FRI-10.326-1-EEEA-10

SOFTWARE TOOLS FOR HYPERSPECTRAL IMAGE PREPROCESSING IN SOIL SAMPLE ANALYSIS [10]

Georgi Manchev
Department of Automatic and Electronics,
"Angel Kanchev" Univesity of Ruse
E-mail: gmanchev@uni-ruse.bg

Assoc. Prof. Stanislav Penchev, PhD
Department of Automatic and Electronics,
"Angel Kanchev" Univesity of Ruse
E-mail: msp@uni-ruse.bg

Principal Asst. Prof. Eleonora Nedelcheva, PhD
Department of Automatic and Electronics,
"Angel Kanchev" Univesity of Ruse
E-mail: ekirilova@uni-ruse.bg

Assoc. Prof. Tsvetelina Georgieva, PhD
Department of Automatic and Electronics,
"Angel Kanchev" Univesity of Ruse
E-mail: cgeorgieva@uni-ruse.bg

Prof. Plamen Daskalov, PhD
Department of Automatic and Electronics,
"Angel Kanchev" Univesity of Ruse
E-mail: daskalov@uni-ruse.bg

***Abstract**: This article presents the sequence of actions implemented in the Orange3 product related to the pre-processing of hyperspectral images of synthetic soil samples. This stage of the analysis precedes the separation of informative features and the construction of models for quantitative assessment of the content of nitrogen-containing compounds in the soil sample. The main stages of the pre-processing of hyperspectral images are: defect removal, normalization, selection of a region of interest and obtaining an average spectral characteristic, removal of uninformative parts of the spectral data, transformation and filtration of the data. The quality of the obtained results shows that the Orange3 software package is suitable for the needs of the study, as it provides the opportunity for a flexible and intuitive way of working, all the necessary tools and good graphical capabilities for visualization of results.*

***Keywords**: Hyperspectral Imaging, Orange3, Spectral preprocessing, Soil analysis.*

INTRODUCTION

Hyperspectral images (HSI) combine both spatial and spectral information about the observed object. They are characterized by spatial (px, py) and spectral (wavelength range, number of wavebands) resolution, with each spatial pixel having its own unique spectrum. The volume of information contained in an HSI many times exceeds that of RGB or multispectral images, and the approaches to processing and analyzing the spectral part deeply contrast with the methods applied to normal digital images. All this seriously complicates their processing by conventional means

[10] The paper was presented on 25 October 2024 in section "Electrical Engineering, Electronics and Automation" with original title in English: SOFTWARE TOOLS FOR HYPERSPECTRAL IMAGE PREPROCESSING IN SOIL SAMPLE ANALYSIS

and requires the use of specialized software tools, through which the relevant data, which are informative for the research being conducted, can be effectively extracted. To address these problems, a number of libraries and modules have been created over the years for software environments such as IDL, GDL, FL, Matlab, SciLab, Python, R, etc. There are also both commercial and free specialized software (Al-Khoder A. et al., 2015, Arslan, M. 2014, Jović A. et al., 2014), aggregating established methodologies and tools that facilitate the data processing process - ENVI, Unscrambler X, eCognition, Pirouette, Orange3, etc. All of them, to a greater or lesser extent, provide the opportunity for interaction and/or direct integration with other environments.

From the point of view of the primary analysis of agrochemical indicators of soils, the spatial part of HSI is not particularly informative and can rather be positioned as an auxiliary characteristic in the preparation and extraction of spectral data, and the methods for their qualitative processing and analysis are strictly focused on the field of spectroscopy. In essence, spectroscopic measurements are monitoring the interaction of light energy (considered as radio-magnetic radiation) with matter (the object of study). This imposes specific requirements on the preparation of samples, instrumentation and the recording process itself. The type of light source, the spectrum in which it emits, as well as its characteristics, the spectral sensitivity of the sensor, preliminary preparation of samples and the method of recording are key factors determining the accuracy of the measurement. Other essential aspects are the methods and tools for limiting the negative effects caused by noise and defects in HSI.

This article presents the sequence of actions related to the pre-processing of hyperspectral images of synthetic soil samples. This stage of the analysis precedes the separation of informative features and the construction of models for quantitative assessment of the content of nitrogen-containing compounds in the soil sample. The algorithms and procedures presented here are implemented in the Orange3 programming environment (Demsar J. et al., 2013) and are part of an approach for assessing some basic agrochemical indicators of soil.

EXPOSITION

Defect removal in HSI

For various reasons (large volume or errors in data transfer, defects in the detector, etc.), defects in HSI may appear. Their manifestations may be of different nature, but in general they have a strongly pronounced negative effect, both in the spectral and spatial part of HSI. There are many methods for their identification and tools for their removal, some of which are mentioned by (Jiang, T. et al., 2018, Manchev G. et al., 2024). It is essential for the quality of the study that HSI with detected defects be corrected in advance or, if correction is impossible, be excluded from the analysis.

Normalization of HSI

In studies involving more than one HSI, it is necessary that the data from all of them be comparable, especially if they are captured under different conditions or over a long-time interval. To address these problems, a normalization process is applied. There are two main types of normalization methods: the first is a simple normalization method that only requires information from the spectrum. The second type includes methods such as Multiplicative Scatter Correction (MSC) (Helland, I. et al., 1995) and Extended Multiplicative Signal Correction (EMSC) (Afseth N. Et al., 2012) that require the presence of a reference spectrum. Orange3 has tools for both types. Since the first type is applicable for this study, its application will be discussed in detail. It essentially represents a linearization of the spectral response obtained from the captured sample with respect to the spectral characteristic of the light source, its own spectral noise and the sensitivity coefficient of the detector. In the general case, the spectral response for a spectral band can be represented in the form (1) from which it is evident that the change in the intensity of the light source and the spectral sensitivity of the detector lead to a change in the recorded spectral response.

$$I_{res} = I_{src} * K_{refl} * K_{det} \qquad (1)$$

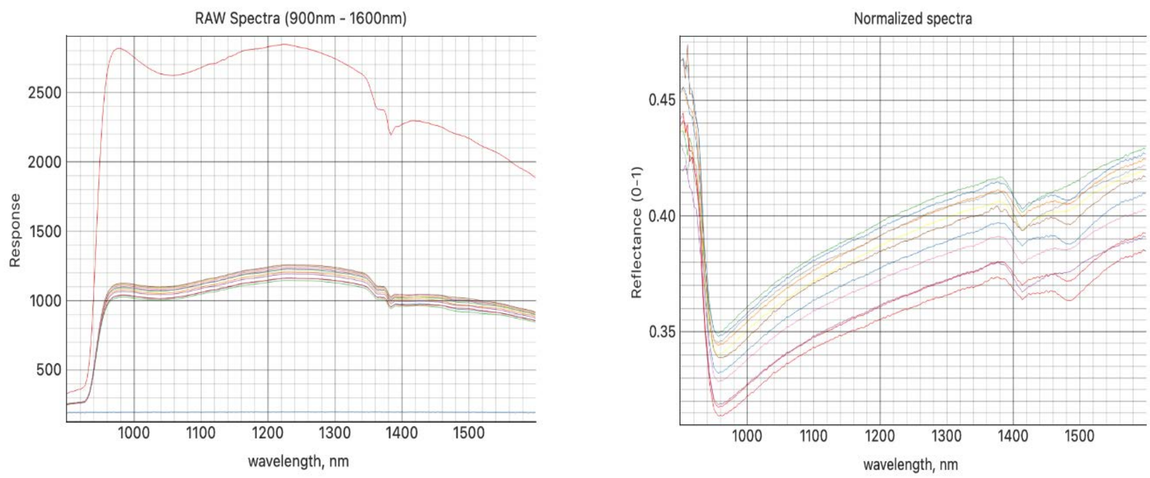
where I_{res} – registered response; I_{src} – source intensity; K_{refl} – reflectance coefficient; K_{det}– sensitivity coefficient.

Normalization (2) can be performed both for each pixel separately and by using average spectral responses depending on whether the focus of the study is on the spatial or spectral part of the HSI, and is applied to each spectral band separately.

$$I_{norm} = \frac{I_r - I_d}{I_w - I_d} \qquad (2)$$

where I_{norm} – normalized intensity; I_w – intensity of white target with >99% reflectance; I_d – intensity of dark (fully covered lens); I_r – intensity of sample.

Fig. 1a shows the average spectral responses of soil samples (middle lines), spectral response from white target (top line) and self-noise spectral response of the detector (bottom line), and Fig. 1b shows the resulting normalized average spectral characteristics of the soil samples.



(a) (b)

Fig. 1. RAW spectral responsens (a); Normalizes spectral responses (b).

The processing was performed using the Preprocess tool in Orange3 shown in Fig. 2. It provides various normalization options, and in this study a classic standard normalization method in the interval 0 … 1 was chosen.

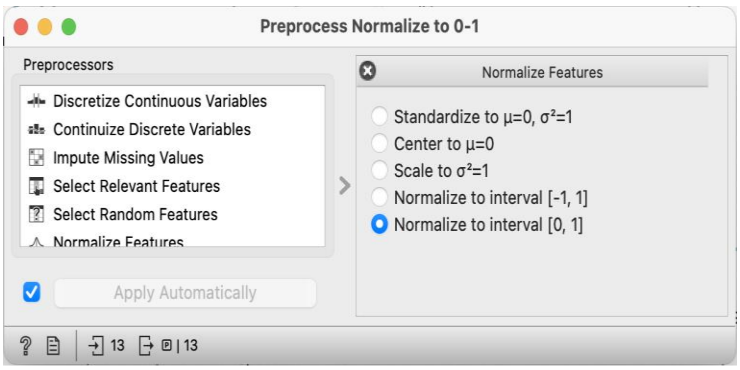


Fig. 2. Preprocess widget of Orange3 software.

Region of interest (ROI) and average spectral characteristics.

When working with average spectral characteristics, the selection of a region of interest is essential. Due to the fact that the soil samples involved in this study are homogeneous, an ROI covering the largest possible area of the sample was selected. The main advantages of such an approach are the extraction of a sufficiently good representative sample and a significant reduction in spectral noise due to the large number of spectral pixels involved in the averaging. Since the result of the used hyperspectral tool (fig. 3) is spatially limited by the ROI hypercube, it was combined with the average spectra tool to generate an average spectral characteristic.

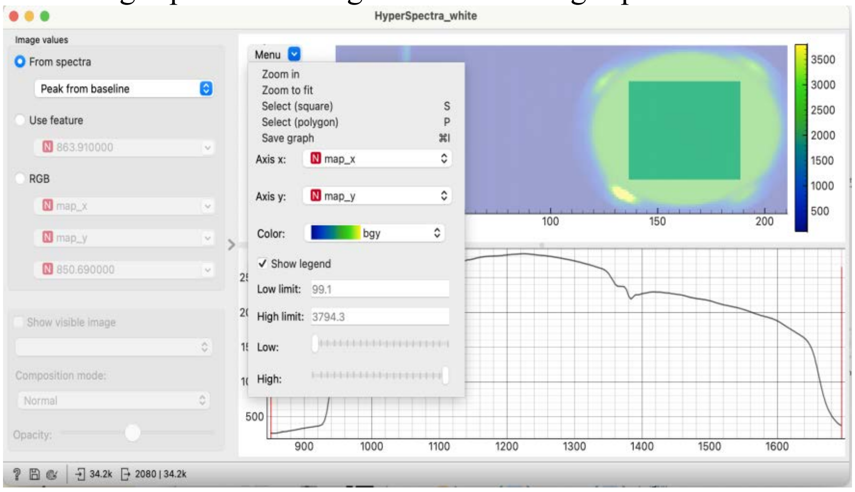


Fig. 3. ROI selection by using hyperspectral widget.

Excluding spectral data

It was found that the hyperspectral camera used in this study has a low spectral sensitivity at the beginning and end of the spectral range. From expressions (1) and (2) it is evident that for the bands with low spectral sensitivity of the sensor, the spectral response registered when capturing the test panel with maximum reflectance has a low amplitude in the absence of a change in the level of the registered intrinsic noise. This leads to a significant amplification of the negative effect of the deteriorated signal-to-noise ratio during normalization. This necessitates the exclusion of these bands from the study. The narrowing of the spectral range was done by selectively excluding spectral regions in the Spectra Preprocess widget.

Data concatenation and merge

The merging of the average spectral characteristics for all soil samples was performed using the Concatenate widget, and the addition of information about sample type and amount of nitrogen input and humidity to them was done using the Merge widget, by loading an external csv file with a data set prepared according to the requirements for describing types and classes of Orange3.

Data transformation, filtration and noise reduction

There are many methods for transforming and filtering spectral data, and Orange3 is equipped with a rich set of tools for this. In order for the effect of working with them to be positive, their application must be justified and as limited as possible. In this study, only the Savitsky–Golay filter from the Spectra preprocess widget was used to remove the difference in baselines, filter the spectral data, and generate the first derivative.

Import and export

Data import and export were done using Load & Save widgets, using a popular universal data transfer format, csv, which provided the ability to exchange data with other software environments and transfer results to other projects in Orange3.

CONCLUSION

The Orange3 software package has a flexible and intuitive environment, enabling quick and easy graphical construction of a project workflow. It combines a set of tools that fully cover the basic needs for pre-processing HSI images of soil samples and the preparation, transformation and extraction of spectral data from them for subsequent analysis. The preliminary application of methods for identification and removal of defects in HSI is essential for the quality of the results obtained. Choosing an ROI covering the largest possible area of the studied homogeneous soil samples leads to a significant reduction in noise in the average spectral characteristic. Normalization is an established and effective method for linearizing and comparing spectral characteristics from different samples, but leads to an increase in the negative effect of the deteriorated signal/noise ratio in the bands with low spectral sensitivity of the sensor, which necessitates their exclusion by narrowing the spectral range. The application of a limited number of tools is essential for limiting the total accumulated error from conversion and transformation.

ACKNOWLEDGMENTS

The report reflects the results of the work on project No. 24-FEEA-05, financed by the "Scientific Research" fund of the University of Ruse, Bulgaria.

REFERENCES

Afseth N. & Kohler, A. (2012). *Extended multiplicative signal correction in vibrational spectroscopy, a tutorial.* Chemometrics and Intelligent Laboratory Systems 117 (2012) 92–99

Al-Khoder A. & Harmouch, H. (2015). *Evaluating four of the most popular Open Source and Free Data Mining Tools.* IJASR International Journal of Academic Scientific Research ISSN: 2272-6446 Volume 3, Issue 1 (February - March), PP 13-23

Arabas, S., Schellens, M., Coulais, A., Gales, J., Messmer, P., others, (2010). *GNU Data Language (GDL) - a free and open-source implementation of IDL.* Poster presented at the European Geosciences Union General Assembly 2010, Vienna, Austria

Arslan, M. (2014). *Free and open source software as a public good: Implications for education.* European Journal of Research on Education, 2014, 2(Special Issue), 158-165

Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, Mozina M, Polajnar M, Toplak M, Staric A, Stajdohar M, Umek L, Zagar L, Zbontar J, Zitnik M, Zupan B (2013) Orange: Data Mining Toolbox in Python, Journal of Machine Learning Research 14(Aug): 2349−2353.

Helland, I., Naes, T., Isaksson, T. (1995). *Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data.* Chemometrics and Intelligent Laboratory Systems 29 (1995) 233-241

InfoMetrix, (©2024). Pirouette - Comprehensive Chemometrics Modeling Software. URL: https://infometrix.com/pirouette/

Jiang, T., Zhuang, L., Huang, T. & Bioucas-Dias, J. (2018). *Adaptive hyperspectral mixed noise removal.* DOI: 10.1109/IGARSS.2018.8519303

Jović, A., Brkić, K. & Bogunović, N. (2014). *An overview of free software tools for general data mining.* 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 2014, pp. 1112-1117, doi: 10.1109/MIPRO.2014.6859735.

Manchev, G., Penchev, S., Georgieva, Ts., Kirilova, E., & Daskalov, P. (2024). *Automatic Detection and Removal of Spiked Points in Hyperspectral Images.* Eng. Proc. 2024, 70, 32. https://doi.org/10.3390/ engproc2024070032

NV5 Geospatial Solutions, Inc., (2024). IDL – Interactive data language. URL: https://www.nv5geospatialsoftware.com/Products/IDL