

siland: an R package for estimating the spatial influence of landscape

Author details

Florence Carpentier^{1,2}, Olivier Martin³

¹ Université Paris-Saclay, INRAE, AgroParisTech, UMR BIOGER, 78850, Thiverval-Grignon, France

² Université Paris-Saclay, INRAE, UR MaIAGE, 78350, Jouy-en-Josas, France.

³ INRAE, BioSP, 84914, Avignon, France

Corresponding author email : florence.carpentier@inrae.fr

15

16 **Abstract**

17 *Context*

18 The spatial distributions of species and populations are both influenced by local variables and
19 by characteristics of surrounding landscapes. Understanding how landscape features spatially
20 structure the frequency of a trait in a population, the abundance of a species or the species' richness
21 remains difficult specially because the spatial scale effects of the landscape variables are often
22 unknown.

23 *Objectives*

24 Here, we present “siland”, an R package for analyzing the effect of landscape features on
25 georeferenced point observations (described in a Geographic Information System shapefile format).

26 *Methods & Results*

27 “siland” simultaneously estimates the spatial scales and intensities of landscape variable
28 effects. It does not require any information about the scale of effect. Two methods are available:
29 one is based on focal sample site (*Bsiland method*, b for buffer) and one is distance weighted using
30 Spatial Influence Function (*Fsiland method*, f for function). ‘siland’ allows for effects tests, effects
31 maps and models comparison.

32 *Conclusions*

33 Adaptable and user-friendly, the “siland” package is a very practical tool to perform
34 landscape analysis.

35

36

37 **Keywords:** landscape size, spatial scale, Kernel smoothing, buffer, scale of effect, spatial extent,
38 scale of response, multiscales, multiyear, multisites

39

40 **Introduction**

Numerous studies demonstrate that the distribution of species richness and abundance depend on both local and landscape variables (García et al., 2011; Rusch et al., 2016; Remm et al., 2017). However, studying the relationships between landscape and species distributions remains challenging because the shape and the scale of landscape effects are unknown (Miguet et al., 2016) and can be missed if assessed at an incorrect scale (Smith et al., 2011). The studied data usually contains georeferenced observations at point sites, named response variables hereafter, and the description of several landscape spatial variables, named landscape variables hereafter. Their studies are often referred to as focal patch studies (Thornton et al., 2011). To identify the scale of landscape variable effects, the common approach consists of the following: (i) *a priori* defining a set of scales; (ii) creating summary variables by computing measures of the landscape variables within discs or rings of radii equal to each scale centered on the observation sites (named buffers hereafter); and (iii) applying a regression model to the response variable with the summary variables as the explanatory variables, for example, a linear model or a random forest algorithm (Bradter et al., 2012). For each landscape variable, the scale of effect is then considered to be the size of the buffer best explaining the response variable.

The main disadvantage of this method is that the number of explanatory variables artificially increases with the number of spatial scales considered. One then faces a complex statistical dilemma, which is dealing with numerous explanatory variables that by their construction are highly correlated. Consequently, the potential scales chosen are often too few and their ranges are too limited (Jackson & Fahrig, 2015). Finally, the effect of a landscape variable is modelled as uniform within the buffer and as null outside it (Chandler & Hepinstall-Cymerman, 2016), which is unrealistic and biologically unjustified as a continuously decreasing effect is expected (Moilanen & Hanski, 2001). Several new methods based on distance-weighted effects have been proposed to model a distance-decreasing effect (Aue et al., 2011; Henry et al., 2012; Serckx et al., 2016), but they explore a limited predefined set of spatial scales for predictors. Other methods quantified the scale of landscape effects without an *a priori* choice of spatial scales (Walsh & Webb, 2014;

Chandler & Hepinstall-Cymerman, 2016). However, none of these methods are yet implemented in a ready-to-use software. Huais (2018) proposed a very convenient R function “multifit” to select scales but with some limitation (it is not distance weighted, requires the choice of a set of scales), while calling for further developments of such a method to generalise this type of automated analysis.

Here, we present “siland”, a package for the R statistical computing environment dedicated to landscape effect analysis. Two methods are available. Both estimate the scales of effect of each landscape variable using maximum likelihood estimation. In the first method, Bsiland method hereafter (B for buffer), the effect of a landscape variable is modeled as in classic methods based on focal sample site, i.e. considered as constant over a disc centered on the observation point. But contrarily to the previous method, it does not require a first definition of tested radii since the optimal buffer radii are estimated. In the second method the effect of landscape is based on a weighted distance, as in the framework proposed by Chandler & Hepinstall-Cymerman (2016). The decrease in weight with distance is modeled by a Spatial Influence Function (SIF). The parameter of the SIF defined the scale of effect of a landscape variable. In this second method, named Fsiland (F for function), the parameters of the SIFs are estimated for each landscape variable. The main functions of siland allow the user (i) to estimate the intensity of local and landscape effects and the scale parameter of each landscape variable, (ii) to test these effects and (iii) to plot landscape effects on maps. We exemplified the package use by analysing the landscape effects of conventional and organic orchards on the density of codling moth larvae per apple tree.

Models and methods

We consider a response variable measured at n different sites denoted Y_i (i stands for a site), L local variables which can be continuous or discrete and are denoted as x_i^l (l stands for a local variable and i for a site) and K landscape variables denoted as z_r^k (k stands for a landscape variable and r for a

polygon in the landscape). In the Bsiland method, the effect of landscape variables is modelled using buffers with p_{i,δ_k}^k , the percentage of the landscape variable k in a buffer of radius δ_k and centered on site i . Since the Bsiland model is based on the generalized linear model framework of generalized linear models, the expected value of the response variable Y_i is modelled as follows :

$$\mu_i = \mu + \sum_{l \in L} \alpha_l x_i^l + \sum_{k \in K} \beta_k p_{i,\delta_k}^k$$

where μ is the intercept, α_l and β_k are the effects of local and landscape variables, respectively. All parameters, μ , $\{\alpha_l, \dots, \alpha_K\}$, $\{\beta_l, \dots, \beta_K\}$ but also $\{\delta_l, \dots, \delta_K\}$ radii of the buffers of the landscape variables are simultaneously estimated by likelihood maximization.

The Fsiland method is based on Spatial Influence Functions (SIFs) in a similar framework to Chandler & Hepinstall-Cymerman's (2016). The entire study area is rasterized, *i.e.* pixelated on a regular grid, named R . The value of each landscape variable k at a pixel r is described in z_r^k . For instance, if the landscape variable k is a presence/absence variable, z_r^k is equal to one or zero. The

$$\mu_i = \mu + \sum_{l \in L} \alpha_l x_i^l + \sum_{k \in K} \beta_k \sum_{r \in R, r \neq i} f_{\delta_k}(d_{i,r}) z_r^k$$

expected value of the response variable Y_i is then modelled as follows :

where $f_{\delta k}(\cdot)$ is the **SIF** associated with the landscape variable k and $d_{i,r}$ is the distance between the center of pixel r and the observation at site i . The SIF is a density function decreasing with the distance. The scale of effect of a landscape variable k is calibrated through the parameter δ_k , the mean distance of f_{δ} . Two families of SIF are currently implemented in the siland package, exponential and Gaussian families defined as $f_{\delta}(d)=2/(\pi\delta^2)\exp(-2d/\delta)$ and $f_{\delta}(d)=1/(2\delta\sqrt{\pi})\exp(-d^2/2\delta^2)$, respectively (Austerlitz et al., 2004). The effect of a landscape variable k is modelled by two parameters : an intensity parameter, β_k describing its strength and its direction and a scale parameter, δ_k , describing how this effect declines with distance. Each pixel potentially has an effect on the response variable at any observation site. No set of scales of effects is initially determined.

Package description

The siland package is written entirely in the scientific computing language R (R Core Team, 2019). It is available on CRAN (<https://cran.r-project.org/web/packages/SILand/index.html>) and new developments are available on <https://github.com/silandpackage/siland>. The analyses presented here were performed using the package siland 2.0.

Case study

We illustrated the abilities of siland on an example previously described and analyzed in Ricci et al. (2009) : the study of codling moth densities, an insect pest specialized on apple orchards in the Basse Durance Valley in southeastern France (see figure 1). The datasets can be extracted from the package using the commands `data(dataCmoth)` and `data(landCmoth)`. The complete analysis script and outputs are available in Supplementary Information (SI 1 and SI 2, respectively). `dataCmoth` is a data frame with two columns named `x` and `y` containing the observation locations, a column `Cmoth`, containing the response variable of the study (the mean number of codling moths in the orchards), and a column `trait`, describing a local variable (the number of insecticide treatments applied in the orchard). `landCmoth` is a `sf` object (from package `sf` (Pebesma,2018)) describing the landscape variables : conventional tree orchards (`conv`), organic tree orchards (`org`) where `conv` is equal to 1 if the land use is associated to orchard with conventional practice and 0 otherwise, and so is it for `org`.

Main functions

The main functions are described in Table 1, while a full description of package functions is available at <https://cran.r-project.org/web/packages/siland/siland.pdf>. The package siland requires two objects containing data. The first object is a data frame composed by two columns named `X` and `Y` containing the observation locations, a third column representing the response variable and eventually other columns representing local variables. The second object is a `sf` object describing landscape variables. It can be obtained directly from shape files of landscape map by using the

function `st_read()` of the package `sf` (see `vignette(siland)` for more details).

Model estimation is performed using the function `Bsiland()` for `Bsiland` method and the function `Fsiland()` for `Fsiland` method. The arguments of the both functions are similar : formula of the model, `land` the `sf` object describing landscape variables and `data`, the observation data frame. The syntax of the formula “`y ~ model`” is similar to `lm()` function of the `stat` package. In the model term of the formula, the explanatory variables are added using the symbol “`+`”. The explanatory variables described in the data frame `data` are considered as local in the model, those described in the `sf` object `land` are considered as landscape variables. Local effects can be modelled as fixed or random (using the syntax `(1|x)`, see `vignette(siland)` for more details). Interaction terms can be considered using the usual symbols “`*`” or “`:`”. Notice that only interactions between local x local and local x landscape variables are considered. Various types of response variables can be considered using the argument `family` which describes the assumed distribution of the response variable and can take the values “`gaussian`”, “`poisson`” and “`binomial`” for data of continuous, counting or presence-absence types, respectively (and their associated link functions identity, log and logit respectively).

Using the argument `border`, the spatial effect of landscape variable can be considered from the observation locations (`border=F`) or from the border of the polygon where observations are located (`border=T`) (see Figure 2). For `Fsiland()`, the additional argument `sif` indicates the family of SIF chosen “`exponential`” (by default) or “`gaussian`”.

The object returned by the functions `Bsiland()` and `Fsiland()` displays the parameters estimation and a test of the global landscape effect. The function `summary()` applied on the result object provides significance tests of the intensity of the effect of explanatory variables (local or landscape).

The two methods are based on likelihood maximization. The functions `Bsiland.lik()` and `Fsiland.lik()` aim to point out some optimization problems. They provide representations of the minus loglikelihood in function of buffer radius or mean distance of the SIF, respectively (see SI 1).

Values of minus loglikelihood lower than the estimated one indicate that the estimation procedure did not perform correctly. In such cases, the estimation needs to be reiterated using different initial values of effects scales (with argument `init` of functions `Fsiland()` and `Bsiland()`).

Graphics outputs

The package `siland` proposes various functions for graphic representation of estimated landscape effects. The function `plotFsiland.land` displays landscape effect map for the estimations of the `Fsiland` method (Figure 3). The influence of each landscape variable is plotted over all the study area. The global influence of the landscape, i.e. the sum of the effects of all landscape variables can also be plotted. The function `plotBsiland.land()` displays landscape effect map for the estimation of the `Bsiland` method (Figure 2). For each landscape variable, the estimated area of effect (buffer of estimated radius) is plotted around each observation locations. The color of the area represents the intensity of the effect.

Interpreting spatial influence functions (SIF)

The SIF function describes how the influence of a pixel/cell of a landscape variable is spatially distributed. We assume that the influence is maximal at the pixel location and decreases with the distance. The greater the estimated mean distance of the SIF, the greater the scale of effect of the landscape variable. The estimated SIF can be displayed using the function `plotFsiland.sif()` (see Figure 1). The function `quantile.sif()` allows to quantify the area of medium influence and significant influence of a landscape variable, that we defined as the disc containing 50% and 95% of the influence of the SIF (neglecting 50% and 5% of its broader effect) respectively. They can be compared to the landscape variables distribution in the study area using function `plotFsiland()` (see Figure 1).

Conclusion

Estimating scales of effect of landscape variables currently remains a great challenge, and

consequently so does the estimation of landscape effect. So far no common methodology has emerged. With the siland package, we propose a tool that we believe is useful for all landscape ecologists who wish to investigate this type of question. Here, we have illustrated how with only few siland functions and limited knowledge of R, it is possible to conduct a reproducible and detailed study of multi-scale landscape effect including estimations but also tests and graphs illustrating the results obtained. The “siland” package is very adaptable, it integrates two methods and can handle various types of data. Applications of siland in a multiyear or multisite framework is presented for the buffers method in the `vignette(siland)`. However future developments are still needed to handle the increasing complexity of questions and data (numerous sites, years and dynamic data).

Authors' contributions

F.C. and O.M. conceived the ideas and designed methodology, programmed the R package. All authors contributed critically and equally to the drafts and gave final approval for publication.

Acknowledgements

We are grateful to Claire Lavigne and Pierre Franck for their case study, which illustrates the use of the SILand package, and for their useful advice. We are thankful to Etienne Klein for his relevant comments. This work was financially supported by the Metaprogram SMACH (Sustainable Management of Crop Health; <http://www.smach.inra.fr/>) of the French National Agronomic Institute (INRA).

References

- 222 Aue, B., Klemens, E., Stefan, H., & Volkmar, W. (2011). Distance weighting avoids erroneous
223 scale effects in species–habitat models. *Methods in Ecology and Evolution*, 3(1), 102–111.
224 doi:10.1111/j.2041-210X.2011.00130.x
- 225 Austerlitz, F., Dick, C. W., Dutech, C., Klein, E. K., Oddou–Muratorio, S., Smouse, P. E., & Sork,
226 V. L. (2004). Using genetic markers to estimate the pollen dispersal curve. *Molecular*
227 *Ecology*, 13(4), 937–954. doi:10.1111/j.1365-294X.2004.02100.x
- 228 Bradter, U., Kunin, W. E., Altringham, J. D., Thom, T. J., & Benton, T. G. (2012). Identifying
229 appropriate spatial scales of predictors in species distribution models with the random forest
230 algorithm. *Methods in Ecology and Evolution*, 4(2), 167–174. doi:10.1111/j.2041-
231 210x.2012.00253.x
- 232 Chandler, R., & Hepinstall-Cymerman, J. (2016). Estimating the spatial scales of landscape effects
233 on abundance. *Landscape Ecology*, 31(6), 1383–1394. doi:10.1007/s10980-016-0380-z
- 234 García, D., Zamora, R., & Amico, G. C. (2011). The spatial scale of plant–animal interactions:
235 effects of resource availability and habitat structure. *Ecological Monographs*, 81(1), 103–
236 121. doi:10.1890/10-0470.1
- 237 Henry, M., Fröchen, M., Maillet-Mezeray, J., Breyne, E., Allier, F., Odoux, J.-F., & Decourtye, A.
238 (2012). Spatial autocorrelation in honeybee foraging activity reveals optimal focus scale for
239 predicting agro-environmental scheme efficiency. *Ecological Modelling*, 225, 103–114.
240 doi:10.1016/j.ecolmodel.2011.11.015
- 241 Huais, P. Y. (2018). multifit: an R function for multi-scale analysis in landscape ecology.
242 *Landscape Ecology*, 33(7), 1023-1028. doi:10.1007/s10980-018-0657-5.
- 243 Jackson, H. B., & Fahrig, L. (2015). Are ecologists conducting research at the optimal scale?
244 *Global Ecology and Biogeography*, 24(1), 52–63. doi:10.1111/geb.12233
- 245 Miguet, P., Jackson, H. B., Jackson, N. D., Martin, A. E., & Fahrig, L. (2016). What determines the
246 spatial extent of landscape effects on species? *Landscape Ecology*, 31(6), 1177–1194.
247 doi:10.1007/s10980-015-0314-1

248 Moilanen, A., & Hanski, I. (2001). On the use of connectivity measures in spatial ecology. *Oikos*,
249 95(1), 147–151. doi:10.1034/j.1600-0706.2001.950116.x

250 Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. *The R*
251 *Journal* 10 (1), 439-446, doi: 10.32614/RJ-2018-009

252 R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria:
253 R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org>

254 Remm, J., Hanski, I. K., Tuominen, S., & Selonen, V. (2017). Multilevel landscape utilization of
255 the Siberian flying squirrel: Scale effects on species habitat use. *Ecology and Evolution*,
256 7(20), 8303–8315. doi:10.1002/ece3.3359

257 Ricci, B., Franck, P., Toubon, J. F., Bouvier, J. C., Sauphanor, B., & Lavigne, C. (2009). The
258 influence of landscape on insect pest dynamics: a case study in southeastern France.
259 *Landscape ecology*, 24(3), 337-349. doi: 10.1007/s10980-008-9308-6

260 Rusch, A., Chaplin-Kramer, R., Gardiner, M. M., Hawro, V., Holland, J., Landis, D., ...
261 Bommarco, R. (2016). Agricultural landscape simplification reduces natural pest control: A
262 quantitative synthesis. *Agriculture, Ecosystems & Environment*, 221, 198–204.
263 doi:10.1016/j.agee.2016.01.039

264 Serckx, A., Huynen, M.-C., Beudels-Jamar, R. C., Vimond, M., Bogaert, J., & Köhl, H. S. (2016).
265 Bonobo nest site selection and the importance of predictor scales in primate ecology.
266 *American Journal of Primatology*, 78(12), 1326–1343. doi:10.1002/ajp.22585

267 Smith, A. C., Fahrig, L., & Francis, C. M. (2011). Landscape size affects the relative importance of
268 habitat amount, habitat fragmentation, and matrix quality on forest birds. *Ecography*, 34(1),
269 103–113. doi:10.1111/j.1600-0587.2010.06201.x

270 Thornton, D. H., Branch, L. C., & Sunkist, M. E. (2011). The influence of landscape, patch, and
271 within-patch factors on species presence and abundance: a review of focal patch studies.
272 *Landscape Ecology*, 26(1), 7–18. doi:10.1007/s10980-010-9549-z

273 Walsh, C. J., & Webb, J. A. (2014). Spatial weighting of land use and temporal weighting of

274 antecedent discharge improves prediction of stream condition. *Landscape Ecology*, 29(7),
275 1171–1185. doi:10.1007/s10980-014-0050-y

276

277 Table 1 **Main functions of siland package. Detailed information about these functions are**
 278 **given in <https://cran.r-project.org/web/packages/siland/siland.pdf>.**

Functions for buffer approach	Functions for approach with FIS	Description
Bsiland	Fsiland	Estimation of model parameters for local and landscape variables and for scale of landscape effects
Bsiland.lik	Fsiland.lik	Graphical representation of profiled likelihood function of scales of effect of landscape variables
plotBsiland.land	plotFsiland.land	Graphical representation of landscape effects
	plotFsiland.sif	Graphical representation of the spatial influence functions
	summary	As for functions lm() and glm(), summary gives estimates and values for tests.
	AIC	AIC value of the estimated model
	BIC	BIC value of the estimated model
	residuals	Vector of residual values for the estimated model
	fitted	fitted values of the estimated model

279

280

281

282

FIGURES

Fig. 1: Map of the observations in the study site and the estimated spatial influence functions (SIFs)

The figure A is obtained with `plotFsiland(resF, landCmoth, data=dataCmoth)`. Black points represent locations of response variable observations. Yellow and blue squares are the pixels where conventional and organic orchards respectively are present. At the right margin, the light and dark discs represent area of medium influence and significant influence, respectively. Blue and yellow discs represent conventional and the organic orchards, respectively. The figure B is obtained with `plotFsiland.sif(resF)`. The blue and orange lines represent the estimated SIF for the organic and conventional orchards, respectively. The vertical lines represent the SIF median.

Fig. 2: Maps of the predicted effect estimated by the Bsiland method obtained with the function `plotBsiland.land()`

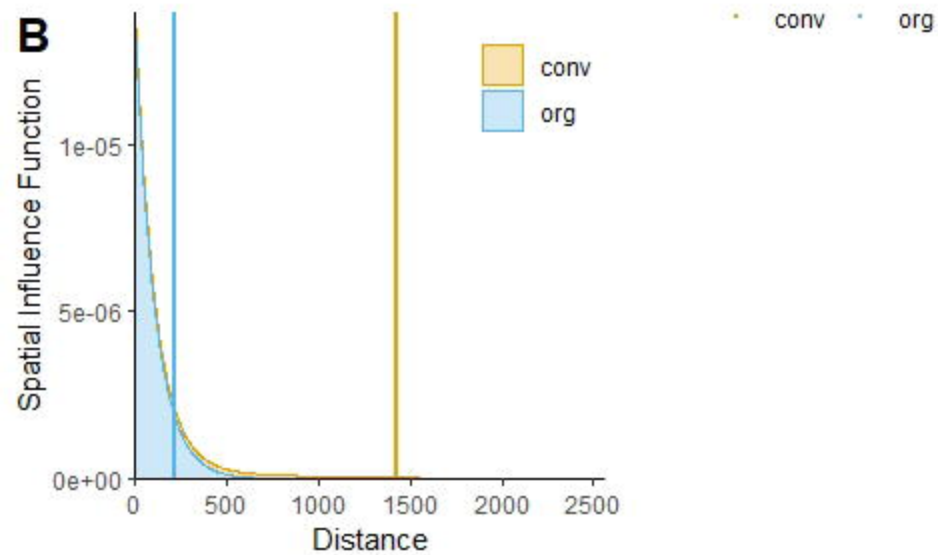
The buffer model was estimated using the command : `resB2=Bsiland(Cmoth~trait+conv+org, land=landCmoth, data=dataCmoth, border=border)`. The colored area represents the buffer around each observation location estimated for the organic orchards effect. At the bottom margin, the bar color gives the effect intensity. Buffers are modeled in graphic A, from the observation locations (`border=FALSE`) and in graphic B from the border of the orchard of each observation (`border=TRUE`).

Fig. 3: Maps of the predicted effect estimated by the Fsiland method obtained with the function `plotFsiland.land()`

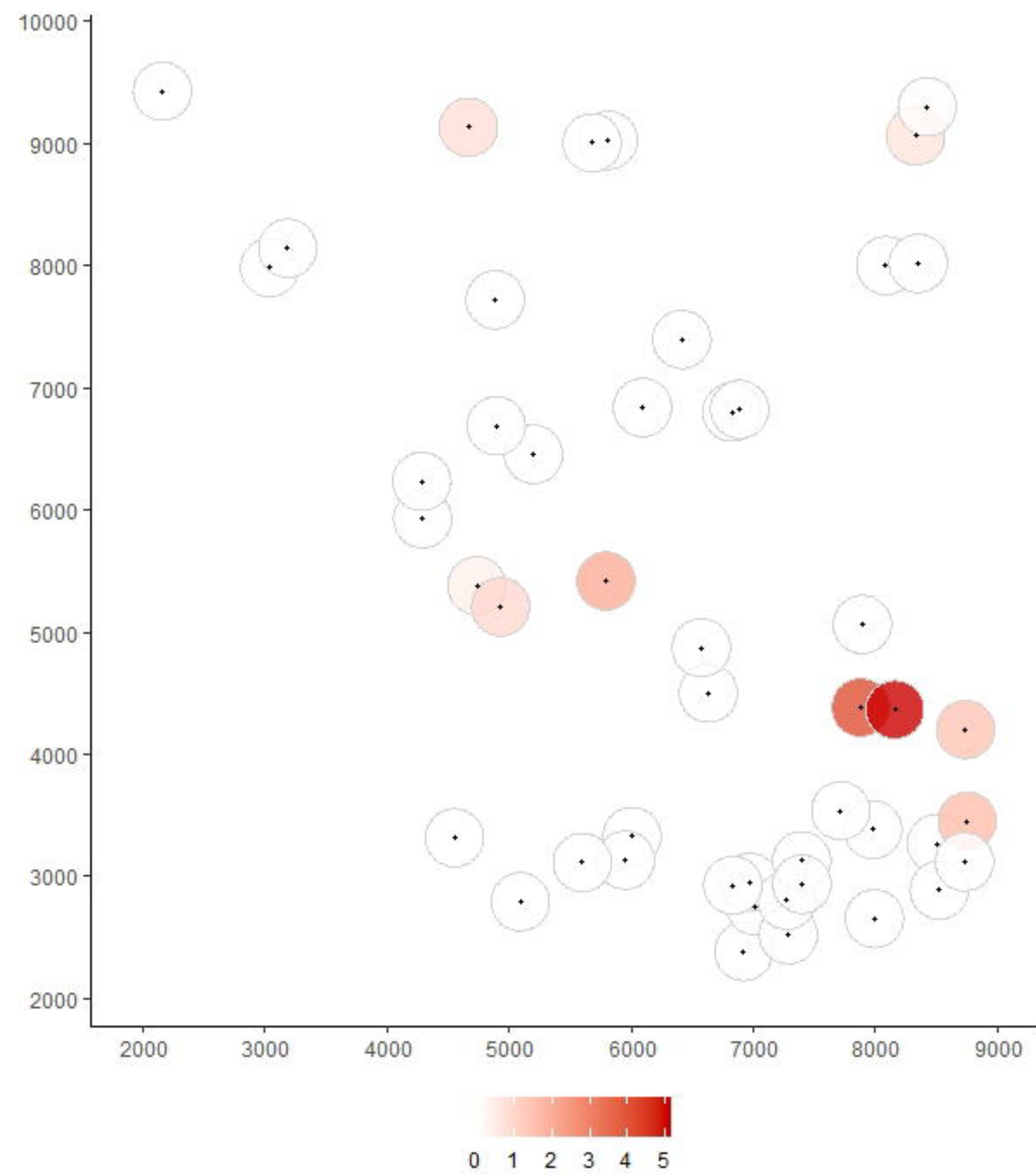
The SIF model was estimated using the command : `resF = Fsiland(Cmoth ~ trait+conv+org, land=landCmoth, data=dataCmoth)`. The three maps A, B and C were

309 obtained with the command `plotFsiland.land(resF, land=landCmoth,`
 310 `data=dataCmoth, var=var), var` equal to 1,2 and 0 respectively. The black points
 311 represent locations of response variable observations. In graphics A and B, the gray points are the
 312 pixels where conventional and organic orchards are present, respectively. The response surface
 313 represents the cumulative effect of the conventional orchards in graphic A, the organic orchards in
 314 graphic B and the global landscape in graphic C. At the bottom margin, the bar color gives the
 315 effect intensity. In the figure A, conventional orchards had negative effects at large scale. In figure
 316 B, organic orchards had positive effects positive effects at small scale. In figure C, global
 317 landscape had an overall negative effect except on spotty areas.

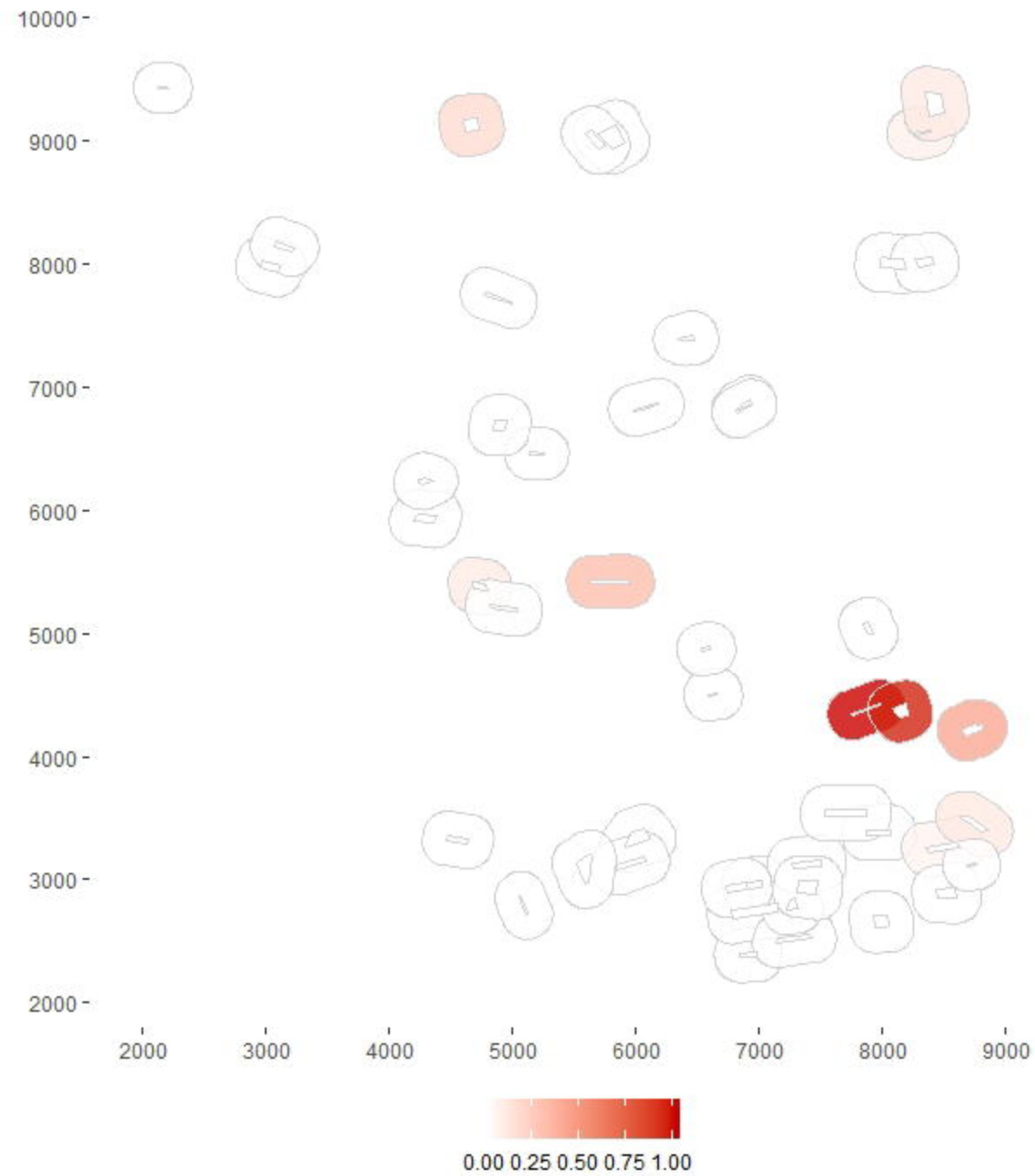
318



A



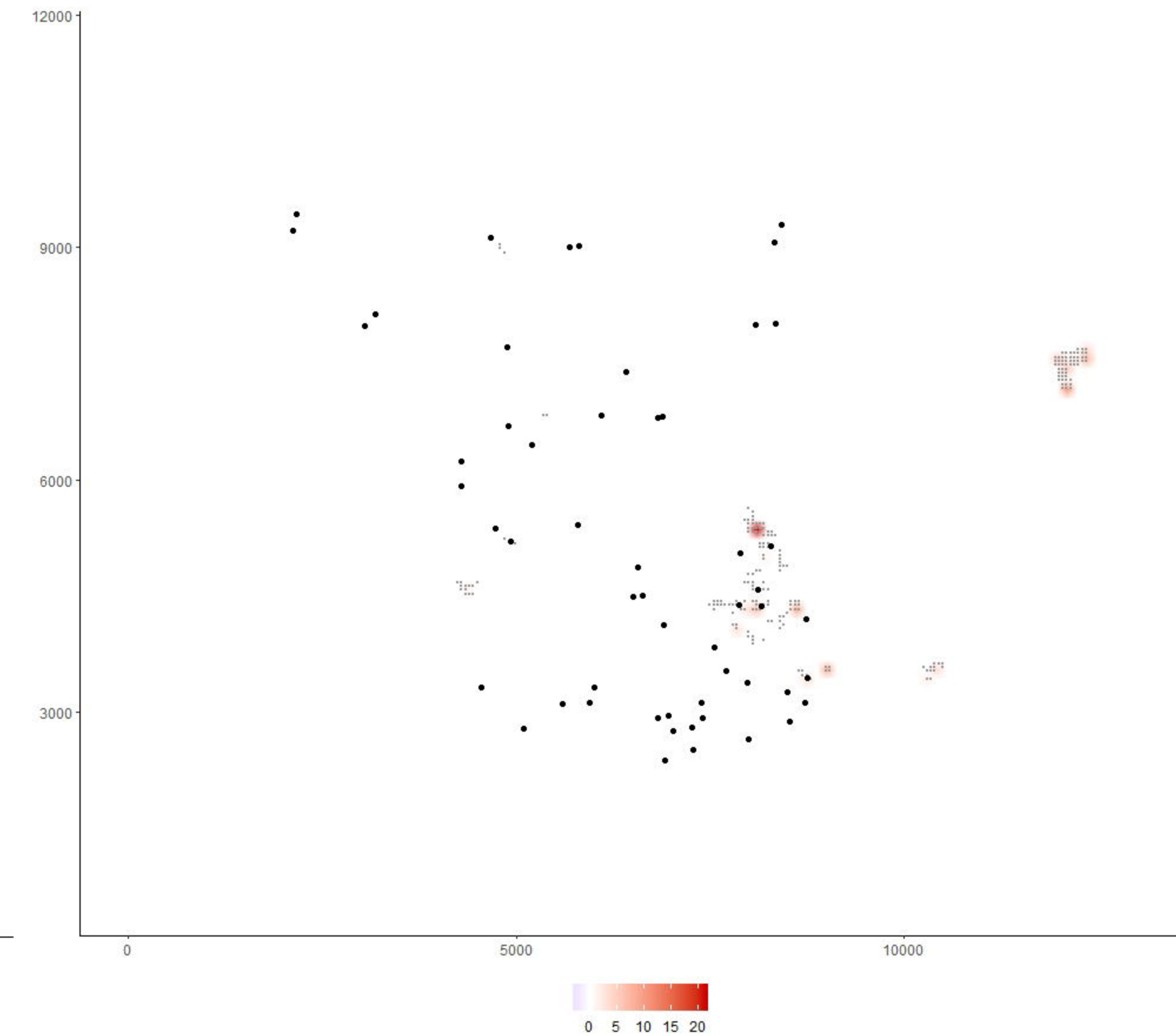
B



A



B



C

