Spatial scale dependence in agricultural drivers of field margin plant communities

Isis Poinas[a,b,*], Guillaume Fried[b], Laura Henckel[c], Christine N Meynard[a]

[a] CBGP, INRAE, CIRAD, Institut Agro, IRD, Univ Montpellier, Montferrier-sur-Lez, France

[b] Anses, Laboratoire de la Santé des Végétaux, Unité Entomologie et Plantes invasives, Montferrier-sur-Lez, France

[c] INRAE Dijon, UMR Agro-écologie, Dijon, France

Authorship statement: I.P., G.F. and C.M. planned and designed the research; L.H. contributed to data processing; I.P. analyzed the data and wrote the first draft of the manuscript; G.F., L.H. and C.M. contributed substantially to revisions. All authors gave final approval for publication.

Data accessibility statement: Data available via the Data INRAE Repository, at https://doi.org/10.57745/NMIPCY

Number of words in abstract: 300

Number of words in main text: 6287

Number of cited references: 48

Number of tables: 4

Number of figures: 3

[*] Corresponding author. Tel: +33 6 86 53 41 80. Corresponding address: 755 avenue du campus Agropolis, CS 30016, 34988 Montferrier-sur-Lez Cedex, France
E-mail address: isis.poinas@inrae.fr.

Abstract

In recent decades, agricultural intensification has led to a strong decline in biodiversity. Field margins are defined as the uncultivated herbaceous vegetation strip located between the cropland and an adjacent habitat. Field margins act as shelters and dispersal corridors for biodiversity in highly disturbed landscapes, and are critical to the maintenance of ecosystem services. However, they are also impacted by agricultural practices in neighbouring fields. Agricultural impacts are often studied at field to landscape scales, and rarely include national spatial extents. One of the challenges in large-scale studies is the lack of standardized monitoring schemes including both biodiversity and accurate estimation of agricultural practices. Here, we take advantage of a national monitoring scheme in 462 sites in France, to assess the effects of agricultural practices on field margin flora at different extents and resolutions. We used spatial simultaneous autoregressive and generalized dissimilarity models to assess the response of plant richness and composition to climatic, soil and landscape conditions, and to agricultural (fertilization, herbicides) and margin management drivers. Analyses were repeated at the site-level, 40 and 75 km resolutions, and at regional and national extents. We found that the impact of agricultural practices on species richness was important at the site-level, whereas climate and crop diversity were important at the 75 km resolution. It was the opposite for compositional variations, with climate being more important at the site-level, and fertilization and crop diversity at the coarsest resolution. There was a strong variation in the variance explained by models among regions, but climate effects were weaker within biogeographic units compared to the national level. Regional analyses revealed interactions between local and regional effects suggesting that specific agricultural practices have contrasting impacts depending on the regional context. To efficiently conserve biodiversity, we therefore suggest implementing an agricultural regulatory framework adapted to each region.

Keywords: field margin, arable plants, agroecosystem, spatial scale, resolution, extent, species pool, crop diversity, agricultural practices

Introduction

Since the second half of the 20th century, agricultural intensification has resulted in steady declines in biodiversity across a wide range of taxa and habitats (Emmerson et al., 2016). The mechanization of agriculture and the resulting aggregation of cropped fields has contributed to the loss of semi-natural areas, such as herbaceous field margins, defined here as uncultivated vegetated strips located between the cropland and the adjacent habitat. Despite their modest area in agroecosystems, field margins play an important ecological role as sources of food, habitat and dispersal corridors for endangered species (Storkey et al., 2011), pollinators (Zamorano et al., 2020) and pest predators (Woodcock et al., 2016). Locally, they also buffer pesticide and fertilizer drift into adjacent terrestrial and aquatic habitats (Haddaway et al., 2018). Their importance has been increasingly recognized in recent European agricultural policies, for example through the conditioning of a green payment provided that margin strips are preserved (Matthews, 2013). Field margins, as ecotones, contain a mix of weed and ruderal species well adapted to agricultural disturbances, and some species typical of grasslands and forests (Aavik & Liira, 2009), making them valuable for studying the unintended impacts of agriculture on biodiversity.

At the field scale, agricultural practices play a key role in shaping the species distribution and diversity of plants in field margins (Marshall & Moonen, 2002). Chemical inputs such as pesticides and fertilizers often drift to the margins creating unintended disturbances in field margin communities. Practices such as sowing date, fertilization level, herbicide use and management of the margin flora using mowing, herbicides or grazing, vary by crop and production type, selecting species that are adapted to these disturbances (Fried et al., 2018; Bassa et al., 2011). Agricultural practices also impact dispersal dynamics (Tscharntke et al., 2005). At the landscape scale, the proportion and diversity of non-crop habitats (Rader et al., 2014) and crop diversity (Sirami et al., 2019) have positive effects on field margin biodiversity. At the regional and national scale, cropping systems are defined by the sequence of crop successions along with the cultural practices associated to each crop (Leenhardt et al., 2010), which

are themselves correlated to climate and topography for example. In France, this has been the basis to define stable uniform agricultural units called “agricultural regions” (Richard-Schott, 2009). Long-term practices over these homogeneous regions could potentially influence the available species pool at larger spatial and temporal scales. It is therefore possible that a given agricultural practice has different local effects on biodiversity depending on the regional species pool (Cornell & Harrison, 2014). Considering local and regional effects at different scales is therefore important to have a better understanding of agricultural practices that could favor biodiversity.

In ecology, spatial scale is usually defined by two attributes: resolution (or grain) and extent, both of them having a strong importance on the processes unveiled. Spatial grain is the size of the unit of analysis. It is often assumed that coarser resolutions allow ignoring local stochasticity to focus on the main patterns structuring communities (Chase, 2014). For example, landscape is expected to have the greatest influence when the grain size is closest to the average area of the different habitats. However, using excessively coarse resolutions may lead to uninformative patterns (Viana & Chase, 2019). Spatial extent is the total area covered by the sampling. It determines the range of variability in some environmental gradients, as well as the potential of inclusion of dispersal processes (Viana & Chase, 2019). Naturally, local filters such as farming practices are easier to detect at small extents and within biogeographic regions, where different study sites are more likely to represent homogeneous conditions for large-scale gradients such as climate (Viana et al., 2016).

Few studies have examined the integration of several spatial scales in agroecosystems. Some have suggested that agricultural effects could prevail at fine spatial grains. For instance, Guerrero et al. (2014) found that weed species richness in Spain decreased linearly with field-level intensification, but remained constant at the landscape level. However, Berquer et al. (2021), focusing at a resolution under 1 km, found the opposite pattern, with field margin communities showing a greater influence of landscape than agricultural practices. Concepción et al. (2012) revealed that extensive farming has a weaker positive impact on weed diversity at the field center and inner field edge when the landscape

is less diversified within a 500 m radius. This suggests that landscape-scale simplification or intensification may result in a depletion of the regional species pool, which then becomes primarily composed of species adapted to agricultural disturbances. Regarding spatial extent, although some studies have addressed national (Rader et al., 2014), continental (Billeter et al., 2008) or regional scales (Nascimbene et al., 2012), they have included a small sample size or lacked standard survey protocols across plots, and often used proxies rather than direct measures of the targeted agricultural practices. Therefore, issues of scale in agroecosystems remain largely under-represented in the literature.

Here we aimed at studying the effects of spatial resolution and extent, for a better understanding of the effects of agricultural practices on plant diversity and composition in field margins. We relied on a standardized national monitoring effort in 462 agricultural field margins covering France between 2013 and 2019 and including monitoring of plant communities as well as agricultural practices. Initial analyses of this dataset at a site-level showed that vegetation composition was primarily structured by landscape and soil, and secondarily by agricultural intensification (Andrade et al., 2021; Fried et al., 2018). Composition was driven by margin management and fertilization, while species richness responded to herbicide use. Building on these first findings, we investigated the potential influence of spatial extent and resolution of analysis. Regarding spatial resolution, our first hypothesis was that the effect of local agricultural practices on field margin plant communities would be easier to detect at finer grains because farmers' practices are specific to each field and can be very different between two neighbouring fields. Our second hypothesis was that landscape, soil and climatic conditions, which are more heterogeneous at larger spatial grains, would predominate over local practices in explaining field margin plant communities at coarser resolutions. Regarding spatial extent, our hypothesis was that the effects of soil and climate on field margin plant communities would dominate at the national extent and become less important within biogeographic regions, providing a clearer picture of the influence of agricultural practices at the regional extent.

Materials and methods

Vegetation surveys

We used vegetation data from a national monitoring effort, the 500-ENI network, which is funded by the French Ministry of Agriculture (see details in Andrade et al., 2021). Note that raw data access requires a request to the Ministry and is conditional on confidentiality of some information, such as site coordinates; however, all datasets used in the following analyses are available in a repository (see the data accessibility statement). Agricultural field margins covering continental France, were surveyed yearly since 2013, representing a total of 543 unique sites between 2013 and 2019 (including some site turnover between years, see **Appendix A**, **Fig. SA.1**). Here we selected a subset of sites that had at least five years of botanical data between 2013 and 2019, resulting in a total of 462 sites for the rest of the analysis. These survey sites were located in field margins representing four main crop types: annual crops (with winter wheat or maize as the main crop production, but including other crops in the rotation), market gardening crops (mainly lettuce) and vineyards. The proportion of sites under organic farming was roughly 20%, but agricultural practices covered a wide range of pesticide application, fertilizers and soil management. Botanical surveys were performed at peak flowering (between the end of April and the beginning of August, depending on the region). At the national scale, this represented 3079 observations (year x site) and 689 plant species. The transect line within each site was located in the middle of the field margin, equidistant from the cropland and the adjacent habitat. Along this transect, plant species were identified in ten 1 m^2 sub-plots (**Appendix A**, **Fig. SA.2**). Presence-absence was recorded for each species and observation. Here we used frequency of occurrence per species and site averaged across years (0 = was never detected in that site; 1 = was registered for all surveyed years in that site), as an index of relative abundance in the species-by-site matrices.

Explanatory variables

We gathered two sets of explanatory variables: the first set came directly from the 500-ENI network and reflects agricultural practices assessed directly on the sites (see Andrade et al., 2021); the second set was compiled from external open access databases. These include soil, climate and landscape data (see below).

Agricultural practices were reported yearly from interviews of farmers into an online database according to a national standardized acquisition protocol. This data related to fertilization, pesticide use, tillage and boundary management (**Appendix A**, **Fig. SA.3**). Soil characteristics in the 0-5 cm layer were acquired at a resolution of 250 m from SoilGrids (Hengl et al., 2014). Climatic data by year and season were obtained from the Chelsa database for the period 1979-2013, at a resolution of 1 km (Karger et al., 2017). We extracted landscape descriptors from OSO land cover maps at a 10 m resolution for 2016 (Inglada et al., 2017). This is a median year with respect to the temporal period covered by our sampling, and we assumed that landscape cover did not change dramatically between 2013 and 2019. We considered percent cover of eight landscape classes (**Appendix A**, **Fig. SA.3**) within buffers of 1 km radius, and used Shannon diversity index (SHDI) (Turner & Gardner, 2015) to characterize compositional landscape heterogeneity. A Shannon's diversity index was also estimated for crop landscape diversity within a buffer of 1 km, considering 17 types of crops (**Appendix A**, **Fig. SA.3**) retrieved from the external databases OSO (Inglada et al., 2017), TOPO (IGN, 2019) and the Graphical Parcel Register (IGN & ASP, 2019).

Our dataset contained missing values for agricultural drivers that were occasionally left empty by the local observers. We used a multivariate imputation method based on a random forest algorithm (package MICE of R v.4.0.0, Buuren & Groothuis-Oudshoorn, 2011) to complete these data (see **Appendix B** for details).

For compositional analyses, all explanatory variables were standardized to have equal mean and standard deviation. We limited the number of explanatory variables to seven (**Table 1**) so that we could include the same ones at different resolutions (see section on data analysis below). We aimed at

representing the essential variables in different categories (climate, soil, landscape and practices) ensuring that they were not highly correlated (Pearson correlation coefficient < 0.6). We retained mean annual temperature and soil pH, since they are tightly related to floristic composition and richness in agroecosystems (Fried et al., 2008). Landscape and crop diversity were selected as indicators of landscape heterogeneity (Rader et al., 2014; Sirami et al., 2019). Finally, we included the herbicide treatment frequency index, the nitrogen dose applied in fertilizers (within adjacent agricultural fields) and the number of margin management events (**Table 1**). These variables have consistently been reported to have a significant effect on field margin communities (Aavik & Liira, 2009; Bassa et al., 2011; Fried et al., 2018).

Data analysis

We used two statistical methods: spatial simultaneous autoregressive models (SAR) with species richness as response variable, and generalized dissimilarity models (GDM) to study species composition while dealing with nonlinear relationships between environmental distance and species turnover (Ferrier et al., 2007). These two types of analyses were applied at the national level using different resolutions, and then at regional and national extents using the site-level resolution.

SARs were implemented (package SPDEP, function *errorsarlm*) using a correlated spatial error term, which is added as a predictor and is a function of the spatial neighborhood of the response variable (**Appendix C**) (Cressie, 2015). We can thus model linear relationships that take into account spatial autocorrelation in the data, i.e. the tendency of nearby points to have more similar values than expected by chance. Hereafter, the term "spatial structuring" will be used interchangeably with "spatial autocorrelation", which can result from an effect of the environment or from an internal process linked to community assembly, such as dispersal limitations (Borcard et al., 2018). Species richness was transformed with a square root to comply with normality. We assessed the residual autocorrelation of each model with a Moran's I test (Thioulouse et al., 2018). We also computed partial

193 regressions using the Nagelkerke pseudo-R^2 to quantify the relative importance of each predictor
194 (Lichstein et al., 2002; **Appendix C**).

195 The use of GDMs (R package $_{GDM}$, function *gdm*) to model species composition is appropriate when
196 species turnover rate varies along the environmental gradients and hence when the response is non-
197 linear. The Bray-Curtis dissimilarity among all pairs of sites was modeled as a function of environmental
198 distance (computed for each factor) and geographical distance (Ferrier et al., 2007). Geographical
199 distances here account for spatial structure. The environmental table was permuted 500 times to
200 assess variable significance, as appropriate for analyses with distance matrices. We then evaluated the
201 importance of each variable in the model with the same procedure as above (partial regressions based
202 on the explained deviance; Maestri et al., 2017).

203 The different analyses detailed in the above sections were reiterated at different spatial scales. The
204 first approach was to increase the study resolution by aggregating data into grid cells. We present here
205 results from three resolutions: site-level, 40 km, and 75 km, (**Fig. 1A,** but see **Appendix D & E** for other
206 resolutions). To avoid cases where single sites with extreme values drive patterns at the aggregated
207 scales, cells with sites located at less than 5 km within the same grid cells were eliminated, as were
208 cells containing only one site and cells in the land - sea borders with more than half of their surface
209 covered by water. We also shifted grid placement in three different iterations, to make sure results
210 were robust to the grid placement (**Appendix D**). Overall, this led to some variations in the number of
211 sites and grid cells across analyses, which were investigated by resampling (**Table 2**, **Appendix D**).
212 Whenever possible, explanatory data were directly extracted at the scale of interest (climate, soil and
213 landscape). All other factors were averaged by cell from site-level data, including the frequency of
214 occurrence by site. Species richness in each grid cell was obtained by aggregating site-level species
215 lists. Centroid coordinates of the grid cells were used in all spatial analyses for aggregated datasets.

216 Lastly, we carried out the same analysis using site-level data at two different extents: national and
217 regional. For this purpose, sites were grouped within ten biogeographical regions standing for the wide

218 pedo-climatic and floristic gradients in France (**Fig. 1B**, based on the VégétalLocal map; Office français
219 de la biodiversité, 2021). Only six regions had a sufficient number of sampling sites to be considered in
220 our analyses (>40 sites): Northern Parisian Basin, Southern Parisian Basin, Armorican Massif,
221 Mediterranean Zone, North-Eastern France and South-Western France.

222

223 **Results**

224 *Effects of spatial resolution*

225 At the site-level resolution, the predictors explained 27.8% of the variance in species richness.
226 Herbicides and nitrogen dose were the main drivers and explained respectively 1.5 ($p = 0.002$) and
227 1.3% ($p = 0.004$) of the total variance in species richness, both having a negative effect on richness.
228 With respect to plant community composition, the GDM explained 29.5% of deviance, the main drivers
229 being mean annual temperature (6.9%, $p < 0.001$) and soil pH (1.2%, $p < 0.001$).

230 With increasing spatial resolution, the importance of climate and landscape for species richness
231 increased, while agricultural effects remained constant and small (**Fig. 2A-C**). Temperature had a
232 positive correlation with richness ($p = 0.004$), while soil pH had no significant effect. Crop diversity was
233 the predominant landscape factor at all aggregated scales, showing a consistent positive correlation
234 with richness, but only significant at the coarsest resolution (75 km, $p = 0.017$, **Table 3A**). On the
235 contrary, landscape diversity did not seem to have any effect (**Table 3A**). Both herbicides and nitrogen
236 dose reduced the number of species at the site-level, but these effects were not significant at coarser
237 resolutions (**Table 3A**). Similarly, the positive correlation of margin management with species richness
238 was mainly apparent at the site-resolution ($p = 0.018$, **Table 3A**). The proportion of richness that can
239 be explained by spatial autocorrelation was higher at the site-resolution (11.4%, $p < 0.001$) compared
240 to coarser resolutions where it was non-significant (**Fig. 2A-C**).

241 With regard to composition, the R^2 of the models increased at coarser spatial grains (**Appendix E**, **Fig.**
242 **SE.1**). Climate and landscape (mainly through crop diversity) together contributed to overall
243 differences in species composition among sites, much more strongly than farming practices and soil at
244 all scales (**Fig. 2D-F**). Climate had less effect on composition at coarser resolutions while the effect of
245 landscape increased. Fertilization was more strongly related to species composition at the coarsest
246 resolution (75 km, 4.3%, p = 0.010) than at the finest, for which it was not significant (**Fig. 2D-F, Table**
247 **3B**). Herbicides did not have any effect on species composition. Also, spatial autocorrelation was
248 relatively important at all aggregated scales (**Fig. 2E-F**).

249 *Effects of spatial extent*

250 The influence of environmental filters on species richness and composition varied greatly among
251 regions (**Fig. 3**). For example, the high R^2 of the model for the Mediterranean zone (approximately
252 38.8%) contrasted with the lower one for the Armorican Massif (15.1%).

253 Some landscape and agricultural effects emerged at a regional extent as more significant than at the
254 national extent. For example, landscape diversity was negatively correlated with richness (6.5%, p =
255 0.016) in the northern Parisian Basin, a relationship that was not visible at the national extent with all
256 regions analyzed together (**Table 4A**). Agricultural effects on richness were often more predominant
257 at the regional extent (**Fig. 3A**). This scale allowed for a clearer identification of the effects of nitrogen
258 fertilization, herbicides, and margin management. The direction of the relationships between richness
259 and predictors remained consistent with analyses at the national level, except for south-western
260 France, which depicted a negative relationship of species richness with temperature, (p = 0.003, **Table**
261 **4A**).

262 With respect to composition, climatic conditions appeared to have higher explanatory values at a
263 national rather than regional extent (**Fig. 3B**). As with richness, new and contrasting effects of
264 landscape and practices were highlighted within regions, such as the effect of landscape diversity on
265 composition in the Mediterranean zone (2.2%, p = 0.020), and the effects of the number of margin

management events in the Mediterranean zone (5.6%, $p < 0.001$) and Armorican Massif (6.9%, $p < 0.001$) (**Fig. 3B**, **Table 4B**).

Discussion

In line with our initial hypotheses, the effects of agricultural practices on field margin flora were more prominent within biogeographic regions. Herbicides and fertilization were important to explain species richness at the site-level, whereas fertilization stood out in explaining species composition at coarse resolutions. Conversely, climate and landscape variables always dominated when data were aggregated, and when analyzed at the national extent. Climate also explained a large percentage of variation in species composition, but not in species richness, at the site-level.

Influence of spatial resolution

Composition was influenced by large-scale landscape and agricultural effects, such as crop diversity at the landscape level and high fertilization in the region, while richness was mostly impacted by field-scale land use intensification, such as high herbicide use and nitrogen fertilization. This is consistent with some studies reporting that the effects of local agricultural practices on field margin plant communities are at least as important to landscape effects (Bassa et al., 2011; Gabriel et al., 2006; but see Martin et al., 2020). Negative impacts of herbicides and nitrogen on field margin richness have already been reported (Aavik & Liira, 2009; Fried et al., 2018), with nitrogen fertilization affecting both richness and composition, and herbicides only impacting richness (Fried et al., 2018). Our results revealed a large-scale effect of nitrogen use on composition, which contrasts with previous studies showing its effect on richness (Billeter et al., 2008; Kleijn et al., 2009). Storkey et al. (2011) demonstrated that large-scale eutrophication is a major explanatory factor of the threat status of arable species in Europe and even more than herbicide use, although the two factors are difficult to disentangle.

290 Regarding landscape variables, crop diversity was the most important predictor of richness and
291 composition at large scales. While the response of plant species richness to Shannon crop diversity is
292 inconsistent among studies (Martin et al., 2020; Sirami et al., 2019), we found here a consistently
293 positive effect regardless of the scale. This can be viewed as landscape complementation for species
294 that are specialists of a single crop (Storkey et al., 2011). Another explanation would be that crop
295 diversity is negatively related to field size (Martin et al., 2020), and this latter might have an additional
296 effect on richness (Fahrig et al., 2015). In contrast, habitat diversity did not show any effect on richness,
297 but only on composition at the site level. Although a diverse landscape does not necessarily support
298 more species, it may favor functionally more diverse communities, supporting the idea that species
299 richness alone cannot capture all aspects of biodiversity (Aavik et al., 2008; Fried et al., 2018). The
300 effect of crop diversity on composition increased with resolution, while this was not the case for
301 landscape diversity, likely because landscapes became more homogeneous in terms of habitat
302 diversity, while they still remained different in terms of crops (**Appendix F**, **Fig. SF.1**).

303 *Influence of spatial extent*

304 More interestingly, our results pointed to strong regional specificities, underlining complex
305 interactions between local and regional scales. For example, the influence of landscape diversity on
306 richness was very high in the northern Parisian Basin, a region mainly composed of large cereal open-
307 fields (**Appendix F**, **Fig. SF.2-3**). The direction of this effect was unexpected, since landscape diversity
308 decreased the number of species in field margins. Given the high level of local intensification in this
309 region, the species from natural habitats are likely too vulnerable to establish locally. As diverse
310 landscapes have fewer croplands and more natural habitats, species adapted to agricultural
311 disturbances are less prevalent in the landscape species pool. The discrepancy between the low level
312 of landscape disturbance and the high level of local disturbance ultimately results in a local loss of
313 species. This apparently counter-intuitive finding could thus be explained by interactions between the
314 regional pool and local conditions (Cornell & Harrison, 2014; Myers & Harms, 2009). Likewise, the

315 detrimental effect of nitrogen fertilization on richness was only significant in the Mediterranean Zone.
316 In this region, soils had low nitrogen content (**Appendix F**, **Fig. SF.3**), and the species pool is mostly
317 composed of stress-tolerant species adapted to nutrient-poor soil conditions. Due to its regional
318 composition, Mediterranean flora would thus be more sensitive to soil enrichment. This supports the
319 results of Kleijn et al. (2009) who concluded that the impact of nitrogen is greater in less intensive
320 regions. Clarifying regional specificities should thus be a strong priority for future agroecological
321 studies, as it paves the way for proposing concrete agricultural management practices that minimize
322 impacts on biodiversity while considering regional characteristics.

323 *Implications for management in agricultural context*

324 Our results have practical implications for crop management. Crop diversity stands out with a major
325 positive impact on plant diversity in field margins. The importance of crop diversity has just begun to
326 be recognized in public policies (Galán-Martín et al., 2015), and our results suggest that this effect will
327 extend to a large scale. Promoting a wide range of crops at the landscape level could therefore slow
328 the decline of some highly specialized arable plant species, particularly in landscapes with large semi-
329 natural cover (Sirami et al., 2019). Fertilization significantly altered floristic assemblages and is also
330 more impactful at a large scale, which is of major interest for national agricultural policies.

331 Beyond large-scale agricultural impacts, the effects of local agricultural practices are also regionally
332 dependent. The region in which measures are implemented can determine their cost-benefit (Kleijn et
333 al., 2009). For instance, a drastic reduction of fertilizers in highly intensive regions mainly composed of
334 nitrophilous species, such as the Parisian Basin here, will have little impact on plant communities of
335 field margins if this action is not coupled with landscape restoration. In contrast, implementing stricter
336 fertilization control in the Mediterranean zone, where plants are adapted to poor soils, would have a
337 much higher impact on the native flora of this region. Diversified areas will also be more susceptible
338 to respond to less costly mitigation measures and are of higher conservation concern (Stevens et al.,
339 2010). We therefore encourage agricultural stakeholders and scientists involved in monitoring

programs to further evaluate the changes in field margin flora and its protection measures across different spatial scales, and implement regulatory measures that incorporate landscape-level management planning adapted to each biogeographic and agricultural region.

Acknowledgements

The 500-ENI network is supported by the French Ministry of Agriculture under the Ecophyto framework with funding from the French Biodiversity Agency (Office Français de la Biodiversité). We would like to thank everyone that has collected data in the field, the farmers who provided information on their practices, and everyone involved in the coordination of the 500-ENI data network.

Funding: This work was supported by an INRAE-ANSES thesis fellowship and two Ecophyto II+ projects: GT STEP 500 ENI and GTP 500 ENI (OFB-21-1642).

Appendix A-F. Supplementary data

Supplementary data associated with this article can be found, in the online version, at XXXXX.

References

Aavik, T., Augenstein, I., Bailey, D., Herzog, F., Zobel, M., & Liira, J. (2008). What is the role of local landscape structure in the vegetation composition of field boundaries? *Applied Vegetation Science*, *11*(3), 375–386. https://doi.org/10.3170/2008-7-18486

Aavik, T., & Liira, J. (2009). Agrotolerant and high nature-value species—Plant biodiversity indicator groups in agroecosystems. *Ecological Indicators*, *9*(5), 892–901. https://doi.org/10.1016/j.ecolind.2008.10.006

Andrade, C., Villers, A., Balent, G., Bar-Hen, A., Chadoeuf, J., Cylly, D., Cluzeau, D., Fried, G., Guillocheau, S., Pillon, O., Porcher, E., Tressou, J., Yamada, O., Lenne, N., Jullien, J., & Monestiez, P. (2021). A real-world implementation of a nationwide, long-term monitoring program to assess the impact of agrochemicals and agricultural practices on biodiversity. *Ecology and Evolution*, *11*(9), 3771–3793. https://doi.org/10.1002/ece3.6459

Bassa, M., Boutin, C., Chamorro, L., & Sans, F. X. (2011). Effects of farming management and landscape heterogeneity on plant species composition of Mediterranean field boundaries. *Agriculture, Ecosystems & Environment*, *141*(3–4), 455–460. https://doi.org/10.1016/j.agee.2011.04.012

Berquer, A., Martin, O., & Gaba, S. (2021). Landscape Is the Main Driver of Weed Assemblages in Field Margins but Is Outperformed by Crop Competition in Field Cores. *Plants*, *10*(10), 2131. https://doi.org/10.3390/plants10102131

Billeter, R., Liira, J., Bailey, D., Bugter, R., Arens, P., Augenstein, I., Aviron, S., Baudry, J., Bukacek, R., Burel, F., Cerny, M., Blust, G. D., Cock, R. D., Diekötter, T., Dietz, H., Dirksen, J., Dormann, C., Durka, W., Frenzel, M., … Speelmans..., M. (2008). Indicators for Biodiversity in Agricultural Landscapes: A Pan-European Study. *Journal of Applied Ecology*, *45*(1), 141–150. https://doi.org/10.1111/j.1365-2664.2007.01393.x

Borcard, D., Gillet, F., & Legendre, P. (2018). *Numerical Ecology with R*. Springer.

Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(1), 1–67. https://doi.org/10.18637/jss.v045.i03

Chase, J. M. (2014). Spatial scale resolves the niche versus neutral theory debate. *Journal of Vegetation Science*, *25*(2), 319–322. https://doi.org/10.1111/jvs.12159

Concepción, E. D., Díaz, M., Kleijn, D., Báldi, A., Batáry, P., Clough, Y., Gabriel, D., Herzog, F., Holzschuh, A., Knop, E., Marshall, E. J. P., Tscharntke, T., & Verhulst, J. (2012). Interactive effects of landscape context constrain the effectiveness of local agri-environmental management: Landscape constrains the effectiveness of local management. *Journal of Applied Ecology*, *49*(3), 695–705. https://doi.org/10.1111/j.1365-2664.2012.02131.x

Cornell, H. V., & Harrison, S. P. (2014). What Are Species Pools and When Are They Important? *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 45–67. https://doi.org/10.1146/annurev-ecolsys-120213-091759

Cressie, N. (2015). *Statistics for Spatial Data*. John Wiley & Sons.

Emmerson, M., Morales, M. B., Oñate, J. J., Batáry, P., Berendse, F., Liira, J., Aavik, T., Guerrero, I., Bommarco, R., Eggers, S., Pärt, T., Tscharntke, T., Weisser, W., Clement, L., & Bengtsson, J. (2016). How Agricultural Intensification Affects Biodiversity and Ecosystem Services. In *Advances in Ecological Research* (Vol. 55, pp. 43–97). Elsevier. https://linkinghub.elsevier.com/retrieve/pii/S0065250416300204

Fahrig, L., Girard, J., Duro, D., Pasher, J., Smith, A., Javorek, S., King, D., Lindsay, K. F., Mitchell, S., & Tischendorf, L. (2015). Farmlands with smaller crop fields have higher within-field biodiversity. *Agriculture, Ecosystems & Environment*, *200*, 219–234. https://doi.org/10.1016/j.agee.2014.11.018

Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, *13*(3), 252–264. https://doi.org/10.1111/j.1472-4642.2007.00341.x

Fried, G., Norton, L. R., & Reboud, X. (2008). Environmental and management factors determining weed species composition and diversity in France. *Agriculture, Ecosystems & Environment*, *128*(1–2), 68–76. https://doi.org/10.1016/j.agee.2008.05.003

Fried, G., Villers, A., & Porcher, E. (2018). Assessing non-intended effects of farming practices on field margin vegetation with a functional approach. *Agriculture, Ecosystems & Environment*, *261*, 33–44. https://doi.org/10.1016/j.agee.2018.03.021

Gabriel, D., Roschewitz, I., Tscharntke, T., & Thies, C. (2006). Beta diversity at different spatial scales: Plant communities in organic and conventional agriculture. *Ecological Applications*, *16*(5), 2011–2021. https://doi.org/10.1890/1051-0761(2006)016[2011:BDADSS]2.0.CO;2

Galán-Martín, Á., Pozo, C., Guillén-Gosálbez, G., Antón Vallejo, A., & Jiménez Esteller, L. (2015). Multi-stage linear programming model for optimizing cropping plan decisions under the new Common Agricultural Policy. *Land Use Policy*, *48*, 515–524. https://doi.org/10.1016/j.landusepol.2015.06.022

Guerrero, I., Carmona, C. P., Morales, M. B., Oñate, J. J., & Peco, B. (2014). Non-linear responses of functional diversity and redundancy to agricultural intensification at the field scale in Mediterranean arable plant communities. *Agriculture, Ecosystems & Environment*, *195*, 36–43. https://doi.org/10.1016/j.agee.2014.05.021

Haddaway, N. R., Brown, C., Eales, J., Eggers, S., Josefsson, J., Kronvang, B., Randall, N. P., & Uusi-Kämppä, J. (2018). The multifunctional roles of vegetated strips around and within agricultural fields. *Environmental Evidence*, *7*(1), 14. https://doi.org/10.1186/s13750-018-0126-2

Hengl, T., Jesus, J. M. de, MacMillan, R. A., Batjes, N. H., Heuvelink, G. B. M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J. G. B., Walsh, M. G., & Gonzalez, M. R. (2014). SoilGrids1km—Global Soil Information Based on Automated Mapping. *PLOS ONE*, *9*(8), e105992. https://doi.org/10.1371/journal.pone.0105992

IGN. (2019). *BD TOPO*. https://geoservices.ign.fr/bdtopo

IGN, & ASP. (2019). *Registre Parcellaire Graphique*. https://geoservices.ign.fr/rpg

Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., & Rodes, I. (2017). Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sensing*, *9*(1), 95. https://doi.org/10.3390/rs9010095

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*(1), 170122. https://doi.org/10.1038/sdata.2017.122

Kleijn, D., Kohler, F., Báldi, A., Batáry, P., Concepción, E. d, Clough, Y., Díaz, M., Gabriel, D., Holzschuh, A., Knop, E., Kovács, A., Marshall, E. j. p, Tscharntke, T., & Verhulst, J. (2009). On the relationship between farmland biodiversity and land-use intensity in Europe. *Proceedings of*

the Royal Society B: Biological Sciences, *276*(1658), 903–909. https://doi.org/10.1098/rspb.2008.1509

Leenhardt, D., Angevin, F., Biarnès, A., Colbach, N., & Mignolet, C. (2010). Describing and locating cropping systems on a regional scale. A review. *Agronomy for Sustainable Development*, *30*(1), 131–138. https://doi.org/10.1051/agro/2009002

Lichstein, J. W., Simons, T. R., Shriner, S. A., & Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, *72*(3), 19. https://doi.org/10.1890/0012-9615(2002)072[0445:SAAAMI]2.0.CO;2

Maestri, R., Shenbrot, G. I., & Krasnov, B. R. (2017). Parasite beta diversity, host beta diversity and environment: Application of two approaches to reveal patterns of flea species turnover in Mongolia. *Journal of Biogeography*, *44*(8), 1880–1890. https://doi.org/10.1111/jbi.13025

Marshall, E. J. P., & Moonen, A. C. (2002). Field margins in northern Europe: Their functions and interactions with agriculture. *Agriculture, Ecosystems & Environment*, *89*(1–2), 5–21. https://doi.org/10.1016/S0167-8809(01)00315-2

Martin, A. E., Collins, S. J., Crowe, S., Girard, J., Naujokaitis-Lewis, I., Smith, A. C., Lindsay, K., Mitchell, S., & Fahrig, L. (2020). Effects of farmland heterogeneity on biodiversity are similar to—Or even larger than—The effects of farming practices. *Agriculture, Ecosystems & Environment*, *288*, 106698. https://doi.org/10.1016/j.agee.2019.106698

Matthews, A. (2013). Greening agricultural payments in the EU's Common Agricultural Policy. *Bio-Based and Applied Economics*, *2*(1), 1–27.

Myers, J. A., & Harms, K. E. (2009). Seed arrival, ecological filters, and plant species richness: A meta-analysis. *Ecology Letters*, *12*(11), 1250–1260. https://doi.org/10.1111/j.1461-0248.2009.01373.x

Nascimbene, J., Marini, L., & Paoletti, M. G. (2012). Organic Farming Benefits Local Plant Diversity in Vineyard Farms Located in Intensive Agricultural Landscapes. *Environmental Management*, *49*(5), 1054–1060. https://doi.org/10.1007/s00267-012-9834-5

Office français de la biodiversité. (2021). *Carte des régions d'origine—Référentiel technique de la marque Végétal local—Version 2021*. www.vegetal-local.fr

Rader, R., Birkhofer, K., Schmucki, R., Smith, H. G., Stjernman, M., & Lindborg, R. (2014). Organic farming and heterogeneous landscapes positively affect different measures of plant diversity. *Journal of Applied Ecology*, *51*(6), 1544–1553. https://doi.org/10.1111/1365-2664.12344

Richard-Schott, F. (2009). Le Recensement Général de l'Agriculture de 1955, une référence pour les géographes ? *Géocarrefour*, *2009*(4), Article Vol. 84/4. https://doi.org/10.4000/geocarrefour.7567

Sirami, C., Gross, N., Baillod, A. B., Bertrand, C., Carrié, R., Hass, A., Henckel, L., Miguet, P., Vuillot, C., Alignier, A., Girard, J., Batáry, P., Clough, Y., Violle, C., Giralt, D., Bota, G., Badenhausser, I., Lefebvre, G., Gauffre, B., … Fahrig, L. (2019). Increasing crop heterogeneity enhances multitrophic diversity across agricultural regions. *Proceedings of the National Academy of Sciences*, *116*(33), 16442–16447. https://doi.org/10.1073/pnas.1906419116

Stevens, C. J., Duprè, C., Dorland, E., Gaudnik, C., Gowing, D. J. G., Bleeker, A., Diekmann, M., Alard, D., Bobbink, R., Fowler, D., Corcket, E., Mountford, J. O., Vandvik, V., Aarrestad, P. A., Muller, S., & Dise, N. B. (2010). Nitrogen deposition threatens species richness of grasslands across Europe. *Environmental Pollution*, *158*(9), 2940–2945. https://doi.org/10.1016/j.envpol.2010.06.006

Storkey, J., Meyer, S., Still, K. S., & Leuschner, C. (2011). The impact of agricultural intensification and land-use change on the European arable flora. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1732), 1421–1429. https://doi.org/10.1098/rspb.2011.1686

Thioulouse, J., Dray, S., Dufour, A.-B., Siberchicot, A., Jombart, T., & Pavoine, S. (2018). Analysing Spatial Structures. In J. Thioulouse, S. Dray, A.-B. Dufour, A. Siberchicot, T. Jombart, & S. Pavoine (Eds.), *Multivariate Analysis of Ecological Data with ade4* (pp. 239–260). Springer. https://doi.org/10.1007/978-1-4939-8850-1_12

Tscharntke, T., Klein, A., Kruess, A., Steffan-Dewenter, I., & Thies, C. (2005). Landscape perspectives on agricultural intensification and biodiversity—Ecosystem service management. *Ecology Letters*, *8*(8), 857–874. https://doi.org/10.1111/j.1461-0248.2005.00782.x

Turner, M. G., & Gardner, R. H. (2015). Landscape Metrics. In M. G. Turner & R. H. Gardner (Eds.), *Landscape Ecology in Theory and Practice: Pattern and Process* (pp. 97–142). Springer. https://doi.org/10.1007/978-1-4939-2794-4_4

Viana, D. S., & Chase, J. M. (2019). Spatial scale modulates the inference of metacommunity assembly processes. *Ecology*, *100*(2), e02576. https://doi.org/10.1002/ecy.2576

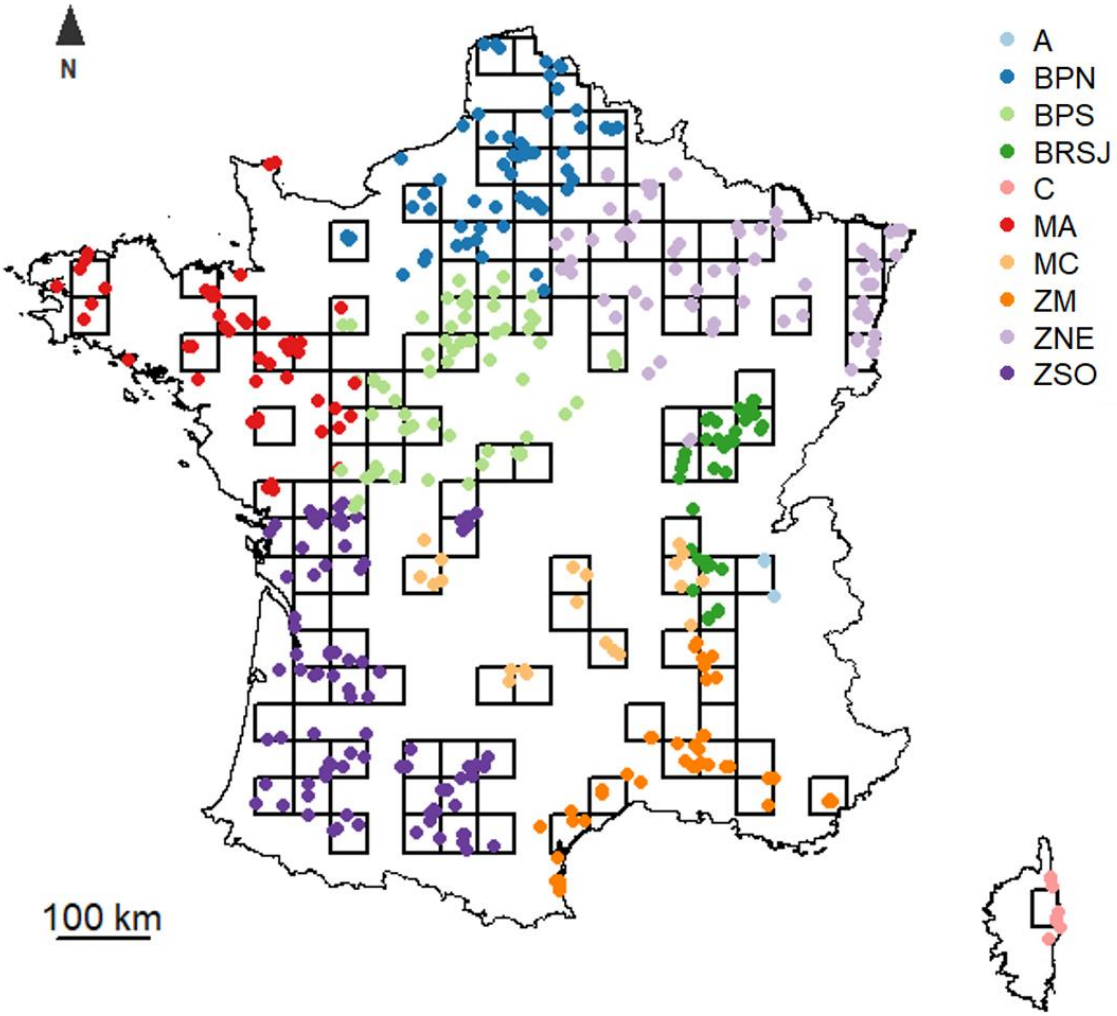
Viana, D. S., Figuerola, J., Schwenk, K., Manca, M., Hobæk, A., Mjelde, M., Preston, C. D., Gornall, R. J., Croft, J. M., King, R. A., Green, A. J., & Santamaría, L. (2016). Assembly mechanisms determining high species turnover in aquatic communities over regional and continental scales. *Ecography*, *39*(3), 281–288. https://doi.org/10.1111/ecog.01231

Woodcock, B. A., Bullock, J. M., McCracken, M., Chapman, R. E., Ball, S. L., Edwards, M. E., Nowakowski, M., & Pywell, R. F. (2016). Spill-over of pest control and pollination services into arable crops. *Agriculture, Ecosystems & Environment*, *231*, 15–23. https://doi.org/10.1016/j.agee.2016.06.023

Zamorano, J., Bartomeus, I., Grez, A. A., & Garibaldi, L. A. (2020). Field margin floral enhancements increase pollinator diversity at the field edge but show no consistent spillover into the crop field: A meta-analysis. *Insect Conservation and Diversity*, *13*(6), 519–531. https://doi.org/10.1111/icad.12454

510 **Figures**

511



512

513 **Fig. 1.** Distribution map of the field margins surveyed in France distributed in a grid of 40 km and
514 colored according to their biogeographic region. Only cells selected for analyses are presented. A =
515 « Alps », BPN = « Northern Parisian Basin », BPS = « Southern Parisian Basin », BRSJ = « Rhone-Saone-
516 Jura Basin », C = « Corsica », MA = « Armorican Massif », MC = « Massif Central », ZM =
517 « Mediterranean Zone », ZNE = « North-Eastern France » and ZSO = « South-Western France ». Note
518 that the regions C, A, MC and BRSJ have been removed from analyses because of their small number
519 of sites.

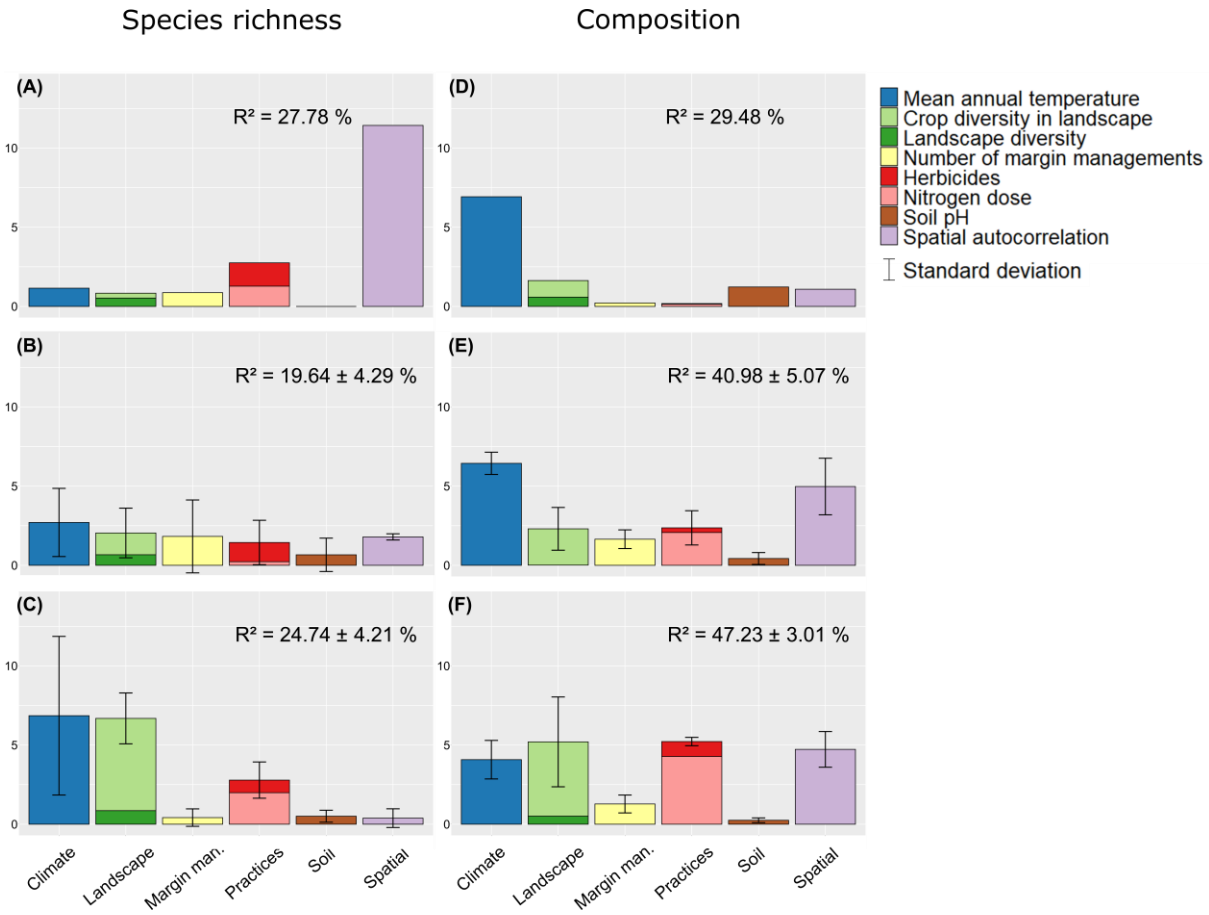


Fig. 2. Percentage of explained variance of species richness and composition by each factor for each spatial resolution. (A, D) 0 km; (B, E) 40 km; (C, F) 75 km. Standard deviations are computed from the three resamplings of grids. Percentages of explained variance are reported in **Appendix E, Table SE.1**.

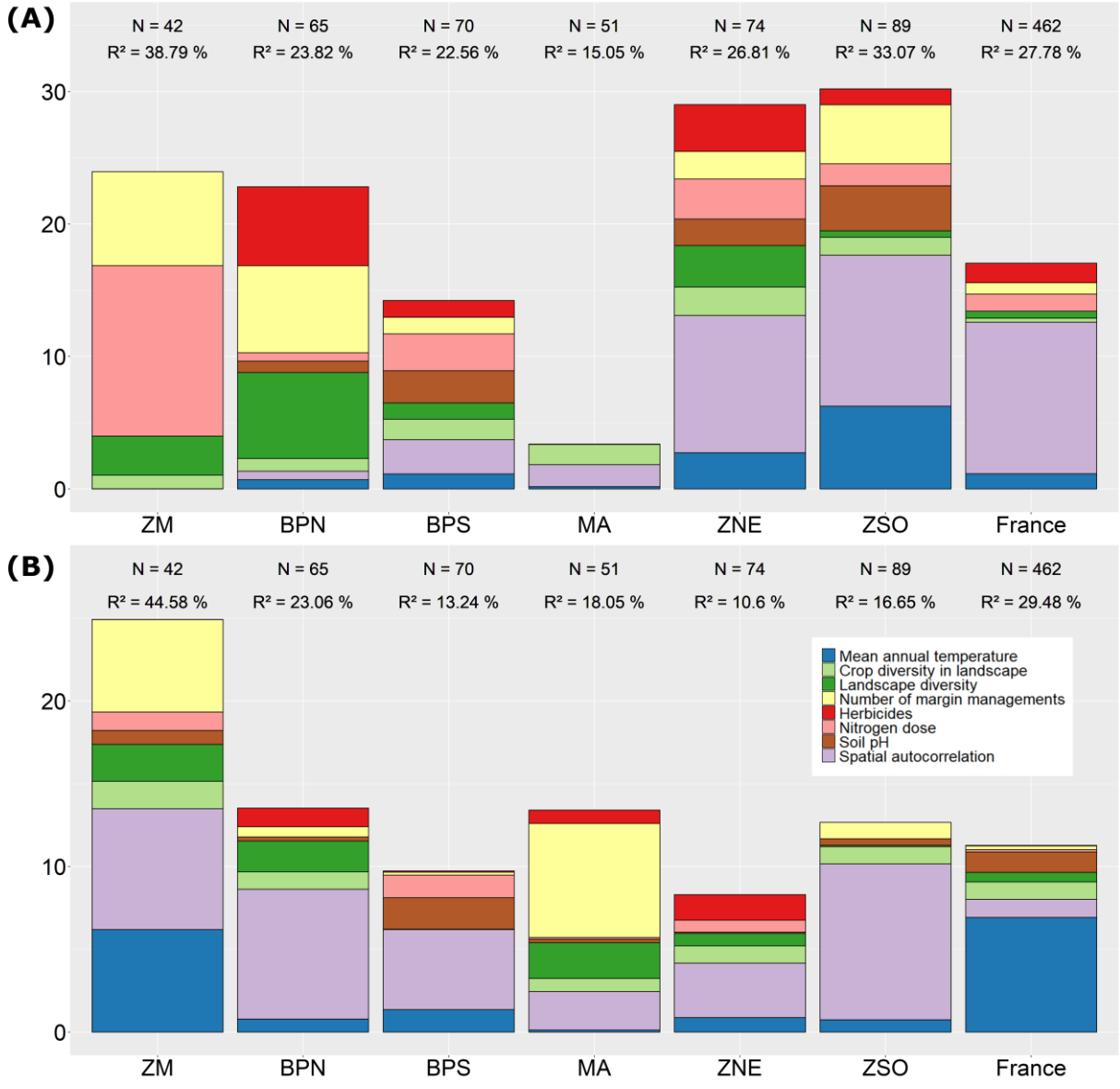


Fig. 3. Percentage of explained variance at the site-scale resolution for (A) richness and (B) composition for each biogeographic region and national extent. Regions are ordered by area from left to right. Percentages of explained variance are reported in **Appendix E, Table SE.2**.

528 **Tables**

529 **Table 1.** Description of explanatory factors used in analyses, with their abbreviations and scale of

530 computation. The variables computed by observation and then averaged by sites are underlined.

Factors	Abbreviations	Units	Type of factor	Data source	Scale of computation
Mean annual temperature	TEMP	°C x 10	Climate	Chelsa Climate	1 km
Soil pH (water)	PH	pH x 10	Soil	SoilGrids	250 m
Shannon's habitat diversity index	SHDI		Landscape	OSO	1 km
Shannon's crop diversity index	SHDI_C		Landscape	RPG	1 km
Dose of nitrogen (fertilization and amendments)	N_DOSE	kg/ha	Agricultural practices	500-ENI	Site
Treatment Frequency Index of herbicides	HERB		Agricultural practices	500-ENI	Site
Number of management events	MAN		Margin management	500-ENI	Site
Spatial structure	SPAT	m (Lambert 93)	Spatial	500-ENI	Site

531

Table 2. Sample size for each resolution. Grid positions in the aggregated resolutions were shifted to obtain three different positions of the grid system. Means and standard deviations below reflect variations across these grids within each resolution.

Spatial resolution	Site-level	40 km	75 km
Mean number of sites by cell		3.3 ± 0.1	5.8 ± 0.1
Mean number of sites	462	387 ± 9	428 ± 1
Mean number of cells		105 ± 3	71 ± 2

536 **Table 3.** Significance of each predictor (in row) for each spatial resolution (in column, in km). For
537 species richness in A), the direction of the relationship is indicated by +/-. The number of +/-/*
538 indicates the number of times the predictor was significant over the three grid shifts. We only report
539 significant relationships with p-values in brackets (the average p-value when significant in multiple
540 grids).

541 **A) Species richness**

	0	**40**	**75**
Temperature	+ (0.004)	+ (0.010)	++ (0.011)
Ph			
Landscape diversity			
Crop diversity			+++ (0.017)
Herbicides	- (0.002)	- (0.031)	
Nitrogen dose	- (0.004)		- (0.044)
Margin management	+ (0.018)	+ (0.019)	
Spatial structure	(< 0.001)		

542

543

544 **B) Composition**

	0	**40**	**75**
Temperature	(< 0.001)	*** (< 0.001)	*** (0.020)
Ph	(< 0.001)		
Landscape diversity	(< 0.001)		
Crop diversity	(< 0.001)	** (< 0.001)	*** (0.020)
Herbicides			
Nitrogen dose		** (< 0.001)	*** (0.010)
Margin management		* (< 0.001)	

Spatial structure	(< 0.001)	*** (< 0.001)	*** (< 0.001)

546 **Table 4.** Significance of each predictor (in row) for each biogeographic region and for national extent

547 (in column, see **Fig. 1** for abbreviations of region names). For species richness in A), the direction of

548 the relationship is indicated by +/-. We only report significant relationships with p-values in brackets.

549 **A) Species richness**

	BPN	BPS	MA	ZM	ZNE	ZSO	France
Temperature						- (0.003)	+ (0.004)
Ph						- (0.048)	
Landscape diversity	- (0.016)						
Crop diversity				+ (0.021)			
Herbicides	- (0.034)						- (0.002)
Nitrogen dose				- (0.001)			- (0.004)
Margin management	+ (0.023)			+ (0.017)		+ (0.013)	+ (0.018)
Spatial structure					(0.003)	(0.002)	(< 0.001)

550

551 **B) Composition**

	BPN	BPS	MA	ZM	ZNE	ZSO	France
Temperature				(< 0.001)			(< 0.001)
Ph							(< 0.001)
Landscape diversity	0.040			0.020			(< 0.001)
Crop diversity							(< 0.001)
Herbicides							
Nitrogen dose							
Margin management			(< 0.001)	(< 0.001)			
Spatial structure	(< 0.001)		(< 0.001)	(< 0.001)	(0.040)	(< 0.001)	(< 0.001)

552

Appendix

Appendix A. Details about the sampling protocol and available data in the 500-ENI network.

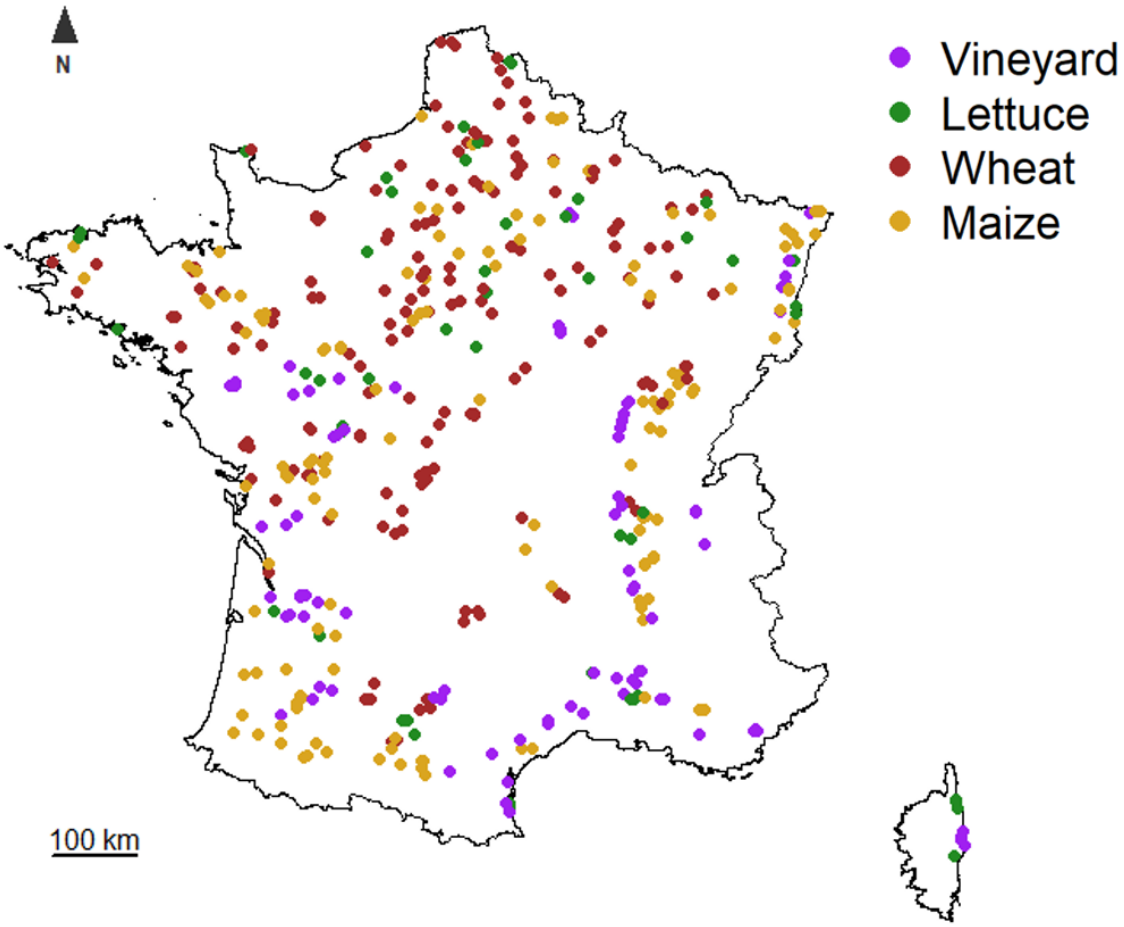


Fig. A.1. Distribution of fields monitored at least five years between 2013 and 2019 in continental France. Purple: vineyards (n = 93), green: market gardening crops (n = 50), brown: winter wheat in rotation (n = 178), yellow: maize in rotation (n = 141).

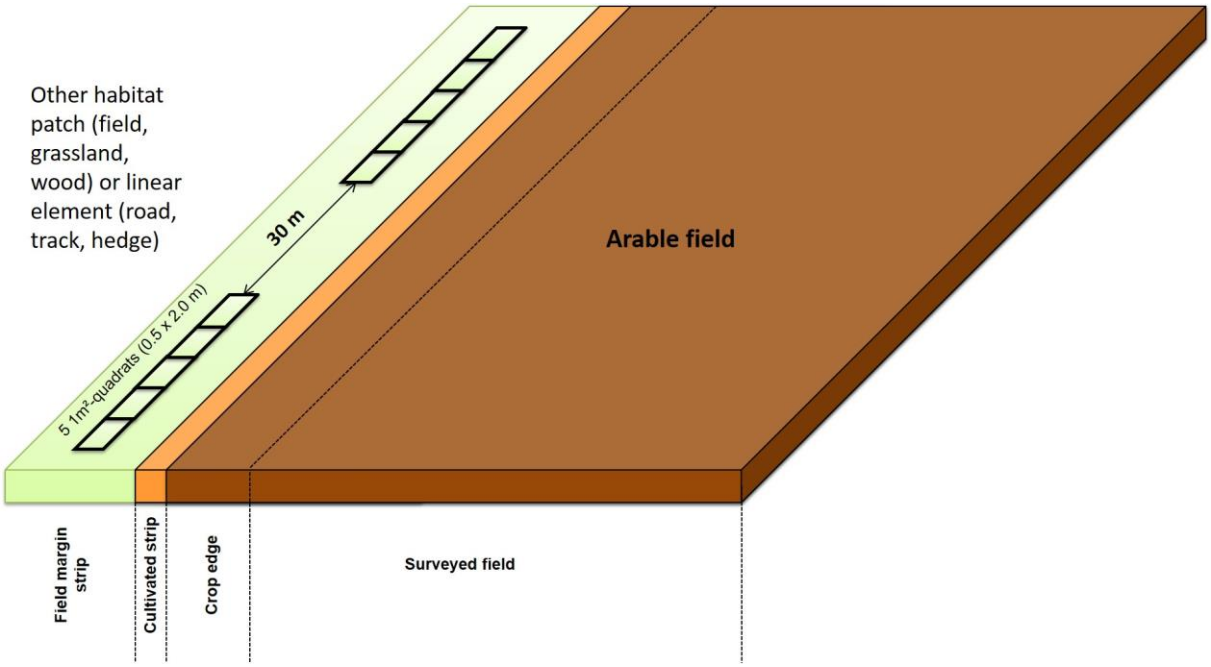


Fig. A.2. Details about the sampling protocol within each field margin (Andrade et al., 2021). The *crop edge* is the 1-6 first meters of the crop. The *cultivated strip* or *crop strip* is outside the last row of crops and is mostly composed of bare soil usually colonized by weed species from the field. The *field margin strip* is the uncultivated herbaceous strip between the *cultivated strip* and the *adjacent habitat*.

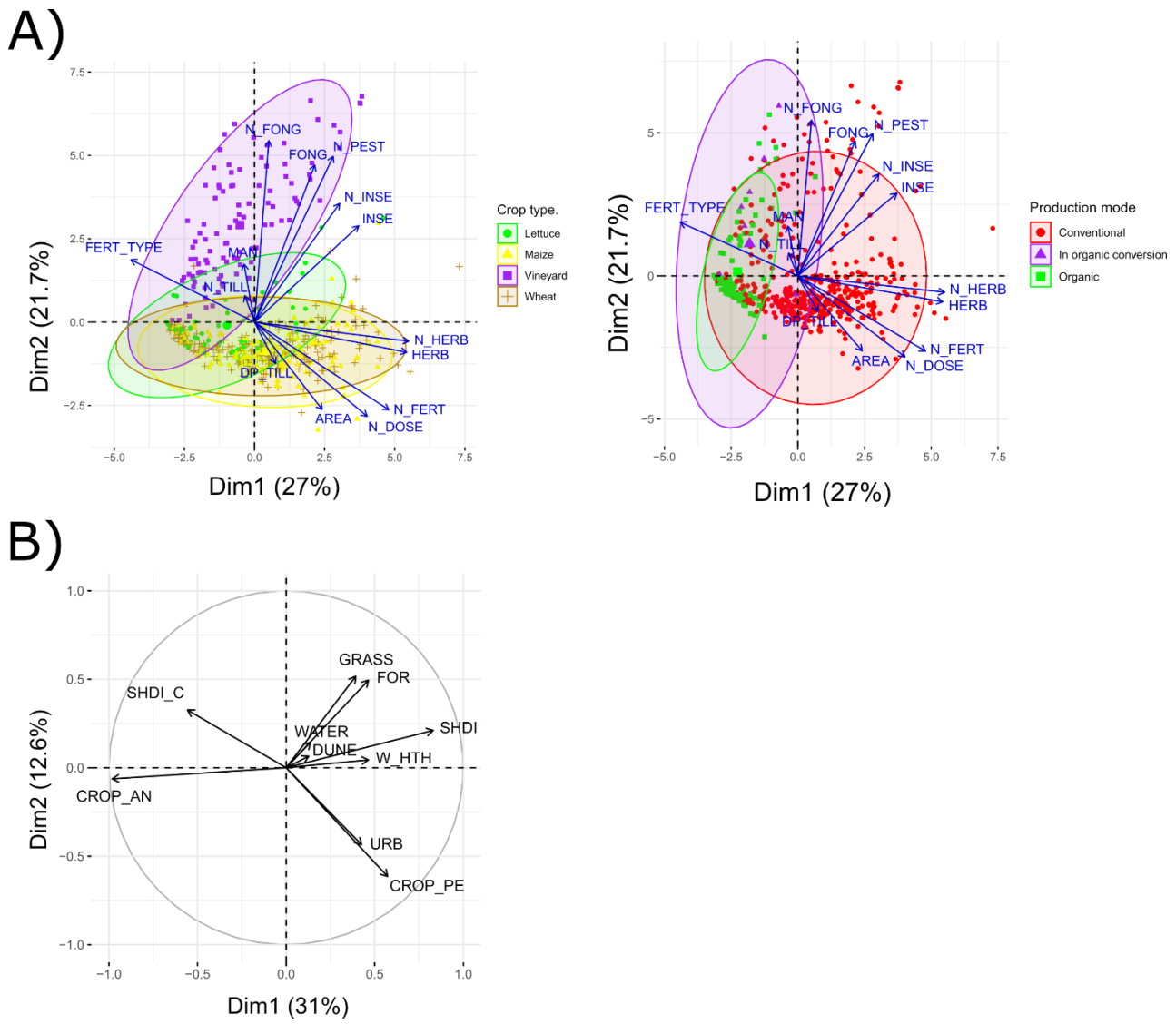


Fig. A.3. PCA on the main types of factors. Points stand for the 543 sites. A) Agricultural practices, B) Landscape. In A), agricultural practices are discriminated by crop type and production mode depicted by colored ellipses. For landscape, the SHDI was computed on 8 habitat classes (annual crops, perennial crops, forests, urban, grasslands, water, beaches and dunes, woody heaths) and the SHDI_C on 17 crop groups (maize, winter cereals, spring cereals, winter rapeseed, spring rapeseed, sunflower, beet, oilseed, grapevine, flax, arboriculture, fiber plants, vegetables or flowers, aromatic/medicinal/perfume plants, other crops, fallows and grasslands).

576 **Abbreviations for agricultural practices**

577 **AREA =** Field area

578 **DP_TILL =** Maximum depth of tillage

579 **FERT_TYPE =** Type of fertilization (mineral VS organic)

580 **FONG =** Fungicide treatment frequency index

581 **HERB =** Herbicide treatment frequency index

582 **INSE =** Insecticide treatment frequency index

583 **MAN =** Number of margin managements

584 **N_DOSE =** Nitrogen dose (fertilization and amendments)

585 **N_FERT =** Number of fertilizations and amendments

586 **N_FONG =** Number of fungicide applications

587 **N_HERB =** Number of herbicide applications

588 **N_INSE =** Number of insecticide applications

589 **N_PEST =** Number of pesticide applications

590 **N_TILL =** Number of tillage events

591

592 **Abbreviations for landscape**

593 **CROP_AN =** Annual crops

594 **CROP_PE =** Perennial crops (vineyards and orchards)

595 **DUNE =** Beaches and dunes

FOR = Forests

GRASS = Grasslands and lawns

SHDI = Shannon landscape diversity index

SHDI_C = Shannon crops diversity index

URB = Urban, roads, mineral surfaces

W_HTH = Woody heaths

WATER = Streams, rivers, lakes

Table A.1. Details about calculation of each explanatory factor included in analyses (see **Table 1**).

Factors	Calculation
Mean annual temperature	
Soil pH (water)	
Shannon's habitat diversity index	$-\sum_{i=1}^{n} p_i \ln p_i$ With $p_i$the proportion of the habitat i in the landscape
Shannon's crop diversity index	$-\sum_{i=1}^{n} p_i \ln p_i$ With $p_i$the proportion of crop i in the agricultural area
Dose of nitrogen (fertilization and amendments)	$\frac{Quantity \times N\ of\ the\ formulation}{100}$
Treatment Frequency Index of herbicides	$\frac{Applied\ dose\ of\ used\ products \times treated\ area}{certified\ dose \times plot\ area}$ For herbicides in vineyards, the treated area accounts for one third of the plot area if the treatment is done on the rows and two thirds if it is done on the inter-rows. In case of missing information, the treatment is supposed to involves only the rows (i.e. the most frequent practice).
Number of management events	All types of management (mowing, grazing…)
Spatial structure	Based on geographical coordinates, see **Appendix C** for more details

References

Andrade, C., Villers, A., Balent, G., Bar-Hen, A., Chadoeuf, J., Cylly, D., Cluzeau, D., Fried, G., Guillocheau, S., Pillon, O., Porcher, E., Tressou, J., Yamada, O., Lenne, N., Jullien, J., &

Monestiez, P. (2021). A real-world implementation of a nationwide, long-term monitoring program to assess the impact of agrochemicals and agricultural practices on biodiversity. *Ecology and Evolution*, *11*(9), 3771–3793. https://doi.org/10.1002/ece3.6459

613 **Appendix B.** Method of imputation for missing values.

614 We used the R package MICE (function of the same name) to impute the missing values from our dataset.
615 This method relates on multiple imputations by chained equations in which each incomplete variable
616 is imputed by a separate model. The algorithm imputes an incomplete variable using other variables
617 as predictors. Variables are imputed in a random order. If predictors are incomplete themselves, the
618 most recently generated imputations are used to complete the predictors prior to imputation of the
619 target variable. The number of iterations is the number of times each variable is computed (until
620 convergence). Given the large number of incomplete predictors in our dataset, we fixed this number
621 to 10 (instead of 5 by default). The function offers a wide range of available methods depending on
622 the nature of the data to be imputed. Here, we chose the Random forest algorithm, as a reliable
623 method to avoid over-fitting by ensemble learning (Shah et al., 2014). We imputed the missing data
624 100 times to control for the variability of imputations. The ultimate imputed value for each observation
625 is the one that occurs most frequently among the 100 imputations. To further improve the imputation,
626 we included some complete factors not used in the analyses (climate, soil, landscape, adjacent habitat,
627 width margin, field area, geographical coordinates, and all practices including the total number of
628 fertilizations, insecticide, herbicide and fungicide applications, tillage and management events one
629 year before the observation). The final dataset given in input included 44 variables and 13119
630 observations (observations from other groups than plants were added to improve the learning ability
631 of the algorithm).

632 The robustness of the method has been tested for 3013 observations. We ensured that the percentage
633 of missing values in each factor did not change from the full dataset. We checked the differences
634 between observed and imputed values. We saw that the error rate of a variable is partly correlated to
635 its percentage of missing values but never exceeds 10% of the total values imputed, and 2.2% of the
636 total dataset (3013 observations).

We also ensured that the imputations by observation did not change significantly the mean values by site. We focused on a subset of 58 sites for which we were able to compute the difference in percentage of the imputed value with the correct value (sites for which at least 5 years of data were available for all variables of interest). It results in 91% of the sites having an estimate of nitrogen dose within 10% of the correct value (100% of the sites for other variables). When differences were tested over the whole dataset (means before VS after imputation), it results in a change of less than 25% from the original value for 98% of the sites, except for nitrogen dose for which it was 76% of sites.

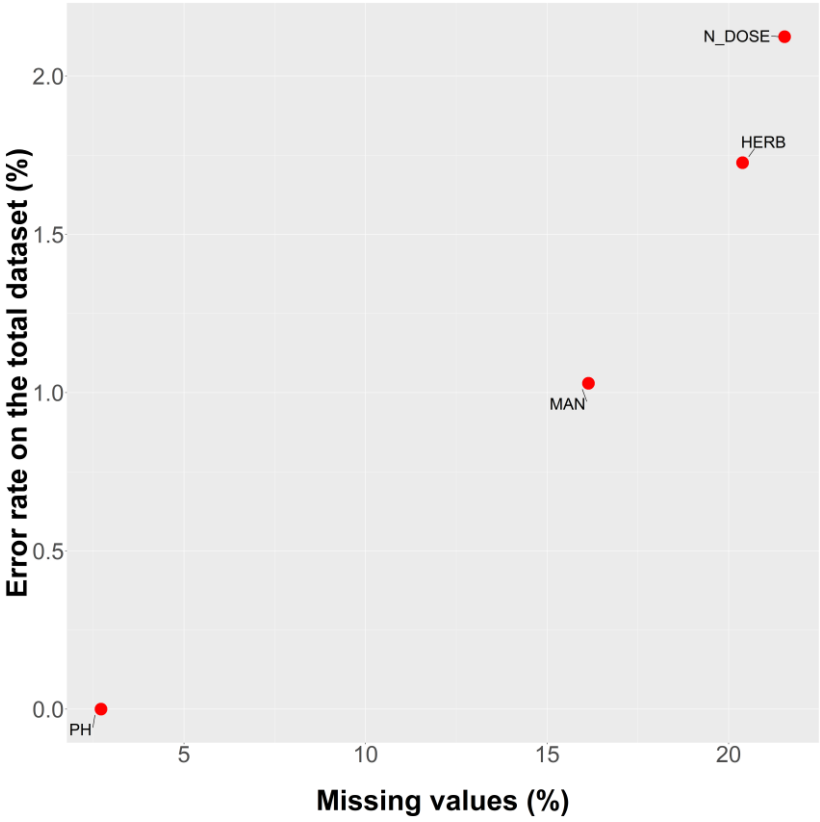


Fig. B.1. Error rate of imputed values estimated for the total used dataset according to the percentage of missing values of each variable.

References

Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of Random Forest and Parametric Imputation Models for Imputing Missing Data Using MICE: A CALIBER Study. *American Journal of Epidemiology*, *179*(6), 764–774. https://doi.org/10.1093/aje/kwt312

Appendix C. Spatial parameters for SAR and methodology for partial regression analyses.

Spatial correlograms were used to choose the maximal distance for the spatial neighborhood matrix at each resolution. We tested several distances around the visual estimate for each spatial resolution and selected the one that produced the lower AIC in a spatial regression with no other predictors (Kissling & Carl, 2008; **Table SC.1**). Similarly, we used AIC to select one of two weight functions (1/x and $1/x^2$) for the neighborhood matrix.

Table C.1. Maximum distances for spatial neighborhood matrix. The selected weight function was always 1/x.

	0	**25**	**40**	**60**	**75**
Maximum distance (in km)	25	25	40	80	100

We followed the methodology described in Borcard *et al.* (1992) to partition the R^2 of the model in several parts, each of them attributed to a single explanatory variable or to the spatial component itself (spatial autocorrelation). For GDM, we used the explained deviance, whereas we used the pseudo-R^2 of Nagelkerke for SAR. This allows us to tease out the effects of each factor and their interaction by building the following models :

1) M1 = model with all variables
2) M2 = model with all variables minus the variable of interest

The R^2 attributed solely to the variable of interest is the R^2 of M1 minus M2.

Then, we can tease out the spatial component from the effects of environmental predictors by building the following models :

1) M3 = model with all variables (with the spatial component)

2) M4 = model with all variables (without the spatial component)

In the case of SAR, the spatial component cannot be removed directly, so M4 was a linear model. The R^2 of Nagelkerke is equivalent to the classical R^2 of a linear regression (Nagelkerke, 1991), so we used this latter for M4. To test for the significance of spatial structure, we performed a Moran test on residuals of M4.

In the case of GDM, the spatial structure is directly included as a predictor by a matrix of geographic distances between sites/cells. Significance can be assessed by permutation tests as for other predictors.

Due to approximations in a few cases, the total R^2 of a complete model may appear to be slightly below the total of environmental and spatial parts of variance explained. In the opposite case (R^2 greater than the sum of variances explained by environment and spatial structure), it implies that there are some interaction parts among components, i.e. variance explained jointly by several factors (these parts are not shown here). Also, some fractions of variance can be negative in case of non-linear dependencies and are set to zero for the sake of interpretation.

References

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the Spatial Component of Ecological Variation. *Ecology*, *73*(3), 1045–1055. https://doi.org/10.2307/1940179

Kissling, W. D., & Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, *17*(1), 59–71. https://doi.org/10.1111/j.1466-8238.2007.00334.x

Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*(3), 691–692.

Appendix D. Details about the sampling at each spatial resolution and test of the robustness of our analyses to sampling biases related to the spatial resolution.

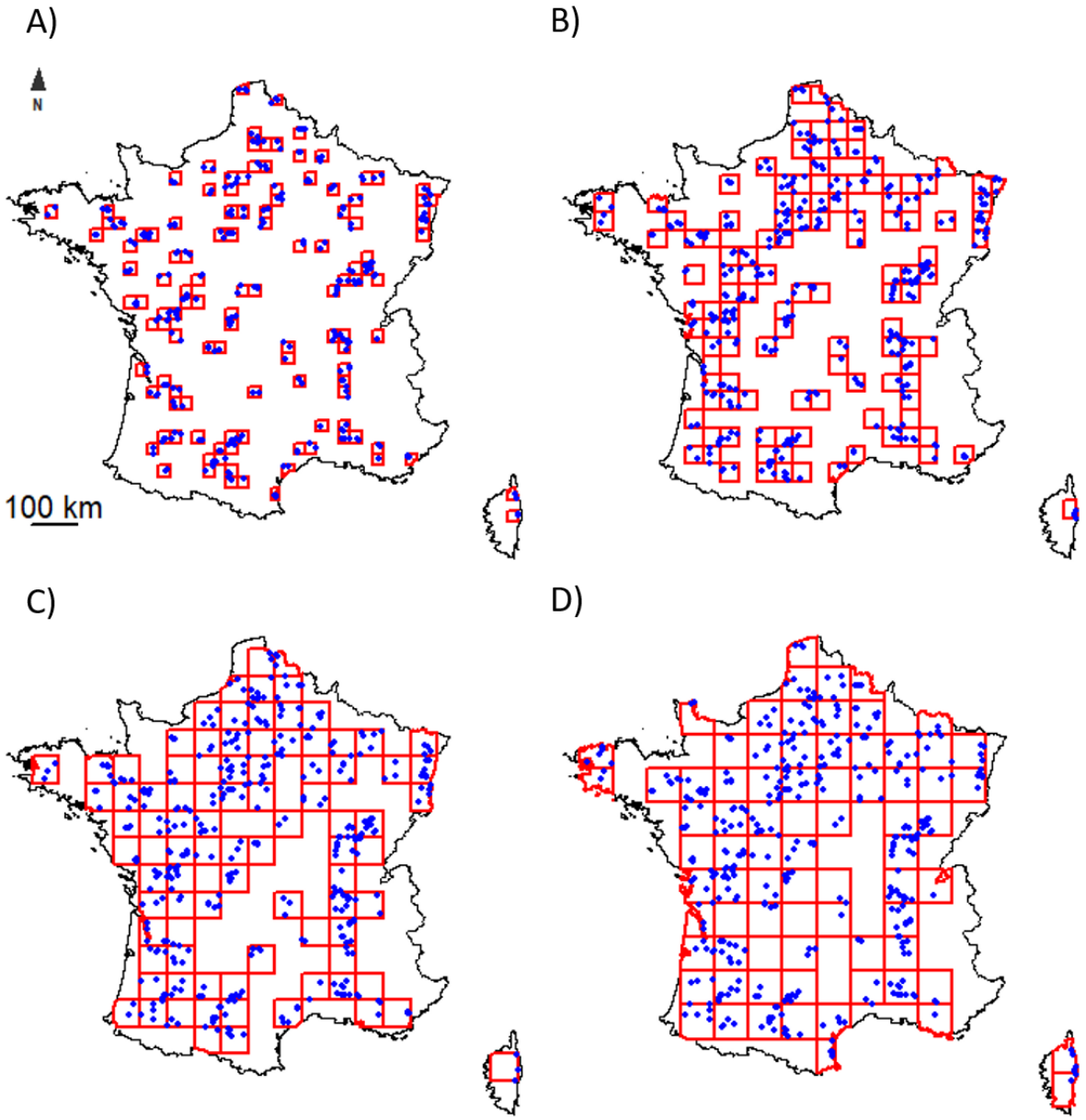


Fig. D.1. Distribution maps of the field margins (blue points) surveyed in grids (red) of different resolutions. A) 25 km, B) 40 km, C) 60 km and D) 75 km. The grid position has been shifted three times in analysis for each resolution. We shifted the initial grid position towards the northeast by half the length of a cell and towards the northwest by a quarter the length of a cell.

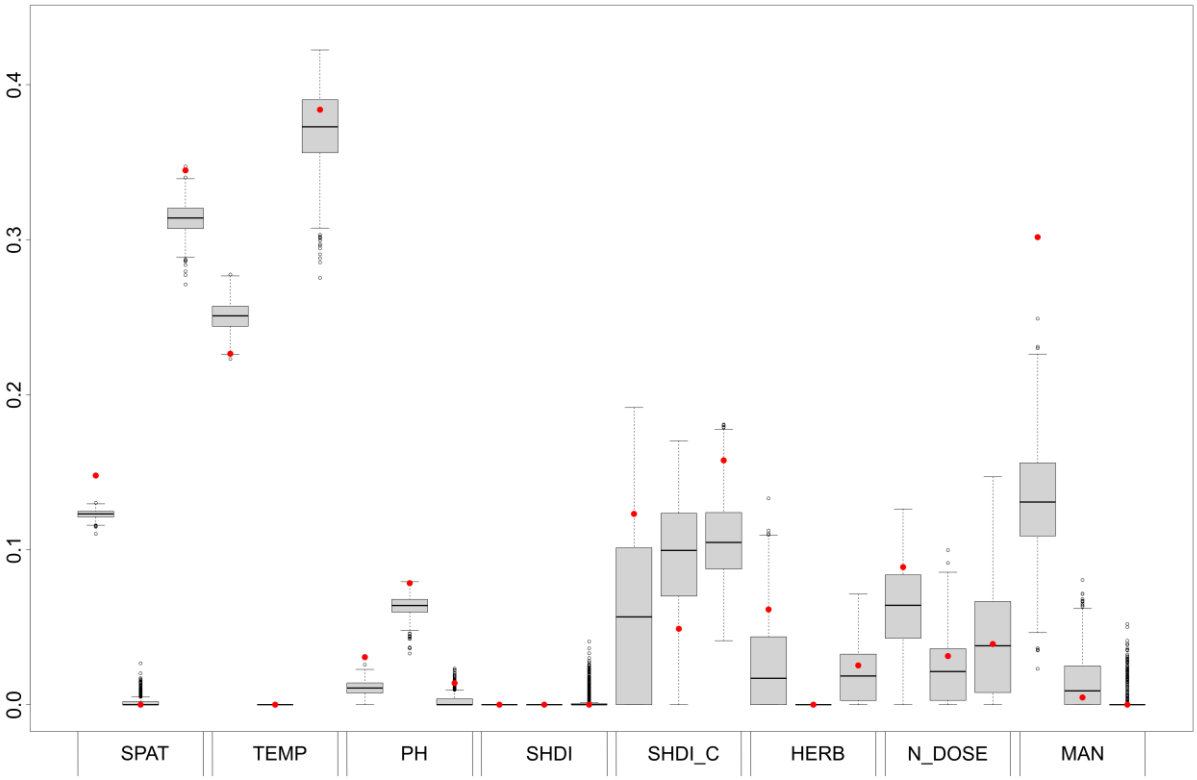
To help the reader understand our selection of resolutions, we clarify the factors that influenced our decision. First, we set the minimum and maximum resolutions at the site-level and 75 km, respectively, as larger grid cells would have resulted in too few cells for analysis. A Mantel correlogram on composition and a variogram on species richness (at the site-resolution) revealed that spatial autocorrelation was significant up to about 40-50 km, so we found it useful to include some resolutions below and above this threshold, while trying to keep resolution intervals more or less equally spaced. We wanted to include at least a couple of intermediate resolutions so that we made sure that the results were a logic progression between both ends, and not just some random variations due to a sampling / aggregation artifact. This is why we ended up with the 5 following resolutions: plot-level, 25, 40, 60 and 75 km. For simplicity, and because the 25 and 60 km resolutions only displayed minor differences with resolutions that preceded and followed them, we decided to present only the site-level, 40 km, and 75 km resolutions in the text.

Different biases can result from this analysis method built around the influence of spatial resolution. Indeed, the number of sites differs between cells and are not uniformly distributed in the cell. This could potentially lead to a lack of precision in cells containing fewer sites or sites that are not evenly distributed spatially. To test for this bias, we randomly drew 2 sites in each cell of the 25 km resolution, before running analyses and repeated the procedure 1000 times. We focused on the 25 km resolution as it is the most concerned by the low number of sites by cell (**Table D.1**). For each type of model (GDM and SAR), we plotted the distribution of parameters over these 1000 runs and compared it to the observed values. We can see that the bootstrap procedure induces little variation of the model parameters and that it is particularly true for SAR models (**Fig. D.2**).

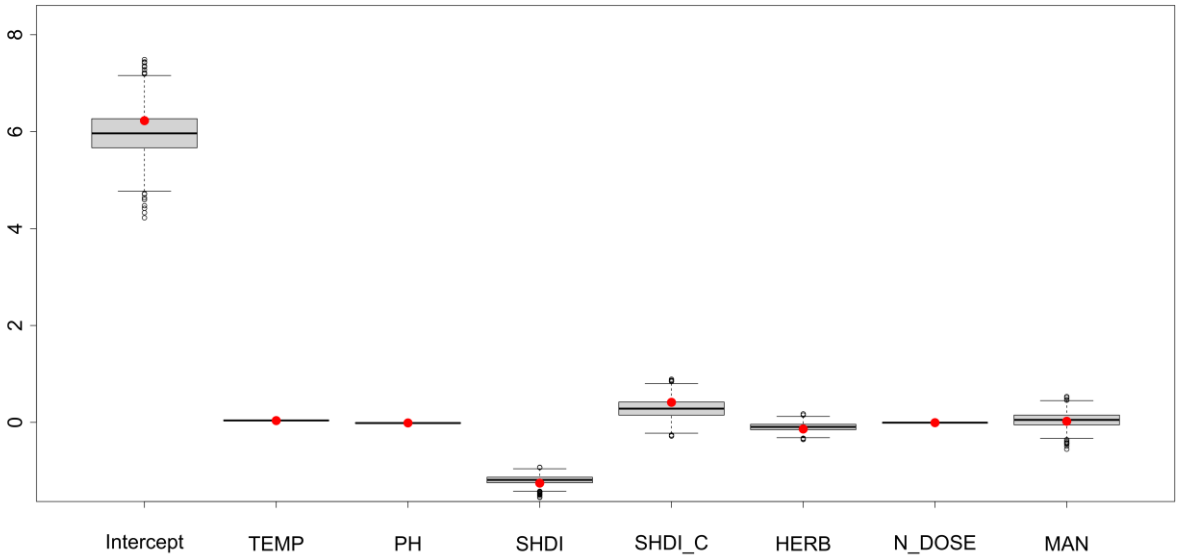
Table D.1. Sample size for each resolution. Grid positions in the aggregated resolutions were shifted to obtain three different positions of the grid system. Means and standard deviations below reflect variations across these grids within each resolution.

Spatial resolution	Site-level	25 km	40 km	60 km	75 km
Mean number of sites by cell		2.7 ± 0,1	3.3 ± 0,1	4.6 ± 0,1	5.8 ± 0,1
Mean number of sites	462	306 ± 10	387 ± 9	418 ± 12	428 ± 1
Mean number of cells		87 ± 7	105 ± 3	86 ± 2	71 ± 2

A) Composition (GDM)



742 **B) Species richness (SAR)**



743

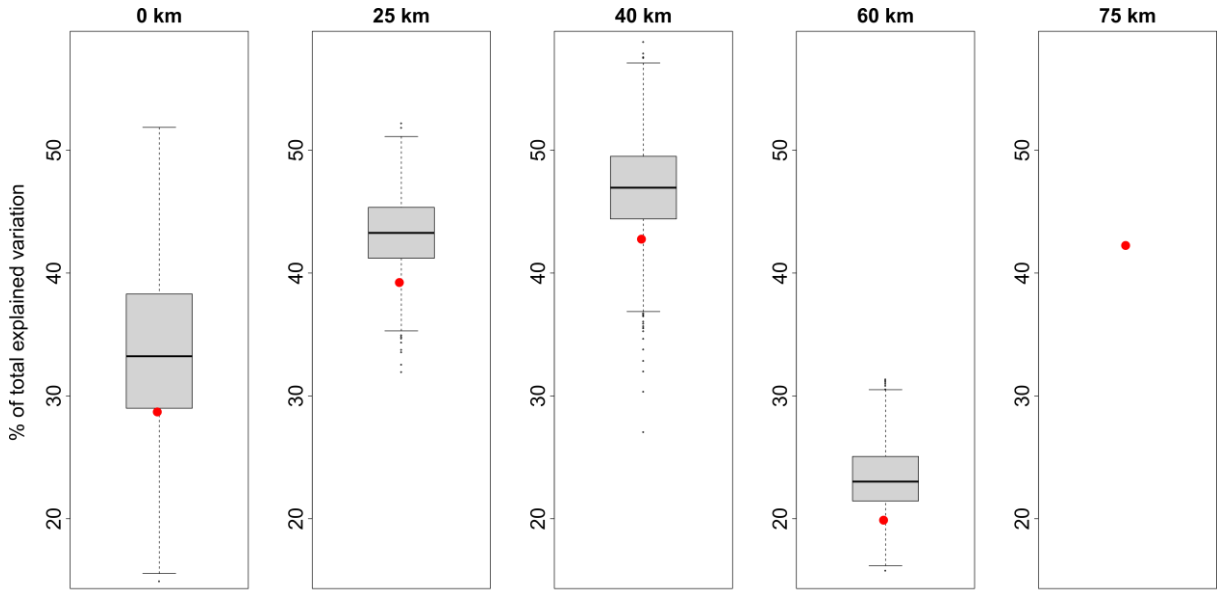
744 **Fig. D.2.** Variation of the A) GDM and B) SAR parameters over the 1000 resamplings. The red points
745 are the observed coefficients. See **Table 1** for coefficient abbreviations. A) For GDM, each explanatory
746 factor has three parameters, corresponding to three I-spline basis functions (see Ferrier et al., 2007).

747

748 A second bias that we wanted to test is the influence of the number of cells on the coefficients and R^2/
749 explained deviance of the models. Indeed, the R^2 of a model can be artefactually inflated when the
750 sampling power is too low, which is a problem when we aim to compare spatial resolutions with
751 different number of spatial units. Consequently, we randomly draw 71 cells at each resolution (except
752 for the coarsest one) as it is the mean number of cells for the resolution with the least one (75 km
753 resolution). We performed analyses for the resolutions 0, 25, 40 and 60 km and repeated the
754 procedure 1000 times. For each type of model (GDM and SAR), we plotted the distribution of the global
755 R^2 / explained deviance over these 1000 runs and compared it to the observed values. We can see that
756 the bootstrap procedure induces little variation of the total explained variance with a slight bias
757 towards overestimation (**Fig. D.3**). However, the relative differences in explained variance between

spatial resolutions are preserved with these resamplings, which comforts us in the fact that we can reliably compare spatial resolutions despite a different number of samples.

A) Composition (GDM)



B) Species richness (SAR)

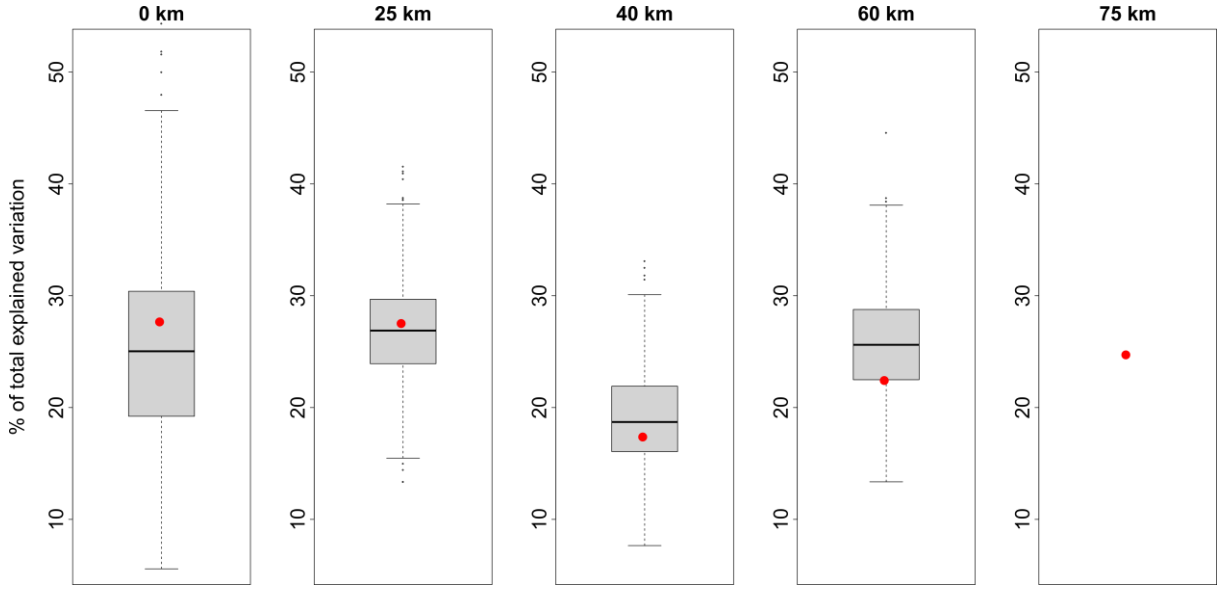
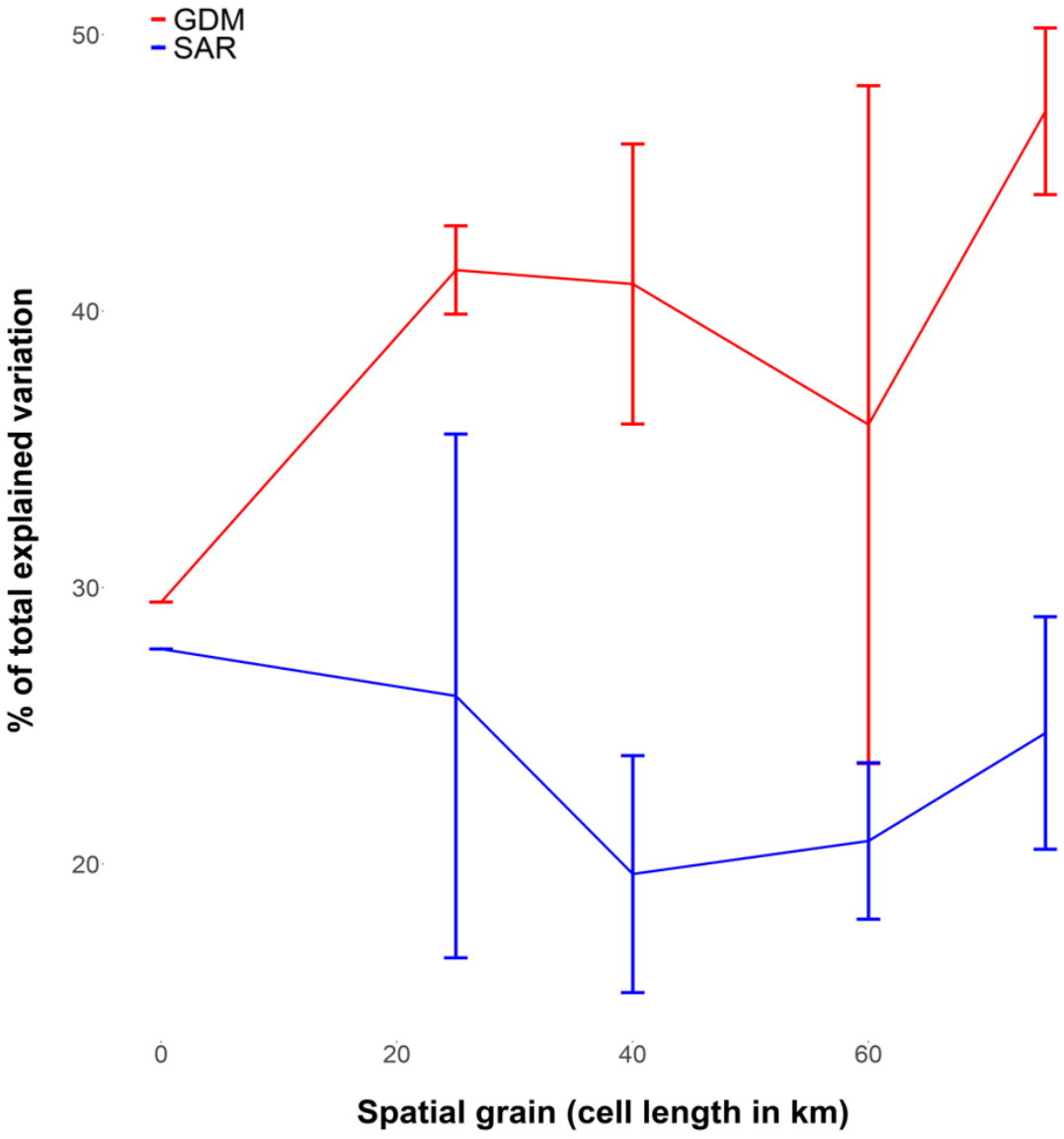


Fig. D.3. Variation of the A) GDM and B) SAR explained variance over the 1000 resamplings. The red points are the observed R^2 / explained deviance.

Another potential shortcoming was that the means computed within aggregated grid cells might not fully represent the regional pool or agricultural region. Indeed, the monitoring sites were stratified to represent the main crops and practices at the national level and not at the landscape or regional level. Checking the representativeness of the sites with the regional context is thus key to assess the relevance of the results. Unfortunately, we were unable to perform this validation with the available dataset.

775 **Appendix E**. Detailed outputs of GDM and SAR.

776



777

778 **Fig. E.1.** Total percentage of explained variance according to the spatial resolution of analyses (for

779 GDM and SAR). Standard deviations were computed from the three resamplings of grids.

780

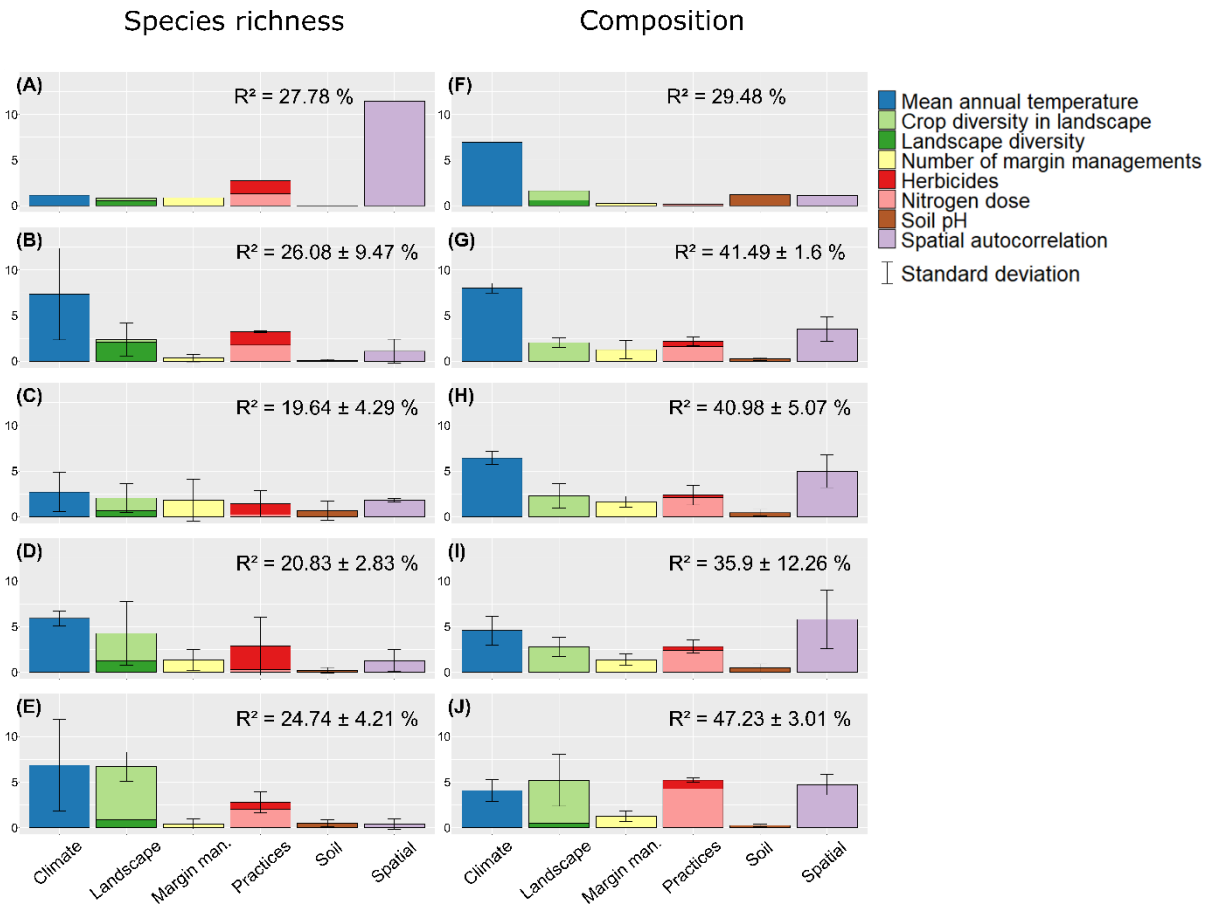


Fig. E.2. Percentage of explained variance of species richness and composition by each factor for each spatial resolution. (A, F) 0 km; (B, G) 25 km; (C, H) 40 km; (D, I) 60 km; (E,J) 75 km. Standard deviations are computed from the three resamplings of grids.

Table E.1. Significance of each predictor (in row) for each spatial resolution (in column, in km). For species richness in A), the direction of the relationship is indicated by +/-. The number of +/-/* indicates the number of times the predictor was significant over the three grid shifts. We only report significant relationships with p-values in brackets (the average p-value when significant in multiple grids).

A) Species richness

	0	**25**	**40**	**60**	**75**
Temperature	+ (0.004)	++ (0.002)	+ (0.010)	+++ (0.012)	++ (0.011)
Ph					
Landscape diversity		- (0.041)			
Crop diversity				++ (0.027)	+++ (0.017)
Herbicides	- (0.002)	- (0.048)	- (0.031)	- (0.006)	
Nitrogen dose	- (0.004)				- (0.044)
Margin management	+ (0.018)		+ (0.019)		
Spatial structure	(< 0.001)				

B) Composition

	0	**25**	**40**	**60**	**75**
Temperature	(< 0.001)	*** (< 0.001)	*** (< 0.001)	*** (0.013)	*** (0.020)
Ph	(< 0.001)				
Landscape diversity	(< 0.001)				
Crop diversity	(< 0.001)	** (< 0.001)	** (< 0.001)	** (0.010)	*** (0.020)
Herbicides					
Nitrogen dose		*** (0.030)	** (< 0.001)	** (0.020)	*** (0.010)
Margin management		* (0.020)	* (< 0.001)		
Spatial structure	(< 0.001)	*** (< 0.001)	*** (< 0.001)	*** (< 0.001)	*** (< 0.001)

Table E.2. Mean percentages ± standard deviations of explained variance of A) species richness and B) composition by each factor (in row) for each spatial resolution (in column, in km). Means and standard deviations are computed from the three grids. We included the portion of variance that results from the combined influence of multiple factors, which is commonly referred to as "Interactions."

A) Species richness

	0	**25**	**40**	**60**	**75**
Temperature	1.15	7.35 ± 5.02	2.71 ± 2.15	5.90 ± 0.82	6.85 ± 5.01
Ph	0	0.09 ± 0.12	0.66 ± 1.05	0.21 ± 0.3	0.51 ± 0.38
Landscape diversity	0.53	2.08 ± 1.42	0.67 ± 0.65	1.21 ± 0.98	0.86 ± 0.95
Crop diversity	0.31	0.30 ± 0.44	1.37 ± 0.97	3.05 ± 2.65	5.83 ± 1.14
Herbicides	1.47	1.47 ± 1.46	1.21 ± 1.53	2.56 ± 3.29	0.80 ± 0.81
Nitrogen dose	1.28	1.76 ± 1.53	0.23 ± 0.16	0.29 ± 0.3	1.99 ± 1.95
Margin management	0.87	0.33 ± 0.41	1.83 ± 2.3	1.32 ± 1.16	0.42 ± 0.55
Spatial structure	11.43	1.09 ± 1.32	1.8 ± 0.19	1.28 ± 1.22	0.38 ± 0.59
Interactions	10.74	11.61 ± 3.05	9.16 ± 6.49	5.02 ± 4.49	7.11 ± 6.93
R^2	27.78	26.08 ± 9.47	19.64 ± 4.29	20.83 ± 2.83	24.74 ± 4.21

B) Composition

	0	**25**	**40**	**60**	**75**
Temperature	6.93	8.01 ± 0.53	6.44 ± 0.7	4.56 ± 1.59	4.08 ± 1.22
Ph	1.23	0.25 ± 0.15	0.43 ± 0.38	0.46 ± 0.45	0.25 ± 0.15
Landscape diversity	0.59	0 ± 0	0.01 ± 0.02	0.04 ± 0.04	0.51 ± 0.59
Crop diversity	1.05	2.04 ± 0.52	2.29 ± 1.37	2.73 ± 1.06	4.69 ± 2.27
Herbicides	0.04	0.55 ± 0.56	0.29 ± 0.26	0.40 ± 0.29	0.95 ± 0.54
Nitrogen dose	0.14	1.63 ± 0.30	2.07 ± 0.91	2.42 ± 0.44	4.28 ± 0.79
Margin management	0.22	1.26 ± 1.01	1.65 ± 0.59	1.37 ± 0.64	1.28 ± 0.57
Spatial structure	1.09	3.55 ± 1.34	4.98 ± 1.79	5.79 ± 3.22	4.73 ± 1.13

Interactions	18.2	24.2 ± 3.31	22.81 ± 4.47	18.13 ± 6.26	26.48 ± 6.92
Explained deviance	29.48	41.49 ± 1.6	40.98 ± 5.07	35.9 ± 12.26	47.23 ± 3.01

Table. E.3. Percentages of explained variance of A) species richness and B) composition by each factor (in row) for each biogeographic region and for national extent (in column, see **Fig. 1** for abbreviations of region names). We included the portion of variance that results from the combined influence of multiple factors, which is commonly referred to as "Interactions."

A) Species richness

	BPN	**BPS**	**MA**	**ZM**	**ZNE**	**ZSO**	**France**
Temperature	0.70	1.13	0.18	0	2.74	6.25	1.15
Ph	0.86	2.44	0	0	2.01	3.41	0
Landscape diversity	6.49	1.23	0	2.95	3.13	0.49	0.53
Crop diversity	0.96	1.55	1.53	1.03	2.14	1.35	0.31
Herbicides	5.95	1.27	0	0	3.54	1.20	1.47
Nitrogen dose	0.64	2.79	0.02	12.88	3.02	1.65	1.28
Margin management	6.58	1.25	0	7.07	2.07	4.46	0.87
Spatial structure	0.64	2.58	1.65	0.02	10.36	11.4	11.43
Interactions	1.00	8.32	11.67	14.84	0	2.86	10.74
R^2	23.82	22.56	15.05	38.79	26.81	33.07	27.78

B) Composition

	BPN	**BPS**	**MA**	**ZM**	**ZNE**	**ZSO**	**France**
Temperature	0.79	1.36	0.13	6.19	0.88	0.74	6.93
Ph	0.26	1.90	0.21	0.84	0.08	0.41	1.23
Landscape diversity	1.86	0.02	2.17	2.23	0.75	0.07	0.59
Crop diversity	1.05	0	0.79	1.65	1.04	1.04	1.05
Herbicides	1.12	0.08	0.81	0.01	1.54	0	0.04

Nitrogen dose	0	1.37	0.11	1.11	0.72	0	0.14
Margin management	0.62	0.17	6.87	5.58	0	0.98	0.22
Spatial structure	7.83	4.83	2.32	7.30	3.30	9.41	1.09
Interactions	9.53	3.51	4.64	19.67	2.29	4.00	18.19
Explained deviance	23.06	13.24	18.05	44.58	10.60	16.65	29.48

Appendix F. Range of values of explanatory variables within spatial resolutions and biogeographic regions.

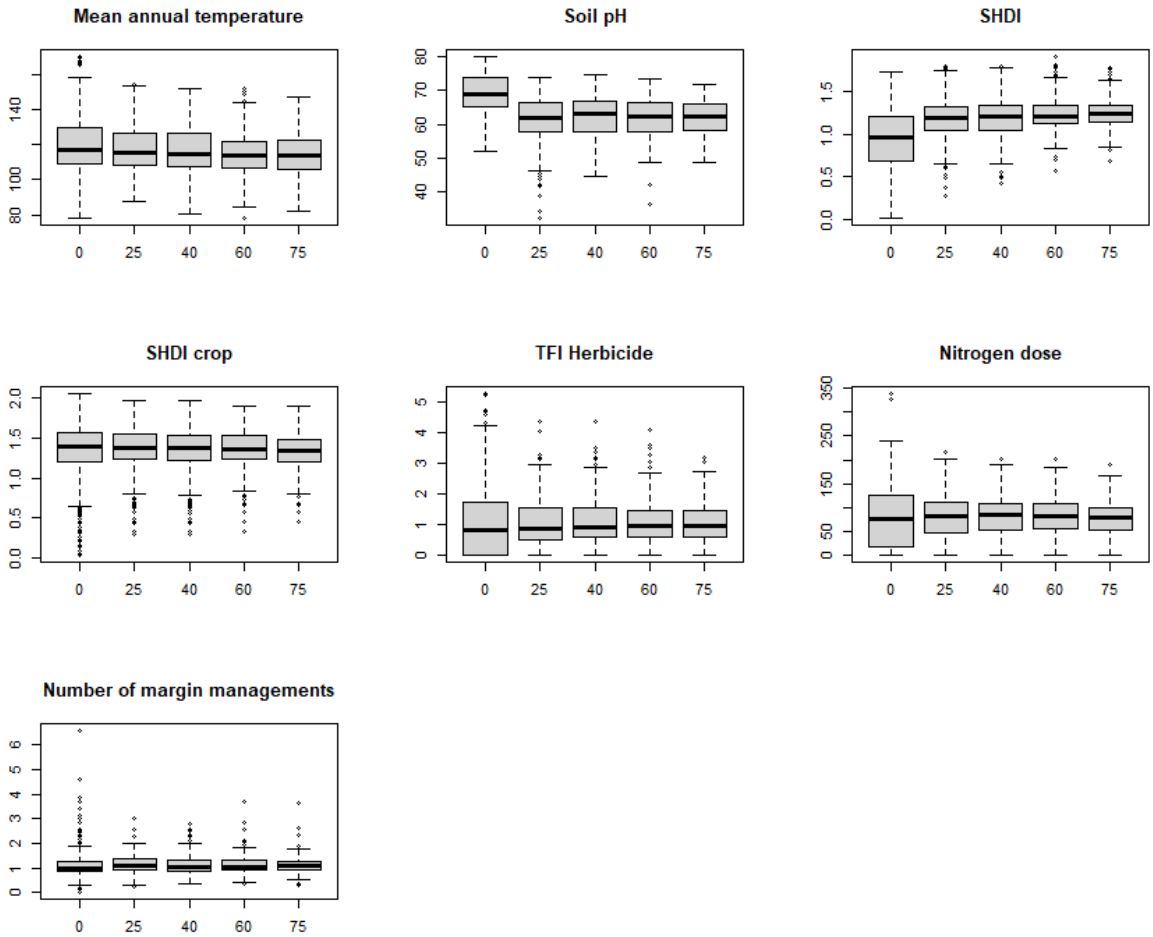


Fig. F.1. Boxplot of environmental predictors by spatial resolution (in km). See **Table 1** for units.

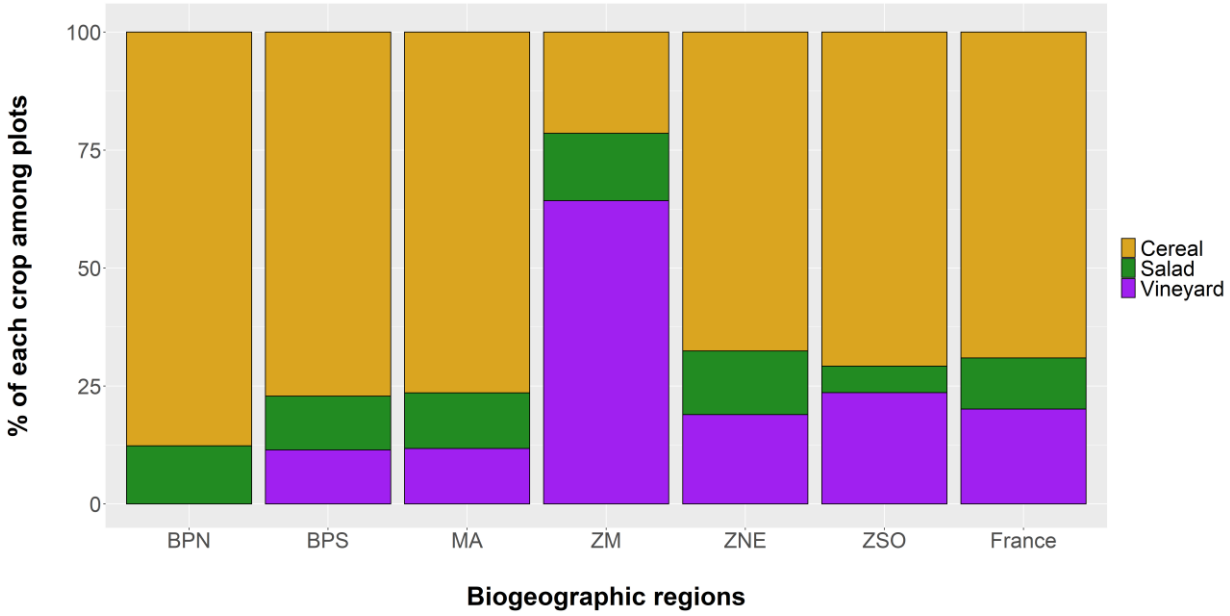
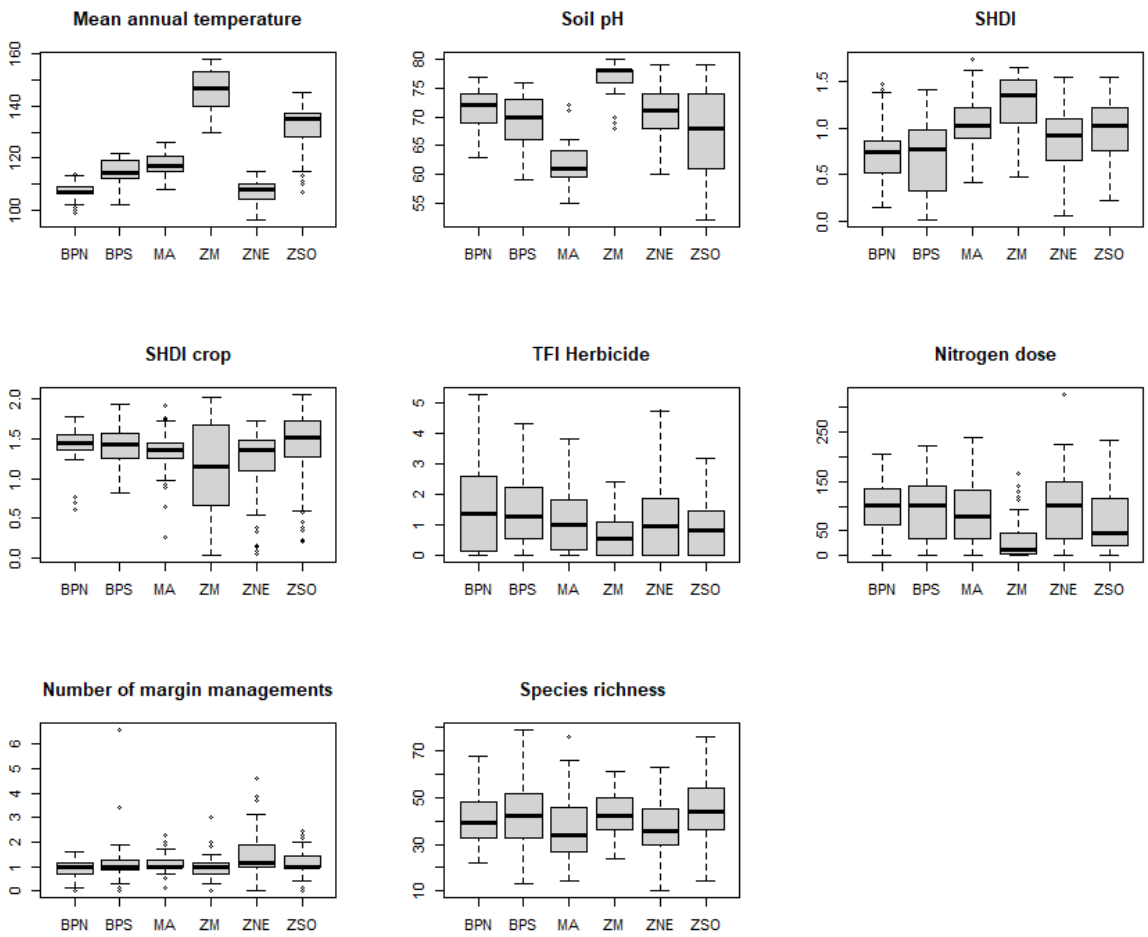


Fig. F.2. Percentages of each crop type among sites surveyed within each biogeographic region. See **Fig. 1** for abbreviations of region names.



825

826 **Fig. F.3.** Boxplot of environmental predictors by biogeographic region. See **Fig. 1** for abbreviations of

827 region names and **Table 1** for units.