An Assessment of Ecotrust's Relative Importance Indicators: Comparisons with Logbook Data for the Market Squid Fishery

By

James Wilen
Joshua Abbott

Report submitted to the California Marine Life Protection Act Initiative
In partial fulfillment of Contract #2006-0014M

6/8/2006

An Assessment of Ecotrust's Relative Importance Indicators: Comparisons with Logbook Data for the Market Squid Fishery

This report assesses the methods used by Ecotrust to develop "importance indices" for fisheries off the Central Coast of California. The survey methods and the indices that were constructed by Ecotrust are summarized in the report by entitled "Commercial fishing grounds and their relative importance off the Central Coast of California", Report to the California Marine Life Protection Act Initiative, Contract #2005-0067M, submitted 20 April, 2006. The survey involved interviewing a non-random sample of fishermen who were responsible for disproportionately larger shares of each fishery's harvests. Fishermen were asked to identify areas that were most important to them and indicated these using an assignment of 100 points or "pennies". These individual responses were then aggregated into group-wide indices, giving higher weights to areas identified by multiple fishermen.

The assessment reported here is designed to validate, to the extent possible with a limited time and money budget, whether the importance indices elicited by surveys are related in any meaningful way to actual behavior. To do this, the squid fishery was selected as a test case, mainly because it has logbook records that are believed to be relatively accurate. Our methods and findings are detailed below in four sections.

I. Introduction

The purpose of this report is to utilize selectively available fisheries data in order to assess the reliability of the relative importance indicators (RICs) described in the Ecotrust report and to draw some inferences about their usefulness for an economic impact analysis of a network of marine protected areas off the central coast. Our comments on the methodology employed in deriving these indicators were included in a previous report and so will not be reiterated here except to inform the results of the empirical analysis.

In assessing the reliability of the RICs and their usefulness as indicators of economic impact to the fishing community there are primarily four important questions that need to be investigated. Firstly, do difference in RICs across fishing sites correlate consistently with differences in the observed intensity of fishing effort by interviewed fishermen? Secondly, how robust is this relationship to changes in biological and economic conditions across successive years? Thirdly, what is the relationship (if any) between the importance indicator for a site and its productivity in terms of expected catch? Finally, how well do the RICs serve in predicting the distribution of fishing effort by the fishermen that were not included in the interview process?

The next section describes the data sources that were made available to us for this study and enumerates the steps taken to prepare it for analysis. The third section addresses each of the aforementioned questions in turn, and the fourth section summarizes our findings.

II. The Data

The empirical analysis within this report is based entirely on data for the market squid fishery. Ideally the results we obtain may have portability with respect to other fisheries, and we will comment on the broader implications of our findings whenever possible, but the lack of reliable spatial effort data for other fisheries precludes any attempt at overall validation. Indeed, the choice of the market squid fishery for this analysis is largely based on the availability of vessel logbook data, although the widespread spatial nature of the squid fishery and its economic importance in the study region make it a worthy case study regardless.

The vessel logbook data employed in this analysis contains a set-by-set record of the fishing activities of both the interviewed and non-interviewed fishermen from the years 2000 to 2005. This period precedes and includes the survey dates in 2005. This data is recorded directly by fishermen and contains information on the spatial location of each set, the date on which it occurred, the depth of the water, and the catch associated with that set. The interviewed and non-interviewed fishermen's records were provided in such a way that we can separate the two groups from one another.

The second data source is a raster dataset of the fishing "grounds" (as defined in the Ecotrust report) for the interviewed fishermen. This dataset reports the relative importance value for each 100m square grid cell. The derivation of these values is described at length in the Ecotrust report but is essentially as follows. First, each sampled fisherman draws his/her fishing sites on a map and assigns some fraction of 100 pennies to each site. Each of these sites is then broken down into 100m squares with each grid receiving an equal share of its site's allocation of pennies. These values are then added across all fishermen to provide indicators of the relative importance of each 100m grid to the fishery (albeit with a weighting scheme that downgrades the contribution of fishermen with smaller than average fishing grounds to the importance index). The interview process resulted in separate raster files for each individual fisherman – information that would have been, ideally, used to correlate survey results with that fisherman's vessel logbook data. However, for confidentiality reasons, we were not allowed access to individual raster files of importance indices but instead were constrained to only use the aggregated indices for the whole fleet sample in our analysis.

In order to constrain the study to the Central Coast region defined in the Ecotrust report, we eliminated all sets conducted beyond the northern or southern extent of the fishing grounds as defined in the raster dataset (bounded roughly at 34.42 and 37.29 decimal degrees latitude). However, after analyzing the location of the remaining sets within ArcGIS, it became apparent that there were a substantial number of recorded sets in logbooks with locations that were definitely unrealistic (i.e. they appeared to occur on land or well out at sea) given the nearshore nature of the fishery. In many of these situations the problem seemed to lie in imprecise recording of the haul location by fishermen (and in particular, rounding to the nearest minute) although a small number of precisely recorded sets were obviously incorrect as well. Obvious spatial outliers were dropped from the data, and the resulting distribution of points can be seen for the interviewed and non-interviewed groups in Figures 1 and 2. (See the Appendix for a more detailed discussion of data accuracy issues.)

All told, there are nine fishermen incorporated in logbook database of interviewed fishermen. This is considerably smaller than the 16 fishermen that were interviewed in the original survey and that contribute to the RIC raster. According to CDFG personnel, five of the interviewed fishermen do not appear at all in the logbook database. Four of these either worked on a seiner (but were not the fishermen of record for the logbook) or ran a light boat (for which logbooks are not available). One of these four fishermen no longer fishes, and another fisherman appears to have never fished squid at all. The remaining two excluded fishermen were dropped due to an apparent lack of participation in the central Californian fishery. The "snowball" approach that was used to select squid fishermen for the survey may be partially to blame for this less than ideal sample for the purposes of matching importance indices with logbook data. More careful sample selection methods were employed for most of the other fisheries, so hopefully the samples are more representative of bonafide fishermen for those associated fisheries

Table 1 shows the distribution of sets across interviewed fishermen and years in the trimmed dataset. It is immediately apparent that there are marked differences in the levels of participation across fishermen as well as substantial variation in total effort from year to year. It is important to note that the squid fishery extends far to the south of our study area with the majority of sets occurring in the southern California region, so zeros in our dataset may not be indicative of entry/exit behavior to the broader fishery.[1]

III. Empirical Analysis

This section addresses each of the research questions presented in the introduction in the order they were posed. In each case we utilize a variety of methods to make our point and try also in each case to proceed from relatively simple graphical/statistical methods to those involving more rigorous and complex statistical tools.

Question 1: Do difference in RICs across fishing sites correlate consistently with differences in the observed intensity of fishing effort by interviewed fishermen?

For the relative importance indicators to serve as useful measures of the importance of particular sights within the fishery, we would like to observe a consistent and strong positive association between the numbers of sets (or, more correctly, the proportion of total sets) fished at a site and the reported value of the site. To address this question empirically, however, requires that we define what is meant by a "site". We could utilize the raster grid itself, but these grids are so small and the degree of uncertainty in assigning a set to a particular grid is so large that this level of spatial disaggregation is likely too susceptible to noise to be of much use. Instead, we have chosen to revert to the

[1] A similar breakdown for non-interviewed fishermen is not possible due to the lack of fisherman identifiers in our dataset. However, the pattern of total effort across years is quite similar, although sets by non-interviewed fishermen typically outnumber those of their interviewed counterparts by a factor of three to five.

underlying motivation for the raster grid itself to form polygon aggregations of the raster cells. The raster grid is essentially a discretized version of an overlay of sixteen distinct polygon layers; as a result there are numerous neighboring grids with identical importance values. We reduce the dimension of the problem by grouping these identical cells into a total of 2103 polygons (where equality of values is measured to the third decimal place) and then counting the number of sets that occurred over the entire six year sample period. Finally, we divide this count by the area of the polygon (in square kilometers) so as to normalize the measure and to provide a measure of the "intensity" of effort at the site.[2] Figure 3 gives an example of the raster to polygon conversion.

Figure 4 summarizes the distribution of importance values in the fishing grounds. Clearly the majority of the grounds are accorded a value very close to zero. There is then a relatively smooth falloff (notwithstanding the spike at 0.7) with a few very small "hotspots" at RIC values greater than four. Figure 5 is a scatterplot showing the relationship between the observed intensity of effort and the reported value indicator for the polygon. There does appear to be some weak positive association between value and set intensity. As would seem reasonable, there is precious little fishing within sites with very low values, this despite their dominance of the fishing grounds in terms of area. The proportion of sites clustered around zero intensity also declines as the value index increases.

There are elements of this graph that raise serious questions, however. First, the explanatory power of the regression line is low with an R-squared of just over 5%.[3] Secondly, there are obvious patterns in the scatterplot which seems to suggest that something other than the value index is behind the observed location of sets. This can be seen in the fact that sites with a RIC of around 2 are fished less overall than many sites with lower values. Also, sites extremely high RIC values appear almost completely unfished. A third problem is evident in the increasing scatter of the points with increasing index values. This uneven variance (heteroskedasticity) need not be a problem in and of itself but it could be indicative of a significant omitted factor driving set intensity.

To investigate this question more closely, we need a statistical model. One relatively simple and common method to model the number of events (sets) in a geographic area is Poisson regression. Under this model the mean of the distribution is specified as an exponential function of the regressors and the parameters are then estimated by maximum likelihood. For instance, if we wish to estimate the average annual intensity of sets per unit area for site *i* at year *t* as a function of their RIC value, we could specify a model of expected set intensity as follows.

$$E(count/km^2 \mid RIC_i)_{it} = \alpha_t \cdot \exp(\beta_0 + \beta_1 RIC_i) \quad (1)$$

[2] It should be noted that 117 sets from the interviewed and 261 sets from the non-interviewed data (about 8% and 7% of each trimmed sample, respectively) fell a little to the east or west of the fishing grounds defined by the raster data but otherwise seemed reasonable. These sets are not included in the analysis since they are not part of the defined grounds (i.e. they have zero reported value). Whether this speaks to measurement error or underestimation of the scope of the fishing grounds by the survey instrument is not clear.

[3] Higher order polynomial regressions did little to improve the fit and were quite sensitive to outliers.

The variable α_t is a measure of the total intensity of fishing in year *t* – for our purposes it is defined as the number of sets by interviewed fishermen in *t* divided by the number executed in the base year of 2000 – and it serves to shift the entire distribution of sets across space according to the level of effort in a year while preserving the relative intensities between sites. Modeling the intensity of fishing in this way has a number of attractive properties. Predictions from the model are bounded from below by zero (a property not shared by linear regression) and the maximum likelihood estimates are valid under very minimal assumptions (Wooldridge, 1997).

Table 2 shows the estimated parameters and standard errors for a variety of model specifications. Model 1 in this table is equivalent to that expressed in (1) but with the natural logarithm of the RIC in place of its level. (Using the natural log consistently improved the fit of the models and helped to prevent excessively high predictions for high RIC values.) The parameter on ln(RIC) can be interpreted as the percentage change in annual sets per square kilometer for a percentage change in its relative importance value. From Model 1 we might expect a one percent change in a site's RIC value to be associated with an increase in annual visitation of 1.53%. The sign and statistical significance of this estimate seems to support the use of the RIC as a predictive tool for the intensity of effort at a particular fishing site. However, the apparent lack of fit of this model (as indicated by the small *R*-squared value) is troubling.[4]

One very important factor determining where fishermen fish, particularly when fishermen engage in short trips, is the distance of the fishing grounds from their home port. This observation has been noted in numerous empirical studies of fishermen's spatial decisions (Eales and Wilen, 1986; Smith and Wilen, 2003). Much of the central California squid fishery is conducted by fishermen who are at least seasonally based out of Monterey (Pomeroy and FitzSimmons, 2001).[5] Models 2 and 3 in Table 2 investigate this question by including the distance from a particular site to Monterey harbor (measured by their overlap with concentric 5 km buffers). Model 2 only includes this distance variable, and it is apparent from the likelihood value and *R*-squared value that distance from port matters a great deal. Furthermore, when the RIC regressor is included in Model 3, its effect is statistically significant and of the predicted sign, but the increase in model fit is quite small and the estimated impact of RIC values is now significantly less than when modeled in isolation. These results suggest that failure to include distance from port effects may result in overstated impacts from closures, a result that is likely applicable to other fisheries since a fisherman's choice of port will depend at least in part on the proximity of desirable fishing grounds.

Models 1 through 3 are all fairly restrictive in that they presume constant effects for changes in RIC value and distance. It is entirely possible that fishermen will behave quite differently according to the values of the independent variables. For example, the increase in number of sets as areas' relative importance indices increase from 1 to 2 may be different than as indices rise from 4 to 5. To investigate this possibility, we modify

[4] The *R*-squared values reported here are defined as the square of the correlation coefficient between the predicted values from the model and the true dependent variable. This provides a measure of fit analogous to the usual *R*-squared statistic for linear regression.

[5] It is likely that fishing at the southern extent of our study region is conducted by vessels based out of Ventura, San Pedro or Channel Islands as part of the southern Californian squid fishery. However, we have chosen to ignore this possibility here to focus on the much more intense harvest near Monterey.

Model 3 by estimating a spline regression. A spline regression allows different values for the parameters over predetermined discrete intervals but ensures that the resulting expectation function is continuous at these breakpoints. A variety of breakpoints are possible for this analysis, but we have decided to break the effect of distance into a short and long-trip effect (within 10km and ≥10km respectively) and have set the breakpoints for the logarithm of RIC by its quintiles.

The results of the estimation are given under Model 4 in Table 2. First of all, it is clear that the effect of distance from port is a very local phenomenon. Beyond 10km there appears to be no effect of distance whatsoever, but within this bound the effect is far stronger than previously estimated. Fishermen *strongly* prefer to stay very close to port and this is evident by steadily declining set density gradient from Monterey harbor outward. The story told by the five RIC parameters is less clear. Three of the five coefficients have the "correct" sign and all the estimates are statistically significant. There also seems to be evidence of diminishing responsiveness to changes in RIC as RIC increases. However, two of the parameters (corresponding to RIC values of 0 to .35 and .87 to 1.46 respectively) indicate a *decreasing* relationship between RIC value and site visitation. The estimate for the lower range should probably be discounted – almost all sites in this range experience negligible visitation and so the estimate is driven by a very small number of outliers. The other negative parameter (*ln(RIC)_Q4*), which corresponds to the gap between the "humps" in Figure 5, is not so easily explained, however. One possible explanation may be the fact that the RIC values reflect the aggregated ratings of sixteen fishermen while we are only able to observe the choices of nine. The rankings of sites in this range may thus have been driven by fishermen not represented here with corresponding logbook data.

The evidence for the ability of the RIC index to predict the intensity of fishing effort is mixed. There does appear on the whole to be a positive but diminishing linkage between the two, but this relationship does not appear to be robust over the entire range. It is difficult to say whether this lack of robustness would carry over to other fisheries. Also, the proximity to port of a fishing site is actually far more successful as a single predictor than the value index, although combining the two variables in a model is preferable. We would expect a similar affinity to be displayed in a variety of other coastal fisheries, although the degree of port affinity for squid fishermen is likely extreme due to the unusual productivity of the Monterey bay ecosystem and the ability of squid fishermen to attract their harvest to them via the use of lights during night fishing.

Question 2: How robust is the relationship between RIC value and the intensity of sets across years?

This question is critical because it speaks to the applicability of the survey results for making long run predictions and to the possible time interval at which such surveys may need to be repeated. In the current context we are fortunate to have logbook data for the five years prior to the survey in 2005 and the year of the survey itself. To investigate this question, we revert to the basic setup of Model 3 in Table 2, only now we interact each variable with an indicator variable for each of the six years in the survey. Conceptually this is identical to estimating the regression for each year separately. We have also restructured the distance variable so that it only goes up to 10 kilometers, given the strong evidence in the previous section that distances beyond this point have no appreciable impact on set intensity. Due to the large number of parameters that result from this estimation we have elected to only display the estimates for the coefficients on the log(RIC) values in Table 3. The effect of distance from port was consistently of the correct (negative) sign and showed little evidence of annual variation for all but the terminal year of the sample.

The results in Table 3 are revealing. There is little statistical evidence of *any* linkage between the 2005 RIC values derived from the survey and the site preferences of fishermen from 2000 to 2003. There is, nonetheless, a very strong association in the year prior to the survey (2004) and a smaller and borderline statistically significant connection in 2005. This relationship cannot be explained by fluctuations in total effort from year to year as this effect is accounted for in the multiplicative factor α shown in equation 1. The increase in *R*-squared and increase in the log-likelihood value over all previous models seems to reinforce the importance of accounting for temporal shifts. At first glance it seems strange that the RIC values appear to reveal more about fishermen's behavior in the year prior to the survey than they do about their contemporaneous behavior. However, given that the interviews occurred roughly in the middle of the central coast squid season, it is likely that fishermen's perceptions of the grounds were still more strongly colored by the previous year's fishing experience than by their limited fishing just prior to the survey.[6] This is despite the fact that interviewers attempted to frame the questions about relative importance as indicative of "lifetime" importance to each fisherman.

These results suggest that the short term linkage between fishing behavior and the value indicators may be more substantial than suggested by the pooled analyses reported in the previous section. On the other hand, the statistical basis for employing RICs for longer-term prediction appears suspect. It is difficult, given the recentness of the survey, to evaluate just how quickly the informational content of the RIC decays through time. Zeidberg et al. (2006) found evidence of periodicity in squid catch per unit effort at lags of one year, 4.5 and 7.5 years (presumably due to the oscillatory patterns in climatic variables) that may provide some basis for the durability of the RIC information on a periodic basis.

The recruitment of squid is notorious for its dependence on environmental variables – a dependence that is magnified by their short life cycles (Zeidberg et al., 2006). For this reason we hesitate to draw any strong inferences from this analysis about the durability of the RICs for use in other fisheries. The indicators for "*K* selected" species at higher trophic levels may have a much longer "shelf life" than those for species such as squid, anchovy, and sardines. What is clear, however, is that the nature of the species matters for how impacts are derived and for the degree of confidence that might realistically be attached to them.

Question 3: What is the relationship (if any) between the importance indicator for a site and its productivity in terms of expected catch?

[6] Alternatively, given the relatively low value of old information relative to new, it could be that fishermen made a calculated decision to suppress their recent idiosyncratic "fine grain" information in favor of the information that could be inferred from older, but still applicable, information from the previous year (Wilson, 1990).

This question is important because it points to the underlying motivations of fishermen when they allocate their "pennies" to the fishing grounds. Numerous economic studies of fishing behavior have revealed (other things equal) that fishermen tend to favor sites that yield higher revenues – a variable which is usually highly correlated with CPUE. To the extent that fishing behavior follows the stated preferences implicit in the RIC index (and we have uncovered some evidence, albeit limited, that it does) then we would expect the index values to have some positive, and perhaps lagged, relationship to catch. Practically speaking, the linkage between catch and RIC value is also important because it provides the mapping from "lost effort" due to a closure to "lost catch" which can then be monetized for a measure of economic impact.

The ideal way to investigate this question would be to utilize fisherman-specific data on catch per unit effort and to then match this data with the fisherman's own assessments of each site's value. We were unable to pursue this approach for a couple of reasons, however. First, for confidentiality reasons, we were not allowed to observe a specific fisherman's ranking of a site; instead we only see the weighted average of sixteen interviews. Secondly, we lack any reliable measures of effort for the fishery. Squid fishing is often a cooperative exercise with light boats assisting seiner vessels. The presence of other fishing operations nearby may have positive (from the additional attractive power of a competitor's lights) or negative impact (congestion effects and gear interference) on the effectiveness of one's effort. Also, there is some heterogeneity within the fleets in terms of vessels and gear used. All of these factors make it difficult to deduce the latent productivity of fishing grounds from data that simply report a location and a total catch for a set.

These matters aside, one would nevertheless expect to observe in aggregate what one would hope to see on an individual basis – namely, that the value ascribed to a site is determined to a significant degree by the expected catch and the expected cost of accessing the site (proxied for by distance to port). To test this hypothesis we estimate the following equation:

$$\ln(RIC) = \beta_0 + \beta_1 Avg\_Catch2000 + ... + \beta_6 Avg\_Catch2006 + \beta_7 Dist\_km + \beta_8 Dist\_km\text{^}2 + \varepsilon.$$

This equation allows for annual variation in the relationship between RIC and the average catch at a site and also allows for a nonlinear effect of distance. We elect to specify the equation in terms of the logarithm of RIC due to the relative superiority of this specification in terms of model fit. To avoid a distortion of the estimates of the parameters on distance from port we only consider sites within 200km of Monterey so as to avoid the inclusion of sets by southern California based vessels. This equation is estimated via OLS and is shown in Table 4.

The effects of distance from port are predictably negative and decreasing over a large portion of the sample range.[7] Fishermen do appear to downgrade their rating of a site based on its distance from port. The evidence for the effect of the average catch

[7] The marginal impact of increased distance reaches zero at around 75km by our estimates. Only 16 observations (3% of our estimation sub-sample) lie beyond this distance.

recorded at a site in a year is less reassuring. Indeed, the lack of significance for any of the coefficients associated with average catch suggests that the RIC values contain *no* information on the relative productivity of fishing sites! This basic result proved remarkably consistent over a range of alternative specifications and estimation approaches (not reported).[8]

There are a number of factors that may be driving this result. First, our estimates are based on fleet-wide averages of annual catch and therefore may be a poor proxy for a true measure of CPUE. Second, the RIC measure is based on the nonlinear aggregation of ratings from sixteen fishermen. Given the large number of fishermen missing from our sample and the complexity underlying the relative weighting of individual preferences in the RIC measure, we might expect the association between site productivity and the value index to be weakened. Third, congestion effects and gear interference may serve to level profit differentials across sites so that fishing decisions are largely based on cost minimization. This seems quite likely given that most squid fishing occurs within a very small geographic area and at night so that competing fishermen can easily locate one another. Finally, fishermen may care not only about the *expected* level of catch at a site but they may also care a great deal about the *risk* (both in terms of variation in catch and exposure to unsafe fishing conditions) that they may face in fishing. If sites with higher expected payoffs also involve more risk then the two effects may cancel each other out.

We strongly doubt that this lack of association between the relative importance indicators and observed catch would carryover to many other fisheries in the sample. Provided that fishermen responded to the survey truthfully, then it seems reasonable to posit that they would consider the relative productivities of fishing sites in their assignment of value to those sites.

Question 4: How well do the RICs predict the distribution of fishing effort for the fishermen that were not included in the interview process?

Ultimately, the purpose of the RICs is to aid in predicting the lost effort (and thus the lost economic benefits) from a series of closures. These must be calculated over the *entire* fleet, not just the portion that was interviewed and so it is important to assess just how well the indicator values predict the behavior of the fishermen that were not included in the survey.

A casual glance at the distribution of sets in Figures 1 and 2 suggests that the two groups of fishermen had roughly the same spatial distribution over the sample period. This supposition is confirmed in Figure 6 which shows that the sets of the two groups exhibit very similar degrees of spatial scatter and orientation, although the average set location of the non-interviewed fishermen is slightly more to the north than that of their interviewed counterparts. Furthermore, the kernel density graphs (essentially smoothed histograms) in Figure 7 show that the two groups have similar distributions of effort in terms of the relative importance indicators. All this evidence would seem to suggest that

[8] A word of caution is due here. Given our specification, we can only estimate this equation for a relatively small number of sites that had a catch record (i.e. were visited at least once) for every sample year, possibly biasing the results. We partially accounted for this by estimating a separate regression for each year, but the results remained essentially the same.

the non-interviewed fishermen (in this case at least) would have described their fishing grounds, and the values they assigned to them, quite similarly to those who were surveyed.

To confirm this intuition, we replicate the first three models from Table 2, only now with the dependent variable being the annual intensity of sets by the non-interviewed fishermen. These results are shown in Table 5. The annual scaling parameters α are also changed to reflect the relative numbers of sets between the two groups so that the parameter estimates are directly comparable to those for the interviewed fishermen. The relationships between the variables are quite similar to those noted previously for the interviewed fishermen. All of the qualitative results remain the same and most of the parameter values are statistically indistinguishable in magnitude across equations. The fit of the models is a bit lower than for the interviewed fishermen but this is to be expected given that the preferences of the non-interviewed fishermen are not factored in the RIC values.[9] Results of a spline regression like Model 4 in Table 2 (not reported) showed a qualitative pattern analogous to that for the interviewed fishermen.

These results provide support for the use of RIC in predicting the distribution of fishing effort by squid fishermen that were not surveyed. Although the predictive power of the value index is not quite as high as it is for the interviewed fishermen, it appears that the two groups can be treated as roughly equivalent in their spatial behavior. Given the different sampling approach used for this fishery compared to most of those in the Ecotrust study, it is difficult to say whether this close similarity of behavior between groups can be expected for other fisheries.

IV. Conclusion

The purpose of this analysis was to use the squid fishery as a test case to examine the usefulness of Ecotrust's "importance indices" (RIC) for assessing the potential impacts of reserves. The usefulness of the method depends upon two issues. The first issue is whether the importance indices computed via Ecotrust sampling/survey/mapping procedures correspond to actual measures of behavior. The second issue is whether the measures can be used to illuminate how a reserve program might impact fishermen.

The squid fishery was selected to do this test case assessment because its logbooks are believed accurate. While the squid fishery may be the best test case because its logbooks are relatively accurate, we faced some special problems related to confidentiality of the RIC data. We spent several days negotiating with Ecotrust and CDFG over how we might access and examine data and still preserve the confidentiality of that data. The compromise we reached in the end was imperfect. It called for preserving the confidentiality of individual fishermen's maps (raster files) about their most important sites, and instead aggregating all of the interviewees' individual files into one group-wide importance index for the Central Coast. We then assembled another data set pertaining to the interviewed fishermen's logbooks, scrambled to protect

---

[9] *R*-squared actually decreases in the movement from Model 2 to 3. This is possible with Poisson regression since maximum likelihood estimation (unlike least squares) does not maximize *R*-squared. The substantial increase in the log likelihood value supports Model 3 over 2, however.

confidentiality. A third data base was constructed of the logbook data for individuals who were not interviewed.

Unfortunately, the compromises to protect confidentiality precluded undertaking the most illuminating investigation, namely examining whether each interviewed fisherman's most visited or highest yield sites were also ranked as his/her most important sites with the "pennies" allocation experiment. Instead we were only able to test whether the importance index surface for the whole interviewed **group** was correlated with the whole group's pattern of hauls. Working only with the aggregated data sets led to other matching problems that were difficult to resolve. Only 9 of the 16 interviewees used to construct the aggregated squid spatial importance index actually have corresponding logbook data. Of those 9, 4 have logbook data with a degree of rounding that makes it difficult to accurately associate all logbook sets with specific GIS data points (Table 6). The resulting sample over which we are able to assess the importance indices method thus is small and possibly non-representative.

Despite the fact that our test case is not ideal, we report here a thorough analysis of the aggregated importance indices for the interviewed group and the net-set data by both the interviewed and non-interviewed groups, over periods that include the interview period and years back to 2000. We believe that we have exhausted the possible ways to test the usefulness of the methods, given the data and statistical tools that are available and intuitively understandable. Our findings are as follows:

1. for the sample of interviewed squid fishermen, there is a weak correlation (explaining about 5% of the variation) between the aggregated importance indices and the sites actually visited by those with logbook data in the interview sample (Figure 5, Table 1; models 1-3).
2. the relationship is not monotonically increasing (as we might expect) with sets per unit area rising as the importance index rises (Table 1; model 4). Our suspicion is that this is likely due to the raster file/logbook mismatch; of those 16 used to make up the aggregated importance index, only 9 have matching logbook data. The non-monotonic relationship thus likely reflects situations for which certain fishermen were interviewed and had their importance indices recorded, but for whom logbook data is missing. We could have excluded them if we were not prevented by confidentiality from individually associating raster files with logbooks.
3. while there is an association between effort (hauls) and the importance index (Table 3), there is no statistical association between the index and catch (Table 5). Thus while fishermen seem to be attracted to areas they designate as important, it is not necessarily because they produce higher catches per net haul. This is a surprising finding and one that is at variance with other economic studies of fishermen's location choices. There are several explanations for this. First, the catch per haul data may be too noisy. Logbooks filled out at sea are estimates and if these are not corrected by processors at the point of landing, they may be too inaccurate to be useful. Second, catch per haul may not reflect abundance or CPUE but instead contractual relationships to deliver certain quantities to processors. Third, squid may aggregate around the lights in relatively constant densities so that

catch per haul does not vary over space systematically. In this case, fishermen would choose areas that minimize travel distance, explaining the fact that distance is the most important predictor of the RIC. Fourth, fishermen may be rapidly depleting local squid populations down to levels that are just profitable, equalizing distance-adjusted catch per haul everywhere.

4. The strongest relationship between the aggregate importance index surface and effort is associated with the 2004 pre-survey year catch (Table 3). This suggests that interviewees reported importance indices that reflect most recent information rather than sites with highest "lifetime importance" as intended by interviewers. The use of most recent information is not particularly surprising given that the squid fishery varies temporally and spatially and given that expectations often reflect the recent past. But it casts some doubt on the usefulness of these indices for gauging permanent reserve impacts for a spatially variable species. Unfortunately, since we were only able to examine this one (spatially variable) test case, we cannot examine the opposite hypothesis, namely that fishermen fishing sedentary species would accurately associate high importance indices to areas with high permanent productivity.
5. Non-interviewed fishermen behave similarly to interviewed fishermen in this test case, even though the interviewed sample is non-random. While the statistical mass of non-interviewed fishermen's hauls is shifted slightly to the south compared with the smaller interviewed sample (Figures 1, 2 and 6), the median hauls are close to identical. In addition, the same basic relationships explain the density of sets per square kilometer, namely distance from port and (marginally) the RIC (Table 4).

Overall, we conclude that for the squid fishery test case, the index computed by Ecotrust's sampling/survey/mapping procedures is associated in an expected manner with actual behavior on the part of sampled fishermen. High values ascribed to importance indices are related to higher effort levels in those areas, although the statistical association is weak and not monotonic (Figure 7). Moreover, even with the non-random sampling, the group of fishermen who were not sampled seems to exhibit similar spatial choices as its sampled counterpart. Although we caution that one cannot make too much out of analysis of a small and imperfectly disaggregated sample, we suspect that the attention the Ecotrust gave to sampling protocol, and the involvement of fishermen in the data gathering design, led to honest survey answers and reliable data. As to the remaining question about whether information about effort behavior can be translated into economic impacts, there is less that can be said with confidence from this case study. As this test case suggests, importance indices computed for species with variable spatial patterns may rely on ad hoc projection of most recent information rather than some longer term historical averaging.

References

Eales, J. and J.E. Wilen. 1996. "An Examination of Fishing Location Choice in the Pink Shrimp Fishery." *Marine Resource Economics* 2(4): 331-351.

Smith, M.D. and J.E. Wilen. 2003. "Economic Impacts of Marine Reserves: The Importance of Spatial Behavior." *Journal of Environmental Economics and Management* 46: 183-206.

Pomeroy, C. and M. FitzSimmons. 2001. "Socio-Economic Organization of the California Market Squid Fishery: Assessment for Optimal Resource Management." California Sea Grant Project. Available online: http://www.psmfc.org/efin/docs/otherpublications /Pomeroy_&_FitzSimmons_2001.pdf.

Wilson, J.A. 1990. "Fishing for Knowledge." *Land Economics* 66(1): 12-29.

Wooldridge, J.M. 1997. "Quasi-Likelihood Methods for Count Data," in *Handbook of Applied Econometrics*, Vol. 2, ed. M.H. Pesaran and P. Schmidt. Oxford: Blackwell, 352-406.

Zeidberg, L.D. et al. 2006. "The Fishery for California Market Squid (*Loligo opalescens*) (Cephalopoda: Myopsida) from 1981 through 2003." *Fisheries Bulletin* 104(1): 46-59.

Appendix: A Discussion of Data Inaccuracy and its Possible Effects on the Analysis

As mentioned in the body of this report, a number of sets had to be dropped from the analysis due to their questionable location coordinates. In all 223 (13%) sets were dropped from the interviewed fishermen's logs while 1179 (24%) were dropped from those of the non-interviewed fishermen. The effects of this trimming do not fall equally across all fishermen. This appears to be driven by a propensity on the part of some fishermen to excessive rounding of their spatial coordinates. Such behavior may be due to limited GPS or other navigational technology onboard the vessel or may occur deliberately as fishermen attempt to hide their choice fishing spots from regulators.

The propensity of different fishermen to round their set locations can be seen in Table 6 which presents the number of rounded locations (those which are not recorded at the maximum resolution of a hundredth of a minute) for each interviewed fishermen in the trimmed dataset.

The implications of imprecise set locations are twofold. First, the fact that some fishermen only record locations to the nearest minute of latitude or longitude may lead to the systematic exclusion of some fishermen over others. If some of the "heavy rounders" exhibit markedly different spatial preferences from their competitors, then we might expect the increased exclusion of their records to distort the relative spatial distribution of fishing effort, possibly causing a divergence between the pattern of effort suggested by interviews and the actual one. Also, differences in the propensity to round between the interviewed and non-interviewed groups can lead to a greater rate of exclusion for one group than another, distorting the relative amounts of total effort (and catch) contributed by each group. This second issue can be corrected by looking at the untrimmed data but the first is irreconcilable unless further assumptions are employed to recover the true location of imprecisely recorded sets.

A second implication of rounding of coordinates is that some of the set locations in the non-trimmed data may be suspect. In many cases it is impossible to separate sets that were rounded from a more precise location from a set that truly occurred at that particular coordinate. This is especially true in the case of latitude since the squid fishery occurs over a wide range of latitudes but is limited to a fairly narrow longitudinal band along the coast. By our calculation, the grid unit of 100 meters is approximately equal to 3 seconds; therefore, spatial coordinates must be recorded to at least the level of the second to be able to reliably match a set to a specific cell. In decimal terms, it is only when latitude and longitude are recorded to the hundredth of a minute (0.6 seconds) that such a level of accuracy is achieved. Anything less precise may result in a possible misattribution of points to raster cells (or their associated value-aggregated polygons). Table 6 suggests that an upper bound of 23% of set locations may suffer from this problem, although the actual number is likely far less. The impacts of this problem are difficult to ascertain, but we have elected to retain these questionable observations rather than discard them for the simple reason that to do so would result in a substantially smaller sample and the complete elimination of two fishermen from the analysis.

Fisherman ID	2000	2001	2002	2003	2004	2005	Total
L00000	0	0	12	0	9	0	21
L00001	0	0	0	10	19	0	29
L00012	2	14	18	0	0	0	34
L00123	0	0	0	51	37	0	88
L00321	38	43	138	114	56	48	437
L01234	2	58	143	102	7	27	339
L04321	4	2	6	50	17	50	129
L12345	0	33	25	25	23	0	106
L54321	0	2	93	49	70	37	251
Total	46	152	435	401	238	162	1434

Table 1: Number of sets (after trimming) for interviewed fishermen along the central coast.

Dependent Variable: Annual Sets per Square Kilometer (Interviewed Fishermen)				
Independent Variable	Model (1)	Model (2)	Model (3)	Model (4)
ln(RIC)	1.5319 (.1710)*		.7533 (.2192)*	
ln(RIC)_Q1				-2.0184 (.4080)*
ln(RIC)_Q2				5.7578 (1.3971)*
ln(RIC)_Q3				2.6622 (1.2325)*
ln(RIC)_Q4				-3.4282 (1.0598)*
ln(RIC)_Q5				1.6554 (.4474)*
Dist_km		-.1454 (.0388)*	-.1113 (.0405)*	
Dist_lt10km				-.5863 (.0464)*
Dist_gt10km				.0025 (.0024)
Constant	-1.9558 (.1337)*	.1220 (.2310)	-.3322 (.3192)	-4.3797 (.7577)*
Log-likelihood	-43934.093	-37580.979	-36071.917	-25518.67
R-squared	.0218	.0760	.0796	.1463

* Indicates the parameter is significant at the 5% level

Table 2: Poisson regression results for interviewed fishermen (standard errors in parentheses).

Dependent Variable: Annual Sets per Square Kilometer (Interviewed Fishermen)	
Independent Variable	Parameter Estimate
ln(RIC)_2000	.1875
	(.8026)
ln(RIC)_2001	-.3661
	(.3100)
ln(RIC)_2002	.3953
	(.2982)
ln(RIC)_2003	.1910
	(.2124)
ln(RIC)_2004	2.1635
	(.5554)*
ln(RIC)_2005	1.1514
	(.6743)**
Log-likelihood	-24260.855
R-squared	.1579

* Indicates the parameter is significant at the 5% level

** Indicates the parameter is significant at the 10% level

Table 3: Estimated link between relative importance indices and annual set intensity when the effect is allowed to vary by year.

Dependent Variable: Annual Sets per Square Kilometer (Non-interviewed Fishermen)			
Independent Variable	Model (1)	Model (2)	Model (3)
ln(RIC)	1.7229		1.2980
	(.1394)*		(.2392)*
Dist_km		-.0827	-.04024
		(.0226)*	(.0206)*
Constant	-1.9032	-.0630	-.8997
	(.1269)*	(.2074)	(.3247)*
Log-likelihood	-119590.02	-116163.62	-104251.18
R-squared	.0201	.0538	.0483

* Indicates the parameter is significant at the 5% level

Table 4: Poisson regression results for non-interviewed fishermen

Dependent Variable: ln(RIC) (Interviewed Fishermen)**	
Independent Variable	Parameter Estimate
Avg_Catch_2000	.0059
	(.0240)
Avg_Catch_2001	-.0097
	(.0085)
Avg_Catch_2002	.0017
	(.0047)
Avg_Catch_2003	.0046
	(.0054)
Avg_Catch_2004	.0044
	(.0093)
Avg_Catch_2005	-.0040
	(.0385)
Dist_km	-.0307
	(.0031)*
Dist_km^2	.0002
	(.00003)*
Constant	.5662
	(.0626)*
R-squared	0.2520

* Indicates the parameter is significant at the 5% level

**Regression only includes sites within 200km of Monterey

Table 5: Linear regression investigating the link between RIC value, expected catch and distance from port (heteroskedasticity-robust standard errors in parentheses).

Fisherman ID	Not Rounded	Rounded	% Rounded
L00000	0	21	100%
L00001	28	1	3%
L00012	0	34	100%
L00123	86	2	2%
L00321	418	19	4%
L01234	333	6	2%
L04321	7	122	95%
L12345	21	85	80%
L54321	216	35	14%
Total	1109	325	23%

Table 6: A comparison of rounding practices by interviewed fisherman in the trimmed dataset. (Rounding is defined as any spatial coordinate not recorded to the hundredth of a minute of latitude or longitude.)

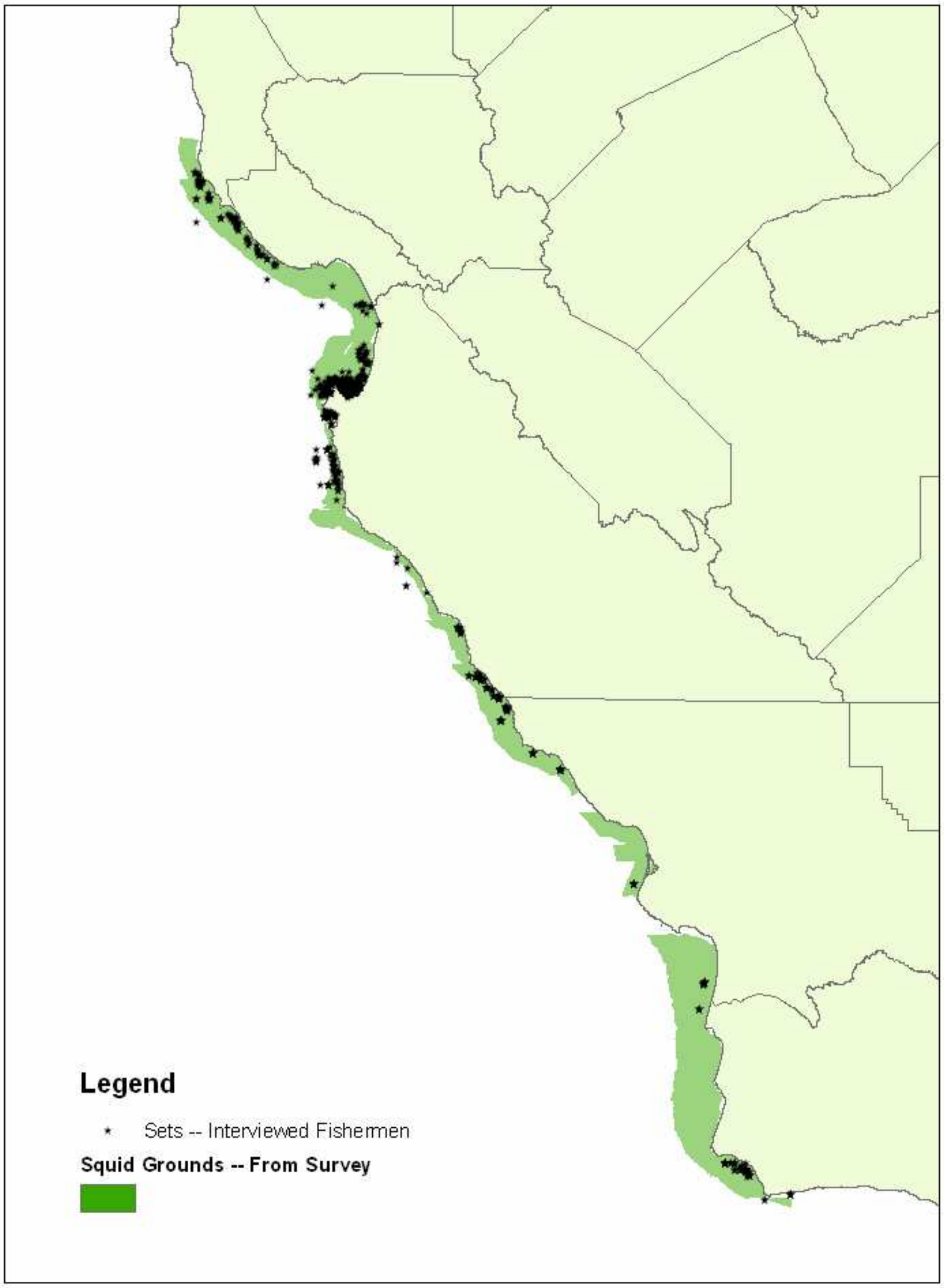

Figure 1: Distribution of sets for interviewed squid fishermen 2000-2005.

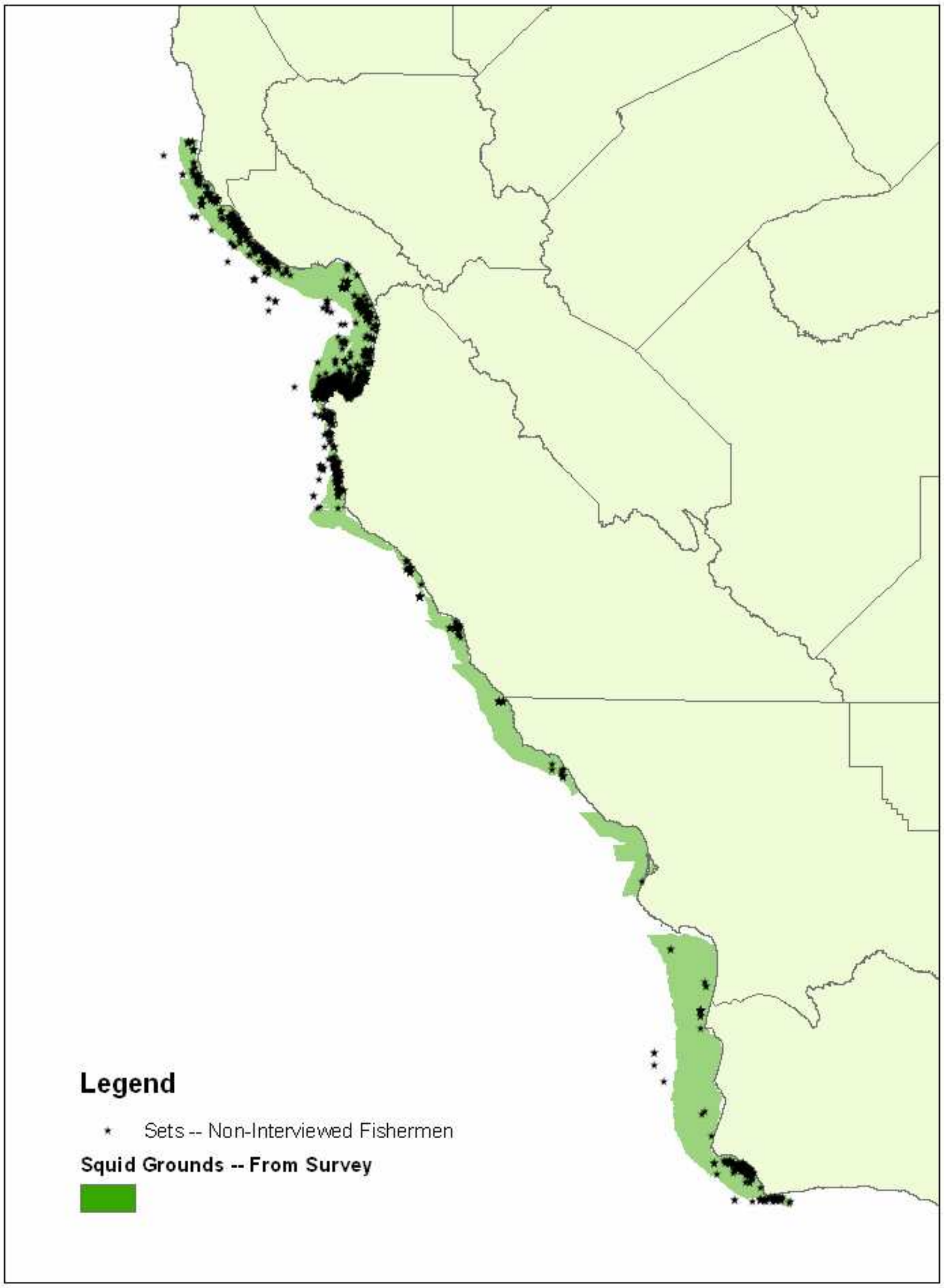

Figure 2: Distribution of sets for non-interviewed fishermen 2000-2005.

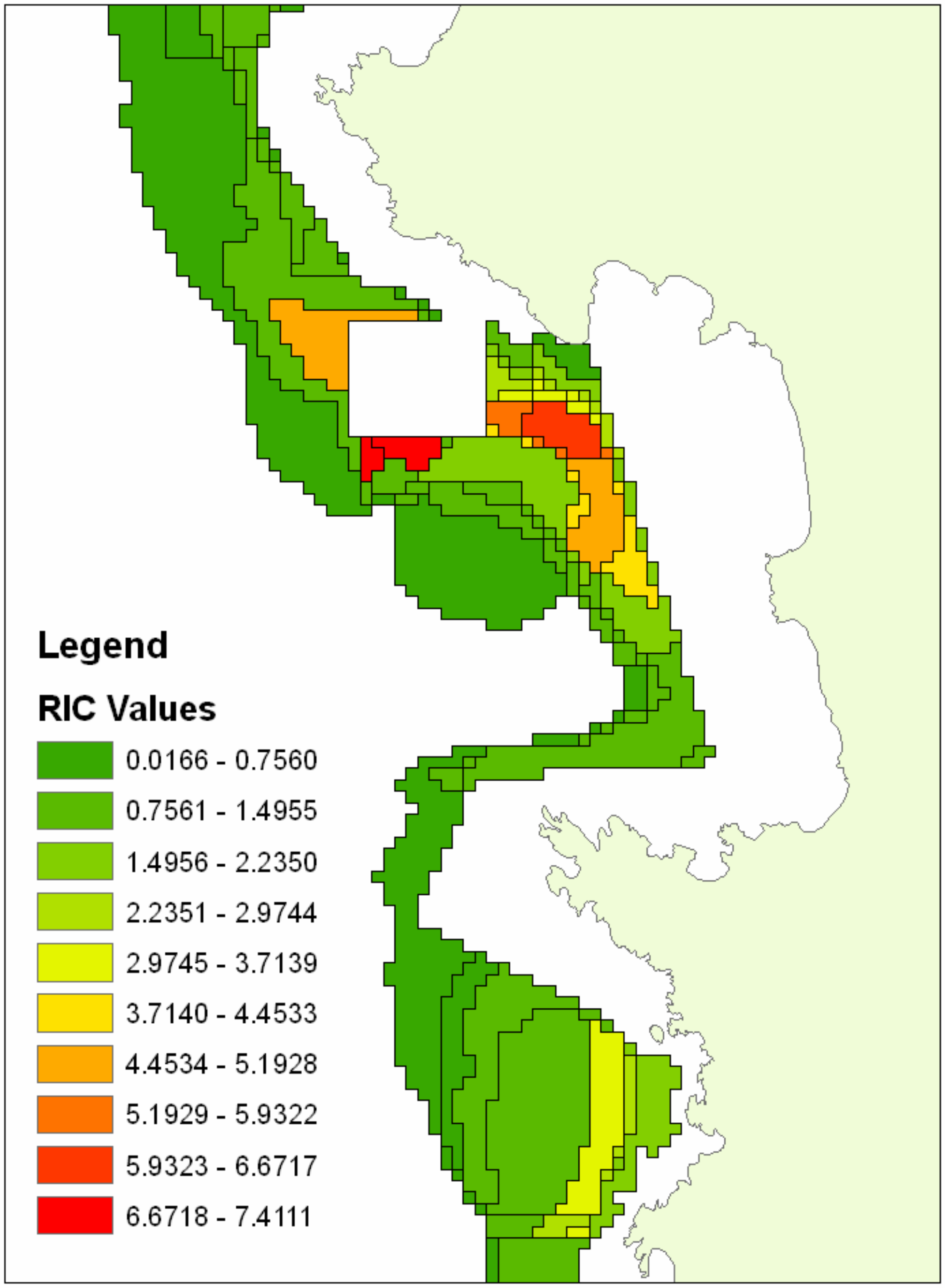

Figure 3: Example of raster to polygon conversion for an area of coast south of Monterey Peninsula.

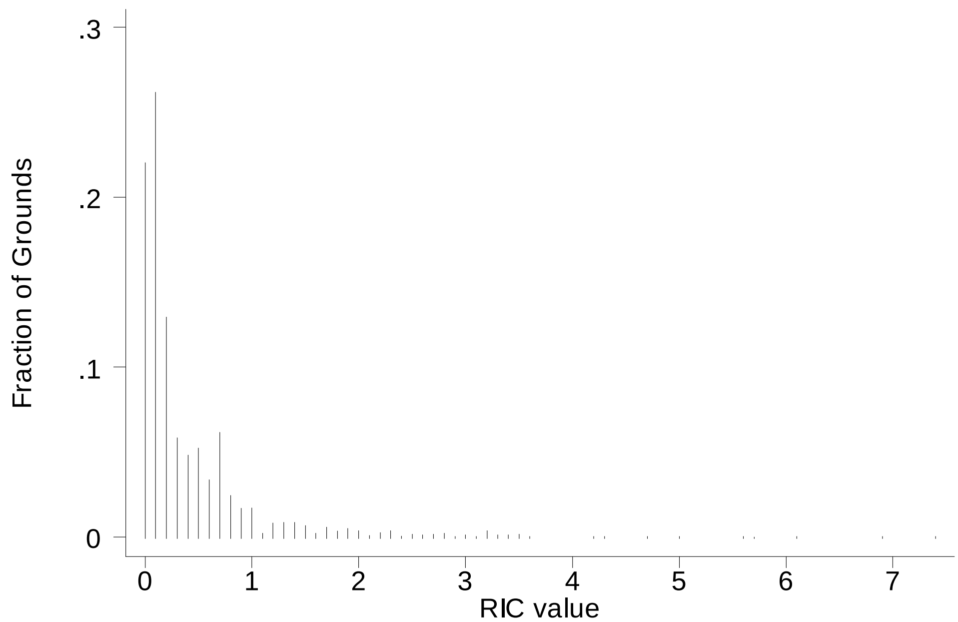

Figure 4: Areal distribution of relative importance indicator values across the fishing grounds

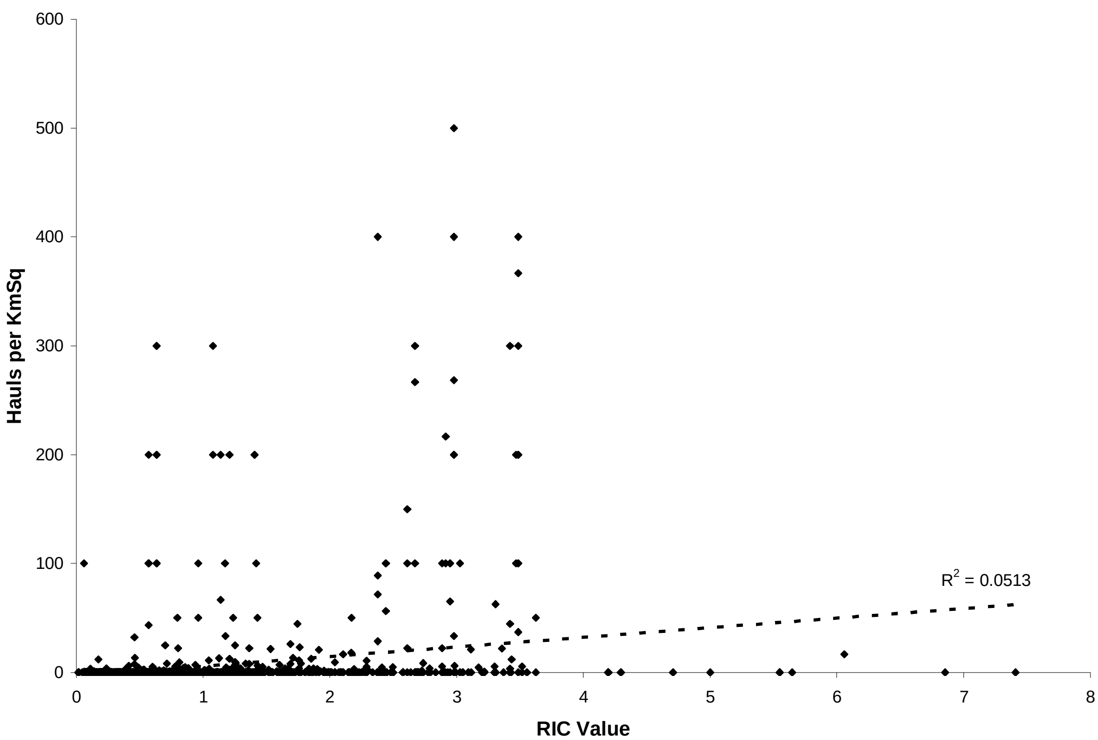

Figure 5: Scatterplot of hauls per square kilometer vs. the relative importance indicator for interviewed fishermen.

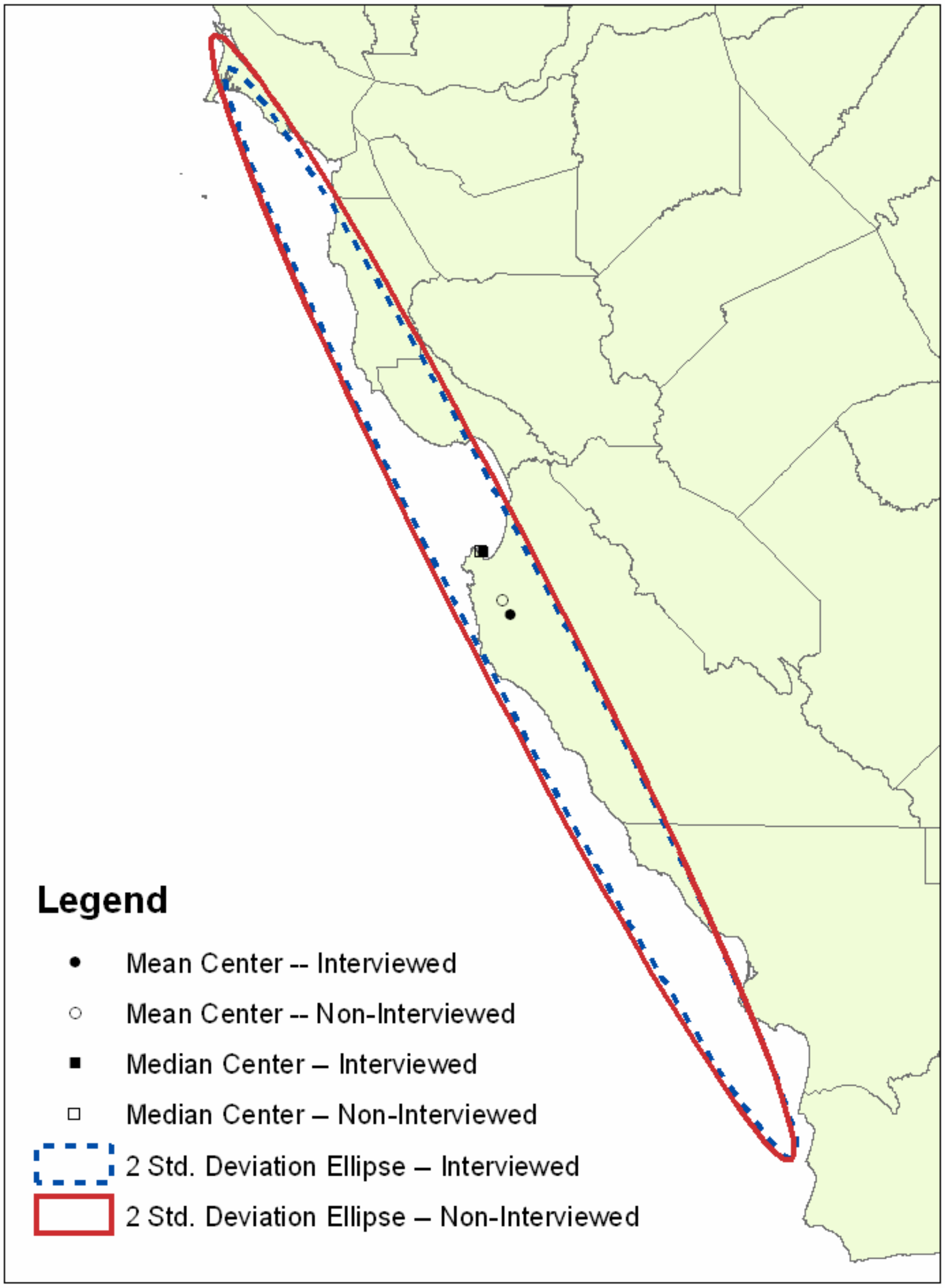

Figure 6: The relative spatial distributions of the interviewed and non-interviewed fishermen

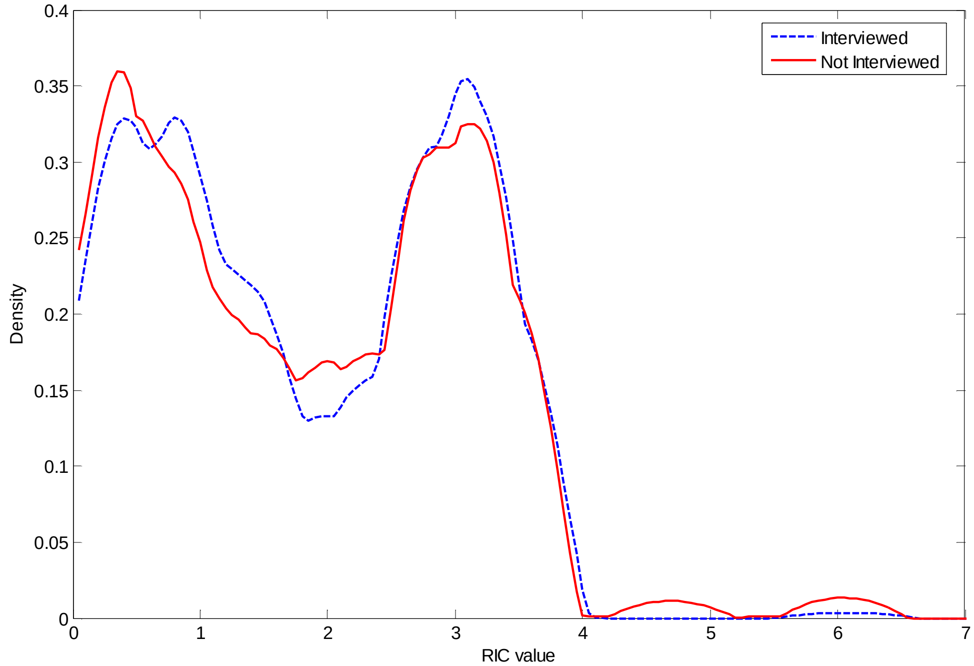

Figure 7: Kernel density distribution showing the proportion of sets executed in areas of particular RIC values for both interviewed and non-interviewed fishermen.