Assessment of Some Parameters of the Topsoil Layer from Multispectral Sentinel 2 Data in Conditions of the Novosibirsk Region

2021 International Conference on Information Technology and Nanotechnology (ITNT) | 978-1-6654-3217-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ITNT52450.2021.9649092

N. V. Rodionova
Institute of Radioengineering and Electronics RAS
Fryazino, Russia
rnv1948123@yandex.ru

S.Ya. Kudryashova
Institute of Soil Science and Agrochemistry SB RAS
Novosibirsk, Russia
sya55@mail.ru

A. S. Chumbaev
Institute of Soil Science and Agrochemistry SB RAS
Novosibirsk, Russia
chas30@mail.ru

***Abstract*—The paper considers the use of Sentinel 2 (S2) optical data for 2019-2020 to assess the content of humus, clay and moisture in the upper (0-10 cm) soil layer on the example of chernozems and gray forest soils of the Novosibirsk region. Special attention is paid to the selection of satellite images, it is necessary to meet the conditions for the soil to be dry and bare. The humus content was estimated for five test sites based on a regression model [1] and S2 spectral band B6 (740 nm). The model parameters were adjusted for the conditions of the study area separately for chernozems and gray forest soils. The clay content in the soil was estimated using S2 optical data, ground-based measurements, and regression models with an exponential dependence of the clay content on the reflection coefficients from the soil at the SWIR channels of S2 ([2], [3]). The change in the percentage of humus and clay content in the test sites soils per the year is shown.**

Keywords— Multispectral data, reflection coefficient, soil humus, soil clay content, soil moisture

I. INTRODUCTION

The optical properties of the soil are mainly influenced by four important factors: mineral composition, soil moisture, organic matter content, and soil texture. Various groups of spectral indices are used for remote reconstruction of these parameters.

In this paper, we consider the possibility of using the optical data of the Sentinel 2 satellite for 2019-2020 to estimate the content of moisture, humus (H) and clay (Clay) in the upper soil layer (0-10 cm) on the example of chernozems and gray forest soils of the Novosibirsk region. Special attention is paid to the selection of satellite images, because for such an assessment there are certain requirements for the soil, namely, the soil must be dry and without vegetation.

II. STUDY AREA

The object of the study is the soil of six test sites located in the Novosibirsk region. Soil samples were selected to determine their basic physical and chemical properties.

Fig. 1 shows the location of the test sites (highlighted by a red rectangle) at a distance of about 55 km east of Novosibirsk. The image was obtained according to Sentinel 2 image dated 23.4.2019 in natural colors (a combination of channels B4-B3-B2). Fig. 1 also shows a detailed map of the location of chernozems (slope of the southwest orientation) and gray forest soils (slope of the northwest orientation) with test site numbers.

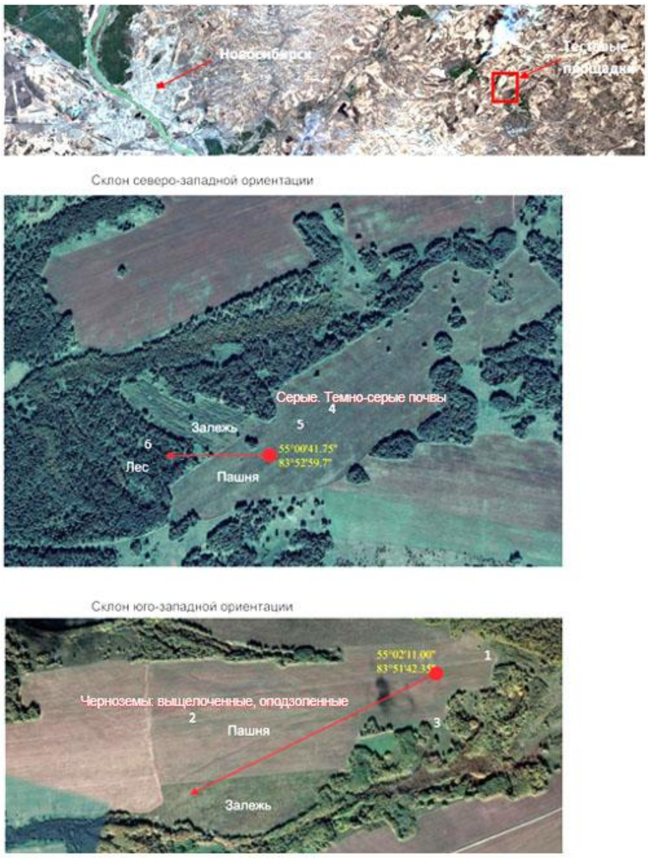


Fig. 1. Test sites in the Novosibirsk region

Table I shows the coordinates of the test sites, as well as the measured values of the percentage of humus and the granulometric composition of soil samples at the depth 0-10 cm (2019 year).

III. INPUT DATA AND RESEARCH METHODS

A. Sentinel 2 multispectral data

The ESA Sentinel 2A satellite was launched in June 2015, and the second Sentinel 2B in March 2017. The multispectral camera has 13 spectral bands spanning from the visible and near infrared to the short wave infrared. The spatial resolution varies from 10 m to 60 m depending on the spectral band. The temporal resolution one of S2 is 10 days, and two satellites – 5 days. Image processing was performed by SNAP [4].

The work was carried out within the framework of the state assignment of FIRE RAS and IPA SB RAS.

TABLE I. DESCRIPTION OF TEST SITES IN THE NOVOSIBIRSK REGION

Test Site	Coordinates (N, E)	Clay, %	Sand, %	Silt, %	Humus, %
1.Leached chernozem, unwashed arable land	55°02'12.3'', 83°51'47.8''	25.6	53.7	20.7	9.9
2.Leached chernozem slightly washed arable land	55°01'42.8'', 83°50'41.8''	19.8	56.8	23.4	8.3
3. Leached chernozem, slightly washed virgin soil	55°01'55.4'', 83°51'29.6''	22.1	51.8	26.1	8.6
4. Gray forest soil, unwashed arable land	55°00'42.0'', 83°53'01.1''	22.9	54.8	22.3	6.0
5. Gray forest soil, slightly washed arable land	55°00'40.5'', 83°52'54.1''	24.4	55.1	20.5	5.2
6. Gray forest soil, forest	55°00'37.8'', 83°52'31.8''	25.0	54.2	20.8	6.6

B. *Selecting S2 images*

For satellite assessment of the organic matter and clay content in the soil, restrictions are imposed on the choice of multispectral images associated with the selection of dry soils without vegetation (bare soils) in the images. In the paper [5] conditions for S2 channels are presented as follows [6]: 1) zero cloud cover in the study area, 2) the value of the vegetation index NDVI=(B8-B4)/(B8+B4)<0.35 to exclude green vegetation, 3) the difference in reflection coefficients (RC) between channels B3 and B2 and channels B4 and B3 should be greater than 0 (using these filters improves soil selection [5]), 4) the value of NBR=(B11-B12)/(B11+B12) should be NBR <=0.05, which allows to select pixels with dry bare soil. Soil moisture increases the absorption of light, and the RC is sharply reduced. Spectral channels B11 and B12 strongly correlate with soil moisture [7], and their difference allows distinguish between the spectra of dry, wet soil, as well as the spectra associated with vegetation. The choice of the threshold for NBR strongly affects the number of pixels in the image that meet the conditions of dry soil. Increasing the NBR threshold to 0.15 reduces the number of 'needed' pixels by a factor of 2 [6], i.e. an increase in NBR leads to decrease in the accuracy of models for determining soil parameters from satellite data.

As a result for the study area, the optical images S2 L2 dated 23.4.2019 and 22.4.2020 were selected, for which the condition of no clouds is met, NDVI changes from 0.17 to 0.24, the difference between the channels B3 and B2 and channels B4 and B3 is greater than 0, the NBR values slightly exceed the threshold of 0.05, namely, NBR changes from 0.1 to 0.158. These conditions are met for 5 test sites, except for site No.6 (forest)

C. *Estimation of humus content in the topsoil layer according to S2 data*

SOC losses (SOC – soil organic carbon) are one of the main causes of arable land degradation. Thus, spatial and temporal monitoring of SOC is an extremely important task, many works are devoted to solving the problem ([1], [6], [8], [10] and many others). These authors have obtained regression models with an exponential relationship between the SOC content and the reflection coefficient (RC) with a different degree of negative correlation between the SOC content and the RC values, depending on the spectral channel. Satellite and ground-based data are needed to create a regression model of the RC-H (humus) relationship, and the number of soil samples must be at least 20 to determine the correlation. In addition, the test areas under study should have a similar granulometric composition. It is the fact that the finer the soil particles, the greater the RC from these soils [9]. Since the models for determining humus are local, so the use of models available in the literature requires adjustments for the study area.

In the works [6], [8] it was shown that the best correlation between the H content in the soil and RC is found in the spectral channels S2 B4-B6 and B11, B12. For a qualitative assessment of changes in the H content in the soils of the test sites under study, we present graphs of the difference in the RC values of the spectral channels S2 B4-B6, B11, B12 for the survey sessions 23.4.2019 and 22.4.2020 (Fig. 2) for each test site, and estimate the changes in the RC for the year. Taking into account the fact that the RC of these channels has a negative correlation with the H content in the soil, we will estimate by the sign of the difference in which direction the H content changed during the year. The difference in RC values is positive for site No. 1 (chernozem, unwashed arable land) for all channels except B12. This indicates that the RC values for site No. 1 increased in 2020 compared to 2019, i.e., the H content decreased.

For the washed-out chernozem (site No. 2), there is a decrease in the difference for four channels and an increase in the difference for one. For site No. 3 (virgin land), the difference in RC for all channels is negative, which may indicate an increase in the content of H. The same situation is for site No. 4 (gray forest soils not washed). For weakly washed gray soils (site No. 5), there is a positive difference for three channels, and a negative difference for two. To quantify H, you need a local model for the test area, or use existing models with adjustments for the local area with its own granulometric composition of the soil.

In 1927, Pokrovsky first proposed an exponential equation that determines the relationship between the H content and the RC values (according to [10]):

$$\rho_{750} = \rho_{750,h} + \mathrm{A}\,\mathrm{e}^{-\kappa H} \quad , \qquad (1)$$

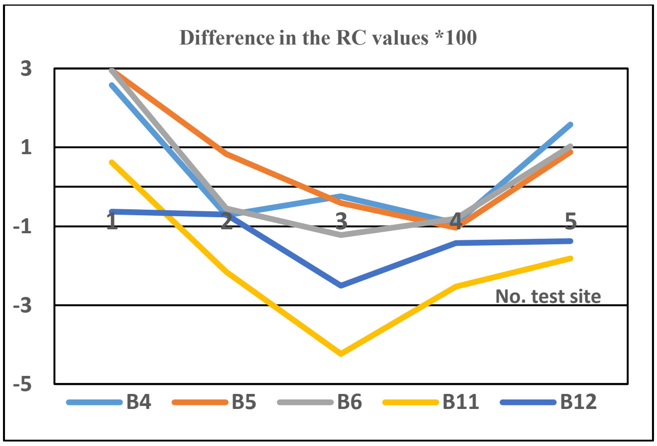


Fig. 2. The difference in the RC values from the soils of the five test sites for the survey sessions 22.4.2020 and 23.4.2019.

where ρ_{750}-reflectance at a wavelength of 750 nm, $\rho_{750,h}$ - reflectance multi-humus soil, $\rho_{750,0} = (\rho_{750,h} + \mathrm{A})$ – reflectance humus-free soil, H- humus content, k is the coefficient that determines the steepness of the exponential graph.

When using the exponential dependence of the RC on the H content in dry bare soil, three important points should be taken into account: 1) with a small H content (up to 3%), its determination is most accurate, but the spread of the RC values corresponding to this humus content is large, 2) with H content values of 6-7 % or more, the determination of the humus content from the RC is inaccurate, the influence is exerted with the same H content by soil moisture, processing conditions [9], 3) the decisive factor is the choice of satellite survey dates that allow identifying pixels with dry bare soil.

In this paper, it is not possible to find out the correlation between the RC and the humus content in 5 test sites of the Novosibirsk region due to insufficient number of ground-based measurement points. Judging by the available ground-based measurements of the humus content in the soils of the test sites (Table I), we assume that the curves of the dependence of RC on H are represented by the receding areas, where the accuracy of determining the content of H from RC is low.

In this paper, to determine the H content from satellite data, we use an exponential model with parameters for chernozems and gray forest soils, presented in [1].

Note that to match model we need to take the optical data for the wavelength of 750 nm. For S2, it is a B6 channel with a wavelength of 740 nm, a bandwidth of 15 nm, and a spatial resolution of 20 m. To use the parameters of the exponent equation [1] in the local conditions of the test areas under study, it is necessary to adjust these parameters. One of the possible options for assessing the humus content in the soils of the test sites is to use the following parameters for chernozems: $\rho_{750,h}$=8.0, A=29.1 and k=0.1256. For gray forest soils, $\rho_{750,h}$=8.5, A=40.5 and k=0.28. Table 3 shows the calculated values of the humus content for five test sites with these coefficients of exponential equation.

The difference between the obtained values of the percentage of H according to the equation (Table II) and the values obtained under laboratory conditions (Table I) is not more than 1.5% in absolute value. The sources of error lie primarily in the insufficient amount of ground data. For each type of soil (chernozems, gray forest soils), about 20 or more ground-based measurements are required, one part of which is used to obtain the exponent parameters and the second part to validate the resulting equation. The more such ground data, the more reliable the formula is. The second point that leads to errors is the choice of the original optical image, for which the most important parameter is the value of the NBR spectral index. Increasing the NBR value >0.05 significantly reduces the number of pixels that meet the condition of dry bare soil. Nevertheless, some conclusions about the quantitative content of humus in the soil can be made, first of all, about the change in H values for the year (2019-2020). The greatest difference is obtained for the site No. 1 - a decrease in the humus content for the year by 2.4 %. For the other sites, the changes for the year are insignificant from 0.3 to 0.9 %.

D. Estimation of the topsoil clay content from S2 data

Numerous studies have been devoted to the assessment of soil texture components using remote optical data ([2], [3], [8], [11], [12] and others). Thus, in the work [8], it was shown that the best correlation with the clay content in the soil is the reflection coefficient of the channel B7, as well as the spectral indices V=B8/B4 (Vegetation index), SAVI=1.5*(B8-B4)/(B8-B4+0.5) (Soil Adjusted Vegetation Index) and others. Weak correlation of S2 with silt and sand is noted. The authors [2] found that channels B11 (SWIR1) and B12 (SWIR2) S2 are most sensitive to changes in the clay content in the soil, and have a negative correlation with this content. Moreover, for the use of these channels, the condition for obtaining a quantitative assessment of the clay content is dry, bare soil. The paper [13] introduced the clay spectral index CI=B11/B12 (Clay Index), which has a strong negative correlation with the clay content in the soil. It should be noted that the quantitative assessment of the clay content in the soil is carried out locally for the study area, and the accuracy of the assessment is directly related to the number of test measurements of soil samples.

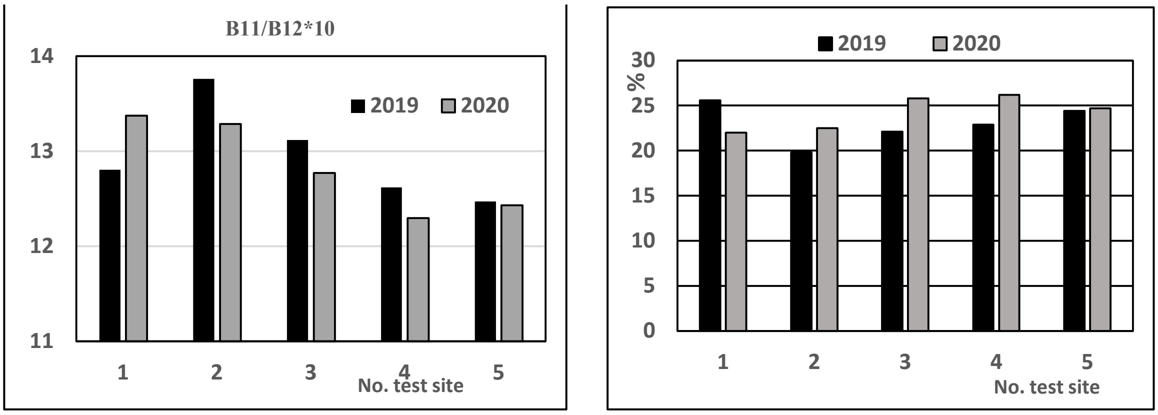
Fig. 3 (a) shows graphs of CI=B11/B12 values for two S2 images dated 23.04.2019 and 22.04.2020, which allows to make a qualitative comparison of changes over the year in the clay content in the topsoil layer of the test sites No. 1-No. 5. The changes in the clay content for the year did not affect site No. 5, for sites No. 2-No. 4 there is a slight increase in the clay content, and a slight decrease for site No. 1.

The regression model [6] of the exponential relationship of the SOC content in the soil with RC may be used to quantify the percentage of clay in the soil. For the case of clay, we use the clay index CI=B11/B12 as a variable in the exponent, which has a negative correlation with the clay content in the soil.

The equation for quantifying the percentage of clay in the soil is as follows: Clay(%)=802*exp(-2.69*CI) – for chernozems, Clay(%)=5123.6*exp(-4,29*CI)- for gray forest soils. Part of the ground-based measurements of the clay content in the soil of the test sites in the Novosibirsk region were used to obtain the exponent parameters, and the remaining part was used for validation. It should be noted that the available number of ground-based measurements is not sufficient for the reliability of the application of the obtained equations.

TABLE II. EXPONENT EQUATION PARAMETERS AND THE HUMUS PERCENTAGE IN TEST SITE SOILS DETERMINED BY THE EQUATION

	$\rho_{750,h}$	A	k	Test site	H, %, 2019	H, %, 2020
Gray forest soil	8.5	40.5	0.28	4	6.0	6.4
				5	5.8	5.4
Leached chernozem	8.0	29.1	0.126	1	9.9	7.5
				2	6.8	7.1
				3	7.7	8.6



(a) (b)

Fig. 3. Assessment of changes in the values of the clay index CI for the year (a) and changes in the values of the percentage of clay in the soils of test sites (b).

Checking for the presence of a correlation of CI with elements of the soil texture showed that there is no negative correlation of CI with silt and physical clay.

Fig. 3 (b) shows graphs of the percentage of clay content in the soils of the five test sites (measured values) and calculated using the equations for 2020. The qualitative assessment of the clay content in the soils of the test plots based on the clay index value does not contradict the quantitative assessment based on the regression model.

E. Estimation of the topsoil layer moisture from S2 data

The paper [14] uses the NDDI (normalized difference drop index) to estimate soil moisture:

$$NDDI = \frac{NDVI-NDWI}{NDVI+NDWI} \quad (2)$$

where $NDWI = \frac{B8A-B11}{B8A+B11}$ or $NDWI = \frac{B8A-B12}{B8A+B12}$ - normalized difference wet index, B8A, B11 и B12- spectral channels S2. The authors show that higher NDDI values correspond to lower values of soil moisture. In this paper, the channel B12 was used to calculate the NDWI. The NDDI graph for the studied sites with chernozems (sites No.1, No. 2, No. 3 (virgin land)) and gray forest soils (No.4, No.5, No.6 (forest)) is shown in Fig. 4 for the optical survey sessions for April (20, 27, and 30) and May (10, 12, and 17) 2020.

There was a strong differentiation of NDDI values for different sites for 20.4.2020, and the grouping of test sites by close values of soil moisture.

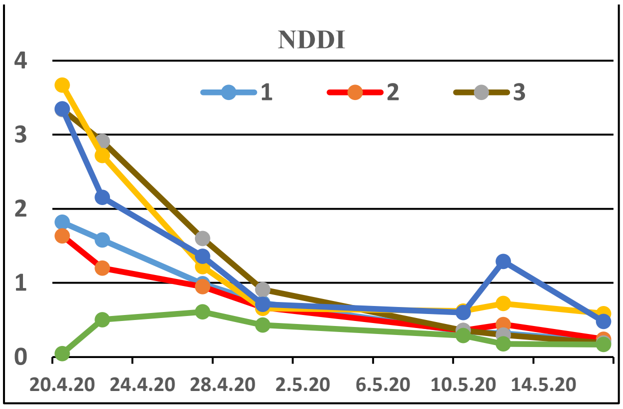


Fig. 4. NDDI graphs for the study sites with chernozems and gray forest soils for April and May 2020.

The soil of site No. 6 (forest) has the highest humidity. Sites No. 4, No. 5 (gray forest soils, arable land) and No. 3 (chernozems, virgin land) are characterized by the lowest humidity. Between these extreme values, sites No. 1 and No. 2 are grouped (black earth, washed and washed away, arable land). During the transition from April to May, the differentiation of sites by NDDI values significantly decreases, leading to the proximity of the soil moisture values for all test sites.

IV. CONCLUSION

The paper estimates the percentage of humus and clay in the topsoil layer of 5 test sites with chernozems and gray forest soils in the Novosibirsk region in 2019-2020 based on the use of ground-based measurements, regression models and satellite S2 data. The change in the percentage of H and clay in the soils of the test sites over the year is shown. For site No. 1 (leached chernozem, unwashed arable land), the humus content decreased by 2.4 % over the year. For the other sites, the changes for the year are insignificant from

0.3 to 0.9 %. The difference between the obtained values of the percentage of H, determined by the exponential equation (Table II), and the values obtained in laboratory conditions (Table I) for 2019, is not more than 1.5% in absolute value. The sources of error lie primarily in the insufficient amount of ground data for each type of soil (chernozems, gray forest soils). The second point that leads to errors is the choice of the original optical image, for which the most important parameter is the value of the spectral index NBR. A similar situation holds for estimating the percentage of clay.

For the percentage of clay content, the changes for the year according to S2 data did not affect site No. 5 (gray forest soils, slightly washed arable land). For chernozem (site No.1), a decrease in the clay content was noted for the year, while for the other sites (No.2- No. 4), a slight increase was noted.

REFERENCES

[1] E.I. Karavanova, D.S. Orlov, “Assessment of humus content in soils by their spectral reflectivity,” Agrokhimiya, no. 1, pp. 3-9, 1996.

[2] S. Bousbih, M. Zribi, Ch. Pelletier, A. Gorrab, Z. Lili-Chabaane, N. Baghdadi, N. Ben Aissa, and B. Mougenot, “Soil Texture Estimation Using Radar and Optical Data from Sentinel-1 and Sentinel-2,” Remote Sens., vol. 11, pp. 1-20, 2019, doi:10.3390/rs11131520.

[3] M. Shabou, B. Mougenot, Z. Lili-Chabaane, C. Walter, G. Boulet, N. Aissa, and M. Zribi, “Soil Clay Content Mapping Using a Time Series of Landsat TM Data in Semi-Arid Lands,” Remote Sens., vol. 7, pp. 6059–6078, 2015.

[4] The european space agency. Sentinel-1 toolbox. URL: https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-1 (accessed 26.11.2021).

[5] J.A.M. Demattêa, C. T. Fongaroa, R. Rizzob, and J. L. Safanellia, Remote Sensing Environ., vol. 212, pp. 161–175, 2018.

[6] F. Castaldi, S. Chabrillat, A. Don, and B. van Wesemael, “Soil Organic Carbon Mapping Using LUCAS Topsoil Database and Sentinel-2 Data: An Approach to Reduce Soil Moisture and Crop Residue Effects,” Remote Sensing, vol. 11, pp. 1-15, 2019, doi: 10.3390/rs11182121.

[7] H. B. Musick, R. E. Pelletier, “Response to soil moisture of spectral indexes derived from bidirectional reflectance in thematic mapper wavebands,” Remote Sens. Environ., vol. 25, pp. 167–184, 1988.

[8] A. Gholizadeh, D. Žižala, M. Saberioon, and L. Borůvka, “Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging,” Rem. Sens. of Environment, vol. 218, pp. 89-103, 2018.

[9] E.I. Karavanova, Optical properties of soils and their nature. MGU, Moscow, 2003.

[10] D. S. Orlov, N. I. Sukhanova, M. S. Rozanova, Spectral reflectivity of soils and their components. MGU, Moscow, 2001.

[11] P.A. Ukrainskii, O.A. Chepelev, “Study of the granulometric composition of the soils of Poskolya according to the data of decoding satellite images,” Izv. Samarskogo nauch. tsentra RAN, vol. 13, pp. 1225-1229, 2011.

[12] E. Vaudour, C. Gomez, Y. Fouad, and P. Lagacherie, “Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems,” Remote Sens. Environ., vol. 223, pp. 21–33, 2019.

[13] T. Hengl, A Practical Guide to Geostatistical Mapping of Environmental Variables. Luxembourg: Publications Office, 2007.

[14] K. Burapapol, R. Nagasawa, “Mapping Soil Moisture as an Indicator of Wildfire Risk Using Landsat 8 Images in Sri Lanna National Park, Northern Thailand,” Journal of Agricultural Science, vol. 8, pp.107-119, 2016.