Appendix S2

1- The StDM basic unit



a) The StDM basic unit: the rectangle represents a state variable; parameters or constants are small circles; sinks and sources are cloudlike symbols, flows are thick arrows, and all the relations between state variables and other variables are fine arrows. b) An example of a conventional multiple regression equation, containing the statistics used to estimate the ecological parameters, is also showed. n is the number of independent variables (X), a is the Y intercept, and b1, b2, b3, …, bn are the partial regression coefficients or ecological parameters.

2 - General overview

The proposed Spatially Explicit StDM framework is a sequential process initiated by the analysis of landscape composition (Fig. 2A) in order to determine the suitable descriptors regarding the land use and the infrastructures at the point level (survey points and associated surrounding habitats in a circular buffer with a radius of 250 meters, Appendix A). This characterization was used to support an information-theoretic statistical method (Fig. 2B), based on Generalized Linear Models (e.g. Nelder and Wedderburn 1972), in order to establish the cause-effect interaction criteria between the influence of explanatory variables on little bustard’s abundances. The explanatory variables’ trends were simulated in a system dynamics model (e.g. Ford 1999), as well as the resulting final outputs concerning the little bustard abundance responses for each point, i.e., local prevailing environmental conditions (Fig. 2C). These simulations, when projected into a geographic space (e.g. Johnston 1998) correspondent to a site (Fig. 2D) and submitted to an appropriate geostatistical method (e.g. Urban 2006), create an integrative picture of the ongoing habitat changes and little bustard’s abundances in response to current modifications (Fig. 2E). Since the dynamics of the species is driven by complex socio-ecological systems, the combined use of such statistical modelling and geostatistical techniques was considered a promising approach to address complex emergent assessments that arise from the individual habitat patch dynamics to the whole landscape (Bastos et al. 2012; Santos et al. 2013).

3 – Statistical analyses

In order to avoid multicolinearity, the 10 predictors selected were tested for pairwise correlation using Spearman’s rho correlation coefficient and only predictors with correlation lower than 0.7 (Elith et al. 2006; Wisz and Guisan 2009) and Generalized Variance Inflation Factor lower than 5 were considered (Neter et al. 1996). A generalized linear model (GLM) with a log-link function to fit the model (count data with an approximate Poisson distribution, O’Hara & Kotze, 2010) was used so that the effect of each explanatory variable in the presence of all others could be examined prior to the dynamic model construction (Fig. 2B). This procedure was applied to assess the statistical relation between the abundance of males (response variable) and the area occupied by the principal land uses and linear infrastructures considered (explanatory variables). When compared to other statistical modelling methodologies such as general additive models or generalized linear mixed models, GLMs are more intuitive namely in mathematical terms, providing easy explanations for the underlying relations between explanatory and response variables. Additionally, increasing the complexity of a regression model by including additional terms and interactions among then usually increases the accuracy of the regression for the training data but will also tend to decrease the accuracy of the model when it is used for prediction, affecting the reliability of interpretations of the fitted model (Venables and Dichmont 2004). We assessed the fit of each candidate model using the Akaike Information Criteria (AIC) value (Akaike 1974; Hurvich and Tsai 1989), by comparing all possible combinations using the Akaike weights (AICwi, Anderson et al. 2000). In order to reduce complexity, we have selected the “best model” supported by the data (lower AIC and higher adjusted R2; e.g. Santos et al. 2011). Other strategies, like multi-model inference could be used to calculate the average model to incorporate in the StDM (Bastos et al. 2012). To deal with over-dispersion we used the mean regression function and the variance function from the Poisson GLM, leaving the dispersion parameter unrestricted (quasi-Poisson). Thus, dispersion was not assumed to be fixed at 1 but was estimated from the data. This strategy results in the same coefficient estimates as the standard Poisson model, however, inference is adjusted for over-dispersion (O’Hara and Kotze 2010). Model residuals were assessed using Moran’s I for exploring possible spatial patterns (Moran, 1950). In fact, spatial autocorrelation in GLM residuals could indicate that there are other factors (not included in the statistical models) partially playing a role shaping our predictions (little bustard abundances).

4 – Converting dynamic simulations in spatially explicit dynamic projections

Although the point data considers the combinations of the principal habitats that characterize the sites, the output only represents preliminary independent contributions (Fig 2D). Since the circumscribed dynamic projections neglect spatial relationships, namely possible spatial autocorrelation among point data, a kriging GIS interpolation method (Cressie and Ver Hoef 1993; Sherman 2011) was applied to project and integrate the attributes for each site, by incorporating spatial autocorrelation among land use percentages and little bustard abundances modelled per point (Zhang and Murayama 2011) (Fig.2E). Ordinary kriging interpolation was selected taking into consideration the type of data that resulted from the StDM (i.e. with continuous, non-discrete distribution and the absence of normally distributed data) and the expected spatial relations between point data (Walker et al. 2008). For every time frame selected, this procedure included the adjustment of the semivariogram by a sensitivity analysis of the most relevant model parameters (i.e. nugget, partial sill, lag size and anisotropy) (Dormann 2011). Finally, we tested the fitting of the interpolated results (in relation to the modelled data) by cross validating the predictions of the spatial model with the results from the StDM simulations for each point, extracting a set of spatial indicators related to the land uses distribution/densities and temporal variation of the little bustard predicted abundances. The interpretation of the spatial changes in the landscape and in the species distribution was based on the use of contour surfaces and the corresponding predicted percentages/densities (Fig. 2E): land uses distribution and little bustard densities were calculated considering a continuous distribution function based on a simple kriging and its temporal variation and the distribution area was calculated based in the aggregation of the point referenced data

5 - Dynamic model conceptualization

The initial values for the land use/cover and infrastructural state variables assumed the initial situation in t0 by point (Appendix E, Land use/cover and infrastructures process equations, e.g. ARTIFICIAL\_\_SURFACES (t)). The processes that affect these state variables are described by difference equations (Appendix E, Land use/cover and infrastructures difference equations, e.g. ANNUAL\_INCREASE\_AS). The inflows (e.g. Appendix E, ANNUAL\_INCREASE\_AS) affecting the state variables (e.g. Appendix E, ARTIFICIAL\_\_SURFACES(t)) were built in order to simulate an increase in the area occupied by a land use or length in the case of infrastructures. State variables were also affected by outflows (e.g. Appendix E, ANNUAL\_DECREASE\_AS) to simulate decreases in the area occupied by a land use or length in the case of infrastructures. In order to calculate the annual rate of change, some conversions were introduced, denominated associated variables (Appendix E, Associated variables, e.g. PERIOD\_RATE\_AS). The annual rate of change (Appendix E, Associated variables, e.g. ANNUAL\_RATE\_AS) was calculated using the actual and predicted areas/lengths occupied by land use / cover and infrastructures: linear rates were calculated using the changes detected by comparing the same plots during two time frames, 2003 and 2010 and 2004 and 2010 for Cuba and Airoso respectively. This information can be downloaded from a spreadsheet or inserted manually (Appendix E, Constants). Variables resulting from simple mathematical operations between associated variables or logical and mathematical operations were used to complete the model output and were assumed to be composed variables and other variables (Appendix E – Composed variables and Other variables).

The abundance of males was determined in accordance with the GLM selected explanatory variables (Appendix E - StDM equation). The abundance was calculated in logarithms, defined by a log-link GLM (Appendix E – LN TETRAX). For process compatibilities and a realistic comprehension of model simulations, a conversion was introduced (Appendix E – TETRAX, Associated variables). We obtained this conversion using an inverse transformation (anti-logarithmic), which transforms logarithms back into abundance (Appendix E – TETRAX, Associated variables).

References

Akaike, H. (1974) A new look at the statistical model identification. Automatic Control IEEE Transactions on 19(6): 716-723.

Anderson, D. R., Burnham, K. P. and Thompson, W. L. (2000) Null hypothesis testing: problems, prevalence, and an alternative. The Journal of Wildlife Management 64 (4): 912-923.

Bastos, R., Santos, M., Ramos, J. A., Vicente, J., Guerra, C., Alonso, J., Honrado, J., Ceia, R.S., Timóteo, S. and Cabral, J. A. (2012) Testing a novel spatially-explicit dynamic modelling approach in the scope of the laurel forest management for the endangered Azores bullfinch (Pyrrhula murina) conservation. Biological Conservation 147(1): 243-254.

Cressie, N. O. E. L. and Ver Hoef, J. M. (1993) Spatial statistical analysis of environmental and ecological data. Environmental Modeling with GIS. Oxford University Press, Oxford 404-413.

Dormann, C. (2011) Modeling species’distributions. In: Jopp, F., Reuter, H. and Breckling, B. (Eds.), Modeling Complex Ecological Dynamics. Springer, pp. 179–196.

Elith, J., Graham, C.H., Anderson, R.P., Ferrier, M., Dudik, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC Overton, J., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. and Zimmermann, N.E. (2006) Novel methods improve of species’ distributions from occurrence data. Ecography 29: 129– 151.

Ford, F. A. (1999) Modeling the environment: an introduction to system dynamics modeling of environmental systems. Island press.

Hurvich, C. M. and Tsai, C. L. (1989) Regression and time series model selection in small samples. Biometrika 76(2): 297-307.

Johnston, C. A. (1998) Geographic information systems in ecology. Blackwell Science.

Moran. P.A.P. (1950) Notes on continuous stochastic phenomena. Biometrika 37: 17-23.

Nelder, J. A. and Wedderburn, R. W. (1972) Generalized linear models. Journal of the Royal Statistical Society Series A (General): 370-384.

Neter, J., Kutner, M., Nachtshei, C. and Wasserma, W. (1996) Applied Linear Regression Models, third ed. McGraw-Hill, Boston.

O’Hara, R. B. and Kotze, D. J. (2010) Do not log‐transform count data. Methods in Ecology and Evolution 1(2): 118-122.

Santos, M., Bastos, R. and Cabral, J. A. (2013) Converting conventional ecological datasets in dynamic and dynamic spatially explicit simulations: Current advances and future applications of the Stochastic Dynamic Methodology (StDM). Ecological Modelling 258: 91-100.

Santos, M., Freitas, R., Crespí, A. L., Hughes, S. J. and Cabral, J. A. (2011). Predicting trends of invasive plants richness using local socio-economic data: An application in North Portugal. Environmental research 111(7): 960-966.

Sherman, M. (2011) Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties. Wiley.

Urban, D. L. (2006) Landscape ecology. Encyclopedia of Environmetrics.

Venables, W. N., and Dichmont, C. M. (2004) GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. Fisheries research 70(2): 319-337.

Walker, J., Balling, R., Briggs, J., Katti, M., Warren, P. and Wentz, E. (2008) Birds of a feather: interpolating distribution patterns of urban birds. Computers, Environment and Urban Systems 32: 19–28.

Wisz, M. and Guisan, A. (2009) Do pseudo-absence selection strategies affect geographic predictions of species? A virtual species approach. BMC Ecology 9: 8. doi:10.1186/ 1472-6785-9-8.

Zhang, C. and Murayama, Y. (2000) Testing local spatial autocorrelation using k-order neighbours. International Journal of Geographical Information Science 14(7): 681-692.