Published in final edited form as:
Adv Parasitol. 2006 ; 62: 37–77. doi:10.1016/S0065-308X(05)62002-7.

Global Environmental Data for Mapping Infectious Disease Distribution

S.I. Hay[1,2], **A.J. Tatem**[1], **A.J. Graham**[1], **S.J. Goetz**[3], and **D.J. Rogers**[1]

[1] TALA Research Group, Tinbergen Building, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

[2] Malaria Public Health & Epidemiology Group, Centre for Geographic Medicine, KEMRI, P.O. Box 43640, 00100 Nairobi GPO, Kenya

[3] The Woods Hole Research Center, P.O. Box 296, Woods Hole, MA 02543-0296, USA

Abstract

This contribution documents the satellite data archives, data processing methods and temporal Fourier analysis (TFA) techniques used to create the remotely sensed datasets on the DVD distributed with this volume. The aim is to provide a detailed reference guide to the genesis of the data, rather than a standard review. These remotely sensed data cover the entire globe at either 1 × 1 or 8 × 8 km spatial resolution. We briefly evaluate the relationships between the 1 × 1 and 8 × 8 km global TFA products to explore their inter-compatibility. The 8 × 8 km TFA surfaces are used in the mapping procedures detailed in the subsequent disease mapping reviews, since the 1 × 1 km products have been validated less widely. Details are also provided on additional, current and planned sensors that should be able to provide continuity with these environmental variable surfaces, as well as other sources of global data that may be used for mapping infectious disease.

1. INTRODUCTION

The growth in the use of remote sensing (RS) and geographic information systems (GIS) has been fuelled, in part, by scientific demands to address many environmental issues at the global scale. The focus of RS and GIS in public health has been on infectious disease mapping (Hay *et al.*, 1997, 2000; Rogers *et al.*, 2002; Tatem *et al.*, 2004; Rogers, this volume, pp. 1–35). This has rarely been attempted at global scales primarily due to a lack of readily available environmental data for epidemiologists (Hay *et al.*, 1996, 1997; Hay, 2000). The RS data provided with this volume, and described in this review, go some way towards resolving this problem. These wide-area RS and other environmental data will be of utility to a variety of applications, but our emphasis here is entirely on infectious disease mapping.

Infectious diseases that have poikilothermic arthropod intermediate hosts are very sensitive to environmental conditions (Rogers and Packer, 1993). This is often highlighted by explaining sensitivities to temperature, rainfall and humidity of components of the basic reproductive number (R_0) of vector-borne diseases (Hay *et al.*, 1997; Rogers *et al.*, 2002; Rogers and Randolph, this volume, pp. 345–381), which mathematically describes the potential for parasite/pathogen persistence through time (Anderson, 1993). The RS-based data reviewed here have been developed to obtain information or “surrogates” of the more traditional climatic variables of relevance to the transmission of vector-borne diseases and thus infectious disease mapping (Hay and Lennon, 1999; Goetz *et al.*, 2000; Hay, 2000; Green and Hay, 2002). We focus primarily on the Advanced Very High Resolution Radiometer (AVHRR) which, more than any other satellite sensor, has found considerable application in large area epidemiology (Hay, 2000; Rogers, this volume, pp. 1–35). We do

Europe PMC Funders Author Manuscripts

not describe the various applications of these data as this is done elsewhere (Hay *et al.*, 2000; Rogers *et al.*, 2002) and is amply illustrated throughout this volume.

2. THE AVHRR SENSOR

The basic principles of RS and satellite sensor systems have been reviewed previously (Hay *et al.*, 1997; Hay, 2000; Tatem *et al.*, 2004). Here we provide a brief overview of the AVHRR sensor, a series of which have collected the RS data distributed on the accompanying DVD. It is beyond the scope of this review to extend this treatment to other sensors.

2.1. History and Overview

The National Oceanographic and Atmospheric Administration's (NOAA) series of polar-orbiting Television Infrared Observation Satellites (TIROS) has been operational since 1978 (Cracknell, 1997). TIROS-N (later renamed NOAA-6) was the first satellite to carry the AVHRR, originally for meteorological purposes, and has been followed by eleven satellites, each with an operational lifetime of 2–4 years. The definitive description of the NOAA polar-orbiting satellites, their radiometer payloads and the data they generate are given in Cracknell (1997) and Kidwell (1998). Information on the NOAA-AVHRR satellite series is updated regularly by theNOAA Satellite and Information Services (NOAASIS) [http://noaasis.noaa.gov/].

The NOAA-AVHRR satellites complete 14.1 near-polar, Sun-synchronous orbits every 24 hours at an altitude of 833–870 km. The NOAA-AVHRR can view a 2800 km swath of the Earth and so, at this orbital frequency, daily data are recorded for the entire Earth surface. Radiation is currently measured in six wavebands (channels) of the electromagnetic spectrum (five on platforms prior to NOAA-15) so that six images are recorded for each area sensed. The visible (channel 1) and near infrared (channel 2) measure reflected solar radiation, whereas the thermal channels (4 and 5) measure emitted thermal infrared. Channel 3 (split into 3A and 3B on the latest platforms) senses the middle-infrared (MIR) and is sensitive to a combination of both reflected and emitted radiances.

The nominal 1.1 × 1.1 km spatial resolution data are transmitted continuously and may be received by stations near to the satellite's path, where they are referred to as High Resolution Picture Transmission (HRPT) data (Cracknell, 1997). On request to NOAA, these data may also be recorded on an on-board tape storage system and later transmitted to Earth as the satellites pass over a network of receiving stations. The data are then referred to as Local Area Coverage (LAC). These data have found application in a very wide range of disciplines that have been reviewed by Ehrlich *et al.* (1994) and Cracknell (1997).

Two processing steps reduce the spatial resolution of most of the NOAA-AVHRR data available to the research community. Since the on-board tape system is incapable of holding global coverage data at 1.1 × 1.1 km spatial resolution, the information from each area of five (across-track) by three (along-track) pixels is stored as a single value corresponding to the average of the first four pixels only, of the first row of the 5 × 3 block. The resulting imagery is referred to as Global Area Coverage (GAC) data. GAC data, with a stated nominal spatial resolution of 4 × 4 km; are far from ideal representations of the raw data (Justice *et al.*, 1989; Robinson, 1996). Nevertheless, the GAC data are the only form in which the NOAA-AVHRR archive was and continues to be collected. Reasonable quality global datasets are available at a variety of spatial resolutions (4 × 4 km and coarser) starting in the early 1980s (James *et al.*, 1994). When further processed, these data are often re-sampled to an 8 × 8 km spatial resolution before distribution (James *et al.*, 1994).

2.2. AVHRR Archives

A series of 1 × 1 km spatial resolution NOAA-AVHRR data are available at the "Global Land 1-KM AVHRR Project" homepage on the United States Geological Survey (USGS), EROS Data Center, website [URL: http://edcdaac.usgs.gov]. The dataset was generated to obtain a standard year of observations for the global land cover mapping project by the International Geosphere Biosphere Programme – Data and Information System (IGBP-DIS) (Eidenshink and Faundeen, 1994; Townshend *et al.*, 1994; Teillet *et al.*, 2000).

The 8 × 8 km spatial resolution NOAA-AVHRR data are available at the Global Land Biosphere Data and Information Web Site at the Goddard Space Flight Center's Distributed Active Archive Center [http://daac.gsfc.nasa.gov/]. The data were archived with the purpose of providing a long-term database for Earth observation, with a particular emphasis on tropical deforestation (Townshend, 1994). These images are referred to as the Pathfinder AVHRR Land (PAL) dataset (James *et al.*, 1994).

2.3. From Digital to Environmental Data

2.3.1. Temporal Range—The 1 × 1 km NOAA-AVHRR data are available by decad (10 day unit) from April to December 1992 (9 months, 27 decads, 162 files); January–September 1993 (9 months, 27 decads, 162 files); February–December 1995 (11 months, 33 decads, 198 files) and January–April 1996 (2 months, 6 decads, 36 files) (see Table 1).

The 8 × 8 km NOAA-AVHRR data are available from August 1981 to September 2001 inclusive; a 20-year time series. While the 1 × 1 km data are obviously more detailed spatially, the longer time series of the 8 × 8 km imagery provides opportunities for examining changing land-surface patterns over a critical period of the Earth's history. These issues are explored in detail in Sections 3.2 and 3.3 respectively.

2.3.2. Data Obtained and File Sizes—For each decad of the 1 × 1 km NOAA-AVHRR the following parameters were downloaded: channels 3, 4, 5, the normalized difference vegetation index (NDVI), satellite and solar zenith angles (see Hay, 2000 for definitions). This amounts to ~680 gigabytes (Gb) of compressed data. When each global image is uncompressed it is 17 347 lines/rows by 40 031 samples/columns. For 8-bit imagery the file size is therefore (17 347 × 40 031 = 694 417 757 bytes) or ~678 Mb. For 16-bit data the file size is ~1.36 Gb.

The global 8 × 8 km data, in contrast, are considerably smaller, at 2168 lines/rows by 5004 samples/columns, and so for 8-bit data the files (2168 × 5004 = 10 848 672 bytes) are ~10.6 Mb and double that for 16-bit files. The 20-year duration of the 8 × 8 km data archive makes the total multitemporal database close to 3.5 Gb when compressed. The spatial resolution of the 1 × 1 km NOAA-AVHRR data and the temporal duration of the 8 × 8 km NOAA-AVHRR data result in such large volumes of data, that distribution on DVD in their original form is not possible, an issue addressed in Section 3.

2.3.3. Geo-Registration and Projection—All RS data represent conditions on a spherical Earth as a regular, gridded raster array of picture elements (pixels). Capturing features of a 3-dimensional object (the globe) on a 2-dimensional map involves some compromise in accurately representing direction, distance, shape and area (Snyder and Voxland, 1994; Snyder, 1997) and different map projections are used for applications that need to show one or other of these map features as accurately as possible. The projection chosen for both the 1 × 1 km and 8 × 8 km NOAA-AVHRR data at source is the Interrupted Goode's Homolosine, which is a combination of a Sinusoidal projection covering the tropical regions of the Earth (to latitudes of 40°44′ North and South) and the Mollweide

projection elsewhere. This combination of projections was considered to be the best at maintaining the shape and area of the continents (Steinwand, 1994). Despite the geographical superiorities of the Goode's projection the majority of potential non-specialist RS or GIS users, and the software with which they are familiar, use latitude and longitude data in an equi-rectangular grid technically known as the Plate-Carrée projection. After temporal Fourier processing (see Section 3) of the data in their original projection we resampled the resulting files using a bilinear interpolation algorithm (Mather, 1999) that ignored sea pixels along coastlines (i.e. sea pixels were given weights of zero in the interpolation algorithm), and distribute both 1 × 1 and 8 × 8 km NOAA-AVHRR products in the Plate-Carreé projection, at nominal resolutions of 0.01° and 0.1°, respectively. The resulting images are therefore 36000 columns × 18 000 rows or 3600 columns × 1800 rows, respectively. Table 2 gives the full details that are required for manipulating these data in RS and GIS software packages.

2.3.4. Rescaling—Values of geophysical variables are usually rescaled before storage as image files, and represented as either 8-bit (i.e. byte) or 16-bit (i.e. integer) binary numbers. This saves on hard disk storage space (a limiting component in earlier computers) and facilitates image compression, which further saves storage space. Eight-bit numbers are always positive, and span the range from 0 to 255. Sixteen-bit numbers may be stored either in signed 16-bit format (range between −32 767 and +32 767) or unsigned 16-bit format (range between 0 and 65 535). Much confusion arises when unsigned 16-bit integers (exceeding +32767) are regarded as signed integers, and vice versa, and users must take care to specify to their software the correct format (i.e. 'signed' or 'unsigned') of 16-bit integer image data. Parts of the full digital range of stored data are often 'reserved' for mask values (indicating the sea, image quality, or some detail of the map projection used) and these should be specified by the providers of the imagery.

The original images were stored at source in either 8-bit or 16-bit format, with certain values reserved to indicate masks, sea or projection interruptions. Full details of this conversion process for the 1 × 1 km imagery are given as examples. The satellite zenith angle (SaZA) data have units of degrees and were stored in 8-bit files. In the original files, the binary min/max=10/190 were used to represent the geophysical min/max=−90/90 and mask values were 0=missing data over land; 1=ocean; 2=Goode's interrupted area. The satellite zenith angle was calculated from SaZA=(DN–100) where DN is the original stored 8-bit number. The solar zenith angle (SoZA) also has units of degrees and was originally stored as 8-bit files. The binary min/max were 10/190 and the geophysical min/max were 0/180. The mask and other values were exactly the same as for the SaZA imagery, so SoZA=(DN–10). The NDVI is a ratio and has no units. It was archived as 8-bit files, the binary min/max of which were 10/210 with the geophysical min/max=−1/1 and mask and other values as before; so NDVI=(DN–110) × 0.01. Finally, the (channels radiance data 3, 4 and 5) had units of degrees Kelvin and were originally stored as unsigned 16-bit data with binary min/max=10/1018 and the geophysical min/max=160/340. Mask and other values were as before, so channel 3, channel 4 or channel 5=(DN+886.32)/5.602.

In general, before the processing steps outlined below, the imagery was stored in signed 16-bit format with the mask and other values (0, 1, 2) being stored as −999, −998 and −997, respectively. The 8 × 8 km NOAA-AVHRR data were archived similarly with further details and (different) rescaling values given by James *et al.* (1994).

2.3.5. Quality Control and Data Pre-Processing—Ancillary data provided with the original imagery were used to exclude unreliable pixels in both the 1 × 1 and 8 × 8 km datasets. Quality control removed pixels taken at satellite zenith angles (SaZA) greater than 40°, to reduce bi-directional effects and parallax errors due to image pixels acquired at

viewing angles far from nadir (Hay and Lennon, 1999; Hay, 2000). Pixels with solar zenith angles (SoZA) exceeding 80° were also excluded, to eliminate data retrieved at low sun elevations (i.e. at dusk and dawn) that result in shadows that affect image quality and therefore interpretation (Hay and Lennon, 1999; Hay, 2000). In addition, the 8 × 8 km NOAA-AVHRR data included a data layer that could be used to mask those pixels determined as cloudy by the 'Clouds from AVHRR' (CLAVR) algorithm (Stowe *et al.*, 1991; Hay and Lennon, 1999).

The NDVI data were then maximum-value composited (Holben, 1986) into monthly files. The (incomplete) set of monthly 1×1 km NOAA-AVHRR data for 1992–1996 were further subjected to maximum value composition to generate a single set of 12 monthly images representing the average (i.e. synoptic) year using all the available data. The particular files contributing to each synoptic month are shown in Table 1. These data were then rescaled conditionally (if ≤ −996 = 0, else = ((NDVI + 1) × 1000 and stored as signed 16-bit integers ready for temporal Fourier analyses (TFA).

MIR (channel 3) data were also maximum-value composited in the same way (Lambin and Ehrlich, 1996). The subsequent processing procedure was identical to that applied to the NDVI. Data values indicating a temperature estimate of 350 K or greater were screened out at this stage, as they were clearly in error. These data were then conditionally rescaled (if ≤ −999 = 0, else = (MIR × 10 for storage as signed 16-bit integers.

Land surface temperature (LST) data were derived using a split window algorithm (Price, 1983), using quality-controlled channel 4 and 5 data and maximum-value composited (Lambin and Ehrlich, 1996). Subsequent processing was identical to that of the channel 3 data. These data were conditionally rescaled (if≤ −990=0, else=(LST × 10)) for storage as signed 16-bit integers.

The 1 × 1 and 8 × 8 km datasets were not corrected systematically for the effects of satellite orbit drift over the lifespan of each satellite, which can affect the inter-comparability of these multitemporal data (Gutman, 1999). This can be particularly problematic in the thermal channels because satellite orbit decays progressively delay the timing of the daily measurement, which is obviously critical for a variable such as temperature that shows strong diurnal variation (Gleason *et al.*, 2002). For these and other reasons, the AVHRR data for 1981 and all data after 1999 were excluded from the 8 × 8 km NOAA-AVHRR time series before TFA (Nemani *et al.*, 2003). Many small artefacts in satellite data are smoothed by TFA (see Section 3) but we emphasize that these uncorrected data are not appropriate for the analyses of temporal change. Datasets that systematically deal with such artefacts for monitoring temporal change are becoming available (Tucker *et al.*, 2005). Finally, only NDVI, MIR and LST data products are processed and distributed here. Although it is possible to derive air temperature and vapour pressure deficit (Goetz *et al.*, 2000), we have avoided these indices as they have been tested less widely, can require further geophysical data inputs and show co-linearity with these existing products, a number of which are used in their generation. Data provided on the accompanying DVD are in the format and with the new scaling as outlined above, with further details in Table 3 and the DVD README file.

3. TEMPORAL FOURIER ANALYSIS (TFA)

3.1. History and Application

Monthly composite imagery usually shows strong serial correlations (i.e. seasons wax and wane in a relatively predictable manner) and this data redundancy may be eliminated in two different ways. Traditionally, the data are subjected to principal components analysis (PCA), and the resultant significant principal components are used in analyses (Eastman and Falk,

1993; Lillesand and Kiefer, 2000). Alternatively, the data may be subjected to TFA, which describes natural environmental cycles such as temperature and vegetation growth in terms of annual, bi-annual, tri-annual and other cycles with shorter or longer wavelengths (periods) (Rogers, 2000). The great attraction of TFA is that it produces a set of orthogonal (i.e. uncorrelated) outputs while retaining a description of seasonality (lost in PCA) that is of vital interest in vector and disease mapping (Rogers *et al.*, 1996; Rogers, 2000; Rogers and Robinson, 2004). One disadvantage of both PCA and TFA is that they both assume stationarity (i.e. constant mean and variance) of the data over time. Trends in data can first be removed by differencing the time series from a moving average spanning a number of annual cycles, and then analysing the de-trended time series. Alternatively, if the trend itself is of interest, a windowing approach can be applied, whereby the data are analysed in a series of (usually overlapping) windows in time. Within each window the data are assumed to be mathematically stationary, and the windowed TFA results can be compared to look for changes in the Fourier components (means, amplitudes, phases etc.) through time.

The origins, mathematical basis and arguments for the biological appropriateness of TFA are developed in detail elsewhere (Rogers, 2000; Rogers *et al.*, 2002; Rogers and Robinson, 2004). In brief, the trajectory through the year (the sequence of 12 monthly images) of every picture element (pixel) in the environmental time series is described by a series of orthogonal sine curves (cycles/harmonics) with different frequencies. Each cycle is described by its amplitude (the maximum variation of the cycle around the mean) and phase (its timing). A total of six cycles is required to describe a dataset of 12 months perfectly (with cycle periods of 12, 6, 4, 3, 2.4 and 2 months). We find that around 90% or more of the variation in the original images is usually captured by the first three (i.e. annual, bi-annual and tri-annual) cycles, and much of the remainder is often simply ‘noise’, so we normally use only these to capture important features of the seasonal variation at each site.

During TFA the raw time series of data were first examined for obvious drop-out values, that arise either from absent imagery or else from pixels masked out in the quality control algorithm of image processing. Data for such months were linearly interpolated from data for the months before and after the dropout months (with data wrap-around when dropout months occurred right at the start or right at the end of the sequence). The resulting time series were subjected to TFA and the raw data were then compared with the recomposed result (i.e. the sum of the Fourier-fitted annual, bi-annual and tri-annual cycles only). Months where the absolute difference between the raw and recomposed data exceeded user-determined threshold values were also regarded as incorrect (these threshold values were generously set so that only obvious outliers would be trapped at this stage). These months were therefore linearly interpolated as before, using data from adjacent months, and TFA carried out again on the corrected data. This process was repeated until no further outliers were identified (generally requiring no more than one or two rounds of interpolation of a few data points). Finally, no 8 × 8 km imagery was available for the months of September–December 1994. Images for these months were first created using the averages of the same months from the 1993 and 1995 images, and these were used in TFA.

3.2. Fourier Data Products

The TFA algorithm developed by the TALA group produces 14 different products for each input satellite channel; the overall mean (a0); the amplitude of the annual (a1), bi-annual (a2) and tri-annual (a3) cycles; the phase (i.e. peak timing) of the annual (p1), bi-annual (ph2) and tri-annual (p3) cycles in months (starting at zero in January); the proportion of the variance in the original time series that is described by the annual (d1), bi-annual (d2), tri-annual (d3) and all three cycles combined (da); the maximum (mx) and minimum (mn) of the seasonal cycle recomposed from the first three harmonics only; and finally the variance (vr) of the original (i.e. not the fitted) time series. These products are made available in full

for the 8 × 8 km imagery but only a subset is given for the 1 × 1 km products due to space constraints (see DVD README file). Specifications for each 8 × 8 km NOAA-AVHRR TFA file are given in Table 3 and examples for each band given in Figure 1a–n. Note that equivalent details for the 1 × 1 km NOAA-AVHRR TFA data are only in the DVD README file. Figure 2a and b provides a vignette of the processed 1 × 1 and 8 × 8 km imagery for the lower reaches of the Nile river in Egypt, Africa. The following section briefly discusses the inter-comparability of these data.

3.3. Inter-Comparison of TFA Surfaces

We have not assumed that 1 × 1 and 8 × 8 km TFA surfaces can be used interchangeably, because of the rather different origins, time periods and processing chains to which the data were subjected. Regression tests between 1 × 1 and 8 × 8 km surfaces were therefore performed at different latitudes to examine whether the overpass times of the various satellites, and varying solar and viewing illumination levels produced regional effects. The global area represented by the NOAA-AVHRR imagery was divided into eight latitude bands: 90–70°N, 70–50°N, 50–30°N, 30–10°N, 10°N–10°S, 10–30°S, 30–50°S, 50–90°S (see Figure 3). The 1 × 1 km NOAA-AVHRR images were resampled using a mean filter to match the spatial resolution of the 8 × 8 km imagery. A stratified random sample of 10000 points within each region was created and values of the pixels representing these points extracted from the mean, maximum and minimum MIR, LST and NDVI. Figure 4a–f shows a sample of these scatter plots of mean MIR, LST and NDVI 8 × 8 km pixel values against mean MIR, LST and NDVI 1 × 1 km pixel values for the 50°N–30°N and 10°N–10°S latitudinal segment. Regression analysis was carried out between the two sets of extracted pixel values. No significant differences were observed between the fit of a linear model (see Table 4) and various non-linear alternatives.

Figure 4 highlights scatterplots for two of the most epidemiologically important latitudinal segments; the 50–30°N segment encompassing the US, Southern Europe and much of Asia, and the 10°N–10°S segment encompassing the equatorial tropical regions of the World. Figure 4a shows that there is correspondence between the 1 × 1 and 8 × 8 km imagery for all three environmental variables at 50–30°N, with the vast majority of sample pixels clustered along the one-to-one lines (i.e. indicating little bias). These results are reflected in Table 4, showing r^2 values above 0.69 for all three variables. An area of concern is the clusters of rogue NDVI pixels which exhibit very low values in the 1 × 1 km NOAA-AVHRR imagery. These represent small water-bodies and coast pixels not distinguishable at the 8 × 8 km scale (but certainly affecting the values of the pixels at this scale). Figure 4b shows that, for MIR and NDVI, little bias between the 1 × 1 km and 8 × 8 km imagery exists at tropical latitudes, and that large differences again exist in a small handful of NDVI pixels. A more significant concern is that the 8 × 8 km NOAA-AVHRR LST data are showing apparently higher temperature than the corresponding 1 × 1 km imagery. This may be an artefact of the difference in time periods over which each TFA product was produced, with the 8 × 8 km imagery providing a more reliable synoptic estimate given the much greater numbers of contributing years. Table 4 shows that, despite this bias, at tropical latitudes all three mean environmental variables exhibit relatively strong correlations (all r^2>0.58). We have focused on the 8 × 8 km imagery in our studies, but we have included a smaller subset of the 1 × 1 km imagery for wider experimentation which should, nevertheless, bear these comparisons in mind.

Europe PMC Funders Author Manuscripts

4. FUTURE GLOBAL ENVIRONMENTAL DATA

4.1. Terra, Aqua and Modis

The Terra (EOS AM-1) and Aqua (EOS PM-2) satellites, launched in December 1999 and May 2002 respectively, are a part of (NASA) Earth Observing System (Parkinson, 2003). A range of onboard sensors capture a variety of image types, but most relevant in this context are the MODerate Resolution Imaging Spectroradiometer (MODIS) and, exclusively for Terra, the Advanced Spaceborne Thermal Emission and Reflection (ASTER) radiometer.

MODIS is particularly attractive for epidemiological applications due to; (i) a better spectral resolution than AVHRR, with 36 spectral channels with smaller waveband ranges and significantly improved signal-to-noise ratios (Justice *et al.*, 2002), (ii) a one to two-day repeat time temporal resolution at significantly higher spatial resolution (250 × 250 to 1000 × 1000 m depending on the channel) than AVHRR (Townshend and Justice, 2002) and (iii) fully processed and quality assessed data products, giving unparalleled, rapid access to contemporary and reliable data on large-area ecosystem processes. MODIS is also potentially attractive to the public health community thanks to the availability of its products at no charge to users, and its longer mission lifespan (Tatem *et al.*, 2004). A drawback of MODIS data, when compared to AVHRR, is the considerably greater computing resources needed to cope with the larger data volumes of some of its products. These constraints are likely to diminish rapidly with the exponential increases in computer power and storage capacities, but these spatial resolutions at the global scale will still present a significant challenge to the majority of users.

ASTER is an alternative data source for studies that have traditionally used SPOT-HRV or Landsat TM sensors (see below). The spatial resolution of ASTER varies with wavelength, yielding 15, 30 and 90 m resolutions at visible–near infrared (VNIR), short wave infrared (SWIR) and thermal infrared (TIR), respectively (Yamaguchi *et al.*, 1998), and the images may prove a powerful tool for studying local disease processes (Tatem *et al.*, 2004).

Spectral bands within the MODIS 36 band array are broadly similar to those of AVHRR, which may permit temporal continuity of datasets and thus the potential for extending the AVHRR time series (Friedl *et al.*, 2002), providing funding for the Terra and Aqua satellites continues (a significant consideration given the uncertainty of most USA-based Earth observation systems (Lawler, 2005)). Likewise, ASTER uses spectral channels similar in characteristics to those of Landsat TM. Significant ongoing efforts in RS are seeking to overcome issues that arise from differences in radiometric and spatial resolutions between old and new sensors (NOAA – NPOESS, 2003). We develop these issues further in Section 4.3.

4.2. Other Satellite Sensors

This section focuses on those sensors that can provide information at the global scale. Many sensors that have not proven useful in epidemiological studies and those with evident constraints, such as difficulties of image costs or data access, are not considered. A more complete account of satellite sensors is available in Campbell (2002) and Verger *et al.* (2003). There are many planned enhancements to the existing sensors which, due to the frequency of modification, are best reviewed at their relevant web sites. These, and the basic technical specifications of the sensors reviewed here, are detailed in Table 5.

4.2.1. Geostationary Satellites—The principal payload of Meteosat is the Meteosat Visible and Infra-Red Imager (MVIRI) (EUMETSAT, 2000). The radiometer operates in a broad visible waveband, a water vapour absorption infrared waveband and a thermal

infrared waveband. The Meteosat satellites were designed for meteorological applications, so part of their spectral range is located in the thermal infrared area of maximal water vapour absorption, making it ideal for monitoring clouds. At nadir the spatial resolution is 2.5 × 2.5 km for the visible images and 5 × 5 km for the thermal infrared and water vapour images. Further from the equator, the spatial resolution decreases so that over northern Europe, for example, it is 4 × 4 km in the visible wavebands and 8 × 8 km in the thermal infrared and water vapour wavebands. Each image is transmitted to the Earth in real time as each scan line is completed, and new images are generated at 30-min intervals.

Meteosat Second Generation (MSG) satellites are designed to give continuity to Meteosat missions with improved spatial, spectral and temporal resolutions (EUMETSAT-ESA, 1998; EUMETSAT, 2001; Schmetz *et al.*, 2002). MSG-1 was launched in August 2002 and, after a period of commissioning and validation, routine operations started in January 2004 (EUMETSAT, 2004). Of particular potential on board this satellite is the Spinning Enhanced Visible and Infrared Imager (SEVIRI) sensor, designed to acquire images every 15 min from 12 different bands at a spatial resolution of 1.4 × 1.4 km for the visible, and 3 × 3 km for all other bands. MSG high-quality datasets have raised expectations as powerful tools for studying temporal and spatial tropical disease patterns in Africa (Hay, 2000). These have largely not been realized due to data archival and distribution restrictions in Europe.

It is important to stress that despite the fact that Meteosat products have found wide application in malaria studies in Africa (Hay *et al.*, 1998, 2003b; Rogers *et al.*, 2002), they cannot be used in the Americas, Oceania or much of Asia as the satellite's sensors capture only that part of the Earth's disc visible from a stationary orbit positioned over the equator at the Greenwich meridian (i.e. longitude 0°). This area includes Africa, Europe and the Middle East. The satellites with equivalent capabilities for the American and Pacific region are NOAA's Geostationary Operational Environmental Satellite (GOES) (NASA, 1999). The satellite stationed at 75° W is known as GOES-E while that at 135°W is called GOES-W. GOES has been used mainly for weather observations and forecasts for the last 25 years (NASA, 1999), and its products offer great potential for epidemiological applications. The general properties of GOES-12 imagery are included in Table 5 for reference.

Two other satellite systems complete the constellation of five satellites that provide geostationary sensor coverage for the entire globe. Indian Ocean Data Coverage (IODC) is provided by Meteosat 5, which has found continued use in a different geostationary orbit position. The sensor details have already been discussed in this section. The final satellite and sensor to mention is the Geostationary Meteorological Satellite (GMS) series which sits in orbit above 140°E. GMS-1 was launched in 1977, the first in a series of satellites culminating in GMS-5 which completed its observational mission in May 2003. GOES-9 has been used as a backup to GMS-5 that had to shut down its imager due to a technical fault. The replacement for the GMS series (and GOES-9) is a new Japanese satellite called the Multi-Functional Transport Satellite (MTSAT-1R), which became operational in 2005 and is to be followed by MTSAT-2. These satellites carry sensors that image in 5 wavebands: visible, near-infrared, two infrared channels and a water vapour channel. The visible images have a spatial resolution of 1 × 1 km while all other channels are sensed at 4 × 4 km spatial resolution at nadir.

4.2.2. Landsat—The launch of Landsat-1 in 1972 heralded a new era of high resolution RS (Lauer *et al.*, 1997; Markham *et al.*, 2004). Since then, the Landsat programme has generated a continuous supply of high resolution imagery for the entire globe, from the first Multispectral Scanner (MSS) aboard Landsat-1 to the latest Enhanced Thematic Mapper (ETM+) on board Landsat-7 (Mika, 1997). During this time, there has been a substantial evolution in the quality of the radiometers (Mika, 1997), their calibration (Chander *et al.*,

2004; Thome *et al.*, 2004) and the development of multi-spectral data analysis techniques developed to process captured data (Landgrebe, 1997). The novelty and conspicuous success of the Landsat programme forced issues regarding data distribution and cost (Draeger *et al.*, 1997) and the feasibility of commercial RS (Williamson, 1997) to be considered seriously for the first time. Many countries have emulated and extended features of the Landsat programme, and other high-resolution RS data sources are becoming increasingly available (Campbell, 2002).

On May 2003, the scan line corrector (SLC) subsystem on board Landsat-7 developed an anomaly (Markham *et al.*, 2004). Later that year the problem was identified as a permanent mechanical failure and Landsat-7 resumed its mission with the SLC turned off. The device was designed to compensate for the forward motion of the satellite. This malfunction causes data loss at the edges of images, with an effective 30% loss of information (Markham *et al.*, 2004). Simple interpolation techniques have been applied to the datasets with promising results, but further validation is needed. Since the remaining pixels, comprising 30 km^2 in the center, are unaffected, Landsat-7 is still able to yield useful imagery in the middle of each scene (USGS, 2003).

4.2.3. SPOT—The French Satellite Pour l'Observation de la Terre (SPOT) programme began in 1986 with the launch of SPOT-1, carrying the High Resolution Visible (HRV) payload. There were many similarities to Landsat-TM imagery, but, essentially, the SPOT-HRV achieved a slightly higher spatial resolution with fewer spectral channels. Data collection has continued with SPOT-4, which carries the High Resolution Visible and Infrared (HRVIR) sensor and the multi-spectral VEGETATION (VGT-1) instrument. The more recent SPOT-5 has a VGT-2, similar to its predecessor, and a High Resolution Geometric (HRG) camera that achieves spatial resolutions of up to 5 m (Campbell, 2002). Of particular relevance from an epidemiological perspective is the VGT instrument, due to the high-quality vegetation index it offers, the availability of fully corrected imagery at a constant 1 × 1 km spatial resolution, and an almost daily global coverage in four spectral bands (0.45–1.75 μm) (Campbell, 2002; Maisongrande *et al.*, 2004). The VGT sensor data can be provided as a number of different products: VGT-P products (the physical values); VGT-S1 products (daily maximum value composited syntheses), VGT-S10 products (10-day maximum value composited syntheses) and VGT–D10 products (10-day BiDirectional Composite syntheses). The VGT-S10 and VGT-D10 products are also available in degraded resolutions of 4 × 4 and 8 × 8 km.

4.2.4. ADEOS—The Japanese polar orbiting Advanced Earth Observation Satellite (ADEOS) program provides global observation datasets, albeit with limited temporal coverage due to instrument problems. The ADEOS-1 satellite, launched in August 1996, went out of service in June 1997 due to structural damage to its solar array. The ADEOS-II platform, launched in December 2002 went out of service in late October 2003, also due to loss of power and subsequent sensor failure. Both platforms carried a Global Imager (GLI) with a spectral range spanning the visible to thermal infrared portion of the spectrum. The 36 GLI spectral bands acquired data at 10:30 local time from an altitude of 800 km in a Sun-synchronous polar orbit. The additional spectral information provided by GLI data has potential for improving estimates of land surface variables, including fire disturbance mapping and "hot spot" monitoring, as well as more visible channels useful for monitoring surface waters. A swath width of 1600 km provided observational data suitable for global land surface monitoring with a repeat interval of four days. The GLI data have a spatial resolution of 250 × 250 m in six channels, comparable to MODIS, and 1 × 1 km in other spectral bands. Improved atmospheric water vapour absorption bands aided estimation of vertical humidity profiles and near-surface water content, which were being used for improved weather forecasting and have utility in disease vector mapping (Goetz *et al.*,

2000). ADEOS science teams have provided higher order datasets of environmental variables derived from the raw GLI data. Production of vegetation cover type, density, productivity and change have experienced severe set-backs from instrument failures, but refined datasets of potential utility to disease applications exist for most of 2003.

4.3. Data Continuity and NPOESS

Continuity between NOAA-AVHRR and MODIS, the current "work horses" of global RS observing systems, requires both comparability assessments between sensors and the development of future new technology sensors. The latter will have improved radiometric and spectral properties, often provided at finer spatial resolution. Inter-sensor comparability requires assessments of the ability to retrieve consistent surface environmental information from instruments with different spectral, spatial, geometric, radiometric and orbital acquisition properties. NASA has recently funded several studies to address these issues in order to ensure continuity of the primary surface reflectance, as well as spectral vegetation indices and other data products that are used to derive many of the environmental variables relevant to epidemiological research. These efforts include empirical comparisons with sensor cross calibrations and associated transformation statistics, similar to what has been done with higher resolution sensors (Goetz, 1997). More analytical assessments are planned based on modelling radiative transfer and incorporating relative sensor responses across the energy spectrum within the sensor band passes. Both approaches will lead, either separately or in combination, to releases of improved datasets compared to those distributed on the accompanying DVD, in addition to providing an extended observational record.

The primary satellite platforms that will provide data continuity with AVHRR, Terra and Aqua over the next two decades of Earth observation are the National Polar-orbiting Operational Environmental Satellite Systems (NPOESS). An NPOESS Preparatory Project (NPP) satellite is scheduled for launch in May 2006 [URL: http://jointmission.gsfc.nasa.gov/]. It will carry four primary sensors including the Visible Infrared Imager Radiometer Suite (VIIRS), the Cross-Track Infrared Sounder, the Advanced Technology Microwave Sounder, and the Ozone Mapping and Profiler Suite. The sensor of primary interest to epidemiological research is the VIIRS, which is intended to provide global observations of land, ocean and atmosphere parameters on a near-daily basis. VIIRS is a 22-band instrument, with a spectral range from the visible through the thermal infrared wavelengths (0.4–12 μm). Each image swath will extend 3000 km, from a satellite orbit of 833 km, by scanning fore and aft ±56° of nadir. Image data will be acquired at two spatial resolutions, 370 and 740 m. A unique feature of VIIRS is that it will reduce pixel size across track (along scan), which will therefore compensate for the pixel size expansion that typically accompanies off-nadir viewing. The ground track of NPP will mimic that of Terra, with a 16-day repeat interval consistently acquiring data at 10:30 local time in its descending node (the ascending node data are acquired at night). On the planned series of NPOESS satellites that follow NPP, data will be acquired at 09:30, 13:30 and 17:30, respectively, in order to provide measures of diurnal variability and continuity with Aqua. These satellites are currently planned for launch after 2010. Although delays are expected, the life expectancy of the instruments will ensure there are periods of overlap between VIIRS sensors. A broad range of environmental variables will be provided from VIIRS observational data by NPP and NPOESS project science teams, including vegetation indices, surface temperature, land cover type classifications, energy and heat fluxes, atmospheric water vapour and soil moisture dynamics. Much of this work will be based on algorithms already developed by the MODIS science teams. The production of VIIRS continuity datasets will also make use of a wide range of geometric, radiometric and atmospheric corrections of the raw image data.

4.4. Other Data Sources

This brief section is not intended to be a comprehensive list of the digital data available to the epidemiological community but simply selected highlights that we have found to be particularly useful. A more comprehensive listing can be found at [http://www.eden-fp6project.net/].

4.4.1. Digital Elevation—Elevation influences temperature, rainfall and humidity and has been used widely in the mapping of infectious disease (Guerra *et al.*, this volume, pp. 157–179). The 1 km Global Land One-kilometre Base Elevation (GLOBE) Digital Elevation Model (DEM) is provided on the accompanying DVD. Full documentation is provided elsewhere [URL: http://www.ngdc.noaa.gov/seg/topo/report/]. We have resampled these data to be compatible with the 1 × 1 km and 8 × 8 km RS datasets. Table 6 summarizes the important features of the data.

The Shuttle Radar Topography Mission (SRTM) obtained elevation data on a near-global scale to generate the most complete high-resolution digital topographic database of Earth. The SRTM consisted of a specially modified radar system that flew on board the space shuttle Endeavour during an 11-day mission in February of 2000 (Rabus *et al.*, 2003; Smith and Sandwell, 2003; Sun *et al.*, 2003). The SRTM is an international project led by the National Geospatial-Intelligence Agency (NGA) and NASA. Elevation data at 1, 3 and 30 arc-second spatial resolution (USGS, 2004) are freely available from Global Land Cover Facility [http://www.landcover.org/data/srtm/].

4.4.2. Gridded Climatologies—The Climate Research Unit at the University of East Anglia, UK, produces a range of global gridded climatologies (New *et al.*, 1999, 2000, 2002; Mitchell and Jones, 2005) derived from interpolated meteorological station data [http://www.cru.uea.ac.uk/cru/data/hrg.htm]. These include 0.5 × 0.5° time series from 1901 to 2000 (New *et al.*, 1999, 2000) and 10 × 10 arc-second climatology of synoptic months (1961–1990) (New *et al.*, 2002). They include useful comparative information on precipitation, wet day frequency, daily mean temperature, diurnal temperature range, cloud cover, frost day frequency, vapour pressure, wind speed, relative humidity and sunshine hours. These data have found extensive use in epidemiological studies (Rogers and Randolph, 2000; Hay *et al.*, 2002; Shanks *et al.*, 2002; Small *et al.*, 2003).

4.4.3. The Africa Data Dissemination Service (ADDS)—The ADDS is operated by the US Agency for International Development (USAID) as part of its Famine Early Warning System Network (FEWS NET) [http://igskmncnwb015.cr.usgs.gov/adds/]. It provides a regularly updated archive of AVHRR-derived NDVI for Africa, as well as a suite of climatic products, including rainfall estimate data related to disease risk and food insecurity. The advantage of these data is that they are available in near real-time and hence useful for malaria epidemic monitoring (Hay *et al.*, 2003a,b).

4.4.4. Gridded Human Population Data—Demographic data are crucial if visually striking maps are to be turned into useful operational tools. The “state-of-the-art” in population mapping is detailed extensively elsewhere in this volume (Balk *et al.*, this volume, pp. 119–156) and the data can also be found on the accompanying DVD (see Table 3 and Balk *et al.*, this volume).

4.4.5. Administrative Boundary Data—One frequent difficulty in interpreting disease and other data that are mapped to local administrative boundaries, and given local administrative names, is that there is no agreed global dataset of administrative boundaries beyond admin level 1 (the major division below country level zero). Although currently

Europe PMC Funders Author Manuscripts

incomplete, the best current prospects for a global documented standard seems to be the United Nations funded Second Level Administrative Boundaries project [http://www3.who.int/whosis/gis/salb/salb_home.htm]. Data for many countries are already available and the site is frequently updated.

4.4.6. Global Landcover Facility (GLCF)—The GLCF at the University of Maryland, USA holds a vast array of freely downloadable satellite imagery and other data sources [http://www.glcf.umiacs.umd.edu/index.shtml]. Satellite imagery available includes global coverage of MODIS and Landsat MSS, TM and ETM+, with ASTER data made available recently. A variety of vegetation, land cover, forest and burned area products, all derived from satellite imagery, are available globally. We include it here because it is an unparalleled resource.

5. CONCLUSIONS

The 8 × 8 km TALA TFA RS dataset provides the most stable synoptic surfaces our group has used for monitoring global scale environmental conditions of relevance to infectious disease mapping. The data represent a significant spatial resolution advance on synoptic climatologies and complement the suite of data these surfaces provide. The 1 × 1 km TALA TFA data streams are less comprehensive temporally, show some latitudinally dependent differences with the 8 × 8 km TFA data and are hence more experimental. They should therefore be used with greater caution. Ancillary DEM, population and human settlement data are also distributed on the DVD. These collective data represent a contemporary snapshot of environmental conditions of use to those engaged in infectious disease mapping and wider environmental studies. It is difficult to predict their useful lifespan but we hope the continuity missions outlined will make their regular updating a necessity.

Acknowledgments

We are grateful to Simon Brooker, Archie Clements, Marlies Craig, Benjamin McCormick, Bethan Purse, Victoria Sanderson and William Wint for their comments on this manuscript and testing beta versions of the DVD. SIH and AJT are funded by a Research Career Development Fellowship (to SIH) from the Wellcome Trust (#069045). SJG acknowledges support from the NASA Applied Sciences program. AJG was funded by EU grant GOCE-2003-010284 EDEN and the paper is catalogued by the EDEN Steering Committee as EDEN0009 (http://www.eden-fp6project.net/). The contents of this publication are the sole responsibility of the authors and do not reflect the views of the European Union. The datasets discussed in this review were generated using resources from a range of grants supplied principally by the UK Department for International Development and the Wellcome Trust, for which we are grateful. The following formal acknowledgement is requested of those who use Pathfinder AVHRR Land data:

> "Data used by the authors in this study include data produced through funding from the Earth Observing System Pathfinder Program of NASA's Mission to Planet Earth in co-operation with the National Oceanic and Atmospheric Administration. The data were provided by the Earth Observing System Data and Information System (EOSDIS) Distributed Active Archive Center at Goddard Space Flight Center, which archives, manages and distributes this dataset."

We provide the information on the DVD that accompanies this article without restriction, and request that when the data are used in further studies, due acknowledgement is given to the TALA Research Group, Oxford University, UK, for the TFA images and this review is referenced as the source. All reasonable efforts have been made to check the fidelity of these data but the authors can accept no liability for subsequent use and application of these datasets.

REFERENCES

1. Anderson, RM. Epidemiology. In: Cox, FEG., editor. Modern Parasitology. 2nd ed. Oxford; Blackwell: 1993. p. 75-116.
2. Campbell, JB. Introduction to Remote Sensing. 3rd ed. Taylor & Francis; London: 2002.

3. Chander G, Helder DL, Markham BL, Dewald JD, Kaita E, Thome KJ, Micijevic E, Ruggles TA. Landsat-5 TM reflective-band absolute radiometric calibration. IEEE Transactions on Geoscience and Remote Sensing. 2004; 42:2747–2760.
4. Cracknell, AP. The Advanced Very High Resolution Radiometer. Taylor & Francis; London: 1997.
5. Draeger WC, Holm TM, Lauer DT, Thompson RJ. The availability of Landsat data: past, present and future. Photogrammetric Engineering & Remote Sensing. 1997; 63:869–875.
6. Eastman JR, Falk M. Long sequence time series evaluation using standardized principal components. Photogrammetric Engineering & Remote Sensing. 1993; 59:1307–1312.
7. Ehrlich D, Estes JE, Singh A. Applications of NOAA-AVHRR 1 km data for environmental monitoring. International Journal of Remote Sensing. 1994; 15:145–161.
8. Eidenshink JC, Faundeen JL. The 1 km AVHRR global land data set - 1st stages in implementation. International Journal of Remote Sensing. 1994; 15:3443–3462.
9. EUMETSAT. The Meteosat System. European Organisation for the Exploitation of Meteorological Satellites; Darmstadt: 2000.
10. EUMETSAT. Meteosat Second Generation—MSG System Overview. European Organisation for the Exploitation of Meteorological Satellites; Darmstadt: 2001.
11. EUMETSAT. Meteosat-8 (MSG-1) Dissemination Trial. European Organisation for the Exploitation of Meteorological Satellites; Darmstadt: 2004.
12. EUMETSAT. The Meteosat Second Generation Programme. European Organisation for the Exploitation of Meteorological Satellites; Darmstadt: 1998.
13. Friedl MA, McIver DK, Hodges JCF, Zhang XY, Muchoney D, Strahler AH, Woodcock CE, Gopal S, Schneider A, Cooper A. Global land cover mapping from MODIS: algorithms and early results. Remote Sensing of Environment. 2002; 83:287–302.
14. Gleason ACR, Prince SD, Goetz SJ, Small J. Effects of orbital drift on land surface temperature measured by AVHRR thermal sensors. Remote Sensing of Environment. 2002; 79:147–165.
15. Goetz SJ. Multi-sensor analysis of NDVI, surface temperature, and biophysical variables at a mixed grassland site. International Journal of Remote Sensing. 1997; 18:71–94.
16. Goetz SJ, Prince SD, Small J. Advances in satellite remote sensing of environmental variables for epidemiological applications. Advances in Parasitology. 2000; 47:289–307. [PubMed: 10997210]
17. Green RM, Hay SI. The potential of Pathfinder AVHRR data for providing surrogate climatic variables across Africa and Europe for epidemiological applications. Remote Sensing of Environment. 2002; 79:166–175. [PubMed: 22581983]
18. Gutman GG. On the monitoring of land surface temperatures with the NOAA/AVHRR: removing the effect of satellite orbit drift. International Journal of Remote Sensing. 1999; 20:3407–3413.
19. Hay SI. An overview of remote sensing and geodesy for epidemiology and public health application. Advances in Parasitology. 2000; 47:1–35. [PubMed: 10997203]
20. Hay SI, Cox J, Rogers DJ, Randolph SE, Stem DI, Shanks GD, Myers MF, Snow RW. Climate change and the resurgence of malaria in the East African highlands. Nature. 2002; 415:905–909. [PubMed: 11859368]
21. Hay SI, Lennon JJ. Deriving meteorological variables across Africa for the study and control of vector-borne disease: a comparison of remote sensing and spatial interpolation of climate. Tropical Medicine & International Health. 1999; 4:58–71. [PubMed: 10203175]
22. Hay SI, Packer MJ, Rogers DJ. The impact of remote sensing on the study and control of invertebrate intermediate hosts and vectors for disease. International Journal of Remote Sensing. 1997; 18:2899–2930.
23. Hay, SI.; Randolph, SE.; Rogers, DJ., editors. Remote Sensing and Geographical Information Systems in Epidemiology. 2000. Advances in Parasitology 47
24. Hay SI, Renshaw M, Ochola SA, Noor AM, Snow RW. Performance of forecasting, warning and detection of malaria epidemics in the highlands of western Kenya. Trends in Parasitology. 2003a; 19:394–399. [PubMed: 12957515]
25. Hay SI, Snow RW, Rogers DJ. From predicting mosquito habitat to malaria seasons using remotely sensed data: practice, problems and perspectives. Parasitology Today. 1998; 14:306–313. [PubMed: 17040796]

26. Hay SI, Tucker CJ, Rogers DJ, Packer MJ. Remotely sensed surrogates of meteorological data for the study of the distribution and abundance of arthropod vectors of disease. Annals of Tropical Medicine and Parasitology. 1996; 90:1–19. [PubMed: 8729623]
27. Hay SI, Were EC, Renshaw M, Noor AM, Ochola SA, Olusanmi L, Alipui N, Snow RW. Forecasting, warning, and detection of malaria epidemics: a case study. Lancet. 2003b; 361:1705–1706. [PubMed: 12767739]
28. Holben BN. Characteristics of maximum-value composite images from temporal AVHRR data. International Journal of Remote Sensing. 1986; 7:1417–1434.
29. James ME, Kalluri SNV, Townshend JRG. The Pathfinder AVHRR land data set: an improved coarse resolution data set for terrestrial monitoring. International Journal of Remote Sensing. 1994; 15:3347–3363.
30. Justice CO, Markham BL, Townshend JRG, Kennard RL. Spatial degradation of satellite data. International Journal of Remote Sensing. 1989; 10:1539–1561.
31. Justice CO, Townshend JRG, Vermote EF, Masuoka E, Wolfe RE, Saleous N, Roy DP, Morisette JT. An overview of MODIS Land data processing and product status. Remote Sensing of Environment. 2002; 83:3–15.
32. Kidwell, KB. NOAA Polar Orbiter Data User's Guide. US Department of Commerce, NESDIS, NOAA, National Climate Data Center, Satellite Data Service Division; Washington DC: 1998.
33. Lambin EF, Ehrlich D. The surface temperature-vegetation index space for land cover and land-cover changes analysis. International Journal of Remote Sensing. 1996; 17:463–487.
34. Landgrebe D. The evolution of Landsat data analysis. Photogrammetric Engineering & Remote Sensing. 1997; 63:859–867.
35. Lauer DT, Morain SA, Salomonson VV. The Landsat Program: its origins, evolution and impacts. Photogrammetric Engineering & Remote Sensing. 1997; 63:831–838.
36. Lawler A. Earth observation program "at risk", academy warns. Science. 2005; 308:614–615. [PubMed: 15860598]
37. Lillesand, TM.; Kiefer, RW. Remote Sensing and Image Interpretation. Wiley; New York: 2000.
38. Maisongrande P, Duchemin B, Dedieu G. VEGETATION/SPOT: an operational mission for the Earth monitoring; presentation of new standard products. International Journal of Remote Sensing. 2004; 25:9–14.
39. Markham BL, Storey JC, Williams DL, Irons JR. Landsat sensor performance: history and current status. IEEE Transactions on Geoscience and Remote Sensing. 2004; 42:2691–2694.
40. Mather, PM. Computer Processing of Remotely sensed Images: an Introduction. 2nd ed. Wiley; Chichester: 1999.
41. Mika AM. Three decades of Landsat instruments. Photogrammetric Engineering & Remote Sensing. 1997; 63:839–852.
42. Mitchell TD, Jones PD. An improved method of constructing a database of climate observations and associated high-resolution. International Journal of Climatology. 2005; 25:693–712.
43. NASA. The GOES Program. 1999. http://www.earth.nasa.gov/history/goes/goes.html. 2004
44. Nemani RR, Keeling CD, Hashimoto H, Jolly WM, Piper SC, Tucker CJ, Myneni RB, Running SW. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. Science. 2003; 300:1560–1563. [PubMed: 12791990]
45. New M, Hulme M, Jones P. Representing twentieth-century space-time climate variability. Part I: Development of a mean 1961–90 monthly terrestrial climatology. Journal of Climate. 1999; 12:829–857.
46. New M, Hulme M, Jones P. Representing twentieth-century space-time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. Journal of Climate. 2000; 13:2217–2238.
47. New M, Lister D, Hulme M, Makin I. A high-resolution data set of surface climate over global land areas. Climate Research. 2002; 21:1–25.
48. NOAA – NPOESS. National Polar-orbiting Operational Environmental Satellite System. 2003.
49. Parkinson CL. Aqua: an earth-observing satellite mission to examine water and other climate variables. IEEE Transactions on Geoscience and Remote Sensing. 2003; 41:173–183.

50. Price JC. Estimating surface temperatures from satellite thermal infrared data – a simple formulation for the atmospheric effect. Remote Sensing of Environment. 1983; 13:353–361.
51. Rabus B, Eineder M, Roth A, Bamler R. The shuttle radar topography mission – a new class of digital elevation models acquired by spaceborne radar. ISPRS Journal of Photogrammetry and Remote Sensing. 2003; 57:241–262.
52. Robinson TP. Spatial and temporal accuracy of coarse resolution products of NOAA-AVHRR NDVI data. International Journal of Remote Sensing. 1996; 17:2303–2321.
53. Rogers DJ. Satellites, space, time and the African trypanosomiases. Advances in Parasitology. 2000; 47:128–171.
54. Rogers DJ, Hay SI, Packer MJ. Predicting the distribution of tsetse flies in West Africa using temporal Fourier processed meteorological satellite data. Annals of Tropical Medicine and Parasitology. 1996; 90:225–241. [PubMed: 8758138]
55. Rogers DJ, Packer MJ. Vector-borne diseases, models and global change. Lancet. 1993; 342:1282–1284. [PubMed: 7695661]
56. Rogers DJ, Randolph SE. The global spread of malaria in a future, warmer world. Science. 2000; 289:2283–2284. [PubMed: 11041794]
57. Rogers DJ, Randolph SE, Snow RW, Hay SI. Satellite imagery in the study and forecast of malaria. Nature. 2002; 415:710–715. [PubMed: 11832960]
58. Rogers, DJ.; Robinson, TP. Tsetse distribution. In: Maudlin, I.; Holmes, PH.; Miles, MA., editors. The Trypanosomiases. CAB International; Wallingford: 2004. p. 139-179.
59. Schmetz J, Pili P, Tjemkes S, Just D, Kerkmann J, Rota S, Ratier A. An introduction to Meteosat Second Generation (MSG). Bulletin of the American Meteorological Society. 2002; 83:977–992.
60. Shanks GD, Hay SI, Stern DI, Biomndo K, Snow RW. Meteorologic influences on *Plasmodium falciparum* malaria in the highland tea estates of Kericho, western Kenya. Emerging Infectious Diseases. 2002; 8:1404–1408. [PubMed: 12498655]
61. Small J, Goetz SJ, Hay SI. Climatic suitability for malaria transmission in Africa, 1911–1995. Proceedings of the National Academy of Sciences of the USA. 2003; 100:15341–15345. [PubMed: 14663146]
62. Smith B, Sandwell D. Accuracy and resolution of shuttle radar topography mission data. Geophysical Research Letters. 2003; 30 art. no.-1467.
63. Snyder, JP. Map projections: a working manual. USGS Professional Paper 1395. United States Geological Survey (USGS), United States Government Printing; Boulder, CO: 1997.
64. Snyder, JP.; Voxland, PM. An album of map projections. USGS Professional Paper 1453. United States Geological Survey (USGS), United States Government Printing; Boulder, CO: 1994.
65. Steinwand DR. Mapping raster imagery to the Interrupted Goode Homolosine Projection. International Journal of Remote Sensing. 1994; 15:3463–3471.
66. Stowe LL, McClain EP, Carey R, Pellegrino P, Gutmann GG, Davia P, Long C, Hart S. Global distribution of cloud cover derived from NOAA/AVHRR operational satellite data. Advances in Space Research. 1991; 11:51–54.
67. Sun G, Ranson KJ, Khairuk VI, Kovacs K. Validation of surface height from shuttle radar topography mission using shuttle laser altimeter. Remote Sensing of Environment. 2003; 88:401–411.
68. Tatem AJ, Goetz SJ, Hay SI. Terra and Aqua: new data for epidemiology and public health. International Journal of Applied Earth Observation and Geoinformation. 2004; 6:33–46. [PubMed: 22545030]
69. Teillet PM, El Saleous N, Hansen MC, Eidenshink JC, Justice CO, Townshend JRG. An evaluation of the global 1-km AVHRR land dataset. International Journal of Remote Sensing. 2000; 21:1987–2021.
70. Thome KJ, Helder DJ, Aaron D, Dewald JD. Landsat-5 TM and Landsat-7 ETM+ absolute radiometric calibration using the reflectance-based method. IEEE Transactions on Geoscience and Remote Sensing. 2004; 42:2777–2785.
71. Townshend JRG. Global data sets for land applications from the advanced very high resolution radiometer: an introduction. International Journal of Remote Sensing. 1994; 15:3319–3332.

72. Townshend JRG, Justice CO. Towards operational monitoring of terrestrial systems by moderate-resolution remote sensing. Remote Sensing of Environment. 2002; 83:351–359.
73. Townshend JRG, Justice CO, Skole D, Malingreau JP, Cihlar J, Teillet P, Sadowski F, Ruttenberg S. The 1 km resolution global data set: needs of the International Geosphere Biosphere Program. International Journal of Remote Sensing. 1994; 15:3417–3441.
74. Tucker CJ, Pinzon JE, Brown ME, Slayback DA, Pak EW, Mahoney R, Vermote EF, El Saleous N. The global inventory mapping and monitoring 1981–1999 AVHRR 8-Km dataset. An extended AVHRR 8-Km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. International Journal of Remote Sensing. 2005; 26:4485–4498.
75. United States Geological Survey (USGS). Landsat 7 SLC Anomaly Investigation. 2003; Vol. 2003
76. United States Geological Survey (USGS). (1, 3, 30) Arc second SRTM elevation, reprocessed to GeoTIFF. The Global Land Cover Facility; College Park, MD: 2004.
77. Verger, F.; Sourbès-Verger, I.; Ghirardi, R. The Cambridge Encyclopaedia of Space: Missions, Applications and Exploration. Cambridge University Press; Cambridge: 2003.
78. Williamson RA. The Landsat legacy: remote sensing policy and the development of commercial remote sensing. Photogrammetric Engineering & Remote Sensing. 1997; 63:877–885.
79. Yamaguchi Y, Kahle AB, Tsu H, Kawakami T, Pniel M. Overview of advanced spaceborne thermal emission and reflection radiometer (ASTER). IEEE Transactions on Geoscience and Remote Sensing. 1998; 36:1062–1071.

Europe PMC Funders Author Manuscripts

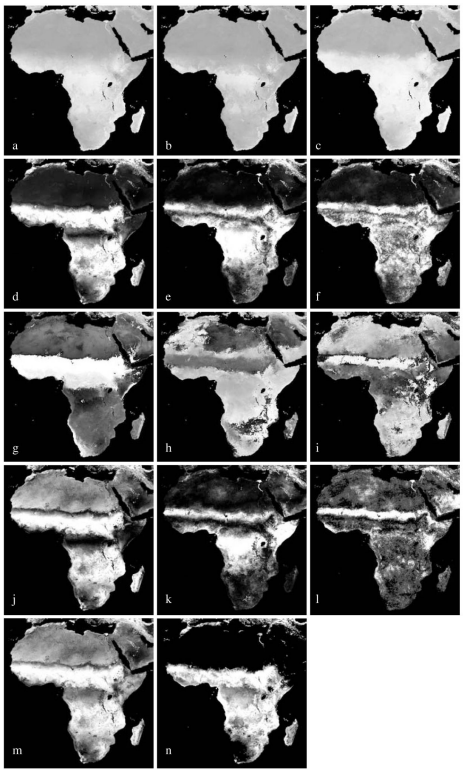

Figure 1.

a–n Panel of the fourteen 8 × 8 km NOAA-AVHRR NDVI TFA images (a) a0, mean, (b) mn, minimum and (c) mx, maximum signal recomposed from the first three Fourier cycles. The amplitude of the (d) a1, annual cycle, (e) a2, bi-annual and (f) a3, tri-annual are also shown, in addition to the (g) p1, phase of annual, (h) p2, bi-annual, (i) and p3, tri-annual cycle in months. The proportion of the variance in the original time series described by the (j) d1, the annual, (k) d2, the bi-annual, (l) d3, the tri-annual and (m) da, all three cycles combined is also shown with (n) vr, the variance of the original data time series. Data are histogram equalized stretched from the minimum data value (black, also water mask) to maximum (white) for display.

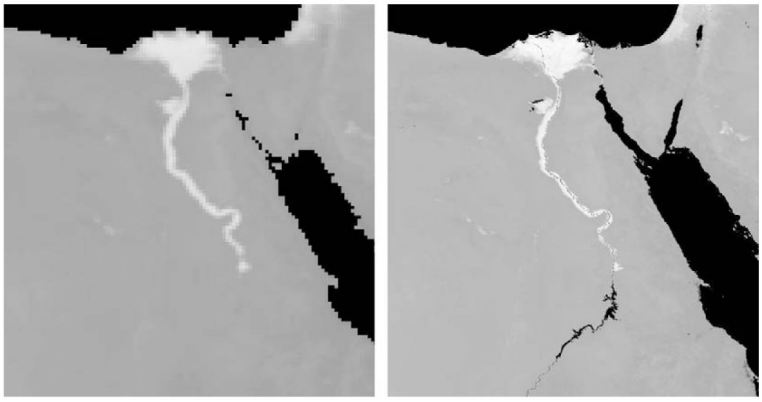

Figure 2.
a,b The lower reaches of the Nile river in Egypt illustrate the difference between the 8 × 8 km spatial resolution imagery (on the left) and the 1 × 1 km imagery (right). Images are of the mean NOAA-AVHRR NDVI TFA a0 product (Table 3). Data are histogram equalized stretched from the minimum data value (black, also water mask) to maximum (white) for display.

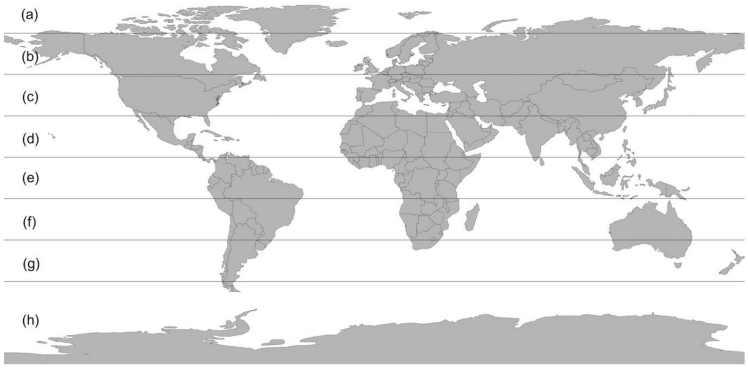

Figure 3.
Latitudinal sample regions, (a) 90–70°N, (b) 70–50°N, (c) 50–30°N, (d) 30–10°N, (e) 10N–10°S, (f) 10–30°S, (g) 30–50°S, (h) 50–90°S.

Europe PMC Funders Author Manuscripts

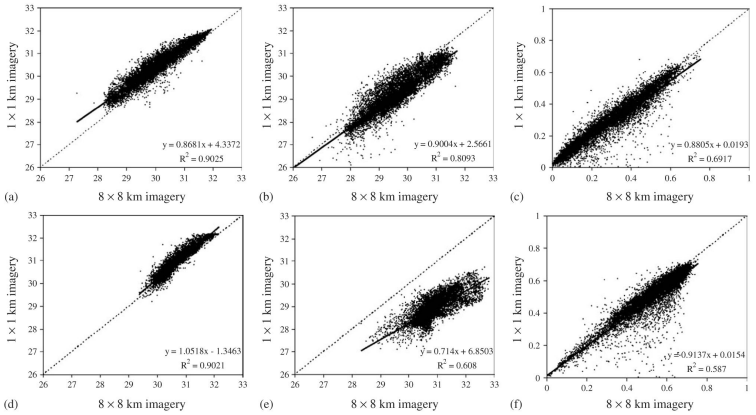

Figure 4.

a–f Scatterplots for latitudinal segments shown in Figure 3 of 8 × 8 km imagery sample points against 1 × 1 km imagery sample points for 50°N–30°N for MIR (a), LST (b) and NDVI (c) and for 10°N–10°S for MIR (d), LST (e) and NDVI (f). One-to-one lines (dashed) are added for ease of interpretation. Solid lines (and equations) are least squares linear fits to the data.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Dates of the NOAA-AVHRR 1 × 1 km data acquisition archived for IGBP-DIS. A dot indicates acquisition of all decads

Month	1992	1993	1995	1996
January		●		●
February		●	●	
March		●	●	
April	●	●	●	●
May	●	●	●	
June	●	●	●	
July	●	●	●	
August	●	●	●	
September	●	●	●	
October	●		●	
November	●		●	
December	●		●	

Europe PMC Funders Author Manuscripts

Table 2

Geo-referencing information for the TALA RS data

Details	1 × 1 km imagery	8 × 8 km imagery
Projection	Geographic (latitude/longitude)	Geographic (latitude/longitude)
Image size, columns	36 000	3600
Image size, rows	18 000	1800
Upper-left coordinate, pixel centre	−179.995	−179.95
Upper-left coordinate, pixel centre	89.995	89.95
Units	decimal degrees	decimal degrees
Pixel size: x	0.01	0.1
Pixel size: y	0.01	0.1
Spheroid	WGS84	WGS84
Datum	WGS84	WGS84

Europe PMC Funders Author Manuscripts

Table 3

DVD imagery contents with details of scaling factors to be applied to the data, the data units and expected values

File name	Imagery type (°)	Image	Scaling	Units	Geophysical min.	Geophysical max.
wd1003a0	TFA 0.1	mean MIR	(x/10) −273	°C	−7.8	48.9
wd1003a1	TFA 0.1	MIR annual amplitude	(x/10)	°C	0	30.8
3.1wd1003a2	TFA 0.1	MIR bi-annual amplitude	(x/10)	°C	0	1.0
wd1003a3	TFA 0.1	MIR tri-annual amplitude	(x/10)	°C	0	0.6
wd1003d1	TFA 0.1	MIR pvs described by annual cycle	No scaling	%	0	94
wd1003d2	TFA 0.1	MIR pvs described by bi-annual cycle	No scaling	%	0	45
wd1003d3	TFA 0.1	MIR pvs described by tri-annual cycle	No scaling	%	0	13
wd1003mn	TFA 0.1	Maximum MIR	(x/10) −273	°C	−26	32.4
wd1003mx	TFA 0.1	Maximum MIR	(x/10) −273	°C	−3.6	50.9
wd1003p1	TFA 0.1	MIR phase of annual cycle	(x/10)	Months	0	12
wd1003p2	TFA 0.1	MIR phase of bi-annual cycle	(x/10)	Months	0	6
wd1003p3	TFA 0.1	MIR phase of tri-annual cycle	(x/10)	Months	0	4
wd1003vr	TFA 0.1	MIR variance	(x/10)	%	0	563
wd1007a0	TFA 0.1	Mean LST	(x/10) −273	°C	−30.7	55.4
wd1007a1	TFA 0.1	LST annual amplitude	(x/10)	°C	0	3.7
wd1007a2	TFA 0.1	LST bi-annual amplitude	(x/10)	°C	0	1.0
wd1007a3	TFA 0.1	LST tri-annual amplitude	(x/10)	°C	0	0.5
wd1007d1	TFA 0.1	LST pvs described by annual cycle	No scaling	%	0	96
wd1007d2	TFA 0.1	LST pvs described by bi-annual cycle	No scaling	%	0	48
wd1007d3	TFA 0.1	LST pvs described by tri-annual cycle	No scaling	%	0	11
wd1007mn	TFA 0.1	Minimum LST	(x/10) −273	°C	−43.5	54.3
wd1007mx	TFA 0.1	Maximum LST	(x/10) −273	°C	−26.5	58.9
wd1007p1	TFA 0.1	LST phase of annual cycle	(x/10)	Months	0	12
wd1007p2	TFA 0.1	LST phase of bi-annual cycle	(x/10)	Months	0	6
wd1007p3	TFA 0.1	LST phase of tri-annual cycle	(x/10)	Months	0	4
wd1007vr	TFA 0.1	LST variance	(x/10)	%	0	758
wd1014a0	TFA 0.1	Mean NDVI	(x/1000)−1	No units[a]	−0.086	0.803
wd1014a1	TFA 0.1	NDVI annual amplitude	(x/1000)	No units[a]	0	0.445

Europe PMC Funders Author Manuscripts

File name	Imagery type (°)	Image	Scaling	Units	Geophysical min.	Geophysical max.
wd1014a2	TFA 0.1	NDVI bi-annual amplitude	(x/1000)	No units[a]	0	0.251
wd1014a3	TFA 0.1	NDVI tri-annual amplitude	(x/1000)	No units[a]	0	0.131
wd1014d1	TFA 0.1	NDVI pvs described by annual cycle	No scaling	%	0	94
wd1014d2	TFA 0.1	NDVI pvs described by bi-annual cycle	No scaling	%	0	80
wd1014d3	TFA 0.1	NDVI pvs described by tri-annual cycle	No scaling	%	0	39
wd1014mn	TFA 0.1	Minimum NDVI	(x/1000)−1	No units[a]	−0.12	0.764
wd1014mx	TFA 0.1	Maximum NDVI	(x/1000)−1	No units[a]	−0.066	0.865
wd1014p1	TFA 0.1	NDVI phase of annual cycle	(x/10)	Months	0	12
wd1014p2	TFA 0.1	NDVI phase of bi-annual cycle	(x/10)	Months	0	6
wd1014p3	TFA 0.1	NDVI phase of tri-annual cycle	(x/10)	Months	0	4
wd1014vr	TFA 0.1	NDVI variance	(x/100)00		0	0.116
wd1030dm	Globe D.E.M.	Globe DEM	No scaling	Meters	−295	6815
afpop00	Pop. surf.	UNEP-GRID Population for Africa	No scaling	Persons	0	1 580 304
afpopd00	Pop. d. surf.	UNEP-GRID Population for Africa	No scaling	Persons/km^2	0	84 923
gpw00	Pop. surf.	SEDAC Gridded population of the World	No scaling	Persons	0	1 467 471
gpw00d	Pop. d. surf.	SEDAC Gridded population of the World	No scaling	Persons/km^2	0	122 755
grump-2000	M. pop. surf.	SEDAC Gridded population of the World	No scaling	Persons	0	1 346 722

TFA=Temporal Fourier analysis; pvs=proportion of variance in original signal;

[a]no units because it is a ratio. Geophysical minimum and geophysical maximum values do not include masks. Globe DEM—digital elevation model. Pop. surf.=population surface, Pop. d. surf.=population density surface, M. pop. surf.=modelled population surface.

Europe PMC Funders Author Manuscripts

Table 4

Values of r^2 produced through linear regression analysis between 8 × 8 and 1 × 1 km imagery by latitudinal segment (latitudinal segments shown in Fig. 3)

Latitude	Mean MIR	Mean LST	Mean NDVI
70–90°N	0.368	0.778	0.135
50–70°N	0.705	0.755	0.523
30–50°N	0.903	0.809	0.692
10–30°N	0.950	0.510	0.822
10S–10°N	0.902	0.608	0.587
10–30°S	0.869	0.434	0.860
30–50°S	0.931	0.858	0.651
50–90°S	0.737	0.819	0.546

Europe PMC Funders Author Manuscripts

Table 5

Technical specifications of satellite sensors that have been used in epidemiology or that show significant potential for epidemiological applications

Satellite[a]	Sensor	Bands[b]	Bandwith (μm)	Spatial res. (m)	Temporal res.	Swath (km)	URL
Landsat-5	MSS[c]	1	0.500–0.600	80	16 days	185	http://landsat.gsfc.nasa.gov
		2	0.600–0.700	80			
		3	0.700–0.800	80			
		4	0.800–1.100	80			
	TM	1	0.450–0.520	30	16 days	185	
		2	0.520–0.600	30			
		3	0.630–0.690	30			
		4	0.760–0.900	30			
		5	1.550–1.750	30			
		6	10.40–12.50	120			
		7	2.080–2.350	30			
Landsat-7	ETM+	1	0.450–0.520	30	16 days	185	
		2	0.530–0.610	30			
		3	0.630–0.690	30			
		4	0.780–0.900	30			
		5	1.550–1.750	30			
		6	10.40–12.50	60			
		7	2.090–2.350	30			
		8 (P)	0.520–0.900	15			
NOAA-16 & 17	AVHRR	1	0.580–0.680	1090	12 hours	2800	http://edc.usgs.gov/products/satellite/avhrr.html
		2	0.725–1.000	1090			
		3A	1.580–1.640	1090			
		3B	3.550–3.930	1090			
		4	10.300–11.300	1090			
		5	11.500–12.500	1090			
SPOT-4	HRVIR	1	0.500–0.590	20	1–4 days	60	http://www.spotimage.fr/
		2	0.610–0.680	20			
		3	0.780–0.890	20			

Europe PMC Funders Author Manuscripts

Satellite[a]	Sensor	Bands[b]	Bandwith (μm)	Spatial res. (m)	Temporal res.	Swath (km)	URL
		7	8.300–9.100	4800			
		8	9.380–9.940	4800			
		9	9.800–11.800	4800			
		10	11.000–13.000	4800			
		11	12.400–14.400	4800			
		12 (HRV)	0.400–1.100	1670			

[a] Operational and/or latest satellite of the series in orbit.

[b] P=Panchromatic.

[c] Also onboard Landsat-1, -2, -3 and -4. For all acronyms refer to text. Res.=Resolution.

Europe PMC Funders Author Manuscripts

Table 6

Geo-referencing information for the GLOBE DEM data

Details	30 arc second imagery
Projection	Geographic latitude/longitude)
Image size, columns	43200
Image size, rows	21600
Upper-left coordinate, pixel edge	−80
Upper-left coordinate, pixel edge	90
Units	Decimal degrees
Pixel size: x	0.00833333
Pixel size: y	0.00833333
Spheroid	WGS84
Datum	WGS84
Vertical (z) units	Metres above mean sea level