

<https://doi.org/10.15407/ggcm2024.195-196.100>

УДК 550.43:553.97

Юрій ХОХА¹, Мирослава ЯКОВЕНКО²

Інститут геології і геохімії горючих копалин НАН України, Львів, Україна,
e-mail: ¹khoha_yury@ukr.net; ²myroslavakoshil@ukr.net

**ЗАСТОСУВАННЯ ХЕМОМЕТРИЧНИХ МЕТОДІВ ТА
РЕГРЕСИВНИХ МОДЕЛЕЙ В ОБРОБЦІ NIR СПЕКТРІВ ТОРФУ
ДЛЯ КІЛЬКІСНОГО ВИЗНАЧЕННЯ
ЙОГО ХІМІКО-ТЕХНОЛОГІЧНИХ ПОКАЗНИКІВ**

Розглянуто теоретичні та практичні аспекти застосування NIR спектрометрії, поєднаної з хемометрією, для експрес-аналізу торфу. Спектрометрія в ближньому інфрачервоному діапазоні забезпечує значний обсяг інформації про складні органічні системи, зокрема нерегулярні полімери, до яких належить торф. Порівняно з класичними аналітичними методами NIR-спектрометрія дозволяє досліджувати зразок без складної пробопідготовки, а сам аналіз триває декілька хвилин. Водночас, позаяк результати виражаються як інтенсивність відбиття випромінювання в діапазоні обертонів фундаментальних частот, їхня обробка вимагає застосування спеціальних математичних та статистичних методів. Показано, що такі методи надає хемометрія. Розглянуто та проаналізовано базові способи препроцесінгу спектрів відбиття. За результатами обробки експериментальних даних доведено, що цей метод можна використовувати для встановлення відповідності торфу чинним нормам, стандартам та технічним умовам щодо вологості, вмісту зольного (неорганічного) залишку та кислотності (*pH*).

Ключові слова: ближня інфрачервона спектроскопія (NIR), аналіз торфу, прогнозні моделі, багатофакторний аналіз, метод регресії часткових найменших квадратів (PLS), ефект попередньої обробки.

Вступ. Торф – це суміш складних нерегулярних полімерів рослинного походження, які пройшли бар'єр біодеградації і є на початкових стадіях катагенезу. Ця суміш умовно поділяється на целюлозний, геміцелюлозний та лігніновий залишок, гумінові та фульвові кислоти та мінеритарні компоненти. Наявність хелатних структур сприяє захопленню металів з різними ступенями окислення органічними компонентами торфу, а рухливість водню в поліфенолах впливає на кислотність/основність та редокс-потенціал. Оскільки торф є сумішшю, то він містить неорганічні компоненти, переважно глинисті мінерали, рідше – карбонати, фосфати, пірит тощо.

© Юрій Хоха, Мирослава Яковенко, 2024

ISSN 0869-0774. Геологія і геохімія горючих копалин. 2024. № 3–4 (195–196)

Зазначені особливості будови торф'яного матриксу зумовлюють складність методів його дослідження. Традиційні способи оцінки якості торфу вимагають трудомістких і тривалих лабораторних досліджень з визначення фізичних і хімічних властивостей зразків торфу за допомогою муфельних печей, сушарок, іоніметрів та інших приладів. Проте сучасні умови промислового торфовидобутку вимагають швидких рішень для класифікації торфу, оцінки його якості, бажано в польових умовах, для раціонального вибору найбільш перспективної сировини. Тому існує великий інтерес до пошуку інструментальних методів для визначення характеристик торфу за допомогою швидкої та надійної процедури з мінімальною підготовкою зразків.

Метою дослідження є пошук альтернативної методики, здатної дати швидку хіміко-технологічну характеристику торфу (вологість, зольність, pH та ін.) за допомогою спектроскопії в ближньому інфрачервоному діапазоні (NIR, $3950\text{--}7500\text{ см}^{-1}$) з використанням ручних польових спектрометрів та хемометричних методів аналізу.

Торф, як і інші складні суміші органічних речовин, зокрема полімерної структури, може бути якісно та кількісно охарактеризований спектрами ближнього інфрачервоного діапазону (Яковенко & Хоха, 2024). Наше завдання – дослідити, та за потреби модифікувати спектри ближнього ІЧ-діапазону, та зіставити їх (знайти кореляції) з кількома характеристиками торфу за допомогою багатофакторного калібрування. Спектроскопія ближнього інфрачервоного діапазону використовується як техніка керування виробничим процесом в режимі онлайн завдяки простоті вимірювань та кількості інформації, що містить у непрямій формі кожен спектр (McClure, 1994). З іншого боку, інформація, представлена в спектрі, має недоліки: у спектрах є фізичні ефекти (шум приладу, нерівність поверхні зразка, неоднорідна структура), які спотворюють дані та надають неточні або зміщені результати (під результатами маються на увазі поглинання/відбиття в ділянці обертонів фундаментальних частот, їхнє взаємне розташування та інтенсивність).

Торф ми розглядаємо як складну, різноманітну та гетерогенну систему, а тому віднесення та кількісне визначення окремих функціональних груп до певних категорій не дає точного опису його складу, характеристик його якості та продуктивності. Для цього повний спектр необхідно розглядати як єдине ціле, що передбачає обробку тисяч змінних. Поєднання складних матеріалів з потребою у швидких, надійних, точних визначеннях спонукає до розробки та використання методів багатофакторного калібрування.

У проведених дослідженнях спектральні дані були отримані в одиницях відбиття (см^{-1}), а також було проведено дослідження різних попередніх обробок (корекція базової лінії, нормалізація, мультиплікативна корекція сигналу, перші та другі похідні тощо) для зразків торфу.

У 1970-х роках було введено і представлено нову наукову концепцію – хемометрика/хемометрія – наука, пов'язана з виконанням обчислень вимірювань, зроблених у хімічному процесі або системі з метою отримання інформації про її стан за допомогою математичних або статистичних методів (Nunes et al., 2012).

Сьогодні, завдяки розвитку обчислювальної техніки, ускладненням та узагальненням математичного апарату, досягненням та важливим успіхам

в обчисленнях, хеометрика набуває популярності в розмаїтих галузях науки, пов'язаних з хімією, включаючи науки про Землю, зокрема ґрунтознавство (забруднення ґрунтів, Mostert et al., 2010), екологію (аналіз навколишнього середовища), геологію (аналіз вугілля, Andrés & Vona, 2005), та промисловості, зокрема контроль якості продукції (Супрунович та ін., 2022). За останні десятиріччя розроблено багато методів та обчислювальних інструментів для хеометричних розрахунків. Хеометричні методи особливо важливі в аналітичних галузях для вдосконалення методів аналізу, аналітичних приладів та загальноаналітичних методик.

Відбір проб, методи та аналізи досліджень. Для проведення дослідження використовували проби торфу, відібрані з різних глибин (від 0 до 140 см) на п'яти представницьких ділянках Львівської області, виокремлених як самостійні торфові родовища (Гончари – з інтервалами: 0–20, 20–40, 40–60, 60–80, 80–100, 100–120 та 120–140 см; Артюхівське – 0–30 см; Полоничне – 0–40 см; Білогорща – 0–40, 40–80 см; Скнилівок – 0–40 см).

У лабораторних умовах проби торфу висушували за кімнатної температури, уникаючи потрапляння прямого сонячного світла, подрібнювали до розміру частинок, що проходять крізь сито з отворами діаметром 5 мм, і розтирали в агатовій ступці.

Проби торфу проаналізували на вологість, зольність, *pH* згідно зі стандартами ДСТУ та ГОСТ (Інститут ґрунтознавства та агрохімії імені О. Н. Соколовського Української академії аграрних наук, 2008; Технічний комітет стандартизації «Ґрунтознавство», 2016; Інститут сільськогосподарської мікробіології та агропромислового виробництва НААН, 2016) для всіх досліджуваних зразків торфу. Для кожної проби виконували трикратні аналітичні визначення вказаних показників, результати яких усереднювали.

NIR-спектри реєстрували на ІЧ-спектрометрі ближнього поглинання *Neo-Spectra (Si-Ware System)* з використанням програмного забезпечення *Spectro-MOST Micro™*. Для цього дослідження автори виготовили кювету з акрилового полімеру з прорізом на дні, яку накладали/встановлювали на ефектор спектрометра і заповнювали зразком торфу висотою не менш ніж 10 мм. Перед зняттям спектра ми реєстрували базову лінію з використанням адаптера максимального розсіювання.

Реєстрація спектрів. Кювету для зразка заповнювали кожною пробкою торфу, поверхню вирівнювали шпателем і поміщали в насадку для дифузного відбиття та отримували спектр. Отримані спектри були результатом спільного додавання п'яти повторних сканувань спектрів у діапазоні 1300–2600 нм ($3950\text{--}7500\text{ см}^{-1}$), виконаних із цифровою роздільною здатністю від 4 до 15 см^{-1} залежно від частини спектра. Така характеристика приладу зумовлена будовою чипа MEMS.

Обробку та інтерпретацію основних показників, отриманих за результатами проведених аналітичних визначень та спектральних даних, проводили за допомогою програм MS Excel, STATISTICA. Багатовимірна статистика (*Multivariate statistics*) – аналіз головних компонентів (*Principal Component Analysis, PCA*), регресія методом часткових найменших квадратів (*Partial Least Squares Regression, PLS-R*) і дискримінантний аналіз методом часткових найменших квадратів (*Partial Least Squares Discriminant Analysis, PLS-DA*),

була застосована і розроблена за допомогою безкоштовного програмного забезпечення *Chemoface версії 1.67*, виконаного в середовищі MATLAB (Nunes et al., 2012).

Усі регресійні моделі базувалися на діапазоні довжин хвиль (1300–2600 нм). Для зменшення спектральних шумів та аберацій, спричинених відмінностями в структурі поверхні та розмірі частинок, використовували математичну попередню обробку.

Результати досліджень. У табл. 1 та 2 наведено вихідні усереднені дані за результатами проведених аналітичних визначень та деякі статистичні параметри розподілу показників у досліджуваних пробах торфу Львівської області.

Коефіцієнти кореляції для розглянутих аналізованих характеристик досліджуваного торфу відображені в табл. 3: показники вологості (W , %) та зольності (A , %) торфу мають позитивну кореляцію між собою (сила зв'язку $r = 0,62$); pH не корелює з жодним показником.

Т а б л и ц я 1. Вологість, зольність та pH досліджуваних проб торфу Львівської області

Родовище/поклад	Вологість, W , %	Зольність, A , %	pH
Скнилівок	51,02	29,65	4,04
Полоничне_1	65,42	17,5	7,78
Полоничне_2	62,02	25,53	5,76
Полоничне_3	16,11	10,57	5,37
Артюхівське_0–30	15,92	35,67	6,23
Білогорща_0–40	56,69	59,97	3,77
Білогорща_40–80	59,73	56,76	3,59
Гончари_0–20	18,19	8,4	5,51
Гончари_20–40	19,6	9,82	5,41
Гончари_40–60	19,02	10,21	5,43
Гончари_60–80	18,16	8,84	5,38
Гончари_80–100	17,97	8,97	5,33
Гончари_100–120	17,55	9,2	5,48
Гончари_120–140	17,01	8,75	5,43
Гончари	39,84	60,8	7,51

Т а б л и ц я 2. Статистичні параметри розподілу показників у досліджуваних пробах торфу Львівської області

Показник	Середнє	Медіана	Мін	Мак	Дисперсія	Ст. відхил.	Коеф. варіації
Вологість, W , %	32,95	19,02	15,92	65,42	403,73	20,09	60,98
Зольність, A , %	24,04	10,57	8,40	60,80	404,02	20,10	83,60
pH	5,47	5,43	3,59	7,78	1,332	1,154	21,12

Т а б л и ц я 3. Кореляційна матриця основних характеристик досліджуваного торфу Львівської області

Показник	Середнє	Ст. відхил.	W , %	A , %	pH
W , %	32,950	20,093	1,00	1,00	1,00
A , %	24,043	20,100	0,62	-0,17	
pH	5,466	1,154	-0,08		

Примітка: півгрубий шрифт – кореляція значуща за рівня $p < 0,05$

Спектральні профілі. Спектри відбиття зразків торфу, використаних у цьому дослідженні, наведено на рис. 1.

Інфрачервоні спектри досліджуваних зразків подібні за інтенсивністю спектральних ліній (див. рис. 1). На ІЧ-спектрограмах ближньої інфрачервоної області виявлено три піки (інтенсивні смуги поглинання) органічних сполук з такими усередненими значеннями довжин хвиль (cm^{-1}): 3960, 5200, 6890. Найбільш інтенсивні спектральні максимуми відповідають довжині хвилі 5200 cm^{-1} , що, імовірно, пов'язано з вільною, зв'язаною та/або адсорбованою водою, а також з гідроксильними групами. Декілька фізичних (геометрія приладу, розмір частинок зразка, форма та розподіл, показник заломлення) та хімічних ефектів призводять до перекриття і часто до широких піків. Спектр ближнього ІЧ-діапазону (NIR-спектр) є результатом поглинання обертонів і комбінацій поглинання декількох функціональних груп, таких як C–H, N–H і O–H. Тому для виявлення конкретної та корисної інформації необхідне широке використання методів аналізу багатовимірних даних.

Усі спектри, які будуть використовуватися для кількісних вимірювань, слід досліджувати у форматі, де вісь ординат є лінійною відносно концентрації зразка. Прості методи вимагають, щоб ордината була лінійною з концентрацією (*Multiple Linear Regression, MLR*), натомість більш просунуті процедури, такі як регресія головних компонентів і метод часткових найменших квадратів, можуть впоратися з даними з невеликою нелінійністю, що виникає внаслідок хімічних взаємодій у зразку торфу.

До спектральних даних застосовуються кілька математичних методів попередніх обробок (*pre-processing methods*), щоб мінімізувати фізичні особливості зразків.

Попередня обробка даних – це те, що відбувається з даними перед тим, як вони потрапляють в алгоритм моделювання. Вона допомагає усунути варіацію, тобто безлад у вимірюваннях, перешкоди та шум, і дає можливість аналізу зосередитися на важливих коливаннях.

Джерелами безладу можуть бути: систематична фонові мінливість; фізичні ефекти, такі як розсіювання через частинки; зміни T , P , змінна матриця вибірки; дисперсія; змінна базова лінія або посилення; нелінійність, насиченість; несистематичний випадковий шум та ін.

Попередня обробка дає можливість зменшити відхилення від сторонніх джерел, зробити відповідну дисперсію більш очевидною, покращити статистичну обробку та інтерпретацію, уникнути числових проблем.

Серед усіх можливих попередніх обробок даних для корекції спектральних варіацій найчастіше використовують такі:

– *Зміщення базової лінії (Baseline offset)*: розмір частинок зразків торфу та процедура наповнення чашки викликають варіації в кількості розсіяного/відбитого світла або у фокусній висоті. Тому застосовують цю попередню обробку, яка забезпечує зміщення базової лінії, що може частково виправити відхилення, спричинені цими ефектами.

– *Масштабування змінних (Variable Scaling)*: виконується для зміни дисперсії змінних і, отже, ваги/значення, які їм надаються під час моделювання. Найбільш поширеним методом є автомасштабування (*Autoscale*), при якому дисперсія змінних стає одиничною, а середнє значення – нульовим. Авто-

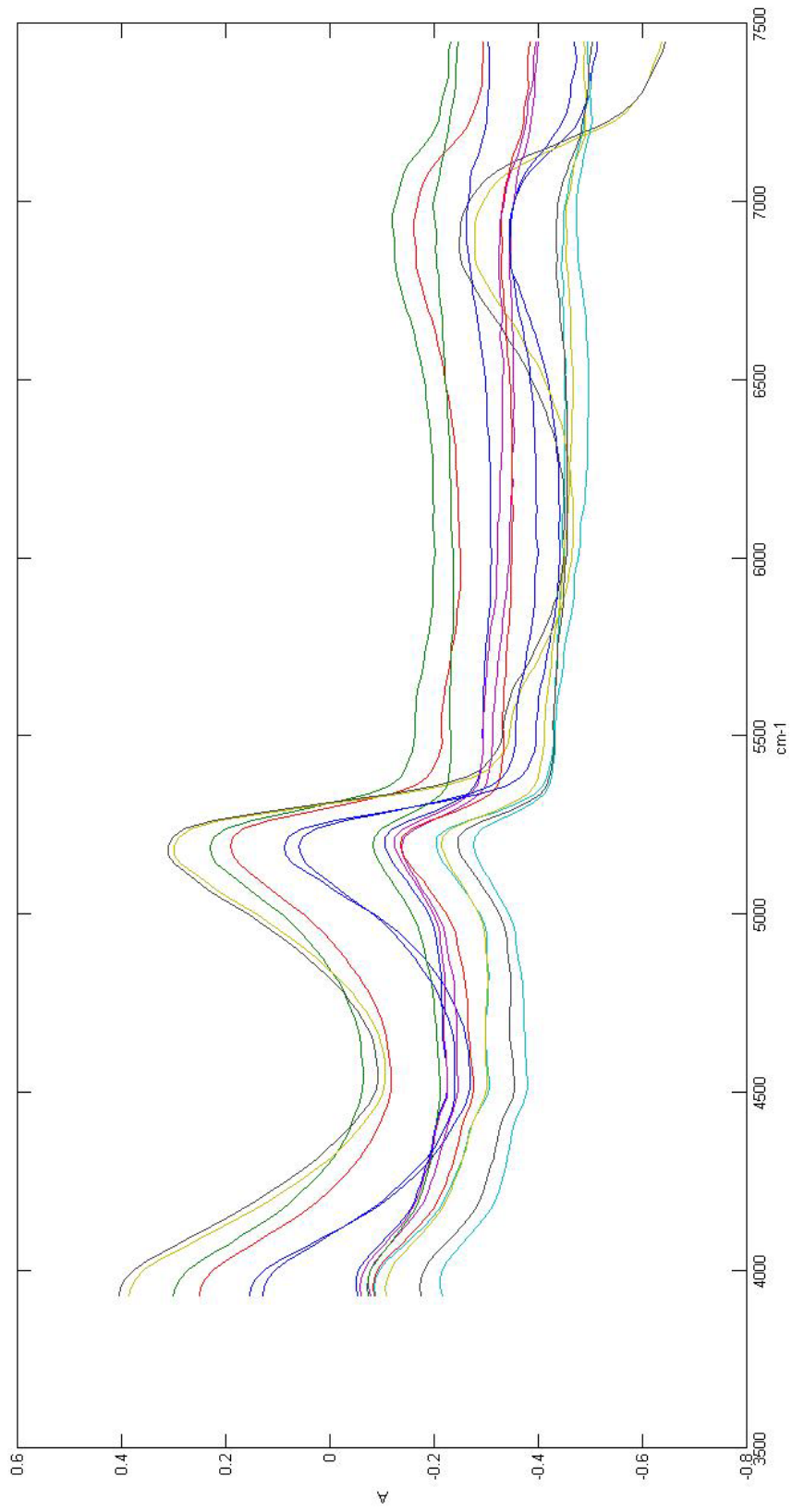


Рис. 1. ІЧ-спектрограма ближньої інфрачервоної області (від 3900 до 7400 cm^{-1}) досліджуваних зразків торфу Львівської області. По осі X – хви-
льове число, cm^{-1} ; по осі Y – одиниці поглинання

масштабування видаляє всю інформацію про масштаб, залишаючи лише те, як змінні корелюють одна з одною – «кореляційна матриця» (рис. 2).

– *Середнє значення (Mean)* на ІЧ-спектрограмі ближньої інфрачервоної області (від 3900 до 7400 cm^{-1}) досліджуваних зразків торфу Львівської області наведено на рис. 3.

– *Нормалізація діапазону (Range normalization)*: усі спектри масштабуються до загального діапазону, приймаючи максимальне значення кожного спектра за 1 (рис. 4), а відсутність сигналу відповідає нулю на осі ординат. Це дозволяє скоригувати незначні відмінності у вазі між зразками.

– *Зміщення базової лінії та нормалізація діапазону (Baseline offset and range normalization)*: цей метод є результатом комбінації двох попередніх. Таким чином, кожен спектр масштабується від 0 до 1. Застосування цієї попередньої обробки особливо помітно в зразках торфу, де показані більш точні та інтенсивні смуги.

– *Мультиплікативна корекція розсіювання (Multiplicative Scatter Correction, MSC)*: здійснюється шляхом коригування базових ліній і нахилів. Кожен спектр порівнюється з еталонним, зазвичай, із середнім спектром, надаючи йому той самий зсув і середній нахил, що й еталонному спектру. Це дозволяє зберегти конкретну хімічну інформацію, за винятком масштабування (Geladi et al., 1985). Враховуються мультиплікативні та адитивні ефекти (рис. 5).

– *Перша та друга похідні (First and Second derivatives)*: метою похідної спектроскопії є виділення смуг поглинання, вирівнювання спектральної базової лінії та одночасно корекція ефектів зміщення розсіювання. Таким чином, спектри першої похідної усувають адитивну базову лінію, коригуючи зміщення базової лінії (рис. 6), тоді як друга похідна використовується для обробки ефектів розсіювання (рис. 7). Незважаючи на те, що це простий і швидкий метод, основним недоліком є те, що похідні спектри не так інтуїтивно очевидні для інтерпретації, як вихідні дані. Крім того, використання похідних збільшує спектральний шум та знижує роздільну здатність. Як перша, так і друга похідні, отримані нами шляхом застосування диференціювання Савицького–Голея, використовуючи вікно з дев'яти точок і поліном четвертого порядку для коригування.

– *Стандартна нормальна варіація (Standard Normal Variate, SNV)* на ІЧ-спектрограмі ближньої інфрачервоної області (від 3900 до 7400 cm^{-1}) досліджуваних зразків торфу Львівської області наведена на рис. 8.

Нормалізація, перша та друга похідні (First and Second derivatives) та стандартна нормальна варіація (Standard Normal Variate, SNV) на ІЧ-спектрограмі ближньої інфрачервоної області (від 3900 до 7400 cm^{-1}) досліджуваних зразків торфу Львівської області наведена на рис. 9.

З наведеного вище випливає, що різноманіття методів попередньої обробки спектральних даних дозволяє компенсувати шуми, аберації та окремі помилки вимірювання та отримати для подальшого опрацювання надійні числові послідовності з мінімальними втратами інформації.

Калібраційні моделі. Модуль багатовимірного калібрування (*Multivariate Calibration*) програмного забезпечення *Chemoface* виконує багатовимірне калібрування з використанням множинної лінійної регресії (*Multiple Linear Regression, MLR*), регресії головних компонентів (*Principal Component*

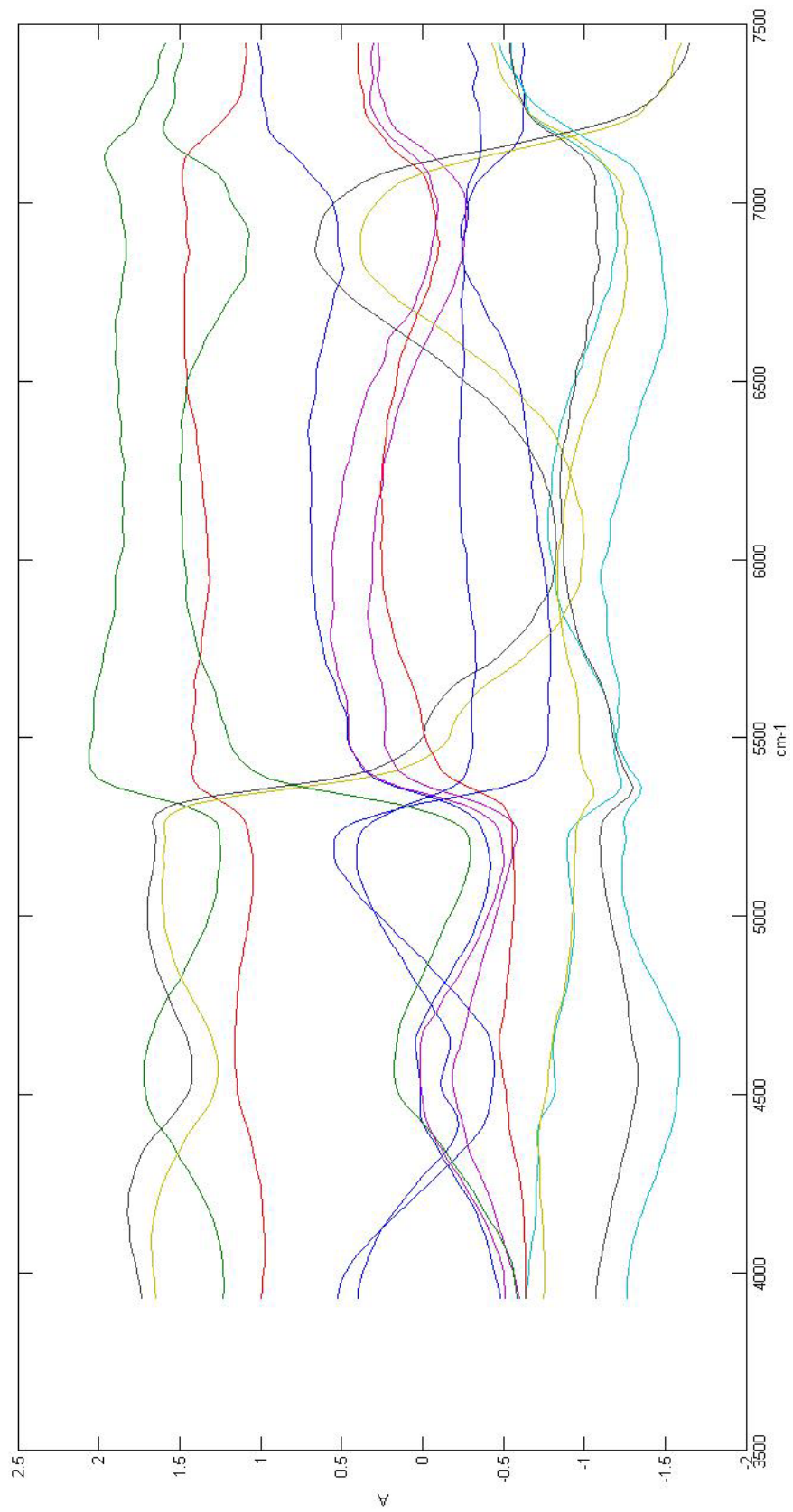


Рис. 2. Попередня обробка даних для корекції спектральних варіацій досліджуваних зразків торфу Львівської області за допомогою автомасштабування (*Autoscale*)

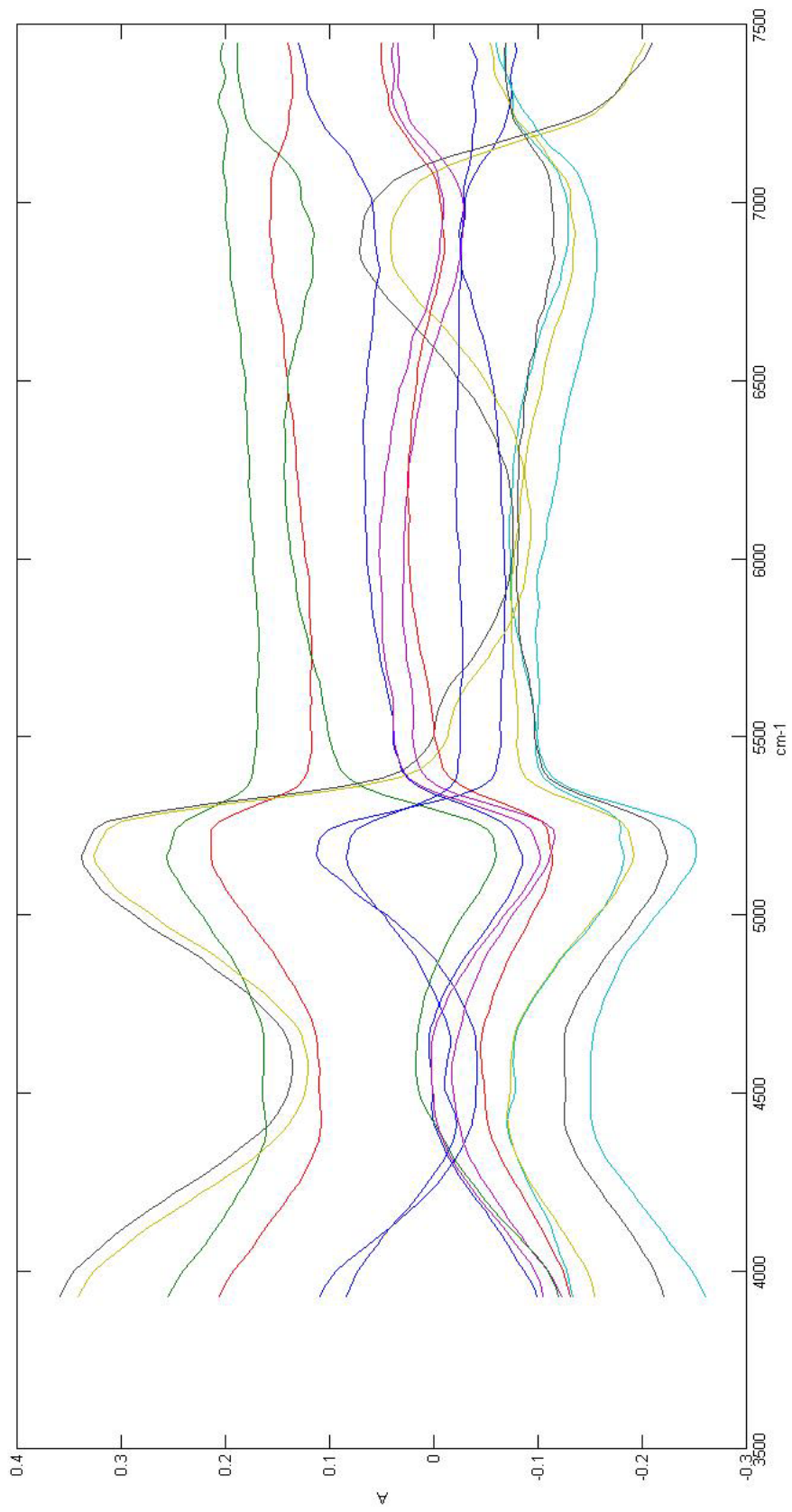


Рис. 3. Попередня обробка даних для корекції спектральних варіацій/змін досліджуваних зразків торфу Львівської області за допомогою середнього значення (*Mean*)

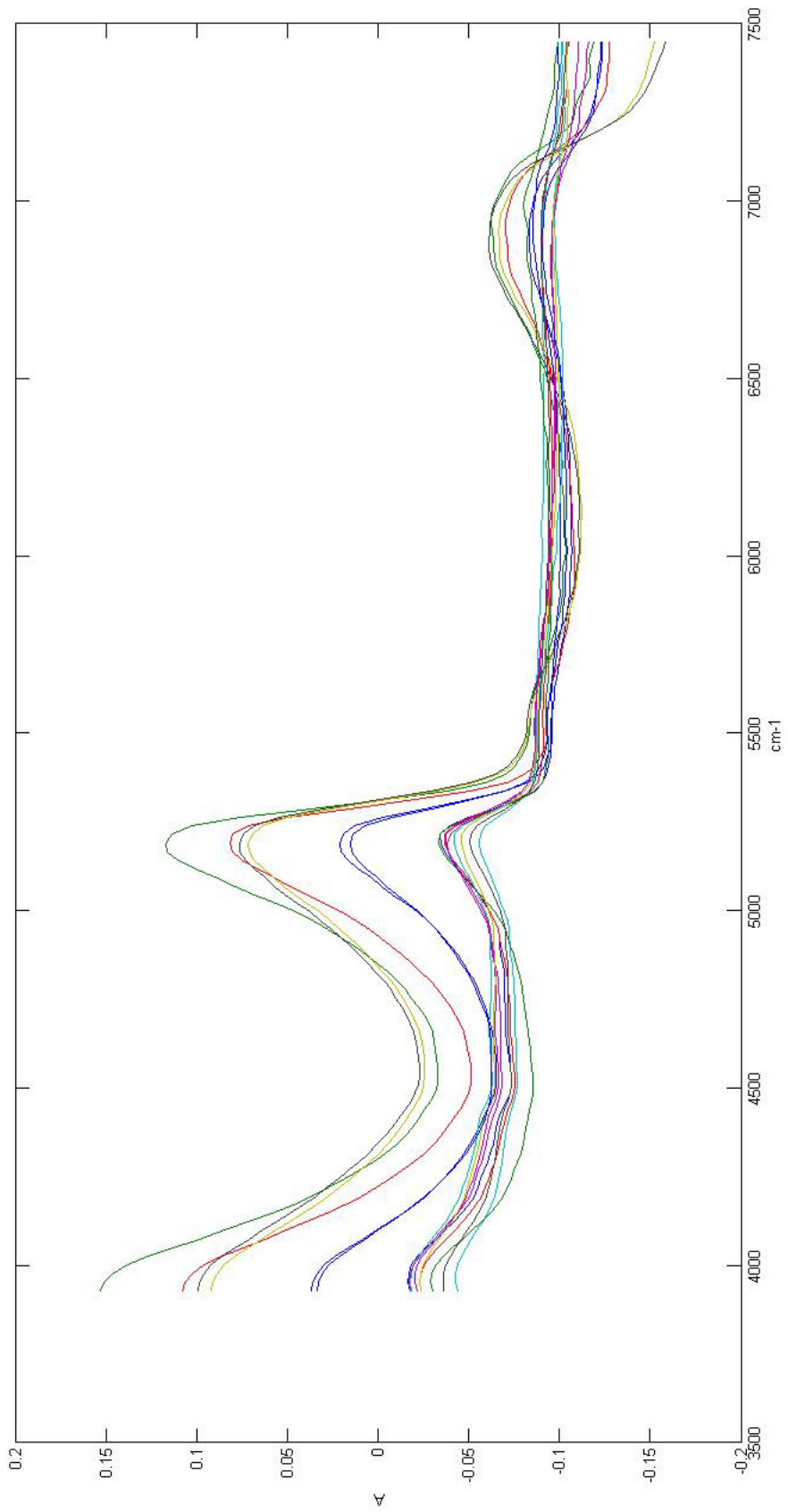


Рис. 4. Попередня обробка даних для корекції спектральних варіацій досліджуваних зразків торфу Львівської області за допомогою нормалізації діапазону (*Range normalization*)

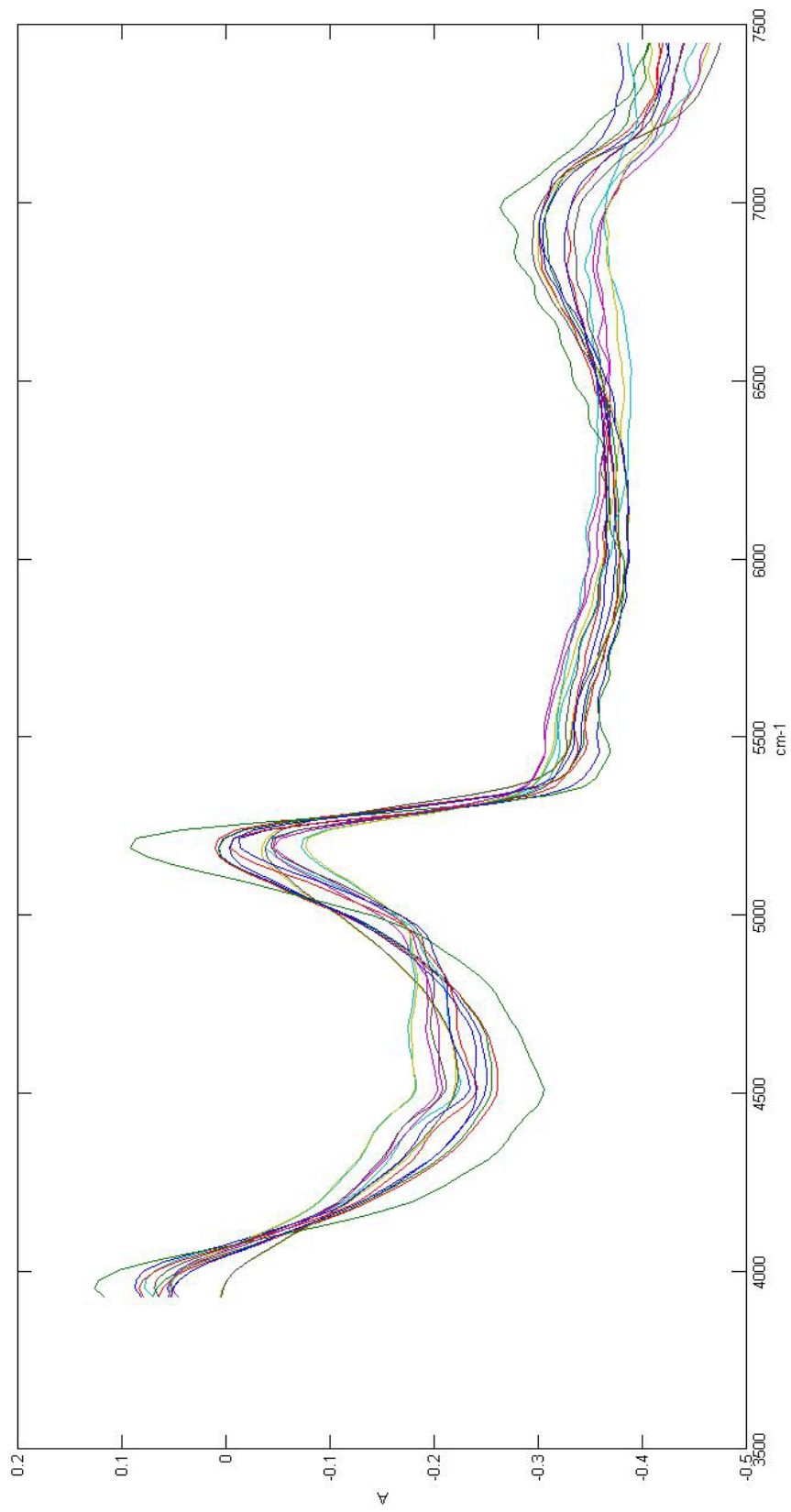


Рис. 5. Попередня обробка даних для корекції спектральних варіацій досліджуваних зразків торфу Львівської області за допомогою мультиплікативної корекції розсіювання (*Multiplicative Scatter Correction, MSC*)

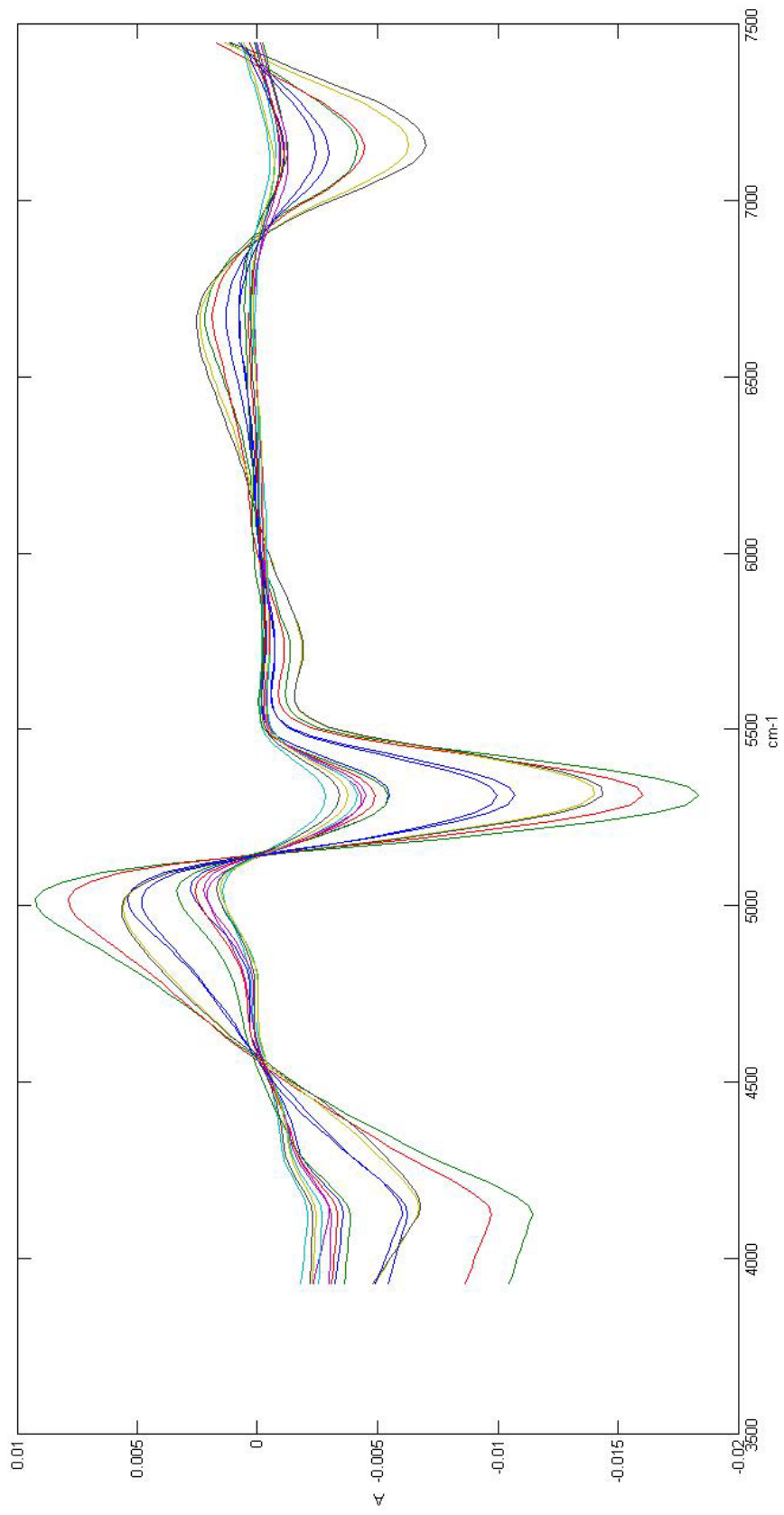


Рис. 6. Попередня обробка даних для корекції спектральних варіацій/змін досліджуваних зразків торфу Львівської області за допомогою нормалізації та першої похідної (*First derivatives*)

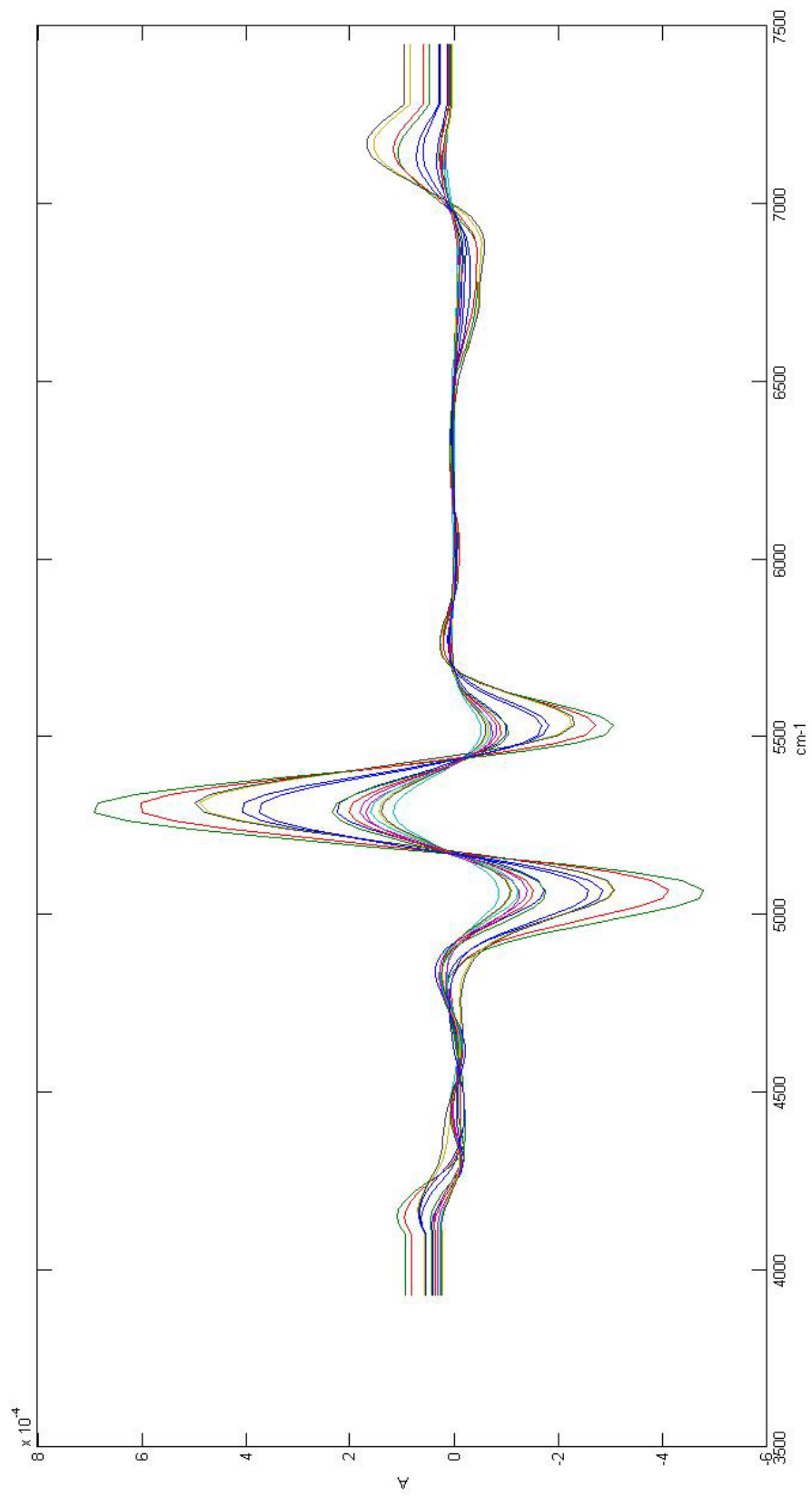


Рис. 7. Попередня обробка даних для корекції спектральних варіацій/змін досліджуваних зразків торфу Львівської області за допомогою нормалізації та другої похідної (*Second derivatives*)

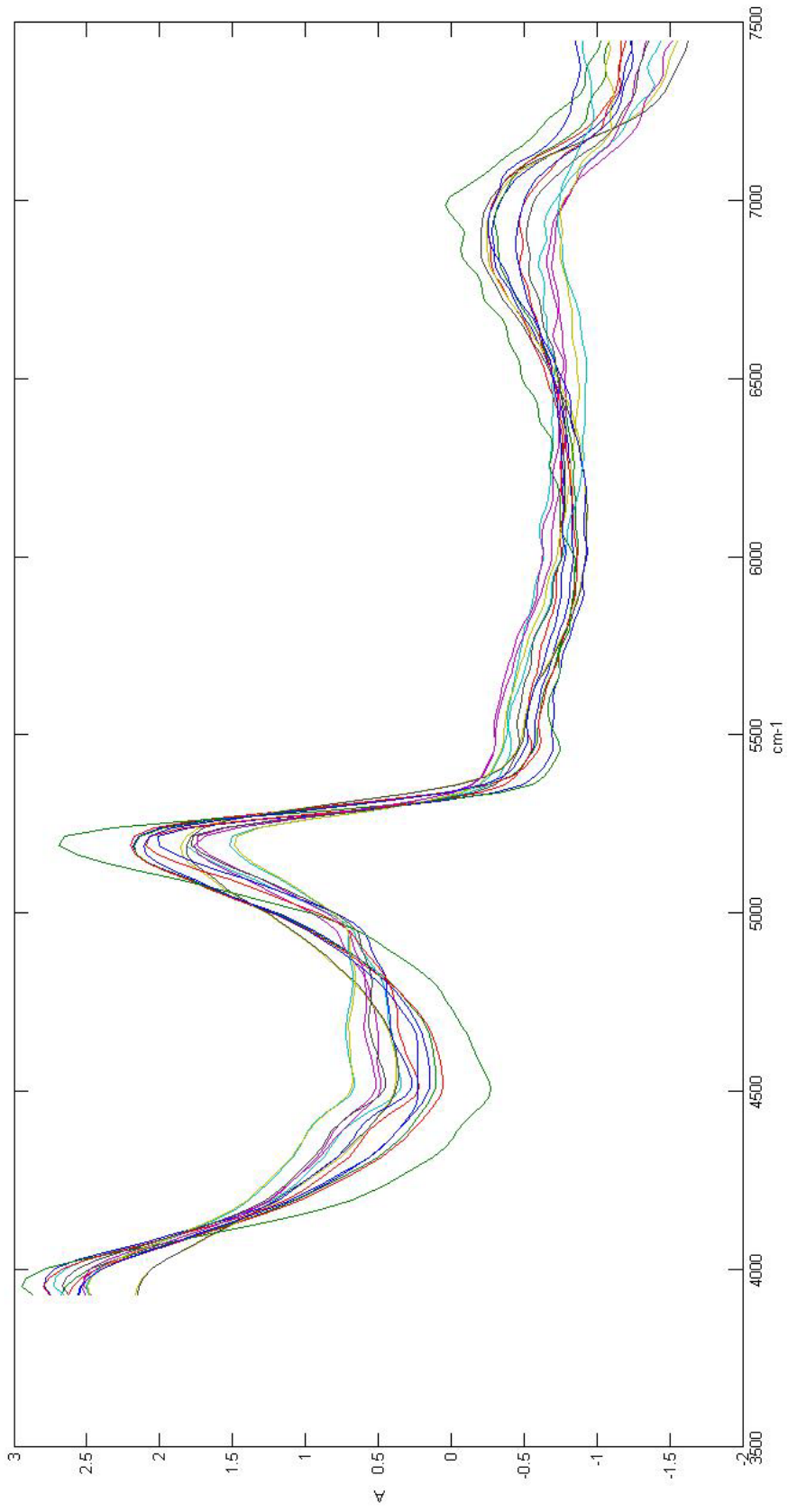


Рис. 8. Попередня обробка даних для корекції спектральних варіацій/змін досліджуваних зразків торфу Львівської області за допомогою стандартної нормальної варіації (*Standard Normal Variate, SNV*)

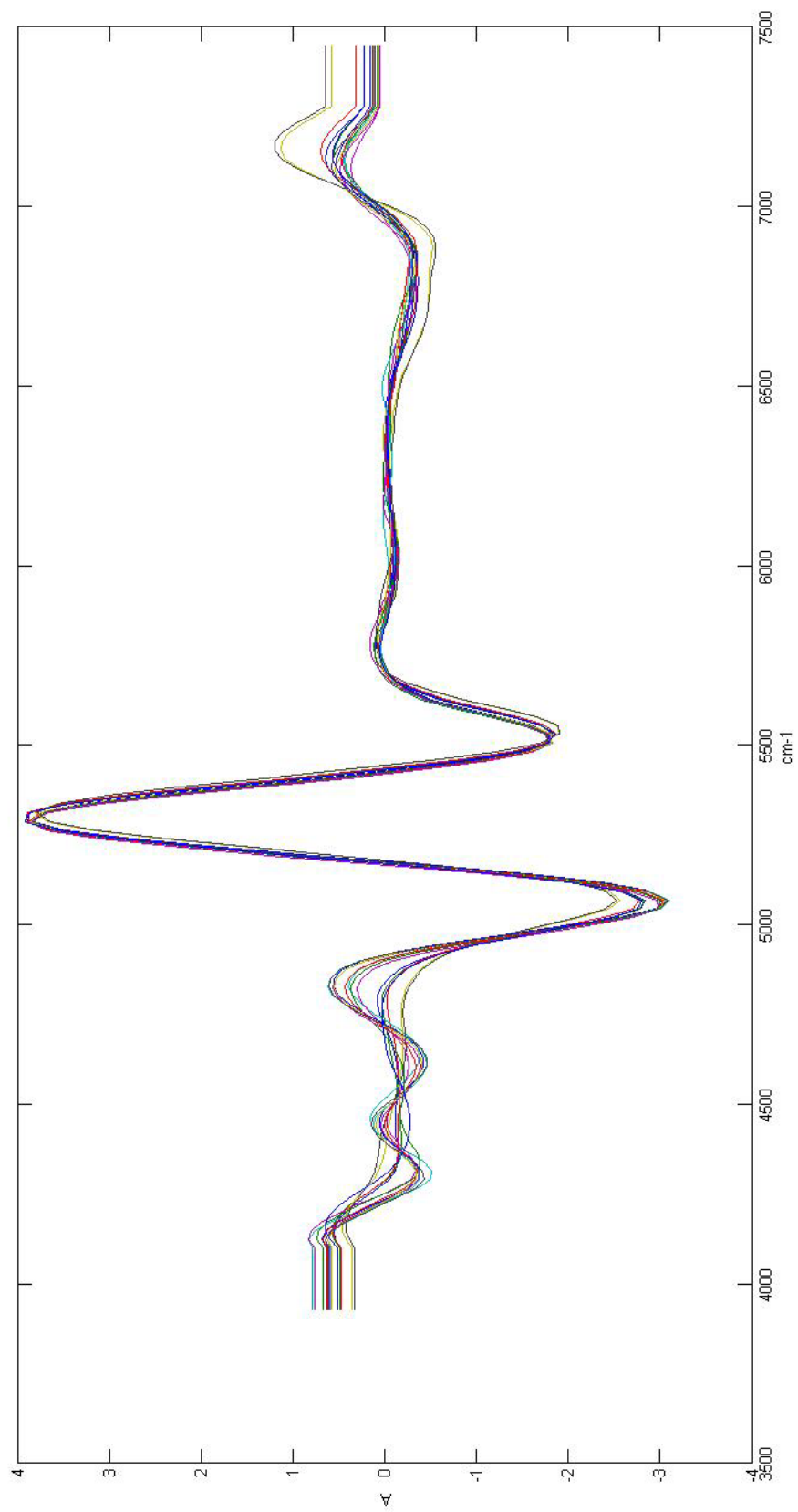


Рис. 9. Попередня обробка даних для корекції спектральних варіацій досліджуваних зразків торфу Львівської області за допомогою нормалізації, першої та другої похідної (*First and Second derivatives*) та стандартної нормальної варіації (*Standard Normal Variate, SNV*)

Regression, PCR), регресії часткових найменших квадратів (*Partial Least Squares Regression, PLS*) та моделювання для класифікації за допомогою дискримінантного аналізу (*PLS-DA, PCR-DA* і *MLR-DA*). Також виконано перехресну перевірку з виключенням по одному (*Leaveone-out, LOO-CV*). Параметри продуктивності для моделей, такі як середньоквадратична помилка (*Root Mean Square Error, RMSE*) та коефіцієнт кореляції (*Correlation Coefficient, R²*), розраховуються для перехресної перевірки, калібрування та тестових наборів.

Щоб знайти будь-який вид лінійного зв'язку між змінними, було проведено крос-кореляцію між довжинами хвиль для спектральних даних. У спектрах торфу в ближньому ІЧ-діапазоні є високий рівень колінеарності, тому прості багатовимірні методи, такі як множинна лінійна регресія (*Multiple Linear Regression, MLR*), не придатні для отримання рішення. Отже, можна вивчати методи калібрування – *регресія головних компонентів (Principal Component Regression, PCR)* або *часткових найменших квадратів (Partial Least Squares, PLS)*, віддаючи перевагу останньому методу, оскільки він враховує залежну змінну в коваріаційній матриці.

Для розробки методу калібрування обрано процедуру повної перехресної перевірки. Значення середньоквадратичної похибки калібрування (*The root mean squared error of calibration, RMSEC*) та прогнозу (*The root mean squared error of prediction, RMSEP*), отримані цим методом, вважаються оптимістичними та достатніми для порівняння моделей.

Виявлення викидів під час калібрування є важливим кроком для підвищення спроможності (передбачувальної здатності) прогнозування розрахункових оціночних коефіцієнтів калібрування. Щоб забезпечити вибір викидів, розглянуто графіки оцінок аналізу головних компонентів, зразків з високою дисперсією (плече/важіль, *leverage* – міра того, наскільки далеко значення незалежної змінної спостереження від значень інших спостережень, ступінь впливу, яку окремий вимір/запис має на рівняння регресії) та ст'юдентизованими залишковими значеннями ± 3 (рис. 10–12) і було підтверджено відсутність викидів у наборі даних. Система виявлення викидів має тенденцію вибирати унікальні вибірки серед усього набору.

RMSE (Root Mean Squared Error of Calibration, середньоквадратична похибка калібрування) досліджуваних характеристик торфу для перехресної перевірки підтверджують сім латентних змінних (*LV – Latent variables*, латентні змінні – приховані змінні, які не спостерігаються безпосередньо, а виводяться за допомогою математичної моделі і являють собою «спільну» дисперсію, або ступінь, до якого змінні «рухаються» разом) як відповідну кількість компонентів PLS (рис. 13–16).

Проаналізовані похибки отримані після застосування *методу часткових найменших квадратів (Partial Least Squares, PLS)* для оригінального спектру даних, де попередня обробка не застосовувалася, а також для наступних попередніх обробок: зміщення базової лінії, нормалізація діапазону, базова лінія/нормалізація, мультиплікативна корекція розсіювання, перша та друга похідні. Це встановлювалося для кожної досліджуваної характеристики.

Зазначимо, що застосування *автомасштабування (Autoscale)*, *середнього значення (Mean)*, *першої та другої похідних (First and Second derivatives)*

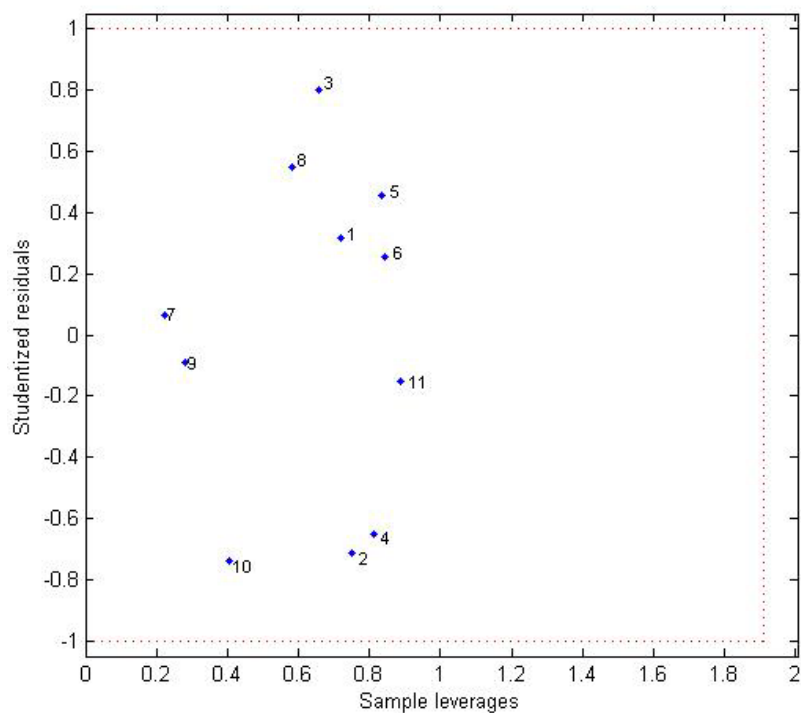


Рис. 10. Тест на виявлення викидів (вологість торфу). Графік оцінки зразків/вибірки: плече/важіль (leverage) \times ст'юдентизовані залишкові значення (Studentized residuals)

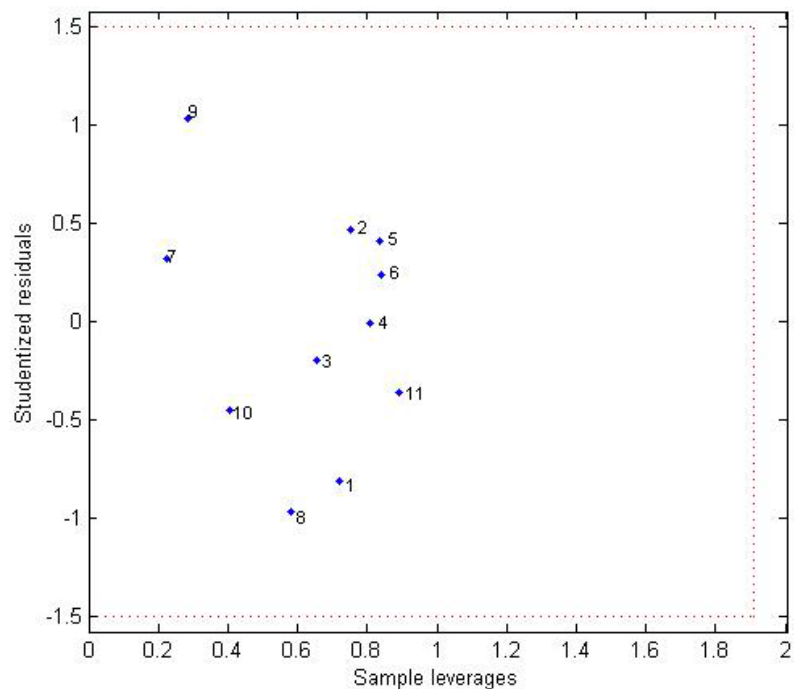


Рис. 11. Тест на виявлення викидів (вміст зольного (неорганічного) залишку торфу). Графік оцінки зразків/вибірки: плече/важіль (leverage) \times ст'юдентизовані залишкові значення (Studentized residuals)

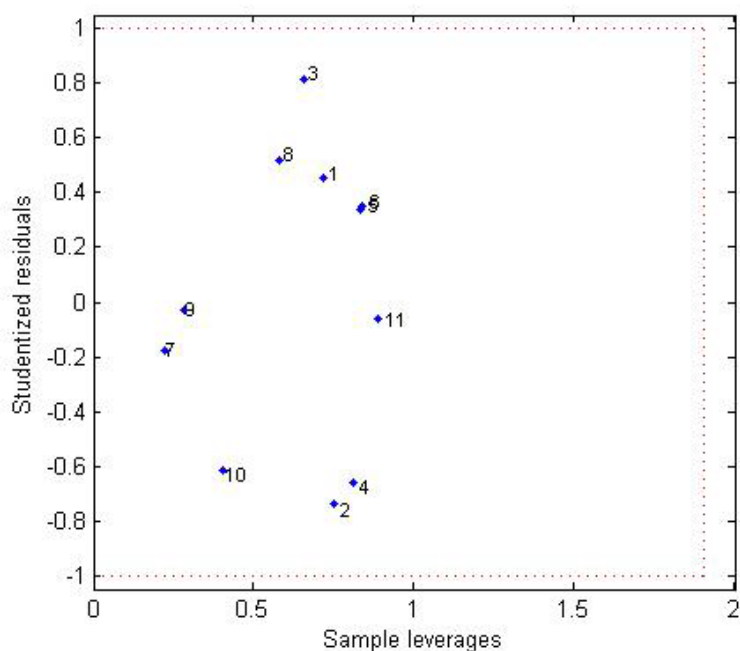


Рис. 12. Тест на виявлення викидів (кислотність (pH) торфу). Графік оцінки зразків/вибірки: плече/важіль (leverage) \times ст'юдентизовані залишкові значення (Studentized residuals)

для корекції зміщення і дрейфів базової лінії не тільки не покращує похибку, отриману без будь-якої попередньої обробки, але й погіршує, тому вони вважаються непридатними як попередня обробка, натомість нормалізація діапазону (*Range normalization*), мультиплікативна корекція розсіювання (*Multiplicative Scatter Correction, MSC*) та стандартна нормальна варіація (*Standard Normal Variate, SNV*) придатні як попередня обробка. Найкращою попередньою обробкою виявилася нормалізація діапазону (*Range normalization*) + перша та друга похідні (*First and Second derivatives*) + стандартна нормальна варіація (*Standard Normal Variate, SNV*) – спостережувані похибки є нижчими, ніж отримані без будь-якої модифікації спектрів.

Отже, застосування в комплексі: нормалізація діапазону (*Range normalization*) + перша та друга похідні (*First and Second derivatives*) + стандартна нормальна варіація (*Standard Normal Variate, SNV*), забезпечує найкращі результати, оскільки робить їх незалежними від ваги зразка, покращуючи кореляцію з відносними властивостями.

Графіки вимірених та передбачених властивостей досліджуваних характеристик торфу для навчального, перехресного контролю та тестового набору свідчать про хорошу передбачувану здатність моделі *PLS* (рис. 17–19).

Висновки:

1. Застосування модулів програмного комплексу *Chemoface* з використанням методу *PLS* для кількісного аналізу техніко-хімічних властивостей торфу на основі даних спектроскопії в ближньому інфрачервоному діапазоні (NIR) дозволяє отримати калібраційні моделі для швидкого та достовірного аналізу сировини, зокрема в польових умовах.

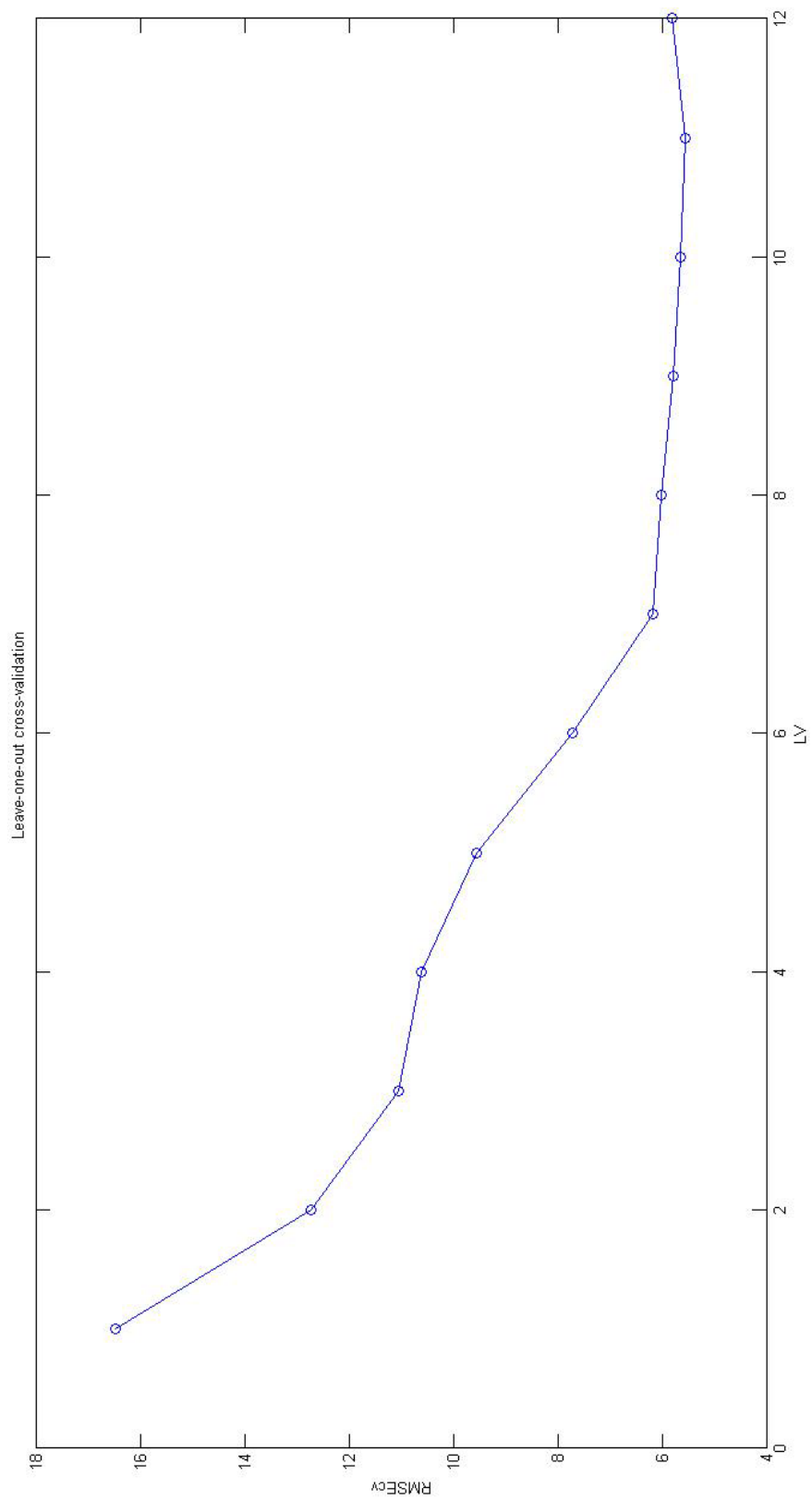


Рис. 13. RMSE (Root Mean Squared Error of Calibration, середньоквадратична похибка калібрування) у перехресній перевірці моделі на основі PLS (вологість торфу)

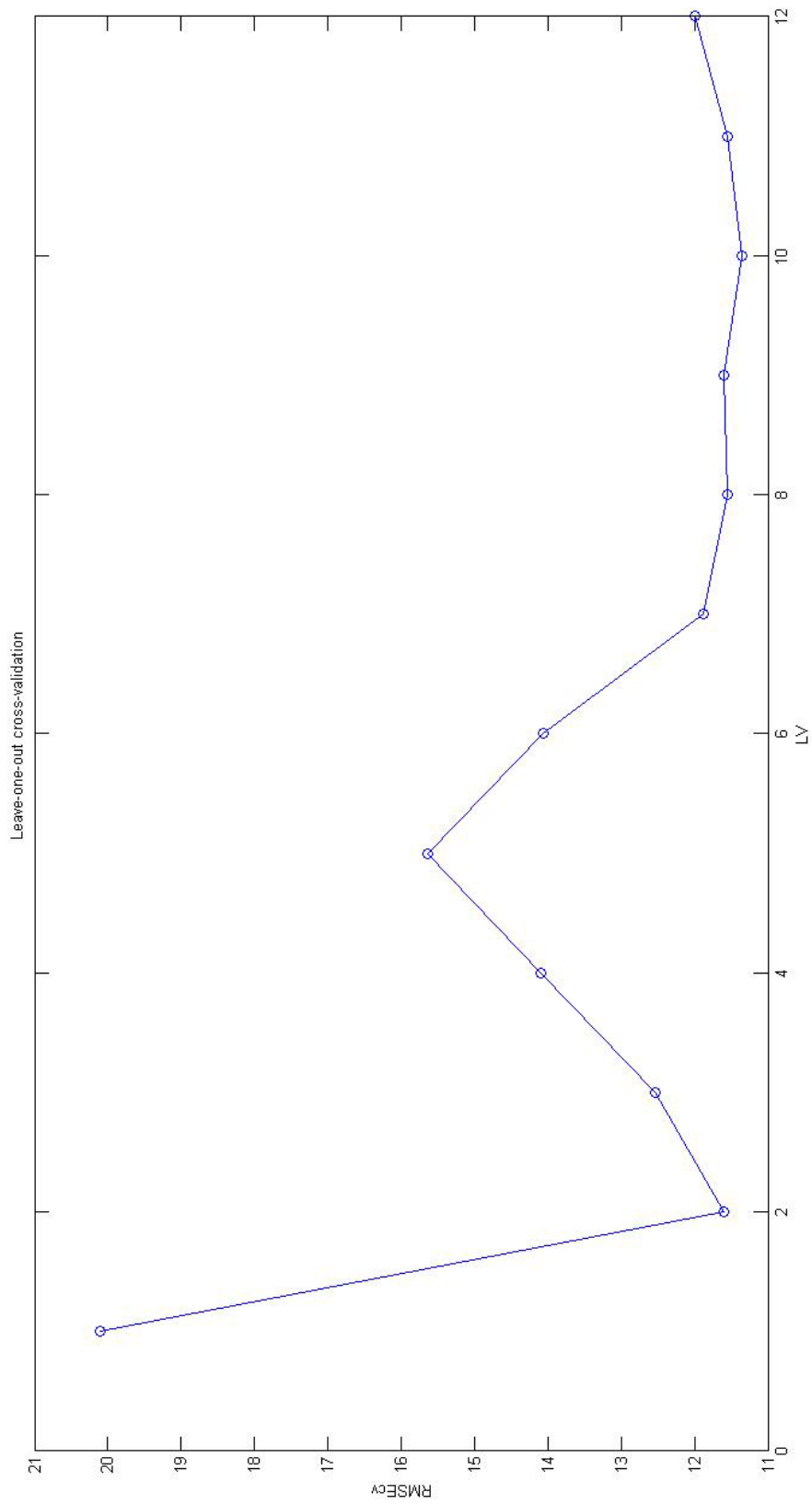


Рис. 14. RMSE (Root Mean Squared Error of Calibration, середньоквадратична похибка калібрування) у перехресній перевірці моделі на основі PLS (вміст зольного (неорганічного) залишку торфу)

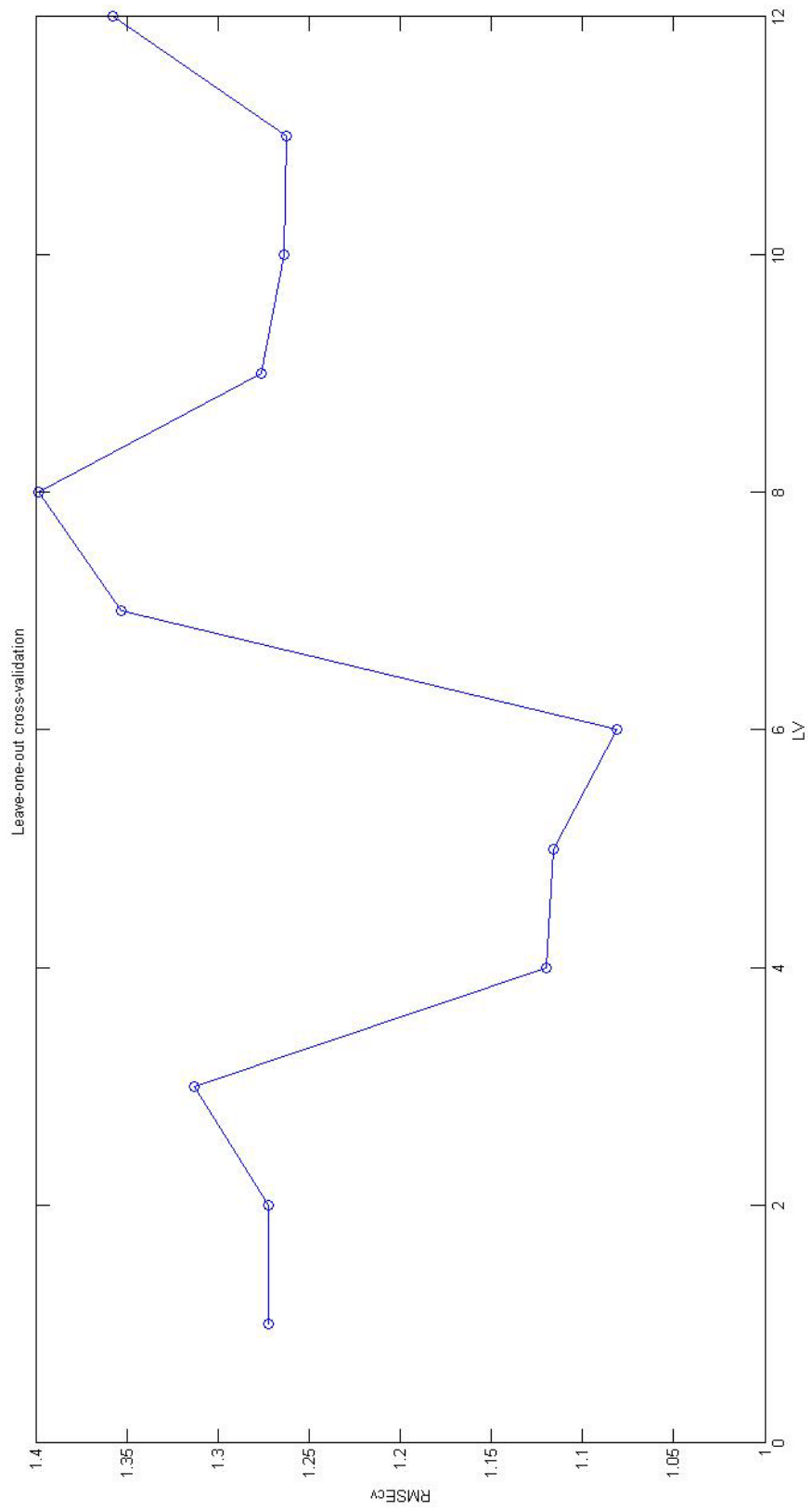


Рис. 15. RMSE (Root Mean Squared Error of Calibration, середньоквадратична похибка калібрування) у перехресній перевірці моделі на основі PLS (кислотність (pH) торфу)

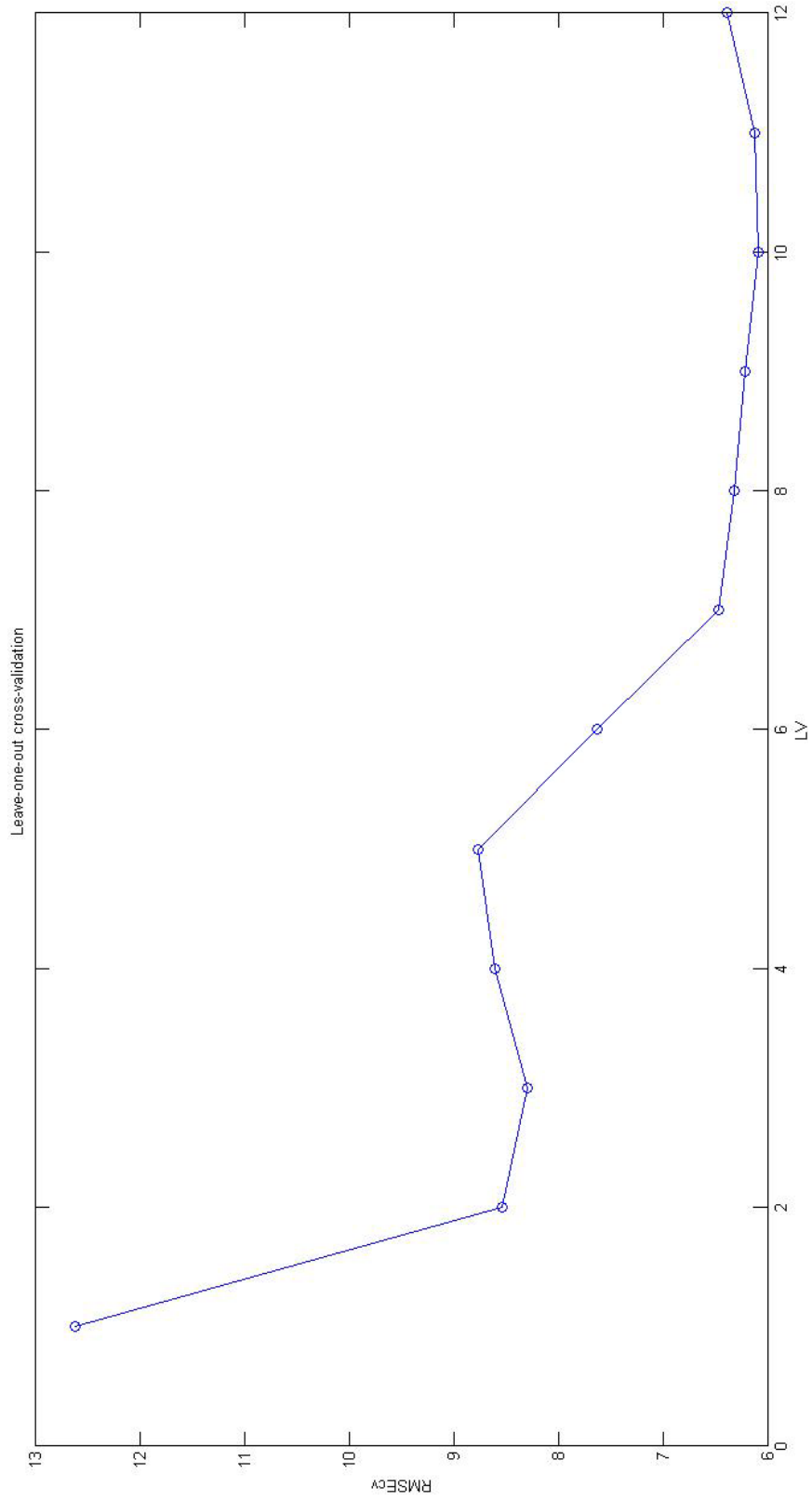


Рис. 16. RMSE (Root Mean Squared Error of Calibration, середньоквадратична похибка калібрування) у перехресній перевірці моделі на основі PLS (усереднені значення досліджуваних характеристик торфу)

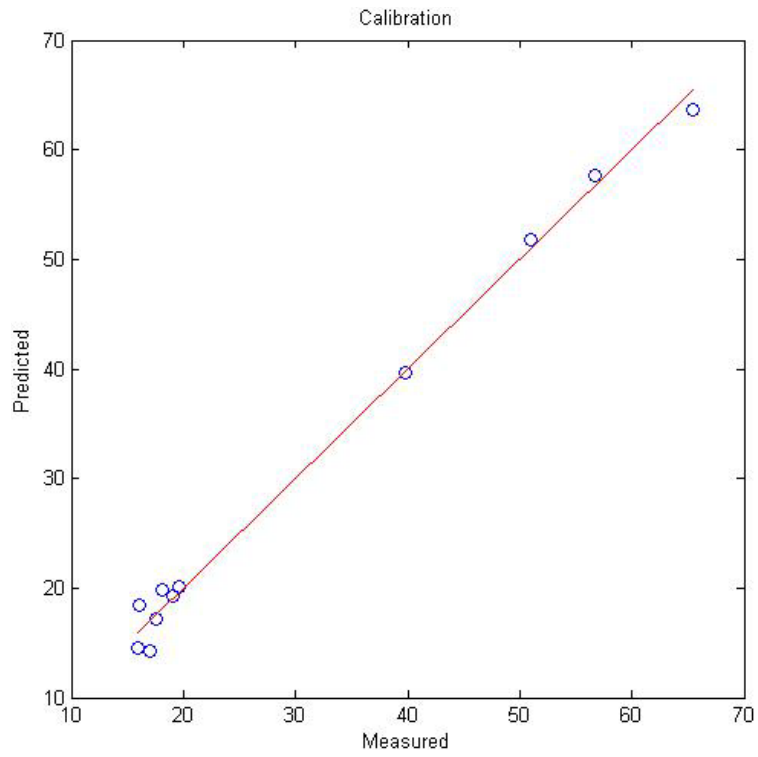


Рис. 17. Багатофакторне калібрування для моделі на основі *PLS* (вологість торфу)

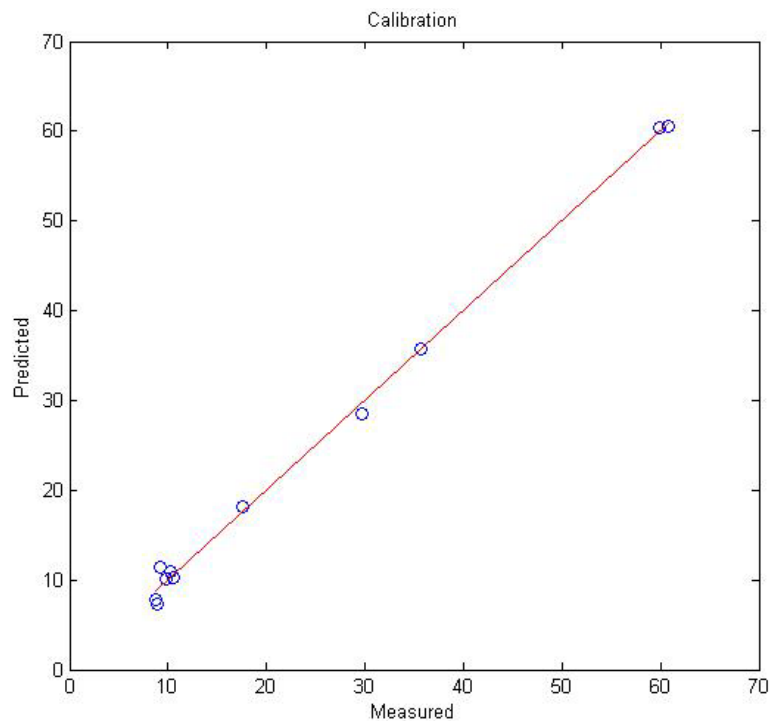


Рис. 18. Багатофакторне калібрування для моделі на основі *PLS* (вміст зольного (неорганічного) залишку торфу)

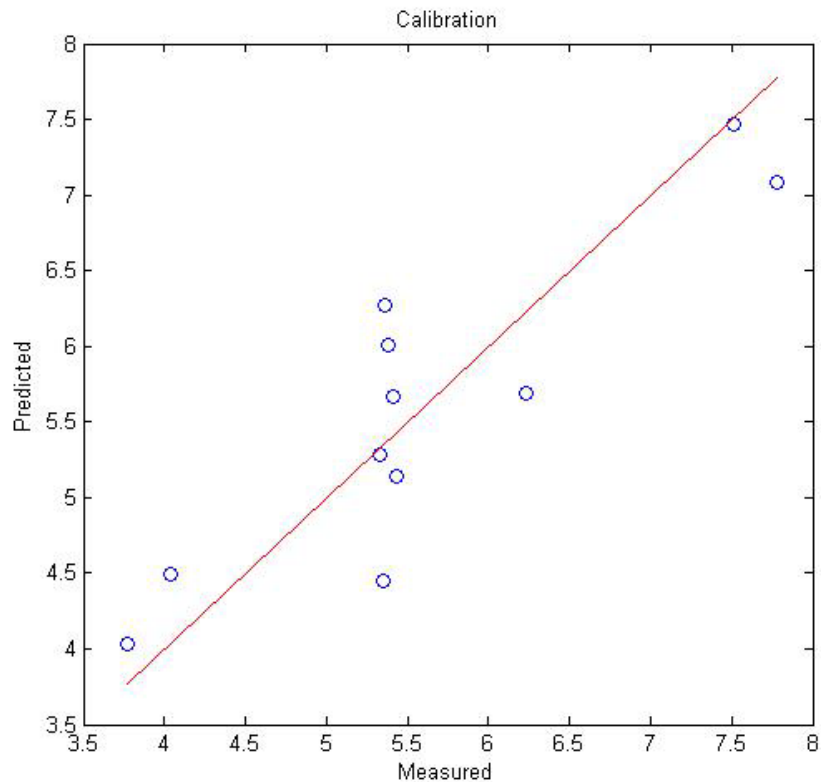


Рис. 19. Багатофакторне калібрування для моделі на основі *PLS* (кислотність (*pH*) торфу)

2. Дослідження показали, що за допомогою спектрометра, який аналізує відбите випромінювання в ділянці ближнього інфрачервоного спектра, та на основі усередненої спектральної характеристики відбитого випромінювання і за допомогою хемометричного програмного забезпечення можливо обчислювати хіміко-технологічні характеристики торфу. Процес аналізу складається з таких етапів: відбір зразка, який репрезентує всю партію сировини; опромінення зразка випромінюванням, яке містить значну частку енергії в ближній інфрачервоній ділянці; аналіз детектором відбитого (поглиненого) випромінювання та побудова інтегральної спектральної характеристики зразка; складання калібраційної моделі; обробка отриманого спектра хемометричним програмним забезпеченням з подальшим обчисленням якісних та кількісних характеристик сировини.

3. Запропонований спосіб (експрес-аналіз) швидкого визначення якісних та кількісних характеристик викопної вуглецевої сировини органічного походження, а саме низинного та верхового торфу різного ступеня розкладання, може бути використаний при встановленні його відповідності чинним нормам, стандартам і технічним умовам щодо вологості, вмісту зольного (неорганічного) залишку та кислотності (*pH*).

- Інститут ґрунтознавства та агрохімії імені О. Н. Соколовського Української академії аграрних наук. (2008). *Меліоранти ґрунту та середовища росту. Готування проб до хімічного та фізичного аналізу, визначення вмісту сухої речовини, вмісту вологи та лабораторно ущільненої насипної щільності (EN 13040:1999, IDT) (ДСТУ EN 13040:2005).*
- Інститут сільськогосподарської мікробіології та агропромислового виробництва НААН. (2016). *Торф і продукти його перероблення для сільського господарства. Методи визначення обмінної й активної кислотності (ДСТУ 7882:2015).*
- Супрунович, С. В., Кормош, Ж. О., & Сливка, Н. Ю. (2022). *Статистичні та хемометричні методи в хімії: навчальний посібник для студентів вищих навчальних закладів.* Луцьк: ВНУ імені Лесі Українки.
- Технічний комітет стандартизації «Ґрунтознавство». (2016). *Якість ґрунту. Визначення зольності торфу і торфового ґрунту (ДСТУ 7942:2015).*
- Яковенко, М., & Хоха, Ю. (2024). Використання методів інфрачервоної спектроскопії для дослідження торфу (родовище Гончари, Львівська область). *Геологія і геохімія горючих копалин*, 1–2(193–194), 113–129. <https://doi.org/10.15407/ggcm2024.193-194.113>
- Andrés, J. M., & Bona, M. T. (2005). Analysis of coal by diffuse reflectance near-infrared spectroscopy. *Analytica chimica acta*, 535(1–2), 123–132. <https://doi.org/10.1016/j.aca.2004.12.007>
- Geladi, P., MacDougall, D., & Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied spectroscopy*, 39(3), 491–500. <https://doi.org/10.1366/0003702854248656>
- McClure, W. F. (1994). Near-infrared spectroscopy: The giant is running strong. *Analytical chemistry*, 66(1), 42A–53A. <https://doi.org/10.1021/ac00073a730>
- Mostert, M. M., Ayoko, G. A., & Kokot, S. (2010). Application of chemometrics to analysis of soil pollutants. *TrAC Trends in Analytical Chemistry*, 29(5), 430–445. <https://doi.org/10.1016/j.trac.2010.02.009>
- Nunes, C. A., Freitas, M. P., Pinheiro, A. C. M., & Bastos, S. C. (2012). Chemoface: a novel free user-friendly interface for chemometrics. *Journal of the Brazilian Chemical Society*, 23(11), 2003–2010. <https://doi.org/10.1590/S0103-50532012005000073>

Стаття надійшла:
20.10.2024 р.

Yurii KHOKHA¹, Myroslava YAKOVENKO²

Institute of Geology and Geochemistry of Combustible Minerals
of National Academy of Sciences of Ukraine, Lviv, Ukraine,
e-mail: ¹khoha_yury@ukr.net; ²myroslavakoshil@ukr.net

**USE OF CHEMOMETRIC METHODS AND REGRESSION MODELS
IN PROCESSING NIR SPECTRA OF PEAT
FOR QUANTITATIVE DETERMINATION
OF ITS CHEMICAL AND TECHNOLOGICAL INDICATORS**

The article discusses theoretical and practical aspects of the use of near infrared (NIR) spectroscopy combined with chemometrics for express analysis of peat. Near infrared spectroscopy provides a significant amount of information about complex organic systems, including irregular polymers such as peat. Compared to classical analytical methods, NIR spectrometry allows analysis without complex sample preparation with analysis time

measured in minutes. Since the results represent the intensity of radiation reflection in the overtone range of fundamental frequencies, their processing requires the use of special mathematical and statistical methods. The use of the Chemoface software package modules (PLS method) for quantitative analysis of the technical and chemical properties of peat based on NIR spectroscopy data has demonstrated the possibility of obtaining calibration models that allow for the quick and reliable analysis of this raw material, including in field conditions. The conducted studies have shown that using a spectrometer that analyzes reflected (absorbed) radiation in the near-infrared spectrum and based on the averaged spectral characteristics of the reflected (absorbed) radiation and using chemometric software, it is possible to calculate the chemical and technological characteristics of peat. The analysis procedure consists of the following stages: selection of a sample representing the entire batch of raw materials; irradiation of the sample with radiation containing a significant proportion of energy in the near-infrared spectrum; analysis with a detector of reflected (absorbed) radiation and construction of an integral spectral characteristic of the sample; compilation of a calibration model using chemometric software; processing of the obtained spectrum using chemometric software with subsequent calculation of the qualitative and quantitative characteristics of the raw materials. The proposed method (express analysis) for rapid determination of qualitative and quantitative characteristics of fossil carbon raw materials of organic origin, namely lowland and highland peat of various degrees of decomposition, can be used to establish its compliance with current norms, standards and technical conditions for moisture content, ash (inorganic) residue content and acidity (*pH*).

Keywords: near-infrared reflectance (NIR) spectroscopy, peat analysis, predictive models, multivariate analysis, Partial Least Squares Regression (PLS), pre-treatments effect.