УДК 004.942+004.383.3
В.О. Яценко
Інститут космічних досліджень НАНУ-НКАУ
В.В. Гніденко
Національний технічний університет України «Київський політехнічний інститут»
М.В. Наливайчук
Національний технічний університет України «Київський політехнічний інститут»
Н.І. Адамчук-Чала
Інститут мікробіології та вірусології ім. Д.К. Заболотного

MEASURING-COMPUTING HYPERSPECTRAL LYDAR SYSTEM FOR DETECTING CHEMICAL AND BIOLOGICAL AGENTS

Створена вимірювально-обчислювальна система дистанційного виявлення хімічних та біологічних речовин. Система перевірена вимірами спектрів поглинання та відбиття забруднюючих компонент. Виміри показали на можливість отримання достатньо сильного розсіювання від різноманітних об'єктів на земній поверхні. Отримані дані показали можливість дистанційного гіперспектрального виявлення хімічних та біологічних речовин в різні моменти часу.

Ключові слова: *вимірювально-обчислювальна система, лідар, лазер, дистанційне зондування, алгоритм, хімічні та біологічні речовини*

Форм. – 2. Табл. – 3. Рис. – 8. Літ. – 7.

1. Introduction

Light detection and ranging (LIDAR) provides active optical remote sensing in backscattering [1-3]. A pulsed laser transmitter is used and light backscattered from molecules and particles is collected by an optical telescope and is detected and range resolved in radar like mode. Chemical and biological content in high plant leaves is an important characteristic showing on plant status. For example, pigment level correlates with power of photosynthetic apparatus and may be used for yield prognosis. Influence of various stresses on photosynthetic apparatus, first of all, induces changes in chlorophyll content. This phenomenon allows to develop effective approaches to fast revealing stresses in phytocenosis of various types, using remote estimates of chlorophyll content and its dynamics. It is well known that reflectance spectra of plant leaves in an optical range are the most informative in respect to chlorophyll content due to the fact that they are formed because of specific spectrum of the pigment. Moreover, these spectra are the suitable characteristic for remote measurements.

Therefore, an idea to apply this characteristic for realization of agrocenosis monitoring and other types of plant canopies seems to be very attractive. Now there are numerous formulae based on regression relations to calculate chlorophyll concentration in leaves by various spectral coefficients of reflectance or their combinations. Practically, all of them are effective enough when chlorophyll concentration is measured under laboratory conditions. When chlorophyll determination is performed for «soil-vegetation» system then estimation results may be distorted. The main cause is a contribution of soil reflectance. There are several approaches to minimize this type of noise by using complex vegetation indices that include components for correction of soil interference. Their application demands measurements of additional parameters of the system, such as leaf index and spectral coefficients of soil reflectance. This condition makes the procedures low effective and moreover, the problem is not solved completely because of variability of soil reflectance, for example, on dependence on its moisture. Therefore, it is very important to develop effective methods acceptable for chlorophyll measurement in the "soil-vegetation" system.

Application of high resolution spectroscopy for remote sensing of vegetation has displayed new characteristics correlating with chlorophyll content. It appeared that a position of the, so called, red edge in spectral curves of reflectance and the shape of this spectral region were dependent on pigment concentration. Various characteristics of red edge region (680-800 nm) have been used. Position of maximum in the 1-st derivative plot of reflectance spectral curve [11] and position of red edge were found to be in regression relation with chlorophyll content. We have revealed a high correlation between the ratio of two maxima in the 1-st derivative plot and chlorophyll. A high resistance of this characteristic to soil contribution has been shown as well. A shortage of this approach is its sensitivity to instrumental noises, which may considerably distort the results, influencing, for example, intensity in maxima and their position. Thus, to develop a device for remote measurements of chlorophyll content in vegetation using the 1st derivative characteristics it needs to solve the following problems: 1) to have such level of noise in the 1st derivative plots which permits to distinguish two maxima, 2) to develop an algorithm for

automatic determination of maximum positions which vary depending on chlorophyll concentration as well, 3) to select appropriate characteristics and computing procedure to calculate chlorophyll content, 4) to develop the procedure to minimize a soil reflection contribution. This work is devoted to solving the problems 1-3.

Our research focuses on a correction of methodology for early detection and identification of biological and chemical agents including various toxins and viruses using network of sensors. The ecological interest is in response to a proliferation of agents development and threats for human health. Additionally, there is a pervasive interest across diverse application areas such as medicine, environmental protection, and vegetation processing to achieve a rapid detection and identification capability of various agents.

2. Methodology

2.1. Polymeric films technology

The following technologies have been used:

- A new plant substratum of total resolution
- Polymeric films technology for direct microbial community observation
- Advantages & resolutions of polymeric films technology
- Observation of total microbial community architecture
- Visualization of individual components of soil microbial communities (Fig. 1)
- Reconstruction of 3D view of soil microbial communities architecture

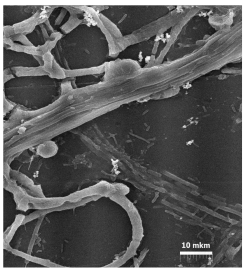

Fig. 1

2.2 Plant object, spectral and biochemical measurements

Leaves of two varieties of winter wheat were studied during the experiments, and appropriate measurements had been collected. Variations of chlorophyll content were caused by differences in mineral nutrition and plant age during vegetation. Reflectance spectra were measured with spectrophotometer SF-10 (USSR) equipped with integrating sphere. The recording system allowed to obtain the spectra in digital form. The spectra were recorded with the increment of 1 nm in the range of 680-750 nm. The pieces of the wheat leaves used for recording of reflectance spectra, and the data were used for chlorophyll determination by chemical method by Arnon.

2.3 Data treatment

2.3.1 Preliminary processing of experimental data

The preliminary data processing is based on application of smoothing procedure proposed in [4-7]. To perform data smoothing we used several moving windows of different size: 5, 7 and 9 point windows. After data smoothing the 1-st derivative was computed using 2-d order polynomial estimated by making use of LS procedure. For this purpose another moving window was applied to smoothed data. Next, the resulting first derivative curve was used to determine estimates of chlorophyll content.

2.3.2 Estimation of chlorophyll content in vegetation

To determine estimates of chlorophyll content in vegetation the following approaches were applied: regression equations of different structure and neural nets.

3. Experimental details of the Lydar system

In the Lydar approach, a laser pulse is transmitted into the atmosphere and the backscattered radiation is detected as a function of time by an optical receiver (Fig.2).

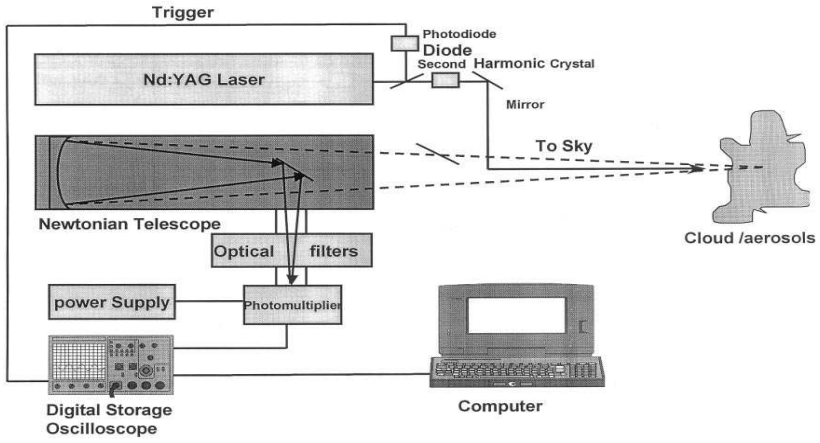

Fig. 2. Scheme of Lydar system

Laboratory tests of the spectrometer (Fig.3) was conducted. Our research has been concentrated on technologies such as optical spectroscopy measurements, laser induced fluorescence, pattern recognition and optimization methods are used in our device. Optical device is intended to be used for ground measurements or can be installed on small airplanes for remote sensing. Spectra training set samples are pre-processed and fed into the SVM algorithm followed by Leave-One-Out Cross-Validation (LOOCV).

Fig.3. Optical device

4. RESULTS

New algorithms for earlier detection of biochemical agents using network of sensors have been tested [4-7]. Our algorithms are based on synergetic approach, SVM-classification and SVM-regression, Sparse Proximal Support Vector Machines, twin Support Vector Machines, Independent Component Analysis (ICA), non-linear dynamic classification, and K-Means Clustering with SAM distance (Table 1, Fig. 4).

Table 1. Comparative analysis of new algorithms

Method	Classification Accuracy (LOOCV)	Flexibility in choosing # of features	Original feature space

SVM	97%	–NA-	Yes
PCA + SVM	97%	No	No
PCA + LDA	97%	No	No
RLR	94%	No	Yes
CFS + SVM	100%	No	Yes
FFS + SVM	100%	Yes	Yes

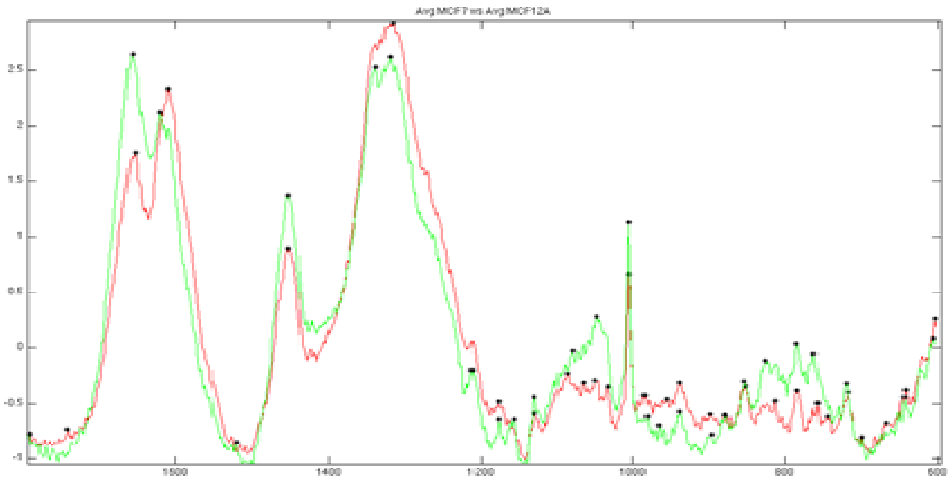

Fig. 4. Comparative analysis of new algorithms

The software for detection of chemical and biological agents in spectral mixture has been created (Fig.5). Dependence between the sensor equivalent spatial resolution and parameters of vegetation spectra mixing

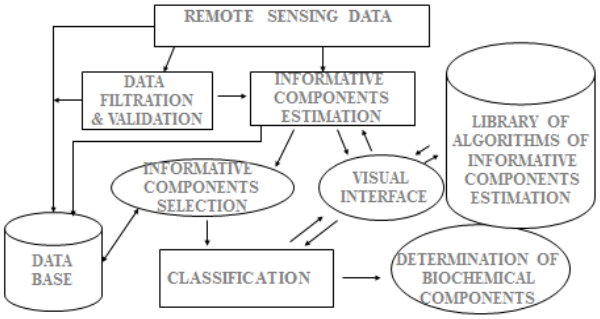

Fig. 5. Software for detection of chemical and biological agents

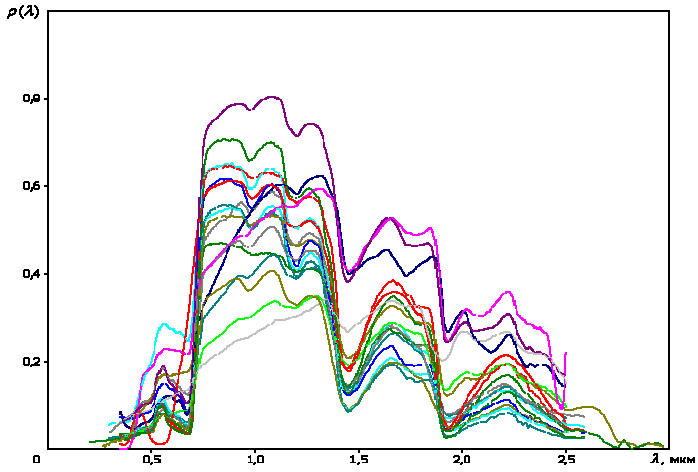

Fig.6. Vegetation spectra

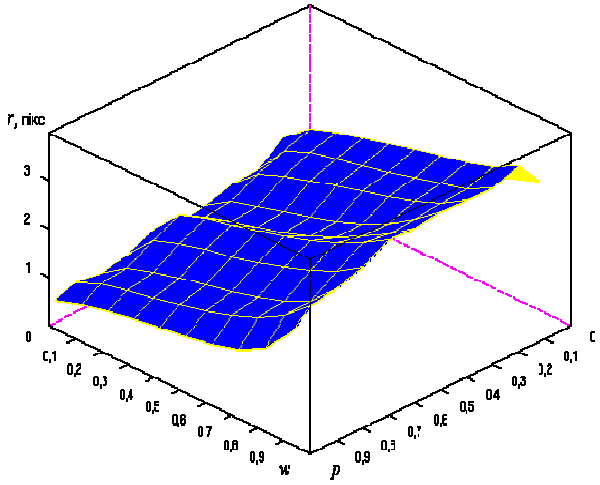

Fig.7. Relation between equivalent spatial resolution *r* and mixing ratio

It was established during the experiments that the noise level inherent to our measuring system was about 0.15% of maximum signal. After computing the 1-st derivative of spectral curves, the level of noise increased depending on the type of calculation procedure applied. Using of formulae with 5- and 9-piont moving windows and approximation by 2-nd degree polynomials [19] we could see about 4% and 1% noise at maximum signal level, respectively. Preliminary smoothing with 5, 7 and 9-point windows revealed their low effectiveness: the noise levels were 0.7; 0.5 and 0.5 of maximum signal for these procedures, respectively. As the 9-point window procedure for computing the 1-st derivative appeared acceptable with respect to distinguishing of maxima, it was selected for processing a large set of spectral curves collected with spectrometer. Fig. 1 shows several plots of the 1-st derivative of reflectance spectral curves recorded for winter wheat leaves with various chlorophyll concentration Visual estimation of the plots shows that the ripples are insignificant. As it has been reported earlier [15], there are two maxima in the 1-st derivative plots. Ratio of maximum intensity is appreciably dependent on chlorophyll content in leaves. Its value approaches 1 for chlorophyll content of about 5 mg/dm^2; it becomes greater than 1 for larger pigment concentrations and less than 1 for lower concentrations.

3.1 Calculation of chlorophyll content

Regression equation 1 (paired regression). The ratio of intensities in two maxima of the 1-st derivative plots of reflectance spectra was used to construct a regression equation with dependent variable in left-hand-side. The simple regression equation used is

$$z = a_0 + a_1 x\,, \qquad (1)$$

where z is estimate of chlorophyll; $x = I_l / I_s$; I_l and I_s are long-wavelength and short-wavelength maxima, respectively. The regression coefficients are easily estimated using LS procedure and known values of $\hat{z}_i$, and $x_i, i = 1,2,....$ Here $\hat{z}_i$ are estimates of chlorophyll concentration found by chemical analysis. The regression coefficients were estimated with standard deviations of their values less than 2%, and determination coefficient was $R^2 = 0.93$. Results for chlorophyll concentrations estimated by spectral and chemical methods for experiments in Fig. 1 are shown in Table 2.

To determine positions of maximum necessary for computing x in regression (1), genetic algorithm (GA) was applied, which allows to find the maxima even if the data curve is ill-conditioned, i.e. contains substantial ripples. GAs do not impose restrictions on the problem being solved such as continuity of goal functional, unimodality or smoothness of its surface. Also, the structure of optimizing GA does not depend on the solved problem context. Generally the estimation problem could be divided into two following sub-problems: implementation of genetic algorithm and interpreting of the results received.

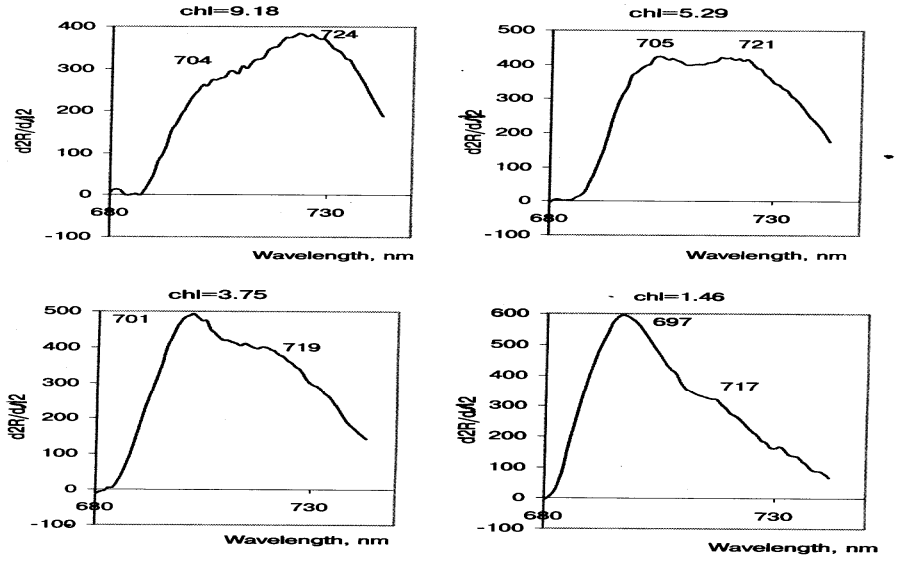

Fig. 8. Plots of first derivatives of reflectance spectra for winter wheat leaves versus wavelength.

Table 2. Maximum positions in 1-st derivative plots and chlorophyll values estimated by various methods.

# of experiment	Maximum positions, nm		Chlorophyll content, mg/dm^2		$\Delta_{\text{spectral-chemical}}$
	simple search	by genetic algorithm	spectral method	Chemical method	
1	704, 724	704, 725	9.35	9.18	0.17
2	705, 721	705, 720	5.33	5.29	0.04
3	701, 719	702, 719	3.80	3.75	0.05
4	697, 717	698, 717	1.54	1.46	0.08

The solution form that is generated by GA remains unchanged for any kind of a problem – this is a sequence of bits of predetermined length. The sequence length and type of its interpreting depends on the problem nature. An advantage of GA is in its low sensitivity to noises of various origins. Computational expenses for implementing GA are much lower than for, say, classic optimization procedures such as modified Newton-Raphson or FP (Fletcher-Powell) procedure. Comparison of results is demonstrated by Table 2. There is no substantial difference observed between maxima found by simple search procedure and GA. This is explained by effectiveness of the smoothing achieved due to application of selected procedure of 1-st derivative computing.

Regression equation 2 (multiple regression). The second regression equation was constructed in the form:

$$z(l) = a_0 + a_1 x_1(l) + a_2 x_2(l) + ... + a_m x_m(l), \quad (2)$$

where $x_i, i = 1,...,m$ are samples of smoothed reflectance spectrum taken with appropriate interval that could be varied depending on the number of experiments available; l is a number of experiment, with $l = 1,2,...,N$. Here

one experiment is referred to one spectrum measurement session and one estimate of chlorophyll content found by applying chemical test. Number of independent variables in equation (2) is much higher than in equation (1) what makes it logically suitable to description of the first derivative plots. The larger is the number of independent variables the higher should be degree of adequacy of equation (2), but there are computational limits to expanding the right-hand side, and shortage of an interval between samples of spectra may result in linearly dependent vectors of measurement matrix and, consequently, in the co-linearity problem. To find estimates of coefficients a_i, $i = 0,1,...,m$ we need at least $m+1$ measurements of reflected spectra and respective estimates of chlorophyll content. Application of equation (2) to estimation of chlorophyll content by our spectral curves, provides the results given in Table 3.

Table 3. Results of regression estimation of chlorophyll content.

# of independent variables	Equation 1 (mse of estimation)	Equation 2 (mse of estimation)
1	0.198	-
5	-	0.985
10	-	0.329
12	-	0.231

Note: mse = mean squared error of estimates.

According to table 2, estimates of chlorophyll, computed via equation (2), exhibit lower quality than the estimates computed via equation (1), though with the number of dependent variables equal to 12 mean squared error of estimation is comparable to the MSE computed for equation (1): 0.198 and 0.231, respectively. Further study of regression approach to estimation is required, this method seems to be promising with increasing the number of independent variables.

Conclusions

1. The main purpose of experimental research is to investigate the influence of small concentrations of chemical and biological agents on reflectance spectra with particular attention paid to the effects of laser induced optical response. Laser experiments are being carried out to give fundamental insight into laser induced fluorescence phenomena. Technologies such as an optical spectroscopy measurement, laser induced fluorescence and microscopy methods have been used in our experiments.
2. New algorithms for earlier detection of biochemical agents using network of sensors have been tested. Our algorithms are based on synergetic approach, SVM-classification and SVM-regression, Sparse Proximal Support Vector Machines, twin Support Vector Machine, Independent Component Analysis (ICA), non-linear dynamic classification, and K-Means Clustering with SAM distance.
3. The experiments performed showed that the ratio of maxima in the 1-st derivative plot from reflectance spectral curve is a parameter, which is highly correlated with chlorophyll content. At 0.15% noise of measuring system the 9-point procedure with approximation by 2-d order polynomials by Savitsky and Golay, is acceptable to distinguish two main maxima in the 1-st derivative plot. Their positions may be determined with high efficiency using a genetic algorithm. Paired linear regression equation permits to determine chlorophyll content with the error not exceeding 4%. Multiple regression equation (2) was less effective, but increasing the number of independent variables improves the estimates substantially. A neural net approach could be used for chlorophyll estimation as well, however it provided lower quality estimates than regression in our particular case. This result could be explained by the fact that we used conventional net structure. Perhaps, special types of nets could be adjusted for this specific problem to obtain better results, and further research in this direction is required. Another promising approach to approximation of the first derivative curves could be the use of orthogonal polynomials, and application of image recognition techniques for estimation of chlorophyll content in vegetation.
4. *T*he algorithms of rapid search and earlier detection of biological/chemical agents using an unmanned aerial vehicle (UAV) have been prepared. Detection-based techniques might be effective in collaboration with classification and clustering techniques for sensor networks (SNs).The development of anomaly detection techniques suitable for SNs is therefore regarded as an essential

research area, which will enable SNs to be much more reliable. In this report, a few of the key design principles and methods relating to the development of anomaly detection techniques in SNs are discussed. Informative criteria for anomaly detection from spectral data can be assisted by classification and clustering. The resulting order parameter set is further processed to obtain the minimum number of most informative parameters which represent the most discriminating pattern space for classification. We propose a general class of models for classification and clustering of spectral data in relational domains that capture probabilistic dependencies between related instances. Tests showed insufficient quality gluing interferometry cubes and large absorption in semitransparent mirror interferometer. The work of the new coating deposition and fabrication of a new interferometer.

1. R.M. Measures, Laser Remote Sensing: Fundamentals and Applications, Wiley, New York, 1984.
2. E.D. Hinkley, Laser Monitoring of the Atmosphere, Springer Verlag, Berlin, 1976.
3. K. Killinger, A. Mooradian, Optical and Laser Remote Sensing, Springer Verlag, Berlin, 1986.
4. V. Yatsenko, N. Kozyrovska et. al. Putative mechanism of bacteria effects on plant photosystem under stress // Biopolymers and Cell. - 2012. - Vol. 28. № 6.
5. Kharytonov M.M., Khlopova V.M., Stankevich S.A., Titarenko O.V. Remote and ground-based sensing of air polluted by nitrogen dioxide in the Dnepropetrovsk region (Ukraine) // Disposal of Dangerous Chemicals in Urban Areas and Mega Cities / Ed. by I. Barnes, K.J. Rudziński.– Dordrecht: Springer, 2013.– P.291-298.
6. V. Yatsenko. Classification, clustering and anomaly detection in sensor networks. Moscow.- December 2012.
7. V. Yatsenko. Active remote sensing of chemical and biological agents: computational models, toxity testing, and human health risk assessment.-2nd International conference on computational Biomedicine. January 24-26, 2013, University of Florida, USA.-P.7.