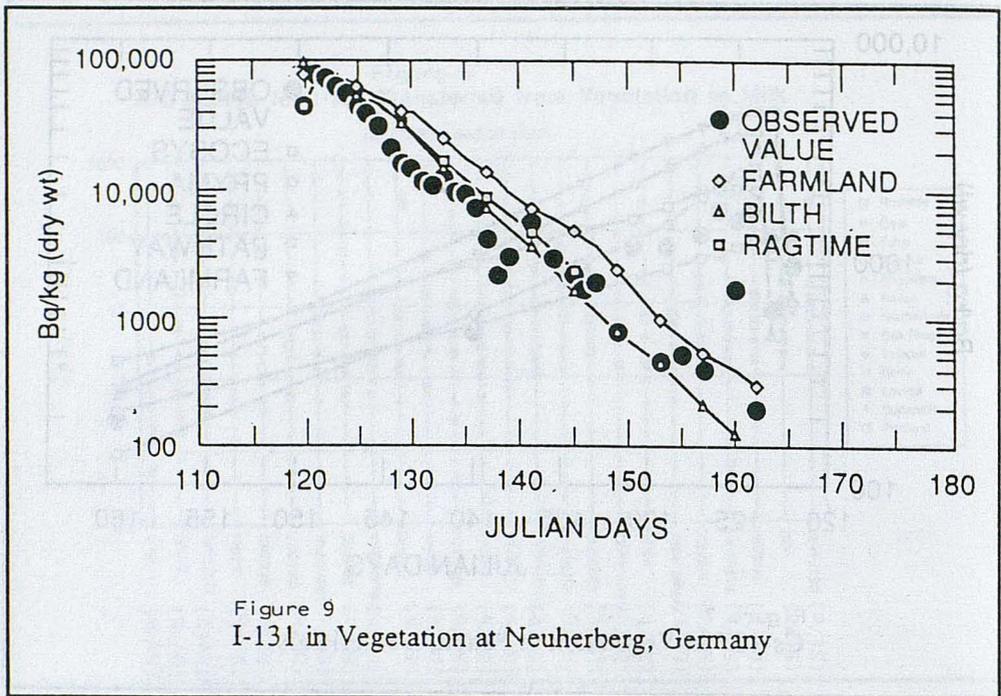
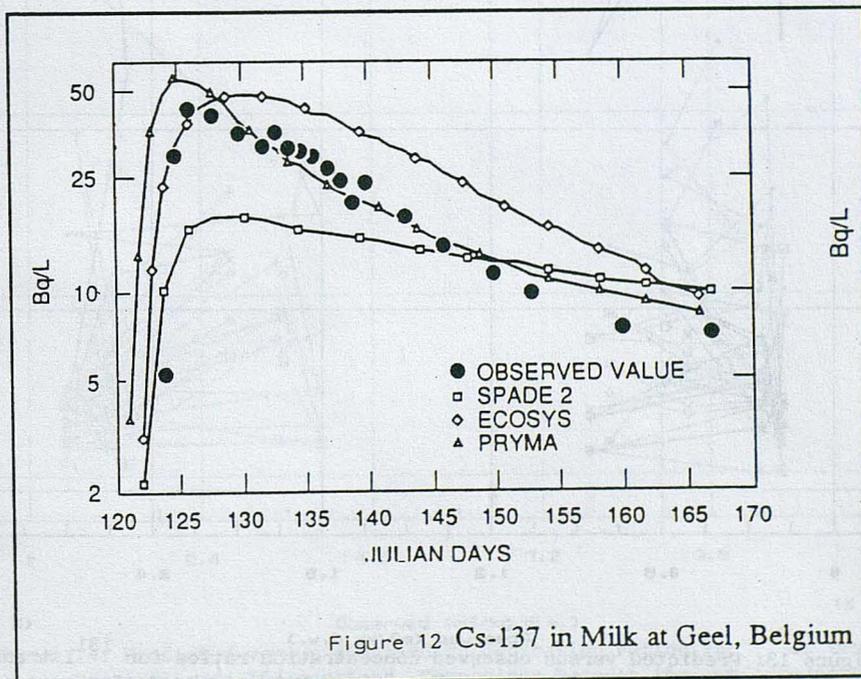
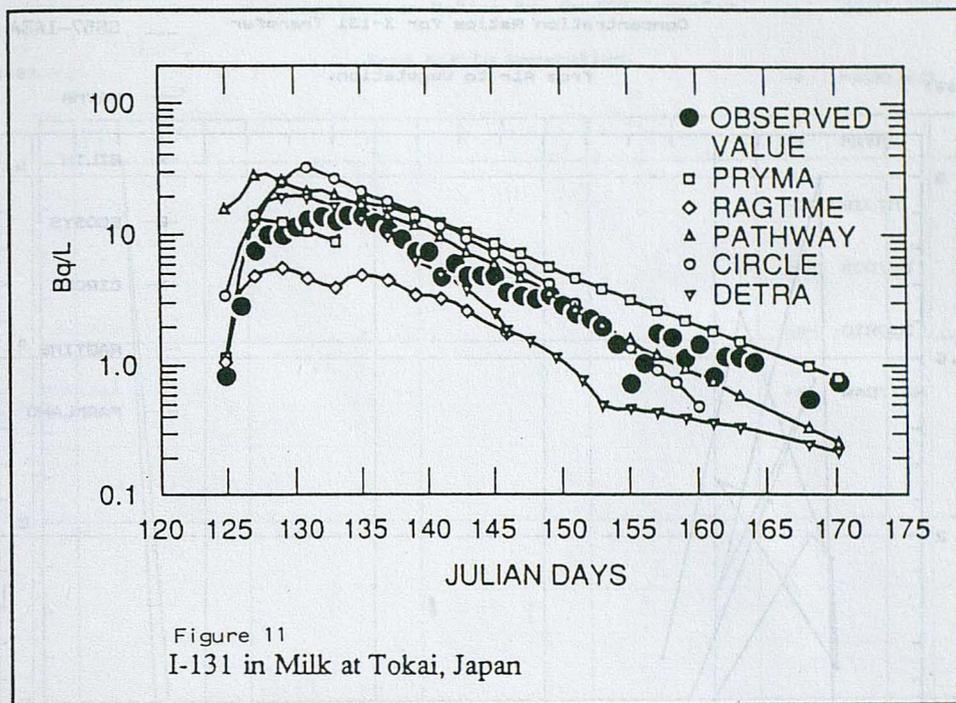


Figure 8  
Cs-137 in Vegetation at Roskilde, Denmark





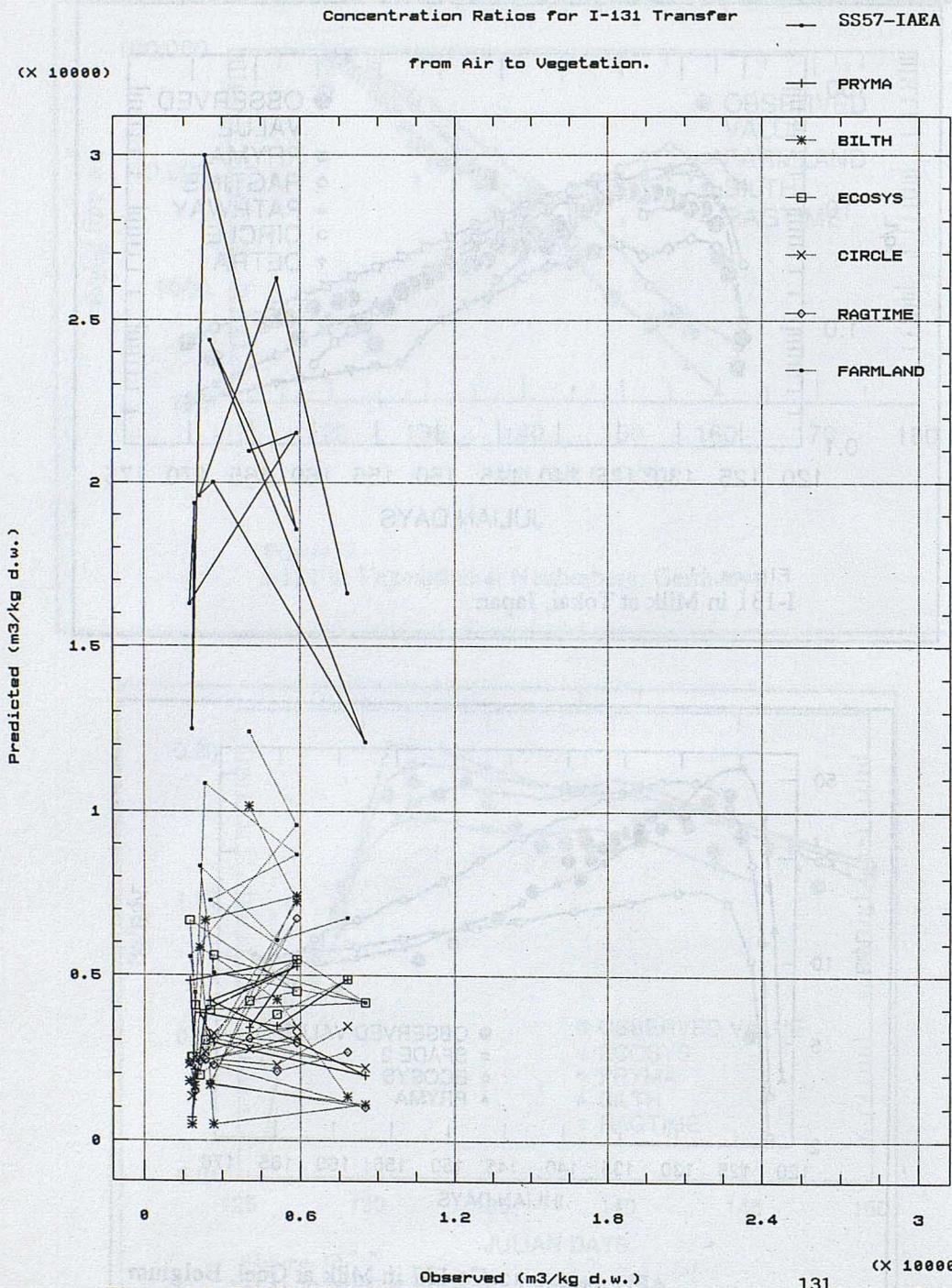


Figure 13: Predicted versus observed concentration ratios for  $^{131}\text{I}$  transfer from Air to Vegetation in 13 locations. Comparison between the performance of a steady conservative model (SS57-IAEA) and the six best-estimate dynamic models studied.

Concentration Ratios for Cs-137 Transfer

SS57-IAEA

(X 10000)

from Air to Vegetation.

+ FARMLAND

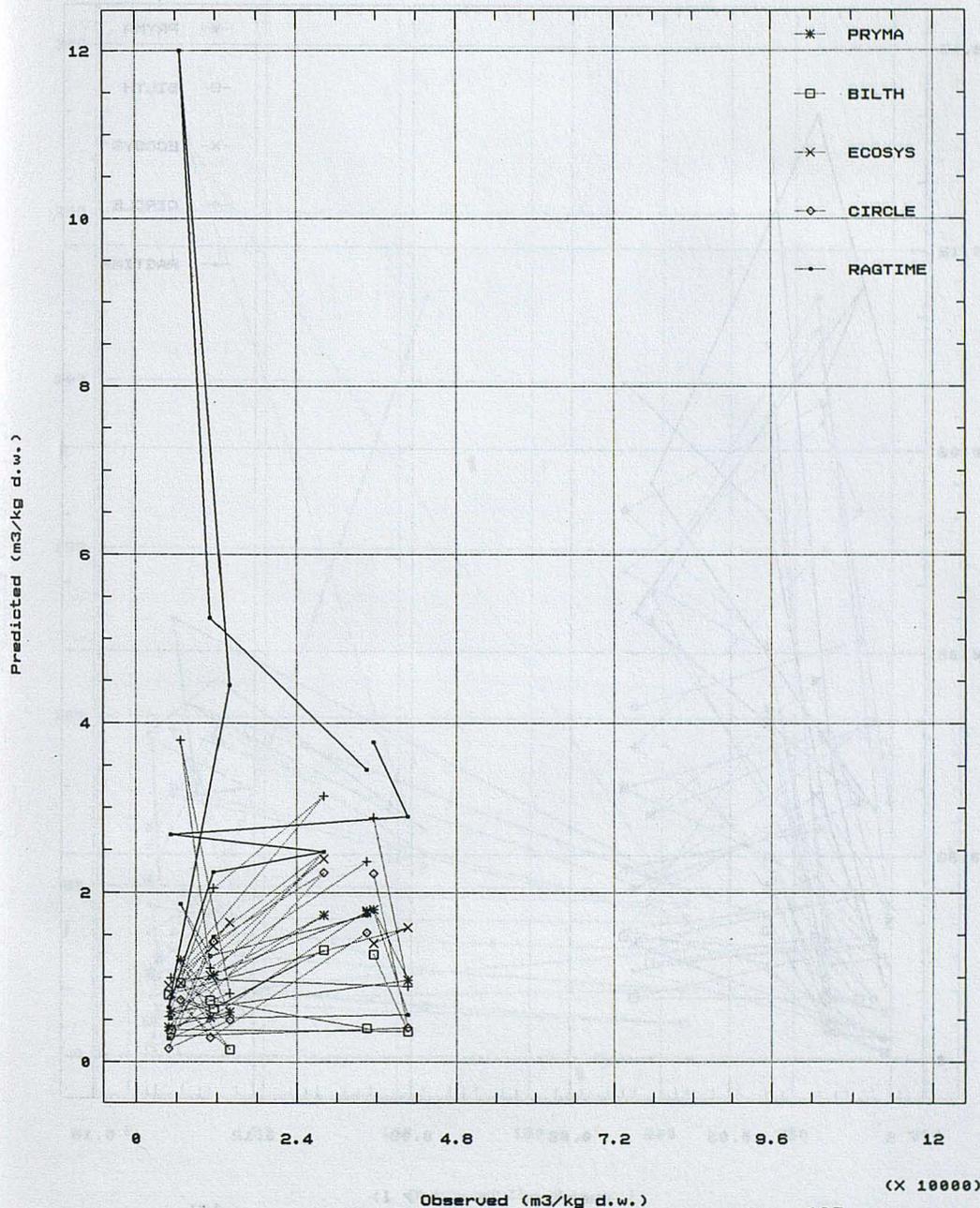


Figure 14: Predicted versus observed concentration ratios for <sup>137</sup>Cs transfer from Air to Vegetation in 10 locations. Comparison between the performance of a steady conservative model (SS57-IAEA) and the six best estimate dynamic models studied.

Concentration Ratios for I-131 Transfer

from Vegetation to Milk.

— SS57-IAEA

+ FARMLAND

\* PRYMA

□ BILTH

-X- ECOSYS

◇ CIRCLE

• RAGTIME

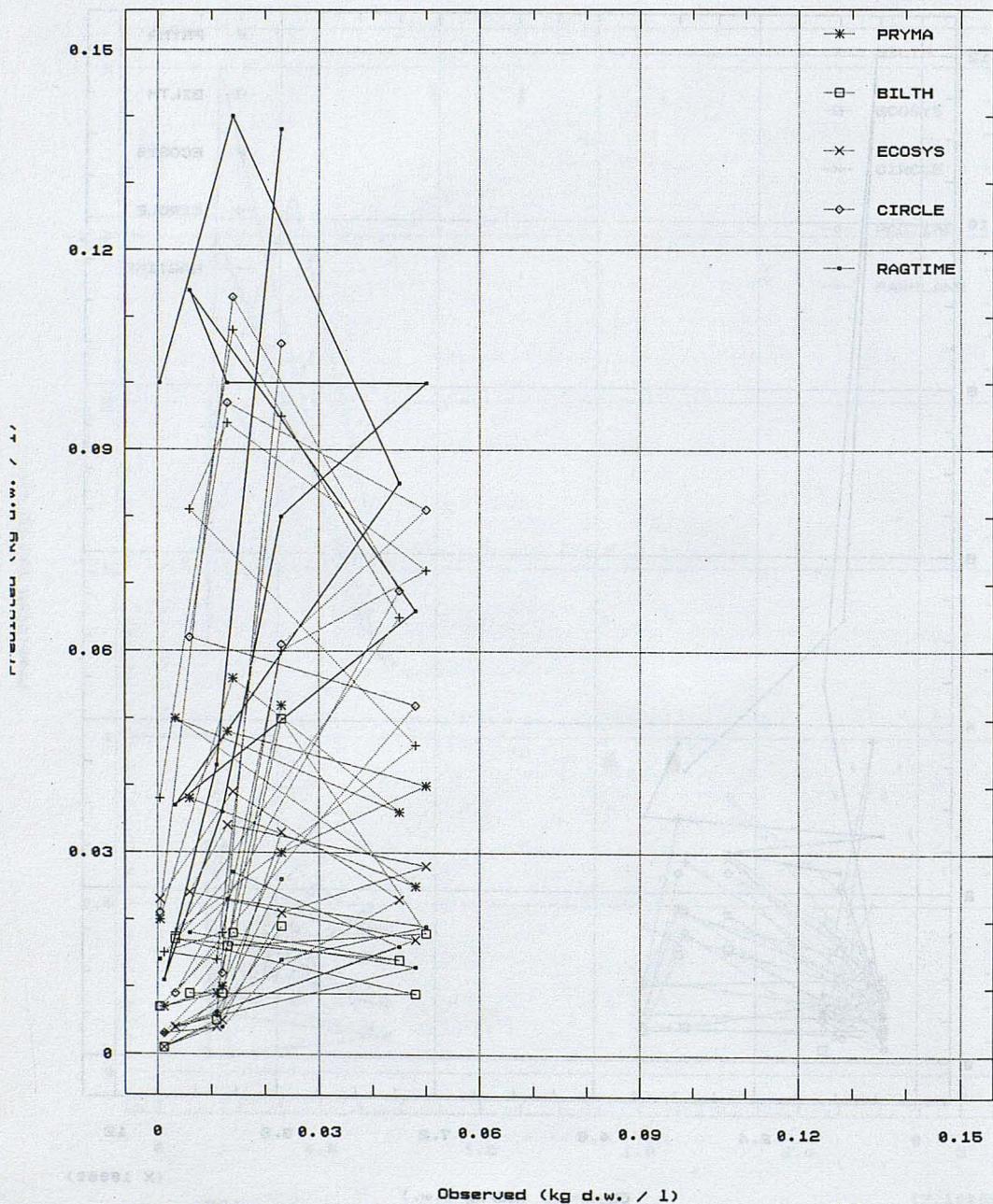


Figure 15: Predicted versus observed concentration ratios for  $^{131}\text{I}$  transfer from Vegetation to Milk in 13 locations. Comparison between the performance of a steady conservative model (SS57-IAEA) and the six best-estimate dynamic models studied.

Concentration Ratios for Cs-137 Transfer  
from Vegetation to Milk

SS57-IAEA

FARMLAND

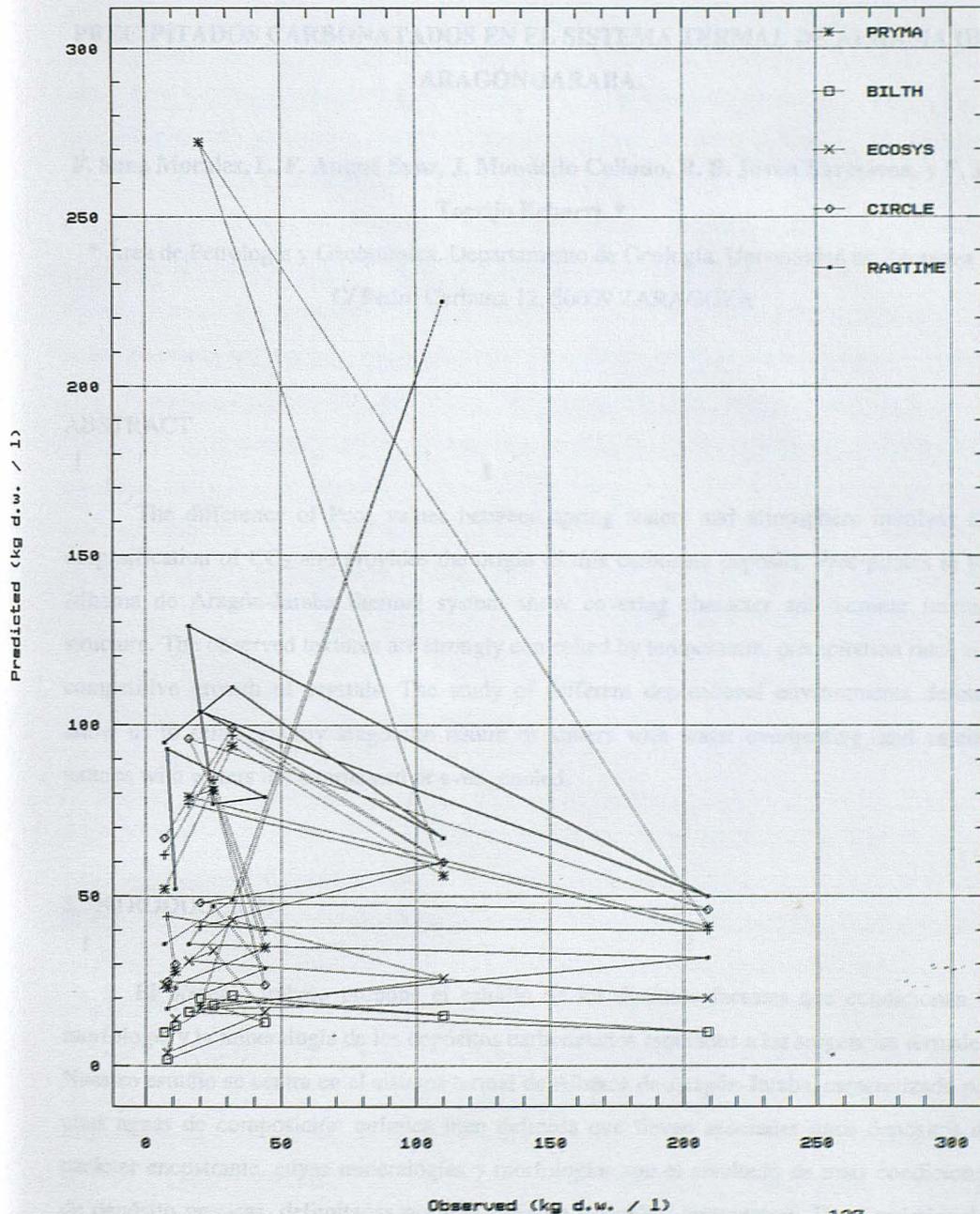


Figure 16: Predicted versus observed concentration ratios for  $^{137}\text{Cs}$  transfer from vegetation to milk in 10 locations. Comparison between the performance of a steady conservative model (SS57-IAEA) and the six best-estimate dynamic models studied.

**PRECIPITADOS CARBONATADOS EN EL SISTEMA TERMAL DE ALHAMA DE  
ARAGÓN-JARABA.**

**F. Sanz Morales, L. F. Auqué Sanz, J. Mandado Collado, R. B. Jovén Barcelona, y F. J.  
Torrijo Echarri. \***

\* Área de Petrología y Geoquímica. Departamento de Geología. Universidad de Zaragoza  
C/ Pedro Cerbuna 12, 50009 ZARAGOZA

**ABSTRACT**

The difference of  $P_{CO_2}$  values between spring waters and atmosphere involves the degasification of  $CO_2$  and provides the origin of this carbonate deposits. Precipitates in the Alhama de Aragón-Jaraba thermal system show covering character and laminar internal structure. The observed textures are strongly controlled by temperature, precipitation rates and competitive growth of crystals. The study of different depositional environments defined allow us to relate mainly aragonitic nature of sinters with water overheating, and calcitic natures with waters not overheated or even, cooled.

**1. INTRODUCCIÓN**

El presente trabajo propone el estudio de los distintos factores que condicionan la morfología y la mineralogía de los depósitos carbonatados asociados a las surgencias termales. Nuestro estudio se centra en el sistema termal de Alhama de Aragón-Jaraba, caracterizado por unas aguas de composición química bien definida que llevan asociadas unos depósitos de carácter encostrante, cuyas mineralogías y morfologías son el resultado de unas condiciones de depósito precisas, delimitadas por una serie de variables: temperatura,  $P_{CO_2}$ , química del agua, contexto hidrodinámico, que han sido elegidas como más representativas, y que son el objeto de este análisis.

## 2. SITUACIÓN GEOGRÁFICA Y GEOLÓGICA

La zona de estudio se sitúa próxima a las localidades de Alhama de Aragón y Jaraba, en un sistema termal de reconocidas propiedades balneoterápicas.

Desde el punto de vista geológico, se encuentra en el sector suroccidental de la Rama Castellana de la Cordillera Ibérica, en el contacto con la cuenca terciaria de Almazán. El claro control estructural de dirección Ibérica (noroeste-sureste) condiciona la disposición de los manantiales en dicha dirección, destacando como principales puntos de surgencia los presentes en Alhama de Aragón, Jaraba, Mochales, Cimballa, Deza y Embid de Ariza.

El modelo de funcionamiento hidrogeológico propuesto considera como área de recarga fundamental del sistema la definida por los materiales carbonatados liásicos predominantes en la Sierra de Solorio. Existen tres direcciones de circulación preferente: a) dirección E-O (Guadarrama) que funciona en el valle del río Mesa antes de Jaraba. b) dirección SO-NE: la del río Mesa, que recoge el flujo no drenado en Mochales y Cimballa y lo canaliza hacia los manantiales de Jaraba. c) dirección SE-NO (ibérica) a favor de la cual se produce un flujo convergente hasta la zona de menor cota en el valle del Jalón, a la altura de Alhama de Aragón, y que juega un importante papel en el termalismo de las aguas

## 3. METODOLOGÍA

a) *metodología de campo*: Se ha efectuado una recogida de muestras de agua representativa del conjunto de surgencias termales de la zona o del complejo balneoterápico en concreto, midiendo *in situ* parámetros tales como temperatura, pH y conductividad. Las muestras de agua se han recogido estabilizando el contenido en cationes mediante la adición de un 1% de HNO<sub>3</sub> suprapuro y otra submuestra no modificada para el análisis de aniones. En los casos en que es posible se recogieron muestras del depósito encostrante procurando obtener aquellas que no mostraban signos de presencia de materia orgánica (tapices, tonalidades verdosas, etc.). Pese a que no en todos los casos se ha podido disponer de una muestra de agua con su correspondiente depósito, al menos, sí se ha dispuesto de la muestra combinada para cada sector o complejo balneoterápico de la zona.

b) *metodología de laboratorio*: En un periodo de tiempo no superior a 24 horas se procede a medir temperatura, pH y alcalinidad, determinando además las concentraciones de los principales aniones y cationes mayores. Las metodologías analíticas utilizadas incluyen absorción atómica para la determinación de cationes e ICP para los menores y trazas. También se han realizado análisis morfológicos y mineralógicos mediante microscopía electrónica con EDAX, difracción de rayos X y microscopía petrográfica.

#### 4. GEOQUÍMICA DE LAS AGUAS TERMALES

Las características hidroquímicas de los manantiales termales de la zona responden a tipologías bicarbonatado-sulfatadas (Jaraba) y sulfatado-bicarbonatadas (Alhama de Aragón), ambas con caracteres comunes cálcico-magnésicos. Adicionalmente, en Alhama de Aragón se aprecian tendencias clorurado-sódicas.

En los manantiales de Jaraba se aprecian variaciones composicionales y de temperatura entre las distintas surgencias que inducen a pensar en procesos secundarios que modifican con intensidad variable el quimismo de la solución en profundidad. A priori puede pensarse en un proceso de mezcla de aguas o en fenómenos de disolución o precipitación. La verificación y caracterización del proceso de mezcla fue efectuada por Auqué (1993) mediante modelización en sistema cerrado con el código PHREQEE.

#### 5. FACTORES INVOLUCRADOS EN EL PROCESO DE PRECIPITACIÓN. REVISIÓN.

##### 5.1. Punto de partida

La situación de desequilibrio de la presión parcial de  $\text{CO}_2$  en las aguas respectó al valor atmosférico da lugar a una pérdida por desgasificación de  $\text{CO}_2$ . Este proceso lleva consigo un aumento del grado de sobresaturación del carbonato en solución, provocando así su precipitación. La ecuación que describe este proceso es la siguiente:



cuanto mayor sea la pérdida de  $\text{CO}_2$ , más rápidamente se alcanzará la sobresaturación en carbonato, hecho que tiene una importante incidencia en el hábito, la fábrica y mineralogía de los precipitados resultantes.

La premisa fundamental sobre la que se basa este estudio propone que durante la recogida de las muestras (de agua y depósitos), la composición del fluido es representativa de aquella bajo la que tuvo lugar la precipitación. Se ha verificado la estabilidad en la composición química de las aguas analizadas al contrastar los análisis con otros datos (Auqué, 1993). El intervalo de tiempo en el que se han formado los depósitos encostrantes, que comprende desde meses a pocos años, se engloba en el periodo de tiempo comprendido entre los primeros y los últimos análisis de las aguas, por lo que se confirma que éstas son suficientemente representativas y reflejan la composición química de los precipitados.

## 5.2. Factores que influyen en la tasa de precipitación y que afectan a la mineralogía de los depósitos .

La precipitación de calcita o aragonito depende de las variaciones temporales de la presión de  $\text{CO}_2$ , tal como mostraron Jones y Renaut (1996) en sistemas de alta entalpía. Esta idea se puede generalizar también para los sistemas de media-baja entalpía, como el que es objeto de este estudio, y así Given y Wilkinson (1985) proponen como factor fundamental para la precipitación de calcita o aragonito la tasa de aporte de  $\text{CO}_3^{=}$  sobre la superficie cristalina en crecimiento, de modo que altos índices de saturación favorecen la formación de aragonito y valores menores producen precipitados de calcita.

Además del índice de saturación hay otros factores que deben ser tenidos en cuenta. Así, Burton (1993) hace notar la importancia de factores tales como temperatura, contenido en  $\text{SO}_4^-$ , relación  $A_{\text{Mg}^{2+}}/A_{\text{Ca}^{2+}}$ , salinidad del sistema o contenido en amonio como favorecedores o inhibidores de la fase mineral específica que precipita, y considera la importancia de los aspectos cinéticos en el desarrollo de los cristales planteando el efecto de los factores

hidrodinámicos sobre la mineralogía mediante el control del aporte de iones a la superficie cristalina en crecimiento.

Todos estos aspectos permiten darnos cuenta de que la precipitación de una u otra mineralogía carbonatada está controlada por numerosos factores que influyen en el control cinético o termodinámico del proceso de precipitación, por lo que se plantea la imposibilidad de separar un único parámetro que haga referencia a un determinado aspecto de los precipitados, como su hábito, fábrica o mineralogía.

De forma pormenorizada, los factores que influyen en la mineralogía de los depósitos son:

\* *La temperatura:* La tasa de precipitación de aragonito respecto a calcita se incrementa fuertemente con el aumento de la temperatura, y no está afectado en gran medida por cambios en el estado de saturación (Burton y Walter, 1987). La temperatura influye en la mineralogía de los precipitados de calcita, aumentando la concentración de  $MgCO_3$  que se incorpora al cristal conforme ésta aumenta, hecho que tiene una importante implicación termodinámica, ya que el aumento de la concentración de  $MgCO_3$  (>4% molar) en calcita aumenta su solubilidad, lo que incidiría en el control cinético de su precipitación respecto a aragonito.

\* *Presencia de agentes inhibidores, como  $SO_4^{2-}$ ,  $Mg^{2+}$  y  $PO_4^{3-}$ :* Los iones sulfato y magnesio presentes en el medio funcionan como agentes inhibidores en la precipitación de calcita ya que al incorporarse en la estructura aumentan significativamente su solubilidad. Esto provoca que sean necesarios mayores valores de sobresaturación para que comience la precipitación. (He y Morse, 1993). El ión fosfato también funciona como agente inhibidor de la formación de calcita al nuclearse en las caras de ésta impidiendo su crecimiento.

\* *Viscosidad del medio:* Influye en la tasa de precipitación del carbonato debido a que modifica la tasa de difusión iónica (Buczynski y Chafetz, 1991). En medios menos viscosos, la tasa de difusión iónica es alta y los iones son transportados rápidamente a los puntos de nucleación del cristal. Esta rápida precipitación favorece la formación de aragonito respecto a calcita.

\* *Salinidad*: Estudios realizados por Zhong y Mucci (1989) demuestran que la tasa de precipitación de aragonito aumenta con la salinidad para un amplio rango estudiado (5-40%) mientras que no hay cambios apreciables en la tasa correspondiente a la calcita.

## 6. HÁBITO Y FÁBRICA DE LOS PRECIPITADOS.

El hábito de los precipitados depende en primer lugar del grado de saturación, así como de factores inherentes a su estructura cristalina, aunque también hay que tener en cuenta el efecto que produce la competencia cristalina sobre el mismo (Jones y Renaut, 1996).

La fábrica de los depósitos espeleotémicos estudiados muestra un crecimiento progresivo del tamaño de los cristales desde el sustrato, que indicaría un descenso del grado de saturación, y en consecuencia un crecimiento más lento, y una clara direccionalidad perpendicular al sustrato, marcado fundamentalmente por la competencia cristalina.

La tasa de aporte de reactantes controla la velocidad de crecimiento cristalino, y puede modificarse mediante variaciones en el grado de sobresaturación o bien mediante cambios en la velocidad de flujo (González *et al.*, 1992). Esta última variable no había sido tenido en cuenta y permite explicar un cambio en la fábrica de los espeleotemas para el caso hipotético de que el grado de sobresaturación se mantuviese constante.

La velocidad de flujo también condiciona la rugosidad del sustrato y la mayor o menor participación de nucleación heterogénea en el proceso de precipitación (Rodríguez-Clemente, 1982).

## 7. EL SISTEMA TERMAL DE ALHAMA DE ARAGÓN-JARABA. DELIMITACIÓN DE VARIABLES.

Los factores que intervienen en el proceso de precipitación son, por tanto, muchos, y no hay motivo para pensar que no puedan interactuar todos juntos. Pero si nos centramos

ahora en el sistema concreto que es objeto de este estudio, podemos realizar algunas simplificaciones:

Partimos de unas aguas termales que están en desequilibrio con la atmósfera, y que llevan asociados en los puntos de surgencia depósitos carbonatados encostrantes. De acuerdo con Chafetz y Folk (1984), que proponen una distribución espacial del predominio del proceso inorgánico y del orgánico, el estudio se centró en los puntos de surgencia, donde el proceso inorgánico es dominante, y de esta forma, evitar de forma sistemática la presencia de actividad orgánica que introduce un mayor número de variables al problema. De acuerdo con un medio con poca influencia orgánica, la viscosidad del medio no constituye un factor significativo, por lo que no se tendrá en cuenta.

Los análisis químicos realizados muestran un muy bajo contenido en fosfatos por lo que la influencia de este factor también puede obviarse.

Ya que se trata de aguas termales, el efecto de la temperatura es significativo y favorece la precipitación, implicándose en el proceso de desgasificación, factor que se considera como el desencadenante del proceso. Así mismo, el aumento de la temperatura de las aguas, de manera artificial, ha condicionado el predominio de unas mineralogías aragoníticas en los depósitos encostrantes encontrados en el interior de las tuberías por las que circula el agua sobrecalentada.

También se han considerado como relevantes las condiciones hidrodinámicas de cada microambiente en particular; es decir, si en el proceso de precipitación del carbonato, el mecanismo que favorece la sobresaturación así como el aporte del carbonato ha sido: goteo, flujo continuo, evaporación, flujo confinado, ascenso de temperatura, mezcla de aguas, agitación mecánica, etc.

Son estas dos últimas variables las que nos han permitido plantear los diferentes ambientes genéticos y el tipo de morfologías y mineralogías de los depósitos a que dan lugar.

## 8. LOS DEPÓSITOS ENCOSTRANTES DEL SISTEMA TERMAL DE ALHAMA DE ARAGÓN-JARABA.

### 8.1. Medios de depósito. Ubicación y relación genética.

Existe una gran variedad de medios en los que se produce la precipitación de estos depósitos encostrantes de carbonato: fuentes interiores, pequeñas cascadas naturales, acequias, paredes de piscinas y lagos, fondo de bañeras de las galerías de baños, grifos y tuberías de conducción del agua termal, etc.

De todos esos ambientes podemos extraer una sencilla clasificación de los diferentes condiciones de precipitación y su relación con las morfologías y mineralogías encontradas:

a) Precipitación en la interfase agua-aire en condiciones de absoluto reposo. La desgasificación de  $\text{CO}_2$  es lenta pero significativa, dando lugar a un entramado de pequeños cristales de naturaleza calcítica que generan laminaciones de espesor micrométrico, que se mantienen en equilibrio en la superficie del agua.

En el contacto entre la interfase agua-aire y el sustrato (pared) también tiene lugar la precipitación. Pequeñas variaciones del nivel del agua dan lugar a un precipitado de laminación característica y naturaleza calcítica.

b) Desgasificación de  $\text{CO}_2$  a partir de una solución que fluye débilmente por mecanismo de goteo, que da lugar a una lenta precipitación y a la génesis de estructuras verticalizadas condicionadas por efecto de la gravedad. Corresponden a los desarrollos estalactíticos, cónicos y tubulares. Su mineralogía es predominantemente calcítica, aunque también está presente aragonito así como yeso secundario.

c) Precipitación rápida en condiciones de importante agitación, favorecida por la caída libre del agua y choque con el sustrato. Los precipitados generados presentan una gran dureza y una estructura interna laminar compacta muy característica. La naturaleza de los precipitados es calcítica.

d) Condiciones de flujo confinado al interior de una tubería y circulación a cierta velocidad. La cinética del flujo y la improbabilidad de que la sección sea colmatada totalmente favorecen la desgasificación de  $\text{CO}_2$  y la consiguiente precipitación de carbonato. Se generan así láminas que tapizan el interior de la sección. La intermitencia del flujo en este ambiente permite que se desarrollen laminaciones o zonaciones durante el crecimiento del precipitado, así como superposición de éstos.

Se han encontrado mineralogías aragoníticas y calcíticas. Es posible que además del control cinético del flujo de agua exista un control térmico que puede correlacionarse con unas u otras mineralogías, ya que se ha verificado que en algunas tuberías, el agua que circulaba fue previamente recalentada. Se ha establecido una relación entre los desarrollos de cristales aciculares de naturaleza predominantemente aragonítica con un aumento de temperatura de las aguas en circulación (conducción en el Balneario de Sicilia, Jaraba), así como los precipitados de naturaleza calcítica con aguas no recalentadas, e incluso enfriadas (conducción de agua hacia la localidad de Embid de Ariza).

La diferente naturaleza de las tuberías observadas (PVC, plomo, cobre, hierro) no condiciona el que exista o no una precipitación, aunque influirá en la velocidad con que se generan estos depósitos.

e) Condiciones de flujo libre de la corriente con importantes variaciones del nivel del agua que permiten, de forma temporal, la exposición de los precipitados a condiciones subaéreas. Se han encontrado formas mamelonares y agregados de aspecto arbustivo. La naturaleza de estos depósitos es calcítica.

## 8.2. Rasgos morfoestructurales y texturales de los precipitados.

La morfología externa de los precipitados es muy variada. Se han observado formas botroidales, arbustivas y mamelonares, desarrollos estalactíticos, laminaciones micrométricas, bandeados fibrosos, etc. La naturaleza y tipo de sustrato sobre el que se generan estos depósitos condiciona en gran medida esta amplia variedad de formas.

Dentro de esta variedad de formas posibles, se pueden diferenciar una serie de rasgos comunes en cuanto a la estructura interna de estos precipitados y los hábitos de los cristales que los constituyen y su relación con el sustrato sobre el que se desarrollan.

Se ha establecido una clasificación de las texturas observadas en los depósitos mediante análisis con lupa binocular, microscopía petrográfica convencional y microscopía electrónica:

- Textura laminar, constituida por desarrollos de pequeños cristales calcícos de hábito romboédrico, escalenoédrico a lenticular. El gran número de impurezas presentes en el medio dificultan el normal crecimiento cristalino, dando lugar, en ocasiones, a una importante porosidad intrapartícula.
- Textura fibrosa, presente en los recubrimientos internos de tuberías y conducciones, constituida por agrupaciones de cristales aciculares de naturaleza predominantemente aragonítica, aunque también están presentes cristales calcícos. En ocasiones, estos últimos presentan hábitos tabulares. Esta textura está asociada a un sobrecalentamiento de las aguas por encima de su temperatura termal natural.
- Textura dendrítica o esferulítica, constituida por agrupaciones de cristales aciculares de naturaleza predominantemente calcítica, que adoptan una disposición en abanico. Los cristales aciculares que forman parte de esta textura presentan segmentaciones, hecho que ha descartado la posibilidad de que sean cristales de aragonito, aunque es posible que coexistan ambas mineralogías.

Según Jones y Renaut (1996) esta textura nos indicaría el desarrollo de cristales en condiciones de fuerte desequilibrio. Se localiza en la envuelta externa de los desarrollos estalactíticos, asociada a signos de presencia de materia orgánica y en depósitos enconstrantes asociados a aguas fuertemente recalentadas (65°C).

- Los desarrollos estalactíticos presentan una estructura interna que responde a dos etapas de desarrollo diferenciadas (Mathews *et al.*, 1991). El menor grado de sobresaturación y un consiguiente crecimiento cristalino más lento caracterizan la segunda etapa de crecimiento de estas estructuras, dando lugar a una textura de tipo drúscico, constituida por cristales visibles a simple vista, con caras cristalinas bien definidas, y que crecen con una orientación común marcada por la competencia cristalina.

- Texturas masivas o ausencia de texturas cristalinas reconocibles, generadas por crecimientos fuertemente competitivos o influencia de la materia orgánica sobre los carbonatos en crecimiento, desarrollándose agregados amorfos sobre los filamentos algales.

Podemos establecer como conclusión, que todos los depósitos estudiados tienen un carácter encostrante y presentan una estructura interna laminar más o menos definida, al margen de las peculiaridades morfológicas y texturales de cada precipitado.

### 8.3. Mineralogía

En Jaraba se ha aludido a un proceso de mezcla de las aguas termales con aguas meteóricas más frías (Aiqué, 1993). Esto implica una disminución de la temperatura en los puntos de surgencia, así como un menor desequilibrio de la  $P_{CO_2}$  entre las aguas de mezcla y la atmósfera, lo que conducirá a que la desgasificación sea menos enérgica. Según los datos aportados por los autores mencionados anteriormente, esto favorecerá la precipitación de calcita frente a la de aragonito, hecho que se comprueba con los datos de análisis de rayos X realizados para algunos precipitados de carbonato de características tobáceas, en los que se observa el predominio de una naturaleza calcítica en Jaraba y una existencia conjunta de calcita y aragonito en Alhama de Aragón.

La precipitación preferente de aragonito sugiere la posibilidad de que la salinidad sea un factor relevante en el proceso de precipitación para la zona de estudio. El aumento de la concentración de sulfato y  $Mg^{2+}$  que conlleva este aumento progresivo de la salinidad a lo largo de la pauta de flujo, podrían ser también factores relevantes, dado el carácter inhibitorio de estos agentes sobre la calcita, como ya se explicó en apartados anteriores.

La relación entre los valores de concentración de Mg y Sr en las muestras sólidas, así como la relación entre éstos y la temperatura, nos permite establecer una conexión entre las mineralogías encontradas y los medios de depósito correspondientes:

- Los precipitados asociados a aguas fuertemente sobrecalentadas (70°C), incluyendo los medios donde las aguas estaban confinadas al interior de una tubería, son de naturaleza predominantemente aragonítica y presentan los valores más altos en Sr y más bajos en Mg.

- Los precipitados asociados directamente a puntos de surgencia son de naturaleza calcítica (LMC) y presentan los valores más altos en Mg y más bajos en Sr.

- Los precipitados asociados a medios donde el agua presenta una fuerte agitación pero no un sobrecalentamiento, presentan naturaleza calcítica (LMC), con valores intermedios para las concentraciones de Mg y Sr.

## 9. CONCLUSIONES

Se han detectado precipitados carbonatados de carácter encostrante en todos los puntos de muestreo asociados a aguas termales, dado el notable desequilibrio entre éstas y la atmósfera.

No podemos establecer ningún tipo de generalización entre las localidades de Alhama de Aragón y Jaraba si intentamos englobar todos los microambientes estudiados correspondientes a todos los puntos de muestreo, pues como se indicó al inicio, en algunos de ellos han sido alterados algunos parámetros, como es la temperatura. Pese a que ello no nos permite establecer generalizaciones, podemos interpretar que:

El incremento en la temperatura de las aguas, una mayor desgasificación condicionada por la cinética del sistema, y la influencia de la salinidad, especialmente el contenido en Mg, son los tres factores principales que favorecen el predominio de naturalezas aragoníticas en los precipitados.

Variaciones en el grado de saturación por mezcla con aguas más frías propicia el desarrollo de mineralogías calcíticas.

Finalmente, la naturaleza y características del sustrato condicionan la morfología externa de estos precipitados.

## 10. AGRADECIMIENTOS

Este trabajo ha sido realizado dentro del Proyecto de Investigación PCB01/93, financiado por la Diputación General de Aragón.

Los autores quieren agradecer a la Dirección y personal técnico de los balnearios de Termas Pallarés, Guajardo, San Roque y Cantarero de Alhama de Aragón, y La Virgen y Sicilia en Jaraba, las facilidades proporcionadas para la toma de muestras de aguas y precipitados.

## 11. BIBLIOGRAFÍA

Auqué, L.F. (1993) *Estudio de sistemas geotermales en Aragón. Pautas de especiación y reacción aplicadas a la modelización de sistemas de baja-media entalpía*. Área de Petrología y Geoquímica del Dep. de Geología. Universidad de Zaragoza. Tesis Doctoral. 509 pp.

Buczynski, Ch. and Chafetz, H. (1990) Habit of bacterially induced precipitates of calcium carbonate and the influence of medium viscosity on mineralogy. *J. Sediment. Petrol.*, **61**(2), 26-233.

Burton, E. (1993) Controls on marine carbonate cement mineralogy: review and reassessment. *Chem. Geology*, **105**, 163-179.

Burton, E. & Walter, L.M. (1987) Relative precipitation rates of aragonite and Mg calcite from seawater: Temperature or carbonate ion control?. *Geology*, Vol. **15**, 111-114.

Chafetz, H. and Folk, R. (1984) Travertines: Depositional morphology and the bacterially constructed constituents. *J. Sediment. Petrol.*, **54**(1), 289-316.

Given, K. & Wilkinson, B.H. (1985) Kinetic control of morphology, composition and mineralogy of abiotic sedimentary carbonates. *J. Sediment. Petrol.*, **55**(1), 109-119.

González, L., Carpenter, S. and Lohmann, K. (1992) Inorganic calcite morphology: Roles of fluid chemistry and fluid flow. *J. Sediment. Petrol.*, **62**(3), 382-399.

He, Sh. and Morse, J.W. (1993) The carbonic acid system and calcite solubility in aqueous Na-K-Ca-Mg-Cl-SO<sub>4</sub> solutions from 0 to 90°C. *Geoch. Cosmoch. Acta*, **57**, 3533-3554.

Jones, B. and Renaut, R. (1996) Morphology and growth of aragonite crystals in hot-spring travertines at Lake Bogoria, Kenya rift Valley. *Sedimentology*, **43**, 323-340.

Matthews, M., Matthews, A. and Ayalon, A. (1991) Environmental controls of speleothem mineralogy in a karstic dolomitic terrain (Soreq Cave, Israel). *Jour. of Geology*, **99**, 189-207.

Rodríguez-Clemente, R. and Sunagawa, I. (1982) Crystal growth processes in sedimentary environments: *Estudios Geológicos*, v. **38**, 125-377.

Zhong, S. and Mucci, A. (1989) Calcite and aragonite precipitation from seawater solutions of various salinities: Precipitation rates and overgrowth compositions. *Chem. Geology*, **78**, 283-299.

## **NÓDULOS Y CONCRECIONES. MODELOS GENÉTICOS.**

**F. J. Torrijo Echarri, J. Mandado Collado, R. B. Joven Barcelona y F. J. Sanz Morales\***

\* Área de Petrología y Geoquímica. Departamento de Geología. Universidad de Zaragoza. C/Pedro Cerbuña, 12. 50009-Zaragoza.

### **Abstract**

The genesis of nodules and concretions is a controversial point in the scientific bibliography, without a generic model accepted. This paper presents an intent of grouping the existing models over the theme to clear the existing controversies in the scientific literature.

We are made a identification of distinct models to group them in the principal genetic types defined, according to the moment of formation of the nodule or concretion, and according to the mechanism of growth, considering also a third group that would complete the two prior ones.

### **Introducción**

Los nódulos y concreciones son prácticamente omnipresentes en casi todas las rocas sedimentarias del registro geológico y su identificación, estudio y valoración ha dado lugar a un sinnúmero de controversias científicas aún no resueltas, sin que pueda establecerse, por tanto, un modelo universalmente admitido para cada tipo de nódulo o concreción.

El objetivo básico de los trabajos realizados sobre nódulos y concreciones es fundamentalmente científico, por cuanto el estudio de estas estructuras facilita la reconstrucción de la historia diagenética de las rocas en que aparecen. El conocimiento del mecanismo de movilización

de los materiales constituyentes del nódulo o concreción, momento de emplazamiento y factores genéticos, son sumamente útiles para la reconstrucción del entorno y condicionantes diagenéticos de una formación geológica, permitiendo su aplicación en temas tan diversos como paleoclimatología (Yemane y Kelts, 1996), hidrogeología (Wilkinson, 1993), explotación de gas e hidrocarburos (Vysotskiy y Sidorenkov, 1987), explotación de uranio (Meschter, 1958), etc.

### **Modelos genéticos**

Antes de comenzar a describir los distintos modelos genéticos que se han dado para explicar el crecimiento de cuerpos concrecionales hay que remarcar que un modelo simple no sirve para todos los tipos de crecimiento concrecionario, y que, seguramente, combinando dos o más modelos conseguiremos obtener, con bastante seguridad, el modelo de crecimiento para cada tipo de nódulo o concreción.

Vamos a sistematizar el estudio de los modelos recopilando en primer lugar los definidos según el momento de formación, centrando después la atención en aquellos en los que el parámetro fundamental considerado es el mecanismo de crecimiento; finalmente, consideramos en un tercer grupo otros modelos complementarios de los anteriores, aunque no claramente inscritos en ninguno de los dos tipos genéricos.

1. Modelos según el momento de formación: Son aquellos que consideran la relación del cuerpo concrecional con la roca encajante.

1.1. Modelos de crecimiento parasedimentario: Consideran que el cuerpo concrecional se ha generado fuera de la roca encajante, y por ello no tiene ninguna relación genética con ésta. Según nuestro punto de vista estas estructuras no deben ser consideradas como auténticos nódulos o concreciones al tener un mecanismo de formación previo a la sedimentación.

Este modelo encaja perfectamente para las denominadas por Pettijohn (1975) como "pseudoconcreciones": bolas armadas (*armored mud balls*), bolas lacustres (*lake balls*), y oncoides.

1.2. Modelos de crecimiento sinsedimentario: Postulan que el crecimiento concrecionario se ha producido durante el depósito del sedimento que dará lugar a la roca encajante. Presentan claramente una marcada relación genética entre ambos procesos. Dentro de este modelo se incluyen las concreciones calcáreas en ciertas arcillas glaciares (*Imatra stones*), algunos nódulos de sílex en rocas carbonatadas y sulfatadas, algunos nódulos piriticos en lutitas, arcillas y rocas carbonatadas, nódulos y concreciones yesíferas en arcillas y material evaporítico, algunas

concreciones con fósiles, nódulos fosfatados en materiales detríticos y algunas concreciones carbonatadas en lutitas.

1.3. Modelos de crecimiento precompactacional: Defienden que el cuerpo concrecional ha crecido después del depósito del sedimento que dará lugar a la roca encajante, pero su crecimiento ha finalizado antes de un grado de compactación perceptible. Dentro de ellos se describen: concreciones carbonatadas en lutitas y arcillas, concreciones carbonatadas en arenas y limolitas (*Kugelsandstein* y *sand crystals*), concreciones manganesíferas en arcillas, algunos nódulos de sílex en rocas sulfatadas y carbonatadas, nódulos piríticos en lutitas, arcillas y rocas carbonatadas, concreciones con fósiles, nódulos de yeso y/o anhidrita en material detrítico y concreciones carbonatadas en material silíceo.

1.4. Modelos de crecimiento sincompactacional: Consideran que, por lo menos, una parte importante del crecimiento concrecionario es contemporáneo en relación con la compactación del sedimento que dará lugar a la roca encajante. Ejemplos de este modelo son: nódulos carbonatados en areniscas, concreciones carbonatadas en arcillas, nódulos y concreciones ferruginosas en rocas detríticas, y nódulos de cuarzo en material detrítico (Navas y Tena, 1987).

1.5. Modelos de crecimiento postcompactacional: Son aquellos que consideran que el crecimiento del cuerpo concrecional es posterior a la completa compactación del sedimento que dará lugar a la roca encajante. No presentan, por tanto, relación genética. En este modelo se incluyen nódulos de sílex en rocas carbonatadas, oolitos y peloides en calcretas de origen pedogénico, y nódulos y concreciones ferruginosas en rocas detríticas.

2. Modelos según el mecanismo de crecimiento: Son aquellos que tratan el crecimiento concrecionario desde el punto de vista del modo de emplazamiento.

2.1. Modelo de crecimiento por relleno de porosidad: El crecimiento concrecionario se produce por la precipitación y el relleno de huecos del sedimento (Lipmann, 1955) (ver figura 1C). Dentro de este marco se pueden dar dos situaciones:

- 1) Los huecos están originados por la porosidad inicial de la roca encajante.
- 2) Los huecos son debidos a la existencia de una porosidad secundaria, creada por la disolución previa del mineral preexistente, y en la cual se produce la precipitación.

El primer autor que propuso este modelo fué Lippman en 1955. Este autor sugiere un modelo de crecimiento para concreciones carbonatadas en el cual la precipitación de carbonato ocurre en los espacios de los poros del sedimento, sin sufrir cambios significativos de tamaño; por tanto, defiende que la porosidad inicial del sedimento durante el crecimiento

concrecionario podría ser calculada a partir de la relación volumétrica entre minerales carbonatados y no carbonatados dentro de la concreción. En 1970, Girin demuestra que existe una distribución regular en el contenido de ciertos elementos, entre el centro y el margen de la concreción. Posteriormente, Raiswell (1971 y 1976) asentó definitivamente las bases de este modelo:

- Las concreciones crecen por cementación en los huecos del sedimento produciéndose un relleno de porosidad.

- Las variaciones en el contenido de cemento carbonatado entre el centro y el margen de la concreción reflejan los cambios en la porosidad del sedimento durante el crecimiento de la concreción.

- El crecimiento concrecionario puede ocurrir tanto en sistemas abiertos como cerrados al agua marina. Datos basados en la porosidad estimada, contenido en  $\text{Sr}^{2+}$  y mineralogía podrían indicar la naturaleza del sistema.

- La morfología de las concreciones depende de la difusión, mecanismo de flujo y permeabilidad del sedimento.

El modelo se completa en 1982 con las aportaciones de Gautier. Este autor estudia los posibles errores que pueden plantearse a la hora de estimar la porosidad del sedimento. Propone que añadidos a los errores analíticos (generalmente menores al 2 %) hay cuatro efectos que pueden complicar la relación entre el porcentaje volumétrico de carbonato y la porosidad inicial del sedimento:

- Los cristales de carbonato pudieron haber desplazado al sedimento durante el crecimiento concrecionario.

- La porosidad podría no haber sido completamente llenada por la precipitación de carbonato.

- Los granos detríticos pudieron haber sido reemplazados por carbonato durante la formación de la concreción.

- Los granos carbonatados detríticos o biogénicos originales del sedimento no pueden diferenciarse analíticamente de los depósitos concrecionales e incrementan, por lo tanto, el porcentaje volumétrico aparente de carbonato autigénico medido.

Más recientemente, Wilkinson (1991) obtiene unos resultados que apoyan claramente la hipótesis de que la permeabilidad controla la formación de horizontes concrecionarios.

2.2. Modelo de crecimiento reemplazante: Los cuerpos concrecionales se emplazan por reacciones químicas entre la solución madre y la roca encajante. Estos crecen reemplazando

cementos tempranos o por corrosión y reemplazamiento del armazón de granos de la roca encajante (Van Tuyl, 1918) y Hennessy y Knauth, 1985 entre otros muchos). En este modelo de crecimiento toma gran relevancia la difusión en estado sólido. Básicamente es un modelo para situaciones en las que tenemos en juego sílice y carbonato (ver figura 1D) y tiene lugar como respuesta a variaciones de pH en las aguas intersticiales (Correns, 1950).

Dentro de este modelo existe una variante denominada "de reemplazamiento reversible" entre sílex y minerales carbonatados (Walker, 1962). Su causa no está clara, pero Walker (1962) sugiere que este emplazamiento puede llegar a darse cuando las fluctuaciones de pH se producen en medios altamente alcalinos (por encima de pH 9), además de poder estar influenciado por variaciones en la temperatura de los fluidos intersticiales. Los efectos de la temperatura, sin embargo, deberían tener menor importancia en la explicación de este reemplazamiento reversible en rocas que no han sufrido un acusado enterramiento. El mismo autor, en 1960, comenta que valores de pH por encima de 9 no son comunes en la naturaleza, aunque en ese mismo año, Garrels muestra que el agua pura en equilibrio con  $\text{CaCO}_3$  tiene un valor teórico de pH entre 9.9 y 10 y que valores del mismo orden de magnitud pueden obtenerse experimentalmente.

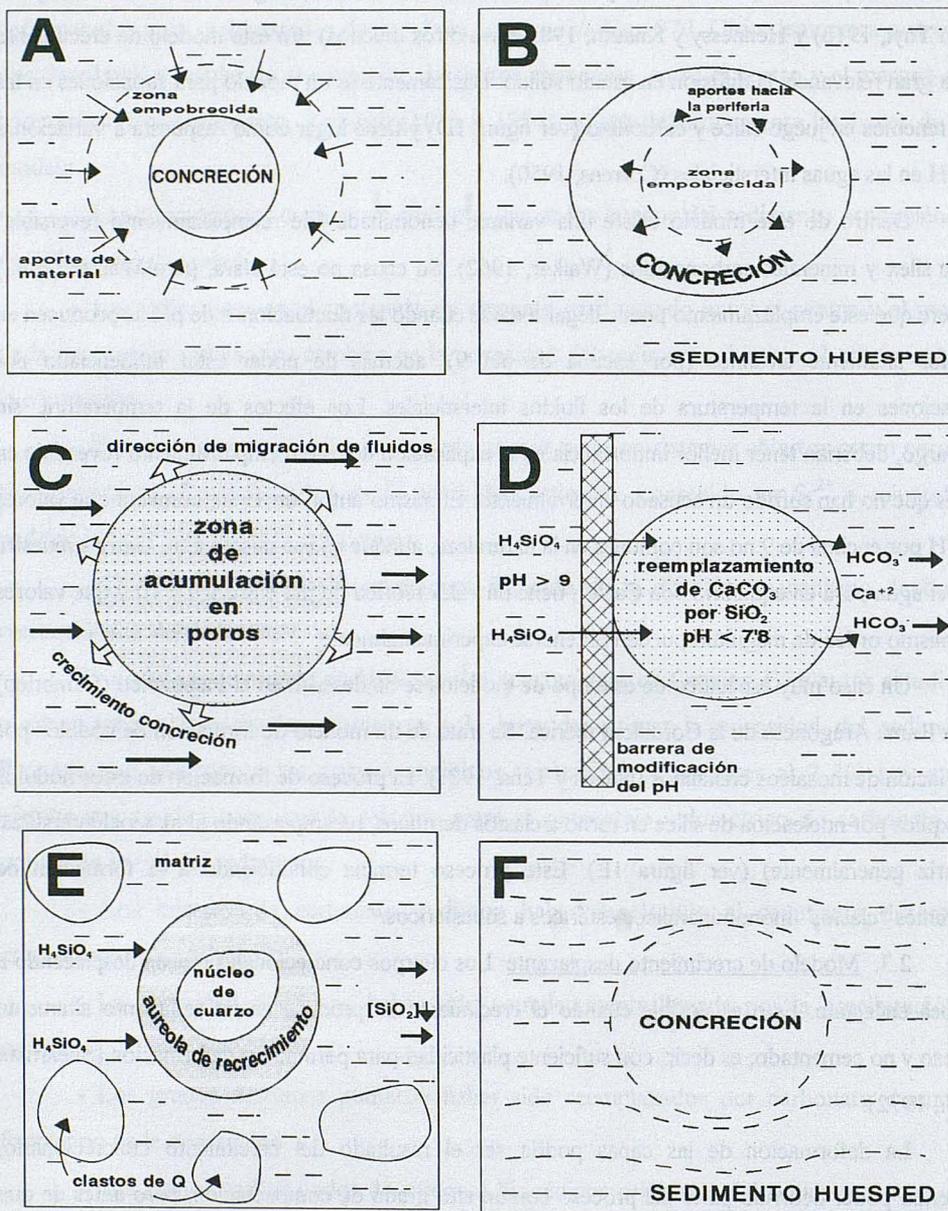
Un caso muy específico de este tipo de modelos se ha descrito en el Paleozoico (Cámbrico) de la Rama Aragonesa de la Cordillera Ibérica. Se trata de un modelo de formación de nódulos por asociación de mosaicos cristalinos (Navas y Tena, 1987). El proceso de formación de estos nódulos se explica por nucleación de sílice en torno a clastos detriticos, reemplazando al material intersticial (matriz generalmente) (ver figura 1E). Este proceso termina conduciendo a la formación de aparentes "clastos" monominerálicos esféricos a subsféricos.

**2.3. Modelo de crecimiento desplazante:** Los cuerpos concrecionales crecen desplazando a la roca encajante. Esto es posible cuando el crecimiento se produce en un sedimento altamente poroso y no cementado; es decir, con suficiente plasticidad para permitir su deformación (Shearman *et al.*, 1972).

La deformación de las capas podría ser el resultado del crecimiento concrecionario, habiendo poder ocurrido parte del proceso con un alto grado de compactación, pero antes de que quedara completamente rígido (Sellés-Martínez, 1996) (ver figura 1F).

La fuerza necesaria para desplazar el sedimento es denominada por numerosos autores como fuerza de cristalización (Dewers y Ortoleva, 1990).

**3. Otros modelos de crecimiento:** Son modelos que nos ayudan a completar los anteriores, por ello no los hemos incluido en ninguno de ellos.



**Figura 1:** (A) Modelo de crecimiento en equilibrio. (B) Modelo de crecimiento concéntrico convencional. (C) Modelo de crecimiento por relleno de porosidad. (D) Modelo de crecimiento reemplazante para concreciones silíceas y carbonatadas. (E) Modelo de crecimiento reemplazante de Navas y Tena (1987). (F) Modelo de crecimiento desplazante.

3.1. Modelo de crecimiento concéntrico convencional: El crecimiento se produce por adición de material desde el centro al exterior del cuerpo concrecional (ver figura 1B); por ello, las zonas más tempranas se sitúan en el centro y las zonas más tardías cerca del límite exterior. Varias evidencias sostienen este modelo:

- La desviación de las láminas dentro y alrededor de algunas concreciones sugiere que la porción central de la concreción fué cementada antes que las porciones exteriores (ver figura 2A) (Raiswell, 1971).

- El aumento sistemático en la cantidad de material detrítico desde el centro al exterior sugiere que el grado de compactación experimentado previo a la cementación fué menor en el centro que en los bordes (Criss *et al.*, 1988).

- Los cambios en la composición química de carbonato desde el centro al exterior son coherentes con el crecimiento concrecionario, éstos reflejan cambios en la química de las aguas intersticiales durante la formación del cuerpo concrecionario (Mozley y Burns, 1993).

- En algunas concreciones pueden observarse evidencias macroscópicas de la estructura concéntrica convencional, tales como zonación interna concéntrica o cambios bruscos en la mineralogía, tamaño cristalino y habito (Scotchman, 1991).

3.2. Modelo de crecimiento concéntrico complejo: Es uno de los modelos más recientes y propone que el crecimiento y la estructura de algunas concreciones carbonatadas, que a priori se explicarían con el modelo anterior, podrían no deberse exclusivamente a éste, y que la adición de material desde el centro al borde de la concreción sufriría pequeñas variaciones que darían como resultado la existencia de zonaciones internas complejas (Mozley, 1996). La zonación química compleja observada ocurre solo a muy pequeña escala (comunmente 10 mm o menos entre zonas) (ver figura 2B).

Este modelo sólo es valido para concreciones encontradas dentro de rocas carbonatadas, ya que no se ha observado zonaciones internas complejas en las encontradas en areniscas. Esta observación tiene dos posibles explicaciones: no existe zonación compleja en concreciones ubicadas en areniscas, o existe pero no se ha observado porque hasta la fecha sólo hay unos pocos estudios detallados de estas concreciones (McBride, 1988).

Este modelo surge por la observación en algunas concreciones carbonatadas de zonaciones de pequeña escala, incompatibles con el modelo anterior (Hennessy y Knauth, 1985), realizada con fluorescencia azul / UV y petrografía de alta resolución, los cuales comienzan a ser actualmente técnicas comunes (la zonación interna compleja es muy difícil o imposible de observar usando microscopios de luz plana y polarizada estandar).

3.3. Modelo de crecimiento en equilibrio: En este modelo el aporte de material se produce hacia el centro generándose de esta manera una zona empobrecida en estos materiales alrededor del nódulo o la concreción (ver figura 1A).

La zonación mineralógica radial es explicada comúnmente como un reflejo de la variación temporal de la química de las aguas intersticiales durante el crecimiento concrecionario (Berner, 1969 y Gautier, 1982, por ejemplo).

En el caso de nódulos y concreciones carbonatadas, su distribución regular podría explicarse representando los caminos de migración de las aguas intergranulares, y observando donde se produce el depósito aislado de carbonato, en forma de nódulo o concreción, ya que este depósito reflejaría la supersaturación de las aguas intergranulares móviles (Raiswell, 1971).

3.4. Modelo de crecimiento dinámico: Este modelo pretende explicar la zonación mineral de algunas concreciones carbonatadas, las cuales presentan solamente un control espacial. Según este modelo, la precipitación de carbonato se produce en lugares extremadamente confinados, mientras la disolución de carbonato se produciría en otros lugares.

Coleman y Raiswell (1995) definen este modelo en concreciones carbonatado-piríticas, planteando para su génesis que las posiciones relativas de las zonas supersaturadas de calcita y pirita vienen determinadas por los rangos relativos de reducción de sulfato, los rangos de adición por difusión y las pérdidas entre el microambiente y las aguas intersticiales que le rodean, y los rangos de precipitación de calcita y pirita (ver figura 2C). Estos son controles cinéticos y no necesitan producir la coincidencia de zonas de precipitación de calcita y pirita.

Según estos autores existen unas observaciones que presentan dificultades para encajar con el modelo, ya que éste necesita:

- Un ambiente intergranular muy específico. Este ambiente sólo se ha encontrado en un rango muy estrecho dentro de ambientes modernos ricos en materia orgánica.
- Una considerable cantidad de sulfuro de estado de oxidación intermedio ( $S^0$ ). El mecanismo exacto para llegar a este estado de oxidación sigue sin esclarecerse.
- Una fuente de materia orgánica que sea realmente metabolizable para obtener reducción de sulfatos en el lugar del crecimiento concrecionario. Las evidencias de este proceso no son observables dentro de las concreciones.
- Una fuente de hierro continua y abundante. Las fases más reactivas del hierro debieron ser, probablemente, utilizadas antes del crecimiento concrecionario.

3.5. Modelo de crecimiento concrecionario sincompactacional (OUC model): Es el más reciente y se podría aplicar para los modelos de crecimiento de relleno de porosidades primarias

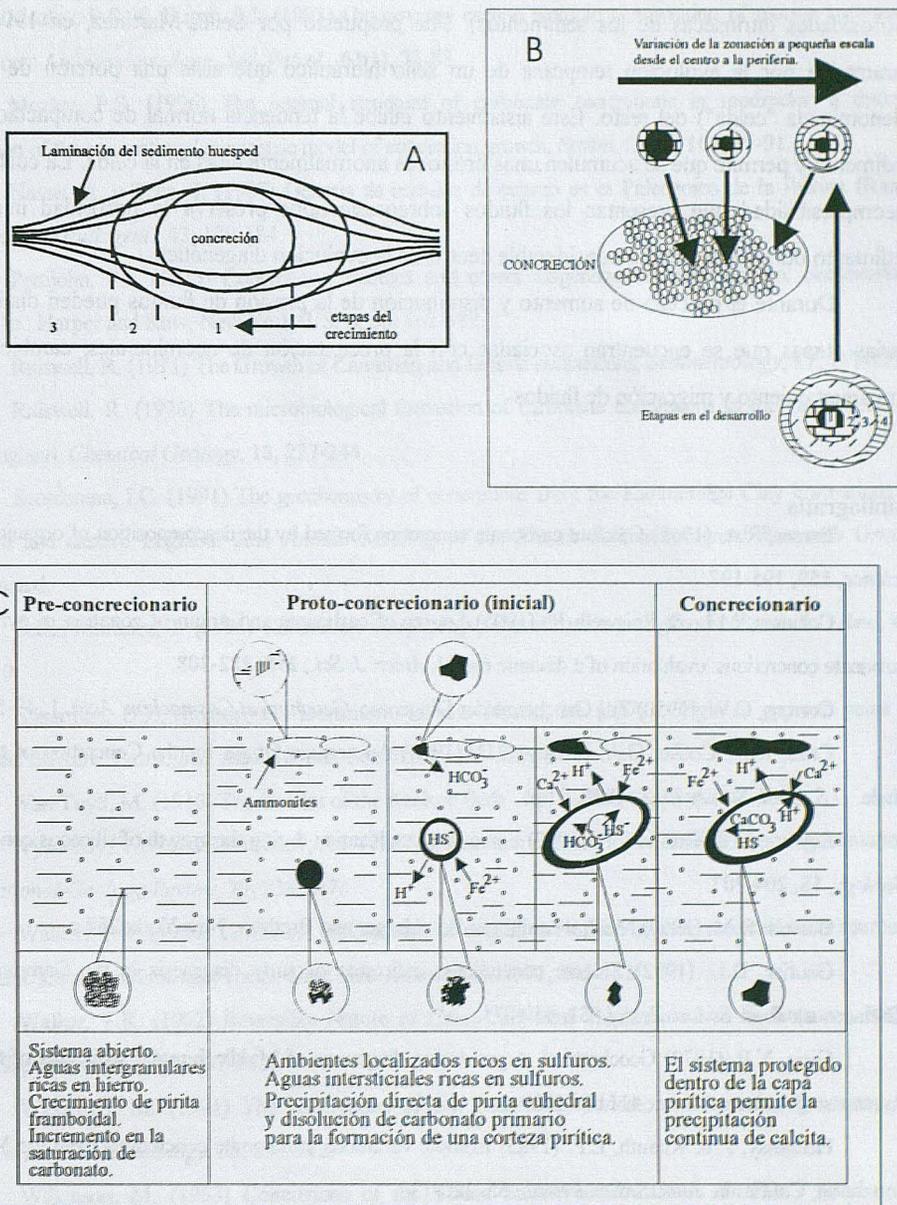


Figura 2: (A) Etapas en el crecimiento de un cuerpo concrecional que presenta una estructura laminar deformada (modificada de Raiswell, 1971). (B) Modelo de crecimiento concéntrico complejo mostrando una hipotética zonación química de pequeña escala en una concreción carbonatada (modificada de Mozley, 1996). (C) Modelo de crecimiento dinámico.

(porosidades intrínsecas de los sedimentos). Fue propuesto por Sellés-Martínez, en 1996, y se caracteriza por la evolución temprana de un sello hidráulico que aísla una porción de cuenca (denominada "celda") del resto. Este aislamiento inhibe la tendencia normal de compactación del sedimento y permite que se acumulen unas presiones anormalmente altas en la celda. La cualidad de incompresibilidad que presentan los fluidos sobrepresionados preserva la porosidad inicial del sedimento durante un tiempo considerable dentro de la evolución diagenética.

Durante el proceso de aumento y disminución de la presión de fluidos pueden distinguirse varias etapas que se encuentran asociadas con la precipitación de neominerales, cambios en el empaquetamiento y migración de fluidos.

### **Bibliografía**

- Berner, R.A. (1969) Calcium carbonate concretion formed by the decomposition of organic matter. *Science*, **159**, 195-197.
- Coleman, M.L. & Raiswell, R. (1995) Source of carbonate and origin of zonation in pyritiferous carbonate concretions: evaluation of a dynamic model. *Amer. J. Sci.*, **298**, 282-308.
- Correns, C.W. (1950) Zur Geochemie der Diagenese. *Geochim. et Cosmochim. Acta*, **1**, 49-54.
- Criss, R.E., Cooke, G.A. & Day, S.D. (1988) An organic Origin for the Concretion of the Ohio Shale. *U.S. Geol. Survey Bull.*, **1836**, 21 p.
- Dewers, T. & Ortoleva, P. (1990) Force of crystallization during the growth of siliceous concretions. *Geology*, **18**, 204-207.
- Garrels, R.M. (1960) Mineral equilibria. Ed.: Harper and Brothers, New York, 254 p.
- Gautier, D.L. (1982) Siderite concretions: indicators of early diagenesis in the Gammon Shale (Cretaceous). *Jour. Sedim. Petrol.*, **52**, 859-871.
- Girin, Y.P. (1970) Geochemical stages during diagenesis of Middle Jurassic sediments of the High Caucasus. *Geochem. Intern.*, **4**, 1146-1158.
- Hennessy, J. & Knauth, L.P. (1985) Isotopic variations in dolomite concretions from the Monterey Formation, California. *Jour. Sedim. Petrol.*, **55**, 120-130.
- Lippman, F. (1955) Ton, Geoden und Minerale des Barrême von Hoheneggelsen. *Geol. Rundschau*, **43**, 475-503.
- Mc Bride, E.F. (1988) Contrasting diagenetic histories of concretions and host rock, Lion Mountain Sandstone (Cambrian), Texas. *Geol. Soc. Am. Bull.*, **100**, 1803-1810.
- Meschter, D.Y. (1958) A study of concretions as applied to the geology of uranium deposits, v. TM-D-1-14, U.S. Atomic Energy Commission Technical Memorandum Report, 10 p.

Mozley, P.S. & Burns, S.J. (1993) Oxygen and carbon isotopic composition of marine carbonate concretions: An overview. *Jour. Sed. Petrol.*, **63**(1), 73-83.

Mozley, P.S. (1996) The internal structure of carbonate concretions in mudrocks: a critical evaluation of the conventional concentric model of concretion growth. *Sedim. Geol.*, **103**, 85-91.

Navas, A. y Tena, J. (1987) Génesis de nódulos de cuarzo en el Paleozoico de la Ibérica (Rama Aragonesa). *Estud. geol.*, **43**, 179-184.

Pettijohn, F.J. (1975) Concretions, nodules and others diagenetic segregations. En: *Sedimentary rocks*. De.: Harper and Row, New York (USA), pp. 462-482.

Raiswell, R. (1971) The Growth of Cambrian and Liassic concretions. *Sedimentology*, **17**, 147-171.

Raiswell, R. (1976) The microbiological formation of carbonate concretions in the Upper Lias of N.E. England. *Chemical Geology*, **18**, 227-244.

Scotchman, I.C. (1991) The geochemistry of concretions from the Kimmeridge Clay Formations of southern and eastern England. *28th Annual Meeting of the British Sedimentological Research Group*, unpaginated.

Sellés-Martínez, J. (1996) Concretion morphology, classification and genesis. *Earth-Sci. Rev.*, **41**, 177-210.

Shearman, D.J., Mossop, G., Dunsmore, H. and Martin, M. (1972) Origin of gypsum veins by Hydraulic fracture. *Trans. Inst. Min. Metall.*, **Sec. B181**, 148-155.

Van Tuyt, M. (1916) The Geodes of the Keokuc Beds. *Am. J. Sci.*, **42**, 34-42.

Vysotskiy, V.N. and Sidorenkov, A.I. (1987) Concretion cyclicity of oil- and gas-bearing formations. *International-Geology-Review*, **29**, 238-247.

Walker, T.R. (1960) Carbonate replacement of detrital crystalline silicate minerals as a source of authigenic silica in sedimentary rocks. *Geol. Soc. America Bull.*, **71**, 145-152.

Walker, T.R. (1962) Reversible Nature of Chert-Carbonate Replacement in Sedimentary Rocks. *Geol. Soc. Amer. Bull.*, **73**, 237-242.

Wilkinson, M. (1991) The concretions of the Bearerraig Sandstone Formation: geometry and geochemistry. *Sedimentology*, **38**, 899-912.

Wilkinson, M. (1993) Concretions of the Valtos Sandstone Formation of skye - geochemical indicators of palaeo-hydrology. *Jour. Geol. Soc.*, **150**, 57-66.

Yemane, K. & Kelts, K. (1996) Isotope geochemistry of Upper Permian early diagenetic calcite concretions: Implications for Late Permian waters and surface temperatures in continental Gondwana. *Paleog. Palaeocl. Palaeoecol.*, **125**, 51-73.

## **RASGOS EDÁFICOS EN LA PERIFERIA DEL LAGO TERCIARIO CENTRAL DE LA CUENCA DE CALATAYUD.**

**R. B. Joven Barcelona, J. Mandado Collado, F. J. Sanz Morales y J. Torrijo Echarri\***

\*Área de Petrología y Geoquímica. Departamento de Geología. Universidad de Zaragoza.  
C/ Pedro Cerbuna 12, 50009 ZARAGOZA.

### **ABSTRACT**

Calatayud Tertiary Basin is constituted by a sedimentologic sequence with more or less concentric disposition around a central lake, that present rapid lateral changes of facies.

In the SE zone of Range of Armantes, determined points show pedogenetic features. The presence of retraction crevices, preserved by calcareous filling with singular textures, no described until the date, is one of pedogenetic features most interesting and spectacular of the zone. These structures generate by processes of desiccation and evaporation in the argillaceous sediments of edge of lake and have a great utility as paleogeographic indicators of the periphery of the same.

### **INTRODUCCIÓN**

La Cuenca de Calatayud es una depresión alargada de dirección NO-SE que se extiende desde Calatayud (Zaragoza) hasta Montalban (Teruel) y cuyas dimensiones son aproximadamente de 130 km de longitud y 20 km de anchura. El estudio que se va a realizar se desarrollará en el margen Oriental de la Sierra de Armantes, situada al N de Calatayud.

El objetivo fundamental de este trabajo consiste en la identificación y estudio de rasgos edáficos desarrollados en los materiales lacustres de la cuenca. Los rasgos edáficos que se van a estudiar son costras síliceas, sílex tobáceos y grietas de retracción, que se localizan en materiales margosos del margen del lago. El interés del estudio de estos rasgos radica en su utilidad como marcadores paleoambientales y la ausencia de referencias anteriores. El antecedente más conocido sobre sílex tobáceos es el estudio presentado por Bustillo *et al.* (1980) en el Mioceno inferior continental de la provincia de Cuenca, en el cual hace una descripción textural de los diferentes tipos encontrados y plantea las etapas que han dado lugar a su formación. En cuanto a las grietas de retracción rellenas por carbonato, no se ha encontrado ninguna referencia previa de un fenómeno similar que pueda explicar su formación, debido al gran tamaño que presentan y al relleno poco habitual.

#### GEOLOGÍA GENERAL DE LA CUENCA

La Cuenca de Calatayud es una fosa tectónica formada por un proceso de rifting en el Mioceno medio mediante la reactivación de fracturas de zócalo. Así, Julivert (1954) señala que la Depresión de Calatayud, en el área de Daroca, tiene un carácter de semigraben, con margen activo al NE por la reactivación de cabalgamientos antiguos, como las fallas normales, mientras que a la altura de Calatayud se puede considerar una verdadera fosa tectónica.

La Cuenca está limitada al SO por la llamada falla de Jiloca, la cual está constituida por un sistema de fallas de dirección NO-SE escalonadas de forma levógira (Colomer *et al.*, 1988), y al NO por una discordancia, la cual pone en contacto los conglomerados terciarios y los materiales del Paleozoicos.

Sobre los materiales paleozoicos se depositaron los materiales mesozoicos y paleógenos en un ambiente continental endorréico, que continúa en el Mioceno con el depósito de evaporitas en el centro de la cubeta completándose el relleno con depósitos predominantemente calcáreos.

El relleno de la cuenca está representado por una gradación de sedimentos desde los detríticos groseros (brechas y conglomerados) hasta evaporitas (yesos y calizas), con los términos intermedios de areniscas, lutitas y margas. La morfología de la cuenca condiciona la distribución espacial de los diferentes materiales, observándose una disposición en bandas casi

concéntricas constituidas por materiales progresivamente más groseros cuanto más próximos están a los bordes.

En la vertical se observa una gran variedad de facies, con una tendencia general a la disminución en el tamaño de grano hacia el techo. También son característicos los bruscos cambios laterales que se aprecian, llegando a situar en el mismo plano los yesos y facies groseramente detríticas, en distancias relativamente cortas.

El aporte de materiales y su depósito está condicionado por la topografía de los bordes y el progresivo hundimiento de la cuenca hacia el SE. La granoselección se realiza en pocos metros, lo que da idea de la rápida pérdida de energía de las corrientes que transportan el material.

Los materiales de gran tamaño de grano se localizan cerca del borde de la cuenca. Así, las brechas se mantienen prácticamente constantes en el borde SO, mientras que el borde NE es variable, pudiendo estar sustituidas las brechas por un material detrítico más grosero, como pueden ser los derrubios de ladera. De la misma manera los conglomerados son típicos de facies de borde, llegando en algún caso a facies más centrales debido a algunas tormentas de gran energía.

El paso a facies internas se produce con materiales de grano más fino. Los conglomerados que siguen apareciendo son de tamaño de grano menor que el de los bordes, con cantos más redondeados de caliza, cuarcitas y pizarras, cementados por carbonato o matriz arenoso-lutítica.

Los sedimentos detríticos y lutítico-carbonatados, que se localizan en el interior de la cuenca, son los materiales más abundantes, observándose un paso brusco de unos a otros. Se presentan en unidades de gran espesor y como matriz de muchas areniscas, de las cuales pasan gradual o bruscamente a formar facies independientes.

Los yesos se encuentran en la parte central de la depresión, fundamentalmente desde la Sierra de Armantes hacia el SE. Normalmente aparecen en grandes unidades, pero también son frecuentes lechos aislados de espesores variables, interestratificadas en las lutitas rojas y las margas. También es frecuente la presencia de yesos nodulares intercalados en lutitas y margas. La mineralogía de los materiales evaporíticos es muy variada, ya que además del yeso-anhidrita y halita descritos hasta la fecha Ortí *et al.* (1994) sugiere la posible existencia de un importante depósito de glauberita-halita en la superficie, gracias a la observación de estratos de yeso secundario con pseudomorfos de glauberita.

Según Collantes *et al.* (1982), la formación de estos depósitos tuvieron lugar en una cuenca continental endorréica mediante procesos de reemplazamiento y desplazamiento dentro de materiales porosos, como lutitas y arenas de grano fino de borde de lago, o por la precipitación por evaporación de agua hipersalinas.

La formación central evaporítica con yesos y algún nivel de halita y epsomita pasan lateralmente a arcillas, margas rojas y conglomerados en la parte baja, y a niveles de calizas lacustres en la parte alta, atribuyéndoles una edad *Aquitaniense-Vindoboniense*. En la Sierra de Armantes se sitúan sobre facies margo-arcillosas, y sobre ellos se sitúa una formación calcárea de edad *Aquitaniense-Pontiense*, además de materiales margosos y limos.

Los sedimentos calcáreos están abundantemente representados, situándose en las planas que ocupan el centro de la cuenca. Estos depósitos carbonatados se encuentran también intercalados entre los materiales detríticos finos o interdigitándose con los yesos.

Los depósitos carbonatados de las planas tienen su origen en la precipitación dentro del lago central en condiciones menos salobres, aunque también se considera la influencia de algas para su depósito. Los materiales calcáreos que se encuentran digitados entre los materiales detríticos se consideran debidos a la precipitación en zonas de llanura circundante al lago sulfatado por procesos de concentración evaporativa.

Una de las características de las calizas encontradas en la Sierra de Armante es la presencia de silicificaciones en forma de nódulos y capas continuas de sílex. Es importante señalar la evolución lateral de nódulos de sílex a una capa continua que desaparece cuando se produce el tránsito de la facies carbonatada a las facies lutíticas. El proceso que se interpreta para la formación de estas silicificaciones es la sustitución de los carbonatos preexistente por la sílice, habiéndose observado en lámina delgada el proceso señalado.

La cronoestratigrafía se interpreta gracias a la presencia de restos de mamíferos en diferentes puntos de la cuenca. Crusafont *et al.* (1957) señala unidades que abarcan desde el *Vindoboniense medio* hasta el *Pontiense superior*, para todo el paquete sedimentario de la depresión. Señala que la fauna *Vindoboniense*, en la zona N, tiene un carácter centroespañol, mientras que en la zona S no aparecen restos en los niveles rojos inferiores. Las asociaciones pontienses que aparecen se diferencian tanto en las litofacies como en las biofacies, lo que impide llegar a conclusiones definitivas.

Trabajos más actuales han establecido la cronoestratigrafía del Terciario para materiales continentales. Freudenthal (1988) establece la cronoestratigrafía del Neógeno por el contenido de restos de familias de roedores dentro de los materiales de la Cuenca de Calatayud-Teruel.

La cronología que se utiliza para la cuenca, de más antiguo a más moderno, abarca los subpisos *Ramblense*, *Aragoniense*, *Vallesiense*, *Turoliense* (*Pontiense*) y *Rusciniense*.

Tena y Mandado (1984) señalan la presencia en las calizas y margas del centro de la cuenca con restos algáceos (en especial caráceas), ostrácodos y gasterópodos pulmonados. También señalan la existencia de lechos lignitíferos de escaso espesor intercalados en los tramos margosos.

### COLUMNA LITOSTRATIGRÁFICA

Los materiales aflorantes en el NE de la S<sup>a</sup> de Armantes se recogen de forma esquemática en la columna litoestratigráfica realizada entre la Ermita del Cristo de Ribota y la Ermita de La Santa Cruz. La zona se localiza en el cuadrante II de la hoja 409 (Calatayud) del Mapa Topográfico Nacional a escala 1:50.000. Una segunda columna, levantada a unos dos km al SO (figura 1B) permite evidenciar las marcadas variaciones de facies que se registran en esta zona.

En la serie de la Ermita de Ribota se diferencian básicamente cuatro tramos (figura 1A):

- Tramo basal lutítico con intercalaciones margosas y areníticas.
- Calizas margosas con intercalaciones carbonatas y margosas .
- Tramo carbonatado con intercalaciones de calizas margosas.
- Tramo terminal carbonatado arenoso con intercalaciones de materiales más detríticos.

El tramo inferior, de potencia decámetrica, está constituido por lutitas rojizas con intercalaciones de capas de arenisca de color rojizo a verdoso y calizas margosas blanquecinas en la base. El término lutítico contiene concreciones carbonatadas de color verdoso, con tendencia a desaparecer hacia el techo del tramo. También se puede observar, dentro de lutitas oscuras con alto contenido en materia orgánica, la presencia de grietas de retracción, siendo interesante el gran tamaño que presentan, así como el material carbonatado que las rellena (estructura que será descrita e interpretada posteriormente). Este tramo presenta un cambio lateral significativo, ya que se observa un cambio a lutitas rojizas, sin presencia de capas carbonatadas ni areniscosas, a pocos metros de distancia.

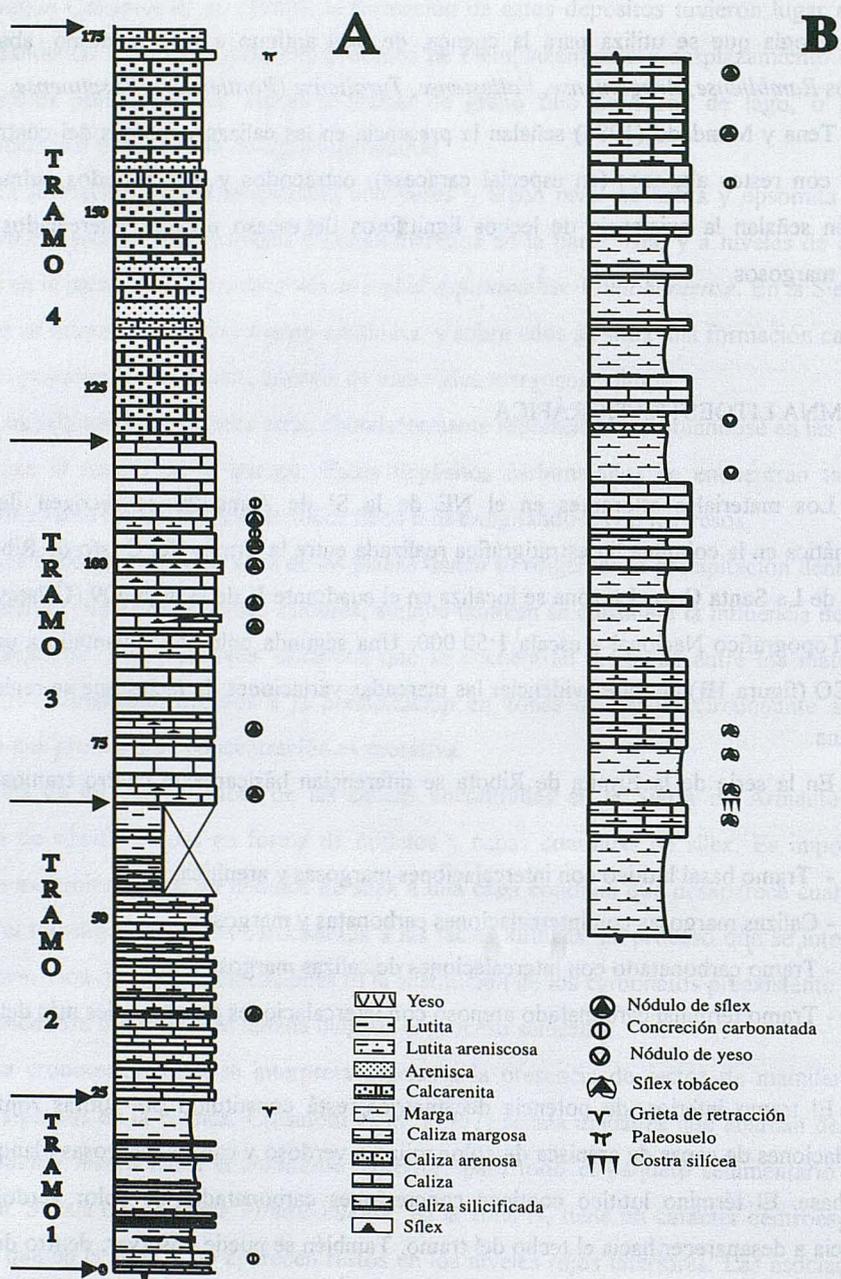


Figura 1. Columnas litoestratigráficas levantadas en el margen oriental de la Sierra de Armantes. A) Columna de la Ermita de Ribota. B) Columna a 1'5 km al SO de la anterior, mostrando la rápida variación de facies en la zona.

La presencia progresivamente creciente de carbonatos, desarrollo de concreciones carbonatadas, marmorizaciones, incremento en el contenido de materia orgánica y, finalmente, las grietas de retracción marcan la transición de una llanura aluvial a un medio lacustre.

Estos materiales aparecen recubiertos por unos niveles de calizas margosas blanquecinas con intercalaciones carbonatadas grisáceas y, a techo, lutitas de color rojizo. Dentro del término margoso es interesante señalar la presencia de estructuras de escape de fluidos y nódulos carbonatados, en la base, y la silicificación de capas concretas del tramo, en la zona intermedia. Se puede observar un importante cambio lateral de estos materiales, que pasan a otros de naturaleza lutítica, rojizos, hacia el NO, y a margas blanquecinas, hacia el SO.

El paso progresivo de materiales carbonatados a materiales lutíticos indicarían la transición de un medio lacustre, más o menos somero, a un medio de llanura aluvial.

El tercer tramo, carbonatado, tiene potencia decamétrica, contiene un material calcáreo ocre-amarillento con intercalaciones de calizas margosas de color ocre a blanquecino. Los carbonatos presentan importantes silicificaciones a lo largo del tramo, llegando incluso a aparecer capas continuas de sílex. Las calizas margosas, a techo del tramo, presentan concreciones carbonatadas de color ocre. Hacia el SO este tramo se relaciona con alternancias de materiales carbonatados y margosos en los cuales también se pueden observar silicificaciones, aunque de tipo tobáceo.

La presencia de material carbonatado sin estructuras internas apreciables ni laminaciones, durante todo el tramo, indicaría que se formaron en la zona central de un lago poco profundo.

El tramo superior, de potencia decamétrica, corresponde a materiales carbonatados arenosos de color oscuro con intercalaciones de areniscas y calcarenitas ocre oscuras, así como algún nivel más calcáreo de color ocre. En el techo del tramo se puede observar los restos de un paleosuelo, dentro de un material calcáreo. El cambio lateral de este tramo hacia NO muestra materiales detríticos y lutíticos de color rojizo, mientras que hacia el SO muestra materiales carbonatados y margosos.

La presencia de material carbonatado y el aumento del contenido en material detrítico indicaría un medio lacustre, más o menos profundo, donde existiría un transporte de mayor energía.

## INDICADORES EDÁFICOS

Los suelos, por ser formaciones superficiales, suelen ser erosionados en el momento que se produce un cambio dinámico en la corteza terrestre. No obstante, algunos suelos presentan niveles duros resistentes a la erosión que hacen posible su conservación. También existen niveles blandos que pueden conservarse, en el caso de la existencia de un proceso rápido de sedimentación sobre los materiales del suelo. En general, se puede decir que la conservación de rasgos edáficos en el registro estratigráfico está principalmente condicionada a la existencia de procesos diagenéticos de reducida intensidad.

La representación en el registro estratigráfico de estructuras que indican paleosuelos es variada en cuanto al tipo y proceso que las han generado. Los principales rasgos edáficos que se pueden encontrar, como marcadores de paleosuelos, son las huellas de raíces, las texturas moteadas y concreciones e, incluso, las estructuras características de suelos salinos. Petrológicamente, los rasgos de suelos que se observan, se pueden agrupar en rasgos pedológicos (cutanes, cristales intercalados y nódulos o concreciones) y rasgos en la matriz del suelo (material coloidal, fragmentos detríticos y huecos generados en el proceso edáfico) (Brewer, 1964).

Los principales fenómenos edáficos, tanto físicos como químicos, que se pueden dar en carbonatos lacustres son la desecación y brechificación, la nodulización y recristalización de la calcita, la neoformación de minerales y la presencia de revestimientos, cementos y sedimentos internos Freytet (1984). Este autor señala también la acción de factores biológicos por la ausencia de laminaciones, bioturbación en el fondo, y las modificaciones químicas que generan con su actividad vital, y que favorecen la formación de determinados minerales.

En la Cuenca de Calatayud, y más concretamente en la Sierra de Armantes, se han encontrado diversos marcadores edáficos, tanto a escala macro como microscópica. De ellos, los que consideramos más interesantes y describimos a continuación son: grietas de retracción y sílex tobáceos.

1. *Los sílex con estructura tobácea* se encuentran dispuestos en superficies planas irregulares de dimensiones decimétricas a métricas y grosor escasamente centimétrico. Se encuentran asociados a costras calcáreas y caliches desarrollados sobre lutitas con intercalaciones de yesos. La estructura tobácea se evidencia por la existencia de conductos subverticalizados rellenos por sílex microcristalino de gran pureza, encajados en un material

también silíceo, de reemplazamiento de lutitas y yesos lenticulares. Un rasgo complementario del carácter superficial de estos depósitos es la presencia en ellos de grietas de retracción abiertas en la masa silíceo, generadas por un proceso de desecación en superficies libres.

2. *Las grietas de retracción* estudiadas en la Sierra de Armantes se encuentran en el interior de un material lutítico de color pardo, muy oscuro, debido al elevado contenido en materia orgánica. En campo aparecen como tabiques centimétricos, muy compactos, que dibujan una estructura poligonal de radio métrico. El material que los constituye es carbonatado y debido a la mayor dureza que presenta respecto al encajante permite que se visualicen fácilmente, al aparecer como resaltes en la superficie de meteorización. Desde un punto de vista mesoscópico el material del relleno de las grietas presenta texturas diferenciables en dos zonas:

- La zona interna del relleno es un carbonato, de composición calcítica, color anaranjado, con una textura esparítica drúsica y en el que se observan contactos netos entre los cristales. Localmente aparecen restos de material lutítico, orientado paralelamente al relleno de la grieta, apreciándose por lo general en sus inmediaciones una disminución en el tamaño de los cristales.

- La zona externa está formada por dos bandas que rodean al relleno interno y que están constituidas por un carbonato, también calcítico, de color oscuro y de aspecto masivo. La observación microscópica permite apreciar una textura con morfología de abanicos, de aproximadamente 1-2 mm, y esferulitos, de aproximadamente 0.5 mm, con crecimiento prismático de los cristales. La orientación de estos abanicos es del interior del relleno hacia el material lutítico que contiene las grietas, siendo más marcada en el contacto entre las dos zonas del relleno. También se observa, en cortes transversales al relleno, que en el interior del mismo existen formas esferulíticas de carbonato donde localmente se puede observar un núcleo de cristales carbonatados. En estos cristales es importante destacar su carácter reemplazante, evidenciado porque el carbonato engloba gran cantidad de material lutítico y clastos de cuarzo, que en ocasiones dificulta el normal desarrollo de los mismos.

Lateralmente a estos afloramientos de grietas de retracción aparecen otros rellenos de grietas en los cuales la zona interna se reduce a una fina capa, casi imperceptible. Esto hace que todo el relleno tenga un aspecto de material carbonatado de color oscuro y masivo. En

lámina delgada este material sigue presentando texturas en abanico y esferulitos que engloban gran cantidad de material lutítico y clastos de cuarzo.

El mecanismo de formación de los rellenos carbonatados de las grietas de retracción encajadas en materiales lutíticos de la llanura aluvial cercana al lago se sintetiza en una serie de pasos que se comentan a continuación:

FASE 1. Inicialmente, el material lutítico correspondería a un depósito marginal del lago, bajo lámina de agua, en la zona de desarrollo de vegetación. Por retracción del lago, esta superficie sufre una etapa de desecación, en la cual se produce el cuarteamiento del sedimento y se forman las grietas de retracción.

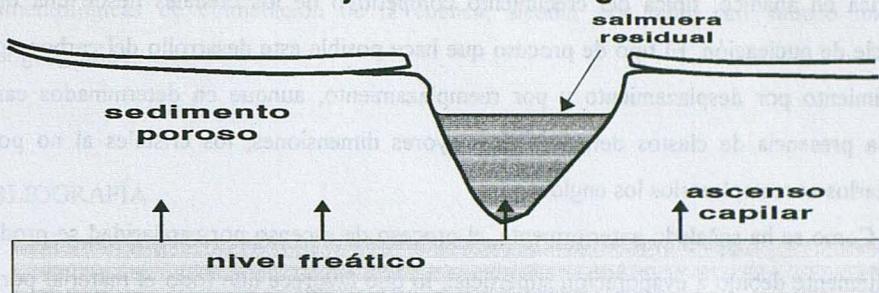
Este episodio supone para el sedimento la pérdida de fluido, en la superficie, y un aumento de la porosidad, esto último debido a la disminución de volumen que sufre la materia orgánica que contiene. El agua residual da lugar a una salmuera concentrada que se localiza en la zona abierta de la grieta, el punto topográficamente más bajo. La zona superior del sedimento se deseca totalmente quedando la escasa agua libre en las grietas, mientras que por debajo de ella se localiza el nivel freático, que favorece al aporte de agua concentrada para la posterior precipitación del material carbonatado por un mecanismo de ascenso capilar por bombeo evaporítico (Fase 1 de la figura 2).

FASE 2. La concentración de la salmuera llega al punto de saturación de la calcita y se produce la precipitación de un cemento esparítico en la superficie libre de la grieta. El crecimiento competitivo de los cristales se da en todo el espacio nucleando sobre las superficies del sedimento y aumentando progresivamente su tamaño hacia las superficies libres según el clásico esquema de crecimiento en mosaico drúsico.

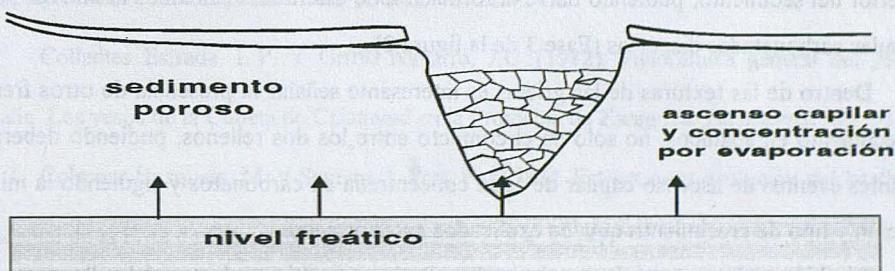
El material carbonatado que precipita en este proceso procede de la salmuera que ha quedado estancada en el espacio libre de la grieta, pero también se produce un aporte por ascenso capilar desde el nivel freático. Este ascenso se ve favorecido por la evaporación que está teniendo lugar en la superficie del sedimento y en la zona abierta de la grieta (Fase 2 de la figura 2).

FASE 3. En el momento en el que se ha formado el relleno central de la grieta se produce una mayor diferenciación en la permeabilidad entre éste y el material poroso que lo rodea. Esto hace que en el momento en el que se producen eventos de ascenso capilar, el agua circula en el contacto entre los dos materiales, favoreciendo la nucleación y formación de cristales de carbonato en esta zona. La dirección de crecimiento de los cristales se produce hacia la zona más permeable, en este caso el sedimento poroso, y da lugar a una textura

### FASE 1: Deseccación y desarrollo de grietas de retracción.



### FASE 2: Concentración evaporítica y precipitación de cemento en las grietas.



### FASE 3: Crecimientos desplazantes a partir del agua que asciende por capilaridad.

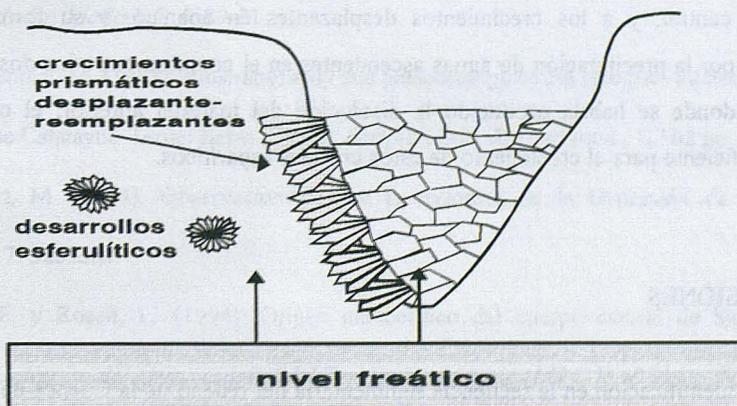


Figura 2. Modelo de formación de los rellenos carbonatados de las grietas de retracción intercaladas en depósitos lutíticos de periferia del lago carbonatado de la Cuenca terciaria de Calatayud.

prismática en abanico, típica del crecimiento competitivo de los cristales desde una única superficie de nucleación. El tipo de proceso que hace posible este desarrollo del carbonato es el crecimiento por desplazamiento y por reemplazamiento, aunque en determinados casos, como la presencia de clastos detríticos de mayores dimensiones, los cristales al no poder desplazarlos ni reemplazarlos los engloban.

Como se ha señalado anteriormente, el proceso de ascenso por capilaridad se produce frecuentemente debido a evaporación superficial, lo que favorece que todo el material poroso que rodea a la grieta pueda estar embebido de agua pero en condiciones vadosas, lo que favorece la evaporación y concentración de la salmuera capilar. Esto facilita la precipitación en el interior del sedimento, pudiendo darse la formación de esferulitos calcíticos al nuclear sobre partículas carbonatadas dispersas (Fase 3 de la figura 2).

Dentro de las texturas de las grietas es interesante señalar la presencia de otros frentes de crecimiento en abanicos, no sólo en el contacto entre los dos rellenos, pudiendo deberse a diferentes eventos de ascenso capilar de agua concentrada en carbonatos y siguiendo la misma dirección y tipo de crecimiento que los explicados anteriormente.

También es interesante la presencia de cristales esparíticos, dentro del relleno central, con una orientación que indica el crecimiento desde el contacto con la zona de abanicos hacia el interior. Estos cristales se formaron en etapas posteriores a la formación del relleno esparítico central, y a los crecimientos desplazantes en abanico, y su formación podría explicarse por la precipitación de aguas ascendentes en el contacto entre los dos rellenos y en una zona donde se habría producido la disolución del material anterior, el cual dejaría el espacio suficiente para el crecimiento de estos cristales esparíticos.

## CONCLUSIONES

La identificación en la secuencia sedimentaria del relleno de la Cuenca de Calatayud de rasgos edáficos en los depósitos químicos o de transición de las facies centrales de la misma permite ampliar y mejorar el esquema sedimentológico de esta cuenca. Especialmente adquieren una gran relevancia los rasgos indicativos de periferia del lago carbonatado central, como grietas de retracción y sílex tobáceos, especialmente los primeros, cuyas características texturales permiten establecer un modelo genético claramente correlacionable con las etapas

sedimentológicas de colmatación de la cuenca, siendo, por tanto, un valioso marcador paleogeográfico.

## BIBLIOGRAFÍA

- Brewer, R. (1964). *Fabric and mineral analysis of soils*. Wiley, 490 pp.
- Bustillo Revuelta, M.A. y Díaz Molina, M. (1980). Sílex "tobáceos" en el Mioceno inferior continental (provincia de Cuenca). Un ejemplo de silicificaciones de paleosuelos en ambiente de lago-playa. *Bol. R. Soc. Esp. Hist. Nat. (Geol.)*, **78**: 227-241.
- Collantes Estrada, L.P. Y Griffó Navarro, J.L. (1982). Panorámica general del yeso en España. Los yesos de la Cubeta de Calatayud en la Provincia de Zaragoza. *Tecniterrae*, **46**: 53-63.
- Colomer Busquets, M. y Santanach Prat, P. (1988). Estructura y evolución del borde sur-occidental de la Fosa de Calatayud- Daroca. *Geogaceta*, **4**: 29-31.
- Crusafont Pairó, M., Villalta Comella, J.F. y Truyols Santoja, J. (1957). Definición estratigráfico-paleontológica de la cuenca terciaria de Calatayud-Teruel. *Cursillos y Conferencia del Instituto "Lucas Mallada"*, **4**: 89-91.
- Freudenthal, M. (1988). Biostratigraphy and paleoecology of the Neogene micromammalian faunas from the Calatayud-Teruel Basin (Spain). *Scripta Geol., Special issue*, **1**, 302 pp.
- Julivert, M. (1954). Observaciones sobre la tectónica de la Depresión de Calatayud. *Arrahona*, 1-17. Sabadell.
- Orti, F. y Rosell, L. (1994). Origen glauberítico del cuerpo central de los yesos de Calatayud. *II Congreso del grupo español del Terciario, simposio (I.G.C.P. n° 324)*: 255-258.
- Tena, J.M. y Mandado, J.A. (1984). Estudio geológico de la Cuenca terciaria de Calatayud-Daroca. *Rev. Acad. Ciencias Zaragoza*, **39**, 69-78.

## **ANALISIS DE LA ESTABILIDAD DE LADERAS EN SUELOS UTILIZANDO LOS ABACOS DE HOEK Y BRAY: UN CASO REAL EN CAMARENA DE LA SIERRA (TERUEL).**

**Adolfo Maestro González y Angel Luis Cortés Gracia**

Departamento de Geología. Facultad de Ciencias. Universidad de Zaragoza.

### **ABSTRACT**

Hoek and Bray's charts allow to realize simple analysis of the stability conditions for soil and soft rock slopes when a circular failure can occur. This method enables the user to carry out an easy and rapid check on the factor of safety of a slope as well as to study the changes of the factor of safety depending on the groundwater conditions of slope. The utility of this method is shown from three unstable slopes in Camarena de la Sierra (Teruel). We used the slope stability charts for obtaining tables and diagrams showing the influence of changes in the slope (cohesion and friction angle of the soil and groundwater conditions) on the factor of safety. The applied study shows the instability of slopes in fully saturated conditions and the different stability conditions for possible values of cohesion and friction angle.

### **1. INTRODUCCION**

La construcción de una carretera en una zona montañosa constituida por materiales blandos (yesos y arcillas) presenta graves problemas geotécnicos. Cuando el estudio de estabilidad no es tenido en cuenta en el proyecto de realización de la obra pueden surgir graves problemas, que acarrear importantes pérdidas económicas y daños, en ocasiones, irreparables o de difícil solución. Un ejemplo de obra civil ya realizada y que actualmente presenta problemas de inestabilidad es la carretera que une las localidades de La Puebla de Valverde y Camarena de la Sierra (Teruel), en las proximidades de esta última localidad. Una primera aproximación a las condiciones de estabilidad de estas laderas o taludes artificiales se puede realizar mediante los ábacos de Hoek y Bray (1977).

En este trabajo se han realizado una serie de estudios geológicos y geotécnicos que han permitido caracterizar los materiales que constituyen los taludes y laderas naturales que bordean la carretera de Camarena de la Sierra. A partir de los parámetros obtenidos en estos estudios y mediante la utilización de los ábacos de Hoek y Bray se han construido tablas y diagramas que muestran las variaciones del factor de seguridad en función de las condiciones del talud:

cohesión y ángulo de fricción interno del suelo, altura y pendiente del talud y grado de saturación en agua.

## 2. SITUACION GEOGRAFICA

El área de estudio se encuentra localizada al Sur de la provincia de Teruel, concretamente en las estribaciones septentrionales de la Sierra de Javalambre. Es una zona montañosa, con desniveles que oscilan entre los 500 y 600 metros y pendientes medias entre  $15^\circ$  y  $20^\circ$ , destacando las cumbres de Peña Blanca (1.752 m) y El Cerro (1.543 m) inmediatamente al Norte del macizo de Javalambre, que alcanza los 2.020 m en el pico del mismo nombre. La localidad de Camarena de la Sierra, a 1.294 m de altitud, está situada junto al río Camarena y tiene una población de apenas 200 habitantes (Fig. 1).

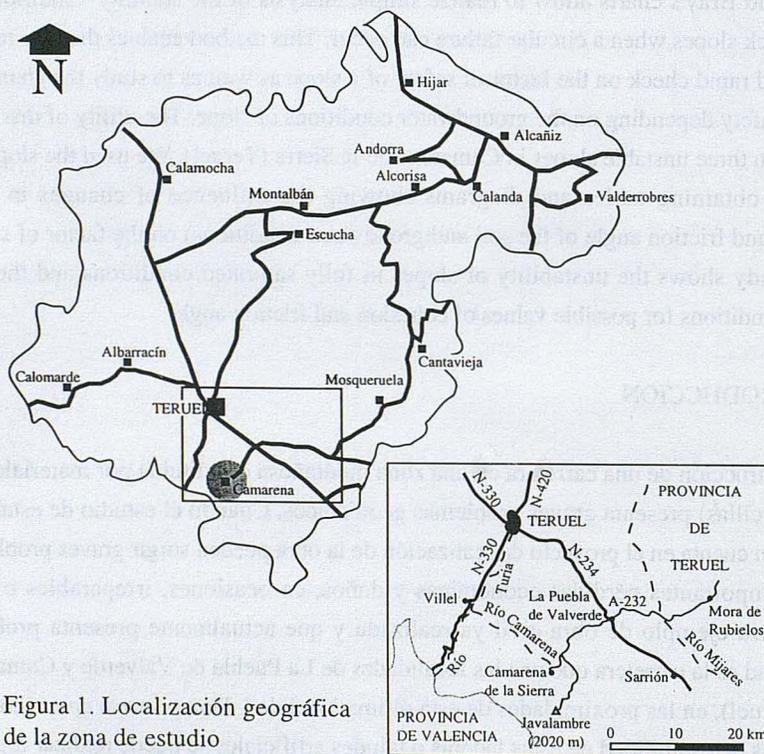


Figura 1. Localización geográfica de la zona de estudio

## 3. MARCO GEOLOGICO

El macizo del Javalambre está constituido fundamentalmente por formaciones carbonatadas de edad jurásica, aflorando el Triásico dolomítico (Facies Muschelkalk) y arcilloso (Keuper) a favor de los fuertes encajamientos de la red fluvial. En su conjunto presenta una estructura en domo atravesada por numerosos sistemas de fracturas, cuyas orientaciones dominantes son NW-SE y NE-SW (Fig. 2).

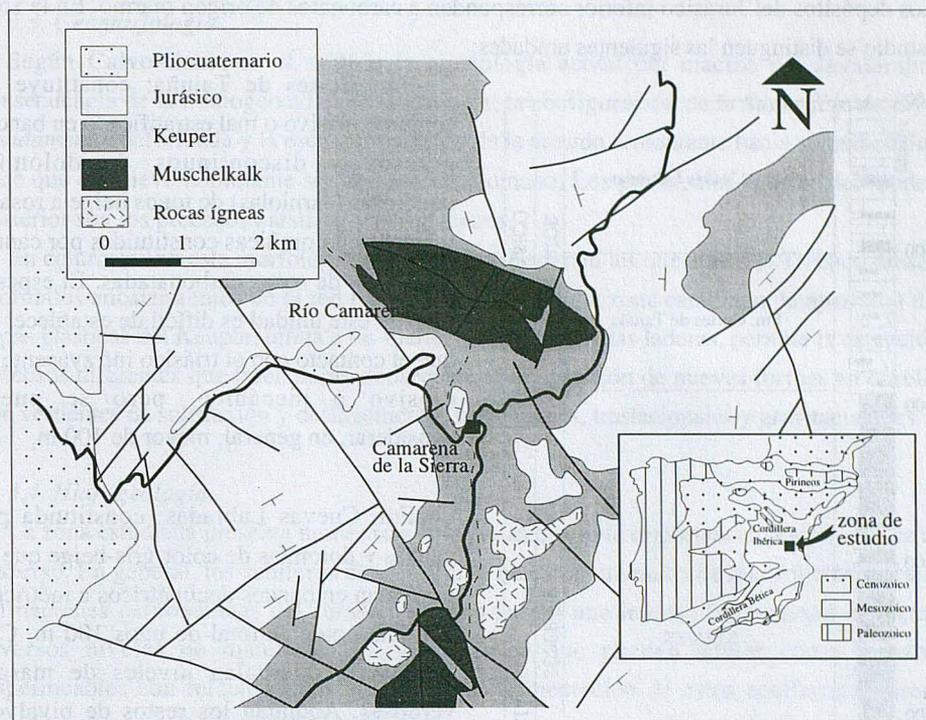


Figura 2. Mapa geológico esquemático de los alrededores de Camarena de la Sierra.

### 3.1. Estratigrafía

Los materiales dominantes en la zona corresponden a sedimentos mesozoicos de edad Triásico medio-superior y Jurásico. Por su importancia en el estudio que nos ocupa, destacaremos las arcillas de las facies Keuper (Triásico superior) y las brechas dolomíticas y carniolas de la base del Jurásico (Fig. 3), cuyas características describimos seguidamente.

El Triásico de esta región presenta materiales en "Facies Germánicas" (Virgili *et al.*, 1977) y se diferencia:

-Triásico medio (facies Muschelkalk): está formado por dolomías grises que intercalan niveles de limolitas, margas y dolomías arenosas. Presentan una potencia superior a 100 m, con tramos masivos y tableados.

-Triásico superior (facies Keuper): está formado fundamentalmente por arcillas y yesos que intercalan niveles arenosos, margosos y margo-dolomíticos. Diversos autores han medido en esta región potencias de más de 300 m para estos materiales. Con bastante frecuencia afloran rocas volcánicas o subvolcánicas groseramente interestratificadas entre los materiales del Keuper. Se trata de rocas granudas con textura diabásica que presentan un avanzado estado de alteración.

Los depósitos del Jurásico inferior corresponden a carbonatos de origen marino. En la zona de estudio se distinguen las siguientes unidades:

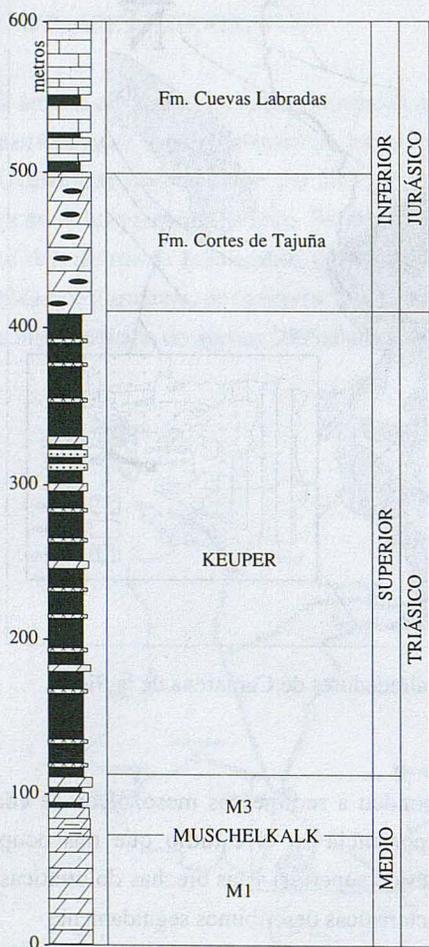


Figura 3. Columna estratigráfica general de la zona estudiada

(Sabarís, 1978; Peña *et al.*, 1984; Simón, 1984).

Hacia el Noroeste, la Sierra de Camarena presenta pliegues de orientación WNW-ESE que afectan a materiales triásicos y jurásicos. Al Norte, la cobertera mesozoica presenta una estructura compleja, ya que los materiales carbonatados del tramo superior del Muschelkalk aparecen despegados y cabalgan en ocasiones a los materiales del Keuper. En aquellas zonas donde no llega a cabargar, el Muschelkalk presenta pliegues con menor longitud de onda que los que afectan al Jurásico, lo que provoca una disarmonía entre los pliegues existentes por debajo y por encima del Keuper. Los materiales del Jurásico, suavemente plegados, muestran una importante fracturación a todas las escalas, con grandes fracturas kilométricas y a escala mesoestructural de direcciones NW-SE, NE-SW, N-S y E-W (Liesa, 1993).

-Fm. Cortes de Tajuña: constituye un conjunto masivo o mal estratificado en bancos gruesos y discontinuos de dolomías oquerosas (carniolas) de tonos beige a rosado y brechas dolomíticas constituidas por cantos angulosos de rocas carbonatadas. El espesor total de esta unidad es difícil de establecer ya que el contacto con el triásico infrayacente es erosivo o mecánico, pero se puede considerar, en general, mayor de 100 m.

-Fm. Cuevas Labradas: constituida por calizas y dolomías de color gris-beige que se presentan en bancos decimétricos a métricos, con una potencia total de unos 100 m. Con frecuencia intercalan niveles de margas verdosas. Abundan los restos de bivalvos, crinoides y braquiópodos.

### 3.2. Estructura

El macizo de Javalambre constituye un gran domo estructural afectado por una serie de fracturas de dimensiones kilométricas. Esta estructura es el resultado de una tectónica de componente vertical que deformó durante el Plioceno superior la *Superficie de Erosión Fundamental* de la Cordillera Ibérica (Solé

### 3.3. Geomorfología

Según Calvo-Cases *et al.* (1983) la morfología actual del macizo del Javalambre es consecuencia de la homogeneidad de la litología, la configuración de la *Superficie de Erosión Fundamental* deformada y la escasa progresión de la erosión remontante hacia su núcleo, lo que hace que el relieve dominante sea de carácter alomado. Los principales agentes del modelado posterior son los procesos kársticos y periglaciares.

En contraste con esta morfología, las áreas periféricas en las que aflora el Triásico presentan profundos encajamientos de la red fluvial. En el caso de la zona estudiada, la aparición de las capas plásticas del Keuper, unida a las fuertes pendientes de las laderas, permite la actuación de procesos diferentes que traen como consecuencia la aparición de nuevas formas en el relieve, con vertientes de soliflucción y deslizamientos rotacionales, traslacionales y gravitacionales.

### 3.4. Hidrogeología

La zona estudiada presenta numerosas surgencias de cierta importancia aunque con carácter disperso. En general, los acuíferos de esta región muestran un marcado control estructural. Las formaciones carbonatadas del Jurásico marino sufren una intensa fracturación e intercalan diversos niveles de materiales margoarcillosos que pueden actuar como horizontes impermeables con relación a los acuíferos. La alimentación de estos acuíferos se produce generalmente a través de las fracturas.

En el caso que nos ocupa, las carnioles y brechas dolomíticas de la base del Jurásico presentan una elevada permeabilidad, con grandes posibilidades de almacenamiento y circulación de fluidos. Además, por debajo aparecen las arcillas de las facies Keuper, que constituyen un nivel impermeable de gran importancia. Por todo ello, los materiales carbonatados del Jurásico basal constituyen un excelente acuífero que, en ocasiones, queda colgado apareciendo las surgencias en el contacto entre estos materiales y las arcillas del Keuper o dentro de este último, a través de los niveles yesíferos y arenosos.

## 4. ESTUDIO GEOTECNICO

### 4.1. Observaciones de campo

La zona estudiada se sitúa junto a la carretera que une las localidades de La Puebla de Valverde y Camarena de la Sierra (Fig. 4). En los alrededores de esta última localidad, la carretera atraviesa materiales margoarcillosos y evaporíticos del Triásico superior sobre los que se apoyan las dolomías y calizas dolomíticas del Triásico terminal-Jurásico inferior.

El trabajo en esta zona conllevó un reconocimiento del terreno, tratando de localizar distintos rasgos geomorfológicos que ayudasen a determinar las zonas inestables. De esta manera, se han localizado deslizamientos de bloques del Jurásico inferior sobre las facies Keuper, formación de grietas en los arcones de la carretera, numerosos fenómenos de reptación (*creep*) en los suelos

herbáceos que cubren las laderas, antiguos deslizamientos, así como algunas cicatrices de arranque. También se han reconocido morfologías lobuladas y/o abombadas en las laderas.

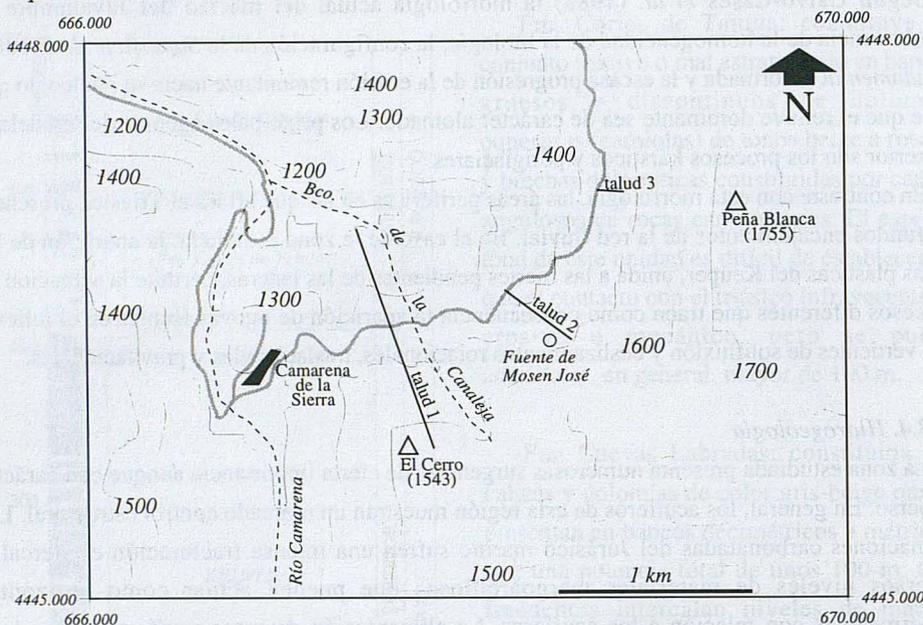


Figura 4. Esquema topográfico de la zona estudiada y situación de los taludes analizados.

Se ha prestado especial atención a la localización de surgencias de agua (procedentes del acuífero carbonatado situado por encima de los yesos y arcillas del Triásico superior) que tienen gran importancia en los fenómenos de deslizamiento de suelos.

Muestra	Y-Ca	Ma-Ca	Br-Ca	Ar-Ca
Resistencia a la compresión uniaxial (t/m <sup>2</sup> )	105,40	207,66	158,88	
Resistencia a la tracción (ensayo brasileño) (t/m <sup>2</sup> )	13,47	17,26	15,76	
Cohesión (t/m <sup>2</sup> ) (envolvente de Mohr)	26	36	32	
Cohesión (t/m <sup>2</sup> ) en corte directo	46,13	18,61	88,54	
Angulo de fricción (°) en compresión uniaxial	53/78	82	48	
Angulo de fricción (°) (envolvente de Mohr)	14	50	37,6	
Angulo de fricción (°) entre sup. de discontinuidad	31	30	35	
Densidad seca (g/cm <sup>3</sup> )	2,25	2,20	2,68	1,93
Densidad húmeda (g/cm <sup>3</sup> )	2,25	2,40	2,76	1,97
Límite de Atterberg / líquido				36%
Límite de Atterberg / plástico				27,96%
Índice de plasticidad				8,04

Tabla 1. Características geotécnicas de los materiales estudiados. Y: yesos (facies Keuper); Ma: margas (facies Keuper); Br: brechas (Fm. Cortes de Tajuña); Ar: areniscas (facies Keuper).

#### 4.2. Caracterización geotécnica de los materiales

A partir de varias muestras recogidas en el campo se han realizado ensayos mecánicos (compresión uniaxial, corte directo y ensayo brasileño) sobre diferentes rocas del Triásico y

Jurásico inferior que afloran en los alrededores de la carretera donde se realizó el estudio. Asimismo, para los materiales en facies Keuper, se ha realizado el cálculo de los límites de Atterberg en las arcillas. En la tabla 1 se representan de forma sintética todas las características geotécnicas obtenidas de los análisis realizados.

#### 4.3. Análisis de estabilidad de laderas

En aquellos casos en que la superficie de deslizamiento se puede asimilar a una superficie cilíndrica cuya sección transversal se asemeja a un arco de círculo, hablaremos de rotura circular (Fig. 5). Este tipo de deslizamiento se suele producir en terrenos homogéneos, ya sea suelos o rocas altamente fracturadas sin direcciones predominantes de fracturación, en los que además ha de darse la condición de que las partículas de suelo o roca tengan tamaño muy pequeño en comparación con las dimensiones del talud (Ayala *et al.*, 1991).

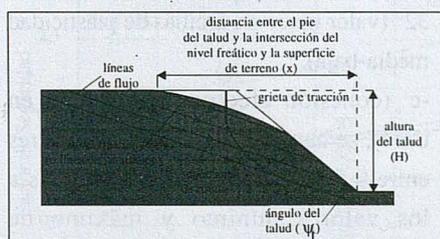


Figura 5. Ejemplo de rotura circular con los parámetros que intervienen en el cálculo del factor de seguridad del talud.

proporcionan un límite inferior del factor de seguridad del talud, obtenido en la suposición de que las tensiones normales en la superficie de deslizamiento se concentran en un único punto. El límite máximo se obtendría suponiendo que éstas se encuentran en los dos puntos extremos de la superficie. A pesar del carácter irreal de esta distribución de tensiones, se ha comprobado que el factor de seguridad real está razonablemente cerca del valor mínimo, quedando el error cometido del lado de la seguridad.

Los deslizamientos que aparecen en los alrededores de Camarena de la Sierra han sido considerados como deslizamientos rotacionales con rotura circular. En este caso, se cumplen las condiciones descritas anteriormente, ya que se trata de un terreno homogéneo (arcillas masivas) y con un tamaño de partículas muy pequeño (arcilla-limo) en relación al talud.

Para este tipo de deslizamientos se utilizan los ábacos de Hoek y Bray (1977), que

proporcionan un límite inferior del factor de seguridad del talud, obtenido en la suposición de que las tensiones normales en la superficie de deslizamiento se concentran en un único punto. El límite máximo se obtendría suponiendo que éstas se encuentran en los dos puntos extremos de la superficie. A pesar del carácter irreal de esta distribución de tensiones, se ha comprobado que el factor de seguridad real está razonablemente cerca del valor mínimo, quedando el error cometido del lado de la seguridad.

Además, se asumen las siguientes simplificaciones:

- El material constitutivo del talud se considera homogéneo en toda la extensión del mismo.
- El círculo de rotura se hace pasar siempre por el pie del talud.
- Se considera la existencia de una grieta de tracción que puede estar por encima o por debajo de la cresta del talud.

Se han estudiado tres taludes a lo largo de la carretera (Figura 4) con diferentes valores de pendiente según la altura total de los mismos. Para cada uno de ellos se han considerado fijos los siguientes parámetros:

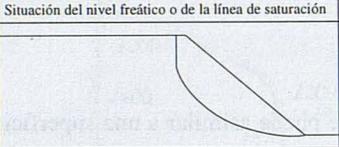
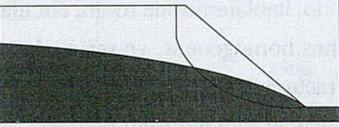
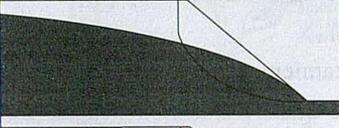
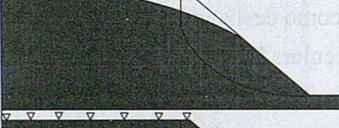
Situación del nivel freático o de la línea de saturación	caso
	Terreno seco (ábaco 1)
	Terreno saturado por debajo de la grieta de tracción: $x = 8H$ (ábaco 2)
	Terreno saturado por debajo de la grieta de tracción: $x = 4H$ (ábaco 3)
	Terreno saturado afectando la grieta de tracción: $x = 2H$ (ábaco 4)
	Terreno saturado (ábaco 5)

Figura 6. Situaciones de la línea de saturación consideradas en los ábacos de Hoek y Bray para un talud con rotura circular.

tracción y terreno completamente saturado. Los pares de valores Cohesión (C) / Factor de Seguridad (Fs), para cada uno de los taludes, pueden observarse en la figura 7.

TIPO DE SUELO	Cohesión seca	Cohesión saturada	Angulo de rozamiento interno ( $\phi'$ )	$\tan \phi'$
limos inorgánicos - limos arcillosos	6,83	0,93	32°	0,62
arcillas inorgánicas baja-media plasticidad	8,79	1,32	28°	0,54
arcillas inorgánicas muy plásticas	10,5	1,12	19°	0,35
mezcla de limos inorgánicos y arcillas	6,59	2,24	32°	0,62

Tabla 2. Valores de cohesión ( $t/m^2$ ) y ángulo de rozamiento interno efectivo para los distintos tipos de suelos (rocas lutíticas) del sector de Camarena de la Sierra.

Los valores de cohesión y ángulo de rozamiento interno efectivo estimados para los distintos tipos de lutitas observadas en el área estudiada han condicionado el rango de estos parámetros utilizado en la aplicación de los ábacos para el cálculo del factor de seguridad (Tabla 2). En

-H (altura del talud en metros), diferente en cada uno de los casos.

- $\gamma$  (peso específico del terreno en  $t/m^3$ ), se ha obtenido experimentalmente el valor  $1,97 t/m^3$ .

- $\psi_t$  (ángulo o pendiente del talud), diferente en cada uno de los casos.

- $\phi'$  (ángulo de rozamiento interno efectivo), se han escogido 20° (valor típico de arcillas de alta plasticidad) y 32° (valor típico de arcillas de plasticidad media-baja).

-c' (cohesión efectiva del terreno en  $t/m^2$ ), se han tomado diferentes valores entre 0,5 y 10  $t/m^2$  correspondientes a los valores mínimo y máximo de cohesión para las arcillas dependiendo del grado de saturación de las mismas.

El cálculo de la estabilidad se ha realizado para condiciones de terreno seco, terreno saturado por debajo de la grieta de

condiciones secas y saturadas por debajo de la grieta de tracción se han tomado valores de cohesión entre 4 y 10 t/m<sup>2</sup>. En condiciones completamente saturadas se han utilizado valores de cohesión entre 0,5 y 3 t/m<sup>2</sup>.

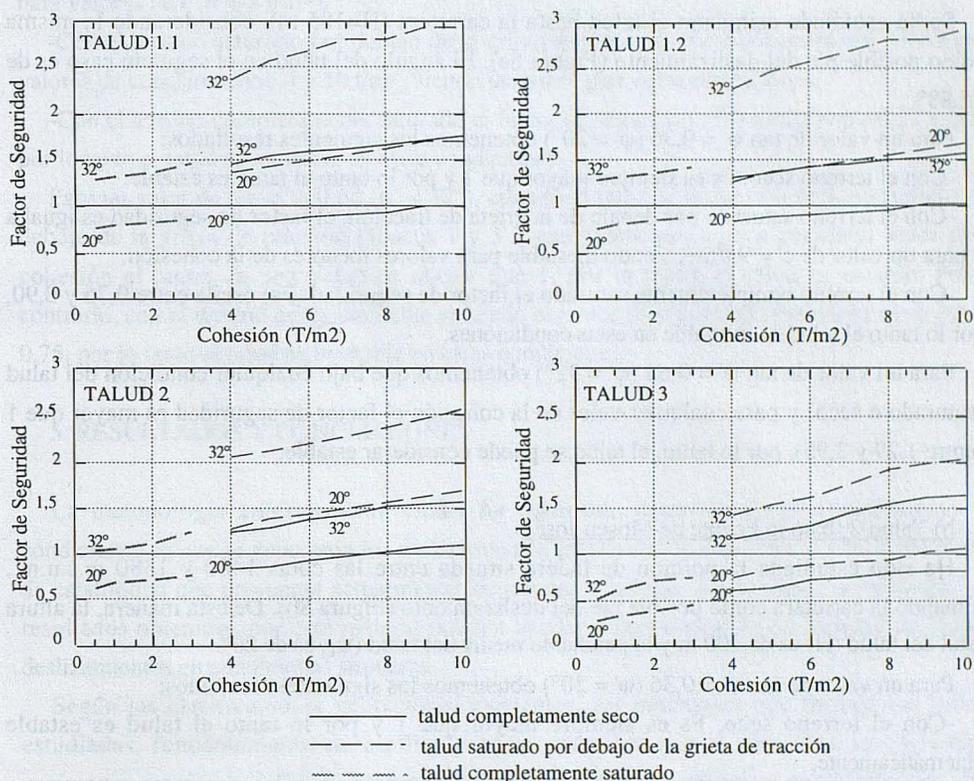


Figura 7. Relaciones Cohesión / Factor de Seguridad para cada uno de los taludes estudiados.

**a) Talud 1 (Barranco de la Canaleja)**

Se sitúa entre las cotas 1.525 (contacto con los niveles carbonatados) y 1.200 m.s.n.m. (base del barranco), obteniendo una altura total del talud (H) de 325 m. En este caso, la pendiente del talud ( $\psi_t$ ) es de 15,15°.

Para un valor de  $\tan \phi' = 0,36$  ( $\phi' = 20^\circ$ ) obtenemos los siguientes resultados:

-Con el terreno seco,  $F_s$  es siempre mayor que 1 y por lo tanto el talud es estable cinemáticamente.

-Con el terreno saturado por debajo de la grieta de tracción,  $F_s = 1$  para un valor de  $c' = 9$  t/m<sup>2</sup>, siendo inestable para valores menores de la cohesión.

-Con el terreno completamente saturado el factor de seguridad ( $F_s$ ) oscila entre 0,76 y 0,87, por lo tanto el talud debe considerarse inestable en estas condiciones.

Para un valor de  $\tan \phi' = 0,62$  ( $\phi' = 32^\circ$ ) obtenemos que bajo cualquier condición del talud (saturado o seco) y para cualquier valor de la cohesión el factor de seguridad es mayor que 1,

por lo tanto, el talud es estable para ese valor del ángulo de fricción (condiciones más favorables).

Se ha estudiado asimismo el talud hasta la carretera ( $H=185$  m), considerando la misma como posible pie del deslizamiento (Figura 8a). El ángulo del talud en el segundo caso es de  $15,88^\circ$ .

Para un valor de  $\tan \varphi' = 0,36$  ( $\varphi' = 20^\circ$ ) obtenemos los siguientes resultados:

- Con el terreno seco,  $F_s$  es siempre mayor que 1 y por lo tanto el talud es estable.
- Con el terreno saturado por debajo de la grieta de tracción, el factor de seguridad es igual a 1 para un valor de  $c' = 9 \text{ t/m}^2$ , siendo inestable para valores menores de la cohesión.
- Con el terreno completamente saturado el factor de seguridad ( $F_s$ ) oscila entre 0,76 y 0,90, por lo tanto el talud es inestable en estas condiciones.

Para un valor de  $\tan \varphi' = 0,62$  ( $\varphi' = 32^\circ$ ) obtenemos que bajo cualquier condición del talud (saturado o seco) y para cualquier valor de la cohesión el factor de seguridad es mayor que 1 (entre 1,29 y 2,95), por lo tanto, el talud se puede considerar estable.

#### b) Talud 2 (bajo la Fuente de Mosen José)

Ha sido estudiada la porción de ladera situada entre las cotas 1480 y 1380 m.s.n.m., tomando la carretera como posible pie del deslizamiento (Figura 8b). De esta manera, la altura total del talud ( $H$ ) es de 100 m y la pendiente media del talud ( $\psi_t$ ) es de  $20^\circ$ .

Para un valor de  $\tan \varphi' = 0,36$  ( $\varphi' = 20^\circ$ ) obtenemos los siguientes resultados:

- Con el terreno seco,  $F_s$  es siempre mayor que 1 y por lo tanto el talud es estable cinemáticamente.
- Con el terreno saturado por debajo de la grieta de tracción, el factor de seguridad es 1 para un valor de  $c' = 8 \text{ t/m}^2$ , siendo inestable para valores menores de la cohesión.
- Con el terreno completamente saturado el factor de seguridad ( $F_s$ ) oscila entre 0,62 y 0,76, por lo tanto el talud es inestable en estas condiciones.

Para un valor de  $\tan \varphi' = 0,62$  ( $\varphi' = 32^\circ$ ), bajo cualquier condición del talud (saturado o seco) y para cualquier valor de la cohesión el factor de seguridad es mayor que 1, por lo tanto, el talud se puede considerar estable. No obstante, los valores del factor de seguridad se aproximan a 1 con el terreno totalmente saturado, existiendo posibilidad de deslizamiento si la cohesión alcanza valores menores de  $0,5 \text{ t/m}^2$  (situación más desfavorable).

#### c) Talud 3 (al Oeste de la Peña Blanca)

Ha sido estudiada la porción de ladera situada entre las cotas 1460 y 1420 m.s.n.m., tomando la carretera como posible pie del deslizamiento (Figura 8c). De esta manera, la altura total del talud ( $H$ ) es de 40 m y la pendiente del talud  $\psi_t = 36,67^\circ$ .

Para un valor de  $\tan \varphi' = 0,36$  ( $\varphi' = 20^\circ$ ) obtenemos los siguientes resultados:

-Con el terreno seco,  $F_s$  es mayor que 1 para valores de cohesión iguales o superiores a  $9 \text{ t/m}^2$ , debiendo considerar el talud inestable con valores menores de cohesión ( $F_s = 0,75-0,94$  para valores de  $c' = 4-8 \text{ t/m}^2$ ).

-Con el terreno saturado por debajo de la grieta de tracción,  $F_s$  oscila entre 0,6 y 0,85 para valores de cohesión entre 4 y  $10 \text{ t/m}^2$ , siendo inestable para estas condiciones.

-Con el terreno completamente saturado el factor de seguridad ( $F_s$ ) oscila entre 0,27 y 0,45, por lo tanto el talud es inestable en estas condiciones.

Para un valor de  $\tan \phi' = 0,62$  ( $\phi' = 32^\circ$ ), cuando el talud se encuentra seco o saturado por debajo de la grieta de tracción (ábacos 1 y 3 respectivamente) y para cualquier valor de la cohesión el factor de seguridad es mayor que 1, por lo tanto, el talud es estable. Por el contrario, con el terreno completamente saturado el factor de seguridad ( $F_s$ ) oscila entre 0,51 y 0,75, por lo tanto el talud es inestable en estas condiciones.

## 5. RESULTADOS Y CONCLUSIONES

La metodología anteriormente citada ha permitido determinar aproximadamente las condiciones en que se generaron los deslizamientos estudiados en este trabajo y las condiciones de estabilidad que presentan actualmente los taludes y laderas que bordean la carretera. Los resultados obtenidos por este método pueden ayudar a pronosticar la posibilidad de nuevos deslizamientos en condiciones similares.

Según las clasificaciones geotécnicas existentes, los materiales que forman las laderas estudiadas, fundamentalmente arcillas y yesos, para los que se han obtenido diferentes parámetros mecánicos y físicos, presentan como principales características geotécnicas: un valor regular a malo como cimiento de carretera en ausencia de heladas; una respuesta media a muy alta ante la acción potencial de las heladas, frecuentes en la zona; propiedades de retracción o entumecimiento ligeras-medias; y unas características de drenaje regulares a malas.

La existencia de importantes acuíferos en los niveles basales del Jurásico favorece el alto grado de humedad de los materiales infrayacentes y con éste, el aumento de la plasticidad de los mismos, causante de numerosos fenómenos de inestabilidad en las laderas estudiadas. La aparición de grietas transversales a la pendiente de la ladera, que afectan incluso al firme de la carretera, demuestra que estos fenómenos se encuentran activos en la actualidad.

Mediante la utilización de los ábacos de Hoek y Bray se puede constatar (Tabla 3) que para un valor del ángulo de rozamiento interno de las arcillas de  $20^\circ$  los tres taludes estudiados son inestables, en condiciones saturadas, o en condiciones de saturación por debajo de la grieta de tracción para cohesiones efectivas menores de  $8-9 \text{ t/m}^2$ . En condiciones secas, los taludes 1 y 2 son estables mientras que el talud 3 es inestable para cohesiones menores de  $9 \text{ t/m}^2$ . Con un ángulo de rozamiento interno de  $32^\circ$ , los taludes 1 y 2 son estables bajo cualquier condición del terreno mientras que el talud 3 es inestable en condiciones de terreno completamente saturado.

Condiciones geotécnicas	Condiciones de saturación	Talud 1.1	Talud 1.2	Talud 2	Talud 3
$\phi' = 20^\circ$	Saturado	Inestable	Inestable	Inestable	Inestable
	Saturado debajo grieta tracción	Estable / Inestable*	Estable / Inestable*	Estable / Inestable**	Inestable
	Seco	Estable	Estable	Estable	Estable / Inestable*
$\phi' = 32^\circ$	Saturado	Estable	Estable	Estable	Inestable
	Saturado debajo grieta tracción	Estable	Estable	Estable	Estable
	Seco	Estable	Estable	Estable	Estable

Nota: \*:  $c' < 9 \text{ t/m}^2$

\*\* :  $c' < 8 \text{ t/m}^2$

Tabla 3. Resumen de las condiciones de estabilidad obtenidas a partir de los ábacos de Hoek y Bray para los taludes estudiados en distintas condiciones de saturación en agua .

### AGRADECIMIENTOS

Este trabajo ha sido realizado gracias a una Ayuda de Investigación concedida por el Instituto de Estudios Turolenses (C.S.I.C.) y al Proyecto PB93-1218 (DGICYT).

### BIBLIOGRAFIA

- AYALA, F.J.; ANDREU, F.J.; FE, M.; FERRER, M.; DE SIMON, A.; FERNANDEZ, I.; OLALLA, C.; GOMEZ, J.; SAMPEDRO, J. Y CIENFUEGOS, F.J. (1991): *Manual de Ingeniería de taludes*. ITGE, Madrid: 456 p.
- CALVO-CASES, A.; GUTIÉRREZ-ELORZA, M.; PEÑA-MONNÉ, J.L. Y SIMÓN-GÓMEZ, J.L. (1983): Morfología de vertientes y neotectónica en el macizo de Javalambre (Provincia de Teruel). *VI Reunión Grupo Español del Cuaternario, Galicia. Comunicaciones*: 429-448.
- HOEK, E. Y BRAY, J.W. (1977): *Rock Slope Engineering*. The Institution of Mining and Metallurgy, London: 402 p.
- LIESA, C.L. (1993): *Análisis de la deformación frágil en las áreas de mosqueruela y la puebla de Valverde (Teruel): Fases tectónicas y perfurbaciones de esfuerzos*. Tesis de Licenciatura, Universidad de Zaragoza: 207 p.
- PEÑA, J.L., GUTIÉRREZ, M., IBÁÑEZ, M.J., LOZANO, M.V., RODRÍGUEZ, J., SÁNCHEZ, M., SIMÓN GÓMEZ, J.L., SORIANO, M.A. Y YETANO, L.M. (1984): *Geomorfología de la provincia de Teruel*. Instituto de Estudios Turolenses, Teruel: 149 p.
- SIMÓN GÓMEZ, J.L. (1984): *Compresión y distensión alpinas en la Cadena Ibérica Oriental*. Tesis Doctoral, Univ. Zaragoza. Pub. Instituto de Estudios Turolenses, Teruel, 269 p.
- SOLÉ SABARÍS, L. (1978): La Meseta. En: *Geografía de España*, (M. De Terán, Ed.). Ariel, Madrid: 42-62.
- VIRGILI, C.; SOPEÑA, A.; RAMOS, A. y HERNANDO, S. (1977): Problemas de la cronoestratigrafía del Triás en España. *Cuadernos de Geología Ibérica*, 4: 57-88.

## **A Model for the Formation of Ultramafic Rocks in the Crust: Peridotites of Southern Spain and Northern Morocco.**

**V. Sánchez Cela**

Dept. de Geología, Universidad de Zaragoza (Spain)

**A. Aparicio Yagüe**

Dept. de Volcanología, Museo Nacional de Ciencias Naturales, CSIC, Madrid

### **Summary**

The peridotites and related rocks of southern Spain and northern Morocco were formed through polyphasic processes by sialic matter than, at different P-T conditions, transformed sedimentary-epimetamorphic materials.

The sialic matter is related to the origin of gneissic rocks that form the "basement" of the Gibraltar Arc. Such gneissic rocks correspond to the greatest dynamic (P-T) and chemical manifestations that partially outlined a great non outcropping domed granite mass during its origin and growth ("emplacement") in the westernmost part of the Mediterranean Sea.

These gneissic rocks represent active sialic chemical matter, which, at moderate to high P-T conditions, transformed sedimentary epimetamorphic materials of an argillaceous to carbonate nature during Alpine times.

The transformation of argillaceous and marly-argillaceous materials by silica-alkaline matter under moderate P-T conditions produced different gneissic rocks of acid to intermediate compositions (e.g. granitic gneisses, cordieritic gneisses, kinzigites, amphibolites). Gabbroid and dioritic rocks were produced by the transformation of marly materials, whereas peridotites were produced from the marly-dolomite materials. Marbles and calc-silicate rocks correspond to the "impermeable fronts of transformation" of the upper dolomite-limestone Triassic-Jurassic sequence.

The geochemical balance between the sialic endogeneous matter and the nature of wall sedimentary rocks were the main cause of the origin and diversity of the ultramafic rocks (e.g. dunites, lherzolites, harzburgites, pyroxenites, etc).

The physico-chemical behaviour of the Mesozoic wall rocks was also the main cause of the differentiation of the igneous-metamorphic association. So, the granular-granoblastic peridotites resulted from the transformation of the marly-dolomite rocks, whereas the gneissic rocks were produced by the transformation, under stress conditions, of the lutitic-greywackic associated rocks. The lightly schistose intermediate rocks (e.g. diorite-amphibolite) resulted from the transformation of the intermediate marly-greywackic sequence.

## 1. INTRODUCTION

The ultramafics in southern Spain are mainly formed by three massifs: Ronda, Ojen and Carratraca. In northern Morocco, Beni Bousera is the main outcrop. Together with these four main massifs, are others of lesser importance (Sierra de las Aguas, Sierra de la Robla, Sebta, Beni Malek). (Fig. 1).

The lherzolites are the most important type of peridotites. Harzburgites, garnet-pyroxenites, olivine-gabbros are other associated ultramafic-mafic rocks (Hernández Pacheco, 1967; Kornprobst, 1969; Obata, 1980; Reuber et al., 1982). Within lherzolites, three types can be differentiated: garnet-lherzolites, spinel-lherzolites and plagioclase-lherzolites. A gradual transition exists between these lherzolites and other ultramafic to mafic rocks (e.g. gabbros, mafic gneisses, amphibolites, etc).

All the current-petrogenetic models on the origin of the ultramafics rocks in southern Spain and northern Morocco start from an indubitable upper mantle provenance. But the mechanisms concerning the emplacement of these "mantle" rocks in the upper crust are very diverse. So, as a summary we can cite:

- Mantle diapirism (Loomis, 1972, 1975; Frey et al., 1985).
- Unrooted mantle slab (Lundeen, 1978; Tubia & Cuevas, 1987).
- Subsidiary unroofing and polyphasic emplacement (Reuber et al., 1982; Van der Wall & Vissers, 1993).
- Subcontinental lithosphere delamination (Platt & Vissers, 1989).
- Allochthonous sheets detached from a diapir (Casas & Carbó, 1990).
- Detachment of a subducted slab together with mantle diapirism (Zeck et al., 1992).
- Unrooted peridotites slab under extension (Torné et al., 1992).
- Extensional detachment of mantle core complexes (Doblas & Oyarzun, 1989)

Contrary to mantle provenances Van Bemmelen (1933) and Termier & Termier (1956) interpreted the Ronda peridotites as formed by metasomatic transformation processes of sedimentary carbonate rocks. This theory was not taken into account because from a petrogenetic point of view it was not sufficiently founded, and also because geologists always identify the mafic to ultramafic rocks, without any doubt, as being rocks from the upper mantle provenance.

## 2. SOME GEOLOGICAL DATA

The main data in proposing a new petrogenetic model for southern Spain ultramafics were mainly deduced from field-geological data complemented with petrological and some geochemical studies in those outcrops that we have defined to be the most suitable. In these outcrops an evident gradual transition exists between the ultramafics and the "wall rocks".

Together with this important observation, other also important data are:

- The ultramafics only exist when a thick carbonate cover is or was present
- Such rocks only exist when together with the carbonate cover an active gneissic basement is present
- Geophysical-geological data indicate that: a) these ultramafics do not exceed 2 km in thickness; b) In spite of their polyphasic character they form autochthonous outcrops, "without roots".

The best places to study the petrogenesis of the ultramafic rocks are those where the nearest original petrogenetic conditions are better conserved, that is, where they were hardly

altered by the polyphasic tecto-petrogenetic Alpine processes as it occurs in Sierra Blanca. Sierra Blanca is a domed structure covered by about 500 to 800 m of Jurassic carbonate rocks, where the ultramafics and related rocks only crop out at the border of this dome and at some faulted zones of the interior of this dome. The presence of peridotites beneath the thick carbonate cover of Sierra Blanca was mainly deduced by the existence of marbles and gneissic rocks with very similar geological features to other zones with outcropping peridotites. This presence was already deduced by: Blumenthal (1930); Loomis (1975); Kornprobst (1974); Torres Roldán (1979); Sanz de Galdeano & Andreo (1995).

In many places the contact-relationships between the different rocks are altered by later tecto-petrogenetic episodes. A good example of this type of outcrop can be observed at the north of Sierra Bermeja (southern Ronda village).

In other outcrops the contact-relationship between the carbonate cover and ultramafic rocks was obliterated by later tectonic processes. Only some ultramafic-gneissic relationships are present. As an example we can cite the Beni Bousera massif. In Sierra de Alpujata the ultramafic rocks are not covered by carbonate rocks, because they were thrust aside by later Alpine tectonic processes.

In all these types of outcrops the ultramafic and mafic associated rocks are always found lying on a gneissic basement with clear petro-structural relationships.

In many outcrops the ultramafic and associated rocks are thrusting younger partially uprooted stratigraphic sequences from their original basement rocks. This produces an apparent allochthonous character of these ultramafics causing some geologists to consider such ultramafics to be allochthonous of a provenance was related to a supposed "oceanic" zone not present in the western Mediterranean.

In the best outcrops (e.g. Sierra Blanca) the following stratigraphic column can be deduced from top to bottom (Fig. 2A).

- A thick Jurassic carbonate cover, that can reach 650 m in thickness, formed by dolomites and Mg-limestones differentially recrystallized and metamorphized (marbles and calc-silicate rocks). In some zones these rocks show calc-silicate minerals. These latter rocks, that constitute the cover of the underlying ultramafic rocks, were very important in the "petrogenetic efficiency" of the global igneous-metamorphic association in such suitable domed outcrops.

- Below these, in a rapid transition, are the ultramafic rocks, which are very diverse in structure and composition due to the "storage" of several petrogenetic episodes (Jurassic to Miocene times). The spinel peridotites (harzburgites-lherzolites) are the most common rocks in contact with calc-silicate and marble rocks. The garnet and plagioclase peridotites, appear to be below the previous ones. All these ultramafic rocks, in spite of the evidence of later tecto-petrogenetic processes, exhibit low anisotropic structures-textures although the blastic character can be important.

- In a gradual manner these ultramafic rocks transit into various mafic rocks, formed mainly by plagioclase-peridotites, and various gabbros. These intermediate rocks exhibit some foliation features that define such rocks as "schistose gabbros", "mafic gneisses" and "amphibolites". These moderate foliated intermediate rocks, in isolated studies, could be classified either as igneous or metamorphic rocks and the latter, from ortho to paraderivate origin. The structural and compositional differentiation is mainly controlled by the original nature of the wall-sedimentary sequence.

- These "igneous-metamorphic" intermediate rocks gradually transit, towards the base, into more leucocratic and more foliated rocks, mainly various gneisses, and from top to bottom we can differentiate the following:

- a) Cordierite-orthopyroxene gneisses
- b) Cordierite-sillimanite-garnet-gneisses
- c) Sillimanite-garnet-gneisses, and finally forming the basement
- d) Gneisses s.l. having scarcely or no metamorphic minerals.

Between sedimentary, igneous and metamorphic rocks, not only is there a gradual transition between the different petrographic types of rocks but also a gradual concordant structuration between the gneissic basement and many mafic-ultramafic rocks of the cover. Field geological data indicate that the weak foliation and folds of the peridotites is concordant with the main foliation (schistosity) of the basal gneisses, generally arranged according to oval structures. This structuration is concordant from macro to microscale. To a macroscale, both gneissic and mafic-ultramafic rocks, are arranged according to arc morphologies (fig. 1). Within this general structuration the gneisses and their associated mafic-ultramafic rocks show, together with concordant domed morphologies, a concordant regional oval foliation associated with evident compressional features. So, many basement gneisses correspond to blastomylonitic gneisses in gradual transition to other blastomylonitic rocks. Although some blastomylonitic structures could be associated with later Alpine episodes. Other previous structures (e.g. oval schistosity), are recognized as early structures developed during the formation of the various gneisses and of their associated mafic-ultramafic rocks.

### 3. NEW PETROGENETIC MODEL

#### 3.1. Necessary Conditions

Four main suitable conditions were necessary for the formation of ultramafics and related rocks:

- 1<sup>st</sup>- The existence of an active sialic basement which during Alpine sialic thickening, and during the formation of the blastomylonitic gneisses, released sialic matter at moderate-high T conditions and under moderate stresses.
- 2<sup>nd</sup>- The presence, on this active sialic basement, of a thick sedimentary cover formed by suitable materials to be transformed by the sialic matter and arranged in suitable stratigraphic sequences (e.g. lutites-marls-limestones).
- 3<sup>rd</sup>- The formation of oval-domed impermeable covers (carbonate to marble rocks) that enhanced the physico-chemical conditions and therefore the transformation processes in the underlying levels (marly dolomites).
- 4<sup>th</sup>- The permanence of all the cited suitable conditions over a long period of time.

Many gneissic rocks, mainly the blastomylonitic types were formed through several petrogenetic episodes by the transformation of argillaceous-sedimentary-epimetamorphic country rocks. These rocks were affected by silicification and to a minor degree feldspathization processes at moderate-high P-T conditions induced by an active granitic-gneissic basement. During these dynamic-petrogenetic processes the blastomylonitic granites can release fluids which at moderate-high temperatures are efficient mechanisms of transport of many chemical elements, mainly of silica and also alkalis (Etheridge et al., 1983; Sinha et al., 1986; O'Hara, 1988).

This "granitic" matter was released in relation to the formation of a polyphase gneissic orogenic-arc that from North Africa (Beni Bousera) and through the Straits of Gibraltar extend over southern Spain (Ronda-Ojén zone).

We relate the origin of the sialic matter, mainly silica, released from the sialic (granitic-gneissic) basement, to the "prime" cause that governs many geological phenomena: "sialic thickening". We understand sialic thickening as the episodic perhaps cyclic, creation of new sialic matter, mainly at the base of the crust, produced by the transformation of the denser upper mantle matter into a lighter sialic one. This involves a new understanding for the upper mantle (Sánchez Cela, 1990). This episodic juvenil contribution to the crust from the upper mantle, mainly in solid state, took and is taking place according to domed to oval morphologies. These structures are those that demand lesser energy for their formation and emplacement.

At the border of these active sialic domes, mainly under high compressional stresses, is where the most important energetic and chemical processes take place. Such places are generally formed by dynamic granitic rocks (gneisses) that are important endogenic sources. At the top of these arc-orogenic structures important petrogenetic processes can take place if such structures are covered by thick carbonate materials (limestones and dolomites).

These carbonate covers, with suitable properties (ductibility, impermeability), act as impermeable barriers to the mobility of the endogenic sialic matter under moderate P-T conditions. This produces an increase in volume that results in an increase in the pressure, mainly in the nucleus of the domes.

Together with this evident increase in volume of the sialic crust a great accumulation of fluids also takes place. These fluids, mainly CO<sub>2</sub>, were released during the formation of the ultramafic-mafic rocks by the transformation of CO<sub>2</sub>-bearing rocks (dolomitic-limestone rocks). If these fluids do not escape, local overpressures, that greatly overpass the lithostatic conditions, can be formed mainly in the nucleus of the domes.

So, the total pressure in a nucleus of a dome can be approximately deduced from:  $P_t = P_l + P_s + P_f$ , where  $P_t$  = total pressure;  $P_l$  = lithostatic pressure;  $P_s$  = stress pressure;  $P_f$  = fluid pressure.

The lithostatic pressure is, in our opinion, the lesser important factor-pressure on such point x, although this "static pressure" is, in many cases, the only pressure considered by many authors.

The stress-pressure together with fluid-pressure are the most important factors in increasing the pressure in such places. These P stress and P fluid "dynamic pressures" together with the formation of new igneous-metamorphic rocks, below the carbonate domes, can together raise the local pressure conditions at very shallow levels of the crust (e.g. 2 km depth).

The formation of moderate and even high temperatures in such high levels of the crust can be explained by the contribution of various thermal sources:

- 1<sup>st</sup>- By the heat associated with the origin of the granitic-gneissic rocks that constitute the basement of ultramafics.
- 2<sup>nd</sup>- By the heat released at shear friction zones during the dynamic emplacement of the gneissic rocks.
- 3<sup>rd</sup>- By the differential thermal behaviour of the sedimentary cover in relation to the gneissic basement and even between Triassic-Jurassic sedimentary materials.
- 4<sup>th</sup>- By the storage of these thermal sources in suitable places, that is, in the marls covered by the "impermeable" dolomitic-limestone domes.

The most important thermal source in the origin of peridotites is related to the formation of blastomylonitic gneissic rocks and shear friction processes, where the conversion

of mechanical energy into heat can be very important (Graham & England, 1976; Brun & Cobbold, 1980; Lockett & Kuszniir, 1981; England & Thompson, 1984; Molnar & England, 1990; England & Molnar, 1993). This dynamic-thermal energy was considered by some geologists in the petrogenesis of the metamorphic rocks, associated with peridotites (Woodcock & Robertson, 1977).

In order to be efficient the thermal energy must be stored in suitable petrological-structural places, that is, in the nucleus of domed sedimentary rocks formed by Triassic marls covered by thick Jurassic carbonates. In such places, because of the low-conductivity-diffusivity properties of the marls, they act as thermal accumulators. This, can produce high temperatures if a more or less continuous thermal energy is released from the sialic basement. So, from this manner together with moderate-high pressures, moderate-high temperatures can also be achieved in relatively shallow levels of the crust.

The composition and structure of the sedimentary rocks together with the vertical injection of sialic matter, under high compressional environments (orogenic arcs), is the main cause of the petrological differentiation from the bottom (gneisses) to the top (ultramafics). So, at the base, where the argillaceous rocks are abundant, micaceous minerals are formed (biotite, moscovite). Their anisotropic crystalline morphologies develop schistose-gneissic foliations during the injection and deformation of the sialic endogenic matter under moderate-high compression. On the contrary, at the top, the sedimentary materials mainly formed by dolomite-marls, give origin to mafic minerals (pyroxenes, olivines) with lower crystalline anisotropic properties.

The differential behaviour of the Triassic-Jurassic sedimentary sequence (see Fig. 3) could explain the formation of inverted metamorphic gradients that are generally present below many peridotites and ophiolites in the world (Woodcock & Robertson, 1977; Reuber et al., 1982; Searle & Malpas, 1982; Burg et al., 1989; Waters, 1989; England & Molnar, 1993; Gnos & Kurz, 1994, etc). These inverted metamorphic series are generally interpreted as allochthonous processes associated with collisional (Jaupart & Provost, 1985; Jain & Manickavasagan, 1993), thrusting (Patrick et al., 1994) and also subduction processes (Peacock, 1987).

### **Vertical petrogenetic differentiation**

In an initial Alpine petrogenetic episode and due to the active conditions created by the sialic basement, a global and almost simultaneous transformation took place of a great part of the Triassic-Jurassic stratigraphic column. So, together with the formation of marbles and some calc silicate rocks and various ultramafic-mafic rocks at the top, mafic to acid gneisses and schists were formed below (Fig. 2B).

The petrogenetic interpretation of the cal-silicate rocks and gneissic rocks was very important in deducing the thermal and chemical endogenic sources for the origin of the ultramafics and associated rocks.

The formation of cal-silicate rocks, formed for example by carbonates, diopside, wollastonite, and other associate minerals is easily explained by "contact metamorphism" of carbonate rocks by "sialic matter" (granitic) at moderate temperature.

Within the petrogenetic transformation processes the calsilicate and associated marble rocks were the first to be produced. This is important because these metamorphic rocks are more chemically and thermally impermeable than their sedimentary protoliths. This produced an increase in the efficiency of the petrogenetic processes at lower levels, that is, in the formation of ultramafic-mafic rocks by the transformation of dolomite marls, and of basic to

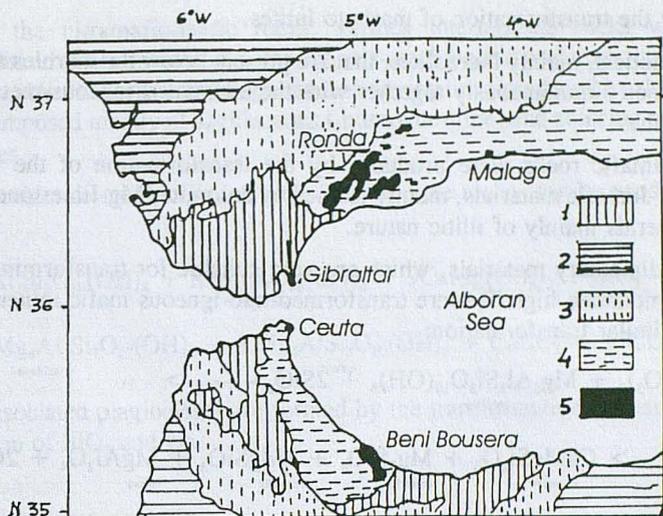


Fig 1 Geological sketch map of southern Spain and northern Morocco. 1 = Cretaceous-Tertiary flysch. 2 = Marly-carbonate rocks (Mesozoic). 3 = Carbonate rocks (Mesozoic). 4 = Metamorphic basement (Paleozoic) with a carbonate Mesozoic cover. 5 = Ultramafic rocks.

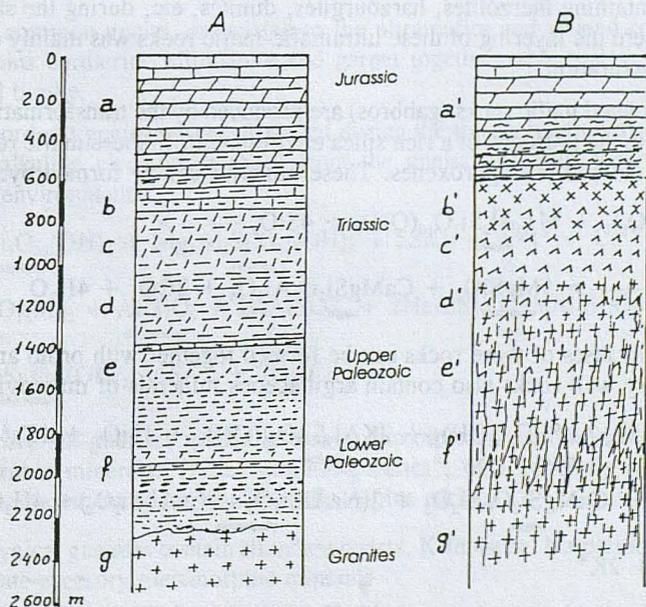


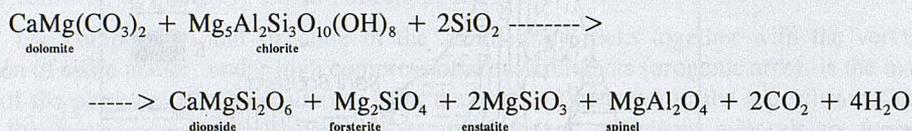
Fig 2 Lithological correlations between the original sedimentary sequence. (A) (after Azema, 1961) and the present igneous and metamorphic sequence (B). a = Carbonate rocks (dolomites); b = Marly dolomites; c = Marls; d = Marly lutites; e = Lutite-greywackes; f = Greywackes with some carbonate rocks; g = Granitic basement. a' = Carbonate rocks (remains) on marbles calc-silicate rocks; b' = Ultramafic rocks; c' = Mafic rocks; d' = Dioritoid-amphibolitic rocks (basic gneisses); e' = High P-T gneisses; f' = Orthogneisses; g' = Reactivated granite basement.

acid gneisses by the transformation of marls to lutes.

The peridotites, mainly lherzolites, that are present below the marbles and calc-silicate rocks, were formed simultaneously together with the associated igneous-metamorphic rocks during Alpine times.

The ultramafic rocks were produced by the transformation of the marly-dolomite Upper Paleozoic-Jurassic materials, mainly formed by dolomite, Mg-limestone, chlorites, and argillaceous minerals mainly of illitic nature.

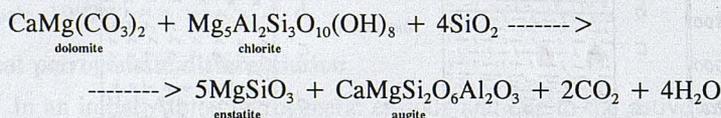
These sedimentary materials, which are very suitable for transforming into a Si-rich environment at moderate-high T, were transformed into igneous mafic minerals through the following or a similar transformation:



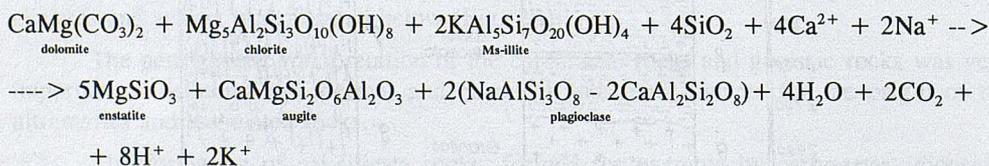
This mineral paragenesis correspond to a lherzolite, the most abundant ultramafic rock in southern Spain.

The different composition of the sedimentary levels (carbonate to marls), together with the differential mobility of chemical elements under moderate-high P-T conditions, could be the main cause of the formation of layering structures in some ultramafic rocks. So, levels were formed containing lherzolites, harzburgites, dunites, etc, during the same petrogenetic episode, and where the layering of these ultramafic-mafic rocks was mainly controlled by the sedimentary structuration.

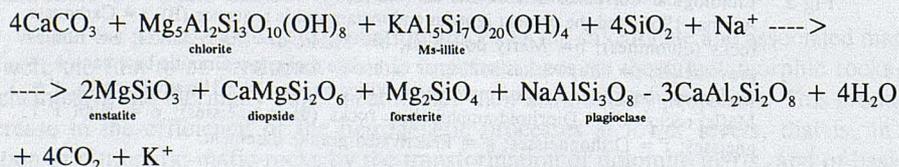
The associated mafic rocks (gabbros) are produced by the transformation of underlying dolomitic marls in the presence of a rich silica environment. In these mafic rocks the presence is frequent of clino and orthopyroxenes. These minerals can be formed by:



The plagioclases of these rocks can be formed together with ortho and clinopyroxene if the sedimentary wall rocks also contain argillaceous minerals of muscovite-illite type



The olivine gabbros associated with lherzolites can be formed by:







that demand less energy for their vertical displacements (Sánchez Cela, 1990). The presence of an extensional zone (Alboran Sea) and compressional orogenic environments (Rif-Betics) can easily be explained by the formation and growth of the Alboran domed granitic mass in western Mediterranean.

The intriguing circular shape of thrusting with the distribution of the associated compressional and extensional environments (Seber et al., 1996) can be explained by our model, that is, by an "in situ" growth of a domed-mushroom granitic mass in such a zone.

Peridotites either alone or associated with other igneous rocks, mainly in Ophiolite complexes, are widely present in many Alpine zones, as occurs in the Tethyan belt from southern Spain to Indonesia: Alps (France-Italy); Dinarides (Yugoslavia); Greece-Aegean Sea, Cyprus, Turkey, Oman, Pakistan, Himalaya until Indonesia. In these places the peridotites or ophiolites are always associated with the presence of orogenic arcs formed by active gneissic-migmatic-mylonitic rocks, and where thick Mesozoic carbonate covers are or were present during the formation of ultramafic and associated igneous rocks.

In all these zones the basement gneissic rocks of the orogenic arcs are possibly partial dynamic manifestations of upper mantle-crustal granitic domes in different stages of growth. So, for example the Hellenic-Aegean arc, the Semail (Oman) arc, the Ronda-Beni Bousera arc and the Western Alps arc and Himalayas correspond to orogenic sialic arcs with crusts from thinner to thicker.

The association of peridotites-serpentinites with "oceanic" to continental orogenic arcs, whose basements are formed by high P-T schist-gneisses, had already been stated by several geologists many years ago (e.g. Hess, 1955).

From the constant association between ultramafic rocks and arc-gneissic basements we deduce that the "oceanic" peridotite-ophiolites are on thin arc-gneissic rocks, which are partially outlining thin domed granitic rocks in various stages of growth.

Similar petrogenetic environments are or were present in some Hercynian ultramafics. So, our knowledge of the Cabo Ortegal (Spain) ultramafics have led us to deduce that such igneous rocks were produced by the conjunction of two favourable conditions: an active Hercynian gneissic basement and the presence of Paleozoic carbonate materials (Aparicio et al., 1987).

Geological-petrological data indicate the existence of a spatial and temporal relationship between the formation of Barrovian and Contact metamorphic rocks in relation to compressional or extensional environments created during the origin and growth of the deduced Alboran granite dome. These metamorphic types together with other intermediate ones, were affected by later tectono-petrogenetic Alpine episodes which partially obliterated the original metamorphic relationship.

An approach to this global relationship between the deduced "Alboran granite dome", formation of the Ronda-Beni Bousera orogenic arcs, igneous and metamorphic rocks, is schematized in the Fig. 3.

As a final consideration we think that the "immutable law" that the ultramafic-mafic rocks are always from upper mantle provenance has impeded us geologists and petrologists from seeing and describing "the true geological environment" of such igneous rocks in many outcrops, mainly in those defined for us as lesser altered by later tectono-petrogenetic processes. Among these more favorable outcrops are the southern Spain and northern Morocco considered here, where many of the original suitable geological conditions for the formation of these ultramafic to mafic rocks are conserved, that is, the presence of an active gneissic

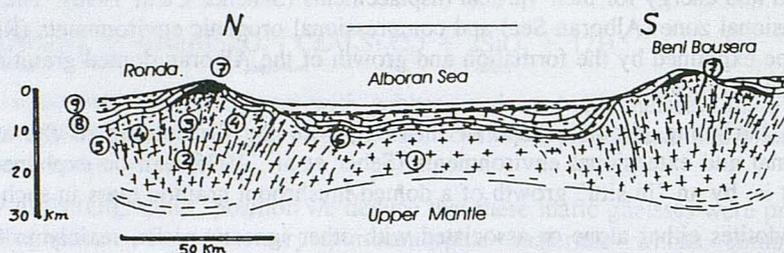


Fig 3 Crustal geological interpretation through Ronda - Beni-Bousera cross-section (see Fig 1). 1 = Extensional granites; 2 = Compressional granites (gneisses); 3-4 = Gneissic-schist rocks (Regional metamorphism); 5 = Sedimentary-epimetamorphic rocks; 6 = Contact metamorphic rocks; 7 = Ultramafic-mafic rocks; 8 = Jurassic carbonate rocks; 9 = Cretaceous-Tertiary flysch.

basement covered by thick carbonate materials in suitable structures (arcuate-oval archs) stable for a long time.

The old idea that the contact between the gneissic basement and the ultramafic-mafic cover is an "inverted paleo-Moho discontinuity" must be therefore ruled out.

## 5. REFERENCES

- Andrieux, J.; Fontbote, J.M. & Mattauer, M. (1971). Sur un modèle explicatif de l'Arc de Gibraltar. *Earth Planet. Sci. Lett.*, 12, 191-198.
- Aparicio, A.; Sánchez Cela, V.; García Cacho, L. (1987). Petrological and geochemical considerations on the Cabo Ortegal complex (NW Spain). *Rev. Acad. Ciencias Zaragoza, Spain*, 42, 131-161.
- Banda, E.; Gallart, J.; García Dueñas, V.; Dañobeitia, J.J. & Makris, J. (1993). Lateral variation of the crust in the Iberian peninsula: new evidence from the Betic Cordillera. In: *Seismic Studies of the Iberian Peninsula*. J. Badal et al. (eds). *Tectonophysics*, 221, 53-66.
- Biermann, C. (1995). The Betic Cordilleras (SE Spain). Anatomy of a dualistic collision-type orogenic belt. *Geol. Mijnbouw*, 74, 167-182.
- Blumenthal, M.M. (1930). Beiträge zur Geologie der betischen Kordilleren beiderseits des Rio Guddalhorce. *Eclogae geol. Helv.*, 23, 41-293.
- Bouillin, J.P.; Duran-Delga, M. & Olivier, Ph. (1986). Betic-Rifian and Tyrrhenian Arcs: Distinctive features, genesis and development stages. In: *The Origin of Arcs*. F.-C. Wezel (ed). *Developments in Geotectonics*, 21. Elsevier, 281-304.
- Brun, J.P. & Cobbold, P.R. (1980). Strain heating and thermal softening in the continental shear zones: a review. *J. Str. Geol.*, 2, 149-158.
- Burg, J.P.; Delor, C.P.; Leyreloup, A.F. & Romney, F. (1989). Inverted metamorphic zonation and Variscan thrusts tectonics in the Rouergue area (Massif Central, France): P-T-t record from mineral to regional scale. In: *Evolution of Metamorphic Belts*. J.S. Daly et al. (eds). *Blackwell*, 423-439.
- Casas, A. & Carbó, A. (1990). Deep structure of the Betic Cordillera derived from the interpretation of a complete Bouguer anomaly map. *J. Geodyn.*, 12, 137-147.
- Doblas, M. & Oyarzun, R. (1989). "Mantle core complexes" and Neogene extensional detachment tectonics in the western Betic Cordilleras, Spain: an alternative model for the emplacement of the

- Ronda peridotite. *Earth Planet. Sci. Lett.*, 93, 76-84.
- England, P.C. & Thompson, A.B. (1984). Pressure-Temperature-Time Paths of Regional Metamorphism; I. Heat Transfer during the Evolution of Regions of Thickened Continental Crust. *J. Petrol.*, 25, 894-828.
- England, P. & Molnar, P. (1993). The interpretation of inverted metamorphic isograds using some simple physical calculations. *Tectonics*, 12, 145-158.
- Etheridge, M.A.; Wall, V.J. & Vernon, R.H. (1983). The role of the fluid phase during regional metamorphism and deformation. *J. met. Geol.*, 1, 205-226.
- Frey, F.A.; Suen, C.J. & Stockman, H.W. (1985). The Ronda high temperature peridotite. Geochemistry and petrogenesis. *Geoch. Cosmoch. Acta*, 49, 2469-2491.
- Gnos, E. & Kurz, D. (1994). Sapphirine-quartz and sapphirine-corundum assemblages in metamorphic rocks associated with the Semail Ophiolite (United Arab Emirates). *Contrib. Mineral. Petrol.*, 116, 398-410.
- Graham, C.M. & England, P.C. (1976). Thermal regimes and regional metamorphism in the vicinity of overthrust faults: An example of shear heating and inverted metamorphic zonation from southern California. *Earth Planet. Sci. Lett.*, 31, 142-152.
- Hernández Pacheco, A. (1967). Estudio petrográfico y geoquímico del macizo ultramáfico de Ojén. *Estudios geol., Spain*, 23, 85-144.
- Hess, H.H. (1955). Serpentinities, Orogeny and Epeirogeny. *Geol. Soc. Am. Spec. Paper*, 62, 391-408.
- Jain, A.K. & Manickavasagan, R.M. (1993). Inverted metamorphism in the intracontinental ductile shear zone during Himalayan collision tectonics. *Geology*, 21, 407-410.
- Jaupart, C. & Provost, A. (1985). Heat focussing granite genesis and inverted metamorphic gradients in continental collision zones. *Earth Planet. Sci. Lett.*, 73, 385-397.
- Kornprobst, J. (1969). Le Massif ultrabasique des Beni Bousera (Rif Interne, Maroc). *Contrib. Mineral. Petrol.*, 23, 282-322.
- Kornprobst, J. (1974). Contribution à l'étude pétrographique et structurale de la zone interne du Rif. *Notes. Mem. Serv. Géol. Maroc*, 251, 338 pp.
- Lockett, J.M. & Kusznir, N.J. (1981). Ductile shear zones: some aspects of constant slip velocity and constant shear stress models. *J. Geophys. Res. Astr. Soc.*, 70, 177-196.
- Loomis, T.P. (1972). Diapiric emplacement of the Ronda high-temperature ultramafic intrusion, southern Spain. *Geol. Soc. Am. Bull.*, 83, 2475-2496.
- Loomis, T.P. (1975). Tertiary mantle diapirism, orogeny and plate tectonics east of the strait of Gibraltar. *Am. J. Sci.*, 275, 1-30.
- Lundeen, M.T. (1978). Emplacement of the Ronda peridotite, Sierra Bermeja, Spain. *Geol. Soc. Am. Bull.*, 89, 172-180.
- Molnar, P. & England, P. (1990). Temperatures, Heat Flux, and Frictional Stress Near Major Thrust Faults. *J. Geophys. Res.*, 95, 4833-4856.
- Obata, M. (1980). The Ronda Peridotite: garnet-spinel-and plagioclase-lherzolite facies and the P-T trajectories of a high-temperature mantle intrusion. *J. Petrol.*, 21, 533-572.
- O'Hara, K. (1988). Fluid flow and volume loss during mylonitization: an origin for phyllonite in an overthrust setting, North Carolina, U.S.A. *Tectonophysics*, 156, 21-36.
- Patrick, B.; Till, A.B. & Dinklage, W.S. (1994). An inverted metamorphic field gradient in the central Brooks Range, Alaska and implications for exhumation of high-pressure/low-temperature metamorphic rocks. *Lithos*, 33, 67-83.
- Peacock, S.M. (1987). Creation and preservation of subduction-related inverted metamorphic gradients. *J. Geophys. Res.*, 92, 12763-12781.
- Platt, J.P. & Vissers, R.L.M. (1989). Extensional collapse of thickened continental lithosphere: A