ERDC/CERL TR-19-20

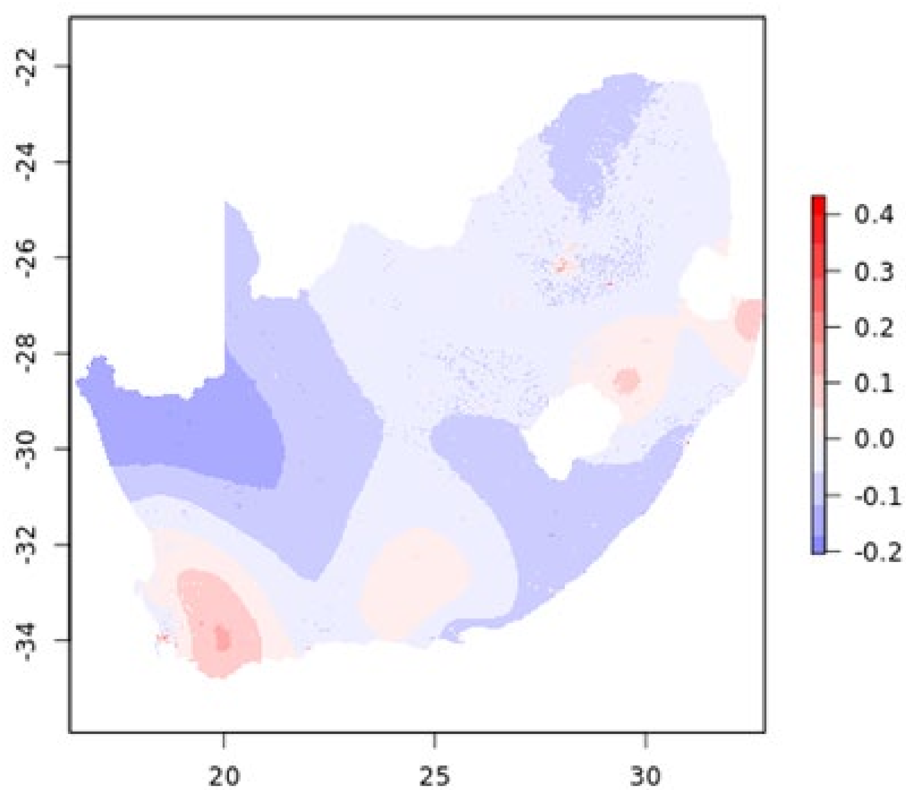
US Army Corps
of Engineers®
Engineer Research and
Development Center

Risk Assessment Planning and Tools for OpeRations (RAPTOR)

Using the Social Vulnerability Index to Forecast Disaster Migration

Chandler M. Armstrong and Lance L. Larkin

September 2019



Construction Engineering
Research Laboratory

Using the Social Vulnerability Index to Forecast Disaster Migration

Chandler M. Armstrong and Lance L. Larkin

U.S. Army Engineer Research and Development Center (ERDC)
Construction Engineering Research Laboratory (CERL)
2902 Newmark Dr.
Champaign, IL 61824

Final Technical Report (TR)

Prepared for Headquarters, U.S. Army Corps of Engineers
Washington, DC 20314-1000

Under Program Element 622784, "Military Engineering Technology";
Project 855; Task 08

Abstract

This report analyzes disaster-driven internal migration, its effects on residence layout by social vulnerability, and demonstrates a modeling procedure for this task. Since the processes underlying disaster-driven migration may unfold at scales below those in the data, especially in vulnerable areas where these models are especially useful, this report focuses on South Africa, which had key flooding events at urban centers over 15 years before the 2011 census. Since only low spatial resolution census data are obtainable for research, we use universal kriging (UK) with a pre-fit model from the Philippines (which used more detailed data) to estimate social vulnerability in South Africa at a higher spatial resolution. With the UK model, we estimate and reaggregate a social vulnerability index (SVI) from South African provincial to municipal boundaries, then fit a model testing the relationship between internal migration, SVI, and flooding risk within each municipality. Results show that, when controlling for flood risk, within-province migration is inversely related to SVI, while neither SVI nor flooding affect migration between provinces. This is consistent with our hypothesis that the socially vulnerable are less likely to leave flood-prone areas, and over time, a positive spatial correlation emerges between SVI and flood risk.

Contents

Tables and Figures

Tables

Figures

Preface

This study was conducted for Headquarters, U.S. Army Corps of Engineers (HQUSACE) under Program Element 622784, “Military Engineering Technology”; Project 855, “Topographical, Image Intelligence and Space”; Task 08. The technical monitor was Angela Rhodes, CEERD-CNC.

The work was performed by the Land and Heritage Conservation Branch, of the Installations Division, of the U.S. Army Engineer Research and Development Center, Construction Engineering Research Laboratory (ERDC-CERL). At the time of publication, Michael L. Hargrave was Chief of the Land and Heritage Conservation Branch; Michelle Hanson was Chief of the Installations Division; and Alan Anderson was the Technical Director for Military Ranges and Lands. The Deputy Director of ERDC-CERL was Dr. Kumar Topudurti and the Director was Dr. Lance D. Hansen.

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

THIS PAGE INTENTIONALLY LEFT BLANK

1 Introduction

1.1 Background

Rising global temperatures have caused a host of impacts worldwide: melting glaciers, a rapid decrease in Arctic ice (and consequent rush by nations to claim new Arctic resources [Shea 2019]), rising sea levels, warming oceans, and extreme weather across many regions of the globe. The intensification of atmospheric anomalies include a growing number of droughts,* devastating hurricanes, and cyclones, which can then combine to cause elevated risks of flooding in populated areas.

Second order impacts of climate change damage infrastructure and displace people in increasing numbers. Also, while studies have shown that climate disaster is not directly linked to interstate or civil war, changing climate and resulting environmental catastrophe can trigger or accelerate internal state conflict (Fleishman and Goodman 2018, Kelley et al. 2018, Moran et al. 2014), which then often causes asylum seekers to leave countries to escape areas of political transformation (Abel et al. 2019). Regardless of whether people are forced to migrate because of climate disaster, conflicts over diminishing resources, or political strife, threats to security and state stability ultimately stem from catastrophic impacts to large populations (Werrell and Femia 2016).

Within the current context of escalating human mass migrations, it is important to consider not only how coercive governments use the movements of displaced people to pressure neighboring states (Greenhill 2010), but also how the effects of climate change impact and destabilize countries. To clarify those effects, there is a need to build the capacity to analyze disaster-driven migration, and more generally, vulnerability to disasters worldwide. This work was undertaken to explore the relationship between disaster impacts and the spatial arrangement of human settlements, specifically, to test the hypothesis that the socially vulnerable are more likely to live in areas that bear a greater impact from natural disasters than surrounding locations.

* Sub-Saharan Africa is extremely vulnerable to drought as rain-fed agriculture provides approximately 90% of the region's food or feed, and it provides livelihoods for 70% of the population (Shiferaw et al. 2014, p 68).

1.2 Objectives

The objective of this work was to explore the relationship between disaster impacts and the spatial arrangement of human settlements, specifically, to test the hypothesis that the socially vulnerable are more likely to live in areas that bear a greater impact from natural disasters than surrounding locations. One mechanism by which this can happen is that more wealthy people will migrate out of a disaster-prone area, while those living in communities with greater social vulnerability are more likely to stay behind because they lack the capital to move. Over time the spatial layout of human settlements by vulnerability may become such that the most vulnerable also live in the locations most impacted by natural disasters.

1.3 Approach

It is generally accepted that the vulnerable and poor often bear the largest impacts of disasters, but we are interested in investigating how this shapes spatial layout of communities, and potentially shapes long-term urban growth. To this end, we examined flood disaster-driven migration to investigate how it might shape residential layout within and around urban areas by developing a method to map inter-regional migration. To examine the flows of people, we drew on the Integrated Public Use Microdata Series (IPUMS)-International census data. In the South African data, we then implemented a Social Vulnerability Index (SVI) initially developed by Ignacio for Filipino IPUMS International census (Ignacio et al. 2015). Finally, we mapped and reaggregated the data to municipal boundaries where we tested a hypothesis relating internal migration, SVI, and risk of flooding.

1.4 Scope

The current analysis focuses on South Africa but is founded on previous spatial analysis of social vulnerability of Filipino communities. The Filipino census provides much greater spatial detail than that of South Africa, and key variables could be downscaled within reasonable limits of uncertainty. South Africa has greater land surface area than the Philippines, yet the census is georeferenced to only 25 second-level administrative units, in contrast to over 700 such units in the Filipino census. A consequence of South Africa's much larger administrative units is that data downscaled from these units have much greater interpolation uncertainty, and hypothesis tests dependent on spatial granularity for power will be much weaker.

However, we are interested in testing the generalizability of spatially based behavioral models to areas of the world where source datasets are much sparser. We have selected South Africa for the quality of their census, combined with its coarse spatial granularity, as a means for testing a spatial model of human behavior in a location where the model will be less statistically powerful. Using regression kriging we demonstrate a method for estimating SVI at spatial supports that are suitable for modeling disaster-driven migration.

2 Developing a Social Vulnerability Index

Social vulnerability is a measure of a community's resilience, its ability to quickly recover from difficulties. More vulnerable communities are less resilient, less able or likely to cope and recuperate from disaster. Social scientists have studied a number of factors that affect a community's resilience, in diverse configurations and locations, and have developed a variety of social vulnerability indices (SVIs), which are planning tools used to evaluate a community's relative social vulnerability and to identify populations at risk. Some indices include, for example, the Baseline Resilience Index for Communities (BRIC) (Cutter et al. 2010); the Community Disaster Resilience Index (CDRI) (Peacock et al. 2010); and the Household SVI (HSVI) (Vincent and Cull 2010). We use the Social Vulnerability Index (SVI) developed by Ignacio et al. (2015) for characterizing social vulnerability to disasters in the Philippines. Ignacio obtained census data from the Philippines Statistics Authority and computed SVI from neighborhood, or barangay, level data. The following procedures use data from IPUMS International (Minnesota Population Center 2018) that are georeferenced to municipalities. SVI is a nebulous and context dependent concept. The important components of an SVI metric may vary for different locations.

Because communities consist of social configurations of people within households, the index is built from concepts of what makes a household vulnerable to hazards. Household heads are presumed to be stronger representatives of household vulnerability; their individual demographics enter directly into SVI calculation. Demographic factors of other household members are aggregated into counts or proportions, and the presence of migrants abroad is also noted. Household assets that round out the SVI include residential ownership, residential building materials, automobiles, household appliances, and presence of overcrowding. Demographic factors include gender, age, and disability, which can combine to indicate less resilience within a patriarchal, capitalist society. Demographic and household factors used to calculate the SVI are socioeconomic status, household composition and disability, and housing. Table 1 lists the factors going into the SVI.

Table 1. Components of the social vulnerability index.

Variable	Unit	Vulnerability Concept
% female	person	gender
% children	person	age
% elderly	person	age
% adults with no secondary degree	person	education
% with no birth registration	person	social dependence
% 20-39 year females with no secondary degree	person	education&gender
% households with female head	household	gender&family
% households with disabled	household	special needs
% heads of household with no secondary degree	household	education
% single heads of household	household	family
average size of household	household	family
% households with no overseas worker	household	social dependence
% households with poor roofing	dwelling	residence
% households with poor walling	dwelling	residence
% households with no tenure	dwelling	social dependence
% houses over 30 years dwelling residence	dwelling	residence
% houses needing repair	dwelling	residence
% houses < 10m2	dwelling	socioeconomic status

Using Principal Components Analysis (PCA), these factors are combined into a single value representing vulnerability. The PCA weights the impact of each component factor on the compressed vulnerability metric based on the influence of each component factor on total variability. The resulting metric represents a difference from the local average of vulnerability; negative values indicate less vulnerability than the local average, and positive values show greater than the local average vulnerability. Because the SVI draws from IPUMS International census data, we can compute the index for other areas of the world and modify it for local contexts. It is also important to note that the model requires only aggregated data; while we use IPUMS microdata, a dataset could be built from aggregated statistics, which often also offer statistics aggregated from more recent surveys.

The factors comprising the SVI are broad and can also be culturally specific. Although the factors may be idiosyncratic, the vulnerability represented by a given SVI can remain universal. For example, both Flanagan et al. (2011) and Ignacio et al. (2015) developed SVI to disaster indices, intended for use with U.S. and Filipino censuses, respectively. Ignacio's SVI uses roofing and walling material as a factor intended to represent vulnerability to disasters that result from poorly constructed or dilapidated residences, while Flanagan's SVI represents a similar aspect with an indicator for mobile home residences. The specific variables used to represent this aspect of vulnerability are unique to the two censuses and also to the contexts they reflect; yet, what they represent for the SVI is congruent or similar. Extrapolation can downscale the SVI to new locations where residential construction and data to measure this aspect of vulnerability may be different from where the SVI model was trained. Nevertheless, the extrapolated SVI continues to represent the concepts it is defined to represent, but whether this representation is faithful is another question that will be discussed below.

Universalizing the SVI is important because we obtain an SVI for South Africa by generalizing a model fit to the Filipino census. The Filipino model aggregated and mapped SVI to municipal boundaries, as determined from IPUMS data, and downscaled SVI based on a set of geographical features that are correlated with the index. These geographical factors, which help us predict an SVI in locations where we cannot observe its value, are a key element for universalizing SVI. When applied to the Philippines, the model downscales SVI to the scale of the correlated geographical features; a scale where the SVI had not been observed. When applied to South Africa, the model is downscaling SVI to a new scale; however the SVI and its relationship to South African geography may not mirror those same dynamics in the Filipino model.

The model predicts SVI from geographical features using regression kriging (RK). The geographical features are represented as rasters and are resampled to a 1-km resolution. The SVI is represented with vector data having full coverage over the geographical area of interest. The Filipino data is available aggregated in smaller polygons than the South African data: municipalities and municipality districts respectively, where municipality districts are a much larger spatial boundary.

The benefit of using a Filipino model to downscale SVI in South Africa is that, because the Filipino model is fit to data aggregated at much smaller spatial boundaries, the Filipino data yields better estimates for a geographical model of SVI. When applied outside the Philippines, however, those estimates may also be biased. We identify two sources of bias, one philosophical in nature and the other well-defined in the statistics.

The first bias stems from differences between SVIs that could be calculated for different locations with different data. The factors used to calculate an SVI for the Philippines with Filipino data may not exist for South Africa. For example, the house construction or specific industry of employment factors used to calculate SVI in the Philippines may not be pertinent to South Africa. The specific data only operationalize the broad concept, however. If these concepts remain pertinent between locations, then this potential source of bias is not necessarily a problem. Even if given factors or whole aspects are not relevant for SVI in South Africa, the phenomenon of social vulnerability to disasters does exist. The model, in theory, represents only social vulnerability to disasters and not the specific manifestations of social vulnerability to disasters, e.g., housing or industry of employment.

The second type of bias is better defined as a statistical concept; this is the potential bias between geographical factors and SVI. The model predicts SVI using parameter estimates for the geographical features associated with the Filipino data. If the correlations between geographical features and SVI is not similar for both the Philippines and South Africa, then a Filipino model applied to South African geography could be biased. This occurs because the model is an extrapolation from correlations between Filipino geography and SVI to South African geography and SVI; one must assume the relationship between geography and SVI is the same for all locations.

3 The Philippines and the SVI

This chapter details the logic of building a RK model of SVI in the Philippines, which will provide the means for estimating SVI in other areas of the world. We chose the Philippines largely as a matter of demand from other scientists. Additional advantages of this selection are that the Philippines census also provides a good degree of geospatial disaggregation and is ideal for fitting a geospatial model of SVI.

The following section describes the problems solved by the model and technical details on model fitting. The section starts with a description of the Change of Support Problem (COSP), a general problem the model must solve. The second subsection gives technical details of the modeling method. The final subsection briefly describes the workflow for building the model.

3.1 Change of support problem

The COSP refers to inference of values at points or areas other than those where values are observed (Gelfand et al. 2001). In cases where observations are available only as areal aggregations and where the goal is to estimate new values at unobserved areas, the problem is more generally referred to as the modifiable areal unit problem (MAUP) (Cressie 1996). This section discusses the logic of a kriging model for downscaling the SVI from polygons (the source support) to a regular grid (the target support).

Kriging was developed to model and estimate spatial distribution of natural resources. The notable feature of kriging is that it models spatial autocorrelation of the data, and not the actual data values. Kriging was originally designed to aggregate data in so-called block-kriging (Cressie 1990, Matheron 1963). Conventionally, kriging analyzes samples of point data — the most disaggregated form of data— to estimate a complete spatial distribution for the purpose of summarizing the resources available over an area of space. With its original design scope, kriging required modifications to estimate a spatial distribution from data aggregated in polygons. The solution is A2A/P kriging (Krivoruchko et al. 2011, Kyriakidis 2004).

The basic technique of A2A/P kriging is to discretize the polygons into a series of regular points, and to estimate the spatial autocorrelation between any two polygons as the average correlation of their discretizing points (Kyriakidis 2004). Similarly, correlation between points and polygons is the average of correlation between the point and discretizing points of the polygon. A2A/P kriging is a type of binomial cokriging, mentioned in the previous section, that cokriges the discretized polygon data with target discretization points from other polygons or a regular grid where one wishes to interpolate data, i.e., the target support. This whole procedure works because, as the next section will discuss, kriging models only variance as a function of distance between points and does not need access to underlying data. The kriging model is simply an autocorrelation model that describes how best —as in best linear unbiased estimate (BLUE)— to interpolate observed data to unobserved supports.

3.2 Regression kriging

An RK model begins with a linear regression model, indexed to space. The SVI is initially downscaled to raster cells by assigning the value of the municipality to each raster cell within municipality borders. Cells that intersect multiple municipalities receive an average of all intersecting municipalities, weighted by the area of each intersection. The regression model fits these downscaled SVI data to the geographical features named in the previous paragraph, and residuals are calculated.

Two methods are available for fitting the regression model of RK (hereafter referred to as the features space model). The first method is kriging with external drift (KED), which fits both the feature and spatial interpolation model in one step. Another option is universal kriging (UK), which fits the features space model separately and uses the estimated parameters to calculate residuals before the spatial interpolation step. If UK fits the feature space model using linear regression, then the results are typically similar or identical to KED (Hengl et al. 2007, Pebesma 2006). We use KED to fit an SVI model to Filipino data and to apply the model to South African data. The next paragraph describes the spatial interpolation step.

UK implements spatial interpolation using simple kriging (SK). SK is identical to a basic Gaussian Process regression, and likewise also to Bayesian Linear Regression with priors over functions. SK assumes a known mean, which we can do safely as the regression model guarantees the mean of residuals is zero. SK fits a covariance function and computes a pairwise covariance matrix for all residuals, along with a set of vectors of covariances between the target points for interpolation and the observed data. With these matrices, RK computes weights to interpolate SVI scores from each municipality to each raster cell. Because we scale from an area (municipality) to a point (center of raster cell), we compute the covariance matrix differently from point-to-point SK. The municipalities are discretized into a number of points, and a covariance function is fit to the discretization of the areal data. Kyriakidis (2004) describes this procedure in greater detail.

RK spatially interpolates the residuals of a georeferenced regression model and proceeds in two steps: a regression model that is subsequently kriged. RK is a multi-equation model and each equation is a model. Both models are spatial but make different assumptions and capture different aspects of space. The kriging model assumes that the data vary continuously over space and are spatially autocorrelated, such that a set of closely clustered locations are more similar to each other than to a set of widely scattered locations. The regression model component of RK ignores spatial factors like shape, size, and distance, and instead assumes that spatial relationships are described entirely by georeferenced features. The accuracy of either model depends, essentially, on its ability to identify spatially homogenous areas (Maantay et al. 2007, Maantay et al. 2008). RK combines both models into one; a topic this section will presently examine.

The kriging interpretation of spatial data is somewhat fraught for demographics, where all data are attached to discrete objects—people—and cannot usefully be viewed as realizations of a smooth and continuous random field. (See Cova et al. [2002] and Kjenstad [2006] for a discussion of the distinction between spatial fields and objects.) Nonetheless, choropleth maps represent a set of expectations for a variable, bounded and constant over an area of space, and it is quite natural to imagine how this could be converted to a smoothly varying function over space. Schmid and Maccannell (1955), and later Tobler (1979), analyzed the production of contour maps from choropleth maps and control points of population density data. Kriging accomplishes the same task within a framework of analysis of spatial autocorrelation.

The input data to RK consists of a matrix of m predictors X and a vector of n observations $\mathbf{y}$ thought to be dependent on X, where typically $m \ll n$. Linear regression analyzes covariance between the columns of X and $\boldsymbol{y}$ using ordinary least squares (OLS) or a Generalized Linear Model (GzLM). Kriging, which models covariance between data points of $\boldsymbol{y}$, assuming these are realizations of random variables with a multivariate Gaussian distribution, is generally equivalent to a Gaussian Process (GP) regression model (Rasmussen and Williams 2006). In addition to data values, kriging requires spatial coordinates of the data, usually written $\boldsymbol{y} = \{y(s_i), \dots, y(s_n)\}$, and $X = \{\boldsymbol{x}(s_i), \dots, \boldsymbol{x}(s_n)\}$ to emphasize that the data are generated by a function with a spatial coordinate. The values of $\boldsymbol{y}$ are measurements at the locations of S, e.g., a volume of mineral, inches of rainfall, or height of a plant. The matrix X is a dataset of co-located measurements that represent independent variables. RK begins with an input of n cases $\{(y(s_i), X(s_i)), \dots, (y(s_i), X(s_n))\}$ and derives a vector of weights for estimating values at an unobserved location s_0.

Regression weights, or regression coefficients, weight the impact of variable $x_j(s_0)$ on the estimate of y_0 at location s_0:

$$\hat{y}(s_0) = \sum_{j=0}^{m} \beta_j x_j(s_0); x_0(s_0) \equiv 1 \tag{1}$$

The constant term x_0 is set to 1. Kriging weights are similar, but rather than weight the effects of variables they instead weight the influence of nearby data points:

$$\hat{y}(s_0) = \sum_{i=0}^{n} \lambda_i y(s_i) \tag{2}$$

The mathematical notation emphasizes that values of $\boldsymbol{y}$ and X are indexed to spatial locations by writing them as functions of location: $y(s_0)$ and $\boldsymbol{x}(s_0)$. We will see that the regression portion of RK does not analyze location or spatial arrangement and is based only on the values generated by those functions. In contradistinction, most forms of kriging do not analyze ancillary data and are based only on analysis of spatial arrangement. RK combines the two methods to analyze both ancillary data and spatial arrangement. We begin by describing the algebra and assumptions of regression before moving on to describe RK.

3.3 Linear regression

OLS finds regression weights that minimize the sum of squared errors:

$$\sum_1^n (y_n - X\boldsymbol{\beta})^2 = \sum_1^n \epsilon^2 \tag{3}$$

The first column of the matrix X are ones for the constant term, the other columns are the data. The weights $\boldsymbol{\beta}$ that minimize $\sum_1^n \epsilon^2$ are a function of the covariance of the data:

$$X^T X \boldsymbol{\beta} = X^T \boldsymbol{y} \tag{4}$$

where $X^T X$ and $X^T \boldsymbol{y}$ are the moment matrices corresponding to the second moments of X and $X^T \boldsymbol{y}$ with $dim(X^T X) = m \times m$ and $dim(X^T \boldsymbol{y}) = m \times 1$.

The matrix $X^T X$ is a system of equations, solving by inverting the moment matrix results in the familiar OLS equation:

$$\boldsymbol{\beta} = (X^T X)^{-1} X^T \boldsymbol{y} \tag{5}$$

The best estimate for an unknown y_0 given input data x_0 is:

$$\hat{y}_0 = \boldsymbol{x}_0 \boldsymbol{\beta} \tag{6}$$

with variance equal to:

$$\hat{\boldsymbol{\epsilon}} = \boldsymbol{y} - \hat{\boldsymbol{y}} \tag{7}$$

$$\sigma^2 = \frac{\hat{\epsilon}^T \hat{\epsilon}}{n-m} \tag{8}$$

OLS imposes a few important assumptions. The first is that $E(\boldsymbol{\epsilon}|X) = 0$, a result of the fact that OLS minimizes the sum of $\boldsymbol{\epsilon^2}$. The first assumption is the most important; it is what guarantees that the model is unbiased and free of confounding effects. Violating the first assumption will invalidate the model. The second assumption is that no variables are linearly dependent; this assumption is required to solve the system of equations (i.e., invert) $X^T X$. A linearly dependent variable is also collinear; hence all but one of the collinear variables may be removed to obtain the assumption of linear dependence. The third assumption is that $\boldsymbol{\epsilon}^2$ has the same conditional variance σ^2 for all observations —i.e., homoscedasticity— and that $\boldsymbol{\epsilon}^2$ is uncorrelated between observations. These two criteria together form the assumption of spherical errors. If the assumption of spherical errors does not obtain, the OLS estimates are still valid but the model is less efficient. This third assumption, in fact, is the essential difference between OLS and kriging methods: kriging relaxes the third assumption by modeling spatial autocorrelation between observations.

Regression weighted dasymetric mapping (RWDM) does not, in general, preserve the original volumes of the source zone data used to estimate values in the target zones. This property is sometimes referenced as coherence or exactitude, and it is an attractive property of kriging. Because RWDM lacks this property, users should usually rescale the estimates so that, when summing across target zones within a source zone, the sums equal the original source zone data. The equation to accomplish this is:

$$\hat{y}_i' = \hat{y}_{i\cdot} \frac{y_i}{\sum_{j=1}^{n_i} \hat{y}_{ij}} \tag{9}$$

where $\hat{y}_i'$ is a rescaled estimate within source zone i, $\hat{y}_{i\cdot}$ is the raw estimate, y_i is the observed value of source zone i, n_i is the number of target zones that intersect source zone i, and $\sum_{j=1}^{n_i} \hat{y}_{ij}$ sums over all j raw estimates in source zone i.

3.4 Kriging

RK analyzes data in two steps. The first step is essentially identical to the OLS model described in the previous section. Rather than keep the fitted values, however, RK will map the residuals:

$$\hat{\boldsymbol{\epsilon}} = \hat{\boldsymbol{y}} - \boldsymbol{y} \tag{10}$$

The second step of RK interpolates the residuals using kriging, with the ultimate effect of allowing the mean of the kriging system to vary based on the OLS model output. Simple kriging (SK) and ordinary kriging (OK) are the two most basic forms of kriging, and both assume the data are generated by an intrinsically stationary process with a constant but known (SK) or unknown (OK) mean over a sample area. Residuals by definition have mean zero, hence RK uses SK to interpolate the residuals over an area. All forms of kriging assume that the observed data $\{y_i, \dots, y_n\} = \{y(s_i), \dots, y(s_n)\}$ are a single sample from a multivariate GP at locations $S = \{s_i | i = 1, \dots, n\}$.

SK is simpler than OK. The following list gives the data generating equations modeled by SK, OK, and RK for comparison purposes:

1. simple: $y_i = e_i$
2. ordinary: $y_i = \mu + e_i$
3. regression: $y_i = x_i^T \beta + e_i$

Although the only difference between SK and OK here is the mean parameter μ, this has important consequences for the objective functions that OK must minimize, and for the properties of the resulting model. OK assumes only intrinsic stationarity and models a constant local mean μ_{x_0}. SK, on the other hand, assumes a known mean and thus a globally stationary constant μ that can be shifted to zero without loss of generality. Lacking a prior known mean, OK also has no way to compute a covariance function and instead uses a variogram to model covariance. Under most circumstances, all major properties of covariance functions carry over to variograms (Gneiting et al. 2000, Wackernagel 2003). A final important difference is that OK uses a Lagrangian multiplier to constrain the kriging weights to sum to one. The role of the Lagrangian in OK equations can be difficult to visualize geometrically. In this case, the Lagrangian can be understood as constraining the optimization of the OK equations to a point in the $dim(\boldsymbol{\lambda}) = n$ dimensional space where the kriging weights sum to one. The kriging weights are similar to those for a weighted mean, and if they do not sum to one then some values will be over- or under-weighted, yielding a result that is biased from the mean. See Wackernagel (2003, chapter 11) for a formal derivation of the OK system of equations using a Lagrangian multiplier.

The regression kriging equation above is oversimplified for comparison purposes; the complete RK equation is:

$$\hat{y}(s_i) = \boldsymbol{x}(s_i)^T\boldsymbol{\beta} + \boldsymbol{\lambda}_i\left(x(s_i)^T\boldsymbol{\beta} - y(s_i)\right) \tag{11}$$

The equations in the list omit the spatial indexes and collapse the second term in the RK equation to e_i: the interpolated residual. RK is further subdivided based on how the β parameters are obtained. If parameters are derived external to the RK model and provided as input, this method is conventionally referred to as UK (Pebesma 2006, Hengl et al. 2007). If, on the other hand, the parameters are fit along with the coordinates of the data locations as part of the kriging equations, then this method is usually referred to as KED. In most cases, the two methods yield identical results; however UK, because it accepts externally derived parameters, can be used to spatially extrapolate data to new locations. The ability to fit a RK model in one location and apply it in another raises important technical and ethical issues that will be revisited for a discussion within the concluding section. For derivations of SK and OK, see Wackernagel (2003) or Gotway and Stroup (1997), and for a comparison of RK, UK, and KED, see

Hengl et al. (2007) or Wackernagel (2003). SK is equivalent to GP regression as described in Rasmussen and Williams (2006).

From the data, SK builds a matrix, a row vector, and scalar:

$$\begin{bmatrix} c(s_1, s_1) & \cdots & c(s_1, s_n) \\ \vdots & \ddots & \vdots \\ c(s_n, s_1) & \cdots & c(s_n, s_n) \end{bmatrix} \tag{12}$$

$$v_0 = [c(s_0, s_1), \dots, c(s_0, s_n)] \tag{13}$$

$$v(s_0) = c(s_0, s_0) \tag{14}$$

The function c is the covariance function —or kernel, in GP terminology. Kriging assumes both first and second order stationarity— so that all data points share V and $v(s_0)$. The matrix V is the $n \times n$ covariance matrix for all observations. The vector v_0 is the $1 \times n$ row vector of variances between the coordinates s_0 and those in S; it is subscripted to emphasize that this vector is recomputed for each s_0. The scalar $v(s_0)$ is the maximum variance σ_y of y_0 at $y(s_0)$, and because we assume a stationary second moment, it also the maximum variance of all the measurements; therefore it is equal to the values on the diagonal of V. Although the covariances of V are functions only of the distance values in S, the kernel c has a number of free hyperparamters (e.g., σ_y, the maximum variance of $\boldsymbol{y}$) that must be estimated from $\boldsymbol{y}$. Indeed, the assumptions of constant mean and variance is equivalent to assuming that the data generation process is sufficiently homogeneous over the area of interest to allow for mean and variance to be estimated from the data; this is what allows the hyperparameters to be estimated from empirical data and to build the matrix V. As mentioned previously, OK does not assume a known mean and models V using a (semi)variogram rather than a covariance function; this distinction is academic for most, however, and most practitioners can treat these two things identically. It is critical to note that, other than hyperparameters, the data in $\boldsymbol{y}$ are not used to compute covariance matrix V and friends; these constructs represent only the spatial correlation in the data.

The kriging weights are derived by solving the system of equations:

$$V\boldsymbol{\lambda}_0 = v_0 \tag{15}$$

$$\boldsymbol{\lambda}_0 = v_0 V^{-1} \tag{16}$$

The probability of a particular estimate for y_0 is given by:

$$y_0|\boldsymbol{y} \sim N(v_0 V^{-1}\boldsymbol{y}, v(s_0) - v_0^T V^{-1} v_0^T) \quad (17)$$

where the best estimate $\hat{y}_0$ is the mean parameter of the normal distribution N (the first parameter.)

Notice that the kriging weights, $\boldsymbol{\lambda}_0 = v_0 V^{-1}$, depend on the location $\boldsymbol{s}_0$ via v_0 and weight the impact of y on the estimate $\hat{y}(s_0)$. The weights are subscripted to emphasize that they are computed for each s_0. A consequence of the assumption of a constant mean over the sample area is that $E[\hat{y}_0 - y_0] = 0$, or that the estimate is unbiased. Also notice that, as the covariance is stationary (the second parameter), it too does not depend on the values of $\boldsymbol{y}$. The two most important aspects of these equations are that (1) the kriging weights only depend on the locations in A and (2) they model covariance between these locations.

The kriging weights $\boldsymbol{\lambda}_0$ are similar in function to the regression weights $\boldsymbol{\beta}$: β weights the effect of the variables in X on y_0, whereas $\boldsymbol{\lambda}_0$ quantify the influence of the observed values in $\boldsymbol{y}$ given their distance from the unobserved location s_0. This difference implies another one of importance: the values of $\boldsymbol{\beta}$ are the same for any given $\boldsymbol{x}(s_0)$, whereas kriging weights depend on the location s_0; kriging fits a regression model for each input point, and must develop the vector v_0 for each point. One might say another difference is that OLS weights imply a varying mean, whereas kriging assumes a constant mean; however, this comparison is ill-posed. Kriging allows for varying deviation from the mean, and this property is a better comparison with OLS' varying mean. Moreover, the constant mean of kriging is more akin to the intercept, or constant, term of an OLS model. A similarity between both kriging estimates and those provided by OLS are that all estimates are assumed to have the same variance, as dictated by kriging's assumption of second order stationarity, and OLS' homoscedasticity of variance. This assumption regarding variance for the system of kriging equations should not be confused, however, with the prediction variance of kriging, given by:

$$v(s_0) - v_0^t V^{-1} v_0 \quad (18)$$

The same as the equation for the variance parameter of N (the second parameter) of the kriging prediction given in equation 3. Prediction variance of kriging expresses the uncertainty of an estimated value as a function of its distance from observed values. Homogeneity of variance, on the other hand, refers to the spherical variance of the covariance matrix V.

The matrix V and friends are also similar to the moment matrices of OLS. The comparison between GP and linear regression is usually drawn between GP and Bayesian Linear Regression (BLR), as BLR is identical to GP with a linear kernel. BLR also relates to OLS via the equation for the posterior distribution $\text{LIKELIHOOD} \times \text{PRIOR}$. OLS and BLR share identical values for maximum likelihood and as n, the number of cases, increases towards infinity, the likelihood "washes out" the priors and the estimates of OLS and BLR converge. What GP does differently is that, whereas BLR analyzes distributions over parameters, GP analyzes distributions over functions (Williams 1997). The distinction is clear with a linear kernel $\sigma^2\phi(X^T)\phi(X)$, where ϕ is a basis function (possibly identity). GP with a linear kernel is BLR with priors over functions. The matrix form of the linear kernel is quite familiar from the OLS equation $\boldsymbol{\beta} = (X^TX)^{-1}X^T\boldsymbol{y}$, from which we may derive the covariance of the weights around their true values to be $E\left[(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\right] = \sigma^2(X^TX)^{-1}$, which we can see is identical to the linear kernel (with identity basis). In essence, GP with a linear kernel would be similar to linear regression that computes moment matrices as $dim(XX^T) = n \times n$ rather than $dim(X^TX) = m \times m$. Kriging rarely uses a linear kernel, however, for which instead conventional kernels include exponential, spherical, Gaussian, and Matern.

The above brief summary of RK, and comparison of the regression and kriging components, highlights that RK combines two supplementary types of spatial models. The regression model analyzes only feature space, georeferenced measurements of landcover, elevation, population, etc.; the kriging model analyzes only geographical space, distance, shape, and size. Regression or kriging in isolation can perform downscaling, but an RK model will outperform either of its components alone (Bourennane et al. 2000).

3.5 Mapping SVI in the Philippines

RK can help researchers to improve visualization of their data and build models that combine data from multiple sources and spatial resolutions. This final section will demonstrate RK in an application to downscale an SVI to the Philippines at a 1-km scale. This section will review the SVI and its construction and consider independent predictors for the SVI. In addition to the SVI factors, the discussion will also analyze how geographical arrangements can be converted into metrics that are predictors of social vulnerability and other socioeconomic concepts.

The unit of analysis of an SVI is the family; however it has limited validity when applied to single families, and the values are more usefully aggregated to a geospatial area so that SVI is a property of a community that encompasses the people, housing, and labor available to a spatial location. The Philippines SVI data are attached to administrative divisions, to which they can be aggregated and geospatially located.

Along with the SVI, we have a set of geographical predictors at the target scale and will use RK to rescale the SVI data to this new scale. Some of these data are observations of the physical environment, often collected and compiled for years or decades, and available preprocessed and in a format designed for analysis. These could include climatology, land cover, crop and livestock, human population density, and nighttime lights, among others. There also exist complex data on the human-built environment, including buildings, roads, utilities, and infrastructure. These latter data facilitate computation of geodemographic metrics reflecting accessibility to social and economic resources. Pebesma (2006) notes that such metrics are “cheap” (quotes in original) because this kind of data is generally freely available, e.g., OpenStreetMap (OpenStreetMap contributors 2018).

Table 2 lists the predictors for a RK model of SVI. All predictors have been resampled to identical grids of 1 km resolution. The landcover predictors are both masks of the associated landcover class from Basevue (MacDonald, Detwiller, and Associates 2013). Nighttime lights uses processed Near Constant Contrast data collected by the Visible Infrared Imaging Radiometer Suite (VIIRS) polar orbitor. Population density data is from LandScan (UT-Battelle 2019) and represents daytime population density rather than population density of residences. Weiss’ accessibility to cities metric (Weiss et al. 2018) is in minutes of travel to the nearest city, with travel times weighted by road availability. Most metrics used in the SVI RK model are not spatial computations or even zonal statistics, but merely observations at a point in space. This also seems to be the case for most spatial models.

Table 2. Input data to an RK model of social vulnerability.

Variable	Source	Vulnerability Relation
Urban landcover	MacDonald, Dettwiller, and Associates Ltd. (2013)	Accessibility
Agricultural landcover	MacDonald, Dettwiller, and Associates Ltd. (2013)	Socioeconomic status
Nighttime lights	Elvidge, Baugh, Zhizhin, Hsu, & Ghosh (2017)	Infrastructure
Population density	UT-Battelle (2019)	Socioeconomic status
Access to cities	Weiss et al. (2018)	Accessibility

Table 2 lists input data to essentially a georeferences OLS model of social vulnerability. The dependent variable, SVI, is georeferenced to polygons, but is rasterized to the same 1 km grid as the predictors. In our case, we will use KED to fit the values, but as discussed, the results are very similar to UK. The following equation would specify the model:

$$y = \alpha + \beta_1 x_{pop} + \beta_2 x_{pop}^2 + \beta_3 x_{urban} + \beta_4 x_{ag} + \beta_5 x_{travel} + \beta_6 x_{nightlights} \quad (19)$$

The equation is quadratic in population density to allow for the effect of population density on SVI to vary with population density. The regression step computes the residuals $x(s_i)^T\boldsymbol{\beta} - y(s_i)$ for a regular, 1-km grid and fits the kriging model to these residuals, using the RK method described in section 3.5.

A block of geographical raster data containing $\{(y(s_i), X(s_i)), \ldots, (y(s_i), X(s_n))\}$ for all s within the entire extent of the Philippines serves as input data for the RK model. The computational procedures are slower than OK or SK, due to the calculation of the area variogram. The RK model described in this section is fit using gstat, and a version of the area variogram modified to run in parallel. A Linux cluster with 46 cores requires about 8 hours to downscale this amount of data, including time for raster preprocessing.

Figure 1 displays the results of downscaling social vulnerability. Associated with the model are interpolation uncertainty quantifications, plotted around the National Capital Region in Figure 2b. This represents uncertainty due to the sizes of the input polygons. Larger polygons and those further from other polygons have greater interpolation uncertainty. Notice that islands and areas associated with Central Luzon have higher uncertainty. The plot margins are the row and column means of the plotted values. These can help highlight isolated and outlying areas that viewers might miss, for example the spots of high vulnerability in Palawan and Muslim Mindinao. Interpolation uncertainty might help with decision making, but ultimately model results are based on assumptions not testable within the framework of the model. If we had the information necessary to assess the accuracy, we would not need to solve the COSP in the first place. The map shown in Figure 1 is a best linear unbiased prediction (BLUP) based on our knowledge of how spatial features may correlate or even cause social vulnerability. By “best” we mean that no linear function exists able to achieve better prediction error. We can improve on it with focused research that tests and establishes relationships between social vulnerability and geographic space and using datasets with wide geographic coverage to expand the applicability of resulting models to broader, hopefully global, ranges.

Figure 1. Social vulnerability, Philippines, 1 km scale.

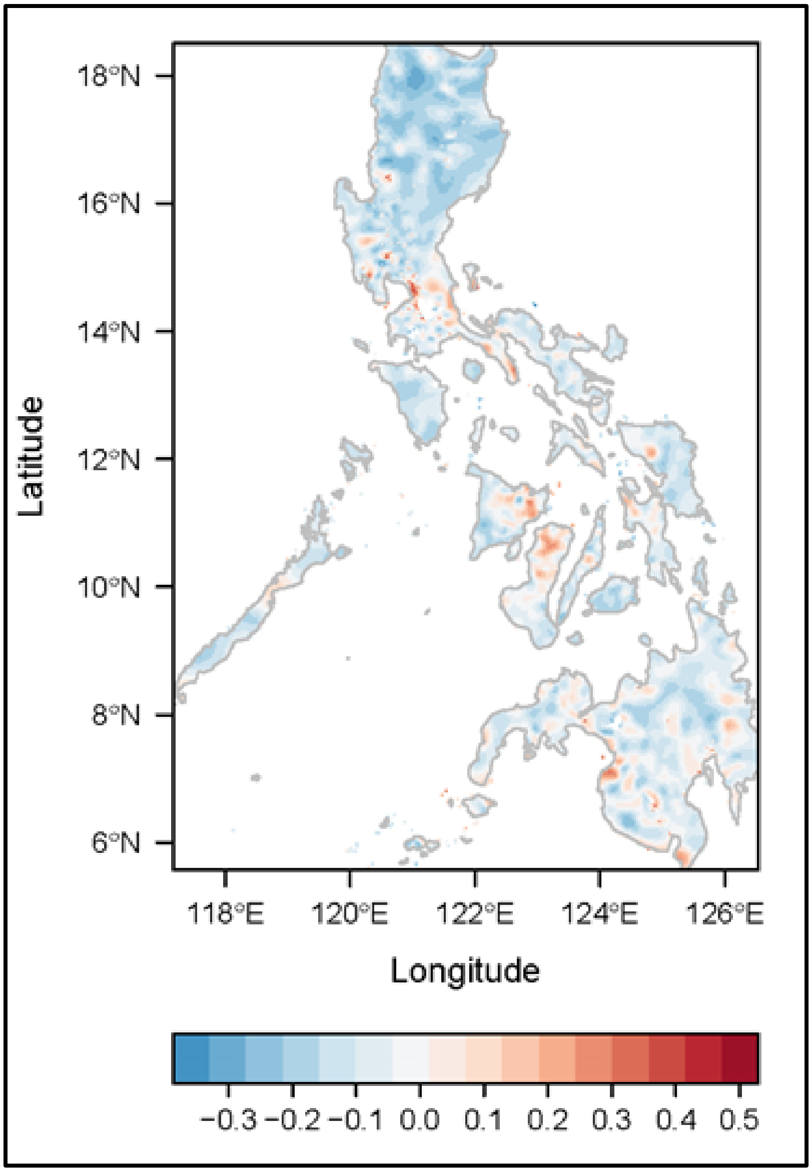
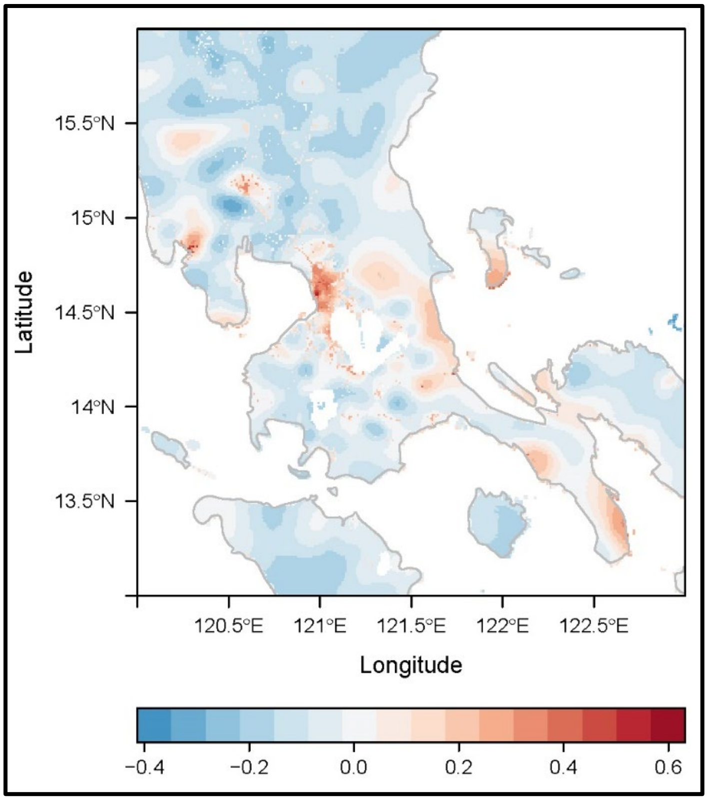
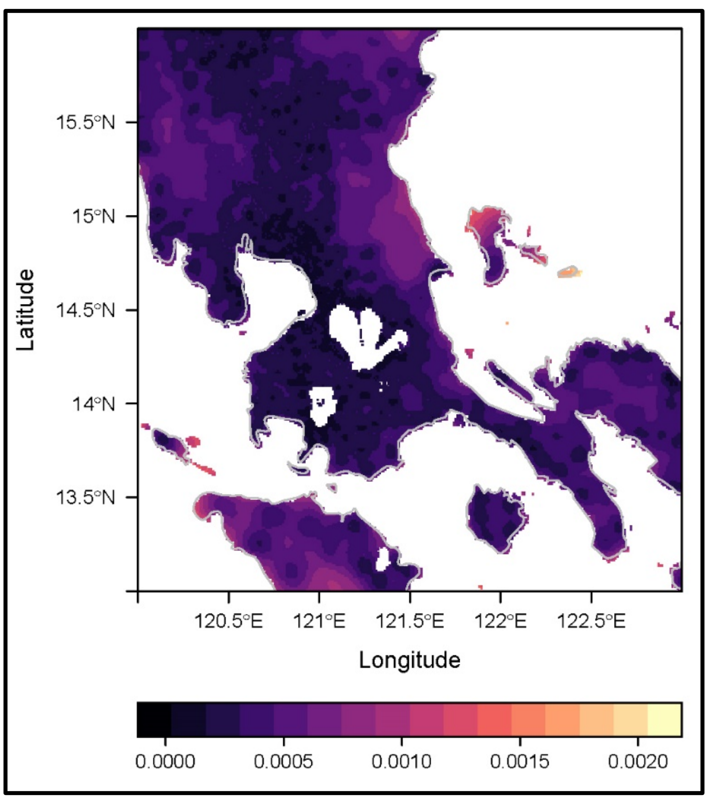


Figure 2. (a) Social vulnerability, 1 km scale (b) Interpolation uncertainty, 1 km scale.



(a)



(b)

Because these residuals are calculated from a model with greater geographic detail, they will share the same levels of geographic variability as the source model, and furthermore will assume the same relationship between the features of the source regression model and the dependent variable. Thus, if some features have a different relationship in the source location, then the predictions of the model may be systematically biased. Despite this additional level of detail in the residuals, the kriging weights remain based on the geographical aggregation of the dependent variable at the target location. This means that the details will be smoothed out by the size of the area to which the dependent variable is aggregated, and the uncertainty of the interpolated values will be greater. In the event a sensible covariance function cannot be devised or fit to the data in the target location, e.g., if the data are georeferenced at too great a level of aggregation, then the covariance function will have a large "nugget" or a variance in the measured value greater than zero as distance to the measurement approaches zero; then the SK step of UK will still fail or be unreliable.

3.6 Developing a geospatial model of the SVI

The RK model links SVI to geographical factors (listed in Table 3), including urban land cover,* population density (UT-Battelle 2017), nighttime lights (Elvidge et al. 2017), and accessibility to cities (Weiss et al. 2018). This model outputs a geographically interpolated SVI; the same SVI as that on which the RK model is trained. Such a model offers two advantages. First, the SVI is downscaled to the resolution of the geographical factors. Second, the SVI can be extrapolated to locations outside those where the RK model was trained.

Table 3. Predictors of SVI for RK model.

Variable	Source	Vulnerability Relation
urban landcover	(MacDonald, Dettwiller and Assoc. Ltd., 2013)	accessibility
agricultural landcover	(MacDonald, Dettwiller and Assoc. Ltd., 2013)	socioeconomic status
nighttime lights	(Elvidge, Baugh, Zhizhin, Hsu, & Ghosh, 2017)	infrastructure
population density	(UT-Battelle, 2019)	socioeconomic status
access to cities	(Weiss et al., 2018)	accessibility

* Visual Navigation Land Cover. Produced by MDA Information Systems, LLC under contract to NGA.

4 South Africa and SVI Case Studies

4.1 Mapping social vulnerability

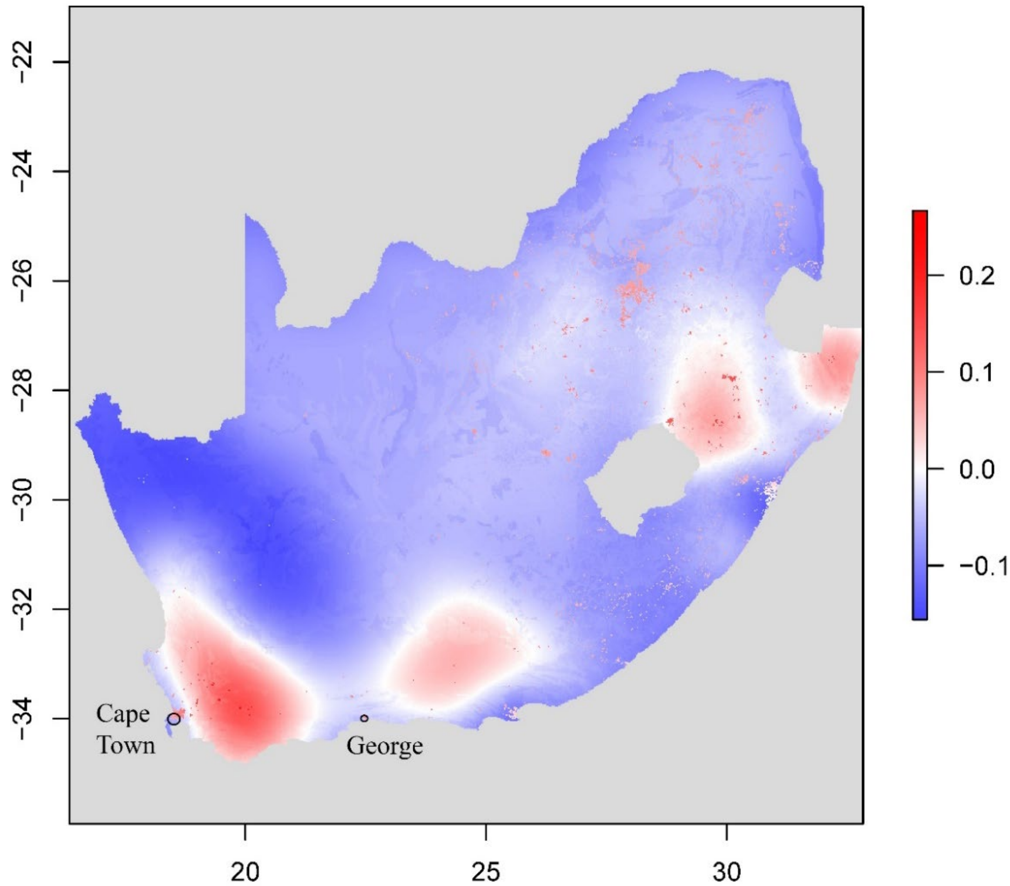
If one is comfortable with the assumptions of RK, it can be extrapolated via UK to locations beyond the area where the model was fit. First, use KED to fit a RK model and retain the beta estimates. Next, supply these estimates to a UK model targeting the new area, which will use them to predict SVI before calculating residuals and kriging these values via SK with available georeferenced data. Looking again at equation **11**, we see that, when extrapolating the model in a new location, the calculated residuals will be the new observed $\boldsymbol{y}$ values, at whatever level of aggregation, subtracted from the predicted value at the target resolution: $x(s_i)^T\boldsymbol{\beta} - y(s_i)$.

Figure 3 shows the results of applying the Filipino model to South African geography. Clearly the downscaling has much less detail than that of the Filipino data, and this reflects the much greater spatial uncertainty for the observed SVI in South African data. Without the Filipino model to make initial estimates, however, the so-called nugget of the spatial kriging models would dominate the results. A nugget is a non-zero value for spatial variance as distance to y_0 approaches zero. A very large nugget leaves very little space for variance between the nugget and maximum variance —a.k.a. sill— resulting in a more uniform interpolation that is not useful for reaggregating the data to alternative boundaries.

4.2 Disaster-driven migration in mixed methods assessment

With the SVI model created and mapped, we tested our hypothesis that the socially vulnerable continue to dwell in disaster-prone areas. For qualitative context, we examined two cities in South Africa: one that has seasonal flooding each year (Cape Town) and one that had catastrophic flooding in 2006 and 2007 (George). Despite annual floods in Cape Flats (Cape Town), there is no mass migration from that area. Instead, residents deal with the flooding the best they can.

Figure 3. Social vulnerability across South Africa.



Cape Flats is a municipal region established during apartheid when black and colored people were moved from the city's core.* Located on low-lying, waterlogged wetlands, most of Cape Town's informal settlements are found here (Waddell 2016). The residents of Cape Flats are often subjected to ponding (pluvial) floods or ground-water floods, rather than catastrophic river flooding. The seasonal ponding results from seepage due to water rising to the ground surface because of a higher water table (Els 2011). The qualitative literature asserts that people cope with the floods in a variety of ways, such as raising communal areas with bricks placed on the ground, digging trenches for drainage, and putting plastic on roofs (Waddell 2016). Residents stay in the flats year-round to maintain access to the nearby city center and the promise of jobs—while remaining in their homes during the flooded winter months for fear their residences will be burglarized if they are relocated to shelters (Drivdal 2016).

* Colored people are a multiracial ethnic group, with Indians/Asians being the fourth ethnicity included in the South African census.

Although Cape Flats consistently floods each year, the lack of outmigration provided a baseline for our study. The residents of the flats acted as our control group because they are a population that does not move despite season floods. To test our hypothesis that the socially vulnerable do not move because of their limited economic means, we needed an experimental group that was subjected to more extreme conditions.

For comparison, we used the example of catastrophic flooding in the smaller city of George, also located in the Western Cape Province. The 2006 and 2007 floods occurred here within the timeframe of migration history available in the South African census and provided a case that we may examine using both quantitative metrics and qualitative research. The damage in 2006 was 1.5 million Rand and the following year's floods created 2.1 million Rand in damages (Faling et al. 2012); both damage assessments were adjusted for inflation. Rainfall for 2006 was nearly twice the previous maximum amount, as measured since 1941 (Mélice and Reason 2007). The deluge, combined with longer drought, resulted in drier conditions with increasing runoff after heavy downpours (Faling et al. 2012). Despite the flooding rivers and landslides, "only in isolated incidents were informal dwellings completely destroyed. However, most informal dwellers had to reconstruct a small or large part of their dwelling" (Benjamin 2016, p 147). The concern of the residents was not the flooding of Skaapkop River, but the overland runoff (Benjamin 2016, p 153).

In examining the quantitative census data to see if there had been migration from the George during this period of flooding, we mapped the proportion of immigrants for each of the provinces (see Figure 4). Although difficult to untangle relocation driven by work or disaster, the circulation of people occurs within provinces rather than between provinces. The qualitative evidence indicates that, at the very least in the case of flooding, victims migrate using networks close to home for assistance (e.g., within municipality).

4.3 Testing a model of SVI and migration from flood-prone areas

Much of the IPUMS data provide information on where people migrated from, along with a time range of when they migrated. The South African data include questions about province or state of residence 5 and 10 years ago. We used this information to test hypotheses regarding the impacts of flooding on migration by vulnerability as measured in the SVI.

Flooding is operationalized using the PREVIEW Global Risk Data Platform. PREVIEW serves flood risk data developed by United Nations Environment Programme/Global Resource Information Database (UNEP/GRID) – Geneva (Peduzzi n.d.). UNEP/GRID sources flood data from the Emergency Events Database (EM-DAT n.d.) and processes the data to calculate a raster of population at risk of flooding.

SVI indices for South Africa are estimated with a UK model, as described in the previous sections. As input, a UK takes both data and parameters for predicting SVI from geographical data using a regression model; our parameters for the UK model are supplied by a KED model previously fit to the Philippines. UK then interpolates the residuals of the regression model, using available South African census data to compute residuals. UK allows us to achieve greater levels of spatial detail when mapping SVI in South Africa, compared to using only South African census data, which again has much less spatial detail than the Philippines data.

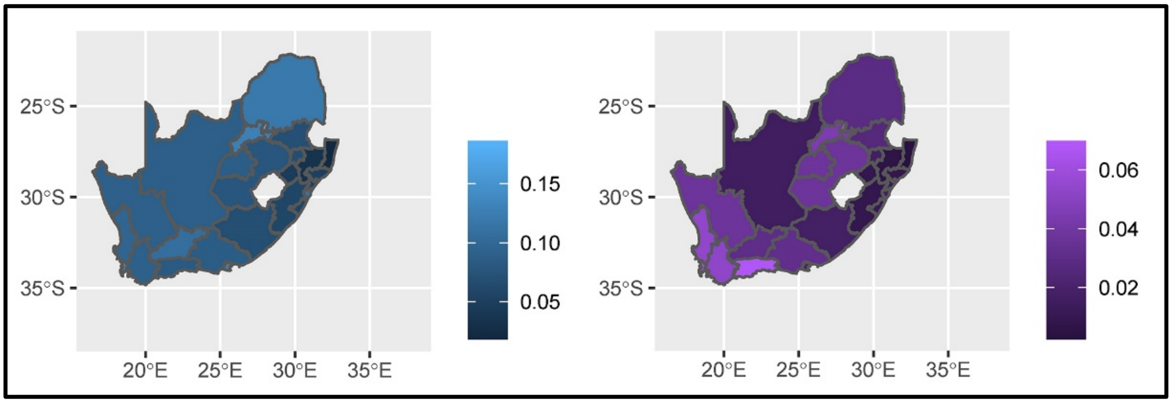
To investigate the impacts of SVI and flooding on migration, we divide migration data into two classes: within- and between-region internal migrants. The dependent variable, migration, is transformed to indicators for within- and between-region migration, and these indicators are aggregated to a municipal boundary. Within-region migrants have moved only within the geographical census units detailed in Table 4. Between-region migrants have moved across the border of these geographical units. The data combine both 2001 and 2011 South African census data but do not model time explicitly.

Table 4. Province divisions for determining within- and between-region migrants.

Census unit	Geopolitical unit(s)
West Cape	West Cape
Free State	Free State
Eastern Cape	Eastern Cape, KwaZulu-Natal
Northern Cape	Gauteng, Limpopo, Mpumalanga, North West, Northern Cape

To ensure that the cases from both censuses are comparable, we modify the 2011 migration history response data to cut-off at 5 years before the census date; any migration history more than 5 years in the past is considered "no migration." This matches the cut-off of the 2001 census data, which has a question that asks about migration within the past 5 years. Figure 4 shows maps of relative proportion of within- and between-region migrant populations in each South African district. The map to the right shows the relative proportion of population moving to the district from other provinces, as a percentage of total movement. The map to the left of Figure 4 shows the percentage of people migrating within a province. To evaluate the hypothesis that these migration patterns are influenced by SVI and flooding, we will inspect the significance, direction, and magnitude of the effects on out-migration associated with the above three independent variables.

Figure 4. Percentage of total migrant movement between-province (right), and within-province (left).



All data —migration, SVI, and flooding— are now represented as a raster. To fit the model, we aggregate all data to South African municipality boundaries, obtained from the South African Municipal Demarcation Board (MDB 2019). Here we highlight the advantage of the UK model. All our data sources use different supports, i.e., they each are georeferenced to different administrative boundaries or raster resolutions. The UK model has provided a coherent best linear unbiased estimate of SVI for each square kilometer across South Africa. Such data are trivial to reaggregate to an arbitrary set of supports, e.g., an alternative set of administrative boundaries. This facilitates our combination of multiple spatial datasets, each with a different support.

The hypothesis test occurs at the level of municipalities obtained from the MDB. The UK model has downscaled the SVI to a 1-km grid across South Africa, and from this grid, SVI values are aggregated to municipalities. Migration metrics are aggregated to municipality districts, which are demarcated by the boundaries in the maps shown in Figure 4. The MBD municipalities are nested with the municipality districts, and obtain values for the migration metrics from their parent district. These two rescaling operations yield estimates for all relevant data at the municipality level, as determined from MBD. The municipality level models are specified by:

$$y_{in} = \alpha + \beta_{svi}x_{svi} + \beta_{fr}x_{fr} + \epsilon \tag{20}$$

$$y_{bw} = \alpha + \beta_{svi}x_{svi} + \beta_{fr}x_{fr} + \epsilon \tag{21}$$

The svi and fr subscripts indicate parameters and data for SVI and flood risk, respectively. The dependent parameters are within and between municipal district migration.

Table 5 lists the results from a model of within- and between-district migration in South Africa. Column two of Table 5 contains results for the within-district migration model, and column three for between-district migration. For our hypothesis, we forecast social vulnerability to have no effect or a negative effect on within-district migration. This supposition is based on the South African case studies in Cape Flats and George that found families do not migrate or that they use short-range, short-term, within-municipality, or within-district migration to cope with disasters. We expect that the socially vulnerable have less capital —either economic or social— and consequently cannot migrate only for the purpose of leaving disaster areas. This is exactly the result we observe; hence we cannot reject the possibility that the SVI is a factor-shaping spatial arrangement of vulnerability. We cannot assert for certain that this occurs because those with higher vulnerability are unable to migrate, but such an interpretation is consistent with the model.

Table 5. Models of internal migration by SVI and flood risk.

Variable	Within-region migration	Between-region migration
Intercept	-2.231(0.003)	-2.862(0.003)
Social vulnerability	-0.044(0.021)*	0.006(0.027)
Flood Risk	-0.001(0.003)	-0.002(0.004)

4.4 Further analysis of the SVI and a lack of migration

Despite the disastrous 2006 flooding in George, which included landslides, burst dams, and 11 people washed away and drowned (Flanagan 2006, Mélice and Reason 2007), there was no significant outmigration from the Eden District. An additional reason for the lack of disaster migration includes the low social vulnerability surrounding George (Figure 3). Although there are wide economic gaps between the rich and poor —as in most of South Africa due to the echoes of apartheid— this region also supports a growing tourist industry. The Eden District draws significant numbers of tourists each year; the population expanded sixfold during the December peak season in 2005 (Tempelhoff et al. 2009, p 98).

The strong influx of tourism combines with a significantly larger proportion of South Africans moving from outside the province to the Eden District (Figure 4). The region, which is often called the Garden District because of the pristine landscape, attracts retirees from other parts of the country (Templehoff et al. 2009, p 98). As the national distribution of income still favors whites (Wilkinson 2017), the breakdown by race in George also indicates the reason behind low social vulnerability of the region.

In 2001, the census data indicate that the population was comprised of 50% Colored (compared to 8.9% nationally), 0.5% Asian (2.4% nationally), 27.5% African (79.2% nationally), and 22.5% White (9.5% nationally). At approximately the same time of the floods, a national analysis of relative capital found that whites were still at the top of per capita incomes. Asians, whose incomes were 60% of the whites' incomes, were the next significant group, while coloreds and blacks had incomes that were 22% and 13% of the whites' incomes, respectively (Leibbrandt et al. 2010). With larger populations of whites and coloreds in George, the higher income levels result in a resilient infrastructure. Even informal settlement dwellers benefit to some degree from a larger tax base, especially as employees for the agricultural and service jobs that result.

5 Conclusions

The research presented in this report is motivated by demands to build the capacity to analyze disaster-driven migration, and vulnerability to disasters more generally, worldwide. A core set of data is available worldwide, but the data resolution varies greatly depending on its source. Other valuable data are available only in specific regions, and we wish to make use of this information where available. This report explored the application of an interpolation model in a location with data of lower resolution, and the use of the data to test hypotheses regarding key variables of interest that may explain how people use, or do not use, migration as a response to disasters.

The observation that people do not leave Cape Town Flats due to ponding or ground-water floods provided this study with a control group by which to empirically evaluate the results of our theoretical model. However, examination of the experimental group in George indicated that victims of catastrophic flooding also stay in the area to access local or regional support networks. Consequently, we must look at larger catastrophes to see if disaster-related flooding (e.g., coastal, river, or flash floods) is a driver of migration that thereby forces people out of a region. The 2019 floods in Mozambique from Cyclone Idai would provide such an example. Although the census data do not currently exist, catastrophic flooding at this level could test our theory that most people do not move out of an area permanently.

The result of hypothesis testing is only one point of interest. We also evaluated our results for their ability to work with spatial vulnerability in areas with high spatial uncertainty. We have demonstrated how analysis proven on data with low spatial uncertainty can be used to improve analysis in areas with higher spatial uncertainty, especially when supported with ground truth through qualitative data.

Another test of disaster migration and social vulnerability would be to look at drought. Longer periods of drought with outbursts of severe precipitation happen more often due to climate change, and Africa is at high vulnerability because of lacking infrastructure and economic capacity (Carrão et al. 2016). Cycles of drought and flooding have occurred in both Kenya and Somalia, providing opportunities for further study. Kenya is currently gathering data for its census. Researchers could use that census data to (1) extrapolate to Somalian refugees who have migrated to the country because of violence and drought, and (2) analyze disaster migration within Kenya. Future research could build on the experience of generalizing between two nations, with two different social contexts and two different levels of spatial uncertainty, and then generalize the procedure to a regional and global scale.

References

Abel, Guy J., Michael Brottrager, Jesus Crespo Cuaresma, and Raya Muttarak. 2019. "Climate, Conflict and Forced Migration." *Global Environmental Change* 54: 239-249. doi:10.1016/j.gloenvcha.2018.12.003.

Amos, B., M. P. McDonald, and R. Watkins. 2017. "When Boundaries Collide: Constructing a National Voting Database of Demographic and Voting Statistics." *Public Opinion Quarterly* 81:385-400.

Barrozo, L. V., R. P. Perez-Machado, C. Small, and W. Cabral-Miranda. 2016. "Changing Spatial Perception: Dasymetric Mapping to Improve Analysis of Health Outcomes in a Megacity." *Journal of Maps* 12 (5):1242-1247.

Benjamin, Mogammad Ameen. 2008. "Analysing Urban Flood Risk in Low-Cost Settlements of George, Western Cape, South Africa: Investigating Physical and Social Dimensions." MSS thesis, University of Cape Town.

Bourennane, H., D. King, and A. Couturier. 2000. "Comparison of Kriging with External Drift and Simple Linear Regression for Predicting Soil Horizon Thickness with Different Sample Densities." *Geoderma* 97(3-4):255-271.

Carrão, Hugo, Gustavo Naumann, and Paulo Barbosa. 2016. "Mapping Global Patterns of Drought Risk: An Empirical Framework Based on Sub-National Estimates of Hazard, Exposure and Vulnerability." *Global Environmental Change* 39:108-124. doi:10.1016/j.gloenvcha.2016.04.012.

Cova, T. J., R. L. Church, and M. F. Goodchild. 2002. "Extending Geographical Representation to Include Fields of Spatial Objects." *International Journal of Geographical Information Science* 16(6):509-532.

Cressie, N. 1996. "Change of Support and the Modifiable Areal Unit Problem." *Geographical Systems* 3:159-180.

———. 1990. "The Origins of Kriging." *Mathematical Geology* 22(3):239-252.

Cutter, Susan L., Christopher G. Burton, and Christopher T. Emrich. 2010. "Disaster Resilience Indicators for Benchmarking Baseline Conditions." *Journal of Homeland Security and Emergency Management* 7(1). doi:10.2202/1547-7355.1732.

Drivdal, Laura. 2016. "Flooding in Cape Town's Informal Settlements: Conditions for Community Leaders to Work towards Adaptation." *South African Geographical Journal* 98(1):21-36. doi:10.1080/03736245.2015.1052839.

Eicher, C. L., and C. A. Brewer. 2001. "Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation." *Cartography and Geographic Information Science* 28(2):125-138.

Els, Zelda. 2011. "Data Availability and Requirements for Flood Hazard Mapping in South Africa." Master of Science (MS) thesis. Stellenbosch, South Africa: Stellenbosch University.

Elvidge, Christopher D., Kimberly Baugh, Mikhail Zhizhin, Feng Chi Hsu, and Tilottama Ghosh. 2017. “VIIRS Night-Time Lights.” *International Journal of Remote Sensing* 38 (21):5860-5879. doi:10.1080/01431161.2017.1342050.

EM-DAT. “The Emergency Events Database” Ottignies-Louvain-la-Neuve, Belgium: Université Catholique de Louvain. www.emdat.be.

Faling, Willemien, Johann W. N. Tempelhoff, and Dewald van Niekerk. 2012. “Rhetoric or Action: Are South African Municipalities Planning for Climate Change?” *Development Southern Africa* 29(2):241-257.

Flanagan, Barry E., Edward W. Gregory, Elaine J. Hallisey, Janet L. Heitgerd, and Brian Lewis. 2011. “A Social Vulnerability Index for Disaster Management.” *Journal of Homeland Security and Emergency Management* 8(1):1-22.

Flanagan, Louise. 2006. “Bodies Recovered After George Floods.” *IOL News*, 2 August 2006. https://www.iol.co.za/news/south-africa/bodies-recovered-after-george-floods-287803.

Fleishman, Rachel, and Sherri Goodman. 2018. “Climate Change and the U.S. Military.” *Defense Dossier* Issue 23.

Gelfand, A. E., L. Zhu, and B. P. Carlin. 2001. “On the Change of Support Problem for Spatiotemporal Data.” *Biostatistics* 2(1):31-45.

Gneiting, T., Z. Sasvari, and M. Schlather. 2000. *Analogies and Correspondences between Variograms and Covariance Functions*. National Research Center for Statistics and the Environment-Technical Report Series (NRCSE-TRS) No. 56. Seattle, WA: NRCSE. http://www.nrcse.washington.edu/pdf/trs56_covar.pdf.

Gotway, C. A., and W. W. Stroup. 1997. “A Generalized Linear Model Approach to Spatial Data Analysis and Prediction.” *Journal of Agricultural, Biological, and Environmental Statistics* 2(2):157-178.

Greenhill, Kelly M. 2010. *Weapons of Mass Migration: Forced Displacement, Coercion, and Foreign Policy*. Ithaca, NY: Cornell University Press.

Hengl, Tomislav, Gerard B. M. Heuvelink, and David G. Rossiter. 2007. “About Regression-Kriging: From Equations to Case Studies.” *Computers and Geosciences* 33(10):1301-1315. doi:10.1016/j.cageo.2007.05.001.

Holt, J. B., and H. Lu. 2011. “Dasymetric Mapping for Population and Sociodemographic Data Redistribution.” In *Urban Remote Sensing: Monitoring, Synthesis and Modeling in the Urban Environment*. New York, NY: John Wiley & Sons.

Ignacio, J. A. F., G. T. Cruz, F. Nardi, and S. Henry. 2015. “Assessing the Effectiveness of a Social Vulnerability Index in Predicting Heterogeneity in the Impacts of Natural Hazards: Case Study of the Tropical Storm Washi Flood in the Philippines.” *Vienna Yearbook of Population Research* 13:91-129.

Kelley, Colin Patrick, Shahrzad Zarrin Mohtadi, Mark A. Cane, Richard Seager, and Yochanan Kushnir. 2015. “Climate Change in the Fertile Crescent and Implications of the Recent Syrian Drought.” *Proceedings of the National Academy of Sciences* 112(11):3241-3246. doi:10.7916/D8MG7NDB.

Kjenstad, K. 2006. "On the Integration of Object-Based Models and Field-Based Models in GIS." *International Journal of Geographic Information Science* 20(5):491-509.

Krivoruchko, K., A. Gribov, and E. Krause. 2011. "Multivariate Areal Interpolation for Continuous and Count Data." *Spatial Statistics* 3:14-19.

Kyriakidis, Phaedon C. 2004. "A Geostatistical Framework for Area-to-Point Spatial Interpolation." *Geographical Analysis* 36(3):259-289. doi:10.1111/j.1538-4632.2004.tb01135.x.

Leibbrandt, Murray, Ingrid Woolard, Arden Finn, and Jonathan Argent. 2010. *Trends in South African Income Distribution and Poverty Since the Fall of Apartheid*. Paris, France: Organization for Economic Co-operation and Development Publishing.

Maantay, J. A., A. R. Maroko, and C. Herrmann. 2007. "Mapping Population Distribution in the Urban Environment: The Cadastral-Based Expert Dasymetric System (CEDS)." *Cartography and Geographic Information Science* 34(2):77-102.

Maantay, J. A., A. R. Maroko, and H. Porter-Morgan. 2008. "Research Note—a New Method for Mapping Population and Understanding the Spatial Dynamics of Disease in Urban Areas: Asthma in the Bronx, New York." *Urban Geography* 29(7):724-738.

MacDonald, Dettwiller, and Associates, Ltd. 2013. *Land Cover Products: BaseVue 2013, 30 meter*. MDA Information Systems LLC. http://www.mdaus.com/products/land-cover-products.

Matheron, G. 1963. "Principles of Geostatistics." *Economic Geology* 58:1246-1266.

McCaa, R., S. Ruggles, M. Davern, T. Swenson, and K. M. Palipudi. 2006. "IPUMS International High Precision Population Census Microdata Samples: Balancing the Privacy-Quality Tradeoff by Means of Restricted Access Extract." *Proceedings of the 2006 CENEX-SDC Project International Conference on Privacy in Statistical Databases*. New York, NY: Springer-Verlag.

MDB ([South African] Municipal Demarcation Board). 2019. *Municipal Demarcation Board*. Web site. http://www.demarcation.org.za/site/.

Mélice, Jean-Luc, and Chris J. C. Reason. 2007. "Return Period of Extreme Rainfall at George, South Africa." *South African Journal of Science* 103(11-12):499-501.

Minnesota Population Center. "Integrated Public use Microdata Series, International: Version 7.1 [Dataset]." *IPUMS International*. Web site. https://doi.org/10.18128/D020.V7.1.

Moran, Ashley McIlvain, Yacob Mulugetta, and Clionadh Raleigh. 2014. *Climate Change and Security in Africa*. The Hague: Institute for Environmental Security.

OpenStreetMap contributors. 2018. *Planet Dump Retrieved from https://Planet.Osm.Org*. https://www.openstreetmap.org.

Peacock, Walter G., Samuel D. Brody, William A. Seitz, William J. Merrell, Arnold Vedlitz, Sammy Zahran, Robert C. Harriss, and Robert R. Stickney. 2010. *Advancing the Resilience of Coastal Localities: Developing, Implementing and Sustaining the use of Coastal Resilience Indicators: A Final Report*. College Station, TX: Hazard Reduction and Recovery Center.

Pebesma, Edzer J. 2006. "The Role of External Variables and GIS Databases in Geostatistical Analysis." *Transactions in GIS* 10(4):615-632. doi:10.1111/j.1467-9671.2006.01015.x.

Peduzzi, Pascal. *Global Estimated Risk Index for Flood Hazard*. http://preview.grid.unep.ch.

Rasmussen, C. E., and C. K. I. Williams. 2006. *Gaussian Processes for Machine Learning*. Cambridge MA: MIT Press.

Schmid, C. F., and E. H. Maccannell. 1955. "Basic Problems, Techniques, and Theory of Isopleth Mapping." *Journal of the American Statistical Association* 27(1):220-239.

Shea, Neil. 2019. "Scenes from the New Cold War Unfolding at the Top of the World." *National Geographic*. https://www.nationalgeographic.com/environment/2018/10/new-cold-war-brews-as-arctic-ice-melts/#close.

Shiferaw, Bekele, Kindie Tesfaye, Menale Kassie, Tsedeke Abate, B. M. Prasanna, and Abebe Menkir. 2014. "Managing Vulnerability to Drought and Enhancing Livelihood Resilience in Sub-Saharan Africa: Technological, Institutional and Policy Options." *Weather and Climate Extremes* 3 (C):67-79. doi:10.1016/j.wace.2014.04.004.

Tempelhoff, Johan, Dewald Van Niekerk, Elize Van Eeden, Ina Gouws, Karin Botha, and Rabson Wurige. 2009. "The December 2004 - January 2005 Floods in the Garden Route Region of the Southern Cape, South Africa." *Jamba: Journal of Disaster Risk Studies* 2(2):93-112. doi:10.4102/jamba.v2i2.18.

Tobler, W. R. 1979. "Smooth Pycnophylactic Interpolation for Geographical Regions." *Journal of the American Statistical Association* 74(367):519-530.

UT-Battelle, LLC. 2017. *LandScan™: High Resolution Global Population Dataset*. Web page. Oak Ridge, TN: Oak Ridge National Laboratory. https://landscan.ornl.gov/.

Wackernagel, H. 2003. *Multivariate Geostatistics: An Introduction with Applications*. Berlin, Germany: Springer.

Waddell, Joy. 2016. "A Nodal Governance Approach to Understanding the Barriers and Opportunities for Disaster Governance: A Case Study on Flood Governance in an Informal Settlement in Cape Town, South Africa." PhD dissertation. Cape Town, South Africa: University of Cape Town.

Weiss, D. J., A. Nelson, H. S. Gibson, W. Temperley, S. Peedell, A. Lieber, M. Hancher, E. Poyart, S. Belchior, N. Fullman, B. Mappin, U. Dalrymple, J. Rozier, T. C. D. Lucas, R. E. Howes, L. S. Tusting, S. Y. Kang, E. Cameron, D. Bisanzio, K. E. Battle, S. Bhatt, and P. W. Gething. 2018. "A Global Map of Travel Time to Cities to Assess Inequalities in Accessibility in 2015." *Nature: International Journal of Science* 553:333-336. https://www.nature.com/articles/nature25181.pdf.

Werrell, Caitlin E., and Francesco Femia. 2016. "Climate Change, the Erosion of State Sovereignty, and World Order." *Brown Journal of World Affairs* 22(2):221-235.

Wilkinson, Kate. 2017. "Black Ownership on SA's Stock Exchange – What We Know." *Africa Check*. https://africacheck.org/factsheets/guide-much-sas-stock-exchange-black-owned-know/.

Williams, C. K. I. 1997. *Prediction with Gaussian Processes: From Linear Regression to Linear Prediction and Beyond* Technical Report No. NCRG/97/012. Birmingham, England: Aston University, Department of Computer Science, Neural Computing Research Group.

Acronyms and Abbreviations

Term	Definition
BLR	Bayesian Linear Regression
BLUE	Best Linear Unbiased Estimate
BLUP	Best Linear Unbiased Prediction
BRIC	Baseline Resilience Index for Communities
CDRI	Community Disaster Resilience Index
COSP	Change of Support Problem
EM-DAT	Emergency Events Database
GP	Gaussian Process
GzLM	Generalized Linear Model
HSVI	Household Social Vulnerability Index
IPUMS	Integrated Public Use Microdata Series
KED	Kriging with External Drift
MAUP	Modifiable Areal Unit Problem
MDB	(South African) Municipal Demarcation Board
OK	Ordinary Kriging
OLS	Ordinary Least Squares
PCA	Principal Components Analysis
RK	Regression Kriging
RWDM	Regression Weighted Dasymetric Mapping
SK	Simple Kriging
SVI	Social Vulnerability Indices
UK	Universal Kriging
UNEP/GRID	United Nations Environment Programme/Global Resource Information Database
VIIRS	Visible Infrared Imaging Radiometer Suite

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
09/01/2019	Final Technical Report (TR)	

4. TITLE AND SUBTITLE
Using the Social Vulnerability Index to Forecast Disaster Migration

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT
622784

6. AUTHOR(S)
Chandler M. Armstrong and Lance L. Larkin

5d. PROJECT NUMBER
855

5e. TASK NUMBER
08

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Engineer Research and Development Center (ERDC)
Construction Engineering Research Laboratory (CERL)
PO Box 9005,
Champaign, IL 61826-9005

8. PERFORMING ORGANIZATION REPORT NUMBER
ERDC/CERL TR-19-20

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Headquarters, U.S. Army Corps of Engineers (HQUSACE)
441 G St., NW
Washington, DC 20314

10. SPONSOR/MONITOR'S ACRONYM(S)
CEERD-CNC

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

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

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report analyzes disaster-driven internal migration, its effects on residence layout by social vulnerability, and demonstrates a modeling procedure for this task. Since the processes underlying disaster-driven migration may unfold at scales below those in the data, especially in vulnerable areas where these models are especially useful, this report focuses on South Africa, which had key flooding events at urban centers over 15 years before the 2011 census. Since only low spatial resolution census data are obtainable for research, we use universal kriging (UK) with a pre-fit model from the Philippines (which used more detailed data) to estimate social vulnerability in South Africa at a higher spatial resolution. With the UK model, we estimate and reaggregate a social vulnerability index (SVI) from South African provincial to municipal boundaries, then fit a model testing the relationship between internal migration, SVI, and flooding risk within each municipality. Results show that, when controlling for flood risk, within-province migration is inversely related to SVI, while neither SVI nor flooding affect migration between provinces. This is consistent with our hypothesis that the socially vulnerable are less likely to leave flood-prone areas, and over time, a positive spatial correlation emerges between SVI and flood risk.

15. SUBJECT TERMS
Disasters, Migration - Internal, Forced migration, South Africa, Risk assessment

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (include area code)
Unclassified	Unclassified	Unclassified	SAR	46	