

# Correlation between normalized difference vegetation index and malaria in a subtropical rain forest undergoing rapid anthropogenic alteration

Nicole M. Wayant<sup>1</sup>, Diego Maldonado<sup>2</sup>, Antonieta Rojas de Arias<sup>3</sup>, Blanca Cousiño<sup>4</sup>, Douglas G. Goodin<sup>5</sup>

<sup>1</sup>Center For Advanced Land Management Information Technologies University of Nebraska – Lincoln Lincoln, NE 68506, USA; <sup>2</sup>Department of Mathematics Kansas State University Manhattan, KS 66506, USA; <sup>3</sup>Centro para el Desarrollo de la Investigación Científica CEDIC/R&D Díaz Gill/Fundación Moisés Bertoni, Paí Pérez 265 es. Mcal Estigarribia, Paraguay; <sup>4</sup>Servicio Nacional de Erradicación y Control de Vectores (SENEPA) Manuel Domínguez c/ Brasil Asuncion, Paraguay; <sup>5</sup>Department of Geography Kansas State University Manhattan, KS 66503, USA

**Abstract.** Time-series of coarse-resolution greenness values derived through remote sensing have been used as a surrogate environmental variable to help monitor and predict occurrences of a number of vector-borne and zoonotic diseases, including malaria. Often, relationships between a remotely-sensed index of greenness, e.g. the normalized difference vegetation index (NDVI), and disease occurrence are established using temporal correlation analysis. However, the strength of these correlations can vary depending on type and change of land cover during the period of record as well as inter-annual variations in the climate drivers (precipitation, temperature) that control the NDVI values. In this paper, the correlation between a long (260 months) time-series of monthly disease case rates and NDVI values derived from the Global Inventory Modeling and Mapping Studies (GIMMS) data set were analysed for two departments (administrative units) located in the Atlantic Forest biome of eastern Paraguay. Each of these departments has undergone extensive deforestation during the period of record and our analysis considers the effect on correlation of active versus quiescent periods of case occurrence against a background of changing land cover. Our results show that time-series data, smoothed using the Fourier Transform tool, showed the best correlation. A moving window analysis suggests that four years is the optimum time frame for correlating these values, and the strength of correlation depends on whether it is an active or a quiescent period. Finally, a spatial analysis of our data shows that areas where land cover has changed, particularly from forest to non-forest, are well correlated with malaria case rates.

**Keywords:** malaria, normalized difference vegetation index, time-series, land cover, Paraguay.

---

## Introduction

Remote sensing has emerged as an important tool for detection, surveillance, forecasting and control

of a variety of arthropod vectors and zoonotic infectious diseases. This approach has been used as a monitoring or predictive tool for a variety of infectious diseases, including malaria (Achee et al., 2006). In disease-related applications, remotely-sensed data have generally been used in two ways (Kazmi and Userly, 2001), namely (i) for the identification and mapping of land cover associated with host and/or organism habitats, and (ii) as a surrogate for meteorological or climatic conditions (e.g. precipitation and surface temperature) known to be

---

Corresponding author:  
Douglas G. Goodin  
Department of Geography  
118 Seaton Hall, Kansas State University  
Manhattan, KS 66506, USA  
Tel. +1 785 532 3411; Fax +1 785 532 7310  
E-mail: dgoodin@ksu.edu

associated with population dynamics of host and/or vector (Hay and Lennon, 1999). The first of these two applications typically rely on moderate or fine resolution imagery (typically varying between 1 and 30 m) that is capable of classifying land use/land cover using a variety of statistical or mathematical pattern-finding techniques (Curran et al., 2000). Often, the data used in these applications are available only at infrequent time intervals.

The second application for remotely-sensed data in infectious disease analysis makes use of image data from sensors such as NOAA-AVHRR, SPOT-VEGETATION or EOS-MODIS (Bergquist and Rinaldi, 2010). The spatial resolution of these sensors (~1-8 km) is frequently too coarse to reveal details about vector habitats, but the data from these sensors are useful in environmental analysis because they are typically available at temporal frequencies as short as one day, yielding a dense time-series of information about the dynamics of the land surface. A common strategy for utilizing these coarse resolution data time-series is to use reflectance data from the visible and near infrared (VNIR) region of the spectrum to calculate the normalized difference vegetation index (NDVI), a simple index formed by dividing the difference between reflectance in the red and near infrared spectral regions by the sum of the reflectance in the same two bands (Tucker, 1979). NDVI correlates to a number of biophysical parameters, including leaf area index (LAI), biomass and fraction of absorbed photosynthetically active radiation (FPAR). Since these canopy parameters are closely linked to the moisture status of the plant environment, which in turn is linked to meteorological/climatic conditions, NDVI is often used as an indirect measure or surrogate for precipitation (Schultz and Halpert, 1993).

The capability of NDVI time-series to monitor and predict vector-borne diseases (particularly those transmitted by mosquitoes) depends on the correlation of both disease incidence and vegetation greenness with precipitation. Although the relationship between precipitation and the life-cycle of the vector is complex and depends on the species of mosquito,

a number of studies have established a positive relationship between precipitation and the breeding cycle of various mosquito vectors (Kazmi and Usery, 2001). NDVI has consistently been shown to be correlated to precipitation – although the exact form and strength of that correlation depends on the ecosystem and cover type in question (Fuller and Prince, 1996). It therefore follows that NDVI should also be correlated to the occurrence of diseases, the vector habitats of which are linked to rainfall. Much of previous work on correlation of NDVI with vector-borne disease has been done to monitor and predict malaria in sub-Saharan Africa. For example, in Kenya, Hay et al. (1998) used Fourier analysis of a 5-year AVHRR time-series to predict malaria transmission seasons. Mushinzimana et al. (2006) used moderate-resolution image data to map vector habitats. Elsewhere in Africa, Dambach et al. (2009) and Ceccato et al. (2007) used time-series of coarse resolution data as climate surrogates and independent variables for predicting malaria transmission.

Much of the research cited above was conducted in sub-Saharan Africa where the land cover is dominated by savanna and grassland. Both of these cover types are associated with more xeric conditions and tend to respond more rapidly to variation in precipitation (Fabricante et al., 2009). Land cover in these areas is also relatively stable. Less is known about the correlation between NDVI and disease rates in tropical or sub-tropical forested areas in South America, where land cover is changing rapidly due to human activities.

In this paper, we report the results of correlating a long (>25-year) malaria time-series with NDVI for a forested, subtropical site in eastern Paraguay undergoing extensive deforestation. In the Regional Office for the Americas of the World Health Organization (EMRO), Paraguay currently ranks 16<sup>th</sup> with respect to the number of malaria cases, while it was a highly endemic country in the past. When normalized to population size, the malaria case rate in Paraguay during the analysis period falls within the top 10 in the region. Like many coun-

tries, Paraguay is vulnerable to increased malaria incidence due to climate change (Rojas de Arias, 2002). The majority of malaria cases in Paraguay occur in a sparsely populated areas along the frontier with Brazil, thus the malaria cases can be quite a burden to the local health services. The impact of malaria in Paraguay during the period of the study was significant, signifying the need for a better understanding of environmental factors which affect disease rates.

In this paper, we investigate the relationship between malaria cases and NDVI time-series by addressing two research questions:

- (i) does NDVI significantly correlate with malaria over longer time periods?; and
- (ii) does land cover change effect the correlation between the two?

## Materials and methods

### Study area

The study area for this research consists of two departments (the principal administrative subunit in Paraguay) within the Republic of Paraguay: Alto Paraná and Canindeyú. Both these departments are located in eastern Paraguay near the common frontier with Brazil and Argentina. The region is within the Atlantic Forest biome, one of two major biomes of Paraguay (Fig. 1).

Historically, the Atlantic Forest was once the most extensive subtropical rain forest in South America, extending from the Brazilian coastal plain inland into eastern Paraguay and Argentina. With over 20,000 plant species (8,000 endemic), 850 species of birds (180 endemic), plus numerous endemic amphibians, reptiles, small mammals and primates, the forest supported one of the most extensive and varied flora and fauna of South America. The Atlantic Forest is recognized as an international biodiversity “hotspot” (Galindo-Leal and Gusmao-Camara, 2003), a designation based on its high diversity and level of anthropogenic threat. Extensive deforestation, driven by timbering, ranch-

ing and agriculture, has reduced the Atlantic Forest from its estimated original area of about 1.2 million km<sup>2</sup> to a current area of approximately 100,000 km<sup>2</sup>. In Paraguay, the remaining Atlantic Forest is highly fragmented, with forest remnants of various sizes being interspersed with pasture and crop land (typically sugar cane, coffee and soy). Human populations in the Atlantic Forest have also increased, resulting in further land clearance and development. Land cover in both departments has been altered by the presence of the Itapúa hydroelectric dam on the Paraná River. Completed in 1984, the dam and resulting lake flooded 1,400 km<sup>2</sup> along the Paraguay-Brazil border.

The departments Canindeyú and Alto Paraná are both located within the region of malaria-endemicity in Paraguay. Malaria cases in the region are generally associated with *Plasmodium vivax*, transmitted by the vectors *Anopheles darlingi*, *An. strodei*, and *An. albitarsis*. Although land cover change affects the number of malaria cases, this has not been directly investigated in eastern Paraguay. The results of similar analyses in the Amazon rain forest

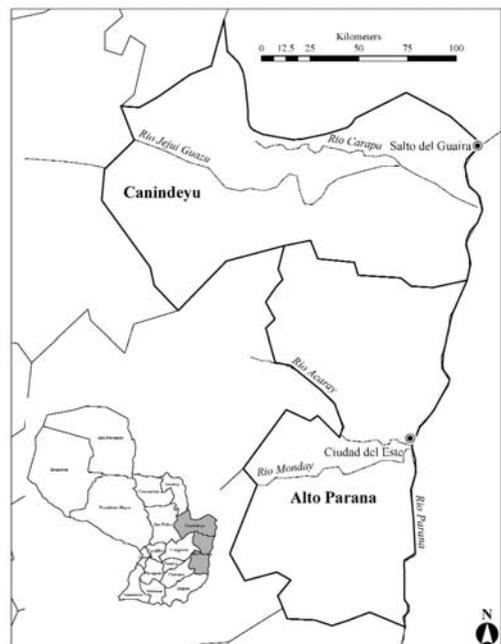


Fig. 1. Map of Paraguay showing the two departments where the current study was carried out.

suggest that certain types of deforestation, particularly those resulting in fragmented forest canopies with substantial shade and proximity to water, may increase the mosquito larval habitat area and possibly the biting frequency of the vector species (Castro et al., 2006; Vittor et al., 2006, 2009).

#### *Remote sensing data*

We used a 260-month time-series (July 1981 - February 2003) of NDVI data subset from the Global Inventory Modeling and Mapping (GIMMS) data set (Tucker et al., 2005). GIMMS is a global NDVI data set derived from 8 km AVHRR Global Area Coverage (GAC) data. GIMMS data consist of pixels from a number of AVHRR instruments, corrected for orbital drift and sensor cross-calibration error using empirical node decomposition (Pinzon et al., 2005). Although the data are available as bimonthly composites, we used only one scene per month. Each department was collected as separate subsets and then stacked into a 260-band image for further analysis. All basic image manipulations were done using version 4.1 of the ENVI image processing environment (ITT Visual Information Solutions; Boulder, CO, USA), and the IDRISI Andes software (Clark Labs; Worcester, MA, USA).

For the land cover analysis, we used the NASA GeoCover dataset (Tucker et al., 2004). The GeoCover dataset provides Landsat imagery for three time periods, circa 1970-1980, 1990 and 2000. Depending on the time period, the data are from either the Landsat Multispectral Scanner (MSS), the Thematic Mapper (TM), or the Enhanced Thematic Mapper (ETM)+ sensors. All GeoCover data are selected for low cloud content, and are orthorectified for high geometric fidelity. For our study area, the data used were acquired in 1975, 1990 and 2000.

#### *Malaria data*

The number of malaria cases for Alto Paraná and Canindeyú was obtained for the same time period as

the NDVI data from the Servicio Nacional de Erradicación y Control de Vectores (SENEPA), Asuncion, Paraguay. To account for population variation over time, the raw data were transformed using a moving window approach, in which the number of cases within each window were determined and then divided by the average number of cases for that window. This procedure was carried out for both departments with window sizes varying from 3 to 7 years.

#### *Correlation analysis*

To address our research questions, we used three related analyses: (i) correlation of the rates of malaria cases with an NDVI time-series consisting of monthly values spatially averaged over each department; (ii) correlation of NDVI averaged over each department with case rates within the moving windows as described below; and (iii) correlation of individual pixel time-series with the case numbers to determine spatial patterns of correlation relative to land cover change during the study period. The methods used for each analysis are explained in further detail below. All image processing were done using code developed with the MATLAB programming language (The Mathworks, Natick, MA, USA, 2007).

As most types of experimentally-collected data, remotely sensed NDVI time-series are unavoidably affected by noise. A reliable, realistic assessment of the correlation of NDVI time-series and any other time-series must account for the presence of noise by implementing some noise-reduction techniques. In our present analysis, we have addressed this issue by means of Fourier analysis, which is justified by the underlying cyclic nature of NDVI data. Indeed, seasonally changing NDVI encodes periodic features, best captured by Fourier analysis, whose main tool, the Fourier Transform, is precisely based on the representation of time-series as a superposition of the periodic sine and cosine functions at different frequencies. To demonstrate the efficacy of the smoothing process, we conducted a correlation

analysis using both the unaltered data series (as received from the original sources) and a Fourier-smoothed NDVI time-series.

To assess the overall correlation between NDVI and case rates in each department, we spatially averaged all NDVI pixel values, and then correlated this averaged value to overall case rates in each department using the Pearson's product moment correlation coefficient ( $r$ ). Pearson's  $r$  assumes a linear relationship between two variables, thus our analysis is not sensitive to any non-linear relationship that might exist. Vegetation activity generally follows precipitation (a major environmental correlate of malaria); we therefore lagged the correlation between NDVI and case rates by values ranging from 0 to 6 months.

As earlier noted, the departments of Alto Paraná and Canindeyú have undergone significant land cover change (especially deforestation) throughout the study period. As these land cover changes would tend to alter the temporal correlation between the two time-series, correlations over shorter time periods might be higher than for the entire series. In order to assess the correlation at shorter time periods, we applied a moving window analysis to each of the smoothed, spatially averaged NDVI time-series. Each moving window was defined by its length, which varied from 3 to 7 years (36 to 84 months). Values of smoothed NDVI and malaria case frequency were correlated within a moving window, after which the window was shifted one month further along the time-series and the correlation was repeated. This analysis necessarily resulted in a series of  $r$ -values across the time-series. Thus, these  $r$ -values represent the temporal evolution of the local (36 to 84 months) correlation between smoothed NDVI and malaria cases.

**Results**

*Moving window correlation*

The correlation values from the moving window analysis applied to area-averaged NDVI values

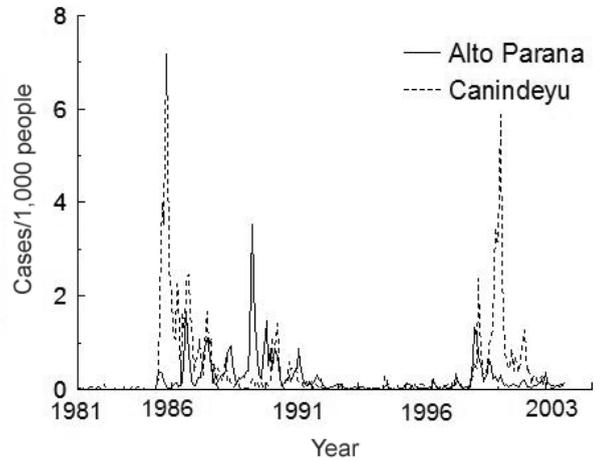


Fig. 2. Malaria case rates for Alto Paraná and Canindeyú, 1981-2003.

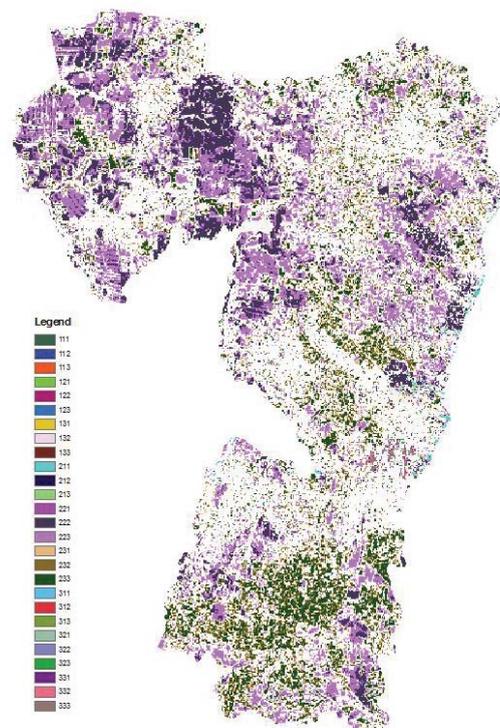


Fig. 3. Land use/land cover change over the period 1975-2000. Classified pixels for three separate years (1975, 1990 and 2000) were classified using a simple scheme (forest, non-forest, water), then sorted into change classes based on transitions from one class to another.

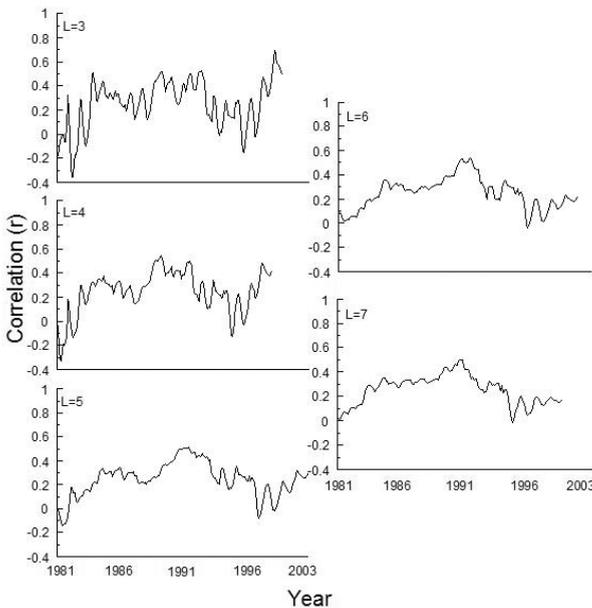


Fig. 4. Correlation values for Alto Paraná using moving window lengths (L) of 3, 4, 5, 6 and 7 years.

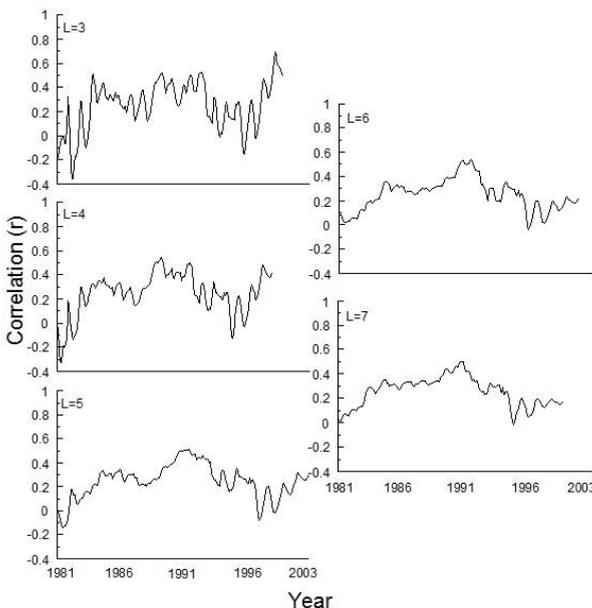


Fig. 5. Correlation values for Canindeyú using moving window lengths (L) of 3, 4, 5, 6 and 7 years.

showed high temporal variability in both departments throughout the study period (Figs. 4 and 5). Comparison of the results from the two departments shows some notable differences. In general, the correlation was higher in Alto Paraná compared to Canindeyú. This is consistent with the results from the previous analysis, where the entire time-series were correlated. Not surprisingly, the range of  $r$ -values over both the Canindeyú and Alto Paraná time-series varied with the length of the moving window. Generally, the highest  $r$ -values were found with the shortest windows; however, the general pattern of temporal variability was similar for all windows. Overall, the best correlations were found with moving windows of 4 years (48 months). Maximum correlations at this window length reached  $r = 0.73$  during periods of high rates of malaria cases. The temporal patterns were not the same for both departments. In Alto Paraná, the highest  $r$ -values were found for moving windows encompassing the years 1987 through 1990. In Canindeyú, the  $r$ -values were generally less (compared to Alto Paraná) and the temporal pattern was slightly more complex. The highest  $r$ -values occurred from 1989 to 1992, and again at the end of the time-series. In quiescent periods (few reported malaria cases) the correlation values became quite low or even negative.

For both departments, the temporal pattern of correlation values at all window sizes mirrors the case rates. In Alto Paraná, the majority of cases occurred during two peak time periods. The first and largest of these (in terms of case rates) occurred between 1985 to 1990-1991, with peak rates occurring in 1989. The second period was both shorter and less active, ranging from 1998-2001 and peaking in 1999. These active periods were surrounded by intervening quiescent periods, where case rates were lower. In Canindeyú, there were three periods of high activity, one ranging from 1984-1986, a second (less pronounced) from 1989-1991, and a third in 1998-2001. Case rates dropped to quiescent levels in intervening times. The active periods in Canindeyú correspond

roughly to those in Alto Paraná, except for the relative quiescent period in 1988. Note that although the peak cases rates varied widely, there is a suggestion of periodicity of case rates in both departments, which appears to be fairly regular and seasonal. These active periods correspond to the times of highest correlation between NDVI and the occurrence of malaria. This is not surprising, since in order for two time-series to correlate, there must be some systematic variation in the magnitude of both. During the quiescent periods, there is little or no variation in malaria case rates in either department. This observation reveals a general weakness in the use of NDVI as a surrogate for correlating environmental variation with malaria (and perhaps diseases); temporal variability in NDVI occurs primarily due to seasonal canopy variability, and will always show an intra-annual pattern of variability. The strength and intensity of this variability may depend on climatic variation (especially precipitation), which may in turn be linked to global or regional circulation patterns (Tsonis et al., 2008) but the seasonal variation will always be present. In contrast, the rates of malaria cases, while somewhat seasonal (see Fig. 2), are characterised to a much greater degree by the pattern of active periods with intervening quiescent periods as discussed above. It is reasonable to expect that the  $r$ -values should be higher when the case rates are higher and show a regular pattern of variability, simply because the dynamic range in case-rates is greater during these times. Similarly, correlations should be lower when case rates are low and variation irregular.

The contrast in correlation values between times of high and low case occurrence also helps to interpret the effects of moving window length. As earlier noted, overall correlations tended to be higher at shorter window lengths in both departments. Selection of a “best” window length is a somewhat subjective exercise, nevertheless based on the criteria of highest mean and maximum correlation, the 4-year window seems to be the best for these data. Systematic change in LULC in the study area might

also play a role in limiting the accuracy of correlation over longer time periods, a factor considered in more detail in the next section.

#### *Spatial-temporal patterns of correlation*

In both the area-averaged and moving window analyses described in the previous two sections, we averaged all smoothed NDVI pixel values for each of the two departments. This was done because malaria case rates data were only available at the department level. However, spatially averaging the NDVI values has the effect of “blurring” the influence of LULC, and potentially lowering the overall correlation in each department. This would be especially true if the temporal NDVI trajectories associated with a particular LULC type correlated more closely with the malaria case rates compared to others. Also, one of the motivating questions behind this analysis is the effect on temporal correlation of NDVI when LULC changes over the course of the analysis period. To address this question, we submitted the correlation values obtained from using a 4-year moving average window for each pixel vector in each department to standardised temporal principal component analysis (PCA). The various components resulting from PCA have the effect of revealing major patterns in temporal data. In particular, the first principal component generally captures the overall trend of a time-series (Eastman and Fulk, 1993) and is therefore well-suited for summarising the overall pattern resulting from using a series of moving window correlations across the entire data set. Application of standardised PCA results in a new image where each pixel value represents the contribution (positive or negative) of that pixel vector to the overall component loading value, thus summarising the major spatial-temporal trends and patterns throughout the study period (see Fig. 6).

The heterogeneous spatial pattern of component loadings in Figure 6 suggests that land cover (and land cover change) plays some role in determining the correlation between temporal NDVI trajectories and malaria cases. To test this association, the

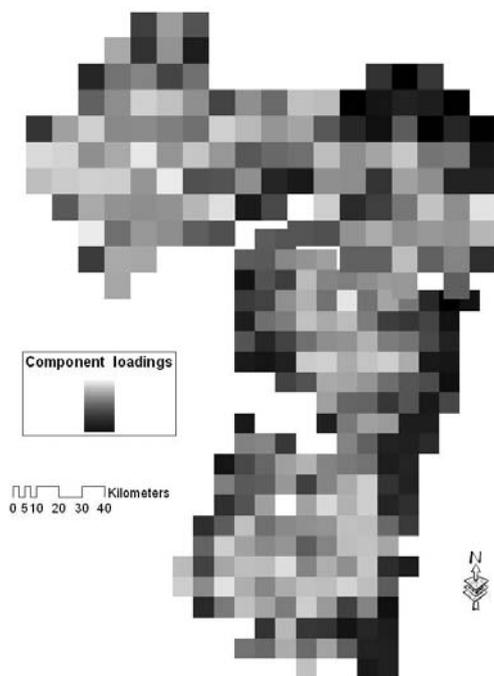


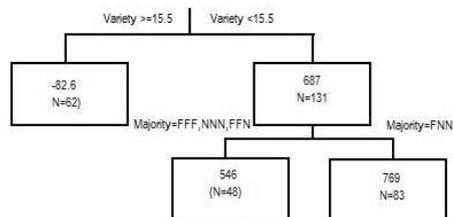
Fig. 6. Map of first principal component for each pixel vector in the study area, showing the contribution of each individual pixel vector to the overall component loading value. Note the pixel size relative to the LULC map (Fig. 3).

$\Delta$ LULC map and the PCA loading map were superimposed and the composition of the various LULC categories present in each PCA loading map pixel determined. The spatial resolution of the PCA loading map was  $8 \times 8$  km (reflecting the fact that it was derived from GIMMS data), while the resolution of the  $\Delta$ LULC map (derived from Landsat MSS and TM data) was  $60 \times 60$  m. Thus, each pixel in the PCA loading map contained about 17,800  $\Delta$ LULC pixels. None of the PCA loading pixels were entirely associated with the same  $\Delta$ LULC class, and in fact all of the  $8 \times 8$  km pixels in the factor loading map contained between 9 and 20 change classes. We therefore determined the most frequently occurring  $\Delta$ LULC class for each loading pixel (*majority*), along with the total number of classes (*diversity*).

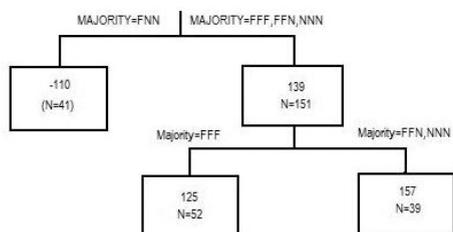
After determining the *majority* and the *diversity* for each loading pixel, a regression tree (Breiman et al., 1984) was used to assess the relationship

between the  $\Delta$ LULC properties of each pixel and its loading value. Regression trees analysis is appropriate in situations such as these where data exploration is the goal and the distributional characteristics of the input data are unknown (De'ath, 2002) as its goal is to create an explanatory rather than a predictive model. Thus, we made no attempt to validate the model other than to note its standard error. Alto Paran and Canindey were analysed separately, yielding separate regression tree models for each department (Fig. 7). Although the resulting tree models appeared similar, there were noteworthy differences between the two that not only showed that land cover change is associated with the occurrence of malaria, but also indirectly shed some light on how  $\Delta$ LULC might be related to the case rates.

In Alto Paran, both the *majority* and *diversity* variables contributed to the categorization of the PCA loading pixels. More homogenous (i.e. less diverse) pixels were associated with high positive loading values (with a mean of 687), whereas more diverse pixels contributed negatively to the PCA loading. This is perhaps not surprising, since a more heterogeneous surface is likely to be more strongly associated with a particular outcome. The second split in the data, involving only the positively associated class, resulted in two terminal nodes associated with differing  $\Delta$ LULC trajectories. The highest positive associations were with the FNN class, implying that deforestation in these pixels occurred sometime between 1975 and 1990. In fact, transition from forest to non-forest is not exclusively associated with deforestation, but the land cover history of this general area suggests that this is the most likely cause for this transformation (Huang et al., 2007). Loading pixels that were associated with no change (NNN, FFF), along with pixels where deforestation occurred later (FFN), were also associated with high positive loading values (and thus positively associated with malaria cases), but the loading values were lower than those in the other terminal node. Note that the majority of homogenous pixels (83 vs. 48) fell into the FNN terminal node. From this we conclude that less homogeneous



7a



7b

Fig. 7. Regression trees for Canindeyú and Alto Paraná. The numerical value in each box is the mean value for pixel loadings belonging to that class. N-values in parentheses are the total number of pixels associated with each terminal node. Annotation on the tree branches shows the variable and values associated with each split. The standard error (SE) for the Canindeyú model was 0.132, for the Alto Paraná model, SE = 0.077.

pixels where deforestation had occurred were more likely to be positively associated with the first principal component of the correlation data, and this with the malaria case rates.

For Canindeyú, the best-fit regression tree did not include *diversity* as a significant variable. The tree had two splits, both determined by *majority* values. Pixels where the  $\Delta$ LULC pattern was FNN were negatively associated with the first principal components, while all other land cover change patterns were positively associated with the first component. The second split in the tree produced two terminal nodes that were both positively associated with the first component but with notable difference in the strength of that association. Pixels that remained forested throughout the study period (FFF) were marginally associated with the first component, whereas those pixels that were deforested between

1990 and 2000 were much more strongly associated with the first component. The NNN class landed in the same terminal node as FFN, but there was only one pixel from the component image in this class.

## Discussion

The results of this research both support and extend the findings of previous work where temporal series of NDVI has been associated with vector-borne infectious diseases. As in previous studies, we found that area-averaged values of NDVI did significantly correlate with malaria occurrence, although neither was especially strong for the NDVI values averaged over the entire department (significant correlations values below 0.30). In evaluating this result, however, it must be remembered that direct correlation of NDVI (a biophysical factor) with malaria case-rates does not explicitly consider all the factors that intervene between the climatic drivers of vector-borne disease, which includes the ecology of the vector and the pathogen, as well as the transmission of the pathogen to humans. Vector control, medical intervention and other factors might act as “hidden” variables in this analysis, affecting the relationship between the primary variables explicitly included in the correlation analysis.

The best correlations were found when the NDVI time-series lagged the disease series. This lagging effect is informative from a biophysical perspective, since it shows that disease responds to environmental drivers, but it is a problem for prediction since it implies that disease may already be occurring (due to increases in precipitation that facilitate the reproductive ecology of the vector) before the NDVI responds. This suggests that further research should concentrate on leading indicators of climate conditions or “trigger events” (Pinzon et al., 2004) within the NDVI time-series, if any can be derived from the data.

The results of the moving window analysis show that the best correlations between the smoothed NDVI and malaria time-series occur with window

sizes of 4 years or 48 months. At this window size, the correlation between the two time-series varies from negative values to about 0.75 (Figs. 4 and 5). Regardless of the moving window size, the correlation values show wide variation across the whole time-series. The reason for this variation is fairly obvious – comparison of the *r*-values with the malaria case time-series (Fig. 2) shows that better correlations occur at times when case rates are higher and are showing a definite trend (i.e. rising or falling). During quiescent periods, when case rates are low, there is no discernible trend in the case data for the smoothed NDVI to correlate with – hence the irregular, low (sometimes negative) correlations. This observation reveals one of the major drawbacks with using NDVI as a precipitation surrogate for surveillance or prediction of infectious diseases linked to climate – NDVI continues to show a seasonal trend (matching the phenology of the underlying vegetation) regardless of the magnitude of precipitation amounts. Thus, the NDVI trend is unlikely to correlate with case occurrence over longer time periods.

The type of land cover present and any trajectory of change in that cover type can also influence the temporal correlation between the series. Much of the previous work relating temporal NDVI to various vectored disease has been done in savanna or grassland areas of sub-Saharan Africa (e.g. Eisele et al., 2003; Ceccato et al., 2007). Grasslands and savanna are more sensitive to climatic fluctuations (especially precipitation), compared to forest, thus NDVI reacts more quickly. Our study area also underwent significant land cover change over the 24-year period of record. This land cover change further complicated the temporal relationship between NDVI and malaria cases, by replacing forest with crop or grassland having different temporal NDVI trajectories.

The results of our spatially-explicit analysis suggest that the strongest associations between case frequency and NDVI occurred when the land cover changed during the period of record, and that the periods of greatest malaria occurrence coincided

with times of active deforestation. In Alto Paraná, pixels that were deforested between 1975 and 1990 were more likely to contribute to higher loadings on the first principal component, indicating higher correlation with malaria cases. Examination of the rates of malaria cases (Fig. 2) show that the highest rates occurred in the late 1980s, during the interval when much of the deforestation in Alto Paraná took place. This result implies that deforestation may be related to malaria case rates, a finding consistent with results in other tropical/subtropical rainforest regions (Vittor et al., 2009). The pattern of change was not entirely the same in Canindeyú, where pixels in which deforestation occurred in the second interval (between 1990 and 2000) were most likely to be more strongly associated with the first principal component. It is not surprising that later deforestation was significant in Canindeyú, since land cover change occurred later there (Huang et al., 2007). However, forest loss does not seem to have been associated with high malaria occurrence in the 1980s. This is likely due to localization of malaria transmission in the department.

Regarding our two original research questions, we can draw the following conclusions. First, temporal trajectories of NDVI are associated with malaria case rates, but over the relatively long time-series considered here, the strength and duration of this association varies widely. The best time interval for correlating malaria cases to NDVI time-series in these data was found to be about 4 years. However, the strength of the correlation varied over time depending on the case rates. The NDVI correlated well with case rates during times of increased case occurrence, but not during quiescent periods. In practical terms, this might limit the efficacy of the NDVI as a predictor of the onset of a period of increased malaria occurrence. The significance of this for long-term monitoring is unclear, but suggests that while coarse resolution NDVI values may help predict the magnitude of case numbers once an active period is underway. Other information (e.g. climatic forecasts, see Pinzon et al., 2005) may be needed in order to understand factors that initiate

active periods. Further research is needed to derive environmental signals that might indicate the onset of increased malaria cases.

Second, land cover change does seem to be associated with malaria case rates, but higher resolution land cover change data are needed in order to define the nature of this relationship. Some of the results shown here suggest that land cover change may be a factor in initiating active disease periods. However, as with most disease processes, it is unlikely that a single variable can completely explain the case dynamics. Our results might also be useful for providing information on what types of LULC might correlate best with the occurrence of malaria. Such analysis could help in understanding associations between land cover change and malaria case rates in eastern Paraguay, ultimately enabling an advance from temporal to spatio-temporal prediction and surveillance.

## Abbreviations

LULC = Land Use / Land Cover

NNN = Not-Forest, Not-Forest, Not-Forest

FFF = Forest, Forest, Forest

FFN = Forest, Forest, Not-forest

## Acknowledgements

We thank the Servicio Nacional de Erradicación y Control de Vectores (SENEPA), Asuncion, Paraguay, for kindly providing the malaria case data used in these analyses. We further acknowledge the Dirección General de Estadísticas, Encuestas y Censos (DGEEC) in Paraguay for providing population data. NMW and DGG acknowledge support from the Kansas NASA Space Grant Consortium (KSGC). NMW was further supported during this project by an Undergraduate Student Scholarship from KSGC. DM acknowledges support from NSF award DMS 0901587.

## References

- Achee NL, Grieco JP, Masuoka P, Andre RG, Roberts DR, Thomas J, Briceno I, King R, Rejmankova E, 2006. Use of remote sensing and geographic information systems to predict locations of *Anopheles darlingi*-positive breeding sites within the Sibun River in Belize, Central America. *J Med Entomol* 43, 382-392.
- Bergquist R, Rinaldi L, 2010. Health research based on geospatial tools: a timely approach in a changing environment. *J Helminthol* 84, 1-11.
- Breiman L, Friedman JH, Olshend RA, Stone CJ, 1984. *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA, USA.
- Castro MCD, Monte-Mor RL, Sawyer DO, Singer BH, 2006. Malaria risk on the Amazon frontier. *Proc Natl Acad Sci USA* 103, 2452-2457.
- Ceccato P, Ghebremeskel T, Jaiteh M, Graves P, Levy M, Ghebreselassie S, Ogbamariam A, Barnston AG, Bell M, del Corral J, Connor SJ, Fesseha I, Brantly EP, Thomson MC, 2007. Malaria stratification, climate, and epidemic early warning in Eritrea. *Am J Trop Med Hyg* 77, 61-68.
- Curran PJ, Atkinson PM, Foody GM, Milton EJ, 2000. Linking remote sensing, land cover and disease. *Adv Parasitol* 47, 37-80.
- Dambach P, Sie A, Lacaux J-P, Vignolles C, Machault V, Sauerborn R, 2009. Using high spatial resolution remote sensing for risk mapping of malaria occurrence in the Nouna district, Burkina Faso. *Global Health Action*, 2.
- De'ath G, 2002. Multivariate regression trees: a new technique for modeling species-environment relationships. *Ecology* 83, 1105-1117.
- Eastman J, Fulk M, 1993. Long sequence time-series evaluation using standardized principal components. *Photo Eng Rem Sens* 59, 991-996.
- Eisele TP, Keating J, Swalm C, Mbogo CM, Githeko AK, Regens JL, Githure JI, Andrews L, Beier JC, 2003. Linking field-based ecological data with remotely sensed data using a geographic information system in two malaria endemic urban areas of Kenya. *Malar J* 2, 44.
- Fabricante I, Oesterheld M, Paruelo JM, 2009. Annual and seasonal variation of NDVI explained by current and previous precipitation across Northern Patagonia. *J Arid Env* 73, 745-753.
- Fuller DO, Prince SD, 1996. Rainfall and foliar dynamics in tropical Southern Africa: potential impacts of global climatic change on savanna vegetation. *Clim Change* 33, 69-96.
- Galindo-Leal C, Gusmao-Camara I, 2003. Atlantic Forest

- Hotspot Status: an Overview. In: The Atlantic Forest of South America: biodiversity status, Threats, and Outlook (Galindo-Leal C, Gusmao-Camara I, Eds.), Atlantic Press, Washington DC, USA.
- Hay SI, Lennon JJ, 1999. Deriving meteorological variables across Africa for the study of vector-borne disease: a comparison of remote sensing and spatial interpolation of climate. *Trop Med Int Health* 4, 58-71.
- Hay SI, Snow RW, Rogers DJ, 1998. From predicting mosquito habitat to malaria seasons using remotely sensed data: practice, problems and perspectives. *Parasitol Today* 14, 306-313.
- Huang C, Kim S, Alstatt A, Townshend JRG, Davis P, Song K, Tucker CJ, Rodas O, Yanosky A, Clay R, Musinsky J, 2007. Rapid loss of Paraguay's Atlantic forest and the status of protected areas - a Landsat assessment. *Rem Sens Env* 106, 460-466.
- Kazmi SJH, Usery EL, 2001. Application of remote sensing and GIS for the monitoring of diseases: a unique research agenda for geographers. *Rem Sens Rev* 20, 45-70.
- Mushinzimana E, Munga S, Minakawa N, Li L, Feng C, Bian L, Kitron U, Schmidt C, Beck L, Zhou G, Githeko AK, Yan G, 2006. Landscape determinants and remote sensing of anopheline mosquito larval habitats in the western Kenya highlands. *Malaria J* 5, 13.
- Pinzon E, Wilson JM, Tucker CJ, 2005. Climate-based health monitoring systems for eco-climatic conditions associated with infectious diseases. *Bull Soc Pathol Exot* 98, 239-243.
- Pinzon JE, Wilson JM, Tucker CJ, Arthur R, Jahrling PB, Formenty P, 2004. Trigger events: enviroclimatic coupling of Ebola hemorrhagic fever outbreaks. *Am J Trop Med Hyg* 71, 664-674.
- Rojas de Arias A, 2002. Primeria comunicación sobre cambio climático en el Paraguay. Sector Salud, Secretaría del Ambiente/PNUD, Paraguay.
- Schultz PA, Halpert MS, 1993. Global correlation of temperature, NDVI and precipitation. *Adv Space Res* 13, 277-280.
- Tsonis AA, Swanson KL, Geli W, 2008. On the role of atmospheric teleconnections in climate. *J Clim* 12, 2990.
- Tucker CJ, 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Rem Sens Environ* 8, 127-150.
- Tucker CJ, Grant DM, Dykstra JD, 2004. NASA's global orthorectified Landsat data set. *Photogram Eng Rem Sens* 70, 312-322.
- Tucker CJ, Pinzon JE, Brown ME, Slayback D, Pak EW, Mahoney R, Vermote E, El Saleous N, 2005. An extended AVHRR 8-km NDVI data set compatible with MODIS and SPOT vegetation NDVI data. *Int J Rem Sen*, 4485-4498.
- Vittor AY, Gilman RH, Tielsch J, Glass G, Shields TIM., Lozano WS, Pinedo-Cancino V, Patz JA, 2006. The effect of deforestation on the human-biting rate of *Anopheles darlingi*, the primary vector of falciparum malaria in the Peruvian Amazon. *Am J Trop Med Hyg* 74, 3-11.
- Vittor AY, Pan W, Gilman RH, Tielsch J, Glass G, Shields T, Sanchez-Lozano W, Pinedo VV, Salas-Cobos E, Flores S, Patz JA, 2009. Linking deforestation to malaria in the Amazon: characterization of the breeding habitat of the principal malaria vector. *Am Trop Med Hyg* 81, 5-12.