



# A framework based on generalised linear mixed models for analysing pest and disease surveys



Lucie Michel <sup>a</sup>, François Brun <sup>b</sup>, David Makowski <sup>c, \*</sup>

<sup>a</sup> ACTA, INRA, UMR Agronomie, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France

<sup>b</sup> ACTA, INRA, UMR AGIR, F-31326 Castanet Tolosan, France

<sup>c</sup> UMR Agronomie, INRA, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France

## ARTICLE INFO

### Article history:

Received 17 February 2016

Received in revised form

28 November 2016

Accepted 9 December 2016

### Keywords:

Alert system

Bayesian model

Disease survey

Generalised linear mixed model

Plant pest

## ABSTRACT

In several countries, regional surveys are carried out to detect the presence of pests and diseases in crops. During these surveys, the incidence of major diseases and the presence of pests are recorded on various dates during the growing season. In this study, we aim to develop a framework to make better use of these regional surveys to estimate pest and disease dynamics, to analyse their variability across sites and years, and to assess uncertainty. Our framework is illustrated in four case studies: Septoria leaf blotch on wheat, downy mildew on grapevine, yellow sigatoka on banana and weevils on sweet potato. We showed that frequentist and Bayesian generalised linear mixed models gave similar results. This type of models is flexible enough to handle different types of data. They can be used to estimate disease and pest dynamics from observations collected in regional surveys and could help regional extension services evaluate risk levels at the regional scale.

© 2016 Elsevier Ltd. All rights reserved.

## 1. Introduction

Plant pest and disease alert systems provide farmers and their advisers with information about plant pest situations. They are designed to help farmers to manage their crop protection practices so as to minimise yield loss and to avoid unnecessary chemical treatments. They are often based on regional annual surveys for major pests and diseases, which have been conducted in several countries for decades. Pest and disease surveys have been performed since 1912 in Kansas (USA (Sim IV et al., 1998)), and since the 1970s in UK (King, 1977; Polley and Thomas, 1991). Pest and disease surveys are now frequently carried out in most American and European countries (Sigvald, 2012; Sine et al., 2010; Verreet et al., 2000) and in Asian countries, including India, Indonesia, Malaysia, and the Philippines (Talukder, 2000).

In France, regional pest and disease surveys became mandatory in 2009 when the government introduced the “Ecophyto” plan to reduce pesticide uses (MAAF, 2008). This plan was updated in 2015 (MAAF, 2015), and its main objective is to detect, identify and track emerging plant health risks. In France, regional surveys are now carried out annually, to detect the presence of pests and diseases in

major crops. During these surveys, the incidence of major diseases and the presence of pests are recorded at different dates (monthly or weekly) during the growing season. Results are published weekly or monthly in plant health bulletins that are distributed free-of-charge to farmers. For example, in 2014, around 13,000 fields were monitored in France and more than 3350 plant health bulletins were published and distributed to farmers (DGAL-SDQPV, 2014).

French plant health bulletins currently summarise survey results solely by presenting simple descriptive statistics (mean, minimum, and maximum values of the collected data) and graphics (histogram, pie chart). No statistical analysis is performed to estimate pest and disease dynamics or to compare current infestation levels with those in previous years. Another limitation of the current version of plant health bulletins is that they do not formally analyse uncertainty in the results of pest surveys.

According to Van Maanen and Xu (2003), models can serve several purposes such as predicting disease infection, estimating the frequency of epidemics, and comparing the performance of different disease management strategies. Mathematical models of plant disease epidemics are useful tools for simulating disease dynamics as a function of local weather conditions, cultivar characteristics, and farmers' practices (Apel et al., 2003; El Jarroudi et al., 2009; Gouache et al., 2012; Penaud et al., 2011; Whish

\* Corresponding author.

E-mail address: [David.Makowski@inra.fr](mailto:David.Makowski@inra.fr) (D. Makowski).

et al., 2015). They usually include however, a high number of input variables and can only be run on a limited number of sites because of the limited availability of input data. For example, in the Champagne-Ardenne region (a major wheat-producing region in France), the mathematical model SeptoLIS (Gouache et al., 2012) is currently run on only four sites by the regional extension service for predicting the dynamics of Septoria leaf blotch (SLB) in wheat crops (CRACA, 2012). Although mathematical models of plant disease epidemics provide useful information to farmers, they cannot be easily used to summarise regional survey data collected by extension services on a large number of sites.

Statistical models have an important role to play because they can be easily fitted to regional survey incidence data with updating as new data become available. We propose here to make better use of regional surveys to estimate pest and disease dynamics, to analyse their variability across sites and years, and to assess uncertainty. We have developed a framework for this purpose based on generalised linear mixed models (glmm). These models are flexible enough to handle both incidence data and count data, and to deal with longitudinal measurements, such as incidence (or count) measurements collected for the same site on different dates (Agresti, 2002 Chap.12). They can easily be fitted to data collected in pest and disease surveys, and can be updated whenever new data become available. Glmm can be used in different ways. They can be used to estimate mean (or median) pest and disease dynamics over a population of sites and years. They can also be used to estimate dynamics from local observations for specific sites and years.

The use of glmm has been promoted in ecology and in agricultural science as a method for analysing nonnormal data such as counts or proportions (Bolker et al., 2009; Gbur et al., 2012; Stroup, 2015), but these models have not been frequently applied in phytopathology (Kriss et al., 2012; Madden et al., 2002; Makowski et al., 2014; Piepho, 1999). Glmm can be fitted to data by frequentist or Bayesian statistical methods (Fong et al., 2010; Makowski et al., 2014; Zhao et al., 2006), but results obtained with these two statistical approaches have been rarely compared in phytopathology. We demonstrate here the utility of using glmm for analysing the pest and disease survey data used in alert systems, and we compare the results obtained with frequentist and Bayesian methods. We expand on the glmm-based work of Kriss et al. (2012) who were concerned with single-time surveys for disease incidence and a small number of linear covariables. Here we are concerned with surveys at multiple times in epidemics and a more complex set of covariables in order to characterize risk over time.

Our framework is illustrated in four case studies covering different crops and pests: (1) *Septoria* leaf blotch (SLB) caused by the fungus *Zymoseptoria tritici* (Fuckel) on wheat (Eyal et al., 1987), (2) downy mildew caused by the oomycete *Plasmopara viticola* (Berk. & Curt.) on grapevine (Jermine et al., 2010), (3) yellow sigatoka caused by the fungus *Mycosphaerella musicola* R. Leach ex. J.L. Mulder on banana (Fouré, 1994), and (4) weevils (*Cylas formicarius*, Coleoptera) on sweet potato (Denon and Mauléon, 2004). These pests and diseases were selected here because they can lead to important yield losses and because regional surveys are conducted in several regions in France to monitor their dynamics every year.

## 2. Materials and methods

### 2.1. General framework

We aimed to fit glmm to data collected in regional surveys for monitoring the levels of a given pest or disease over the growing season in the region concerned. In the type of survey considered here, a pest or disease species is observed in  $N$  different site-years. Depending on the pest/disease species, the observations recorded

( $Y$ ) may correspond to binary data (yes/no, i.e. pest presence/absence), disease incidence (proportion of diseased plants or organs), or count data (pest abundance, e.g., number of insects counted in traps). Observations are collected at each site-year on different dates during the growing season. Several observations are thus available in each site-year. In some surveys, additional data concerning sowing date or crop cultivar, for example, may be collected to characterize the site-years surveyed.

Our framework includes the following steps:

- (i) Definition of a glmm. The characteristics of this model should be specified, taking into account the nature of the observations (binary data, disease incidence, or count data) collected in the regional survey considered.
- (ii) At time  $t$ : fitting of the model defined in step (i) to the data collected up to time  $t$ . Model parameters were estimated using either a frequentist method (maximum likelihood) or a Bayesian method (implemented with a Markov chain Monte Carlo algorithm). Details about the fitting procedures are given in 2.4.
- (iii) Use of the fitted model to estimate pest/disease dynamics up to time  $t$ , and to predict future changes in the pest/disease epidemic up to time  $t+\Delta t$ , where  $\Delta t$  represents a time interval (e.g., week, month).
- (iv) Calculation of confidence or credibility intervals for analysis of the uncertainty of the estimated and predicted dynamics.

Steps (ii) to (iv) can be repeated whenever new data become available during the growing season. For example, the model can be first fitted using data collected before time  $t$ , and used to predict pest/disease dynamics between  $t$  and  $t+\Delta t$ . If new data are collected between  $t$  and  $t+\Delta t$ , the model can be fitted to the new data, and then used to predict pest/disease dynamics after  $t+\Delta t$ . Estimated and predicted dynamics can be displayed graphically, as shown in section 3.

### 2.2. Model specification

The glmm defined in step (i) is expressed as:

$$g(\mu_{ij}) = \alpha_i + \beta_i t_{ij} + \sum_{k=1}^K \gamma_k z_i^{(k)} \quad (1)$$

where  $\mu_{ij}$  is the conditional expected value of the  $j$ th observation  $Y_{ij}$  collected in the  $i$ th site-year at time  $t_{ij}$ ,  $\alpha_i$  and  $\beta_i$  are two random site-year-specific regression parameters (varying across site-years),  $z_i^{(k)}$ ,  $k = 1, \dots, K$ , are  $K$  variables characterising the  $i$ th site-year (e.g., risk level defined from sowing date and cultivar resistance), and  $\gamma_k$  are the  $K$  regression parameters associated with  $z_i^{(k)}$ ,  $k = 1, \dots, K$ .

In Eq. (1),  $\mu_{ij}$  is related to the model input variables via a function  $g$  called the *link function*. Different link functions should be used for the different types of observations ( $Y$ ) available. When  $Y_{ij}$  is a binary variable or corresponds to a proportion of infected plants, the logit function is a popular choice (Agresti, 2002 Chap.4; Gbur et al., 2012; Stroup, 2015). In this case, the function  $g$  is expressed as  $g(\mu_{ij}) = \log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right)$ . When  $Y_{ij}$  corresponds to count data, the log function is often appropriate,  $g(\mu_{ij}) = \log(\mu_{ij})$  (Agresti, 2002 Chap.4).

Full specification of the model requires definition of the conditional distribution of the observations, i.e., the distribution of  $Y_{ij}$  conditional to  $\mu_{ij}$ , i.e.,  $P(Y_{ij}|\mu_{ij})$ . Again, the choice here depends on

Download English Version:

<https://daneshyari.com/en/article/5760916>

Download Persian Version:

<https://daneshyari.com/article/5760916>

[Daneshyari.com](https://daneshyari.com)