The use of jackknifing for the evaluation of geographic profiling reliability

A. Papini, D. K. Rossmo, S. C. Le Comber, R. Verity, M. Stevenson and U. Santosuosso

Abstract

The use of geographic profiling (GP), based on "Rossmo's formula", a technique derived from criminology, has proven to be effective in assessing the origin of invading species of *Caulerpa* in the Mediterranean. The application on *Caulerpa taxifolia* and *C. racemosa* var. *cylindracea* showed the most probable center of spread of the algae. This article discusses a method of assessing the degree of robustness of the results obtained with Rossmo's method using the same species.

To provide an evaluation of the reliability of geographic profiling results we used the jackknife technique; randomly eliminating part of the data set for a given number of replicates (500) in order to analyze the obtained result for each replicate. The results are a series of images with geoprofiling prioritization, each produced with one of the replicates. These images can be summarized in three different ways: (1) OR, depicting all the high probability pixels from the series of replicates; (2) AND, depicting only those high probability pixels present in every replicate; and (3) MEAN, depicting the mean color value for each pixel calculated from all the replicates. We show that jackknifing can be a useful method to increase robustness of GP analysis, both in criminology, epidemiology and biological invasions. Summarizing jackknifing results with the OR logical operator yields highest robustness and worst precision, while the use of the AND operator increases precision but reduces robustness. Using the mean of the pixel values maintains the visualization of the areas of highest priority, while also showing the surrounding area with varying colors, with a meaning analogous to confidence limits.

Introduction

Geographic profiling (GP) is an analytical technique used in criminology, with the aim of calculating the most probable origin of linked crimes, which is usually the offender's home. GP is used by police forces around the world to help focus investigations and prioritize suspects in cases of serial crimes (Rossmo 2012).

GP input consists of spatial data about the locations of linked crimes, which is used to create a probability surface to overlay on the map of interest in the form of a geoprofile (Rossmo 2000). GP does not provide an exact origin location, but, instead, provides a prioritization pattern for investigation based on a descending order of the probability height on the geoprofile (Rossmo 2000).

The model is based on two components: a distance-decay function, such that the probability of a crime (or other events with a localization on a map) decreases with increasing distance from the offender's residence; and a buffer zone, within which the probability increases with distance (Rossmo 2000). The distance-decay function is due to travel costs – both for human criminals and invasive species – (Stevenson *et al.* 2012), in economical or energy terms, respectively. The buffer zone is linked to the avoidance by criminals of locations too close to their residence. In biology, the existence and the extent of the buffer zone should be analyzed case by case. For example, Dramstad (1996), Saville *et al.* (1997), Singh *et al.* (2001) and Stevenson *et al.* (2012) showed evidence of a buffer zone in trees and bees.

GP has been used in biology to analyze the origins of infectious diseases (Le Comber *et al.* 2011, Verity *et al.* 2014), to predict the locations of multiple nest locations of bumble bees (Suzuki-Ohno *et al.* 2010), and to study the patterns of animal foraging (Le Comber *et al.* 2006; Raine *et al.* 2009) and shark hunting (Martin *et al.* 2009).

Stevenson *et al.* (2012) used GP to identify the origin of the invasion of a species, starting from the current known locations of their populations. The places colonized by the invasive populations were considered analogous to crime sites, while the source or sources of the invasion were considered analogous to the criminal's home. The same authors tested GP in comparison to other spatial techniques, such as the center of minimum distance, the spatial mean, the spatial median and a single parameter density model. GP gave better results compared to the other techniques in 52 of the 53 data sets explored for invasive species in Great Britain. Stevenson *et al.* (2012) provided a list of values for the buffer zone radius evaluated as the most appropriate in their analysis. The technique was applied with success on biological invasions of algae (Papini *et al.* 2013) and insects (Cini *et al.* 2014).

Invasive species are considered to be one of the main causes of biodiversity loss (Vitousek *et al.* 1996; Wilcover *et al.* 1998). Invasive species can damage native species through predation and competition, by modifying ecosystem functions and by altering the abiotic environment and by spreading pathogens (Strayer *et al.* 2006; Ricciardi and Cohen 2007 Pimentel et al. 2005).

The invasion of macroalgae, such as some species belonging to genus *Caulerpa* (Caulerpales, Chlorophyta) is one of the main threats to marine natural environments (Meinesz *et al.* 2001). Two species of *Caulerpa* J. V. Lamouroux, *Caulerpa taxifolia* (Vahl) C. Agardh and *C. racemosa* (Forskål) J. Agardh var. *cylindracea* (Sonder) Verlaque, Huisman and Boudouresque, caused severe biological pollution (Piazzi *et al.* 2005) in the Mediterranean. One interesting feature of this invasion is that the origin is known - an accidental release from the aquarium of Monaco in 1984 (Meinesz and Hesse 1991; Meinesz *et al.* 2001; Turan *et al.* 2011). Geographic profiling of the invasive caulerpas spread in the Mediterranean was already used with success by Papini *et al.* (2013), taking advantage of the fact that the spreading origin of *Caulerpa taxifolia* is known, and the related data set is a good starting point for calibrating the technique.

The GP analysis applied to biological invasions has certain limitations: the results obtained with Rossmo's formula are based on the postulate that all spreading events should come from one or a limited number of origins, which is not always true for biological invasions where secondary sites of invasion may frequently derive from the original primary or more independent introductions may occur (Santosuosso and Papini 2016). Furthermore, vegetative propagation and other "slow" ways of environmental spread may obscure the general pattern (Papini *et al.* 2013).

An alternative approach may be a series of geoprofiles using different time periods for the data (e.g., year 1, year 2, year 3, etc.), a technique already used with success in crime analysis (Rossmo and Velarde 2008). Such an approach allows for a better understanding of the spread from secondary sites, reducing the "noise" in the data. This was the approach taken by Stevenson *et al.* (2012), who fitted the parameters of the model using a maximum likelihood approach from a time series. Moreover, almost certainly some of the sites of the invading algae are unknown, making the final result approximate (Papini *et al.* 2013). This is also frequently true in criminology where the accuracy of data used for GP may affect the accuracy of the analysis (Snook *et al.* 2005 and Rossmo 2005).

A possible method to analyze the effect of the errors derived from the limitations linked to the geoprofile is the use of data resampling techniques such as jackknifing or bootstrapping (Miller 1974, Efron 1979, 1972, Efron and Tibshirani 1986, 1993). Both methods are commonly used in other biological analyses (Manly 2006); for example, bootstrapping is commonly used to assess the robustness of phylogenetic analysis (Felsenstein 1985). The two methods are very similar, consisting both in a random deletion of part of the data, with the jackknife using such a reduced data set, while the bootstrap substitutes the deleted data by duplicating some of the remaining data items (Meyer *et al.* 1986).

In this study, we used the jackknife technique (Miller 1974, Efron 1979) to test the robustness of a GP analysis. After van Belle *et al.* (2004), a statistical or an analytical procedure is robust if it performs well when the needed assumptions are not violated "too badly." After the same authors it is not a strictly mathematical definition, but robustness should provide a measure of the confidence limits of the obtained results. Even Umeton et al. (2011) defined robustness as "Robustness is the persistence of a system property respect to perturbations". In the case of the geographic profiling, the jackknife technique should provide an idea of the robustness of the analysis and of the confidence within which we can look for the point of origin of the sites of biological invasion by *C. racemosa* var. *cylindracea.* The assessment of the confidence limits of geographic profiling may be extended to other analyses, including those outside the field of biological invasions.

For assessing robustness, it is possible to create new data sets simply by resampling the observed data. Such an analysis requires to take a series of subsamples from data set a given number of times. Some observations appear once, others twice, others not at all (van Belle *et al.* 2004). In jackknifing, a part of the sample is systematically omitted, for

example by removing one data point at a time, and the analysis is then carried out for each newly constructed subset (Efron 1982; Efron and Tibshirani 1993).

Materials and methods

The model for geoprofiling analysis was described by Rossmo (2000), who compared the use of Manhattan and Euclidean distances, preferring the former to describe criminal movement in urban areas. However, Le Comber *et al.* (2006) and Stevenson *et al.* (2012) suggested that Euclidean distances are more appropriate for animal and plant movements in nature.

The geographic profiling function generates a surface where each pixel has a different priority score indicating the optimal search pattern for the sources of invasive species (Stevenson *et al.* 2012). For each pixel with coordinates (i, j) of the target area, the score function (p) is calculated as follows (Rossmo 2000):

$$p_{ij} = k \sum_{n=1}^{C} [\phi_* / (|x_i - x_n| + |y_j - y_n|)^f + (1 - \phi)(B^{g-f})/(2B - |x_i - x_n| - |y_j - y_n|)^g]$$

where

ϕ_* is equal to 1 if $|x_i - x_n| + |y_j - y_n| > B$; 0 otherwise

In this formula rapresentation it is used Manatthan metric: "$|x_i - x_n| + |y_j - y_n|$" .

For point p with coordinates (i, j), the formula sums the probability across all the locations where the invading organism was found. After Rossmo (2000), Φ functions as a switch that is set to 0 for sites within the buffer zone, and 1 for sites outside the buffer zone. k is an empirically determined constant, which was set to 1 in our study. B is the radius of the buffer zone, and C is the number of events (in this case the reports about the presence in a given locality of the invader). f and g are parameters that control the shape of the distance-decay function on either side of the buffer zone radius. For our analysis we used the same parameters as in Papini *et al.* (2013). The parameters specify the increase in dispersal probability moving away from the source, reaching a maximum value at a distance equal to the radius of the buffer zone. This reflects the reduced probability of dispersal within the buffer zone and the fact that dispersal probability declines with distance (Stevenson *et al.* 2012).

This function produces a search priority surface for the inputted locations on the user-provided map (Rossmo 2000). Rossmo described the equation as a curve, which, when plotted in three dimensions, resembles the shape of a volcano with a caldera. The sum of these 'volcano' shaped decay functions produces a surface describing an optimal search pattern for the location of the origin of the species invasions. To check the robustness of the results we performed a jackknife (Miller 1974) analysis of the data set. We omitted 20% of the observations from the data sets for each of 500 replicates. The number of necessary bootstrap replicates is a controversial issue. We followed the indication by Pattengale *et al.* (2010) proposed for phylogenetic analysis (100-500 replicates). For each replicate, we performed a geoprofiling analysis. The points in the map with highest priority were represented as red pixels (best 5%) and green pixels (best 10%). The final result obtained after the jackknife procedure resulted in a final image where the position of the red pixels and of the green pixels obtained in each replicate were combined with two logical operators (AND, OR) and the mean value of the pixels (see below), using a Python script (jack.py) written by the authors. We used the known invasion origin of *C. taxifolia* (Monaco) to calibrate our model, estimating the critical and map-dependent parameters *B*, *f*, and *g*.

The distribution data for the invasive *Caulerpa* were obtained from Jousson *et al.* (1998) for *C. taxifolia* (since in this case the spreading origin is known) and from Verlaque *et al.* (2003, 2004) and Piazzi *et al.* (2005) for *C. racemosa* var. *cylindracea*.

Our programs were written in Python 2.6.4 (http://www.python.org/) and run on an Ubuntu 9.10 Karmic Koala (http://www.ubuntu.com/) Operating System, Linux kernel 2.6.31. PIL (Python Imaging Library 1.1.6, http://www.pythonware.com/products/pil/) was installed.

The results (Images) were blended with the original map with Python commands (see Software notes in Supplementary material). The Python programs are released under GPL license and available at www.unifi.it/caryologia/PapiniPrograms.html. The maps were downloaded from Open Street Map, available at http://www.openstreetmap.org.

The images obtained with each jackknife replicate were summarized in three ways:

1. OR – the final image shows all the high probability pixels from any of the replicates, resulting in a larger peak probability geoprofile area (corresponding in set theory to the union of the sets of high probability pixels from each replicate).

2. AND – the final image shows only those high probability pixels that were present in every replicate, resulting in a smaller peak probability geoprofile area (corresponding in set theory to the intersection of the sets of high probability pixels from each replicate).

3. MEAN – the final image shows the mean color value for each pixel of coordinates (x, y), based on all the replicates. This last method provides a zone with the RGB color corresponding to the AND zone (the mean of the pixels found in the AND image are the same as they have the same high probability value in every replicate), plus a zone with varying minor priority values obtained from the mean of the different RGB values corresponding to the probability values in each replicate. This second relaxed zone can be considered a fuzzy set (*sensu* Zadeh 1965) of varying priority pixels. This image offers a general representation of the robustness of the analyzed data with respect to the AND or the OR method

In Fig. 1 we show how 3 (example) images deriving from hyphotetical jackknife replicates can be summarized with AND, OR and MEAN methods.

Results

Figure 2 shows a summary of the jackknife images with the OR logical operator applied on the dataset of *C. taxifolia*. As a result the number of red pixels (indicating maximum probability) is quite high, compared to the following figures. Figure 2a is a higher magnification of this zone. Figures 3 and 3a show the result of the AND technique, which showed no red pixels. That is the resampled data sets did not produce overlapping red areas. . Figures 4 and 4a show the results of the MEAN technique. The areas around the peak probability location show degrading RGB values corresponding to varying probability values obtained by averaging the RGB values obtained from the whole jackknifed data set. As a consequence, while there is an area of pure red pixels (RGB values 255,0,0), other shades of red appear as a result of averaging various RGB values in a given point.

Discussion

To provide an evaluation of the geographic profiling results we used the jackknife technique, which involves randomly eliminating part of the data set for a given number of replicates followed by the reanalysis of the remaining data. The results are a series of images with geoprofiling prioritization, each produced with one of the replicates. The images can be summarized in three ways: (1) OR logical operator, showing all the high probability pixels from any of the replicates (the set union); (2) AND logical operator, showing only those high probability pixels present in all the replicates (the set intersection); and (3) MEAN, the mean probability value for each pixel calculated from all the replicates.

The expansion of the red pixels with the OR logical operator is expected; as the robustness of the result increases, the resolution decreases and the precision of the analysis is reduced. This is probably the most common way to integrate the information from a series of bootstrap or jackknife replicates, as seen, for instance, in clinical investigations (Steyerberg *et al.* 2006). The AND logical operator has the opposite effect as it takes into consideration only the red pixels present in all the jackknife replicates. Consequently, the precision of the analysis increases, but not so the robustness. The MEAN method stands somewhere in the middle of the two other methods. The highest priority pixels

found in the AND image are still red (if present), while other pixels prioritized in the OR image are shown in lighter shades of red indicating their lower probability.

All three methods provide useful information, with the OR maximizing robustness and the AND maximizing precision. The MEAN method summarizes the information arising from both the AND and the OR logical operator and provides a more synthetic summary of a jackknife analysis in this context.

In conclusion, jackknifing can be a useful method to increase robustness of GP analysis in criminology, epidemiology and biology. The concept of confidence limits is fundamental for assessing estimates of a parameter (Cox and Hinkley 1973) or a series of parameters, for instance, the phylogenetic reconstruction of the relationships between species (Felsenstein 1985). While in phylogenetic analysis the confidence limits of a phylogenetic reconstruction is expressed as a percentage above branches of phylogenetic trees (for instance Fesenstein 1985 and Simeone et al. 2016), as the results of geoprofiling are images, the confidence limits must be represented on the image itself. Summarizing jackknife results with the OR logical operator yields the highest robustness and the worst precision, increasing the number of highlighted pixels in the image, while the use of the AND operator increases precision but reduces the number of highlighted pixels in the image and hence decreases robustness. Using the mean of the pixel values maintains the visualization of the areas of highest priority, while also displaying the surrounding area using different colors, providing information analogous to confidence limits.

Acknowledgements: This work was supported by the Antonino Caponnetto Foundation.

References

Cini A., Anfora G., Escudero-Colomar L.A., Grassi A., Santosuosso U., Seljak G. Papini A. (2014) Tracking the invasion of the alien fruit pest *Drosophila suzukii* in Europe. Journal of Pest Science 87(4): 559-566.

Cox, D. R. and Hinkley, D. V. (1974) Theoretical Statistics. London: Chapman & Hall.

Dramstad W (1996) Do bumblebees (Hymenoptera: Apidae) really forage close to their nests? J. Insect Behav. 9:163–182.

Efron B (1979) Bootstrap methods: another look at the jackknife. Annals of Statistics 7:1-26.

Efron B (1982) The jackknife, bootstrap and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia.

215 Efron B and Tibshirani R (1986) The bootstrap (with discussion). Statistical Science 1: 54-77.

216 Efron B and Tibshirani R (1993) An introduction to the bootstrap. Chapman & Hall, London.

217 Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791.

218 Hill D, Coquillard P, Vaugelas J, Meinesz A (1998) An algorithmic model for invasive species: application to Caulerpa
219 taxifolia (Vahl) C. Agardh development in the north-western Mediterranean Sea. Ecological Modelling 109:251–
220 265.

221 Jousson O, Pawlowski J, Zaninetti L, Meinesz A, Boudouresque CF (1998) Molecular evidence for the aquarium origin
222 of the green alga *Caulerpa taxifolia* introduced to the Mediterranean Sea. Marine Ecology Progress Series
223 172:275-280.

224 Jousson O, Pawlowski J, Zaninetti L, Zechman FW, Dini F, Di Giuseppe G, Woodfield R, Millar A, Meinesz A (2000)
225 Invasive alga reaches California. Nature 408:157–158.

226 Le Comber SC, Nicholls B, Rossmo DK, Racey PA (2006) Geographic profiling and animal foraging. J. Theor. Biol.
227 240:233–240.

228 Le Comber SC, Rossmo DK, Hassan AN, Fuller DO, Beier JC (2011) Geographic profiling as a novel spatial tool for
229 targeting infectious disease control. International Journal of Health Geographics 10:35.

230 Manly, B. FJ. (2006) *Randomization, bootstrap and Monte Carlo methods in biology*. Vol. 70. CRC Press.

231 Martin RA, Rossmo DK, Hammerschlag N (2009) Hunting patterns and geographic profiling of white shark predation.
232 J. Zool. 279:111–118.

233 Meinesz A, Hesse B (1991) Introduction et invasion de l'algue tropicale *Caulerpa taxifolia* en Méditerranée nord-
234 occidentale. Oceanologica Acta 14:415–426.

235 Meinesz A, Belsher T, Thibaut T *et al.* (2001) The introduced green alga *Caulerpa taxifolia* continues to spread in the
236 Mediterranean. Biological Invasions 3:201–210.

237 Meinesz A (1992) Modes de dissémination de l'algue *Caulerpa taxifolia* introduite en Méditerranée. Rapport de la
238 Commission Internationale sur la Mer Méditerranée 33:44.

239 Meinesz A, Hesse B (1991) Introduction et invasion de l'algue tropicale *Caulerpa taxifolia* en Mediterranée Nord
240 occidentale. Oceanologica Acta 14(4):415-426.

241 Meyer JS, Ingersoll CG, McDonald LL, Boyce MS (1986) Estimating Uncertainty in Population Growth Rates:
242 Jackknife vs. Bootstrap Techniques. *Ecology* 67(5): 1156-1166.

Miller RT (1974) The jackknife-a review. Biometrika 61(1):1-15.

Papini, A., Mosti, S. & Santosuosso, U. (2013). Tracking the origin of the invading *Caulerpa* (Caulerpales, Chlorophyta) with geographic profiling, a criminological technique for a killer alga. *Biological Invasions*, 15(7): 1613-1621.

Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME and Stamatakis A (2010) How many bootstrap replicates are necessary? Journal of Computational Biology 17(3): 337-354.

Piazzi L, Meinesz A, Verlaque M *et al.* (2005) Invasion of *Caulerpa racemosa* var. *cylindracea* (Caulerpales, Chlorophyta) in the Mediterranean Sea: an assessment of the spread. Cryptogamie, Algologie 26:189-202.

Pimentel, D., Zuniga, R., & Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States.*Ecological economics*,*52*(3), 273-288.

Raine NE, Rossmo DK, Le Comber SC (2009) Geographic profiling applied to testing models of bumble-bee foraging. J. R. Soc. Interface 6:307–319.

Relini G, Relini M, Torchia G (2000) The role of fishing gear in the spreading of allochthonous species: the case of *Caulerpa taxifolia* in the Ligurian Sea. ICES Journal of Marine Science 57:1421–1427.

Rossmo DK (2000) Geographic profiling. CRC Press, Boca Raton, FL.

Rossmo, DK (2005) Geographic heuristics or shortcuts to failure?: response to Snook et al. *Applied Cognitive Psychology* 19(5): 651-654.

Rossmo, D. K. (2012). Recent developments in geographic profiling. Policing, 6(2), 144-150.

Rossmo, DK, Velarde, L (2008). Geographic profiling analysis: Principles, methods, and applications. In S Chainey & L Tompson (Eds.), *Crime mapping case studies: Practice and research* (pp. 35-43). Chichester: John Wiley & Sons.

Sant N, Delgado O, Rodriguez-Prieto C, Ballesteros E (1996) The spreading of the introduced seaweed *Caulerpa taxifolia* (Vahl) C. Agardh in the Mediterranean Sea: testing the boat transportation hypothesis. Botanica Marina 39:427–430.

Santosuosso U, Papini A. (2016) Methods for geographic profiling of biological invasions with multiple origin sites. International Journal of Environmental Science and Technology, 13(8): 2037-2044. 10.1007/s13762-016-1032-1.

Saville NM, Dramstad WE, Fry GLA, Corbet SA (1997) Bumblebee movement in a fragmented agricultural landscape. Agric. Ecosyst. Environ. 61:145–154.

Siddall ME (1995) Another monophyly index: revisiting the jack-knife. Cladistics 11: 33–56.

Simeone MC, Grimm GW, Papini A, Vessella F, Cardoni S, Tordoni E, Piredda R, Franc A, Denk T (2016) Plastome data reveal multiple geographic origins of *Quercus* Group Ilex. Peer Journal 4:e1897. doi: 10.7717/peerj.1897.

Singh HP, Batish DR, Kohli RK (2001) Allelopathy in agroecosystems. An overview. J. Crop Prod. 4:1–41.

Snook, B., Zito, M., Bennell, C., & Taylor, P. J. (2005). On the complexity and accuracy of geographic profiling strategies. *Journal of Quantitative Criminology, 21*, 1-26.

Sonder G (1845) Nova algarum genera et species, quas in itinere ad oras occidentales Novae Hollandiae, collegit L. Priess, Ph. Dr. Botanische Zeitung 3:49–57.

Steyerberg EW, Neville BA, Koppert LB, *et al.* (2006) Surgical mortality in patients with esophageal cancer: development and validation of a simple risk score. *J Clin Oncol 24:4277-84.*

Stevenson MD, Rossmo DK, Knell RJ, Le Comber SC (2012) Geographic profiling as a novel spatial tool for targeting the control of invasive species. Ecography 35:1–12.

Strayer DL, Eviner VT, Jeschke JM, Pace ML (2006) Understanding the long-term effects of species invasions. Trends Ecol. Evol. 21:645–651.

Suzuki-Ohno Y, Inoue MN, Ohno K (2010) Applying geographic profiling used in the field of criminology for predicting the nest locations of bumble bees. Journal of Theoretical Biology 265:211–217.

Umeton, R., Stracquadanio, G., Sorathiya, A., Papini, A., Lio`, P., Nicosia, G. (2011) Design of robust metabolic pathways. In: DAC 2011 - Proceedings of the 48th Design Automation Conference, DAC 2011, San Diego, CA, USA, June 5-9, 2011, ACM (2011) 747–752. Isbn: 978-1-4503-0636-2.

van Belle G, Fisher LD, Heagerty PJ and TS Lumley (2004) Biostatistics: A methodology for the Health Sciences. Second Edition. John Wiley and Sons.

Verity R, Stevenson MD, Rossmo DK, Nichols RA and Le Comber SC (2014) Spatial targeting of infectious disease control: identifying multiple unknown sources. Methods in Ecology and Evolution 5.7, 647-655

Verlaque M, Durand C, Huisman JM, Boudouresque CF, Le Parco Y (2003) On the identity and origin of the Mediterranean invasive *Caulerpa racemosa* (Caulerpales, Chlorophyta). European Journal of Phycology 38:325–339.

Verlaque M, Carrillo JA, Gil-Rodriguez MC, Durand C, Boudouresque CF, Le Parco Y (2004) Blitzkrieg in a marine invasion: *Caulerpa racemosa* var. *cylindracea* (Bryopsidales, Chlorophyta) reaches the Canary Islands (north-east Atlantic). Biological Invasions 6:269–281.

Vitousek P, D'Antonio CM, Loope L, Westbrooks R (1996) Biological invasions as global environmental change. Amer. Sci. 84:468–478.

Wilcover DS, Rothstein D, Dubow J, Phillips A, Losos E (1998) Quantifying threats to imperiled species in the United States. Bioscience 48:607–615.

Zadeh L. A. (1965) "Fuzzy sets." Information and Control 8 (3) 338–353.

Figure Captions

Figure 1 - Three (example) images deriving from hypothetical jackknife replicates can be summarized with AND, OR, MEAN and even MODE methods. 1a, 1b, 1c represent three hypothetical different areas of probability (only red pixels reported on figures) obtained from three GP analysis on three datasets derived from Jackknife. 1d: OR summarization: all points that were red in just at least one of the three starting figures are red also in the final figure. 1e: MEAN summarization: in each point of the final figure, the color will correspond to the mean values found in each of the starting figures. Only where the red pixels are present in a position (x, y) in all the jackknife-obtained figures (here three of three), there will be plain red even in the final figure. 1f: AND summarization: Only where the red pixels are present in a position (x, y) in all the jackknife-obtained figures (here three of three), there will be plain red even in the final figure. In the rest of the final image, only white pixels can be found.

Figure 3 – OR image, showing all the high probability pixels from the replicates (corresponding to the union of the high probability sets). Figure 3a – Higher magnification of the peak probability area.

Figure 4 – AND image, showing only those high probability pixels that were present in every replicate (corresponding to the intersection of the high probability sets). Figure 4a – Higher magnification of the peak probability area. No red pixel is present, that is the condtion of presence in all the images is not respected.

Figure 5 – Mean image, showing the mean color value for each pixel (corresponding to the AND image zone, plus a varying RGB color zone corresponding to pixels found with high probability in only some replicates). Figure 5a – Higher magnification of the peak probability area.