

PROTOKÓŁ NR 1/2022/15
Z POSIEDZENIA GRUPY EKSPERCKIEJ DS. METOD FIZYKOCHEMICZNYCH
KOMISJI FARMAKOPEI
W DNIU 14 LISTOPADA 2022 R.

Porządek obrad posiedzenia (wideokonferencja):

1. Otwarcie posiedzenia.
2. Przyjęcie porządku obrad posiedzenia.
3. Przyjęcie protokołu nr 2/2021/14 z posiedzenia Grupy eksperckiej ds. Metod Fizykochemicznych Komisji Farmakopei w dniu 29 listopada 2021 r.
4. Omówienie i weryfikacja zgodności z tekstem Farmakopei Europejskiej polskojęzycznej wersji znowelizowanego^{II} tekstu opublikowanego w Farmakopei Europejskiej 11.1, przeznaczonego do zamieszczenia w części podstawowej Farmakopei Polskiej wydanie XIII (FP XIII 2023).
5.21. Metody chemometryczne stosowane do danych analitycznych ^{II (11.1)}
5. Uchwała Grupy eksperckiej ds. Metod Fizykochemicznych Komisji Farmakopei w sprawie rozdziału *5.21. Metody chemometryczne stosowane do danych analitycznych*.
6. Wolne wnioski.

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

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

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 i Dyrektor Departamentu Farmakopei dr Ewa Leciejewicz-Ziemecka, witając zebranych na spotkaniu Grupy eksperckiej.

Ad 2) Porządek obrad posiedzenia przyjęto bez zastrzeżeń.

Ad 3) Protokół nr 2/2021/14 z posiedzenia Grupy eksperckiej ds. Metod Fizykochemicznych Komisji Farmakopei w dniu 29 listopada 2021 r. przyjęto jednogłośnie.

Ad 4) Na niniejszym posiedzeniu omawiano polskojęzyczną wersję znowelizowanego tekstu podstawowego *5.21. Metody chemometryczne stosowane do danych analitycznych*. Wprowadzony do Farmakopei Europejskiej (Ph. Eur.) w roku 2016 rozdział *5.21. Metody chemometryczne stosowane do danych analitycznych* został w Suplemencie 11.1 Ph. Eur. poddany przez Komisję Farmakopei Europejskiej (KFEur) procesowi pełnej nowelizacji w oparciu o aktualny stan wiedzy w tej dziedzinie. Zmieniony tekst *5.21* zostanie zamieszczony

w części podstawowej Farmakopei Polskiej wydanie XIII (FP XIII 2023). Publikacja FP XIII 2023 planowana jest w listopadzie 2023 r. wraz z wersją elektroniczną.

Dyrektor DF podkreśliła, że w celu usprawnienia przebiegu posiedzenia w formie wideokonferencji załączony do zaproszenia na posiedzenie materiał zawierał wprowadzone przez Departament Farmakopei wstępne weryfikacje zgodności z wersjami oryginalnymi i ustaleniami redakcyjnymi oraz sugestie zapisów (zgodne z poprzednią wersją rozdziału 5.21 oraz w tekstami powiązаныmi 5.24. *Obrazowanie chemiczne* i 5.28. *Wielowymiarowa statystyczna kontrola procesów*, 5.3. *Analiza statystyczna wyników biologicznych oznaczeń zawartości i badań jakościowych*, 5.25. *Technologia analizy procesu*). Następnie Członkowie Grupy przesłali swoje uwagi przed terminem posiedzenia i na niniejszym spotkaniu Dyrektor DF przedstawiła w formie prezentacji zebrane uwagi do tekstu. Przyjęte zmiany, po posiedzeniu wprowadzi Departament Farmakopei. Na niniejszym posiedzeniu omówiono tekst do części 2-11.

USTALENIA OGÓLNE

bias – systematyczny błąd

contibution – wkład

cosine similarity – kosinus podobieństwa

extraction – wydobywanie

ill-conditioned matrix – źle określona macierz

least absolute shrinkage and selection operator (LASSO) – operator najmniejszego absolutnego kurczenia i wyboru

leave-one-out – pomiń jeden obiekt

leave-subset-out – pomiń podzbiór

maximum likelihood estimation – oszacowanie metodą największej wiarygodności

observation – obserwacja

original data – wyjściowe dane

principal components (PC) – główne składowe

primary PC – podstawowe PC

secondary PC – drugorzędne PC

sensor – czujnik

source signal – sygnał źródłowy

variance – wariancja

variation – zmienność

USTALENIA SZCZEGÓŁOWE

Str. 1, wiersz 5-6 i 13 powinno być: „wprowadzenie do zastosowania technik chemometrycznych”

Str. 1, wiersz 24-28 powinno być: „Chemometrię definiuje się jako dyscyplinę chemiczną wykorzystującą matematykę, statystykę i logikę formalną (*formal logic*), aby (a) zaprojektować lub wybrać optymalnie wydajne procedury i eksperymenty, (b) zapewnić uzyskanie maksymalnej istotnej informacji przez analizę danych chemicznych i (c) uzyskać wiedzę o układach chemicznych.”.

Str. 2, wiersz 8-10 powinno być: „pomiędzy zmiennymi w obrębie badanego układu. Jednak należy podkreślić, że chociaż metody bazujące na modelowaniu danych mogą mieć duży potencjał, nie są one w stanie zastąpić zweryfikowanych czy przyjętych teorii, jeżeli takie istnieją.”.

Str. 2, wiersz 16 powinno być: „w wielu dziedzinach naukowych i technologicznych”.

Str. 2, wiersz 18-20 powinno być: „również są stosowane w przemyśle, takim jak petrochemiczny, tekstylny, w czujnikach i kosmetyce, z dalszą możliwością wprowadzania w innych dziedzinach nauki”.

Str. 2, wiersz 29-30 powinno być: „paradygmatowi dużych zbiorów danych”.

Str. 2, wiersz 33 - strona 3, wiersz 3 powinno być: „Algorytmy chemometryczne nadają się również do eksploracji danych ponieważ uwzględniają zmienność (*variations*) wielu źródeł, które mogą być skorelowane (współliniowe zmienne). Rozważania opisane w części 1-2-2 Dane, nadal dotyczą eksploracji danych w celu uzyskania informacji o dobrej jakości.”.

Str. 3, wiersz 5 powinno być: „W chemometrii, w przeciwieństwie do klasycznej chemii analitycznej”.

Str. 3, wiersz 10-12 powinno być: „zależna od założeń o rozkładzie danych niż inne metody statystyczne, gdyż rzadko obejmuje testowanie hipotez. Podczas modelowania, wzmocnieniu mogą ulegać najbardziej czułe badane własności”.

Str. 3, wiersz 22 powinno być: „W jednowymiarowej analizie, zmierzone zmienne układu są analizowane indywidualnie”.

Str. 4, wiersz 9-10 powinno być: „zmienne są rzadko ortogonalne, ale w mniejszym lub większym stopniu współliniowe, co sprzyja użyciu analizy wielowymiarowych danych”.

Str. 4, wiersz 26 powinno być: „Wiele zadań analitycznych odpowiada tej kategorii”.

Str. 4, wiersz 33 powinno być: „Techniki z nadzorem są najbardziej powszechne w zastosowaniach chemometrycznych.”.

Str. 5, wiersz 17 powinno być: „macierz odwrotna (nie zawsze występuje) macierzy X ”.

Str. 5, wiersz 25 powinno być: „1-2-1. Etapy implementacji”.

Str. 5, wiersz 31-32 powinno być: „Zbiór danych powinien pokrywać zmienność eksplorowanej zmiennej (zmiennych)”.

Str. 5, wiersz 33 powinno być: „jeżeli dostępne dane nie pokrywają oczekiwanej zmienności, przygotować i zmierzyć próbki”.

Str. 6, wiersz 4 powinno być: „wstępnego przetwarzania matematycznego”.

Str. 6, wiersz 30-31 powinno być: „wektor zawierający n wartości danych odpowiadających obiektowi. Każdy więc obiekt pojawia się jako punkt w m wymiarowej przestrzeni”.

Str. 7, wiersz 4 powinno być: „oceny każdej zmiennej osobno”.

Str. 7, wiersz 11-12 powinno być: „pomaga uniknąć nadmiarowości”.

Str. 7, wiersz 16 powinno być: „interferencji wyjściowych danych lub błędu pomiarowego”.

Str. 7, wiersz 28-29 i 32-33 powinno być: „błędy procedury porównawczej”.

Str. 7, wiersz 31 powinno być: „wyjaśnia jedynie część wariancji”.

Str. 8, wiersz 5 powinno być: „Zakłócenia mogą wynikać”.

Str. 8, wiersz 7 powinno być: „Zakłócenia wprowadzają zmienność do danych”.

Str. 8, wiersz 13 powinno być: „czyniąc jej wydobywanie bardziej złożonym”.

Str. 8, wiersz 14-15 powinno być: „W celu polepszenia procesu modelowania, do danych X jak i Y można zastosować wiele transformacji przed analizą wielowymiarową.”.

Str. 8, wiersz 31-32 powinno być: „eliminacja szumu, usuwanie linii bazowej/tła (*background*)” (do ustalenia).

Str. 9, wiersz 8 powinno być: „W ten sposób część szumu lub rozproszenia jest usuwana z widma.”.

Str. 9, wiersz 21 powinno być: „druga pochodna również usuwa (*remove*) liniowy efekt”.

Str. 9, wiersz 23 powinno być: „wzmocnić widoczność”.

Str. 9, wiersz 30-31 powinno być: „w sygnale docelowym (*target signal*)”.

Str. 10, wiersz 14 powinno być: „dla danego zbioru danych”.

Str. 10, wiersz 27-29 powinno być: „Diagram przedstawiony na ryc. 5.21.-1 pokazuje jak zbiory danych są zwykle dzielone aby zbudować, ocenić i zwalidować model jako ostateczny cel walidacji zgodnej z regulacjami.”.

Str. 11, wiersz 2 powinno być: „Ryc. 5.21.-1. *Diagram procesu walidacji modelu chemometrycznego z nadzorem*”.

Str. 12, wiersz 5 powinno być: „próbek użytych do budowy modelu kalibracyjnego i obejmować”.

Str. 12, wiersz 14-15 powinno być: „Próbki kalibracyjne (*calibration samples*) mogą być iteracyjnie podzielone na wewnętrzny zbiór testowy i pozostały zbiór uczący.”.

Str. 13, wiersz 5 powinno być: „1-3-5. **Parametry działania**”.

Str. 13, wiersz 11 powinno być: „zależności do odpowiadających danych niezależnych (X) i danych zależnych (Y)”.

Str. 13, wiersz 18-20 powinno być: „które są porównane z obserwowanymi wartościami y (zmierzonymi niezależnie stosując procedurę porównawczą)”.

Str. 14, wiersz 23 powinno być: „wsparcie wyboru modelu, który działa dobrze”.

Str. 14, wiersz 29 powinno być: „specyficzne dla zastosowanej techniki chemometrycznej i charakteru danych analitycznych”.

Str. 14, wiersz 31 powinno być: „Wybrane próbki są albo specyficznie przyporządkowane”.

Str. 14, wiersz 33 powinno być: „ze specyficznym sposobem usuwania próbek”.

Str. 15, wiersz 3-6 powinno być: „nazywana również walidacją krzyżową ‘pomiń podzbiór’ (*leave-subset-out*). Model jest uczony przy użyciu $K-1$ podzbiorów i oceniany za pomocą pominiętego (*left out*) podzbioru. Ta procedura jest powtarzana K -razy, każdorazowo z innymi K podzbiórami pominiętymi.”.

Str. 15, wiersz 8 powinno być: „Przeprowadza się serię walidacji krzyżowych”.

Str. 15, wiersz 12-17 powinno być: „Najprostszy algorytm walidacji krzyżowej o K segmentach to wariant walidacji krzyżowej ‘pomiń jeden obiekt’ (*‘leave-one-out’ cross-validation*, LOOCV), gdzie K jest równe N (liczba próbek w kalibracyjnym zbiorze danych). W LOOCV jedna próbka zostaje wyjęta ze zbioru kalibracyjnego, a model jest budowany dla pozostałych próbek. Uzyskany model jest następnie użyty aby uzyskać przewidywanie dla wybranej (*selected*) próbki.”.

Str. 15, wiersz 20-21 powinno być: „model o najniższym RMSECV”.

Str. 15, wiersz 23-24 powinno być: „nie jest już zdolny do przewidzenia/klasyfikowania poprawnie nowych próbek”.

Str. 15, wiersz 26 powinno być: „losowo dzielony na K segmentów, z której każdy zawiera N/K obserwacji (*observations*)”.

Str. 16, wiersz 5-7 powinno być: „*W tym przykładzie wybrano losowo jedną trzecią pominiętych danych, przy czym każdą próbkę pomijano dokładnie jeden raz.*”.

Str. 16, wiersz 10 powinno być: „1-3-7. **Ocena działania modelu**”.

Str. 16, wiersz 15-16 powinno być: „walidacją wymaganą przez aktualne zalecenia regulacyjne.”.

Str. 17, wiersz 5-7 powinno być: „materiałów, które są przetwarzane w tym samym miejscu wytwarzania i są narażone na zmieszanie. Jeżeli oprócz chemicznej identyfikacji inne właściwości”.

Str. 17, wiersz 10 powinno być: „materiały przetwarzane blisko rozpatrywanego procesu”.

Str. 17, wiersz 22-23 powinno być: „wymaga uwzględnienia aspektów, na które zmiana nie wpływa.”.

Str. 17, wiersz 29-30 powinno być: „W celu oceny specyficzności, klasyfikując zbiór testowy, analizuje się liczbę błędów fałszywie pozytywnych i fałszywie negatywnych.”.

Str. 17, wiersz 34 powinno być: „sprawność aparatury analitycznej”.

Str. 18, wiersz 10-12 powinno być: „Należy ustalić obiekty odległe, które nie są uwzględniane w przestrzeni kalibracyjnej. Można tego dokonać używając współczynnika korelacji pomiędzy próbka, środkiem (*centroid*) zbioru kalibracyjnego (dane X)”.

Str. 19, wiersz 4-5 powinno być: „ocenę odchylenia standardowego pomiarów”.

Str. 19, wiersz 12 powinno być: „Należy zwrócić szczególną uwagę”.

Str. 19, wiersz 14-15 powinno być: „W ocenie tych parametrów może być pomocne użycie planu eksperymentów.”.

Str. 19, wiersz 16 powinno być: „Model chemometryczny może również być badany”.

Str. 20, wiersz 6-7 powinno być: „Złożoność dużych zbiorów lub tabel danych utrudnia ich interpretację”.

Str. 20, wiersz 11 powinno być: „zmienne wpływają w ten sam sposób”.

Str. 21, wiersz 3-7 powinno być: „Procedura prowadzi do konstrukcji jednej macierzy z ortogonalnymi kolumnami wektorów/wektorami kolumnowymi (wyniki – *scores*) i innej macierzy z ortonormalnymi wierszami wektorów/wektorami wierszowymi (ładunki – *loadings*). Główne składowe (*principal components*, PC) lub ukryte zmienne są liniową kombinacją osi wyjściowych zmiennych.” (do ustalenia).

Str. 21, wiersz 31 powinno być: „zmiennosc jest zdefiniowana jako [statystyczna] wariancja”.

Str. 22, wiersz 1-2 powinno być: „składowe są uszeregowane ze względu na malejącą wyjaśnioną wariancję”.

Str. 22, wiersz 5-6 powinno być: „określa się mianem podstawowych PC (*primary PC*)”.

Str. 22, wiersz 10 powinno być: „mogą być następnie pokazane”.

Str. 22, wiersz 13 powinno być: „składowych zachowanych w modelu PCA”.

Str. 22, wiersz 17-18 powinno być: „Procenty reszt i wyjaśniona wariancja, wskazuje jak dobrze model jest dopasowany/pasuje do danych.” (do ustalenia).

Str. 22, wiersz 28-29 powinno być: „może prowadzić do modelu właściwszego dla celu”.

Str. 23, wiersz 5-8 powinno być: „*ukazując oryginalną przestrzeń. Geometryczne przedstawienie trzech różnych zbiorów danych X. Po lewej stronie, obiekty wykreślono w wielowymiarowej przestrzeni o 3 wymiarach (3-dimensional multivariate space)*”.

Str. 24, wiersz 5 powinno być: „metody wstępnego przetwarzania”.

Str. 24, wiersz 16 powinno być: „nowe źródło zmienności”.

Str. 24, wiersz 21 powinno być: „wyjściowe (*original*) zmienne”.

Str. 24, wiersz 31-34 powinno być: „W chemometrii termin ‘źródło’ dotyczy w pierwszej kolejności wkładów *m* poszczególnych składników chemicznych do całkowitego sygnału każdej próbki *n*, analizowanej jako mieszanina. ICA może być uznana jako rozszerzenie (*extension*) techniki PCA.”.

Str. 25, wiersz 1-6 powinno być: „powiązane z czystymi sygnałami źródłowymi (*source signals*). Podczas gdy PCA optymalizuje macierz wariancji danych, która odpowiada statystyce drugiej pochodnej (*second-order statistics*), ICA optymalizuje statystykę wyższego rzędu, taką jak kurtoza. Dlatego PCA ujawnia nieskorelowane składowe, a ICA identyfikuje odrębne źródła, jako składowe.”.

Str. 25, wiersz 20-22 powinno być: „(maksymalizując nie Gaussowski charakter, minimalizując wzajemną informację lub stosując oszacowanie metodą największej wiarygodności (*maximum likelihood estimation*))”.

Str. 25, wiersz 26 powinno być: „Ustalenie liczby niezależnych składowych”.

Str. 25, wiersz 29-31 powinno być: „czystych składowych do kilku wkładów i powstaną zaszumione sygnały. Ponadto, należy podkreślić, że ICA nie szereguje niezależnych składowych”.

Str. 26, wiersz 9-10 powinno być: „ICA jest szczególnie ceniona w obrazowaniu chemicznym i użyteczna do dekompozycji (*decomposition*) danych spektroskopowych”.

Str. 26, wiersz 16-18 powinno być: „Współczynnik korelacji i kosinus podobieństwa (*correlation coefficient and the cosine similarity*), również często nazywane korelacją, są statystycznymi metodami używanymi do porównania danych i wyznaczenia stopnia podobieństwa.”.

Str. 27, wiersz 19 powinno być: „mniej bliskie siebie”.

Str. 28, wiersz 4 powinno być: „odwrotność macierzy wariancji-kowariancji”.

Str. 28, wiersz 9-10 powinno być: „Odległość Mahalanobisa wektora x_i od środka c zbioru danych \bar{x} ”.

Str. 28, wiersz 14-16 powinno być: „Macierz C_x^{-1} jest odwrotnością macierzy wariancji-kowariancji. Jeżeli wymiar m staje się duży, a zmienne pomiarowe są skorelowane, macierz C_x jest źle określona lub osobliwa”.

Str. 28, wiersz 29-30 powinno być: „Jest to tylko zgrubne przybliżenie krytycznej wartości, zwłaszcza dla PLS.”.

Str. 29, wiersz 14-15 powinno być: „zmniejszenie wymiarowości, ortogonalności pomiędzy PC i uszeregowanie PC.”.

Str. 29, wiersz 31-32 powinno być: „przypisanie obiektu testowego x_i do jednej z wstępnie zdefiniowanych K grup (lub klas) w zbiorze danych”.

Str. 30, wiersz 8-10 powinno być: „macierz wariancji-kowariancji dla każdej klasy. Dlatego w QDA wymagane jest oszacowanie dużo większej liczby parametrów, co powinno mieć miejsce jedynie, gdy jest dostępny wystarczająco liczny zbiór danych.”.

Str. 30, wiersz 16 powinno być: „składa się z próbki o znanych cechach”.

Str. 30, wiersz 18 powinno być: „Każda próbka jest przypisana do jednej z najbliższych klas”.

Str. 30, wiersz 24-25 powinno być: „Liczba istotnych głównych składowych może być dostosowana osobno dla każdej grupy obiektów przez walidację krzyżową.”.

Str. 31, wiersz 15-16 powinno być: „dokładność, współczynnik klasyfikacji poprawności (błąd/brak błędu) (*error/non-error rate*) i proporcja nieprzypisanych próbek”.

Str. 31, wiersz 21 powinno być: „PCA do rozróżnienia klas, które trudno rozdzielić”.

Str. 32, wiersz 14-16 powinno być: „Dla wszystkich z nich należy wybrać miarę podobieństwa (odmienności) lub odległość pomiędzy obiektami. Odległość euklidesowa jest typowym miernikiem (*metric*).”.

Str. 32, wiersz 28-31 powinno być: „Oba tryby prowadzą do hierarchicznej struktury klastrów, co oznacza, że każdy klastrow składa się z zestawu podklastrów (*set of sub-clusters*). Algorytmy różnią się sposobem obliczania podobieństwa pomiędzy klastrami, nazywanym połączeniem (*linkage*). Algorytmy całkowitych połączeń”.

Str. 33, wiersz 5 powinno być: „Ten algorytm może tworzyć bardziej zbalansowane/zrównoważone (*balanced*) podziały” (do ustalenia).

Str. 34, wiersz 4-5 powinno być: „klastry blisko połączonych punktów”.

Str. 34, wiersz 12 powinno być: „musi być ustalona *a priori* liczba klastrów”.

Str. 34, wiersz 19 powinno być: „grupowanie usiłuje ustalić najlepsze dopasowanie danych do założonego modelu”.

Str. 35, wiersz 7-10 powinno być: „Końcowy wynik grupowania będzie się różnił w zależności od wybranej liczby klastrów i pozycji środków wybranych przez algorytm. Dlatego ważne jest aby kilkukrotnie uruchomić algorytm, z różnymi wstępnie wybranymi losowo środkami. Mierniki (*metrics*) użyte do obliczenia”.

Str. 35, wiersz 19-22 powinno być: „obiekty są luźno rozłożone pomiędzy wymiarami. W przypadku algorytmu hierarchicznego grupowania nie trzeba wstępnie ustalić liczby klastrów. Ich liczbę uzyskuje się ucinając dendrogram na wstępnie określonym poziomie. Utworzone klastry w dużym stopniu mogą być zależne od wybranego w algorytmie sposobu połączenia”.

Str. 36, wiersz 9-11 powinno być: „Każda ścieżka od korzenia do danego liścia odpowiada serii następujących po sobie węzłów decyzji ulokowanych na ścieżce. Każda ścieżka może być przekształcona w sekwencje kroków”.

Str. 36, wiersz 21 powinno być: „algorytmy drzew decyzyjnych minimalizują stratę”.

Str. 36, wiersz 24 powinno być: „funkcji strat”.

Str. 36, wiersz 30-31 powinno być: „dominującej klasie w przypadku różnicowania”.

Str. 36, wiersz 33-34 powinno być: „Typowym podejściem radzenia sobie z niestabilnością i brakiem równowagi na etapie budowy drzew decyzyjnych”.

Str. 37, wiersz 7-9 powinno być: „RF poszukuje najlepszej zmiennej spośród losowego podzbioru zmiennych. To pociąga za sobą większą różnorodność drzew, która bardziej uwypukla błąd systematyczny dla niższej wariancji.”.

Str. 37, wiersz 14 powinno być: „używa się w trakcie każdego przebiegu”.

Str. 37, wiersz 22-23 powinno być: „ryzyko nadmiernego dopasowania (*overfitting*) zmniejszone, co prowadzi do ogólnie lepszego modelu. Uśrednianie wyników dla wielu drzew utworzonych razem”.

Str. 37, wiersz 28-32 powinno być: „Inne ograniczenie wynika z nieradzenia sobie przez drzewa decyzyjne z brakiem równowagi. Przykładowo, słabe działanie przewidywania dokładności dla mało licznych podzbioru danych uczących nie wpłynie znacząco na końcowy model. Rozwiązaniem może być zrównoważenie klas i zakresu zmiennej odpowiedzi nim drzewo decyzyjne zostanie dopasowane (*fitting*).”.

Str. 38, wiersz 1-3 powinno być: „Jednakże, te ulepszenia odbywają się kosztem dokładności dopasowania do zbioru uczącego podczas procesu modelowania i utraty zdolności interpretacyjnej.”.

Str. 38, wiersz 31-34 powinno być: „W pierwszym etapie, oszacowywana jest macierz C , używając początkowego przybliżenia (*initial estimate*) macierzy S , a następnie przybliżenie dla macierzy C pozwala uzyskać poprawione przybliżenie (*updated estimate*) macierzy S , które następnie polepsza C (stąd określenie ‘naprzemienne’).”.

Str. 39, wiersz 2 powinno być: „układzie fizykochemicznym”.

Str. 39, wiersz 7-8 powinno być: „Często możliwe jest uzyskanie dokładnego oszacowania czystych składowych widma lub profili wkładów”.

Str. 39, wiersz 10-11 powinno być: „Ponadto, wiedza fizyczna i chemiczna o układzie może zostać wykorzystana do weryfikacji wyniku”.

Str. 40, wiersz 3 powinno być: „Jeżeli problemem są liniowość i selektywność”.

Str. 40, wiersz 7 powinno być: „Zastosowanie MCR-ALS”.

Str. 40, wiersz 12 powinno być: „MLR rozszerza modele”.

Str. 40, wiersz 15 powinno być: „Model MLR może również być rozszerzony”.

Str. 40, wiersz 22 powinno być: „W przypadku m niezależnych zmiennych i n obiektów, model ma postać”.

Str. 40, wiersz 28-29 powinno być: „minimalizuje sumę kwadratów błędów pomiędzy oszacowywaną funkcją regresji”.

Str. 41, wiersz 2-6 powinno być: „Oszacowanie parametrów modelu powoduje odwrócenie macierzy ($X^T X$). Jeżeli wymiar m staje się duży, a zmienne pomiarowe są skorelowane, macierz $X^T X$ staje się źle określona lub osobliwa. Źle określona macierz $X^T X$ będzie prowadzić do bardzo zmiennego estymatora \hat{b} ”.

Str. 41, wiersz 11 powinno być: „oszacować parametr modelu b ”.

Str. 41, wiersz 13 powinno być: „nieskorelowanych zmiennych”.

Str. 41, wiersz 14-15 powinno być: „są silnie skorelowane”.

Str. 41, wiersz 21-22 powinno być: „jedynie do prostych przypadków regresji o niskiej wymiarowości (*low-dimensional*)”.

Str. 41, wiersz 24 powinno być: „można łatwo odwrócić ich macierz”.

Str. 41, wiersz 30 powinno być: „w przypadku występowania wielokrotnej współliniowości”.

Str. 41, wiersz 33 powinno być: „W celu uniknięcia nadmiernego dopasowania model MLR często jest stosowany”.

Str. 42, wiersz 1 powinno być: „Wybór optymalnych zmiennych X ”.

Str. 42, wiersz 11 powinno być: „odpowiedni do prostych o niskiej wymiarowości”.

Str. 42, wiersz 30-33 powinno być: „Ponieważ wyjściowa macierz danych X jest zastąpiona ortogonalną macierzą wyników \hat{T} , MLR dobrze nadaje się do oszacowania współczynników regresji, oznaczonych jako macierz \hat{Q}^T . Celem jest ponownie ustalenie współczynników regresji, które minimalizują sumę kwadratów reszt w macierzy \hat{F} .”

Str. 43, wiersz 1 powinno być: „model PCR jest pokazany dla wielowymiarowych (*multivariable responses*) odpowiedzi”.

Str. 43, wiersz 7 powinno być: „oryginalnej przestrzeni danych używając ładunków \hat{P} ”.

Str. 44, wiersz 12-13 powinno być: „w przypadku niektórych schematów walidacji krzyżowej. Liczba PC, dla której obserwuje się minimalną wariancję reszt Y ”.

Str. 44, wiersz 22 powinno być: „zawarta w głównych składowych wyższego rzędu”.

Str. 45, wiersz 12 powinno być: „Niewyjaśniona część danych jest utworzona z reszt”.

Str. 45, wiersz 16-17 powinno być: „czynniki zawierają informację opisującą maksymalną zmienność X ”.

Str. 45, wiersz 29 powinno być: „tym samym redukując wpływ dużej zmienności X ”.

Str. 45, wiersz 33 powinno być: „umożliwiają uniknięcie nadmiernego dopasowania”.

Str. 47, wiersz 13-14 powinno być: „Zwykle PLS2 jest preferowanym podejściem w celach przesiewowych i w przypadkach wysoko skorelowanych zmiennych Y ”.

Str. 47, wiersz 17-18 powinno być: „włącza strukturę danych Y do rozkładu macierzy kalibracji X ”.

Str. 47, wiersz 27 powinno być: „do oszacowania składników pikseli”.

Ad 5) Uchwała dotycząca rozdziału 5.21. *Metody chemometryczne stosowane do danych analitycznych*, zostanie podjęta przez Grupę ekspercką ds. Metod Fizykochemicznych KF po omówieniu całego tekstu.

Ad 6) Ustalono termin kolejnego posiedzenia Grupy, na którym omawiana będzie dalsza część tekstu 5.21., na dzień 19.12.2022 r.

Na zakończenie posiedzenia Przewodniczący Grupy eksperckiej prof. dr hab. Zbigniew Fijałek oraz Dyrektor Departamentu Farmakopei dr Ewa Leciejewicz-Ziemecka podziękowali zebranym za udział w wideokonferencji i merytoryczną dyskusję, a także za przesłane uwagi.

Przewodniczący
Grupy eksperckiej ds. Metod
Fizykochemicznych KF

Prof. dr hab. Zbigniew Fijałek