

# Spatial prediction of coral reef habitats: integrating ecology with spatial modeling and remote sensing

J. R. Garza-Pérez<sup>1</sup>, A. Lehmann<sup>2</sup>, J. E. Arias-González<sup>1,\*</sup>

<sup>1</sup>Coral Reef Ecosystems Ecology Laboratory, Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional Unidad Mérida, Carr. Ant. Progreso km 6 Cordemex, Mérida, Yucatán 97210, México

<sup>2</sup>Swiss Center for the Cartography of Fauna, Terreaux 14, 2000 Neuchâtel, Switzerland

**ABSTRACT:** Spatial prediction of coral reef habitats and coral reef community components was approached on the basis of the 'predict first, classify later' paradigm. Individual community components (biotic and geomorphologic bottom features) were first predicted and then classified into composite habitats. This approach differs from widely applied methods of direct classification based on remote sensing only. *In situ* coral reef community-condition assessment was first used to measure a response variable (percentage cover of habitat). Reef bottom features (topographic complexity, sand-sediment, rock-calcareous pavement and rubble) were then predicted using generalized additive models (GAMs) applied to continuous environmental maps, high-resolution Ikonos satellite images and a reef digital topographic model (DTM). Next, using GAMs on newly created bottom maps, models were fitted to predict coral community components (hard coral, sea-grass, algae, octocorals). At this stage, high-resolution maps of the geomorphologic and biotic components of the coral reef community at an experimental site (Akumal Reef in the Mexican Caribbean) were produced. Coral reef habitat maps were derived using GIS following a hierarchical classification procedure, and the resulting merged map depicting 8 habitats was compared against thematic maps created by traditional supervised classification. This general approach sets a baseline for future studies involving more complex spatial and ecological predictions on coral reefs.

**KEY WORDS:** GAM · GRASP · GIS · Ikonos · DTM · Mexican Caribbean

—Resale or republication not permitted without written consent of the publisher—

## INTRODUCTION

Habitat mapping and community characterization is a prerequisite for condition and resources assessment of any ecosystem. Improved accuracy and higher resolution of output maps are necessary for the optimization of time and monetary resources invested in field research. Prediction of habitat distribution is a powerful tool to help understand ecological processes, and constitutes a baseline for applied research oriented towards policy-resource management.

The current applicability and generalization of remote sensing (RS) research in coral reef habitat mapping remains problematic due to the intrinsic differences in each studied reef. The choice among several data sources available (e.g. Landsat; SPOT — Satellite Pour l'Observation de la Terre; Ikonos; CASI —

Compact Airborne Spectrographic Imager), all of which offer different degrees of accuracy, resolution and spectral characteristics, creates additional disparity. Finally, because of the use of various different habitat classification methods and mapping techniques, it is difficult to make comparisons between approaches, and this makes each mapping of coral reef habitats unique (i.e. Mumby et al. 1997, Chauvaud et al. 1998, 1998, Andrefoüet et al. 2000, 2003, Mumby & Edwards 2002, Purkis et al. 2002).

Spatial prediction models are static and probabilistic, since they statistically relate the geographical distribution of species or communities to their environment. The quantification of such species-environment relationships represents the core of predictive geographical modeling in ecology, and the models are based on hypotheses of environmental factors control-

\*Corresponding author. Email: earias@mda.cinvestav.mx

ling spatial distribution of species and/or communities (Guisan & Zimmerman 2000).

Spatial prediction modeling has been traditionally a land-based research field (e.g. Leathwick et al. 1998, Tappeiner et al. 1998, 2001, Austin 1999, Debinsky et al. 1999, Bio 2000, Heegard et al. 2001), along with GIS applications for environmental managing (Fedra 1998, Stanbury & Starr 1999, Theobald et al. 2000, Woodhouse et al. 2000), with only a few cases applied to marine resources (e.g. Basu & Nalamotu 1997, Bushing 1997). Spatial modeling of underwater communities is slowly being accepted, with recent examples on distribution of submerged vegetation in freshwater ecosystems (e.g. Janauer 1997, Lehmann & Lachavanne 1997, Lehmann et al. 1997, Schmieder 1997, Heegard et al. 2001, Lehmann 1998) and estuarine fishes distribution (Stoner et al. 2001). More examples can be found in terrestrial ecosystems (e.g. Münier et al. 2001, Tappeiner et al. 2001, Aspinall 2002, Cawsey et al. 2002, Ferrier et al. 2002). All these applications have given rise to the development of new techniques that incorporate geographic data-base management, spatial modeling and prediction, based on innovative statistical techniques such as generalized linear models (GLMs: Chambers & Hastie 1993) and generalized additive models (GAMs: Hastie & Tibishirani 1990).

GAMs are non-parametric extensions of GLMs, which originated as a generalization of the classical least square relation (LSR) (Yee & Mitchell 1991). The use of GAMs has proven to be a very useful tool in ecology (Guisan & Zimmermann 2000, Austin 2002, Guisan et al. 2002, Lehmann et al. 2002a). Lehmann et al. (2002b) have developed the Generalized Regression Analyses and Spatial Prediction (GRASP) concept, and its implementation encapsulates this general approach. With GRASP, Lehmann et al. (2002a) were able to predict the pre-human distribution of fern biodiversity in the whole territory of New Zealand, by estimating the relationships between fern distributions and the environment, using sites largely free from human disturbance. The main advantage of the use of GAMs for this type of study is the ability of the models to fit the data, by estimating the response curve with a smoothing function (non-parametric). This allows the description of environmental gradients in better agreement with ecological theory, and the extension of classical regression fitting to other distributions

(binomial, Poisson, gamma) (Hastie & Tibishirani 1990, Yee & Mitchell 1991, Austin 1999, 2002).

The use of these tools (habitat classification, spatial modelling and remote sensing) is now rapidly growing in coral reef studies. Nevertheless, to our knowledge this study is the first attempt to apply the GRASP methodology to predicting habitats in coral reefs. In this paper we also apply the paradigm 'predict first, classify later' proposed by Overton et al. (2000), incorporating a modification of the coral reef habitat-classification methodology proposed by González-Gándara et al. (1999).

## MATERIALS AND METHODS

**Study area.** Akumal Reef is situated in the Northern-Central portion of Quintana Roo State, Mexico, on the east coast of the Yucatan Peninsula. It is a very well developed fringing reef, with diverse habitats. Coral patches, sea grass beds and algae prairies can be found in the back reef lagoon. A well-delineated reef crest, along with calcareous pavement transition zones, spur and groove systems and sand channels is observed in the fore reef, down to an approximate depth of 40 m. We defined the extent of our study area by a  $12 \times 1$  km strip between  $20^\circ 25.995$  N,  $87^\circ 17.310$  W and  $20^\circ 20.005$  N to  $87^\circ 20.917$  W. The reef structures present in this area have a relatively homogeneous geomorphology (Fig. 1). For the definition of the 54 sampling stations, we relied on expert knowledge of the zone and a high-resolution ( $4 \times 4$  m pixel resolution) Ikonos image (Space Imaging), in order to acquire the geographic coordinates of polygons representing

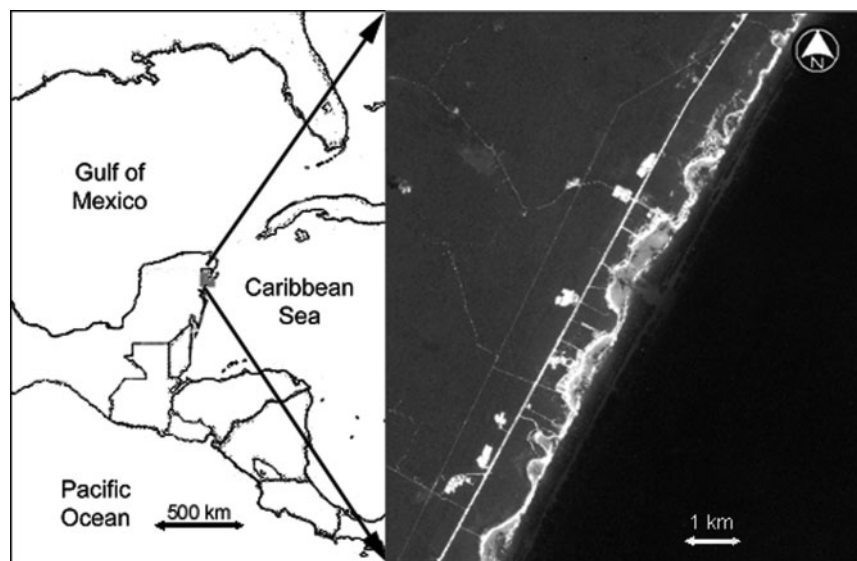


Fig. 1. Study area, Akumal Reef on the coast of Quintana Roo state in Mexico

different substrata types in all the reef zones (reef lagoon, backreef, crest, fore-reef transition, first step and second step).

**Coral reef community characterization.** Characterization and assessment of the benthic communities at each station was performed using a modified version of the Aronson & Swanson (1997) video transect method. We recorded one  $50 \times 0.6$  m transect with a Hi8 Handicam (Sony CCDTR-4000) inside an underwater housing (StingRay, Light and Motion). In each transect, temperature ( $^{\circ}\text{C}$ ), salinity, dissolved oxygen (YSI multi-analyzer, Model 85) and topographic complexity (chain transect) were also recorded. Field work was completed with the collaboration of personnel from the Coral Reef Ecosystems Ecology Laboratory at Centro de Investigación y Estudios Avanzados, Merida.

**Continuous environmental maps creation.** A digital topographic model (DTM) was created at  $4 \times 4$  m resolution from 6752 echo-sounding and satellite image data points, using kriging extrapolation methods in Surfer 8.0 (Golden Software), triangulated irregular networks (TIN) and grid functions in 3D analyst extension for ArcGIS 8.1.2 (Environmental Systems Research Institute). This DTM was exported to ArcView 3.3 (ESRI) for the depth-map query.

The high-resolution Ikonos image (composite image of blue [B], green [G] and red [R] channels) was treated to obtain a depth invariant index with the method proposed by Green et al. (2000), in order to refine the radiometric and visual information extracted from the image once the water column effect was homogenized with it.

The radiometric information of the polygons corresponding to the 53 stations in the study area was extracted from the 4 bands using MultiSpec v.3 software (for MacOS X, Laboratory for Applications of Remote Sensing, Purdue University). Each station was positioned in the image according to the GPS coordinates recorded in the field at the start point of each transect. During the dives, a compass bearing of each transect was recorded, in order to be able to recreate the area extension of the transects within the satellite images (~12 to 15,  $4 \times 4$  m pixels). Another 7 'dummy' stations were defined and included, using expert knowledge and the Ikonos image, in order to have radiometric information on 100% of sand cover and 100% sea-grass bed areas.

In order to obtain the spatial predictions of the biotic components of the reef, we needed continuous environmental maps to correlate the information. The high-resolution satellite images and the bathymetric map provided an excellent platform for the basic continuous maps. Then 2 consecutive modeling runs were performed with GRASP methodology (Lehmann et al. 2002a) in S-PLUS 2000 Professional (Insightful) using binary and Poisson distributions and exporting the

final predictions as lookup tables (LUTs) to be imported by GRASP scripts in ArcView 3.x (ESRI). GRASP (Lehmann et al. 2002a) is a general method for making spatial predictions of response variables (RVs) using point surveys of the RVs and spatial coverage of predictor variables (PVs). GRASP uses GAMs to fit response surfaces as function of predictors in environmental space, in order to use the spatial pattern of the predictor surfaces to predict the response in geographic space. GRASP methodology differs from other geostatistical analysis (surface-fitting algorithms) in that it uses surface estimates in 'predictor space' and not directly in 'geographic space' (Chambers & Hastie 1993, Lehmann et al. 2002a,b,c).

The statistical evaluation of the models was carried out along with a cross-validation for each selected model, either using correlation coefficient values (Gaussian and Poisson distributions) or receiver operating characteristic (ROC, described in Fielding & Bell 1997, for binomial distributions) (Lehmann et al. 2002a,b,c). All models were fitted with a quasi-binomial distribution, using a backwards step-wise procedure to select the significant predictors, and using the *F*-statistic ( $p = 0.05$ ) for assessing changes in residual deviance.

GRASP was first used here to predict the geomorphologic (bottom) features (sand/sediment, rock/calcareous pavement, rubble and topographic complexity — also known as substratum rugosity) using data from the depth map and the 4 satellite image bands as PVs. Topographic complexity and percentage cover of the substrata types were used as RVs.

**Modeling coral reef's biotic components with GRASP.** Coral reef biotic components were then modeled with GRASP as broad functional groups (BFGs). That is percentage cover on the groups: hard corals, octocorals, algae, sea grass and sponges. These were predicted using percent cover value of each BFG by stations as RVs, and information was derived for substrata types, depth, topographic complexity and the depth corrected satellite image band as PVs.

**Coral reef habitat classification.** Using data on percent cover of BFGs, extracted from the video transects, we proceeded to perform a Monte Carlo test for selecting the significantly related variables to the structure of our benthic data (CANOCO v.4, Microcomputer-Power), then we used those significant variables to perform an agglomerative cluster analysis using the Gower similarity coefficient by Weighted Pair Group Average (MVSP v3.12b, Kovach Computing Services); these options were selected because the Gower coefficient is used when mixed data types are present in the matrix, weighted pair group average (WPGMA) because average linking methods provide a more balanced approach to clustering, and the weighted option because it was expected that some of the reef habitats

were less sampled than others. In order to correctly classify the reef lagoon and the fore-reef habitats, we used 2 separate runs of Monte Carlo tests and 2 cluster analyses, with the data of the stations corresponding to either reef lagoon or fore-reef zones. With the obtained dendrograms, the characteristics of the stations (percent coverage of biotic components and substrata, depth, topographic complexity, etc) in each cluster were averaged in order to obtain the 'definition' of each cluster (each cluster now treated as a habitat); Note that the 'cut' in the dendrogram was not performed at a given similarity (i.e. 75%), but at a distance where it 'made sense' for each cluster, again, taking into account expert knowledge of the system. Then a discriminant analysis (Statistica 6, MathSoft) was performed in order to test our 'hand picked' (user defined) classification of clusters of stations corresponding to our habitats (also user defined, taking into account expert knowledge); since the stations are grouped into different categories (habitats), the test aims to establish the accuracy of fit of the stations to their categories.

After obtaining the characteristics (percentage of cover of the different groups, depth, topographic complexity, etc) defining each habitat, we performed a query in the GIS layers (prediction maps) in order to obtain habitat distribution maps, which can be displayed over the original high-resolution satellite image.

The implementation of this proposed methodology is depicted as a flow chart in Fig. 2.

**Supervised classification of Ikonos for comparison with GRASP.** A supervised classification was performed on the Ikonos image, using the colour bands: blue, green, red and the depth corrected band. The classification scheme was defined using the field stations grouped in the habitat classification mentioned above. The classification was repeatedly performed to redundancy by several decision rules, minimum distance (or spectral distance) being the one selected for this analysis.

## RESULTS

### Digital topographic model (DTM)

The DTM layer imported to ArcView was rasterized at the same resolution as the original DTM model ( $4 \times 4$  m pixel resolution) and a continuous depth map (0 to -35 m) of the study area was extracted (Fig. 3).

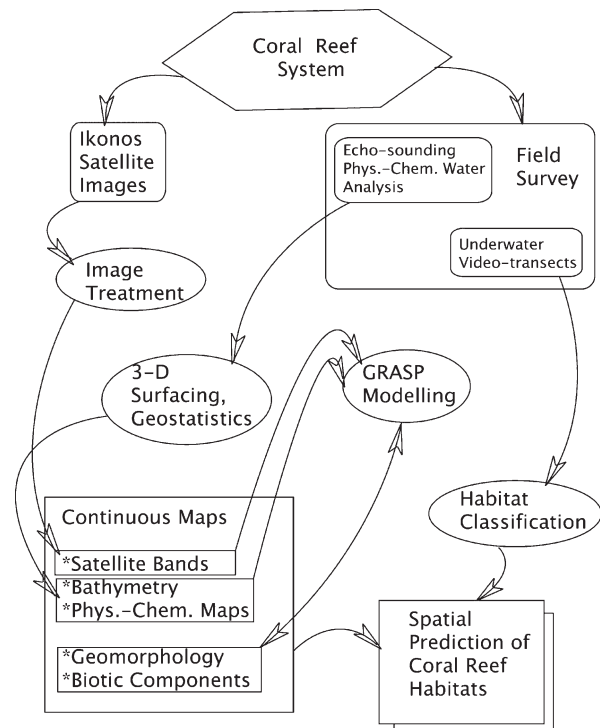


Fig. 2. Workflow of the proposed implementation of coral-reef-habitat spatial prediction. The hexagon is the natural system, rounded rectangles are information collected by diverse methods, ellipsoids are data treatments and rectangles are outputs/results. GRASP: Generalized Regression Analyses and Spatial Prediction

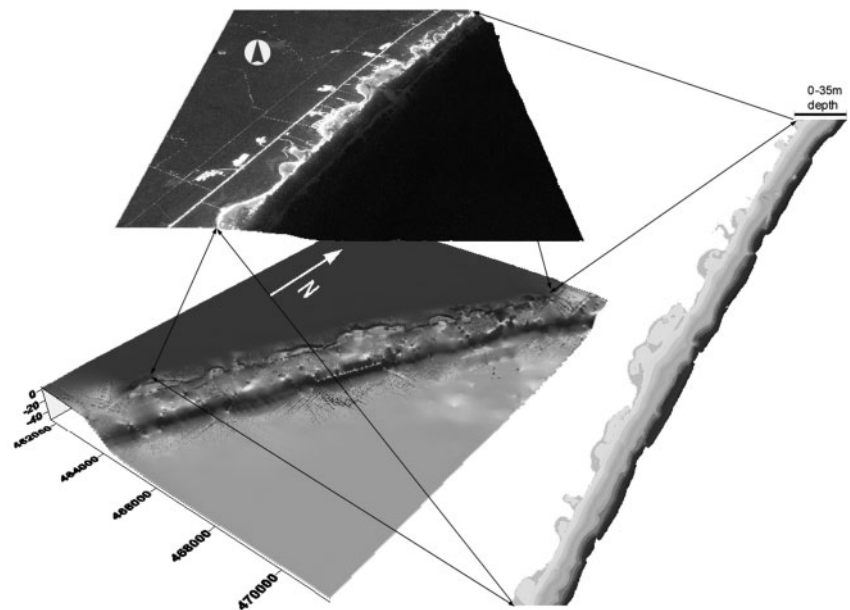


Fig. 3. Akumal digital topographic model (DTM) from echo-sounding survey, Ikonos high-resolution satellite image overlay on top of the DTM and raster extraction (0 to 35 m depth) from 'depth' layer in the GIS

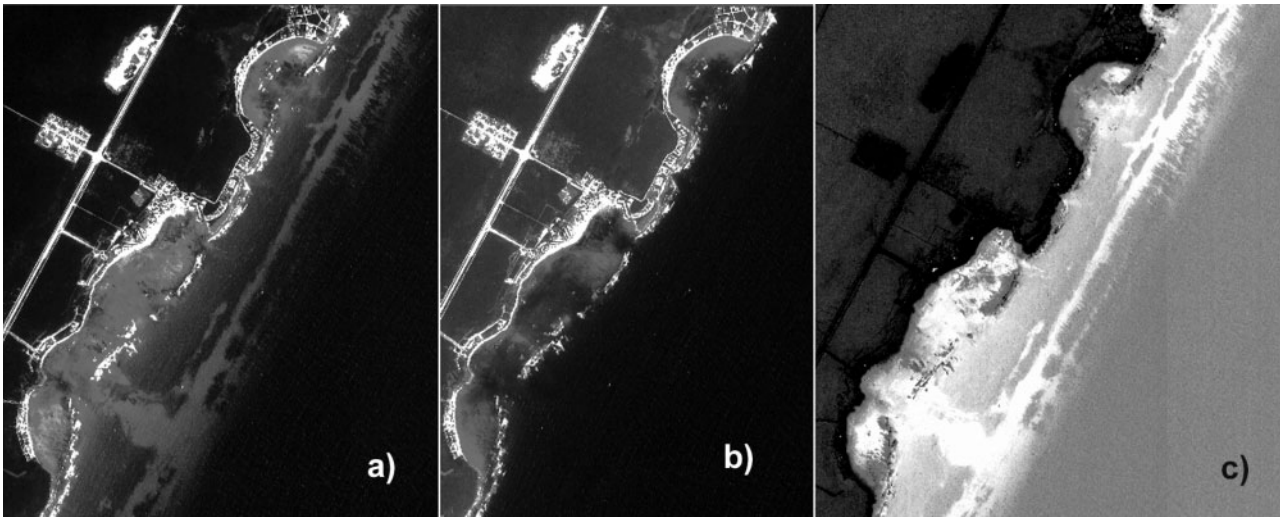


Fig. 4. Satellite band combinations with algebraic transformation, for water-column effect correction (small excerpt from the whole treated image to show better detail). (a) blue and green (B-G), (b) green and red (G-R), and (c) blue and red (B-R) band combinations (this last combination provided the best correction)

### Satellite image enhancement

A linearization of the light absorption in the satellite image was performed with combinations of bands of the image (B-G, G-R, B-R). Applying the algebraic transformation to the band combinations (with MultiSpec v.3 software), we obtained new channels and decided to use B-R because it offered a better definition of the bottom features than the other 2 combinations (Fig. 4).

### Spatial prediction

With the first run of GRASP modeling (validation, V, and cross-validation, CV, values of the models, ranging from 0.7 to 1 ROC), we obtained the prediction and prediction standard error (pSE) maps of topographic complexity (TC) values, rock/calcareous pavement, sand/sediment and rubble cover (Fig. 5), all with a spatial definition of 4 × 4 m pixels. An actual sand/sediment pSE map was included in the figure to illustrate this feature in the analysis.

For the second GRASP run, the model correlation values (V and CV) ranged from 68.3 to 86.2% for hard corals, octocorals and calcareous-articulated algae.

Green algae and brown/filamentous algae models showed ROC values (V and CV) of 0.6 and 0.8, respectively. The prediction and pSE maps of the biotic components of the reef were obtained, also at 4 × 4 m resolution; see Fig. 6, where the brown/filamentous algae pSE map was included in order to illustrate this feature in the analysis.

A separate modeling run was performed for prediction of seagrasses, in a separate set of layers, queried to display only the reef lagoon areas of the study zone (Fig. 6). The evaluation and cross-validation of the model by ROC statistic was 0.94 (see Table 1 for

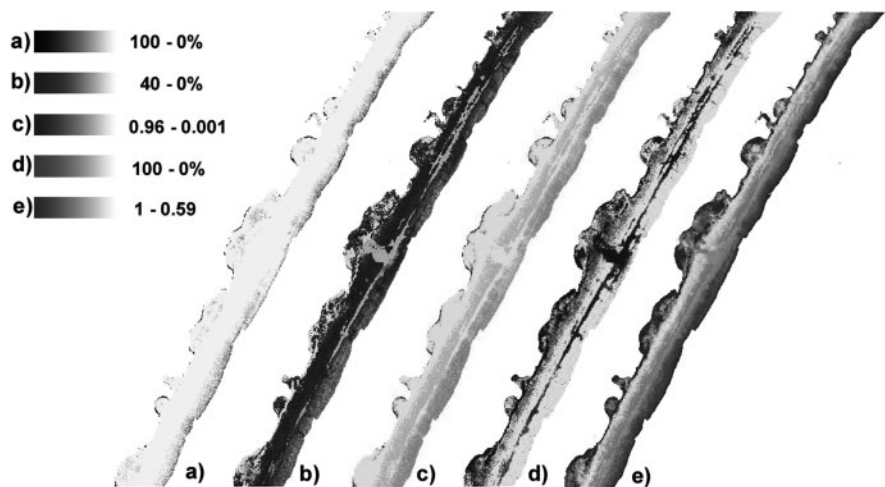


Fig. 5. Base maps (reef geomorphologic features) predicted with GRASP: (a) rubble, (b) rock/calcareous pavement, (c) topographic complexity, (d) sand/sediment, (e) sand/sediment prediction standard error (pSE). Gradient values given as % cover except (c) and (e)

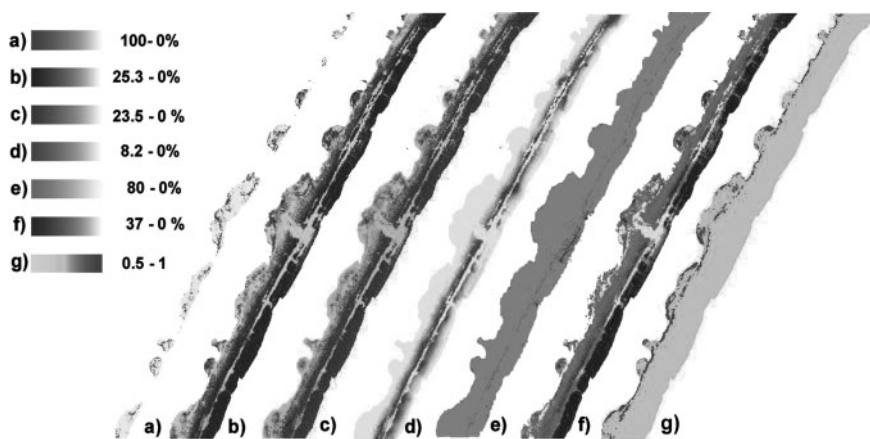


Fig. 6. GRASP of biotic components coverage: (a) seagrass (reef lagoon area extraction), (b) hard corals, (c) octocorals, (d) calcareous-articulated algae, (e) green algae, (f) brown/filamentous algae, (g) brown/filamentous algae prediction standard error (pSE). Gradient values given as % cover, except (g)

initial and final models, degrees of freedom used in the smoothing, validation and cross-validation values).

### Habitat classification

The Monte Carlo tests showed 7 significant variables ( $p < 0.05$ ) out of 14 (hard coral, octocoral, brown/filamentous algae, calcareous/articulated algae, green algae, recent dead coral and sponges), for classifying the fore-reef (slope) habitats, and 7 out of 14 for classifying the reef lagoon (sea grass, brown/filamentous algae, hard

Table 1. Initial (all predictive variables included) and final (selected predictive variables) GAMs for geomorphologic and biotic response variables, as produced by GRASP interface, where s = spline smoother; names in parentheses are the variables introduced in the models: bands 1, 2, 3 and 4, correspond to satellite image bands red, green, blue and water column depth corrected composite; depth = depth; sand = sand/sediment; rock = rock/calcareous pavement; rubble = rubble; ct = topographic complexity; and values after comma are degrees of freedom of the spline smoother. ROC: receiver operating characteristic; COR: correlation. Example: Response Variable: s(Predictive Variable<sub>1</sub>, degrees of freedom for the spline smoother)+ s(Predictive Variable<sub>2</sub>, degrees of freedom for the spline smoother)+...

Variable	Initial model	Final model	Validation	Cross-Validation
Rock/ calcareous pavement	Rock: s(band1, 4) +s(band2, 4)+s(band3, 4) +s(B1, 3Comp, 4)+s(depth, 4)	Rock: s(band1, 4) +s(B1, 3Comp, 4)	ROC:0.7	ROC:0.7
Sand/ sediment	Sand: s(band1, 4)+s(band2, 4) +s(band3, 4)+s(band4, 4) +s(depth, 4)	Sand: s(band1, 4)+s(band2, 1) +s(band4, 4)+s(depth, 4)	ROC:0.875	ROC:0.875
Rubble	Rubble: s(band1, 4)+s(band2, 4) +s(band3, 4)+s(band4, 4) +s(depth, 4)	Rubble: s(band1, 4) +s(band2, 1)+s(band4, 4) +s(depth, 1)	ROC:0.798	ROC:0.798
Topographic complexity	TopogComp.: s(band1, 3) +s(band2, 3)+ s(band3, 3) +s(band4, 3)+ s(depth, 3)	Topog.Comp.: s(band1, 3) +s(band4, 3)	ROC: 1.0	ROC: 1.0
Hard coral	HardCoral: s(band1, 4) +s(band2, 4)+ s(band3, 4) +s(band4, 4)+ s(depth, 4)+ s(ct, 4) +s(rubble, 4)+ s(rock, 4)+s(sand, 4)	HardCoral: s(band1, 1) +s(band3, 1)+s(band4, 1) +s(depth, 4)+ s(sand, 4)	COR:0.862	COR:0.778
Octocoral	Octocoral: s(band4, 4)+s(depth, 4) +s(ct, 4)+s(rubble, 4)+s(rock, 4) +s(sand, 4)	Octocoral: s(band4, 1) +s(ct, 1)+ s(rock, 4)+s(sand, 1)	COR:0.840	COR:0.771
Green algae	GreenAlgae: s(band4, 4) +s(depth, 4)+ s(ct, 4)+ s(rubble, 4) +s(rock, 4)+s(sand, 4)	GreenAlgae: s(band4, 1) +s(depth, 1)+ s(sand, 4)	ROC:0.606	ROC:0.606
Filamentous/ brown algae	Filam.Brwn.Algae: s(band4, 4) +s(depth, 4)+ s(ct, 4)+ s(rubble, 4) +s(rock, 4)+s(sand, 4)	Filam.Brwn.Algae: s(rock, 4)+ s(sand, 4)	ROC: 0.8	ROC: 0.8
Calcareous articulated algae	Calc.Artic.Algae: s(band, 4) +s(depth, 4)+ s(ct, 4)+ s(rubble, 4) +s(rock, 4)+s(sand, 4)	Calc.Arctic.Algae: s(depth, 4) +s(sand, 1)	COR:0.683	COR:0.564
Seagrass	SG: s(band1, 4)+ s(band2, 4) +s(band3, 4)+ s(band4, 4) +s(depth, 4)+s(rubble, 4)+s(Sand, 4) +s(ct, 4)+s(Rock, 4)	SG: s(band2, 4)	ROC: 0.941	ROC: 0.941

coral, octocoral, green algae, rock/calcareous pavement and sand/sediment). From the cluster analysis (lagoon and fore reef) 8 habitats were extracted: (1) back-reef coral patches; (2) reef lagoon sea-grass dominated plains; (3) reef-lagoon algae prairies; (4) high-coral and high-algae cover spurs and grooves; (5) high-coral and low-algae cover spurs and grooves; (6) medium-coral and medium-algae spurs and grooves; (7) medium-coral and high-algae spurs and grooves; (8) sand plains and channels with absent or very low biotic cover (Fig. 7).

The selection of clusters (the similarity level at which the 'cut' was performed, that is, the selection of the groups of stations defining each habitat) was statistically validated by performing a discriminative analysis for each set of stations (Statistica v.6 MathSoft). These discriminant tests showed low p-values ( $p < 0.0001$ ) for our selections of clusters.

With the average and range values for each component of the defined habitats (Table 2), we queried the predicted biotic and substrata GIS layers in order to obtain the 8 habitat maps (Fig. 8), and created a new map merging the layers of the 8 different habitats (Fig. 9).

### Comparison between GRASP and traditional supervised classification outputs

A quantitative comparison between the predicted habitat map and the classified image was performed by landscape metrics, calculating the area from the number of pixels assigned to each habitat/class (Table 3).

## DISCUSSION

Following the recent stream of studies and applications that are possible using GAMs, we obtained high-definition biotic maps of a coral reef, thus implementing a powerful and versatile tool for coral reef ecologists and managers.

The level of accuracy achieved by the GRASP methodology implemented here for coral reefs is acceptable (cross-validation ranging from 0.6 to 1.0 ROC, 68.3 to 86.2% of correlation in the models, and >50% of accuracy for predicted maps), directly related to the amount of information used to define the relationships between the components in the models. It is positive also that all the geomorphologic, and most of the biotic, variables were satisfactorily predicted. The

biotic broad groups of zoanthids and sponges were not predicted, since the fitting of their models was very low due to the low amount of cover and/or absence in the survey stations. In comparison with other studies in-

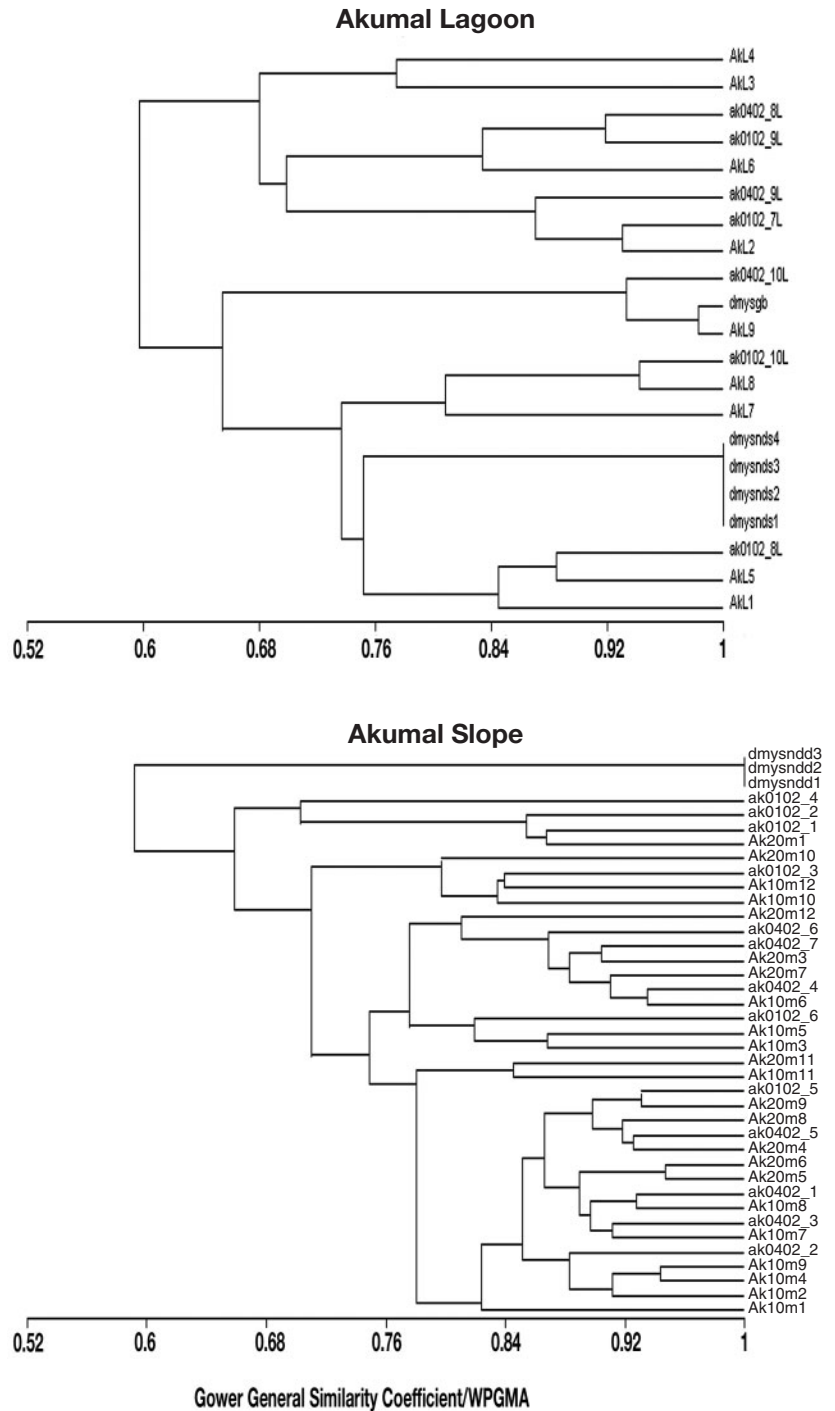


Fig. 7. Cluster analysis (WPGMA, Gower similarity index) of the stations to define habitats (1 to 8). Note that clusters were selected at different similarity, where they made ecological sense. Stations defining Habitats 1 and 2 have a similarity coefficient of at least 0.85. Stations for Habitats 3 to 6 show at least 0.77 similarity coefficient

Table 2. Habitat description. Characteristics taken from classification analysis and used for querying the spatial prediction maps of geomorphologic and biotic components. All values given in % cover except for topographic complexity and depth

Variables	Lagoon			Habitat Slope				Sand
	1	2	3	4	5	6	7	
<b>Biotic cover</b>								
Hard coral	>5 – <20	0	<6	>10 – <25	>25	>10 – <25	>10 – <25	0
Octocoral	<10	0	<10	>5 – <15	>5 – <15	<16	>5 – <25	0
Filamentous/ brown algae	<25	<3	<30	>20 – <30	<20	>15 – <36	<60	0
Green algae	<1	<1	<15	<10	<5	1	<1	0
Calcareous articulated algae	<1	0	0	<22	<10	<6	<20	0
Seagrass	<20	>75	<10	0	0	0	0	0
<b>Geomorphologic features</b>								
Sand/sediment	<20	<25	>30	<5	<1	<30	<15	>95
Rubble	<15	0	<60	<15	<15	<15	<22	0
Rock/calcareous pavement	<30	0	<15	<30	<45	<40	<45	0
Depth	<3	<3	<3	>10	>10	>5	<10	0
Topographic complexity	<0.25	<0.02	<0.25	<0.35	<0.25	<0.40	<0.33	<0.01

volving GRASP, the levels of cross-validation obtained give values of 0.94 ROC (Lehmann et al. 2002a), 0.65 to 0.98 ROC (Lehmann et al. 2002b) and 0.61 to 0.72 ROC correlations (Zaniewski et al. 2002).

As one form of validation for the spatial predictions of the modeled variables, the parallel generation of standard error (pSE) maps for each predicted variable gives spatially explicit validations of the values predicted by the model, with each pixel having a predicted pSE value for its counterpart pixel. On the one hand, these maps can be used to assess the accuracy of predictions, without the necessity of having a great amount of field data to make separate training and testing data sets, which, at the beginning of this study, was one of the goals. On the other hand, we have seen that even though recording more field data could give us the opportunity to perform an accuracy assessment

using traditional methods (comparison between training and test data sets), it must be carried out in the same period of field work, since percentage cover of important biotic components such as algae tends to fluctuate during and between the climatic seasons (Garza-Pérez et al. unpubl.), making *a posteriori* comparisons between data sets difficult.

Nevertheless, the accuracy level achieved in this way (with its modest number of stations) could have been improved further. The water column data recorded for each station (salinity, temperature and dissolved oxygen) could not be used in the final predictive models due to the lack of continuous maps for these heterogeneous variables, thus decreasing the final model's validation and cross-validation values. These variables were highly significant for

some biotic components (i.e. sea-grass, green algae, where the cross-validation values of the final models ranged from 0.85 to 1.0 correlation when these variables were incorporated to the modeling process). The small number of recordings of these variables (54) did not allow us to create extrapolation maps with the desired accuracy all across the study area (with kriging or other geostatistical methods). It is highly recommended that a full physical-chemical spatial analysis survey should be carried out along the bathymetric soundings, instead of sampling discrete sites.

In an overall fashion, considering the low number of sampling stations, the size of the area of study and the complexity of the system, we obtained useful predictions of the biotic and geomorphologic components and a very coherent map of habitats, in accordance to the studied reef system. For this paper a classification scheme of 8 coral reef habitats was selected from the cluster analysis to give a clear example of the habitat differentiation achievable by using the GRASP maps, which will be comparable with traditional remote sensing classification methods in subsequent studies. And as it is a user-defined classification, several other broader or more specific classifications can be selected. As we use a hierarchical system for managing data, modeling and prediction of species assemblages and species spatial patterns can be done easily,

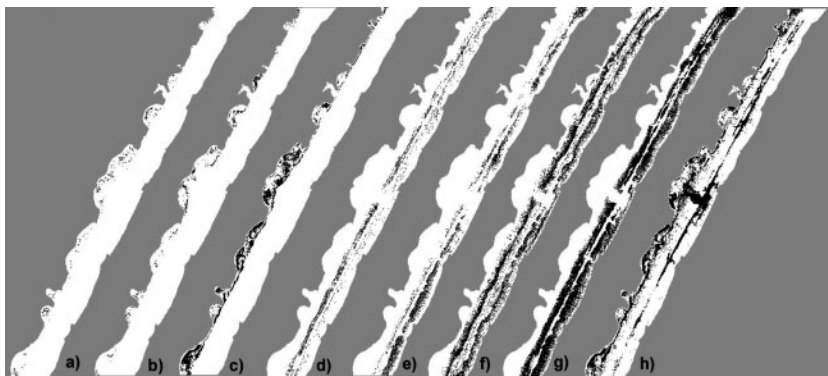


Fig. 8. Habitat maps queried from the geomorphologic and biotic maps. Habitats (a) 1, (b) 2, (c) 3, (d) 4, (e) 5, (f) 6, (g) 7, (e) 8 (for habitat descriptions, see Table 2)



thus giving ecological robustness to the characterization of the system.

The comparison between GRASP and traditional supervised classification outputs using the same image and classification scheme provided, in a sense, another validation of the GRASP maps. The classification scheme of 8 classes was selected, taking into account the results of Andrefoüet et al. (2003), to evaluate coral reef areas with Ikonos images. The 8-class thematic map produced with the supervised classification methodology did not show an accurate spatial definition of classes (see Fig. 9), as it was unable to separate Habitat 4 from Habitat 6 (see Table 2 for habitat description), so a new classification was performed, this time with 9 classes (Fig. 9), separating shallow sand from deep sand stations. The new thematic map was more in accordance with the GRASP map, assigning most classes into similar spatial distribution patterns; nevertheless the GRASP map provided an image more in accordance with reality. The traditional supervised classification method presented a spatial incorrect estimation of habitats, thus affecting the interpretability and usefulness of the classified map, and had less spatial accuracy when delimiting boundaries of classes (Table 3, Fig. 9). This is probably due to the minimum-distance decision rule used, which tends to classify pixels that otherwise should go unclassified, and does not consider class variability (ERDAS 1999). This decision rule provided the best interpretation of our classification scheme and fitted more appropriately to our data set. The GRASP-derived (habitat) map presents 'unclassified' areas corresponding to geomorphologic features and biotic component coverage values not represented in the field data and thus not taken into account in the habitat definition (from the classification). However, the GRASP maps of individual features and components include wider predicted ranges of values, below and above those registered in the field, in order to depict the gradients of coverage all along the study area. Thus GRASP turned out to be more robust for producing predictive maps, validating the methodology proposed here for obtaining useful spatial predictions over large areas from small datasets.

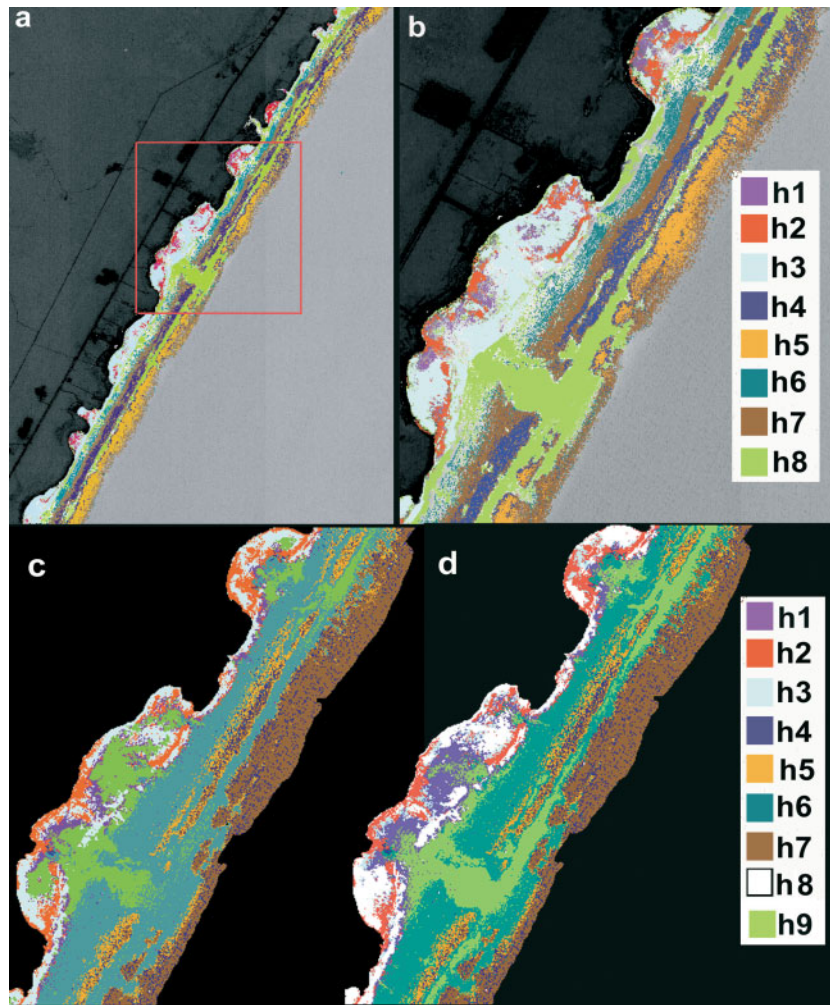


Fig. 9. Classified map of Akumal Coral Reef: (a) whole study area map showing the 8 habitats (h1 to h8) queried from the geomorphologic and biotic maps obtained by generalized regression analyses and spatial prediction (GRASP) methodology, (b) close-up view of a small portion to enhance details of the habitat mapping by GRASP, (c) view of the same zoomed area of the Ikonos image treated by supervised classification (minimum distance), following an 8-class scheme, (d) same portion and procedure as in (c), but following a 9-class scheme (including h9)

Expert knowledge of a system, along with the use of high-resolution satellite images, plays a very important role in this kind of research. In this study this helped us, for instance, to establish a focused sampling design in order to record the greatest amount of possible variation in the coral reef habitats with the least cost.

Other studies, using GAMs for prediction of vegetation distribution in terrestrial ecosystems (Yee & Mitchell 1991, Leathwick et al. 1998, Lehmann & Austin 1999, Cawsey et al. 2002, Ferrier et al. 2002, Lehmann et al. 2002a,b) in New Zealand and Australia, have achieved very good results in the fitting and stability of the models and in the accuracy of predictions, using spatial resolutions ranging from  $1 \times 1$  to  $5 \times 5$  km, regardless of the great areas involved

Table 3. Landscape metrics. Area estimated for each habitat from GRASP habitat map, and from supervised classification image (8- and 9-class scheme). All values expressed in m<sup>2</sup>. All habitats correspond to description in Table 2, except in the 9-class scheme, where former Habitat 8 was divided into shallow and deep sand

Habitat	GRASP	8-class scheme	9-class scheme
1	117 552	391 216	689 792
2	217 696	560 496	524 912
3	1 137 936	1 044 496	229 904
4	704 128	603 424	603 424
5	1 096 608	853 808	853 808
6	705 104	4 424 592	3 551 824
7	2 564 544	3 033 008	3 033 008
8	2 868 240	1 374 288	1 153 840
9	–	–	1 644 816
<b>Total</b>	<b>9 411 808</b>	<b>12 285 328</b>	<b>12 285 328</b>
<b>Total study area</b>	<b>12 285 328</b>	<b>12 285 328</b>	<b>12 285 328</b>

(New South Wales and New Zealand territories). One of the factors involved in the fitting and accuracy of these models is the great amount of data used (sets ranging in the thousands of plots), which are extracted from sources such as national vegetation and forest surveys. Unfortunately these data sets are almost non-existent or not reliable in most third-world coral reef areas, and our implementation offers a good alternative for characterization and condition assessment.

The direct ecological applications for these GRASP continuous maps on coral reefs are: (1) reef habitat assessment, which allows us to identify hot spots of biotic features, (2) use as baseline information to classify and characterize the spatial extent of each defined habitat, (3) helping in management tasks (defining zones within marine reserves, defining the probable distribution and abundance of any given resource), (4) characterization of the spatial relationships between coral reef fish benthic communities (Arias-Gonzalez et al. unpubl.).

One key utility of this implementation of spatial prediction in coral reefs is that from a relatively small area we are able to scale up the predictions to greater areas in the Mexican Caribbean fringing reef system, and the methodology (though not the models) can be easily transported to other reef areas in the world.

Currently we are applying this methodology to seaplain atoll-type reefs in the Caribbean Sea (Chinchorro Bank [Gulf of Mexico] and Alacranes Reef) and the Great Barrier Reef (Davies Reef) (Acosta-González et al. unpubl., Arias-González et al. unpubl.).

In contrast with common remote-sensing classification procedures (non-supervised and supervised classification), in this methodology we do not assign each pixel to individual classes (thus our expression of hard-classification, where a pixel can belong to one class only per image treatment), but each pixel has a

value assigned to each different layer, giving us the possibility to map the subtle variations of the components of the reef as a continuum. The *a posteriori* merging and classification of the continuous maps into habitats has more to do with the human need to break things into smaller comprehensive portions. As Townsend (2000) points out, the delineation of classes on maps is necessary to facilitate the communication of information about spatial patterns of the distribution and abundance of species.

In the technical aspect of this study, Ikonos high-resolution satellite imagery has proven to be more useful for predictive tasks (such as this GRASP

application) than for 'hard' classification by remote sensing procedures. In a recent study, Andrefoüet et al. (2003) stated that for management and scientific applications that need at least 80% of accuracy, only a 4- to 5-class scheme can be used with Ikonos imagery, and that hard coral cover areas are poorly classified most of the time, especially when these areas also include algae coverage. In this case we have been able to obtain spatial predicted coverage values of each one of the most important biotic components (including corals and 3 different types of algae) and geomorphologic characteristics of our study area, thus giving a wider range of uses and applications for the spatial information obtained. Ikonos imagery is referred to as not very effective in cost-benefit terms in comparison with Landsat TM for coarse-level habitat mapping, but its value in defining boundaries of habitat patches over other types of imagery is recognized (Mumby & Edwards 2002). This favorable characteristic allows the use of expert knowledge in order to implement and refine a directed sampling effort towards recording the maximum possible variation in the system.

In order to establish the advantages (or not) of the implementation of this methodology in coral reef ecosystems, we are currently comparing the results of GRASP against traditional 'hard classification' methods (relying only on spectral information) for mapping coral-reef biotic and geomorphologic features, using Ikonos and Landsat 7 ETM+ imagery (for 3 reefs in the Mexican Caribbean). We are also investigating the feasibility of extending the spatial predictions to larger reef areas using Landsat 7 ETM+ images of the northern part of Quintana Roo state in Mexico. The present study can be taken as the introduction of GRASP to coral reef ecology; the comparison between methods would be a second step towards full validation of the

GRASP application and implementation on coral reef systems.

Partial approaches to our methodology are proposed by Mumby & Harbone (1999), who propose a hierarchical classification scheme for the Caribbean reefs; by Andrefouët & Claereboudt (2000), who use a classification scheme of coral reefs that correlates the environmental data with remotely sensed data (SPOT-HRV); by Aspinall (2002), who used logistic regression to validate vegetation maps extracted from a high-definition hyperspectral airborne scanner (128 bands, ESSi Probe 1); by Guisan & Zimmermann (2000) in their review of predictive habitat-distribution modeling; by Overton et al. (2002) in proposing a pyramidal structure of data management; and by Ferrier et al. (2002) who propose a similar implementation and use of community-level modeling with GAMs in terrestrial ecosystems. In this study we merged several proposals from other science fields to implement this methodology in coral reef ecology.

As pointed out by Yee & Mitchell (1991) several years ago, and recently by Lehmann et al. (2002a), GAMs and GRASP are computer tools that demand high processor performance and large memory capacity when used over large areas at high resolution (several million pixels). Using GRASP in coral reef areas at high resolution (4 m), as we did in this exercise, is possible since the extension of reef area is restricted to a narrow zone in the satellite image. For applications over greater reef areas such as bank reefs, i.e. Alacran Reef in the Gulf of Mexico (Arias-Gonzalez et al. unpubl.) and Chinchorro Bank in the Caribbean Sea, the option could be the use of Landsat 7 ETM+ imagery (30 × 30 m pixel resolution), since covering such areas with Ikonos imagery would be very costly. Nevertheless, if money for such investment (imagery and computers) is not a main constraint, the computing power of the latest generation of processors and current memory capacity has reached the point where the size of the data files is becoming less important.

As a conclusion, we can state that the methodology implemented here, incorporating habitat characterization and topology, 3D topographic models, remote sensing techniques, and spatial prediction (GRASP) in coral reefs, is an effective, useful and reliable way to map extensive coral reef areas with a relatively small investment in field surveys. Combining this GRASP application with time series of community assessment in selected stations on the reef, we could implement a forecast of reef condition and analyze changes of its components over time. As this methodology is fully compatible with GIS, new layers of related information could be added as they become available, and they may improve the prediction accuracy and stability of our initial models.

*Acknowledgements.* We would like to thank CONACyT in Mexico for funding this research (proyecto #28386 N, beca crédito 119328 and beca crédito mixta para estancias en el extranjero), also to thank the kind help and support of the Swiss Center for the Fauna Cartography (CSCF, Neuchâtel) for the first stage of this project. This study would not have been possible without the timely collaboration of S. Andrefouët and F. Muller-Karger at the Remote Sensing Laboratory, USF St. Petersburg, and NASA data-buy program for acquiring the Ikonos imagery. Thanks to the Akumal Dive Shop (D. Brewer, G. Arcila and crew) and Akumal CEA, for all the kind support in field activities. Also to the Coral Reef Ecosystems Ecology Lab members (LEEAC Adventure Team), for their valuable time and help in field activities. We would also like to thank to P. J. Mumby for his valuable comments on the original manuscript and the kind observations and comments from 4 anonymous referees.

#### LITERATURE CITED

- Andrefouët S, Claereboudt M (2000) Objective class definition using correlation of similarities between remotely sensed and environmental data. *Int J Rem Sens* 21(9): 1925–1930
- Andrefouët S, Roux L, Chancerelle Y, Boneville A (2000) A fuzzy probabilistic scheme of study for objects with indeterminate boundaries: application to French Polynesia reefs-capes. *IEEE Transactions on Geosci Rem Sens* 38(1):257–270
- Andrefouët S, Kramer P, Torres-Pulliza D, Joyce KE and 10 others (2003) Multi-sites evaluation of Ikonos data for classification of tropical coral reef environments. *Rem Sens Environ* 88:128–143.
- Aronson RB, Swanson DW (1997) Video surveys of coral reefs: uni and multivariate applications. *Proc 8th Int Coral Reef Symp* 2:1923–1926
- Aspinall RJ (2002) Use of logistic regression for validation of maps of the spatial distribution of vegetation species derived from high spatial resolution hyperspectral remotely sensed data. *Ecol Model* 157:301–312
- Austin MP (1999) The potential contribution of vegetation ecology to biodiversity research. *Ecography* 22:465–484
- Austin MP (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. *Ecol Model* 157(2):101–118
- Basu A, Nalamotu C (1997) Marine geographic information system for the exclusive economic zone. *Mar Geod* 20(2-3):255–266
- Bio AMF (2000) Does vegetation suit our models? Data and model assumptions and the assessment of species distribution in space. PhD thesis, Utrecht University
- Bushing WW (1997) GIS-based gap analysis of an existing marine reserve network around Santa Catalina island. *Mar Geod* 20(2-3):205–234
- Cawsey EM, Austin MP, Baker BL (2002) Regional vegetation mapping in Australia: a case study in the practical use of statistical modelling. *Biodiversity Conserv* 11:2239–2274
- Chavaud S, Bouchon C, Marniere R (1998) Remote sensing techniques adapted to high resolution mapping of tropical coastal marine ecosystems (coral reefs, seagrass beds and mangrove). *Int J Rem Sens* 19(18):3625–3639
- Chambers EM, Hastie TJ (1993) *Statistical models*. Chapman and Hall, London
- Debinski DM, Kindscher K, Jakubaukas ME (1999) A remote sensing and GIS-based model of habitats and biodiversity in the Greater Yellowstone Ecosystem. *Int J Rem Sens* 20 (17):3281–3291

- ERDAS (1999) Field guide, 5th edn. ERDAS, Atlanta
- Fedra K (1998) Geographic information systems and spatial analysis in coastal zone management. *Mar Ind Technol* 1,2: 1–12
- Ferrier S, Drielsma M, Manion G, Watson G (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in north-east New South Wales: II. Community level modelling. *Biodiversity Conserv* 11:2309–2338
- Fielding AH, Bell JF (1997) A review of methods for the assessment of predictions errors in conservation presence/absence models. *Environmental Conserv* 24(1): 38–49
- González-Gándara C, Membrillo-Venegas N, Nuñez-Lara E, Arias-González JE (1999) The relationship between fish and reefs in the Alacranes Reef, Yucatan, Mexico: a preliminary trophic functioning analysis. *Vie Milieu* 49(4): 275–286
- Green EP, Mumby PJ, Alasdair JE, Clark CD (2000) Remote sensing handbook for tropical coastal management. UNESCO, Paris
- Guisan A, Zimmerman NE (2000) Predictive habitat distribution models in ecology. *Ecol Model* 135 147–186
- Guisan A, Edwards TC, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol Model* 157(2–3)89–100
- Hastie T, Tibishirani R (1990) Generalized additive models. Chapman Hall, London
- Heegaard E, Birks HH, Gibson CE, Smith SJ, Wolfe-Murphy S (2001) Species-environmental relationships of aquatic macrophytes in Northern Ireland. *Aquat Bot* 70 175–223
- Janauer GA (1997) Macrophytes, hydrology, and aquatic ecotones: a GIS-supported ecological survey. *Aquat Bot* 58: 379–391
- Leathwick JR, Burns BR, Clarkson BD (1998) Environmental correlates of three alpha-diversity in New Zealand primary forests. *Ecography* 21:235–246
- Lehmann A (1998) GIS modelling of submerged macrophyte distribution using generalized additive models. *Plant Ecol* 139:113–124
- Lehmann A, Austin M (1999) Implementing GRASP to predict tree distribution in Southeastern NSW and to test different potential measures of water stress. Report to the Division of Wildlife and Ecology, CSIRO, Canberra
- Lehmann A, Lachavanne JB (1997) Geographic information systems and remote sensing in aquatic botany. *Aquat Bot* 58:195–207
- Lehmann A, Jaquet JM, Lachavanne JB (1997) A GIS approach of aquatic plant spatial heterogeneity in relation to sediment and depth gradients, Lake Geneva, Switzerland. *Aquat Bot* 58:347–361
- Lehmann A, Overton JMC, Leathwick JR (2002a) GRASP: Generalized Regression Analysis and Spatial Predictions. *Ecol Model* 157:189–207
- Lehmann A, Leathwick JR, Overton JMcC (2002b) Assessing New Zealand fern diversity from spatial predictions of species assemblages. *Biodiversity Conserv* 11:2217–2238
- Lehmann AJ, Overton, M, Austin MP (2002c) Regression models for spatial prediction: their role for biodiversity and conservation. *Biodiversity Conserv* 11:2085–2092
- Mumby PJ, Edwards AJ (2002) Mapping marine environments with Ikonos imagery: enhanced spatial resolution can deliver greater thematic accuracy. *Rem Sens Environ* 82:248–257
- Mumby PJ, Harborne AR (1999) Development of a systematic classification scheme of marine habitats to facilitate regional management and mapping of Caribbean coral reefs. *Biological Conserv* 88:155–163
- Mumby PJ, Green EP, Edwards AJ, Clark CD (1997) Coral reef habitat-mapping: how much detail can remote sensing provide? *Mar Biol* 130:193–202
- Mumby PJ, Green EP, Clark CD, Edwards AJ (1998) Digital analysis of multispectral airborne imagery of coral reefs. *Coral Reefs* 17:59–69
- Münier B, Nygaard B, Ejrnæs, Bruun HG (2001) A biotope landscape model for prediction of semi-natural vegetation in Denmark. *Ecol Model* 139:221–233
- Overton JM, Leathwick JR, Lehmann A (2000) Predict first, classify later – a new paradigm of spatial classification for environmental management: a revolution in the mapping of vegetation, soil, land cover and other environmental information. 4th Int Conf Integrating GIS and Environmental Modelling: Problems, Prospects and Research Needs. Banff, Alberta. Available at [www.colorado.edu/research/cires/banff/pubpapers/212/](http://www.colorado.edu/research/cires/banff/pubpapers/212/)
- Overton JM, Stephens RTT, Leathwick JR, Lehmann A (2002) Information pyramids for informed biodiversity conservation. *Biodiversity Conserv* 11:2093–2116
- Purkis S, Kenter JAM, Oikonomou E, Robinson IS (2002) High resolution ground verification, cluster analysis and optical model of reef substrate coverage on Landsat TM imagery (Red Sea, Egypt). *Int J of Rem Sens* 23: 1677–1698
- Schmieder K (1997) Littoral zone-GIS of Lake Constance: a useful tool in lake monitoring and autecological studies with submersed macrophytes. *Aquat Bot* 58:333–346
- Stanbury KB, Starr RM (1999) Applications of geographic information systems (GIS) to habitat assessment and marine resource management. *Oceanol Acta* 22 (6): 699–703
- Stoner AW, Manderson JP, Pessutti JP (2001) Spatially explicit analysis of estuarine habitat for juvenile winter flounder: combining generalized additive models and geographic information systems. *Mar Ecol Prog Ser* 213:253–271
- Tappeiner U, Tasser E, Tappeiner G (1998) Modelling vegetation patterns using natural and anthropogenic influence factors: preliminary experience with a GIS based model applied to an alpine area. *Ecol Model* 113:225–237
- Tappeiner U, Tappeiner G, Aschenwald J, Tasser E, Ostendorf B (2001) GIS based modelling of spatial pattern of snow cover duration in an alpine area. *Ecol Model* 138: 265–275
- Theobald DM, Hobbs NT, Bearly T, Zack JA, Shenk T, Riebsame WE (2000) Incorporating biological information in local land-use decision making: designing a system for conservation planning. *Landscape Ecol* 15:35–45
- Townsend PA (2000) A quantitative fuzzy approach to assess mapped vegetation classifications for ecological applications. *Rem Sens Environ* 72:253–267
- Woodhouse S, Lovett A, Dolman, P, Fuller R (2000) Using a GIS to select priority areas for conservation. *Computers, Environment Urban Syst* 24:79–93
- Yee TW, Mitchell ND (1991) Generalized additive models in plant ecology. *J Veg Sci* 2:587–602
- Zaniewski AE, Lehmann A, Overton JMcC (2002) Predicting species distribution using presence-only data: a case study of native New Zealand fern. *Ecol Model* 157:261–280