

**Monitoring the risk of *Legionella* infection
using graphical independence network
updated from temporal measurements and
applied to agricultural irrigation with reclaimed
wastewater.**

Gaspar Massiot

AgroParisTech Nancy, EcoSilva, UMR Silva

Isabelle Albert

Inrae AgroParisTech Paris-Saclay, SOLsTIS, UMR 518 Mia-Paris

Dominique Courault

Inrae, DREAM, UMR 1114 EMMAH

Applibugs 10 Décembre 2021

- **Projet** : SmartFertiReuse (Smart Ferti-irrigation et RÉUtilisation des eaux USÉes traitées) coordonné par VEOLIA Recherche et Innovation.
- **Objectif du projet** : Concevoir un service de ferti-irrigation.
- Partenaires Académiques :

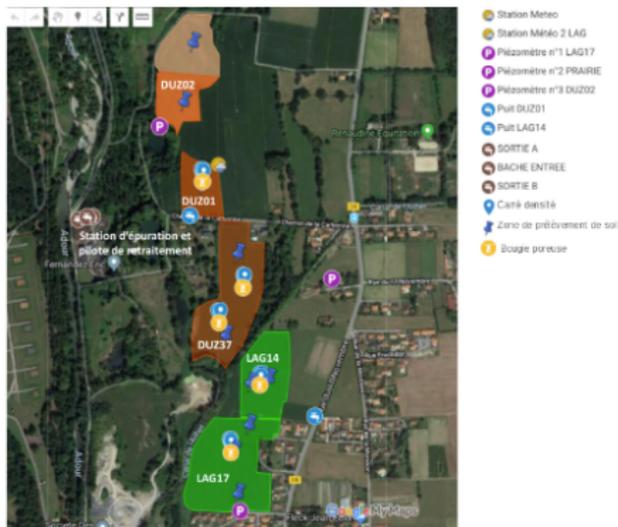


- Partenaires Industriels :



Site d'étude : Aureilhan (65)

Emplacement des parcelles et des pilotes de recherche



- 2 pilotes de recherche
- 3 parcelles d'expérimentation (2 ha ; 3 ha ; 2.5 ha)
- 2 parcelles témoin (4.7 et 9.7 ha)
- 2 puits d'irrigation en eau de nappe
- 3 piézomètres

Objectif : Comparer deux filières de filtration

Filière qualité A (50 m³/h) → en service 2020

Eau usée brute
entrée STEP



Eau usée
traitée sortie
STEP/entrée
pilote



Préfiltres 130
µm



Ultrafiltration



Réacteur
UV

Chloration

Bâche
sortie
pilote
50 m³

Deux modes :

Recherche (rejet Adour sans chloration) et irrigation (avec chloration)

Filière qualité B (50 m³/h → 75 M³/h en lavage) → en service 2019

Filtre Amiad
autonettoyant
40 µm puis en 2021:
80 µm



Réacteurs UV



Chloration

Bâche
sortie
pilote
50 m³



Expérimentations

Prélèvements à différents niveaux:

Eau brute
Entrée de STEP



En sortie
des pilotes



Après stockage en bache



Dans l'air
par impacteur
pendant l'irrigation



À la sortie du canon
d'irrigation

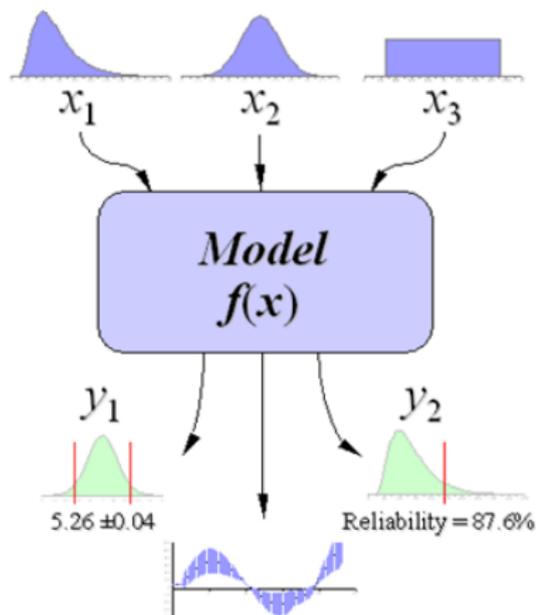
2 parcelles suivies avec des conduites culturales différentes



- Mise en place d'une **méthodologie de suivi des risques sanitaires** liés aux pathogènes émis lors de l'irrigation et propagés :
 - ▷ par aspersion et dispersion dans l'air (bioaérosols) ;
 - ▷ par transfert de la surface irriguée jusque dans la nappe.

La méthode la plus répandue pour l'évaluation du risque utilise la méthode de Monte Carlo :

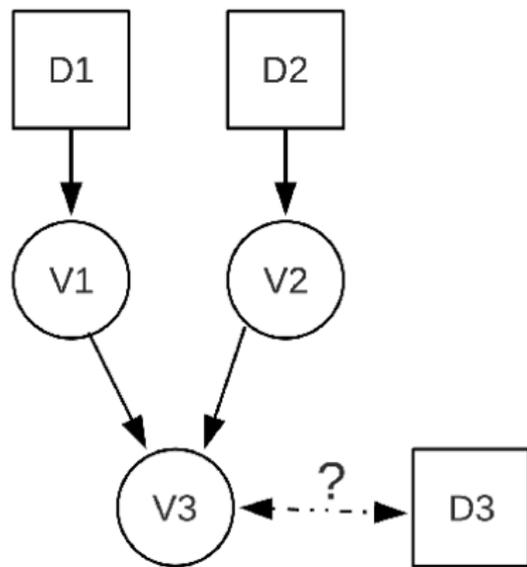
- Un modèle empirique est utilisé pour simuler l'influence de différents paramètres sur le risque et le quantifier ;
- À chaque paramètre du modèle est associée une loi de probabilité correspondant à la prise en compte de la variabilité et/ou l'incertitude associées à ce paramètre ;
- Des simulations de type Monte Carlo permettent d'estimer les distributions marginales des variables d'intérêt.



Les limites de la méthode tiennent à :

- La définition du modèle numérique le plus adéquat pour représenter la réalité ;
- La construction des lois de probabilité sur les paramètres ;
- La non-possibilité de prendre en compte l'information en aval de la chaîne modélisée dans l'évaluation

Exemple



Simulation Monte-Carlo

- D1 : données de contamination de l'eau
- V1 : probabilité de contamination de l'eau
- D2 : données d'exposition à l'eau contaminée
- V2 : intensité d'exposition à l'eau contaminée
- V3 : probabilité de maladie
- D3 : nombre de malades

Propositions

Limite : la définition du modèle numérique le plus adéquat pour représenter la réalité

- **Construction globale du modèle par réseau bayésien**
- **Validation partielle par inférence bayésienne**

Limite : la construction des lois de probabilité sur les paramètres

- **Séparation données / modèle expert**
- **Élicitation de dires d'experts**

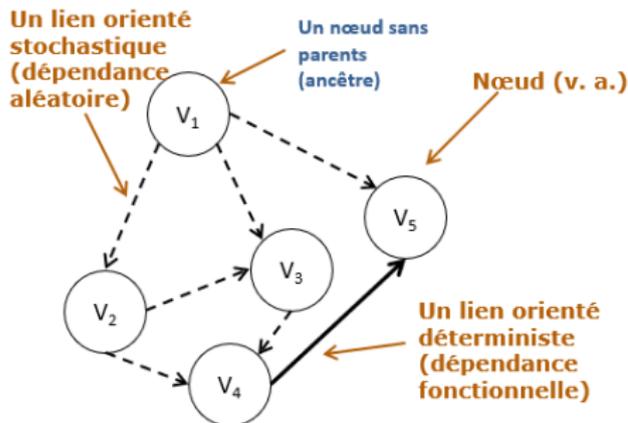
Limite : la non-possibilité de prendre en compte l'information en aval de la chaîne modélisée

- **Inférence bayésienne, synthèse bayésienne**

Graphical independance network / Réseau bayésien à variables continues

Un modèle graphique donnant une représentation des dépendances et indépendances entre les variables aléatoires (v.a.)

$G = (\Gamma, \Sigma)$, graphe orienté acyclique (DAG), composé d'un ensemble de nœuds représentant des v.a. et un ensemble d'arcs orientés sans cycle traduisant les liens entre ces v.a. par des lois de probabilité conditionnelles. Soit $V = (V_i)_{i \in \Gamma}$ l'ensemble des v.a. du DAG



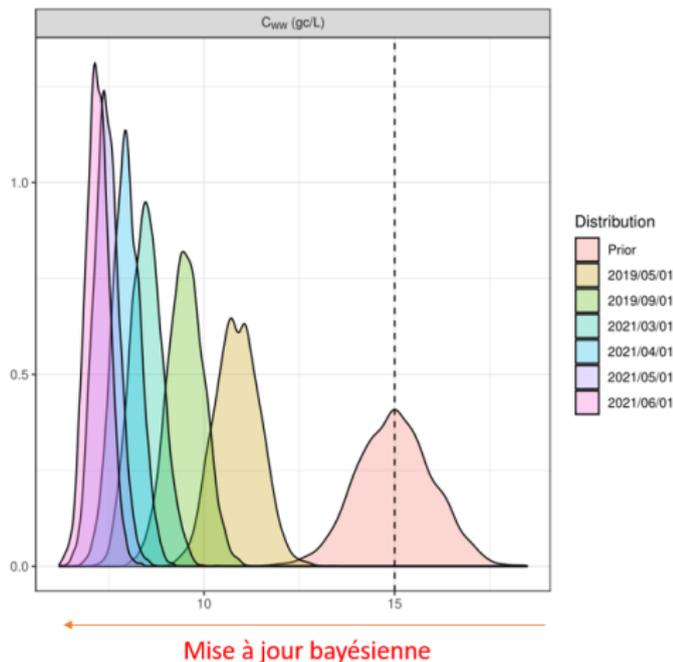
Indépendance conditionnelle

$$V_i \perp\!\!\!\perp V_{\Gamma \setminus \text{desc}(i)} \mid V_{\text{pa}(i)}$$

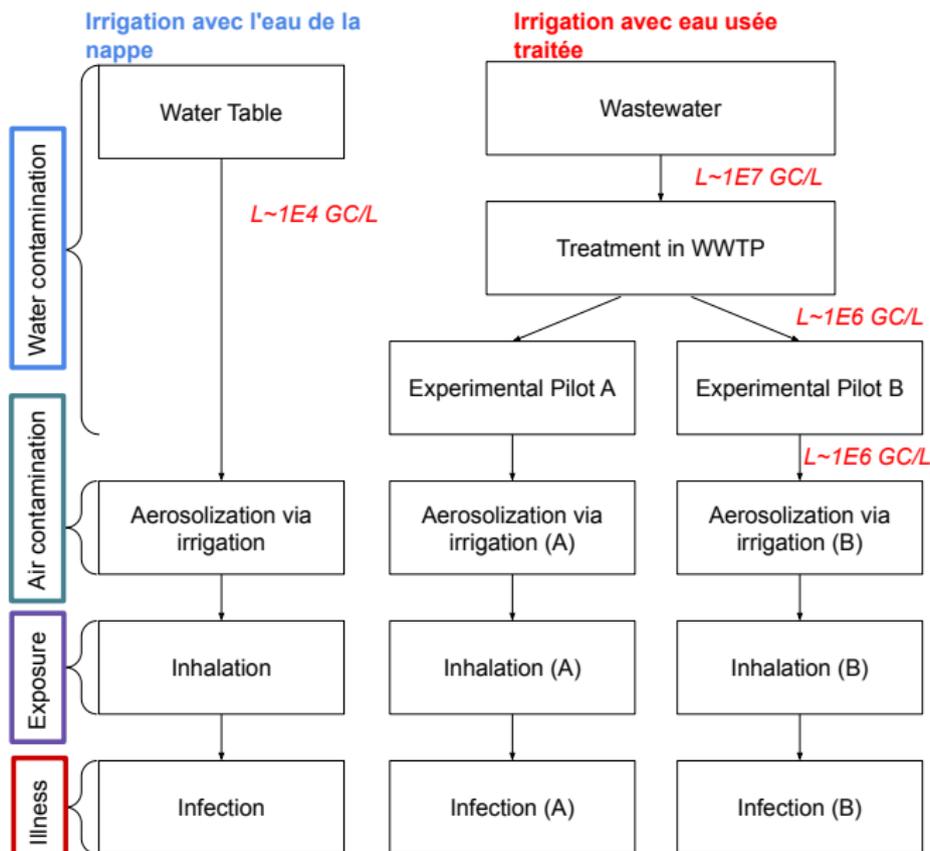
Factorisation de la loi jointe

$$p(V) = \prod_{i \in \Gamma} p(V_i \mid V_{\text{pa}(i)})$$

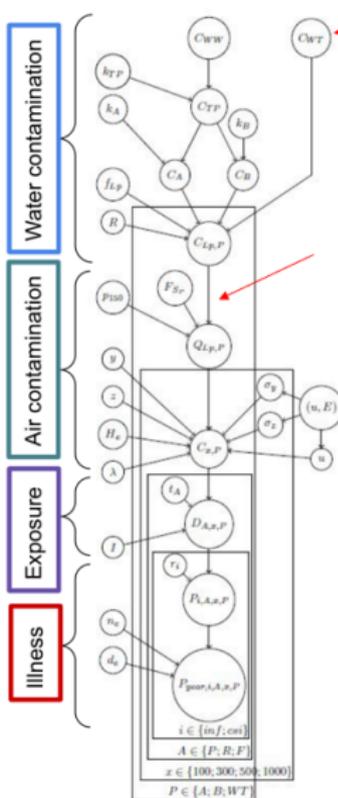
- Approche Bayésienne par MCMC (Markov Chain/Monte Carlo)
 - ▷ Permet d'exprimer ce que l'on sait sur les inconnues à partir des données (ce qui est connu)
 - ▷ Algorithme séquentiel : chaque itération comporte une étape de mise à jour de la chaîne des paramètres
- Part d'une information *a priori* pour en déduire la loi *a posteriori* par les données
- Réduit l'incertitude sur les inconnues.



Construction du modèle



- **Water contamination :**
 - ▷ Modélisation du traitement de l'eau usée avant irrigation
- **Air contamination :**
 - ▷ Modélisation de la dispersion des pathogènes dans l'air
- **Exposure :**
 - ▷ Modélisation de l'exposition de différentes populations (agriculteurs, résidents, passants)
- **Illness :**
 - ▷ Modélisation de la probabilité d'infection par un modèle dose-réponse



La distribution des nœuds ancêtres est déterminée par **élicitation** de dires d'experts

Lien **déterministe** : ex. : modèle de dispersion atmosphérique

- Modèle de dispersion atmosphérique :

$Q_{Lp,P}$ = taux d'émission de Legionella Pneumophila, u : vitesse du vent

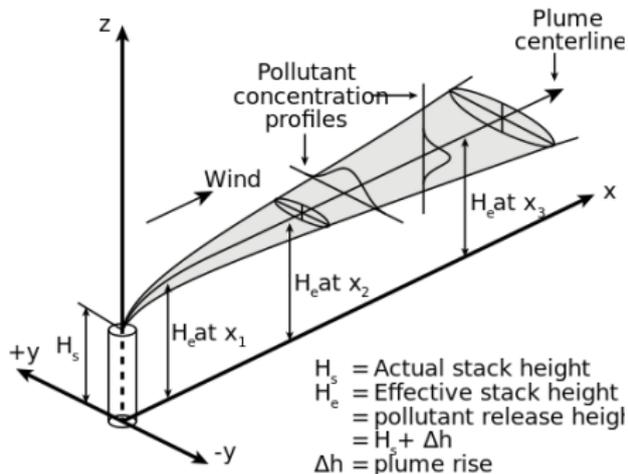
$$C_{x,P} = \frac{Q_{Lp,P}}{2\pi u \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \exp\left(-\frac{\lambda x}{u}\right) \times \left\{ \exp\left(-\frac{(z-H_e)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+H_e)^2}{2\sigma_z^2}\right) \right\}$$

$$\Rightarrow \theta_{disp} = (\lambda, H_e)$$

- Modèle dose-réponse exponentiel :
 D : dose d'exposition

$$P_{int} = 1 - \exp(-r_i D)$$

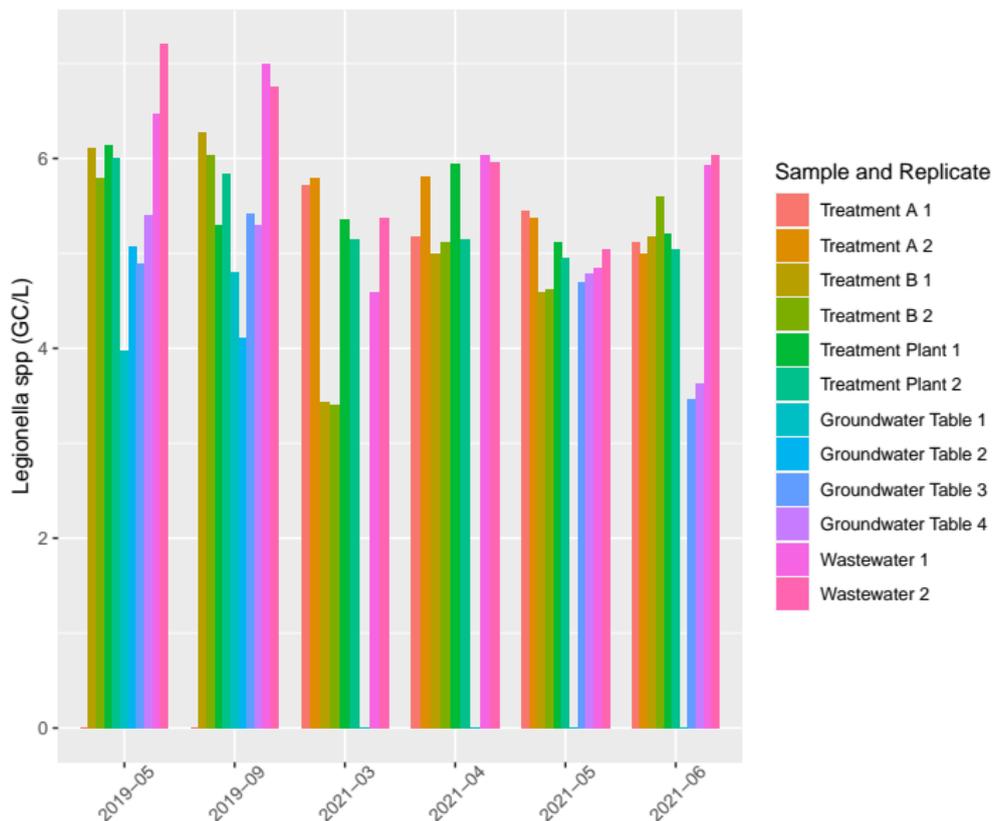
$$\Rightarrow \theta = r_i$$



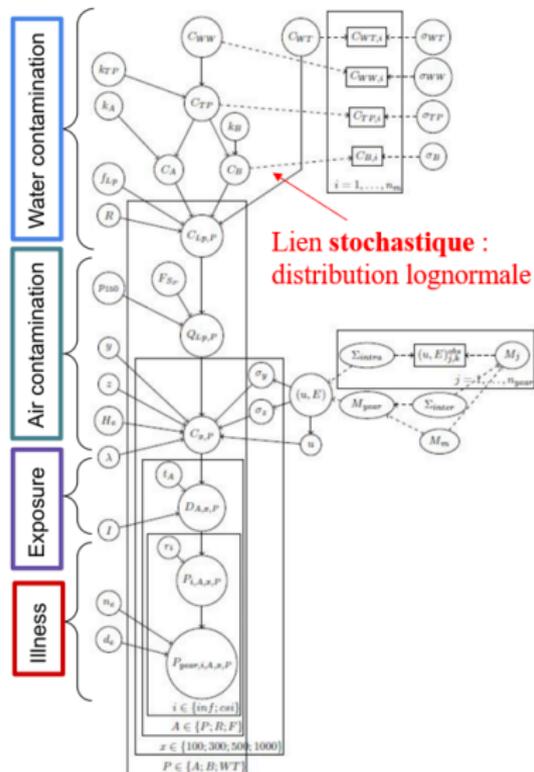
Priors on the Water contamination block of the Core model

Description	Sym- bol	Unit	Value	Distri- bution
Water quality in water table	C_{WT}	gc/L	$\mu = 10,$ $\sigma = 1$	Lognor- mal
Wastewater quality before any treatment	C_{WW}	gc/L	$\mu = 15,$ $\sigma = 1$	Lognor- mal
Log removal value of the treatment in WWTP	k_{TP}	–	$\mu = 8,$ $\sigma = 0.1$	Normal
Log removal value of the treatment by experimental pilot <i>A</i>	k_A	–	$\mu = 4,$ $\sigma = 0.1$	Normal
Log removal value of the treatment by experimental pilot <i>B</i>	k_B	–	$\mu = 3,$ $\sigma = 0.1$	Normal

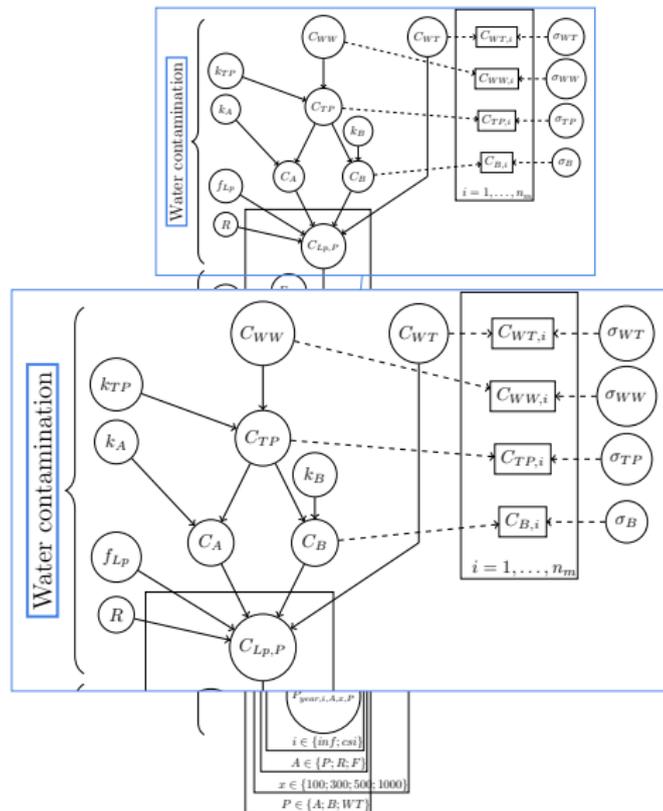
Données de concentration



- Réseau Bayésien
- Suivi "continu" des données
- Données **microbiologiques** de contamination de l'eau (six dates 21 Mai, 17 Septembre 2019, et une date par mois de Mars à Juin 2021)
- Données **météorologiques locales** (Météo France de Mai 2010 à Septembre 2017 et Inrae Avril 2018 à Juin 2020)
- Données **épidémiologiques nationales**



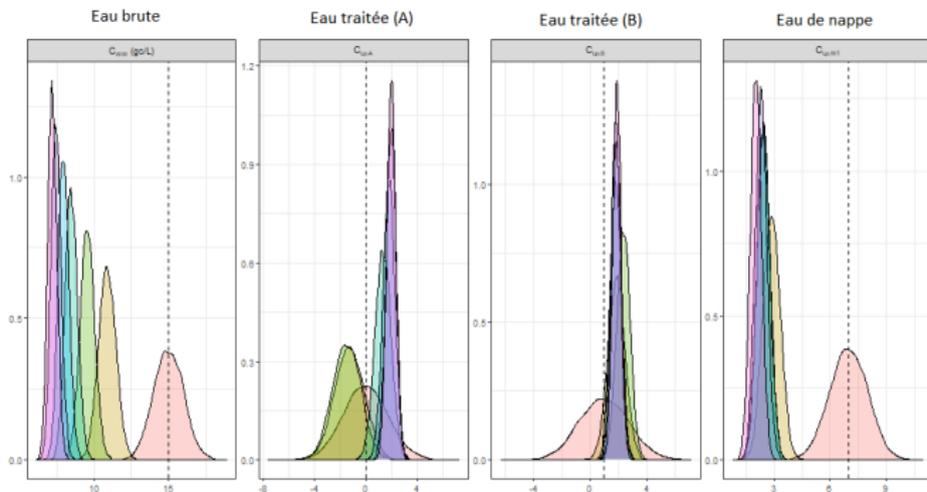
Augmented model



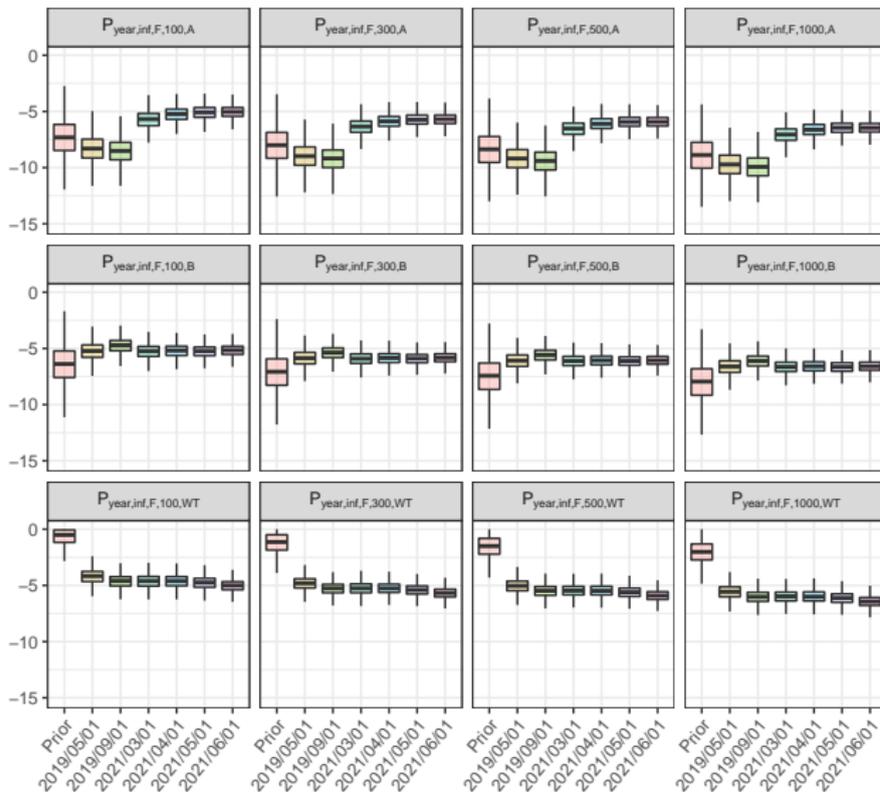
Priors on the **Water contamination** block of the **Augmented model**

Variable X	Distribution
$C_{WT,i}; C_{WW,i}; C_{TP,i}; C_{A,i}; C_{B,i}$ μ_P σ_P	$X \sim \log \mathcal{N}(\mu_P, \sigma_P)$ $X = \log(C_P)$ $1/X^2 \sim \text{Gamma}(10, 1)$

- MCMC simulations



- Modèle dose réponse exponentiel : $Prob = 1 - \exp(-r * dose)$



- Probabilités d'être infecté à différentes distances
100m 300m 500m 1km
Pour 3 catégories de personnes
- Seuil maximal acceptable (US EPA) : probabilité d'infection annuelle de $-4 \log_{10}$.

	$P_{year,inf,F,100,A}$	$P_{year,inf,F,100,B}$	$P_{year,inf,F,100,WT}$
Prior	-7.31 [-8.47;-6.16]	-6.38 [-7.58;-5.22]	-0.51 [-1.17;-0.06]
2019/05/01	-8.3 [-9.14;-7.47]	-5.24 [-5.79;-4.69]	-4.17 [-4.66;-3.75]
2019/09/01	-8.51 [-9.3;-7.75]	-4.72 [-5.21;-4.29]	-4.6 [-5.05;-4.23]
2021/03/01	-5.68 [-6.27;-5.17]	-5.25 [-5.71;-4.82]	-4.61 [-5.06;-4.21]
2021/04/01	-5.23 [-5.7;-4.79]	-5.19 [-5.63;-4.8]	-4.61 [-5.03;-4.22]
2021/05/01	-5.08 [-5.52;-4.65]	-5.25 [-5.65;-4.88]	-4.74 [-5.17;-4.37]
2021/06/01	-5.04 [-5.43;-4.65]	-5.16 [-5.54;-4.81]	-5 [-5.39;-4.67]

Augmented model - Illness

- Probabilités d'être infecté à différentes distances

100m 300m 500m 1km

Pour 3 catégories de personnes

- Seuil maximal acceptable (US EPA) : probabilité d'infection annuelle de $-4 \log_{10}$.

Distance <i>m</i>	+ Farmer			← Resident			- Passerby		
	A	B	WT	A	B	WT	A	B	WT
100	-3.23	-3.36	-3.19	-4.05	-4.17	-4.01	-4.72	-4.84	-4.7
300	-3.86	-3.99	-3.83	-4.69	-4.82	-4.66	-5.36	-5.49	-5.33
500	-4.09	-4.21	-4.05	-5.44	-5.56	-5.41	-5.59	-5.71	-5.55
1000	-4.61	-4.74	-4.59	-5.44	-5.56	-5.41	-6.12	-6.24	-6.09

Conclusion

- Elaboration d'un modèle qui permet d'échanger avec les experts pour la construction des priors,
- Prise en compte de l'incertitude et sa propagation vers les paramètres du modèle
- Suivi continu des données et remise à jour
- Généricité qui permet d'avoir le même schéma de base pour d'autres pathogènes (adaptation aux spores, dose réponse, . . .)

Perspectives

- Inclusion de données épidémios,
- Modélisation du transfert de pathogènes dans le sol jusque dans la nappe (couplage d'un modèle de la plateforme sol virtuel INRAE à un modèle QMRA, travail en cours)
- Adaptation du modèle à des problématiques multi-pathogènes.