Development *of the* Gulf of Mexico Marine Debris Model

Zachary Nixon and Nir Barnea

US Department of Commerce
National Oceanic and Atmospheric Administration
National Ocean Service
Office of Response and Restoration
Marine Debris Division
Silver Spring, MD 20910

National Oceanic and Atmospheric Administration
Technical Memorandum NOS-OR&R-35
January 2010

DEVELOPMENT OF THE GULF OF MEXICO MARINE DEBRIS MODEL

Zachary Nixon[1] and Nir Barnea[2]

[1] Research Planning, Inc.
Columbia, SC 29201, USA

[2] National Oceanic and Atmospheric Administration
Office of Response & Restoration
NOAA Marine Debris Division
Silver Spring, MD 20910, USA

National Oceanic and Atmospheric Administration
Technical Memorandum NOS-OR&R-35
January 2010

This document should be cited as follows:
Nixon, Z., and N. Barnea. 2010. Development of the Gulf of Mexico Marine Debris Model. January 2010. NOAA Technical Memorandum NOS-OR&R-35.

For copies of this document, please contact:
NOAA Marine Debris Division
N/ORR, SSMC-4
1305 East-West Highway
Silver Spring, MD 20910

www.MarineDebris.noaa.gov

ACKNOWLEDGMENTS

The development of this model benefited greatly from the support and guidance of a number of people who worked on survey, assessment, and removal of marine debris after hurricanes Katrina and Rita:

- Steve Heath, Mark Clark, and Nyle Trimble: Alabama Marine Resources Division, Dauphin Island, AL
- Irvin Jackson Grant Larsen, and Kim Denz: Mississippi Department of Marine
- Resources, Biloxi, MS
- Marty Bourgeois, Harry Blanchet, and Jim Hanifen,: Louisiana
- Department of Wildlife and Fisheries, Baton Rouge, LA
- Garrett Broussard, and Sam Martin: Louisiana Department of Natural Resources, Lafayette and Baton Rouge, LA
- Robert Travis: U.S. Coast Guard, New Orleans, LA
- Mark Nickelson: FEMA Inshore Lead, Mandeville, LA

Christopher Barker, Holly Bamford, Debra Simecek-Beatty, Brendan Bray, George Graettinger, Neal Parry, and Glen Watabayashi of NOAA's Office of Response and Restoration, and Michael Carron, of the Northern Gulf Institute at Stennis Space Center reviewed earlier drafts, and their comments greatly improved the document.

Carey Morishige of NOAA's Marine Debris Program edited the final version of the document and turned it into a Technical Memorandum.

We welcome any comments or suggestions you may have to improve the current version of the document. Please send your comments to:

Marine Debris ERP
NOAA OR&R
7600 Sand Point Way N.E.
Seattle, WA 98115
Fax: (206)526-6329
Email: mderp@noaa.gov

Table of Contents

Identified Problem and Opportunity

During the 2005 hurricane season, Hurricane Katrina inflicted severe damage on the Gulf of Mexico (GOM) coastal region and deposited huge amounts of debris over large areas of the Gulf coast. Submerged marine debris poses a hazard to vessel traffic, can cause snags and hang-ups during fishing and can adversely affect viable fishing grounds. To address the submerged debris problem, Congress appropriated funds to survey areas potentially affected by submerged marine debris, tasking the National Oceanic and Atmospheric (NOAA) Office of Coast Survey and Office of Response and Restoration to conduct the surveys, compile and disseminate data to the public and stakeholders in an effective and useable format, carry out outreach activities, and otherwise coordinate for removal of marine debris, identifying both present and future needs. This project, lasting over three and a half years, came to be known as the Gulf of Mexico Marine Debris Project (GOMMDP).

One of the most pressing needs when addressing marine debris dispersion caused by a major storm is to assess and verify debris locations. Knowing where high densities of marine debris are likely to be found could greatly assist in developing survey priorities and in planning for debris removal. Moreover, a simple, inexpensive, rapid, and verifiable model to predict marine debris densities could be very useful to Federal and State managers dealing with marine debris. The data collected by this project comprise one of the largest systematically collected, post-storm investigations of marine debris in existence, and provide the building blocks for assembling such a model. This document describes model construction and performance.

Goal

The purpose of the model is to characterize spatial differences in the distribution of relatively large, storm-mobilized and deposited, anthropogenic marine debris objects across the nearshore seafloor so as to help prioritize these areas for survey. The guidance for developing the model was that it should be conceptually and statistically robust and based upon the data collected by the GOMMDP. In addition, the model should be as applicable as possible to a generic storm event, or a storm event to occur in the future.

Previous Efforts

Previous efforts at quantitatively modeling debris generated by hurricanes have largely focused on terrestrial debris. Notable efforts include the HAZUS-MH family of models maintained by the Federal Emergency Management Agency (FEMA, 2008). The commercially developed HurDET model (Umpierre and Margolies, 2005) is another example. These efforts, while robust and based upon actual data derived from multiple storm events, are limited to the spatial support of existing terrestrial boundaries – municipalities, census blocks, or traffic analysis zones. As such, they are incompatible with the goal of estimating spatial differences in the distribution of marine debris.

Model Overview and Limitations

This effort models gross relative differences in the spatial distribution of all types of non-floating anthropogenic marine debris objects across the nearshore seafloor by statistical correlations via logistic regression. The model focus is limited to debris objects generated from coastal inhabited areas, and mobilized and deposited during a hurricane-level storm event. This model does not attempt to mechanistically describe the sources, movement, and deposition of marine debris, nor will it be able to predict actual concentrations or volumes of debris. Instead, the model attempts to relate spatial data describing the relative densities of observed marine debris to spatial data describing potential influencing factors, and to identify potential hotspots based upon actual data.

The data used to construct the model are the locations of marine debris items related to Hurricanes Katrina and Rita identified via side scan sonar in surveys conducted in 2006 by the GOMMDP. These data do not distinguish between natural or anthropogenic debris, or between storm-generated and non-storm-generated debris. The model, however, will be confined to predicting concentrations of post-storm, anthropogenic debris, which is usually of greatest concern. It is assumed that the data collected by this project represent a realistic picture of storm-mobilized and deposited subsurface anthropogenic marine debris objects.

It is difficult to quantify the time and length scales of the model in the sense used for physical deterministic models. The model itself examines correlations only, and has no inherent time scale, however some generalizations are possible. This model examines trends in the essentially static distribution of marine debris objects at timescales of months to years after the originating storm event. As such, it is intended to provide a snapshot of this distribution after a storm, but before another event capable of mobilizing substantial amounts of debris. All predictor data have been discretized to a common grid of 100 meter (m) by 100m square cells. This can be considered the minimum length scale of the model. The model was constructed using data from a raster grid at this resolution that spans approximately 250 kilometers (km). As such, the maximum length scale of the model can be considered to be hundreds of kilometers. The output of the model is an estimated probability of encountering one or more debris items within that grid cell, given the values of the predictor data in that cell.

Predictor Data

The factors influencing the distribution of anthropogenic marine debris following a storm event are largely intuitive. In general, higher storm energy in the form of wind, waves, and storm surge in the vicinity of concentrations of human structures leads to larger amounts of anthropogenic debris. If these debris sources are close to open water, then there is a higher potential for mobilization of debris from sources and deposition in the nearby marine environment. Thus, selecting candidate predictor data sets for marine debris seems equally intuitive. The critical part in building a model that will be useful in the future, however, is selecting and evaluating candidate predictor data sets that one might reasonably anticipate being available in a consistent format soon after a storm event. Here, we evaluate only data that are spatial in nature and that either exist for the entire country or would be produced in a consistent format after a storm event. Table 1 lists the evaluated predictor data sets.

Storm surge, wind speed, and bathymetric data were evaluated as proxies for debris-generating storm energy. Distance to shoreline, distance to waterways, and nearby onshore structure damage, and offshore oil and gas infrastructure density were measured as proxies for relative potential debris sources. Note that all spatial data analysis, including processing of the predictor data, was carried out using the ArcGIS 9.2 software package (ESRI, 2007a).

TABLE 1. Evaluated debris probability predictor raw data sets.

Source	Title	Format
NOAA	H*Wind Surface Wind Analysis data	ESRI shapefile
NOAA	Experimental probabilistic hurricane storm surge data	GRIB2 grid
NOAA	NGDC 3-Arc second coastal relief model data	ASCII grid
ESRI	Shoreline data	ESRI shapefile
USACE	National waterways network data	BTS text file
LOSCO	Navigated waterways of Louisiana	ESRI Shapefile
FEMA	Remote sensing data	ESRI shapefile
MMS	Oil and gas platforms in federal waters	ESRI Shapefile
GSA	Oil and gas platforms in Alabama State Waters	ESRI Shapefile
MARIS	Oil and gas platforms in Mississippi State Waters	ESRI Shapefile
LDNR	Oil and gas platforms in Louisiana State Waters	ESRI Shapefile

NOAA's H* Wind Surface Wind Analysis products (NOAA, 2007b) describe hurricane surface wind fields by summarizing data from a variety of observation platforms. Powell *et al.* (1998) provide more background on these data products. While multiple datasets representing snapshots representing various time intervals are available, a single dataset consisting of a lattice of vector points representing the peak surficial wind speed was obtained. These data were imported to ArcGIS and interpolated to the project grid spacing of 100 m by 100 m using an inverse distance weighting method. Figure 1 depicts the final maximum wind speed grid.

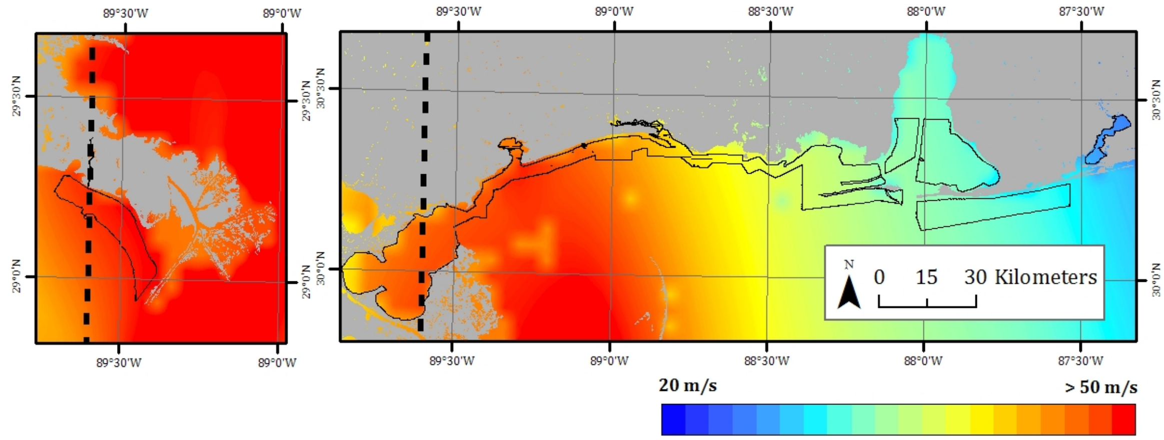


FIGURE 1. 100 m grid of maximum surface wind speeds in meters per second for hurricane Katrina. Dotted line indicates storm eye track.

NOAA's Experimental Probabilistic Hurricane Storm Surge data (NOAA, 2007a) describe the 10 percent chance storm surge exceedance height above normal high tide based upon an ensemble of Sea, Lake, and Overland Surge from Hurricanes (SLOSH) model runs using the last posted National Hurricane Center (NHC) official hurricane advisory. These data were converted from the GRIB2 format to vector points, imported to ArcGIS, and interpolated to the project grid spacing using an inverse distance weighting method. Figure 2 depicts the final storm surge height grid for hurricane Katrina.

We will later note that wind speed as described above and the predicted probabilistic storm surge data are strongly correlated, largely because the SLOSH model ensembles used to generate surge estimates rely mainly upon wind speed as an input. True, independent measures of storm surge (from post-landfall ground surveys; USGS sensors; or remote sensing techniques) or hindcast SLOSH model output from after storm landfall would be better variables to include a part of this model. Currently, however, such products are not generated in consistent or timely manner after all storm events, and so, are not proposed as candidates for inclusion in this effort, which is intended to produce a "nowcast" product.

NOAA NGDC's 3-Arc Second Coastal Relief Model data (Divins and Metzger, 2007) describe integrated bathymetric and topographic information for the coastal zone derived from all of applicable data sources, including topographic maps, hydrographic soundings, and sonar and LIDAR data. These data were converted from the ASCII format to a raster in ArcGIS and resampled to the project grid spacing using a cubic convolution method. Figure 3 depicts the final bathymetric-topographic grid.

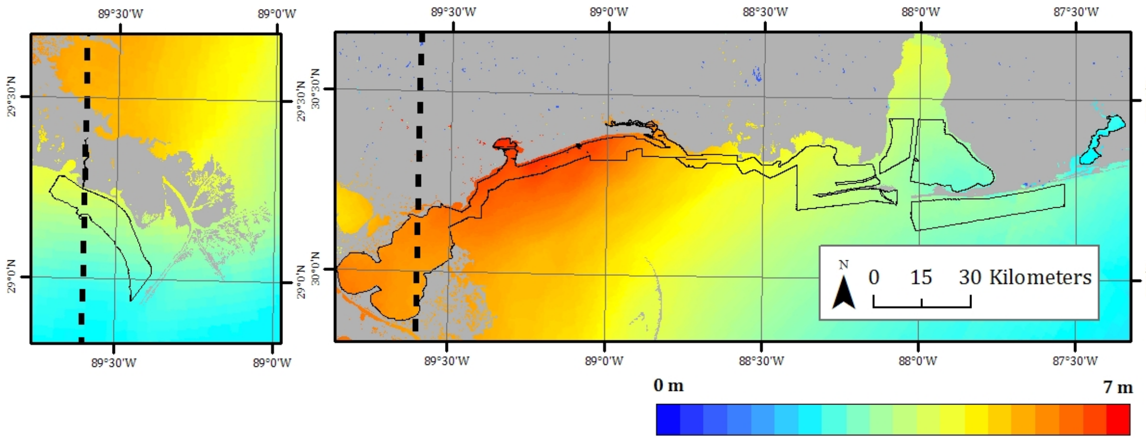


FIGURE 2. 100 m grid of 10 percent chance storm surge exceedance heights in meters for hurricane Katrina.

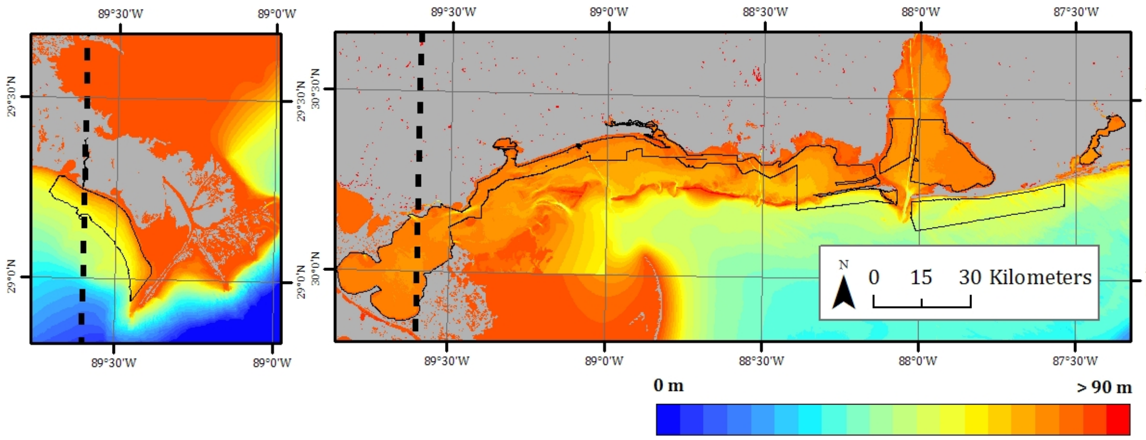


FIGURE 3. 100 m grid of bathymetric/topographic elevation in meters.

The Environmental Systems Research Institute (ESRI) high resolution shoreline (ESRI, 2007b) was converted to a land-water grid using the project grid cell spacing. These same grid dimensions were used for all predictor grids. For each cell, Euclidean distance to land was calculated in meters in ArcGIS. Figure 4 depicts the final distance to land grid.

The United States Army Corps of Engineers (USACE) National Waterways Network Data (USACE, 2007) and Louisiana waterway data (LOSCO, 1999) describe linear waterway routes for the intra-coastal and near-coastal waters of the nation. These data were obtained, imported into ArcGIS and merged to a single dataset of vector lines. For each cell, Euclidean distance to linear waterway was calculated in meters. Figure 5 depicts the final distance to waterway grid.

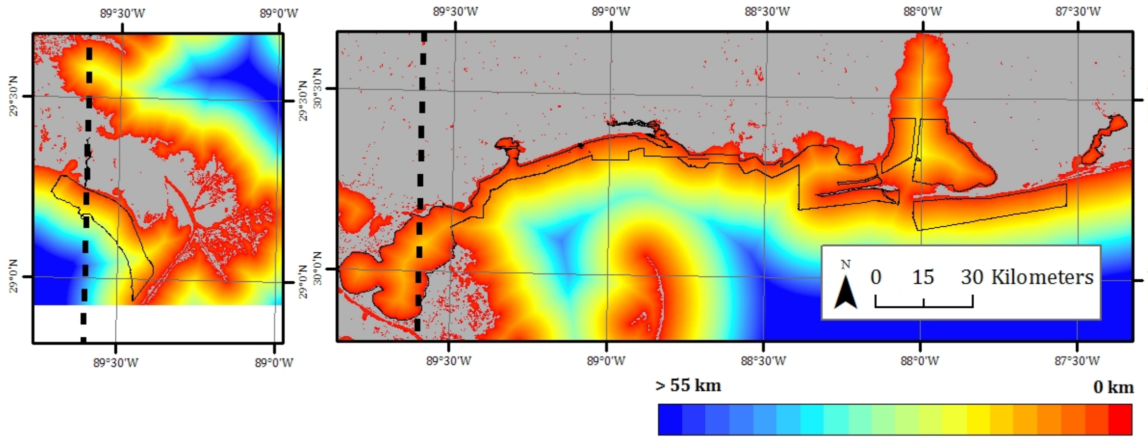


FIGURE 4. 100 m grid of Euclidean distance to shoreline in meters.

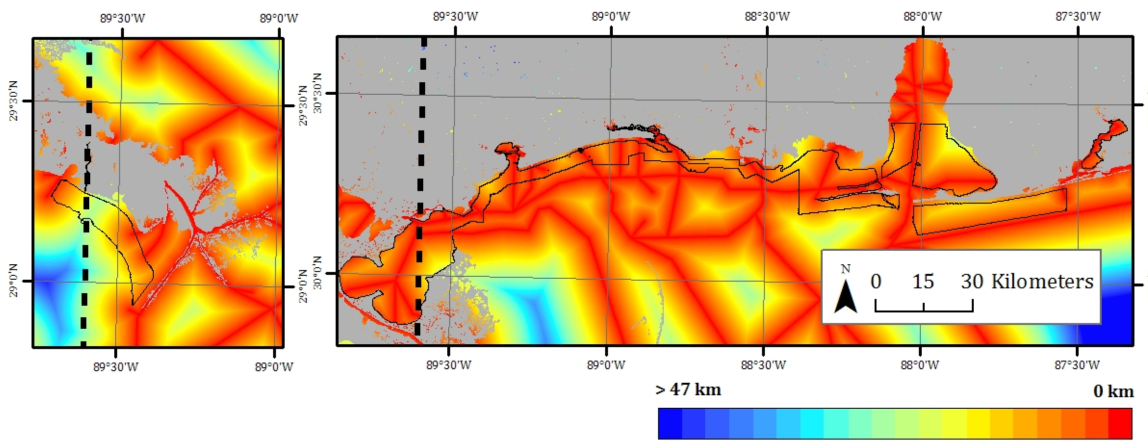


FIGURE 5. 100 m grid of Euclidean distance to waterways in meters.

FEMA damage assessment polygons (FEMA, 2007) describe areas damaged by storm events and are generated by interpretation of remote sensing data after a storm. Polygons are digitized by analysts and assigned one of several damage descriptors. These damage descriptors are qualitative. To assist in using such descriptors in a quantitative model, these descriptors were converted to a set of damage indices ranging from 10 to 80. These indices are an estimate of the percentage of destroyed structures within areas with that descriptor, as per Table 2.

These polygon data were imported to ArcGIS and converted to a damage index grid at the project grid spacing. It is anticipated that marine debris will be more likely to be located proximal to areas with significant storm damage, but it is not known what degree of proximity will be significant. A series of derivative grids were generated to represent local damage indices at three different spatial scales. These derivative grids represent summations of the damage index values at each location within four different distances – 10 km, 5 km, 2.5 km, and 1 km - from that location. These sums serve as proxy variables for the potential amount of debris derived from nearby onshore structures. These three distances were selected as representative of what "nearby" means, in the context of debris mobilization and deposition. Figure 6 depicts the damage index grid and three focal sum grids.

TABLE 2. FEMA hurricane damage descriptors, definitions and assigned damage index values.

Descriptor	Definition	Damage Index
Flood Area	No explicit definition of damage; assumed similar to the Limited Damage category defined below.	10
Limited Damage	Generally superficial damage to solid structures (e.g., loss of tiles or roof shingles); some mobile homes and light structures are damaged or displaced.	10
Moderate Damage	Solid structures sustain exterior damage (e.g., missing roofs or roof segments); some mobile homes and light structures are destroyed, many are damaged or displaced.	25
Extensive Damage	Some solid structures are destroyed; most sustain exterior and interior damage (e.g., roofs missing, interior walls exposed), most mobile homes and light structures are destroyed.	50
Catastrophic Damage	Most solid and all light or mobile structures destroyed.	80

Multiple datasets exist that describe the locations of offshore oil and gas infrastructure in the Gulf of Mexico. Data describing wells and oil and gas platforms in federal waters (MMS, 1998) as well as state waters (GSA, 2008; LDNR, 2007; MARIS, 2004) were obtained and merged into a single dataset of vector points. It is anticipated that marine debris will be more likely to be located proximal to oil and gas infrastructure, but, similar to the onshore damage data, it is not known what degree of proximity will be significant. Again, derivative grids were created at different spatial scales to explore these relationships. These derivative grids represent point densities in counts per square kilometer of oil and gas infrastructure at each location within four different distances – 10 km, 5 km, 2.5 km and 1 km - from that location. These densities serve as proxy variables for the potential amount of debris derived from nearby oil and gas infrastructure. Figure 7 depicts the three density grids.

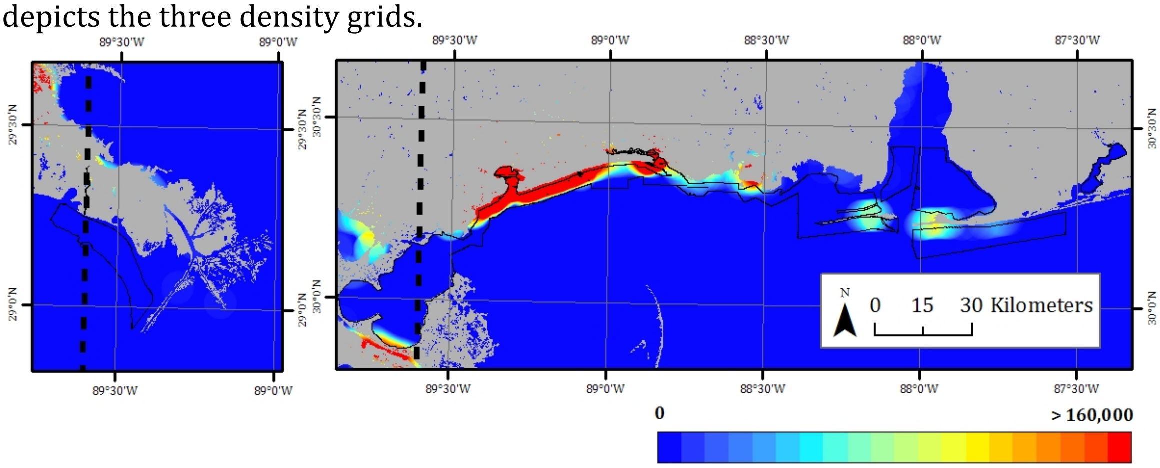


FIGURE 6. 100 m grid of 5 km focal sum of damage index derived from FEMA damage polygons.

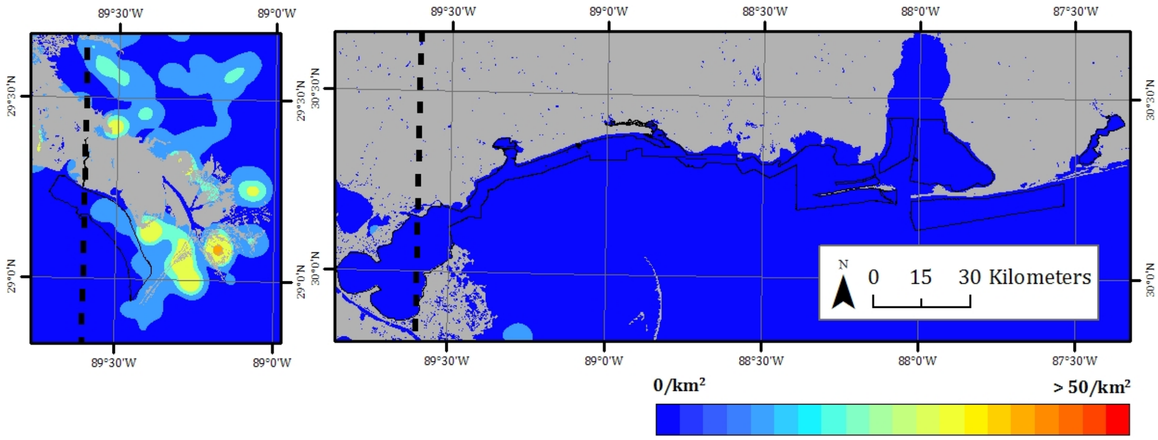


FIGURE 7. 100 m grid of oil and gas infrastructure density evaluated at 5 km kernel size in counts per square km.

Methodology

The goal of the model is to relate some measure of marine debris density or likelihood after a storm event to spatially distributed predictor variables. Maps of these predictor variables can be used in the future to estimate the locations of debris concentration areas for survey prioritization. There are a number of statistical methodologies that might be used to construct such a model. The logistic regression variant of Generalized Linear Modeling (GLM) was selected because it is a relatively simple, robust modeling strategy implementable in a number of software packages that yields a relatively easily understandable output. Venables and Ripley (2003) provide a review of modern logistic regression techniques. The coupling of logistic regression models with spatially distributed predictors derived from data sets stored in Geographical Information System (GIS) is widespread across a number of disciplines, including ecology and hazard modeling.

The model development was based upon the known locations of 5,137 items identified via side scan sonar as potential debris items, as in Figure 8. Of these, roughly 10 percent were located in the eastern portion of the study area and were thought to have been possibly generated by Hurricane Ivan in 2004. These items were removed from the data set, leaving 4,549 debris locations. Logistic regression models the binary probability of a given event – in this case, the probability of an identified item occurring at a given location. The modeling procedure requires a set of locations where items did not occur. These locations are termed pseudo-absences. Accordingly, a set of 4,549 random locations were generated. The model seeks to find differences in the average values of the predictor variables between the actual debris item locations and the average values of the entirety of the surveyed area – as represented by the pseudo-absence locations. Note that, while there are almost identical numbers of actual and random locations, the actual locations appear strongly clustered in certain locations, while the random locations are evenly distributed across the study area.

All statistical modeling was conducted using the open-source statistical computing language R (R Development Core Team, 2007). The predictor variable grids were sampled at each actual and randomly generated location. These data were imported from the ArcGIS environment to the R environment for exploratory analysis and statistical modeling. After construction of the final model in the R environment, the model was implemented as a raster algebra equation in the ArcGIS environment to yield a final prediction grid. The following sections describe the results of the exploratory analysis and the details of model construction and results.

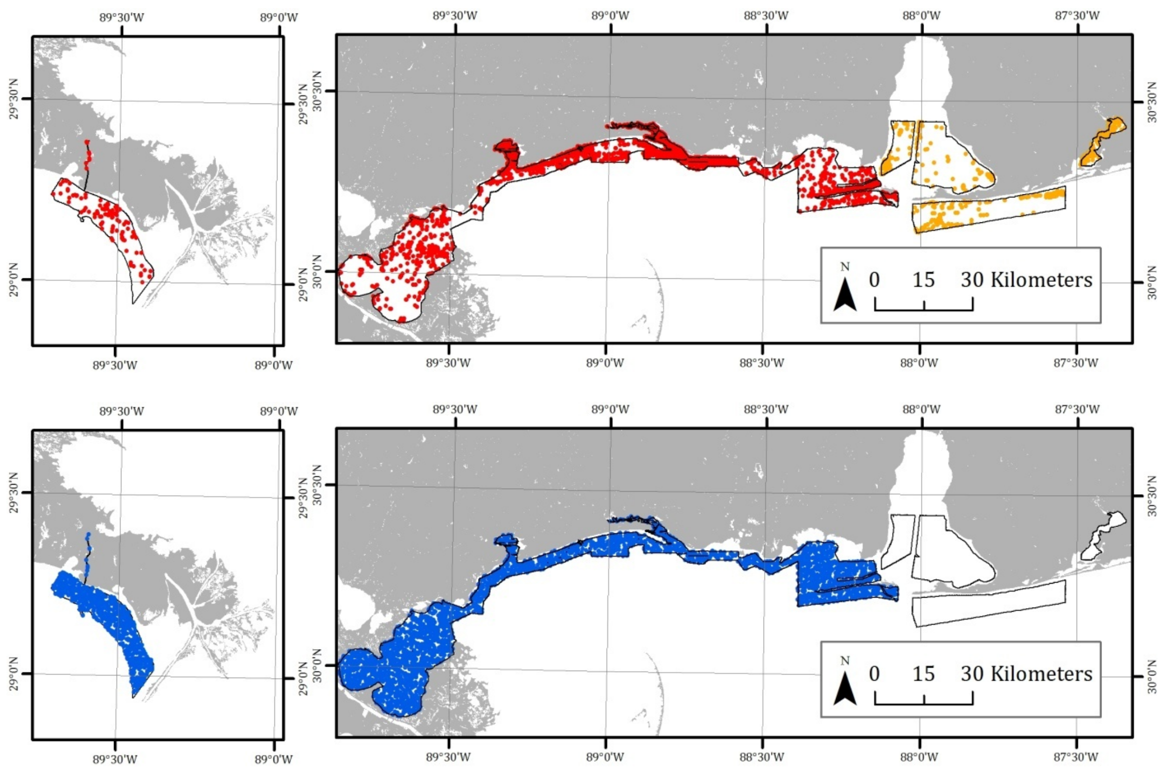


FIGURE 8. Identified items (A) and randomly generated pseudo-locations (B). The project study area boundaries outlined. Debris locations inside the Hurricane Ivan impact area in orange. Note the clustered nature of the actual locations versus the even distribution of random locations.

Exploratory Data Analysis

Model construction began with exploratory analysis of the relationships between the investigated predictor variables. Table 3 contains summary statistics for the predictor variables at all sampled locations – both debris and pseudo-absence. Figure 9 contains pairs plots for all the predictor variables below the diagonal, and absolute correlation coefficients in the upper diagonal. Note the strong correlations between maximum wind speed and maximum storm surge, as well as between the damage indices and infrastructure density metrics.

All potential predictor variables were evaluated for ability to predict the distribution of marine debris in several ways. Figure 10 shows histograms of predictor variable values at all locations by presence/absence of an identified marine debris object. Univariate logistic regression models were constructed for each continuous variable. Parameter significance results via X^2 tests are reported in Table 4. We note that all the variables are significantly related to the presence of debris. While the strength of these relationships is statistically robust, in some cases this is an artifact of the very large sample size. Often, the effects sizes are not very large, implying that there is strong evidence for a relationship, but that relationship itself is weak.

TABLE 3. Summary statistics and codes for predictor variables

Variable	Code	Min	1st Q	Med	Avg	3rd Q	Max.
Max. Wind Speed (m/s)	WSMAX_M	22.5	34.4	42.0	40.8	46.9	50.4
Max. Storm Surge (m)	SURGE_M	1.1	3.0	5.0	4.3	5.4	6.3
Elevation (m)	BATHY_M	-17.6	-3.6	-2.7	-3.8	-2.1	3.0
Dist. to Land (m)	DLAND_M	0	1000	2341	2842	4272	10782
Dist. To Waterway (m)	DWWAY_M	0	781	2138	2842	4327	14227
Damage Index (10 km)	DAM10	0	50	11410	55136	100975	282145
Damage Index (5 km)	DAM25	0	0	100	18514	21715	112210
Damage Index (2.5 km)	DAM50	0	0	0	4385	2651	47920
Damage Index (1 km)	DAM100	0	0	0	417	0	11520
Infr. Dens. (10 km)	DEN10	0.0	0.0	0.0	0.2	0.1	4.9
Infr. Dens. (5 km)	DEN25	0.0	0.0	0.0	0.1	0.1	4.2
Infr. Dens. (2.5 km)	DEN50	0.0	0.0	0.0	0.2	0.1	8.8
Infr. Dens.(1 km)	DEN100	0.0	0.0	0.0	0.2	0.0	16.0

Note that debris items tend to be located closer to land and waterways in areas with higher wind speeds and storm surges, and in areas with higher damage indices. These results are consistent with intuition. We also note that storm surge exceedance and wind speed seem to have different coefficient signs, and that all the infrastructure density variables have negative coefficients

TABLE 4. Univariate model coefficient estimates and parameter *p*-values for logistic regressions of binary debris item presence versus candidate model variables.

Code	Coefficient	Parameter *p*-value
WSPMAX_MS	-0.03774	<0.0001
SURGE_M	0.5596	<0.0001
BATHY_M	0.3202	<0.0001
DLAND_M	-0.0004684	<0.0001
DWWAY_M	-0.0002258	<0.0001
DAM10	0.001428	<0.0001
DAM25	0.0001556	<0.0001
DAM50	0.00003891	<0.0001
DAM100	0.00001257	<0.0001
DEN10	-0.10679	<0.0001
DEN25	-0.54238	<0.0001
DEN50	-1.1999	<0.0001
DEN100	-2.0856	<0.0001

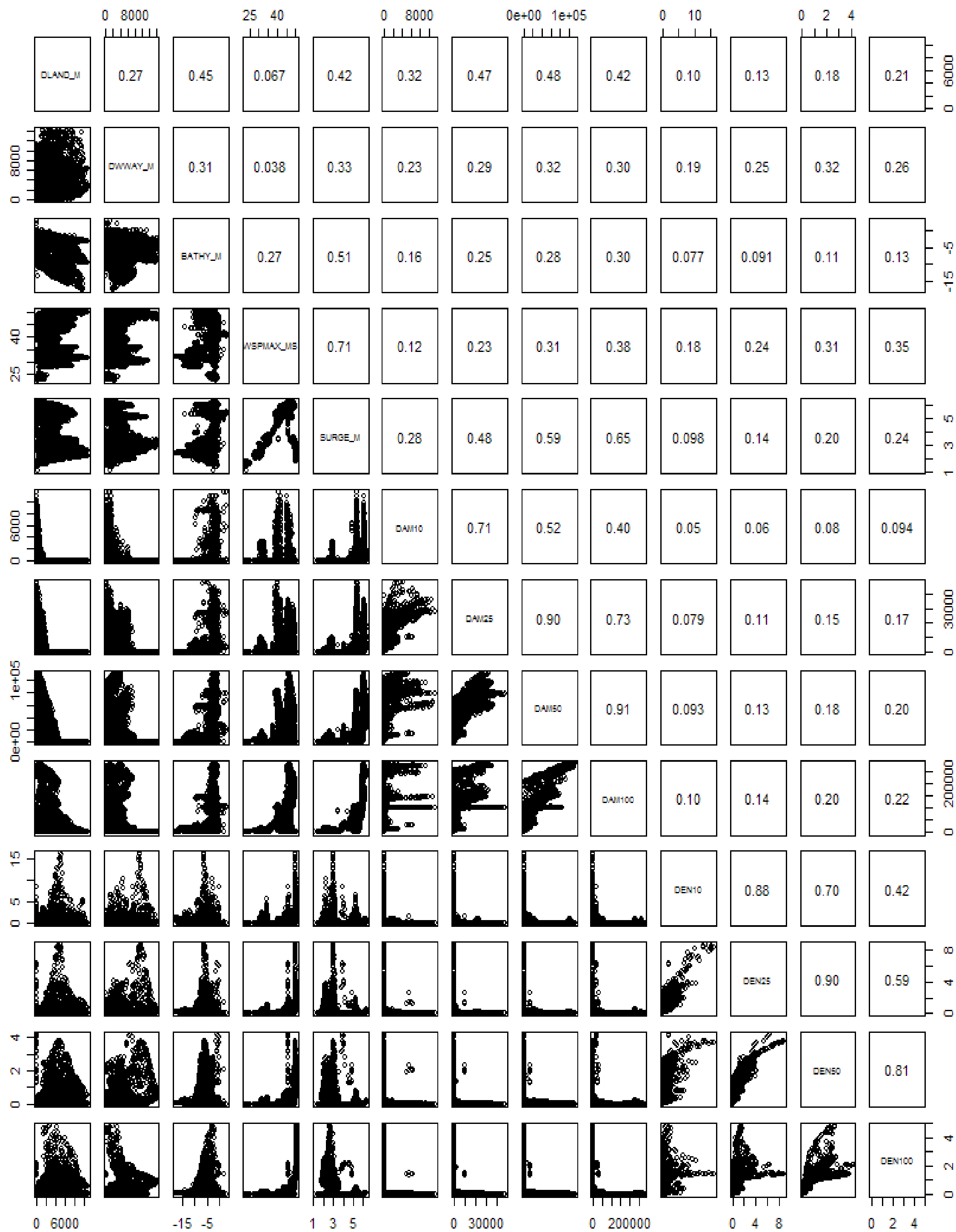


FIGURE 9. Pairs plots (below the diagonal) and absolute correlation coefficients (above the diagonal) for the evaluated predictor variables.

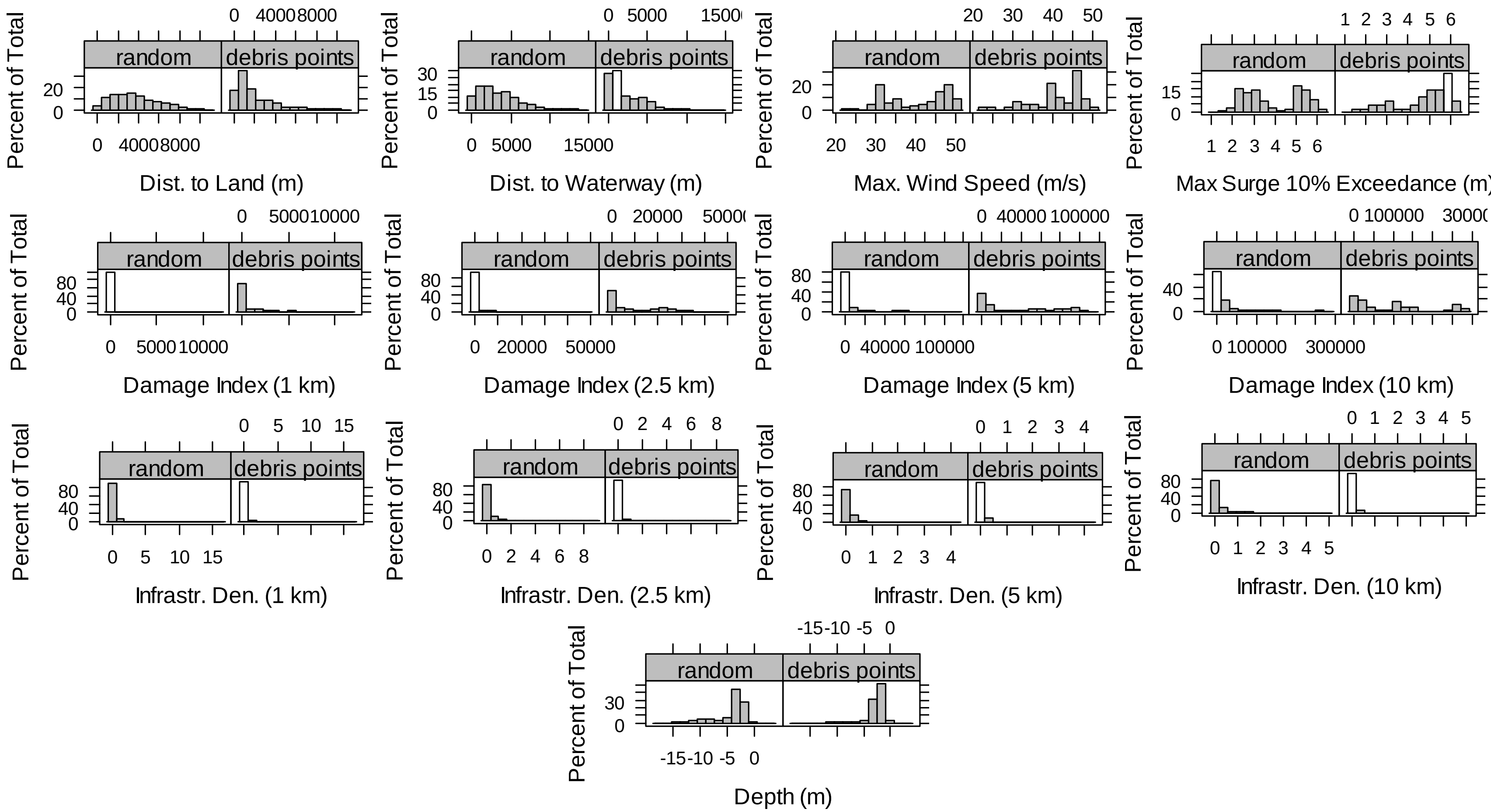


FIGURE 10. Histograms of percentage of observations of item locations and random pseudo absences by each of the evaluated predictor variables.

Model Construction

The strong correlations between some of the candidate predictors were noted above. Strong correlations between predictor variables can be problematic with some modeling strategies. In this application, there is no need to be overly concerned with this multicollinearity, because pure prediction is the goal. However, care should be used in evaluating the sign and significance of individual model coefficients. An initial multivariate logistic regression model was constructed with the *glm* procedure in R. The multivariate model was then constructed via stepwise forward variable selection as implemented via the *stepAIC* function in R (Venables and Ripley, 2003)

Bayesian Model Averaging (BMA) was also evaluated as a candidate for final model construction. BMA accounts for the uncertainty inherent in the model selection process by averaging over many different competing models, rather than attempting to select the unique group of variables that make the best model. BMA incorporates model uncertainty into conclusions about parameters and prediction and avoids the potentially statistically hazardous process of selection from a set of multiple variables. Hoeting *et al.* (1996, 2002) and Raferty *et al.* (1997) provide a review of BMA and its implementation in the R environment. The BMA derived multivariate model was derived via model averaging as implemented by the *bic.glm* function in R.

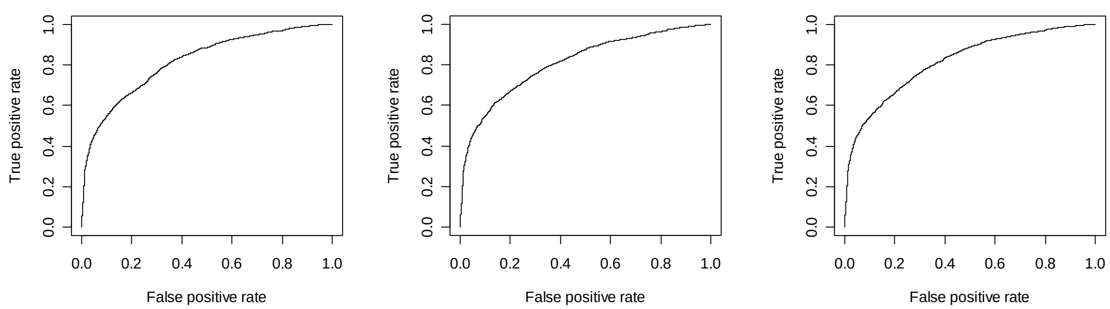
The performance of the classifier model against can be evaluated by examining the Receiver Operator Characteristic (ROC) curve. The estimated probability of a debris item occurring at each location – both the item locations and the random locations – is calculated using the model coefficients. The ROC curve then plots the false positive rate versus the true positive rate. The area under this curve (AUC) ranges from 0 to 1 and indicates the overall performance of the model. Model ROC curves (Figure 11) and performance statistics were calculated in R using the *ROCR* package (Sing *et al.*, 2005).

Initial trials indicated that the inclusion of the infrastructure density variables in the model yields surprisingly unstable results—particularly when extrapolating to prediction areas in southern Louisiana where infrastructure densities are many orders of magnitude higher than the coasts of Alabama and Mississippi. A version of the model using the same stepwise procedure but with no infrastructure variables was also evaluated. Table 5 lists model coefficients and AUC statistics for the BMA derived, stepwise logistic and stepwise logistic with no infrastructure density variables regression models.

With the exception of maximum wind speed, the same predictor variables were retained in each model. Specifically, wind speed is assigned a negative coefficient indicating higher wind speeds are associated with lesser debris probabilities. This is somewhat counterintuitive. Distance to waterway was retained by the stepwise procedure operating on the full model, but was dropped by the stepwise procedure when the infrastructure variables were removed. It was also not included in any of the BMA model ensembles.

TABLE 5. Model coefficients and ROC area under the curve AUC statistics for stepwise, stepwise without infrastructure variables, and BMA derived logistic models.

Coefficient	**Stepwise Model**	**Stepwise Model (No Inf. Vars.)**	**BMA Model**
Intercept	3.317925209	2.857604815	3.4354686055
Max. Wind Speed (m/s)	-0.122662103	-0.098712904	-0.1383307294
Max. Storm Surge (m)	0.380548628	0.273156107	0.4805729010
Dist. to Land (m)	-0.000113435	-0.000119084	-0.0001046976
Dist. to Waterway (m)	-0.000020349	-	-
Elevation (m)	0.072490502	0.078413039	0.0726930590
Damage Index (1 km)	0.000295082	0.000310402	0.0003309315
Damage Index (2.5 km)	0.000295082	0.000041568	0.0000260500
Damage Index (5 km)	0.000011030	0.000012413	0.0000191179
Damage Index (10 km)	0.000002470	0.000001969	0.0000004563
Infrastructure Density (1 km)	0.619554410	NA	0.5790855283
Infrastructure Density (2.5	-1.165087074	NA	-0.9312981047
Infrastructure Density (5 km)	1.317550611	NA	0.8113536116
Infrastructure Density (10	-0.457476689	NA	-
ROC Area Under the Curve	**0.82**	**0.81**	**0.82**



A. B. C.

FIGURE 11. Receiver operator characteristic (ROC) curves for stepwise (A), stepwise with no infrastructure variables, and BMA derived (B) logistic models.

The performance of all three models is good and essentially identical as evaluated via the AUC statistic and ROC curves. The benefits of the BMA method are largely lost in this particular application, as there seems to be small uncertainty about model selection. Also, the inclusion of the variables yields essentially no improved predictive power as measured by AUC (Table 5). As such, the stepwise model without infrastructure density variables was selected as the final model.

Results

The final model can be expressed as an equation that calculates the logit of the debris item encounter probability as a linear function of the predictor variables. Table 6 lists estimated coefficients and standard errors of those coefficients for the final model. Figure 12 displays the final model as implemented in ArcGIS via map algebra. The coefficients were used to calculate the logit, which was then converted into a probability ranging from 0 to 1 for each grid cell. In general, notice that, unsurprisingly, the highest predicted probabilities correspond with areas from Figure 7 with large amounts of debris, and also with areas that sustained high storm energies and resultant damage. Thus, the model corresponds well with the intuitive and actual knowledge of marine debris distributions in the Gulf of Mexico following the passage of Hurricane Katrina.

TABLE 6. Final model estimated coefficient values and standard errors.

Coefficient	Coefficient	Std. Error
Intercept	2.857604815	0.220188877
Dist. to Land (m)	-0.000119084	0.000015887
Max. Wind Speed (m/s)	-0.098712904	0.006942386
Max. Storm Surge (m)	0.273156107	0.042173799
Elevation (m)	0.078413039	0.016030242
Damage Index (1 km)	0.000310401	0.000055248
Damage Index (2.5 km)	0.000041568	0.000041568
Damage Index (5 km)	0.000012413	0.000003820
Damage Index (10 km)	0.000001969	0.000000945

It is in this map form that model output following a future storm event will be most useful. This grid of continuous marine debris item encounter probabilities could be sliced at some arbitrary value to derive hotspot areas of an arbitrary size. Alternatively, if only a fixed amount of survey effort was available after an event, such data could be used to ensure that the highest probability areas were included in available survey effort – maximizing survey efficiency.

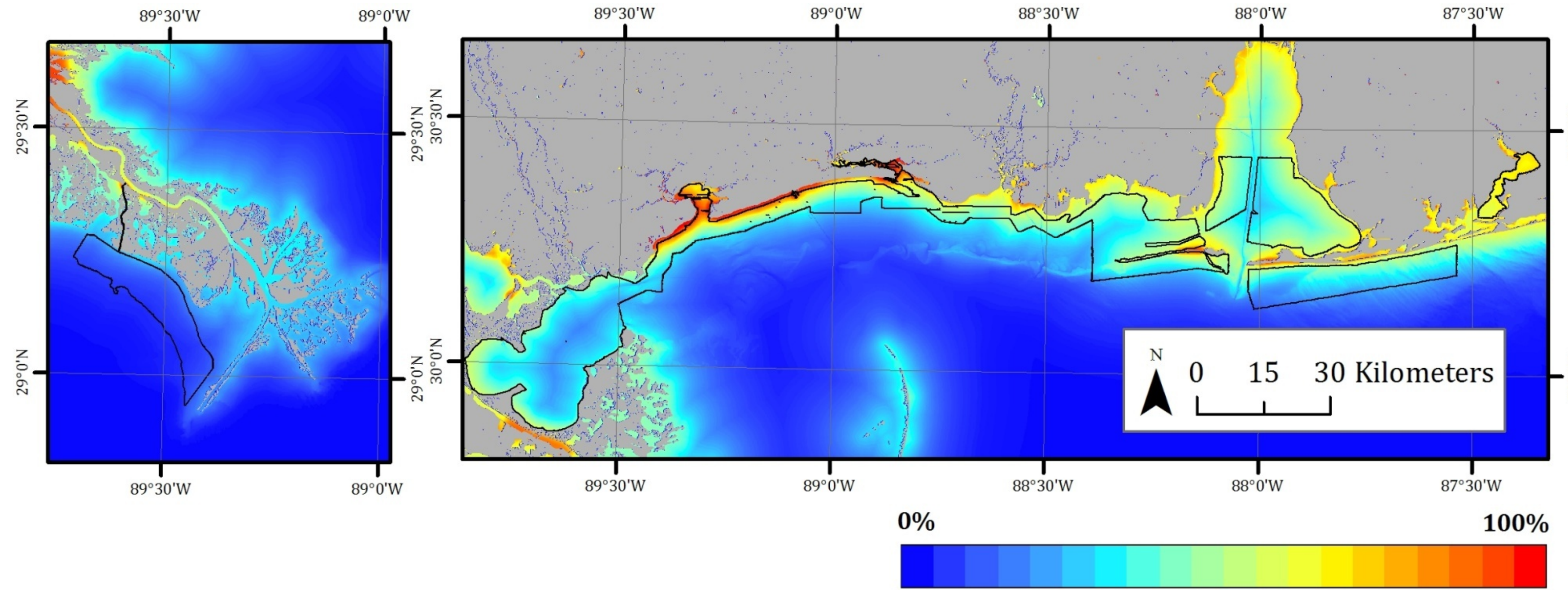


FIGURE 12. Final model prediction map. Reds indicate areas of higher predicted likelihood of marine debris concentrations.

Uncertainty

Sources of uncertainty in this modeling process are manifold. The initial data collection of each predictor data set involves unique assumptions and uncertainty, both spatial and measurement. Additionally, the way these data are spatially referenced – vector points, grids of various resolutions, etc. – was modified as described, resulting in statistical change-of-support issues. Finally, many of the predictors used here are indices (densities or focal sums) based upon other datasets, and were derived based upon logical but unexamined assumptions. All of these issues should be kept in mind when evaluating this model or constructing a similar one. The procedure of logistic regression, however, yields as its result a probability as well as confidence intervals around that probability, rather than a discrete outcome. As such, the model inherently accounts for a large portion of the uncertainty of the relationships used to construct it.

Finally, note that the model output for physical settings very different from the areas where data was collected, in particular inland water bodies such as lakes and reservoirs, is likely to be extreme and not reflective of reality. Model output should only be used in areas similar to those in which these data were collected – nearshore marine and estuarine zones.

Model Validation

Initially, the data from the southern portion of the study area in Plaquemines parish which were collected somewhat later in the process were held aside to conduct an external model validation with independent data. However, there were few debris points in this area, and the area was both remote and homogenous with respect to predictor data. As such, this external validation was less than useful.

Instead, we conducted an internal validation—specifically, a 10-fold cross validation of model accuracy. In this procedure, the data were divided into 10 parts. The stepwise model procedure was carried out 10 times, each time using 9/10^{ths} of the data. The model, derived in each iteration, is tested using the 1/10th of the data held out. Summary statistics are then generated. The 10-fold cross-validated ROC AUC is depicted in Figure 13. The average AUC of these cross-validated runs is 0.81 (95%CI: 0.80, 0.82). This tight grouping indicates that the model is well-calibrated and not overfit. Nonetheless, validation with data from another storm event would be preferable.

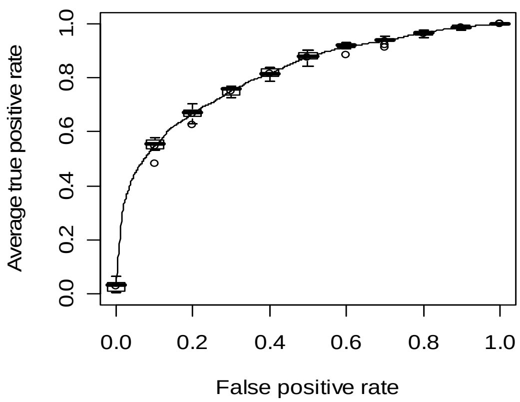


FIGURE 13. Vertical average of 10-fold cross-validated ROC curves with box plots for final logistic model.

Generalization

It is the goal of this effort, as stated above, to construct a model that will be useful in future applications. The key to this process is the ability of the model to generalize well to future unknown storm events. Obviously, the data that were used to construct this model represent a distribution of marine debris from one particular storm in one particular region: Hurricane Katrina in the northern Gulf of Mexico. As such, the model summarizes trends and correlations particular to this event.

The only way to either test or improve the generalization of the model is with quantitative data from other storm events in other locations. No information could be found for other actual quantitative data describing marine debris distributions post-storm that could be used to directly test the ability of this model to generalize at the current time. However, it is believed that the general trends captured by the model – debris tends to be located close to debris sources and in areas of high storm energy – are intuitive and will hold true in other circumstances. Appendix A contains model output products for Hurricane Ivan in 2004 and Hurricane Ike in 2008. Efforts are currently underway to collect marine debris data in the western portion of Louisiana's coast in the impact area of Hurricane Rita. These data may provide valuable insights into how well this model performs for other storm events, as well as the raw materials with which to improve its performance.

References

Divins, D., and D. Metzger. 2007. NGDC coastal relief model. Available at: http://www.ngdc.noaa.gov/mgg/coastal/coastal.html. (10/20/06)

Environmental Systems Research Institute, Inc. (ESRI). 2007a. ArcGIS v9.2. ESRI, Redlands CA. Available at: http://www.esri.com. (8/20/07)

Environmental Systems Research Institute, Inc. (ESRI). 2007b. ArcData. Available at: http://www.esri.com. (8/20/07)

Federal Emergency Management Agency (FEMA). 2006. Hurricane Katrina remote sensing data. Available at: http://www.gismaps.fema.gov/2005pages/rsdrkatrina.shtm. (8/20/07)

Federal Emergency Management Agency (FEMA). 2008. Hazus-MH MR3. Available at: http://www.fema.gov/plan/prevent/hazus/ (1/20/08).

Geological Survey of Alabama (GSA). 2008. Offshore oil and gas wells. Unpublished.

Hoeting, J. A., Raftery, A. E., and D. Madigan. 2002. A method for simultaneous variable and transformation selection in linear regression. Journal of Computational and Graphical Statistics. 11:485-507.

Hoeting, J. A., Raftery, A. E., and D. Madigan. 1996. A method for simultaneous variable selection and outlier identification in linear regression. Computational Statistics and Data Analysis. 22:251-270.

Louisiana Oil Spill Coordinators Office (LOSCO). 1999. Navigated waterways of Louisiana. Available at: http://lagic.lsu.edu/data/. (8/20/07)

Louisiana Department of Natural Resources (LDNR). 2007. Oil, gas, and injection wells in Louisiana. Available at: http://lagic.lsu.edu/data/. (8/20/07)

Minerals Management Service (MMS). 1998. Oil and gas producing platforms in the Gulf of Mexico. Available at: http://lagic.lsu.edu/data/. (8/20/07)

Mississippi Automated Resource Information System (MARIS). 2004. Oil and gas wells. Available at: http://www.maris.state.ms.us/statewide.html. (1/20/08)

National Oceanic and Atmospheric Administration (NOAA). 2007a. Experimental probabilistic storm surge data. Available at: http://www.weather.gov/mdl/psurge/. (8/20/07)

National Oceanic and Atmospheric Administration (NOAA). 2007b. H*Wind surface wind analysis data. Available at: http://www.aoml.noaa.gov/hrd/data_sub/wind.html. (8/20/07)

Powell, M., Houston, S., Amat, L., and N. Morisseau-Leroy. 1998. The HRD real-time hurricane wind analysis system. Journal of Wind Engineering and Industrial Aerodynamics. 77&78:53-64.

R Development Core Team. 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org. (8/20/07)

Raftery, A. E., Madigan, D., and J. A. Hoeting. 1997. Bayesian model averaging for linear regression models. Journal of the American Statistical Association. 92:179-191.

Sing, T., Sander, O., Niko Beerenwinkel, N., and T. Lengauer. 2005. ROCR: visualizing classifier performance in R. Bioinformatics 21(20):3940-3941.

Umpierre, D and G. Margoles. 2005. Broward County's web-based hurricane debris estimation tool (HurDET). In: 2005 ESRI International User Conference Proceedings.

United States Army Corps of Engineers (USACE). 2007. National waterway network. Available at: http://www.iwr.usace.army.mil/ndc/data/data1.htm. (8/20/07)

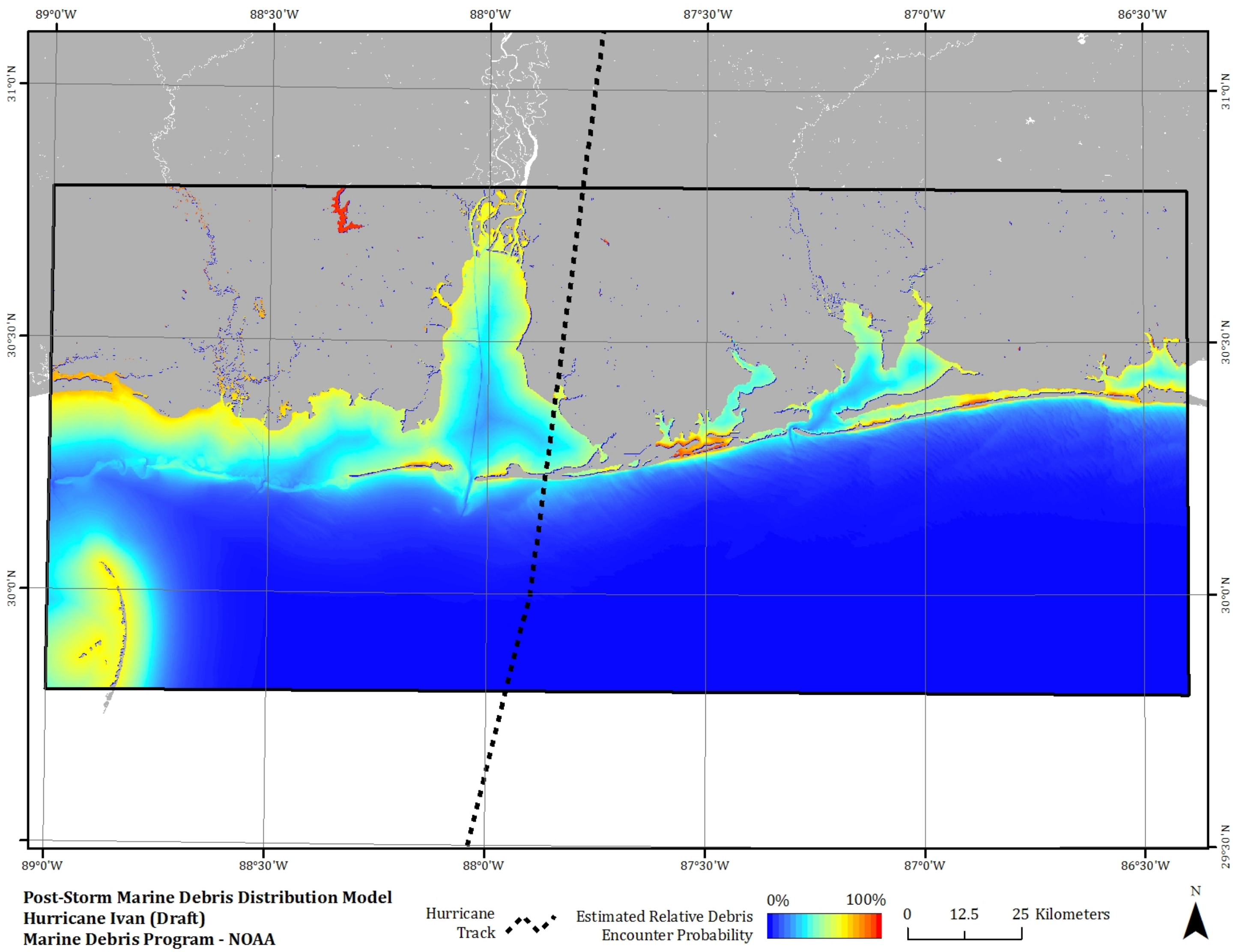
Venables, W.N, and B.D. Ripley. 2003. Modern Applied Statistics with S. Springer, New York, NY. 495 pp.

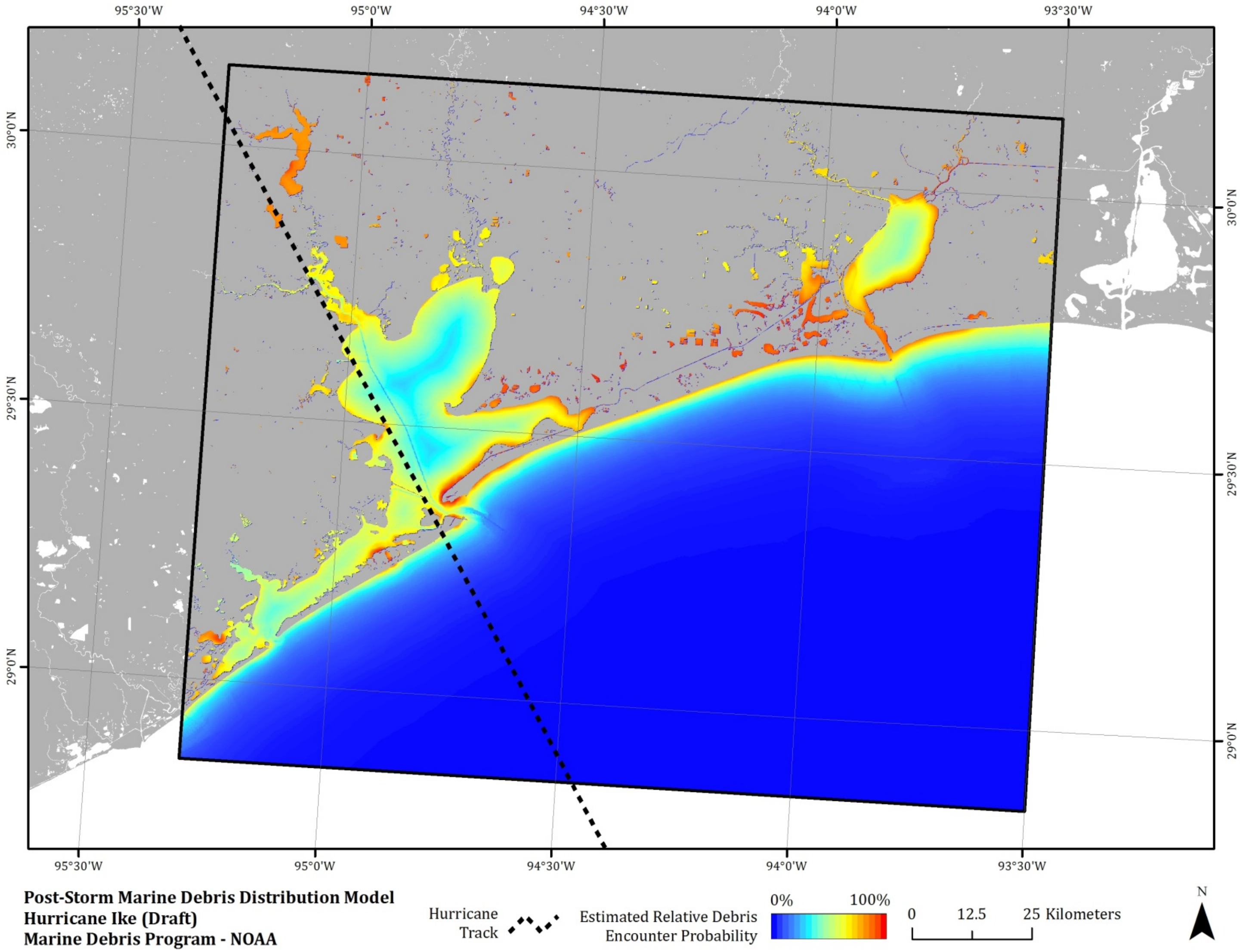
Appendix A – Additional Model Results

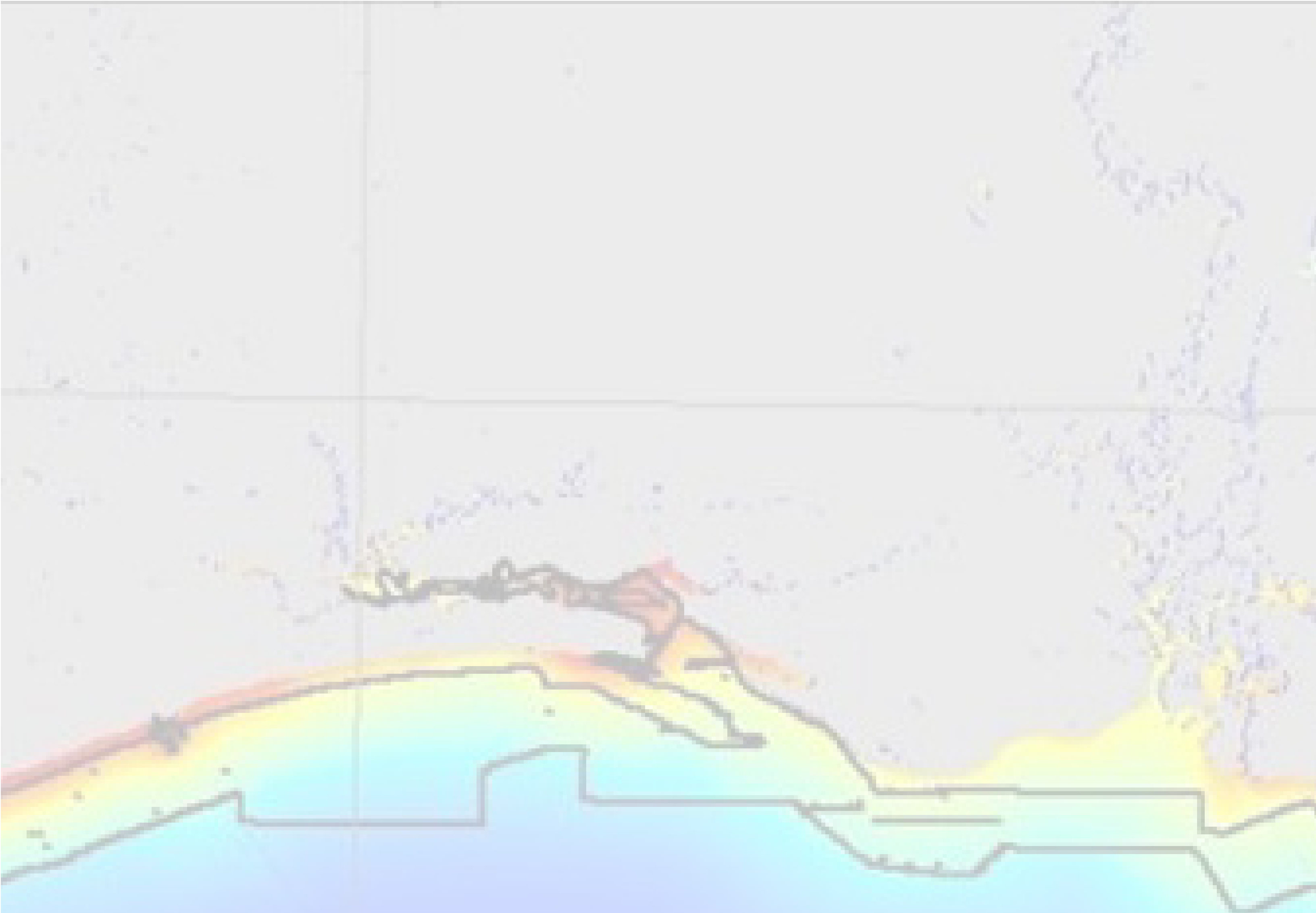
The model described above was used to generate marine debris predictions for coastal areas impacted by the following additional storms:

- Hurricane Ivan on September 16, 2004
- Hurricane Ike on September 13, 2008

The following maps depict the model output for those storms. Model output is also available as a raster layer for use with GIS software in ESRI® GRID format. Contact the NOAA Marine Debris Program at www.MarineDebris.noaa.gov or 301.713.2989 for data or information requests.


89°0'W
88°30'W
88°0'W
87°30'W
87°0'W
86°30'W
31°0'N
30°30'N
30°0'N
29°30'N
Post-Storm Marine Debris Distribution Model
Hurricane Ivan (Draft)
Marine Debris Program - NOAA
Hurricane Track
Estimated Relative Debris Encounter Probability
0%
100%
0 12.5 25 Kilometers
N


95°30'W
95°0'W
94°30'W
94°0'W
93°30'W
30°0'N
29°30'N
29°0'N
Post-Storm Marine Debris Distribution Model
Hurricane Ike (Draft)
Marine Debris Program - NOAA
Hurricane Track
Estimated Relative Debris Encounter Probability
0%
100%
0
12.5
25 Kilometers
N



United States Department of Commerce

Gary Locke
Secretary

National Oceanic and Atmospheric Administration

Jane Lubchenco, Ph.D.
Undersecretary of Commerce for Oceans and Atmosphere
Administrator, National Oceanic and Atmospheric Administration

National Ocean Service

David Kennedy
Acting Assistant Administrator for Ocean Services and
Coastal Zone Management