Zeszyty Naukowe Akademii Morskiej w Gdyni
Scientific Journal of Gdynia Maritime University

Nr 105/2018, 124–135
ISSN 2451-2486 (online)
ISSN 1644-1818 (printed)
DOI: 10.26408/05.11

Złożony/submitted: 01.06.2018
Zaakceptowany/accepted: 08.06.2018
Opublikowany/published: 29.09.2018

ALGORYTM I APLIKACJA W PROGRAMIE EXCEL DLA KROKOWEJ APROKSYMACJI DANYCH DROGĄ ROZWIĄZANIA UKŁADU RÓWNAŃ METODĄ GAUSSA

ALGORITHM AND EXCEL APPLICATION FOR DATA STEPWISE APPROXIMATION BY SOLUTION OF THE SYSTEM OF EQUATIONS USING GAUSS METHOD

Stanisław Polanowski
Uniwersytet Morski w Gdyni, Morska 81-87, 81–225 Gdynia,
Wydział Mechaniczny, Katedra Siłowni Okrętowych, e-mail: s.polanowski@wm.am.gdynia.pl,
ORCID 0000-0003-0501-6266

Streszczenie: W artykule zaprezentowano algorytm oraz aplikację w programie Excel krokowej aproksymacji danych metodą najmniejszych kwadratów dla modelu wielomianowego, liniowego względem współczynników. Pod pojęciem wielomianu aproksymującego rozumiany jest wielomian uogólniony, którego jednomiany są dowolnymi liniowo niezależnymi funkcjami. Współczynniki wielomianu aproksymującego są wyznaczane metodą Gaussa. Specjalnie utworzona tabela informacyjna umożliwia optymalny dobór jednomianów oraz budowanie modelu krok po kroku, drogą włączania do modelu jednomianów najbardziej zmniejszających sumę kwadratów odchyleń. W tej tabeli wskazano sumę kwadratów odchyleń, odchylenia skrajne oraz wartości odchylenia standardowego na każdym kroku aproksymacji. Stanowi to podstawę do podjęcia decyzji o zakończeniu aproksymacji, a także umożliwia wyłonienie punktów o nadmiernych odchyleniach.

Słowa kluczowe: metoda najmniejszych kwadratów, aproksymacja krokowa, model wielomianowy, algorytm Gaussa, aplikacja EXCELA.

Abstract: The paper presents the algorithm and the application in Excel of the stepwise least squares data approximation for a polynomial model linear with respect to coefficients. The term approximating polynomial is understood as a generalized polynomial which monomials are any linearly independent functions. The coefficients of the approximating polynomial are determined by the Gaussian method. Specially created information table enables an optimal selection of monomials and building a model step by step, by incorporating the monomials with most decreasing sums of squared deviations. In this table, the sum of squared deviations, extreme deviations and standard deviation values are indicated at each approximation step. This is the basis for making the decision to complete the approximation, as well as the selection of points with excessive deviations.

Keywords: least squares method, stepwise approximation, polynomial model, Gaussian algorithm, EXCEL application.

1. WSTĘP

Zagadnienie aproksymacji danych pomiarowych metodą najmniejszych kwadratów dla wielomianowego modelu liniowego względem współczynników jest zagadnieniem teoretycznie opisanym. Istnieją liczne narzędzia dla jego rozwiązania z zastosowaniem takich programów jak Statistica, Matematica, Matlab.

Powszechnie dostępnym programem umożliwiającym średniokwadratową aproksymację danych, jest program EXCEL [EXCEL 2016]. Wbudowane w nim programy aproksymacji średniokwadratowej mogą być wykorzystywane jedynie do obróbki jednoowymiarowych zbiorów danych, zasadniczo do celów redakcyjnych i w bardzo ograniczonym zakresie do celów inżynierskich. Brak w nich np. możliwości dokonywania ocen statystycznych.

Ze względu na narzędzia dostępne w programie EXCEL problem aproksymacji średniokwadratowej można rozwiązać dwoma sposobami: drogą rozwiązania układu równań liniowych, np. metodą Gaussa, lub na bazie rachunku macierzowego, którego niezbędne do tego celu funkcje są dostępne w programie EXCEL.

W artykule zaprezentowano przykład rozwiązania problemu aproksymacji pierwszą metodą. Przykład aproksymacji metodą najmniejszych kwadratów z wykorzystaniem funkcji macierzowych programu EXCEL zostanie omówiony w innym artykule. W obydwu przypadkach przyjęto zasadę nie tylko aproksymacji danym modelem, lecz także zasadę krokowego budowania modelu jako sumy członów wnoszących w kolejnym kroku aproksymacji największy wkład w minimalizację sumy kwadratów odchyleń [*Stepwise regression*; Efroymson, Ralston i Wilf 1960; Enslein, Ralston i Wilf 1977].

Istotą problemu jest przybliżenie przyjętym modelem, metodą najmniejszych kwadratów, zbioru danych, np. pomiarowych y_p $(p = 1 \div P)$, określonych na wielowymiarowym zbiorze argumentów (zmiennych niezależnych) $x(x_1, x_2, \ldots, x_q)$ [Hartmann, Lezki i Schäfer 1974].

Zagadnienie jest liniowe jeżeli modelem aproksymacji jest wielomian uogólniony y_K funkcjibazowych f_k (jednomianów), liniowy względem współczynników a_k, co można zapisać następująco:

$$y_K = \sum_{k=1}^{K} a_k f_k \,, \tag{1}$$

gdzie $f_k = f_k(x)$ – liniowo niezależne funkcje bazowe argumentów zbioru danych (zmiennych niezależnych) $x(x_1, x_2, \ldots, x_q)$.

Funkcje bazowe f_k mogą być znane lub zadane z góry albo też poszukiwane w procesie aproksymacji, co ma miejsce w rozpatrywanym przypadku.

Współczynniki a_k wyznacza się z warunku najmniejszych kwadratów. Wymaga się, by suma kwadratów S odchyleń wartości aproksymowanych od wartości uzyskanych z aproksymacji przyjętym modelem (1) osiągnęła minimum:

$$S = \sum_{p=1}^{P}(y_p - y_{Kp})^2 = \text{MIN}, \tag{2}$$

co implikuje następujący warunek:

$$\frac{\partial S}{\partial a_k} = -2\sum_{p=1}^{P}\left(y_p - y_{Kp}\right) \cdot \frac{\partial y_{Kp}}{\partial a_k} = 0. \tag{3}$$

Otrzymuje się układ równań liniowych, którego rozwiązaniem są wartości współczynników a_k. Układ ten można zapisać następująco:

$$\left\{\begin{matrix} a_1 \cdot f_{11} + \cdots + a_k \cdot f_{1k} + \cdots + a_K \cdot f_{1K} = f_1 y \\ \vdots \\ a_1 \cdot f_{1k} + \cdots + a_k \cdot f_{kk} + \cdots + a_K \cdot f_{kK} = f_k y \\ \vdots \\ a_1 \cdot f_{1K} + \cdots + a_k \cdot f_{kK} + \cdots + a_K \cdot f_{KK} = f_K y \end{matrix}\right\}, \tag{4}$$

gdzie elementy lewej strony układu równań (4) wyznacza się ze wzoru:

$$f_{kl} = \sum_{p=1}^{P} f_{kp} \cdot f_{lp} \text{ dla } k, l = 1 \div K \tag{5}$$

Wyrażenie dla wyznaczania wielkości $f_k y$ jest następujące:

$$f_k y = \sum_{p=1}^{P} f_{kp} \cdot y_p. \tag{6}$$

Współczynniki $a_k$wielomianu aproksymującego (1) są wynikiem rozwiązania układu równań (4).

Jeżeli model aproksymacji jest znany, współczynniki aproksymacji można łatwo wyznaczyć, wykorzystując w tym celu, np. funkcje macierzowe programu EXCEL. Na ogół jednak model nie jest znany lub znany częściowo i poszukuje się jego przybliżenia, np. wielomianem potęgowym.

Celem pracy jest nie tylko aproksymacja średniokwadratowa zbioru danych, znanym lub szukanym modelem, lecz także wyłonienie modelu o uzasadnionej statystycznie strukturze.

W artykule zaprezentowano pierwszą, z dwóch opracowanych, aplikację rozwiązania zagadnienia aproksymacji średniokwadratowej z wykorzystaniem programu EXCEL.

Istotą zaprezentowanego przykładu jest krokowe zwiększanie bazy jednomianów równania aproksymującego.

W przypadku prezentowanej aplikacji można zbudować model zawierający do 8 jednomianów, co wymaga rozwiązania układów równań liniowych zawierających do 8 niewiadomych. Ograniczenie programu do rozwiązywania układu maksymalnie 8 równań liniowych zostało wymuszone ograniczeniami prezentacji. Prezentowany program łatwo rozwinąć na większą liczbę funkcji bazowych, np 20 i więcej.

2. WPROWADZANIE DANYCH DO ARKUSZA I ICH WSTĘPNE PRZEKSZTAŁCENIA

Na potrzeby artykułu wykorzystano opublikowane dane pomiarowe z badań morskich charakterystyk napędowych mocy napędu N w funkcji prędkości obrotowej wału n i prędkości statku v, co pokazuje tabela 1 [Giernalczyk i Górski 2011].

Obszar danych podzielono na obszar wczytywania i przygotowania do obróbki. Umożliwia to wstępne przygotowanie danych, np. skalowanie (mnożenie przez współczynniki) zmiennych niezależnych lub też ich normalizację przez sprowadzenie do przedziałów wartości [–1: 1]. Niekiedy może być pożądane przesunięcie osi y.

Tabela 1. Wprowadzanie i wstępne przygotowanie danych

Table 1. *Data input and preliminary data preparation*

	A	B	C	D	E	F	G	H	I
14		Obszar wprowadzania danych				Obszar przygotowania danych			
15		Wiersz początku danych			22	Zmiana położenia osi y			
16		Wiersz końca danych			38	b_1	b_2	b_3	
17						0	0	0	
18		Liczba danych, P			17	Mnożniki skalujące danych			
19		Wprowadzone dane				c_{x1}	c_{x2}	c_{c3}	c_y
20		obr/min	knt		kW	1	1	1	1
21	p	n	v		N	x_1	x_2	x_3	y
22	1	75	9,5		542	75	9,5		542
23	2	95	12		1144	95	12		1144
24	3	115	14		2124	115	14		2124
25	4	125	14		2911	125	14		2911
26	5	135	16		3946	135	16		3946
27	6	48	3,5		270	48	3,5		270
28	7	72	6		694	72	6		694
29	8	95,6	9,2		1535	95,6	9,2		1535
30	9	116	12		2701	116	12		2701
31	10	128	12		3700	128	12		3700
32	11	128	12		3844	128	12		3844
33	12	49	6,5		158	49	6,5		158
34	13	72	9,1		462	72	9,1		462
35	14	93,1	11,5		972	93,1	11,5		972
36	15	115,4	14		2131	115,4	14		2131
37	16	125,9	15		2628	125,9	15		2628
38	17	137,2	16		3570	137,2	16		3570

W rozpatrywanym przykładzie (tab. 1) przyjęto: $c_i = 1; b_i = 0; x_1 = n;\ \ x_2 = v; y = N$. W komórce E15 jest deklarowany numer wiersza początku danych, a w E16 – końca danych. W E18 wyświetla się liczba aproksymowanych danych.

W niektórych przypadkach mogą być pożądane przekształcenia funkcyjne zmiennych x i y, np. logarytmowanie, itd.

Doświadczenie pokazuje, że wstępne rozpoznanie celowości zastosowania powyższych operacji znacznie ułatwia proces aproksymacji i zapobiega przyjęciu nieadekwatnego modelu. Jest to szczególnie ważne, gdy obrabiane są zbiory danych o małej liczebności, co najczęściej ma miejsce w działalności inżynierskiej i naukowej w dziedzinach technicznych.

3. WYZNACZANIE WARTOŚCI ELEMENTÓW TABLICY UKŁADU RÓWNAŃ

W celu wyznaczenia sum (5) iloczynów wartości jednomianów oraz sum (6) iloczynów wartości jednomianów i wartości wyjściowych utworzono kolumny wartości jednomianów dla poszczególnych punktów p (tab. 2).

Tabela 2. Kolumny wartości jednomianów

Table 2. *Columns of monomial values*

	W	X	Y	Z	AA	AB	AC	AD	AE
18	w_{1k}	3	1	1	0	1	0	0	0
19	w_{2k}	0	2	1	0	0	2	0	3
20	w_{3k}	0	0	0	0	0	0	0	0
21	p	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
22	1	421875	6768,75	712,5	1	75	90,25	1	857,375
23	2	857375	13680	1140	1	95	144	1	1728
24	3	1520875	22540	1610	1	115	196	1	2744

W kolumnach X÷AE, począwszy od wiersza 22, są deklarowane funkcje jednomianowe. Wymagana jest ich wzajemna niezależność liniowa i określoność na zbiorze argumentów.

W prezentowanej aplikacji zadeklarowano jednomiany postaci:

$$f_k = x_1^{w_{1k}} \cdot x_2^{w_{2k}}. \tag{7}$$

Przykładowo: |X22|=$F22^X$18*$G22^X$19. Wpis ten należy przeciągnąć w dół arkusza co najmniej do wiersza końca danych. Analogiczne wpisy powinny być wykonane w kolumnach Y–AE, począwszy od wiersza 22.

Dla dodatnio określonych, niezerowych wartości x_1 i x_2 można deklarować dowolne rzeczywiste wartości w_{1k} i w_{2k} (tab. 2).

W przypadku wielomianów potęgowych o potęgach całkowitych dodatnich do wyznaczenia odpowiednich sum można wykorzystać bezpośrednio funkcję programu EXCEL “SUMA.ILOCZYN“. Ograniczyłoby to jednakże możliwości tworzenia jednomianów z potęgami ujemnymi lub niecałkowitymi.

W tabeli 3 zamieszczono wyznaczone wartości elementów tablicy układu równań dla prezentowanego przykładu aproksymacji.

Tabela 3. Układ równań liniowych

Table 3. *System of linear equation*

	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP
20		X	Y	Z	AA	AB	AC	AD	AE	AF
21	T 1	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	yf_k
22	f_1	3,90603E+13	4,87638E+11	35089435927	21739287,4	2588626147	3943767836	21739287,4	5,52E+10	60215755428
23	f_2	4,87638E+11	6305803069	447331701,8	273302,479	32444635,04	51155144,06	273302,479	7,25E+08	736517225,4
24	f_3	35089435927	447331701,8	32444635,04	21059,67	2423017,857	3687700,197	21059,67	51155144	53406575
25	f_4	21739287,4	273302,479	21059,67	17	1725,2	2381,45	17	31155,01	33332
26	f_5	2588626147	32444635,04	2423017,857	1725,2	188820,78	273302,479	1725,2	3687700	3976882,8
27	f_6	3943767836	51155144,06	3687700,197	2381,45	273302,479	422776,6757	2381,45	5892056	5940558,12
28	f_7	21739287,4	273302,479	21059,67	17	1725,2	2381,45	17	31155,01	33332
29	f_8	55235805669	724851461,5	51155144,06	31155,009	3687700,197	5892056,304	31155,009	83823891	82690026,63

Poniżej podano przykłady wyrażeń obliczeniowych dla sum (5) i (6):

|AH22|=SUMA.ILOCZYNÓW(ADR.POŚR($AH20&$E15):ADR.POŚR($AH20&$E$16);ADR.POŚR(AH20&$E$15):ADR.POŚR(AH20&$E$16)),

|AI22|=SUMA.ILOCZYNÓW(ADR.POŚR($AH20&$E15):ADR.POŚR($AH20&$E$16);ADR.POŚR(AI20&$E$15):ADR.POŚR(AI20&$E$16)),

|AP22|=SUMA.ILOCZYNÓW(ADR.POŚR($AH20&$E15):ADR.POŚR($AH20&$E$16);ADR.POŚR("E"&$E$15):ADR.POŚR("E"&$E$16)).

Należy zauważyć, że macierz AH22:AO29 (tab. 3) jest macierzą kwadratową symetryczną.Wartości wyznaczone nad przekątną tablicy wystarczy odpowiednio przepisać do dolnej połowy tablicy.

Symbole kolumn zawarte w komórkach AH20:AP20 służą w tym przypadku ułatwieniu w tworzeniu programu.

4. WYZNACZANIE WARTOŚCI WSPÓŁCZYNNIKÓW WIELOMIANU APROKSYMUJĄCEGO

Po wyznaczeniu wartości elementów tabeli 3 można przystąpić do wyznaczenia wartości współczynników wielomianu aproksymującego (1).

W tym celu w pracy zastosowano algorytm Gaussa.

Tablica T1 (tab. 3) jest redukowana do tablicy T2 (tab. 4), w wyniku czego ulega eliminacji niewiadoma a_1. W następnym kroku jest eliminowana niewiadoma a_2 i powstaje tablica T3 (tab. 4). W ten sposób powstaje osiem kolejnych tablic T1–T8 o stopniowo malejącym wymiarze.

Tabela 4. Rozwiązanie układu równań metodą Gaussa

***Table 4.** The solution of the system of Gaussian equations*

	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP
30		T 2	f_2	f_3	f_4	f_5	f_6	f_7	f_8	
31		f_2	-0,000447087	-1,9003E-05	-3,90445E-09	-2,6167E-07	-3,93774E-06	-3,90445E-09	-7,2E-05	3,12339E-05
32		f_3	-0,000264088	-2,6286E-05	-4,36135E-08	-2,78E-06	-4,12805E-06	-4,36135E-08	-4,4E-05	1,9599E-05
33		f_4	-8,75813E-05	-7,0397E-05	-2,25437E-07	-1,3086E-05	-8,57965E-06	-2,25437E-07	-1,9E-05	8,35104E-06
34		f_5	-4,92925E-05	-3,7683E-05	-1,09896E-07	-6,6698E-06	-4,61194E-06	-1,09896E-07	-1E-05	5,32106E-06
35		f_6	-0,000486893	-3,6729E-05	-4,72937E-08	-3,0272E-06	-6,23495E-06	-4,72937E-08	-8E-05	3,52966E-05
36		f_7	-8,75813E-05	-7,0397E-05	-2,25437E-07	-1,3086E-05	-8,57965E-06	-2,25437E-07	-1,9E-05	8,35104E-06
37		f_8	-0,000638613	-2,7782E-05	-7,47872E-09	-4,9022E-07	-5,7047E-06	-7,47872E-09	-0,0001	4,45753E-05
38			T 3	f_3	f_4	f_5	f_6	f_7	f_8	
39			f_3	-0,0570292	-0,000156415	-0,00994149	-0,006823806	-0,000156415	-0,0038	0,004353164
40			f_4	-0,7612811	-0,002565293	-0,14882998	-0,089154505	-0,002565293	-0,05517	0,025490852

Wyrażenia obliczeniowe dla przykładowych komórek są następujące:
|AI31|=JEŻELI($AH23=0;AI23;AI$22/$AH$22-AI23/$AH23),
|AJ31|=JEŻELI($AH23=0;AJ23;AJ$22/$AH$22-AJ23/$AH23),
|AP31|=JEŻELI($AH23=0;AP23;AP$22/$AH$22-AP23/$AH23).

Zastosowanie funkcji JEŻELI zapewnia istnienie rozwiązania dla macierzy każdej postaci, pomijając przypadki $f_{kk} = 0$, co nie powinno mieć miejsca.

W tabeli 5 zamieszczono wyniki obliczeń współczynników aproksymacji a_k dla kolejnych modeli aproksymacji (1).

Tabela 5. Wyznaczanie współczynników równań aproksymujących

***Table 5.** Calculation of coefficients of approximation equations*

	AG	AH	AI	AJ	AK	AL	AM	AN	AO
2		y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
3	a_1	0,001541612	0,002413773	0,0024418	0,002480176	0,002469332	0,002647813	#DZIEL/0!	#DZIEL/0!
4	a_2		-0,069860923	-0,066616457	-0,059246498	-0,05674851	-0,058698857	#DZIEL/0!	#DZIEL/0!
5	a_3			-0,07633219	-0,263765405	-0,32789756	-1,169670851	#DZIEL/0!	#DZIEL/0!
6	a_4				68,33870299	41,82284963	-65,19898086	#DZIEL/0!	#DZIEL/0!
7	a_5					0,785080496	4,451770067	#DZIEL/0!	#DZIEL/0!
8	a_6						4,146229801	#DZIEL/0!	#DZIEL/0!
9	a_7							#DZIEL/0!	#DZIEL/0!
10	a_8								2,308231

Automatycznie wyznaczane są wartości a_k dla wszystkich K, w tym przypadku dla $K = 1 - 8$.

Zasadę obliczeń współczyników, wynikającą z metody Gaussa (tab. 5), ilustrują następujące przykładowe wyrażenia:
dla $K = 1$: a_1=|AH3|=AP22/AH22,
dla $K = 2$: a_1|AI3|=(AP22-AI4*AI22)/AH22, a_2 =|AI4|=AP31/AI31.

Należy zwrócić uwagę na przypadek 0^0. W takim przypadku EXCEL sygnalizuje status błędu. Status błędów dzielenia przez 0, obserwowany w tabeli 5 (kolumny AN i AO), jest następstwem zadeklarowania w kolumnie AD (tab. 2) jednomianu liniowo zależnego od jednomianu w kolumnie AA. Jeżeli dana kolumna znajduje się poza już przyjętym modelem, sygnalizowane statusy błędu nie mają żadnego wpływu na wyniki aproksymacji. Nie są one maskowane, ponieważ spełniają funkcję informującą o zadeklarowaniu jednomianu, liniowo zależnego od jednego z wcześniej zadeklarowanych jednomianów.

5. KROKOWY PROCES TWORZENIA WIELOMIANU APROKSYMUJĄCEGO

Jeżeli postać modelu aproksymującego jest znana z góry, to wystarczy w tabeli 6 zadać znane wartości wykładników w_{1k} i w_{2k}, właściwe dla kolejno przyjmowanych jednomianów modelu.

Tabela 6. Tabela informacyjna krokowego procesu aproksymacji

Table 6. *Information table of the stepwise approximation process*

	W	X	Y	Z	AA	AB	AC	AD	AE
11		AR	AS	AT	AU	AV	AW	AX	AY
12	Dy_{max}	611,0	421,6	405,3	392,8	392,2	389,3	#DZIEL/0!	#DZIEL/0!
13	Dy_{min}	-448,5	-210,1	-229,0	-243,8	-243,9	-246,0	#DZIEL/0!	#DZIEL/0!
14	Sw	16	15	14	13	12	11	10	9
15	St	293,4	144,6	148,9	153,5	159,7	166,7	#DZIEL/0!	#DZIEL/0!
16	$\|Dy/St\|_{max}$	2,08	2,92	2,72	2,56	2,46	2,34	#DZIEL/0!	#DZIEL/0!
17	S	1,38E+06	3,13E+05	3,10E+05	3,06E+05	3,06E+05	3,06E+05	#DZIEL/0!	#DZIEL/0!
18	w_{1k}	3	1	1	0	1	0	0	0
19	w_{2k}	0	2	1	0	0	2	0	3
20	w_{3k}	0	0	0	0	0	0	0	0
21	p	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
22	1	421875	6768,75	712,5	1	75	90,25	1	857,375
23	2	857375	13680	1140	1	95	144	1	1728
24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

W tabeli 6, w wierszach 12–17 ujęto wartości parametrów charakteryzujących model w aspekcie statystycznym, umożliwiających podjęcie decyzji o końcowej postaci modelu aproksymacji oraz stanowiących podstawę do oceny błędów.

W tabeli 6, po włączeniu do modelu aproksymacji kolejnego jednomianu f_k, są wyliczane i prezentowane wartości natępujących parametrów: Dy_{max}, Dy_{min} – maksymalne i minimalne wartości odchyleń aproksymacji, S – suma kwadratów odchyleń, Sw – liczba stopni swobody, St – standardowe odchylenie, $|Dy/St|_{\max}$ – maksymalna wartość modułu odniesienia. Adresy kolumn w wierszu 11 (tab. 6) mają charakter pomocniczy dla tworzenia programu.

Nawet jeżeli model jest znany, to racjonalniej jest zastosować krokowe włączanie członów, kierując się wkładem dołączonego członu w zmniejszenie sumy kwadratów odchyleń S.

W rozpatrywanym przykładzie przyjęto, że model nie jest znany i zastosowano metodę krokowego poszukiwania zbiorów wartości w_{1k} i w_{2k}, kierując się zasadą najmniejszego S.

Wyniki tego postępowania obrazuje tabela 7.

Tabela 7. Przykład krokowego wyłaniania jednomianów aproksymujących

Table 7. *An example of stepwise emergence of approximating monomials*

Krok 1					f_1		
w_{11}	1	0	1	2	3	4	3
w_{21}	0	1	1	0	0	0	1
S	1,04E+07	1,38E+07	6,30E+06	3,15E+06	1,38E+06	2,04E+06	4,75E+06
Krok 2					f_2		
w_{12}	0	1	0	1	2	3	2
w_{22}	0	0	1	1	1	1	2
S	1,37E+06	1,37E+06	1,19E+06	8,65E+05	3,13E+05	9,63E+05	6,48E+05

Wyłoniono jednomiany f_1 i f_2.

Okazało się, że otrzymano identyczny model jak w przypadku wyłonienia go sposobem analitycznym [Giernalczyk i Górski 2011]. Stosując krokową aproksymację, wyłoniono dodatkowe, równorzędnie statystycznie modele, mające także pewne uzasadnienie teoretyczne (fizyczne) [Polanowski i Pawletko 2015; Charchalis i Polanowski 2016].

Na każdym kroku dokonywana jest analiza wpływu wartości w_{1k}, w_{2k} na wartość S (tab. 6). W rozpatrywanym przypadku powiększenie modelu o kolejny jednomian nie jest uzasadnione. Dla $K = 3$(krok 3) zmniejszenie wartości S jest pomijalnie małe przy zauważalnym wzroście wartości St (tab. 6), spowodowanej malejącą liczbą stopni swobody Sw.

Program umożliwia bieżącą obserwację i porównanie maksymalnych odchyleń aproksymacji Dy_{max}, Dy_{min} z wartościami odchylenia standardowego (tab. 6).

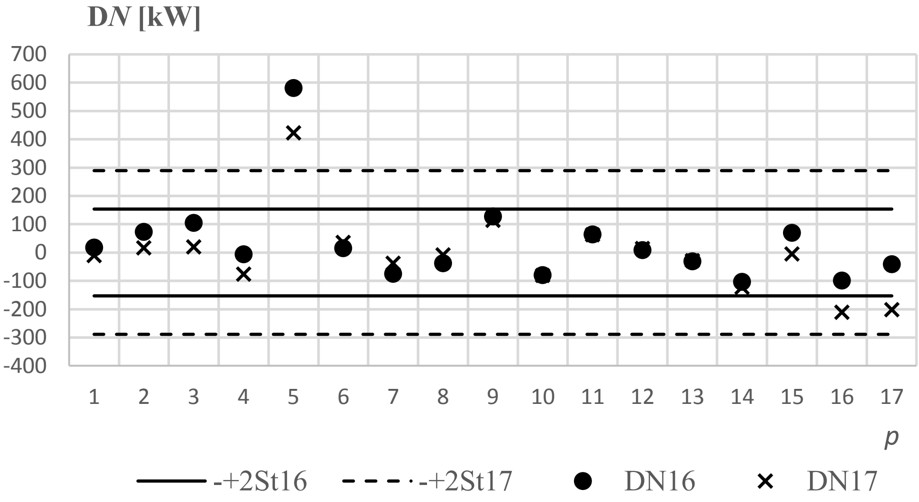
W powyższym celu program generuje tabelę 8 odchyleń Dy w punktach p.

Tabela 8. Kolumny odchyleń aproksymacji

Table 8. *Approximate deviation columns*

	AQ	AR	AS	AT	AU	AV	AW	AX	AY
20		Odchylenia w kolejnych krokach aproksymacji							
21	p	Dy_1	Dy_2	Dy_3	Dy_4	Dy_5	Dy_6	Dy_7	Dy_8
22	1	-108,368	-3,439205657	17,14546007	16,29443095	17,29025522	12,78110208	#DZIEL/0!	#DZIEL/0!
23	2	-177,74	30,18911815	48,75901644	60,40464582	60,57400077	55,47990243	#DZIEL/0!	#DZIEL/0!
24	3	-220,599	27,61875642	34,69592701	43,70126659	43,37439977	43,83357091	#DZIEL/0!	#DZIEL/0!

W rozpatrywanym przypadku analiza odchyleń dla wyłonionego modelu dwuelementowego ujawnia nadmierne odchylenia w przypadku punktu pomiarowego nr 5 (tab. 1).



Rys. 1. Odchylenia aproksymacji dla danych z punktem 5 i bez punktu 5

Fig. 1. *Approach deviations for data with point 5 and without point 5*

Odchylenie DN17 dla punktu $p = 5$ (rys. 1) istotnie przekracza wartość dwóch odchyleń standardowych -+2St17. Po usunięciu punktu 5 z aproksymowanego zbioru wartość odchylenia standardowego St16 zmalała prawie dwukrotnie w porównaniu z St17, a odchylenie DN16 dla punku 5 przekroczyło czterokrotną wartość odchylenia standardowego St16. Punkt 5 wyłączono z aproksymacji, co wymagało reorganizacji kolumny p (tab. 1) oraz zadania nowej wartości końcowego wiersza danych w komórce E16.

6. ZAOKRĄGLANIE WARTOŚCI WSPÓŁCZYNNIKÓW

Jednym z aspektów obróbki danych jest ocena błędów i istotności współczynników modelu oraz liczby miejsc znaczących współczynników.Dla oceny statystycznej współczynników modelu konieczne jest wyznaczenie macierzy kowariancji.

Program EXCEL umożliwia wyznaczenie macierzy kowariancji, ponieważ dostępne są funkcje macierzy transponowanej i odwrotnej. Wykorzystanie tej możliwości zostanie zaprezentowane w następnym artykule, w którym rozwiązanie zagadnienia aproksymacji oparto na wykorzystaniu funkcji macierzowych dostępnych w programie EXCEL.

Tabela 9 prezentuje podprogram zaokrąglania współczynników równania aproksymującego, w którym wykorzystano funkcję ZAOKR.

Tabela 9. Zaokrąglanie wartości współczynników równania aproksymującego

Table 9. *Rounding the values of the approximation equation factors*

	BA	BB	BC	BD	BE	BF
9	Zadanie kolumny (stopnia) wielomianu >			AI		Liczba
10	K	Kolumna	a_k		a_{kz}	miejsc
11	1	AH	a_1	0,0024	0,0024138	7
12	2	AI	a_2	-0,0699	-0,069861	6
13	3	AJ	a_3	0,0000	0,0000000	
14	4	AK	a_4	0,0000	0,0000000	
15	5	AL.	a_5	0,0000	0,0000000	
16	6	AM	a_6	0,0000	0,0000000	
17	7	AN	a_7	0,0000	0,0000000	
18	8	AO	a_8	0,0000	0,0000000	
19	Największa bezwzględna wartość funkcji w przedziale aproksymacji					3780,1
20	Maksymalny bezwzględny błąd zaokrąglenia					0,068
21	p	y_a	y_z	Dy_z		
22	1	545,4392	545,4502313	-0,01103		
23	2	1113,811	1113,833295	-0,02241		
24	3	2096,381	2096,421135	-0,03989		

W komórce BD9 należy zadeklarować wielomian o rozpatrywanym wymiarze *Kz* listy BB11–BB18.

W komórkach BF11 i BF18 zadajemy liczbę miejsc znaczących dla współczynników $a_1 - a_K$. Podejmuje się decyzję, obserwując wartość maksymalną błędu zaokrąglenia w komórce BF20. W rozpatrywanym przykładzie przyjęto, że błąd zaokrąglenia nie może być większy od 0,1 kW.

7. PODSUMOWANIE

Przedstawiony w artykule algorytm aproksymacji danych metodą najmniejszych kwadratów umożliwia krokowe wyłanianie modelu zbudowanego z sumy jednomianów, minimalizujących sumę kwadratów odchyleń po każdym kroku aproksymacji.

Sygnalizacja sumy kwadratów odchyleń, odchyleń skrajnych oraz odchylenia standardowego dla każdego kroku aproksymacji, umożliwia wyłonienie optymalnego modelu aproksymacji z punktu widzenia statystycznego. Jest to szczególnie pomocne w przypadku mało licznych zbiorów danych.

Aplikacja przedstawionego algorytmu aproksymacji w programie EXCEL daje możliwość jego wykorzystania do obróbki danych zarówno w pracach inżynierskich, jak i naukowych. Program EXCEL jest powszechnie dostępny.

Aplikacja aproksymacji średniokwadratowej, wbudowana w programie EXCEL, nie stwarza powyższych możliwości. Umożliwia jedynie aproksymację zbiorów jednowymiarowych z ograniczoną możliwością statystycznej oceny zasadności przyjęcia modelu aproksymacji.

LITERATURA

Charchalis, A., Polanowski, S., 2016, *Modeling of Torque Characteristics and its Derivative for Ship Propultion System with Fixed Pitch Propellers*, Journal of KONES, vol. 23, no. 3.

Efroymson, M.A., Ralston, A., Wilf, H.S. (eds.), 1960, *Multiple Regression Analysis. Mathematical Methods for Digital Computers*, John Wiley, New York.

Enslein, K., Ralston, A., Wilf, H.S., 1977, *Statistical Methods for Digital Computers,* John Wiley & Sons, New York.

EXCEL 2016.

Giernalczyk, M., Górski, Z., 2016, *Siłownie okrętowe*, cz. I, Wydawnictwo Akademii Morskiej w Gdyni, Gdynia.

Hartmann, K., Lezki, E., Schäfer, W., 1974*, Statistische Versuchsplanung und-ausvertung in der Stofftwirtschaft*, VEB Leipzig.

Polanowski, S., Pawletko, R., 2015, *Modelowanie i wyznaczanie charakterystyk mocy układu napędowego statku wypornościowego ze śrubą o stałym skoku*, Zeszyty Naukowe Akademii Morskiej w Gdyni, nr 91, s. 122–131.

Źródła internetowe

Stepwise regression, http: en.wikipedia.org.