

1 **Mapping and explaining wolf recolonization in France using dynamic occupancy models and**  
2 **opportunistic data**

3

4 Julie Louvrier<sup>1,2</sup>, Christophe Duchamp<sup>2</sup>, Eric Marboutin<sup>3</sup>, Sarah Cubaynes<sup>1</sup>, Rémi Choquet<sup>1</sup>, Christian  
5 Miquel<sup>4</sup>, Olivier Gimenez<sup>1</sup>

6 1. CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valéry Montpellier, EPHE,  
7 Montpellier Cedex 5, France

8 2. Office National de la Chasse et de la Faune Sauvage, CNERA prédateurs et animaux déprédateurs,  
9 Parc Micropolis, 05000 Gap, France

10 3. Office National de la Chasse et de la Faune Sauvage, ZI Mayencin, 5 allée de Béthléem, 38610  
11 Gières, France.

12 4. Université Grenoble 1, CNRS, Laboratoire Ecologie Alpine, UMR 5553, 38041 Grenoble 9, France

13

14

15 Corresponding author: Julie Louvrier, Centre d'Écologie Fonctionnelle et Évolutive, CNRS, 1919 route  
16 de Mende, 34090 Montpellier, France. E-mail address: [julie.louvrier@cefe.cnrs.fr](mailto:julie.louvrier@cefe.cnrs.fr)

17 **Abstract**

18 While large carnivores are recovering in Europe, assessing their distributions can help to predict and  
19 mitigate conflicts with human activities. Because they are highly mobile, elusive and live at very low  
20 density, modeling their distributions presents several challenges due to i) their imperfect detectability,  
21 ii) their dynamic ranges over time and iii) their monitoring at large scales consisting mainly of  
22 opportunistic data without a formal measure of the sampling effort. Not accounting for these issues can  
23 lead to flawed inference about the distribution.

24 Here, we focused on the wolf (*Canis lupus*) that has been recolonizing France since the early 90's. We  
25 evaluated the sampling effort *a posteriori* as the number of observers present per year in a cell based on  
26 their location and professional activities. We then assessed wolf range dynamics from 1993 to 2014,  
27 while accounting for species imperfect detection and time- and space-varying sampling effort using  
28 dynamic site-occupancy models.

29 Ignoring the effect of sampling effort on species detectability led to underestimating the number of  
30 occupied sites by 50% on average. Colonization increased with increasing number of occupied sites at  
31 short and long-distances, as well as with increasing forest cover, farmland cover and mean altitude.  
32 Colonization decreased when high-altitude increased. The growth rate, defined as the number of sites  
33 newly occupied in a given year divided by the number of occupied sites in the previous year, decreased  
34 over time, from over 100% in 1994 to 5% in 2014. This suggests that wolves are expanding in France  
35 but at a rate that is slowing down. Our work shows that opportunistic data can be analyzed with species  
36 distribution models that control for imperfect detection, pending a quantification of sampling effort.  
37 Our approach has the potential for being used by decision-makers to target sites where large carnivores  
38 are likely to occur and mitigate conflicts.

39

40 **Key words**

41 *Canis lupus*, gray wolf, large carnivores, occupancy models, opportunistic data, sampling effort, species  
42 detectability, species distribution models, recolonization

43

## 44 Introduction

45 Large carnivores are often considered as key elements for maintaining ecosystems. Because of  
46 their high position in the trophic chain, their extinction can lead to trophic cascades and detrimental  
47 changes in species abundance and functioning of ecosystems (Ripple et al. 2014). Once widespread in  
48 Europe, many populations of large carnivores were extirpated over the last century, mainly due to  
49 interferences with human activities (Breitenmoser 1998, Ripple et al. 2014). Since the 1970s, all large  
50 carnivores have recovered, benefiting from legal protection and the recovery of wild ungulate  
51 populations, resulting in most of the European countries hosting at least one viable population of large  
52 predators (Chapron et al. 2014). Often used as a conservation success story, the recovery of large  
53 carnivores in human-dominated areas comes with challenges, including the question of whether there  
54 are any sufficiently large and functional areas left for viable populations (Packer et al. 2013). Another  
55 issue is how to coordinate management of these species at large scales, possibly across borders (Linnell  
56 and Boitani 2012, Bischof et al. 2015), in particular in the context of international treaties and  
57 directives (e.g. the Habitats Fauna Flora European Directive).

58 In this context, mapping the distribution of a species can help to predict and mitigate conflicts.  
59 Species distribution models (SDMs) have become important tools in the ecological, biogeographical  
60 and conservation fields (Guisan and Thuiller 2005). By correlating presence-only or presence-absence  
61 data of a species to environmental factors, SDMs provide an understanding of habitat preferences and  
62 predictions on future species distribution. This is especially relevant for species involved in conflicts,  
63 since predicting their future presence can help targeting contentious areas and guide management to  
64 reduce conflicts (Guillera-Arroita et al. 2015). However, the monitoring of large carnivores remains  
65 challenging to carry out in the field because these species live at low density and occupy wide areas  
66 (Gittleman et al. 2001). Therefore, assessing the distribution of these species comes with  
67 methodological challenges.

68 First, standard SDMs such as Maxent (Phillips et al. 2006) rely on the assumption of perfect  
69 detection, assuming that the focal species is detected everywhere it is present (Yackulic et al. 2013).  
70 Going undetected at a given site does not necessarily mean that this species is absent from that site, but

71 rather that it may simply be missed for various reasons related to observer abilities, habitat  
72 characteristics or species level of activity (Kéry et al. 2010, Kéry 2011). Ignoring the issue of imperfect  
73 detection can result in false absences that lead to flawed inference in two ways: i) the distribution maps  
74 are biased by underestimating actual presences and can misrepresent certain viable habitat features that  
75 are falsely identified as unfavorable (Kéry and Schaub 2011, Lahoz-Monfort et al. 2014); ii) there may  
76 be confusion in identifying the drivers of the species distribution when detection depends on  
77 environmental explanatory variables that are independent from the variables influencing the species'  
78 actual presence. For instance, if altitude has a negative effect on detection but not on presence, then as  
79 altitude gets higher, the species is less likely to be detected which can lead to erroneous conclusions  
80 that the species prefers lower altitudes (Lahoz-Monfort et al. 2014). To cope with this first issue,  
81 single-season or static site-occupancy models were developed (Mackenzie et al. 2006) and have been  
82 widely used for carnivores (e.g., Thorn et al. 2011 for brown hyenas *Hyaena brunnea*; Long et al. 2010  
83 for black bears *Ursus americanus*, fishers *Martes pennant* and bobcats *Lynx rufus*; Sunarto et al. 2012  
84 for Sumatran Tigers *Panthera tigris sumatrae*). Based on spatial and temporal replicated sampling of  
85 the target species, these models allow assessing the effects of environmental factors on species  
86 occupancy, while making the distinction between non-detections and true absences via the estimation  
87 of species detectability.

88         Second, most SDMs are implicitly based on the ecological niche concept (Grinnell 1917;  
89 Hutchinson 1957) and therefore rely on two main hypotheses: i) the species is present in areas where  
90 environmental conditions are the most favorable and ii) dispersal is not a limiting factor (Jeschke and  
91 Strayer 2006). However, expanding species are often absent from an area not because conditions are  
92 unfavorable but because they have not yet dispersed to this area, or because of geographical barriers or  
93 dispersal constraints (Araújo and Guisan 2006). Hence, static SDMs ignore important dynamic  
94 processes, which may lead to bias in the resulting distributions (Yackulic et al. 2015). Static SDMs  
95 should therefore not be used for predictions (Zurell et al. 2009). To deal with this second issue,  
96 occupancy models have been extended (Mackenzie et al. 2003, Royle and Kéry 2007) to account for  
97 the influence of dynamic processes such as colonization and extinction on the species range dynamics

98 (Mackenzie et al. 2003). So-called multi-season or dynamic site-occupancy models are increasingly  
99 used to assess the range dynamics of expanding or invasive species (e.g., Bled et al. 2011 for the  
100 hadeda ibis *Bostrychia hagedash* and Broms et al. 2016a for the common myna *Acridotheres tristis*),  
101 but remain rarely applied to carnivores (e.g., Marcelli and Fusillo 2012 for the Eurasian otter *Lutra*  
102 *lutra* or Miller et al. 2013 for the grey wolves *Canis lupus*).

103 Third, data collection is particularly costly if not unfeasible for elusive species that need wide  
104 areas due to the large presence area required for sampling. In this context, citizen science is considered  
105 as an efficient source of information to assess changes in a species distribution by covering wide areas  
106 (Schmeller et al. 2009). However, data from citizen science are often collected with protocols that do  
107 not control for variation in the sampling effort i) in time: a site can be sampled by several observers  
108 during a given year and not the following year and ii) in space: given two sites where the species is  
109 present, if the sampling effort is higher in one site, this might lead to recording a false absence in the  
110 site with lower sampling effort (Kéry et al. 2010). As a consequence, if sampling effort is not  
111 controlled for, detectability can be underestimated as well as its variability, leading to an  
112 overestimation of the distribution area (Van Strien et al. 2013).

113 Static and dynamic occupancy models hold promise to analyze population trends from  
114 opportunistic data because the data collection process is formally incorporated (Isaac et al. 2014).  
115 Occupancy models are divided into two types of process; i) the ecological one, governing dynamic  
116 occupancy processes (initial occupancy, local extinction and local colonization) and ii) the observation  
117 process, governing the data collection process (MacKenzie et al. 2003). However, to address the third  
118 issue and apply occupancy models to opportunistic data, one needs to differentiate between a site that  
119 was not sampled and a site that was sampled but the species was not detected. In the case of several  
120 species being monitored, the detection of a species in a site informs about the non-detection of other  
121 species because this site is known to have been sampled (Van Strien et al. 2013). This no longer holds  
122 for single-species settings, and the assumption is sometimes made that all sites where at least one  
123 detection occurred are sampled throughout the whole duration of the study (Molinari-Jobin et al. 2012,  
124 Rich et al. 2013).

25 Here, we considered Grey wolves (*Canis lupus*) as a case study to illustrate the challenges in  
26 using opportunistic data and SDMs to infer the range dynamics of large carnivores. Wolves  
27 disappeared in most of the Western European countries during the twentieth century (Promberger and  
28 Schroder 1993; Boitani 2010) except in Spain, Portugal and Italy (Boitani and Cucci 1993). The species  
29 naturally recolonized the French Alps from the remaining Italian population (Valière et al. 2003, Fabbri  
30 et al. 2007). Because the species is protected by law while being a source of conflicts with  
31 shepherding, its recolonization process needs to be carefully monitored.

32 Our main objective was to describe and determine the drivers of wolves recolonization pattern  
33 in France between 1993 and 2014. To account for imperfect detection, we built a dynamic site-  
34 occupancy model (Mackenzie et al. 2006) and analyzed opportunistic data collected by a network of  
35 trained volunteers since 1992. To do so, we built *a posteriori* the sampling effort to account for biases  
36 in data collected through citizen science. To describe the recolonization process over time, we  
37 addressed two main questions: (i) What are the environmental and biological factors influencing  
38 colonization and extinction probabilities? (ii) How can sampling effort be inferred *a posteriori*, i.e.  
39 after the data were collected, and to what extent does sampling effort correlate with detection  
40 probability?

41

## 42 Methods

43

### 44 **Study species and area**

45 The first wolf (*Canis lupus*) occurrence was detected in France in the early 1990s as a consequence of  
46 the Italian population's expansion (Valière et al. 2003, Ciucci et al. 2009). The species then spread  
47 outside the Alpine mountains to reach the Pyrenees and the Massif Central westward first in 1999, and  
48 the Vosges Mountains northward from 2011. The wolf is an opportunist species that can adapt its diet  
49 depending on available prey species (Pouille et al. 1997, Imbert et al. 2016). In areas with livestock  
50 farming, strong interactions between wolf presence and sheep breeding usually occur. The study area  
51 mostly covered Eastern France and a major part of Central France (Fig. 1).

52

53

[Figure 1 about here]

54

## 55 **Data collection**

56 Wolf detection data were made of presence signs sampled all year long from 1992 to 2014 thanks to a  
57 network of professional and non-professional observers. The network size has increased from a few  
58 hundred people in 1994, up to 3000 wolf experts in 2015. Every observer is trained during a 3-day  
59 teaching course led by the French National Game and Wildlife Agency (ONCFS) to document signs of  
60 the species presence (Duchamp et al. 2012). Presence signs went through a standardized control  
61 process combining genetic identification tools, and validation standards to prevent misidentification  
62 (Duchamp et al. 2012). For every presence sign, the date and location of collection were stored in a  
63 geo-referenced database. These data are considered opportunistic in the sense that monitoring occurs all  
64 year long in an extensive manner without explicitly quantifying the sampling effort.

65

## 66 **Dynamic site occupancy models**

67 To model the colonization dynamics of wolf, we used dynamic site-occupancy models (Mackenzie et  
68 al. 2003). These models allow the quantification of species occupancy while correcting for imperfect  
69 species detectability based on repeated sampling in time and space. We defined sampling units as  
70 10x10km cells, which appears to be the best option in the context of our study (Marboutin et al. 2010)  
71 and also is the recommended surface to produce maps of presence by the European Union (E. C. 2006).  
72 Site occupancy models rely on several assumptions, including the closure assumption which states that  
73 the ecological state of a site (whether it is occupied or not) remains unchanged through occasions (or  
74 surveys)  $j$  within a year  $k$ . During year  $k$ , sites were monitored mainly in winter from December to  
75 March, the most favorable period to detect the species between the two peaks of dispersal events in  
76 spring and fall (Mech and Boitani 2010). We defined the secondary occasions  $j$  as December, January,  
77 February and March and  $y_{i,j,k}$ , the observed state of site  $i$  equal to 1 if at least one sign of presence was  
78 found at site  $i$  during occasion  $j$  in the year  $k$  (and 0 otherwise).

79 We considered a state-space formulation of the dynamic occupancy model (Royle and Kéry  
80 2007) in which the model is viewed as the combination of (i) the ecological process that involves the  
81 latent ecological state of a site, i.e. whether it is occupied or not; (ii) the observation process that leads  
82 to the detections or non-detections by the observer conditional on the state of the system. The  
83 colonization probability  $\gamma_{i,k}$  is the probability that an empty site  $i$  during year  $k$  becomes occupied  
84 during year  $k+1$ , while the extinction probability  $\varepsilon_{i,k}$  is the probability that an occupied site  $i$  during  
85 year  $k$  becomes empty during year  $k+1$ . We define  $z_{i,1}$  as the initial latent state of site  $i$  as being drawn  
86 from a Bernoulli distribution with the success probability being  $\Psi_{i,1}$ ,  $z_{i,1} \sim \text{Bernoulli}(\Psi_{i,1})$ . All other  
87 latent states  $z_{i,k}$  for  $k > 1$  are drawn from a Bernoulli distribution as  $z_{i,k+1} | z_{i,k} \sim \text{Bernoulli}(z_{i,k}(1 - \varepsilon_{i,k}) +$   
88  $(1 - z_{i,k}) \gamma_{i,k})$ . On top of the ecological process stands the observation process, in which the  
89 detections/non-detections are drawn from a Bernoulli distribution  $y_{i,j,k} | z_{i,k} \sim \text{Bernoulli}(z_{i,k} p_{i,j,k})$  where  
90  $p_{i,j,k}$  is the probability that the species is detected at site  $i$  for an occasion  $j$  during year  $k$ . The state-  
91 space formulation is appealing as it makes explicit the latent states  $z_{i,k}$  that can be used to build  
92 distribution maps.

93

#### 94 **Sampling effort**

95 Monitoring the range expansion of wolves at the country level prevented us from implementing any  
96 standardized experimental sampling design. Instead, the presence signs were sampled in an  
97 opportunistic way and the sites were defined *a posteriori*. When dealing with detection-only data,  
98 various approaches have been adopted to infer the non-detections. In the context of species-list  
99 protocols, if other species are detected at a site but not the focal species, one can assume that observers  
00 were present but did not detect the species of interest and, hence a non-detection for this species is  
01 recorded at this given site (Kéry et al. 2010). In the context of single-species monitoring, several  
02 authors have assumed that the observation effort was sufficient enough to make the assumption that all  
03 occupied cells were monitored during the study period in any year (e.g., Molinari-Jobin et al. 2012).  
04 We adopted an original approach to infer the non-detections based on the available qualitative  
05 information on the observers. When entering the network, observers attended a 3-day training session

06 to learn how to identify the species and how it is monitored (Duchamp et al. 2012). During these  
07 training sessions, we recorded the observers' personal and professional address, socio-professional  
08 category and entry date into the network. The entry date was used to quantify how many observers  
09 were present in the network each year. We calculated a circular buffer for the prospection area for each  
10 observer based on a radius specific to his/her socio-professional category and a center located at his/her  
11 address (Supplementary material, table A1). For instance, for an observer belonging to the category 1  
12 (departmental authority) whose address was located in the French Department number 39, his/her  
13 prospection area would be 4 999 km<sup>2</sup>, which is the size of the Department (Supplementary material,  
14 Fig. A1 and Table A2). For this observer, a circular buffer was built with a radius calculated as *Radius*

$$15 = \sqrt{\frac{\textit{prospection area}}{\pi}}.$$

16 For each 10x10km cell, we then calculated the number of observers monitoring the species per  
17 year by summing the number of prospection areas overlapping the cell (supplementary material, Fig.  
18 A2). We set the species detection probability at a site to zero when the sampling effort was null in that  
19 site, i.e. no observers were present in that cell. When at least one observer was found in a cell in a given  
20 year, we considered that sampling occurred, hence concluding that a presence sign found at a particular  
21 occasion this year was a detection, and a non-detection otherwise. We expected that with more  
22 observers per site per year, the species was more likely to be detected, in other words that the sampling  
23 effort had a positive effect on the detection parameter. We performed a sensitivity analysis to assess  
24 how a change in the construction of the sampling effort influenced the model parameter estimates  
25 (Supplementary material, Fig. A3).

26

## 27 **Habitat covariates**

28 Wolves can adapt to a large range of different habitats, which makes it difficult to identify specific  
29 factors that may influence the species' presence at a site (Mech and Boitani 2010). However, we  
30 incorporated proxies of variables that might shape the wolf distribution (Table 1). Vegetation  
31 composition can indirectly influence the probability that a site becomes colonized (Marucco 2009) as

32 well as altitude (Llaneza et al. 2012, Falcucci et al. 2013). Using the CORINE Land Cover ® database  
33 (U.E – SOeS, Corine Land Cover, 2006), we defined 3 covariates to characterize the landscape of the  
34 study area: forest cover, farming cover and rock cover. We used the IGN BD\_ALTI® database (250m  
35 resolution) to calculate the mean altitude of each site as well as the proportion of altitude higher than  
36 2500m. Above this limit, most of the vegetation cover is grassland or rocky area. Altitude may be  
37 linked to colonization. We also predicted a site with a high proportion of high-altitude (>2500m high)  
38 would be less attractive for the species. Forest cover may structure the ungulate distribution (i.e. prey  
39 species). As a consequence, we expected that a site with higher forest cover would have a higher  
40 probability of being colonized and a site with higher rock cover would have a lower probability of  
41 being colonized. We also used the proportion of agricultural area as a covariate combining all types of  
42 farming activities including pastures areas. Those areas can be occupied by sheep, a possible prey to  
43 wolves, and therefore may have a positive influence on the settlement of the species at a site. Altitude  
44 may be linked to colonization. We also predicted a site with a high proportion of high-altitude (>2500m  
45 high) would be less attractive for the species.

46         Dispersal capacity is a key factor to explain the dynamic of wolf colonization (Boyd and  
47 Pletscher 1999, Kojola et al. 2006, Ciucci et al. 2009). Because cells occupied by established packs  
48 may act as a source of dispersers, (Yackulic et al. 2012), the neighborhood of an unoccupied cell may  
49 influence its colonization probability (Veran et al. 2015). On the other hand, wolves' strategies in  
50 colonization aim at avoiding neighbors via long-distance dispersal to avoid territorial competition with  
51 neighboring packs. In that spirit, the presence of individuals at short and long-distance could be  
52 accounted for by using conditional autoregressive models and auto-logistic models (Bled et al. 2013).  
53 However, due to the computational burden and convergence issues, we could not implement this  
54 approach here. We therefore defined two covariates that consisted of the observed number of  
55 contiguous observed occupied cells at both short and long-distances around the focal cell. The short-  
56 distance covariate was defined as the number of observed occupied cells that were directly contiguous  
57 to the focal cell i.e., situated within a distance of 10 km. The limit for the long-distance parameter was  
58 set to avoid a dilution effect due to the small number of observed occupied cells at very long-distances

59 but large enough to account for most long-distance observed occupied cells that could play a role in the  
60 colonization probability. Based on observations of wolf dispersal in the Western Italian Alps (Marucco  
61 and McIntire 2010), we set this limit at 150 km around the focal cell. We expected a positive effect of  
62 these two covariates on the probability of a site to be colonized.

63 Because dispersal could be driven by the presence of physical barriers (Wabakken et al. 2001,  
64 Blanco et al. 2005), we defined a landscape covariate depicting the distance from the center of the site  
65 to the closest barrier defined as highways or rivers (U.E – SOeS, Corine Land Cover, 2006). We  
66 expected this covariate to impact colonization negatively.

67 In the first few years after sites become newly colonized, extinction probability is expected to  
68 be high as long as only isolated individuals use them. Once a pack has settled, pack persistence is the  
69 rule for wolves when other packs are present in the surrounding areas (Mech and Boitani 2010). Pack  
70 splitting may rise from various sources including harvest or poaching of alpha pairs (Gehring et al.  
71 2003, Brainerd et al. 2008) leading to a locally extinct site. Within the distribution of an actively  
72 expanding population, extinct sites might be recovered by surrounding individuals, either by dispersers  
73 or by neighboring packs. We therefore expected the extinction probability to decrease over time, which  
74 was tested by using “year” as a continuous covariate.

75 Finally, in addition to sampling effort, we considered the potential effect of road densities on  
76 the species detectability, first through facilitation of site accessibility for the observers and second,  
77 because cross roads are often used as marking sites (Barja et al. 2004), which can lead to an increase in  
78 the species detection probability. Because presence signs rely partly on track records in the snow, we  
79 considered an ‘occasion effect’ to account for the variation in detection conditions due to weather  
80 variations across the survey months (Marucco 2009).

81 Last, we considered the initial occupancy probability as constant since only two sites were  
82 occupied in the first year of the study, which was not enough to assess the effects of covariates on this  
83 parameter.

84

85 **Model fitting, selection and validation**

86 We performed covariate selection using stochastic search variable selection (SSVS; George and  
87 McCulloch 1993, O’Hara and Sillanpää 2009). In brief, SSVS builds a model that includes all covariate  
88 combinations as special cases. In practice, this is achieved by adding binary indicator variables,  $\alpha_p$   
89 equals 1 or 0, which allows the estimation of the regression parameter  $\beta_p$  or excludes it by setting it to a  
90 constant (Supplementary material Table C1). In a Bayesian framework, we excluded a regression  
91 parameter by constraining it to 0 by specifying an informative prior centered on 0, while we estimated  
92 it by using a flat prior, that is  $\beta_p \sim (1 - \alpha_p)\text{Normal}(0,0.0001) + \alpha_p\text{Uniform}(0,1)$  with  $\alpha_p \sim$   
93  $\text{Bernoulli}(0.5)$ . Prior to model selection, we ran a Spearman test to check for correlations among  
94 covariates.

95 We used the software JAGS (Plummer 2003) and Markov chain Monte Carlo (MCMC)  
96 simulations for model selection and parameter estimation. We ran three MCMC chains with a burn-in  
97 period of 2500 iterations followed by 10000 iterations on which we based our inference. We checked  
98 convergence visually by inspecting the chains and by checking that the R-hat statistic was below 1.2  
99 (Gelman and Shirley 2011). We finally produced distribution maps of the latent states by using *a*  
00 *posteriori* means of the  $z_{i,k}$  from the best model. To assess the fit of our final model, we used the  
01 posterior predictive checking approach (Gelman et al. 1996) that has recently been applied to  
02 occupancy models (Broms et al. 2016b) (Supplementary material Fig. B1).

03

## 04 Results

### 05 **The effect of covariates on detectability and the dynamic of occupancy**

06 The model best supported by the data had detection as a function of sampling effort, road  
07 density and occasion (month) and colonization as a function of forest cover, farmland cover, mean  
08 altitude, proportion of high-altitude and the number of observed occupied cells at a short and long-  
09 distance neighborhood (Supplementary material, Table C1). This model appeared to fit the data  
10 adequately well (Supplementary material, Fig. B1). Posterior medians and 95% credible intervals are

11 given for each parameter. To calculate the effect of a covariate, we set the other covariates to their  
12 mean value.

13 Initial occupancy probability was low, as expected since few sites were detected as occupied at  
14 the beginning of the study (Supplementary material, Table C2).

15 As predicted, forest cover had a positive influence on the probability that a site became  
16 colonized. The proportion of farmland area within a cell also appeared to have a positive but weak  
17 influence on this probability. Below 1500m of altitude, the probability that a site became colonized was  
18 close to zero, whereas above this limit the probability reached up to 0.26 (0.16; 0.41) (Fig. 2). This  
19 probability decreased with the high-altitude proportion in a site. Finally, as predicted, both the short  
20 and long-distance count of observed occupied neighboring cells had a strong influence on the  
21 probability that a site became colonized over a year and was dependent of the early vs. late period of  
22 the wolf recovery trend. Over time, the number of observed occupied neighboring cells increased at  
23 both short and long-distance (Supplementary material, Fig. D1). If all of the 8 neighboring cells were  
24 observed as occupied, the probability that the target site became colonized was 0.37 (0.23; 0.54)  
25 compared to a colonization probability of  $2.71 \times 10^{-3}$  ( $2.11 \times 10^{-3}$ ;  $3.47 \times 10^{-3}$ ) if the target site had only 0  
26 to 2 contiguous neighboring cells observed occupied. As this number increased, the probability that a  
27 site became colonized increased accordingly (Fig. 2).

28 [Figure 2 about here]

29 Sites located within the Alps had the highest number of observed occupied sites at both short  
30 and long-distance. Colonization probability was the highest in this area (Fig.3). The highest part of the  
31 Alps (i.e. sites with the greatest proportions of high-altitude) remained with a low colonization  
32 probability (Supplementary material, Fig. D2). Overall, this probability was higher than zero in  
33 mountainous areas and increased with time as the number of occupied sites increased (Fig. 3).

34 [Figure 3 about here]

35 Finally, and as expected, detection probability varied according to the survey month with the  
36 lowest mean value of 0.07 (0.06; 0.09) in December and the highest value of 0.11 (0.09; 0.14) in  
37 January, and intermediate values of 0.10 (0.08; 0.13) in February and 0.10 (0.08; 0.12) in March when

38 road density and sampling effort were set to their mean values (Fig. 4). As expected, detection  
39 probability increased when the number of observers per site increased but, in contrast with what we  
40 expected, decreased with increasing road density. The sensitivity analysis showed weak effects of  
41 variations in the prospection areas used to build the sampling effort, except for the number of observed  
42 occupied sites (Supplementary Material, Fig. A3).

43 [Figure 4 about here]

#### 44 **Distribution map**

45 From 1992 to 2014, 13,554 presence signs were recorded by the network and used in our  
46 analysis. The species was initially spotted in 2 cells in 1993 and was detected in 143 cells in 2014  
47 (around 70-fold increase, see top panel in Fig. 5). This led to an apparent occupancy (proportion of  
48 occupied sites on the total number of sites in the study area) varying from 0.001 in 1993 to 0.046 in  
49 2014.

50 Accounting for both sampling effort and imperfect detection, we estimated the number of  
51 occupied sites as up to 10 (1; 20) in 1993 and up to 193 (178; 208) in 2014 (top panel in Fig. 5);  
52 overall, the estimates were higher than the naïve estimates of occupancy. When we ignored the  
53 sampling effort in the detection process, we found an estimated number of occupied sites equal to 3 (1;  
54 5) in 1993 and up to 184 (171; 196) in 2014 leading to an average of 9 (8; 10) newly occupied sites per  
55 year. Most discrepancies between the two models (accounting for vs. ignoring the sampling effort)  
56 were found at the early stage of the colonization process when the network of observers was  
57 implemented mainly in the southeastern part of the Alps (compare bottom left and right panels in Fig.  
58 5; see also Supplementary material, Fig. D3). Accounting for the sampling effort allowed us to infer the  
59 species presence on sites that were not prospected or prospected with a low sampling effort. As soon as  
60 the number of observers increased, the network was more homogeneously spread in space and  
61 estimates of the number of occupied sites became similar whether the sampling effort was included in  
62 the model or not (top panel in Fig. 5).

63 [Figure 5 about here]

64 Our results showed that in 1994 the species was found only in the Southern Alps, then actively  
65 colonized towards the Northern Alps at the beginning of the 2000's. The colonization process started to  
66 reach the Pyrenees and Massif Central area in early 2000 and the Vosges area in the very northern part  
67 of France, at the beginning of the 2010's, indicating that the French wolf population is still in a phase of  
68 expansion. On average 9 (6; 12) new sites became occupied per year with a minimum of 0 newly  
69 occupied sites in 2002 and a maximum of 29 newly occupied sites in 2012. This led to an average  
70 growth rate (i.e. number of new sites divided by the total number of sites the previous year) of 18%  
71 (15%; 19%) (Fig. 6). This growth rate decreased over time, from 125% at the early stage of the wolf  
72 colonization in 1994 to 5% in 2014, but the species is still in an expanding phase mainly thanks to the  
73 colonization outside of the alpine range.

74 [Figure 6 about here]

75 The model did not predict absence in places where presence signs were found (Fig. 7). Sites  
76 with high occupancy probability were mainly those close to the sites where the species had been  
77 previously detected, mostly due to the effect of short-distance neighbors. Some sites had a high  
78 probability of being occupied ( $> 0.75$ ), however the uncertainty associated with those predictions was  
79 also high (standard deviation [SD]  $> 0.30$ ). We found sites with high probability of occupancy ( $> 0.75$ )  
80 with low uncertainty (SD  $< 0.20$ ), and some of those sites were observed as occupied in the following  
81 year because the model propagates information backwards in time and so  $z_k$  is informed directly by  $z_{k+1}$ .

82

83 [Figure 7 about here]

84

## 85 Discussion

86 Determining favorable areas is often accomplished by building distribution maps using habitat  
87 suitability models (e.g., Mladenoff et al. 1999) or occupancy models (e.g., Marucco 2009). However,  
88 these studies often rely on a static relationship between the species of interest and its environment

89 (Jedrzejewski et al. 2008). Here, we used dynamic site-occupancy models and brought new insights on  
90 the processes governing the dynamic of recolonization of a keystone carnivore species. By controlling  
91 for species detectability and heterogeneous sampling effort, our approach can be used to assess the  
92 distribution dynamics of any species based on opportunistic data, pending relevant information is  
93 gathered on the people collecting the data.

94

## 95 **Model assumptions**

96

97 Site occupancy models rely on several assumptions that need to be discussed (Mackenzie et al. 2003,  
98 2006). First, the species should not be detected when absent from a site (i.e. no false positives). This is  
99 unlikely to happen in our case since we did not account for presence signs that were rejected because  
00 they did not fulfill the standardized criteria used to avoid species misidentification (Duchamp et al.  
01 2012). If doubts persist about the occurrence of false positives, this assumption could be relaxed by  
02 using site-occupancy models that account for misidentifications (Miller et al. 2011, Rich et al. 2013).  
03 Second, detection histories of all sampling units are assumed to be independent. However, detection  
04 histories were likely dependent in space because of a non-homogeneous spatial sampling effort inherent  
05 to opportunistic data. We partly accounted for this non-independence by quantifying the sampling  
06 effort. Furthermore, by accounting for the number of observed occupied neighboring cells, we made the  
07 detection history of a focal cell dependent partly on the detection histories of the neighboring cells. If  
08 the source of dependence is unknown, spatial autocorrelation can be modeled using geostatistical tools  
09 on occupancy or extinction/colonization parameters and also on detection (Bled et al. 2013). Third, the  
10 status of a site should not change during primary occasions - the closure assumption (Rota et al. 2009).  
11 If movements or mortality occurred inside or outside of the sampling sites, it is likely that the  
12 probability of occupancy in a given time interval did not depend on the occupancy status of a site in the  
13 previous time interval (Mackenzie and Royle 2005). In this situation of so-called random temporary  
14 emigration, the bias in parameter estimates is minimal, but occupancy should be interpreted as use of  
15 the sampling area rather than the proportion of area occupied by the species (Mackenzie et al. 2004).

.16 To prevent for this potential bias, we used the data provided within the winter period from November  
.17 to March as a primary occasion because it corresponds to the most stable period in the social  
.18 organization of the packs, between two main dispersal events in October (pup integration into the  
.19 packs' hunting activities) and the next mating season in March. Fourth, there should be no unmodelled  
.20 heterogeneity in the model parameters. In our study, heterogeneity might remain in species detection  
.21 even after accounting for spatio-temporal variation in the sampling effort or in the colonization  
.22 parameter after accounting for the effect of environmental covariates. Regarding the detection  
.23 probability, some heterogeneity might remain due to a difference of detection in the presence signs, e.g.  
.24 tracks vs. hair (Graves et al. 2011). This was unlikely to occur in our study because the vast majority of  
.25 presence signs are tracks. Regarding the colonization parameter, even though we had data on the  
.26 number of killed preys during the hunting season, we did not have information on wild prey density at  
.27 such a large scale . During winter (i.e. our primary occasions), wild preys consist mainly of Chamois  
.28 (*Rupicapra rupicapra*), mouflons (*Ovis ammon*), roe deer (*Capreolus capreolus*) and red deer (*Cervus*  
.29 *elaphus*) (Duchamp et al. 2012), for which we used characteristics of their habitats as a proxy for their  
.30 presence (Jedrzejewski et al. 2008).

.31 Besides the usual assumptions of occupancy models, we also had to deal with opportunistic data  
.32 that are collected through non-standardized sampling protocols. To cope with opportunistic data, we  
.33 defined a grid of spatial units that was overlaid on the map of detections/non-detections. In doing so,  
.34 the size and shape of these sampling units might have an impact on inference about the wolf  
.35 distribution. Indeed, if the size of the sampling unit is too small, then there is a risk of having very few  
.36 detections within a year, which would make the estimation of the detection probability difficult. On the  
.37 other hand, if the size of the sampling units is too large, then there is at least one detection in any cell,  
.38 which is of little use to estimate the distribution. We used 10x10km cells as sampling units, a choice we  
.39 made in agreement with what was recommended by the European Union and also shown to be the best  
.40 tradeoff between the species home range and sensitivity of the distribution to the size and shape of the  
.41 unit cell (Marboutin et al. 2010). The average size of wolf home ranges vary between 100 and 400 km<sup>2</sup>  
.42 in Western and central Europe (Ciucci 1997, Mech and Boitani 2010, Duchamp et al. 2012). Although

43 these cells might not entirely cover wolves territories, Latham et al. (2014) studied the effect of grid  
44 size to assess wolf's occupancy and found that taking a large grid size may not be appropriate for areas  
45 with moderate to high wolf density as it can overestimate occupancy rate.

46 Last, we assumed that observers were prospecting homogeneously inside the prospection area  
47 we assigned to them. This assumption may have been violated for two reasons. First, we believe that an  
48 observer was more likely to prospect more intensively near the center of the prospecting area, because  
49 it was defined as a home or work location, or near places where she/he already found presence signs.  
50 We also assumed that observers were prospecting homogeneously in time. This hypothesis can also be  
51 violated because observers may show different patterns in sampling frequency and some might not be  
52 prospecting during the months of winter. Finally, we made the assumption that once entered in the  
53 network, observers did not leave it unless we had information indicating the contrary such as a change  
54 of job or social status. Consequently, we might have overestimated the number of observers actually  
55 prospecting in the network for a given year. We therefore recommend recording carefully the activity  
56 of observers within the network to get a realistic picture of the actual sampling effort (Beirne and  
57 Lambin 2013).

58

### 59 **Effects of environmental covariates**

60

61 We used road density as a proxy of human presence and found a negative influence on the detection  
62 probability. When defining the road density covariate, we accounted for all types of roads (except  
63 highways) and assumed that this covariate could be a proxy for site accessibility. Because many  
64 observers from the network are wildlife professionals who are familiar with opportunistic surveys,  
65 main roads may not be used and accessibility to a site may consist mostly in dirt and forest roads or  
66 pathways. The negative influence could be explained by the fact that a high road density may also  
67 affect the spatial distribution of wolves. Because wolves tend to avoid roads (Whittington et al. 2005),  
68 there might be fewer presence marks. As expected, we found that detection probability increased when  
69 sampling effort increased, therefore highlighting the importance to account for imperfect detection

.70 when it is likely to be inhomogeneous in time and space. Finally, detection varied according to the  
.71 month of the survey, which can be explained by the variability in snow conditions in the study area,  
.72 with higher detectability within the alpine range than in the Massif Central or lowlands for instance.

.73 We found that colonization was mainly influenced by the number of observed occupied  
.74 neighbors at short and long-distances, showing that dispersal is a key factor of the dynamic of  
.75 occupancy. These results corroborate those of Adams et al. (2008) who showed that dispersal was the  
.76 main component explaining wolf population dynamics. Several long-distance dispersal events have  
.77 been documented across the alpine area (Wolf Alpine Group. 2014) and in France from the very south  
.78 eastern part to the northern Alps, from the Alps to the Massif Central or the northern Alps to the very  
.79 northeastern part of France (Duchamp et al. unpublished data). Further studies explicitly modeling  
.80 dispersal processes could help to better predict wolves colonization by accounting for factors that could  
.81 enhance or slow down the dispersal rate for instance (Broms et al. 2016a).

.82 We found that mean altitude had a positive effect on colonization probability. Wolves are  
.83 highly flexible and are able to live in various areas from maize cultures to high mountains (Kazcenski  
.84 et al. 2013). Starting from Italy westward to the Alps (Lucchini et al. 2002, Fabbri et al. 2007), wolves  
.85 reached the alpine range via the natural Apennine mountain corridor. Therefore, the effect of mean  
.86 altitude may be related to the history of the wolves' natural recovery process. However, we also found  
.87 a negative effect of the proportion of altitude higher than 2500m, i.e. the higher the proportion of high-  
.88 altitude, the less likely a site was to become colonized. Above 2500m, high vegetation turns from  
.89 forested ecosystems to sparse vegetation, above alpine meadows with rocky covers and snow. In  
.90 contrast, more forest cover associated with lower altitudes (<2500m) increased the probability that a  
.91 site become colonized mainly because these habitats' structure and composition are much more suitable  
.92 to the presence of key prey species like deer (Suter et al. 2004) or mountain ungulates (Darmon et al.  
.93 2012). To a lesser extent, the effect of farmland cover was also found to have a positive but weak  
.94 influence on the colonization probability, possibly because pasture areas host domestic preys during the  
.95 summer period (Meriggi and Lovari 1996). The inclusion of more explicit covariates related to pastoral

96 activity, such as the number of sheep in space, may provide a better understanding of the interaction  
97 between domestic prey and wolf presence, but these were not available to us.

98

### 99 **Trends in wolf recolonization**

00

01 Colonization patterns have been studied during recent decades (Wabakken et al. 2001). It appears that  
02 in Scandinavia, wolves were showing a colonization process that is typical of species with high  
03 dispersal capacities and pre-saturation dispersal (Swenson et al. 1998). This process is characterized by  
04 single long leaps forward and as a consequence, the colonization front is less well defined (Hartman  
05 1994) compared to a stepping stone dispersal strategy. Wolves seem to follow a similar pattern in  
06 France (Fig. 6) with an effect of long-distance neighborhood on colonization probability. This  
07 biological trait used by wolves is mainly known as a mechanism to avoid competition with other packs  
08 (Hayes and Harestad 2000). Once the area becomes more saturated, dispersers may fulfill remaining  
09 gaps in between other established territories, settling at long-distance unoccupied sites with higher risks  
10 of mortality due to an Allee effect (Hurford et al. 2006, Sanderson et al. 2013) or demographic  
11 stochasticity (Vucetich et al. 1997). In line with Marescot et al. (2011) who estimated a positive rate of  
12 increase in abundance, we demonstrated that the spatial dynamic mechanism of the wolves' natural  
13 recovery is still going on, particularly outside the alpine range both northward and westward. However,  
14 this recovery appeared to slow down in proportion with the total number of occupied sites per year,  
15 mainly due to sites becoming saturated within the alpine range and/or a recent increase in official wolf  
16 controls held by the government aimed to deter damage on livestock.

17 We used dynamic occupancy models to assess the *current* and *dynamic* distribution of a species  
18 that is expanding since it returned; there is a temptation to go a step forward and aim at forecasting its  
19 *future* distribution. However, we emphasize the difficulty of achieving this objective because we used  
20 environmental variables to explain colonization of the species that had already occurred; by definition,  
21 we could not incorporate the drivers that may appear relevant to explain future colonization events. For  
22 instance, now that wolves have settled in the alpine range and continue to expand, they are likely to

23 encounter new environments such as lowlands in the next few years, a landscape that may drive future  
24 colonization. Consequently, use our model as a predictive tool should be considered in an adaptive  
25 framework, i.e. by updating the management rules and the distribution maps every year during the  
26 active colonization phase.

27 The outcomes of our analyses have important consequences for managing animal species  
28 because their conservation statuses must be assessed partly through trends in their distributions (see  
29 art.1 of the Habitats Fauna Flora European Directive). Dynamic occupancy models are therefore  
30 relevant tools to the decision-making process by providing maps and spatio-temporal trends. In the case  
31 of the wolf, these models can help in preventing damage to livestock (Miller 2015). The identification  
32 of areas where the species may or may not occur along with the surrounding uncertainty may be used to  
33 target specific sites and determine priorities for implementing mitigation measures.

34

### 35 **Acknowledgements**

36 The authors are sincerely thankful to all the volunteers belonging to the Large Carnivore Network for  
37 the data collection and local investment in the fieldwork. J. Louvrier is thankful to the GDR Ecologie  
38 Statistique, University of Montpellier and ONCFS for grants she received to conduct her work. We  
39 warmly thank Ian Renner for editing our manuscript.

40

### 41 **References**

42 Adams, L. G. et al. 2008. Population dynamics and harvest characteristics of wolves in the Central  
43 Brooks Range, Alaska. - *Wildl. Monogr.* 170: 1–25.

44 Alberto Meriggi, S. L. 1996. A Review of Wolf Predation in Southern Europe: Does the Wolf Prefer  
45 Wild Prey to Livestock? - *J. Appl. Ecol.* 33: 1561–1571.

46 Araújo, M. B. and Guisan, A. 2006. Five (or so) challenges for species distribution modelling. - *J.*  
47 *Biogeogr.* 33: 1677–1688.

- 48 Barja, I. et al. 2004. The importance of crossroads in faecal marking behaviour of the wolves (*Canis*  
49 *lupus*). - *Naturwissenschaften* 91: 489–492.
- 50 Beirne, C. and Lambin, X. 2013. Understanding the Determinants of Volunteer Retention Through  
51 Capture-Recapture Analysis: Answering Social Science Questions Using a Wildlife Ecology  
52 Toolkit. - *Conserv. Lett.* 6: 391–401.
- 53 Bischof, R. et al. 2015. Wildlife in a Politically Divided World : Insularism Inflates Estimates of Brown  
54 Bear Abundance. - *Conserv. Lett.* 0: 1–9.
- 55 Blanco, J. C. et al. 2005. Wolf response to two kinds of barriers in an agricultural habitat in Spain. -  
56 *Can. J. Zool.* 83: 312–323.
- 57 Bled, F. et al. 2011. Assessing hypotheses about site occupancy dynamics. - *Ecology* 92: 938–951.
- 58 Bled, F. et al. 2013. Dynamic occupancy models for analyzing species' range dynamics across large  
59 geographic scales. - *Ecol. Evol.* 3: 4896–909.
- 60 Boitani, L. 2010. Wolf conservation and recovery. - In: Mech, L. D., and Boitani, L. (eds.), *Wolves:*  
61 *Behavior, Ecology, and Conservation.* University of Chicago Press, pp 317-340
- 62 Boitani L. and Ciucci P. 1993. Wolves in Italy: critical issues for their conservation. -In: Promberger,  
63 C., and Schröder, W. (eds.), *Wolves in Europe: status and perspectives.* Munich Wildlife Society,  
64 pp 74-90.
- 65 Brained, S. M. et al. 2008. The Effects of Breeder Loss on Wolves. - *J. Wildl. Manage.* 72: 89–98.
- 66 Breitenmoser, U. 1998. Large predators in the Alps : The fall and rise of man's competitor. - *Biol.*  
67 *Conser.* 83: 279–289.
- 68 Broms, K. M. et al. 2016a. Dynamic occupancy models for explicit colonization processes. - *Ecology*  
69 97: 194–204.

- 70 Broms, K. M. et al. 2016b. Model selection and assessment for multi-species occupancy models.  
71 Ecology. 97: 1759–1770.
- 72 Chapron, G. et al. 2014. Recovery of large carnivores in Europe’s modern human-dominated  
73 landscapes. - Science 346: 1517–1519.
- 74 Ciucci, P. et al. 1997. Home range, activity and movements of a wolf pack in central Italy. - J. Zool.  
75 243: 803-819.
- 76 Ciucci, P. et al. 2009. Long-Distance Dispersal of a Rescued Wolf From the Northern Apennines to the  
77 Western Alps. - J. Wildl. Manage. 73: 1300–1306.
- 78 Council of the European Commission 1992. Council directive 92/43/EEC of 21 May 1992 on the  
79 conservation of natural habitats and of wild fauna and flora. -Official Journal of the European  
80 Communities. Series L 206: 7-49.
- 81 Darmon, G. et al. 2012. Spatial distribution and habitat selection in coexisting species of mountain  
82 ungulates. - Ecography 35: 44–53.
- 83 Diane K. Boyd, D. H. P. 1999. Characteristics of Dispersal in a Colonizing Wolf Population in the  
84 Central Rocky Mountains. - J. Wildl. Manage. 63: 1094–1108.
- 85 Duchamp, C. et al. 2012. A dual frame survey to assess time- and space-related changes of the  
86 colonizing wolf population in France. - Hystrix 23: 1–12.
- 87 European Commission, 2006. Assessment, monitoring and reporting under article 17 of the habitats  
88 directive: explanatory notes and guidelines. 64p.
- 89 Fabbri, E. et al. 2007. From the Apennines to the Alps: colonization genetics of the naturally expanding  
90 Italian wolf (*Canis lupus*) population. - Mol. Ecol. 16: 1661-1671.

- 91 Falcucci, a. et al. 2013. Modeling the potential distribution for a range-expanding species: Wolf  
92 recolonization of the Alpine range. - *Biol. Conserv.* 158: 63–72.
- 93 Gehring, T. M. et al. 2003. Limits to plasticity in gray wolf, *Canis lupus*, pack structure: Conservation  
94 implications for recovering populations. - *Can. Field-Naturalist* 117: 419–423.
- 95 Gelman, A. et al.. 1996. Posterior predictive assessment of model fitness via realized discrepancies  
96 (with discussion). *Stat. Sinica* 6:733–807.
- 97 Gelman, A. and Shirley, K. 2011. Inference from simulations and monitoring convergence. - *Handb.*  
98 *Markov Chain Monte Carlo*: 163–174.
- 99 George, E., and McCulloch, R. E. 1993. Variable selection via Gibbs sampling. *JASA* 88: 881–889.
- 00 Gittleman, J. L. et al. 2001. Strategies for carnivore conservation: lessons from contemporary  
01 extinctions. - *Conserv. Biol. Ser.* 5: 61-92.
- 02 Graves, T. A. et al. 2011. Linking landscape characteristics to local grizzly bear abundance using  
03 multiple detection methods in a hierarchical model. *Anim. Conserv.* 14:652–664.
- 04 Grinnell, J. 1917. The niche-relationships of the California Thrasher. - *The Auk* 427-433.
- 05 Guillera-Arroita, G. et al. 2015. Is my species distribution model fit for purpose? Matching data and  
06 models to applications. - *Glob. Ecol. Biogeogr.* 1–17.
- 07 Guisan, A. and Thuiller, W. 2005. Predicting species distribution: Offering more than simple habitat  
08 models. - *Ecol. Lett.* 8: 993–1009.
- 09 Hartman, G. 1994. Long-Term Population Development of a Reintroduced Beaver (*Castor fiber*)  
10 Population in Sweden. - *Conserv. Biol.* 8: 713–717.

- .11 Hayes, R. D. and Harestad, A. S. 2000. Demography of a recovering wolf population in the Yukon. -  
.12 Can. J. Zool. 78: 36–48.
- .13 Hurford, A. et al. 2006. A spatially explicit model for an Allee effect: why wolves recolonize so slowly  
.14 in Greater Yellowstone. - Theor. Popul. Biol. 70: 244–54.
- .15 Hutchinson, G. E. 1957. Concluding remarks. Cold Springs harbor Symposia on Quantitative Biology  
.16 22: 415-427.
- .17 Imbert, C. et al. 2016. Why do wolves eat livestock?: Factors influencing wolf diet in northern Italy. -  
.18 Biol. Conserv. 195: 156-168.
- .19 Isaac, N. J. B. et al. 2014. Statistics for citizen science: extracting signals of change from noisy  
.20 ecological data. - Methods Ecol. Evol. 5: 1052–1060.
- .21 Jedrzejewski, W. et al. 2008. Habitat suitability model for Polish wolves based on long-term national  
.22 census. - Anim. Conserv. 11: 377–390.
- .23 Jeschke, J. M. and Strayer, D. L. 2006. Determinants of vertebrate invasion success in Europe and  
.24 North America. - Glob. Chang. Biol. 12: 1608–1619.
- .25 Kaczensky, P. et al. 2013. Status, management and distribution of large carnivores–bear, lynx, wolf &  
.26 wolverine–in Europe. - Report to the EU Commission.
- .27 Kéry, M. 2011. Towards the modelling of true species distributions. - J. Biogeogr. 38: 617–618.
- .28 Kéry, M. and Schaub, M. 2011. Bayesian Population Analysis using WinBUGS - a hierarchical  
.29 perspective.
- .30 Kéry, M. et al. 2010. Predicting species distributions from checklist data using site-occupancy models.  
.31 - J. Biogeogr. 37: 1851–1862.

- 32 Kojola, I. et al. 2006. Dispersal in an Expanding Wolf Population in Finland. - *J. Mammal.* 87: 281–  
33 286.
- 34 Lahoz-Monfort, J. J. et al. 2014. Imperfect detection impacts the performance of species distribution  
35 models. - *Glob. Ecol. Biogeogr.* 23: 504–515.
- 36 Latham, M. C. et al. 2014. Can occupancy-abundance models be used to monitor wolf abundance? -  
37 *PLoS One* 9: e102982.
- 38 Linnell, J. D. C. and Boitani, L. 2012. Building biological realism into wolf management policy: The  
39 development of the population approach in Europe. - *Hystrix* 23: 80–91.
- 40 Llaneza, L. et al. 2012. Insights into wolf presence in human-dominated landscapes: the relative role of  
41 food availability, humans and landscape attributes. - *Divers. Distrib.* 18: 459–469.
- 42 Long, R. a. et al. 2010. Predicting carnivore occurrence with noninvasive surveys and occupancy  
43 modeling. - *Landsc. Ecol.* 26: 327–340.
- 44 Lucchini, V. et al. 2002. Noninvasive molecular tracking of colonizing wolf (*Canis lupus*) packs in the  
45 western Italian Alps. - *Mol. Ecol.* 11: 857-868.
- 46 Mackenzie, D. I. and Royle, J. A. 2005. Designing occupancy studies : general advice and allocating  
47 survey effort. - *J. Appl. Ecol.* 42: 1105–1114.
- 48 Mackenzie, D. I. et al. 2003. Estimating Site Occupancy , Colonization , and Local Extinction When a  
49 Species Is Detected Imperfectly. - *Ecology* 84: 2200–2207.
- 50 Mackenzie, D. I. et al. 2004. Investigating species co-occurrence patterns when species are detected  
51 imperfectly. - *J. Anim. Ecol.* 73: 546–555.
- 52 MacKenzie, D. I. et al. 2006. Occupancy estimation and modelling. Inferring patterns and dynamics of  
53 species occurrence. - Academic, Burlington.

- 54 Marboutin, E. et al. 2010. On the effects of grid size and shape when mapping the distribution range of  
55 a recolonising wolf (*Canis lupus*) population. - *Eur. J. Wildl. Res.* 57: 457–465.
- 56 Marcelli, M. and Fusillo, R. 2012. Land use drivers of species re-expansion: inferring colonization  
57 dynamics in Eurasian otters. - *Divers. Distrib.* 18: 1001–1012.
- 58 Marescot, L. et al. 2011. Capture-recapture population growth rate as a robust tool against detection  
59 heterogeneity for population management. - *Ecol. Appl.* 21: 2898-2907.
- 60 Marucco, F. 2009. Spatial population dynamics of a recolonizing wolf population in the Western Alps.  
61 -University of Montana, Missoula.
- 62 Marucco, F. and McIntire, E. J. B. 2010. Predicting spatio-temporal recolonization of large carnivore  
63 populations and livestock depredation risk: wolves in the Italian Alps. - *J. Appl. Ecol.* 47: 789–  
64 798.
- 65 Mech, L. D. and Boitani, L. 2010. *Wolves: Behavior, Ecology, and Conservation: Behavior, Ecology,  
66 and Conservation.* -University of Chicago Press.
- 67 Miller, J. R. B. 2015. Mapping attack hotspots to mitigate human--carnivore conflict: approaches and  
68 applications of spatial predation risk modeling. - *Biodivers. Conserv.* 24: 2887–2911.
- 69 Miller, D. A. W. et al. 2011. Improving occupancy estimation when two types of observational error  
70 occur: non-detection and species misidentification. - *Ecology* 92: 1422–1428.
- 71 Miller, D. A. W. et al. 2013. Determining Occurrence Dynamics when False Positives Occur:  
72 Estimating the Range Dynamics of Wolves from Public Survey Data. - *PLoS One* 8: e65808.
- 73 Mladenoff, D. J. et al. 1999. Predicting gray wolf landscape recolonization: logistic regression vs. new  
74 field data. - *Ecol. Appl.* 9: 37-44.

- .75 Molinari-Jobin, a. et al. 2012. Monitoring in the presence of species misidentification: the case of the  
.76 Eurasian lynx in the Alps. - *Anim. Conserv.* 15: 266–273.
- .77 O’Hara, R. B., and Sillanpää, M. J. 2009. A review of Bayesian variable selection methods: what, how  
.78 and which. *Bayesian Anal.* 4: 85–118.
- .79 Packer, C. et al. 2013. Conserving large carnivores: Dollars and fence. - *Ecol. Lett.* 16: 635–641.
- .80 Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. -*Ecol.*  
.81 *Model.* . 190: 231–259.
- .82 Plummer, M. 2003. JAGS: A program for analysis of bayesian graphical models using gibbs sampling.  
.83 - *Proc. 3rd Int. Work. Distrib. Stat. Comput. (DSC 2003)*. March: 20–22.
- .84 Poulle, M.-L. et al. 1997. Significance of ungulates in the diet of recently settled wolves in the  
.85 Mercantour mountains (southeastern France). - *Rev. d’Ecologie (Terre vie)* 52: 357–358.
- .86 Promberger, C., and Schröder, W. 1993. Wolves in Europe: status and perspectives. -*Munich Wildlife*  
.87 *Society*.
- .88 Rich, L. N. et al. 2013. Estimating occupancy and predicting numbers of gray wolf packs in Montana  
.89 using hunter surveys. - *J. Wildl. Manage.* 77: 1280–1289.
- .90 Ripple, W. J. et al. 2014. Trophic cascades from wolves to grizzly bears in Yellowstone. - *J. Anim.*  
.91 *Ecol.* 83: 223–33.
- .92 Rota, C. T. et al. 2009. Occupancy estimation and the closure assumption. - *J. Appl. Ecol.* 46: 1173–  
.93 1181.
- .94 Royle, J. A. and Kéry, M. 2007. A Bayesian state-space formulation of dynamic occupancy models. -  
.95 *Ecology* 88: 1813–1823.

- 96 Sanderson, C. E. et al. 2013. With Allee effects, life for the social carnivore is complicated. - *Popul.*  
97 *Ecol.* 56: 417–425.
- 98 Schmeller, D. S. et al. 2009. Advantages of Volunteer-Based Biodiversity Monitoring in Europe. -  
99 *Conserv. Biol.* 23: 307–316.
- 00 Sunarto, S. et al. 2012. Tigers need cover: multi-scale occupancy study of the big cat in Sumatran forest  
01 and plantation landscapes. - *PLoS One* 7: e30859.
- 02 Suter, W. et al. 2004. Spatial variation of summer diet of red deer *Cervus elaphus* in the eastern Swiss  
03 Alps. - *Wildlife Biol.* 10: 43–50.
- 04 Swenson, J. et al. 1998. Geographic expansion of an increasing brown bear population, evidence for  
05 presaturation dispersal. - *J. Appl. Ecol.* 67: 708–715.
- 06 Thorn, M. et al. 2011. Brown hyaenas on roads: Estimating carnivore occupancy and abundance using  
07 spatially auto-correlated sign survey replicates. - *Biol. Conserv.* 144: 1799–1807.
- 08 Valière, N. et al. 2003. Long-distance wolf recolonization of France and Switzerland inferred from  
09 non-invasive genetic sampling over a period of 10 years. - *Anim. Conserv.* 6: 83–92.
- 10 Van Strien, A. J. et al. 2013. Opportunistic citizen science data of animal species produce reliable  
11 estimates of distribution trends if analysed with occupancy models. - *J. Appl. Ecol.* 50: 1450–  
12 1458.
- 13 Veran, S. et al. 2015. Modeling spatial expansion of invasive alien species: relative contributions of  
14 environmental and anthropogenic factors to the spreading of the harlequin ladybird in France. -  
15 *Ecography* 38: 001–011.
- 16 Vucetich, J. A. et al. 1997. Effects of Social Structure and Prey Dynamics on Extinction Risk in Gray  
17 Wolves. - *Conserv. Biol.* 11: 957–965.

- 18 Wabakken, P. et al. 2001. The recovery, distribution, and population dynamics of wolves on the  
19 Scandinavian peninsula, 1978-1998. - *Can. J. Zool.* 79: 710–725.
- 20 Whittington, J. et al. 2005. Spatial responses of wolves to roads and trails in mountain valleys. - *Ecol.*  
21 *Appl.* 15: 543-553.
- 22 Wolf Alpine Group 2014. Wolf population status in the Alps: Pack Distribution and trends up to 2012.  
23 -7th « Wolf Alpine Group » Workshop“wolf monitoring over the Alps – towards a unique  
24 approach” 2013 March 19- 20th - Jausiers, FRANCE.
- 25 Yackulic, C. B. et al. 2012. Neighborhood and habitat effects on vital rates: expansion of the Barred  
26 Owl in the Oregon Coast Ranges. - *Ecology* 93: 1953–1966.
- 27 Yackulic, C. B. et al. 2013. Presence-only modelling using MAXENT : when can we trust the  
28 inferences ? - *Methods Ecol. Evol.* 4: 236–243.
- 29 Yackulic, C. B. et al. 2015. To predict the niche, model colonization and extinction. *Ecology*, 96: 16-  
30 23. Zurell, D. et al. 2009. Static species distribution models in dynamically changing systems: How  
31 good can predictions really be? - *Ecography* 32: 733–744.

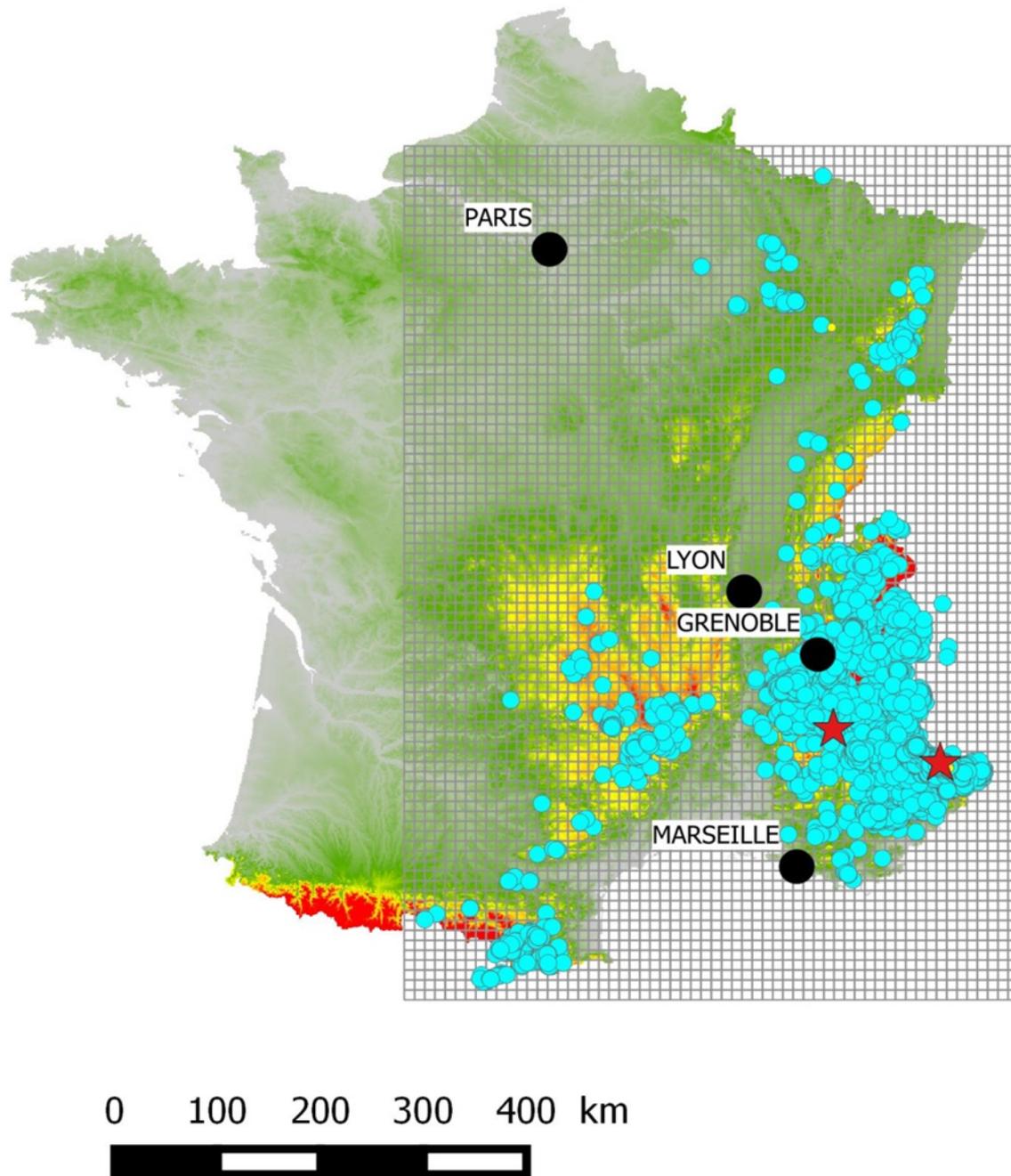


Figure 1: Maps of cumulated species detections (blue dots) for the period 1993-2014. Red stars represent the first detections made in 1992. Sites were defined as 10x10km cells within a grid covering all detections. Red areas represent mountainous areas with an altitude higher than 1500 meters.

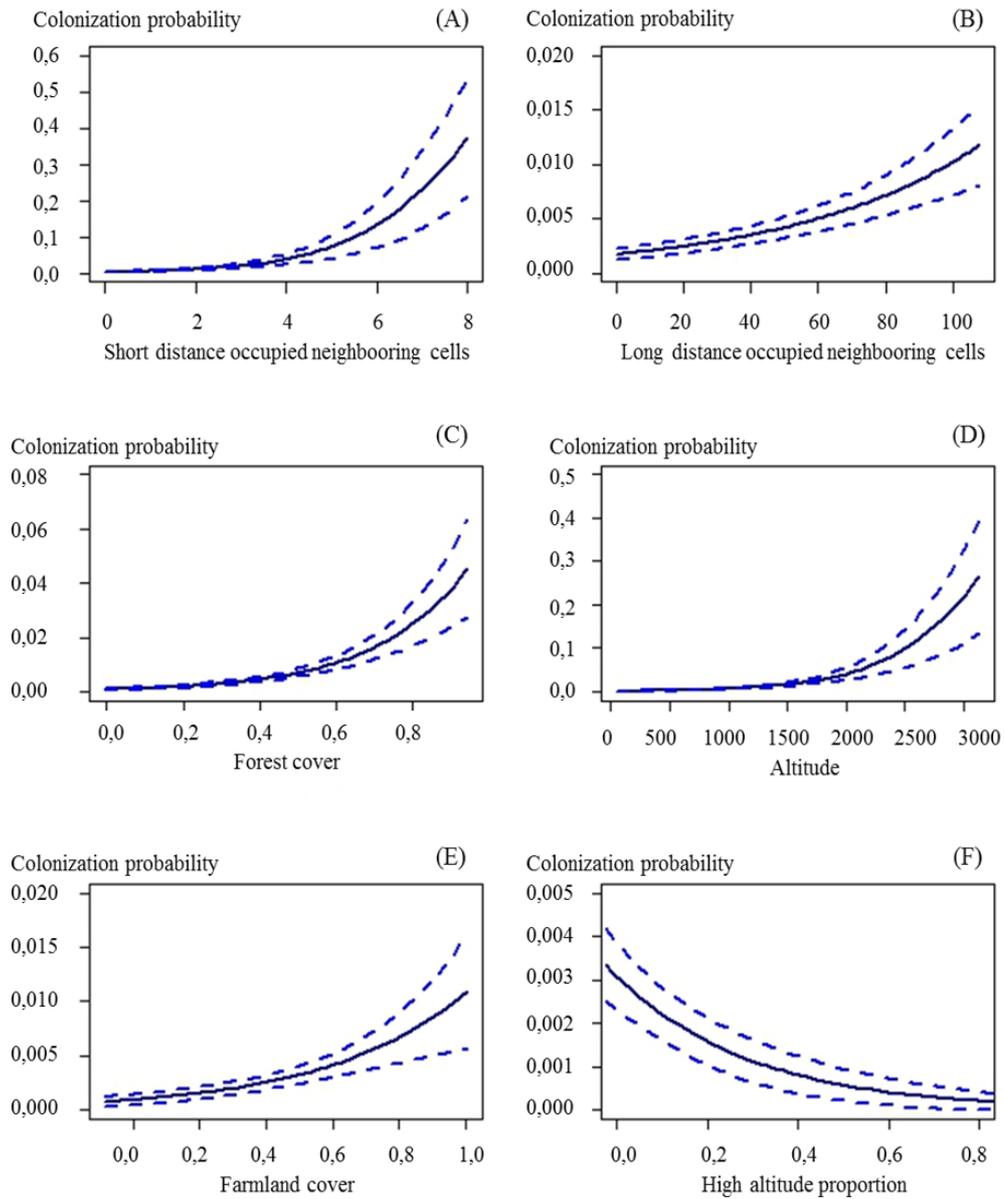


Figure 2: Relationship between the estimated colonization probability and (A) short-distance occupied neighboring cells, (B) long-distance occupied neighboring cells, (C) forest cover, (D) altitude, (E) farmland cover, and (F) site proportion of altitude higher than 2500 m.

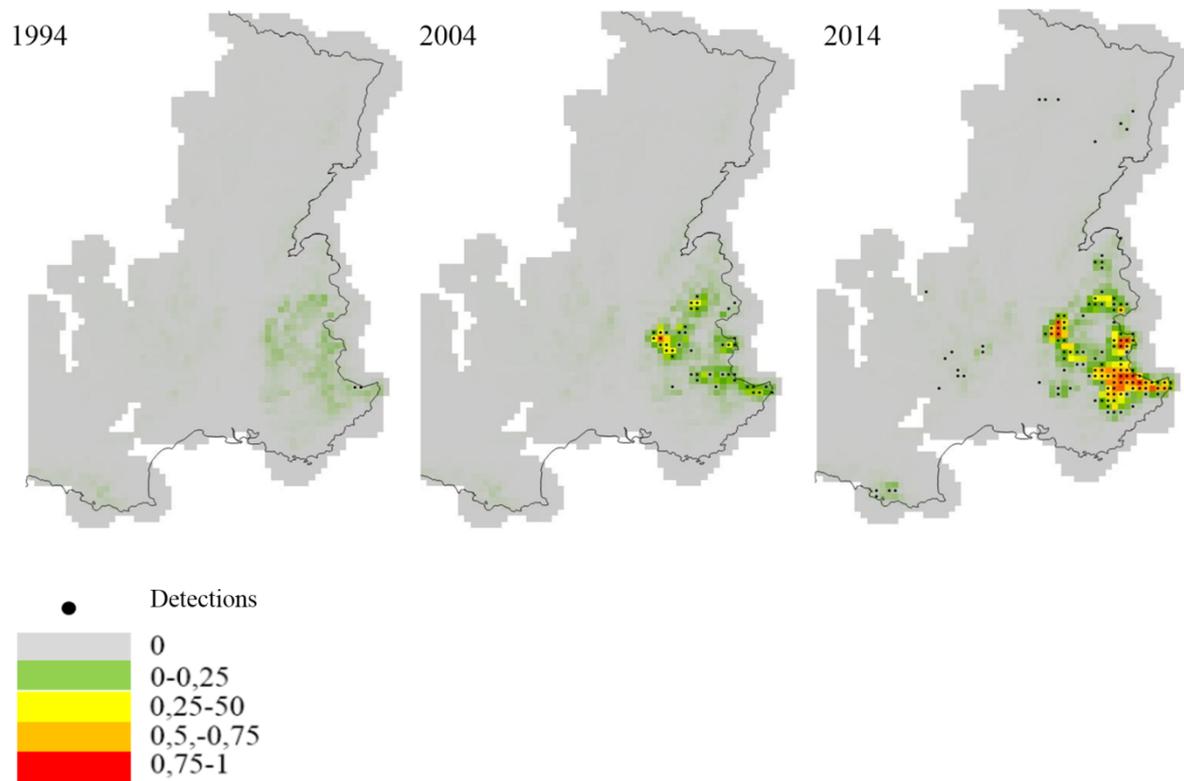


Figure 3: Maps of estimated colonization probability between 1993 and 1994, 2003 and 2004 and 2013 and 2014 from the best model (Table 2). Black dots represent detections made in 1993, 2003 and 2013.

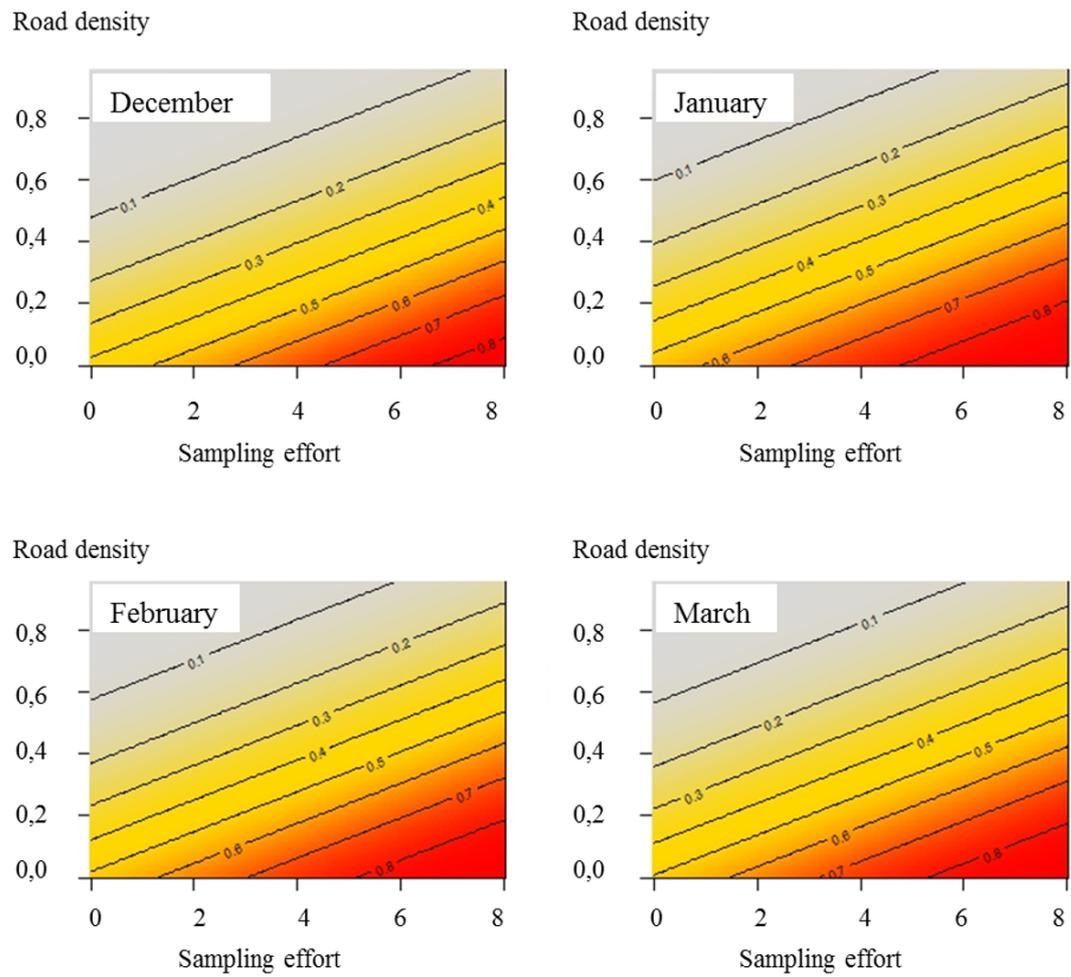


Figure 4: Joint effects of road density, sampling effort and occasion (month) on the species detection probability.

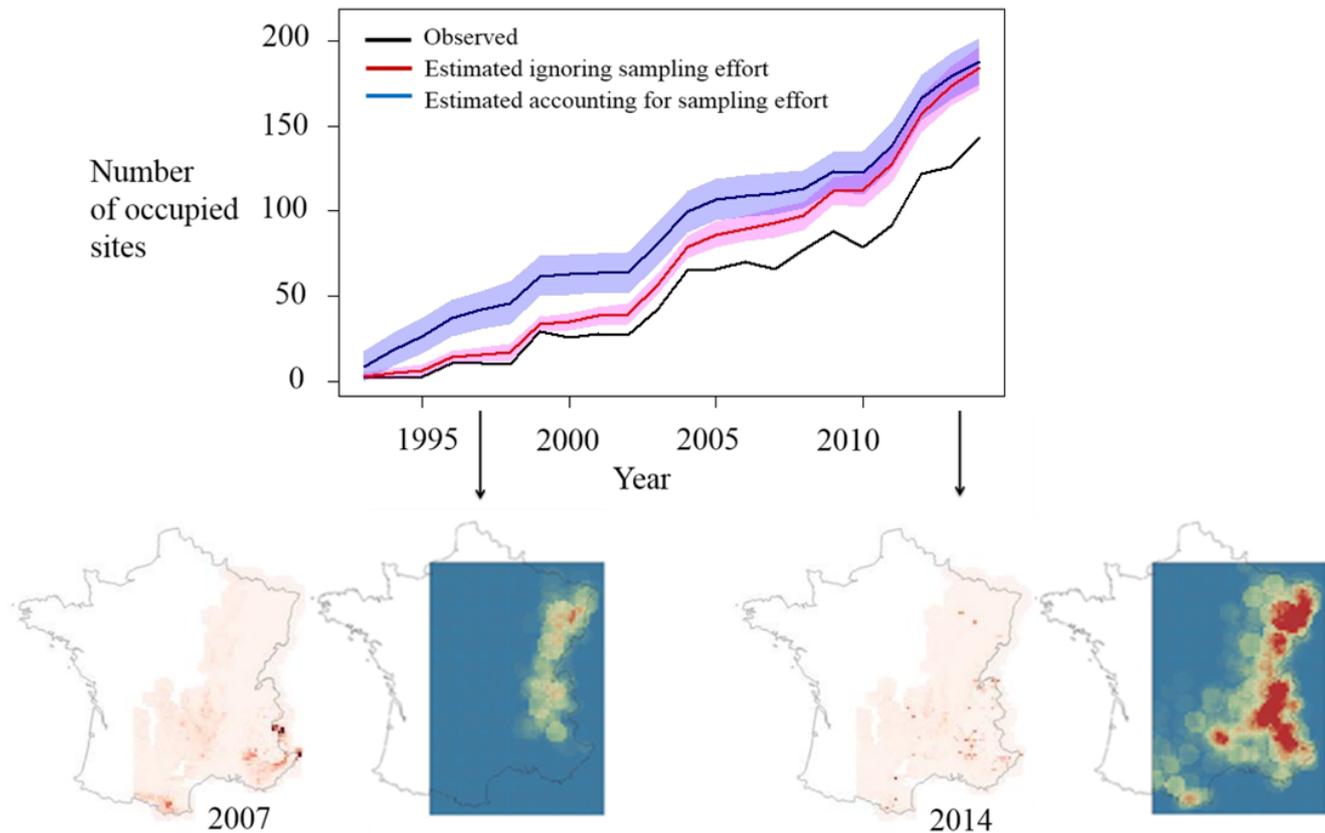


Figure 5:

Top: Number of 10 x 10 km quadrats observed (black), estimated occupied ignoring sampling effort (red) and estimated occupied accounting for sampling effort (blue). Blue and Pink parts represent the 95% credible interval associated to the estimated number of occupied sites.

Bottom: Maps of differences between estimates of occupancy from the model accounting for sampling effort and the model ignoring sampling effort. Dark red sites are sites that appeared estimated occupied by the model accounting for sampling effort but did not appear occupied once ignoring sampling effort. Both maps are associated with maps of the sampling effort for the years 1997 and 2014.

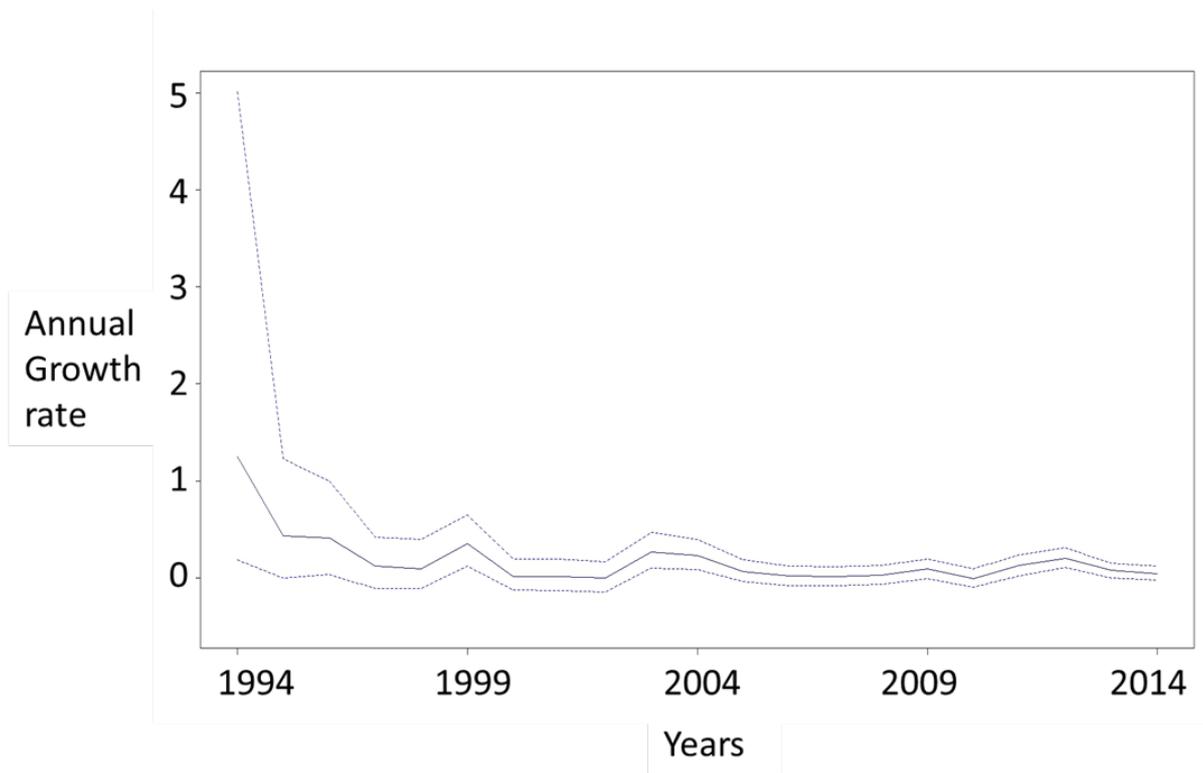


Figure 6: Growth rate (i.e. number of new sites rated by the total number of sites the previous year, multiplied by 100) given for each year from 1994 to 2014

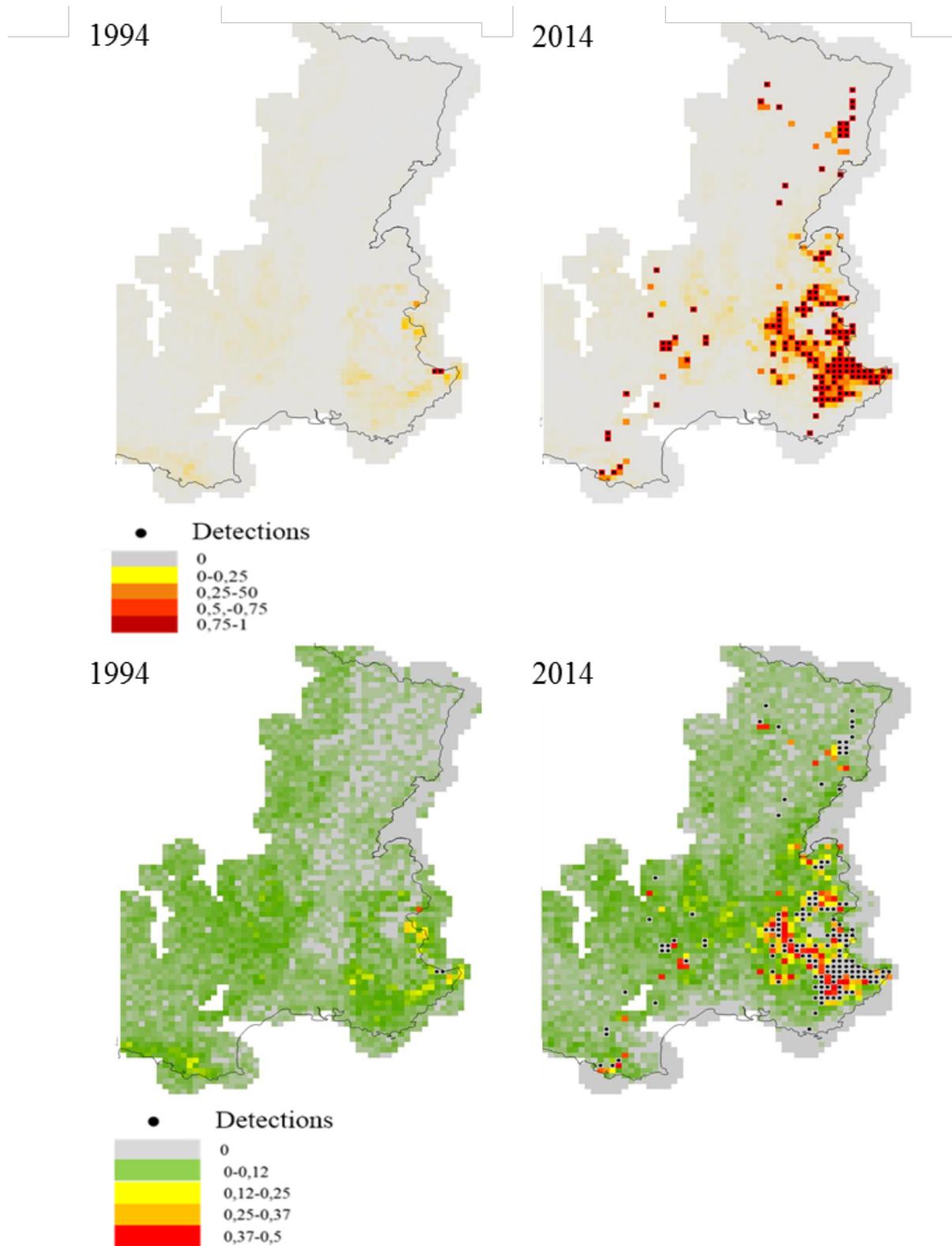


Figure 7: Maps of estimated occupancy (top) and associated standard deviation (bottom) for years 1994 and 2014. Black dots represent detections made in 1994 and 2014.

Table 1: Description and expected effects of covariates used to describe the occupancy dynamics of wolf in France.

Covariate	Abbreviation	Parameter	Description	Expected effect	Reference
Forest cover	Forest	Colonisation ( $\gamma$ )	Percentage of mixt, coniferous or deciduous forests cover	+	Oakleaf <i>et al.</i> 2006, Fechter & Storch 2013
Farmland cover	Agr	Colonisation ( $\gamma$ )	Percentage of pasture lands and other farming activities cover	+/-	Glenz <i>et al.</i> 2001
Rock cover	Rock	Colonisation ( $\gamma$ )	Percentage of rock cover	-	
High altitude	Halt	Colonisation ( $\gamma$ )	Proportion of altitude higher than 2500 meters	-	Glenz <i>et al.</i> 2001
Altitude	Alt	Colonisation ( $\gamma$ )	Mean altitude	+/-	Llanneza <i>et al.</i> 2012 Falcucci <i>et al.</i> 2013
Distance to the closest barrier	Dbarr	Colonisation ( $\gamma$ )	Minimal distance between a highway or one of the five main rivers in France.	-	Falcucci <i>et al.</i> 2013
Short distance occupied neighboring cells	SDAC	Colonisation ( $\gamma$ )	Proportion of observed occupied contiguous cells	+	Bled <i>et al.</i> 2011
Long distance	LDAC	Colonisation ( $\gamma$ )	Proportion of observed occupied cells within a 150		

---

occupied			km radius without the contiguous cells.	+	
neighboring cells					
Year	Trend-year	Extinction ( $\varepsilon$ )	Year as a linear effect	-	Marucco, 2009
(continuous)					
Sampling effort	SEff	Detection ( $p$ )	Number of observers per site per year	+	
Road density	Rdens	Detection ( $p$ )	Percentage of site covered by roads	+	
Month-survey	survey	Detection ( $p$ )	Occasion of survey (categorical)	+/-	Marucco, 2009

---

Table A1: Size of prospection areas as a function of socio-professional category of observers. Observers were classified according to 8 entities to capture the diversity of their professional and personal field activities. People working for the departmental authorities (Category 1) display a field effort all over that departmental area. Observers belonging to the category 2 are state employees affected to the protected area they are working in. Details were not given for Regional Natural Park agents and Natural reserve agents. Their prospection area corresponds to the mean area of the protected area they are affiliated to. ONCFS agents (category 8) are attributed half a French Department as field areas when assigned for species monitoring. ONF agents (category 9) are attributed 1/10 of a French Department. Farmers (category 4) and hunters (category 5) usually focus on the restricted area (“municipality”) where they farm, breed sheep or hunt. Scientists (category 3), members of a naturalist association (category 6) and volunteers (category 7) were given  $\frac{1}{4}$  of their affiliated department as their main activity might not be focused on species monitoring.

Socio-professional category	Prospection area	Number of observers
Administrative (e.g., department authorities) <b>(category 1)</b>	Area of the affiliated French department	230
National Park agent <b>(category 2)</b>	Area of the affiliated National Park	156
Regional Natural Park agent <b>(category 2)</b>	450 km <sup>2</sup>	69
Natural Reserve agent <b>(category 2)</b>	10 km <sup>2</sup>	20
Scientist <b>(category 3)</b>	$\frac{1}{4}$ of the affiliated department’s surface	6
Agricultural profession <b>(category 4)</b>	Area of the affiliated municipality	59
Hunter <b>(category 5)</b>	Area of the affiliated municipality	313
Member of a naturalist association <b>(category 6)</b>	$\frac{1}{4}$ of the affiliated department’s surface	216
Volunteer <b>(category 7)</b>	$\frac{1}{4}$ of the affiliated department’s surface	311
ONCFS agent <b>(category 8)</b>	$\frac{1}{2}$ of the affiliated department’s surface	513
French Forest Agency (ONF) agent <b>(category 9)</b>	1/10 of the affiliated department’s surface	284

Table A2: French departments where observers are present in 2014 along with their area and the number of observers affiliated to each department depending on their socio-professional category. Categories 2, 4 and 5 are not shown because their prospection areas do not depend on the size of the affiliated department. See also Figure A1.

Department code	Surface (km <sup>2</sup> )	Category 1	Category 3	Category 6	Category 7	Category 8	Category 9
1	5762	5	0	8	8	22	11
4	6925	20	0	5	24	12	12
5	5549	8	0	4	17	19	6
6	4299	5	0	0	18	14	1
7	5529	5	0	6	0	12	4
8	5229	0	0	0	3	1	0
9	4890	0	0	4	1	11	7
10	6004	0	1	1	0	2	0
11	6139	0	0	0	0	3	2
12	8735	4	0	0	3	11	4
13	5087	0	0	1	1	10	1
15	5726	5	0	4	3	11	2
19	5857	0	0	0	0	2	0
21	8763	2	0	0	1	11	0
23	5565	0	0	0	0	1	0
24	9060	0	0	0	0	2	0
25	5234	14	0	11	4	22	11
26	6530	9	0	6	10	14	18
30	5853	4	0	3	0	7	1
31	6309	1	1	0	6	12	1
32	6257	0	0	0	1	8	0
34	6101	0	1	1	0	3	0
38	7431	6	1	21	29	28	23
39	4999	5	0	16	7	12	14
42	4781	8	0	1	4	3	3
43	4977	2	0	0	1	4	0
46	5217	0	0	0	0	1	0
47	5361	0	0	0	0	3	0
48	5167	17	0	5	6	14	8
54	5246	5	0	9	3	7	4
55	6211	1	0	1	1	23	3
57	6216	4	0	4	6	17	1
58	6817	0	0	1	0	0	0
59	5743	0	0	0	0	0	0
63	7970	0	1	5	1	6	1
65	4464	0	0	0	0	2	0
66	4116	6	0	7	18	10	11
67	4755	11	0	19	16	31	14
68	3525	28	0	20	44	16	17
69	3249	1	0	1	2	4	0
70	5360	10	0	11	1	12	12
71	8575	0	0	1	2	7	1
73	6028	7	0	5	17	17	33
74	4388	6	0	16	29	20	13
81	5758	0	0	0	0	4	1
82	3718	3	0	0	0	3	0
83	5973	3	1	2	3	15	1
84	3567	2	0	1	0	10	5
87	5520	0	0	0	1	0	0
88	5874	16	0	12	17	25	33
89	7427	0	0	1	0	0	0
90	609	7	0	3	3	9	5



Figure A1: Map of French departments with the identity code used in Table A2.

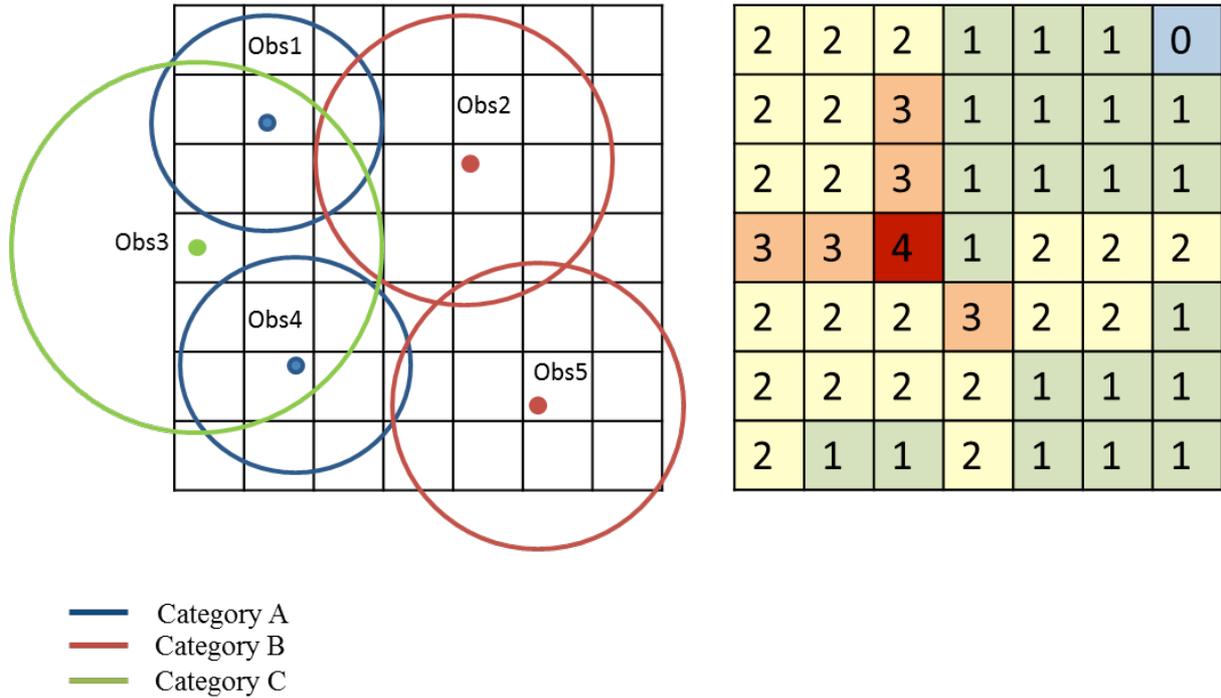


Figure A2: Schematic representation of how the sampling effort was calculated. Left: Observers were plotted according to their address. A circular buffer was affiliated to each observer with a surface equal to the prospection area following Tables A1 and A2. Right: resulting sampling effort calculated as the sum of observers sampling in each cell.

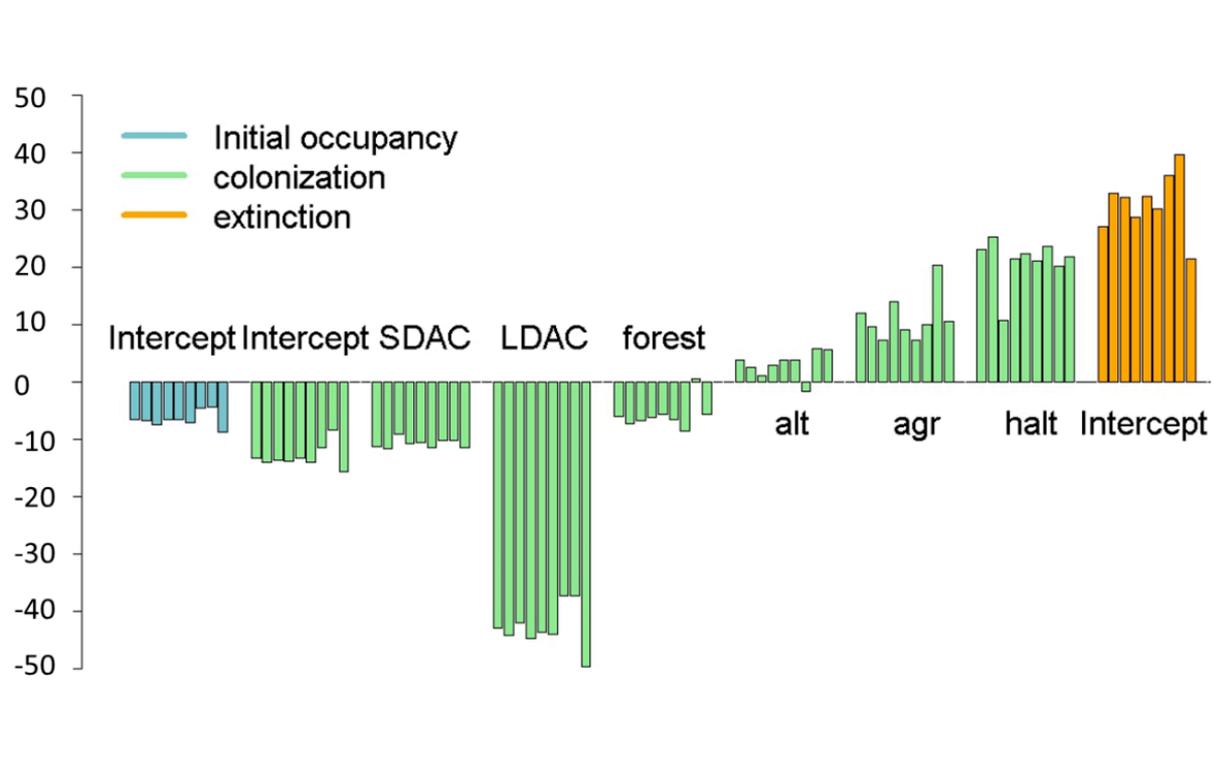


Figure A3: Analysis of sensitivity to the sampling effort definition. Each bar within a barplot represents the percentage of change (in %) in parameter estimates when compared to our original definition of the sampling effort after modification of the prospection area for one category of observers, in the following order: category 6: 1/10 of the affiliated department's surface; category 6: 1/2 of the affiliated department's surface; category 6: 100% of the affiliated department's surface; category 7: 1/10 of the affiliated department's surface; category 7: 1/2 of the affiliated department's surface; category 7: 100% of the affiliated department's surface; category 8: 1/10 of the affiliated department's surface; category 8: 1/4 of the affiliated department's surface; category 8: 100% of the affiliated department's surface. The model best supported by the data is used throughout these analyses.



Table C2: Parameters estimates from the best dynamic site-occupancy model for wolf in France between 1993 and 2014. The median a posteriori is given with the associated standard deviation (SD). Occasions 2, 3 and 4 correspond to January, February and March. Estimates are given on a logit scale except for alpha.psi which is given on its natural scale, i.e. [0, 1].

Parameters	Median	SD	Interpretation
alpha.psi	$3.17 \times 10^{-3}$	$1.91 \times 10^{-3}$	Initial occupancy
alpha.p	-2.17	0.12	Detection intercept
beta.SEffp	0.26	$3.03 \times 10^{-2}$	Effect of sampling effort on detection
beta.Rdensp	-0.93	$5.85 \times 10^{-2}$	Effect of accessibility (road density) on detection
beta.occp2	0.47	$7.97 \times 10^{-2}$	Effect of occasion 2 on detection
beta.occp3	0.37	$8.29 \times 10^{-2}$	Effect of occasion 3 on detection
beta.occp4	0.33	$8.04 \times 10^{-2}$	Effect on occasion 4 on detection
alpha.gamma	-5.87	0.12	Colonization intercept
beta.agamma	0.65	0.11	Effect of farmland cover on colonization
beta.SDACgamma	0.68	$4.13 \times 10^{-2}$	Effect of short distance occupied neighboring cells on colonization
beta.LDACgamma	0.54	$6.02 \times 10^{-2}$	Effect of long-distance occupied neighboring cells on colonization
beta.fgamma	0.90	$7.97 \times 10^{-2}$	Effect of forest cover on colonization
beta.altgamma	0.94	$7.57 \times 10^{-2}$	Effect of mean altitude on colonization
beta.haltgamma	-0.24	$4.30 \times 10^{-2}$	Effect of high altitude proportion on colonization

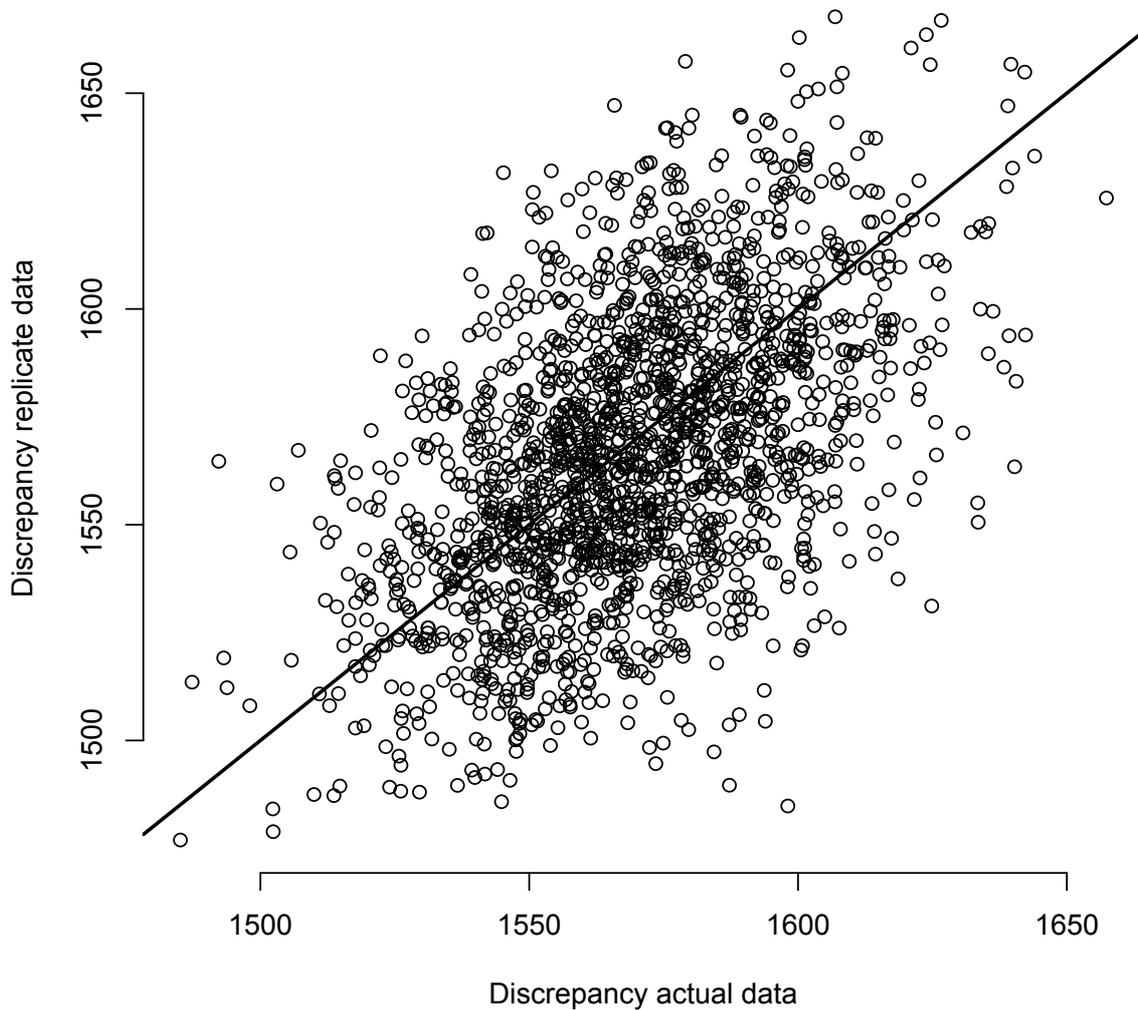


Figure B1: Results from posterior predictive checks for the dynamic occupancy model best supported by the wolf data. We show a scatterplot of the predicted chi-square discrepancy between simulated and expected data (on the Y axis) versus the observed chi-square discrepancy between expected and observed data (on the X axis) across MCMC samples. The Bayesian predictive p-value is 0.46 and represents the proportion of samples above the diagonal. Overall, the fit of the model seems satisfactory.

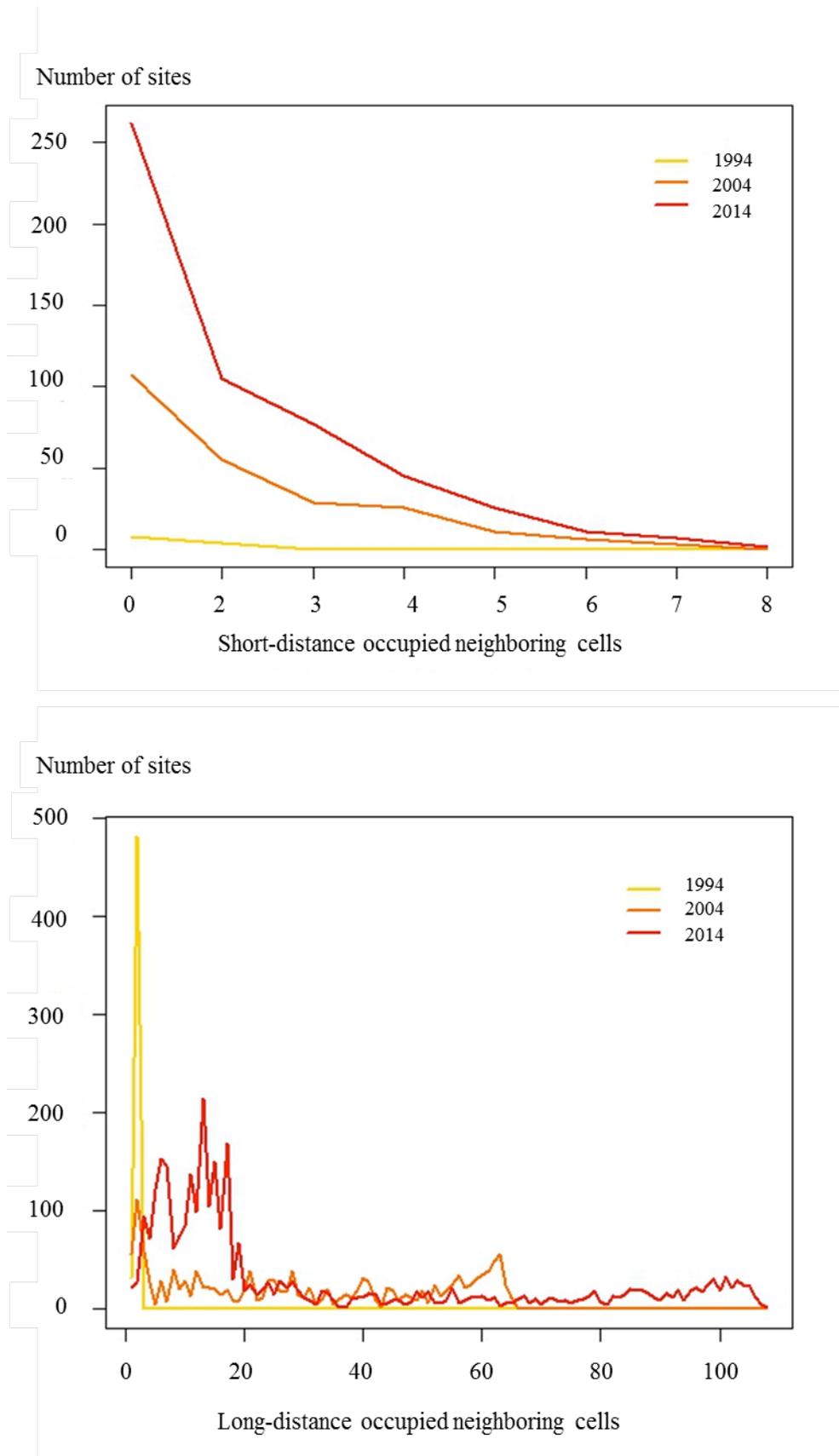


Figure D1: Number of sites having more than 0 observed occupied neighboring cells at short (contiguous cells) and long distance (between 10 km and 150 km) for 1994, 2004, and 2014.

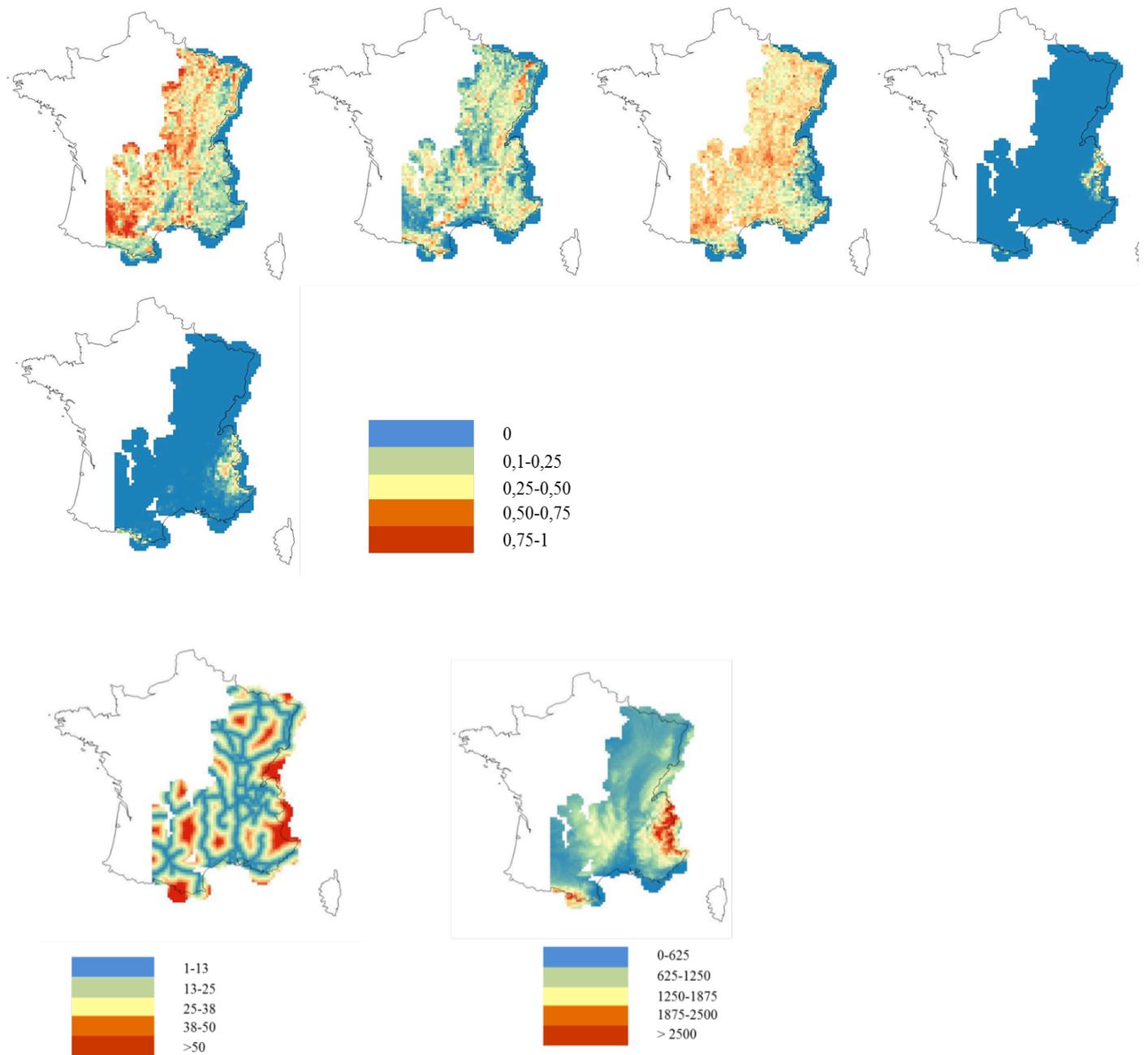


Figure D2: Maps of covariates tested in the study. First row from left to right: proportion of farmland cover (from 0 to 1); proportion of forest cover (from 0 to 1); road density (from 0 to 1); high altitude density (from 0 to 1). Second row: proportion of rock cover (from 0 to 1). Third row: left: distance from the cell center to the closest barrier (highway or river); right: mean altitude (in meters).

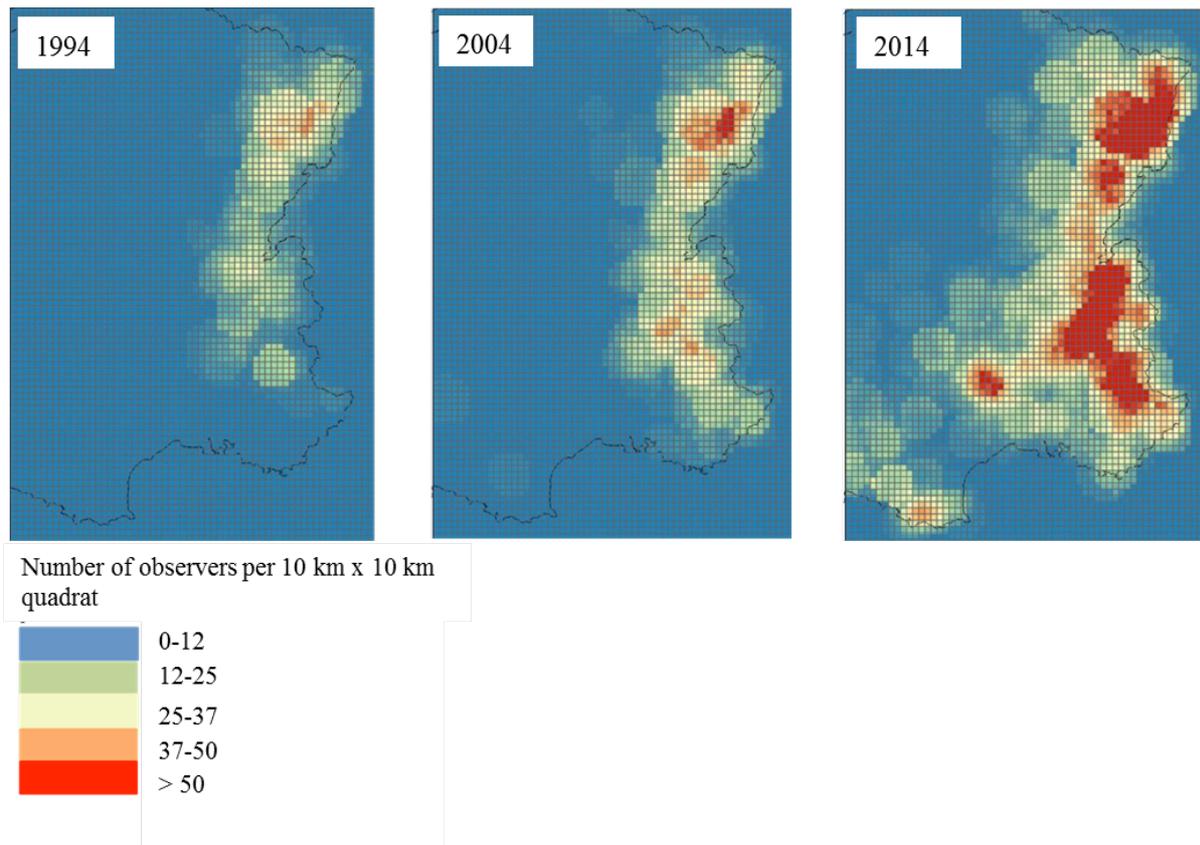


Figure D3: Maps of sampling effort for 1994, 2004, and 2014. Sampling effort was defined as the number of prospecting observers per site per year. In 1994, 1036 sites were prospecting by at least one observer and 3083 were prospecting in 2014. In 1994 only sites in North-Eastern part of France were prospecting and in 2014 all Eastern France was prospecting with some parts in South-West. Size of the prospecting area depended on the socio-professional category of observers: observers from an administrative field (policemen for instance) were assigned a theoretical prospecting area the size of the French department in which they were affiliated, observers from national parks, regional natural parks and natural reserves were assigned a prospecting area the size of those areas. Observers from the farming profession and hunters were assigned a prospecting area the size of the county they work in. Scientists, members of naturalist associations and private observers were assigned an area a quarter the surface of the French department where they found signs. Observers working in the French Game and Wildlife Agency (ONCFS) were assigned an area half the surface of the French department they work in. Observers from the French National Office for Forest (ONF) were assigned an area 1/10 the surface of the department they work in.