PRINCIPAL COMPONENT ANALYSIS OF LUNAR SOIL CHARACTERIZATION CONSORTIUM DATA.

D. Stankevich[1], Y. Shkuratov[1], C. Pieters[2], [1]Astron. Institute of Kharkov National Univ. 35 Sumskaya St., Kharkov, 61022, Ukraine. [2]Geological Sciences, Brown Univ., Providence, RI 02912. USA.

Summary: We analyzed spectra of lunar soils in the range 0.35–2.50 μm. Using the Lunar Soil Characterization Consortium (LSCC) database we show the principal component method (PCA) to be a useful technique to find a correlation between spectral properties of the samples and chemical composition. The number of spectral eigenfunctions important to approximate high-informative RELAB spectra is quite high, near 20.

Introduction: Interpretation of spectral properties of lunar materials relies on fundamental characteristics of diagnostic absorptions that are based on principals derived from mineral physics [1]. Such applications, however, require high precision visible to near-infrared spectra (0.3 – 2.6 μm) of high spectral resolution. Spectroscopic analysis has been highly successful for limited targets on the lunar nearside using instruments developed in the late 1970's [2]. It was shown that "red slope" is controlled by Fe^{2+} and Ti^{4+} ions in lunar mineral; 1 μm band is formed by pyroxenes and olivines; pyroxenes also are responsible for the 2 μm band.

The majority of high spatial resolution data currently available for the Moon comes from multi-spectral imagers that typically consist of a digital framing camera equipped with several filters. The Chandrayaan-1 spacecraft (Indian lunar mission) will be equipped with a scanning imager spectrometer providing high spatial as well as high spectral resolution. For interpretation of future data it is necessary to study relationships between the general spectral properties of lunar soils and their measured compositional characteristics [3]. To probe information in high spectral resolution spectra of the lunar samples obtained in the laboratory over the range 0.35–2.50 μm, a principal-component analysis was used to develop a statistical link between spectral properties and composition.

Initial data. We use the LSCC data on mare and highland soils [4]. The samples are representative of the compositional diversity. We used bulks (<45 μm) and 3 size-particle separates <10, 10-20, 20-45 μm for each sample. High spectral resolution bi-directional spectra were acquired in the RELAB at Brown Univ. Altogether we used 78 independent RELAB spectra, each includes 441 wavelengths. All spectra and composition data are available at: http://www.planetary.brown.edu/pds/LSCCsoil.html . For the analysis presented here, we focus on two compositional parameters that directly influence spectral characteristics of the lunar regolith, these are FeO and TiO_2 contents.

It is important to note that our analysis is strictly statistical by nature. The optical properties of exceptionally well characterized soils have been accurately measured, and we simply identify and use the most highly correlated relations.

PCA spectral analysis. We present the results of linear statistical analysis of the relation between lunar sample albedo with regolith parameters and composition. We used the PCA method. The main goal of this method is to find the system of functions, allowing to describe any given spectrum as a series with minimal number of terms. These terms are an eigenfunctions for the correlation matrix of spectra. Eigenvalues show the relative weight of the eigenfunctions. Figure 1 shows eigenvalues for the studied 78 spectra. One can see that these values fall down very quickly with number increase. This means that a relatively small number of the eigenfunctions enables to describe the spectra. The maximal number of the founded eigenfunctions is smaller than the number of spectra due to computational limitations (stability of the inverse matrix). In our case we determined 40 eigenfunctions.
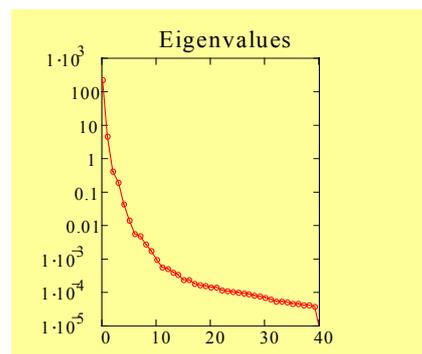

Fig. 1. Dependence of eigenvalues on their number.

Figures 2 and 3 present spectral curves of the first two eigenfunctions. The first component describes general features of the spectral set. This characterizes the general slope of spectra and the NIR bands centered near 1 and 2 μm. The second component demonstrates more subtle structural features: the spectral bend in visible range and the fine structure of the NIR bands.
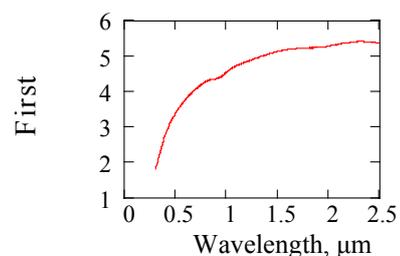

Fig. 2. Spectral dependence of the first eigenfunction.

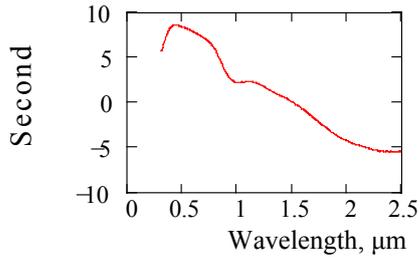The eigenfunctions of higher orders characterize more subtle structure of the spectra.

Fig. 3. Spectral dependence of the second eigenfunction.

An important question is how many eigen-functions should be taken into account for spectra approximation. This number is closely related to the noise of spectra. We developed a new procedure to estimate the number. The main idea is to make the PCA twice. In the first case we used the initial data set. In the second case we used the set with added random noise. Than we compare these two sets keeping the eigenfunctions that remain almost the same. The stable eigenfunctions are important for further analysis, and unstable ones are noise-dependent. In Fig. 4 we plot the correlation coefficients between eigenfunctions of the two sets. The axes are numbered with the eigenfunction numbers. The scale from blue to red colors corresponds to increasing correlation coefficient (from 0 to 1). As can be seen in Fig. 4 the first 18 eigenfunctions remain the same when a random noise with 0.01% amplitude is added. This noise is typical for RELAB data. We note also that increasing the noise amplitude by the factor 3 makes the number of stable eigenfunctions of about 10, i.e. the number is still rather high.
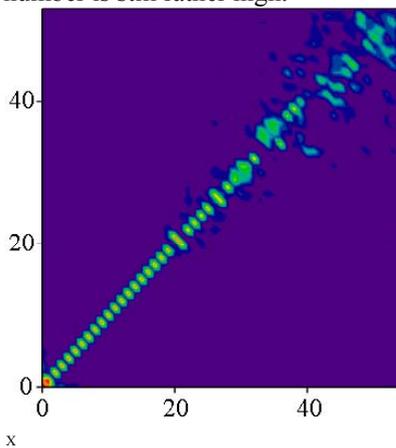

Fig. 4. Map of correlation coeffitient distribution at the comparison of initial and noisy spectra.

Thus any spectrum from the spectral suite can be expanded to series on eigenfunctions: $S_k(\lambda_i) = w_{1k}\varphi_1(\lambda_i)+w_{2k}\varphi_2(\lambda_i)+…$, where φ_i is the i-th eigenfunction, w_{lk} is the l-th loading coefficient for k-th spectrum. We use the loading coefficients as a characteristic of each sample. Then, we find the closest correlations between the coefficients and the

composition of the samples using a linear combination, e.g., for FeO the we have the following: $\log(FeO\%)=a_0+a_1w_1+a_2w_2+…$, where a_m is the coefficients providing such a correlation (m=18). Figures 5 and 6 demonstrate correlations between measured and predicted content of FeO and TiO_2, respectively, for the LSCC samples. The correlations are rather high; in both the cases it is 0.896.
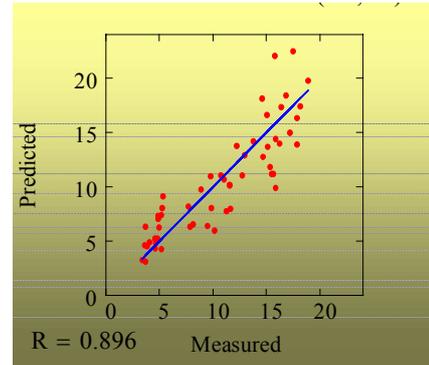

Fig. 5. Correlation between measured and predicted content of FeO for the LSCC samples.

Conclusion: Thus lunar soil spectra in the range $0.3 − 2.5$ μm can be presented with a limited set of spectral functions (on the order of 10) whose a linear combination provides a high accuracy fit to the original spectra The number of the significant eigenfunctions is higher then in the case of Clementine filter data that also were used to map chemical and mineral composition of lunar soils [5].
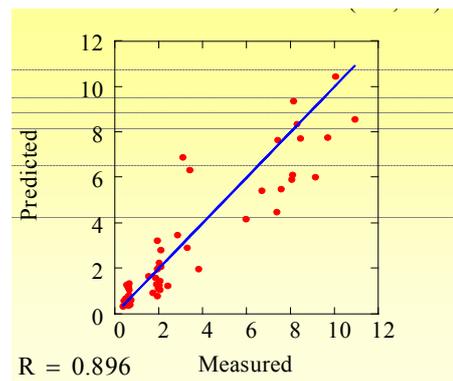

Fig. 5. Correlation between measured and predicted content of TiO_2 for the LSCC samples.

References: [1] Burns, R. 1993. *Mineralogical Application of Crystal Field Theory.* Cambridge Univ. Press, Cambridge, 551 p. [2] Pieters, C. 1993. in: *Elemental and Mineralogical Composition* / C. Pieters and P. Englert eds., 309-339, Cambridge Univ. Press. [3] Pieters, C. et al. 2002. *Icarus* 155, 285–298. [4] Taylor, L. et al. 2003. *LPSC* 34th, Abstract #1774. LPI Houston. [5] Pieters, C. et al. 2005. Icarus (submitted).