# Supplemental Materials

Supplementary Table S1. Main vegetation sub-associations on the BMGR West and their attributes.

Supplementary Appendix S1. Analyzing binary data with the problem of spatial autocorrelation and separation.

**Supplementary Table S1**. **Main** **vegetation sub-associations on the BMGR West and their attributes.** The first digit of the three-digit vegetation sub-association code distinguishes vegetation alliances. The second digit distinguishes vegetation associations, and the third digit, vegetation sub-associations.

†: Used without modification in the analysis

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Vegetation Sub-association Code | Vegetation Sub-association Name | Area (hectares) | Totoal Number of Relevés | Number of Releves with *Brassica tournefortii* Present | Comments |
| 100 | Creosote monotype | 39028 | 41 | 26 | † |
| 110 | Creosote-white bursage | 56943 | 53 | 20 | † |
| 113 | Creosote-fagonia-white bursage on hills | 1350 | 5 | 4 | Not used |
| 115 | Creosote-white bursage-ocotillo on ridges | 1664 | 10 | 0 | † |
| 116 | Creosote-ocotillo-white bursage on plains | 5748 | 9 | 2 | † |
| 117 | Creosote-Spanish needles-white bursage on sands | 10848 | 9 | 6 | † |
| 120 | Creosote-triangle leaf bursage | 5767 | 13 | 4 | † |
| 130 | Creosote-white bursage-triangle leaf bursage | 4102 | 13 | 0 | † |
| 132 | Creosote-white bursage-triangle leaf bursage, burned | 193 | 1 | 1 | Not used |
| 141 | Creosote-white bursage-teddy bear cholla | 3990 | 21 | 1 | † |
| 150 | Creosote-mesquite-triangle leaf bursage floodplain | 835 | 16 | 9 | Combined with 151 and 152 |
| 151 | Creosote-white bursage-blue palo verde floodplain | 1446 | 10 | 7 | Combined with 150 and 152 |
| 152 | Creosote-white bursage playa/floodplain | 250 | 4 | 2 | Combined with 150 and 151 |
| 160 | Creosote-white bursage-big galleta grass | 5508 | 26 | 13 | † |
| 170 | Creosote-triangle leaf bursage-yellow paloverde/ironwood | 1178 | 3 | 0 | Not used |
| 171 | Creosote-white bursage/paloverde-ironwood pavements | 16228 | 20 | 8 | Combined with 177 |
| 175 | Creosote-white bursage/yellow paloverde-ironwood bar/swale | 3549 | 32 | 3 | † |
| 176 | Creosote-white bursage/ironwood-blue paloverde | 3263 | 13 | 7 | † |
| 177 | Creosote-white bursage/ironwood-yellow paloverde | 14956 | 20 | 7 | Combined with 171 |
| 178 | Creosote-white bursage/ironwood-blue paloverde-club cholla | 288 | 4 | 4 | Not used |
| 191 | Creosote-fagonia-white bursage on steep slopes | 4865 | 45 | 6 | † |
| 241 | White bursage-creosote-teddy bear cholla | 767 | 10 | 0 | Combined with 242 |
| 242 | White bursage-creosote-ironwood-teddy bear cholla | 828 | 4 | 0 | Combined with 241 |
| 260 | White bursage-big galleta grass on dunes | 10353 | 32 | 14 | † |
| 261 | White bursage-big galleta grass on fans | 991 | 10 | 10 | † |
| 275 | White bursage/elephant tree on alluvium/pediment | 217 | 7 | 0 | Combined with 276 |
| 276 | White bursage-elephant tree-brittlebush on mountains | 19651 | 51 | 1 | Combined with 275 |
| 280 | White bursage-creosote-ocotillo | 10685 | 25 | 13 | † |
| 291 | White bursage-creosote-yellow paloverde | 1033 | 7 | 5 | † |
| 292 | White bursage-creosote-brittlebush/ironwood | 1270 | 8 | 2 | † |
| 400 | Mormon tea-agave/white bursage | 1165 | 13 | 0 | † |
| 410 | Arrowleaf/sumac-beargreass/mormon tea-lavender | 784 | 11 | 0 | † |
| 631 | Brittlebush-creosote-white bursage/yellow paloverde | 4079 | 28 | 2 | † |
| 670 | Brittlebush-creosote on dark rocks | 1649 | 7 | 3 | † |
| 681 | Brittlebush-white bursage-creosote on fans | 777 | 6 | 2 | Combined with 691 |
| 691 | Brittlebush/ironwood-blue paloverde on fans | 1058 | 9 | 2 | Combined with 681 |
| 710 | Desert holly-white bursage-wandholdback | 60 | 3 | 0 | Not used |
| 800 | Mesquite bosque | 7 | 1 | 1 | Not used |
| 810 | Mesquite/wolfberry-catclaw-cheesebush | 818 | 7 | 5 | † |
| 811 | Ironwood/brittlebush-wolfberry-cheesebush | 2727 | 16 | 7 | Combined with 812 |
| 812 | Ironwood/brittlebush-wolfberry-white bursage | 2274 | 18 | 8 | Combined with 811 |
| 830 | Lavender/hollyleaf bursage-brittlebush | 158 | 12 | 0 | † |
| 900 | Blue paloverde/hollyleaf bursage on mountains | 108 | 3 | 0 | Not used |

# Supplementary Appendix S1. Analyzing binary data with the problem of spatial autocorrelation and separation

## Logistic regressions that account for spatial autocorrelation of the data.

The uneven distribution of relevé locations led to spatial clustering or overdispersion of our data. The resulted spatial autocorrelation may affect the conclusion of logistic regression models with regard to the significance of each explanatory variable and model prediction of invasibility. To evaluate the potential influence of spatial autocorrelation, we used the corrHLfit function in the spaMM package in R (version 2.3.0, Rousset 2018) to perform additional logistic regressions, and compared the results with those inferred by the basic logistic regression models. The spaMM package provides a solution for applying generalized linear mixed models to spatially autocorrelated data. Its use of Laplace or penalized quasi-likelihood (PQL) approximations of likelihood allows for likelihood ratio tests on fixed effects, and introduces less bias in hypothesis testing than the glmmPQL procedure in R, which is often recommended for fitting logistic regression models to autocorrelated data (Rousset and Ferdy 2014).

We treated all data in each regression as one group and modeled the spatial autocorrelation as the Matérn covariance function. We allowed corrHLfit function to estimate two key parameters of the Matérn function: the spatial scale parameter *ρ* and the smoothness parameter *υ*. In this way, the Matérn function can represent a general form of spatial correlation, including the commonly used exponential and squared exponential correlation functions (for *υ*=0.5 and *υ*$\rightarrow \infty $, respectively) (Rousset and Ferdy 2014). A PQL/L method was used for model fit, in which the coefficients of the fixed effects were estimated by PQL, and all dispersion and correlation parameters were estimated based on maximum likelihood (Rousset and Ferdy 2014). To determine the significance of each explanatory variable, we used a likelihood ratio test to compare the full model with an alternative model, in which either of the two explanatory variables was dropped. The maximum likelihood iteration failed to converge on an optimum because of separation in the binary data (six sub-associations had no *Brassica tournefortii* present and one had the species present in all relevés). As a solution, we removed data associated with these sub-associations in the regressions (83 of the 636 original data points).

The corrHLfit function effectively reduced autocorrelation of model residuals (Fig. A1). Likelihood ratio tests based on these logistic regression models gave the same conclusion as that inferred by the basic logistic regression models. Vegetation sub-associations was the only significant factor explaining *B. tournefortii* invasibility (Table A1). Moreover, probability of *B. tournefortii* presence (i.e. invasibility) in each sub-association predicted by this most parsimonious model did not differ from that predicted by the correspondent basic logistic regression model.

**Table A1.** Likelihood ratio tests of logistic regression models (using corrHLfit function in the spaMM package) determine that spatial environments represented by vegetation sub-associations, but not proximity to roads, is significantly associated with the presence/absence of *B. tournefortii*. Each likelihood ratio test compares the full model, in which both explanatory variables were included, with an alternative model, in which one of them was dropped.

|  |  |  |  |
| --- | --- | --- | --- |
| **Models1 in comparison** | **Deviance** | **Degrees of freedom** | **Probability (> χ2)** |
| Full model vs. model where proximity to road was dropped | 0.177 | 1 | 0.674 |
| Full model vs. model where sub-association was dropped | 59.464 | 21 | 1.54×10-5 |

1. All models share the same parameters of the Matérn function of spatial autocorrelation: *υ*=2.95, *ρ* =0.00039.



**Fig. A1.** **A logistic regression model fitted by the corrHLfit function in spaMM package in R effectively reduced residual autocorrelation (B) in comparison to a standard regression where autocorrelation is not modeled (A).** Both models are based on a data subset where data points causing separation of binary data were removed. Presence/absence of *B. tournefortii* is the response variable. Vegetation sub-association and proximity to roads are the explanatory variables. (A) Residual correlation is evident within the first 10 km in a standard logistic model fitted by the glm function in R. (B) Residual autocorrelation is effectively reduced in a model fitted by the corrHLfit function in the spaMM package in R.

Since we removed some of the data to resolve the problem of binary data separation, we performed additional two analyses to confirm that the data subtraction did not alter the conclusion of our inference. In both analyses, we used the full dataset and chose a method that could either handle or avoid the problem of separation of binary data, but were less suited to test our hypothesis. Both analyses gave the conclusion that sub-association, or its surrogate variable, is the only factor significantly associated with *B. tournefortii* presence, supporting the conclusion from the main analysis based on the subtracted dataset. The method and results of these two analyses are described in the following two subsections.

## Applying bias-reduction logistic models

We fitted the full dataset by logistic models based on a bias-reduced estimator (Kosmidis and Firth 2009) instead of the maximum likelihood estimator used by standard logistic regressions. The fit was achieved by using the brglm function in the brglm package (version 0.6.1) in R (Kosmidis 2017). When there is separation in binomial/binary data, model estimates based on the bias-reduced estimator are always finite while those based on the maximum likelihood estimator can be infinite (Kosmidis and Firth 2009). Nevertheless, regressions based on this method cannot model the spatial autocorrelation that existed in the data. We used likelihood ratio tests to compare the full model with alternative models where one of the two explanatory variables was dropped. The tests suggested that sub-association is the only factor significantly associated with *B. tournefortii* presence (Table A2).

**Table A2. Likelihood ratio tests of bias-reduction logistic regression models determine that spatial environments reflected by vegetation sub-association, but not proximity to roads, significantly explaings the presence/absence of *B. tournefortii*.** Each likelihood ratio test compares the full model, in which both explanatory variables were included, with an alternative model, in which one of them was dropped.

|  |  |  |  |
| --- | --- | --- | --- |
| **Models in comparison** | **Deviance** | **Degrees of freedom** | **Probability (> χ2)** |
| Full model vs. model where proximity to road was dropped | 0.70 | 1 | 0.402 |
| Full model vs. model where sub-association was dropped | 190.43 | 28 | < 10-10 |

## Using relevé slope as a surrogate variable of spatial environment

We used relevé slope as a surrogate variable of spatial environment, avoiding the problem of separation of binary data that can arise when using categorical explanatory variables (e.g. sub-association). Environmental factors such as soil type and vegetation are strongly influenced by the terrains over the BMGR West. Rocky, mountainous environments are usually on steep slopes whereas sandy, valley environments are usually on gentle slopes. Therefore, differences in relevé slopes can approximate habitat differences. Nevertheless, it cannot fully represent spatial environments since locations of similar slopes can differ in other biotic and abiotic factors.

 We fitted the original dataset with logistic models using the corrHLfit function. Spatial autocorrelation was models as the Matérn function and the PQL/L method was used for model fit. Relevé slope was measured as the percent slope (the ratio of the rise to the run expressed as a percentage), and was natural log transformed to be used in the regression. 5% was added to all data to handle 0% slopes and to avoid introducing outliers among the log transformed data. Proximity to roads (in meters) was also natural log transformed and used as the other explanatory variable in the regression.

Likelihood ratio tests (performed by the fixedLRT function in the spaMM package) confirm that relevé slope was the only significant factor explaining *B. tournefortii* presence (Table A3). Moreover, the relationship between relevé slope and presence of *B. tournefortii* is negative (*β* = -0.914±0.171), suggesting areas with steeper slopes are less habitable to *B. tournefortii*.

**Table A3. Likelihood ratio tests of logistic regression models (spatial autocorrelation modeled by the Matérn function) determine that relevé slope, but not proximity to roads, significantly explains *B. tournefortii* presence.** Each likelihood ratio test compares the full model with an alternative model, in which one of the explanatory variables was dropped.

|  |  |  |  |
| --- | --- | --- | --- |
| **Models**1 **in comparison** | **Deviance** | **Degrees of freedom** | **Probability (> χ2)** |
| Full model vs. model where proximity to road was dropped | 1.56 | 1 | 0.211 |
| Full model vs. model where relevé slope was dropped2 | 27.96 | 1 | 1.24×10-7 |

1. All models share the same parameters of the Matérn function of spatial autocorrelation: *υ*=1.43, *ρ* =0.00025.

## References

Kosmidis I (2017) brglm: Bias reduction in binomial-response generalized linear models. R package version 061

Kosmidis I, Firth D (2009) Bias reduction in exponential family nonlinear models. Biometrika 96:793–804

Rousset F, Ferdy J (2014) Testing environmental and genetic effects in the presence of spatial autocorrelation. Ecography 37:781–790