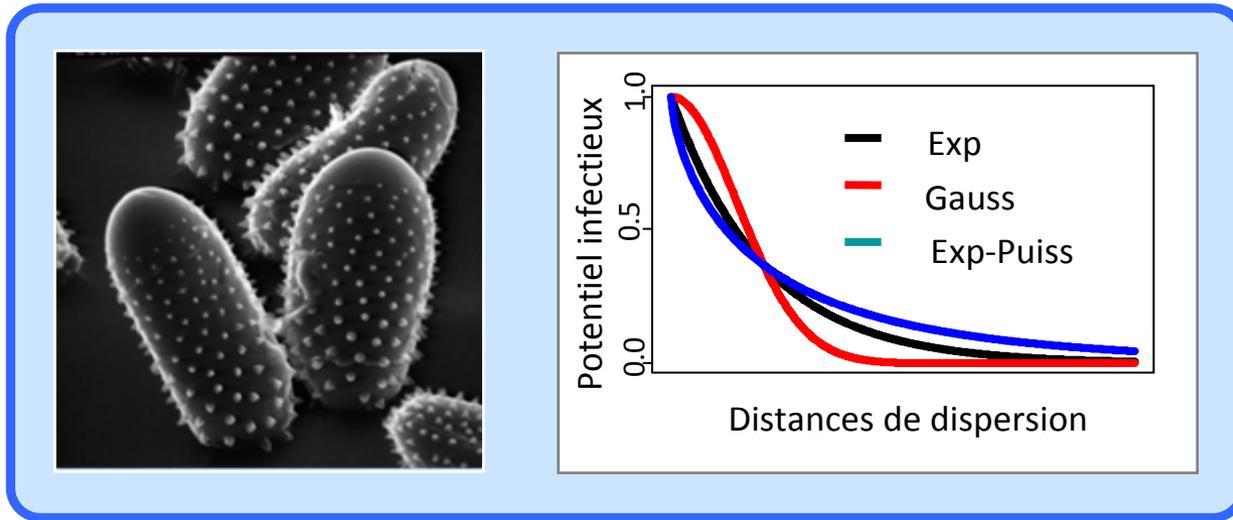


# Estimation de la dispersion de la rouille du peuplier en vallée de la Durance



**Frédéric Fabre, Constance Xhaard, Jérôme Coville, Samuel Soubeyrand  
& Fabien Halkett**

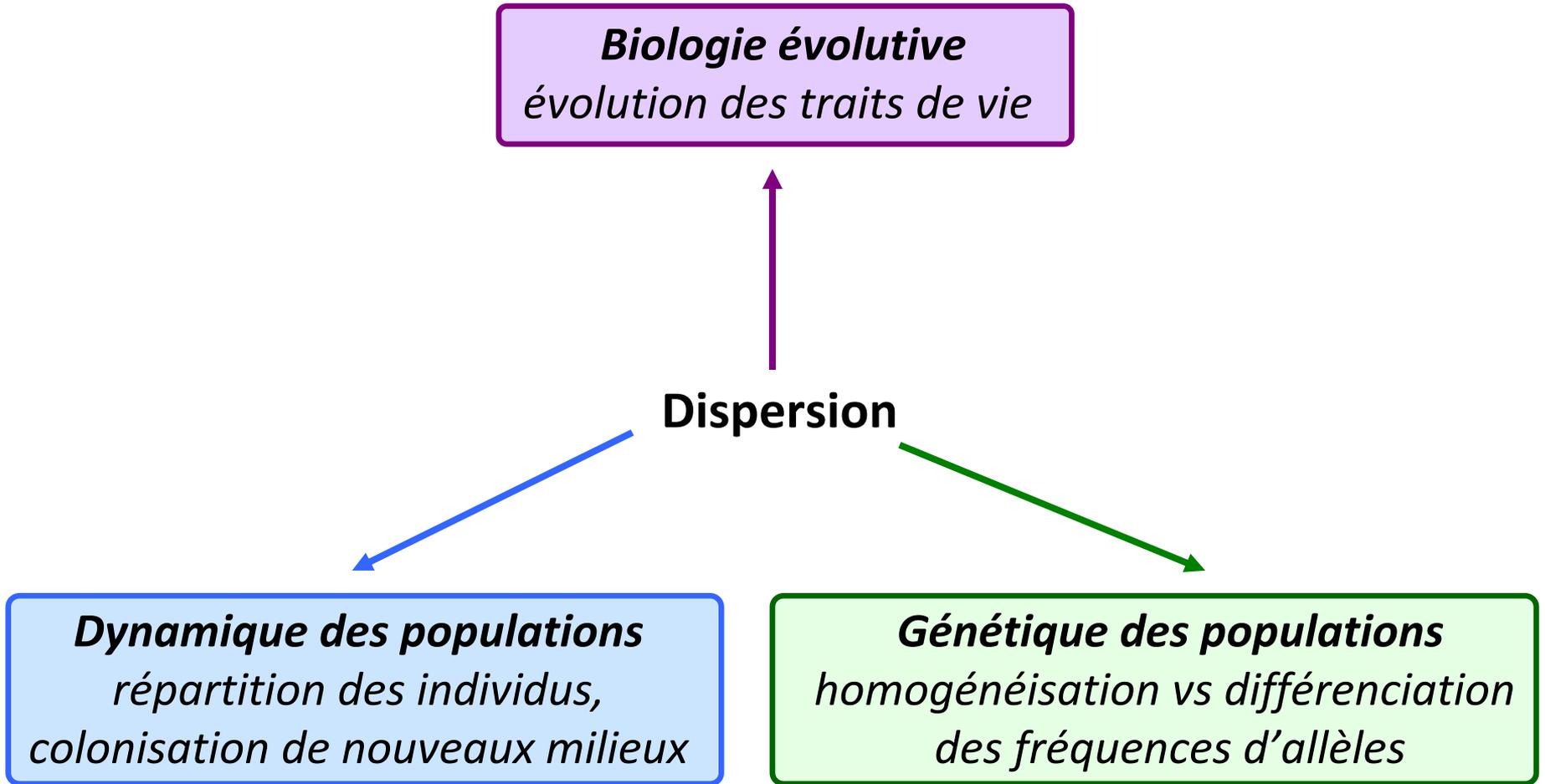
**INRA Avignon (BioSP & PV), INRA Nancy (UMR 1136)**

**Biologie évolutive**  
*évolution des traits de vie*

**Dispersion**

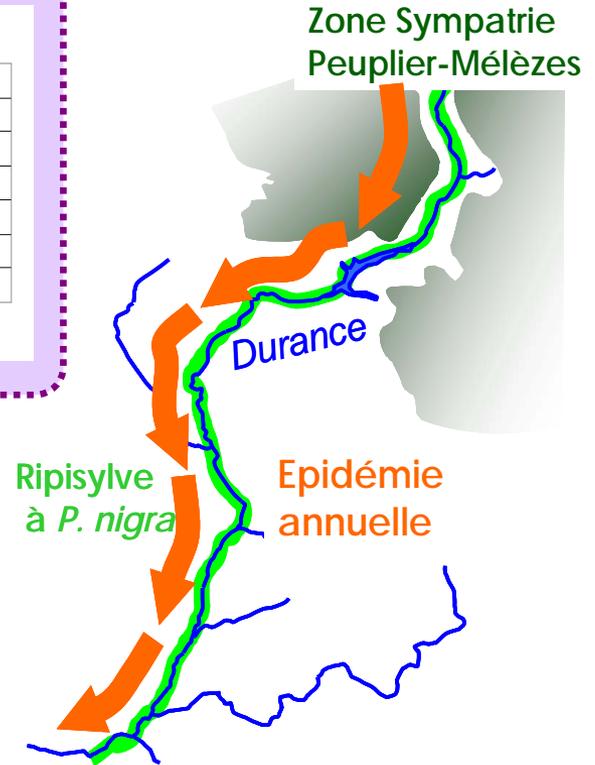
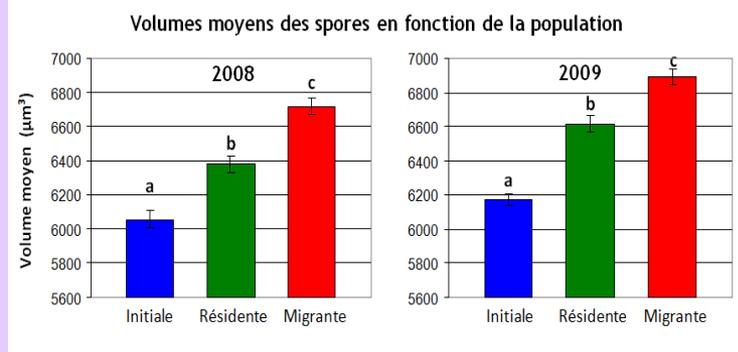
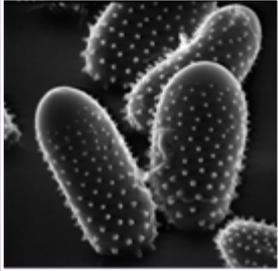
**Dynamique des populations**  
*répartition des individus,  
colonisation de nouveaux milieux*

**Génétique des populations**  
*homogénéisation vs différenciation  
des fréquences d'allèles*

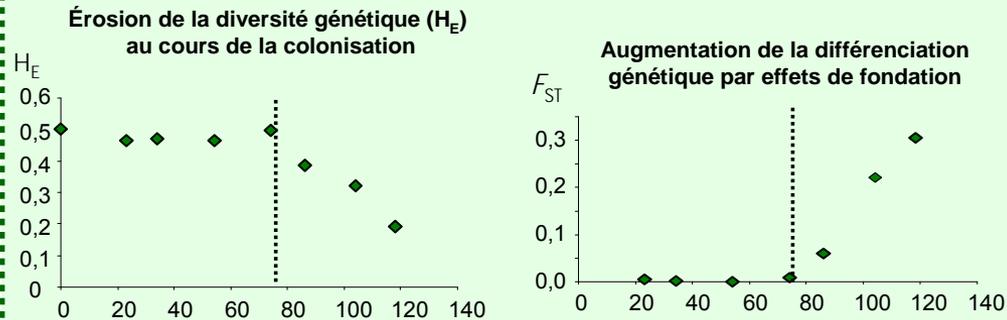


# Approche multidisciplinaire de l'étude de la dispersion

## Biologie évolutive



## Génétique des populations



## Dynamique des populations

Estimation de la dispersion  
lors d'évènement de colonisation



# La vallée de la Durance : un corridor écologique



200 km

Durance

France métr...  
Echelle 1 : 772.170  
0 10 km

IGN

# Modèle d'étude : la rouille du peuplier

## *Melampsora larici-populina*



Champignon parasite foliaire

Nécessite deux hôtes d'espèces différentes pour accomplir son cycle de vie...

# Cycle biologique de *Melampsora larici-populina*

## Reproduction sexuée



écidiospores ( $n+n$ )

*plasmogamie*

Aiguilles de mélèzes

spermogonies  
et spermaties ( $n$ )

basidiospores ( $n$ )

*méiose*

basides ( $2n$ )

## Multiplication clonale



urédiniospores  
( $n+n$ )

Feuilles de  
peupliers

téliospores  
( $n+n$ )

*caryogamie*

Cette particularité biologique fait de la vallée de la Durance une situation écologique propice à l'étude de l'épidémiologie de ce champignon....

# La rouille en vallée de la Durance

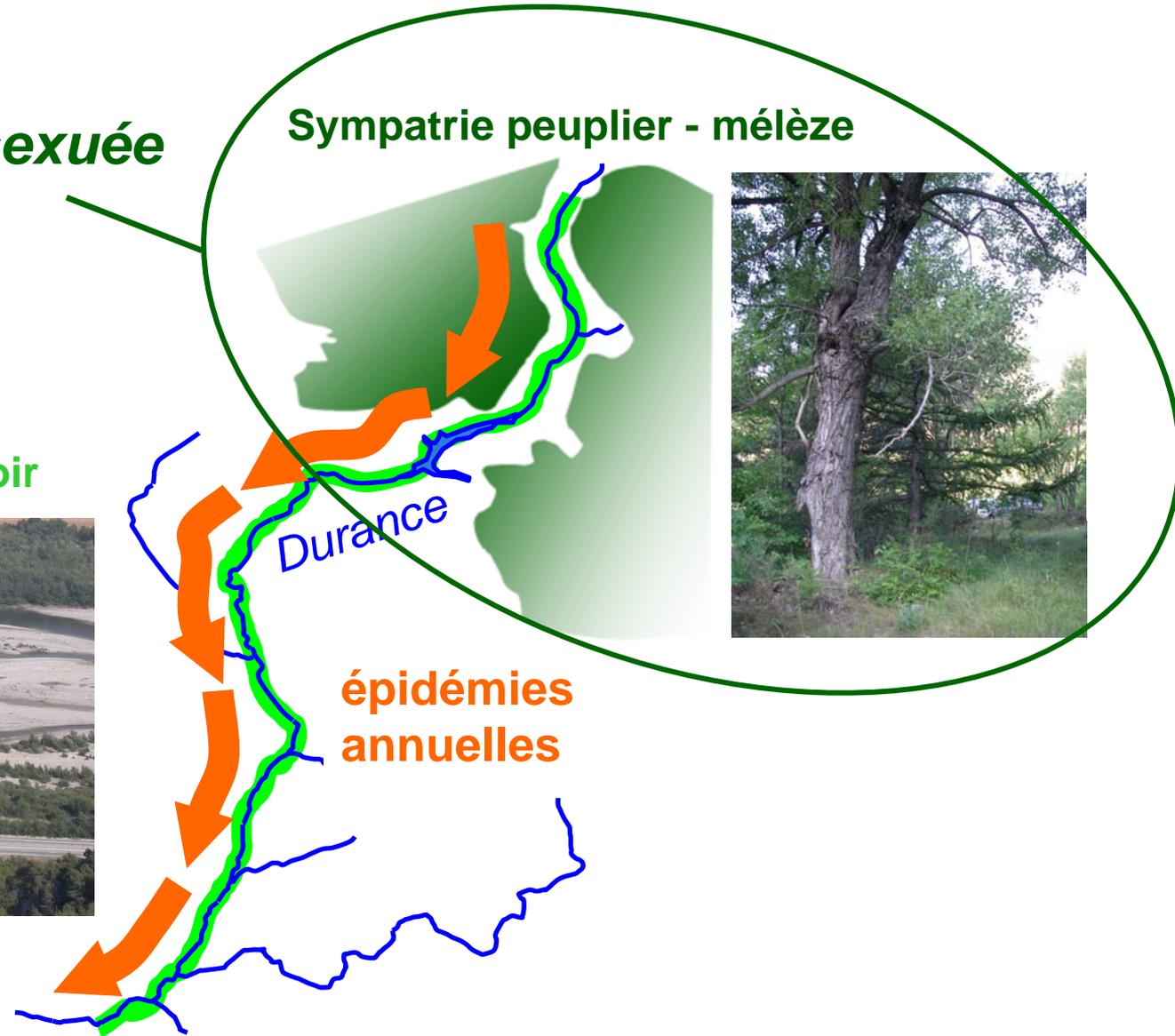
**Reproduction sexuée**

**Sympatrie peuplier - mélèze**

**Ripisylve à peuplier noir**



200 km de long



# Suivi spatio-temporel de l'épidémie

## Méthodes

- 7 campagnes & 13 sites suivis en 2008
- Deux protocoles d'échantillonnage pour mesurer l'incidence de la maladie:
  - 1<sup>er</sup> échantillonnage : détermination du % d'arbres infectés sur des peupliers choisis au hasard
  - 2<sup>ème</sup> échantillonnage : détermination du % de feuilles infectées sur des rameaux collectés au hasard

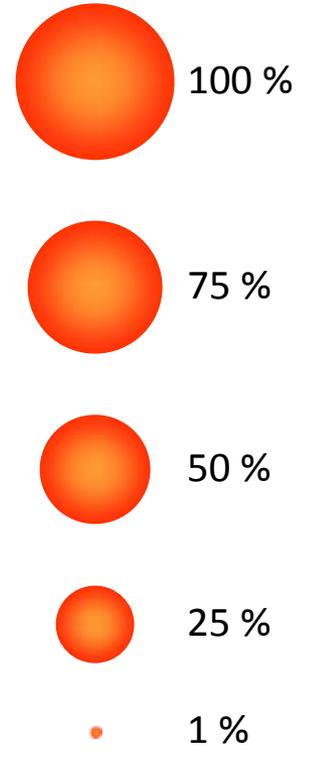
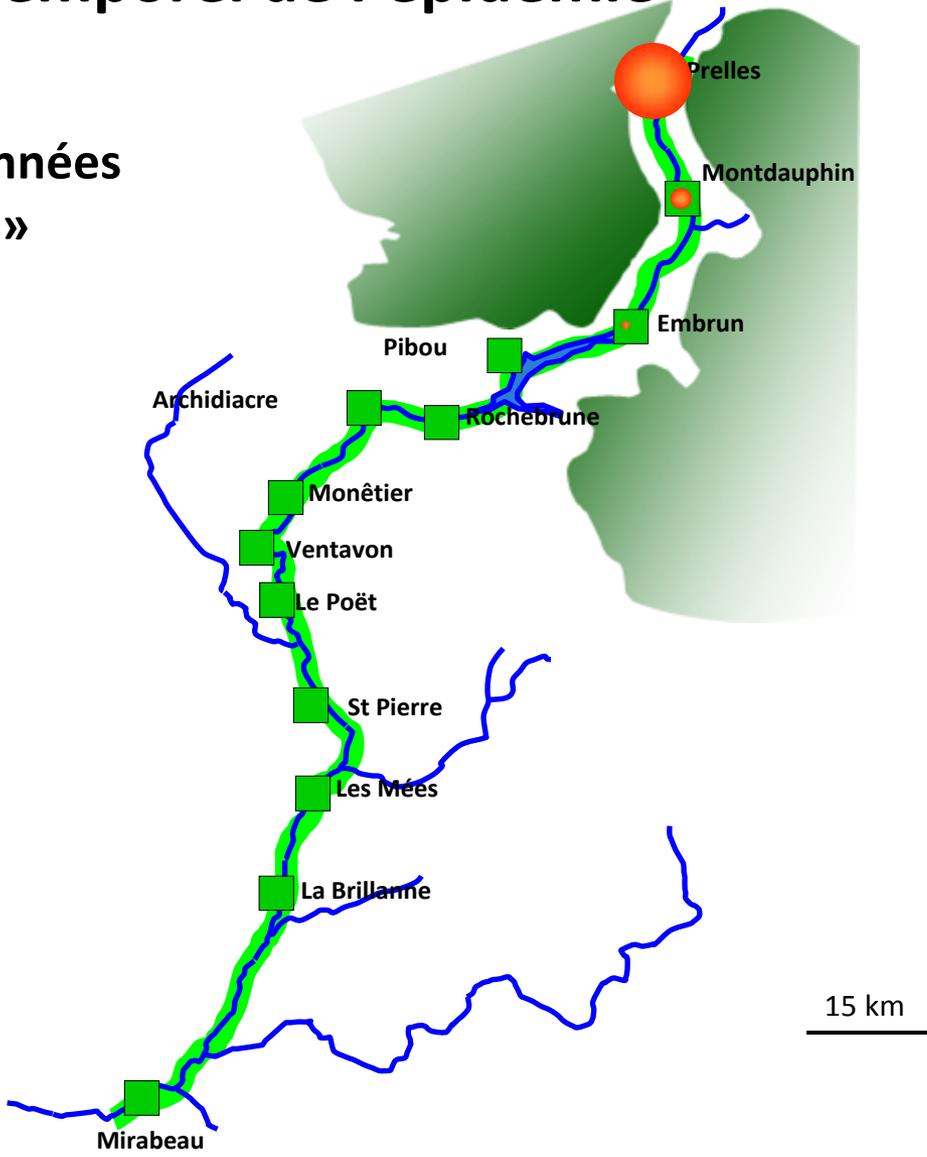


## Objectifs :

1. Décrire le roulement de l'épidémie
2. Estimation la fonction dispersion du pathogène

# Suivi spatio-temporel de l'épidémie

A partir des données  
« rameaux »

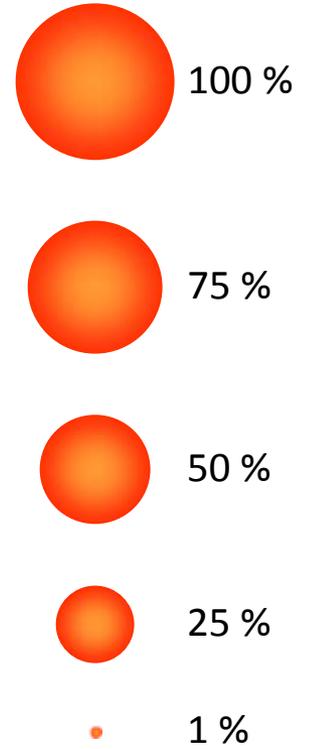
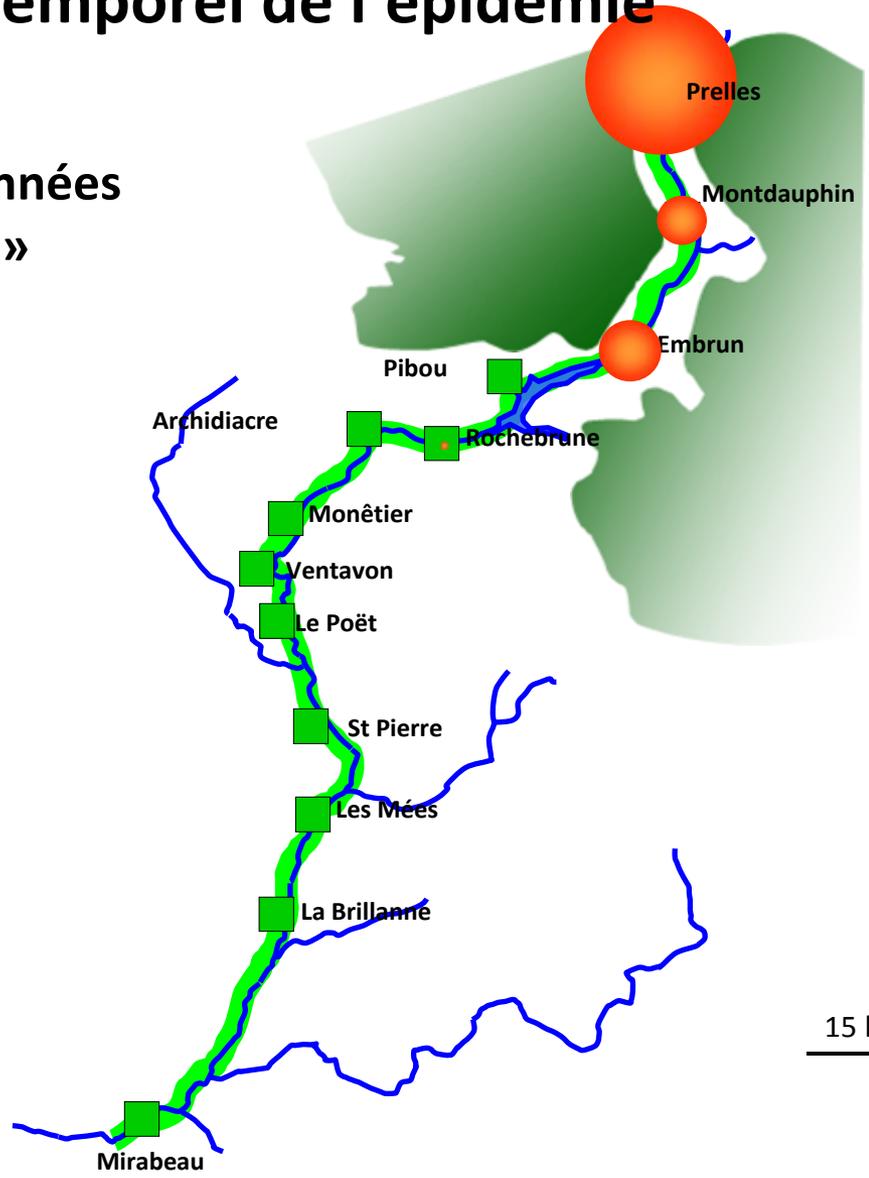


Pourcentage  
d'infection des  
rameaux

Tournée 1 (du 9 au 11 juillet)

# Suivi spatio-temporel de l'épidémie

A partir des données  
« rameaux »



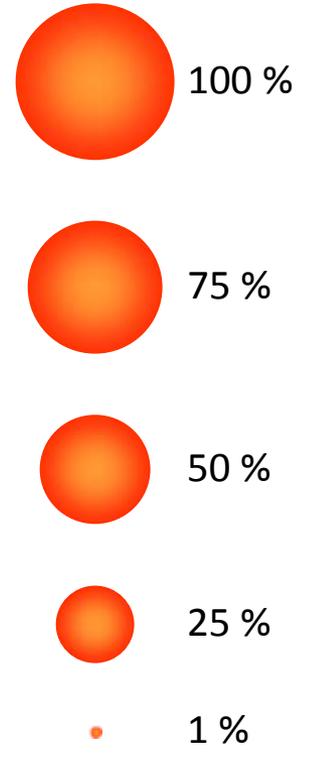
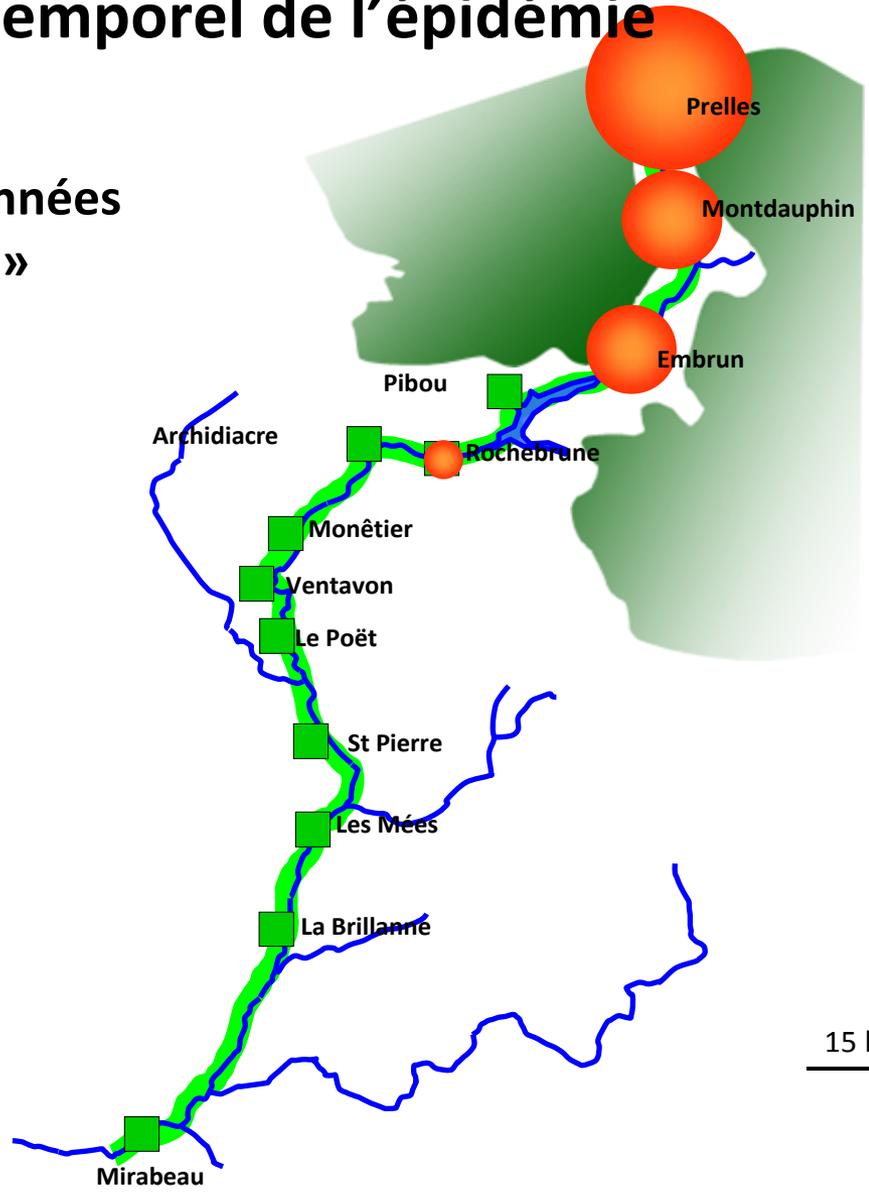
Pourcentage  
d'infection des  
rameaux

15 km

Tournée 2 (du 28 au 30 juillet)

# Suivi spatio-temporel de l'épidémie

A partir des données  
« rameaux »



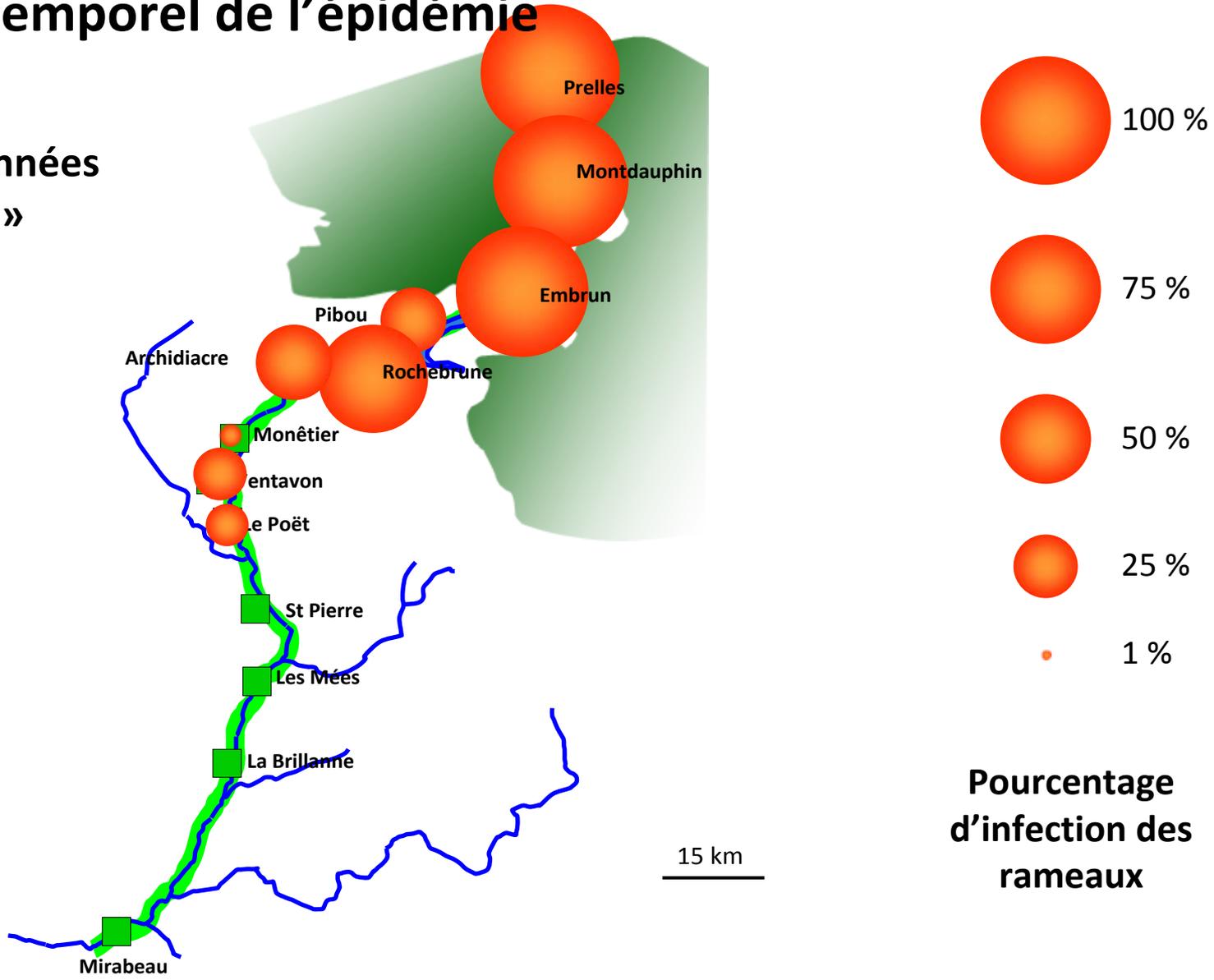
Pourcentage  
d'infection des  
rameaux

15 km

Tournée 3 (du 20 au 22 aout)

# Suivi spatio-temporel de l'épidémie

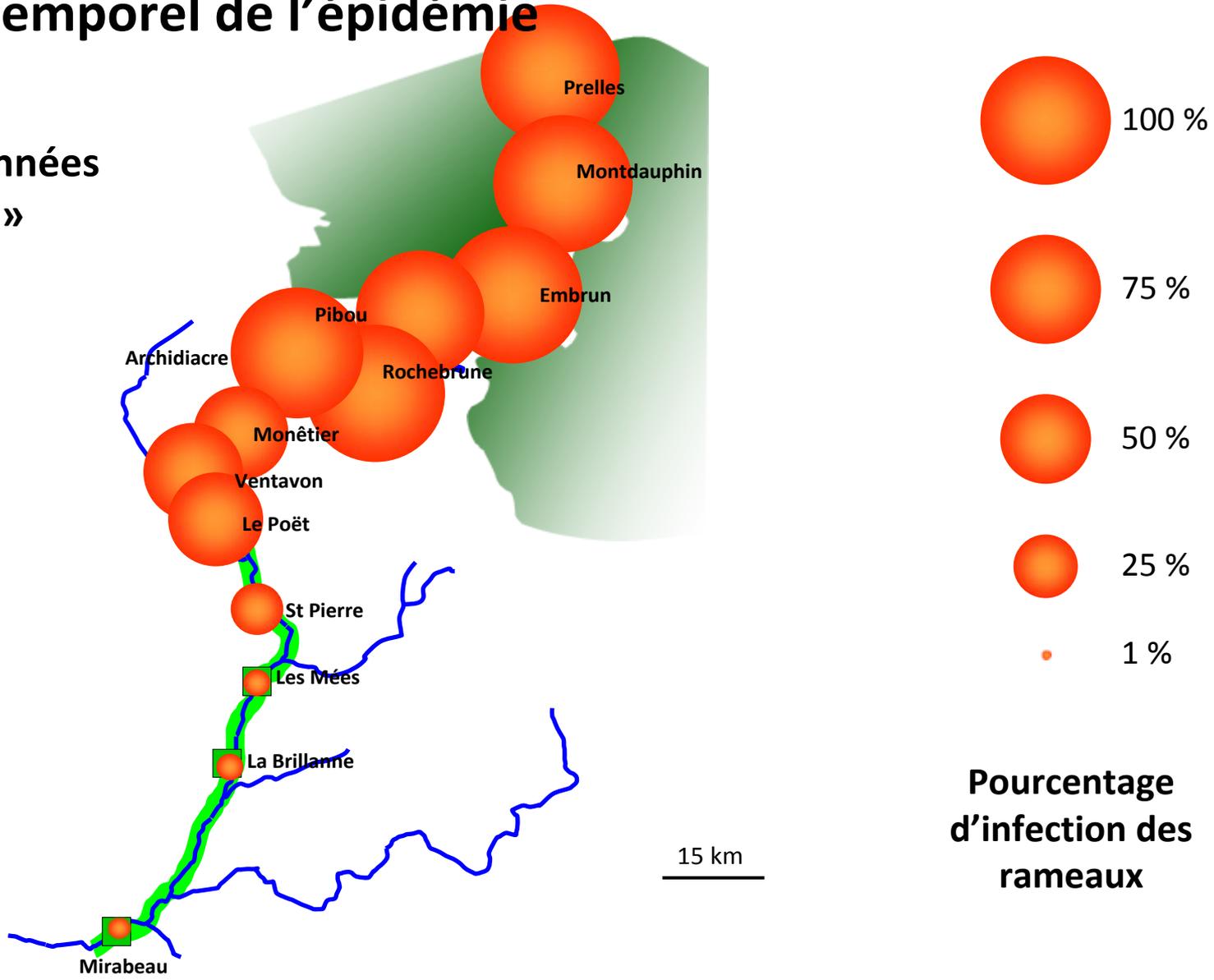
A partir des données  
« rameaux »



Tournée 4 (du 08 au 10 septembre)

# Suivi spatio-temporel de l'épidémie

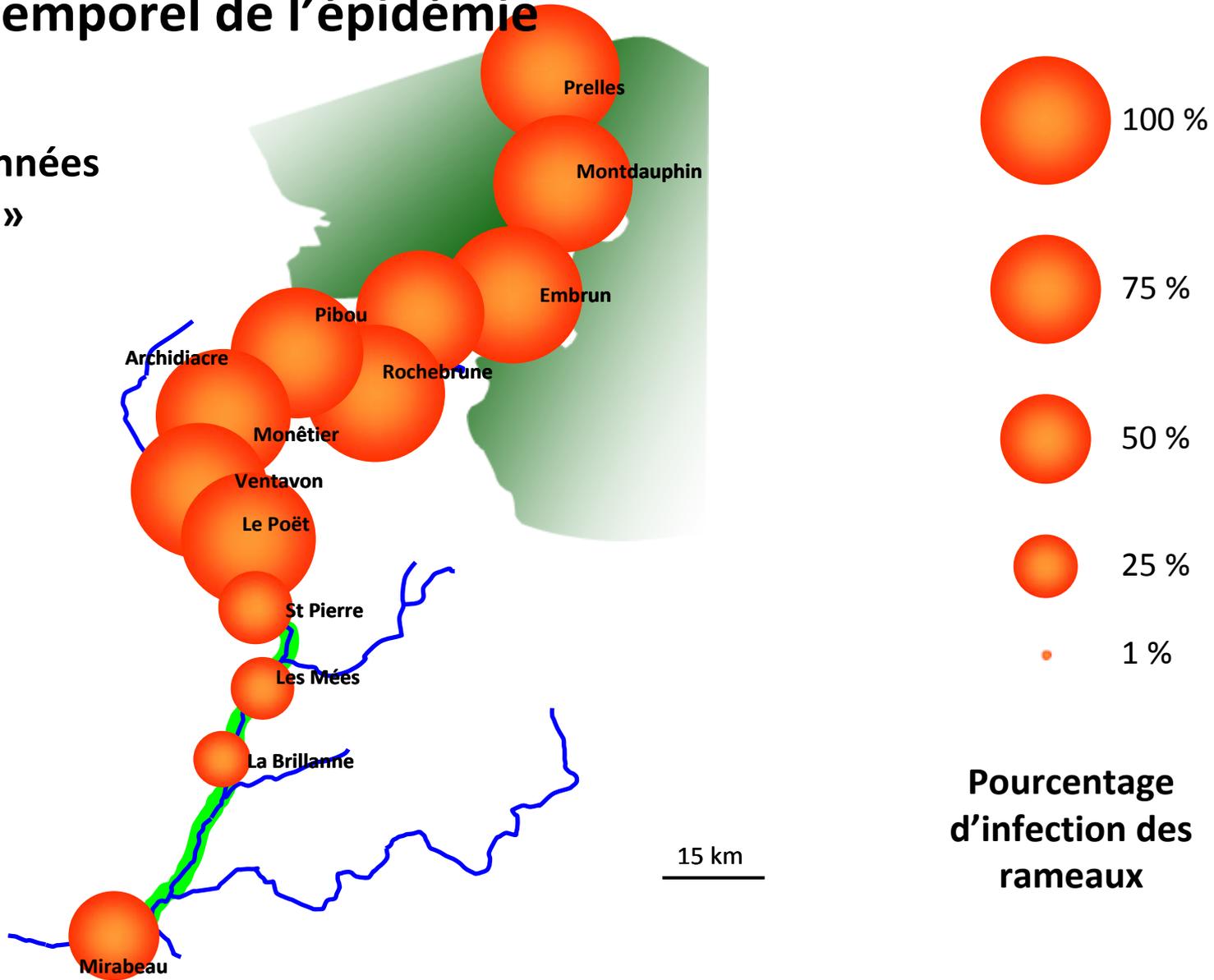
A partir des données  
« rameaux »



Tournée 5 (du 30 septembre au 02 octobre)

# Suivi spatio-temporel de l'épidémie

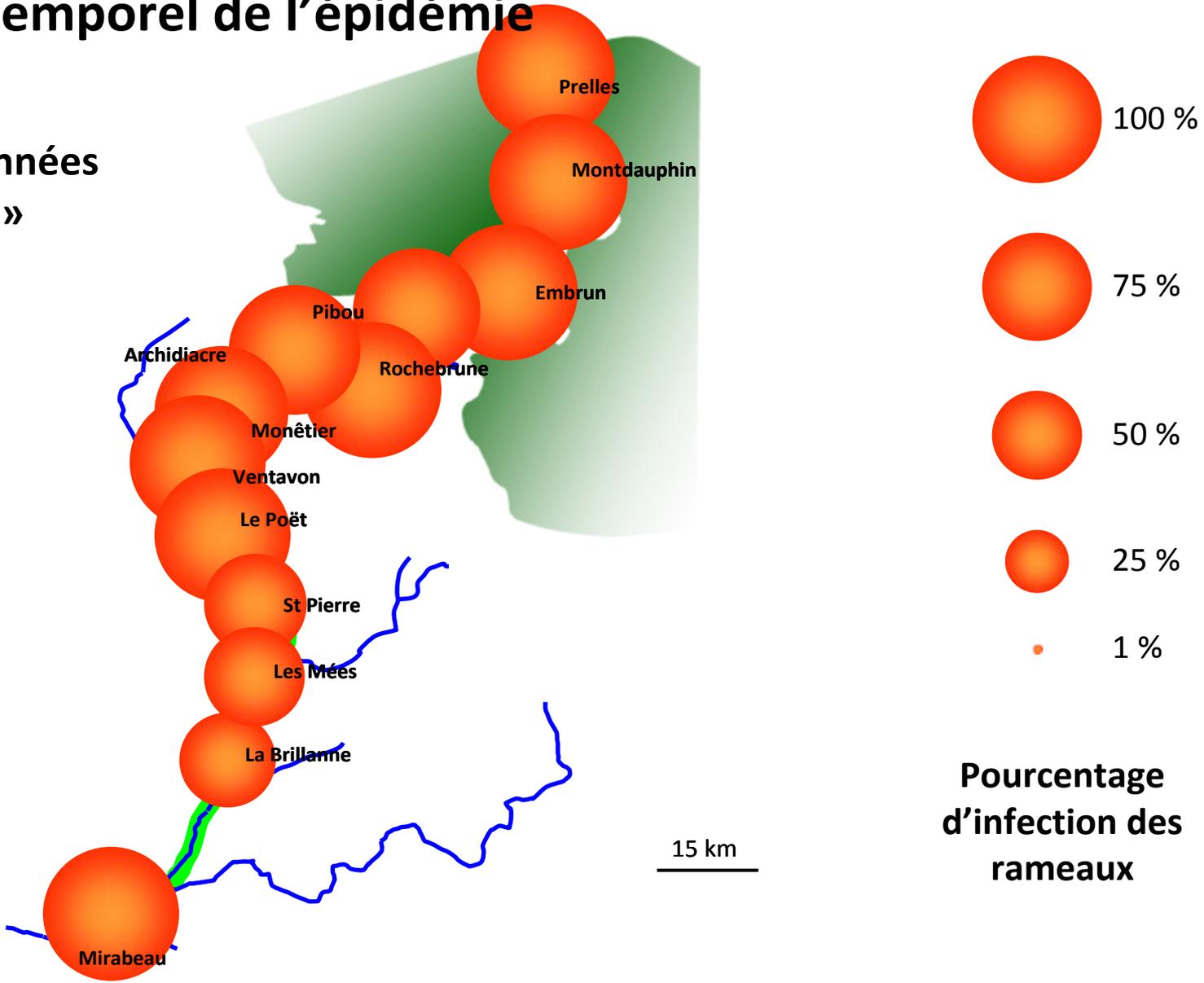
A partir des données  
« rameaux »



Tournée 6 (du 20 au 24 octobre)

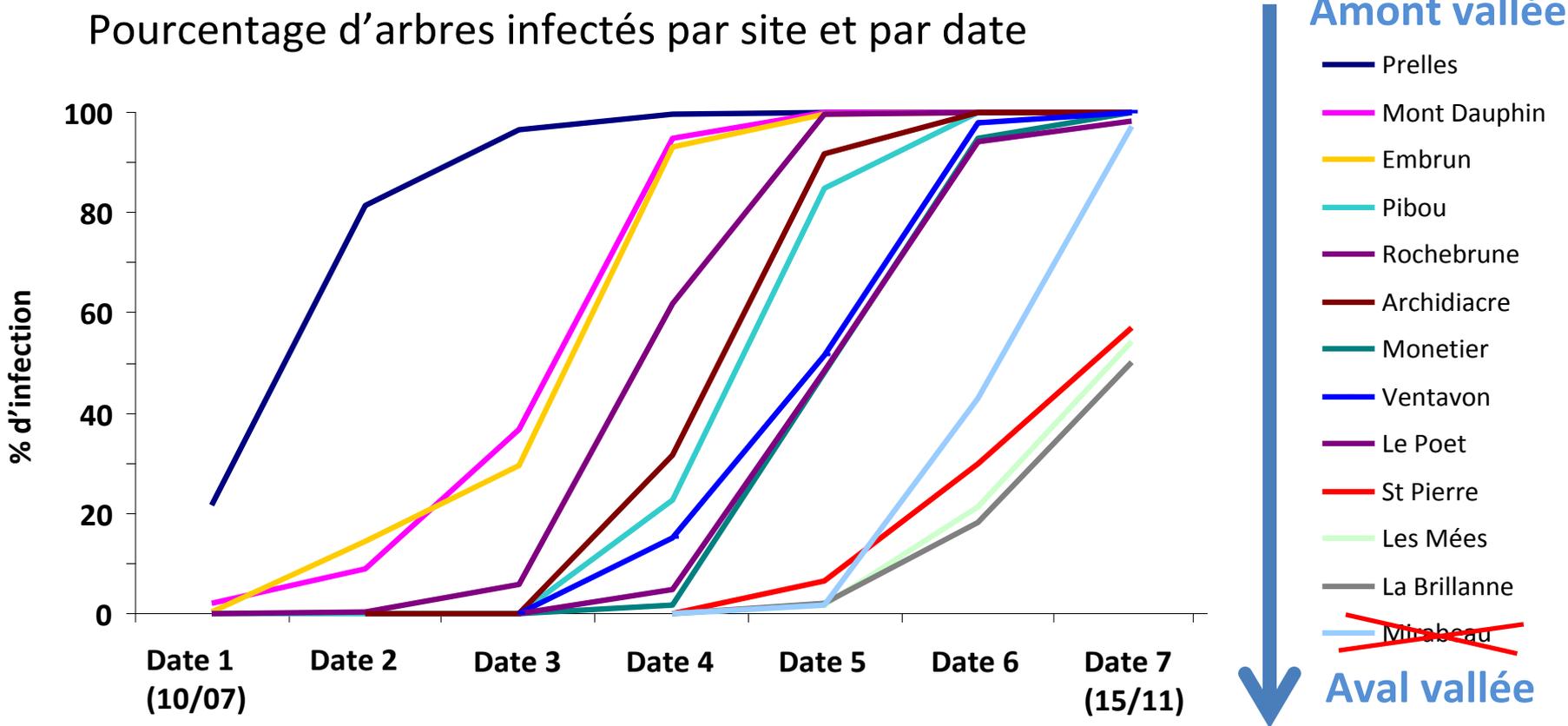
# Suivi spatio-temporel de l'épidémie

A partir des données  
« rameaux »



Tournée 7 (du 12 au 14 novembre)

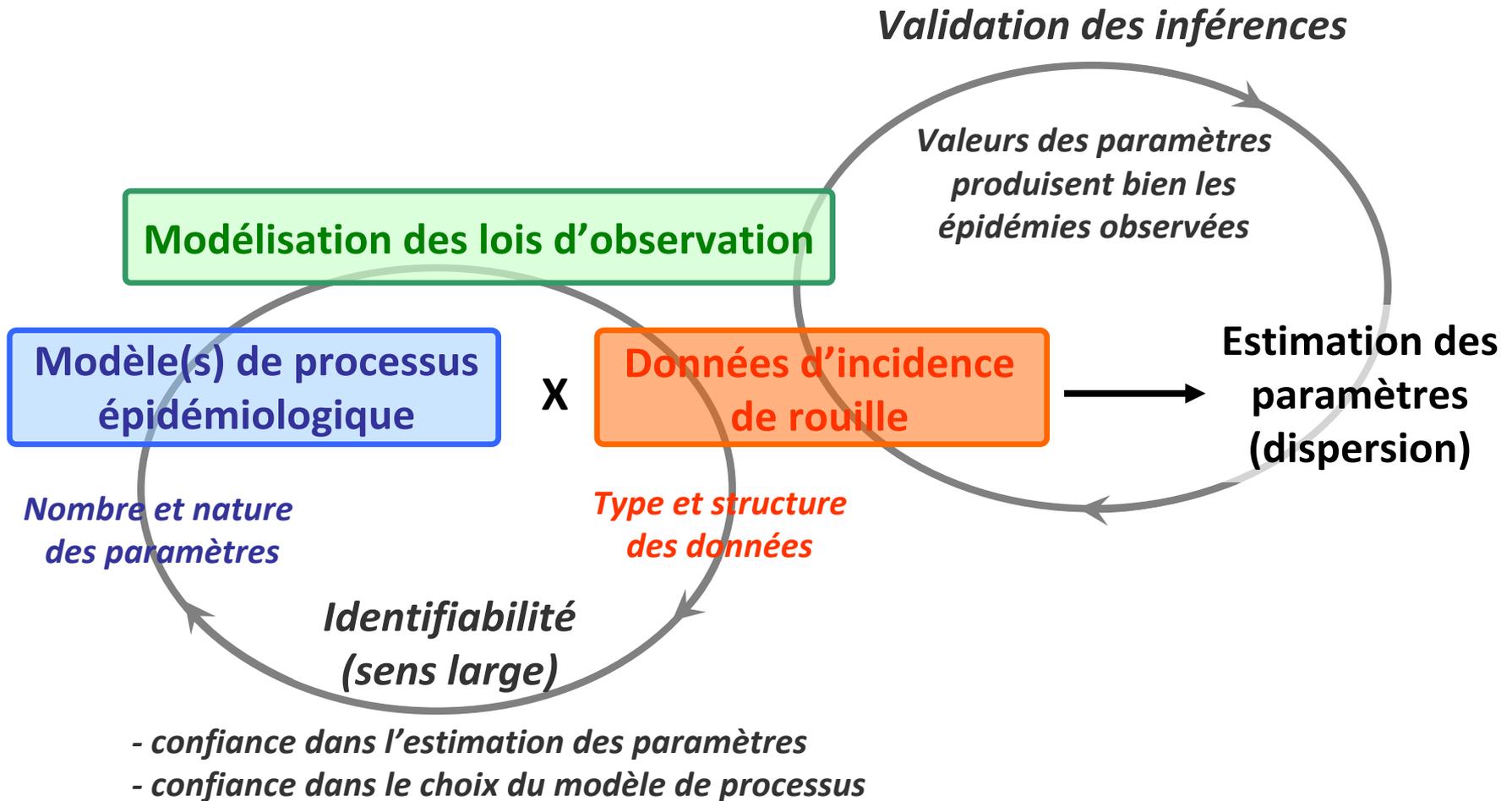
# Suivi spatio-temporel de l'épidémie



- Processus de colonisation globalement homogène entre sites
- Tous les sites sont saturés à la fin de la saison sauf ceux plus au sud

# Démarche pour l'estimation des paramètres

## Approche mécanico-statistique



# Modélisation : Modèle des processus épidémiologiques

Représentation des mécanismes épidémiologiques avec un modèle déterministe

variable «latente » :

$$\partial_t u(x, t) =$$

densité du nuage de spores actives au temps  $t$  au point  $x$

*Dispersion*  
 $D[u](x, t)$   
Opérateur de déplacement

+

*Réaction*  
 $ru(K - u)$   
Croissance locale (modèle logistique, Paramètre  $r$ )

# 4 Modèles de déplacement

## Déplacement diffusif

$$D[u] := \sigma \frac{\partial^2 u}{\partial x^2}(x, t)$$

Paramètre : coefficient de diffusion  $\sigma$

- **Avantages :**
  - Théorie mathématique très bien comprise
  - Numériquement « facile » et peu coûteux
- **Faiblesses :**
  - Modélise des flux de matières (et non des déplacements unitaires)
  - Une seule fonction de dispersion possible (vitesse constante du front de progression)

# 4 Modèles de déplacement

## Déplacement diffusif

$$D[u] := \sigma \frac{\partial^2 u}{\partial x^2}(x, t)$$

Paramètre : coefficient de diffusion  $\sigma$

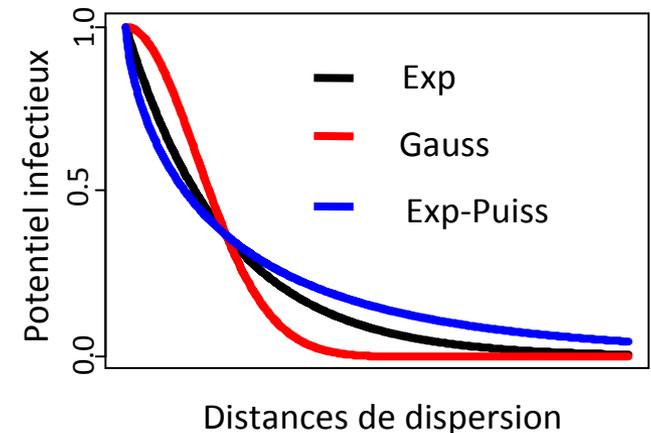
- **Avantages :**
  - Théorie mathématique très bien comprise
  - Numériquement « facile » et peu coûteux
- **Faiblesses :**
  - Modélise des flux de matières (et non des déplacements unitaires)
  - Une seule fonction de dispersion possible (vitesse constante du front de progression)

## Déplacement par noyau

$$D[u](x, t) := \int k_\lambda(y-x) [u(y, t) - u(x, t)] dy$$

Noyau de dispersion  $k_\lambda$  de paramètre  $\lambda$

- **Avantages :**
  - Modélise la re-répartition de la densité en individus (plus proche d'un mouvement)
  - Plusieurs formes de dispersion possible (ici 3)
- **Faiblesses :**
  - Numérique plus difficile à mettre en œuvre
  - Théorie moins bien établie

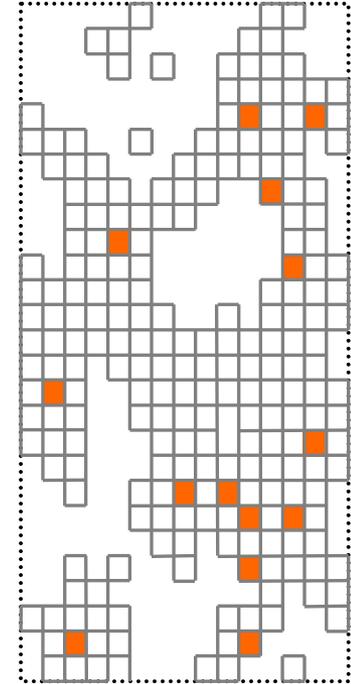


# Modélisation des lois d'observation

Lien entre la variable latente du modèle mécaniste ( $u(x,t)$ ) et les variables observées (manifestations indirectes des processus épidémiologiques)



Site d'étude



□ Habitat vide    ■ Habitat occupé

# Modélisation des lois d'observation

Lien entre la variable latente du modèle mécaniste ( $u(x,t)$ ) et les variables observées (manifestations indirectes des processus épidémiologiques)



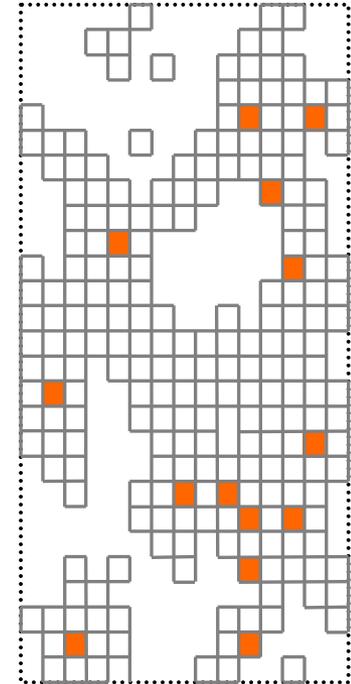
Nombre de lésions sur une feuille  $s$  localisée en  $x_s$

$$N_s(t) | u(x_s, t), R_s(t) \square \text{Pois}(u(x_s, t) R_s(t))$$

Où  $R_s(t) \square \text{Gam}(\sigma^{-2}, \sigma^2)$  décrit la réceptivité de la feuille (microclimat, position de la feuille dans l'arbre, âge...)

La réceptivité permet de modéliser la sur-dispersion des données si un nombre important d'habitats est observé

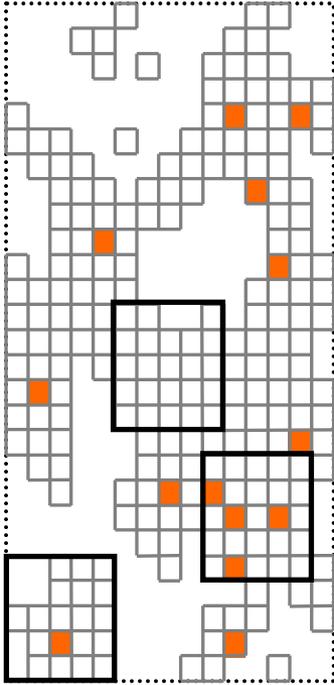
Site d'étude



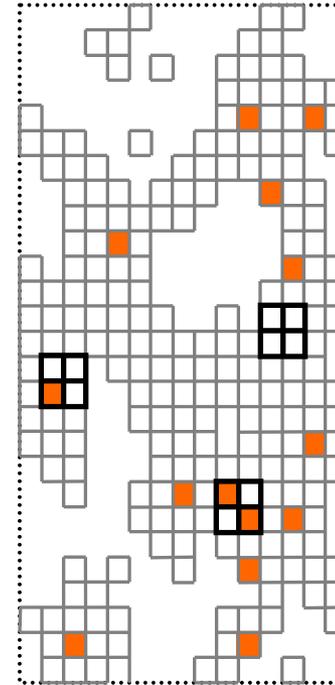
Habitat vide     Habitat occupé

# Un double échantillonnage

Echantillonnage 'Arbre'



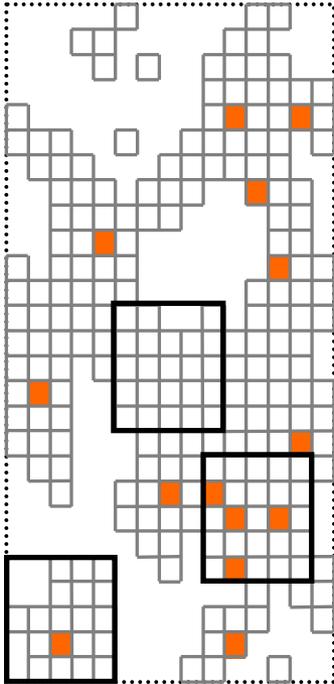
Echantillonnage 'Rameau'



Hypothèses différentes  
sur la variabilité des réceptivités  
au sein d'une unité  
d'échantillonnage

# Loi d'observation de l'infection au niveau 'arbre'

## Echantillonnage 'Arbre'



### Observations:

- Dans le site  $s$  au temps  $t$ , on observe  $M_{st}$  arbres dont  $Y_{st}$  sont infectés
- Sur chaque arbre, on regarde l'état de  $N_f$  ( $\approx 150$ ) feuilles
- Ces  $N_f$  feuilles constituent un échantillon iid de réceptivité (la portée de la dépendance spatiale des réceptivité est faible / à la taille d'un peuplier)

La probabilité  $p_{st}$  qu'une feuille localisée en  $x_s$  soit infectée :

$$\begin{aligned} p_{st} &= P(N_s(t) > 0 | u(x_s, t)) \\ &= 1 - (1 + u(x_s, t))^{-1/\sigma^2} \end{aligned}$$

Le nombre d'arbres infectés sachant  $u(x_s, t)$  suit une loi binomiale

$$Y_{st} | u(x_s, t) \square Bin\left(M_{st}, 1 - (1 - p_{st})^{N_f}\right)$$

# Loi d'observation de l'infection au niveau 'rameau'

## Observations :

- On observe  $G_{st}$  rameaux ayant chacun  $M_{stg}$  feuilles dont  $Y_{stg}$  sont infectés
- Un rameau est considéré comme un groupe d'habitat (feuille) homogène pour sa réceptivité (i.e. qui partage les mêmes facteurs de risques)

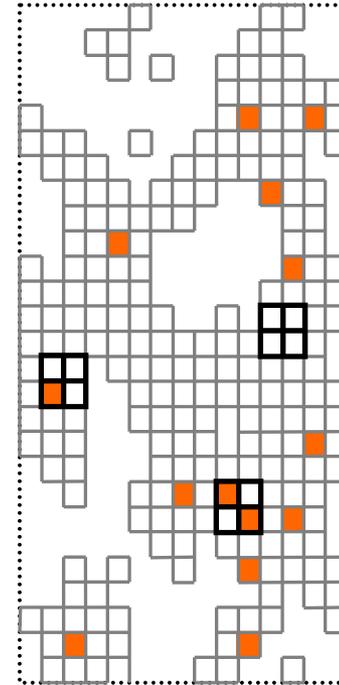
$Y_{stg}$  sachant  $R_g(t)$  et  $u(x_s, t)$  suit une loi binomiale

$$Y_{stg} \mid u(x_s, t), R_g(t) \square Bin(M_{stg}, 1 - e^{-u(x_s, t)R_g(t)})$$

et

$$u(x_s, t)R_g(t) \square Gam(\sigma^{-2}, u(x_s, t)\sigma^2)$$

## Echantillonnage 'Rameau'



**Le nombre de rameaux infectés sachant  $u(x_s, t)$  suit une loi de mélange gamma-binomiale**

$$P[Y_{stg} = y \mid u(x_s, t)] = \int_0^{\infty} f_{Bin}(M_{stg}, 1 - e^{-z})(y) f_{Gam}(\sigma^{-2}, u(x_s, t)\sigma^2)(z) dz$$

**Cette loi est une binomiale sur-dispersée  
(utile en épidémiologie .... mais difficile à manipuler !)**

# Identifiabilité « pratique » du modèle

Peut-on estimer les paramètres du modèle sachant les données dont on dispose (nature et quantité) ?

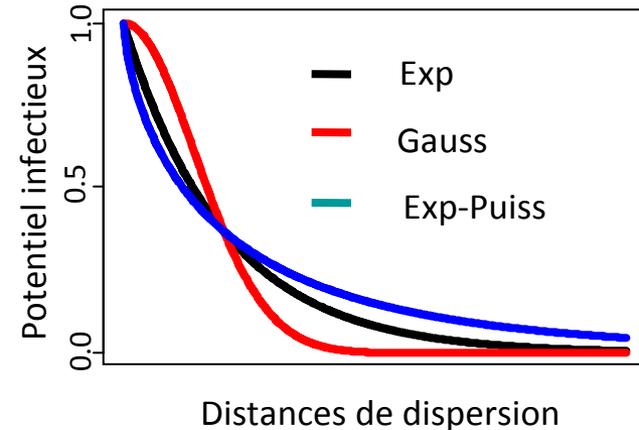
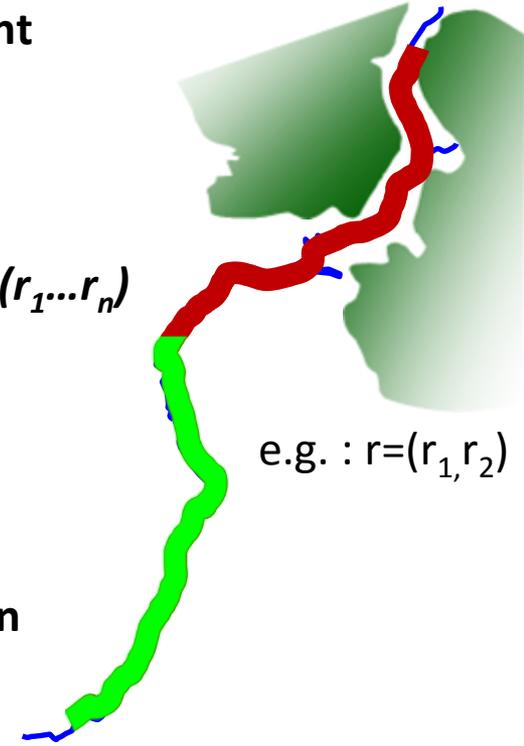
- On tire au hasard des paramètres « vrais »  $\vartheta=(r, \lambda, \tau, \sigma)$

(1) Taux de croissance le long de la vallée :  $r=(r_1 \dots r_n)$

(2) Fonctions de dispersion

(Exp, Gauss, Exp-Puiss, RD) :  $\lambda, \tau$

(3) Paramètre de sur-dispersion de l'infection :  $\sigma$



- On génère un jeu de données « plausible » selon  $\vartheta$  c'est-à-dire générant des épidémies sur 5 mois et 200 km telle que (1) % moyen arbre infectée à 30 jours  $< 50\%$  et  $> 20\%$  à 150 jours et (2)  $> 5\%$  des arbres soient infectées à 200 km à 150 jours

- On estime  $\vartheta$  sachant le jeu de données par maximum de vraisemblance

- On teste la corrélation entre les vraies valeurs et celles estimées (n=100 par modalité)

# Identifiabilité pratique du modèle

**Tableau1 :** Coefficient de corrélation entre  $\vartheta_{vrai}$  et  $\vartheta_{estim}$  ( $n=100$ ) pour  $r$  [0.01, 0.5],  $\lambda$  [0.1, 5],  $\tau$  [0.2, 0.7] et  $\sigma^2$  [0.1, 1000].

	Param	Exp	Gauss	Exp-Puiss	RD
1 taux de croissance $r=(r_1)$	$r$	<b>0.99</b> (0.001)	<b>0.99</b> (0.001)	<b>0.98</b> (0.01)	<b>0.99</b> (0.001)
	$\lambda$	<b>0.99</b> (0.004)	<b>0.99</b> (0.001)	<b>0.99</b> (0.001)	<b>0.99</b> (0.001)
	$\tau$			<b>0.98</b> (0.02)	
	$\sigma^2$	<b>0.99</b> (0.001)	<b>0.99</b> (0.001)	<b>0.99</b> (0.01)	<b>0.99</b> (0.002)
4 taux de croissance $r=(r_1-r_4)$	$r$	<b>0.84</b> (0.06)	<b>0.90</b> (0.06)	<b>0.64</b> (0.09)	<b>0.76</b> (0.1)
	$\lambda$	<b>0.99</b> (0.004)	<b>0.99</b> (0.001)	<b>0.99</b> (0.001)	<b>0.98</b> (0.009)
	$\tau$			<b>0.34</b> (0.2)	
	$\sigma^2$	<b>0.99</b> (0.001)	<b>0.99</b> (0.001)	<b>0.99</b> (0.007)	<b>0.99</b> (0.001)

- Identifiabilité « pratique » très bonne pour 1 taux de croissance
- Plus difficile quand le nombre de taux augmente surtout pour la fonction Exp-Puissance (moins bonne estimation du paramètre de décroissance  $\tau$ )

# Apport de l'échantillonnage multi-types

**Tableau2** : Coefficient de corrélation entre  $\vartheta_{vrai}$  et  $\vartheta_{estim}$  ( $n=100$ ).

	Param	Exp	Gauss	Exp-Puiss	RD
Ech Arbre (200 arbres)	$r$	<b>0.99</b> (0.002)	<b>0.99</b> (0.001)	<b>0.99</b> (0.001)	<b>0.99</b> (0.002)
	$\lambda$	<b>0.98</b> (0.005)	<b>0.99</b> (0.001)	<b>0.96</b> (0.01)	<b>0.98</b> (0.005)
	$\tau$			<b>0.99</b> (0.002)	
	$\sigma^2$	<b>0.93</b> (0.03)	<b>0.98</b> (0.01)	<b>0.96</b> (0.01)	<b>0.86</b> (0.05)
Ech Arbre (100 arbres) + Ech Rameaux (20 rameaux)	$r$	<b>0.99</b> (0.001)	<b>0.99</b> (0.001)	<b>0.98</b> (0.01)	<b>0.99</b> (0.001)
	$\lambda$	<b>0.99</b> (0.001)	<b>0.99</b> (0.001)	<b>0.99</b> (0.001)	<b>0.99</b> (0.001)
	$\tau$			<b>0.98</b> (0.02)	
	$\sigma^2$	<b>0.99</b> (0.001)	<b>0.99</b> (0.001)	<b>0.99</b> (0.001)	<b>0.99</b> (0.002)
Ech Rameaux (40 rameaux)	$r$	<b>0.95</b> (0.01)	<b>0.99</b> (0.005)	<b>0.8</b> (0.1)	<b>0.92</b> (0.04)
	$\lambda$	<b>0.92</b> (0.03)	<b>0.98</b> (0.07)	<b>0.82</b> (0.17)	<b>0.77</b> (0.1)
	$\tau$			<b>0.80</b> (0.24)	
	$\sigma^2$	<b>0.99</b> (0.001)	<b>0.99</b> (0.003)	<b>0.99</b> (0.001)	<b>0.99</b> (0.001)

- L'échantillonnage « rameaux » procure l'information sur  $\sigma^2$  au détriment de  $r$
- L'échantillonnage multi-type améliore sensiblement les inférences

# Confiance dans la sélection du processus de dispersion

**Tableau 3 :** Fréquence des cas où le vrai modèle est sélectionné (AIC) ( $n=50$ ).

Vrai \ Select	Exp	Gauss	Exp-Puiss	RD
Exp	<b>0.70</b>	0.26	0	0.04
Gauss	0.32	<b>0.64</b>	0.02	0.04
Exp-Puiss	0.14	0.02	<b>0.82</b>	0.02
RD	0.02	0.08	0	<b>0.90</b>

- Identification du vrai modèle avec une freq. > 0.6
- La fonction Exp-Puiss (décroissance lente) est bien identifiée, avec des écart d'AIC très importants (>400)
- Confusion plus fréquente entre les fonctions à décroissance rapide (gaussien et exp)
- Très bonne identification du modèle de diffusion / dispersion par noyau

# Inférences sur les données réelles

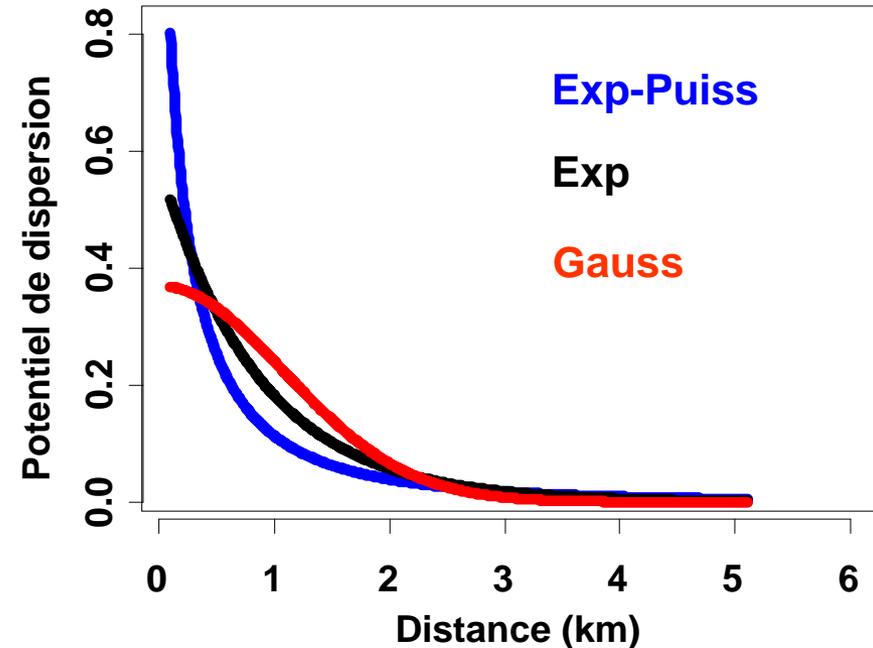
**Tableau 4 :** Sélection de modèle (AIC) sur l'épidémie de rouille de 2008

	Exp	Gauss	Exp-P	RD
$r=(r_1)$	3848	3864	2848	3864
$r=(r_1, r_2)$	3386	3410	2848	3424
$r=(r_1, r_2, r_3)$	3436	3468	<b>2692</b>	3488
$r=(r_1, r_2, r_3, r_4)$	3384	3406	2716	3422

**Tableau 5 :** Estimation des paramètres

	Paramètre	Estimation (med)
<b>Modèle Sélectionné</b>	$r_1$ (aval)	0.061
	$r_2$	0.073
	$r_3$ (amont)	0.05
	$\lambda$	0.078
	$\tau$	0.15
	$\sigma^2$	10.55

Fonction de Dispersion



# Remerciements

Axelle Andrieux

Bénédicte Fabre

Benoît Marçais

Béranger Bertin

Christine Gehin

Claude Hüsson

Pascal Frey

Olivier Caël

