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

FLSGP-T-02-004 C2

Blueways Recreational Boating Characterization for Charlotte Harbor, Florida

Enhancements to a Regression-Based Approach to Estimate Preferred Recreational Boating Destinations in Charlotte Harbor, Florida

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

Introduction and Objectives

This report presents the methods and results of network enhancements to a regression-based framework that estimates preferred recreational boating destinations in Charlotte Harbor, Florida. The methods and findings of the initial modeling effort are presented in a report submitted to the FMRI entitled "A Regression-Based Approach to Estimate Preferred Recreational Boating Destinations in Charlotte Harbor, Florida" (Sidman and Fik, 2002).

The initial modeling effort highlighted the importance of network variables in estimating the preferred destinations of recreational boaters. The primary goal of this study was to determine if enhancements to the network and pivot structure could generate better estimations of preferred destinations than those obtained from the initial modeling effort. A secondary goal was to determine if an application of the model to small areas -- specific origin regions (SOR's) -- would improve the accuracy of estimations. The scope of work for model enhancements included the following objectives:

1) Extend the network structure and pivots to include, where appropriate, the following areas: Charlotte Harbor proper, Pine Island Sound, Matlacha Pass, nearshore coastal waters, Lemon Bay, and the Caloosahatchee River.

2) Test for variability in latitude (x) and longitude (y) estimations for specific origin regions (SOR's).

In addition to the stated objectives, preferred destinations also were estimated for a condensed version of the original, non-network enhanced, data set that was limited to one preferred destination per survey respondent -- the preferred destination calculated to be the farthest from the boaters' launch origin. The assumption was that eliminating multiple preferred destinations would reduce estimation error.

Database Development

The initial modeling effort included, as independent variables, the latitude and longitude of up to three pivots along each travel route from the location of network-travel route intersection. The enhanced model evaluated travel routes through 6 pivots along an expanded network that included 28 pivots and additional segments for Charlotte Harbor proper, Pine Island Sound, Matlacha Pass, nearshore coastal waters, Lemon Bay, and the Caloosahatchee River (Figure 1).

The expanded network and pivot structure was derived from prominent travel corridors that emerged from an analysis of travel route counts by ¼-square-mile grid cell. The ArcInfo INTERSECT overlay function was used to determine the number of travel routes that passed through each grid cell. Each grid cell was assigned a value calculated as the proportion of travel routes from the total that intersect a specific cell. The structure of the enhanced network was validated by data obtained by aerial videography (Sidman and Flamm, 2001), which identifies similar boat traffic flow patterns.

Analysis

Several variants of the initial regression framework were used to establish the best estimations of preferred destinations. Each model variant was based on information obtained from 192 mail surveys distributed to a random sample of boaters who registered their vessels in Lee and Charlotte Counties. Many boaters identified multiple preferred destinations associated with their boating trip. The network-enhanced models (region and SOR specific) were based on the entire sample of preferred destinations associated with boaters' trips (n = 361). The condensed models (region and SOR specific) utilized the original network and pivot structure described in Sidman and Fik, 2002, and were based on the destination determined to be the farthest from the launch origin (n = 186) – six of the 192 mail surveys used did not contain a preferred destination.

Individual equations estimate destination longitude (*xd*) and destination latitude (*yd*), using a forward step-wise procedure with a $p<.10$ level of inclusion/retention. Estimations for latitude were included as an independent variable in the longitude model. The basic models are as follows:

$yd_i = f$ (independent variables); and

$xd_i = f$ (independent variables including estimated yd).

<u>Enhanced Network Model</u>

The forward stepwise regression procedure was applied to a data set that incorporated coordinates of up to 6 pivots along each travel route. The additional variables are described in the initial report. The enhanced network model was first run on a dataset that included all of the preferred destinations (n = 386) identified by mail survey respondents. The dataset was then disaggregated so that individual model runs could be restricted to observations that comprise each of the four identified SOR's (Englewood and Placida Harbor, Punta Gorda Isles and Port Charlotte, Ft. Myers and Cape Coral, and Pine Island) – see Figure 2. Recall, a K-means cluster analysis was used to identify SOR's on the basis of launch origin proximity.

<u>Condensed Data Model</u>

The forward stepwise regression framework was then applied to a condensed version of the original data set (no network enhancements) to account for potential error related to the prediction of multiple destinations from a single origin. Condensed data model estimations are generated on the basis of one preferred destination per survey respondent -- the preferred destination calculated to be the farthest from the boaters' launch origin. In this way, independent variables (i.e. origin latitude, origin longitude, activity, boat type, draft, distance traveled, hours spent on water, etc.) are not repeated for multiple destinations, as was the case in both the initial and network-enhanced models. For comparison purposes the forward stepwise regression procedure was applied to the entire sub-sample of "farthest" preferred destinations (n = 186), and then restricted to observations that comprise each of the four SOR's. A list and description of the variables is provided in Table 1.

Table 1. Condensed Model Variables

Variable	Description
NPY	Y-coordinate of Nearest Pivot
P1X	X-coordinate of Pivot 1 (second-nearest pivot)
P1Y	Y-coordinate of Pivot 1 (second-nearest pivot)
P2X	X-coordinate of Pivot 2 (third-nearest pivot)
P2Y	Y-coordinate of Pivot 2 (third-nearest pivot)
P1Y*X	P1Y*X-coordinate of origin
P2Y*X	P2Y*X-coordinate of origin
P2Y*Y	P2Y*Y-coordinate of origin
X*X	X-coordinate of origin squared
X*Y	X-coordinate of origin * Y-coordinate of origin
WDIS	On-water travel distance
WDISX	WDIS*X-coordinate of origin
WDISY	WDIS*Y-coordinate of origin
ANF	Anchoring as a favorite activity
ANY	Anchoring as a favorite activity * Y-coordinate of origin
FX	Fishing as a favorite activity * X-coordinate of origin
NAF	Nature-viewing as a favorite activity
SA	Sailing as an activity
SAF	Sailing as a favorite activity
SAX	Sailing as an activity * X-coordinate of origin
SAY	Sailing as an activity * Y-coordinate of origin
CRX	Cruising as a favorite activity * X-coordinate of origin
DFF	Deep sea fishing as favorite activity
DFX	Deep sea fishing as an activity * X-coordinate of origin
DIVEX	Diving as an activity * X-coordinate of origin
DIVEY	Diving as an activity * Y-coordinate of origin
WSKIX	Water skiing as favorite activity * X-coordinate of origin
WSKIY	Water skiing as favorite activity * Y-coordinate of origin
HRSX	Hours spent on water * X-coordinate of origin
DHX	Distance to network * hours on water * X-coordinate of origin
DN	Distance to the network from launch site
DNY	Distance to network * Y-coordinate of origin
DRFT	Draft of boat
DAYX	Day-tripper dummy * X-coordinate
DH	Distance from home to network

Results

Enhanced Network Model

The enhanced network models (Table 2) generated an adjusted r-square of .513 for longitude (*xd*) and .825 for latitude (*yd*). These values can be compared to the results from the initial models (Table 3), which produced an adjusted r-square of .643 for longitude (*ad*) and .855 for latitude (*yd*).

Table 2. Enhanced network model summary.

Study Region	**Sample Size (n)**	**Adjusted R-squares** *xd*	*yd*
Entire*	361	.513	.825
Sor1	184	.450	.480
Sor2	38	.714	.539
Sor3	134	.475	.528
Sor4	42	.793	.938

* Sample size varies because some outliers were eliminated from each run.

Table 3. Initial model summary without network enhancements.

Study Region	**Sample Size (n)**	**Adjusted R-squares** *xd*	*yd*
Entire*	384	.643	.855
Sor1	184	.490	.495
Sor2	38	.493	.696
Sor3	134	.519	.605
Sor4	42	.829	.901

* Sample size varies because some outliers were eliminated from each run.

Condensed Model

Goodness of fit measures for longitude *xd* and latitude *yd* models that use the non-enhanced network and pivot structure, and estimate the "most-distant" preferred boating destinations are presented in Table 4. A list and of variables found to be significant at the 95% confidence level is presented in Table 5. The distance (in miles) between observed and estimated preferred destinations is presented in Table 6.

Table 4. Condensed model summary.

Study Region	Sample Size (n)	Adjusted R-squares *xd*	*yd*
Entire*	186	.698	.842
Sor1	92	.754	.701
Sor2	38	.348	.498
Sor3	51	.617	.853
Sor4	20	.931	.928

* Sample size varies because some outliers were eliminated from each run.

Table 5. Condensed model: Listing of variables in descending order of significance (see Table 1 for variable definitions).

Entire Study Region	
X	P2X, P1Y*X, P1X; WDISX, WDIS, SAY, HRSX, SAX, X*X
Y	P2Y, P1Y, SAY, DFX, DIVEF
Sor 1 – specific-origin region 1	
X	P2X, WDIS, P1X*Y, DFF
Y	P2Y, DFF, NPY, WDIS, DIVEY, CRX, SAF
Sor 2 – specific-origin region 2	
X	ANY, NHX, XH, NAF, CRX, WSKIY
Y	P2Y*Y, WDISY, P1Y*X, DNY, P2Y*X, WSKIY, DAYX, DH, FX
Sor 3 – specific-origin region 3	
X	P2X, P1X, SA, XY
Y	P2Y*Y, P1Y, HRSX, DIVEY, SAY, SAF, DFF
Sor 4 – specific-origin region 4	
X	P2X, SAX, P2Y, DIVEX, NPY, DFF
Y	P2Y, WSKI, DN, ANF, DRFT

Discussion

Enhanced Network Model

A regional application of the regression framework did not benefit from network enhancements. Increasing the number of network segments and pivots generally lowered 'goodness of fit' measures (adjusted r-square values) when the model was applied to the entire study region. Network enhancements did improve the adjusted r-square values of either *xd* or *yd* when the models were in applied to certain SOR's, but such improvements are probably not significant (i.e., .938 versus .901 for *yd* in SOR 4; Tables 2 and 3).

Network enhancements to the model included calculating up to 6 pivots for each travel route, along the network. In the expanded network, many of the pivots were peripheral, located along less traveled corridors. Such secondary or tertiary pivots did not prove to be important in estimating preferred destinations because pivots with the highest accessibility to preferred destinations tended to be centrally located, not lying beyond 3 or 4 pivots from the place where travel routes intersected with the network. Repeated measures (redundant pivot information for multiple travel routes), therefore, may account for lower 'goodness of fit' estimations from models that incorporated the expanded network structure.

Table 6. Condensed model: Distance between observed and estimated destinations.

Study Region	Sample Size (n)	Average Error*	St. Dev*.
Sor1	92	3.62	2.35
Sor2	38	2.49	1.55
Sor3	51	3.96	1.92
Sor4	20	1.80	.94

*Average error and St. deviation estimates are in miles

Condensed Model

The condensed data model generated the best results, especially when applied to individual SOR's. The results show that the on-water travel network is the predominant factor in determining (or constraining) movement within the study region (Table 5). It is evident that the geographic location of network pivot 2 (i.e., the third- pivot from where the travel route intersects with the network) is the most important factor in explaining the preferred destinations of boaters in the survey. Of secondary importance is the location of pivot 1 (the second pivot along the network from the location of route-network intersection), and its location along the Y-axis. These results are consistent with those of the initial modeling effort (Sidman and Fik, 2002). It is interesting to note that the second (pivot 1), and third (pivot 2) nearest network pivots are positioned at points that are, in a relative sense, at or near the geographic center of the network. The locations of these pivots are at a distance of approximately one-half the non-enhanced network's "diameter" (as measured in terms of the number of linkages that separate the most-distant points in the network).

The number of hours available per on-water excursion was found to be a significant determinant in the choice of a preferred destination. There is a direct relationship between the number of hours spent on the water and the predicted X-axis value (for the model run on the entire study region). This implies that boaters with more time to spend on the water do not necessarily venture farther west. In fact, the results show that boaters leaving from more easterly locations with more hours to spend on the water actually prefer locations that are less distant from network pivots and somewhat closer to their points of origin as measured along the X-axis. One plausible explanation for this is that boaters tend to spend significantly more time at locations in which they are more familiar with -- destinations nearer to their launch sites. Consequently, "familiarity" with on-water locations closer to the point of origin may actually reduce the propensity to travel over greater distances, despite a large time budget.

When the model was run for SOR's, activity variables also were significant. For instance, preferred destinations of boaters identifying sailing as an activity tend to be dependent upon the location of the launch site. In general, sailing tends to be associated with preferred destinations that are located further west of boaters not engaging in this activity. This is especially true for Charlotte County boaters. Also, boaters from SOR 1 (Ft. Myers and Cape Coral) that identified sailing as a favorite activity tended to prefer destinations that were further north of those that did not identify sailing as a favorite activity. Overall, boaters in SOR 3 (Port Charlotte and Punta Gorda) who listed sailing as a favorite activity tended to prefer locations that were typically south of boaters that did not list sailing as a favorite activity, with the exception of those boaters located in the northern confines of this origin-specific region. In the case of SOR 4 (Englewood and Placida), sailing tended to be an activity that was associated with preferred destinations that were slightly east of boaters engaging in other activities. This makes sense given that SOR 4 is comprised of the barrier islands that make up the western boarder of the study area. Other activities that were shown to play a role in helping to explain preferred destinations of boaters by origin-specific region were deep-sea fishing (in the case of SORs 1,3, and 4), cruising (SORs 1 and 2), water-skiing (SORs 2 and 4), anchoring (SORs 2 and 4), and nature-viewing (SOR 2).

In sum, the identification of prominent travel corridors and network nodes (pivots) is sufficient to estimate preferred destinations for the Charlotte Harbor region. Network enhancements that included secondary network segments and nodes lowered 'goodness of fit' estimates when applied regionally, and did not significantly improve estimations when applied to specific SOR's. However, improvements were seen when the model was applied to a non-network enhanced (original network and pivot structure) condensed data set that contained only the location of the farthest destination. The 'condensed' data models generated even better predictions when applied to specific SOR's (Table 6). For example, when applied to SOR's, the condensed data models generated an average aggregate distance of 2.96 miles, with a standard deviation of 1.69 miles between observed and estimated destinations. This translates to a mapping resolution of 4.65 square miles; 2.35 miles more resolute than the results of the initial modeling effort described in Sidman and Fik, 2002.

Conclusions

The findings of this study reinforce the importance of the effects of spatial structure --the organizational layout and connective properties of the primary travel network and it's geographic foundation. Larger geographic and network forces are found to be predominant in estimating the preferred destinations of boaters (e.g., network structure, pivots, distance, and location-based variables). Behavioral factors are found to be secondary to the overarching influence of network and geographic considerations. Nevertheless, behavioral factors such as activity preference are shown to be significant in determining preferred destinations when the model is applied to specific-origin regions (SOR's).

Hence, modeling efforts that seek to estimate boater destinations in this region should employ both network and distance-based variables in conjunction with use or activity variables. The predictive models should be flexible in design and incorporate both regional and traffic shed-specific information. The models should be sensitive to the localized influence of the SOR's and the intra- and inter-regional variability in activities and behavior while accounting for the overarching impact of the geography of the travel network.

The inclusion of geographic and network considerations offers a fruitful alternative to models which rely solely on attitudinal or behavioral characteristics as the prime determinants of the choice destinations of boaters (Murphy, 1975). In addition, gravity-type variables also may prove extremely useful in helping to improve the explanatory power of the model. For example, further enhancements to the model would include gravity-type variables such as measures of the relative attraction (size, extent, and drawing power of various destination locales) and competitive nature of destinations, as well as, measures of accessibility and intervening opportunities (Fik, Mulligan, and Amey, 1993; Fik and Mulligan, 1998). Ultimately, a multi-faceted approach that incorporates, geographic, network, attitudinal, and behavioral information may offer the greatest potential for estimating the use-potential of boating destinations.

References

Fik, T., Mulligan G., and Amey, R. 1993. "Labor Migration Between Hierarchically Competing and Intervening Origins and Destinations." *Environment and Planning A*, v24; p1271-1290

Fik. T., and Mulligan, G. 1998. "Functional Form and Spatial Interaction Models." *Environment and Planning A*. v30; p1497-1507.

Murphy, P. 1975. " The Role of Attitude in the Choice Decisions of Recreational Boaters". *Journal of Leisure Research, v7;n3: p216-224.*

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

Figure 1. Travel Corridors and Derived Expanded Network

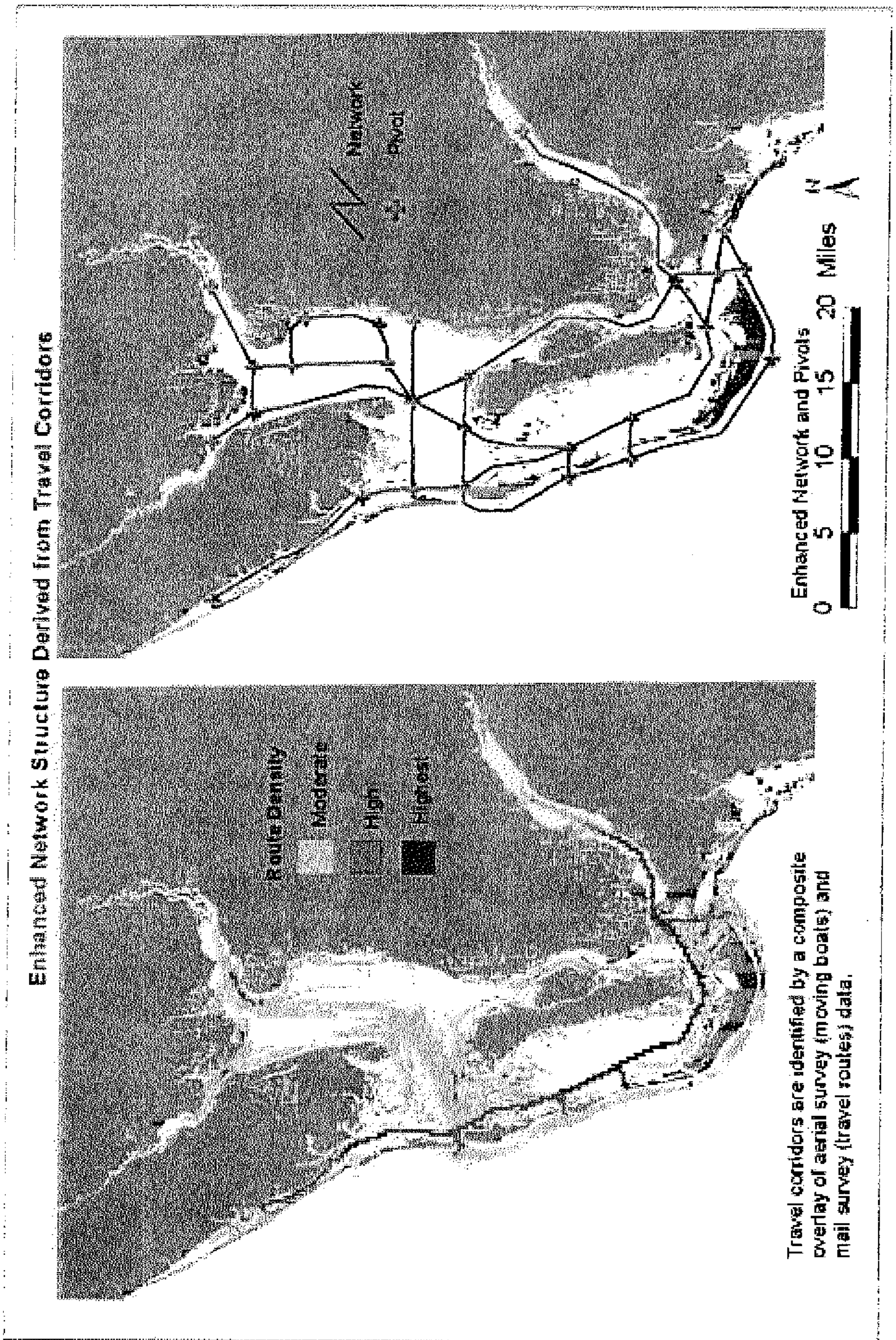

Figure 2. Specific Origin Regions