

Inferred ionic composition and salinity of a Bolivian Quaternary lake, as estimated from fossil diatoms in the sediments

M. Roux¹, S. Servant-Vildary² & M. Servant³

¹CEPE-CNRS, BP 5051, 34033 Montpellier, France; ²Antenne O.R.S.T.O.M., Laboratoire de Geologie du Museum, 43 Rue Buffon, 75005 Paris, France; ³O.R.S.T.O.M., 70-74 Route d'Aulnay, 93140 Bondy, France

Key words: diatoms, past salinities, salt lake, reciprocal averaging, correspondence analysis, multiple linear regression

Abstract

Multiple Linear Regression (MLR) was used to estimate past salinity and ionic composition in sediments of a paleolake in the southernmost Bolivian Altiplano. Modern diatom assemblages were sampled from the subsurface sediments of 14 saline lagoons, and compared quantitatively with fossil assemblages. Only one application of the method is reported: to a Quaternary outcrop located on the border of one of them (Laguna Ballivian). Multivariate analysis was performed on diatom data to check the range of variation of both the fossil and modern flora. It also allowed a reduced set of explanatory variables for regression analysis.

Finally, a number of transformations of the chemical variables was computed prior to the regression process in order to stabilize the residuals variance and obtain a good error estimate. Only ten chemical components were studied, but the method could be applied to other palaeoenvironmental parameters.

Introduction

Diatoms in sediments have been used for many qualitative or semi-quantitative paleolimnological interpretations. Recently, diatom remains have been used to provide a quantified record of past chemical variation (essentially pH) (cf. Charles, 1985; Flower, Battarbee & Appleby, 1977; Brugam, 1980; Renberg & Hellberg, 1982; Van Dam, Suurmond & ter Braak, 1981).

In the present paper, we report their use to estimate past salinities. Our sample sites comprised a set of salt lakes on the Bolivian plateau, an area propitious for this sort of research since many lakes of varied ionic composition and salinity, but under similar geological and climatic

(temperature and rain) settings, occur there. The fossil samples came from an outcrop on the bank of one lake.

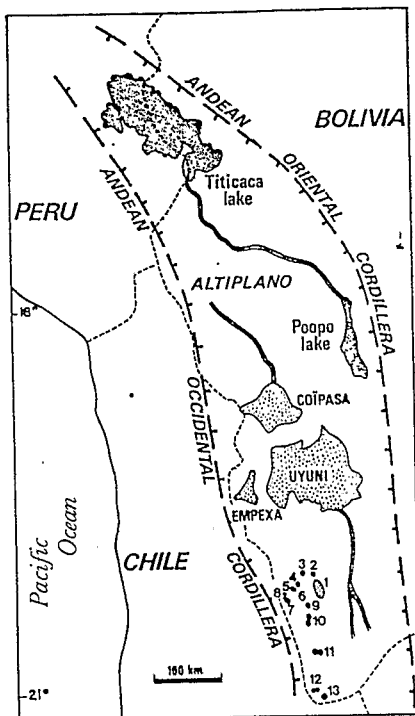
Our methodology was based on the pioneering work of Imbrie & Kipp (1971) whose main hypothesis was that for a given species or assemblage ecological responses to chemical change remain constant through time. Our purpose was to test several transformations of the dependent (chemical) variables to obtain a reliable estimate of the error on the reconstructed chemical composition. It appears that the best transformation improves on both the distribution of residuals and the multiple correlation coefficient (cf. Draper & Smith, 1981).

The diatom data and the chemical environment

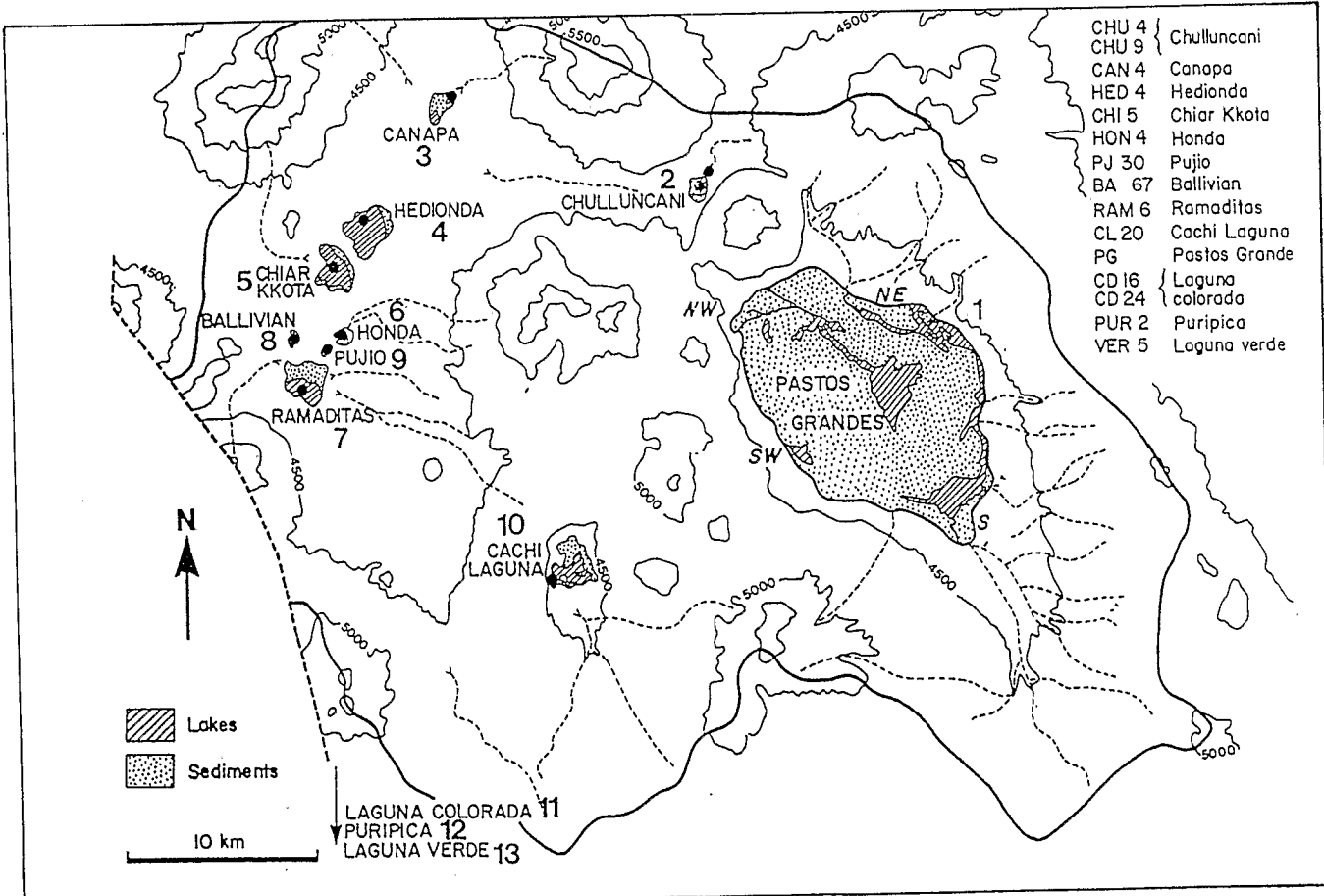
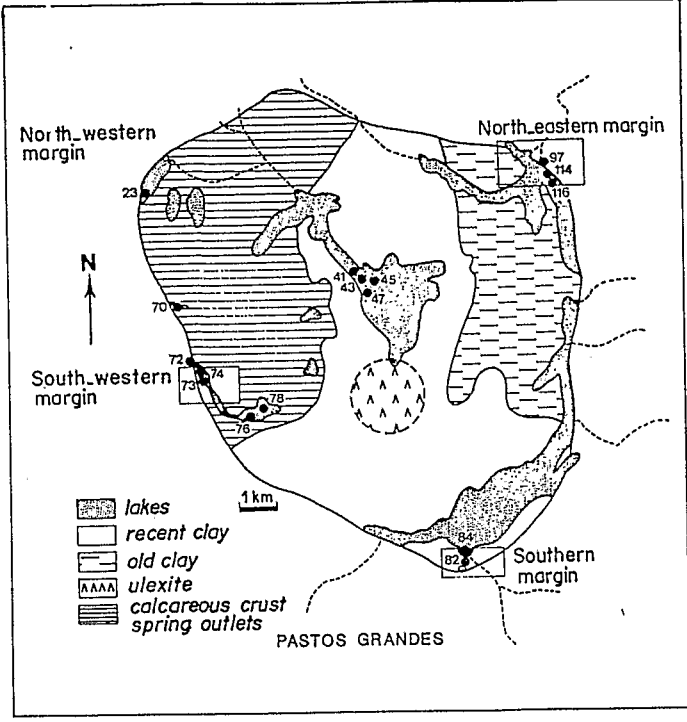


Fonds Documentaire ORSTOM

Cote: Bx21134 Ex: 1



- 1 Pastos grandes
- 2 Chulluncani
- 3 Canapa
- 4 Hedionda
- 5 Chiar Kkota
- 6 Honda
- 7 Ramaditas
- 8 Ballivian
- 9 Pujio
- 10 Cachi Laguna
- 11 Laguna Colorada
- 12 Puripica
- 13 Laguna verde



- CHU 4 } Chulluncani
- CHU 9 } Chulluncani
- CAN 4 } Canapa
- HED 4 } Hedionda
- CHI 5 } Chiar Kkota
- HON 4 } Honda
- PJ 30 } Pujio
- BA 67 } Ballivian
- RAM 6 } Ramaditas
- CL 20 } Cachi Laguna
- PG } Pastos Grande
- CD 16 } Laguna colorada
- CD 24 } Laguna colorada
- PUR 2 } Puripica
- VER 5 } Laguna verde

Fig. 1. Location of the fourteen lakes studied. Codes used for samples are: PG or P for Pastos Grandes, CHU for Chulluncani, CAN for Canapa, HED for Hedionda, CHI for Chiar Kkota, HON for Honda, RAM for Ramaditas, BA for Ballivian, PJ for Pujio, CL for Cachi Laguna, CD for Laguna Colorada, PUR for Puripica, VER for Laguna Verde, PJ or PN for Poopo.

Table 1. Physico-chemical features of study lakes. Data: temperature, °C; alkalinity, meq·l⁻¹; ions, mM l⁻¹.

Lake	Water sample no.	Temp.	Density	pH	Alk ¹	Cl	SO ₄	B	Si	Na	K	Li	Ca	Mg	TDS ²	Sediment no.
Ballivian	BAL 1	5	1.032	8.18	4.88	620	59.40	13.90	0.90	591	43.5	3.67	29.90	24.90	45335	BA67
Ramaditas	RAM5	1	1.020	8.15	2.93	392	32.00	7.12	1.48	330	26.3	1.70	34.20	13.40	27658	RAM6
Laguna Verde	VER 2	2	1.010	8.72	7.25	182	24.00	11.60	1.02	196	7.88	5.26	5.44	10.80	14716	VER5
Hedionda	HED 3	8	1.050	8.50	10.00	693	186.00	21.70	0.98	885	53.7	17.60	13.00	26.70	67099	HED4
Pujio	PUJ 5	1	1.022	8.85	7.22	409	45.00	13.40	0.93	435	26.1	5.33	9.98	8.64	31139	PJ30
Puripica	PUR 4A	4	1.024	8.52	7.80	430	48.50	22.00	0.73	415	44.0	15.70	11.60	11.30	32785	PUR2
Honda	HON3	6	1.015	9.05	4.40	290	27.10	5.27	1.13	293	25.3	6.77	4.99	5.76	31392	HON4
Chiar Kkota	CHI 4	8	1.051	8.28	8.05	1090	42.50	23.10	1.23	900	63.9	25.40	33.40	46.90	69439	CHI5
Canapa	CAN 3	6	1.009	9.18	2.15	63.4	52.80	1.20	1.12	156	5.42	2.81	1.62	1.40	11440	CAN4
Chulluncani	CHU 3	5	1.087	8.80	35.00	1240	277.00	88.70	0.78	1310	327	3.24	18.20	78.20	119261	CHU4
	CHU 2	8	1.008	10.20	11.40	55.5	45.40	13.60	0.67	101	46	0.40	4.99	2.18	11377	CHU9
Cachi Laguna	CAL 19	21	1.029	10.38	355.00	128	38.60	13.20	5.75	460	73	7.80	0.06	0.01	36270	CL20
Laguna Colorada	CLD 4	6	1.081	8.40	31.50	1830	67.10	56.60	1.63	1770	109	28.20	6.48	37.60	120357	CD16
	CLD 33	10	1.040	8.52	12.90	831	62.30	24.30	1.16	865	54	12.50	2.57	15.70	59166	CD24
Pastos Grandes	PAG 69	5	1.009	8.42	5.09	204	2.76	2.77	1.30	174	13.6	6.12	4.99	3.70	12787	PG70
	PAG 72	10	1.000	6.95	3.30	2.5	1.09	0.09	1.33	3.5	0.42	0.07	1.27	0.83	644	PG72
	PAG 73	1	1.009	8.15	5.13	190	3.15	2.36	2.10	170	10.2	6.12	4.74	5.68	12189	PG73
	PAG 74	1	1.001	7.85	4.20	16.9	1.35	0.27	1.08	15.2	1.23	0.56	0.82	2.28	1542	PG74
	PAG 75	10	1.010	8.35	3.85	234	3.18	2.96	1.15	196	14.8	8.29	4.99	5.43	14421	PG76
	PAG 77	10	1.098	7.91	9.42	2420	30.40	29.60	1.22	2000	128	86.50	37.40	49.40	144099	PG78
	PAG 81	10	1.000	9.62	0.52	1.8	0.15	0.07	0.57	1.8	0.1	0.05	0.15	0.13	195	PG82
	PAG 83	15	1.147	7.46	9.70	3770	35.10	37.40	1.18	3350	165	97.30	41.10	51.40	225344	PG84
	PAG 96	10	1.000	8.92	1.36	4.2	0.05	0.10	0.80	3.7	0.32	0.17	0.31	0.31	402	PG97
	PAG 124	10	1.167	6.95	7.68	4340	33.80	50.40	0.60	3520	251	167.00	59.40	105.00	255230	P114
	PAG 115	7	1.171	7.50	13.10	4450	33.1	48.10	1.12	4000	189	124.00	62.30	85.60	267366	P116
	PAG 22	1	1.001	9.35	1.51	19.7	0.75	0.32	1.37	19.6	1.10	0.69	0.61	0.45	1475	PG23
	PAG 40	4	1.020	8.52	4.25	470	4.84	5.55	1.13	403	26.10	16.90	8.98	10.90	28564	PG41
	PAG 30	6	1.010	8.05	3.21	227	2.60	2.63	0.62	196	12.80	7.57	4.99	5.43	13961	PG43
	PAG 44	5	1.073	7.40	9.08	1730	13.20	26.80	0.73	1480	101	72.10	27.40	4.53	103267	PG45
	PAG 48	5	1.211	7.20	22.90	5460	25.60	25.60	1.12	4480	363	236.00	77.30	143.00	324141	PG47

¹ Alkalinity; ² total dissolved salts.

are described first. The use of Correspondence Analysis [or Reciprocal Averaging, see Hill (1973) and Benzécri (1973)] as a preliminary step instead of Principal Components Analysis is then discussed for the processing of diatom data. The results of the various regressions on modern samples are compared. Finally, past salinities are estimated from the best available regression formula.

Materials and methods

Diatom remains from the sediment/water interface are preferable to those from just below the interface because neoformation of smectite from the silica of diatom frustules frequently occurs (Badaut *et al.*, 1979). This may cause many diatom frustules to disappear over a few centimetres, but is not important in the water/sediment interface (Iltis, Risacher & Servant-Vildary, 1984; Servant-Vildary, 1984). The diatom flora sampled from the water column in 1982 and from the water/sediment interface in 1978 showed a Pearson product-moment correlation coefficient [after transformation $\text{Log}(1+x)$] of 0.91 and 0.82 for Honda and Ramaditas lakes, respectively, 0.72 for Canapa, 0.52 for Chulluncani, and 0.87 for Cachi laguna. Good preservation of frustules is illustrated in Servant-Vildary & Roux (in press).

In the southern Bolivian Altiplano occur many small and shallow saline lakes ('salars'). They lie scattered in a volcanic area extending over 10 000 km² in the Occidental Cordillera of the Andes (Fig. 1). The twelve smallest (Fig. 1, 2 to 12) have individually consistent water chemistry (Ballivian & Risacher, 1981), but differ significantly when compared to each other. Their salinity varies from 11.3 to 120 gl⁻¹ (Table 1).

Three samples were obtained from each lake and were collected in May 1978, November 1978, February 1980. All had similar diatom floras, indicating that samples from the water/sediment interface provide an integrated record over at least two years.

Because of the consistency of the diatom flora, in most cases only one sample, corresponding to one chemical analysis at the sample site, was mathematically analysed. However, from the largest lake, Pastos Grandes, because of significant chemical variation in the water column (Table 1), 16 samples were analysed; in this lake salinity is a function of distance from the river and/or some spring outlets, and generally increases towards the lake centre. Another lake, Poopo, provided 10 samples (Servant-Vildary, 1978). In all, 40 samples were analysed.

Chemical methods used were: atomic absorption for Na, K, Li, Ca, Mg; colorimetry by mercury thiocyanate for Cl; indirect colorimetry by methylthymol blue upon excess of Ba after pre-

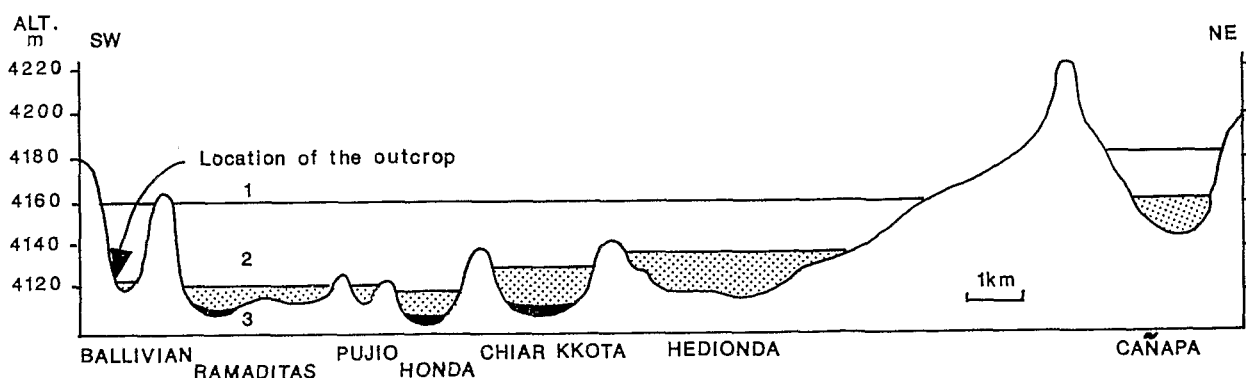


Fig. 2. Cross section of Lipez area where most study lakes are located. 1, water level during Minchin lacustrine episode; 2, water level during Tauca lacustrine episode; 3, present water level. Altitude as metres above sea-level.

Table 2. Diatoms in fossil samples. For explanation of sample numbers, see text. Data as per mille of total fossil flora.

Taxon	Sample number																											
	18	18A	18B	20	22	24	26	28	28A	33	33A	35	36	36A	41	43	44	44A	46	48	48A	50	52	53	55	57	59	61
<i>Achnanthes delicatula</i>	0	0	66	0	0	57	115	9	40	0	0	25	27	9	120	0	0	0	0	0	9	0	0	430	26	9	0	0
<i>A. exigua</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0
<i>A. hauckiana</i>	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	18	0	0
<i>A. lanceolata</i>	11	0	0	0	8	0	0	0	0	0	0	25	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0
<i>A. lemmermani</i>	0	0	351	0	8	96	31	0	0	7	6	0	0	0	0	0	0	0	0	0	9	0	0	0	13	0	0	0
<i>A. minutissima</i>	0	0	0	0	107	0	0	0	40	7	0	0	27	9	0	0	9	0	4	0	0	36	0	8	0	26	33	0
<i>A. speciosa</i>	0	0	335	0	0	57	23	9	40	13	0	0	0	9	160	0	0	0	4	91	0	48	0	0	13	0	0	0
<i>Amphora atacamae</i>	0	0	4	0	0	0	0	0	0	0	0	0	27	17	0	0	0	0	11	0	0	0	0	0	32	26	99	14
<i>A. atacamae minor</i>	6	0	4	0	0	0	8	0	0	0	0	0	0	9	0	0	0	0	4	0	0	0	0	0	135	70	60	0
<i>A. carvajaliana</i>	0	9	37	0	0	13	0	9	0	0	0	625	27	348	0	0	0	0	813	0	9	83	19	8	0	18	0	0
<i>A. coffeaeformis</i>	11	0	8	24	8	0	192	0	0	0	0	0	27	17	0	0	0	0	0	0	0	24	0	0	45	53	13	43
<i>A. frenguelli</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	190	0	0	32	0	0	0
<i>A. libyca</i>	11	0	0	24	8	6	0	9	0	7	6	0	0	0	0	0	9	0	0	0	0	0	0	8	0	0	0	0
<i>A. lineolata</i>	0	0	0	0	0	0	31	9	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	9	0	0
<i>A. platensis</i>	63	60	0	126	69	166	8	171	0	307	746	0	54	17	0	353	0	47	0	442	870	24	0	0	0	0	0	0
<i>A. sp.</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	184	232	72
<i>A. veneta</i>	0	190	0	55	23	64	62	77	0	20	99	0	0	122	0	10	0	9	0	0	19	36	0	0	0	0	0	0
<i>Anomoeoneis sp.</i>	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0
<i>A. sphaer. angusta</i>	6	9	0	8	0	25	15	17	0	7	0	0	0	0	0	39	274	0	0	0	9	0	9	0	0	0	0	0
<i>A. sphaer. platensis</i>	0	9	0	0	0	13	38	77	0	7	0	0	27	9	0	147	442	9	0	156	9	0	0	91	0	0	0	0
<i>Caloneis westii</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0
<i>Ceratoneis arcus</i>	0	9	0	0	0	6	0	9	0	0	0	25	27	26	0	10	9	0	0	0	0	0	0	8	0	0	0	0
<i>Chrysophyceae spp.</i>	0	0	0	0	0	0	0	9	400	0	0	0	0	9	0	0	0	0	0	0	0	0	168	0	0	0	0	0
<i>Cocconeis placentula</i>	611	517	12	307	115	121	192	137	240	67	50	125	595	61	320	59	18	849	4	13	0	179	308	248	13	35	7	14
<i>Cyclotella gamma</i>	34	0	0	0	0	0	0	0	0	413	6	25	0	9	0	10	18	9	0	0	9	0	0	0	0	0	0	0
<i>C. stelligera</i>	0	0	0	8	0	6	0	0	0	0	0	0	27	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0
<i>Cymbella cistula</i>	6	0	0	0	84	0	0	0	0	0	0	0	0	0	0	0	0	9	0	26	0	131	0	8	0	0	0	0
<i>C. lunata</i>	11	0	21	8	8	0	54	9	0	7	6	0	0	26	0	0	53	0	0	26	0	12	0	0	110	281	79	29
<i>C. microcephala</i>	0	0	0	0	0	6	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>C. minuta</i>	0	0	0	0	0	6	0	0	40	0	0	25	0	17	0	10	9	0	0	0	0	0	0	0	0	0	0	0
<i>C. pusilla</i>	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Denticula elegans</i>	91	0	0	283	374	13	0	0	40	7	0	0	0	9	0	10	0	0	0	78	0	36	0	0	6	0	0	0
<i>Diploneis smithii</i>	6	0	0	0	8	13	0	0	0	7	0	25	0	9	0	0	9	0	0	0	0	0	0	8	0	0	0	0
<i>Fragilaria construens</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0
<i>F. pinnata</i>	0	0	37	0	0	13	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	36	159	0	0	0	0	0
<i>Gomphonema angustatum</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	0
<i>G. intricatum</i>	0	9	0	0	15	13	0	9	0	0	0	0	0	9	0	0	0	9	0	0	0	0	0	0	0	0	0	0
<i>G. lanceolata</i>	0	0	0	0	0	0	0	0	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>G. parvulum</i>	0	0	0	0	0	0	0	34	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Mastoglia atacamae</i>	17	0	0	31	53	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0
<i>Melosira octogona</i>	0	0	8	0	0	0	0	0	0	0	0	0	0	0	360	10	0	0	0	0	0	0	9	33	6	0	0	0
<i>Navicula cari cincta</i>	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	6	0	0	0	29
<i>N. cincta</i>	6	9	4	8	0	6	0	51	0	7	0	0	0	9	0	88	115	0	4	0	9	12	37	8	0	0	0	0
<i>N. cryptocephala</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	47	0	0	9	0	0
<i>N. guatemalensis</i>	11	60	0	8	0	0	0	188	0	20	0	0	0	26	0	29	0	9	0	0	0	0	0	0	0	0	0	0

precipitation of BaSO₄ for SO₄; colorimetry by ammonium molybdate for Si.

The paleogeography of deposits in the southern Bolivian Altiplano was studied by Servant & Fontes (1978) who indicated the most lagoons were connected during the 'Minchin' lacustrine episode of 30 000 to 22 000 BP (Fig. 2, level 1). Only Canapa, the highest lagoon, was not connected. After a long dry period between ~20 000 and 13 000 years BP, a new lacustrine phase, the 'Tauca' wet episode, extended to 10 000 years BP. Its water level was lower than that of the former Minchin large lake, and the lagoons were isolated from each other as now (Fig. 2, levels 2 and 3).

To assess quantitatively lake salinity ranges

and chemical composition during the Minchin and Tauca lacustrine periods, an outcrop with well-preserved fossil diatom frustules located on the border of Laguna Ballivian (Fig. 2) was studied in detail. The outcrop includes 11 quite pure diatomite layers intercalated with dominantly sandy-clay and sandy layers lacking or with few diatoms. Twenty-eight samples were obtained from this outcrop at depths between 18 and 61 cm from the surface (numbered 18 to 61 according to depth in Tables 2, 3 and 6; A and B samples are lateral equivalents). Thick sandy layers at levels 36 and 41 contained eolian gypsum, indicating some reworking of sediments by wind. A ¹⁴C date, from carbonated plants remains at

Table 3. Diatoms in the sediments, and their fossil occurrence. Numbers indicate fossil sample depth and, in parentheses, the percentage of the taxon in the total fossil flora.

Taxon	Occurrence in fossil samples
<i>Achnanthes exigua</i>	43(1)
<i>A. lanceolata</i>	18(1.1), 22(0.8), 35(2.5), 46(0.4)
<i>A. minutissima</i>	22(10.7), 28A(4), 33(0.7), 36(2.7), 36A(0.9), 44(0.9), 46(0.4), 50(3.), 53(0.8), 57(2.6), 59(3.3)
<i>Amphora frenguelli</i>	50(19), 55(3.2)
<i>Amphora</i> sp.	57(18.4), 59(23.2), 61(7)
<i>Anomoëis</i> sp.	28(0.9), 44A(0.9)
<i>Cyclotella stelligera</i>	20(0.8), 24(0.6), 36(2.7), 55(0.6)
<i>Cymbella cistula</i>	18(0.6), 22(8.4), 44A(0.9), 48(2.6), 50(13.1), 53(0.8)
<i>C. minuta</i>	24(0.6), 28A(4), 35(2.5), 36A(1.7), 41(1), 44(0.9)
<i>C. pusilla</i>	18B(0.4)
<i>Diploneis smithii</i>	18(0.6), 22(0.8), 24(1.3), 33(0.7), 35(2.5), 36A(0.9), 44(0.9), 53(0.8)
<i>Gomphonema angustatum</i>	50(4.8)
<i>G. intricatum</i>	18A(0.9), 22(1.5), 24(1.3), 33(0.7), 35(2.5), 36A(0.9), 44(0.9), 53(0.8)
<i>G. lanceolatum</i>	28A(4)
<i>Melosira octogona</i>	18B(0.8), 41(36), 43(1), 52(0.9), 53(3.3), 55(0.6)
<i>Navicula guatemalensis</i>	18(1.1), 18A(6), 20(0.8), 28(18.8), 28A(2), 36A(2.6), 43(2.9), 44A(0.9)
<i>N. halophila</i>	18(0.6), 18A(0.9), 20(2.4), 22(9.2), 24(0.6), 28(0.9), 28A(4), 33(1.3), 33A(1.1), 44A(0.9), 50(1.2)
<i>N. neoventricosa</i>	46(0.4), 61(1.4)
<i>N. peregrina</i>	18B(0.8), 48(0.9)
<i>N. placentula</i>	18A(0.9), 24(0.6), 28(0.9)
<i>Nitschia commutata</i>	46(0.4)
<i>N. gracilis</i>	28(5.1), 50(1.2)
<i>N. ingens</i>	18(0.6), 33(0.7), 55(2.6), 57(1.8), 59(13.9), 61(2.9)
<i>N. intermedia</i>	18A(0.9), 28(2.6), 33(0.7), 36(2.7), 36A(1.7), 46(0.4), 48A(0.9), 50(1.2)
<i>N. sigma</i>	33(0.7)
<i>N. sigmoidea</i>	28(0.9)
<i>N. valdestriata</i>	33(0.7), 55(0.6), 57(0.9)
<i>Rhopalodia musculus</i>	24(0.6), 55(1.3), 59(4.3), 61(1.4)
<i>Surirella ovalis</i>	33A(0.6)
<i>Synedra acus</i>	48(1.3)
<i>S. ulna</i>	48A(0.9)

level 61, gave an age of 22000 years BP, which correlates with the Minchin lacustrine period. Accordingly, the uppermost part of the outcrop is assumed to correlate with the Tauca lacustrine period.

Taxonomic details of fossil diatoms (Table 2) are given by Mello E Sousa (1989). Here, we note that only 95 species were present, of which 31 were represented in lake sediments (Table 3). Six were present in only one fossil sample and, since they were below 1% of the total flora in these, were removed in subsequent analysis.

As previous studies (Servant-Vildary & Roux, *in press*) have indicated that diatoms seem more sensitive to ionic proportions than to their absolute ion quantity, the only chemical variables taken into account were the ratios of each ion to total dissolved salts.

Analysis of data

Correspondence analysis

A preliminary analysis used Correspondence Analysis of floristic data. The goals were to overview the modern flora and its relation to the fossil one, and to provide the regression process with some uncorrelated new variables. Correspondence Analyses (rather than Principal Component analysis) was used because it seems more suited for counts of floristic data (Gauch, Whittaker & Wentworth, 1977; Greenacre, 1984). An extra feature of this method, developed in the French school of data analysis was incorporated: supplementary or passive elements (Benzecri, 1973; Greenacre, 1984). If factor coordinates are regarded as a new reference system, it is possible to project on these new axes not only the rows/columns of the main table but also some additional rows and/or columns not taken into account in the computation. These additional rows/columns are referred to as supplementary elements and are mostly used to facilitate interpretation of results linking this external information to the main active elements.

Our data comprised a main matrix consisting of

the percentages of the 113 species (rows) distributed in 40 subliving samples from the sediments (columns). As supplementary elements, we added on the right side of the main table some 28 columns representing the whole set of fossil samples. As passive elements they do not alter the factorial axis, but they can be related to their subliving counterparts.

The detailed interpretation of the results are given by Mello e Souza *et al.* (1989), though on a restricted set of modern samples. Here, we give only the most important results as they relate to the present regression process. The main result of the analysis (Fig. 3) is that there is good correspondence between the subliving samples and the fossil ones. That is, at least for the six first factorial axis (which account for 46% of the variance), all the coordinates of the fossil samples are included in the range of those of the modern ones. Some fossil samples (e.g. 18B, 20, 22) have no true modern analogue, but their intermediate position on the main factorial axis makes it possible to obtain valuable interpolations by the multiple linear regression method (MLR).

Regression analysis

General considerations. In dealing with MLR for paleoclimatic reconstructions, there are at least three objectives: to derive a satisfactory formula relating environmental parameters to biological ones (in the present case, diatom species); to estimate the mean error value; and to detect possibly abnormal observations (so-called 'outliers' in statistical terms) which may indicate some important climatic accident. However, the application of MLR in the usual way is not easy because (1) estimation of the standard error requires a set of residuals approximately gaussian, which in turn calls for some transformation of the variables, and (2) detection of outliers supposes that the regression formula is as accurate for these special observations as for others (in mathematical words, it is necessary to have an homogeneous variance for all the residuals).

To minimise these difficulties, four different

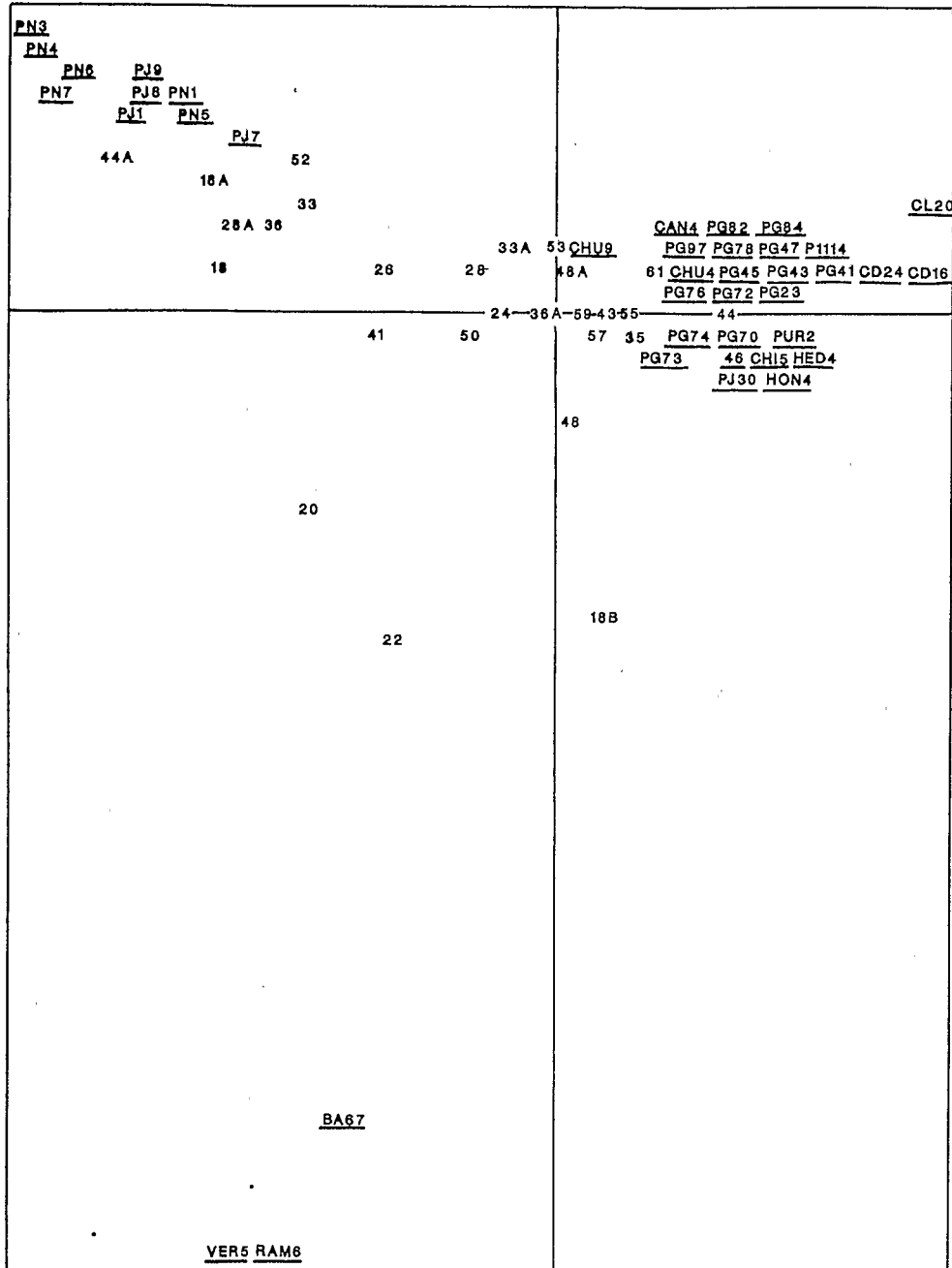


Fig. 3. Correspondence analysis of floristic data. The first (horizontal) factorial axis accounts for 9.36% of variance, while the second one (vertical) explains 9.15% of variance. Modern samples are underlined (see codes in legend of Fig. 1) as they are active elements. Fossil samples (see Table 2) are passive ones.

ways of data preparation were investigated: no transformation; logarithm of the independent variable; rank transformation of the independent variable; and derivation of a new set of explana-

tory variables by reciprocal averaging of a special data table. Some transformations of the explanatory variables were also investigated but yielded no improvement.

In our analysis, each chemical parameter was considered in turn as the independent variable, and each of the above methods was applied to every chemical variable. Thus, 40 multiple linear regressions (10 variables \times 4 methods) were compared. To facilitate comparisons, the usual correlation coefficient between the observed values and the estimated ones were computed. When there was no transformation of variables, it coincided with the multiple correlation coefficient. In the case of variable transformation, the reverse transformation on the output of the regression was undertaken to get the estimated results measured in the same units as the original variable; the correlation coefficient was then computed.

In the first three methods, the same 6 factor coordinates were used as explanatory variables. This number was selected after looking at the 46 per cent of total explained variance, and noting the rapid decrease of variances explained by the subsequent axes. In the last method, a new data table was erected for each parameter needing

Table 4. Comparative results of Regression analysis. Upper half, correlations between estimated and observed values; lower half, standard error (quadratic mean of the residuals).

Variable	No-transf. of var.	Log. transform.	Rank transform.	By class mean percent.
Alk	0.6623	0.5572	0.2856	0.7450
B	0.6329	0.5847	0.6299	0.6987
Ca	0.6168	0.6164	0.5426	0.7696
Cl	0.7475	0.7523	0.7253	0.8682
K	0.5605	0.5089	0.4982	0.6679
Li	0.3941	0.3626	0.3873	0.4184
Mg	0.7108	0.7337	0.6856	0.8168
Na	0.5768	0.5666	0.4779	0.8067
Si	0.5993	0.4422	0.4958	0.7359
So	0.7496	0.7300	0.7307	0.7741
<hr/>				
Alk	57.127	67.358	83.968	50.860
B	2.039	2.157	2.048	1.884
Ca	10.813	11.146	11.788	8.772
Cl	84.541	84.891	89.044	63.140
K	22.357	23.483	23.954	20.091
Li	4.567	4.676	4.652	4.514
Mg	5.729	5.655	6.026	4.699
Na	59.556	60.368	68.307	43.087
Si	29.053	35.396	33.799	24.573
So	61.815	64.917	64.249	59.129

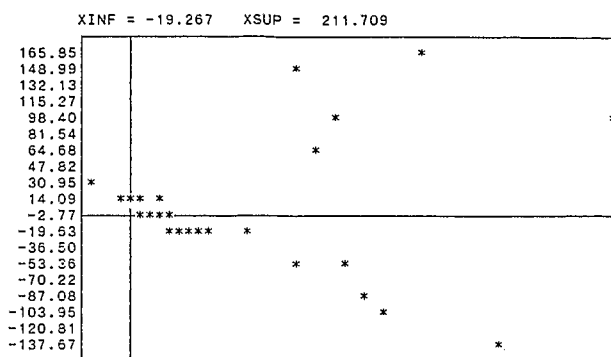


Fig. 4. Plot of the residuals of MLR with non-transformed variable Alk. Horizontal axis: estimated values. Vertical axis: residuals.

explanation. This new table was intended to describe exclusively the relationship between the studied parameter and the diatom flora. A few coordinates axis derived from Correspondence Analysis comprised the new explanatory variables.

Non-transformation of variables. MLR on non-transformed variables provided very low values for the multiple correlation coefficient, denoting poor adequacy of the linear formula (Table 4, col. 1). In addition, many variables had a poor distribution of residuals. For instance, the residual plot for the variable 'Alk' (Fig. 4) clearly shows several outliers and a non-constant variance. Since almost all variables involve one defect or the other, standard MLR appears inapplicable.

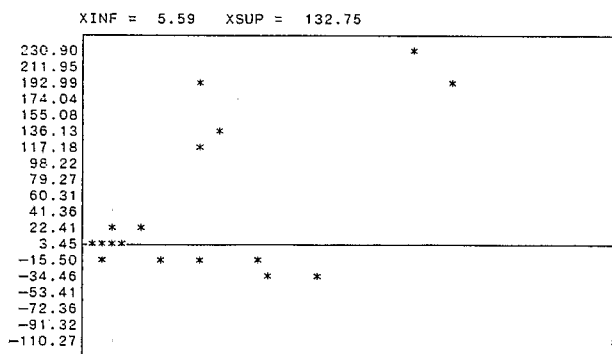


Fig. 5. Variable Alk. Plot of residuals after logarithmic transformation. Horizontal axis: estimated values after reverse transformation. Vertical axis: residuals after reverse transformation.

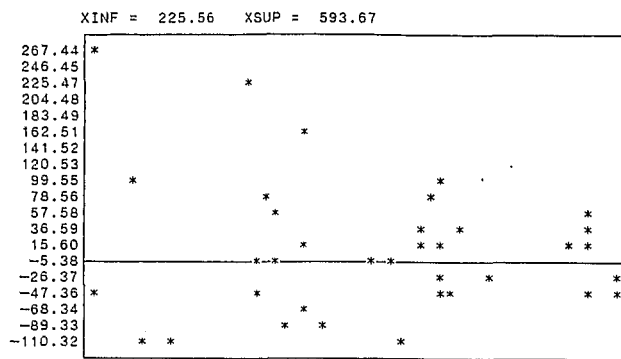


Fig. 6. Variable Cl. Plot of residuals after logarithmic transformation. Horizontal axis: estimated values after reverse transformation. Vertical axis: residuals after reverse transformation.

Logarithmic transformation of chemical variables. Each chemical ratio (Y) was transformed to $\text{Log}(Y + 1)$. The results, however, in terms of correlation coefficients, were not better when the reverse transformation (exponential) was applied to the estimated values (Table 4, col. 2). The residuals plots, though better than the previous ones, still showed outliers. For instance, in the 'Alk' plot (Fig. 5), the residuals appear better balanced, but an outlier remains on the high value side (right) of the graph. The only acceptable plot was the one for Cl (Fig. 6), although, of course, outliers are easily identified since they have very abnormal values. In any event, MLR was derived on subsamples obtained by removing outliers; the residuals plots were then better, but only two variables showed a higher multiple correlation coefficient, namely Li (0.8752) and Mg (0.8402).

Even so, this approach was also abandoned because it was considered inappropriate when past climatic indices were to be estimated. Outlier samples, from a chemical standpoint, are merely the result of some extreme environmental conditions which may well have appeared in the past. However, our preliminary studies on diatoms demonstrated that on floristic grounds both fossil and modern samples showed approximately the same dispersion. Thus, any modern sample could be an analogue for a fossil one, so that all observations of subliving diatoms should be considered.

Rank transformation. According to Iman & Conover (1979), the use of rank transformation is not suited for estimating a mathematical relationship except 'if one is interested in a model that predicts well over a monotonic non-linear surface then the rank regression will work quite well'. However, for our data, this transformation did not outperform the scores of the others (Table 4, col. 3). Only the variables K and Si had a slightly higher correlation coefficient. On the other hand, to obtain the estimated values is difficult since it is necessary to interpolate in order to recover the values out of non-integer 'rank' estimations. In our view, these poor results were probably due to non-monotone relationships. For this reason a new approach was attempted.

Factorizing relationships between floristic and chemical data. Floristic dispersion can be due to many environmental factors (e.g. wind, light) apart from water chemistry. Thus, the link between chemical and floristic data may not be apparent, at least in the first factorial axis. To shed light on this link, a 'by-class mean percentage table' (Gasse & Tekaiia, 1983; Servant-Vildary & Roux, 1985) was erected as follows. Each row represented a species, while columns were categories, or classes, of chemical ratios. Thus, the cell located at row *i* and column *j* contained the mean percentage of species *i* for those samples which fell into category *j*. The first step was then to divide the variables into a multi-state attribute. As the number of modern samples was 40, we found it appropriate to split the range of each variable into 4 intervals, the cut-off points of which were chosen so as to have approximately the same number of observations in each category. An example will aid explanation. The ratio Alk/TDS

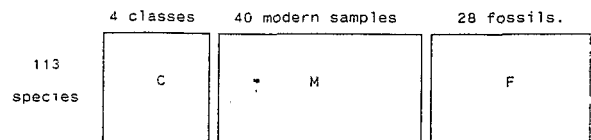


Fig. 7. The union of tables C, M and F is submitted to Reciprocal Averaging, but only table C (by-class mean percentages) is considered as active.

had the following cut off points: 0.006, 0.008, 0.017, so that class 1 (0-0.006) contained 9 samples, class 2 (0.007-0.008) 10, class 3 (0.009-0.017) 11 and class 4 (>0.017) 10 samples. When this was done, from the transformed chemical data table, the partial mean percentages over each class of ratio and for each species were computed. Finally, the finished table (Table C, Fig. 7) was similar to the floristic one, where samples were replaced by clusters.

Correspondence analysis of such a table was highly informative, since it allowed for graphical display of the categories of variable surrounded by the related species. But graphical information was not sufficient for the accurate estimations sought. It is for this reason that we placed, alongside the 'by-class mean percentage' table (Table C,

Fig. 7), the original table M for modern diatom samples and Table F for fossil records (Fig. 7). Columns in these tables were marked as supplementary elements for the computer program so as not to interfere with the processing of the main table (C). But they did have coordinates on the same factorial axis as the categories of variables. Thus, this computation provided two useful outputs: the display of all samples on the same graph as the categories of variables, giving a first appraisal for the fossil samples of their environmental conditions (Fig. 8), and the factorial coordinates of the samples which could then be considered as the new explanatory variables. They made up the input for the regression process. In addition, on the homologous graph displaying the rows of the table, it was possible to project the

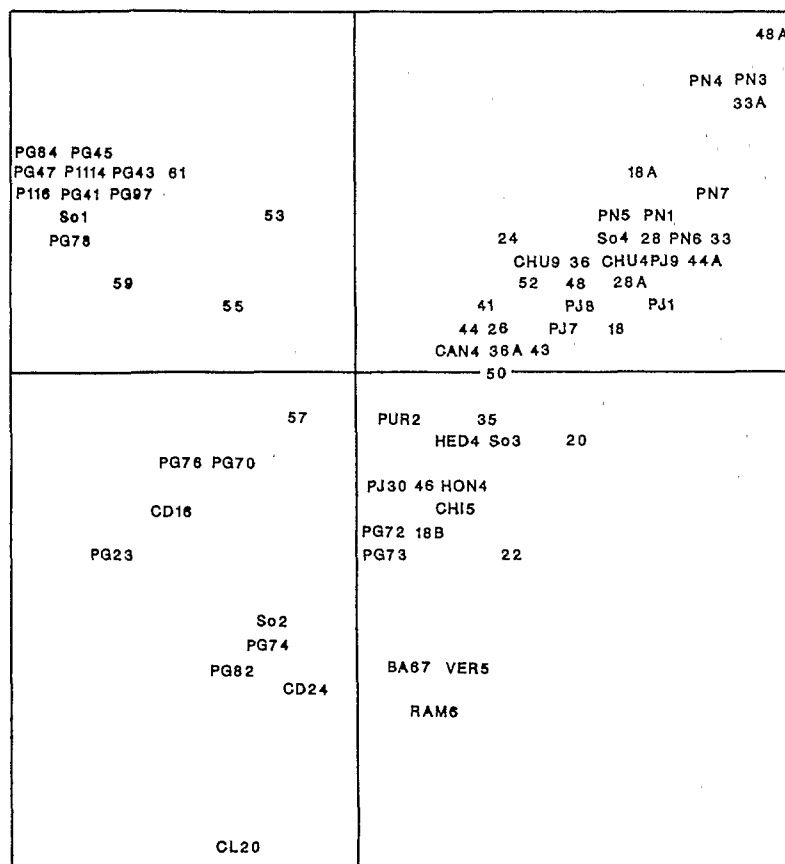


Fig. 8. Correspondence analysis of the 'by-class mean percentage' table for SO_4^{-2} . Representation of first (horizontal) and second (vertical) factorial axis (respectively 48% and 31% of variance). Modern samples as well as fossil ones are passive elements in this analysis.

Table 5. Results of the best regressions.

Variable	Correl.	R ²	Std err.	No. of explan. VAR.
Alk	0.8386	0.70	41.534	3/5 factor comb.
Ca	0.8032	0.65	8.183	3/5 factor comb.
Cl	0.8632	0.75	63.140	3 factors
K	0.7680	0.59	17.291	3 factor comb.
Mg	0.8168	0.67	4.699	3 factors
Na	0.8067	0.65	43.087	3 factors
Si	0.7817	0.61	25.909	5 fact. comb. + Log Y
So	0.7838	0.61	57.997	3/5 factor comb.

categories of variables, to that each appeared surrounded by those species mostly linked to that category.

Since there are four classes of chemical variable, only three factorial axes could be obtained (Greenacre, 1984). Overall this process

was rather complicated because it needed the construction and analysis of as many C-type tables as the number of chemical variables. However, it was worthwhile since all correlation coefficients were improved (Table 4, col. 4), some being much better than those given by the previous transformations (see variables K and Na, for instance). Two variables, B and Li, could not be reliably estimated whatever the transformation (correlations of 0.6987 and 0.4184, respectively); their values were always so low that it was likely that errors in measurement were greater than the values themselves. Study of these variables was discontinued.

Of remaining variables, Alk, Ca, K, Si and SO₄ still showed weak correlation coefficients. For instance, estimated Alk (correlation of 0.745)

Table 6. Best estimations for the fossil samples. See text for explanation.

Fossil sample no.	Variables								
	Alk	Ca	Cl	K	Mg	Na	Si	So	TDS
18	4.1	13.2	336.3	-7.1	18.6	325.2	7.8	185.8	38.7
18A	17.3	14.1	361.6	15.8	19.7	313.8	6.6	204.9	30.0
18B	-4.7	21.4	510.2	30.5	8.8	407.1	1.6	148.3	13.4
20	-3.9	15.4	363.3	-8.8	15.9	274.5	3.0	172.8	45.4
22	-3.2	30.1	413.9	10.7	14.9	278.1	1.6	156.5	42.3
24	6.5	15.1	382.1	37.9	11.1	299.6	2.6	202.5	12.2
26	4.0	16.1	396.9	14.3	16.3	338.1	5.0	186.9	16.0
28	37.3	14.5	292.0	26.0	13.8	235.4	4.1	201.3	42.4
28A	8.1	16.3	365.0	-6.7	21.5	346.3	13.1	191.8	15.4
33	-19.6	22.0	196.0	9.9	17.1	271.4	5.4	188.5	52.2
33A	248.8	23.0	111.0	23.0	5.6	101.9	1.2	191.4	64.0
35	0.8	13.3	402.4	60.4	10.3	334.4	1.7	172.5	64.8
36	8.3	14.7	352.7	11.8	17.2	335.8	7.4	198.3	19.5
36A	0.5	15.1	441.3	47.4	13.6	300.6	2.1	181.3	24.9
41	3.2	15.5	445.1	-27.3	16.7	321.8	3.9	189.8	32.6
43	178.2	14.7	267.8	27.1	10.8	186.0	4.2	185.9	40.5
44	23.8	22.3	400.5	50.8	14.1	314.1	2.6	186.7	26.5
44A	17.7	13.5	312.8	1.0	20.5	335.4	10.5	183.9	46.4
46	-3.1	10.7	427.4	44.6	7.4	318.8	1.8	158.2	121.2
48	125.9	15.0	252.7	25.3	6.1	194.6	2.0	197.1	37.4
48A	335.7	25.5	43.2	25.2	2.5	76.1	1.3	191.3	55.0
50	1.2	15.7	388.0	19.9	14.8	325.7	4.1	182.8	13.1
52	-0.4	15.2	369.9	34.6	19.8	336.1	16.9	197.6	12.1
53	3.6	14.4	511.8	9.1	11.7	290.4	3.3	147.8	14.2
55	3.5	13.3	531.1	27.0	10.6	340.4	2.6	116.0	53.6
57	7.4	16.6	516.5	27.4	11.9	319.1	3.1	128.5	18.9
59	1.6	12.0	586.3	25.2	9.2	338.6	1.9	55.3	92.8
61	-2.1	9.3	482.4	44.7	7.3	459.4	0.1	73.9	148.4

accounted for only 55.5% of the variation of observed Alk. In addition, the plot of residuals often showed a curvilinear dependence of the residuals over estimated values. For these reasons, it was decided to add to the factorial coordinates F1 and F2 their quadratic combinations, $F1^2$, $F1 \cdot F2$ and $F2^2$, as new explanatory variables (Polynomial regression). To avoid the introduction of irrelevant noisy variables, the more sophisticated stepwise linear regression was used, wherein the explanatory variables were included one at a time (cf. Zar, 1984).

Finally (Table 5), three factor combinations were used for Alk, Ca and SO_4 , while 5 were necessary to raise the correlation coefficient of variables K and Si to an acceptable value, e.g. >0.7 . Additionally, Si values required a logarithmic transformation for a best fit of the residuals.

Some of the residuals were still unsatisfactory (e.g. Alk, Fig. 9), but fortunately the best diagrams were provided by the major variables, namely Cl, Na and SO_4 (see Fig. 10 for Cl); these are the variables which account for the biggest fraction of total dissolved salts.

Discussion

To provide information on past salinities, the formulae providing best estimates were applied to fossil samples. These are shown in Table 6. It must be pointed out that in this table each variable was estimated separately from each other (this is why the sum of the estimated values is not equal to 1000-percentages are magnified by a factor of 10). Another feature of the table is that it contains some negative values; these reflect the fact that the regression formulae may have negative coefficients which sometimes give rise to negative estimations. Clearly such values correspond to very low values and should be replaced by zero.

In table 6 two points are noteworthy. First, the overall fluctuations are rather smooth except for some particular levels which show very high values of Alk associated with low values of Cl and Na. The errors on Alk, K and TDS are high, but

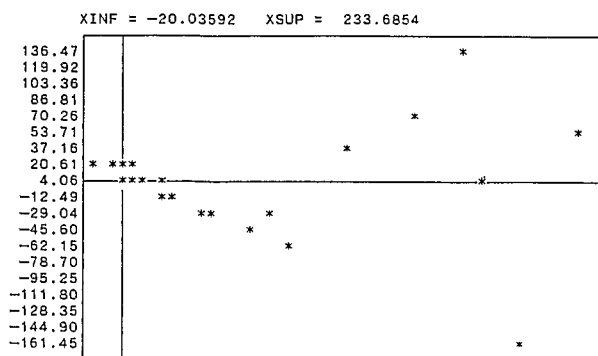


Fig. 9. Variable Alk: plot of residuals of the regression (ordinates) against estimated values (abscissae), using the by-class mean percentage table.

the repeated coincidence of this phenomenon cannot be attributed to chance. These levels are 33A, 43, 48A and possibly 48. Second, the variable SO_4 has a narrow range of variation excepting for two low values at levels 59 and 61.

Through careful use of Regression analysis, therefore, it has been possible to obtain reasonable estimations of salinity ranges and ionic concentrations of paleolake waters. Although the sample grid is not entirely satisfactory, a hydrochemical evolutionary scheme for Minchin and a part of Tauca can now be drawn from our computations (Fig. 11). The available data indicate that Minchin has had three distinct hydrochemical episodes (Servant-Vildary, Roux, Servant & Risacher, 1989):

In the first episode (samples 24 to 41), sedimentation was characterized by quite regular

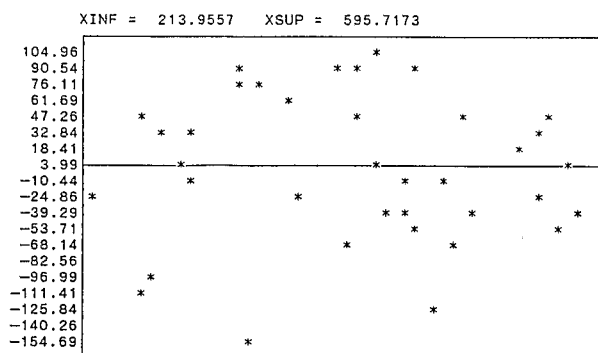


Fig. 10. Variable Cl: plot of residuals of the regression (ordinates) against estimated values (abscissae), using the by-class mean percentage table.

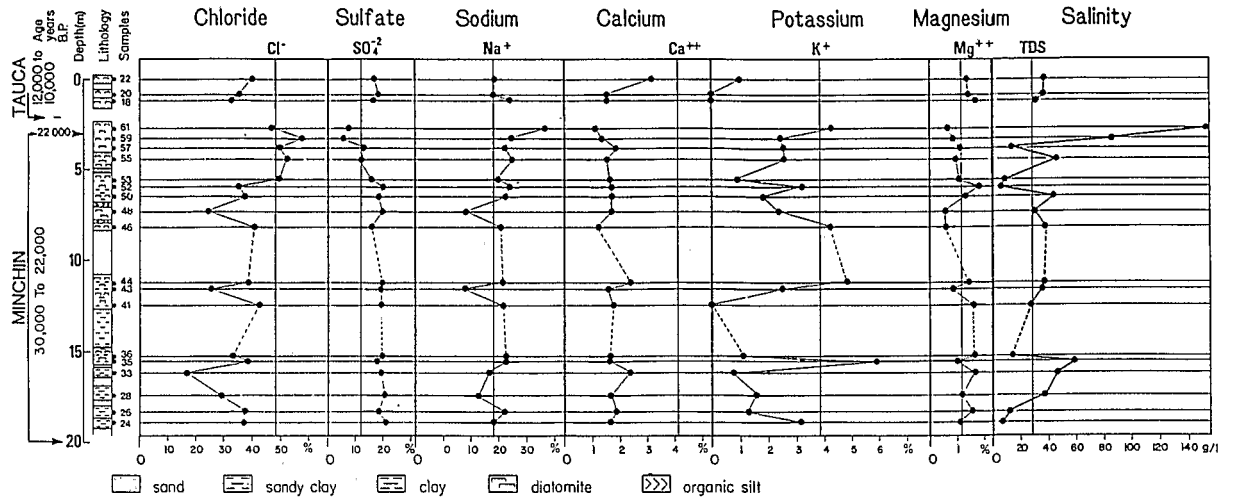


Fig. 11. Variations along the core of ionic contents (in% except salinity in g l^{-1}), with standard error (Table 4). Vertical lines Cl^- , SO_4^{2-} , Na^+ , K^+ , Mg^{2+} , TDS, correspond to the inferred values of these elements for modern sample of Ballivian (BA67).

intervals of sand and diatomaceous silt. The number of diatom frustules is relatively poor, 0.5 to 2×10^6 frustules g^{-1} , except for sample 33 which contains 21×10^6 frustules g^{-1} . This episode can be split into two hydrochemical sequences:

- 1A: samples 24, 26, 28 were HCO_3^- and Na dominate.
 sample 33 with Na, SO_4 and Ca
 sample 35 with Na^+ , SO_4 , Cl and K
- 1B: sample 36 represents a carbonated phase which evolved to a sodium chloride-sulfate phase in sample 41. The presence of reworked gypsum, in association with thick sandy material, between 36 and 41, indicates that lake water level was significantly lowered, giving way to evaporite formation.

The second episode (samples 43 to 46) began with a sodium-bicarbonate level (43) which evolved to a sodium-calcium sulfate (44) then a sodium chloride-sulfate level (46). The eolian origin of the deep sand level between samples 44 and 46 seemed not to have modified the chemical evolution of the waters.

The third episode (samples 48 to 61) was characterized by an increase of salinity and Na-Cl

concentration. The abundance of the diatom frustules in the samples probably indicates enhanced carbonate precipitation due to increased biological activity. Overall, the available information suggests a significant increase in salinity and ionic concentration in a water profile up to 20 metres deep under climatic conditions characterized by high evaporation. Thus, sample 48 is essentially sodium carbonate and sulfate, samples 50 and 52 are sodium and sulfate-chloride with high content in K and Mg, and samples 53 to 61 are sodium-chloride.

In conclusion, it is noted that the object of the present study was to obtain both a good estimate of past salinities and an appraisal of the error on such estimation. Emphasising this last problem, we were in search of data transformations providing an approximate normal distribution of residuals. Two classical transformations of the chemical parameters were considered, logarithm and ranks, together with a new approach to this problem involving the factorization of a special data table. This table, a 'by-class mean percentage' table, summarized the distribution of the diatom flora among the categories of each chemical variable.

We could not afford to eliminate any outlying observations since similar fossil sediments were

likely to occur, revealing a possible interesting event. Using the approach of the 'by-class mean percentages', we did get satisfactory estimations of salinities, but the distributions of residuals were not as homogeneous as they should be for several parameters. Fortunately, the major salts, namely Cl, Na and SO₄, showed a reliable scatter diagram of the residuals. Although the 95% confidence interval is still large, it allows for the detection of some important changes in the ionic water contents that can be related to climatic events.

Acknowledgements

Thanks are due to Dr. A. Arakel and especially Professor W. D. Williams for reviewing the English manuscript, and to Dr. A. Foucault, J. C. Fontes, M. Servant and J. M. Rouchy for helpful discussions. This work was supported by ORSTOM (GEOCIT Program directed by M. Servant) within UMSA La Paz/ORSTOM Paris agreement.

References

- Badaut, D., F. Risacher, H. Paquet, J. P. Eberhardt & F. Weber, 1979. Néof ormation de minéraux argileux à partir de frustules de diatomées: le cas des lacs de l'altiplano bolivien. *C. r. Acad. Sci., Paris*, 289D: 1191-1193.
- Ballivian, O. & F. Risacher, 1981. Los salares del Altiplano boliviano. ORSTOM, Paris. 246 pp. multigr.
- Benzécri, J. P., 1973. *L'Analyse des Données, Tome 2: L'Analyse des Correspondances*. Dunod, Paris.
- Brugam, R. B., 1980. Postglacial diatom stratigraphy of Kirchner marsh, Minnesota. *Quat. Res.* 13: 133-146.
- Charles, D. F., 1985. Relationships between surface sediment diatom assemblages and lakewater characteristics in Adirondack lakes. *Ecology* 66(3): 994-1011.
- Draper, N. R. & H. Smith, 1981. *Applied Regression Analysis*. 2nd ed. John Wiley, New York.
- Flower, R. J., R. W. Battarbee & P. G. Appleby, 1977. The recent palaeolimnology of acid lakes in Galloway, South-West Scotland: diatom analysis, pH trends, and the rôle of afforestation. *J. Ecol.* 75: 797-824.
- Gasse, F. & F. Tekaia, 1983. Transfer functions for estimating paleoecological conditions (pH) from East African diatoms. *Hydrobiologia* 103: 85-90.
- Gauch, H. G., Jr, R. H. Whittaker & T. R. Wentworth, 1977. A comparative study of reciprocal averaging and other ordination techniques. *J. Ecol.* 65: 157-174.
- Greenacre, M. J., 1984. *Theory and Applications of Correspondence Analysis*. Academic Press, London & New York.
- Hill, M. O., 1973. Reciprocal averaging: an eigenvector method of ordination. *J. Ecol.* 61: 340-354.
- Iltis, A., F. Risacher & S. Servant-Vildary, 1984. Contribution à l'étude hydrobiologique des lacs salés du Sud de l'altiplano bolivien. *Revue Hydrobiol. trop.* 19(3): 259-273.
- Iman, R. L. & W. J. Conover, 1979. The use of rank transform in regression. *Technometrics* 21 (4): 499-509.
- Imbrie, J. & N. G. Kipp, 1971. A new micropaleontological method for quantitative paleoclimatology: application to a late pleistocene Caribbean core. In: K. Turekian (ed.) *The Late Cenozoic Glacial Ages*. Yale University Press, New Haven: 71-181.
- Mello E Sousa, S. H., S. Servant-Vildary & M. Roux, 1989. Taxonomy and palaeoecology of Quaternary diatoms of Lake Ballivian, Southern Bolivian Altiplano. Brazilian Paleontology Congress, Sept. 1989, Sao Paulo. (Abstract)
- Renberg, I. & T. Hellberg, 1982. The pH history of lakes in southwestern Sweden, as calculated from the subfossil diatom flora of the sediments. *Ambio* 11: 30-33.
- Servant, M. & J. C. Fontes, 1978. Les lacs quaternaires des hauts plateaux des Andes boliviennes. Premières interprétations paléoclimatiques. *Cah. ORSTOM, sér. Géol.* X(1): 9-23.
- Servant-Vildary, S., 1978. Les diatomées des sédiments superficiels d'un lac salé, chloruré-sulfaté sodique de l'altiplano bolivien, 1e lac Poopo. *Cah. ORSTOM, Sér. Géol.* X(1): 79-90.
- Servant-Vildary, S., 1984. Les diatomées des lacs sursalés boliviens. Sous-classe des Pennatophycidées. I.- Famille des Nitzschiacées. *Cah. ORSTOM, sér. Géol.* XIV(1): 35-53.
- Servant-Vildary, S. & M. Roux, 1985. Recherche d'une fonction de transfert Diatomées/Altitudes/Températures actuelles et application à l'holocène. Poster session. Colloque et Séminaire Climat et Développement. 15-16 Octobre 1985, ORSTOM, Bondy, France.
- Servant-Vildary, S. & M. Roux, in press. Multivariate statistical approach to relate water-sediment diatom assemblages and chemical variables in saline lakes from the Lipez area (Southern bolivian altiplano). *Hydrobiologia*.
- Servant-Vildary, S., M. Roux, M. Servant & F. Risacher, 1989. The use of diatoms to infer bolivian quaternary lake salinity. Climatological implications. International Symposium on Global Change in South America, May 1989.
- Van Dam, H., G. Suurmond & C. J. F. ter Braak, 1981. Impact of acidification on diatoms and chemistry of Dutch moorland pools. *Hydrobiologia* 83: 425-459.
- Zar, J. H., 1984. *Biostatistical Analysis*. Prentice Hall, Englewood Cliffs, New Jersey.