

Etablissement d'une carte de maladie à partir d'un échantillon

La connaissance de la localisation spatiale des cas de maladie permet d'augmenter la précision et l'efficacité des méthodes de lutte. Mais la cartographie exhaustive de l'incidence d'une maladie dans la population hôte est impossible en raison des contraintes de temps et de coût. Il est donc nécessaire d'utiliser des méthodes d'inférence de la distribution spatiale à partir d'un échantillon. Les méthodes de géostatistique comme le krigeage sont parfois utilisées, mais elles ne tiennent compte ni du processus épidémique à l'origine de la structure spatiale de la maladie, ni de l'hétérogénéité de la population hôte. La méthode décrite ci-dessous est basée sur un processus épidémique fonction des distances individuelles entre hôtes.

L'algorithme itératif consiste à minimiser une fonction objectif qui mesure l'ajustement aux données de la carte estimée. La procédure commence par la cartographie exhaustive de tous les sites et par la détermination de l'état (sain ou infecté) d'un échantillon aléatoire de sites. Une probabilité de maladie est affectée à chaque site : 1 pour les sites échantillonnés infectés, 0 pour les sites échantillonnés sains, et une probabilité tirée dans une loi uniforme $[0, 1]$ pour les sites non échantillonnés. La valeur initiale de la fonction objectif est calculée pour cette carte. Puis à chaque pas de l'algorithme un site i est choisi par un tirage équiprobable dans les sites non échantillonnés et

sa probabilité de maladie est mise à jour selon la fonction $P_i = 1 - \exp\left[-\sum_{j \neq i} P_j \exp(\theta - \mu d_{ij})\right]$, où θ

est un paramètre de transmission, μ un paramètre de dispersion, et d_{ij} la distance entre les sites i et j . Si la fonction objectif recalculée avec la probabilité P_i est inférieure à sa valeur au pas précédent, la mise à jour de P_i est acceptée, sinon elle est refusée (une règle de décision basée sur un tirage aléatoire peut être utilisée pour éviter les minimums locaux). L'algorithme se termine lorsqu'aucun changement significatif de la fonction objectif n'a été obtenu après un nombre de pas déterminé. Les paramètres θ et μ sont estimés au cours des itérations par essais et sélection de valeurs qui minimisent la fonction objectif.

La fonction objectif mesure la différence entre les probabilités P_i estimées pour les sites échantillonnés et leurs vraies valeurs p_i (0 ou 1). Cette fonction est la déviance

$$D = \frac{2}{n} \sum_{i=1}^n \left[p_i \log \frac{P_i}{p_i} + (1 - p_i) \log \left(\frac{1 - P_i}{1 - p_i} \right) \right], \text{ où la somme est calculée sur les } n \text{ sites}$$

échantillonnés.

La méthode a été testée sur des relevés exhaustifs effectués sur des parcelles de citrus infectées par le Huanglongbing en Floride. Des échantillons aléatoires simulés sur ces données ont été utilisés pour ré-estimer l'ensemble des données et comparer les cartes estimées avec les vraies cartes. Ces simulations produisent des cartes visuellement similaires aux vraies cartes. Les estimations sont en moyenne meilleures que celles obtenues par krigeage pour des taux d'échantillonnage de 20% ou supérieurs. Elles sont au moins égales pour des taux d'échantillonnage inférieurs à 20%.

La méthode initiale a par la suite été améliorée. Des mesures d'incertitude sur les probabilités de maladie à chaque site ont été développées. Le calcul a été accéléré par une initialisation plus précise des probabilités de maladie des sites non échantillonnés.

W. Luo, S. Pietravalle, S. Parnell, F. Van den Bosch, T. R. Gottwald, M. S. Ireby, and S. R. Parker (2012). An improved regulatory sampling method for mapping and representing plant disease from a limited number of samples. *Epidemics*, 4(2):68–77.

S. Parnell, T. R. Gottwald, M. S. Ireby, W. Luo, and F. van den Bosch (2011). A Stochastic Optimization Method to Estimate the Spatial Distribution of a Pathogen from a Sample. *Phytopathology*, 101(10):1184–1190.