

L'autécologie du bouleau pubescent (*Betula pubescens* Ehrh.)

Auteur : Musson, Martin

Promoteur(s) : Claessens, Hugues; Zeoli, Lorna

Faculté : Gembloux Agro-Bio Tech (GxABT)

Diplôme : Master en bioingénieur : gestion des forêts et des espaces naturels, à finalité spécialisée

Année académique : 2023-2024

URI/URL : <http://hdl.handle.net/2268.2/21256>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

10 Annexe

Annexe 1 : Fiche Terrain, basée sur les fiches Forêt.Nature (2021).

ID de la placette :/.....

Fiche terrain

1. Localisation

Latitude	Longitude	Altitude

2. Relevé dendrométrique

Test essence

☐

B. Pubescent

☐

B. Verruqueux

Htot			
C à 1m30			
H début houppier			
Densité du peuplement			
Age			



3. Influence humaine

Usage précédent du sol

Sylvicole - Agricole - Autre

Origine du peuplement

Plantation - Régénération naturelle

Itinéraire sylvicole

Taillis - Futaie régulière - Futaie irrégulière
Taillis sous futaie - Sylviculture AO - Aucune

ID de la placette :/.....

Compaction du sol

Oui - Non

Orniérage du sol

Oui - Non

Présence de drains

Présence - Absence

4. Topographie

Classe descriptive
de la pente :

Formes de relief	
Dépression géomorphologique	
Vallée	Plaine alluviale
	Terrasse alluviale homogène
	Bourrelet alluvial, berge, île
	Dépression marginale
Vallon	
Versant	A-pic, abrupt, falaise, ligne de crête
	Haut de pente, pente convexe
	Pleine pente
	Replat, terrasse, pente concave
	Bas de pente
Plateau	Plateau étendu, plane
	Dôme, sommet, rebord (légère pente)
	Cuvette, dépression du plateau

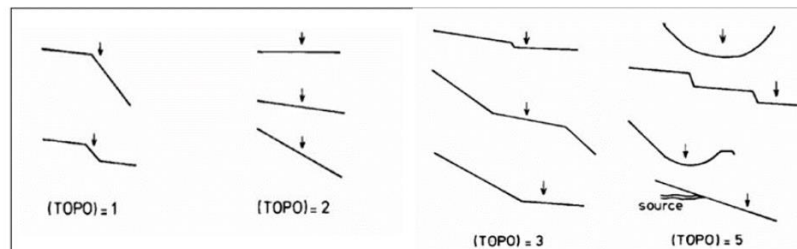
Pente moyenne :

Diagramme de pente moyenne : une ligne horizontale avec un point de départ, un point d'arrivée, et une flèche indiquant la direction de la pente.

Orientation de la pente :

Diagramme d'orientation de la pente : un rectangle avec un point central.

Interprétation en terme
de bilan hydrique :



5. Pédologie

Lithologie :

Type d'humus :

pH de l'horizon humifère :

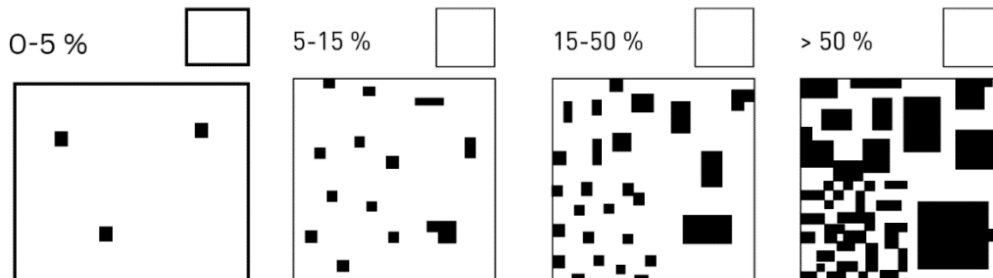
pH en profondeur :

Texture :

ID de la placette :/.....

Charge caillouteuse :

Proportion de la charge caillouteuse :



Nature de la charge caillouteuse :

Phase :

Profondeur :

Profondeur de sondage : cm

Cause de l'arrêt de sondage :

Profondeur utile du sol : cm

Horizon argileux ou compacté : Présence : ☐ Profondeur : cm

Profondeur d'apparition du pseudogley : cm

Profondeur d'apparition du gley : cm

Classe de drainage :

R.U. Totale : mm (voir Annexe 1)

Type de profil :

Affleurements rocheux : Présence : ☐ **Proportion :** %

Cours d'eau voisin : OUI - NON **Placette inondable :** OUI - NON

Indice trophique : **Niveau hydrique :**

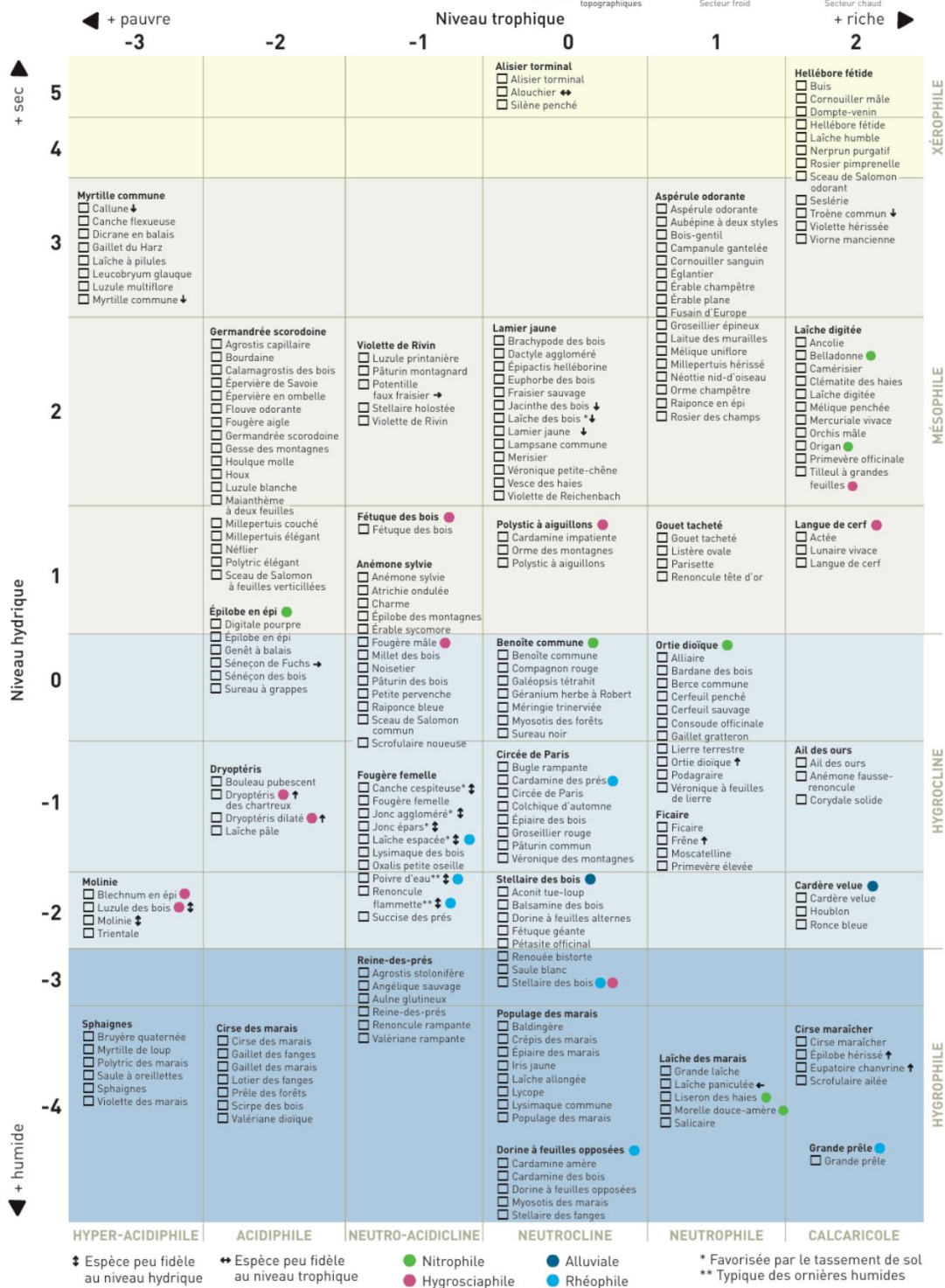
Profondeur (cm)	Pierrosité		
	5-15 %	15-50 %	> 50 %
> 125	0		
80-125	1		
40-79	2	2	3
20-39	4	4	5
< 20		6	6

ÉCOGRAMME DES ESPÈCES (NOMS FRANÇAIS)

ID de la placette :/.....

STATION : DATE :

- ☐ Présence discrète (< 5 %) ☐ Présence significative (5-25 %)
☒ Abondance (26-50 %) ☒ Dominance (> 50 %) ☒ Présence dans les ornières



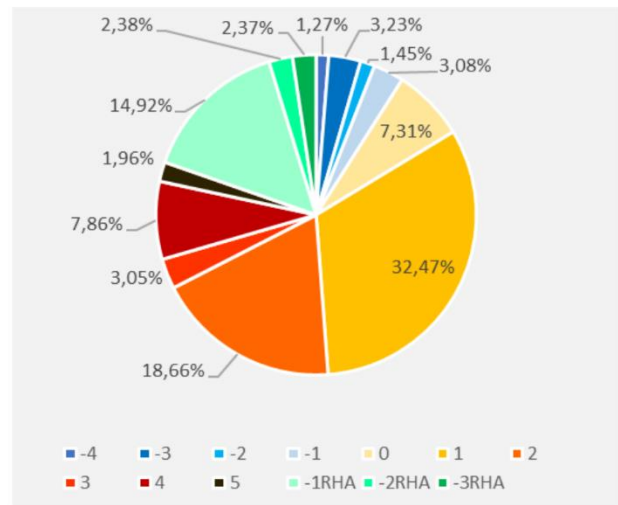
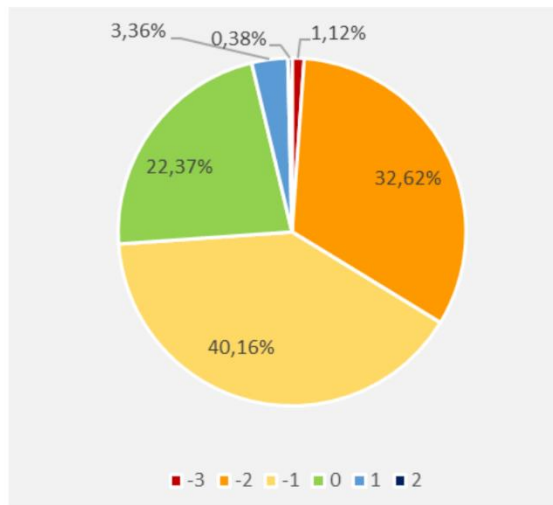
Annexe 2 : Les données de terrain, cartographiques et intégrées, accompagnées de leur code, utilisé lors de l'analyse.

Relevés de terrain		Relevés cartographiques	
Données dendrométriques			
Nom	Code	Nom	Code
• Hauteur totale (m) • Circonférence à 1m30 (cm) • Surface terrière(m²/ha) • Age (années)	• Htot • C_130 • g • age	• /	• /
Données pédologiques			
Nom	Code	Nom	Code
• Classe d’humus • Ph humifère et en profondeur • Texture du sol • Niveau hydrique pédologique • Profondeur utile et totale (cm) • Phase	• type_d_humus • Ph_hum et Ph_prof • texture • NH_pedo • prof_utile et profondeur _de_sondage • phase	• Texture (CNSW) • Contenance du sol en eau (mm) • Principaux types de sol • Niveau trophique cartographique • Niveau hydrique cartographique • Déficit hydrique estival • Déficit en oxygène	• texture_carte • SWC_carto • PTS_carto • NT_carte • NH_carte • Def_hyd_estiv • Def_o2_carte
Données topographiques			
Nom	Code	Nom	Code
• Classification descriptive de la pente • Interprétation en terme de bilan hydrique • Pente moyenne (°) • Orientation de la pente	• classe_descriptive_de_la_pente • Interpretation_bilan_Hydrique • Pente_moyenne • Orientation_Pente	• Pente (°) • TPI à 70 mètres (m) • TPI à 200 mètres (m) • Sous-secteurs radiatifs • Situation d’apport en eau hors précipitations	• Pente • TPI_70m • TPI_200m • ssr_carte • App_eau_Carte

Données phytosociologiques			
Nom	Code	Nom	Code
- Niveau hydrique phytosociologique - Niveau trophique phytosociologique	- NH_phyto - NT_phyto	/	/
Données climatiques			
Nom	Code	Nom	Code
/	/	- Température moyenne sur l'année (°C) - Température moyenne sur la période de végétation (°C) - Somme des degrés jours (°C) - Longueur de la période de végétation (jour) - Evapotranspiration moyenne (mm) - Bilan hydrique estival (mm) - Somme des précipitations de l'année (mm) - Somme des précipitations de la période de végétation (mm)	- tg_30 - tg_410_30 - gdd_30 - long_sais_vege_30 - etp_410_30 - bhe_410_30 - rf_30 - rf_410_30
Données altitudinales			
Nom	Code	Nom	Code
/	/	Modèle numérique de terrain	Altitude
Données d'influence humaine			
Nom	Code	Nom	Code
- Scénario sylvicole appliqué - Présence de drains	- It_sylv - drains	Influence passée : carte de Ferraris (1770-1778)	Us_prec

Données intégrées	
Nom	Code
- Niveau trophique intégratif - Niveau hydrique intégratif - Déficit hydrique estival intégratif - Déficit en oxygène intégratif	- NT_integratif - NH_integratif - Def_hyd_estiv - Def_o2

Annexe 3 : Proportion de surface du sol de la Wallonie selon les différentes classes de niveaux trophiques (à gauche) et hydrique (à droite).



LE BOULEAU PUBESCENT, UNE ESSENCE ENCORE MYSTÉRIEUSE

APPEL À VOTRE CONTRIBUTION POUR MIEUX LE CONNAÎTRE

par Martin Musson

Étudiant à Gembloux Agro-Bio Tech - Université de Liège

À l'heure où de nombreux changements imprévisibles sont à venir, il est opportun de se pencher sur une sylviculture s'ouvrant à une diversité pouvant offrir plus de résilience à la forêt. Cela passe certes par une variété d'âges et de structures au sein des peuplements mais cela dépend également de leur composition. En plus de l'épicéa, du hêtre et du chêne, de nombreuses essences se développent dans nos forêts wallonnes sans que de grandes attentions leur soient prêtées. Parmi ces essences se trouvent les bouleaux verruqueux (*Betula pendula*) et pubescent (*Betula pubescens*).

Héloïse Dubois, dans son guide de sylviculture, a montré l'intérêt d'intégrer ces espèces colonisatrices dans les peuplements pour leurs rôles écologiques, sylvicoles et leur production d'un bois dont la qualité est encore sous-estimée (Dubois et al., 2023). Ces avantages poussent d'autant plus à les considérer qu'à l'heure actuelle nos principales essences forestières sont affaiblies.

Malgré l'importance que les bouleaux vont prendre dans la sylviculture future, leur comportement dans le cadre des changements climatiques est encore à éclaircir. Au premier abord, tout porte à penser que, par leur grande variabilité génétique et leur capacité de dispersion, les bouleaux sont relativement souples face aux changements climatiques. Cependant, un doute persiste quant

à la capacité d'adaptation du bouleau pubescent, plus nordique que le verruqueux. Un travail de fin d'étude débute à Gembloux Agro Bio-Tech, sous la direction de Lorna Zeoli et Hugues Claessens, pour estimer dans quelle mesure le bouleau pubescent, moins connu que le verruqueux, présente aussi une opportunité pour la forêt wallonne.

Etant donné le peu d'études concernant le bouleau pubescent et la difficulté à distinguer les deux espèces, le recensement de cette essence n'a jamais été effectué précisément sur l'ensemble de la Wallonie. Ainsi, un appel à contribution est lancé à tous les propriétaires et gestionnaires forestiers qui possèdent dans leurs peuplements des groupes d'au moins une petite dizaine de bouleaux pubescents, et qui souhaitent

participer à cette recherche. Le cas échéant, nous viendrions y réaliser des mesures dendrométriques et écologiques. Éventuellement, nous aimerions choisir avec eux quelques arbres pour y prélever deux carottes afin de réaliser une étude dendroécologique (analyse de l'évolution de la largeur de cerne en relation avec les variations du climat).

DISTINGUER LES DEUX ESPÈCES

La distinction entre les deux espèces de bouleau n'est pas simple. Afin de différencier le pubescent du verruqueux, il faut prêter attention aux points suivants :

- les feuilles du bouleau pubescent sont plutôt en forme de losange,

L'HOMME ET LA FORÊT

simplement dentées et plus petites que celles du bouleau verruqueux, qui sont plutôt triangulaires, doublement dentées et plus pointues;

- ses jeunes rameaux sont couverts de poils et non verruqueux (attention, ne pas se fier aux gourmands);

- peu de motifs verticaux sont présents sur son écorce; celle-ci ne se craquelle pas fortement à la base en vieillissant;
- les fines branches de son houppier sont peu pendantes.

Tableau comparatif des caractères physiques des bouleaux pubescent et verruqueux

Bouleau pubescent		Bouleau verruqueux
	Allure de l'arbre	
	Base du tronc	
	Découpe de la feuille	

Si vous souhaitez participer à cette recherche, n'hésitez pas à prendre contact avec Martin Musson à l'adresse ci-dessous, par mail, téléphone ou courrier. Les localisations des groupes d'arbres peuvent être transmises selon votre préférence : whatsapp, mail, lettre, etc.

martin.musson@student.uliege.be | 0471 62 60 87 | Gestion des Ressources Forestières
Gembloux Agro-Bio Tech (ULiège) | 2, Passage des Déportés | 5030 Gembloux

Annexe 5 : Jeu de données utilisé pour l’analyse

ID	Altitude	R	SI	SI classe	age	Us. prec	It. sylv	drains	zho	classe descriptive de la pente	pente_moyenne	interpretation_bilan_hydrrique	TRI_70m	TRI_200m
BER 01	564,4	24	11,23	tolerance	83	foret	Futaie reg	non	Haute Ardennes	plane	0	2	-3,31	-10,12
BER 02	561,4	27	12,95	tolerance	71	ouvert	Futaie reg	non	Haute Ardennes	plane	0	2	-0,96	-8,16
BRB 03	595,4	21	3,06	tolerance elargie	118	foret	Futaie reg	non	Haute Ardennes	plane	2,5	2	0,08	-4,01
BUL 01	659,2	21	10,12	tolerance elargie	69	foret	Futaie irr	non	Haute Ardennes	pente	5	1	0,31	-4,25
CHAM 01	338,1	16	9,66	tolerance elargie	73	champs	Aucune	non	Haute Ardennes	plane alluviale	0	2	-0,16	-2,25
CHAM 02	338,2	24	13,8	tolerance	53	champs	Aucune	non	Haute Ardennes	plane alluviale	0	2	-0,84	-10,05
CHIN 01	325,8	22	18,43	optimum	23	foret	AO	oui	Haute Ardennes	plane	0	2	0,39	0,24
CHIN 02	325,4	26	17,93	optimum	34	foret	AO	oui	Haute Ardennes	plane	0	2	0,59	-4,12
DOR 01	232,1	22	19,83	optimum	47	champs	Futaie reg	non	Condroz-Sambre et Meuse	plane	0	2	-0,31	8,36
EHE 01	360,5	18	14,3	tolerance	31	foret	Futaie reg	non	Basse et moyenne Ardennes	bourrelet alluvial	6,2	2	-13,65	-121,80
EHE 02	354,7	17	10,56	tolerance elargie	49	foret	Aucune	non	Basse et moyenne Ardennes	terrasse alluviale	0	5	-21,27	-142,43
EHE 03	354,5	16	11,04	tolerance	55	foret	Aucune	non	Basse et moyenne Ardennes	bourrelet alluvial	0	5	-26,67	-145,40
EHE 04	409,4	19	12,09	tolerance	72	foret	Futaie irr	non	Ardennes centro-orientale	rebord	4,1	2	-0,94	6,64
EHE 05	429,3	23	16,28	optimum	25	foret	Futaie reg	oui	Ardennes centro-orientale	plane	3,3	2	-0,69	-3,78
FRA 01	195	19	14,45	tolerance	42	foret	Futaie irr	non	Fagne-Famenne-Calestienne	bourrelet alluvial	3,1	2	-4,59	-10,20
FRA 02	195,5	21	15,2	optimum	42	foret	Futaie irr	oui	Fagne-Famenne-Calestienne	plane	0	2	-0,80	-9,56
GES 01	195,4	28	15,22	optimum	51	foret	Futaie reg	non	Condroz-Sambre et Meuse	plane	4,7	2	0,06	6,88
HAUL 01	418,2	28	9,46	tolerance elargie	79	foret	Futaie reg	non	Ardennes centro-orientale	plane alluviale	0	2	0,49	-3,84
HAUL 02	421,5	27	7,47	tolerance elargie	89	foret	Futaie reg	non	Ardennes centro-orientale	plane alluviale	0	2	0,75	-1,98
HAUL 03	422,4	24	7,62	tolerance elargie	96	foret	Futaie reg	non	Ardennes centro-orientale	plane alluviale	0	2	-3,20	-8,33
HAUL 04	423,4	25	15,03	optimum	28	foret	Futaie reg	oui	Ardennes centro-orientale	plane	0	2	0,27	-6,64
HAUT 01	334,6	24	10,66	tolerance elargie	77	foret	Futaie reg	non	Basse et moyenne Ardennes	rebord	0	3	-2,59	-33,59
HAUT 02	333,9	17	8,86	tolerance elargie	82	foret	Aucune	non	Basse et moyenne Ardennes	plane alluviale	0	5	-7,31	-45,69
HAV 01	257,6	20	15,59	optimum	56	champs	Futaie reg	non	Condroz-Sambre et Meuse	pente	9,6	1	-0,02	-66,21
HAV 02	262,3	20	16,85	optimum	55	champs	tail s. fut	non	Fagne-Famenne-Calestienne	plane alluviale	0	5	6,29	-6,62
HAV 03	293,9	22	18,98	optimum	45	foret	Futaie reg	non	Condroz-Sambre et Meuse	pente	0	3	8,08	-1,16
MARE 01	249	16	18,24	optimum	18	foret	tailles	oui	Basse et moyenne Ardennes	terrasse alluviale	0	5	-4,61	-23,28
MARE 02	249,2	18	21,1	optimum	18	foret	tailles	non	Basse et moyenne Ardennes	plane alluviale	0	5	-4,94	-21,42
MARE 03	250,1	16	20,62	optimum	18	foret	tailles	oui	Basse et moyenne Ardennes	terrasse alluviale	3,6	3	-2,49	-19,80
NEUF 01	347,7	22	14,84	tolerance	56	foret	Futaie reg	non	Basse et moyenne Ardennes	plane	0	2	0,08	1,04
NEUF 02	347,7	22	14,84	tolerance	56	foret	Futaie reg	non	Basse et moyenne Ardennes	plane	0	2	-0,92	-0,61
NEUF 03	347,7	22	14,84	tolerance	56	foret	Futaie reg	non	Basse et moyenne Ardennes	plane	0	2	-0,92	-0,61
ORN 01	170,4	19	17,32	optimum	60	foret	tail s. fut	non	Heshino-Brabant	plane	0	2	-0,35	-1,78
OUF 01	301,6	18	20,21	optimum	35	champs	Futaie reg	non	Condroz-Sambre et Meuse	rebord	2,9	2	2,84	13,82
REI 01	532	28	9,23	tolerance elargie	91	foret	AO	non	Haute Ardennes	pente	5,3	1	1,00	1,57
REI 02	539	20	8,61	tolerance elargie	73	foret	Futaie reg	oui	Haute Ardennes	dom	3,6	2	-0,31	3,50
REI 03	531,5	26	9,17	tolerance elargie	65	foret	Aucune	non	Haute Ardennes	plane	1,4	2	-0,20	-2,94
RIDA 01	381,1	20	17,98	optimum	33	foret	AO	non	Basse et moyenne Ardennes	rebord	5,5	2	-0,53	3,03
SAMB 01	371,6	19	18,11	optimum	32	foret	AO	non	Basse et moyenne Ardennes	plane	4,8	2	-1,20	-8,58
SAMB 02	230,1	24	21,83	optimum	49	foret	Futaie reg	non	Condroz-Sambre et Meuse	plane	0	2	2,02	11,40
SAMB 03	228,9	26	16,56	optimum	40	foret	Futaie reg	non	Condroz-Sambre et Meuse	plane	0	2	0,63	1,65
SAMB 04	229,9	25	18,72	optimum	49	foret	Futaie irr	non	Condroz-Sambre et Meuse	plane	0	2	1,27	6,53
SAMB 05	275,4	19	15,46	optimum	67	foret	Futaie irr	non	Condroz-Sambre et Meuse	plane	0	2	2,63	8,85
SIMM 01	89,4	25	17,7	optimum	52	ouvert	Aucune	non	Plaines et Vallées Scaldisiennes	plane	0	2	-1,20	-0,43
SIMM 02	92,9	25	20,46	optimum	36	ouvert	Futaie reg	non	Plaines et Vallées Scaldisiennes	plane	0	2	6,88	15,83
SIMM 03	93	24	22,37	optimum	37	ouvert	Futaie reg	non	Plaines et Vallées Scaldisiennes	plane	4	2	1,16	3,69
SIMM 04	456,1	27	13,8	tolerance elargie	30	foret	Aucune	oui	Ardennes centro-orientale	bourrelet alluvial	0	5	-3,78	-33,39

ID	Type d'humus	texture	phase	profondeur de sondage	prof. utile	SMC carto	Def. o2	Def. hyd. estiv	NT. Intégratif	ble. 410_30	etp. 410_30	gdd_30	long. sais. wege_30	r _f _30	r _f _410_30	tg_30	tg_410_30
BER_01	paratourbeux	Argilea	2	95	45	89	0	0	-2	114,38	503,62	677,98	176,03	1107,59	618,00	7,94	12,13
BER_02	paratourbeux	Argilea	2	85	45	85,2	-3	3	-2	114,38	503,62	677,98	176,03	1107,59	618,00	7,94	12,13
BER_03	tourbeux	MO	0	150	150	85,2	4	3	-3	114,38	503,62	677,98	176,03	1107,59	618,00	7,94	12,13
BUJ_01	moder	Argilea	2	75	65	85,2	-3	3	-3	181,52	485,51	583,70	160,83	1248,43	667,03	7,23	11,45
CHAAN_01	tourbeux	MO	0	150	150	85,2	4	0	0	18,60	551,61	864,36	186,60	1105,86	570,21	9,13	13,37
CHAAN_02	tourbeux	MO	0	150	150	85,2	4	0	0	6,99	553,57	852,76	187,37	1093,32	560,55	9,05	13,31
CHIN_01	moder	Argileu	4	40	40	100	3	0	-2	107,38	531,98	766,08	180,93	1279,41	639,36	8,56	12,77
CHIN_02	moder	Argileu	4	35	35	100	3	3	-2	107,38	531,98	766,08	180,93	1279,41	639,36	8,56	12,77
DOR_01	moder	Argilea	2	65	25	90	-3	3	1	6,61	530,34	838,69	193,67	941,06	536,95	9,22	13,22
DOR_02	moder	Argilea	4	30	30	99	2	0	0	69,89	516,17	789,45	192,57	1037,25	586,07	8,78	12,89
ERE_01	moder	Argilea	2	45	45	99	1	0	0	69,89	516,17	789,45	192,57	1037,25	586,07	8,78	12,89
ERE_02	moder	Argilea	2	50	50	99	2	0	0	69,89	516,17	789,45	192,57	1037,25	586,07	8,78	12,89
ERE_03	moder	Argilea	4	50	30	89	4	0	-2	69,89	516,17	789,45	192,57	1037,25	586,07	8,78	12,89
ERE_04	moder	Argilea	4	85	40	97	4	0	-2	69,89	516,17	789,45	192,57	1037,25	586,07	8,78	12,89
ERE_05	moder	Argilea	2	70	40	168	2	0	-1	-6,85	527,51	886,26	197,00	943,57	520,66	9,49	13,50
FRA_01	moder	Argilea	2	65	35	168	-2	2	0	-6,85	527,51	886,26	197,00	943,57	520,66	9,49	13,50
FRA_02	moder	Argilea	2	50	50	89	-2	2	-1	49,65	540,06	956,63	201,63	853,22	490,41	9,88	13,91
GES_01	moder	Argilea	3	80	55	817	4	0	-3	100,19	510,30	775,68	182,13	1128,40	610,49	8,39	12,49
HAUL_01	paratourbeux	Argileu	1	150	100	817	4	0	-3	100,19	510,30	775,68	182,13	1128,40	610,49	8,39	12,49
HAUL_02	paratourbeux	Argileu	2	80	50	89	4	0	-2	100,19	510,30	775,68	182,13	1128,40	610,49	8,39	12,49
HAUL_03	paratourbeux	Argilea	2	80	80	89	-3	3	-2	100,19	510,30	775,68	182,13	1128,40	610,49	8,39	12,49
HAUL_04	moder	Argilea	2	90	30	83	3	0	-1	103,35	516,87	742,52	183,97	1186,38	620,23	8,52	12,62
HAUTF_01	moder	Argilea	2	80	55	242	0	1	-1	103,35	516,87	742,52	183,97	1186,38	620,23	8,52	12,62
HAUTF_02	moder	Argilea	0	150	150	250	0	1	-1	29,18	528,99	864,22	196,87	1004,88	558,17	9,24	13,34
HAV_01	moder	Argilea	0	150	65	176	-1	1	-1	29,18	528,99	864,22	196,87	1004,88	558,17	9,24	13,34
HAV_02	moder	Argilea	2	70	70	85	0	0	-2	29,18	528,99	864,22	196,87	1004,88	558,17	9,24	13,34
HAV_03	moder	Argilea	2	50	50	242	3	0	-2	64,52	514,89	801,85	188,60	1098,15	579,41	8,91	12,99
MAIE_01	moder	Argileu	4	38	38	242	2	0	0	64,52	514,89	801,85	188,60	1098,15	579,41	8,91	12,99
MAIE_02	moder	Argilea	4	30	30	242	1	0	0	64,52	514,89	801,85	188,60	1098,15	579,41	8,91	12,99
MAIE_03	moder	Argileu	2	90	60	97	-1	1	-3	139,55	508,00	811,61	190,53	1167,37	667,54	8,91	13,01
NEUF0_01	moder	Argilea	2	65	65	219	-1	1	-3	235,45	490,40	637,97	201,57	1270,42	702,76	7,91	12,10
ORH_01	moder	Argilea	0	150	80	261	-2	2	0	54,474	964,93	201,57	201,57	856,56	491,34	9,92	13,96
OUT_01	moder	Argilea	0	150	150	85	0	0	0	31,35	529,24	844,30	195,50	1006,28	560,59	9,12	13,23
REI_01	moder	Argilea	4	35	35	83	0	1	-2	208,38	494,37	676,69	178,53	1270,42	702,76	7,91	12,10
REI_02	moder	Argilea	4	30	30	83	-1	1	-3	208,38	494,37	676,69	178,53	1270,42	702,76	7,91	12,10
REI_03	moder	Argilea	1	90	90	89	-1	1	-3	208,38	494,37	676,69	178,53	1270,42	702,76	7,91	12,10
RIDA_01	moder	Argilea	2	60	55	204	0	1	-1	54,33	517,85	828,50	191,17	1071,88	572,18	9,09	13,16
RIDA_02	moder	Argilea	2	60	50	204	-2	2	-2	-29,89	534,74	908,89	196,67	900,49	504,85	9,58	13,64
RIDA_03	moder	Argilea	0	150	150	259	0	1	-1	-29,89	534,74	908,89	196,67	900,49	504,85	9,58	13,64
SAMB_01	moder	Argilea	2	85	55	259	-2	2	0	-29,89	534,74	908,89	196,67	900,49	504,85	9,58	13,64
SAMB_02	moder	Argilea	2	85	55	259	-2	2	0	-29,89	534,74	908,89	196,67	900,49	504,85	9,58	13,64
SAMB_03	moder	Argilea	2	85	55	259	-2	2	0	-29,89	534,74	908,89	196,67	900,49	504,85	9,58	13,64
STAM_01	moder	Argilea	0	150	150	139	0	0	-3	61,80	542,20	1025,50	205,43	855,22	480,40	10,35	14,32
STAM_02	moder	Argilea	0	150	150	139	0	0	-3	61,80	542,20	1025,50	205,43	855,22	480,40	10,35	14,32
STAM_03	moder	Argilea	0	150	150	139	0	0	-3	61,80	542,20	1025,50	205,43	855,22	480,40	10,35	14,32
STAM_04	moder	Argilea	0	150	150	139	0	0	-3	61,80	542,20	1025,50	205,43	855,22	480,40	10,35	14,32
WIT_01	moder	Argileu	4	35	35	242	3	0	-1	97,69	522,36	689,39	172,30	1225,42	620,06	8,00	12,24

ID	Def_o2_carte	Def_hyd_estiv_carte	NT_carte	s_s_1_carte	App_eau_carte	penie	texture_carte
BER_01	-3	3	-2	sous-secteur_neutre	sans_apport_eau	27,27	G
BER_02	4	0	-3	sous-secteur_neutre	Apport permanent	21,97	V
BER_03	4	0	-3	sous-secteur_neutre	Apport permanent	11,65	V
BUL_01	-3	3	-2	sous-secteur_neutre	sans_apport_eau	42,10	G
CHAN_01	4	0	0	sous-secteur_neutre	Apport permanent	2,86	V
CHAN_02	4	0	0	sous-secteur_neutre	Apport permanent	2,26	V
CHIN_01	2	0	-1	sous-secteur_neutre	Apport permanent	9,04	A-G
CHIN_02	2	0	-1	sous-secteur_neutre	Apport permanent	5,05	A-G
DOR_01	0	2	1	sous-secteur_neutre	sans_apport_eau	6,76	G
ERE_01	2	0	0	sous-secteur_froid	Apport permanent	28,62	G
ERE_02	2	0	0	sous-secteur_froid	Apport permanent	32,56	G
ERE_03	2	0	0	sous-secteur_froid	Apport permanent	32,51	G
ERE_04	-1	1	-2	sous-secteur_neutre	sans_apport_eau	31,79	G
ERE_05	-2	2	-1	sous-secteur_neutre	sans_apport_eau	42,03	G
FRA_01	1	0	0	sous-secteur_froid	Apport variable	21,27	E
FRA_02	1	0	0	sous-secteur_froid	Apport variable	23,50	E
GES_01	-1	1	-2	sous-secteur_neutre	sans_apport_eau	29,92	G
HAUL_01	4	0	-3	sous-secteur_neutre	Apport permanent	18,94	V
HAUL_02	4	0	-3	sous-secteur_neutre	Apport permanent	19,49	V
HAUL_03	-3	3	-2	sous-secteur_neutre	sans_apport_eau	18,45	G
HAUL_04	-3	3	-2	sous-secteur_neutre	sans_apport_eau	16,97	G
HAUT_01	0	0	-2	sous-secteur_neutre	Apport permanent	39,04	G
HAUT_02	3	0	0	sous-secteur_froid	Apport permanent	22,59	G
HAV_02	3	0	0	sous-secteur_froid	Apport permanent	15,47	A-G
HAV_01	0	1	-1	sous-secteur_neutre	sans_apport_eau	55,75	G
HAV_03	0	2	-1	sous-secteur_neutre	sans_apport_eau	48,87	G
MAAE_01	2	0	0	sous-secteur_froid	Apport permanent	9,93	G
MAAE_02	2	0	0	sous-secteur_froid	Apport permanent	19,40	G
MAAE_03	2	0	0	sous-secteur_froid	Apport permanent	23,01	G
NEUF_01	-3	3	-2	sous-secteur_neutre	sans_apport_eau	18,78	G
NEUH_01	4	0	-3	sous-secteur_neutre	Apport permanent	29,33	G
ORN_01	0	0	-2	sous-secteur_neutre	sans_apport_eau	14,10	A
OUF_01	-1	1	0	sous-secteur_neutre	sans_apport_eau	14,72	G
RE_01	0	2	-2	sous-secteur_neutre	sans_apport_eau	44,21	G
RE_02	0	2	-2	sous-secteur_neutre	sans_apport_eau	35,14	G
RE_03	-3	3	-2	sous-secteur_neutre	sans_apport_eau	29,50	G
RIDA_01	0	0	-1	sous-secteur_neutre	sans_apport_eau	28,26	G
RIDA_02	-1	1	-1	sous-secteur_neutre	sans_apport_eau	26,06	G
SAMB_01	-2	2	-1	sous-secteur_neutre	sans_apport_eau	13,80	A
SAMB_02	-3	3	-1	sous-secteur_neutre	sans_apport_eau	16,05	A
SAMB_03	-2	2	-1	sous-secteur_neutre	sans_apport_eau	17,18	A
SAUB_01	-2	2	-1	sous-secteur_neutre	sans_apport_eau	6,46	A
SIAM_02	0	2	-3	sous-secteur_neutre	sans_apport_eau	6,38	Z
SIAM_03	0	2	-3	sous-secteur_neutre	sans_apport_eau	33,95	Z
SIAM_04	0	1	-2	sous-secteur_neutre	sans_apport_eau	18,37	Z
WIT_01	2	0	0	sous-secteur_froid	Apport permanent	30,99	(G)

Annexe 6 : Code de l'analyse (RStudio)

```
#=====
===

#           Analyse des donnees

#=====
===

#Auteur : Martin Musson

#Date : Juin 2024

#Cadre : TFE sur l'Autecologie du Bouleau pubescent


# 1 Ouverture des librairies, session et dossiers

#=====

rm(list=ls())

{ # ouverture des librairies

  library(FactoMineR)

  library(factoextra)

  library(missMDA)

  library(randomForest)

  library(descr)

  library(caret)

  library(e1071)

  library(DALEX)

  library(shapper)

  library(iBreakDown)

  library(ggplot2)

  library(pander)

  library(rstatix)

  library(car)

  library(ggrepel)

  library(corrplot)

  library(RColorBrewer)

  library(randomForestExplainer)
```

```
library(boot)
library(fastshap)
library(dplyr)
library(tidyr)
library(viridis)
library(patchwork)
library(BlandAltmanLeh)
```

```
#Ouverture de la session
```

```
setwd("chemin d'accès")
```

```
#Ouverture du dossier data
```

```
data <- read.table(file = "bande de donnees .csv", header= T, sep = ";", dec=",")
```

```
rownames(data) <- as.character(unlist(data[,1]))
```

```
rownames(data)
```

```
data <- data[,-1]
```

```
#Choix de l'expression de la variable
```

```
data$Alpha <- Alpha- as.numeric(data$Alpha)
```

```
data$Beta <- as.factor(data$Beta)
```

```
}
```

```
# Utilisation de subset pour retirer les lignes avec type_d_humus égal à "tourbeux" ou "paratourbeux"
```

```
#data <- subset(data, !(type_d_humus %in% c("tourbeux", "paratourbeux")))
```

```
# description du jeu de données (moyenne, mediane, minimum, maximum, écart type)
```

```
descr(data)
```

```
sd(data$g)
```

```
# 2 Presentation de l'échantillon
```

```
#=====
```

```
# description du jeu de données (moyenne, mediane, minimum, maximum, écart type)
```

```

descr(data)

sd(data$g)

hist(data$Alpha, breaks=10, col="lightgreen" , xlab="Alpha") # Observation de la distribution d'une
variable

# Distribution de la variable explicative. Recherche d'une distribution de forme normale.

summary(descr(data$SI))

hist(data$SI, breaks=10, col="lightgreen" , xlab="Age")

descr(data$SI)

hist(sqrt(data$SI), breaks=10, col="lightgreen" , xlab="Age") #application d'une racine

hist(log(data$SI), breaks=10, col="lightgreen" , xlab="Age") #application d'un logarithme

# Finalement, les transformations n'apportent pas plus de normalite. Il n'y a donc pas de nécessité à
redéfinir la variable à expliquer.

# Relations univariées entre les variables explicatives ou entre la variable à expliquer et une variable
explicative

# Regression univariée entre une variable numerique Alpha et l'indice de productivite (SI)

reg_1 <- lm(SI~Alpha, data=data)

pander(summary(reg_1))

# Analyse des résidus de la regression univariée

par(mfrow = c(1, 1))

plot(reg_1)

residuals <- residuals(reg_1)

# Test de normalité de Shapiro-Wilk

shapiro.test(residuals)

#Visualisation de la regression univariée

ggplot(data, aes(x = Alpha, y = SI)) +

  geom_point() + # Ajouter les points du scatter plot

  geom_smooth(method = "lm", se = FALSE, color = "orange") + # Ajouter la ligne de régression

  labs(x = "Alpha (.) " ,

    y = "SI (m)") +

  theme_minimal() +

  theme(

    axis.line = element_line(size = 1, colour = "black"),

```



```

axis.text = element_text(size = 12, colour = "black"),
axis.title = element_text(size = 14, face = "bold"),
axis.ticks = element_line(size = 1, colour = "black")
)

# Comparaison des moyennes de SI (ou d'une variable explicative numerique) en fonction d'une
variable factorielle (Beta)

# Visualisation de la relation par boites a moustaches.

boxplot(SI ~ Beta, data = data,
        main = "Boîtes à Moustaches du SI par Beta",
        xlab = "Beta",
        ylab = "SI",
        col = c("lightblue", "lightgreen", "lightcoral", "lightyellow"))

# Précision de la relation par anova

# Condition de normalité par le test de Shapiro
by(data$SI, data$Us_prec, shapiro.test)

# Condition de l'homogeneite des variances par le test de Levene
pander(leveneTest(SI~Us_prec, data=data))

#L'Anova
anova1 <- aov(SI~Us_prec, data=data)
summary(anova1)


# 3.      Pré-traitements du jeu de donnees pour l'analyse
#=====

# 3.1.      Nombre de lignes avec des NA
#-----

nrow(data)

nrow(data)-nrow(na.omit(data)) #Soustraction du nb de lignes total par le nombre de lignes ayant
des NA's

```

```

# 3.2.      Selection des variables conservees
#-----

# Deux modeles sont effectues sur des donnees de terrain tandis que deux autres sont construit sur
des donnees uniquement cartographiques.

# Le choix des variables entre les deux premiers ou les deux derniers modeles se fait ici.

# Preparation du jeu de donnees pour la suite.

data_famd <- data[,-c(2,5,6,7,8,9,10,12,17,32,33,34,35,36,37,38)]


# 3.3.      La FADM
#-----

# Objectif : recherche ou de correlations ou de valeurs aberrantes

# Qu est-ce qu une AFDM : une derivee de PCA permettant de realiser une analyse en composante
principale sur des variables et qualitatives et quantitatives.

# 3.3.1      La FADM : Calcul et valeurs propres

# Calcul de la FAMD

res.famd <- FAMD(data_famd[, -3], ncp = 6, sup.var = data_famd$SI, ind.sup = NULL, graph = TRUE)

# Nombre de variables actives

num_active_vars <- nrow(res.famd$var$coord)

print(num_active_vars)

# Valeurs propres

eig.val <- get_eigenvalue(res.famd)

head(eig.val)

# Visualisation des valeurs propres (scree plot)

fviz_screplot(res.famd, addlabels = TRUE, ylim = c(0, 50)) # 2 Dimensions (composantes
principales) sont conservees

# 3.3.2      L AFDM : Graphique des variables

# Toutes les variables

var <- get_famd_var(res.famd)

var

# Coordonnees des variables

head(var$coord)

```

```

# Cos2 des variables

head(var$cos2)

var$coord

# Contributions aux dimensions

head(var$contrib)

var$contrib

# Visualisation de la FAMD

fviz_famd_var(res.famd, axes = c(1,2), choice = "quanti.var", col.var.sup = "lightgrey", repel = TRUE)

my_colors <- c( "purple", "darkgreen", "brown", "brown", "brown", "blue", "blue",
"blue", "blue", "purple", "purple", "purple",
"purple", "purple", "purple", "purple", "purple", "blue", "blue", "blue", "blue") # ajustez les couleurs
selon vos besoins

fviz_famd_var(res.famd, col.var = my_colors, axes = c(1,2), col.var.sup = "lightgrey", repel = TRUE)

# 3.3.3 La AFDM : Graphique des individus

# Extraire les coordonnées des individus

ind_coords <- as.data.frame(res.famd$ind$coord)

# Extraire la qualité de représentation (cos2) des individus

ind_cos2 <- as.data.frame(res.famd$ind$cos2)

# Ajouter les noms des individus

ind_coords$names <- rownames(data_famd)

# Representation des individus selon leur classe de productivite
{
  palette_colors <- c( "#006600", "#99CC66", "#FF9900")

  # Créer un graphique avec ggplot2 en colorant selon SI_classe

  ggplot(ind_coords, aes(x = Dim.1, y = Dim.2, color = SI_classe, fill = SI_classe)) +

    geom_point(size = 2) +

    geom_text_repel(aes(label = names), size = 4, max.overlaps = 10) +

    stat_ellipse(aes(group = SI_classe), type = "norm", level = 0.95, alpha = 0.2, geom = "polygon") + #
Ajouter des ellipses pour chaque catégorie

    scale_color_manual(values = palette_colors) + # Ajuster les couleurs des points selon les niveaux
de SI_classe

    scale_fill_manual(values = palette_colors) + # Ajuster les couleurs des ellipses selon les niveaux de
SI_classe
}

```

```

labs(x = "Dim 1",
     y = "Dim 2",
     color = "SI_classe",
     fill = "SI_classe") +
theme_minimal() +
theme(legend.position = "bottom")
}

# 3.3.4 La AFDM : Regroupement (clustering)

# Creation de trois groupes sur base de l'affinite de chaque variables avec les composantes de la
FADM

hc <- HCPC(res.famd,nb.clust=3)

plot(hc, choice="bar")

# verifier sur quelles dimensions la projection des clusters est la plus pertinente.

hc$desc.axes # La 1 et la 2.

# Clusering sur base de l'AFDM :

# Étape 1 : Effectuer HCPC avec 3 clusters

hc <- HCPC(res.famd, nb.clust = 3, graph = TRUE)

# Étape 2 : Extraire les données du clustering

cluster_data <- hc$data.clust

# Étape 3 : Ajouter les clusters aux données FAMD

famd_coords <- data.frame(res.famd$ind$coord) # res.famd$ind$coord doit etre un data frame
famd_coords$Cluster <- cluster_data$clust # Ajoutez la colonne Cluster

# Étape 4 : Diviser les données par cluster

cluster_1 <- famd_coords[famd_coords$Cluster == 1, ]
cluster_2 <- famd_coords[famd_coords$Cluster == 2, ]
cluster_3 <- famd_coords[famd_coords$Cluster == 3, ]

# Identifier les noms des individus pour chaque cluster

noms_cluster_1 <- rownames(cluster_1)
noms_cluster_2 <- rownames(cluster_2)
noms_cluster_3 <- rownames(cluster_3)

# Sélectionner les lignes correspondantes dans data pour chaque cluster

```

```

data_cluster_1 <- subset(data, rownames(data) %in% noms_cluster_1)
data_cluster_2 <- subset(data, rownames(data) %in% noms_cluster_2)
data_cluster_3 <- subset(data, rownames(data) %in% noms_cluster_3)
# Vérification des résultats
head(data_cluster_1)
# pour obtenir une information particuliere sur les clusters (groupes).
# Ici, l'indice de productivite moyen du groupe 1 est demande, ainsi que son ecart type
mean_SI_1 <- mean(data_cluster_1$SI)
sd_SI_1 <- sd(data_cluster_1$SI)
# verifier sur quelles dimensions la projection des clusters est la plus pertinente.
hc$desc.axes
# Etape 5 : Projection des clusters
# Ajouter les noms de lignes
famd_coords$name <- rownames(data_famd)
# Calculer les moyennes des coordonnées pour chaque cluster
centroids <- famd_coords %>%
  group_by(cluster) %>%
  summarize(across(starts_with("Dim"), mean))

# Visualiser les clusters avec des ellipses representant 95 % de la distribution des groupes dans la
FAMD
p <- fviz_cluster(hc,
  geom = "point",
  data = data_famd,
  main = FALSE,
  axes = c(1, 2),
  ellipse.type = "norm",
  ellipse.level = 0.95,
  repel = FALSE,
  pointsize = 1,
  labelsize = 2)

```

```

p +
  geom_text_repel(data = famd_coords, aes(x = Dim.1, y = Dim.2, label = name), size = 3) +
  geom_point(data = centroids, aes(x = Dim.1, y = Dim.2, color = as.factor(cluster)), size = 5) +
  geom_text(data = centroids, aes(x = Dim.1, y = Dim.2, label = paste("Cluster", cluster)), vjust = -1) +
  stat_ellipse(data = famd_coords, aes(x = Dim.1, y = Dim.2, group = cluster), type = "norm", level =
0.95)

# Etape 6 : Validation de la difference entre les groupes

# Préparation des données pour l'ANOVA : Combiner les données des trois clusters en un seul data
frame

data_cluster_1$Cluster <- as.factor(1)
data_cluster_2$Cluster <- as.factor(2)
data_cluster_3$Cluster <- as.factor(3)
combined_data <- rbind(data_cluster_1, data_cluster_2, data_cluster_3)

# l'ANOVA

anova_result <- aov(SI ~ Cluster, data = combined_data)

summary(anova_result)

# Validation de l'Anova :

# Extraction des résidus

residuals <- residuals(anova_result)

# Test de Shapiro-Wilk pour la normalité des résidus

shapiro.test(residuals)

# Q-Q plot pour les résidus

qqnorm(residuals)
qqline(residuals)

# Homoscedasticite

leveneTest(SI ~ Cluster, data = combined_data)

# Test post-hoc de Tukey pour vérifier quelles paires de clusters sont significativement différentes

tukey_result <- TukeyHSD(anova_result)

print(tukey_result)

# 3.4. La correlation entre les variables

#-----

```



```

# Une graphique de correlation n est possible qu entre des donnees quantitatives

# Selection des datas numerique uniquement avec la librairie dplyr

data_cor <- data_famd %>%
  select_if(is.numeric)

# Calculer la matrice de corrélation

correlation_matrix <- cor(data_cor)

# Afficher la matrice de corrélation avec corrplot

corrplot(correlation_matrix, method = "color", type = "lower",
  tl.col = "black", tl.srt = 70,
  addCoef.col = "black", number.cex = 0.4, # Ajuster la taille des coefficients
  col = colorRampPalette(c("#BB4444", "#EE9988", "#FFFFFF", "#77AADD", "#4477AA"))(100),
  title = "Matrice de Corrélation", mar = c(0,0,1,0))

```

4. Selection des variables par Foret aleatoire pour le modele A - definition du modele predictif.

```

#=====
=====

```

4.1. La Choix parmi les ensembles de variables correlees (Modele A)

```

#-----

```

Defintion d'hyperparamètres par defaults

```
seed<-1001 # pour le reproduction
```

```
trees <- 500 # nombre d'arbres aleatoires consideres
```

```
ndsz <- 2 # La division s arrete lorsque le groupe est egal ou inferieur a deux.
```

```
nbtry <- 3 # nombre de variables piochees aleatoirement et presentees a chaque noeuds de l'arbres
aleatoire (seule la plus efficace est retenue)
```

```
# selection des variables de terrain (pour les modèle A et B) ou cartographiques (pour les modèles C
et D)
```

```
data_rf_select <- data.frame(data[, -c("variables cartographiques ou de terrain")]) # selection des
variables de terrain pour le modele A
```

```
# Selection des variables cliamtiquesou de l'altitude
```

```

data_rf_sel_c <- data_rf_select[,-c("toutes les variables climatique + altitude, sauf celle conservée")]
# l'altitude est selectionnee par default pour le modele A

# Selections de variables topographique et pedologique

data_rf_sel <- data_rf_sel_c[,-c("topographiques et pedologiques non retenues")]

# Sur bases des variables choisies, des modeles sont formulees par foret aleatoire.

set.seed(seed) # pour la reproduction

a_rf_final <- randomForest(SI ~ ., data = data_rf_sel, importance = TRUE, ntree = trees, nodesize =
ndsz, mtry = nbtry) # Random forest du modele A

print(a_rf_final)

# Verification de l independance des variables selectionnees.

res.famd_sel <- FAMD(data_rf_sel, ncp = 6, sup.var = data_rf_sel$SI, ind.sup = NULL, graph = TRUE)

fviz_famd_var(res.famd_sel, axes = c(1,2), col.var.sup = "lightgrey", repel = TRUE)


# 4.2. Selection entre les variables restantes ; presentaiton de leur importance (Modele
A)
#-----

# Obtenir l importance des variables

importance <- importance(a_rf_final)

print(colnames(importance))

importance_df <- data.frame(Feature = rownames(importance),
                           IncMSE = importance[, "%IncMSE"],
                           IncNodePurity = importance[, "IncNodePurity"])

# Trier les variables par importance (IncMSE)

importance_df <- importance_df[order(importance_df$IncMSE, decreasing = TRUE), ]

print(importance_df)

# Visualiser l'importance par %IncMSE

ggplot(importance_df, aes(x = reorder(Feature, IncMSE), y = IncMSE)) +
  geom_bar(stat = "identity", width = 0.8) +
  coord_flip() +
  xlab("Facteurs") +
  ylab("Augmentation de l'erreur quadratique moyenne (MSE)") +
  theme(

```

```
axis.text.y = element_text(size = 11),
axis.title.y = element_text(size = 14),
axis.title.x = element_text(size = 14))
```

```
X<-data_rf_sel
y<-data_rf_sel$SI
```

4.2. Selection entre les variables restantes ; modele final et hyperparametres (Modele A)

```
#-----
```

```
# Definir un seuil pour choisir uniquement les variables les plus importante
```

```
top_n <- 2
```

```
{
```

```
selected_features <- importance_df$Feature[1:top_n]
```

```
selected_features <- c("Altitude","SWC_carto","NT_integratif") # pour choisir d'autres variables
que juste les plus importantes
```

```
print(selected_features)
```

```
# Filtrer les données pour inclure seulement les variables sélectionnées
```

```
X_selected <- X[, selected_features]
```

```
y <- data_rf_sel$SI
```

```
# Créer un nouveau data frame avec les variables sélectionnées et la variable cible
```

```
a_var_sel <- data.frame(X_selected, SI = y)
```

```
head(a_var_sel)
```

```
# Generating a random forest of 5000 trees
```

```
set.seed(seed)
```

```
SIRF.T5000 <- randomForest(SI ~ ., data = a_var_sel, ntree = 5000)
```

```
print(SIRF.T5000)
```

```
# Extraction du taux d'erreur "out-of-bag"
```

```
oob_error_rate <- SIRF.T5000$mse
```

```
# Graphique du taux d'erreur "out of bag"
```

```
plot(1:5000, oob_error_rate, type = "l",
```

```

xlab = "Number of trees", ylab = "OOB MSE",

main = "Evolution of the OOB error rate as a function of the number of trees")

# Add horizontal lines representing the OOB error at specific tree counts
abline(h = oob_error_rate[300], col = "red")
abline(h = oob_error_rate[500], col = "green")
abline(h = oob_error_rate[250], col = "blue")
abline(h = oob_error_rate[800], col = "purple")
abline(h = oob_error_rate[400], col = "yellow")
}

trees <- 300 # definition du nombre d'arbres utilises dans la foret

# Definiton de la valeur du nodesize
# =====
{
  set.seed(seed)

  # Initialize a list to record OOB errors
  oob.node <- NULL

  # Define different values of nodesize to test
  nodesize_values <- c(1, 2, 5, 10, 20, 50)

  # Loop to train a model for each nodesize value
  for (ns in nodesize_values) {
    rf <- randomForest(SI ~ ., data = a_var_sel, nodesize = ns, ntree = trees)
    oob.node <- rbind(oob.node, c(ns, rf$mse[trees]))
  }

  # Convert oob.node to a data frame for easier manipulation
  oob.node <- as.data.frame(oob.node)
  colnames(oob.node) <- c("nodesize", "OOB_MSE")
  print(oob.node)

```

```

# Plot the evolution of the OOB error with respect to nodesize
plot(oob.node$nodesize, oob.node$OOB_MSE, type = "b",
     xlab = "Minimum node size", ylab = "OOB MSE",
     main = "OOB MSE vs. Minimum node size")

# Find and display the minimum OOB MSE
min_oob_node <- min(oob.node$OOB_MSE)
print(min_oob_node)
}

ndsz <- 2 # definition du node size

# Definition du mtry
# =====
{
  set.seed(seed)

  # Initialize a list to record OOB errors
  oob.mtry <- NULL

  # Define the maximum number of mtry to test (number of predictor variables)
  max_mtry <- ncol(a_var_sel) - 1

  # Loop to train a model for each mtry value
  for (mt in 1:max_mtry) {
    rf_mtry <- randomForest(SI ~ ., data = a_var_sel, nodesize = ndsz, mtry = mt, ntree = trees)
    oob.mtry <- rbind(oob.mtry, c(mt, rf_mtry$mse[trees]))
  }

  # Convert oob.mtry to a data frame for easier manipulation
  oob.mtry <- as.data.frame(oob.mtry)
  colnames(oob.mtry) <- c("mtry", "OOB_MSE")

```

```

print(oob.mtry)

# Plot the evolution of the OOB error with respect to mtry
plot(oob.mtry$mtry, oob.mtry$OOB_MSE, type = "b",
     xlab = "Number of randomly selected features", ylab = "OOB MSE",
     main = "OOB MSE vs. Number of randomly selected features (mtry)")

# Find and display the minimum OOB MSE
min_oob_mtry <- min(oob.mtry$OOB_MSE)
print(min_oob_mtry)
}
nbtry <- 2

# FIN de la selection des hyper-parametres du random forest -----

# Defintion du modele final avec les meilleurs hyper parametres
set.seed(seed)
a_rf <- randomForest(SI ~ ., data = a_var_sel, importance = TRUE, ntree = trees, nodesize = ndsz,
mtry = nbtry)
print(a_rf)
print(selected_features)

# 4.3. Validation du modele par Bootstrap (Modele A)
#-----

# Reprise du modele initial
set.seed(seed)
a_rf <- randomForest(SI ~ ., data = a_var_sel, importance = TRUE, ntree = trees, nodesize = ndsz,
mtry = nbtry)
print(a_rf)

```

```

# Fonction pour entraîner le modele et calculer le R^2

train_rf_model <- function(data, indices) {

  train_data <- data[indices, ] # Sélectionne les échantillons bootstrap
  test_data <- data[-indices, ] # Utilise les échantillons restants pour le test

# Entraîner le modele

  rf_model <- randomForest(SI ~ ., data = train_data, importance = TRUE, ntree = trees, nodesize =
ndsz, mtry = nbtry)

# Prédire sur les donnees de test

  predictions <- predict(rf_model, newdata = test_data)

# Calculer le R^2

  y_true <- test_data$SI
  ss_res <- sum((y_true - predictions)^2)
  ss_tot <- sum((y_true - mean(y_true))^2)

# Eviter la division par zéro et les valeurs constantes

  if (ss_tot == 0 || length(unique(predictions)) == 1) {
    return(NA) # Retourne NA si ss_tot est 0 ou si les prédictions sont constantes
  }

  r_squared <- 1 - (ss_res / ss_tot)
  return(r_squared)
}

# Appliquer la validation bootstrap

set.seed(seed) # Utilisez votre propre seed si nécessaire

bootstrap_results <- boot(data = a_var_sel, statistic = train_rf_model, R = 1000)

# Afficher les résultats

print(bootstrap_results)

# Calculer le R^2 moyen et filtrer les NA

bootstrap_r2_values <- bootstrap_results$t[!is.na(bootstrap_results$t)]

bootstrap_r2_mean <- mean(bootstrap_r2_values)

sd(bootstrap_r2_values)

# Calculer l'intervalle de confiance manuellement si boot.ci échoue

if (length(bootstrap_r2_values) > 0) {

```

```

bootstrap_ci <- quantile(bootstrap_r2_values, c(0.025, 0.975))
print(paste("R^2 moyen :", bootstrap_r2_mean))
print("Intervalle de confiance à 95% :")
print(bootstrap_ci)
} else {
  print("Erreur: Pas assez de valeurs bootstrap valides pour calculer l'intervalle de confiance.")
}
# 4.3.      Valeur de Shapley - MODELE A
#-----
# Definition de la fonction prédictive pour le modèle Random Forest
predict_rf <- function(model, newdata) {
  predict(model, newdata)
}
# Reprise du model de Random Forest
set.seed(seed)
rf_final_select <- randomForest(SI ~ ., data = a_var_sel, importance = TRUE, ntree = trees, nodesize =
ndsz, mtry = nbtry)
print(rf_final_select)
# Calcul des SHAP values avec fastshap
shap_values <- explain(rf_final_select, X = a_var_sel[, -which(names(a_var_sel) == "SI")],
  pred_wrapper = predict_rf, nsim = 100)
# Conversion des SHAP values en format long
shap_df <- as.data.frame(shap_values)
shap_long <- shap_df %>%
  mutate(index = 1:nrow(shap_df)) %>%
  pivot_longer(cols = -index, names_to = "Variable", values_to = "SHAP_value")
# Séparation des colonnes numériques
numeric_cols <- a_var_sel %>%
  select(where(is.numeric)) %>%
  mutate(index = 1:nrow(a_var_sel))

```



```

# Standardisation des valeurs pour chaque variable numérique

numeric_standardized <- numeric_cols %>%
  mutate(across(-index, ~ (. - mean(.))/sd(.)))

# Conversion en format long

numeric_long <- numeric_standardized %>%
  pivot_longer(cols = -index, names_to = "Variable", values_to = "value")

# Fusion des valeurs normalisées avec les données SHAP long

shap_long_numeric <- shap_long %>%
  inner_join(numeric_long, by = c("index", "Variable"))

# Création du graphique pour les variables numériques

plot_numeric <- ggplot(shap_long_numeric, aes(x = SHAP_value, y = Variable, color = value)) +
  geom_violin(scale = "width", draw_quantiles = c(0.25, 0.5, 0.75)) +
  geom_jitter(width = 0.2, height = 0.2, size = 12, alpha = 0.4) +
  scale_color_viridis_c(option = "turbo") +
  theme_minimal() +
  labs(title = "SHAP Summary Plot - Numeric Variables",
       x = "SHAP Value",
       y = "Feature")
print(plot_numeric)

# Séparation des colonnes qualitatives

categorical_cols <- a_var_sel %>%
  select(where(is.factor)) %>%
  mutate(index = 1:nrow(a_var_sel))

# Conversion en format long

categorical_long <- categorical_cols %>%
  pivot_longer(cols = -index, names_to = "Variable", values_to = "value")

# Fusion des valeurs avec les données SHAP long

shap_long_categorical <- shap_long %>%
  inner_join(categorical_long, by = c("index", "Variable"))

# Création du graphique pour les variables qualitatives

```

```

plot_categorical <- ggplot(shap_long_categorical, aes(x = SHAP_value, y = Variable, color = value)) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(width = 0.2, height = 0.2, size = 10, alpha = 0.5) +
  scale_color_viridis_d(option = "turbo") +
  theme_minimal() +
  labs(title = "SHAP Summary Plot - Categorical Variables",
        x = "SHAP Value",
        y = "Feature") +
  theme(legend.position = "right", legend.box = "horizontal",
        legend.key.size = unit(1, "cm"), # Optional
        legend.text = element_text(size = 8),
        legend.title = element_text(size = 10, face = "bold")) +
  guides(color = guide_legend(ncol = 1)) # Légende en deux colonnes
print(plot_categorical)

```

5. Comparaison des modèles avec des données de terrain et des données uniquement cartographiques

```

#=====
=====

```

5.1. Bland et Altman, modèle A et C : Méthode 1

```
#-----
```

Créer un dataframe avec les prédictions de chaque modèle

```

df_corr_a_c <- data.frame(
  first = predict(a_rf, newdata = a_var_sel), # a_rf : modèle final de random forest pour A. a_var_sel :
variables sélectionnées pour le modèle A
  second = predict(c_rf, newdata = c_var_sel) # c_rf : modèle final de random forest pour C. c_var_sel
: variables sélectionnées pour le modèle C
)

```

Création d'un graphique de Bland-Altman

```

bland.altman.plot(df_corr_a_c$first, df_corr_a_c$second, main = "Bland-Altman Plot",
  xlab="mean measurement",
  ylab="differences",
  mode=1)

```

```

B_A_plot<-bland.altman.plot(df_corr_a_c$first, df_corr_a_c$second, graph.sys = "ggplot2",
conf.int=.95)

B_A_plot

# Ajout d un ligne de regression

pander(bland.altman.stats(df_corr_a_c$first,df_corr_a_c$second, two = 1.96, mode =1, conf.int =
0.95))

# 5.1. Bland et Altmann, modele A et C : Methode 2

#-----

# Faire les prédictions avec les deux modèles

pred_rf_a <- predict(a_rf, newdata = a_var_sel)

pred_rf_c <- predict(c_rf, newdata = c_var_sel)

# Calculer la moyenne et la différence des prédictions

moyenne_pred <- (pred_rf_a + pred_rf_c) / 2

difference_pred <- pred_rf_a - pred_rf_c

# Créer un data frame pour le tracé

df_bland_altman <- data.frame(

Moyenne = moyenne_pred,

Différence = difference_pred

)

# Calculer les limites d'accord ( $Moyenne \pm 1.96 * \text{ecart-type}$ )

mean_diff <- mean(df_bland_altman$Différence)

sd_diff <- sd(df_bland_altman$Différence)

upper_limit <- mean_diff + 1.96 * sd_diff

lower_limit <- mean_diff - 1.96 * sd_diff

# Tracer le diagramme de Bland-Altman

ggplot(df_bland_altman, aes(x = Moyenne, y = Différence)) +

geom_point() +

geom_hline(yintercept = mean_diff, color = "blue", linetype = "dashed") +

geom_hline(yintercept = upper_limit, color = "red", linetype = "dashed") +

geom_hline(yintercept = lower_limit, color = "red", linetype = "dashed") +

labs(title = "Diagramme de Bland-Altman pour les modeles avec altitude ",

x = "Moyenne des prédictions de SI",

```

y = "Différence des prédictions de SI A et C") +

theme_minimal()

FIN

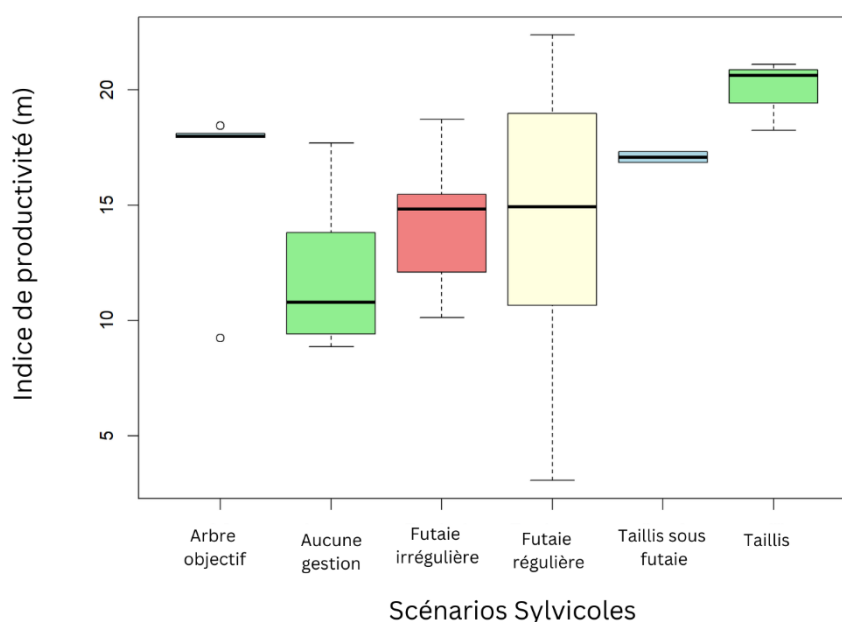
Annexe 7 : Influence des actions anthropiques sur l'indice de fertilité (H30).

Scénarios sylvicoles

L'analyse de similarité des moyennes est observée par le biais d'une anova (ci-dessous).

	Degrés de liberté	La somme des carrés	La moyenne des carrés	La valeur du test de Fisher	La p-valeur du test de Fisher
Cluster	5	178,4	35,69	1,931	0,111
Résidus	40	739,1	18,48		

Les résultats indiquent qu'aucune moyenne n'est significativement différente, et que le scénario sylvicole n'a pas d'impact sur l'indice de fertilité. Cependant, le test de Shapiro-Wilk, qui vérifie la normalité des résidus, n'a pas pu être réalisé en raison du faible nombre de représentants pour chaque itinéraire sylvicole. Bien que le test de Levene, qui évalue l'homogénéité des variances, ait été validé, ses résultats doivent être interprétés avec prudence. Une analyse visuelle des boîtes à moustaches (voir ci-dessous) confirme néanmoins l'indépendance de l'indice de fertilité.



La présence de drains

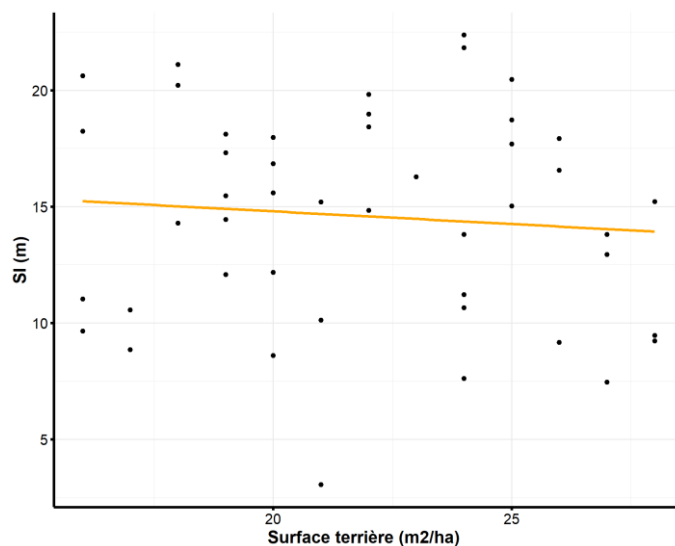
L'analyse de similarité des moyennes est observée par le biais d'une anova (ci-dessous).

	Degrés de liberté	La somme des carrés	La moyenne des carrés	La valeur du test de Fisher	La p-valeur du test de Fisher
Cluster	1	13,9	13,85	0,674	0,416
Résidus	44	903,7	20,54		

Le test ANOVA confirme que l'hypothèse nulle est respectée, indiquant que les moyennes sont identiques. La présence de drains n'a donc pas d'influence sur l'indice de fertilité. Les tests de Shapiro-Wilk et de Levene valident également les conditions d'application de l'ANOVA.

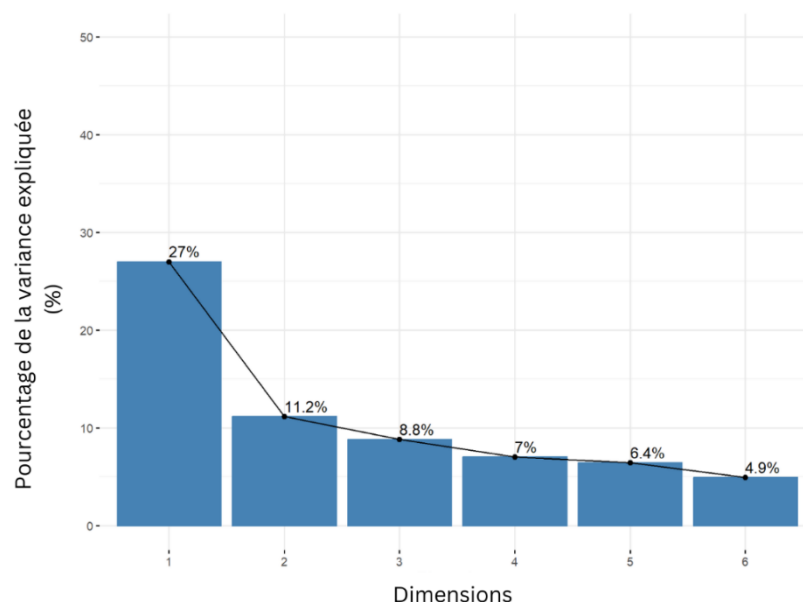
La surface terrière

La surface terrière reflète le scénario sylvicole. Pour vérifier l'indépendance de l'indice de fertilité par rapport à celui-ci, une régression linéaire simple a été effectuée entre l'indice de fertilité (SI) et la surface terrière (voir ci-dessous). Les conditions d'application de la régression, à savoir la normalité et l'homoscédasticité des résidus, ont été vérifiées. La régression est caractérisée par un coefficient de détermination de 0,00774, indiquant que la surface terrière explique très peu la variabilité de l'indice de fertilité et peut donc être considérée comme négligeable.



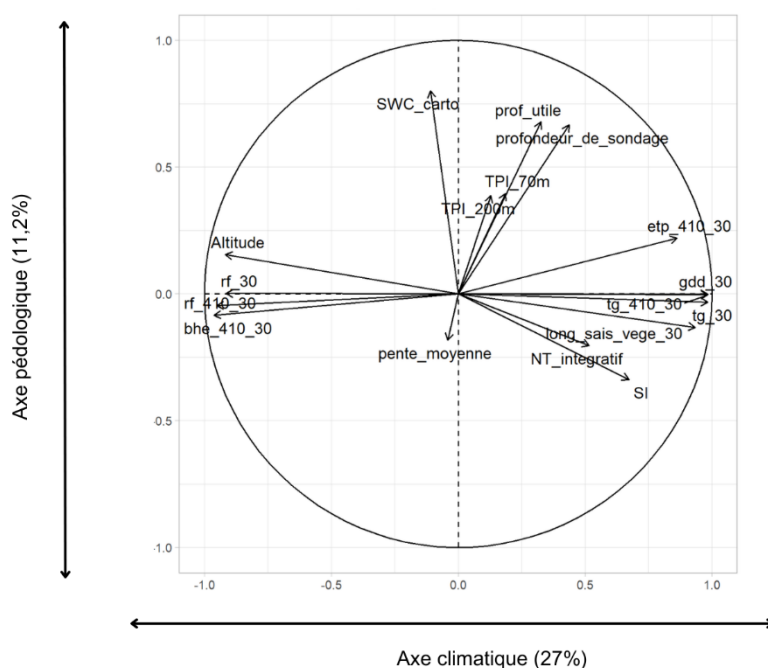
Annexe 8 : Prétraitements complémentaires sur les données des modèles A et B (avec les données de terrain).

Diagramme d'éboulis de la FAMD



L'analyse du diagramme d'éboulis permet de définir le nombre de composantes principales de la FAMD à considérer. Il présente les différentes dimensions en abscisse et le pourcentage de variance expliquée par celles-ci en ordonnée. Ce pourcentage diminue au fil des dimensions, ce donne à cet histogramme une forme de coude. Les composantes à conserver sont celles allant jusqu'au ce point d'inflexion du coude. Ici, deux composantes sont conservées.

Représentation des variables quantitatives sur la FAMD

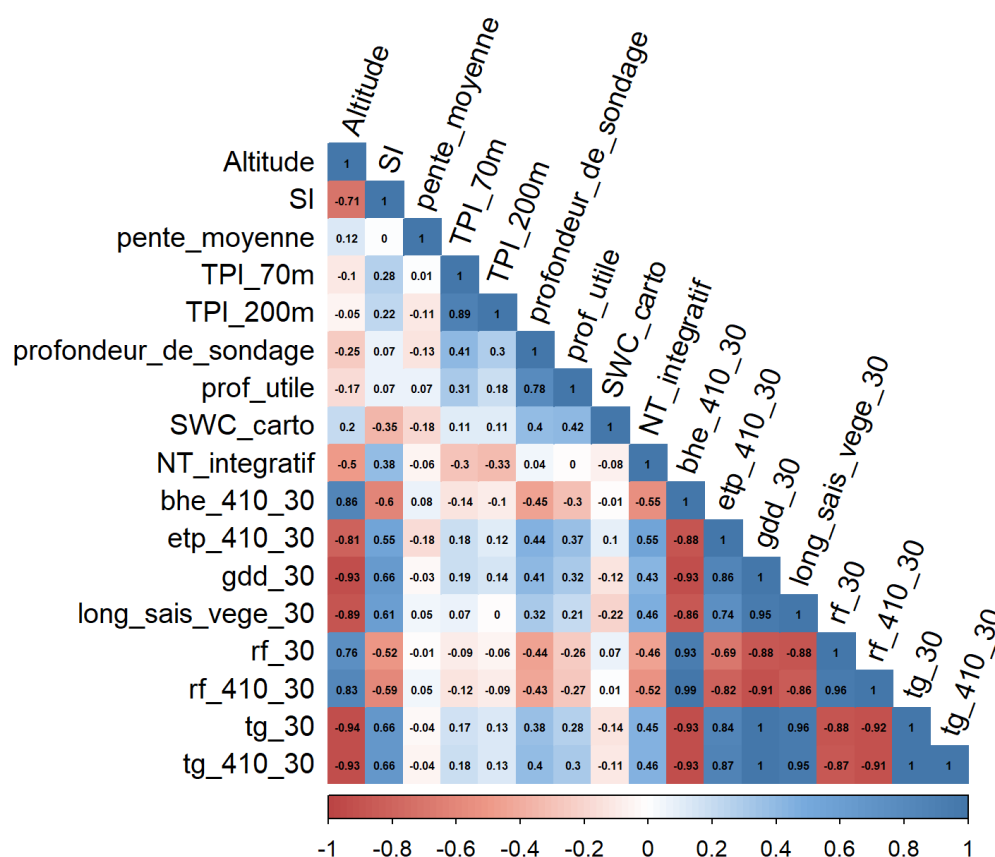


La FAMD quantitative présente les variables numériques et renseigne sur leurs corrélations. Celles-ci peuvent être avec les dimensions principales, entre variables explicatives, ou avec la variable à expliquer (SI). Plus deux flèches sont proches, plus les variables représentées sont corrélées. Des flèches opposées à 180° indiquent des variables inversement proportionnelles. De plus, plus une flèche a une grande projection sur un axe, plus elle y est fortement corrélée.

Ainsi, on peut observer la formation de deux axes de variables : un axe climatique et un axe pédologique. Il est également possible de déduire une série de corrélations entre les variables explicatives. La variable de niveau trophique (NT) apparaît corrélée aux ensembles de données climatiques, tout comme SI, la variable à expliquer.

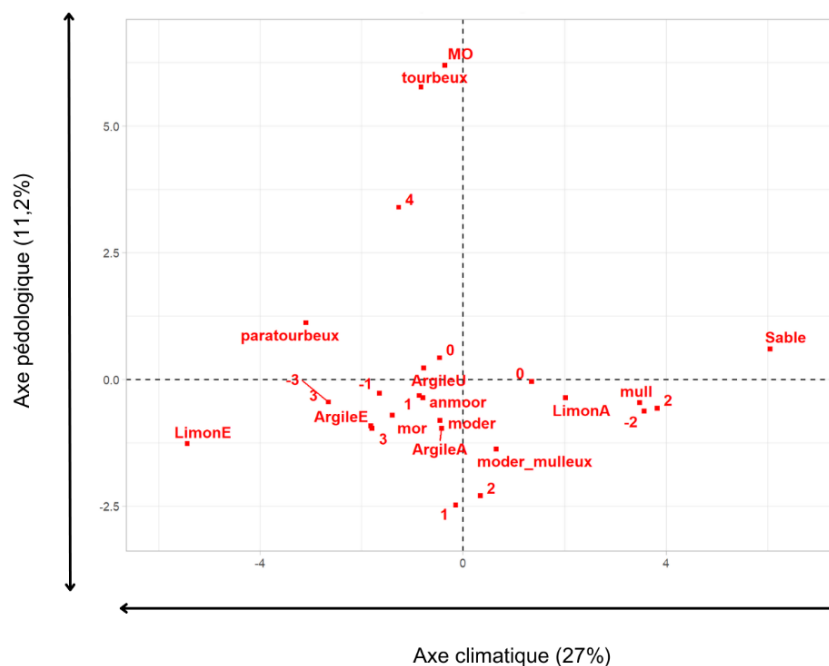
Matrice de corrélation

La matrice de corrélation (ci-dessous) permet de poser des valeurs chiffrées sur ces relations entre variables. Alors que les corrélations entre les variables climatiques, avec la variable d'altitude, sont entre 70 et 100 %, celles avec le niveau trophique sont de maximum 55 %. Par exemple, la corrélation est de 45 % entre le niveau trophique et la température moyenne de la saison de végétation.



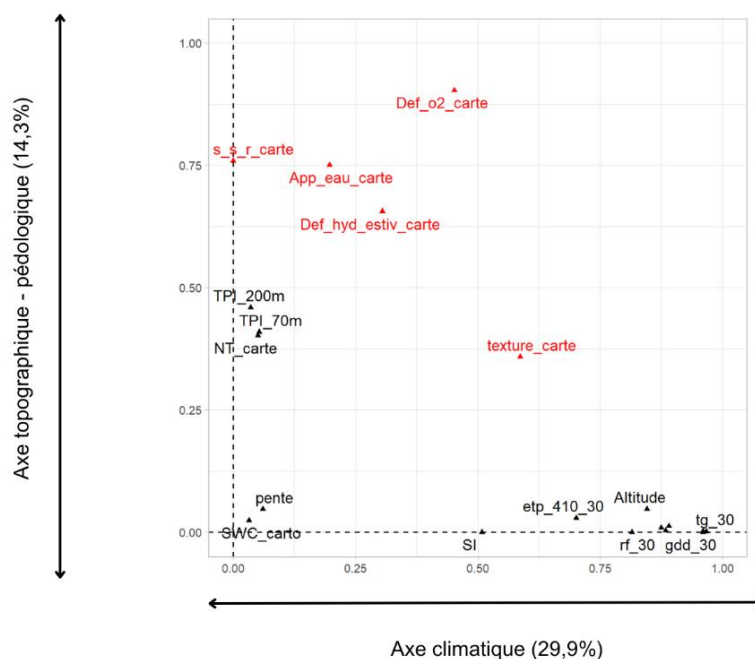
Représentation des variables qualitatives sur la FAMD

Les variables qualitatives voient leur différentes valeurs être également représentées selon leur importance pour les dimensions 1 et 2 de la FAMD (ci-dessous). Il est notamment possible d'observer les influences des textures et des humus sur les dimensions principales. Celles du sable (Sable) et du limon (LimonE) semblent fortement corrélées avec l'axe climatique. Les sol tourbeux semblent quant à eux déterminants pour l'axe pédologique (tourbeux – MO).



Annexe 9 : Prétraitements sur les données des modèles C et D (avec les données cartographiques uniquement).

FAMD

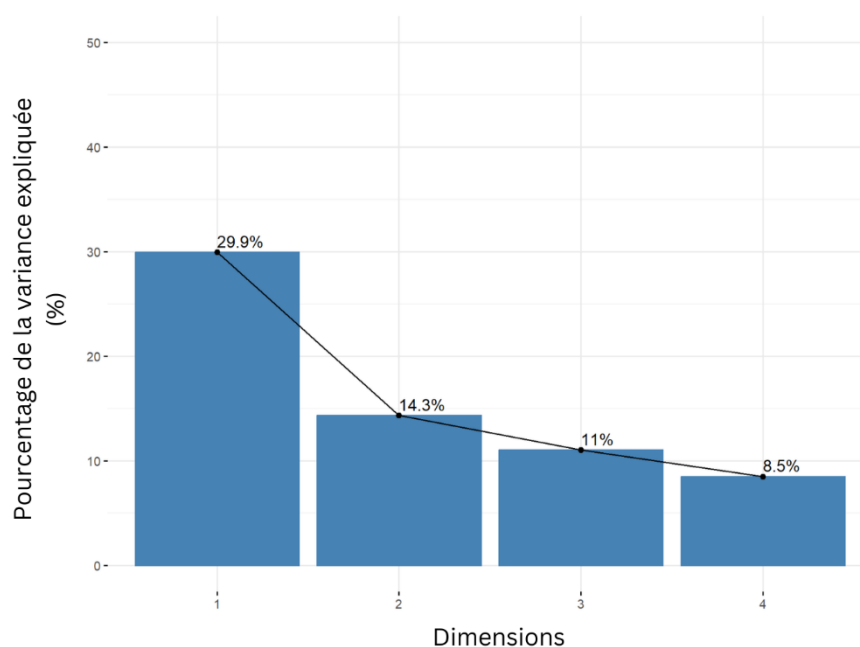


La FAMD ci-dessus représente les variables qualitatives et quantitatives sur ses axes principaux. Elle permet d'identifier plusieurs ensembles de variables corrélées, de par leur proximité. Un premier ensemble reprend les données climatiques et altitudinale. Un deuxième ensemble est celui des indices

de position topographique. Finalement, un troisième ensemble est identifié comme reprenant les ensembles de sous-secteurs radiatifs, d'apport en eau hors précipitation et de déficit hydrique estival.

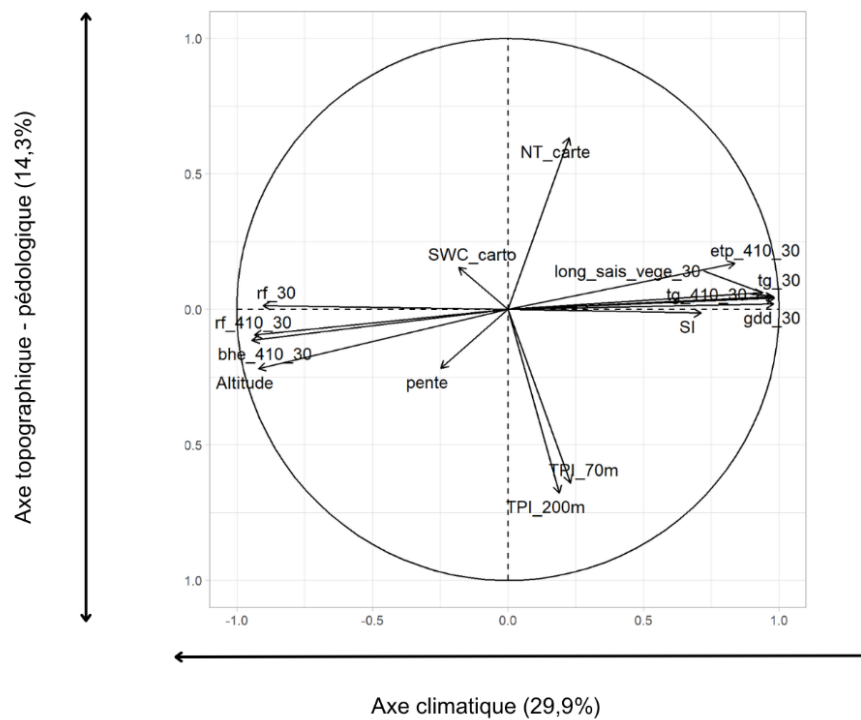
A nouveau, la première dimension est celle d'un axe corrélé aux données climatiques. Le second est autant représenté par des données topographiques que pédologiques.

Diagramme d'éboulis de la FAMD



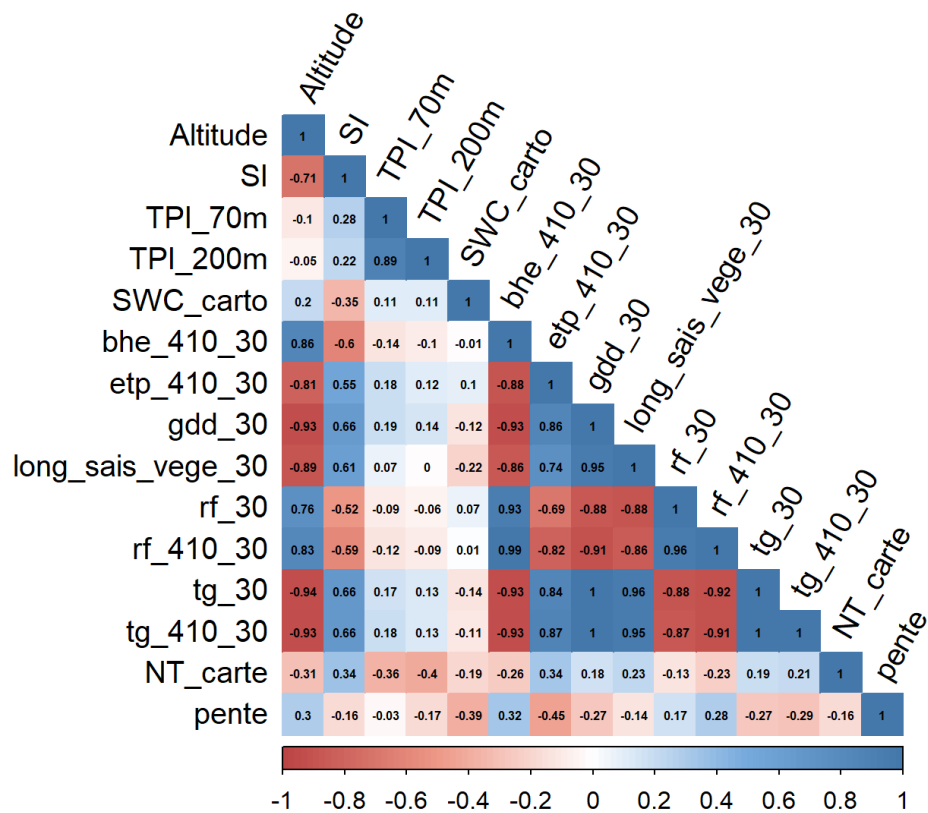
Le point d'inflexion du diagramme en éboulis se situant au-dessus de la deuxième dimension, seules les deux premières dimensions sont conservées.

Représentation des variables quantitatives sur la FAMD



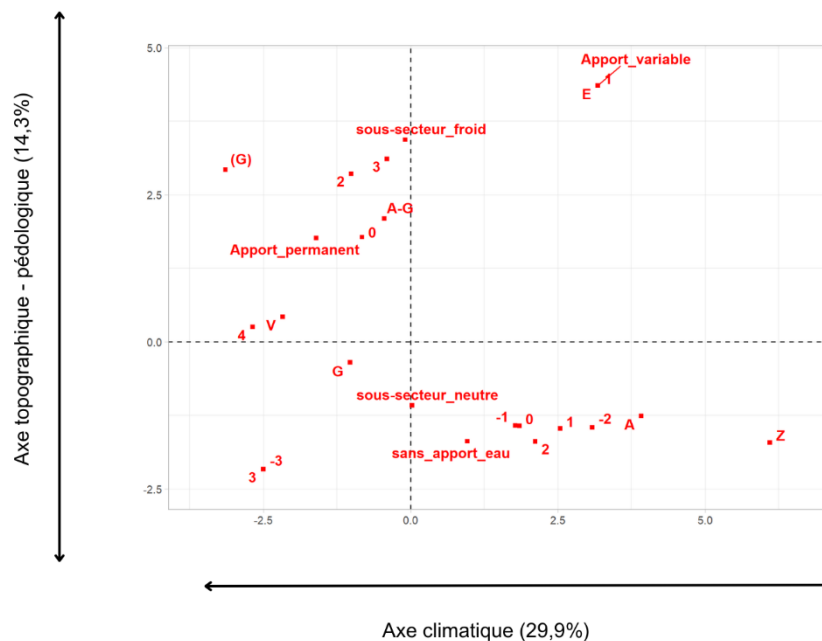
La représentation des variables quantitatives sur la FAMD ci-dessus permet de confirmer les ensembles de variables corrélées climatiques et topographiques. Il est à noter que le niveau trophique, cette fois-ci, n'est pas corrélé avec les données climatiques.

Matrice de corrélation



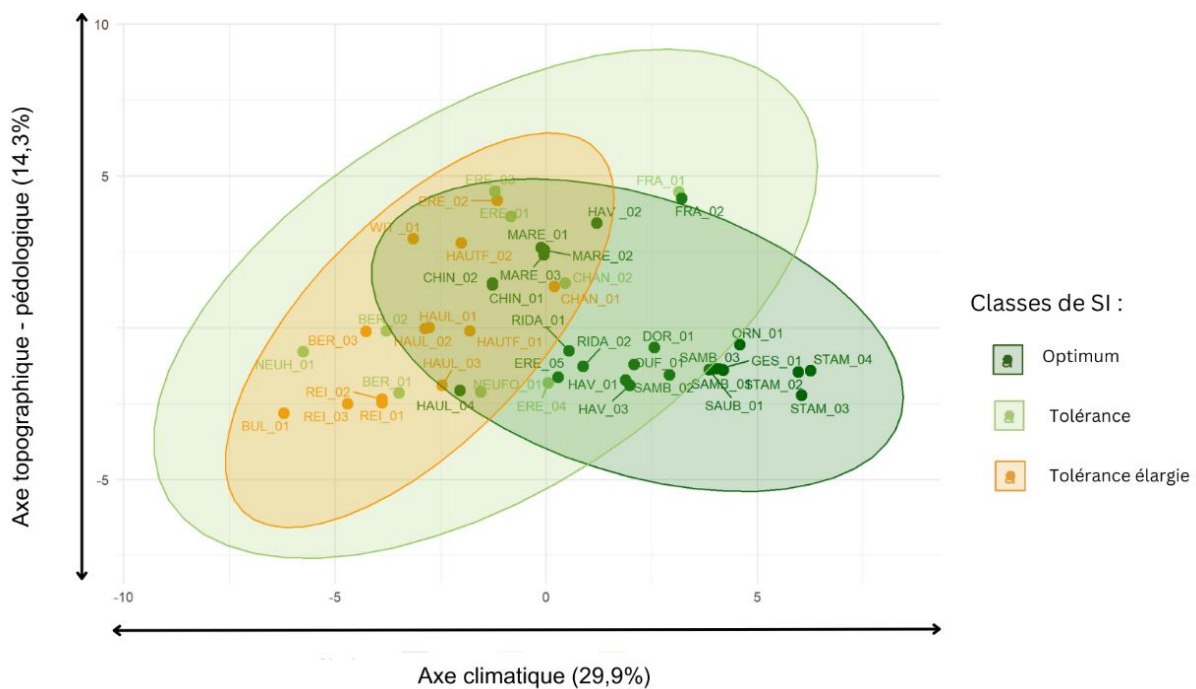
Une corrélation significative est observée dans la matrice de corrélation, entre les variables climatiques, y compris l'altitude.

Représentation des variables qualitatives sur la FAMD



Les variables qualitatives voient, ci-dessus, leur différentes valeurs être également représentées selon leur importance pour les dimensions 1 et 2 de la FAMD. Au sein des catégories de texture du sol, il est, entre autre, relevé que le sable (Z) semble très fortement corrélé avec les hautes températures (à droites sur la FAMD).

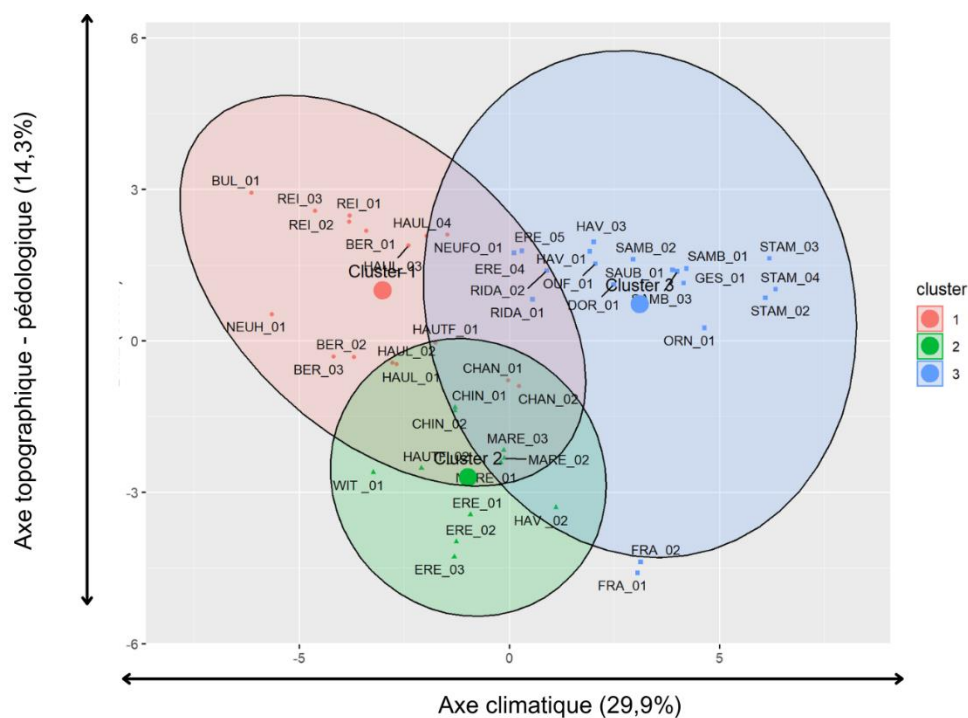
Représentation des individus sur la FAMD



Ci-dessus, la distribution des individus sur les composantes principales de la FAMD des variables utilisées pour la construction des modèles C et D. Chaque individu est identifié par le nom de

la placette. Elles sont également renseignée par leur classe d'aptitude. Une ellipse renseigne la zone de présence associée à chaque classe d'aptitude.

Regroupement des individus sur la FAMD



Un regroupement des placettes selon les composantes principales de la FAMD des modèles C et D permet de distinguer trois ensembles de placettes (clusters), qui se chevauchent partiellement.

Les indices de productivité moyens des différents clusters ont des écart types qui se chevauchent partiellement.

/	SI moyen (m)	Ecart – type du SI (m)
Cluster 1	10,12	3,13
Cluster 2	14,9	3,83
Cluster 3	18	2,57

L'anova distingue au moins deux clusters aux indices de fertilité moyens statistiquement différents.

	Degrés de liberté	La somme des carrés	La moyenne des carrés	La valeur du test de Fisher	La p-valeur du test de Fisher
Cluster	2	479	240	23,5	1,25 ^e -07 ***
Résidus	43	438	10,2		

Une comparaison de leurs moyennes avec le test de Tukey permet de distinguer significativement les différents clusters, deux à deux.

	Différence des moyennes	Borne inférieure	Borne supérieur	Valeur de p ajustée
2-1	4,79	1,91	7,67	0.000632
3-1	7,88	5,08	10,7	0.0000001
3-2	3,09	0,347	5,84	0.0241

Annexe 10 : Sélection parmi chaque ensemble de variable corrélées

Modèle A : Choix parmi les variables pédologiques et topographiques

Combinaisons des variables	R ² de calibration
prof_utile + TPI_70m	38,71
Profondeur_de_sondage + TPI_200m	40,44
Profondeur_de_sondage + TPI_70m	41,53
prof_utile + TPI_200m	42,74

Modèle B : Choix parmi les variables climatiques

Variable climatique	R ² de calibration
Altitude	43,15
bhe_410_30	41,16
etp_410_30	35,94
gdd_30	48,53
long_sais_vege_30	42,14
rf_30	43,38
rf_410_30	39,17
tg_30	45,95
tg_410_30	48,62

Modèle B : Choix parmi les variables pédologiques et topographiques

Combinaisons des variables	R ² de calibration
Profondeur_de_sondage + TPI_200m	47,32
Profondeur_de_sondage + TPI_70m	43,71
prof_utile + TPI_70m	43,6
prof_utile + TPI_200m	48,59

Modèle C : Choix parmi les variables topographiques

Variables topographiques	R ² de calibration
ssr_carte	47,97
App_eau_carte	48,97
def_hyd_est	46,53
TPI_200m	48
TPI_70m	49,33

Modèle C : Choix parmi les variables climatiques

Variables climatiques	R ² de calibration
Altitude	49,71
bhe_410_30	48,19
etp_410_30	45,83
gdd_30	56,81
long_sais_vege_30	53,03
rf_30	53,87
rf_410_30	49,93
tg_30	53,89
tg_410_30	56,82

Modèle D : Choix parmi les variables topographiques

Variables topographiques	R ² de calibration
Ssr_carte	51,98
def_hyd_est	53,29
TPI_70m	53,31
TPI_200m	53,77
app_eau_carte	54,53

Annexe 11 : Sélection parmi les variables non corrélés

Chaque tableau présente les différentes combinaisons de variables testées pour les modèles A, B, C et D. La première colonne indique le nombre de variables utilisées dans le modèle, tandis que la deuxième colonne affiche le R^2 de calibration. Si ce dernier est suffisamment élevé, une validation par Bootstrap est effectuée. La moyenne du R^2 de validation est reportée dans la quatrième colonne, avec son écart-type dans la cinquième. Enfin, la dernière colonne liste les hyperparamètres du modèle de forêt aléatoire utilisé.

Modèle A

Nombre de variables	Variables du modèle	R2_c	R2_v : moyenne	R2_v : écart-type	Hyper-paramètres
2	Altitude + NT_integratif	56,33	0,3758	0,2886	300-2-2
3	Altitude + NT_integratif + TPI_200m	49,96			400-2-2
3	Altitude + NT_integratif + type_d_humus	54,04	0,3742	0,2215	250-2-2
3	Altitude + TPI_200m + type_d_humus	56,88	0,3851	0,2268	250-2-2
3	Altitude + SWC_carto + type_d_humus	52,11			400-2-2
3	Altitude + SWC_carto + Def_o2	41,39			800-2-2
3	Altitude + NT_integratif + Def_o2	44			500-2-2
3	Altitude + SWC_carto + NT_integratif	58,13	0,4071	0,2293	300-2-2
3	Altitude + SWC + TPI_200m	53,33			400-2-2
3	Altitude + NT_integratif + prof_utile	52,48			800-2-2
3	Altitude + SWC_carto + prof_utile	54,31			800-2-2
3	Altitude + NT + prof_utile	52,48			800-2-2
3	Altitude + TPI_200m + prof_utile	52,08			800-2-2
4	Altitude + NT_integratif + TPI_200m + prof_utile	49,01			800-2-2
4	Altitude + NT_integratif + TPI_200m + type_d_humus	55,04	0,4092	0,2042	300-2-2
4	Altitude + tpye_d_humus + TPI_200m + prof_utile	53,59	0,3944	0,2086	300-2-2

Modèle B

Nombre de variables	Variables du modèle	R2_c	R2_v : moyenne	R2_v : écart type	Hyper-paramètres
2	tg_410_30 + TPI_200m	50,98			900-2-2
3	tg_410_30 + TPI_200m + type_d_humus	57,51	0,4323	0,1908	1000-2-2
3	tg_410_30 + TPI_200m + NT_integratif	54,47			500-2-2
3	tg_410_30 + type_d_humus + prof_utile	57,87	0,4291	0,1842	1000-2-2
3	tg_410_30 + TPI_200m + prof_utile	55,36			900-2-2
3	tg_410_30 + NT_integratif + type_d_humus	55,74			500-2-2
3	tg_410_30 + NT_integratif + SWC_carto	58,54	0,4356	0,2109	900-2-2
3	tg_410_30 + TPI_200m + SWC_carto	52,47			300-2-2
3	tg_410_30 + type_d_humus + SWC_carto	51,65			1100-2-2
3	tg_410_30 + TPI_200m + prof_utile	55,22			500-2-2
3	tg_410_30 + NT_integratif + Def_o2	51,35			500-2-2
3	tg_410_30 + TPI_200m + Def_o2	50,09			500-2-2
3	tg_410_30 + SWC_carto + prof_utile	57,81	0,4776	0,183	300-2-2
4	tg_410_30 + type_d_humus + SWC_carto + TPI_200m	55,13			1100-2-2
4	tg_410_30 + NT_integratif + SWC_carto + TPI_200m	55,85			300-2-2
4	tg_410_30 + NT_integratif + prof_utile + TPI_200m	56,58			300-2-2
5	tg_410_30 + NT_integartif + SWC_carto + prof_utile + TPI_200m	59,83	0,4677	0,1879	900-2-3

Modèle C

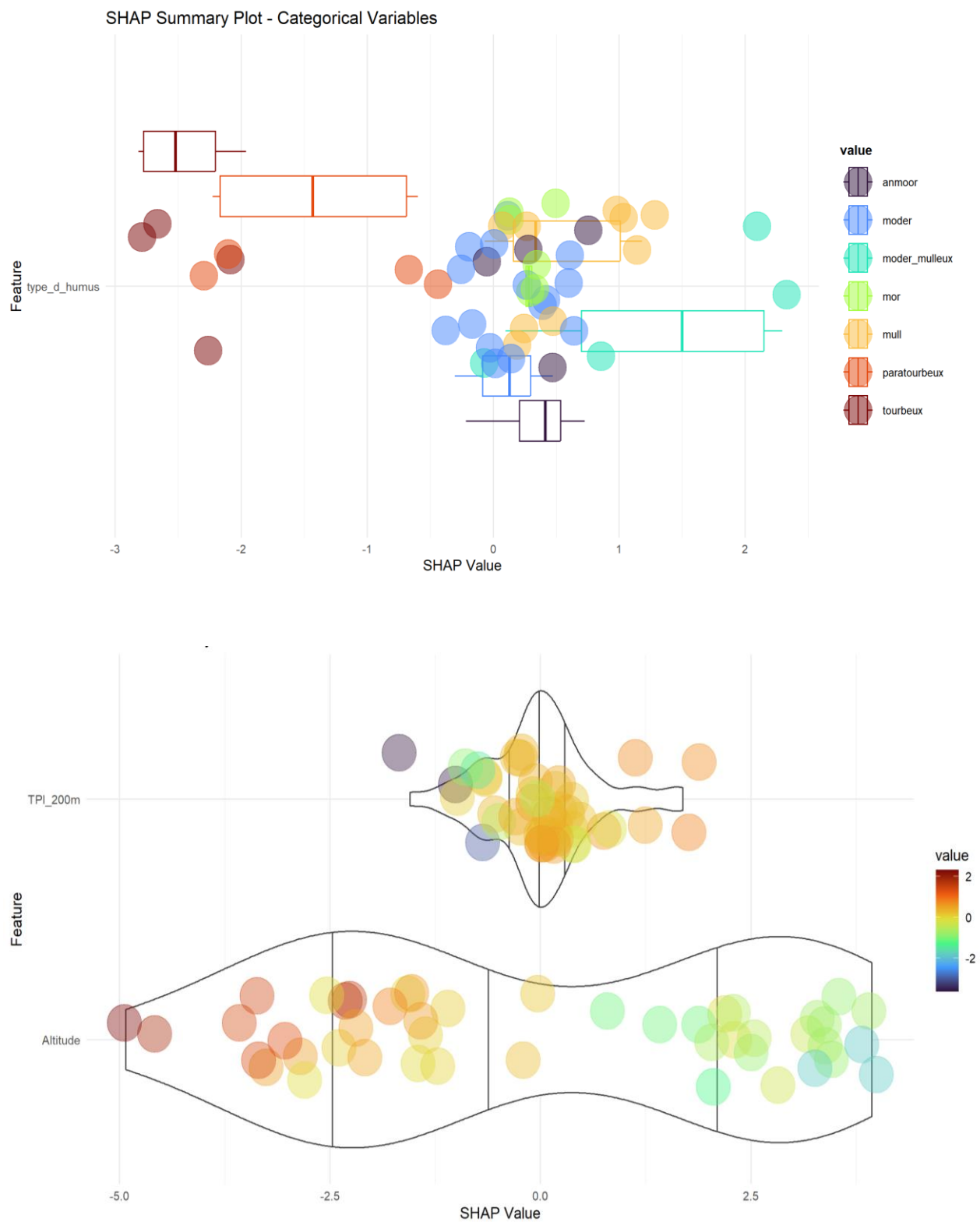
Nombre de variables	Variables du modèle	R2_c	R2_v : moyenne	R2_v : écart-type	Hyper-paramètres
2	Altitude + Def_o2_carte	45,98			100-2-2
2	Altitude + NT_carte	58,12	0,3988	0,2697	400-2-2
2	Altitude + TPI_70m	55,72			400-2-2
2	Altitude + SWC_carto	54,36			300-2-2
2	Altitude + pente	51,81			300-2-2
2	Altitude + texture_carte	55,68	0,3988	0,2697	400-2-2
3	Altitude + NT_carte + texture_carte	58,19	0,4481	0,218	400-2-2
3	Altitude + NT_carte + SWC_carto	54,45			400-2-2
3	Altitude + NT + def_o2_carte	49,56			1000-2-2
3	Altitude + NT_carte + pente	51,78			300-2-2
3	Altitude + NT_carte + TPI_70m	57,65	0,4105	0,2192	400-2-2
3	Altitude + TPI_70m + Def_o2_carte	50,67			400-2-2
3	Altitude + TPI_70m + SWC_carto	56,08	0,4195	0,2156	250-2-2
3	Altitude + TPI_70m + texture_carte	59,56	0,4369	0,2189	300-2-2
3	Altitude + texture_carte + pente	48,98			300-2-2
3	Altitude + texture_carte + SWC_carto	53,68			300-2-2
3	Altitude + texture_carte + def_o2_carte	45,19			400-2-2
3	Altitude + SWC_carto + Def_o2_carte	48,86			500-2-2
4	Altitude + NT_carte + TPI_70m + Def_o2_carte	55,69			400-2-2
4	Altitude + texture_carte + TPI_70m + Def_o2_carte	50,35			500-2-2
4	Altitude + texture_carte + TPI_70m + NT_carte	60,63	46,23		1000-2-2
4	Altitude + SWC_carto + TPI_70m + NT_carte	56,95			1000-2-2

Modèle D

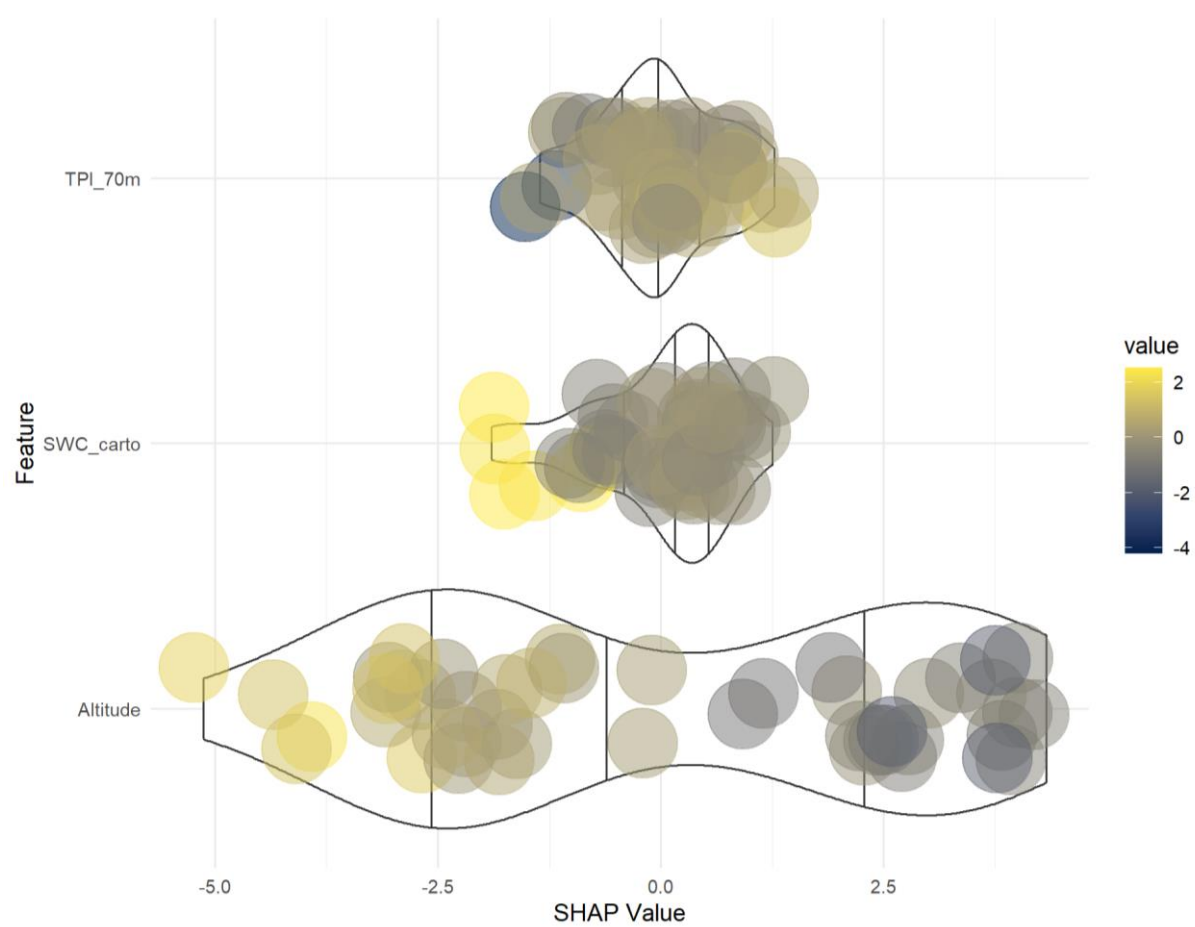
Nombre de variables	Variables du modèle	R2_c	R2_v : moyenne	R2_v : écart-type	Hyper-paramètres
2	tg_410_30 + Def_o2_carte	51,78			500-2-2
2	tg_410_30 + pente	45,73			800-2-2
2	tg_410_30 + texture_carte	54,72			500-2-2
2	tg_410_30 + SWC_carto	48,86			250 - 2 -2
2	tg_410_30 + NT_carte	50,84			1000-2-2
2	tg_410_30 + App_eau_carte	48,06			300-2-2
3	tg_410_30 + Def_o2_carte + pente	54,47	0,4354	0,1893	700-2-2
3	tg_410_30 + Def_o2_carte + texture_carte	57,02	0,4171	0,227	250-2-2
3	tg_410_30 + Def_o2_carte + SWC_carto	51,88			1000-2-2
3	tg_410_30 + def_o2_carte + NT_carte	53,62	0,381	0,2454	1000-2-2
3	tg_410_30 + def_o2_carte + App_eau_carte	53,07			1000-2-2
3	tg_410_30 + NT_carte + pente	50,45			500-2-2
3	tg_410_30 + NT_carte + SWC_carto	53,02			5000-2-2
3	tg_410_30 + NT_carte + texture_carte	58,42	0,4619	0,2246	3000-2-2
3	tg_410_30 + NT_carte + App_eau_carte	52,58			300-2-2
3	tg_410_30 + texture_carte + pente	52,2			700-2-2
3	tg_410_30 + texture_carte + SWC_carto	52,78			250-2-2
3	tg_410_30 + texture_carte + SWC_carto	53,67	4239	0,2109	700-2-2
3	tg_410_30 + SWC_carto + App_eau_carte	51,92			500-2-2
4	tg_410_30 + Def_o2_carte + texture_carte + NT_carte	54,71	0,4271	0,1806	3000-2-2
4	tg_410_30 + SWC_carto + texture_carte + NT_carte	56,43			700-2-2
4	tg_410_30 + SWC_carto + texture_carte + Def_o2_carte	55,65			250-2-2
4	tg_410_30 + def_o2_carte + NT_carte + SWC_carte	54,05			700-2-2
4	tg_410_30 + def_o2_carte + NT_carte + pente	55,13	0,4124	0,2112	3000-2-3
4	tg_410_30 + def_o2_carte + SWC_carto + pente	52,17			300-2-3

Annexe 12 : Valeurs de Shapley pour les modèles A et C.

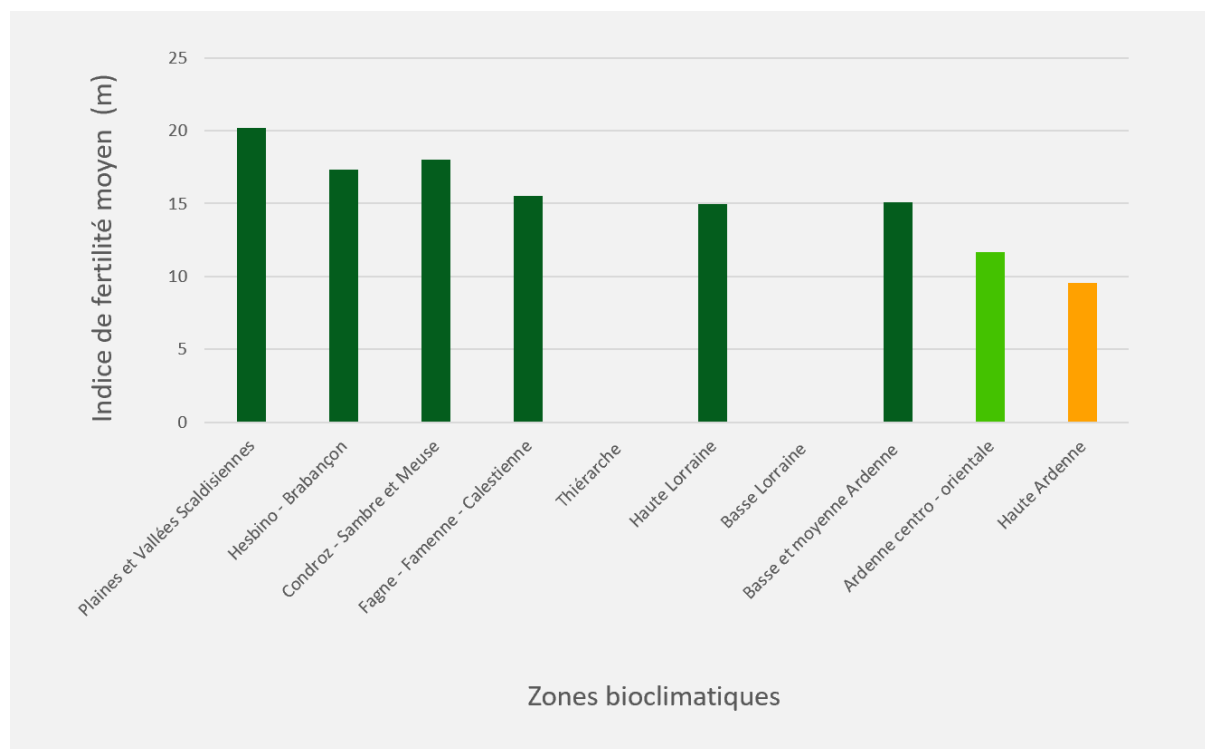
Modèles de Shapley pour le modèle A



Modèles de Shapley pour le modèle C



Annexe 13 : Productivité du bouleau pubescent selon les zones bioclimatiques.



Annexe 14 : Nombre d'échantillons par écotope

NT	NH	Zones bioclimatiques								Hors Ardenne	Ardenne
		Ardenne centro-orientale	Basse et moyenne Ardenne	Condroz-Sambre et Meuse	Fagne - Famenne - Calestienne	Haute Ardenne	Haute Lorraine	Hesbino - Brabanton	Plaines et Vallées Scaldiennes		
-2 et -3	-3RHA	1	0	0	0	2	0	0	0	0	3
-2 et -3	-2RHA	0	1	0	0	0	0	0	0	0	1
-2 et -3	-1RHA	0	1	0	0	3	0	0	1	1	4
-2 et -3	-4	5	0	0	0	1	0	0	0	0	6
-2 et -3	-2 et -3	0	1	0	0	0	2	0	0	2	1
-2 et -3	0 et -1	0	0	1	0	1	0	0	0	1	1
-2 et -3	1	0	0	0	0	1	0	0	0	0	1
-2 et -3	2	0	0	0	0	0	0	0	1	1	0
-1 et 0	-3RHA	0	0	0	0	0	0	0	0	0	0
-1 et 0	-2RHA	0	0	4	1	0	0	1	0	6	0
-1 et 0	-1RHA	0	0	0	1	0	0	0	0	1	0
-1 et 0	-4	0	0	0	0	0	0	0	0	2	0
-1 et 0	-2 et -3	1	4	0	1	0	0	0	0	1	5
-1 et 0	0 et -1	0	2	1	0	0	0	0	1	2	2
-1 et 0	1	0	2	2	0	0	0	0		2	2
1	-3RHA	0	0	1	0	0	0	0	0	1	0