Internship progress

Abdourahmane Diallo

2024-03-18

1 Setting

1.1 Packages

- Code

1.2 Functions

- Code

2 Plan

- Setting
- Database import
- Database exploration
- Earthworms data
- Soil data extraction
- Climate data extraction
- To do next

3 Database import

- Import of database **LandWorm_dataset_site_V1.9.xlsx** (february 22, 2024)
- The database contains **8019** rows and **481** columns

3.1 Data selection: EcoBioSoil

	Numbers
cp	227
dc	5520
gp	299
mh	867
sg	545
NA's	561

- The database therefore changes from **8019** to **5520** observations.

4 Database exploration

- CR = Completion rate

4.1 Complete columns

▶ Code

	Variables	CR
79	ID	100%
80	Protocole	100%
82	owner	100%
83	AB_tot	100%
127	AB_Allolobophora_chlorotica_chlorotica	100%
167	AB_AD	100%
168	AB_JV	100%

	Variables	CR
169	AB_SA	100%

▶ Code

4.2 Non-complete columns

	Variables	CR
93	AB_Lumbricus_castaneus	99.5%
90	AB_Aporrectodea_rosea	99%
87	AB_Aporrectodea_caliginosa	98.9%
123	AB_Lumbricus_terrestris	98.6%
121	AB_Aporrectodea_icterica	98%
145	BM_Aporrectodea_icterica	96.4%
109	BM_Lumbricus_castaneus	96.3%
1	Programme	96.1%

	Variables	CR
2	Annee	96.1%
4	ID_Site	96.1%
106	BM_Aporrectodea_rosea	96%
97	AB_Octolasion_cyaneum	95.8%
100	AB_Satchellius_mammalis	95.6%
11	clcm_lvl1	95.4%
12	clcm_lvl2	95.4%
13	clcm_lvl3	95.4%
16	code_clcm_lvl1	95.4%
17	code_clcm_lvl2	95.4%

	Variables	CR
18	code_clcm_lvl3	95.4%
113	BM_Octolasion_cyaneum	95.4%
173	AB_Aporrectodea_longa_longa	94.9%
120	AB_Aporrectodea_giardi	94.5%
116	BM_Satchellius_mammalis	94.3%
174	AB_Aporrectodea_nocturna	93.5%
147	BM_Lumbricus_terrestris	93.2%
8	gps_x	92.8%
9	gps_y	92.7%
144	BM_Aporrectodea_giardi	92.1%

	Variables	CR
197	BM_Aporrectodea_nocturna	91.5%
96	AB_Murchieona_muldali	90.7%
112	BM_Murchieona_muldali	90.5%
187	AB_Lumbricus_sp	90.5%
210	BM_Lumbricus_sp	89.7%
196	BM_Aporrectodea_longa_longa	89.4%
88	AB_Aporrectodea_caliginosa_meridionalis	89%
186	AB_Lumbricus_rubellus_castanoides	89%
209	BM_Lumbricus_rubellus_castanoides	88.8%
103	BM_Aporrectodea_caliginosa	87.9%

	Variables	CR
182	AB_indéterminable	87.3%
205	BM_indéterminable	87.3%
104	BM_Aporrectodea_caliginosa_meridionalis	86.9%
177	AB_Aporrectodea_sp	86.9%
200	BM_Aporrectodea_sp	86.3%
188	AB_Octolasion_sp	85.5%
211	BM_Octolasion_sp	85.5%
131	AB_Lumbricus_rubellus_rubellus	84.2%
151	BM_Allolobophora_chlorotica_chlorotica	83.9%
155	BM_Lumbricus_rubellus_rubellus	83.2%

	Variables	CR
81	Code_Parcelle	77.8%
180	AB_Dendrobaena_sp	77%
203	BM_Dendrobaena_sp	77%
125	AB_Lumbricus_festivus	75.5%
149	BM_Lumbricus_festivus	75.5%
170	AB_Allolobophora_chlorotica_postepheba	74.3%
193	BM_Allolobophora_chlorotica_postepheba	73.9%
89	AB_Aporrectodea_cupulifera	72.1%
105	BM_Aporrectodea_cupulifera	72.1%
94	AB_Lumbricus_friendi	72%

	Variables	CR
172	AB_Aporrectodea_indéterminable	69.2%
195	BM_Aporrectodea_indéterminable	68.9%
110	BM_Lumbricus_friendi	68.7%
141	AB_Octolasion_lacteum_lacteum	68.6%
165	BM_Octolasion_lacteum_lacteum	68.6%
135	AB_Lumbricus_centralis	67.2%
92	AB_Eiseniella_tetraedra	67%
108	BM_Eiseniella_tetraedra	66.9%
159	BM_Lumbricus_centralis	66.2%
37	clay	64.1%

	Variables	CR
181	AB_Eisenia_fetida	62.1%
204	BM_Eisenia_fetida	62.1%
175	AB_Aporrectodea_ripicola	61.9%
198	BM_Aporrectodea_ripicola	61.7%
31	fine_sand	61.6%
32	coarse_sand	61.6%
34	fine_silt	61.6%
35	coarse_silt	61.6%
225	AB_Eisenia_andrei	61.4%
230	BM_Eisenia_andrei	61.4%

	Variables	CR
184	AB_Lumbricus_castaneus_disjonctus	61.1%
207	BM_Lumbricus_castaneus_disjonctus	60.9%
237	AB_Microscolex_phosphoreus	60.3%
239	BM_Microscolex_phosphoreus	60.3%
21	ph_eau	54.1%
142	AB_Microscolex_dubius	53.6%
166	BM_Microscolex_dubius	53.6%
26	om	52.6%
171	AB_Allolobophora_sp	51.9%
194	BM_Allolobophora_sp	51.7%

	Variables	CR
224	AB_Dendrodrilus_rubidus	51.3%
229	BM_Dendrodrilus_rubidus	51.3%
245	AB_Lumbricus_friendi_lineatus	50%
253	BM_Lumbricus_friendi_lineatus	49.9%
24	n_tot	49.4%
192	AB_Scherotheca_sp	48.8%
215	BM_Scherotheca_sp	48.8%
236	AB_Aporrectodea_limicola	48.5%
238	BM_Aporrectodea_limicola	48.5%
190	AB_Prosellodrilus_sp	48.2%

	Variables	CR
213	BM_Prosellodrilus_sp	48.1%
122	AB_Aporrectodea_longa	46.8%
146	BM_Aporrectodea_longa	46.8%
226	AB_Proctodrilus_antipai_antipai	44.9%
231	BM_Proctodrilus_antipai_antipai	44.7%
232	AB_Eisenia_veneta	44.7%
233	BM_Eisenia_veneta	44.7%
183	AB_Indéterminable	44%
206	BM_Indéterminable	44%
23	c_org	43.7%

	Variables	CR
101	AB_Scherotheca_savignyi_indéterminable	42.6%
117	BM_Scherotheca_savignyi_indéterminable	42.6%
291	AB_Scherotheca_savignyi_savignyi	42.6%
293	BM_Scherotheca_savignyi_savignyi	42.6%
84	BM_tot	41.2%
10	Altitude	40.4%
246	AB_Octodrilus_complanatus	40.1%
254	BM_Octodrilus_complanatus	40%
179	AB_Aporrectodea_tuberculata	39.9%
202	BM_Aporrectodea_tuberculata	39.9%

	Variables	CR
178	AB_Aporrectodea_trapezoides	39%
201	BM_Aporrectodea_trapezoides	39%
290	AB_Allolobophora_burgondiae	37.6%
136	AB_Lumbricus_rubellus_friendoides	37.5%
295	AB_Scherotheca_aquitana	37.5%
292	BM_Allolobophora_burgondiae	37.4%
297	BM_Scherotheca_aquitana	37.4%
160	BM_Lumbricus_rubellus_friendoides	37.3%
15	land_cover_detail	36.6%
189	AB_Prosellodrilus_amplisetosus_amplisetosus	36.4%

	Variables	CR
212	BM_Prosellodrilus_amplisetosus_amplisetosus	36.4%
280	AB_Eisenia_fetida_indéterminable	35.8%
282	BM_Eisenia_fetida_indéterminable	35.8%
278	AB_Bimastos_eiseni	32%
279	BM_Bimastos_eiseni	32%
91	AB_Dendrobaena_octaedra	28.7%
107	BM_Dendrobaena_octaedra	28.7%
3	Date_Prelevement	24.4%
119	AB_Aporrectodea_caliginosa_indéterminable	22.9%
294	AB_Aporrectodea_rubra_acidicola	22.8%

	Variables	CR
296	BM_Aporrectodea_rubra_acidicola	22.8%
143	BM_Aporrectodea_caliginosa_indéterminable	22.3%
286	AB_Prosellodrilus_occidentalis_occidentalis	21.2%
289	BM_Prosellodrilus_occidentalis_occidentalis	21.2%
133	AB_Prosellodrilus_fragilis_fragilis	20.6%
157	BM_Prosellodrilus_fragilis_fragilis	20.6%
7	postal_code	20%
185	AB_Lumbricus_meliboeus	18.6%
208	BM_Lumbricus_meliboeus	18.6%
33	sand	18.2%

	Variables	CR
36	silt	18.2%
138	AB_Dendrodrilus_rubidus_subrubicundus	18.1%
162	BM_Dendrodrilus_rubidus_subrubicundus	18.1%
5	Modalite	17.4%
85	AB_STAD_X	17.1%
223	AB_Dendrobaena_attemsi	16.4%
228	BM_Dendrobaena_attemsi	16.4%
22	c_tot	15.8%
99	AB_Prosellodrilus_fragilis_indéterminable	15.2%
115	BM_Prosellodrilus_fragilis_indéterminable	15.2%

	Variables	CR
301	AB_Prosellodrilus_amplisetosus	15.1%
307	BM_Prosellodrilus_amplisetosus	15.1%
298	AB_Hemigastrodrilus_monicae	14.7%
299	AB_Octodrilus_indéterminable	14.7%
300	AB_Proctodrilus_antipai_indéterminable	14.7%
302	AB_Prosellodrilus_praticola	14.7%
303	AB_Scherotheca_porotheca	14.7%
305	BM_Octodrilus_indéterminable	14.7%
306	BM_Proctodrilus_antipai_indéterminable	14.7%
308	BM_Prosellodrilus_praticola	14.7%

	Variables	CR
309	BM_Scherotheca_porotheca	14.7%
304	BM_Hemigastrodrilus_monicae	14.6%
234	AB_Haplotaxis_sp	13.9%
235	BM_Haplotaxis_sp	13.9%
38	type_tillage	13.2%
28	cu_EDTA	12.4%
222	AB_Avelona_ligra	11.1%
227	BM_Avelona_ligra	11.1%
27	cu_tot	10.8%
263	AB_Pheretima_indéterminable	10.2%

	Variables	CR
271	BM_Pheretima_indéterminable	10.2%
51	herbicide_freq	10%
259	AB_Dendrobaena_cognettii	9.4%
267	BM_Dendrobaena_cognettii	9.4%
50	insecticide_freq	9.3%
6	Bloc	8.9%
43	fertilisation	8.4%
260	AB_Dendrobaena_hortensis	8.4%
268	BM_Dendrobaena_hortensis	8.4%
49	fungicide_freq	7.5%

	Variables	CR
284	AB_Microscolex_sp	7%
287	BM_Microscolex_sp	7%
285	AB_Pheritima_Diffringens	6.6%
288	BM_Pheritima_Diffringens	6.6%
44	ferti_min_product	6.4%
46	ferti_orga_product	6.2%
78	grassland_type	5.6%
63	rotation_plant_div	5.1%
281	AB_Lumbricus_rubellus_indéterminable	5.1%
283	BM_Lumbricus_rubellus_indéterminable	5.1%

	Variables	CR
56	tfi_herbicide	4.5%
73	herbage_use	4.3%
86	AB_Allolobophora_chlorotica_indéterminable	4%
102	BM_Allolobophora_chlorotica_indéterminable	4%
176	AB_Aporrectodea_rubra	4%
191	AB_Scherotheca_dinoscolex	4%
199	BM_Aporrectodea_rubra	4%
214	BM_Scherotheca_dinoscolex	4%
40	tillage_frequency_intra	3.5%
20	ph_kcl	3.3%

	Variables	CR
52	molluscicide_freq	3.2%
45	ferti_min_qtty	3.1%
47	ferti_orga_qtty	3.1%
66	crop_residues_management	2.7%
59	total_tfi	2.5%
75	herb_age	2%
76	animal_loading	2%
60	mecanical_weed_control	1.9%
65	rotation_grassland	1.8%
258	AB_Dendrobaena_alpina_zeugochaeta	1.8%

	Variables	CR
261	AB_Eisenia_sp	1.8%
262	AB_Flabellodrilus_bartolii	1.8%
264	AB_Prosellodrilus_pyrenaicus	1.8%
265	AB_Scherotheca_nivicola	1.8%
266	BM_Dendrobaena_alpina_zeugochaeta	1.8%
269	BM_Eisenia_sp	1.8%
270	BM_Flabellodrilus_bartolii	1.8%
272	BM_Prosellodrilus_pyrenaicus	1.8%
273	BM_Scherotheca_nivicola	1.8%
216	AB_Aporrectodea_nocturna_nocturna_cistercianus	1.6%

	Variables	CR
217	AB_Scherotheca_mifuga	1.6%
218	AB_Scherotheca_rhodana	1.6%
219	BM_Aporrectodea_nocturna_nocturna_cistercianus	1.6%
220	BM_Scherotheca_mifuga	1.6%
221	BM_Scherotheca_rhodana	1.6%
55	tfi_insecticide	1.5%
98	AB_Octolasion_lacteum	1.1%
114	BM_Octolasion_lacteum	1.1%
77	trampling_nature	1%
314	AB_Octolasion_lacteum_gracile	0.9%

	Variables	CR
315	BM_Octolasion_lacteum_gracile	0.9%
242	AB_Ethnodrilus_lydiae	0.8%
243	AB_Hemigastrodrilus_monicae_magnus	0.8%
244	AB_Hormogaster_praetiosa	0.8%
247	AB_Prosellodrilus_indéterminable	0.8%
248	AB_Scherotheca_corsicana_corsicana	0.8%
249	AB_Zophoscolex_graffi	0.8%
250	BM_Ethnodrilus_lydiae	0.8%
251	BM_Hemigastrodrilus_monicae_magnus	0.8%
252	BM_Hormogaster_praetiosa	0.8%

	Variables	CR
255	BM_Prosellodrilus_indéterminable	0.8%
256	BM_Scherotheca_corsicana_corsicana	0.8%
257	BM_Zophoscolex_graffi	0.8%
74	mowing_frequency_yr	0.7%
274	AB_Aporrectodea_georgii	0.5%
275	AB_Panoniona_leoni	0.5%
276	BM_Aporrectodea_georgii	0.5%
277	BM_Panoniona_leoni	0.5%
240	AB_Aporrectodea_balisa	0.4%
241	BM_Aporrectodea_balisa	0.4%

	Variables	CR
310	AB_Scherotheca_minor	0.4%
311	BM_Scherotheca_minor	0.4%
14	clcm_lvl4	0.3%
19	code_clcm_lvl4	0.3%
312	AB_Orodrilus_paradoxus_paradoxus	0.3%
313	BM_Orodrilus_paradoxus_paradoxus	0.3%
25	c/n	0%
29	soil_temperature	0%
30	soil_humidity	0%
39	tillage_depth	0%

	Variables	CR
41	tillage_frequency_inter	0%
42	tillage_date	0%
48	ferti_orga_freq	0%
53	nematicide_freq	0%
54	tfi_fungicide	0%
57	tfi_mollucicide	0%
58	tfi_nematicide	0%
61	thermal_weed_control	0%
62	crop_rotation_yr	0%
64	intercrop_div	0%

	Variables	CR
67	amdmt_orga_freq	0%
68	amdmt_orga_names	0%
69	amdmt_orga_qtty	0%
70	amdmt_calcic	0%
71	amdmt_calcic_names	0%
72	amdmt_calcic_qtty	0%
95	AB_Lumbricus_herculeus	0%
111	BM_Lumbricus_herculeus	0%
118	Parcelle	0%
124	AB_Allolobophora_chlorotica	0%

	Variables	CR
126	AB_A._muldali/rosea	0%
128	AB_Aporrectodea_longa/giardi	0%
129	AB_Indéterminable_epigeic	0%
130	AB_Lumbricus_friendi/centralis	0%
132	AB_Octolasion_indéterminable	0%
134	AB_Dendrobaena_pygmea	0%
137	AB_indéterminable_endogeic	0%
139	AB_Lumbricus_indéterminable_anecic	0%
140	AB_Eisenia_indéterminable	0%
148	BM_Allolobophora_chlorotica	0%

	Variables	CR
150	BM_A._muldali/rosea	0%
152	BM_Aporrectodea_longa/giardi	0%
153	BM_Indéterminable_epigeic	0%
154	BM_Lumbricus_friendi/centralis	0%
156	BM_Octolasion_indéterminable	0%
158	BM_Dendrobaena_pygmea	0%
161	BM_indéterminable_endogeic	0%
163	BM_Lumbricus_indéterminable_anecic	0%
164	BM_Eisenia_indéterminable	0%
316	AB_Ethnodrilus_zajonci	0%

	Variables	CR
317	BM_Ethnodrilus_zajonci	0%
318	AB_Hormogaster_sp	0%
319	AB_Octodrilus_lisseansis	0%
320	BM_Hormogaster_sp	0%
321	BM_Octodrilus_lisseansis	0%
322	AB_Scherotheca_michaelseni	0%
323	AB_Scherotheca_occidentalis	0%
324	AB_Scherotheca_occitanica	0%
325	AB_Aporrectodea_haymozi	0%
326	AB_Dendrobaena_alpina	0%

	Variables	CR
327	AB_Scherotheca_corsicana	0%
328	AB_Octolasion_tyrtaeum	0%
329	AB_Lumbricus_rubellus	0%
330	AB_Aporrectodea_terrestris	0%
331	AB_Aporrectodea_rubicunda	0%
332	AB_Diporodrilus_omodeoi	0%
333	AB_Eisenia_parva	0%
334	AB_Scherotheca_albomaculata	0%
335	AB_Bimastos_rubidus	0%
336	AB_Scherotheca_portonana	0%

	Variables	CR
337	AB_Scherotheca_brevisella	0%
338	AB_Proctodrilus_antipai	0%
339	AB_Octodrilus_juvyi	0%
340	AB_Dendrobaena_byblica	0%
341	AB_Dendrodrilus_subrubicundus	0%
342	AB_Prosellodrilus_albus	0%
343	AB_Kritodrilus_tetryae	0%
344	AB_Lumbricus_klarae	0%
345	AB_Aporrectodea_haymoziformis	0%
346	AB_Kritodrilus_micrurus	0%

	Variables	CR
347	AB_Allolobophora_satchelli	0%
348	AB_Ethnodrilus_aveli	0%
349	AB_Aporrectodea_zicsii	0%
350	AB_Diporodrilus_pilosus	0%
351	AB_Eumenescolex_emiliae	0%
352	AB_Dendrobaena_pantaleonis	0%
353	AB_Dendrobaena_veneta	0%
354	AB_Lumbricidae_f	0%
355	AB_Murchieona_minuscula	0%
356	BM_Lumbricidae_f	0%

	Variables	CR
357	BM_Murchieona_minuscula	0%
358	BM_Octolasion_tyrtaeum	0%
359	AB_Dendrobaena_alpina_indéterminable	0%
360	BM_Dendrobaena_alpina_indéterminable	0%
361	AB_Oligochaeta_so	0%
362	BM_Oligochaeta_so	0%
363	AB_Adult	0%
364	AB_cocon	0%
365	AB_indéterminé	0%
366	AB_Juvenile	0%

	Variables	CR
367	AB_Sub.adult	0%
368	AB_Allolobophora_delitescens	0%
369	AB_Amynthas_indicus	0%
370	AB_Aporrectodea_arverna	0%
371	AB_Aporrectodea_cuendeti	0%
372	AB_Aporrectodea_gogna	0%
373	AB_Aporrectodea_sineporis	0%
374	AB_Aporrectodea_velox	0%
375	AB_Aporrectodea_voconca	0%
376	AB_Bimastos_parvus	0%

	Variables	CR
377	AB_Boucheona_corbierensis	0%
378	AB_Boucheona_rosae	0%
379	AB_Ethnodrilus_gatesi	0%
380	AB_Ethnodrilus_setusmonsanus	0%
381	AB_Flabellodrilus_luberonensis	0%
382	AB_Gatesona_chaetophora	0%
383	AB_Gatesona_lablacherensis	0%
384	AB_Gatesona_rutena	0%
385	AB_Haplotaxis_gordioides	0%
386	AB_Helodrilus_oculatus	0%

	Variables	CR
387	AB_Hormogaster_insularis	0%
388	AB_Hormogaster_samnitica_lirapora	0%
389	AB_Kritodrilus_calarensis	0%
390	AB_Lucquesia_tiginosa	0%
391	AB_Lumbricus_bouchei	0%
392	AB_Lumbricus_improvisus	0%
393	AB_Octodrilus_hemiandrus	0%
394	AB_Panoniona_satchelli	0%
395	AB_Proctodrilus_tuberculatus	0%
396	AB_Prosellodrilus_alatus	0%

	Variables	CR
397	AB_Prosellodrilus_biserialis	0%
398	AB_Prosellodrilus_fragilis_polythecosus	0%
399	AB_Prosellodrilus_idealis	0%
400	AB_Scherotheca_altarocca	0%
401	AB_Scherotheca_betharramensis	0%
402	AB_Scherotheca_boccaverhju	0%
403	AB_Scherotheca_capcorsana	0%
404	AB_Scherotheca_chicharia	0%
405	AB_Scherotheca_darioi	0%
406	AB_Scherotheca_gigas_gigas	0%

	Variables	CR
407	AB_Scherotheca_haymozi	0%
408	AB_Scherotheca_minor_minorissima	0%
409	AB_Scherotheca_monspessulensis_idica	0%
410	AB_Scherotheca_monspessulensis_monspessulensis	0%
411	AB_Scherotheca_orbiensis	0%
412	AB_Scherotheca_pereli	0%
413	AB_Scherotheca_qiui	0%
414	AB_Scherotheca_sanaryensis	0%
415	AB_Scherotheca_trezencensis	0%
416	AB_Vignysa_callasensis	0%

	Variables	CR
417	AB_Vignysa_teres	0%
418	AB_Vosgesia_zicsii	0%
419	AB_Zophoscolex_atlanticus	0%
420	AB_Zophoscolex_micellus	0%
421	BM_Allolobophora_delitescens	0%
422	BM_Amynthas_indicus	0%
423	BM_Aporrectodea_arverna	0%
424	BM_Aporrectodea_cuendeti	0%
425	BM_Aporrectodea_gogna	0%
426	BM_Aporrectodea_sineporis	0%

	Variables	CR
427	BM_Aporrectodea_velox	0%
428	BM_Aporrectodea_voconca	0%
429	BM_Bimastos_parvus	0%
430	BM_Boucheona_corbierensis	0%
431	BM_Boucheona_rosae	0%
432	BM_Dendrobaena_byblica	0%
433	BM_Diporodrilus_omodeoi	0%
434	BM_Diporodrilus_pilosus	0%
435	BM_Ethnodrilus_aveli	0%
436	BM_Ethnodrilus_gatesi	0%

4.3 Focus on GPS coordinates

- There is **398** NA (CR = 92.8%) in **GPS_X**
- There is **401** NA (CR = 92.7%) in **GPS_Y**

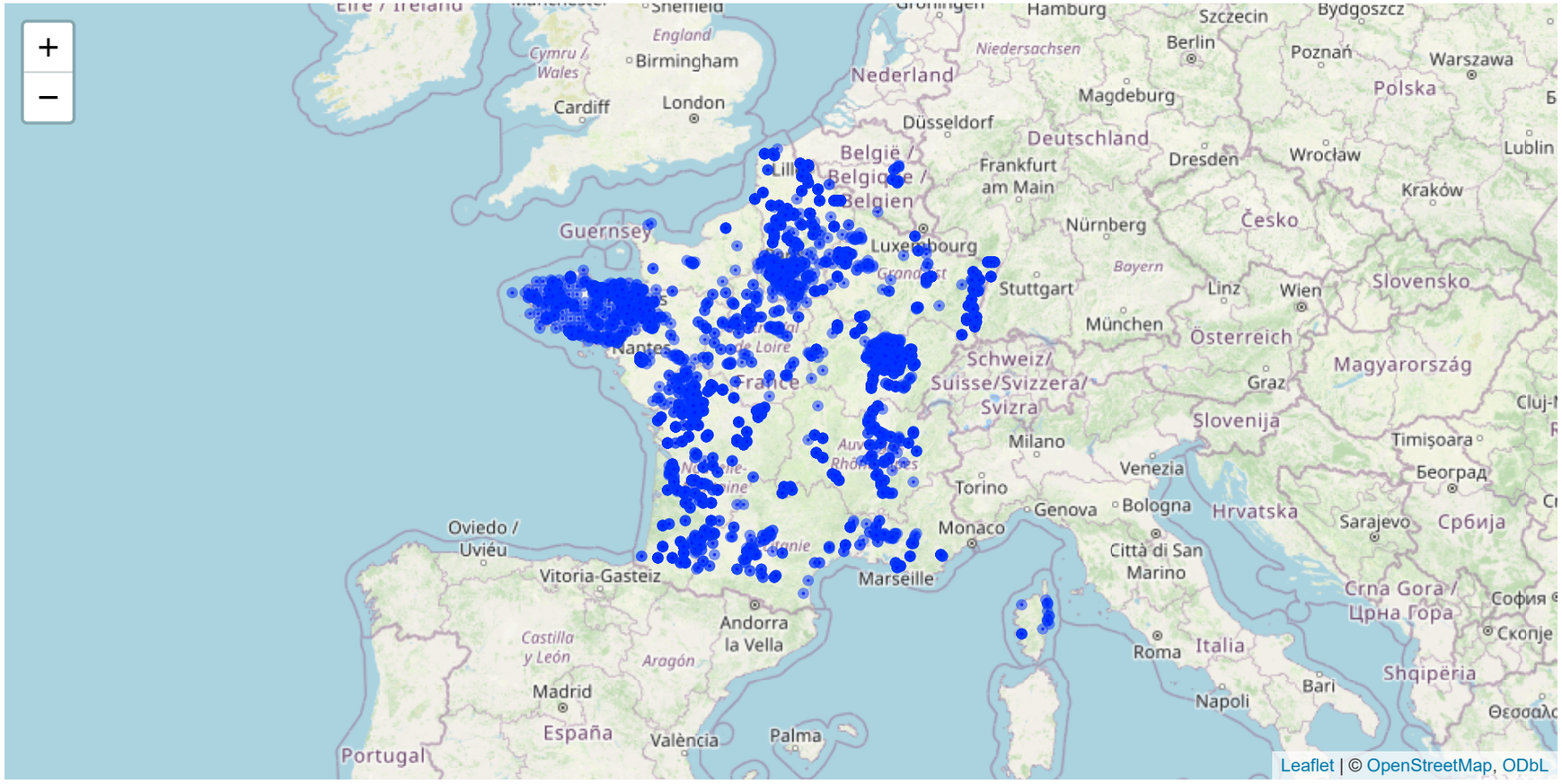
▶ Code

- We delete the *NA* lines in the GPS coordinates
- The database therefore changes from **5520** to **5119** observations.
- Merging database and climat database

▶ Code

4.4 Cartography

▶ Code



- We delete points outside France (**22**)
- The database therefore changes from **5120** to **5098** observations.

4.5 Focus on years

- Cleaning the Annee column
- CR of Annee = **96.1%** (33 levels)

▶ Code

	Numbers
1990	19
1991	23
1992	22
1993	15

	Numbers
1994	29
1995	6
1996	7
1997	8
1998	15
1999	30
2000	24
2001	10
2002	20
2004	9

	Numbers
2005	47
2006	57
2007	78
2008	24
2009	52
2010	67
2011	69
2012	127
2013	285
2014	542

	Numbers
2015	261
2016	508
2017	287
2018	372
2019	506
2020	353
2021	832
2022	344
2023	50

4.6 Focus on protocols

- List of protocols available on the database (5 levels)

▶ Code

	Numbers
F	51
F_HS	872
HS	2940
M	1166
M_HS	69

- Selection of protocols: **F_HS, HS**

▶ Code

	Numbers
F_HS	872
HS	2940

- The database therefore changes from **5098** to **3812** observations.

4.7 Focus on clcm_lvl1

- CR of clcm_lvl1 = **95.4%** (5 levels)

▶ Code

- Merging levels

▶ Code

	Numbers
Forest and semi natural areas	204
Agricultural areas	2732

	Numbers
Artificial surfaces	860
NA's	16

- Update **code_clcm_lvl1**

▶ Code

- For the moment, we will keep the NA of **clcm_lvl1**

4.8 Focus on clcm_lvl2

- CR of clcm_lvl2 = **95.4%** (11 levels)

▶ Code

- Merging levels

▶ Code

	Numbers
Arable land	1496
Artificial, non-agricultural vegetated areas	667

	Numbers
Forests	117
Heterogeneous agricultural areas	107
Industrial, commercial and transport units	168
Mine, dump and construction sites	25
Open spaces with little or no vegetation	1
Pastures	372
Permanent crops	757
Scrub and/or herbaceous vegetation associations	85
NA's	17

4.9 Focus on clcm_lvl3

- CR of clcm_lvl3 = **95.4%** (23 levels)

▶ Code

	Numbers
Agro-forestry areas	89
Airports	44
Beaches, dunes, sands	1
Broad-leaved forest	25
Complex cultivation patterns	13

	Numbers
Coniferous forest	4
Construction sites	21
Fruit trees and berry plantations	18
Green urban areas	648
Industrial or commercial units and public facilities	10
Mixed forest	88
Moors and heathland	7
Natural grasslands	65
Non-irrigated arable land	1493
Other artificial, non-agricultural vegetated areas	12

	Numbers
Other heterogeneous agricultural areas	5
Other mine, dump and construction sites	4
Other scrub and/or herbaceous vegetation associations	1
Pastures, meadows and other permanent grasslands under agricultural use	372
Road and rail networks and associated land	114
Sport and leisure facilities	7
Transitional woodland-shrub	12
Vineyards	739
NA's	20

4.10 Land use selection (clcm_lvl3)

▶ Code

	Numbers
Broad-leaved forest	25
Coniferous forest	4
Green urban areas	648
Mixed forest	88
Natural grasslands	65
Non-irrigated arable land	1493

	Numbers
Pastures, meadows and other permanent grasslands under agricultural use	372
Vineyards	739

- **Maybe, we can merge the three types of forest ?**

4.11 Land use & protocol overview

▶ Code

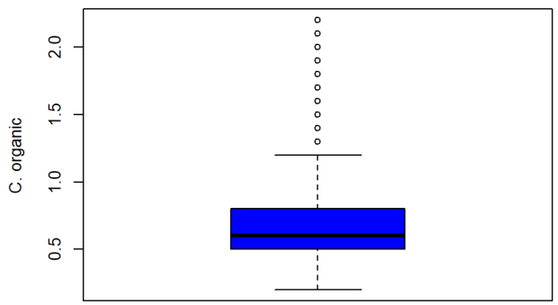
	F_HS	HS
Broad-leaved forest	9	16
Coniferous forest	3	1
Green urban areas	0	648
Mixed forest	11	77
Natural grasslands	3	62
Non-irrigated arable land	276	1217

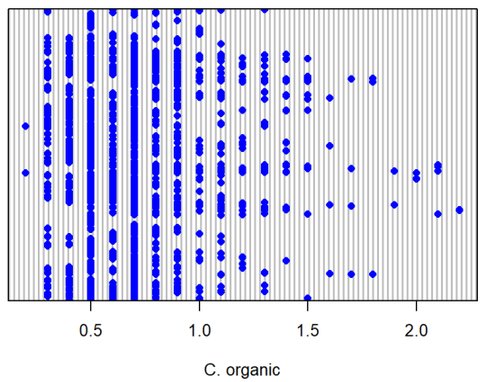
	F_HS	HS
Pastures, meadows and other permanent grasslands under agricultural use	116	256
Vineyards	373	366

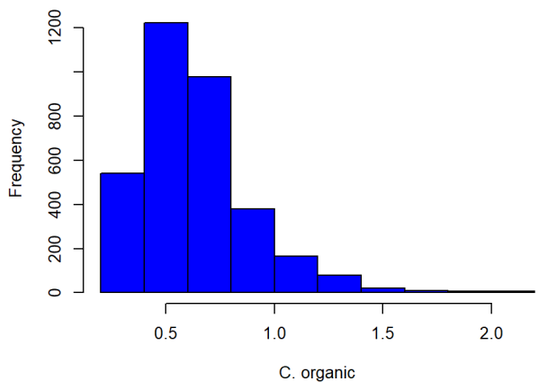
5 Soil data extraction

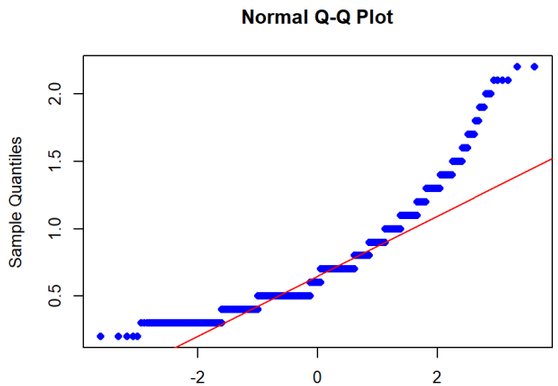
5.1 Soil organic carbone (g/kg)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 0.2000  0.5000  0.6000  0.6641  0.8000  2.2000       9
```



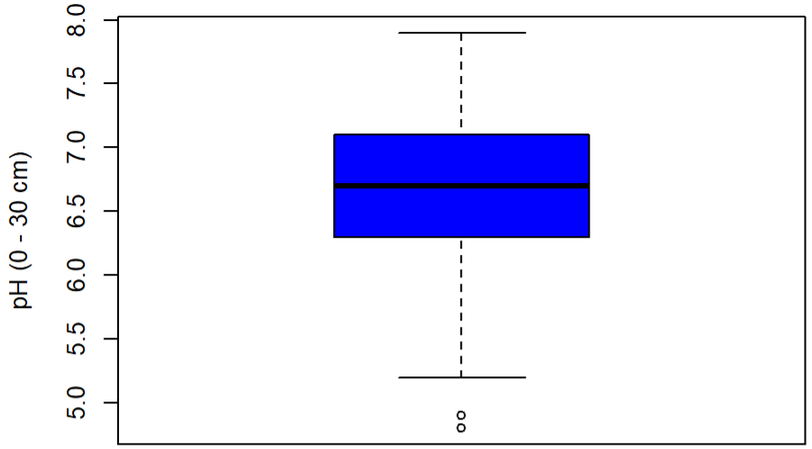


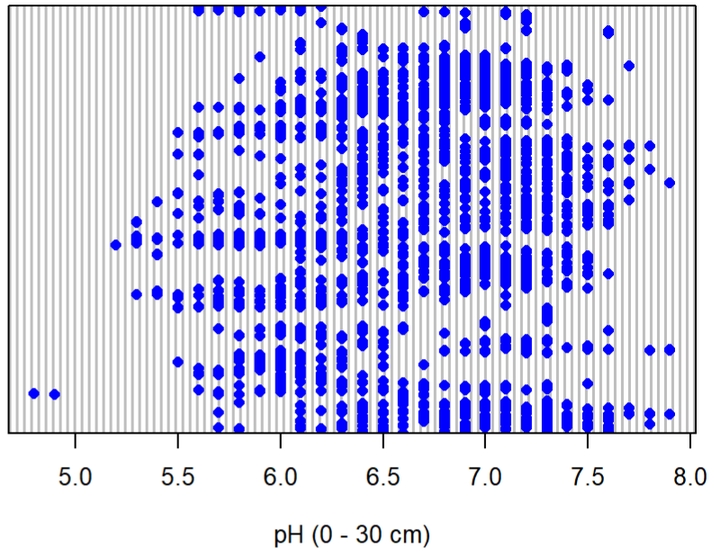


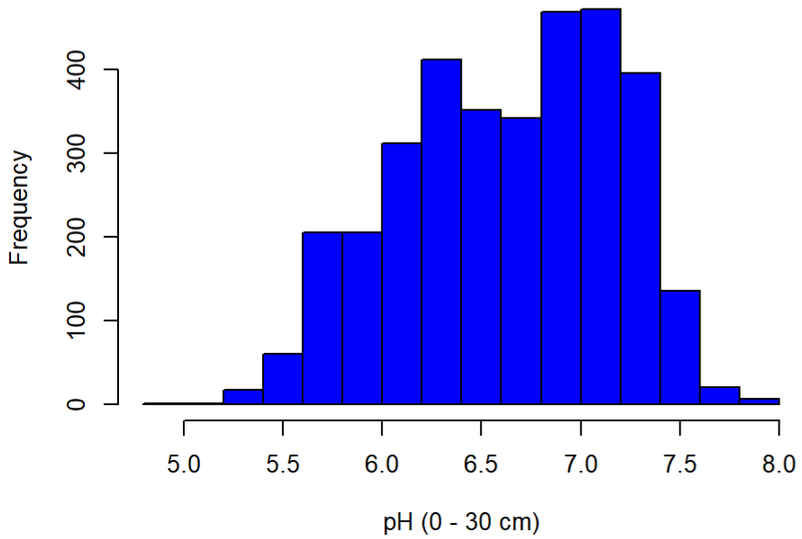


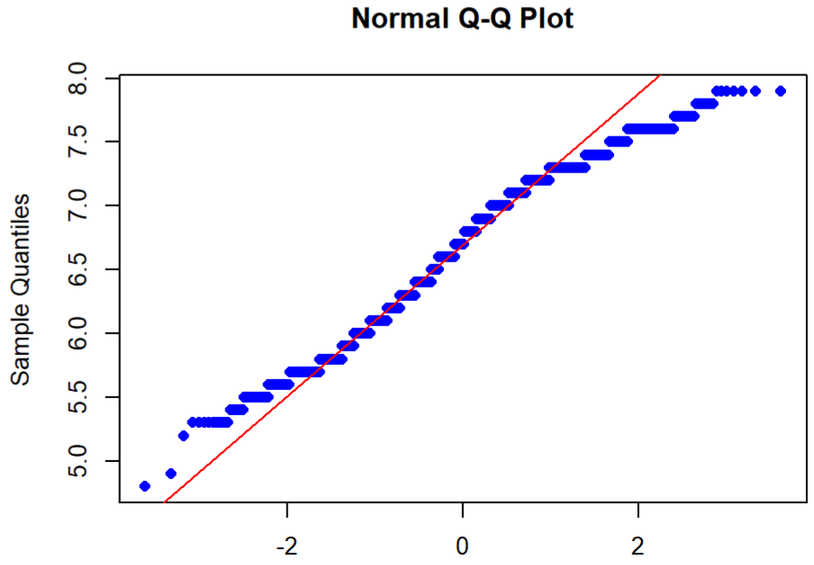
5.2 pH

Extracted values


pH (0 - 30 cm)
5.0
5.5
6.0
6.5
7.0
7.5
8.0


5.0
5.5
6.0
6.5
7.0
7.5
8.0
pH (0 - 30 cm)


Frequency
0
100
200
300
400
5.0
5.5
6.0
6.5
7.0
7.5
8.0
pH (0 - 30 cm)


Normal Q-Q Plot
Sample Quantiles
5.0
5.5
6.0
6.5
7.0
7.5
8.0
-2
0
2

Measured values & extracted values

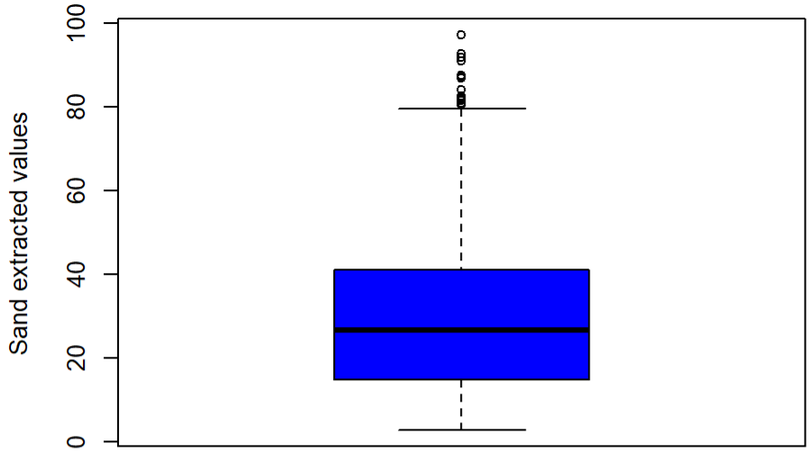
- Clean pH column

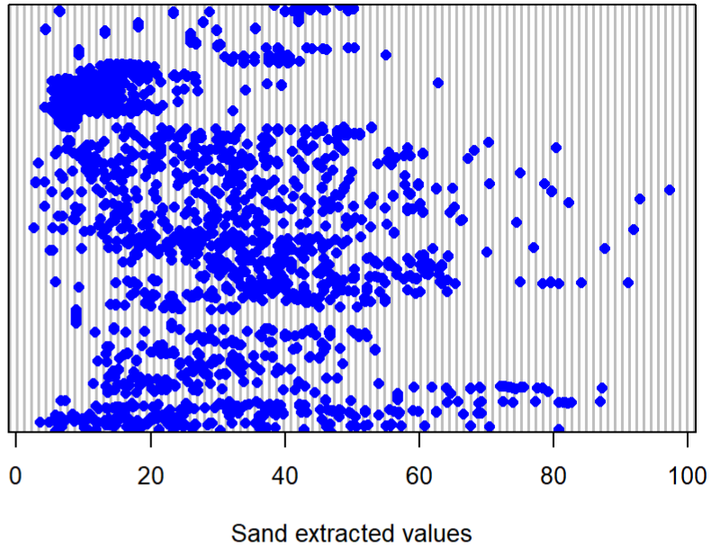
▶ Code

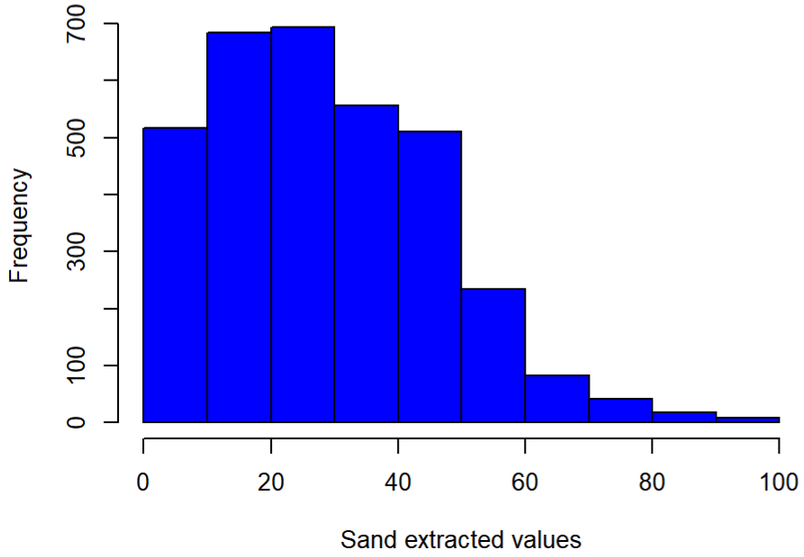
▶ Code

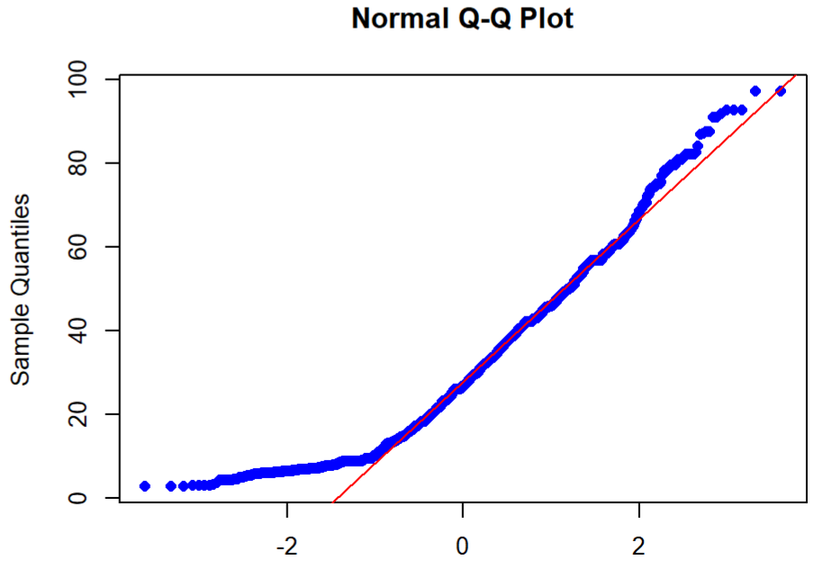
5.3 Sand

Extracted values (g/kg, 0 - 30 cm)


Sand extracted values
0
20
40
60
80
100


0
20
40
60
80
100
Sand extracted values


Frequency
0
100
300
500
700
0
20
40
60
80
100
Sand extracted values


Normal Q-Q Plot
Sample Quantiles
0
20
40
60
80
100
-2
0
2

Measured values & extracted values

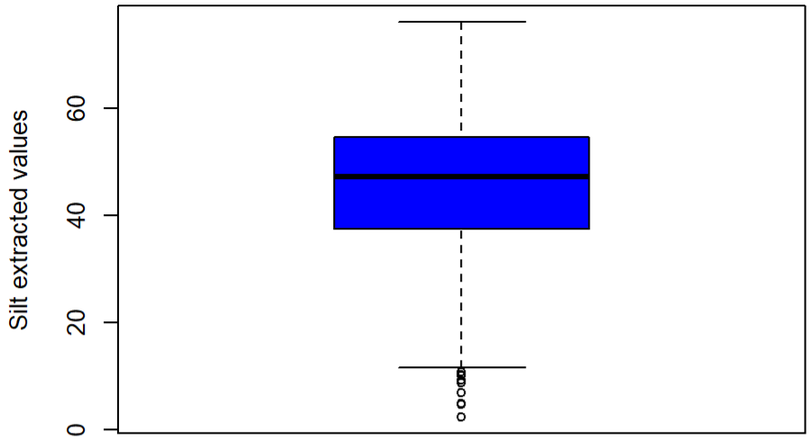
- Clean sand column

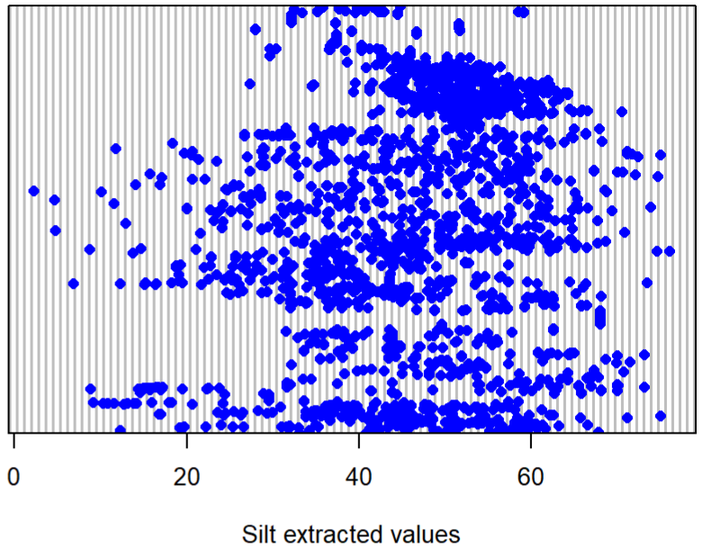
▶ Code

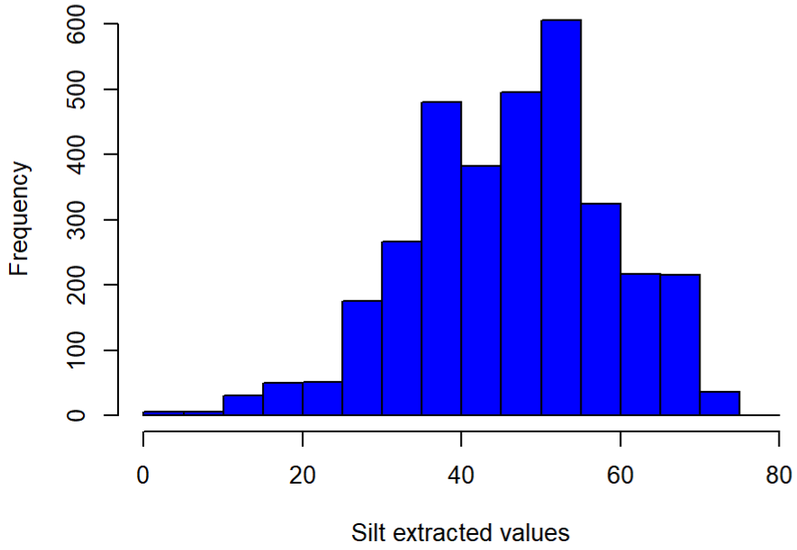
▶ Code

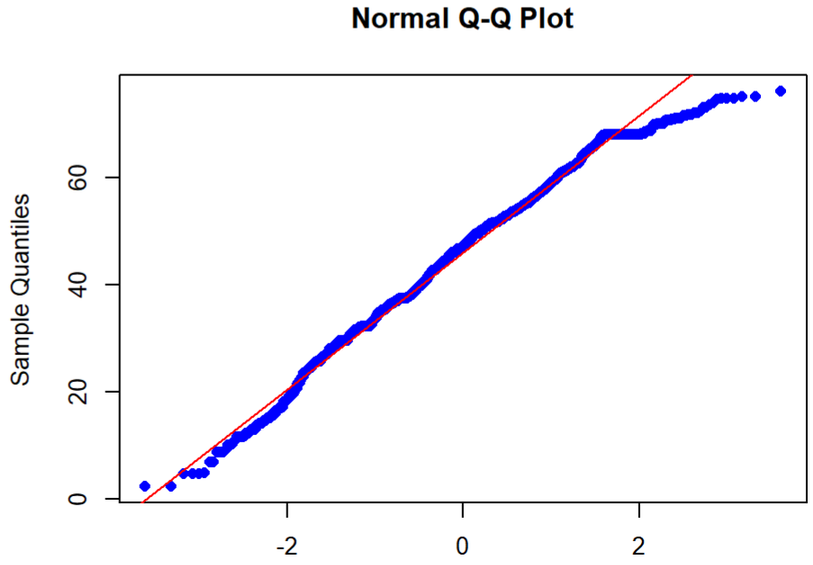
5.4 Silt

Extracted values (g/kg, 0 - 30 cm)


Silt extracted values
0
20
40
60


0
20
40
60
Silt extracted values


Frequency
0
100
200
300
400
500
600
0
20
40
60
80
Silt extracted values


Normal Q-Q Plot
Sample Quantiles
0
20
40
60
-2
0
2

Measured values & extracted values

- Clean silt column

▶ Code

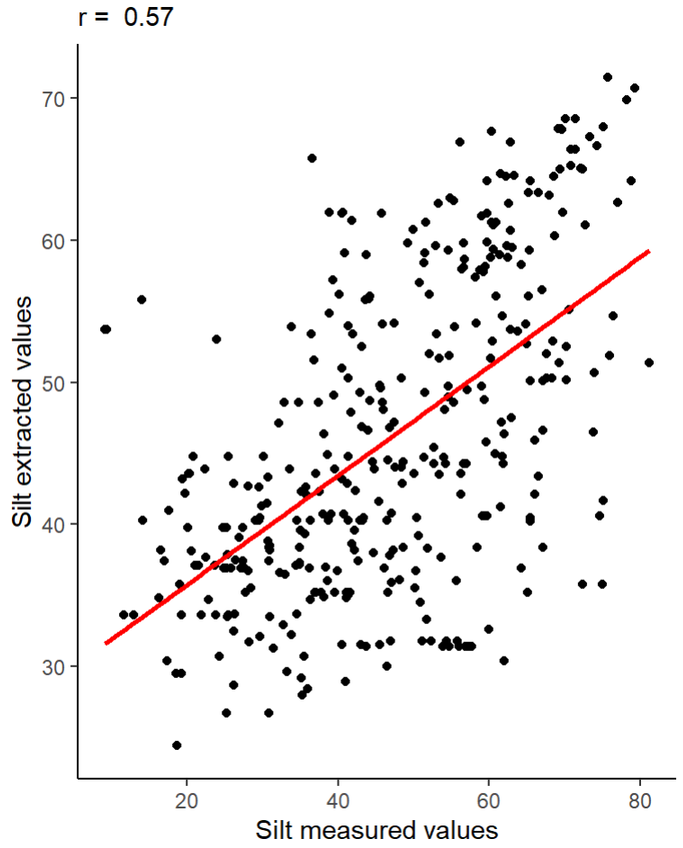
▶ Code

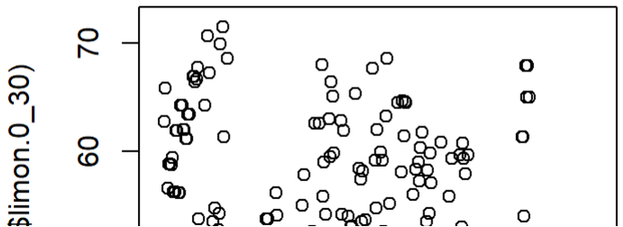
- Method ?
- Depth ?
- Measured values (CR = 28.6%)

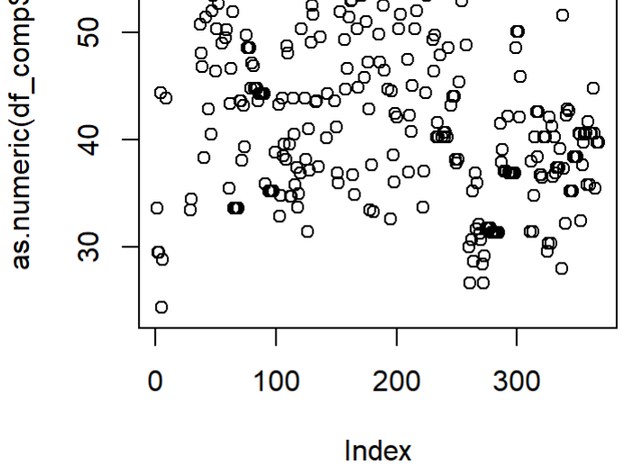
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  9.102  34.947  46.550  46.810  59.850  81.200
```

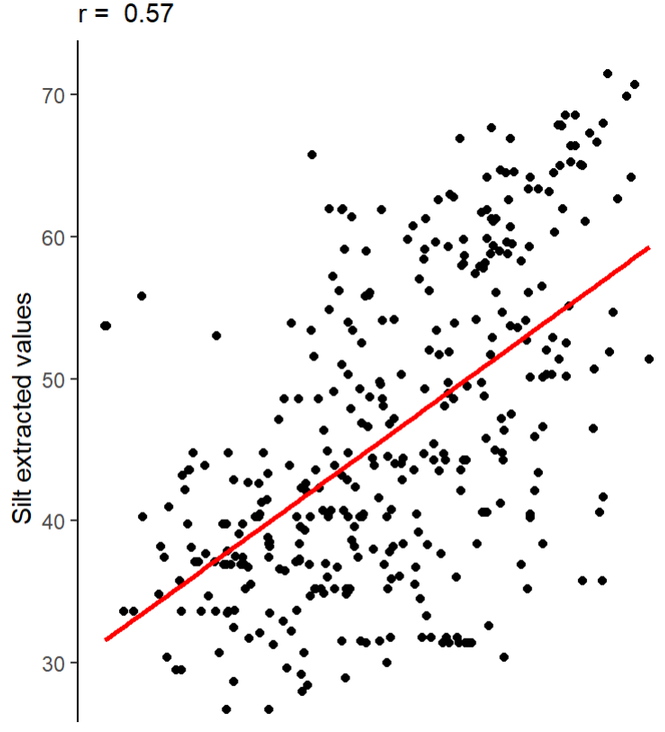
- Extracted values

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  24.40   37.10   43.90   46.06   54.20   71.50
```





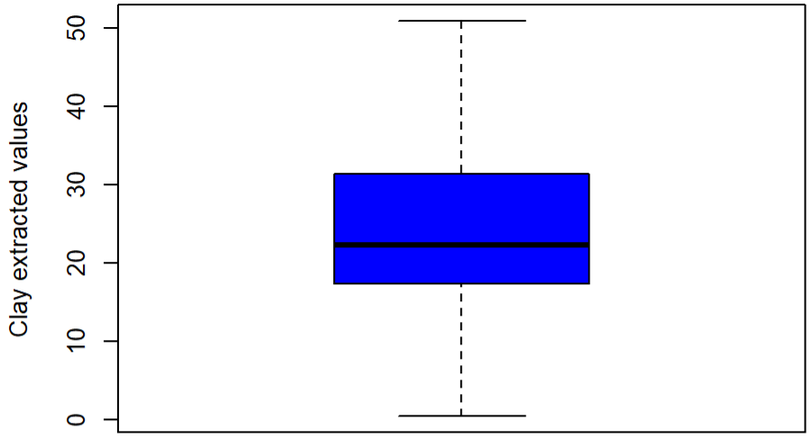

as.numeric(df_comp$
30
40
50
0
100
200
300
Index

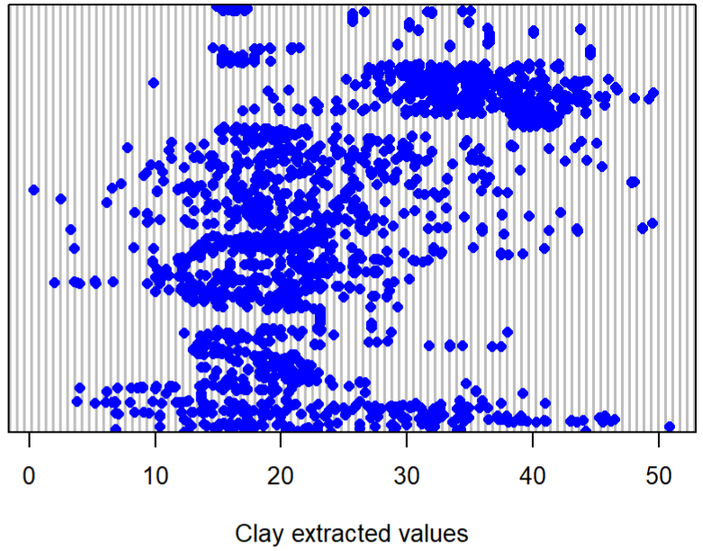

r = 0.57
Silt extracted values
30
40
50
60
70

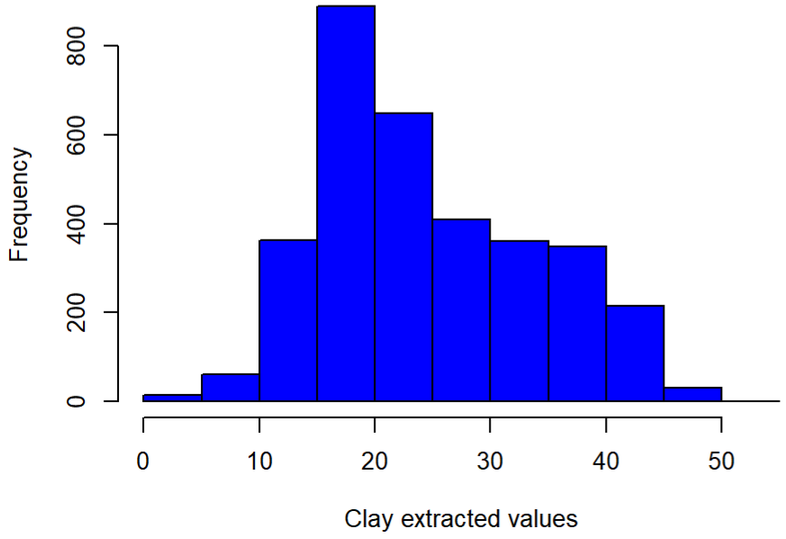
5.5 Clay

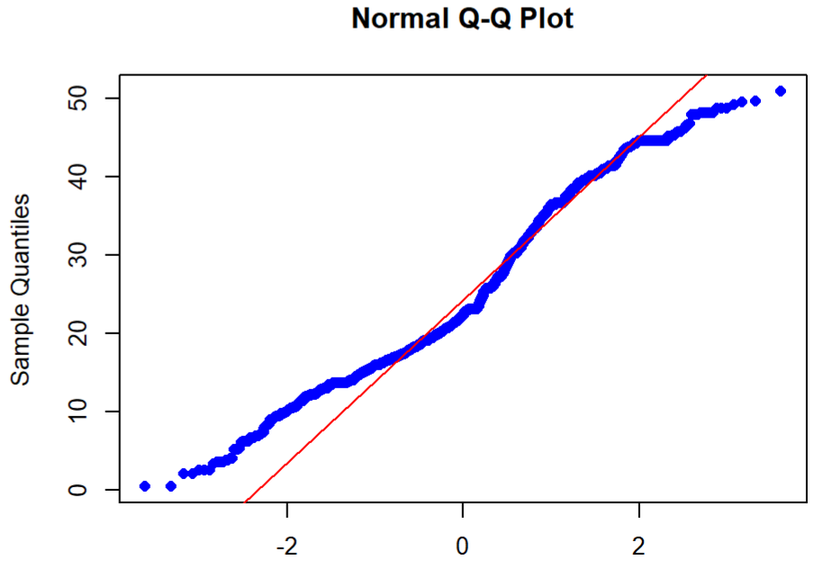
Extracted values (g/kg, 0 - 30 cm)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.40   17.30   22.30   24.48   31.32   50.90      73
```


Clay extracted values
0
10
20
30
40
50


0
10
20
30
40
50
Clay extracted values


Frequency
0
200
400
600
800
0
10
20
30
40
50
Clay extracted values


Normal Q-Q Plot
Sample Quantiles
0
10
20
30
40
50
-2
0
2

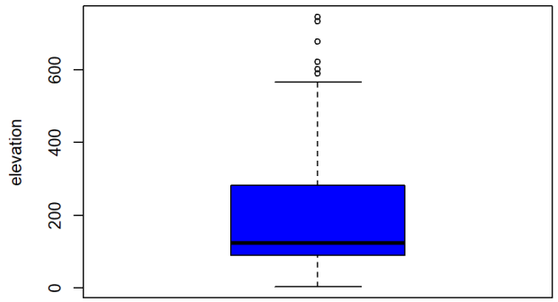
Measured values & extracted values - Clean clay column

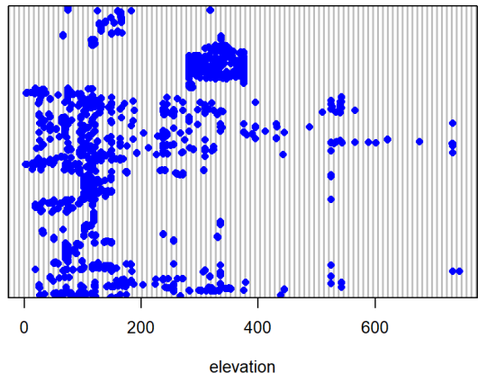
- Code

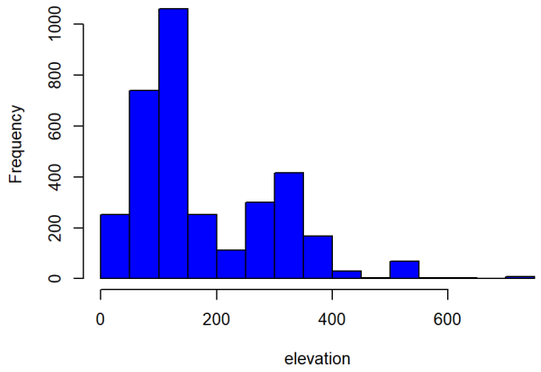
- Code

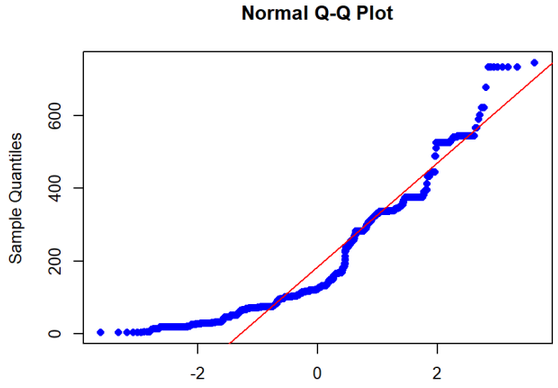
5.6 Elevation

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  2.763  88.661 123.132 176.089 281.520 744.715 
```



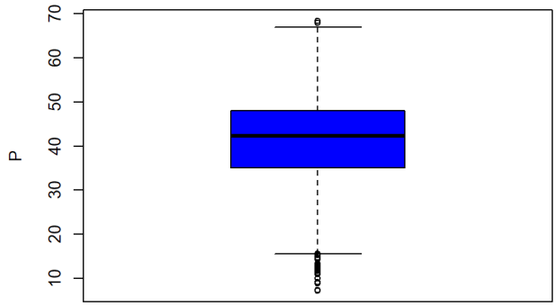


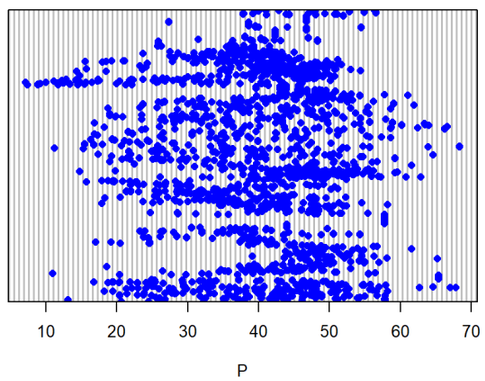


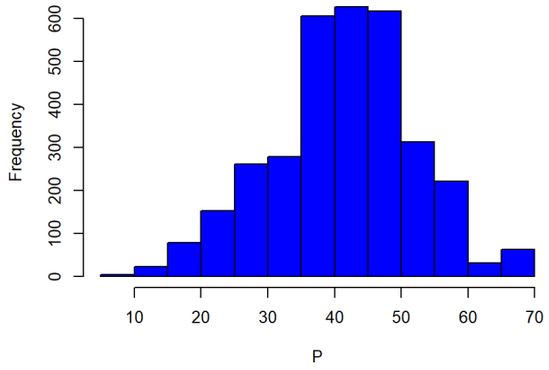


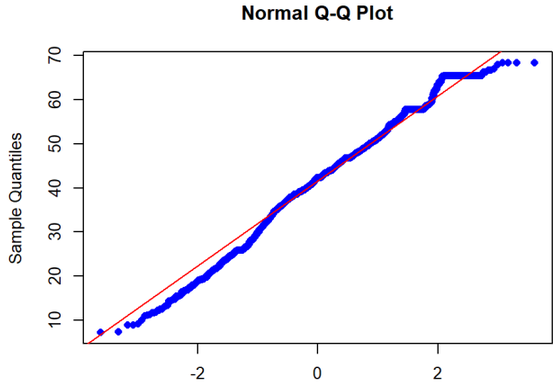
5.7 Phosphore (P, mg/kg)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
7.182	35.083	42.377	41.286	48.091	68.392	127



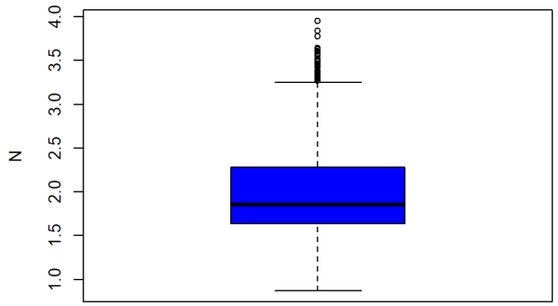


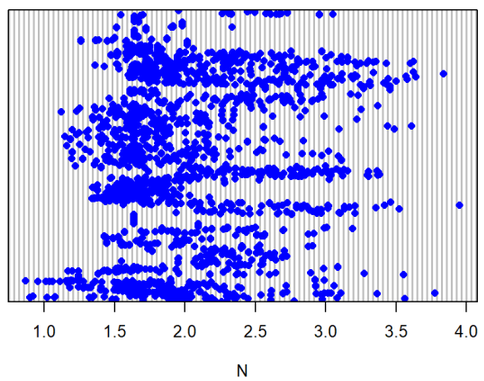


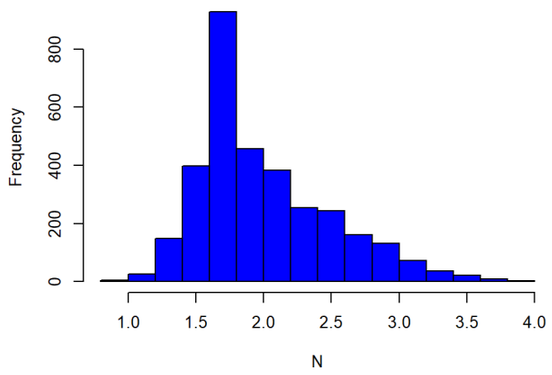


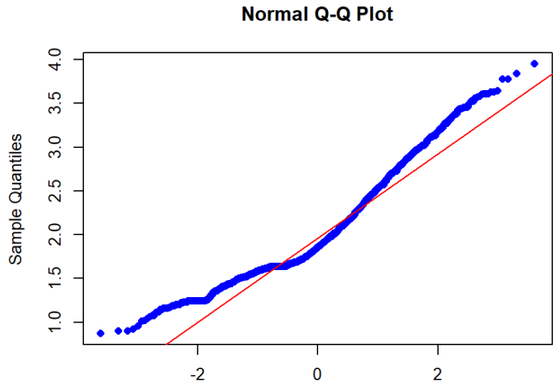
5.8 Azote (N, g/kg)

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 0.8697  1.6357  1.8565  1.9894  2.2847  3.9521     128
```



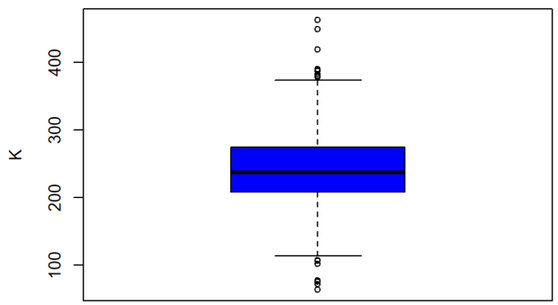


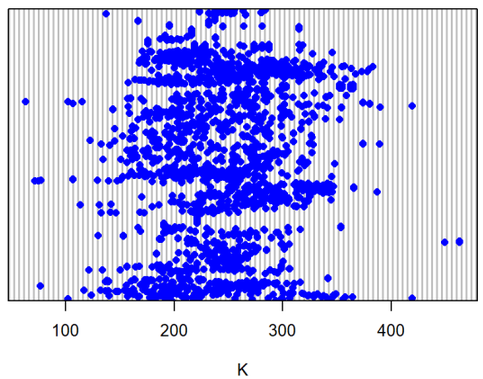


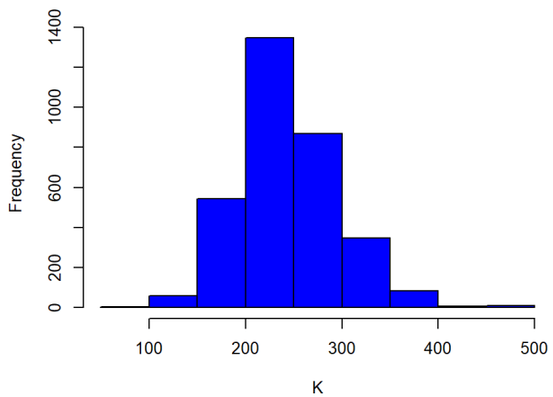


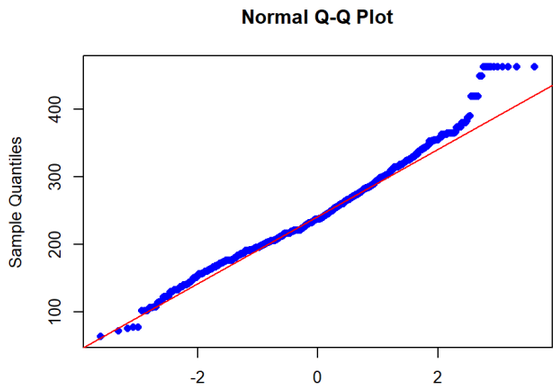
5.9 Potassium (K, mg/kg)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
63.28	207.85	237.38	243.61	275.03	463.07	127



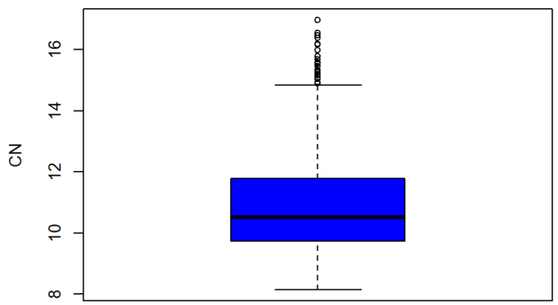


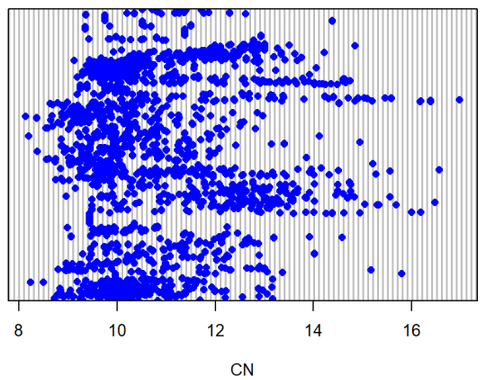


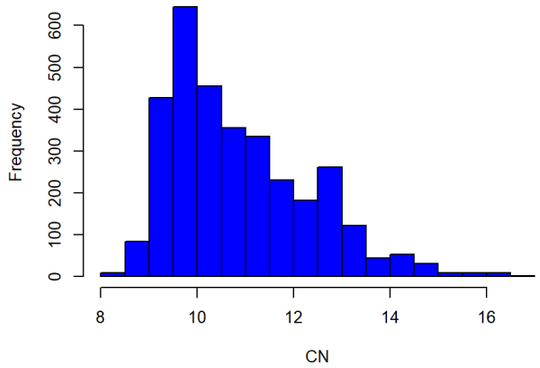


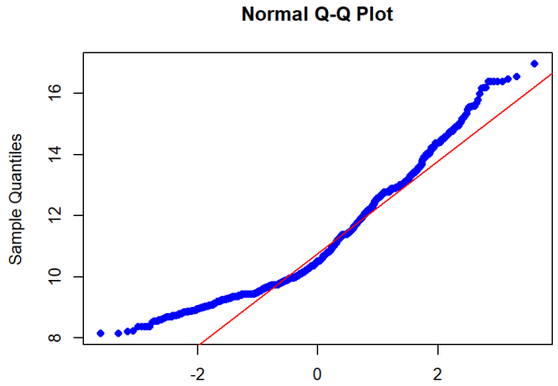
5.10 C/N

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  8.140   9.746  10.512  10.885  11.788  16.979     128
```







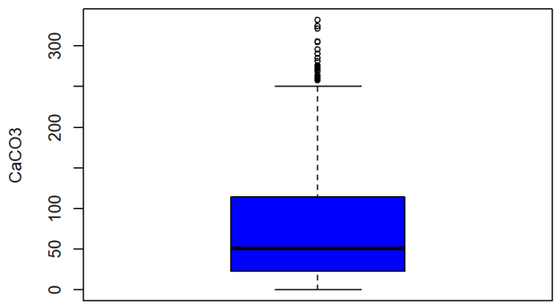


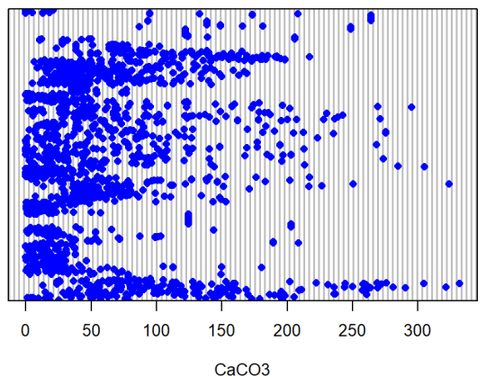
5.11 Capacité d’échange de cations (CEC, cmol/kg)

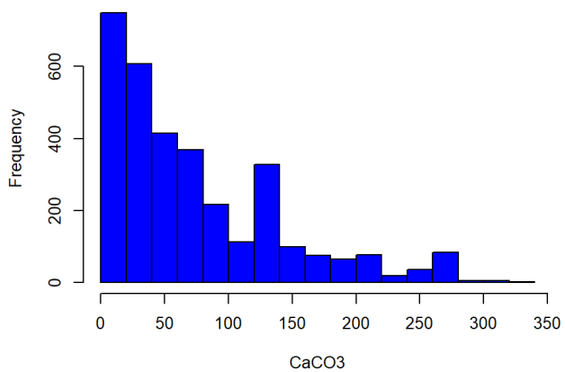
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  6.206  11.909  15.586  15.973  20.063  31.112     128
```

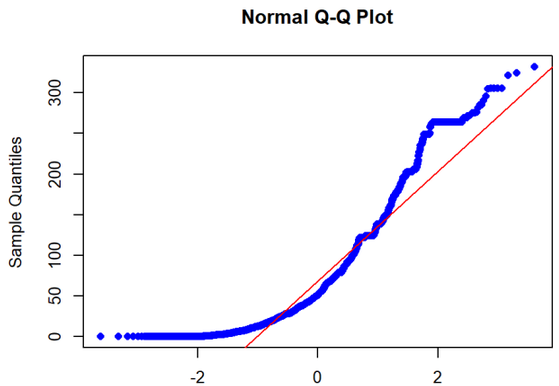
5.12 Carbonates de calcium (CaCO3, g/kg)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0.00	22.95	51.07	73.63	114.06	332.01	127









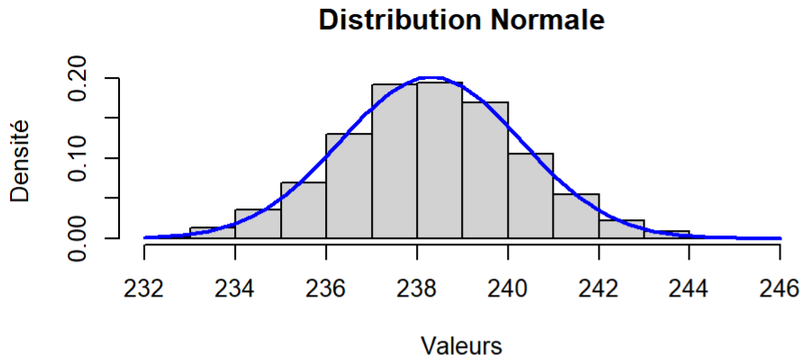
6 Analyses explorations

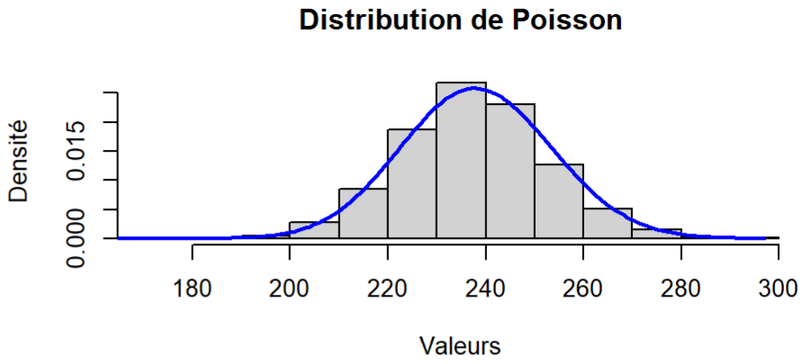
Réduction du jeu de donnée

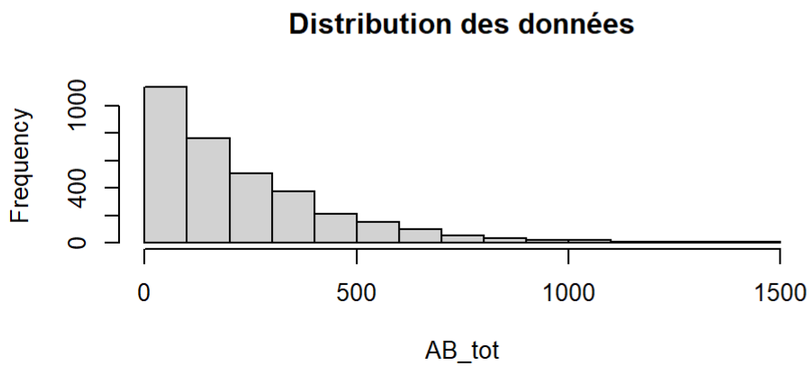
▶ Code

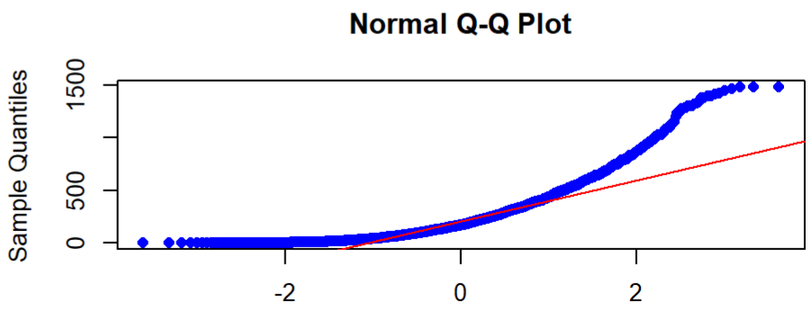
7 Distributions des variables de réponse

7.1 Distributions: AB_tot

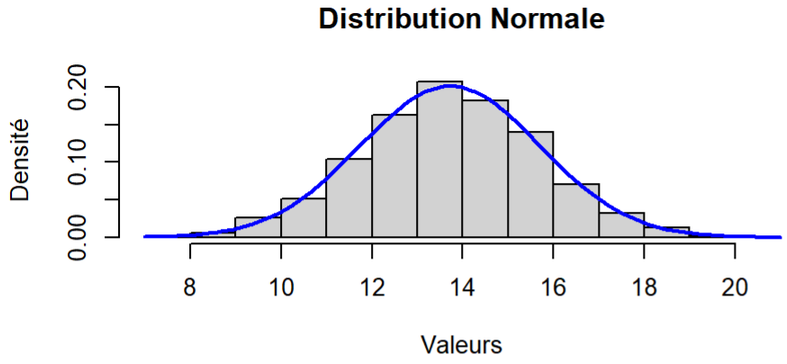


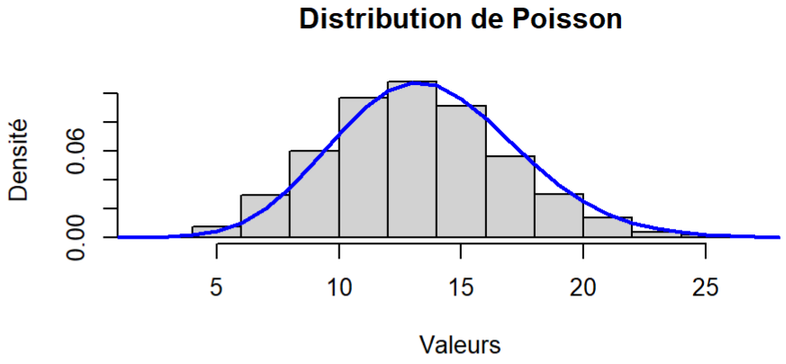


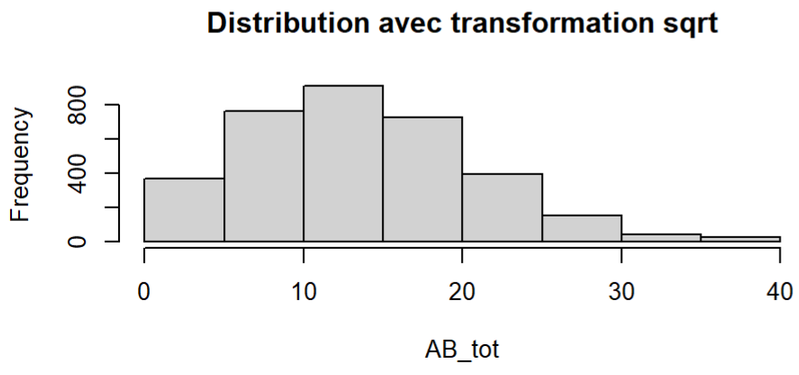


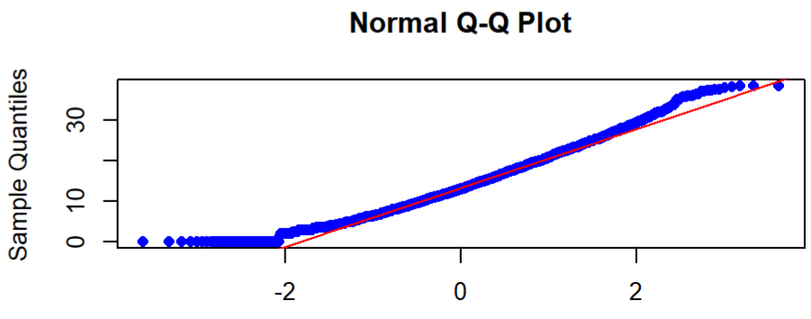


- Transformation sqrt

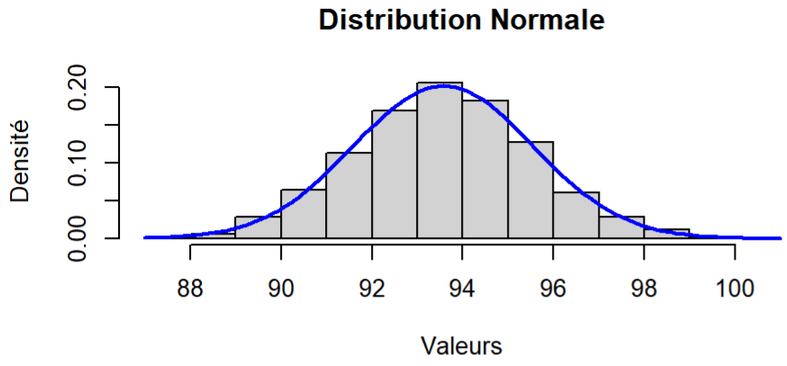

Distribution Normale
Densité
0.00
0.10
0.20
8
10
12
14
16
18
20
Valeurs

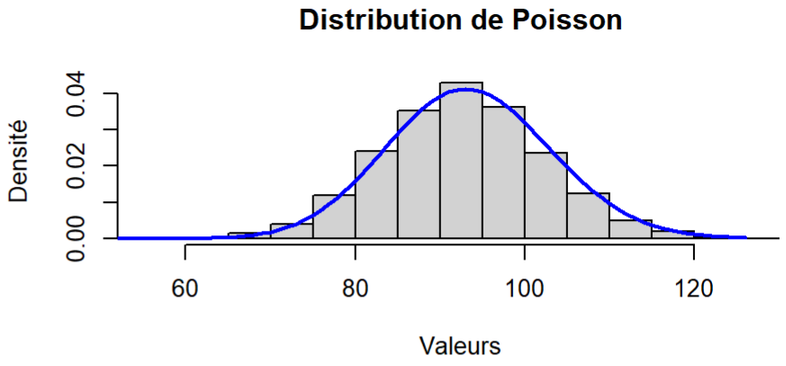

Distribution de Poisson
Densité
0.00
0.06
5
10
15
20
25
Valeurs

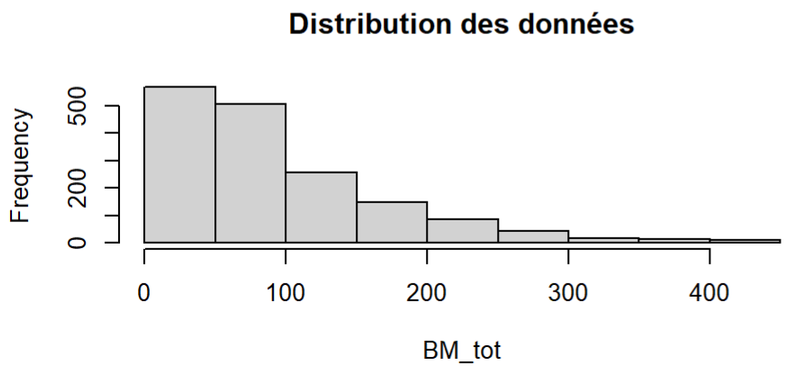

Distribution avec transformation sqrt
Frequency
0
400
800
0
10
20
30
40
AB_tot

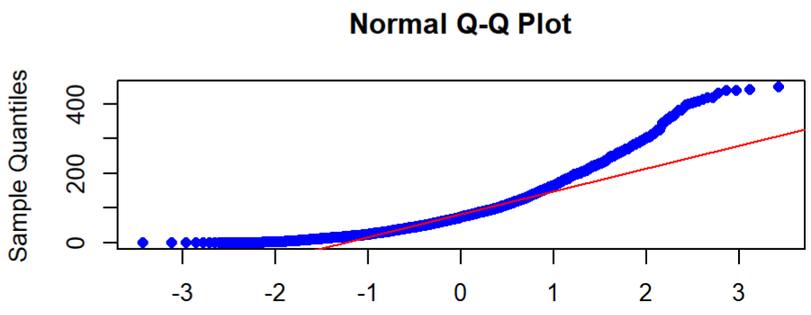

Normal Q-Q Plot
Sample Quantiles
0
10
30
-2
0
2

7.2 Distributions: BM_tot

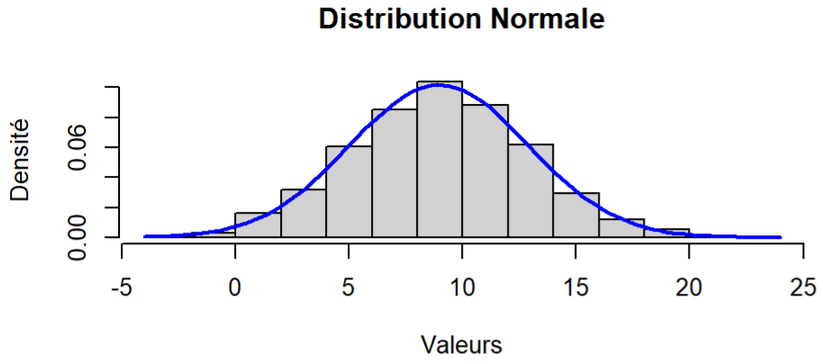


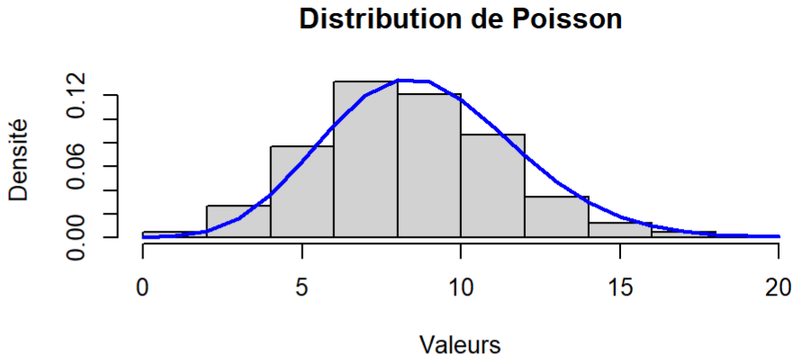


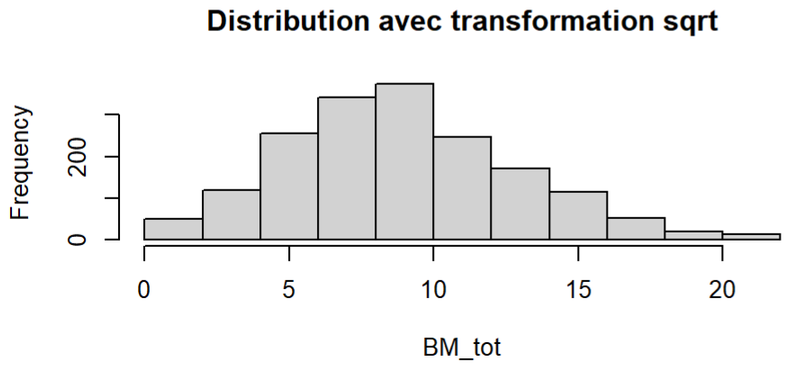


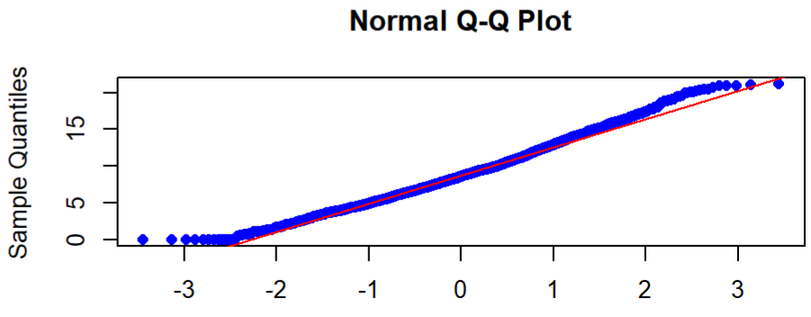


- Transformation sqrt

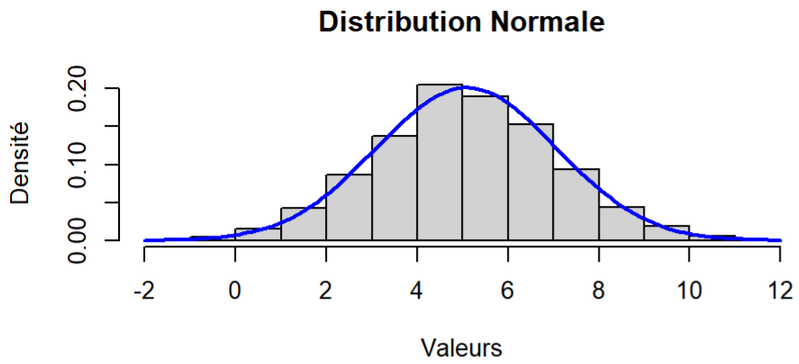

Distribution Normale
Densité
0.00
0.06
-5
0
5
10
15
20
25
Valeurs

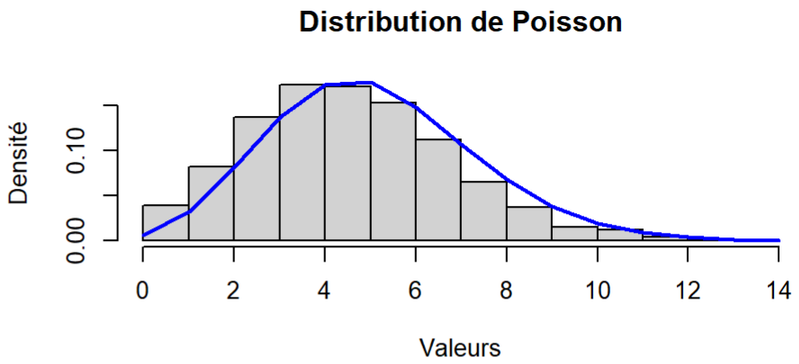

Distribution de Poisson
Densité
0.00
0.06
0.12
0
5
10
15
20
Valeurs

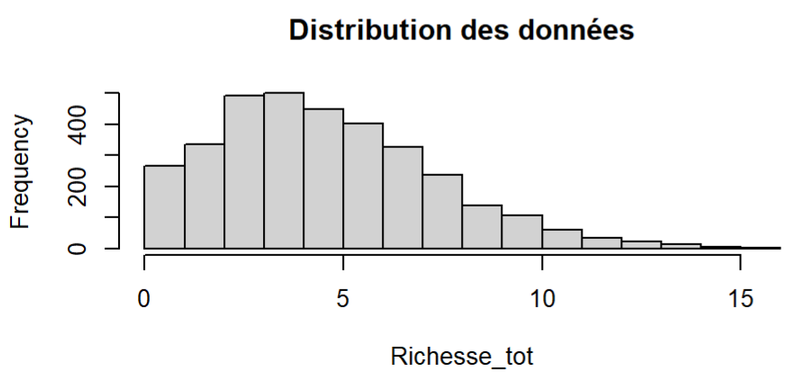

Distribution avec transformation sqrt
Frequency
0
200
0
5
10
15
20
BM_tot

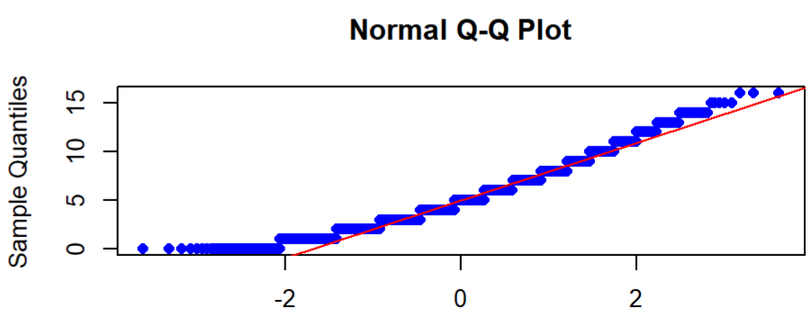

Normal Q-Q Plot
Sample Quantiles
0
5
15
-3
-2
-1
0
1
2
3

7.3 Distributions: Richesse_tot

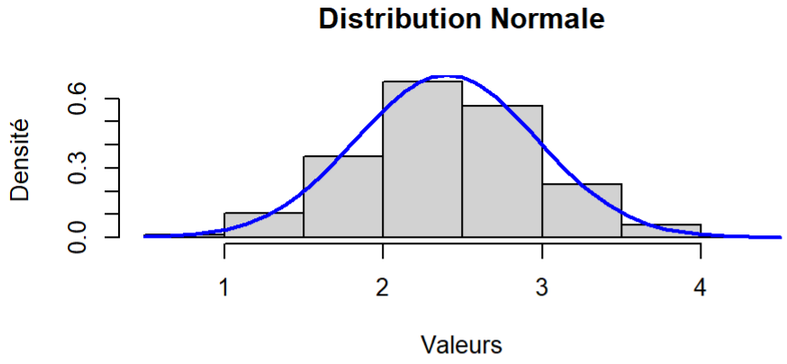


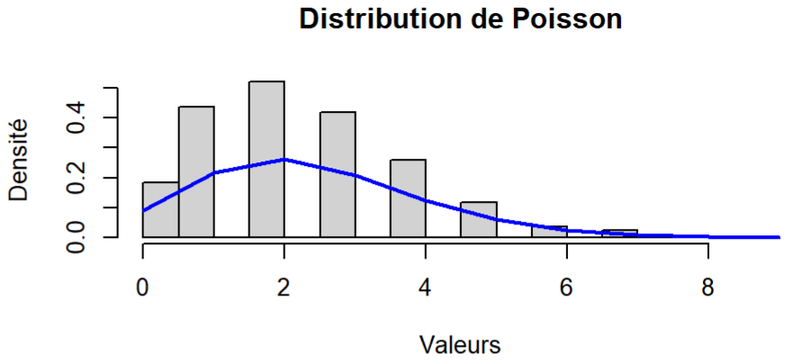


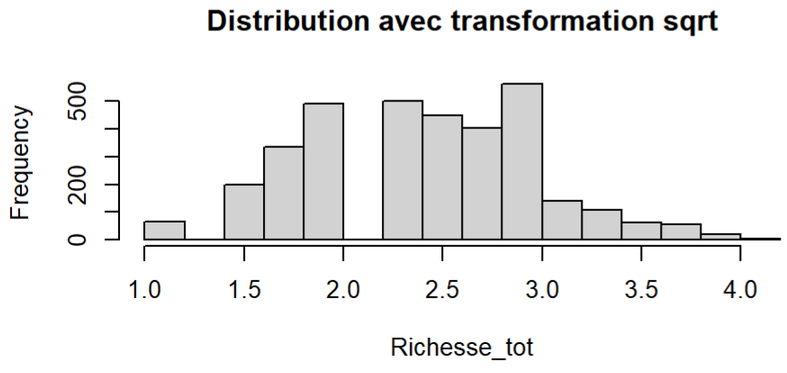


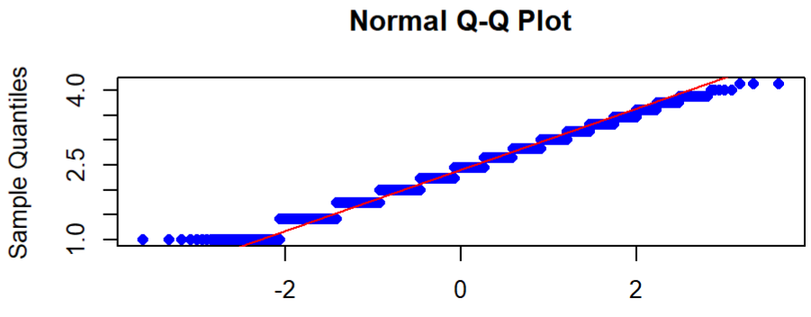


- Transformation sqrt





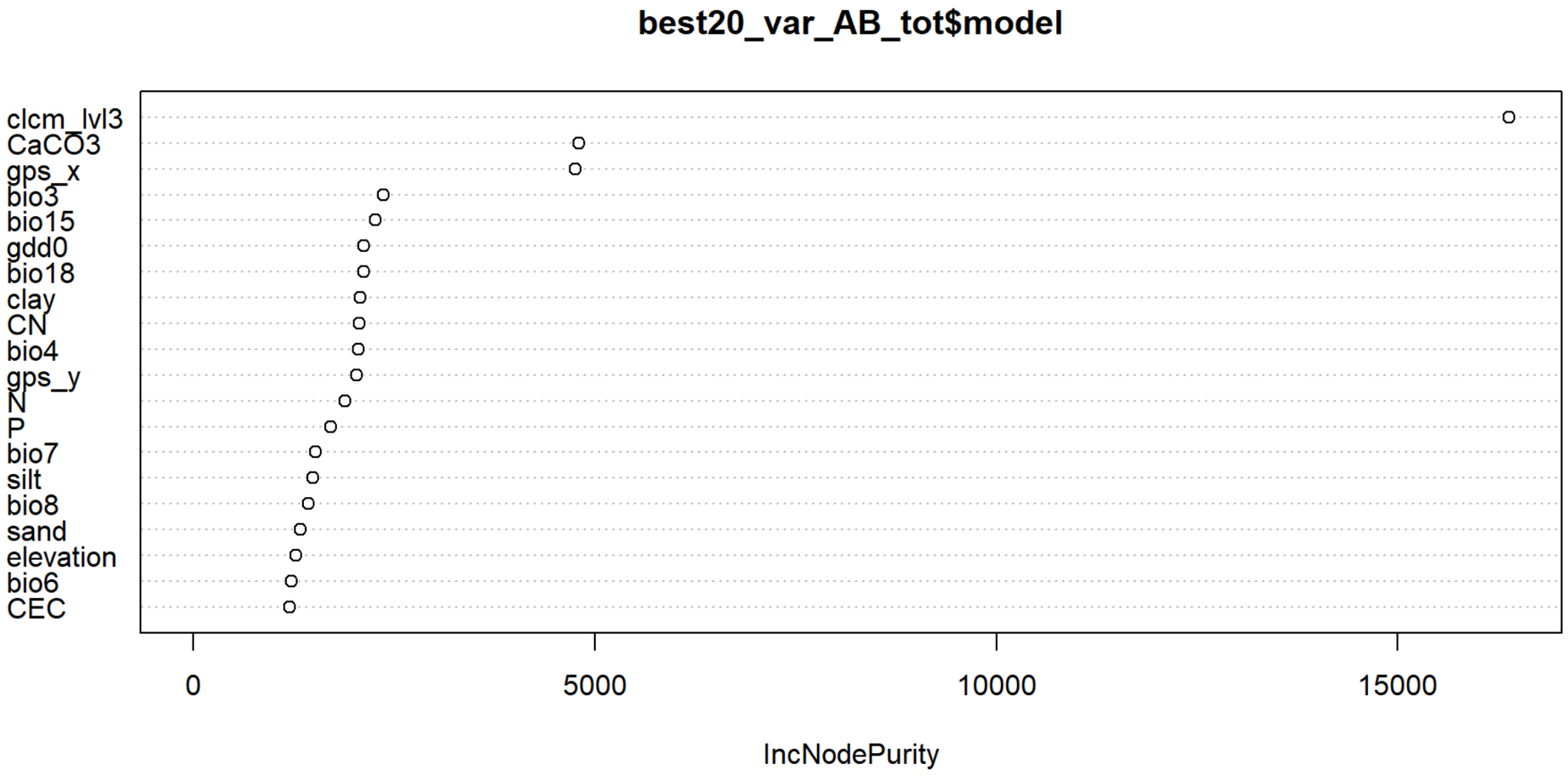




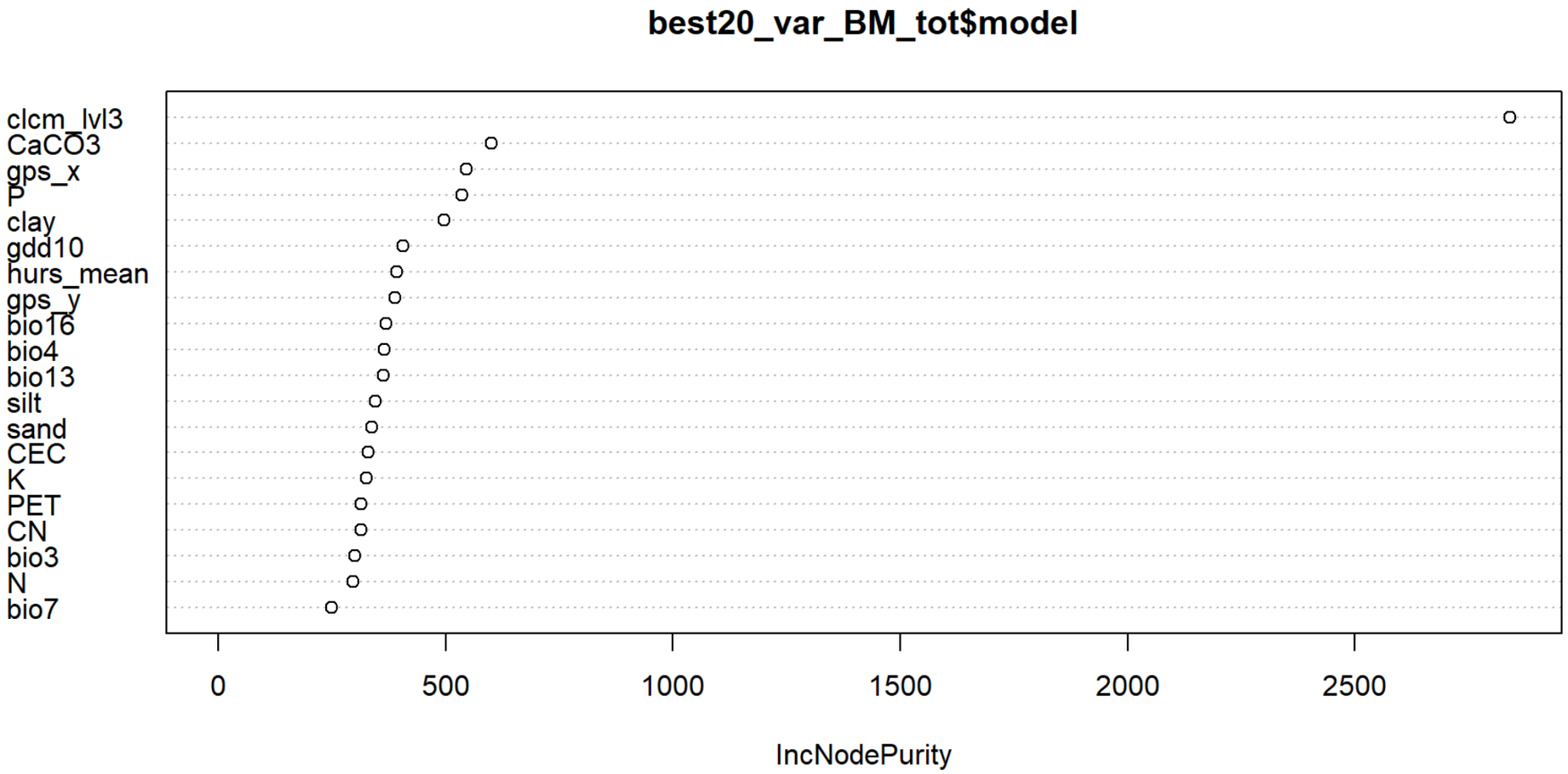
Tranformation non satisfaisante

8 Importance des variables

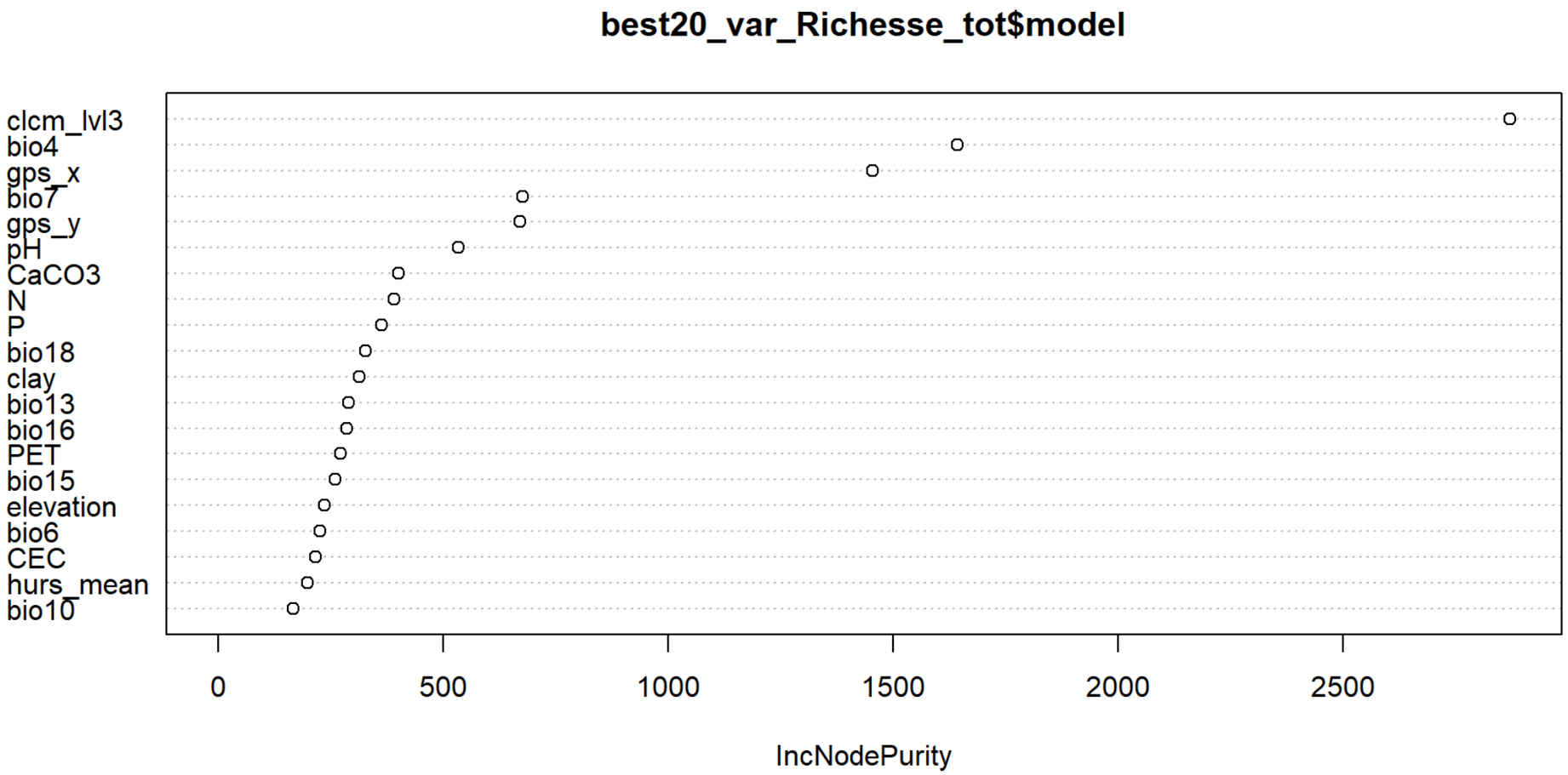
8.1 Importance des variables pour AB_tot



8.2 Importance des variables pour BM_tot



8.3 Importance des variables pour Richesse tot



9 Modélisation

▶ Code

AB_tot :

- Data partition (3219, 26):
  - train data (80 %) = 2353, 26
  - test data (20 %) = 866, 26
- Nombre de simulation = 30

Richesse_tot :

- Data partition (1654, 26):
  - train data (80 %) = 1212, 26
  - test data (20 %) = 442, 26
- Nombre de simulation = 30

Richesse_tot :

- Data partition (3268, 26):
  - train data (80 %) = 2390, 26
  - test data (20 %) = 878, 26
- Nombre de simulation = 30

9.1 GLM

- Code

9.2 GAM

- Code

9.3 RF

- Évaluer le modèle avec le paramètre par défaut

- Tuning the RF model par grid
- ntree = $100, 300, 500, 700, 900, 1000, 1300, 1500, 1700, 2000$
- mtry = $2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24$
- maxnodes = $10, 20, 30, 40, 50, 60, 70, 80, 90, 100$

Nombre totale de model = $ntree * mtry * maxnode = 960$

- Validation des models sur les données de test

▶ Code

9.4 GBM

- Évaluer le modèle avec le paramètre par défaut

- Tuning the GBM model par grid
- n.trees = $1000, 1500, 1700, 2000, 3000$
- shrinkage = $2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24$
- interaction.depth = $3, 5, 6, 8, 10$
- n.minobsinnode = $2, 5, 10, 30, 50, 70$

Nombre totale de model =
$n.trees * shrinkage * interaction.depth * n.minobsinnode = 900$

- Validation des models sur les données de test

▶ Code

9.5 Compilation pour chaque algoritme

- GLM • GAM • RF • GBM • ANN AB_tot

```
Epoch 1/100
30/30 - 2s - loss: 154.5313 - mae: 10.3588 - val_loss: 74.2300 - val_mae: 6.7177 - 2s/epoch -
53ms/step
Epoch 2/100
30/30 - 0s - loss: 77.2036 - mae: 6.9098 - val_loss: 40.7713 - val_mae: 4.7978 - 284ms/epoch -
9ms/step
Epoch 3/100
30/30 - 0s - loss: 66.7403 - mae: 6.3957 - val_loss: 36.4680 - val_mae: 4.5783 - 157ms/epoch -
5ms/step
Epoch 4/100
30/30 - 0s - loss: 65.7970 - mae: 6.4241 - val_loss: 37.1321 - val_mae: 4.6334 - 136ms/epoch -
5ms/step
Epoch 5/100
30/30 - 0s - loss: 61.8237 - mae: 6.1633 - val_loss: 36.8751 - val_mae: 4.6338 - 152ms/epoch -
5ms/step
Epoch 6/100
28/28 - 0s - 154ms/epoch - 6ms/step
```

- ANN BM_tot

```
Epoch 1/100
31/31 - 1s - loss: 90.9128 - mae: 8.6923 - val_loss: 93.1033 - val_mae: 8.7465 - 1s/epoch -
38ms/step
Epoch 2/100
31/31 - 0s - loss: 75.3880 - mae: 7.7322 - val_loss: 67.1164 - val_mae: 7.1876 - 134ms/epoch -
4ms/step
Epoch 3/100
31/31 - 0s - loss: 46.7940 - mae: 5.6455 - val_loss: 32.8317 - val_mae: 4.7379 - 125ms/epoch -
4ms/step
Epoch 4/100
31/31 - 0s - loss: 30.3963 - mae: 4.3875 - val_loss: 23.7723 - val_mae: 3.8818 - 120ms/epoch -
4ms/step
Epoch 5/100
31/31 - 0s - loss: 26.5200 - mae: 4.0852 - val_loss: 19.1167 - val_mae: 3.4201 - 123ms/epoch -
4ms/step
Epoch 6/100
14/14 - 0s - 92ms/epoch - 7ms/step
```

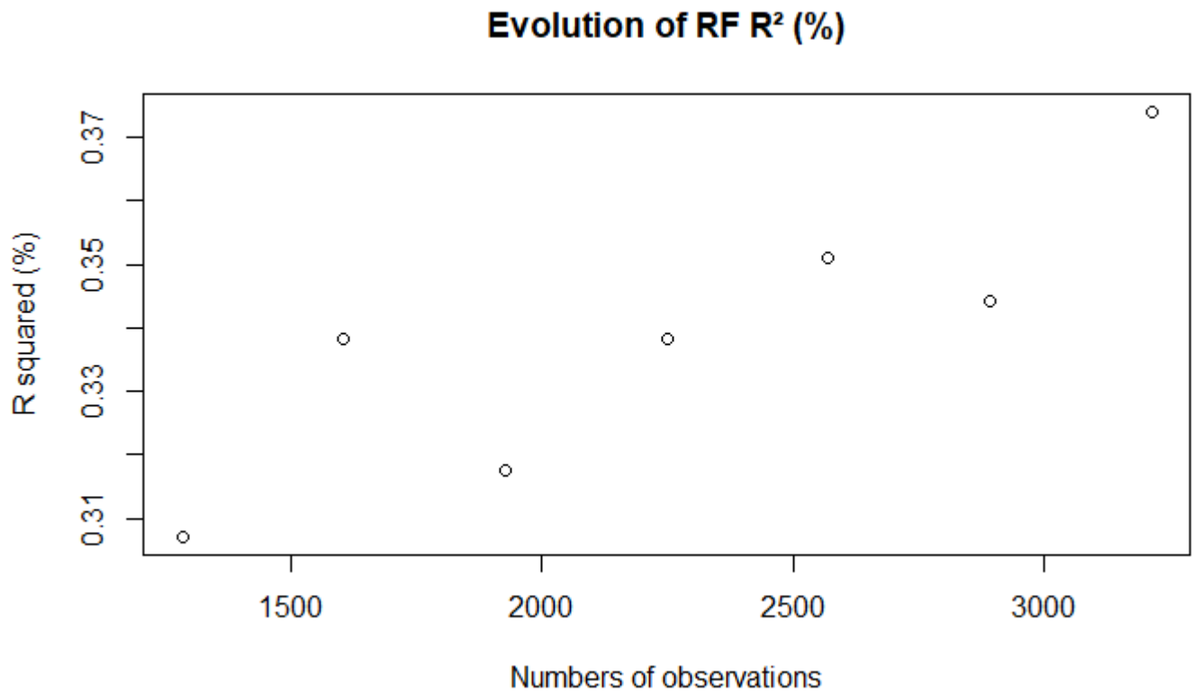
- ANN Richesse_tot

```
Epoch 1/100
30/30 - 1s - loss: 24.4488 - mae: 4.1566 - val_loss: 23.5875 - val_mae: 4.0806 - 1s/epoch -
39ms/step
Epoch 2/100
30/30 - 0s - loss: 16.5966 - mae: 3.2396 - val_loss: 15.7100 - val_mae: 3.2281 - 129ms/epoch -
4ms/step
Epoch 3/100
30/30 - 0s - loss: 14.2293 - mae: 2.9395 - val_loss: 14.4026 - val_mae: 3.0849 - 125ms/epoch -
```

```
4ms/step
Epoch 4/100
30/30 - 0s - loss: 12.6673 - mae: 2.7648 - val_loss: 13.1256 - val_mae: 2.9510 - 139ms/epoch -
5ms/step
Epoch 5/100
30/30 - 0s - loss: 12.6827 - mae: 2.7501 - val_loss: 13.1620 - val_mae: 2.9360 - 136ms/epoch -
5ms/step
28/28 - 0s - 115ms/epoch - 4ms/step
```

10 Resultats

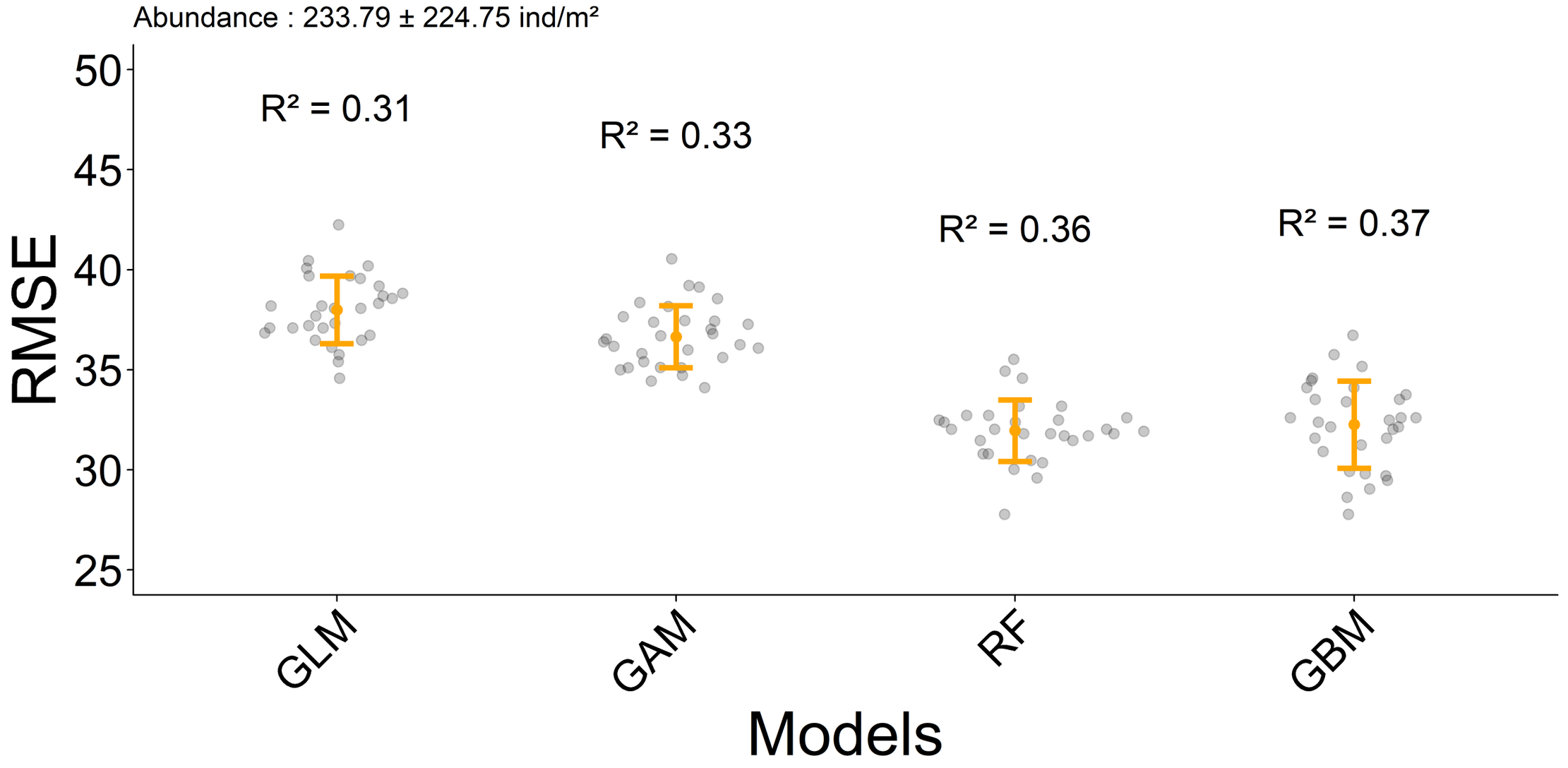
10.1 Sansibilité au nombre d’observation



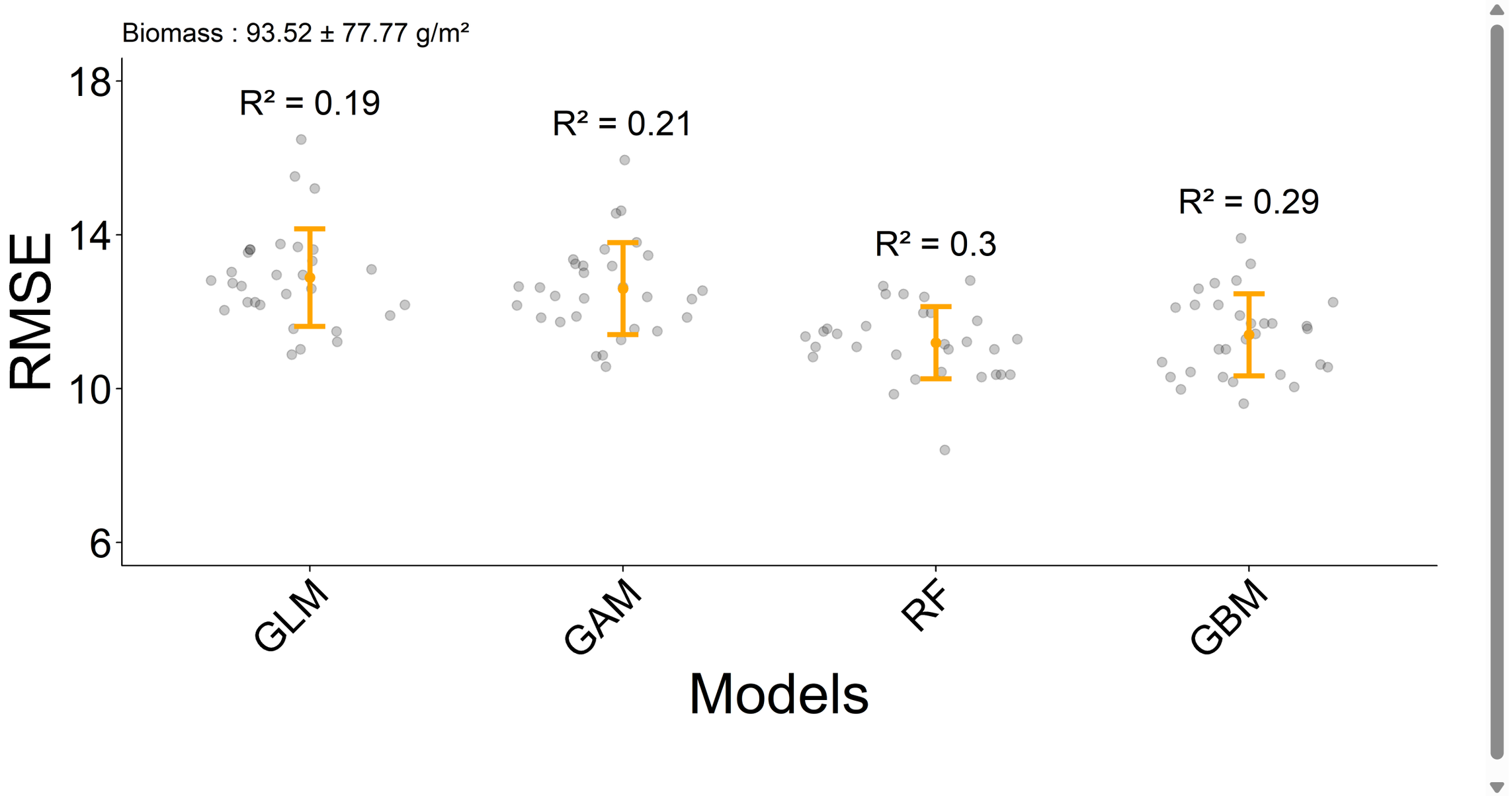
{width=“1200”,aligne=“center”}

10.2 RMSE sur le JDD test

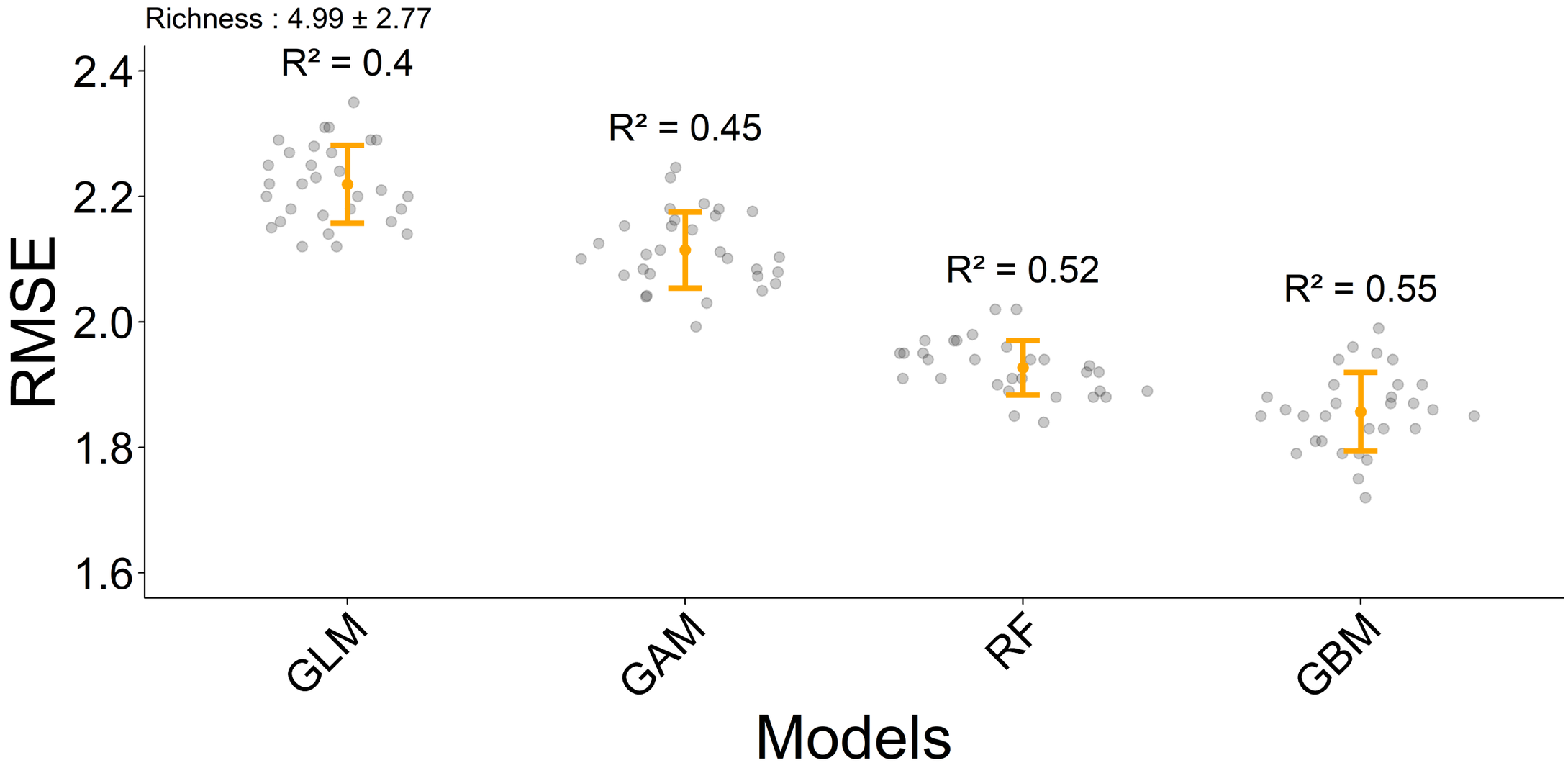
- AB_tot


Abundance : 233.79 ± 224.75 ind/m²
RMSE
50
45
40
35
30
25
R² = 0.31
R² = 0.33
R² = 0.36
R² = 0.37
GLM
GAM
RF
GBM
Models

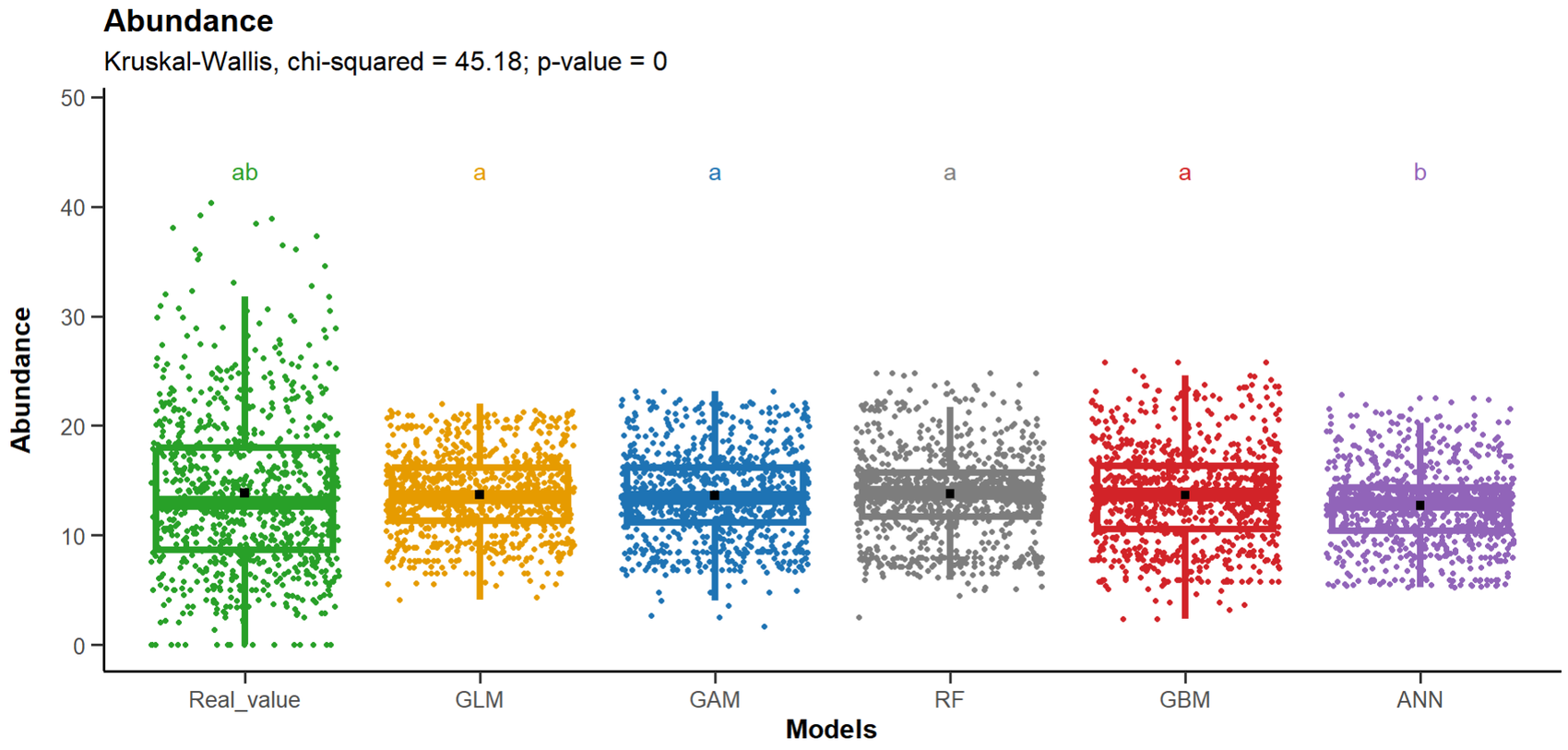
- BM_tot

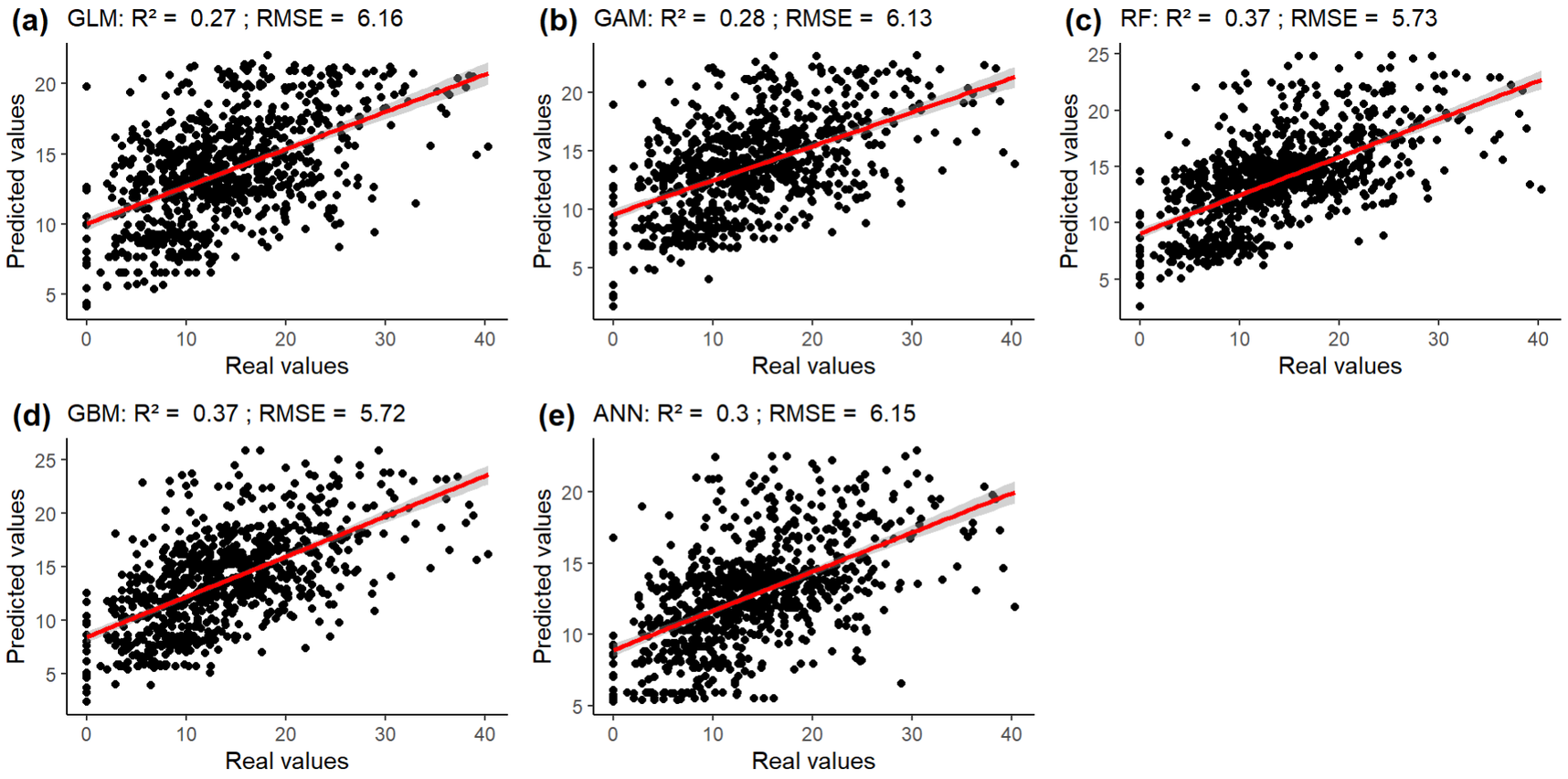

Biomass : 93.52 ± 77.77 g/m²
R² = 0.19
R² = 0.21
R² = 0.3
R² = 0.29
RMSE
18
14
10
6
GLM
GAM
RF
GBM
Models

- Richesse_tot

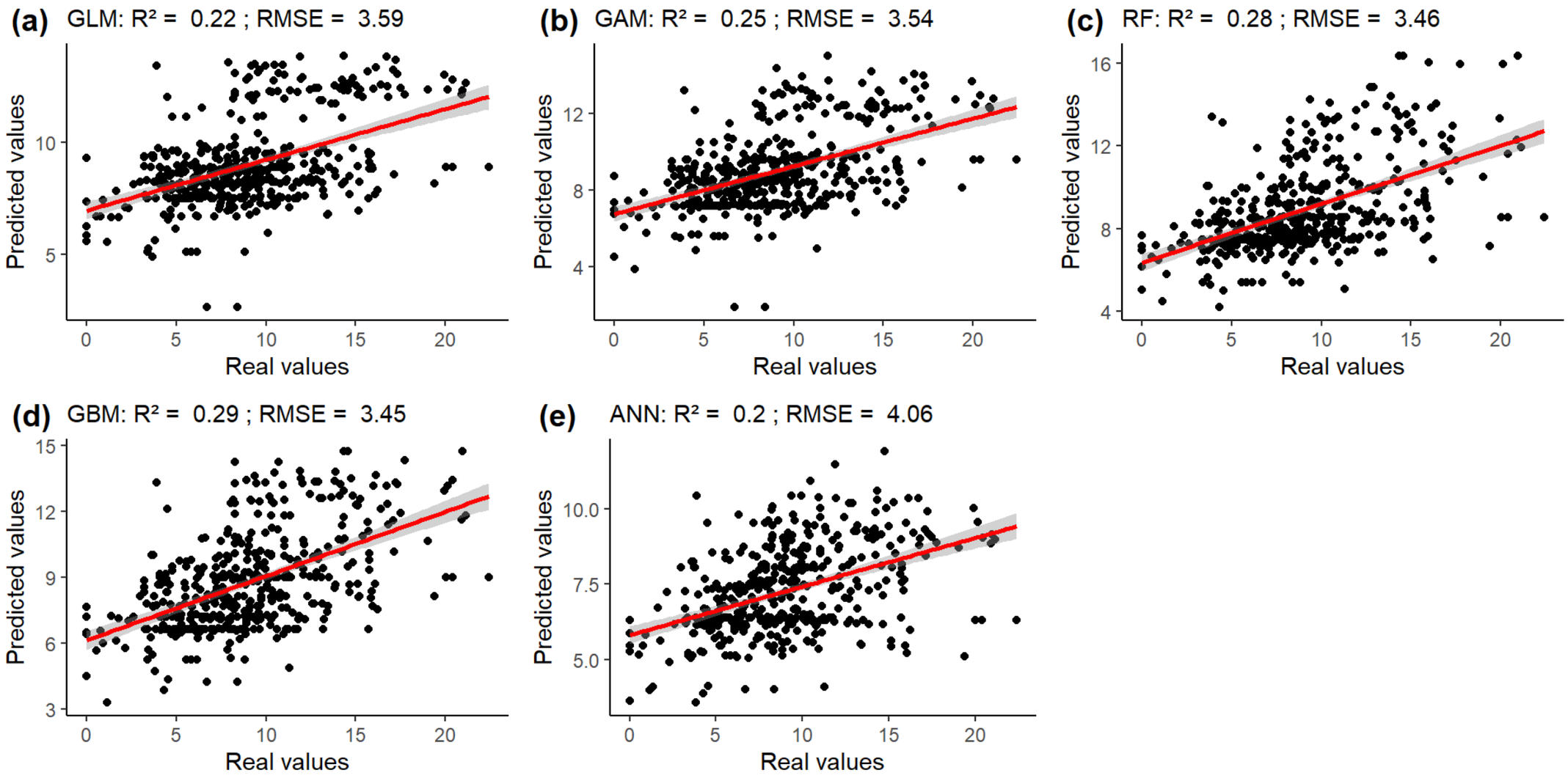


10.3 Prediction: AB_tot

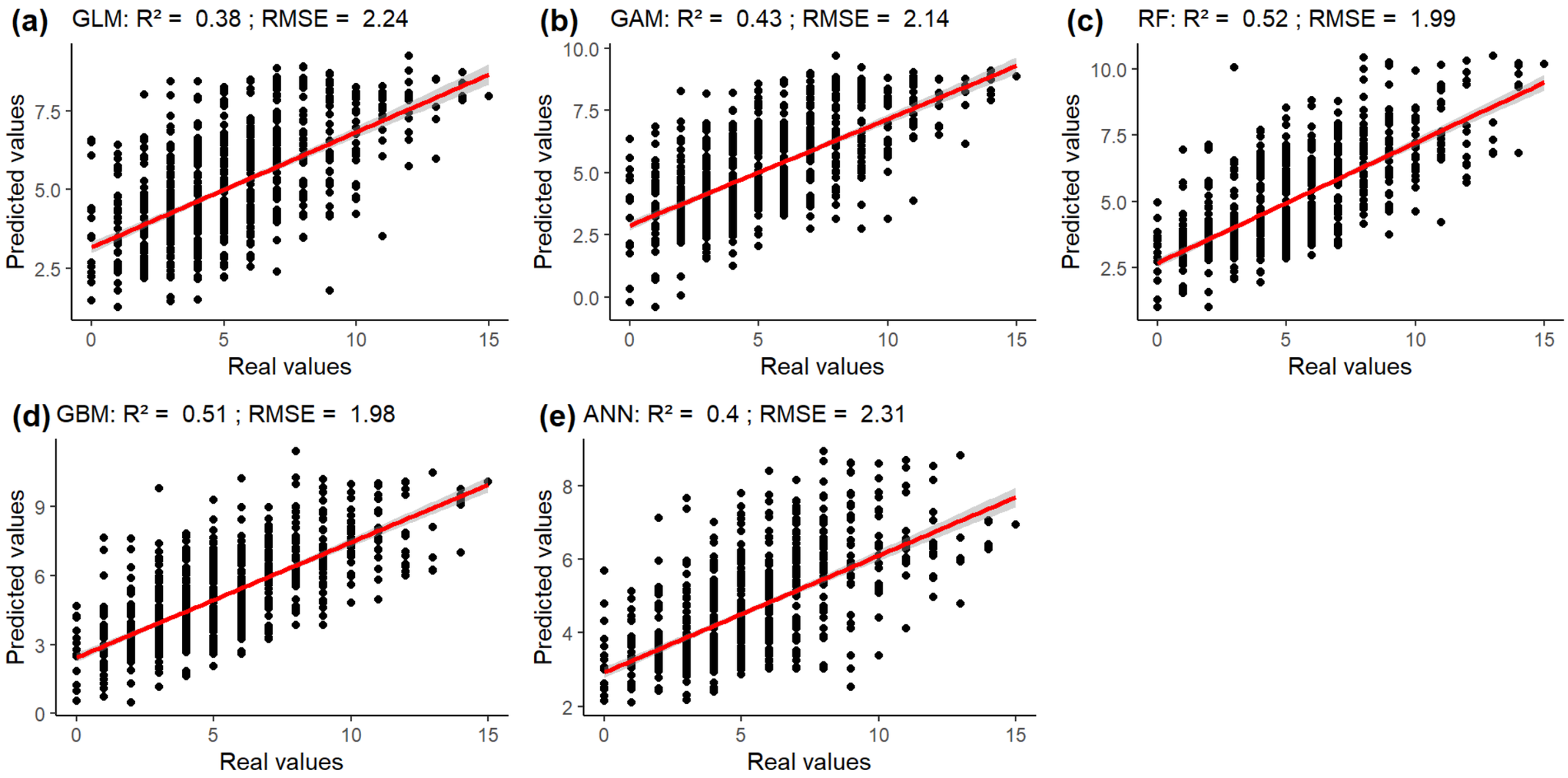



(a) GLM: R² = 0.27 ; RMSE = 6.16
(b) GAM: R² = 0.28 ; RMSE = 6.13
(c) RF: R² = 0.37 ; RMSE = 5.73
(d) GBM: R² = 0.37 ; RMSE = 5.72
(e) ANN: R² = 0.3 ; RMSE = 6.15
Predicted values
Real values

10.4 Prediction: BM_tot



10.5 Prediction: Richesse_tot



11 Questions

12 Idées améliorations models:

- Reconversture les données vdt pour diminuer la disperssion (/25 ?)
- Création des models par OS ou equilibre des levels des OS
- Cas 2: models sans repetition temporelle des données