

Utilisation conjointe de l'analyse factorielle et de l'analyse de variance dans l'étude de la contamination des parcelles bocagères par les insectes ailés

Odile LEDRU et P. UNGEMACH (1)

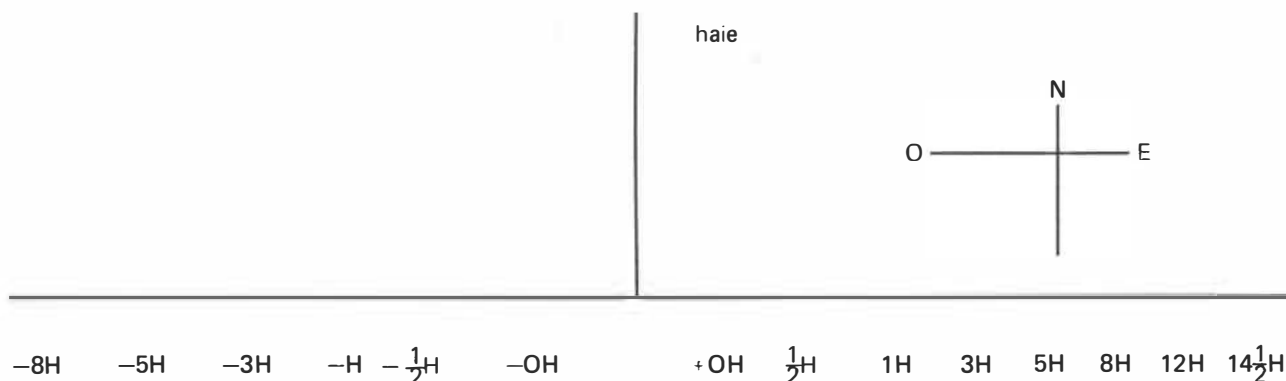
I – ÉTUDE DES DONNÉES 1974 : PRÉSENTATION.

1 – Les facteurs de la variabilité connus a priori.

On veut étudier la répartition de différentes espèces de pucerons en phase ailée sur une parcelle bocagère; la culture est la pomme de terre, le site est

celui de Réguiny.

Les pucerons sont capturés par des pièges jaunes à eau situés selon un transect perpendiculaire à la haie NS, à différentes distances de part et d'autre de cette haie. Plus précisément, en prenant comme unité la hauteur H de la haie (H = 10 m environ) le schéma de piégeage est le suivant :



Chaque station de piégeage consiste en un piège bas (au sol) et un piège haut (à 70 cm); pour chaque station on a des relevés de température (mini et maxi) et de vitesse moyenne du vent pendant l'intervalle de temps entre deux relevés successifs. L'expérience s'est déroulée du jour 135 au jour 205 de l'année 1974 (c'est-à-dire de la mi mai vers la fin juillet). Les relevés ont eu lieu en général trois jours par semaine, une fois par jour, sauf lors des cinquième et neuvième semaines où il y a eu deux ou trois relevés par jour (du lundi au vendredi).

2. Contraintes sur les relevés.

a) Du fait du peu d'homogénéité du facteur temps (les relevés aux différentes dates ont des poids complètement différents, l'intervalle de temps du relevé variant entre quelques heures et plusieurs jours), il était indispensable de travailler sur des relevés hebdomadaires (10 au total). Cette procédure a l'avantage de fournir des données plus facilement comparables et plus commodés à étudier, aussi bien pour les analyses de variance que pour les analyses factorielles des correspondances

(1) I.N.R.A., Laboratoire de Biométrie, C.N.R.Z., Domaine de Vilvert 78350 JOUY-EN-JOSAS.

(A.F.C.). En revanche, on perd la signification des données climatiques concernant notamment la vitesse moyenne du vent entre deux relevés (mais avaient-elles plus de sens entre le relevé du vendredi matin et celui du lundi, voire du mardi ?). Il est clair que les données 1974 sont totalement inadaptées à une analyse pour de faibles variations du temps.

b) Un autre difficulté a été celle des distances à la haie; on a piégé des pucerons à 15 distances différentes, mais, pour des raisons matérielles, les pièges n'ont été installés à certaines distances que pendant la deuxième ou quatrième semaine. On a préféré ne pas tenir compte de ces distances (au moins dans une première étape), pour garder le bénéfice d'une prise en compte simplifiée du facteur temps, plutôt que de garder toutes les distances et d'éliminer quatre semaines sur dix, ou de déterminer des estimations pour le moins hasardeuses.

3. Les facteurs de variabilité contrôlée retenus.

Finalement les données se présentent sous la

Tableau (1) Liste des pucerons ailés capturés en bacs jaunes.

(Les noms des espèces prises en compte dans le texte sont soulignés).

1 - <u>Myzus persicae</u>	31 - <u>Myzaphis rosarum</u>
2 - <u>Macrosiphum euphorbiae</u>	32 - <u>Myalopterus pruni</u>
3 - <u>Aulacorthum solani</u>	33 - <u>Myzus cerasi</u>
4 - <u>Brevicoryne brassicae</u>	34 - <u>Macrostelus rubis-nigri</u>
5 - <u>Cavariella aegopodii</u>	35 - <u>Sitobion avenae</u>
6 - <u>Asphorophora rubi</u>	36 - <u>Sitobion frugivorus</u>
7 - <u>Microlophium evansi</u>	37 - <u>Tubaphis rannunculi</u>
8 - <u>Metolophium festucae</u>	38 - <u>Macrostelus fumosus</u>
9 - <u>Metolophium dirhodum</u>	39 - <u>Macrostelapha</u>
10 - <u>Myzocallis castanicola</u>	40 - <u>Rhopalosiphum ascalonicus</u>
11 - <u>Myzocallis coryli</u>	41 - <u>Rhopalosiphoninus</u>
12 - <u>Tuberculoides annulatus</u>	42 - <u>Elatobius abietinus</u>
13 - <u>Myzocallis sp.</u>	43 - <u>Myzus ornatus</u>
14 - <u>Macrostelus viciae</u>	44 - <u>Cavariella theobaldi</u>
15 - <u>Capitophorus homi</u>	45 - <u>Periphyllus</u>
16 - <u>Acyrtosiphon pisum</u>	46 - <u>Pesphigien</u>
17 - <u>Acyrtosiphon malvae</u>	47 - <u>Brachycaudus lychmidis</u>
18 - <u>Phorodon</u>	48 - <u>Myzaphis foeniculi</u>
19 - <u>Rhopalosiphum insertum</u>	49 - <u>Pterolaphis fagi</u>
20 - <u>Rhopalosiphum padi</u>	50 - <u>Lisomaphis berberidis</u>
21 - <u>Thelaxus quercicola</u>	51 - <u>Anoea corni</u>
22 - <u>Acyrtosiphon loti</u>	52 - <u>Lechmidés</u>
23 - <u>Aphis fabae</u> groupe	53 - <u>Neomyzus circumflexus</u>
24 - <u>Brachycaudus helichrysi</u>	54 - <u>Holopaphis</u>
25 - <u>Brachycaudus cardui</u>	55 - <u>Ovatus</u>
26 - <u>Dysaphis</u>	56 - <u>Corylobium arellanae</u>
27 - <u>Dactynotus uroleucon</u>	57 - <u>Myzus cerasi</u>
28 - <u>Dactynotus uroleucon</u>	58 - <u>Myzus lythri</u>
29 - <u>Cryptomyzus galeopsidis</u>	59 - <u>Rhopalosiphoninus lathyrum</u>
30 - <u>Cryptomyzus ribis</u>	60 - <u>Rhopalosiphum sativae</u>

4. Méthodes d'analyse choisies.

On a choisi deux méthodes complémentaires : l'analyse de variance, qui donne des résultats analytiques, et l'analyse factorielle des correspondances, qui donne des résultats synthétiques ; leurs conclusions de-

forme d'un tableau de dénombrements à quatre entrées, ces entrées étant :

l'espèce «E» : 49 ont piégées au moins une fois ; en fait on ne gardera que les plus représentées.

la hauteur «H» de piégeage : deux niveaux, haut et bas, notés H7 et HO (1)

la semaine «S» de piégeage : 10 niveaux, notés S1,..., S10

la distance «D» de piégeage : huit niveaux, notés D1,..., D8 (correspondant à -5H, -1H, -1/2H, 0H, 1/2H, 1H, 3H, 8H) (1)

On dispose ainsi de 160 dénombrements par espèce. Les espèces sont notées Ex où x est le code numérique affecté à l'espèce considérée (tabl. 1 et 2).

Tableau (2) Tableau synthétique des données

Données.	E1	E3	E4	E5	E6	E10	E12	E13	E23	E36	E38	E8	E9	E17	E21	E35	Total sur les espèces	% des niveaux d'un même facteur
H7	150.	167.	531.	2161.	97.	185.	253.	52.	548.	174.	34.	142.	108.	100.	207.	188.	5097	33
HO	793.	401.	450.	5302.	360.	55.	105.	370.	1143.	225.	299.	247.	214.	231.	93.	140.	10428	67
S1	238.	299.	18.	1731.	1.	0.	0.	66.	19.	7.	11.	132.	36.	49.	3.	21.	2621	16,8
S2	135.	111.	32.	2405.	3.	0.	1.	69.	79.	27.	1.	154.	42.	38.	28.	27.	3152	20,1
S3	157.	63.	20.	2369.	34.	1.	0.	94.	114.	40.	29.	76.	35.	85.	118.	30.	3225	20,8
S4	20.	23.	10.	437.	40.	11.	1.	17.	76.	57.	13.	15.	44.	43.	75.	35.	917	6
S5	97.	35.	42.	301.	94.	39.	13.	68.	145.	89.	99.	1.	68.	62.	34.	70.	1257	8,1
S6	239.	35.	475.	195.	252.	36.	47.	131.	614.	112.	170.	10.	69.	52.	28.	100.	2565	16,4
S7	25.	1.	36.	7.	30.	62.	43.	5.	228.	47.	19.	1.	24.	2.	4.	27.	561	3,7
S8	15.	1.	271.	1.	0.	23.	39.	7.	330.	13.	1.	0.	4.	0.	8.	6.	719	4,7
S9	11.	0.	61.	17.	3.	37.	67.	2.	53.	6.	0.	0.	0.	0.	0.	9.	266	1,8
S10	6.	0.	16.	0.	0.	31.	147.	13.	33.	1.	0.	0.	0.	0.	2.	3.	242	1,6
D1	88.	82.	46.	990.	19.	18.	19.	34.	137.	43.	11.	52.	38.	32.	33.	29.	1671	10,8
D2	123.	100.	87.	853.	68.	20.	20.	62.	150.	44.	56.	81.	51.	45.	27.	35.	1822	11,8
D3	161.	88.	352.	1559.	97.	25.	38.	78.	523.	88.	115.	80.	87.	64.	46.	70.	3471	22,3
D4	5.	6.	3.	310.	7.	11.	15.	13.	36.	37.	1.	8.	18.	13.	24.	17.	524	3,3
D5	131.	63.	176.	730.	62.	34.	57.	44.	286.	68.	64.	41.	37.	38.	37.	62.	1930	12,4
D6	162.	119.	150.	757.	129.	24.	39.	66.	215.	54.	51.	31.	38.	51.	31.	41.	1958	12,6
D7	112.	58.	82.	793.	53.	79.	136.	50.	156.	31.	22.	33.	31.	38.	47.	33.	1754	11,3
D8	161.	52.	85.	1471.	22.	29.	34.	75.	188.	34.	13.	63.	22.	50.	55.	41.	2395	15,5
Total par espèce	943	568	981	7463	457	240	358	422	1691	399	333	389	322	331	300	328	15525	
% des espèces	6	3,6	6,3	46	2,9	1,5	2,3	2,7	11	2,4	2,2	2,5	2,1	2,2	1,8	2,2	100	

vraient se recouper ou se compléter efficacement. On étudiera le comportement des quatre entrées vues plus haut les unes par rapport aux autres et on cherchera à dégager, par exemple, s'il y a des espèces ou groupes d'espèces caractéristiques (au sens de «apparaissant plus tôt ou plus tard», «apparaissant près ou loin de la haie», ...) et de même pour les hauteurs, les semaines ou les distances.

(1) la lettre H a été utilisée pour désigner deux facteurs différents :

- la hauteur de piégeage,
- la distance du point de relevé à la haie, H étant dans ce cas la hauteur de la haie boisée.

Afin d'éviter les confusions :

- H = hauteur de piégeage est toujours désignée par cette lettre H seule ou suivie des symboles 7 ou 0 (H7 = 0,70 m au-dessus du sol et HO = niveau du sol).
- H = distance à la haie est toujours précédée d'une valeur chiffrée, par ex. : -5H = point d'observation situé en amont de la haie, à 5 fois sa hauteur ; 1/2H = point d'observation situé en aval à une demie fois la hauteur (voir BAUTRAIS *et al.*, 1976*, fig. 2).

II – DONNÉES 74 : ANALYSE FACTORIELLE DES CORRESPONDANCES.

1. Justification de l'emploi de l'analyse factorielle des correspondances

On dispose des quatre entrées vues précédemment, dont on veut analyser la structure : espèce, hauteur, semaine, distance (E, H, S, D). Avec elles on constituera les tableaux individus x variables caractéristiques de l'analyse factorielle des correspondances.

L'avantage des techniques d'analyse factorielle des correspondances réside, comme on le sait, dans la parfaite symétrie existant entre individus et variables et dans la pondération de ces éléments qui permet d'«équilibrer» l'analyse. En outre, aucune hypothèse restrictive n'est nécessaire et on peut traiter un grand nombre d'observations à la fois.

Cette méthode d'analyse est donc particulièrement adaptée à notre problème puisque la distinction individu-variable y est arbitraire et que par exemple (parmi les espèces choisies) certaines espèces sont beaucoup plus représentées que d'autres et que l'on peut envisager grâce à la propriété d'équivalence distributionnelle de l'analyse factorielle des correspondances un regroupement objectif des espèces ayant des comportements voisins.

2. Ce qu'est l'analyse factorielle des correspondances et ce qu'elle peut dire.

L'analyse factorielle des correspondances consiste à chercher des combinaisons linéaires de variables, ces combinaisons (appelées facteurs) étant à variance maximum et non corrélées entre elles.

Géométriquement, ces facteurs représentent les axes principaux de l'ellipsoïde des observations, situé dans un espace à k dimensions, k étant le nombre de variables. Les graphiques obtenus sont la projection sur deux de ces axes des observations. La distance choisie, qui est fondamentale dans l'analyse factorielle des correspondances, correspond à ce que l'on appelle la «métrique du χ^2 »; c'est elle qui assure à l'analyse factorielle des correspondances un certain nombre d'avantages bien connus (LEBART et FENELON, 1975).

Les facteurs, qui sont des variables synthétiques, sont donc expliqués en plus ou moins grande part par chacune des variables : selon que celles-ci sont respectivement plus ou moins éloignées sur l'axe factoriel (en projection) et plus ou moins proches de cet axe. De même, on expliquera les variables, leurs regroupements, leurs oppositions ..., en fonction des individus voisins

et vice-versa, en permutant individus et variables. On pourra calculer, d'après les coordonnées factorielles, les distances intervariables (ou inter-individus), pour tenter une classification sommaire.

3. Tableaux choisis.

Nous avons pris uniquement les 16 espèces les mieux représentées. Quatre découpages individus x variables ont été choisis, chacun permettant d'étudier mieux l'entrée désignée comme variable. Les quatre découpages sont notés :

(1) H.S.D x E, (2) H.D.E. x S, (3) S.E x H.D., (4) (H+S+D) x E.

Explication des notations :

(1) H.S.D. représente 160 individus qui sont les niveaux de l'interaction hauteur-semaine-distance, (H.S.D.) x E est donc un tableau 160 x 16. Les individus seront notés de 1001 à 1160 dans l'analyse factorielle des correspondances (1), avec le codage suivant, D variant d'abord, puis S, puis H :

H7	S1	D1	1001
•	•	•	•
•	•	•	•
•	•	•	•
•	S1	D8	1008
•	•	•	•
•	•	•	•
•	S10	D1	1073
•	•	•	•
•	•	•	•
H7	S10	D8	1080
•	•	•	•
HO	S1	D1	1081
•	•	•	•
•	•	•	•
•	S1	D8	1088
•	•	•	•
•	•	•	•
•	S10	D1	1133
•	•	•	•
•	•	•	•
HO	S10	D8	1160

Ainsi H7 sera représentée par les points 1001 à 1080

S1 sera représentée par les points 1001 à 1008 et 1081 à 1088 etc. (2) et (3) sont codés de façon

analogue.

(4) correspond à l'analyse factorielle des correspondances (1), où l'on a cumulé les individus sur chaque niveau de la hauteur, la semaine et la distance; cela devrait permettre une interprétation synthétique de toutes les variables. C'est un tableau 20 x 16.

4 — Interprétation rapide des résultats

Dans l'analyse factorielle des correspondances (1), les trois premiers facteurs représentent 69,3 p. 100 de la variabilité totale, ce qui permet d'avoir une excellente visualisation à trois dimensions et facilite l'interprétation. Il en est de même dans les analyses

factorielles des correspondances (2) et (4) : 78,7 p. 100 et 85,7 p. 100 (respectivement) pour les trois premiers facteurs. Un peu moins bien dans l'analyse factorielle des correspondances (3) : 54,9 p. 100 (63,7 p. 100 si l'on rajoute le quatrième facteur). Passons maintenant à l'interprétation rapide facteur par facteur, dans chacune des quatre analyses factorielles des correspondances.

a) Tableau (1) H.S.D. x E. (tabl. 3)

C'est le tableau qui par sa structure se rapproche le plus de l'analyse de variance exposée par la suite; il étudie les caractéristiques spatio-temporelles des différentes espèces.

Tableau (3) Analyse des espèces par l'A.F.C. du tableau H.S.D. x E.

Explication tableau 3 :

- 1) JI sont inscrits en colonne les variables
- 2) QLT : qualité de représentation. C'est la somme (comprise entre 0 et 1) des COR sur la ligne correspondante. C'est une mesure de la proximité de l'observation au sous espace engendré par les 7 facteurs présents.
- 3) POID : poids relatif de l'observation. Pour une variable c'est la somme des différents poids sur les individus, divisée par le poids total.
- 4) INR : inertie du point relativement à l'inertie totale du nuage.
- 5) $n \neq F$: coordonnée de l'observation, sur le $n^{\text{ième}}$ facteur
- 6) COR : contribution relative de la variable au $n^{\text{ième}}$ facteur. C'est le carré du coefficient de corrélation entre le $n^{\text{ième}}$ facteur et l'observation considérée, compris entre 0 et 1. Plus cette quantité est grande, plus l'observation en question explique le facteur.
- 7) CTR : contribution absolue de la variable au $n^{\text{ième}}$ facteur, en pourcentage de la valeur propre de ce facteur.

J	JI	QLT	POID	INR	1 ^{er}	COR	CTR	2 ^{ème}	COR	CTR	3 ^{ème}	COR	CTR	4 ^{ème}	COR	CTR	5 ^{ème}	COR	CTR	6 ^{ème}	COR	CTR	7 ^{ème}	COR	CTR
11	E11	613	.61	191	46	.01	199	77	.01	247	116	.221	352	238	591	199	76	241	177	80	241	148	42	241	
21	E31	888	.37	341	309	163	151	14	.71	30	1	.31	208	45	211	249	39	231	401	102	121	912	52	8481	
31	E41	999	.63	1341	1241	425	1551	683	129	741	1154	367	5111	97	2	41	404	64	1471	206	12	331	21	0	01
41	E51	991	.481	1181	596	812	2621	230	175	641	104	26	321	41	0	01	21	2	51	14	0	11	105	26	981
51	E61	956	.29	501	405	126	171	824	233	501	471	323	1681	267	25	171	587	118	1011	205	14	151	221	17	261
61	E101	986	.15	961	1044	356	931	1385	181	751	173	3	1456	202	2591	393	15	241	1491	209	4211	464	20	601	
71	E121	1000	.23	2301	2715	436	2751	2069	513	5111	254	4	91	688	24	841	101	1	21	335	17	811	129	1	71
81	E131	600	.27	131	42	2	01	330	130	71	422	211	291	355	150	271	249	74	171	74	7	21	147	26	111
91	E231	997	.109	1171	884	423	1361	715	277	1401	2	0	21	193	18	291	700	265	3321	158	14	341	20	0	11
101	E361	609	.26	341	401	104	101	421	76	127	593	157	591	74	221	1001	4	0	01	76	3	21	318	45	461
111	E381	798	.21	391	558	102	111	886	249	421	752	179	741	457	66	351	689	150	1011	113	4	31	307	48	581
121	E81	511	.25	181	615	325	161	269	58	51	196	31	61	49	27	01	0	0	01	102	9	31	328	86	481
131	E91	479	.21	151	143	17	11	277	84	61	517	221	341	316	93	181	89	7	21	37	1	01	176	26	121
141	E171	269	.21	121	177	31	11	138	19	11	444	197	291	142	19	31	57	3	11	138	19	51	34	1	01
151	E211	788	.19	461	103	3	01	4	0	01	349	30	141	1295	411	2531	328	27	211	1125	310	3001	169	7	101
161	E511	578	.21	261	469	106	71	348	59	61	312	67	181	735	260	891	57	2	11	120	7	41	400	77	601
I 15525.0 10001 10001 10001 10001 10001 10001 10001 10001 10001																									

Premier axe (36,5 p. 100 de la variabilité totale)

Il exprime l'effet temps en opposant *C. aegopodii*, *M. festucae* et *A. solani* (E5, E8 et E3) (avec les premières semaines, donc précocité) à *T. annulatus*, *M. castanicola* et *B. brassicae* (E12, E10 et E4) qui sont plus tardives (dernières semaines); la contribution la plus forte est apportée par *C. aegopodii* (E5) (81,2 p. 100) qui n'apparaîtra plus dans les autres facteurs.

Deuxième axe (25,2 p. 100)

Il exprime également le temps, mais dans une différenciation des dernières semaines, qui dans l'ensemble sont plus caractéristiques de l'abondance des espèces choisies. Il oppose *T. annulatus* (E12) (semaine S10) à *A. fabae* groupe, *M. funestum*, *A. rubi* et *B. brassicae* (E23, E38, E6 et E4) (semaines S5, S6, S7 et S8).

Troisième axe (9,6 p. 100)

Il traduit l'effet hauteur par opposition de *A. rubi*, *M. funestum*, *S. fragariae*, *Hyperomyzus sp.*, *M. dirhodum* (E6, E38, E36, E13, E9) (avec H0) à *B. brassicae*, *M. festucae* et *S. avenae* (E4, E8 et E35) (avec H7), la hauteur caractéristique correspondant à un plus fort piègeage.

b) Tableau (2) : H.D.E. x S.

Il met en évidence les moments les plus caractéristiques pour l'ensemble des espèces considérées, ce qui confirme la première analyse :

- le premier axe (44,4 p. 100) oppose les premières semaines aux dernières;

- le deuxième axe (22,7 p. 100) oppose semaines S5 et S6 à semaines S9 et S10.

c) Tableau (3) : S.E. x H.D.

C'est le tableau le plus approprié pour étudier les effets hauteur et distance, l'effet temps n'apparaîtra que faiblement dans le premier axe.

Premier axe (26,2 p. 100)

Il oppose les différentes distances de H7 à celles de HO. *M. castanica*, *T. annulatus*, *T. quercicola* et *S. avenae* (E10, E12, E21 et E35) sont plus abondantes en H7 alors que *M. persicae*, *Hyperomyzus sp.* et *S. fragariae* (E1, E13 et E36) sont en HO.

Deuxième axe (16 p. 100)

Il caractérise la distance en opposant les stations proches de la haie (D3 (- 1/2H), D2(- 1H) et D5 (1/2H)) à celles qui en sont éloignées (D1 (- 5H), D7 (3H) et D8 (8H)). En ce qui concerne les espèces l'opposition a lieu entre *M. funestum* (E38) fortement piégée en

D3 et *M. castanica*, *T. annulatus* (E10 et E12) plus abondantes en D7.

Troisième axe (12,7 p. 100)

Il oppose les distances extrêmes (D1, D8) à celles qui sont à moyenne distance derrière la haie (D5, D6), avec *M. castanica* et *T. annulatus* (E10 et E12) contre *S. fragariae* (E36).

d) Tableau (4) : (H+S+D) x E

Ce tableau permet une prise en compte globale de tous les niveaux qui ne sont plus représentés que par un seul point. En contrepartie, la dispersion des individus correspondant à un même niveau n'est plus prise en ligne de compte; on ne dispose plus que d'un moyen. Il faudra donc être prudent dans l'interprétation des graphiques (en particulier pour les deux hauteurs, qui cumulent chacun 80 points-individus de l'analyse factorielle des correspondances (1)).

Tableau (4) Exemples de sorties de l'analyse de variance en GENSTAT pour les espèces E5 (en haut) et E36 (en bas)

***** ANALYSIS OF VARIANCE *****

VARIATE: Y1

SOURCE OF VARIATION	DF	SS	SS*	MS	VR
UNITS STRATUM					
HAUT.FUR	1	9.9710	1.52	9.9710	53.333
SEMAINE	9	596.1130	89.63	65.1237	393.837
LTN	1	530.5713	81.14	530.5713	3239.642
QUAD	1	0.8547	0.13	0.8547	5.169
CUR	1	29.7488	4.55	29.7488	179.976
DEVIAT IONS	6	24.9382	3.81	4.1564	25.136
DISTANCE	7	15.6311	2.39	2.2330	13.524
HAUTEUR,SEMAINE	9	9.4147	1.44	1.0451	5.376
DEV.LIN	1	5.2904	0.81	5.2904	31.994
DEV.CUR	1	0.0140	0.00	0.0140	0.000
DEV.CUR	1	1.1254	0.17	1.1254	5.806
DEVIAT IONS	6	2.9849	0.46	0.4975	3.009
HAUT.FUR,DISTANCE	7	4.5574	0.70	0.6511	3.937
SEMAINE,DISTANCE	63	17.8109	2.72	0.2827	1.712
LTN,DEV	7	4.9673	0.76	0.7096	4.291
QUAD,DEV	7	1.4980	0.23	0.2140	1.204
CUR,DEV	7	2.5923	0.40	0.3733	2.243
DEVIATIONS	42	8.7533	1.34	0.2084	1.200
RESIDUAL	63	10.4175	1.59	0.1654	
TOTAL	159	653.9153	100.00	4.1127	
GRAND TOTAL	159	653.9153	100.00		
GRAND MEAN		2.241			
TOTAL NUMBER OF OBSERVATIONS		160			

Premier axe (51,7 p. 100)

Oppose une fois de plus *A. solani*, *C. aegopodii*, *M. festucae* (E3, E5, E8) à *M. castanica*, *T. annulatus* (E10, E12) parallèlement à l'opposition entre premières semaines et dernières, hauteur 0 et hauteur 7, distances 1 et 8 à distances 5 et 7. La hiérarchie des effets qu'on avait déjà pu observer se trouve confirmée.

Deuxième axe (24,8 p. 100)

Oppose *M. castanica*, *T. annulatus* (E10, E12) à *M. funestum*, *A. rubi*, *Hyperomyzus sp.* (E38, E6, E13) et en ce qui concerne les observations H7 à HO, D7 et D4 à D3 et D6 (respectivement) semaines S9 et S10 à semaines S5 et S6.

***** ANALYSIS OF VARIANCE *****

VARIATE: Y1

SOURCE OF VARIATION	DF	SS	SS*	MS	VR
UNITS STRATUM					
HAUT.FUR	1	0.3308	0.31	0.3308	1.300
SEMAINE	9	62.1394	57.52	6.9344	27.305
LTN	1	4.8135	4.46	4.8135	19.036
QUAD	1	46.3051	42.86	46.3051	193.125
CUR	1	0.8815	0.82	0.8815	3.406
DEVIAT IONS	6	10.1394	9.39	1.6899	5.683
DISTANCE	7	7.2516	6.71	1.0359	4.097
HAUTEUR,SEMAINE	9	4.5199	4.18	0.5072	1.990
DEV.LIN	1	0.6746	0.62	0.6746	2.660
DEV.CUR	1	0.0965	0.09	0.0965	0.381
DEV.CUR	1	0.5116	0.47	0.5116	2.023
DEVIATIONS	6	3.2373	3.00	0.5395	2.130
HAUT.FUR,DISTANCE	7	2.4648	2.28	0.3521	1.393
SEMAINE,DISTANCE	63	15.3942	14.25	0.2444	0.960
LTN,DEV	7	0.6901	0.64	0.0986	0.390
QUAD,DEV	7	2.4353	2.25	0.3479	1.346
CUR,DEV	7	1.6829	1.58	0.2404	0.951
DEVIATIONS	42	10.5858	9.80	0.2520	0.992
RESIDUAL	63	15.9302	14.75	0.2520	
TOTAL	159	108.0308	100.00	0.6794	
GRAND TOTAL	159	108.0308	100.00		
GRAND MEAN		0.899			
TOTAL NUMBER OF OBSERVATIONS		160			

Troisième axe (9,1 p. 100)

Oppose *A. rubi*, *M. funestum*, *M. dirhodum*, *A. malvae*, *S. fragariae* (E6, E38, E9, E17, E36) à *B. brassicae* (E4) (et *A. fabae* groupe, *M. festucae* (E23-E8)) puis semaines S5, S4, S10 à semaine S8; D6 (et D4) à D3; H7 à HO).

5 — Ébauche de classification des espèces

Les graphiques des analyses factorielles des correspondances font apparaître entre variables des proximités qui traduisent des corrélations de même grandeur avec des facteurs dont nous venons de tenter l'explication, et nous mènent à une ébauche de classification inter-espèces qui sera donc faite principalement à travers le facteur temps, puis celui de la hauteur;

(*A. rubi* - *M. funestum* (E6 - E38)) par semaines S6 et S7, HO:

(*S. avenae* - *S. fragariae* (E35 - E36)) par semaines S5 et S6;

(*A. solani* - *C. aegopodii* - *M. festucae* (E3 - E5 - E8)) par semaines S1 et S2, H₀;

(*M. persicae*, *M.dirhodum* - *Hyperomyzus* sp.
- *A. malvae* - *T. quercicola* (E1 - E9 - E13 - E17 - E21))
par semaines S4 et S5;

A. fabae groupe (E23) se situe entre (*S. avenae* - *S. fragariae* (E35 - E36)), (*A. rubi* - *M. funestum* (E6 - E38)) et *B. brassicae* (E4);

M. castanicola et *T. annulatus* (E10 et E12) sont isolées.

Ces résultats ne sont cependant basés que sur les deux ou trois premiers axes qui expriment 50 à 60 p. 100 de la variabilité totale, aussi avons-nous utilisé une autre méthode de classification afin de juger des distorsions dans l'espace, l'arbre de longueur minimum, qui relie les points voisins par un lien simple. Ces liaisons ont été portées sur le graphe des deux premiers axes de l'analyse factorielle des correspondances du tableau H.S.D. x E (fig. 1).

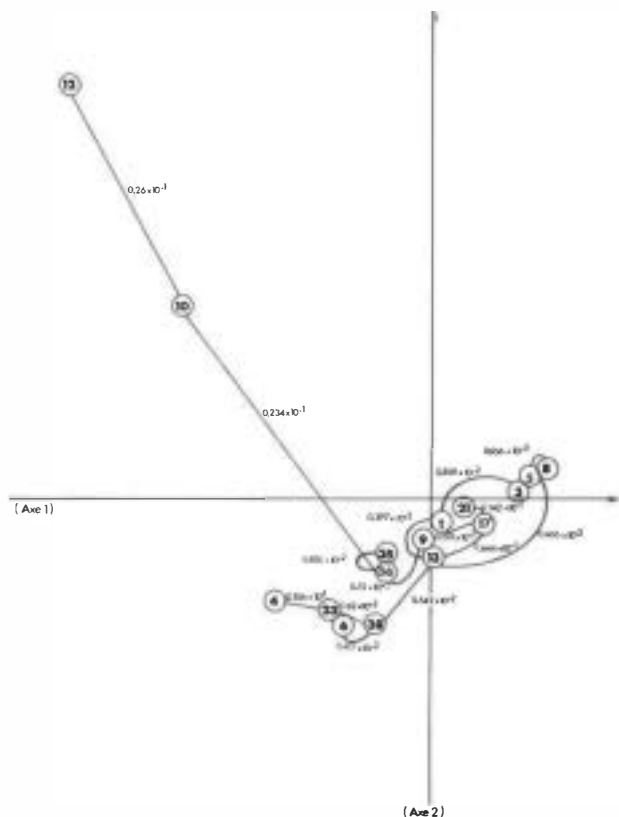


Figure (1) Premier plan factoriel et tracé de l'arbre de longueur minimum entre espèces (étude H.S.D. x E)

Elle confirme les regroupements (*A. rubi* - *M. funestum* (E6 - E38)), (*S. avenae* - *S. fragariae* (E35 - E36)), (*C. aegopodii* - *M. festucae* (E5 - E8)), (*M. persicae* - *M. dirhodum* - *Hyperomyzus* sp. - *A. malvae* - *T. quercicola* (E1 - E9 - E13 - E17 - E21)) qui sont respectivement reliés par lien(s) simple(s); *A. solani* (E3) n'est pas relié directement au groupe (*C. aegopodii* - *M. festucae* (E5 - E8)) mais par l'intermédiaire de *M. persicae* et *Hyperomyzus* sp. (E1 et E13) qui en sont très proches (voir distances portées sur la figure 1), on gardera donc le groupe (*A. solani* - *C. aegopodii* - *M. festucae* (E3 - E5 - E8)); *A. fabae* groupe (E23) est proche de (*M. funestum* - *A. rubi* (E38 - E6)) et de *B. brassicae* (E4), mais pas de (*S. avenae* - *S. fragariae* (E35 - E36)).

6. Analyse factorielle des correspondances avec espèces éclatées.

Les premières analyses factorielles des correspondances ont fait apparaître des regroupements entre espèces qui s'expliquent par un effet semaine ou hauteur, etc., c'est-à-dire entre espèces vivant ensemble aux mêmes endroits. Il n'apparaît pas au contraire de liaison entre espèces ayant des comportements différents et c'est ce qu'on a tenté de faire ressortir d'une analyse factorielle des correspondances H.S.D. $\times$ E où l'on a «découpé» chaque variable E en quatre variables, grâce au codage suivant :

- 1) «О»

- 2) «peu de pucerons» (55 p. 100 des dénombrements)

- ### 3) «moyennement» (1/3)

- 4) «beaucoup» (1/9)

Les résultats permettent de mieux apprécier certaines relations; les regroupements apparaissant dans les premières analyses factorielles des correspondances se retrouvent ici sur les classes 3 et 4 des espèces concernées. Mais des rapprochements nouveaux apparaissent : en particulier les espèces *T. annulatus* et *M. castanicola* (E12 et E10) qui étaient isolées dans les premières analyses ressortent plus nettement :

–*T. annulatus* (E12) moyenne avec absence de
C. aegopodii (E5);

– peu de *T. annulatus* (E12) avec beaucoup de *Hyperomyzus* sp. (E13);

— beaucoup de *M. castanicola* (E10) avec peu de *M. persicae* et *M. funestum* (E1 et E38) ; absence de *A. solani*, *S. fragariae*, *A. rubi* (E3, E36, E6) ;

— peu de *S. fragariae* (E36) avec beaucoup de *Hyperomyzus* sp. (E13), toutes deux étant corrélées avec semaines S4, S1 et S2 ce qui explique l'opposition entre ces deux espèces;

— absence de *M. persicae* (E1) avec peu de *A. solani* et *C. aegopodii* (E3 et E5), expliqués par semaines S6 et S5.

Le codage choisi apparaît ainsi comme un moyen de révéler des liaisons que les dénombrements initiaux utilisés tels quels ne permettaient pas d'apprécier; c'est un moyen d'apprécier des relations non linéaires entre variables initiales.

III — DONNÉES 74 : L' ANALYSE DE VARIANCE (ANOVA).

Soit N la variable aléatoire de dénombrement des pucerons d'une espèce et $\{N_{ijk}\}$ ($i = 1, 2; j = 1 \dots 10, k = 1 \dots 8$) un échantillon de cette variable, où N_{ijk} est le dénombrement :

à la i ème hauteur ($i = 1, 2$)

au j ème temps ($j = 1, 10$), c'est-à-dire semaine j

à la k ème distance à la haie ($k = 1, \dots, 8$)

$$N_{ijk} = \mu_{ijk} + \varepsilon_{ijk}$$

où $\mu_{ijk} = E(N_{ijk})$: valeur moyenne de N_{ijk}

ε_{ijk} résiduelle

avec l'hypothèse que les ε_{ijk} sont indépendants de même variance σ^2 .

Pour les espèces choisies, les estimateurs de σ^2 notés $\hat{\sigma}^2$ varient de 893,3 à 2,3; les variances sont donc très hétérogènes et approximativement proportionnelles au nombre de pucerons puisque le rapport σ^2/n varie dans un intervalle de longueur 0,11

$$(0,007 < \frac{\sigma^2}{n} < 0,119) \text{ où } n = \sum_{i,j,k} N_{ijk}$$

La transformation pour homogénéiser les variances les mieux adaptées à ces données est celle qui consiste à prendre leur logarithme (SCHEFFE, 1959); on effectuera donc l'ANOVA sur $y_{ijk} = \text{Log } N_{ijk}$

Dans ce cas $\hat{\sigma}^2$ est compris entre 0,1264 et 0,2724, et peut être considérée constante.

Un autre problème se pose, celui de la normalité de ε sur laquelle est fondée la validité des tests des effets. A chaque triplet (i, j, k) des indices, correspond une seule observation; l'absence de répétition, rend impossible le test de normalité des ε_{ijk} .

Considérons $\{\varepsilon_{ijk}\}_{ijk}$ comme un échantillon de la variable ε dont on étudie la normalité; les histogrammes des résidus $(N_{ijk} - \hat{\mu}_{ijk})_{ijk}$ montrent une distribution non normale, avec plusieurs maxima et minima locaux et parfois asymétrie (fig. 2). Les résidus des données transformées y_{ijk} présentent également des extrema locaux prononcés, une étude plus poussée, qui n'a pas pu être réalisée ici s'avère donc nécessaire pour mieux connaître la validité des tests effectués.

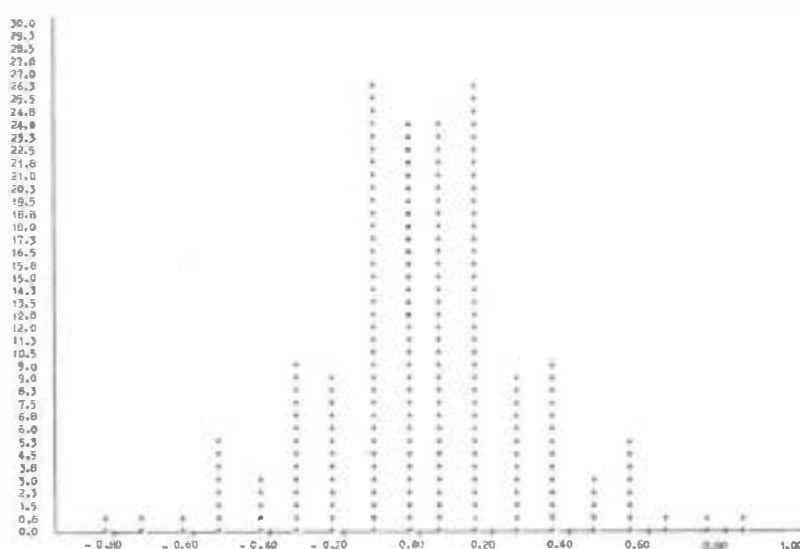


Figure (2) Histogramme des résidus de l'analyse de variance portant sur les données brutes de l'espèce E3.

Néanmoins et «faute de mieux» l'ANOVA, sur les données y_{ijk} , montre que le temps représente l'effet le plus important pour presque toutes les espèces (sauf *M. persicae*, *M. castanicola*, *Hyperomyzus sp.* et *M. funestum* (E1, E10, E13, et E38) allant jusqu'à 89,6 p. 100 de la somme des carrés totale et en général plus de 40 p. 100. Ceci vient du fait que les données concernent une longue période pendant laquelle le comportement des pucerons varie beaucoup puisqu'ils ne sont présents que pendant cinq ou six semaines dont deux ou trois de façon importante.

Il a été fait une analyse avec un sous-modèle de régression polynomiale sur le temps; le modèle étudié décompose la valeur moyenne de y_{ijk} en la somme

— d'une fonction constante, μ ;

— de fonctions de i (resp. j et k), α_i (resp. β_j et γ_k), qui représentent l'effet du i ème (resp. j ème et k ème) niveau de la hauteur (resp. de la semaine et de la distance);

— de fonctions de i et j (resp. i et k , j et k) qui sont les termes d'interaction entre les niveaux correspondants de la hauteur et de la semaine (resp. de la hauteur et de la distance; de la semaine et de la distance) :

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijk}$$

$$\text{avec } \beta_j = f(s_j) + d_j \quad \text{ou } f(s) = a_1x_1 + a_2x_2 + a_3x_3$$

X_i polynômes orthogonaux de degré i et

s_j = j ème niveau de la semaine

d_j : déviation de β_j à la régression polynomiale.

Ce sous-modèle permet de mieux maîtriser la variabilité du facteur temps et l'évolution des dénombrements. Pour les espèces *M. persicae*, *A. solani*, *C. aegopodii*, *M. festucae*, *M. dirhodum*, *M. castanicola*, *T. annulatus*, *Hyperomyzus sp.* *A. malvae* (E1, E3, E5, E8, E9, E10, E12, E13, E17), l'effet linéaire est de beaucoup le plus important alors que pour les espèces *A. rubi*, *A. fabae* groupe, *B. cardui*, *S. fragariae*, *M. funestum* (E6, E23, E25, E36, E38), c'est l'effet quadratique; ce qui traduit une tendance croissante ou décroissante pour les premières, un passage par un extrema pour les autres.

L'effet hauteur a le rapport de variance le plus important pour les espèces *M. persicae*, *M. castanicola*, *Hyperomyzus sp.* et *M. funestum* (E1, E10, E13 et E38), c'est-à-dire que pour ces espèces, la répartition des pucerons est surtout caractérisée par une différence entre les hauteurs qui se traduit par une présence plus précoce ou plus tardive aux différentes distances ainsi que plus abondante, les moyennes sur les hauteurs allant jusqu'à 0,357, en H7 et 1,1219 en H0 pour *Hyperomyzus sp.* (E13); ce sont les seules espèces pour lesquelles le temps n'est pas l'effet dominant.

Si on excepte les espèces *S. fragariae* (E36), *B. brassicae* (E4) et *M. festucae* (E8) pour lesquelles l'hypothèse $\alpha_i = 0$ est acceptée, c'est-à-dire qu'il n'y a pas d'effet hauteur de piégeage, pour les autres espèces, la hauteur de piégeage est plus importante que la distance à la haie. Mais on peut se demander si la croissance des plants de pomme de terre ne perturbe pas la répartition verticale des pucerons et ne donne pas trop de poids aux différences de hauteurs.

Les interactions d'ordre 2 ne sont jamais «très» significatives. Remarquons toutefois que pour *Hyperomyzus sp.* (E13) qui a un effet hauteur très important, l'interaction semaine-hauteur est supérieure à l'effet distance.

L'effet distance à la haie est peu significatif, mais fait apparaître pour l'ensemble des espèces un minimum en +H0, un maximum devant la haie (−0,5H) et un autre généralement moins important en 8H pour *C. aegopodii*, *Hyperomyzus sp.*, *A. fabae* groupe, *M. festucae* (E5, E13, E23, E8);

3H pour *M. castanicola*, *T. annulatus*, *T. quer-cicola* (E10, E12, E21);

1H pour *M. persicae*, *S. fragariae*, *M. funestum*, *A. solani*, *A. rubi*, *M. dirhodum*, *T. ranunculina* (E1, E36, E38, E3, E6, E9, E17);

0,5H pour *B. brassicae*, *S. avenae* (E4, E35).

L'étude détaillée des résultats d'analyse de variance, qu'il nous est impossible de transcrire ici, permet de mettre en évidence certains rapprochements; ainsi *A. rubi*, et *M. funestum* (E6 et E38) qui montrent une grande similitude pour tous les effets. D'autres espèces ayant même comportement par rapport au temps se regroupent de la façon suivante :

(*A. solani*, *C. aegopodii*, *M. festucae* (E3, E5,

E8)), (*M. persicae*, *M. dirhodum*, *Hyperomyzus* sp., *A. malvae* (E1, E9, E13, E17)), (*S. avenae*, *S. fragariae* (E35, E36)).

Pour ces deux premiers groupes, on remarque des évolutions en fonction de la hauteur et de la distance à la haie très voisines, il semble donc qu'il n'y ait pas de phénomènes de compétition entre ces espèces. Au contraire les espèces *S. avenae* et *S. fragariae* (E35 et E36) qui présentent même évolution par rapport au temps et aux distances, se situent à des hauteurs différentes, tandis que *M. castanicola* et *T. annulatus* (E10 et E12) ont un comportement spatial très proche mais n'apparaissent pas en même temps (tabl. 5). Les résultats des tests F (au niveau 5 p. 100).

Tableau (5) Classements par ordre décroissant de l'importance des effets pour chaque espèce.

Espèces	1	2	3	4	5	6	8	9	10	12	13	17	21	23	25	36	38
H	1	2	3	2	2	4	2	1	2	1	2	2	2	2	2	5	1
S	2	1	1	1	1	1	1	2	1	2	1	1	1	1	1	1	2
SD	3	3	2	3	3	2	3	3	3	3	4	3	4	3	3	2	3
SH	5	5	4	4	4	4	4	4	4	3	4	5	4	5	4	3	4
DS	4	4	5	5	6	6	5	6	6	5	5	3	6	4	4	5	5
DS	6	6	6	6	5	5	6	5	5	6	6	6	5	6	6	6	6
H	7	2	2	1	2	2	2	2	1	2	2	1	2	1	2	2	2
S	1	2	1	8	3	8	2	6	10	10	3	8	10	7	8	6	6
SD	2	4	2	5	1	6	1	4	10	7	2	6	4	6	6	6	6
SH	3	3	3	7	2	4	3	7	8	10	5	1	1	5	4	5	3
DS	4	7	6	10	4	3	4	5	7	7	6	5	2	7	3	3	5
DS	5	5	4	4	5	2	6	3	3	6	2	3	4	2	2	2	2
DS	6	1	5	1	6	1	5	2	1	2	1	4	5	1	1	1	1
DS	7	6	7	6	7	5	6	8	2	5	8	7	7	3	5	4	4
DS	8	9	7	2	10	10	10	1	6	4	7	10	6	2	8	7	6
DS	9	8	10	3	7	6	10	10	5	3	10	10	10	8	8	9	10
DS	10	10	10	9	10	10	10	10	4	1	9	10	9	9	10	10	10
D	1	8	4	7	4	7	5	4	6	7	7	8	6	5	5	7	7
S	2	3	3	4	3	3	2	2	7	6	1	5	6	7	6	4	2
SD	3	1	2	1	1	1	1	1	4	3	2	1	1	1	1	1	1
SH	4	7	8	8	8	8	8	8	8	8	1	8	7	8	8	7	6
DS	5	6	5	2	7	4	4	5	2	2	5	6	3	3	2	3	4
DS	6	2	1	3	6	2	6	3	5	4	4	2	5	2	3	2	3
DS	7	5	6	5	5	5	7	6	1	1	6	4	2	5	7	8	5
DS	8	4	7	6	2	6	3	7	3	5	3	3	4	4	4	6	8

— l'interaction hauteur-distance est nulle pour les espèces *A. rubi*, *M. festucae*, *M. castanicola*, *T. annulatus*, *A. malvae*, *S. fragariae* (E6, E8, E10, E12, E17, E36).

— l'interaction semaine-hauteur, est nulle pour *A. malvae*, *T. quercicola*, *S. avenae*, *S. fragariae* (E17, E21, E35, E36) et linéaire pour *M. persicae*, *A. solani*, *C. aegopodii*, *A. rubi*, *M. festucae* (E1, E3, E5, E6, E8).

— l'interaction ($\alpha\beta$) hauteur-semaine est nulle pour *T. quercicola*, *S. avenae*, *S. fragariae* (E21, E35, E36), linéaire pour *T. annulatus*, *M. persicae*, *A. solani* (E12, E1, E3).

L'effet hauteur est nul pour *M. festucae*, *S. fragariae* (E8, E36)

L'effet distance est nul pour *T. quercicola* (E21).

Ces résultats portant sur $\{v_{ijk}\} = \text{Log}(N_{ijk})$ cela signifie que le dénombrement s'exprime par un produit de fonction de la hauteur, semaine, distance.

Pour E36 où α , ($\alpha\beta$), (α) et (β) sont significativement nuls à 5 p. 100 on pourrait écrire

$$v_{ijk} = \mu + \beta_j + \gamma_k + \varepsilon_{ijk}$$

Soit $N_{ijk} = e^{\varepsilon_{ijk}} (e^{\mu} e^{\beta_j} e^{\gamma_k}) = e^{\varepsilon_{ijk}} (m \cdot b_j \cdot e_k) = m \cdot b_j \cdot c_k + e_{ijk}$ lorsque le résidu est assez petit.

Exemples d'analyses : *S. fragariae* (E36). *C. aegopodii* (E5), pour laquelle on rejette toutes les hypothèses de nullité d'effets (tabl. 4).

On pourrait poursuivre l'analyse en faisant des tests de comparaisons multiples quand les hypothèses précédentes sont rejetées et, par exemple, regrouper les semaines significativement semblables en les considérant comme des répétitions, mais il faut pour cela que les effets SD, SH et S soient identiques. Ceci suggère la recherche d'une transformation des données qui les exprimerait par un modèle additif (c'est-à-dire linéaire en fonction des effets H, S et D). Parmi les transformations classiques l'une semble particulièrement intéressante, elle consiste à prendre N^p où p est déterminé de façon à ce que les N_{ijk} soient à la fois de loi normale, additifs et de variances égales (BOX et COX, 1964).

On utilisera les regroupements qui apparaissent dans l'Analyse Factorielle des Correspondances et qui traduisent des corrélations importantes entre espèces, pour essayer de poursuivre l'étude quantitative en considérant les espèces proches comme des répétitions du dénombrement.

L'intérêt de cette méthode réside dans le fait qu'on a une précision d'autant plus grande sur les estimations que le nombre de répétitions est important et biologiquement : elle permettrait d'étudier la répartition de plusieurs espèces à partir de l'une d'elles.

Pour tester si des différences sont significatives entre les espèces d'un même groupe on a fait des ANOVA à quatre facteurs (hauteur, semaines, distances, espèce) sur les données définies par:

$$V_{ijkl} = \frac{\text{Log } N_{ijkl}}{\sum_{ijk} \text{Log } N_{ijkl}}$$

Les indices i, j, k , étant définis comme précédemment et l notant les «niveaux» des espèces.

Cette transformation des données élimine la différence entre espèces due à des effectifs différents sur ces espèces, de façon à étudier un comportement individuel moyen des pucerons.

Il apparaît alors des interactions significativement non nulles dans lesquelles le facteur espèce entre en jeu, on ne peut donc accepter de le considérer comme une simple répétition.

CONCLUSION.

On voit ici se confronter deux points de vue, l'un descriptif dans l'étude des classifications par l'A.F.C. et l'arbre de longueur minimum, et l'autre quantitatif, qui rejette les précédents résultats en ce sens qu'il n'accepte pas l'hypothèse d'un même comportement pour des espèces d'abord jugées voisines. Mais plutôt que de s'exclure l'un l'autre, ils vont de pair,

le premier suggérant des hypothèses que le second approche et ce, pour permettre des prédictions dans des situations analogues.

Par ailleurs des résultats se recoupent, aussi bien par l'analyse de variance que pour l'analyse factorielle des correspondances on constate une hiérarchisation des effets : temps, puis hauteur et enfin distance; les interactions d'ordre 2 (semaine-hauteur, semaine-distance et hauteur-distance) sont peu importantes.

L'effet distance, le moins apparent, devrait être mieux mis en évidence (ainsi que l'effet hauteur) sur les données 75 où les relevés sont beaucoup plus rapprochés dans le temps et effectués durant une période plus courte. Le facteur temps sera ainsi moins prépondérant. En outre, les relevés ayant lieu plusieurs fois par jour, on pourra suivre l'évolution au cours d'une journée et aborder les phénomènes de concurrence; enfin, les données météorologiques seront intégrées.

Le facteur temps, qui jusqu'ici était un facteur contrôlé comme la hauteur et la distance, sera distingué : on considérera les données comme des fonctions du temps dont on étudiera les paramètres à l'aide des autres facteurs.

