



# MONITOROWANIE I STEROWANIE JAKOŚCIĄ PROCESÓW WSADOWYCH (OKRESOWYCH)

**Tomasz Demski**

*StatSoft Polska Sp. z o.o.*

## Wprowadzenie

Procesy wsadowe (nazywane też okresowymi, ang. *batch process*) to procesy przebiegające przez pewien określony czas, w których pewne surowce ładowane są do zbiornika, następnie przetwarzane w zmieniających się warunkach (podgrzewanie, chłodzenie, zmiana ciśnienia, dodawanie i odprowadzanie substancji itd.), a na koniec finalny produkt (i ewentualne odpady) pobierane są ze zbiornika (zob. [1]). Produkt poddawany jest zazwyczaj wielu skomplikowanym badaniom laboratoryjnym, które określają jego jakość. Przykładami procesów wsadowych są: wytwarzanie polimerów, fermentacja, destylacja czy krystalizacja. Procesy takie stosuje się w przemyśle chemicznym, petrochemicznym, farmaceutycznym, papierniczym i spożywczym.

Procesy wsadowe są często wykorzystywane w gałęziach przemysłu, gdzie wymagania odnośnie właściwości produktu są bardzo wysokie, a ich spełnienie jest niezwykle ważne (np. w produkcji leków). Ponadto koszty powstające na skutek wytworzenia nieodpowiedniej partii produktu są zazwyczaj bardzo duże.

Sterowanie procesem wsadowym jest ważne i trudne również dlatego, że wiele procesów jest bardzo czułych nawet na drobne zmiany parametrów, zachodzące w toku procesów (np. fermentacja, zob. [2] i [3]). Odchylenia od optymalnych wartości parametrów technologicznych wpływają na jakość i ilość wytwarzanego produktu.

Dodatkową trudnością jest częste wymaganie zagwarantowania stosowania konsystentnych i udokumentowanych procedur wytwarzania. Niespełnienie tego warunku przekreśla możliwość sprzedaży produktu.

Najczęstszym sposobem sterowania procesem jest wykorzystanie wiedzy i doświadczenia operatora, który korzystając z informacji o dotychczasowym przebiegu procesu (wartościach parametrów) steruje systemem tak, aby uniknąć problemów i zapewnić jego optymalne działanie. Wady takiego podejścia to niewielkie wykorzystanie danych historycznych, arbitralność podejmowanych decyzji i uzależnienie się od umiejętności operatora. Ponadto przy uruchamianiu nowych procesów czas nabycia przez operatorów odpowiedniego doświadczenia może być długi.



Historyczne dane wykorzystywane są przez drugie podejście do sterowania procesami. W tym wypadku na podstawie danych historycznych opracowuje się reguły decydujące, jakie powinny być optymalne wartości poszczególnych parametrów i w jaki sposób należy je korygować, jeżeli wystąpią odchylenia. Podejście to jednak niezbyt dobrze sprawdza się w monitorowaniu procesów.

Trzecie podejście wykorzystuje ideę sprawdzoną w wielu dziedzinach przemysłu i usług: statystyczne sterowanie procesem (*Statistical Process Control, SPC*; zob. [5] lub [6]).

Często wyniki testów laboratoryjnych wytworzonych substancji uzyskiwane są po kilku godzinach, a przetwarzanie kolejnej partii rozpoczyna się wkrótce po zakończeniu poprzedniego procesu. Informacja, iż poprzedni proces przebiegł źle, jest dostępna dopiero, gdy produkcja kolejnej partii jest już zaawansowana lub nawet zakończona. Skutkiem tego mogą być całe serie partii produktu niespełniające wymagań. Zwiększa to korzyści, jakie dają ocenianie procesów na podstawie ich parametrów i wykrywanie rozregulowań, a tak właśnie działa SPC – jest to bardzo ważny argument przemawiający za stosowaniem podejścia SPC.

Faktycznie od dawna stosowano SPC do monitorowania właściwości finalnego produktu. Jednak w przypadku parametrów samego procesu pojawiają się trudności, których rozwiązanie zaproponowano dopiero w latach dziewięćdziesiątych (zob. [1]).

Pierwszą z trudności jest specyficzny układ danych. Otóż w przypadku procesów wsadowych dla każdej partii mamy ciąg pomiarów właściwości procesów (np. temperatury), dokonywanych w pewnych odstępach czasu. W efekcie dla każdej partii produktu mamy tabelę z wartościami cech procesu zmierzonymi w różnym czasie (tak jak na rysunku poniżej). Ponadto mamy informacje o partii produktu jako całości: surowcach użytych na wejściu procesu, parametrach finalnego produktu, czy partia została zaakceptowana, czy nie itd.

Dane: Arkusz1 (11 zm. \* 110 prz.)

	Czas	Temperatura wewnętrzna 1	Temperatura wewnętrzna 2	Temperatura medium	Ciśnienie 1	Ciśnienie 2	Strumień 1	Strumień 2
--	------	--------------------------	--------------------------	--------------------	-------------	-------------	------------	------------

Dane: Arkusz2 (11 zm. \* 110 prz.)

	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	

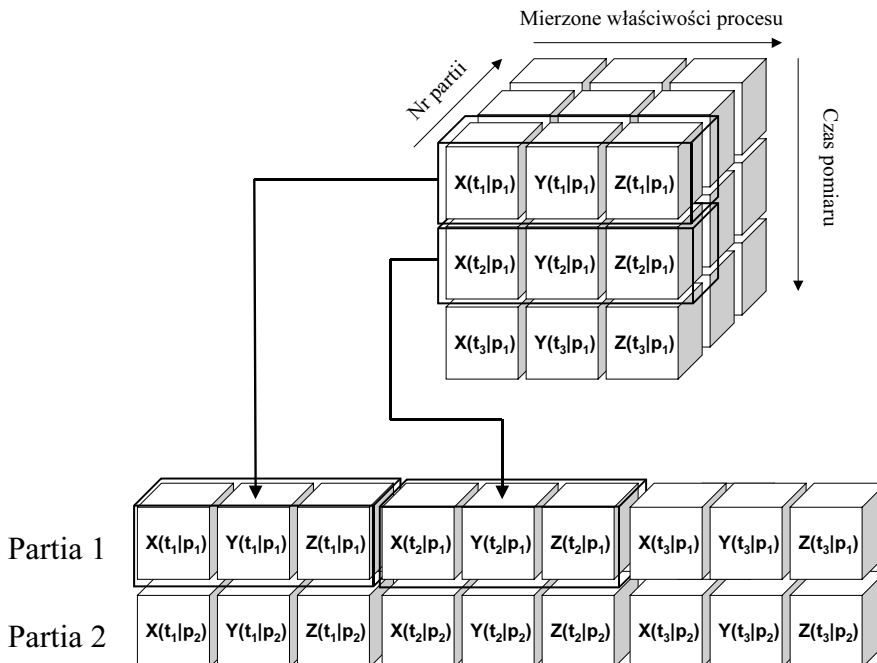
Dane: Arkusz3 (11 zm. \* 110 prz.)

	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

Taki układ danych nie nadaje się do typowej analizy statystycznej, w której obiektom podlegającym analizie odpowiadają przypadki (wiersze tabeli), a badane właściwości obiektów zapisywane są w zmiennych (kolumnach tabeli). Przed analizą musimy w odpowiedni sposób przekształcić dane. Kluczową sprawą jest określenie, co będziemy traktowali jako obiekty, a co jako zmienne (badane cechy obiektów). W monitorowaniu i sterowaniu procesem wsadowym zazwyczaj obiektami są poszczególne partie.

Co będzie właściwością partii? Na pewno będą to wspomniane wcześniej ogólne cechy partii (np. czy spełnia ona wymogi jakościowe). Poza tym partię będą charakteryzowały wartości parametrów procesu, mierzone w toku procesu. Innymi słowy zmiennymi będą kolejne pomiary wszystkich parametrów procesu dla badanej partii.

Wymagane przekształcenie danych przebiega następująco: najpierw tworzymy tabelę (macierz dwuwymiarową), w której każdej partii odpowiada jeden wiersz. W kolejnym kroku bierzemy wartości właściwości procesu dla danej partii zmierzone w pewnym momencie i umieszczamy je jako wartości w kolejnych kolumnach w wierszu odpowiadającym danej partii. Czynności te powtarzamy dla wszystkich czasów pomiaru. Sposób tworzenia tabeli danych przedstawiony jest na rysunku poniżej. 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 stosowania SPC dla procesów wsadowych jest wielka liczba otrzymanych właściwości procesu. Zauważmy, że opisane powyżej przekształcenie danych prowadzi do uzyskania ogromnej liczby zmiennych, które powinniśmy uwzględniać przy monitorowaniu procesu. 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. Trudno sobie wyobrazić praktyczne stosowanie tysięcy jednowymiarowych 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 [7]).

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

Podejście polegające na połączeniu techniki redukcji danych (PCA lub PLS) z przekształceniem danych przedstawionym powyżej nosi nazwę *Multiway Multivariate Statistical Process Control*, a bardziej szczegółowo *Multiway PCA* (w przypadku zastosowania PCA) i *Multiway PLS*.

Zwróćmy uwagę na jeszcze jeden ważny problem. Otóż 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 powinniśmy wykonać jeszcze jedno przekształcenie: podzielić odchylenia od średniej przez odchylenie standardowe zmiennej wejściowej. 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.).

## Monitorowanie procesu

Przed określeniem kryterium sygnalizacji o rozregulowaniu procesu powinniśmy przeanalizować dane historyczne, w celu podziału partii na typowe i nietypowe oraz sprawdzenia, na ile nasze podejście jest w stanie opisać proces (por. [1], [2] i [3]) oraz wykrywać rozregulowania. Jako kryterium, że dana partia jest typowa, możemy przyjąć tylko

zaklasyfikowania jej na podstawie właściwości produktu ([2] i [4]) albo uwzględnić również wyniki analizy danych wskazujących, że pewna partia była nietypowa ([1]).

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.

Do wprowadzenia wskaźników jakości potrzebujemy kilku wzorów, a więc przedstawimy teraz odrobinę matematyki.

Wartość zmiennej  $x_i$  dla partii numer  $j$  oznaczmy jako  $x_i(j)$ . Oryginalne wartości zmiennych z wartościami składowych głównych ( $t$ ) łączy następująca relacja:

$$x_i(j) = \sum_{s=1}^S t_s(j) p_{si} + e_i(j) \quad (1)$$

W powyższym wzorze  $t_s(j)$  oznacza wartość składowej głównej numer  $s$  dla partii numer  $j$ ,  $p_{si}$  to ładunek czynnikowy dla składowej  $s$  i zmiennej  $i$ ,  $e_i(j)$  to reszta (rozbieżność między modelem a wartościami obserwowanymi) dla zmiennej  $i$  oraz partii  $j$ .  $S$  oznacza liczbę użytych składowych głównych.

W bardziej formalnym zapisie macierzowym dane  $\mathbf{X}$  przedstawiamy jako:

$$\mathbf{X} = \sum_{s=1}^S \mathbf{t}_s \mathbf{p}'_s + \mathbf{E} \quad (2)$$

Gdzie  $\mathbf{t}_s$  jest wektorem wartości składowej głównej nr  $s$  dla poszczególnych partii,  $\mathbf{p}_s$  jest wektorem ładunków dla składowej  $s$  (' oznacza transponowanie tego wektora, czyli zamianę z kolumny na wiersz),  $\mathbf{E}$  oznacza macierz błędu (odchyłeń wartości obserwowanych i wynikających z modelu), a  $S$  to liczba wyodrębnionych składowych.

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 formuluje 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; więcej informacji o doborze liczby składowych znajduje się w pracach [1] i [7].

Do stwierdzenia, które partie są typowe, a które nie, używamy wykresów rozrzutu dla składowych głównych, ponieważ można się spodziewać, iż nietypowe partie będą odstawać od innych. Zauważmy, że dla normalnych partii wektory  $\mathbf{t}_s$  powinny mieć wielowymiarowy rozkład normalny. Wizualną kontrolę możemy przeprowadzić za pomocą wykresów rozrzutu lub karty kontrolnej Hottelina ([6]).

Wróćmy na chwilę do wzoru (1). Naturalnym kryterium rozregulowania procesu wydaje się być różnica między wartością wynikającą z PCA a faktycznie obserwowaną. Ponieważ



odchylenia wystąpią dla wszystkich zmiennych, to monitorujemy sumę kwadratów odchyłeń wartości obserwowanych od wynikających z modelu (dla wszystkich zmiennych):

$$Q(j) = \sum_{i=1}^N e_i^2(j) \quad (3)$$

Suma kwadratów odchyłeń  $Q(j)$  od wartości przewidywanych na podstawie modelu będzie miała rozkład  $\chi^2$  (zob. [1] i [3]), i bazując na tym rozkładzie można wyznaczyć granice kontrolne.

Po wstępnej selekcji partii i wybraniu typowych przeprowadzamy właściwą analizę MPCA, tzn. wyznaczamy transformację (1) i stosujemy ją dla nowych partii, tworząc karty kontrolne Hottelina i dla sum kwadratów odchyłeń  $Q(j)$ . Czasami dobiera się jeszcze odpowiednie położenie granic kontrolnych na kartach, tak aby wszystkie partie uznane za prawidłowe mieściły się w obrębie granic kontrolnych (zob. [2] i [3]).

Powyższą procedurę stosujemy, jeśli monitorowanie ma dotyczyć zakończonych partii (tzn. mamy pomiary właściwości procesu dla wszystkich czasów). Bardzo często będziemy jednak chcieli prowadzić monitorowanie w toku procesu dla konkretnej partii. Stosujemy wtedy jeden z czterech schematów uzupełniania danych:

1. Przyjmujemy, że brakujące obserwacje (przyszłe pomiary) będą równe średniej wartości dla typowych partii. Innymi słowy macierz  $X$  dla kolejnych pomiarów wypełniamy zerami.
2. Przyjmujemy, że odchylenie od średniej w przyszłości będzie równe odchyleniu obserwowanemu dla ostatniej serii pomiarów (tzn. wypełniamy puste miejsca macierzy  $X$  ostatnimi pomiarami właściwości procesu).
3. Łączymy podejścia 1 i 2, zakładając, że odchylenia będą spadały od bieżących wartości do zera w pewien sposób (np. liniowo lub wykładniczo).
4. Korzystając z wyników PCA dla typowych partii, przewidujemy wartości przyszłych obserwacji.

Każda z tych metod ma zalety i wady. Podejście nr 4 wydaje się być najlepszym, jeśli dysponujemy co najmniej 1/10 wszystkich pomiarów. Dokładną analizę i porównanie metod uzupełniania danych można znaleźć w pracy [1].

## Modelowanie cech procesu wsadowego

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.

Model PLS dla macierzy zmiennych niezależnych  $\mathbf{X}$  i macierzy zmiennych zależnych  $\mathbf{Y}$  wygląda następująco:

$$\mathbf{X} = \sum_{s=1}^S \mathbf{t}_s \mathbf{p}'_s + \mathbf{E}_X \quad (4)$$

$$\mathbf{Y} = \sum_{s=1}^S \mathbf{u}_s \mathbf{q}'_s + \mathbf{E}_Y \quad (5)$$

Macierze  $\mathbf{E}_X$  i  $\mathbf{E}_Y$  to macierze błędów,  $S$  to liczba składowych (taka sama dla  $\mathbf{X}$  i  $\mathbf{Y}$ ), a  $\mathbf{t}_s$  i  $\mathbf{u}_s$  to składowe główne.

Wektory  $\mathbf{t}_s$  i  $\mathbf{u}_s$  wyznaczone są tak, aby uzyskać największą kowariancję dla każdej ich pary. Dla każdej pary składowych wykonuje się regresję liniową w celu wyznaczenia współczynników regresji  $a_k$ :

$$\mathbf{u}_k = a_k \mathbf{t}_k + \mathbf{e}_k \quad (6)$$

Przy budowie modelu nie musimy ograniczać się wyłącznie do PLS (który jest modelem liniowym). Możemy skorzystać z innych technik, które są w stanie wychwycić zależności nieliniowe, np. sieci neuronowych (przykład zastosowania sieci RBF przedstawiono w pracy [3]), drzew klasyfikacyjnych lub innych technik predykcyjnego *data mining*. Zazwyczaj przed zastosowaniem tych technik również będziemy musieli zredukować liczbę wymiarów.

## Przykład

Przedstawimy teraz przykład stosowania MPCA dla rzeczywistych danych o przebiegu procesów wsadowych. Do dyspozycji mamy informacje o produkcji 55 partii i o jakości wytworzonej substancji<sup>10</sup>.

W toku wytwarzania jednej partii mierzono 10 właściwości procesu. Łącznie wykonano po 100 pomiarów każdej właściwości. Dwie właściwości od pewnego momentu przyjmują stałą wartość dla wszystkich partii - zostały one pominięte w analizie.

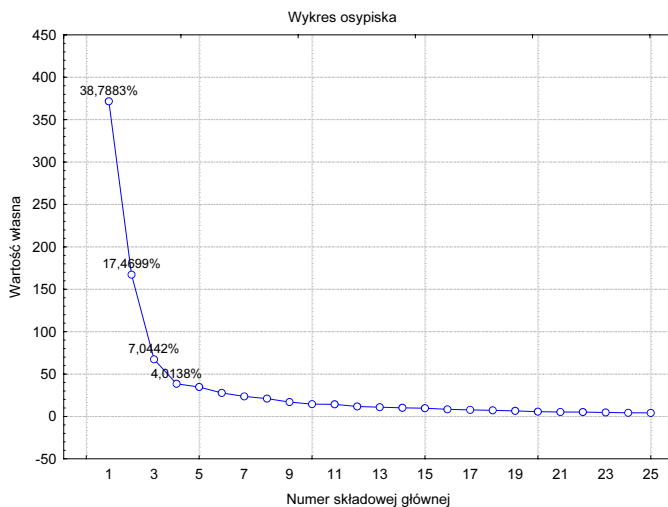
W analizie pominiemy dane o dwóch partiach: jedna z nich przeszła testy jakości, a druga nie. Informacje te wykorzystamy do sprawdzenia, jak uzyskany model działa dla nowych danych.

<sup>10</sup> W przykładzie wykorzystano nieco przekształcone dane omawiane w [1]. Za udostępnienie danych dziękujemy prof. Johnowi F. Macgregorowi.

Jak wspomnieliśmy wcześniej, do budowy modelu typowych partii możemy wykorzystać dane o wszystkich procesach, które dały produkt spełniający specyfikację (zob. [2] i [3]). Zastosujemy jednak inne, bardziej restrykcyjne podejście, takie jak w pracy [1]: najpierw zbudujemy model dla wszystkich obserwacji i wykryjemy obserwacje nietypowe. W ostatecznej analizie (modelu referencyjnym) pominiemy nie tylko wadliwe partie, ale również partie, dla których proces przebiegał nietypowo.

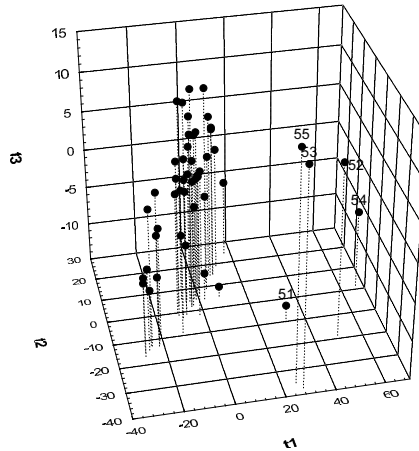
Ponieważ dysponujemy danymi przekształconymi do odpowiedniej postaci, do budowy modelu wykorzystamy moduł *NIPALS* pakietu *STATISTICA*. (Na surowych danych analizy *MPCA* i *MPLS* wykonuje pakiet *STATISTICA MSPC*).

Na początku analizy musimy odpowiedzieć na pytanie, ile składowych głównych należy wyodrębnić. Jako wstępne kryterium zastosujemy warunek, aby wartości własne odpowiadające składowym głównym były większe od 1 (interpretujemy je w ten sposób, że każda składowa główna przynosi więcej informacji niż poszczególne zmienne). Kryterium to prowadzi do uzyskania rozwiązania z 25 składowymi głównymi. Zobaczmy, jak dużo zmienności wyjaśniają kolejne składowe główne.

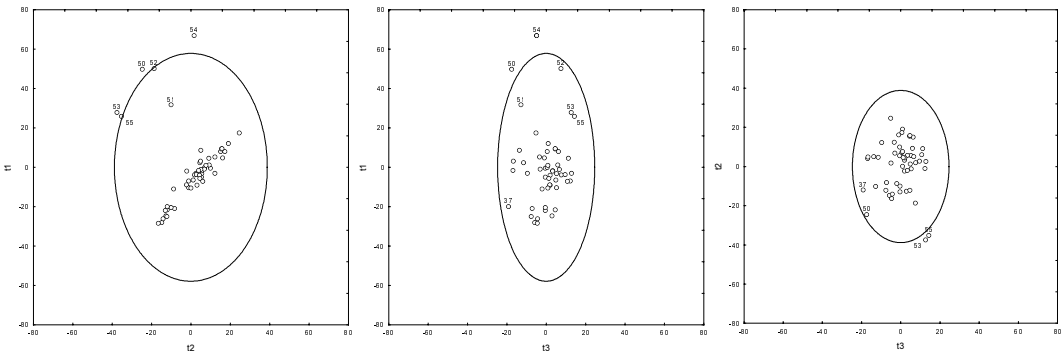


Powyższy wykres nosi nawet wykresu osypiska. Widać na nim wyraźnie, że zdecydowanie najwięcej zmienności wyjaśniają pierwsza i druga składowa (łącznie ok. 56%), trzecia składowa przenosi 7% zmienności, a począwszy od czwartej mamy tzw. „osypisko”: ciąg składowych głównych, które wyjaśniają podobną, niewielką część zmienności oryginalnych danych. Jednym z kryteriów doboru liczby wyodrębnianych składowych jest wybranie takiej liczby składowych, po której wykres osypiska się wypłaszcza (zaczyna się osypisko). W naszym przypadku odpowiednią liczbą będzie 3 (aczkolwiek można by zastanowić się nad przyjęciem liczby 2).

Po wykonaniu analizy *MPCA* dla 3 składowych głównych przyjrzymy się rozkładowi punktów w przestrzeni składowych głównych. Na poniższym rysunku bardzo wyraźnie widać 5 odstających punktów. Odpowiadają one partiom od 51 do 55.



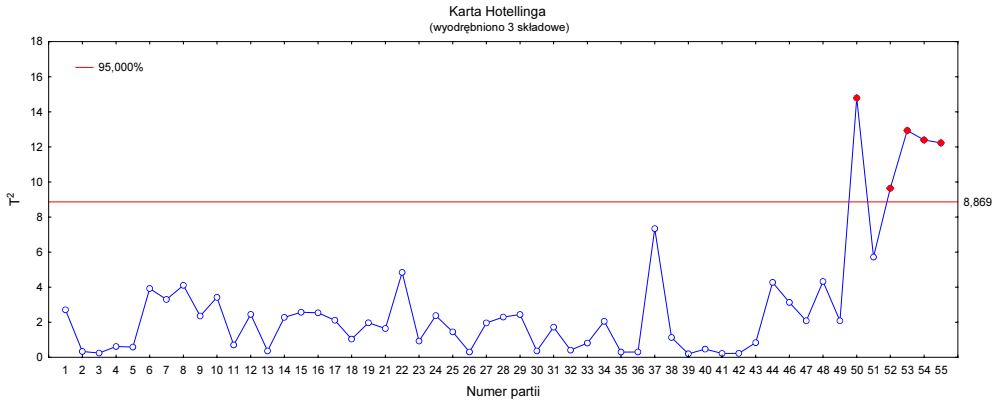
Dokładniejszy wgląd uzyskamy za pomocą dwuwymiarowych diagramów korelacyjnych dla par składowych głównych:



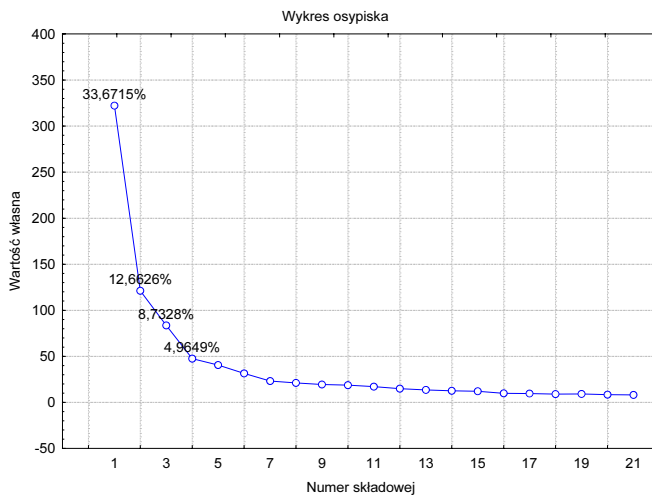
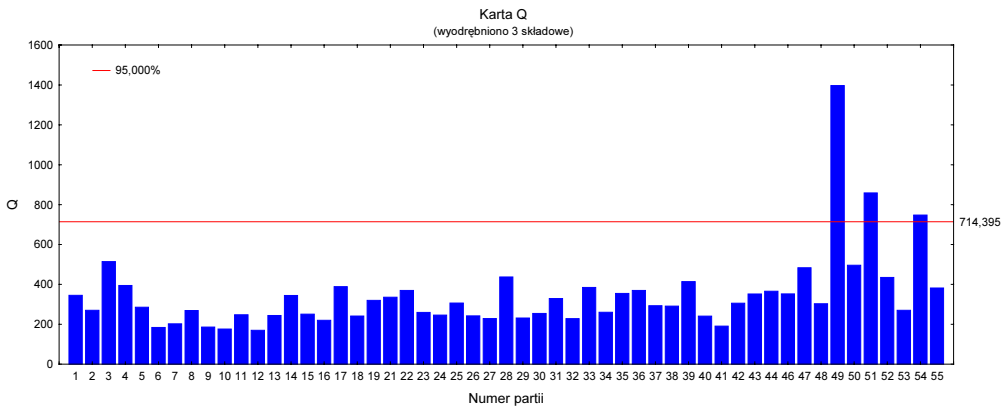
Na powyższych wykresach widzimy, że nietypowe są również partie 50 (odstaje na wszystkich wykresach) i 37 (odstaje na wykresach dla składowej 1 i 3 oraz 2 i 3). Zwróćmy uwagę, że na tym etapie analizy elipsa odpowiadająca trzem odchyleniom standardowym jest tylko pewnym punktem odniesienia, ważniejszy jest układ punktów i to, czy nie tworzą one wyraźnie oddzielonych „wysp”.

Warto zauważyć, że na wykresie dla dwóch pierwszych składowych wyraźnie widać, że po usunięciu odstających przypadków, składowe pierwsza i drugą będą skorelowane. Wynika z tego, że jeśli pominiemy te punkty w analizie, uzyskamy wyraźnie inne składowe główne (uzyskiwane za pomocą PCA składowe główne są z definicji nieskorelowane).

Zobaczmy teraz na karty kontrolne: Hotellinga (wielowymiarowa karta dla składowych głównych) i sumy kwadratów odchyłeń (Q).



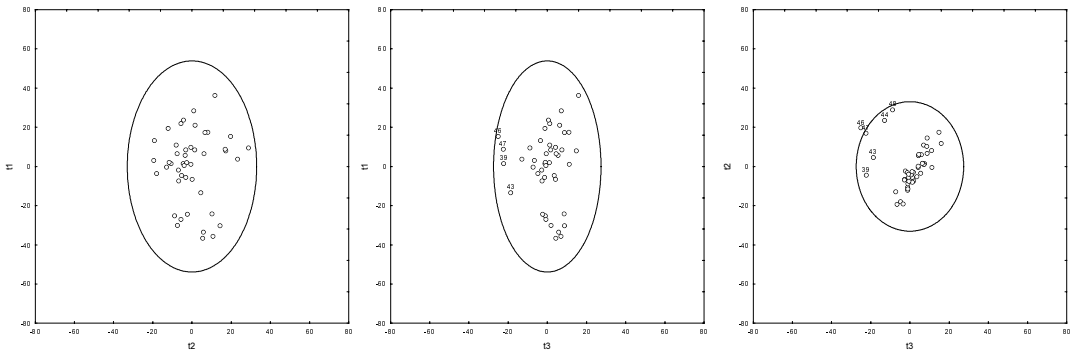
Na karcie Hotellinga sygnały o rozregulowaniu występują dla partii numer 50, od 52 do 55. Ponadto partia 37 jest stosunkowo blisko granicy kontrolnej. Natomiast karta Q bardzo silnie sygnalizuje rozregulowanie dla 49 partii, poza tym mamy sygnały dla partii 51 i 54.





Po analizie powyższych wykresów za nietypowe uznamy partie 37 i od 49 do 56. Liczbę składowych dobieramy tak samo jak w poprzedniej analizie. Patrząc na wykres osypiska widzimy, że struktura składowych uległa zmianie, różnica w zmienności wyjaśnianej przez składowe 2 i 3 zmniejszyła się. Jednak i tym razem możemy przyjąć liczbę wyodrębnianych składowych równą 3.

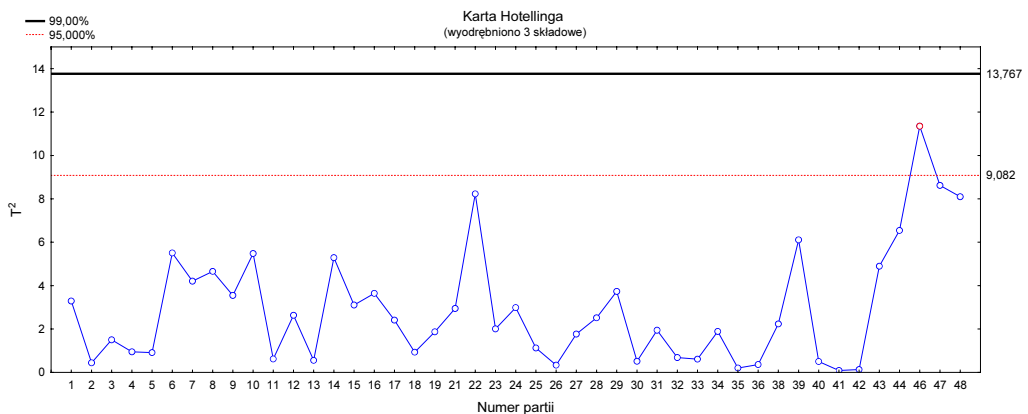
Podobnie jak przy pierwszej analizie do wykrywania odstających obserwacji użyjemy diagramów korelacyjnych dla wartości składowych głównych.



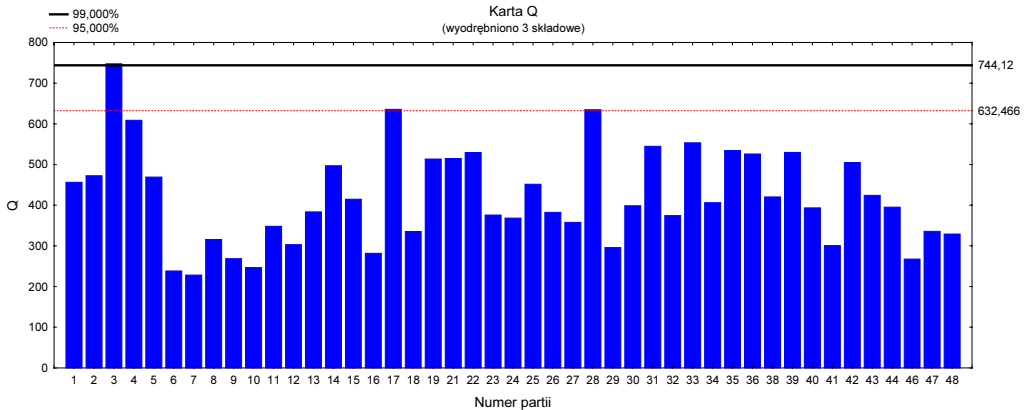
Powinniśmy zwrócić uwagę na następujące partie:

- ◆ Na wykresie pierwszej i trzeciej składowej: 39, 43, 44, 46, 47 i 48
- ◆ Na wykresie drugiej i trzeciej składowej: 39, 43, 44, 46 i 48

Kolejnym sprawdzianem będą karty kontrolne. Tym razem wykreśliły granice kontrolne dla poziomu  $p$  równego 95% i 99%.



Na karcie Hotellinga mamy przekroczenie granicy kontrolnej dla  $p=95\%$  dla partii numer 46, nie przekracza ona jednak granicy kontrolnej dla  $p=99\%$ . Natomiast na karcie Q mamy jedno przekroczenie granicy dla  $p=99\%$  (partia nr 3) i dwa minimalne przekroczenia granicy kontrolnej dla  $p=95\%$ .



Do tej pory nie korzystaliśmy z danych o jakości produktu. Skonfrontujmy je teraz z naszymi dotychczasowymi badaniami. Otóż faktycznie za nieodpowiednie uznano partie 38, 40 do 42, 46, 49 do 55 oraz partia 45, której nie uwzględnialiśmy w analizie, ponieważ chcemy użyć jej do testowania modelu. Spośród błędnych partii jako nietypowe zidentyfikowaliśmy do tej pory partię 46 i partie od 49 do 55. Z drugiej strony wykryliśmy kilka odstających przebiegów procesu, które dały produkt spełniający specyfikację.

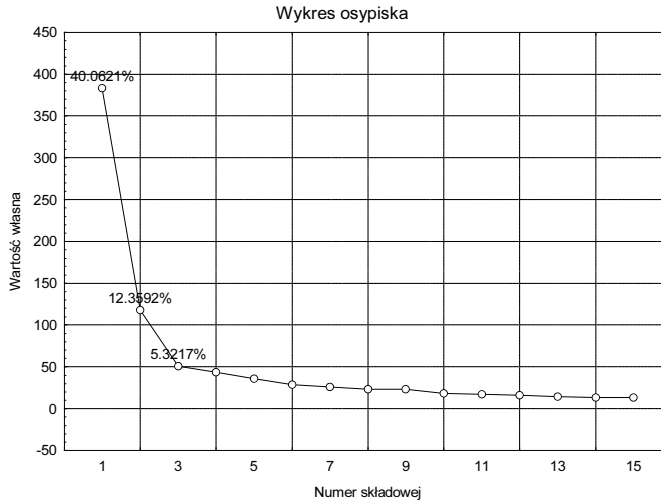
To, że nie wykryliśmy wszystkich rozregulowań nie powinno nas dziwić. Korzystaliśmy przecież tylko z części informacji o procesie. Przede wszystkim nie uwzględniliśmy parametrów wejściowych substancji, które z pewnością bardzo silnie wpływają na jakość produktu.

Z drugiej strony fałszywe sygnały pojawiają się dla partii sąsiadujących z partiami o nieodpowiedniej jakości. Wyjaśnieniem może być np. to, że operator wiedząc, że coś jest nie tak, „ratował” proces dla kolejnych partii i nietypowo sterował przebiegiem przetwarzania.

Zwróćmy uwagę na jeszcze jedną sprawę – położenie granicy kontrolnej. Na karcie Q mamy 3 sygnały o rozregulowaniu przy  $p=95\%$  i tylko jeden przy  $p=99\%$ . Wszystkie one odpowiadają produktom o zadowalającej jakości. W związku z tym jako właściwy poziom  $p$  przyjmujemy  $99\%$ .

W modelu referencyjnym wykorzystamy dane niezawierające partii niespełniających wymogów jakościowych oraz partii wskazanych jako nietypowe w dotychczasowej analizie. Łącznie wyeliminujemy partie: 3 i od 37 do 55. Co prawda partia numer 3 nie została zdyskwalifikowana przez testy jakości ani nie sąsiaduje z taką partią, jednak została wskazana przez kartę Q, a w wypadku naszej analizy gorzej jest „zepsuć” model referencyjny nietypowymi danymi, niż utracić nieco informacji, pomijając jedną dobrą partię.

Również budując model referencyjny, powinniśmy określić liczbę składowych. Tym razem z wykresu osypiska wynika, iż wystarczy użyć dwóch składowych.



Model referencyjny z dwiema składowymi dobrze opisuje dane o partiach wykorzystanych w jego budowie: nie ma żadnych sygnałów o rozregulowaniu. Zastosujemy go teraz dla „nowych” danych (dwóch procesów, których nie braliśmy pod uwagę przy tworzeniu modelu referencyjnego).

Moduł *STATISTICA* NIPALS umożliwi zapisanie modelu PCA w języku PMML i potem zastosowanie go dla nowych danych (generowany jest także kod w języku C do stosowania w zewnętrznych aplikacjach). Po zastosowaniu modelu dla dwóch partii odłożonych do próby testowej uzyskamy następujący wynik:

Partia	Klasyfikacja	T <sup>2</sup>	Granica kontrolna T <sup>2</sup>	Q	Granica kontrolna Q
20	Dobra	0,338	11,4	698	756
45	Zła	0,930	11,4	5318	756

Partia 45 nie spełniała wymogów jakościowych i mamy dla niej bardzo silny sygnał o rozregulowaniu na karcie Q: wartość tej statystyki wynosi 5318 przy granicy kontrolnej równej 756.

## Literatura

- [1]. Nomikos P., MacGregor J.F., “Multivariate SPC Charts for Monitoring Batch Process”, Technometrics 37(1), 41-59.
- [2]. Lennox B., Hiden H.G., Montague G.A., Kornfeld G., Goulding P.R., “Application of multivariate statistical process control to batch operations”, Computers & Chemical Engineering 24, 291-296.



- [3]. Lennox B., Montague G.A., Hiden H.G., Kornfeld G., Goulding P.R., “Process monitoring of an Industrial Fed-batch Fermentation”, *Biotechnology & Bioengineering* 74(2), 125-135.
- [4]. van Sprang E.N.M, Ramaker H-J, Gurden S.P., Westerhuis S.P., Smilde A.K, Poster dostępny na stronie Katedry Inżynierii Chemicznej Uniwersytetu w Amsterdamie [www-its.chem.uva.nl](http://www-its.chem.uva.nl).
- [5]. Greber T., “Statystyczne sterowanie procesami – doskonalenie jakości z pakietem *STATISTICA*”, StatSoft Polska 2000.
- [6]. Montgomery D. C., “Introduction to Statistical Quality Control”, wyd. III, John Wiley & Sons, Inc.
- [7]. Mazerski J., “Podstawy chemometrii”, Wydawnictwo Politechniki Gdańskiej 2000.