



HAL
open science

Identifying pesticide mixtures at country-wide scale

Milena Cairo, Anne-Christine Monnet, Stephane S. Robin, Emmanuelle Porcher, Colin Fontaine

► **To cite this version:**

Milena Cairo, Anne-Christine Monnet, Stephane S. Robin, Emmanuelle Porcher, Colin Fontaine. Identifying pesticide mixtures at country-wide scale. 2022. hal-03815557v1

HAL Id: hal-03815557

<https://hal.science/hal-03815557v1>

Preprint submitted on 14 Oct 2022 (v1), last revised 27 Mar 2023 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Identifying pesticide cocktails at country-wide scale

Milena CAIRO¹, Anne-Christine MONNET¹, Stéphane ROBIN^{1,2}, Emmanuelle PORCHER¹, Colin FONTAINE¹

¹ Centre d'Écologie et des Sciences de la Conservation (CESCO), Muséum national d'Histoire naturelle, Centre National de la Recherche Scientifique, Sorbonne Université, CP 135, 57 rue Cuvier 75005 Paris, France

² Sorbonne Université, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, F-75005 Paris, France

Corresponding author: Milena Cairo, milena.cairo1@mnhn.fr, Centre d'Écologie et des Sciences de la Conservation (CESCO), Muséum national d'Histoire naturelle, CP 135, 57 rue Cuvier 75005 Paris, France

ABSTRACT

1 Wild organisms are exposed to complex cocktails of pesticides owing to the
2 considerable diversity of substances and agricultural practices. The study of pesticide
3 cocktails is essential because of potentially strong synergistic effects, making cocktails
4 effects not predictable from the effects of single compounds. In addition, little is known
5 about the exposure of organisms to pesticide cocktails *in natura*.

6 We aimed to identify the number and composition of pesticide cocktails potentially
7 occurring in French farmland, using a database of pesticide purchases listed at the
8 postcode level. We developed a statistical method based on a mixture model to cluster
9 postcodes according to the identity, purchase probability and quantity of 279 active
10 substances.

11 We found that the 5,631 French postcodes can be clustered into 18 postcode
12 groups characterized by a specific pattern of pesticide purchases, that is a particular
13 pesticide cocktail. Substances defining cocktails can be sorted into “core” substances
14 highly probable in most postcode groups and “discriminating” substances are specific
15 to and highly probable in some postcode groups only, thus playing a key role in the
16 identity of pesticide cocktails. We found 13 core substances: two insecticides
17 (deltamethrin and lambda-cyhalothrin), seven herbicides (glyphosate, diflufenican,
18 fluroxypyr, MCPA, 2,4-d, triclopyr, pendimethalin) and four fungicides (fludioxonil,
19 tebuconazole, difenoconazole, thiram). The number of discriminating substances per
20 postcodes group ranged from 2 to 80. These differences in substance purchases
21 seemed related to differences in crop composition but also potentially to regional
22 effects.

23 Overall, our analyses return (1) sets of molecules that are likely to be part of the
24 same pesticide cocktails, for which synergetic effects should be investigated further
25 and (2) areas within which biodiversity might be exposed to similar cocktail
26 composition. This information will hopefully be of interest for future ecotoxicological
27 studies to characterise the actual impact of pesticide cocktails on biodiversity in the
28 field.

29 **Keywords:** Active Substances, Cluster, mixture model, expectation-maximization
30 algorithm, risk assessment

INTRODUCTION

31 Since the mid-20th century, pesticides have become of common use in agriculture
32 and their effects on both the environment and human health are now a concern. For
33 example, systemic pesticides are known to affect a broad range of organisms, from
34 invertebrates, both terrestrial and aquatic, to amphibians or birds (Humann-Guillemint
35 et al., 2019; Mahmood et al., 2016; Yang et al., 2008), thereby questioning the
36 sustainability of agroecosystem functioning and related services (Deguines et al.,
37 2014; Dudley et al., 2017; Furlan et al., 2018; Geiger et al., 2010). Pesticides are also
38 identified as a concern for human health, with numerous pesticide poisonings reported
39 across developing countries (Boedeker et al., 2020) and recent evidence of
40 relationships between diseases such as Parkinson's or cancers and exposure to
41 organophosphate insecticides (Sheahan et al., 2017; Tassin de Montaigu and
42 Goulson, 2020).

43 Pesticides effects on biodiversity are usually demonstrated with a focus on a
44 single substance or a limited set of substances in general (e.g. thiamethoxam,
45 clothianidin, imidacloprid, thiacloprid or glyphosate; (Botías et al., 2015; Busse et al.,
46 2001; Rundlöf et al., 2015; Van Bruggen et al., 2018). Yet, wild organisms are exposed
47 to complex cocktails (Dudley et al., 2017), owing to the diversity of substances
48 available and used in farmlands. Hence, studying substance cocktails is considered a
49 central task for environmental risk assessment (Lydy et al., 2004a), notably because
50 the effects of pesticide cocktails can strongly exceed the additive effects of single
51 compounds (Bopp et al., 2016; Junghans et al., 2006). Laboratory experiments
52 demonstrate synergetic interactions among substances within cocktails, affecting the
53 effect of the cocktail in non-additive ways (Cedergreen, 2014; Hernández et al., 2017;
54 Heys et al., 2016). While the importance of studying the effects of cocktails beyond
55 those of single substances was highlighted as soon as the late sixties (Keplinger and
56 Deichmann, 1967), and their evaluation is mandatory in the European Union since
57 2009 (EC No 1107/2009), few attempts to do so exist outside laboratories (Gibbons et
58 al., 2015).

59 Studies examining the effects of substance cocktails use two approaches:
60 bottom-up or top-down (Altenburger et al., 2013; Hernández et al., 2017; Relyea,

61 2009). The bottom-up approach aims at testing all possible cocktail compositions,
62 starting from pairs of substances to more complex combinations. This method makes
63 it challenging to consider more than a handful of substances. For example, ten
64 substances represent 45 possible pairs and over a thousand possible combinations of
65 three or more substances (Lydy et al., 2004a). Moreover, such approach might be
66 more suited to experiments in controlled rather than natural environments, as the latter
67 are recognized as strongly contaminated (Tang et al., 2021), making the control of
68 cocktail composition difficult. The top-down approach proposes to compare the effect
69 of cocktails, starting from potentially frequent cocktails including a high number of
70 substances but at the cost of not testing all combinations. In addition, the few existing
71 field studies generally focused on the effects of pesticide cocktails composed of a
72 restricted number of substances, on specific crops or on restricted spatial extent,
73 thereby limiting a broad understanding of cocktail effects. (e.g. Brittain et al., 2010;
74 Hallmann et al., 2014; Millot et al., 2017, but see Schreiner et al., 2016 & (Fritsch et
75 al., 2022). The top-down approach makes it critical to identify relevant cocktail
76 compositions, i.e. those actually occurring in the fields. The number of actual cocktails
77 encountered in agroecosystems should be much lower than the number of possible
78 combinations of substances because each substance is intended for a limited set of
79 crops only and because agricultural production is regionally specialised on particular
80 crops. Such regional specialisation implies that existing cocktails are likely to be
81 spatially structured. However, we still miss an overall picture of the cocktail composition
82 and spatial structure over large spatial extents.

83 Here, we introduce a new statistical method to identify relevant pesticide
84 cocktails, i.e. actual combinations of substances potentially co-occurring in
85 agroecosystems across Metropolitan France. We overcame the general problem of
86 limited availability of data on temporal and spatial use of pesticides (Navarro et al.,
87 2021) by taking advantage of the recent publication of an up-to-date database on
88 pesticide purchase in France, the French national bank of pesticide sales database.
89 This database has registered mandatory declarations of quantities of active substance
90 purchased in France since 2013 (law n°2006-1772) at a relatively fine spatial grain
91 (postcode of the buyer). France is also the seventh largest users of pesticides in the
92 world (FAO 2020) and has a wide range of agricultural types (Urruty et al., 2016), which
93 makes it a well-suited case country to try and identify pesticide cocktails encountered

94 in the field by wild organisms, as well as their spatial variation. Applying an
95 Expectation/Maximization algorithm to a mixture model, we obtained a clustering of
96 French postcodes on the basis of the composition of active substances purchased. We
97 show that the number of clusters, i.e. groups of postcodes with a similar composition
98 of purchased pesticides, is reasonably low and that this clustering is spatially coherent
99 and related to crop planting, as expected with regional specialisation. We show how
100 such clustering can be used to identify potentially important pesticide substances and
101 cocktails deserving further investigation.

METHODS

102 **1.1 Pesticide data**

103 Data on active substances were obtained from the French national bank of
104 pesticide sales (BNV-d; <https://bnvd.ineris.fr>). The BNV-d database registers
105 mandatory declarations of active substances sold in France. For each active substance
106 sale, the seller indicates the amount and the postcode of the buyer in the database.
107 This database thus indicates the quantity of active substances purchased at the spatial
108 resolution of the postcode of the buyer. Substances are identified with their generic
109 name and a unique identifier, the Chemical Abstracts Service number. We modified
110 generic names when synonyms were found. We only retained substances with a
111 license fee (i.e. under compulsory declaration) because we can expect thorough
112 reporting for these.

113 The years registered in the database ranged from 2013 to 2020. We discarded
114 the year 2013 because of incomplete data during the first reporting year, and the two
115 last years of the time series (2019 and 2020) because additions and changes in the
116 database are allowed for two years after a declaration. Also, note that the legislation
117 has kept changing until 2016, with consequences for the mandatory nature of
118 declaration for some substances or treatments. In particular, until 2016 the
119 geographical information associated with seed coating substances was that of the
120 seed coating company, not of the buyer. Hence, 2017 can be considered the most
121 accurate and thorough year within the period 2013-2020.

122 The data provides the total mass (in g) bought per substance with mandatory
123 declaration, of which in 2017 there were 279. We studied the data at the postcode
124 scale, assuming that substances purchased in a given postcode would be used within

125 the same postcode. In metropolitan France, postcode areas range from 0.17 km² to
126 614.39 km² (median = 62.79 km², Q1 = 19.59 km², Q3=140.36 km²). Using specific
127 postcodes (CEDEX) that enable the identification of private companies, we discarded
128 the data related to the national railroad company (SNCF): SNCF is a major buyer with
129 central purchasing bodies that do not use the substances within the postcode of
130 purchase. We converted all remaining CEDEX codes to their corresponding regular
131 postcode. We were thus left with 5,631 postcodes with information about the quantities
132 (in g) of 279 active substances purchased in 2017. We classified these substances
133 into fungicides, herbicides, insecticides following the Pesticide Properties Data Base
134 (PPDB) (Lewis et al., 2016) and the European commission pesticide database
135 (ec.europa.eu/food/plant/pesticides/eu-pesticides-database/active-substances).
136 There were also substances belonging to other categories (32 other target groups;
137 Table S2 for a complete list) that we classified as “other”.

138 To relate the use of active substances to the area of arable land in postcodes, we
139 extracted the total area of cropland from the 2016 French Land Parcel Identification
140 System (LPIS, [Agence de Services et de Paiements, 2015](#)). This database is a
141 geographic information system developed under the European Council Regulation No
142 153/2000, for which the farmers provide annual information about their fields and crop
143 rotation. We grouped the 16 categories of cropland types used in LPIS into 11 sub-
144 groups (Figure S9) (Cantelaube and Carles, 2010; Levavasseur et al., 2016). We
145 summed the area of all types of cropland but meadows to obtain the total crop area
146 per postcode.

147 **1.2 Mixture model**

148 **1.2.1 Input data**

149 As described above, the dataset consisted of n (=5,361) postcodes and p (=279)
150 substances. For each postcode i ($1 \leq i \leq n$) and substance j ($1 \leq j \leq p$), we denoted
151 by X_{ij} the presence/absence variable, which is 1 if substance j is bought in postcode
152 i and 0 otherwise, and by Y_{ij} the log of the quantity of substance j bought in postcode
153 i (when used) normalized with the cropland area of postcode i :

$$154 \quad Y_{ij} = \log \left(\frac{\text{quantity of substance } j \text{ bought in postcode } i}{\text{cropland area of postcode } i} \right)$$

155 (Y_{ij} is NA when substance j is not bought in postcode i).

156 **1.2.2 Model**

157 We aimed to provide a clustering of the postcodes according to the quantity of
158 the various substances bought. Mixture models (McLahan and Peel, 2000) provide a
159 classical framework to achieve such a clustering. The model we consider assume that
160 the n postcodes are spread into K groups and that their respective use of the different
161 substances depends on the group they belong to. Mixture models precisely aim at
162 recovering this unobserved group structure from the observed data.

163 **1.2.2.1.1 Groups definition**

164 We denoted by Z_i the group to which postcode i belongs. We assumed the Z_i are
165 all independent and that each postcode i belongs to group k ($1 \leq k \leq K$) with
166 respective proportions π_k :

$$167 \quad \pi_k = \Pr\{Z_i = k\}. \quad (1)$$

168 Note that the π_k consists of only $K - 1$ independent parameters, as they have to sum
169 to one ($\sum_{k=1}^K \pi_k = 1$).

170 **1.2.2.1.2 Emission distribution**

171 The model then describes the distribution of the observed data conditional on the
172 group to which each postcode belongs. The distribution of the presence/quantity pair
173 (X_{ij}, Y_{ij}) is built in two stages: first, if postcode i belongs to group k , substance j is used
174 in the postcode with probability γ_{kj} :

$$175 \quad \gamma_{kj} = \Pr\{X_{ij} = 1 | Z_i = k\}, \quad (2)$$

176 then, if substance j is used in postcode i , its log-quantity is assumed to have a
177 Gaussian distribution:

$$178 \quad (Y_{ij} | X_{ij} = 1, Z_i = k) \sim \mathcal{N}(\mu_{kj}, \sigma_{kj}^2). \quad (3)$$

179 with μ_{kj} and σ_{kj}^2 the mean and variance of the log-quantity of substance j used in a
180 postcode from group k , provided that the substance is bought in the postcode. In
181 addition to the $(K - 1)$ proportions π_k and the $K \times p$ probabilities γ_{jk} , this model
182 involves $K \times p$ mean parameters μ_{kj} and as many variance parameters σ_{kj}^2 . This
183 makes a total of $K - 1 + 3Kp$ parameters to be estimated.

184 Combining Equations (2) and (3), we defined the conditional distribution f_{jk} for
185 substance j in a postcode from group k :

$$186 \quad f_{jk}(x_{ij}, y_{ij}) = x_{ij}\gamma_{kj}\phi(y_{ij}; \mu_{kj}, \sigma_{kj}^2) + (1 - x_{ij})(1 - \gamma_{kj})$$

187 denoting by $\phi(\cdot; \mu, \sigma^2)$ the probability density function of the Gaussian distribution
188 $\mathcal{N}(\mu, \sigma^2)$.

189 To avoid over-parametrization, we also considered models with constrained variance,
190 assuming either that the variance depends on the substance but not on the group:
191 $\sigma_{kj}^2 \equiv \sigma_j^2$, or that the variance is the same for all substances in all groups: $\sigma_{kj}^2 \equiv \sigma^2$.

192 **1.2.3 Inference**

193 Mixture models belong to incomplete-data models, i.e. they can deal with
194 situations where part of the relevant information is missing. For the sake of brevity, we
195 denoted by Y the set of observed variables (i.e. all the (X_{ij}, Y_{ij})) and by Z the set of
196 unobserved variables (i.e. the Z_i). We further denoted by θ the whole set of parameters
197 to be estimated: $\theta = (\{\pi_k\}, \{\gamma_{kj}\}, \{\mu_{kj}\}, \{\sigma_{kj}^2\})$.

198 A classical way to estimate the set of parameters θ is to maximize the log-
199 likelihood of the data $\log p(Y; \theta)$ with respect to the parameters. An important feature
200 of incomplete-data models is that this log-likelihood is not easy to compute, and even
201 harder to maximize, as its calculation requires integrating over the unobserved variable
202 Z . However, the so-called 'complete' log-likelihood, which involves both the observed
203 Y and the unobserved Z , $\log p(Y, Z; \theta)$ is often tractable.

204 **1.2.3.1.1 Expectation-Maximization algorithm**

205 The Expectation-maximization (EM) algorithm (Dempster;A.P et al., 1977) resorts
206 to the complete log-likelihood to achieve maximum-likelihood inference for the
207 parameters. More specifically, because $\log p(Y, Z; \theta)$ can not be evaluated (as Z is not
208 observed), EM uses the conditional expectation of the complete likelihood given the
209 observed data, namely $\mathbb{E}[\log p(Y, Z; \theta)|Y; \theta]$, as an objective function, to be maximized
210 with respect to θ .

211 The EM algorithm alternates the steps 'E' (for expectation) and 'M' (for
212 maximization) until convergence. It can be shown that the likelihood of the data

213 $\log p(Y; \theta)$ increases after each EM step. The reader may refer to Dempster et al.
 214 (1977) or McLahan and Peel (2000) for a formal justification of the procedure.

215 **1.2.3.1.2 E step**

216 This step aimed at recovering the relevant information to evaluate the objective
 217 function. In the case of mixture models, the E steps only amounts to evaluating the
 218 conditional probability τ_{ik} for the postcode i to belong to group k given the data
 219 observed for the postcode and the estimate of the parameter θ_{ik} after iteration $h - 1$:

$$220 \quad \tau_{ik}^{(h-1)} = \Pr\{Z_i = k | \{(X_{ij}, Y_{ij})\}_{1 \leq j \leq p}; \theta^{(h-1)}\}$$

221 The calculation of τ_{ik} simply resorts to Bayes formula. In the following, we drop the
 222 iteration superscript (h) for the sake of clarity, and we use the notation $\hat{\theta}$ to indicate
 223 the current estimate. Because the substance are assumed to be independent, we get

$$224 \quad \hat{\tau}_{ik} = \hat{\pi}_k \prod_{j=1}^p \hat{f}_{jk}(x_{ij}, y_{ij}) / \left(\sum_{\ell=1}^K \hat{\pi}_\ell \prod_{j=1}^p \hat{f}_{j\ell}(x_{ij}, y_{ij}) \right).$$

225 **1.2.3.1.3 M step**

226 The M step updates the parameter estimate by maximizing
 227 $\mathbb{E}[\log p(Y, Z; \theta) | Y; \theta^{(h-1)}]$ with respect to θ . The objective function can be calculated
 228 using the conditional probabilities τ_{ik} s

$$229 \quad \mathbb{E}[\log p(Y, Z; \theta) | Y; \theta^{(h)}] = \sum_{i=1}^n \sum_{k=1}^K \hat{\tau}_{ik} (\log \pi_k + \sum_{j=1}^p \log f_{kj}(x_{ij}, y_{ij})).$$

230 The maximization of this function yields in close-form update formulas for all
 231 parameters. All estimates can be viewed as weighted versions of intuitive proportions,
 232 means or variance. Let us first define

$$233 \quad \hat{N}_k = \sum_{i=1}^n \hat{\tau}_{ik}, \hat{M}_{kj} = \sum_{i=1}^n \hat{\tau}_{ik} x_{ij}.$$

234 \hat{N}_k is the current estimate of the number of entities belonging to group k ; \hat{M}_{kj} is the
 235 current estimate of the number of entities from group k where substance j is bought.
 236 For the proportions and probability of use, we get the following updates:

$$237 \quad \hat{\pi}_k = \hat{N}_k / n, \hat{\gamma}_{kj} = \hat{M}_{kj} / \hat{N}_k.$$

238 For the quantitative part of the model, we get additionally:

$$239 \quad \hat{\mu}_{kj} = \frac{1}{\hat{M}_{jk}} \sum_{i=1}^n \hat{\tau}_{ik} x_{ij} y_{ij} \hat{\sigma}_{kj}^2 = \left(\frac{1}{\hat{M}_{jk}} \sum_{i=1}^n \hat{\tau}_{ik} x_{ij} y_{ij}^2 \right) - (\hat{\mu}_{kj})^2.$$

240 Similar estimates of σ_j^2 and σ^2 can be derived for the models with constrained
 241 variances.

242 **1.2.4 Model selection**

243 To select the number of groups K and to choose between the models with
244 unconstrained and constrained variances, we used the Bayesian Information Criterion
245 (BIC, Schwarz, 1978). We adopted the same form as in Fraley and Raftery [1999], that
246 is:

$$247 \quad BIC = \log p(Y; \hat{\theta}) - \frac{n}{2} \log(\# \text{independent parameters}).$$

248 As indicated above, the number of independent parameters is:

- 249 • $K - 1 + 3Kp$ with unconstrained variances σ_{jk}^2 ,
- 250 • $K - 1 + 2Kp + p$ with constant variance for each substance $\sigma_{jk}^2 \equiv \sigma_j^2$,
- 251 • $K + 2Kp$ with constant variance $\sigma_{jk}^2 \equiv \sigma^2$.

252 **1.2.5 Estimated parameters**

253 The output of the mixture model yielded K groups with their corresponding
254 estimated parameters, that is $\hat{t}_{ik}, \hat{\gamma}_{kj}, \hat{\mu}_{kj}, \hat{\sigma}_{kj}^2$, with k one of the K groups obtained, j
255 an active substance and i a postcode. These estimated parameters gave information
256 on groups of postcodes and substances bought per group.

257 \hat{t}_{ik} was the conditional probability that a postcode i belong to each group k given the
258 quantities of substances bought in the postcode. We used this probability to associate
259 each postcode to its most probable group.

260 $\hat{\gamma}_{kj}$ was the probability of a substance j to be used in a postcode of group k . We used
261 this probability to study the composition of active substances in each group k .

262 $\hat{\mu}_{kj}$ and $\hat{\sigma}_{kj}^2$ were the estimated mean and variance of the log-quantity of substance j
263 per square meter of cropland purchased in a postcode from group k . These quantities
264 were used to refine our understanding of the substance composition of postcode
265 groups.

266 **1.3 Analyses on estimated parameters**

267 **1.3.1 Spatial structure of the groups**

268 To characterise the spatial structure of postcode groups, we quantified the spatial
269 spread of postcodes belonging to a same group via the area of the convex hull of the
270 group. The convex hull of a group is the smallest convex set that contains all postcodes
271 of the group.

272 Regardless of their spatial aggregation, most groups contain a few scattered
273 postcodes, such that the convex area of all groups generally contains most of France,
274 making comparisons of the area irrelevant. To circumvent this difficulty, we merged all
275 contiguous postcodes within a group into single polygons and retained only the largest
276 polygons, representing 80% of the total area of a group. This eliminated the scattered
277 postcodes outside the main core of postcodes within a group.

278 We also characterized the similarity among the K groups in terms of substance
279 use via hierarchical clustering on distances between groups. To obtain a matrix of
280 between-group distances, we used results from the mixture model and calculated a
281 maximum-likelihood inference when two randomly chosen groups were merged (see
282 method in 1.2). We repeated this step for each possible group pair. We thus obtained
283 a matrix of between-group distances, characterized as differences in likelihood
284 between mixture models. Using this matrix, we computed an agglomerative nesting
285 clustering, using Ward criterion, implemented in the R package *cluster* (Maechler et
286 al.,2019, R Core Team 2021).

287 **1.3.2 Searching for the drivers of the substance composition of groups**

288 We tried to identify some of the possible drivers of the substance composition of
289 groups using two complementary approaches. First, we tested whether the groups
290 obtained with the mixture model, which by construction differ in terms of active
291 substances purchased, also differed in terms of crop composition. To compare the
292 proportion of area covered with different crops among groups, we performed a log-
293 ratio analysis (LRA). This approach was implemented in the R package *easyCODA*
294 (Greenacre, 2019, R Core Team 2021). Second, we used Mantel tests (Mantel &
295 Valand 1970) to estimate the correlations between three distance matrices among
296 postcode groups: distances in the composition of substances purchased in the group
297 (see above), distances in crop composition, and geographic distances. We used a
298 spearman method and used 9999 permutations, computed with the *vegan* package
299 (Oksanen and Simpson, 2022)

300 **1.3.3 Test of the temporal robustness of the mixture model**

301 To test robustness of the results of the mixture model based on the pesticide
302 purchase data from the year 2017, we also run the mixture model on BNV-d data over
303 the period 2015 to 2018. To do so, we aggregated all purchase data from 2015 to 2018
304 and analysed these data in the same way as those from 2017. In the following, the
305 groups obtained with the mixture model applied on the 2017 data (respectively 2015-
306 2018 data) are referred to as the “2017 groups” (respectively the “2015-2018 groups”).

307 We used postcode probabilities to be in group k (i.e. $\hat{\tau}_{ik}$) to compare results from
308 the two mixture models, with the 2017 groups as a reference. We compared each 2017
309 group with all 2015-2018 groups by calculating the proportion of postcodes in each
310 2017 group that belong to each 2015-2018 group. We thus obtained a matrix with the
311 percentage of postcodes from 2017 groups that were found in the various 2015-2018
312 groups (Gelbard et al., 2007).

RESULTS

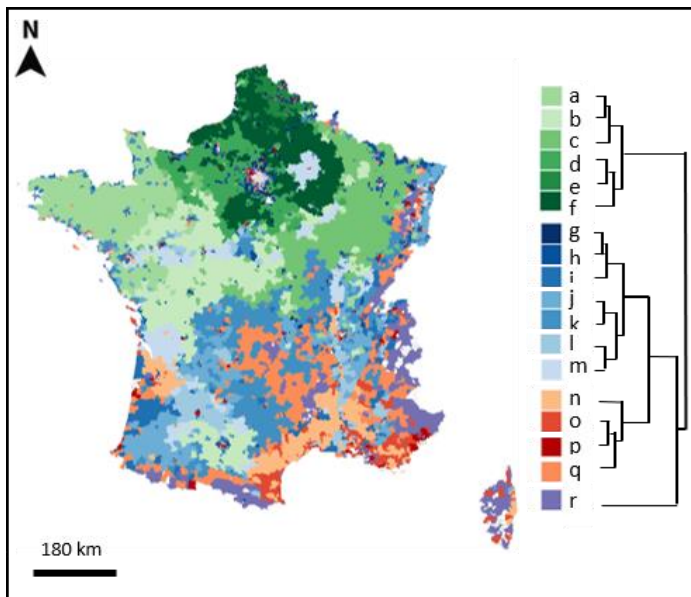
313 **1.4 The mixture model yields a small number of groups of postcodes**

314 The mixture model with unconstrained variances had the highest BIC and
315 classified the 5,631 postcodes into 18 groups on the basis of 2017 purchase data for
316 279 active substances (Figure S2). Most postcodes were unambiguously attributed to
317 a single of these groups, as shown by the bimodal distribution of the probability for a
318 postcode i to belong to group k , with most values close to 0 or 1 (Figure S3). Only 17
319 out of 5,631 postcodes had a maximum probability to be in a group lower than 0.7.

320 Most groups of postcodes identified by the mixture model were spatially
321 aggregated, albeit of contrasting sizes (Figure 1). The number of postcodes per group
322 ranged from 159 to 585 (median = 294, Q1 = 239, Q3= 362), which translated into a
323 cropland area per group ranging from 81.45 km² to 30858.86 km² (median = 4372.32
324 km², Q1 = 1761.08 km², Q3= 12863.21 km²). The cropland area of groups was
325 negatively related to the area of the convex envelop encompassing it, such that groups
326 with the largest cropland area tended to be the most spatially clustered (Figure 2).
327 Such a spatial clustering of postcodes purchasing similar pesticide substances was

328 expected as agricultural practices are spatially structured (see below) but keep in mind
329 that the mixture model did not incorporate spatial information.

330



331

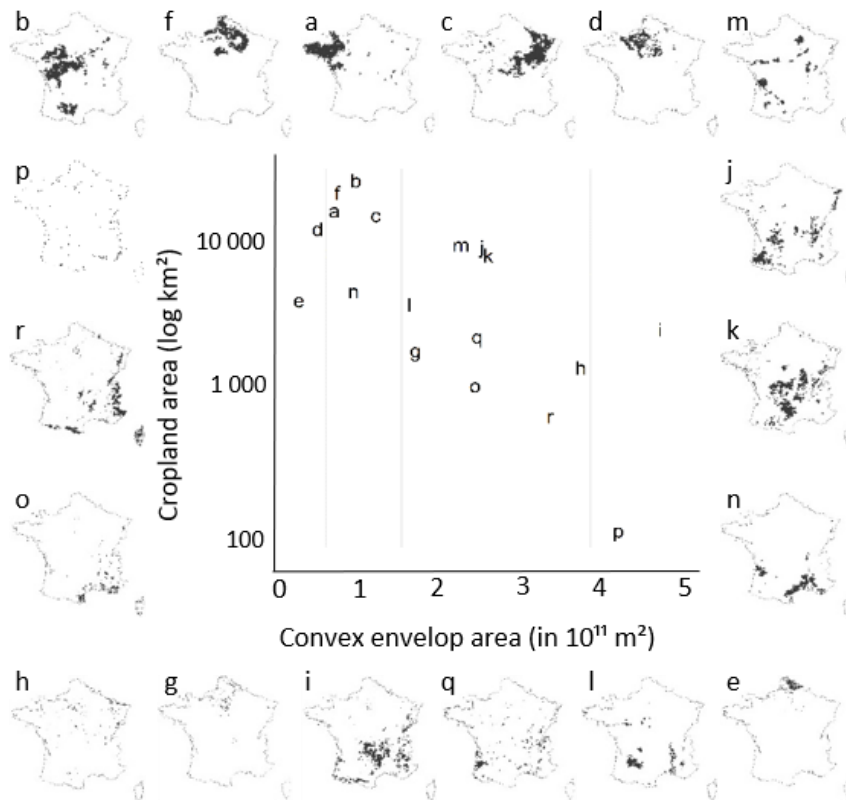
332 *Figure 1: Map of France split into postcode groups obtained from the mixture model on the basis*
333 *of active substances purchased in postcodes. Postcodes within a group share the same colour.*
334 *The dendrogram was obtained using an agglomerative hierarchical*

335

336 Postcode groups corresponded to specific geographical and/or agricultural
337 regions. For example, group a corresponded mostly to Brittany (the western peninsula)
338 and group c was predominantly located in North-eastern France. Groups m and o were
339 more scattered across the country but overlapped almost perfectly with wine regions
340 (Figure 2). Note that a few groups were composed of a limited number of postcodes
341 spatially scattered across France (e.g. groups i, h, and p, Figure 2). In particular, group
342 p represented less than 2 km² of cropland and is generally discarded in the following.

343 The groups identified by the mixture model were relatively robust to a change in
344 the temporal range of the data, as shown by the results of the mixture model on the
345 2015-2018 data (Figure S7). This second clustering yielded 24 groups and the
346 percentage of shared postcodes between the 2017 groups and their most similar 2015-
347 2018 groups varied between 38% and 83% (median= 64%, Q1=54%, Q3= 74%). For
348 example, groups corresponding to Northern France (group e vs. group 6) or the
349 Champagne region (group c vs. 2) were stable over time (Figure S7). The higher
350 number of groups obtained with the 2015-2018 mixture model (24 vs. 18) was often
351 due to the split of some 2017 groups into two 2015-2018 groups. For example, for 2017

352 group *r*, there was 48% similarity with 2015-2018 group 21 and 44% similarity with
 353 group 23. Similarly, Brittany was covered by 2017 group *a* vs. 2015-2018 groups 16
 354 and 20 (*Figure S7*). Because of this temporal consistency in the clustering, we only
 355 present in the following the analyses on the 2017 dataset, which is thought to be more
 356 accurate (see 1.1).

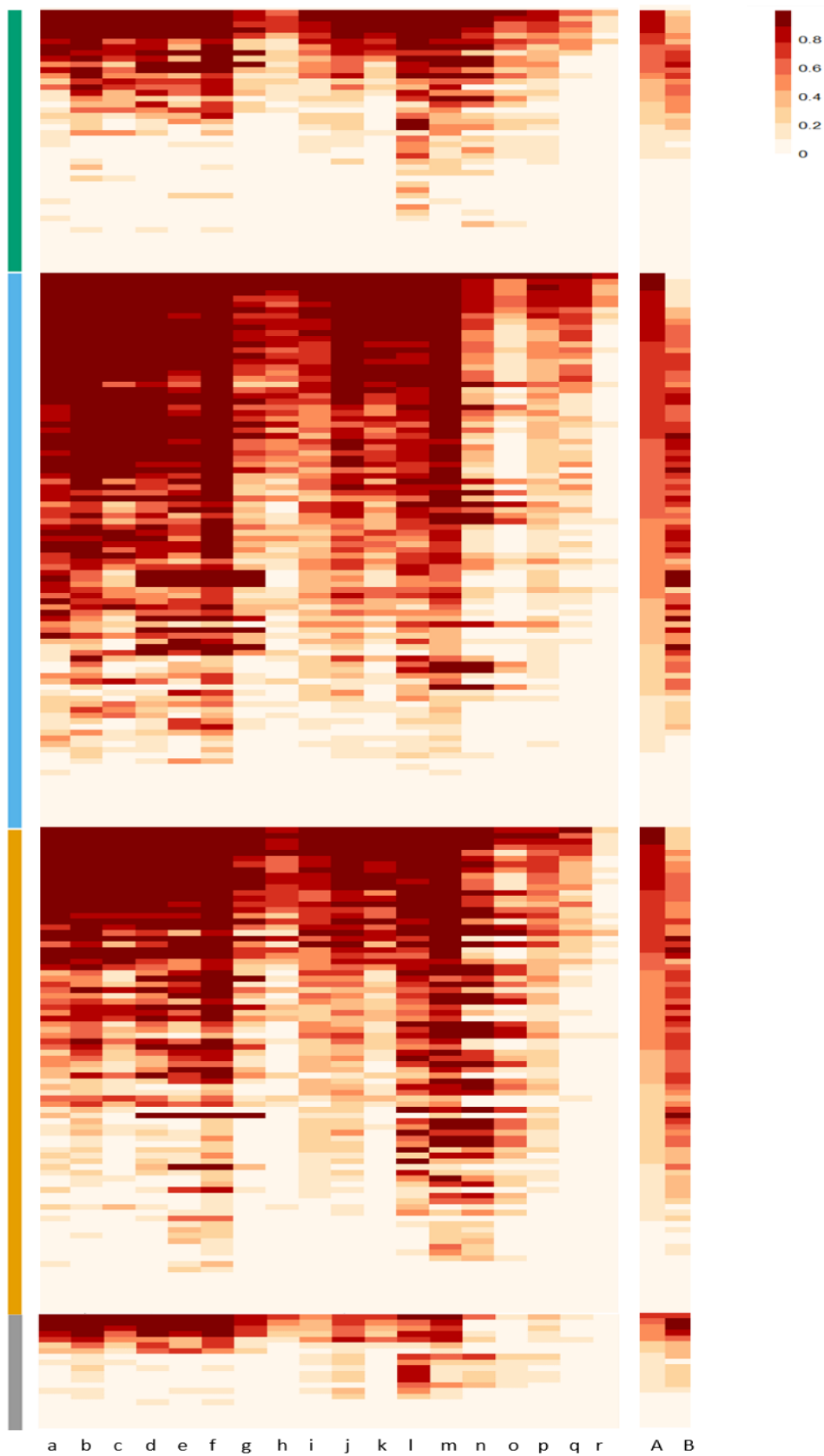


357 *Figure 2: Relationship between cropland area (log scale) and convex area, a proxy for spatial*
 358 *extent, of groups. The spatial distribution of each group is plotted around the relationship, with*
 359 *one map of France per group, in which postcodes forming each group are highlighted in black.*
 360 *Groups are ordered clockwise from top left in decreasing cropland area. Note that the focus on*
 361 *cropland area (not total area) in a postcode makes some groups with little cropland (e.g.*
 362 *mountain areas, *q* or *m*) appear with a relatively large area on the maps, although they are*
 363 *ranked low in terms of cropland area.*

364 **1.5 Substance composition of postcode groups: core and discriminating** 365 **substances**

366 Postcode groups differed in terms of the composition of substance purchased
 367 (*Figure 3*), as expected from the clustering algorithm, but may also share common
 368 substances. Group composition was inferred, and can be characterised by, (1) the

369 probability of a substance to be purchased by a postcode from a group ($\hat{\gamma}_{kj}$), and, if
370 the substance is purchased, (2) the estimated mean quantity purchased ($\hat{\mu}_{kj}$) and (3)
371 the estimated variance in the latter quantity (σ_{jk}^2). In the following, for the sake of
372 simplicity, we chose to focus on the probability of substances to be purchased, knowing
373 that this probability was positively related with the estimated mean quantity (Figure S4
374 & Figure S6, $r = 0.2$) and negatively related with the estimated variance (Figure S4, r
375 = -0.15). For a given substance, this probability can also vary substantially across
376 groups, and we used this variability to distinguish two main types of substances with
377 interest for the definition of postcode groups, for the identification of relevant pesticide
378 cocktails: core substances and discriminating substances (*Figure 4*).



379

380

Figure 3: Heatmap of the probability γ_{kj} in each group, in each of four categories of substances: insecticides (green), herbicides (blue), fungicides (orange), other targets (grey). Within each category, substances are ordered in increasing average probabilities of use across groups. For readability, substance names are not displayed and can be found in Figure S8. On the right of the figure, column A corresponds to the mean probability of use and column B corresponds to the scaled (0,1) variance in probability of use across groups.

381 Core substances, defined as substances with a high average and low variance
382 of probability to be purchased across groups, were by definition found in most groups;
383 they were widespread molecules that were likely to form the backbone of cocktails
384 encountered by living organisms in farmland. Using an arbitrary threshold value of
385 mean purchase probability of 0.85, we found 13 such core substances with high
386 probabilities (*Figure 3 & Figure S5*): two pyrethroid insecticides (deltamethrin, lambda-
387 cyhalothrin), seven herbicides of different chemical families (glyphosate, diflufenicanil,
388 fluroxypyr, MCPA, 2,4-d, triclopyr, pendimethalin) and four fungicides (fludioxonil,
389 tebuconazole, difenoconazole and thiram). Because they were found with high
390 probability in most groups, these substances were unlikely to weight strongly in the
391 definition of postcode groups, although they can contribute via differences in the mean
392 quantities used across groups. For example, the average estimated amount of
393 glyphosate purchased ranged from 30 to 634 kg/cropland m² (median= 43, Q1= 36,
394 Q3 = 77) among groups.

395 Discriminating substances are defined as substances with medium to high mean
396 probability of purchase, mechanically associated with a large variance across groups
397 in this probability (*Figure S5*). Because of their contrasting probability of purchase
398 across groups, discriminating substances were likely to contribute greatly to the
399 formation of groups. We used the arbitrary range of average probabilities from 0.5 to
400 0.85 to define discriminating substances. Using these thresholds, we found a set of 85
401 discriminating substances, including 42 herbicides, 28 fungicides, 10 insecticides and
402 5 with other targets (*Supplementary information 2*). In the following, we focus on
403 discriminating substances that are highly probable ($\hat{y}_{kj} > 0.85$) in at least one postcode
404 group, i.e. substances that are likely major components of pesticide cocktails occurring
405 in a given group. We found five widespread discriminating substances purchased with
406 a probability higher than 0.85 in 13 out of 18 groups: azoxystrobin, iodosulfuron-
407 methyl-sodium, mesotrione, metsulfuron-methyl and prothioconazol. These
408 substances are very close to core substances. Conversely, eight substances were
409 highly specific, being purchased with high probability (>0.85) in three groups only (e.g.
410 dimethachlore in groups *d*, *c*, and *f*). Within a group, the number of discriminating
411 substances with high probability of purchase (> 0.85) varied strongly among groups,
412 from 2 for group *h* to 80 for group *f* (mean= 42 ± 26). This cross-group variation in the
413 number of highly probable discriminating substances has implication for the
414 composition and complexity of the pesticide cocktails in French agroecosystems: from

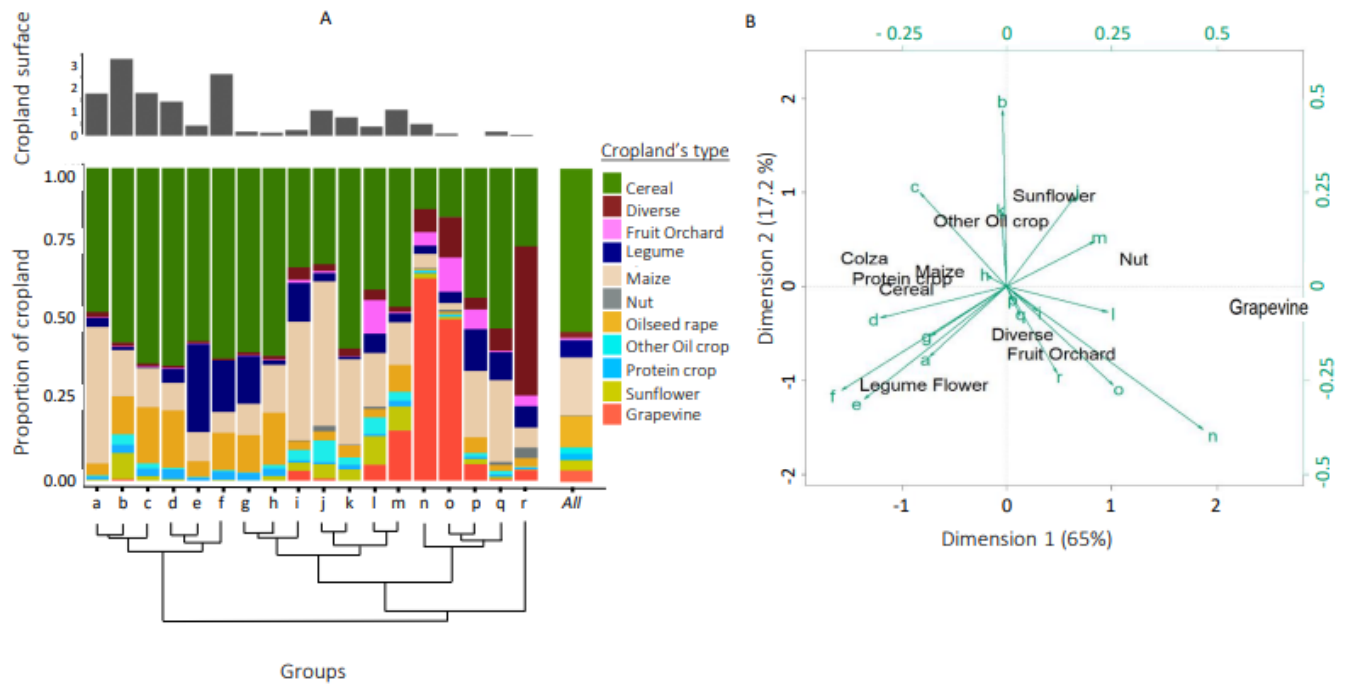
415 relatively “simple” (13 core substances and two discriminating substance in group *h*)
416 to highly complex (13 core substances and 80 discriminating substances in group *f*).

417 The 181 remaining substances, with a low average probability to be purchased
418 (<0.5), also had a role in group identification, but were seldom purchased and will not
419 be described further (*Figure 3*).

420 **1.6 Postcode groups differ in terms of crop composition, but active** 421 **substance purchase may not be solely driven by crop identity**

422 Groups of postcodes, which by construction are composed of different cocktails
423 of substances, also differed in terms of proportions of cropland grown with various
424 crops, such that groups with close pesticide composition sometimes, but not always,
425 also exhibited similar crop usage (*Figure 4*). The possible relations between pesticide
426 composition and crop composition can be visualized either on *Figure 4*, where crop
427 composition of groups similar in terms of pesticides purchases are plotted next to each
428 other, or on the biplot of the log ratio analysis (*Figure 5*), in which groups with similar
429 crop composition are plotted next to each other. For example, groups *o* and *n*,
430 characterized by a large proportion of vineyards, were close to each other both in the
431 log-ratio analysis, which is indicative of similar crop compositions (*Figure 5*) and in the
432 hierarchical clustering, which is indicative of similar pesticide purchases (*Figure 4*).
433 The same was true for groups *e* and *f*, and, to a lesser extent, *d*, characterized by an
434 appreciable proposition of crops from the legumes/flowers category. However, some
435 groups such as *f* and *g* were different in terms of substances (not in the same sub-
436 group, *Figure 1 & 4*) while exhibiting comparable proportions of crop types (*Figure 4*).
437 Alternatively, some groups that were closely related in terms of substance purchases,
438 such as groups *a* and *b*, could be characterized by dissimilar crop compositions. The
439 latter patterns may suggest regionalisation of substance use, such that neighbouring
440 regions tend to use similar products or substances even with variations in crops grown
441 (e.g. *a* and *b*), or distant regions tend to use different products or substances even with
442 similar crops (e.g. *a* and *k*).

443



444

445 *Figure 4: A. Distribution of crop type area across groups. The top grey histogram shows the*
 446 *distribution of total cropland area across groups (in 10⁴ km²). The dendrogram was obtained*
 447 *using an agglomerative hierarchical clustering on the basis of Ward's method among groups*
 448 *(see 2.2.1). B. Biplot of the log ratio analysis relating the proportion of crop types in each group.*
 449 *Only groups identified as spatially coherent are displayed (see 3.2). For readability, the groups*
 450 *and crop types are displayed on two different scales: black for crop types, green for groups. The*
 451 *size of arrows corresponds to the contribution of each group. Groups that appear close to each*
 452 *other on the biplot have similar crop composition, which can be inferred from the contribution*
 453 *of each crop type to the axes.*

454

455 Despite the abovementioned associations between the crop composition and
 456 active substance compositions of groups, we found no significant correlation between
 457 distance matrices: the distance in substance composition among groups was not
 458 correlated with the distance in crop composition (Mantel test, $\rho = 0.12$, $P = 0.12$).
 459 However, we found a correlation between the geographic distance and active
 460 substance compositions of groups (Mantel test, $\rho = 0.12$, $P = 0.12$) indicating that
 461 spatially closer postcodes groups are more similar in composition of actives
 462 substances.

DISCUSSION

463 A major challenge in pesticide risks assessment is to characterise cocktails of
464 pesticides used in the field (Lydy et al., 2004), partly because of the large number of
465 substances used but also because of the limited information on the combinations of
466 used substances contaminating the environment. Here, we developed a methodology
467 to analyse a newly available database on pesticide purchases across France. It aimed
468 to identify groups of postcodes with similar compositions of pesticide purchases and
469 characterise their spatial structure, two critical pieces of information to unravel the
470 composition of pesticide cocktails. Our method resulted in the clustering of the 5,631
471 French postcodes into a relatively low number of groups. These groups represent as
472 many potential pesticide cocktails, which is much lower than the possible combinations
473 among the 279 substances included in the data. In the following, we discuss how our
474 findings can help understand the impacts of pesticides in the environment (e.g. by
475 identifying relevant pesticide cocktails) and how this approach can be improved in the
476 future and the possible mechanisms underlying the groups.

477 ***1.7 Significance of the identification of highly probable active substances and*** 478 ***of cocktails of active substances characteristic of postcode groups for the*** 479 ***study of the impacts of pesticides in the environment***

480 The identification of active substances that are purchased with high probability in
481 all (core substances) or a subset (discriminating substances) of postcode groups might
482 contribute to reducing the potential street light effect, whereby most research efforts
483 focus on molecules that are either easy to study (Hendrix, 2017) or that were
484 popularized by previous studies (Tsvetkov and Zayed, 2021). Unsurprisingly, most
485 core substances identified here are already well-known, widely-used substances.
486 Glyphosate is the most widely used broad-spectrum herbicide (Jatinder Pal Kaur Gill
487 et al. 2017; Myers et al. 2016), with associated concerns regarding pervasive direct
488 and indirect effects (Van Bruggen et al., 2018). Tebuconazole and difenoconazole, two
489 triazole fungicides, are widely used and studied (Zubrod et al., 2019). Deltamethrin and
490 lambda-cyhalothrin, two pyrethroids impacting nervous systems (Ray and Fry, 2006;
491 Soderlund and Bloomquist, 1989), are known to have adverse effects on a large range

492 of non-target species such as fish, birds and amphibians (Ali et al. 2011). Yet, a
493 preliminary literature search on these 13 core substances suggests that the research
494 effort on their adverse effects on biodiversity is still highly variable. For core herbicides,
495 a simple search of the molecule name together with “biodiversity” or “ecotoxicology” in
496 the abstract of articles on ISI Web of Science yields more than two hundred research
497 articles for glyphosate and around seventy for 2,4-d, but only 2 to 17 articles for
498 diflufenican, fluroxypyr, MCPA, triclopyr and pendimethalin. For core insecticides, the
499 same search returns ca. 40 articles for lambda-cyhalothrin and deltamethrin. The four
500 core fungicides were no exception, with a number of research articles of less than ten
501 for thiam, fludioxonil and difenoconazole and around thirty for tebuconazol. Ultimately,
502 our method eases the bottom-up approach in the laboratory by providing a selection of
503 understudied substances deserving further attention.

504 Studying all possible (combinations of) substances is prohibitive (Wolska et al.,
505 2007); beyond the identification of single substances, our approach chiefly contributes
506 to identifying combinations of active substances that are likely to be encountered in
507 farmland environments, i.e. pesticide cocktails. The mixture model identified a
508 relatively small number of postcode groups (18 to 24 depending on the temporal
509 coverage of pesticide data). Each group is characterized by a specific combination of
510 purchases of active substances and can be interpreted as potential cocktails of
511 pesticides occurring in the location of the postcodes, under the assumption that all
512 purchased substances are used within the buying area during the year of purchased
513 (see “Limitations” below). Among the 269 active substances considered in these
514 analyses, we highlighted the core substances included in most cocktails and the
515 discriminating substances specific to particular cocktails. Within each postcode group,
516 both types of substances might be a good starting shortlist of substances within which
517 one can investigate potential interactive effects on biodiversity. Indeed, these
518 substances are purchased with high probability in at least some large groups of
519 postcodes, hence potentially part of widespread cocktails. Although this list is much
520 shorter than the total list of authorized active substances, it still contains 13 core
521 substances, plus 2 to 80 discriminating substances depending on the postcode group.
522 Since our approach to identifying core and discriminating substances was based on
523 probability of purchase only, this shortlist of substances could be narrowed down
524 further by selecting active substances bought in large quantities (see also “Limitations
525 and perspectives”, 1.8) or with high toxicity. The appreciable number of core and

526 discriminating substances composing cocktails is anyway consistent with surveys
527 showing that active substances are rarely found alone in the environment (Silva et al.,
528 2019). It also further substantiates the need for a broader assessment of the synergistic
529 effects of pesticides on biodiversity, often completed on a limited set of substances
530 only (Schreiner et al., 2016; Silva et al., 2019). For core substances, for example, some
531 cocktails effects have already been studied. but mostly on pairs of substances (Brodeur
532 et al., 2014; Peluso et al., 2022) and more rarely for cocktails of three or more
533 substances (Cedergreen, 2014), but in any cases, many more combinations remain
534 untested. Focusing on the reasonable number of relatively complex cocktails identified
535 by the present approach would contribute to improve our understanding of the
536 synergistic effects of realistic cocktails on organisms.

537 **1.8 Limitations & perspectives**

538 **1.8.1 Limited spatio-temporal resolution of the BNV-d data**

539 The first limitation of our study is associated with the BNV-d database, which
540 provides information on quantity and year of pesticide purchase, as well as on the
541 administrative location of the buyer, but not on the actual date and location of pesticide
542 treatments, nor on the actual pesticide contamination of the various postcodes. For
543 simplicity, we assumed that the pesticides were used in the year of purchase and in
544 the postcode of purchase. These assumptions may not be verified under all
545 circumstances because farmers are sometimes known to store some pesticide
546 products despite their high prices, e.g. to anticipate increased taxes, and because
547 farms are sometimes spread across several postcodes. Yet, there are a couple of
548 indications that the assumption of immediate and local use of pesticides is generally
549 correct. For example, our results are consistent with those of an extensive European
550 study on soil contamination (Silva et al., 2019) which identified glyphosate and the
551 fungicides boscalid, epoxiconazole, and tebuconazole as the most frequent and most
552 abundant contaminants. These substances either belong to the core substances we
553 identified (glyphosate and tebuconazole) or to discriminant substances (boscalid and
554 epoxiconazole) with a high probability of being used over half of the postcode groups.

555 Although our estimation of pesticide cocktail composition may be roughly correct
556 at the resolution of a postcode and of a year, the actual use of pesticides in space and
557 time varies at much finer scales than those of available data. Pesticide substances
558 bought within a given postcode and year may be spread in contrasting fields and times
559 and may not be found together in the environment, depending on their half-life and
560 transport in the environment. The actual cocktail composition of a site hence depends,
561 among others, on the crop cover in the landscape and associated farming practices.
562 Downscaling the BNV-d database to the field scale is challenging (Cahuzac et al. 2018;
563 Ramalanjaona, 2020), but it might reveal other patterns than the ones we highlighted
564 here, probably decreasing the number of substances that are part of local cocktails.
565 Such fine-grained data on pesticides might be more relevant to assess the impact of
566 pesticide contamination on biodiversity.

567 ***1.8.2 Going beyond the use of purchase probabilities and arbitrary thresholds*** 568 ***to identify the substances of interest for risks assessment***

569 The method we developed is continuous, with quantitative estimates of purchase
570 probabilities, as well as mean and variance of quantities purchased per postcode
571 group. Still, we used arbitrary thresholds to identify core and discriminating
572 substances. The cocktails composition we highlighted here are then dependent on the
573 chosen thresholds. Depending on the question of interest, these thresholds can and
574 should be adapted. For example, by changing the threshold to 0.80, there are nine
575 more core substances, and among these substances there are, for example,
576 imidacloprid and boscalid, both known for high use and effects on biodiversity (Lopez-
577 Antia et al., 2015; Qian et al., 2018; Simon-Delso et al., 2017; Yang et al., 2008).

578 In addition, most of our interpretation of pesticide cocktail composition relies on
579 the estimated purchase probabilities, but these cocktails were also identified using
580 information on the mean and variance of purchased amounts within postcodes, hence
581 cocktails differ for these variables as well. For example, glyphosate, a core substance
582 with high purchase probability in all postcode groups, was bought in contrasting
583 quantities across postcode groups: the average amount was 8.8×10^7 Kg/m² and
584 ranged from 3×10^7 Kg/m² in group *q* to 6×10^8 Kg/m² in group *r*. Although the purchase
585 probability was positively correlated to the mean purchased quantity and negatively to

586 its variance, the correlation is not strong, and further analysis is needed to fully uncover
587 variation in substance quantities within the cocktails we identified.

588 ***1.8.3 Taking into account the yearly variation of pesticides use***

589 Our analysis appeared relatively robust to yearly variation in pesticide purchase
590 when comparing the postcode groups obtained between the 2017 and the 2015-2018
591 datasets. This strong correlation between the 2017 and the 2015-2018 analysis is not
592 entirely surprising because of the presence of the 2017 data in both analyses. Yet,
593 adding three years of data into the analysis did not affect much the composition of
594 postcode groups, which suggests relatively stable patterns of pesticide purchase in
595 France over a short time period. Nonetheless, we observed some differences, mainly
596 due to the split of some groups, which were also expected due to climatic variation,
597 changes in legislation on pesticide use (Urruty et al., 2016) or changes in crop areas
598 (Levavasseur et al., 2016). A better integration of the temporal dynamics of pesticide
599 purchases in the characterisation of pesticide cocktails is needed if we are to monitor
600 pesticide cocktails across France. This can be achieved by applying the mixture model
601 to each year of data separately. Investigating the spatial stability of groups and
602 cocktails compositions across years would contribute to either estimate annual
603 cocktails or to find temporarily stable cocktails. Finding recurrent cocktails could
604 facilitate risk assessment over years. Indeed, this could provide key information on
605 frequency of cocktails encountered by organisms as repeated contact might increase
606 risks (Stuligross and Williams, 2021).

607 ***1.9 Postcode groups are related to the crop they grow but also to other*** 608 ***regional factors, but the underlying mechanisms remain to be fully*** 609 ***identified***

610 Although no spatial information was included in the mixture model analysis, the
611 postcode groups exhibited a strong spatial structure, in which most groups are strongly
612 aggregated and only a few small groups are scattered across France. Such spatial
613 structure was expected since pesticide use is strongly crop-dependent. For example,
614 acetamiprid, a substance used to protect fruit trees or grapevine against aphids, is
615 bought with high probability in group / only, a group with a high proportion of fruit

616 orchard and grapevine. Similarly, cyproconazole, a substance with a broader spectrum
617 of use, is bought with high probability in several groups with contrasting crop
618 compositions (*a, b, c, d, e, f, g, j, k, m*) (Figure 4). However, differences from this pattern
619 were found: some postcode groups spatially close can have different set of crops but
620 similar substance purchases or some postcode groups spatially distant can have
621 similar set of crops but different substance purchases. This result suggests that local
622 conditions, such as climate or pests, or some regional patterns in the pesticide market
623 and/or distribution, can drive the purchase of active substances more than the set of
624 crops grown (Silva et al., 2019; Storck et al., 2017). Hence, the differences among
625 postcode groups were related to a combination of crop identity effects and other
626 regional effects that will need additional analysis to be identified.

CONCLUSION

627 This study shows that a finite reasonably low number of cocktails of substances
628 can be identified at the scale of France. It is important to pursue ecotoxicological
629 studies on the synergistic effects of mixtures to identify risks and better understand the
630 effects of pesticide products on organisms. The mapping of these pesticide cocktails
631 makes it possible to identify the regions under different regimes of pesticide
632 contamination. This might be particularly useful to plan *in situ* tests for both pesticide
633 contamination and effects on biodiversity. Here we did not investigate the effects of
634 cocktails on wild organisms, and further work should be done on this aspect.

Acknowledgement

635 This project was funded and supported by ANSES (grant agreement 2019-CRB-
636 03_PV19) via the tax on sales of plant protection products. The proceeds of this tax
637 are assigned to ANSES to finance the establishment of the system for monitoring the
638 adverse effects of plant protection products, called 'phytopharmacovigilance' (PPV),
639 established by the French Act on the future of agriculture of 13 October 2014. We wish
640 to thank the steering committee of the project: Fabrizio Botta, Sandrine Charles, Marc
641 Girondot, Olivier Le Gall, Thomas Quintaine, and Lynda Saibi-Yedjer. Milena Cairo

642 was supported by ANR project VITIBIRD (ANR-20-CE34-0008) while working on this
643 project. This work also benefitted from the support of the project ECONET (ANR-18-
644 CE02-0010)

Conflict of interest

None.

SUPPLEMENTARY MATERIALS

Supplementary materials to this article can be found online at

<https://doi.org/10.5281/zenodo.7198832>

REFERENCES

- 645 Ali, S. F., Shieh, B. H., Alehaideb, Z., Khan, M. Z., Louie, A., Fageh, N., & Law, F. C.
646 (2011). A review on the effects of some selected pyrethroids and related
647 agrochemicals on aquatic vertebrate biodiversity. *Canadian Journal of Pure &*
648 *Applied Sciences*, 5(2), 1455-1464.
- 649 Altenburger, R., Backhaus, T., Boedeker, W., Faust, M., Scholze, M., 2013.
650 Simplifying complexity: Mixture toxicity assessment in the last 20 years. *Environ.*
651 *Toxicol. Chem.* 32, 1685–1687. <https://doi.org/10.1002/etc.2294>
- 652 Boedeker, W., Watts, M., Clausing, P., Marquez, E., 2020. The global distribution of
653 acute unintentional pesticide poisoning: estimations based on a systematic
654 review. *BMC Public Health* 20, 1–19. [https://doi.org/10.1186/s12889-020-09939-](https://doi.org/10.1186/s12889-020-09939-0)
655 [0](https://doi.org/10.1186/s12889-020-09939-0)
- 656 Bopp, S.A.K., Klenzier, A., van der Linden, S., Lamon, L., Paini, A., Parissis, N.,
657 Richarz, A.-N., Triebe, J., Worth, A., 2016. Review of case studies on the human
658 and environmental risk assessment of chemical mixtures.
659 <https://doi.org/10.2788/272583>
- 660 Botías, C., David, A., Horwood, J., Abdul-Sada, A., Nicholls, E., Hill, E., Goulson, D.,
661 2015. Neonicotinoid Residues in Wildflowers, a Potential Route of Chronic
662 Exposure for Bees. *Environ. Sci. Technol.* 49, 12731–12740.
663 <https://doi.org/10.1021/acs.est.5b03459>
- 664 Brittain, C.A., Vighi, M., Bommarco, R., Settele, J., Potts, S.G., 2010. Impacts of a
665 pesticide on pollinator species richness at different spatial scales. *Basic Appl.*
666 *Ecol.* 11, 106–115. <https://doi.org/10.1016/j.baae.2009.11.007>
- 667 Brodeur, J.C., Poliserpi, M.B., D'Andrea, M.F., Sánchez, M., 2014. Synergy between
668 glyphosate- and cypermethrin-based pesticides during acute exposures in

669 tadpoles of the common South American Toad *Rhinella arenarum*.
670 *Chemosphere* 112, 70–76. <https://doi.org/10.1016/j.chemosphere.2014.02.065>

671 Busse, M.D., Ratcliff, A.W., Shestak, C.J., Powers, R.F., 2001. Glyphosate toxicity
672 and the effects of long-term vegetation control on soil microbial communities.
673 *Soil Biol. Biochem.* 33, 1777–1789. [https://doi.org/10.1016/S0038-](https://doi.org/10.1016/S0038-0717(01)00103-1)
674 [0717\(01\)00103-1](https://doi.org/10.1016/S0038-0717(01)00103-1)

675 Cantelaube, P., Carles, M., 2010. Le registre parcellaire graphique : des donn é es g
676 é ographiques pour d é crire la couverture du sol agricole.

677 Cedergreen, N., 2014. Quantifying synergy: A systematic review of mixture toxicity
678 studies within environmental toxicology. *PLoS One* 9.
679 <https://doi.org/10.1371/journal.pone.0096580>

680 Deguines, N., Jono, C., Baude, M., Henry, M., Julliard, R., Fontaine, C., 2014. Large-
681 scale trade-off between agricultural intensification and crop pollination services.
682 *Front. Ecol. Environ.* 12, 212–217. <https://doi.org/10.1890/130054>

683 Dempster, A.P., Laird, N., Rubin, D., 1977. Maximum Likelihood from Incomplete
684 data via the EM Algorithm.

685 Dudley, N., Attwood, S.J., Goulson, D., Jarvis, D., Bharucha, Z.P., Pretty, J., 2017.
686 How should conservationists respond to pesticides as a driver of biodiversity loss
687 in agroecosystems? *Biol. Conserv.* 209, 449–453.
688 <https://doi.org/10.1016/j.biocon.2017.03.012>

689 Fritsch, C., Appenzeller, B., Burkart, L., Coeurdassier, M., Scheifler, R., Raoul, F.,
690 Driget, V., Powolny, T., Gagnaison, C., Rieffel, D., Afonso, E., Goydadin, A.C.,
691 Hardy, E.M., Palazzi, P., Schaeffer, C., Gaba, S., Bretagnolle, V., Bertrand, C.,
692 2022. Pervasive exposure of wild small mammals to legacy and currently used
693 pesticide mixtures in arable landscapes. *Sci. Rep.* 1–22.
694 <https://doi.org/10.1038/s41598-022-19959-y>

695 Furlan, L., Pozzebon, A., Duso, C., Simon-Delso, N., Sánchez-Bayo, F., Marchand,
696 P.A., Codato, F., Bijleveld van Lexmond, M., Bonmatin, J.M., 2018. An update of
697 the Worldwide Integrated Assessment (WIA) on systemic insecticides. Part 3:
698 alternatives to systemic insecticides. *Environ. Sci. Pollut. Res.* 1–23.
699 <https://doi.org/10.1007/s11356-017-1052-5>

700 Geiger, F., Bengtsson, J., Berendse, F., Weisser, W.W., Emmerson, M., Morales,
701 M.B., Ceryngier, P., Liira, J., Tschardt, T., Winqvist, C., Eggers, S.,
702 Bommarco, R., Pärt, T., Bretagnolle, V., Plantegenest, M., Clement, L.W.,
703 Dennis, C., Palmer, C., Oñate, J.J., Guerrero, I., Hawro, V., Aavik, T., Thies, C.,
704 Flohre, A., Hänke, S., Fischer, C., Goedhart, P.W., Inchausti, P., 2010.
705 Persistent negative effects of pesticides on biodiversity and biological control
706 potential on European farmland. *Basic Appl. Ecol.* 11, 97–105.
707 <https://doi.org/10.1016/j.baae.2009.12.001>

708 Gelbard, R., Goldman, O., Spiegler, I., 2007. Investigating diversity of clustering
709 methods: An empirical comparison. *Data Knowl. Eng.* 63, 155–166.
710 <https://doi.org/10.1016/j.datak.2007.01.002>

711 Gibbons, D., Morrissey, C., Mineau, P., 2015. A review of the direct and indirect
712 effects of neonicotinoids and fipronil on vertebrate wildlife. *Environ. Sci. Pollut.*

713 Res. 22, 103–118. <https://doi.org/10.1007/s11356-014-3180-5>

714 Greenacre, M., 2019. Variable Selection in Compositional Data Analysis Using
715 Pairwise Logratios. *Math. Geosci.* 51, 649–682. <https://doi.org/10.1007/s11004->
716 018-9754-x

717 Hallmann, C.A., Foppen, R.P.B., Van Turnhout, C.A.M., De Kroon, H., Jongejans, E.,
718 2014. Declines in insectivorous birds are associated with high neonicotinoid
719 concentrations. *Nature* 511, 341–343. <https://doi.org/10.1038/nature13531>

720 Hendrix, C.S., 2017. The streetlight effect in climate change research on Africa. *Glob.*
721 *Environ. Chang.* 43, 137–147. <https://doi.org/10.1016/j.gloenvcha.2017.01.009>

722 Hernández, A.F., Gil, F., Lacasaña, M., 2017. Toxicological interactions of pesticide
723 mixtures: an update. *Arch. Toxicol.* 91, 3211–3223.
724 <https://doi.org/10.1007/s00204-017-2043-5>

725 Heys, K.A., Shore, R.F., Pereira, M.G., Jones, K.C., Martin, F.L., 2016. Risk
726 assessment of environmental mixture effects. *RSC Adv.* 6, 47844–47857.
727 <https://doi.org/10.1039/c6ra05406d>

728 Humann-Guillemot, Ségolène, Binkowski, Ł.J., Jenni, L., Hilke, G., Glauser, G.,
729 Helfenstein, F., 2019. A nation-wide survey of neonicotinoid insecticides in
730 agricultural land with implications for agri-environment schemes. *J. Appl. Ecol.*
731 56, 1502–1514. <https://doi.org/10.1111/1365-2664.13392>

732 Humann-Guillemot, S., Tassin de Montaigu, C., Sire, J., Grünig, S., Gning, O.,
733 Glauser, G., Vallat, A., Helfenstein, F., 2019. A sublethal dose of the
734 neonicotinoid insecticide acetamiprid reduces sperm density in a songbird.
735 *Environ. Res.* 177, 108589. <https://doi.org/10.1016/j.envres.2019.108589>

736 Junghans, M., Backhaus, T., Faust, M., Scholze, M., Grimme, L.H., 2006. Application
737 and validation of approaches for the predictive hazard assessment of realistic
738 pesticide mixtures. *Aquat. Toxicol.* 76, 93–110.
739 <https://doi.org/10.1016/j.aquatox.2005.10.001>

740 Keplinger, M.L., Deichmann, W.B., 1967. Acute toxicity of combinations of pesticides.
741 *Toxicol. Appl. Pharmacol.* 10, 586–595. <https://doi.org/10.1016/0041->
742 008X(67)90097-X

743 Levavasseur, F., Martin, P., Bouty, C., Barbottin, A., Bretagnolle, V., Théron, O.,
744 Scheurer, O., Piskiewicz, N., 2016. RPG Explorer: A new tool to ease the
745 analysis of agricultural landscape dynamics with the Land Parcel Identification
746 System. *Comput. Electron. Agric.* 127, 541–552.
747 <https://doi.org/10.1016/j.compag.2016.07.015>

748 Lewis, K.A., Tzilivakis, J., Warner, D.J., Green, A., 2016. An international database
749 for pesticide risk assessments and management. *Hum. Ecol. risk Assess.* 22,
750 1050–1064. <https://doi.org/10.1017/CBO9781107415324.004>

751 Lopez-Antia, A., Ortiz-Santaliestra, M.E., Mougeot, F., Mateo, R., 2015. Imidacloprid-
752 treated seed ingestion has lethal effect on adult partridges and reduces both
753 breeding investment and offspring immunity. *Environ. Res.* 136, 97–107.
754 <https://doi.org/10.1016/j.envres.2014.10.023>

755 Lydy, M., Belden, J., Wheelock, C., Hammock, B., Denton, D., 2004a. Challenges in
756 regulating pesticide mixtures. *Ecol. Soc.* 9. <https://doi.org/10.5751/ES-00694->

757 090601

758 Lydy, M., Belden, J., Wheelock, C., Hammock, B., Denton, D., 2004b. Challenges in
759 Regulating Pesticide Mixtures. *Ecol. Soc.* 53, 1689–1699.

760 Mahmood, I., Sameen, R.I., Shazadi, K., Alvina, G., Hakeem, K.R., 2016. Effects of
761 Pesticides on Environment. *Plant, Soil Microbes Vol. 1 Implic. Crop Sci.* 1–366.
762 <https://doi.org/10.1007/978-3-319-27455-3>

763 Millot, F., Decors, A., Mastain, O., Quintaine, T., Berny, P., Vey, D., Lasseur, R., Bro,
764 E., 2017. Field evidence of bird poisonings by imidacloprid-treated seeds: a
765 review of incidents reported by the French SAGIR network from 1995 to 2014.
766 *Environ. Sci. Pollut. Res.* 24, 5469–5485. <https://doi.org/10.1007/s11356-016-8272-y>

767

768 Navarro, J., Hadjikakou, M., Ridoutt, B., Parry, H., Bryan, B.A., 2021. Pesticide
769 toxicity hazard of agriculture: regional and commodity hotspots in Australia.
770 *Environ. Sci. Technol.* 55, 1290–1300. <https://doi.org/10.1021/acs.est.0c05717>

771 Oksanen, J., Simpson, G.L., 2022. Package ‘vegan.’

772 Peluso, J., Furió Lanuza, A., Pérez Coll, C.S., Aronzon, C.M., 2022. Synergistic
773 effects of glyphosate- and 2,4-D-based pesticides mixtures on *Rhinella*
774 *arenarum* larvae. *Environ. Sci. Pollut. Res.* 29, 14443–14452.
775 <https://doi.org/10.1007/s11356-021-16784-0>

776 Qian, L., Qi, S., Cao, F., Zhang, J., Zhao, F., Li, C., Wang, C., 2018. Toxic effects of
777 boscalid on the growth, photosynthesis, antioxidant system and metabolism of
778 *Chlorella vulgaris*. *Environ. Pollut.* 242, 171–181.
779 <https://doi.org/10.1016/j.envpol.2018.06.055>

780 Ramalanjaona, L., 2020. Mise à jour du calcul des coefficients de répartition spatiale
781 des données de la BNVD Note méthodologique 95.

782 Ray, D.E., Fry, J.R., 2006. A reassessment of the neurotoxicity of pyrethroid
783 insecticides. *Pharmacol. Ther.* 111, 174–193.
784 <https://doi.org/10.1016/j.pharmthera.2005.10.003>

785 Relyea, R.A., 2009. A cocktail of contaminants: How mixtures of pesticides at low
786 concentrations affect aquatic communities. *Oecologia* 159, 363–376.
787 <https://doi.org/10.1007/s00442-008-1213-9>

788 Rundlöf, M., Andersson, G.K.S., Bommarco, R., Fries, I., Hederström, V.,
789 Herbertsson, L., Jonsson, O., Klatt, B.K., Pedersen, T.R., Yourstone, J., Smith,
790 H.G., 2015. Seed coating with a neonicotinoid insecticide negatively affects wild
791 bees. *Nature* 521, 77–80. <https://doi.org/10.1038/nature14420>

792 Schreiner, V.C., Szöcs, E., Bhowmik, A.K., Vijver, M.G., Schäfer, R.B., 2016.
793 Pesticide mixtures in streams of several European countries and the USA. *Sci.*
794 *Total Environ.* 573, 680–689. <https://doi.org/10.1016/j.scitotenv.2016.08.163>

795 Schwarz, G., 1978. Estimating the dimension of a model. *Ann. Stat.* 6, 461–464.

796 Sheahan, M., Barrett, C.B., Goldvale, C., 2017. Human health and pesticide use in
797 Sub-Saharan Africa. *Agric. Econ. (United Kingdom)* 48, 27–41.
798 <https://doi.org/10.1111/agec.12384>

799 Silva, V., Mol, H.G.J., Zomer, P., Tienstra, M., Ritsema, C.J., Geissen, V., 2019.
800 Pesticide residues in European agricultural soils – A hidden reality unfolded. *Sci.*

801 Total Environ. 653, 1532–1545. <https://doi.org/10.1016/j.scitotenv.2018.10.441>
802 Simon-Delso, N., San Martin, G., Bruneau, E., Hautier, L., Medrzycki, P., 2017.
803 Toxicity assessment on honey bee larvae of a repeated exposition of a systemic
804 fungicide, boscalid. *Bull. Insectology* 70, 83–90.
805 Soderlund, D.M., Bloomquist, J.R., 1989. Neurotoxic actions of pyrethroid
806 insecticides. *Annu. Rev. Entomol.* 34, 77–96.
807 <https://doi.org/10.1146/annurev.en.34.010189.000453>
808 Storck, V., Karpouzas, D.G., Martin-Laurent, F., 2017. Towards a better pesticide
809 policy for the European Union. *Sci. Total Environ.* 575, 1027–1033.
810 <https://doi.org/10.1016/j.scitotenv.2016.09.167>
811 Stuligross, C., Williams, N.M., 2021. Past insecticide exposure reduces bee
812 reproduction and population growth rate. *Proc. Natl. Acad. Sci. U. S. A.* 118, 1–
813 6. <https://doi.org/10.1073/pnas.2109909118>
814 Tang, F.H.M., Lenzen, M., McBratney, A., Maggi, F., 2021. Risk of pesticide pollution
815 at the global scale. *Nat. Geosci.* 14, 206–210. [https://doi.org/10.1038/s41561-](https://doi.org/10.1038/s41561-021-00712-5)
816 [021-00712-5](https://doi.org/10.1038/s41561-021-00712-5)
817 Tassinde Montaigu, C., Goulson, D., 2020. Identifying agricultural pesticides that may
818 pose a risk for birds. *PeerJ*.
819 Tsvetkov, N., Zayed, A., 2021. Searching beyond the streetlight: Neonicotinoid
820 exposure alters the neurogenomic state of worker honey bees. *Ecol. Evol.* 11,
821 18733–18742. <https://doi.org/10.1002/ece3.8480>
822 Urruty, N., Deveaud, T., Guyomard, H., Boiffin, J., 2016. Impacts of agricultural land
823 use changes on pesticide use in French agriculture. *Eur. J. Agron.* 80, 113–123.
824 <https://doi.org/10.1016/j.eja.2016.07.004>
825 Van Bruggen, A.H.C., He, M.M., Shin, K., Mai, V., Jeong, K.C., Finckh, M.R., Morris,
826 J.G., 2018. Environmental and health effects of the herbicide glyphosate. *Sci.*
827 *Total Environ.* 616–617, 255–268.
828 <https://doi.org/10.1016/j.scitotenv.2017.10.309>
829 Wolska, L., Sagajdakow, A., Kuczyńska, A., Namieśnik, J., 2007. Application of
830 ecotoxicological studies in integrated environmental monitoring: Possibilities and
831 problems. *TrAC - Trends Anal. Chem.* 26, 332–344.
832 <https://doi.org/10.1016/j.trac.2006.11.012>
833 Yang, E.C., Chuang, Y.C., Chen, Y.L., Chang, L.H., 2008. Abnormal foraging
834 behavior induced by sublethal dosage of imidacloprid in the honey bee
835 (Hymenoptera: Apidae). *J. Econ. Entomol.* 101, 1743–1748.
836 <https://doi.org/10.1603/0022-0493-101.6.1743>
837 Zubrod, J.P., Bundschuh, M., Arts, G., Brühl, C.A., Imfeld, G., Knäbel, A.,
838 Payraudeau, S., Rasmussen, J.J., Rohr, J., Scharmüller, A., Smalling, K.,
839 Stehle, S., Schulz, R., Schäfer, R.B., 2019. Fungicides: An Overlooked Pesticide
840 Class? *Environ. Sci. Technol.* 53, 3347–3365.
841 <https://doi.org/10.1021/acs.est.8b04392>

Appendix

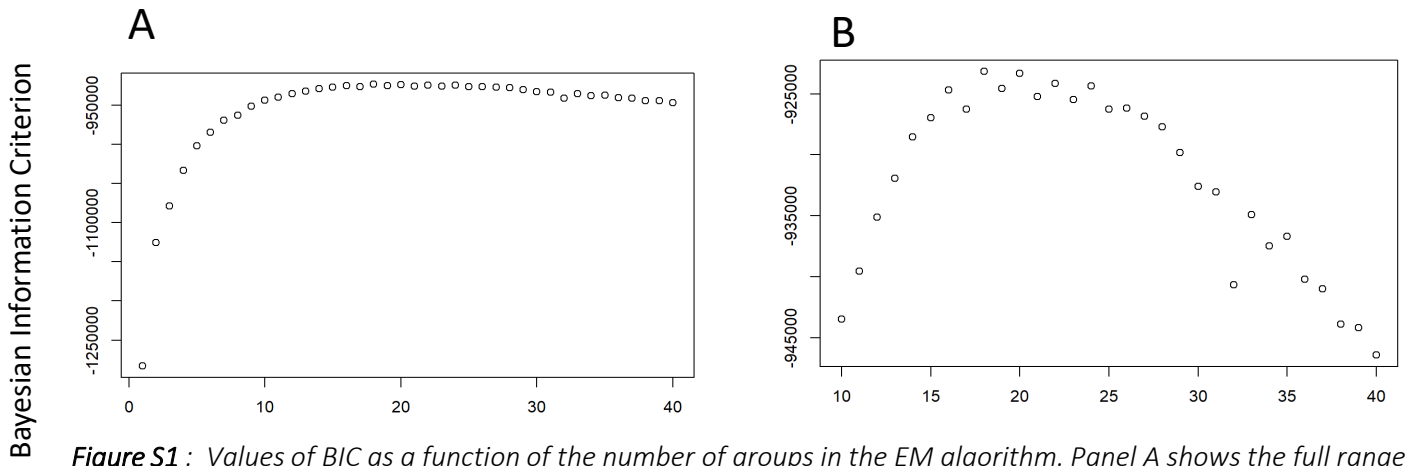


Figure S1 : Values of BIC as a function of the number of groups in the EM algorithm. Panel A shows the full range of number of groups tested (from 1 to 40). Panel B is a closeup around the maximum BIC value

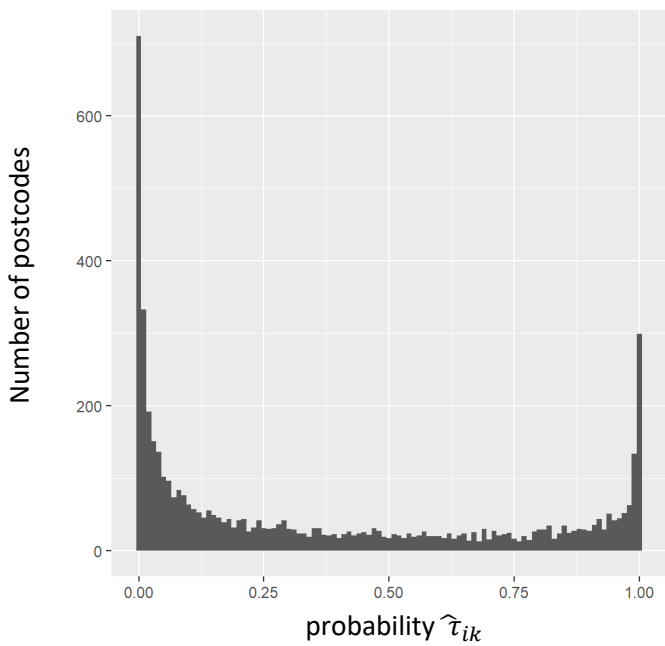


Figure S2: Distribution of the, $\hat{\tau}_{ik}$ probability of a postcode i to be in a group k

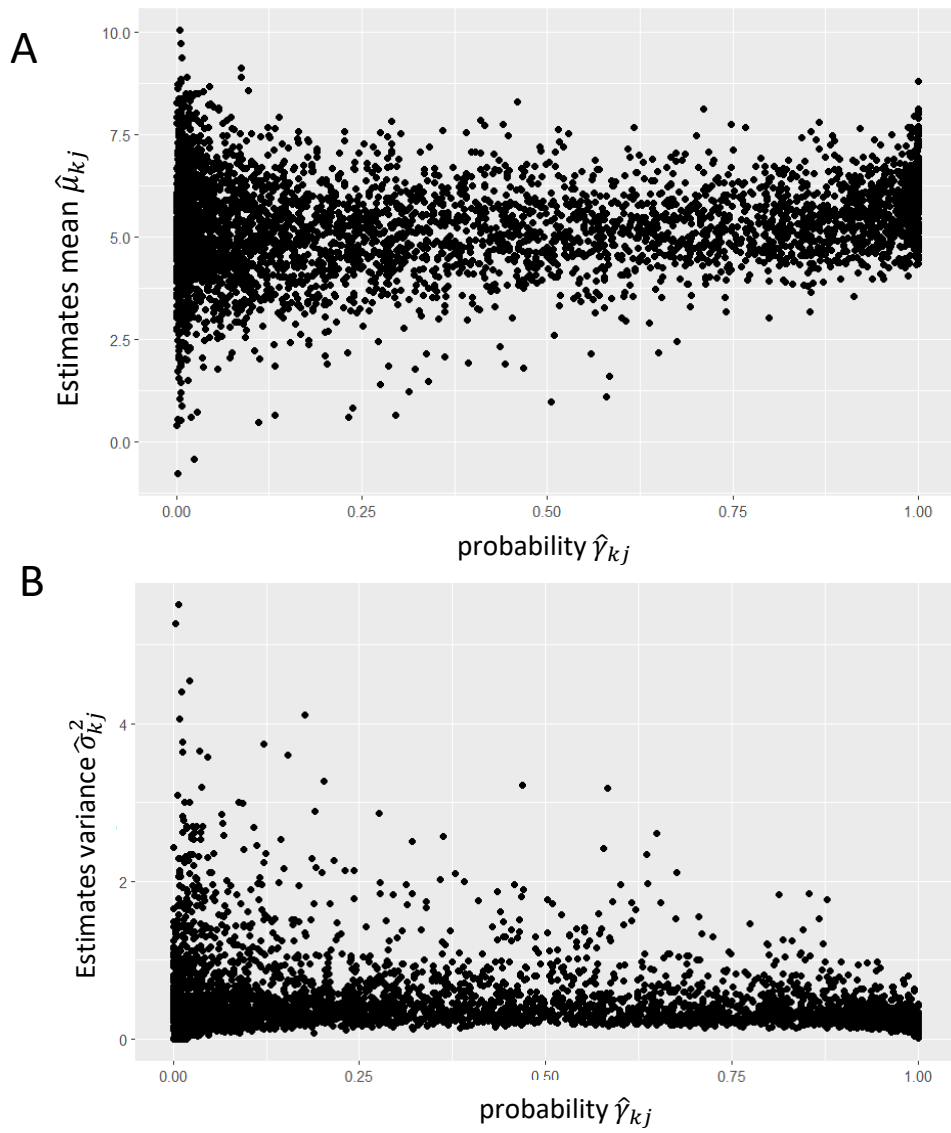


Figure S3: Estimated mean ($\hat{\mu}_{kj}$, panel A) and variance ($\hat{\sigma}_{kj}^2$, panel B) of substance quantities purchased in a group as a function of the probability of a substance j to be in a group $\hat{\gamma}_{kj}$

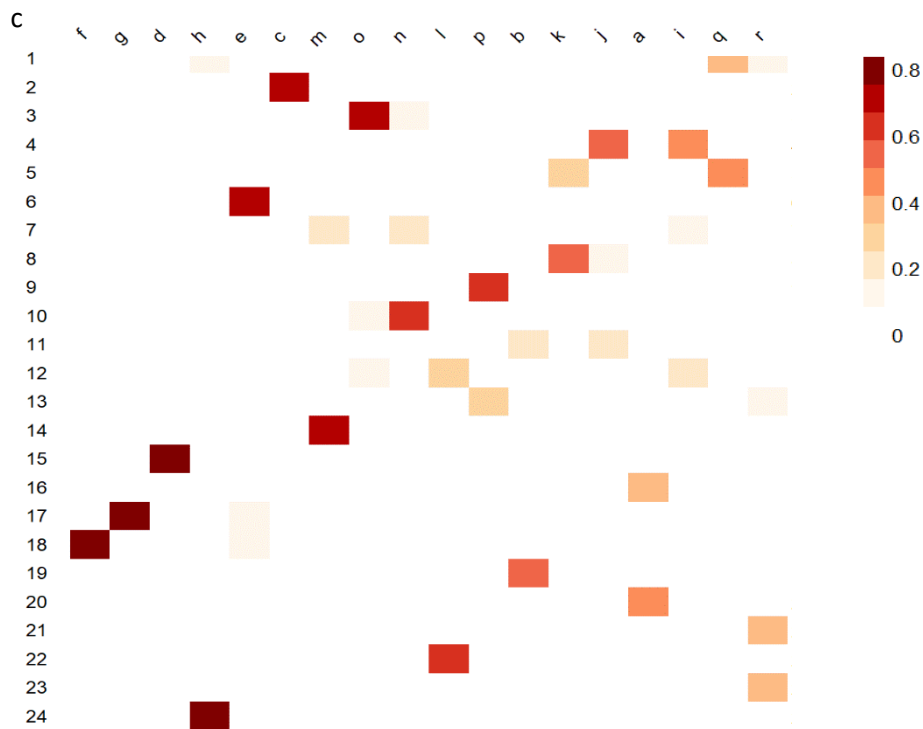
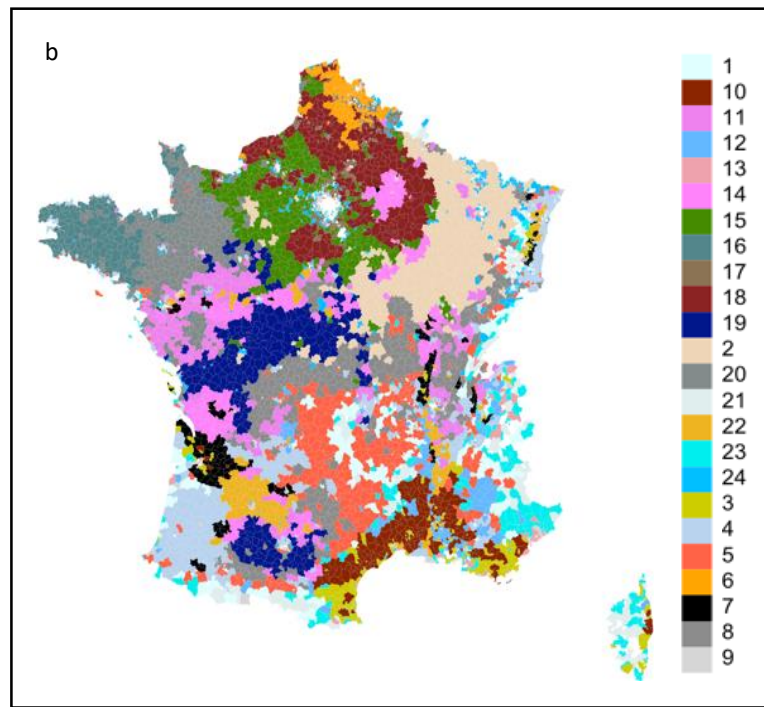
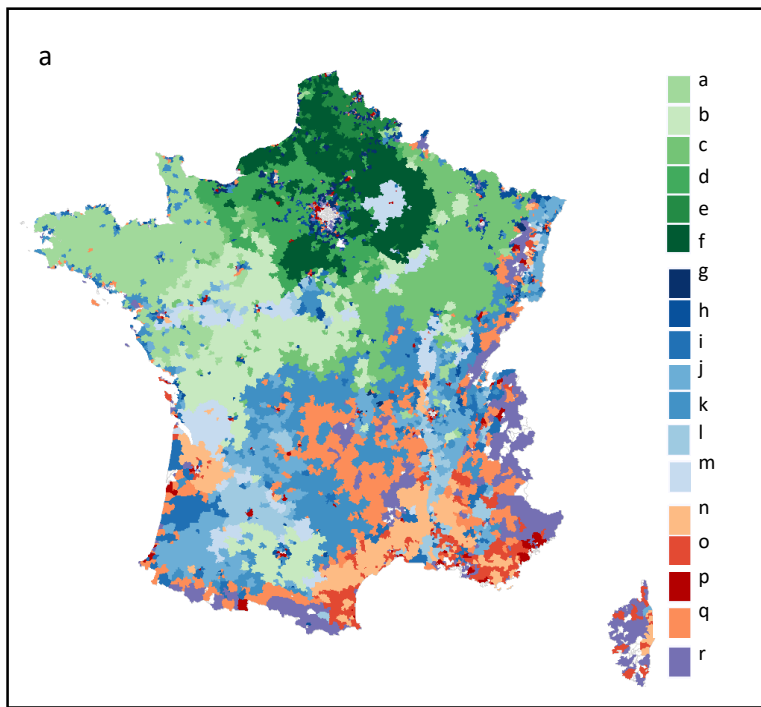


Figure S4: Differences and similarities in the clustering of postcodes produced by the mixture model with only 2017 substance purchase data (a) or 2015-2018 data (b). Postcode within a group share the same colour. Panel c shows proximity of the 2017 groups with 2015-2018 groups on a heatmap, expressed as the percentage of postcodes from 2017 groups that were found in the various 2015-2018 groups. The graph should be read vertically: for example, 2017 group r is split mostly into 2015-2018 groups 21(38%) and 23 (37%), with a small fraction of postcodes also found in 2015-2018 groups 1 (12%) and 13 (10%). In contrast, virtually all postcodes of 2017 group f are found in 2015-2018 group 18.

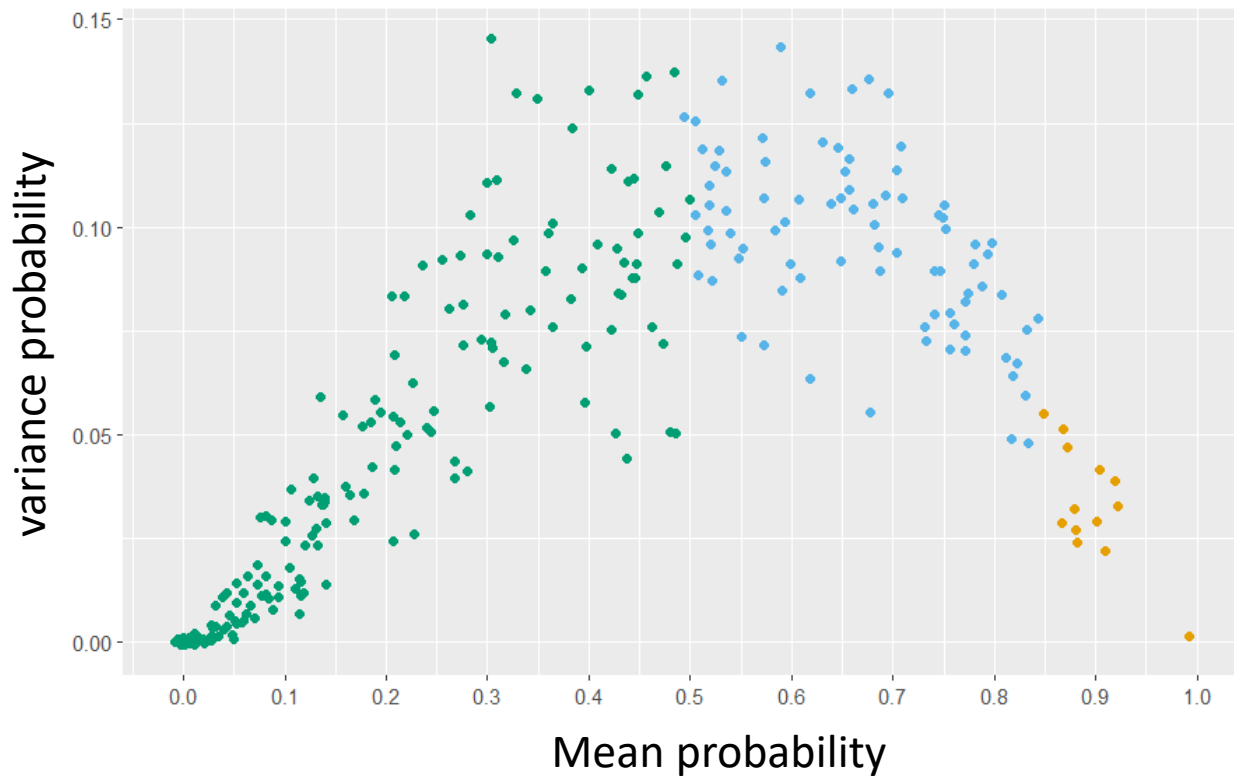
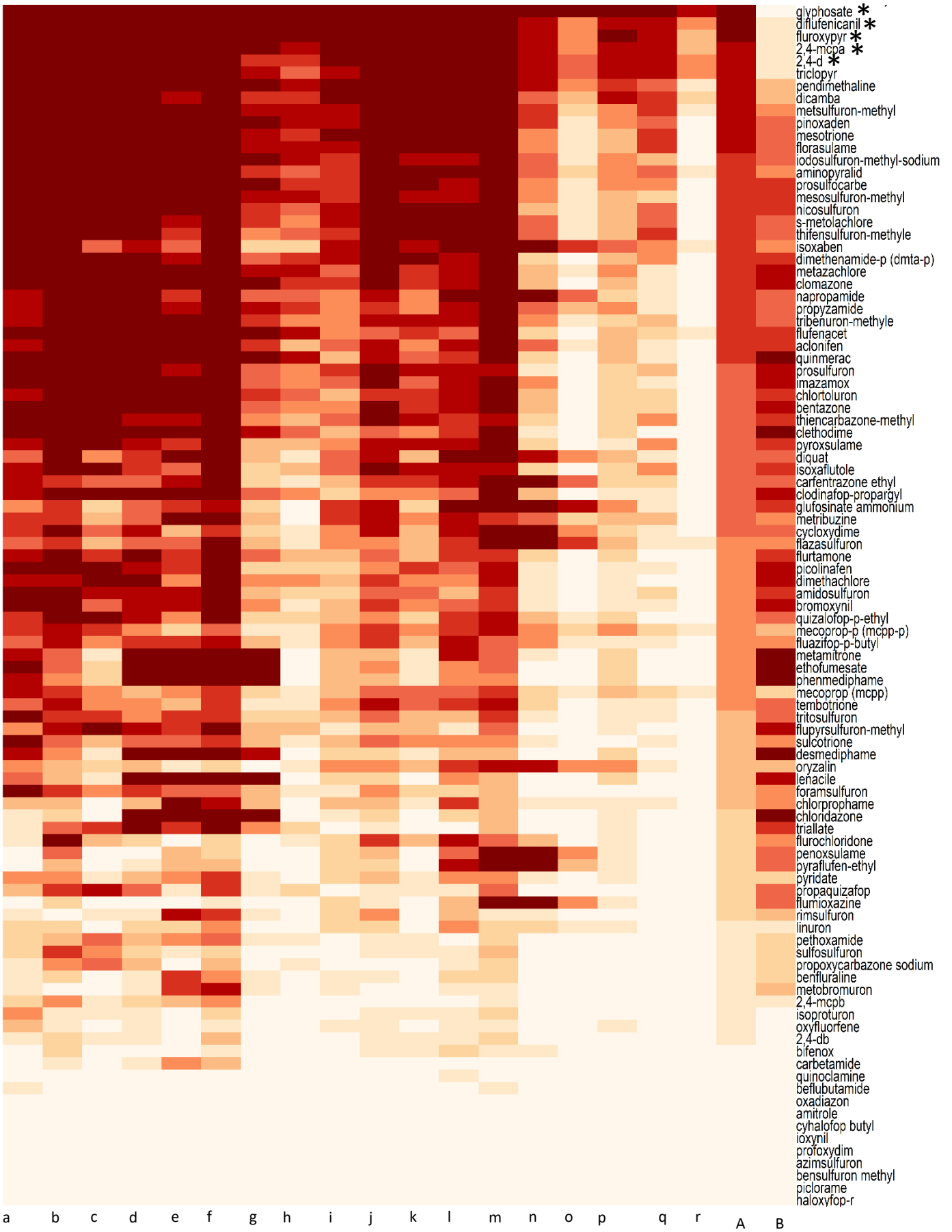
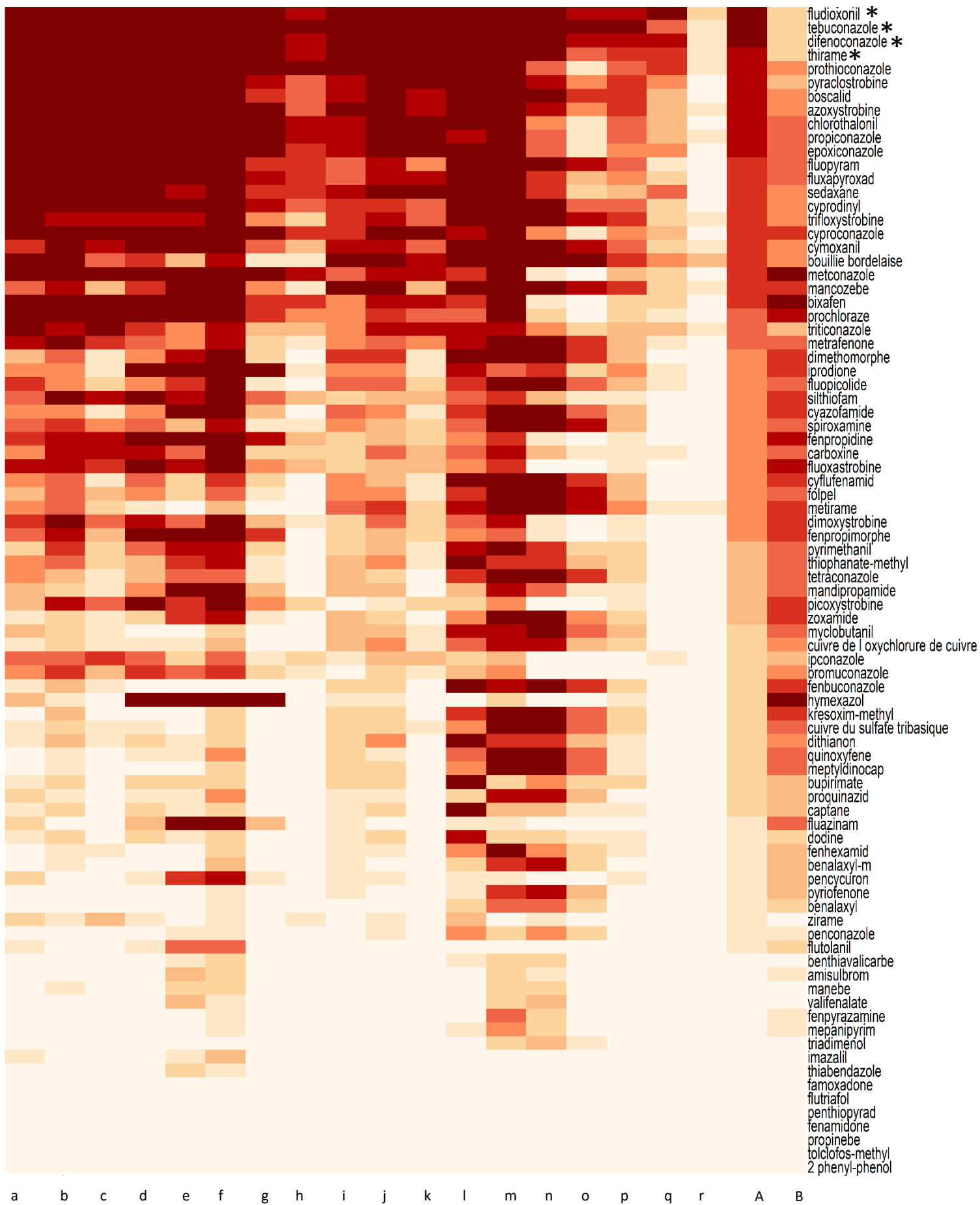


Figure S5: Variance in probabilities of substances to be in a group as a function of their mean probability to be in a group. Colours indicate core (orange), discriminant (blue) and other (green) substances.

Herbicides



Fungicides

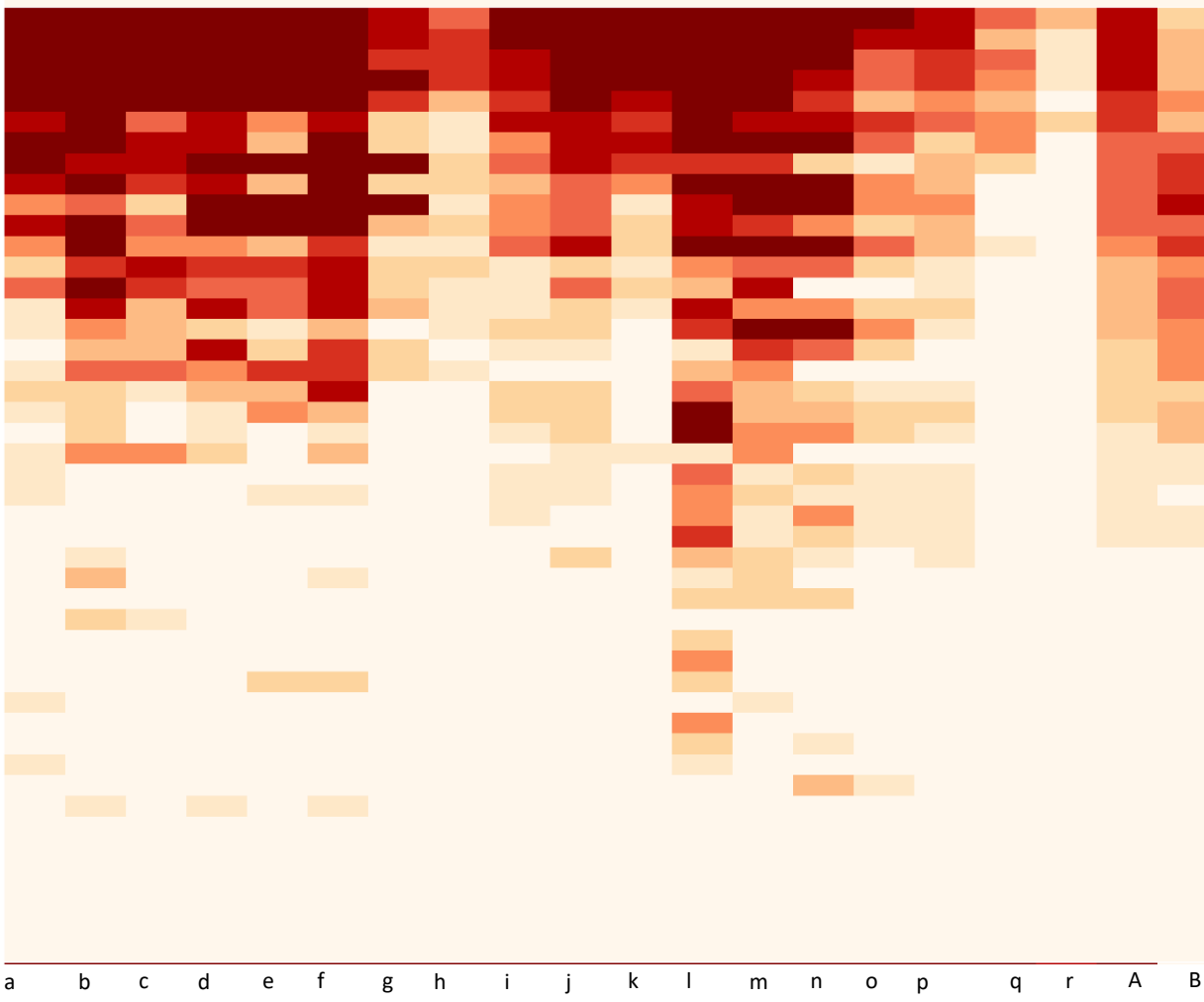


Other targets



- ethephon
- mepiquat-chlorure
- trinexapac-ethyl
- prohexadione-calcium
- isoxadifen-ethyl
- paclobutrazol
- imazaquine
- acrinathrine
- tebufenpyrad
- acide alpha naphtylacetique (ana)
- 1-dodecanol
- 1-tetradecanol
- fenazaquin
- acibenzolar-s-methyl
- decane-1-ol
- hypochlorite de sodium
- 1,3-dichloropropene
- alpha naphtyl acetamide (nad)
- forchlorfenuron
- bromadiolone
- cintofen

Insecticides



- deltamethrine *
- lambda-cyhalothrine *
- cypermethrine *
- imidaclopride *
- thiaclopride
- acetamipride
- chlorpyrifos-methyl
- tefluthrine
- tau-fluvalinate
- thiamethoxam
- pyrimicarbe
- chlorantraniliprole
- alphamethrine
- zetacypermethrine
- phosmet
- indoxacarbe
- betacyfluthrine
- esfenvalerate
- pymetrozine
- spirotetramat
- fenoxycarbe
- etofenprox
- hexythiazox
- bifenazate
- spinetoram
- pyriproxyfene
- diflubenzuron
- methiocarbe
- etoxazole
- pyrimiphos-methyl
- fosthiazate
- spirodiclofen
- clothianidine
- chlorpyrifos-ethyl
- acequinocyl
- tebufenozide
- ethoprophos
- methoxyfenozide
- gamma-cyhalothrine
- phosphure d aluminium
- oxamyl
- phosphure de magnesium
- cyfluthrine
- pyridabene
- malathion
- fluorure de sulfuryle

Figure S8: Heatmap of probability γ_{kj} , that substance j is used in postcode k . Groups were obtained from a mixture models optimised by maximum likelihood with an iterative method: Expectation Maximisation. Groups were ordered by similar composition of substance purchases. Substances belong to four categories: herbicides, fungicides, insecticides and other targets. Within each category of substances, substances were ordered in increasing number of groups in which they were used. Column A corresponds to the mean probability of use and column B corresponds to the scaled (0,1) variance in probability of use across groups. Asterisks (*) highlight core substances.

Table S1: Complete list of substance's targets name associated with the "other" category

Substance's targets	Number of substances
Acaricide	5
Algicide	1
Attractant	2
Bactericide	1
Nematicide	1
Plant activator	1
Plant growth regulator	11
Rodenticide	2
Safener	1

Table S2 : Correspondence table between crop categories from the Land Parcel Identification System (LPIS) and aggregated crop categories used in our analyses

CATEGORY FROM RPG	CATEGORY USED
Common wheat	Cereals
Barley	Cereals
Other cereals	Cereals
Miscellaneous	Miscellaneous
Arboriculture	Orchard
Olive tree	Orchard
Fruit Orchard	Orchard
Legumes/Flowers	Legumes/Flowers
Maize	Maize
Nut	Nut
Other oil crops	Other oil crops
Protein crop	Protein crop
Rapeseed oil	Rapeseed oil
Sunflower	Sunflower
Grapevine	Grapevine