SMART + SUSTAINABLE

FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

International Conference of the Chilean Computer Science Society (CCSS)

A flag system to label boat with suspicious behaviour

A thesis from the Magister en Ciencias de la Ingeniería (MCI-UAI)

Rodrigo Martinez
Sebastián Moreno
FACULTAD DE INGENIERÍA Y CIENCIAS
UNIVERSIDAD ADOLFO IBÁÑEZ

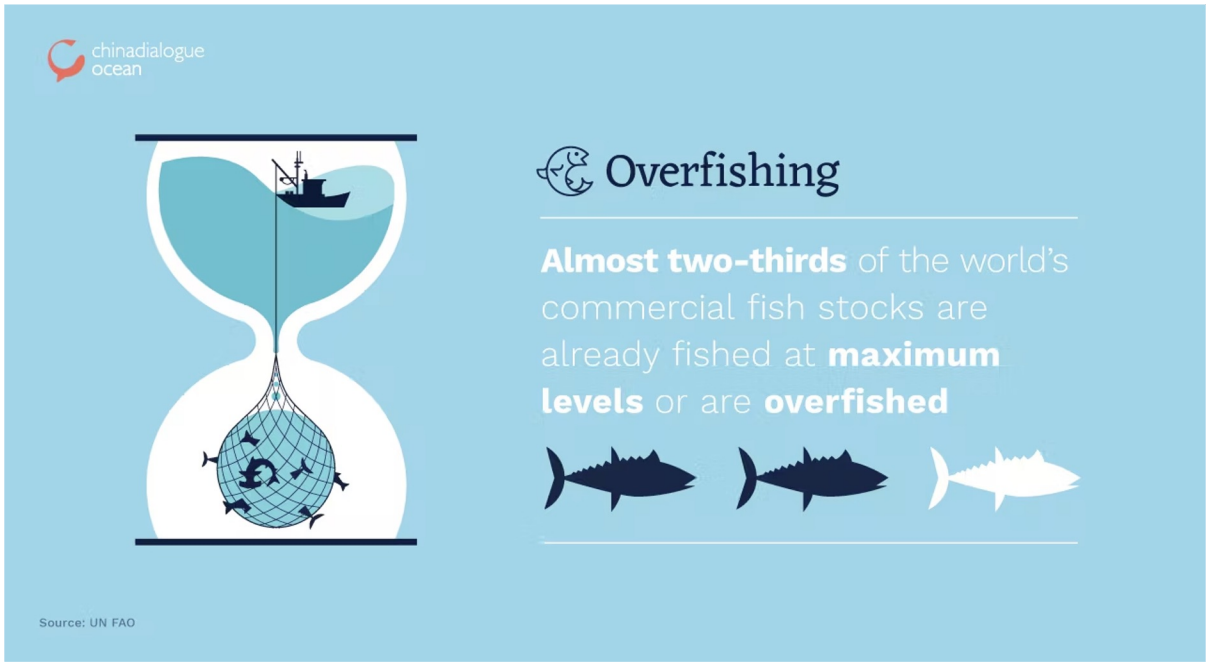
Gob_Lab UAI
UNIVERSIDAD ADOLFO IBÁÑEZ

SERNAPESCA
Ministerio de Economía, Fomento y Turismo
Gobierno de Chile

Sebastián Moreno
Universidad Adolfo Ibáñez

- Ph.D. in computer Science, Purdue University, 2014.
- Associate professor (2022, at UAI since 2015).
- Head of Master of Science in Data Science (2022).
- Associate director postgraduate academic programs (2021)
  - Magister en Ciencias de la Ingeniería
  - Master of Science in Data Science
  - Doctorado en Data Science
- More than 30 conference papers (KDD, ICDM, WWW) and WOS journals.
- Program committee in more than 40 conference including NeurIPS, KDD, WWW, ICDM, and SCCC.
- Research interest: Machine learning, neural networks, relational learning, statistical network analysis.

Problem definition



- Fish consumption per capita increased from 9 kg in 1961 to 20 kg in 2015
- Possible ecological damage to biodiversity
- FAO proposed limiting the extraction of marine resources through fishing quotas

Problem definition

- 26 types of fishes in Chile
  - 11 in overfishing
  - 8 in depletion/collapse
- SERNAPESCA oversees fishing quotas through random inspection, but these are evaded
  - Reports with omissions
  - Change in the type of fish
  - Others
- Consequences
  - Fishing boats prolongate their extraction
  - Negatively affects the marine biodiversity

Problem definition

- SERNAPESCA oversees, on-site, the reported forms
- SERNAPESCA can not be on-site in all places
  - On-site reports
  - Documented reports (not on-site)
- The absence of SERNAPESCA could be used to omit information in documented reports
- SERNAPESCA has an idea of possible fraudulent fishing boats

SERNAPESCA needs to focalize the fiscalization resource to avoid evasions in fishing quotas

Solution

- Create a “flag system” to determine fraudulent ships.

Data exploration

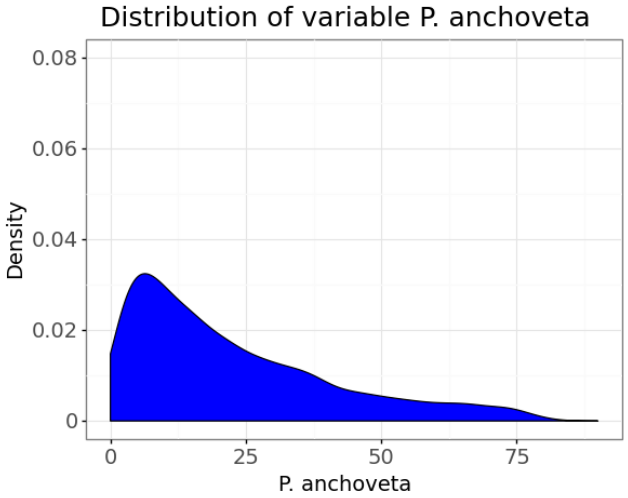
Defining “behavior”

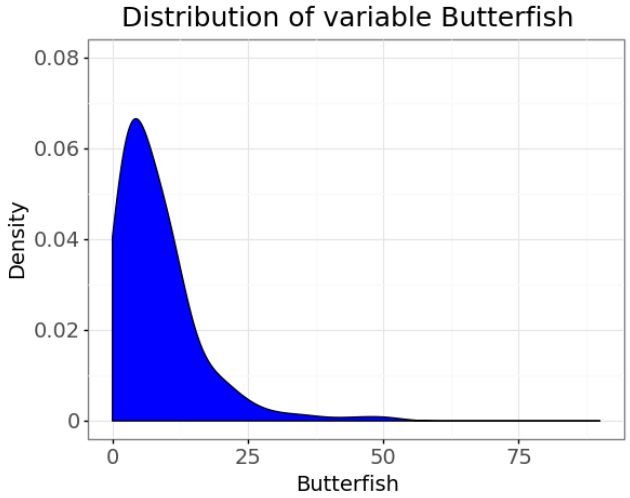
Report with possible fraudulent fishing boats

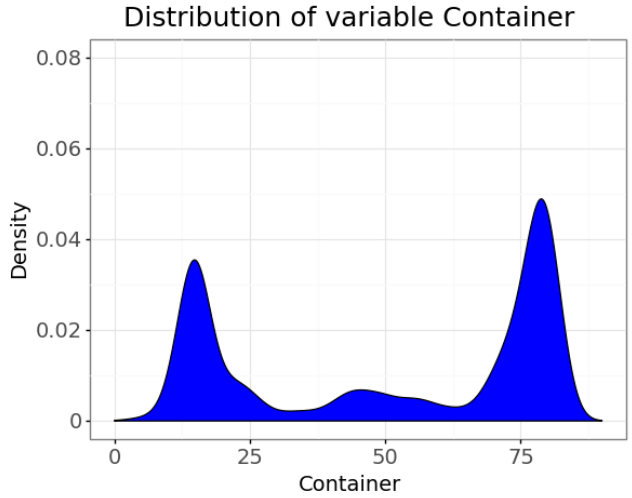
Data analysis

- 2,009 on-site records and 5,952 documented records of artisan fishing boats.
- ID of the boat (380 fishing boats).
- Tonnage for 5 type of fishes
- Tonnage capacity per boat

ID	Fiscalization	P. anchoveta	cod	M. sardinella	butterfish	E. pilchard	Tonnage
10	Presencial	20.5	10.3	0	0	5.2	30 m^3







Statistical analysis

- We applied four statistical analysis to answer the following questions:

1. Does it change the average fishing behavior per fish (on-site vs reported)?
2. Does it change the average fishing behavior per fish and boat (on-site vs reported)?
3. Does it change the average tonnage per fish (on-site vs reported)?
4. Does it change the average tonnage per fish and boat (on-site vs reported)?

Statistical analysis

1. Average fishing behavior per fish (on-site vs documented)

- We compare the general probability of catching each type of fish (on-site vs documented)

$$P(F = f | C = c) = \frac{\sum_{b=1}^{NB} \sum_{i=1}^{ND_b} (\mathbf{1}_c(O_{b,i}) \cdot \mathbf{1}_{\mathbb{Z}^+}(W_{b,f,i}))}{\sum_{b=1}^{NB} \sum_{i=1}^{ND_b} \mathbf{1}_c(O_{b,i})}$$

f: type of fish (P. anchoveta, cod, M. sardinella, butterfish, and, E. pilchard)
c: type of certification (0 or 1)
$O_{b,i}$: i-report for fishing boat b (on-site (1)/documented (0))
$W_{b,f,i}$: tonnage of fish f captured by fishing boat b in its i-report.
Two-proportion Z-test, for H_0: p1=p2, H_1: p1≠p2

Statistical analysis

1. Average fishing behavior per fish (on-site vs documented)

- We compare the general probability of catching each type of fish (on-site vs documented)

Fish	On-site probability	Reported probability	p-value
P. anchoveta	**0.91**	**0.57**	**5.82e-169**
cod	**0.09**	**0.25**	**7.10e-53**
M. sardinella	0.06	0.07	0.13
butterfish	**0.05**	**0.12**	**3.42e-19**
E. pilchard	**0.75**	**0.89**	**3.19e-52**

- Statistical significance decrease for P. anchoveta
- Statistical significance increase for cod, butterfish, and E. pilchard

Statistical analysis

2. Average fishing behavior per fish and boat (on-site vs s documented)

- We compare the probability of catching each type of fish per boat (on-site vs reported)

$$P(F = f | C = c, B = b) = \frac{\sum_{i=1}^{ND_b} (\mathbf{1}_c(O_{b,i}) \cdot \mathbf{1}_{\mathbb{Z}^+}(W_{b,f,i}))}{\sum_{i=1}^{ND_b} \mathbf{1}_c(O_{b,i})}$$

f: type of fish (P. anchoveta, cod, M. sardinella, butterfish, and, E. pilchard)
c: type of certification (0 or 1)
$O_{b,i}$: i-report for fishing boat b (on-site (1)/documented (0))
$W_{b,f,i}$: tonnage of fish f captured by fishing boat b in its i-report.
Two-proportion Z-test, for H_0: p1=p2, H_1: p1≠p2

Statistical analysis

2. Average fishing behavior per fish and boat (on-site vs s documented)

- For each boat we compare the fishing probability for each fish (on-site vs reported).
- We applied 5 * 239 = 1,195 hypothesis test Rejection could imply a behavior change.
- Several boats with 1 or 2 rejections, which could show a fraudulent behavior.
- 30 fishing boat with a high number of rejections.

Test rejected	Fishing boats (α=5%)
0	91
1	73
2	45
3	25
4	5
5	0

Statistical analysis

2. Average fishing behavior per fish and boat (on-site vs documented)

- For each boat we compare the fishing probability for each fish (on-site vs documented).
- Example of a fishing boat with 4 rejections 18 on-site, 26 documented.
- High variability for on-site reports.
- Change of behavior between P. anchoveta and E. pilchard.
- Important increase for M. sardinella and butterfish.

Fish	Average frequency on-site	Average frequency documented
P. anchoveta	51% ± 31%	25% ± 9%
cod	26% ± 25%	26% ± 8%
M. sardinella	0%	31% ± 9%
butterfish	0%	29% ± 7%
E. pilchard	20% ± 15%	54% ± 21%

Statistical analysis

3. Average tonnage behavior per fish (on-site vs documented)

- We compare the general container utilization per type of fish (on-site vs documented)

$$T_{(F=f|C=c)} = \frac{\sum_{b=1}^{NB}\sum_{i=1}^{ND_b} \frac{(\mathbf{1}_c(O_{b,i}) \cdot (W_{b,f,i}))}{B_b} \cdot 100}{NB}$$

f: type of fish (P. anchoveta, cod, M. sardinella, butterfish, and, E. pilchard)
c: type of certification (0 or 1)
$O_{b,i}$: i-report for fishing boat b (on-site (1)/documented (0))
$W_{b,f,i}$: tonnage of fish f captured by fishing boat b in its i-report.
Two-proportion T-test, for H_0: p1=p2, H_1: p1≠p2

Statistical analysis

3. Average tonnage behavior per fish (on-site vs documented)

- We compare the general container utilization per type of fish (on-site vs documented)

Fish	Container utilization on-site	Container utilization documented	p-value
P. anchoveta	69% ± 65%	40% ± 37%	2.708e-65
cod	35% ± 43%	27% ± 34%	0.015
M. sardinella	33% ± 44%	80% ± 83%	6.549e-16
butterfish	15% ± 21%	26% ± 29%	1.511e-05
E. pilchard	44% ± 52%	66% ± 52%	0.997

- Statistical significance decrease for P. anchoveta.
- Statistical significance increase for M. sardinella and butterfish.

Statistical analysis

4. Average tonnage behavior per fish and boat (on-site vs s documented)

- We compare the container utilization per type of fish and boat (on-site vs documented)

$$T_{(F=f|C=c,B=b)} = \sum_{i=1}^{ND_b} \frac{\mathbf{1}_c(O_{b,i}) \cdot W_{b,f,i}}{B_b} \cdot 100$$

f: tipo de pez
c: tipo de certificación
o: variable binaria para indicar si existe información de pesca
w: cantidad de tonelaje del barco b, pez f, viaje i
Z-test de dos proporciones, agrupado para H_0: p1=p2, H_1: p1≠p2

Statistical analysis

4. Average tonnage behavior per fish and boat (on-site vs s documented)

- We compare the container utilization per type of fish and boat (on-site vs documented).
- We applied 5 * 222 = 1,110 hypothesis test Rejection could imply a behavior change.
- Most rejections happen with a single fish.

Test rejected	Fishing boats (α=5%)
0	181
1	38
2	3
3	0
4	0
5	0

Statistical analysis

4. Average tonnage behavior per fish and boat (on-site vs s documented)

- We compare the container utilization per type of fish and boat (on-site vs documented).
- Example of a fishing boat with 2 rejections 20 on-site, 36 documented.
- High variability for on-site fishing.
- Zero/One on-site report for butterfish, cod and M. sardinella, but several documented reports for them.

Fish	Container utilization on-site	Container utilization documented
P. anchoveta	76% ± 45%	NA
cod	12% ± NA	65% ± 9%
M. sardinella	6% ± NA	64% ± 32%
butterfish	NA	53% ± 9%
E. pilchard	65% ± 53%	117% ± 35%

Flag system

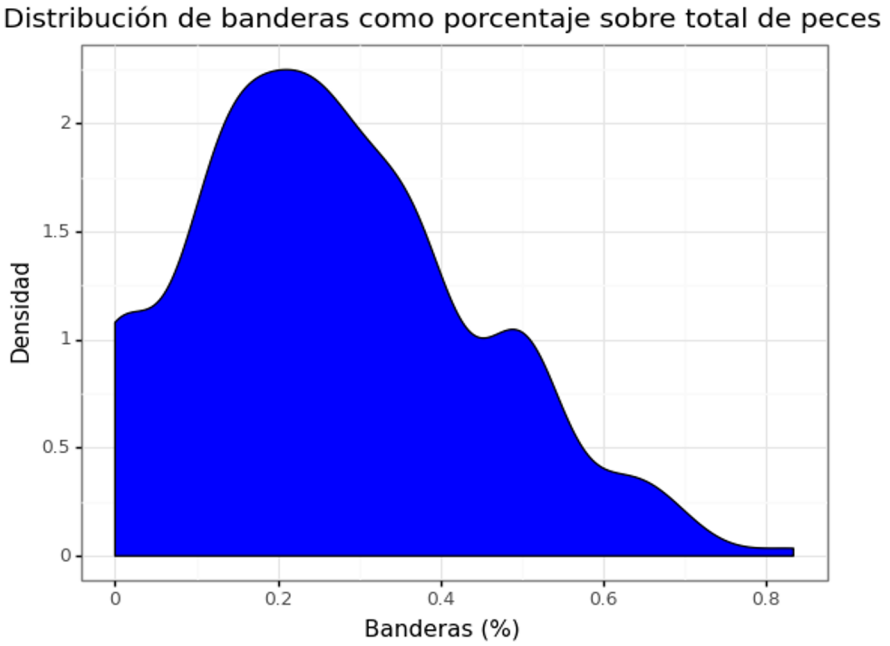
- We applied at most 10 statistical hypothesis tests per fishing boat (fishing and tonnage behavior, per fish). If no data is available we could not apply the test.

- Each test is a “flag”, and each rejection will raise the flag (possible fraudulent behavior).

- Considering that different boat will have different number of possible flags, we calculate the ratio between raise flags/total flags.
  Example:
  Fishing boat 1: 3 raise flag from 9 possible flags => 33%
  Fishing boat 2: 2 raise flag from 3 possible flags => 66%
  Fishing boat 2 has a higher “percentage of fraudulence” than fishing boat 1.

Flag system

- We calculate the “fraudulence percentage” for each fishing boat.
- We observe 3 different groups.
- 0% → No flags raised (no rejections) (29 fishing boats)
- 20% → Low fraudulence percentage (158 fishing boats)
- >=50% → High fraudulence percentage (49 fishing boats)



Flag system

- We calculate the “fraudulence percentage” for each fishing boat.
- Fishing boat with the highest fraudulence percentage.
- Just three type of fishes fraudulence percentage = 83% (5/6 flags).
- Statistical significance decrease for P. anchoveta
- Statistical significance increase for cod

Fish	Prob. Onsite	%Con. Onsite		Prob. Doc.	%Con. Doc.
P. anchoveta	100%	60%		28%	15%
cod	0%	0%		67%	14%
M. sardinella	NA	NA		NA	NA
butterfish	NA	NA		NA	NA
E. pilchard	64%	11%		94%	32%

Conclusions

- We propose a flag system to determine fraudulence of fishing boats. This will help SERNAPESCA to focus their control resources to reduce evasion of fishing quotas.
- We analyzed 7,961 records from 380 fishing boats (5,952 documented and 2,009 on-site), including the captured tonnage of 5 types of fishes and their container capacity.
- Over 2,000 hypothesis tests were performed. The type of certification implies a change in behavior, both in frequency and container use.
- We calculated the “fraudulence percentage” for each fishing boat, where a higher index could show a suspicious behavior.
- A final report is generated with the percentage of rejected flags for each fishing boat

SMART + SUSTAINABLE

FACULTAD DE INGENIERÍA Y CIENCIAS

UNIVERSIDAD ADOLFO IBÁÑEZ

Questions?

Sebastián Moreno

Sebastian.moreno@uai.cl

Gob_Lab UAI

UNIVERSIDAD ADOLFO IBÁÑEZ

SERNAPESCA

Ministerio de Economía, Fomento y Turismo

Gobierno de Chile