Multitemporal, Multichannel AVHRR Data Sets for Land Biosphere Studies—Artifacts and Corrections

Josef Cihlar, Hung Ly, Zhanqing Li, Jing Chen, Hartley Pokrant, and Fengting Huang

Temporal compositing of daily optical satellite data has become an accepted methodology for obtaining frequent images of large areas for studies of the land surface. However, such composite data sets sometimes contain large "artifacts" (i.e., errors due to sources unrelated to the land surface itself). The goal of this study was to develop a series of operations and algorithms that would identify and remove as many of these errors as possible. The specific objective was to obtain, every ten days, surface reflectance in advanced very high resolution radiometer (AVHRR) channels 1 and 2 and normalized difference vegetation index (NDVI)—all referenced to a constant viewing geometry—and surface temperature. The processing steps of the resulting methodology, dubbed "ABC3," for atmospheric, bidirectional, and contamination corrections of the Canada Centre for Remote Sensing (CCRS), includes atmospheric corrections; the identification of pixels "contaminated" by clouds, snow (including subpixel); and shadows from optically thick clouds; bidirectional reflectance and thermal emissivity corrections; and the replacement of the contaminated pixels in the composites through interpolation. The resulting procedures yield surface reflectance and NDVI fully corrected for bidirectional effects; a version of NDVI corrected for solar zenith only; and surface temperature corrected for atmospheric and surface emissivity effects. An evaluation of the resulting data set shows that the procedures provide significantly improved data products compared with the raw composites, but they do not approximate a single-date image sufficiently closely, especially for AVHRR channel 1, whose reflectance values are generally low. This is attributed to the limitations of the input data and knowledge of atmospheric (and partly bidirectional) characteristics applicable to each composite pixel. © Government of Canada, 1997

INTRODUCTION

Medium-resolution optical satellite data have become an important source of information about the characteristics and dynamics of the land biosphere at regional and global scales. The unique strengths of such data types afforded by daily coverage and sensitivity to the ecosystem state have been demonstrated in numerous studies with data from the advanced very high resolution radiometer (AVHRR) on board National Oceanic and Atmospheric Administration (NOAA) satellites. These successes have led to the construction of several similar sensors to be launched in the late 1990s: VEGETATION on board SPOT 4 in 1997 (Saint, 1992), MODIS on the first Earth Observing System platform (Salomonson, 1988; Bunning et al., 1994a), and MERIS on the European Envisat.

An intrinsic element in using data from these sensors for land studies is the compositing process, which allows the production of data sets with minimum cloud contamination (Holben, 1986). The compositing process implies that neighboring pixels are typically acquired at various dates and with different acquisition geometries. The composite pixels may thus contain "artifacts"—that is, effects of variables or characteristics in the measurement process that do not represent the surface target itself.
These effects fall into two basic categories: atmospheric and geometric. The two principal atmospheric effects are 1) decrease in transmissivity due to aerosols or water vapor (and to a smaller extent gaseous absorption), and 2) residual cloud contamination—both complete obscuration of the surface within the sensor’s field of view and subpixel clouds that cannot be readily discerned but nevertheless significantly modify the surface signal. The key geometric effects are due to the nonuniform bidirectional reflectance function, which causes variations in surface reflectance as a function of the relative positions sun-target-sensor, and the variable sensor field of view, which increases nonlinearly with the increasing scan angle. The geometrical configuration may also cause differences in the local solar time, which may be important from the viewpoint of plant physiology and surface energy budget. Other effects (radiometric calibration, geometric misregistration) also may be present, although the current production systems endeavor to minimize these effects during AVHRR processing (e.g., James and Kalluri, 1994; Robertson et al., 1992).

The aforementioned atmospheric and geometric effects may be reduced through the choice of the pixel selection algorithm. For example, it has been shown that both types of effects are minimized when the maximum value of the normalized difference vegetation index (NDVI) is used as the compositing criterion (Holben, 1986). In a previous study (Cihlar et al., 1994), it was found that, although this criterion also yields artifacts such as preferential selection in the forwardscatter direction, it is comparable to other alternatives because of its robustness and a fairly balanced performance under various conditions. Thus, although other compositing criteria have been tested (Vivy et al., 1992; Qi and Kerr, 1994; Cihlar et al., 1994), it seems inevitable that the artifacts will continue to be present in future compositing data products, although, it is hoped, in significantly smaller amounts.

In AVHRR composite data, the geometric and atmospheric effects are significantly reduced when band ratios are computed, such as various vegetation indices employing combinations of channels 1 and 2 (Holben, 1986; Lee and Kaufman, 1986). However, individual channel measurements show these effects more clearly and thus hinder applications of the composite images that require the use of individual channels (e.g., Townshend et al., 1994; Hame et al., 1992; Quarmby et al., 1992).

The production of composite data sets has been reported by various authors (Blumel and Tonn, 1986; Derrien et al., 1983; Goward et al., 1994; see also Townshend, 1994). Work described here addresses the processing steps after compositing. Its purpose was to develop a set of procedures to minimize artifacts in AVHRR composites and to evaluate the effect of the corrections. The specific objective was to obtain, for each ten-day compositing period, values for surface reflectance in AVHRR channels 1 and 2, NDVI, and surface temperature, as free of residual errors (i.e., effects not representing the surface under uniform illumination and viewing conditions) as possible. Such a data set would then approximate nadir images obtained under cloud-free conditions during the growing season. The data set would portray land surface features only, without being obscured by interfering effects. The rationale for the processing and the specific steps are described in the next section.

**METHODOLOGY**

The basic initial data set consisted of 20 ten-day composites of NOAA-11 AVHRR data between 21 April and 31 October 1993 covering the Canadian landmass. The composites were prepared by using the GEOcoding and COMPosing System, or GEOCOMP (Robertson et al., 1992). GEOCOMP performs sensor calibration for the five AVHRR bands, using time-dependent gain and offset coefficients for channels 1 and 2 (Cihlar and Teillet, 1995) and on-board calibration for channels 3–5. Satellite data are registered to ground control points by high-resolution image chips and resampled using a modified Kaiser 16-point kernel. The registered images are input to the compositing process, and a compositing criterion (maximum NDVI value for the 1993 data set) is employed to select the most cloud-free pixel over the ten-day period. No view zenith (θ) restriction was used in the 1993 production; that is, all data [θ, range: −69° (backscatter) to +69° (forescatter)] were permissible candidates for selection. No atmospheric corrections were made prior to the compositing steps, because earlier tests showed that such corrections result in the preferential selection of pixels at large θ, when maximum NDVI is used as the compositing criterion (Cihlar and Huang, 1994). For each ten days, GEOCOMP produces ten composite channels of data: AVHRR radiance for bands 1–5, NDVI (in this paper, defined as the difference divided by the sum of the reflectance in channels 2 and 1), view zenith angle (θv), solar zenith angle (θs), relative azimuth angle (φ) between θv and θs, and the date of acquiring the selected pixel. All data are stored in 16 bits to ensure that there is no loss of accuracy in the processing.

As noted earlier, the initial composite data set contains various artifacts. The procedure to deal with these is dubbed “ABC3” (atmospheric, bidirectional, and contamination corrections of CCRS) and consisted of the following steps (Fig. 1), which are discussed in detail in the remainder of this section:

- **Computation of the top-of-the-atmosphere (TOA) reflectance**
- **Atmospheric correction of channels 1 and 2**
- **Identification of cloudy, partly cloudy, and snow-covered pixels to produce a contamination mask**
Figure 1. Processing sequence for AVHRR composite images leading to the production of corrected channels 1 and 2 and separately corrected channel 4 and NDVI. TOA, top of the atmosphere; SZA, solar zenith angle; BRDF, bidirectional reflectance distribution factor.

- Corrections for bidirectional reflectance effects in channels 1 and 2
- Interpolation of channels 1 and 2 data to replace “contaminated” pixels (i.e., pixels containing signals confusing the analysis of land surface characteristics)
- Corrections of the NDVI data for solar zenith angle effects
- Interpolation of NDVI data to replace contaminated pixels
- Atmospheric and surface emissivity corrections for AVHRR channel 4
- Interpolation of channel 4 data to replace contaminated pixels

Table 1 shows a comparison of ABC3 with other recent AVHRR processing strategies.

**Top-of-the-Atmosphere (TOA) Reflectance**

TOA reflectance for channel 1 or 2 is calculated from radiance, \( L^*_s \), with the following formula (Teillet, 1992):

\[
\rho^*_i = \frac{\pi d^2 L^*_s}{E_{0i}\cos\theta_i},
\]

\[
L^*_s = \frac{D_i - O_i}{G_i},
\]

where:

- \( \rho^*_i \) is the top-of-the-atmosphere reflectance (dimensionless);
- \( \theta_i \) is solar zenith angle (degrees);
- \( d \) is the Earth-Sun distance (astronomical units);
- \( E_{0i} \) is the exoatmospheric solar irradiance (Wm\(^{-2}\)μm\(^{-1}\));
- \( L^*_s \) is the radiance (Wm\(^{-2}\)sr\(^{-1}\)μm\(^{-1}\));
- \( D_i \) is the digital signal level (counts);
- \( O_i \) is the calibration offset coefficient (counts);
- \( G_i \) is the calibration gain coefficient (counts/Wm\(^{-2}\)sr\(^{-1}\)μm\(^{-1}\)).

The subscript \( i \) refers to AVHRR channel number (1 or 2).

Values of \( G_i \) and \( O_i \) were calculated with consideration of postlaunch sensor degradation (Teillet and Holben, 1994; Cihlar and Huang, 1993).

Because of the large \( \theta_i \) values encountered at high latitudes, the \( \rho^*_i \) values were often inflated by the low \( \cos\theta_i \) in Eq. (1). These values were not changed at this stage, because they were expected to be identified by the cloud-screening procedure.

**Atmospheric Correction of AVHRR Channels 1 and 2**

**Considerations**

Satellite measurements of the radiation reflected by Earth’s surface are affected by the scattering and absorption effects of the atmospheric constituents. The most important atmospheric effects are molecular (Rayleigh) and aerosol scattering and absorption by ozone and water vapor. The problems of atmospheric corrections of satellite optical data are well understood, and effective algorithms have been developed (Kneizys et al., 1988; Tanré et al., 1990). At the present time, however, rigorous corrections of medium spatial resolution data are not practical, principally owing to the lack of accurate data for input to atmospheric models. The most important among these are ozone, water vapor, aerosols, and bidirectional surface reflectance distribution (Tanré et al., 1992; Townshend et al., 1994).

In the absence of adequate pixel-specific ancillary data over large areas and on a daily basis, climatological values are often used or the effect is neglected. For example, the AVHRR Pathfinder data set corrects for Rayleigh scattering and ozone effects but not for aerosol or water vapor (Aguio, 1993). Similarly, no corrections for aerosols or water vapor are made in the global 1 km data set project (Townshend et al., 1994; Eidenshink and...
Table 1. Comparison of Recent AVHRR Processing Strategies

<table>
<thead>
<tr>
<th>Step</th>
<th>PATHFINDER (GAC)</th>
<th>FASIR (GAC)</th>
<th>IGBP data set (HRPT)</th>
<th>ABC3 (HRPT)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C1, C2, C3, C4, C5</td>
<td>C1, C2</td>
<td>C1, C2, C3, C4, C5</td>
<td>C1, C2, C3, C4, C5</td>
</tr>
<tr>
<td>Registration, resampling</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Compositing</td>
<td>Yes (8-11 days)</td>
<td>Yes (28-31 days)</td>
<td>Yes (8-11 days)</td>
<td>Yes (8-11 days)</td>
</tr>
<tr>
<td>Converting C1, C2 into TOA reflectance</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Atmospheric correction</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>C1, C2, C4</td>
</tr>
<tr>
<td>Surface emissivity correction</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>C4</td>
</tr>
<tr>
<td>Coarse cloud mask</td>
<td>Obtained, not applied</td>
<td>Yes*</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Identify subpixel contamination</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>BRDF corrections</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>C1, C2</td>
</tr>
<tr>
<td>Replacement of contaminated data</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>C1, C2, C4</td>
</tr>
</tbody>
</table>

2 Los et al. (1994).
3 Eldershink and Faundeen (1994).
4 This paper.
5 Approximate correction for sensor degradation (applied to NDVI).
6 CLAVR algorithm.
7 Constant C4 threshold.

Faundeen, 1994). Other simplifications are often introduced to make operational use of atmospheric correction practicable. Teillet (1992) designed a look-up table approach based on precomputation of intermediate variables needed for pixel-based corrections and a data flow similar to that used for geometric data corrections. More recently, Rahman and Dedieu (1994) developed a simplified version of the 5S algorithm (Tanré et al., 1990) for use with composite images. The algorithm, called SMAC (simplified method for atmospheric correction), is much faster than more-detailed radiative transfer models because it uses semiempirical formulations and coefficients that depend on the sensor/spectral band of interest. It accounts for two-way gaseous transmission and for Rayleigh as well as aerosol scattering. Tests by Rahman and Dedieu (1994) showed that the errors introduced by the parameterization are small.

Implementation

SMAC was employed in the processing reported here. The program requires vertically integrated gaseous contents, $\theta$, $\theta$, and $\phi$, and the value of aerosol optical depth at 550 nm for each pixel. The pixel-based angular information is provided in the basic GEOCOMP data set. The processing was carried out with assumed water-vapor content of 2.3 g/cm$^2$ and ozone content of 0.319 cm-atm. In the absence of actual data for optical depth at 550 nm, a constant value of 0.05 was used (Gerard Dedieu, 1994, personal communication, LERTS Toulouse). This value has also been recommended by Teillet (1992) for clear sky conditions, based on extensive atmospheric transmission measurements during the growing season in eastern Canada (Ahern et al., 1991). Because most pixels to be corrected should represent clear sky conditions, the aerosol content at middle to northern latitudes over land can be expected to be low a significant percentage of the time. A major aerosol increase occurs during forest fires, but then its spatial variability is very large and no specific information on the concentrations is available. Thus, the ABC3 procedure currently does not account for the effects of varying aerosols, except in extreme cases when the aerosol effect starts to approximate that of clouds and the CECANT algorithm detects it (see below). It should be noted that the accuracy of 5S decreases if $\theta$, and $\theta$, are above 60° and 50°, respectively, or if the optical depth is above 0.8 (Rahman and Dedieu, 1994). SMAC was applied sequentially to all 20 composites. The processing resulted in surface reflectance for AVHRR channels 1 ($\rho_1$) and 2 ($\rho_2$).

From the foregoing description, it is evident that the atmospheric correction is a "nominal" one. The optimum correction would have to be based on all parameters being specific to the actual AVHRR pixel's acquisition time and atmospheric composition, not just pixel geometry. No such atmospheric composition data are currently available anywhere. The complete corrections are thus not feasible at present, but it is hoped that they will become so with data from the Earth Observing System (EOS).

Identification of Contaminated Pixels

Principle

Because the goal of the processing was to produce a data set for the study of land biosphere, pixels were considered to be contaminated if they contained targets other than vegetation or bare soil, rock, or open water. Sources of contamination were thus clouds, partial (subpixel) clouds, cloud shadows, smoke or other heavy aerosols, snow, or ice.

A new procedure was developed to identify such contaminated pixels (Cihlar, 1996). The procedure, dubbed CECANT (cloud elimination from composites
using albedo and NDVI trend), is based on the high sensitivity of NDVI to the presence of clouds (clouds themselves and shadows from optically thick clouds that cause a "large enough" decrease of the surface NDVI compared with the NDVI of the same surface under clear sky), aerosol, and snow. Three features of the annual surface reflectance trend are used: the high contrast between the albedo (represented by AVHRR channel 1) of land, especially when fully covered by green vegetation, and clouds or snow/ice; the average NDVI value (expected value for that pixel and period based on the measured data); and the monotonic trend in NDVI. The procedure builds on algorithms developed by Sellers et al. (1994) and Cihlar and Howarth (1994), respectively, and requires values of $R$ and $Z$ to be computed for each pixel $(i,j)$ and compositing period $t$:

$$R(i,j,t) = \frac{\text{NDVI}(i,j,t) - \text{Median}_{i,j}}{\text{NDVI}_{\text{avg}}(i,j,t)}$$

$$Z(i,j,t) = \frac{\text{NDVI}_{\text{med}}(i,j,t) - \text{NDVI}(i,j,t)}{\text{NDVI}_{\text{med}}(i,j,t)}$$

where $M(i,j)$ is the median for the set $\{\text{NDVI}(i,j,t) - \text{NDVI}_{\text{med}}(i,j,t)\}$ for all values $t$; $\text{NDVI}$ is the average curve fitted to measure NDVI values; and $\text{NDVI}_{\text{med}}(i,j,t)$ is the value of the upper envelope for the NDVI values of pixel $(i,j)$.

$R$ thus measures the deviation of $\text{NDVI}(i,j,t)$ from the expected average value in units of "average scatter" about the seasonal curve. The median is chosen because it is less sensitive to large deviations from the average curve than other measures of central tendency (Los et al., 1994). Conversely, $Z$ tests for the monotonic behavior of the seasonal NDVI profile. It implies that a decrease in NDVI below the expected value by the monotonic trend is caused by pixel contamination.

Although $R$ and $Z$ respond to short-term pixel contamination, they are not likely to detect all cloud and snow/ice pixels because these pixels may have consistently high albedo and therefore low $R$ and $Z$. An albedo threshold is therefore used, represented in CECANT by AVHRR channel 1 ($\rho_a = C1$). The three parameters ($C1$, $R$, $Z$) are used in combination to separate pixels with a relatively high albedo that could Nevertheless be plausibly caused by intrinsically high surface reflectance, from pixels with a similar albedo caused by a mixture of a dark surface and cloud/snow/ice.

**Implementation**

Four thresholds are required in CECANT to identify partially contaminated pixels (Cihlar, 1996):

$C1$: the maximum channel 1 reflectance of a clear-sky, snow- or ice-free land pixel in the data set

$R_{max}$: the maximum acceptable deviation of the measured value $\text{NDVI}(i,j,t)$ below the estimated $\text{NDVI}_{\text{med}}(i,j,t)$

$Z_{max}$: the maximum acceptable deviation of the measured value $\text{NDVI}(i,j,t)$ above the estimated $\text{NDVI}_{\text{med}}(i,j,t)$

$M(i,j)$, $\text{NDVI}_{\text{med}}(i,j,t)$, and $\text{NDVI}(i,j,t)$ were calculated by using the FASIR model (Sellers et al., 1994), which approximates the seasonal NDVI curve with a third-order Fourier transform. Before the NDVI computation, missing values between first and last measurements were replaced through linear interpolation after finding the seasonal peak for each pixel, using the rationale and algorithm of Cihlar and Howarth (1994), to minimize the effect of the missing values on NDVI estimation; and the temporal series was extended by adding two time periods with $\text{NDVI}=0$ at each end of the seasonal curve to better anchor the fitted curve. Using the 1993 data set, it was found that the identification of contaminated pixels was not highly sensitive to the $C1$ threshold and a constant value of $\rho_a = 0.30$ was therefore used. The upper and lower limits for $R$ and $Z$ were determined separately for each composite period, using mean values of $R$ and $Z$ histograms. The histograms were prepared for all Canada after subsampling the 1 km data to retain every sixth line and sixth pixel [refer to Cihlar (1996) for details]. Given the thresholds channel 1, $R(t)$, and $Z(t)$, a contamination mask was prepared for each date. Figure 2 shows the threshold $R$ and $Z$ values for the individual periods.

$R$ and $Z$ are sensitive to the accuracy of estimating NDVI, and $R_{max}$, although the procedure does allow compensation for systematic bias in the estimation of $R$ and $Z$ in different periods. The Fourier series approach seems to provide an effective method for obtaining these
NDVI values, especially for pixels with one distinct seasonal NDVI peak. It performs less satisfactorily for broad peaks, such as coniferous forest, and further improvement in this case is needed. It should also be noted that the foregoing procedure eliminates pixels with permanent snow or ice coverage, and the resulting data set is thus not appropriate for cryospheric studies.

**Corrections for Bidirectional Reflectance Effects in Channels 1 and 2**

**Considerations**

Given the wide range of $\theta$, values in the composite data set, bidirectional variations have a significant effect on the data (Hoben and Fraser, 1984; Gallo and Eiden-shink, 1988; Gutman et al., 1989; Cihlar et al., 1994; Wu et al., 1995; Li et al., 1996). The knowledge and practice of routine corrections of satellite data for these effects are not yet well developed, and few attempts have been made at bidirectional reflectance distribution factor (BRDF) corrections of large data sets. Much of the current research is aimed at future sensors where it is hoped that the BRDF will be sampled in sufficient detail during a few days so that a bidirectional model may be inverted (Strahler, 1994; Rahman et al., 1993). Such an approach would not be applicable to current data sets, because few clear-sky pixels are available during ten-day periods at northern latitudes where frequent clouds are present during the summer months (Cihlar and Howarth, 1994). For bidirectional corrections over large areas, empirical approaches are more appropriate at the present (Gutman et al., 1989; Gutman, 1992, 1994). Various formulations have been used, from a second-order polynomial (Gutman, 1992) to a semiparametric model (Roujean et al., 1992; Wu et al., 1995). Cihlar et al. (1994) found that the model of Walthall et al. (1985) provided a good fit to AVHRR data from central Canada obtained in midsummer of 1988. However, in initial experiments, the model did not perform well with the 1993 data, possibly because it does not readily accommodate the changing vegetation structure during the season. We found that the model of Roujean et al. (1992) as employed by Wu et al. (1995), performed more satisfactorily for this data set. The original model uses leaf area index as a measure of changing canopy structure. Wu et al. (1995) parameterized this dependence through NDVI using AVHRR data from lower latitudes in North America. When testing its performance over Canada, we found (Li et al., 1996) that the model substantially lessened, but did not completely eliminate, $\theta$, dependence of reflectance. To remove such a dependence as much as possible, the coefficients of the Wu et al. model were retuned, using the data set for Canada.

**Implementation**

The model of Roujean et al. (1992) is given by:

$$
\rho_{i}(\theta, \phi) = K_0 + K_1 f_1(\theta, \phi) + K_2 f_2(\theta, \phi),
$$

where $\rho_{i}$ denotes the reflectance in channel 1 ($i=1$) and channel 2 ($i=2$). $f_1$ and $f_2$ are functions of three angles: solar zenith $\theta_s$, viewing zenith $\theta$, and relative azimuth $\phi$. $f_1$ gives the bidirectional dependence due to diffuse reflection from opaque reflectors on the ground and to the shadowing effects of the leaves. $f_2$ accounts for the contribution of the volume scattering by a collection of randomly dispersed facets of canopies and bare soils. $K_0$ represents the bidirectional reflectance for $\theta=\phi=0$, $K_1$ and $K_2$ are coefficients reflecting the relative contributions of $f_1$ and $f_2$. They were fitted as functions of NDVI, using clear-sky AVHRR TOA reflectance obtained in the conterminous United States and parts of Canada (Wu et al., 1995).

A new set of values for $K_0$ and $K_1$ was obtained, using atmospherically corrected AVHRR composite data. Retuning was done by retaining the same functional dependences on NDVI as in Wu et al. (1995) but adjusting their coefficients. Similar to Table 2 of Wu et al. (1995), the new coefficients were tuned by using the following formats differentiated by four basic land cover types:

**Channel 1**

<table>
<thead>
<tr>
<th></th>
<th>$K_0$</th>
<th>$K_1$</th>
<th>$K_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>barren</td>
<td>$C_0$</td>
<td>$C_1$</td>
<td>$C_2$</td>
</tr>
<tr>
<td>cropland</td>
<td>$C_0$</td>
<td>$C_1$</td>
<td>$C_2$</td>
</tr>
<tr>
<td>forest</td>
<td>$C_0$</td>
<td>$C_1$</td>
<td>$C_2$</td>
</tr>
<tr>
<td>grassland</td>
<td>$C_0$</td>
<td>$C_1$</td>
<td>$C_2$</td>
</tr>
</tbody>
</table>

**Channel 2**

<table>
<thead>
<tr>
<th></th>
<th>$K_0$</th>
<th>$K_1$</th>
<th>$K_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>barren</td>
<td>$C_0$</td>
<td>$C_1$</td>
<td>$C_2$</td>
</tr>
<tr>
<td>cropland</td>
<td>$C_0$</td>
<td>$C_1$</td>
<td>$C_2$</td>
</tr>
<tr>
<td>forest</td>
<td>$C_0$</td>
<td>$C_1$</td>
<td>$C_2$</td>
</tr>
<tr>
<td>grassland</td>
<td>$C_0$</td>
<td>$C_1$</td>
<td>$C_2$</td>
</tr>
</tbody>
</table>

The coefficients $C_0$, $C_1$, $C_2$, and $C_3$ were determined by means of multiple linear regression, using sample pixels from clear-sky AVHRR composite data after the atmospheric correction. To reduce the data volume and the correlation among the data (Li and Leighton, 1992), a 10% random sample was obtained from the entire data set (April to October 1993) for each cover type. The data were eroded by one pixel (i.e., the pixels at the boundary of all land-cover polygons were excluded) prior to the selection of pixels for deriving BRDF models to minimize misregistration errors. Although the four land-cover types listed represent the most distinct bidirectional characteristics, differences in bidirectional reflectance distribution exist among the subclasses of land cover, such as different forests (Li et al., 1996). Therefore, the regression was carried out for each of the nine land-cover types, on the basis of the land-cover map of Canada derived from 1-km AVHRR data (Pokrant, 1991).

Table 2 shows values of the coefficients $C_0$, $C_1$, $C_2$, $C_3$, and $C_4$ for nine land-cover types in two channels. With these coefficients, one can normalize Eq. (5) by dividing $K_i (C_i$ in Table 2). The resulting function can be
Table 2. New Coefficients for the Bidirectional Reflectance Model for Various Land Cover Types (Wu et al., 1995)

<table>
<thead>
<tr>
<th>Land Cover</th>
<th>$C_0$</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>$C_3$</th>
<th>$C_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel 1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed wood</td>
<td>0.05626</td>
<td>0.00341</td>
<td>0.20925</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deciduous</td>
<td>0.05545</td>
<td>0.00240</td>
<td>0.24120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitional</td>
<td>0.06084</td>
<td>0.00622</td>
<td>0.28591</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coniferous</td>
<td>0.05162</td>
<td>0.00467</td>
<td>0.23477</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tundra</td>
<td>0.08440</td>
<td>0.18325</td>
<td>0.90561</td>
<td>-3.18811</td>
<td>3.72133</td>
</tr>
<tr>
<td>Barren</td>
<td>0.12545</td>
<td>0.00940</td>
<td>0.22983</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cropland</td>
<td>0.09090</td>
<td>0.00685</td>
<td>0.47905</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rangeland</td>
<td>0.08356</td>
<td>0.18757</td>
<td>0.71980</td>
<td>-1.92178</td>
<td>1.96764</td>
</tr>
<tr>
<td>Built-up</td>
<td>0.10531</td>
<td>0.00876</td>
<td>0.31200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Channel 2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed wood</td>
<td>0.24190</td>
<td>0.04897</td>
<td>0.29825</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deciduous</td>
<td>0.24380</td>
<td>0.06237</td>
<td>0.25644</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitional</td>
<td>0.16472</td>
<td>0.00027</td>
<td>0.30519</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coniferous</td>
<td>0.16211</td>
<td>0.01173</td>
<td>0.31275</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tundra</td>
<td>0.17830</td>
<td>0.13877</td>
<td>0.45717</td>
<td>-0.9908</td>
<td>1.02401</td>
</tr>
<tr>
<td>Barren</td>
<td>0.16661</td>
<td>0.03034</td>
<td>0.34747</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cropland</td>
<td>0.28780</td>
<td>0.03106</td>
<td>0.43316</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rangeland</td>
<td>0.23094</td>
<td>0.31250</td>
<td>-0.47362</td>
<td>4.48017</td>
<td>-4.58914</td>
</tr>
<tr>
<td>Built-up</td>
<td>0.25967</td>
<td>0.03542</td>
<td>0.33321</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

used to normalize a measurement made for a specific illumination and viewing geometry to a common sun-target-satellite geometry and to compute an albedo defined over the upper hemisphere from the normalized reflectance (Wu et al., 1995). [Note that there is a typographic error in the equation for computing albedo in Wu et al. (1995); the correct version is given in Li et al. (1996)]. The normalized reflectances are particularly useful for monitoring the temporal evolution of land surfaces because the dependence on the illumination and viewing geometry are removed or alleviated. Hence, changes in normalized reflectance are due primarily to the variations of surface properties, provided that the atmospheric effects are completely removed. In accord with procedure, a pair of channel 1 and 2 images were obtained with a constant viewing geometry ($\theta_i=0^\circ$, $\theta_v=45^\circ$) for each of the 1993 compositing periods.

The overall efficiency of the correction can be evaluated by following the methodology proposed by Li et al. (1996). First, the original and corrected reflectances are fitted with polynomials against view zenith with its sign determined by relative azimuth. The nonflatness in each polynomial indicates the degree of bidirectional dependence, without being confused by natural reflectance variation within each cover type. It is quantified by computing the absolute mean deviation of the polynomial curve from its mean value across the range of view zenith angles. The ratio of the deviation against the mean reflectance value indicates the relative variation resulting from the bidirectional dependence.

The resulting values for surface reflectance before and after the BRDF corrections are given in Tables 3a and 3b. It is seen that the relative changes associated with bidirectional dependence are reduced by 33% to 66%. The mean improvement in surface reflectance was determined by using the fraction of land cover occupied by each cover type according to the data of Pokranc (1991). Overall, 50.7% (channel 1) and 61.5% (channel 2) of the BRDF-dependent variability was removed by the corrections described above. It should be noted that, especially for channel 1, the reflectance values are low (refer to Fig. 10) and small errors in estimation may have significant effect.

Replacement of Contaminated Pixels for AVHRR Channels 1 and 2

Considerations

Prior to this step, contaminated pixels were identified, but no corrections were made. To obtain the final channel data in each compositing period, new values were needed for the contaminated pixels. Two cases were recognized, during (i.e., between the first and last clear-sky measurements) and at the end (before the first or after the last clear-sky measurement) of the growing season.

For pixels during the growing season, it was assumed that no rapid changes would occur between adjacent ten-day periods, and the missing values could therefore be interpolated by using the seasonal trajectory for that pixel/channel. We tested both linear and cubic spline interpolation routines and found that, though cubic spline could better adjust to a local trend, it tended to exaggerate temporal fluctuations of the signal (e.g., due to inaccurate BRDF correction). Linear interpolation appeared
Table 3a. Improvements of Estimated Surface Reflectance due to the BRDF Corrections: Channel 1

<table>
<thead>
<tr>
<th>Cover-Type No.</th>
<th>Fraction</th>
<th>Y mean</th>
<th>dY</th>
<th>dmY</th>
<th>Y mean</th>
<th>dY</th>
<th>dmY</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>0.044</td>
<td>0.073</td>
<td>0.023</td>
<td>0.307</td>
<td>0.054</td>
<td>0.066</td>
<td>0.107</td>
<td>65.689</td>
</tr>
<tr>
<td>2.000</td>
<td>0.010</td>
<td>0.073</td>
<td>0.023</td>
<td>0.302</td>
<td>0.052</td>
<td>0.066</td>
<td>0.119</td>
<td>60.525</td>
</tr>
<tr>
<td>3.000</td>
<td>0.114</td>
<td>0.096</td>
<td>0.027</td>
<td>0.285</td>
<td>0.065</td>
<td>0.090</td>
<td>0.136</td>
<td>52.345</td>
</tr>
<tr>
<td>4.000</td>
<td>0.139</td>
<td>0.073</td>
<td>0.022</td>
<td>0.287</td>
<td>0.051</td>
<td>0.066</td>
<td>0.105</td>
<td>64.766</td>
</tr>
<tr>
<td>5.000</td>
<td>0.182</td>
<td>0.135</td>
<td>0.042</td>
<td>0.110</td>
<td>0.080</td>
<td>0.017</td>
<td>0.166</td>
<td>39.845</td>
</tr>
<tr>
<td>6.000</td>
<td>0.215</td>
<td>0.135</td>
<td>0.040</td>
<td>0.293</td>
<td>0.110</td>
<td>0.021</td>
<td>0.196</td>
<td>33.136</td>
</tr>
<tr>
<td>7.000</td>
<td>0.242</td>
<td>0.115</td>
<td>0.029</td>
<td>0.249</td>
<td>0.095</td>
<td>0.009</td>
<td>0.101</td>
<td>39.512</td>
</tr>
<tr>
<td>8.000</td>
<td>0.028</td>
<td>0.116</td>
<td>0.039</td>
<td>0.333</td>
<td>0.080</td>
<td>0.008</td>
<td>0.094</td>
<td>71.673</td>
</tr>
<tr>
<td>9.000</td>
<td>0.003</td>
<td>0.050</td>
<td>0.033</td>
<td>0.246</td>
<td>0.093</td>
<td>0.016</td>
<td>0.169</td>
<td>49.761</td>
</tr>
<tr>
<td>Weighted improve</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30.732</td>
</tr>
</tbody>
</table>

Table 3b. Improvements of Estimated Surface Reflectance due to the BRDF Corrections: Channel 2

<table>
<thead>
<tr>
<th>Cover-Type No.</th>
<th>Y mean</th>
<th>dY</th>
<th>dmY</th>
<th>Y mean</th>
<th>dY</th>
<th>dmY</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>0.223</td>
<td>0.064</td>
<td>0.286</td>
<td>0.202</td>
<td>0.019</td>
<td>0.096</td>
<td>66.432</td>
</tr>
<tr>
<td>2.000</td>
<td>0.251</td>
<td>0.067</td>
<td>0.266</td>
<td>0.238</td>
<td>0.023</td>
<td>0.098</td>
<td>63.293</td>
</tr>
<tr>
<td>3.000</td>
<td>0.195</td>
<td>0.048</td>
<td>0.247</td>
<td>0.154</td>
<td>0.014</td>
<td>0.089</td>
<td>64.297</td>
</tr>
<tr>
<td>4.000</td>
<td>0.186</td>
<td>0.046</td>
<td>0.248</td>
<td>0.151</td>
<td>0.010</td>
<td>0.064</td>
<td>74.142</td>
</tr>
<tr>
<td>5.000</td>
<td>0.218</td>
<td>0.049</td>
<td>0.226</td>
<td>0.174</td>
<td>0.020</td>
<td>0.117</td>
<td>48.334</td>
</tr>
<tr>
<td>6.000</td>
<td>0.206</td>
<td>0.052</td>
<td>0.249</td>
<td>0.184</td>
<td>0.019</td>
<td>0.119</td>
<td>52.349</td>
</tr>
<tr>
<td>7.000</td>
<td>0.291</td>
<td>0.065</td>
<td>0.225</td>
<td>0.259</td>
<td>0.018</td>
<td>0.069</td>
<td>69.450</td>
</tr>
<tr>
<td>8.000</td>
<td>0.278</td>
<td>0.069</td>
<td>0.249</td>
<td>0.225</td>
<td>0.023</td>
<td>0.101</td>
<td>59.411</td>
</tr>
<tr>
<td>9.000</td>
<td>0.256</td>
<td>0.050</td>
<td>0.232</td>
<td>0.227</td>
<td>0.026</td>
<td>0.116</td>
<td>49.949</td>
</tr>
<tr>
<td>Weighted improve</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>61.490</td>
</tr>
</tbody>
</table>

Ym = (Y_1 + Y_2 + … + Y_n)/n, df = (1 - 1/n) * Σ(Y - Y_m)(dY - dmY), where Y is the estimated surface reflectance by a fourth-order polynomial fitted to a random sample of pixels selected for the cover type from across Canada for view zenith ranging between -70° and 70°; n=141, and dmY is the amount of BRDF-related variability, (as a fraction of the mean reflectance).

a viable alternative, although it tends to smooth out local variations in the temporal trajectory.

A different problem occurred at the beginning and the end of the growing season where pixels were often missing because of snow cover. Taking advantage of the cyclical behavior of reflected radiation measured by AVHRR at northern latitudes, we assumed that the annual trajectory for individual channels as well as for NDVI could be approximated by a second-degree polynomial. This assumption is based on the fact that the second-degree polynomial provides a good approximation of the seasonal cycle of this process.

Implementation
The pixels under the contaminated mask were replaced through interpolation, provided that at least three uncontaminated values were available. During the growing season (i.e., between the first and last clear-sky measurements), the new values were found through linear interpolation for channels 1 and 2. A second-order polynomial form was used to calculate values for the end of the growing season. In this case, the polynomial was fitted to the plot p(x,t) for all clear-sky (i.e., contaminated) periods t, starting with the first clear-sky composite period after 1 August 1993. This step was used to ensure that the polynomial could be anchored by several measured values at each end. Interpolation was done only if at least three uncontaminated pixels were available (this condition was not met in some cases in the far north). After determining the best-fit coefficients for the polynomial p_{x,t}(x,t) and p_{y,t}(x,t), the new values p_{x,t}(x,y,t) and p_{y,t}(x,y,t) were calculated by using the polynomial coefficients to replace contaminated pixels prior to the first clear pixel or after the last such pixel. For this calculation, only clear-sky measurements were plotted on the x-axis (time), which spanned all of 1993.

**NDVI Processing**
It would be desirable to compute the NDVI from BRDF-corrected channels 1 and 2. This should lessen the effect of viewing geometry on NDVI (Li et al., 1996). However, because the difference in bidirectional dependence between channel 1 and channel 2 may not be significantly larger than the uncertainties of the BRDF correction, we decided to retain the atmospherically corrected NDVI (after the step described in the section on atmospheric correction of AVHRR channels 1 and 2). This is supported by previous results (Holben, 1986; Lee
Table 4: Coefficients for the Solar Zenith Angle Correction of NDVI

<table>
<thead>
<tr>
<th>Type</th>
<th>SIB2 type*</th>
<th>NDVI-98</th>
<th>NDVI-5</th>
<th>k1</th>
<th>k2</th>
<th>k3</th>
<th>k4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coniferous</td>
<td>Needle-leaf evergreen trees</td>
<td>0.696</td>
<td>0.039</td>
<td>0.19</td>
<td>1.18</td>
<td>0.52</td>
<td>1.04</td>
</tr>
<tr>
<td>Deciduous</td>
<td>Broadleaf deciduous trees</td>
<td>0.721</td>
<td>0.039</td>
<td>0.32</td>
<td>1.38</td>
<td>0.52</td>
<td>1.04</td>
</tr>
<tr>
<td>Mixed</td>
<td>Broadleaf and needle-leaf trees</td>
<td>0.721</td>
<td>0.039</td>
<td>0.28</td>
<td>1.35</td>
<td>0.52</td>
<td>1.04</td>
</tr>
<tr>
<td>Transitional forest</td>
<td>Needle-leaf evergreen trees</td>
<td>0.696</td>
<td>0.039</td>
<td>0.19</td>
<td>1.18</td>
<td>0.52</td>
<td>1.04</td>
</tr>
<tr>
<td>Cropland</td>
<td>Agriculture or C4 grassland</td>
<td>0.674</td>
<td>0.039</td>
<td>0.38</td>
<td>1.45</td>
<td>0.52</td>
<td>1.04</td>
</tr>
<tr>
<td>Rangeland</td>
<td>C4 grassland</td>
<td>0.611</td>
<td>0.039</td>
<td>0.15</td>
<td>2.80</td>
<td>0.52</td>
<td>1.04</td>
</tr>
<tr>
<td>Tundra</td>
<td>Short vegetation</td>
<td>0.674</td>
<td>0.039</td>
<td>0.38</td>
<td>1.45</td>
<td>0.52</td>
<td>1.04</td>
</tr>
<tr>
<td>Barren</td>
<td>Short vegetation</td>
<td>0.674</td>
<td>0.039</td>
<td>0.38</td>
<td>1.45</td>
<td>0.52</td>
<td>1.04</td>
</tr>
<tr>
<td>Built-up</td>
<td>Short vegetation</td>
<td>0.674</td>
<td>0.039</td>
<td>0.38</td>
<td>1.45</td>
<td>0.52</td>
<td>1.04</td>
</tr>
</tbody>
</table>

From Sellers et al. (1994): the first column shows cover types used in this study (Pokrant, 1991). NDVI 98 and NDVI 5 are the 98 and 5 percentiles, respectively; k are coefficients for Eqns. (6)-(8).

and Kaufman, 1986; Cihlar et al., 1994) that showed that the θ dependence for atmospherically corrected NDVI may be relatively weak. However, corrections for θ were desirable in view of the known dependence of the NDVI on θ (Holben et al., 1986; Lee and Kaufman, 1986). Another advantage of the direct NDVI processing as carried out here is that interpolation for missing values can be made for NDVI more reliably than for channels 1 and 2, especially where very few uncorrected pixels are available (e.g., in the far north). As a result, two NDVI data sets are obtained, corrected through different procedures.

Sun Angle Correction

Considerations. Sellers et al. (1994) derived a set of correction coefficients for the AVHRR Global Area Coverage (GAC) data set as part of their FASIR procedure. They identified high and low NDVI limits for four vegetation categories (grass/crops, broadleaf deciduous forests, evergreen forests, and coniferous forests) and θ, based on a global data set between latitudes 20° and 70°. After combining similar data for the Northern and Southern Hemispheres, they determined 98 NDVI percentiles and fitted a function of θ, to these values for each category. They used a similar approach to obtain θ dependence for bare soil, assuming that 5% values of the cumulative NDVI distribution for bare soil, shrubs, and desert represent no-vegetation conditions. The resulting coefficients are shown in Table 4.

Although the data have different spatial resolution from those of interest in this paper, the coefficients are considered applicable because they are based on data for the same sensor and for the entire globe. A partial check of the coefficients was made for coniferous forest and cropland, using two composites from the 1993 data set. The pixels for each cover type were selected by using the digital land-cover map of Pokrant (1991). For each AVHRR composite and cover type, cumulative histograms were produced for the following θ ranges: 30–40°, 40–50°, 50–60°, and 60–70°. The 2 and 98 percentiles were determined from these histograms. Figure 3 shows the results of this analysis as well as comparable values for the global data set of Sellers et al. (1994). For conifer, the two sets of thresholds correspond well where the 1993 data set had a good spatial distribution of pixels. For example, the 45° and 55° histograms represent...
pixels from across Canada and match the values computed with coefficients of Sellers et al. well, whereas most of the July, 35° data are from western Canada and differ the most. Most of the 65° data are from eastern Canada. For cropland (Fig. 3b), the agreement was good at low $\theta_0$ for the 98 percentile, but the decrease with $\theta_0$ was less pronounced than for FASIR (the 65° group had only 383 pixels, mostly in western Canada). In Figures 3a and 3b, the 35° data have somewhat lower NDVI than do those at 45°, which would not normally be expected. This is again related to the geographic distribution of the pixels included, because those farther south (drier region in central Canada) tend to have lower NDVI. The 2-per-

centile values were much higher in most cases than those for FASIR, probably owing to the fact that the 1993 data used represent peak green period. Given the generally good correspondence between the FASIR and the 1993 thresholds (where a spatially representative data set was available) and the need to correct data from various parts of the growing season, the FASIR procedure was employed to correct all 1993 clear-sky pixels for sun angle variations.

Implementation
To apply the $\theta_0$ corrections, a correspondence table was established between the land-cover types of Pokrant et al. (1991) and those of Sellers et al. (1994), as shown in Table 4. The new set of NDVI values was then computed for a reference $\theta_0=45°$ based on the equations of Sellers et al. (1994):

$$\text{NDVI}_{45°} = \frac{(\text{NDVI}_{95°} - \text{NDVI}_{45°}) \times (\text{NDVI}_{45°} - \text{NDVI}_{35°})}{\text{NDVI}_{35°} - \text{NDVI}_{35°}} + \text{NDVI}_{35°},$$

$$\text{NDVI}_{35°} = \text{NDVI}_{35°} \times \left(1 - k_1 \left(\frac{\theta_0 \pi}{180} - \frac{\pi}{6}\right)^2\right),$$

$$\text{NDVI}_{35°} = \text{NDVI}_{35°} \times \left(1 + k_1 \left(\frac{\theta_0 \pi}{180} - \frac{\pi}{6}\right)^2\right),$$

where:

$\text{NDVI}_{35°}$ is the normalized value for $\theta_0=45°$;
$\text{NDVI}_{35°}$ is the measured value;
$\text{NDVI}_{35°}$, $\text{NDVI}_{35°}$, $\text{NDVI}_{35°}$, and $\text{NDVI}_{35°}$ are calculated from the Eq. (6)-(8);
$\text{NDVI}_{35°}$, $\text{NDVI}_{35°}$, $k_1$, $k_2$, $k_3$, and $k_4$ are constants.

Eqs. (6)-(8) were applied to all land pixels of each composite.

NDVI Replacement in Contaminated Pixels
Considerations. Assuming that no rapid changes occur between adjacent compositing periods, the missing values could be interpolated by using cubic spline or a similar algorithm. Because NDVI curves in this geographic region are typically bell shaped, the interpolation should preserve the concave shape as much as possible. This requirement would favor more-advanced interpolation routines. We have experimented with various approaches: linear interpolation, cubic spline interpolation, and the use of the Fourier series estimate $\text{NDVI}_{\text{asf}}(i,j,t)$. In general, the more-advanced algorithms performed better than did linear interpolation. However, they may introduce significant errors, depending on the configuration of the adjacent values. Linear interpolation does not suffer from this shortcoming, but it tends to underestimate (overestimate) the expected NDVI value when the local shape of the NDVI curve is concave (convex). For the analysis reported here, we therefore employed a combination of the three methods.

Implementation
For the Fourier estimate, we employed the knowledge of the idealized NDVI seasonal curve $\text{NDVI}_{\text{asf}}(i,j,t)$ computed by the FASIR procedure (Sellers et al., 1994). Briefly, this idealized curve results from first identifying pixels/periods below the average seasonal NDVI trajectory and then discounting these pixels in proportion to the degree of likely contamination [measured by $\text{R}(i,j,t)$, Eq. (3)] to produce a “best-fit” seasonal NDVI curve without cloud contamination. The detailed procedure for obtaining the $\text{NDVI}_{\text{asf}}(i,j,t)$ values for the 1993 data set was described by Cihlar (1996). Standard computing methods were used for cubic spline and linear interpolations.

For the final NDVI data set, the NDVI surface $\theta_0$-corrected value (see the section on sun angle correction) was used for uncontaminated pixels. If contaminated, the three interpolated values were compared, the two extreme values were discarded, and the intermediate value was retained. It was found that, in most cases, the linear interpolation was chosen. As Figure 4 shows, in >95% of cases, linear interpolation provided the intermediate estimates. Linear interpolation was chosen most often near the ends of the growing season, possibly because the NDVI trajectories are typically flat at that time. Linear interpolation was also the most frequent choice in the middle of the growing season, although here the other interpolations were chosen more often. Cubic spline was a distant second choice, although the difference between Fourier and cubic spline interpolations was relatively small. Although the nonlinear interpolations offer the possibility of a better approximation of a rapid temporal change, they are also more susceptible to the noise in the NDVI measurements. This may result in introducing errors in the interpolated data. An additional threshold was therefore used to account for the imperfections in the interpolation procedure by setting the maximum NDVI value to 0.85; this was based on the inspection of NDVI histograms, which showed smaller values in the vast majority of cases. In the future, using linear interpolations may be the preferred conservative choice.
Channel 4 Correction

The goal of the processing of thermal channels was to obtain surface temperature information. Surface temperature is an important environmental parameter that can serve as a source of insight into the surface radiation budget and the physiological vegetation processes (Nemani et al., 1993; Goward et al., 1994). It may also be a useful additional attribute for land-cover mapping (Running et al., 1994b; Zhu and Evans, 1992).

The extraction of precise surface temperature information from AVHRR composite images is complex because of atmospheric attenuation, surface emissivity effects, and possibly the differences in the local solar time for the composite pixels acquired on different days. Cloud contamination plays a major role at northern latitudes (Cihlar, 1987). As a result, little work has been carried out over large areas, and the present study appears to be the first attempt to produce a seasonal land surface temperature data set over a large area.

Considerations

Atmospheric attenuation in the thermal infrared channels (AVHRR channels 4 and 5) is dominated by water vapor in the atmospheric column (Price, 1983), but haze effect can also be significant (McClelland et al., 1971). Comprehensive radiative transfer solutions to the total attenuation have been formulated (Kneizys et al., 1988) that usually require atmospheric sounding data to obtain sufficient accuracies. Other investigators focused on semiempirical approaches, among which is the split-window method (Price, 1983; 1984; Rustana et al., 1992; Coll et al., 1994). It is based on the differential sensitivity of channels 4 and 5 to the atmospheric water vapor (McClain et al., 1985). The split-window approach also accounts for the sun angle dependence of the thermal signal (Sobrino et al., 1991).

Cooper and Asrar (1989) compared various methods for retrieving surface temperature over a tall-grass prairie and found that Price’s (1984) model performed well in most cases, meeting their criteria in seven of eight cases. It should be noted that the accuracy of the split-window method depends, among other factors, on the magnitude of the surface emissivity difference between AVHRR channels 4 and 5. Because there usually is such a difference over land, the split-window approach provides an imperfect approximation of the land surface temperature. More sophisticated approaches have been proposed, making use of modeling, additional spectral channels, day and night data, and so forth (Becker and Li, 1990; Sobrino et al., 1991; Li, 1990; Li and Becker, 1992). However, these approaches require data that are not readily available at northern latitudes. Recently, Coll et al. (1994) described a semiempirical method for determining surface temperature, using the split-window method and separate accounting for atmospheric attenuation and soil emissivity:

\[ T_s = T_a + (a_1 + a_2(T_s - T_a)) - (a_1 - T_s) + B(e) \]  \( (9) \)

where \( T_a, T_s \) are brightness temperatures in AVHRR channels 4, 5 (K); and \( a_1, a_2 \) are constant coefficients related to the atmospheric transmittances in the channels used.

\( B(e) \) accounts for the effect of surface emissivity \( e \) on the split-window equation and is thus a function of both atmospheric conditions and surface emissivity:

\[ B(e) = a^e (1 - e_1) - \beta (e_2 - e_1) \]  \( (10) \)

where \( e_1 \) and \( e_2 \) are the surface emissivities in AVHRR channels 4 and 5; \( a \) and \( \beta \) are parameters that depend on the atmosphere type and surface temperature.

Coll et al. (1994) used LOWTRAN 7 to calculate \( a \) and \( \beta \) for a set of atmospheric profiles (standard atmospheres and actual representative midlatitude sonde measurements), accounting for important atmospheric gases. They found that the influence of spectral variation of emissivity (represented by \( \beta \)) is larger for dry atmospheres and in general is more important than the effect of \( e_2 \) represented by \( a \).

Implementation

The approach outlined by Coll et al. (1994) was employed to obtain surface temperature values. Values of \( a_1 \) [Eq. (9)] were derived by Coll et al. from simulation, using various surface and air temperatures and surface emissivities for midlatitude conditions and thus considered applicable to this data set: \( a_1 = 1.29, a_2 = 0.28 \ \text{K}^{-1} \). Values of \( a \) and \( \beta \) [Eq. (10)] also were determined through simulation by Coll et al. and obtained from Coll’s Figure 2 for an atmospheric water-vapor content near 2.3 g/cm²: \( a = 45 \ \text{K}, \beta = 40 \ \text{K} \). Emissivity values were
obtained from a combination of sources, as detailed below.

In the thermal infrared region, most land pixels have emissivity above 0.9, but significant differences exist between soil and vegetation. Surfaces fully covered with green vegetation have a high emissivity; the value 0.98 often being quoted as representative (Cooper and Asrar, 1984; Chen and Zhang, 1986; Chen et al., 1989) but appears to have a wider range depending on species, plant component, and the wavelength range (Salisbury and D’Aria, 1992). The emissivity of soils and rocks varies about 0.95 (somewhat less for rocks), depending on the mineralogical composition and wavelength (Salisbury and D’Aria, 1992; 1994); soil roughness appears less important (Chen et al., 1989; Nerry et al., 1990b).

In light of the practical constraints and the lack of proved techniques for estimating emissivity of large pixels with a partially covered soil surface, an empirical approach was used with NDVI as a measure of the degree of vegetation cover. The correlation between spectral emissivity and NDVI was observed experimentally (Becker and Li, 1991; van de Griend and Owe, 1993). Van de Griend and Owe (1993) found a close relation between emissivity and the logarithm of NDVI, and this form was employed here as well. Data from various authors (van de Griend and Owe, 1993; Salisbury and D’Aria, 1994; Wan and Dozier, 1989; Nerry et al., 1990a) were used to derive a log-linear relation between NDVI and emissivity in the AVHRR channel 4, ε4, and the emissivity difference (εi−εc). On the basis of these data, it was assumed that the emissivity would not exceed the bounds <0.955,0.985>. Thus the final relation has the form

\[ ε_i = 0.9897 + 0.0299 \cdot \ln(NDVI). \]  

(11)

where NDVI is corrected for atmospheric and sun angle effects.

Similarly, it was assumed that (εi−εc) follows a log-linear relation with NDVI. The relation has the form

\[ ε_i - ε_c = 0.01019 + 0.01344 \cdot \ln(NDVI). \]  

(12)

It should be noted that these relations may be considered representative only in a statistical sense, especially for soils whose emissivities vary widely and can attain values up to 0.98 (Nerry et al., 1990a; Salisbury and D’Aria, 1992, 1994).

Figure 5 shows surface temperatures computed with various split-window formulas. It is evident that Eqs. (11) and (12) produce estimates that are close to those of Price (1984) and McClain et al. (1985) but significantly higher than those of Singh (1984). This is consistent with the results of Cooper and Asrar (1989), who found that the first two models performed well over a tall-grass prairie, but Singh’s model underestimated the actual values. The major difference between the above approach and the formulas of Price or McClain is the variation with NDVI, which yields higher or lower surface temperatures, depending on the NDVI value. Given the known dependence of emissivity on the surface composition, Eqs. (9)–(12) appear to produce acceptable results.

**Contaminated Pixel Replacement**

Only surface temperature values [computed with Eqs. (9)–(12)] that were contaminated during the “growing season” were replaced by interpolation. These were the values between the first and the last uncontaminated pixels. The interpolation was made in the same way as for channels 1 and 2.

**RESULTS AND DISCUSSION**

The processing steps described herein were intended to produce a portrait of the changing landscape during the growing season, without being obscured by snow or atmospheric effects. Given the complex nature of the relation between ecosystem characteristics and the AVHRR measurements as well as the pervasive lack of detailed information needed to correct for the ensuing errors or uncertainties, numerous assumptions had to be made. The temporal dimension is the key that helps unlock this complex set of problems and permits a first approximation of a “clean” data set to be produced.

A validation of the procedure cannot be made in a strict sense because no equivalent independent data set exists (or could reasonably be expected to be collected). The AVHRR sensor is currently unique, and the data used here are based on all the available measurements during the period. Thus, in addition to a before-after comparison for individual pixels (see below), we also used a single-date, near-nadir image from a temporally adjacent period that was not part of the composite. Specifically, we produced a full-country composite for 6 July

![Figure 5. A comparison of surface temperatures computed with various split-window formulas. The stepped behavior of the line from the present study [Eqs. (9)–(12)] reflects the effect of surface emissivity. C4, AVHRR channel 4.](image-url)
Figure 6 A comparison of corrected clear-sky composite pixels with single-date near-nadir clear-sky pixels: channels 1 (C1) and 2 (C2) and NDVI. The general equation for the line is $y = a + bx$, $R^2$ is the coefficient of determination, and s.e. is the standard error of the estimate. (a) Pixel-by-pixel; (b) 3x3 pixel means.
1993; carried out atmospheric and bidirectional reflectance corrections as described herein; prepared a cloud mask for this image, using CECANT thresholds derived for that period in 1993; and located an area with mostly cloud free pixels obtained with $\theta_0 < 10.01^\circ$. For these pixels (located in northwest Ontario, south, and southwest of Hudson Bay, dominantly coniferous forest), we located corresponding uncontaminated composite pixels for the period 11–20 July 1993, assuming that surface reflectance change between the two periods would be small near peak green.

Figure 6a shows the correspondence between the two data sets for individual pixels. The relation was closest for channel 2, likely in part because of the higher dynamic range. The fit for channel 1 was quite poor, evidently because of the low dynamic range of the pixels included in the comparison (mostly coniferous forest). In all cases, the lack of correspondence between the two data sets was enhanced by differences in pixel sizes and locations (reflecting \( \theta_0 \) and registration variations) and by imperfect atmospheric corrections. The deviation from a 1:1 slope is likely due to the increase of the signal from dark targets (coniferous forest, wetlands) through larger mixed pixels in the composite data (with brighter surface cover); note that this effect is substantially stronger in channel 1. To partly reduce these effects, both data sets were averaged, using $5 \times 5$ mean filter. The scatter of points was much reduced (Fig. 6b); for example, the standard error of the estimate decreased from 0.014 (0.04) to 0.006 (0.017) for channel 1 (2). The $r^2$ value improved significantly in the relative sense but was large only for channel 2.

The effect of the corrections on composite data is shown in Figure 7 for individual pixels representing various cover types. Only uncontaminated pixels show the symbol (square, triangle) at a given measurement period. In general, channel 1 values are considerably smoother and lower than the original data. As examination of the curves shows, the improved smoothness is a result of bidirectional corrections (smoother shape) as well as of cloud screening (removal of large anomalies). This is also true for channel 2; however, the corrections were larger because of the stronger angular dependence of channel 2 radiation. It is worth noting that small increases in reflectance (i.e., subpixel clouds, haze, or small bidirectional effects resulting in decreased NDVI) were also identified as contaminated by the CECANT procedure, in addition to the severely contaminated pixels near the end of the growing season. The interpolation procedures for contaminated pixels appeared to produce consistent trends, both during and at the end of the growing season—even for pixels with few uncontaminated measurements (note the last two examples in Fig. 7).

NDVI and surface temperature trajectories for the same pixels are shown in Figure 8. Overall, NDVI values are higher as a result of the atmospheric corrections (Holben, 1986; Cihlar et al., 1994). The seasonal trajectory became smoother after elimination of the contaminated pixels, but it also became more irregular as a result of the atmospheric corrections. This suggests that, for applications where the absolute NDVI value is not required (e.g., cover-type mapping), it may be preferable to use TOA NDVI values after screening for (and replacing) contaminated pixels. The interpolation for contaminated pixels again produced realistic trends, both during and at the end of the growing season. Surface temperature values also increased in comparison with TOA, as expected. The data show a tendency (most obvious for cropland) toward higher values in spring and early summer, before latent heat loss starts to dominate in regulating the surface heat budget. Figure 8 also illustrates the differences in the snow-free growing season (note temperature curves for the barren and tundra pixels). The final data set does not contain cloud and emissivity effects, but it does retain the information on the length of the growing season because, in contrast with the reflective channels and NDVI, no interpolation was made at the ends of the measurement period. However, it should be noted that the temperature values at the ends of the growing season could be lower than those shown as a result of contamination. Such cases can be identified by using the contamination mask (see examples in Fig. 8).

Mean values for several parameters representing the entire data set are shown in Figures 9 to 14 for 1) clear, 2) contaminated, and 3) all pixels. The mean $\theta$ shows a predictable seasonal trajectory for clear pixels with a minimum near summer solstice (Fig. 9a). The values for the contaminated pixels were distorted in all the figures by missing pixels in the Arctic (i.e., no image data in the composite) during periods 1 and 4. Figure 9a illustrates the challenge of optical remote sensing at northern latitudes: the average $\theta$ was $>60^\circ$ after 31 August 1993 (standard deviation for clear pixels was $<5^\circ$ for all periods). The average view zenith varied considerably (Fig. 9b). It is interesting to note that, although the overlap between the two groups was substantial (standard deviation about 20° in most compositing periods), the clear pixels had smaller view zenith angles during most of the growing season. This confirms that the CECANT procedure was sensitive not only to clouds or snow, but also to bidirectional effects, screening out pixels farther from nadir because atmospheric effects have been taken into account at that stage (Fig. 1). The overall $\theta$ curve ("all," Fig. 9b) is closer to that of clear pixels in midsummer and of contaminated pixels near the ends of the growing season; this reflects the seasonally varying proportions of clear and contaminated pixels. $\theta$ for "cloudy" pixels also shows the effect of missing data, especially for periods 1 and 4.

The effect of corrections on AVHRR channel 1 is presented in Figure 10. Figure 10a shows the measured TOA reflectance values. Figure 10b shows the measured values (corrected as discussed above) for clear pixels and
Figure 7. A comparison of original (top-of-the-atmosphere, dotted line) and corrected/interpolated (surface, solid line) AVHRR channel 1 and 2 data for selected pixels. Only uncontaminated pixels are denoted by a symbol (triangle for channel 1, square for channel 2). The pixel/line number is in the top left corner of each figure.
interpolated values for contaminated pixels. Figure 10c shows the mean values for pixel-by-pixel difference between the two data sets. At TOA, the strong effect of snow in early and late parts of the season (included among “cloudy” pixels) is evident. In midsummer, the contamination is mainly due to clouds, with mean values twice that of clear pixels (Fig. 10a). Surface channel 1 values decreased consistently for clear pixels (Fig. 10b) compared with TOA as a result of atmospheric corrections. The seasonal trend is related to the cover types for which clear pixels were available (dominated by the four types of forest and cropland in the last part of the season). The increase in channel 1 surface reflectance at the end of the season may be caused by errors in detecting contaminated pixels. This is because the low surface NDVI values may approximate those of snow or clouds while channel 1 reflectance is still below the CECANT threshold (0.30); coniferous forest and cropland had the most clear-sky pixels in period 20. The channel 1 reflectance for contaminated pixels (Fig. 10b) was higher than for clear pixels, although greatly reduced from the TOA values. The means for clear and contaminated values reflect differences in land cover. For example, most of the barren and tundra pixels were contaminated during periods 1–7, thus causing the “cloudy” pixels to have a higher mean value. In Figure 10c, the “clear” curve represents the average of pixel-by-pixel differences between TOA and surface values (i.e., the combined effect of atmospheric and bidirectional corrections). The overall reflectance of clear pixels decreased owing to atmospheric and bidirectional corrections (it decreased because of $\theta$, but increased in most cases because of $\theta_{\text{in}}=45^\circ$), about 0.05. It is thus evident that the ABC3 corrections are very significant, even for clear sky (the TOA values of channel 1 are about twice as large as for surface). Not surprisingly, the corrections are much larger for contaminated pixels, not only in midsummer when contamination is mostly due to clouds (about 100% of the surface value on the average, Figs. 10a and 10b) but also near the ends of the growing season when snow is the most significant factor (up to 400% of the mean surface reflectance).

Figure 11 shows the results for channel 2. Qualitatively, the results are similar to those for channel 1 except for surface reflectance (Fig. 11b), which exhibits a seasonal peak in all three groups. The relatively low correction for channel 2 (about 0.025 for most periods, Fig. 11c) reflects the small $\theta$ values for clear pixels. There was a major difference between TOA and surface values for the contaminated pixels (Fig. 11c), which represents the combined effect of all ABC3 corrections. The lower overall average in periods 1 and 4 was again due to missing data. It is worth noting that, in midsummer, the contaminated pixels had relatively low mean reflectance; that is, most of the contamination was due to clouds (full pixel or subpixel).

Figure 12 presents the results for temperature data. In all cases, the curves rise to a midseason maximum and subsequently decrease, as expected. For TOA, the mean seasonal increase was smallest in clear pixels and significantly larger in pixels contaminated by clouds or snow. The overall average increase of surface temperature above its TOA value was about 5 (K) for clear pixels (Fig. 12c). The effect of NDVI-dependent emissivity was approximately 2 (K), as evident from a comparison of midseason and late-season clear pixels. The inverse behavior of channel 4 cloudy pixels (compared with channels 1 and 2) indicates the seasonal temperature trend at northern latitudes and the latitudinally varying length of the growing season. The peak temperatures for cloudy pixels (e.g., periods 8 and 10) also reflect missing data that have been replaced by interpolation.

Figure 13 shows the effect of the atmospheric, bidirectional, and $\theta$ corrections on NDVI. For clear pixels, the mean seasonal NDVI range was about 0.18 at the TOA (Fig. 13a), 0.16 with $\theta$ correction (Fig. 13b), and 0.22 with full bidirectional correction (Fig. 13c). The rapid variations for cloudy pixels (Fig. 13a) are caused by missing pixels. The difference in seasonal trend between clear and cloudy pixels for surface data is due to land cover. At northern latitudes (higher snow/cloud contamination), the NDVI has a characteristic bell-shaped curve. Farther south, the coniferous forest has a strong effect on the mean NDVI. In the last part of the season (periods 14–19), the four forest types contained the most clear pixels, thus leading to increased NDVI. In period 20, most of the clear pixels were in cropland with a lower NDVI (Fig. 13c). The small increase in NDVI in period 19 is most likely due to the effect of the $\theta$ (compare Figs. 9a, 13b, and 13c).

Figure 14a shows mean curves for pixel-by-pixel differences between surface (corrected for $\theta$, effects only) and TOA NDVI. For clear pixels, the NDVI increased by 0.1 or more on the average. For clear pixels, the average correction in the first part of the season was about 0.1 (standard deviation 0.04), attributable mostly to atmospheric corrections (the mean $\theta$ for clear pixels was about 50°, Fig. 9a). The correction increased up to 0.2 in the later part of the season because of the increasing $\theta$ (Fig. 14a). The increase for contaminated pixels was higher, especially where missing pixels were replaced through interpolation.

Average NDVI values computed from BRDF-corrected channel 1 and 2 data are shown in Figure 14b. The major difference is evident for clear pixels, especially near the end of the growing season where the BRDF-corrected NDVI was considerably higher. The difference between clear and contaminated pixels at the end of the season (periods 17–20, Fig. 14b) is due to the land-cover types represented. The effect of $\theta$, only and full BRDF corrections is shown in Figure 14c as mean values computed from the NDVI(j) data for the two NDVI versions. During most of the season, the mean
Figure 8. A comparison of original (top-of-the-atmosphere, dotted line) and corrected/interpolated (surface, solid line) NDVI and temperature for the same pixels as in Figure 7 (the pixel line number is in the top left corner of each figure). Only uncontaminated pixels are denoted by a symbol (triangle for NDVI, square for temperature).
difference was about 0.03, for both clear and contaminated pixels. The large difference near the end of the season (periods 15–20) is due to the fact that the \(\theta\) correction was kept constant for \(\theta > 60^\circ\), consistent with Sellers et al. (1994), whereas the full BRDF correction was applied to all angles. During most of the season, the difference between the two versions is not large, although the fully corrected NDVI is consistently higher (Fig. 14c). Because the \(\theta\), corrected NDVI contains residual \(\theta\), errors (albeit small) that would decrease its value, the higher values for the fully corrected NDVI are expected. It thus appears that the NDVI derived from BRDF-corrected channels 1 and 2 is a better estimate of the actual NDVI value. However, the procedure employing \(\theta\) correction alone may offer more flexibility in replacing missing values, especially in the far north where few uncontaminated pixels may be available (less than three per season in some cases).

**SUMMARY AND CONCLUSIONS**

Compared with single-date images, the use of multitemporal AVHRR composites presents special challenges re-
Figure 11. Mean values per period for 1) uncontaminated, 2) contaminated, and 3) all (247,000) pixels for AVHRR channel 2: (a) top of the atmosphere; (b) surface; (c) pixel-by-pixel difference between top-of-the-atmosphere and surface reflectance.

Figure 12. Mean values per period for 1) uncontaminated, 2) contaminated, and 3) all (247,000) pixels for AVHRR channel 4: temperature: (a) top of the atmosphere; (b) surface; (c) pixel-by-pixel difference between top-of-the-atmosphere and surface temperatures.
Figure 13. Mean values per period for 1) uncontaminated, 2) contaminated and 3) all (247,000) NDVI: (a) top of the atmosphere; (b) NDVI corrected for SZA effect (FASIR); (c) NDVI computed from fully corrected AVHRR channel 1 and 2 (ABC3).

Figure 14. Mean values per period for 1) uncontaminated, 2) contaminated, and 3) all (247,000) pixels: (a) pixel-by-pixel difference between surface (based on FASIR procedure) and TOA NDVI; (b) pixel-by-pixel difference between surface (based on ABC3 procedure) and TOA NDVI; (c) pixel-by-pixel difference between surface NDVI computed in two ways, FASIR and ABC3. SZA, solar zenith angle; TOA, top of the atmosphere.
lated to residual atmospheric, bidirectional, geometric, and seasonal effects. The approach described here, dubbed ABC3 (atmospheric, bidirectional and contamination corrections of CCRS), consists of a set of procedures that have been developed to significantly reduce some of these effects. The aim is to produce a data set that portrays the land surface itself and its seasonal dynamics in the shortwave and longwave region of the electromagnetic spectrum. ABC3 yields equivalent nadir values with a constant solar zenith angle for surface reflectance in AVHRR channels 1 and 2, NDVI, and surface temperature corrected for atmospheric and surface emissivity effects for each compositing period. The analysis of the corrected data set indicated that the procedures removed a large proportion of variations associated with residual clouds (full and subpixel), bidirectional effects, and thermal emissivity. Considering mean values across Canada, the degree of contamination decreased approximately in the order snow>clouds>bidirectional atmospheric effects. It should be noted that the atmospheric corrections did not fully correct for the varying aerosol concentration, except indirectly when the effect approximated that of a cloud and could be eliminated through CECANT. Compared with clear-sky pixels, the presence of snow or clouds greatly increased the mean surface reflectance—for example, by 400% (for snow/clouds) and 100% (clouds only in midsummer). It is therefore critically important that these sources of contamination be identified and screeened out.

The two versions of NDVI (FASIR, corrected for solar zenith effect only, and ABC3, based on fully corrected channels 1 and 2) differed significantly. The difference was smaller during most of the growing season (about 0.03), with ABC3 yielding higher values. It increased much more after mid-September because of the high solar zenith angles during that period. This indicates that full BRDF correction is required at northern latitudes, even for applications based on NDVI.

The major remaining sources of error appear to be due to geometric errors, both misregistration and varying pixel sizes in the composites; incomplete atmospheric corrections because only nominal parameter values could be used at this stage; and imperfect bidirectional corrections, which are to a significant extent related to the preceding two. These deficiencies are due to a lack of accurate ancillary data and not the ABC3 procedure itself. The geometric errors will only decrease when higher resolution data and more precise geolocation information become available with future sensors such as MODIS. Improved atmospheric corrections depend primarily on a better knowledge of aerosols and water vapor at the time of image acquisition, which, again, will have to await future sensors. The extent to which these improvements will also resolve the residual bidirectional errors remains to be established. From this data set and related work (Li et al., 1996), it can be concluded that the simple modeling approach used here provides a good first approximation.

The resulting data set represents a substantial improvement over the “raw” composite data for applications to vegetation studies. However, the resulting data set is not a sound substitute for single-date, near-nadir images, where such uncontaminated images are available and where neither timeliness nor the multitemporal observations are required.

**REFERENCES**


Li, Z. (1990), L’emissivité spectrale et teledetection infrarouge thermique: mesure relative, analyse spectrale et impacts sur la determination de la temperature de surface. These de l’Université Louis Pasteur, Strasbourg, France.


