# Analysis of impacts of climate variability on malaria transmission in Sri Lanka and the potential for an early warning system

Foundation for Environment, Climate and Technology, SRI LANKA

Report sponsored by MacArthur Foundation under a grant to the University of Peradeniya

**Contact:**
**Principal Investigator:**
Lareef Zubair,
Foundation for Environment, Climate and Technology,
C/o Maintenance Office, Mahaweli Authority of Sri Lanka,
Digana Village, Rajawella, Sri Lanka
Tel: +94-81-2376746
lareefzubair@gmail.com

**Citation:** Zubair, L., Connor, S.J., Ward, N., Galapaththy, G.N.L., Mahanama, S., Faizal, H.M., Yahiya, Z., Siriwardhana, M., Tennakoon, M., Chandimal, J., Razick, S., Nawarathna, B., Shaman, J., Ndiaye, O. and Yang, H. Analysis of Impacts of Climate Variability on Malaria Transmission and the Development of Early Warning System. Project Report: Foundation for Environment, Climate and Technology, Digana Village, November 2011.

**Website:**
http://www.climate.lk/
http://www.tropicalclimate.org/

**Sponsors:**
The preparation of this document was supported by a grant from the John D. and Catherine T. MacArthur Foundation to the University of Peradeniya to launch a new Master's in Development Practice Programme based on a model of rigorous, inter-disciplinary training and the curriculum recommended by the International Commission on Education for Sustainable Development Practice.

# Table of Contents

# 1. Annex

## Data Quality Control

With sensors becoming cheaper, with much data being digitized, and with affordable computing resources, it has now become possible to find answers to many questions from large sets of data from multiple sources. Alongside such large volumes of data, it is the need to ensure that the data are not contaminated by human error, instrument error, copying errors and other forms of error. The interpretation of the data should be guided by the quality, precision and uncertainty in obtaining data.

Data quality control is the broad field of controlling the usage of data with known quality measurement for an application or a process. The best way of ensuring data quality is to ensure that measurements are taken precisely and carefully. After the data is acquired, further quality control is usually done after a data quality assurance process, which consists of checks for data inconsistency and correction. The primary purpose of the quality control of observational data is missing data detection, error detection and possible error corrections in order to ensure the highest possible reasonable standards of accuracy for the optimum use of these data by all possible users. The data quality control process uses the information from the quality assurance process, and then it decides to use the data for analysis or in an application or business process. For example, if a data quality control process finds the data contains too much error or inconsistency, it rejects the data to be processed. The usage of incorrect data could crucially impact outputs and interpretation.

As an example the following figure 01 shows the annual observed mean temperature and the reconstructed mean temperature in Kandy. Blue line shows the annual observed mean temperature in Kandy from 1869 to 1998. The dotted line shows the reconstruction from neighboring stations. When these two lines are similar the data is likely to be more reliable. The data records in Kandy were impaired due to the station being shifted around 1920 and 1950.

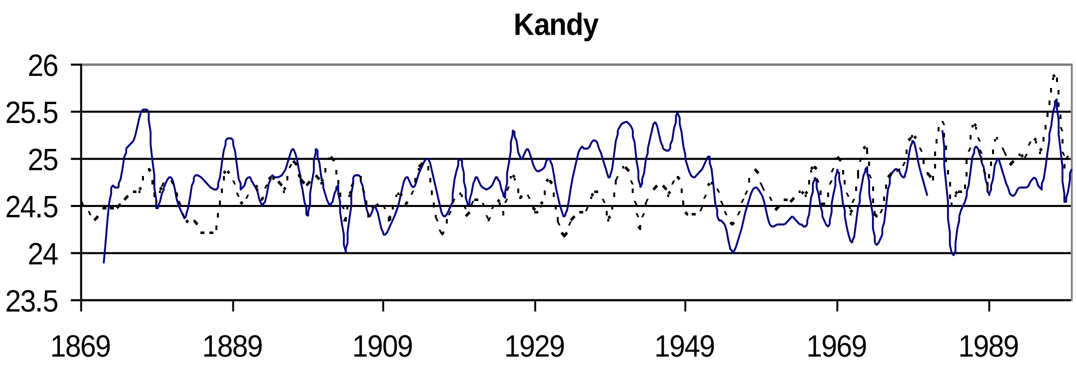


Figure 01: Annual observed mean temperature and the reconstructed mean temperature in Kandy for the duration of 1869 to 1998

The importance of data quality control is that data quality involves ensuring the accuracy, timeliness, completeness and consistency of data used by an organization while also making sure that everyone who uses the data has a common understanding of what the data represents. Poor quality data can waste both time and the money.

## Shuttle Radar Topography Mission

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. SRTM consisted of a specially modified radar system that flew onboard the Space Shuttle Endeavour during a 11-day mission on February 11 of 2000.

The elevation models are arranged into tiles, each covering one degree of latitude and one degree of longitude, named according to their south western corners (i.e. n45e006 stretches from 45°N 6°E to 46°N 7°E). With the development of science, the resolution of the cells of the source data of the SRTM was changed with time. Global land one kilometer base elevation (GLOBE) was initially spearheaded in 1990 and its resolution of the cells was thirty arc seconds which is approximately 1km. Today, there are two levels of resolutions used in the SRTM. One is one arc second, but 1" (approx. 30 meter) data have only been released over United States territory; for the rest of the world, only three-arc-second -3" (approx. 90-meter) data are available. Each one arc second tile has 3,601 rows, consisting of 3,601 16 bit bigendian cells. The dimensions of the three-arc-second tiles are 1201 x 1201.

SRTM data is being processed at the Jet Propulsion Laboratory (JPL) in Pasadena, California into research-quality digital elevation models. The data is 90 meter averaged from the original 30 meter data. As each continent is completed, the data is being sent to NIMA for finalizing (final editing, verification, and conformance to National Map Accuracy standards). Each continent is being processed in turn beginning with North America, then, South America, Australia, Eurasia, Africa, North & South Pacific, and North & South Atlantic.

The SRTM radar contained two types of antenna panels, C-band and X-band. The near-global topographic maps of Earth called Digital Elevation Models (DEMs) are made from the C-band radar data. These data were processed at the JPL and are being distributed through the United States Geological Survey's EROS Data Center. Data from the X-band radar are used to create slightly higher resolution DEMs but without the global coverage of the C-band radar. The SRTM X-band radar data are being processed and distributed by the German Aerospace Center.

## Normalized Difference Vegetation Index

The Normalized Difference Vegetation Index (NDVI) is a numerical and simple graphical indicator that uses the visible and near infrared bands of the electromagnetic spectrum which is used to analyze remote sensing measurements, typically but not necessarily from a space platform, and assess whether the target is being observed contains live green vegetation or not. The NDVI has been in use for many years to measure and monitor plant growth (vigor), vegetation cover, and biomass production from multi-spectral satellite data.

The NDVI may be calculated from the amount of light that the observed area reflects in the near-infrared and red regions of the light spectrum. Live green plants absorb solar radiation in the photosynthetically active radiation (PAR) spectral region, which the plants use as a source of energy in the process of photosynthesis. Live green plants appear relatively dark in the PAR and relatively bright in the near-infrared. By contrast, clouds and snow tend to be rather bright in the red and quite dark in the near-infrared. NDVI is calculated using the following formula;

$$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{VIS})}{(\mathrm{NIR} + \mathrm{VIS})}$$

NIR – Value of near infra red radiation from a pixel

VIS – Value of photosynthetically active radiation from a pixel

This formula yields a value that ranges between -1 (i.e. usually for water) and +1. Practically negative values represent water; values around zero represent bare soil; and values over 6 represent dense green vegetation.

A low reflectance in the region increases the probability that the observed area contains vegetation. Generally healthy vegetation will absorb most of the visible light that falls on it, and large portion of the near infrared light. Bare soils on the other hand reflect moderately in both red and infrared portion of the electromagnetic spectrum.

NDVI is widely applied in vegetative studies as it is been used to estimate crop yields, pasture performances etc. Also NDVI is directly related to other ground parameters such as percent of ground cover, photosynthetic activity of the plant, surface water, leaf area index and the amount of bio mass.

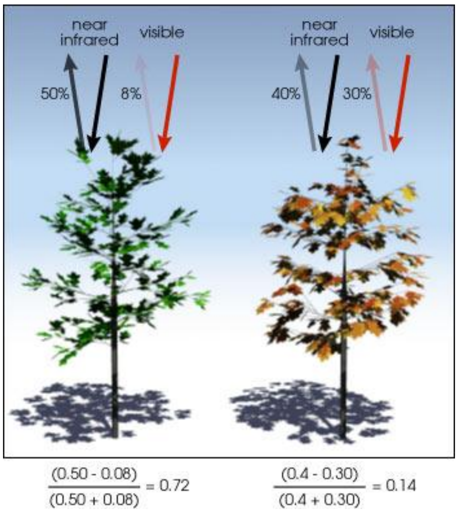


Figure 02: Graphical representation of the NDVI calculation

## El-Nino Southern Oscillation Indices

El Niño/La Niña-Southern Oscillation (ENSO) is a quasiperiodic climate pattern that occurs across the tropical Pacific Ocean roughly every five years. It is characterized by variations in the temperature of the surface of the tropical eastern Pacific Ocean; warming or cooling known as *El Niño* and *La Niña* respectively; and air surface pressure in the tropical western Pacific is called as the *Southern Oscillation*. The two variations are coupled: the warm oceanic phase, El Niño; accompanies high air surface pressure in the western Pacific, while the cold phase, La Niña; accompanies low air surface pressure in the western Pacific.

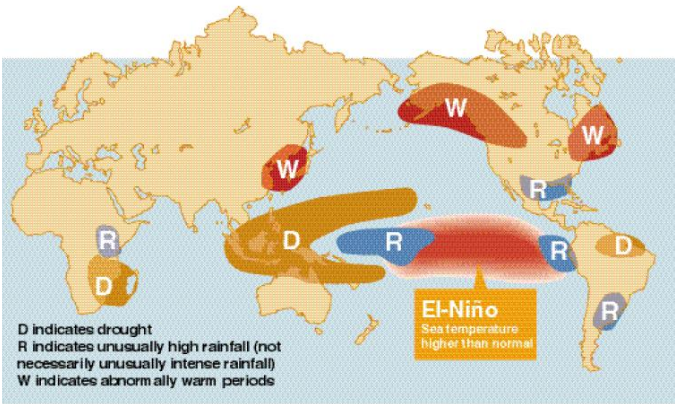


Figure 03: Graphical representation of the air surface temperature

ENSO causes extreme weather conditions in many regions of the world; and developing countries that depend upon agriculture and fishing, particularly those bordering the Pacific Ocean are the most affected (Ex. The damages from floods and landslides caused by very high rainfall in Peru and southern California; forest fires in Indonesia that have caused air pollution problems; crop failures and sometimes famine from droughts in southern Africa and the collapse of the Peruvian anchovetta fisheries because of warmer coastal waters).

Although ENSO is not responsible for all of the variability of climate, in many regions of the world it does have an impact on seasonal climate. Thus, overall, the damages caused in terms of loss of human lives and economic losses have been quite high.

## Geographical Information System

A geographic information system (GIS), geographical information science, or geospatial information studies is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographically referenced data. In the simplest terms, GIS is the merging of cartography, statistical analysis, and database technology. Following figure 04 simply illustrates how the GIS work; and the inputs and the output devices of the GIS.

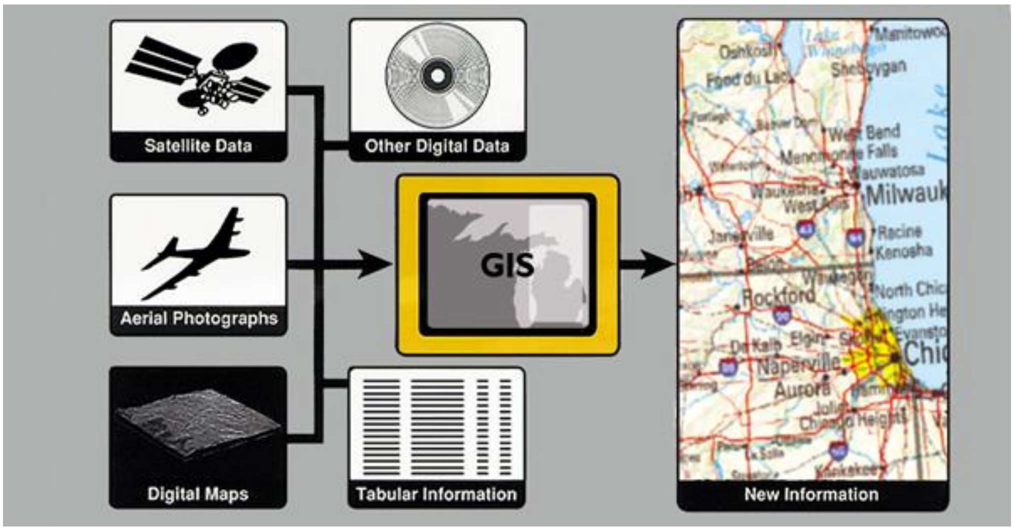


Figure 04: Graphical representation of how GIS works in producing important outputs

Modern GIS technologies use digital information for various digitized data creation.  The most common method of data creation is digitization, where a hard copy map or survey plan is transferred into a digital medium through the use of a computer-aided design (CAD) program, and geo-referencing capabilities. With the wide availability of ortho-rectified imagery (both from satellite and aerial sources), heads-up digitizing has become the main avenue through which geographic data is extracted.

GIS is used to view, understand, question, interpret, and visualize data in many ways in which relationships, patterns, and trends in the form of maps, globes, reports, and charts are revealed. Further, GIS helps to answer questions and solve problems by looking at the data in a way that is quickly understood and easily shared.

## Indian Ocean Dipole

Indian Ocean Dipole (IOD) is an irregular oscillation of sea-surface temperatures (SST) in which the western Indian Ocean becomes alternately warmer and then colder than the eastern part of the ocean. The IOD phenomenon was first identified by climate researchers in 1999. Yet evidence from fossil coral reefs demonstrates that the IOD has functioned since at least the middle of the Holocene period, 6500 years ago. An average of four; each positive/negative IOD events occur during each 30 year period with each event lasting around six months.

The IOD involves an aperiodic oscillation of SST, between "positive", "neutral" and "negative" phases. A positive phase sees greater-than-average SST and greater precipitation in the western Indian Ocean region, with a corresponding cooling of waters in the eastern Indian Ocean, which tends to cause droughts in adjacent land areas of Indonesia and Australia. The negative phase of the IOD brings about the opposite conditions, with warmer water and greater precipitation in the eastern Indian Ocean, and cooler and drier conditions in the west (figure 05).

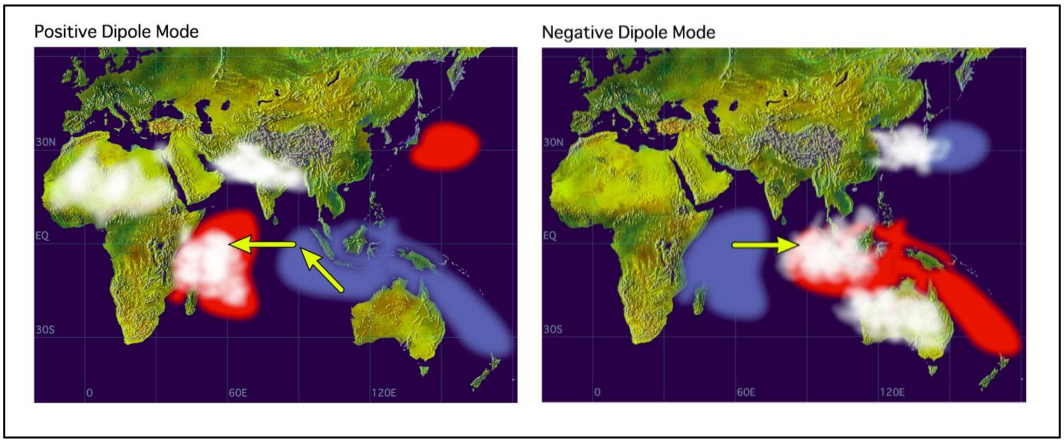


Figure 05: Graphical representation of positive and negative dipole mode.

SST anomalies shaded in red for warm anomalies and blue for cold. White patches indicate increased convective activities, and arrows indicate anomalous wind directions during IOD events.

The IOD also affects the strength of monsoons over the Indian subcontinent. The IOD is one aspect of the general cycle of global climate, interacting with similar phenomena like the El Niño-Southern Oscillation (ENSO) in the Pacific Ocean. However, studies show that the IOD has a much more significant effect on the rainfall patterns in south-east Australia than the ENSO in the Pacific Ocean.

## Canonical Correlation Analysis

Canonical Correlation Analysis (CCA) is a statistical method used to analyze the multivariate linear models. It is used in scenarios where there are more than one X variables and more than one Y variables. It identifies, analyzes and describes the relationship between the two sets of variables. In other words, CCA studies the relationship between a set of predictor (independent variables) and a set of criterion (dependant) variables, or between two pairs of vectors.

Compared to other statistical approaches this is more widely used in areas such as chemistry, biology, meteorology, demography, economics, and business management. CCA can identify the nature of the relationship between two sets of variables. Further, it has the possibility to identify number of Statistically Significant relations between the two sets. {For example, assume from a given data set it derives equations for accuracy of reading (x1), speed of the reading (x2), clearness of the reading (x3), and few measures. Say x4, x5, x6. In practice it is important to identify the relationships between accuracy, speed, clearness and other variables, and this can be done using CCA}.

Figure 06 expresses the correlations among the X variables.

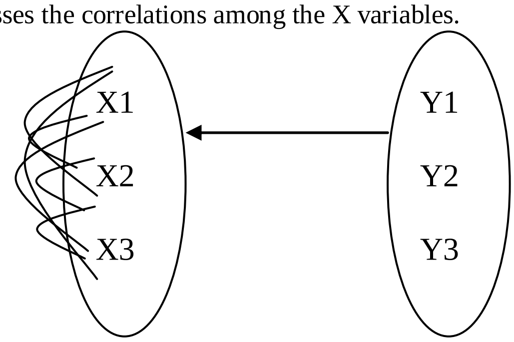


Figure 06: Correlation within a data set.

If the independent variables are in a Matrix format CCA is the most useful technique to apply in order to assess the relationship between the matrix dependant variable and the multiple dependant measures.

CCA generally expresses as follows:-

| |
|---|
| X1+ X2+ X3+X4 = Y1+ Y2+ Y3+ Y4 |
| (metric, nonmetric) (metric, nonmetric) |

## Land Surface Models (LSMs)

Land surface models are used to provide a boundary condition to hydrological models, representing the interactions between the land surface and the atmosphere. Consequently the land surface model must predict the radiation, water, heat and carbon exchanges, with explicit representation of vegetation and soil types.

Most of LSMs are dynamic, discrete-time, non-linear, state-space, deterministic and parametric. Normally LSMs comprise of a vertical and a horizontal structure (eg: can simulate a grid/box to a depth of several meters from the surface of the earth as well as represent several 'tiles' of vegetation etc. on the surface) . However characteristics used for each application varies from model to model.

Meteorological data such as precipitation, wind, snow, long wave radiation, short wave radiation, temperature, pressure, humidity are necessary driving inputs for information. Almost all LSMs need the same input of met data. Additional point/area information on soil characteristics and vegetation – type, Leaf Area Index (LAI), canopy information etc. are also needed for driving the model. The exact requirements vary from model to model. Usual outputs, though not restricted to, are – soil moisture, soil temperature, surface runoff, drainage from lower boundaries, surface fluxes of $CO_2$ and heat fluxes, plant transpiration, soil evaporation, and plant growth. Some models are more detailed while some comprises fewer details, but all of them have the first two as outputs. The time-step of the simulation varies from model to model. Some simulate 10 seconds or less while some models go for 30-minute or 60-minute time steps. LSMs are normally coupled to GCMs (Global Climate Models) but easily act as stand-alone models.

## Digital Elevation Model

A digital elevation model (DEM) is a digital model or 3-D representation of a terrain's surface. DEMs are data files that contain the elevation of the terrain over a specified area, usually at a fixed grid interval over the "Bare Earth". The intervals between each of the grid points will always be referenced to some geographical coordinate system. This is usually either latitude-longitude or UTM (Universal Transverse Mercator) coordinate systems. The closer or together the grid points are located the more detailed the information will be in the file. The details of the peaks and valleys in the terrain will be better modeled with small grid spacing than the grid intervals that are very large. Elevations other than at the specific grid point locations are not contained in the file. As a result peak points and valley points that donot coincide with the grid will not be recorded in the file. DEMs are commonly built using remote sensing techniques; however, they may also be built from land surveying. DEMs are often used in geographic information systems and serve as the most common basis for digital relief maps. The term Digital Elevation Model is often used as a generic term for Digital Surface Models (DSMs) and Digital Terrain Models (DTMs), only representing height information without any further definition about the surface.

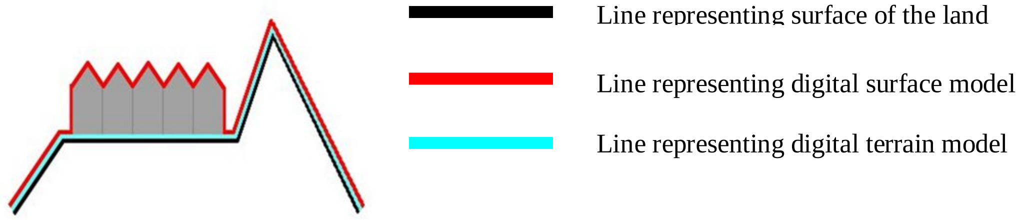


Figure 07: Schematic diagram representing the difference between DSMs and DTMs

The quality of a DEM is the measure of how accurate the elevation is at each pixel (absolute accuracy) and how accurately the morphology is presented (relative accuracy). Terrain roughness, sampling density (elevation data collection method), grid resolution or pixel size, interpolation algorithm, vertical resolution and terrain analysis algorithm are the factors that count for the quality of the DEM-derived products.

The DEM is used for new venture planning, modeling surface and ground water flow patterns, logistics planning, seismic planning, well site planning, pipeline route planning, facility planning, drainage analysis, cut-fill operations, creation of relief maps, extraction of terrain parameters, rendering of 3D visualizations and rectification of aerial photography or satellite imagery etc.

# 2. References


Anguluri Suryachandra. (2008). Indian Ocean Dipole (IOD). Available at: http://www.jamstec.go.jp/frsgc/research/d1/iod/ (accessed on August 22, 2011).

Ansari N. B. (n.d.). Data quality control. Available at: http://www.pitt.edu/~super1/lecture/lec7171/001.htm (accessed on September 26, 2011).
ESRI. (2011). Geographic information system. Available at: http://www.esri.com/what-is-gis/index.html (accessed on August 27, 2011).

Food Security and Nutrition Analysis Unit (n.d). Understanding the normalized difference vegetation index (NDVI). Available at: http://www.fsnau.org/downloads/Understanding_the_Normalized_Vegetation_Index_NDVI.pdf (accessed on September 22, 2011).

Glantz, M. H. (2001). Currents of change: Impacts of El-Nino & La-Nina on climate & society. 2nd edition. Cambridge University Press. UK.

Greenyer A. (2007). The Importance of Quality Control: How Good Is Your Data? Available at: http://www.customerthink.com/article/importance_quality_control_how_good_data (accessed on September 26, 2011).

National Oceanic and Atmospheric Administration. (n.d.). 2.B. Brief history of GLOBE. Available at: http://www.ngdc.noaa.gov/mgg/topo/report/s2/s2B.html (accessed on October 3, 2011)

Rabus, B., Eineder, M., Roth, A. & Balmer, R. (2003). The shuttle radar topography mission-a new class of digital elevation models acquired by a spaceborne radar. *Journal of Photogrammetry & remote sensing*, 57: 241-262.

Ramirez E. (2009). Shuttle Radar Topography Mission: The mission to map the world. Available at: http://www2.jpl.nasa.gov/srtm/index.html (accessed on September 27, 2011).

Sarachik, E. S. and Cane, M. A. (2010). The El Nino southern oscillation phenomenon Cambridge University Press. UK

Spatial Energy. (2011). Imager & Data Digital Elevation Models (DEMs). Available at: http://www.spatialenergy.com/products_digital.html (Accessed on September 26, 2011).

The International Research Institute for Climate and Society. (2003). Overview of ENSO impacts. Available at: http://iri.columbia.edu/climate/ENSO/societal/impact/index.html (accessed on September 27, 2011).

U. S. geological Survey. (2007). Geographic information systems. Available at: http://egsc.usgs.gov/isb/pubs/gis_poster/ (accessed on August 27, 2011).

U.S. Geological Survey. (2010). Shuttle Radar Topography Mission: Processing flow. Available at: http://srtm.usgs.gov/data/processing.php (accessed on September 27, 2011).

Wikipedia. (2011a). Canonical Correlation. Available at: http://en.wikipedia.org/wiki/Canonical_correlation (accessed on August 22, 2011).

Wikipedia. (2011b). Data Quality Control. Available at: http://en.wikipedia.org/wiki/Data_quality_control (accessed on August 22, 2011).

Wikipedia. (2011c). Digital Elevation Model (DEM). Available at: http://en.wikipedia.org/wiki/Digital_Elevation_Model (accessed on August 23, 2011).

Wikipedia. (2011d). El-Nino Southern Oscillation. Available at: http://en.wikipedia.org/wiki/El_Ni%C3%B1o-Southern_Oscillation (accessed on September 23, 2011).

Wikipedia. (2011e). Geographic information system. Available at: http://en.wikipedia.org/wiki/Geographic_information_system (accessed on August 22, 2011).

Wikipedia. (2011f). Land Surface Model (LSM version 1.0). Available at: http://en.wikipedia.org/wiki/Land_Surface_Model_%28LSM_version_1.0%29 (accessed on August 23, 2011).

Wikipedia. (2011g). Normalized Difference Vegetation Index. Available at: http://en.wikipedia.org/wiki/Normalized_Difference_Vegetation_Index (accessed on August 22, 2011).

Wikipedia. (2011h). Shuttle Radar Topography Mission. Available at: http://en.wikipedia.org/wiki/Shuttle_Radar_Topography_Mission (accessed on August 22, 2011).

World Meteorological Organization. (2004). Guidelines on quality control procedures for data from automatic weather stations. Available at: http://www.wmo.int/pages/prog/www/IMOP/meetings/Surface/ET-STMT1_Geneva2004/Doc6.1%282%29.pdf (accessed on September 26, 2011).

Foundation for Environment, Climate and Technology,
C/o Maintenance Office, Mahaweli Authority of Sri Lanka,
Digana Village, Rajawella, Sri Lanka
Tel: +94-81-2376746