

PROTOKÓŁ NR 2/2016/7
Z POSIEDZENIA GRUPY EKSPERCKIEJ DS. METOD FIZYKOCHEMICZNYCH
KOMISJI FARMAKOPEI
W DNIU 18 LUTEGO 2016 R.

Porządek obrad posiedzenia:

1. Otwarcie posiedzenia.
2. Przyjęcie porządku obrad posiedzenia.
3. Przyjęcie protokołu nr 1/2016/6 z posiedzenia Grupy eksperckiej ds. Metod Fizykochemicznych Komisji Farmakopei w dniu 21 stycznia 2016 r.
4. Omówienie i weryfikacja zgodności z tekstem Suplementu 8.7 Farmakopei Europejskiej dalszej części polskojęzycznej wersji nowego tekstu podstawowego 5.21. *Metody chemometryczne stosowane do danych analitycznych* (od części 2-3-4), przeznaczonego do zamieszczenia w Suplemencie 2016 do Farmakopei Polskiej wydanie X.
5. Wolne wnioski.

Obecni na posiedzeniu członkowie Grupy eksperckiej ds. Metod Fizykochemicznych Komisji Farmakopei:

Przewodniczący	- prof. dr hab. Zbigniew Fijałek
Zastępca Przewodniczącego	- prof. nadzw. dr hab. Anna Gumieniczek
Członkowie:	- prof. dr hab. Tomasz Bączek
	- prof. nadzw. dr hab. Jan Maurin

Nieobecni na posiedzeniu członkowie Grupy eksperckiej ds. Metod Fizykochemicznych Komisji Farmakopei:

prof. dr hab. Zenon Kokot, prof. dr hab. Jerzy Łukasiak

Obecni na posiedzeniu pracownicy Urzędu Rejestracji Produktów Leczniczych, Wyrobów Medycznych i Produktów Biobójczych:

Dyrektor Departamentu Farmakopei	- Ewa Leciejewicz-Ziemecka
Departament Farmakopei	- Elżbieta Sadowska

Omówienie przebiegu posiedzenia:

Ad 1) Posiedzenie otworzył Przewodniczący Grupy eksperckiej prof. dr hab. Zbigniew Fijałek witając przybyłych. W imieniu Urzędu zebranych powitała Dyrektor Departamentu Farmakopei dr Ewa Leciejewicz-Ziemecka.

Ad 2) Porządek obrad posiedzenia przyjęto bez zastrzeżeń.

Ad 3) Protokół nr 1/2016/6 z posiedzenia Grupy eksperckiej ds. Metod Fizykochemicznych Komisji Farmakopei w dniu 21 stycznia 2016 r. przyjęto jednogłośnie.

Ad 4) Na niniejszym posiedzeniu kontynuowano omawianie polskojęzycznej wersji nowego tekstu podstawowego opublikowanego w Suplemencie 8.7 Farmakopei Europejskiej 5.21. *Chemometric methods applied to analytical data (Metody chemometryczne stosowane do danych analitycznych)* (od części 2-3-4). W trakcie weryfikacji ww. tekstu stosowane były także ustalenia podjęte na wcześniejszych posiedzeniach (8.12.2015 r. i 21.01.2016 r.); tekst

omówiono szczegółowo do części 2-8-4 (*Potencjalne zastosowania*) włącznie. Ustalenia zostaną wprowadzone w Departamencie Farmakopei.

USTALENIA OGÓLNE

Ustalono ostateczne tłumaczenie terminu *scores* jako wyniki.

USTALENIA

Str. 27, wiersz 17 powinno być: „Potencjalne zastosowania”.

Str. 27, wiersz 19 powinno być: „spektroskopia w bliskiej podczerwieni”.

Str. 27, wiersz 20 powinno być: „SIMCA jest bardziej odpowiednia niż PCA”.

Str. 28, wiersz 4 powinno być: „Klaster (grupa) składa się z grupy obiektów lub punktów danych”.

Str. 28, wiersz 12 powinno być: „gdzie obiekt nie przypisany zostaje przyporządkowany do grupy”.

Str. 28, wiersz 15 i cała monografia powinno być: „Podstawy”.

Str. 28, wiersz 18 powinno być: „graficznym przedstawieniem danych, a niehierarchiczne grupowanie identyfikuje klastry”.

Str. 28, wiersz 22-24 powinno być: „Grupowanie hierarchiczne rekursywnie identyfikuje klastry w trybie aglomeracyjnym (podejście ‘wstępujące’ (*bottom-up*)) lub podziałowym (podejście ‘zstępujące’ (*top-down*)), aby zbudować strukturę drzewa (dendrogram). Tryb aglomeracyjny”.

Str. 28, wiersz 26-28 powinno być: „zostanie sklasyfikowany (ryc. 5.21.-7). Tryb podziałowy zakłada na początku, że cały zbiór danych to jeden klaster, który następnie jest dzielony rekursywnie dopóki uzyska się klastry zawierające tylko jeden punkt danych.”.

Str. 29, wiersz 2 powinno być: „odległości pomiędzy dwoma obiektami”.

Str. 29, wiersz 5 powinno być: „dwóch najbardziej podobnych klastrów”.

Str. 29, wiersz 6 powinno być: „Poziom braku podobieństwa”.

Str. 29, wiersz 8-9 powinno być: „*Dendrogram grupowania hierarchicznego w trybie aglomeracyjnym do uzyskania klastrów zawierających pojedynczy punkt danych*”.

Str. 29, wiersz 17-19 powinno być: „(w metodach ‘opartych na modelach’) i algorytm DBSCAN (*‘density-based’ algorithms*, ‘oparty na gęstości danych’) i metody ‘oparte na siatce’ reprezentowane przez algorytm statystycznej sieci informacyjnej (*statistical information grid*, STING).”.

Str. 30, wiersz 22 powinno być: „oszacowanie największego prawdopodobieństwa”.

Str. 30, wiersz 25 powinno być: „Grupowanie w oparciu o gęstość danych (*density-based*, DB), jak algorytm DBSCAN”.

Str. 30, wiersz 30 powinno być: „dzielią przestrzeń danych na skończoną liczbę komórek”.

Str. 31, wiersz 1-2 powinno być: „komórek podrzędnych z poziomu niższego”.

Str. 31, wiersz 5 powinno być: „Algorytmy są wrażliwe na warunki wyjściowe użyte do zainicjowania grupowania danych”.

Str. 31, wiersz 7-8 powinno być: „Mierniki użyte do obliczenia”.

Str. 31, wiersz 11-13 powinno być: „poprzez wstępne przetwarzanie danych przed ich grupowaniem. Algorytmy DB mogą radzić sobie z klastrami o arbitralnych kształtach, ale ich słabością jest ograniczenie w obchodzeniu się z wielowymiarowymi danymi, gdzie obiekty są luźno rozłożone w przestrzeniach.”.

Str. 31, wiersz 15 powinno być: „algorytmy takie jak grupowanie oparte na gęstości danych”.

Str. 31, wiersz 26 powinno być: „nieprzetworzonych danych w poszukiwaniu połączeń, trendów i zależności pomiędzy zmiennymi”.

Str. 32, wiersz 1 powinno być: „*WIELOWYMIAROWA DEKONWOLUCJA/ROZPLATANIE KRZYWYCH*”.

Str. 32, wiersz 5 i cała monografia powinno być: „profile udziałów (*contribution profiles*)”.

Str. 32, wiersz 6 powinno być: „profile czystych składowych (tj. ładunki MCR)”.

Str. 32, wiersz 8-9 powinno być: „Podczas optymalizacji parametrów MCR wykorzystuje się często algorytm naprzemiennych najmniejszych kwadratów (*alternating least squares, ALS*).”.

Str. 32, wiersz 12 powinno być: „MCR-ALS oszacowuje profile udziałów C i profile czystych składowych S ”.

Str. 32, wiersz 16 powinno być: „na składowe/czynniki. Przykładowo, ani udziały ani absorbancja”.

Str. 32, wiersz 17-19 powinno być: „Ten fakt może być użyty do ekstrakcji profili czystych składowych i ich udziałów w dobrze funkcjonującym zbiorze danych. Istnieją także inne typy ograniczeń, które mogą być zastosowane, np. równość, jednomodalność, zamknięcie i zbilansowanie masy.”.

Str. 32, wiersz 24 powinno być: „uzyskane profile udziałów czystych składowych”.

Str. 32, wiersz 30-31 powinno być: „Do obliczania ALS ważny jest wybór odpowiedniej liczby składowych, aby uzyskać stabilne rozwiązanie i dobre oszacowanie co można osiągnąć używając np. rozwijającej analizy czynnikowej (EFA)”.

Str. 33, wiersz 2-4 powinno być: „rygorystycznie wymagane, a ‘miękkie’ ograniczenia dają możliwość odstępstw od ustalonych wartości”.

Str. 33, wiersz 5-7 powinno być: „muszą być wyrażone jako np. stężenie substancji czynnej, używając prostej regresji liniowej. Oznacza to, że rzeczywista zawartość musi być znana co najmniej dla jednej próbki. Gdy zmienności dwóch lub kilku składników chemicznych”.

Str. 33, wiersz 9-10 powinno być: „składnik gdy inny zanika, lub dwa składniki zanikają w tym samym tempie tworząc trzeci. W wyniku tego, zmienność dla określonej substancji jest istotnie maskowana”.

Str. 33, wiersz 19-21 powinno być: „Jeżeli dane nie spełniają wymagań liniowości i selektywności, do kalibracji może być wymagana większa liczba wzorców na analiz. Jeżeli nie obserwuje się jednoznacznej odpowiedzi sygnału”.

Str. 33, wiersz 23 powinno być: „Zastosowanie MCR z użyciem ALS”.

Str. 33, wiersz 29-31 powinno być: „Wielokrotna regresja liniowa (*multiple linear regression, MLR*) jest klasyczną metodą wielowymiarową, która używa zbiorów zmiennych (wektorów) x (macierz danych X) w postaci liniowych kombinacji tak, aby były one dopasowane tak bardzo jak to tylko możliwe do odpowiadającej pojedynczej zmiennej (wektora) y ”.

Str. 34, wiersz 1 powinno być: „MLR rozszerza liniową regresję do przypadku wyboru więcej niż jednej zmiennej”.

Str. 34, wiersz 13 powinno być: „równania linii prostej”.

Str. 34, wiersz 20 powinno być: „który najlepiej minimalizuje”.

Str. 34, wiersz 22 powinno być: „reszty y członu błędu f ”.

Str. 35, wiersz 1-2 powinno być: „MLR, jeżeli istnieje, nie będzie stabilny.”.

Str. 35, wiersz 15 powinno być: „następujące ograniczenia i właściwości”.

Str. 35, wiersz 18 powinno być: „występowania współliniowości pomiędzy zmiennymi X ”.

Str. 35, wiersz 28 powinno być: „wysoki stopień specyficzności i pełny ich rząd”.

Str. 36, wiersz 5 powinno być: „REGRESJA GŁÓWNYCH SKŁADOWYCH”.

Str. 36, wiersz 8 powinno być: „macierz kalibracji X ”.

Str. 36, wiersz 9-10 powinno być: „macierzy wyników \hat{T} i ładunków \hat{P} ”.

Str. 36, wiersz 14 powinno być: „Jak w przypadku PCA, macierz kalibracji zostaje rozłożona do macierzy wyników”.

Str. 36, wiersz 19 powinno być: „może być optymalnie korelowana stosując wielokrotną regresję liniową”.

Str. 37, wiersz 13 powinno być: „liczba wartości dla próbki”.

Str. 37, wiersz 16 powinno być: „przewidywana wartość dla nieznannej próbki”.

Str. 37, wiersz 19 powinno być: „*Rozkład macierzy w regresji głównych składowych (PCR)*”.

Str. 38, wiersz 2-3 powinno być: „względem wariancji reszt Y . W większości przypadków, dodatkowe PC poza ten punkt, nie poprawiają działania predykcyjnego, lecz model kalibracji staje się nadmiernie dopasowany.

Pomimo jego korzyści/zalety jako ważnego narzędzia radzenia sobie z współliniowością/wzajemną korelacją danych X , słabość PCR polega na jego niezależnym rozkładzie macierzy X i Y . Podejście to może brać pod uwagę zmienność danych X , co niekoniecznie jest istotne/odpowiednie dla optymalnej regresji danych Y ”.

Str. 38, wiersz 5-6 powinno być: „Etapowy wybór głównych składowych (np. wybór PC2 zamiast PC1) może okazać się przydatny, do poprawy działania modelu kalibracyjnego”.

Str. 38, wiersz 8 powinno być: „Potencjalne zastosowania”.

Str. 38, wiersz 9 powinno być: „PCR jest wielowymiarową techniką z dużą liczbą narzędzi diagnostycznych”.

Str. 38, wiersz 10 powinno być: „detekcję błędnych pomiarów”.

Str. 38, wiersz 11-12 powinno być: „stabilnych rozwiązań dla danych kalibracyjnych w postaci kompletnych widm lub ich obszernych zakresów/obszarów widma”.

Str. 38, wiersz 14 powinno być: „omawiane powyżej ograniczenia oraz wady metody”.

Str. 38, wiersz 17 powinno być: „**METODA CZĄSTKOWYCH NAJMNIEJSZYCH KWADRATÓW**”.

Str. 38, wiersz 21 powinno być: „algorytmem wielowymiarowej regresji”.

Str. 38, wiersz 24 powinno być: „kowariancję pomiędzy blokami”.

Str. 38, wiersz 26 powinno być: „zostaje wykorzystana do odszukania głównych składowych”.

Str. 39, wiersz 3-4 powinno być: „jednoczesny rozkład macierzy X i Y aby wytworzyć składowe (najlepiej zwane czynnikami PLS, czynnikami lub ukrytymi zmiennymi)”.

Str. 39, wiersz 7 powinno być: „przewidywania wartości Y ”.

Str. 39, wiersz 12-14 powinno być: „elementami w procesie iteracji. Podczas gdy macierz Y jest rozkładana do macierzy ładunków i wyników, odpowiednio \hat{Q} i \hat{U} , rozkład macierzy X ”.

Str. 39, wiersz 17-18 powinno być: „Do połączenia macierzy Y z rozkładem macierzy X , do pierwszego oszacowania wartości wyników \hat{T} , wykorzystuje się dane Y aby prowadzić rozkład macierzy X ”.

Str. 39, wiersz 22 powinno być: „wariancje reszt wykorzystuje się podczas walidacji do ustalenia optymalnej liczby czynników”.

Str. 40, wiersz 14 powinno być: „liczba wartości dla próbki”.

Str. 40, wiersz 17 powinno być: „przewidywana wartość dla nieznannej próbki”.

Str. 40, wiersz 20 powinno być: „*Rozkład macierzy danych w regresji PLS*”.

Str. 41, wiersz 2-3 powinno być: „Krytycznym etapem w PLS jest wybór liczby czynników. Wybór zbyt małej liczby czynników spowoduje nieodpowiednie wyjaśnienie zmienności uczącego zbioru danych”.

Str. 41, wiersz 4-5 powinno być: „niestabilności wynikającej z kalibracji (ryc. 5.21.-10). Optymalna liczba czynników jest szacowana podczas walidacji kalibracji. Rycina 5.21.-10 ukazuje zmiany błędu kalibracji”.

Str. 41, wiersz 7-10 powinno być: „W przypadku B błąd przewidywania ujawnia brak minimum; jednak minimum obserwuje się w przypadku C. W sytuacji braku minimum, liczbę składowych można ustalić jeżeli nie obserwuje się istotnych spadków różnic/błędu”.

Str. 41, wiersz 13-14 powinno być: „dla każdej zmiennej Y . Zwykle, PLS2 jest preferowanym podejściem w celach przesiewowych”.

Str. 41, wiersz 19-22 powinno być: „*Efekt dodawania czynników do modelu: kalibracja staje się coraz bardziej dokładna, tj. zmniejsza się błąd resztowy, podczas gdy działanie*

predykcijne modelu może się pogorszyć, jak dla przypadku C. Optymalna liczba czynników jest kompromisem”.

Str. 42, wiersz 3-6 powinno być: „PLS stał się preferowaną alternatywą ilościowej kalibracji dla PCR, gdyż włącza wpływ struktury danych Y na rozkład macierzy kalibracji X . W konsekwencji, zostaje zgromadzona informacja o najbardziej ważnych/istotnych czynnikach i możliwy jest opis maksymalnej zmienności danych X ”.

Str. 42, wiersz 8-10 powinno być: „zapewnia wizualizację w celu optymalizacji działania kalibracji. Dodatkowo, PLS jest skuteczna w obecności szumu”.

Str. 42, wiersz 12-13 powinno być: „poddaje się regresji macierz X do transformacji w macierz Y , złożoną z zer i jedynek. Jedyneki i zera wskazują klasę”.

Str. 42, wiersz 14-15 powinno być: „do oszacowania układu pikseli”.

Ad 5) Ustalono termin kolejnego posiedzenia Grupy, na którym omawiana będzie dalsza część tekstu 5.21., na dzień 8 marca 2016 r.

Na zakończenie posiedzenia Przewodniczący Grupy eksperckiej prof. dr hab. Zbigniew Fijałek oraz Dyrektor Departamentu Farmakopei dr Ewa Leciejewicz-Ziemecka podziękowali zebranych za przybycie i merytoryczną dyskusję.

*Przewodniczący
Grupy eksperckiej ds. Metod
Fizykochemicznych KF*


Prof. dr hab. Zbigniew Fijałek