Potential of multiangular spectral measurements to characterize land surfaces: Conceptual approach and exploratory application

Nadine Gobron, Bernard Pinty, Michel M. Verstraete, John V. Martonchik, Yuri Knyazikhin, and David J. Diner

Abstract. New sensors exhibiting advanced technical specifications motivate the development of improved algorithms to take advantage of the enhanced performances of these sensors. In the particular case of the Multiangle Imaging Spectroradiometer (MISR) instrument, the angular sampling of the scattered radiance field, coupled with high spatial resolution and accurate radiometric calibration, justifies the implementation of physically based algorithms to optimally interpret the data and extract high-level information. This paper proposes a new approach to the reliable and accurate characterization of vegetated areas on the basis of data gathered in space and to the delivery of improved products to meet increasingly demanding user requests. An exploratory study based on advanced very high-resolution radiometer (AVHRR) data shows the potential of approaches based on advanced models but also points out the limitations associated with the use of data from monodirectional instruments. By contrast, a preliminary investigation conducted with synthetic MISR-like multiangular data illustrates the potential of analyzing data of high radiometric quality with advanced models to move toward a more complete characterization of terrestrial surfaces.

1. Introduction
Over the past decades, significant advances have taken place in the understanding of the scattering of light by the atmosphere, plant canopies, and soil surfaces [e.g., Ross, 1981; Simmer and Gerstl, 1985; Shultzis and Myneni, 1988; Marshak, 1989; Knyazikhin et al., 1992; Liang and Strahler, 1993; Govaerts and Verstraete, 1997; Pinty and Verstraete, 1998]. Progress in our understanding of the radiation regime of plant canopies has not resulted yet, however, in decisively better applications at the global scale. This is due, in large part, to the limited spatial, temporal, spectral, and directional sampling afforded by existing instruments, as well as by their relatively low radiometric accuracy. Advances in remote sensing technology during the past decade has changed this situation. Several instruments were designed specifically for multiangular spectral measurements, namely the Along-Track Scanning Radiometer (ATSR-2: two observation angles and a spatial resolution of 1 km) was launched in 1995 [Stricker et al., 1995], and the Polarization and Directionality of Earth Reflectances (POLDER) instrument (up to 14 observation angles and a spatial resolution of 7 km), which was operated for an 8-month period [Deschamps et al., 1994]. The Multiangle Imaging Spectroradiometer (MISR) instrument will provide multiangle and multispectral surface reflectance data at nine observation angles and a spatial resolution of 275 m–1.1 km [Diner et al., 1998]. These developments create new opportunities to conduct quantitative global and regional investigation of terrestrial surfaces. In this context we explore a new two-step approach to the quantitative characterization of land surfaces. The method consists first in creating a large look-up table (LUT) of simulated spectral and directional reflectance fields for a wide range of typical terrestrial surfaces, characterized by a small but critical set of field-measurable physical parameters. Remote sensing data are then analyzed through comparisons of the measurements with the entries of this LUT. When an acceptable match is found between a string of spectral and directional measurements and a corresponding string in the LUT for identical illumination and viewing geometries, the values of all the physical parameters associated with the generation of that entry in the LUT are considered as a possible solution to the radiative transfer problem. Clearly, it may well be that more than one combination of surface parameters may yield fields of spectral and directional reflectances similar to those retrieved from space measurements. In this case we report multiple solutions and recognize that there is a basic ambiguity in the satellite observations, in that they cannot differentiate between those geophysical situations. In the absence of any additional constraint or observation these equivalent solutions are used as a measure of the degree of ambiguity or accuracy of the retrieval. However, with complementary models and/or measurements it may be possible to decrease further the uncertainty about the state of the geophysical system.

We document the feasibility and limitations of this approach by conducting a couple of studies. The first one investigates to what extent quantitative information on terrestrial environments can be derived from multiangular spectral measurements acquired with AVHRR. We have thus constructed a first LUT suitable to describe the spectral and directional reflec-
tances of all major biomes and analyzed global AVHRR data accumulated over time. It will be seen that the number of combinations of surface parameters that may explain a particular set of such measurements is rather large. The wide range of surface parameter values identified as potential solutions are interpreted in part as a basic indicator of the limitations inherent in these AVHRR data.

In the second study we wanted to investigate to what extent MISR-like data would provide a better constraint on this problem and yield more reliable results when analyzed in the same manner. Advanced interpretation tools are being developed to fully exploit the high performances of this instrument [Diner et al., 1997, 1998b, c; Martonchik et al., 1998a, b]. It turned out that the predicted radiometric accuracy as well as the spectral and directional sampling of the MISR instrument were of sufficient quality that we increased the difficulty of the test and modified the study design to investigate the capability of MISR data to discriminate between surface conditions differing only in relatively minor ways from each other. Specifically, since the leaf area index (LAI) of plant canopies so largely controls the reflectance of the surface, we fixed that parameter to a single value ($\text{LAI} = 3$) and investigated to what extent MISR data could, in principle, differentiate between canopies with the same LAI but different leaf size and orientations, heights, or optical properties of the leaves and of the underlying soil. This is a more stringent test than was used for AVHRR data, and it will be seen that the availability of multiangular data of high radiometric accuracy is critical to limit the range of variability that each state variable can assume, given the constraints of the observations and the method of inversion.

This paper thus explores a new approach to the problem of land cover characterization and investigates the performance and limitations of this approach by documenting to what extent the availability of high radiometric accuracy multiangular data permits the identification of a unique set of surface parameters capable of accounting for the observations.

2. On the Role and Importance of Models

Conceptually, all approaches to remote sensing data interpretation rely on a model. The model can be explicit, as in the case of radiation transfer models, or implicit, as for vegetation indices that imply or assume an underlying model [Tucker et al., 1985; Kaufman and Tanré, 1992; Pinty and Verstraete, 1992; Myneni et al., 1995; Verstraete et al., 1996]. In any case models provide a link between the radiation measurements and the variables and processes controlling these observations. Consequently, remote sensing data can yield information only on those variables and processes that are explicitly represented in the model, independently of their intrinsic relevance for a particular application. However, other types of information can be derived if there are additional auxiliary or ancillary data sources, as well as additional models to derive this information.

The selection of a model depends (1) on the nature of the application and the accuracy required by the ultimate user of the derived information, (2) on the availability and quality of the data, and (3) on the allowable cost of producing the desired information. Clearly, accuracy requirements and computing constraints restrict the type of approaches that can be followed in any particular case. The selection of a specific model to interpret a given data set implies that the chosen model is capable of representing the observed radiative situation at the level of accuracy defined a priori. There must be a balance between the number of state variables required by the selected model and the characteristics of the available data sets in terms of radiometry, signal to noise ratio, spectral and angular sampling, etc. However, the higher the number of free state variables, the larger the probability of fitting well the data set and the larger the probability of finding multiple solutions to the radiation transfer problem. The latter point constitutes a key stumbling block, since identifying too many possible solutions is useless for the end user, but retrieving wrong information, as may be the case when applying very simplistic models, is also unsatisfactory.

A large panoply of data interpretation schemes has been suggested in the literature, from classical vegetation indices to computer graphics models enabling the representation of the radiation regime in very complex three-dimensional heterogeneous situations. Since the number of state variables remains a key issue for inversion purposes, the dimensionality of the model to be applied remains a critical choice to be made by the user early in the investigation. Typically, one-dimensional models are appropriate for spatially homogeneous surfaces, when average values are sufficient for the intended purpose, or when only limited detail is needed: such models typically use 5 to 7 state variables. These one-dimensional models can be combined to derive mixing models suitable for idealized heterogeneous surfaces, for which the radiative interactions between the components of the scene can be neglected. In this case the models may require 5 to 7 state variables per geophysical medium present in the scene. Only three-dimensional models are able to describe complex spatially heterogeneous surfaces and to provide a very detailed description of the scene; they imply a very large number of state variables on the order of the number of cell properties times the number of cells. As mentioned previously, the optimal choice of a model results from a compromise between the nature and the accuracy of the desired information, the type and quality of measurements, and the allowable exploitation cost. In the context of land surface applications, the desired information ranges from state variables of the radiation transfer problem, such as the leaf area index (LAI), to ecological mapping and monitoring to address land cover issues. The implication of choosing a particular interpreting tool is illustrated in Table 1.

Table 1 links the desired information to the input knowledge used in the development of the interpretation tool, which itself controls the nature of the retrieved output. The last column outlines some of the main drawbacks of each approach, in relation to the assumptions made about the geophysical medium being studied. The specificity and accuracy of the desired information increase downward in the table, as does the complexity and dimensionality of the model. However, the gain in accuracy of the description and therefore the knowledge obtained on the scene are also increasing downward. For instance, the broad patterns of vegetation distribution can be easily and economically described with vegetation indices, without requiring the formal inversion of sophisticated models. Since such multiple indices have been proposed [e.g., Huete, 1988; Price and Bausch, 1995; Pinty and Verstraete, 1992], the selection of a particular formula will be controlled at least in part by the type and amount of a priori knowledge available for the system. Nevertheless, the actual interpretation of index values must always be made in light of the explicit and implicit assumptions associated with the chosen index.

The importance of this tool selection problem is also related to the fact that the degree of general applicability of the in-
Table 1. Selection of a Particular Interpretation Tool Depends on the Information Requirements and the a Priori Knowledge Available About the Observed Target, Affects the Type and Reliability of Products Generated, and Automatically Defines the Possible Sources of Error

<table>
<thead>
<tr>
<th>Desired Information</th>
<th>Input Knowledge</th>
<th>Interpretation Tool</th>
<th>Output Retrieved</th>
<th>Sources of Data Misinterpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of vegetation</td>
<td>leaf spectral properties above + soil line above + atm</td>
<td>NDVI*</td>
<td>single value to be correlated with desired info</td>
<td>atm, soil, aniso atmosphere, compactness and heterogeneity</td>
</tr>
<tr>
<td></td>
<td>oriented point particles uniformly distributed</td>
<td>SAVI*</td>
<td>same as above but leaf R_L, T_L and orientation</td>
<td></td>
</tr>
<tr>
<td>Quantification of vegetation attributes</td>
<td>oriented plates uniformly distributed</td>
<td>GEMI*</td>
<td>same as above, plus canopy height leaf size and number</td>
<td>spatial heterogeneity</td>
</tr>
<tr>
<td>Land cover identification</td>
<td>specification of radiative properties in the 3-D space</td>
<td>inclusion of spatial distribution</td>
<td>explicit scene</td>
<td>validity of the representation of the scene</td>
</tr>
</tbody>
</table>

Atm and aniso designate the atmospheric and anisotropic perturbing effects, respectively. LAD, R_L, and T_L correspond to the leaf angle distribution functions, the leaf reflectance, and transmittance models, respectively.

*Normalized difference vegetation index [Rouse et al., 1973].
*Soil adjusted vegetation index [Huete, 1988].
*Global environment monitoring index [Pinty and Verstraete, 1992].

terpretation tool decreases downward in the table. Thus a specific three-dimensional model may resolve the radiation transfer problem in great detail and very accurately but for a very limited number of specific cases. This approach will provide a very good local solution, but the number of model implementations required to address global issues may be very large. On the contrary, a more generic one-dimensional model may describe the radiation transfer problem globally but at the cost of a poorer information content in any given case. Since the retrieval of information from remote sensing data always reduces to the fitting of a model to the data set, it is critical that model complexity and data accuracy and reliability be matched. As better constraints can be provided by new instruments, we are naturally using more advanced models to exploit these new opportunities. In other words the justification for using increasingly complex models depends on the quantity and quality of the data obtained by sampling the radiation field in the spectral and angular domains. Clearly, data of higher radiometric quality, acquired under a more extensive sampling scheme, should provide a better constraint on the model inversion process and should thus lead to a finer discrimination between competing representations of the geophysical environments of interest. It is in this context that the improved Earth Observing System strategy implemented through the new generation of sensors will probably yield improved environmental information and new applications.

3. Exploratory Application Using Global AVHRR Data

Most land surface applications at the global scale are based on AVHRR data and use vegetation indices as the main interpretation tool. As seen from Table 1, vegetation indices may provide an easy way to detect signatures related to the presence of vegetation, but it is only recently that they have become optimized to retrieve quantitative information [e.g., Gobron et al., 1999], and despite the many limitations associated both with AVHRR data and with the simplicity of the algorithm, important results have been obtained [e.g., Sellers, 1985; Townshend et al., 1987; Lambin and Ehrlich, 1995; Myneni et al., 1997]. Our objective here is to show that physically based approaches permit a much clearer assessment of the accuracy and reliability of the retrieved information and require less a priori knowledge about the observed environment. In particular, such an approach can take advantage of sources of variability (for instance, because of anisotropy of the radiation field) which are necessarily considered as noise in empirical, index-basis analyses.

3.1. Data Processing

Our initial exploration uses an existing data set extracted from the global vegetation index (GVI) database [Kidwell, 1990]. These data have been further processed to avoid cloud contamination and to reduce some of the most perturbing atmospheric effects [Berthelot et al., 1997]. The spectral measurements assigned to a GVI pixel are obtained by selecting, within a window of five lines by three columns in an global area coverage (GAC) data set, the spectral values associated with the pixel for which the NDVI has the highest value. The AVHRR/GAC pixel values, in turn, are obtained by averaging the spectral values of the first four pixels within a line and ignoring the fifth pixel as well as the following two lines within a local area coverage (LAC) full resolution AVHRR data set. Hence since the GVI data used here are heavily subsampled and averaged compared to the original data, their spatial resolution is nominally about 15 km. However, global data sets are available and can be processed in a reasonable amount of time.

The spectral bidirectional reflectance factors or radiances derived from AVHRR measurements are controlled by a series of radiative processes (including water vapor absorption as well as molecular and aerosol scattering) with which the solar radiation has interacted before it reaches the sensor. The successful application of advanced methods and algorithms to interpret surface bidirectional reflectance fields is conditioned by the availability of (1) high-quality data gathered by space sensors and (2) a reliable and accurate inversion technique to retrieve the radiative surface properties from the former. The AVHRR sensor is a multispectral but single view angle instrument which renders the handling of aerosol effects over land very problematic. In effect, AVHRR data do not provide enough constraints, specially in the angular domain [Marion-
Chik et al., 1998a], for solving accurately the inverse atmospheric problem. This, in turn, translates into a high degree of ambiguity to permit the estimation of accurate surface bidirectional reflectance factors [Lyapustin, 1999]. These limitations are not, however, compensated by any additional source of fundamental aerosol load and properties available all over the globe and at temporal and spatial scales compatible with the original AVHRR measurements. Therefore the derivation of accurate surface retrievals for global applications remains a key issue which cannot be performed completely satisfactorily on the basis of actual AVHRR data and/or in situ measurements.

Nevertheless, low-level algorithms can be applied to account for at least some typical effects due, for instance, to molecular scattering. In the particular case of AVHRR/GVI data, the radiative effects due to absorption and scattering by the main atmospheric constituents have been tentatively reduced by application of a procedure described by Berthelot et al. [1997]. These corrections were implemented with the Simplified Method for the Atmospheric Correction (SMAC) tool [Rahman and Dedieu, 1994], a simplified but fast parametric radiative transfer code in the atmosphere adapted from SS [Tanré et al., 1990] to correct AVHRR spectral observations. In this endeavor, Rayleigh scattering is prescribed following the profiles of a standard atmosphere, vertically integrated water vapor amounts are assumed to be those provided by the climatology of Oort [1983], and total ozone amounts are derived from Total Ozone Mapping Spectrometer (TOMS) data [McPeters et al., 1998]. The latter two components are varying spatially and temporally. Finally, the latitudinal profile of aerosol load is imposed as suggested by Berthelot and Dedieu [1997] and Berthelot et al. [1997]. Since there is no unique way to address these atmospheric effects, the procedure implemented by the team generating the Land Surface Reflectances (LASUR) data sets must be considered a first attempt to provide a useful solution to the problem of correcting AVHRR data for these atmospheric effects.

The AVHRR/GVI/LASUR data have been further composited in time to limit the effect resulting from the presence of clouds. Thus exactly 2 AVHRR spectral measurements (red and near infrared) are available for each pixel of about 15 by 15 km, every 7 days. Since AVHRR is a monodirectional instrument, the only way to emulate multangular observations was to accumulate data in time, assuming the surface did not change during that period. Our analysis is based on a monthly compilation of these 7-day periods to accumulate sufficient angular sampling to apply the algorithm.

3.2. Selection of a Particular Model and Inversion Technique

The application of a complex three-dimensional model would have required a complete documentation of the structure and properties of a large number of terrestrial environments. For the purpose of this exploratory research we have opted, instead, to select a flexible one-dimensional radiation transfer model [Gobron et al., 1999b] to represent the interaction of light with plant canopies. It implements a statistical description of the discrete nature of the canopy to relax the continuous (turbid) medium assumption and includes an explicit representation of architectural effects for homogeneous canopies.

In this model, the first 2 orders of scattering (by the soil and by the leaves) are calculated in three-dimensional space after an adaptation of a discrete model originally developed by Verstraete [1987] to account for the extinction of the direct incoming solar radiation. This statistical description of the canopy architecture permits an explicit representation of the enhanced backscattering effects due to the finite size of leaves, also known as the hot spot phenomenon in the first 2 orders of scattering [Verstraete et al., 1990]. The multiple-scattering contribution is calculated with a Discrete Ordinates Method using an azimuthally averaged expression of the anisotropic scattering phase function proposed by Shultis and Myneni [1988]. Extensive comparisons against a ray-tracing model simulating homogeneous canopies have demonstrated the high performances of this semidiscrete model [e.g., Gobron et al., 1997b].

Another significant advantage in this choice of model is the possibility to generate large look-up tables (LUTs) relatively quickly, on the basis of well-identified canopy parameters. In its basic version the model requires seven input state variables, namely, the leaf angle distribution function, the leaf reflectance and transmittance coefficients, the soil albedo or BRF values, and any three of the four following information items: the height of canopy, the leaf area index, the equivalent leaf diameter, and the number of leaves per unit volume. For most practical purposes the use of LUTs storing a large number of bidirectional spectral reflectance factors is more effective than classical inversion techniques allowing a wide range of solutions, possibly distinguished only by numerical variations dependent on the precision of the computer. The latter technique has however proven useful for methodological research, where operational constraints are largely absent [e.g., Pinty et al., 1989; Camillo, 1987; Goel and Strebel, 1983; Privette et al., 1995].

When implementing a LUT approach, the operator selects a priori the number and type of radiative solutions that will be considered. This implies that if no constraint is placed on the maximum allowed value of the cost function, at least one solution will always be chosen, but the accuracy of the solution depends on the level of discretization of the LUT. This is a definite advantage over classical techniques seeking the minimum of a cost function with few constraints on the solutions. Since the solution of an inverse problem by searching a match between measurements and simulations available in a LUT requires the initial identification and characterization of all possible cases, the discretization of the LUT must be sufficiently fine to ensure that at least one reasonable solution can be associated with any likely set of measurements. On the other hand, the generation of very highly discretized LUTs will often result in the identification of multiple simulated cases which could explain the observations and thus may lead to the identification of nonunique solutions.

For this exploratory study, based on the AVHRR/GVI/LASUR data, 35 "radiative biome types" supposed to span the range of possible solutions for each GVI pixel have been defined through the model parameter values given in Table 2. Only canopies with standard healthy green leaves have been considered, and little attention has been given to the representation of bare soil conditions. Canopy spectral reflectance and transmittance properties have been extracted from the data set provided by Price [1995]. The results shown below thus depend in part on this specific hypothesis. In practice, it is obviously feasible to increase the size of the LUT by adding more leaf spectral properties, but this mainly leads to the identification of multiple solutions to the inverse problem. After performing a number of sensitivity experiments, we concluded that the set of
Table 2. Definition of 35 Radiative Biome Types

<table>
<thead>
<tr>
<th>LAI</th>
<th>Diameter of a Single Leaf</th>
<th>Height of the Canopy</th>
<th>Leaf Angle Distribution</th>
<th>Brightness of Soil</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>small</td>
<td>low</td>
<td>erectophile</td>
<td>dark</td>
</tr>
<tr>
<td>1</td>
<td>small</td>
<td>low</td>
<td>planophile</td>
<td>bright and dark</td>
</tr>
<tr>
<td>1</td>
<td>large</td>
<td>low</td>
<td>erectophile</td>
<td>dark and bright</td>
</tr>
<tr>
<td>1</td>
<td>large</td>
<td>low</td>
<td>planophile</td>
<td>dark</td>
</tr>
<tr>
<td>1</td>
<td>large</td>
<td>medium</td>
<td>erectophile</td>
<td>dark</td>
</tr>
<tr>
<td>1</td>
<td>large</td>
<td>medium</td>
<td>planophile</td>
<td>dark</td>
</tr>
<tr>
<td>2</td>
<td>small</td>
<td>low</td>
<td>erectophile</td>
<td>dark</td>
</tr>
<tr>
<td>2</td>
<td>small</td>
<td>low</td>
<td>planophile</td>
<td>dark and bright</td>
</tr>
<tr>
<td>2</td>
<td>large</td>
<td>low</td>
<td>planophile</td>
<td>dark</td>
</tr>
<tr>
<td>2</td>
<td>large</td>
<td>medium</td>
<td>erectophile</td>
<td>dark</td>
</tr>
<tr>
<td>2</td>
<td>large</td>
<td>medium</td>
<td>planophile</td>
<td>dark</td>
</tr>
<tr>
<td>3</td>
<td>small</td>
<td>low</td>
<td>erectophile</td>
<td>bright</td>
</tr>
<tr>
<td>3</td>
<td>small</td>
<td>low</td>
<td>planophile</td>
<td>bright and dark</td>
</tr>
<tr>
<td>3</td>
<td>large</td>
<td>low</td>
<td>planophile</td>
<td>bright and dark</td>
</tr>
<tr>
<td>3</td>
<td>large</td>
<td>low</td>
<td>planophile</td>
<td>bright and dark</td>
</tr>
<tr>
<td>5</td>
<td>small</td>
<td>low</td>
<td>planophile</td>
<td>bright</td>
</tr>
<tr>
<td>5</td>
<td>small</td>
<td>low</td>
<td>planophile</td>
<td>dark and bright</td>
</tr>
<tr>
<td>5</td>
<td>large</td>
<td>medium</td>
<td>erectophile</td>
<td>dark</td>
</tr>
<tr>
<td>5</td>
<td>large</td>
<td>medium</td>
<td>planophile</td>
<td>bright</td>
</tr>
<tr>
<td>5</td>
<td>large</td>
<td>medium</td>
<td>planophile</td>
<td>bright and dark</td>
</tr>
<tr>
<td>0</td>
<td>large</td>
<td>low</td>
<td>planophile</td>
<td>dark</td>
</tr>
<tr>
<td>0</td>
<td>large</td>
<td>low</td>
<td>planophile</td>
<td>very bright</td>
</tr>
</tbody>
</table>

“Small” and “large” leaf diameters correspond to values of 1 and 5 cm, respectively; “low and medium” heights of the canopy correspond to values of 0.5 and 2 m, respectively. The soil albedo values are fixed at 0.07 for dark and 0.16 for bright soils in the visible band and at 0.09 for dark and 0.21 for bright soils in the near-infrared band, respectively. Note that the two bare soil cases correspond to anisotropic surfaces with a single-scattering albedo of 0.1 for dark and 0.7 for very bright soils in the visible band and 0.2 for dark and 0.8 for very bright soils in the near-infrared band. The other two model parameters are the hot spot parameter (0.1) and the asymmetry factor (~0.2).

35 radiative biome definitions described in Table 2 were adequate to produce an acceptable trade-off between an accurate but sufficiently unambiguous solution, without overtaxing the computer resources available. Again, the strategic priority was to identify at least one reasonable solution to the inverse problem for each pixel. The selection of model properties and the design of the LUT were based on the assumptions that (1) all pixel values do correspond to surface conditions which can be effectively simulated by the semidiscrete model mentioned above, (2) land surfaces do not evolve substantially during the period of data accumulation (one calendar month), and (3) all measured bidirectional reflectance factors were accurately decontaminated from atmospheric effects in the process of generating the LASUR product. These three assumptions clearly affect the scope and representativity of this exploratory study, but it should be noted that the severity of the assumptions will be largely reduced with the progressive availability of better data from the new generation of sensors, and in particular MISR, which will provide four spectral measurements in each of nine observation directions for each pixel within a period of a few minutes.

This being said, more than one solution may be found to describe terrestrial environments even when analyzing MISR data, depending on the discretization level of the LUT and on the spectral and bidirectional sampling of the reflectance field. Moreover, any set of environmental conditions (soil, canopy, and atmospheric properties) which result in sufficiently similar observed reflectance distributions, meaning sets of spectral and directional reflectances not statistically different at the level of precision of the measurements or of the computer, will be indistinguishable and therefore constitute equally acceptable solutions to the radiation transfer problem.

In addition, there are inherent limits to the retrievability of some model parameters, independently from the approach followed for data analysis or the selection of the interpretation model. For instance, our understanding of the transfer of radiation in terrestrial environments leads to the conclusion that the leaf area index of a plant canopy cannot be reliably retrieved from any analysis of remote sensing data when its value exceeds about 4 or so, or when the spectral and directional properties of the underlying soil are too close to those of the canopy [e.g., Privette et al., 1995; Gobron et al., 1997a; Pinty et al., 1998]. Thus it is conceivable that two or more geophysical situations lead to indistinguishable reflectance observations, and the LUT should have a sufficient number of entries to include at least one of these situations as a solution.

3.3. Design of the Inversion Algorithm

The algorithm applied to identify the most probable solutions to the inverse problem among the set of possible solutions predefined in the LUT operates in three basic steps. The first one consists in calculating, for each of the 35 radiative biomes \( b \) defined in Table 2, for each couple of AVHRR spectral measurements \( \lambda \) at a given date \( t \), and for each pixel \( p \), the following cost function:

\[
\delta^2(p, b, \lambda, t) = 4 \frac{[\rho_{\text{lu}}(\theta_u^p, \theta_u^p, d\phi^p, \lambda) - \rho_{\text{lu}}(\theta_u^b, \theta_u^b, d\phi^b, b, \lambda)]^2}{[\rho_{\text{lu}}(\theta_u^b, \theta_u^b, d\phi^b, \lambda) + \rho_{\text{lu}}(\theta_u^b, \theta_u^b, d\phi^b, b, \lambda)]^2},
\]

where \( \theta_u^p \) and \( \theta_u^b \) are the illumination and observation zenith angles, respectively, and \( d\phi^p \) is the relative azimuth angle.
between the solar and the viewing planes. In this equation, 
\( \rho_{wp}(\theta_0, \theta_v, d \phi_v, \lambda) \) correspond to the measured bidirectional reflectance factors and 
\( \rho_{wp}(\theta_0, \theta_v, d \phi_v, b, \lambda) \) correspond to 
the simulated bidirectional reflectance factors implemented in 
the LUT for the biome \( b \) in the same spectral band and under 
identical geometrical conditions.

This cost function measures the quadratic distance between 
the bidirectional spectral reflectance factors available with the 
AVHRR instrument and the model simulations. The pixels 
associated with a \( \delta^2 \) value larger than 0.2 (~30% of the relative 
reflectance values) were disregarded from further analysis. 
These spectral time-dependent cost functions are averaged 
over a period of typically 1 month but with the constraint that 
no more than one individual cost function may exceed 0.2. 
If this constraint is not verified, the biome is not considered 
a possible solution of the problem.

In the second step, the algorithm ranks the biomes in 
increasing order of these averaged monthly cost functions 
separately in the two channels, with a rank zero for the best solu-
tion(s). In the third and last step, a quality index is evaluated to 
identify which entries in the LUT lead to the smallest cost 
function in all spectral channels and thus represent the most 
probable solutions. In this application all predefined biomes in 
the LUT leading to a change in the \( \delta^2 \) value of less than 0.0008 
(~2% of the relative reflectance values) are considered as 
indistinguishable and therefore generate multiple solutions in 
that spectral band. In other words, if the spectral reflectances 
that populate the LUT do not differ by at least 2% in at least 
one spectral band, the corresponding biomes are given the 
same rank.

This quality function is evaluated as follows:

\[
Q(b, n_b) = 1 - \frac{1}{n_b k} \sum_{\lambda} a_\lambda R(b, \lambda_k) \quad R(b, \lambda_k) < R_{max},
\]

where \( k \) is the total number of spectral bands available for 
the analysis, \( n_b \) is the number of biomes defined as possible solutions of 
the inverse problem, \( a_\lambda \) is the spectral weight associated 
with the corresponding spectral band (set to 1 in our 
initial application), and \( R_{max} \) specifies the highest allowed rank 
beyond which the probability to find a given biome \( b \) is 
considered very low. \( R(b, \lambda_k) \) is the rank associated to the biome 
\( b \) for each spectral band \( \lambda_k \). Finally, the radiative biome(s) for 
which \( Q(b, n_b) \) is closest to 1 is (are) selected as the most 
probable solution(s). Since multiple combinations of spectral 
ranks may give the same value for the function \( Q(b, n_b) \), there 
is no guarantee of finding a unique solution [see Gobron, 
1997].

Other inversion algorithms may be considered, and different 
cost functions and quality indices combining the results in all 
available spectral bands may be proposed. For instance, 
approaches based on the minimization of cost functions defined 
as sums over all spectral channels of quadratic distances be-
tween data and model simulation may certainly be valuable. 
However, their main drawback is to generally yield results 
baised by the spectral and angular conditions responsible for 
the largest differences between actual and simulated data. 
Such a bias obviously affects the analysis of AVHRR data, 
since the correction scheme for atmospheric effects assumes 
simplified latitudinal aerosol profiles, which may not be accu-
rate enough. This will translate, in turn, into overestimated 
bidirectional reflectance factors emerging the land surfaces in 

the visible channel, which can themselves be understood as 
a vegetation canopy less dense than is actually the case. In our 
initial exploration the quality of the information derived from 
each and every individual measurement is considered to be 
equally reliable and independent of the spectral band or the 
angular conditions of illumination and observation. The role of 
the parameter \( R_{max} \) in (2) is to restrict the search for solutions 
among those which have a sufficient degree of quality in each 
band separately. Our experience shows that this approach 
leads to the identification of at least one solution for each pixel 
and that solutions become highly improbable when at least one 
of the spectral ranks becomes too high. Restricting the search 
in this way thus reduces the time and computing resources 
needed to identify the most probable solutions.

3.4. Results

The algorithm described above has been applied to the analysis 
of composite weekly data sets derived from the AVHRR/ 
GVI/LASUR database. Only a detailed case-by-case investiga-
tion of the geometries of illumination and observation for each 
pixel and for the periods under investigation could establish 
whether merging data from multiple consecutive periods ef-
fectively results in a useful increase in angular sampling. How-
ever, this assumption has been made and the algorithm has 
been applied to data sets corresponding to composited 
AVHRR/GVI/LASUR data for the entire calendar month, 
namely January (5 weeks) 1989.

The solid line in Figure 1 shows the proportion of pixels in 
these global data sets, analyzed as described above, which 
yielded the indicated number of acceptable solutions (i.e., bi-
omes whose radiative properties matched the observations in 
the sense given in the previous sections). The dotted line in 
the same figure shows the cumulative number of pixels for which 
the indicated number or fewer solutions were found. Hence 
the accumulation of weekly AVHRR/GVI data for a period of 
a month is sufficient to identify a unique biome type for about 
40% of the terrestrial areas, including the warm and cold 
desertic regions. In this preliminary study we set \( R_{max} = 15 \), so 
no pixel data ever led to the identification of more than 15 
acceptable biomes, out of the 35 possible solutions identified 
in Table 2.
These first results illustrate that it is not possible to find a unique solution for as much as 60% of the vegetated pixels on the sole basis of information provided by radiation transfer models and AVHRR/GVI/LASUR data. Moreover, we found that these nonunique radiatively equivalent solutions can sometimes span a large range of LAI values. This latter result is shown on the global maps in Plate 1 which exhibits the minimum and maximum values of LAI corresponding to the radiatively equivalent biomes identified by our algorithm as potentially accounting for the observed variability in the data, for the month of January 1989. The proper interpretation of these maps is that on the basis of the AVHRR data analyzed here, and with the proposed approach, the estimated global distribution of LAI lies between the two extremes shown. The implied LAI values are reasonable in many places, but the range of possible values is largest over equatorial regions that are highly contaminated by smoke and clouds. In our experience, a wide range of values for retrieved parameters such as the LAI is generally associated with a large number of radiatively equivalent solutions. A global map of the geographical distribution of the number of acceptable solutions does not show random patterns but rather suggests a relation between this number and the type of biogeophysical system. However, this latter result is difficult to interpret since it is strongly dependent on the discretization of the possible solutions in the LUT.

A careful inspection of the global maps in Plate 1 shows that the determination of the land cover type for a particular pixel is not trivial and often leads to nonunique solutions: typically, even an important environmental parameter such as LAI may vary by one unit or more among the radiatively acceptable solutions. This result is determined in part by technical limitations (including the 8 bits of coding of the AVHRR/GVI/LASUR data and the relatively low signal to noise ratio of the instrument), but also by fundamental physical limits, as can occur when two or more different biogeophysical systems exhibit, for a given set of illumination and observation geometries, essentially similar spectral and directional reflectances [e.g., Gobron et al., 1997a]. This problem is compounded by the limitations and inaccuracies associated with the atmospheric correction of the data. Indeed, the contamination of measurements by inaccurate corrections for water vapor, clouds, and aerosols, especially in tropical regions, may lead to systematic biases and therefore the selection of solutions corresponding to relatively high (low) reflectances in the visible (near infrared) channels and thus to an underestimation of the LAI. Clearly, these problems limit the detailed interpretation of the results, since for instance, it is difficult to guarantee that the

Plate 1. Range of LAI values using weeks 1–5 of GVI/LASUR data in January 1989. The map at the top corresponds to the minimal values and the bottom one to the maximal values that could account for the observations.
accumulation of GVI data over time (1 month in that case) does not add more noise than reliable information in the constructed data sets.

Nevertheless, the five color codes used to represent LAI in Plate 1 (the black pixels identify targets for which no solution was implemented in the LUT a priori) illustrates the great spatial and temporal variabilities encountered over the Earth surface and the sharpness of the major boundaries between various biogeophysical systems over various continents.

As a preliminary conclusion from this experiment, it appears that a significant amount of ancillary information may be required to narrow down the number of acceptable solutions when analyzing AVHRR/GVI/LASUR data. This information may come from additional models of terrestrial cover dynamics in order to increase the constraints on the solutions to be found with the filtered temporal signals, or from data of much better quality with respect to such issues as the calibration, the digitization level, the signal to noise ratio, and the atmospheric correction procedures. Indeed, with the limited spectral, directional, spatial, and temporal sampling of AVHRR/GVI/LASUR data, variations in the radiation field induced by a change in the leaf spectral properties often cannot be distinguished from a possible concurrent change in canopy LAI, for instance. Section 4 explores the benefits associated with the second option in the context of the exploitation of MISR data on the Terra platform.

### 4. Preliminary Investigation With Synthetic MISR-like Data

The exploratory study described earlier shows the limits inherent to the use of low-quality data, accumulated over time from a monodirectional sensor. It has been seen that when data quality is poor, in the sense that a significant part of the variability of the signal cannot be definitely assigned to known causes, it does not make sense to have a dense discretization of the LUT, as multiple combinations of predefined geophysical conditions will regularly result in indistinguishable solutions. On the contrary, as users demand higher accuracy and reliability in the products generated from remote sensing data, instruments must be designed and operated in such a way that finer distinctions must be made, so the level of discretization of the LUT must be adjusted to the quality of data. In the remainder of this section we will focus on exploring to what extent the characteristics of the MISR instrument permit an effective reliable and accurate determination of the global land cover distribution. Our analysis supports the feasibility of a physically reliable and accurate determination of the global land cover characteristics of the MISR instrument permit an effective exploration to what extent additional spectral bands could also help reduce the number of acceptable solutions, while the combined use of six MISR cameras results in the selection of, at most, six biome characterizations. Figure 2 shows the number of radiometrically acceptable solutions retrieved in this experiment, as a function of the number and type of angular combinations that have been used. It can be seen that the number of acceptable solutions decreases with the number of cameras used in the inversion scheme. For instance, the use of a single camera may yield from 3 to 15 acceptable solutions, while the combined use of six MISR cameras results in the selection of, at most, six biome characterizations. Figure 2 also shows, in a very preliminary way, the advantage of analyzing multiangular data: given data in two spectral bands, the addition of further measurements in other view directions significantly restricts the range of biomes that could be considered probable solutions. In this respect, we chose the radiative biome 127 as the target environment and assessed to what extent simulated MISR measurements of this biome could be confused for any one of the 160 radiative biomes described in the LUT. This biome corresponds to a canopy where the leaf angle distribution is erectophile with the height of 2.0 m, the leaf diameter is 5 cm. The spectral leaf properties correspond to the first one in the Table 3, and the underneath soil is the dark one. The algorithm was applied sequentially to the data generated by each of the MISR cameras separately, all possible combinations of data produced by 2 to 8 MISR cameras, and the data generated by all nine cameras together. Figure 2 shows the number of radiometrically acceptable solutions retrieved in this experiment, as a function of the number and type of angular combinations that have been used. It can be seen that the number of acceptable solutions decreases with the number of cameras used in the inversion scheme. For instance, the use of a single camera may yield from 3 to 15 acceptable solutions, while the combined use of six MISR cameras results in the selection of, at most, six biome characterizations. Figure 2 also shows, in a very preliminary way, the advantage of analyzing multiangular data: given data in two spectral bands, the addition of further measurements in other view directions significantly restricts the range of biomes that could be considered probable solutions, given the observed or simulated data set. Clearly, it will be important to document to what extent additional spectral bands could also help reduce the number of acceptable solutions.

The quality of the retrieval can be examined with Figures 3 and 4, which display the frequency of finding specific biomes (i.e., the probable solutions) for each inversion experiment and the location in spectral space (red versus near infrared) of the

<table>
<thead>
<tr>
<th>Diameter of a single leaf</th>
<th>1 and 5 cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height of the canopy</td>
<td>0.5, 2, 5, and 10 m</td>
</tr>
<tr>
<td>Leaf angle distribution</td>
<td>planophile, erectophile, uniform, plagiophile, and extremophile</td>
</tr>
<tr>
<td>Leaf reflectance and transmittance red domain</td>
<td>(0.0876, 0.0465) or (0.06520, 0.0465)</td>
</tr>
<tr>
<td>near infrared domain</td>
<td>(0.46849, 0.41511) or (0.41849, 0.40511)</td>
</tr>
<tr>
<td>Soil albedo red domain</td>
<td>0.0743 (dark) and 0.16339 (bright)</td>
</tr>
<tr>
<td>near infrared domain</td>
<td>0.0913 (dark) and 0.2125 (bright)</td>
</tr>
</tbody>
</table>
angular signatures of all biomes predefined in the LUT, respectively. From Figure 3 it can be seen that (1) the “true” solution (biome 127) is always selected as one of the most probable solutions, (2) the progressive addition of cameras in the inversion scheme reduces considerably the number of acceptable solutions (i.e., significantly contributes to the discrimination between biome types), (3) the same five solutions are systematically selected when data from six or more cameras are used in the analysis, and (4) a single inversion applied using the nine cameras together is enough to retrieve these five solutions with an equal probability.

It must be emphasized that the identification of five equally acceptable solutions in this particular exercise is largely controlled by the number and type of possible solutions that were included a priori in the LUT and represents the range of possibilities allowed by the semidiscrete model under these conditions. Figure 4 is very informative in this regard, since it shows that the spectral reflectance fields generated by these five solutions and sampled by the MISR instrument are indeed very similar to each other and to the radiance fields corresponding to the “true” solution. Although this conclusion is quite obvious from a radiation transfer perspective, it underscores the difficulty of discriminating different geophysical systems which exhibit similar spectral and directional signatures. In this particular application the only differences among the five solutions retrieved concerned the leaf diameter and the canopy height. When duplicating the experiment just described

---

**Figure 2.** Number of solutions retrieved by applying the LUT inversion procedure using two spectral bands (red and near infrared), as a function of the number of angular measurements used. The numbers in the diagram indicate how many combinations of the chosen number of cameras, taken among the nine possible MISR cameras, yield the specified number of solutions. The sum of all these numbers in a given vertical column corresponds to the total number of combinations of $m$ elements taken nine at a time.

**Figure 3.** Retrieval frequency of the biome type as a function of the number and combinations of the MISR cameras used in the inversion procedures. The “true” solution (B127) is indicated by an arrow.
with synthetic MISR data generated in the principal plane, the inversion procedure identified only two biomes (biomes 7 and 127) instead of 5. These two biomes differ only by the values of the height of the canopy and the leaf diameter equal respectively to 0.5 m and 1 cm for biome 7. However it can be noticed that the ratio of the leaf diameter by the height of the canopy remains almost the same for biomes 7 and 127. This result is consistent with the idea that the sampling of the radiance fields in the principal plane allows a better discrimination of architectural properties than is achievable from other azimuthal observation conditions.

The models used in the interpretation of the data play a critical role in determining the type, amount, accuracy, and reliability of the information retrieved from an analysis of remote sensing data, as discussed earlier. The exploratory retrieval exercise just described is clearly based on a rather ideal situation, in the sense that the same model is used for generating the possible solutions in the LUT and for identifying the acceptable solution from an analysis of a particular data set. However, further investigations in this direction would not be justified if this approach could not be shown to work under these circumstances.

5. Conclusions

One purpose of this research was to investigate a physically based approach to the interpretation of multiangular remote sensing data and to evaluate its advantages and drawbacks in the context of the characterization of land surfaces.

Since physical models encapsulate our understanding of the spectral and angular reflectance fields, they can objectively document the state of these surfaces in terms of relevant measurable geophysical variables. Clearly, the ability to take full advantage of the performance of such models depends critically on the quality of the data and on the density of the directional and spectral sampling. This was illustrated by applying the same algorithm first to a heavily averaged and subsampled AVHRR data set typical of the global observations currently available and then to an idealized simulation of the MISR instrument. The various sources of uncertainties in the AVHRR data (e.g., calibration, radiometric, and spectral resolution), as well as the difficulty to account for atmospheric corrections and surface anisotropy effects, result in large errors and therefore in potentially frequent confusions between biome types. Thus the LUT must be rather small, and only a few different biogeophysical systems can be effectively distinguished. Conversely, the high-quality data generated by instruments such as the MISR sensor will permit a much more refined and reliable characterization of land surfaces.

This exploratory study suggests that the feasibility of discriminating between biomes of the basis of reflectance measurements alone critically depends on the quality of the data and specifically on the availability of multangular surface level observations, on the radiometric resolution of the measurements, and on the proper calibration of the data.

The specific advantages of the MISR instrument were explored through a simulation study. The analysis of data acquired almost simultaneously by multiple cameras under different observation angles appears to allow a much better discrimination between surface types and a more detailed, reliable, and accurate description of the corresponding biomes than is feasible from monodirectional instruments.

Acknowledgments. Some of the results shown in this paper were presented at the workshop entitled “Multiangular Remote Sensing Progress and Advancing BRDF Usefulness in the EOS Era and Beyond” (organized by NASA in January 1997), at the Seventh International Symposium on Physical Measurements and Signatures in Remote Sensing (organized by ISPRS in April 1997), and at the MISR Science Team Meeting (organized by JPL in June 1997). The authors acknowledge the support of JPL, of the VEGETATION preparatory programme financed by the European Union, France, Belgium, Sweden, and Italy, as well as of the NASDA, under the GLI programme. Part of this work has been achieved in the framework of Nadine Gobron’s Ph.D. thesis (May 1997), who currently benefits from an JRC postdoctoral fellowship.

References


