

Derivation of a soil albedo dataset from MODIS using principal component analysis: Northern Africa and the Arabian Peninsula

Liming Zhou, Robert E. Dickinson, and Yuhong Tian

School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA

Received 22 August 2005; revised 23 September 2005; accepted 5 October 2005; published 8 November 2005.

[1] This paper analyzes MODIS 1 km albedo kernels of 7 spectral bands over Northern Africa and the Arabian Peninsula and through these kernels develops a new high quality dataset that provides a simple statistical method to scale up spectral and broadband albedos from pixel to arbitrary coarse resolution grid square for use in climate models. This dataset significantly improves characterization of spatial and spectral variability and solar zenith angle dependence of soil albedo relative to simple grid means from MODIS data. The statistical method based on minimum noise fraction rotation transforms is able to not only successfully capture most of the MODIS albedo variance but also to extract large-scale spatial structures of albedo patterns from the original MODIS data while improving the data quality and reducing the number of parameters needed to represent the data. **Citation:** Zhou, L., R. E. Dickinson, and Y. Tian (2005), Derivation of a soil albedo dataset from MODIS using principal component analysis: Northern Africa and the Arabian Peninsula, *Geophys. Res. Lett.*, 32, L21407, doi:10.1029/2005GL024448.

1. Introduction

[2] Albedo determines the amount of solar radiation absorbed by the Earth's surface and thus surface fluxes and climate [Dickinson *et al.*, 1993]. Current climate models generally represent the land surface albedo by two-stream approximations for vegetated surfaces and by a limited number of prescribed values for non-vegetated surfaces. For example, the recently developed Common Land Model [Zeng *et al.*, 2002] and NCAR Community Land Model [Bonan *et al.*, 2002] specify bare soil albedos by 8 soil colors globally from dark to light. Each soil color has prescribed albedos that assume a near-infrared to visible albedo ratio of 2 and are independent of solar zenith angle (SZA).

[3] Such simple representation produces the largest albedo biases over Northern Africa and the Arabian Peninsula in climate models [Zhou *et al.*, 2003; Oleson *et al.*, 2003; Tian *et al.*, 2004]. Soils, sands, and rock are typically classified as a single land cover type. However, the solar shortwave diffuse albedos vary by a factor of about 2.5 from the darkest volcanic terrains to the brightest sand sheets [Tsvetsinskaya *et al.*, 2002]; the ratio of near-infrared to visible albedos observed in MODIS over deserts is in the range 1.6 to 2.7 [Zhou *et al.*, 2003]; and soil albedos increase significantly with SZA [Wang *et al.*, 2005].

[4] Satellites provide information of global spatial sampling at regular temporal intervals and thus can be used to

characterize the model albedo more accurately. Standard treatments that use satellite data spatially average their values from high-resolution pixels to lower-resolution model grids. Can we characterize spectral and spatial variations and SZA dependence of soil albedos with more detail than by simple grid means from the latest MODIS albedo products for use in coarse resolution climate models? To provide such a characterization, this paper analyzes MODIS 1 km albedo kernels of 7 spectral bands over Northern Africa and the Arabian Peninsula and through these kernels develops a new dataset that provides a simple economical way to scale up albedos from pixel to arbitrary coarse resolution grid square.

2. Data Processing and Methods

[5] We use the MODIS 1 km albedo product (MOD43B1, Level V004) that consists of 7 spectral bands over our study region for the period of 2000–2005. The MODIS albedo algorithm adopts a semiempirical, kernel driven linear Bidirectional Reflectance Distribution Function (BRDF) model to best characterize the anisotropy of the global surface. For each of the MODIS spectral bands, the BRDF model relies on the weighted sum of three kernels that are retrieved from the multivariate multiangular cloud-free atmospherically corrected surface reflectances. These kernels can be used to compute direct beam albedos at any given SZA and diffuse albedos based on simple cubic polynomials of SZA [Schaaf *et al.*, 2002]. Kernels for three broadband albedos, 0.3–0.7 μm (visible), 0.7–5.0 μm (near-infrared), and 0.3–5.0 μm (shortwave) as used in climate models can be easily obtained through simple spectral to broadband conversions [Liang *et al.*, 1999].

[6] The MODIS albedos represent the best quality retrieval possible over each 16-day period and quality assurance values (QA) are attached to the data. The QA data can be used to tell whether or not the data are of good quality. To ensure the best quality used in this study, we average only the pixels flagged with “good quality” QA (bits 0~7) in all 7 spectral bands over dust-free periods from November through January between 2000 to 2005 to generate a climatology of albedo kernels. Because vegetation is heterogeneous and varies seasonally, vegetated pixels are removed from this study according to the MODIS 1 km land cover map [Friedl *et al.*, 2002]. The resulting 1 km image consists of 13,653,383 barren pixels and each pixel has 21 kernels (3 kernels by 7 spectral bands).

[7] Our objective is to generate a high quality albedo dataset in a simple economical way that is useful for climate models. The albedo measurements provided by MODIS can be highly redundant because multispectral data bands

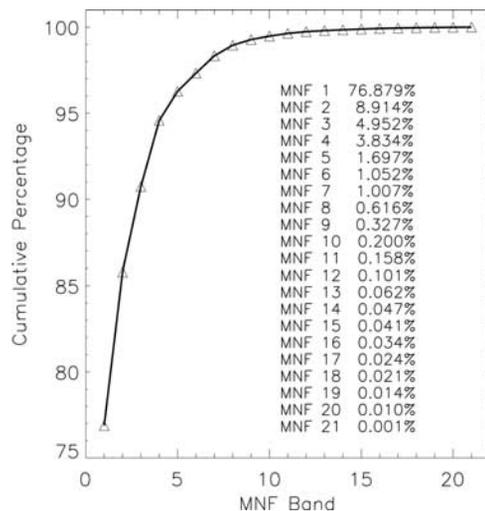


Figure 1. Cumulative percentage of eigenvalues for the MNF transforms. The percentage of eigenvalues for the 21 MNF bands is listed.

are often highly correlated and MODIS BRDF kernels are not orthogonal. They contain noise due to residual atmospheric effects and uncertainties and errors from observations and retrievals. Consequently, MODIS data can be made more useful with further statistical analysis. MODIS data can be modeled statistically by dividing into spatial patterns of large-scale, local-scale and noise. How such a division is made depends on the spatial resolution needed and how much detail one wants to characterize at such resolution. For coarse resolution climate models that typically have resolution from about 100–200 km, we may regard some local-scale variations as noise that is discarded to capture larger-scale albedo patterns.

[8] Here we use minimum noise fraction (MNF) rotation transforms to extract spatial structures, segregate noise, and reduce the dimensionality of the MODIS data using ENVI [Research Systems, Inc., 2004]. The MNF transform is essentially a sequence of two Principal Components Transformations (PCTs). ENVI assumes that the data for each pixel consists of signal and noise and that adjacent pixels contain the same signal but different noise. Because the noise of MODIS data is unknown over the study region, a “most homogeneous” subregion is chosen and a shift difference is performed over this subregion by differencing adjacent pixels to the right and above each pixel and averaging the results to estimate a spectral noise covariance matrix. The first PCT, based on the obtained noise matrix and applied to the entire image, decorrelates and rescales the data such that the transformed data has a noise with unit variance. The second PCT rotates the noise-whitened data to produce uncorrelated output bands by finding a new set of orthogonal axes that have their origin at the data mean and for which the data variance is maximized. The resulting uncorrelated MNF bands are linear combinations of the original spectral bands. The first MNF band contains the largest percentage of data variance (referred to as eigenvalue) and the highest spatial coherence; the second MNF band contains the second largest eigenvalue and second highest coherence, and so on, and the last MNF band is

noise-dominated and has the least variance. Therefore, inversion of the MNF transform using only the coherent bands provides a noise-filtered dataset.

[9] Figure 1 shows the cumulative percentage of eigenvalues for the MNF transforms, with the first 4 (10) MNF bands explaining 94.6% (99.5%) of the total variance of MODIS data over the study region analyzed. Evidently, the first several MNF bands are sufficient to represent the large-scale spatial patterns of albedos in climate models. Further analysis focuses on the first 7 MNF bands that explain 98.5% of the total variance. The amplitudes of these bands at 1 km are stored for further aggregation.

[10] We develop a new dataset for use in both regional and global climate models that provides a set of grid mean amplitudes by spatially averaging the 1 km MNF amplitudes for each model (square) grid cell at spatial resolution from 0.5° to 5° with an interval of 0.5°. For each resolution, seven MNF composites, i.e., between MNF band 1 to MNF band n , where $n = 1, 2, \dots, 7$, is considered in this dataset to provide an option to choose albedo patterns at various spatial scales. For example, one can use all the 7 MNF bands to represent all the large-scale albedo patterns or the first 3 MNF band to represent only continental scale albedo structures depending on spatial resolution and how much detail needed at such resolution as previously discussed. Since the grid mean amplitudes from the MNF data are anomalies relative to the MNF band means (a single value per band), a simple linear transformation (i.e., an inverse MNF transform) is needed to generate kernels with the absolute values for the MNF data that are used to calculate spectral and broadband albedos as done in MODIS [Schaaf *et al.*, 2002]. Note that our dataset is derived for dry soil. A climate model in addition would include a dependence of soil albedo on soil moisture.

[11] We also create an alternative but similar dataset for square grids at 10 km resolution that allows further aggregation into any coarser model resolution. A 10 km grid is fine enough to maintain the pixel information and coarse enough to be easily scaled up to coarser resolution model grids. Our statistics indicates that the grid mean values aggregated from 1 km data differ little from those aggregated from the 10 km data (see more discussion in Section 3).

[12] We quantify the performance of our new dataset in characterizing spectral and spatial variability and SZA dependence of soil albedo at four coarse resolution (square) grids of 50, 100, 150, and 200 km. For each resolution, we create two grid means of 21 kernels for each pixel following the above procedures, one from the 1 km MODIS data (referred to as the “MODIS based”) and the other from the 7 MNF bands (referred to as the “MNF based”), and calculate diffuse albedos and direct albedos at SZA of 0°, 25°, 50°, 75° for 10 bands (7 MODIS spectral bands plus 3 broadbands) as done in MODIS. The grid mean albedos obtained directly from MODIS versus those obtained from the 7 MNF bands are assessed using ordinary least squares to quantify the proportion of the total variance in MODIS data explained by the MNF method (referred to as R^2). For each band, seven R^2 are calculated for the MNF based albedos generated from the MNF bands 1, 1–2, 1–3, 1–4, 1–5, 1–6, and 1–7,

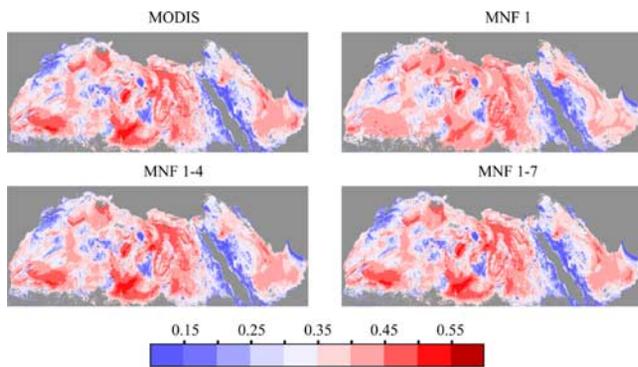


Figure 2. Spatial distribution of grid mean shortwave diffuse albedos obtained from MODIS and those obtained from the MNF bands 1, 1–4, and 1–7, respectively, at 10 km resolution.

respectively, which are denoted as $R^2(n)$, $n = 1, 2, \dots, 7$. Here n is used as a threshold to represent how much detail needed at a given resolution. As previously discussed, the first several MNF bands have the highest spatial coherence and represent the large-scale spatial patterns and thus R^2 represents the spatially correlated variance and $1-R^2$ the spatially uncorrelated variance. We assess the dependence of R^2 on spatial resolution and SZA. In total, there are 1400 values of R^2 (4 resolutions by 10 bands by 7 MNF composites by 4 direct albedos plus 1 diffuse albedo).

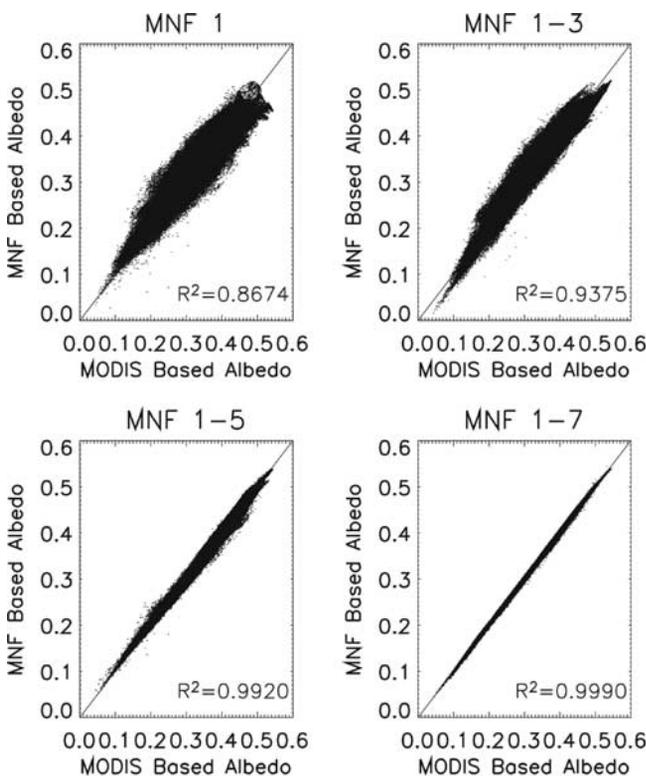


Figure 3. Scatter plots for MODIS based versus MNF based grid mean shortwave diffuse albedos at 10 km resolution. R^2 represents the proportion of the total variance in MODIS data explained by the MNF based method. There are in total 151,520 grids.

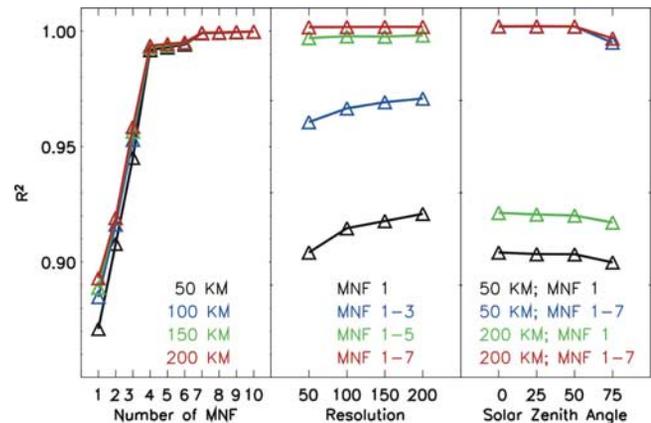


Figure 4. R^2 as a function of (left) the number of MNF bands at resolution of 50, 100, 150, and 200 km, (middle) spatial resolution for the MNF bands 1, 1–3, 1–5, and 1–7, and (right) solar zenith angle for the MNF bands 1 and 1–7 at resolution of 50 and 200 km, for shortwave diffuse albedos. R^2 is defined as Figure 3.

Direct albedos behave similarly to diffuse albedos and thus are not shown for most cases here.

3. Results and Discussion

[13] Figures 2 and 3 show spatial patterns and scatter plots of MODIS based and MNF based grid mean shortwave diffuse albedos at 10 km resolution. Evidently, a few MNF bands are able to capture most of the MODIS based albedo variances at this (and lower) resolution but adding more MNF bands captures yet more detailed albedo spatial structures and thus increases R^2 . For example, the first MNF band explains 86.7% of the MODIS variance and the first 7 MNF bands explain 99.9% of the variance at 10 km. Figure 4 shows how R^2 varies as a function of the number of MNF bands, spatial resolution, and SZA for shortwave diffuse albedo. It indicates that more correlated variance is obtained by keeping more MNF bands or using a coarser resolution grid. As expected, the relative contribution of each MNF band to R^2 significantly decreases as more high-ranked MNF bands are included such that the contribution after the MNF band 7 can be ignored (which is why we use only the first 7 MNF bands). For the MNF band 1 alone, R^2 increases significantly as the resolution becomes coarser but it varies little with resolution using all 7 MNF bands because they characterize most of the more local-scale albedo variations that are sensitive to resolution. R^2 decreases slightly as SZA increases but this decrease is not very sensitive to resolution and the number of MNF bands. The three kernels in MODIS data have different magnitudes, with a larger value for the isotropic kernels than that for the volumetric and geometric kernels; thus normalizing these kernels will slightly increase (decrease) the R^2 values for larger (smaller) SZA because both volumetric and geometric kernels determine the SZA dependence of albedo.

[14] We list in Table 1 the R^2 values of diffuse albedos for the 7 MODIS spectral bands and 3 broadbands at two coarse resolution grids of 50 and 200 km for 1 to 7 MNF bands. The R^2 shows a nonlinear relationship to the number of

Table 1. R^2 for Diffuse Albedos at Resolution of 50 km and 200 km

Band ^a	MODIS α^b	Number of MNF ^c						
		1	2	3	4	5	6	7
<i>50 km</i>								
1	0.14	3748	8360	9467	9675	9667	9702	9992
2	0.24	5924	8667	9438	9948	9969	9967	9991
3	0.36	7922	8466	9013	9789	9896	9899	9983
4	0.44	8916	9295	9387	9947	9950	9968	9978
5	0.52	9649	9737	9876	9973	9974	9984	9986
6	0.56	9347	9404	9913	9980	9983	9984	9988
7	0.50	8370	8398	9659	9980	9980	9982	9993
8	0.23	6619	8494	9277	9847	9881	9886	9993
9	0.46	9084	9258	9603	9982	9982	9986	9992
10	0.34	8710	9079	9451	9919	9930	9943	9991
<i>200 km</i>								
1	0.14	4316	8264	9527	9702	9691	9703	9993
2	0.24	6424	8689	9542	9957	9972	9971	9993
3	0.35	8260	8642	9229	9839	9913	9911	9986
4	0.44	9155	9437	9546	9956	9960	9973	9983
5	0.52	9703	9769	9912	9980	9981	9989	9991
6	0.56	9417	9455	9938	9987	9988	9990	9992
7	0.50	8512	8528	9753	9984	9984	9986	9995
8	0.23	7092	8559	9410	9871	9896	9895	9995
9	0.45	9226	9350	9713	9986	9987	9989	9994
10	0.34	8930	9191	9586	9937	9945	9951	9993

^aSpectral/broad bands (μm): 1: 0.459–0.479, 2: 0.545–0.565, 3: 0.62–0.67, 4: 0.841–0.876, 5: 1.23–1.25, 6: 1.628–1.652, 7: 2.105–2.155, 8: 0.3–0.7, 9: 0.7–5.0, 10: 0.3–5.0.

^bGrid mean albedo, α , over our study region.

^c R^2 values are multiplied by 10000.

MNF bands as also seen in Figure 4. For example, R^2 for the shortwave albedo (band 10) at 50 km increases from 87.1% for the first MNF band to 90.8% for the first 2 MNF bands but only increases by about 0.5% from the first 6 MNF bands to the first 7 MNF bands. For the fixed number of MNF bands, R^2 increases with decreasing resolution as also shown in Figure 4. For example, if a single MNF band is used, R^2 for the shortwave albedo increases from 87.1% at 50 km to 89.3% at 200 km. In addition, the R^2 values differ somewhat among the 10 spectral and broad bands and generally are larger for those with higher albedos. Since the MODIS average spectral albedos generally increase with wavelength, so do the R^2 values.

[15] As previously mentioned, an intermediate dataset at 10 km is also generated to further scale up to any lower resolution. The albedo biases produced from this dataset relative to those directly from the 1 km data are negligible at 50, 100, 150, and 200 km resolutions. For example, the largest albedo differences averaged over our study region obtained using the MNF band 1 alone at 50 km resolution are $\sim 10^{-4}$ (1.5, 2.8, 4.6, 5.5, 6.3, 6.5, 6.0, 2.8, 5.7, 4.3), respectively, for the 10 spectral and broad bands shown in Table 1.

[16] These results indicate that the MNF based specification of MODIS albedo kernels can significantly improve characterization of spatial and spectral variability and SZA dependence of albedo beyond the simple soil class related albedo scheme currently used in climate models. Such improvements are more pronounced for higher albedos, more MNF bands, and lower spatial resolution. For coarser resolution models, fewer MNF bands are needed.

[17] A compromise must be made between how much detailed information is wanted versus how many MNF bands are retained.

[18] Evidently, this study also provides a statistical method that is able to not only successfully capture most of the MODIS albedo variance but also to extract large-scale spatial patterns from the MODIS albedos while improving data quality and reducing the number of parameters needed to represent the data. The MNF bands can be also used to predict soil albedos for MODIS pixels with vegetation cover or missing data. For vegetated pixels, MODIS observations contain the contribution from both vegetation and its underlying soil while radiation models need to know the soil albedo as a boundary condition. Because vegetation is sparsely distributed over semi-arid regions, it can be treated as a disturbance on the large-scale patterns and these patterns can be used to interpolate soil albedos for vegetated pixels or ones with missing data.

[19] Since current climate models use only grid mean values of albedos while subgrid variability is not considered, reproducing grid means of albedos will be of greatest priority to modelers as done in this study. However, a corresponding dataset of grid variance for each model grid cell is also attached to our new dataset for possible use in describing the subgrid variability. Our further work will explore some possible applications of this variability in climate models.

[20] **Acknowledgment.** This work was funded by the NASA grants NNG04GK87G and NNG04GO61G.

References

- Bonan, G. B., et al. (2002), The land surface climatology of the NCAR community land model coupled to the NCAR Community Climate Model, *J. Clim.*, 15, 3123–3149.
- Dickinson, R. E., A. Henderson-Seller, and P. J. Kennedy (1993), Biosphere-Atmosphere Transfer Scheme (BATS) version 1e as coupled to the NCAR Community Model, *NCAR Tech. Note, NCAR/TN-387+STR*, 72 pp., Natl. Cent. for Atmos. Res., Boulder, Colo.
- Friedl, M. A., et al. (2002), Global land cover from MODIS: Algorithms and early results, *Remote Sens. Environ.*, 83, 287–302.
- Liang, S., A. H. Strahler, and C. Walthall (1999), Retrieval of land surface albedo from satellite observations: A simulation study, *J. Appl. Meteorol.*, 38, 712–725.
- Oleson, K. W., et al. (2003), Assessment of global climate model land surface albedo using MODIS data, *Geophys. Res. Lett.*, 30(8), 1443, doi:10.1029/2002GL016749.
- Research Systems, Inc. (2004), ENVI, The Environment for Visualizing Images, user's guide, version 4.0, Boulder, Colo.
- Schaaf, C. B., et al. (2002), First operational BRDF, albedo, and nadir reflectance products from MODIS, *Remote Sens. Environ.*, 83, 135–148.
- Tian, Y., et al. (2004), Land boundary conditions from MODIS data and consequences for the albedo of a climate model, *Geophys. Res. Lett.*, 31, L05504, doi:10.1029/2003GL019104.
- Tsvetinskaya, E. A., et al. (2002), Relating MODIS-derived surface albedo to soils and rock types over northern Africa and the Arabian peninsula, *Geophys. Res. Lett.*, 29(9), 1353, doi:10.1029/2001GL014096.
- Wang, Z., M. Barlage, X. Zeng, R. E. Dickinson, and C. B. Schaaf (2005), The solar zenith angle dependence of desert albedo, *Geophys. Res. Lett.*, 32, L05403, doi:10.1029/2004GL021835.
- Zeng, X., M. Shaikh, Y. Dai, R. E. Dickinson, and R. B. Myneni (2002), Coupling of the Common Land Model to the NCAR Community Climate Model, *J. Clim.*, 14, 1832–1854.
- Zhou, L., et al. (2003), Comparison of seasonal and spatial variations of albedos from MODIS and Common Land Model, *J. Geophys. Res.*, 108(D15), 4488, doi:10.1029/2002JD003326.

R. E. Dickinson, Y. Tian, and L. Zhou, School of Earth and Atmospheric Sciences, Georgia Institute of Technology, 311 Ferst Drive, Atlanta, GA 30332, USA. (lmzhou@eas.gatech.edu)