A PRACTICAL ANALYTICAL APPROACH FOR PREDICTING SAND SPECTRAL SIGNATURES

Bradley W. Kimmel and Gladimir V. G. Baranoski

Natural Phenomena Simulation Group
School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
bwkimmel@curumin.cs.uwaterloo.ca

ABSTRACT

The authors have recently proposed a model, based on Monte Carlo methods, to simulate light interaction with sand [1]. In this paper, principal component analysis (PCA) and regression techniques are applied to yield a compact analytical representation of the spectral reflectance signatures produced by the model. This analytical formulation compares well with the original model and is appropriate for applications demanding interactive rates.

1. INTRODUCTION

Sand is an ubiquitous material found in diverse and remote environments, from Earth deserts and costal regions to extraterrestrial landscapes. Accordingly, airborne or satellite-based equipment is often used to measure the spectral signatures of these sandy areas in order to infer their intrinsic properties without the need for a field survey. Predictive simulations of light interaction with sand can provide a substantial contribution to these efforts. Notably, their application in conjunction with traditional measurement procedures can accelerate the hypothesis generation and validation cycles of remote sensing research frameworks aimed at this widespread type of soil.

In remote sensing studies involving soils, principal component analysis (PCA) and regression techniques are often applied to the classification and interpretation of satellite images by using spectral data associated with different soil targets [2, 3, 4, 5], to assist spectral band selection for multispectral remote sensor systems [6, 7], and to reduce the dimensionality of soil hyperspectral data sets [8, 9]. They have also been used to establish the spectral similarity among soil samples and soil types, which is related to major soil constituents, such as mineral (e.g., iron oxides and nitrate) and organic matter contents, and assist the prediction and analysis of these soil constituents [10, 11, 12].

This work was partially supported by the Canadian Foundation for Innovation (CFI Grant 33418) and the Natural Sciences and Engineering Research Council (NSERC Grant 238337).

Recently, the authors introduced a *sp*ectral *li*ght *t*ransport model for *s*and, called SPLITS [1]. The model employs Monte Carlo techniques to simulate light interaction with a sand sample, and may be used to predict the reflectance of the sample, given its physical and mineralogical characteristics. The goal of the research presented here is to to provide an analytical approximation to the SPLITS model. This proposed representation for the model makes use of PCA and regression techniques to yield a compact alternative to SPLITS.

2. BACKGROUND

The purpose of the SPLITS model is to simulate the spectral and spatial properties of light interaction with sand, given its physical and mineralogical characteristics [1]. For the present work, we concern ourselves with the spectral aspect.

Within the SPLITS modeling framework, a sand medium is represented by randomly oriented and randomly sized spheroidal particles distributed throughout the half-space below a plane boundary [1]. The parameters to the model include the mass concentrations of the iron oxides hematite, goethite, and magnetite: three major factors contributing to soil reflectance; as well as the amount of water present in the sand, expressed as the *degree of saturation*, the fraction of pore space occupied by water [13]. Additional parameters describe the geometrical arrangement of the mineral constituents. These include the fraction of particles by volume that consist of a single mineral (pure), of a mixture of the parent material and an iron oxide (mixed), and of the parent material coated by an iron oxide mixture (coated). In the case where coated particles are present, another parameter describes the thickness of the coating relative to the size of the particle.

The Monte Carlo simulation consists in tracing a path from an incident beam of light, through a simulated sand medium, until the ray is either scattered or absorbed. However, rather than storing the locations of individual sand particles, the particles are generated as required during the sim-

ulation. Light interaction with the particle is simulated, and the particle is subsequently discarded.

As is typical with Monte Carlo simulations, many trials (on the order of 10^8 paths in this case) are required to determine the overall light transport behavior of a given sample. Such simulations can take much time, which makes experimentation with the model difficult for applications that demand high interactive rates. For such applications, it is therefore be preferable to have an analytic alternative to this model.

3. METHODOLOGY

The analytical formulation for the proposed approximation to SPLITS is derived as follows. Physical and mineralogical data representing a set of 500 hypothetical sand samples (the *training* set) were generated randomly from within the domain of the SPLITS model. The model is then applied to yield the corresponding spectral reflectance curves for those samples. Specifically, the directional-hemispherical reflectance [14] was evaluated using an incident angle of $0°$. Other incident angles could be used as required for particular applications.

Principal component analysis (PCA) was performed on these reflectance curves (Section 3.1), yielding a small set of basis spectra along with the principal components associated with each sample. Regression analysis (Section 3.2) was then performed to obtain a predictor mapping the physical and mineralogical data characterizing a given sand sample to the corresponding principal components. The results of these analyses are then combined to yield an analytical formulation for the spectral reflectance signatures simulated by the SPLITS model.

3.1. Principal Component Analysis

Principal component analysis [15] was performed on these reflectance curves. The spectral curves from the training set were aggregated into an $\ell \times n$ matrix R, where ℓ is the size of the training set and n represents the number of wavelengths sampled per set. The rows $\mathbf{r}_i$ of R correspond to the samples in the training set. The analysis consists in performing an eigen-decomposition of R^tR. This yields $R^tR = U\Lambda U^t$, where Λ is the diagonal matrix with the descending eigenvalues λ_i of R^tR along the main diagonal, and U is an orthogonal matrix having the corresponding eigenvectors $\mathbf{u}_i$ for its columns.

The eigenvectors $\mathbf{u}_i$ denote the directions of decreasing variance in the training set, with $\mathbf{u}_1$ being the direction of maximum variance [15]. The eigenvalues λ_i indicate the variance in the direction of $\mathbf{u}_i$. This property allows us to ignore variance beyond a given threshold by projecting the data into the space spanned by the first k eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$, for some $k < n$. That is, we reexpress the data as

$$\tilde{\mathbf{r}}_i = U_k^t \mathbf{r}_i, \tag{1}$$

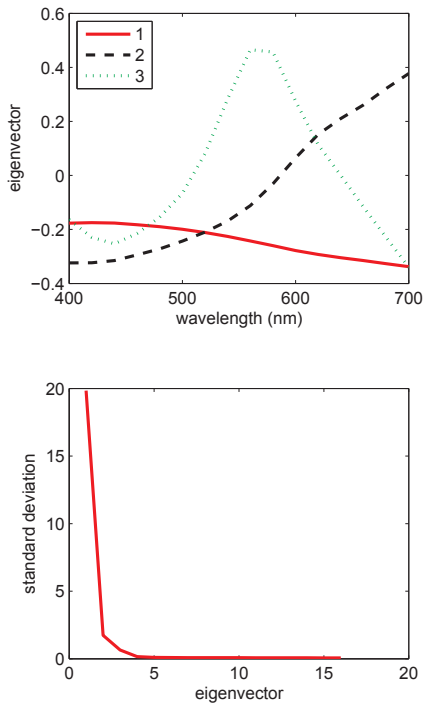

Fig. 1. The results dervied from principal component analysis (PCA) of the reflectance spectra simulated by the SPLITS model. *Top*: The first three basis spectra. *Bottom*: The standard deviation ($\sqrt{\lambda_i}$) of the reflectance data in the direction of the eigenvectors $\mathbf{u}_i$ derived from the analysis, indicating that three eigenvectors are sufficient to capture all but an negligible portion of the variance in the simulated spectra.

where U_k is the matrix formed by dropping all but the first k columns of U [15].

This allows us to represent the spectral curves generated by the SPLITS model using a small set of basis spectra (three were found to be sufficient to represent over 99.9% of the variance in the spectral reflectance data from the training set) along with the principal components $\tilde{\mathbf{r}}_i$ associated with each sample (see Figure 1).

3.2. Regression

To relate the principal components $\tilde{\mathbf{r}}_i$ of a training datum to its corresponding physical characteristics, a regression analysis [15] was performed. The physical and mineralogical characterization data $\mathbf{x}_i$ describing the sample were first mapped to an m-dimensional feature space via a function $\phi(\mathbf{x})$ in the components of $\mathbf{x}$. Linear regression was then performed in this feature space to yield an $m \times k$ matrix W. This gives a nonlinear predictor

$$\tilde{\mathbf{r}} \approx W^t \phi(\mathbf{x}) \tag{2}$$

	Training Set	Test Set
Mean	0.0051	0.0068
95^{th} Percentile	0.0120	0.0179
Maximum	0.0474	0.0551

Table 1. Statistics characterizing the root-mean-squared errors (RMSE) between spectral reflectance signatures simulated by the SPLITS model and those predicted by the proposed analytical formulation.

for the principal components corresponding to a given set of characterization data.

3.3. Summary

By combining Equations (1) and (2), we obtain a predictor for the corresponding spectral curve,

$$\mathbf{r} \approx U_k W^t \phi(\mathbf{x}).$$

For a set of physical data corresponding to a given sand sample, therefore, the analytical approximation to the SPLITS model is evaluated by applying the predictor function obtained via regression to that data, yielding principal components. These principal components are multiplied by the corresponding basis spectra and summed to obtain the reflectance spectrum of the sand sample in question.

4. RESULTS

To evaluate the accuracy of this approximation for a given sample, the spectral reflectance curves produced by the SPLITS model are compared to those given by the proposed analytical formulation. This comparison was performed for all of the samples in the training set, yielding a mean root-mean-squared (RMS) training error of 0.0051.

Also, characterization data corresponding to an independent set of 500 hypothetical sand samples (the *test* set) was generated randomly within the domain of the SPLITS model. The above evaluation was also performed using this test set, yielding a mean RMS error of 0.0068. In comparison with results obtained by applying similar techniques to reconstruct the spectral signatures of other natural materials [16], these relatively low error values indicate that the analytical approximation provides a good spectral reconstruction of the curves generated from the original model. This aspect is further illustrated by a summary of the RMS errors presented in Table 1.

The RMS errors for the individual samples in the test set are provided in Figure 2, shown against each of the model parameters that were varied. Note the lack of pattern in the errors as a function of any model parameter. Plots demonstrating typical comparisons between the reflectance curves simulated by SPLITS with those provided by the proposed formulation are shown in Figure 3. These indicate that the proposed formulation accurately represents the reflectance curves simulated by the SPLITS model.

5. CONCLUSION

We have derived a compact analytical formulation for the spectral directional-hemispherical reflectance signatures of sand samples. It is based on principal component and regression analysis of SPLITS: a light transport model recently introduced by the authors [1]. The proposed formulation compares well with the original model, and is efficient enough to use when high interactive rates are required.

6. REFERENCES

[1] B.W. Kimmel and G.V.G. Baranoski, "A novel approach for simulating light interaction with particulate materials: application to the modeling of sand spectral properties," *Optics Express*, vol. 15, no. 15, pp. 9755–9777, 2007.

[2] F. Chen, D.E. Kissel, L.T. West, and W. Adkins, "Field scale mapping of surface soil clay concentration," *Precision Agriculture*, vol. 7, pp. 7–26, 2004.

[3] G.A. Fox and R. Metla, "Soil property analysis using principal component analysis, soil line and regression models," *Soil Science Society of America Journal*, vol. 69, pp. 1782–1788, 2005.

[4] Y. Ge, J.A. Thomasson, R. Sui, and J. Wooten, "A comparison of regression and regression-kriging for soil characterization using remote sensing imagery," in *Annual Meeting of the American Society of Agricultural and Biological Engineers*, 2007.

[5] S. Murchie, L. Kirkland, S. Erard, J. Mustard, and M. Robinson, "Near-infrared spectral variations of Martian surface materials from ISM imaging spectrometer data," *Icarus*, vol. 147, pp. 444–471, 2000.

[6] A. Palacios-Orueta and S.L. Ustin, "Multivariate statistical classification of soil spectra," *Remote Sensing of Environment*, vol. 57, pp. 108–118, 1996.

[7] A. Palacios-Orueta and S.L. Ustin, "Remove sensing of soil properties in the Santa Monica mountains I. spectral analysis," *Remote Sensing of Environment*, vol. 65, pp. 170–183, 1998.

[8] M.O. Smith, P.E. Johnson, and J.B. Adams, "Quantitative determination of mineral types and abundances from reflectance spectra using principal component analysis," *Journal of Geophysical Research*, vol. 90, pp. C797–C804, 1985.

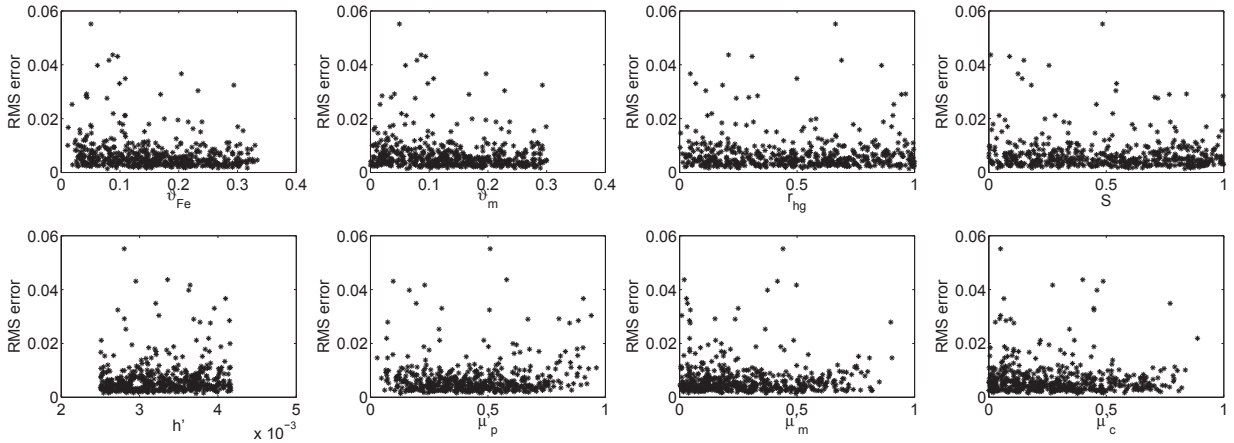

Fig. 2. The root-mean-square (RMS) error (across all wavelengths sampled) is shown against each of the model parameters varied. *Upper row (left to right)*: Total iron oxide concentration by mass (ϑ_{Fe}), concentration of magnetite by mass (ϑ_m), relative concentration of hematite vs. goethite by mass (r_{hg}), degree of saturation ($\mathcal{S}$). *Lower row (left to right)*: Relative thickness of grain coating (h'), fraction of pure particles by mass (μ'_p), fraction of mixed particles by mass (μ'_m), fraction of coated particles by mass (μ'_c).

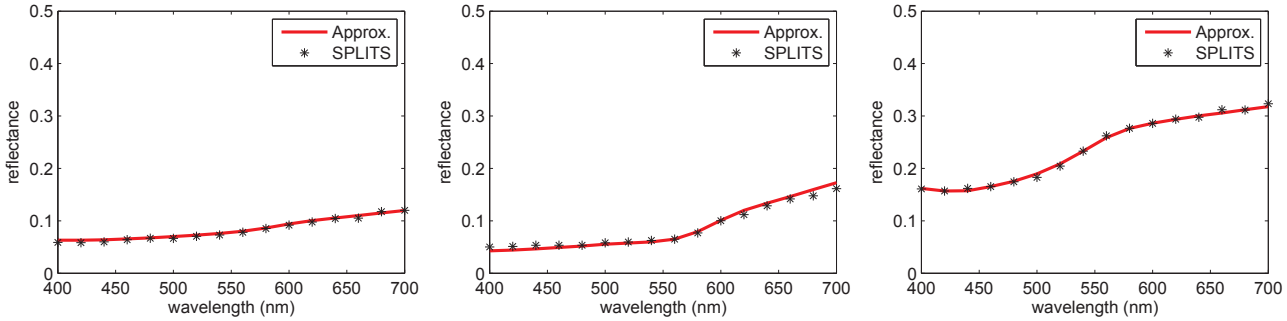

Fig. 3. Comparisons between the directional-hemispherical reflectance [14] as simulated by the SPLITS model and the approximation presented in this work for three hypothetical sand samples. The asterisks represent the reflectance computed using the SPLITS model. The solid line indicates the approximation using principal component analysis (PCA) and regression analysis.

[9] J. Iqbal, P. Owens, and J.L. Willers, "Remote sensing of surface carbon and water contents using bare soil imagery," in *14th World Congress of Soil Science*, Philadelphia, Pennsylvania, USA, 2006.

[10] L.S. Galvão and I. Vitorello, "Variability of laboratory measured soil lines of soil from Southeastern Brazil," *Remote Sensing of Environment*, vol. 63, pp. 166–181, 1998.

[11] R. Linker, I. Shmulevich, A. Kenny, and A. Shavi, "Soil identification and chemometrics for direct determination of nitrate in soils using FTIR-ATR mid-infrared spectroscopy," *Chemosphere*, vol. 61, pp. 652–658, 2005.

[12] C. van Waes, I. Mestdagh, P. Lootens, and L. Carlier, "Possibilities of near infrared reflectance spectroscopy for the prediction of organic carbon concentrations in grassland soils," *Journal of Agricultural Science*, vol. 143, pp. 487–492, 2005.

[13] D.B. Lobell and G.P. Asner, "Moisture effects on soil reflectance," *Soil Science Society of America Journal*, vol. 66, no. 3, pp. 722–727, 2002.

[14] F.E. Nicodemus, J.C. Richmond, J.J. Hsia, I.W. Ginsberg, and T. Limperis, *Geometrical Considerations and Nomenclature for Reflectance*, National Bureau of Standards, United States Department of Commerce, 1977.

[15] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.

[16] I.E. Bell and G.V.G. Baranoski, "Reducing the dimensionality of plant spectral databases," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 570–576, 2004.