Étude de l'influence de la topographie et des caractéristiques des vergers sur l'efficacité du programme d'épandage d'insectes stériles pour le carpocapse de la pomme (*Laspeyresia pomonelle*)

par

Gaétan Daigle, Thierry Duchesne, Emmanuelle Reny-Nolin et Louis-Paul Rivest
Département de mathématiques et de statistique
Université Laval, Ste-Foy
Québec
Canada G1K 7P4
Corresponding author's email: lpr@mat.ulaval.ca

1 Traitement préliminaire des données

Les caractéristiques des trappes dans les vergers sont les variables explicatives les plus importantes pour modéliser l'abondance des carpocapses et l'étendue des dommages. Cette section décrit les transformations appliquées à ces variables pour faciliter leur interprétation.

1. La densité du verger. Pour caractériser la densité des arbres dans le verger nous avons utilisé deux variables associées à la disposition des arbres dans le verger, Within Row Spacing, WRSP, Between Row Spacing, BRSP, et deux variables associées à la taille des arbres eux-mêmes, Tree height, TRHE et Basal Diameter BADIA. Pour les combiner en une seule nous avons effectué une analyse en composantes principales de la matrice de variances-covariances sur l'échelle logarithmique. Les corrélations étaient toutes positives ce qui indique la présence d'un effet taille. En effet, les grands arbres ont un diamètre élevé et sont plantés dans des vergers où les arbres sont relativement espacés. La première composante principale explique 86% de la variabilité. L'exponentielle de cette première composante, TR_AR=$\text{WRSP}^{0.59}\text{BRSP}^{0.15}\text{TRHE}^{0.17}\text{BADIA}^{0.60}$ a été utilisée dans les analyses qui suivent. Une grande valeur de TR_AR correspond à une faible densité d'arbres autour d'une trappe.

2. La latitude et la longitude. Les coordonnées X et Y donnent la longitude et la latitude des trappes. Dans ce système d'axes la rive du lac Okanagan peut être représentée par une droite de pente -1.37. La direction de cette droite est un vecteur unitaire de $\arctan(-1.37) = -54\,\text{deg}$. Une rotation

de -36 degrés est requise pour faire coïncider la rive du lac avec l'axe y. Le changement d'axe s'effectue sur les données centrées, X_c et Y_c. Les nouvelles coordonnées X_N et Y_N sont alors données par

$$\begin{pmatrix} X_N \\ Y_N \end{pmatrix} = \frac{1}{1000} \begin{pmatrix} \cos(-36) & -\sin(-36) \\ \sin(-36) & \cos(-36) \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \end{pmatrix}.$$

La division par 1000 fait en sorte que X_N et Y_N sont en kilomètres. Ils varient dans l'intervalle (-4.4,6). La variable X_N est associée à la distance au lac; $X_N > 2.5$ donne un point dans le lac. La variable Y_N détermine le déplacement le long de la rive du lac. Les nouvelles coordonnées sont à peu près non corrélées, avec $r(X_N, Y_N) = 0.03$, où r représente la corrélation de Pearson. Avec les données originales $r(X, Y) = -0.47$. Il faut noter que la rive du lac Okanagan est pentue et que l'altitude GPSH est corrélée avec X_N et Y_N, $r(X_N, \text{GPSH}) = -0.75$ et $r(Y_N, \text{GPSH}) = 0.41$. L'interprétation de cette première corrélation est claire: lorsque X_N augmente, on s'approche de la rive du lac Okanagan et l'altitude diminue.

3. La variable ASPECT. ASPECT donne la direction, en degrés, de la pente du terrain où se trouve une trappe. Pour une pente orientée franc nord ASPECT=0 deg, ou 360 deg, alors que pour une orientation franc est ASPECT=90 deg. La raideur de la pente est évaluée par la variable SLOPE qui donne essentiellement la tangente de l'angle de l'inclinaison de la pente par rapport à l'horizontale. Ainsi si SLOPE=0, ASPECT n'est pas défini. L'importance de la variable ASPECT est donc liée à SLOPE. Ainsi dans les analyses les variables SL_SIN_ASP = SLOPE$\times$ sin(ASPECT) et SL_COS_ASP = SLOPE$\times$cos(ASPECT) sont utilisées. Ces variables prennent la valeur 0 si SLOPE=0.

Nous avons jusqu'ici traité de 7 variables explicatives, TR_AR, X_N, Y_N, GPS_HGT, SLOPE, SL_SIN_ASP, SL_COS_ASP. Les autres variables décrivant les trappes sont STYLE, une variable dichotomique, GRD_CULTIV et IRRIGATION, deux variables qualitatives décrivant le mode de culture du verger, et trois variables associées à la distance de l'agglomération urbaine la plus proche, URB200, URB100 et DISTANCE. Dans les analyses, des variables centrées sont utilisées et Z dénote le vecteur de ces variables pour une trappe.

2 Modélisation temporelle et spatiale de l'abondance des carpocapses

Pour décrire l'arrivée des carpocapses dans les trappes on utilise un processus de Poisson non stationnaire, avec le temps t dans l'intervalle (0,1). Le temps $t = 0$ correspond à une semaine avant le début de la période de collecte et $t = 1$ est le temps de la dernière visite. Le tableau 1 donne les dates correspondant à ces deux évènements pour les trois années de l'étude. Si Z dénote le vecteur

des variables géographiques et physiques mesurées pour une trappe, le modèle utilisé postule que l'intensité du processus de Poisson est donnée par

$$\lambda(t|Z,\alpha,\psi) = \lambda_0(t)\exp(Z'\beta + \alpha + \psi) \quad t \in (0,1), \tag{1}$$

où β un vecteur de paramètres inconnus associés aux variables explicatives, α représente un effet de trappe résiduel et ψ est associé au verger.

La notation suivante est utilisée:

- l'indice i, $i = 1,\ldots,I$ représente le verger, en 2001 le nombre total de vergers est de $I = 236$; $j = 1,\ldots,N_i$ est associé aux trappes. Le nombre de trappes N_i par verger varie entre 1 et 14. Le total des trappes utilisées est $\sum N_i$. Le vecteur des variables explicatives centrées de la trappe j du verger i, ou plus simplement de la trappe (i,j), est appelé Z_{ij}.

- $k = 1,\ldots,N_{ij}$ représente les visites effectuées à la trappe (i,j). Nous appelons $t_{ijk} \in (0,1)$ le temps de la $k$ième visite à la trappe (i,j); de plus par définition $t_{ij0} = 0$ ce qui correspond à la date d'apparition des carpocapses. En 2001, le 22 avril est le temps 0.

- $m_{ijk} = (t_{ijk} + t_{ij(k-1)})/2$ est le point milieu de l'intervalle de temps entre la $(k-1)$ième et la $k$ième visite à la trappe (i,j);

- n_{ijk} représente, pour $i = 1,\ldots,I$, $j = 1,\ldots,N_i$, et $k = 1,\ldots,N_{ij}$ le décompte des carpocapses fait au temps t_{ijk}; ce nombre représente l'effectif des insectes capturés entre les temps $t_{ij(k-1)}$ et t_{ijk}.

- r_{ijk} est une variable indicatrice prenant la valeur 1 si, lors de la $k$ième visite à la trappe (i,j), la règle du 40 stériles pour 1 sauvage n'est pas respectée.

Nous sommes ainsi en présence d'un processus de dénombrement dans chaque trappe. Le processus augmente de un lorsqu'un carpocapse est capturé. Les temps exacts d'arrivée des insectes sont inconnus. En fait ces temps sont censurés par intervalle car à chaque insecte capturé est associé un intervalle de temps $(t_{ij(k-1)}, t_{ijk})$. L'arrivée de carpocapses sauvages étant un évènement relativement rare, plusieurs techniques d'analyse peuvent être envisagées.

Les modèles à risques proportionnels. Les modèles présentés dans le chapitre 8 de Therneaux et Grambsch(2000) auraient pu être utilisés en imputant les temps exacts d'arrivée par les points milieux m_{ijk} des intervalles $(t_{ij(k-1)}, t_{ijk})$. Nous nous sommes intéressés à modéliser le temps de la première capture d'un carpocapse sauvage. La variable dépendante pour une trappe est donc donnée par l'intervalle $(t_{ij(k-1)}, t_{ijk})$ de première capture. Cette variable est censurée à droite par le temps de la dernière visite lorsqu'aucun insecte

sauvage n'est capturé. La forme (1) du modèle correspond alors au modèle avec risques proportionnels de Cox. Plusieurs approches ont été présentées dans la littérature pour estimer les paramètres de ce modèle à partir de données censurées par intervalle, voir par exemple Alioum et Commenges (1996). Pour simplifier les analyses nous avons cependant imputé les temps d'arrivée par les milieux m_{ijk} des intervalles de première observation. Les résultats pour le modèle de Cox présentés au tableau 2 ont été obtenus à l'aide de la procédure PHREG de SAS.

Les modèles de régression paramétrique pour des durées de vie. Les modèles de régression paramétrique sont des alternatives au modèle de Cox qui permettent d'incorporer explicitement la censure par intervalle. En utilisant, dans (1), l'expression paramétrique $\lambda_0(t) = ct^\alpha$, on peut utiliser la régression Weibull, avec données censurées par intervalle, pour estimer les paramètres. Pour transformer les paramètres estimés $\hat{\gamma}$ de la régression Weibull en paramètres loglinéaires, on utilise la transformation $\hat{\beta} = -\hat{\alpha}\hat{\gamma}$ (voir Klein et Moeschberger, 1997, p. 378). L'erreur-type de $\hat{\beta}$ est ensuite calculée par linéarisation. La procédure LIFEREG de SAS a été utilisée pour ajuster une loi Weibull.

La régression Poisson. Cette méthode s'applique tant aux insectes stériles qu'aux insectes sauvages. La plupart des analyses sont basées sur cette technique. Pour modéliser l'effet temporel, une série de Fourier d'ordre m a été utilisée,

$$\log \lambda_0(t) = a_0 + \sum_{j=1}^{m} \{a_j \sin(2\pi jt) + b_j \cos(2\pi jt)\}. \tag{2}$$

Des séries jusqu'à l'ordre 5 ont été ajustées. L'ordre 4 a offert un bon compromis biais-variance pour la modélisation des carpocapses tant sauvages que stériles.
En vertu du modèle postulé, la distribution du nombre de carpocapses observés dans la trappe (i, j) au temps t_{ijk}, n_{ijk}, est une Poisson de paramètre

$$\int_{t_{ij(k-1)}}^{t_{ijk}} \lambda(t|Z_{ij}, \alpha_{ij}, \psi_i)dt = \exp(Z'_{ij}\beta + \alpha_{ij} + \psi_i) \int_{t_{ij(k-1)}}^{t_{ijk}} \lambda_0(t)dt.$$

L'intégrale de $\lambda_0(t)$ est une expression complexe qui dépend des paramètres a et b de la série de Fourier. Lors de l'ajustement des modèles statistiques, nous approximons cette intégrale par la méthode du rectangle,

$$\int_{t_{ij(k-1)}}^{t_{ijk}} \lambda_0(t)dt \approx (t_{ijk} - t_{ij(k-1)})\lambda_0\left(m_{ijk}\right).$$

Étant donné que la majorité des visites sont hebdomadaires ou bimensuelles, les différences $t_{ijk} - t_{ij(k-1)}$ sont en grande majorité inférieures à 0.1 et ces approximations sont précises.

Les paramètres d'intérêt sont les coefficients a et b des séries de Fourier et le vecteur β. La caractérisation des composantes intra-trappe et intra-verger des erreurs n'est pas un objectif de l'étude. Ainsi, lors des analyses nous avons utilisé la régression Poisson standard, sans effet aléatoire pour les trappes et les vergers. La régression Poisson s'exprime sous la forme du modèle log-linéaire suivant,

$$\log E(n_{ijk}) = \log(t_{ijk} - t_{ij(k-1)}) + a_0 + \sum_{\ell=1}^{m} \{a_\ell \sin(2\pi\ell m_{ijk}) + b_\ell \cos(2\pi\ell m_{ijk})\} + Z'_{ij}\beta,$$

où $\log(t_{ijk} - t_{ij(k-1)})$ est une variable offset. Les variances ont été estimées à l'aide d'estimateurs sandwich qui mesurent la variabilité inter-verger. L'estimateur de variances-covariances, $\hat{\Sigma}$, de $(\hat{a}, \hat{b}, \hat{\beta})$ s'écrit en fonction de Z^*_{ijk}, le vecteur de toute les variables explicatives, y compris celles associées à la modélisation du temps, et de $\hat{\mu}_{ijk}$, la valeur prédite pour n_{ijk} de la façon suivante

$$\hat{\Sigma} = \{\sum_{ijk} Z^*_{ijk} Z^{*\prime}_{ijk} \hat{\mu}_{ijk}\}^{-1} [\sum_i \{\sum_{jk} Z^*_{ijk}(n_{ijk} - \hat{\mu}_{ijk})\}\{\sum_{jk} Z^*_{ijk}(n_{ijk} - \hat{\mu}_{ijk})\}'] \{\sum_{ijk} Z^*_{ijk} Z^{*\prime}_{ijk} \hat{\mu}_{ijk}\}^{-1}$$

(voir Cameron et Trivedi, 1998, p.320). Pour étudier la présence d'une interaction entre le temps et les variables associées aux trappes, des interactions $\sin(2\pi m_{ijk}) \times Z_{ij}$ et $\cos(2\pi m_{ijk}) \times Z_{ij}$ ont été incluses dans le modèle. Puisque $\beta_c Z_{ij} \cos(2\pi m_{ijk}) + \beta_s Z_{ij} \sin(2\pi m_{ijk})$ peut s'écrire $r_{max} Z_{ij} \cos\{2\pi(m_{ijk} - t_{max})\}$ où r_{max} est la valeur maximale de l'interaction et t_{max}, le temps où cette interaction est maximale, les interaction temporelles sont rapportées comme des couples (r_{max}, t_{max}) au Tableau 3. La majorité des calculs ont été faits à l'aide de la procédure GENMOD de SAS.

3 Détermination des facteurs associés à l'abondance des carpocapses

L'année cible pour les analyses est 2001. La sélection de modèle a été faite à partir des données de 2001 avec la régression Poisson pour l'abondance des carpocapses, sauvages et stériles, et avec la régression logistique pour la conformité à la règle du 40 pour 1. Toutes les variables explicatives disponibles ont été considérées dans la procédure de choix des modèles. Le modèle sélectionné, pour chaque type d'analyse, a ensuite été ajusté aux données de 1999 et 2000. De plus, les paramètres de 2001 pour l'abondance des carpocapses sauvages ont été réestimés à partir des temps de première capture en utilisant la régression Weibull et le modèle de Cox.

Année	$\sum N_i$	Sauvages	Stériles	$t = 0$	$t = 1$	nb jours	% visite
1999	506	1069	282144	01/05	25/9	148	90%
2000	500	290	151740	24/04	18/9	148	77%
2001	454	263	113401	22/04	2/9	134	65%

Tableau 1 : Quelques statistiques descriptives concernant la capture des insectes aux trois années de l'étude. On y retrouve le nombre total de trappes ($\sum N_i$), le nombre total d'insectes sauvages et stériles capturés, les dates correspondant à $t = 0$ et à $t = 1$, le nombre maximal de jours durant lesquels une trappe a été en opération, et la proportion moyenne, par rapport au nombre maximal, des jours d'utilisation d'une trappe (% visite). Les dates sont données dans le format jj/mm.

Année	1999	2000	2001		
Modèle	Poisson	Poisson	Poisson	Weibull	Cox
Variable	$\hat{\beta}$ $(e.t.)$	$\hat{\beta}$ $(e.t.)$	$\hat{\beta}$ $(e.t.)$	$\hat{\beta}$ $(e.t.)$	$\hat{\beta}$ $(e.t)$
X_N	0.17 (0.10)	0.17 (0.15)	0.58 (0.20)	0.37 (0.14)	0.41 (0.14)
X_N^2	-0.28 (0.09)	-0.18 (0.07)	-0.21 (0.08)	-0.24 (0.10)	-0.23 (0.10)
TR_AR	0.36 (0.11)	0.66 (0.17)	0.60 (0.24)	0.17 (0.16)	0.17 (0.16)

Tableau 2 : Estimations des paramètres du modèle final pour la capture des carpocapses sauvages et leurs erreurs-types.

3.1 Carpocapses sauvages

L'analyse des données de 2001 a débuté par une sélection pas à pas parmi les 13 variables explicatives disponibles. Puisque PROC GENMOD de la version 8 de SAS n'a pas de procédure automatique pour faire une telle sélection, nous avons utilisé PROC LOGISTIC en prenant comme variable dépendante une variable dichotomique indiquant la présence d'au moins un insecte sauvage. En fixant les seuils d'entrée et de sortie de la procédure de sélection à .30 et .10 respectivement cette procédure permet d'identifier toutes les variables potentiellement associées à la présence des insectes.
La régression Poisson, avec estimateur de variance Sandwich, a ensuite été utilisée pour faire la sélection définitive des variables explicatives à un seuil de 5%. L'ajustement du modèle a été vérifié en incluant des termes quadratiques, des interactions entre variables explicatives et avec le temps. Cette étape a permis l'ajout de X_N^2. Finalement des trappes extrêmes, ayant de grands résidus pour la régression Poisson, ont été identifiées et retirées de l'analyse. Ceci a permis de vérifier la stabilité des estimations présentées au tableau 2.
L'interprétation des paramètres pour 2001 révèle que pour des trappes près du lac (avec $X_N \approx 1.5$) l'intensité du processus de Poisson augmente de 50% par rapport aux trappes avec $X_N = 0$. Elle diminue de plus de 50% loin du lac

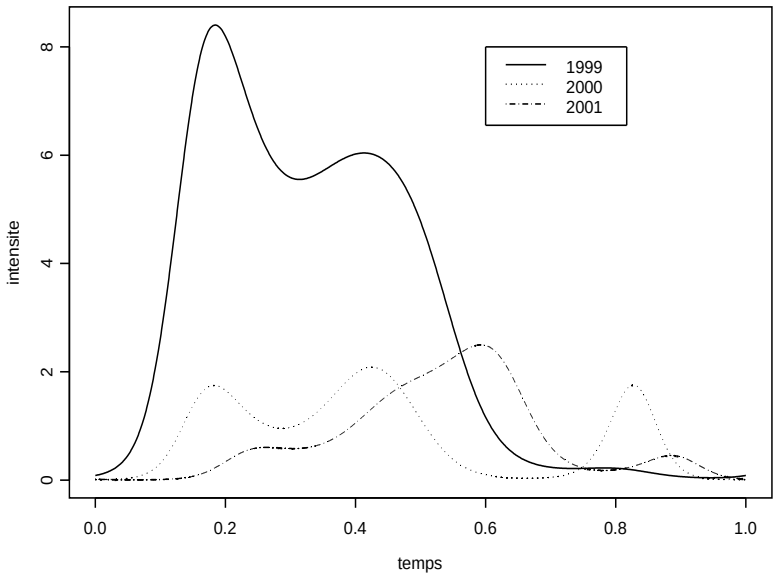

Figure 1: Estimation de l'intensité $\lambda_0(t)$ d'arrivée des carpocapses sauvages pour trois années

(avec $X_N < -1$). De plus une augmentation d'un écart-type ($\sigma = 0.77$) de la variable TR_AR augmente $\lambda(t)$ de 58 %. Les insectes sauvages sont donc plus abondants dans les vergers moins denses.
La régression Weibull et le modèle de Cox ont été utilisés, tel que décrit à la section 2 pour valider les estimations faites en 2001. La différence la plus importante est le paramètre pour TR_AR qui est beaucoup plus faible pour les analyses de temps de première apparition des carpocapses sauvages. Une analyse, non présentée ici, sur les données de 1999 explique ce phénomène. En 1999, une interaction significative TR_AR$\times t$, montre que l'importance de la variable TR_AR dépend du temps. Pour t voisin de 0 son paramètre est essentiellement nul; il atteint son maximum, 0.65, à $t = 0.4$. Une faible interaction TR_AR$\times t$ est donc une explication possible pour les écarts entre les analyses basées sur les données complètes et celles qui utilisent seulement les temps de première apparition.
Les paramètres estimés pour 1999 et 2000, présentés au tableau 2, sont relativement semblables à ceux de 2001. L'estimation pour X_N est cependant plus petite en 1999 et en 2000 qu'en 2001. Ainsi, en 2001, la proximité au lac est associée à une augmentation des carpocapses sauvages pour la première fois.

La figure 1 compare les estimations de $\lambda_0(t)$, l'intensité temporelle de l'abondance

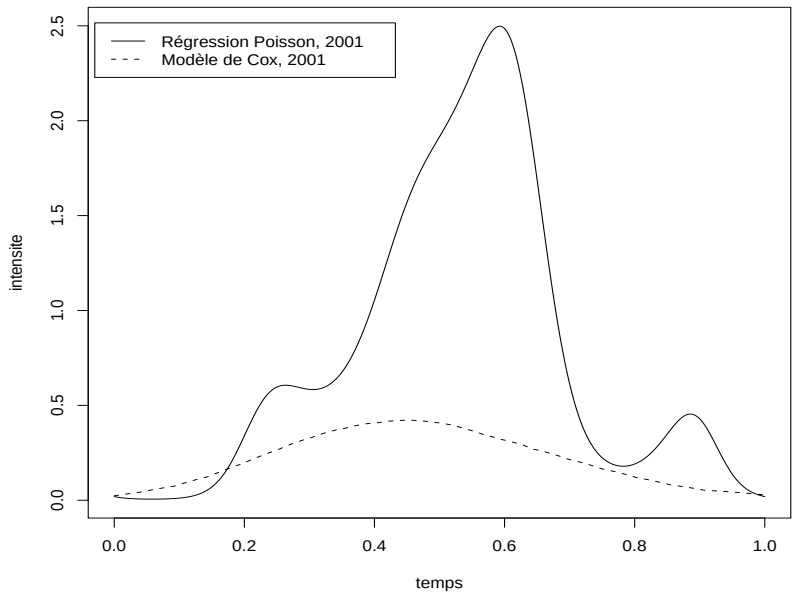

Figure 2: Estimation de l'intensité $\lambda_0(t)$ en obtenue à l'aide de la régression Poisson et du modèle de Cox

des carpocapses sauvages, pour les trois années de l'étude. Même si les périodes d'observation pour les trois années ne sont pas rigoureusement identiques il est intéressant de mettre les courbes sur le même graphique. Les carpocapses sauvages étaient plus abondants en 1999. En 1999 et 2000 on voit deux maxima locaux à $t = 0.2$ et $t = 0.42$ qui correspondent au 19 mai et au 17 juin environ. En 2001 l'arrivée des carpocapses sauvages est plus tardive avec un maximum à $t = 0.61$, ce qui correspond au 12 juillet. La figure 2 compare l'estimation de l'intensité $\lambda_0(t)$ obtenu avec les données complètes et celle qui utilise seulement les temps de première apparition. Cette dernière utilise la méthode de Ramlau-Hansen (voir Klein et Moschberger, 1977, section 6.2) avec le noyau de Epanechnikov et une fenêtre de $b = 0.3$ pour lisser l'estimation du risque. Les deux estimations ont la même forme cependant le maximum aux environs de $t = 0.6$ est beaucoup plus accentué pour la régression Poisson que pour le modèle de Cox. La présence, dans (1) d'effets aléatoires α et ψ pour la variabilité inter-trappes et inter-vergers est sans doute responsable de ces différences. La fonction de risque estimée par la régression Poisson est en effet égale à $\lambda_0(t)\exp(Z'\beta)E\{\exp(\alpha+\psi)\}$, alors que le risque cumulé pour le modèle de Cox s'écrit

$$-\log E[\exp\{-\int_0^t \lambda_0(s)ds\exp(Z'\beta+\alpha+\psi)\}]$$

$$\leq \int_0^t \lambda_0(s)ds \exp(Z'\beta)E\{\exp(\alpha+\psi)\}.$$

L'inégalité précédente est une conséquence de Jensen: si X est une variable alétoire et a est une constante, $E(e^{-aX}) \geq e^{-aE(X)}$. Il y a donc une explication théorique aux écarts observés à la Figure 2: en présence d'effets aléatoires inter verger et inter trappe, l'ajustement d'un modèle avec risques proportionnels sous-estime l'intensité $\lambda_0(t)$.

3.2 Carpocapses stériles

Une régression Poisson telle que présentée à la section 2 a été utilisée pour les analyses. Vu l'abondance de données, seules les composantes du modèle qui sont significatives au seuil 1% ont été retenues. Le STYLE de trappe est le facteur le plus important; l'emploi des trappes WING, qui représentent environ 90% des trappes disponibles, multiplie l'intensité de capture par 4. En fait la performance des trappes DELTA dépend fortement de l'altitude GPS_HGT. Une faible altitude est associée à des probabilités de capture plus importantes. La deuxième variable en importance pour 2001 est X_N. La relation entre la distance au lac est, en moyenne, l'inverse de celle pour les sauvages. Il y a plus de stériles loin du lac (lorsque $X_N < 0$); à $X_N = -1.5$, l'intensité augmente de 50%. Les deux interactions significatives $X_N \times t$ et $X_N \times$GPS_HGT soulignent que cette relation varie dans le temps et avec l'altitude. Cette association négative entre X_N et l'intensité des stériles est unique à 2001. Ainsi en 2000, la proximité de la rive du lac est liée à une augmentation des stériles.
La figure 3 présente les intensités des processus de Poisson pour les stériles de 1999 à 2001. Il y a avait plus d'insectes stériles en 1999; les courbes pour 2000 et 2001 sont très semblables.

3.3 Respect de la règle du 40 stériles pour 1 sauvage

La variable dépendante pour cette analyse est r_{ijk} qui vaut 1 si la règle n'est pas respectée lors de la visite de la trappe i au temps t_{ijk} et 0 sinon, pour $i = 1, \ldots, I$, $j = 1, \ldots, N_i$ et $k = 1 \ldots, N_{ij}$. La probabilité de succès de cette variable Bernoulli dépend à la fois du temps d'observation et des caractéristiques des trappes. Cela suggère de considérer le modèle suivant

$$\text{logit}E(r_{ijk}) = \lambda_0(t_{ijk}) + Z'_{ij}\beta.$$

où $\lambda_0(t)$ représente une fonction lisse du temps écrite à l'aide d'une série de Fourier comme dans les sections précédentes. Cette formulation du modèle n'inclut pas un effet de trappe résiduel. Pour éviter qu'un tel effet ne biaise les estimations d'erreurs-types à la baisse, ces derniers sont estimés à l'aide d'un estimateur de variance de type sandwich associé à la variabilité

	1999	2000	2001
Var.	$\hat{\beta}$	$\hat{\beta}$	$\hat{\beta}$
X_N	-0.03	0.09*	-0.13*
X_N^2	0.01	0.08*	0.12*
GPS_HGT	-0.0004	0.002*	-0.002*
STYLE DELTA	-1.32*	-1.42*	-1.40*
STYLE WING	0 (0)	0 (0)	0 (0)
$X_N \times$GPS_HGT	0.0008*	0.001*	0.0013*
GPS_HGT$\times$STYLE DELTA	-0.002	-0.004*	-0.008*
Inter.	(r_{max}, t_{max})	(r_{max}, t_{max})	(r_{max}, t_{max})
$X_N \times t$	(-0.03,0.13)	(.05,0.42)	(0.12,0.33)*
GPS_HGT$\times t$	(0.002,0.15)*	(.003,0.48)*	(0.002,0.08)*

Tableau 3 : Estimations des paramètres du modèle final pour la capture des carpocapses stériles. Les interactions avec le temps sont rapportées comme une paire (r_{max}, t_{max}) où r_{max} donne la valeur maximale de l'interaction et t_{max} donne le temps où cette interaction est maximale. L'interaction est minimale au temps $t_{max} + 1/2$; elle est alors donnée par $-r_{max}$. Un paramètre suivi d'un astéristique signifie que ce dernier est significativement non nul au seuil 1%.

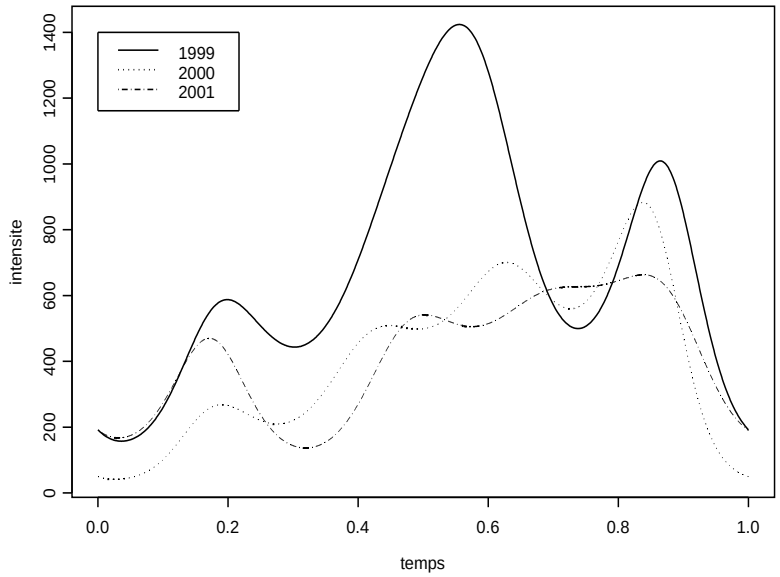

Figure 3: Intensité estimée $\hat{\lambda}_0(t)$ pour les stériles

	1999	2000	2001
Var.	$\hat{\beta}$ $(e.t.)$	$\hat{\beta}$ $(e.t.)$	$\hat{\beta}$ $(e.t.)$
X_N	0.22 (0.07)	0.19 (0.12)	0.47 (0.14)
X_N^2	-0.22 (0.05)	-0.19 (0.06)	-0.24 (0.07)
TR_AR	0.34 (0.10)	0.57 (0.15)	0.54 (0.17)

Tableau 4 : Estimations des paramètres du modèle pour le respect de la règle du 40 pour 1.

inter-verger. Les variables explicatives importantes sont les mêmes que celles pour l'abondance des stériles présentée au tableau 2 et les estimations des paramètres des tableaux 2 et 4 sont très voisines.
Pour étudier les fluctuations temporelles de la probabilité de violer la règle du 40 pour 1, on utilise la probabilité de non-conformité évaluée pour des variables explicatives nulles, $\exp \hat{\lambda}_0(t)/(1 + \exp \hat{\lambda}_0(t))$. La figure 4 présente le graphique de ces probabilités estimées pour les trois années. L'allure générale des courbes est similaire à celles de la figure 1. L'année 2001 se caractérise par un maximum relativement tardif, à $t = 0.62$ ce qui correspond au milieu de juillet. Les probabilités de la figure 3 étant relativement faibles, les paramètres associés aux variables explicatives peuvent s'interpréter sur le mode multiplicatif comme à la section 3.1. Augmenter d'un écart-type la variable TR_AR ou se rapprocher du lac en passant de $X_N = 0$ à $X_N = 1.5$ augmente la probabilité de violer la règle du 40 pour 1 de 50%.

4 Association entre les mesures de dommages des pommes et les caractéristiques du verger

En 1999, 2000, et 2001 des dommages ont été constatés dans respectivement 10, 5 et 28 vergers. Pour modéliser l'ampleur des dommages en fonction des caractéristiques des vergers nous utilisons un modèle Tobit. Ce modèle est un modèle normal censuré que l'on peut définir à l'aide d'une variable latente $Y_i* = \mu + Z_i'\beta + \epsilon_i$ où β est un vecteur de paramètres inconnus et ϵ_i est une variable aléatoire de distribution $N(0, \sigma^2)$. On observe $Y_i = Y_i*$ si $Y_i* > 0$ et $Y_i = 0$ sinon. Les données égales à 0 sont donc considérées comme étant censurées à gauche. On a modélisé, pour les trois années de l'étude, les dommages en fonction des 2 variables explicatives retenues aux sections 3.1 et 3.3, TR_AR et X_N. Les résultats apparaissent au tableau 5.
En 2000 et 2001, la variable TR_AR est la seule variable significative dans les analyses des dommages. L'estimation du paramètre est cependant beaucoup plus élevée en 2001. De plus en 2001, le seuil observé pour la distance au lac X_N, est de 10%. Cette variable semble donc être associée aux dommages.

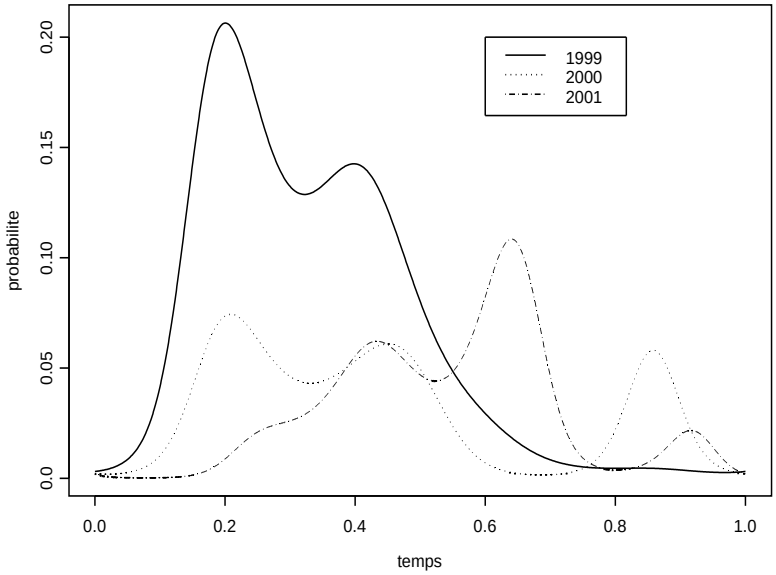

Figure 4: Probabilité estimée de non-conformité à la règle du 40 stériles pour 1 sauvage

	1999	2000	2001
Var.	$\hat{\beta}$ $(e.t.)$	$\hat{\beta}$ $(e.t.)$	$\hat{\beta}$ $(e.t.)$
μ	-0.0074 (0.0013)	-0.0077 (0.0022)	-0.0218 (0.0041)
X_N	-0.0008 (0.0005)	0.0011 (0.0007)	0.0022 (0.0014)
TR_AR	0.0007 (0.0009)	0.0025 (0.0010)	0.0086 (0.0028)
σ^2	0.0062 (0.0009)	0.0046 (0.0012)	0.016 (0.0024)

Tableau 5 : Régression tobit pour les dommages.

5 Conclusion

En 2000 et 2001 les captures de carpocapses sauvages et stériles sont semblables. Cependant, il y a eu beaucoup plus de dommages dans les vergers en 2001. Peut-on expliquer ce phénomène? Les analyses présentées ici permettent de formuler certaines hypothèses. En 2001, l'arrivée tardive des carpocapses sauvages fait en sorte que la règle du 40 stériles pour 1 sauvage n'était pas respectée en juillet dans un nombre important de vergers. En juillet les pommes ont atteints une grosseur appréciable et elles sont peut-être plus sensibles aux carpocapses. Une deuxième caractéristique de 2001 est une diminution des insectes stériles près du lac. Cette diminution est associée à une augmentation des insectes sauvages et des dommages aux vergers près du lac. Soulignons également qu'en 2000 et 2001 les vergers moins denses, correspondant à de grandes valeurs de TR_AR, sont favorables à la présence de carpocapses sauvages et sont ainsi associés à des dommages plus importants. Notre dernière hypothèse concerne le monstre Ogopogo un résident du lac Okanagan qui aurait, paraît-il, favoriser la prolifération des carpocapses sauvages près du lac en 2001. Peut-être faudrait-il prendre des moyens musclés pour éloigner Ogopogo de la rive du lac Okanagan dans les prochaines années.

RÉFÉRENCES

A. Alioum et Commenges, D. (1996). A proportional hazard model for arbitrarely censored and truncated data. *Biometrics*, **52**, 512-524.

A. C. Cameron et Trivedi, P.K.(1998)*Regression Analysis of Count Data*. New York: Cambridge University Press.

J. P. Klein et Moeschberger, M. L. (1997). *Survival Analysis. Techniques for Censored and Truncated Data.*. New York: Springer Verlag.

T. M. Therneau et Grambsch, P. M. (2000). *Modeling Survival Data. Extending the Cox Model*. New York: Springer Verlag.