# Estimating Violation Risk for Fisheries Regulations

Fred S. Roberts
DIMACS and CCICADA
Rutgers University

LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY



CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Fisheries Rules

- The United States sets rules for fishing with the goal of maintaining healthy fish populations.
- Rules depend on specific species and include
  - Allowable locations to fish
  - Allowable seasons to fish
  - Catch quotas
- Violations of the rules leads to fines – sometimes quite large

Endangered Atlantic Cod

LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY

CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Enforcement of Fisheries Rules

- Many agencies are involved in enforcing the fisheries rules and regulations.
- One of those agencies is the US Coast Guard.
- Through the Laboratory for Port Security at Rutgers and the CCICADA Center, we have been working with the Coast Guard to define and enhance scoring rules to lead to better enforcement of fisheries rules.

LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY

United States Coast Guard
U.S. Department of Homeland Security

CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Enforcement of Fisheries Rules

- This work has gotten me to some interesting places.



LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY



CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Fisheries Law Enforcement

- The US Coast Guard District 1 (based in Boston) uses a ***scoring system called OPTIDE to determine which commercial fishing vessels to board to look for violations.***
- The OPTIDE rule was built based on expert judgment and intuition.
- They asked us if their success rate in finding violations by boarding could be improved by use of sophisticated methods of data analysis.
- Goal: refine the ability to determine the risk profile of vessels.

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Fisheries Law Enforcement

## Examples of Scoring Rule Components

- Points for current or past negative intelligence reports
- Points depending upon date last boarded
- Points based on information about the type of boat
- Points for having found violations in past boardings – depending upon type of violations
- ***Board if total score (number of points) exceeds a threshold***

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Many Goals of Fisheries Law Enforcement

- While our project is concerned with increasing success rates from boarding, there are many other goals of fisheries law enforcement:
  - Balanced deterrent
  - Balanced policing
  - Balanced maintenance of safe operations


Modeling
Technology
Policy
LABORATORY FOR PORT SECURITY

Command, Control, and Interoperability
Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Many Goals of Fisheries Law Enforcement

- The project started with the following definition of the goal: ***Find a decision rule for deciding whether or not to board that leads to as large a percentage of times as possible in which boarding leads to finding a violation.***


LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY



CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Many Goals of Fisheries Law Enforcement

- Complication: There are different types of violations:
  - Fisheries violations
  - Safety violations
  - It's hard to compare seriousness of a fishery violation vs. a safety violation
- Complication: Some violations are more serious than others
  - Categories of violations (e.g., simply severe or not severe)
  - Ranking of violations (from least to most severe)

CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Many Goals of Fisheries Law Enforcement

- This leads to need for metrics for scoring violations and a change of goal to maximize scores obtained when boarding.
- Let V be the ***violation score*** obtained from a boarding.
- In work to date, we have been letting V = 0 or 1, depending on whether or not a violation is found.
- Alternatives:
  - V = number of violations found
  - V = weighted sum $\Sigma\, w_i x_i$ where $x_i = 1$ if the $i^{th}$ violation is found and 0 otherwise, and $w_i$ is some weighting factor

# Data from the Coast Guard

- Coast Guard provided us with 11 years of available data on USCG activities and violations incurred by commercial fishing vessels
- Anonymized data
- Partial information allowing us to infer (most) of the scoring rule components.

# Project Challenges

- **Complex Domain**
  - Fishery regulations, quotas, migration patterns, weather, regional variations, close-to-shore/ offshore, etc.
  - Complex human behavior influenced by economic, environmental and human factors
  - Game-theoretic aspects: deterrent, adversarial behavior

# A Variety of Relevant Features

- ***In addition to working with Coast Guard features, we have looked at introducing other features***, such as:
  - Weather
  - Seasonality
  - Fish migration
  - Key fish species
  - Home port
  - Detailed vessel history

- Economic data (e.g., fish prices)
- Socioeconomic factors (such as type of family boat vs. large commercial fishing boat, or attitudes toward law enforcement)

# Looking at Data

- Data is Challenging
- Sample Challenge: Fish Price
- Premise: When fish price is high, there will be more violations
- Fish price data available from NOAA (National Oceanic & Atmospheric Admin)

- Challenges we have encountered:
  - Data comes in text form
  - Data doesn't have a standard format.
  - The names, the order of fish types and presentation of price are irregular, change from one text file to another.

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Sample NOAA Fish Price Data

BOSTON, MASS. NEW ENGLAND FISH EXCHANGE AUCTION
FISH LANDINGS AND PRICES IN 1,000 LBS & $/CWT DATE 8/10/12
NO LANDINGS REPORTED
BASE NEW ENGLAND SEAFOOD DISPLAY AUCTIONS FISH LANDINGS & PRICES IN 1,000 LBS & $/CWT DATE 8/9/12
PRICES INCLUDES DEALERS FEES
1/ 0 MEANS LESS THAN 100 POUNDS

| SPECIES | LBS | MIN | HIGH |
|---|---|---|---|
| COD WHALE | 1/0 | 416 | 416 |
| COD LGE | 1.0 | 398 | 536 |
| COD MKT | 3.1 | 352 | 392 |
| COD SCRD | 0.8 | 344 | 354 |
| COD MIXED | 0 | 209 | 238 |
| GILLNET LGE COD | 0 | 419 | 419 |
| GILLNET MKT COD | 0.2 | 356 | 371 |
| GILLNET SCRD COD | 0 | 296 | 296 |
| HADDOCK | 1.2 | 412 | 442 |
| HADDOCK SCRD | 2.1 | 339 | 359 |
| STRIPED BASS | 4.6 | 292 | 302 |
| POLLOCK | 0.7 | 77 | 105 |
| POLLOCK MED | 6.1 | 79 | 106 |
| POLLOCK SCRD | 2.9 | 79 | 79 |
| CUSK | 0.1 | 37 | 37 |
| HAKE LGE | 1.0 | 112 | 179 |
| HAKE MED | 0.3 | 112 | 127 |
| HAKE SML | 0 | 62 | 62 |
| HAKE SOW | 0.1 | 272 | 324 |
| BLUEFISH LGE DRESSED | 0 | 63 | 63 |
| BLUEFISH MED DRESSED | 0.2 | 65 | 75 |
| BLUEFISH MED RND | 0.1 | 88 | 88 |
| OCN PRCH | 3.9 | 62 | 100 |
| YELLOWTAIL LGE | 1.3 | 244 | 263 |
| YELLOWTAIL SML | 1.7 | 190 | 208 |
| YELLOWTAIL MIXED | 0.5 | 62 | 194 |

# RIPTIDE: Rule Induction OPTIDE: Automatically Learning Violation Prediction Rules

# Goal

- Build a predictive model that given a ship's current features and history predicts the likelihood of a violation
  - Outperform current Coast Guard rule's predictive power
  - Demonstrate model's applicability to increase Coast Guard's boarding efficiency

# Machine Learning Approaches

- Looked at machine learning methods to see if other features, or combination of present features and new ones, can lead to decision rules that obtain higher success rate from boardings.
- Represent boarding activities by a set of features
- Aim to learn a classifier that will output “board” or “don’t board” based on the features
- Choosing the features: Combination of data analysis, intuition, and a lot of trial and error

LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY

CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Machine Learning Approaches

- Our approach: boosted decision tree
- Useful for comparison to rule-based approach like OPTIDE.
- In boosting, instead of learning a single decision tree, we learn multiple decision trees on different training sets.
- We then learn the “best” weights for combining results of individual decision trees into an overall boosted decision tree

LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY

CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Machine Learning Approaches: RIPTIDE

- RIPTIDE is our classifier obtained from USCG data
- RIPTIDE = Rule Induction OPTIDE
- Our best model for RIPTIDE uses some new features, such as type of vessel (General, Trawler, Pot/Trap) and prior violations per boarding
- Much experimentation.
- ***Best model for RIPTIDE found so far outperforms OPTIDE up to 87% in an experiment***
- This model uses some features not used in OPTIDE, e.g., distance to coast, vessel subtype

# Machine Learning Approaches: RIPTIDE

- ***Best model for RIPTIDE found so far outperforms OPTIDE up to 87% in an experiment***
- Experiment:
  - Choose random set of $k$ vessels, rank elements according to the model (OPTIDE, RIPTIDE), test whether top-ranked vessel has a violation. Repeat experiment many times
  - If $k = 30$, RIPTIDE does 87% better than OPTIDE
  - If $k = 20$, RIPTIDE does 76% better than OPTIDE
  - If $k = 10$, RIPTIDE does 38% better than OPTIDE
- RIPTIDE is best at larger k, maybe larger than what Coast Guard would use

# Machine Learning Approaches: RIPTIDE

- Theoretical implications of these findings remain to be explicated in future work, which our USCG partners are currently undertaking in exploration of our new ideas.
- In practice, if use RIPTIDE, would need to "retrain" models at regular intervals (e.g., annually)
- RIPTIDE software was delivered to Coast Guard District 1; it is being evaluated for extension to entire Coast Guard system

# Next Steps

- **Connect additional data**
  – Fish prices, weather, "quota reset days", etc.

- **Improve prediction models**
  – Better individualized violation behavior models
  – Link additional data sources to improve prediction

LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY



CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# De-OPTIDE: Data-Enhanced OPTIDE: Regression Methods

# Regression Models

- Looked at regression models to derive alternative weights for the same features used in OPTIDE, based on some of the data
- Developed decision rules based on derived weights
- Tested those decision rules on rest of the data

Modeling
Technology
Policy
LABORATORY FOR PORT SECURITY

Command, Control, and Interoperability
Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# OPTIDE Approach

- Underlying assumption:
  - A violation is related to an underlying score S which is a weighted sum of some predictor variables

$$\text{Boarding: } \boldsymbol{B} = \begin{cases} \text{Yes,} & \text{if } S \geq \boldsymbol{d} \\ \text{No,} & \text{if } S < \boldsymbol{d} \end{cases} \qquad \text{where } \boldsymbol{d} = \text{threshold}$$

# Statistical Latent Model

- The *same* assumption that violation is related to an underlying score S, plus a potential error term, leads us directly to a statistical model: a logistic regression model.
- This LR model directly studies the probability of having a violation

$$\text{Violation:}\quad V = \begin{cases} \text{Yes,} & \text{if } \mathbf{S} \geq d \\ \text{No,} & \text{if } \mathbf{S} < d \end{cases}$$

LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY



CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Logistic Regression Model

$$S = W_1X_1 + W_2X_2 + \ldots + W_nX_n + \text{error}$$

- $X_i$ predictor variables (OPTIDE features)
- $W_i$ = weights
- Error is normally distributed with mean 0, variance $\sigma^2$
- We use the data set available to us to determine (estimate) the weights $W_i$ that will be used to create a new decision rule

CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# New Decision Rule

- New score $S_{new}$ is a weighted sum of the same set of predictor variables, but with a new set of weights $W_i$ determined by our data analysis using the logistic regression model.

- Then,

$$\text{Boarding: } \hat{B} = \begin{cases} \text{Yes, } & \text{if } S_{new} \geq \boldsymbol{d}_{new} \\ \text{No, } & \text{if } S_{new} < \boldsymbol{d}_{new} \end{cases} \quad \text{where } \boldsymbol{d}_{new} = \text{threshold}$$

LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# New Decision Rule

- Choosing new threshold $d_{new}$:
  - Several ways. One way: Choose a required % of vessels for which rule recommends boarding
  - In our case, 10% seemed useful because as get lower, success rate at finding violations increases.
- Because new decision rule is determined from weights obtained from historical data, we call the new rule data-enhanced OPTIDE: DE-OPTIDE

# A Simulation Study

- We randomly split the boarding data set available to us into two subsets:
  - 50% used for training; 50% used for validating
- Fit a LR model to the 50% data; then, based on the data analysis results, we come up with a set of new weights and a new decision rule.
- We apply the new rule to the remaining 50% of data and see how effective we are.
- To control variation due to random split into two classes, repeated this 10 times.
- Thus, not testing a single decision rule, but a method

Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Results

- Measured against the percentage of cases where boarding is recommended, DE-OPTIDE always has a better boarding efficiency than OPTIDE.
- The advantage of DE-OPTIDE over OPTIDE is larger when the percentage of boats boarded is small.
- We found efficiency increasing as we increase threshold.
- There is a 33% increase in efficiency over OPTIDE (32% success rate vs. 24% success rate) when thresholds are adjusted to achieve 10% of present boarding rate.
- Unlikely USCG would cut boarding rate so significantly.
- Probably need a combination of DE-OPTIDE and a randomized or mixed strategy for boarding.

# Other Approaches

# Other Approaches

The optimal strategy is not necessarily to board the boat with the highest predicted probability of being in violation.  Here are three reasons why:

1. A desire to check on every boat in the fishery at least once a year

2. Consideration of the time it will take to board

3. Search/wait and improve?

# Other Approaches

The optimal strategy is not necessarily to board the boat with the highest predicted probability of being in violation.  Here are three reasons why:

1. **A desire to check on every boat in the fishery at least once a year**

2. Consideration of the time it will take to board

3. Search/wait and improve?

# Checking all Boats

- In order to check all boats at least once a year, at times the Coast Guard must choose to board a boat predicted to have a small chance of being in violation.
- When is a good time to perform such boardings?
- Are there more general concerns about the number of prior boardings? For example, consider:

| | Probability of Violation | Boardings in Past Year |
|---|---|---|
| Boat A | 13% | 2 |
| Boat B | 15% | 6 |

- Might board Boat A

LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY



CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# A Simple Model for Tradeoff between Balanced Deterrence and Violation Yield

- Score S(v) of vessel v should be combination of violation yield y(v) and days since last boarded D(v)
- Let α be a model parameter

$$S(v) = y(v) + \alpha D(v)$$

- If b past boardings, u unsuccessful past boardings, take

$$y(v) = f(b,u) + .05Z$$

- f(b,u) comes from observed data, Z uniformly distributed between -1 and +1

# A Simple Model for Tradeoff between Balanced Deterrence and Violation Yield

- Ran simulations of this model:
  - 5 candidates per day, selected uniformly at random from 100 vessels with highest scores at start of day
  - Can't take top 5 – they may not all be out in area where Coast Guard is checking
- Run model for period of time (e.g. 3 years) with varying values of $\alpha$
- Comparing average number of observed violations over entire period to average number boarded in last period can help compare scoring rubrics

# Other Approaches

The optimal strategy is not necessarily to board the boat with the highest predicted probability of being in violation. Here are three reasons why:

1. A desire to check on every boat in the fishery at least once a year

2. **Consideration of the time it will take to board**

3. Search/wait and improve?

# Alternative Efficiency Measures

Consider the following hypothetical:

| | Probability of Violation | Predicted Time Boarding will Take |
|---|---|---|
| Boat A | 12% | 4 hours |
| Boat B | 15% | 6 hours |

- ***If we measure efficiency as violations per boarding*** (VPB), then the efficiency of boarding Boat A is .12 VPB, and the efficiency of boarding Boat B is .15 VPB. Prefer Boat B.
- ***If we measure efficiency as violations per hour*** (VPH), then efficiency of boarding Boat A is .12/4=.03 VPH, & efficiency of boarding Boat B is .15/6=.025 VPH. Prefer Boat A.
- It should be noted that the best we may be able to do is predict approximate boarding time; there is no way to know it for certain. This further complicates decision-making.

# Other Approaches

The optimal strategy is not necessarily to board the boat with the highest predicted probability of being in violation.  Here are three reasons why:

1. A desire to check on every boat in the fishery at least once a year

2. Consideration of the time it will take to board

3. **Search/wait and improve?**

LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY



CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Will "Better" Ships Appear?

- Let us consider a hypothetical:

| | Probability of Violation | Time Until we can Board |
|---|---|---|
| Boat A | 10% | 0 |
| Boat B | 15% | 1 hour |

- Of course in the real world there is no way to KNOW Boat B will appear in an hour (although it's conceivable we know where Boat B is, but it is an hour away). The best we might be able to do is predict the probability a Boat like Boat B will appear.
- Back to the hypothetical: If we simply want to maximize the probability of finding a violation, and only have time to board one boat a day, and if waiting one hour is acceptable, then it would be best to bypass boat A and board boat B in an hour.

# Will “Better” Ships Appear? (2)

- Let us continue to consider the hypothetical:

| | Probability of Violation | Time Until it Appears |
|---|---|---|
| Boat A | 10% | 0 |
| Boat B | 15% | 1 hour |

- Let us consider violations per hour (VPH) as the criterion when deciding which boat to board.
- 1 hour to board. Boat A: .1/1=.1 VPH. Boat B: .15/2 = . 075 VPH. Prefer Boat A.
- 4 hours to board. Boat A: .1/4=.025 VPH. Boat B: .15/5=. 03 VPH. Prefer Boat B.
- At times the best strategy may be to search or wait for “better” ships, meaning ships our model judges more likely to be in violation.

CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Will "Better" Ships Appear? (3)

- A simple model: Suppose meet a fishing vessel every T minutes. (A simplifying assumption of random rate of encountering a fishing vessel.)
- Immediately decide whether or not to board.
- Suppose yield p varies uniformly from 0 to 1
- Suppose boarding takes time tT
- What value of p should be threshold for boarding?
- Under certain assumptions:

$$p = \frac{(2t+2) - \sqrt{(2t+2)^2 - 4t^2}}{2t}$$

- As boarding time $tT$ increases, the threshold $p$ increases.
- Confirms intuition that the longer boarding takes, the pickier one must be in boarding.
- Need more realistic models for t, T, p

# Will "Better" Ships Appear? (4)

- Making use of analogies from ecology.
- Body of literature: Animals are very efficient users of energy in obtaining food.
- Animals may evolve so as to be efficient in use of energy in foraging
- Animals may develop feeding preferences that maximize their caloric intake per unit time.
- Expected caloric intake corresponds to expected violation found.
- Variety of models to capture these ideas might be relevant to the Coast Guard's problem

# Pure Pursuers & Pure Searchers

- Based on analogies with ecology, we study two types of strategies for Coast Guard Vessels: Pure Pursuers and Pure Searchers.
  - ***Pure Pursuers***: They “know” the entire array of vessels they might board on a given day, and for each know the OPTIDE score and the distance away, the expected “violation score” (1 or 0, or more generally number of violations found) and the expected time required to reach and board/inspect the vessel.
  - For these, we examine some algorithms for deciding which vessels to board.

LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY



CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Pure Pursuers & Pure Searchers

- ***Pure pursuers*** expend little or no energy in moving about in search of food; they wait until food is sighted and then pursue their prey.
- Examples: anolis lizards, kingfishers, many frogs, certain preying mantises, ambush bugs, certain predatory cats, certain owls.

LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY

CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Pure Pursuers & Pure Searchers

- Based on analogies with ecology, we study two types of strategies for Coast Guard Vessels: Pure Pursuers and Pure Searchers.
  - ***Pure Searchers***: They spend a great deal of time searching for vessels to board, but when they encounter them, have to decide whether to board or search for a vessel with a higher probability of having a violation. (In ecology, will a later prey encountered offer more energy/calories?)
  - For these, we examine 2 kinds of strategies for boarding decisions: the patient and the impatient strategies.

LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY



CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Pure Searchers

- ***Patient Strategy***: Wait for the fishing vessel that is most likely to have a violation among the vessels it is possible to encounter (prey that offers highest energy value)
- Our methods aim to provide guidance as to when it is better to be patient or to board the first vessel that has a score exceeding threshold (be patient or attempt to eat the first prey encountered that offers sufficiently high energy value).

# Pure Searchers

- ***Impatient Strategy***: Wait a certain number of encounters while waiting for a vessel that is very likely to have a violation, but then after a while, board the next vessel encountered that has a sufficiently high probability of having a violation (e.g., an OPTIDE score over threshold, even if not very high)
- Our methods aim to provide guidance as to when it is better to board the first vessel that has a score exceeding threshold or adopt the impatient strategy of waiting for "awhile"

# Pure Searchers

- ***Pure searchers***: Spend a great deal of time and energy searching for food, but when food is sighted, very little time on pursuit.
- Examples: warblers, kinglets, titmice, some lizards, many skinks


LPS Modeling Technology Policy
LABORATORY FOR PORT SECURITY



CCICADA
Command, Control, and Interoperability Center for Advanced Data Analysis
A Department of Homeland Security Center of Excellence

# Questions for Ecological Analogy

- How do we calculate the "expected violation score" (expected energy value) E(i) given an OPTIDE (or other) score of i (assessed physical characteristics of the prey)?
- Which of the measures of efficiency used in the ecological literature are most appropriate to investigate?
- Should the Coast Guard adopt a pure pursuer or pure searcher model, or some hybrid?
- If the pure pursuer model, how bring in uncertainty about distribution of vessels and their OPTIDE scores?
- If the pure searcher model, when is the patient strategy best and when is the impatient strategy best?
- If it is the impatient strategy, which variation of that strategy is best?

# Potential Future Work

- Randomization:
  - Makes it much more difficult to outguess law enforcement
  - Smart randomization methods such as PROTECT system developed by the CREATE Center at USC
  - Randomization of goals: hybrid strategies use different goals at different times, at random
- RIPTIDE
  - **Connect additional data:** interactions between market conditions & fishery violations
  - **Improve prediction models:** Better individualized violation behavior models; link additional data sources to improve prediction
  - **Implement a rank learner:** Use ranked sets of boardings as training examples

# Potential Future Work

- DE-OPTIDE
  - Use LASSO-type penalized regression to provide automatic learning; to identify additional important features and eliminate unimportant ones.
  - Provide computing customized software to periodically update the DE-OPTIDE rule
- Beyond Predictive (OPTIDE) Modeling
  - Explore other objectives such as violations found per hour, desire to check each vessel annually
  - Explore models to aid in search strategies
- Comparing Violations
  - Distinguish fisheries and safety violations
  - Consider severity of a violation
  - Develop Violation Scores rather than just counts

# Potential Future Work

- Data Collection and Addition of Features
  - Examine pattern of violation in the set of vessels that were boarded more than once within a very short amount of time (e.g., three days)
  - Connect additional data such as fish prices, weather, “quota reset days”, and others
  - Improve prediction models by using better individualized violation behavior models & additional data sources to improve prediction
  - Updated data set to include additional features, including:
    - Fishery/fisheries for which each vessel is permitted
    - Location of boarding
    - Amount of time per boarding

# Conclusions

- OPTIDE, though based mostly on intuition, does quite well based on the features it uses.
- Both RIPTIDE and DE-OPTIDE improve over OPTIDE, but may require changes in number of Coast Guard vessels patrolling
- Many alternative approaches are needed to formalize all the multitude of goals in fisheries law enforcement.

# Thanks to my Coauthors

- Hans Chalupsky, USC
- Bobby DeMarco, Rutgers
- Ed Hovy, Carnegie-Mellon
- Paul Kantor, Rutgers
- Alisa Matlin, Rutgers
- Priyam Mitra, Rutgers
- Birnur Ozbas, Rutgers
- James Wojtowicz, Rutgers
- Minge Xie, Rutgers

# Additional Thanks

- LCDR Ryan Hamil, US Coast Guard
- Lt. Ryan Kowalske, US Coast Guard
- Andrew Philpot, USC
- William Strawderman, Rutgers