

Estimation des paramètres d'un modèle de prévision du risque septoriose

François Brun (ACTA)

David Gouache, Arnaud Bensadoun, Christian Pagé, David
Makowski, Daniel Wallach

R
 M
 T **Modélisation & Analyse de Données
pour l'Agriculture**

Journée Modélisation 11 décembre 2015

**Intégration de différentes sources de
connaissances pour valoriser nos données**

Intérêt des approches bayésiennes

Septoriose et CC



- Question : quel impact du CC sur les épidémies de septoriose ?

*Scénario GIEC 2007
A1B*

**Référence
1970–2000**



**Futur lointain
2070–2100**

État sanitaire « actuel »

État sanitaire « à venir »

Mais les incertitudes ?

De nombreuses sources...

Incertain
émissions des GES

Scénario GIEC 2007
A1B

Référence
1970–2000



Futur lointain
2070–2100

État sanitaire « actuel »

État sanitaire « à venir »

Incertain
biologiques
+ équations
+ paramètres

Incertain
climatiques
+ hypothèses
+ descente d'échelle

Proposition pour les incertitudes

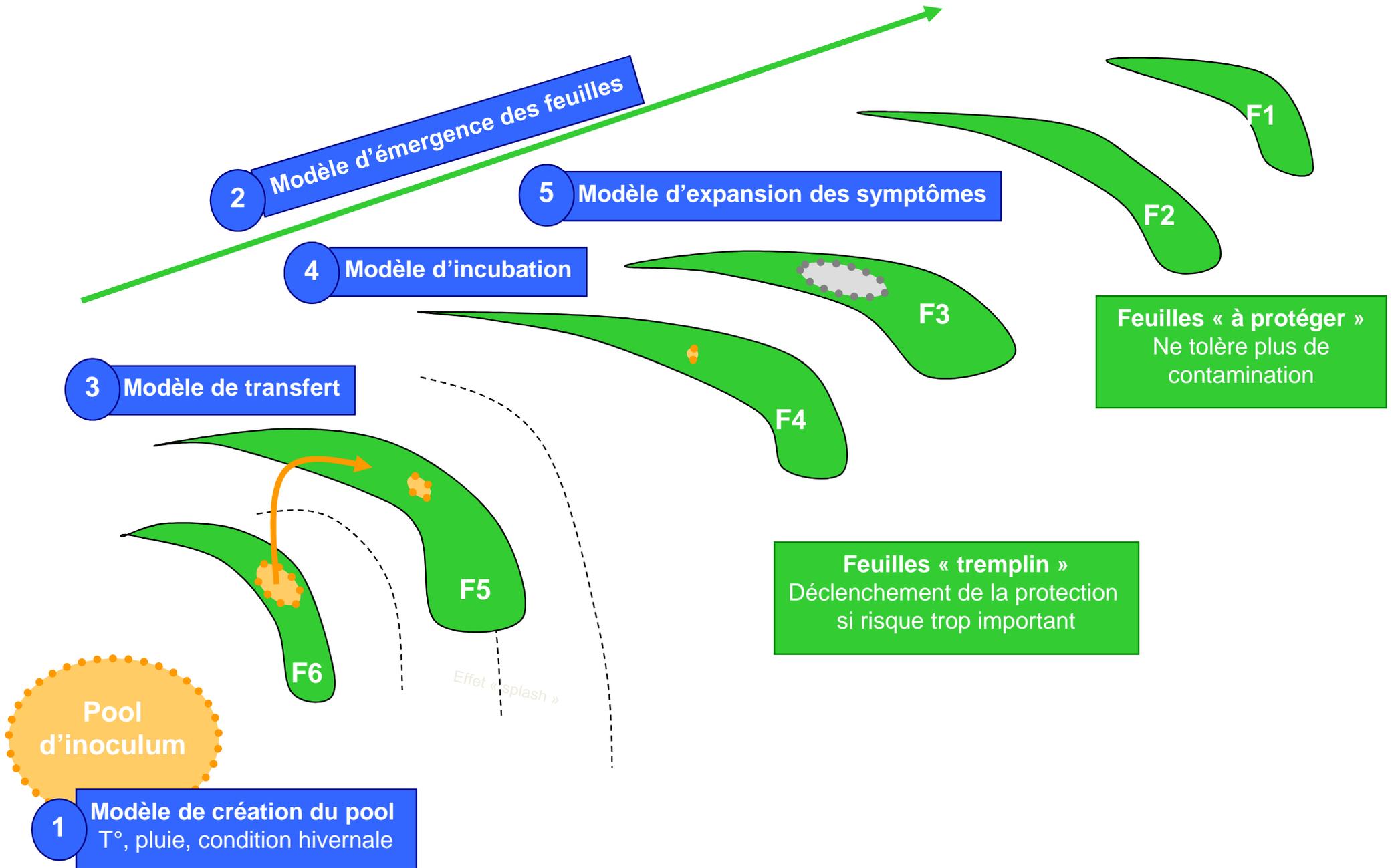
- Emission GES : **Un seul scénario considéré : A1B**
- climatiques
 - Modèles circulation globale associés à descente échelle
 - ⇒ **Utilisation de 5 modèles** (GISS, NCAR, CCCMA, MRI, ARPEGE)
- biologiques
 - Modèle Septo-LIS
 - ⇒ **Incertitude sur les paramètres de Septo-LIS**
 - ⇒ **Bayésien**



Démarche

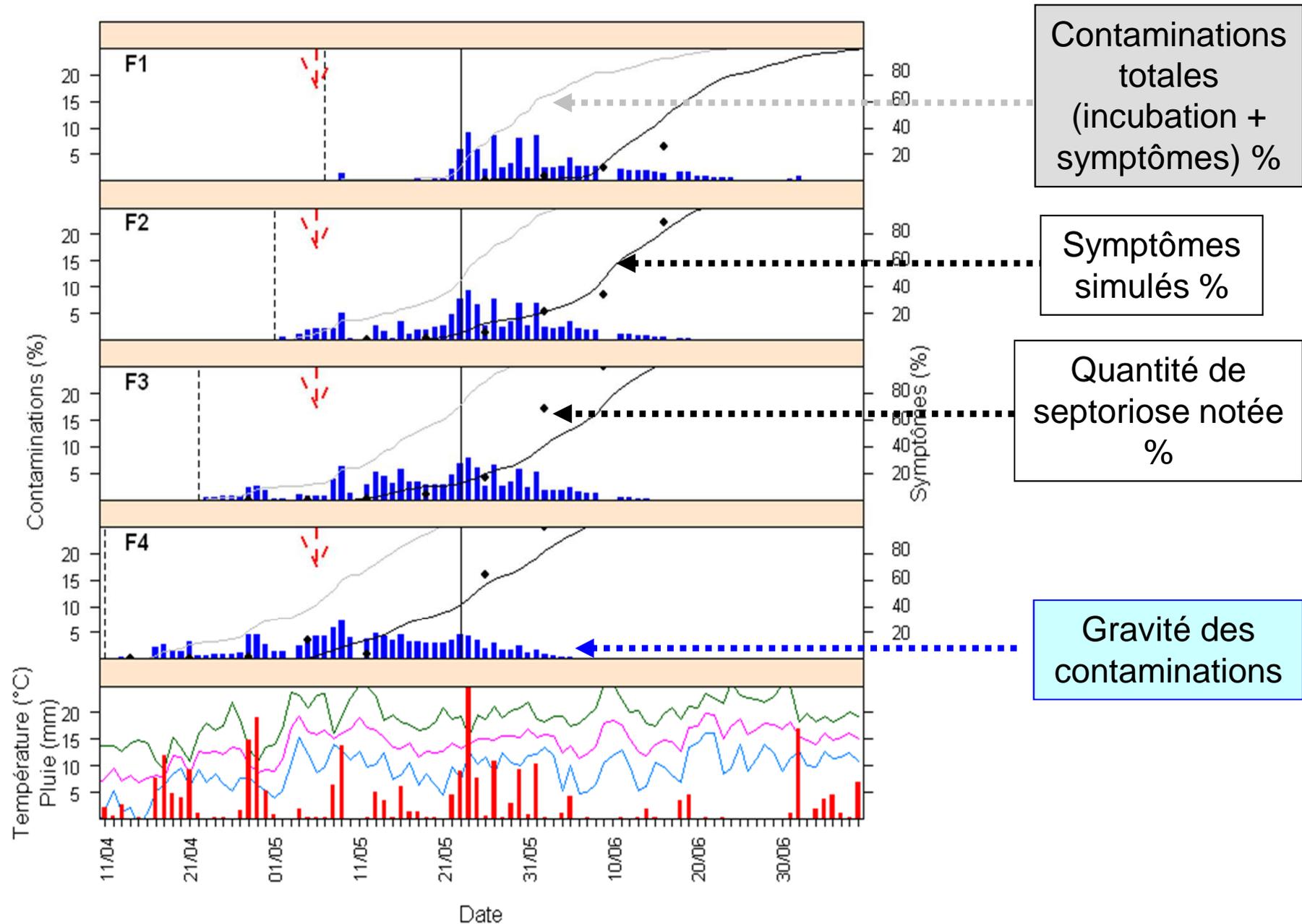
1. Modèle Septo-LIS décrivant la biologie de la septoriose
2. Estimation des paramètres avec leurs incertitudes
3. Projection CC en prenant en compte l'incertitude sur le modèle et le CC

Modèle épidémiologique sous Septo-LIS



Exemple de sortie

DINOSOR 12/11/2007 BIGNAN 5695



Démarche

1. Modèle Septo-LIS décrivant la biologie de la septoriose
2. Estimation des paramètres avec leurs incertitudes
3. Projection CC en prenant en compte l'incertitude sur le modèle et le CC

Estimation des paramètres

distribution *a posteriori* (théorème de Bayes)

$$f(\theta, \sigma^2 | Y^{\text{obs}}) \propto f(Y^{\text{obs}} | \theta, \sigma^2) \pi(\theta, \sigma^2)$$

vraisemblance *a priori*

$$y_i^{\text{obs}} = y_i + \varepsilon_i$$

$$\varepsilon_i \underset{\text{iid}}{\sim} N(0, \sigma^2)$$

avec

$$y_i^{\text{obs}} = \sin^{-1}(\sqrt{S_i^{\text{obs}}/100})$$

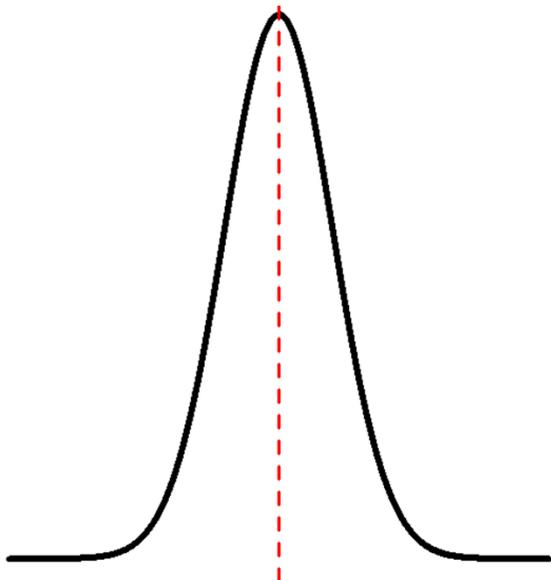
$$y_i = \sin^{-1}(\sqrt{S_i(\theta)/100})$$

Quelles informations *a priori* ?

- 17 paramètres : difficile d'avoir une information *a priori* à partir des experts du modèle.

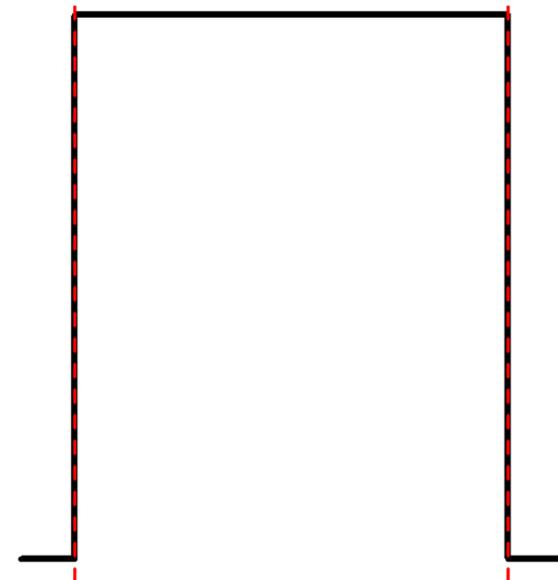
Soit valeur du paramètre provenant d'une estimation précédente (fréquentiste) à partir de données.

⇒ Distribution gaussienne



Soit valeurs des paramètres considérées comme mal connu, mais des bornes peuvent être définies.

⇒ Distribution uniforme



distributions *a priori*

Parameter	Units	Prior shape	Prior mean	Reference
θ_{mult}	% necrosis/% necrosis	Uniform	8.25	(1), (4)
$\theta_{\text{winter.th}}$	% necrosis	Gaussian	20.70	(4)
θ_{scale}	–	Uniform	50.50	(4)
$\theta_{\text{min.inoc}}$	% necrosis	Gaussian	5.50	(4)
$\theta_{T \text{ min}}$	°C	Gaussian	2.50	(4)
θ_{deplete}	% necrosis/% necrosis	Uniform	7.50	(4)
θ_{rain}	mm ⁻¹	Gaussian	-0.35	(1)
θ_{height}	cm ⁻¹	Gaussian	0.55	(1)
$\theta_{\text{prod.inoc}}$	% necrosis/% necrosis	Uniform	5.50	(1)
$\theta_{\text{expansion}}$	% necrosis/°C	Gaussian	0.047	(1), (2)
θ_{initial}	% necrosis	Gaussian	0.26	(1), (2)
$\theta_{\text{min.today}}$	°C	Gaussian	2.50	(4)
$\theta_{\text{min.yesterday}}$	°C	Gaussian	2.50	(4)
$\theta_{\text{min.rain}}$	mm	Gaussian	1.53	(1), (4)
$\theta_{\text{limit } F1}$	mm	Gaussian	7.50	(4)
$\theta_{\text{leaf.growth}}$	°C-days/°C-days	Gaussian	1.65	(3)
$\theta_{\text{rain.infect}}$	mm	Gaussian	7.50	(1), (4)
variance of residual error		Jeffrey		

Références : (1) Audsley et al. (2005); (2) Robert et al. (2008); (3) Li (2005); (4) Gouache (unpublished).

Les données

- Beaucoup de données disponibles,
 - des centaines de sites-années... (témoin non traité de différents essais)
- mais temps de calcul rédhibitoire

⇒ **Sélection de 8 sites années uniquement**

– 2000, 2001, 3 sites en 2003, 2006, 2007

(Calvados, Côte d'Or, Côte d'Armor, Morbihan, Seine et Marne, Seine Maritime, Somme)

– Observations répétées à différentes dates

Méthode

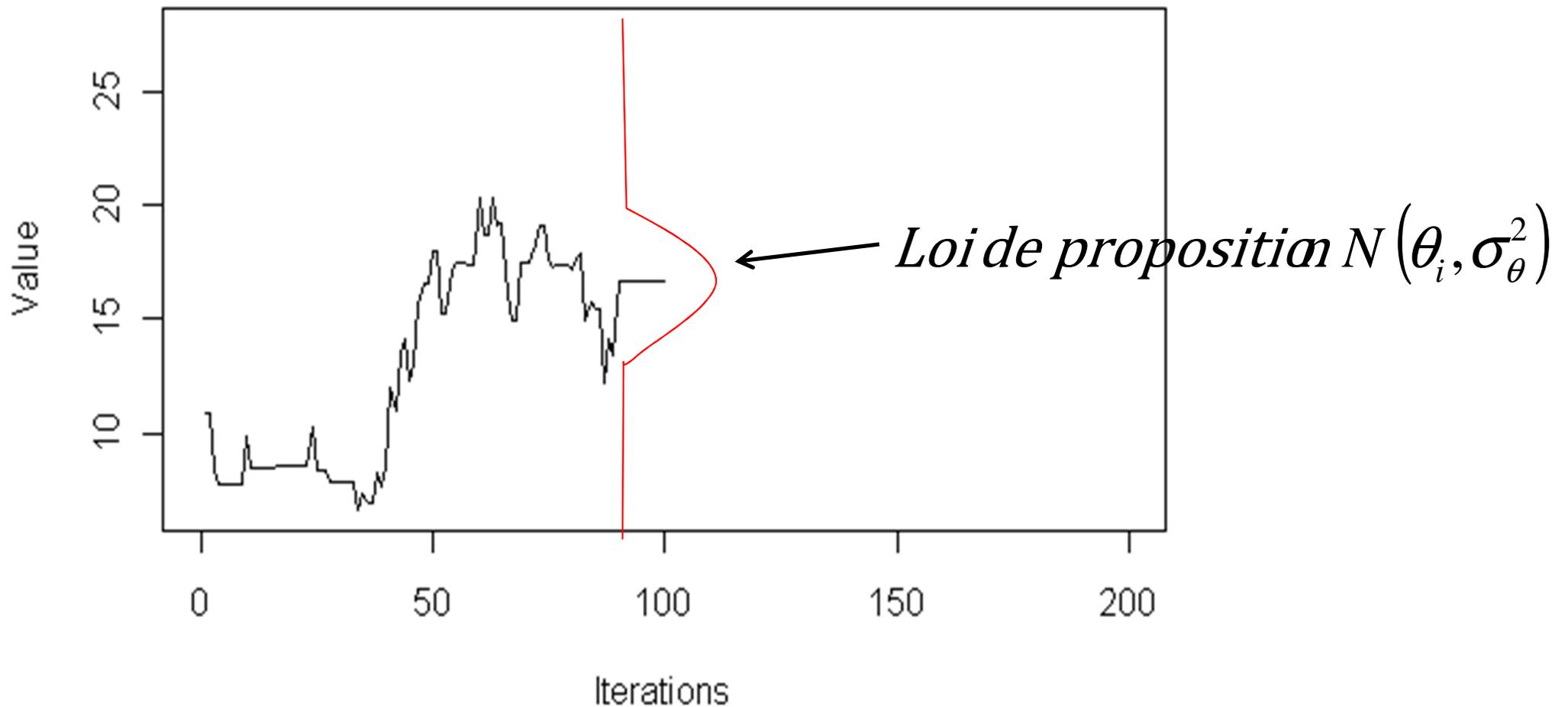
- Méthode de Monte Carlo par Chaîne de Markov (MCMC) avec algorithme de Metropolis-Hastings within Gibbs
 1. Set $p = 0$. Choose an initial value $\theta^{(0)}$ for the disease model parameters. Draw an initial value $\sigma^{2(0)}$ of the residual variance as in step 3, with $\theta = \theta^{(0)}$.
 2. Metropolis Hastings step, which draws values from $f(\theta|Y^{obs}, \sigma^2) \propto f(Y^{obs}|\theta, \sigma^2)\pi(\theta)$
 - a. Draw a proposed parameter vector $\theta^{*(p+1)}$ using the proposal distribution $\theta^{*(p+1)} \sim N(\theta^{(p)}, tune * \Sigma)$ where $tune$ is a constant and Σ is a covariance matrix (see below).
 - b. Calculate the ratio
$$r = \frac{f(Y^{obs}|\theta^{*(p+1)}, \sigma^{2(p)})\pi(\theta^{*(p+1)}, \sigma^{2(p)})}{f(Y^{obs}|\theta^{(p)}, \sigma^{2(p)})\pi(\theta^{(p)}, \sigma^{2(p)})}$$
If $r > 1$, $\theta^{(p+1)} = \theta^{*(p+1)}$
If $r < 1$, $\theta^{(p+1)} = \theta^{*(p+1)}$ with probability r and $\theta^{(p+1)} = \theta^{(p)}$ with probability $1 - r$.
 3. Gibbs step. Draw values from $f(\sigma^2|Y^{obs}, \theta) \propto f(Y^{obs}|\theta, \sigma^2)\pi(\sigma^2)$, which can be recognized as an inverse gamma distribution ($1/\sigma^2$ has a gamma distribution).

Draw

$$\sigma^{2(p+1)} \sim \text{InvGamma}(\alpha, \beta)$$
$$\alpha = 73/2 + 2$$
$$\beta = \sum_{i=1}^{73} [y_i - y_i^{obs}]^2 / 2$$
 4. Repeat steps 2 and 3 until convergence

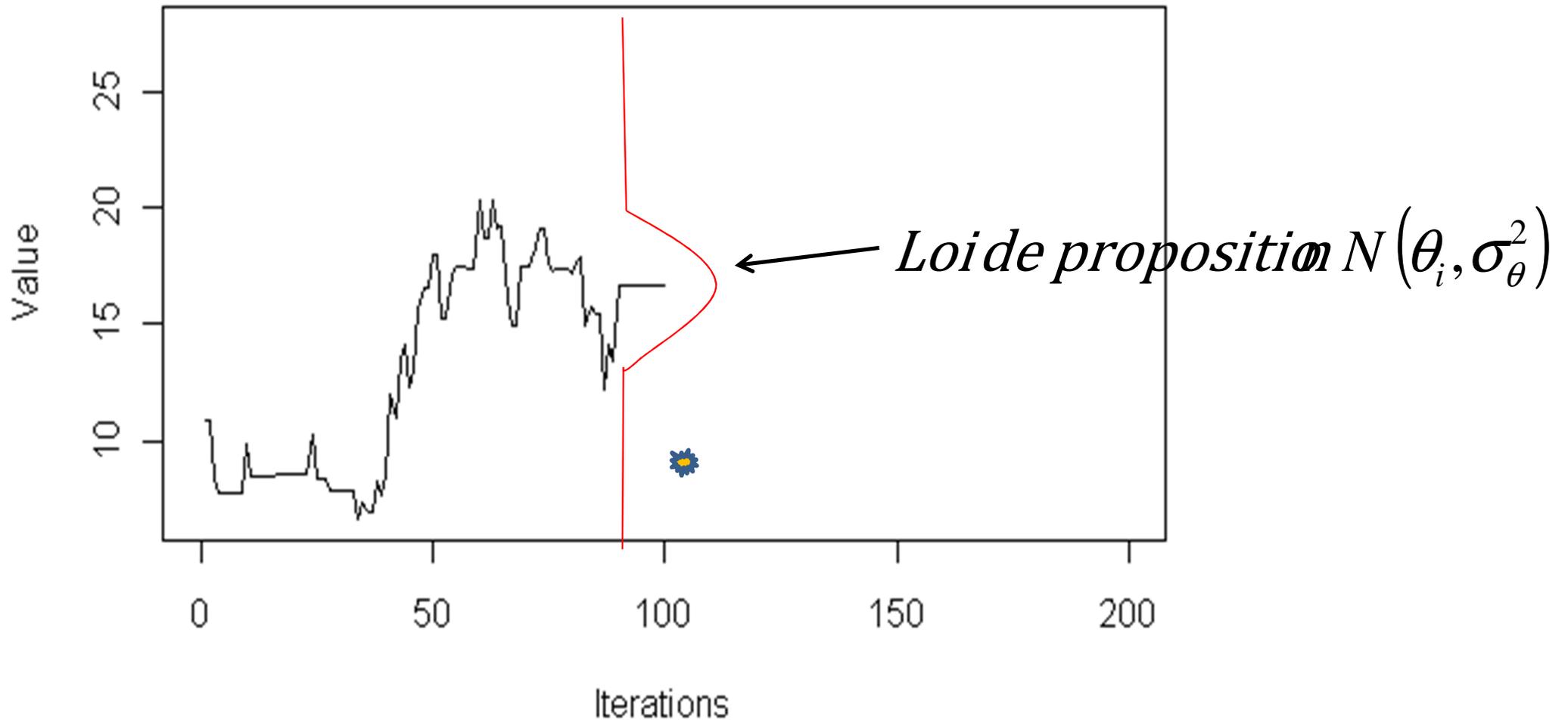
Algorithme – illustration – 1/5

Trace de la chaîne de Markov



Algorithme - illustration – 2/5

Trace de la chaîne de Markov

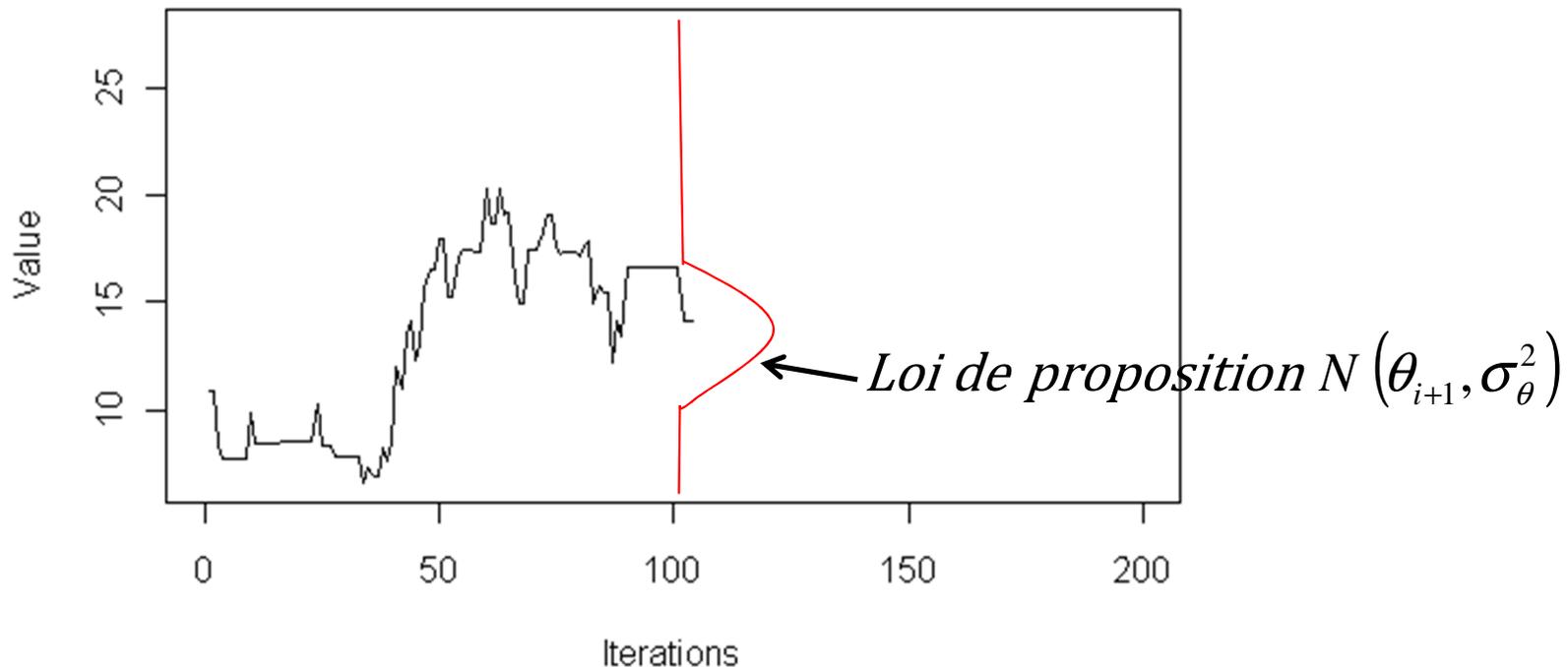


Algorithme - illustration – 3/5

$$\frac{P(y|\theta_{i+1}) \times \pi(\theta_{i+1})}{P(y|\theta_i) \times \pi(\theta_i)} \geq 1$$

$$P(y|\theta_{i+1}) \times \pi(\theta_{i+1}) \geq P(y|\theta_i) \times \pi(\theta_i)$$

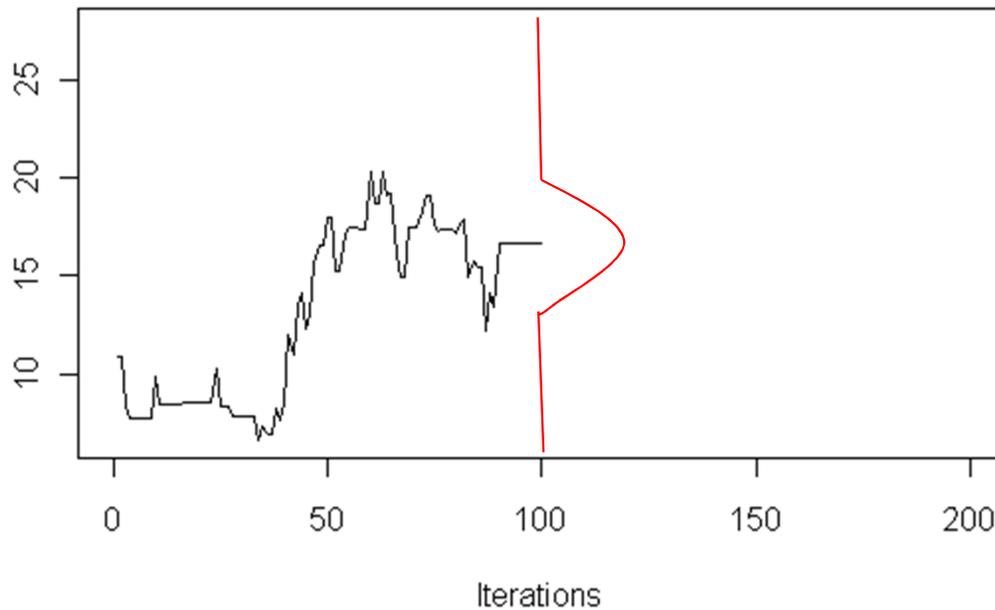
Trace de la chaîne de Markov



Algorithme - illustration – 4/5

$$\frac{P(y|\theta_{i+1}) \times \pi(\theta_{i+1})}{P(y|\theta_i) \times \pi(\theta_i)} < 1$$

Trace de la chaîne de Markov

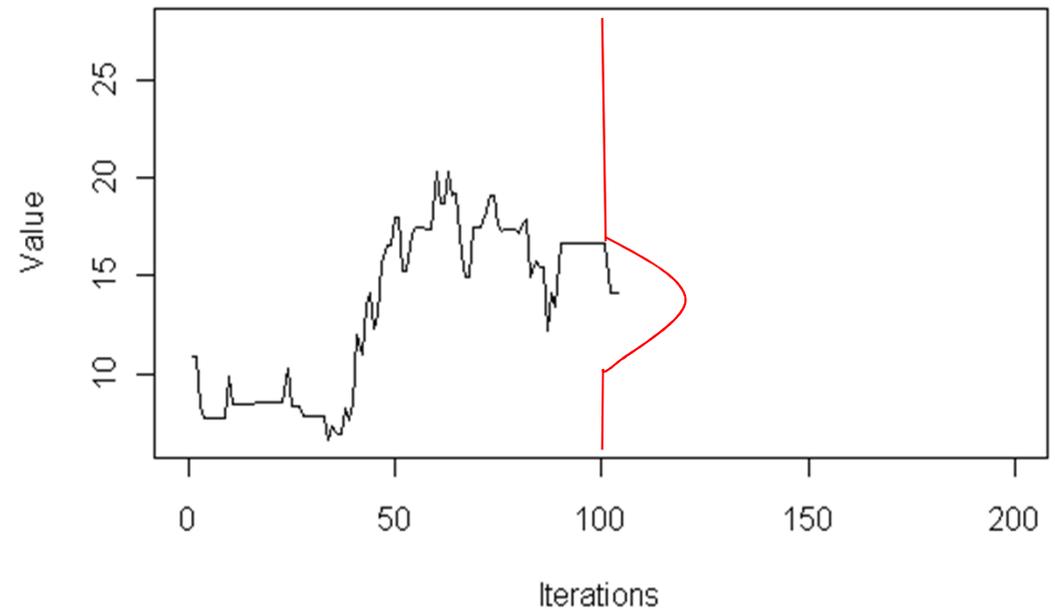
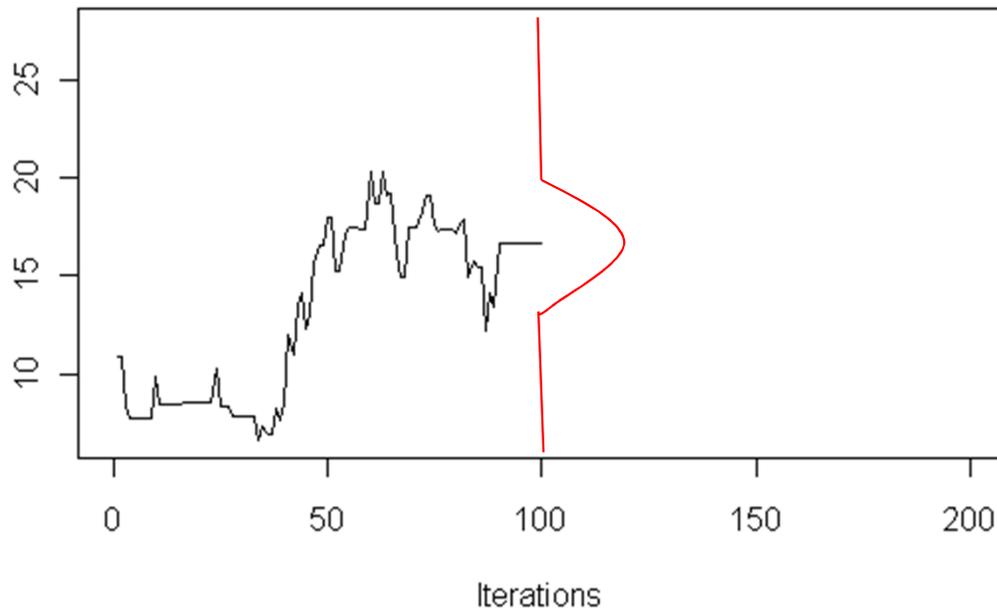


Algorithme - illustration – 5/5

$$\frac{P(y|\theta_{i+1}) \times \pi(\theta_{i+1})}{P(y|\theta_i) \times \pi(\theta_i)} < 1$$

Trace de la chaîne de Markov

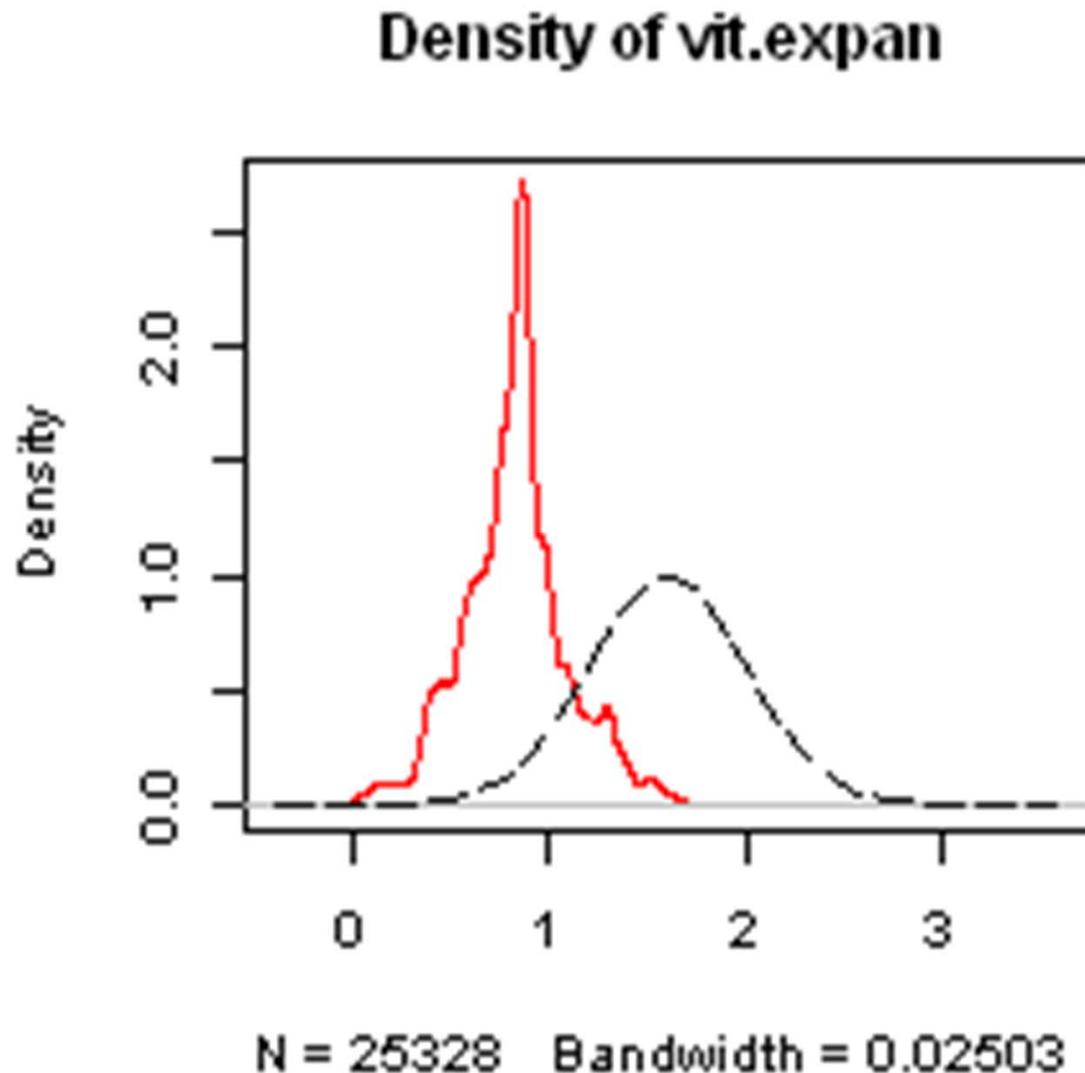
Trace de la chaîne de Markov



Après 27 jours de calcul...

- > 150000 itérations
- Convergence OK

**=> Distribution
résumée par 516
vecteurs**

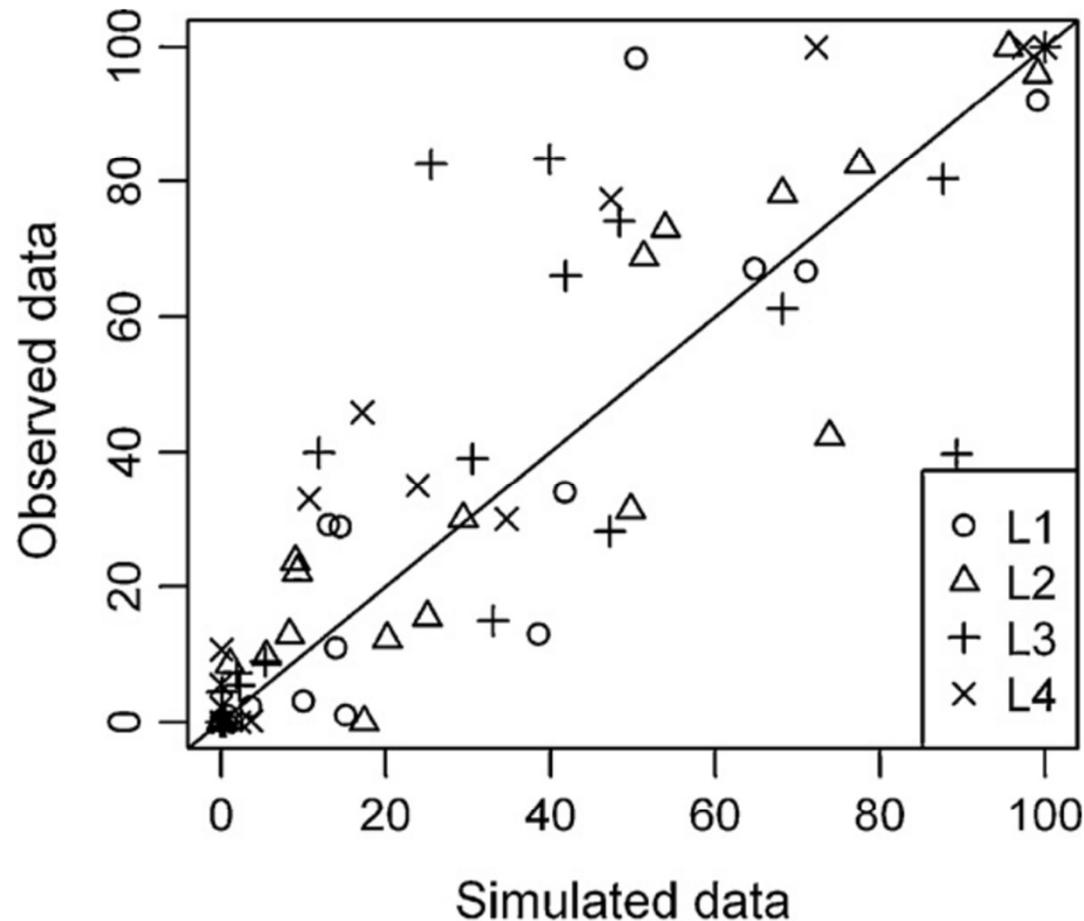


distributions *a posteriori*

Parameter	Units	Prior shape	Prior mean	Prior std. dev.	Posterior Mean	Posterior std. dev.
θ_{mult}	% necrosis/% necrosis	Uniform	8.25	3.90	8.89	2.32
$\theta_{\text{winter.th}}$	% necrosis	Gaussian	20.70	5.17	20.97	5.26
θ_{scale}	–	Uniform	50.50	28.58	5.42	3.75
$\theta_{\text{min.inoc}}$	% necrosis	Gaussian	5.50	2.25	7.26	2.15
$\theta_{T\text{min}}$	°C	Gaussian	2.50	2.25	2.16	2.17
θ_{deplete}	% necrosis/% necrosis	Uniform	7.50	3.75	5.00	3.81
θ_{rain}	mm ⁻¹	Gaussian	-0.35	0.08	-0.38	0.07
θ_{height}	cm ⁻¹	Gaussian	0.55	0.63	0.26	0.17
$\theta_{\text{prod.inoc}}$	% necrosis/% necrosis	Uniform	5.50	2.60	5.71	2.46
$\theta_{\text{expansion}}$	% necrosis/°C	Gaussian	0.047	0.028	0.030	0.015
θ_{initial}	% necrosis	Gaussian	0.26	0.12	0.21	0.10
$\theta_{\text{min.today}}$	°C	Gaussian	2.50	2.25	2.53	2.29
$\theta_{\text{min.yesterday}}$	°C	Gaussian	2.50	2.25	2.41	2.39
$\theta_{\text{min.rain}}$	mm	Gaussian	1.53	0.74	1.44	0.43
$\theta_{\text{limit } F1}$	mm	Gaussian	7.50	2.25	5.59	1.46
$\theta_{\text{leaf.growth}}$	°C-days/°C-days	Gaussian	1.65	0.08	1.63	0.08
$\theta_{\text{rain.infect}}$	mm	Gaussian	7.50	2.25	6.36	2.21
variance of residual error		Jeffrey			0.054	0.010

Qualité d'ajustement

$$\text{eff} = 1 - \left(\sum_{m=1}^{73} (S_m^{\text{obs}} - S_m)^2 / \sum_{m=1}^{73} (S_m^{\text{obs}} - \bar{S}^{\text{obs}})^2 \right) = 0.76$$



$$\text{RMSE} = \sqrt{1/n \sum_{m=1}^{73} (S_m^{\text{obs}} - \hat{S}_m)^2} = 16.86\%$$

Fig. 1. Scatter plot of observed versus the posterior mean of predicted severity for the eight site-years used for parameter estimation for leaf layers 1–4 (L1 to L4 respectively) where L1 corresponds to uppermost leaf.

Démarche

1. Modèle Septo-LIS décrivant la biologie de la septoriose
2. Estimation des paramètres avec leurs incertitudes
3. Projection CC en prenant en compte l'incertitude sur le modèle et le CC

Simulations et variables d'intérêt

- 3 sites × 58 années × 5 modèles climatiques × 516 vecteurs de paramètres
⇒ 448920 simulations
- **Sévérité “effective”** : pourcentage de nécrose
 - moyennée sur les 3 dernières feuilles
 - période de remplissage de grain (BBCH 55 à BBCH 75)
- calcul de 4 variables d'intérêt
 - **Q1: sévérité moyennée sur les années**
 - Q2: variabilité interannuelle de la sévérité
 - Q3: risque d'avoir des années avec très forte sévérité
 - Q4: risque d'avoir des années avec très faible sévérité

Q1 : Sévérité moyenne



Effet médian :
diminution septoriose

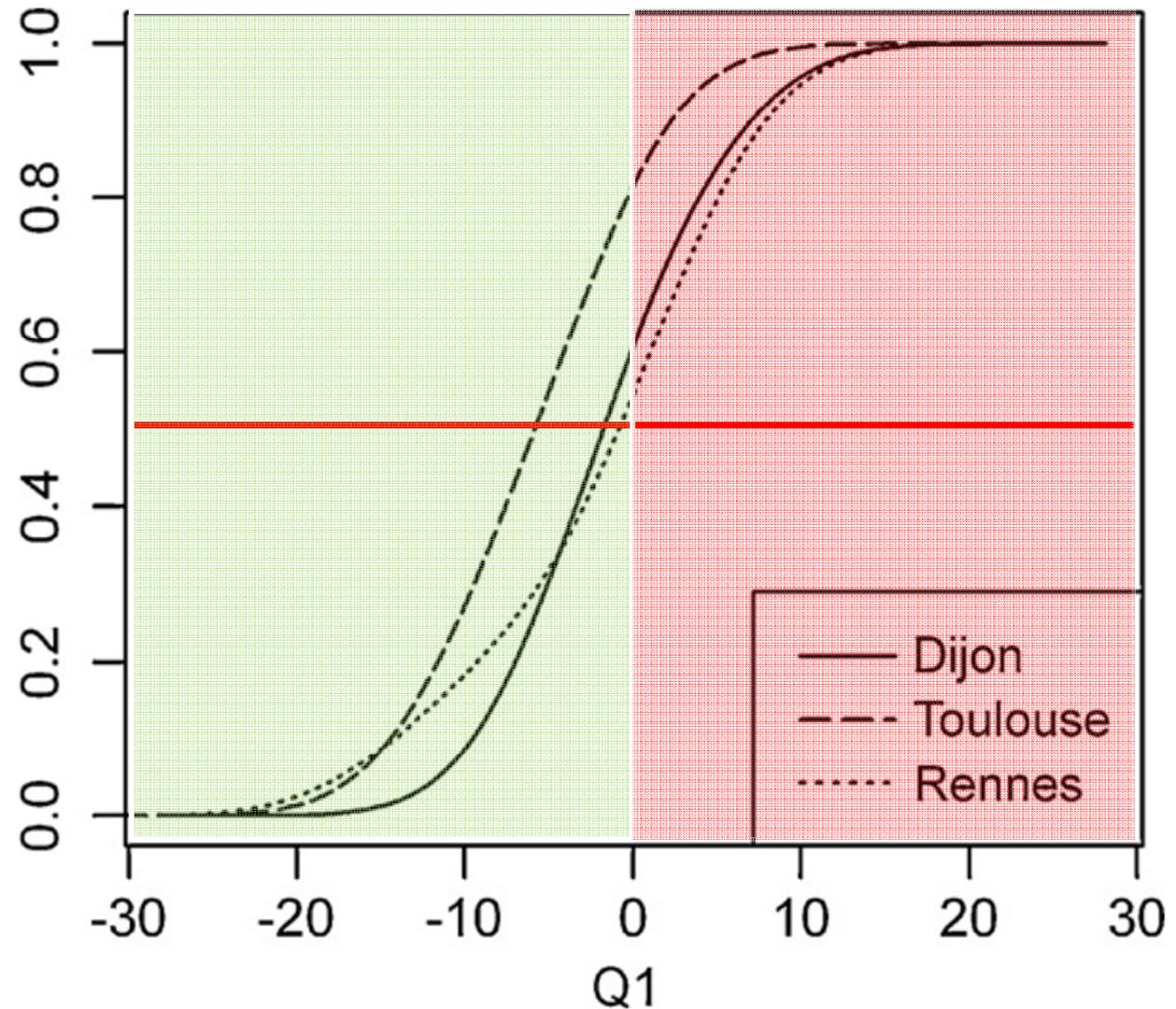
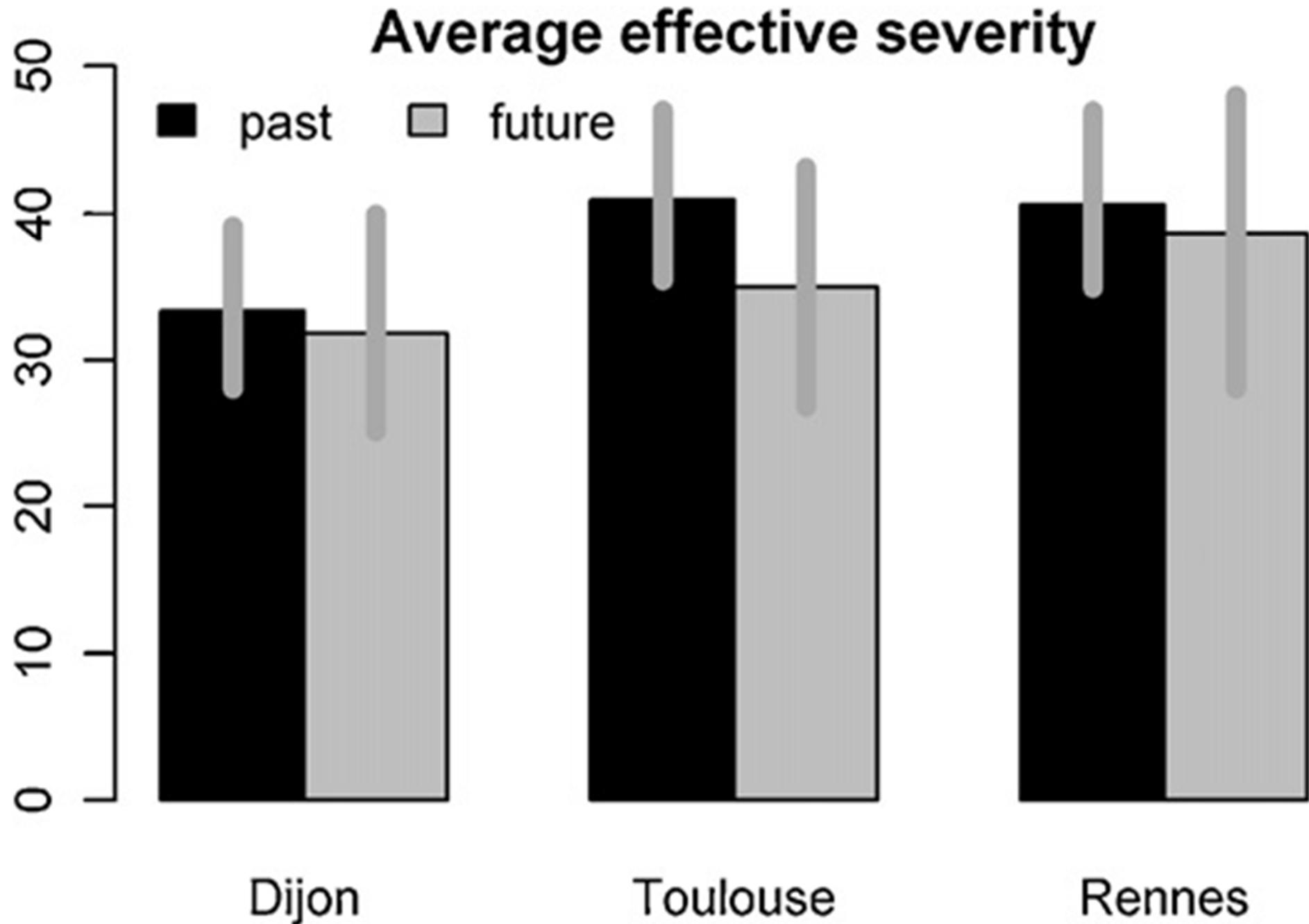


Fig. 3. Cumulative probability distribution functions for criterion Q1 for the 3 locations.

Q1 : Sévérité moyenne



Conclusion

- voir si les méthodes bayésiennes sont utiles ?
- Difficultés
 - ... finalement, assez peu d'information *a priori*
 - Mise en œuvre des algorithmes sur ce type de modèle
 - Besoin de recoder les algorithmes : risque d'erreur !
 - Temps de calcul monstrueux !
 - Valorisation que d'une partie des données existantes...

perspectives

- Des applications opérationnelles ?
 - Analyse fréquentielle souvent utilisé, **y intégrer l'incertitude du modèle serait assez naturel**
 - Nécessité **d'exploiter toutes les données** disponibles
 - Temps de calcul : ici approche 100% R, mais Septo-LIS disponible via agrobox
- Changement climatique
 - On va vers des approches multi-modèles...
 - Mais difficulté d'avoir différents modèles. Notre approche, un substitue ?



Modelling climate change impact on *Septoria tritici* blotch (STB) in France: Accounting for climate model and disease model uncertainty

David Gouache^{a,*}, Arnaud Bensadoun^b, François Brun^c, Christian Pagé^d, David Makowski^e, Daniel Wallach^f

^a Arvalis – Institut du Végétal, rue de Noetzlin, Bât. 630, 91405 Orsay, France

^b Arvalis – Institut du Végétal, UMR AGIR B.P. 52627, F-31326 Castanet Tolosan, France

^c ACTA, UMR AGIR B.P. 52627, F-31326 Castanet Tolosan, France

^d SUC au CERFACS, URA CERFACS/CNRS No 1875, 42 Avenue Gaspard Coriolis, 31057 Toulouse, France

^e INRA, UMR 211. 78850 Thiverval-Grignon, France

^f INRA, UMR AGIR B.P. 52627, F-31326 Castanet Tolosan, France

A R T I C L E I N F O

Article history:

Received 28 February 2011

Received in revised form 11 January 2012

Accepted 27 April 2012

Keywords:

Climate change

Disease model

Bayesian method

Uncertainty

Global warming

Metropolis Hastings

Parameter estimation

Triticum aestivum L.

Septoria tritici

A B S T R A C T

We calculate the impact of climate change on the effective severity of *Septoria tritici* blotch (STB) of winter wheat (*Triticum aestivum* L.) at three representative locations in France. The calculation uses climate models for climate prediction, and a disease model to link disease severity to weather. Four impact criteria are considered: the change in average (over years) severity, the change in interannual variance of severity, the change in number of years with particularly high severity and the change in the number of years with particularly low severity.

We also calculate the uncertainty associated with those impact criteria. Three different uncertainty sources are considered: uncertainty in predicting climate, uncertainty in the values of the disease model parameters and uncertainty due to residual error of the disease model. Uncertainty in climate is considered by using different global climate models and downscaling methodologies to produce five different climate series for greenhouse gas emission scenario A1B, for a baseline period comprising harvest years 1971–1999 and a future period spanning 2071–2099. A Bayesian approach, using a Metropolis Hastings within Gibbs algorithm, is used for parameter estimation. This gives a posterior distribution both for the 17 model parameters that were considered and for the variance of residual error.

Climate change is predicted to reduce the average severity of STB by 2–6%, depending on location, and to result in more low severity years and fewer high severity years. There is appreciable uncertainty. For example, the probability that average severity will increase rather than decrease is 40%, 18% and 45% for the three locations. We calculated first order sensitivity indices for climate model, parameter vector and residual error considered as three factors. The climate model factor has by far the largest sensitivity index. However, interactions between factors also make a major contribution to overall variance.