**US Army Corps
of Engineers**®
Engineer Research and
Development Center
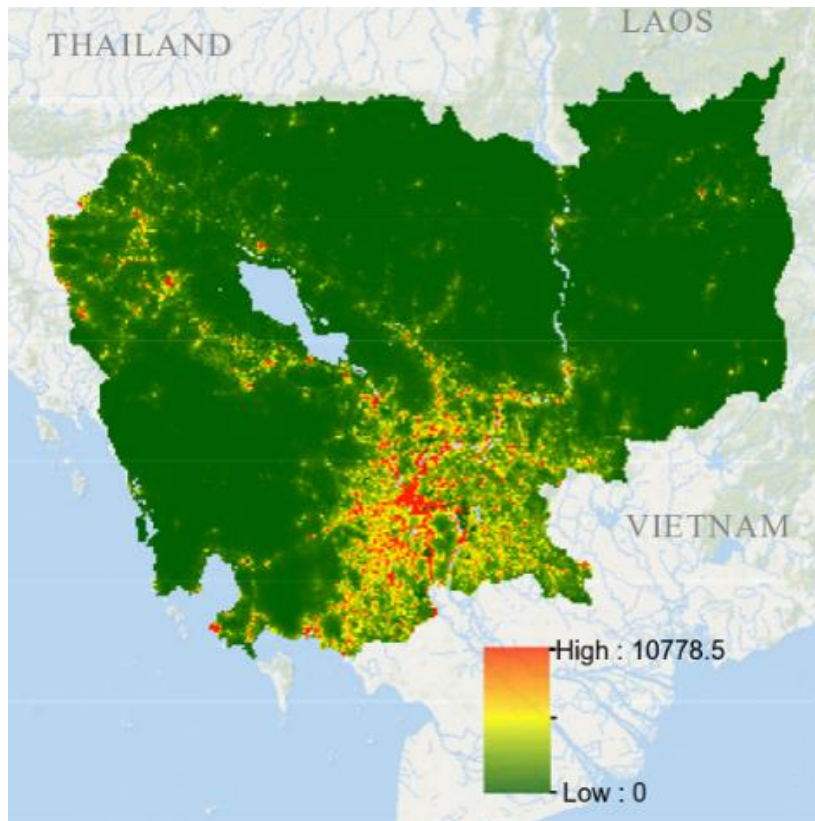
ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

Geospatial Research Laboratory

ERDC Geospatial Research and Engineering (GRE) 6.2 Portfolio

New and Enhanced Tools for Civil Military Operations (NET-CMO)

Nicole M. Wayant, Sarah J. Becker, Joshua Parker, S. Bruce Blundell, Susan Lyon, Megan Maloney, Robin Lopez, Sean Griffin, and John A. Nedza

February 2020

Force Health Risk Assessment

The U.S. Army Engineer Research and Development Center (ERDC) solves the nation's toughest engineering and environmental challenges. ERDC develops innovative solutions in civil and military engineering, geospatial sciences, water resources, and environmental sciences for the Army, the Department of Defense, civilian agencies, and our nation's public good. Find out more at www.erdc.usace.army.mil.

To search for other technical reports published by ERDC, visit the ERDC online library at http://acwc.sdp.sirsi.net/client/default.

New and Enhanced Tools for Civil Military Operations (NET-CMO)

Nicole M. Wayant, Sarah J. Becker, Joshua Parker,
S. Bruce Blundell, Susan Lyon, Megan Maloney, Robin Lopez,
Sean Griffin and John A. Nedza

Geospatial Research Laboratory (GRL)
U.S. Army Engineer Research and Development Center
7701 Telegraph Road
Alexandria, VA 22315

Final report

Approved for public release; distribution is unlimited.

Abstract

Civil Military Operations (CMO) associated geospatial modeling is intended to enable increased knowledge of regional stability, assist in Foreign Humanitarian Assistance (FHA), and provide support to Force Health Protection (FHP) operational planning tasks. However, current geoenabled methodologies and technologies are lacking in their overall capacity to support complex mission analysis efforts focused on understanding these important stability factors and mitigating threats to Army soldiers and civilian populations. CMO analysts, planners, and decision-makers do not have a robust capability to both spatially and quantitatively identify Regions of Interest (ROI), which may experience a proliferation in health risks such as vector-borne diseases in areas of future conflict. Additionally, due to this general absence of geoenabled health assessment models and derived end-products, CMO stakeholders are adversely impacted in their Military Decision Making Process (MDMP) capabilities to develop comprehensive area studies and plans such as Course of Action (COA). The NET-CMO project is focused on fostering emerging geoenabling capabilities and technologies to improve military situational awareness for assessment and planning of potential health threat-risk vulnerabilities.

Contents

Figures and Tables

Figures

Tables

Preface

This study was conducted for the Engineering Research and Development Center (ERDC) under Project 855, "New and Enhanced Tools for Civil Military Operations (NET-CMO)."

The work was performed by multiple branches of the Geospatial Research Division, U.S. Army Engineer Research and Development Center, Geospatial Research Laboratory (ERDC-GRL). At the time of publication, Mr. Jeffrey Murphy was Chief of the project lead's branch; Ms. Martha Kiene was Chief; and Mr. Ritch Rodebaugh was the Technical Director for the Geospatial Research and Engineering (GRE) business portfolio. The Deputy Director of ERDC-GRL was Ms. Valerie Carney and the Director was Mr. Gary Blohm.

COL Teresa A. Schlosser was the Commander of ERDC, and Dr. David W. Pittman was the Director.

1 Introduction

1.1 Background

The New and Enhanced Tools for Civil Military Operations (NET-CMO) FY19 6.2/6.3 research project primarily addresses three challenge problem areas. These problems affect Civil Military Operation (CMO) ability to ensure long-term regional stability, assist in foreign humanitarian assistance (FHA), increase public health, or provide situational awareness. Solutions to these problems will not only increase stability and mitigate threats to the civilian population, but will also ensure force readiness and health for the Army at large.

1.2 Army challenge problems

1.2.1 Problem 1 - Uniform pixel size for spatial disparate datasets

Commonly used geoenabled products, such as disease hazard maps and prediction of conflict prone regions, require raster data with varying spatial resolutions. These models require their raster inputs to have a uniform pixel size, which often requires resampling. Currently, there is not guidance on how to choose a uniform pixel size when working with spatially disparate data. In previous research, the common approach has been to resample to the smallest or largest pixel size of the dataset without taking into consideration the amount of error introduced, the processing time needed for analysis, or the preservation of spatial patterns, all of which can have an impact on a model's overall results. Thus, a statistically valid way of selecting a uniform pixel size for spatially disparate raster data needs to be developed.

1.2.2 Problem 2 - Resolution of data in denied regions

Many regions of the world are data poor in terms of environmental, disease, demographic, and social data. Relevant data sources that may be available typically have coarse spatial and temporal resolution – with key predictive incidence datasets potentially only available annually at either a country or provincial level. Information with such a low fidelity provides for inhibited situational awareness of a region of interest (ROI) and makes it difficult for CMO planners to develop course of action (COA) based end-products. Also, new techniques, scaling methods, and geoprocessing

algorithms are needed to conflate and merge ancillary geospatial data sets across these spectrums of scale and fidelity.

1.2.3   Problem 3 - Mosquito-borne disease prediction

Over half of the world's population lives in a region where health risk from mosquito-borne diseases are endemic. With changing temperatures and increased extreme weather events, the global reach of these diseases is expanding. Current CMO practices do not possess the capability to predict when and where these mosquito-borne diseases will occur or the portion of the population that could become infected; therefore, they do not have the best situational awareness as to which preventative techniques to use, whom to vaccinate, the amount of medical supplies needed, or areas where civilians should be removed. Additionally, without understanding how a disease will spread, Soldiers could potentially be unable to protect themselves from these predictable health risks directly impacting mission readiness and effectiveness. As seen in the West African Ebola epidemic that occurred between 2013-2016, infectious disease can severely disrupt not only the public health of a region, but also its social and economic stability. With the increasing spatial distribution of mosquito-borne diseases and the growing resistance of these diseases to medicines, it is imperative that CMO work efforts focus on a better understanding of when and where these diseases will occur, spread, and their severity. Project outcomes will serve to safeguard a region's public health, help stabilize populations, and protect Soldier's well-being.

1.3   Approach

The quantitative tools and methodologies developed within the NET-CMO project will serve to enhance current and future mission analysis for CMO. The project was successful due to building upon scientific research across an interdisciplinary range of academic disciplines and employing a sound scientific framework. The tools and methodologies developed through research and development (R&D) work efforts were well-researched, constructed, documented, tested, and more importantly subjected to rigorous statistical validation to take into account the uncertainty that exists with any type of analysis. Additionally, team members working on the project had the requisite academic backgrounds and diverse professional expertise (physics, mathematics, human geography, environmental science, etc.) required to tackle the stated Army challenge problems.

The leveraging of outcomes from the prior ERDC 6.2 project Vulnerability Assessment Software Toolkit (VAST) drove the distinctive tasks of this project. The applied statistical approaches and planned quantitative tools and methodologies as mapped to problem areas are summarized below:

1. Uniform pixel size for spatial disparate datasets

   a. Test and validate previously developed VAST tools to find an optimal uniform pixel size when working with remotely sensed data with disparate spatial resolution.

      (1) Tool 1: Multiple Semivariograms
      (2) Tool 2: Local Spatial Dispersion

2. Resolution of data in denied regions

   a. Develop methodology to down sampling annual records of vector-borne disease to a monthly time scale.
   b. Develop a model that can downscale and optimize provincial or country level vector-borne disease statistics to a 1 km or smaller pixel size.

3. Mosquito-borne disease prediction

   a. Complete VAST developed mosquito-borne disease simulation.

1.4   Objective

The intended objectives of the NET-CMO project and goal of the R&D work efforts were to deliver Stakeholders the following geospatial processing mechanisms to enhance interpretation of Operational Environment (OE) and mission analysis capabilities.

- Tool(s) capable of finding a uniform pixel size when working with data with varying spatial resolutions.
- The capability to downscale coarse temporal and spatial resolution vector-borne disease data to provide a more defined situational awareness of a region.
- Mapping and health risk analysis of the spatial and temporal spread of any mosquito-borne disease and the proportion of the population that will become infected.

2 Uniform Pixel Size

2.1 Literature review

Terrestrial features in remotely sensed imagery or geospatial data have inherent and quantifiable spatial variability and heterogeneity. The spatial resolution of a remotely sensed image represents the scale of sensor observations on the land surface (i.e. the pixel size). Other types of spatially sampled environmental data (e.g. precipitation) can be represented in gridded or raster form. The selection of an appropriate scale depends on the type of information desired as well as the size and variability of the land phenomena under examination. In modeling processes on the Earth's surface, their spatial resolution must be considered. If the process is affected by detail at a finer scale than that provided by the data, the model's output will be misleading (Goodchild 2011).

The relationship between the size of objects or features in an image and spatial resolution helps determine the spatial structure of the image. Fine resolution, relative to scene object size, results in high correlation of neighboring pixels, reducing the local spatial variance. Large pixel size, relative to scene objects, results in a mixing of response from different kinds of objects, also depressing local variance. The pixel size that results in a maximum variance would then best capture the spatial variation in the image (Rahman et al. 2003; McCloy and Bøcher 2007). As seen in this study, this general principle may not hold for images with heterogeneous spatial structure having a broad range of spatial frequency of variation for image objects.

Two common approaches used when resampling to a common spatial resolution are upscaling and downscaling. Upscaling refers to resampling all the data to a larger pixel size, while downscaling refers to resampling a raster to a smaller pixel size. The challenges associated with each resampling approach are listed below:

1. Resampling raster data from a small pixel size to a larger pixel size may cause useful information to be lost.
2. Downscaling to an arbitrary pixel size introduces user bias and false precision. Pixels that are significantly smaller than the target objects they represent contain redundant information (Wulder and Boots 2001). If redundant pixels are overly present, minor background pixels

could become overly represented and skew geospatial models (Rodriguez-Carrion et al. 2014; Costanza and Maxwell 1994). Thus, it is preferable to size pixels at the point where only relevant information is preserved (Fisher et al. 2017).

Little research has been carried out on how to effectively determine an ideal pixel size to resample to when working with disparate rasters, especially those that may be internally heterogeneous, such as land cover or population density. However, multiple techniques have been assessed by previous studies to identify a pixel size for single raster datasets that best represent the spatial information of objects present within a raster scene. The techniques include local variance (McCloy and Bøcher 2007; Woodcock and Strahler 1987; Rahman et al. 2003; Rodriguez-Carrion et al. 2014; Sharma et al. 2011; Hyppänen 1996), fractal analysis of digital elevation models (Sharma et al. 2011; Lam and Quattrochi 1992), and analysis of semivariograms (Atkinson and Curran 1997; Wu et al. 2006; Rahman et al. 2003; Hyppänen 1996; Cohen et al. 1990).

Of these methods, two were explored in NET-CMO: semivariograms and local spatial variance. A semivariogram is a statistical method that efficiently characterizes the structure of spatial patterns (de Oliveira Silveira et al. 2017) and spatial continuity (de Lima Guedes et al. 2015) of a raster image. One of the most attractive qualities of semivariograms is their ability to render the spatial autocorrelation that occurs when evaluating rasters. Spatial autocorrelation, the tendency of phenomena to be similar to nearer points than farther points, is present in any dataset that describes a spatially dependent phenomenon, but can lower overall model accuracy in models that assume spatial independence (Dormann et al. 2007; Kühn and Dormann 2012; Legendre 1993). The semivariogram is a measure of spatial dependence between two observations as a function of distance between them and provides a graphical representation of spatial autocorrelation when working with spatial data. The semivariogram graph shows the distance at which pixels in a raster are no longer spatially autocorrelated, which is known as the 'range' of the semivariogram. Pixels before the range are similar to each other and therefore contain redundant information. An ideal pixel size for a single image is based on this location in the semivariogram. This ideal pixel size for a single image balances the maximization of variance between neighboring pixels and the ability to maintain spatial patterns throughout the dataset.

The second method used to identify an ideal uniform pixel size across spatially disparate rasters was local spatial dispersion (LSD). Rahman et

al. (2003) assessed image spatial structure of similar vegetation by analyzing the mean local variance of pixel values at varied spatial resolutions. The authors found that a maximum value for this function may be related to an optimum pixel size for the segmentation of a particular land surface process or feature type. Two competing concerns are involved: finding a balance between reducing the correlation among neighboring pixels having sizes smaller than the spatial structure, and reducing effects of different spatial objects intermixed within a given pixel (pixel mixing). The balance between these concerns is obtained by finding the sample size associated with the maximum mean local variance of a feature when plotted against pixel size (Woodcock and Strahler 1987). This size will be tuned to the particular spatial structure of scene elements that make up the feature(s) under investigation.

2.2 Geoanalytic data and methods

2.2.1 Semivariograms

2.2.1.1 Data

The proposed semivariogram method is applied to tree cover, population, precipitation, and wind raster data available over Cambodia (Table 1). Cambodia was selected because its diverse spatial patterns in land cover, including multiple types of forest cover and woodlands, deforested land, plains, agriculture, grasslands, wetlands, and urban areas.

All datasets were retrieved and processed in Google Earth Engine (GEE). The original application of the data was to provide monthly representations for each dataset. Landsat-derived tree cover and population data from WorldPop[*] are available in GEE as annual products and assumed static for each month of 2010. Wind was available as an averaged monthly product and required no additional processing. Precipitation from the Climate Hazards group Infrared Precipitation with Stations was the only dataset that required temporal reduction by summing the six collections for the month of June 2010 to provide the total monthly precipitation in millimeters (Table 1).

---

[*] WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton) (2013) Cambodia 100m Population. Alpha version 2010, 2015 and 2010 estimates of numbers of people per pixel (ppp) and people per hectare (pph), with national totals adjusted to match UN population division estimates (http://esa.un.org/wpp/) and remaining unadjusted. DOI: 10.5258/SOTON/WP00040.

Table 1. Overview of all data types from Cambodia used in the study.

Dataset	Source	Temporal Resolution	Date	Units	Resolution	Source
Tree cover	Landsat-5 and Landsat-7	Annual	2010	% Tree cover (0 - 100)	30 m	Sexton et al. 2013
Population	Worldpop	Annual	2010	Estimated number of people per hectare	1,000 m	Worldpop 2013
Precipitation	Tropical Rainfall Measuring Mission and gauge data	Five days	2010-June	Millimeters (mm)	5,000 m	Funk et al. 2015
Wind	Global Land Data Assimilation System (GLDAS)	Monthly	2010-June	Wind speed in meters/second	25,000 m	GLDAS 2010

2.2.1.2 Methods

The output of a semivariogram is a graph that plots semivariance against grouped distances between pixels in a raster. Nearby pixels often exhibit lower semivariance than pixels that are farther away. This is because nearby pixels are more likely to contain similar pixel values. Pixels farther away from each other are likely to contain different features, and thus show high semivariance.

An example of a semivariogram is shown in Figure 1. The x-axis represents the distance between pixels and the y-axis shows the amount of semivariance between pixels based on their physical distance from each other. The range, represented by a vertical dashed line, is the location in the semivariogram where pixels are no longer spatially autocorrelated. Pixels are spatially correlated before the range is reached and where the sill begins, represented by a horizontal hashed line. Hyppänen (1996) and Cohen et al. (1990) showed the range and sill relate to actual features on the ground.

Figure 1. A sample semivariogram showing the location of the sill and range (ESRI 2019a). Pixels are less correlated as distance between pixels increases.
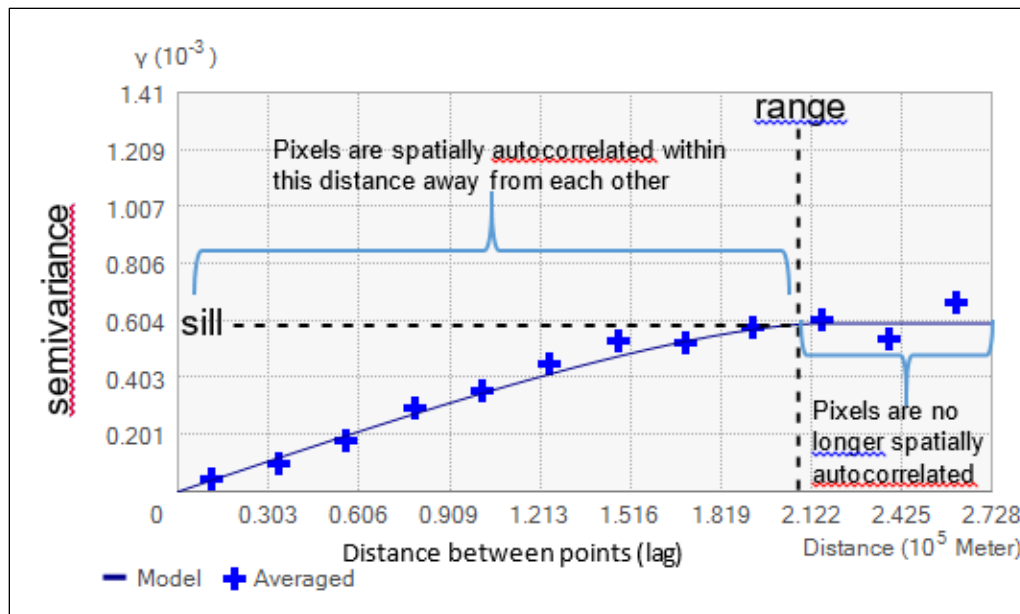
Previous research indicates the range as the ideal pixel size in a single raster type dataset (Hyppänen 1996; Curran 1988; Löw and Duveiller 2014). However, according to framework based on Nyquist's sampling theorem (Nyquist 1928), half the semivariogram range is more appropriate. Nyquist's sampling theorem is based on his framework for converting one-dimensional telegraph analog data to digital format, and it states that the sampling rate must be equal to one-half the signal bandwidth. If the semivariance between pixels at varying ranges is thought of as a signal, then one-half the range can be thought of as the signal's sampling rate, and therefore, the ideal pixel size for a single raster. This theory is supported by more recent research, which demonstrated that the distance above half the semivariogram range indicates the size at which spatial elements are not related (Rahman et al. 2003; Modis and Papaodysseus 2006).

The workflow to extend the semivariogram to multiple rasters is listed below. The workflow was built in ArcMap and ModelBuilder in ArcGIS[*] (Environmental Systems Research Institute, Redlands, CA).

---

[*] ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license. Copyright © Esri. All rights reserved. For more information about Esri® software, please visit www.esri.com."

1. Each raster used in the analysis was projected into a common projection.
2. The semivariogram and its range for each raster were calculated using the ArcGIS Geostatistical Wizard. The lag distance of the semivariogram was manually selected to be smaller than most visible objects in the raster. No accepted method exists to determine lag in continuous raster data.
3. Because anisotropy (where the data show higher spatial autocorrelation in one direction than another) was present in the data, ArcGIS Geostatistical Wizard calculated the major and minor ranges. Half the minor range for the ideal pixel size for each raster is used as its optimal pixel size. The minor range was selected because it was the direction that contained the most variance in the raster.
4. All of the raster images were upscaled or downscaled to all of the calculated ideal pixel sizes using a bilinear interpolation.
5. Nearest neighbor interpolation was then used to resample and match the pixel position and resolution from the original projected raster for the purpose of calculating raster error introduced from resampling.

The Mean Absolute Error (MAE) was calculated between the original and resampled raster data at each pixel size. The MAE values from this calculation were averaged at the resampled pixel size, which quantified the error caused by resampling. The average MAE values at each ideal pixel size were ranked from lowest to highest; the ideal pixel size with the lowest average MAE was selected as the best uniform pixel size.

2.2.2   Local spatial dispersion

An algorithmic approach was developed that addressed the creation of a spatial data model and a set of methods for performing internal calculations to arrive at an optimal sample size. These methods were required to compute and optimize LSD within the model before calculating the optimal sample size based on the discovered set of local maxima. A graphical user interface was then created in the MATLAB environment to perform the calculations and display the results of LSD optimization.

2.2.2.1 Data

The first step in the process is to create the spatial data model by populating it with resampled versions of the original image or spatial dataset with progressively lower spatial resolution. To do this, a

resampling method must be chosen to create these pixel aggregation levels calculated by a neighborhood function. Three resampling methods were allowed: block processing of the mean of each set of neighborhood values, bilinear interpolation, and bicubic interpolation.

The next step is to compute the local dispersion at each cell location for each resampled image in the spatial data model. Common techniques for this purpose include local spatial variance (LSV) and Mean Absolute Deviation (MAD). The result is a set of LSD "images," each one having the spatial resolution of the resampled image from which it was created. However, in order to proceed further with the matrix algebra necessary to find the set of LSD maxima in this multidimensional space, the spatial data model must have uniform granularity along all three orthogonal directions *(r,c,s)*. This is required to have a uniform distribution of LSD values. Due to the nature of the pixel aggregation process, this granularity decreases in the sample size *(s)* direction. Therefore, an interpolation scheme must be applied to each LSD image to achieve the same spatial resolution as the original image. The nearest neighbor interpolation will be employed for this purpose. This results in a uniform distribution of LSD values throughout LSD space.

2.2.2.2 Methods

- Hessian Matrix Optimization:

    r = row
    c = column
    s = sample size

$$H_{rcs} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} = \begin{bmatrix} \partial^2 f/\partial r^2 & \partial^2 f/\partial r \partial c & \partial^2 f/\partial r \partial s \\ \partial^2 f/\partial c \partial r & \partial^2 f/\partial c^2 & \partial^2 f/\partial c \partial s \\ \partial^2 f/\partial s \partial r & \partial^2 f/\partial s \partial c & \partial^2 f/\partial s^2 \end{bmatrix}$$

In order to find the Hessian for each **x** vector, each element must be evaluated numerically. A finite divided-difference approximation method will be used for this purpose. The values of **x** in the row, column, and sample dimensions will be perturbed by some small fractional value δ to

generate the partial derivatives. δ cannot be too small or too large. Too small a value may not provide enough variation in the variable to capture the functional trend at that location, too large may cause excess inaccuracy in the estimate for the derivative. Nominally, each δ increment in LSD space can be taken as an adjoining raster grid cell (one pixel) along one of the orthogonal axes *r,c,s*.

In employing the divided-difference method to approximate the partial derivatives, one can normally choose from equations for a "forward," "centered," or "backward" sampling scheme for the δ increment. Since the centered difference equations are considered a more accurate representation of the derivative, this approach will be used to estimate the Hessian matrix elements. This requires adding and subtracting δ for each independent variable in the approximation equations, maintaining a consistent approach. However, because outside image boundaries cannot be sampled, Hessian elements for pixels within a distance δ of the *r,c* edges for each image will not be estimated. Normally, this limitation would also apply along the *s* axis as well. However, because any higher resolution images with sample sizes between *s = 1* and *s = δ* may contain a large amount of LSD local maxima information, these images will be retained by substituting delta increments that yield LSD samples in the positive *s* direction.

The result of these divided-difference calculations is an estimated Hessian matrix for each location in LSD space. The centered approximation equations for the 9 Hessian elements h_{ij} *(i=1,2,3; j=1,2,3)* are provided below. If assumed that the partials are continuous in the region surrounding each location **x** in LSD space, the mixed partials will be equivalent, e.g. $\partial^2 f/\partial r \partial c = \partial^2 f/\partial c \partial r$.

Centered Divided Difference:

$h_{11} = \partial^2 f/\partial r^2 = [f(r+\delta r,c,s) - 2f(r,c,s) + f(r-\delta r,c,s)] / (\delta r)^2$

$h_{22} = \partial^2 f/\partial c^2 = [f(r,c+\delta c,s) - 2f(r,c,s) + f(r,c-\delta c,s)] / (\delta c)^2$

$h_{33} = \partial^2 f/\partial s^2 = [f(r,c,s+\delta s) - 2f(r,c,s) + f(r,c,s-\delta s)] / (\delta s)^2$

$h_{21} = \partial^2 f/\partial r \partial c = \partial^2 f/\partial c \partial r = [f(r+\delta r,c+\delta c,s) - f(r+\delta r,c-\delta c,s) - f(r-\delta r,c+\delta c,s) + f(r-\delta r,c-\delta c,s)] / 4\delta r \delta c$

$h_{31} = \partial^2f/\partial r\partial s = \partial^2f/\partial s\partial r = [f(r+\delta r,c,s+\delta s) - f(r+\delta r,c,s-\delta s) - f(r-\delta r,c,s+\delta s) + f(r-\delta r,c,s-\delta s)] / 4\delta r\delta s$

$h_{32} = \partial^2f/\partial c\partial s = \partial^2f/\partial s\partial c = [f(r,c+\delta c,s+\delta s) - f(r,c+\delta c,s-\delta s) - f(r,c-\delta c,s+\delta s) + f(r,c-\delta c,s-\delta s)] / 4\delta c\delta s$

where

$h_{12} = h_{21}$; $h_{13} = h_{31}$; and $h_{23} = h_{32}$.

The next step in the process is testing each Hessian for the property of negative definiteness. Every location **x** in LSD space for which $H(\boldsymbol{x})$ is negative definite will define a local maximum for $f(r,c,s)$. To perform this test the determinants of three subset matrices H_1, H_2, H_3 of the Hessian must be found, starting from the upper left position (h_{11}). These are:

$H_1 = \quad h_{11} \quad$ (a 1x1 matrix)

$\det(H_1) = h_{11} = \partial^2f/\partial r^2$

$H_2 = \quad h_{11} \quad h_{12} \quad$ (a 2x2 matrix)

$\qquad h_{21} \quad h_{22}$

$\det(H_2) = h_{11} h_{22} - h_{12} h_{21}$

$\qquad = \partial^2f/\partial r^2 \, \partial^2f/\partial c^2 - \partial^2f/\partial r\partial c \, \partial^2f/\partial c\partial r$

Under the assumption that the partials are continuous in the region surrounding location **x** in LSD space,

$\det(H_2) = \partial^2f/\partial r^2 \, \partial^2f/\partial c^2 - (\partial^2f/\partial r\partial c)^2$

$H_3 = H$ (the full 3x3 matrix)

$\det(H_3) = h_{11} h_{22} h_{33} - h_{11} h_{23} h_{32} - h_{12} h_{21} h_{33} + h_{12} h_{23} h_{31} + h_{13} h_{21} h_{32} - h_{13} h_{22} h_{31}$

$\det(H_3) = \partial^2f/\partial r^2 \, \partial^2f/\partial c^2 \, \partial^2f/\partial s^2 - \partial^2f/\partial r^2 \, \partial^2f/\partial c\partial s \, \partial^2f/\partial s\partial c - \partial^2f/\partial r\partial c \, \partial^2f/\partial c\partial r \, \partial^2f/\partial s^2 + \partial^2f/\partial r\partial c \, \partial^2f/\partial c\partial s \, \partial^2f/\partial s\partial r + \partial^2f/\partial r\partial s \, \partial^2f/\partial c\partial r \, \partial^2f/\partial s\partial c - \partial^2f/\partial r\partial s \, \partial^2f/\partial c^2 \, \partial^2f/\partial s\partial r$

Again, assuming that the partials are continuous in the local region,

$$\det(H_3) = \partial^2 f/\partial r^2\, \partial^2 f/\partial c^2\, \partial^2 f/\partial s^2 - \partial^2 f/\partial r^2\, (\partial^2 f/\partial c\partial s)^2 - \partial^2 f/\partial s^2$$
$$(\partial^2 f/\partial r\partial c)^2 + 2(\partial^2 f/\partial r\partial c\, \partial^2 f/\partial c\partial s\, \partial^2 f/\partial r\partial s) - \partial^2 f/\partial c^2\, (\partial^2 f/\partial r\partial s)^2$$

The following conditions are necessary and sufficient for *H(**x**)* to be negative definite:

$$\det(H_1) < 0$$

$$\det(H_2) > 0$$

$$\det(H_3) < 0$$

This test is applied to every location vector **x** in LSD space, ultimately transforming LSD space into a "local maximum" space. **x** is a local maximum of *f(r,c,s)* wherever *H(**x**)* is negative definite. The output from these operations is, in theory, the set of optimal sample sizes associated with the subset of **x** vectors defined by the negative definiteness property of *H(**x**)* across the image or spatial dataset as determined by the LSD approach. These may be mapped to particular feature objects in the data with relatively uniform spatial frequencies to determine the optimal sample sizes generated by different features. If a single optimal sample size for the full dataset is desired, a weighted mean may be taken of the full set of derived sample sizes.

The mean of the set of sample sizes associated with the set of LSD local maxima determined by the above procedure will be weighted by the LSD value associated with each local maximum. Because every location in the dataset's LSD space is investigated for a possible local maximum, this single average sample size will be implicitly weighted by the area of individual feature objects that generate similar optimal sample sizes due to a relatively uniform spatial frequency response in the data.

• Peakedness and Optimal Sample Size

This complete set of local maxima may not be of uniform quality in terms of the robustness of each maximum found for *LSD = f(r,c,s)*. That is, there may be some very weak or "shallow" maxima that are barely included in the set because they meet the requirements for negative definiteness near the limits of precision for the floating point numbers used in the

calculations. These maxima may have spurious accuracies and may not represent the spatial frequencies of the underlying image or spatial data feature. It may be useful, therefore, to apply a threshold to exclude these lower-quality maxima. The term "peakedness" will be used to describe the strength or quality of the LSD local maximum.

The peakedness of each local maximum will be calculated using the Laplacian of the function $LSD = f(r,c,s)$ evaluated at each point determined by the Hessian matrix calculations. From vector analysis, the Laplacian means the "divergence of the gradient" of a scalar function, and is itself a scalar quantity. For a local maximum of a multivariate function, the Laplacian will be a negative number. The more "peaked" the local maximum is, the more negative the number. In this way the range of Laplacian values can be calculated for the initial full set of local maxima, then a chosen threshold expressed as a percentage of that range to include only those maxima with Laplacian values more negative than the threshold can be applied. The full set of local maxima in LSD space is equivalent to a threshold of zero.

For the purposes here, the scalar function is $LSD = f(r,c,s)$. The Laplacian at any point (r,c,s) is then given by

$$\nabla^2 f = \partial^2 f/\partial r^2 + \partial^2 f/\partial c^2 + \partial^2 f/\partial s^2$$

Fortunately, these second-order partial derivatives were already estimated numerically when calculating the Hessian matrix for each location in LSD space, and comprise the principal diagonal of the matrix. They are now available to calculate the Laplacian for the set of local maxima determined by the Hessian matrix analysis. To do this, the trace (the sum of elements of the principal diagonal) of each Hessian matrix tr(H_{rcs}) in LSD space is found. The full range of Laplacian values, or peakedness, in LSD space can then be found.

The final step is the calculation of optimal sample size. Using peakedness, the effect of different thresholds on the process of finding an optimal sample size for the whole image can be explored. The optimal size is defined as the mean of the set of sample sizes associated with the LSD space locations of the set of local maxima after applying a chosen Laplacian threshold, if desired. This mean is weighted by the number of

local maxima and their associated LSD values at each sample size. It is given by

$$S_{opt} = \sum_{i,j=1}^{n,m} LSD(lmax_{i,j})S_i \Bigg/ \sum_{i.j=1}^{n,m} LSD(lmax_{i,j})$$

where

S_{opt} = optimal sample size
i = resampled image number
n = total number of resampled images
m = total number of LSD local maxima in resampled image i
$LSD(lmax_{i,j})$ = for image i, the LSD value for each j of m local maxima with peakedness above a given threshold
S_i = sample size of image i

2.3   Results and discussion

2.3.1   Semivariograms

2.3.1.1 Results

The ideal pixel sizes for each raster dataset are listed in Table 2, along with the MAE for each resampling of the rasters. The average MAE at 470 m was the lowest among the four optimal pixel sizes, indicating that 470 m is the optimal uniform pixel size to resample each raster. The higher average MAE values at the larger pixel sizes indicated more error was introduced when resampling to those sizes.

Table 2. Ideal pixel size for each raster and its corresponding MAE. The average MAE was lowest when all of the rasters were resampled to 470 m, making 470 m the ideal pixel size.

Data Type	Original Resolution (m)	Ideal pixel size (m)	MAE after resampling to each ideal pixel size			
			470 m	20,457 m	94,901 m	96114 m
Tree cover	30	470	4.2073	8.6604	12.0908	11.4899
Population	1,000	20,457	0.0342	0.3322	0.3504	0.309
Precipitation	5,000	94,901	0.4525	12.5672	21.8766	23.1783
Wind	25,000	96114	0.0002	0.0086	2.47E-08	0.0321
Average MAE:			1.1736	5.3921	8.5795	8.7523

Major and most minor features in each raster remained distinct when the rasters were resampled from their original pixel sizes to best pixel size for the dataset, 470 m. Resampling the population, precipitation, and wind pixels to the best pixel size for the entire dataset at 470 m introduced

redundancy to the dataset, but prevented problems that arose when resampling to each individual raster's ideal pixel size (see Figure 2).

Tree cover (30 m original resolution). The semivariogram calculated for the tree cover dataset resulted in a half range of 470 m. Some local variation in percent tree cover remained apparent at the 470 m pixel size, while linear features among the tree cover, such as waterways, that were visible at the native original of 30 m were no longer visible at 470 m (see Figure 2a).

Population (1,000 m original resolution). The semivariogram calculated for the population dataset resulted in a half range of 20,457 m. Because the population of Cambodia is low throughout most of the country, resampling the raster to the ideal pixel size resulted in substantially reduced higher population values in urban areas that were present at the original spatial resolution (see Figure 2b).

Precipitation (5,000 m original resolution). The semivariogram calculated for the precipitation dataset resulted in a half range of 96,114 m, which is much larger than its native pixel size of 5,000 m. The resulting pixels at 96,114 m were so large that there were gaps where there were not enough pixels on the border of the country (see Figure 2c).

Wind (25,000 m original resolution). The semivariogram calculated for the wind dataset resulted in a half range of 94,901 m. The variation in wind speed was low and when resampled to its optimal pixel size, areas of high and low wind speed remained clear throughout the country; however, resampling to that pixel size left gaps as pixels were removed along the border of the country (see Figure 2d).

Figure 2. Each raster at its original resolution, resampled to its ideal pixel size, and to the pixel size, 470 m, with the lowest average MAE for (a) tree cover; (b) population; (c) precipitation; and (d) wind.
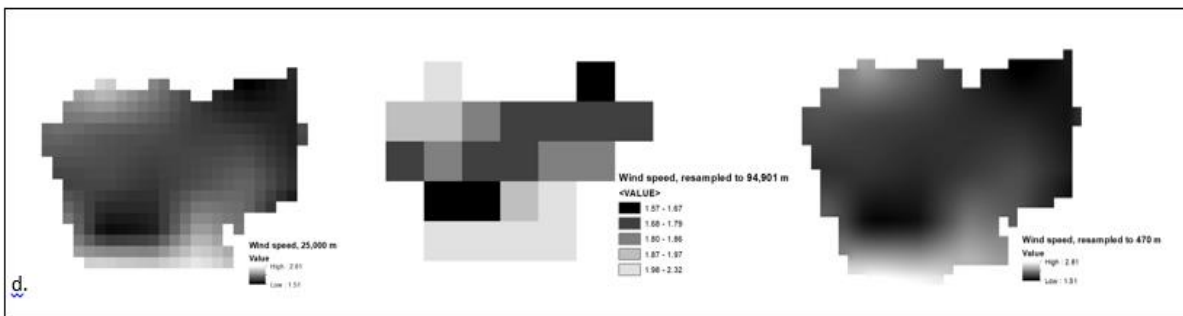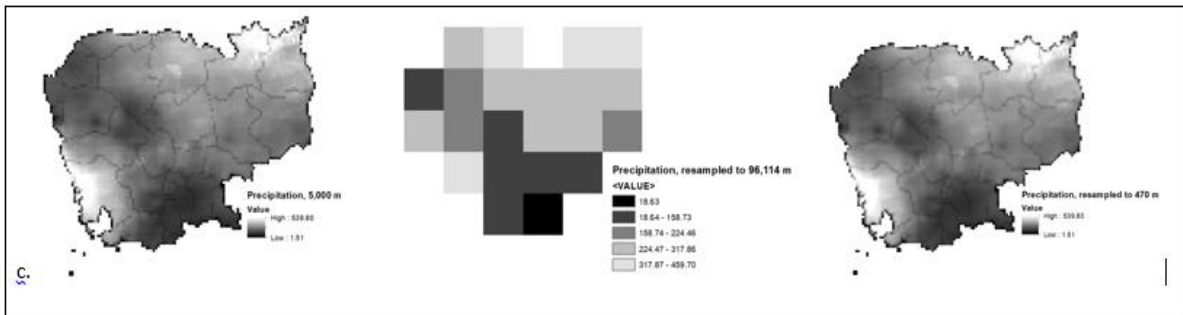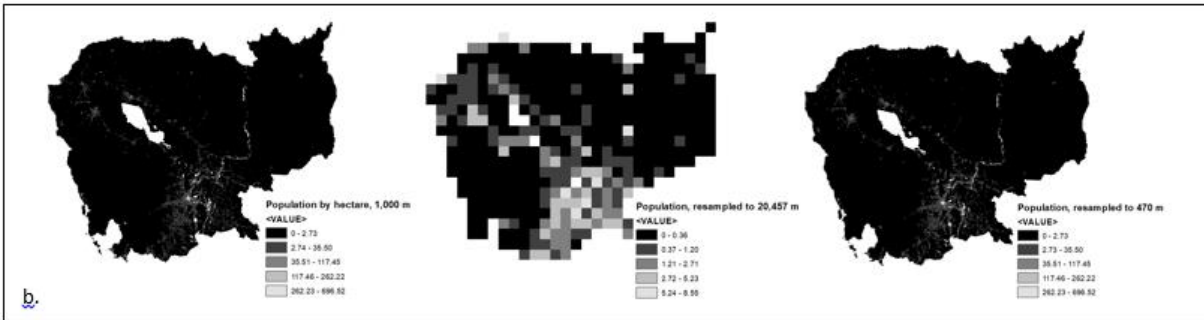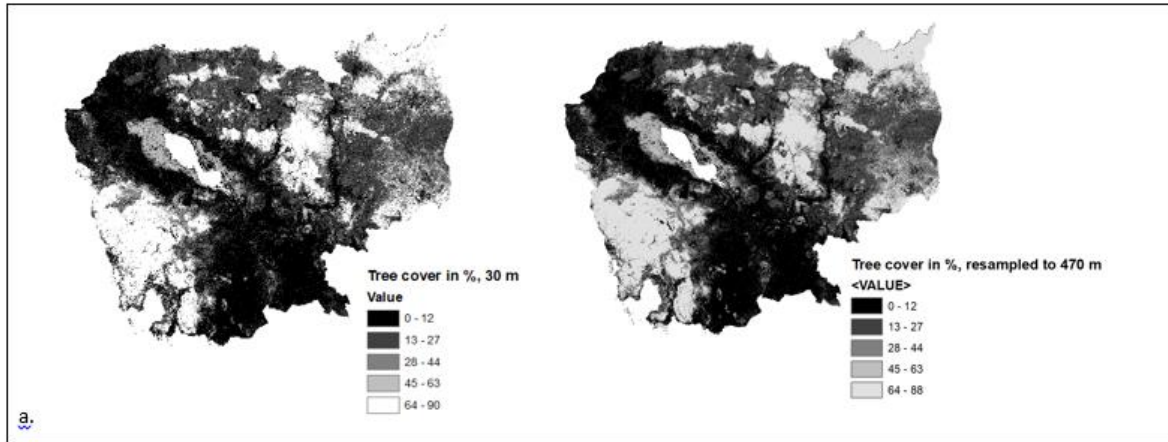
2.3.1.2 Discussion

This paper expands on semivariogram methods used to identify a target uniform pixel size that can be used when analyzing rasters of differing spatial resolutions. This is achieved by finding a pixel size that minimizes both overall error and spatial autocorrelation while maintaining information from the original rasters. The test case of four rasters between 30 m and 25,000 m resolutions resulted in an ideal pixel size of 470 m, which preserves the primary features shown in the data and useful in cases when mapping individual features on the ground is unnecessary.

This approach yielded an optimal pixel size that was smaller than three of the four rasters, thus resampling the three rasters to the optimal pixel required resampling them to resolutions smaller than their original resolutions. As discussed in the introduction, resampling to a smaller pixel size introduces redundancy into the data, which can increase spatial autocorrelation. Spatial autocorrelation violates the assumption of independent observations (Dormann 2007; Legendre 1993; Kühn and Dormann 2012) and redundancy can slow down processing times and take up unnecessary storage (Fisher et al. 2017).

This semivariogram technique would be most useful in species distribution models that require a uniform pixel size, like the maximum entropy model (Phillips 2017; Phillips et al. 2006; Nezer et al. 2017) and the ecological-niche factor analysis (Hirzel and Arlettaz 2003; Hirzel et al. 2002; Santini et al. 2019).

The resulting ideal pixel size is consistent with pixel sizes used in species distribution models. Hao et al. (2019) reviewed data used in species distribution studies and found that data ranged from 5 m in small-scale studies to 110,000 m in global studies. Data that are analyzed on the local level are often resampled to the smallest pixel size in the dataset (Soucy et al. 2018), which may be necessary in modeling presence of disease vectors, but unnecessary in other scenarios that do not require fine detail. For example, Santini et al. (2019) modeled species abundance at a large pixel size based on the theory that assumes a pattern arises at a geographic scale, irrespective of local variations. Fisher et al. (2017) found that higher resolution imagery performed better in characterizing environmental quality variables comprising a watershed model, but also did a cost-benefit analysis and determined costs to be higher utilizing high-resolution imagery.

2.3.2 Local spatial dispersion

2.3.2.1 Results

An overview of results from the multiscale LSD processing algorithm of several examples of geospatial data will be provided: a WorldView2 image (Figure 3) over Florida processed as Normalized Difference Vegetation Index (NDVI) values and three environmental datasets from Cambodia for tree cover (Figure 4), population density (Figure 5), and precipitation (Figure 6). These datasets have a wide disparity of spatial resolutions: 1.3 m, 30 m, 991 m, and 4,954 m, respectively.

Dataset statistics from the LSD optimization processing are shown in Table 3, including the optimal sample size results with and without the peakedness threshold. Example peakedness thresholds were chosen for each dataset to display an appreciable fraction of the total number of local maxima. The number of pixels available in each original image acts as an upper limit on the number of resampled images that can be created for optimization. This is controlled by varying the maximum percentage of edge pixels. For consistency in comparison, all processing was performed with the following parameters in common: computation kernel size, 3x3; resample method, pixel block mean value; LSD statistic, MAD; and finite difference equation delta value in pixels, 1.

Figure 3. WorldView2 image (Florida).

Figure 4. Tree cover (Cambodia).

Figure 5. Population density (Cambodia).

Figure 6. Precipitation (Cambodia).
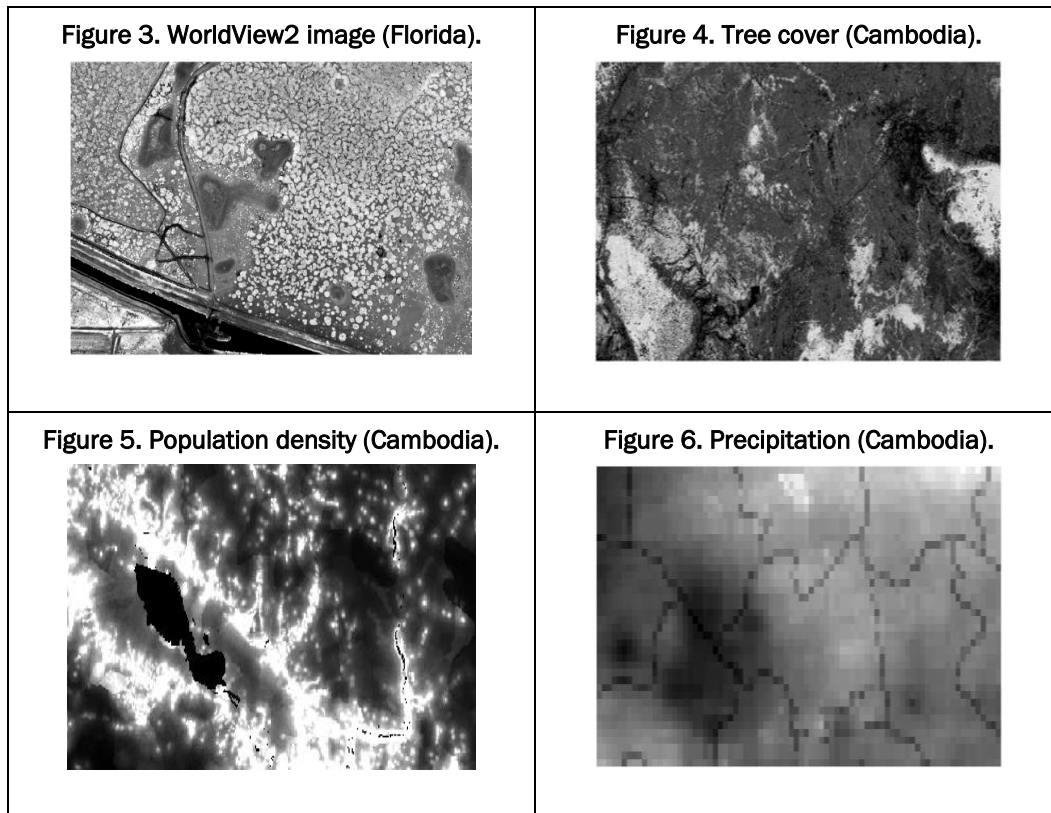
Table 3. Dataset parameters and optimal sizes.

Dataset	native res. (m)	rows	cols.	# images	peak thresh. (%)	# local maxima		optimal size (m)	
						thresh.	no thresh.	thresh.	no thresh.
Florida WorldView2	1.3	908	1304	28	25	311534	707546	3.71	4.78
Cambodia tree cover	29.7	1470	3277	21	20	81936	3026162	50.1	96.8
Cambodia population	991	197	395	29	1	6601	40344	2180	3212
Cambodia precipitation	4954	40	80	7	15	868	1480	10984	11572

• Florida WorldView2 Dataset

For this high-resolution dataset depicting a mix of canopy, linear features, and open ground, Figure 7 shows a plot of the mean and median of the chosen LSD statistic (in this case, MAD) for each resampled image. Both measures of central tendency reach a maximum at a sample size of approximately 6.5 m. This value agrees with the optimal sizes given by the LSD optimization process. Figure 8 shows the frequency distribution of local maxima across the series of resampled images. The maxima become less frequent in the resample size dimension, except for a slight increase at

the first resample size of 2.6 m. This image contains the highest fraction of local maxima. The thresholded subset of these is depicted in Figure 9, showing their distribution across the image resampled to 2.6 m. It is apparent that they are spatially associated with different features in the images, such as the pattern of canopy and the edges of the canal in the lower left. The full distribution of thresholded local maxima in LSD space is shown as a point cloud in perspective view in Figure 10. Note the influence of the image's linear features in the vertical distribution of local maxima.

Figure 7. Mean, median MAD vs. sample size.
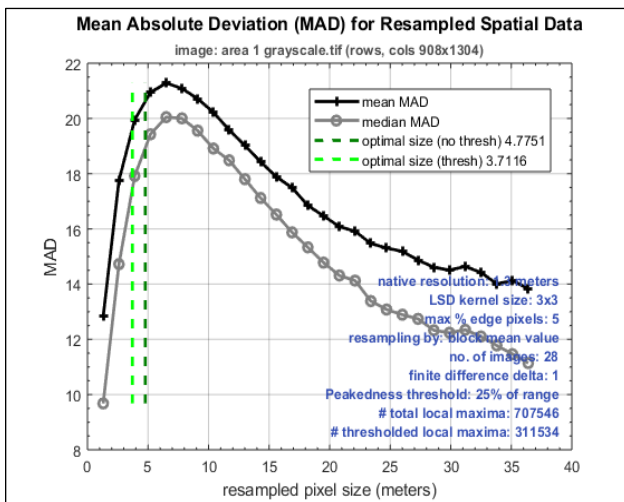
Figure 8. Local maxima vs. sample size.
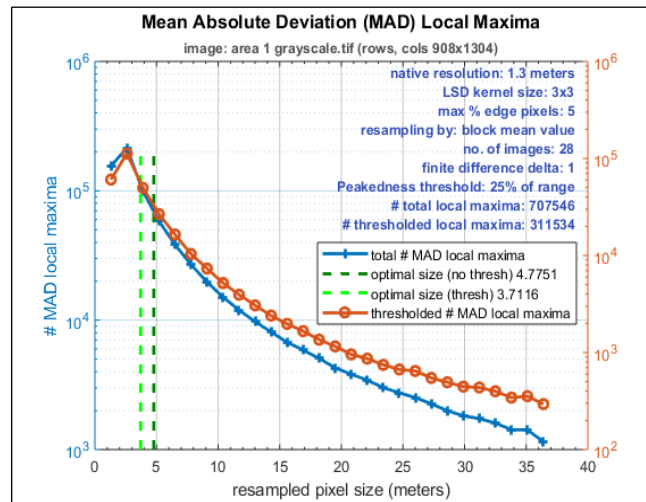
Figure 9. Local maxima distribution at 2.6 m.
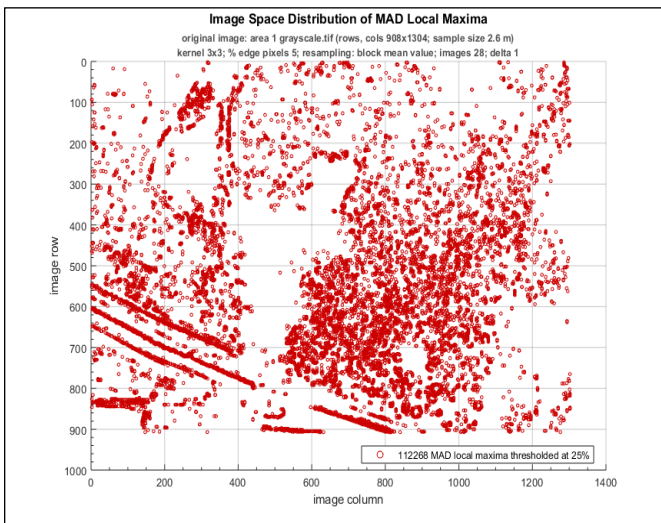
Figure 10. Local maxima distribution in LSD space.
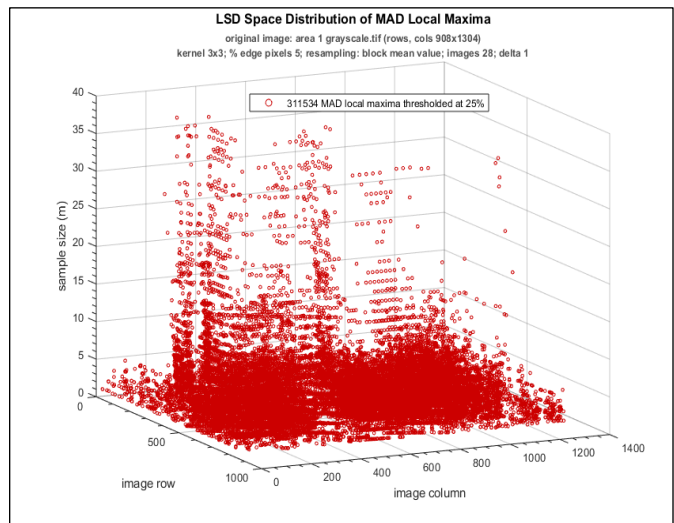
•   Cambodia Tree Cover Dataset

Figure 11 shows a plot of the MAD mean and median for each resampled image. In this case, their plots do not reach a local maximum against resample size, so an indication of an optimal size is not given. In spite of this, the LSD optimization method provides optimal sizes of 97 m (unthresholded) and 50 m (thresholded) for a native resolution of 30 m. Figure 12 shows a heat map of MAD values at the sample size 89 m. This sample size is the closest in the series to the calculated optimal size of 97 m. Figure 13 depicts the distribution of thresholded local maxima derived from the MAD heat map distribution for the 89 m resampled image. It is apparent that the local maxima arrange themselves at locations where there are sudden changes in MAD values across the image space as seen in Figure 13. Figure 14 shows a peakedness histogram for the total set of local maxima. Since they were thresholded at 20% of the peakedness range, it is apparent that the remaining maxima represent a small fraction of the total. Table 3 shows that this figure is 81936/3026162 or 2.7%. Of these, 6,068 local maxima are found at sample size 89 m, but this is sufficient to reveal their distribution according to the change of variance across the image space.

Figure 11. Mean, median MAD vs. sample size.
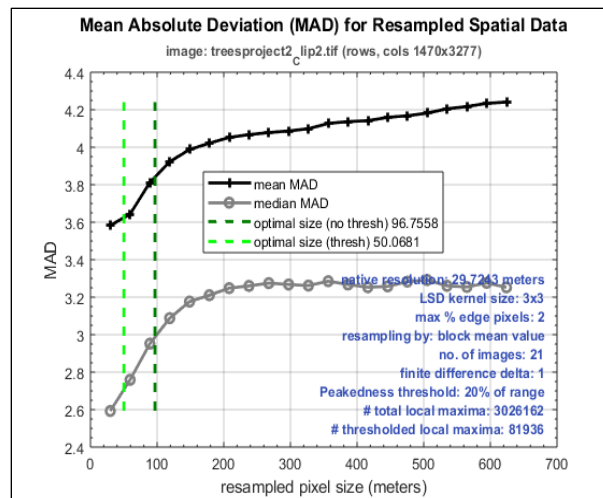
Figure 12. MAD heat map at sample size 89 m.
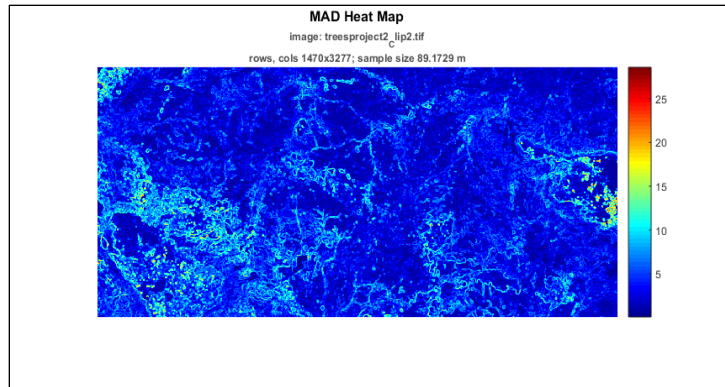
Figure 13. Local maxima distribution, sample size 89 m.
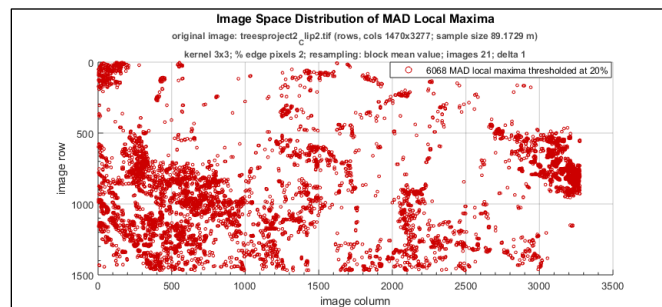
Figure 14. Peakedness histogram.
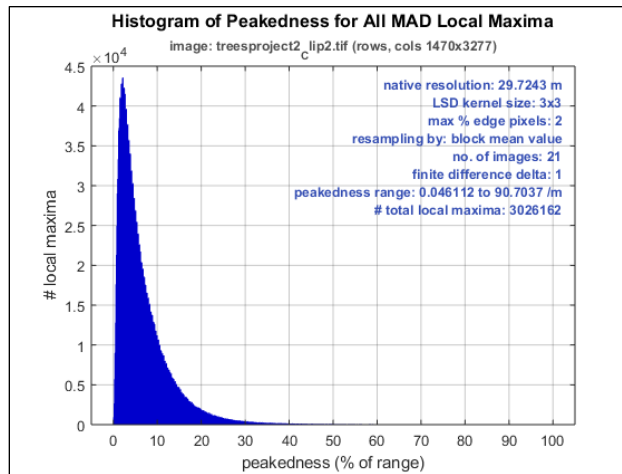
2.4   Cambodia population density dataset

This dataset contains a large body of water where there are no values. Higher population densities surround the lake and line the watercourses that empty into it. As reported in Table 3, the calculated thresholded and unthresholded optimal sample sizes are 2,180 and 3,212 m, respectively, given the native spatial resolution of 991 m. Figure 15 shows an upward

trend in the MAD mean and median plots for the lower sample sizes in the series of 29 images, along with the computed optimal sizes of 2,180 m and 3,212 m for thresholded and unthresholded peakedness, respectively. Figure 16 shows the MAD heat map for the resample size 2,972 m, the size closest to the unthresholded optimal value. The full point cloud distribution of thresholded local maxima in LSD space derived from the MAD values. However, this view looks straight down along the sample size axis at the local maxima found in the entire resampled image series. Figure 17 shows a scatter plot of MAD values for all local maxima in LSD space, plotted against their peakedness values. This plot gives the user a sense of how the maxima are distributed across the peakedness range as well as the range of dispersion from which they were derived. Figure 18 is a plot of mean peakedness for each image in the resample series, showing that it is highest at the original spatial resolution and then drops down to a relatively constant value as sample size increases.

Figure 15. Mean, median MAD vs. sample size.
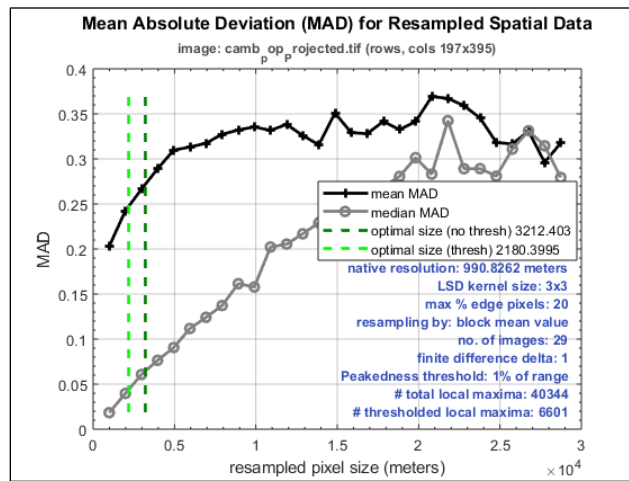
Figure 16. MAD heat map, sample size 2972 m.
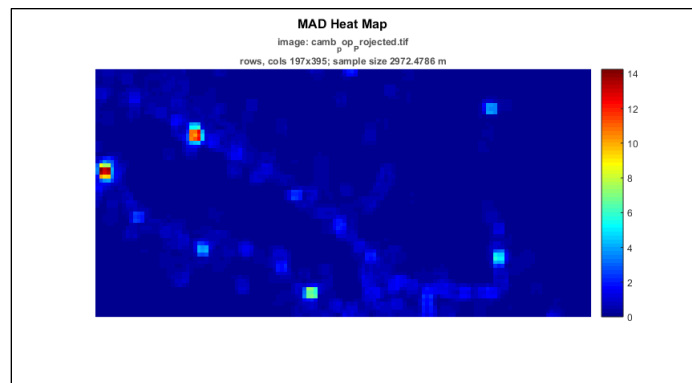
Figure 17. Local maxima distribution in LSD space.
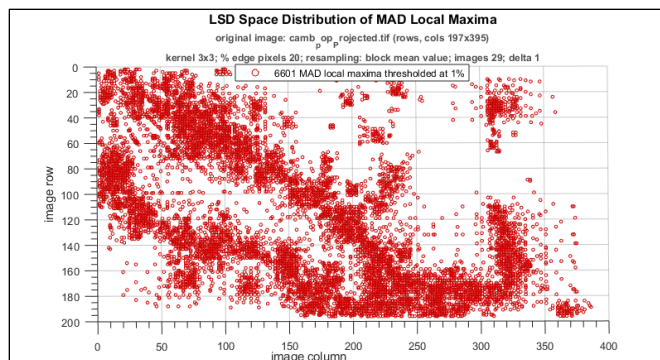
Figure 18. Scatter plot of MAD vs. peakedness.
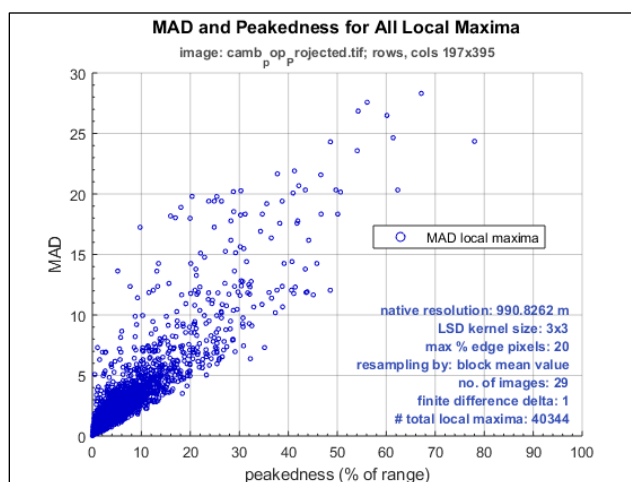
- Cambodia Precipitation Dataset

This dataset has the largest native resolution of 4,954 m, a plot of the MAD mean and median is shown for each resampled image (Figure 19). In this case, the plots not only do not reach a local maximum against resample size, but also continue an upward trend through the resample size series. Yet, the LSD optimization approach still provided reasonable optimal sizes of 11,572 m (unthresholded) and 10,984 m (thresholded). The sample size in the resampled image series closest to the calculated optimal sizes is 9,909 m. Figure 20 is a histogram of the frequency of MAD values in the image with that spatial resolution, showing a maximum at a MAD value of about 10-12 m. A MAD heat map is provided for sample size 9,909 m in Figure 21. Here, it can be seen that the higher dispersion values are associated with transition zones with higher spatial frequencies in the original image. Finally, Figure 22 shows a perspective view of the point cloud of thresholded local maxima throughout LSD space. Their distribution appears more homogeneous at higher levels in the space.

Figure 19. Mean, median MAD vs. sample size.
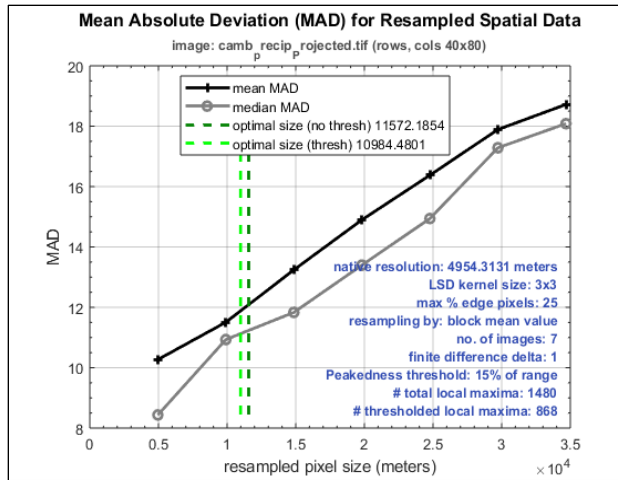
Figure 20. MAD value frequency histogram, 9,908 m.
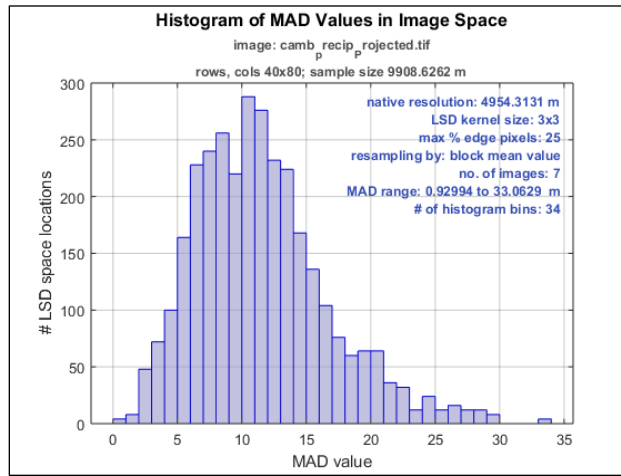
Figure 21. MAD heat map, sample size 9,908 m.
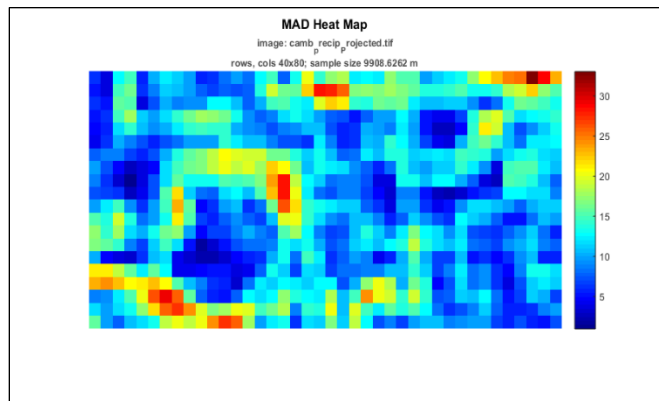
Figure 22. Local maxima distribution in LSD space.
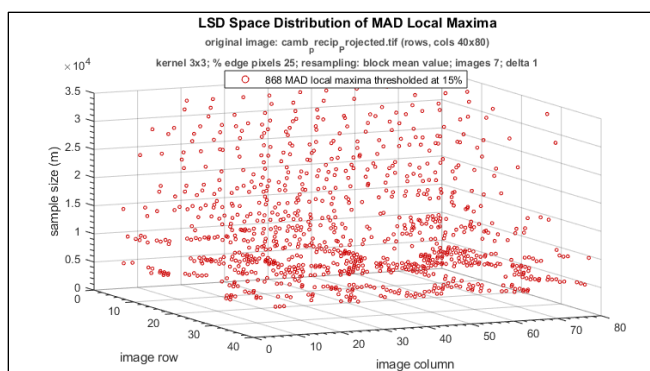
2.4.1.1 Discussion

The spatial characteristics of continuously varying phenomena on the Earth's surface directly inform remotely sensed data or other types of environmental information collected in a geospatial context. The spatial domain, or structure of this data, can be used to optimize its interpretation or extraction of spatial information. Effective mapping or modeling of spatially dependent information requires capturing the spatial variation patterns of features of interest. A key consideration in image analysis is the relationship between spatial resolution and the spatial frequency structure of features found in the image data. In this methodology, optimal sample size results were driven by the number and distribution of LSD local maxima as well as the LSD values associated with each local maximum. If a peakedness threshold is chosen, the set of local maxima is first winnowed by a minimum peakedness value.

The setting of a peakedness threshold can be a useful tool for exploring the distribution and peakedness of the local maxima set in LSD space by examination of various plotting options in the LSD Analysis Tool. A threshold is required if the retention of only high-value LSD optima for optimal sample size calculations is indicated. However, a general strategy has not been identified for choosing a threshold and, absent a supporting rationale for its use, selecting the unthresholded optimal size as a default procedure is recommended. In this work, a multiscale modeling approach to determine an optimal sample size for raster images containing remotely sensed or other environmental data with variable spatial structure was successfully examined. Resampling an image dataset in this way can increase the efficiency of image processing functions such as feature segmentation or of geospatial models such as that employed in the NET-CMO project at ERDC-GRL.

- Graphical User Interface

A useful tool and user interface was also created, called the LSD Analysis Tool, to exercise the algorithmic approach and allow a user to interactively process a dataset while in control of particular processing parameters (Figure 23). Various plotting options display relationships among LSD values, local LSD maxima, maxima peakedness, and LSD space locations. These output features and level of user control provide for repeated experimentation and better understanding of the spatial data structure.

Figure 23. Local spatial dispersion analysis tool.
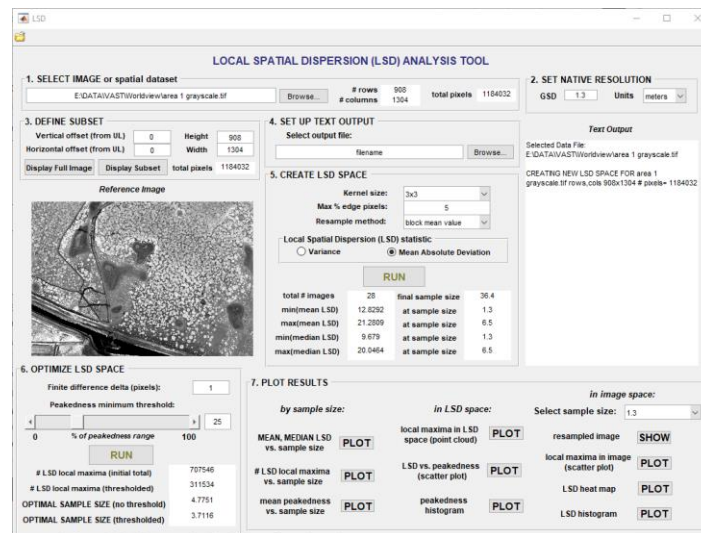
3   Spatial Downscaling

3.1   Literature review

Mosquito-borne illnesses are a significant public health concern, both to the Department of Defense (DoD) and the broader national and international public health community. To truly understand these diseases and their threats, a thorough grasp of their spatial distribution, patterns, and determinants is needed (Pages et al. 2010). This information, when available, is often only at a sub-national to regional scale. Such data availability fails to meet tactical-level applications when diseases exhibit high local variation (Rytkonen 2004; Linard and Tatem 2012). Additionally, finer spatial resolution is also required to successfully target disease burden within the population and reduce exposure.

Previous research has applied spatial downscaling techniques to meet specific epidemiological study needs requiring more localized statistics. Examples include downscaling malaria incidence rates from regional to urban centers through multivariate regression, hand-foot-mouth disease from national to township levels using generalized linear models, and applying hierarchical Bayesian frameworks to develop 5 km gridded risk maps of malaria, *Plasmodium falciparum*. (Gething 2012; Wang et al. 2017; Altamiranda-Saavedra et al. 2018). While these studies were able to improve coarse-scale information, they still failed to meet a spatial resolution relevant to tactical-level epidemiological mapping applications or the processing time required to support time-sensitive operations.

3.2   Geoanalytic data and methods

The research presented in this report focused on dengue, a mosquito-borne viral disease transmitted by female mosquitoes, mainly of the species *Aedes aegypti*, the same vector responsible for transmitting chikungunya, yellow fever, and Zika infection. Dengue is endemic to the tropical belt and greatly influenced by rainfall, temperature, and unplanned rapid urbanization, with the severest form of disease being the leading cause of hospitalization and death among children and adults in Latin America and Asia (Brady et al. 2012). While oral prophylaxis can prevent mosquito-vector diseases such as malaria, there are no specific vaccines or antiviral treatments against dengue fever (Hesse et al. 2017).

This lack of treatment not only puts local populations at risk, but can also adversely impact military operations.

3.2.1   Data

Researchers at the Geospatial Research Laboratory (GRL) queried provincial-level dengue incidence rates at monthly intervals between 1998 and 2010 from Project Tycho, a global health research database maintained by the University of Pittsburgh (Panhuis et al. 2018). Cambodia served as the ROI due to the endemicity of dengue, high local variation in disease incidence, and availability of administrative-level statistics. The data were reformatted to CSV and spatially joined in ESRI ArcMap to the Large-Scale International Boundary (LSIB)[*] shapefile (Humanitarian Information Unit 2017).

GEE served as the high-performance cloud computing (HPC) environment used to process monthly composites of environmental, demographic, and landscape covariates between 1998 and 2010. GEE combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities that includes vector and raster data processing, machine-learning classifiers, and time series algorithms (Gorelick et al. 2017).

3.2.2   Methods

The methods in this research followed spatial downscaling principles found in similar studies that include improving coarse population and demographic data, and remotely sensed products such as precipitation, soil moisture, and surface temperature (Gaelle et al. 2016; Zhang et al. 2016; Ezzine et al. 2017; Pang et al. 2017). The downscaling methods use a statistical algorithm to determine a relationship between a coarser response variable and finer spatial resolution covariates. This study chose to apply the random forests (RF) regression algorithm because of its demonstrated ability to yield higher accuracy compared to linear modeling techniques, albeit more difficult to interpret than a traditional linear regression (Couronne et al. 2018). RF is an ensemble classifier that constructs multiple de-correlated random regression trees that are bootstrapped and aggregated using the mean predictions from all

---

[*] LSIB: Large Scale International Boundary Polygons, Simplified. U.S. Department of State, Office of the Geographer at **https://catalog.data.gov/dataset/global-lsib-lines-simplified-2017mar30**

regression trees (Breiman 2001). RF models also provide a quantitative measurement of each variable's contribution to the regression output, which is useful in evaluating the importance of each variable concerning dengue prevalence and conditions that affect disease vector suitability.

In this case, the monthly dengue incidence rates previously compiled in ESRI ArcMap serve as the response variable. The monthly composites of environmental, landscape, and demographic geospatial data serve as the covariates used to develop a response function and model incidence rates to a user-defined output pixel size; this study selected 1,000 m output grid cells because it met the high-resolution criteria of previous fine-scale epidemiology studies (Sturrock et al. 2014; Delmelle et al. 2014). As previously stated, rainfall, temperature, and urbanization significantly affect the presence of dengue, primarily due to influences on habitat suitability for the mosquito vector, *Aedes aegypti*. The spatial covariates used in this study included precipitation, land surface temperature, NDVI, population, land cover and land use, and elevation (Table 4, Figure 24).

Table 4. Spatial covariate types and data sources used in the epidemiological.

Type	Spatial Covariate	Source
Environmental	Precipitation sum mean	CHIRPS
	Land Surface Temperature (Day and Night) min mean max	MODIS
	NDVI*	MODIS
Landscape	Elevation	SRTM
	Annual Land Cover Product	MODIS
Demography	Human Population	WorldPop
* Normalized Difference Vegetation Index, measure of vegetation cover and vigor		

Figure 24. Spatial covariates representing environmental, landscape, and demographic determinants.
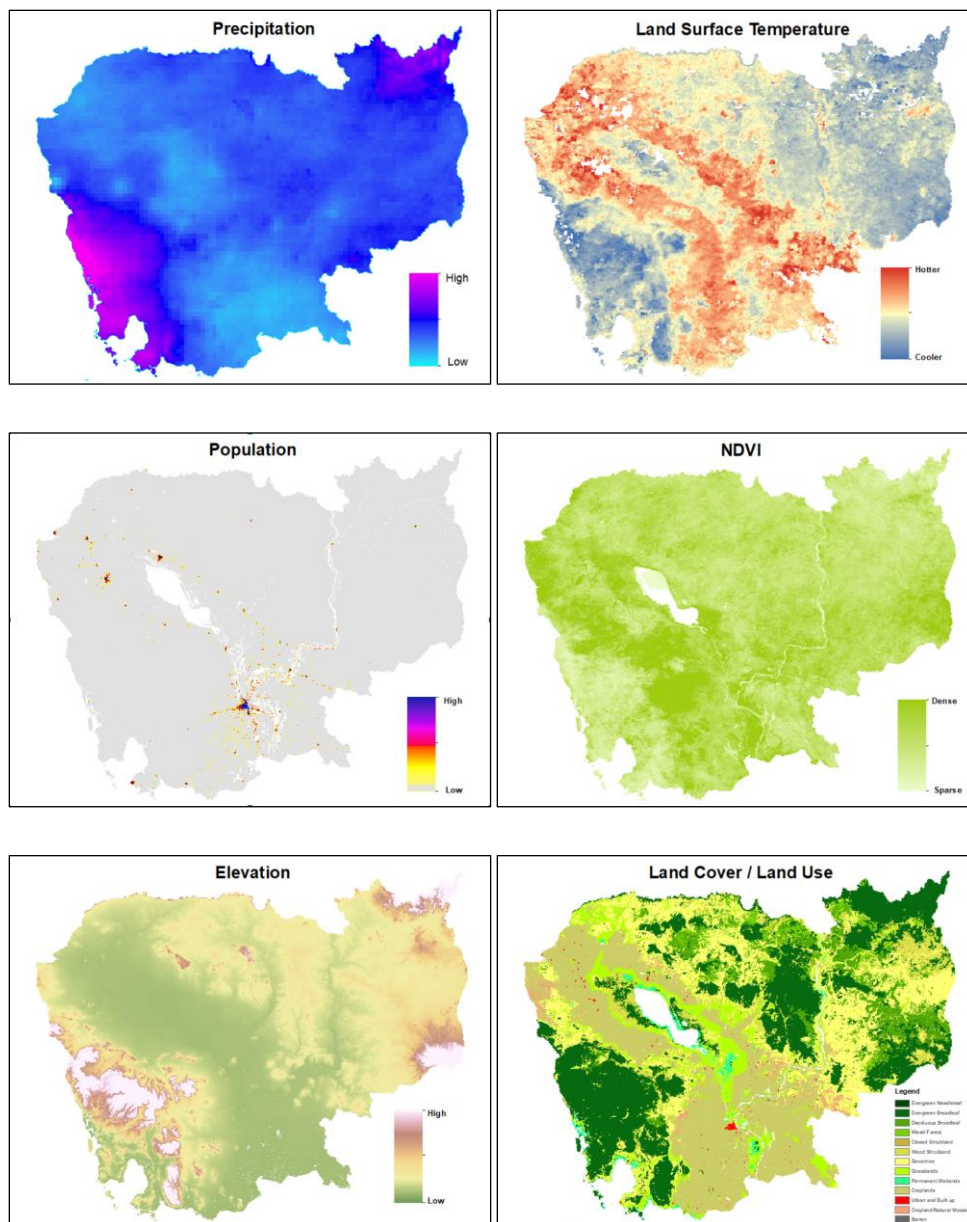
The spatial downscaling methodology is summarized in the sequential steps below:

1. Query and download administrative-level monthly dengue incidence rates from Project Tycho.
2. Spatially join dengue incidence rates to Large-scale International Boundaries (LSIB) shapefile and upload to GEE as a table asset.
3. Query environmental, landscape, and demographic spatial covariates in GEE and spatially reduce to monthly composites.

4. Select month and year to model.
5. Create a stratified sampling scheme in GEE and extract observed
    incidence rates (response variable) and environmental/landscape
    variables (covariates) for the time period.
6. Build RF classifier using regression and run prediction; Validate
    regression outputs by aggregating predicted grid cell values to the
    provincial boundary and compare to observed administrative-level
    incidence rates.

3.3   Results and discussion

Figure 25 provides a visual comparison between the gridded values
derived from the RF regression downscale model and the observed
provincial-level incidence rates for June 2010. The gridded output clearly
shows a much higher spatial fidelity that meets any number of tactical,
operational needs. The gridded output can serve as a disease risk map that
provides an understanding of the spatial variability in dengue and
locations of higher risk to exposure. Also, the HPC environment of GEE
made it possible to develop a gridded model for the entire nation within
minutes, a task that would be computationally intensive and time-
consuming if duplicated in a desktop PC environment.

Figure 25. (a) Results of 1,000-m downscaled product compared to (b) provincial-level
statistics.

(a)   (b)
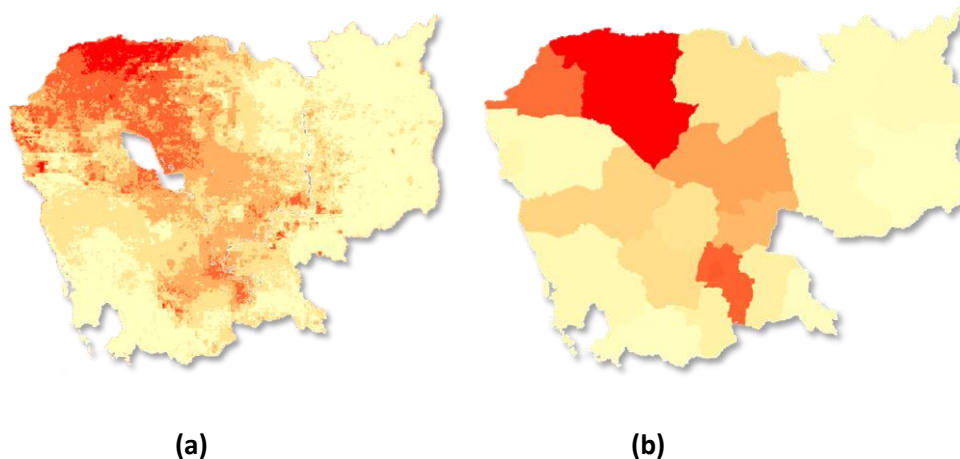
3.4   Discussion

Figure 26 lists the RF spatial covariates in order of importance for June 2010. Population, temperature, vegetation cover, and precipitation were the most important variables, respectively, for describing the model, which coincides with epidemiological literature. The order of variable importance remained relatively consistent regardless of the chosen month and year.

Figure 26. Spatial covariates in order of importance to the June 2010 downscale model.
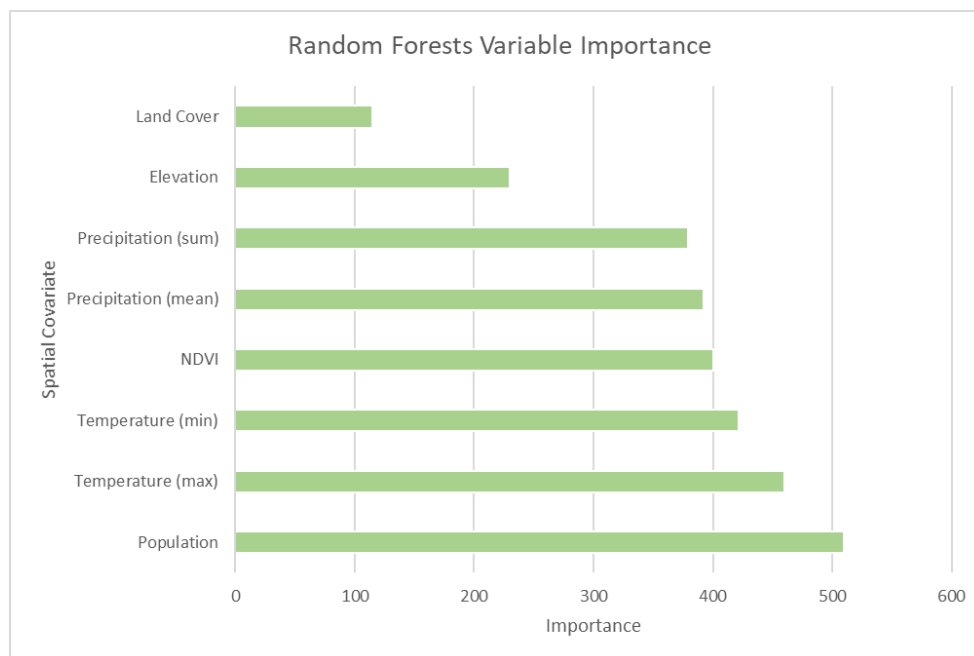
Figure 27 provides an example of model validation results for June 2010 using the spatial aggregation technique described in Step 8 of the methodology summary. Grid cell values of predicted disease incident rates were averaged within each administrative boundary and compared to the observed incidence rate for that given province. The absolute minimum and maximum difference between observed and downscaled data was 0.92 and 16.6 with the root mean square error (RMSE) being 5.64. A scatterplot was also used to compare observed and downscaled values yielding an R^2 of 0.87 (Figure 28).

Figure 27. Comparison between observed and downscaled output per province.
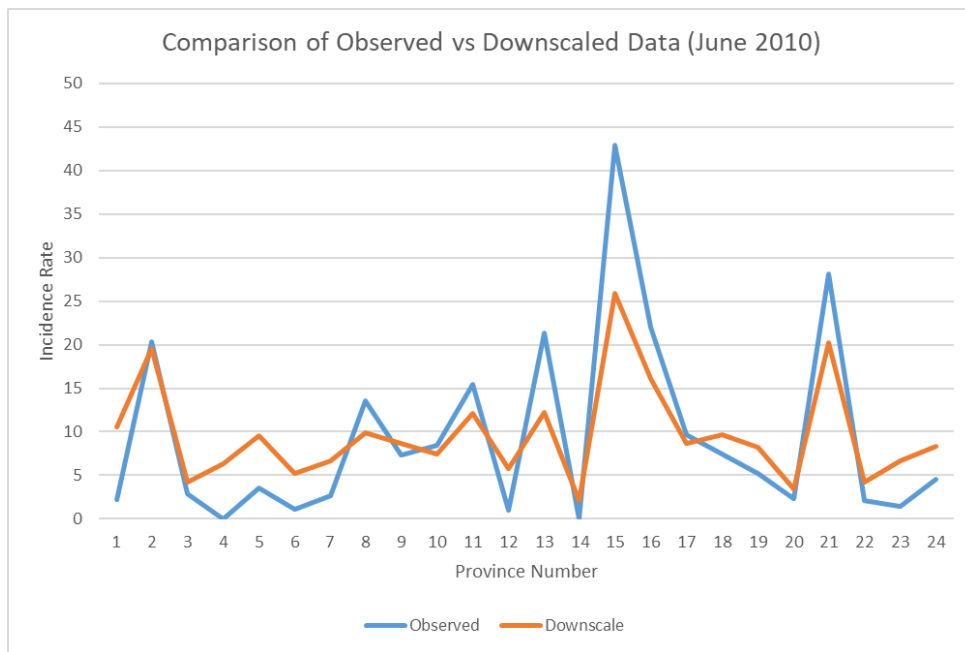
Figure 28. Scatterplot of observed and downscaled output per province.
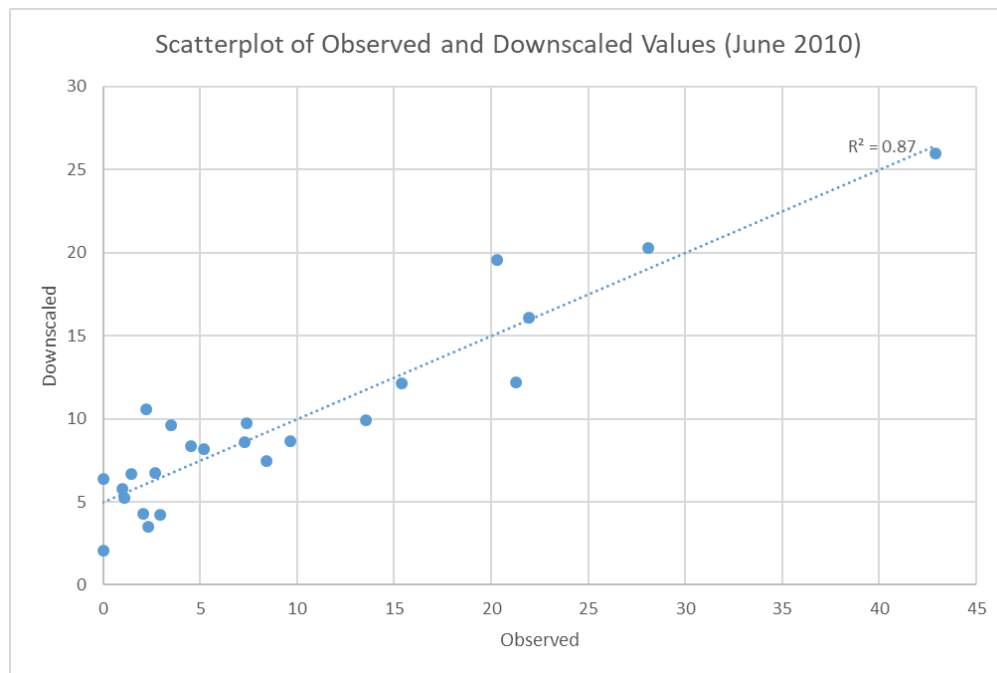
Spatial downscale models were developed for each month between 1998 and 2010, totaling 156 geospatial disease risk products. All models showed significant agreement between downscaled and observed data with the highest RMSE being 10.25 and the lowest being 1.22. The lowest calculated r^2 for the scatterplot comparisons was .72, and the highest was .94.

RF regression proved to be a high performing predictive algorithm that required little knowledge of machine-learning to achieve good results. However, the RF model is not as easily interpretable as a traditional linear regression or classification/regression tree (CART), mainly due to the ensemble technique that creates hundreds of random, independent tress and then combines the average into a single result.

GEE provided a HPC environment that met the standards required for tactical and operational standards. Gridded products at 1,000 m spatial resolution were processed at national-levels within minutes as opposed to several hours on a desktop environment. Further advantages to GEE include reduced local resources, both related to computation and data accessibility. The main disadvantage to GEE is that it is not aimed at the novice user since it requires programming knowledge of either JavaScript or Python.

Future contributing work to this study would explore the local spatio-temporal dynamics of the downscaled models. Dengue is known to be influenced by seasonal variables such as precipitation and surface temperatures. Identifying strong temporal signals within a time-series could provide a further understanding of risk trends overtime and possible associations with climate-disease teleconnections.

In conclusion, this study improved coarse, administrative-level disease data by downscaling to a 1,000 m grid cell using RF regression and spatial covariates. The generated output provides the level of tactical precision required to support Civil-Military Operations (CMO) targeting human health initiatives at a local scale. The output also provides a detailed geospatial product of disease risk that can be used to inform doctrine related to force health protection and force readiness during deployments.

4 Temporal Disaggregation

4.1 Literature review

Lack of disease incidence data is a common problem within the field of epidemiology (Beale et al. 2008). In order to study the distribution, frequency, patterns, and predictors of disease, epidemiologists need to have a sufficient record of past disease cases. Unfortunately, for many areas of the world, disease incidence data is extremely lacking. At best, data may be available at country level or on an annual basis. From a temporal point of view, this is unacceptable for identifying periods of extreme disease activity, the seasonal patterns of disease, or prediction. Since the global collection of higher fidelity disease data is unlikely, a way to temporally disaggregate annual disease incident data must be identified.

Temporal disaggregation is the process of taking a low frequency time series, such as annual disease cases, and dividing the time series into a higher frequency time series, such as monthly disease cases. This process of down sampling is typically done by 1) dividing the lower frequency time series into equal portions of the higher frequency time series or 2) using one or more related higher frequency time series to model the desired signal (Chamberlin 2010). For example, if disease transmission is dependent on the presence of mosquitoes, then infections are expected to occur more frequently when conditions are good for mosquito survival and reproduction. High frequency indicator time series of mosquito presence, such as temperature and precipitation, are often recorded and readily available.

Within epidemiological literature, disaggregation of disease data has focused on the spatial distribution of disease (Rahman 2017). RF, spatial scan statistic, and neural networks (Khan et al. 2005; Kitron et al. 2006; Mendes and Marengo 2009; Rahman 2017) have used disease related variables, such as population density, land cover, and climatic factors, to spatially distribute disease incidents from a large region of interest to several smaller geographic areas. This process provides a finer resolution of the physical location of disease. The process of temporally disaggregating disease incidents, however, has not yet been explored.

Some popular methods for temporal disaggregation include neural networks, splines, and regression (Huth 2002; Kumar et al. 2012; Herath et al. 2016). However, the most commonly used methods are those developed specifically for temporal disaggregation, mainly when dealing with economic data (Chamberlin 2010; Sax and Steiner 2013). The basic framework for these methods, known as Denton, Denton-Cholette, Chow-Lin, Fernandez, and Litterman, can be broken into three separate steps (Chamberlin 2010; Sax and Steiner 2013). The mathematical notation of these steps and their descriptions are in Table 5.

Table 5. Variable notations and their definitions.

Notation	Description
y	The unknown time series of interest
y_l	The known low-frequency version of y
$\hat{y}$	The disaggregated time series of y_l
p	A preliminary estimate of y
X	The matrix of related time series
C	The matrix to convert high frequency to low frequency
D	The distribution matrix
Σ	The variance-covariance matrix

- Step 1: Estimate the time-space variance-covariance matrix

  Estimate the variance-covariance matrix, Σ, in the high-frequency time series space. Within this space, each dimension corresponds to a single point in time in the high-frequency time series. How Σ is calculated differs for each disaggregation method (Sax and Steiner 2013).

- Step 2: Compute a preliminary estimate, p, of the desired signal, y

  Denton and Denton-Cholette (Sax and Steiner 2013) simply estimate p as p = X, which only works when there is a single indicator series or no indicator series. Chow-Lin, Fernandez, and Litterman (Sax and Steiner 2013) compute p as a Generalized Least Squares (GLS) estimate of the desired signal y (Sax and Steiner 2013). Specifically, p = X$\hat{B}$ where

$$\hat{B} = [X^T C^T (C \Sigma C^T)^{-1} C X]^{-1} X^T C^T (C \Sigma C^T)^{-1} y_l$$

- Step 3: Adjust p using y_l to get a final estimate $\hat{y}$

When computing p, no restrictions are imposed in relation to the known low-frequency time series y_l. In particular, for the final estimate $\hat{y}$, disaggregation requires that $C\hat{y} = y_l$. This can be computed using a right-pseudoinverse D of C and the low-frequency error $u_l := y_l - Cp$ in the preliminary estimation of y:

$$\hat{y} = p + Du_l$$

where

$$D = \Sigma C^T (C\Sigma C^T)^{-1}$$

From here it can be seen that $C\hat{y} = Cp + u_l = y_l$, which is the original signal. It follows that if these techniques can be applied to economic data, they can be applied to vector-borne disease data as well.

4.2 Geoanalytic data and methods

4.2.1 Data

Monthly records of dengue from 2005- 2010 for each province in Cambodia were collected from the University of Pittsburgh's Project Tycho (van Panhuis et al. 2014). Dengue is a vector-borne disease that is spread by the *Aedes aegypti* and *Aedes albopictus* mosquitoes and is typically found within tropical and sub-tropical areas of world. There is currently no vaccine for dengue. While early detection and proper treatment can severely lower fatality rates, dengue has become a leading cause of hospitalization and death among both children and adults in most Asian and Latin American countries (WHO 2019).

Environmental risk to dengue, or any mosquito-borne disease, is related to the parameters of precipitation, temperature, land cover, and vegetation health, all of which affect the habituation of mosquitoes (Chitnis et al 2012; Stuckey et al. 2014; Parham et al. 2015; Siraj et al. 2015). Monthly records of these parameters were obtained from remotely sensed data (Table 6), with NDVI considered a measure of vegetation health. Once collected, the environmental parameters were sub-setted to each Cambodian province and averaged in order to form a non-spatial time series.

Table 6. Remotely sensed environmental variables and their sources. Variables were spatially averaged for each province for every month of analysis.

Environmental Variable	Data Source
Normalized Difference Vegetation Index (NDVI)	MODIS (MOD43A4)
Precipitation Sum	CHIRPS
Precipitation Mean	CHIRPS
Precipitation Sum Lagged	CHIRPS
Precipitation Mean Lagged	CHIRPS
Land Surface Temperature (LST) Day Mean	MODIS (MOD11A1)
LST Day Minimum	MODIS (MOD11A1)
LST Day Maximum	MODIS (MOD11A1)
LST Night Mean	MODIS (MOD11A1)
LST Night Minimum	MODIS (MOD11A1)
LST Night Maximum	MODIS (MOD11A1)

Within the dengue time series, there were five provinces that did not have continuous dengue records. These provinces are as follows: Kracheh, Stueng Traeng, Mondol Kiri, and Krong Pailin. In order to only use continuous time series, these provinces were deleted from the dataset. A map of the Cambodian provinces used in the analysis can be seen in Figure 29.

Figure 29. Cambodian provinces used for temporal disaggregation.
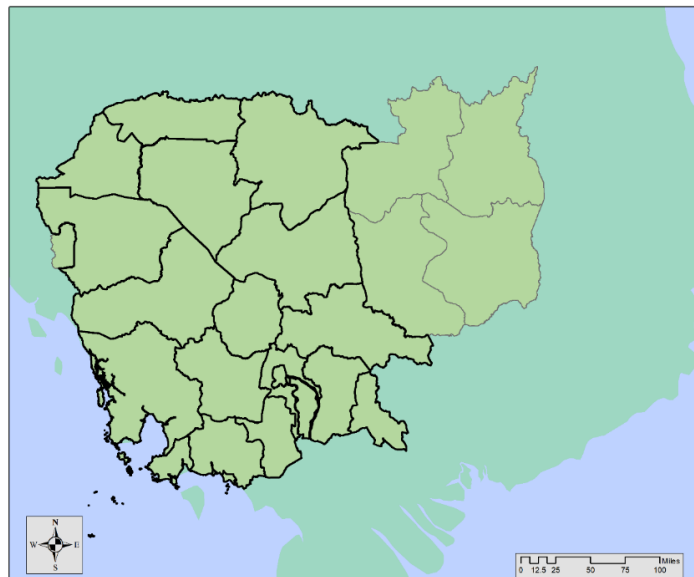
4.2.2   Methods

The first step of the analysis was determining which environmental variables were temporally correlated to dengue. To do this, the dengue and environmental time series of each Cambodian province were averaged together to create a mean, country level time series of dengue, and its predictors. The mean environmental time series were then correlated to the mean dengue time series. Environmental variables that had an R-coefficient of 0.3 or higher, and were significant at α=0.05, were kept for disaggregation. Additionally, variables were tested for collinearity. If a pair of variables were found to have a collinearity of at least 0.90, the variable that was least correlated to dengue was discarded.

Next was determining if the mean dengue time series was stationary. A stationary time series is one whose mean and variance are constant throughout time. A stationary time series is relatively easy to predict and most forecasting models assume that a time series is either stationary or can be made stationary. For temporal disaggregation, whether or not a time series is stationary affects which disaggregation method is used. For stationary series, the Chow-Lin disaggregation method is the best choice (Chamberlin 2010; Sax and Steiner 2013). For non-stationary time series, the Fernandez and Litterman models are more optimal (Chamberlin 2010; Sax and Steiner 2013). Please note that the Denton and Denton-Cholette disaggregation methods are only used when zero or one predictor variables are available. Since there is more than one environmental variable related to dengue, these methods were not explored.

The augmented Dickey-Fuller test was run on the mean dengue time series to determine if the series was stationary. The null hypothesis of the augmented Dickey-Fuller test is that the time series is non-stationary. If the augmented Dickey-Fuller statistic is significant at α=0.05, then the time series is considered stationary. For this study, the augmented Dickey-Fuller statistic was -3.5237 and had a p-value of 0.046. Therefore, the mean dengue time series is stationary and the appropriate disaggregation method is Chow-Lin.

After stationarity had been determined, the mean dengue time series was aggregated from monthly cases to annual cases. This was done by summing all of the dengue cases within one year. Using Chow-Lin, a model was developed for every possible combination of environmental variables. A RMSE was calculated for each model and significance levels

for each model's coefficients were recorded. The model with the lowest RMSE, most significant coefficients, and most visually pleasing graph was selected as the optimal model for disaggregation. This model was then run on the annually aggregated dengue cases for each individual province of Cambodia and an average RMSE was calculated. All calculations and model development was completed in Mathwork's Matlab.

4.3 Results and discussion

4.3.1 Results

The R-coefficients of the correlations between the mean dengue time series and the mean environmental predictors are listed in Table 7. The variables with an absolute R-coefficient greater than 0.3 are as follows: NDVI, precipitation sum, precipitation mean, and LST night maximum. Of these variables, precipitation sum and precipitation mean were highly correlated to each other. Due to this, precipitation mean was not kept as an environmental predictor since precipitation sum was more correlated to dengue.

Table 7. R-coefficient between the mean dengue time series and the mean environmental variables time series.

Environmental Variable	R-Coefficient
Normalized Difference Vegetation Index (NDVI)	-0.35
Precipitation Sum	0.48
Precipitation Mean	0.47
Precipitation Sum Lagged	0.25
Precipitation Mean Lagged	0.25
Land Surface Temperature (LST) Day Mean	0
LST Day Minimum	0.13
LST Day Maximum	-0.09
LST Night Mean	-0.29
LST Night Minimum	-0.16
LST Night Maximum	-0.36

A total of seven models, plus a null model that evenly distributed dengue cases throughout the year, were created from every possible combination of dengue correlated environmental variables. Of these models, only three models, 1, 4, and 6, had significant environmental coefficients (Table 8).

This significance occurred in the models that only used NDVI, precipitation sum, and land surface temperature (LST) night maximum as variables. Of these three models with significant coefficients, model 4, the model with the precipitation sum predictor, had the lowest RMSE.

Table 8. Model name, variables, the p-value of each model's coefficients and the RMSE. A * denotes variables that have significant coefficients at α=0.05.

Model Name	Model Variables	Environmental Variable Significance	RMSE
Null	Null	NaN	12.210
1	NDVI*	0.0236	13.833
2	NDVI Precipitation Sum	0.980 0.896	10.71990
3	NDVI Precipitation Sum LST Night Maximum	0.679 0.978 0.553	18.88614
4	Precipitation Sum*	0.0233	10.760
5	Precipitation Sum LST Night Maximum	0.314 0.431	24.16393
6	LST Night Maximum*	0.0312	12.80885
7	LST Night Maximum NDVI	0.444 0.336	20.54409

The graphs of the models provide a better picture of how well the mean dengue time series has been disaggregated (Figure 30). Of the three models with significant coefficients, model 4 does the best job of capturing the general seasonality of dengue. Model 6, in contrast, closely mirrors the null model and does not capture the peaks and valleys of the dengue time series. Both the null model and model 6 were only able to capture the step increases and decreases in dengue cases from year to year. Lastly, model 1, which only uses NDVI as a predictor variable, predicts a large number of dengue cases when a lower number occurred and predicts periods low disease activity when extreme dengue activity happened. The precipitation model (model 4), however, was able to capture every step change, peak and valley in the dengue time series, and only missed the extreme amplitude of cases in 2007.

Figure 30. Graph of (a) Mean Cambodian dengue time series (b) Null Model, (c) Model 1, (d) Model 4 and (e) Model 6.
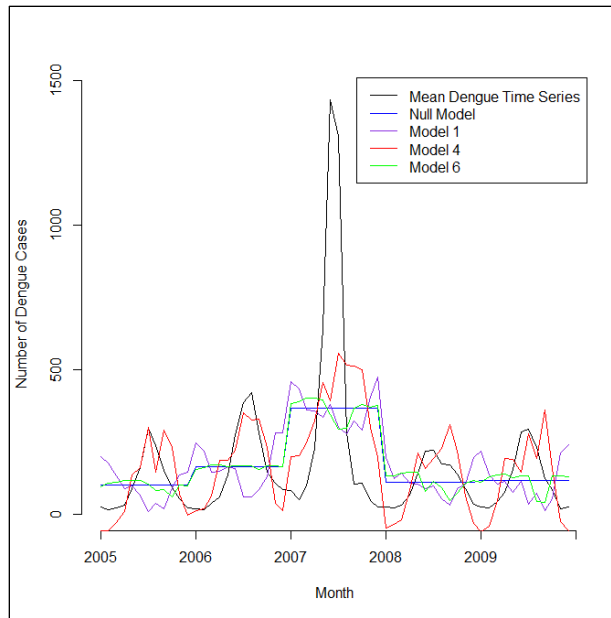
Based on its low RMSE score, significant predictor variable, and its ability to capture the general seasonality of the mean Cambodian time series, the Chow-Lin disaggregation method with the monthly time series of an area's total precipitation was applied to each individual province in Cambodia. The mean RMSE among the disaggregation of all of these time series was 12.925. The disaggregation of four Cambodian provinces' dengue time series are in Figure 31.

Figure 31. Graphs of the dengue time series and annual dengue disaggregated using model 4 for the Cambodian provinces of (a) Battambang (RMSE = 3.67), (b) Banteay Meanchey (RMSE = 18.53367), (c) Kandal (RMSE = 31.1175), (d) Siem Reap (RMSE = 20.53035).
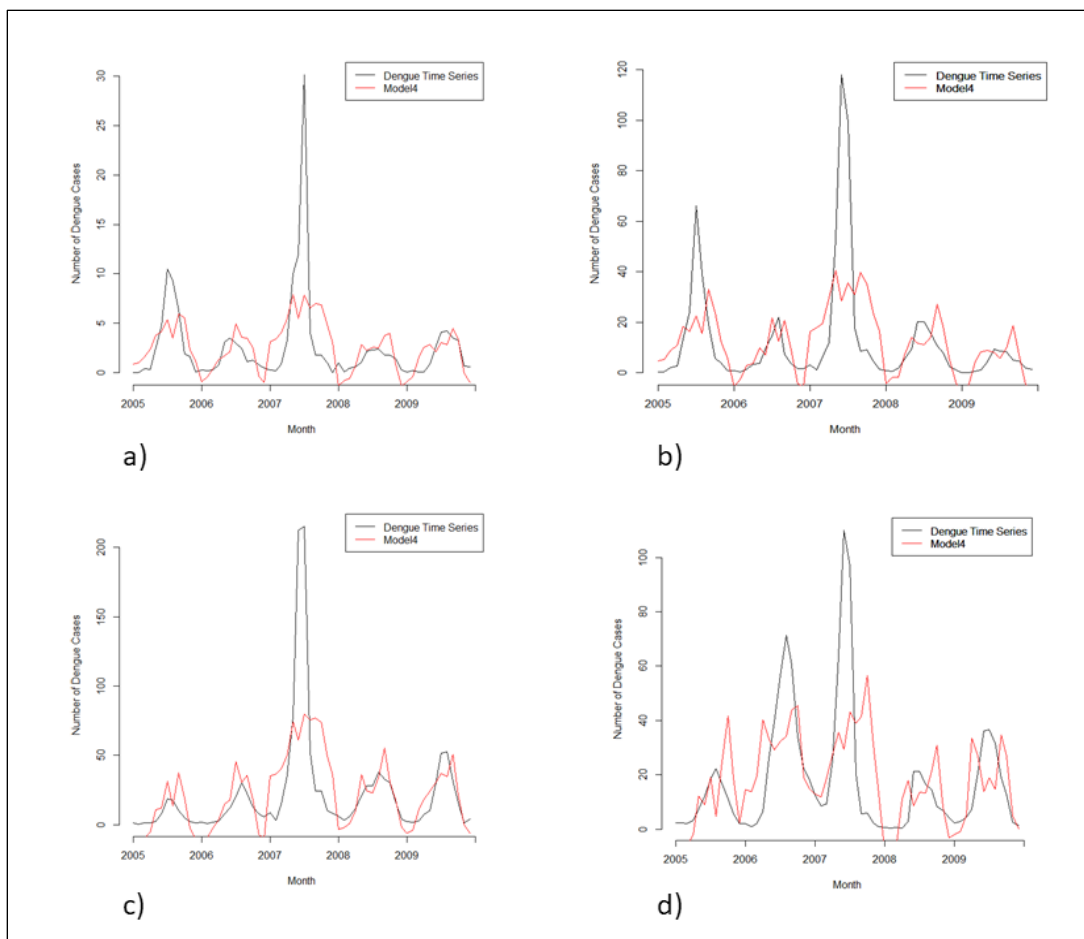
4.3.2 Discussion

The variables that were most significantly correlated to Cambodian dengue were NDVI, precipitation, and maximum night LST. The temporal disaggregation models created with these predictors did not all achieve RMSEs lower than the null model and many produced variable coefficients that were not significant. Of all the models tested, model 4, which only used precipitation data as a predictor, performed the best. Model 4 had a lower RMSE than the null model, the coefficient of the precipitation variable was significant at $\alpha=0.05$, and visually disaggregated the dengue time series better than any other model. Thus, temporal disaggregation of dengue cases can be achieved using the Chow-Lin disaggregation method with the total sum of an area's precipitation as a predictor variable. More research is need in this area to see if the same technique can be used for other types of vector and non-vector-borne diseases.

5   Mosquito-Borne Disease Simulation

5.1   Literature review

Traditionally, density-based mathematical models of mosquito-borne disease divide the human and vector populations into "containers:" susceptible humans (S) are healthy but could potentially acquire the disease, exposed humans (E) have been bitten by an infected mosquito and undergoing incubation, infected humans (I) are experiencing the symptoms of the disease, and those that recover (R) are both healthy again as well as immune to re-infection. The total population, N, is merely the sum of these containers. Similarly, the adult mosquito population is made up of mosquitoes that are susceptible to infection, s, and those infected, i. Because of the short life-cycle of mosquitoes and the biomechanics of most mosquito-borne diseases, it's assumed that the mosquito dies before ever recovering.

The progression of mosquito-borne diseases in these models require continuous interaction between mosquitoes and humans – a susceptible human is bitten by an infected mosquito or a susceptible mosquito bites an infected human. To model this interaction, it is generally assumed that the populations are well-mixed – each mosquito in the population has an equal probability of interacting with each human, and vice versa. This makes the number of newly exposed human and mosquito individuals merely a matter of dimensional analysis. Let a be the number of bites a mosquito gives in a day, b be the probability that an individual mosquito bite leads to a susceptible human becoming "exposed," and define $\lambda_{S \to E}$ to be the "exposure rate," the rate at which susceptible humans become exposed to the disease. Since the probability of each member of the human and vector populations are equal, the exposure rate can be written simply as

$$\lambda_{S \to E} = \frac{ab\,i}{N} \tag{1}$$

A standard disease model of this type results in two phases: (1) disease free equilibrium, where the death of infected mosquitos outpaces the ability for them to infect new humans and (2) persistent presence of disease, where the temporal dynamics of the disease progression determine stable proportions of the human and mosquito populations that are infected (Shah and Gupta 2013).

While most vector-borne disease simulation research has been completed in rural areas, recent work has been given to modeling areas of high population density, such as a dense urban city. In a dense city, physical barriers to interaction and the sporadic location of vector breeding sites makes it so the human and vector populations interact heterogeneously. To address this shortcoming, Kong et al. (2018) modeled the interactions of mosquito and humans to be a random variable, with some humans more likely to interact with mosquitoes than others (Kong et al. 2018). They derived the exposure rate to have the form

$$\lambda_{S \to E} = k \ln \left(1 + \frac{ab\,i}{k\,N}\right) \tag{2}$$

where k is a heterogeneity parameter and ln is the natural logarithm. For very small values of k,

$$\lim_{k \to 0} k \, ln \left(1 + \frac{ab\,i}{kN}\right) = 0 \tag{3}$$

which represents the mosquito and human populations being completely quarantined from each other. Conversely, for large values of k,

$$\lim_{k \to \infty} k \, ln \left(1 + \frac{ab\,i}{kN}\right) = \frac{ab\,i}{N} \tag{4}$$

which is what is given under a well-mixed assumption .

While all of these models incorporate the heterogeneous interaction of human and mosquito populations, they do not incorporate the full life cycle of the mosquito. Specifically, the models do not include a quiescent period, where, after achieving a blood meal, a mosquito ceases host-seeking and finds an appropriate habitat to lay their eggs. This lack of temporal variation in host-seeking could be ignoring vital predictions of mosquito-borne disease.

5.2 Geoanalytic data and methods

The focus of the developed mathematical model was on the interaction of two different populations: human and mosquito. A container approach, the binning of a population into different categories based on certain characteristics, was applied to both human and mosquito populations to model how a disease moves within and between each population. The

containers for the human and mosquito populations are very similar, but different based on key differences in population life cycles.

5.2.1 Data

For the human population, the population is divided into four different containers: susceptible (S), exposed (E), infected (I), and recovered (R). Susceptible individuals are not immune to the disease and have yet to be exposed. Exposed individuals have been bitten enough by one or more infected mosquitoes for the disease to begin to incubate. After a sufficient period of time incubating, individuals become sick and are moved into the infected container. It is during this period of time that mosquitoes who are not infected themselves can bite the infected individuals and the disease can transfer from the human population back to the mosquito population. Finally, after a period of time spent infected, the individual recovers and spends a period of time temporarily immune from infection before being susceptible again to exposure.

Similar to the human population, the mosquito population is divided into containers representing their full life cycle: egg, larvae, pupae, and adult. The mosquito population is further divided to reflect two separate dynamics: the progression of the disease in the mosquito and the mosquito's feeding cycle. The first part progresses similarly to the human population, being susceptible, exposed, and infected (mosquitoes do not recover). However, adult female mosquitoes seek out a blood meal for the purpose of providing energy to lay eggs, and after biting enough individuals to acquire a sufficient blood meal, go off to lay eggs and are no longer biting. This means that the three containers are further divided into active and quiescent, meaning the adult population has six total containers: susceptible active (s_a), and quiescent (s_q), exposed active (e_a), and quiescent (e_q), and infected active (i_a) and quiescent (i_q)[*]. Each stage of the mosquito's pre-adult life is considered to last for 2-5 days, and as an adult, the female mosquitoes actively search for blood meals and then spend 2-3 days laying eggs. Mosquitos total lifespan is 3-4 weeks, meaning that female mosquitoes will on average achieve 3-4 blood meals before dying.

---

[*] Without loss of generalization, an incubation period is not included for the mosquitos, considering mosquitoes to be infected immediately after digestion of an infected blood meal.

5.2.2 Methods

The probability dynamics of interaction driven state transitions (Kong et al 2016) is now considered. Let X_i be the number of effective contacts[*] between an individual in one population and the ith individual in the other population. It is assumed that for each individual, this number is Poisson distributed $\pi(\theta_i)$, with θ_i being the mean number of effective contacts for the ith individual. If the system was well mixed, θ_i is uniform, with every member of both populations having equal probability to interact with each other. However, θ_i is considered to be a random variable itself with a Gamma distribution, $\Gamma(\theta_i, \{k, m\})$ where k is a shape parameter and m is the rate parameter. The marginal distribution for X_i is therefore

$$P(X_i = x) = \binom{x + k - 1}{x} \left(\frac{m}{m+1}\right)^k \left(\frac{1}{m+1}\right)^x \tag{5}$$

which is the probability density of a negative binomial distribution with mean k/m and variance $k(1 + m)/m^2$. The probability of having at least one effective contact is given by

$$p = 1 - P(X_i = 0) = 1 - \left(1 + \frac{1}{m}\right)^{-k} \tag{6}$$

Identifying Equation 2 with the relationship between risk and rate to be $risk = 1 - e^{rate}$, the transition rate can be derived to

$$\lambda(t) = k \ln\left(1 + \frac{1}{m}\right) \tag{7}$$

For the human population, the only transition of interest is caused by a blood meal being acquired by an infected mosquito -- blood meals by uninfected mosquitoes have no effect on the susceptible human population. Therefore, an "effective contact" is identified for the human population to be the blood meal acquisition by a member of the active infected mosquito population, i_q. If the mosquitos have a bite rate, a, with the probability that bite from an infected mosquito results in an infection is b, then the mean number of infected bites per person is $ab\, i_a/kN$.

---

[*] Here, "effective contact" means a contact that leads to a state transition.

Recalling that the mean of the Gamma distribution is $\frac{k}{m}$, the susceptible to exposed transition rate can be written as

$$\lambda_{S \to E}(t) = k_H ln\left(1 + \frac{ab\, i_a(t)}{kN}\right) \tag{8}$$

where k_H is a heterogeneity parameter for the interactions of the susceptible human population and the active infected mosquito population.

For the mosquito population, the definition of an "effective contact" is more complex. The full life cycle of the mosquito population would need to be modeled, including the quiescent periods of time adult female mosquitos spend laying eggs and not biting individuals. So, an effective contact means an active mosquito successfully achieving a blood meal, not just by those that achieve infected blood meals. For active mosquitos, the average number of effective contacts then is the product of the number of bites each mosquito takes and the probability a single bite will result in a blood meal[*]. Identifying the probability bite results in a sufficient blood meal as c, the transition rate of mosquitos from active blood meal seeking to quiescent is given by

$$\lambda_{a \to q}(t) = k_v\, ln\left(1 + \frac{ac}{k_v}\right) \tag{9}$$

where k_v is the parameter controlling the heterogeneity of blood meal acquisition from the entire human population. For the susceptible mosquito population, there are two types of transitions that can occur: from s_a to i_q via interaction with the infected proportion of the human population, $\hat{I} = I/N$, or from s_a to s_q via interactions with the uninfected proportion of the human population, $1 - I$. This means that the probability of these transfers is given by

$$\lambda_{s_a \to s_q} = \lambda_{a \to q}(t)\hat{I}(t) = k_v ln\left(1 + \frac{ac}{k_v}\right)\hat{I}(t) \tag{10}$$

$$\lambda_{s_a \to i_q} = \lambda_{a \to q}(t)\left(1 - \hat{I}(t)\right) = k_v ln\left(1 + \frac{ac}{k_v}\right)\left(1 - \hat{I}(t)\right) \tag{11}$$

---

[*] Or, equivalently, how many bites are necessary for a blood meal

In general, the heterogeneity factor, k, doesn't need to be equal between these two populations, but here it will be considered $k_V = k_H = k$ for simplicity.

5.2.3 Human and vector master equations

Now that the dynamic state transition rates for the human and vector populations have been derived, the full master equations can be written for the stochastic model. For the human population, we have

$$\partial_t S(t) = -k \, ln \left(1 + \frac{ab \, i_a(t)}{k \, N}\right) S(t) + \lambda_{R \to S} R(t) \tag{12}$$

$$\partial_t E(t) = -\lambda_{E \to I} E(t) + k \, ln \left(1 + \frac{ab \, i_a(t)}{k \, N}\right) S(t) \tag{13}$$

$$\partial_t I(t) = -\lambda_{I \to R} I(t) + \lambda_{E \to I} E(t) \tag{14}$$

$$\partial_t R(t) = -\lambda_{R \to S} R(t) + \lambda_{I \to R} E(t) \tag{15}$$

where $\lambda_{R \to S}$ is the immunity loss rate, $\lambda_{E \to I}$ is the incubation rate, and $\lambda_{I \to R}$ is the recovery rate. To write the master equation for the immature mosquito population, first note that eggs are laid by quiescent adults. Identifying the number of eggs as N_{eggs} and the transition from quiescent to active as $\lambda_{q \to a}$, the immature master equations will be written as

$$\partial_t n_e(t) = -\lambda_{e \to l} n_e(t) + \lambda_{q \to a} N_{eggs} \left(s_q(t) + i_q(t)\right) - \mu_e n_e(t) \tag{16}$$

$$\partial_t n_l(t) = -\lambda_{l \to p} n_l(t) + \lambda_{e \to l} n_e(t) - \mu_l n_l(t) \tag{17}$$

$$\partial_t n_p(t) = -\lambda_{p \to adult} n_p(t) + \lambda_{l \to p} n_l(t) - \mu_p n_p(t) \tag{18}$$

$\lambda_{p \to adult}$ is written to identify the rate of pupae maturing into adults, both male and female. Assuming the population is evenly split, only half of those mosquitos will be added to the active susceptible mosquito containers. With this last piece in place, the adult mosquito master equations can be written as

$$\partial_t s_a(t) = -k \, ln \left(1 + \frac{ac}{k}\right) s_a(t) + \lambda_{q \to a} s_q(t) - \mu_a s_a(t) \tag{19}$$

$$\partial_t s_q(t) = \lambda_{q \to a} s_q(t) + k \, ln \left(1 + \frac{ac}{k}\right) (1 - \hat{I}) s_a(t) - \mu_q s_q(t) \tag{20}$$

$$\partial_t i_a(t) = -k \ ln \left(1 + \frac{ac}{k}\right) i_a(t) + \lambda_{q \to a} i_q(t) - \mu_a i_a(t) \tag{21}$$

$$\partial_t i_q(t) = -\lambda_{q \to a} i_q(t) + k \ ln \left(1 + \frac{ac}{k}\right) \hat{I} s_a(t) + k \ ln \left(1 + \frac{ac}{k}\right) i_a(t) - \mu_q i_q(t) \tag{22}$$

To explore the dynamics of this model, these equations were implemented into a fully stochastic model in Java using a human population of $N = 3000$ and an initial mosquito population of $n = 300$. Containers were initialized with a random proportion of the overall population and simulations were run for 10 years of simulation time to reach dynamic steady states. So, as to keep temporal parameters identical across the phase space sweep, the overall mosquito population was controlled by modifying the adult death rate. Other parameters were kept constant (see Table 9 for values).

Table 9. Parameter values.

Parameter	Description	Value
k	Heterogeneity Factor	Varies (day$^{-1)}$
a	Mosquito bite rate	3 day^{-1}
b	Probability of exposure per bite, human	0.3
c	Probability of exposure per bite, vector	0.3
N_{eggs}	Number of eggs laid at once	300
$\lambda_{E \to I}$	Intrinsic incubation rate	1/12 day^{-1}
$\lambda_{I \to R}$	Recovery rate	1/7 day^{-1}
$\lambda_{R \to S}$	Immunity Loss Rate	1/225 day^{-1}
$\lambda_{q \to a}$	Re-activation rate	1/3 day^{-1}
μ_a, μ_q	Adult mosquito death rates	1/28 day^{-1}
μ_p, μ_l, μ_p	Immature mosquito death rates	1/10 day^{-1}
$\lambda_{e \to l}$	Egg to larvae transition rate	1/3 day^{-1}
$\lambda_{l \to p}$	Larvae to pupae transition rate	1/7 day^{-1}
$\lambda_{p \to adult}$	Pupae to adult transition rate	1/3 day^{-1}

5.3 Results and discussion

5.3.1 Results

The implementation of the described container model via simulation demonstrated two expected limiting dynamic phases (Figure 32a): disease

free equilibrium (for small k, which represents quarantine) and persistent disease (for large k, representing complete mixture). Additionally, it is found that there exists a region in between these limiting cases where the system undergoes aperiodic outbreaks and large swings in the population proportions. This lack of temporal stability is evidence of stochastic resonance, wherein the mosquito population behaves like a colored noise source to the human population. This causes small fluctuations in the populations of the mosquito containers to translate into large population swings in the human containers. This dynamic phase is made more apparent by considering the relative deviation of the system parameters, i.e. the standard deviation divided by the average. As seen in Figure 33, both the mean value and the relative variance demonstrate the aperiodic phase.

Further, an order of magnitude increase in mosquito population causes this dynamic phase to disappear as the more mosquitos present in a region per person, the more likely the systems dynamics will mimic a well-mixed system. Thus, the aperiodic outbreak region appears as a balance of disease presence and population screening.

Figure 32. (a) Time trajectories of the susceptible (S), exposed (E), infected (I), and recovered (R) populations for the three states of the epidemic model: disease free equilibrium, aperiodic outbreak, and persistent recovery. (b) Mean population proportion values as a function of the heterogeneity parameter, k. Each point is a time average of 10 years.
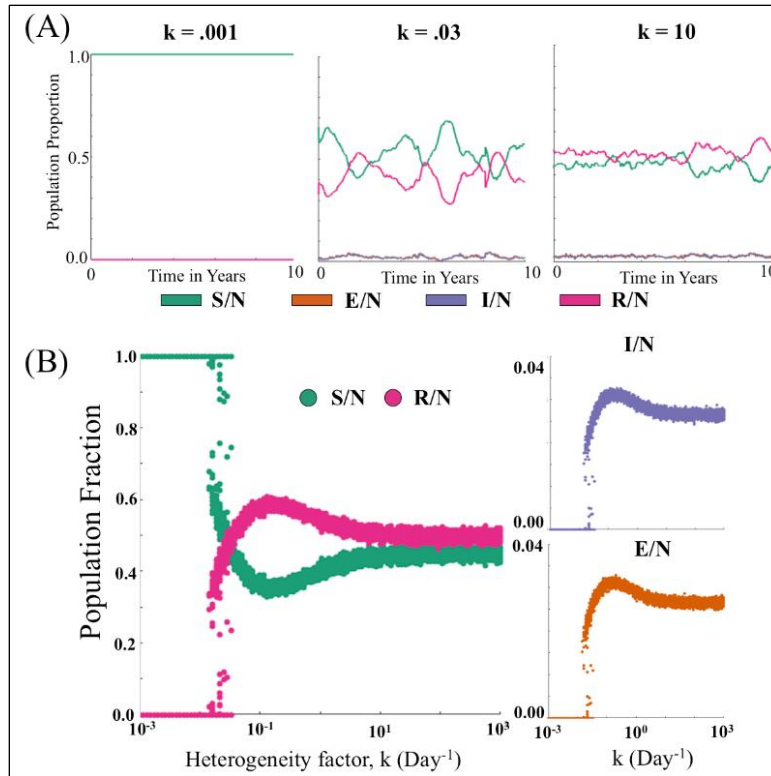
Figure 33. (Left) average and (right) relative variance of susceptible proportion as a function of both heterogeneity parameter and human-vector population proportion. Values were calculated over time and then average across 100 simulations.
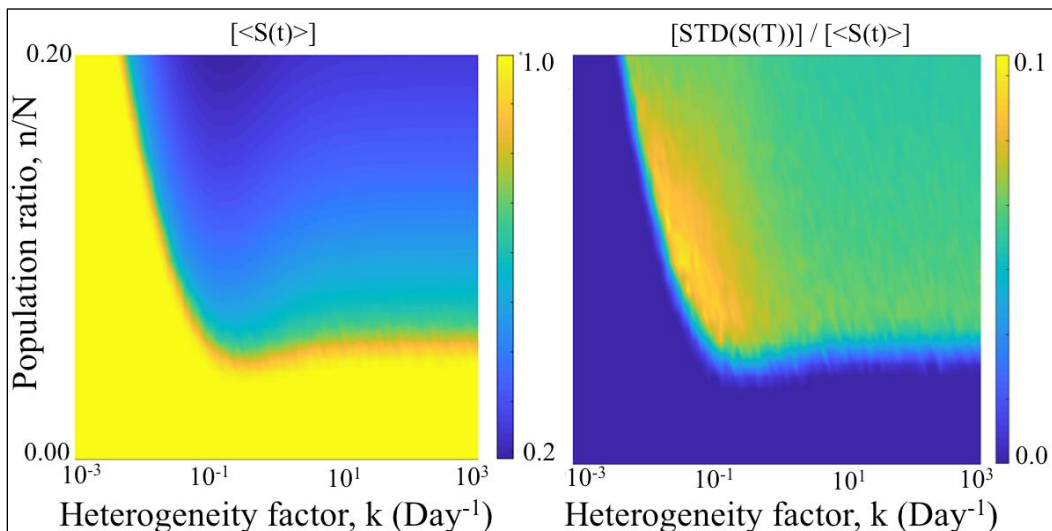
5.3.2   Discussion

In this report, a container-based mosquito-borne disease model that includes heterogeneous interactions between the host-vector population and the oscillatory behavior of mosquito activity due to temporary cessation of blood meal seeking has been motivated to lay eggs. These two additions lead to a model with three dynamic phases: disease-free equilibrium, aperiodic outbreak, and persistent disease presence. The aperiodic outbreak phase is a result of stochastic fluctuations in infected mosquito populations propagating to the human population via non-linear stochastic resonance, leading to high relative variance of infection cases. This resonance disappears with overall mosquito population increases as well, and has probabilistic qualities that cement the aperiodic outbreak stage is a distinct phase.

These results suggest that interventions to screen human populations from mosquitoes or reduce the overall mosquito population (to prevent outbreak) can cause increasing temporal volatility, leading a population to be more vulnerable to full on outbreaks.

6   Project Summary

6.1   Army mission support outcomes

Upon successful closure of the NET-CMO research project, innovative geoenabling tools and methodologies focused on enhancing mission analysis for Army CMO were developed. Metrics for success set forth by GRL management that align with Army Force Health Protection (FHP) doctrine, emerging Army Modernization priorities, and Joint Operational Planning requirements were the focus of these enhancements. Project metrics fundamentally tracked the following narrative: "develop critical enhancements to the suite of methods and tools supporting mission analysis for civil-military operations to enhance stability and mitigate health threats-risks to the civilian population."

Specified operational planning tasks within the Army mission support area of FHP and the Multi Domain Common Operating Picture (COP) have a focus on potential threats to conditions of soldier's health. The require-ments stated to address these threats include protection of warfighting function, preventive medicine measures, and unit health assessments of combat and operational stress. The evaluation of environmental influences (i.e. regions of high risk of vector-borne diseases) are also necessary to maintain force readiness and for development of COAs. Attainment of a sufficient level of situational awareness within a region of interest that are prone to disease is important for Commanders to provide timely assessment of these health threats-risks to personnel under their command.

The products and methodologies developed by NET-CMO specifically address the identification of regions prone to vector-borne disease. This is vital for FHP, force readiness, and operational planning. Additionally, tools and techniques were also developed to support the identification of these disease prone regions. The project's noteworthy outcomes and the Army requirements they addressed are listed below.

- Army Doctrine: Protection Warfighting Function
    - o   Commanders and unit leaders are responsible for protecting and preserving Army personnel against injury or loss that may result from risks of vector-borne diseases. Army policy stated in AR 11-35 requires Commanders to provide timely assessment of these risks to personnel under their command (Joint Publication 4-02.8)

* NET-CMO Outcome

~ Disease Simulation process
~ Health Threat-Risk products

- Army CFT Priorities: Common Operating Environment

  o Focus on addressing near-term capability gaps to enable a shared geo-spatial foundation across networked computing environments and warfighting functions. With primary emphasis upon the handling and manipulation of complex, big data sources for geoenabled mission planning and analysis (Army Modernization Strategy)[*]

    * NET-CMO Outcomes

      ~ Uniform Pixel Sizing tools (Semivariogram and LSD)
      ~ Temporal Data Disaggregation methods
      ~ Spatial Downscaling methods

- Army Campaign Plan: Joint Operation Planning Process

  o Preparation of COA or COP (health related) to conduct a map-based environmental assessment of situational mission space to evaluate associated Soldier/Squad health threats-risks. Product Visualization of vector-borne health influences to generate a geospatial overlay of operational AOI to enable strategic planning initiatives (Joint C2 CDD)[†]

    * NET-CMO Outcomes

      ~ Health Threat-Risk products

6.2 Conceptual workflow

The downrange goal of the *collective* NET-CMO project components is to assimilate the various tools and methodologies into a connected operational workflow (Figure 34). Providing a user (i.e. analyst, planner, decision-maker) with an enhanced geoprocessing mechanism to transform complex, big geospatial data sources into situationally aware, actionable

---

[*] Army Modernization Strategy (Apr 2018), Cross Functional Team (CFT) Priorities: Army Futures Command (AFC).

[†] Joint C2 CDD (Apr 2013), sections 1.2.2 Concept Development Framing of the visualization capabilities needed to enable solutions that address 1.2.1 Strategic Guidance, 1.2.4 Global Force Management (GFM) Support to Planning, and 1.2.5 Readiness Support.

end-products. This would *geo*enable the very time-consuming and computationally intensive process of generating outcomes for predicting and understanding health threat-risk related factors impacting Soldiers.

Figure 34. Functional concept mapping of *collective* NET-CMO workflow.
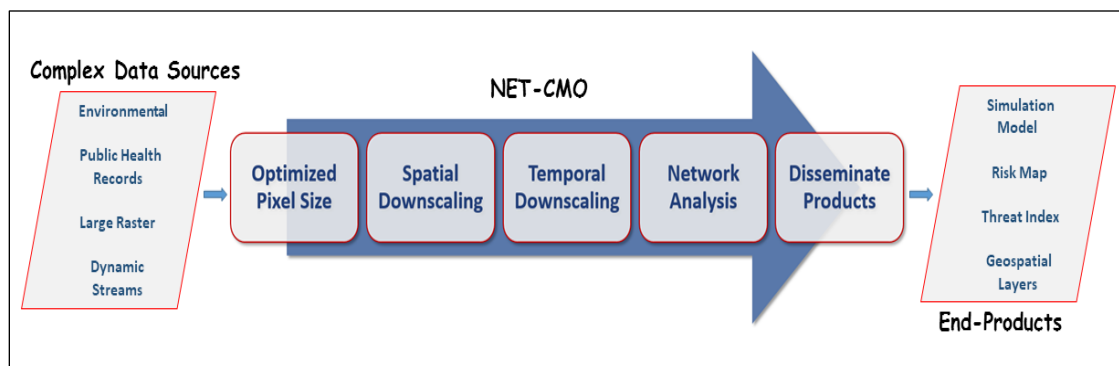
6.3 Stakeholder engagements

6.3.1 U.S. Army Walter Reed Biosystematics Unit (WRBU)

Using RF and environmental raster datasets, NET-CMO produced monthly maps of potential *aedes aegypti* activity in the southwest United States to WRBU. WRBU included these maps in a monthly report on mosquito risk sent to the Global Emerging Infections Surveillance (GEIS). This information is used by U.S. troops stationed along the Southern U.S. border (Figure 35).

Figure 35. Mosquito activity risk map for SW United States.
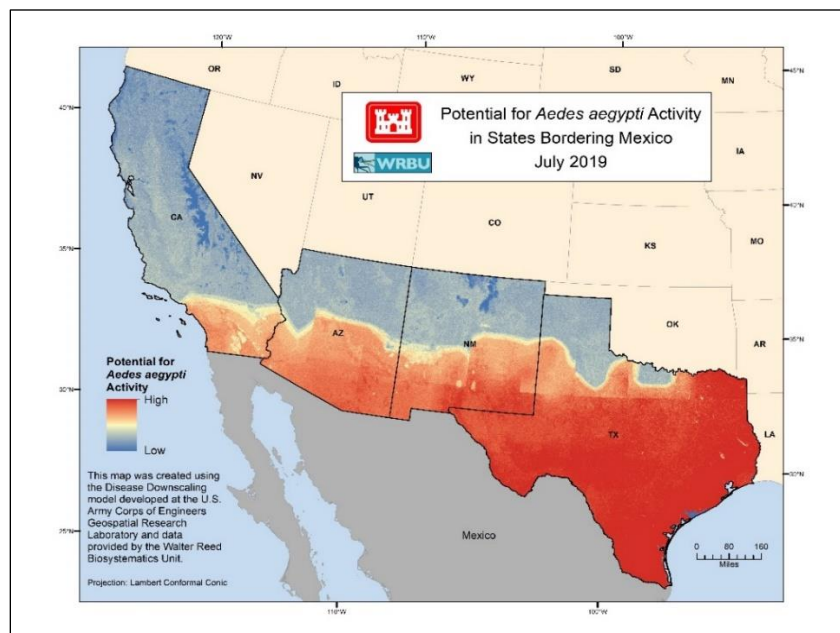
6.3.2 Army Public Health Command (APHC): Tick-Borne Diseases Lab (TBDL)

Using the machine learning technique maximum entropy, NET-CMO developed presence maps of the tick *ixodes scapularis* (Figure 36). APHC TBDL representatives requested this work to help them identify regional hotspots for tick-borne disease activity. The identification of these regions is useful for determining training regions. Preventing soldiers from contracting a tick-borne disease, such as Lyme disease, during training is vital in order for the U.S. Army to maintain force readiness. Reference *Appendix B* for a summary of this ERDC technical note.

Figure 36. Ixodes scapularis disease map for NE United States.
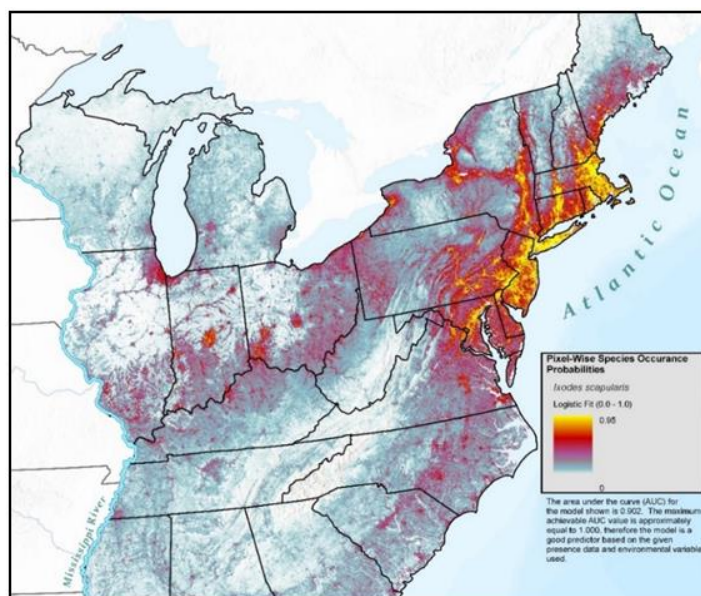
6.3.3 Army Force Health Protection and Preventive Medicine: MEDDAC-Korea/65th Medical Brigade

NET-CMO researchers generated a map product for evaluation of mosquito larvae habitat in South Korea (Figure 37). The creation of this map will be detailed in an ERDC technical note, 'Understanding the Disease Vector Operational Environment by Predicting Presence of *Anopheles* Mosquito Breeding Sites Using Maximum Entropy Modeling: Maxent Software Platform' by Susan Lyon. A summary of this ERDC technical note is in *Appendix C*.

Figure 37. Potential breeding sites of *Anopheles* mosquitos shown in purple.
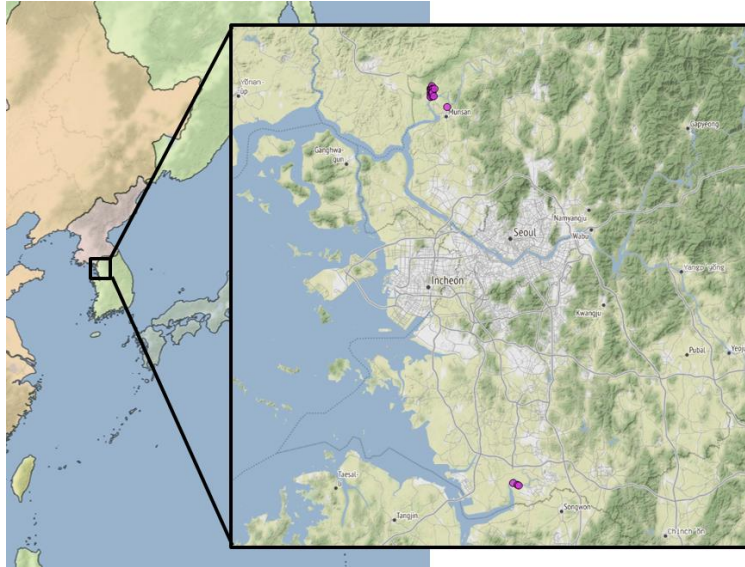
6.4 Potential for transition

6.4.1 Map based planning system (MBPS)

Significant effort has been directed towards transitioning NET-CMO technologies and VAST disease simulation workflow into the ERDC-GRE 6.2-6.3 MBPS operational planning framework. The R&D team prepared detailed Feature Requirement Specification (FRS) documentation to include User Interface (UI) design requirements, component technical descriptions, and development of notional workflows to enable transition to MBPS and FY20 successor program GEOD (Figure 38).

Figure 38. Notional UI concept for VAST Workflow within MBPS.
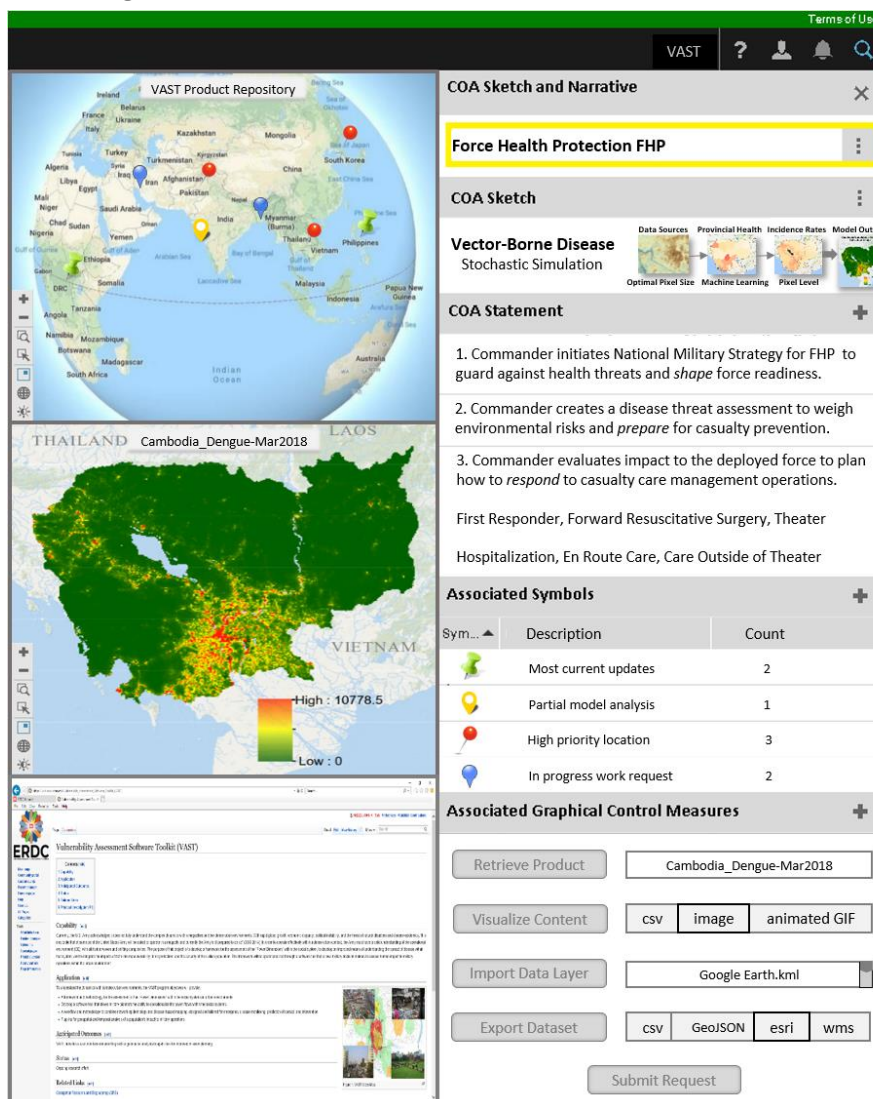
6.4.2 Harris ENVI

ERDC-GRL has established a Cooperative Research And Development Agreement (CRADA) with the creator of the ENVI software suite, L3Harris Geospatial Solutions ("Harris"), to exchange R&D work and technologies.[*] As part of this CRADA, ERDC-GRL will transition the ArcMap semivariogram model (Figure 39) to Harris to further develop within the ENVI framework. The overall goal of this transition is the implementation

---

[*] ENVI: Environment for Visualizing Images. L3Harris Geospatial Solutions.
https://www.harrisgeospatial.com/Software-Technology/ENVI

of the semivariogram tool Distributed Common Ground System–Army (DCGS-A).

Figure 39. Uniform pixel size ArcMap model builder model.
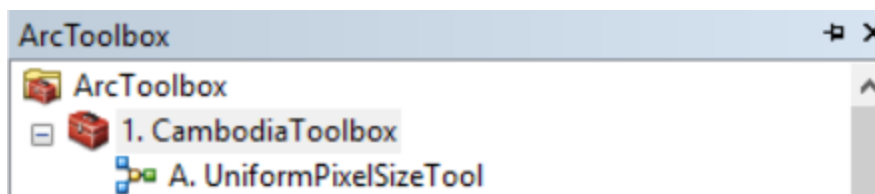
6.5 Next generation of NET-CMO

6.5.1 Systemic integration

Building upon R&D project marketing initiatives (*Appendix A*) and stated outcomes with specified customer requirements, the future potential of NET-CMO related work efforts is very positive. ERDC-GRE program areas such as Geo-Enabled Operational Design (GEOD) emphasizing geospatial approaches to Army mission command and analysis will be able to further leverage and transition these NET-CMO workflows, tools, and methodologies into their program-systems of record.

Other mission support organizations that may benefit from further development and collaboration of geoenabled health threat-risk assessments include the U.S. Army Medical Command MEDCOM Public Health Center (USAPHC). The integration of NET-CMO work efforts to assimilate their technologies into operational command level geoenabled systems would enhance Army readiness posture by identifying and assessing current and emerging health threats.

6.5.2 Product space

Continuation of the NET-CMO work area would focus on expanding the portfolio of health threat-risk type evaluations of potential diseases along with more rigorous quantification (i.e. statistical analysis) of environmental induced stressors impacting Army readiness, Soldiers well-being, and civil stability mitigation efforts.

Example 1-Spatial and temporal prediction of tick-borne diseases: There is potential to generalize the spatio-temporal modeling of mosquito-spread disease to diseases spread by ticks. This would require a marked amount of model complexity because ticks have a much longer life span,

sometimes living over 1-2 years. During this lifespan, they survive off interactions with a variety of other animals, so the model would have to incorporate several other species. Thus, modeling efforts would incorporate other population interactions and zoological lifecycles, and the environmental factors affecting them. Inverting this relationship, the prevalence of tick-borne disease in a region would also be used to infer the local population of otherwise unaccounted for animal populations.

Example 2-Risk to the Warfighter of non-vector-borne diseases: The expansion of mapping and being predictive of disease capabilities beyond vector-borne diseases would be beneficial to the Army. Current methodologies for predicting the spread of infectious diseases focus on human-to-human interactions, without considering the role the physical environment plays. Inclusion of environmental parameters into non-vector-borne infectious disease models would allow for a strong identification of regions prone to disease and the ability to better protect U.S. soldiers.

Example 3-Optimal Uniform Pixel Size for Categorical Data: Continuous raster and gridded numerical pixel data are important to the Army; however, the Army also works heavily in the domain of gridded categorical data. This includes land cover mapping involving classifying imagery into categorical land cover classes. The same questions arise when working with multiple categorical rasters of disparate pixel sizes of what pixel size is the optimal size. Continuation of NET-CMO could include development of a uniform pixel technique with categorical data.

6.6   Concluding remarks

The NET-CMO project took 1 year of fast R&D work effort leveraging previous work that had been completed within the ERDC 6.2 project VAST. Given this accelerated time frame, in coincidence with specified metrics established by management primarily focused upon proof-of-concept activities to assess, test, validate (i.e. statistically), and establish transition paths for enhanced geoenabling technologies and capabilities, the project was highly successful. As proposed in the initial NET-CMO Project Management Plan (PMP), established exit criteria and team performance metrics were effectively accomplished.

The culmination of these new geoanalytic methods, geoprocessing tools and, value-added end-products will have a positive impact to further

support FHP domain (i.e. disease threats-risks) of military operations. To include protecting soldiers and assets through the ability to identify vector-borne disease, understand stability vulnerabilities in a region, and mitigate threats to a civilian population. NET-CMO deliverables will provide the U.S. Army with an increased situational awareness capability and provides planners with the resources needed to help resolve these potential health threats-risks.

References

Altamiranda-Saavedra, M., X. Porcasi, C. M. Scavuzzo, and M. M. Correa. 2018. "Downscaling incidence risk mapping for a Colombian malaria endemic region." *Tropical Medicine and International Health* 23:10, 1101-1109. doi: 10.1111/tmi.13128

Atkinson, P. M., and P. J. Curran. 1997 "Choosing an appropriate spatial resolution for remote sensing investigations." *Photogrammetric Eng Remote Sensing* 63(12):1345-1351.

Beale, L., J. J. Abellan, S. Hodgson, and L. Jarup. 2008. "Methodologic issues and approaches to spatial epidemiology." *Environmental Health Perspect*, 116: 1105-1110.

Brady, O. J., P. W. Gething, S. Bhatt, J. P. Messina, J. S. Brownstein, A. G. Hoen, C. L. Moyes, A. W. Farlow, T. W. Scott, and S. I. Hay. 2012. "Refining the global spatial limits of dengue virus transmission by evidence-based consensus." *PLoS Neglected Tropical Diseases* 6:8 e1760. doi: 10.1371/journal.pntd.0001760

Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5-32. doi: 10.1023/A:1010933404324

Chamberlin, G. 2010. "Temporal disaggregation." *Economic and Labour Market Review*, Nov. 2010: 106-121.

Chitnis, N., D. Hardy, and T. Smith. 2012. "A periodically-forced mathematical model for the seasonal dynamics of malaria in mosquitoes." *Bull Math Biol*, 74: 1098-1124.

Cohen, W. B., T. A. Spies, and G. A. Bradshaw. 1990. "Semivariograms of digital imagery for analysis of conifer canopy structure." *Remote Sensing Env* 34(3):167-178.

Costanza, R., and T. Maxwell. 1994. "Resolution and predictability: an approach to the scaling problem." I 9(1):47-57.

Curran, P. J. 1988. "The semivariogram in remote sensing: an introduction." *Remote Sensing Env* 24(3):493-507.

de Lima Guedes, I. C., J. M. de Mello, E. M. de Oliveira Silveira, C. R. de Mello, A. A. dos Reis, and L. Rezende Gomide. 2015. "Continuidade espacial de características dendrométricas em povoamentos clonais de Eucalyptus sp. avaliada ao longo do tempo." *Cerne* 21(4):527-533

Delmelle, E. M., H. Zhu, W. Tang, and I. Casas. 2014. "A web-based geospatial toolkit for the monitoring of dengue fever." *Applied Geography* 52, pp. 144-152. doi: 10.1016/j.apgeog.2014.05.007

De Oliveira Silveira, E. M., J. M. Mello, F. W. Acerbi Jr, A. A. dos Reis, K. D. Withey, and L. A. Ruiz. 2017. "Characterizing landscape spatial heterogeneity using semivariogram parameters derived from NDVI image." *Cerne* 23(4):413-422.

Dormann, C. F., J. M. Mcpherson, M. B. Araújo, R. Bivand, J. Bolliger, G. Carl, R. G. Davies, A. Hirzel, W. Jetz, W. D. Kissling, and I. Kühn. 2007. "Methods to account for spatial autocorrelation in the analysis of species distributional data." *Ecography* 30(5):609-628.

Dormann, C. F. 2007. "Effects of incorporating spatial autocorrelation into the analysis of species distribution data." *Global Ecology and Biogeography* 16(2):129-138.

Environmental Systems Research Institute (ESRI). 2019a. Understanding a semivariogram: The range, sill, and nugget. https://pro.arcgis.com/en/pro-app/help/analysis/geostatistical-analyst/understanding-a-semivariogram-the-range-sill-and-nugget.htm

Environmental Systems Research Institute (ESRI). 2019b. Choosing a lag size. http://desktop.arcgis.com/en/arcmap/latest/extensions/geostatistical-analyst/choosing-a-lag-size.htm

Fisher, J. R., E. A. Acosta, P. J. Dennedy-Frank, T. Kroeger, and T. M. Boucher. 2017. "Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality." *Remote Sensing in Ecology and Conservation* 4(2):137-149.

Goodchild, M.F. 2011. "Scale in GIS: An overview." *Geomorphology* 130:5-9.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment.

Griffin, S. P. 2017. "Tactical-level Disease Vector Hazard Mapping using the Maximum Entropy (Maxent) Algorithm and ArcGIS Standard Operating Procedures (SOP)." SOP, Geospatial Research Laboratory, Engineer Research and Development Center 2019.

Hao, T., J. Elith, G. Guillera-Arroita, and J. J. Lahoz-Monfort. 2019. "A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD." *Diversity and Distributions* 25(5):839-852.

Herath, S. M., P. R. Sarukkalige, and V. T. V. Nguyen. 2016. "A spatial temporal downscaleing approach to development of IDF relations for Perth airport region in the context of climate change." *Hydrological Sciences Journal*, 61(11): 2061-2070.

Hesse, E. M., L. J. Martinez, R. G. Jarman, A. G. Lyons, K. H. Eckels, R. A. De La Barrera, and S. J. Thomas. 2017. "Dengue Virus Exposures Among Deployed U.S. Military Personnel." *American Journal of Tropical Medicine and Hygiene* 96:5 pp. 1222–1226. doi:10.4269/ajtmh.16-0663

Hirzel, A. H., and R. Arlettaz. 2003. "Modeling habitat suitability for complex species distributions by environmental-distance geometric mean." *Environmental Management* 32(5):614-623.

Hirzel, A. H., J. Hausser, D. Chessel, and N. Perrin. 2002. "Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data?" *Ecology* 83(7):2027-2036.

Huth, R. 2002. "Statistical downscaling of daily temperature in Central Europe." *Institute of Atmospheric Physics, Prague, Czech Republic.*

Hyppänen, H. 1996. "Spatial autocorrelation and optimal spatial resolution of optical remote sensing data in boreal forest environment." *Int J Remote Sens* 17(17):3441-3452.

Joint Chiefs of Staff. 2016. *Doctrine for Health Service Support in Joint Operations: Force Health Protection (FHP).* Joint Publication 4-02.8. Washington, DC: Joint Chiefs of Staff.

Khan, M. S., P. Coulibaly, and Y. Dibike. 2005. "Uncertainty analysis of statistical downscaling methods." *Journal of Hydrology,* 319: 357-382.

Kitron, U., J. A. Clennon, M. C. Cecere, R. E. Gurtler, C. H. King, and G. Vazquez-Prokopec. 2006. "Upscale or downscale: Applications of fine scale remotely sensed data to Chagas disease in Argentina and schistosomiasis in Kenya." *Geospatial Health*, 1(1): 49-58.

Kong, L., J. Wong, W. Han, and Z. Cao. 2016. "Modeling Heterogeneity in Direct Infectious Disease Transmission in a Compartmental Model." *Int J Environ Res Public Health*. 13(3): 253.

Kong, L., J. Wang, Z. Li, S. Lai, Q. Lio, H. Wu, and W. Yang. 2018. "Modeling the Heterogeneity of Dengue Transmission in a City." *Int. J. Environ. Res. Public Health*. 15(6): 1128.

Kühn, I., and C. F. Dormann. 2012. "Less than eight (and a half) misconceptions of spatial analysis." *J Biogeography* 39(5):995-998.

Kumar, J., B. G. Brooks, P. E. Thornton, and M. C. Dietze. 2012. "Sub-daily statistical downscaling of meteorological variables using neural networks." *Proceeding Computer Science*, 9(2012): 887-896.

Lam, NS-N, and D. A. Quattrochi. 1992. "On the issues of scale, resolution, and fractal analysis in the mapping sciences." *The Professional Geographer* 44(1):88-98.

Legendre, P. 1993. "Spatial autocorrelation: Trouble or new paradigm?" *Ecology* 74(6):1659-1673.

Linard, C., and A. J. Tatem. 2012. "Large-scale spatial population databases in infectious disease research." *International Journal of Health Geographics* 11:7. doi: 10.1186/1476-072X-11-7

Löw, F., and G. Duveiller. 2014. "Defining the spatial resolution requirements for crop identifycation using optical remote sensing." *Remote Sensing*, 6(9):9034-9063.

McCloy, K. R., and P. K. Bøcher. 2007. "Optimizing image resolution to maximize the accuracy of hard classification." *Photogrammetric Engineering & Remote Sensing* 73(8):893-903.

Mendes, D., and J. A. Marengo. 2009. "Temporal downscaling: A comparison between artificial neural network and autocorrelation techniques over the Amazon Basin in present and future climate change scenarios." *Theoretical and Applied Climatology*, 100(3-4): 413-421.

Modis, K., and K. Papaodysseus. 2006. "Theoretical estimation of the critical sampling size for homogeneous ore bodies with small nugget effect." *Mathematical Geology* 38(4):489-501.

Nezer, O., S. Bar-David, T. Gueta, and Y. Carmel. 2017. "High-resolution species-distribution model based on systematic sampling and indirect observations." *Biodiversity and Conservation* 26(2):421-437.

Nyquist, H. 1928. "Certain topics in telegraph transmission theory." *Transactions of the American Institute of Electrical Engineers* 47(2):617-644 systematic sampling and indirect observations. *Biodiversity Conservation* 26(2):421-437.

Pages, F., M. Faulde, E. Orlandi-Pradines, and P. Parola. 2010. " The past and present threat of vector-borne diseases in deployed troops." *Clinical Microbiology and Infectious Diseases* 16: 209–224. doi:10.1111/j.1469-0691.2009.03132.x

Pang, B., J. Yue, G. Zhao, and Z. Xu. 2017. "Statistical Downscaling of Temperature with the Random Forest Model." *Advances in Meterology* 17. doi: 10.1155/2017/7265178

Parham, P. E., J. Waldock, G. K. Christophides, D. Hemming, F. Agusto, K. J. Evans, N. Fefferman, H. Gaff, Ab. Gumel, S. LaDeau, S. Lenhart, R. M. Mickens, E. N. Naumova, R. S. Ostfeld, P. D. Ready, M. B. Thomas, J. Velasco-Hernandez, and E. Michale. 2015. "Climate, environmental and socio-economic change: weighing up the balance in vector-borne disease transmission." *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 370: 1665.

Phillips, S. J. 2017. A Brief Tutorial on Maxent. http://biodiversityinformatics.amnh.org/open_source/maxent/ (date accessed: 08 August 2018.)

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. "Maximum entropy modeling of species geographic distributions." *Ecological Modelling* 190(3-4):231-259.

Phillips, S. J., P. A. Robert, and R. E. Schapire. 2006. "Maximum entropy modeling of species geographic distributions." *Ecological Modelling* 190 (3-4): 231-259. Accessed July 29, 2019. https://www.cs.princeton.edu/~schapire/papers/ecolmod.pdf.

Rahman, A. 2017. "Estimating small area health-related characteristics of populations: a methodological review." *Geospatial Health*, 12(495): 3-14.

Rahman, A. F., J. A. Gamon, D. A. Sims, and M. Schmidts. 2003. "Optimum pixel size for hyperspectral studies of ecosystem function in southern California chaparral and grassland." *Remote Sensing Env* 84(2):192-207.

Rodriguez-Carrion, N. M., S. D. Hunt, M. A. Goenaga-Jimenez, M. Vélez-Reyez. 2014. Determining optimum pixel size for classification. In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XX* (Vol. 9088, p. 90880X). International Society for Optics and Photonics.

Rytkonen, M. J. 2004. "Not all maps are equal: GIS and spatial analysis in epidemiology." *International Journal of Circumpolar Health* 63:1, pp. 9-24. doi: 10.3402/ijch.v63i1.17542.

Santini, L., S. Pironon, L. Maiorano, and W. Thuiller. 2019. "Addressing common pitfalls does not provide more support to geographical and ecological abundant-centre hypotheses." *Ecography* 42(4):696-705.

Sax, C., and P. Steiner. 2013. "Temporal disaggregation of time series." *The R Journal*, 5(2): 80-87.

Shah, N. H., and J. Gupta. 2013. "SEIR Model and Simulation of Vector Borne Diseases." *Applied Mathematics*, 2013, 4, 13-17

Sharma, A., K. N. Tiwari, and P. B. S. Bhadoria. 2011. "Determining the optimum cell size of digital elevation model for hydrologic application." *J Earth System Sci* 120(4):573-582.

Siraj, A. S., M. J. Bouma, M. Santos-Vega, A. K. Yeshiwondim, D. S. Rothman, D. Yadeta, P. C. Sutton, and M. Pascual. 2015. "Temperature and population density determine reservoir regions of seasonal persistence in highland malaria." *Proceedings Royal Society B*, 282: 20151383.

Soucy, J. P. R., A. M. Slatculescu, C. Nyiraneza, N. H. Ogden, P. A. Leighton, J. T. Kerr, and M. A. Kulkarni. 2018. "High-resolution ecological niche modeling of Ixodes scapularis ticks based on passive surveillance data at the Northern Frontier of Lyme disease emergence in North America." *Vector-Borne and Zoonotic Diseases* 18(5):235-242.

Stuckey, E. M., T. Smith, and N. Chitnis. 2014. "Seasonally dependent relationships between indicators of malaria transmission and disease provided by mathematiccal model simulations." *PLoS Comput Biol*, 19(9): e1003812. *Trends in Parasitology*, 29(10): 477-482.

Sturrock, H., J. M. Cohen, P. Keil, A. J. Tatem, A. L. Menach, N. E. Ntshalintshali, M. S. Hsiang, and R. D. Gosling. 2014. "Fine-scale malaria risk mapping from routine aggregated case data." *Malaria Journal* 13:421. doi: 10.1186/1475-2875-13-421.

van Panhuis, W. G., J. Grefenstette, S. Y. Jung, Ni. S. Chok, A. Cross, H. Eng, B. Y. Lee, V. Zadorozhny, S. Brown, D. Cummings, and D. S. Burke. 2014. "Contagious diseases in the United States from 1888 to the present." *New England Journal of Medicine*, 369(22): 2152-2158.

Wang, J. X., M. G. Hu, S. C. Yu, G. X. Xiao. 2017. "Downscaling research of spatial distribution of incidence of hand foot and mouth disease based on area-to-area Poisson Kriging method." *Europe PMC* 38:9, pp. 1201-1205. doi: 10.3760/cma.j.issn.0254-6450.2017.09.012.

Woodcock, C. E., and A. H. Strahler. 1987. "The factor of scale in remote sensing." *Remote Sensing of Environment* 21:311-322.

Wu, S. S., B. Xu, and L. Wang. 2006. "Urban land-use classification using variogram-based analysis with an aerial photograph." *Photogrammetric Engineering & Remote Sensing* 72(7):813-822.

Wulder, M., and B. Boots. 2001. "Local spatial autocorrelation characteristics of Landsat TM imagery of a managed forest area." *Canadian J Remote Sensing* 27(1):67-75.

Zhang, Q., P. Shi, V. P. Singh, K. Fan, and J. Huang. 2016. "Spatial downscaling of TRMM-based precipitation data using vegetative response in Xinjiang, China." *International Journal of Climatology* 37:10, pp. 3895-3909. doi: 10.1002/joc.4964.

Appendix A: NET-CMO Summary

New and Enhanced Tools for Civil-Military Operations (NET-CMO)

U.S. ARMY CORPS OF ENGINEERS   BUILDING STRONG

Background:

Current civil military operations (CMO) tools that are used to ensure long-term regional stability, assist in foreign humanitarian assistance (FHA), increase public health, or provide situational awareness need to be either enhanced or new tools developed. Current capabilities are lacking in their ability to support mission analysis that enhance stability and mitigates threats to the civilian population. CMO planners do not have the capability to identify quantitatively regions which will experience violence or areas that will experience an increase in vector-borne diseases. Additionally, due to a lack of data, CMO planners are unable to develop detailed area studies and plans.

Current Work:

Three different problem sets for CMO have been identified below:

1. **Mosquito-borne disease prediction:** Over half of the world's population lives in an area of the world for which mosquito-borne diseases are endemic. With changing temperatures and increased extreme weather events, the global reach of these diseases is expanding. Currently, CMO do not possess the capability to predict when and where these mosquito-borne diseases will occur or the portion of the population who could become infected. With the increasing spatial distribution of mosquito-borne diseases and the growing resistance of these diseases to medicines, it is imperative that CMO be able to understand when and where these diseases will occur, spread, and their severity. This will protect a region's public health and help to stabilize the population.

2. **Environmental Security Index:** Food security, water availability, and a region's temperature, affect the stability of a region. A validated methodology currently does not exist that is capable of predicting environmental impact on a populous, food security, water availability, and regional stability. Such a tool would prove useful for CMO to provide FHA and to help shape the operational environment after a crisis.

3. **Low resolution data in data denied regions:** Many regions of the world are data poor in terms of environmental, disease, demographic, and social data. Data that are available typically have coarse spatial and temporal resolution – with data only available annually at either a country or province level. Information with such a low fidelity provides poor situational awareness of a theater and makes it difficult for CMO planners to develop area or functional-oriented studies.

Capabilities:

- Map and analysis of the spatial and temporal spread of any mosquito-borne disease and the proportion of the population that will become infected
- Tool capable of finding a uniform pixel size when working with data with varying spatial resolutions
- The capability to downscale coarse temporal and spatial resolution data to provide a stronger situational awareness of a region
- Environmental security analysis and the ability to produce a map that details which regions will experience higher levels of volatility due to water and food shortages and other climate related issues

POC:
Project Manager (PM): Nicole M. Wayant, (703) 428-7434, Nicole.M.Wayant@usace.army.mil

U.S. ARMY CORPS OF ENGINEERS – Geospatial Research Laboratory
7701 TELEGRAPH RD, ALEXANDRIA, VA 22315
www.usace.army.mil

Appendix B: Species Distribution Modeling of *Ixodes scapularis* and Associate Pathogens in States East of the Mississippi River Summary

Introduction

The Tick-Borne Disease Laboratory (TBDL) of the Army Public Health Center requested that the ERDC Geospatial Research Laboratory analyze tick observation data and model the presence of the vector, *I. scapularis,* and the associated pathogens carried by this vector, using a modeling methodology similar to the one used by St. John, et al. 2016. Ecological niche models were developed using the maximum entropy (Maxent) methodology (Phillips 2006), which creates a probability map predicting where a target species will be found based on a dataset of presence observations of the species, combined with multiple environmental datasets believed to describe habitat suitability.

Methodology

Using location and species information for *I. scapularis* and environmental input raster data, a Maxent model was generated using a toolbox and standard operating procedure (SOP) developed for use in ArcGIS (Griffin 2017). The Maxent methodology generates measures to assess the suitability of the model's prediction in the form of a sensitivity plot. Once the pathogen presence probability models were complete, the models were compared using the Raster Calculator tool in the Map Algebra toolbox under Spatial Analyst Tools in ArcGIS Pro 2.3.3. Areas where overlapping probabilities are shown could indicate locations where co-infections are likely to occur. As *I. scapularis* initially acquires pathogens from previous host species, such as rodents or deer, before passing them on to humans; probability distributions could also indicate regions where higher occurrences of host species carrying disease

End Product

The result was a map of *Ixodes scapularis* habitat prediction. The map's color ramp indicates the probability that *Ixodes scapularis* is the probability that at least one tick is in that pixel, derived from the presence locations and environmental variables.

Ixodes scapularis
life cycle.

LARVA   NYMPH   ADULT MALE   ADULT FEMALE
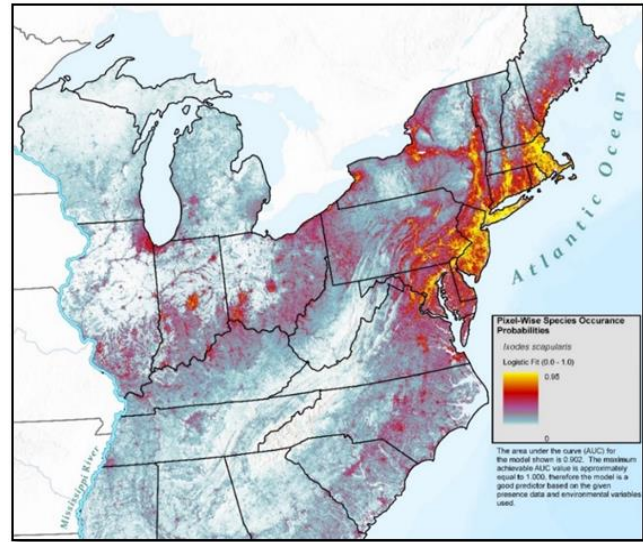
Appendix C: Understanding the Disease Vector Operational Environment by Predicting Presence of *Anopheles* Mosquito Breeding Sites Using Maximum Entropy Modeling Summary

Understanding the Disease Vector Operational Environment by Predicting Presence of *Anopheles* Mosquito Breeding Sites Using Maximum Entropy Modeling: Maxent Software Platform

Introduction

The purpose of this study is to help address the need for new and innovative methods to promote military readiness through better understanding of vector-borne disease threats in both familiar and unfamiliar operational environments. The *Anopheles* genus is comprised of several hundred species of mosquito, dozens of which have the ability to transmit parasites that cause malaria in humans (Kim, et al. 2011). Malaria is of particular interest to medical planners, because of the long history of adverse effects on combat operations (U.S. Dept of the Army 2015).

Methodology

Maximum entropy is a machine learning method that makes predictions based on incomplete information (Phillips, Anderson and Schapire 2006) Since the maximum entropy method is not usually implemented in standard statistical software platforms, a dedicated software application called Maxent can be used for analysis. Maxent is a Java-based open source Windows application that was developed to provide a way to use the maximum entropy method to model species geographic distributions (Phillips, Dudík and Schapire 2004). The software uses data from known species presence locations, along with environmental data that describe aspects of the habitat that may cause the species to prefer those locations, and builds a probability map predicting the likelihood that the target species will be found throughout the area of interest (Phillips, Anderson and Schapire 2006).

End Product The Maxtent output shows the higher predicted likelihood of the presence of *Anopheles* breeding sites as shown in red, and cells with lowest predicted likelihood are shown in blue. The presence locations used to build the model are indicated by white squares on the prediction map.
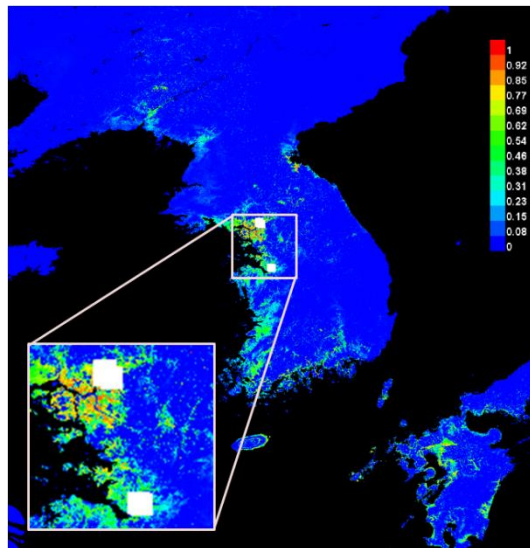
Acronyms

CART	classification/regression tree
COA	Course of Action
COP	Common Operating Picture
CMO	Civil Military Operations
CRADA	Cooperative Research and Development Agreement
DCGS	Distributed Common Ground System
DoD	Department of Defense
FHA	Foreign Humanitarian Assistance
FHP	Force Health Protection
FRS	Feature Requirement Specification
GFM	Global Force Management
GEE	Google Earth Engine
GEOD	Geospatially Enabled Operational Design
GLS	Generalized Least Squares
GRL	Geospatial Research Laboratory
HPC	High Performance Computing
LSD	Local Spatial Dispersion
LSIB	Large-Scale International Boundary
LST	Land Surface Temperature
LSV	Local Spatial Variance
MAD	Mean Absolute Deviation
MBDS	Mosquito-Borne Disease Simulation
MBPS	Map Based Planning Services
MDMP	Military Decision Making Process
MAE	Mean Absolute Error
NDVI	Normalized Difference Vegetation Index
NET-CMO	New and Enhanced Tools for Civil-Military Operations
OE	Operational Environment
PMP	Project Management Plan
R&D	Research and Development
RF	Random Forests
RMSE	Root Mean Square Error
ROI	Region of Interest
USAPHC	U.S. Army Public Health Center
TBDL	Tick-Borne Diseases Lab
UI	User Interface
VAST	Vulnerability Assessment Software Toolkit Readiness

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
February 2020	Final	

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
New and Enhanced Tools for Civil Military Operations (NET-CMO)	
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
	62784

6. AUTHOR(S)	5d. PROJECT NUMBER
Nicole M. Wayant , Sarah J. Becker, Joshua Parker, S. Bruce Blundell, Susan Lyon, Megan Maloney, Robin Lopez, Sean Griffin and John A. Nedza	855
	5e. TASK NUMBER
	22
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Geospatial Research Lab (GRL) U.S. Army Engineer Research and Development Center 7701 Telegraph Road Alexandria, VA 22315	ERDC/GRL TR-20-3

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
U.S. Army Corps of Engineers Washington, DC 20314-1000	
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Civil Military Operations (CMO) associated geospatial modeling is intended to enable increased knowledge of regional stability, assist in Foreign Humanitarian Assistance (FHA), and provide support to Force Health Protection (FHP) operational planning tasks. However, current geoenabled methodologies and technologies are lacking in their overall capacity to support complex mission analysis efforts focused on understanding these important stability factors and mitigating threats to Army soldiers and civilian populations. CMO analysts, planners, and decision-makers do not have a robust capability to both spatially and quantitatively identify Regions of Interest (ROI), which may experience a proliferation in health risks such as vector-borne diseases in areas of future conflict. Additionally, due to this general absence of geoenabled health assessment models and derived end-products, CMO stakeholders are adversely impacted in their Military Decision Making Process (MDMP) capabilities to develop comprehensive area studies and plans such as Course of Action (COA). The NET-CMO project is focused on fostering emerging geoenabling capabilities and technologies to improve military situational awareness for assessment and planning of potential health threat-risk vulnerabilities.

15. SUBJECT TERMS		
United States – Armed Forces – Stability operations	Humanitarian assistance Geospatial data Geographic information systems	Mosquitoes as carriers of disease Epidemiology

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	SAR	86	
Unclassified	Unclassified	Unclassified			**19b. TELEPHONE NUMBER** *(include area code)*