

ФАКТОРНЫЙ АНАЛИЗ ДЛЯ ОПРЕДЕЛЕНИЯ БАЗОВЫХ ПАРАМЕТРОВ ПОЧВЕННЫХ ГОРИЗОНТОВ

Предлагается методика по определению факторным анализом в многопараметрическом исследовании базовых параметров, с помощью которых можно описать все параметры почвенного горизонта. Оценка правильности метода осуществляется по характеристикам регрессионных моделей параметров исследования.

Одной из важных проблем в многопараметрических исследованиях почв является проблема определения главных, базовых параметров, число которых может быть существенно меньше исходного числа показателей.

Уменьшение числа параметров при изучении почв при полной адекватности описания имеет важное значение для практики исследования, так как позволяет значительно снизить затраты на отбор и анализ проб почвенных горизонтов.

В связи с этим на базе факторного анализа нами был разработан алгоритм, позволяющий из всех почвенных характеристик выделить базовые, с помощью которых можно определять остальные по соответствующим регрессионным моделям.

Как известно, в факторном анализе основным предположением является равенство [1,2]:

$$X_i = \sum_{r=1}^k a_{ir} \times F_r + e_i, \quad (1)$$

(i = 1, 2, ..., p)

где X_i -ая переменная, F_r – r-ый фактор, a_{ir} – факторная нагрузка, k – количество факторов, e_i – остатки, которые представляют источники отклонений, действующие только на X_i .

Эти p случайных величин e_i предполагаются независимыми как между собой, так и с k величинами F_r .

Уравнение (1) нельзя проверить непосредственно, поскольку p переменных X_i выражены в них через $(p+k)$ ненаблюдаемых переменных.

Но эти уравнения заключают в себе гипотезу о ковариациях и дисперсиях X_i , которую можно проверить.

Когда число факторов $k > 1$, то ни факторы, ни нагрузки не определяются однозначно, поскольку в уравнении (1) факторы F_r могут быть заменены любым ортогональным преобразованием их с соответствующим преобразованием нагрузок. Это свойство использовано для преобразования или вращения факторов

[3], полученных в каком-либо практическом исследовании.

Вращение подбирается так, чтобы переменные, которые в большей или меньшей степени измеряют некоторые легко опознаваемые стороны, имели бы достаточно высокие нагрузки на один фактор и нулевые или почти нулевые на другие факторы.

Если отправной точкой является корреляционная матрица R с единицами на главной диагонали, то говорят о компонентном анализе, чья модель отлична от модели классического факторного анализа и приводит к дескриптивным факторам. Если в матрице R используют оценки общностей, то получают модель факторного анализа.

Алгоритм метода главных компонент:

1. Расчет корреляционной матрицы R :

$$R_{jk} = S_{jk} / (\sqrt{S_{jj} \times S_{kk}}), \quad (2)$$

где $S_{jk} = \sum (X_{ij} - T_j) \times (X_{ik} - T_k) -$

$$- \frac{1}{n} \sum_{i=1}^n (X_{ij} - T_j) \times \sum_{i=1}^n (X_{ik} - T_k),$$

$$T_j = \frac{\sum_{i=1}^n X_{ij}}{n}$$

$i = 1, 2, \dots, n$ – наблюдения; $j = 1, 2, \dots, m$ – переменные.

2. Вычисление собственных значений, собственных векторов корреляционной матрицы.

3. Вычисление накопленных отношений собственных значений корреляционной матрицы, больших или равных заданной пользователем константы.

4. Вычисление матрицы факторных нагрузок по собственным значениям и соответствующим собственным векторам корреляционной матрицы.

5. Ортогональное вращение матрицы факторов.

На базе метода главных компонент предлагается следующий алгоритм минимизации

числа параметров исследования для многопараметрических объектов:

1. Строим матрицу исследования (строчки – наблюдения, столбцы – параметры исследования).

2. Методом главных компонент находим матрицу факторных нагрузок. Осуществляем варимаксное вращение в пространстве факторов (строчки в матрице факторных нагрузок – параметры исследования, столбцы – гипотетические переменные, факторы).

3. В каждой строчке матрицы факторных нагрузок, то есть для каждого параметра исследования, находим максимальную по модулю факторную нагрузку.

4. Определяем по каждому фактору попадание в этот фактор параметров с максимальной по модулю факторной нагрузкой (пункт 3). То есть тем самым определяем объединение параметров по факторам.

5. В объединившихся в каждом факторе параметрах выбираем один параметр с максимальной по модулю факторной нагрузкой. Число таких выбранных параметров будет равно, очевидно, числу факторов.

6. Строим для всех параметров исследования полиномиальные модели, аргументами в которых будут выбранные в пункте 5 параметры.

7. По построенным моделям для каждого параметра осуществляем определение вкладов параметров-аргументов (оценку количественной обусловленности параметров выбранными параметрами).

8. Сравниваем качественные групповые обусловленности, объединения параметров по факторам, с количественными обусловленностями параметров, полученными в пункте 7.

Если групповые и количественные обусловленности для всех параметров исследования не будут сильно отличаться по числу совпадений, то выбранные в пункте 5 параметры могут быть приняты за базисные при описании данного многопараметрического объекта, матрица исследования которого была взята за основу в данном алгоритме. То есть тем самым осуществляем минимизацию количества параметров исследования, потому что число факторов меньше числа параметров.

Рассмотрим использование этого алгоритма для горизонта A_n чернозема обыкновенного среднегумусного среднемошного глинистого и тяжелосуглинистого.

Результаты факторного анализа:

Таблица 1. Объединение по фактору 6

| Номер | Название параметра | Нагрузка |
|-------|----------------------------|----------|
| 1 | Гигроскопическая влажность | -0,9730 |

в факторе 6 базовый параметр – 1 (гигроскопическая влажность)

Таблица 2. Объединение по фактору 5

| Номер | Название параметра | Нагрузка |
|-------|--|----------|
| 2 | Мощность гумусового горизонта, см | -0,7078 |
| 3 | Сумма поглощенных оснований, мг-экв на 100 г почвы | 0,8896 |

в факторе 5 базовый параметр – 3 (сумма поглощенных оснований)

Таблица 3. Объединение по фактору 3

| Номер | Название параметра | Нагрузка |
|-------|--|----------|
| 4 | Общий запас гумуса, т/га | 0,5707 |
| 5 | Сумма фракций менее 0,01 мм | -0,5847 |
| 8 | Подвижный P_2O_5 по Чирикову (в мг на 100 г почвы) | 0,9092 |

в факторе 3 базовый параметр – 8 (подвижный P_2O_5 по Чирикову (в мг на 100 г почвы))

Таблица 4. Объединение по фактору 2

| Номер | Название параметра | Нагрузка |
|-------|------------------------------------|----------|
| 6 | pH водной суспензии по Алямовскому | 0,9314 |

в факторе 2 базовый параметр – 6 (pH водной суспензии по Алямовскому)

Таблица 5. Объединение по фактору 1

| Номер | Название параметра | Нагрузка |
|-------|---------------------|----------|
| 7 | Гумус по Тюрину в % | -0,9433 |

в факторе 1 базовый параметр – 7 (гумус по Тюрину, %)

Таблица 6. Объединение по фактору 4

| Номер | Название параметра | Нагрузка |
|-------|---|----------|
| 9 | Обменный K_2O по Чирикову (в мг на 100 г почвы) | 0,9464 |

в факторе 4 базовый параметр – 9 (обменный K_2O по Чирикову (в мг на 100г почвы))

По результатам факторного анализа, табл. 1–6, определены базовые параметры для построения регрессионных моделей:

1 – гигроскопическая влажность;
 3 – сумма поглощенных оснований, мг-экв на 100г почвы;
 6 – pH водной суспензии по Алямовскому;
 7 – гумус по Тюрину, %;
 8 – подвижный P_2O_5 по Чирикову, мг на 100 г почвы;
 9 – обменный K_2O по Чирикову, мг на 100 г почвы.

Регрессионные модели на базовых параметрах:

**Зависимый параметр – 2
 (мощность гумусового горизонта, см)**

Таблица 7. Вклады параметров-аргументов в модели

| Номер | Название параметра | Вклад в модель |
|-------|--|----------------|
| 1 | Гигроскопическая влажность | 0,0344 |
| 3 | Сумма поглощенных оснований, мг-экв на 100 г почвы | 0,0508 |
| 8 | Подвижный P_2O_5 по Чирикову (в мг на 100 г почвы) | 0,0268 |
| 6 | pH водной суспензии по Алямовскому | 0,0699 |
| 7 | Гумус по Тюрину в % | 0,0844 |
| 9 | Обменный K_2O по Чирикову (в мг на 100 г почвы) | 0,7336 |

Модель – 2 Мощность гумусового горизонта, см

значимый по вкладу параметр – 7 (гумус по Тюрину, %) – вклад = 0,0844360;

значимый по вкладу параметр – 9 (обменный K_2O по Чирикову (в мг на 100г почвы) – вклад = 0,7336345.

$$y = -244,6701 + 38,6856x_7 + 4,6238x_9^2 - 0,6675x_7^3 + 6,7184x_9 - 0,2614x_9^2 + 0,0029x_9^3 \quad (3)$$

графики исходного и модельного изменений

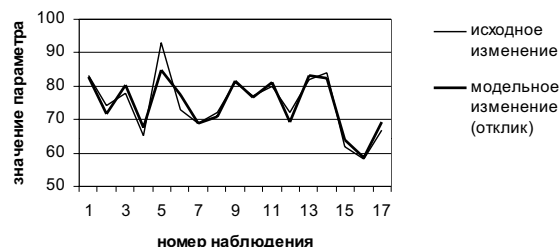


Рисунок 1. Графики исходного и модельного изменений мощности гумусового горизонта

Таблица 8. Характеристики модели

| Характеристики модели | Значения |
|----------------------------|----------|
| Коэффициент детерминации | 0,95 |
| Средняя абсолютная ошибка | 1,95 |
| Средняя ошибка в процентах | 2,61 |

**Зависимый параметр – 4
 (общий запас гумуса, т/га)**

Таблица 9. Вклады параметров-аргументов в модели

| Номер | Название параметра | Вклад в модель |
|-------|--|----------------|
| 1 | Гигроскопическая влажность | 0,0443 |
| 3 | Сумма поглощенных оснований, мг-экв на 100 г почвы | 0,0278 |
| 8 | Подвижный P_2O_5 по Чирикову (в мг на 100 г почвы) | 0,1000 |
| 6 | pH водной суспензии по Алямовскому | 0,0934 |
| 7 | Гумус по Тюрину в % | 0,6886 |
| 9 | Обменный K_2O по Чирикову (в мг на 100 г почвы) | 0,0459 |

Модель – 4 Общий запас гумуса, т/га значимый по вкладу параметр – 8 (подвижный P_2O_5 , мг на 100г почвы)

вклад = 0,0999775;

значимый по вкладу параметр – 6 (pH водной суспензии по Алямовскому)

вклад = 0,0934471;

значимый по вкладу параметр – 7 (гумус по Тюрину, %) вклад = 0,6885791.

$$y = 15096,743 + 5,0305x_8 + 0,9348x_8^2 + 0,0130x_8^3 - 4710,143x_6 + 240,1996x_6^2 + 9,3165x_6^3 + 974,9214x_7 - 90,4667x_7^2 + 2,3656x_7^3 \quad (4)$$

графики исходного и модельного изменений

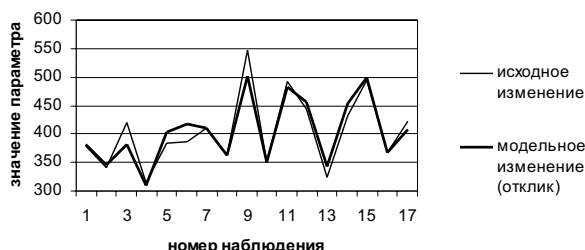


Рисунок 2. Графики исходного и модельного изменений общего запаса гумуса

Таблица 10. Характеристики модели

| Характеристики модели | Значения |
|----------------------------|----------|
| Коэффициент детерминации | 0,95 |
| Средняя абсолютная ошибка | 14,05 |
| Средняя ошибка в процентах | 3,47 |

**Зависимый параметр – 5
(сумма фракций менее 0,01 мм)**

Таблица 11. Вклады параметров-аргументов в модели

| Номер | Название параметра | Вклад в модель |
|-------|---|----------------|
| 1 | Гигроскопическая влажность | 0,0169 |
| 3 | Сумма поглощенных оснований, мг-экв на 100 г почвы | 0,0373 |
| 8 | Подвижный P ₂ O ₅ по Чирикову (в мг на 100 г почвы) | 0,8869 |
| 6 | pH водной суспензии по Алямовскому | 0,0130 |
| 7 | Гумус по Тюрину в % | 0,0231 |
| 9 | Обменный K ₂ O по Чирикову (в мг на 100 г почвы) | 0,0227 |

Модель – 5 Сумма фракций менее 0,01 мм значимый по вкладу параметр – 8 (подвижный P₂O₅, мг на 100 г почвы) вклад = 0,8869079.

$$y = 54,6267 + 0,6030x_8 - 0,0805x_8^2 - 0,0007x_8^3 \quad (5)$$

Список использованной литературы:

1. Харман Г. Современный факторный анализ. - М.: Статистика, 1972.
2. Иберла К. Факторный анализ. - М.: Статистика, 1980.
3. Kaiser. H.F. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 23, 187 - 200 (1958).

графики исходного и модельного изменений



Рисунок 3. Графики исходного и модельного изменений суммы фракций <0,01

Таблица 12. Характеристики модели

| Характеристики модели | Значения |
|----------------------------|----------|
| Коэффициент детерминации | 0,95 |
| Средняя абсолютная ошибка | 1,37 |
| Средняя ошибка в процентах | 2,49 |

Все модели имеют высокие коэффициенты детерминации и небольшие средние абсолютные и относительные ошибки.

То есть для определения параметров горизонта A_n надо определить только базовые параметры, а остальные будут определены по построенным регрессионным моделям.