NASA's Global Orthorectified Landsat Data Set

Compton J. Tucker, Denelle M. Grant, and Jon D. Dykstra

Abstract

NASA has sponsored the creation of an orthorectified and geodetically accurate global land data set of Landsat Multispectral Scanner, Thematic Mapper, and Enhanced Thematic Mapper data, from the 1970s, circa 1990, and circa 2000, respectively, to support a variety of scientific studies and educational purposes. This is the first time a geodetically accurate global compendium of orthorectified multi-epoch digital satellite data at the 30- to 80-m spatial scale spanning 30 years has been produced for use by the international scientific and educational communities. We describe data selection, orthorectification, accuracy, access, and other aspects of these data.

Background

Spatial variability exists for our planet's land surface at dimensions of meters to tens of meters, due to local terrain variability and associated microclimatic influences upon vegetation types and associations. Accordingly, spatial data at tens of meters are required to accurately map many areas, because of the low spatial autocorrelation of land-surface features (Townshend and Justice, 1988; Townshend and Justice, 1990). In addition, a variety of natural and human land-use changes, such as wild fires, deforestation, wetland conversion, and urbanization, represent alterations of landscapes, which also occur at spatial scales of tens of meters. These are important perturbations of the global environment and require similar spatial scale data for quantification. Currently, we lack information regarding where environmental change is occurring, what the changes are, and what the post-change properties of the altered areas are (Townshend *et al.*, 1991).

Understanding environmental or land-cover dynamics represents an important challenge in the study of the global environment, because many land-cover changes take place at fine scales of resolution, requiring Landsat-type imagery for accurate measurement. Uses for such data range from biodiversity and habitat mapping for localized areas, to specifying parameters for large-scale numerical models simulating biogeochemical cycling, hydrological processes, and ecosystem functioning. These needs have been recognized in the International Geosphere Biosphere Programme, the World Climate Research Programme, and the International Satellite Land Surface Climatologic Project, among others (Becker *et al.*, 1988; IGBP, 1990; WMO, 1992). Responding to these needs, the Scientific Data Purchase Program of NASA's Stennis Space Center (http://www.esa.ssc.nasa.gov/datapurchase) has directed the production of global orthorectified and co-registered Landsat Multispectral Scanner (MS), Thematic Mapper (TM), and Enhanced Thematic Mapper (ETM+) data for three periods: the late 1970s, circa 1990, and circa 2000, respectively.

Landsat 80-, 30-, and 15-m satellite data are the only record of global land-surface conditions at a spatial scale of tens of meters spanning the last 30 years. They constitute an indispensable history of land-surface state. Data at these spatial resolutions can provide a high potential mapping accuracy of natural vegetation and alterations to it, if and only if highly accurate scene-to-scene within- and among-date registration is achieved. Otherwise, misregistration errors between or among dates are confused with land-cover change and resulting interpretations are meaningless (Townshend *et al.*, 1992).

The best solution to eliminate or minimize misregistration errors is to precision orthorectify each scene within each of the three epochs. The process of orthorectification removes erroneous image displacements caused by the interaction between terrain relief or local elevation changes and sensor orientation variations. The orthorectification process results in remotely sensed image products that possess both the image-based information of the original satellite data and the geometric information of a geodetically accurate map.

When Landsat data are assembled into a mosaic to cover an area of interest, an underlying assumption is that the data have a consistent geometry throughout the image. This has been demonstrated for the MS, TM, and ETM+ instruments with increasing accuracy, respectively (Desachy *et al.*, 1985; Welch *et al.*, 1985; Malaret *et al.*, 1985; Bryant *et al.*, 1985; Storey and Choate, 2000). If the Earth's surface were the same elevation over the geographical area of interest and the satellites in question were in identical orbits for the period of interest, a Landsat data mosaic could be "tied" together using a linear or affine mapping, and the resulting mosaic would be an accurate representation of the surface with accurate distances among all surface features. When the topography is irregular, as is normally the case, it is necessary to correct localized horizontal displacements created by perspective view distortions around areas of local relief. This process, called orthorectification, combines knowledge of the elevation of each image point with the precise viewing geometry at that point to calculate a horizontal correction to the satellite data. The result is an image product that appears as if the satellite or the viewer is looking normal to the Earth at every location. In such orthogonal views, the horizontal position of any feature directly beneath the viewer would not be effected by local terrain variations.

Correcting satellite imagery for variations in topography and/or satellite viewing perspectives is best accomplished by using ground coordinates, also known as ground control points. This enables digital elevation data to be associated with the respective satellite data by matching coordinates. Accurate association of the digital elevation data with the satellite image data is necessary to compensate for topography and/or viewing perspective variations. When the satellite data have been corrected for terrain and/or satellite viewing perspectives, these data are referred to as having been "orthorectified." In this paper we use orthorectified to mean the satellite data have been corrected for terrain displacements, corrected for any

C.J. Tucker and D.M. Grant are with the Laboratory for Terrestrial Physics, NASA/Goddard Space Flight Center, Greenbelt, MD 20771 (compton@ltpmailx.gsfc.nasa.gov).

J.D. Dykstra is with Earth Satellite Corporation, 6011 Executive Blvd., Suite 40, Rockville, MD 20852.

Photogrammetric Engineering & Remote Sensing
Vol. 70, No. 3, March 2004, pp. 313–322.

0099-1112/04/7003–0313/$3.00/0

satellite viewing variations, and have accurate geodetic coordinates associated with the data.

The benefits of common orthorectification of global Landsat data for three epochs are a very high geodetic mapping accuracy for all epochs, minimizing between- and among-scene registration errors within each epoch, and minimizing between- and among-date registration errors. The latter two considerations minimize classification errors due to misregistration; the former consideration provides the basis for a highly accurate geodetic global mapping standard. The data set we describe achieves not only highly accurate within- and among-scene common orthorectification globally, but has also been "tied" to ground control points to result in a high absolute geodetic accuracy. These data represent a potentially valuable data source for the study of the terrestrial environment which has heretofore not been available.

Orthorectification Basis

Initially, the Landsat data set we describe herein included assembling and orthorectifying Landsat MSS data from the 1970s and TM data for circa 1990. In the third year (2000) of the project, NASA decided to add circa 2000 data from the ETM+ instrument. We refer to these three periods hence forth as "epochs." A consequence of the evolution of the project is that Landsats 4 and 5 data became the orthorectification basis for the three epochs. In retrospect, it would have been preferable to use the Landsat 7 ETM+ instrument as the mapping basis for all epochs, because these data have a better location accuracy (Storey and Choate, 2000). The only exception to the Landsat 4 and 5 orthorectification basis is a block of 505 scenes from northern South America, where nonsystematic geometric errors in Landsat 4 and 5 data from receiving stations in Brazil and Ecuador prevented these data from providing the orthorectification basis for this area.

Extent of Landsat Coverage

The circa 1990 epoch data for this project includes 7,600 Landsat-4 and -5 TM scenes which cover continental areas and selected islands. The corresponding 1970s epoch images from Landsats 1, 2, and 3 comprise 8,190 MSS scenes, due to different area coverage for Landsats 1, 2, and 3 versus Landsats 4 and 5. A total of 7,550 MSS scenes were acquired for 6,976 unique path-rows (e.g., there were 574 duplicate scenes due to cloud cover, etc.)(Plate 1a). Consequently, there are 1,214 MSS path-rows for which no data exist; 600 are from Brazil and adjacent areas and 614 elsewhere. We estimate that about 60 percent of the missing 1970s MSS scenes could be obtained if a sustained effort were made at data recovery, because many of these scenes exist with individual researchers.

The 600 missing MSS scenes from the reception area of the Brazil Landsat receiving station are not available from this receiving station because the magnetic media containing these data has degraded. Similar problems exist at many other Landsat receiving stations. This is a continuing problem which must be addressed in a systematic fashion to avoid additional losses of irreplaceable satellite data. The data set we describe is a major step towards the preservation of unique Earth resources data.

The circa 1990 data includes 7,413 Landsat 4 and 5 scenes for 7,037 unique path-rows, with 377 duplicate scenes for cloud cover mitigation. The missing 563 scenes were largely over Siberia and were obtained from Landsat 7 (Plate 1b). This complete "best-of-station" data set now resides at EDC and represents a significant contribution to preservation and distribution of baseline global Landsat TM data for the circa 1990 time period.

A total of 8,500 Landsat 7 ETM+ scenes were obtained for the circa 2000 epoch. An additional 900 Landsat 7 path-rows were acquired over the circa 1990 epoch data, to cover most islands and parts of Antarctica, in addition to covering the same areas as the circa 1990 epoch (Plate 1c).

Data Selection Criteria

The absence of catalogues of Landsat 1, 2, 3, 4, and 5 data from all ground receiving stations made data availability determination a major task, because many receiving stations are the best, and frequently the only, source of Landsat data from their reception regions. Lists of all possible Landsat scenes available from the Earth Resources Observation Systems (EROS) Data Center (EDC) and foreign ground receiving stations were compiled and reviewed. Thirteen receiving stations, including EDC, were utilized for the circa 1990 data (Figure 1 and Plate 2). MSS data for the 1970s epoch were only available from the Canadian, European, and American archives, thus limiting the total number and quality of late 1970s images available for orthorectification.

For each available scene from all epochs, an image analyst examined the satellite data for sensor problems, missing scan lines, clouds and haze, and any other manifestation of poor data quality. These image characteristics were weighed against the year and the season the data were acquired. Preference for scene selection was based on the following:

- *Year of Acquisition.* (1) MSS scenes were acquired largely for dates between 1972 and 1980, due to a limited amount of data. A limited amount of Landsat 4 and 5 MSS data were also acquired (Plate 1a). The lack of availability of MSS data is exacerbated by the loss of much of the foreign MSS archives through degradation of magnetic media. (2) There was limited acquisition and archiving of images from Landsat 4 and 5 globally by the Earth Observation Satellite Company during the "commercialization" period of the Landsat program from 1985 through 1999. The absence of an aggressive global data acquisition and archiving strategy for Landsat 4 and 5 TM data was exacerbated by cloud cover, data quality, and changing surface phenological conditions. It was thus not feasible to use data only from 1990. Landsat TM images acquired for this epoch generally range in year from 1987 to 1993, with preference given to the best scene available that also met the other image approval criteria. A consequence of the "commercialization" of Landsat from 1985 through 1999 was in order to

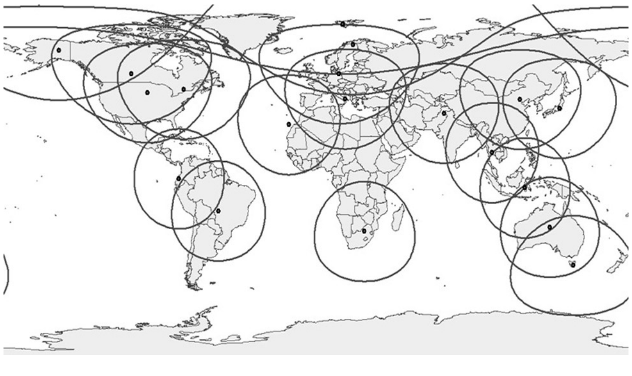

Figure 1. Location of the Landsat receiving stations in operation for the Landsat satellites. From 50 to 100 percent of the data required for all epochs was obtained from the EROS Data Center, Sioux Falls, South Dakota. Multispectral scanner data were acquired from five foreign receiving stations for the 1970s epoch to complete the coverage for this period. It was necessary to purchase data from 11 foreign receiving stations to complete the TM coverage for the circa 1990 epoch. All Landsat 7 data for the circa 2000 epoch were acquired from the EROS Data Center.

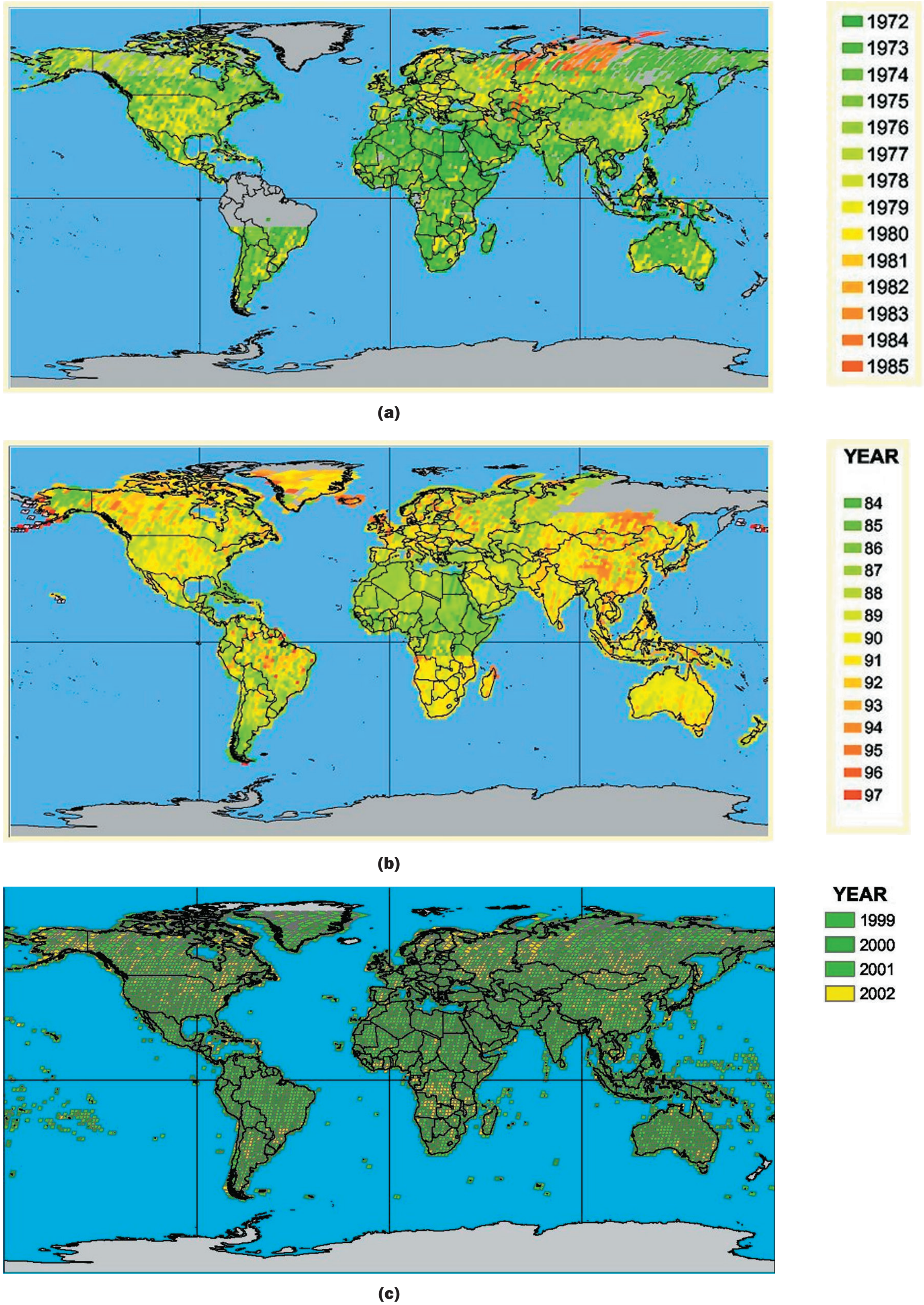

Plate 1. Distribution of the Landsat data for the three epochs comprising the global orthorectified data as described in this paper. (a) Circa 1970s MSS imagery largely from Landsat 1, 2, and 3 with a few MSS scenes from Landsat 4 and 5. (b) Circa 1990 TM imagery from Landsat 4 and 5 with a block of Landsat 7 data from Siberia. (c) Circa 2000 ETM+ data from Landsat 7. Several regional gaps exist with no data for the 1970s, largely because these early data from Landsat 1, 2, and 3 have been lost due to tape degradation.

minimize expenses, the Earth Observation Satellite Company usually collected Landsat 4 and 5 data if someone was willing to purchase it. Fortunately, many Landsat receiving stations (Figure 1) collected and archived large quantities of Landsat 4 and 5 data during this period. (3) Landsat 7 data were usually from 2000, although some data were acquired from 1999 and some from 2001 and 2002. The farsighted Landsat 7 global data acquisition strategy resulted in large numbers of ETM+ imagery globally over a short time period (Goward *et al.*, 1999; Arvidson *et al.*, 2001). This is a major accomplishment of the Landsat 7 Project and should be the standard by which future Landsat-type missions are judged.

- *Cloud Cover.* Clouds cover about 60 percent of the land surface at any given time (Rossow and Schiffer, 1999). Although it would be desirable to use only cloud-free Landsat scenes, this is impossible to do in many parts of the world due to persistent cloudiness. The percentage of cloud cover was approximated by scene from the meta-data produced by the EROS Data Center, and priority was given to those scenes that had the least cloud cover. All candidate imagery was visually examined, and the best scenes were chosen. In cases where multiple scenes had similar cloud cover, a decision based on cloud location and vegetation phenology was made. Over tropical rain forests, where persistent cloudiness occurs, scenes with a higher cloud cover were occasionally selected, because they represented the best data. Multiple scenes for some areas were also acquired and processed where those scenes together provided supplementary information missed by only selecting one scene or the other.
- *Data Quality.* Landsat data are prone to a number of errors: missing scan lines, pixel drop outs, saturated or missing bands, etc. These result from instrument malfunctions and/or problems at ground receiving stations. These data errors are sometimes noted in the metadata and sometimes not. Priority for data quality was given only to the reflective channels. Thermal-band data quality issues for TM and ETM+ data did not preclude images from these epochs. Final data quality judgment was based upon visual and radiometric inspections of every image used.
- *Phenology.* The TM and ETM+ spectral bands are optimized for studying vegetation; thus, growing season imagery is more useful than is imagery during periods of senescence (Tucker and Sellers, 1986). To take advantage of this, Landsat 7 acquisition times were selected during times when the historical normalized difference vegetation index (NDVI) (Tucker, 1979) data were at peak values (Goward and Williams, 1997). A similar approach was adopted in this project for the selection of all imagery. In very humid areas, peak greenness was frequently inappropriate because significant cloud cover was also present. The low acquisition frequency of Landsat 4 and 5 data was also a problem in parts of the world where no ground receiving stations existed (Figure 1). When clear scenes were not available during the period of maximum NDVI, substitute scenes were selected based upon cloud-cover criteria. At high latitudes and elevations, care was made to insure presence of green vegetation and the absence of snow and ice. This was accomplished by visual inspection of all questionable candidate images, frequently by more than one analyst.
- *Single Data Runs.* Obtaining consecutive scenes within a path collected on the same date increases the ease with which information can be derived from the scenes, because atmospheric conditions and solar zenith angle variations will change gradually for the block of data in question. Thus, geophysical values derived from an assemblage of rows from the same path acquired on the same day will be easier to interpret than those from a mosaic constructed from scenes acquired at different times. A priority was given to acquire scenes from the same path for a given day where possible. If cloud or data quality was an issue, data from different dates were selected for the scenes in question.

Orthorectification

The best available geodetic and elevation control data were used to correct all imagery for positional accuracy. The National Imagery and Mapping Agency (NIMA) provided geodetic control points to the Earth Satellite Corporation where they were used in data processing. Because of the sensitive nature of the ground control points provided, these data cannot be released to the public. We made the decision to use this approach to achieve the highest geodetic accuracy possible for the three data epochs. If users of the data we describe wish to have other Landsat data orthorectified to the three Landsat epochs described herein, they may attempt this themselves or contract this work to the Earth Satellite Corporation.*

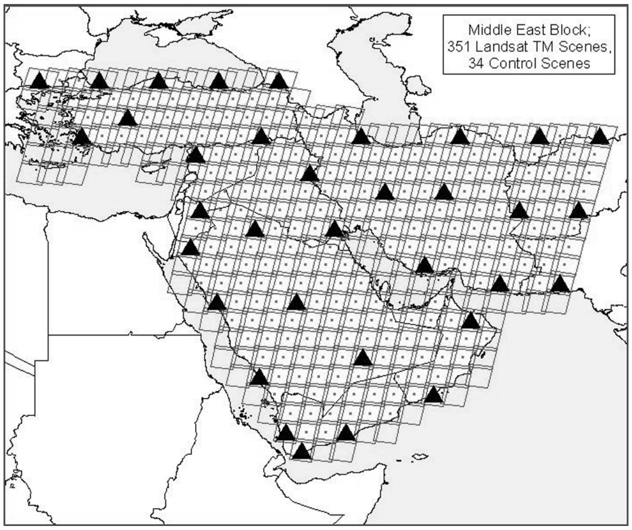

Figure 2. The Asia Minor adjustment block was comprised of 351 Landsat 4 and 5 scenes. A pixel-to-pixel correlation process was used to "tie" juxtaposed scenes together within their areas of overlap for several hundred-thousand common points. More than 350 geodetic control points were located manually on 34 TM scenes in this block as indicated by the black triangles (Δ). The appropriate digital elevation models were associated with the geodetic coordinates, and the entire block was then simultaneously orthorectified, resulting in an absolute positional accuracy of better than 50 meters root-mean-square error. See also Plates 3 and 4.

The first step in the orthorectification process was to assemble the satellite data in large blocks comprising 400 to 1,500 Landsat 4 and 5 scenes (Plate 3). A photogrammetric program called "Mospoly" (Earth Satellite Corporation patent number 6125329) was used to perform a pixel-to-pixel correlation to accurately acquire tie-points within the overlap between adjacent TM scenes for the block in question. A six parameter affine transformation was used to "tie" the images together in all areas of overlap. This included translation, rotation, differential scale, and affinity or "skew."

Once the images were "tied" together at common overlapping points in all overlapping areas, geodetic control points were located manually in the TM scenes where they occurred. The next step was to associate the respective digital elevation model data with the satellite data block (Plate 4) by matching geodetic coordinates. The digital terrain elevation data sets had a spatial resolution ranging between 3.0 and 30.0 arc seconds (30 m and 1 km, respectively). A typical block of data will contain over 300,000 tie points and over 300 ground control points (Figure 2).

---

*NASA does not endorse any private entities for providing services to individuals, groups, or other organizations.

The result of the triangulation process using the tie points and the ground control points is a mathematical mapping between the Earth's geoid and row and column space of the raster Landsat image. The precision orthorectified image is then produced by simply stepping through the ground space at the final geodetic projection while referencing the digital elevation model to project the *x*, *y*, and *z* location of the image pixel to a fractional pixel location within the output two-dimensional Landsat TM image. A pixel interpolation technique, either nearest neighbor or cubic-convolution, was then used to calculate the intensity values to assign to the image pixels within the final orthorectified image. In the case of the individual Landsat TM products, a nearest-neighbor interpolation was used to preserve as much of the original spectral information as possible. Each TM reflective band was resampled using a nearest-neighbor algorithm to 28.5-m resolution pixels; the associated thermal channel, when available, was resampled to 114 m. The final orthorectified TM reflective channel data have a root-mean-square (RMS) geodetic accuracy of better than 50 m.

In order to insure an accurate image-to-image among-epoch registration to the circa 1990 orthorectified data, the horizontal control was passed directly from the orthorectified Thematic Mapper imagery to the geometrically "raw" MSS 1970s epoch data. Vertical control was provided by the digital elevation model data. An automatic pixel cross-correlation approach was used to collect several hundreds of control points within each MSS scene. After clearing the parallax along scan lines for each point, a thin-plate spline approach was used to fit the MSS data to the *circa* 1990 epoch data. The MSS-to-ground mapping was then used to orthorectify the MSS data to a 57-m pixel using the same digital elevation model data as was used for the *circa* 1990 TM orthorectified data. This process was used for all four MSS bands and resulted in a root-mean-square error (RMSE) less than 100 m. The thin-plate spline approach was necessary because of the inherent geometric nonlinear ties of the 1970s epoch MSS data.

A "bundle" adjustment was used to orthorectify the *circa* 2000 Landsat 7 reflective spectral band data at a 28.5-m spatial resolution, using the same digital elevation model data as was used for the *circa* 1990 TM orthorectified data. Because of the improved internal mapping accuracy of the circa 2000 ETM+ imagery, horizontal control was provided by pixel correlation to orthorectify the *circa* 2000 data to the *circa* 1990 data. Vertical control was provided by the digital elevation model data. The Landsat 7 panchromatic band was orthorectified to 14.25 m while the thermal channel was orthorectified to 57 m. This process was used to geo-register all eight ETM+ bands and resulted in an RMSE less than 50 m. All scenes acquired during this project have been geo-registered and orthorectified so that each Landsat image has the high spatial accuracy required for quantitative land surface studies. Five data products were produced:

- Orthorectified MSS images, consisting of all four spectral bands, with a native 57-m pixel size (Table 1);
- Orthorectified TM circa 1990 data consisting of the six reflective and one thermal bands, at their native pixel resolutions of 28.5 and 114 m, respectively. No thermal data are available for a limited number of scenes (Table 2);
- Orthorectified ETM+ circa 2000 data, consisting of all spectral bands at their native spatial resolutions of 14.25, 28.5, and

Table 1. Landsat 1, 2, and 3 Multispectral Scanner Scene Product Specifications for the Mid-1970s Data Epoch. It Takes 8,190 MSS Scenes to Cover the Same Area as 7,600 Landsat 4, 5, and 7 Scenes, Due to Different Area Coverage. A Total of 7,550 MSS Scenes Were Acquired for 6,976 Unique Path-Rows (See also Plate 1a)

Number of scenes used	7,550 MSS scenes from 6,976 path-rows. A limited number of Landsat 4 and 5 MSS scenes were also used.
Number of spectral bands	All four MSS bands
Cloud cover	90% of input imagery will be 10% or less cloud cover
Image format	GeoTIFF
Resampling interpolation method	Nearest-neighbor (no interpolation)
Pixel size	57 m for all four bands
Projection	UTM zone determined by scene center
Datum/Spheroid	WGS 84
Horizontal control	Control passed from the orthorectified thematic mapper circa 1990 epoch imagery
Vertical control	U.S. Government provided 1-arc-second digital elevation model data where available; otherwise, GTOPO30 30-arc-second digital elevation model data
Positional accuracy	<100 meters, root-mean-square error

Table 2. Landsat 4 and 5 TM Scene Product Specifications for the circa 1990 Epoch. This Includes 7,413 Landsat 4 and 5 TM Scenes which Cover most of the Earth's Continental Areas. Copyright Restrictions for Landsat 4 and 5 Images Acquired within the Last Ten Years Were Negotiated for Unrestricted Distribution of the Final Orthorectified Image Products. However, the Europe and Thailand Station Operators Prohibited the Redistribution of the Raw Data Used as Input to the Orthorectification Process. These Restrictions Apply only to the Original Landsat Data and Do Not Apply to the Orthorectified Image Products. There Were 477 Scenes Missing from This Epoch for which no Data Were Found. Consequently, Data from Landsat 7 Were Used (See also Plate 1b)

Number of scenes used	7,413 scenes over 7,037 path/rows. Imagery acquired from the EROS Data Center and a wide collection of foreign ground stations
Number of spectral bands	All seven TM bands; a few scenes lack the thermal channel
Cloud cover	90% of input imagery will be 10% or less cloud cover
Image format	GeoTIFF
Resampling interpolation method	Nearest-neighbor (no interpolation)
Pixel size	28.5 m for the six reflective bands and 114 m for the thermal channel
Projection	UTM zone determined by scene center
Datum/Spheroid	WGS 84.
Horizontal control	U.S. Government provided control; sub-pixel accuracy
Vertical control	U.S. Government provided 1-arc-second digital elevation model data where available; otherwise, GTOPO30 30-arc-second digital elevation model data
Positional accuracy	<50 meters, root-mean-square error

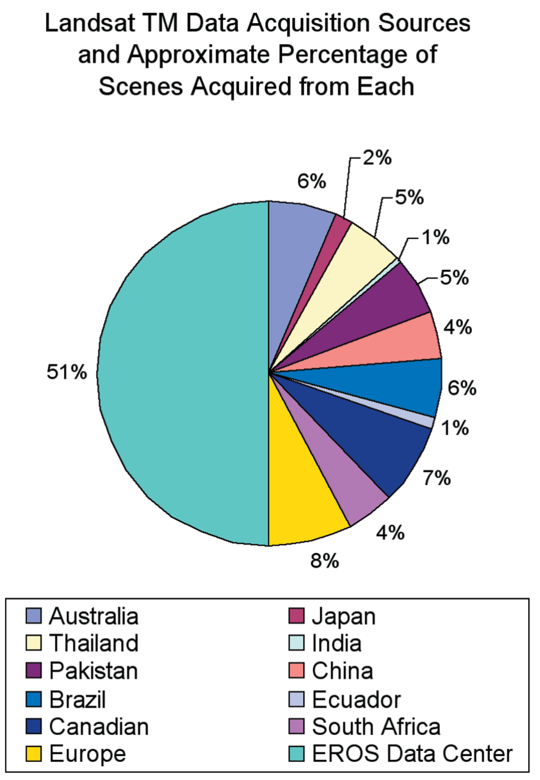

Plate 2. Breakdown of Landsat scenes by receiving station for the circa 1990 epoch. Without the assistance of the foreign Landsat receiving stations, it would have been impossible to obtain near global Landsat 4 and 5 thematic mapper coverage. See also Plate 1b for the global distribution of data for the circa 1990 epoch.

57.0 m, for the panchromatic, reflective, and thermal bands, respectively (Table 3);

- Three-band TM and three-band ETM+ Thematic Mapper "blended" mosaics covering areas of 5° by 6° for all continental landmasses for the circa 1990 epoch (28.5-m resolution) (Table 4) and circa 2000 epoch data (14.25-m resolution) (Table 5), respectively. The 14.25-m spatial resolution was achieved by panchromatic band "sharpening" of the three bands used in the circa 2000 epoch mosaics.

Mosaic Products

Five- by six-degree mosaics were created from both the circa 1990 epoch TM data with 28.5-meter pixels and the circa 2000 epoch ETM+ with 14.25-meter pixels (Tables 4 and 5). The 0.52- to 0.60-, 0.76- to 0.90-, and 2.08- to 2.35-μm bands for both the circa 1990 and circa 2000 epochs were used and were assigned blue, green, and red colors, respectively, to produce a three-band color image in GeoTIFF format. In order to eliminate possible scene-to-scene discrepancies in image quality, year, season, and cloudiness, a flexible color palate technique was developed, by which digital histograms of adjacent scenes were compared and "feathered." These data were projected into UTM coordinates and resampled to 28.5 m for the circa 1990 epoch and to 14.25 m for the circa 2000 epoch using a cubic-convolution process to enhance the visual quality of the mosaic (Koeln *et al.*, 1999) (Plate 5). In tropical regions with persistent cloud cover, color and intensity balancing between scenes is difficult and produces visible disjunctions between images. This is most pronounced if the number of available scenes is low for the area in question and data from widely different phenological times are used.

Data Validation

Once images were orthorectified and geodetic coordinates were assigned, all data were transferred to the the Stennis Space Center where they underwent a process of independent evaluation for data quality and geopositional accuracy. Data were ingested into a mass storage system where scene identification and scene metadata were verified. Software automatically produced image statistics on cloud cover using a reduced-resolution three-band composite of each scene. An analyst reviewed image statistics and also examined at full resolution all scenes flagged to have questionable data, significant amounts of data dropout, cloud contamination, or any other problems. After initial quality checks, geopositional accuracy statistics were assessed using a small sample (40 to 100 per data block) of independent ground control points to

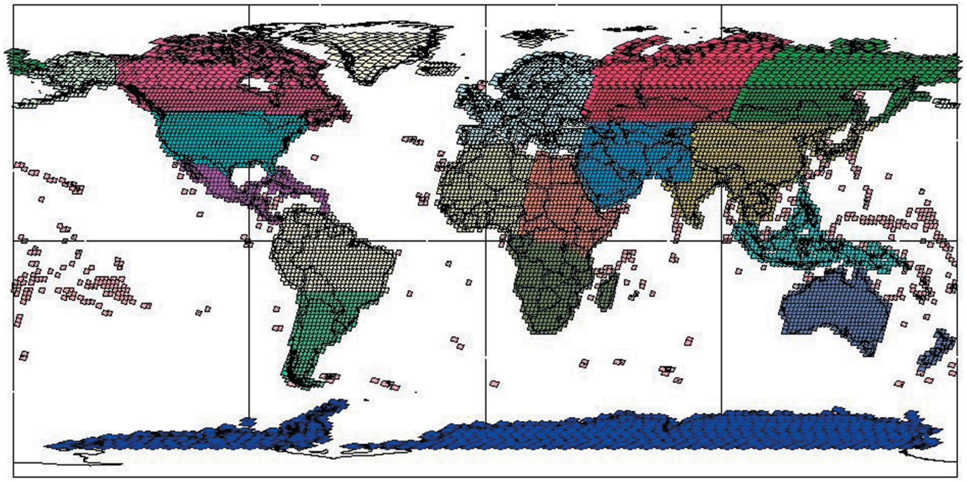

Plate 3. Regional aggregations of data were used to adjust the blocks of satellite data described in this paper. The number of Landsat scenes per block varied from 400 to 1,500. See Figure 2 for an example of one of the blocks.

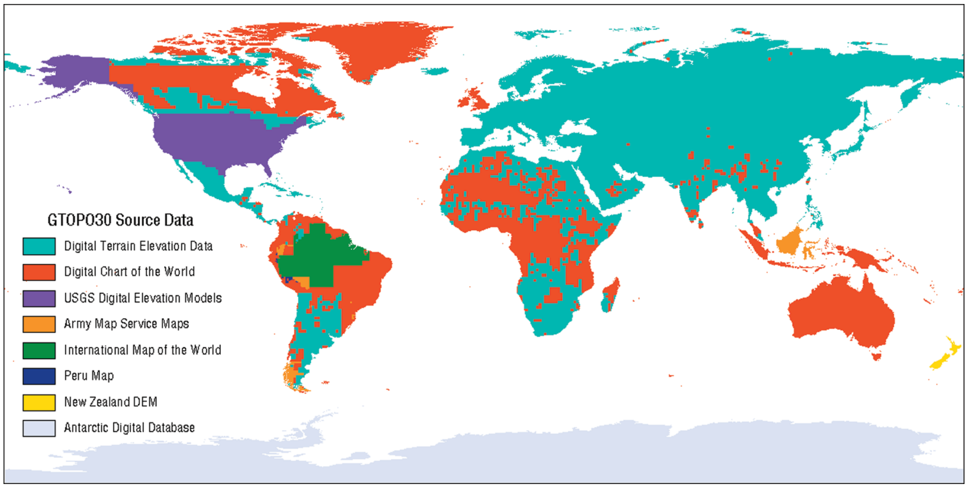

Plate 4. The digital elevation data used for the orthorectification process of all three global epochs consisted of a combination of 30-arc-second GTOPO30 (1-km) and, where available, a 3-arc-second (30-m) digital elevation model data. The light blue and dark purple areas in this figure indicate where the higher resolution 3-arc-second digital elevation model data were used.

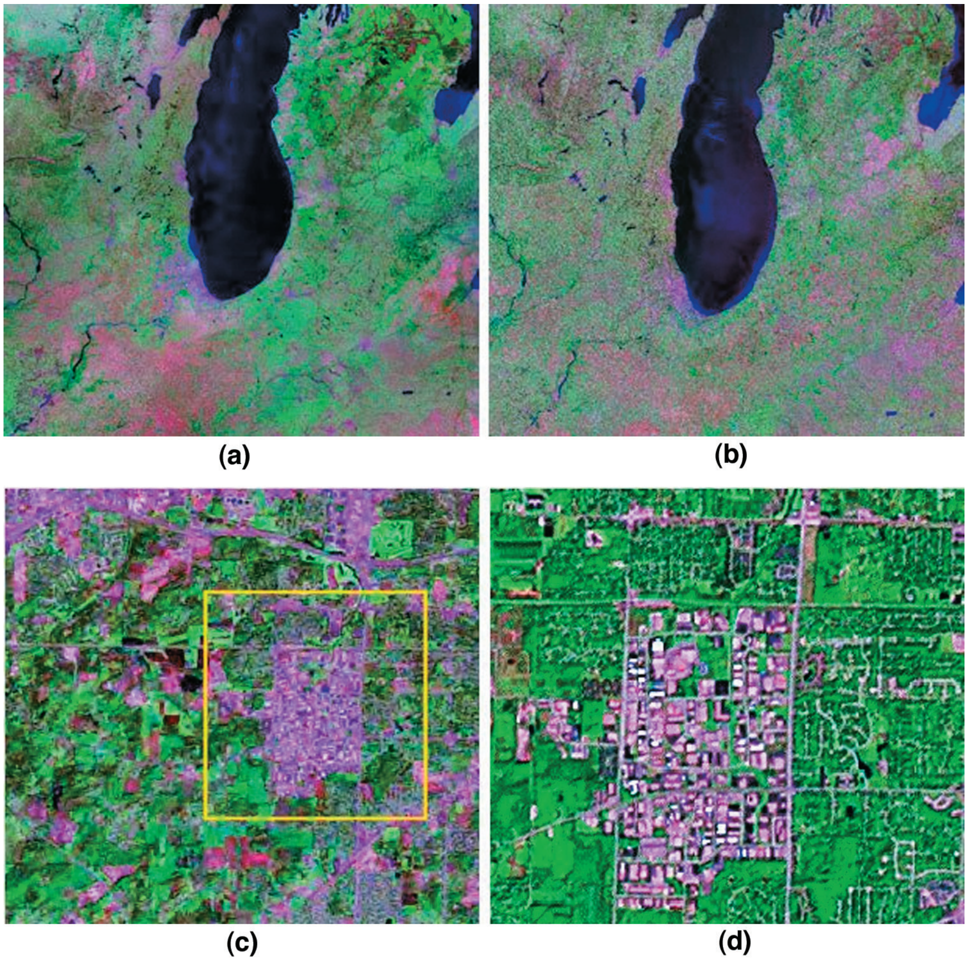

Plate 5. Circa 1990 epoch (a) and circa 2000 epoch (b) Landsat mosaic products over the southern portions of Lake Michigan and adjacent areas of the United States. A full-resolution 28.5-m subset of (a) for the circa 1990 epoch is shown as (c) and a full-resolution 14.25-m subset of (b) for the circa 2000 epoch is shown as (d).

TABLE 3. LANDSAT 7 ENHANCED THEMATIC MAPPER SCENE PRODUCT SPECIFICATIONS FOR CIRCA 2000 EPOCH. A TOTAL OF 8,500 LANDSAT 7 SCENES WERE OBTAINED FOR THE CIRCA 2000 EPOCH. AN ADDITIONAL 900 LANDSAT 7 PATH-ROWS WERE ACQUIRED OVER THE CIRCA 1990 EPOCH DATA, TO COVER MOST ISLANDS (SEE ALSO PLATE 1C)

Number of scenes used	8,500 scenes, all provided by the EROS Data Center
Number of spectral bands	All eight enhanced thematic mapper bands
Cloud cover	90% of imagery will have 10% or less cloud cover
Image format	GeoTIFF
Resampling interpolation method	Nearest-neighbor (no interpolation)
Pixel size	28.5 m for six reflective bands, 14.25 m for the panchromatic band, and 57 m for the thermal band
Projection	UTM zone determined by scene center
Datum/Spheroid	WGS 84
Horizontal control	Control passed from the orthorectified thematic mapper circa 1990 epoch imagery
Vertical control	U.S. Government provided 1-arc-second digital elevation model data where available; otherwise, GTOPO30 30-arc-second digital elevation model data
Positional accuracy	<50 meters, root-mean-square error

TABLE 4. ORTHORECTIFIED 5° BY 6° LANDSAT 4 AND 5 THEMATIC MAPPER MOSAIC PRODUCT. EACH MOSAIC COVERS AN AREA OF APPROXIMATELY 5 DEGREES OF LATITUDE BY 6 DEGREES OF LONGITUDE. MOSAICKED DATA ARE AVAILABLE IN BOTH GEOTIFF AND MRSID COMPRESSED FORMAT. THE HISTOGRAMS OF THE INDIVIDUAL SCENES WERE MATCHED AND THE OVERLAPPING AREAS BETWEEN SCENES FEATHERED TO CREATE A VIRTUALLY SEAMLESS MOSAIC. THESE DATA WERE PROJECTED TO A UTM PROJECTION WITH THE WGS84 DATUM AND RESAMPLED TO 28.5 M USING A CUBIC-CONVOLUTION PROCESS TO ENHANCE THE VISUAL QUALITY OF THE MOSAIC. EXAMPLES OF THESE MOSAIC DATA APPEAR IN PLATES 5A AND 5C

Number of scenes used	7,413 Landsat 4 and 5 thematic mapper orthorectified scenes used to derive mosaics
Number of spectral bands	Thematic mapper bands 7, 4, and 2 used as red, green, and blue colors, respectively
Cloud cover	90% of imagery will have 10% or less cloud cover
Image format	GeoTIFF or MrSid compressed
Resampling interpolation method	Cubic convolution
Pixel size	28.5 meters
Projection	UTM and blocked into 5 degree N-S partitions
Datum/Spheroid	WGS 84
Horizontal control	Same as Table 2
Vertical control	Same as Table 2
Positional accuracy	<50 meters, root-mean-square error.

TABLE 5. ORTHORECTIFIED 5° BY 6° LANDSAT 7 ENHANCED THEMATIC MAPPER MOSAIC PRODUCT. EACH MOSAIC COVERS AN AREA OF APPROXIMATELY 5 DEGREES OF LATITUDE BY 6 DEGREES OF LONGITUDE. THE HISTOGRAMS OF THE INDIVIDUAL SCENES WERE MATCHED AND THE OVERLAPPING AREAS BETWEEN SCENES WERE FEATHERED TO CREATE A VIRTUALLY SEAMLESS MOSAIC. THESE DATA WERE PROJECTED TO A UTM PROJECTION WITH THE WGS84 DATUM AND RESAMPLED TO A 14.25-M SPATIAL RESOLUTION USING A CUBIC-CONVOLUTION PROCESS TO ENHANCE THE VISUAL QUALITY OF THE MOSAIC. EXAMPLES OF THESE MOSAIC DATA APPEAR AS PLATES 5B AND 5D

Number of scenes used	8,500 Landsat 7 Enhanced Thematic Mapper orthorectified scenes used to derive mosaics
Number of spectral bands	Enhanced Thematic Mapper bands 7, 4, and 2 used as red, green, and blue colors, respectively
Cloud cover	90% of imagery will have 10% or less cloud cover
Image format	GeoTIFF or MrSid compressed
Resampling interpolation method	Cubic convolution
Pixel size	14.25 meters
Projection	UTM and blocked into 5 degree N-S mosaics
Datum/Spheroid	WGS 84
Horizontal control	Same as Table 3
Vertical control	Same as Table 3
Positional accuracy	<50 meters, root-mean-square error

estimate the overall geopositional accuracy of the *circa* 1990 epoch imagery. Accuracy estimations were then extrapolated to all images within the processing block in question.

MSS and ETM+ data geo-positional accuracy validation was similar to the procedure used to estimate geo-locational accuracy for the Landsat TM circa 1990 epoch product, except that the analysts had more ground control points to use in the computation of RMS error. The Earth Satellite Corporation did not directly use the NIMA-supplied ground control points for their MSS and ETM+ processing. These thus could also be used to verify geopositional accuracy of the earliest and latest Landsat data epochs. Control point documentation included path/row references, diagram sketches, and descriptive remarks. These control points were entered into a text file and converted to vector point coverage using an image processing software package. This UTM projected coverage served as reference data for evaluating overall accuracy. An analyst then located each of the ground control points in the respective Landsat scene, using all available information. If a ground control point was unable to be discerned, the point was then omitted from verification.

A first-order polynomial technique was used to determine the cumulative geolocation RMSE for points sampled across a number of images from a given data block. Only TM and ETM+ data blocks with an RMSE less than 50 m and MSS data blocks with an RMSE less than 100 m were accepted. Data not accepted were returned to the Earth Satellite Corporation for reprocessing.

The absence of a sufficient amount of ground control points in some blocks precluded this step of verification. The following blocks were not specifically evaluated for RMS geolocation error: Australia, North Africa, northeast Asia, southern South America, and Arctic regions 1, 2, and 3 (Plate 3). There was no geolocation accuracy assessment for the 5° by 6° mosaics because these data were derived from individual orthorectified scenes.

The spectral feathering and radiometric adjustments used to generate the satellite mosaics compromises some scientific uses. However, the mosaics have proven to be very useful for preliminary scientific analyses, education, visualization, and geographic information system "backdrop purposes." Adjacent scenes in the same UTM zone were geographically assembled and an analyst "hand-blended" image bands between and among scenes. This process maximizes spatial information at the expense of spectral integrity of pixels. By virtue of the seamless "feathering" between adjacent scenes within areas of overlap, it is frequently impossible to determine which pixel has come from which scene in areas of overlapping data.

Data Access and Use

Scene selection, acquisition, and processing of Landsat data as described were completed by the Earth Satellite Corporation (http://www.earthsat.com) of Rockville, Maryland, under NASA contracts NAS13-98046 and NAS13-02032. Through the terms of this contract, NASA has unrestricted right to copy, use, and distribute the orthorectified Landsat MSS, TM, ETM+, and mosaic products for scientific and educational purposes. Some "raw" data acquired for orthorectification have selective redistribution restrictions (see Table 2). Archiving and redistribution of scenes and mosaics will be done by the USGS EROS Data Center, Sioux Falls, South Dakota, and regional data centers supported by NASA. Data will be available to all requesters for a small fee to recover copying and mailing expenses. Large blocks of these data are also available from the University of Maryland's Global Land Cover Facility (http://glcf.umiacs.umd.edu/index.shtml).

Individual orthorectified Landsat 4 and 5 TM scenes include all seven bands and are in UTM coordinates based on the WGS84 datum. Landsat 7 circa 2000 data include eight bands. Each scene is available in a GeoTIFF format, a single file for each band. A typical GeoTIFF file is usually less than 60 megabytes, with the exception of the Landsat 7 panchromatic band, which is about 240 mb. Orthorectified Landsat 1, 2, and 3 MSS data include all four bands and are also projected into a UTM projection, similar to the TM data sets. Mosaics are distributed as either an uncompressed three-band GeoTIFF (approximately 1.5 gigabytes) or as a "MrSid" compressed file (www.lizardtech.com/solutions/geospatial). Mosaics are approximately 50 megabytes when compressed with the "MrSid" compression software.

Preliminary use of the data we describe indicates a high research potential of the 1970s, circa 1990, and circa 2000 Landsat data epochs with a high geodetic accuracy. Forest mapping work for all of Madagascar has found average positional accuracy to be better than 40 m for the 1990 and 2000 data layers and better than 80 m for the MSS data from the 1970s. Furthermore, very little time was needed to achieve coregistration among the three time periods (Steininger, personal communication, 2003). Variation between GPS coordinates and circa 2000 epoch ETM+ geodetic coordinates has been reported to average 20 m for 30 locations in Central America (Grant Harris, personal communication, 2003).

The potentially substantial time savings by avoiding extensive coregistration among three Landsat data epochs cannot be overstated while retaining excellent geodetic accuracy. The data set we describe represents a major achievement in the preservation and availability of global Landsat data for three time periods spanning almost 30 years. It is also the first time a globally consistent orthorectified satellite data set at a spatial scale of tens of meters with highly accurate within-scene and among-scene geodetic accuracy has been available to educators and researchers at minimal cost. Should this effort be repeated henceforth at five-year intervals, the international research community would continue to have unprecedented opportunities to monitor the land cover of our planet and document changes to it. The data set we describe is an important first step towards this end.

Acknowledgments

We thank Fritz Pollicelli, Kern Witcher, and Troy Frisbe of NASA/Stennis Space Center and Edwin Sheffner of NASA Headquarters for their work to ensure that the data set we describe came into existence.

References

Arvidson, T., J. Gasch, and S. N. Goward, 2001. Landsat-7's long-term acquisition plan—An innovative approach to building a global imagery archive, *Remote Sensing of Environment*, 78:13–26.

Becker, F., H. J. Bolle, and P. R. Rowntree, 1988. *The International Satellite Land-Surface Climatology Project*, ISLSCP-Secretariat, Free University of Berlin, Federal Republic of Germany, 100p.

Bryant, N. A., A. L. Zobrist, R. E. Walker, and B. Gokhman, 1985. An analysis of Landsat Thematic Mapper p-product internal geometry and conformity to Earth surface geometry, *Photogrammetric Engineering & Remote Sensing*, 51:1435–1448.

Desachy, J., G. Begni, B. Boissin, and J. Perbos, 1985. Investigation of Landsat-4 Thematic Mapper line-to-line and band-to-band registration and relative detector calibration, *Photogrammetric Engineering & Remote Sensing*, 51:1291–1298.

Goward, S. N., and D. L. Williams, 1997. Landsat and Earth system science: Development of terrestrial monitoring, *Photogrammetric Engineering & Remote Sensing*, 63:887–900.

Goward, S. N., J. Hasket, D. L. Williams, T. Arvidson, J. Gasch, R. Lonigro, M. Reely, J. Irons, R. Dubayah, S. Turner, K. Campera, and R. Bindschadler, 1999. Enhanced Landsat capturing all the Earth's land areas, *EOS Transactions,* 80(26):289–293.

IGBP, 1990. *The International Geosphere-Biosphere Programme: A Study of Global Change (IGBP)*, Report No. 12, The Initial Core Projects, Stockholm, Sweden, 330p.

Janetos, A. C., and C. O. Justice, 2000. Land cover and global productivity: A measurement strategy for the NASA programme, *International Journal of Remote Sensing*, 21:1491–1512.

Koeln, G. T., J. D. Dykstra, and J. Cunningham, 1999. Geocover and Geocover-LC: Orthorectified Landsat TM/MSS data and derived land cover for the world, *Proceedings, International Symposium on Digital Earth*, 29 November–02 December, Beijing, China (Science Press, Beijing, China), unpaginated CD ROM.

Malaret, E., L. A. Bartolucci, D. F. Lozano, P. E. Anuta, and C. D. McGillem, 1985. Landsat-4 and Landsat-5 Thematic Mapper data quality analysis, *Photogrammetric Engineering & Remote Sensing*, 51:1407–1416.

Rossow, W. B., and R. A. Schiffer. 1999. Advances in understanding clouds from ISCCP, *Bulletin of the American Meteorological Society*, 80:2261–2287.

Storey, J., and M. Choate, 2000. Landsat-7 on-orbit geometric calibration and performance, *Algorithms for Multi-Spectral, Hyperspectral, and Ultra-Spectral Imagery* (S.S. Chen and M.R. Descour, editors), Proc. of International Society for Optical Engineering, Bellingham, Washington, pp. 143–154.

Townshend, J. R. G., and C. O. Justice, 1988. Selecting the spatial resolution of satellite sensors required for global monitoring of land transformations, *International Journal of Remote Sensing*, 9:187–236.

———, 1990. The spatial variation of vegetation changes at very coarse scales, *International Journal of Remote Sensing*, 11:149–157.

Townshend, J. R. G., C. O. Justice, W. Li, C. Gurney, and J. McManus, 1991. Global land cover classification by remote sensing: Present capabilities and future possibilities, *Remote Sensing of Environment*, 35:243–256.

———, 1992. The impact of misregistration on the detection of changes in landcover, *IEEE Transaction on Geoscience and Remote Sensing*, 30:1054–1060.

Tucker, C. J., 1979. Red and near-infrared linear combinations for monitoring vegetation, *Remote Sensing of Environment*, 8:127–150.

Tucker, C. J., and P. J. Sellers, 1986. Satellite remote sensing of primary production, *International Journal of Remote Sensing*, 7:1395–1416.

WMO, 1992. *WCRP Scientific Plan for the GEWEX Continental Scale International Project (GCIP)*, Report Number WCRP-67, WMO/TD-461, World Meteorological Organization and the International Council of Scientific Unions, Geneva, Germany, 65p.

Welch, R., T. R. Jordan, and M. Ehlers, 1985. Comparative evaluations of the accuracy and cartographic potential of Landsat-4 and Landsat-5 Thematic Mapper image data, *Photogrammetric Engineering & Remote Sensing*, 51:1249–1262.

(Received 18 September 2002; accepted 14 January 2003; revised 28 March 2003).