

**AFPP – ONZIÈME CONFÉRENCE INTERNATIONALE SUR LES MALADIES DES PLANTES
TOURS – 7 AU 9 DÉCEMBRE 2015**

**ELABORATION D'UN MODELE D'ALERTE DE MALADIE FONGIQUE SUR GRANDES CULTURES :
APPLICATION AU PATHOSYSTEME *PUCCINIA STRIIFORMIS*/BLE TENDRE D'HIVER**

A. PIERRE⁽¹⁾, E. GOURDAIN⁽¹⁾, G. COULEAUD⁽¹⁾, X. LE BRIS⁽²⁾, F. MOREAU⁽²⁾, O. DEUDON⁽¹⁾,
E. HERITIER⁽¹⁾, D. SIMONNEAU⁽¹⁾, F. PIRAUX⁽¹⁾

⁽¹⁾ ARVALIS-institut du végétal, 91720 Boigneville, France

⁽²⁾ ARVALIS-institut du végétal, 44370 La Chapelle Saint Sauveur, France
amandine.pierre.03@gmail.com ; e.gourdain@arvalisinstitutduvegetal.fr

RÉSUMÉ

La rouille jaune du blé tendre (*Puccinia striiformis*) est une maladie fongique à dispersion aérienne qui a été parfois la cause d'une nuisibilité de plus de 60 q/ha comme observé en 2014. Un modèle statistique, de type modèle de survie, à l'échelle de la parcelle a été élaboré à partir d'un jeu de données nationales comprenant sept années de notation de la maladie sur le terrain. Ce modèle agro-climatique considère un effet du climat de l'année précédent la campagne, ainsi que du climat en cours de campagne, sur les délais d'apparition des premiers symptômes de la maladie. La notion de résistance variétale de la plante hôte est elle aussi intégrée au modèle. L'utilisation de cet outil permet de fournir aux utilisateurs un calendrier de visite des cultures, en vue de prévenir les épidémies. Ceci aidera les agriculteurs à objectiver leur prise de décision d'éventuels traitements, ce qui s'inscrit dans le cadre du projet EcoPhyto.

Mots-clés : modèle de survie, rouille jaune, épidémiologie, inoculum primaire, outil d'aide à la décision.

ABSTRACT

DEVELOPMENT OF A FUNGAL DISEASE WARNING MODEL ON FIELD CROPS: APPLICATION TO THE PATHOSYSTEM *PUCCINIA STRIIFORMIS* / WINTER WHEAT

Yellow rust (*Puccinia striiformis*) is an airborne fungal disease of wheat, which has sometimes caused yield loss of more than 60 q / ha in year 2014. A statistical survival model was developed from a national database of seven years of field data. This agro-climatic model considers an effect of the climate of the year preceding the campaign, therefore the primary inoculum, and of the climate of the current campaign on the time of the symptom emergence. Varietal resistance of the host plant is also included in the model. This work aims at providing users with a scheduled survey of crops, in order to prevent epidemics. Using this tool will help farmers in decision-making for potential sprayings, which are part of EcoPhyto programm.

Keywords: survival model, yellow rust, epidemiology, primary inoculum, decision-making tool.

INTRODUCTION

La rouille jaune *Puccinia striiformis* est une maladie foliaire du blé, qui en est l'unique hôte connu à ce jour. Cette maladie est présente principalement dans les régions nord et côtières où elle a pu engendrer une nuisibilité de plus de 60 q/ha en 2014 (essais variétés et fongicides Arvalis 2014). Au printemps, pour des températures entre 7°C et 10°C (Coakley *et al.*, 1981, 1984 ; Chen *et al.*, 2012) et en cas de pluie, le champignon produit des urédospores qui se disséminent via le vent et contaminent de nouvelles plantes. Les symptômes caractéristiques apparaissent au bout d'une semaine sous la forme de petites pustules jaunes alignées le long des nervures de la feuille, appelées urédies ; les plantes touchées sont regroupées en foyers dans la parcelle. Les filaments de mycélium pénètrent à l'intérieur de la feuille et détruisent les tissus verts de l'hôte, ce qui engendre des perturbations dans le cycle de croissance de celui-ci. Deux semaines après l'infection, chaque urédie contient des milliers d'urédospores prêtes à être disséminées à nouveau (Chen, 2005). En plus d'avoir un cycle court et explosif de reproduction, les souches de rouille évoluent très rapidement. Ceci induit des modifications importantes de la virulence de la maladie suivant les années et nécessite une vigilance permanente des agriculteurs. En effet, chaque semaine ils doivent surveiller l'apparition des symptômes et juger s'il est temps d'effectuer un traitement phytosanitaire ou non. L'élaboration d'un modèle de surveillance des parcelles permet alors de réduire et optimiser le nombre de sorties au cours de la campagne.

Lorsque les conditions climatiques ne sont pas favorables à la germination (De Vallavieille-Pope *et al.*, 2002 ; Garrett *et al.*, 2000 ; Mboup *et al.*, 2012), le champignon se maintient sous forme de mycélium au sein des tissus vivants jusqu'à des températures pouvant atteindre -10°C, en-deça desquelles l'hôte et le parasite sont détruits (Ma *et al.*, 2015). Les repousses peuvent ainsi être contaminées et servir de refuge tout au long de l'interculture. Il n'existe pas de traitement préventif contre la rouille jaune, ainsi, ces travaux de modélisation ont pour finalité d'estimer un délai d'apparition des symptômes à l'échelle de la parcelle.

Les modèles de survie sont des modèles statistiques qui apparaissent dès le XVII^e siècle et sont issus de questionnement démographique (Lorino, 2002). Ils permettent de calculer le temps restant avant la mort pour des organismes biologiques (l'espérance de vie) ou, plus tard avec l'industrialisation, le temps restant avant l'échec ou la panne dans les systèmes artificiels (analyse de fiabilité). Les modèles de Cox, nommés d'après leur auteur, David Cox, voient le jour en 1972, et sont des modèles statistiques semi-paramétriques. Chaque individu est caractérisé par une courbe de survie qui dépend du temps et de une ou plusieurs variables explicatives qualitatives ou quantitatives. Il est également possible d'intégrer des processus aléatoires dans ces modèles. Les modèles de Cox sont à ce jour les modèles les plus utilisés en analyse de survie (Lorino, 2002). Ces modèles semblent adaptés à la problématique agronomique de la rouille jaune. En effet, il est possible de considérer l'organisme vivant comme étant une parcelle, où l'on calcule le délai d'apparition de l'évènement maladie. Ceci prend en compte des variables agro-climatiques mesurées sur la parcelle et inclut un effet aléatoire année. L'originalité de la démarche est d'utiliser un modèle de survie pour proposer un nombre variable de sorties de contrôle de la maladie (entre 0 et 3 maximum). Le nombre de visites a été optimisé et la pertinence des résultats en fonction du nombre maximal de visites est étudié.

Vigicultures® est un portail Internet de saisie, partage et valorisation des observations des grandes cultures en temps réel sur tout le territoire de France métropolitaine. En place depuis 2008, et harmonisé depuis 2010, il a été mis en place par 4 instituts (ARVALIS-Institut du végétal, le Centre Technique Interprofessionnel des Oléagineux Métropolitains, l'Institut technique de la Betterave et l'Institut Technique du Lin) et est désormais présent dans 17 régions administratives (Sine *et al.*, 2010). Les travaux de modélisation sont réalisés à partir de ces données, sur les parcelles de blé tendre d'hiver et de printemps, des campagnes 2009 à 2014. Les parcelles sont observées de manière hebdomadaire selon un protocole de notation harmonisé à l'échelle nationale en libre accès sur internet.

Dans un premier temps, le pathosystème blé tendre/*P. striiformis* sera présenté, ainsi que le jeu de données utilisé et les principales particularités de celui-ci. Dans un second temps, le choix des variables agro-climatiques pertinentes dans la dynamique de la maladie sera explicité et la méthode de sélection des variables à implémenter sera développée. Puis, une partie sera consacrée à l'étude de

l'optimisation des seuils de risque attribués à chaque visite. Les limites et améliorations possibles du modèle seront évaluées en dernière partie.

MATERIEL ET MÉTHODE

PATHOSYSTEME BLE TENDRE/*P. STRIIFORMIS*

La rouille jaune est une maladie fongique de type aérien causée par le champignon basidiomycète *Puccinia striiformis*. C'est un champignon parasite biotrophe, polycyclique dont le blé est l'unique hôte connu à ce jour bien que des infections aient pu être observées sur orge, triticales et seigle en cas de forte pression de la maladie. Au cours des années, les différentes variétés de blé ont hérité de résistances plus ou moins efficaces contre la rouille et le traitement phytosanitaire n'est nécessaire que sur des variétés dites sensibles ou en cas de contournement par les races de rouille plus virulentes. Les races de rouille sont peu nombreuses en France et les principales différences entre celles-ci se retrouvent sur la période d'incubation au sein de l'hôte et sur la précocité des attaques (De Vallavieille-Pope, 2012). Seuls les traitements curatifs sont efficaces contre la rouille jaune, ainsi les agriculteurs doivent surveiller l'apparition des symptômes toutes les semaines de février à juin.

JEU DE DONNEES ET VARIABLES AGROCLIMATIQUES

Le jeu de données est issu des observations Vigicultures® décrites en introduction. Il est constitué de notations sur 2500 parcelles de blé tendre réparties sur tout le territoire de France métropolitaine de 2009 à 2014. Chaque parcelle est caractérisée par une variété, une date de semis, des coordonnées géographiques, un type de sol, un précédent et anté-précédant cultural ainsi que le type de travail du sol effectué. Les observations hebdomadaires sont des notations de stade de la plante et de maladie sur les trois dernières feuilles du moment que l'on concatène à l'échelle de la parcelle en présence/absence de symptômes. Grâce aux coordonnées géographiques, il est possible de récupérer les données des stations météorologiques les plus proches et de calculer les variables climatiques nécessaires à l'élaboration du jeu de variables exhaustif de départ. Chaque parcelle est ainsi caractérisée par un ensemble de variables agro-climatiques et l'information retenue est, soit la date d'apparition des symptômes pour les parcelles ayant subi une infestation, soit la dernière date d'observation pour les parcelles saines. Ceci constitue la base de l'élaboration des modèles de survie.

La rouille jaune est sensible aux conditions climatiques et suite à des recherches bibliographiques (Coakley *et al.*, 1982, 1988) et à l'expertise de spécialistes de la maladie, un premier jeu de 200 variables climatiques est élaboré selon les périodes temporelles décrites dans le tableau I.

Des variables d'incubation (quantité de spores) et un indice de contamination calculés à partir d'un modèle préexistant appelé Yello (Rouzet et Jacquin, 1995) au cours de ces mêmes périodes sont aussi retenus. A travers l'utilisation des données de l'année précédent la campagne, on tend à prendre en compte l'impact d'un inoculum primaire en début de campagne. Pour finir, la note de résistance variétale (nombre entier entre 1 et 9) complémente ce jeu de variables de départ.

Une première série d'analyses (corrélations deux à deux, ACP, ACM, ANOVA ou test de Chi² (Cornillon *et al.*, 2012), ainsi qu'une expertise de spécialistes, a permis de réduire la taille du jeu de variables par deux. Les variables restantes, ainsi que leurs interactions deux à deux, sont ensuite intégrées dans un modèle de Cox. Une seconde étape de sélection de variables a été effectuée à partir de ce modèle.

Tableau I : Périodes de calcul des variables climatiques du jeu de variables initial - *Calculation periods of climatic variables of the original set of variables.*

Campagne précédente				Campagne en cours	
Hiver	Sortie hiver	Printemps	Été	Levée-épi 1 cm	Epi 1 cm-2 Noeuds
15 dec. - 31 jan.	1er fev. - 15 mars	15 mars - 31 avr.	1er août - 15 sept.	-	-

SELECTION DE VARIABLES ET MODELE DE SURVIE DE TYPE MODELE DE COX

Sélection de variables – méthode LASSO

La sélection se fait à partir d'un jeu de 2500 variables qui comprend les variables retenues en §II.2 et les interactions deux à deux de ces variables. Celles-ci sont sélectionnées par une méthode de régression de type LASSO (Tournoud *et al.*, 2015). La méthode LASSO conduit à des coefficients de régression nuls pour les variables peu explicatives. Ces variables ne sont donc pas retenues dans le modèle. Le tableau II présente les variables sélectionnées (avec des coefficients non nuls).

Tableau II : Variables agronomiques et climatiques retenues dans le modèle de survie par méthode de sélection Lasso - *Agronomic and climatic variables included in the model of survival by the Lasso selection method.*

Variables	Définition
note_variete	note de résistance variétale entre 1 et 9
date_epi1cm_recalcule	date épi 1 cm en nombre de jours depuis le 1 ^{er} janvier
sTMiniinf0_leveeepi1cm	Somme des T°min > 0°C Levée - Epi 1cm Campagne N
sTMoynTMoysup7_leveeepi1cm	Somme des T°moyenne >7°C /Nombre de jours où T°moyenne >7°C Levée - Epi 1cm Campagne N
mTMini_leveeepi1cm	Moyenne des T°minimales Levée - Epi 1cm Campagne N
sIncubation_Yello_leveeepi1cm	Somme incubation (indice Yello entre 0 et 10) Levée - Epi 1cm Campagne N
sContamination_Yello_leveeepi1cm	Somme contamination (indice Yello entre 0 et 100) Levée - Epi 1cm Campagne N
mTMini_epi1cm2N	Moyenne des T°minimales Epi 1cm - 2 noeuds Campagne N
sContamination_Yello_hiver_precedent	Somme contamination (indice Yello entre 0 et 100) Hiver Campagne N-1
sTMoynTMoysup0_hiver_precedent	Somme des T°moyenne >0°C /Nombre de jours où T°moyenne >0°C Hiver Campagne N-1
sPETPsup0_sortiehiver_precedent	Somme pluie - ETP >0°C Sortie hiver Campagne N-1
nPluieinf02_sortiehiver_precedent	Nombre de jours de sécheresse Sortie hiver Campagne N-1
c2Pluieinf02_printemps_precedent	Nombre de séquence de 2 jours de sécheresse Printemps campagne N-1
sPluiePluiesup01_printemps_precedent	Somme des pluies/nombre de jours de pluie Printemps campagne N-1
sIncubation_Yello_printemps_precedent	Somme incubation (indice Yello entre 0 et 10) Printemps campagne N-1
sTMaxi_ete_precedent	Somme des T°maximales Été Campagne N-1

Ces variables sélectionnées ont ensuite été intégrées à un modèle de Cox défini ci-après.

Modèles de Cox

Les modèles de survie comptent plusieurs sous-famille de modèles (Lorino, 2002). Dans cet article uniquement les modèles de types modèles de Cox sont présentés. Les modèles de Cox sont des modèles semi-paramétriques à risques proportionnels. Ils sont descriptifs et prédictifs et peuvent intégrer des variables qualitatives, quantitatives et des facteurs à effets aléatoires. Le modèle de Cox spécifie que le risque instantané pour l'individu i dépend du temps et s'écrit :

$$\alpha_i(t) = \alpha_0(t) \cdot \exp(\beta_i Z_i)$$

$\alpha_0(t)$ est la fonction de risque de base, β est le vecteur de paramètres de régression et Z_i le vecteur de covariables pour l'individu i (Lorino, 2002). L'intégrale du risque instantané est le risque cumulé $\Lambda_i(t)$ à partir duquel on calcule la fonction de survie $S_i(t) = \exp(-\Lambda_i(t))$ (Saint-Pierre, 2015). Les modèles de survie sont régis par deux hypothèses principales de fonctionnement : la fonction de survie dépend du temps plus une ou plusieurs covariables qualitatives ou quantitatives et les risques sont proportionnels (les covariables ne dépendent pas du temps) (Abrahamowicz *et al.*, 2004). D'autres sorties du modèle sont possibles, mais dans le cadre de l'étude, seule la fonction de répartition est considérée. Cette fonction est égale à $1 - S_i(t)$, et illustre la probabilité d'apparition d'un évènement avant la date t, ici l'infestation de la parcelle, avant chaque date d'observation. La fonction de répartition est bornée entre 0 et 1 et la durée d'observation est celle du jeu de données d'ajustement (ici du 1^{er} janvier au 31 juillet). Dans le cadre de l'étude, seules les variables quantitatives sont retenues et la variable année est

un facteur aléatoire qui permet de prendre en compte d'autres effets spécifiques à chaque campagne non pris en compte par les variables quantitatives.

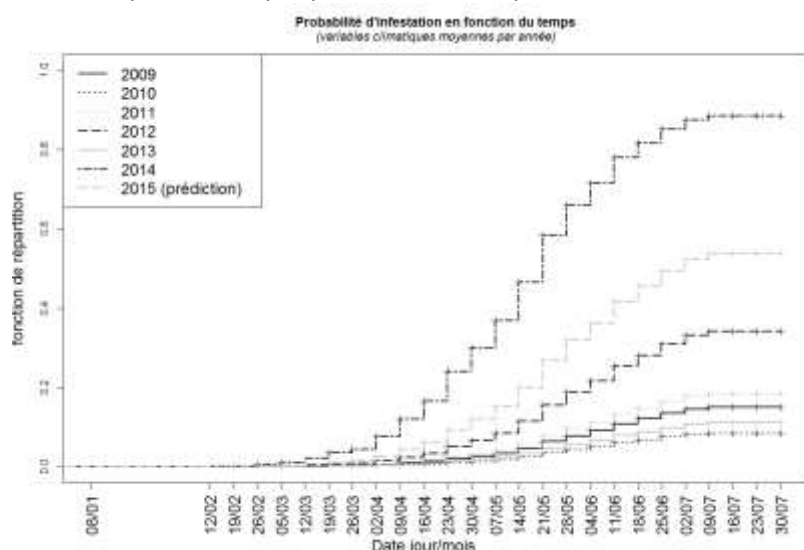


Figure 1 : Graphique des fonctions de répartition de chaque année. Les fonctions de répartition sont calculées sur le jeu de données d'ajustement (2009-2014) ainsi que l'année 2015 (prédiction), à partir des valeurs moyennes annuelles des variables climatiques - *Graph of annual distribution functions. The distribution functions are calculated on the set of adjustment data (2009-2014) and 2015 (prediction) from the annual average values of climate variables*

Cette représentation graphique des fonctions de répartition issue de l'ajustement du modèle de survie illustre parfaitement les données observées et ce, même en dehors du jeu de données d'ajustement (prédiction pour l'année 2015 : voir résultats et robustesse du modèle).

Utilisation du modèle de survie pour définir des dates de visite

La fonction de répartition calculée sur la parcelle i donne la probabilité d'apparition de l'infestation avant la date t . Trois seuils de probabilité ont été définis et, lorsqu'ils sont atteints sur une parcelle, une visite de cette parcelle est déclenchée. Sur certaines parcelles, le seuil le plus faible n'est jamais atteint, et donc aucune visite n'est déclenchée. Sur d'autres parcelles, si le premier seuil est atteint, cela déclenche une première visite. Ensuite, si la maladie n'est pas détectée à la première visite, et que plus tard dans la saison le deuxième seuil est atteint, une deuxième visite est déclenchée, et il en va de même pour la troisième visite. Les trois seuils de probabilité ont été choisis de manière à maximiser un critère d'acceptabilité défini au paragraphe suivant.

Notion d'acceptabilité des prédictions et optimisation des seuils de décision

Afin d'évaluer la qualité prédictive du modèle, un critère d'acceptabilité des résultats est défini spécifiquement pour le pathosystème rouille jaune/blé tendre : on considère qu'à partir d'un certain seuil de probabilité, l'évènement peut apparaître. On dispose alors d'une date d'apparition des symptômes prédite par le modèle. Cette prédiction est considérée comme acceptable lorsqu'elle est comprise entre une semaine avant et une semaine après la date d'apparition réelle. Ce critère d'acceptabilité est conforme à la réalité observée sur le terrain, c'est à dire au temps de réaction nécessaire à l'agriculteur pour une prise de décision. Il a été défini des critères d'acceptabilité de sensibilité et de spécificité du modèle. Ceci permet d'évaluer la qualité prédictive du modèle pour les parcelles infestées (sensibilité) et la capacité du modèle à ne pas prédire la maladie pour les parcelles saines (spécificité). Le critère d'acceptabilité totale est la moyenne des deux précédents critères, et des méthodes d'optimisation statistique automatisées permettent de retenir le seuil de décision pour lequel ce critère est maximal. La qualité des prédictions en fonction du nombre de seuil à définir est évaluée dans la partie résultats.

RESULTATS

Les résultats de trois stratégies de suivi des parcelles sont présentés : un seuil (approche 1), deux seuils (approche 2) et trois seuils (approche 3). Puis le détail des prédictions de l'approche 3 sont abordés. Les critères d'acceptabilités sont tous exprimés en pourcentages.

QUALITE DES PREDICTIONS CONCERNANT LES TROIS APPROCHES

La figure 2 présente les résultats de simulation des trois approches (Modèle de survie 1 visite (MS 1), MS 2 et MS 3) sur le jeu de données d'ajustement.

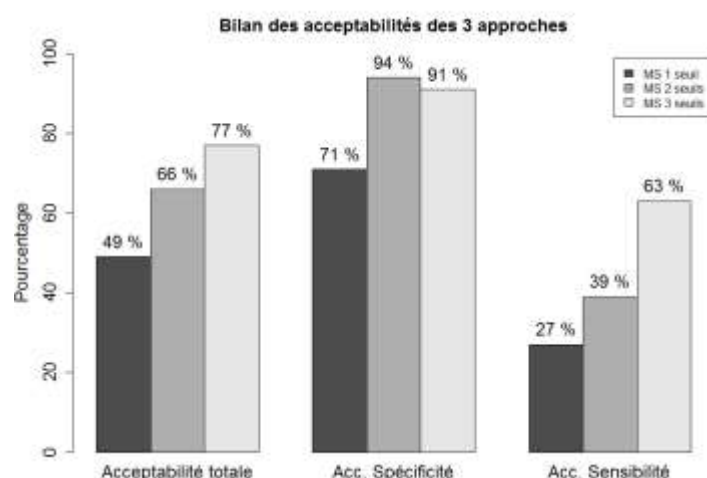


Figure 2 : Histogramme d'acceptabilité des résultats de simulation effectuées sur le jeu de données d'ajustement Vigicultures 2009-2014 d'un modèle de survie unique pour lequel un à trois seuils de décisions sont appliqués.

Histogram of acceptability of simulation results performed on the adjustment dataset Vigicultures 2009-2014 a single survival model for which 1-3 decision thresholds are applied.

L'approche 3 apparaît comme étant la plus acceptable selon les critères définis auparavant : par rapport à l'approche 1, un tiers de parcelles supplémentaires sont correctement détectées (soit deux fois plus de parcelles infestées). Plus le nombre de visite augmente, plus l'on a de chance de détecter l'apparition des symptômes au moment acceptable. En pratique, lorsqu'un seuil n'est pas atteint, la visite n'est pas effectuée. L'approche 3 propose ainsi un maximum de trois visites réparties sur toute la campagne, ce qui est considéré comme viable par les acteurs de terrain. C'est donc cette approche qui est retenue par la suite.

QUALITE DES PREDICTIONS CONCERNANT L'APPROCHE 3

Détail des prédictions des visites

Le jeu de données d'ajustement est constitué de 80 % de parcelles saines et 20 % de parcelles infestées. Le tableau ci-dessous présente les résultats de prédiction du modèle de survie avec l'approche 3 où il est observé que 70 % des parcelles totales sont visitées au moins une fois, 30 % deux fois et seulement 10 % trois fois.

Tableau III : Résultats des prédictions sur le jeu de données d'ajustement détaillées pour chaque seuil de décision atteint - *Results predictions on the detailed adjustment data set for each decision threshold reached.*

Parcelles et prédictions	Total	Détail
Parcelles saines	63% sont vues au moins une fois	0 Visite : 37% Visite 1 : 63% Visite 2 : 26% Visite 3 : 7%
Parcelles infestées détectées à temps	70% des parcelles infestées sont détectées au bon moment	Visite 1 : 25% Visite 2 : 30% Visite 3 : 15%
Parcelles infestées non détectées à temps	30% des parcelles infestées ne sont pas détectées au bon moment	0 Visite : 5% Visite 1 : 20% Visite 2 : 5% Visite 3 : 5%

La visite 2 est plus de deux fois discriminante que la visite 1 pour les parcelles saines, c'est à dire que l'observateur se déplace deux fois moins pour rien. C'est aussi la visite pour laquelle les résultats sont les plus acceptables pour les parcelles infestées détectées à temps, et celle pour laquelle il y a le moins de parcelles infestées non détectées à temps. De plus, cette visite concerne trois fois plus de parcelles

que la troisième visite, c'est la visite à ne pas manquer en pratique. Les écarts entre les visites 1 et 2 sont en moyenne de 4 semaines et de 3 semaines entre visite 2 et 3. Les dates de visites du modèle permettent un recouvrement temporel de toute la période d'activité de la maladie : du 26 avril pour les attaques précoces au 25 juin pour les attaques plus tardives. Les parcelles saines sont visitées en moyenne une fois par campagne et les parcelles infestées 1.7 fois.

La figure 3 est un boxplot qui présente les erreurs de prédiction liée à chaque visite pour les parcelles infestées non détectées à temps à l'échelle de la semaine. La largeur des boxplot est proportionnelle au nombre de parcelles concernées.

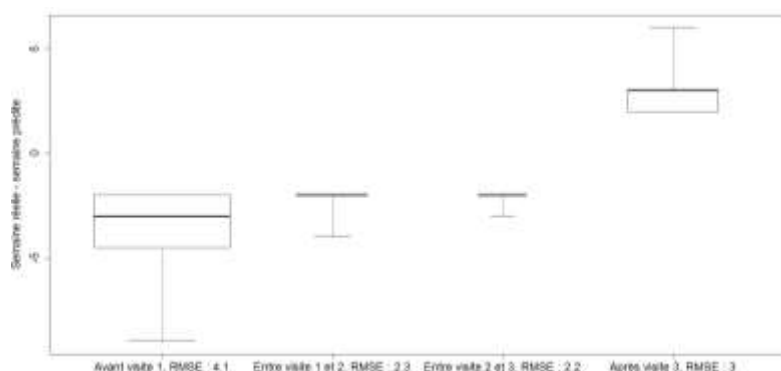


Figure 3 : Boxplot des écarts entre semaine réelle et semaine prédite et RMSE pour chacune des visites des parcelles infestées non détectées à temps - *Boxplot of differences between the actual week of apparition and predicted week and RMSE for each visit concerning infested field undetected in time.*

C'est lors de la première visite que la majorité des parcelles infestées non détectées à temps sont les plus représentées avec en moyenne un mois de retard de la prédiction par rapport à la date d'apparition réelle des symptômes. Entre les visites 2 et 3, les parcelles sont décelées avec 2 semaines de retard environ, mais ceci peut engendrer des pertes minimales en cas d'épidémie lente. La troisième visite détecte les parcelles avec 3 semaines d'avance en moyenne ce qui cause plus de dommages car l'observateur ne retournera plus sur la parcelle par la suite.

Robustesse du modèle sur un jeu de données extérieur aux données d'ajustement

Un jeu de données Vigicultures® blé tendre d'hiver 2015 a été extrait au 27/05/2015 et des prédictions ont été réalisées à partir du modèle de survie ajusté précédemment. L'acceptabilité totale de 77 % sur le jeu d'ajustement diminue de 20 % sur le jeu de données 2015. Ceci est dû à une diminution d'un tiers de la spécificité d'initialement 91 %, et la sensibilité de 66 % au départ n'est pas modifiée.

DISCUSSION

VARIABLES AGRO-CLIMATIQUES RETENUES

Les travaux de Coakley *et al.* (1982, 1988) ont orienté le choix des périodes de calcul des variables climatiques de départ, et les résultats de cette étude et de la sélection de variable confortent l'hypothèse que le climat de l'année précédant la campagne peut avoir un impact sur les épidémies de rouille jaune sur une parcelle.

DETAIL DES PREDICTIONS DES VISITES PAR ANNEE

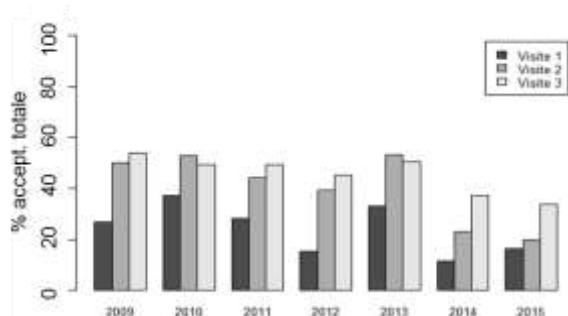


Figure 4 : Détail des % d'acceptabilité totale des résultats par visite et par année (2015 données extérieures au jeu d'ajustement) - *Detail of % total acceptability of results per visit per year (2015 data outside the data fitting)*

L'acceptabilité des visites est assez constante, excepté en 2012, 2014 et 2015 où l'acceptabilité est la plus faible (presque moitié moins que les autres), et qui sont les trois années où le taux d'infestation a été exceptionnel : respectivement 24 %, 50 % et 30 % des parcelles du jeu

d'observation infestées contre 10 % pour les années plus standards. D'autre part, il est possible de fournir une définition de l'utilisation de chacun des seuils en pratique : le seuil 1 permettrait de détecter les attaques précoces de maladie même si la plupart des erreurs de prédiction sont situées à ce niveau, surtout en 2014 où les attaques ont été très précoces (dès février selon les données terrain). Le deuxième seuil est le seuil pour lequel le plus de visites des parcelles infestées sont correctement détectées quelles que soient les années ainsi que le seuil pour lequel la plupart des parcelles saines ne sont pas revues (le plus spécifique, Tableau III). Le troisième seuil est celui pour lequel les infestations les plus tardives où dites les plus explosives sont détectées comme en 2015 par exemple. Le concept des trois visites permet donc de répartir au mieux les prédictions des infestations en fonction des épidémies selon les années. Cette approche pourrait permettre de parer les modifications de prédiction dues aux changements de race de rouille (De Vallavieille-Pope *et al.*, 2012) avec par exemple, la race *Warrior* apparue en 2011, et aux contournement des résistances variétales.

DISCUSSION CONCERNANT LE NOMBRE DE SEUIL A DEFINIR

L'étude présente un modèle de survie où trois seuils de décision ont été optimisés. L'optimisation de 4 et 5 seuils a aussi été analysée. La figure 5 présente l'acceptabilité des résultats du modèle de Cox en fonction du nombre de seuil définis.

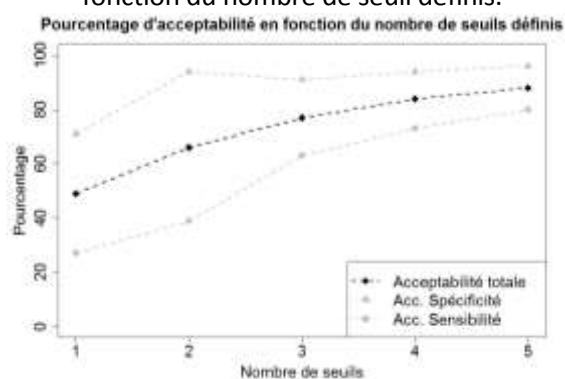


Figure 5 : Acceptabilité totale, spécifique et sensible en fonction du nombre de seuils– *Total, specific and sensitive acceptability to the number of thresholds.*

L'acceptabilité totale des résultats augmente en fonction du nombre de visites, mais ce jusqu'à une certaine limite. La spécificité du modèle semble atteindre un maximum et se stabiliser dès la deuxième visite, la sensibilité augmente progressivement mais dès le troisième seuil défini, son accroissement diminue de moitié à chaque seuil supplémentaire. L'acceptabilité totale est la moyenne des deux précédentes. S'il n'y avait pas de contrainte de terrain, il serait intéressant de visualiser le nombre optimal de seuils à définir pour que le modèle de Cox atteigne un maximum de sensibilité et spécificité. Dans le cadre de ces travaux et de l'implémentation dans un outil d'aide à la décision pour les agriculteurs, le nombre de trois visites au maximum semble le meilleur compromis.

VALIDATION DE LA DISTRIBUTION DES VARIABLES CLIMATIQUES

Lors de l'utilisation d'un modèle prédictif, il est intéressant de vérifier que les variables du jeu de données d'ajustement sont bien représentatives de la variabilité existante (Gouache *et al.*, 2015). L'analyse en composantes principales (ACP) est une approche qui permet d'étudier les ressemblances entre individus du point de vue de l'ensemble des variables (Cornillon *et al.*, 2012). Les 15 variables retenues dans le modèle ont été calculées sur un jeu de données de la France entière sur la période 1995-2015, et en considérant les deux variétés de blé sensible et résistante les plus représentées sur le territoire. Une ACP a été réalisée sur ce jeu de données et les observations sont représentées sur le premier plan factoriel en superposant les observations du jeu d'ajustement en tant qu'individus supplémentaires :

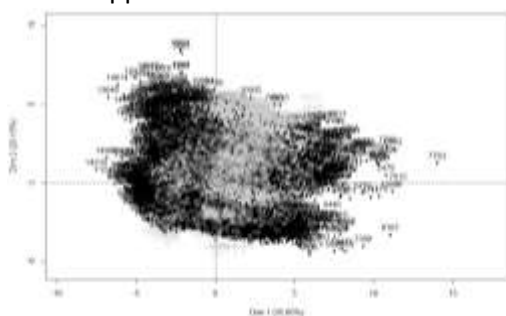


Figure 6 : ACP des individus du jeu de données d'ajustement (en gris) et d'un jeu de données France entière depuis 1995 (variété Apache/Alixan) (en noir) – *PCA of individuals in the adjustment dataset (gray) and a set of data all over France since 1995 (variety Apache / Alixan) (black).*

Les répartitions des individus du jeu d'ajustement et du jeu France entière semblent se superposer sur les deux premières dimensions. Le jeu d'ajustement semble donc représentatif de la

variabilité existante des 20 dernières années ; le modèle peut être utilisé sur les années futures.

CONCLUSION

Le modèle présenté dans cet article est l'un des premiers modèles ajustés sur un jeu de données Vigicultures® et les résultats de ces travaux encouragent l'utilisation future des données de cet outil de surveillance.

D'après la sélection de variables retenues dans le modèle de Cox, il semblerait que les épidémies de rouille jaune soient sensibles à la résistance variétale, au stade de développement variétal, mais aussi à un ensemble de variables climatiques tels que la pluviométrie et la température principalement. La somme de pluie, le nombre de jours de sécheresse, le nombre de jours de gel, la somme des températures maximale en été ou minimales en hiver ainsi que des interactions de ces variables entre elles par exemple, ont un impact sur le délai d'apparition des symptômes. De plus, les périodes de calcul de ces variables climatiques montrent que le climat de l'année précédant la campagne actuelle, jusqu'à l'hiver précédent, a un impact sur le délai d'apparition des symptômes.

Les modèles de survie se révèlent être des outils fonctionnels pour répondre à la problématique de départ : délai d'apparition de symptômes sur une parcelle. D'une part, les résultats de simulation concernant les taux d'acceptabilité sont suffisamment pertinents pour pouvoir envisager une utilisation en campagne, d'autre part, le modèle semble plus robuste sur des prédictions utilisant des données extérieures aux données d'ajustement que tous les modèles rouille jaune existant à ce jour chez Arvalis. En effet, l'acceptabilité totale diminue de 20 % sur la prédiction des parcelles 2015, en raison d'une perte de spécificité du modèle (la capacité du modèle à ne pas prédire la maladie lorsqu'il n'y en a pas). Or, sur le terrain, il s'avère préférable que les acteurs explorent une parcelle saine plutôt qu'ils ne se déplacent pas alors que la parcelle était infestée. De plus, il apparaît que les données climatiques d'ajustement sont représentatives de la variabilité existante sur toute la France depuis les 20 dernières années ; ainsi, on peut penser que le modèle restera valide d'ici les quelques années à venir.

Le pathosystème rouille jaune/blé tendre d'hiver sert de support à cette étude, or les modèles de survie se révèlent assez performants et il sera intéressant de généraliser ces travaux à d'autres pathosystèmes.

REMERCIEMENTS

Toute la team RJ : François Piroux, Emmanuelle Gourdain, Gilles Couleaud, Xavier le Bris, Fabrice Moreau, Olivier Deudon, Emmanuelle Heritier ; mais aussi Claude Pope, les observateurs Vigicultures® et Danièle Simoneau, Therneau Terry.

BIBLIOGRAPHIE

- ❖ Abrahamowicz M., Du Berger R., Krewski D., Burnett R., Bartlett G., Tamblyn M.R., Leffondré K., 2004 - Bias due to aggregation of individual covariates in the Cox regression model. *American Journal of Epidemiology*, 160, 696-706.
- ❖ Chen X.M., 2005 - Epidemiology and control of stripe rust [*Puccinia striiformis* f. sp. *tritici*] on wheat. *Plant Pathology*, 27, 314-337.
- ❖ Chen J., Chu C., Souza A.J., Guttieri M.J., Chen X., Xu S., Hole D., Zemetra R., 2012 - Genome-wide identification of QTL conferring high-temperature adult-plant (HTAP) resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in wheat. *Molecular Breeding*, 29, 791-800.
- ❖ Coakley S.M., Boyd W.S., Line R.F., 1982 - Statistical models for predicting stripe rust on winter wheat in the Pacific Northwest. *Phytopathology*, 72, 1539-1542.
- ❖ Coakley S.M., Boyd W.S., Line R.F., 1984 - Development of regional models that use meteorological variables for predicting stripe rust disease on winter wheat. *Journal of Climate and Applied Meteorology*, 23, 1234-1240.

- ❖ Coakley S.M., Line R.F., 1981 - Quantitative relationships between climatic variables and stripe rust epidemics on winter wheat. *Phytopathology*, 71, 461-467.
- ❖ Coakley S.M., Line R.F., McDaniel L.R., 1988 - Predicting stripe rust severity on winter wheat using an improved method for analyzing meteorological and rust data. *Phytopathology*, 78, 543-550.
- ❖ Cornillon P.A., Guyader A., Husson F., Jégou N., Josse J., Kloareg M., Matzner-Løber E., Rouvière L., 2012 - Statistiques avec R. *Presses Universitaires de Rennes*, 296 p.
- ❖ De Vallavieille-Pope C., Ali S., Leconte M., Enjalbert J., Delos M., Rouzet J., 2012 - Virulence dynamics and regional structuring of *Puccinia striiformis* f. sp. *tritici* in France between 1984 and 2009. *Plant Disease*, 96, 131-140.
- ❖ De Vallavieille-Pope C., Huber L., Leconte M., Bethenod O., 2002 - Preinoculation effects of light quantity on infection efficiency of *Puccinia striiformis* and *P. triticina* on wheat seedlings, *Phytopathology*. 92, 1308-1814.
- ❖ Garrett K.A., Mundt C.C., 2000 - Effects of planting density and the composition of wheat cultivar mixtures on stripe rust: An analysis taking into account limits to the replication of controls. *Phytopathology*, 90, 1313-1321.
- ❖ Gouache D., Léon M.S., Duyme F., Braun P., 2015 - A novel solution to the variable selection problem in Window Pane approaches of plant pathogen – climate models: Development, evaluation and application of a climatological model for brown rust of wheat. *Agricultural and Forest Meteorology*, 205, 51-59.
- ❖ Lorino T., 2002 - Modèles statistiques pour des données de survie corrélées. Life Sciences [q-bio]. Institut National Agronomique Paris-Grignon - INA P-G.
- ❖ Ma L., Qiao J., Kong X., Zou Y., Xu X., Chen X., Hu X., 2015 - Effect of low temperature and wheat winter-hardiness on survival of *Puccinia striiformis* f. sp. *tritici* under controlled conditions. *PLoS ONE*, 10:e0130691.
- ❖ Mboup M., Bahri B., Leconte M., De Vallavieille-Pope C., Kaltz O., Enjalbert J., 2012 - Genetic structure and local adaptation of European wheat yellow rust populations: the role of temperature-specific adaptation. *Evolutionary Applications*, 5, 341-352.
- ❖ Rouzet J., Jacquin D., 1995 - Aide à la décision : modèles de prévision et lutte raisonnée. *Phytoma*, 475, 29-32.
- ❖ Saint-Pierre P., 2015 - Introduction à l'analyse des durées de survie. Cours Université Pierre et Marie Curie.
- ❖ Siné M., Morin E., Simonneau D., Brochard M., De Cosnac G., Escriou H., 2010 - VIGICULTURES - An early warning system for crop pest management. Scientific and Technical Information and Rural Development, IAALD XIIIth World Congress, Montpellier, 26-29 avril 2010.
- ❖ Tournoud M., Larue A., Cazalis M.A., Venet F., Pachot A., Monneret G., Lepape A., Veyrieras J.B., 2015 - A strategy to build and validate a prognostic biomarker model based on RT-qPCR gene expression and clinical covariates, *BMC Bioinformatics*, 16:106.