MACHINE LEARNING PROCEDURES: AN APPLICATION TO BYCATCH DATA OF THE MARINE TURTLES *CARETTA CARETTA*

Maite Pons[1], Soledad Marroni[2], Irene Machado[3], Badih Ghattas[4] & Andrés Domingo[1]

SUMMARY

The objective of this study is to estimate unreported data on C. caretta bycatch in the Western South Atlantic Ocean between 1998 and 2007. We apply Machine Learning procedures: Classification and Regression Trees, Random Forests, CForest and Support Vector Machines to determine the best method that predicts the total catch of loggerheads turtles by uruguayan longline fleet. Random Forests was the method selected because it presents the minor predictive error rate. This method predicted a total capture of 12.958 loggerheads for the study period which is extremely relevant for this species placed on the IUCN RedList as vulnerable. Machine Learning procedures appear to be useful in the case where access to information is limited, particularly in fisheries where the information of the total captures recorded in logbooks is under-reported or missing altogether.

KEYWORDS

Machine Learning, Pelagic longline, Loggerhead turtles, CPUE

[1] Recursos Pelágicos, Dirección Nacional de Recursos Acuáticos (DINARA), Constituyente 1497, CP 11200, Montevideo, Uruguay, pons.maite@gmail.com; adomingo@dinara.gub.uy

[2] Sección Limnología, Facultad de Ciencias, Montevideo, Uruguay, solemar_11@hotmail.com

[3] Sección Oceanología, Facultad de Ciencias, Montevideo, Uruguay, imachado@fcien.edu.uy

[4] Institut de Mathématiques de Luminy (IML), Marseille, France, ghattas@univmed.fr

INTRODUCTION

Five species of marine turtles occur in the South Western Atlantic Ocean: *Caretta caretta*, *Dermochelys coriacea, Chelonia mydas*, *Lepidochelys olivacea* and *Eretmochelys imbricata*. All of these species are listed in Appendix I of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) and are classified as endangered or critically endangered on The World Conservation Union RedList (IUCN 2007).

One of the main causes of mortality of juveniles and adult turtles is the incidental capture in different fishing gears especially those of pelagic longline (Spotila et al. 2000). In the Western South Atlantic Ocean interaction between marine turtles and pelagic longline has been reported in various studies (Achaval et al. 2000, Kotas et al. 2003, Domingo et al. 2003, 2004, 2005, 2006, Sales et al. 2004, Lopez Mendilaharsu et al. 2007, Giffoni et al. 2008) focusing on the two species that represent the highest capture rates in the region, loggerhead (*C. caretta*) and leatherback (*D. coriacea*).

The Uruguayan longline fleet has been operating in the South Atlantic Ocean since 1981 (Rios et al. 1986) targeting swordfish (*Xiphias gladius*), tunas (*Thunnus spp.*) and some shark species, such as *Prionace glauca* (Domingo et al. 2002). Capture information of target species are recorded in onboard logbooks by the captain, excluding bycatch data.

With the implementation of the “Programa Nacional de Observadores a bordo de la Flota Atunera” (PNOFA) by the Dirección Nacional de Recursos Acuáticos (DINARA) in 1998 (Mora & Domingo 2006) marine turtle bycatch data started to be recorded. Scientific observers collect data on local environmental conditions, details of fishing operations and catch by species (target, bycatch, discards and lost catch). Spanning 1998-2004, this program covers an average of 31% of the total activity of the fleet (Mora & Domingo 2006, Pons et al. 2007). Bycatch and discards that occur in the remainder of the fleet’s vessels, those not observed, are up till now unknown.

Consistent data are required for the effective management of fisheries. Considering the critical status of the populations of marine turtles and the small amount of reliable data available, estimates of this species total catch is of utmost importance.

In biology the use of Machine Learning procedures are principally focused in DNA studies (Cho & Won 2003, Bhardwaj et al. 2005, Díaz-Uriarte & Alvarez de Andrés 2006, Strobl et al. 2007), the information on the use of these procedures in analyzing fishery data is still poor (Watters & Deriso 2000, Tserpes *et al.* 2006, Lennert-Cody & Berk 2007). Lennert-Cody & Berk (2007) implemented Random Forest to determine the unreported data of dolphin bycatch in purse-seine fisheries in the Pacific Ocean; Tserpes et al. (2006) analyzed data from the Greek swordfish fishing fleets in the eastern Mediterranean, by means of machine-learning approaches, in order to define differences in exploitation patterns and fishing strategies; and Watters & Deriso (2000) used regression tree methods to analyze catches per unit of effort from the Japanese longline fishery for bigeye tuna in the central and eastern Pacfic Ocean.

The objective of this study is to estimate unreported data on *C. caretta* by-catch in the fraction of the Uruguayan longline fleet that is not observed in the Western South Atlantic Ocean. We apply Machine learning procedures: Classification And Regression Trees (Breiman et al. 1984), Random Forests (Breiman 2001), CForest (Horton et al. 2006) and Support Vector Machines (Vapnik 1995) to determine the best method that predicts the total catch of loggerheads by uruguayan longliners. We also use Random Forests and CForest to determine which variables are the most important in the prediction of loggerheads bycatch.

MATERIALS AND METHODS

1. Data source

1.1. Observer data

Data was collected by scientific observers of the PNOFA between April 1998 and November 2007. The date, geographic position (latitude and longitude), effort (number of hooks), sea surface temperature (ºC) and catch (number of individuals by specie), have been record for each set. A total of 2.858.513 hooks were observed in 1449 sets between 18º and 41º S and 20º and 54º W (**Fig. 1**).

As there exists two types of fishing gear, we considered two categories: 1) polyamide monofilament and 2) nylon multifilament.

A total of six variables were including in the analysis for the prediction of the CPUE of loggerhead turtles (**Table 1**). CPUE was the response variable, and since it is continuous, we considered regression procedures. Capture Per Unit of Effort (CPUE) is calculated as the number of *C. caretta* per 1000 hooks (ind./1000 hooks).

1.2. Logbook data

For the same period and same area as mentioned above, the data from the fraction of the Uruguayan longline fleet, not observed, were analyzed. The information of the predictor variables and the effort data were obtained from logbooks provide by DINARA. Only the sets with temperature data and geographical position were used. A total of 6.637 sets were analyzed with an effort of 6.652.481 hooks.

2. Data analysis

2.1 Classification And Regression Trees (CART)

CART (Breiman *et al.*, 1984) is a binary splitting method, which partitions recursively the data set into disjoint subgroups, the leafs. It uses two algorithms. The first splits iteratively the data set into two sub-samples according to a binary rule like "temperature < 24°". The splitting rule is based on one of the explanatory variables and on a threshold on this variable. It is chosen as the one that minimizes the heterogeneity of the obtained subsamples; when the output variable is discrete, classification trees are constructed and the criteria used are the "entropy" or the "gini" one, and for continuous outcome, regression trees are constructed using the "deviance" criterion.

The two obtained sub-samples are then partitioned by the same way recursively until there are too few observations (usually five) in the obtained samples (other stopping rules are available). This gives a tree whose terminal nodes' number may be too high. The mean value of the output variable is assigned to each leaf, computed over the observations within the corresponding region.

To avoid overfitting the data when using this tree a pruning algorithm is used to select an optimal sub-tree.

These models have been widely studied in Machine Learning and applied statistics and present many advantages like their representation in form of a binary tree, working in high dimension and variable ranking (Nerini & Ghattas, 2007).

2.2 Random forests (RF)

Random Forests (Breiman, 2001) is an ensemble method which aggregates K trees similar to the ones constructed with CART, each one being grown using a bootstrap sample of the original data set. Each tree in the forest uses at each node only a subset of the explanatory variables (in our case 2, as suggested by Liaw & Wiener (2002)). Besides the trees are not pruned.

The prediction given by a RF is the mean of the predictions given by the K trees in the forest when using regression trees, or the majority vote for classification trees.

2.3 CForest (CF)

This method is also an ensemble method like RF, but the trees constructed are different from CART and follow the conditional inference trees (Horton et al. 2006). These trees optimize the splitting rule using statistical tests in contrast with CART which uses heterogeneity measures.

2.4 Support Vector Machines (SVM)

SVM (Vapnik, 1995) are learning methods for classification and regression. Their concept is based on looking for a linear separator between the classes, which should give a perfect classification, and which should maximize the "margins", that is stay the farthest possible from the classes borders. When the data set is not linearly separable, it is mapped using a non linear function into a higher dimensional space (called feature space) where a linear separation should be possible. When using SVM, the non linear mapping need not to be known explicitly as the Kernel function (scalar product between mapped observations) is sufficient to construct the separating function (Hastie *et. al*. 2001).

3. Model selection

In order to determine the best method that predicts the total catch of loggerheads by uruguayan longliners, the original data are randomly divided into two datasets or samples. The first sample, 2/3 of the original data, is the training set or learning sample and is used to train the models. The second sample, 1/3 of the original data, is the test sample, or the evaluation set, which is used to validate the models. The validating process is developed by calculating the Mean Square Error (MSE), which is used when a continuous response like CPUE is modeled.

This procedure (training and validation) was repeated one-hundred times. Finally we calculated the average of the MSE of each model. The model which obtains the minor predictive error rate is selected for the prediction of CPUE of the unreported loggerheads by-catch data.

4. Variables importance

The RF and CF also provide a measure of variable importance. Importance is derived from the contribution of each variable that is accumulated along all the nodes and all the trees where it was used (Breiman, 2002). The measure of variable importance, in RF, was calculated for the increase in percentage of MSE, and in CF, this measure was calculated for Mean Decrease Accuracy.

5. Software

All analysis was carried out in R free software (R Development Core Team, 2007) using different packages: tree (Ripley, 2007.) for CART, randomForest (Liaw & Wiener, 2007) for RF, party (Hothorn et al., 2007) for CF and e1071 (Dimitriadou et al., 2007) for SVM.

RESULTS

The percentage of annual coverage of PNOFA, in relation to the total effort realized by the longline fleet, varied between 3 and 71% within the study period. The smallest coverage was registered in 2000, while the largest coverage was registered in 2007 (**Table 2**).

1. Machine Learning procedures

RF and CF, preformed similarly, however, RF presented the minor predictive error rate. SVM was the model that presented the major predictive error (**Table 3**). For this reason, the chosen model was RF to predict the un-reported incidental capture of loggerhead sea turtles.

RF predicted a total capture of 10.756 turtles and in addition to those reported by PNOFA observers, totaled 12.958 loggerheads for the period between April 1998 and November 2007. RF explained 46% of the variability of the data. The estimated loggerhead CPUE with their standard deviation can be observed in the **figure 2**, and the f**igure 3** shows this CPUE and the observed by PNOFA. Both curves demonstrate a similar tendency with exception of years 1999, the period 2003-2005 and 2007 where tendencies are reversed. Except for 1998, all the CPUE values were greater than those observed by PNOFA (**Fig. 3**). The biggest differences occurred in 2006 with a observed CPUE of 1.3 ind./1000 hooks and a corresponding estimated value of 3.8 ind./1000 hooks.

2. Variables Importance

The year, temperature, and fishing gear, were the most important variables estimated by RF y CF related with incidental capture of loggerhead (**Fig. 4**), although in different orders. In order to visualize the relationship between these variables and the loggerhead CPUE, we constructed a tree using CART (**Fig. 5**). The tree was not prune.

The multifilament fishing gear presented the smallest mean values of loggerhead CPUE (0,3 ind/1000 anz.) and is shown as the first leaf of the tree. With respect to the monofilament gear, there exists a large range of CPUE values (0,8-13,4 ind/1000 anz.) and includes different associated variables. The largest CPUE values occur where temperatures are greater than 24º C and there exists a large variability between years and months. The geographic position appears as split variables in the last nodes with a division at 35º S and 52º W (**Fig. 5**).

DISCUSSION

There was little difference in the prediction error between the RF and CF. This shows that both methods are useful in the generation of a model for our data set. As results indicated, SVM presented the largest prediction error. In general SVM show the best performance in the majority of cases of classification (Culhane et. al. 2002); however, Díaz Uriarte & Alvarez de Andrés (2006) also obtained in their study a better performance of RF in comparison to other methods (SVM, K Nearest Neighbors) for real and simulated DNA data.

The predicted CPUE by RF follows the general tendencies of those observed by PNOFA. It is not clear what is happening in years where the tendencies were reversed. We did not find that the percent coverage of PNOFA explained this situation.

RF and CF similarly ranked the variables in the importance analysis. Strobl et al. (2007) suggested the use of variable importance of unbiased CF procedure for the evaluation of variable importance if the potential predictors vary in their number of categories or scale level, as in our case. Consequently, the gear appears to be the variable most important in the variability of loggerhead CPUE in the uruguayan longline fleet.

However, we have considered that it is not the gear in itself the major variable that influences loggerhead CPUE, since the properties of the gears do not influence in the catchabilty of sea turtles. The vessels utilizing the two types of gear operate in different zones for the South Western Atlantic Ocean. The vessels using multifilament gear are at sea longer and operate further from the coast in the open ocean (Domingo et al. 2007). For this reason, it is possible that characteristics of the area rather than those of the multifilament gear generate a smaller capture of loggerhead turtles, but we don’t know why it is not reflected in the importance of latitude and longitude. The categorization of spatial variables (latitude and longitude) in zones, with respect to the operational activities and effort distribution, would be useful in determining the impact of the area on turtle bycatch.

The relationship between the largest values of CPUE and temperature can be explained by sea turtle biology. Loggerhead turtles are ectotherms (animals that obtain body temperature from their surroundings), therefore they often prefer warmer waters, and are commonly found near the surface (Bolton & Witherington 2003).

Machine Learning procedures seem to be useful in the case where access to information is limited, for example in fisheries, where information of the total captures recorded in logbooks is under-reported or missing altogether, which often is the case with bycatch data. The application of a prediction model with said characteristics, that estimates the un-reported catch, is very important for the implementation of fisheries management and conservation of species. Our study estimated that the total capture of loggerheads is six times greater than what is currently reported by observers, which is extremely relevant for this species placed which is placed on the IUCN RedList as vulnerable (IUCN, 2007). Also this value is under-estimated since there were sets that were not considered in the study that were set by the Uruguayan fleet during this period.

The quality of reported data is important to assess population status. Machine Learning procedures and the implementation of a well rounded observer program can help produce reliable data. As an example of this, Lennert-Cody and Berk (2007) implemented RF to determine the unreported data of dolphin bycatch in purse-seine fisheries in the Pacific Ocean.

In the most case of fishing management, the evaluations of stocks populations are based on logbooks data. Then, we suggest that Machine Learning approximations can be used to correct logbook information in relation to more reliable sources such as data collected by on-board observers not only for bycatch but also for target species.

BIBLIOGRAPHY

ACHAVAL F., Y.H. Marin & L.C. Barea. 2000. Captura incidental de tortugas marinas con palangre pelágico oceánico en el Atlántico Sud- occidental. En: G. Arena & M. Rey. (Eds.). Captura de grandes peces pelágicos (pez espada y atunes) en el Atlántico Sud-occidental, y su interacción con otras poblaciones. INAPE – PNUD URU/92/003. Pp. 83-88. Montevideo, Uruguay.

BHARDWAJ N., R. E. Langlois, G. Zhao & H. Lu. 2005. Kernel-based machine learning protocol for predicting DNA-binding proteins. *Nucleic Acids Research*, 2005, Vol. 33, No. 20.

BOLTON, A. B., & B. E. Witherington. (2003). Loggerhead Sea Turtles. Washington: Smithsonian Books.Bolton & Witherington 2003.

BREIMAN L., Friedman J.H., Olshen R.A. & C.J Stone. 1984. Classification and Regression Trees. New York: Chapman and Hall. New York.

BREIMAN L. 2001. Random forest. *Machine Learning Journal*. 45, 5-32.

BREIMAN L. 2002. Manual on Setting Up, using, and understanding Random Forests v3.1. http://www.stat.berkeley.edu/users/breiman/RandomForests/cc.home.htm.

CHO S-B. & H-H. Won. 2003. Machine Learning in DNA Microarray Analysis for Cancer Classification. Conferences in Research and Practice in Information Technology, Vol. 19.

CULHANE, A. C. Perriere, G. Considine, E. C. Cotter, T. G. & D. G. Higgins. 2002. Between-group analysis of microarray data. Vol. 18, No. 12, 1600–1608.

DIAZ-URIARTE R. & S. Alvarez de Andrés. 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*. 7:3.

DIMITRIADOU E., K. Hornik, F. Leisch, D. Meyer & A. Weingessel. 2007. e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. [http://cran.r-project.org/src/contrib/Descriptions/e1071.html] [R package version 1.5-17].

DOMINGO, A., O. Mora & M. Cornes. 2002. Evolución de las capturas de elasmobranquios pelágicos en la pesquería de atunes de Uruguay, con énfasis en los tiburones azul *(Prionace glauca)*, moro *(Isurus oxyrinchus)* y porbeagle *(Lamna nasus)*. Col. Vol. Sci. Pap. ICCAT, 54(4): 1406-1420.

DOMINGO A., A. Fallabrino, R. Forselledo & V. Quiricci. 2003. Incidental capture of loggerhead (*Caretta caretta*) and leatherback (*Dermochelys coriacea*) sea turtles in the Uruguayan long-line fishery in Southwest Atlantic. En Seminoff, J.A. (Comp.) 2003. Proceedings of the 22nd Annual Symposium on Sea Turtle Biology and Conservation. NOAA Technical Memorandum NMFS – SEFSC – 503, PP. 308.

DOMINGO A., A. Fallabrino & M. Laporta. 2004. Incidental capture and space-time distribution of sea turtles in the Uruguayan longline fishery in South West Atlantic. En: Abstracts of the 4th World Fisheries Congress, Vancouver, Canada. Pag. 58.

DOMINGO A., L. Bugoni & L Prosdocimi. 2005. Diagnóstico de Situación: Las Pesquerías y su Interacción con las Tortugas Marinas en el océano Atlántico Sur Occidental (ASO), Grupo de Especialistas de Tortugas Marinas del ASO. Montevideo. Uruguay. 22pp.

DOMINGO, A., G. Sales, B. Giffoni, P. Miller, M. Laporta & G. Maurutto. 2006. Distribución y composición de tallas de las tortugas marinas (C*aretta caretta* y *Dermochelys coriacea*) que interactúan con el palangre pelágico en el atlántico sur. Col. Vol. Sci. Pap. ICCAT. 59(3): 992-1002.

DOMINGO A., M. Pons, P. Miller, C. Passadore, O. Mora & G. Pereyra. 2008. Estadísticas del atún aleta amarilla (*Thunnus albacares*) en la pesquería de palangre pelágico de Uruguay (1981-2006). Col. Vol. Sci. Pap. ICCAT. SCRS/2007/122.

GIFFONI B., A. Domingo, G. Sales, F. Niemeyer Fiedler, P. Miller. 2008. Interacción de tortugas marinas (*Caretta caretta y Dermochelys coriacea*) con la pesca de palangre pelágico en el Atlántico Sudoccidental: una perspectiva regional para la conservación. Col. Vol. Sci. Pap. ICCAT. SCRS/2007/168.

HASTIE T.J., Tibshirani R. & J. Friedman. 2001. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction.* Springer.

HOTHORN T., K. Hornik & A. Zeileis. 2006. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15:651-674.

HOTHORN T., K. Hornik & A. Zeileis. 2007. party: A Laboratory for Recursive Part(y)itioning [http://cran.r-project.org/src/contrib/Descriptions/party.html] [R package version 0.9-93].

IUCN, 2007. IUCN 2007 List of Threatened Species. A global species assessment. Available at [http://www.redlist.org]

KOTAS, J. E., Dos Santos, S., de Azevedo, V. G., Gallo & B. M. G., Barata. P. C. R., 2003. Incidental capture of loggerhead (*Caretta caretta*) and leaderback (*Dermochelys coriacea*) sea turtles by the pelagic longline fishery off southern Brazil. *Fish. Bull.* 102:393-399.

LENNERT-CODY C. E. & R. A. Berk. 2007. Statistical learning procedures for monitoring regulatory compliance: an application to fisheries data. *J. R. Statist. Soc. A.* 170 (3) 671-689.

LIAW A. & M. Wiener. 2007. randomForest: Breiman and Cutler's random forests for classification and regression [http://cran.r-project.org/src/contrib/Descriptions/ randomForest.html] [R package version 4.5-22].

LIAW A. & M. Wiener. 2002. Classification and regression by random forest. R News 2/3, 18–22.

LÓPEZ-MENDILAHARSU M., G. Sales, B. Giffoni, P. Miller, F. Niemeyer Fiedler & A. Domingo. 2007. Distribución y composición de tallas de las tortugas marinas (*Caretta caretta* y *Dermochelys coriacea*) que interactúan con el palangre pelágico en el Atlántico Sur. Col. Vol. Sci. Pap. ICCAT. 60(6): 2094-2109.

MORA, O. & A. Domingo. 2006. Informe sobre el Programa de Observadores a bordo de la flota atunera uruguaya (1998-2004). Col. Vol. Sci. Pap. ICCAT, 59(2): 599-607.

NERINI D. & B. Ghattas. 2007. Classifying densities using functional regression trees: Applications in oceanology. *Computational Statistics & Data Analysis*. In press. Doi:10.1016/j.csda.2006.09.028.

PONS M**.**, A. Domingo, O. Mora**,** & G. Pereyra. 2007. Informe del proyecto: "Corrección de la Serie Histórica de datos de Esfuerzo y Capturas de la Flota Atunera Uruguaya (1981-2004)". ICCAT-Departamento de Recursos Pelágicos DINARA. Col. Vol. Sci. Pap. ICCAT. SCRS/2007/156.

R DEVELOPMENT CORE TEAM. 2007. R: A Language and Environment for Statistical Computing. [http://CRAN.R-project.org/]. R Foundation for Statistical Computing, Vienna, Austria.

RIPLEY B. 2007. tree: Classification and regression trees. [http://cran.r-project.org/src/contrib/Descriptions/tree.html] [R package version 1.0-26].

SALES G., B.B. Giffoni, P.C.R. Barata & G. Maurutto. 2004. Interação de tartarugas marinhas com a pesca de espinhel pelágico na costa brasileira – 1999-2003. Resúmenes del la II Reunión Sobre Investigación y Conservación de Tortugas Marinas del Atlántico Sur Occidental-ASO. San Clemente del Tuyú, Argentina.

SPOTILA J. R., R. R. Reina, A. C. Steyermark, P. T. Plotkin & F. V. Paladino. 2000. Pacific leatherback turtles FACE extinction. *Nature*. 405: 529-530.

STROBL C., A. L. boulesteix, A. Zeileis & T. Hothorn. 2007. Bias in random variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*. 8:25.

TSERPES G., D. K. Moutopoulos, P. Peristeraki, G. Katselis & C. Koutsikopoulos. 2006. Study of swordfish fishing dynamics in the eastern Mediterranean by means of machine-learning approaches. *Fisheries Research*. 78. 196-202.

VAPNIK V., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.

WATTERS G. & R. Deriso. 2000. Catch per unit of effort of bigeye tuna: a new analysis with regression trees and simulated annealing. Inter-American Tropical Tuna Commission (IATTC). Vol. 21, 8.

Acknowledgements:

A los observadores científicos del PNOFA, capitanes, marineros y armadores de los barcos. A Mathias Bourel y Caren Barceló.

TABLES

Table 1. Predictors used in the analysis of CPUE loggerhead data.

Variable	Type	Observations
Year	Categorical (10)	Period: 1998-2007
Month	Categorical (12)	January-December
Surface Water Temperature	Continuous	In ºC (range: 9-30º)
Latitude	Continuous	In decimal scale
Longitude	Continuous	In decimal scale
Fishing gear	Categorical (2)	1: monofilament (polyamide) 2: multifilament (nylon)

Table 2. Percentage of the total effort of fleet observed by PNOFA. The effort is in thousands of hooks.

year	Total effort	% observed by PNOFA
1998	567	10,2
1999	553	12,0
2000	386	3,1
2001	520	6,4
2002	564	11,6
2003	1274	33,0
2004	1747	45,2
2005	1934	25,9
2006	1345	33,9
2007	645	70,8

Table 3. Averages of MSE obtained by the model selection procedure. The model which obtains the minor error average was selected (indicated in green).

Procedure	CART	Random Forest	cForest	SVM
MSE	0.01263635	0.01038265	0.01165472	0.17016555

Table 4. Number of turtles captured by the uruguayan longline fleet in the South Western Atlantic Ocean during April 1998 through November 2007.

year	**Nº of turtles estimated**	**Turtles observed by PNOFA**	**Total turtles capture by the fleet**
1998	481	82	563
1999	721	56	777
2000	458	7	465
2001	1150	66	1216
2002	863	95	958
2003	983	163	1146
2004	1074	338	1412
2005	1511	282	1793
2006	3026	261	3228
2007	489	911	1400
total	10756	2261	**12958**

FIGURES

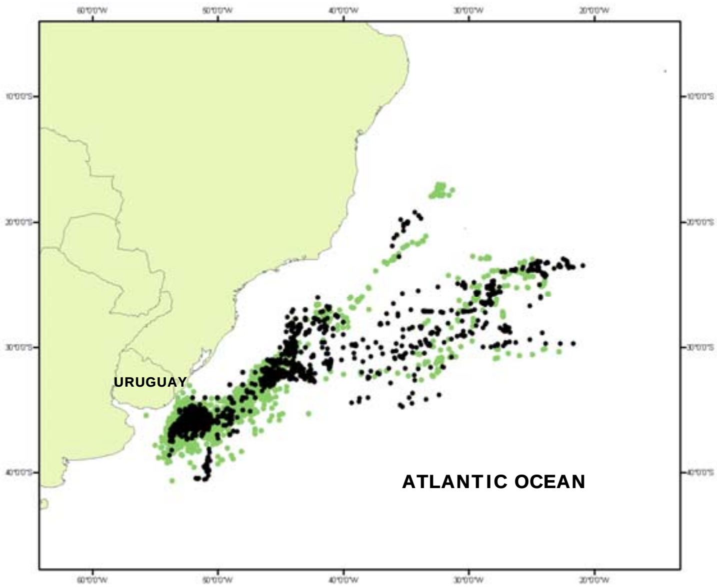

Figure 1. Location of the total sets (green and black dots) realized for the uruguayan longline fleet between April 1998 and November 2007 in South Western Atlantic Ocean. The black dots correspond to the sets observed by PNOFA.

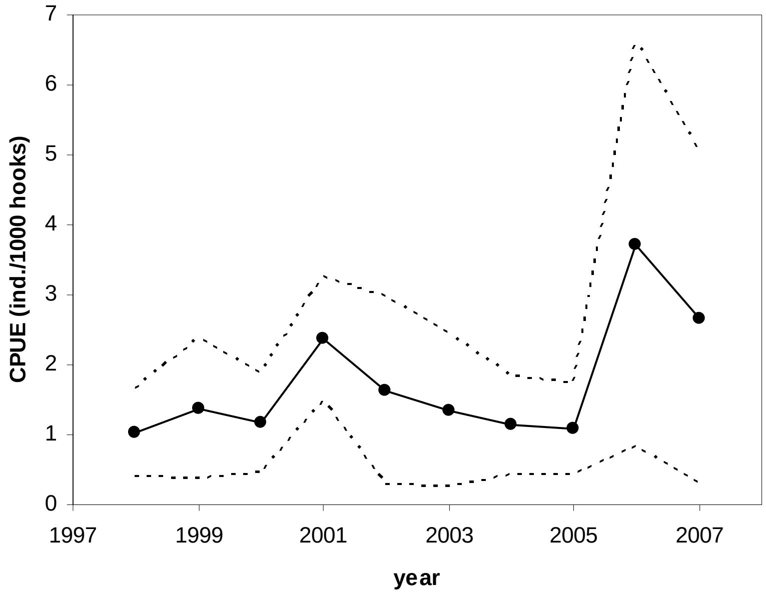

Figure 2. Mean and standard deviation of the loggerheads CPUE predicted by RF.

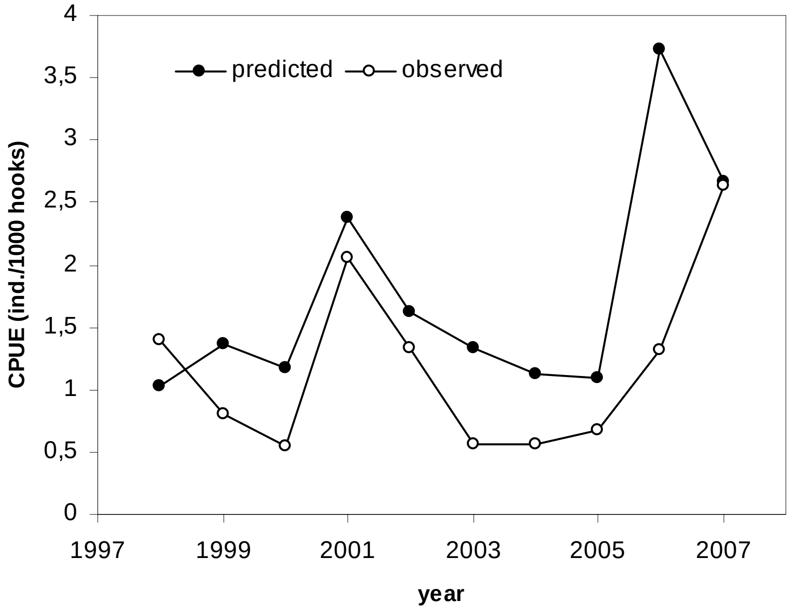

Figure 3. Mean of the loggerheads CPUE (ind./1000 hooks) predicted by RF and observed by PNOFA.

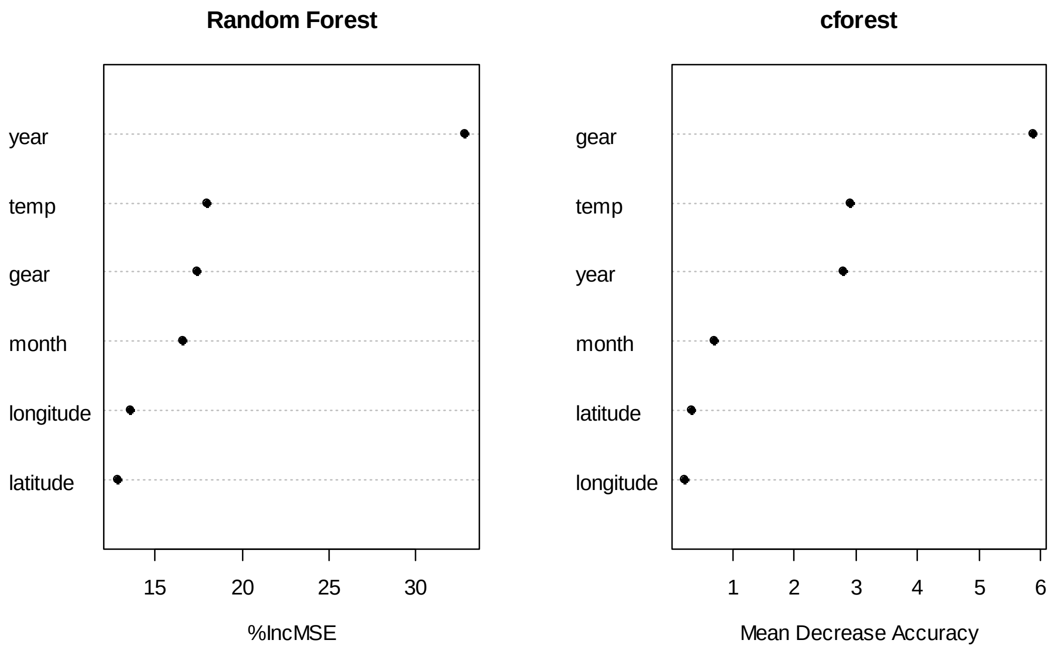

Figure 4. Variable importance plot generated by RF (left) and by CF (right). The ranked variable importance is measured by the increased mean square error (IncMSE) in RF and by the mean decrease accuracy in CF.

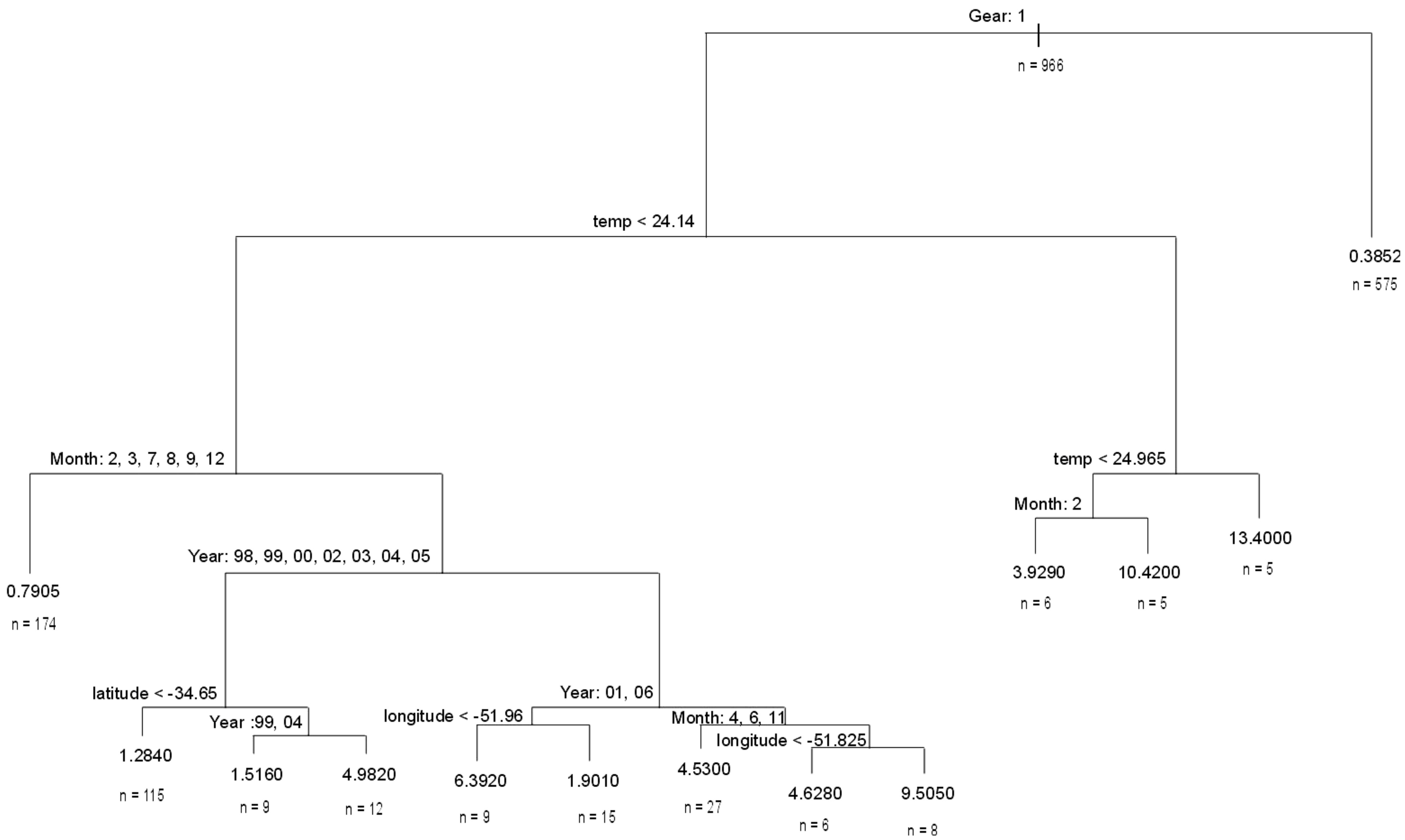

Figure 5. Optimal tree for loggerhead turtles in the South Western Atlantic Ocean from April 1998 to November 2007. For the numerical predictors, data with values of less than the splitting point go to the left daughter node. For the categorical variables, data pertaining to that splitting criterion go to the left daughter node. “n” is the number of data in each node.