

Jacek Domostawski\*

## **Pakiet programów komputerowych dla identyfikacji parametrów wybranych modeli chronobiologicznych**

### **Streszczenie**

Prezentowany pakiet programów komputerowych jest przeznaczony do analizy zależności czasowych występujących w żywych organizmach. Działa on na danych liczbowych otrzymanych w wyniku pomiarów przeprowadzanych cyklicznie np. co 2–4 godz. w ciągu kilku dni, tak aby ogólna liczba punktów czasowych nie przekraczała 50. Sposób traktowania tych wyników polega na znalezieniu modelu, który w statystycznie istotny sposób będzie opisywał dane doświadczalne. W obecnej wersji pakietu zastosowano modele: liniowy, kwaziliniowy i dwa modele nieliniowe. Pakiet ma budowę modułową, która pozwala zarówno na zróżnicowanie przebiegu analizy zgodnie z potrzebami użytkownika, jak i na rozszerzenie jej zakresu. Może być stosowany dla analizy danych z szerokiego zakresu badań właściwych chronobiologii. Działa na mikrokomputerach typu IBM – PC.

### **Wprowadzenie**

Znaczna liczba doświadczeń chronobiologicznych wymaga przeprowadzenia analizy danych eksperymentalnych w oparciu o pewne modele matematyczne, które możliwie poprawnie opisują czasowe

---

\*Zakład Fizjologii Zwierząt. Pracownia Chronobiologii, Instytut Zoologii Uniwersytetu Jagiellońskiego

zależności występujące w badanych procesach. Analiza taka powinna pozwolić na stwierdzenie czy badany proces biologiczny jest procesem periodycznym, a jeżeli tak, to znaleźć oceny parametrów tego procesu takie, jak np. mediana, akrofaza, okres. Wymaga to szerokiego stosowania techniki komputerowej wspomaganej odpowiednim oprogramowaniem. Prezentowany pakiet programów stanowi narzędzie, przy pomocy którego chronobiolog może przeprowadzić analizę danych doświadczalnych. Pakiet obejmuje moduł edytora ułatwiający przygotowanie danych, moduł wstępnej analizy statystycznej oraz dwa moduły realizujące periodyczny model liniowy i kwaziliniowy (linearyzowany) badanego przebiegu oraz dwa okresowe modele nieliniowe. Sterowanie poszczególnymi modułami pakietu odbywa się przy pomocy wyboru określonych opcji, czyli tzw. menu. Upraszcza to komunikację pomiędzy komputerem a użytkownikiem i spełnia postulat oprogramowania przyjaznego (user friendly).

## **Przygotowanie danych**

Zasadniczym elementem – zwanym dalej punktem czasowym – jest wartość liczbowa pomiaru wielkości biologicznej (np. poziom enzymu w krwi lub stężenia elektrolitu) dokonanego w określonym czasie. Bardzo często w tym samym czasie dokonuje się kilku lub kilkunastu pomiarów (np. pobranie materiału od wielu osobników) tej samej wielkości. Dla pracy pakietu programowego niezbędny jest zbiór dostatecznej liczby punktów czasowych. Liczba punktów czasowych niezbędnych dla realizacji wybranego modelu jest związana z liczbą parametrów charakterystycznych. Przy pomocy modułu edytora rezultaty pomiarów wpisujemy do zbioru dyskowego. Edytor pozwala na utworzenie zbiorów dyskowych zawierających dwie postacie danych doświadczalnych:

1. Zbiór wszystkich danych dla każdego punktu czasowego, w takiej postaci, w jakiej zostały bezpośrednio zmierzone;
2. Zbiór średnich i odchyłeń standardowych, jeśli wyniki pomiarów zostały wcześniej opracowane statystycznie.

Oba sposoby zapisu odnoszą się do pomiarów dokonywanych nie tylko w równych odstępach czasu (np. co 2 godziny). Równomierną skalę czasu wprowadza się przez podanie wartości początkowej i odstępu pomiędzy kolejnymi pomiarami (tzw. krok czasowy pomiaru). W przypadku nierównomiernej skali czasu musimy wprowadzić każdą wartość czasową oddzielnie. Ze względu na poprawne obliczanie akrofazy, w przypadku badania rytmów okołodobowych, przyjęto konwencję, że czas pomiaru jest liczony w stosunku do północy pierwszej kolejnej doby, czyli czas pomiarów dokonywanych przed północą jest podawany jako liczby ujemne (zamiast czasu zegarowego – liczba godzin pozostająca do północy). Istnieją ograniczenia dotyczące liczby punktów czasowych. Ze względu na wymagania statystyczne nie może ich być mniej niż 6. Z drugiej strony ograniczone rozmiary pamięci programów nie pozwalają na opracowywanie tymi modułami więcej niż 50 punktów czasowych. Doświadczenie pokazuje, że ta ostatnia liczba jest w większości przypadków wystarczająca.

## **Wstępna analiza statystyczna**

W przypadku, gdy wprowadzone zostały wartości wszystkich dokonanych pomiarów (a nie tylko średnie i odchylenia standardowe) musimy przeprowadzić wstępną analizę statystyczną. Ze względu na istnienie bardzo dużej liczby znakomitych programów statystycznych moduł ten w prezentowanym pakiecie programowym został ograniczony do niezbędnego minimum. Obejmuje ono analizę wariancji (ANOVA), wyliczenie wartości średnich i odchyleń standardowych dla każdego punktu czasowego oraz przeprowadzenie testowania istotności różnic pomiędzy tymi średnimi (metodą każdy z każdym w oparciu o kryterium Studenta) [8]. Przeprowadzenie testu ANOVA ma podstawowe znaczenie dla dalszego działania. Jeżeli w jego wyniku uzyskamy informację, że czas nie różnicuje naszych pomiarów w odpowiednim stopniu (tzn. wariancja wewnątrzgrupowa jest większa lub równa wariancji międzygrupowej), to dalsza praca z tymi danymi traci sens z punktu widzenia chronobiologii. W przypadku pozytywnego

wyniku testu ANOVA zostaje automatycznie utworzony dodatkowy zbiór dyskowy (oprócz zbioru zapisanego przez użytkownika przy pomocy edytora danych), który zawiera wartości średnie i odchylenia standardowe dla każdego punktu czasowego. Format tego zbioru jest dostosowany do dalszego przetwarzania przez moduły realizujące wybrane modele matematyczne dla procesów okresowych (rytmicznych).

## Proste modele kosinorowe

Najprostszym modelem rytmu biologicznego był model kosinora zaproponowany przez Halberga [4]. Zakładał on, że cykliczny proces biologiczny może być opisany funkcją:

$$F(t) = M + A \cos\left(\frac{2\pi t}{T} + \Phi\right) \quad (1)$$

gdzie:  $F(t)$  – oznacza funkcję czasu

$M$  – przeciętną wartość badanego przebiegu – mediana

$A$  – maksymalną wartość w badanym przebiegu – amplituda

$T$  – okres badanego przebiegu – ustalony arbitralnie na 24 godziny

$\Phi$  – akrofazę – wystąpienie pierwszego maksimum badanego przebiegu w stosunku do północy.

Model ten charakteryzowały dwie cechy – z punktu widzenia chronobiologii opisywał on tzw. rytmy dobowe, matematycznie zaś stanowił równanie liniowe, które można rozwiązać przy pomocy metody regresji. Wraz z rozszerzeniem zakresu badań chronobiologii na systemy wykazujące rytmy okołodobowe, infra- lub ultradienne, oraz sezonowe, liniowy model kosinora okazał się niewystarczający. Naturalną drogą rozwoju tego modelu stało się uzmiennienie okresu (tzn. okres z wielkości stałej stawał się parametrem zmiennym). Matematycznie było to równoważne odejściu od liniowego modelu i przejściu do modelu nieliniowego. Rozpatrywanie modeli nieliniowych wiąże

się ze znaczną komplikacją obliczeń. Ostatecznie w tym wariancie dla opisu rytmu biologicznego przyjęto formułę:

$$F(t) = M + Bt + A \cos\left(\frac{2\pi t}{\tau} + \Phi\right) \quad (2)$$

w której wielkości  $F(t)$ ,  $M$ ,  $A$ ,  $\Phi$  – mają takie samo znaczenie jak poprzednio, zaś  $\tau$  – jest okresem badanego przebiegu wyliczonym z danych doświadczalnych. Liniowy czynnik  $Bt$  opisuje możliwy wpływ czynników wolnozmiennych – długookresowych na badany przebieg. Wartości parametrów modelu oceniano na podstawie danych doświadczalnych stosując metodę najmniejszych kwadratów (MNK), tj. minimalizując formę kwadratową:

$$Q^2 = \sum_{i/1}^N [Y_i - E_i(\theta)]^2 \frac{1}{\sigma_i^2} \quad (3)$$

przy czym:  $Q^2$  – oznacza minimalizowaną formę kwadratową

$N$  – liczbę punktów czasowych

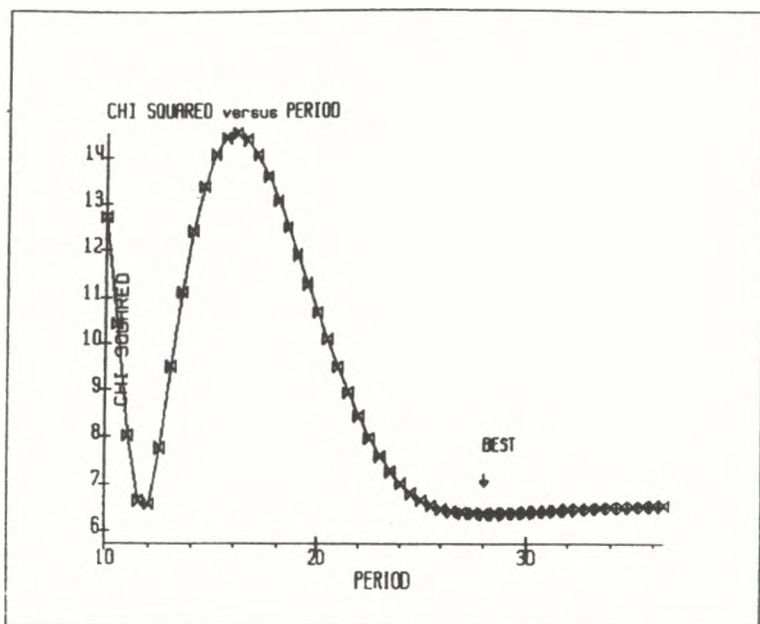
$Y_i$  – wartości pomiarów (średnie) w odpowiednich punktach czasowych

$E_i$  – wartości teoretyczne (wyliczone z modelu) w tych samych punktach czasowych

$\theta$  – zespół parametrów modelu

$\sigma_i^2$  – uśrednione wariancje pomiarów.

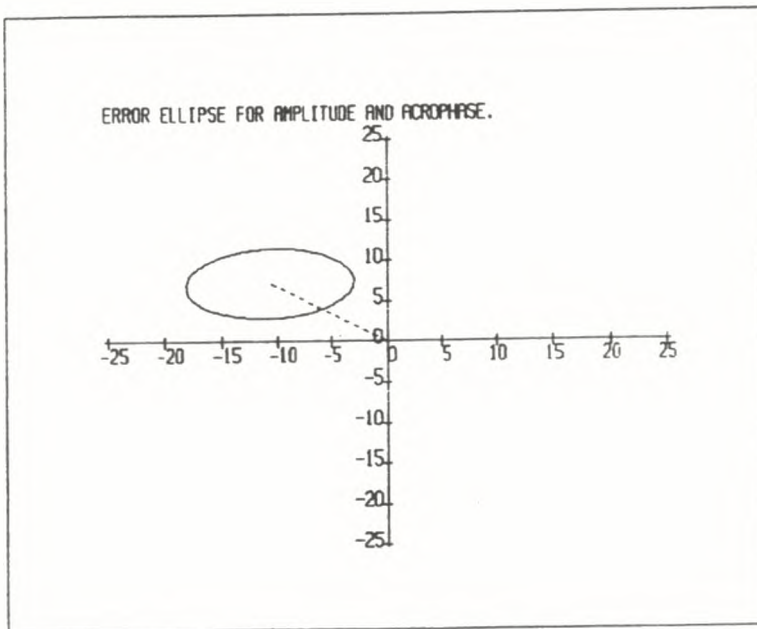
Ponieważ funkcja (2) jest nieliniowa tylko ze względu na jeden parametr ( $\tau$  – okres), rozwiązanie problemu można nieco uprościć stosując przeszukiwanie zadanego na wejściu zakresu zmienności okresu przy optymalnie dobranych pozostałych parametrach. Wynik tego działania zostaje przedstawiony w postaci wykresu (ryc. 1) wartości formy kwadratowej (3) jako funkcji parametru  $\tau$ . Jeżeli zostanie znalezione minimum, to jest ono traktowane jako punkt startowy udokładniania iteracyjnego. Otrzymany w ten sposób kwaziliniowy problem MNK jest rozwiązywany przy pomocy koncepcji SINGULAR VALUE DECOMPOSITION (SVD). Termin ten nie ma właściwego polskiego odpowiednika, a oznacza uogólnienie



Ryc. 1. Wykres funkcji  $\chi^2$  w zależności od parametru okres – znaleziono minimum dla wartości  $\tau = 28$  godz.

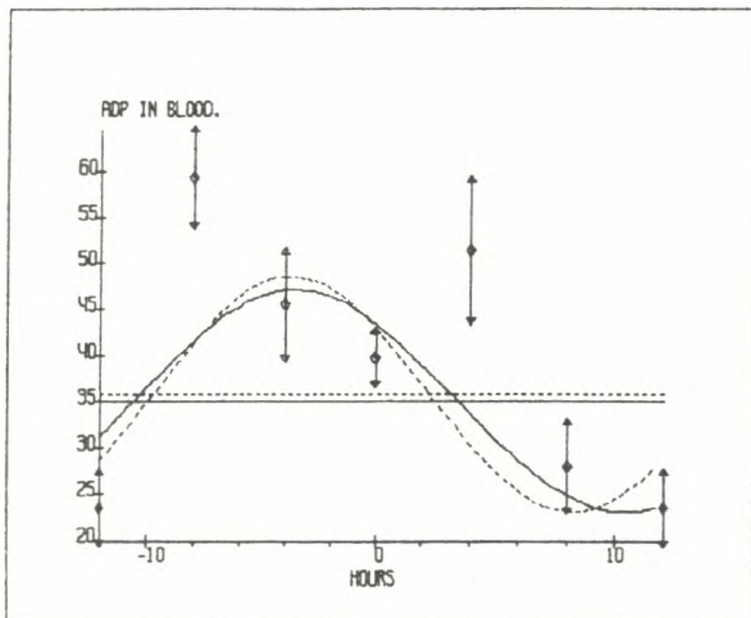
problemu własnego na macierze prostokątne, których rząd nie musi być koniecznie równy mniejszemu wymiarowi. – Szczegółowy opis koncepcji i zastosowań SVD można znaleźć, np. w pracy [3]. Zasadniczym motywem porzucenia tradycyjnego podejścia, związanego z rozwiązywaniem układu równań normalnych, była chęć zapewnienia kontroli nad numeryczną stabilnością wyników. Dalszym argumentem za stosowaniem SVD jest często kiepska jakość i niewielka liczebność danych pomiarowych otrzymanych w trakcie badania rytmów biologicznych.

W rezultacie działania modułu uzyskujemy (w przypadku gdy forma kwadratowa (3) osiąga minimum) wartości parametrów  $M$ ,  $A$ ,  $B$ ,  $\tau$ ,  $\Phi$  oraz oceny statystyczne ich błędów. Oprócz tego jest podawana ocena statystyczna testu  $\chi^2$  jako miara jakości dopasowania modelu do wielkości doświadczalnych. Dodatkowo zostaje obliczona tzw.



Ryc. 2. Elipsa błędów dla parametrów amplitudy (oś pionowa) i akrofazy (oś pozioma)

„elipsa błędów” dla parametrów amplituda ( $A$ ) i akrofaza ( $\Phi$ ) (ryc. 2). Powszechnie przyjmuje się, że parametry rytmu nie są istotne statystycznie w przypadku gdy „elipsa błędów” obejmuje punkt o współrzędnych, 0,0 na płaszczyźnie tych parametrów. Na zakończenie pracy tego modułu otrzymujemy rysunek danych, ich odchyłeń standardowych oraz dopasowanych modelowych krzywych (ryc. 3), przy czym linią przerywaną zaznaczono przebieg modelu z ustalonym (24-godzinnym) okresem.



Ryc. 3. Przykład dopasowania modelu 1 (krzywa ciągła) do punktów eksperymentalnych

Linia punktowana oznacza dopasowanie modelu z ustalonym 24 – godzinnym okresem (model Halberga). Jakość dopasowania jest zła

## Modele nieliniowe

Rozwój metodyki pomiarów w chronobiologii doprowadził do wniosku, że w wielu przypadkach proste modele typu funkcji (1) lub (2) nie mogą wystarczyć do zadowalającego opisu wyników pomiarów. Spowodowało to powstanie bardziej rozbudowanych w pełni nieliniowych matematycznych modeli rytmów biologicznych. Warto tutaj zauważyć, że z punktu widzenia metody najmniejszych kwadratów, doskonałą zgodność eksperymentu z teorią możemy uzyskać powiększając liczbę doбираalnych (swobodnych) parametrów modelu. Jednakże są przynajmniej dwa powody, aby zwiększanie liczby swobodnych parametrów modelu stosować bardzo ostrożnie:



1) ponieważ wartości parametrów są oceniane na podstawie danych eksperymentalnych, więc powiększanie liczby parametrów wiąże się ze zwiększeniem liczby zmierzonych punktów czasowych – czyli ze zwiększeniem objętości przeprowadzonego doświadczenia;

2) dla każdego parametru należy znać jego tzw. interpretację – czyli inaczej mówiąc, każdy parametr matematycznego modelu powinien być związany z biologiczną rzeczywistością.

Moduł programowy obecnie realizuje dwa nieliniowe modele rytmów biologicznych. Pierwszy z nich dotyczy procesów, które możemy rozpatrywać jako superpozycję dwu wzajemnie niezależnych oscylatorów. Model ten można sobie wyobrazić jako sytuację, gdy na wewnętrzny endogeny rytm jakiegoś czynnika biologicznego nakłada się rytm egzogeny otoczenia zewnętrznego w stosunku do tego czynnika. Mamy wtedy do czynienia z przypadkiem maskowania rytmu wewnętrznego przebiegiem zewnętrznym [7]. Matematycznie sytuacja ta daje się opisać funkcją:

$$F(t) = M + A_1 \cos\left(\frac{2\pi t}{\tau_1} + \Phi_1\right) + A_2 \cos\left(\frac{2\pi t}{\tau_2} + \Phi_2\right) \quad (4)$$

gdzie:  $M$  – mediana,

$A_1$  i  $A_2$ ;  $\tau_1$  i  $\tau_2$ ;  $\Phi_1$  i  $\Phi_2$  – amplitudy, okresy i akrofazy części endogennej i egzogennej odpowiednio.

Drugi z modeli może być stosowany w przypadku, gdy stan dynamiczny układu biologicznego można przedstawić jako oscylator podlegający z jednej strony pewnym zaburzeniom (tłumieniu) zakłócającym rytm, a z drugiej strony działaniu zewnętrznej periodycznej siły, która próbuje narzucić układowi własny rytm (w postaci takiego czynnika wymuszającego można rozważać np. cykliczne zmiany światła i ciemności). Z matematycznego punktu widzenia jest to model oscylatora harmonicznego z drganiem gasnącymi działającego w zewnętrznym polu periodycznej siły wymuszającej. Proces tego typu można opisać funkcją:

$$F(t) = M + A_1 e^{-ht} \sin(\omega_1 t + \Phi_1) + \frac{H \sin(\omega_2 t + \Phi_2)}{\sqrt{(\omega_1^2 - \omega_2^2)^2 + 4\omega_2^2 h^2}} \quad (5)$$

przy czym:  $M$  przedstawia – medianę,

$A_1$  i  $\Phi_1$  – odpowiednio amplitudę i akrofazę składowej własnej rytmu;

$\omega_1$  i  $\omega_2$  – częstotliwości kołowe części własnej rytmu i siły wymuszającej;

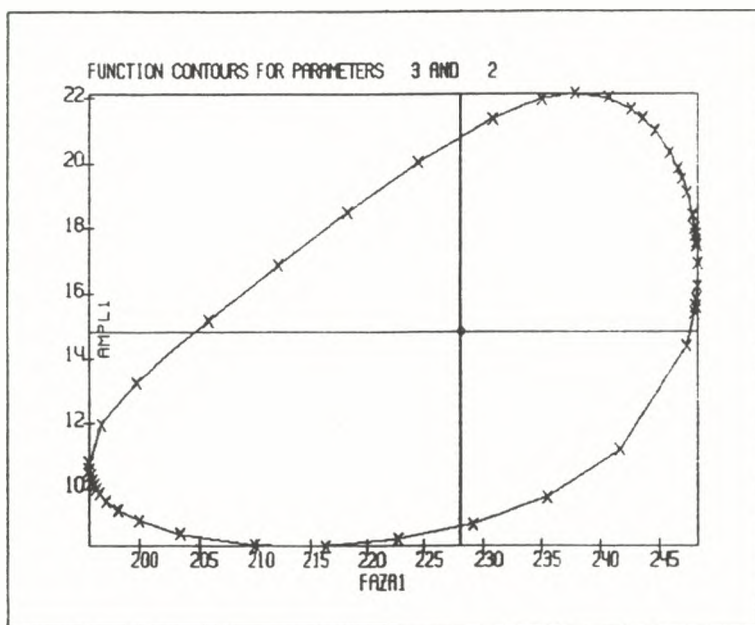
$H$  i  $\Phi_2$  – amplitudę i akrofazę zewnętrznej siły wymuszającej;

natomiast  $h$  – stałą tłumienia.

Wiele innych bardziej skomplikowanych matematycznych modeli rytmów biologicznych zostało omówionych w pracach [5, 6]. Programową realizację powyższych modeli zapewnia rozbudowany zespół wyspecjalizowanych procedur komputerowych spełniających warunki metody najmniejszych kwadratów, która, tak jak i poprzednio, sprowadza się do minimalizacji formy kwadratowej postaci (3). Forma tak zapisana wyraża matematyczną postać metody MNK i nie jest związana z konkretnym modelem matematycznym. Na wstępnym etapie minimalizacji jest stosowana metoda SIMPLEKSÓW – dająca szybko przybliżone oszacowanie obszaru minimum udokładnianego następnie przy pomocy metody gradientów. Ta ostatnia metoda daje (w przypadku gdy minimum zostało znalezione) końcowe wartości parametrów modelu wraz z oszacowaniem ich błędów. Prócz procedur minimalizacyjnych moduł realizujący modele nieliniowe został wyposażony w wiele dodatkowych możliwości, z których warto tutaj wspomnieć dwie.

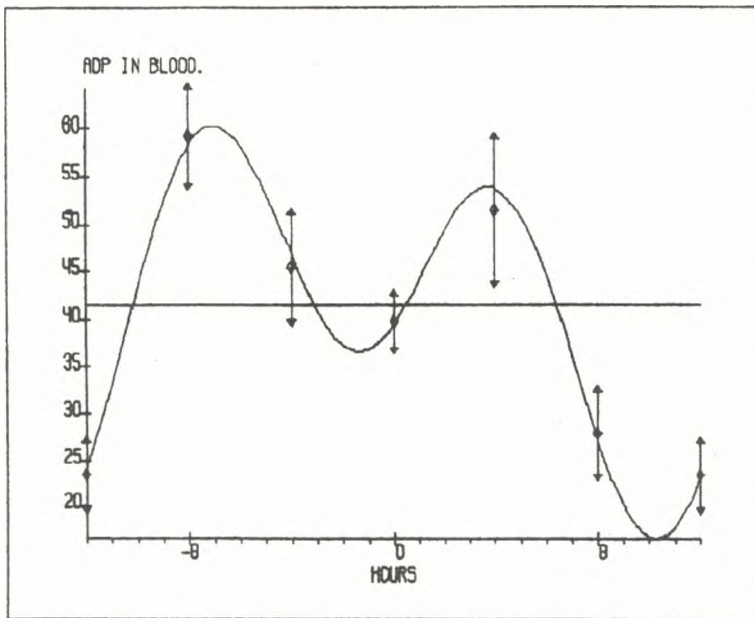
Jedna jest związana z możliwością otrzymywania konturów dla dowolnej pary parametrów (np. dla pary amplituda – akrofaza jest to uogólniony analog „elipsy błędów”). Polega ona na konstrukcji konturu jako śladu przecięcia hiperelipsoidy, utworzonej wokół punktu minimum w niezredukowanej przestrzeni parametrów i płaszczyzny wyznaczonej przez dwa wybrane parametry (np. płaszczyzny amplituda – akrofaza, ryc. 4). Statystyczna interpretacja konturów jest analogiczna do interpretacji „elipsy błędów”, o której wspomniano poprzednio.

Druga bierze pod uwagę fakt, że zazwyczaj dla funkcji wieloparametrowych, istnieje wiele tzw. minimów lokalnych. Zdecydowana większość programów stosowanych dla minimalizacji funkcji po znalezieniu pierwszego minimum kończy pracę i nie ma praktycznie



Ryc. 4. Przykład konturu dla współrzędnych faza i amplituda.  
Widać asymetrię konturu

pewności, czy znalezione minimum jest najlepszym – globalnym w danym obszarze zmienności parametrów. W proponowanym rozwiązaniu istnieje procedura, która w przypadku znalezienia jednego minimum pozwala w systematyczny sposób poszukiwać następnych, a w przypadku natrafienia na lepsze (mniejsza wartość liczbową formy kwadratowej (3)) automatycznie powraca do procedur minimalizacyjnych w celu udokładnienia nowo znalezionych parametrów. Bardziej szczegółowy opis idei matematycznych i stosowanych algorytmów minimalizacji funkcji można znaleźć w pracach [1, 2]. Podobnie jak dla modułu realizującego model kwaziliniowy – pracę modułu nieliniowego kończy utworzenie wykresu zawierającego dane doświadczalne wraz z błędami (odchyleniami standardowymi) i zaznaczonym przebiegiem dopasowanej krzywej modelowej (ryc. 5). Osobny zbiór dyskowy



**Ryc. 5.** Przykład dopasowania modelu 2 do danych doświadczalnych identycznych jak w przykładzie na ryc. 3.

Jakość dopasowania nie budzi zastrzeżeń

zawiera (tab. 1) zebrane wartości parametrów modelu wraz z błędami, a także wielkość oceny testu  $\chi^2$  oraz odpowiadający jej poziom istotności będące miarą zgodności rozpatrywanego modelu z doświadczeniem.

Tabela 1

Wyniki dla modelu nieliniowego			Wyniki dla modelu kwazilinowego	
Parametr	Wartość	Błąd	Wartość	Błąd
Mediana	41,4630	± 5,1218	35,004	± 3,702
Ampl 1	14,8360	± 7,9927	12,050	± 4,617
Faza 1	228,0769	± 25,5040	45,125	± 34,389
Okres 1	12,0707	± 1,1190	28,086	± 8,454
Ampl 24	10,6133	± 6,6905	–	–
Faza 24	42,1190	± 39,7755	–	–
Okres 2	24,000	± ,000	–	–

Czynnik CHI kwadr. =	,039665	6,356
Poziom istotności =	,000001	0,6155

## Podsumowanie

Z przytoczonego przykładu (ryc. 1 – 5) widać, że dla wielu typów danych chronobiologicznych proste modele kosinorowe nie wystarczają dla ich interpretacji. Pomimo znalezienia minimum formy kwadratowej (3) względem okresu  $\tau$  istotności wyników amplitudy i akrofazy ocenionych na podstawie „elipsy błędów” (ryc. 2) całość dopasowania oceniona na podstawie kryterium  $\chi^2$  jest bardzo słaba ( $\chi^2 = 6.356$  co odpowiada poziomowi istotności  $p = 0.616$ ) – porównaj ryc. 3. Wynika to prawdopodobnie z faktu, że jak sugeruje przebieg punktów doświadczalnych, badana wielkość posiada dwa nierównoważne (o różnych wysokościach) maksima, a taki przebieg nie może być opisany pojedynczym kosinorem. Zastosowanie do tych samych danych procedury nieliniowej – model opisany funkcją (4) z ustaloną na 24 godziny wartością okresu  $\tau_2$  daje wysoko istotne dopasowanie ( $\chi^2 = 0.04$ , co odpowiada poziomowi istotności lepszej niż  $p \ll 0.001$  – ryc. 5). Przykład ten pokazuje, jak ważne jest, aby zespół prowadzący

badania w dziedzinie chronobiologii dysponował możliwie dużą liczbą wariantów opracowania danych.

Przedstawiony w niniejszej pracy pakiet programów komputerowych wydaje się spełniać postulat szerokich możliwości, tym bardziej, że stosunkowo prosto można go uzupełnić innymi potrzebnymi w danym przypadku modelami. Pakiet programów został napisany w językach programowania C i FORTRAN i współpracuje z mikrokomputerami typu IBM PC z dyskiem twardym, drukarką i dowolną popularną kartą graficzną.

## Literatura

- Eadie W.T., et. al., 1971, *Statistical Methods in Experimental Physics*. North-Holland Publ.Co. Amsterdam
- Fletcher R. and Powell M.J.D., 1963, *A rapidly converging descent method for minimization* Comput. J., 6: 163
- Forsythe G.E. et.al., 1977, *Computer Methods for Mathematical Computations* Englewood Cliffs. New York
- Halberg F., 1974, *Glossary of Selected Chronobiologic Term. in Chronobiology*. Georg Thieme Publishers Stuttgart
- Moore-Ede M.C., Sulzman F.M. and Fuller C.A., 1982, *The Clocks that Time Us*. Harvard University Press, Cambridge, MA
- Wever R.A., 1984, *Toward a Mathematical Model of Circadian Rhythmicity*. Raven Press, New York
- Minors D.S., Waterhouse J.M. 1992, *Investigating the Endogenous Component of Human Circadian Rhythms: A Review of Some Simple Alternatives to Constant Routines*. Chronobiology International, Vol. 9, 55
- Byrkit D.R., 1987, *Statistics Today. A Comprehensive Introduction*. The Benjamin/Cummings Publishing Company In. Menlo Park, California

## **Computer Programs Package for Some Chronobiology Models**

### **Summary**

All-purpose experimental data processing package for Chronobiology is described. It allows to analyze the data obtained from the short measurement series (up to 50 time points). The way of treating the kind of data consist in finding a model that describes the experimental data in a statistically significant manner. In that version a linear model, quasilinear model, and two non-linear models are applied. The package has modular structure which enables both the differentiation of the analysis procedure according to the needs of the user and the extension of the scope of analysis. The package can be used in chronobiology – oriented analysis such as, for example, measuring the levels of enzymes in animal organs, blood pressure, and body temperature in human and animals, affected by various environmental and pharmacological factors. The software utilize an IBM – PC type microcomputers.