Supporting Information

Lawson *et al. "Active management of protected areas enhances metapopulation expansion under climate change"*

Contents

S1: Null model structure

We used the top-fitting models from a previous study (Lawson et al. 2012) as "null models" for the colonization and survival of *Hesperia comma* populations in this study. Terms indicating management and protection status were subsequently added to assess their effects over and above those explained by other variables. The structures of the null models are given below, and incorporate the following variables:

- Patch area in hectares: the area of suitable habitat <10 cm in height was calculated for each patch by calculating the total area of each patch (using digitized patch polygons) and multiplying this figure by the proportion of the patch which contained turf less than 10 cm in height (estimated during field surveys).

- Host plant cover, representing the proportion of turf <10 cm in height which contained suitable *Festuca ovina* host plants.

- Bare ground cover, representing the proportion of turf <10 cm in height which contained bare ground. Bare ground is important for *H. comma* because it heats up more than longer vegetation in direct sunlight and provides warm microclimates for egg-laying (Davies et al. 2006). The survival model presented below includes a squared bare ground term to account for the fact that patches with very high bare quantities of bare ground are less suitable for *H. comma* populations (Lawson et al. 2012).

- Direct connectivity, representing the expected number of adult butterflies to immigrate into the focal patch from surrounding habitat patches.

- Indirect connectivity, representing the connectivity of the focal patch to empty but suitable habitats, which facilitate immigration events.

The two connectivity measures were calculated using the following formula:

$$S_i = \sum_{i \neq j} p_j \exp\left(-\alpha d_{ij}\right) A_j^b$$

Where i is the focal patch and j all other patches, which have area A_j and are separated from i by distance d_{ij}. Here, A_j is effective area <10 cm (ha) of patch j, and d_{ij} edge-to-edge distances between patches i and j (km). α (a negative exponential dispersal kernel) and b (a scaling function for patch emigration) are estimated from a previous study (Wilson et al. 2010). For direct connectivity, p=1 for occupied patches and 0 for unoccupied patches. For indirect connectivity, p_j is calculated is $\sum_{i \neq j} S_j A_j$ (*i.e.* the connectivity of each patch j weighted by its area A_j). The original analysis which demonstrates the importance of the variables in the above variables for the establishment and survival of *H. comma* populations can be found in Lawson et al. (2012).

Colonization

$$logit(c_i) = \alpha + \beta_1 CD_i + \beta_2 CI_i + \beta_3 HO_i$$

$$y_i \sim Bernoulli(c_i)$$

Where c_i indicates the probability of colonizing patch i, CD_i indicates the "direct connectivity" of patch i to patches that were occupied in 2000, and CI_i indicates the "indirect connectivity" of patch I, reflecting the availability of suitable but uncolonised habitat surrounding the patch (see Lawson et al. 2012), and HO indicates the proportion of the patch that was covered by the host plant *Festuca ovina*. y_i is a binary variable indicating whether patch i was colonised between 2000 and 2009.

Survival

$$logit(s_j) = \alpha + \beta_1 CD_j + \beta_2 AR_j + \beta_3 IN_j + \beta_4 MC_j + \beta_5 BG_j + \beta_6 BG_j^2$$

$$z_j \sim Bernoulli(s_j)$$

Where s_j indicates the probability of survival in patch j, CD indicates the "direct connectivity" of patch j to patches that were occupied in 2000, AR indicates the areal extent of the patch, IN indicates the solar index of the patch (a combination of aspect and slope), MC indicates the "macroclimate" of the patch (mean daily August maximum temperature from 2000-2009), and BG indicates the proportion of the patch that was bare ground. z_j is a binary variable indicating whether the population in patch j remained present in 2009.

S2: Investigating the influence of spatial autocorrelation

Spatial autocorrelation can have important effects on the conclusions derived from models of species distribution dynamics (Dormann et al. 2007). We examined our data for spatial autocorrelation, and assessed the spatial scale at which spatial autocorrelation occurs, using a semivariance analysis (Meisel and Turner 1998). Because the analyses presented in this manuscript are based on null models (see section S1) which incorporate spatially-explicit biological processes (e.g. connectivity effects), we used a semivariance analysis to examine the spatial patterns in the residuals of these null models. The results of this analysis are presented in Figure S1.

In the null models for colonization (Fig. S1a), the semivariance reaches a sill (leveling-off point; Meisel and Turner 1998) between 10 and 15 km, possibly extending up to 20km. This indicates that spatial autocorrelation is strongest within this range (most notably at distances within 10km; Fig S1a). For the survival null models (Fig. S1b), a sill is reached within 10km, but there is an apparent spike in semivariance at just below 25km. Note, however, that our dataset contains few pairs of patches separated by this distance range (number of pairs = 91), so our estimate of the semivariance at this point is somewhat uncertain. Overall, we can be confident that spatial autocorrelation in the residuals of the null survival model occurs within 25km. For this reason, we choose to examine the effects of spatial autocorrelation on our conclusions regarding management and protection status at a distances ranging from 5 to 25 km (see below).

To assess the extent to which our results were robust to spatial autocorrelation effects, we repeated our analyses using generalized linear mixed models. Instead of a single intercept for all sites (α in the equations above), we fitted a random intercept which grouped patches within grid squares of sizes ranging from 5km to 25km, using the *lme4* package in *R* (Bates et al. 2011; R Development Core Team 2011). To calculate patch groups, we constructed *R* code (R Development Core Team 2011) which superimposed a grid of each resolution (5km, 10km, 15km, 20km or 25km) across the British

Isles, with the origin (southwest corner with spatial coordinates [0,0]) taken to be the origin of the UK Ordnance Survey National Grid. We then classified patches into grid square groups based on whether they were within the same grid square. As such, the fixed intercept α was replaced with a random intercept:

$$\alpha_k \sim Normal(\mu_\alpha, \sigma_\alpha^2)$$

Where α_k is the intercept for patches in grid cell group k, μ_α is the mean probability of colonization or survival across all patches, and σ_α^2 represents the variance in the probability of colonization or survival among grid cells.

1. Mixed model results: Colonization

Table S3 displays model selection tables for colonization models with and without spatial autocorrelation effects. The rank order of the best models and the direction of predictions remain consistent whether or not a random intercept is used, indicating that the findings of this analysis are robust to spatial autocorrelation.

2. Mixed model results: Survival

Table S4 displays model selection tables for colonization models with and without spatial autocorrelation effects. The finding that primary management improves the probability of survival remains consistent whether or not the effects of spatial autocorrelation are considered (Table S4). However, the evidence that populations were more likely to survive in protected patches weakens once spatial autocorrelation is accounted for, such that the effects of protection status on survival may have been exaggerated by spatial autocorrelation effects.

S3: Detail on method for predicting colonization and survival probabilities

Table 2 (main text) displays predicted colonization and survival probabilities for a patch in each of the management categories. Because our models also incorporated effects of other patch attributes (e.g. patch size and connectivity; Lawson et al. 2012), we needed to choose values for these variables to produce colonization and survival predictions for an "average" patch. We chose to use the mean values based on all patches used in each analysis (colonization and survival). The values of these variables are given in Tables S1 and S2. Note that only variables which entered models (see section S1) are given.

S4: Relationships between management categories and variables in null models

Some management categories were associated with improvements in colonization and survival probabilities (Fig. 1, main text), but had apparently little benefit once important environmental variables were controlled for (Table 1, main text; see section S1 of this supplementary material for environmental variables included in models). To further explore why this might be, we investigated whether management or protection designation was associated with the environmental variables included in our null models. In Figures S2 and S3 we plot the distributions of environmental variables for patches used in (a) the colonization analysis (Fig. S2) and (b) the survival analysis (Fig. S3). In the following paragraphs, we briefly discuss differences in environmental variables among protection and management categories.

1. Colonization data

Amongst sites that were unoccupied by *H. comma* in 2000, patches that were close to existing *H. comma* populations (*i.e.* patches that had higher direct connectivity) and were in more well-connected networks of habitat (i.e. had higher indirect connectivity; see Lawson et al. 2012) tended to be under primary management by conservation bodies, rather than voluntarily managed under agri-environment schemes or unmanaged. Our models therefore suggest that primary management greatly improved colonization chances of patches over and above the benefits of their higher connectivity (Table 1a, main text).

Voluntarily managed sites tended to have higher indirect connectivity than unmanaged sites, which might explain why we found little evidence for positive effects of voluntary management on colonization once the effects of connectivity had been accounted for.

There was no overall tendency for protected sites to have higher connectivity (direct or indirect) or host plant cover than unprotected sites, supporting our conclusion that protected areas improved colonization independently of these variables (Table 1a, main text).

2. Survival data

Amongst habitat patches occupied by *H. comma* in 2000, voluntarily managed sites tended to be larger and in more well-connected habitat networks (higher indirect connectivity) than either managed or unmanaged sites. Thus, patch size and connectivity variables may have played a role in influencing land managers' decisions to "opt-in" to agri-environment schemes (AES), and could explain why we found no positive effect of AES once these variables had been accounted for (Table 1b, main text).

On average, sites with higher direct connectivity were more likely to be protected as Sites of Special Scientific Interest (SSSIs). This reflects the fact that the 2000 distribution of *H. comma* was concentrated around protected areas. Protected sites also tended to be found in warmer regions of Britain (*i.e.* with higher mean August maximum temperatures). Both of these variables may have exaggerated the impacts of protected areas on population survival between 2000 and 2009 (Fig. 1b, main text), explaining why we found only relatively weak evidence that protection enhanced population survival (Table 1b, main text) despite survival in protected areas being higher than in unprotected areas (Fig. 1b, main text).

Supplementary references

Bates D., Maechler M., Bolker B. (2011) lme4: Linear mixed-effects models using S4 classes. *http://CRANR-projectorg/package=lme4*.

Davies Z.G., Wilson R.J., Coles S., Thomas C.D. (2006) Changing habitat associations of a thermally constrained species, the silver-spotted skipper butterfly, in response to climate warming. *Journal of Animal Ecology* **75**, 247-256.

Dormann C.F., McPherson J.M., Araújo M.B. *et al.* (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* **30**, 609-628.

Lawson C.R., Bennie J.J., Thomas C.D., Hodgson J.A., Wilson R.J. (2012) Local and landscape management of an expanding range margin under climate change. *Journal of Applied Ecology* **49**, 552-561.

Meisel J.E., Turner M.G. (1998) Scale detection in real and artificial landscapes using semivariance analysis. *Landscape Ecology* **13**, 347-362.

R Development Core Team. (2011) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Wilson R.J., Davies Z.G., Thomas C.D. (2010) Linking habitat use to range expansion rates in fragmented landscapes: a metapopulation approach. *Ecography* **33**, 73-82.

Supplementary tables

Variable	Value
Direct connectivity	2.0
Indirect connectivity	4.9
Host plant cover (%)	16

Table S1: *values of other environmental variables in colonization models that were used to predict colonization probabilities for habitat patches in different management and protection categories.*

Variable	Value
Direct connectivity	12
Areal extent of patch (ha)	2.3
Solar index	230
Macroclimate (mean daily maximum temperature during August)	22
Bare ground cover (%)	9.4

Table S2: *values of other environmental variables in survival models that were used to predict survival probabilities for habitat patches in different management and protection categories.*

Variables in model	K	δAICc	LL	Predicted colonization probability (%)					
				protected			unprotected		
				primary	voluntary	unmanaged	primary	voluntary	unmanaged
(a) fixed intercept									
primary * protection	7	0.0	-172.2	28.2	5.1	5.1	9.3	8.3	8.3
primary	5	5.8	-177.2	23.2	7.5	7.5	23.2	7.5	7.5
(b) 5km									
primary * protection	8	0.0	-159.4	20.5	2.5	2.5	5.3	4.0	4.0
primary	6	2.9	-162.9	15.9	3.5	3.5	15.9	3.5	3.5
(c) 10km									
primary * protection	8	0.0	-158.1	23.0	2.6	2.6	7.1	5.4	5.4
primary	6	4.6	-162.4	18.5	4.5	4.5	18.5	4.5	4.5
(d) 15km									
primary * protection	8	0.0	-161.2	25.9	3.3	3.3	7.6	6.2	6.2
(e) 20km									
primary * protection	8	0.0	-164.4	24.3	3.9	3.9	6.5	5.8	5.8
voluntary * protection	8	4.5	-166.7	19.5	2.1	19.5	5.5	5.5	5.5
primary + protection	7	5.2	-168.1	21.9	7.6	7.6	12.8	4.2	4.2
(f) 25km									
primary * protection	8	0.0	-157.7	21.4	3.1	3.1	8.0	5.4	5.4
primary	6	2.9	-161.2	17.0	4.6	4.6	17.0	4.6	4.6
voluntary * protection	8	5.7	-160.6	16.9	1.7	16.9	6.2	5.3	6.2

Table S3: Comparison of colonization models with (a) fixed intercept, and models with random intercepts among squares of size (b) 5km (c) 10km (d) 15km (e) 20km and (f) 25km. K=number of parameters; δAICc = difference in AICc between model and top model; LL=log-likelihood. The six right-hand columns show the predicted probabilities of colonization for an "average" patch in each of the different management categories, assuming mean values for other patch attributes. Only models with δAICc ≤6 are shown; models with AICc scores lower than the AICc scores of simpler (nested) models have been excluded (note that in the case of (d), only a single best model exists within the top six AICc units).

Variables in model	K	δAICc	LL	Predicted survival probability (%)					
				protected			unprotected		
				primary	voluntary	unmanaged	primary	voluntary	unmanaged
(a) fixed intercept									
primary + protection	9	0.0	-78.7	91.7	81.8	81.8	83.2	67.0	67.0
primary	8	0.6	-80.1	90.7	77.2	77.2	90.7	77.2	77.2
protection	8	2.3	-80.9	88.7	88.7	88.7	73.7	73.7	73.7
voluntary	8	4.7	-82.1	88.2	75.9	88.2	88.2	75.9	88.2
management	8	4.8	-82.2	87.8	87.8	78.1	87.8	87.8	78.1
null	7	5.1	-83.4	86.1	86.1	86.1	86.1	86.1	86.1
(b) 5km									
primary	9	0.0	-77.7	94.3	81.3	81.3	94.3	81.3	81.3
voluntary	9	4.2	-79.8	92.1	79.5	92.1	92.1	79.5	92.1
protection	9	4.4	-79.8	90.9	90.9	90.9	80.0	80.0	80.0
management	9	4.7	-80.0	91.4	91.4	82.4	91.4	91.4	82.4
null	8	4.8	-81.1	90.1	90.1	90.1	90.1	90.1	90.1
(c) 10km									
primary	9	0.0	-78.8	92.9	80.5	80.5	92.9	80.5	80.5
management	9	3.9	-80.7	90.5	90.5	81.7	90.5	90.5	81.7
null	8	4.0	-81.9	88.9	88.9	88.9	88.9	88.9	88.9
(d) 15km									
primary	9	0.0	-78.5	92.7	79.1	79.1	92.7	79.1	79.1
unmanaged	9	3.7	-80.3	90.3	90.3	80.4	90.3	90.3	80.4
null	8	4.2	-81.7	88.3	88.3	88.3	88.3	88.3	88.3
(e) 20km									
primary	9	0.0	-78.2	92.2	78.6	78.6	92.2	78.6	78.6
voluntary	9	4.0	-80.2	89.8	77.2	89.8	89.8	77.2	89.8
null	8	4.4	-81.5	87.9	87.9	87.9	87.9	87.9	87.9

(f) 25km									
primary + protection	10	0.0	-78.7	91.7	81.7	81.7	83.3	67.0	67.0
primary	9	0.3	-79.7	90.9	76.5	76.5	90.9	76.5	76.5
protection	9	2.3	-80.9	88.7	88.7	88.7	73.7	73.7	73.7
voluntary	9	4.6	-82.1	88.2	75.5	88.2	88.2	75.5	88.2
unmanaged	9	4.7	-82.1	87.8	87.8	77.7	87.8	87.8	77.7
null	8	5.0	-83.4	86.1	86.1	86.1	86.1	86.1	86.1

Table S4: *Comparison of survival models with (a) fixed intercept, and models with random intercepts among squares of size (b) 5km (c) 10km (d) 15km (e) 20km and (f) 25km. K=number of parameters; δAICc = difference in AICc between model and top model; LL=log-likelihood. The six right-hand columns show the predicted probabilities of survival for an "average" patch in each of the different management categories, assuming mean values for other patch attributes. Only models with δAICc ≤6 are shown; models with AICc scores lower than the AICc scores of simpler (nested) models have been excluded.*

Supplementary figures

(a) colonisation
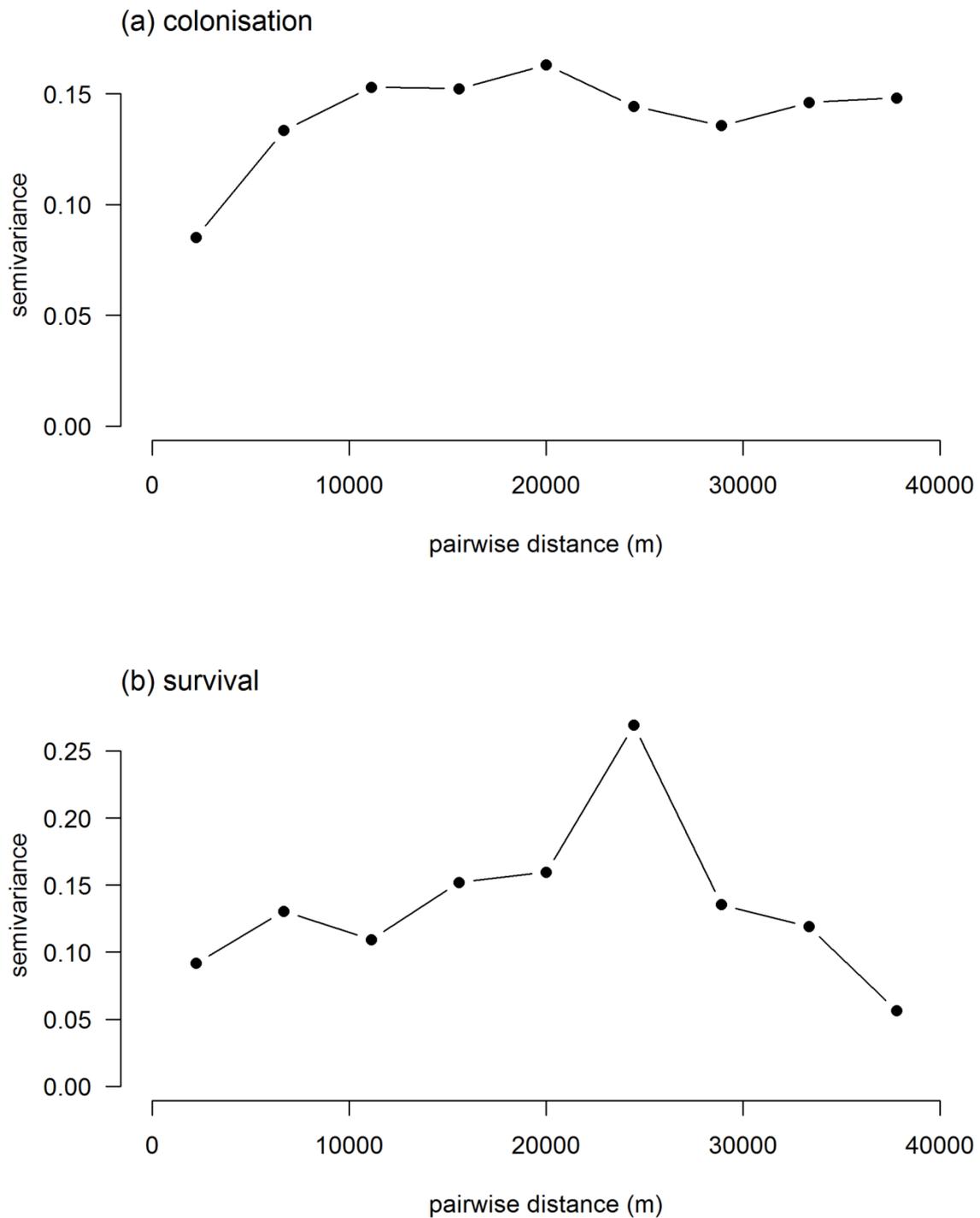
(b) survival

Figure S1: *Semivariograms showing how the variance contributions of the residuals of null models for (a) colonization and (b) survival change with distance between patches. For both colonization and survival model residuals there exists a sill in the semivariance between 10 and 25 km.*
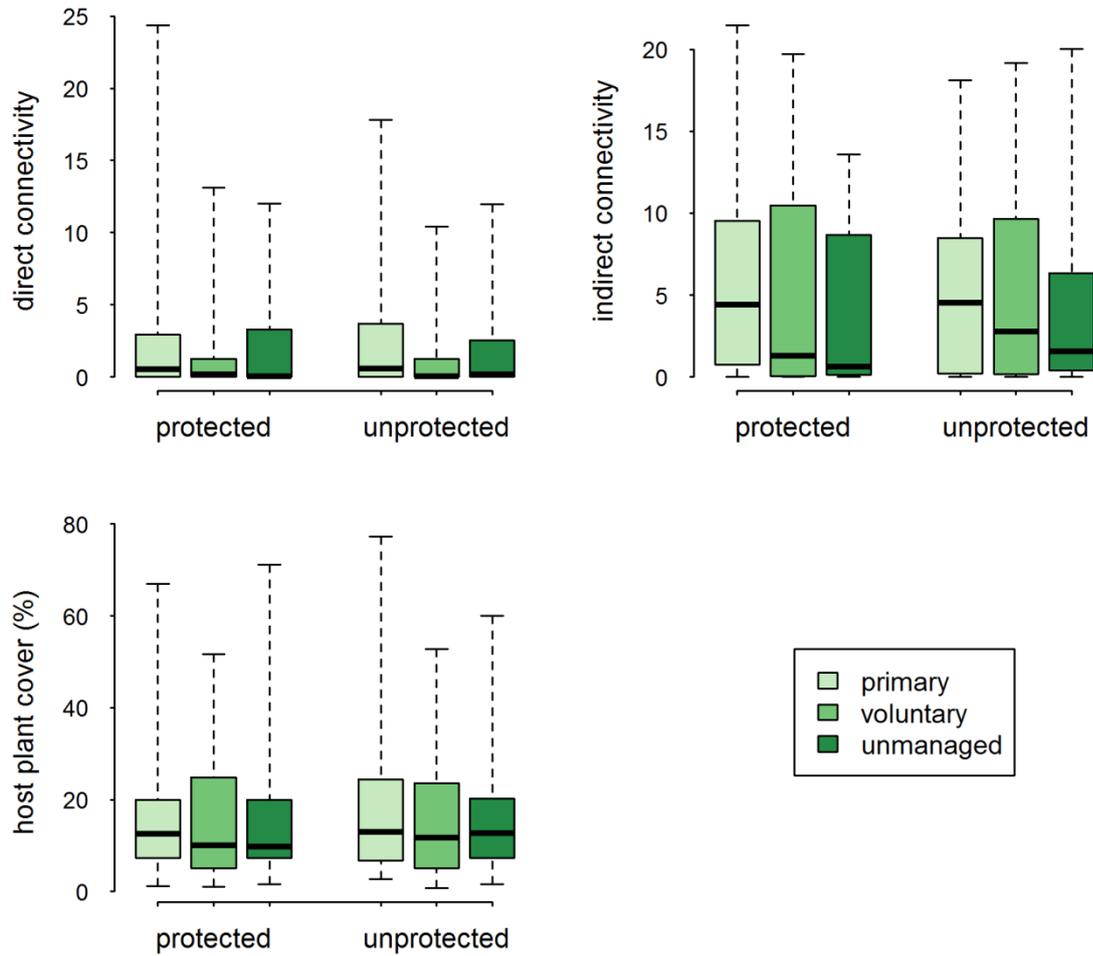
Figure S2: Boxplots showing values of environmental variables for sites in colonization analyses (i.e. for patches that were unoccupied in 2000) in each of the different protection and management categories.
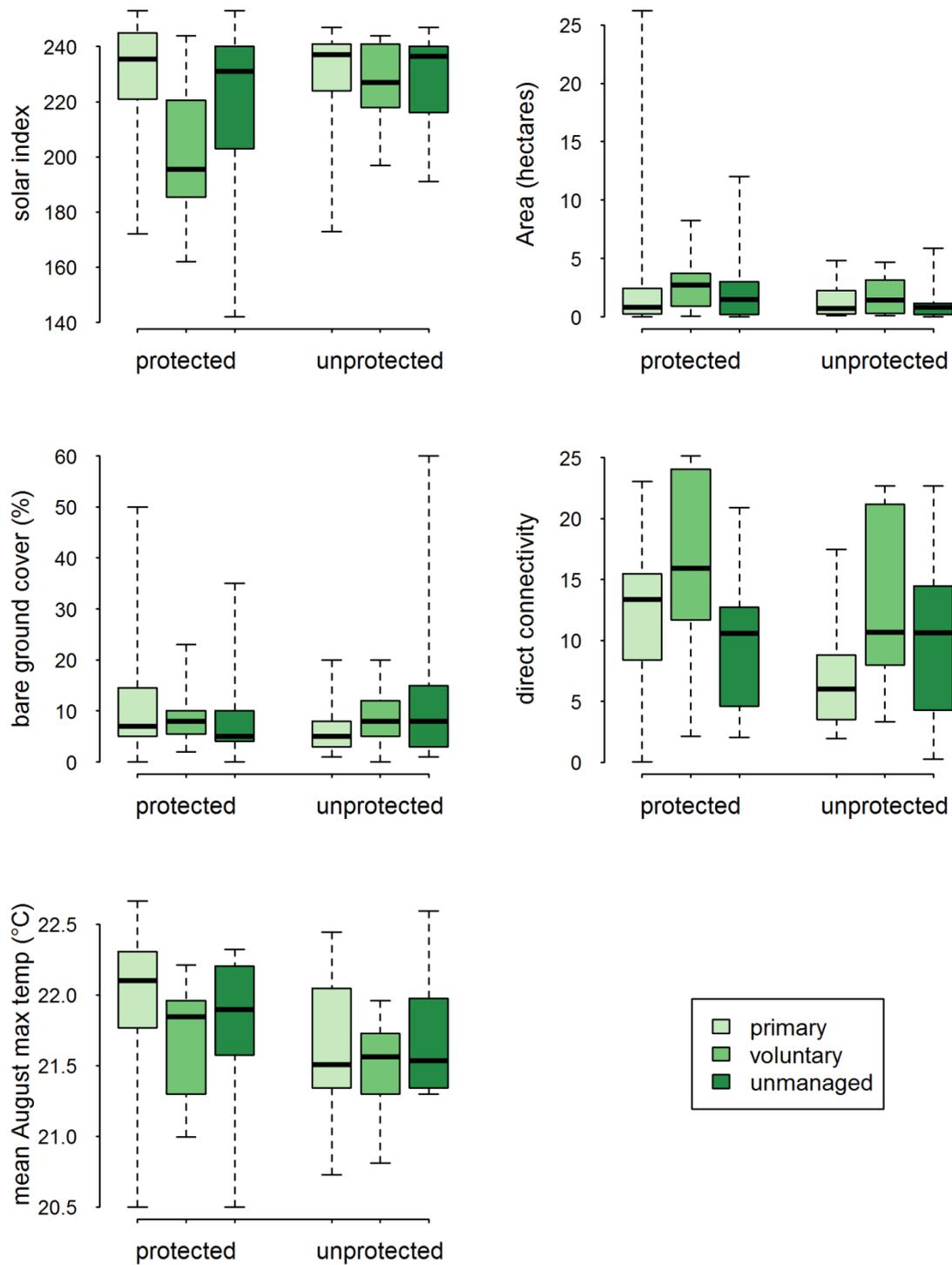
Figure S3: *Boxplots showing values of environmental variables for sites in survival analyses (i.e. for patches that were occupied in 2000) in each of the different protection and management categories.*