A Compact Framework to Efficiently Represent the Reflectance of Sand Samples

Bradley W. Kimmel, *Student Member, IEEE*, and Gladimir V. G. Baranoski, *Senior Member, IEEE*

***Abstract*—The authors have recently proposed a model, based on Monte Carlo methods, to simulate light interaction with sand. In this paper, principal component analysis 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. Examples are provided comparing the original model to the proposed formulation for three hypothetical sand samples. The effect of water content on reflectance is demonstrated for these samples. Additionally, examples are provided comparing the original and proposed models for three sand samples from the U.S. Army Topographic Engineering Center spectral database.**

***Index Terms*—Nonlinear regression, principal component analysis (PCA), reflectance, sand.**

I. Introduction

SAND is an ubiquitous material found in diverse and remote environments, from Earth deserts and coastal 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 classify satellite images by using spectral data associated with different soil targets [2]–[5], to assist spectral-band selection for multispectral remote sensor systems [6], [7], and to reduce the dimensionality of hyperspectral data sets [8], [9], including those associated with soils [10], [11]. 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 [12]–[14].

Manuscript received September 28, 2008; revised April 18, 2009. First published September 18, 2009; current version published October 28, 2009. This work was supported in part by the Canadian Foundation for Innovation (CFI) under CFI Grant 33418 and in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) under NSERC Grant 238337.

The authors are with the Natural Phenomena Simulation Group, David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: bwkimmel@curumin.cs.uwaterloo.ca; gvgbaran@curumin.cs.uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2009.2028610

Fig. 1. Photograph of sand showing a close-up of sand particles.

Recently, the authors introduced a SPectral LIght Transport model for Sand, 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 in this paper, which is an extended and upgraded version of a conference presentation [15], is 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.

II. Background

Sand is a particular type of soil composed of particles immersed in a medium of air and water (the *pore space*). According to the system developed by the U.S. Department of Agriculture [16], at least 85% of these particles are between 0.05 and 2 mm, and the remainder have a smaller size (Fig. 1).

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 this paper, we concern ourselves with the spectral aspect.

Within the SPLITS modeling framework, a sand medium is represented by randomly oriented and 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, which is the fraction of pore space occupied by water [17]. 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

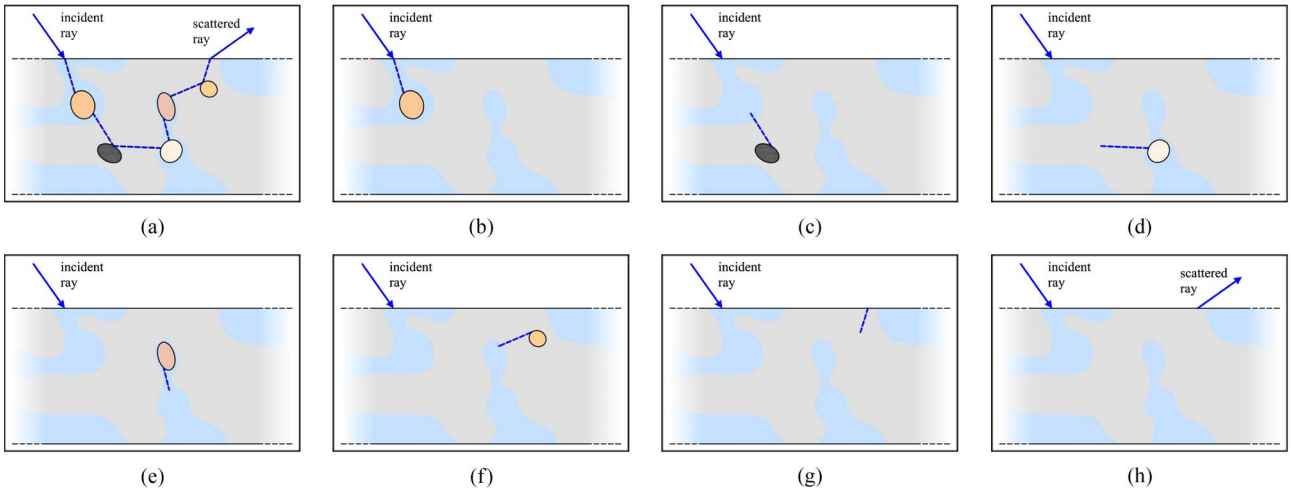

Fig. 2. Comparison between a traditional ray-tracing approach and that employed by the SPLITS model [1] is depicted. (a) For a traditional ray-tracing approach, the particles are explicitly stored. (b) For the SPLITS model, an incident ray first interacts with the extended boundary. If the path enters the medium, a new particle is generated randomly. Light interaction with that particle is simulated. (c) If the path is not absorbed in the particle, it eventually exits the particle. The particle is discarded and a new one is generated. (d)–(f) This process is repeated. (g) If the intersection with the generated particle lies above the extended boundary, the ray instead interacts with the boundary. (h) Ray may be internally reflected, or it may be scattered out of the medium. In the latter case, the process terminates.

(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 simulation (Fig. 2). 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 preferable to have an analytic alternative to this model.

III. Methodology

The analytical formulation for the proposed approximation to SPLITS is derived as follows. Physical and mineralogical data representing a set of 3000 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. For this framework, the reflectance curves were sampled at 16 wavelengths, regularly spaced between 400 and 700 nm. Specifically, the directional hemispherical reflectance [18] was evaluated using an incident angle of 0°. Other incident angles could be used as required for particular applications.

PCA was performed on these reflectance curves (Section III-A), yielding a small set of basis spectra along with the principal components associated with each sample. Regression analysis (Section III-B) was then performed to obtain a predictor mapping the physical and mineralogical data characterizing a given sand sample to the corresponding

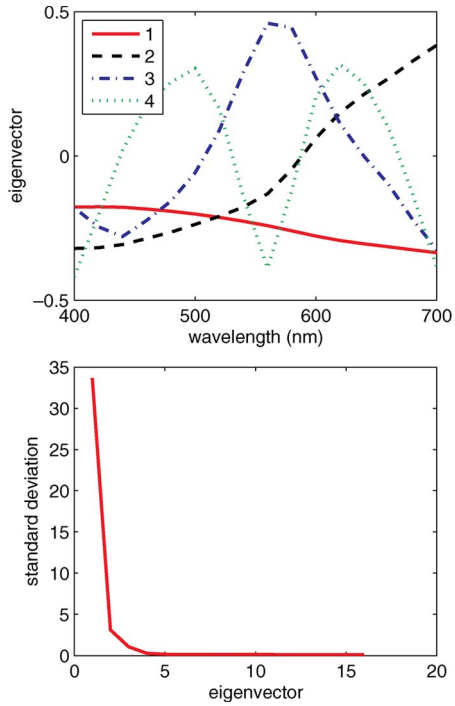

Fig. 3. Results derived from PCA of the reflectance spectra simulated by the SPLITS model. (Top) First four basis spectra. (Bottom) Standard deviation $(\sqrt{\lambda_i})$ of the reflectance data in the direction of the eigenvectors $\mathbf{u}_i$ derived from the analysis, indicating that four eigenvectors are sufficient to capture all but a negligible portion of the variance in the simulated spectra.

TABLE I
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

	Training Set	Test Set
Mean	0.0067	0.0069
95^{th} Percentile	0.0153	0.0158
Maximum	0.0785	0.0478

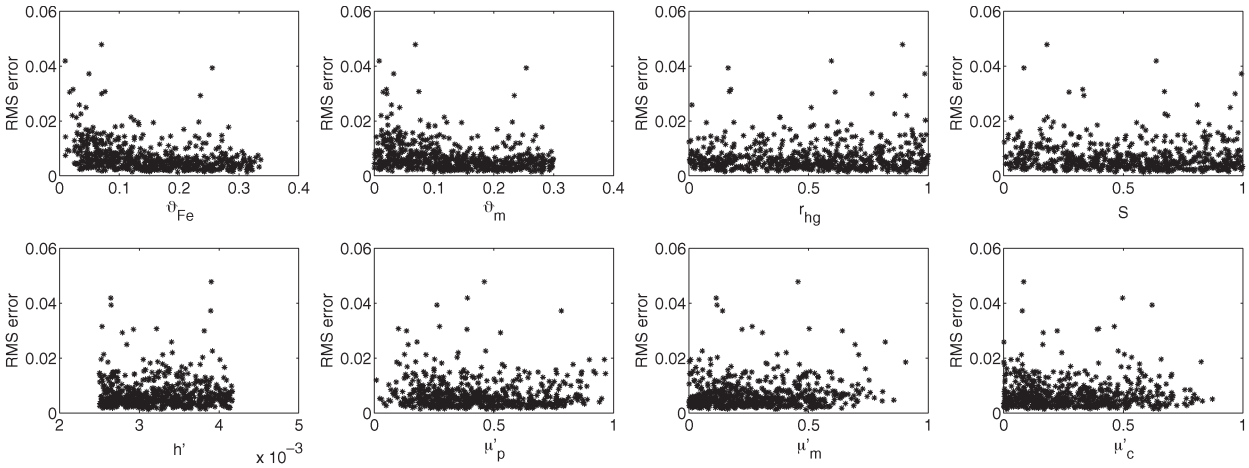

Fig. 4. RMS error (across all wavelengths sampled) is shown against each of the model parameters varied. (Upper row) (left to right) Total iron oxide concentration by mass (ϑ_{Fe}), concentration of magnetite by mass (ϑ_m), relative concentration of hematite versus goethite by mass (r_{hg}), and degree 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).

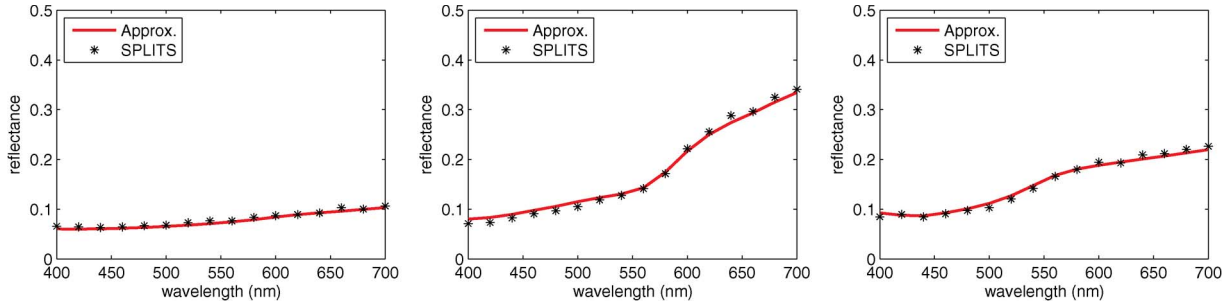

Fig. 5. Comparisons between the directional hemispherical reflectance [18] as simulated by the SPLITS model and the approximation presented in this paper for three hypothetical sand samples. The asterisks represent the reflectance computed using the SPLITS model. The solid line indicates the approximation using PCA and regression analysis.

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.

A. PCA

PCA [19] 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 $n = 16$ is the number of wavelengths sampled per set. The rows $\mathbf{r}_i$ of R correspond to the samples in the training set. To keep the reconstruction process simple, the sample mean was not subtracted. The analysis consists in performing an eigendecomposition 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 [19]. 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

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

where U_k is the matrix formed by dropping all but the first k columns of U [19].

This allows us to represent the spectral curves generated by the SPLITS model using a small set of basis spectra (four 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 Fig. 3).

B. Regression

To relate the principal components $\tilde{r}_i$ of a training datum to its corresponding physical characteristics, a regression analysis [19] 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 $\mathbf{x}$. For this framework, ϕ maps the characterization data $\mathbf{x}$ to the cubic monomial basis in the components $\mathbf{x}_i$. That is, given the components $\mathbf{x}_i$, $i = 1, \ldots, d$, of $\mathbf{x}$, $\phi(\mathbf{x})$ has components consisting of the values $\mathbf{x}_{i_0}\mathbf{x}_{i_1}\mathbf{x}_{i_2}$ for every triple (i_0, i_1, i_2) satisfying $0 \le i_0 \le i_1 \le i_2 \le d$, where $\mathbf{x}_0 = 1$. Linear regression was then performed in this feature space to yield an $m \times k$ matrix W. This gives a nonlinear predictor

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

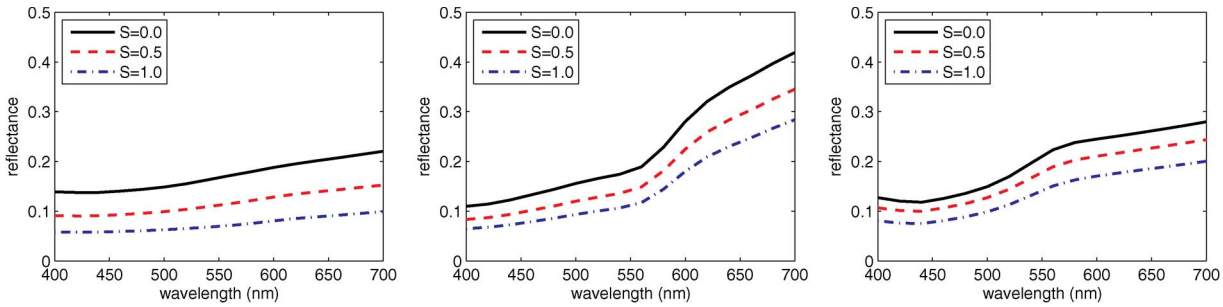

Fig. 6. Effect on the directional hemispherical reflectance [18] of the three hypothetical sand samples is depicted as the degree of saturation is adjusted. The curves were obtained using the analytical approach presented in this paper.

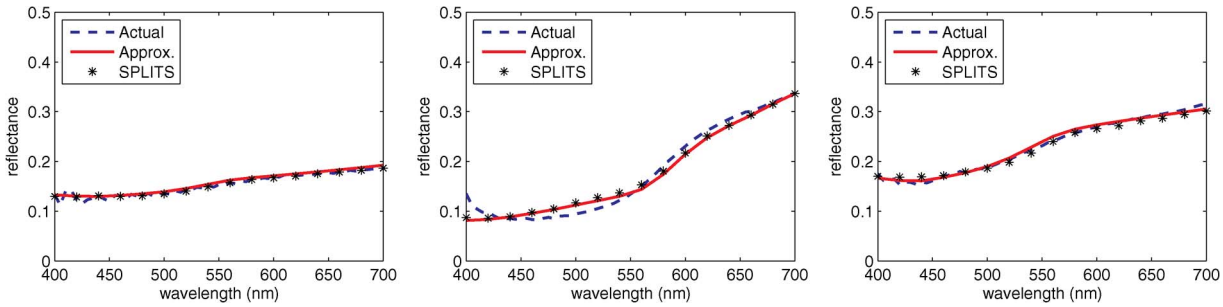

Fig. 7. Comparisons are shown depicting the reflectance of three sand samples from the TEC spectral database, the corresponding reflectance predicted by the SPLITS model, and the results of the proposed analytical approximation of the SPLITS model. (Left) Magnetite-rich beach sand from central Peru (TEC #10039240). (Middle) Dune sand from Saudi Arabia (TEC #13j9823). (Right) Sample from a dike outcrop in San Bernardino county, California (TEC #13au9815).

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

C. Summary

By combining (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.

IV. 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 0.0067.

In addition, characterization data corresponding to an independent set of 600 hypothetical sand samples (the test set) were generated randomly within the domain of the SPLITS model. The earlier evaluation was also performed using this test set, yielding a mean rms error of 0.0069. In comparison with results obtained by applying similar techniques to reconstruct the spectral signatures of other natural materials [20], 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 I.

The rms errors for the individual samples in the test set are shown in Fig. 4, 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 Fig. 5. These indicate that the proposed formulation accurately represents the reflectance curves simulated by the SPLITS model.

Fig. 6 shows the variation in the directional hemispherical reflectance [18] as the water content is varied. The water content, expressed as the degree of saturation—the fraction of the pore space occupied by water, is varied from zero to one. This behavior is consistent with the original model [1] and with the qualitative reflectance properties of sand reported in the literature [21].

SPLITS, as originally formulated, was evaluated [1] using several sand samples from the U.S. Army Topographic Engineering Center (TEC) spectral database [22]. Fig. 7 shows this comparison for three samples, along with the corresponding reflectance curve predicted by the analytical formulation proposed in this paper.

V. Conclusion

We have derived a compact analytical formulation for the spectral directional hemispherical reflectance signatures of sand samples. It is based on PCA and regression analysis of SPLITS:

a light-transport model recently introduced by the authors [1]. The proposed formulation compares well with the original model, as indicated by several comparisons using hypothetical, as well as real, sand samples. The analytical formulation is efficient enough to use when high interactive rates are required.

Although not specifically addressed in this paper, the potential exists for efficient land-surface models to be used for the retrieval of soil physical properties, given remotely sensed reflectance data. This possibility is suggested by Pauwels *et al.* [23] and corresponds to inversion of the framework described in this paper. Although several numerical techniques are suitable for model inversion, this can be a highly nontrivial process [24].

In future work, we intend to extend the techniques described in this paper to support other types of soil, such as those that include organic compounds. Another important line of research would involve extending the framework described in this paper to include other portions of the electromagnetic spectrum. This would allow the framework to capture important absorption features of water and other mineral constituents. Additionally, although the framework proposed describes a formal relationship between the model parameters and the basis spectra, attributing a physical meaning to these eigenvectors would allow one to estimate more readily the principal components associated with a given sand sample. This is also an important avenue for future research. We also plan to expand these techniques to support the prediction of the bidirectional reflectance distribution function [18].

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments and criticisms.

REFERENCES

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

[2] F. Chen, D. Kissel, L. West, and W. Adkins, "Field scale mapping of surface soil clay concentration," *Precis. Agric.*, vol. 5, no. 1, pp. 7–26, Feb. 2004.

[3] G. Fox and R. Metla, "Soil property analysis using principal component analysis, soil line, and regression models," *Soil Sci. Soc. Amer. J.*, vol. 69, no. 6, pp. 1782–1788, 2005.

[4] Y. Ge, J. Thomasson, R. Sui, and J. Wooten, "A comparison of regression and regression–kriging for soil characterization using remote sensing imagery," in *Proc. Annu. Meeting Amer. Soc. Agric. Biol. Eng.*, 2007. Paper 071001.

[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, no. 2, pp. 444–471, Oct. 2000.

[6] A. Palacios-Orueta and S. Ustin, "Multivariate statistical classification of soil spectra," *Remote Sens. Environ.*, vol. 57, no. 2, pp. 108–118, Aug. 1996.

[7] A. Palacios-Orueta and S. Ustin, "Remote sensing of soil properties in the Santa Monica mountains. I. Spectral analysis," *Remote Sens. Environ.*, vol. 65, no. 2, pp. 170–183, Aug. 1998.

[8] A. Mukherjee, M. Velez-Reyes, and B. Roysam, "Interest points for hyperspectral image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 748–760, Mar. 2009.

[9] N. Renard and S. Bourennane, "Dimensionality reduction based on tensor modeling for classification methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1123–1131, Apr. 2009.

[10] M. Smith, P. Johnson, and J. Adams, "Quantitative determination of mineral types and abundances from reflectance spectra using principal component analysis," *J. Geophys. Res.*, vol. 90, pp. C797–C804, Feb. 1985.

[11] J. Iqbal, P. Owens, and J. Willers, "Remote sensing of surface carbon and water contents using bare soil imagery," presented at the 14th World Congr. Soil Sci., Philadelphia, PA, 2006.

[12] L. Galvão and I. Vitorello, "Variability of laboratory measured soil lines of soil from Southeastern Brazil," *Remote Sens. Environ.*, vol. 63, no. 2, pp. 166–181, Feb. 1998.

[13] 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, no. 5, pp. 652–658, Nov. 2005.

[14] 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," *J. Agric. Sci.*, vol. 143, pp. 487–492, 2005.

[15] B. Kimmel and G. Baranoski, "A practical analytical approach for predicting sand spectral signatures," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2008, vol. 3, pp. 668–671.

[16] Soil Science Division Staff, *Soil Survey Manual. Soil Conservation Service*, 1993. U.S. Department of Agriculture Handbook 18.

[17] D. Lobell and G. Asner, "Moisture effects on soil reflectance," *Soil Sci. Soc. Amer. J.*, vol. 66, no. 3, pp. 722–727, 2002.

[18] F. Nicodemus, J. Richmond, J. Hsia, I. Ginsberg, and T. Limperis, *Geometrical Considerations and Nomenclature for Reflectance*. Washington, DC: Nat. Bureau Standards, U.S. Dept. Commerce, 1977.

[19] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[20] I. Bell and G. Baranoski, "Reducing the dimensionality of plant spectral databases," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 570–576, Mar. 2004.

[21] M. Baumgardner, L. Silva, L. Biehl, and E. Stoner, "Reflectance properties of soils," *Adv. Agron.*, vol. 38, pp. 1–43, 1985.

[22] J. Rinker, C. Breed, J. McCauley, and P. Corl, "Remote sensing field guide—Desert," U.S. Army Topographic Eng. Center, Fort Belvoir, VA, Tech. Rep. ETL-0588, Sep. 1991.

[23] V. Pauwels, A. Balenzano, G. Satalino, H. Skriver, N. Verhoest, and F. Mattia, "Optimization of soil hydraulic model parameters using synthetic aperture radar data: An integrated multidisciplinary approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 2, pp. 455–467, Feb. 2009.

[24] B. Combal, F. Baret, M. Weiss, A. Trubuil, D. Macé, A. Pragnère, R. Myneni, Y. Knyazikhin, and L. Wang, "Retrieval of canopy biophysical variables from bidirectional reflectance—Using prior information to solve the ill-posed inverse problem," *Remote Sens. Environ.*, vol. 84, no. 1, pp. 1–15, Jan. 2002.

Bradley W. Kimmel (S'08) received the B.Math. degree in pure mathematics and computer science and the M.Math. degree in computer science from the University of Waterloo, Waterloo, ON, Canada, where he is currently working toward the Ph.D. degree in computer science in the Natural Phenomena Simulation Group, David R. Cheriton School of Computer Science.

His research interests include simulating light interaction with natural environments.

Gladimir V. G. Baranoski (SM'07) received the Ph.D. degree in computer science from the University of Calgary, Calgary, AB, Canada, in 1998.

He is currently an Associate Professor with the David R. Cheriton School of Computer Science and the Leader of the Natural Phenomena Simulation Group, University of Waterloo, Waterloo, ON, Canada. The results of his multidisciplinary research on the modeling of light interactions with natural materials (e.g., plants, soil, and human tissues) have been made available to the scientific community through the publication of articles in remote sensing, applied optics, and computer graphics journals. He has also organized and presented several tutorials and courses in international conferences (e.g., Annual Conference of the European Association for Computer Graphics—EUROGRAPHICS, International Conference and Exhibition on Computer Graphics and Interactive Techniques, and International Geoscience and Remote Sensing Symposium) on topics related to his research contributions.