

**AFPP – 10^e CONFÉRENCE INTERNATIONALE SUR LES MALADIES DES PLANTES
TOURS – 3, 4 et 5 DÉCEMBRE 2012**

**COMMENT COMBINER RESEAU D'OBSERVATIONS, EXPERTISE ET MODELISATION
POUR ELABORER LE BULLETIN DE SANTE DU VEGETAL ?**

F. BRUN⁽¹⁾, D. MAKOWSKI⁽²⁾ ET F. PIRAUX⁽³⁾

⁽¹⁾ ACTA, 31320 Auzeville, France, francois.brun@acta.asso.fr

⁽²⁾ INRA, Agronomie, 78850 Thiverval-Grignon, France, david.makowski@grignon.inra.fr

⁽³⁾ Arvalis - Institut du végétal, 91720 Boigneville, France, f.piroux@arvalisinstitutduvegetal.fr

RÉSUMÉ

Le Bulletin de Santé du Végétal (BSV) fournit une information sur l'évolution de l'état sanitaire des cultures. Des outils de modélisation existent pour certaines maladies ou ravageurs et peuvent apporter une information complémentaire aux observations de terrain et à l'expertise. Pour aboutir au diagnostic complet inclus dans le BSV, les acteurs doivent intégrer ces différentes informations (données d'observation, de modélisation et d'expertise). Dans le cadre du RMT modélisation (www.modelia.org), nous avons identifié quelques pistes potentielles pour mieux formaliser et améliorer le processus d'intégration des données en profitant d'innovations en terme de techniques statistiques et de nouvelles technologies. Il s'agit 1) des méthodes d'assimilation de données et 2) des méthodes de spatialisation des simulations et des erreurs de prédiction. Nous illustrerons ces possibilités sur des exemples.

Mots-clés : modélisation, réseau d'observation, analyse de risque, BSV, méthodologie.

SUMMARY

HOW TO COMBINE OBSERVATIONS, MODELING AND EXPERTISE TO ELABORATE A PLANT HEALTH DIAGNOSIS?

Report on Plant Health (BSV in France) provides information on the evolution of the health status of crops. Modeling tools exist for specific diseases or pests and may provide additional information to field observations and expertise. To reach a diagnosis included in the full report, actors must integrate these different information (observational data, modeling and expertise).

As part of the French network modeling for Agriculture (www.modelia.org), we identified a few potential ways to better formalize and improve the process of data integration by taking advantage of innovations in terms of statistical techniques and new technologies. These are 1) methods of data assimilation and 2) methods for spatialisation of simulation and prediction errors. We illustrate these possibilities with examples.

Key words: modeling, observation network, risk analysis, plant health, methodology.

INTRODUCTION

Avec le plan ECOPHYTO 2018, le système des avertissements pour les attaques de maladies ou de ravageurs des cultures (Avertissements agricoles®) a été fondamentalement transformé depuis le 1er janvier 2010. Ainsi, le nouveau Bulletin de Santé du Végétal (BSV, <http://agriculture.gouv.fr/consultez-les-bulletins-de-sante>) est un outil important du plan ECOPHYTO 2018 et de son axe 5 « Renforcer la surveillance sur les bio-agresseurs et les effets non intentionnels des pesticides ». Ce sont désormais les BSV qui, sous la responsabilité des Chambres Régionales d'Agriculture ont pour mission d'informer les acteurs du monde agricole de l'état sanitaire des principales cultures sur leur territoire régional. La demande pour des résultats de modélisation est forte : dans le cahier des charges des BSV, il est demandé que les bulletins incluent la « Dynamique des organismes nuisibles estimée par l'animateur et/ou prédite par les modèles épidémiologiques, illustrée par graphiques, cartes. » (Comité National Epidémiosurveillance du 7 juillet 2009).

Les modèles utilisés pour la protection des cultures sont nombreux et largement utilisés par les acteurs du développement agricole (FNLON, 2009). Un modèle permet de simuler un niveau de risque de présence d'un bio-agresseur (maladie ou ravageur) à partir de données climatiques, d'anticiper son développement et, ainsi, de contribuer à l'analyse de risque pour le raisonnement des traitements par les acteurs agricoles. Ces modèles ont des avantages importants puisqu'ils permettent de calculer un risque dans un grand nombre de contextes (modes de conduites, météo,...). Certains modèles calculent également des variables difficiles à mesurer mais pertinentes pour évaluer un risque (par exemple, le nombre d'infections non observable).

Néanmoins, dans la pratique, lorsque des modèles sont disponibles, leurs utilisateurs ne se basent pas exclusivement sur leurs simulations car celles-ci sont toujours incertaines et potentiellement entachées d'erreurs. Les utilisateurs prennent donc en compte non seulement les résultats d'un modèle mais aussi des résultats d'essais, de réseaux d'observations, notamment sur des parcelles de témoins non traités, et d'expertises. Dans un tel contexte, il est important d'étudier non pas l'apport de la seule modélisation, mais l'apport de la modélisation comme partie intégrante d'une démarche globale.

Dans la quasi-totalité des travaux méthodologiques portant sur les modèles utilisés pour la protection des cultures, les problèmes liés à l'intégration des simulations des modèles avec d'autres sources d'information sont ignorés. Dans la pratique, ce travail d'intégration se fait déjà dans chaque dispositif, de façon souvent implicite et sans qu'il soit forcément formalisé.

Certaines éditions du BSV sont effectivement basées sur la combinaison des résultats de modèles, d'un réseau d'observation et de l'expertise. Le travail d'intégration des différentes sources d'information doit se faire très fréquemment, à chaque réunion de rédaction en cours de campagne (une fois par semaine). Il nous semble particulièrement important de mieux formaliser et d'améliorer le processus d'intégration des données en profitant d'innovation en terme de techniques statistiques et de nouvelles technologies afin de profiter au mieux des informations sur l'état sanitaire des cultures afin d'améliorer la qualité et la représentativité de l'information contenue dans le BSV. Ces améliorations contribueront à l'atteinte des objectifs de réduction des produits phytosanitaires liée à l'optimisation du nombre de traitements et de leur positionnement. Dans le cadre du RMT modélisation (www.modelia.org), nous avons commencé à mener une réflexion sur ce sujet (RMT modélisation, 2012)

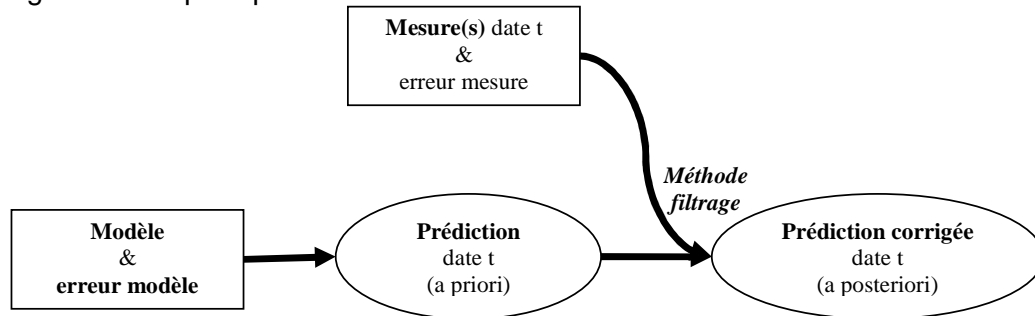
L'objectif de ce papier est **de présenter des pistes pour mieux formaliser et améliorer le processus d'intégration des données dans le cadre du BSV** en profitant d'innovations en terme de techniques statistiques et de nouvelles technologies. Il s'agit **1) des méthodes d'assimilation de données** qui permettent de corriger les prédictions d'un modèle à partir d'observation(s) en cours de campagne, **2) des méthodes de spatialisation des simulations et des erreurs de prédiction** qui permettent de profiter d'un réseau d'observation régional pour corriger les prédictions.

Nous illustrerons les méthodes d'assimilation de données sur un exemple à vocation pédagogique, mais non opérationnel, afin de faciliter la compréhension de ces méthodes.

LES METHODES D'ASSIMILATION DE DONNEES

Ces méthodes permettent de prendre en compte une information spécifique à une situation (un lieu et/ou une année par exemple) pour modifier le modèle (Makowski et al, 2006). Typiquement, cette information peut être une ou plusieurs mesures en cours de saison correspondant à une variable d'état simulée, comme par exemple des observations de sévérité. Le principe est alors de modifier la variable d'état afin de mieux correspondre à l'observation d'un jour j (par exemple issu d'un réseau d'observation) et d'utiliser ce modèle modifié pour les prédictions futures (par exemple prédiction sur la base d'un scénario météo à 10 jours) (Figure 1).

Figure 1 : principe de l'assimilation de données.



Les potentiels de correction dépendent notamment de la relation entre les variables mesurées et celles simulées. A noter que le modèle corrigé ne peut être légitimement utilisé que dans la situation où l'on a procédé à cette assimilation de données.

Différentes familles de méthodes existent, mais nous limiterons notre étude aux méthodes de « filtrage » des séries temporelles.

MATERIELS ET METHODES

L'ensemble des calculs ont été réalisés dans le logiciel R (R Development Core Team, 2011) et l'ensemble des scripts sont disponibles sur demande.

Exemple pédagogique : un modèle simplifié de dynamique des populations

Comme modèle simplifié nous proposons le modèle de croissance logistique de Verhulst (1840) utilisé en dynamique des populations. Ce modèle peut s'écrire sous la forme d'une équation différentielle de Y, la taille de la population, par rapport à t, le temps.

$$(m1) \quad \frac{dY}{dt} = a \cdot Y \cdot \left(1 - \frac{Y}{k}\right)$$

Avec a, k deux paramètres positifs : a le taux de croissance et k la capacité d'accueil. Si la condition initiale $0 < Y_0 < k$ (à $t=1$), cela implique que $Y < k$ pour tout t. L'équation se résout analytiquement (solution logistique), mais pour faciliter la généralisation à d'autres modèles de maladie climatique, nous choisissons la solution numérique avec une discrétisation d'un jour.

$$(m2) \quad Y(t=1) = Y_0$$

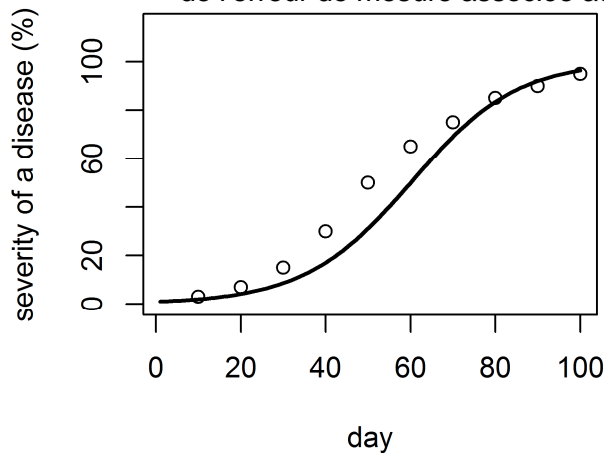
$$(m3) \quad dY = a \cdot Y \cdot \left(1 - \frac{Y}{k}\right)$$

$$(m4) \quad Y(t+1) = Y(t) + dY$$

D'un point de vue biologique, dans le cas des maladies foliaires, on peut considérer que y est la quantité de surface présentant des lésions, que K est la surface foliaire totale, constante au court du temps, et que t est un nombre de jour depuis l'infection initiale (à $t=1$, $Y=Y_0$). On travaille en pourcentage avec $k=100\%$, $Y_0=1\%$ et $a=0.08 \cdot \text{jour}^{-1}$ (Figure 1).

Nous faisons l'hypothèse que nous avons en parallèle des données. Pour l'illustration, ce sont des données fictives (Figure 1), qui diffèrent légèrement du modèle (dans notre cas d'école, une surestimation lié à un démarrage plus rapide de la maladie).

Figure 1 : Dynamique de la sévérité d'après le modèle de Verhulst (ligne) et données fictives utilisées pour l'illustration (cercle). On a une information sur l'écart-type de l'erreur de mesure associée aux observations.



Jour	Y_{obs} cas 1: $\sigma_{Y_{obs}} = 15\%$ cas 2: $\sigma_{Y_{obs}} = 3\%$
10	3
20	7
30	15
40	30
50	50
60	65
70	75
80	85
90	90
100	95

Filtre de Kalman

Nous présentons de manière concise le principe du filtre de Kalman (Kalman et Bucy, 1961) qui est un des algorithmes les plus simples et les plus connus (pour le détail, voir Makowski et al, 2006). Le filtre s'applique pour un modèle défini par une relation linéaire entre les variables d'état au temps t et au temps $t+1$ et incluant un terme résiduel aléatoire distribué selon une loi normale.

$$(k1) \quad Y_{t+1} = a \cdot Y_t + b_t + \varepsilon_t \quad \varepsilon_t \sim N\{0, \sigma_\varepsilon\}$$

Avec ce modèle, il est possible de calculer analytiquement l'espérance et la variance de $Y(t)$ à tout t .

Supposons qu'on dispose d'une mesure reliée directement à la variable d'état Y , avec une erreur de mesure τ tel que :

$$(k2) \quad M_t = Y_t + \tau_t \quad \tau_t \sim N\{0, \sigma_\tau\} \text{ (indépendant de } \varepsilon)$$

On peut alors écrire la distribution de Y_t conditionnellement à M_t (K_t est le gain de Kalman) :

$$(k3) \quad Y_t | M_t \sim N\{E(Y_t) + K_t \cdot (M_t - E(Z_t)), (1 - K_t) \cdot \text{var}(Y_t)\}$$

$$K_t = \frac{\text{var}(Y_t)}{\text{var}(Y_t) + \text{var}(\tau_t)}$$

Si l'erreur de mesure est nulle ($\text{var}(\tau_t) = 0$), la prédiction est remplacée par la mesure.

Cette approche peut être adaptée aux modèles incluant plusieurs variables d'état (Y_t étant un vecteur des variables d'état), aux modèles dont les équations sont non linéaires avec la version étendue du filtre de Kalman et aux relations linéaires plus complexes entre les mesures M et l'état du modèle Y .

Le filtre de Kalman peut se mettre en œuvre avec le logiciel R (R Development Core Team, 2011) et la librairie dlm pour travailler avec les modèles linéaires dynamiques (Petris, 2010), en utilisant les fonctions dlmModPoly, dlmModReg, dlmMLE, dlmFilter. Il s'agit de techniques très utilisées pour l'analyse de série temporelle et l'on peut alors considérer le modèle de Verhulst comme une variable d'entrée de ces modèles statistiques d'analyse.

Marche aléatoire. Pour commencer, on peut considérer un modèle de base de marche aléatoire sans tendance (θ est la variable d'état de la marche aléatoire). Dans ce cas,

l'information issue du modèle dynamique de Verhulst n'est pas utilisée et seule les données d'observation vont être mobilisée (il ne s'agit pas de correction).

On définit ainsi le modèle de marche aléatoire.

$$(k4) \quad M_t = \theta_t + \tau_t \quad \text{avec} \quad \tau_t \sim N\{0, \sigma_\tau\}$$

$$(k5) \quad \theta_{t+1} = \theta_t + \varepsilon_t \quad \text{avec} \quad \varepsilon_t \sim N\{0, \sigma_\varepsilon\}$$

Avec σ_ε un paramètre inconnu à estimer (argument dV sous dlm).

Voici la mise en œuvre sous R:

```
library(dlm)
# M : observations considérées correspondant aux valeurs Y (NA : valeurs manquantes)
# spécification du modèle de marche aléatoire avec sigma_e inconnu
# var_tau : variance des observations
buildFun=function(x) {return(dlmModPoly(order=1, dV=Vemp, dW=exp(x), C0=400, m0=1))}
# Estimation de sigma_e
fit<-dlmMLE(M, parm=c(0.5), build=buildFun, method = c("Nelder-Mead"))
# définition du modèle avec sigma_e estimé
fitted.modY=buildFun(fit$par)
# filtre de Kalman appliqué aux variables d'état et aux données
modFilter=dlmFilter(M, mod=fitted.modY)
# Ycor : variable d'état corrigée
Ycor=modFilter$m[-1]
```

Si le modèle dynamique épidémiologique n'est que faiblement explicatif, la marche aléatoire donnera de bons résultats, puisqu'il permet uniquement de filtrer les données.

Modèle de régression. En fait, on a une information *a priori* sur notre système qui est la simulation du modèle dynamique de Verhulst (Y_t). On va donc définir une relation entre cette information et nos données mesurées (régression locale).

$$(k6) \quad M_t = \alpha_t + \beta_t \cdot Y_t + \tau_t \quad \text{avec} \quad \tau_t \sim N\{0, \sigma_\tau\}$$

$$(k7) \quad \alpha_{t+1} = \alpha_t + \varepsilon_t \quad \text{avec} \quad \varepsilon_t \sim N\{0, \sigma_\varepsilon\}$$

$$(k8) \quad \beta_{t+1} = \beta_t + \varepsilon'_t \quad \text{avec} \quad \varepsilon'_t \sim N\{0, \sigma_{\varepsilon'}\}$$

Avec σ_ε et $\sigma_{\varepsilon'}$ deux paramètres inconnus à estimer

Voici la mise en œuvre sous R:

```
library(dlm)
# Y : simulation pour de la variable d'état pour chaque jour (matrice d'une colonne)
# M : observations considérées correspondant aux valeurs Y (NA : valeurs manquantes)
# spécification du modèle de régression avec sigma_e et sigma_e' inconnus
# var_tau : variance des observations
buildFun=function(x) {return(dlmModReg(Y, dV=var_tau, dW=c(exp(x[1]), exp(x[2])), m0=c(0,1)))}
# Estimation de sigma_e et sigma_e'
fit=dlmMLE(M, parm=c(0.5,0.5), build=buildFun, method = c("Nelder-Mead"))
# définition du modèle avec sigma_e et sigma_e' estimés
fitted.modY=buildFun(fit$par)
# filtre de Kalman appliqué aux variables d'état et aux données
modFilter=dlmFilter(M, mod=fitted.modY)
# Ycor : variable d'état corrigée
Ycor=modFilter$m[-1,1] + modFilter$m[-1,2]*Y
```

Filtre particulière

Le filtrage particulière permet ainsi d'obtenir le même type de résultat qu'avec le filtre de Kalman, mais l'avantage du filtrage particulière est que cette méthode peut être appliquée à des modèles non linéaires et pas seulement à des modèles linéaires. Le filtre particulière est une méthode qui est facile à appliquer avec une grande diversité de modèles, mais qui nécessite de nombreuses simulations. Comme le filtre de Kalman, l'application du filtrage particulière se déroule en deux étapes : 1) l'étape de propagation à travers le modèle avec notamment le terme d'erreur ε du modèle (f est la fonction de mise à jour du modèle dynamique) et 2) l'étape de correction.

A chaque pas de temps :

1) on propage N fois Y au temps (t+1), les ε étant tirés dans une distribution normal. On fixe l'écart type de ε à 3% (ce choix est arbitraire pour l'exemple, mais son choix pose une réelle difficulté).

$$(p1) \quad Y_{t+1,i} = f(Y_{t,i}) + \varepsilon_{t,i} \quad \text{pour } i=1\dots N$$

2) s'il n'y a pas de mesure, $Y_{cor,t+1} = Y_{t+1}$

3) s'il y a une ou plusieurs mesures

3a) on calcule les poids (avec comme choix de loi de proposition celui de l'a priori ce qui simplifie le calcul d'importance sampling).

$$(p2) \quad w_{t+1,i} = \frac{p(Y_{t+1,i} | M_{t+1,i})}{p(Y_{t+1,i})} \quad \text{puis } w_{t+1,i}^* = \frac{w_{t+1,i}}{\sum_{k=1}^N w_{t+1,k}}$$

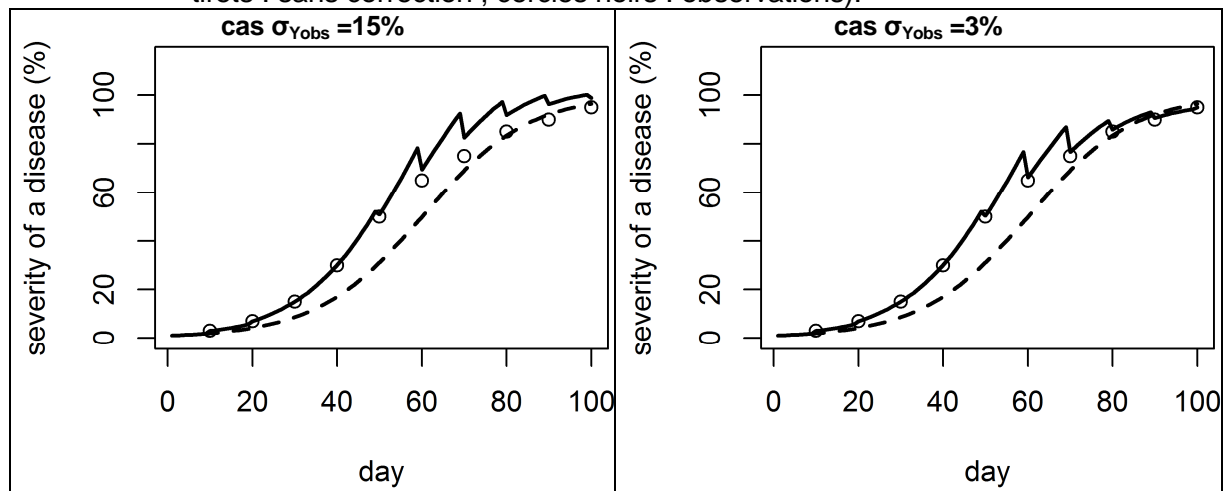
3a) on ré-échantillonne N valeurs de $Y_{cor}(t+1)$ dans la distribution discrète définie par l'ensemble des N $(Y_{t+1,i}, w_{t+1,i}^*)$ (les valeurs de Y associées à un faible poids vont « disparaître » au profit des valeurs associées à un poids important).

Cela correspond à l'application d'une méthode d'estimation de probabilité par importance sampling séquentiellement à chaque date de mesure de manière à estimer la distribution a posteriori $Y_t | M_t$ pour toutes les dates t où une mesure est disponible. On n'a alors une approximation des distributions a posteriori et la qualité de cette approximation dépend du niveau de complexité du modèle et de la taille de l'échantillon N utilisée.

RESULTATS ET DISCUSSION

Le filtre de Kalman permet de prendre en compte l'information sur les observations pour estimer la valeur de la sévérité (Figure 3). Les résultats correspondent à la définition d'une régression linéaire dynamique entre les observations et le modèle de Verhulst (équation k6,7,8). La correction avec le modèle de marche aléatoire n'a pas donné de bons résultats, le modèle de Verhulst étant assez explicatif dans notre cas d'étude.

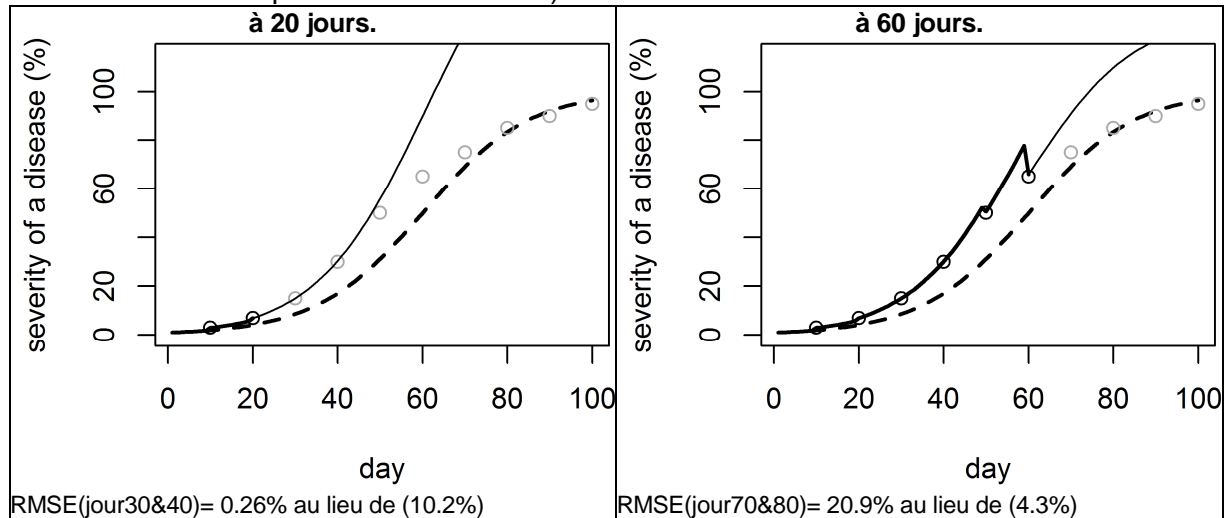
Figure 3 : Correction avec le filtre de Kalman du modèle de Verhulst (ligne : correction ; tirets : sans correction ; cercles noirs : observations).



On remarque nettement l'effet de l'erreur de mesure des observations : plus cette erreur est réduite, plus de poids est donné aux observations.

Maintenant, dans la pratique, à un moment donné, on va vouloir corriger le modèle avec les données recueillies jusqu'à ce moment et utiliser cette correction pour des prédictions suivantes (Figure 4). On peut alors évaluer la qualité de la correction (ici sur les deux dates suivantes).

Figure 4 : Correction avec le filtre de Kalman du modèle de Verhulst (pour $\sigma_{Y_{obs}}=3\%$) (ligne : correction ; tirets : sans correction ; cercles noir : observations prises en compte pour la correction, cercle gris : observations suivantes pour l'évaluation de la qualité de la correction).

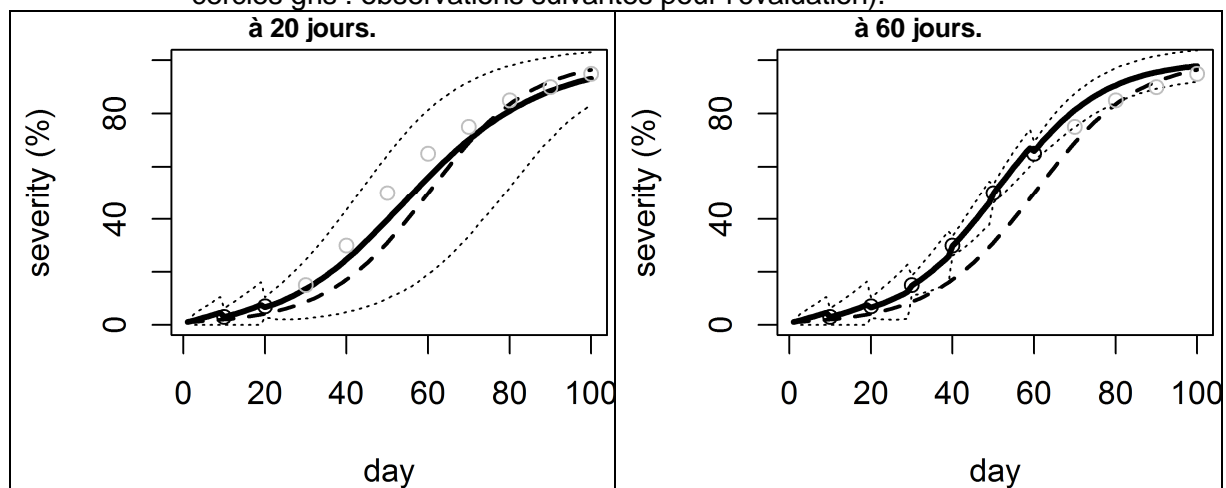


Dans cette exemple, cette correction peut sembler très efficace en début de campagne (on est dans une phase relativement linéaire du modèle avec $dY \approx a \cdot Y$ lorsque $Y \ll k$) et nettement moins par la suite.

De manière similaire, **le filtre particulaire** permet de prendre en compte l'information sur les observations pour estimer la valeur de la sévérité. Lorsque l'on va évaluer les qualités de prédictions du modèle sachant les mesures précédentes, on voit une certaine amélioration dans le cas à 60 jours, mais on fait légèrement moins bien à 20 jours (Figure 5).

Cette amélioration partielle est intéressante et certainement liée au fait que les techniques de filtres particulaires sont plus adaptées aux modèles fortement non linéaires. Si l'on se place en début de campagne (20 jours), cette correction de la variable d'état montre aussi une forte variance des prédictions après correction liée à la définition du terme d'erreur du modèle qui se propage. Par ailleurs, la technique du filtre particulaire est plus coûteuse en temps de calcul et d'adaptation du modèle, puisqu'il faut rajouter un terme d'erreur à la variable d'état et, donc, intervenir dans le code du modèle.

Figure 5 : Correction avec le filtre particulaire du modèle de Verhulst (pour $\sigma_{Y_{obs}} = 3\%$). (ligne : correction ; pointillés : intervalle de confiance [10%-90%]; tirets : sans correction ; cercles noirs : observations prises en compte pour la correction, cercles gris : observations suivantes pour l'évaluation).

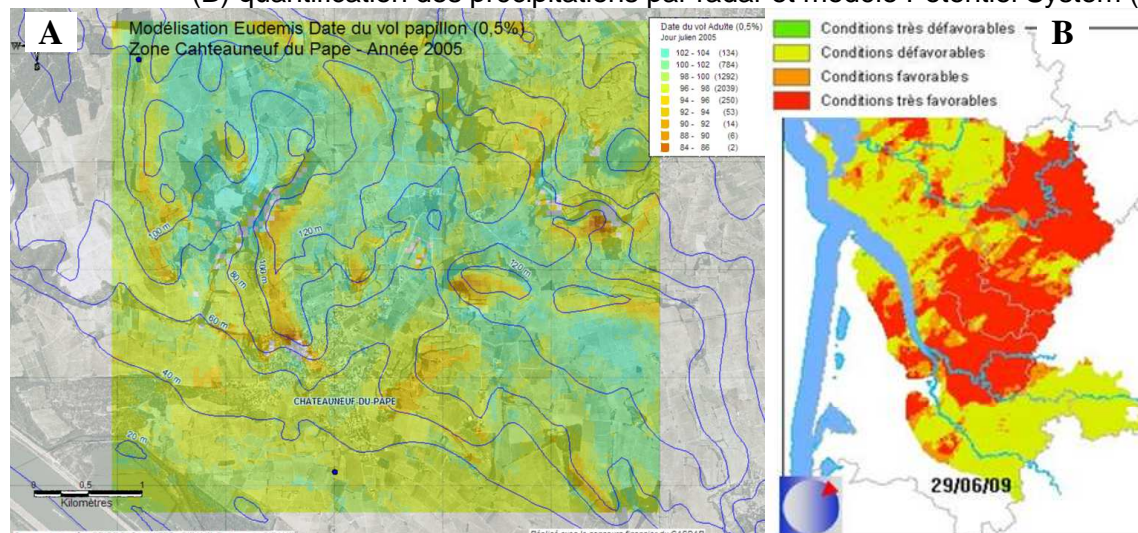


LES METHODES DE SPATIALISATION

SPATIALISATION DES DONNEES METEOROLOGIQUES ET DES RESULTATS DE MODELES

Les données météorologiques utilisées dans les modèles pour la protection des cultures peuvent être estimées en d'autres lieux que les stations de mesure. Pour cela, il existe différentes méthodes d'interpolation. Une méthode d'interpolation géostatistique bien connue est le krigeage qui permet, à partir de données dispersées, d'obtenir une représentation homogène de l'information en prenant en compte la corrélation spatiale. Ces méthodes peuvent prendre en compte le relief ou non et mobiliser aussi des données issues de radar pour quantifier finement les précipitation (une résolution inférieure au km²). Ainsi, pour une zone considérée, on obtient les données météo interpolées à chaque point pour chaque jour et il est alors relativement facile d'avoir une prédiction du modèle pour chacun des points de la zone (Figure 5, RMT modélisation, 2012).

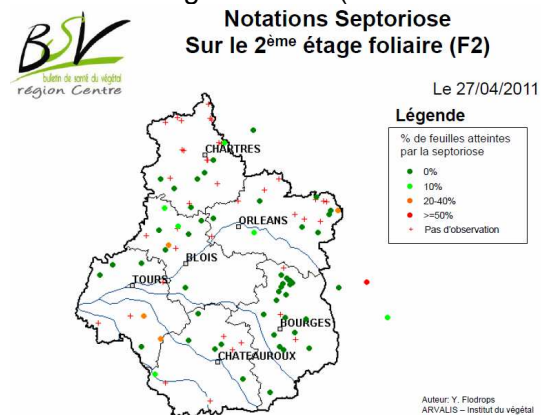
Figure 5 : Spatialisation des données météo puis des simulations. (A) prise en compte du relief sur les températures et modèle Eudemis en région PACA (CIRAME). (B) quantification des précipitations par radar et modèle Potentiel System (IFV).



DES RESEAUX D'OBSERVATIONS AVEC DES DONNEES EXPLOITABLES

Par ailleurs, des données issu des réseaux d'observation du système de Surveillance Biologique du Territoire sont disponibles (Figure 6).

Figure 6 : Représentation des données d'observation de Septoriose du blé à une date donnée en région centre (Arvalis - Institut du Végétal) issues de Vigicultures®.



Plusieurs systèmes informatiques centralisés de remontée, de validation et de stockage de ces informations existent. On peut citer Vigicultures® en grandes cultures (Arvalis - Institut du Végétal, Cetiom, ITB; Sine et al., 2010).

COMMENT COUPLER CES OBSERVATIONS ET LES SIMULATIONS SPATIALISEES ? : PRINCIPE D'UNE CORRECTION PAR SPATIALISATION

Pour l'exemple, prenons une zone d'intérêt carrée de 10km*10km avec une maille de 1km, on va avoir une matrice $N^2=10*10$ de données météo journalières interpolées (Meteo).

$$(s1) \quad Meteo(jour) = \begin{pmatrix} m_{1,1} & \dots & m_{1,10} \\ \vdots & \ddots & \vdots \\ m_{10,1} & \dots & m_{10,10} \end{pmatrix}$$

Avec $m_{i,j}$ le vecteur des données météo journalières jusqu'au jour pour la position (i, j).

Puis une matrice des prédictions du modèle pour ces mêmes points (Y).

$$(s2) \quad Y(jour) = \begin{pmatrix} Y_{1,1} & \dots & Y_{1,10} \\ \vdots & \ddots & \vdots \\ Y_{10,1} & \dots & Y_{10,10} \end{pmatrix}$$

Avec $Y_{i,j}$ le vecteur des prédictions du jour pour la position (i, j).

Par ailleurs, on peut avoir un réseau de k observations géolocalisées sur la zone considérée associé aux N mailles, mais qui ne correspondent pas forcément aux stations météo.

$$(s3) \quad M_{1..k}(jour) = (M_1, \dots, M_k) \quad \text{avec } k \ll N \text{ et chaque } M \text{ géolocalisée}$$

Pour combiner, ces observations et les simulations, nous proposons les étapes suivantes :

- 1) interpoler les données météo sur la zone considérée
- 2) simuler le modèle sur la zone considérée
- 3) calculer les erreurs de prédiction du modèle aux points d'observations

$$(s4) \quad \Sigma_{1..k}(jour) = M_{1..k}(jour) - Y_{1..k}(jour) = (\Sigma_1, \dots, \Sigma_k)$$

- 4) spatialiser les erreurs de prédiction par krigeage sur la zone considérée.

$$(s5) \quad \Sigma kri(jour) = \begin{pmatrix} \Sigma kri_{1,1} & \dots & \Sigma kri_{1,10} \\ \vdots & \ddots & \vdots \\ \Sigma kri_{10,1} & \dots & \Sigma kri_{10,10} \end{pmatrix}$$

- 5) corriger le modèle avec cette information (on peut évaluer la qualité de correction par validation croisée)

$$(s6) \quad Ycor(jour) = Y(jour) + \Sigma kri(jour) = \begin{pmatrix} Ycor_{1,1} & \dots & Ycor_{1,10} \\ \vdots & \ddots & \vdots \\ Ycor_{10,1} & \dots & Ycor_{10,10} \end{pmatrix}$$

- 6) représenter spatialement ces valeurs.

CONCLUSION

Ce papier avait pour objectif de présenter et d'illustrer les potentialités des principales pistes pour combiner simulations de modèles et observations dans le cadre des dispositifs de Surveillance Biologique du Territoire et de l'édition des BSV.

Les deux pistes présentées semblent des voies complémentaires de combinaison de résultats de modélisation et d'observation : la première permet de mieux prédire la dynamique temporelle et la seconde permet de prendre en compte la structure spatiale des erreurs du modèle afin de corriger les prédictions au niveau d'une région pour des lieux sans d'observation.

Il reste désormais à appliquer ces méthodes sur des cas réels du BSV et de lever les difficultés méthodologiques supplémentaires de mise en œuvre (nature des sorties des modèles et des observations, homogénéité du réseau d'observation, maillage en données météo, ...). Ainsi, dans certains cas, les sorties (par exemple, un indice de risque) du modèle sont d'une nature très différentes des données disponibles (par exemple une sévérité) : il faudra alors expliciter et formaliser la relation entre le modèle et les données, ce qui n'est pas toujours évident. Le problème de l'homogénéité du réseau d'observation est une autre difficulté : il faudra parfois regrouper différents types de données (variétés différentes par exemple) afin de pouvoir exploiter les données existantes.

Si l'amélioration des qualités de prédiction est significatif, il faudra aussi évaluer leur caractère opérationnel dans le cadre relativement contraint de l'édition des BSV (édition hebdomadaire avec peu de temps disponible,...). A une autre échelle de temps, certaines améliorations pourraient être disponibles pour la campagne suivante.

REMERCIEMENTS

Nous remercions l'ensemble des participants au groupe de travail « Mieux profiter de la synergie entre réseau d'observations, expertise et modélisation pour l'élaboration du Bulletin de Santé du Végétal » mené dans le cadre du Réseau Mixte Technologique Modélisation et Agriculture (www.modelia.org, volet 2. Synergies entre modélisation, expertise, expérimentation). Le RMT modélisation est financé par le « Compte d'affectation spécial pour le développement agricole et rural » du Ministère de l'Agriculture de l'Agroalimentaire et de la Forêt.

BIBLIOGRAPHIE

FNLON, 2009. Document de synthèse du projet CASDAR FNLON-APCA-ACTA « surveillance biologique du territoire ».

Kalman R. E. and R. S. Bucy, 1961. New results in linear filtering and prediction theory. Journal of Basic Engineering, 83 : 95-108.

Makowski D., Guérif M., Jones J.W., and Graham W., 2006. Data assimilation with crop models. In; Working with dynamic crop models: Evaluation, analysis, parameterization, and applications. Elsevier, Amsterdam. pp 151-172.

Makowski D., 2010. How to synthesize different types of knowledge using models? Journée internationale du RMT modélisation "What's new, what's next in dynamic system modeling in agronomy.", 7 septembre 2010.

Plan ECOPHYTO 2018, 2009. Plan ECOPHYTO 2018 - Fiches-Actions. Comité national d'orientation et de suivi Écophyto 2018. Version au 22 avril 2009

Petris G., 2010. dlm: Bayesian and Likelihood Analysis of Dynamic Linear Models. R package version 1.1-1, URL <http://CRAN.R-project.org/package=dlm>

RMT modélisation, 2012. Séminaire Mieux profiter de la synergie entre réseau d'observations, expertise et modélisation pour l'élaboration du Bulletin de Santé du Végétal. 9 mars 2012, Paris. <http://www.modelia.org/moodle/course/view.php?id=55>.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. <http://www.R-project.org>

Sine M., Morin E., Simonneau D., Brochard M., De Cosnac G., Escriou H, 2010. VIGICULTURES® – An early warning system for crop pest management. IAALD XIIIth World Congress organized by Agropolis International, 26-29 April 2010, Montpellier, France.