



JAK UGOTOWAĆ DOBRĄ ZUPĘ, CZYLI MODELOWANIE PROCESÓW WSADOWYCH Z WYKORZYSTANIEM MSPC NA PRZYKŁADZIE PROCESU POLIMERYZACJI

Tomasz Demski, StatSoft Polska Sp. z o.o.

Ogólne przedstawienie zagadnienia

W bardzo wielu dziedzinach przemysłu występują tzw. procesy wsadowe (nazywane też okresowymi, ang. *batch process*). Są to procesy przebiegające przez pewien czas, w jednym urządzeniu (zbiorniku, reaktorze, piecu itp.) lub na jednym stanowisku pracy. Najczęściej w takich procesach pewne surowce ładowane są do zbiornika, następnie przetwarzane w zmieniających się warunkach, a na koniec uzyskany produkt pobierany jest ze zbiornika (zob. [1]). Zazwyczaj co jakiś czas dokonujemy pomiarów właściwości procesów, takich jak temperatura lub ciśnienie w różnych miejscach zbiornika, czy też strumienie wpływających i odprowadzanych z niego substancji. Przykładami takich procesów wsadowych są: ważenie piwa, polimeryzacja, fermentacja, produkcja drożdży, trawienie wafli półprzewodnikowych, lakierowanie samochodów. Warto zauważyć, że podobny układ danych i zadanie pojawiają się nie tylko w przemyśle, ale np. w badaniach klinicznych oraz przedklinicznych, gdzie poddajemy ludzi lub zwierzęta pewnym działaniom i przez określony czas badamy, co się z nimi dzieje, wykonując wiele testów, badań itp. (zob. [2]).

W codziennym życiu przykładem takiego procesu jest gotowanie zupy: do garnka wkładamy produkty spożywcze, zalewamy je wodą, podgrzewamy, ściągamy szumowiny, gotujemy, dodajemy przyprawy, potem śmietanę. Oczywiście kucharz na bieżąco sprawdza, czy wszystko przebiega dobrze, próbuje zupę i koryguje proces. Niemniej jednak zdarza się, że zupa nie wyjdzie, będzie za słona albo śmietana się zważy.

Wadliwy przebieg procesów wsadowych zazwyczaj wiąże się z bardzo dużymi kosztami, także dlatego, że zwykle po zakończeniu produkcji jednej partii, szybko rozpoczyna się kolejną, a ocena jakości produktu w laboratorium trwa zazwyczaj długo. W konsekwencji mamy całe serie zepsutych partii produktu, zanim w laboratorium wykryty zostanie zły parametr.

Procesy wsadowe są ponadto często bardzo skomplikowane i trudno jest znaleźć prostą regułę prowadzącą zawsze do dobrego produktu. Mamy dużo właściwości procesu, przebiega



on dynamicznie i niewielkie zmiany w procesie mogą prowadzić do zupełnie różnego wyniku.

Warto też zwrócić uwagę, że procesy wsadowe są często wykorzystywane w gałęziach przemysłu, w których zła jakość produktu może stanowić zagrożenie dla ludzkiego życia i zdrowia (np. w przemyśle farmaceutycznym i spożywczym).

Wszystko to powoduje, że bardzo duże korzyści dają metody analizy danych zaprojektowane z myślą o takich właśnie procesach: wielokierunkowa analiza składowych głównych (*Multiway PCA, MPCA*) i wielokierunkowa cząstkowych najmniejszych kwadratów (*Multiway PLS, MPLS*).

Analiza danych o procesach wsadowych

Podstawową trudnością w stosowaniu statystyki do badania właściwości procesów wsadowych jest układ danych. Dla każdej partii mamy ciąg pomiarów właściwości procesów (np. temperatury) dokonywanych w różnym czasie. Oznacza to, iż dla każdej partii mamy tabelę z wartościami cech procesu zmierzonymi w różnym czasie (tak jak na rysunku poniżej). Ponadto możemy mieć dane o partii produktu jako całości: surowcach użytych na wejściu procesu, parametrach finalnego produktu, zaakceptowaniu (lub nie) danej partii itd.

Partia 1	Czas	Temperatura wewnętrzna 1	Temperatura wewnętrzna 2	Temperatura medium	Ciśnienie 1	Ciśnienie 2	Strumień 1	Strumień 2
1	1							
2	2							
3	3							
4	4							
5	5							
6	6							
7	7							
8	8							

Partia 2	1	2	3	4	5	6	7	8
Czas	Temperatura wewnętrzna 1	Temperatura wewnętrzna 2	Temperatura medium	Ciśnienie 1	Ciśnienie 2	Strumień 1	Strumień 2	
1								
2								
3								
4								
5								
6								
7								
8								

Partia 3	1	2	3	4	5	6	7	8
Czas	Temperatura wewnętrzna 1	Temperatura wewnętrzna 2	Temperatura medium	Ciśnienie 1	Ciśnienie 2	Strumień 1	Strumień 2	
1	1	120,05	120,08	160,07	1,968	2,150	1,005	0,999
2	2	120,06	120,08	160,09	2,128	2,141	1,015	0,993
3	3	120,07	120,10	160,04	2,088	1,840	0,994	0,977
4	4	120,06	120,10	160,06	2,120	1,915	0,988	1,008
5	5	120,06	120,10	160,06	1,903	2,037	0,988	0,994
6	6	120,06	120,08	160,06	1,873	1,965	1,008	0,993
7	7	120,06	120,10	160,06	2,099	1,929	1,006	1,016
8	8	120,04	120,09	160,09	2,011	1,905	1,006	0,984

Przed analizą dane należy przekształcić tak, aby obiektom podlegającym analizie odpowiadały przypadki (wiersze tabeli), a badane właściwości obiektów zapisywane są w zmiennych (kolumnach tabeli). Stosuje się tu dwa podejścia:

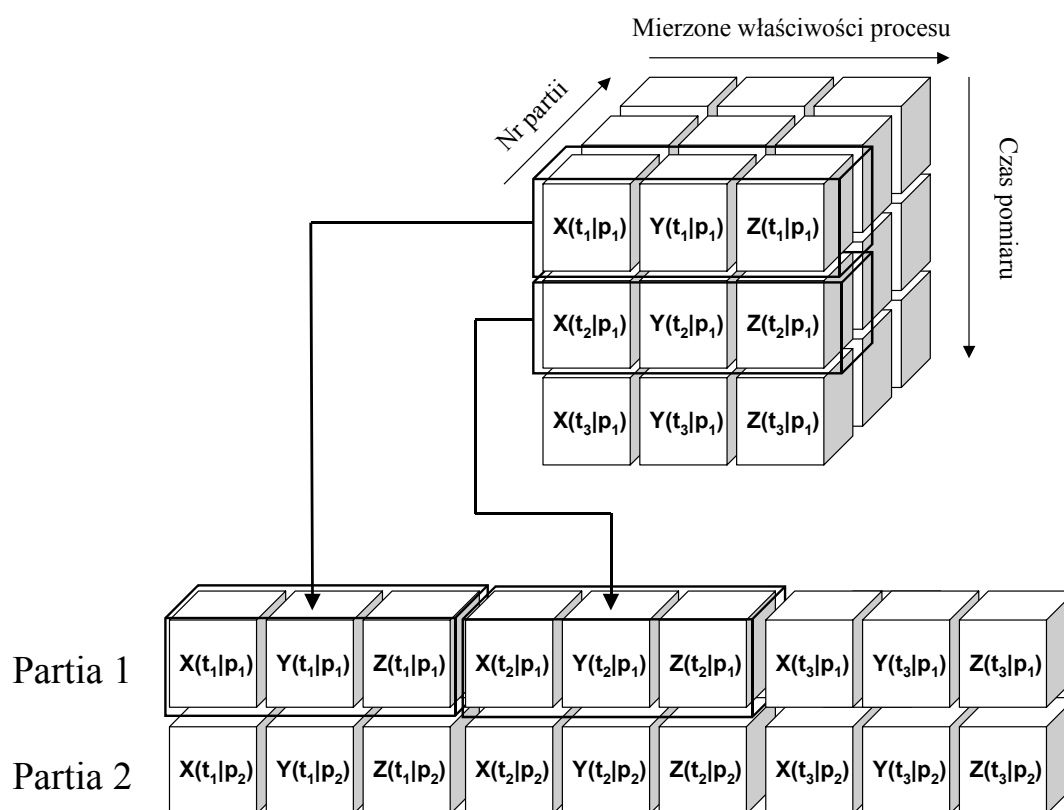
1. wielokierunkowa analiza według czasu (ang. *time-wise multi-way PCA* i *PLS*, w skrócie *TMPCA* i *TMPLS*),
2. wielokierunkowa analiza według partii (ang. *batch-wise multi-way*, w skrócie *BMPLS*).



Poniżej przedstawimy pokrótce oba te podejścia. Zostały one dokładnie przedstawione w artykułach [1] i [2]. Porównaniu obu metod poświęcony jest artykuł [3]. Zwróćmy uwagę, iż oba te podejścia możemy zrealizować w *STATISTICA*.

Wielokierunkowa analiza według czasu

W tym wypadku tworzymy tabelę (macierz dwuwymiarową), w której każdej partii odpowiada jeden wiersz. Następnie w zmiennych (kolumnach) umieszczamy wartości wybranej właściwości procesu dla danej partii zmierzone w pewnym momencie (np. temperatura na początku procesu będzie pierwszą zmienną, temperatura po 15 minutach drugą itd., aż do zakończenia procesu). W ten sposób postępujemy dla wszystkich mierzonych właściwości procesu (np. 10 pierwszych zmiennych będą to kolejne pomiary temperatury, 10 następnych będzie zawierało ciśnienia itd.). Całą procedurę przedstawia poniższy rysunek, a dokładniejszy i bardziej formalny opis transformacji dostępny jest w pracy [1]



W razie potrzeby po wykonaniu powyższego przekształcenia dane uzupełniamy o ogólne właściwości danej partii, takie jak: klasa jakości produktu i jego właściwości, uzyskana wydajność i parametry surowców.

Dużym utrudnieniem analizy danych dla procesów wsadowych jest ogromna liczba otrzymanych zmiennych. Jeśli bierzemy pod uwagę tylko 10 właściwości procesu i mierzymy je 100 razy (a nie jest to wcale bardzo często – w praktyce pomiary mogą odbywać się co kilka sekund), to uzyskamy łącznie 1000 zmiennych. Stosowanie w takim wypadku



jednowymiarowych technik analiz danych (np. kart kontrolnych Shewharta) raczej nie wchodzi w grę. Trudno sobie wyobrazić praktyczne stosowanie tysięcy kart kontrolnych każdej zmiennej z osobna, a poprawność i celowość takiego podejścia jest bardzo wątpliwa.

Rozwiązaniem jest zastosowanie wobec zmiennych z kolejnymi pomiarami parametrów procesu techniki redukcji wymiarów: analizy składowych głównych (PCA) lub metody uogólnionych najmniejszych kwadratów (PLS). Metody te umożliwiają uzyskanie kilku składowych (zmiennych ukrytych), które niosą praktycznie całą interesującą informację zawartą w dużej liczbie zmiennych wejściowych. Metodę składowych głównych stosujemy w monitorowaniu procesu, a PLS w budowie modeli predykcyjnych służących do przewidywania właściwości produktu. W przypadku obu metod składowe uzyskuje się jako kombinacje liniowe oryginalnych zmiennych (więcej informacji o analizie składowych głównych można znaleźć w pracy [4]).

Ze względu na dużą liczbę zmiennych do wykonania PCA i PLS stosuje się algorytm NIPALS (zob. [1]).

Związki między zmiennymi są zazwyczaj nieliniowe i dynamiczne, a proces jest sterowany, tak aby parametry zmieniały się według pewnej trajektorii. Najprostszym rozwiązaniem jest przeprowadzenie analizy na odchyleniach wartości parametrów od średniego przebiegu, innymi słowy przed analizą od każdej uwzględnianej zmiennej odejmujemy jej średnią.

Na koniec wykonujemy jeszcze standaryzację, tj: dzielimy odchylenia od średniej przez odchylenie standardowe tej zmiennej. Dzięki temu sprowadzimy wszystkie mierzone wielkości do jednej skali, co jest konieczne, ponieważ mierzone parametry są wyrażone w różnych jednostkach (temperatura w kelwinach, ciśnienie w paskalach itp.)

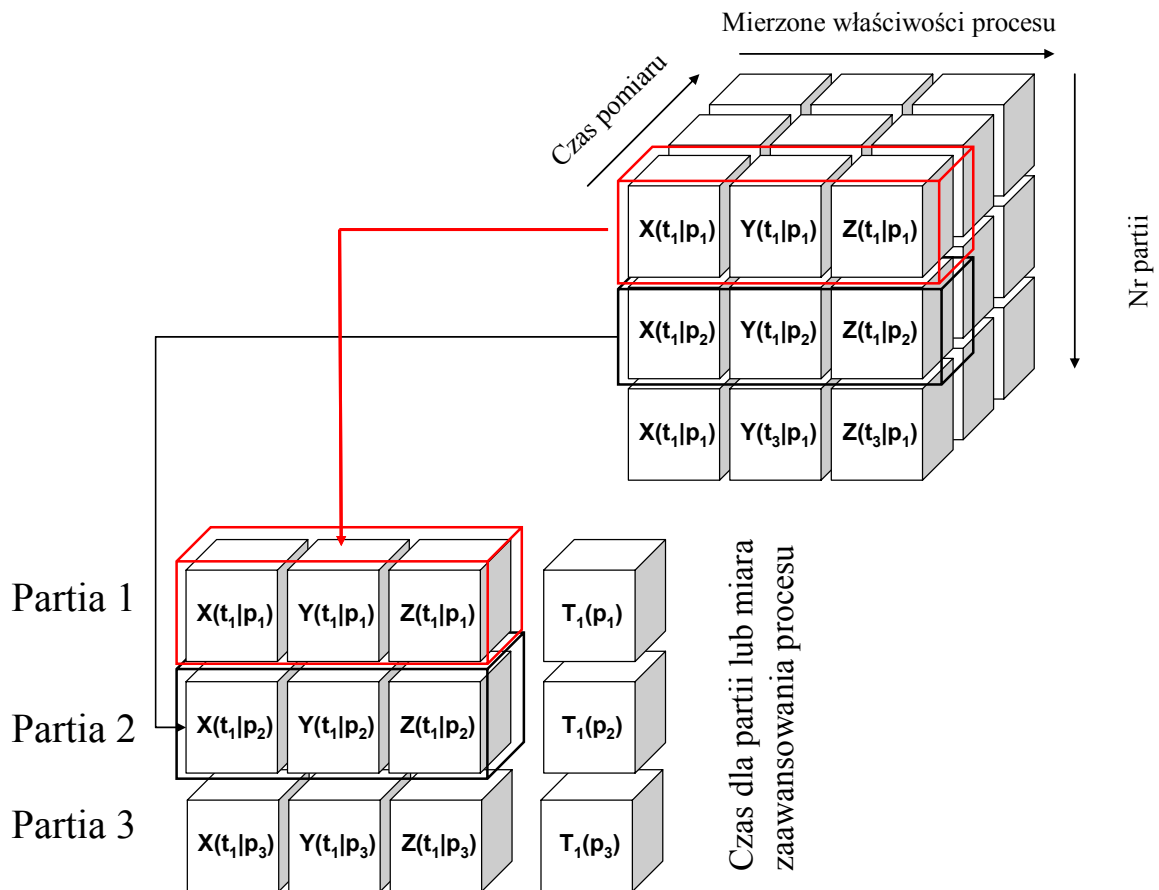
Wielokierunkowa analiza według partii

Podejście to zostało zaproponowane w pracy [2]. W tym podejściu wyniki pomiarów łączone są według partii: łączymy wyniki pomiarów tych samych właściwości dokonane na różnym etapie procesu. Przykładowo założmy, że pierwszą mierzoną właściwością jest temperatura. Pierwszą kolumnę tabeli do analizy utworzą wartości temperatury: najpierw będą to pomiary temperatury na początku procesu dla pierwszej partii, potem dla drugiej, trzeciej itd., aż do ostatniej partii. Następnie do pierwszej kolumny wpisujemy pomiary temperatury w drugim kroku procesu dla pierwszej partii, drugiej itd. Można powiedzieć, że sklejamy dane w kierunku partii, podczas gdy w omówionym wcześniej podejściu sklejaliśmy dane w kierunku czasu. Ponadto dane uzupełniamy zmienną zawierającą czas pomiaru dla partii (czas lokalny) lub odpowiednią dla konkretnego procesu miarą jego zaawansowanie (przykładem takiej zmiennej jest zawartość alkoholu dla fermentacji alkoholowej). Tworzenie tabeli danych do analizy pokazuje rysunek poniżej.

Zmienne uzyskanej w ten sposób tabeli są następnie centrowane (poprzez odjęcie średniej dla zmiennych) i standaryzowane (poprzez podzielenie przez odchylenie standardowe). Zauważmy jednak, iż tym razem średnia i odchylenie standardowe są obliczane dla wszystkich partii i wszystkich czasów, w efekcie analizujemy dane o odchyleniach temperatury



od średniej temperatury dla **całego** procesu, natomiast w podejściu według czasu badaliśmy odchylenie od średniej w konkretnym momencie, innymi słowy odchylenie od typowej trajektorii. Taka transformacja danych oznacza, iż odchylenie od średniej trajektorii musi być wydobyte w inny sposób.



Dla tak uzyskanej tabeli danych budujemy model PLS, w którym zmienną zależną jest czas lub miara zaawansowania procesu. Wyniki modelu są następnie analizowane tak, aby uzyskać typową trajektorię dla „dobrych” partii oraz zbudować model pozwalający przewidywać właściwości partii na podstawie warunków początkowych i wyników modelu PLS dla tej partii. Szczegółowy opis obliczeń można znaleźć w [2]

Zaletą wielokierunkowej analizy według partii jest możliwość bezpośredniego stosowania jej dla procesów o różnej długości oraz do monitorowania procesu w jego toku. W przypadku analizy według czasu, aby utworzyć karty kontrolne i obliczyć statystyki procesu, powinniśmy zebrać dane o całym procesie (oczywiście trudność tę można rozwiązać, np. przewidując dalszy przebieg procesu, zagadnienie to omówione jest w [1]), natomiast analiza według partii w naturalny sposób działa dla dowolnego etapu procesu. Jak piszą autorzy [2], analiza według partii jest bardziej skupiona na śledzeniu przebiegu procesu, tego co się dzieje z konkretną partią, natomiast podejście według czasu dotyczy bardziej rozróżniania dobrych i złych partii jako całości.



Analiza danych

Jeśli wykonujemy wielokierunkową analizę według partii, to przekształcone dane możemy analizować metodą składowych głównych (PCA) lub cząstkowych najmniejszych kwadratów (PLS). PCA stosujemy, gdy brak jest zmiennych zależnych, a PLS gdy mamy takie zmienne, tj. wielkości, które chcemy przewidywać na podstawie przebiegu procesu.

W wyniku PCA ze zmiennych charakteryzujących proces (tj. kolejnych pomiarów) uzyskujemy zestaw nowych zmiennych, które nazywamy składowymi głównymi. Monitorowanie procesu będziemy prowadzić na podstawie wartości składowych głównych dla poszczególnych partii oraz rozbieżności między modelem składowych głównych a obserwowanymi wartościami zmiennych.

Pierwszą decyzją, jaką należy podjąć, jest ustalenie odpowiedniej liczby wyodrębnianych składowych. Do tego celu stosuje się bardzo wiele różnych metod, np. przyjmuje się, że składowa główna powinna wyjaśniać co najmniej tyle zmienności oryginalnych danych, ile „średnio” przypada na zmienną (kryterium to formułuje się najczęściej w ten sposób, że wartość własna odpowiadająca danej składowej głównej ma być większa od 1. W MPCA często stosuje się metody bazujące na sprawdzianach krzyżowych.

Sprawdzian krzyżowy polega na wielokrotnym tworzeniu modeli, przy czym za każdym razem jedną partię (lub losowo wybraną grupę partii) odkładamy do testowania, a składowe główne wyznaczamy na podstawie pozostałych partii. Następnie dla modelu uwzględniającego określoną liczbę składowych głównych sprawdzamy, jak wygląda błąd dla partii nieużywanej przy wyznaczaniu składowych. Jeśli w modelu zaczniemy uwzględniać zbędne lub „przypadkowe” składowe, to błąd dla partii testowej przestaje spadać, a nawet zaczyna rosnąć. Zwykle taki sposób doboru liczby składowych jest bardzo skuteczny. Więcej informacji o doborze liczby składowych znajduje się w pracach [1] i [4].

Często użyteczne jest przewidywanie właściwości produktu lub procesu, jak również końcowych właściwości produktu w czasie przetwarzania danej partii, jeszcze przed zakończeniem produkcji. W takim wypadku wyróżniamy zmienne przyczyn (zmienne niezależne) i zmienne wyników (zmienne zależne lub odpowiedzi).

Przed utworzeniem modelu wykonujemy transformację danych opisaną powyżej, podobnie jak w przypadku monitorowania procesu. Jako technikę modelowania najczęściej stosuje się PLS, który buduje model nie dla obserwowanych zmiennych, ale dla wyodrębnionych składowych głównych. Ponieważ proces charakteryzuje wiele cech (które będą zmiennymi niezależnymi), redukcja wymiarów jest niezwykle użyteczna. Ponadto model PLS uwzględnia wiele zmiennych zależnych.

W przypadku wielokierunkowej analizy według partii zawsze wykonujemy model PLS, bo zawsze mamy zmienną zależną.



Przykład

Dane dotyczą procesu polimeryzacji opisanego dokładniej w pracy [1]. Dysponujemy następującymi właściwościami procesu mierzonymi w 100 punktach czasu:

- ♦ temperatura w 3 punktach reaktora (zmiennie $Temp1$, $Temp2$, $Temp3$),
- ♦ temperatura medium chłodzącego/ogrzewającego (zmiennie $TM1$, $TM2$),
- ♦ ciśnienie w trzech punktach (zmiennie $P1$, $P2$, $P3$),
- ♦ strumień dwóch substancji.

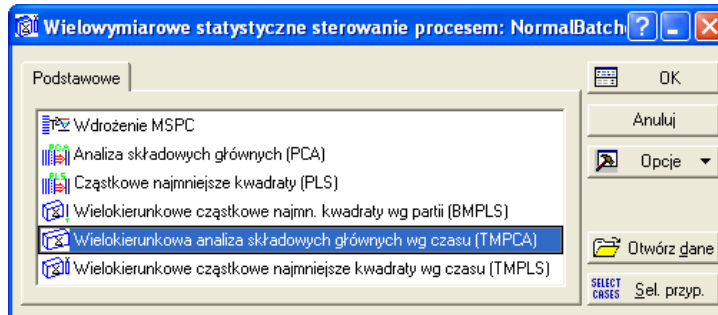
Wyniki badań laboratoryjnych wytworzonego polimeru otrzymywane są po około 12 godzinach od zakończenia wytwarzania. Oczywiście w tym czasie produkowane są kolejne partie i jeśli wystąpił błąd w przebiegu procesu nie ma możliwości jego skorygowania dla następnych partii – w efekcie one również są nieodpowiednie.

Dane: NormalBatches.sta (12 zmn. * 3000 prz.)												
	1	2	3	4	5	6	7	8	9	10	11	12
	Partia	Temp1	Temp2	Temp3	P1	Str1	TM1	TM2	P2	P3	Str2	Czas
1	B1	0,57039	0,887247	0,546655	0,98417	0,5202	0,989112	0,933139	0,954723	0,727997	1,387073	0
2	B1	0,576384	0,862227	0,552975	0,979343	0,7248	0,989199	0,93337	0,956171	0,72632	1,44286	1
3	B1	0,582171	0,823841	0,559798	0,975297	0,7837	0,989719	0,933601	0,956344	0,728416	1,487092	2
4	B1	0,589163	0,795793	0,566655	0,976823	0,8186	0,989834	0,933804	0,95481	0,734842	1,511825	3
5	B1	0,597086	0,779892	0,573479	0,975901	0,7854	0,988592	0,934498	0,954781	0,740011	1,523741	4
6	B1	0,605835	0,770938	0,581277	0,975901	0,8064	0,988477	0,935511	0,955707	0,739033	1,538545	5
7	B1	0,614619	0,76567	0,589345	0,977285	0,8139	0,989574	0,936524	0,956257	0,740011	1,537281	6
8	B1	0,622714	0,761555	0,597076	0,975581	0,7811	0,990239	0,93716	0,957039	0,74518	1,515797	7
9	B1	0,630671	0,759415	0,604807	0,974942	0,7909	0,990556	0,937536	0,959008	0,748952	1,490522	8
10	B1	0,639454	0,758625	0,612538	0,976149	0,7979	0,990874	0,937912	0,960252	0,751746	1,462719	9
11	B1	0,64996	0,757868	0,620269	0,97732	0,773	0,991163	0,938317	0,960513	0,752864	1,435458	10
12	B1	0,662671	0,757671	0,629479	0,976575	0,7385	0,99148	0,938693	0,960745	0,752026	1,406752	11
13	B1	0,676415	0,758889	0,63916	0,974481	0,7168	0,991654	0,93907	0,962395	0,750629	1,376061	12
14	B1	0,688609	0,760403	0,650252	0,973771	0,7098	0,991683	0,939446	0,962105	0,750908	1,337967	13
15	B1	0,701113	0,762576	0,661311	0,97402	0,7068	0,991711	0,940429	0,96251	0,755518	1,281098	14
101	B2	0,561848	0,885173	0,537042	0,941331	0,1247	0,988564	0,932068	0,954694	0,723945	1,29536	0
102	B2	0,568978	0,891526	0,541479	0,984845	0,4529	0,988332	0,9323	0,954578	0,722688	1,369561	1
103	B2	0,576556	0,869963	0,547832	0,979201	0,7034	0,988332	0,932531	0,954463	0,724923	1,427333	2
104	B2	0,583342	0,831084	0,555429	0,976398	0,7981	0,988881	0,93311	0,95481	0,728975	1,471565	3
105	B2	0,590128	0,799677	0,563429	0,97764	0,8053	0,988101	0,93392	0,953826	0,734004	1,502257	4
106	B2	0,597361	0,782098	0,57163	0,975368	0,8165	0,98839	0,934701	0,953884	0,739452	1,515617	5
107	B2	0,604905	0,772584	0,580034	0,976859	0,8347	0,989314	0,935453	0,955621	0,740151	1,527352	6
108	B2	0,613103	0,765934	0,588403	0,975936	0,8001	0,98969	0,936176	0,957329	0,74071	1,529879	7
109	B2	0,621611	0,76195	0,596269	0,974729	0,819	0,990036	0,936697	0,957531	0,740989	1,525005	8
110	B2	0,630085	0,759481	0,604941	0,976646	0,8158	0,990412	0,937073	0,956663	0,742805	1,507131	9
111	B2	0,640074	0,757638	0,615664	0,975404	0,7826	0,990787	0,93742	0,958342	0,747835	1,473551	10
112	B2	0,651752	0,757177	0,627193	0,975581	0,7824	0,991163	0,937507	0,959587	0,75007	1,44304	11
113	B2	0,663257	0,75744	0,63758	0,976539	0,757	0,990903	0,938115	0,960397	0,748533	1,417043	12
114	B2	0,675864	0,758263	0,646756	0,974765	0,725	0,99021	0,938722	0,961237	0,750629	1,378227	13
115	B2	0,687265	0,759942	0,65563	0,973807	0,7162	0,990065	0,938896	0,960542	0,751187	1,340134	14
116	B2	0,698564	0,762312	0,664471	0,974161	0,7143	0,990383	0,939966	0,961208	0,751746	1,293555	15
117	B2	0,714374	0,764551	0,673647	0,974516	0,7089	0,99073	0,940458	0,96251	0,752864	1,237588	16
118	B2	0,735386	0,76725	0,683059	0,974871	0,7011	0,991134	0,940053	0,963871	0,752026	1,188121	17
119	B2	0,755778	0,771135	0,692168	0,975226	0,6932	0,990989	0,939359	0,963582	0,749092	1,153638	18

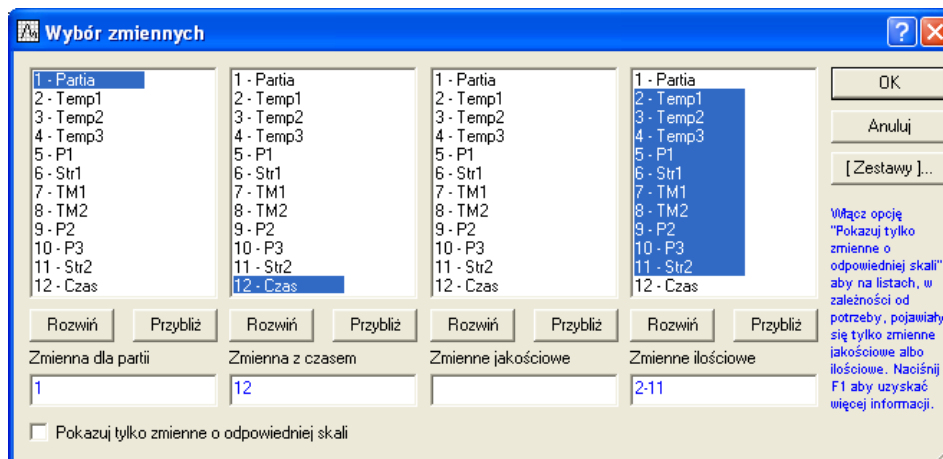
Do diagnostyki procesów zastosujemy metodę wielokierunkowego PCA według czasu. Zaczniemy od zbudowania modelu prawidłowego procesu. Dysponujemy danymi o 30 takich seriach. Wycinek wejściowego arkusza danych widzimy powyżej.



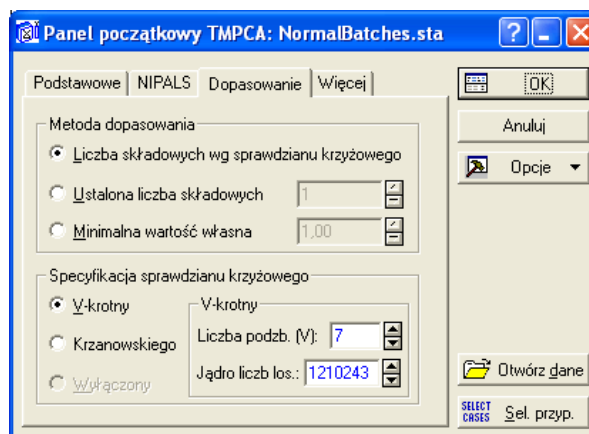
Analizę wykonamy w pakiecie *STATISTICA* MSPC. Umożliwia on analizę danych surowych – automatycznie wykonuje przekształcenia opisane powyżej. Po uruchomieniu programu z menu *Statystyka* wybieramy polecenie *PCA, PLS, wielowymiarowe SPC*. Następnie jako typ analizy wskazujemy *Wielokierunkową analizę składowych głównych według czasu*. Zwróćmy uwagę, że do dyspozycji mamy również inne metody, w tym dla przekształconych już danych.



W kolejnym oknie naciskamy przycisk *Zmienne* i wybieramy zmienne do analizy tak jak na rysunku poniżej.



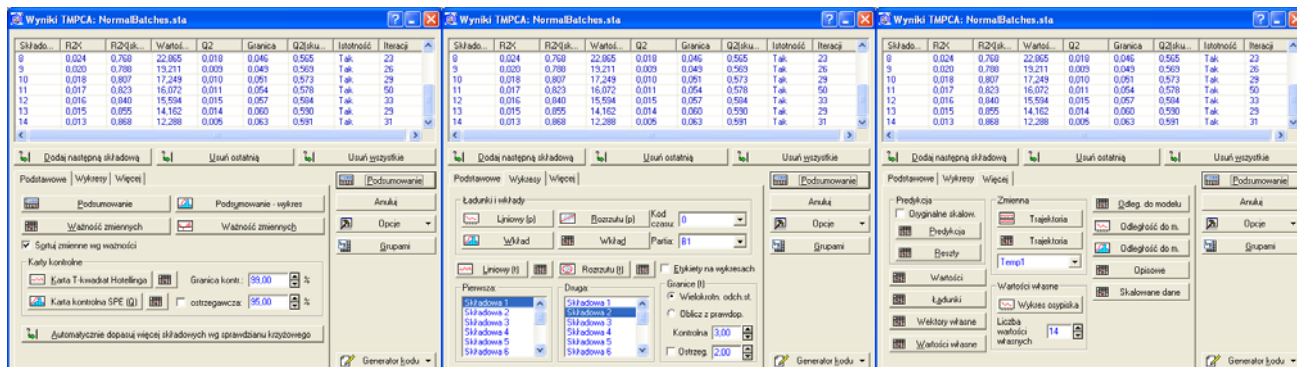
Domyślne ustawienia programu są odpowiednie dla naszego zadania i nie będziemy ich zmieniać. Warto jednak zwrócić uwagę na sposób określania liczby składowych głównych tworzących model.





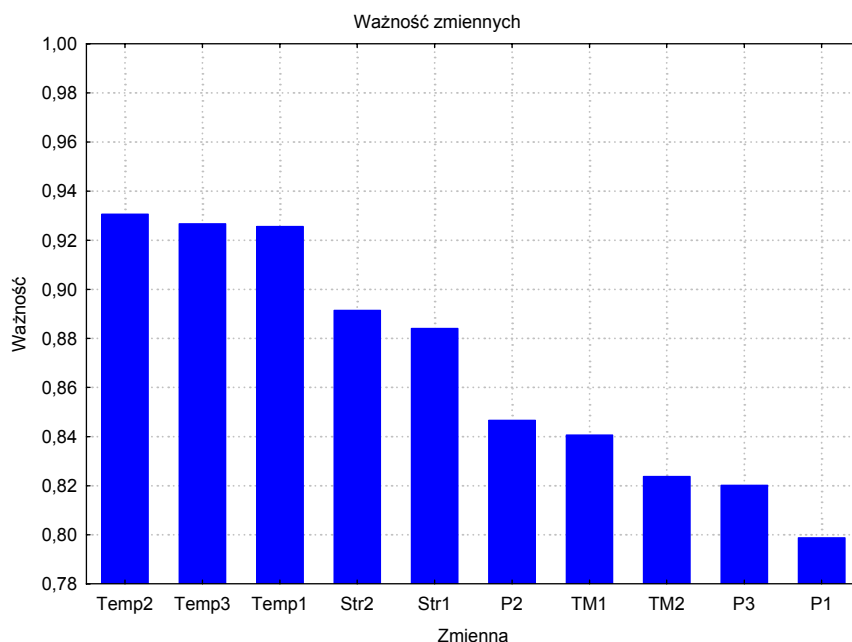
Na karcie *Dopasowanie* (zob. rysunek powyżej) do dyspozycji mamy trzy sposoby dobru liczby składowych, domyślny to sprawdzian krzyżowy i właśnie to podejście zastosujemy.

Za pomocą *STATISTICA* możemy tworzyć mnóstwo rozmaitych wykresów i statystyk podsumowujących przebieg procesu (zob. rysunek poniżej).

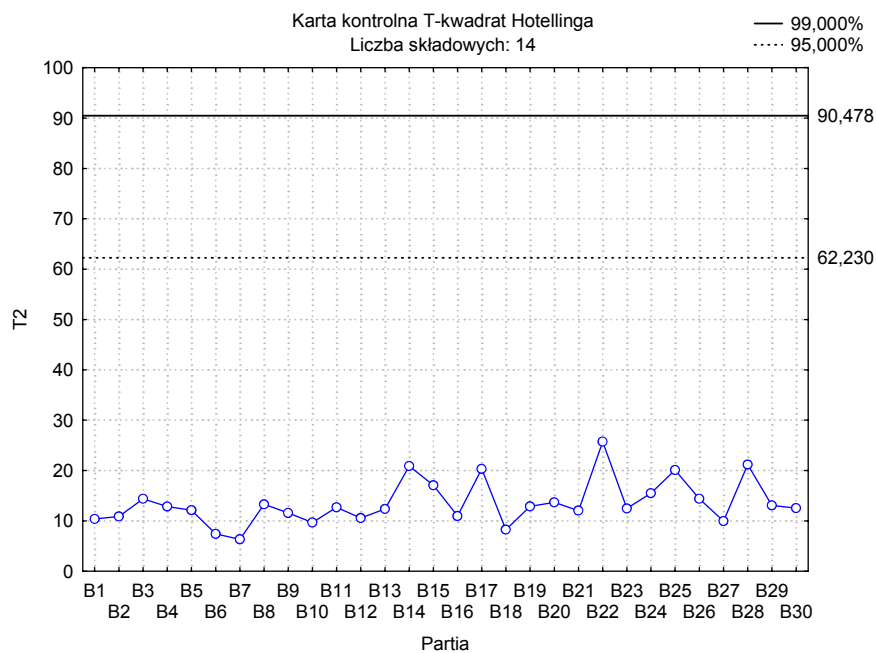


Pierwszy ważny wynik to liczba składowych głównych i procent wyjaśnianej zmienności danych. W naszym przypadku model zawiera 14 składowych, które łącznie wyjaśniają około 87% całej zmienności danych. Zauważmy, iż wyjściowa liczba zmiennych to aż 1000 (10 właściwości procesów * 100 momentów pomiaru).

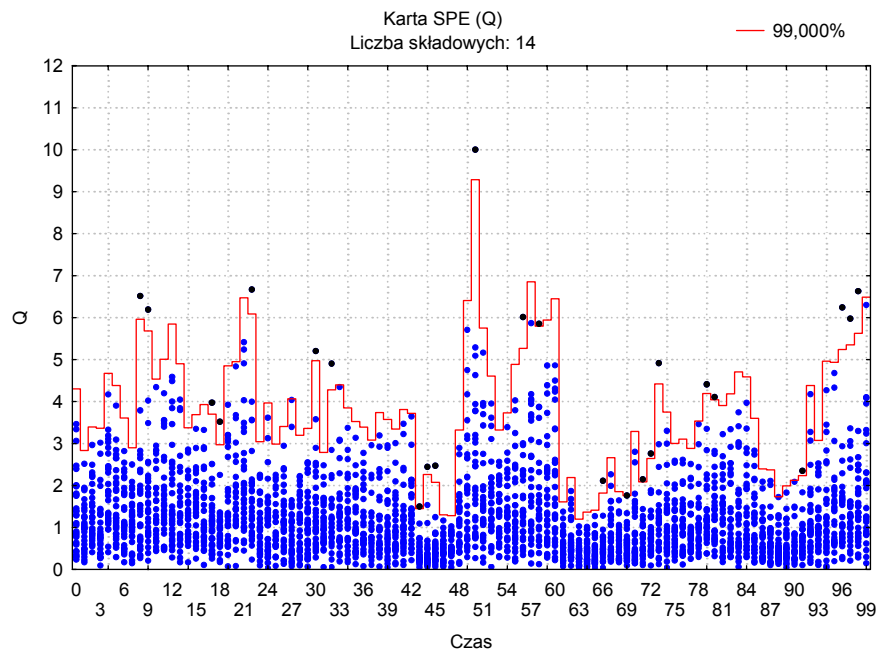
Warto sprawdzić, które właściwości są ważne, być może którąś z nich można pominąć. Jak widzimy na poniższym wykresie, wszystkie właściwości są dość ważne, a najbardziej istotne zmienne to temperatury w trzech miejscach reaktora.



Popatrzmy teraz na kartę kontrolną Hotellinga. Jest to wielowymiarowa karta dla wyodrębnionych przez nas składowych głównych. Pokazuje ona, czy któraś z partii nie jest wyraźnie inna od pozostałych. Jak widać na rysunku poniżej, karta nie sygnalizuje żadnego rozregulowania.



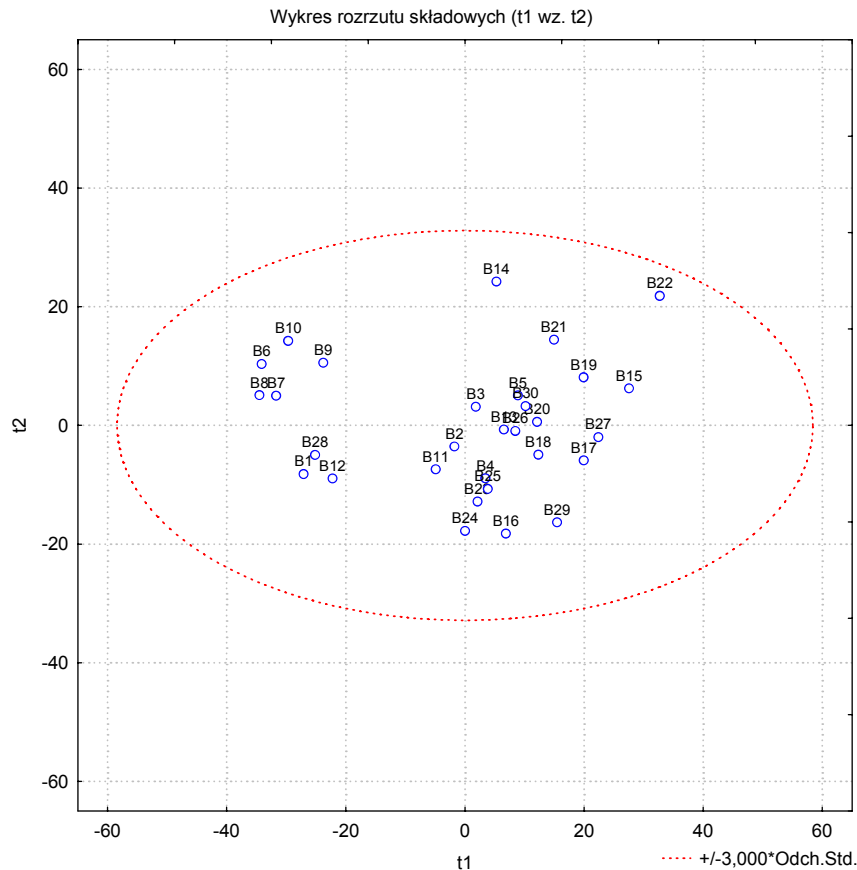
Do sprawdzenia, jak w czasie zmieniały się poszczególne partie służy karta SPE(Q). Pokazuje ona, jak w danym momencie wyglądała odległość właściwości procesu dla danej partii, od wartości wynikających z modelu. Kartę SPE(Q) widzimy poniżej i tym razem wystąpiły na niej 24 sygnały o rozregulowaniu. Zauważmy jednak, że karta przedstawia 3000 punktów (100 pomiarów dla 30 partii), a prawdopodobieństwo fałszywego sygnału o rozregulowaniu wynosi 1%, więc dla prawidłowego procesu powinniśmy mieć mniej więcej tyle fałszywych sygnałów.



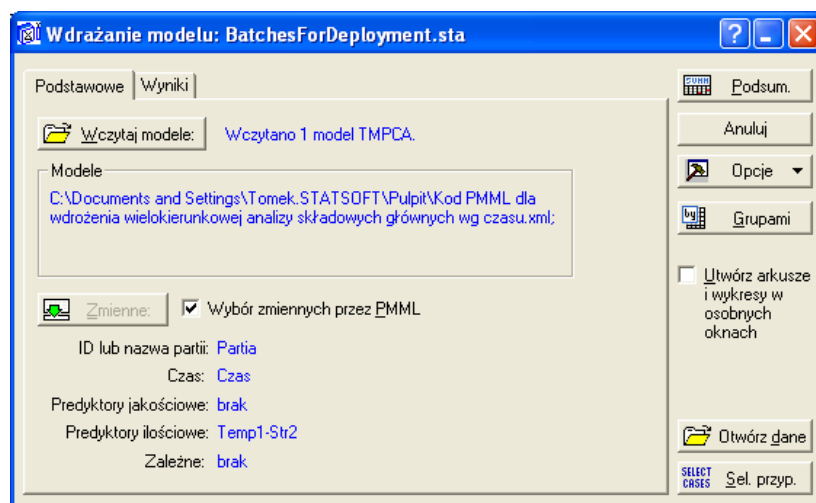
Jednorodność badanych partii możemy sprawdzić za pomocą wykresu wartości dwóch najważniejszych składowych dla poszczególnych partii. Można powiedzieć, że wykres taki



stanowi mapę partii i w skondensowany sposób pokazuje ich wzajemne relacje, które są podobne, które różne, czy występują jakieś zgrupowania partii. Wykres wartości składowych znajduje się na poniższym rysunku. Wszystkie partie mieszczą się w obszarze ± 3 odchyłeń standardowych, nie ma też wyraźnych, odległych od siebie zgrupowań punktów.



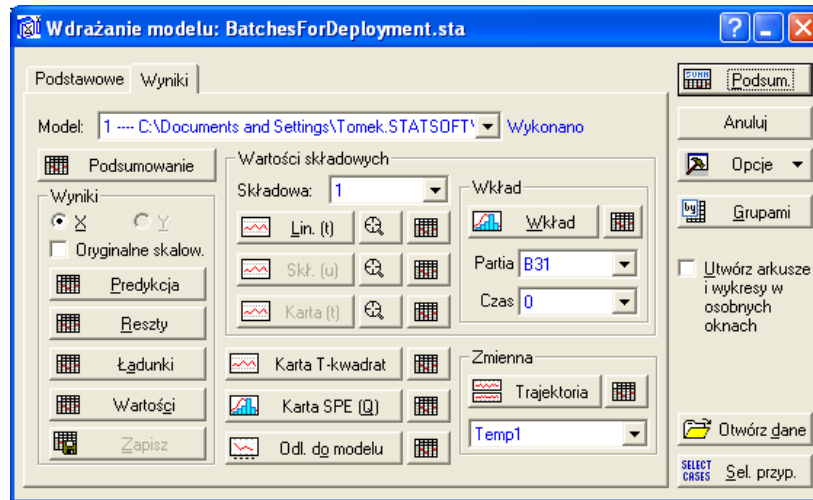
Na podstawie dotychczasowych analiz możemy powiedzieć, że badane partie są jednorodne, nie występują dla nich rozregulowania procesu. Ponadto wiemy, że uzyskany produkt spełnia wymogi jakościowe. Możemy zastosować nasz model dla nowych danych.



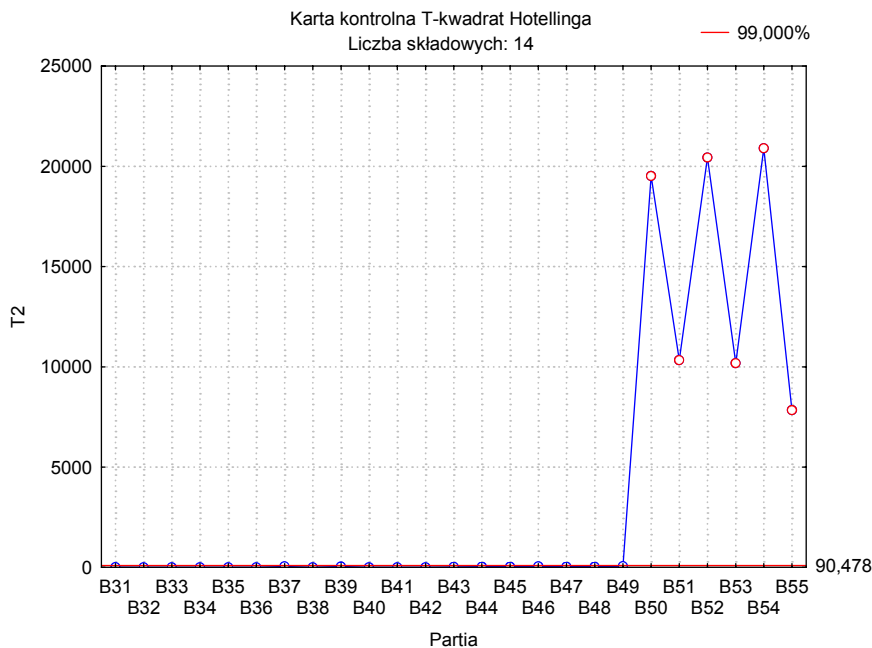


Model zapisujemy, naciskając przycisk *Generator kodu*. Do wyboru mamy *Wdrożenie w STATISTICA Enterprise* lub utworzenie zbioru z kodem PMML. Zapisujemy nasz model jako plik PMML, następnie otwieramy dane do wdrożenia w *STATISTICA*. Z menu *Statystyka* wybieramy polecenie *PCA, PLS, wielowymiarowe SPC*. Następnie jako typ analizy wskazujemy *Wdrożenie MSPC*. W oknie *Wdrażanie modelu* wskazujemy plik z kodem po kliknięciu przycisku *Wczytaj modele* – i możemy rozpocząć badanie nowych partii.

Na karcie *Wyniki* mamy do dyspozycji rozmaite wykresy i statystyki.



Tym razem, jeśli utworzymy kartę Hotellinga, od razu zauważymy, że począwszy od 50 partii wystąpiły bardzo duże problemy w procesie. Faktycznie partie te dały nieodpowiedni produkt.





Podsumowanie

Zastosowanie wielokierunkowej analizy danych pozwala diagnozować skomplikowane procesy wsadowe. W szczególności umożliwia szybsze wykrywanie rozregulowań – jeszcze w toku procesu lub tuż po jego zakończeniu, dzięki czemu nie musimy czekać na wyniki badań laboratoryjnych dla partii produktu. Jednocześnie metody te wymagają dosyć złożonych przekształceń danych i zaawansowanych metod analizy: składowych głównych i cząstkowych najmniejszych kwadratów. Jednak w praktyce wykonanie samej analizy, a potem jej stosowanie jest stosunkowo łatwe, dzięki wdrożeniu tych metod w programie *STATISTICA*.

Literatura

1. Nomikos P., MacGregor J.F., „Multivariate SPC Charts for Monitoring Batch Process”, *Technometrics* 37(1), 41-59.
2. Wold S., Kettaneh N., Froden H., Holmberg A., „Modeling and diagnostics of batch processes and analogous kinetic experiments”, *Chemometrics and intelligent laboratory systems* 44 (1998) 331-340.
3. Gracia S. „Batch Process Analysis using MPCA/MPLS. A Comparison Study”, <http://www.cheric.org/ippage/p/ipdata/2001/05/file/ip2001-03.pdf>.
4. Mazerski J., „Podstawy chemometrii”, Wydawnictwo Politechniki Gdańskiej 2000.