
METHODOLOGY FOR MODELING THE RELATIONSHIP OF COLOR CHARACTERISTICS OF DIGITAL IMAGES AND AGROCHEMICAL SOIL INDICATORS⁴

Eng. Antonina Mihaylova – PhD Student

Department of Automatics and Mechatronics,

“Angel Kanchev” University of Ruse

Tel.: +359 82 888 676

E-mail: amihaylova@uni-ruse.bg

Abstract: In this report, a methodology is proposed for researching the relationship and deriving a regression mathematical model for determining the agrochemical parameters of the soil - content of organic matter (humus) and phosphorus through colour characteristics obtained from digital images of soil from optical devices. The methodology describes the modelling of a regression relationship between the selected colour characteristics from a previous study and the indicators of soil humus and phosphorus, measured in laboratory conditions using statistical methods. The stages of research of the studied object are indicated in accordance with the "black box" principle and the need to carry out mandatory checks of statistical hypotheses for the adequacy of the model and the significance of its coefficients. The methodology illustrates the application of univariate linear regression analysis LR and multivariate linear regression analysis MLR Stepwise. The models were compared and evaluated using evaluation criteria – coefficient of determination (R^2), root mean square error (RMSE) and residual prediction deviation (RPD). It was found that the regression models for obtaining humus and phosphorus obtained by the MLR Stepwise method improved the results obtained for the camera device by increasing the coefficient of determination (R^2) by an average of 13% and reducing the root mean square error (RMSE) by an average of 5%.

Keywords: Regression models, Regression analysis, Prediction of agrochemical soil indicators, express monitoring, colour characteristics of digital soil images, remote sensing of soil indicators, digital images.

ВЪВЕДЕНИЕ

Бързите и лесни техники за определяне на агрохимични показатели на почвата, основани на компютърно зрение са предпочитани алтернативни решения за устойчиво управление в съвременното прецизно земеделие (Georgieva, 2020). Предвид характерната променливост на свойствата на почвата в рамките на дадено поле за различни райони на земеползване, прилагането на алтернативни инструменти за своевременно оценяване качеството на почвата е необходимост за съвременния фермер. Известни подобни решения за допълване и подобряване на конвенционалните методи са оптичните техники с висока разделителна способност като спектрометрия и мобилни сензори (Bünemann, 2018), които може да имат по-ниска точност от методите на лабораторните измервания, но предлагат бързи решения, недеструктивни са и са ориентир за определяне плодородието на обработваемото поле (Heil, 2022; Lavanya, 2020).

За изследването на сложни многофакторни обекти, каквато е почвата, връзката между оптичните данни и качествените параметри се търси основно чрез статистически методи (Mitkov, 2023). Те са свързани с извършване на анализ на големи масиви от данни чрез методите на математическата и приложна статистика и позволяват лесно и ефективно да бъдат вградени в специализирани софтуерни системи и да бъдат реализирани чрез програмни инструменти в различни измервателни устройства и приложения. При изследване на връзки между количествени (измерими) данни в областта на машинното обучение един от препоръчителните методи е регресията. Прилагането на регресионния анализ се оказва сложна задача, предимно поради изискването за спазване на точно определени етапи и

⁴ Докладът е представен в секция Електротехника, електроника и автоматика на 27 октомври 2023 с оригинално заглавие на български език: МЕТОДИКА ЗА МОДЕЛИРАНЕ НА ВРЪЗКАТА НА ЦВЕТОВИ ХАРАКТЕРИСТИКИ НА ЦИФРОВИ ИЗОБРАЖЕНИЯ И АГРОХИМИЧНИ ПОКАЗАТЕЛИ НА ПОЧВАТА

извършване на проверки за изпълнението на статистически критерии. За изследване на агрохимичните показатели на почвата регресионен анализ може да се приложи чрез намиране на зависимостта между измерени цветови характеристики на изображения за представителна група почвени проби с известен състав и свойства. Предложеният алгоритъм за провеждане на регресионен анализ за определяне на почвен органичен въглерод-хумус и усвоими форми на фосфор ще допринесе за по-лесното решаване на задачи за намиране на математически модели за определяне на показателите на качеството на почвата и изследване на наличието или отсъствие на връзка между тях. Методите за моделиране, които изучават подобни процеси са част от химичната дисциплина хеометрия "chemometrics", която използва различни математически и статистически методи за максимално извличане на пълноценна информация при анализа на химични данни (Mevic and Wehrens, 2007).

Целта на тази статия е да се предложи методика за моделиране на регресионна връзка на цветови характеристики на цифрови изображения на почвата, получени от избрани оптични устройства документ-камера-Dd, фотоапарат- Df, камера на мобилен телефон- Dt, колориметър Dk и показателите на почвата - съдържание на органично вещество и усвоими форми на фосфор.

ИЗЛОЖЕНИЕ

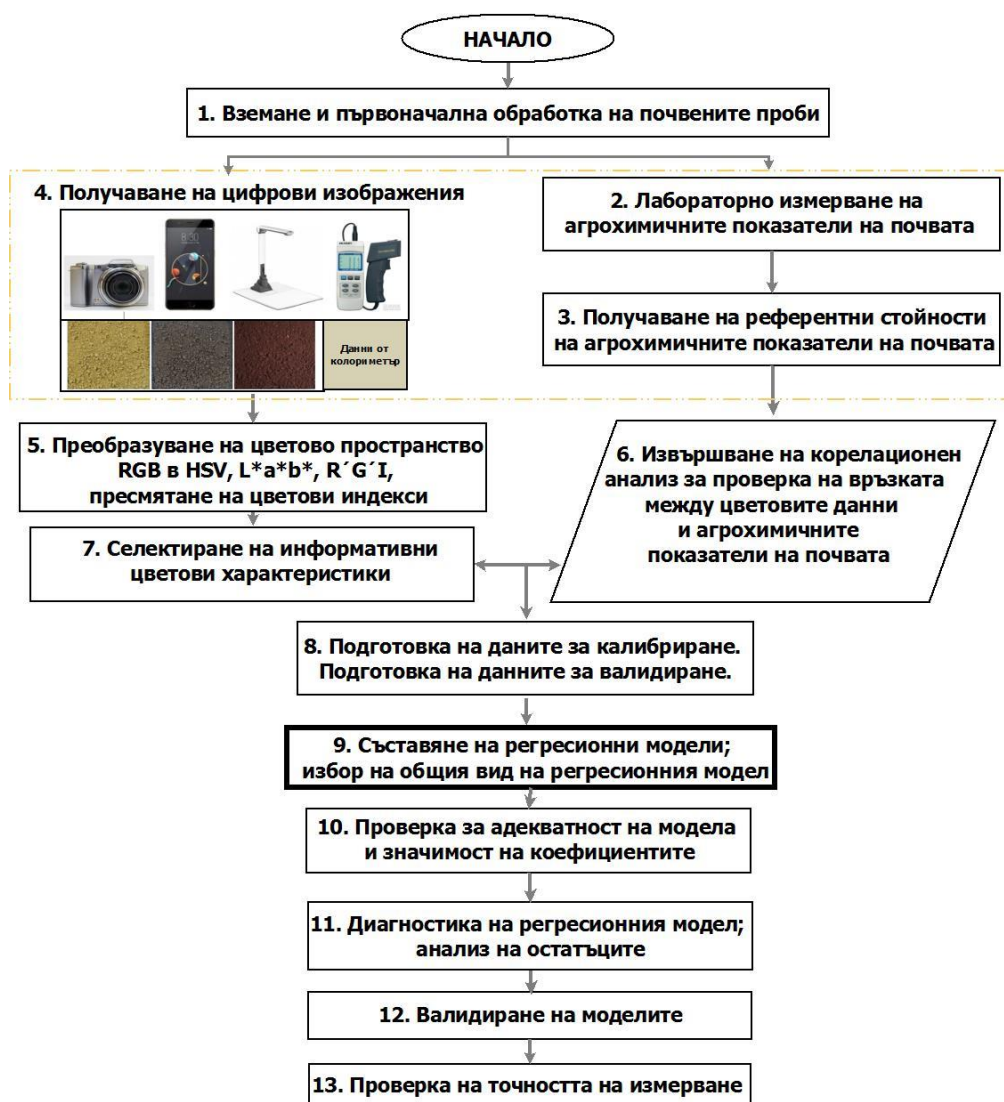
С практически пример са разгледани етапите на решение на регресионна задача и извеждане на регресионни математически модели за определяне на съдържание на органичен въглерод- хумус Н и фосфор Р в почвата. Схемата на етапите на моделиране е показана на фиг. 1.

Първият етап е вземане, подготовка и първоначална обработка на почвените проби. Необходимо е работните проби да са с широк обхват на съдържание от изследваните агрохимични показатели и различна запасеност с хранителни вещества. В този пример почвените проби са взети от три обработваеми земеделски полета в района на град Русе и Добрич. Препоръчително е пробите да се изсушат на въздух 2 седмици, да се смелят и пресеят през сито 2 mm. Това е стандартна процедура при подобни изследвания и се прави с цел поставяне на почвените проби в еднакви експериментални условия.

Етапи две, три и четири от методиката се извършват паралелно. Данните от лабораторното измерване на агрохимичните показатели служат за референтни стойности. В примера са измерени съдържание на органичен въглерод-хумус Н по метод на Тюрин и усвоими форми на фосфор Р, измерен със спектрофотометър по ацетатно- лактатен метод. Получаването на цифровите изображения на пробите се прави последователно с избраните оптични устройства. В посочения пример почвените проби са заснети с документ-камера-Dd, фотоапарат- Df, камера на мобилен телефон- Dt, и са измерени данни за цвета с цифров колориметър Dk.

Целесъобразно е цифровите изображения от камерите да се получават в JPEG формат. Данните от колориметъра се получават след предварително калибриране на уреда в цветови пространства RGB и HSL. Тези цветови пространства са най-използвани при този вид изследвания и поради нелинейната зависимост на показателите с цвета на почвата се препоръчва преобразуване на данните в различни цветови пространства (Rossel, 2007). В посочения пример са избрани RGB, HSV, L*a*b* 1976 и RGI (Rasras, 2006), (Wang, 2015). Данните от цифровите изображения от всички оптични устройства са преобразувани в избраните цветови пространства в програмна среда Matlab. Обработката на изображенията, сегментиране, изрязване и получаване на цветови данни в избрани цветови пространства е следващ етап 5 от разглежданата методика. Целесъобразно е за допълнително отстраняване на смущения от засенчване, отблясъци и несъвършенства на осветлението при заснемане, преобразуването на данни да включва въвеждане на индекси, които са препоръчани за използване при изображения (Танеја, 2022). В посочения пример са използвани индекс на оцветяване C index, индекс на зачервяване B index, индекс на нюанса S index, получени чрез математически пресмятания (Rossel, 2022). Съвкупността от определените цветови

характеристики и данните от пресметнатите индекси представляват входните променливи данни за моделиране на регресионната връзка.



Фиг. 1 Схема на етапите на моделиране

В етап 6 следва извършване на корелационна проверка за връзката на цветовите данни с показателите на почвата, в който освен цветовите данни се включват и данните за агрохимичните показатели на почвата. Зависимостите, при които се получава коефициент на корелация над 0.5 са значими и определящи за получаване на регресионните модели.

Следващият етап 7 е селектиране на информативни цветови характеристики. Необходимо е чрез статистически методи за подбор на информативни цветови признаци да се подберат тези входни променливи данни, които имат значителна корелация със изследваните параметри на почвата и същевременно са слабо корелирани помежду си. Това в примера е постигнато с прилагане на метод на сумите и корелационен анализ (Mitkov, 2011) с програма Statistica StatSoft. Извършва се проверка за скрита вътрешна корелация /мултиколинеарност/ между входните регресионни данни, което е основно изискване при прилагане на регресионни методи. Още етапът включва дескриптивното изясняване естеството на данните, тяхното статистическо разпределение по средна стойност и медиана, при които може да се отстранят данни с груби грешки, проби, отличаващи се от останалите над 75%. По този начин информацията от цветовите данни се концентрира в малък брой входни променливи, наричани регресионни фактори.

На фиг. 2 е показана схемата на примерния регресионен експеримент със селектираните регресионни фактори в предходно изследване, (Mihaylova, 2023), представен като обект на изследване на принципа на „черната кутия“.



Фиг. 2 Схемата на примерен регресионен експеримент

В етап 8 е необходимо да се подготвят /изберат/ данните, които ще бъдат включени в моделирането на регресионния модел. Обикновено се спазва правилото, че за съставяне на модела се използват 2/3 от работните данни, а останалата 1/3 част служи за проверка на модела и се явява предварителна подготовка за етап 12-валидиране на получените модели. Друг вариант е като набор за калибриране да се използват всички проби, а за валидиране и проверка на точността да се получи нова извадка с големина 30% от тях. В посочения пример е избран втори вариант. Извадката за валидиране е определена по метод Монте Карло с пакет Statistica.

При първоначалното съставяне на регресионните модели е целесъобразно да се използва линейна регресия (Linear Regression) като се приложи еднофакторен и многофакторен регресионен анализ MLR за по-точното определяне на влиянието на цветовете фактори върху показателите на почвата (Bratоеv, 2020), след което моделът се усложнява. За определяне на връзката между цветовете данни и химичната информация за почвата обикновено първоначално се търси някакъв вид линеен модел, тъй като тези модели са по-бързи за пресмятане, по-лесни за интерпретация и теоретично по-добре изяснени (Todorova, 2009). В посочения пример за регресионен фактор при еднофакторен регресионен анализ се използва цвят компонента „нормализирана стойност на червения цвят по данни от всяко оптично устройство R'n. Тя е избрана по „метод на сумите“ въз основа на най-високата корелационна зависимост между цветовете данни от използваните оптични устройства и търсените почвени показатели в етап 6. При многофакторен регресионен анализ MLR е целесъобразно да се приложат няколко независими информативни цвят фактори. В примера са използвани селектирани цвят характеристики, предмет на предходно изследване, (Mihaylova, 2023), посочени на схемата на регресионния експеримент фиг. 2 като допълнително е приложена техника Stepwise за отстраняване на съществуваща вътрешна корелация.

Следващ етап 9 е получаване на регресионен модел, който показва връзката между селектираните цвят характеристики и търсените показатели на почвата хумус и фосфор.

Оценката на регресионния модел, етап 10 от методиката включва оценка на параметрите на модела и оценка на неговата адекватност. Най-широко прилаганият метод за оценка на параметрите на регресионния модел е Методът на най-малките квадрати. Този метод се прилага при линейните регресионни модели и неговата цел е да се намерят(оценят) стойностите на параметрите на модела, за които функцията на остатъците приема минимална стойност. Проверява се с нулева хипотеза и критерий на Фишер с ниво на значимост p-value

$< 0,05$ и отхвърляне на нулевата хипотеза. Незначимите коефициенти, тези за които стойностите на $p > 0,05$, нулевата хипотеза не се отхвърля и те не се включват в моделите. Регресионният анализ от посочения пример за определяне на показателите хумус и фосфор е извършен с пакет Statistica. Получени са уравнения на регресия по два метода- LR и MLR Stepwise, обобщени в таблица 1.

Таблица 1. Регресионни уравнения и характеристики за показатели на почвата хумус и фосфор, получени по два метода LR и MLR Stepwise.

Метод линейна регресия LR	Критерии за оценка на модел			Регрес. модел, получен с метод MLR Stepwise	Критерии за оценка на модел		
Хумус	R	R ²	RMS E	Хумус	R	R ²	RMS E
Ht = 30,41. R't - 8,63	0.85	0.73	0.32	Ht = 1,99+0,04.bt	0.87	0.75	0.31
Hd =3,89-13,42. R'd ³	0.84	0.71	0.33	Hd =3.57-2.35 Sd	0.84	0.71	0.33
Hk =4,34-5,44. R'k + 5,99. R'k ³	0.81	0.66	0.36	Hk=3.10-1.07Skix	0.81	0.66	0.36
Hf = 43,84-142,36. R'f + +247,82. R'f ³	0.76	0.57	0.40	Hf=4,40-0,07.bf+7,6. Sfix -3,89. Sf	0.82	0.68	0.35
Фосфор				Фосфор			
Pf = -29,89+107,59. R'f	0.60	0.36	3.58	Pf =0,63.bf	0.62	0.38	3.58
Pk = 7,80+16,69. R'k	0.58	0.34	3.64	Pk =13,07 + 5,83.Skix	0.59	0.34	3.64
Pd = 29.11. R'd	0.49	0.24	3.93	Pd =11.24 +15.8.Cdix	0.58	0.33	3.67
Pt = 60.06 -123.56. R't	-0.48	0.23	3.94	Pt = 16,39 - 0,26.bt	-0.50	0.25	3.90

В етап 11 диагностика на регресионния модел и анализ на остатъците (грешките) е необходимо да се провери дали грешките, имат нормално разпределение и дали е изпълнено условието за хомоскедастичност- остатъците да имат асимптотично нормално разпределение със средна стойност 0 и постоянна дисперсия. В разглеждания пример тази проверка е реализирана, чрез пакета Statistica като са сравнени хистограмата на разпределение на данните със съответната крива на нормално разпределение .

Следващ етап 12 от методиката е извършване на проверка на моделите, чрез заместване с данните от валидиращата извадка, получена в етап 8 и определяне на статистическите характеристики на модела. Тази проверка показва дали има статистически значима разлика между средните стойности от всеки набор от данни и стандартното отклонение на всяка група. Проверката в разглеждания пример е извършена с функцията t-test в пакета Statistica и нулева хипотеза за равенство на средни стойности. Изборът на най- добър калибровъчен модел за определяне на даден параметър се прави на база на статистическите параметри, получени от регресионната процедура и са основни критерии за оценка на построените регресионни модели. Получените резултати от примера са обобщени в таблица 2 а) и б), като най-добрите калибровъчни модели за получаване на хумус и фосфор от разглеждания пример са посочени с червен цвят.

Като критерии за оценка на построените регресионни модели се прилагат:

- Най-ниска стойност на стандартна грешка на калибровка или корен от средноквадратична грешка (**RMSE**), (Mihaylova, 2023)
- Най-висока стойност на остатъчното отклонение на прогнозата **RPD**, като стойностите са в интервала над 3 до 8 (Танеја, 2022), формула (Mihaylova, 2023).

Най-добрите калибровъчни модели са посочени с червен цвят.

Таблица 2. Характеристики на регресионни модели за определяне на хумус и фосфор преди и след валидиране

Калибровъчен модел				След валидиране		
Модел	R	R ²	RMSE модел	R ²	RMSE модел	RPD
HtSw	0.87	0.75	0.31	0.97	0.56	5.02
Hd	0.84	0.70	0.33	0.93	0.81	3.74
Hk	0.81	0.66	0.36	0.93	0.84	3.59
Hf	0.76	0.57	0.40	0.94	0.77	3.89
HdSw	0.84	0.71	0.33	0.92	0.85	3.54
HkSw	0.81	0.66	0.36	0.94	0.75	4.00

Калибровъчен модел				След валидиране		
Модел	R	R ²	RMSE модел	R ²	RMSE модел	RPD
Pd	0.60	0.36	3.58	0.90	4.73	4.56
Pf	0.60	0.36	3.58	0.87	5.34	4.03
Pf Sw	0.62	0.38	3.58	0.90	4.69	4.59
PdSw	0.63	0.40	3.50	0.89	4.81	4.48
PdSw	0.62	0.38	3.53	0.90	4.74	4.55
PkSw	0.59	0.34	3.64	0.90	4.73	4.56

За получаване действителната точност на така получените модели се провежда процедура крос-валидация с независим набор от цветови данни, която има за цел да тества до колко добре получените модели измерват показателите на почвата съдържание на органично вещество - хумус и фосфор като се сравнят действително измерените лабораторни стойности с получените стойности по модела.

ЗАКЛЮЧЕНИЕ

Предложената методика позволява моделиране на връзката на цветови характеристики на цифрови изображения на почва за определяне съдържание на органично вещество и усвоими форми на фосфор, чрез използване на регресионни методи.

Като научен статистически метод за изследване на връзки и зависимости регресионният анализ се явява подходяща техника за намиране на точни зависимости за прогнозно определяне показателите на качеството на почвата.

С цел изследване влиянието на цветовите фактори, които са статистически значими при определяне показателите на почвата, е предложено провеждане на еднофакторен и многофакторен регресионен анализ. Установено е, че точността на прогностичния резултат се повишава при прилагане на многофакторен регресионен анализ.

Калибровъчните модели от посочения пример, базирани на техниката MLR Stepwise постигат по-висока точност на определяне на изследваните почвени параметри -средно с 13% и намаление на корен от средноквадратична грешка (RMSE) средно с 5%.

Получените резултати показват, че предложената методика би била в помощ за полесното моделиране и оценяване на връзката между агрохимичните показатели и цветовите характеристики на почвата, получени от оптични устройства.

Получените по предложената методика регресионни математически модели са подходящи за внедряване в мобилни приложения за алтернативен, бърз, неструктивен метод за анализ на агрохимичните показатели на почвата.

БЛАГОДАРНОСТ

Настоящото изследване се подкрепя от Национална научна програма „Интелигентно растениевъдство“ на Министерството на образованието и науката, одобрена с Решение на Министерския съвет № 866 /26.11.2020 г.

REFERENCES

Bratov, K., Beloev, H., Mitkov A., & Mitev, G. (2020). "On the possibility of conducting fast and reliable soil tests (in Bulgarian)", *Mechanization in agriculture & Conserving of the resources*, vol. 66, Issue 2, pp. 71-76.

- Bünemann, K., Bongiorno, G., Bai, Z., Creamer, R., & De Deyn, G. (2018). "Soil quality – A critical review", ELSEVIER, *Soil Biology and Biochemistry*, vol. 120, pp. 105–125.
- Georgieva, Tsv., Mihaylova, A., Stefanov, E., Daskalov, P., & Sigrimis, N. (2020). "Knowledge engineering in smart agriculture to secure food and the environment", *Proceedings of University of Ruse*, vol. 59, book 3.3, pp. 69-76.
- Heil, J., Jörges, C., & Stumpe, B. (2022). "Evaluation of using digital photography as a cost-effective tool for the rapid assessment of soil organic carbon at a regional scale", ELSEVIER, *Soil Security*, vol. 6, pp. 1-10.
- Lavanya, G., Rani, C., & Ganesh Kumar, P. (2020). "An automated low cost IoT based Fertilizer Intimation System for smart agriculture", ELSEVIER, *Sustainable Computing: Informatics and Systems*, vol. 2, pp. 1–12.
- Levin, N., Ben-Dor, E., & Singer, A. (2005). "A digital camera as a tool to measure colour indices and related properties of sandy soils in semi-arid environments", *International Journal of Remote Sensing*, Vol. 26, No. 24, pp. 5475–5492.
- Masakazu K. & Shibusawa, S. (2020) "In-situ multiple parameter calibration and mapping using a mobile soil sensor", *Soil Science and Plant Nutrition*, 66:6, 837-846, DOI: 10.1080/00380768.2020.1834828.
- Mihaylova, A., Georgieva, Tsv., & Daskalov, P. (2023). „Impact evaluation of the optical devices for obtaining soil color characteristics”, in press.
- Mitkov A., & Bratov, K. (2023). "Statistical methods in agriculture and agricultural technology, Ruse University Academic Publishing House (**Оригинално заглавие:** Митков А., К. Братов, 2023г. **Статистически методи в земеделието и земеделската техника.** Русе: Академично издателство "Русенски университет".
- Mitkov, A. (2011). Theory of experiment (in Bulgarian), Ruse, Danube press (**Оригинално заглавие:** Митков А. 2011. Теория на експеримента. Русе: Издателство: Дунав прес.
- Mouazen, A., Maleki, M., De Baerdemaeker J., & Ramon, H. (2007). "On-line measurement of some selected soil properties using a VIS–NIR sensor", ELSEVIER, *Soil & Tillage Research*, vol. 93, pp. 13–27.
- Rasras, R., Ibrahim, M., El Emary, M., & Skopin D. (2006). "Developing a New Color Model for Image Analysis and Processing", *Computer Science and Information Systems, Research Gate*, vol. 4, Number 1, pp. 43-55.
- Rossel V., & Bouma, J. (2022). "Soil sensing: A new paradigm for agriculture", ELSEVIER, *Agricultural Systems* vol. 148, pp. 71–74, 2016.
- Rossel, V., Minasny, B., Roudier, P., & McBratney, A. (2007). "Colour space models for soil science", ELSEVIER, *Geoderma*, vol. 133, pp. 320– 337.
- Taneja, P., Kumar, H., Vasava, B., Fatholoulumi, S., Daggupati, P., & Biswas, A. (2022) . "Predicting soil organic matter and soil moisture content from digital camera images: comparison of regression and machine learning approaches", *Canadian Journal of Soil Science*, <https://doi.org/10.1139/cjss-2021-0133>.
- Todorova. M., Atanassova, S., & Ilieva, R. (2009). Determination of soil organic carbon using near-infrared spectroscopy“, *Agricultural science and technology*, vol.1, №2, pp 1-6.
- Wang, Y., Huang, T., Liu, J., Lin, Z., Li, S., Wang, R., & Ge, Y. (2015). "Soil pH value, organic matter and macronutrients contents prediction using optical diffuse reflectance spectroscopy", ELSEVIER, *Computers and Electronics in Agriculture*, vol. 111, pp. 69-77.