Spatiotemporal Modeling of a Spread of Apple Scab

Rémi Créte [1]
(1) remi.crete@etud.univ-angers.fr

07 septembre 2010

Thesis :

Taking into account the spatial heterogeneity in modeling the dynamics of development of apple scab.

Thesis Supervisor : Besnik PUMO

With funding from the region Pays-de-Loire

MODEMAVE Collaborative project :

INRA
Angers France

LAREMA
Angers University
France

Collaborators :

V. Caffier[2], F. Didelot[2], B. Pumo[1],[3], P. Santagostini[3]

(1) LAREMA, UMR 6093, 2 Boulevard Lavoisier, 49045 Angers, France

(2) INRA, Centre d'Angers-Nantes, UMR77, IFR149, 42 rue Georges Morel, F-49071 Beaucouzé, France

(3) Agrocampus Ouest - Centre d'Angers, 2 rue Le Nôtre, 49045, Angers, France

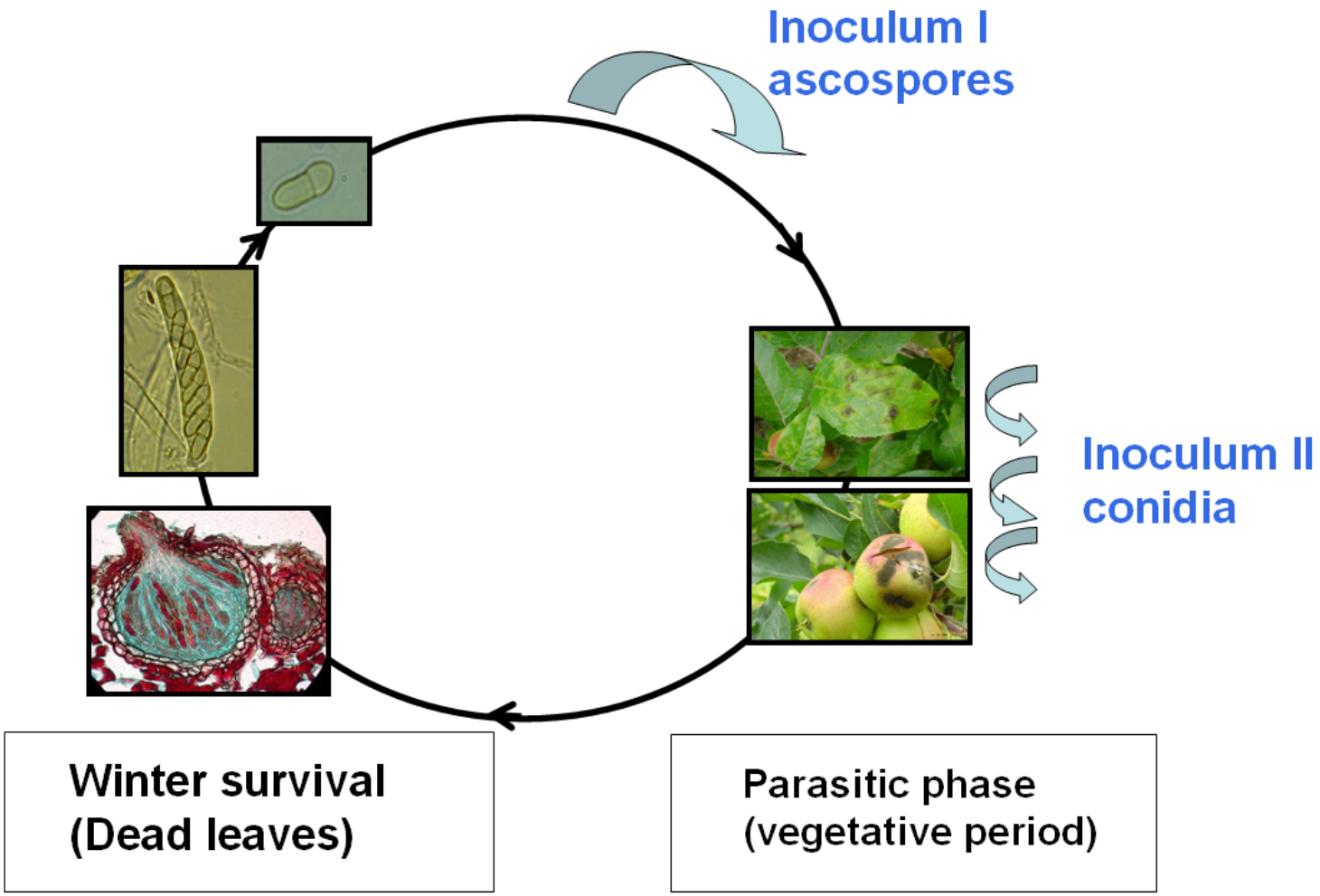
Plan

I - Motivation of this work and study framework

Apple Scab

- Caused by Venturia Inaequalis.
- Affected fruits are excluded from the sale.

Life cycle of Venturia inaequalis



Control measures

- Removing leaf litter and trimmings containing infected tissue from the orchard and incinerating them.
- Chemical controls including a great variety of compounds (fungicides made from benzene, copper, wettable sulfur), 10-20 times a year.
- To mix the susceptible variety with an other, resistant, inside the same orchard.

Problems

- Leaf litter can't be totally removed.
- The current context of sustainable development imposes limits on the use of chemical fungicides by 2018.
- It happens that resistant varieties become circumvented, and then behave like susceptible ones.

Solutions ?

- To adress this problem of circumvention, new resistance factors are under selection in breeding programs of apple, and should result in the future to create new varities resistant to scab.
- The effectiveness and sustainability of these new resistance should be enhanced by the simultaneous deployment of several varieties resistant to the scale of the orchard.
- Indeed, the varietal mixtures (mixed cultivars with different levels of susceptibility in the same orchard) can significantly reduce the impact of the disease, their effectiveness depending on the mode of dispersal of the pathogen.

Biological questions

- **What are the terms of dispersion of virulent strains over time within a plot ?**
  - **Depending on :**
    - the type of plot : pure or mixed : (spatial homogeneity)
    - the type of inoculum : I or II, and the period of the year (temporal heterogeneity)
- **What are the influences between plots on the development of virulent strains ?**

  (spatial and temporal heterogeneity)

Experimental protocol

Bypassing of the variety Ariane *(Vf)* :

- **INRA orchard area including two types of plots :**
  - 3 pure plots
  - 3 mixed plots

- **Before 2004: No disease on Ariane trees.**
- **In 2004: Appearance of a disease outbreak.**

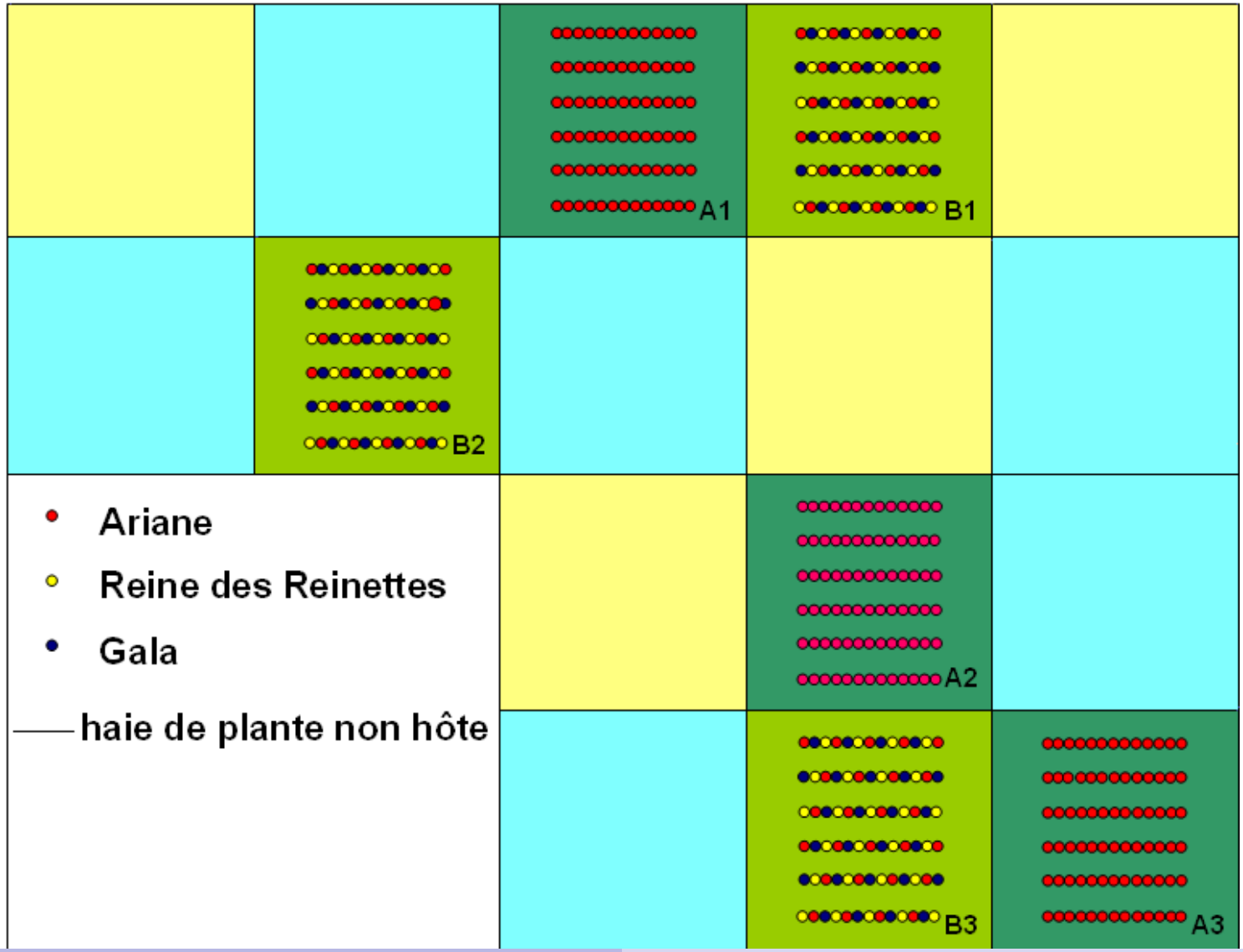
→ Study of spatio-temporal development of scab on Ariane :

Notation of the disease on all Ariane trees

- From 2004 to 2007 (mixed plots)

- From 2007 to 2008 (pure plots)

Experimental orchard



Severity scale

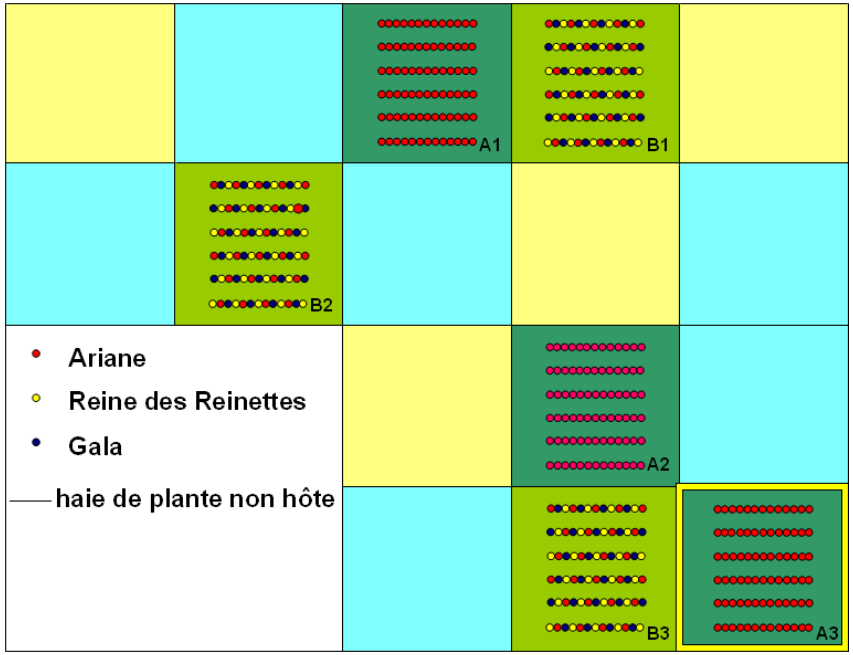
Note of illness	% scabbed leaves (sporulating)
1	0 %
2	0 to 1 %
3	1 to 5 %
4	*Intermediate state*
5	± 25 %
6	*Intermediate state*
7	± 50 %
8	± 75 %
9	> 90 %

Healthy tree

Diseased tree

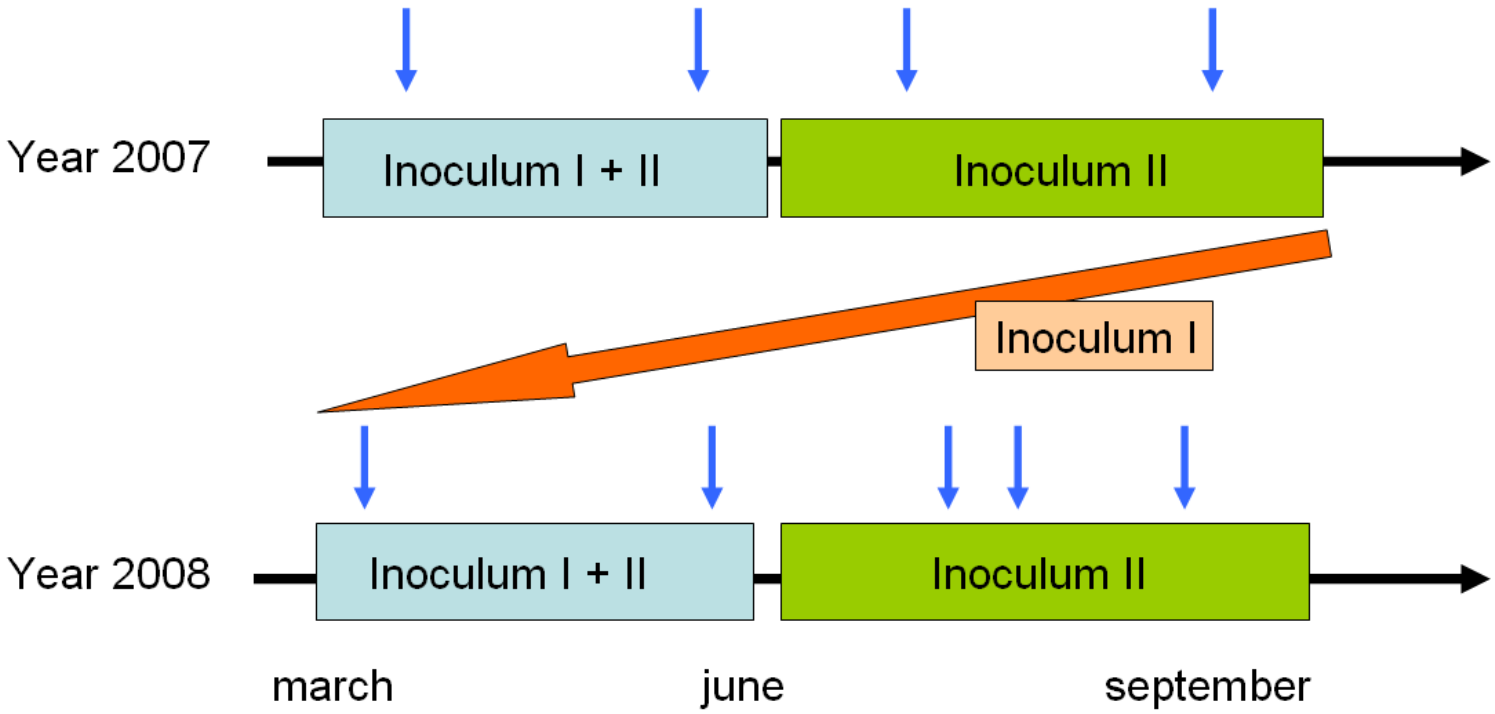
II - Statistical description of a plot

Statistical description of the plot A3

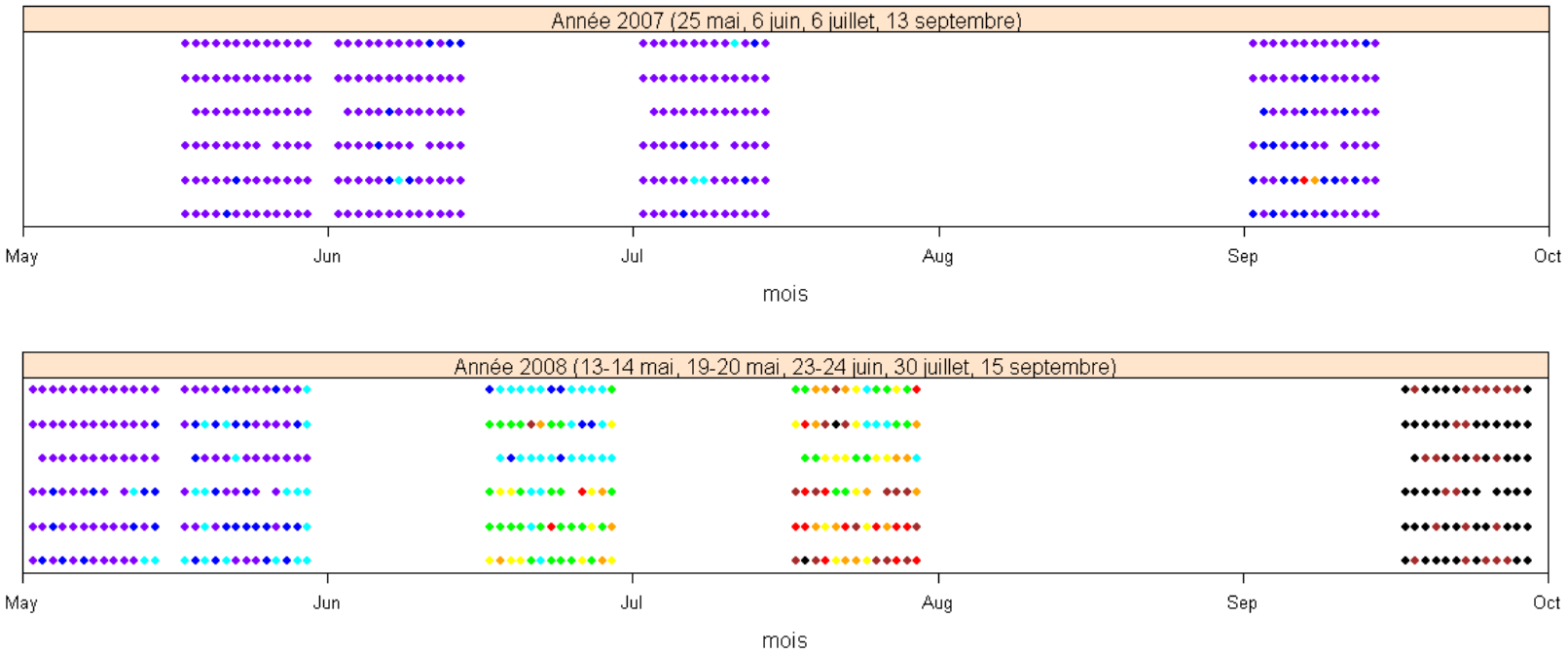


- This is a pure plot including infection with scab is most marked.
- In the case of mixed plots, the development of the disease is different, and less in terms of severity.
- The covered years are 2007 and 2008 since it is a pure plot.

Dates of measurements (2007 - 2008)

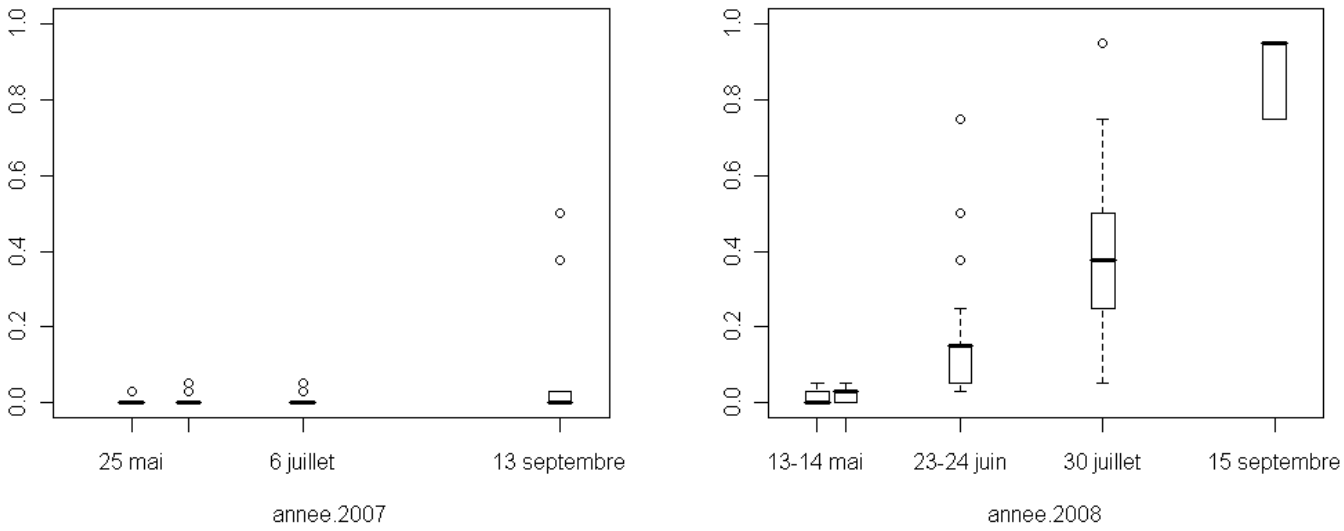


Severity data (1)



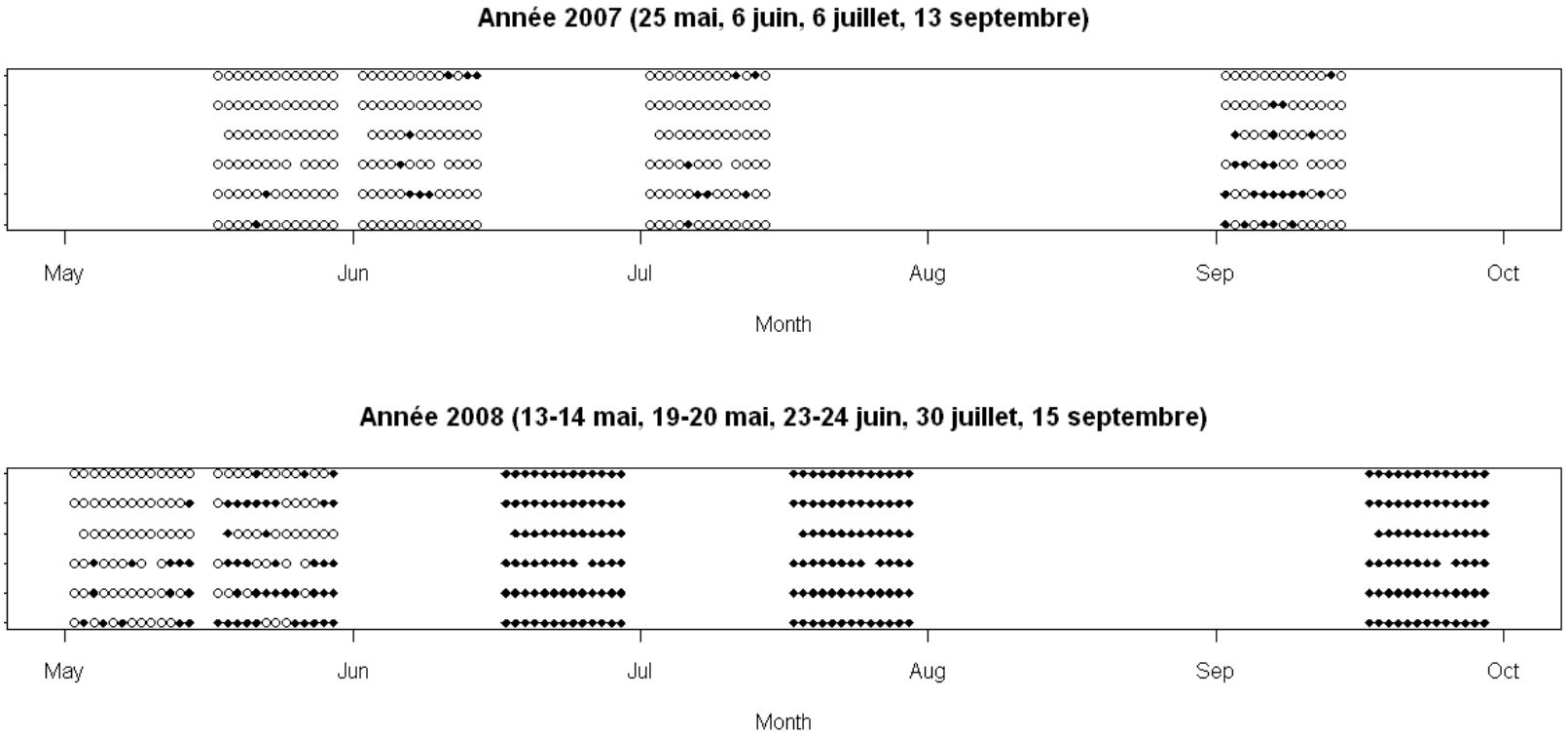
Evolution of the epidemic during 2007 and 2008

Severity data (2)



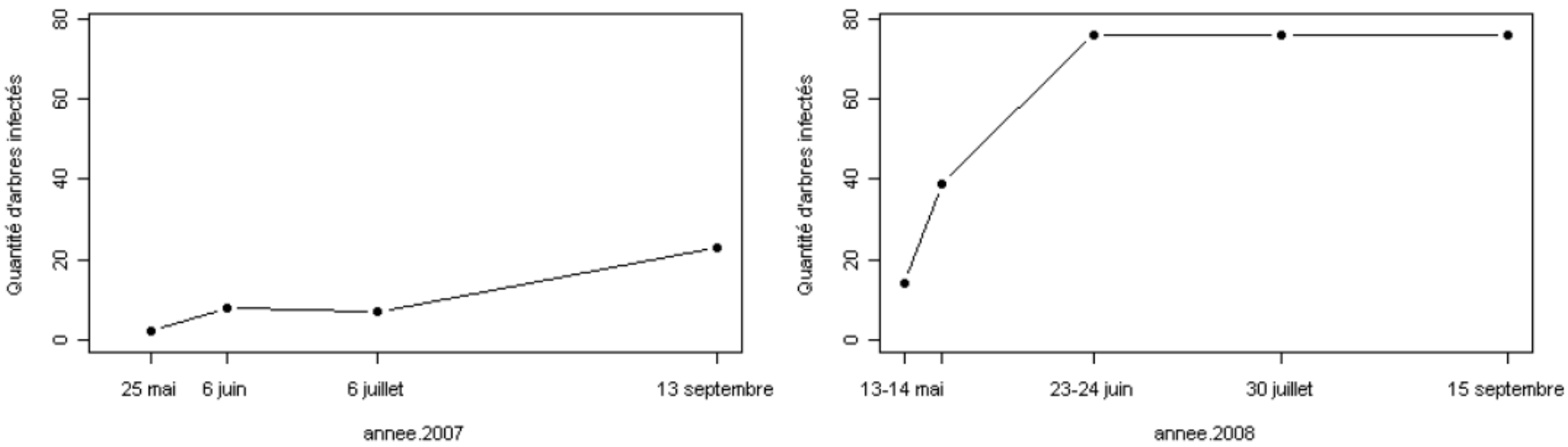
Evolution of the epidemic during 2007 and 2008

Binary data (1) :



Evolution of the epidemic during 2007 and 2008

Binary data (2) :



Number of infected trees 2007-2008

Conclusions

Large differences in behavior of the plot between 2007 and 2008 :

- In terms of number of infected sites
- In terms of states of these sites.

The analysis by year seems most appropriate :

- A model taking into account only the phenomenon of disease occurrence for 2007.
- A model taking into account the severity in 2008.

III - A Model for disease occurrence

Why this kind of model ?

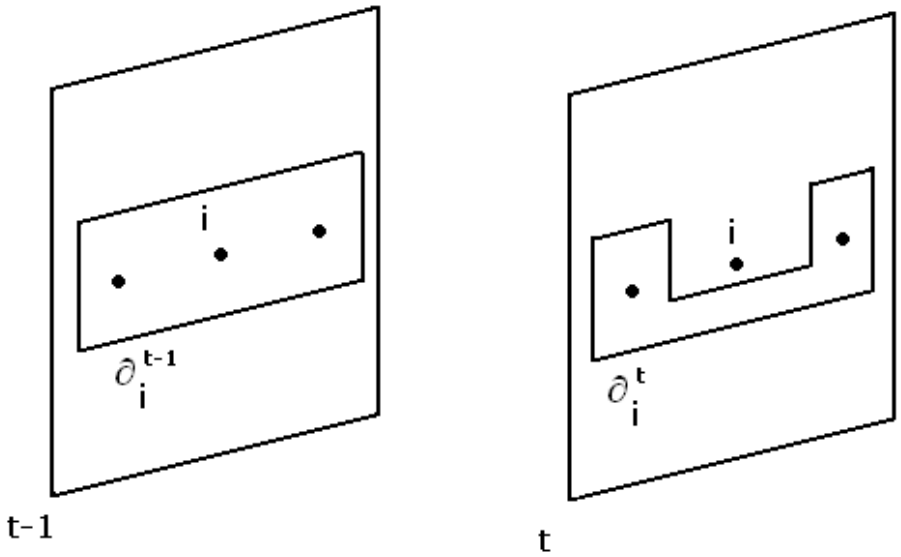
- Using classic for decades in many areas (ecology, meteorology, epidemiology ...)
- Advantage of simplifying data relevant here in 2007 : The trees are very slightly contaminated.
- This phenomenon is available for all other plot of the orchard.
- Study of a plot between two consecutive time : spatial and temporal homogeneity.
- 2007 is the only year where we have the data of all plots (pures and mixed) : This model is necessary for taking into account the spatial heterogeneity.

The dynamic autologistic model :

- Binary spatiotemporal model.
- Local spatial dependence of sites conditionally to the other sites of the field (Markov Field).
- Local time dependence of sites conditionally to the other sites of the field (Markov Chain of Markov Field).
- By the Hammersley-Clifford's theorem, this dependence is reflected in the existence of a neighborhood graph for the sites.

Spatiotemporal graph : Case of the two nearest neighbours

This amounts to the study of sites placed on a line.



We got explicitly :

$$\partial_{\mathbf{i}}^{\mathbf{t}} = \{i-1, i+1\}$$

$$\partial_{\mathbf{i}}^{\mathbf{t-1}} = \{i-1, i, i+1\}$$

Analytical expression

Temporal transition of the field :

For two states x and y of the field, consecutives in time, the transition is given by :

$$P(x, y) = Z^{-1}(x).exp\{U(y \mid x)\}$$

- where Z is the normalizing constant of P
- and where energy U equals to :

$$U(y \mid x) = \sum_{i \in S} \alpha_i(x).y_i + \sum_{<i,j>} \beta_{ij}.y_i.y_j$$

Conditional law of a site :

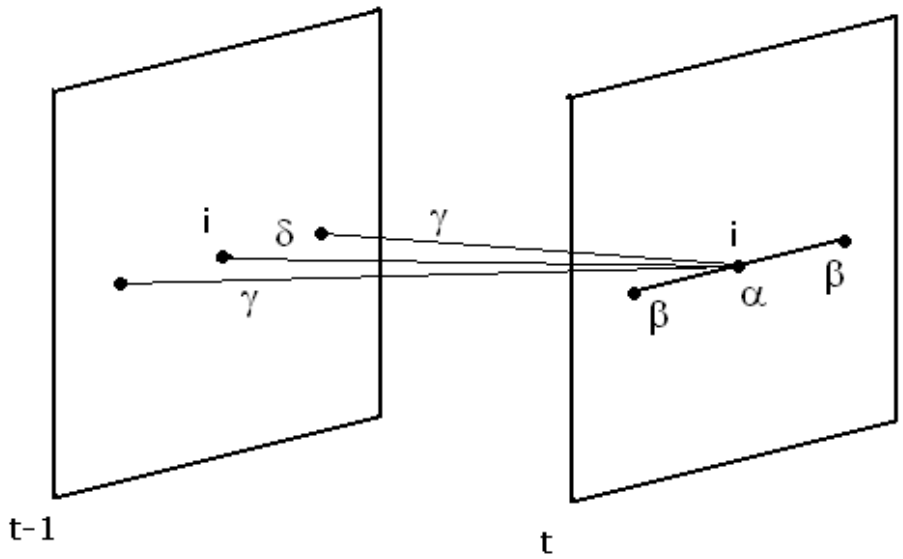
For two states x and y of the field, consecutives in time, the conditional transition is given by :

$$P_i(y_i \mid y^i, x) = C^{-1}(x).exp\left\{U(y_i \mid y^i, x)\right\}$$

- where C is the normalizing constant of P_i
- and where we have :

$$U(y_i \mid y^i, x) = y_i.[\alpha_i(x) + \sum_{<i,j>} (\beta_{ij}.y_j)]$$

Case of the two nearest neighbours :



$$U(y_i \mid y^i, x) = y_i.[\alpha + \delta.x_i + \sum_{<i,j>} (\beta.y_j + \gamma.x_j)]$$

- α and β are the instantaneous parameters
- γ and δ are the time delay parameters

Case of any neighborhood :

$$U(y \mid x) = \sum_{i \in S} \alpha_i(x).y_i + \sum_{<i,j>} \beta_{ij}.y_i.y_j$$

- $\alpha_i(x)$ is the first order potential in i :

$$\alpha_i(x) = \alpha + \delta.x_i + \sum_{l \in \partial_i^- \setminus \{i\}} \gamma_{i,l}.x_l$$

- β_{ij} is the second order potential of i and j.

Parameters estimation (1) :

- We search the parameter $\theta = (\alpha, \beta, \gamma, \delta)$ which maximize $P(x, y)$:

$$P(x, y) = Z^{-1}(x).exp\{U(y \mid x)\}$$

- Problem : The normalizing constant is often intractable.
- Possible solution : To replace $P(x, y)$ with a pseudo-likelihood function.

Parameters estimation (2) :

The parameter $\theta_0 = (\alpha_0, \beta_0, \gamma_0, \delta_0)$ which maximize the following expression is the conditional pseudo-likelihood estimator of θ :

$$LogCPL(\theta) = \sum_{i \in S} ln \left(P_i(y_i \mid y^i, x; \theta) \right)$$

Improvement of the model

Expression of potential in the general case :

- The network structure requires the use of a distance function $\varphi(i,j)$ between two sites at a given time :

$$\beta_{ij} = \varphi(i,j,\beta)$$

(isotropy of the interaction potential)

- The heterogeneity of the chain X requires finding a function of time difference *psi* acting on the parameters *delta* and *gamma* in the following expression :

$$\alpha_i(x) = \alpha + \psi(\delta).x_i + \sum_{l \in \partial_i^- \setminus i} \psi(\gamma).\varphi(i,l).x_l$$

Other perspectives :

In the homogeneous case :

- Using other methods of estimation (MCMC), Theoretical study of the parameters (asymptotic law of estimators and tests).
- Study of the severity by MCMF, extending binary case.
- Simulation of models on large fields.

In the heterogeneus case :

- Study of multiple plots, taking into account covariates.
- Consideration of the temporal heterogeneity in more detail.

The basis idea

We just want to study the state of a plot (in terms of severity), conditionally to the past :

$$P(Z_t \mid Z_{t-1}) = \prod_{i \in S} P(Z_{i,t} \mid Z_{t-1})$$

(Similarly to a HMCM)
That means :

- Trees evolve simultaneously, compare to the previous time.

Study of both occurrence and severity :

$$P([Z_t, X_t] \mid [Z_{t-1}, X_{t-1}]) = \prod_{i \in S} P([Z_{i,t}, X_{i,t}] \mid Z_{t-1})$$

And for the conditionnal law of a tree :

$$P([Z_{i,t}, X_{i,t}] \mid Z_{t-1}) = P(Z_{i,t} \mid X_{i,t}, Z_{t-1}).P(X_{i,t} \mid X_{i,t-1}, Z_{t-1})$$

This define the 2-stage model, where :

- The first stage is to model the occurrence of the disease on a tree, knowing this phenomenon at the previous time with the severity of the plot.
- The second stage is to model the severity of a tree, knowing its state of contamination (yes/no) and the state of severity of the plot at the previous time.

Epidemic assumptions

We make the following assumptions :

- The contaminated state is an absorbing state.
- No influence from the rest of the orchard.
- The quantity of spores released by an infected tree is proportionnal to its estimated level of contamination.

The infection potential

It represents the risk of a tree i to be contaminated by all the other trees , between $t-1$ to t (Mollison 1977) :

$$W_{i,t} = \sum_{j \neq i} z_{j,t-1}.p(i,j,\gamma)$$

where :

- $z_{j,t}$ is the severity state of j at time t.
- $p(i,j,\gamma)$ is the probability that a spore released by j is deposited on i.
- γ is the parameter of dispersion we want to determine.
  For instance, $p(i,j,\gamma) = \frac{1}{\gamma.\sqrt{2\pi}}.exp(\frac{|i-j|^2}{2\gamma^2})$

The first stage model

If $X_{i,t-1} = 1$, then $X_{i,t} = 1$.

$X_{i,t} \propto \mathcal{B}(1 - exp(-\alpha.W_{i,t}))$
where :

- α is a parameter depending on environmental condition.
- $\alpha.W_{i,t}$ is the expected quantity of spore which have been succeeded on establishing lesions on i.
- We assume here that the number of lesions on i, which were healthy at the previous time, follows a poisson law with means equals to $\alpha.Wi, t$.

The second stage model

If $X_{i,t} = 0$, then $Z_{i,t} = 0$.

$Z_{i,t} \propto Beta(\mu_i, \sigma)$

with μ_i its mean value, and σ its scale parameter :
$\mu_i \propto \beta^{(1)} + \beta^{(Y)}.Y_{i,t-1} + \beta^{(W)}.W_{i,t} + \beta^{(C)}.C_{i,t}$

and where :

- $C_{i,t}$ depends on environmental conditions (AIC selection of several covariates)

Bibliography

Besag, J.E.(1974) Spatial interaction and statistical analysis of lattice system. Journal of the Royal Statistical Society, Series B 76, 192-236.

Chadoeuf, J., Nandris, D., Geiger, J.P., Nicole, M., et Pierrat, J.C.(Décembre 1992). Modélisation spatio-temporelle d'une épidémie par un processus de Gibbs : Estimation et Tests. Biometrics 48, 1165-1175.

Soubeyrand S., Held L., Höhle M., Sache I. (2008). Modelling the spread in space and time of an airborne plant disease. Appl. Statist. 57, Part 3, 253-272.

Mollison D. (1977) Spatial contact models for ecological and epidemic spread, with discussion, J. Roy. Stat. Soc., B, 39, 283-326.

Caffier V., Didelot F., Orain G., Lemarquand A., Parisi L. (2009). Efficiency of association of scab control methods on resistance durability of apple : the case study of cultivar Ariane. IOBC wprs Bulletin, sous presse.

Guyon X., Pumo B. (2007) Space-time estimation of a particle system model. Statistics, 41 (5), 395-407.

Guyon X. (1995). Random Fields on a Network : Modeling, Statistics, and Applications. Springer-Verlag

Guyon X., Hardouin C. (2007) Markov Chain Markov Field Dynamics : Models and Statistics.

THANK YOU VERY MUCH