

**A NEW MODIFIED GAUSSIAN MODEL (MGM) USING THE CROSS-VALIDATION METHOD.** N. Tsuboi<sup>1</sup>, S. Sugita<sup>1</sup>, T. Hiroi<sup>2</sup>, K. Nagata<sup>1</sup> and M. Okada<sup>1</sup>, <sup>1</sup>Department of Complexity Science and Engineering, Graduate School of Frontier Sciences, the University of Tokyo, Kashiwa 277-8561, JAPAN ([tsuboi@astrobio.k.u-tokyo.ac.jp](mailto:tsuboi@astrobio.k.u-tokyo.ac.jp)), <sup>2</sup>Department of Geological Sciences, Brown University, Providence, RI 02912, U.S.A.

**Introduction:** Visible and near-infrared spectroscopy is a very powerful tool for measuring the mineral distribution of a solid planetary surface and has often been employed by spacecraft missions. The reflectance spectra of a planetary surface show multiple absorption bands due to silicate minerals, such as olivine and pyroxene. The modified Gaussian model (MGM) is a useful method for deconvolving the complex absorption bands into Gaussians and a background continuum [1]. It has been used successfully to characterize both mineralogical and elemental compositions of planetary surface materials [e.g., 1, 2, 3].

In order to conduct MGM analysis using a large quantity of spectral data, such as the Spectral Profiler (SP) onboard KAGUYA and the Moon Mineralogy Mapper (M<sup>3</sup>) onboard Chandrayaan-I, automation is highly desired. However, the automation of MGM analysis is not easy at present. When we analyze the reflectance spectrum of a sample of unknown mineralogical composition, it is very difficult to find how many Gaussians are needed to fit the spectrum. Thus, in the conventional MGM analyses, we had to adjust the number of Gaussians manually by looking at the quality of spectral fit. In this study, we propose a new MGM scheme for automatically determining the optimum number of Gaussians, using an approach called the cross-validation method.

**Cross-validation method:** The cross-validation is a standard tool for model selection and for avoiding overfitting [e.g., 4, 5]. Overfitting refers to fitting an observed dataset with too many model parameters. When a dataset is overfitted, the obtained optimum parameter values may be meaningless. For example, as the number of parameters (e.g., the number of Gaussians) is increased, the square error between the observed and the model becomes smaller. This trend continues until the number of model parameters is equal to the number of data points. In such an extreme case, the error between the observed and fit model will be zero. Nevertheless, such an apparent perfect fit does not indicate that the model parameters are correct but that the number of model parameters is too large. Thus, it is very difficult to know the optimum number of Gaussians based only on the square error. This is why the cross-validation method is very important. Here we use a reflectance spectral dataset to explain the cross-validation method briefly.

**Procedure:** The cross-validation method randomly splits a spectral dataset (total of  $m$  wavelength bands,

for instance) into two sets; a training set and a validation set. Then, a conventional MGM is used to fit the training set to find the optimum parameter values, such as the central wavelength of Gaussians. Then the validation set is used to test if the obtained optimum model parameter values are good by computing the root mean square (RMS) error between the obtained model and the validation dataset. After repeating this procedure using different numbers of Gaussians, we can find the optimum number of Gaussians, which gives the smallest error for the validation dataset. However, different ways of splitting the original dataset to training and validation sets might lead to different results. Thus, this whole process is usually repeated many times with different ways of splitting the dataset using different series of random numbers. Then, the average of the results is adopted to find the real optimum number of model parameters.

**Application to silicate spectra:** Pyroxene is one of the major components of planetary crusts and is observed on the moon, Mars, and asteroids. Also, olivine is estimated to be a dominant mineral in planetary mantles and is seen in very basic lava flows. Thus, olivine is a very important mineral in planetary remote sensing studies. However, the detection of olivine is not easy, particularly in mixtures of olivine and pyroxene. Pyroxene has two absorption bands near 1 and 2  $\mu\text{m}$ , whereas olivine has a broad complex absorption band near 1  $\mu\text{m}$  and no band in the 2  $\mu\text{m}$ . The 1  $\mu\text{m}$  absorption band position of olivine and pyroxene is very close, thus the absorption bands of those two minerals overlap significantly. In addition, the absorption coefficient of olivine is usually much smaller than that of pyroxene. Thus, the presence of olivine may not be recognized unless a very careful analysis is made [e.g., 6].

As a test of the cross-validation method, it is applied to the complex reflectance spectra of olivine and pyroxene mixtures for examining whether spectra of olivine (OLV) and pyroxene (PYX) mixtures can be deconvolved into their component bands. Also employed are orthopyroxene (OPX) and clinopyroxene (CPX) mixtures because olivine-pyroxene mixtures are difficult to distinguish from OPX-CPX mixtures. More specifically examined is whether there is a difference in the optimum number of Gaussians and the central wavelength between OLV-OPX mixtures and OPX-CPX mixtures.

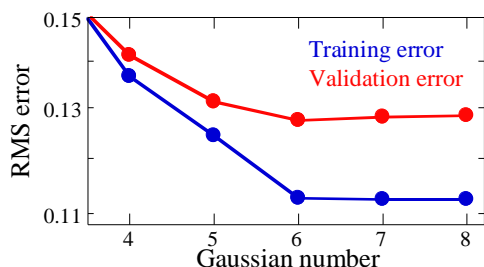


Figure 1. The result of cross-validation analysis applied to the spectrum of an olivine-orthopyroxene mixture (1:3 mixing ratio). The RMS error for training set (training error) and the RMS error for validation set (validation error) are shown in blue and red, respectively. For example, although eight Gaussians give a good fit to the training set, validation error is actually worse than the six-Gaussian case, which is the optimum case.

**Spectral data:** All the reflectance spectra are taken from the NASA/Keck RELAB database located at Brown University. The OLV was purchased from WARD'S Natural Science Establishment, Inc., and OPX from Chichi-jima, Japan for OLV-OPX mixtures (45-75  $\mu\text{m}$ ), and OPX was from Webster, North Carolina and CPX from Mauna Kea, Hawaii for OPX-CPX mixtures (<45  $\mu\text{m}$ ).

**Result and Discussion:** Figure 1 is the result of a cross-validation analysis applied to the spectrum of an OLV-OPX mixture (25:75 mixing ratio). The result indicates that the optimum number of Gaussians is 6 based on validation error. The same number 6 was obtained for the 50:50 mixture of OLV and OPX. Figure 2 shows the reflectance spectra of these mixtures fit with six Gaussians. It is noted that an absorption band around 1.05  $\mu\text{m}$  is identified in the MGM analysis. This band is observed in pure olivine samples but not in pure pyroxene samples. In both OLV-OPX mixtures (25:75 and 50:50 mixing ratios), this absorption band around 1.05  $\mu\text{m}$  can be detected. This result indicates that an olivine component can be detected in OLV-OPX mixtures if olivine content is over 25 wt%.

In the reflectance spectrum of the 50:50 mixture of OPX and CPX, the optimum number of Gaussians is found to be 5, which is different from OLV-OPX mixtures. The reflectance spectrum of the OPX-CPX mixture fit with five Gaussians is shown in Figure 3. It is noted that no band is identified around the 1.05  $\mu\text{m}$  region. This result strongly suggests that the presence/absence of an MGM-deconvolved absorption band around 1.05  $\mu\text{m}$  would be a useful indicator for olivine. Also, the accurate detection of additional absorption band for olivine strongly suggests that this cross-validation method can accurately find the optimum number of absorption bands.

**Conclusion:** In this study, MGM analyses of the reflectance spectra of OLV-OPX mixtures and OPX-CPX mixtures were conducted using the cross-validation method. This approach provides the most appropriate number of Gaussians for fitting a given spectra without subjective judgment. Thus, this method will be useful for developing an automatic MGM analysis algorithm. Furthermore, the result of the analysis suggests that the presence of olivine in OLV-OPX mixtures can be detected down to 25 wt%. Because this new analytical method is expected to be applicable to more complex mineral mixtures, it can become a very useful tool for analyzing real planetary spectral data.

**References:** [1] Sunshine J. M. *et al.* (1990) *JGR*, 95, 6855. [2] Sunshine J. M. and Pieters C. M. (1993) *JGR*, 98, 9075. [3] Sunshine J. M. and Pieters C. M. (1998) *JGR*, 103, 13675. [4] Stone (1974) *Biometrika*, 64, 29. [5] Faber and Rajko (2007) *Anal. Chim. Acta*, 595, 98. [6] Cloutis E. A. *et al.* (1986) *JGR*, 91, 11641.

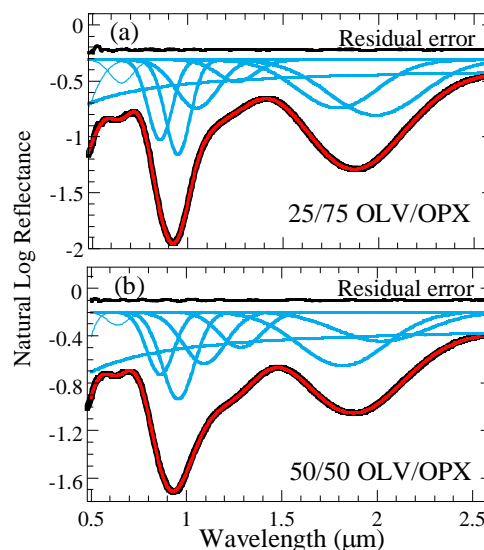


Figure 2. MGM fits to the olivine-opx mixtures spectra with six Gaussians, which is obtained the cross-validation analysis. (a) 25/75 OLV/OPX. (b) 50/50 OLV/OPX.

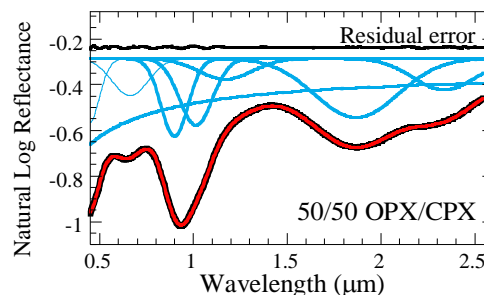


Figure 3. MGM fits to the spectrum of OPX-CPX mixtures (1:1 mixing ratio) with five Gaussians, which is obtained the cross-validation analysis.