

Dokumentation
MacMAT
– Modern Analog Technique –

Dr. Rainer Sieger
Alfred-Wegener-Institut für
Polar- und Meeresforschung
Postfach 12 01 61
27515 Bremerhaven
e-mail: rsieger@pangaea.de

14. Juni 1995

Inhaltsverzeichnis

1	Theoretische Grundlagen	3
2	Berechnung der Ähnlichkeitsbeziehung	3
3	Berechnung des Paläoumweltparameters	4
4	Standardabweichung	4
5	Bedienung	5
5.1	Referenzdatensatz	5
5.2	Probendatensatz	6
5.3	Umweltparameterdatensatz	6
5.4	Berechnung	6
6	Literatur	8

1 Theoretische Grundlagen

Die *Modern Analog Technique* benutzt zur Bestimmung von Paläotemperaturen Ähnlichkeitsbeziehungen zwischen zwei Objekten. Die Ähnlichkeit wird dabei als *similarity index* bezeichnet. Die Objekte bestehen je aus einer gleichgroßen Anzahl von Variablen (z.B. Foraminiferen-Arten), die normiert auf die Gesamtanzahl angegeben werden. Um eine hohe Ähnlichkeit zu erzielen, genügt die Übereinstimmung in der proportionalen Zusammensetzung zwischen der Referenzprobe auf der einen und der Kernprobe auf der anderen Seite.

Die Berechnung der Paläotemperatur wird in zwei Schritte unterteilt.

1. Schritt Berechnung des *similarity index* (S_i)
2. Schritt Berechnung des Umgebungsparameters, basierend auf dem S_i -Wert

Da bei dieser Methode keine Regression über einen gesamten Datensatz erfolgt, sondern nur eine Beziehung von Objekt (Referenzprobe) zu Objekt (Kernprobe), ist diese Methode besonders flexibel in der Anwendung. Sowohl der Probandatensatz, als auch der Referenzdatensatz können schnell ergänzt bzw. verändert werden (Meinecke, 1992).

2 Berechnung der Ähnlichkeitsbeziehung

Der Winkel θ zwischen zwei Vektoren gilt als Maß für die Ähnlichkeit.

Es gilt

- $\theta = 0^\circ$ ($\cos \theta = 1$), die Vektoren sind collinear, d.h. die Probenzusammensetzung ist identisch.
- $\theta = 90^\circ$ ($\cos \theta = 0$), die Vektoren sind orthogonal zueinander, d.h. beide Proben sind vollständig verschieden.

Bei diesen Betrachtungen werden die Vektoren durch die relative Anzahl der Individuen pro Art gebildet, d.h., für die Berechnung des S_i ($S_i = \cos \theta$) werden die Rohzählraten der Proben normalisiert. Anschließend kann die Formel von Imbrie und Purdy (1962) angewendet werden. Es ergibt sich die Beziehung

$$S_i = \frac{\sum_{i=1}^n P_i \cdot R_i}{\sqrt{\sum_{i=1}^n P_i^2 \cdot \sum_{i=1}^n R_i^2}} \quad (1)$$

mit

n	Anzahl der Arten
n_r	Anzahl der Referenzproben
n_k	Anzahl der Kernproben
i	Zählindex für Arten ($i = 1, \dots, n$)
j	Zählindex für Referenzprobe ($j = 1, \dots, n_r$)
k	Zählindex für Kernprobe ($k = 1, \dots, n_k$)
P	relatives Zählergebnis in der Probe
R	relatives Zählergebnis in der Referenzprobe.

Diese Berechnung wird für alle Kombinationen von Referenz- und Kernproben durchgeführt. Die Ergebnisse der Berechnung bilden somit eine Matrix in der jeder Kernprobe (P_k) ein Si -Wert je Referenzprobe (R_j) zugeordnet wird. Statt des Si -Wertes wird bei Imbrie und Purdy (1962) der Wert θ' angegeben. Dieser berechnet sich durch

$$\theta' = \frac{45 - \theta}{45}. \quad (2)$$

Durch diese Transformation lässt sich erreichen, dass für einen *similarity index* von 0 der Wert $\theta' = 1$ bzw. für $Si = 0$ der Wert $\theta' = -1$ ausgegeben wird.

	A	B	C	D	E	Si	θ'
a (Referenzprobe)	10	20	30	40	0	1	+1.000
b	5	10	15	20	0	1	+1.000
c	5	10	15	50	0	0.92	+0.498
d	10	30	20	4	0	0.71	+0.002
e	40	30	20	10	0	0.67	-0.071
f	0	0	0	0	80	0	-1.000
Kernprobe	5	13	33	38	10	0.97	+0.679

Tabelle 1: Die Tabelle zeigt einige mit einem Excel-Worksheet gerechnete Beispiele. Zu Bemerkem ist, dass die Referenzprobe zu sich selbst ähnlich ist ($\theta' = 1$).

3 Berechnung des Paläoumweltparameters

Zur Berechnung eines Paläoumweltparameters (Y_{est}) wird jeder rezente Oberflächenprobe (R_j) ein aktueller Umweltparameter (Y_R) zugeordnet, der an der geographischen Position der Oberflächenprobe herrscht. Der Umweltparameter wird damit durch eine definierte Vergesellschaftung von Foraminiferenarten am darunterliegenden Meeresboden repräsentiert. Jeder Kernprobe (P_k) kann damit auf der Basis seiner Foraminiferenvergesellschaftung und einer Gewichtung durch den Ähnlichkeitskoeffizienten (Si), ein modernes Analogon zugeordnet werden. Für die Berechnung des Paläoumweltparameters Y_{est} werden nur eine vom Anwender bestimmbare Anzahl (n_s) von Referenzproben herangezogen. Diese Referenzproben weisen dabei die bezüglich der untersuchten Kernprobe n_s höchsten Si -Werte auf.

Die Berechnung von Y_{est} erfolgt dann nach Houston (1980) als gewogenes arithmetisches Mittel

$$Y_{\text{est}} = \frac{\sum_{j=1}^{n_s} Si_{R_j,P} \cdot Y_{R_j}}{\sum_{j=1}^{n_s} Si_{R_j,P}}. \quad (3)$$

Eine Verschärfung lässt sich erreichen, wenn nur die Referenzproben benutzt werden deren Ähnlichkeit zur untersuchten Kernprobe sehr groß ist. Sinnvoll ist demnach die Angabe eines $\epsilon \in (0, 1]$. Eine Referenzprobe R_j wird nur dann zur Berechnung herangezogen, wenn gilt

$$Si_{R_j,P} > 1 - \epsilon \quad (4)$$

4 Standardabweichung

Neben dem berechneten Umweltparameter Y_{est} werden für jede Station auch die Umweltparameter der n_s besten Referenzproben ausgegeben. Aus diesen n_s besten Umweltparametern

wird jeweils eine gewichtete Standardabweichung berechnet. Für die Standardabweichung σ gilt dabei (z.B. Bartsch, 1992)

$$\sigma_k = \frac{1}{N-1} \sum_{j=1}^{n_s} (Y_j - \bar{Y})^2 \cdot Si_{j,k} \quad (5)$$

mit

j	Zählindex für die n_s besten Referenzprobe ($j = 1, \dots, n_s$)
k	Zählindex für Kernprobe ($k = 1, \dots, n_k$)
N	$\sum_{j=1}^{n_s} Si_{j,k}$
Y	die Umweltparameter der n_s besten Referenzproben
\bar{Y}	arithmetisches Mittel $\left(\bar{Y} = \frac{\sum_{j=1}^{n_s} Y_j}{n_s}\right)$
Si	Ähnlichkeiten der Kernprobe zu den n_s besten Referenzproben

5 Bedienung

Die Bedienung des Programms *MacMAT* gliedert sich in zwei Teile,

- die Erstellung der Datensätze
- der Aufruf des Programms

Das Programm lädt zur Laufzeit drei Datensätze. Es sind dies

- der Referenzdatensatz
- der Probandatensatz
- der Umweltparameterdatensatz

5.1 Referenzdatensatz

Der Referenzdatensatz besteht aus einer Kopfzeile und maximal 100 Datenzeilen (= 100 Arten). Die Kopfzeile enthält die Namen der Referenzproben. Die Anzahl der Referenzproben ist auf 1000 begrenzt. Jeder Name darf maximal 10 Zeichen lang sein und keine Leerzeichen enthalten¹. Maximal dürfen sich in dieser Zeile nur 4000 Zeichen befinden. Die Namen sind durch Tabulatorstops voneinander zu trennen.

Jede der dann folgenden Datenzeilen beginnt mit dem Namen der durch diese Zeile repräsentierten Species. Der Speciesname darf dabei maximal 63 Zeichen lang sein und keine Leerzeichen enthalten. In den weiteren maximal 1000 Spalten wird für jede Referenzprobe und jede Species die Anzahl der gezählten Individuen eingetragen. Wurde kein Individuum gefunden, so muSS eine 0 eingetragen werden. Die einzelnen Spalten sind durch Tabulatorstops voneinander zu trennen. Jede Datenzeile darf maximal 4000 Zeichen enthalten.

¹Sinnvollerweise ersetzt man die Leerzeichen vorher mit einem Texteditor.

Beispiel (Das Zeichen → steht für den Tabulatorstop)

Taxa/_Station_(PS)	→	1141	→	1146	→	1147	→	1160
Actinocyclus_actinochilus	→	3	→	6	→	8	→	0
Chaetoceros_spp.	→	348	→	363	→	386	→	593
Nitzschia_cylindrus	→	7	→	2	→	1	→	17
Odontella_weissflogii	→	2	→	7	→	22	→	0

Eine solche Datei lässt sich am besten mit einem Tabellenkalkulationsprogramm erstellen und verändern. Die Datei muss jedoch als Textdatei abgespeichert werden !

5.2 Probendatensatz

Der Probendatensatz hat den gleichen Dateiaufbau wie der Referenzprobendatensatz. Aus Speicherplatzgründen wurde die Anzahl der Proben (= Spalten) auf maximal 250 begrenzt. Ansonsten lässt sich das oben gesagte wiederholen.

5.3 Umweltparameterdatensatz

Der Umweltparameterdatensatz enthält die den Referenzproben zugeordneten Umweltparameter. Der Umweltparameterdatensatz besteht wiederum aus einer Kopfzeile und maximal 250 Datenzeilen (= Anzahl der Proben). In der ersten Spalte kann beliebiger Text stehen (z.B. die Stationsnummer). Die folgenden maximal 10 Spalten enthalten die Umweltparameterdatensätze. Jede Spalte erzeugt später einen zugehörigen Ergebnisdatensatz. Der Name der Ergebnisdatei wird dabei vom Spaltentitel gebildet. Damit ist es also möglich, in einem Rechenlauf verschiedene Betrachtungen des selben Probendatensatzes anzustellen. Im Beispieldatensatz ist z.B. der Umweltparameter einfach durch die Stationsnummer repräsentiert bzw. durch die Temperatur in 10 m Wassertiefe.

Beispiel (Das Zeichen → steht für den Tabulatorstop)

Station	→	bestStation	→	Temp.10m
1141	→	1	→	0.678
1146	→	2	→	0.023
1147	→	3	→	0.043
1160	→	4	→	0.562
1248	→	5	→	0.499

5.4 Berechnung

Das Programm *MacMAT* wird durch einen Doppelklick gestartet. Nach dem Start des Programms wird über den Standardfile-Dialog der Dateiname des Referenzdatensatzes eingegeben. Anschliessend wird auf gleiche Weise der Probendatensatz sowie der Umweltparameterdatensatz spezifiziert. Sind alle Datensätze geladen, beginnt die Berechnung der Ähnlichkeitskoeffizienten für alle Referenzen und alle Proben. Dieser Vorgang kann einige Zeit in Anspruch nehmen. Nach der Berechnung der S_i beginnt die Abarbeitung des Umweltparameterdatensatzes. Das Programm zeigt jeweils an, in welches Verzeichnis die Ergebnisdateien geschrieben werden. Nach der Berechnung erwartet das Programm die erneute Eingabe von Dateinamen

für einen weiteren Rechenlauf. Diese Enlosschleife kann durch Wahl des Knopfes *Abbruch* jederzeit verlassen werden.

Das Programm schreibt in das Verzeichnis aus dem der Probandatensatz geladen wurde die folgenden Dateien:

- *Probandatensatz.Si* – die Matrix der Ähnlichkeitskoeffizienten. Diese Matrix kann sehr groSS werden und ist nur für den wirklichen Experten von Interesse.
- *Spaltentitel* – Ergebnisdatei. Diese Textdatei kann zur Weiterverarbeitung von jedem Tabellenkalkulationsprogramm (z.B. Excel) gelesen werden.
- *Spaltentitel.tex* – die gleiche Ergebnisdatei. Diesmal im L^AT_EX 2_ε-Format. Damit kann mit einer bestehenden T_EX-Installation sofort ein formatierter Ausdruck der Ergebnisdatei erzeugt werden (siehe die nachfolgenden Beispiele).

Eine Ergebnisdatei einer Berechnung enthält für jede Probe den berechneten Umweltparameter sowie die Umweltparameter der n_s besten assoziierten Referenzproben. Über diese wird eine Standardabweichung gerechnet die in der letzten Spalte aufgeführt wird.

Ein Beispiel

Probe	berechnete Temperatur	Temperaturen der 5 besten Referenzproben					Standardabweichung
440	2.51	4.986	4.795	1.402	0.251	0.849	2.281
450	3.40	-0.038	2.861	4.923	4.795	4.599	2.138
460	3.45	4.986	1.402	3.971			1.887
470	3.37	2.861	4.795	4.923	3.971	0.251	1.926
480	1.40	1.402					0.000
510	2.52	1.402	4.986	4.795	-1.148		2.971
520	Fehler						Fehler
530	3.44	2.861	3.971	4.923	0.849	4.599	1.654

Die leeren Felder entstehen immer dann, wenn innerhalb der von ϵ vorgegebenen Schranken keine ähnliche Referenzprobe gefunden wurde. Ist eine Zeile völlig leer, so wurde überhaupt keine ähnlich Probe gefunden. *MacMAT* schreibt dann in diesen Fällen das Wort *Fehler* in diese Zeile (siehe Probe 1276).

Einen Sonderfall für eine Berechnung stellt der Vergleich eines Referenzdatensatzes mit sich selbst dar. Das Programm erkennt an den Dateinamen für Referenzdatensatz und Probandatensatz das hier Referenz gegen Referenz gerechnet werden soll. Der Ergebnisdatsatz enthält dann zusätzlich den ursprünglichen Umweltparameter und die Differenz zum berechneten Wert. In die Berechnung des Umweltparameters und der zugehörigen Standardabweichung geht die zu sich selbst am ähnlichste Referenzprobe nicht mit ein.

Beispiel

Probe	berechnete Temperatur	Temperaturen der 5 besten Referenzproben					original Temperatur	Differenz	Standard-abweichung
1181	-0.14	0.023	-0.638	-0.533	0.007	0.448	0.026	0.170	0.447
1182	-0.60	-0.638	-0.622	-0.533			0.023	0.621	0.057
1193	-1.58	-1.641	-1.673	-1.790	-1.132	-1.669	-1.790	-0.208	0.257
1195	-1.62	-1.764	-1.471				-1.790	-0.172	0.217
1197	-1.78	-1.777					-1.710	0.067	0.000
1225	0.08	0.091	0.695	0.298	-0.047	-0.621	-0.110	-0.194	0.483
1273	-0.98	-1.790	-1.022	-1.341	-0.533	-0.190	-0.120	0.857	0.637
1276	Fehler						-1.684	997.316	Fehler
1277	-1.40	-1.790	-1.101	-0.976	-1.790	-1.341	-1.632	-0.233	0.381

6 Literatur

Bartsch, H.-J. (1992): Mathematische Formeln.- 504 S., BZ, Köln.

Houston, W.H. (1980): The Agulhas Current During the Late Pleistocene: Analysis of Modern Faunal Analogs.- Science, **207**: 64-66.

Imbrie, J. und Purdy, E.G. (1962): Classification of modern Bahamian carbonate sediments.- Mem. Assoc. Petr. Geol., **I**: 253-272.

Meinecke, G. (1992): Spätquartäre Oberflächenwassertemperaturen im östlichen äquatorialen Atlantik.- 181 S., Berichte aus dem Fachbereich Geowissenschaften der Universität Bremen, **29**, Bremen.