Monitoring the risk of Legionella infection using a general Bayesian network updated from temporal measurements in agricultural irrigation with reclaimed wastewater

Gaspar Massiot, Dominique Courault, Pauline Jacob, Isabelle Albert

Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

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

Gaspar Massiot*[a,b], Dominique Courault[c], Pauline Jacob[d], Isabelle Albert[b]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Reuse of reclaimed wastewater for agricultural irrigation is an expanding practice worldwide, especially in water-stressed areas. This practice needs to be monitored, partly because of pathogens that water may still contain after treatments. More particularly, sprinkler irrigation is known to generate aerosols which may lead to severe health risks on the population close to irrigated areas in the case of presence of *Legionella* bacteria in the water. Therefore, a pilot experiment was conducted on two corn fields in South-Western France, irrigated with wastewater undergoing two different water treatments (ultra-filtration and UV). Water analyses have shown high levels of *Legionella* in the water even after these two tertiary treatments (up to 10^6 GC/L). In this context, we proposed to use an updated Bayesian Network (BN) in Quantitative Microbial Risk Assessment (QMRA) to monitor the risk of *Legionella* infection in the vicinity of the irrigated plots. The model's originality is based on i) a graphical probabilistic model that describes the pathway of *Legionella* from the WasteWater Treatment Plant (WWTP) to the population using observed and latent variables and ii) the model inference updating at each new available measurement *in situ*. Different scenarios are simulated according to the exposition time of the persons, taking into account various distances from the emission source and a large dataset of climatic data. From the learning process included in the Bayesian principle, quantities of interest (contaminations before and after water treatments, inhaled dose, probabilities of infection) can be quantified with their uncertainty before and after the inclusion of each new data collected *in situ*. Thus, this approach gives a rigorous tool that allows to monitor the risks, facilitates discussions with reuse experts and progressively reduces the uncertainty quantification through field data accumulation. For the two pilot treatments analyzed in this study, median annual risk of *Legionella* infection did not exceed the US EPA annual infection benchmark of 10^{-4} for any of the population at risk during the past monthes of the pilot experiment. Nevertheless the risk still bears watching with support from the method shown in this work.

1 Introduction

In the context of climate change, growing global water scarcity and more severe and frequent drought events have intensified the need to find alternative water resources [1,2]. Reclaimed wastewater is a renewable resource from which renewable energy can be produced, but also materials and water can be recovered [3,4]. It is a possible alternative supply to address such water shortages [5].

The combination of climate change, population growth and urbanisation has lead the European Parliament to foster increased water reuse to alleviate the stress on freshwater supply and prevent the fall in groundwater levels, due in particular to agricultural irrigation [6] which represented more than half of the water used annually in Europe in 2017 according to the European Environment Agency. [7]

However, the wastewater reuse still remains uncommon in some countries. Several factors can explain this observation among them the pathogens that water may contain, which resulted in strict reuse guidelines. Adegoke, *et al.* (2018) [8] have reported the different pathogens found in wastewaters and the associated risks. Quality indicators are commonly used, but they are more and more debated on their representativeness (particularly for

*corresponding author.
[a]UMR AgroParisTech/INRAE Silva, INRAE-AgroParisTech-Université de Lorraine, 54000, Nancy, France.
[b]UMR 518 AgroParisTech/INRAE MIA-Paris Saclay, INRAE-AgroParisTech-Université Paris-Saclay, Cedex 05, 75231 Paris, France.
[c]UMR 1114 EMMAH, INRAE, Site Agroparc, CS 40509, 84914 Avignon, France.
[d]Veolia Recherche & Innovation, Chemin de la Digue, 78600 Maisons-Laffitte, France.

viruses and *Legionella e.g.*, they are not explicitly taken into account as indicators despite their wide occurrence in the environment). Pragmatic recommendations and guidelines were defined in WHO (1989) and revised in 2006 on the basis of health risk considerations. The new European regulations on minimum requirements for water reuse for agricultural irrigation will apply from 26 June 2023 (EC, 2021).

Sprinkler irrigation with reclaimed wastewater is widely practiced throughout the world but may enhance hazards relative to the dispersion of pathogens by bioaerosols.[9–11] A collective expertise appraisal has been performed by ANSES in 2012, which assessed various secondary risks (e.g. ingestion) and pointed out the lack of knowledge on health risks, particularly due to pathogen inhalation.

An experiment has been conducted in South-Western France to evaluate the impact of two different treatments on the water quality, reused for sprinkler irrigation of two corn fields (SmartFertiReuse project, https://www.smartfertireuse.fr/) controlling the population risks. Among the different analyzed pathogens in the water, *Legionella* was found at high level both at the entrance and output of the WasteWater Treatment Plant (WWTP) (up to 10^7 GC/L). In the new EU regulations, criteria have been defined for water quality. A minimum required values for *Legionella* has been set to <1000 CFU/L for a reclaimed water class A used for food culture irrigation. These rules are expected to stimulate and facilitate water reuse in the EU and shall also be accompanied by appropriate monitoring of the health risks linked to this practice.

Legionella is an opportunistic bacteria which may be inhaled by human through water aerosol.[12–14] *Legionella* has around 50 species among them *Legionella pneumophila* which is responsible of the most severe infections such as Legionnaire's disease and pneumonia illness (Pontiac fevers)[15]. These infections due to *Legionella pneumophila* are particularly problematic for people who have a certain fragility or those who are exposed for a long period to these pathogens and may cause in some mortality cases. In France, an increasing trend of Legionellosis is observed since 1995*. A lot of cases were linked to cooling towers[16] but some studies have shown contaminated aerosols by *Legionella pneumophila* in the vicinity of WWTP[17] and associated diseases have been observed for example in Netherlands.[18]

Different approaches have been proposed to quantify health risk applied to *Legionella* (reviewed by Hamilton, *et al.* (2018) and Hamilton, *et al.* (2016)[13,14]). Often there are few observations and numerous authors have shown a wide variability in both space and time of the punctual measurements.[12,19] It is therefore important to take into account this variability and the uncertainty in risk assessment.

Quantitative Microbial Risk Assessment (QMRA) is an established framework recommended by multiple international organizations (WHO, US EPA, EFSA) to monitor such risks of infection.[20] QMRA models have been widely used in the risk assessment of waterborne disease,[21–23] for bioaerosol risk[9,24–27] and for food safety management[28] and are often coupled with Monte Carlo (MC) simulation approaches which consist of drawing values for all the inputs in the model from distributions defined by external knowledge (e.g., experimental or historical data, expert knowledge) and then propagating these values to all the intermediate variables and outputs through the system in a way that takes the variability and uncertainty of the inputs into account. It is, by definition, a unidirectional approach that prevents the inversion of the dependency relationship between the variables.

Bayesian Networks (BNs), known also as Graphical Independence Networks in the context of continuous random variables, are an innovative technique in the context of evidence synthesis[29,30]. These models can be applied to a wide range of application domains such as environmental modelling,[31] artificial intelligence[32] and also to quantitative microbial risk assessment (for a review, see Beaudequin, *et al.* (2015) and Verbyla, *et al.* (2016)[33,34]). Bayesian networks are based on directed acyclic graph (DAG) where conditional dependencies between the variables are qualitatively represented by the directed arcs on the graph[35] and quantitatively defined with local transition probabilities or deterministic equations. Bayesian Networks approaches in QMRA provide a powerful framework where the dependency relationship between variables can be inverted ('backward calculation', by which the states of the model's variables are updated using 'downstream' data) and that allows good communication with experts through their graphical representation.[36–39]

This study objective was the development and application of an updated graphical independence network to monitor health risks associated with exposure to *Legionella pneumophila* during irrigation events with reclaimed wastewater. The impact of two types of water treatments were analyzed (described in the following section). Keeping the steady monitoring aspect in mind, the constructed Bayesian network is reevaluated with each arrival of new data. At each step we computed posterior distributions for all the variables in the model using priors defined as the posterior distributions at the previous step. The final purposes were to assess different risk scenarios according to the population activities, accounting their distance from the irrigated areas and the environmental conditions in order to get infection probabilities for each configuration.

In the following, material and methods section describes the study site, the data and the models implemented for the developed risk assessment monitoring tool. In the results section, we show how the risk quantification can be updated whenever desired and discussed with the experts. The discussion section discusses the benefits of the developed approach for the continuous monitoring of health risks in the context of QMRA, emphasizing over the necessity of informed use of such tools when few data are available.

2 Material and methods

2.1 Study site and data

A pilot experiment was conducted on an area near Tarbes (43°15'45.78"N, 0° 5'16.95"E) where irrigated corn fields from

* https : / / www . santepubliquefrance . fr / maladies-et-traumatismes / maladies-et-infections-respiratoires/legionellose (accessed on April 6th 2022)

two farms having different agricultural managements are monitored for various quality criteria since 2018 in the framework of the SmartFertireuse project (FUI SmartFertiReuse project†). Irrigation generally occurs from end of June up to the end of August, with 3 or 4 water supplies of 30-40 mm/ha. Two different water treatment pilots (ultra-filtration (A) and UV (B)) were installed at the wastewater treatment plant outlet in order to reach the required log reduction after standard treatment of the initial WWTP (stopping only at the secondary level). The different water compartments were sampled at each irrigation dates in order to monitor the classical indicators for water quality[40] and to quantify *Legionella* and additional pathogens such as enteric viruses. For each sample we collected: 100 ml at the entrance of WWTP (6 samples with 2 replicates each), 250 ml after standard WWTP treatments (6 samples with 2 replicates each), 250 ml after A treatment (4 samples with 2 replicates each), 250 ml after B treatment (6 samples with 2 replicates each), and 400 ml at the groundwater (4 samples at each of 2 different sites on the field and with 2 replicates each) for PCR analysis. Additionally, 2 biosamplers (AGI4-Ace Glass incorporated USA, connected to a 12 L/min pump) containing 40 ml of ultrapure Milli-Q water, were used to sample bioaerosols in the air above corn during irrigation (see Figure 1).

All samples were analyzed to quantify *Legionella* from molecular biology (qPCR from the standard NFT90-471‡). All *Legionella* species are considered potentially pathogenic for humans, but *Legionella pneumophila* is the etiological agent responsible for most reported cases of community-acquired and nosocomial legionellosis.[41,42] Additionnaly, some cultivations were made for samples showing high values. Figure 2 shows the data collected for *Legionella spp.* presence using qPCR method and differentiated by sampling zone. We can see a large variability both between the different analyzed compartments and for the three studied years. As expected the raw water at the entrance of the WWTP presented the highest values of Legionella and the groundwater the lowest values from 2019 to 2021. The impact of the two tertiary treatments (A and B) tend to decrease the quantities compared to values found in raw water. A large variability was also observed according to the dates in 2021 which is difficult to explain. Many factors among them climatic and water storage conditions, can lead variability. These points will be discussed in the results.

In parallel to theses analyses, the meteorological conditions are monitored thanks to a climatic station set up on the area, which records continuously wind speed and direction, global radiation, air temperature and moisture. To complete this dataset, daily data acquired from the Meteo-France weather station (Tarbes (43°11'12"N, 0°00'00"E) have been used between May and September over the period from 04/29/2010 to 06/18/2020 in order to better take into account the temporal variability of environmental conditions. The crop height was also measured during the irrigation events to assess the aerodynamic roughness (z_0) occurring on the atmospheric transport. Values varied from 80 cm to 3.2 m for the studied period.

† https://www.smartfertireuse.fr/les-moyens (accessed on April 6th 2022)
‡ https://tinyurl.com/5n6v4y38

2.2 The QMRA model

The first objective of our modelling is to construct a QMRA model that describes the transmission of *Legionella* along the water exposure pathway described in Figure 3.

2.2.1 Water quality model

We modeled 3 different pathways for the water used for the irrigation of the fields:

1. Irrigation directly from the groundwater table with quality C_{GT} [genome copies per liter (GC/L)], currently in use;

2. Irrigation from wastewater sequentially treated by standard WWTP treatments and experimental pilot A with quality C_A [GC/L];

2b. Irrigation from wastewater sequentially treated by standard WWTP treatments and experimental pilot B with quality C_B [GC/L].

In the process currently used (part 1 above), the groundwater is not treated in WWTP before being used for irrigation. In the experimental processes (parts 2 and 2b above) reclaimed wastewater is used. The initial wastewater coming to the WWTP is sequentially treated by standard WWTP treatments and experimental pilot (A or B). These treatments effects on the water quality are modelled by the following logarithmic decays formulas:

$$C_A = C_{TP}/10^{k_A}, \tag{1}$$

$$C_B = C_{TP}/10^{k_B}, \tag{2}$$

$$\text{with } C_{TP} = C_{WW}/10^{k_{TP}}, \tag{3}$$

where C_{TP} denotes the concentration of *Legionella* in wastewater treated by standard WWTP treatments [GC/L], C_{WW} denotes the concentration of *Legionella* in the raw water at the entrance of the WWTP (wastewater) [GC/L], k_{TP} denotes the decay rate of the filtration process of the WWTP [unitless], and k_A (resp. k_B) denotes the decay rate of the filtration process of experimental pilot A (resp. B) [unitless].

Concentration in *Legionella pneumophila* $C_{Lp,P}$ [GC/L] in each pathway $P \in \{GT, A, B\}$ was deduced from the concentrations C_P above by the following formula[13]:

$$C_{Lp,P} = C_P f_{Lp}, \tag{4}$$

where f_{Lp} denotes the fraction of *Legionella pneumophila* among the measured *Legionella spp.*

2.2.2 Spray irrigation

Two different sprinklers are used on the fields. They are characterized by their respective flow rate F_1 and F_2 [m^3/h] from which a portion p_{150} of particles is smaller than $150\mu m$ (portion of aerosols in respirable range, $< 150\mu m$). Experimental pilot A is associated with sprinkler 1 and experimental pilot B is

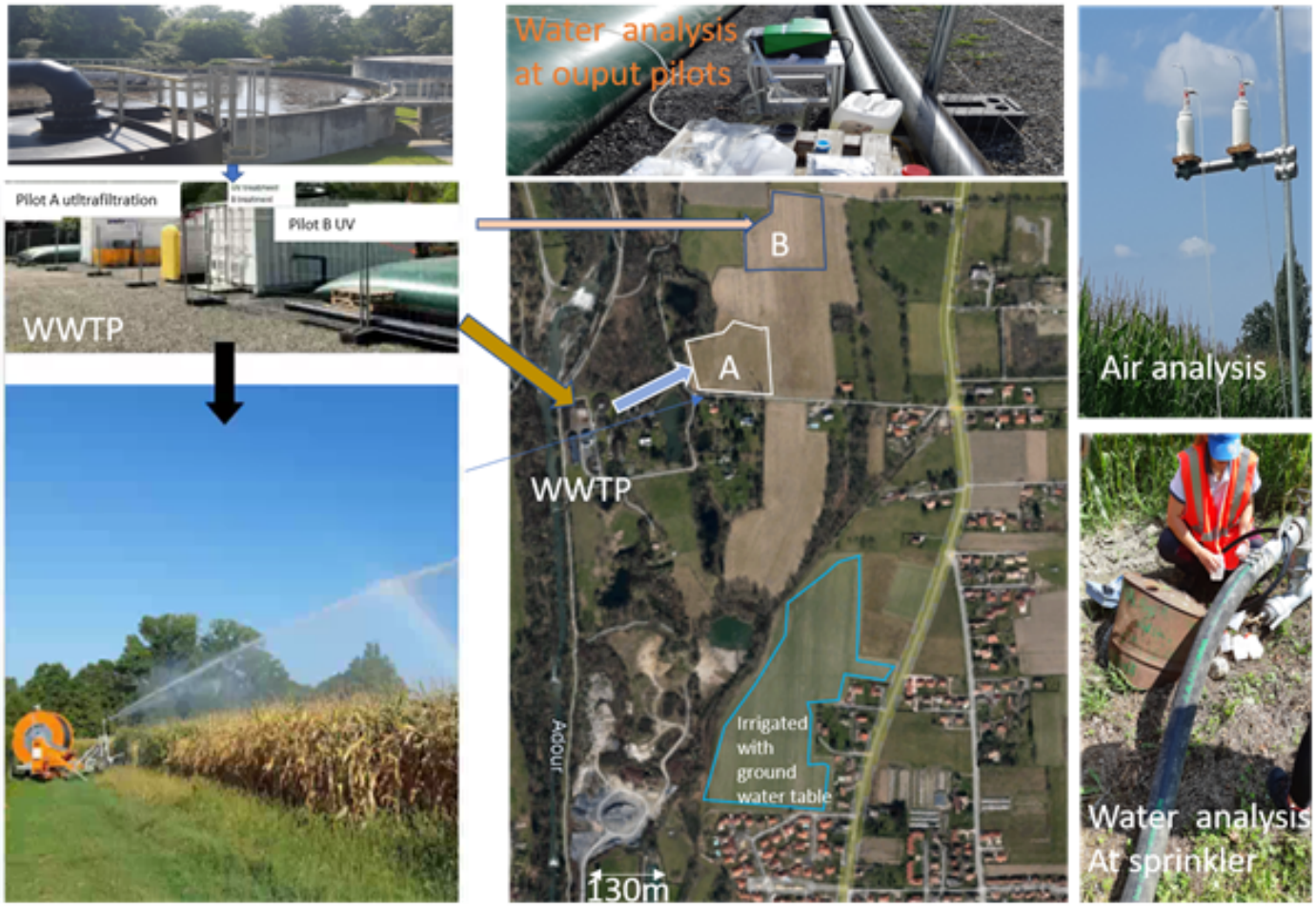


Fig. 1 Wastewater pathway towards irrigation in the vicinity of Tarbes, France. Fields A and B are monitored for wastewater usage.

associated with sprinkler 2. The following formula models the quality $Q_{Lp,P}$ $[GC/m^3]$ of the water used for the irrigation after aerosolization:

$$Q_{Lp,P} = F_{S_P} p_{150} C_{Lp,P} 3600/10^3, \tag{5}$$

where $C_{Lp,P}$ is the concentration of *Legionella pneumophila* in the water used for irrigation defined in equation (4) above $[GC/L]$, $Q_{Lp,P}$ denotes the concentration of *Legionella pneumophila* in the aerosolized water used for irrigation $[GC/m^3]$, S_P denotes the sprinkler associated to the pathway $P \in \{GT, A, B\}$, with $F_{S_{GT}} = (F_1 + F_2)/2$, $F_{S_A} = F_1$ and $F_{S_B} = F_2$ and the 10^3 and 3600 multipliers ensure the unit for $Q_{Lp,P}$ (in GC/m^3).

2.2.3 Atmospheric dispersion model for spray irrigation

We used a modified Gaussian plume atmospheric transport model to describe the concentration of *Legionella pneumophila* $C_{x,P}$ $[GC/m^3]$ at $x \in \{100; 300; 500; 1000\}$ meters downwind from the irrigation sprinkler for each pathway P:

$$C_{x,P} = \frac{Q_{Lp,P}}{2\pi u \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \left\{ \exp\left(-\frac{(z-z_0)^2}{2\sigma_z^2}\right) + \exp\left(-\frac{(z+z_0)^2}{2\sigma_z^2}\right) \right\} \exp\left(-\frac{\lambda x}{u}\right), \tag{6}$$

where $Q_{Lp,P}$ is defined in equation (5) $[GC/m^3]$, y is the horizontal distance perpendicular to wind $[m]$, z is the downwind receptor breathing zone height $[m]$; z_0 is the estimated aerodynamic rugosity $[m]$, u is the wind speed $[m/s]$; λ is the microbial decay coefficient $[s^{-1}]$, and σ_y, σ_z denote respectively the horizontal and vertical dispersion coefficients $[m]$ and are calculated by the following equations: $\sigma_y = R_y x^{r_y}$ and $\sigma_z = R_z x^{r_z}$, where R_y, r_y, R_z and r_z are constants depending on the wind speed u $[m/s]$ and the insolation value E $[J/cm^2]$ (see Seinfeld, *et al.* (2016)[43] and Tables 1 and 2 in Supplementary Materials).

2.2.4 Exposure model

Three exposure models were constructed depending on the activity of the population at risk:

- passersby with a supposed average $t_{passerby}$ of 1 minute of

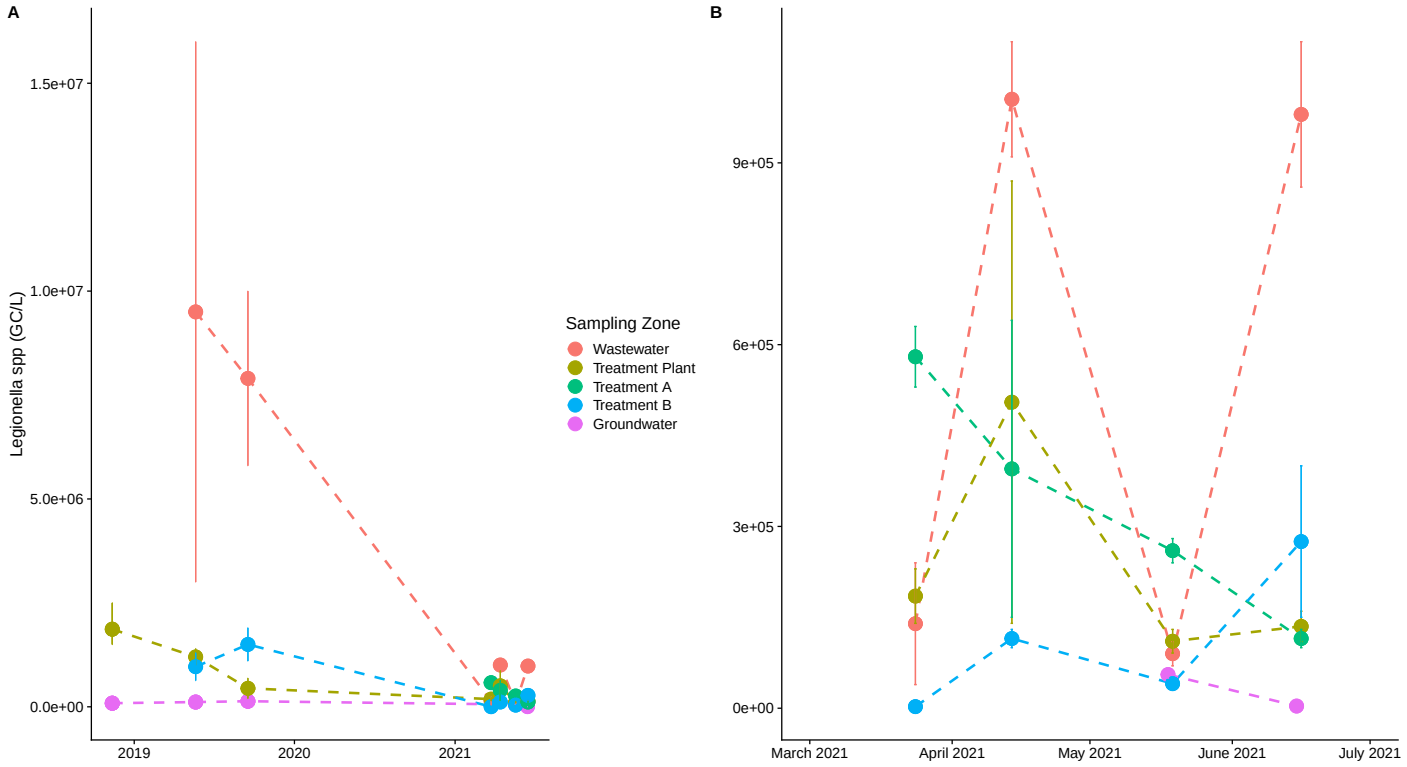


Fig. 2 Data collected on site differentiated by sampling zone. Points represent mean concentration of *Legionella spp.* for replicates and vertical bars represent their variability. A Complete data from November 2018 to June 2021. B Zoom on year 2021.

exposure per irrigation day[44],

- residents with a supposed average $t_{resident}$ of 2 seconds of exposure over the 2.27 hours spent outside per irrigation day[44,45], and
- farmers with a supposed average t_{farmer} of 30 minutes of exposure per irrigation day (chosen according to the local farmer behavior knowledge).

For each population, we computed the fraction of day this population was exposed and considered an uncertainty of a half minute around it. For example $t_{passerby}$ was taken as a Gaussian distribution centered around $1/(60\times 24) = 6.94\text{e-}4$ with standard deviation $1/(\sqrt{2}\times 60\times 24) = 4.91\text{e-}4$ truncated on $[0;1]$.

The inhaled dose $D_{A,x,P}$ [genome copies per day] was calculated for each pathway P, distance x and activity A taking values in $\{passerby; resident; farmer\}$ as follows:

$$D_{A,x,P} = C_{x,P} I t_A, \tag{7}$$

where $C_{x,P}$ is defined in equation (6) $[GC/m^3]$, I denotes the mean inhalation rate [m^3 air/day], and t_A the fraction of day each population at risk A is exposed [unitless].

2.2.5 Risk characterization

In our dose-response model, we distinguished clinical severity infections *csi* (corresponding to an infection requiring a clinical visit) from other infections *inf* (corresponding to subclinical infection or potentially Pontiac Fever endpoint). The daily probability $P_{i,A,x,P}$ was calculated for each pathway P, distance x, activity $A \in \{passerby; resident; farmer\}$ and type of infection $i \in \{inf, csi\}$ using the exponential dose response model for *Legionella pneumophila* from Armstrong, *et al.* (2007)[46] defined as follows:

$$P_{i,A,x,P} = 1 - e^{-r_i D_{A,x,P} f_{CFU}}, \tag{8}$$

where $D_{A,x,P}$ is defined in equation (7) $[CFU/day]$, r_i is the probability of the bacteria bypassing the host defenses and initiating response [unitless], and f_{CFU} denotes the relation between colony-forming units (CFU) and genome copies (GC) for *Legionella pneumophila*. The annual risk, which is the probability to be infected at least once during one irrigation day in a year was calculated as per the following equation:

$$P_{year,i,A,x,P} = 1 - (1 - P_{i,A,x,P})^{n_e d_e}, \tag{9}$$

where $n_e d_e$ is the total annual number of days of irrigation with n_e the number of episodes of irrigation over a year and d_e the duration of one irrigation episode in days.

Figure 4 represents the directed acyclic graph (DAG) of the QMRA model constructed with the equations above. The input distributions (or prior distributions in the Bayesian framework) for each parameters are defined in Table 1.

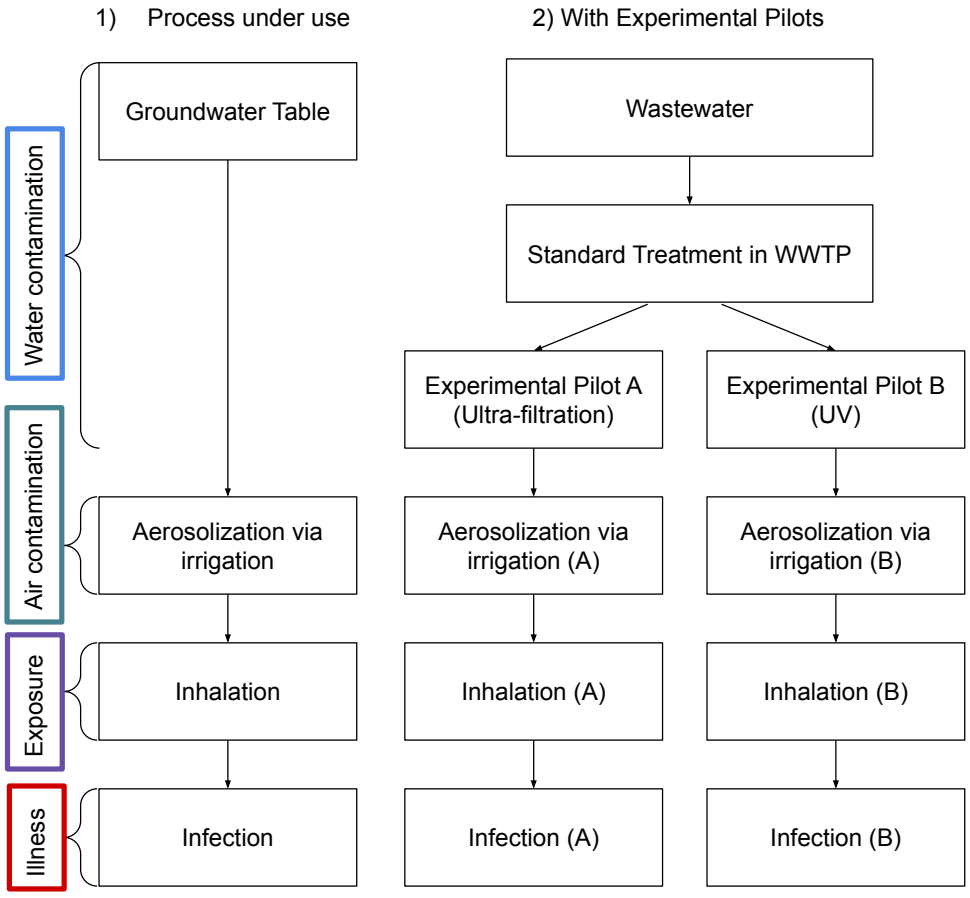


Fig. 3 Overview of the exposure pathway of *Legionella pneumophila* in the context of agricultural irrigation of the experimental plots in Tarbes.

2.3 Coupling the datasets

To account for the experimental data described in Section 2.1 we built an augmented model presented in Figure 5 (on the right, the QMRA model from Figure 4 is augmented by the data, in quantity n_m for the microbiological data and in quantity n_{day} and n_{year} for the climatic data, represented by rectangles using graphical conditional independence links).

This new model takes into account the measure uncertainty for the microbiological data and the inter and intra annual variabilities for the climatic data. In the case of the microbiological data, let $C_{Q,i}$ be the ith measured concentration of *Legionella* (in GC/L) at step $Q \in \{WW, GT, TP, A, B\}$. This random variable is linked to the concentration of *Legionella* C_Q defined in section 2.2.1. These random variables are linked through the following model:

$$\log_{10}(C_{Q,i}) \sim \mathcal{N}(\log_{10}(C_Q), \sigma_Q),$$

where $\mathcal{N}(,)$ is the normal distribution and σ_Q represents the measure uncertainty. We assigned a Gamma distribution to the square of its inverse *i.e.* $1/\sigma_Q^2 \sim \Gamma(100,1)$ which represents a vague prior confidence interval of $[0.09; 0.11]$ at 95% for σ_Q (in GC/L).

In the case of the climatic data, let $(u,E)_{j,k}^{obs}$ denote the vector of the observed wind speed (in m/s) and insolation value (in J/cm^2) at the jth day ($j = 1,\ldots,n_{day}$) of the kth year ($k = 1,\ldots,n_{year}$). This random vector is linked through the following model to the vector (u,E) of the QMRA model :

$$\log\left[(u,E)_{j,k}^{obs}\right] \sim \mathcal{N}_2(M_k, \Sigma_{intra}), \quad (10)$$

$$M_k \sim \mathcal{N}_2(M_m, \Sigma_{inter}), \quad (11)$$

$$M_{year} \sim \mathcal{N}_2(M_m, \Sigma_{inter}), \quad (12)$$

$$\text{and } \log[(u,E)] \sim \mathcal{N}_2(M_{year}, \Sigma_{intra}), \quad (13)$$

where $\mathcal{N}_2(,)$ is the bivariate normal distribution, Σ_{intra} and Σ_{inter} are the measure variability between days of each year and the measure variability between each year respectively, M_k and M_m are the mean of the kth year and the global mean over the years respectively. M_{year} is the predicted mean vector for a randomized year given by $M_{year} \sim \mathcal{N}_2(M_m, \Sigma_{inter})$. (u,E) is the vector of a predicted wind speed and insolation value in a randomized year. We assigned prior distributions to the variance and mean parameters

Description	Symbol X	Unit	Value	Reference
Water contamination	.....	.....	.....	.....
Water quality in groundwater table in *Legionella*	C_{GT}	GC/L	$\log_{10}(X) \sim \mathcal{N}(10,1)$	Expert knowledge and literature data
Water quality before treatment in *Leg.*	C_{WW}	GC/L	$\log_{10}(X) \sim \mathcal{N}(15,1)$	Expert knowledge and literature data
Log decay in WWTP by standard treatment	k_{TP}	$\log_{10}(GC/L)$	$X \sim \mathcal{N}(8,1)$	Project expert knowledge
Log decay by experimental pilot A	k_A	$\log_{10}(GC/L)$	$X \sim \mathcal{N}(4,1)$	Project expert knowledge
Log decay by experimental pilot B	k_B	$\log_{10}(GC/L)$	$X \sim \mathcal{N}(3,1)$	Project expert knowledge
Portion of *Legionella pneumo.*	f_{Lp}	Unitless	$X \sim \mathcal{N}(1e\text{-}3, 2e\text{-}4)\mathcal{T}(0,)$	From testing lab analytical techniques
Air contamination	.....	.....	.....	.....
Flow rate of sprinkler 1	F_1	m^3/h	$X \sim \mathcal{N}(44,1)$	Sprinklers properties [47]
Flow rate of sprinkler 2	F_2	m^3/h	$X \sim \mathcal{N}(42,1)$	Sprinklers properties [47]
Portion of aerosols in respirable range ($< 150\mu m$)	p_{150}	Unitless	$X \sim \mathcal{U}(5e\text{-}4, 7e\text{-}4)$	Sprinklers properties [47]
Horizontal distance perpendicular to wind	y	m	$X \sim \mathcal{U}(0, 2.5)$	From literature [13,48]
Downwind receptor breathing zone height	z	m	$X \sim \mathcal{U}(1, 1.7)$	Height of breathing zone
Estimated aerodynamic rugosity on the field	z_0	m	$X \sim \mathcal{U}(0.1, 0.45)$	Prior knowledge on the field
Wind speed and Insolation rate	(u,E)	m/s and J/cm^2	$\log(X) \sim \mathcal{N}_2\left(\binom{2}{8}, \frac{1}{8}I_2\right)$	Vague climatic prior
Microbial decay coefficient	λ	s^{-1}	$X \sim \mathcal{N}(1.32e\text{-}4, 3.44e\text{-}4)$	[49,50]
Exposure	.....	.....	.....	.....
Inhalation rate	I	m^3/day	$X \sim \mathcal{N}(20, 2)$	From literature [51,52]
Fract. of day a passerby is exposed	t_P	Unitless	$X \sim \mathcal{N}(6.94e\text{-}4, 4.91e\text{-}4)\mathcal{T}(0,1)$	From literature [44]
Fraction of day a resident is exposed	t_R	Unitless	$X \sim \mathcal{N}(3.15e\text{-}3, 4.91e\text{-}4)\mathcal{T}(0,1)$	From literature [44,45]
Fraction of day a farmer is exposed	t_F	Unitless	$X \sim \mathcal{N}(2.08e\text{-}2, 4.91e\text{-}4)\mathcal{T}(0,1)$	Knowledge on local farmer behavior
Illness	.....	.....	.....	.....
Portion of CFU	f_{CFU}	Unitless	$X \sim \mathcal{N}(1e\text{-}3, 2e\text{-}4)\mathcal{T}(0,)$	From testing lab analytical techniques from uncensored data
Dose response parameter for *Legionella pneumophila* infection endpoint	r_{inf}	Unitless	$\log(X) \sim \mathcal{N}(-2.934, 0.488)$	From literature [13,46,53]
Dose response parameter for *Legionella pneumophila* clinical severity infection endpoint	r_{csi}	Unitless	$\log(X) \sim \mathcal{N}(-9.688, 0.296)$	From literature [13,46,54]
Number of irrigation episodes per year	n_e	Unitless	$X \sim Multi\{3,4,5\}$	Process knowledge
Irrigation episodes duration	d_e	Days	$X \sim Multi\{2,3,4,5,6,7\}$	Process knowledge

Table 1 Model input parameters and distributions for the QMRA model. $\mathcal{N}(\mu,\sigma)$ denotes the normal distribution of mean μ and standard deviation σ, eventually truncated on the interval $T(,)$, $\mathcal{N}_2(M,\Sigma)$ denotes the bivariate Gaussian distribution of mean M and variance Σ. I_2 denotes the identity matrix of order 2. $\mathcal{U}(a,b)$ denotes the uniform distribution on the interval (a,b). $X \sim Multi\{a_1,a_2,\ldots\}$ stands for **equiprobable** sampling from the values $a_1,a_2,\ldots$ (multinomial distribution)

described above:

$$\Sigma_{intra}^{-1} \sim W_2(I_2;24), \tag{14}$$

$$\Sigma_{inter}^{-1} \sim W_2(I_2;24), \tag{15}$$

$$\text{and } M_m \sim \mathcal{N}_2\left(\binom{2}{8}, \frac{1}{24}I_2\right), \tag{16}$$

where I_2 is the identity matrix of order 2, $W_2(,)$ is the bivariate Wishart distribution, and $N_2(,)$ denotes the bivariate Gaussian distribution chosen to recover the vague climatic priors defined in Table 1 for (u,E) vector.

We used the R [55] package `rjags` [56] to obtain the Bayesian inference of the augmented model. MCMC sampling algorithms were applied as implemented in the software JAGS 4.3.0. [57]

3 Results

We ran the QMRA model (*Prior*), and the augmented model at each month for which data was collected (May and September 2019 and each month from March to June 2021) (*Posteriors 1 to 6*). For each model, we ran two independent MCMC chains (using different initial values for the parameters) of $1,400,000$ simulations with a burn-in period of $1,000,000$. Convergence of the MCMC run was assessed by graphical inspection of the chains.

In the following, we describe the results and discuss the update of prior beliefs in our model module by module.

3.1 Water contamination

The priors described in Table 1 yielded the following 95% credible intervals for water quality concentrations expressed here in log: $k_A \in [2;6]$, $k_B \in [1;5]$, $k_{TP} \in [6;10]$, $f_{Lp} \in [0.06, 0.14]$, $\log_{10}(C_A) \in [-0.4;6.5]$, $\log_{10}(C_B) \in [0.6;7.3]$, $\log_{10}(C_{GT}) \in [8;12]$, $\log_{10}(C_{WW}) \in [13;17]$, $\log_{10}(C_{TP}) \in [4;10]$, $\log_{10}(C_{Lp,A}) \in [-5;1]$, $\log_{10}(C_{Lp,B}) \in [-4;2.5]$, and $\log_{10}(C_{Lp,GT}) \in [3;7]$. These credible intervals are derived from the distributions described in Table 1 and the propagation of the uncertainty according to Figure 5 for non terminal variables (or parameters).

Figure 6 as well as Table 2 illustrate that the groundwater ($C_{Lp,GT}$) and the water after treatment by the pilot A or B ($C_{Lp,A}$, $C_{Lp,B}$) show similar contamination levels ($\sim 10^2$ GC/L) indicating that the new irrigation process is as safe as the previous one (with groundwater without treatment). As expected, we also notice a significant log reduction between contamination of raw and treated wastewater for the posterior distributions (~ 3 $\log_{10}$ reduction). Nevertheless, one observes that the parameters k_{TP}, k_A and k_B are smaller in the posterior distributions than in the prior distribution indicating that the data introduced at the C_{TP}'s, C_A's and C_B's level (see Figure 5) are more contaminated than expected. This may be due to the storage of the water between treatments in large reservoirs or poor control of the treatment chain as discussed below.

From the prior distributions (in red in Figure 6) to the posterior ones, one observes reduced uncertainties (narrower distributions in posterior). The posterior distribution at June 2021 (in pink in

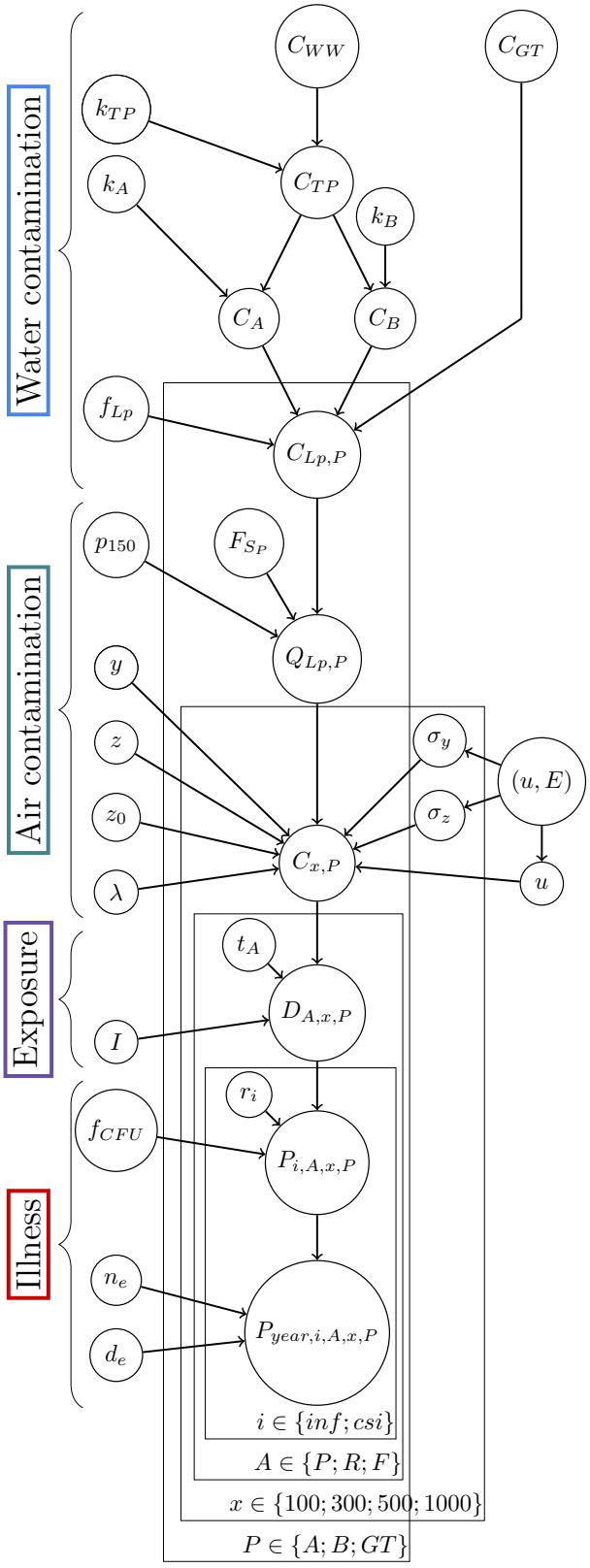


Fig. 4 Directed acyclic graph of the Bayesian network model.

Figure 6) shows reduction of the concentration after treatments: 1 $\log_{10}$ reduction (from 10^6 to 10^5 GC/L) for the wastewater after treatment in treatment plant and treatment A and similarly for treatment B. The transition to *Legionella pneumophila* from *Legionella* measurements analysed by PCR reduces again the contamination by 3 $\log_{10}$.

Although we observe a significant log reduction between raw and treated wastewater, the differences between treatments A and B and groundwater are very slim compared to what would have been expected by the prior knowledge of processes A and B which should eliminate most of the bacteria. This may be due to storage issues or lack of control in the chain of experimentation and will be discussed in the following.

Distribution	$C_{Lp,A}$	$C_{Lp,B}$	$C_{Lp,GT}$
Prior	-0.03 [-3.46;3.34]	1.03 [-2.32;4.39]	6.99 [5.01;8.9]
2018/11/01	-0.68 [-2.92;1.57]	0.34 [-1.93;2.6]	2.29 [1.53;3.08]
2019/05/01	-1.03 [-3.14;1.05]	1.94 [0.76;3.1]	2.11 [1.46;2.73]
2019/09/01	-1.23 [-3.26;0.9]	2.4 [1.47;3.33]	2.06 [1.52;2.61]
2021/03/01	1.4 [0.23;2.57]	1.82 [1.03;2.6]	2.06 [1.51;2.59]
2021/04/01	1.84 [0.92;2.74]	1.87 [1.17;2.56]	2.06 [1.51;2.59]
2021/05/01	1.98 [1.2;2.76]	1.82 [1.18;2.42]	2.02 [1.5;2.52]
2021/06/01	2 [1.31;2.66]	1.89 [1.3;2.46]	1.86 [1.37;2.35]

Table 2 Medians and 95% confidence intervals for concentrations in *Legionella pneumophila* ($\log_{10}$ GC/L)

3.2 Air contamination

The air contamination module output is the concentration $C_{x,P}$ (in GC/m^3) of *Legionella pneumophila* in the air at distances $x \in \{100\,m, 300\,m, 500\,m, 1000\,m\}$ from the source due to the irrigation using water of quality $P \in \{A, B, GT\}$ (see Figure 7). One observes reduced uncertainty from the prior to the posterior distributions of the concentration of *Legionella pneumophila* in the air for all treatments and distances. As expected, the distributions are globally decreasing with the distance: from 10^{-3} at $100\,m$ to 10^{-5} at $1000\,m$ for wastewater after treatment in TP and treatment A and similarly for treatment B and groundwater GT. Final predicted air concentrations are of similar intensity for irrigation using water from treatments A and B and groundwater which comforts the previous remarks on the quality of the proposed irrigation process.

3.3 Illness

Figures 8 to 10 represent the probabilities of clinical severity infection over one year, for farmers (see Figure 8), for residents

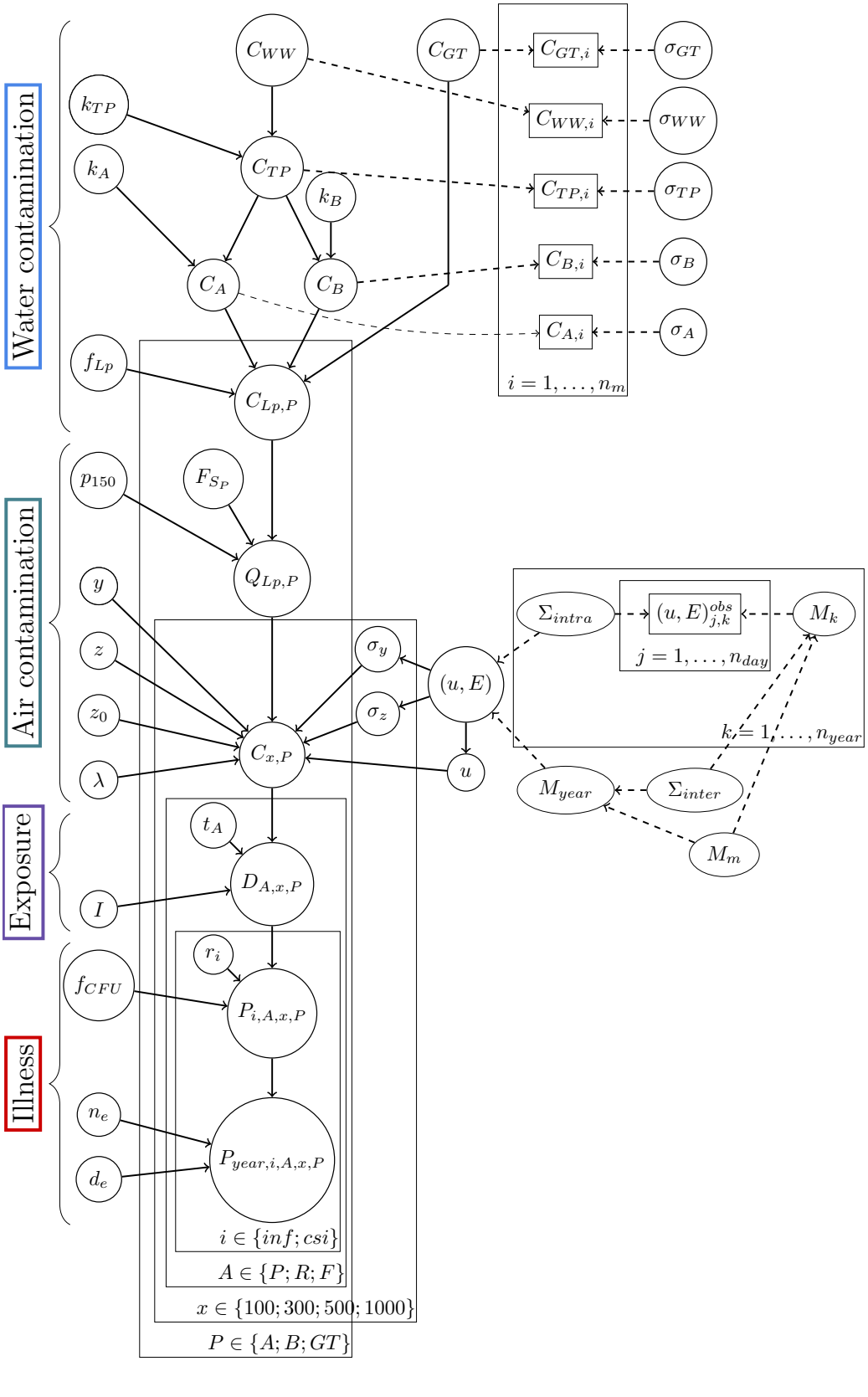


Fig. 5 Directed acyclic graph of the augmented model.

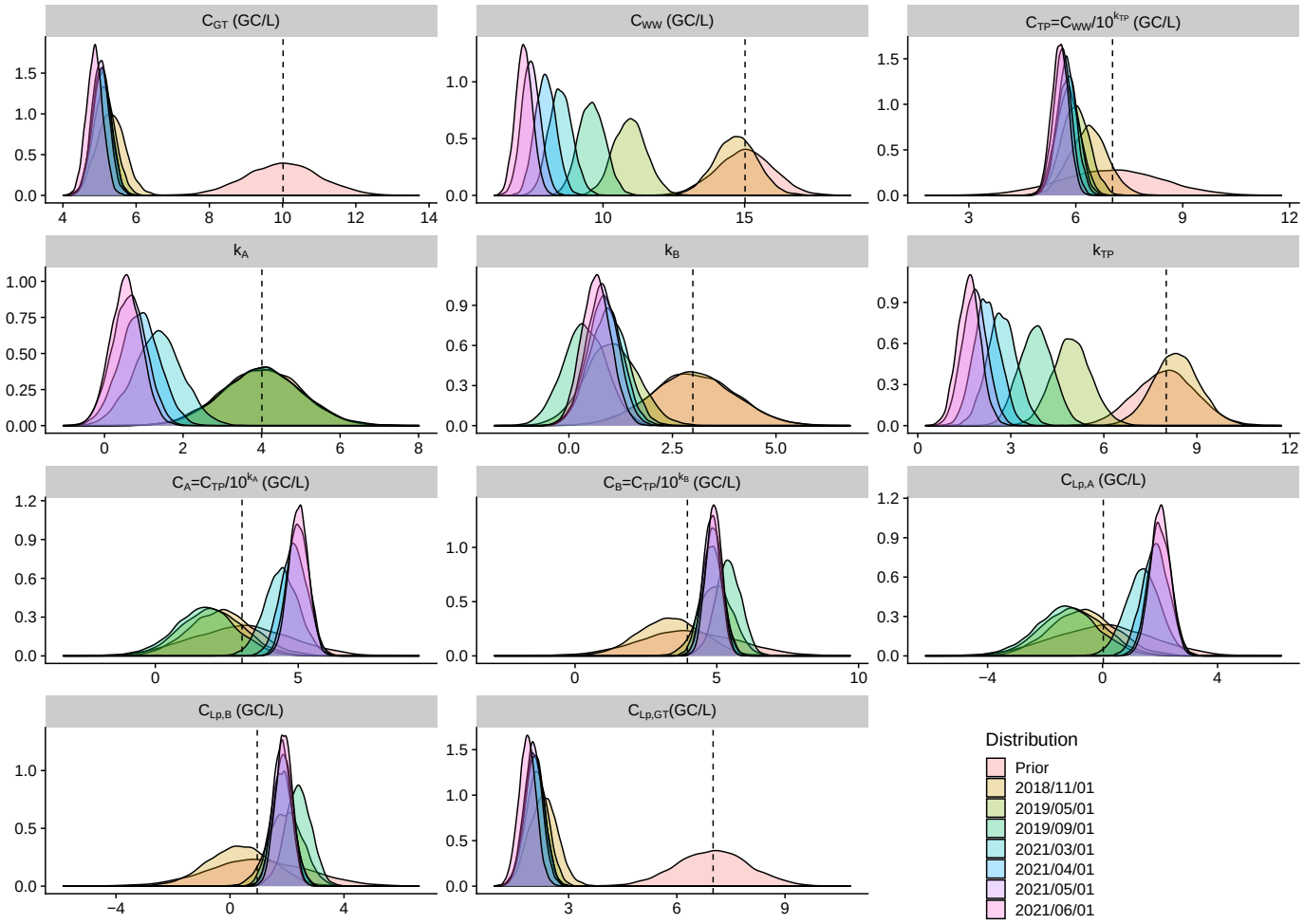


Fig. 6 Prior (red) versus posterior marginal distributions of selected parameters and variables of the water contamination module of the augmented Bayesian network. Dashed lines represent the median of the prior marginal distributions. Distributions represented are in $\log_{10}(CFU/L)$

(see Figure 9) and for passersby (see Figure 10). For both clinical severity infections (csi) and others (inf) we observe a decrease in the risk of contamination with the distance for each category of people at risk: from 10^{-10} at $100m$ to 10^{-13} at $1000m$ for the risk of clinical severity infection for the farmers exposed to treatment A (and after treatment TP). We observe a decay of approximately 1 $\log_{10}$ between farmers and residents for the probability of clinical severity infection (which is again the case between residents and passersby). Treatments A and B and groundwater give similar probabilities of infection with a slightly higher risk for groundwater which confirms the previous results that the new irrigation process is slightly safer than the one under use (with groundwater, without further treatment). It is also notable that the upper bound of the 95% credibility intervals never exceeds the US EPA annual infection benchmark of 10^{-4} infections per year (for drinking water) for clinical severity infections.

4 Discussion

A Bayesian network methodology was applied to a QMRA problematic to monitor the risk of *Legionella* infection in the vicinity of agricultural plots irrigated with two experimental water treatment pilots. Bayesian networks approach allows for simple accounting for variability and uncertainty in a context of complex modelling such as QMRA models. In the developed approach a modified Gaussian plume dispersion model was used to compute health risks according different scenarios, using knowledge of meteorological conditions over long periods (> 20 years here) and distinguishing three categories of persons at risk, two dose-response endpoints and different downwind distances from the sprinkler.

Both for annual infection and clinical severity infection, no major difference was observed between risk induced by irrigation using water treated by treatment A (ultra fine filtration, after standard treatment plant), treatment B (UV, after standard treatment plant) or groundwater. This comforts our prior belief that treatments A and B give water quality at least comparable to the quality of the groundwater. Risks were observed to decrease with distance from the source, which was expected according to the atmospheric dispersion modelling. Approximately 1 log difference was observed between each of the studied categories of people (farmers, residents and passersby), with passersby being the population least at risk and farmers the population most at risk

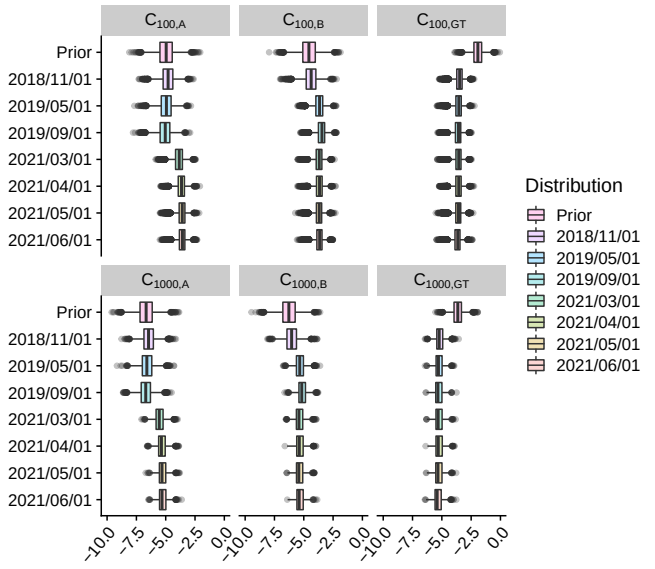


Fig. 7 Log10 of concentrations in *Legionella pneumophila* in the air due to the irrigation process for groundwater (GT), and after treatments A and B at $100m$ and $1000m$ from the source.

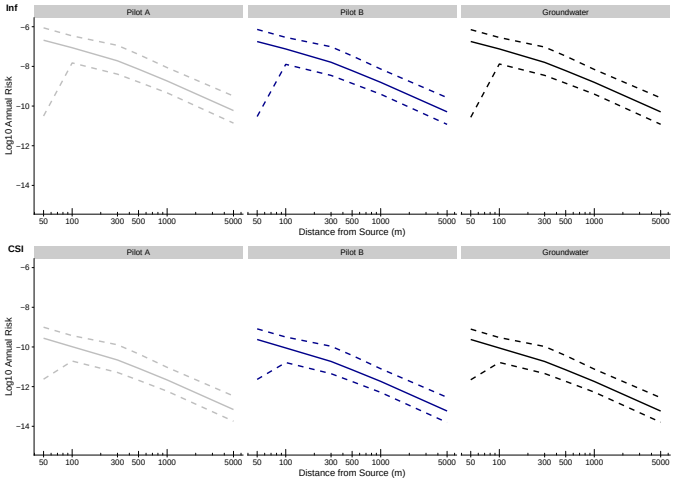


Fig. 8 Log10 annual infection risks for L. pneumophila for the **farmers** due to sprinkler exposure at downwind distances ranging from 50 to 5000 meters from the source. The median (solid line) and interquartiles confidence interval (dashed lines) are shown. Infection (Inf) or clinical severity infection (CSI) dose response model endpoints.

according to the model (logical order according to the exposure time).

High values of Legionella in GC/L were found in the different analyzed compartments even after the two studied treatments. Impact of the storage time in large reservoirs after the two pilots can in part explain some high observed levels. The cleaning of these tank covers appears as an important point to check for operational applications. A such observation reveals the interest of working in the first steps with GC for an operational monitoring of the water quality, GC are uncensored data and PCR analysis are less time consuming and can be done routinely. The values reported in GC were first introduced in our model until the risk characterization, then converted in CFU applying a significant decrease of 10e3 between GC and UFC, as quantified from our laboratory analytical results and as often observed in bibliography. Let us mention that if one applies this decrease factor (f_{CFU}) on the contamination after the tertiary treatment A and B, i.e. on C_A and C_B, the two treatments are class A reclaimed water because only 0.1% are above the new European limit (<1000 CFU/L).

It is notable that the upper bound of the interquartiles credibility intervals of the posterior never exceeded the US EPA annual infection benchmark of 10^{-4} infections per year for drinking water. Although, a slow decrease of risk is observed with the distance to the source which is the consequence of the meteorological conditions in the vicinity of the agricultural plots. This observation is consistent with previous studies that have predicted long-range transport of *Legionella*.[13,16,58] In our local climatic context where the decrease is slow it seems essential that the concentration have to be already low (below 10^4) at $100m$ from the sprinklers.

The developed tool is ready to be updated especially by air measured concentrations on the plots by the irrigations with reclaimed water that will take place starting in Summer 2021. This tool will then be used to monitor the potential risks in the vicinity of the experimental plots and thus meet the public health demands of population protection.

In a more general QMRA context, the model we developed gives an operational tool, and theoretically stable methodology to maintain a continuous monitoring of the risks induced by the irrigation practices on the agricultural plots under study. This methodology can adapt to new sets of data and easily update the model if given new information. This model can easily be adapted to other well known pathogen through the enlightened adaptation of a few priors such as the ones on the decay parameter λ, or the dose-response parameter r, ...

A challenge would lie in the usage of the model to answer simultaneous multipathogen risk analysis. Also, one of our next main interest lies in the addition of a modelisation of the impact of the irrigation with reclaimed wastewater on the quality of the groundwater to simultaneously quantify the inhalation induced risk and the groundwater table contamination by reclaimed wastewater transport in soil. Furthermore, it could be very interesting to link the final probability of infection over the year to epidemiological data over the region. But for the moment, the epidemiological data at our disposal are only available at a large geographical scale (surface of Occitanie region is approximately $70,000km^2$ compared to the surface of our study zone which is approximately $6km^2$) and it does not seem reasonable to link this data to our very local modeling.

5 Conclusions

- First developed Bayesian networks approach in QMRA for wastewater surveillance. The main asset is that risks are quantified with their uncertainties at each desired time taking into account new and past data.
- The QMRA model describes the pathway of the water contamination from the entrance of the WWTP to the infection

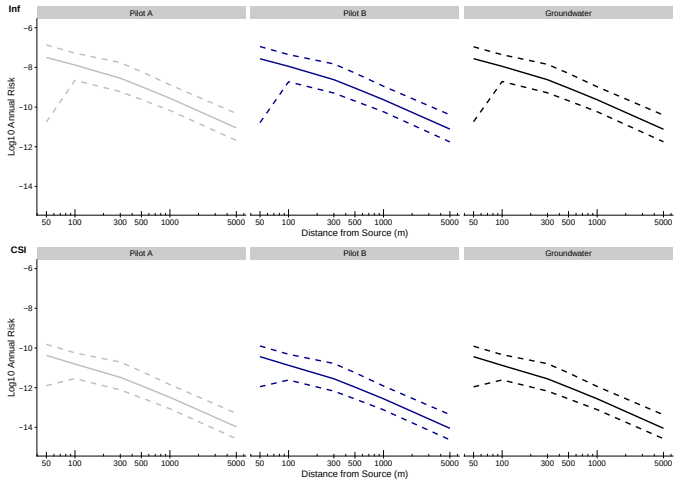


Fig. 9 Log10 annual infection risks for L. pneumophila for the **residents** due to sprinkler exposure at downwind distances ranging from 50 to 5000 meters from the source. The median (solid line) and interquartiles confidence interval (dashed lines) are shown. Infection (Inf) or clinical severity infection (CSI) dose response model endpoints.

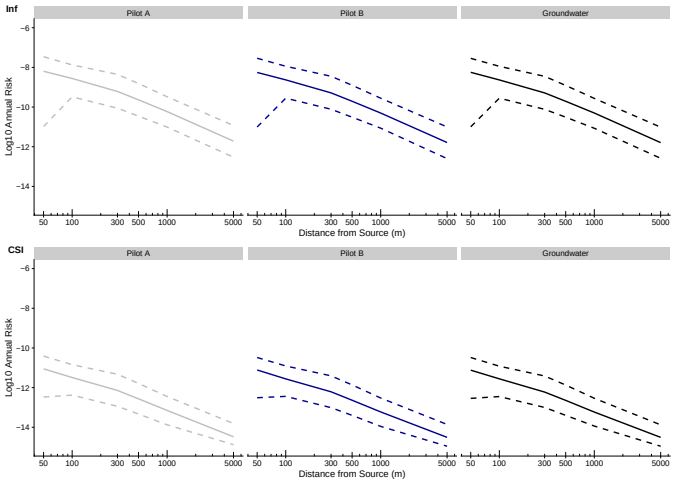


Fig. 10 Log10 annual infection risks for L. pneumophila for the **passersby** due to sprinkler exposure at downwind distances ranging from 50 to 5000 meters from the source. The median (solid line) and interquartiles confidence interval (dashed lines) are shown. Infection (Inf) or clinical severity infection (CSI) dose response model endpoints.

risk using pathogen decay models in the WWTP, a Gaussian plume model for the air contamination, an inhalation model for the population exposure and finally a dose-response model for risk characterization.

- The uncertainty of the probabilistic QMRA model is reduced by the introduction of observed data (pathogens concentrations and regional meteorological data) along the contamination pathway to obtain the distribution of the number of pathogens before and after treatment in WWTP, of the number of aerosolised pathogens, of the concentration of pathogens in the air at different distances, and finally the exposure and illness distributions for three different categories of population at risk and two illness endpoints.
- *Legionella* annual subclinical infection risk and annual clinical severity infection risk linked to the agricultural irrigation using the groundwater table or two experimental water treatment pilots are inferred below 10^{-4} tolerable limit defined by the US EPA for farmers, passersby and residents at distances between $100\,m$ and $1000\,m$ away from sprinklers.
- Such dynamic approach can be applied to various pathogens in the context of wastewater reuse

Conflicts of interest

There are no conflicts to declare.

Acknowledgment

The partners of the project are Polymem, Bio-UV, Ecofilae, VERI, Veolia Eau, Sede, and Inrae-Transfert Narbonne.

Established with the authorization of the Tarbes-Lourdes-Pyrénées Syndicate, owner of the treatment plant, this project involves farmers as well as the Departmental Federation of Farmers' Unions of Hautes-Pyrénées (FDSEA 65) and the Chamber of farming Hautes-Pyrénées.

The project is FUI funding, co-funded by the Sud-West region and Occitania region (France).

Notes and references

1 P. T. Vo, H. H. Ngo, W. Guo, J. L. Zhou, P. D. Nguyen, A. Listowski and X. C. Wang, *Science of The Total Environment*, 2014, **494-495**, 9–17.

2 H. Elbasiouny, H. El-Ramady and F. Elbehiry, 2021.

3 J. S. Guest, S. J. Skerlos, J. L. Barnard, M. B. Beck, G. T. Daigger, H. Hilger, S. J. Jackson, K. Karvazy, L. Kelly, L. Macpherson, J. R. Mihelcic, A. Pramanik, L. Raskin, M. C. M. Van Loosdrecht, D. Yeh and N. G. Love, *Environmental Science & Technology*, 2009, **43**, 6126–6130.

4 J. Fito and S. W. Van Hulle, *Environment, Development and Sustainability*, 2021, **23**, 2949–2972.

5 J. Hristov, J. Barreiro-Hurle, G. Salputra, M. Blanco and P. Witzke, *Agricultural Water Management*, 2021, **251**, 106872.

6 EP, *European Parliament*, 2020.

7 Food and Agriculture Organization of the United Nations, *The state of the world's land and water resources for food and agriculture: Managing systems at risk*, Earthscan, 2011.

8 A. A. Adegoke, I. D. Amoah, T. A. Stenström, M. E. Verbyla and J. R. Mihelcic, *Frontiers in Public Health*, 2018, **6**, 337.

9 D. Courault, I. Albert, S. Perelle, A. Fraisse, P. Renault, A. Salemkour and P. Amato, *Science of the Total Environment*, 2017, **592**, 512–526.

10 T. Paez-Rubio, E. Viau, S. Romero-Hernandez and J. Peccia, *Applied and Environmental Microbiology*, 2005, **71**, 804–810.

11 B. Teltsch and E. Katzenelson, *Applied and Environmental Microbiology*, 1978, **35**, 290–296.

12 M. Blanky, Y. Sharaby, S. Rodríguez-Martínez, M. Halpern and E. Friedler, *Water Research*, 2017, **125**, 410–417.

13 K. A. Hamilton, M. T. Hamilton, W. Johnson, P. Jjemba,

Z. Bukhari, M. LeChevallier and C. N. Haas, *Water Research*, 2018, **134**, 261–279.
14 K. A. Hamilton and C. N. Haas, *Environ. Sci.: Water Res. Technol.*, 2016, **2**, 599–613.
15 B. Diederen, *J Infect*, 2008, **56(1)**, 1–12.
16 S. M. Walser, D. G. Gerstner, B. Brenner, C. Höller, B. Liebl and C. E. Herr, *International journal of hygiene and environmental health*, 2014, **217**, 145–154.
17 P. Xu, C. Zhang, X. Mou and X. C. Wang, *Water Science and Technology*, 2020, **82**, 1547–1559.
18 A. D. Loenenbach, C. Beulens, S. M. Euser, J. P. van Leuken, B. Bom, W. van der Hoek, A. M. de Roda Husman, W. L. Ruijs, A. A. Bartels and A. Rietveld, *Emerging infectious diseases*, 2018, **24**, 1914.
19 J. W. Tang, *Journal of the Royal Society Interface*, 2009, **6**, S737–S746.
20 C. N. Haas, J. B. Rose and C. P. Gerba, *Quantitative microbial risk assessment*, John Wiley & Sons, 2014.
21 S. Petterson and N. Ashbolt, *Journal of Water and Health*, 2016, **14**, 571–589.
22 C. E. Owens, M. L. Angles, P. T. Cox, P. M. Byleveld, N. J. Osborne and M. B. Rahman, *Water Research*, 2020, **174**, 115614.
23 D. Mara, P. Sleigh, U. Blumenthal and R. Carr, *Journal of water and health*, 2007, **5**, 39–50.
24 R.-n. Wang, X. Li and C. Yan, *Environmental Science and Pollution Research*, 2021, 1–18.
25 A. Simhon, V. Pileggi, C. A. Flemming, G. Lai and M. Manoharan, *Water Research*, 2020, **183**, 116121.
26 A. Carducci, G. Donzelli, L. Cioni and M. Verani, *International journal of environmental research and public health*, 2016, **13**, 733.
27 C. Yan, Z.-c. Gui and J.-t. Wu, *Environmental Science and Pollution Research*, 2021, **28**, 8140–8150.
28 D. P. Janevska, R. Gospavic, E. Pacholewicz and V. Popov, *Food Research International*, 2010, **43**, 1915–1924.
29 S. Højsgaard, *Journal of Statistical Software*, 2012, **46**, 1–26.
30 I. Albert, E. Espié, H. de Valk and J.-B. Denis, *Risk Analysis: An International Journal*, 2011, **31**, 1141–1155.
31 P. A. Aguilera, A. Fernández, R. Fernández, R. Rumí and A. Salmerón, *Environmental Modelling & Software*, 2011, **26**, 1376–1388.
32 K. B. Korb and A. E. Nicholson, *Bayesian artificial intelligence*, CRC press, 2010.
33 D. Beaudequin, F. Harden, A. Roiko, H. Stratton, C. Lemckert and K. Mengersen, *Environment International*, 2015, **80**, 8–18.
34 M. E. Verbyla, E. M. Symonds, R. C. Kafle, M. R. Cairns, M. Iriarte, A. Mercado Guzman, O. Coronado, M. Breitbart, C. Ledo and J. R. Mihelcic, *Environmental science & technology*, 2016, **50**, 6803–6813.
35 T. D. Nielsen and F. V. Jensen, *Bayesian networks and decision graphs*, Springer Science & Business Media, 2009.
36 I. Albert, E. Grenier, J.-B. Denis and J. Rousseau, *Risk Analysis: An International Journal*, 2008, **28**, 557–571.
37 J. Smid, D. Verloo, G. Barker and A. Havelaar, *International Journal of Food Microbiology*, 2010, **139**, S57–S63.
38 C. Rigaux, S. Ancelet, F. Carlin, C. Nguyen-thé and I. Albert, *Risk Analysis*, 2013, **33**, 877–892.
39 D. Beaudequin, F. Harden, A. Roiko and K. Mengersen, *Science of The Total Environment*, 2016, **541**, 1393–1409.
40 EC, *European Commission*, 2021.
41 M. Palusińska-Szysz and M. Cendrowska-Pinkosz, *Archivum immunologiae et therapiae experimentalis*, 2009, **57**, 279–290.
42 N. P. Cianciotto, *International Journal of Medical Microbiology*, 2001, **291**, 331–343.
43 J. H. Seinfeld and S. N. Pandis, *Atmospheric chemistry and physics: from air pollution to climate change*, John Wiley & Sons, 2016.
44 ANSES, *Saisine n°2009-SA-0329*, 2012.
45 US EPA, *U.S. Environmental Protection Agency*, 2011.
46 T. W. Armstrong and C. N. Haas, *Risk Analysis*, 2007, **27**, 1581–1596.
47 B. Molle, L. Huet, S. Tomas, J. M. Granier, P. Dimaiolo and C. Rosa, *PhD thesis*, irstea, 2009.
48 T. Paez-Rubio, A. Ramarui, J. Sommer, H. Xin, J. Anderson and J. Peccia, *Environmental science & technology*, 2007, **41**, 3537–3544.
49 P. Hambleton, M. Broster, P. Dennis, R. Henstridge, R. Fitzgeorge and J. Conlan, *Epidemiology & Infection*, 1983, **90**, 451–460.
50 P. Dennis and J. Lee, *Journal of Applied Bacteriology*, 1988, **65**, 135–141.
51 P. Stellacci, L. Liberti, M. Notarnicola and C. N. Haas, *Desalination*, 2010, **253**, 51–56.
52 J. P. Brooks, M. R. McLaughlin, C. P. Gerba and I. L. Pepper, *Journal of Environmental Quality*, 2012, **41**, 2009–2023.
53 D. Muller, M. L. Edwards and D. W. Smith, *Journal of Infectious Diseases*, 1983, **147**, 302–307.
54 R. Fitzgeorge, A. Baskerville, M. Broster, P. Hambleton and P. Dennis, *Epidemiology & Infection*, 1983, **90**, 81–89.
55 R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020.
56 M. Plummer, *rjags: Bayesian Graphical Models using MCMC*, 2019.
57 M. Plummer, Proceedings of the 3rd international workshop on distributed statistical computing, 2003, pp. 1–10.
58 K. Borgen, I. Aaberge, Ø. Werner-Johansen, K. Gjøsund, B. Størsrud, S. Haugsten, K. Nygård, T. Krogh, E. Høiby and D. Caugant, *Eurosurveillance*, 2008, **13**, 18985.

1 Supplementary Material

1.1 Pasquill stability classes

Stability class	R_y	r_y	R_z	r_z
A	0.469	0.903	0.017	1.380
B	0.306	0.885	0.072	1.021
C	0.230	0.855	0.076	0.879
D	0.219	0.764	0.140	0.727

Table S1: Coefficients for Pasquill Stability Classes

A : extremely unstable B : moderately unstable		C : unstable D : neutral	
Wind speed (u) at ground level (10m)	Insolation value (E)		
	strong $> 2000\ J/cm^2$	moderate $\in [1000; 2000]$	light $< 1000\ J/cm^2$
< 2 m/s	A	A-B	B
2-3 m/s	A-B	B	C
3-4 m/s	B	B-C	C
4-6 m/s	C	C-D	D
> 6 m/s	C	D	D

Table S2: Pasquill Stability Classes definition

1.2 Graphical representation of the distributions

In the following, we present the Prior (red) versus posterior marginal distributions of parameters and variables represented in Figure 5 of main article and not represented in the main content of the article. Dashed lines represent the median of the prior marginal distributions.

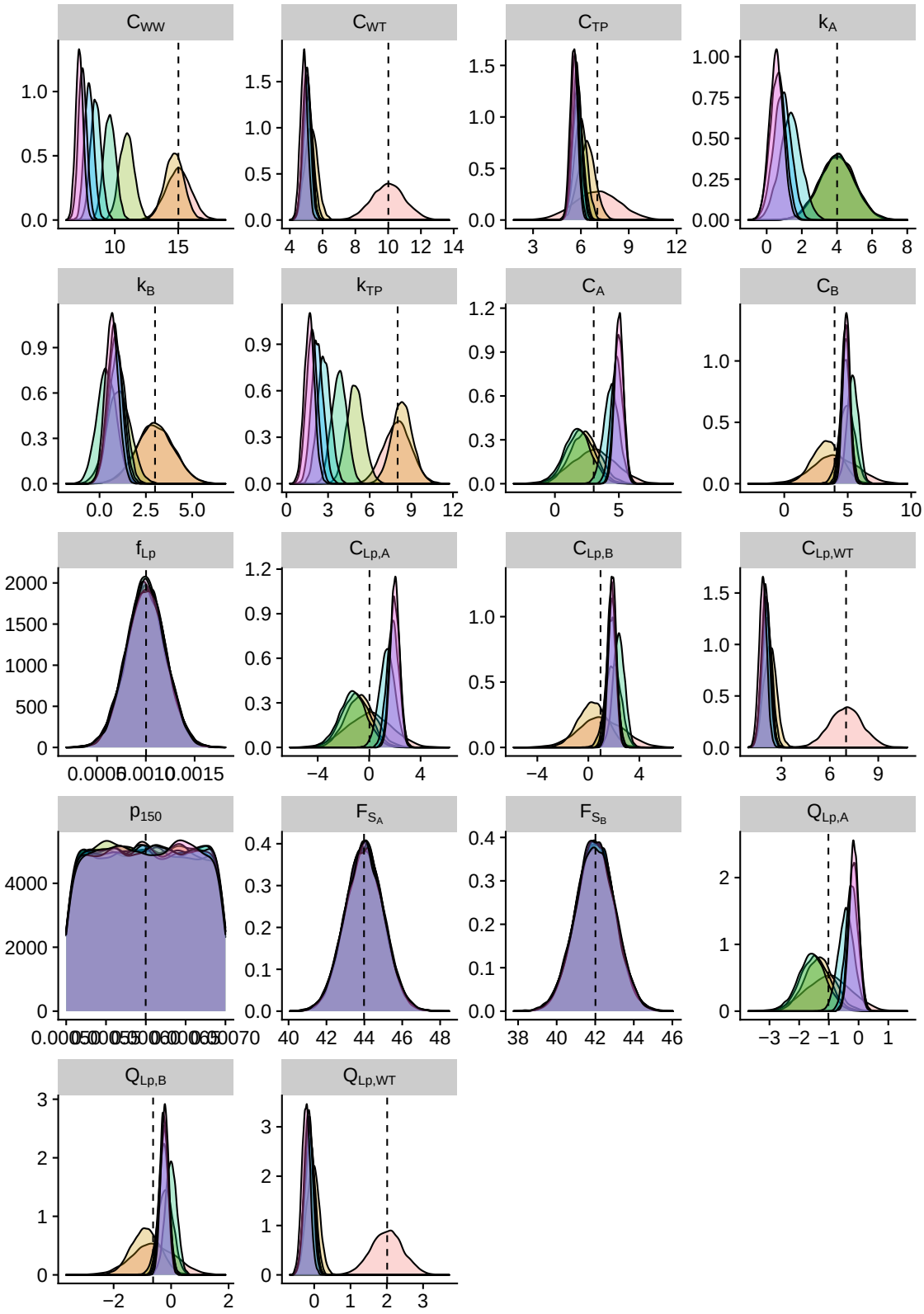


Figure S1: Prior (red) versus posterior marginal distributions of selected parameters and variables of the water contamination module of the augmented Bayesian network. Dashed lines represent the median of the prior marginal distributions. Distributions represented are in GC/CFU for R, $\log_{10}(CFU/L)$ for $C_{Lp,A}$ and $C_{Lp,GT}$, m^3/h for F_{S_A} and F_{S_B}, $\log_{10}(CFU/m^3)$ for $Q_{Lp,A}$, $Q_{Lp,B}$ and $Q_{Lp,GT}$ and unitless for f_{Lp} and p_{150}.

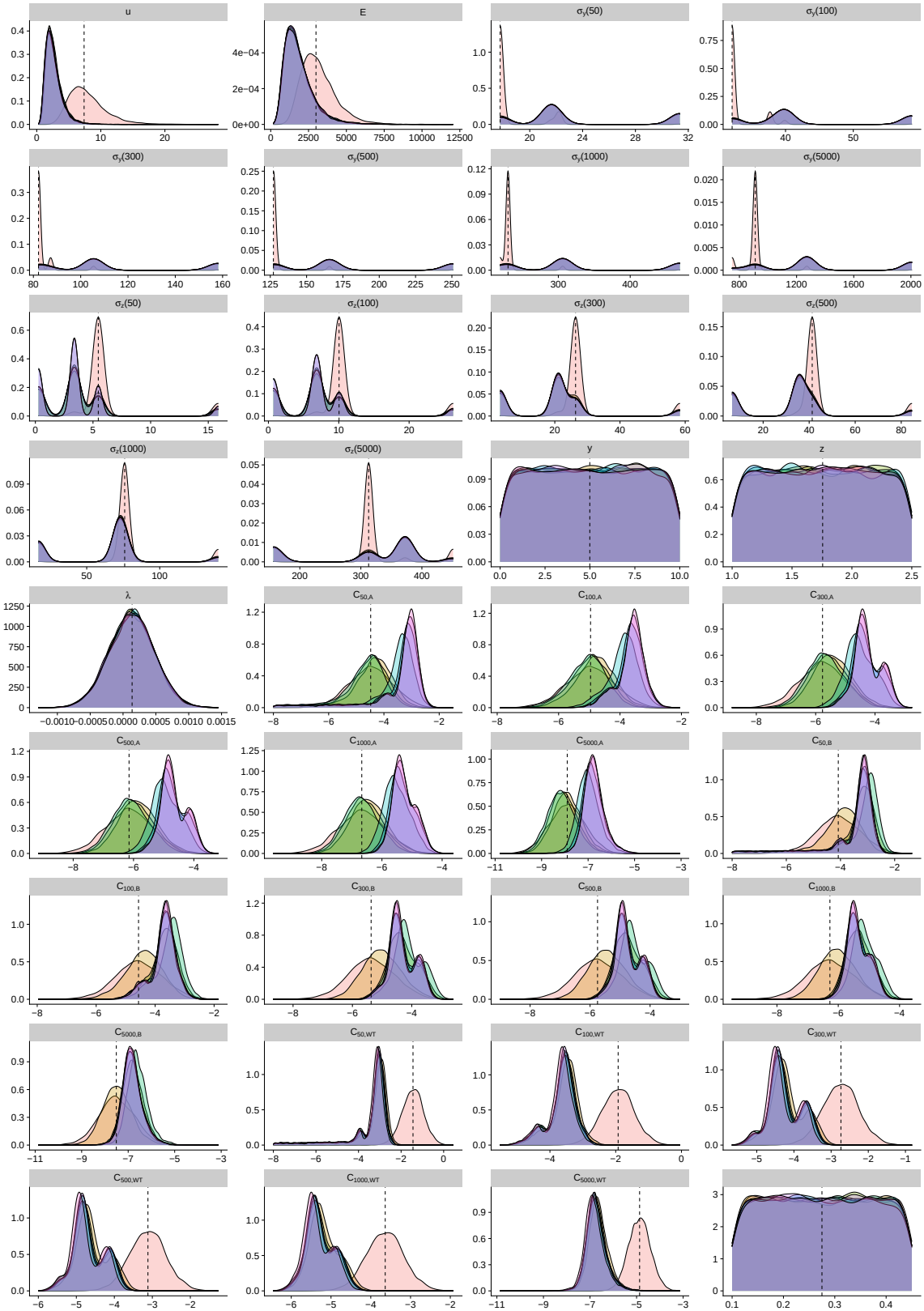


Figure S2: Prior (red) versus posterior marginal distributions of selected parameters and variables of the water contamination module of the augmented Bayesian network. Dashed lines represent the median of the prior marginal distributions. Distributions represented are in m/s for u, J/cm^2 for E, m for σ_y, σ_z, y, z and z_0 and s^{-1} for λ.

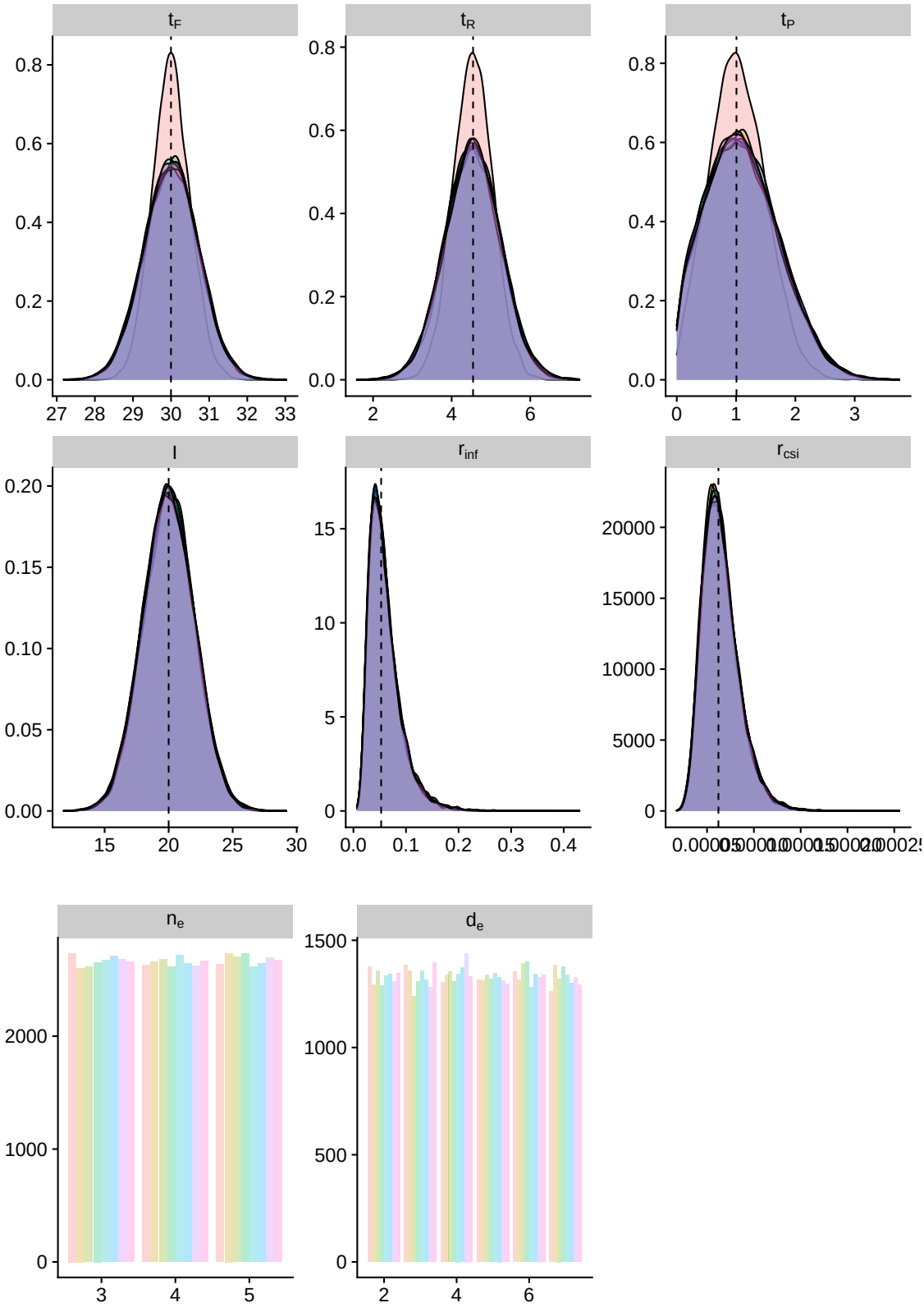


Figure S3: Prior (red) versus posterior marginal distributions of selected parameters and variables of the water contamination module of the augmented Bayesian network. Dashed lines represent the median of the prior marginal distributions. Distributions represented are unitless for $t_{passerby}$, $t_{resident}$, t_{farmer}, r_{csi}, r_{inf} and n_e, m^3/day for I and $days$ for d_e.

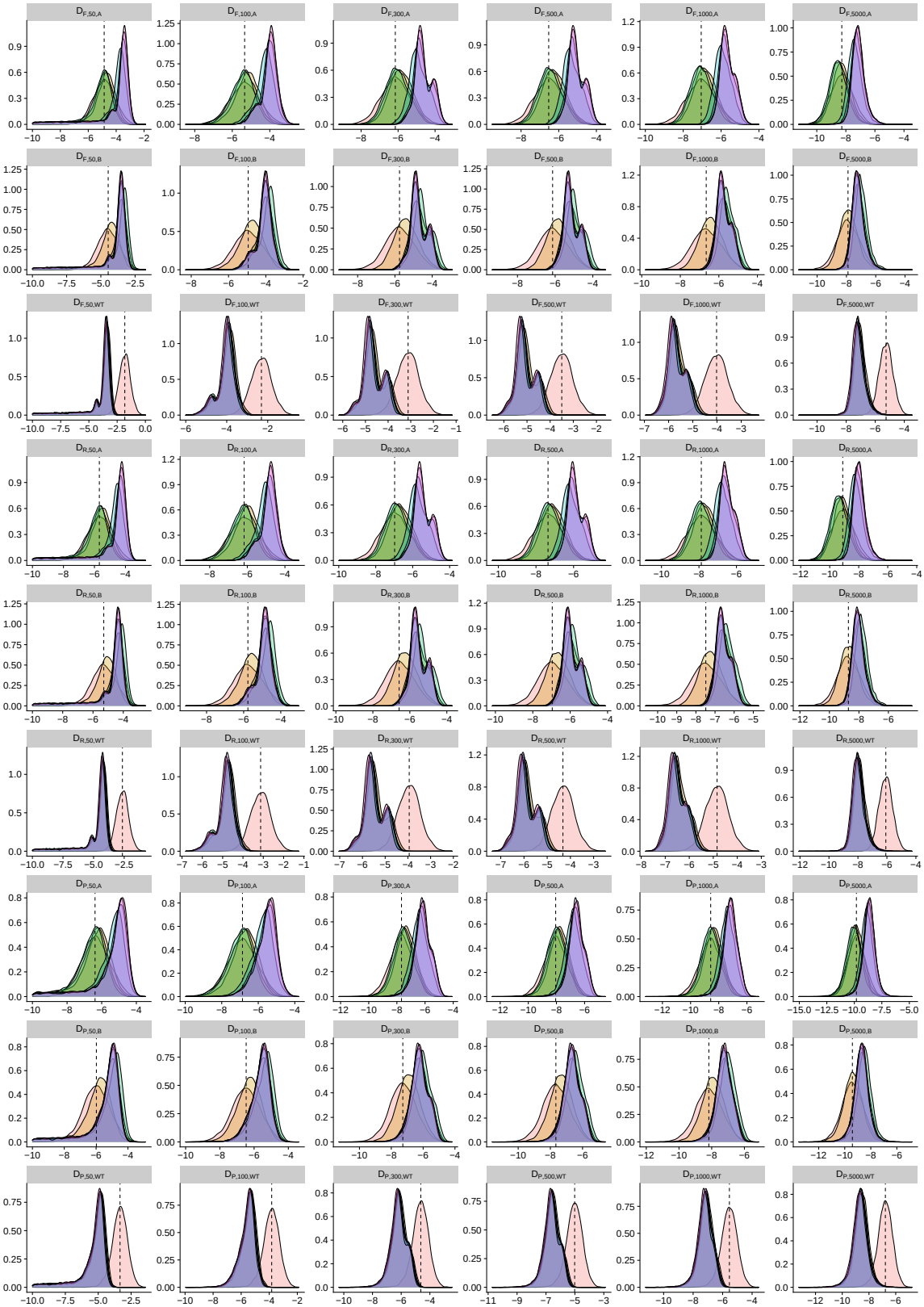


Figure S4: Prior (red) versus posterior marginal distributions of selected parameters and variables of the water contamination module of the augmented Bayesian network. Dashed lines represent the median of the prior marginal distributions. Distributions represented are in $\log_{10}(CFU/day)$.

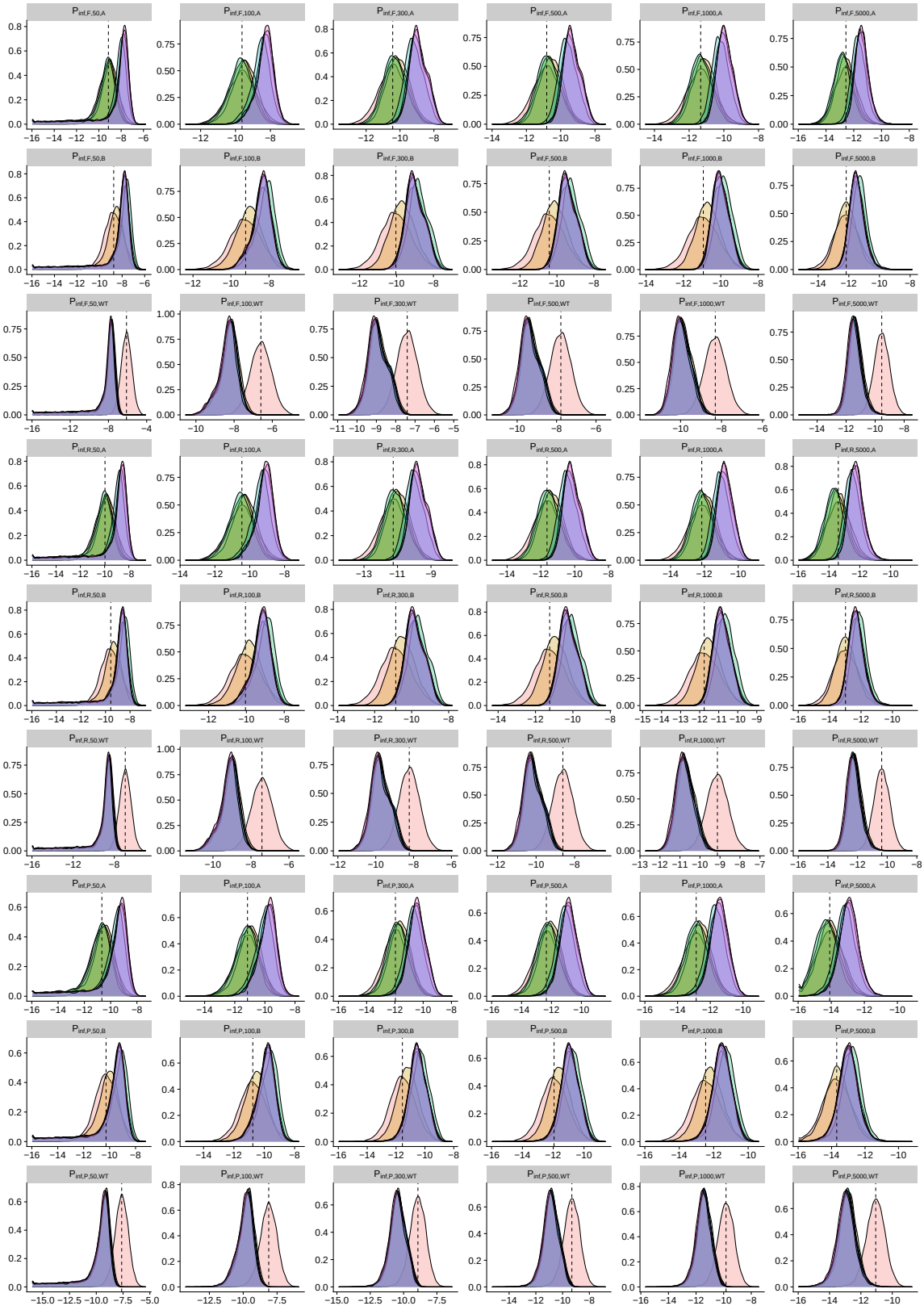


Figure S5: Prior (red) versus posterior marginal distributions of selected parameters and variables of the water contamination module of the augmented Bayesian network. Dashed lines represent the median of the prior marginal distributions. Distributions represented are in $\log_{10}$ of the instantaneous probability of subclinical severity infection for the farmers, residents and passersby.

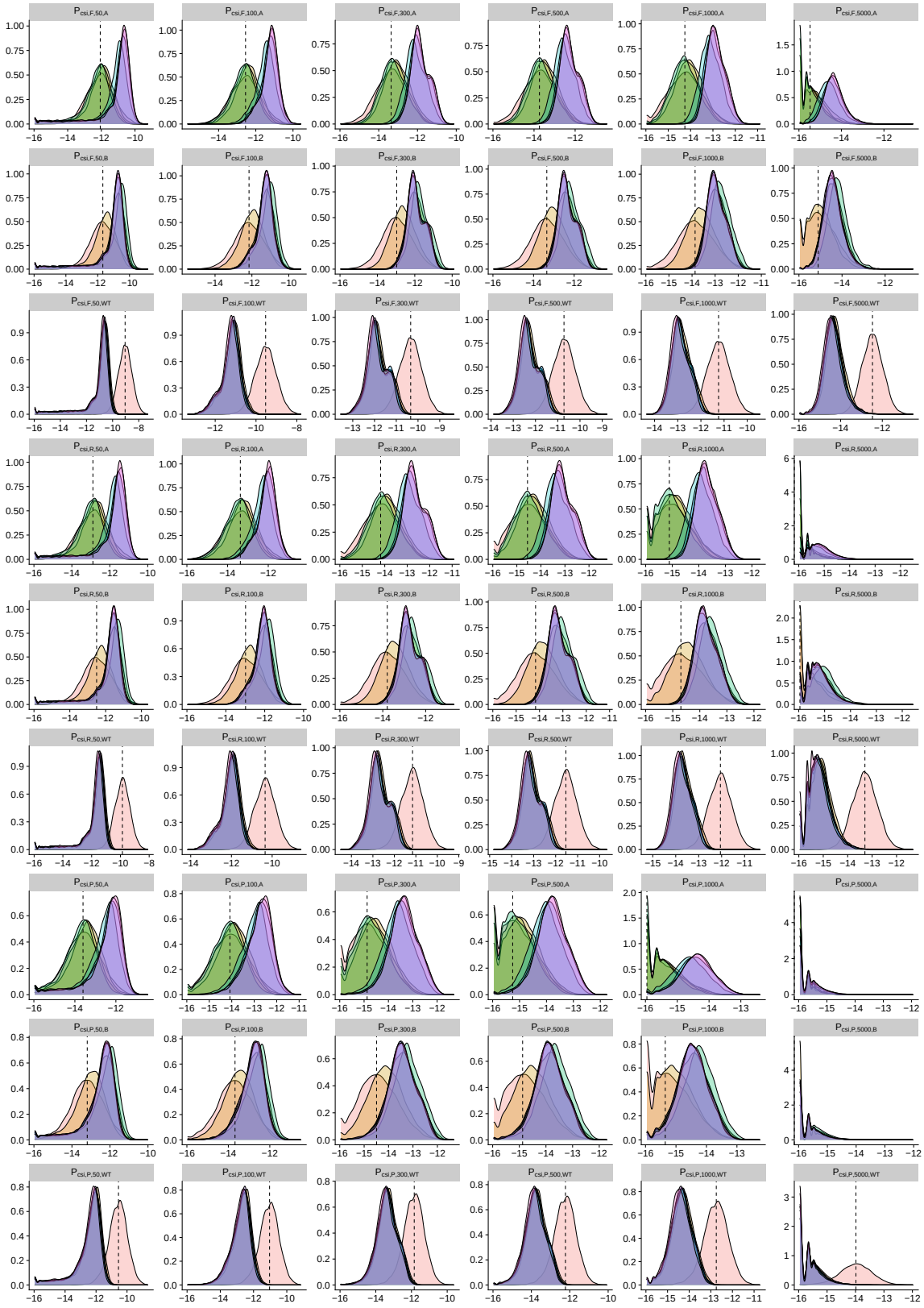


Figure S6: Prior (red) versus posterior marginal distributions of selected parameters and variables of the water contamination module of the augmented Bayesian network. Dashed lines represent the median of the prior marginal distributions. Distributions represented are in $\log_{10}$ of the instantaneous probability of clinical severity infection for the farmers, residents and passersby.

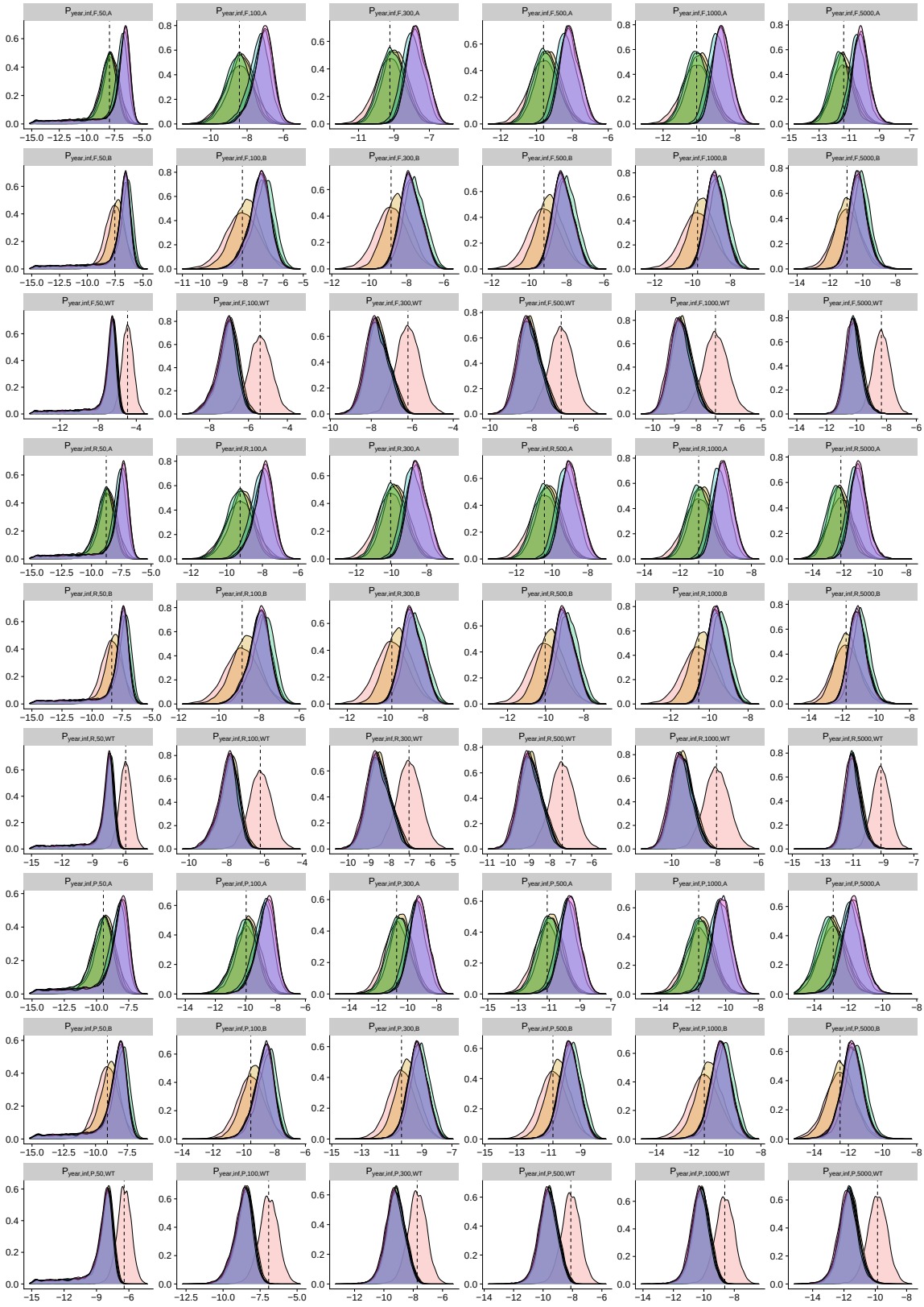


Figure S7: Prior (red) versus posterior marginal distributions of selected parameters and variables of the water contamination module of the augmented Bayesian network. Dashed lines represent the median of the prior marginal distributions. Distributions represented are in $\log_{10}$ of the yearly probability of subclinical severity infection for the farmers, residents and passersby.

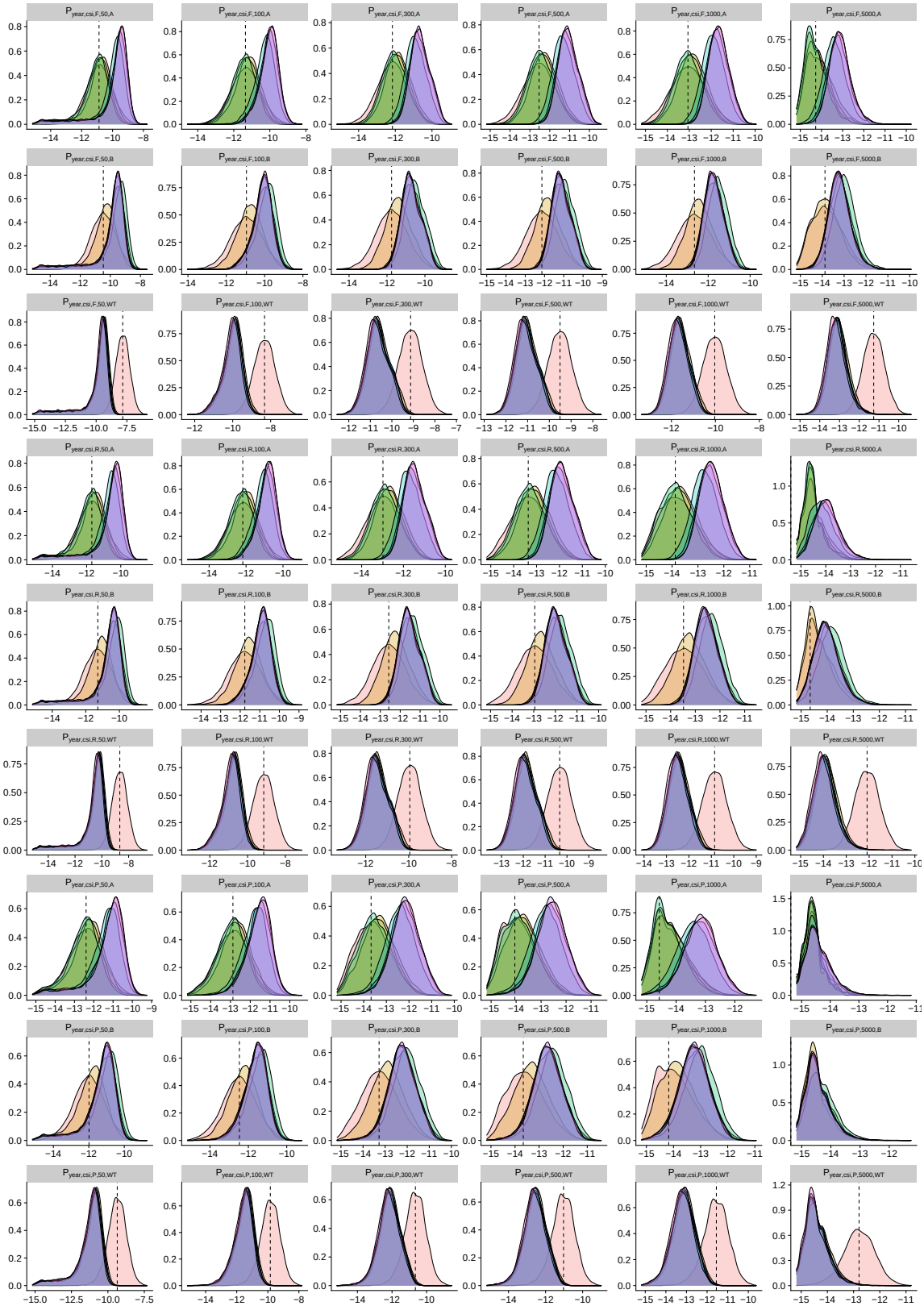
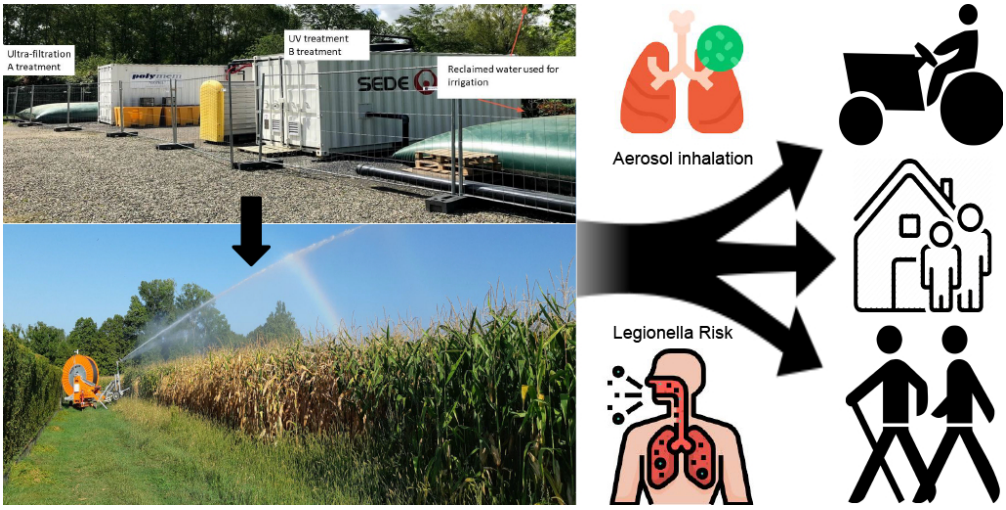


Figure S8: Prior (red) versus posterior marginal distributions of selected parameters and variables of the water contamination module of the augmented Bayesian network. Dashed lines represent the median of the prior marginal distributions. Distributions represented are in $\log_{10}$ of the yearly probability of clinical severity infection for the passersby.



43x21mm (600 x 600 DPI)

This paper presents a model for the risk assessment of pathogen inhalation from sprinkler irrigation using treated wastewater.