Technical Appendix: Protecting Open Space & Ourselves: Reducing Flood Risk in the Gulf of Mexico Through Strategic Land Conservation

To identify the most effective watersheds for land conservation, we took the following approach:

- **Identified watersheds with a high probability of future flood damages**, using a random forests classification model, and a suite of physical-based predictor variables.
- **Identified high conservation opportunity watersheds** using the Partnership for Gulf Coast Land Conservation vision footprint and the Protected Areas Database.
- **Mapped priority multi-objective watersheds with high likelihood of future flood damage and high conservation opportunity**

About this study

The study area consisted of 12th order watersheds (based on the USGS Hydrological Unit Code (HUC) encompassing 144 counties and parishes adjacent to the Gulf of Mexico. This area stretches from the Florida Keys to the Southern tip of Texas, which includes watersheds from the following six states: Florida, Georgia, Alabama, Mississippi, Louisiana, and Texas. Watersheds were selected for analysis if they intersect one of the NOAA-designated coastal counties/parishes and if NFIP claim data was available. In total, we analyzed 2,651 adjacent watersheds ringing the Gulf coast (Figure 1).

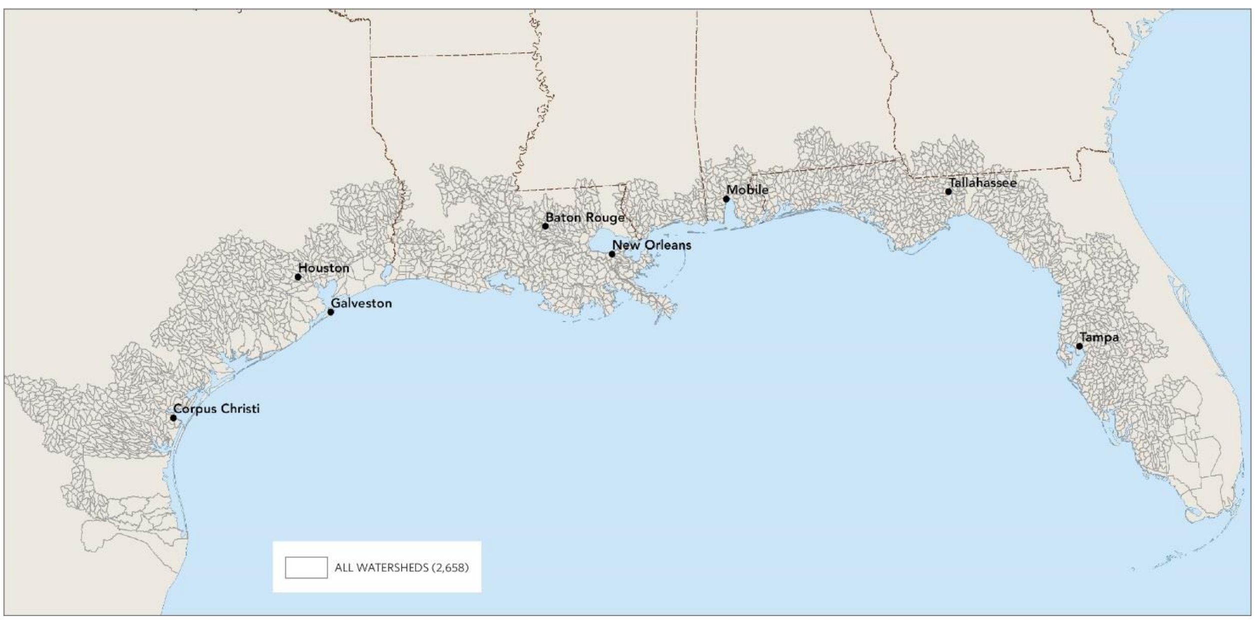

Figure 1. Study area and watershed boundaries.

The coastal margin associated with the Gulf of Mexico is an ideal region to examine the relationship between undeveloped open space and flooding for several reasons. First, this low-lying coastal margin contains large amounts of floodplain (on average, approximately 32 percent of the study area lies within the 100-year floodplain), making it extremely vulnerable to the adverse effects of rainfall and surge-based flooding events. For example, from 1996 to 2007, communities within the study area experienced the largest amount of insured property damage in the U.S. (Brody et al., 2011b). Second, the Gulf coast has a legacy of rapid population growth and associated land use change, creating a diversity of land use and land cover patterns across watersheds. Third, there remain large tracts of undeveloped lands along the Gulf coast that if protected, will help prevent future flood losses to local communities.

To predict the probability of future flood claims and identify watersheds in which open space protection is likely to reduce future flood risk, we engaged in exploratory modelling using a host of techniques to explore the relationships between the amount and configuration of open space and flood damage. Flood damage was measured as the dollar amount (contents and building damage) of claims paid per household under the National Flood Insurance Program (NFIP) from 2008-2014. While insured loss does not cover all damage incurred from a flood event, it remains a strong and viable indicator of flood damage. The dollar value of individual insured claims was aggregated at the 12th order HUC level and log-transformed to better approximate a normal distribution. We recorded and analyzed approximately $1 billion dollars in insured flood losses across the study area.

Due to the relatively short time period of the study data, it is difficult to know if the 1360 watersheds which had insurance policies but no reported damage are 'true zeros' (ie had a storm event large enough capable of causing flooding, but did not cause damages), or 'false zeros' (ie did not have a storm event which could have caused flooding). To address this issue, we queried the NOAA storm events database for all flood, flash flood, and heavy rain events from 2008-2014, and intersected the paths of these 1411 events with the watersheds dataset. This resulted in a dataset of 1201 watersheds which had either reported storm events or reported flood damage.

After identifying watersheds with reported storm events, there were still 182 watersheds which had from 100 – 11,171 insurance policies, yet reported no flooding damage from 2008-2014. A majority of these watersheds were located in Florida, and only experienced heavy rain events, not flooding or flash flooding. A reasonable hypothesis is that watersheds with a large number of insurance policies actually have a significant risk of flooding, but just didn't experience flooding conditions during the study period. Although these areas show up in the data as having zero damages, they could be considered 'false zeros'. Therefore, we also modeled scenarios which removed these watersheds, to see how it affected the predictive ability of the model.

Several contextual control variables were also measured and analyzed in the statistical models. Under socioeconomic and built environment characteristics, we included the median household income based on U.S. Census data for 2010. Census block groups were aggregated to the watershed unit. These variables were included in the models to account for the number of structures per watershed and their estimated value (structure and contents) based on household income. Three development-based development variables were measured based on the percent of impervious surface cover using NOAA CCAP data: high-intensity, low-intensity, and developed open space.

We also included several environmental baseline controls. Precipitation, the biggest predictor of flooding and flood damage, was measured based on the number of days with greater than 1 inch of rainfall. Annual summaries from National Climatic Data Center precipitation stations were interpolated using a kriging estimator for each year in the study period and summarized by the watershed unit. As a measure of risk exposure, we calculated the percent of each HUC in the 100-year floodplain using FEMA Q3 data. On average, 32 percent of watersheds in the sample lie within the 100-year floodplain. Topographic features can also be important when predicting flood impacts as low-lying areas along the coast tend to pool water and result in inundation. The average slope of each watershed was calculated using 30 meter digital elevation models (DEM) from the National Elevation Dataset.

Hydraulic conductivity of the soil, which measures the speed and ability with which the soil can infiltrate water was also included as an environmental control variable in the statistical models. If surface runoff cannot penetrate into the soil, flooding can be exacerbated, especially in developed areas. The Gulf coast study area contains a range of soil characteristics, from clay/clay-loam to sand. The hydraulic conductivity (KSAT) in each watershed was determined by matching soil data from NRCS (http://soildatamart.nrcs.usda.gov/) to the accepted Green and Ampt parameters for soil texture classes (Rawls et al., 1982; Rawls et al., 1983). The KSAT variable was derived from the STATSGO2 dataset (Soil Survey Staff, Natural Resources Conservation Service, USDA, Web Soil Survey, http://websoilsurvey.nrcs.usda.gov/, accessed 05/11/2013). KSAT was calculated using a weighted average across component soils for the surface layer. The hydraulic conductivity values in the watershed ranged from 0-108.73 micrometers per second.

Finally, we included the size of each watershed in square meters. Larger watersheds should contain more drainage area and greater stream length that could affect flood impacts, particularly in down-stream areas. We also included a variable measuring the number of NFIP policies as an additional measure of risk perception and exposure.

Identifying watersheds with high probability of future flood damages

The watersheds dataset consisted of 2658 HUC12 watersheds, with 1873 having at least 1 NFIP policy. Of these 1873 watersheds, 513 reported damages from 2008-2014, and 1360 reported no damage.
There were a number of difficulties modeling the connection between flooding and predictor variables associated with the watersheds dataset, including:

1. Strongly unbalanced dataset -- high number of watersheds with $0 damage
2. Distinguishing between watersheds with ‘true’ $0 damage and those which simply didn’t have a significant enough storm event over the 8 years of study data to cause flooding.
3. Outlier watersheds with very high damage
4. Heavily skewed distributions of both predictor and response variables, violating assumptions of traditional regression techniques
5. Spatial autocorrelation between watersheds
6. Spatial non-stationarity of response – the relationships between the variables and response changes across the Gulf.
7. High correlation between response variables

Due to the relatively short time period of the study data, it is difficult to know if the 1360 watersheds which had insurance policies but no reported damage are ‘true zeros’ (ie had a storm event large enough capable of causing flooding, but did not cause damages), or ‘false zeros’ (ie did not have a storm event which could have caused flooding). To estimate this, we queried the NOAA storm events database for all flood, flash flood, and heavy rain events from 2008-2014, and intersected the paths of these 1411 events

with the watersheds dataset. This resulted in a dataset of 1201 watersheds which had either reported storm events or reported flood damage.
After identifying watersheds with reported storm events, there were still 182 watersheds which had from 100 – 11,171 insurance policies, yet reported no flooding damage from 2008-2014. A majority of these watersheds were located in Florida, and only experienced heavy rain events, not flooding or flash flooding. A reasonable hypothesis is that watersheds with a large number of insurance policies actually have a significant risk of flooding, but just didn't experience flooding conditions during the study period. Although these areas show up in the data as having zero damages, they could be considered 'false zeros'. Therefore, we also modeled scenarios which removed these watersheds, to see how it affected the predictive ability of the model.

Given the many ways skewed data can affect the inference we make about the relationship between variables, we took an exploratory approach towards statistical modeling, employing 4 different techniques to assess the relationship between predictor and response variables: 1) multilinear regression, using AIC model selection, 2) geographically-weighted regression, 3) classification and regression trees, and 4) random forests. For final modeling scenarios, we used random forests to model both probability of flooding and amount of damages.

Random forests modeling
Random forests (RF) is a non-parametric statistical classification technique that combines classification and regression tree (CART) techniques with a bootstrap method which prevents overfitting problems commonly associated with CARTs. RF is not as susceptible to problems with underlying data distributions as traditional parametric techniques such as linear regression, and can also better handle non-linear responses, making it a better approach for the data problems associated with the dataset. RF can perform both classification (ie probability of flooding occurrence) and regression (extent of damages), in a fashion analogous to logistic and linear regression, respectively.

We performed all RF modeling in R using the random Forest library, growing 2000 trees per model, and using a minimum node size of 10 for classification RF models, and 20 for regression models. Node size determines the minimum number of watersheds which must be clustered together to form a node within a random forest. Increasing the minimum node size from the default values of 1 for classification and 5 for regression helps prevent RF from overfitting the data.

To evaluate the overall predictive performance of models, we used the Area Under the Curve (AUC) of receiver operating characteristic (ROC) curves. The area under the ROC curve (AUC) provides a measure of discrimination ability, varying from 0.5 for discrimination ability no better than random, to 1.0, indicating perfect discrimination between presence and absence of flooding damage. AUC is preferred over an overall percentage of correctly classified sites, both because correct classification rate is dependent on the prevalence of an event (proportion of watersheds which had reported flooding damage) and because correct classification rate is threshold dependent (Fielding & Bell 1997).

Exploratory modeling using a variety of parametric and non-parametric techniques revealed mostly weak correlations between open space patch metrics (size and configuration of open space) and flood damage, and strong correlation with physical variables, so we built our final models using random forests. Random forests (RF) is a non-parametric statistical classification technique that combines classification and regression tree (CART) techniques with a bootstrap method which prevents overfitting problems commonly associated with CARTs.

The final RF model best predicting probability of flooding damage included:

1) soil conductivity,
2) precipitation,
3) percent of watershed in floodplains,
4) mean slope,
5) mean distance to coast,
6) spatial coordinates, and
7) developed land cover.

In all models, including geographic coordinates as predictor variables improved model accuracy. Longitude was the second most important predictor after human footprint (either developed area or insurance policies), and correlated with a geographic split in reported damage between the western and eastern halves of the Gulf of Mexico. This geographic split seems to coincide with both precipitation and with soil conductivity. However, because longitude is a stronger predictor than both soil conductivity and precipitation, it could also be correlated with uncaptured geographic determinants of flooding, such as broader weather patterns.

Classification random forests models performed moderately well at predicting probability of flooding damage, with AUC ranging from 0.84 to 0.88 for models using all watersheds with a reported storm event or damage, and 0.90 to 0.94 for models using a subset of this data that omitted 'false zeros' (watersheds which did not experience damage but had more than 100 insurance policies). This reveals that using the dataset with false zeros removed resulted in substantially better predictive ability, and subsequent results are reported using this approach. The final model had an overall accuracy of 85% and we used this model to identify watersheds with a high probability of flood damage. 730 watersheds (nearly one third of the watersheds) were classified as high flooding potential, meaning the model predicts a 70 percent or higher chance of flooding in the future (Figure 2).

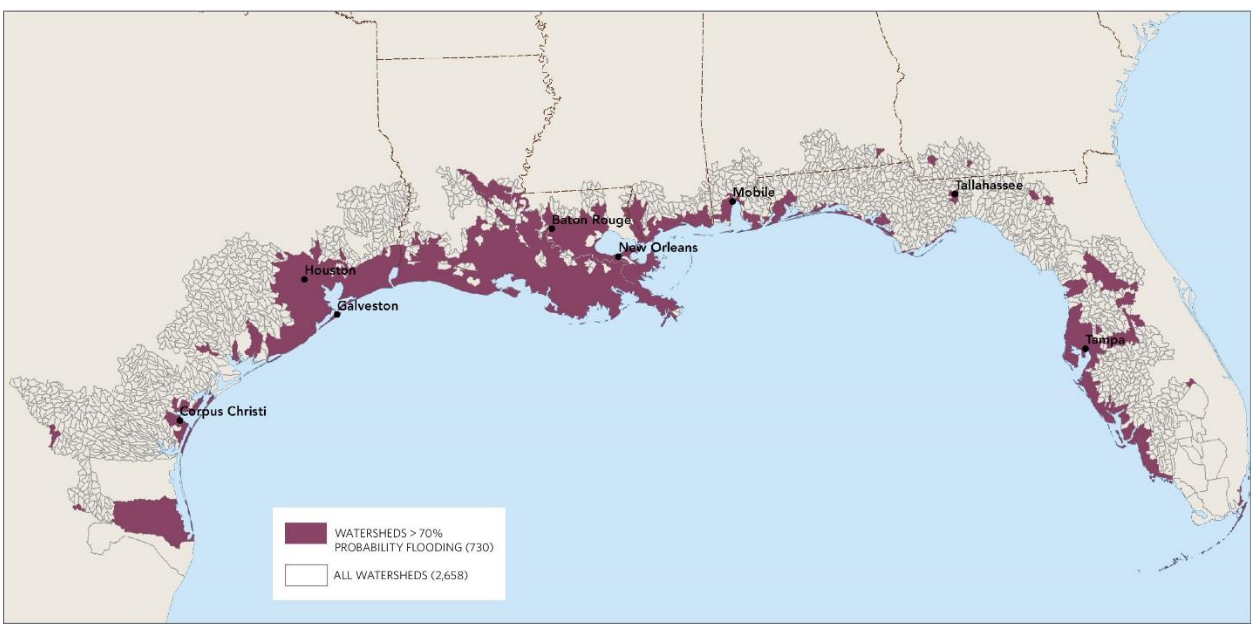

Figure 2. 730 watersheds with modeled high probability of flood damages.

Identifying unprotected, high conservation opportunity watersheds

In October, 2014, the Partnership for Gulf Coast Land Conservation, in conjunction with The Nature Conservancy and The Conservation Fund, released 'A Land Conservation Vision for the Gulf of Mexico Region' which is a series of maps that identify high-value geographic areas for land conservation. These maps were created by the partner organizations and include:

1. Focus areas identified by the partners that reflect local community values;
2. Wetlands
3. Migratory bird habitat
4. Scenic rivers
5. Longleaf pine habitat

We used the Conservation 'vision footprint' maps to indicate conservation lands that are a high priority for land trusts across the Gulf Coast region. Because we are focused on open space protection, we screened out developed lands within the 'vision footprint' maps using the 2011 National Land Cover Dataset. To identify watersheds with opportunities for conserving land within the "vision footprint', we used the Protected Areas Database to identify open spaces that are not currently protected. 1168 watersheds were classified as high conservation opportunity, meaning 20% or more of the watershed area is undeveloped, unprotected and within the PGCLC conservation vision footprint (Figure 3).

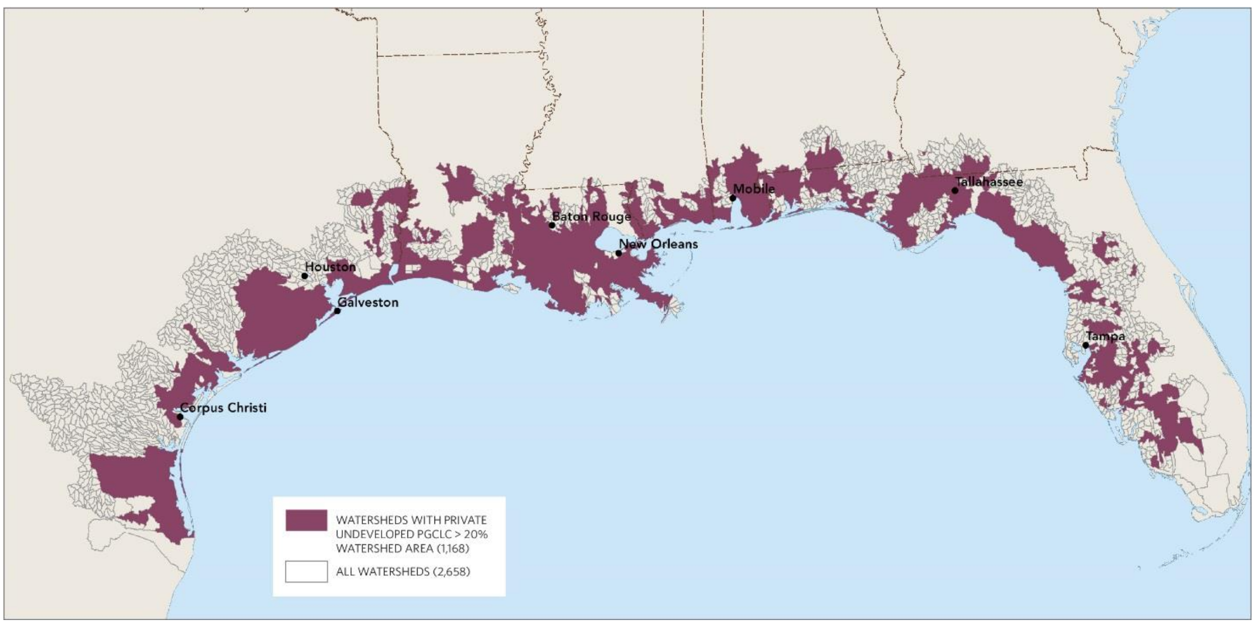

Figure 3. 1168 watersheds classified as high conservation opportunity.

Mapping High Priority Multi-Objective Watersheds
To map the multi-objective watersheds, we clipped the high flood damage watersheds using the high conservation opportunity watersheds to select watersheds that are **both** high risk for flood damage and high conservation value. Our analysis identified **421 watersheds** along the Gulf of Mexico that have both high likelihood of flood damages and high conservation opportunity (Figure 4).

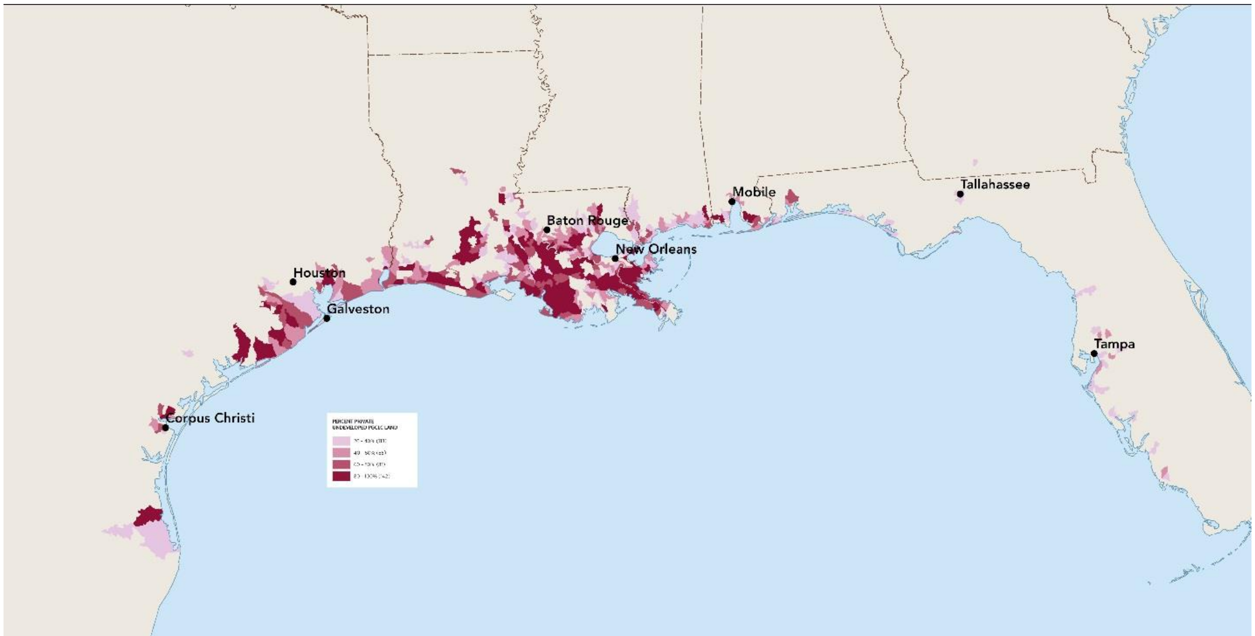

Figure 4. Watersheds along the Gulf of Mexico that have both high likelihood of flood damages and high conservation opportunity.

How to use these maps
Open space protection has long been a cornerstone of land use planning and policy across the U.S. This designation is commonly embedded within local land use or comprehensive plans and often implemented through voluntary land acquisition in fee and easement, and, to a lesser extent, through subdivision, zoning, and building ordinances including transfers of development rights, mitigation, and buffers and setbacks. Open space is employed as a land use strategy for multiple purposes, including establishing public parks and recreation areas, separating conflicting land uses, protecting naturally occurring wetlands and riparian corridors, and providing water retention/detention

The 421 watersheds identified in our analysis are appropriate locations for further consideration of open space protection as a flood risk reduction and conservation strategy. Municipalities, land trusts and government agencies engaged in land acquisition within these watersheds are best suited to identify specific land protection opportunities within each watershed but should consider opportunities to fund these acquisitions to meet multiple objectives.

FEMA's voluntary Community Rating System (CRS) provides an opportunity to use open space protection to meet multiple objectives by more tightly linking land conservation and flood risk reduction. The CRS is a voluntary program in which communities get points for flood mitigation activities and can earn discounts on their residents' flood insurance premiums. Open space protection has long been a creditable CRS activity but is underutilized despite its likely effectiveness in reducing losses. CRS communities that protect open space (even for reasons other than flood risk reduction) can earn their citizen's an insurance discount.

343 municipalities, counties, and parishes are within or partially within the high priority watersheds. As of 2014, only 100 of these communities were enrolled in the CRS program. Of the 100 CRS communities, the average number of points each community earned for open space was only 131 out of a possible 2020 points for open space preservation. This indicates that there is substantial opportunity to improve CRS scores within key target watershed communities by protecting open space which can reduce flood risk, conserve species and habitats and provide an economic benefit in the form of reduced insurance premiums.

*To further explore the 421 priority watersheds and the findings in this report, please visit www.maps.coastalresilience.org/gulfmex

Contact:
Dr. Christine C Shepard
Director of Science, Gulf of Mexico Program
The Nature Conservancy
cshepard@tnc.org

Original report:
Shepard C, Majka D, Brody S, Highfield W, and J Fargione. Protecting Open Space & Ourselves: Reducing Flood Risk in the Gulf of Mexico Through Strategic Land Conservation. Washington DC: The Nature Conservancy. December 2016. 12 pages.