

СПОСІБ ПЕРЕВІРКИ УЗГОДЖЕНОСТІ КЛАСИФІКОВАНИХ ДАНИХ ПІД ЧАС ОЦІНЮВАННЯ ПОКАЗНИКІВ ЯКОСТІ ҐРУНТУ

Семенюк Р.,

e-mail: SemeniyukRoman@i.ua,
<https://orcid.org/0000-0001-5139-6330>,

Яремчук Н., д-р. техн. наук, проф.,
<https://orcid.org/0000-0003-0661-1243>

НТУУ «КПІ ім. І.Сікорського»

Гусар І.,

email: gusaririna19@gmail.com,
<https://orcid.org/0000-0002-58724672>

ДНУ «УкрНДІПВТ ім. Л. Погорілого»

Анотація

Мета роботи - розглянути спосіб побудови шкали класифікації показників якості ґрунту з урахуванням невизначеності аналізу проб ґрунту. Як характеристику якості класифікації обрано матрицю відповідності шкали. Запропонувати спосіб перевірки узгодженості класифікованих вибіркового даних за інформаційною мірою узгодженості з застосуванням граничної міри узгодженості, яку отримують на основі матриці відповідності шкали класифікації.

Методи досліджень: теоретичні – аналіз і синтез інформаційних ресурсів та аналітичні – побудова математичної моделі способів перевірки узгодженості класифікованих даних під час оцінювання показників якості ґрунту.

Результати. Для оцінювання якості ґрунтів за окремими показниками родючості використовуються впорядкована шкала класифікації, тобто впорядкований ряд рівнів якості, які можуть бути використані під час моніторингу земельних ресурсів, для поліпшення якості ґрунтів внесенням добрив, у визначенні придатності земель для певних потреб.

Першим етапом визначення якості ґрунтів є отримання проб ґрунту за прийнятою програмою, другим – визначення показників якості, третім – прийняття відповідних рішень. Для реалізації другого етапу необхідна шкала класифікації, яка заснована на процедурі аналізу проб і тому є індивідуальною для кожного способу аналізу і, крім того, повинна враховувати невизначеність вимірювання показників якості, що є досить значною для всіх способів аналізу. Тому в роботі використано, як важливу характеристику шкали, матрицю відповідності між сферами визначення окремих класів еквівалентності шкали і отриманими результатами класифікації, в комірках якої знаходяться ймовірності правильного і неправильного віднесення до певних класів еквівалентності.

Висновки. Отримані вибірки класифікованих даних підлягають подальшому опрацюванню. Як показав аналіз літературних джерел, вибірки аналізують для визначення кореляції або узгодженості з опорними даними, перевірки розсіювання, збіжності вибіркового даних або їх узгодженості. Наявність такої перевірки сприяє впевненості у прийнятті подальших рішень. Тому в роботі запропоновано спосіб перевірки вибірки класифікованих даних із використанням міри узгодженості або збіжності класифікованих даних із застосуванням прийнятої шкали класифікації. Оскільки невизначеність вимірювання приводить до розсіювання даних відносно центру вибірки, то для визначення границі можливої неузгодженості використовується матриця відповідності шкали класифікації.

Ключові слова: шкала класифікації, міра узгодженості, клас еквівалентності.

Постановка проблеми. Визначення та контроль стану родючості ґрунтів або якості земельних ділянок потрібні для виконання таких завдань як тривалий моніторинг земельних ресурсів, поліпшення якості ґрунтів внесенням добрив, визначення придатності земель для оптимізації різних способів використання тощо [1]. Для виконання цих завдань можна виділити такі етапи:

- отримання проб ґрунту за детально визначеною програмою;
- визначення показників родючості ґрунтів за стандартизованими методиками;
- прийняття рішень щодо стану якості ґрунту або способів поліпшення стану ґрунту за внесенням необхідної кількості добрив.

Точність визначення запасів поживних речовин значною мірою залежить від площі елементарної ділянки — найменшої площі, яку можна охарактеризувати однією об'єднаною пробою ґрунту.

На розмір елементарної ділянки впливає тип ґрунту, кількість внесених добрив тощо[2].

Предмет дослідження. Для отримання необхідних даних використовується шкала класифікації, яка дозволяє представити дані, отримані за метричною шкалою, як категоризовані дані, що є певними рівнями лінгвістичної змінної «показник якості ґрунту» (наприклад, «низький», «середній», «високий»).

Якщо звернутись до першого етапу, що стосується програми відбору проб, то вона, в певній мірі, пов'язана зі способом внесення добрив на досліджуваній ділянці ґрунту. Під час внесення добрив за їх розпорошенням на ділянці може бути застосований рандомізований відбір проб без застосування певної системи. Для рідких форм добрив, які вважають на теперішній час високотехнологічними, але й такими, що можуть спричинити негативні зміни, необхідна певна система отримання проб ґрунту. Ось у роботі [3] зазначено, що оскільки застосування безводного аміаку можливе лише його локальним внесенням у стрічки на певну глибину, зміни,

які відбуваються у стрічках його внесення, можуть бути доволі контрастними порівняно з показниками ґрунту у міжрядді внаслідок формування осередків з високою концентрацією солей. Тому в роботі [3] було запропоновано і досліджено відбір проб за схемою зигзагоподібного перетинання, що забезпечувало однако-ву кількість проб у стрічці і у міжрядді. Як критерій вибору оптимальної програми відбору проб в роботі [3] було обрано коефіцієнт парної кореляції між вмістом рухомого фосфору та амонійного азоту у стрічці та у змішаному зразку, який складався з 20 індивідуальних проб, відібраних за схемою зигзагоподібного перетинання. Отримані значення коефіцієнту кореляції були достатньо високими у разі вибірки 20-ти проб, але зі зменшенням їхньої кількості до 10 або навпаки зі збільшенням до 30-40 не було виявлено чіткого зв'язку між показниками родючості, визначеними у стрічці внесення добрив та у зразках, відібраних за зигзагоподібною схемою. Також не було виявлено кореляції за умов рандомізованого відбору.

З розглянутого вище витікає, що, по-перше, є необхідність у перевірці узгодженості вихідних даних під час прийняття рішень щодо показників якості ґрунту, а по-друге, необхідно мати декілька критеріїв, які дають змогу перевірити узгодженість саме між класифікованими даними. Тому в представленій роботі запропоновано використання інформаційних критеріїв перевірки узгодженості класифікованих даних з використанням прийнятої для класифікації шкали.

Викладення основного матеріалу.

1. Побудова шкали класифікації показників якості ґрунту з урахуванням невизначеності вимірювання.

Показниками якості ґрунту є властивості ґрунту, які мають певні розміри, можуть бути представлені числовими значеннями і посиленням. Це посилення за VIM-2008 [4] може бути одиницею вимірювання, вимірювальною процедурою, еталонним матеріалом або їх комбінацією. Відповідно до цього шкали показників

якості ґрунту пов'язані з процедурою, яка використовується під час їх вимірювання. Наприклад, визначення рухомих сполук фосфору в ґрунті може виконуватись за методом Кірсанова (ДСТУ 4405:2005) і за методом Мачигіна (ДСТУ 4114:2002). Відповідні цим методам шкали будуть відрізнятись. Тому ілюстрацією побудови шкали в цій роботі буде шкала, заснована на визначенні рухомих сполук фосфору за методом Мачигіна. За цим методом лінгвістичні характеристики терм-множини шкали «вміст рухомого фосфору в пробах ґрунту за методом Мачигіна в мг/кг» такі: T_1 – «низький» (менше 15), T_2 – «середній» (16-30), T_3 – «підвищений» (31-45), T_4 – «високий» (46-60), T_5 – «дуже високий» (більше 60). Зважаючи на відносну похибку вимірювання вмісту фосфору, встановлену як границю довірчого інтервалу з ймовірністю 0.95, що становить 30% для вмісту P_2O_5 до 15 мг/кг і 20% для вмісту більше 15 мг/кг [5], було побудовано шкалу для нечіткої лінгвістичної змінної з використанням трапецієподібних і трикутних функцій приналежності окремих термів [6]:

$$\mu_{T_1}(x) = \begin{cases} 1, & \text{для } 0 \leq x \leq 12,5, \\ (18,5 - x) / 6 & \text{для } 12,5 < x \leq 18,5, \\ 0 & \text{в інших випадках} \end{cases}$$

$$\mu_{T_2}(x) = \begin{cases} (x - 12,5) / 6 & \text{для } 12,5 < x \leq 18,5, \\ 1 & \text{для } 18,5 < x \leq 24,5, \\ (36,5 - x) / 12 & \text{для } 24,5 < x \leq 36,5, \\ 0 & \text{в інших випадках} \end{cases}$$

$$\mu_{T_3}(x) = \begin{cases} (x - 24,5) / 12 & \text{для } 24,5 < x \leq 36,5, \\ 1 & \text{для } 36,5 < x \leq 38, \\ (53 - x) / 15 & \text{для } 38 < x \leq 53 \\ 0 & \text{в інших випадках} \end{cases}$$

$$\mu_{T_4}(x) = \begin{cases} (x - 38) / 15 & \text{для } 38 < x \leq 53, \\ (68 - x) / 15 & \text{для } 53 < x \leq 68 \\ 0 & \text{в інших випадках} \end{cases}$$

$$\mu_{T_5}(x) = \begin{cases} (x - 53) / 15 & \text{для } 53 < x \leq 68, \\ 1 & \text{для } 68 < x \leq 75 \\ 0 & \text{в інших випадках} \end{cases}$$

Враховуючи переріз функцій приналежності сусідніх термів побудовано матрицю відповідності встановленої шкали, де в комірках наведені значення, що є ймовірностями віднесення до певних класів еквівалентності P_{ij} . Діагональні значення відповідають ймовірностям правильного віднесення ($i = j$), а сусідні відповідають ймовірностям неправильного віднесення ($i \neq j$), які спричинені наявністю невизначеності вимірювання. Наприклад, для першого класу

$$\text{еквівалентності } P_{11} = \left(\int_0^{x_{кр}} \mu_{T_1}(x) dx \right) / \left(\int_0^{x_{max}} \mu_{T_1}(x) dx \right),$$

де $x_{кр}$ – критична точка функції приналежності, під якою розуміють точку зі ступенем приналежності 0,5 [7]. Для першого класу еквівалентності $x_{кр} = 15,5 \text{ мг / кг}$,

$$P_{11} = 0,952, P_{12} = 0,048.$$

Отримана матриця відповідності може бути використана за заданих границь термів множини як показник якості шкали класифікації, а також для оптимізації вибору границь, тому що, як видно з таблиці 1, в центральній частині зони визначення шкали класифікації ймовірність правильного віднесення зменшується.

Таблиця 1 – Матриця відповідності шкали класифікації

Границі значень вимірюваного показника (мг/кг)	T1	T2	T3	T4	T5
0 – 15,5	0,952	0,048	0	0	0
15,5 – 30,5	0,050	0,850	0,1	0	0
30,5 – 45,5	0	0,100	0,775	0,125	0
45,5 – 60,5	0	0	0,125	0,75	0,125
60,5 – 75	0	0	0	0,129	0,871

Якщо врахувати невизначеність вимірювання і варіацію вмісту рухомого фосфору в стрічці і за рандомізованого вибору точок відбору проб (за даними, отриманими в [3]), то можна провести моделювання з використанням отриманого середньоквадратичного відхилення і прийнятого розподілу значень отриманих проб ґрунту. Для нормального розподілу вибірки, сумарного середнього квадратичного відхилення 10% і об'єму вибірки $n=20$ було проведено моделювання в межах третього класу еквівалентності з середнім значенням вмісту рухомого фосфору 38 мг/кг. Вибір цього класу еквівалентності був обумовлений тим, що форма ФП цього класу відповідає трикутній, що зі свого боку свідчить про максимальне розсіювання. Результати класифікації були отримані за функціями приналежності $\mu_{T_i}(x)$ для двох алгоритмів роботи нечіткого класифікатора: за максимумом ступеня приналежності перерізу (стовпчик 3 табл.2) і з урахуванням всіх ступенів приналежності перерізу результату вимірювання сім'єю з ФП $\mu_{T_i}(x)$ (стовпчик 4 табл.2).

Багаторазове повторення моделювання для вибірки $n=20$ показало високу збіжність загальних результатів класифікації. Водночас у процедурі моделювання

враховано розсіювання реальних даних за рандомізованим відбором (перший стовпчик) і невизначеність вимірювання (результати класифікації за шкалою). Якщо порівняти дані моделювання з даними, наведеними в табл.1, то можна зробити висновок, що дані таблиці відповідності можуть бути використані для визначення максимальної границі міри збіжності даних відбору проб ґрунту, за яких програму відбору проб можна вважати задовільною.

2. Критерій перевірки узгодженості (збіжності) класифікованих даних.

Авторами даної роботи було запропоновано такий критерій перевірки узгодженості (збіжності) класифікованих даних за інформаційною мірою узгодженості Cns (за [8] Cns – consensus measure) або неузгодженості Dns (за [8] Dns – dissention measure), де $Cns(x) = 1 - Dns(x)$:

$$Cns(x) = 1 + \sum_{i=1}^n p_i \log_2 \left(1 - \frac{|x(T_i) - x(T_y)|}{x_{\max} - x_{\min}} \right), \quad (1)$$

де $x(T_i)$ – середина класу еквівалентності T_i за областю визначення шкали класифікації, $x(T_y)$ – середина класу еквівалентності, яка відповідає центру вибірки, p_i – оцінка ймовірності перебування в класі еквівалентності T_i .

Таблиця 2 – Результати класифікації за моделюванням процедури відбору проб за вибіркою $n=20$ за нормального розподілу результатів вимірювання

Ранжовані результати вимірювання (мг/кг)	Кількість результатів	Результат класифікації за алгоритмом 1	Результат класифікації за алгоритмом 2
28	1	T_2	$T_2 0,71; T_3 0,29$
35	1	T_3	$T_2 0,12; T_3 0,87$
36	3	T_3	$T_2 0,04; T_3 0,96$
37	3	T_3	$T_3 1$
38	3	T_3	$T_3 1$
39	2	T_3	$T_3 0,93; T_4 0,07$
40	1	T_3	$T_3 0,87; T_4 0,13$
41	1	T_3	$T_3 0,8; T_4 0,2$
42	1	T_3	$T_3 0,73; T_4 0,27$
43	2	T_3	$T_3 0,67; T_4 0,33$
44	1	T_3	$T_3 0,6; T_4 0,4$
47	1	T_4	$T_3 0,4; T_4 0,6$
51	1	T_4	$T_3 0,13; T_4 0,87$
Загальний результат нечіткої класифікації		$T_2 0,05; T_3 0,85; T_4 0,1$	$T_2 0,048; T_3 0,792; T_4 0,16$

Граничне значення Cns для заданої невизначеності можна отримати із застосуванням матриці відповідності (табл.1) за центром вибірки, що відповідає певному класу еквівалентності. Наприклад, для $T_y = T_3$, $x(T_y) = 38 \text{ мг/кг}$, отримуємо

$$(Cns)^{zp} = 1 + P_2 \log_2 \left(1 - \frac{|x(T_2) - x(T_y)|}{x_{\max} - x_{\min}} \right) + P_4 \log_2 \left(1 - \frac{|x(T_4) - x(T_y)|}{x_{\max} - x_{\min}} \right) =$$

$$= 1 + 0,1 \cdot \log_2 \left(1 - \frac{|23 - 38|}{75} \right) + 0,125 \cdot \log_2 \left(1 - \frac{|53 - 38|}{75} \right) = 0,928.$$

Якщо міра узгодженості даних вибірки більша або дорівнює $(Cns)^{zp}$, вибірка вважається узгодженою. Для даних таблиці 2 $Cns = 0,933$. Отже, ця вибірка вважається узгодженою. Якщо програма отримання проб є недосконалою, що є причиною отримання аномальних результатів, міра узгодженості даних зменшується, що є сигналом для удосконалення програми.

У таблиці 3 наведено результати моделювання процедури класифікації, за якої до нормально розподілених даних додано суміш з більшим розсіюванням.

Дані таблиці 3 не можна вважати достатньо узгодженими, про що свідчить значення міри узгодженості: $Cns = 0,866$, що менше за граничне значення, яке обчислене з урахуванням невизначеності вимірювання. Перевагою формули 1 для

міри узгодженості є те, що її можна використовувати при різних областях визначення окремих термів, тобто за нерівномірної шкали класифікації.

Першим етапом визначення якості ґрунтів є отримання проб ґрунту за прийнятою програмою, другим – визначення показників якості, третім – прийняття відповідних рішень. Для реалізації другого етапу необхідна шкала класифікації, яка заснована на процедурі аналізу проб і тому є індивідуальною для кожного способу аналізу і, крім того, повинна враховувати невизначеність вимірювання показників якості, що є досить значною для всіх способів аналізу. Тому в роботі використано, як важливу характеристику шкали, матрицю відповідності між областями визначення окремих класів еквівалентності шкали і отриманими результатами класифікації, в комірках якої знаходяться ймовірності правильного і неправильного віднесення до певних класів еквівалентності.

Висновки. Отримані вибірки класифікованих даних підлягають подальшому опрацюванню. Як показав аналіз літературних джерел, вибірки аналізують з метою визначення кореляції або узгодженості з опорними даними, перевірки розсію-

Таблиця 3 – Результати класифікації за моделюванням засміченої вибірки

Ранжовані результати вимірювання (мг/кг)	Кількість результатів	Результат класифікації за алгоритмом 1	Результат класифікації за алгоритмом 2
23	1	T_2	$T_2 1$
24	1	T_2	$T_2 1$
25	1	T_2	$T_2 0,96; T_3 0,04$
26	2	T_2	$T_2 0,88; T_3 0,12$
27	2	T_2	$T_2 0,79; T_3 0,21$
28	1	T_2	$T_2 0,71; T_3 0,29$
30	1	T_2	$T_2 0,52; T_3 0,48$
36	2	T_3	$T_2 0,04; T_3 0,96$
37	2	T_3	$T_3 1$
38	2	T_3	$T_3 1$
39	2	T_3	$T_3 0,93; T_4 0,07$
40	1	T_3	$T_3 0,87; T_4 0,13$
41	1	T_3	$T_3 0,8; T_4 0,2$
42	1	T_3	$T_3 0,73; T_4 0,27$
Загальний результат		$T_2 0,05; T_3 0,85; T_4 0,1$	$T_2 0,38; T_3 0,583; T_4 0,037$

вання, збіжності вибірових даних або їх узгодженості. Наявність такої перевірки сприяє впевненості у прийнятті подальших рішень. Тому в роботі запропоновано спосіб перевірки вибірки класифікованих даних з використанням міри узгодженості або збіжності класифікованих даних з використанням прийнятої шкали класифікації. Оскільки невизначеність вимірювання приводить до розсіювання даних відносно центру вибірки, то для визначення границі можливої неузгодженості використовується матриця відповідності шкали класифікації.

Література

1. ДСТУ 4362:2004. Якість ґрунту. Показники родючості ґрунтів. Чинний від 2006-01-01. Київ: Держспоживстандарт України, 2005, 20 с.

2. Гусар І., Любченко С., Купчик В. Визначення строкатості показників родючості та оптимізація розмірів елементарних ділянок для відбору зразків ґрунту в технологіях керованого землеробства: зб. наук. пр. / УкрНДІПВТ ім. Л.Погорілого. Випуск 18(32). – Дослідницьке, книга 2, 2014 с.102-104.

3. Мірошніченко М.М., Гладких Є.Ю., Ревтьє А.В. Особливості відбору проб ґрунту за локального внесення мінеральних добрив // Агроекологічний журнал, №4, 2016 с. 75-80.

4. JCGM 200: 2008. International vocabulary of metrology – Basic and general concepts and associated terms (VIM). JCGM 2008 – 88p.

5. ДСТУ 4114-2002. Ґрунти. Визначення рухомих сполук фосфору і калію за модифікованим методом Мачигіна. Київ. Держстандарт України, 2002, 10с.

6. Семенюк Р.С. Способы определения терм-множества лингвистической переменной с неопределенностью измерения // 15-й Міжнародний науково-технічний семінар «Неопределенность измерений: научные, нормативные и методические аспекты», УМ-2018, Созополь, 10 сентября 2018.

7. Пегат А. Нечеткое моделирование и управление / А. Пегат; с англ. – М.: Бином. Лаборатория знаний. 2009 – 798с.

8. Taste W.J., Wierman M.J. An information theoretic measure for the evaluation of ordinal scale data // Behavior Research Methods, 2006, 38(3), 487-494.

Literature

1. DSTU 4362: 2004. Soil quality. Indexes of fertility of soils. Operating from 2006-01-01. Kyiv: Derzhspozhyvstandart of Ukraine, 2005.20 p.

2. I. Gusar, S. Lyubchenko, V. Kupchik. Determination of variegated fertility rates and optimization of elementary plot sizes for soil sampling in managed farming technologies: Coll. Sciences. Ave. / UkrNIIPVT them. L.Pogoryloho. Issue 18 (32). - Doslidnytske, book 2, 2014 Pp.102-104.

3. Miroshnichenko MM, Gladkikh EY, Revtev AV Features of soil sampling at local yield of mineral fertilizers // the Agroecological magazine, №4, 2016 p.s 75-80.

4. JCGM 200: 2008. International Vocabulary of Metrology - Basic and General Concepts and Associated Terms (VIM). JCGM 2008 - 88p.

5. DSTU 41142002. Soils. Determination of movable phosphorus and potassium compounds is after the modified method of Machigin. Kyiv. National standard of Ukraine, 2002,10s.

6. Semeniyuk P.C. Ways of determining the term linguistic setvariable with measurement uncertainty // the 15th International scientific and technical seminar

7. Pegat A. Fuzzy modeling and control / of A. Pegat; with eng - M.: Binom. Laboratory known.2009 - 798p.

8. Taste W.J., Wierman M.J. An information theoretic measure for the evaluation of ordinal scale data // of Behavior Research Methods, 2006, 38 (3), 487-494.

Literatura

1. ISO (IV)CCCLXII: MMIV. Qualis est in solo. Indicators de solo ubertatem. 2006-01-01 ut efficaciora. Kyiv: State Committee de Ucraina, MMV, 20c.
2. Husar I. Lubchenco S. Kupchik B. variis Indicators optimize et ubertatis magnitudinem repulsi-technology culturae elementa exordii insidias solo sampling: Coll. Scientiae. pr. / UkrNDIPVT eos. L. Pohoriloho. Libri XVIII (XXXII). - Doslidnytske, libri 2: 2014. s.102-104.
3. M Miroshnichenko incompti tenuem lecture EY, A. Revtye Sampling features pro loci solo fecundatio // Agroecology magazine, №4, MMXVI p. 75-80.
4. CC JCGM: 2008. copiam Romani sermonis metrology International - Latin quod communia verbis, et consociata (LABOR). JCGM MMVIII - 88p.
5. ISO 4114-2002. Nata. Determinatio mobili de phosphoro et kalium mutata Machyhina suis Revolutionibus componit. Kyiv. Latin publica de Russia, MMII, 10c
6. Semeniuk Rs Ad determinandum modi multitudo lynhvystycheskoy variabilis-terminus et mensura neopredelennostyu // Acta Internationalis 15. Workshop «dubitationem mensurae nauchnye, normatyvnye et facies Methodological», MMXVIII-um, Sozopol, September X MMXVIII.
7. Pehat Nechetkoe A. sculpturae et Management / A. Pehat; cum Anglis. - M.: Binom. Laboratory scientia. MMIX - 798s.
8. Tastee W.J., Wierman M.J. An contemplativa est mensura ad iudicium notitia ex notitia ordinal scale // Moribus Research Methods, MMVI, XXXVIII (III), 487-494.

UDC 681.2

METHOD OF VERIFICATION OF CONFORMITY OF CLASSIFIED DATA IN THE EVALUATION OF SOIL QUALITY INDICATORS

Semeniuk R.,

e-mail: SemeniukRoman@i.ua,

<https://orcid.org/0000-0001-5139-6330>,

Yaremchuk N., Doctor of Technical Sciences, Professor,

<https://orcid.org/0000-0003-0661-1243>

NTUU «KPI them. I. Sikorsky»,

Gusar I.,

email: gusaririna19@gmail.com,

<https://orcid.org/0000-0002-58724672>

SSO L. Pogorely UkrNIPIPT

Summary

Goal. The purpose of the paper is to consider a method of constructing a scale of classification of soil quality indicators, taking into account the uncertainty of soil sample analysis. Scale matching matrix is selected as a characteristic of classification quality. To propose a method of checking the consistency of the classified sample data against the information measure of consistency using the consistency limit obtained on the basis of the classification scale of the classification scale.

Research methods: theoretical - analysis and synthesis of information resources and analytical - construction of a mathematical model of methods for checking the consistency of classified data when evaluating soil quality indicators.

Results. To assess the quality of soils by individual fertility indicators, an orderly classification scale is used, that is, an orderly series of quality levels that can be used for monitoring land resources, improving soil quality for fertilizer application, and determining the suitability of land for specific purposes.

The first stage of determining the quality of soils is to obtain soil samples according to the adopted program, the second is to determine the quality indicators, and the third is to make appropriate decisions. The second stage requires a classification scale that is based on a sample analysis procedure and therefore is individual to each method of analysis and, in addition, must take into account the uncertainty of measuring quality indicators, which is significant enough for all methods of analysis. Therefore, in the paper we used, as an important characteristic of the scale, a matrix of correspondence between the areas of determination of individual classes of equivalence of the scale and the obtained classification results, in which cells are the probability of correct and incorrect assignment to certain equivalence classes.

Conclusions. Obtained samples of classified data are subject to further processing. As revealed by the analysis of the literature, the samples are analyzed for the purpose of determining correlation or consistency with the reference data, scatter test, sampling of data or their consistency. The presence of such a check contributes to the confidence in future decisions. Therefore, the paper proposes a method of verifying a sample of classified data using a measure of consistency or convergence of classified data using an accepted classification scale. Since measurement uncertainty leads to scattering of data relative to the sample center, a classification scale matching matrix is used to determine the margin of possible inconsistency.

Key words: classification scale, measure of consistency, equivalence class.

УДК 681.2

СПОСОБ ПРОВЕРКИ СОГЛАСОВАННОСТИ КЛАСИФИЦИРОВАННЫХ ДАННЫХ ПРИ ОЦЕНКЕ ПОКАЗАТЕЛЕЙ КАЧЕСТВА ПОЧВЫ

Семенюк Р.,

e-mail: SemeniukRoman@i.ua,

<https://orcid.org/0000-0001-5139-6330>,

Яремчук Н., д-р.техн. наук, проф.,

<https://orcid.org/0000-0003-0661-1243>

НТУУ «КПИ им. И.Сикорского»

Гусар И.,

email: gusaririna19@gmail.com, <https://orcid.org/0000-0002-58724672>

ГНУ «УкрНИИПИТ им. Л. Погорелого»

Аннотация

Цель работы - рассмотреть способ построения шкалы классификации показателей качества почвы с учетом неопределенности анализа проб почвы. Как характеристику качества классификации избрана матрица соответствия шкалы. Предложить способ проверки согласованности классифицированных выборочных данных по информационной степени согласованности с применением предельной степени согласованности, которую получают на основе матрицы соответствия шкалы классификации.

Методы исследований: теоретические анализ и синтез –информационных ресурсов и аналитические – построение математической модели способов проверки согласованности классифицированных данных при оценке показателей качества почвы.

Результаты. Для оценки качества почв по показателям плодородия используется упорядоченная шкала классификации, то есть упорядоченный ряд уровней качества, которые могут быть использованы при мониторинге земельных ресурсов, при улучшении качества почв по внесенным удобрениям, при определении пригодности земель для определенных целей.

Первым этапом определения качества почв является получение проб почвы по принятой программе, вторым - определение показателей качества, третьим - принятие соответствующих решений. Для реализации второго этапа необходима шкала классификации, которая основана на процедуре анализа проб и поэтому является индивидуальной для каждого способа анализа и, кроме того, должна учитывать неопределенность измерения показателей качества, является весьма значительной для всех способов анализа. Поэтому в работе использованы, как важную характеристику шкалы, матрицу соответствия между областями определения отдельных классов эквивалентности шкалы и полученными результатами классификации, в ячейках которой находятся вероятности правильного и неправильного отнесения к определенным классам эквивалентности.

Выводы. Полученные выборки классифицированных данных подлежат дальнейшей обработке. Как показал анализ литературных источников, выборки анализируют с целью определения корреляции или согласованности с опорными данными, проверки рассеивания, сходимости выборочных данных или их согласованности. Наличие такой проверки способствует уверенности в принятии дальнейших решений. Поэтому в работе предложен способ проверки выборки классифицированных данных с использованием степени согласованности или сходимости классифицированных данных с использованием принятой шкалы классификации. Так как неопределенность измерения приводит к рассеиванию данных относительно центра выборки, то для определения границы возможной несогласованности используется матрица соответствия шкалы классификации.

Ключевые слова: шкала классификации, мера согласованности, класс эквивалентности.