Time-series analysis of Soil Moisture index: an examination of relationships between environmental conditions and Ebola emergence

Kevin T Gabelman

Masters of Science in Geographic Information Science (MSGISci)

Department of Geography, California State University, Long Beach

Introduction

Ebola virus transmission to humans, though uncommon, remains a mystery to researchers. Its geographic extent has traditionally been constrained to sub-saharan Africa, but its potential habitat, environmental or biological reservoirs, remain unconfirmed. It is important, therefore, to begin to understand the mechanisms that guide its behavior to generate a predictive framework which includes the types of conditions that contribute to its emergence, as well as how those might interplay or change over time. This study sought to examine a time series of remotely sensed soil moisture in localities that experienced an Ebola outbreak. It tested 14 cases wherein Ebola to human transmission location and timing are known. A wavelet transform analysis revealed significantly anomalous soil moisture preceding these Ebola outbreaks. Scale-wise anomalies manifested at multiple time-scales, 100-110 days prior to 11 of 14 sites, suggesting a causal link between dynamics governing soil moisture, vegetative behavior, intermediate animal hosts, and viral populations. This study provides evidence for the value of remote sensing and wavelet analysis in discovering temporal correlates of disease transmission.

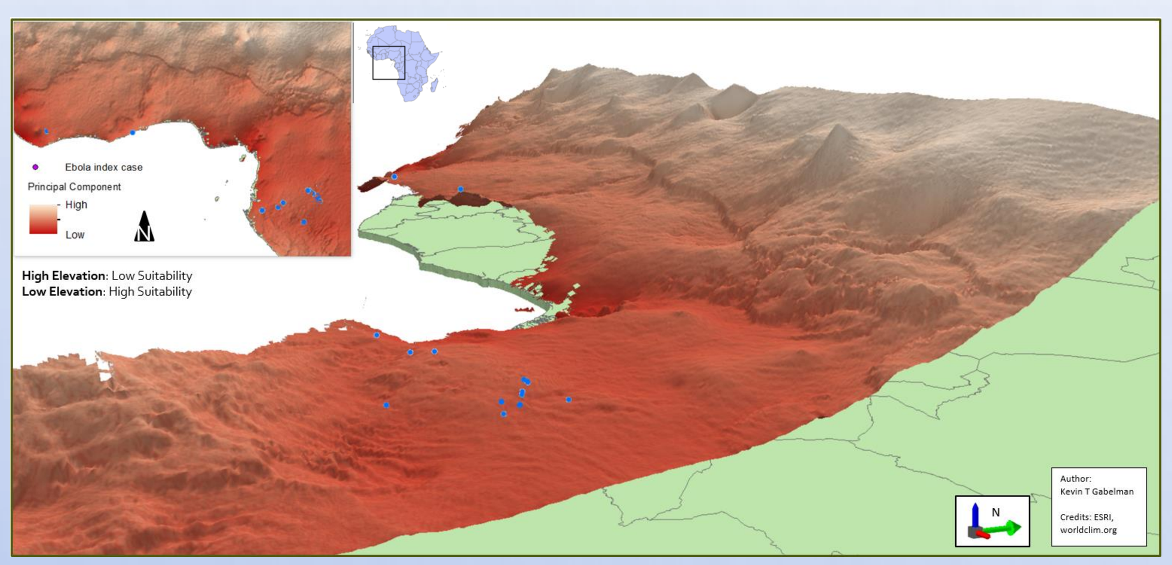

Figure 3. *Ebola outbreak occurrences alongside habitat suitability within equatorial Africa*

Data and Data Sources

Ebola occurrences taken from Mylne et al. (2014) and georeferenced

- ASCAT-METop satellite and sensor suite
- 1/10° resolution soil moisture index (SWI) raster datasets
- Originally developed to measure wind speed, satellite scatterometers actively transmit electromagnetic pulses to measure backscatter to their antennae.

Data (Table 1) were organized and stored in solid state drives to facilitate data decompression and speedy analysis.

Table 1. *List of data and data sources used in the project*

Dataset	Source
SWI Daily Raster Datasets	Copernicus FTP access
Ebola index case occurrences	Mylne et al. 2014

Methodology

The methods required include multi-phased data acquisition, preparation, dual-staged analysis, and data visualization. Soil water index data were acquired via FTP download from the European Copernicus portal. These arrived as compressed daily global coverage WGS 1984 HDF arrays of raster datasets. Ebola index case site locations were acquired from Mylne et al (2014).

The first of two analysis phases statistically analyzed the raster datasets using the Math function in ArcPy - a Python module for Esri ArcGIS mapping software. After computing wavelet coefficients, at each time position, from the signal data, I created a time-series graph, which displays wavelet coefficient at each time position. Then I computed the inverse transform of the anomalous scale coefficients. This generated a discrete wavelet transform.

Database Aquisition/Organization → Raster Algebra Analysis → Wavelet Analysis → Data Dissemination

Figure 2. *Shows the basic framework for this study*

Methods Visualization

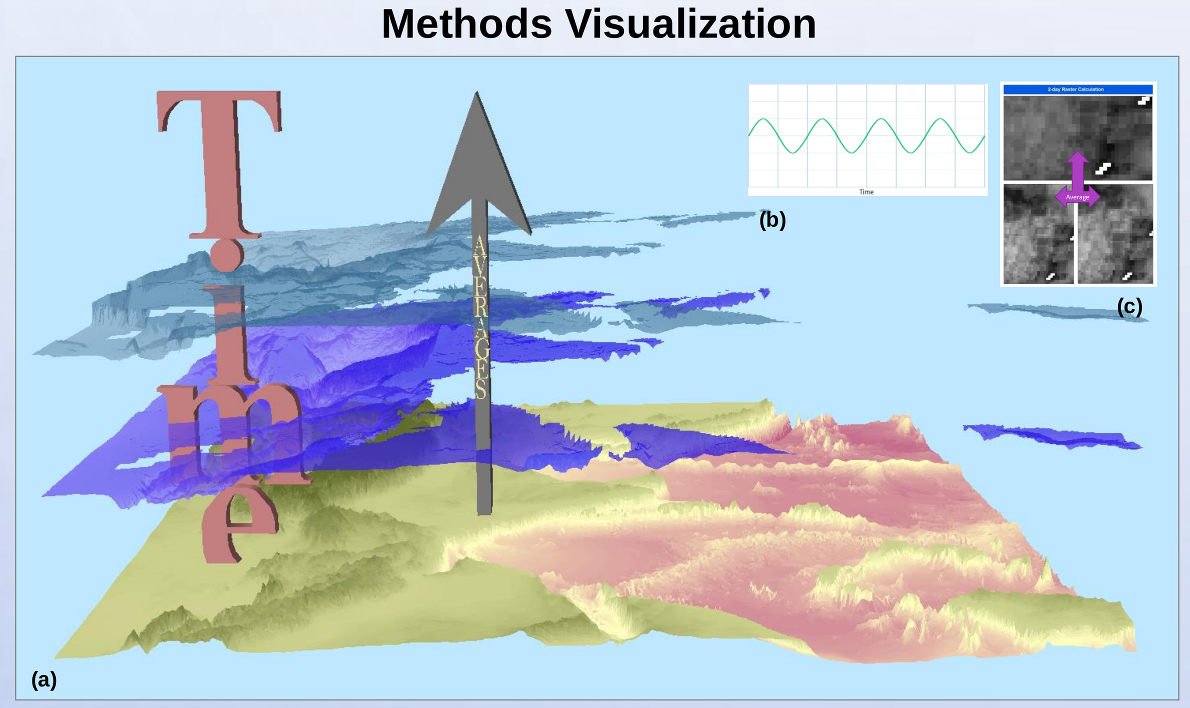

Figure 3. *A soil moisture layer stack (a), produced from daily SWI readings, results in a time series of means (b), which can be used in a wavelet analysis. Previous literature relied on 10-day composite layers, often represented as 10-day averages or maximum pixel values (c).*

Figure 4. *Ebola Outbreak Locations. Circular outlines show sub-set climate outbreak regions. The gradient shading shows regional climatic variation.*

Ebola tends to occur in tropical regions, but rapid human transit and intermediate animal host migration can spread the disease elsewhere.

This study looked at how climate (soil) varied, leading up to an outbreak, within these zones.

Timeline

Database Organization	Apply for Funding	Projection/ Extraction	Finding Soil Moisture Spatial Means	Data organization	Wavelet Analysis	Create Website
using Arcpy, cmd, 7z January - February 2016	January - April 2016	Using Arcpy March - May 2016	Use Raster Math in Arcpy and zonal statistics May - June 2016	Excel June 2016	• MATLAB Wavelet Analysis • subset in February and May 2016 • final analysis July 2016	July - August 2016

Results

The decomposed second or third order soil moisture signals (Figure 5, Figure 6) show the strongest signal occurs at approximately 540 days, or, 60 days before a known Ebola virus index case.

Anomalous soil moisture manifests between 520 and 560 days. This signal remains pronounced in the second and third reconstructed signals.

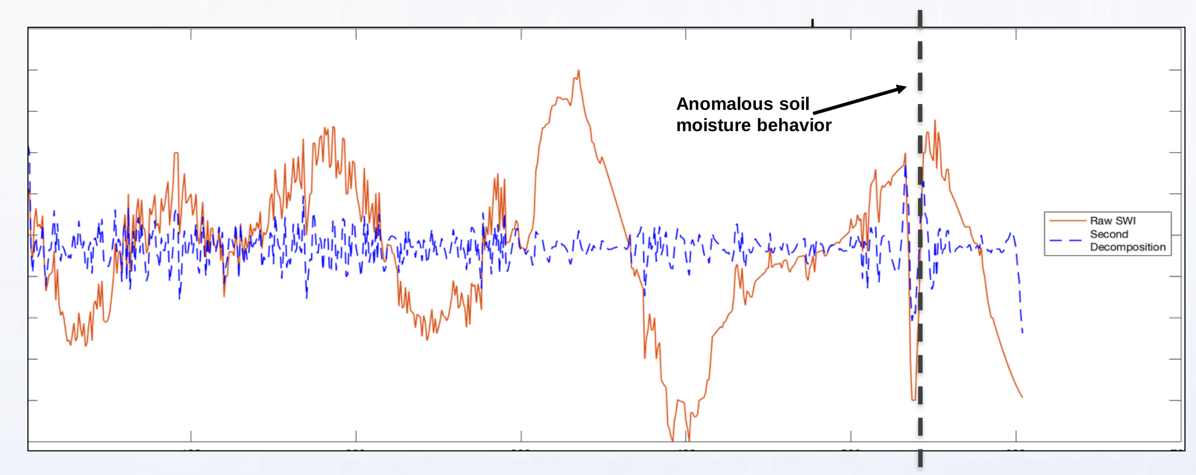

Figure 5. *shows(a) raw SWI signal for 600 days preceding an Ebola outbreak and (b) the reconstructed signal following a third-order wavelet decomposition.*

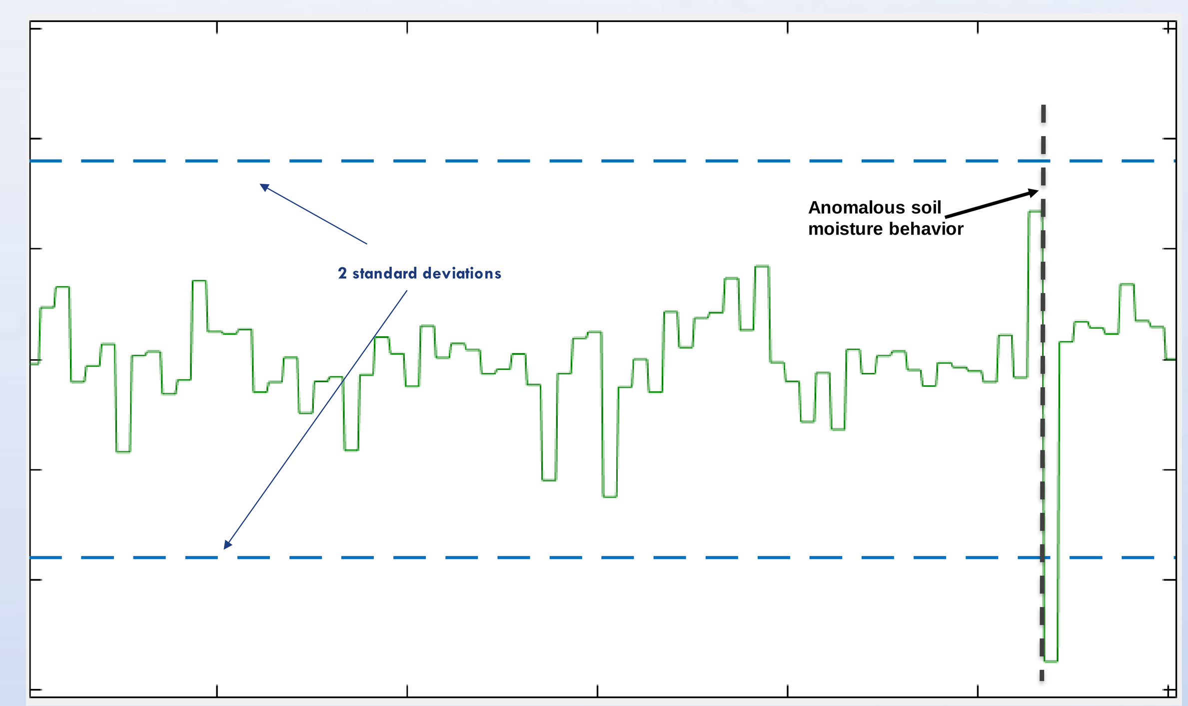

Figure 6. *This shows the wavelet coefficients for the decomposition time-series shown in the previous graph. Wavelet coefficient values indicate statistical significance above 2 standard deviations from their mean.*

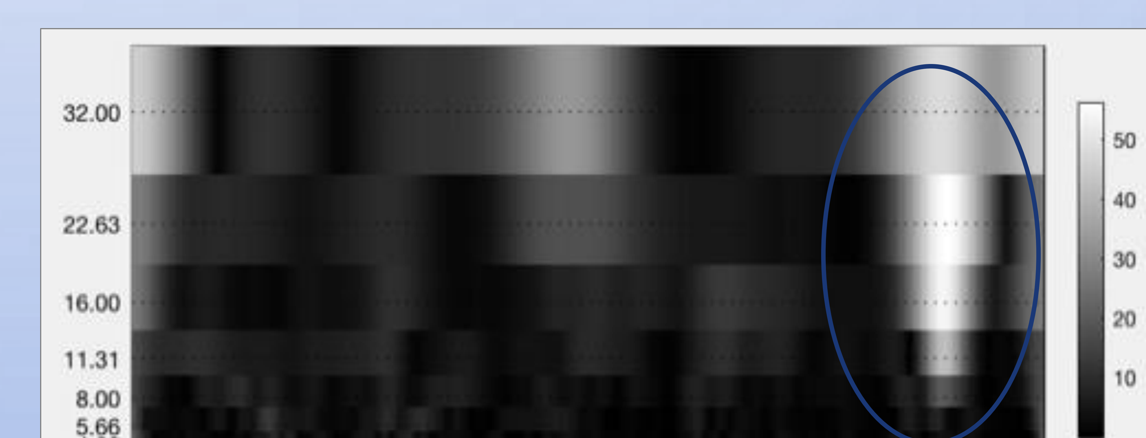

Figure 7. *This displays wavelet coefficient power (grayscale gradient) variation as absolute time (horizontal axis) and time-scale (vertical axis) vary. In the second decomposition, there is a visible signal starting at the 8-day time-scale, which increases in magnitude to a scale of 22 days. We expect to see longer time-scale intervals display larger coefficient value areas, which is particularly visible at 32 days.*

Additional levels of signal reconstruction result in depressed signal strength. The best signal, therefore, occurs in the third reconstruction, between time scale 8 and 22.

These data together show anomalous and statistically significant soil moisture preceding an Ebola virus outbreak. It leads the same signal from vegetative data in Lash et al. (2008).

Discussion

Ebola virus represents a uniquely extreme case of high host mortality. That it kills nearly all of its known mammalian hosts on a population scale suggests an uncanny ability to mutate or hibernate in alternate hosts, or in the environment. This paper sought to better understand if rates of change or variability in soil moisture, one measureable environmental indicator, correspond with Ebola transmitting to human hosts.

Though the specific format of the analysis performed as part of this study is not predictive, the mathematical methods and software techniques employed could be repurposed to provide advance warning of Ebola virus entering human populations. The soil moisture wavelet analysis examines almost a decade of environmental data. Ebola outbreak events conform to a singularly unique soil moisture time-series decomposition. This reveals a "temporal niche" that precedes the virus changing from purely zoological to human infection behavior.

The time-series wavelet analysis performed on the soil moisture in regions surrounding Ebola outbreaks represents an iterative step forward in quantitative complexity of spatially explicit epidemiological study. Spatial analysis is becoming a powerful tool in epidemiological research. It is able to utilize epidemiological assumptions and incorporate types of data previously unsuited to more traditional epidemiological models. The statistical and descriptive power of spatial modelling and the increasing availability and diversity of remotely sensed data compel exploration into disease dynamics that could not be otherwise studied.

Conclusion

This study sought to test whether soil moisture changes may underlie Ebola transmission to humans. This is important because scientists do not yet understand how Ebola mutates and spreads among human populations. Understanding is the beginning of being able to predict these tragic events.

A study of wider scope and greater statistical power would generate a chi-square significance test of all "suitable" regions. Then, in addition to summarizing remote sensing data from known Ebola index occurrence localities, researchers could randomly sample from within environmentally suitable regions at a set density. The researcher could run a similar analysis of each region and perform a chi-square test which examines absence as well as presence of an Ebola outbreak within a particular time scale to provide a control sample.

Submitted in partial fulfillment of the requirements of the Masters of Science in Geographic Information Science(MSGISci), August 13, 2016.

For additional information please contact: Kevin Gabelman
kevintgabelman@gmail.com