FLORIDA SEA GRANT COLLEGE PROGRAM
R/C-P-24

FLSGP-T-02-005 C2

Blueways Recreational Boating Characterization for Charlotte Harbor, Florida

Trend Surface Analysis of Boating Use and Tests of Statistical Association Between Use and Various Environmental and Occurrence Variables

Report to the FMRI in Fulfillment of January – June, 2002 Services

June 10, 2002

Charles Sidman
Florida Sea Grant
University of Florida

Timothy J. Fik
Department of Geography
University of Florida

Richard Flamm
Manatee Section
Florida Marine Research Institute

Introduction and Objectives

The focus of this component of the ongoing Blueways recreational boating characterization was to conduct exploratory data analyses that test the statistical association between boating use and various environmental and occurrence variables. The report is divided into two parts, consistent with the following objectives.

1. Map boating use by one-quarter-mile square grid cell. This required the following tasks:

    a. Improve the accuracy of digitized travel routes by incorporating sea grass and navigation aids to ensure the best placement of interpreted travel routes from mail surveys.

    b. Integrate mail and aerial survey data themes.

    c. Estimate trend surfaces for recreational boating routes and destination locales.

2. Explore the associations between boating use, accidents, and environmental variables. This involved the following tasks.

    a. Conduct a GIS analysis to establish buffers around the network, manatee sighting and mortality locations, and mail and aerial survey boat locations and travel routes. Identify buffer areas that intersect with the boating use trend surfaces for routes and destinations as input for association analyses.

    b. Explore the spatial associations between boating use and boating accidents, sea grass scarring, manatee sightings, and manatee mortality.

Part 1. Trend Surface Analysis of Boating Probability

Data Compilation

The trend-surface model was developed to estimate boating use for those areas (one-quarter-mile square grid cells) within the study region for which no aerial reconnaissance or mail survey data were available. The following GIS data were used as model input:

- Travel routes obtained from mail surveys (identify primary travel network).
- Destination points obtained from mail surveys (identify preferred boating locales).
- Moving boats identified from aerial reconnaissance (correspond to primary

travel network).

- Stationary boats obtained from aerial reconnaissance (correspond to preferred boating locales).
- One-quarter-mile-square grid generated within the GIS, clipped to conform to the study-area shoreline.

Data for the trend surface analysis were generated within the Arc/Info geographic information system (GIS). The statistical analysis was conducted within the Number Cruncher Statistical System (NCSS) 2001 statistics software.

Travel Route Adjustments

The initial digitization effort located travel routes on the basis of general bathymetry and a one-mile square grid; the grid size depicted on the map that was included with the mail survey questionnaire. Digitized travel routes were re-evaluated for positional accuracy to meet the higher mapping resolution (one-quarter-mile square) selected for this analysis. Sea grass and navigational aids were incorporated as background themes within the GIS to obtain the most accurate representation of travel routes (Figure 1).

Boater travel routes and destinations were digitized into the Arc/Info GIS from mail surveys conducted by Sidman and Flamm (2001). Mail survey respondents were asked to trace their last two travel routes and identify favorite destinations along those routes. As such, mail survey data reflect boating patterns during the survey months of April – June 2000. Moving and stationary vessels were identified by aerial reconnaissance and digitized into the Arc/Info GIS (Gorzelaney, 1998). Aerial flights for Lee County took place during sixteen randomly selected dates during January – December 1997, and January – December 2000 for Charlotte County. The trend analysis used a composite representation of all mail and aerial survey data.

Data Integration

Boat data for routes and destinations (mail survey), and moving and stationary vessels (aerial reconnaissance) were aggregated to one-quarter-mile by one-quarter-mile grid cells using an Arc/Info INTERSECT overlay process (Appendix A). The study area contains n=7577 one quarter-mile square cells, covering an area equivalent to approximately 43 square miles.

Trend Surface Analysis

A trend-surface analysis was employed to model and generate estimates of route and destination use in one-quarter-mile-square grid cells for which no mail or aerial survey data were collected. Estimates were mapped by linking boating-use results from the trend surface analysis back to each one-quarter-square mile grid cell (Figure 2). This was accomplished by the use of a key (relate) field that contained a unique number for each grid cell.

A least-squares stepwise regression approach was utilized to "best fit" the trend surface as opposed to employing an interpolation technique that relies solely on localized averaging or smoothing. The regression approach to trend surface estimation is preferred to a simple descriptive approach for a variety of reasons. Specifically, trend-surface estimation is appropriate given that the spatial distribution of the average proportion of either moving boats or stationary boats by cell represents only a sample or one possible realization from an infinite (or indefinitely large) number of possible samples that could have been drawn from the respective populations. In addition, the trend surface estimation is able to distinguish between a complex variety of spatial trends in the estimated surface that can be local and/or regional in nature.

Methods

The trend surface analysis was carried out using a 4^{th} order polynomial model. The model was estimated using a stepwise-regression procedure, allowing for all possible interactions between independent variables (up to an implied 4^{th} order). The stepwise regression procedure reduces the likelihood of multi-collinearity amongst explanatory variables. A list of the variables used in this analysis is given in Table 1. Routes and destinations were modeled separately. Each model generated four estimates for each grid cell: 1) an estimated value; 2) an average estimated value; 3) a lower estimated value and, 4) an upper estimated value. In cases where the model generated a negative probability estimate for a given cell, the estimated number was restricted to a value very close to zero (.000001). The lower limit of the mean estimation interval for these cases was also reset to a value of .0000001 (with a minimum upper limit of .000002).

Individual trend surfaces were constructed from the average estimated values of routes (RPAVG) and destinations (DPAVG) for each one-quarter-mile square grid cell. These values were then divided by the maximum observed value within the respective surfaces to produce separate boating use composite coefficients for routes (BUCR) and destinations (BUCD). The BUCR and BUCD indices have a minimum value of zero and a maximum value of 1.0 (Figure 3).

Results and Discussion

The trend surface model for routes generated an R-square of 0.4759, and an R-square for non-zero values of 0.5301 (Table 2). The network variable (NET) derived from a one-quarter-mile GIS buffer of the main travel network was the most significant explanatory variable. Other significant variables include moving boats (MB), and the interaction of moving boats and the network (NET*MB). Model results for destinations show an R-square of 0.3701, and an R-square for non-zero values of 0.4015 (Table 3). Network variables proved to be the most significant predictors of destination use. Many cells in the sample display an average proportion value of zero (i.e., no boat was observed either passing through the cell or occupying locations within the cell as a final destination in either the mail or aerial surveys). Given that we are dealing with a sample, a proportion value of zero may not accurately reflect a given cell's potential for on-water boating use. In other words, it is possible that boaters could utilize a cell with an average proportion value of zero in a traveling or destination capacity despite the fact that no activity was observed in that cell in either survey. This is true, of course, unless controlling depth or water restrictions prohibit travel through or to that cell. Just as proportion values may be understated for some cells, proportion values may also be overstated in various cells within the samples. Trend surface modeling can help overcome this problem of over or under estimation by generating an expected probability value on a cell-specific basis.

Trend surface models have the added advantage of allowing one to test the statistical significance of multiple influences and/or variables. This quality serves a three-fold purpose. First, it helps to verify that the observed trend surface is not simply a random spatial pattern. Second, it allows for statistical validation of the model and comparison of the relative strength of each independent variable and the degree to which each independent variable explains variation in the dependent variable and how the average proportion of survey routes or stationary boats per cell vary over space. Third, the trend-surface analysis provides a statistic that can be used to evaluate the overall accuracy of the model to replicate the observed spatial pattern. More importantly, perhaps, is the fact that trend-surface estimation allows for the prediction of values by geographic location and provides confidence bands (upper and lower limits) about those predicted values. This is useful in assessing whether cells with values of zero (or values close to zero) within the sample are underestimated given their geographic locations, characteristics, and the exposed spatial trend. It also allows one to examine the low- and high-end estimates of the proportions on a cell-by-cell basis.

The estimation of individual models captured the 'general' spatial trends unique to the use type (BUCR versus BUCD). Nevertheless, the models did not fully replicate the respective trend surfaces for three reasons. First, there is a fairly high percentage of cells with average proportion values of zero in each database; something which is expected given the highly resolute nature of the analysis (i.e., the size of grid cells) and the relatively small number of vessels observed in each sample compared to the total number of vessels in the study region. Second, the sample-specific nature of the observed trend surfaces limits the degree to which the data can be used to estimate the expected values per cell. Lastly, the number of explanatory variables is limited to a handful of geographic variables and by no means represents an exhaustive list of all possible influences that affect variations in on-water use within the study region. Nonetheless, the estimation procedure generated statistically valid and reasonable estimates for the expected proportions for both the network/moving boat surface and the destination/stationary boat surface.

Conclusions

Trend surface estimation revealed distinct spatial patterns in route and destination use. Despite the differences in explanatory variables, network and geographic variables were consistently found to be significant and influential in explaining variation in the observed proportions. ***The findings of this more comprehensive trend surface approach corroborates a previous study that highlighted the overarching importance of network variables in estimating preferred recreational boating destinations (Sidman and Fik, 2002).***

A boating trend surface provides the basis for measuring correlations or identifying "hotspots" – localized areas where BUCR or BUCD values and the values of environmental or occurrence variables are highly spatially covariant. The identification of hotspots could then be used to demarcate zones where recreational boating activities have the greatest consequences to either the environment or other users.

Table 1. List of Trend Surface Analyses Variables

Variable	Description
ID	cell identification number
Analysis Variables	
MRP	proportion of mail survey routes per cell
ARP	proportion of aerial survey routes (moving boats) per cell
MDP	proportion of mail survey favorite destinations per cell
ADP	proportion of aerial destinations (stationary boats) per cell
RPAVG*	route proportion average = (MRP + ARP) / 2
DPAVG*	destination proportion average = (MDP + ADP) / 2
X	location coordinate of cell denoting longitude (scaled: divided by 1,000)
Y	location coordinate of cell denoting latitutude (scaled: divided by 1,000)
NET	network buffer dummy (NET=1 if cell is within ½ mile of main on-water travel network; NET=0 if not)
MB	moving-boat buffer dummy (MB=1 if cell is within ¼ mile of moving boat corridors; MB=0 if not)
FAV	favorite destination dummy (FAV=1 if cell is within ¼ mile of favorite destination area; FAV=0 if not)
SB	stationary boat buffer dummy (SB=1 if cell is within ¼ mile of stationary boat area; SB=0 if not)
TF**	dummy variable to denote presence (or absence) of tidal flat: TF=1 (TF=0)
SCAR**	dummy variable to denote moderate or severe sea grass scarring (SCAR=1) versus little or no scarring (SCAR=0)
ACC**	proportion of accidents per cell
MAN**	estimated proportion (relative density) of manatees per cell

* dependent variables used in trend surface models
** derived variables from map comparison/correlation analyses

Table 2. Results Trend-Surface Model for Travel Routes/Moving Vessels

Dependent Variable: RPAVG
Mean = 9.883141E-03
Standard Deviation = 2.104318E-02
Minimum = 0; Maximum = 0.2070056

Procedure: Stepwise Regression w/interactive variables and polynomial (4th order) using inclusion/exclusion criteria of .05/.10

Independent Variable(s)	Regression Coefficient	t-value(prob level)	
Intercept	1.670709	2.0971	.03601
X	2.401089E-04	2.8748	.00405
Y	-6.924920E-03	-2.2871	.02223
NET	175.11610	20.9763	.00000
MB	0.316997	3.0419	.00235
Y*Y	7.350677E-03	2.3357	.01953
X*X*X	-2.365574E-04	-3.2700	.00108
Y*Y*Y	-2.866066E-03	-2.3754	.01755
NET*X	-0.531749	-19.8224	.00000
NET*Y	-0.210360	-20.8632	.00000
NET*(X*Y)	0.485994	21.4868	.00000
NET*(X*X)	0.501275	17.6641	.00000
NET*(Y*Y)	6.005638E-02	16.0212	.00000
NET*(X*X*X)	-1.424969E-04	-13.7706	.00000
NET*(X*X*Y)	-2.609713E-04	-21.8552	.00000
NET*(Y*Y*X)	-8.839209E-05	-14.9820	.00000
MB*X	2.984959E-05	1.7345	.08287
MB*Y	-7.456802E-04	-3.0612	.00221
MB*(Y*Y)	4.167216E-04	2.9506	.00318
NET*MB	1.336674E-03	6.2850	.00000
NET*MB*FAV	-1.805973E-03	-2.0956	.03614

n=7577 observations (cells)
R-square = .4759
Recovered R-square for non-zero values = .5301
RMSE = 1.5250E-02
F = 343.0581 (.00000)

Table 3. Results of Trend-Surface Model for Destinations/Stationary Vessels

Dependent Variable: DPAVG
Mean = 1.268851E-04
Standard Deviation = 3.804947E-04
Minimum = 0; Maximum = 0.012303

Procedure: Stepwise Regression w/interactive variables and polynomial (4^{th} order) using inclusion/exclusion criteria of .05/.10

Independent Variable(s)	Regression Coefficient	t-value(prob level)	
Intercept	--	--	--
ARP	0.3180652	18.8695	.00000
RPAVG	-1.321826E-03	-4.5103	.00000
FAV	3.708957E-02	6.7882	.00000
SB	8.385341E-04	5.5243	.00000
NET*(X*Y)	1.434331E-05	5.0764	.00000
NET*(Y*Y)	-1.141261E-05	-5.5260	.00000
NET*(X*X*X)	-9.532606E-09	-4.5154	.00000
NET*(Y*Y*Y)	4.790915E-06	5.4548	.00000
SB*X	-1.252952E-06	-5.0933	.00000
FAV*X	-1.188960E-04	-6.6485	.00000
FAV*(X*Y)	1.379913E-05	2.8982	.00376
FAV*(X*X)	8.655321E-05	6.7614	.00000
FAV*(Y*Y)	-5.108842E-06	-2.9959	.00274

n=7577 observations (cells)
R-square = .3701
Recovered R-square for non-zero values = .4015
RMSE = 3.846269E-04
F = 194.5909 (.00000)

Note: The destination/stationary boat trend-surface model does not contain an intercept term given that it did not prove to be significantly different from zero at the 90% confidence level.

Figure 1. Travel Routes, Sea grass and Navigational Aids.

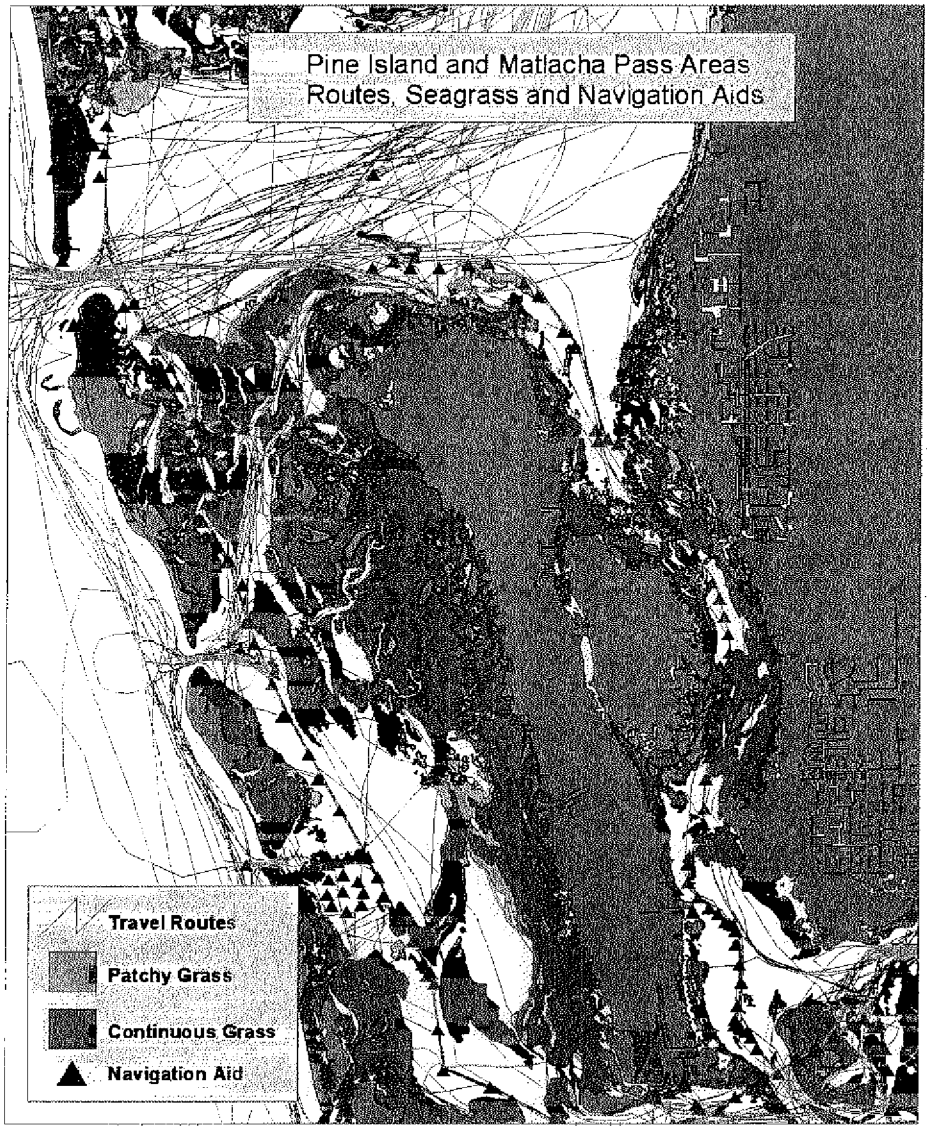


Figure 2. Trend Surfaces for Routes and Destinations.

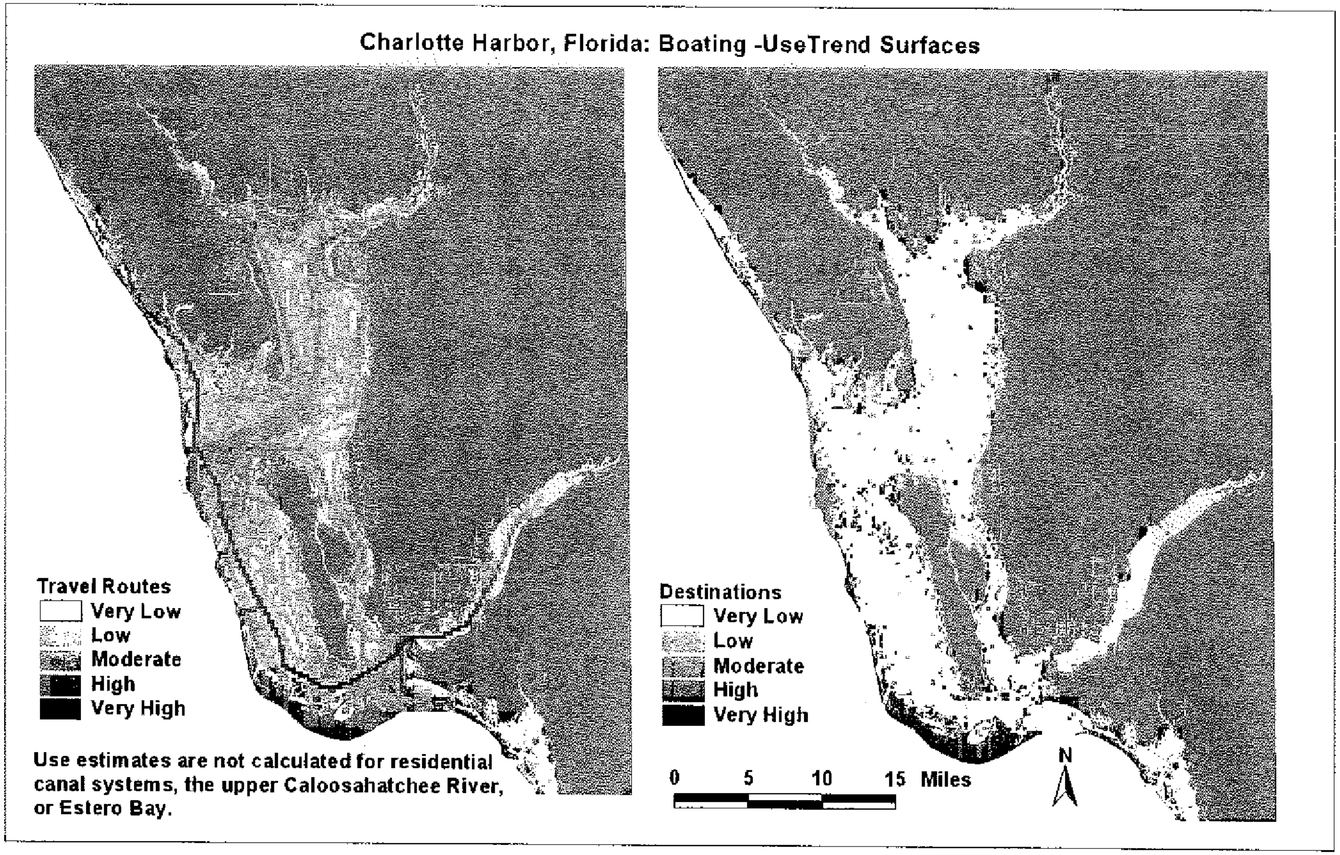
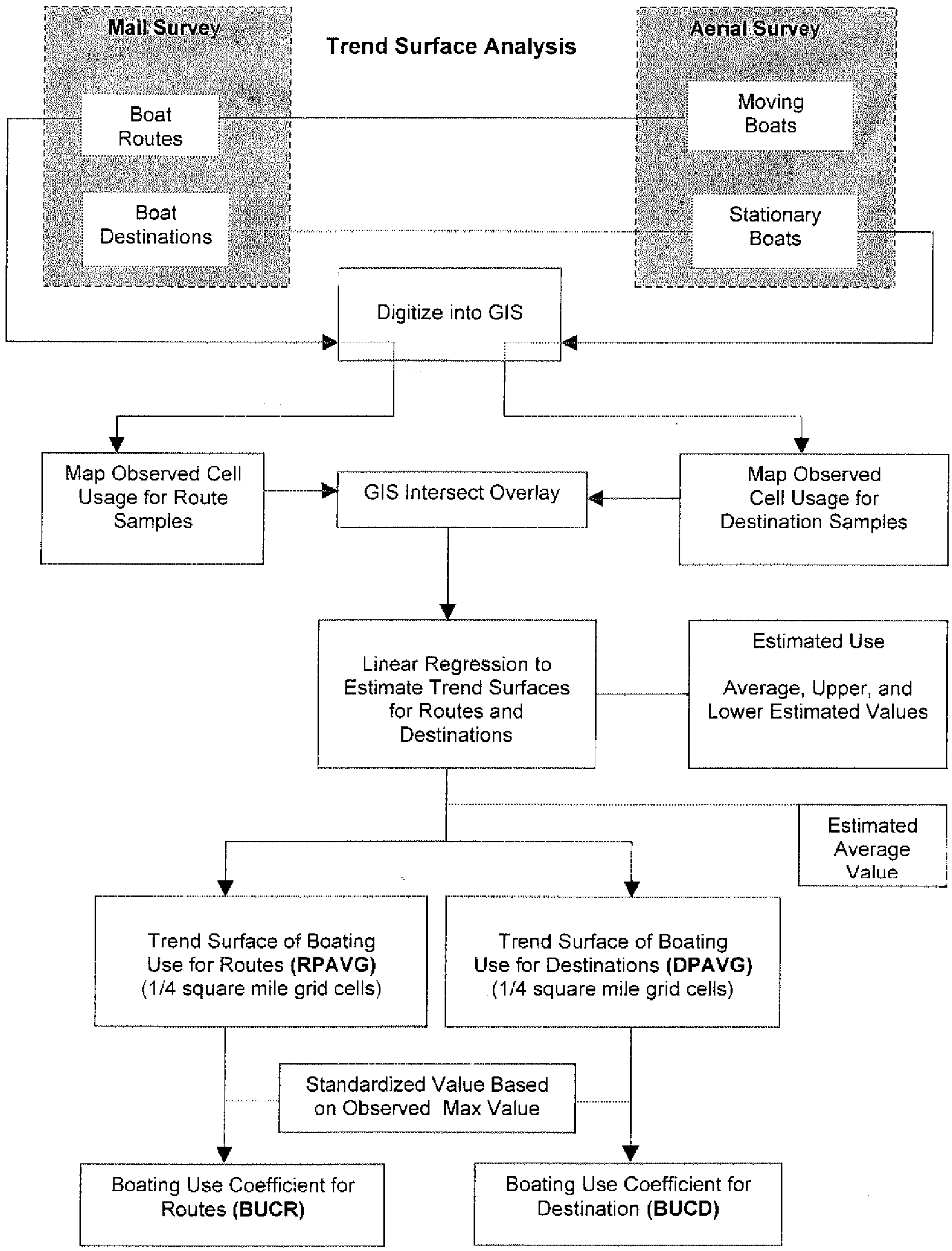


Figure 3. Trend Surface Analysis Flowchart.



Part 2. Exploring Associations Between Boating Use and Various Environmental and Occurrence Variables

Data Compilation

The second research component explores the extent to which there are potential 'regional' and/or 'local' associations between recreational boating use (as represented by the BUCR and BUCD trend surfaces) and various environmental and occurrence variables.

The analysis was restricted to cells with above-average boating route and destination use (i.e., cells in which BUCR and BUCD values exceeded their respective means), to test the statistical relationships between variables within higher-use areas of the trend surfaces. The sample size of the BUCR trend surface consisted of n=2060 total observations, while the sample size of the BUCD trend surface consisted of n= 1580 total observations. The following occurrence variables were selected for association analyses:

- **ACC** – the number of boating accidents per cell within the last year ACC=1 if accident was observed; ACC=0 if otherwise);
- **SCAR** – the observance of sea grass scaring within a cell (SCAR=1 if sea grass scarring is observed; SCAR=0 if otherwise);
- **MANP** – the proportion of manatee sightings by cell; and
- **MAND** – the proportion on manatee deaths by cell.

A variety of statistical procedures were chosen to test the relationships between boating use (BUCR and BUCD) indices, and the variables listed above. The specific choice of a technique was determined by data type, limiting assumptions, and the nature of the hypotheses to be tested. In many cases, non-parametric tests were employed given that these procedures require that fewer pretest assumptions be met.

Boating Use and Accidents

A two-sample Mann-Whitney U/Wilcoxon Rank-Sum Test was used to examine the relationship between boating accidents and the magnitude of the boating use coefficients (BUCR and BUCD). This Rank-Sum test is the non-parametric substitute for the equal-variance t-test when the assumption of normality is not valid (as in this study) and the distributions are found to be identical. The test requires that the BUCR and BUCD data be converted to the ordinal scale (ranking values from highest to lowest value and applying the usual convention for ties). The ranks of the two independent samples are then compared based upon the sum of the ranks in each sample (with the mean sum of ranks evaluated). The test hypotheses are as follows:

Null hypothesis: The difference in medians of ranked BUCR and BUCD values for cells in which accidents were observed (ACC=1) versus cells in which accidents were not observed (ACC=0) is equal to zero. In other words, there is an "equality of medians".

Alternative hypothesis: The difference in medians of ranked BUCR and BUCD values between where accidents were observed and not observed is less than zero, given that the mean values of BUCR and BUCD for ACC=1 are greater than the mean values of BUCR and BUCD for ACC=0. In other words, there is an inequality of medians and the median value for the sample when ACC=1 is significantly less than the median value for the sample when ACC=0.

Results

The results of the Rank-Sum tests are presented in Table 1. For BUCR, we "fail to reject" the null hypothesis that the difference in rank is equal to zero at the 95% confidence. For BUCD, we "reject" the null hypothesis at the 95% confidence. The latter result suggests that a statistical relationship does exist between boating use and accidents. Note that in both cases, the identical distribution assumption has been validated using the Kolmogorov-Smirnov D statistic or KS test (at the 95% confidence level).

Conclusion

There is evidence to support the contention that boating accidents are more likely to be observed in higher-use destination locales (where BUCD values are high). The results suggests that boating accidents are more likely to occur within cells or locations that are destination oriented rather than route oriented and/or vicinities were there are more observed stationary vessels. There is no evidence to support an assertion that accidents are associated with high-use travel corridors.

Discussion

This analysis was limited by the small sample size of recorded accidents (n=38 observations). This analysis can be improved by obtaining a larger historical sample. For example, a sequence of historical trend surfaces of boating accidents could be developed to further investigate the relationships between boating use and the occurrence of boating accidents and the flow patterns within high accident occurrence locales over time. Temporal accident information could be used to evaluate the potential effectiveness of on-water use restrictions and/or management tools (speed zones, channel markers, etc) to reduce the likelihood of boating accidents.

Table1. Mann-Whitney U/Wilcoxon Rank Sum Test for Differences in Medians for two independent samples (where ACC=1 versus ACC=0)

		Descriptive Statistics			
Variable: BUCR	*count(n)*	*mean*	*s*		*Norm. test decision**
for ACC=0	2048		0.147551	0.15432	reject normality
for ACC=1	12		0.170672	0.17112	reject normality
	n=2060				
Variable: BUCD					
for ACC=0	1570		0.045021	0.05417	reject normality
for ACC=1	10		0.104110	0.11546	reject normality
	n=1580				

*as based on skewness, kurtosis, ominbus normality test (at 95% confidence level)

Mann-Whitney U/Wilcoxon Rank Sum Test Results

Variable: BUCR

Group variable (sample)	U	Wilcoxon Sum of Ranks	Mean (W)
ACC=0	9469.5	2107646	2110464
ACC=1	15106.5	15184.5	12366

Standard deviation of W = 2051.566; Number of ties = 153

Alternative Hypothesis	Z-Value**	Prob-value	Decision (one-tail test)
Diff<0	1.3738	0.084747	Fail to reject null at 95% confidence

Kolmogorov-Smirnov (KS) test for different distributions: D = .292 < D critical (.3647):

Variable: BUCD

Group variable (sample)	U	Wilcoxon Sum of Ranks	Mean (W)
ACC=0	5424.0	1238659	1241085
ACC=1	10276.0	10331	7905

Standard deviation of W = 1416.81; Number of ties = 50

Alternative Hypothesis	Z-Value**	Prob-value	Decision (one-tail test)
Diff<0	1.7123	0.04342	Reject null at 95% confidence

Kolmogorov-Smirnov (KS) test for different distributions: D = .3101 < D critical (.3995):

** with correction factor for ties

Boating Use and Sea Grass Scarring

Non-parametric Rank-Sum and Chi-square test procedures were employed to examine the associations between boating use (BUCR and BUCD) and the occurrence of sea grass scarring. The Rank-Sum procedures were used to test for differences in medians for the two independent samples (cells where SCAR = 1 versus cells where SCAR = 0) given that the normality assumption was violated and the distributions were found to be identical using the KS test. The Rank-Sum test hypotheses are as follows:

Null hypothesis: The difference in medians of ranked BUCR and BUCD values for cells in which scarring was observed (SCAR=1) versus cells in which scaring was not observed (SCAR=0) is equal to zero. In other words, there is an "equality of medians".

Alternative hypothesis: The difference in medians of ranked BUCR and BUCD values for cells in which scarring was observed (SCAR=1) versus cells in which scarring was not observed (SCAR=0) are greater than the mean values of BUCR and BUCD for SCAR=0. In other words, there is an inequality of medians and the median value of the sample when SCAR=1 is significantly less than the median value for the sample when SCAR=0.

Next, a series of 2x2 contingency table analyses (cross-tabulations) were carried out to test the associations between scaring (SCAR) and extreme use-intensity values under the null hypothesis of statistical independence. To complete this test, two dichotomous variables: BUCRSTD and BUCDSTD were constructed to denote extreme use-intensity values (along travel routes and for destinations). The variables are as follows:

- **BUCRSTD**=1 for cells along the travel route where the BUCR value is greater than one standard deviation from its mean (and BUCRSTD=0 if otherwise); and

- **BUCDSTD**=1 for destination cells where the BUCD value is greater than one standard deviation from its mean (BUCDSTD =0 if otherwise).

The Chi-square test hypotheses are as follows:

Null hypothesis: There is no statistical association (positive or negative) between BUCRSTD or BUCDSTD and SCAR.

Alternative hypothesis: There is a statistical association (positive or negative) between BUCRSTD or BUCDSTD and SCAR.

Results

The null hypothesis of equal medians for the two samples is 'rejected' at the 95% confidence level, in favor of the alternative hypothesis which states that the median or mean "ranked value" of BUCR is higher when SCAR=1. The results suggest that sea grass scarring occurs in cells with lower travel use-intensities based upon the inverse relationship between SCAR and BUCR values. Stated another way, the actual value of BUCR is typically higher for cases where SCAR=0. Note also that the identical distribution assumption holds using the KS test at the 95% confidence level.

Chi-square results are outlined in Table 3. For routes (BUCRSTD) we "reject" the null hypothesis of no statistical association between BUCRSTD and SCAR at the 95% confidence level. For destinations, we "reject" the null hypothesis of no statistical association between BUCDSTD and SCAR at the 95% confidence level.

Conclusions

There is statistical evidence of an inverse relationship between boating use along travel routes (BUCR) and sea grass scarring (SCAR). This suggests that sea grass scarring is more likely to be observed in areas which are less traveled along the travel network. In terms of the boating use values at destinations (BUCD), we find no evidence of a verifiable statistical association between BUCD and scarring. At the chosen level of spatial resolution, sea grass scarring is not found to be associated with high-use destinations.

The chi-square result indicates that while an inverse statistical relationship may exist between BUCR and SCAR in general, there is an association between scarring in areas that are more highly traveled. The opposite is true for high use destinations as there is an inverse correlation between BUCDSTD and SCAR.

While Rank-Sum results show that sea grass scarring is more likely to be observed in less traveled network segments, chi-square results show that sea grass scarring is more likely to be associated with heavily traveled network sections. Taken as a whole, sea grass scarring is statistically associated with the lowest and highest traveled grid cells on the network.

Discussion

BUCR and BUCD values were observed to be higher in cells were sea grass scarring did not occur. This suggests an inverse relationship between use-intensity and sea grass scarring (Table 1). In other words, the results indicate that sea grass scarring occurs in cells with lower travel use-intensities based upon the inverse relationship between SCAR and BUCR values. Stated another way, the actual value of BUCR is typically higher for cases where SCAR=0. Note that the identical distribution assumption holds using the KS test at the 95% confidence level.

The chi-square statistics reveals a weak positive, yet significant correlation between sea grass scarring (SCAR) and high use intensity (as defined by

BUCRSTD). This finding points to an association between sea grass scarring and the heaviest traveled portions of the network (i.e., on routes where travel is equal to or exceeds one standard deviation from the mean).

Table 2. Mann-Whitney U/Wilcoxon Rank Sum Test for Differences in Medians for Two Independent Samples (where SCAR=1 versus SCAR=0)

		Descriptive Statistics		
Variable: BUCR	*count(n)*	*mean*	*s*	*Norm. test decision**
for SCAR=0	1953	0.149891	0.15553	reject normality
for SCAR=1	107	0.107440	0.12572	reject normality
	n=2060			
Variable: BUCD				
for SCAR=0	1198	0.046847	0.05788	reject normality
for SCAR=1	382	0.040841	0.04403	reject normality
	n=1580			

*as based on skewness, kurtosis, ominbus normality test (at 95% confidence level)

Mann-Whitney U/Wilcoxon Rank Sum Test Results
Variable: BUCR

Group variable (sample)	U	Wilcoxon Sum of Ranks	Mean (W)
SCAR=0	131213.5	2039295	2012567
SCAR=1	77757.5	83535.5	110263.5

Standard deviation of W = 5982.365; Number of ties = 153

Alternative Hypothesis	Z-Value**	Prob-value	Decision (one-tail test)
Diff>0	-4.4678	0.00004	Reject null at 95% confidence

Kolmogorov-Smirnov (KS) test for different distributions: D = .004 < D critical (.1350):

Variable: BUCD

Group variable (sample)	U	Wilcoxon Sum of Ranks	Mean (W)
SCAR=0	235586	953787	947019
SCAR=1	222050	295203	301971

Standard deviation of W = 7649.301; Number of ties = 50

Alternative Hypothesis	Z-Value**	Prob-value	Decision (one-tail test)
Diff>0	-0.8848	0.18813	Fail to reject null at 95% confidence

Kolmogorov-Smirnov (KS) test for different distributions: D = .0766 < D critical (.0799):
Fail to reject null hypothesis of identical distributions at 95% confidence level

---

** with correction factor for ties

Table 3. Cross-Tabulation Reports and Chi-square Analyses of Boating Use and Sea Grass Scarring

Original Variable: **BUCR** (use intensity along routes) transformed to BUCRSTD:
where BUCRSTD = 1 if BUCR > 1 standard deviation from its mean
and BUCRSTD = 0 if BUCR is otherwise.

	SCAR		
BUCRSTD	0	1	Total
0	1311	62	1373
1	642	45	687
	1953	107	2060

Chi-square = 3.849 (prob. = .049775)
Test result: Reject null hypothesis of no statistical association between BUCRSTD and SCAR at the 95% confidence level.

Correlation: Phi = .043226 (weak positive, yet significant correlation)

Original Variable: **BUCD** (use intensity along routes)
...transformed to BUCDSTD:

where BUCDSTD = 1 if BUCD > 1 standard deviation from its mean
and BUCDSTD = 0 if BUCD is otherwise.

	SCAR		
BUCDSTD	0	1	Total
0	656	262	918
1	542	120	662
	1198	382	1580

Chi-square = 22.752 (prob. = .000002)
Test result: Reject null hypothesis of no statistical association between BUCDSTD and SCAR at the 95% confidence level.

Correlation: Phi = -0.120001 (weak negative, yet significant correlation)

Boating Use and Manatee Sightings

Manatee sightings were aggregated to one-quarter-square mile grid cells using the Arc/Info INTERSECT overlay process described in Appendix A. The cell-by-cell proportion values for manatee sightings (MANP) were compared to trend surface boat use indices BUCR and BUCD for cells where use is estimated to be above the average. The distributions of BUCR, BUCD and MANP variables were shown to be identical and of equal variance (using a standard variance-ratio, equal-variance test), and non-normal. The non-normal quality of these distributions precluded the use of the Pearson's Product Moment correlation. The Test hypotheses are as follows:

Null Hypothesis: There is no statistical association between the ranked-ordered values of either BUCR or BUCD versus MANP.

Alternative Hypothesis: There is a statistical relationship between the ranked-ordered values of either BUCR or BUCD versus MANP.

<u>Results</u>

The results of the Spearman rank correlation are summarized in Table 4. For routes (BUCR), we 'reject' the null hypothesis of no statistical association between the rank values of MANP and BUCR in favor of the alternative hypothesis, revealing statistical evidence of a weak negative, yet significant correlation at the 95% confidence level. For destination areas (BUCD), we 'fail to reject' the null hypothesis of no statistical association between the rank values of MANP and BUCD at the 95% confidence level.

<u>Conclusion</u>

The results of this regional analysis point to a weak negative, yet significant relationship between manatee sightings (MANP) and boat use (BUCR) along travel routes. There is no statistical evidence of a regional association between manatee sightings and boating destinations (BUCD). However, 'hot-spots' – areas of potential conflict between boating and manatees -- may still persist at specific locations (Figure 4).

<u>Discussion</u>

There is an inverse relationship between manatee sightings and boating use in highly traveled areas - the highest proportion of manatee sightings occur in areas that are less traveled along the network by boaters. Future modeling efforts should incorporate management data themes (e.g., manatee speed zones, marine protected area boundaries, etc.) and manatee flight reconnaissance information (flight paths) into the analysis to assess the degree to which management efforts reduce the likelihood of boats and manatees to share common space.

Table 4. Spearman Rank Correlations of Manatee Sightings (MANP) and Use-Intensity (BUCR and BUCD)

Spearman's r (MANP vs. BUCR) = -0.1297

t = - 5.937 (probability = .00000)
d.f. = 2058

Test Result: Reject null hypothesis of no statistical association between the rank values of MANP and BUCR in favor of the alternative hypothesis, revealing statistical evidence of a weak negative, yet statistically significant correlation (an inverse relationship) at the 95% confidence level.

Spearman's r (MANP vs. BUCD) = 0.0343

t = 1.364 (probability = .08623)
d.f. = 1578

Test Result: Fail to reject the null hypothesis of no statistical association between the rank values of MANP and BUCD at the 95% confidence level.

Boating Use and Manatee Deaths

Manatee death locations were aggregated to one-quarter-square mile grid cells using the Arc/Info INTERSECT overlay process described in Appendix A. The proportion of manatee deaths by cell (MAND) are compared to trend surface estimated boating use along routes (BUCR) and destinations (BUCD). A Spearman rank correlation procedure was chosen as the distributions of values were found to be non-normal, of unequal variance, and non-identical. The test hypotheses are as follows:

Null Hypothesis: There is no statistical relationship between the ranked-ordered values of either BUCR or BUCD and MAND.

Alternative Hypothesis: There is a statistical relationship between the ranked-ordered values of either BUCR or BUCD and MAND.

<u>Results</u>

The estimated Spearman rank correlation coefficient that measures the statistical association between the ranked values of MAND versus BUCR and BUCD, respectively, are shown in Table 5. For BUCD, we 'fail to reject' the null hypothesis of no statistical association between the rank values of MAND and BUCD at the 95% confidence level. For BUCR, we 'reject' the null hypothesis of no statistical association between the rank values of MAND and BUCR in favor of the alternative hypothesis, revealing statistical evidence of a weak positive, yet statistically significant correlation at the 95% confidence level.

Conclusion

The correlation analysis reveals a weak, positive and statistically significant association between the rank-ordered proportion of manatee deaths and boating use along highly traveled routes in the study region. By contrast, there is no evidence of a statistical association between manatee deaths and the boating at high-use destinations (at the 95% confidence level).

Discussion

Although a correlation is found between MAND and BUCR, one cannot assume that a causal relationship exists. Many other factors may be used to explain variations in the proportion of manatee deaths across the region. To explore these possibilities, a multi-variate regression model was developed. The model examined the extent to which variations in MAND (the dependent variable) are explained by a series of independent variables (based upon available data). Given the limited nature of the bi-variate correlation analysis a multi-variate framework was constructed to examine potential factors that might account for variability in manatee deaths. The multi-variate analysis is presented of the following section.

Table 5. Spearman Rank Correlation of Manatee Deaths and Use-Intensity Coefficients of Routes and Destinations

Spearman's r (MAND vs. BUCR) = 0.1185

t = 5.414 (probability = .00000)
d.f. = 2058

Test Result: Reject null hypothesis of no statistical association between the rank values of MAND and BUCR in favor of the alternative hypothesis, revealing statistical evidence of a weak positive, yet statistically significant correlation at the 95% confidence level.

Spearman's r (MAND vs. BUCD) = 0.0402

t = 1.598 (probability = .0550)
d.f. = 1578

Test Result: Fail to reject null hypothesis of no statistical association between the rank values of MAND and BUCD at the 95% confidence level.

Destination Boat use and Manatee Deaths: Multi-Variate Analysis

A regression analysis was carried out to model the proportion of manatee deaths by cell (MAND) along highly traveled routes (UICSTD) in the study area, as a function of a series of interactive explanatory variables; recall, that the Spearman rank correlation analysis found no statistical association between destination use (BUCD) and manatee deaths (MAND). Sample observations were restricted to cells where route traffic is high (i.e., BUCRSTD). The independent variables used in the model include the {X,Y} location coordinates of the geographic centers of one–quarter-mile square grid cells (representing longitude and latitude, respectively). Information on the absolute locations of cells was included to account for possible spatial variability in parameter estimates and the geographic variations in the influences of explanatory variables. A 4^{th} order polynomial expression was used to test all possible interactions between variables (up to an implied exponent of 4). The list of independent variables included BUCR, BUCD, ACC, MANP, BUCRSTD, BUCDSTD, RPAVG, and DPAVG (as described earlier).

A forward-stepwise regression procedure was run to avoid the likelihood of redundancy or multi-collinearity amongst the explanatory variables of the model (with an inclusion/exclusion criteria of .10 and .15, respectively). Only those variables with coefficients that tested significant at the 90% confidence level were retained in the final run. The test hypotheses are as follows:

Null hypothesis: The selected independent variables do not significantly explain spatial variation in manatee deaths.

Alternative hypothesis: At least one independent variable significantly explains spatial variation in manatee deaths.

Results

The results of the stepwise-regression model are summarized in Table 6. We 'reject' the hypothesis that selected independent variables do not significantly account for spatial variation in manatee deaths. Numerous variables were found to be significant in explaining manatee deaths including absolute location (as a function of X,Y coordinates) and locales where highly traveled routes and destinations overlap. These preliminary results indicate a spatial regularity or non-randomness in the manatee death pattern.

Conclusion

The regression model accounted for roughly 29% of the variation in manatee deaths (MAND) within the study region. Although the model contained a limited number of geographic and use-intensity variables several interesting statistical associations can be observed (as implied by the magnitude and sign of estimated coefficients). First, manatee deaths tend to increase as X increases, that is, as one moves east across the study region (a factor that might be explained by the increased availability of freshwater sources in easterly locations. Secondly, and in direct contrast to the results of the rank correlation analysis, manatee deaths are negatively related to use-intensity along routes (see negative coefficients associated with BUCR and implied interactive relationships associated with RPAVG).

Discussion

Unlike the correlation analysis, the regression framework allows for one to measure the influence of one variable on the dependent variable holding the effects of other variables constant. Consider how manatee proportions across various locations also help explain variations in manatee deaths (as represented by the numerous interactive variables comprised of location coordinates and the manatee proportions variable MANP, which are found to be significant in the estimation of the model). Surprisingly, the combined influence of both routes and destination proportions as obtained from the averaging of data from the aerial and mail surveys and as represented by the interaction variable RPAVG*DPAVG (a composite use-intensity index), is shown to be negatively related to the proportion of manatee deaths. This is true, of course, holding all other effects accounted for by this model constant.

The regression model is not without its' shortcomings. The model only accounted for roughly 28% of the variation in manatee deaths within the region. Undoubtedly, the regression model is under-specified, as it excludes many variables that might prove useful in accounting for the remaining unexplained variation. For instance, the presence or absence of speed zones, the width of channels and/or water depth, and vital traffic-shed information such as the predominant types of boating activities and/or boat types found within the locality could be added to the database. In addition, boat-density measures (at proximate launch sites), boating and on-water use restrictions, shed-type (e.g., marina versus canal systems), the attractiveness of cells or destinations within the network or study region, distance to major or nearby network pivots or favored destinations, not to mention boater and demographic profiles could significantly improve the model. The current inability of the regression model to more adequately account for variations in manatee deaths is most likely a byproduct of the limitations of the data. This could be one possible explanation for the error terms testing non-normal (Figure 5).

Table 6. Results of a Forward Stepwise Regression to Explain Variations in Manatee Death

Dependent Variable: MAND

Estimated Regression Equation

Independent Variable	regression coefficient	standard error	t-stat (Ho: B=0)	prob. level	decision Ho: B=0
Intercept	44.09969	20.60434	2.1403	0.032448	Reject Ho
X	-0.3577774	5.293645E-02	-6.7586	0.000000	Reject Ho
Y	0.1173771	4.960952E-02	2.3660	0.018074	Reject Ho
BUCR	-30.56437	6.961354	-4.3906	0.000012	Reject Ho
ACC	4.887598E-02	2.037649E-02	2.3986	0.016545	Reject Ho
X*X	4.324871E-04	7.731438E-05	5.5939	0.000000	Reject Ho
Y2	-1.977379E-04	4.781386E-05	-4.1356	0.000037	Reject Ho
XY	1.676589E-04	3.981449E-05	4.2110	0.000027	Reject Ho
X*X*X	-2.187379E-07	4.12377E-08	-5.3043	0.000000	Reject Ho
Y*Y*Y	9.813972E-08	1.61397E-08	6.0806	0.000000	Reject Ho
Y*Y*X	-9.354905E-08	2.252691E-08	-4.1528	0.000034	Reject Ho
X*MANP	-2.362659E-02	9.097466E-03	-2.5971	0.009471	Reject Ho
Y*MANP	1.733592E-02	6.598463E-03	2.6273	0.008672	Reject Ho
X*Y*MANP	-2.870447E-05	1.063041E-05	-2.7002	0.006987	Reject Ho
X*X*MANP	3.929285E-05	1.456346E-05	2.6980	0.007032	Reject Ho
BUCRSTD*X	3.978043E-04	1.232644E-04	3.2272	0.001270	Reject Ho
BUCRSTD*Y	5.966567E-04	1.200464E-04	4.9702	0.000001	Reject Ho
BUCRSTD*X*Y	-1.461596E-06	2.839301E-07	-5.1477	0.000000	Reject Ho
RPAVG*DPAVG	-25616.06	3908.793	-6.5534	0.000000	Reject Ho
RPAVG*X	0.1647902	5.196768E-02	3.1710	0.001542	Reject Ho
RPAVG*X*Y	4.907444E-04	1.250287E-04	3.9251	0.000090	Reject Ho
RPAVG*DPAVG*X	20.20111	3.130727	6.4525	0.000000	Reject Ho
RPAVG*DPAVG*Y	16.66408	3.104581	5.3676	0.000000	Reject Ho
DPAVG*X	0.2296026	6.469677E-02	3.5489	0.000396	Reject Ho
DPAVG*Y	-0.1629043	4.498962E-02	-3.6209	0.000301	Reject Ho
RPAVG*Y*Y	-5.106652E-05	2.171649E-05	-2.3515	0.018792	Reject Ho
RPAVG*X*X*Y	-5.365332E-07	1.248566E-07	-4.2972	0.000018	Reject Ho

R-Square = 0.289116

Analysis of Variance Section

Source	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level
Intercept	1	2.378641	2.378641		
Model	26	3.942937	0.1516514	31.8008	0.000000
Error	2033	9.694953	4.768792E-03		
Total(Adjusted)	2059	13.63789	6.62355E-03		

Root Mean Square Error 6.905644E-02
Mean of Dependent 3.398059E-02
Adjusted R-Square = .281

Future Research Opportunities

Future research could expand on the frameworks and methods employed here to better understand the complex spatial patterns and relationships such as those explored in this preliminary study. A more thorough analysis is worth pursuing. This, however, would require the incorporation of additional management themes and more detailed information on manatees and accidents. Two promising areas to explore are (1) a ranking of conflict potential by small area based upon a multi-variate composite indexing scheme, and (2) the incorporation of management data themes into the association analyses to evaluate current management practices.

Conflict Potential (Composite Indexing)

The weak yet significant associations between boating use and environmental variables point to the need for a more comprehensive and in-depth analysis. This analysis would identify hot-spot composite zones (based on the strength of statistical associations) where conflict potential among a multitude of human and environmental uses are at their highest. A composite conflict potential index could be constructed by a multi-overlay analysis to target specific geographic areas where human use is high in environmentally sensitive or degraded areas. For example, figure 6 represents a composite of 'greater than average' incidence of manatee sightings, manatee deaths, boat route use (BUCR) and destinations (BUCD). Composite indices can be used to identify localities where a current management initiative may be falling short of its intended objective (i.e., not targeting the areas of greatest need), or where a particular management intervention may be needed (i.e., no wake zone, manatee speed zone, channel marker).

Incorporation of Management Data Themes

The association analyses presented in this report focused on human-environmental relations without examining the management component of the equation. The preliminary findings revealed significant (albeit weak in some cases) relationships between recreational boating use and selected environmental and occurrence variables. Analyses such as these have direct implications for an assessment of current management practices. For example, the data analysis revealed an inverse relationship between boating traffic and manatee sightings. This suggests that management practices such as the placement of speed zones, no wake zones, marine protected areas, or channel markers, etc., may be effectively channeling traffic from these environmentally sensitive areas. Further studies might attempt to differentiate between the relative effectiveness of specific management alternatives by type and/or by location.

References

Sidman, C., and Flamm, R. 2001. A Survey of Methods for Characterizing Recreational Boating in Charlotte Harbor, Florida. Florida Sea Grant Technical Paper 109. Florida Sea Grant, University of Florida. Gainesville, Florida.

Sidman C, and Fik, T. 2002. A Regression-Based Approach to Estimate Preferred Recreational Boating Destinations in Charlotte Harbor, Florida. Final Report submitted to the Florida Marine Research Institute. St. Petersburg, Florida.

Gorzelaney, J,F., 1998. Evaluation of Boat Traffic Patterns and Boater Compliance in Lee County, Florida. Final Report submitted to the Florida DEP Bureau of Protected Species Management. Tallahassee, Florida.

Figure 4. Overlap of Greater than Average Manatee Sightings and Boating Destinations.

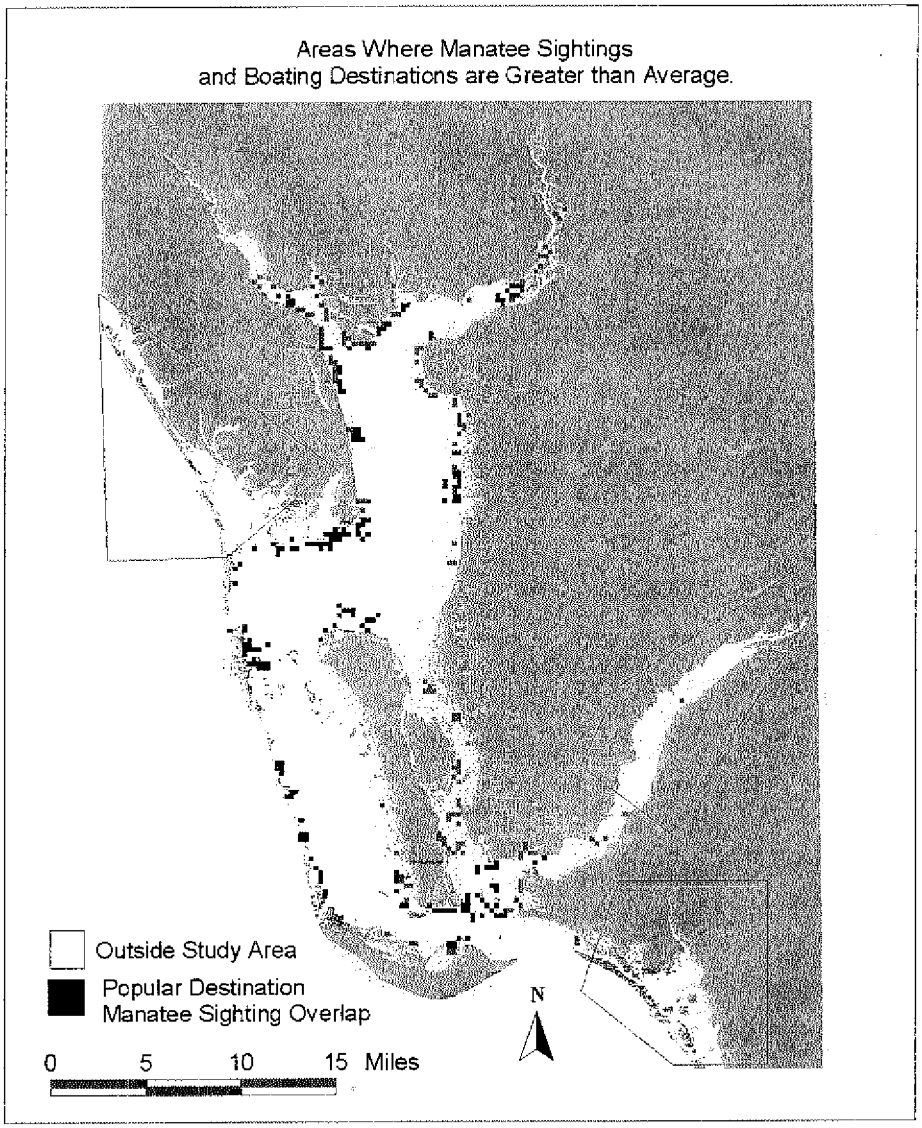
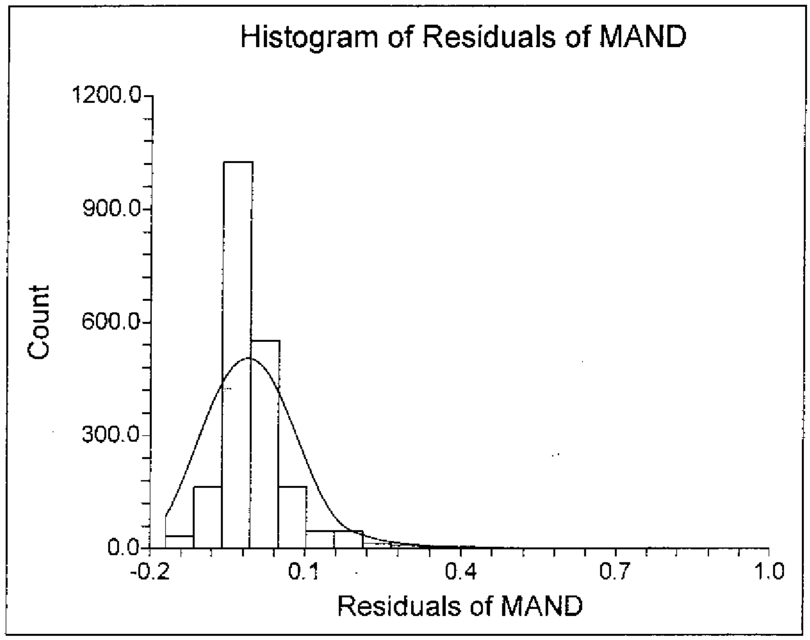


Figure 5. Distribution of Error Terms from Regression Model and Tests of Normality.

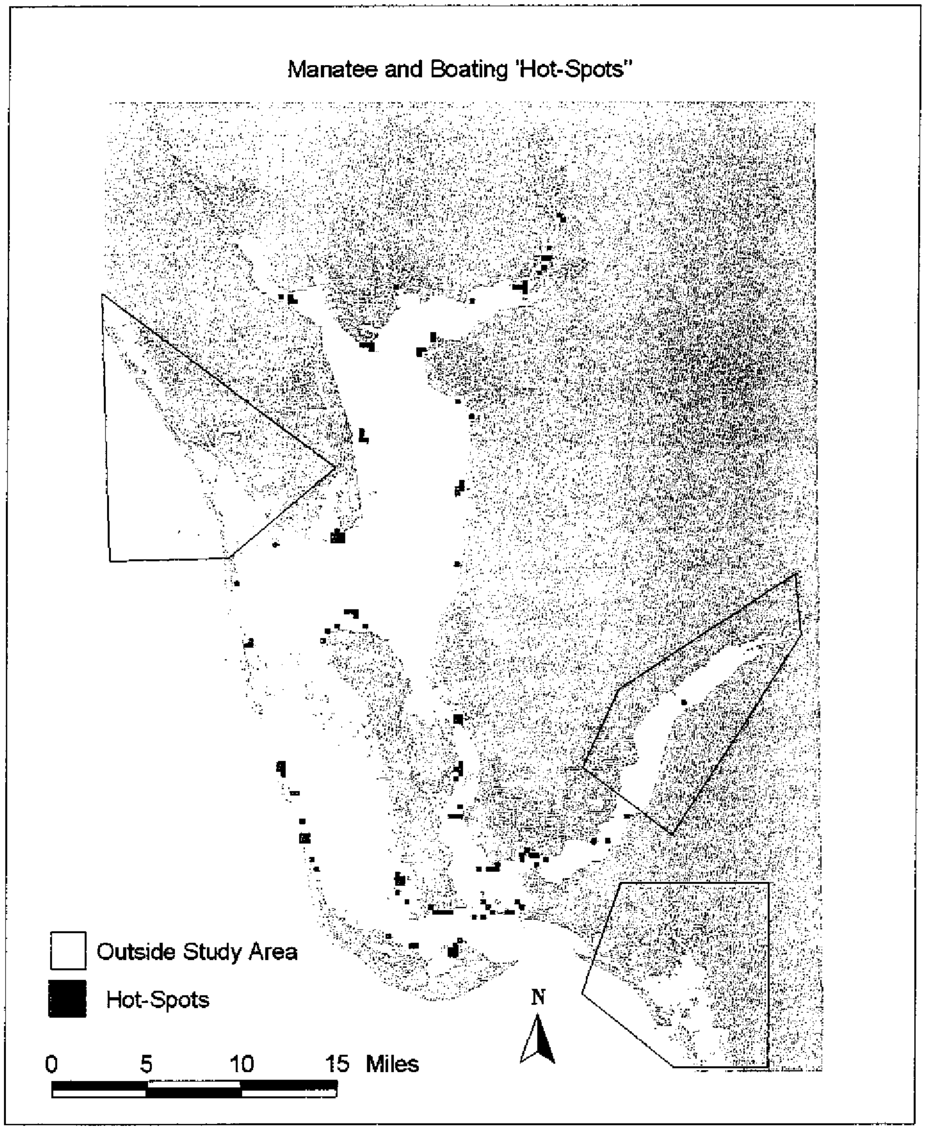


Normality test section:

Shapiro-Wilk statistic = .6356 (probability .0000);
Reject Ho: Normality of error terms at 95% confidence level.

KS test statistic = .2728 > KS critical (.0220) at .05 significance level;
Reject Ho: Normality of error terms at 95% confidence level.

Figure 6. Manatee-Boating Hot-Spots.



Appendix A

ARC/INFO INTERSECT Overlay Analysis

Assumptions:

Route coverage name = Routecov
Grid coverage name = Gridcov
Output coverage name = Outcov (from INTERSECT)
Frequency output filename = Freqout

The following GIS operations were performed.

1. Run the INTERSECT command.

    ARC: INTERSECT Routecov Gridcov Outcov line

2. Run the FREQUENCY command on Outcov.aat

    The frequency item is the internal id item for the grid coverage (i.e. Gridcov#). No summary item is necessary.

    ARC: FREQUENCY Outcov.aat Freqout

    Enter Frequency item names (type END or a blank line when done):
    ==
    Enter the 1st item: Gridcov#
    Enter the 2nd item: END

    Enter Summary item names (type END or a blank line when done):
    ____
    Enter the 1st item: END

3. Run the JOINITEM command

    ARC: JOINITEM Gridcov.pat Freqout Gridcov.pat Gridcov#

    The output Gridcov contains an item named FREQUENCY that identifies the number of routes that cross each ¼ mile grid cell.