Kronos: A Software System for the Processing and Retrieval of Large-Scale AVHRR Data Sets

Zengyan Zhang, Joseph JáJá, David A. Bader, Satya N.V. Kalluri, Huiping Song, Nazmi El Saleous, Eric Vermote, and John R.G. Townshend

Abstract

Raw remotely sensed satellite data have to be processed and mapped into a standard projection in order to produce a multi-temporal data set which can then be used for regional or global Earth science studies. However, traditional methods of processing remotely sensed satellite data have inherent limitations because they are based on a fixed processing chain. Different users may need the data in different forms with possibly different processing steps; hence, additional transformations may have to be applied to the processed data, resulting in potentially significant errors. In this paper, we describe a software system, Kronos, for the generation of custom-tailored products from the Advanced Very High Resolution Radiometer (AVHRR) sensor. It allows the generation of a rich set of products that can be easily specified through a simple interface by scientists wishing to carry out Earth system modeling or analysis. Kronos is based on a flexible methodology and consists of four major components: ingest and preprocessing, indexing and storage, a search and processing engine, and a Java interface. After geo-location and calibration, every pixel is indexed and stored using a combination of data structures. Following the users' queries, data are selectively retrieved and secondary processing such as atmospheric correction, compositing, and projection are performed as specified. The processing is divided into two stages, the first of which involves the geo-location and calibration of the remotely sensed data and, hence, results in no loss of information. The second stage involves the retrieval of the appropriate data subsets and the application of the secondary processing specified by the user. This scheme allows the indexing and the storage of data from different sensors without any loss of information and, therefore, allows assimilation of data from multiple sensors. User specified processing can be applied later as needed.

Introduction

Over the last few years, satellite-based sensors have become the primary source of information at regional scales for geographical, meteorological, and environmental studies, because they provide the needed dynamic temporal view of the Earth's surface. A particularly important sensor is the Advanced Very High Resolution Radiometer (AVHRR), on board the National Oceanic and Atmospheric Administration (NOAA) series of satellites, which has been used for monitoring the terrestrial environment at resolutions ranging from 1 km to very coarse resolutions of 15 km and greater (Townshend, 1994). Its coarse resolution data sets — the only data available globally on a daily basis — have been collected over a time period of more than 16 years. Reflectance and temperature measurements from the AVHRR have been widely used for a variety of studies at both regional and continental scales, which include monitoring seasonal land-cover dynamics (Justice *et al.*, 1985; Justice, 1986; Townshend and Justice, 1986), classifying land cover (Tucker *et al.*, 1985; Townshend *et al.*, 1987; Defries and Townshend, 1994; Defries *et al.*, 1995), estimating variables to model surface fluxes and energy balance (Price, 1982; Seguin and Itier, 1983; Kustas *et al.*, 1994; Seguin *et al.*, 1994; Kalluri *et al.*, 1998), monitoring biomass burning (Matson and Holben, 1987; Kaufman *et al.*, 1990; Kennedy, 1992), modeling net primary production (Goward and Dye, 1987; Prince and Goward, 1995), asssessing crop condition and predicting yield (Boatwright and Whitehead, 1986; Tucker and Choudhury, 1987; Teng, 1990; Quarmby *et al.*, 1993; Doraiswamy and Cook, 1995), and monitoring sea surface temperature operationally (McClain *et al.*, 1985).

The Global Area Coverage (GAC) multi-channel AVHRR data have been available since 1981, and the acquisition of these data is expected to continue well into this century. However, the GAC data are not in a format which can be readily used in regional or global scientific studies. The raw AVHRR data are also referred to as level 1B data, and they contain uncalibrated digital numbers from the five bands of AVHRR, calibration and Earth location information, satellite and solar geometry, and telemetry information (Kidwell, 1997). The level 1B data have to be processed and mapped to a standard projection and co-registered to produce multi-temporal data sets.

Thus far, several global data sets have been produced from the AVHRR instrument for the study of land-cover dynamics. These data sets have found widespread use in a variety of global and regional land-science applications, such as regional agricultural monitoring, interannual variation in vegetation and changes in the length of growing season, global deforestation, land-cover classification, and climate modeling (Tucker *et al.*, 1994; Townshend, 1994; Myneni *et al.*, 1997). Continental Normalized Difference Vegetation Index (NDVI) data sets are

Z. Zhang and J. JáJá are with the Institute for Advanced Computer Studies (UMIACS) and Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 (zengyan@umiacs.umd.edu).

D. Bader is with the Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87131.

S. Kalluri, H. Song, N. El Saleous, and E. Vermote are with the Department of Geography, University of Maryland, College Park, MD 20742.

J.R.G. Townshend is with UMIACS and the Department of Geography, University of Maryland, College Park, MD 20742.

Photogrammetric Engineering & Remote Sensing
Vol. 66, No. 9, September 2000, pp. 1073–1082.

0099-1112/00/6609–1073$3.00/0

produced by the Global Inventory Monitoring and Modeling Studies (GIMMS) group at NASA's Goddard Space Flight Center (GSFC) (Holben, 1986; Los *et al.*, 1994), and 1 km global land data sets are produced by the Earth Resources Observation Systems (EROS) Data Center (EDC) (Eidenshink and Faundeen, 1994). Several versions of the Global Vegetation Index (GVI) products (Kidwell, 1990; Goward *et al.*, 1994a; Goward *et al.*, 1994b) are also produced. In 1994, the Pathfinder AVHRR land data set processing was initiated to provide the users with a well calibrated and consistently processed land data set for global change research studies (James and Kalluri, 1994; Smith *et al.*, 1997). However, there are some inherent limitations in these efforts. The above-mentioned data sets are each in a fixed geographic projection and have been processed using a fixed compositing method over a certain time period. Some of the data sets also have had atmospheric correction applied. It has been shown (Steinwand *et al.*, 1995) that a significant loss of information occurs during reprojection of raster imagery. Prince and Goward (1996) reviewed the suitability of the Pathfinder I data set for the modeling of Net Primary Production and have shown that the full potential of the AVHRR data could not be realized due to the compositing of thermal channels into 10-day composites using a maximum NDVI criterion, the failure of the cloud screening procedure, and the absence of ocean data. Different methods of compositing multi-temporal data could have significant impacts on the calculation of biophysical variables from the satellite data (Choudhury *et al.*, 1994; Roy, 1997). Because the requirements of individual users change depending upon their specific application and use of the data, it is desirable to have a system that can generate custom tailored data sets rapidly and efficiently.

Furthermore, for a comprehensive modeling and monitoring of Earth system processes, it is necessary to combine data from several remote sensing instruments (e.g., King, 1999). This requires a software infrastructure that allows the archiving and retrieval of data from multiple sensors for a given geographic location as well as combining and correlating the resulting multiple datasets. The system methodology adopted in Kronos can provide the foundation of this infrastructure with added functionality as needed. Following the Pathfinder AVHRR land data set processing (James and Kalluri, 1994) in 1994, research was carried out under NASA's Pathfinder Data Set Product Generation Algorithm Development project (NASA NRA-94-MTPE-06) to develop a processing system that would generate custom-tailered data products from AVHRR data (El Saleous *et al.*, 2000). However, an efficient hierarchical indexing and storage data structure is required to deal with the large archive of AVHRR data which extends from 1981 to the present. Storing and indexing multi-temporal data from several years is an important research area which is being addressed by the scientific and computing community (Shock *et al.*, 1996; Wolfe *et al.*, 1998). To allow multi-temporal and multi-resolution analysis, algorithms that rapidly search and retrieve the processed data based on user defined spatio-temporal queries are needed. An efficient processing system should also allow the users to perform interactive computation and analysis of the retrieved data. The focus of this paper is to describe an efficient indexing scheme that we have developed to archive and retrieve geo-located and calibrated AVHRR IFOV (instantaneous field of view) data. Our goal in this work is to substantially expand the potential utility of the AVHRR data by allowing an on-demand capability for processing and generating user-specified subsets of AVHRR data while providing well-tested and widely accepted standard processing software modules.

We have used algorithms developed by El Saleous *et al.* (2000) to convert GAC level 1B data to physical measurements suitable for scientific applications. These processed data are then indexed and stored following an efficient hierarchical spatial-temporal data structure.

In particular, the prototype system described in this paper provides an easy-to-use interface that allows a user to request AVHRR data products based on a region of interest, a time period, a compositing method, and a projection. The system will efficiently generate the data products from the AVHRR GAC level 1B data without introducing errors due to reprojection or re-gridding. A library consisting of various map projections, compositing functions, and atmospheric correction modules (El Saleous *et al.*, 2000) provides a programming environment that enables the user to generate with ease a wide variety of products. The system has been optimized for the efficient access and generation of user-specified products based on the AVHRR GAC data.

The rest of the paper is organized as follows: (1) an overview of the system architecture is given; (2) the ingest and preprocessing of the AVHRR GAC 1B raw data using algorithms from El Saleous *et al.* (2000) is described; (3) a brief description of the indexing scheme and the storage of the data on a standard disk array is given; (4) the library functions for projection, compositing, and atmospheric correction are listed; and (5) a sample of products generated using Kronos and related performance results are given.

System Architecture

Our starting point is the AVHRR GAC level 1B orbit data which consists of a set of IFOV records organized according to the scan lines of each orbit (Figure 1). Daily GAC level 1B data contain about 14 files for the entire globe, each file corresponding to one orbit and holding all the IFOVs generated for that orbit. Each IFOV record contains the reflectances registered through the five channels (two visible channels and three thermal channels), and estimates of its geo-locations, time information, and some miscellaneous information. Our goal was to design a system that ingests AVHRR level 1B data and allows the handling of high-level tasks of the following types:

- *Type I.* Given a region specified on a global map, a time period, a projection selected from a list of standard projections, a compositing function selected from a list of the most commonly used functions, and an atmospheric correction algorithm selected from a list of algorithms, generate the corresponding data products with the specified resolution.

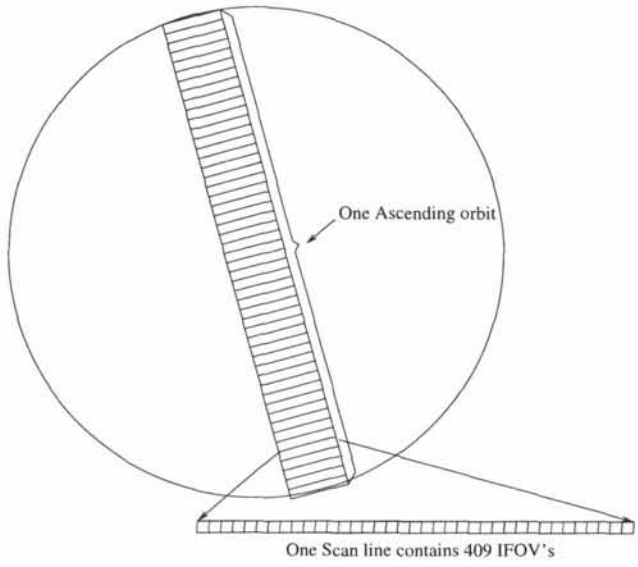

Figure 1. A single file of GAC data represents one ascending satellite orbit, about 7000 scan lines.

- *Type II.* Same as Type I except that the user can substitute his/her own processing algorithms rather than using the particular algorithms provided through the library.

The architecture of Kronos consists of four major components as shown in Figure 2. They are ingest/preprocessing, indexing and storage, a search and processing engine, and a Java interface.

- *Ingest/Preprocessing.* This component ingests AVHRR level 1B data and applies procedures for geo-location, calibration, and cloud screening.
- *Indexing and Storage.* This component deals with indexing the data and laying it out across a disk array.
- *Search and Processing Engine.* This component consists of a search engine and a library of basic processing functions such as computing different projections, atmospheric correction algorithms, and compositing functions. The purpose of this component is to fetch the appropriate data and to apply the required processing tasks.
- *Java Interface.* This component allows users to easily formulate their queries according to their research needs, and it generates a multi-layer output image as requested.

In the next sections, we provide details regarding these components.

Ingest and Preprocessing Phase

Raw AVHRR level 1B data are given in a packed format in orbit files (Kidwell, 1997). Each orbit has about 13,000 scan lines and each scan line has 409 pixels. Each pixel is an IFOV corresponding to a patch of the Earth and is recorded as a vector of digital numbers (DN) corresponding to the measurements in the five spectral bands. Additional scan line information for time, location geometry, and calibration are also appended. The main preprocessing tasks involve geo-location, calibration, and cloud screening. The material presented in this section is a summary of the AVHRR processing algorithms reported by El Saleous *et al.* (2000).

During the ingestion and preprocessing phase, observation data are extracted from the AVHRR GAC level 1B orbits and accurate geo-location and calibration are performed. For each IFOV, the precise latitude and longitude of the center using the geo-location scheme of Baldwin and Emery (1993), Patt and Gregg (1994), and Rosborough *et al.* (1994) are derived, and the solar zenith angle, view zenith angle, and relative azimuth angle are computed. At this stage, a set of pixel quality flags are also computed. Finally, the details of this phase are provided.

Geo-Location

Although Earth location points and solar and view angles are provided in the AVHRR GAC level 1B data set, observed errors of several GAC pixels are commonly found (James and Kalluri, 1994). Therefore, the data are geo-located using an orbital model and updated ephemeris data (also called orbital reference information which is updated regularly and maintained on file at Satellite Services Branch of the National Climatic Data Center). That is, a satellite clock correction algorithm and an orbital model are used to determine the position of the spacecraft from the latest available ephemeris data, followed by a determination of the most accurate geo-location information, such as the latitude and longitude of the IFOV's center or the geometry of the IFOV's perimeter (Baldwin and Emery, 1993; Patt and Gregg, 1994; Rosborough *et al.*, 1994). The revised Earth locations and solar/view geometry, such as solar zenith angle, view zenith angle, and relative azimuth angle, are calculated for each pixel.

Calibration

Calibration is required due to the fact that the sensor response changes shortly after launch and during the sensors' lifetime.

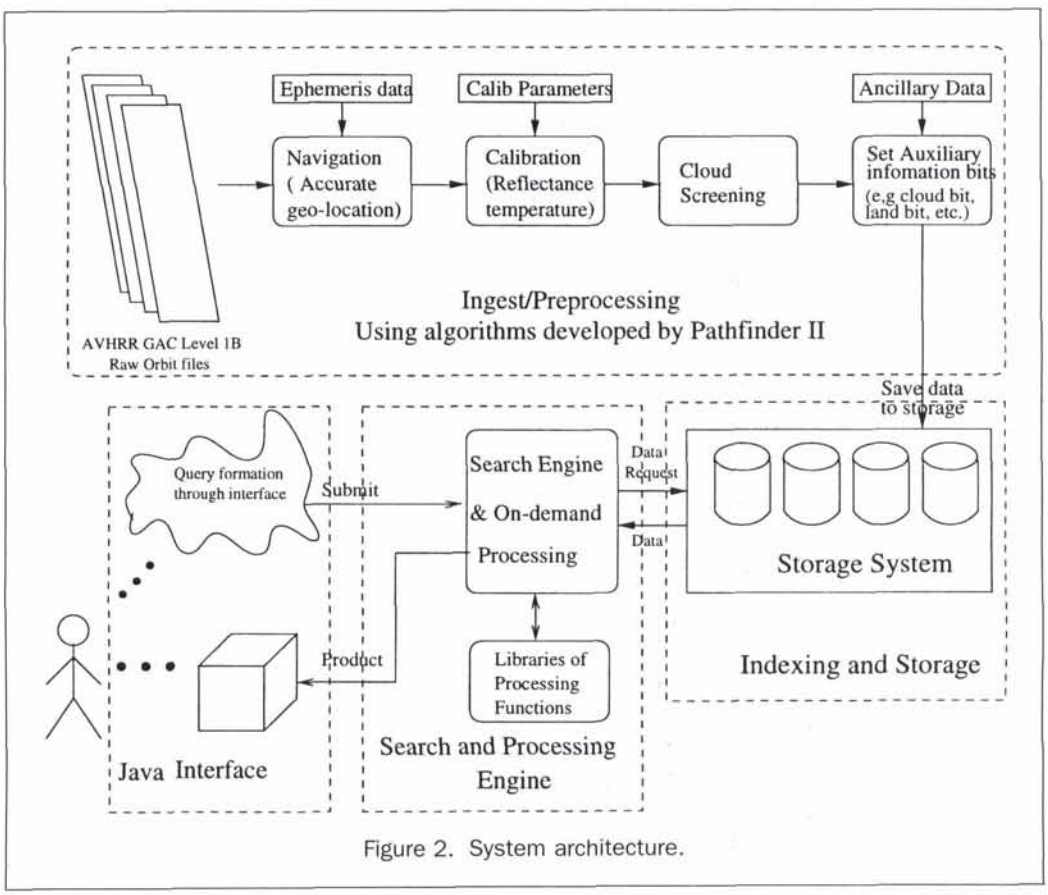

Figure 2. System architecture.

TABLE 1. AVHRR GAC OBSERVATION RECORD. THE FLOATING POINT DATA WERE PROPERLY SCALED AND CASTED TO THE DESIRED DATA TYPES (USHORT MEANS UNSIGNED SHORT). THERE ARE 28 BYTES TOTALLY IN EACH IFOV RECORD.

AVHRR GAC Observation	Type	Byte	Range
Latitude of the IFOV center	Long	4	−90.0000000–90.0000000
Longitude of the IFOV center	Long	4	−180.0000000–180.0000000
Greenwich Mean Time (GMT)	Short	2	00.000–24.000
Calibrated band 1 reflectance	Short	2	000.00–100.00
Calibrated band 2 reflectance	Short	2	000.00–100.00
Calibrated band 3 temperature	Ushort	2	000.00–373.00
Calibrated band 4 temperature	Ushort	2	000.00–373.00
Calibrated band 5 temperature	Ushort	2	000.00–373.00
Solar zenith angle	Short	2	00.00–90.00
View zenith angle	Short	2	−60.00–60.00
Relative azimuth angle	Ushort	2	000.00–360.00
Pixel quality flags: NOAA QC, land/sea, night/day, CLAVR, cloud shadow	Ushort	2	Bitmask

Channels 1 and 2 are the visible and near-infrared channels, respectively, and are calibrated to produce at-satellite radiances using a time-dependent correction that accounts for sensor degradation, and it intercalibrates among the satellites. The approach used relies on calibration coefficients derived empirically from the data following an in-flight calibration method (Vermote and Kaufman, 1995). Channels 3 to 5 are calibrated using a nonlinear function based on the internal calibration targets, baseplate temperatures, instrument-dependent response curves, and NOAA-provided gains and offsets (Weinreb *et al.*, 1990; Rao, 1993). Channel 3 is calibrated using the gains and offsets in the GAC data record. Thermal channels are converted to equivalent brightness temperatures using a lookup table based on the inverse Planck function convolved with the instrument response.

Pixel Quality Flags

A set of pixel quality flags to identify cloud condition, cloud shadow, land/water, night/day, and NOAA Quality Control flags in the original GAC 1B data are computed. A detailed description of these flags can be found in Ouaidrari *et al.* (1997) and El Saleous (2000). A pixel is classified as mixed, cloudy, or clear using the CLAVR (Clouds from AVHRR) algorithm (Stowe *et al.*, 1991), which performs a series of cloud screening tests based on thresholds in the five AVHRR bands derived from sample data over a variety of surface types, including deserts and ice fields. Once clouds are identified, then cloud shadows are determined by a series of trigonometric functions (El Saleous *et al.*, 2000). After the preprocessing is completed, a record for each IFOV containing the information shown in Table 1 is saved.

Indexing and Storage

To generate a two-dimensional (2D) gridded output image following a user's spatial and temporal specifications, such as a task of Type I introduced earlier, the system should be able to identify, retrieve, and determine which IFOV should be binned into an output grid box. Given the large quantity of data available, we clearly need to build an indexing structure that allows an efficient way to retrieve the requested data without having to access any data outside the region of interest. The problem of building efficient spatial data structures has been extensively studied in the literature. A comprehensive overview can be found in Samet (1990). However, none of these methods seems to be directly applicable to our case because we want to search by both temporal and spatial bounds in order to find all the nearest-neighbor IFOVs from different orbits covering the specified output grid box. We also want the capability to co-register our data with other types of remotely sensed data as accurately as possible. Thus, we have developed an indexing structure that combines an equal-angular grid and a spatial data structure called the k-d trees (k = 2) (Bentley, 1975) for all the IFOVs within a cell of the grid. We elaborate on this indexing structure next.

Consider an equal-angular global grid D with resolution of 1 degree by 1 degree indexed by the latitude and longitude representing one day of global AVHRR data. Each cell in D will contain all the geo-located IFOVs whose spatial coordinates (latitude and longitude) fall within this cell. Such a set of IFOVs is referred to as the bucket corresponding to the specified cell. Figure 3 illustrates the variance in sizes of the different buckets for a specific set of daily data (day 121 of 1989). The mean number of IFOVs in a bucket is 530 and the maximum is 940. Given a region R specified by a query, we want to be able to retrieve all the IFOVs in the global grid of D that falls within R. These IFOVs will be used by the processing engine to build a 2D output grid whose grid point latitude and longitude are computed according to the map projection and resolution. For each such output grid point, we want to be able to easily find all the nearest neighbors from different orbits and then choose a representative IFOV from these nearest neighbors using some compositing criterion (such as maximum NDVI, minimum channel one, etc). To accomplish this, we may need to selectively access certain IFOVs within a bucket, and hence we build a k-d trees (k = 2) (Bentley, 1975) on top of the IFOVs in each bucket. This is a very efficient spatial data structure for selectively accessing a portion of static spatial data. For more details regarding k-d trees, see Bentley (1975). The size of the overall indexing structure per day is approximately 0.5 MB, which can easily fit in main memory.

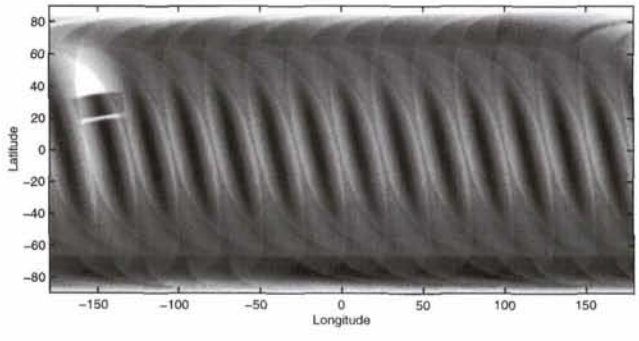

Figure 3. IFOV's distribution of the geo-located one-day AVHRR data (1 deg by 1 deg) for day 121 of 1989. The darker the data, the more IFOVs reside in the grid cell. The mean number of IFOVs per cell is 530 and the maximum is 940. (Note: There are some missing scanlines in the upper left part.)

TABLE 2. AVAILABLE MAP PROJECTIONS

Geographic (GEO)	Universal Transverse Mercator(UTM)
State Plane Coordinates (SPCS)	Albers Conical Equal Area (ALBERS)
Lambert Conformal Conic (LAMCC)	Mercator (MERCAT)
Polar Stereographic (PS)	Polyconic (POLYC)
Equidistant Conic (EQUIDC)	Transverse Mercator (TM)
Stereographic (STEREO)	Lambert Azimuthal Equal Area (LAMAZ)
Azimuthal Equidistant (AZMEQD)	Gnomonic (GNOMON)
Orthographic (ORTHO)	General Vertical Near-Side Perspetive (GVNSP)
Sinusiodal (SNSOID)	Equirectangular (EQRECT)
Miller Cylindrical (MILLER)	Van der Grinten (VGRINT)
(Hotline) Oblique Mercator (HOM)	Robinson (ROBIN)
Space Oblique Mercator (SOM)	Alaska Conformal (ALASKA)
Interrupted Goodes Homolosine (GOOD)	Mollweide (MOLL)
Interrupted Mollweide (IMOLL)	Hammer (HAMMER)
Wagner IV (WAGIV)	Wagner VII (WAGVII)
Oblated Equal Area (OBEQA)	Satellite Perspective

To incorporate the temporal index, the two-dimensional grid is extended into a three-dimensional structure in which the third dimension is used to represent time. As we are expecting consecutive days of data, such a structure will be quite efficient. In fact, given a multi-day query, we can easily get the addresses of the buckets falling in the specified region, after which we use the k-d trees to retrieve the appropriate IFOVs.

We now turn our attention to the method used to place the data on a disk array. Given the three-dimensional grid indexing, we partition it into equal size sub-cubes, each of size m by n by t corresponding to latitude, longitude, and time. We then stripe the data in the sub-cubes across the disk array. In each bucket, we build an optimal k-d tree and save it as a complete tree with IFOV record on the node of the tree. This scheme is quite efficient in retrieving data because it is highly likely that the data will be spread almost equally among the disks.

Search and Processing Engine

This component computes a multi-layered output image that conforms to the specifications supplied by the query, using the IFOV records indexed and stored by the indexing and storage component. These specifications, provided by the users through a Java interface to be described in the next section, include the following parameters: (1) region of interest, (2) time period, (3) desired resolution of the output image, (4) map projection, (5) desired atmospheric correction, and (6) compositing function. In particular, this component includes libraries of standard map projections (Snyder, 1987) (Table 2), atmospheric correction routines (Vermote and Tanre, 1992; Tanre *et al.*, 1992; Vermote *et al.*, 1996; Holben *et al.*, 1998; El Saleous *et al.*, 2000) (Rayleigh/Ozone, Water Vapor, and Stratospheric Aerosol), and compositing functions (e.g., Maximum NDVI, Minimum Channel 1 reflectance, and Maximum Channel 4 temperature). These libraries can easily be augmented with additional functions and allow the user to select any desired combination.

The basic procedure used by the Search and Processing Engine consists of the following sequence of steps:

(1) Generate a 2D uniform grid whose dimensions are derived from the spatial region and the resolution specified by the query;
(2) For each pixel (i,j) of the desired output grid, compute the latitude and the longitude according to the map projection and the resolution specified by the query;
(3) For each day in the specified time period, do the following for each pixel (i,j):
   (a) Retrieve all the IFOVs from different orbits that are nearest to (i,j) within a rectangle P determined by the resolution and its center (i,j),
   (b) Apply atmospheric correction as appropriate to each of the retrieved IFOVs, and
   (c) Compute a unique set of layer values for pixel (i,j) using the specified compositing function on the retrieved IFOVs;
(4) For each pixel (i,j) of the output grid, compute the unique set of layer values for (i,j) using the compositing function over all the daily values covering the specified time period; and
(5) Generate the desired set of layers of the output grid.

Notice that, in our implementation, Steps 3 and 4 are combined in such a way that, once daily data are generated, they are composited immediately with the previous data, and, hence, we only need to store a single intermediate copy of the output grid.

Steps 1, 4, and 5 can be done in a straightforward way using the information supplied by the query. Step 2 amounts to invoking the appropriate routine from the library of standard projections (Snyder, 1987). We now elaborate on the details regarding Step 3.

Using our indexing grid D described in the previous section, for each pixel (i,j) all the buckets intersecting with rectangle P are retrieved. A nearest-neighbor search is then applied from different orbits on the k-d tree associated with the buckets of (i,j). Once these IFOVs are found, an atmospheric correction is applied to each such IFOV as specified by the query.

Atmospheric correction can be applied because the radiances measured by the instrument on board the satellite are affected by the presence of the atmosphere between the sensor and the target due to atmospheric scattering and absorption (Tanre *et al.*, 1992). Channels 1 and 2 are affected by ozone and water vapor absorption as well as molecular and aerosol scattering. However, the Rayleigh scattering mainly affects channel 1. Aerosol scattering is the most challenging term for atmospheric correction because of the spatial and temporal variability of both the type and the amount of particles in the atmosphere. The option of performing the atmospheric corrections for the effects of Rayleigh, ozone, water vapor, and stratospheric aerosols as described by Vermote and Tanre (1992), Vermote *et al.* (1996), Holben *et al.* (1998), El Saleous *et al.*, (2000) is provided.

To minimize disk accesses, our system applies Steps 3a through 3c to all the pixels in each row of the output image in order from the highest latitude to the lowest. The maximum number of buckets in descending order of latitude in main memory is fitted, and the corresponding pixels are processed as described above. Also, if the output image is too big to be kept in main memory, the image is divided into horizontal stripes and an output image is generated stripe by stripe.

Finally, this component provides routines to generate several data products, including the Normalized Difference Vegetation

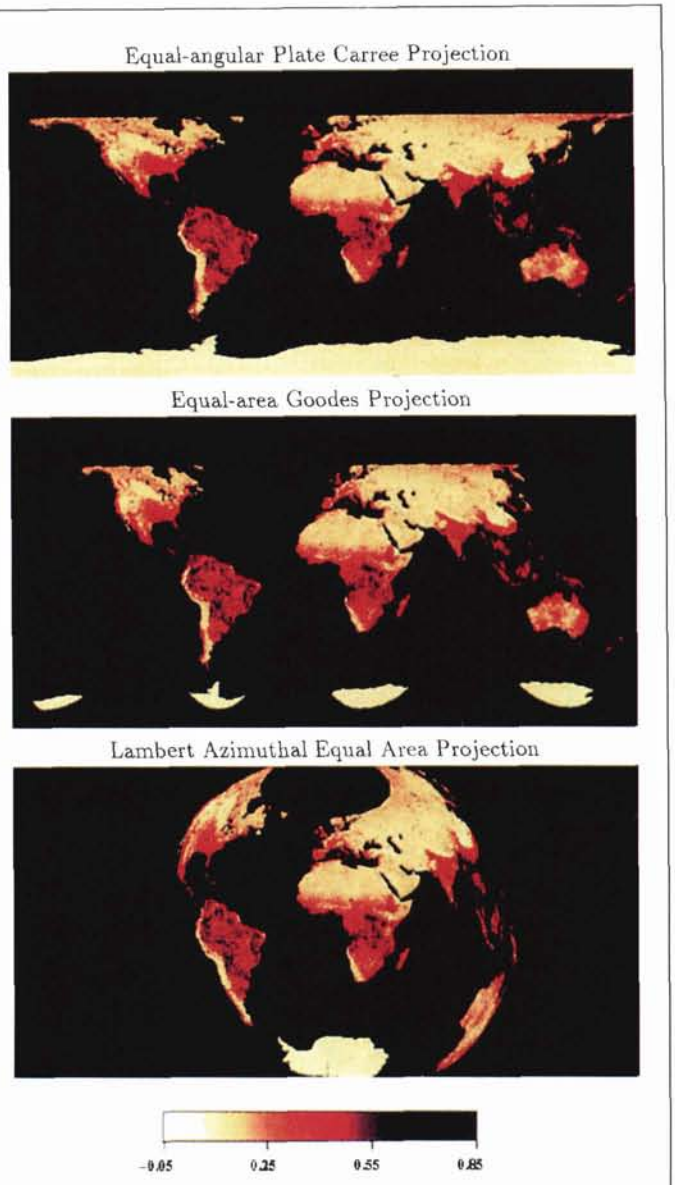

Plate 1. Global 10-day composite (day 1 to 10 of 1989) NDVI in different projections.

Index (NDVI), Channel 1 and 2 reflectances, and Channels 3 to 5 brightness temperature. The data products can either be returned as a collection of flat raster image files (with one image layer per file) along with a process description text (e.g., compositing function chosen, resolution used, image size, etc), or as a standard packaged format such as NCSA (National Center for Supercomputing Applications) Hierarchical Data Format (HDF) or CDF which can easily pass layers of imagery and descriptive comments of the processing from the processing system to the user.

User Interface

A Java interface is provided, as shown in Figure 4, which allows users with different interests to formulate their queries and interact easily with the system. There are three methods for the user to specify the spatial bounds of the query. The first consists of drawing a rectangle over the desired region on a global map which can be panned and zoomed by the user. The second method is to select a region from a pop-up menu containing a list of continents or regions. The corresponding bounding latitude and longitude are displayed in the query formulation worksheet. The third method allows user to enter the numerical coordinates directly into the appropriate boxes.

From the temporal period list, the user can choose any temporal bounds to perform multi-day compositing. Users can select a preferred map projection from the list shown in Table 2. Because map projections transform the three-dimensional surface of the Earth (an oblated spheroid) into a planar surface, distortions of area, conformality, distance, direction, and scale may be introduced. Different map projections minimize different distortion measures, and, hence, their selection depends on the application. The users can also select compositing functions from a pop-up menu, such as Maximum NDVI, Minimum channel 1 reflectance, and Maximum Channel 4 temperature. Atmospheric correction in the reflective bands is also an option, including Rayleigh/Ozone, Water Vapor, and Stratospheric Aerosol. Users can choose any combination of these algorithms to be applied or can exclude all of them.

By specifying the spatial-temporal bounds, compositing methods, map projection, resolution, atmospheric correction, and output list, a query is formed which can then be submitted to the system through a submit button.

Implementation and Results

A prototype system written in C has been run on an IBM RS6000 Model 3CT workstation with a 67MHz superscaler POWER2 processor and an approximately 60 GB of IBM SSA Disk Subsystem (sustained data rates up to 35 MB/s). For one day's data (day 121 of 1989), it takes about 130 minutes to geo-locate the data, and about 30.1 minutes to build the index with the optimal k-d trees (k = 2). The size of the resulting data set for one day is 960MB. As mentioned before, the size of the index for one day is approximately 520KB. Plate 1 shows the global NDVI generated by Kronos for one day using Plate Carree, Goodes, and Lambert Azimuthal Equal Area Projections. Table 3 gives the running time for some sample queries tested on Kronos. Additional data products generated by Kronos are shown in Plate 2. The spatial distribution of the reflectances and brightness temperatures are consistent with known variations across different land-cover types. The geometry and angular information which is preserved for each pixel could be used for studying land surface anisotropy (Zhang *et al.*, 1998). In spite of using the maximum NDVI compositing criterion, significant cloud contamination can be seen in the southern part of Africa.

We are currently developing a multiprocessor version of Kronos. Most of the operations can be easily mapped into a multiprocessor system with very little or no communication required. One way to accomplish this is to divide the output image into equal-size horizontal stripes and let each node process one stripe. In fact, we have already developed the code for this strategy, and users are able to achieve an almost linear speedup in terms of the number of available processors.

Summary

In this paper, we describe an integrated software processing system for archiving, retrieving, and processing AVHRR remotely sensed data. We developed a hierarchical indexing scheme to efficiently access large amounts of spatial and temporal data to process, store, and archive so that pertinent information can be extracted with ease. A platform independent Java interface provides an easy way for defining the spatial and temporal queries. The system could also be used for other satellite sensors as well so that the complete original data records of each IFOV can be retained without sampling. This system provides the capability for the fusion of multi-resolution, multi-sensor, and multi-temporal data.

Storing the satellite data in this form has distinct advantages over storing the data in flat image format because it

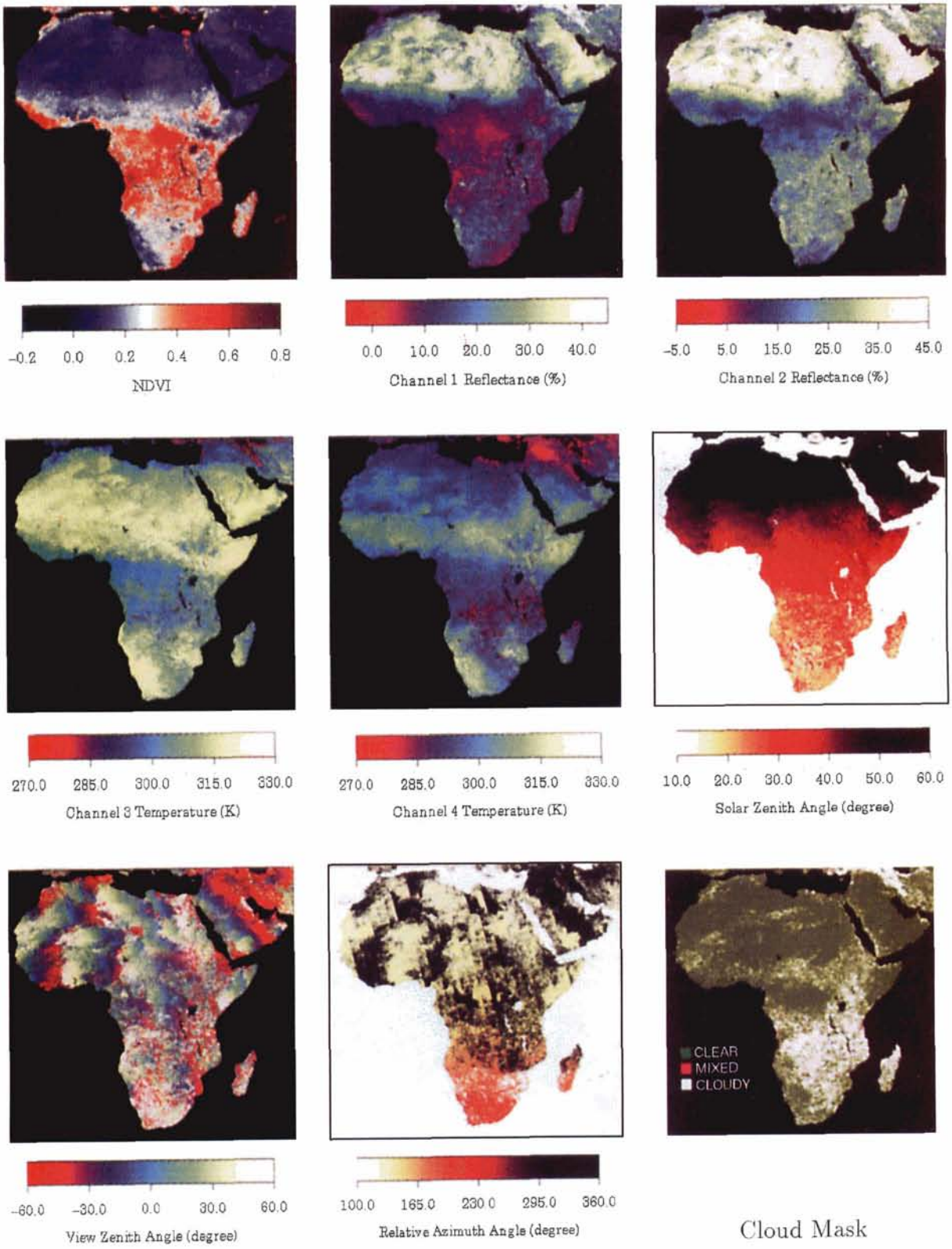

Plate 2. Some 10-day composite products (day 1 to 10 of 1989) for the continent of Africa in the Plate Carree Projection.

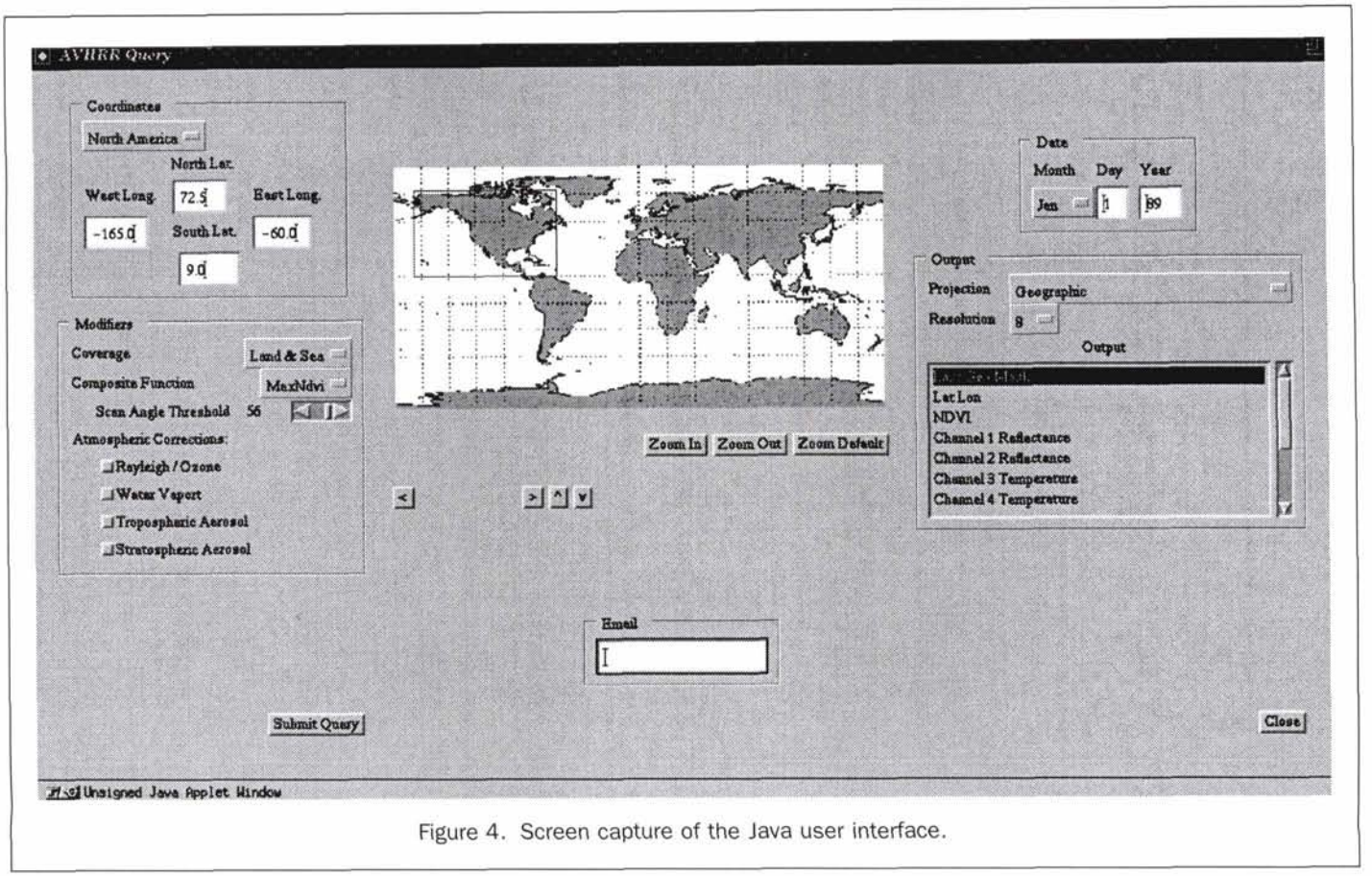

Figure 4. Screen capture of the Java user interface.

TABLE 3. RUNNING TIME IN MINUTES FOR GENERATING NDVI PRODUCTS FOR THE SPECIFIED REGION AND PROJECTION FOR DAY 121 OF 1989

Place/Size	Plate Carree	Goodes	Lambert Azimuthal	Azimuthal Equidistant
Global (5004 × 2502)	66.7	34.7	43.7	61.4
Africa (1131 × 1061)	3.5	3.2	4.1	4.0
N America (1460 × 883)	5.2	3.6	4.2	4.5
Eastern US (320 × 390)	0.49	0.40	0.44	0.46

allows the data to be repeatedly processed without resampling and and it allows different multi-resolution multi-temporal data products to be generated without introducing reprojection errors. The system is designed in such a way that it can be easily ported to a multi-processor environment such as the IBM SP2. This can be achieved by using parallel inter-processor communication primitives such as the message passing interface (MPI) and a parallel hierarchical storage system such as the High Performance Storage System (HPSS). A parallel implementation of this system will further reduce the processing time considerably. This system will not only allow "operational" processing and ingesting of satellite data, but will also enhance the capability to reprocess historic data to provide consistently processed error-free data for the Earth science community. Our research work is expected to have a major impact on the methodology of processing satellite data, and it is also expected to help in expanding the use of the NASA Earth science data and information beyond basic research to a broader user community, including both private and public sectors.

Acknowledgment

We would like to thank the UMIACS parallel system staff, Jim Humphries, Allan Tong, Jim Kukla, and Gary Jackson, for their system support and their development of the Java interface. This work was supported by NASA ESIP (NCC5300), NSF NPACI (FCPO PO10152408), and NSF Grand Challenge (BIR 9318183) grants.

References

Baldwin, D., and W. Emery, 1993. A systemized approach to AVHRR image navigation, *Annals of Glaciology*, 17:414–420.

Bentley, J.L., 1975. Multidimensional binary search trees used for associative searching, *Communications of the ACM*, 18(9):509–517.

Boatwright, G.O., and V.S. Whitehead, 1986. Early warning and crop condition assessment research, *IEEE Transactions on Geoscience and Remote Sensing*, GE-24:54–64.

Choudhury, B.J., N.E. Digirolamo, and T.J. Dorman, 1994. A comparison of reflectances and vegetation indices from three methods of compositing the AVHRR-GAC data over northern Africa, *Remote Sensing Reviews*, 10:245–263.

Defries, R.S., and J.R.G. Townshend, 1994. NDVI-derived land cover classifications at a global scale, *International Journal of Remote Sensing*, 15(17):3567–3586.

Defries, R.S., M. Hansen, and J.R.G. Townshend, 1995. Global discrimination of land cover types from metrics derived from AVHRR pathfinder data, *Remote Sensing of Environment*, 54:209–222.

Doraiswamy, P.C., and P.W. Cook, 1995. Spring wheat yield assessment using NOAA AVHRR data, *Canadian Journal of Remote Sensing*, 21:43–51.

Eidenshink, J.C., and J.L. Faundeen, 1994. The 1km AVHRR global data set: First stages in implementation, *International Journal of Remote Sensing*, 15(17):3443–3462.

El Saleous, N., E.F. Vermote, C.O. Justice, J.R.G. Townshend, J. Tucker, and S.N. Goward, 2000. Improvements in the global biospheric record from the Advanced Very High Resolution Radiometer (AVHRR), *International Journal of Remote Sensing*, 21(6):1251–1278.

Goward, S.N., and D.G. Dye, 1987. Evaluating North American net primary productivity with satellite observations, *Advances in Space Research*, 7:165–174.

Goward, S.N., D.G. Dye, S. Turner, and J. Yang, 1994a. Objective assessment of the NOAA global vegetation index data product, *International Journal of Remote Sensing*, 14:3365–3394.

Goward, S.N., S. Turner, D.G. Dye, and S. Liang, 1994b. The University of Maryland improved global vegetation index product, *International Journal of Remote Sensing*, 15(17):3365–3395.

Holben, B.N., 1986. Characteristics of maximum-value composite images from temporal AVHRR data, *International Journal of Remote Sensing*, 7(11):1417–1434.

Holben, B.N., T.F. Eck, I. Slutsker, D. Tanre, J.P. Buis, A. Setzer, E.F. Vermote, J.A. Reangan, Y.J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, and A. Smirnov, 1998. Aeronet—A Federal instrument network and data archive for aerosol characterization, *Remote Sensing of Environment*. 66(1):1–16.

James, M.E., and S.N.V. Kalluri, 1994. The Pathfinder AVHRR land data set: An improved coarse resolution data set for terrestrial monitoring, *International Journal of Remote Sensing*, 15(17):3347–3363.

Justice, C.O., 1986. Monitoring the grasslands of semi-arid Africa using NOAA-AVHRR data, *International Journal of Remote Sensing*, 7:1383–1622.

Justice, C.O., J.R.G. Townshend, B.N. Holben, and C.J. Tucker, 1985. Analysis of the phenology of global vegetation using meteorological satellite data, *International Journal of Remote Sensing*, 6:1271–1381.

Kalluri, S.N.V., J.R.G. Townshend, and P. Doraiswamy, 1998. A simple single layer model to estimate transpiration from vegetation using multi-spectral and meteorological data, *International Journal of Remote Sensing*, 19:1037–1053.

Kaufman, Y.J., A. Setzer, C. Justice, C.J. Tucker, M.C. Pereira, and I. Fung, 1990. Remote sensing of biomass burning in the tropics, *Fire in the Tropical Biota. Ecosystem Processes and Global Challenges* (J.G. Goldammer, editor), Springer Verlag, Berlin, pp. 371–399.

Kennedy, P.J., 1992. Biomass burning studies: The use of remote sensing, *Ecological Bulletins*, 42:133–148.

Kidwell, K.B., 1990. *Global Vegetation Index User's Guide*, NOAA/NESDIS National Climatic Data Center, U.S. Department of Commerce, Washington, D.C., 214 p.

———, 1997. *NOAA Polar Orbiter Data Users Guide*, NOAA/NESDIS National Climatic Data Center, U.S. Department of Commerce, Suitland, Maryland, 236 p.

King, M.D., 1999. *EOS Science Plan*, NP-1998-12-069-GSFC, NASA, Greenbelt, Maryland, 397 p.

Kustas, W.P., E.M. Perry, P.C. Doraiswamy, and M.S. Moran, 1994. Using satellite remote sensing to extrapolate evapotranspiration estimates in time and space over a semiarid rangeland basin, *Remote Sensing of Environment*, 49:275–286.

Los, S., C.O. Justice, and C.J. Tucker, 1994. A global 1° × 1° NDVI data set for climate studies: Part I: Derivation of a reduced resolution data set from the GIMMS Global Area Coverage product of the AVHRR, *International Journal of Remote Sensing*, 15:3493–3518.

Matson, M., and B.N. Holben, 1987. Satellite detection of tropical burning in brazil, *International Journal of Remote Sensing*, 8:509–516.

McClain, E.P., W.G. Tichel, and C.C. Walton, 1985. Comparative performance of AVHRR-based multichannel sea surface temperatures, *Journal of Geophysical Research*, 90(C6):11587–11601.

Myneni, R.B., C.D. Keeling, C.J. Tucker, G. Asrar, and R.R. Nemani, 1997. Increased plant growth in the northern high latitudes from 1981 to 1991, *Nature*, 386:698–702.

Ouaidrari, H., E. Vermote, N. El Saleous, and D. Roy, 1997. AVHRR Pathfinder II data set: Evaluation and improvements, *Physical Measurements and Signatures in Remote Sensing* (G. Guyot and T. Phulpin, editors), A.A. Balkema, Rotterdam/Brookfield, pp. 131–137.

Patt, F., and W. Gregg, 1994. Exact closed-form geolocation algorithm for earth survey sensors, *International Journal of Remote Sensing*, 15(18):3719–3734.

Price, J.C., 1982. Estimation of regional scale evapotranspiration through analysis of satellite thermal infrared data, *IEEE Transactions on Geoscience and Remote Sensing*, GE20:286–292.

Prince, S.D., and S.N. Goward, 1995. Global primary production: A remote sensing approach, *Journal of Biogeography*, 22:815–835.

———, 1996. Evaluation of the NOAA/NASA pathfinder AVHRR land data set for global primary production modeling, *International Journal of Remote Sensing*, 17(1):217–221.

Quarmby, N.A., M. Milnes, and T.L. Hindle, 1993. The use of multitemporal NDVI measurements from AVHRR data for crop yield estimation and prediction, *International Journal of Remote Sensing*, 14:199–210.

Rao, C.R.N., 1993. *Non-Linearity Corrections for the Thermal Infrared Channels of the Advanced High Resolution Radiometer: Assessment and Recommendations*, NOAA Technical Report NESDIS-69, NOAA/NESDIS National Climatic Data Center, Washington, D.C., 31 p.

Rosborough, G., D. Baldwin, and W. Emery, 1994. Precise AVHRR image navigation, *IEEE Transactions on Geoscience and Remote Sensing*, 32(3):644–657.

Roy, D.P., 1997. Investigation of the maximum normalized difference vegetation index (NDVI) and the maximum surface temperature (t-s) AVHRR compositing procedures for the extraction of NDVI and t-s over forest, *International Journal of Remote Sensing*, 18(11):2383–2401.

Samet, H., 1990. *The Design and Analysis of Spatial Data Structures*, Addison-Wesley, Reading, Massachusetts, 493 p.

Seguin, B., and B. Itier, 1983. Using midday surface temperature to estimate daily evaporation from satellite thermal IR data, *International Journal of Remote Sensing*, 4:371–383.

Seguin, B., D. Courault, and M. Guerif, 1994. Surface temperature and evapotranspiration: Application of local scale methods to regional scales using satellite data, *Remote Sensing of Environment*, 49:287–295.

Shock, C., C. Chang, L. Davis, S. Goward, J. Saltz, and A. Sussman, 1996. A high performance image database system for remotely sensed imagery, *Proceedings of Euro-Par'96*, 22–29 August, Lyon, France, Springer, Berlin, 2:109–122.

Smith, P.M., S.N.V. Kalluri, S.D. Prince, and R. DeFries, 1997. The NOAA/NASA Pathfinder AVHRR 8-km land data set, *Photogrammetric Engineering & Remote Sensing*, 63(1):12–13, 27–32.

Snyder, J., 1987. *Map Projections - A Working Manual*, U.S. Geological Survey Professional Paper 1395, U.S. Geological Survey, Washington, D.C., 383 p.

Steinwand, D.R., J.A. Hutchinson, and J.P. Snyder, 1995. Map projections for global and continental data sets and an analysis of pixel distortion caused by reprojection, *Photogrammetric Engineering & Remote Sensing*, 61(12):1487–1497.

Stowe, L.L., E.P. McClain, R. Carey, P. Pellegrino, G.G. Gutman, P. Davis, C. Long, and S. Hart, 1991. Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data, *Advances in Space Research*, 3:51–54.

Tanre, D., B.N. Holben, and Y.J. Kaufman, 1992. Atmospheric correction algorithm for NOAA-AVHRR products: theory and applications, *IEEE Transactions on Geoscience and Remote Sensing*, 30(2):231–248.

Teng, W.L., 1990. AVHRR monitoring of U.S. crops during the 1988 drought, *Photogrammetric Engineering & Remote Sensing*, 56:1143–1146.

Townshend, J.R.G., 1994. Global data sets for land applications from the advanced very high resolution radiometer: An introduction, *International Journal of Remote Sensing*, 15(17):3319–3332.

Townshend, J.R.G., and C.O. Justice, 1986. Analysis of the dynamics of African vegetation using the normalized difference vegetation index, *International Journal of Remote Sensing*, 7:1189–1207.

Townshend, J.R.G., C.O. Justice, and V.T. Kalb, 1987. Characterization and classification of South American land cover types using satellite data, *International Journal of Remote Sensing*, 8:1189–1207.

Tucker, C.J., J.R.G. Townshend, and T.E. Goff, 1985. African land-cover classification using satellite data, *Science*, 227:369–375.

Tucker, C.J., and B.J. Choudhury, 1987. Satellite remote sensing of drought conditions, *Remote Sensing of Environment*, 23:243–252.

Tucker, C.J., W.W. Newcomb, and A.E. Dregne, 1994. AVHRR data sets for determination of desert spatial extent, *International Journal of Remote Sensing*, 15(17):3519–3545.

Vermote, E., and D. Tanre, 1992. Analytical expressions for radiative properties of planar Rayleigh scattering media including polarization contribution, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 47(4):305–314.

Vermote, E., and Y.J. Kaufman, 1995. Absolute calibration of AVHRR visible and near infrared channels using ocean and cloud views. *International Journal of Remote Sensing*, 16(13):2317–2340.

Vermote, E.F., N.Z. El Saleous, and B.N. Holben, 1996. Aerosol retrieval and atmospheric correction, *Advances in the Use of NOAA AVHRR Data for Land Applications* (G. D'Souza, A.S. Belward, and J.P. Malingreau, editors), Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 93–124.

Weinreb, M.P., G. Hamilton, S. Brown, and R.J. Koczor, 1990. Nonlinearity corrections in calibration of advanced very high resolution radiometer infrared channels, *Journal of Geophysical Research*, 90(C5):7381–7388.

Wolfe, R.E., D.P. Roy, and E. Vermote, 1998. MODIS land data storage, gridding, and compositing methodology: Level 2 grid, *IEEE Transactions on Geoscience and Remote Sensing*, 36(4):1342–1338.

Zhang, Z., S.N.V. Kalluri, J. JáJá, S. Liang, and J.R.G. Townshend, 1998. Models and high performance algorithms for global BRDF retrieval, *IEEE Computational Science and Engineering*, 5(4):16–29.

(Received 05 March 1999; accepted 24 August 1999; revised 02 November 1999)