Journal Pre-proof

Environmental variables database from a Miocene marine stratigraphical section: a multivariate statistical analysis

Edgardo L. Navarro , M. Sol González Estebenet , M. Verónica Guler , Sabrina Fuentes , José Ignacio Cuitiño , Luis Palazzesi , Juan Pablo Pérez Panera , Viviana Barreda

PII: S0895-9811(21)00086-9
DOI: https://doi.org/10.1016/j.jsames.2021.103239
Reference: DIB 106975

To appear in: *Data in Brief*

Received date: 18 February 2021
Revised date: 12 March 2021
Accepted date: 15 March 2021

Please cite this article as: Edgardo L. Navarro , M. Sol González Estebenet , M. Verónica Guler , Sabrina Fuentes , José Ignacio Cuitiño , Luis Palazzesi , Juan Pablo Pérez Panera , Viviana Barreda , Environmental variables database from a Miocene marine stratigraphical section: a multivariate statistical analysis, *Data in Brief* (2021), doi: https://doi.org/10.1016/j.jsames.2021.103239

Article Title

Environmental variables database from a Miocene marine stratigraphical section: a multivariate statistical analysis

Authors

Edgardo L. Navarro[1], M. Sol González Estebenet[2], M. Verónica Guler[2], Sabrina Fuentes[2], José Ignacio Cuitiño[3], Luis Palazzesi[4], Juan Pablo Pérez Panera[5], Viviana Barreda[4]

Affiliations

[1]Comisión de Investigaciones Científicas (CIC)-CGAMA, Departamento de Geología, Universidad Nacional del Sur (UNS), San Juan 670, 8000 Bahía Blanca, Buenos Aires, Argentina.

[2]Instituto Geológico del Sur (INGEOSUR-CONICET), Universidad Nacional del Sur (UNS), San Juan 670, 8000 Bahía Blanca, Argentina.

[3]Instituto Patagónico de Geología y Paleontología, CCT CONICET-CENPAT, Av. Almirante Brown 2915, U9120ACD Puerto Madryn, Chubut, Argentina.

[4]Museo Argentino de Ciencias Naturales "Bernardino Rivadavia", Av. Ángel Gallardo 470, C1405DJR Ciudad Autónoma de Buenos Aires, Argentina.

[5]CONICET- Laboratorio de Bioestratigrafía, YPF Tecnología (Y-TEC) SA. Avenida del Petróleo Argentino s/ n (e/ 129 y 143), Berisso. 1923, Buenos Aires, Argentina

Corresponding author(s)

Edgardo L. Navarro (enavarro@criba.edu.ar)

Abstract

The data presented here are related to the research article "Miocene Atlantic transgressive-regressive events in northeastern and offshore Patagonia: a palynological perspective" (Guler et al. in press). A total of 60 drilled cutting samples from a 580 m-thick subsurface stratigraphic section (YPF.Ch.PV.es-1 borehole) in Península Valdés, Chubut Province, Argentina, collected every 10 m, were processed for palynological analysis. The quantitative data were statistically evaluated. In detail, the database contain: 1) raw palynological data - proxy data – from counting under transmitted light microscope; 2) four paleoenvironmental variables selected to conduct a multivariate analysis: terrestrial/marine ratio, acritarchs, outer neritic dinocyst taxa and warm-water dinocyst taxa; 3) transformed variables used for the Principal Component Analysis (PCA) and the PC scores obtained, stratigraphically ordered from the top to bottom of the borehole. Data from future studies in new sites combined with here presented data, can be useful to refine paleoenvironment models applied to basin analysis.

Keywords

Palynological –proxies- data, Paleoenvironmetal variables, Principal Component Analysis, Miocene marine events

Subject	Geology
Specific subject area	Palynological – proxies- data applied in stratigraphy using Principal Components Analysis.
Type of data	Tables
How data were acquired	Transmitted light microscope (Leica DM 2500), XLSTAT software.
Data format	Raw Analyzed Transformed
Parameters for data collection	The aleatory counting of the original palynological data –palynological remains - were done in organic residues sieved through meshes of 10 µm. Counts of 250 palynomorphs were considered a representative population sampling for the statistical analysis. In particular, dinocysts were counted from 25 µm sieved palynological residue throughout the slide (Table 1).
Description of data collection	Data come from 60 drilled cutting samples collected every 10 m from a 580 m-thick Miocene subsurface section. Samples were processed for palynological analysis. Aleatory counting of palynological data were done using transmitted light microscope Leica DM 2500. Slides were stored at the Laboratorio de Palinología, Universidad Nacional del Sur, Bahía Blanca (LPUNS).
Data source location	YPF.Ch.PV.es-1 borehole, Península Valdés, Chubut Province, Argentina (42° 44´ 46.26´´S 63° 39´ 26.58´´W).
Data accessibility	Raw data are available as supplementary tables (Word file) at Mendeley data repository: http://dx.doi.org/10.17632/wjyjdz2hn6.1

Related research article	M.V. Guler, M.S. González Estebenet, E.L. Navarro, S. Fuentes, J.I. Cuitiño, L. Palazzesi, J.P. Pérez Panera, V. Barreda (in press). Miocene Atlantic transgressive- regressive events in northeastern and offshore Patagonia: a palynological perspective. J. South Am. Earth Sci.

Value of the Data

- The raw data provide a detailed palynological composition for the Miocene at middle latitudes from the Southern Hemisphere (Table 1), useful to integrate with other palynological dataset from coeval sites to contribute to the basin analysis.
- The transformed data – Principal Components scores – allowed us to recognize and differentiate paleoenvironments, transgressive – regressive trends and the identification of major marine event (e.g. maximum flooding – key features useful for sequence stratigraphy).
- These dataset can be useful for researchers from other disciplines, who can combined our data with their own to refine paleoenvironment models applied to basin analysis.

1. Data Description

These data support the research article entitled "Miocene Atlantic transgressive- regressive events in northeastern and offshore Patagonia: a palynological perspective", by Guler et al., in press [1] . The dataset here reported include:

(1) The number of palynomorphs –proxy data- classified in seven categories (dinoflagellate cysts, pollen grains, spores, acritarchs, chlorococcalean algae, prasinophycean algae, foraminiferal linings), counted from 10 µm sieved palynological residues, reaching a total of 250 individuals. Additionally, the number of specimens of dinocyst taxa, counted throughout each slide from 25 µm sieved palynological residues. . The 60 analyzed samples from the YPF.Ch.PV.es-1 borehole are stratigraphically ordered from top to bottom. (Table 1, doi: 10.17632/wjyjdz2hn6.1).

(2) The environmental variables used in the Principal Component Analysis (PCA). The environmental variables were obtained from selected ecologically important individuals and groups of dinocyst taxa as well as groups of palynomorphs. Variables used in the multivariate analysis were log10 transformed in order to find the best-fit statistical model (Table 2).

(3) The correlation matrix, the eigenvalues (linear combinations of the original variables) which reflect the quality of the projection from the N-dimensional initial to a lower number of dimensions, and the link of the variable with the PCs axis (Table 3).

(4) The Principal Component (PC) scores related to each sample, estimated with the multivariate analysis (Table 4). Each sample in its PC -PC scores- was represented in a biplot graph, 1PC and 2PC (Fig. 6 in Guler et al., in press) and also, stratigraphically from top to bottom through the YPF.Ch.PV.es-1 borehole (Fig. 7 in Guler et al., in press).

2. Experimental Design, Materials and Methods

2.1 Source of raw palynological data

Data come from 60 drilled cutting samples collected every 10 m from a 580 m-thick Miocene subsurface sedimentary succession (YPF.Ch.PV.es-1 borehole), Valdes Basin. Samples were processed for palynological analysis using hydrochloric and hydrofluoric acids in order to remove carbonates and silicates, respectively. Organic residuies were sieved through meshes of 10 and 25 µm and dehydrated with ethanol in order to prepared strew mounts using UV-curable acrylates (Trabasil ® NR2) as a mounting medium. Slides scanning was undertaken under light microscopy Leica DM 2500. With exception of six samples, the raw data were obtained from counting of more than 250 palynomorphs larger than 10 µm per sample (Table 1). Additionally, dinocysts were counted from 25 µm sieved palynological residue throughout the sample (Table 1).

2.2 Selection of paleoenvironmental variables and multivariate statistical analysis

In order to extract paleoenvironmental variables, it was critical to establish meaningful paleoecological and climatic information from fossil palynological remains considered as proxy data. The criteria to select the variables to include in the Principal Components Analysis (PCA) were comprehensively discussed in Guler et al., in press.The variables terrestrial/marine, acritarchs, outer neritic dinocyst taxa and warm-water dinocyst taxa, were used to analyze the environmental conditions represented in the set of samples from the subsurface stratigraphic section (YPF.Ch.PV.es-1 borehole). To reduce the dimensionality of the space of the variables, a PCA was performed. The logarithmic transformation, to base 10 (log10), of variables before PCA was done (Table 2). In order to avoid most of the information of any of the variables be expressed by only one of the components, the correlation matrix was used. The dissimilarity between samples was established using: (1) the PC scores (Table 4) corresponding to the highest percentage of the associated variance (first and second PC) (Table 2), and (2) the percentage of correlation of the variables with each PC (Table 3). In a biplot graph (first and second PC values), samples become grouped according to similar environmental conditions (Fig. 3, Guler et al., in press). Next, the first and second PC values were represented through the stratigraphical log (Fig. 4., Guler et al., in press), showing the evolution of the environmental conditions, allowing the recognition of key features related to the sequence stratigraphy.

CRediT author statement

Edgardo L. Navarro: Writing - Original Draft, Methodology, Format Analysis. **M. Sol González Estebenet**: Writing - Original Draft, Methodology, Investigation. **M. Verónica Guler**: Writing - Original Draft, Investigation. **Sabrina Fuentes:** Investigation**. José Ignacio Cuitiño**: Conceptualizacion. **Luis Palazzesi:** Conceptualizacion. **Juan Pablo Pérez Panera:** Resources, Conceptualizacion. **Viviana Barreda:** Conceptualizacion.

Acknowledgments

This work was partially supported by the Universidad Nacional del Sur – UNS – (grant PGI 24/ZH26-UNS) and, YPF Tecnología – T-TEC (grant Y-TEC I+D+i 620). Palynological processing of samples was appreciatively prepared by Pablo Díaz and Luciano Baraldi.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

Table 2: Environmental variables used in the Principal Component Analysis expressed as percentages and log10 *transformation*. Outer: outer neritic to oceanic dinocyst taxa; T/M: terrestrially-derived/ marine palynomorphs ratio; Acrit: acritarch over total dinocysts; Warm: warm-water dinocyst taxa. Depth in meters below ground surface. Samples 2, 4 and 11 were considered outliers and therefore eliminated from the analysis to obtain a better fit of the statistical model.

MUESTRAS		%0UTER	T/M	%ACRIT	%CALIDOS	log%CAL	log%OUTER	Log%T/M	log%ACRIT
1	60/65	0.53	0.43	23.68	5.79	0.762639	-0.278754	-0.364699	1.374459
3	70/75	0.01	1.09	192.31	0.01	-2.000000	-2.000000	0.037789	2.283997
5	80/85	0.01	1.41	271.43	8.93	0.950782	-2.000000	0.148063	2.433656
6	85/90	0.01	0.44	84.03	5.04	0.702604	-2.000000	-0.356547	1.924453
7	90/95	1.37	2.31	252.05	4.11	0.613798	0.136677	0.364082	2.401495
8	95/100	8.70	1.19	91.30	0.01	-2.000000	0.939302	0.076388	1.960491
9	100/105	0.01	2.26	25.81	4.03	0.605548	-2.000000	0.354404	1.411728
10	115/120	0.01	2.10	65.38	1.28	0.107905	-2.000000	0.322219	1.815476
12	130/135	2.39	1.94	0.01	0.48	-0.320146	0.378824	0.288796	-2.000000
13	145/150	0.55	0.19	12.30	2.60	0.414243	-0.262451	-0.725912	1.089731
14	150/155	1.27	3.61	40.51	7.59	0.880524	0.102373	0.557350	1.607523
15	155/160	9.11	0.31	0.01	10.44	1.018552	0.959431	-0.507084	-2.000000
16	165/170	9.39	0.04	0.01	22.18	1.345925	0.972737	-1.411245	-2.000000
17	170/175	3.73	0.47	0.01	20.04	1.301839	0.571460	-0.325697	-2.000000
18	175/180	4.69	0.47	0.01	12.03	1.080399	0.671464	-0.328353	-2.000000
19	185/190	3.11	0.25	0.01	17.49	1.242716	0.492594	-0.595221	-2.000000
20	205/210	9.23	2.15	0.01	10.77	1.032185	0.965238	0.333215	-2.000000
21	210/215	1.59	1.12	0.01	10.48	1.020203	0.200659	0.047292	-2.000000
22	220/225	0.94	0.85	0.01	30.63	1.486076	-0.028029	-0.071882	-2.000000
23	235/240	0.01	1.68	0.01	12.08	1.082086	-2.000000	0.224734	-2.000000
24	240/245	0.89	2.81	0.01	21.43	1.330993	-0.049218	0.448158	-2.000000
25	245/250	1.82	2.22	0.01	20.00	1.301030	0.259637	0.345508	-2.000000
26	255/260	1.49	0.67	0.01	32.84	1.516348	0.173925	-0.171292	-2.000000

27	270/275	0.91	0.75	0.01	25.00	1.397940	-0.041393	-0.123818	-2.000000
28	275/280	1.68	1.02	0.01	24.28	1.385228	0.226005	0.007488	-2.000000
29	280/285	3.42	0.63	0.01	12.07	1.081811	0.533627	-0.198046	-2.000000
30	285/290	1.08	0.83	0.01	5.38	0.730487	0.031517	-0.081160	-2.000000
31	290/295	1.59	1.42	0.01	11.11	1.045757	0.200659	0.153483	-2.000000
32	295/300	1.92	1.42	0.01	13.94	1.144335	0.283997	0.152967	-2.000000
33	300/305	1.20	0.41	0.01	14.40	1.158362	0.079181	-0.390430	-2.000000
34	310/315	4.07	1.50	0.01	18.70	1.271823	0.609065	0.176091	-2.000000
35	315/320	0.48	0.70	0.01	13.94	1.144335	-0.318063	-0.154902	-2.000000
36	320/325	1.47	1.37	0.01	12.45	1.095316	0.165897	0.137080	-2.000000
37	325/330	0.01	3.70	0.01	25.64	1.408935	-2.000000	0.568202	-2.000000
38	330/335	1.36	0.29	0.01	19.73	1.295081	0.133713	-0.530566	-2.000000
39	335/340	3.33	0.65	0.01	16.67	1.221849	0.522879	-0.184865	-2.000000
40	345/350	5.88	2.73	0.01	26.47	1.422764	0.769551	0.435729	-2.000000
41	355/360	0.01	1.27	0.01	24.00	1.380211	-2.000000	0.104735	-2.000000
42	360/365	0.01	0.96	0.01	14.00	1.146128	-2.000000	-0.017585	-2.000000
43	365/370	9.52	0.69	0.01	13.10	1.117113	0.978811	-0.162996	-2.000000
44	375/380	4.74	0.09	0.01	21.33	1.328930	0.675718	-1.029963	-2.000000
45	380/385	4.43	0.32	0.01	15.76	1.197654	0.646746	-0.501390	-2.000000
46	385/390	4.07	0.09	0.01	21.95	1.341459	0.609065	-1.049218	-2.000000
47	390/395	1.39	0.06	0.01	36.38	1.560835	0.144010	-1.221228	-2.000000
48	395/400	1.05	0.14	0.01	21.20	1.326422	0.019997	-0.855317	-2.000000
49	400/405	1.45	0.25	0.01	12.58	1.099802	0.160667	-0.599744	-2.000000
50	405/410	1.23	0.01	0.01	13.09	1.116884	0.088142	-1.910269	-2.000000
51	410/415	24.75	0.09	0.01	35.83	1.554281	1.393596	-1.060698	-2.000000
52	415/ 420	28.27	0.04	0.01	38.19	1.581900	1.451326	-1.424155	-2.000000
53	425/ 430	26.77	0.09	0.01	50.00	1.698970	1.427675	-1.033807	-2.000000
54	430/ 435	8.64	0.21	0.01	27.75	1.443243	0.936451	-0.678285	-2.000000

55	435/ 440	8.62	0.45	0.01	16.67	1.221849	0.935542	-0.347655	-2.000000
56	445 /450	11.29	0.33	0.01	13.71	1.137027	1.052706	-0.486827	-2.000000
57	475/480	1.05	0.49	0.01	15.66	1.194733	0.022276	-0.308903	-2.000000
58	510/515	1.81	0.17	0.01	6.14	0.787969	0.256490	-0.780214	-2.000000
59	550/555	8.00	0.23	0.01	18.29	1.262112	0.903090	-0.638272	-2.000000
60	580/585	7.29	0.61	0.01	14.06	1.148063	0.862827	-0.212303	-2.000000

(*)The 0 (zero) values were approximated to 0.01 in order to apply the logarithm transformation.

Table 3: a) Matrix correlation of the variables used in the Principal Component Analysis. b) Eigenvalues corresponding to the factors and percentages amount of variance retained by each principal component (PC), individual and cumulated. c) Contributions (in percentages) of the variables to the principal components.

a) Correlation matrix

Variables	Log10 %Warm	Log10 %Outer	Log10 %t/m	Log10 %Micrh
Log10 %Warm	1.000	0.268	-0.271	-0.643
Log10 %Outer	0.268	1.000	-0.418	-0.478
Log10 %t/m	-0.271	-0.418	1.000	0.282
Log10 %Micrh	-0.643	-0.478	0.282	1.000

b) Eigenvalues

	1PC	2PC	3PC	4PC
Eigenvalue	2.196	0.893	0.611	0.300
Variability (%)	54.900	22.325	15.271	7.505
Cumulative percentage	54.900	77.225	92.495	100.000

c) Contribution of the variables (%)

	1PC	2PC	3PC	4PC
Log10 %Warm	-26.083	-28.492	-11.725	33.700
Log10 %Outer	-23.878	17.747	45.690	12.685
Log10 %t/m	17.677	-41.067	39.031	2.224
Log10 %Micrh	32.362	12.693	-3.554	51.391

Table 4: Principal Component scores (PC) related to each stratigraphically distributed sample.

Sample	1PC	2PC	3PC	4PC
1	1.341	-0.817	-0.562	0.950
3	-4.750	-2.019	0.013	-1.500

5	2.858	0.417	-1.381	1.150
6	2.458	-0.254	-1.771	0.548
7	2.212	-0.499	0.507	1.673
8	3.405	-3.149	2.111	-0.587
9	2.885	0.635	-0.836	0.410
10	3.394	0.104	-0.671	0.168
12	1.037	-0.364	1.626	-1.215
13	1.206	-1.459	-0.754	0.416
14	1.869	0.149	0.673	1.557
15	-0.884	-0.492	0.424	-0.070
16	-1.842	-1.308	-0.777	-0.031
17	-0.767	0.113	0.224	0.085
18	-0.650	-0.108	0.402	-0.071
19	-0.892	-0.219	-0.111	-0.069
20	-0.245	0.511	1.391	0.176
21	-0.080	0.489	0.544	-0.189
22	-0.413	0.815	0.012	0.097
23	1.100	1.687	-0.790	-0.881
24	0.120	1.317	0.677	0.099
25	-0.090	1.040	0.785	0.157
26	-0.613	0.636	0.020	0.169
27	-0.379	0.690	-0.012	0.002
28	-0.400	0.721	0.329	0.124
29	-0.482	0.106	0.457	-0.084
30	0.124	0.180	0.427	-0.536
31	-0.016	0.635	0.653	-0.138
32	-0.133	0.677	0.659	-0.023
33	-0.464	0.133	-0.115	-0.234

34	-0.373	0.667	0.844	0.211
35	-0.074	0.570	-0.108	-0.325
36	-0.050	0.670	0.585	-0.112
37	1.119	2.353	-0.560	-0.504
38	-0.704	0.052	-0.309	-0.135
39	-0.572	0.237	0.394	0.037
40	-0.365	1.026	1.176	0.470
41	0.781	1.782	-1.080	-0.657
42	0.864	1.451	-1.102	-0.892
43	-0.702	-0.015	0.784	0.117
44	-1.386	-0.744	-0.532	-0.048
45	-0.861	-0.210	0.125	-0.026
46	-1.377	-0.728	-0.606	-0.067
47	-1.447	-0.559	-1.234	-0.093
48	-0.924	-0.259	-0.778	-0.239
49	-0.623	-0.196	-0.271	-0.313
50	-1.618	-1.703	-1.842	-0.686
51	-1.936	-0.908	-0.190	0.397
52	-2.268	-1.341	-0.584	0.341
53	-2.042	-0.775	-0.210	0.541
54	-1.328	-0.348	-0.005	0.241
55	-0.903	-0.132	0.488	0.141
56	-1.005	-0.414	0.451	0.072
57	-0.400	0.282	-0.078	-0.201
58	-0.574	-0.698	-0.255	-0.597
59	-1.143	-0.430	0.110	0.084
60	-0.706	0.001	0.632	0.088