

DOI:

ОПЫТ ИСПОЛЬЗОВАНИЯ ЯЗЫКА ПРОГРАММИРОВАНИЯ R ДЛЯ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ

В. М. Буре^{1,2}, О. А. Митрофанова¹

¹ *ФГБНУ «Агрофизический научно-исследовательский институт» (АФИ);*

² *Санкт-Петербургский государственный университет (СПбГУ)*

Рассмотрена задача статистического анализа данных, полученных при помощи беспилотного летательного аппарата, с применением языка программирования R в системе точного земледелия. Задача заключается в анализе цветовых характеристик посевов пшеницы с различными качественными показателями. Представлен опыт анализа наличия линейной зависимости между цветом растений и дозой азота в зависимости от различных наборов качественных факторов.

EXPERIENCE OF USING R PROGRAMMING LANGUAGE FOR STATISTICAL PROCESSING OF REMOTE SENSING DATA

V. M. Bure^{1,2}, O. A. Mitrofanova¹

¹ *Agrophysical Research Institute;*

² *St. Petersburg State University*

The article deals with the problem of statistical analysis of data obtained with an unmanned aerial vehicle, using R programming language in the system of precision farming. The task is to analyze the colour characteristics of wheat crops with different quality indicators. The experience of the analysis of the existence of a linear relationship between the colour of plants and the dose of nitrogen depending on different sets of qualitative factors is presented.

Введение

В ходе решения многочисленных задач точного земледелия с использованием данных дистанционного зондирования (ДДЗ) возникает необходимость применения статистического анализа. В настоящее время исследователям и специалистам агропромышленного комплекса доступен широкий выбор программ и пакетов, позволяющих осуществлять обработку ДДЗ.

Для решения рассмотренной в работе задачи использовался язык программирования R (среда разработки RStudio). Выбор средства обработки данных основан на следующих главных преимуществах языка R:

– открытый исходный код, лицензия GNUGPL (GeneralPublicLicense);

- удобная программная среда разработки Rstudio;
- активное развитие проекта (программное обеспечение регулярно обновляется, добавляются новые пакеты);
- широкий выбор статистических и численных методов.

В последние годы применение языка R в задачах точного земледелия становится все более актуальным и перспективным (Panayi, Peters, Kyriakides, 2017; Plant, 2012). Далее рассмотрен пример решения подобной задачи.

Объекты и методы.

Для статистической обработки использовались аэрофотоснимки тестовых площадок пшеницы, полученные с помощью автоматизированного беспилотного летательного аппарата Геоскан-401. Всего было заложено 168 площадок с известными дозами внесения азота (0, 60, 90 кг действующего вещества на 1 га) и качественными показателями посевов (размер зерна – крупные и мелкие; наличие сорняков – есть или нет; нормы высева – 6 или 5 млн на 1 га). В рамках задачи необходимо было проанализировать наличие линейной зависимости между цветом растений и дозой азота в зависимости от различных качественных показателей.

Таким образом, исходными данными задачи являлись параметры цвета растений L , a , b для восьми вариантов набора качественных показателей, которые представляются в виде обобщенной характеристики цвета (Буре, Канаш, Митрофанова, 2017):

$$C_{Lab} = \alpha L + \beta_1 a + \beta_2 b, \quad (1)$$

где коэффициенты $\alpha \geq 0$, $\beta_1 \geq 0$ и $\beta_2 \geq 0$ подбираются эмпирически, причем

$$\alpha + \beta_1 + \beta_2 = 1.$$

Основным критерием подбора коэффициентов является обеспечение максимально выраженной линейной связи между обобщенной характеристикой цвета (1) и дозой азота на тестовых площадках. Исходя из того, что необходимо построить 8 уравнений линейных регрессий (для каждого набора качественных показателей), данное условие можно записать в следующем виде (Bure, Mitrofanova, 2017):

$$\max_{(\alpha, \beta_1)} \min_{i=1, \dots, 8} R_i^2(\alpha, \beta_1), \quad (2)$$

где R_i^2 – коэффициенты построенных линейных регрессий.

Результаты и обсуждение.

Решение задачи можно представить поэтапно следующим образом:

1. Анализ распределений выборок параметров цвета по каждому показателю. Входными данными являются две выборки соответствующей пары качественных показателей. Сначала проводится анализ соответствия каждого набора данных нормальному распределению с помощью R-функции *lillie.test* (тест Колмогорова-Смирнова в модификации Лиллиефорса). Если выборки одновременно будут иметь нормальное распределение, то осуществляется переход к этапу 2, если они одновременно будут иметь ненормальное распределение, то осуществляется переход к этапу 3.

2. Проводится сравнение распределений выборок параметрическим методом. Внутри алгоритма осуществляется перебор коэффициентов α , β_1 , β_2 ; используются критерий Фишера с помощью R-функции *var.test* и критерий Стьюдента с помощью R-функции *t.test*.

3. Проводится сравнение распределений выборок непараметрическим методом. Внутри алгоритма также осуществляется перебор коэффициентов α , β_1 , β_2 ; используется критерий Вилкоксона с помощью R-функции *wilcox.test*.

4. Решается основная задача — осуществляется поиск максимина (2), а также построение линейных регрессионных моделей с помощью R-функции *lm*.

Для выполнения предварительной обработки (этапы 1–3) было разработано отдельное веб-приложение (рис. 1) с помощью пакета Shiny. Входными данными является текстовый файл, содержащий выборки коэффициентов α , β_1 , β_2 для одной пары качественных показателей. В качестве выходных данных приложение выводит результат сравнения распределений, а также содержимое загруженного файла.

Для автоматизации решения основной задачи и проведения последующих аналогичных опытов также было разработано веб-приложение (рис. 2) с помощью пакета Shiny. В приложении пользователь загружает текстовый файл с исходными данными, а в итоге выводятся содержимое файла, величины

найденного максимина и подобранных коэффициентов, а также результаты построения линейных регрессий.

Check of distributions

Choose CSV File

Browse...

data_MK.txt

Upload complete

☒ Header

Separator

☐ Comma

☐ Semicolon

☐ Tab

Quote

☐ None

☐ Double Quote

☐ Single Quote

Result

Distributions are different

File content

i	xL	xa	xb	yL	ya	yb
1	49.25	-17.77	17.49	53.27	-18.95	21.14
2	51.45	-19.83	16.99	53.34	-18.09	20.18
3	54.19	-19.57	17.07	54.02	-18.98	20.92
4	54.01	-22.46	21.67	54.42	-21.47	24.56
5	48.35	-17.23	25.76	55.35	-23.40	25.66
6	53.58	-19.12	26.03	54.10	-18.73	20.14
7	51.45	-12.04	13.34	51.69	-11.89	15.62
8	49.59	-13.57	15.14	47.91	-14.52	18.10
9	48.92	-13.36	14.27	55.35	-14.09	12.44
10	52.87	-13.33	12.80	56.71	-11.81	10.86
11	49.16	-9.52	7.06	53.47	-13.66	13.90
12	46.60	-12.95	9.36	52.60	-16.51	16.33
13	52.74	-8.89	16.23	51.23	-15.12	20.64
14	50.40	-13.61	20.60	53.60	-16.67	23.47
15	47.93	-9.16	16.72	53.48	-15.10	26.14
16	50.24	-10.26	18.16	49.93	-15.07	23.34
17	55.09	-12.24	23.46	51.92	-15.60	27.06
18	51.28	-10.83	19.03	52.81	-13.53	22.65

Рис. 1. Веб-приложение Check of distributions.

MaxMin Searcher

Analysis of plants color characteristics using aerophots with different factors of qualitative indicators

Choose CSV File

Browse...

regres.txt

Upload complete

☒ Header

Separator

☐ Comma

☐ Semicolon

☐ Tab

Quote

☐ None

☐ Double Quote

☐ Single Quote

i	L1	a1	b1	L2	a2	b2	L3	a3	b3	L4	a4	b4	L5	a5	b5	L6	a6	b6	L7
1	53.27	-18.95	21.14	51.69	-11.89	15.62	49.25	-17.77	17.49	51.45	-12.04	13.34	47.00	-18.31	24.72	48.95	-15.51	20.67	57.98
2	53.34	-18.09	20.18	47.91	-14.52	18.10	51.45	-19.83	16.99	49.59	-13.57	15.14	48.06	-18.10	24.17	45.47	-14.80	20.41	59.46
3	54.02	-18.98	20.92	55.35	-14.09	12.44	54.19	-19.57	17.07	48.92	-13.36	14.27	47.73	-19.12	26.14	46.11	-16.46	22.07	52.31
4	54.42	-21.47	24.56	56.71	-11.81	10.86	54.01	-22.46	21.67	52.87	-13.33	12.80	46.72	-16.49	23.48	46.96	-17.90	23.47	49.09
5	55.35	-23.40	25.66	53.58	-19.12	16.33	48.35	-17.23	26.76	49.16	-9.52	7.06	45.19	-18.94	24.48	47.34	-17.63	21.79	49.96
6	54.10	-18.73	20.14	52.60	-16.51	16.33	53.58	-19.12	26.03	46.60	-12.95	9.36	44.64	-15.33	21.62	45.99	-17.69	22.74	55.20
7	51.23	-16.12	20.64	52.60	-7.68	11.07	52.74	-8.89	16.23	50.32	-10.83	12.32	49.63	-13.99	26.76	52.31	-8.30	20.45	51.67
8	53.60	-16.67	23.47	52.20	-11.82	14.05	50.40	-13.61	20.60	53.82	-9.57	11.13	54.83	-10.96	23.94	47.32	-12.64	24.34	57.15
9	53.48	-15.10	26.14	52.88	-13.09	16.40	47.93	-9.16	16.72	50.28	-8.50	7.46	50.51	-12.65	24.60	46.44	-11.07	22.81	51.80
10	49.93	-15.07	23.34	53.00	-9.86	15.02	50.24	-10.26	18.16	45.65	-9.80	8.66	48.70	-12.29	26.69	51.76	-7.75	20.49	56.20
11	51.92	-15.60	27.06	53.94	-12.22	15.97	55.09	-12.24	23.46	50.64	-11.88	13.73	47.36	-12.25	26.09	52.69	-10.10	22.69	53.25
12	52.81	-13.53	22.65	54.36	-12.66	17.73	51.28	-10.83	19.03	56.83	-9.78	10.96	52.76	-10.88	25.01	49.48	-11.31	24.92	49.16

Results

MaxMin	Alpha	Beta	Gamma
0.43	0.00	0.61	0.59

First regression

Call:

lm(formula = y1_res ~ p\$N)

Residuals:

Mn 10 Median 30 Max

-1.22494 -0.28816 0.07258 0.41434 0.97986

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.454212 0.380368 0.1179 0.861

p\$N -0.059941 0.007134 -8.402 7.65e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7414 on 10 degrees of freedom

Multiple R-squared: 0.8799, Adjusted R-squared: 0.8635

F-statistic: 70.59 on 1 and 10 DF, p-value: 7.647e-06

Рис. 2. Веб-приложение MaxMinSearcher

Заключение

В работе представлен опыт анализа цветовых параметров растений по данным дистанционного зондирования с помощью языка программирования R. Разработаны и протестированы алгоритмы проведения статистической обработки, а также осуществлена автоматизация решения задачи с помощью пакета Shiny в виде веб-приложений.

Выбранный инструментарий решения задачи является достаточно эффективным средством статистической обработки данных, полученных с помощью аэрофотосъемки, в системе точного земледелия. Автоматизация в виде веб-приложений представляется актуальным и перспективным подходом, который позволит в будущем осуществлять автоматизированное решение широкого круга задач точного земледелия в одном веб-сервисе (Janahshiri, MohdShariff, 2014).

Список литературы

1. Буре В. М., Канааш Е. В., Митрофанова О. А. Анализ характеристик цвета растений по аэрофотоснимкам с различными факторами качественных показателей // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2017. Т. 13. № 3. С. 278–285.
2. Bure V. M., Mitrofanova O. A. Analysis of color characteristics of plants using aerial photography // Constructive nonsmooth analysis and related topics (dedicated to the memory of V. F. Demyanov) (CNSA). 2017. 7973945.
3. Jahanshiri E., Mohd Shariff A. R. Developing web-based data analysis tools for precision farming using R and Shiny // IOP Conf. Series: Earth and Environmental Science. 2014. № 20. 012014.
4. Panayi E., Peters G. W., Kyriakides G. Statistical modelling for precision agriculture: A case study in optimal environmental schedules for *Agaricus Bisporus* production via variable domain functional regression // PloS ONE. 2017. 12(9): e0181921.
5. Plant R. E. Spatial data analysis in ecology and agriculture using R. Boca Raton: CRC Press. 2012. 648 p.