Counting visitors in Marine Protected Areas and after? A statistical modelling experiment to estimate the spatial and temporal distributions of recreational coastal activities

Damien Le Guyader, *Terra Maris, France, damien.leguyader@terramaris.fr*
Nicolas Jannic, *Agence française pour la biodiversité (AFB)-Parc Naturel Marin des Estuaires Picards et de la Mer d'Opale (PNM EPMO), France.*
Xavier Harlay, *Agence française pour la biodiversité (AFB)-Parc Naturel Marin des Estuaires Picards et de la Mer d'Opale (PNM EPMO), France.*

Introduction

Sustainable management of marine and coastal environments implies having a realistic and dynamic knowledge of human activities, particularly in marine protected areas (MPAs). Coastal recreational activities are difficult to monitor due to the diversity of practices involved, and to the fact that this population is often highly mobile. Recently, a three years survey on recreational shore fishing has been conducted in French MPAs during European LIFE project. This survey has been extended to other concomitant recreational activities in the "*Parc Naturel Marin des Estuaires Picards et de la Mer d'Opale (PNM EPMO)*" (Meirland et al., 2015). Results of this survey has provided amounts of count data for diverse activities and have drawn precise snapshots of their localisation during the days and for the sites monitored. However, the extraction of a knowledge providing a homogeneous vision (in space and time) of recreational coastal activities from such count data still remain a challenge.
This contribution aim to estimate the distribution over time (by season) and space (by counting site) of leisure activities on the PNM EPMO territory using statistical modelling of counting data.

Material and method

Count data

Counting operations (n = 5,415) were conducted from 15-05-2014 to 12-31-2016 and spread over 36 sites. The database contains 185,533 observations of diverse recreational activities such as recreational shore fishing (for Blue mussels, Brown shrimps, Common cockles, marine worms..), walking and swimming, horse riding, dog walking, kayaking, kite-surfing, wind-surfing, surfing, sailing...

Covariates

The explanatory variables tested (n = 17) are related to meteorological conditions (wind speed and direction, hours of sunlight…), tidal coefficient and height, time dependant variables (month of year, day of the week, holidays …), demographic pressure, or accessibility proxies by car.

Algorithm selection

Eight different algorithms were trained and evaluated for the blue mussels' hand fishing activity. The best model explaining the classified relative density of fishers was the random forest algorithm (Breiman, 2001). This algorithm was selected for modelling the other activities.

Modelling

We performed repeated (n = 3) k-fold (k = 10) cross validations (Hastie et al., 2009) on data divided in a training set (80% of data) and in a test set (20%).
In order to reduce the risk of error in the statistical models, criteria of null variance (or near zero variance) and criteria of multi-collinearity between co-variables (threshold > 0.75) were checked and corrected (Zuur et al., 2010). To account for unbalanced multi-class problems (Wang and Yao, 2012), the class distribution was modified by the up-sampling method.
A random forest model is fitted for each activity (n = 22). Each model aims to minimize the multi-class logloss index.

Models evaluation

The importance of explanatory variables is identified for each model. In addition, each model is evaluated with the Area Under the Curve (AUC) value. The training set AUC ranges from 0.57 for the aquatic hiking activity to 0.9 for the marine plants shore fishing.

Estimates of the spatial and temporal distribution

A matrix (866,448 rows) is created to associate each co-variable (n = 17) with each activity (n = 22), with each site (n = 36) and with each day (from 01/01/2014 to 31/12/2016). Relative abundance classes are then predicted for each activity. The observed median relative abundance for each class and for each activity is assessed and associated with the estimates. Finally, daily abundances and associated median absolute deviation (MAD) are summed by season and then averaged over the three years, for each site and for each activity.

Results

The spatial distribution of abundance for each leisure activity, for each site and for each day of the period between 01/01/2014 and 31/12/2016, is modelled using 22 statistical models. These models can be considered with a good predictive quality (median AUC = 0.75). The mean spatial distribution (by season) of abundance is mapped for each of these activities (*e.g.* Figure 1). A measure of the uncertainty of the results is also proposed.

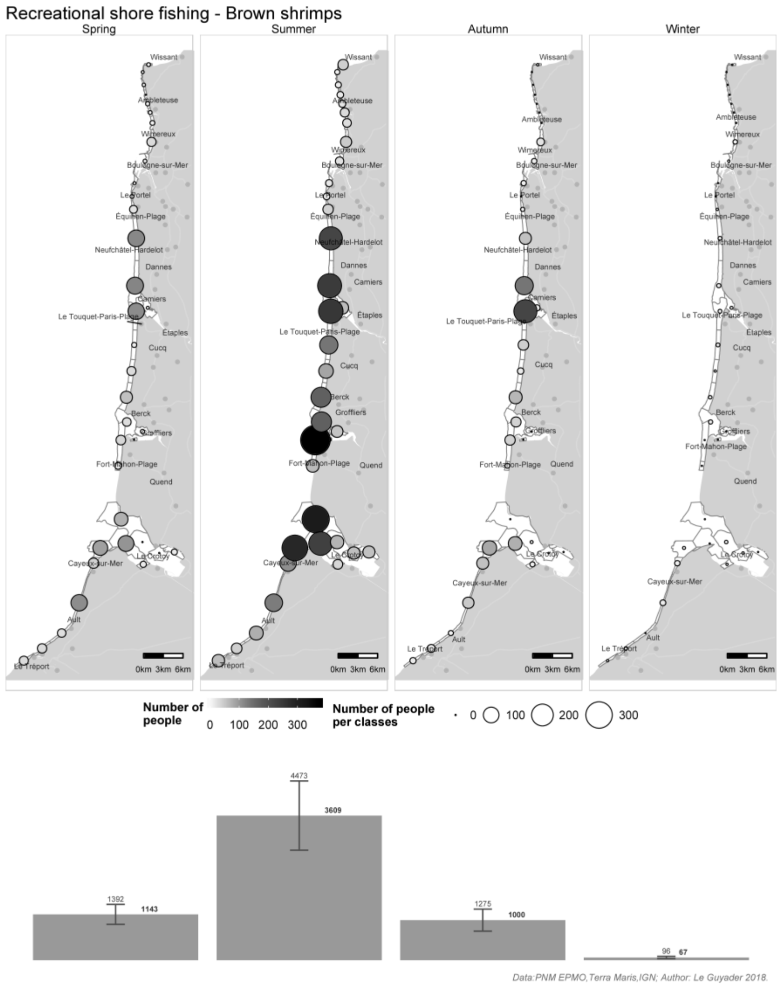

Figure 2.Mean seasonal abundance distribution for the brown shrimps recreational shore fishing (mean sum for 3 years from 01/01/2014 to 31/12/2016).

Conclusion

Results of the study provide estimates of the distribution over time (by season) and space of leisure activities on the PNM EPMO territory using statistical modelling of counting data. Potential contributions and limitations of such approach in providing knowledge on recreational activities in marine protected areas will be discussed.

References

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, Springer Series in Statistics. Springer New York, New York, NY.

Meirland, A., Beck, F., De Cubber, L., 2015. La pêche à pied récréative du Tréport (76) au Cap Gris Nez (62) (Premier rapport de diagnostic du projet life + pêche à pied de loisir). Agence des aires marines protégées.

Wang, S., Yao, X., 2012. Multiclass Imbalance Problems: Analysis and Potential Solutions. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 42, 1119–1130. https://doi.org/10.1109/TSMCB.2012.2187280

Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution 1, 3–14.