Time Series Processing of MODIS Satellite Data for Landscape Epidemiological Applications

Markus Neteler
ITC-irst, SSI/MPBA, Via Sommarive, 18, 38050 Povo (Trento), Italy
E-mail: neteler@itc.it

Abstract

This paper reports on the processing of time series of MODIS NDVI/EVI and LST satellite data in a Geographical Information System (GIS). The required data preparations for the integration of MODIS data in GIS is described with focus on the reprojection from MODIS/Sinusoidal projection to national coordinate systems. To remove low quality pixels, the MODIS quality maps are utilized. We explain subsequent filtering of Land Surface Temperature maps with an outlier detector to eliminate originally undetected cloud pixels. Further analysis of time series is briefly discussed as well as related landscape epidemiological applications in the field of tick-borne diseases.

1. Introduction

In epidemiological modeling, survey data are usually collected at sampling sites and then regionalised in Geographical Information Systems (GIS). To enhance the spatial data density, continuous field data such as land surface temperatures (LST), snow coverage, and vegetation indices are commonly derived from satellite data. The launches of the new satellite systems *Terra* (December 1999) and *Aqua* (May 2002) significantly improve the situation of data availability for scientific purposes and predictive epidemiological studies. The *Moderate Resolution Imaging Spectroradiometer* (MODIS) is a key instrument on both Terra and Aqua satellites. Each delivers twice a day global coverages at 250 m (Red, NIR), 500m (MIR) and 1000m resolution (TIR) which make them most interesting to support epidemiological studies. Usually one week after acquisition the data sets are made available to the public.

The orbit of Terra around the Earth is scheduled to pass from north to south across the equator in the morning, while Aqua passes with reverted direction from south to north over the equator in the afternoon. Terra, crossing the equator at about 10:30 a.m. and 10:30 p.m. local solar time, is in a sun-synchronous orbit with a delay of 30 minutes with respect to LANDSAT-7. The further orbital parameters are equal to those of LANDSAT-7. Aqua is crossing around 01:30 a.m. and 01:30 p.m. local time.

MODIS is a whisk-broom sensor with 36 channels ranging from visible to thermal-infrared (GSFC/NASA, 2003). Data are delivered at 250m (2 channels), 500 m (5 channels) and 1000m resolution (29 channels). MODIS can be considered as a much enhanced successor of the AVHRR instrument onboard the NOAA series of satellites. MODIS improves upon the performance of AVHRR by providing both higher spatial resolution and greater spectral resolution, and has been used as a major source for the design for AVHRR's planned operational successor (VIIRS, http://www.ipo.noaa.gov/Technology/viirs_summary.html), currently anticipated for launch late in 2006. This paper

ISSN 1686-6576/© Geoinformatics International

focuses on two of the numerous MODIS data products, the Land Surface Temperature (LST) and Vegetation Index 16-day composites (NDVI and EVI).

2. Epidemiological Risk Maps and Study Region

The methods described in this paper have been implemented to extend an epidemiological study about the exposure risk to Lyme and tick-borne encephalitis (TBE) diseases transmitted by ticks in the Autonomous Province of Trento, Italian Alps. The predictions are carried out through the analysis of the distribution of *Ixodes ricinus* (L.) nymphs infected with *Borrelia burgdorferi* s.l. with a model based on tree-based classifiers. This model is supported by a Geographical Information System (Merler et al., 1996; Rizzoli et al., 2002; Furlanello et al., 2003). In this project data on *I. ricinus* (L.) density, assessed by dragging the vegetation in 438 sites during 1996 as well as data collected in 2000-2003 were cross-correlated with the digital cartography of a GIS.

The area includes the Autonomous Provinces of Bolzano and Trento and the Province of Belluno (Italy), a region of approximately 18,000 km^2 area size. The complex terrain has an elevation range from slightly above sea level to 3,800 m (Bolzano: elev. range 3,700 m, mean elev. 1,800 m, mean slope 26°; Trentino: elev. range 3,700 m, mean elev. 1,400 m, mean slope 26°; Belluno: elev. range 3,200 m, mean elev. 1,500 m, mean slope 27°). The density of available meteorological stations is highly varying and often concentrates in the valleys. To improve the risk mapping at high resolution, other sources of climatic and vegetation data are required. This demand can be fulfilled by using remote sensing data. The integration of satellite data into epidemiological research enhances the spatio-temporal resolution of climatological data in particular in mountainous regions where climatic stations and ground surveys are unavailable or sparse (Hess et al., 2002).

3. MODIS Data Pre-Processing for GIS Usage

MODIS data sets are delivered as “Base Level Swath Data” as well as “gridded data”. Grid data were initially projected in Integerised Sinusoidal (ISIN, Level V003), later in Sinusoidal (SIN, Level V004) projections. Both data types require further pre-processing before they can be used in a GIS. Especially ISIN is usually unsupported in most of “off the shelf” and free GIS and image processing software packages. The “MODIS Reprojection Tool” (MRT, U.S. Geological Survey 2004) can be used to reproject both ISIN and SIN to a more common projection (e.g., UTM) or to national grid systems such as Gauss-Boaga (Italy). Furthermore MRT allows for geographical subsetting. It writes the output to standard data formats such as Geo TIFF, and is executable on various operating systems. In April 2004 the V003 product were removed from the USGS archives as all existing MODIS data were reprocessed to V004. The V004 data quality is significantly higher, especially due to changes in the processing of inland water pixels (Wan, 2003).

Further pre-processing steps after the reprojection comprise the pixel-wise application of the quality map provided along each data product. This is followed by an outlier detection for certain MODIS products to minimise the presence of low quality pixels not properly indicated in the quality maps. The method is explained in greater detail in the following part.

3.1 Pre-Processing of Land Surface Temperatures (LST) data

MODIS Land Surface Temperature and Emissivity (LST/E) products are mapping land surface temperatures and emissivity values. The underlying algorithms use other MODIS data as input, including geolocation, radiance, cloud masking, atmospheric temperature, water vapour, snow, and land cover. Temperatures are extracted in Kelvin; an accuracy of 1 Kelvin is yielded for materials with known emissivities (Wan, 1999). After reprojection, we applied pixel-wise the related quality maps to the reprojected LST maps. This step is required due to limitations in the official cloud detection algorithm as used to create the land surface temperature quality maps. In particular, cloud detection during the night pass is error-prone at cloud corners. Also other quality variations occur due to aerosol and thin cloud presence which are not always indicated in the quality map. To overcome this problem, a simple outlier detection for the minimum temperatures was integrated into the proposed procedure:

$$lower\_boundary = 1st\_Quartile - 1.5 * (3rd\_Quartile - 1st\_Quartile)$$

Equation 1

Here we considered only LST maps with a sufficient number of pixels. To avoid that a few deviating pixels influence the overall outlier statistics, we were choosing as minimum the availability of at least 25% valid LST pixels in a map. In case of less usable pixels the entire map was rejected and not considered for the outlier detection. Quartiles and lower boundaries were calculated monthly-wise for all accepted LST maps. The LST maps were filtered then on a monthly base with the mean of the monthly lower boundary threshold values.

Due to the temperature differences in the day and night passes of MODIS the thresholds were considered separately. The proposed outlier filter aims at removal of all pixels which contain cloud top surface temperatures. Especially in the night overpasses these cloud contaminated pixels sometimes remain undetected by the NASA quality algorithm.

3.2 Pre-processing of Vegetation Indices (NDVI/EVI)

The two MODIS vegetation indices, the classical *Normalised Difference Vegetation Index* (NDVI) and a newly developed *Enhanced Vegetation Index* (EVI) are spectral measures of the amount of vegetation present on the ground. The MODIS-NDVI describes the relative “greenness” of the Earth’s vegetation on a scale from minus one (-1) to plus one (+1) and is intended to continue the time series more of than 20 years of NOAA/AVHRR-derived NDVI. The magnitude of NDVI is related to the level of photosynthetic activity in the observed vegetation. The EVI is MODIS-specific and offers improved sensitivity in high biomass regions while being less sensitive to atmospheric aerosol scattering (especially smoke from burning vegetation). It also minimises the influence of background interference caused by bare soil reflecting off the ground (Huete et al., 1999). These vegetation indices can be integrated into epidemiological models to reflect vegetation dynamics. The 16-day composite product minimises cloud cover problems by substituting a cloud covered pixel with a later uncontaminated pixel within a 16-days period. These composites reflect the current vegetation status at a sufficient temporal resolution. In order to use these MODIS map products in epidemiological applications, the beforehand reprojected NDVI/EVI maps have to be filtered using the related quality bit pattern maps. These quality maps contain information about pixel quality, aerosol contents, cloud, water or snow presence etc. As they are encoded in 16 bit, a bit pattern module was developed for GRASS GIS to perform this operation (r.bitpattern).

4. MODIS Time Series Analysis

The study has been carried out with GRASS GIS software (Neteler and Mitasova, 2004). The recent implementation of general time series processing (r.series) for GRASS raster maps supports univariate statistics for a series of MODIS scenes. By selecting various time ranges and operators, various indicators can be calculated.

4.1 Validation and Applications of Daily LST Data

To validate the usability of MODIS/Terra data in epidemiological studies as an enhancement of data availability, the monthly mean temperatures of selected meteorological stations and the related MODIS data at the same coordinates have been investigated. Figure 1 shows monthly mean temperatures of the Cavedine meteorological station (Trentino, Italy) compared to mean LST results from MODIS/Terra. Despite missing pixels due to cloud and high aerosol presence the curves are matching well except for January (1), February (2), and July (6). Table 1 reports on the number of observations used for the calculation of the monthly mean temperatures. Only in case of months with nearly continuous cloud cover the mean temperatures deviate significantly. Figure 2 shows a boxplot which confirms that the data distribution of both data types is similar. A two-sample Kolmogorov-Smirnov test on both distributions results in D=0.1667 and p-value=0.9985, confirming that the distributions significantly match. It is important to note that LST temperatures are not identical to air temperatures as measured by meteorological stations. However, efforts have been undertaken to derive air temperatures from surface temperatures (e.g., Goetz et al., 2000).

An application related to the distribution of tick-borne diseases is the calculation of "autumnal cooling". This index is calculated by linear regression to describe the autumnal temperature decline from August to October (northern hemisphere). Sites of tick-borne encephalitis (TBE) appear to be characterized by a high rate of autumnal cooling, relative to the annual maximum of the monthly mean LST level in midsummer (Randolph et al., 2000). Using the raster map time series calculator of GRASS (r.series), this index can be easily derived for all years of available MODIS data.

4.2 Applications of NDVI/EVI 16-day Composites

The temporal dynamics of vegetation are an important indicator of vegetation type and, consequently, moisture conditions on the ground. This is an important predictor of suitable habitat for *I. ricinus* (L.) (Randolph, 2001). Two applications can be identified: the direct integration of NDVI/EVI values into presence/absence models for infectious diseases. Then it is also promising to calculate temporal NDVI/EVI

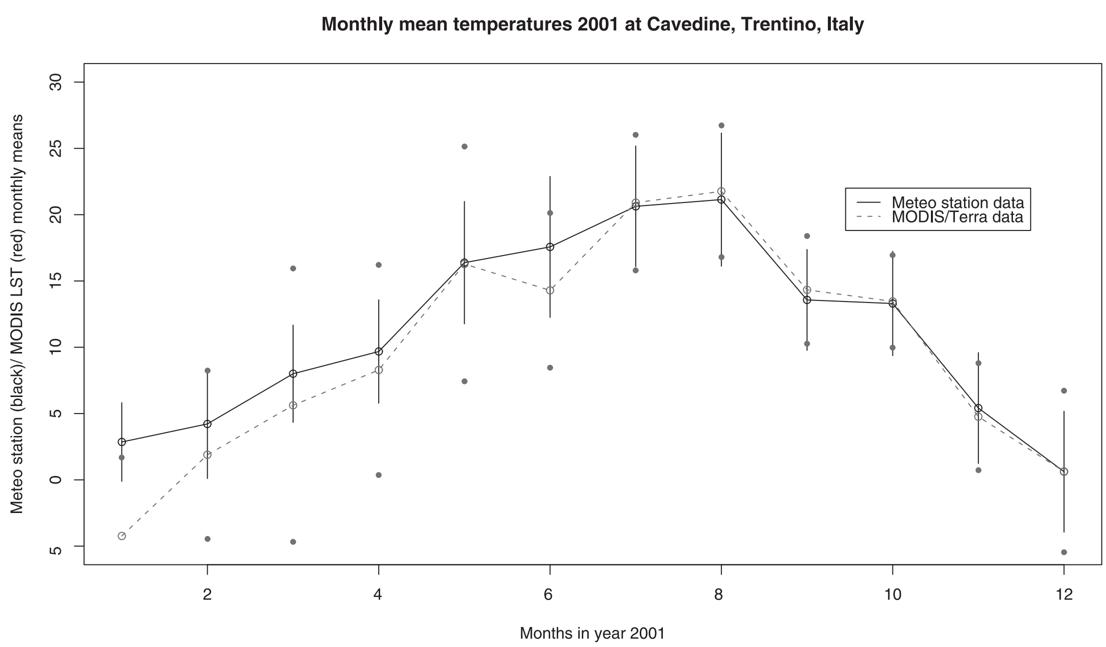

Figure 1: Comparison of monthly mean temperatures (2001) from Cavedine meteorological station (hourly data) and related MODIS/Terra V004 data (at max. two values per day). The standard deviation is indicated (bar for station, dot for MODIS)

Figure 2: Boxplot of monthly mean temperatures (2001) from Cavedine meteorological station (hourly data) and related MODIS/Terra data (zero, one or two values per day depending on the cloud cover). The solid line error bars belong to the meteorological stations, the dot error bars belong to the MODIS values.

Table 1: Number of observations for calculation of monthly mean temperatures (2001) from Cavedine meteorological station and related MODIS/Terra data

Data source	Month of year 2001												Act./Pot. Observations
	1	2	3	4	5	6	7	8	9	10	11	12	
Meteo	721	649	722	697	721	697	721	721	697	720	697	721	8484/8760
MODIS	10	32	15	29	23	8	29	35	24	35	32	20	292/730

differences in order to determine the spatial patterns of spring duration etc. These patterns may have influence on both the life-cycle of (small) mammals as hosts and ticks as vectors.

5. Conclusions and Future Research

The usability of MODIS LST and NDVI/EVI data for epidemiological studies appears to be promising. More that 3 years of data are available now, extending the 20 years of time series of AVHRR/NOAA. All available MODIS/Terra data sets have been processed for the study area. To enhance the temporal resolution of LST maps, the integration of the MODIS/Aqua data is ongoing which then provides four maps per day from May 2002 onwards (in case of cloud-free conditions). The study on TBE presence and its relationship to "autumnal cooling" is currently ongoing for the Trentino (Northern Italy) study area.

Further work is ongoing to relate the density of rodents as tick hosts to the temporal variations of the EVI over the year.

Acknowledgements

M. Neteler was supported by the FUR-PAT Project WebFAQ. The meteorological data have been provided from the Autonomous Province of Trento. Dr. A. Rizzoli/CEA is acknowledged for her support with ticks related data of the ECODIS project.

References

Furlanello, C., Neteler, M., Merler, S., Menegon, S., Fontanari, S., Donini, A., Rizzoli, A. and Chemini, C., 2003, GIS and the Random Forest Predictor: Integration in R for Tick-borne Disease Risk Assessment. In *Proc. 3rd International Workshop on Distributed Statistical Computing* (DSC2003), Edited by K. Hornik and F. Leisch, Vienna, Austria, CDROM.

Goetz, S., Prince, S. and Small, J., 2000, Advances in satellite remote sensing of environmental variables for epidemiological applications. In: *Remote sensing and Geographical Information Systems in epidemio-logy,* Edited by S.I. Hay and S.E. Randolph and D.J. Rogers (Academic Press), 289-307.

GSFC/NASA, 2003. MODIS technical specifications. Electronic document: http://modis.gsfc.nasa.gov/about/specs.html.

Hess, G., Randolph, S., Arneberg, P., Chemini, C., Furlanello, C., Harwood, J., Roberts, M. and Swinton, J., 2002, Spatial aspects of disease dynamics. In *The Ecology of Wildlife Diseases,* Edited by P.J. Hudson, A. Rizzoli, B.T. Grenfell, H. Heesterbeek and A.P. Dobson (Oxford University Press), 102-118.

Huete, A., Justice, C. and Leeuwen, W., 1999, MODIS *vegetation index (MOD 13)*. Algorithm theoretical basis document ATBD13. Electronic document: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod13.pdf.

Merler, S., Chemini, C. F. C. and Nicolini, G., 1996, Classification tree methods for analysis of mesoscale distribution of Ixodex ricinus (Acari: Ixodidae) in Trentino, Italian Alps. *Journal of Medical Entomology,* 33(6), 888-893.

Neteler, M. and Mitasova, H., 2004, *Open Source GIS: A GRASS GIS Approach. 2nd edition,* (Boston: Kluwer Academic Publishers).

Randolph, S., 2001, The shifting landscape of tick-borne zoonoses: tick-borne encephalitis and Lyme borreliosis in Europe. *Philosophical Transactions of the Royal Society of London B,* 356, 1045-1056.

Randolph, S., Green, R., Peacey, M. and Rogers, D., 2000, Seasonal synchrony: The key to tick-borne encephalitis foci identified by satellite data. *Parasitology,* 121, 15-23.

Rizzoli, A., Merler, S., Furlanello, C. and Genchi, C., 2002, Geographical information system and bootstrap aggregation (bagging) of tree-based classifiers for Lyme disease risk assessment in Trentino, Italian Alps. *Journal of Medical Entomology,* 39(3), 485-492.

U.S. Geological Survey, 2004. *MODIS Reprojection Tool V3.1*. Software: http://edc.usgs.gov/programs/sddm/modisdist/index.shtml.

Wan, Z., 1999. *MODIS Land-Surface Temperature*. Algorithm theoretical basis document (LST ATBD): LST calculations. Electronic document: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod11.pdf.

Wan, Z., 2003. *MODIS Land-Surface Temperature products users" guide*. Electronic document: http://www.icess.ucsb.edu/modis/LstUsrGuide/usrguide.html.