

1 **Title:** How far away do you keep your equines? Estimation of the equine population's spatial distribution in  
2 France

3

4 **Author names and affiliations:** Halifa FARCHATI <sup>a,c,d</sup>, Benoit DURAND <sup>b</sup>, Maud MARSOT <sup>b</sup>, David GARON <sup>c</sup>,  
5 Jackie TAPPREST <sup>a</sup>, Carole SALA <sup>d</sup>

6<sup>a</sup> Laboratory for Animal Health in Normandy, Physiopathology and Epidemiology of Equine Diseases Unit,  
7 French Agency for Food, Environmental and Occupational Health & Safety (ANSES) F14430 Goustranville,  
8 France

9<sup>b</sup> Epidemiology Unit, Laboratory for Animal Health, French Agency for Food, Environmental and Occupational  
10 Health and Safety (ANSES), 94700, Maisons-Alfort, France

11<sup>c</sup> Normandy University, UNICAEN, Centre F. Baclesse, UR ABTE EA 4651, 14000 Caen-France

12<sup>d</sup> University of Lyon- Epidemiology and Support to Surveillance Unit, French Agency for Food, Environmental  
13 and Occupational Health & Safety (ANSES), 31, avenue Tony Garnier, F69364 Lyon Cedex 07, France

14

15 **Keywords:** Spatial distribution, Equine, Owner, Bayesian model, Surveillance

16

## 17 **Acknowledgments**

18 We would like to thank the French horse and riding institute (IFCE) and the French Agency for Food,  
19 Environmental and Occupational Health & Safety (ANSES) for their financial support. We are also very  
20 grateful to the data providers: IFCE, Le Trot, France Galop, the French ministry for agriculture and food. We  
21 would also like to thank the owners and keepers who answered our surveys. We need to thank various  
22 people having contributed in different ways to the completion of this work: Jean-Philippe Amat, Marie  
23 Delerue, Mathilde Dhollande, Romain Dhollande, Xavier Dornier, Joelle Dop, Arnaud Duluard, Bénédicte  
24 Ferry, Giovanni Lo Lacono, Carole Margelidon, Aurélie Merlin, Marion Wujek Moreau, Julie Schneider,  
25 Caroline Teyssier, and Sonia Wittreck.

26

1 **Title:** How far away do you keep your equines? Estimation of the equine population's spatial distribution in

2 France

3

4

5 **Keywords:** Spatial distribution, Equine, Owner, Bayesian model, Surveillance

6

## 7 Abstract

8 It is essential to have an accurate picture of the spatial distribution of equines to be able to monitor equine  
9 health events effectively. In France, this information is only available for certain categories of live equines  
10 kept in professional structures and for dead equines removed by renderers. This limits the surveillance,  
11 prevention and control methods able to be used to prevent the spread of equine diseases. Our study aimed  
12 to provide a realistic estimate of the spatial distribution of the French equine population at the detailed scale  
13 of the French *commune* (France's smallest administrative unit). For this purpose, we adapted the Bayesian  
14 method used by Lo Lacono et al., based on the distance between the owner's location and the location of  
15 his/her equines, and on the percentage of urban coverage. To assess whether the location of dead equines  
16 could be representative of the location of live equines, the distribution of distances between equines and  
17 owners was calculated from a sample of live equines on the one hand, and a sample of dead equines on the  
18 other, both accurately located. We also tested two different assignment methods for locating equines:  
19 Method 1 assigned to each owner a single holding *commune*, while Method 2 allowed more variability in  
20 holding *communes* for owners associated with multiple equines. A marked difference was observed between  
21 Methods 1 and 2 regardless of the sample used, with only 2.4% and 4.3% respectively of the *communes*  
22 having the same number of equines. Conversely, little difference was observed in the results whether the  
23 live or dead equine sample was used, with approximately 45% of the *communes* having the same number of  
24 equines. Regarding differences in spatial distribution, Method 1 based on the live equine sample estimated  
25 higher local densities of equines without considering urban areas. In contrast, Method 2 provided more  
26 dispersed maps, with low densities in the densest urban areas. In conclusion, dead equines appeared to be  
27 representative of live equines and some of our estimates are consistent with the information collected by  
28 the French horse and riding institute (IFCE). These results now have to be compared with field data to test  
29 their relevance.

30

## 31 1 Introduction

32 Data on geographic location and interspecies contacts are often considered key factors in developing  
33 accurate predictive models of pathogen spread or establishing effective and appropriate methods for the  
34 surveillance, prevention or control of animal diseases (House and Keeling, 2008a, b; M. J Keeling et al., 2010).

35 The major influence of the population's spatial structure on the spread of infectious diseases is well known in  
36 epidemiology, whether for humans or domestic/wild animals (Ferguson et al., 2005; Ferguson et al., 2001;  
37 Gudelj and K.A.J., 2004; Keeling et al., 2001; Swinton et al., 1998). Depending on the different ways that  
38 diseases can spread and the density of animal populations, a disease can spread locally then extend either  
39 slowly or quickly over very long distances, especially when sick animals are moved around.

40 The effectiveness and success of measures to control the speed and spread of disease depend on the  
41 existence of high quality spatial and epidemic data, in addition to the traceability of animal movements. The  
42 example of the equine influenza outbreak in Australia in 2007, when nearly 4,500 equine premises were  
43 infected in less than two months (Brendan et al., 2009; Callinan, 2007) demonstrated how the lack of  
44 accurate spatial data on the equine population and its movements can greatly complicate the  
45 implementation of surveillance methods and control measures to combat and control diseases. It also  
46 showed how traceability information is helpful in rapidly implementing effective control measures (Garner et  
47 al., 2010). More recently, in the South of France, management of the nine outbreaks of equine infectious  
48 anemia virus (EIAV) occurring between 2017 and 2019 systematically required use of a specific  
49 epidemiological survey to identify neighboring structures and test equines that had potentially been in  
50 contact with equines already identified as being sick (Amat et al., 2018; Merlin et al., 2021). In the absence of  
51 information on precise location of French equines, these investigations are time consuming and delay both  
52 the identification of sick animals and implementation of control measures.

53 This example shows that the traceability of equines in France needs to be improved in order to effectively  
54 take measures against diseases. Currently, only the movements of racehorses in training have to be traced  
55 for anti-doping tests. For all other equines, current French regulations only require registration of equine  
56 premises in the SIRE database, which is France's national identification system. They do not, however require

57 the registration of the equines kept there. The spatial distribution of French equines is currently estimated  
58 from surveys of professional structures, and identification checks carried out by the French horse and riding  
59 institute (IFCE) (Ecus-Ifce). However, the representation obtained is partial, as the majority of equines are  
60 kept in non-professional structures, often on the horse owners' premises.

61 All around the world, surveillance plan strategies and outbreak management efforts in the equine sector are  
62 hampered by the lack of information on the precise location of equines and their movements. The recent  
63 enforcement in April 2021 of the European Union animal health law (EU's AHL) on transmissible animal  
64 diseases (Regulation (EU) 2016/429 (Anonyme, 2016)) should improve the traceability of equines, imposing  
65 mandatory notification of equines' usual location. Nevertheless, the operational implementation of this  
66 measure should take many years, even in countries with near-complete equine identification and a  
67 centralized identification database. The regulatory registration of the location of all equines is not expected  
68 to be effective for some time.

69 In the meantime, it would be very useful to managers, epidemiologists, and researchers to have realistic  
70 estimates of equine densities at detailed spatial scales in order to improve checks, surveillance, and equine  
71 disease modeling. With this in mind, Giovanni Lo Lacono et al. (Lo lacono et al., 2013) implemented in their  
72 research a Bayesian method to estimate the spatial distribution of horses in Great Britain, based on the  
73 distribution of owner-equine distance. They used a sample of owner-equine pairs that were both precisely  
74 located to estimate the probable distance between an owner and his/her equine. Additionally, an  
75 environmental component (proportion of urban coverage) modulated the location probability so as to assign  
76 to the equine the most plausible location(s). The final result was a theoretical map of the British equine  
77 population.

78 Adapting the approach of Giovanni Lo Lacono (Lo Lacono et al., 2013), our objective was to estimate the  
79 spatial distribution of the equine population in metropolitan France at the *commune* scale (a *commune* being  
80 the smallest French administrative unit). In our case, for comparison purposes and to assess whether the  
81 location of dead equines can be representative of that of live equines, two samples of live equines and dead  
82 equines were used like training datasets. To assign a location to live equines registered in the central

83 database from the SIRE database, two different assignment methods were realized. The first one assigned a  
84 single holding *commune* for all equines of each owner and the second method allowed one, two or three  
85 holding *communes* of owners, depending on the number of linked equines (one *commune* for one equine,  
86 two *communes* for two equines, or three *communes* for three or more equines). Different data sources and  
87 hypotheses that will be detailed later were used for this purpose, to obtain multiple results that are  
88 thereafter discussed. We will discuss in particular the value of combining multiple data sources and the  
89 contribution of using two samples and two methods.

90

## 91 2 Materials and Methods

### 92 2.1 Data sources and information

93 Several data sources were used to train the model and estimate the spatial distribution of the French equine  
94 population.

#### 95 2.1.1 Training datasets

96 The model was calibrated using information from three datasets:

- 97 - A survey sample dataset of 2,353 precisely located live equine-owner pairs. This sample came from a  
98 2019 national survey of owners and keepers registered in the SIRE database (Farchati et al., 2020).  
99 The main characteristics of this sample are available in Table 1. The sample was obtained by linking  
100 each owner with one or more *communes* of equine ownership, with as many pairs as different  
101 *communes* of ownership for each owner. Racehorses in training were excluded from this dataset  
102 because they were considered differently in this study. This dataset is referred to as the live equine  
103 sample in the following.
- 104 - A sample of 13,595 dead equine-owner pairs extracted in April 2019 from the fallen stock data  
105 interchange (FSDI) database, which has been centralizing information on all removed (dead)  
106 animals—animal identification, *commune* of death, and information on the person who requested  
107 the removal (name, surname or company name, *commune* of residence)—since 2011. The pairs  
108 included were those for which i) the equine identification number was traceable in the SIRE  
109 database (n=97,327, i.e. 33% of 298,733 initial records in the extraction of FSDI), ii) there was a high  
110 probability that the applicant was the owner of the dead equine (n=44,771 among the 97,327) and  
111 iii) the applicant had filled out the information on the cadaver online (n=13,595 among the 44,771)  
112 (through ATM<sup>1</sup> services managed by the IFCE). The selection was made by cross-referencing the FSDI  
113 and SIRE databases, and comparing the identification and location information for owners, keepers  
114 and removal applicants. The intention in limiting the sample to owners who used the online system

---

<sup>1</sup> ATM-équidés ANGEE (ATM) is an association created in 2009, following privatization of the public rendering service, to organize cadaver removal. It offers owners negotiated prices for cadaver removal in order to limit illegal burial and provide rendering companies with a guaranteed payment for removals.

115 was to increase the likelihood that the owners' home *communes* as registered in the SIRE database  
116 were up to date (as the recording of information at ATM allows it to update SIRE). This sample of  
117 dead equine-owner pairs is referred to as the dead equine sample in the following and is described  
118 in Table 1.

119 - Data extracted from the CORINE Land Cover (CLC) (Cover) geographic database, to calculate the  
120 proportion of urban coverage for each *commune* in metropolitan France. CLC is a biophysical  
121 inventory of European land use and its changes according to a nomenclature of 44 items. To take  
122 into account the urban/rural land use aspect for equines, we calculated the proportion of  
123 continuous and discontinuous urban coverage relative to the total surface area of each *commune*.

### 124 2.1.2 Mapping datasets

125 In order to obtain realistic equine density maps, we assigned to each equine presumed to be living in  
126 metropolitan France a "holding" *commune* (where it was kept) out of the most probable ones, based on the  
127 model results. An initial identification of the equine presumed to be living in France was thus necessary. The  
128 SIRE database, managed since 1976 by the IFCE, centralizes declarative information on almost all of the  
129 equines living in France (equine identifying number, sex, breed, date of birth, date of death declaration (if  
130 relevant)), along with their owners (owner identifying number, *commune* for the owner's residence, date on  
131 which the equine owner became its owner, and date on which information was last updated) and in parallel  
132 information of keepers and equine premises (keeper identifying number, *commune* for the keeper's  
133 residence, *commune* of the equine premises and the date of the premises were opened and closed (if  
134 applicable)) without specific traceable link between keepers and the equines they keep, and therefore a lack  
135 of centralized equine location information in France. This database has also difficulty in identifying with  
136 certainty whether an equine is still alive or dead (it has been estimated that nearly 30% to 40% of owners  
137 have never reported the death of their dead equine(s) to the IFCE, or reported it long after the death (IFCE,  
138 personal communication)). The living equine population is consequently not well known. Four data sources  
139 were cross-checked to constitute the set of equine-owner pairs for which the equine and its owner were  
140 presumed to be living in Metropolitan France:



- 141 - Data were extracted from the SIRE database in late March 2019 as a basis for selecting appropriate  
142 data. This dataset included 3,031,446 equines owned by 790,051 owners located in 35,359  
143 *communes*.
- 144 - A dataset for 2015-2019 extracted from Europe's Trade Control and Expert System (TRACES) was  
145 used to identify equines that were no longer living in France because they had been exported.  
146 TRACES is a computerized system for international trade in live animals, animal products and certain  
147 non-animal products. It records, among other things, the SIRE identifier, transponder or  
148 international number and the date of departure of exported equines. The dataset extracted  
149 concerned 31,336 equines.
- 150 - The FSDI database was used to identify the dead equines collected by rendering companies that had  
151 not been declared dead in the SIRE database. The whole database since its creation was used i.e.,  
152 298,733 removals.
- 153 - A dataset of racehorses with the exact location of the horse corresponding to the commune in which  
154 it was being trained. This set included 37,776 trotters recorded as being in training at least once  
155 between August 2018 and August 2019, and 13,290 gallopers recorded as being in training at least  
156 once between July 2020 and October 2020. The dataset was compiled from monthly extractions of  
157 the databases for the two parent companies of horse racing in France: Le Trot (for trotters) and  
158 France Galop (for gallopers). For each horse in pre-training or training, each company records the  
159 SIRE identifier, the *commune* of the horse's location (premises where it is kept or where it is trained),  
160 and date of its entry/exit. Because these equines were accurately located, and the location of the  
161 owner(s) does not necessarily correspond to the location of the equine, due to the need for a  
162 specific (and rare) training structure, we did not apply the location algorithm to this population.  
163 However, we used the racehorse dataset to remove horses in training from the SIRE dataset and live  
164 equine sample.

165 From these four datasets, we obtained:

- 166 - A dataset of 1,068,142 equine-owner pairs corresponding to 1,068,142 equines presumed to be  
 167 alive and belonging to 397,039 owners living in 32,837 *communes* in metropolitan France and  
 168 resulting from the SIRE extraction after removing dead animals and cross-checking with the other  
 169 data sources. The details of the different steps are shown in Figure 1. The dataset obtained is  
 170 referred to as the equine dataset in the following.
- 171 - A dataset of 44,768 racehorses in training (12,785 gallopers and 31,983 trotters) along with their  
 172 location, resulting from the removal of dead equines then the selection of a location for each equine  
 173 in metropolitan France on the date closest to April 1st, 2019 (date on which the data were extracted  
 174 from the SIRE database). This dataset is referred to as the racehorse dataset in the following.

## 175 2.2 Modeling

176 Using the approach of Giovanni Lo Lacono (Lo Lacono et al., 2013), we estimated the probability  $P(r/r_0)$  that  
 177 an equine is kept in *commune* of centroid  $r = (x, y)$  knowing that its owner resides in *commune* of centroid  
 178  $r_0 = (x_0, y_0)$ .  $P(r/r_0)$  was expressed by the following equation:

$$179 \quad P(r / r_0) = \frac{\text{equines in the owner's commune}}{\delta(|r - r_0|)P_0(r_0)} + \frac{\text{equines in a different commune from the owner}}{[1 - \delta(|r - r_0|)][1 - P_0(r_0)]P_{sep}(r, r_0)} \quad (1)$$

180 with  $P_0(r_0)$  the probability that the equine and owner are in the same *commune*,  $|r - r_0| =$   
 181  $\sqrt{(x - x_0)^2 + (y - y_0)^2}$  the Euclidean distance between  $r$  and  $r_0$ ,  $P_{sep}(r, r_0)$  the probability that the equine  
 182 and the owner are localized in *communes* remote from  $|r - r_0| \neq 0$ , and  $\delta$  a Dirac function equal to 1 if the  
 183 owner and the equine are in the same *commune*, and equal to 0 otherwise.

184 The probability  $P_0(r_0)$  was related only to the urban coverage of the owners' *commune* of residence and  
 185 obtained using the following equation:

$$186 \quad P_0(r_0) = \begin{cases} n_1 * \frac{1}{u(r_0) * \sigma\sqrt{2\pi}} * e^{\left(-\frac{[\ln(u(r_0)) - \mu]}{2\sigma^2}\right)} & \text{if } u(r_0) \neq 0 \\ n_2 * \mathfrak{I} & \text{if } u(r_0) = 0 \end{cases} \quad (2)$$

187 with  $u(r_0)$  the proportion of urban coverage of the owner's location,  $\mu$  (the expectation) and  $\sigma$  (the standard  
 188 deviation) the parameters of a lognormal distribution to be determined (Figure S-1 of supplementary

189 material),  $n_1$  and  $n_2$  the normalization constants that ensure that  $\int_1^N P_0(r_0) = \alpha$  (N being the number of  
 190 potential *communes* less one in metropolitan France and  $\alpha$  the proportion of equines kept in the same  
 191 location as the owner) and  $\zeta$  the proportion of owners keeping their equines in their place of residence and  
 192 with zero urban coverage.

193 Furthermore, both the live and dead equine samples showed that the distribution of distances between the  
 194 owners' location and their equine was well described by a power law (Figure 2). The probability that the  
 195 owner and the equine were distant from  $|r - r_0|$  was modeled by:

$$196 \quad P_{sep}(r, r_0) = A * \|r - r_0\|^{-\beta} * P_0(r) \quad (3)$$

197 with  $|r - r_0|^{-\beta}$  the probability that the owner and the equine are distant by  $|r - r_0|$ ,  $A$  a constant set by the  
 198 constraint  $\int_{D-r_0} P_{sep}(r, r_0) = 1$  (with  $D - r_0$  the set of all potential *communes* in France except the one of  
 199 the owner's choice) and  $P_0(r)$  defined in the equation (2).

### 200 2.3 Estimation of the parameters

201 Parameters  $\sigma$  and  $\mu$  of the lognormal distribution of the distances in equation (2) were estimated using the  
 202 least squares adjustment method. Parameter  $\beta$  of equation (3) was estimated using a Bayesian approach,  
 203 using Markov chain Monte Carlo (MCMC) methods to generate a distribution for the parameter. The prior  
 204 considered was a non-informative uniform distribution and the likelihood was given by:

$$205 \quad P(\text{data} | \beta) = \prod_{i=1}^{N_{obs}} P(r_i/r_{0i}) \quad (4)$$

206 where  $N_{obs}$  was the number of owner-equine pairs in the sample, and  $r_{0i}$  and  $r_i$  the respective *communes*  
 207 of the owner and his/her equine(s) for the  $i^{\text{th}}$  pair.

208 We used the MCMCpack R package (Martin et al., 2011) to generate a chain of 6,000 iterations with a burn-  
 209 in of 1,000 iterations. The chain's convergence was monitored by Geweke's convergence diagnostic  
 210 (`geweke.diag` (Geweke, 1991)).

211 We estimated two distributions of parameters  $\beta$  and two values for each of parameters  $\sigma$ ,  $\mu$ , from both the  
212 live and the dead equine samples.

#### 213 2.4 Assignment of holding *commune(s)* to owners and equines

214 For each of the 1,068,142 owner-equine pairs from the equine dataset, we assigned one or more probable  
215 holding *communes* for equines using the previously calculated parameters and the urban coverage of the  
216 owners' *commune* of residence. The data (owner-equine pairs) did not allow us to quantify exactly the  
217 number of different locations for the equines of the same owner, however, the survey showed us that  
218 almost (83%) of the owners hosted their equines in one *commune*, 12% in two *communes* and the rest in 3  
219 *communes* (Farchati et al., 2021). Thus, we tested two different assignment approaches. In the first one, we  
220 assigned each owner a single holding *commune* regardless of the number of equines linked with that owner,  
221 i.e. the holding *commune* selected was the same for all the equines paired with the same owner. The second  
222 approach allowed more variability in the holding *communes* for owners linked with more than one equine.  
223 For them, we assigned up to two or three holding *communes*, depending on the number of linked equines  
224 (two *communes* for two equines, or three *communes* for three or more equines).

225 The steps were similar in both methods: for each owner, located in *commune* C, we randomly sampled a  
226 commune. This latter could be C (same *commune*) with the probability  $P_0(C)$  or C' (different to C) with the  
227 probability  $(1 - P_0(C)) * P_{sep}(C, C')$ . In Method 1, this draw was directed once for each owner, and for  
228 Method 2, once for owners with one equine, twice for owners with two equines and three times for owners  
229 with three or more equines. Thus, the equines of each owner were randomly assigned to the selected  
230 *communes*. More detailed steps are available in supplementary material Method 1 and Method 2.

#### 231 2.5 Estimation of the number of equines per *commune*

232 For each owner, we applied both methods and used the two distributions of the parameters obtained from  
233 both the live and the dead equine sample. A distribution of their equines locations was estimated using the  
234 MCMC method. Then, the procedure of drawing lots for a location (*commune*) described above was used to  
235 calculate a realization of this distribution (i.e. a random draw of a *commune* in the distribution of locations):  
236 it was therefore necessary to make repetitions to add variability. We performed 100 iterations for each

237 method and parameter set. Finally, the number of equines assigned per *commune* for each method and  
238 parameter set was obtained by averaging the numbers obtained over the 100 iterations and adding the  
239 number of racehorses from the racehorse sample information.

240

## 241 **3 Results**

### 242 **3.1 Characteristics of the samples and datasets**

243 A description of the live equine sample (n=2,353), the dead equine sample (n=13,595), and the equine  
244 dataset (n=1,068,142) is available in Table 1. The distribution of sex among the dead equines was almost  
245 identical to those of both the live equine sample and the equine dataset, with a slight under-representation  
246 of males and over-representation of females in the dead equine sample. Saddle horses were over-  
247 represented in the live equine sample, while draft horses were under-represented in both the live and dead  
248 equine samples. The age distribution differed in the dead equine sample, with an overrepresentation of old  
249 equines (over 20 years of age) and very young equines (less than two years old). The number of equines per  
250 owner differed between the three datasets, with more owners linked to more than one equine in the dead  
251 equine sample (74% vs. 59% in the live equine sample, and 36% in the equine dataset).

252 The proportion of owners who kept their equines in their *commune* of residence decreased as the  
253 percentage of the *commune's* urban coverage increased (Figure 2A). In addition, the distribution of the  
254 percentage of urban coverage was the same in both samples (live and dead equines, Wilcoxon test,  $p > 0.05$ ).  
255 Most of the owners kept their equines in their commune of residence, but when they did keep them in a  
256 different *commune*, it was less than 30 km away (Figure 2B), with no difference in the distance distributions  
257 between the live and dead equine samples (Wilcoxon test,  $p > 0.05$ ).

### 258 **3.2 National estimation of the spatial distribution of equines in France**

259 The estimations obtained for  $\beta$  differed from 0.11 (95% credibility interval: 0.11-0.12) for the estimation  
260 resulting from the live equine sample to 0.05 (95%CI: 0.05-0.06) for the estimation from the dead equine

261 sample (supplementary material Figure S-1). However, the estimations obtained for  $\sigma$  and  $\mu$  were quite  
262 similar with  $\mu = -3.06$  and  $\sigma = 1.07$  (supplementary material Figure S-2).

263 The average number of equines per *commune* varied greatly between *communes* depending on the method  
264 and sample used to estimate the distribution parameters (Table 2, Figure 3).

265 A big difference between assignment methods 1 and 2 was observed on the live and dead equine samples  
266 (Table 3; Figure 3A, B). Using the live equine sample, only 2.4% of the 32,837 *communes* had the same  
267 number of equines with both methods (Figure 3A). Method 1 assigned more equines in 66 % of the  
268 *communes*, with a mean difference of 28.9 equines and a median difference of 13 [IQR 5-31] equines. This  
269 indicates an asymmetric distribution of differences and large differences in some *communes* (maximum  
270 difference of 10.823 equines). In the 32% of *communes* where Method 2 gave a higher number of equines,  
271 the difference was more marked, with 75% of *communes* having more than 22 equines minimum compared  
272 with Method 1. Results were similar when using the dead equine sample, with 4.3% of *communes* having the  
273 same number of equines, and 73% having more equines with Method 1, though the difference is smaller  
274 than that observed with the live equine sample (median of 7 [IQR 2-17] and mean of 16.4) (Figure 3B). The  
275 distribution of the difference when Method 2 assigned more equines was quite similar to that observed with  
276 the live equine sample.

277 On the other hand, little difference was observed in the results according to the sample used when the same  
278 method was applied, with around 45% of *communes* having the same number of equines (Table 3; Figure 3C,  
279 D) and less variability with Method 2. This limited variability was confirmed by both the small differences  
280 observed in the number of equines, whether the live or dead equine sample was used, and the near-  
281 symmetric distributions of these differences (whether positive or negative) (median= 2/-2 [IQR 1-4] / [IQR 1-  
282 3 ] and mean =2.9/-2.9) (Figure 3D). With Method 1, greater variability was observed between the two  
283 samples. When using the live equine sample, Method 1 assigned more equines to a *commune*, the average  
284 difference being 36.4 equines with a maximum of 10,822 (Figure 3C). Negative differences were more  
285 limited, with an average and maximum difference of 21.4 and 37 equines respectively.

286 Regarding differences in spatial distribution, Method 1 resulted in higher local equine densities and did not  
287 seem sensitive to urban areas, where estimated equine densities were high (Figure 4A, C). The results were  
288 somewhat improved with the dead equine sample, which factored in urban areas better (Figure 4C). In  
289 contrast, Method 2 resulted in a more dispersed picture of equines, with low densities in most dense urban  
290 areas regardless of the sample used (Figure 4B, D). At a larger scale (*département* or *région*), regardless of  
291 the method or sample used, the estimates showed areas with higher densities in Normandy, Brittany, Pays-  
292 de-la-Loire, Bouches-du Rhône, Pyrénées Atlantiques and a lower density in the Massif Central (Figure 4,  
293 Figure S-4). Finally, the addition of racehorse data to the estimates did not change the estimated spatial  
294 distributions, the areas of highest density in our estimates being those with the most racehorses  
295 (supplementary material Figure S-3).

## 296 4 Discussion

297 To our knowledge, this study is the first estimation of a national spatial distribution of French equines,  
298 including those kept by non-professionals, at the *commune* level. It supplements the studies conducted each  
299 year by the Economics Department of the IFCE, providing a map of the equine population (mostly kept in  
300 professional structures (Ecus-Ifce)) by *région* (the largest French administrative unit) and/or *département*  
301 (the medium French administrative unit). While promising, our results showed some sensitivity to the  
302 sample and the method used, and some of the spatial estimates obtained were quite unrealistic, with a high  
303 density of equines in highly urban areas. We discuss here the results obtained with regard to the imprecision  
304 of data used, the influence of the hypotheses and the methodology used.

### 305 4.1 Potential effect of data bias on the density estimates

306 Two samples were used as training data for estimating the parameters of the model, including the  
307 distribution of distances between owners and equines. The live equine sample had the same limitations as  
308 the survey itself, already mentioned in previous studies (Farchati et al., 2020, 2021). These limitations were  
309 mainly related to the bias in the representativeness of the respondents in terms of location and type of  
310 owners. Indeed, most owners in the sample kept their equine at home. Urban owners keeping horses on the  
311 outskirts of cities were thus very poorly represented, whereas they are in the majority in certain densely

312 populated urban areas. This probably explains the excessive number of equines obtained in urban areas,  
313 since it was considered that most owners— including those in the city—kept their equine at home. This error  
314 was exacerbated by the fact that Method 1 located all the equines of a single owner in the same *commune*.  
315 For example, 10,828 equines were located in Ile-de-France (Figure S-4) probably due to the presence in the  
316 Paris of the national guard equines. This hypothesis is supported by the fact that the estimate based on the  
317 dead equine sample better factored in the urban area with both methods. The dead equine sample was  
318 more variable than the live equine sample, with a few more distances over 30 km and probably more retired  
319 equines further away from their owners. Nevertheless, this sample probably overrepresented professionals  
320 and multiple owners, as we aggregated data from several years. Statistically, owners with more equines  
321 were more likely to i) be included, ii) use the ATM system for dead equine removal, iii) provide their equine  
322 identification number and iv) update their information in the SIRE database. This is confirmed by the  
323 difference in the number of equines kept, with only 25% of owners of dead equines having only one equine  
324 and 40% having more than six, compared with 41% of owners of live equines having only one equine and  
325 18% having more than six. Additionally, despite close cross-checking of data and information to differentiate  
326 owners from keepers among the people requesting cadaver removal, some keepers (as opposed to owners)  
327 may have been retained in the dead equine sample, participating in the overrepresentation of multiple  
328 owners. This probably explained the very high density of equine estimates in French breeding areas  
329 (Normandy, Brittany, Bouches-du-Rhône and Pyrénées-Atlantique (Figure S-4)) when using Method 1 with  
330 the dead equine sample.

331 Finally, the quality of the equine dataset used to assign the equines may have an impact on spatial estimates  
332 of equines. Indeed, a previous study of SIRE data quality (Farchati et al., 2020) indicated that nearly one-third  
333 of owners did not update their home address when they moved, while the location assignment of equines is  
334 based on this information. However, most of the owners who moved remained in the same *département*  
335 (Farchati et al., 2021), so the equine location error remains spatially localized (the equines distribution on a  
336 *département* scale (scale greater than the *commune*) is more robust). In addition, although we cross-  
337 checked the data, it is likely that many equines in this dataset were actually dead. It is possible that the  
338 probability of death is spatially homogeneously distributed (30-40% of owners did not report the death of



339 their equine), in which case the overestimation of the number of equines is also likely to be evenly  
340 distributed across France.

#### 341 4.2 Effect of the model, hypotheses and assignment methods

342 The main limitation in our adaptation of the work of Lo Lacono was the approximation of the location of  
343 owners and equines at the centroid of the *communes* because the exact (x, y) locations were not available.  
344 This probably had a strong effect on distance distribution estimates because equines and owners in the same  
345 *commune* can be far apart, depending on the size of the *commune*. Similarly, an owner and equine(s) located  
346 in different *communes* may be very close if they are both close to the inter-communal “border”. In the first  
347 case the distance will be underestimated, while in the second case it will be overestimated. The effect of this  
348 approximation on the estimation of model parameters is not clear. The estimates of parameters  $\sigma$  and  $\mu$   
349 were quite similar regardless of the sample used. This similarity could be explained by the similarity of the  
350 two distributions of urban coverage for the owners who kept their equines in their *commune* of residence. In  
351 contrast, the estimates of the  $\beta$  parameter were very different from one sample to another. This difference  
352 could possibly be explained by the use of a Bayesian approach, which was sufficiently sensitive to the sample  
353 size and especially to the fact that its calculation took into account not only the urban coverage of the  
354 owner's *commune* of residence, but also the urban coverage of the equine holding *commune*, which differed  
355 for each owner-equine pair in the sample. When estimated from the dead equine sample, parameter  $\beta$  was  
356 much larger than when estimated from the live equine sample. It appears to be more realistic as fewer  
357 equines were assigned to high density urban areas.

358 Furthermore, the estimation of the spatial distribution of equines was very sensitive to the assignment  
359 method used, despite a first random step to select, for each owner, a holding *commune* from a list of all  
360 plausible holding *communes*. Assignment method 1 assumed that all the equines of a given owner were kept  
361 in the same *commune*. This approach limited the variability of holding *communes*, and exacerbated the  
362 effect of the bias in the samples used to estimate  $\beta$ . The results showed areas of very high local equine  
363 density, including the densest urban areas such as large cities. The bias in the spatial estimation was more  
364 limited when using the dead equine sample, the dataset including more variability. Assignment method 2

365 allowed more variability in the holding *commune* for owners having more than one equine. This approach  
366 resulted in a highly dispersed spatial distribution regardless of the sample used. Interestingly, the high  
367 density observed in large cities with Method 1 when using the live equine sample was no longer observed.  
368 Besides, Method 2 appeared less sensitive to the sample used, as differences between the estimations were  
369 limited (around two equines on average when a difference was in fact observed). The variability probably  
370 counterbalanced the bias in the samples used to estimate the distribution of distances.

## 371 5 Conclusion and prospects

372 Three of the four estimations of the spatial distribution of French equines are consistent with information  
373 collected by the IFCE (Ecus-Ifce). It is probable that the reality lies somewhere between the results obtained  
374 using the two assignment methods. With very little difference between the results obtained using Method 2  
375 with both samples and a more realistic map with Method 1 and the dead equine sample, our results show  
376 that the location of dead equines could be used to estimate the location of live equines. This could be  
377 extremely useful in a disease outbreak situation in order to effectively control the spread prior to  
378 implementation in the field of the EU's Animal Health Law. This work could be a source of inspiration for  
379 many countries affected by the same problematic as ours. An evaluation of the relevance of our estimates is  
380 therefore planned in the near future by confronting our estimates with the surveys carried out in the field to  
381 check the traceability of equines in France, thanks to the censuses carried out by the IFCE.

382

383 **TABLE 1.** Description of the dead equine sample, live equine sample and equine dataset

Information	Modalities	Dead equine sample (n=13,595)		Live equine sample (n=2,353)		Equine dataset (n=1,068,142)	
		n	%	n	%	n	%
Sex	NA	0	0.0	0	0	52	0.0
	Female	7 938	58.4	1234	52.4	575099	53.8
	Gelding	3 444	25.3	640	27.2	205426	19.2
	Male	2 213	16.3	479	20.4	287565	26.9
Age*	NA	5011	0.0	448	0.0	0	0.0
	0	115	1.3	0	0.0	0	0.0
	[1-2]	938	10.9	144	7.6	81324	7.6
	[3-10]	1850	21.6	678	35.6	402158	37.7
	[11-20]	2545	29.6	853	44.8	412812	38.6
	[21-30]	2820	32.9	226	11.9	151357	14.2
	>30	316	3.7	4	0.2	20491	1.9
Group of breeds	Donkey	822	6.0	70	3.0	46194	4.3
	Racehorse	1923	14.1	341	14.5	216041	20.2
	Pony	3178	23.4	486	20.7	150962	14.1
	Saddle horse	6360	46.8	1362	57.9	467460	43.8
	Draft horse	1312	9.7	94	4.0	187485	17.6
Number of equine(s) per owner	1	1689	25.4	826	41.1	251939	63.5
	[2-5]	2222	33.5	814	40.5	112468	28.3
	[6-20]	1801	27.1	290	14.4	26875	6.8
	[21-100]	828	12.5	76	3.8	5437	1.4
	>100	98	1.5	2	0.1	320	0.1

Total number of different owners	6638	100	2008	100	397039	100
----------------------------------	------	-----	------	-----	--------	-----

384 \* In years; age at death for a dead equine, or age at the date of data extraction for a live equine

385

386 **TABLE 2.** Distribution of the average number of equines (among the 100 simulations) per *commune* depending  
 387 on the assignment method and sample used to estimate the parameters

Sample	Method	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
Live equines	1	0	6	13	30.7	33	10828
	2	0	2	7	30.6	36	1313
Dead equines	1	0	5	24	30.8	37	2810
	2	0	2	8	30.6	37	1418

388

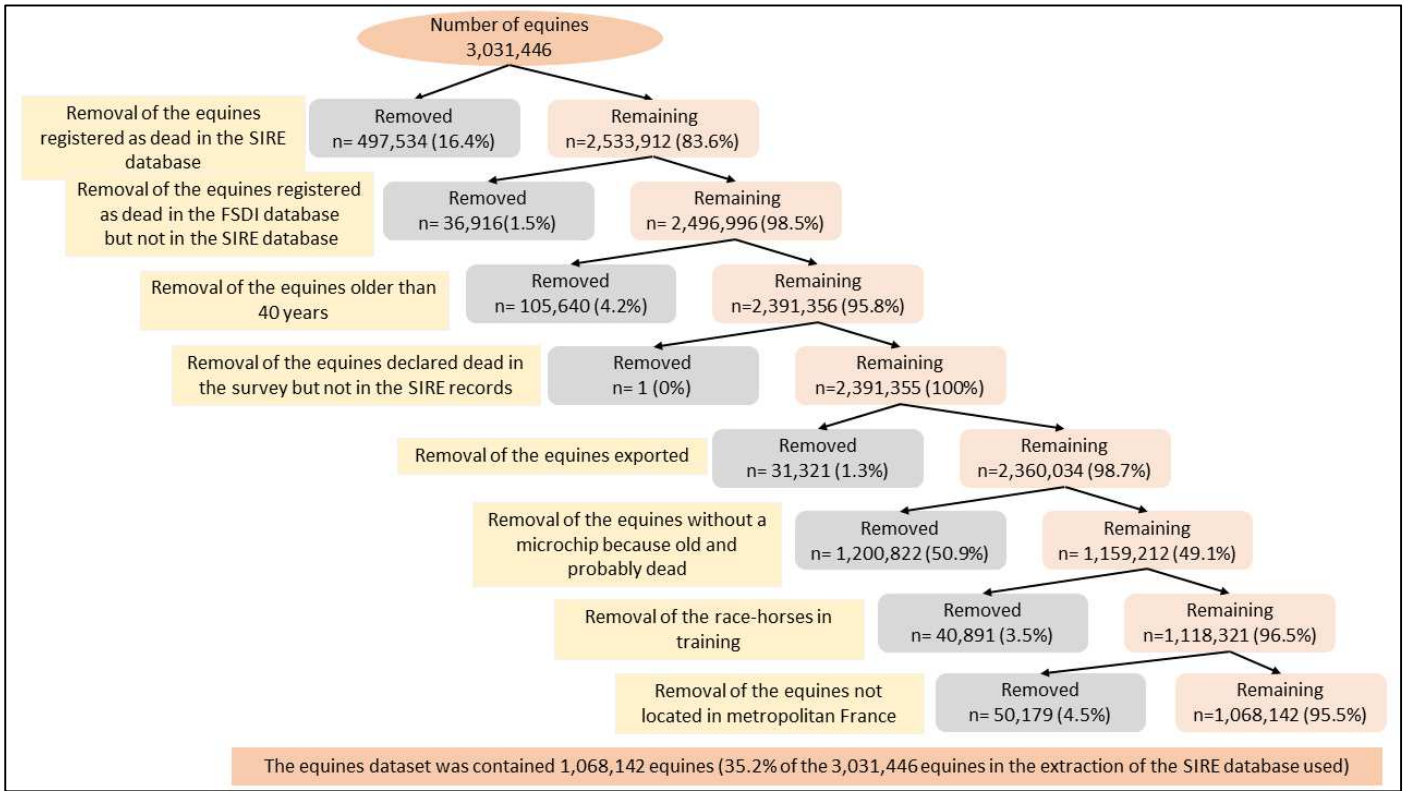
389 **TABLE 3.** Comparison of the results of the models: distributions of the difference in the average number of  
 390 equines assigned to each *commune* according to the samples and models.

Comparison level	Method or sample concerned	Distribution of the difference	No difference	Positive difference	Negative difference
Live equine vs. dead equine sample	Method 1	%(number of <i>communes</i> )	43.2 (15051)	20.8 (7257)	36.0 (12571)
		Minimum		1	-1
		1st quartile		5	-22
		Median		15	-24
		Mean		36.5	-21.4
		3rd quartile		38	-26
		Maximum		10822	-37
	Method 1	%(number of <i>communes</i> )	47.9 (16718)	26.0 (9073)	26.1 (9088)

		Minimum	1	-1
		1st quartile	1	-1
		Median	2	-2
		Mean	2.9	-2.98
		3rd quartile	4	-3
		Maximum	50	-143
Method 1 vs. Live equine		% (number of <i>communes</i> )	2.4 (841)	65.7 (22910) 31.9 (11128)
Method 2	sample	Minimum	1	-1
		1st quartile	5	-22
		Median	13	-42
		Mean	28.98	-59.2
		3rd quartile	31	-74
		Maximum	10823	-457
	Dead	% (number of <i>communes</i> )	4.3 (1514)	73.4 (25612) 22.2 (7753)
	equine	Minimum	1	-1
	sample	1st quartile	2	-14
		Median	7	-33
		Mean	16.4	-53.5
		3rd quartile	17	-66
		Maximum	1392	-423

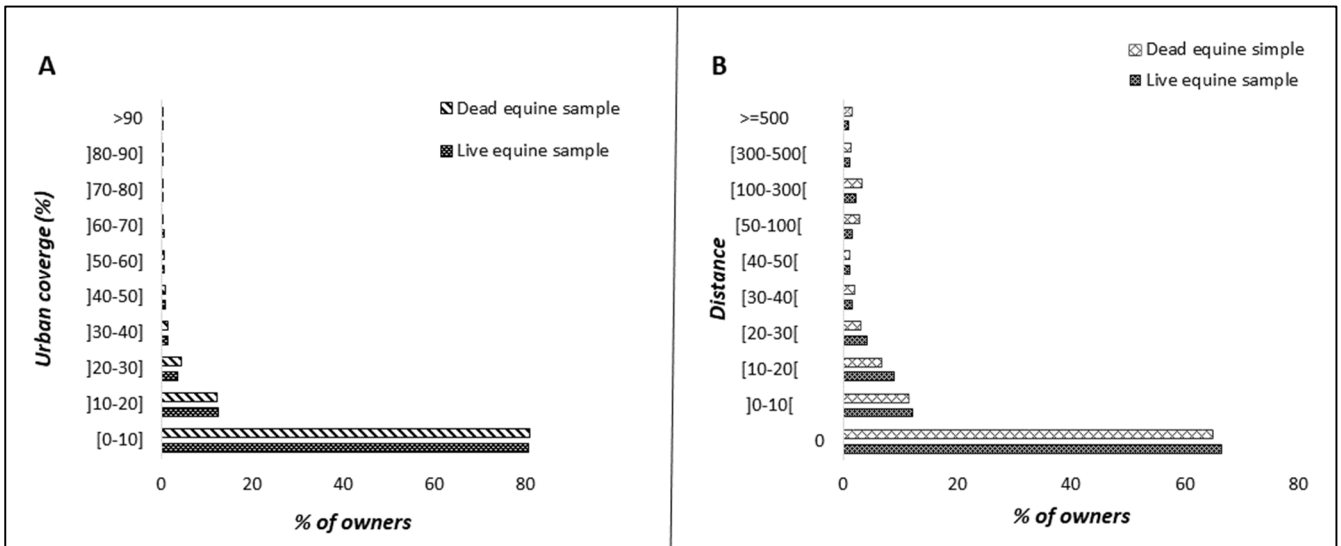
391

392



394

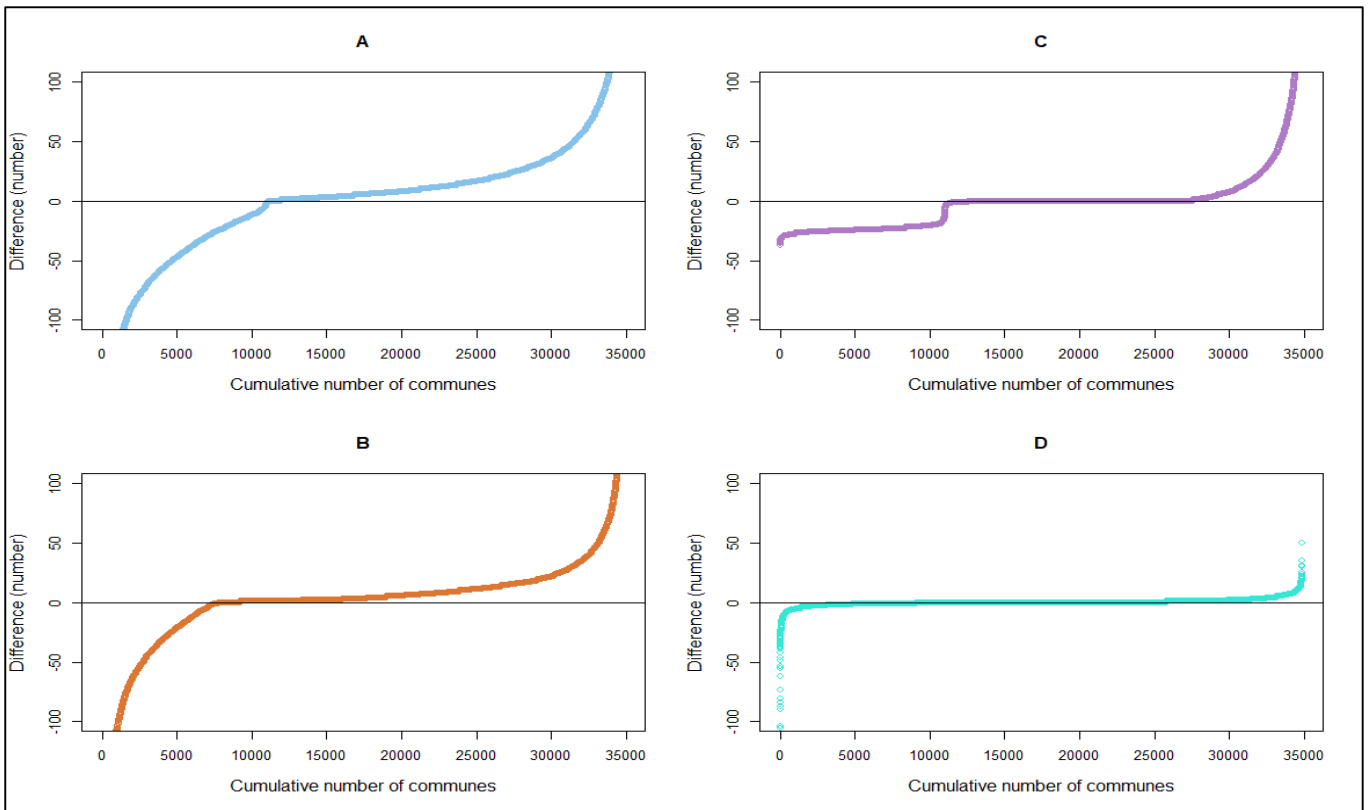
395 **Figure 1.** Different steps determining the number of equines assumed alive and used to build the equines  
396 dataset.



397

398 **Figure 2.** Distribution of the proportion of owners keeping their equine in their *commune* of residence  
399 according to the urban coverage (A) and distribution of the proportion of owners (B) according to the  
400 distance between the residence and holding *communes* in the live and dead equine samples

401

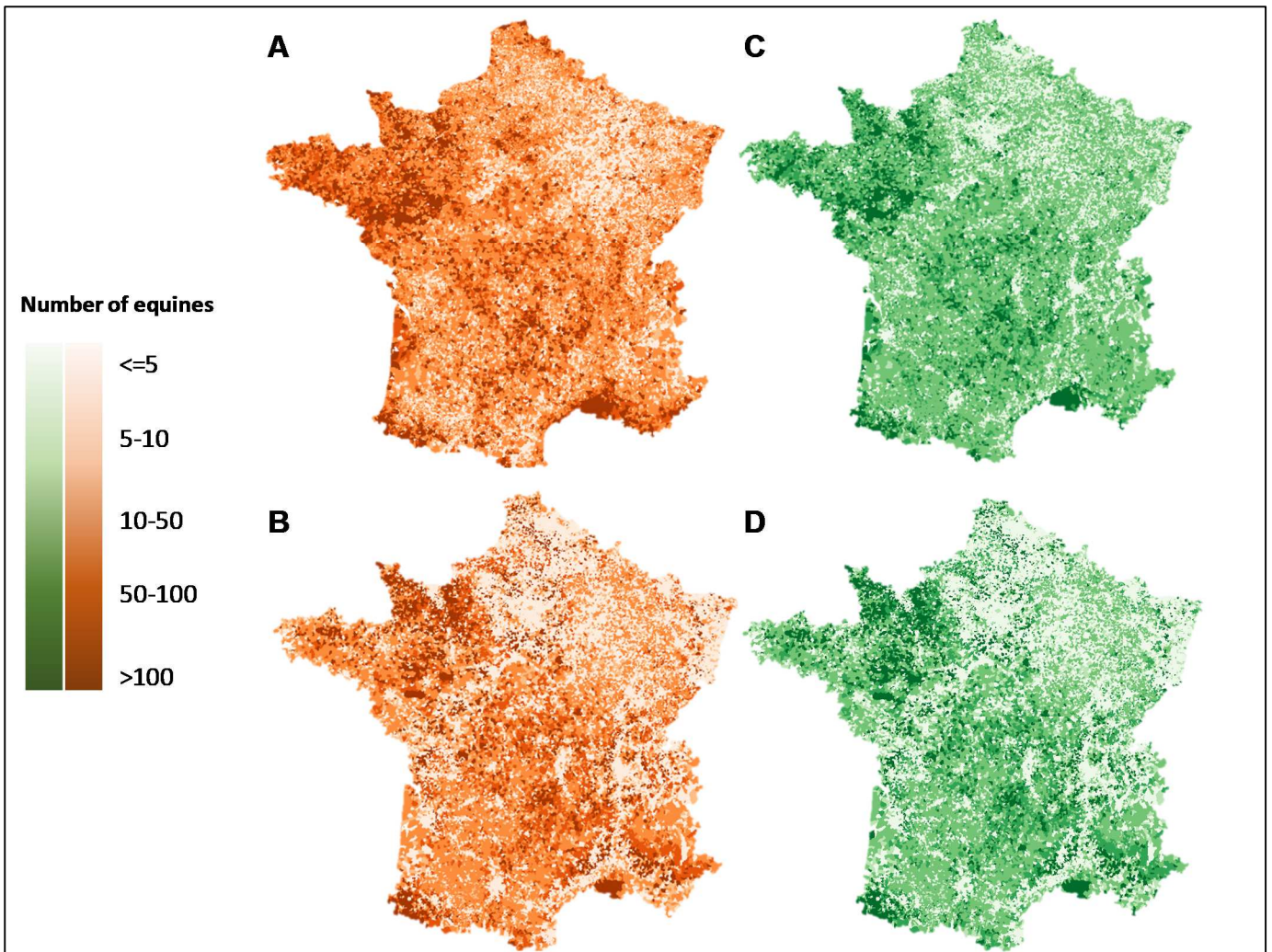


402

403 **FIGURE 3.** Difference in the average number of equines per *commune* according to the sample and method  
 404 used (restricted to values between 100 and -100): (A) live equine sample Method 1 vs. Method 2, (B) dead  
 405 equine sample Method 1 vs. Method 2, (C) Method 1 live equine sample vs. dead equine sample and (D)  
 406 Method 2 live equine sample vs. dead equine sample

407

408



409

410 **FIGURE 4.** Spatial distribution of the equine population in France estimated from (A) Method 1 and live equine  
 411 sample, (B) Method 2 and live equine sample, (C) Method 1 and dead equine sample and (D) Method 2 and  
 412 dead equine sample (D) after inclusion of racehorse data.

413

414



415

416 **Reference**

417 Amat, J.P., Tapprest, J., Gaudaire, D., Merlin, A., Strugar, S., Coulibaly, E., Grandcollot-Chabot, M., Hans, A.,  
418 2018. Deux cas d'anémie infectieuse des équidés détectés dans le sud-est de la France : point de situation  
419 n°2 (27/03/2018) Plateforme ESA.

420 Anonyme, 2016. Règlement 2016/429 du Parlement Européen et du Conseil du 9 mars 2016 publiée au  
421 journal officiel de l'Union Européenne le 31 mars 2016.

422 Brendan, C., Michael, P.W., Samuel, H., Greame, G., 2009. The equine influenza epidemic in Australia: spatial  
423 and temporal descriptive analyses of a large propagating epidemic. *Prev. Vet. Med.* 92, 60-70.

424 Callinan, I.H., 2007. Equine influenza. The August 2007 outbreak in Australia. [https://nla.gov.au/nla.obj-  
425 961843962/view?partId=nla.obj-962288923#page/n5 /mode/1up](https://nla.gov.au/nla.obj-961843962/view?partId=nla.obj-962288923#page/n5 /mode/1up), (accessed on 30 September 2021).

426 Cover, C.L., Corine Land Cover : données d'inventaire biophysique de l'occupation des sols et de son  
427 évolution selon une nomenclature en 44 postes. [https://www.statistiques.developpement-  
428 durable.gouv.fr/corine-land-cover-0](https://www.statistiques.developpement-durable.gouv.fr/corine-land-cover-0), (accessed on 30 September 2021).

429 Ecus-Ifce, Annuaire écus, IFCE, Institut Français du Cheval et de L'équitation. [https://www.ifce.fr/recherche-  
430 transverse/?q=ANNUAIRE+%C3%89CUS&cb=https://www.ifce.fr/](https://www.ifce.fr/recherche-transverse/?q=ANNUAIRE+%C3%89CUS&cb=https://www.ifce.fr/), (accessed on 30 September 2021).

431 Farchati, H., Merlin, A., Sassac, M., Dornier, X., Dhollande, M., Garon, D., Tapprest, J., Sala, C., 2020. Is the  
432 French SIRE equine information system a good basis for epidemiological and modeling research? Quality  
433 assessment using two surveys. *Res. Vet. Sci* 134, 96–101.

434 Farchati, H., Merlin, A., Sassac, M., Dornier, X., Dhollande, M., Garon, D., Tapprest, J., Sala, C., 2021. Home  
435 Sweet Home: New Insights Into the Location of Equine Premises in France and Keeping Habits to Inform  
436 Health Prevention and Disease Surveillance. *Front Vet Sci.* 23, 8:701749.

437 Ferguson, N.M., D.A.T., C., S., C., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S., Burke, D.S., 2005.  
438 Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437, 209–214.

439 Ferguson, N.M., Donnelly, C.A., Anderson, R.M., 2001. Transmission intensity and impact of control policies  
440 on the foot and mouth epidemic in Great Britain. *Nature* 413, 542–548.

441 Garner, M.G., Cowled, B., East, I.J., Moloney, B.J., Kung, N.Y., 2010. Evaluating the effectiveness of early  
442 vaccination in the control and eradication of equine influenza – a modelling approach. *Prev. Vet. Med.* 99,  
443 15-27.

444 Geweke, J.F., 1991. Evaluating the accuracy of sampling-based approaches to the calculation of posterior  
445 moments. In *Bayesian Statistics 4* (ed JM Bernardo, JO Berger, AP Dawid and AFM Smith). Clarendon Press,  
446 Oxford, UK.

447 Gudelj, I., K.A.J., W., 2004. Spatial heterogeneity, social structure and disease dynamics of animal  
448 populations. *Theor Popul Biol*, 66, 139–149.

449 House, T., Keeling, M.J., 2008a. Deterministic epidemic models with explicit household structure. *Math.*  
450 *Biosci.* 213, 29–39.

451 House, T., Keeling, M.J., 2008b. Household structure and infectious disease transmission. *Epidemiol. Infect.*  
452 137, 654–661.

453 Keeling, M.J., Danon, L., Vernon, M.C., House, T.A., 2010. Individual identity and movement networks for  
454 disease metapopulations. *Proc. Nat.l Acad. Sci. U.S.A.* 107, 8866–8870.

455 Keeling, M.J., Woolhouse, M.E., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J.,  
456 Kappey, J., Wilesmith, J., Grenfell, B.T., 2001. Dynamics of the 2001 UK foot and mouth epidemic: Stochastic  
457 dispersal in a heterogeneous landscape. *Science* 294, 813–817.

458 Lo Iacono, G., Robin, C.A., Newton, J.R., Gubbins, S., Wood, J.L., 2013. Where are the horses? With the sheep  
459 or cows? Uncertain host location, vector-feeding preferences and the risk of African horse sickness  
460 transmission in Great Britain. *J. R. Soc. Interface* 10, 20130194.

461 Martin, A.D., Quinn, K.M., Park, J.H., 2011. MCMCpack: Markov chain Monte Carlo in R. *J. Stat. Software* 42,  
462 323.

463 Merlin, A., Gaudaire, D., Letorey, L., Delerue, M., Amat, J.P., Lecouturier, F., Madeline, A., Deshiere, A.,  
464 Aubert, C., Poudevigne, F., Quinio, C., Martin, F., Strugar, S., Coulibaly, E., Depecker, A., Grancollot-Chabot,  
465 M., Hans, A., 2021. Synthèse des foyers d'anémie infectieuse des équines reportées en France entre 2017 et  
466 2019.

467 Swinton, J., Harwood, J., Grenfell, B.T., Gilligan, C.A., 1998. Persistence thresholds for phocine distemper  
468 virus infection in harbour seal *Phoca vitulina* metapopulations. *J Anim Ecol* 67, 54–68.

469 TRACE, The Trade Control and Expert System. [https://ec.europa.eu/food/animals/traces\\_en](https://ec.europa.eu/food/animals/traces_en), (accessed on 30  
470 September 2021).