

METHOD OF ANALYSIS OF MEASUREMENT RESULT COHESION WITH THE THEORY OF MATHEMATICAL RECORD USING AND CONCENTRATIONS ANALYSIS

*Katarzyna Topolska, *Wojciech Walkowiak, **Mariusz Topolski

Politechnika Wroclawska

**Wydział Mechaniczny*

Instytut Konstrukcji i Eksploatacji Maszyn

***Wydział Elektroniki*

Katedra Systemów i Sieci Komputerowych

Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland

e-mail: katarzyna.topolska@pwr.wroc.pl, wojciech.walkowiak@pwr.wroc.pl

mariusz.topolski@pwr.wroc.pl.pl

Abstract

This study presents method of selection the most coherent, repeatable in time, results of researches. Firstly, a possible of application of concentration analysis for detection of variable or dependent variable, which results are different from other variables, is presented. Because some results of the same variable in time periods could influence on increase of results standard error, for such points detection, measure using mathematic filing system theory, Spearman's correlation ordinal ranks and weight of information contradiction, in sense of Dempster-Shafer theory, are proposed.

The obtained measure cause decrease of results standard errors by selection and removal of measure points which influence on validity scales decrease.

Example-results of the measurement, the analysis of concentrations for variable Y1, Y2, Y3 and Y4, results of the normalization for each result of measurements, average values of measurement for all received results BZ and for selected MODELS, results of measurement marked curves of average data for all BZ measurements and measurement selected of MODEL least standard-errors are presented in the paper.

Keywords: *modelling, durability, Dempster – Shafer, statistica*

METODA ANALIZY SPÓJNOŚCI WYNIKÓW POMIARÓW Z WYKORZYSTANIEM TEORII EWIDENCJI MATEMATYCZNEJ I ANALIZY SKUPIEŃ

Streszczenie

W artykule została przedstawiona metoda doboru najbardziej spójnych wyników pomiarów powtarzanych w czasie. W pierwszej części zaprezentowano możliwość zastosowania analizy skupień do wykrycia zmiennej bądź zmiennych zależnych, których wyniki różnią się od pozostałych zmiennych. Ponieważ niektóre wyniki pomiarów tej samej zmiennej w odstępach czasu mogą wpływać na zwiększanie błędu standardowego wyników, więc w celu wykrycia takich punktów zaproponowano miarę wykorzystującą teorię ewidencji matematycznej, korelację rang porządkowych Spearmana i wagę sprzeczności informacji w sensie teorii Dempstera-Shafera. Uzyskana miara pozwoliła na obniżenie błędów standardowych pomiarów, przez wyselekcjonowanie i usunięcie punktów pomiarowych, które wpływają na zmniejszenie rzetelności skal.

W szczególności w artykule przedstawiono przykładowe wyniki pomiaru, analiza skupień dla zmiennych Y1, Y2, Y3 i Y4, wyniki normalizacji dla poszczególnych wyników pomiarów, uśrednione wartości pomiarów dla wszystkich otrzymanych wyników BZ i dla wyselekcjonowanych MODELI, wyniki pomiarów wraz z naniesionymi krzywymi uśrednienia danych dla wszystkich pomiarów BZ i pomiarów wyselekcjonowanych o najmniejszych błędach standardowych MODELI.

Słowa kluczowe: *modelowanie, trwałość, Dempster- Shafer, statystyka*

1. Wprowadzenie

Badając różne obiekty, np. elementy silnika, aby uwiarygodnić wyniki pomiaru wykonuje się badania na większej liczbie tych samych elementów w odstępie czasu. Oczywiście jest, że te zmienne tak mierzone są od siebie zależne. Po wykonaniu takich eksperymentów pojawia się istotny problem odwzorowania wszystkich pomiarów w jeden spójny i wiarygodny przebieg zmienności. Bardzo często błędnie uśrednia się wyniki takich pomiarów, uzasadniając to powtarzalnymi warunkami pomiaru. Niestety w rzeczywistości różne czynniki zewnętrzne, których nie bierzemy pod uwagę, wpływają na poprawność wyników. Należy więc zastosować metody pozwalające na wykrycie pewnych punktów pomiarów, istotnie statystycznie różniących się od skupień poprawnych wyników. Błędy mogą pojawiać się w obrębie pojedynczych pomiarów albo całych mierzonych zmiennych. W dalszej części pracy zostanie zaprezentowana metoda doboru wyników badań do wyznaczania charakterystyk zmienności.

Teoria Dempstera - Shafera, zwana jest inaczej teorią ewidencji matematycznej, czy też teorią funkcji przekonania. Wyznaczane są prawdopodobieństwa z jakimi dane hipotezy można udowodnić. Teoria umożliwia rozróżnienie wiedzy od niewiedzy i ma zastosowanie w przypadkach niepełnej informacji, składania ewidencji i aktualizacji przekonań. W teorii Dempstera – Shafera zdaniom przypisuje się wielkości $Bel(A)$, która nazywana jest stopniem przekonania. Teoria wykorzystuje matematyczne prawdopodobieństwa do subiektywnej oceny ekspertów i daje możliwość przypisywania przekonań zarówno dla pojedynczych przekonań jak i grup odpowiedzi, a także nie uzupełnia się w teorii Dempstera – Shafera niepełnej specyfikacji.

Przez funkcję masy w sensie teorii DS rozumie się funkcję $m:2^{\{\theta\}} \rightarrow [0,1]$ spełniającą warunki:

$$\sum_{A \in 2^{\theta}} m(A) = 1, \quad (1)$$

$$m(\phi) = 0, \quad (2)$$

$$\forall_{A \in 2^{\theta}} m(A) \geq 0. \quad (3)$$

Dowodzono [5], że dla każdej funkcji przekonania **Bel** istnieje dokładnie jedna funkcja masy **m** taka, że $Bel(A) = \sum_{B \in A} m(B) = 1$, natomiast dla zbiorów większej liczności funkcję $m(A)$ można traktować jako wyrównanie "ignorancji" podzbiorów danego zbioru.

Przez funkcję przekonania w sensie teorii DS rozumie się taką funkcję $Bel:2^{\{\theta\}} \rightarrow [0,1]$, że

$$Bel(A) = \sum_{B \in A} m(B) = 1, \quad (4)$$

gdzie $m(B)$ jest funkcją masy w sensie teorii DS.

Funkcja przekonania mierzy wiarygodność poszlak na rzecz A .

θ - dyskretny niepusty skończony zbiór, przy czym dla pewnego naturalnego n

$$\theta = \theta_1 \times \theta_2 \times \theta_3 \times \dots \times \theta_n. \quad (5)$$

Zmienna A - przyjmuje wartości ze zbioru θ

Rozpatrując dwa rozkłady m_1 i m_2 , można dokonać ich połączenia, otrzymując nowy rozkład bazowy m według reguły

$$m(C) = \frac{\sum_{A \cap B = C} m_1(A) \cdot m_2(B)}{1 - \sum_{A \cap B = \phi} m_1(A) \cdot m_2(B)}. \quad (6)$$

W pracy wykorzystano teorię ewidencji matematycznej gdyż wyniki pomiaru są często nieprecyzyjne i niepewne oraz brak jest czasami całkowitej specyfikacji modelu. Szersze informacje odnośnie teorii ewidencji matematycznej można odnaleźć w pracach [1,2,3,4,5,6].

2. Metoda selekcjonowania zmiennych i pojedynczych punktów pomiarowych

W celu zobrazowania metody przytoczony zostanie przykład, na podstawie którego konsekwentnie będą prezentowane poszczególne etapy wyznaczania charakterystyki.

Przyjmijmy dla uproszczenia, że na pewnym elemencie mechanicznym wykonano dla pewnej zadanej wartości $X=\{1,2,\dots,10\}$ pomiar wartości $Y(X)$. W celu uzyskania spójnych wyników pomiar powtórzono na czterech różnych tej samej klasy elementach.

Wyniki eksperymentu zestawiono w tabeli 1.

Tab. 1. Przykładowe wyniki pomiaru
Tab. 1. Example-results of the measurement

lp	X	Y1	Y2	Y3	Y4
1	1	1	<u>1,24</u>	0,84	<u>0,61</u>
2	2	1,57	1,36	1,73	<u>1</u>
3	3	2,23	<u>2,56</u>	2,06	<u>1,6</u>
4	4	2,88	2,6	3,14	<u>2</u>
5	5	3,35	3,77	3,51	<u>2,74</u>
6	6	4	4,38	3,82	<u>3,04</u>
7	7	5,16	5,04	5,39	<u>3,7</u>
8	8	6,51	6,92	6,33	<u>5,07</u>
9	9	8,15	7,97	8,63	<u>4,89</u>
10	10	10,4	10,74	11,01	<u>11,97</u>

W tabeli 1 wytłuszczono pewne wyniki, co zostanie wyjaśnione w dalszej części rozważań.

Krok 1. Wykonanie analizy skupień

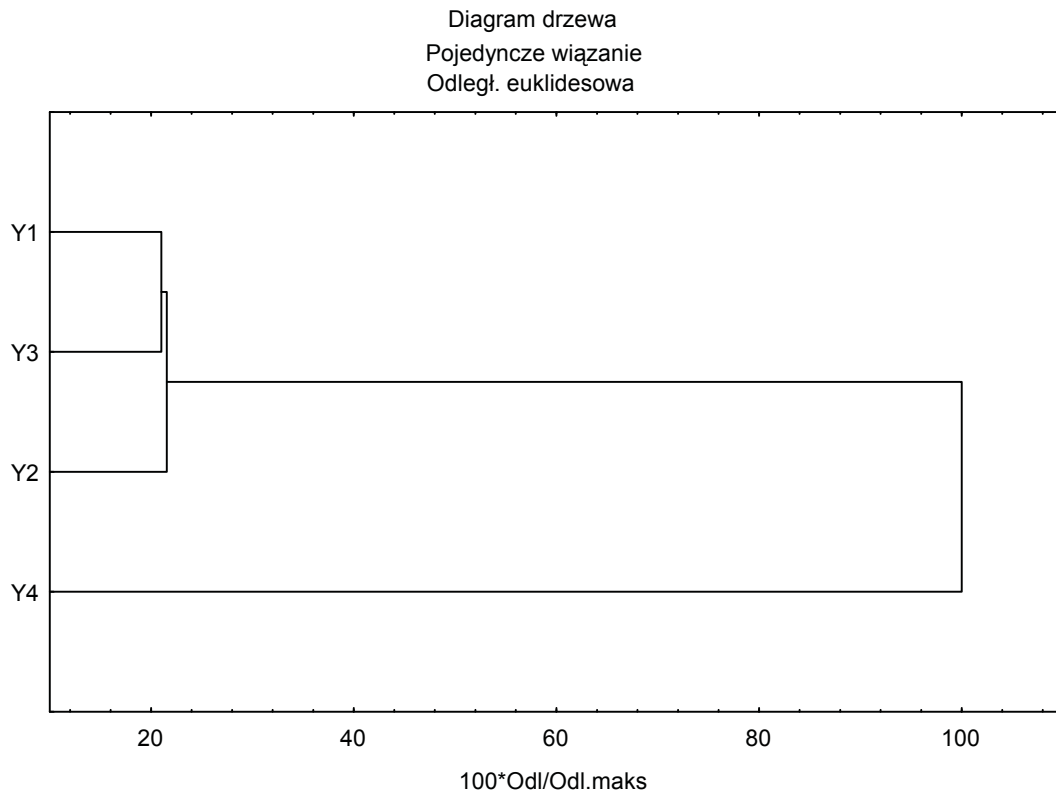
W kroku tym należy dokonać eksploracyjnej analizy danych, której celem jest ułożenie zmiennych w grupy w taki sposób, aby stopień ich powiązania był jak największy, a ze zmiennymi pozostałych grup jak najmniejszy. Analiza taka wykrywa struktury z danych bez wyjaśniania dlaczego one występują.

Krok 2. Selekcja pojedynczych pomiarów, które istotnie wnoszą błędy do całości badań.

Ponieważ zmienne $Y=\{1,2,3,4\}$ są zależne od siebie, więc w pierwszej kolejności dokonuje się normalizacji ich wartości, tak aby kolumny wyników spełniały warunek ortogonalności:

$$\forall k \in K, \quad m(Y_k^{(n)}) = \frac{n_k^{(n)}}{\sum_{i=1}^N n_k^{(i)}}, \quad (7)$$

gdzie: k - oznacza k -tą zmienną ze wszystkich K zmiennych. W analizowanym wypadku $k=\{1,2,3,4\}$,
 $n=\{1,2,\dots,N\}$ oznacza n -ty pomiar spośród N pomiarów dla k -tej zmiennej.



Rys. 1. Analiza skupień dla zmiennych Y1, Y2, Y3 i Y4
Fig.1. The analysis of concentrations for variable Y1, Y2, Y3 and Y4

Z rysunku 1 można wywnioskować, że zmienne Y1 i Y3 oraz Y2 są ze sobą bardzo powiązane, ale zmienna Y4 osłabia nasze kryterium tego, na ile jest ona istotna w badaniach. Ponieważ zmienna Y4 zdecydowanie tworzy własne skupienie, więc można przypuszczać, że należy wyniki Y4 odrzucić. Jednakże odrzucenie wszystkich wyników może powodować stratę cennych punktów pomiarów. W następnym punkcie zostanie przedstawiony tok postępowania, za pomocą którego wyodrębnione zostaną pomiary wprowadzające istotne błędy do charakterystyk pomiarów.

Po wykonaniu normalizacji przedstawionej we wzorze 7, dokonuje się normalizacji wierszy (każdego pomiaru) wyników dla wszystkich zmiennych:

$$\forall n \in N \quad m(Y_k^{(n)}) = \frac{m(Y_k^{(n)})}{\sum_{k=1}^K m(Y_k^{(n)})} \quad (8)$$

Wykonując działania zgodnie ze wzorami 7 i 8 otrzymuje się macierz, której wiersze spełniają warunek ortogonalności i stanowią elementy ogniskowe dla funkcji przekonania w sensie teorii Dempstera-Shafera:

$$\begin{aligned} a) \quad & \forall n \in N \quad \sum m(Y_k^{(n)}) > 0 \\ b) \quad & \forall n \in N \quad \sum m(Y_k^{(n)}) = 1 \end{aligned} \quad (9)$$

Wyniki takiej normalizacji zestawiono w tabeli 2.

Tab.2. Wyniki normalizacji dla poszczególnych wyników pomiarów
 Tab. 2. Results of the normalization for each result of measurements

lp	X	Y1	Y2	Y3	Y4
1	1	0,26	0,32	0,22	0,20
2	2	0,27	0,23	0,29	0,21
3	3	0,26	0,29	0,23	0,23
4	4	0,26	0,23	0,28	0,23
5	5	0,24	0,27	0,25	0,25
6	6	0,25	0,27	0,24	0,24
7	7	0,26	0,25	0,26	0,23
8	8	0,25	0,26	0,24	0,24
9	9	0,27	0,26	0,28	0,20
10	10	0,22	0,23	0,23	0,32

Niech

$$Y_k^{(n)} = \begin{cases} 1 & m(Y_k^{(n)}) \\ 0 & 1 - m(Y_k^{(n)}) \end{cases} \quad (10)$$

oraz

$$\alpha = 1 - \sum_{Y_k^{(n)} \cap Y_{k+1}^{(n)} = \emptyset} m(Y_k^{(n)}) \cdot m(Y_{k+1}^{(n)}) \quad (11)$$

Można określić miary sprzeczności informacji, które zaprezentowano we wzorach 12 i 13:

$$a) \gamma_1 = \frac{1 - 0.5R}{\frac{1}{K} \sum_{k=1}^K m(Y_k^{(n)})} + \frac{1 - \sum_{Y_k^{(n)} \cap Y_{k+1}^{(n)} = \emptyset} m(Y_k^{(n)}) \cdot m(Y_{k+1}^{(n)})}{\frac{2}{K} \left[\sum_{k=1}^K m(Y_k^{(n)}) \right]^2} \text{ dla} \quad (12)$$

$$0.5R + \frac{1 - \sum_{Y_k^{(n)} \cap Y_{k+1}^{(n)} = \emptyset} m(Y_k^{(n)}) \cdot m(Y_{k+1}^{(n)})}{\frac{2}{K} \left[\sum_{k=1}^K m(Y_k^{(n)}) \right]^2} < 1,$$

$$b) \gamma_2 = \frac{1 - \sum_{Y_k^{(n)} \cap Y_{k+1}^{(n)} = \emptyset} m(Y_k^{(n)}) \cdot m(Y_{k+1}^{(n)})}{\frac{2}{K} \left[\sum_{k=1}^K m(Y_k^{(n)}) \right]^2} - \frac{R}{\frac{1}{K} \sum_{k=1}^K m(Y_k^{(n)})} \text{ dla} \quad (13)$$

$$0.5R + \frac{1 - \sum_{Y_k^{(n)} \cap Y_{k+1}^{(n)} = \emptyset} m(Y_k^{(n)}) \cdot m(Y_{k+1}^{(n)})}{\frac{2}{K} \left[\sum_{k=1}^K m(Y_k^{(n)}) \right]^2} > 1.$$

W powyższych analizach wykorzystano tylko funkcje alokacji prawdopodobieństwa i nie obliczano funkcji przekonania, gdyż dla zadania pojedynczych zdarzeń mamy do czynienia z bayesowską funkcją przekonania, która jest tożsama z funkcją alokacji prawdopodobieństwa:

$$\text{Bel}(Y_k^{(n)}) = \sum_{Y_k^{(n)} \in Y_k^{(n)}} m(Y_k^{(n)}) . \quad (14)$$

We wzorach 12 i 13 R jest korelacją Spermiana między jedną zmienną zależną, a pozostałymi zmiennymi.

Jeżeli kolejny pomiar n (wiersze danych) spełnia warunek $\gamma_2, \gamma_2 < 0.1$ wówczas pomiar taki można uznać za spójny, niesprzeczny z pozostałymi wynikami pomiarów dla określonej wartości X . W wypadku, gdy niespełniony jest powyższy warunek pomiar jest usuwany.

Po wykorzystaniu zależności 12 i 13 w tabeli 1 i 2 zaznaczono wytłuszczonym drukiem pomiary, które nie spełniają wyżej zaproponowanego warunku niesprzeczności. W związku z powyższym są one usunięte z dalszej analizy.

W ten sposób uzyskano uśrednione wyniki pomiarów zestawione w tabeli 3 z obliczonymi błędami standardowymi dla wypadku wzięcia pod uwagę wszystkich pomiarów i wziętych tylko tych najbardziej istotnych według zaproponowanej metody selekcji.

Analiza wyników pomiarów z wykorzystaniem opracowanego modelu sprawiła, że po uśrednieniu istotnych wyników dla poszczególnych n pomiarów K zmiennych uzyskano uogólnioną wartość zmiennej Y , której błąd standardowy dla średniej jest dla każdego pomiaru istotnie mniejszy niż gdyby uśredniono wszystkie wyniki BD.

Tab. 3. Uśrednione wartości pomiarów dla wszystkich otrzymanych wyników BZ i dla wyselekcjonowanych MODEL I
Tab. 3. Average values of measurement for all received results BZ and for selected MODELS

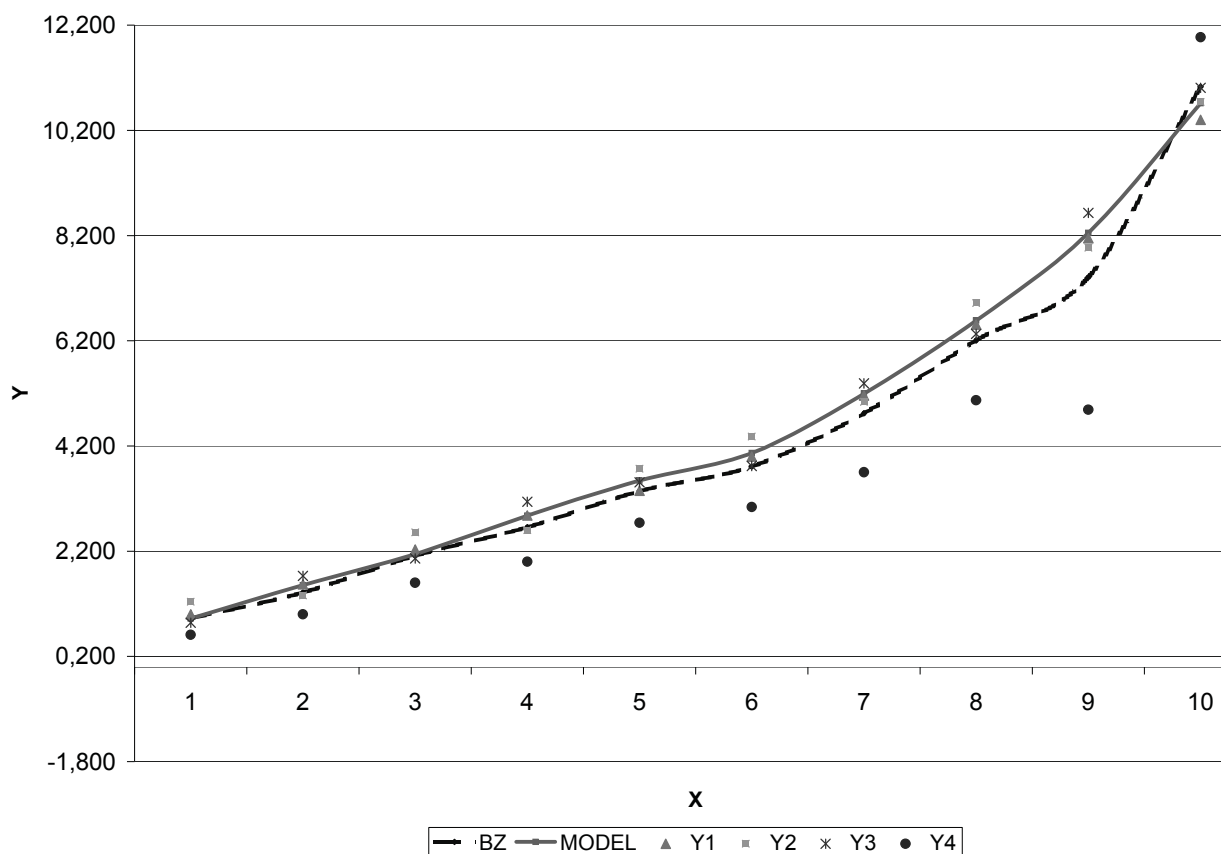
Zmienna X	Y Średnia	Y Średnia	Y Odch.Std.	Y Odch.Std.	Y Stand. Błąd	Y Stand. Błąd
	BZ	MODEL	BZ	MODEL	BZ	MODEL
1	0,923	0,920	0,265	0,113	0,133	0,080
2	1,415	1,553	0,315	0,186	0,158	0,107
3	2,113	2,145	0,400	0,120	0,200	0,085
4	2,655	2,873	0,489	0,270	0,245	0,156
5	3,343	3,543	0,437	0,212	0,219	0,122
6	3,810	4,067	0,564	0,286	0,282	0,165
7	4,823	5,197	0,762	0,178	0,381	0,103
8	6,208	6,587	0,798	0,302	0,399	0,175
9	7,410	8,250	1,703	0,341	0,851	0,197
10	11,030	10,717	0,675	0,306	0,337	0,176

W ten sposób można teraz wykreślić funkcje zależności zmiennej Y od X , co przedstawiono na rysunku. Na tym rysunku przedstawiono naniesione punkty pomiarowe dla trzech badań, każde po 10 pomiarów. Ponadto nakreślono uśrednione wyniki tych pomiarów dla dwóch przypadków 1) wszystkie wyniki pomiarów BZ, 2) wyselekcjonowane pomiary z wykorzystaniem MODELU zaproponowanego w niniejszej pracy. Drugi wypadek został nakreślony dla całkowitego usunięcia zmiennej Y_4 i dwóch punktów dla zmiennej Y_2 zawartych w tabelach 1, 2.

3. Podsumowanie

W pracy zaprezentowano metodę selekcjonowania zmiennych oraz pojedynczych punktów pomiarowych najlepiej związanych ze wszystkimi pomiarami. Analiza skupień została wykorzystana do oszacowania zmiennych najbardziej związanych ze sobą i dała ona możliwość usunięcia zmiennej najbardziej oddalonej w sensie podobieństwa (odległość euklidesowa) od pozostałych. Wykorzystano teorię ewidencji matematycznej do oszacowania stopnia przynależności punktu pomiarowego do grupy punktów. W tym celu wykorzystano miarę R korelacji rang porządkowych Spearmana, między każdą zmienną a pozostałymi zmiennymi, aby ustalić moc powiązania tych zmiennych. Następną wielkością wykorzystaną we wzorach 12 i 13 jest miara sprzeczności w sensie teorii Dempstera-Shafera. Dzięki zastosowaniu unifikacji tych trzech informacji uzyskano miarę $\gamma = \{\gamma_1, \gamma_2\}$, której dziedziną jest przedział liczb rzeczywistych $[0,1]$. Im wartość jest bliższa zero, tym poszczególne pomiar jest bardziej spójny z grupą pomiarów. Jeżeli wartość tego współczynnika dąży do 1, oznacza to, że punkt pomiarowy jest coraz słabiej związany z grupą pomiarów. W pracy przyjęto dla 4 pomiarów istotność tego współczynnika 0,1. Jeżeli wyniki $\gamma = \{\gamma_1, \gamma_2\}$ są mniejsze od 0,1, są one uznane za istotne dla całości badań.

Zaletą takiego podejścia jest możliwość wychwycenia tych pomiarów, które w istotny sposób powiększają błąd standardowy wartości średniej. Daje ono możliwość uzyskania spójnej skali wyników i tym samym istotnie poprawia ich wiarygodność.



Rys. 2. Wyniki pomiarów wraz z naniesionymi krzywymi uśrednienia danych dla wszystkich pomiarów BZ i pomiarów wyselekcjonowanych o najmniejszych błędach standardowych MODEL I

Fig. 2. Results of measurement marked curves of average data for all BZ measurements and measurement selected of MODEL least standard-errors

4. Literatura

- [1] Durham, S. D., Smolka, J. S., Valtorta, M., *Statistical consistency with Dempster's rule on diagnostic trees having uncertain performance parameters*, International journal of Approximate Reasoning, 6, 67-81, 1992.
- [2] Gordon, J., Shortliffe E. H., *The Dempster-Shafer Theory of Evidence*, [w:] Schafer G. Pearl J., [red.] Readings in Uncertain Reasoning. Morgan Kaufmann Publ., Inc., San Mateo, California, 1991.
- [3] Straszecka, E., Straszecka. J., *Interpretation of Medical Symptoms Using Fuzzy Focal Elements*, 4 th International Conference on Computer Recognition Systems - CORES2005, Rydzyna (Poland) 22-25 May, 2005.
- [4] Straszecka, E., Straszecka, J., *Medical Reasoning with Fuzzy and Uncertain Symptoms* - Proc. European Symposium on Intelligent Techniques, pp. 101-103 ESIT 2000 Aachen, Germany, 2000.
- [5] Wierzchoń, S., *Metody reprezentacji i przetwarzania informacji niepewnej w ramach teorii Dempstera-Shafera*, Instytut Podstaw Informatyki Polskiej Akademii Nauk, Warszawa, 1996.
- [6] Wang, A., *A Defect in Dempster-Shafer Theory*, Proc. Conf. Uncertainty in Artificial Intelligence, UA196, URL=<ftp://cogsci.indiana.edu/pub/wang.dempster.ps>.