

KOMPUTEROWA BAZA DANYCH ZLICZEŃ ZIARN PYŁKU

Data base for pollen counts

Adam WALANUS

Summary. The computer program POLPAL (POLish-PALynological...) is generally described. Reasons for using computer for storing and manipulating data are emphasized. POLPAL offer three systems of data input, including counting. As an output different types of diagrams are produced. Data may be transformed by the standard numerical methods. Some algorithms for tables comparison are also available. Export of the data to TILIA [5] format and other is possible.

Key words: data base, data handling, pollen counts, pollen diagram

Dr Adam Walanus, Instytut Fizyki, Politechnika Śląska, ul. Bolesława Krzywoustego 2, 44-100 Gliwice

O CELOWOŚCI ZASTOSOWANIA KOMPUTERA DO PRACY Z TABELAMI PYŁKOWYMI

Przydatność komputera w pracy jest dziś już oczywista. Na przykład przy pisaniu tekstów łatwość wprowadzania poprawek i otrzymywania wydruków w nowych wersjach jest nową jakością. Tabela palinologiczna podlega zawsze, co najmniej dwóm działaniom arytmetycznym. Sumuje się liczby w wierszach, by następnie za pomocą dzielenia otrzymać procenty. Nawet średniej wielkości tabela wymaga ogromnej ilości operacji na kalkulatorze. A przede wszystkim wielokrotnego wpisywania tych samych liczb, które do komputera wpisuje się tylko raz.

Biorąc pod uwagę możliwości graficzne dzisiejszego sprzętu komputerowego trzeba stwierdzić, że ręczne rysowanie diagramów jako zdecydowanie nieekonomiczne odchodzi w przeszłość. Mając dane w pamięci komputera można drukować diagramy w różnych wersjach z łatwością nieporównywalną z pracą ręczną. Oczywiście osiągnięcie sprawności w pracy z kom-

puterem wymaga pewnego przygotowania i pewnej, niewielkiej wprawy.

Zupełnie nierealną, bez komputera, rzeczą jest wykonywanie bardziej zaawansowanych analiz statystycznych czy numerycznych. W literaturze coraz częściej spotyka się tego typu przetwarzanie danych. Celem jego jest podniesienie stopnia obiektywności przedstawianych wniosków.

Na świecie jest wiele programów napisanych z myślą o pracy z tabelami palinologicznymi. Czasem palinolodzy sami piszą programy, jest to sytuacja najlepsza, jeśli założyć odpowiednie umiejętności programistyczne. Najłatwiej jest posługiwać się programem napisanym przez siebie, można go modyfikować i ulepszać według własnych pomysłów. Jeżeli palinolog nie ma zamiaru pisać programu komputerowego, to jest skazany na program napisany przez kogoś innego. Wtedy dobrze jest mieć przynajmniej kontakt z programistą.

Dlatego częstą sytuacją jest, że laboratoria mają własne programy. Ma to wiele zalet w po-

równaniu z posiadaniem obcego programu. Przy szkoleniu użytkowników programu lepsze jest słowo mówione, niż niepełna zwykle instrukcja. Ponadto dobrze jest mieć możliwość zlecenia modyfikacji programu. W pracy naukowej z zasady pojawiają się sytuacje nowe.

Program POLPAL powstawał wiele lat [9], jednocześnie był eksploatowany przez kilka osób z różnych laboratoriów. Posiada on możliwości wbudowane na życzenie różnych użytkowników, zawiera w sobie część ich doświadczenia.

Pojęcie o stanie oprogramowania w omawianej dziedzinie można wyrobić sobie najlepiej przeglądając *Newsletter INQUA-Commission for the Study of the Holocene, Working Group on Data-Handling Methods*, gdzie znaleźć można też informację o POLPAL'u [10, 12].

WPROWADZANIE DANYCH

Dane do bazy POLPAL wprowadzać można na trzech poziomach. Poziomy te odpowiadają różnym stopniom wstępnego, "na papierze" przetworzenia danych. Poziom zerowy, czyli zupełny brak przetworzenia przedkomputerowego jest wtedy, gdy zaobserwowane w okularze mikroskopu szczątki rejestruje się bezpośrednio w komputerze naciskając odpowiednie klawisze. Poziom pierwszy polega na przepisywaniu do komputera protokołów ze zliczania szczątków w poszczególnych próbkach. Jeśli natomiast palinolog posiada gotową tabelę, ułożoną ręcznie z protokołów, to ona oczywiście również może być wprowadzona do pamięci komputera.

Osobnym zagadnieniem z dziedziny wprowadzania danych jest lista taksonów. W programie POLPAL obowiązkowo musi być utworzona lista nazw taksonów. W standardowej wersji może ona liczyć do 1000 nazw o długości do 25 liter. Lista taksonów może być do pewnego stopnia modyfikowana. W zasadzie, każdy użytkownik bazy danych może mieć własną listę taksonów, jednak trzeba tu mieć na uwadze ewentualną wymianę danych pomiędzy różnymi osobami, lub przesłanie danych do centralnej bazy danych. W takim wypadku posiadanie zupełnie

odmiennej listy taksonów poważnie utrudni wymianę danych na poziomie komputerowym. Zagadnienie obowiązywania standardowej, krajowej czy regionalnej listy taksonów wykracza poza ramy niniejszego tekstu.

ZLICZANIE SZCZĄTKÓW Z KOMPUTEREM

W momencie podjęcia decyzji do jakiego taksonu należy zaobserwowane w próbce ziarno pyłku (sporomorfa, szczątek makroskopowy, wioślarka, okrzemka itp.), trzeba tę decyzję zarejestrować i przejść do następnego szczątku. Tak więc pojawia się kwestia jak najszybszego zapisania faktu, że zaobserwowano jedno ziarno pewnego taksonu. Taksonów jest kilkaset. Jednak tylko kilka z nich pojawia się często. Sytuacja jest dość złożona, toteż i optymalne rozwiązanie nie jest całkiem proste [12]. Zakłada się tutaj, że mając przed sobą dziesiątki tysięcy ziarn do zliczenia, warto poświęcić czas na zapoznanie się z programem, by zaoszczędzić później wielokrotnie więcej czasu dzięki sprawniej pracy.

Dziesięć najczęściej występujących taksonów zakodować można za pomocą tzw. klawiszy funkcyjnych F1, F2,... F10. O tym, które to będą taksony decyduje dowolnie użytkownik bazy danych. Rzadziej pojawiającym się taksonom przypisuje się kody dwuliterowe. Kodów takich mogłoby być $26 \times 26 = 676$, jednak nie wszystkie kombinacje liter nadają się do zapamiętania, np. „iy”. Ile taksonów zapamiętać przez kody dwuliterowe, należy do decyzji użytkownika. Pozostałe taksony kodowane są trzyliterowo.

W każdej chwili bardzo łatwo można wywołać na ekran komputera słownik przypominający znaczenie kodów.

Zliczanie polega na naciśnięciu jednego z klawiszy funkcyjnych lub kolejno dwóch liter lub odstępu i trzech liter (odstęp jest wskazówką dla komputera, że ma czekać na trzecią literę). Po naciśnięciu kodu, komputer wyświetla nazwę taksonu, informację o stanie zliczania i czeka na następny kod. Pomyłki można kasować. Sekwencje liter, które nie kodują żadnego taksonu nie wpisują się.

Zliczanie można w każdej chwili przerwać i kontynuować później.

WPISYWANIE PROTOKOŁÓW ZE ZLICZANIA

Protokół ze zliczania szczątków w próbkę ma postać następującego typu: *Pinus* 123, *Betula* 45, *Alnus* 67 itd.

Wprowadzanie takiego zapisu do komputera polega na wpisywaniu kodu literowego taksonu i liczby ziarn. W powyższym przykładzie naciskać trzeba następujące klawisze: pi123be45 aln67 itd. Kod trzyliterowy nie wymaga specjalnego zaznaczania gdyż kody są rozdzielone liczbą i komputer odróżniający cyfry od liter potrafi właściwie odczytać kod. W każdej chwili można przywołać na ekran słownik kodów.

WPISYWANIE TABEL

Tabele wpisuje się kolumnami, czyli taksonami. Tak więc najpierw należy wybrać z listy taksonów odpowiednią nazwę. Następnie komputer oczekuje na podaną wcześniej ilość liczb. Wpisywanie dziesiątków lub setek liczb jest pracochłonne, dlatego w programie POLPAL zastosowano dwa ułatwienia wykorzystujące pewną charakterystyczną cechę tabel palinologicznych. Chodzi o przewagę małych liczb i zer w tych tabelach. Wiele taksonów jest reprezentowanych przez pojedyncze ziarna w pojedynczych próbkach. Duże ilości zer wpisuje się podając ich ilość z minusem. Np. wpisanie liczby 22 spowoduje wstawienie w kolumności dwudziestu dwóch zer.

Mając do wpisania szereg liczb jednocyfrowych, np. 1, 2, 0, 4, 1, 8 itd. można naciskając kropkę przełączyć program na opcję "one digit". Komputer czeka wtedy na pojedynczą cyfrę w związku z czym nie trzeba naciskać klawisza ENTER, co daje 50% oszczędności uderzeń w klawiaturę. Powrotu do normalnego wpisywania długich liczb dokonuje się przez ponowne naciśnięcie kropki.

Kolejne wpisane taksony umieszcza się na dowolnej pozycji względem już wpisanych. Nie można przerwać wpisywania taksonu w połowie, wyłączenie komputera spowoduje stratę wpisanej już części kolumny. Istnieje wiele możliwości poprawiania wpisanych danych.

TABELA I DANE UZUPEŁNIAJĄCE

Na wstępie trzeba zaznaczyć, że wpisywane protokoły zliczania nie stają się automatycznie tabelą. Tabelę można z nich dopiero utworzyć. Proces taki przebiega automatycznie jeśli podawane numery protokołów rosną z góry na dół, jak głębokość w [cm]. W przeciwnym wypadku trzeba poinformować komputer o właściwej kolejności próbek.

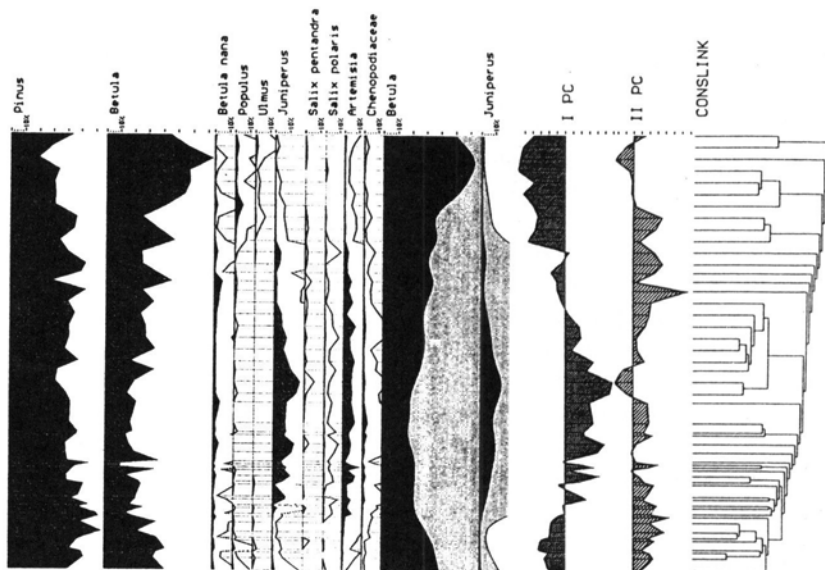
Kolejność taksonów w tabeli ustawiana jest automatycznie. Służą do tego jednoliterowe kody przyporządkowane taksonom, oznaczające tzw. formy życiowe. W przypadku palinologii na początku tabeli umieszczone będą drzewa, w porządku alfabetycznym. Dalej krzewy, krzewinki, rośliny zielne itd. O tym czy *Salix* jest krzewem czy drzewem komputer jest informowany właśnie poprzez jednoliterowy kod, który musi być wprowadzony razem z taksonem.

Wyniki zliczania pojedynczych ziarn wpisują się do zbioru protokołów, tak więc przechodzenie do tabeli odbywa się w wyżej opisanym sposób.

Tabela identyfikowana jest w programie poprzez swój numer. Tabela ma swoją nazwę, o którą komputer w odpowiednim momencie zapyta, prosząc o jej wprowadzenie. Tabela zapamiętywana jest w sposób skondensowany, zajmuje mało miejsca na dysku twardym, czy elastycznym [11].

Tabela zawiera jedynie liczby szczątków, przy czym kolumny tabeli są opatrzone nazwami taksonów. Wiersze natomiast są w tabeli jedynie ponumerowane kolejnymi liczbami. Wszelkie inne informacje są traktowane jako dodatkowe względem tabeli. Pierwszym przykładem danych dodatkowych jest zestaw numerów próbek lub ich głębokości.

Liczby te mieszczą się w specjalnym pliku, który trzeba w programie utworzyć i wypełnić. Dane dodatkowe, jak sama nazwa wskazuje, nie są konieczne do pracy nad tabelą. Dane te mogą mieć różne wersje, można mieć np. kilka plików z głębokościami próbek, odróżniały się będą one jednoliterowym, dodatkowym symbolem. Inne dane dodatkowe to ilości dodanego do próbek tzw. wskaźnika egzotycznego lub odstępu międ-



Ryc. 1. Diagram palinologiczny wydrukowany na drukarce igłowej z programu POLPAL, z wykresami danych przetworzonych. Obok zwykłych krzywych procentowych wydrukowane są dwie krzywe wygładzone, dwie pierwsze składowe główne i dendrogram CONSLINK.

Fig. 1. The palynological diagram produced by the POLPAL program at the 24-needle printer. Besides of the usual %-curves the smoothed version is given, also principal components and CONSLINK dendrogram.

dzy próbkami na diagramie, podane w milimetrach.

Wspomniane trzy typy plików z danymi dodatkowymi powiązane są z próbkami. Ponadto wprowadzić można do pamięci zonację profilu oraz zestawy taksonów przydatne przy wielokrotnym druku diagramów według jednego schematu.

OBLICZENIA

Najprostszą, podstawową operacją arytmetyczną wykonywaną w programie, jest obliczanie procentowego udziału pyłku poszczególnych taksonów w próbkach. Procenty obliczyć można na bazie dowolnej sumy, program standardowo proponuje pewną sumę, można ją jednak zmienić. Część możliwości obliczeniowych łączy się z sumowaniem taksonów, można np. zsumować wszystkie drzewa i wszystkie rośliny zielne. Łatwo wtedy obliczyć np. procent drzew względem zielnych (co niekoniecznie musi mieć jakiś sens) lub procent drzew względem sumy drzew i zielnych.

Jeżeli w próbkach był liczony wskaźnik egzotyczny, dodawany do znanej objętości próbki, to można obliczyć bezwzględną koncentrację szczątków, a nie tylko ich udział procentowy [3]. Koncentrację można obliczać na jednostkę objętości, jednostkę masy, suchej masy itd. Bardzo skomplikowane obliczenia przygotować można za pomocą specjalnego podprogramu dodającego, mnożącego, dzielącego i wykonującego operację uzupełnienia do 100% na plikach zawierających informację o objętościach, masach itp. próbek [8].

Wyniki wszystkich obliczeń mogą być przedstawione na ekranie i wydrukowane. Wyświetlić i wydrukować można dowolne taksony w dowolnej kolejności i dowolny fragment profilu.

DRUK DIAGRAMÓW

Tabela palinologiczna (i inne), odpowiednio przeliczona przedstawiana jest najczęściej graficznie w postaci diagramu.

Zważywszy, że rozpowszechnione obecnie urządzenia drukujące potrafią rysować kropki

o wielkości rzędu dziesiątej części milimetra, można spodziewać się przewagi techniki komputerowej nad rysownikiem. Gwoli sprawiedliwości trzeba jednak dodać, że typowa drukarka laserowa rysuje tylko na zwykłej kartce A4, a rysownik może kreślić po arkuszu dowolnej wielkości. Jednak już stosunkowo tania drukarka 24-igłowa z „długim wałkiem” może wydrukować diagram o wysokości 34 cm i dowolnej długości (zawierający dowolnie dużo taksonów). Program POLPAL ma możliwość rysowania diagramów w częściach – otrzymanie diagramu o rozmiarach np. 68 cm na 1 m jest więc tylko kwestią sklejenia dwóch arkuszy papieru.

Jak zostało to już zasugerowane, POLPAL rysuje diagramy na drukarce laserowej i 24-igłowej, ponadto za pomocą drukarki 9-igłowej można otrzymać diagramy „robocze” o mniejszej precyzji.

Istnieje również wersja programu obsługująca ploter. Na rysunku (Ryc. 1) przedstawiony jest przykład diagramu otrzymanego z wykorzystaniem różnych możliwości programu.

ANALIZY NUMERYCZNE

Nowe narzędzie, jakim jest komputer umożliwia nie tylko obliczanie procentów ale i wykonywanie o wiele bardziej skomplikowanych operacji matematycznych. Obecnie można efektywnie zastosować stare metody statystyczne oraz nowe metody analiz numerycznych opracowane po pojawieniu się elektronicznej techniki obliczeniowej. Nie każdy przyrodnik da się przekonać do obliczeń, których jedynym celem jest zatrudnienie komputera i wykazanie się numerycznymi wynikami. Metody statystyczne jednak dają pewien dodatkowy wgląd w tabele danych. Oczywiście trzeba rozumieć, na czym to inne spojrzenie polega. Przy czym nie chodzi tu o zrozumienie działania danej metody matematycznej, a jedynie o uchwycenie istoty wyniku jej działania. Można polecić tu metodę przykładów. Najlepiej zastosować metodę do prostych, dobrze znanych danych. Wtedy powinno być oczywiste co ujawniła metoda statystyczna. Trzeba podkreślić jedną z niewątpliwych za-

let metod numerycznych – ich obiektywność. Obliczenia komputera w najmniejszym stopniu nie zależą od intencji badacza. Wynik końcowy zależy jedynie od danych wejściowych i, czasami, od pewnych parametrów, które mogą być ustawiane stosownie do preferencji bardziej doświadczonego interpretatora metody komputerowej.

Program POLPAL zawiera cztery typowe metody numeryczne.

CONSLINK

Celem tej metody jest ułatwienie przeprowadzenia obiektywnego podziału profilu, tzw. zonacji [1]. Wynikiem działania programu jest dendrogram, którego gałęzie łączą próbki lub grupy wcześniej połączonych próbek na odpowiednich poziomach podobieństwa. Nazwa metody oznacza CONstrained Single LINK, ograniczone stratygraficznie łączenie próbek na podstawie podobieństwa pomiędzy najbardziej podobnymi pojedynczymi (single) próbkami. Do obliczeń wchodzi nie cała tabela a jedynie do 30 najliczniej reprezentowanych taksonów. Przykład zastosowania metody przedstawiony jest na rysunku.

ANALIZA SKŁADOWYCH GŁÓWNYCH (PRINCIPAL COMPONENTS ANALYSIS)

Jest to podstawowa, podręcznikowa metoda [7] pozwalająca wychwycić główne cechy zmienności składu taksonomicznego w profilu. Duża ilość taksonów (do 30) zamieniana jest na kilka *quasi*-taksonów zawierających w sobie najważniejszą informację. Pierwsza składowa główna (*quasi*-takson) niesie najwięcej informacji, kolejne mniej, jednak każda składowa główna zawiera informację nową, niezależną od wcześniejszych składowych.

Podstwowe fakty „dziejące” się w profilu są więc widoczne w pierwszej składowej głównej, bardziej subtelne są wychwycone przez drugą składową. Czasem jakiegoś sensu dopatrzeć można się również w trzeciej składowej, ostrożny interpretator zwykle jednak widzi tam już tylko szum informacyjny.

ANALIZA KORESPONDENCYJNA

To najbardziej złożona z oferowanych przez POLPAL analiz [4]. Jej wyróżniającą cechą jest połączenie próbek z taksonami (wierszy tabeli z kolumnami). Na jednym wykresie otrzymuje się tu profil, w postaci linii łamanej biegnącej od pierwszej do ostatniej próbki, oraz punkty reprezentujące taksony. Bliskość taksonu i próbki ma prostą interpretację, którą sprawdzić można na konwencjonalnym diagramie. Jest to metoda, która pozwala przedstawić tabelę w sposób bardzo skondensowany.

RAREFACTION ANALYSIS

Skomplikowana i pracochłonna dla komputera metoda, nie mająca nawet polskiej nazwy, ma bardzo prosty cel [2]. Ważną wielkością charakteryzującą poziom stratygraficzny jest bogactwo taksonów, ilość taksonów jaką zaobserwowano w próbce. Dość oczywiste jest, że taksony, których szczątków w próbce jest mało (ale są) będą znalezione dopiero po zliczeniu odpowiednio dużej całkowitej ilości szczątków. Tak więc próbka, w której zliczono 1500 ziarn pyłku będzie zawierała najprawdopodobniej więcej różnych taksonów niż identyczna próbka ale policzona tylko do 600 ziarn. Jeżeli próbki w profilu zawierają różne sumy szczątków a badacz interesuje bogactwo taksonów to trzeba wykonać właśnie rarefaction analysis aby wszystkie próbki sprowadzić do tej samej (niestety najmniejszej) sumy. Wtedy dopiero zmienność liczebności taksonów wzdłuż profilu będzie mogła być interpretowana. Tym bardziej, że metoda ta dostarcza też informacji o precyzji otrzymanej informacji.

PODZIĘKOWANIE

Autor dziękuje za stymulację rozwoju programu POLPAL i wkład pracy w jego budowę Prof. M. Ralskiej-Jasiewiczowej, Dr D. Nalepce i Mgr E. Madeyskiej.

LITERATURA

- [1] BIRKS H. J.B. 1986. Numerical zonation, comparison and correlation of Quaternary pollen – stratigraphical data: 743–774 W: B. Berglund (red.), *Handbook of Holocene Palaeoecology and Palaeohydrology*. Wiley & Sons, Chichester – New York – Brisbane – Toronto – Singapore, ss. 869.
- [2] BIRKS H. J. B. LINE J. M. 1992. The use of rarefaction analysis for estimating palynological richness from Quaternary pollen-analytical data. *The Holocene* 2(1): 1–10.
- [3] FAEGRI K., IVERSEN J. 1989. Textbook of Pollen Analysis, 4th Edition (by K. Faegri, P. E. Kaland and K. Krzywiński)... Wiley & Sons. Chichester – New York – Brisbane – Toronto – Singapore, ss. 328.
- [4] GREENACRE M. J. 1984. Theory and application of correspondence analysis. Academic Press, London and Orlando.
- [5] GRIMM E. C. 1990. *Tilia* and *Tilia-Graph*: PC spreadsheet and graphics software for pollen data. INQUA-Comission for the Study of the Holocene, Working Group on Data-Handling Methods, *Newsletter* 4.
- [6] MAKOHONIENKO M. WALANUS A. 1991. Analizy numeryczne wyników badań palinologicznych osadów Jeziora Lednickiego. W: K. TOBOLSKI (red.) *Wstęp do Paleokologii Lednickiego Parku Krajobrazowego*, Wyd.Nauk. UAM, Poznań: 71–79.
- [7] MORRISON D. F. 1990. Wielowymiarowa analiza statystyczna. PWN.
- [8] NALEPKA D. WALANUS A. 1994. Arytmetyka w diagramach pyłkowych, *Wiad. Bot.* 39(1/2).
- [9] RALSKA-JASIEWICZOWA M., WALANUS A. 1989. Projekt palinologicznej bazy danych. *Zesz. Nauk. Pol. Śląskiej, Ser. Mat.-Fiz., 61, Geochronometria* 6: 189–192.
- [10] RALSKA-JASIEWICZOWA M., WALANUS A. 1991. Polish palynological database (POLPAL) in course of building. INQUA-Comission for the Study of the Holocene, Working Group on Data-Handling Methods, *Newsletter* 5.
- [11] WALANUS A. 1989. Saving computer memory in storing tables of pollen counts. *Pollen et Spores*, 31(1–2): 161–164.
- [12] WALANUS A. 1994. Optimizing taxon codes in pollen counting, INQUA-Comission for the Study of the Holocene, Working Group on Data-Handling Methods, *Newsletter* 11: 6.