

## **MIEUX VALORISER LES RÉSEAUX D'ÉPIDÉMIOLOGIE<sup>1</sup> LORS DE L'ÉLABORATION DU BULLETIN DE SANTÉ DU VÉGÉTAL**

par Lucie MICHEL

Chantal de Fouquet<sup>2</sup>. – Le plan Ecophyto prévoit la surveillance sanitaire des cultures afin de raisonner les applications de produits phytosanitaires. Mais les nombreuses données des réseaux d'observation sont encore très mal valorisées. La thèse présentée par Lucie MICHEL a pour objet :

(1) de tester différentes méthodes statistiques pour mieux caractériser, à l'aide de ces données, l'état sanitaire de plusieurs cultures et en prévoir l'évolution ;

(2) de proposer une méthodologie d'échantillonnage afin de réduire les coûts d'observations sans trop pénaliser la précision des estimations ;

(3) grâce à l'automatisation des calculs, de mettre en œuvre les outils statistiques retenus pour évaluer l'incidence des maladies et proposer des prévisions, ces résultats étant destinés à être publiés dans le Bulletin de Santé du Végétal (BSV).

Rendre la méthodologie opérationnelle, en impliquant étroitement les nombreux acteurs (y compris les agriculteurs destinataires du BSV) fait partie des objectifs de la thèse.

Le mémoire commence par un rappel synthétique sur les pesticides et l'évolution réglementaire, notamment les actions promues par le Plan Ecophyto. Le Bulletin de Santé du Végétal et les acteurs qui interviennent sont présentés. Le contexte de la thèse étant posé, le plan du mémoire est donné.

Cette introduction est bien rédigée.

Le premier chapitre pose les trois problématiques de la thèse, et présente les quatre cas d'étude :

- le cas d'étude principal est la septoriose du blé en Champagne-Ardenne. Les parcelles d'observation sont constituées de variétés différentes. De mars à juin, les observations sont effectuées sur les trois dernières feuilles déployées de chacune des 20 plantes prélevées aléatoirement dans la parcelle. Pour chaque feuille, la note d'incidence (sur 20) indique le nombre d'individus positifs.

Au printemps, un BSV hebdomadaire est publié pour les céréales.

- les trois autres cas sont le mildiou de la vigne en Midi-Pyrénées, et en Guadeloupe, la cercosporiose jaune du bananier et le charançon de la patate douce.

Le chapitre deux est un article en anglais, publié dans l'European Journal of Plant Pathology en 2016. Il présente les "modèles linéaires généralisés" (à effets fixes) et les "modèles mixtes

---

<sup>1</sup> Thèse soutenue le 30 mai 2016 pour obtenir le grade de docteur délivré par l'Institut des Sciences et Industries du Vivant et de l'Environnement (AgroParisTech), spécialité Sciences agronomiques.

<sup>2</sup> Membre correspondant de l'Académie d'Agriculture de France. Directeur de recherche à l'École des Mines de Paris, centre de géosciences-géostatiques.

linéaires généralisés" (pour introduire des effets aléatoires) retenus pour les ajustements. Deux classes de méthodes sont comparées, les statistiques fréquentistes (les paramètres étant ajustés par maximum de vraisemblance) et les statistiques bayésiennes.

L'étude porte sur la septoriose du blé en Champagne-Ardenne, les observations des trois feuilles étant traitées séparément. Les effets pris en compte sont le niveau de risque de la parcelle (trois modalités selon la date de semis et la variété) et l'effet (aléatoire) "site-année", soit quatre cas en modèle fréquentiste et bayésien respectivement (aucun effet, ou l'un des deux seulement, ou les deux).

Trois critères sont utilisés pour évaluer la qualité de l'ajustement : le critère d'information d'Akaike, la racine de l'erreur quadratique moyenne (RMSE) ou de l'erreur de prédiction (RMSEP) ainsi que les critères AIC (cas fréquentiste) ou DIC (cas bayésien).

Le nombre de feuilles infectées est modélisé par une variable binomiale, dont les probabilités s'expriment (à une transformation logistique près) par une combinaison linéaire de la date, et si ces effets sont pris en compte, des indicatrices pour le risque (effet fixe), et de coefficients aléatoires pour l'effet site-année.

A mêmes effets pris en compte, les résultats apparaissent analogues par modélisation fréquentiste ou bayésienne, cette dernière fournissant une évaluation de l'incertitude mais nécessitant des temps de calcul beaucoup plus élevés.

Le modèle incluant les deux effets (risque et site-année) est utilisé pour prévoir l'incidence de la septoriose, une à quatre semaines après les dernières observations. La qualité de ces prévisions est évaluée par validation croisée (les dernières observations ne sont pas prises en compte, ce qui permet de comparer les prévisions à la "réalité" de l'échantillonnage). L'auteur discute et justifie les différences de qualité des ajustements et de précision des prévisions, très variables entre les trois feuilles.

La discussion des résultats est bien menée et convaincante (faible apport de la prise en compte du risque lié à la parcelle, par exemple).

Le lecteur non familier avec les modèles linéaires généralisés regrettera cependant l'absence d'explications de certaines grandeurs utilisées (la fonction logit, le critère d'information d'Akaike, les critères AIC et DIC...). Ces éléments pouvaient être reportés en annexe, pour faciliter la lecture du mémoire.

L'application de ces méthodes a nécessité leur programmation (codes fournis en annexe).

Une étude de sensibilité aux hypothèses (pourquoi pas la loi uniforme comme loi de probabilité non-informative par exemple) aurait été intéressante.

Le troisième chapitre (soumis pour publication) applique le modèle linéaire mixte généralisé aux quatre cas d'étude, ainsi qu'à celui de la septoriose du blé en région Centre. Les deux critères de qualité des ajustements sont le critère d'information d'Akaike et la racine de l'erreur quadratique moyenne (RMSE).

Pour la septoriose du blé, l'incidence est caractérisée par la médiane. Dans le cas de la région Centre, les prévisions à une ou deux semaines sont satisfaisantes.

L'ajustement en modèle bayésien s'avère problématique dans le cas du mildiou, pour lequel une variante du modèle de référence est utilisée.

Pour la cercosporiose jaune du bananier, l'estimation de la probabilité de dépassement d'un seuil fixé semble surestimer les symptômes, par comparaison aux comptages.

Enfin, dans le cas du charançon de la patate douce, le modèle tend à sous-estimer le nombre d'insectes en début de saison. Appliquée à un site particulier, la modélisation bayésienne indique une incertitude élevée.

Une discussion plus détaillée des résultats et plus systématique de la concordance (ou non) des deux critères (résultats présentés dans un tableau) aurait été intéressante, mais trouve difficilement place dans un article.

Au chapitre quatre, une méthodologie de réduction du nombre de parcelles observées est proposée à l'échelle régionale. Le principe consiste à réduire l'échantillonnage à la deuxième date d'observations, et à examiner les conséquences de cette réduction sur les prévisions à huit dates ultérieures. Le nombre d'observations est réduit suivant un critère de stratification (trois ou sept strates) effectuée à partir des résultats du modèle mixte linéaire généralisé ; deux autres critères de réduction sont comparés : la stratification (trois strates) selon le risque associé à la parcelle, et un sous-échantillonnage aléatoire. Selon le critère d'erreur quadratique moyenne, la stratification en sept strates à partir du modèle linéaire généralisé (recalé aux observations supposées disponible) s'avère préférable. Ce critère de stratification permet (y compris avec seulement trois strates) une réduction sensible du nombre d'observations, sans trop de perte de précision.

Le dernier chapitre présente la mise en œuvre opérationnelle de l'outil, en coopération avec de nombreux acteurs. Dans le cas de la septoriose du blé en Champagne-Ardenne, il est proposé d'utiliser les données pour représenter l'évolution de l'incidence à partir des observations, et à l'échelle régionale, pour estimer chaque semaine l'incidence de la maladie par groupe de risque, ainsi que son évolution les semaines suivantes. Les cartes d'observations et les synthèses de l'incidence, avec le report des résultats des années précédentes, ont effectivement été publiées dans le BSV de Champagne-Ardenne, en respectant de strictes contraintes de temps. A la fin de la rédaction, l'automatisation du "rapportage" était en cours dans la région Centre, ainsi que pour le mildiou de la vigne en Midi-Pyrénées.

Ce chapitre laisse transparaître un certain enthousiasme : "pari gagné" pour les premières applications effectives de la méthode, avec à terme une meilleure prévision de l'évolution des maladies ou des ravageurs, permettant de réduire l'utilisation des produits phytosanitaires.

La conclusion reprend la synthèse des principaux résultats et de leurs applications.

La rédaction du mémoire est claire et synthétique. Les chapitres en français comportent toutefois d'assez nombreuses "coquilles" et maladresses de style.

On regrettera également l'absence de résumé en français des deux articles en anglais. Le premier article, publié dans l'European Journal of Plant Pathology présente en effet le cadre méthodologique (fréquentiste ou bayésien), le choix du modèle (modèle linéaire mixte généralisé), le paramétrage, les critères d'ajustement, ainsi que l'examen détaillé des résultats dans le cas de la septoriose du blé et la comparaison pour les trois feuilles, qui sont des points techniques essentiels de la démarche.

Malgré ces quelques maladresses, ce travail présente un grand intérêt, car il permet de valoriser les nombreuses données de surveillance acquises ou dont l'acquisition est prévue. Les résultats des modélisations statistiques sont discutés sans que leur utilisation en agronomie ne soit jamais perdue de vue. Les applications sont pragmatiques. Cette thèse présente un bon exemple d'application aboutie de méthodes statistiques assez récentes.