QMRA by farm to fork modelling

Isabelle Albert

Risk Assessmentôs 10th Anniversary Seminar
Finnish Food Safety Authority Evira, Helsinki, Finland
27 October 2011

QMRA by farm-to-fork modelling

Isabelle Albert

Researcher in statistics, Met@risk Unit, Food Risk Analysis Methodologies, MIA, INRA, France

DIET
AGRICULTURE
ENVIRONMENT

Outlines

Quantitative Microbial Risk Assessment by farm-to-fork modelling

> Introduction: Aims of a farm-to-fork modelling. Main case study: the fate of *Bacillus cereus* in a cooked, pasteurized and chilled food

> Part I: How to model the food pathway?

> Part II: Which distributions for the model inputs?

> Part III: How to validate the model?

Introduction

Aims of a farm-to-fork modelling

> In order to assess an unknown risk through the modelling of known stages (into Codex guidelines)

> Very unknown risk: large consuming population and large variability of the pathogen effects, many unobservable contaminations

> But well-known farming and industrial processes, partially known bacterial dynamics, expert knowledge and sparse data through the food chain

> Following the contamination of a product from farm-to-fork and beyond gathering sparse data and expert knowledge through the food chain

> In order to do sensitivity analyses: for example, scenarios

> Re-engineering, in order to elaborate control plans, to evaluate new policy

> A powerful tool if including uncertainty and variability of all quantities

Main case study
A product-pathogen pair

Bacillus cereus in a cooked, pasteurized and chilled food

Courgette purée (Refrigerated Processed Food of Extended Durability)

Each contamination is a mixture of 6 types of strains with their own characteristics of survival, growth potential and virulence (nothing to death)

<u>Aim:</u> to assess the contamination in a package before consumption according the type of bacteria

Part I: How to model the food pathway?

How to model the food pathway?
Introduction

Which product? Which population? Where to start? Where to finish?

> Sources of a foodborne disease? Where does the serving at risk come from (raw products?, firms?)? Which population of interest?

> To start: initial contamination is required (probability of contamination of the raw product, quantity of bacteria (cfu) with its associated uncertainty and variability

> To finish: product contamination at a process step, population exposure (if consumption data), number of diseases (if a dose-response model)

> A time unit for exposure, number of disease: per day, yearé

> A huge task! So, a well-defined framework, a representative case for decision

How to model the food pathway?
Case study

Bacillus cereus in a cooked, pasteurized and chilled food

A specific food/pathogen couple: 6 types of *B. cereus* strains/a 400g package of courgette purée

The beginning: the contamination in a batch of raw courgettes (344 kg) in the factory.

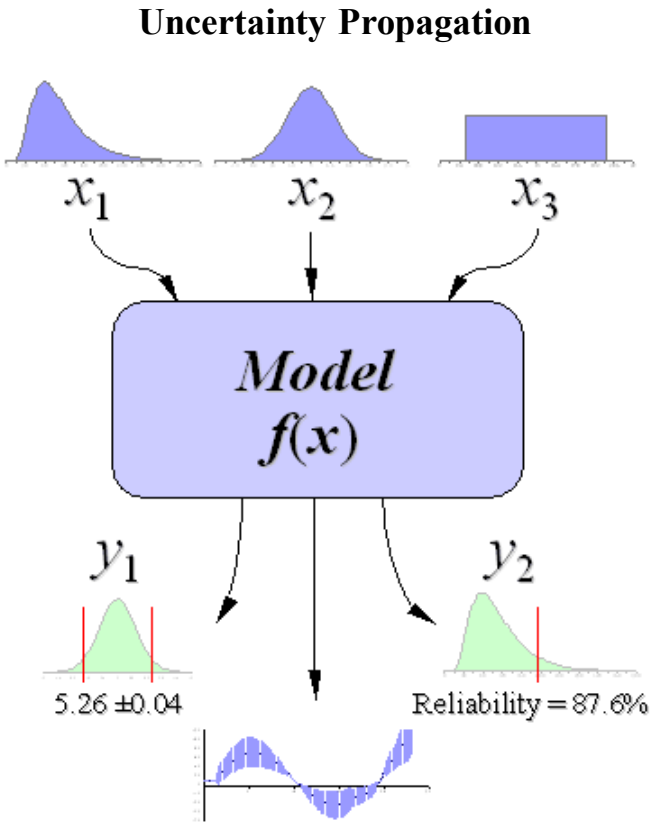
The end: the contamination in a package (400g) after home storage in refrigerators

The numeric tool

> Standard approach: Monte Carlo simulation

> In order to take into account the uncertainty associated with the knowledge and the variability of the phenomenon, and because no analytic solution exists to the final distributions of interest

> To sample from random distributions (pseudo-random sampled numbers from a software)



Monte Carlo simulation

> Number of iterations? (*N*, *i* indexes the iteration)

> Depending on the desired precision

> For a mean: (law of large numbers)

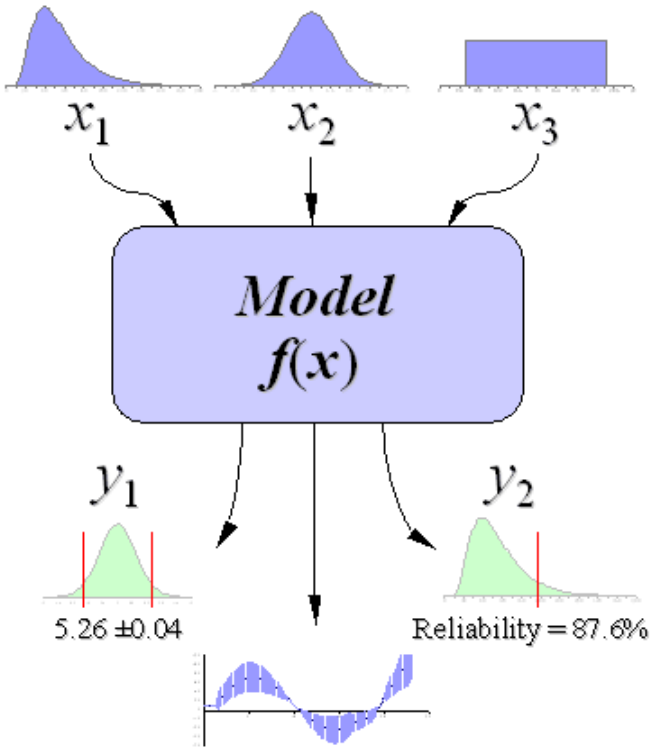
$$P\left(\left|\frac{1}{N}S_N(f)-E(f)\right|\geq t_\alpha\sqrt{\frac{V_{estim}(f)}{N}}\right)\leq\alpha$$

$$S_N(f)=\frac{1}{N}\left(f(\mathbf{x}^1)+.....f(\mathbf{x}^N)\right)$$

$$V_{estim}(f)=\frac{1}{N}\left(f\left(\mathbf{x}^i\right)-S_N(f)\right)^2$$

t_α α-th quantile of a Normal(0,1)

> Stability of the results comparing to their standard error

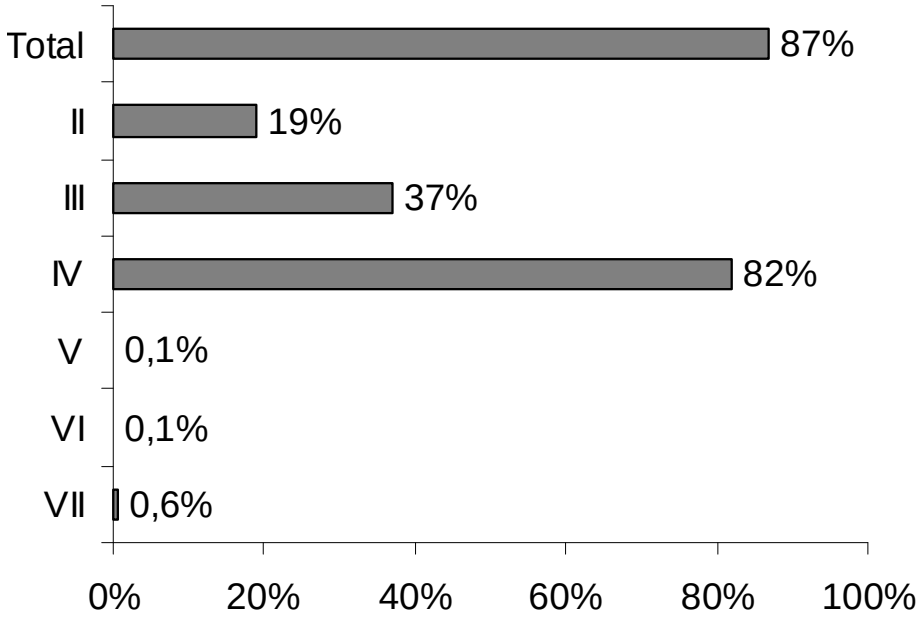


How to model the food pathway?

Outputs in the case study

Bacillus cereus in a cooked, pasteurized and chilled food

After iterating *2000 batches and 100 packages by batch*: % of packages containing at least one *B. cereus* spore (per genetic group or not (Total)) at the end of home storage in refrigerators

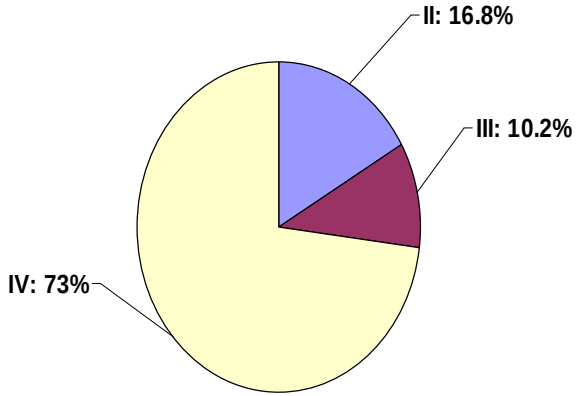


How to model the food pathway?

Outputs in the case study

Bacillus cereus in a cooked, pasteurized and chilled food

After iterating *2000 batches and 100 packages by batch*: mean proportion of each genetic group relatively to the whole *B. cereus* population in the contaminated packages at the end of home storage in refrigerators

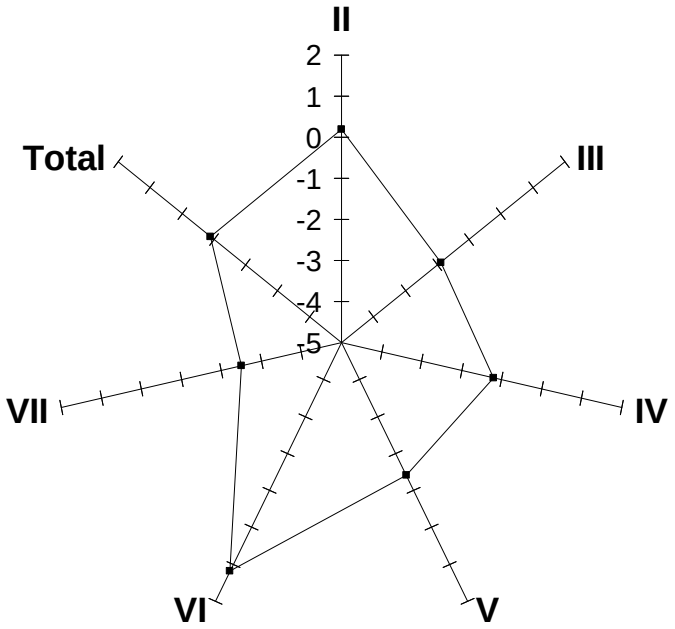


How to model the food pathway?

Outputs in the case study

Bacillus cereus in a cooked, pasteurized and chilled food

After iterating *2000 batches and 100 packages by batch*: mean concentration (log CFU/g) into the contaminated packages (per genetic group or not (Total)) at the end of home storage in refrigerators



How to model the food pathway?

Outputs in the case study

Bacillus cereus in a cooked, pasteurized and chilled food

After iterating *2000 batches and 100 packages by batch*: confidence interval at 95% of the *B. cereus* concentration into the contaminated packages (Total or per genetic group) at the end of home storage in refrigerators (and at process stages)

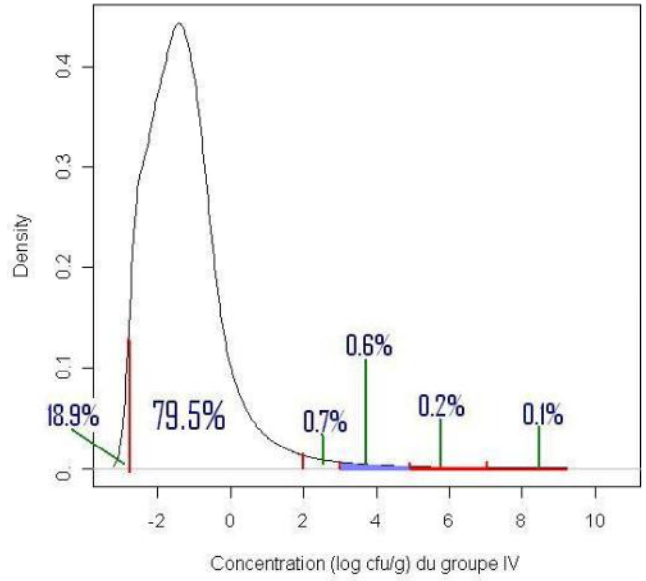
	Initial contamination of 300-kg batches of raw courgettes	After cooking of 300-kg batches of raw courgettes	After mixing with ingredients and partitioning into 400-g packages	After pasteurisation of 400-gpackages	At the end of home storage in refrigerators
II	[-1.5;0]	[-5.5;-0.6]	[-2.6;-0.6]	[-2.6;-0.8]	[-2.6;6.7]
III	[-4;-2.4]	[-5.5;-2.8]	[-2.6;-1]	[-2.6;-1.1]	[-2.6;0]
IV	[-4;-2.4]	[-5.5;-2.9]	[-2.1;-0.4]	[-2.6;-0.6]	[-2.6;1.7]
V	[-4;-2.5]	[-5.5;-3.1]	[-2.6;-2.3]	[-2.6;-2.6]	[-2.6;3.4]
VI	[-1.6;0]	[-5.5;-1.5]	[-2.6;-1.8]	[-2.6;-2.3]	[-2;7.8]
VII	[-4;-2.4]	[-5.5;-2.7]	[-2.6;-2.1]	[-2.6;-2.1]	[-2.6;-2.1]
Total	[-1;0.2]	[-5.5;-0.6]	[-1.8;-0.4]	[-2.6;-0.5]	[-2.6;3.5]

How to model the food pathway?

Outputs in the case study

Bacillus cereus in a cooked, pasteurized and chilled food

After iterating *2000 batches and 100 packages by batch*: distribution of the concentration of the group IV and % of packages at risk if the bacteria will not be killed during the preparation of the product (by microwave cooking)



high risk (more than 5 log cfu/g)

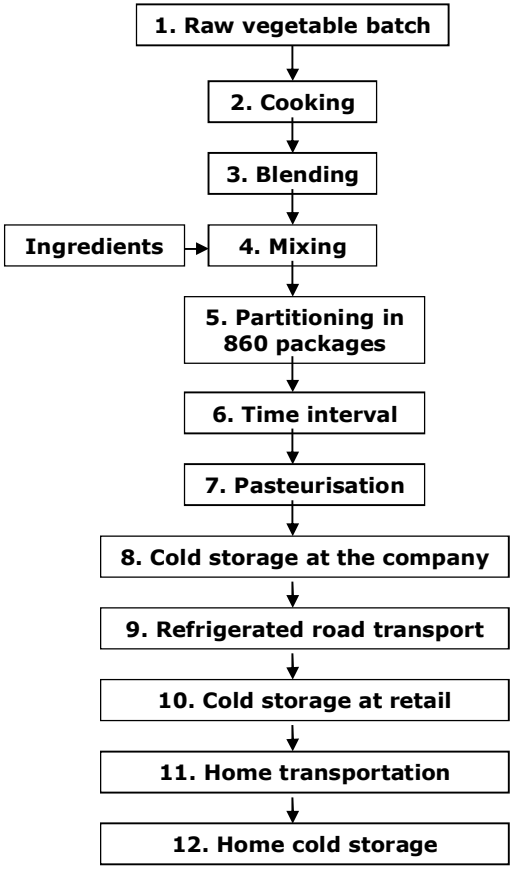
moderate risk (between [3 ; 7] log cfu/g)

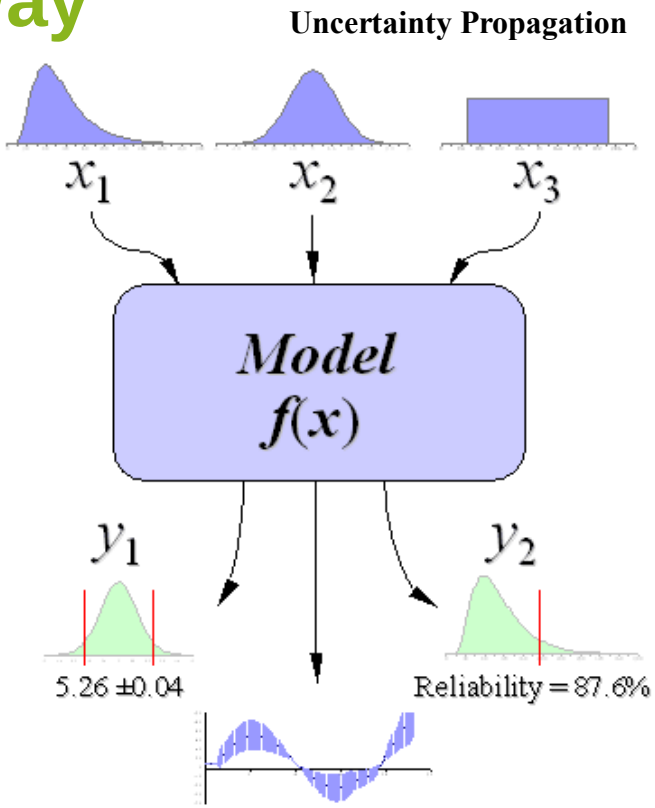
low risk (below 3 log cfu/g)

How to model the food pathway?

Model

The model: the contamination change through the food pathway





What level of model?

> Continuous time or not (modelling the main steps)?

> Continuous time necessitates differential equation of the variation of contamination

For example: bacterial growth model

$$\begin{cases} \dfrac{dy(t)}{dt} = 0, \quad t \le \lambda(T) \\ \dfrac{dy(t)}{dt} = \mu_{\max}(T)(1-\exp(y(t)-y_{\max})), \quad t > \lambda(T) \\ y(0) = y_0, \end{cases}$$

What level of model?

> Continuous time or not (modelling the main steps)?

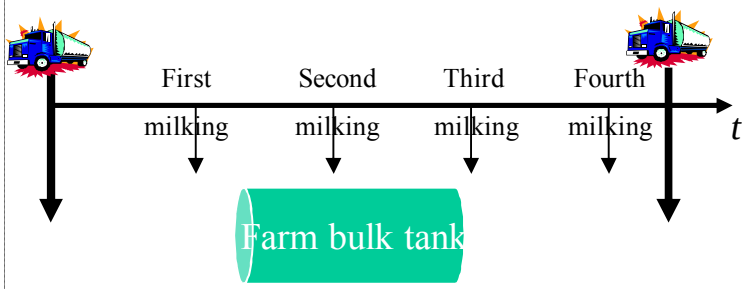
> Continuous time can necessitate differential equation of the factors which influence the contamination. For example: temperature in milk depending on milk volume

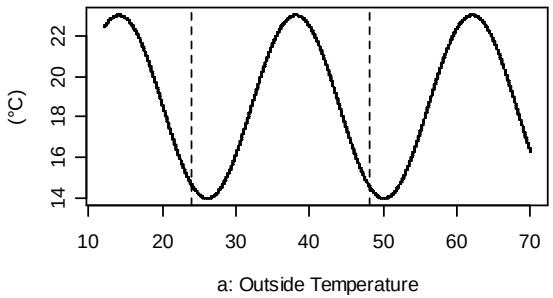
$$dT_a = \frac{VT + T_0 dV}{V + dV} - T$$

> Could be computational intensive because of a long pathway. Does one want to assess the contamination at each time (*t*)? Which benefits to consider the variation of the influencing factors at each time?

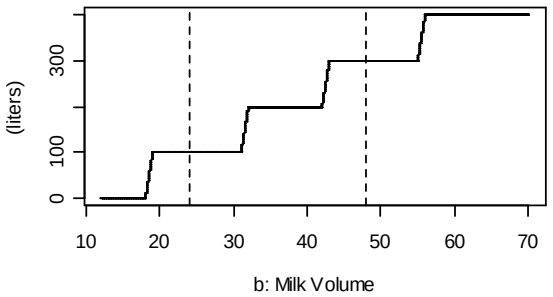
How to model the food pathway?
Model
Example

Modelling *Listeria monocytogenes* growth in farm tank milk

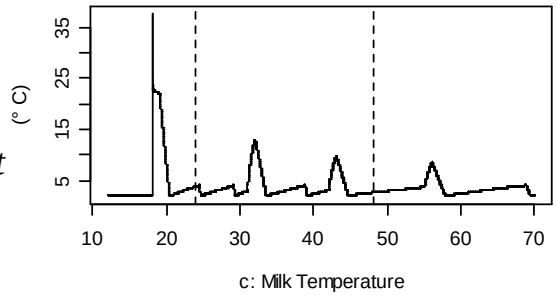




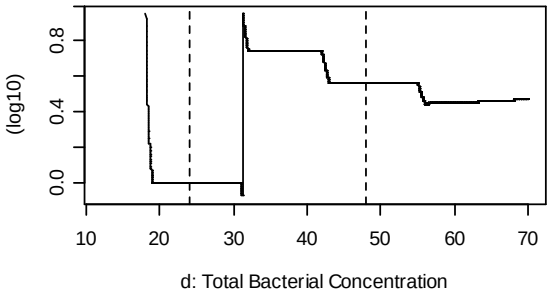
a: Outside Temperature



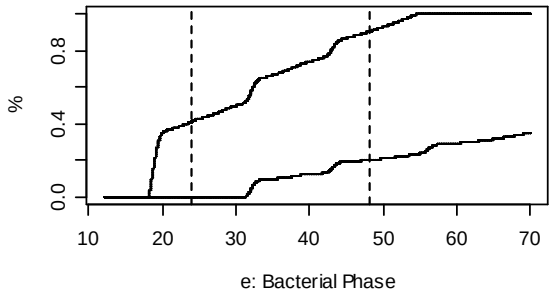
b: Milk Volume



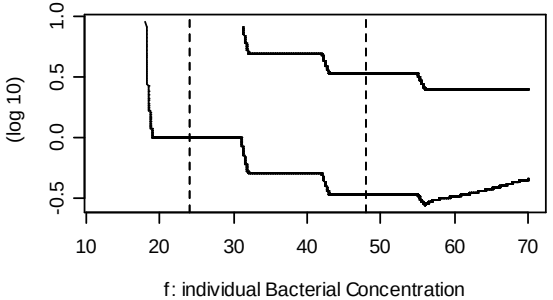
c: Milk Temperature



d: Total Bacterial Concentration



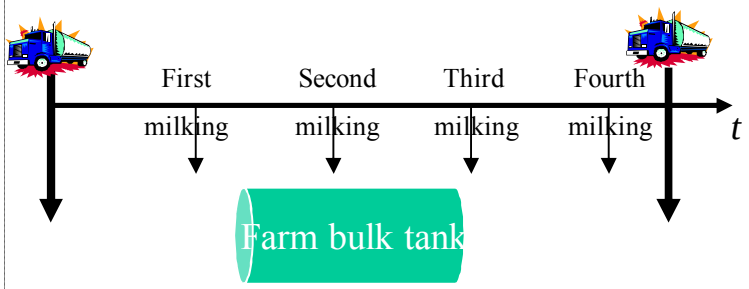
e: Bacterial Phase



f: individual Bacterial Concentration

I. Albert, R. Pouillot et J.-B. Denis.
Stochastically modeling Listeria monocytogenes growth in farm tank milk, Risk Analysis, 25 : 1171-1185, 2005.

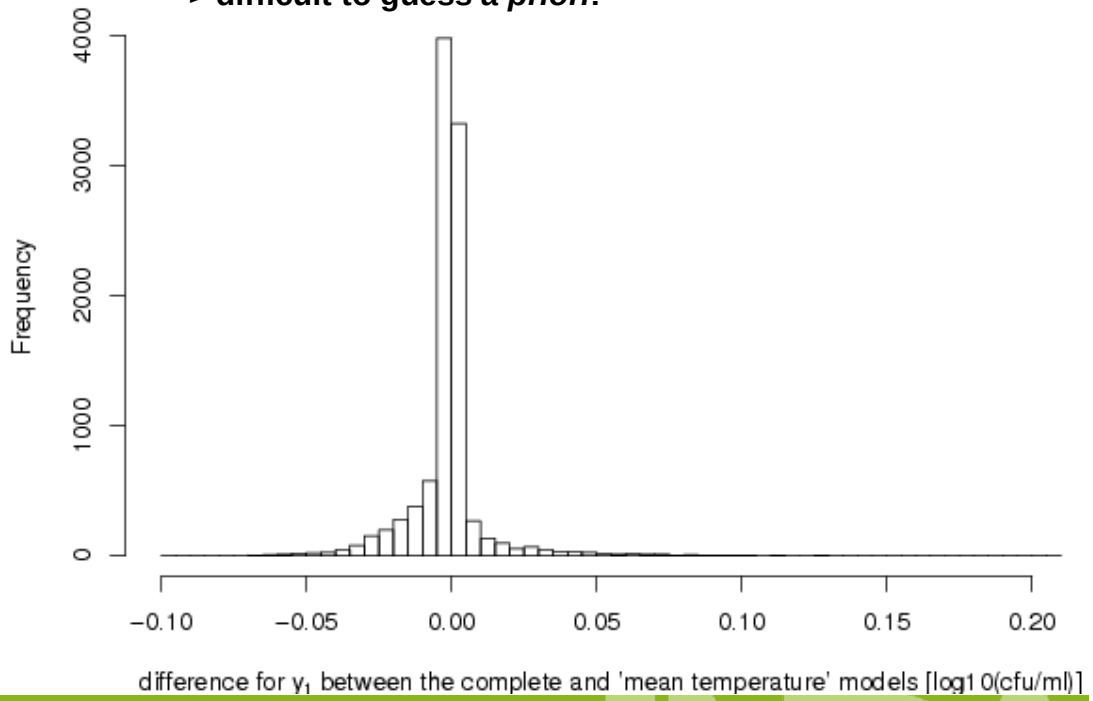
Modelling *Listeria monocytogenes* growth in farm tank milk



I. Albert, R. Pouillot et J.-B. Denis.
Stochastically modeling Listeria monocytogenes growth in farm tank milk, Risk Analysis, 25 : 1171-1185, 2005.

> No difference between the final concentration (y_1) in the complete model (temperature varying at each time) and in the model with a constant temperature (mean of the varying temperature)

> difficult to guess *a priori*!



What level of model?

> If modelling by step, which steps?

> Do not forget the objectives

> Steps where results are needed

> For example: exposure

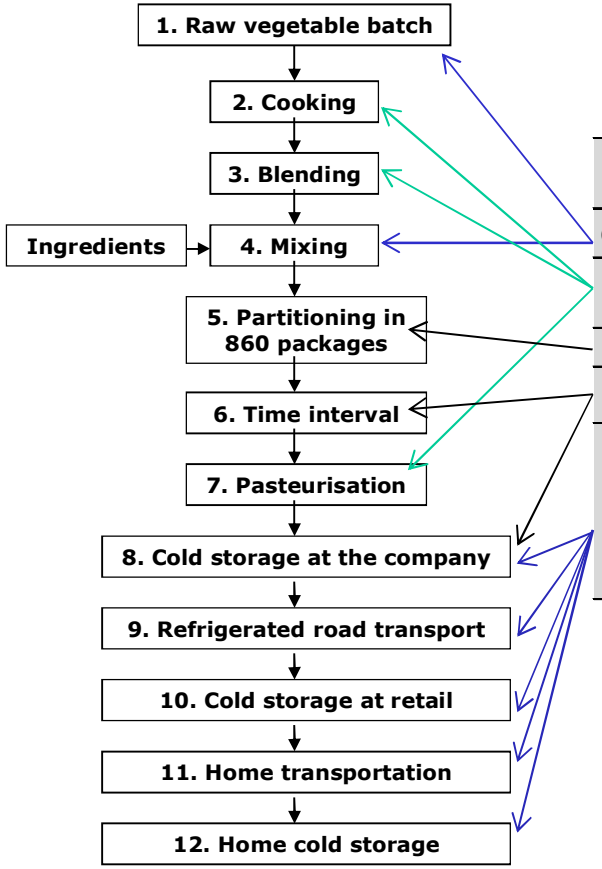
> Steps where impacting mechanisms are: new contaminations, bacteria dynamics changes, portioning, hygiene, consumption, dose-response effectsé

> Steps where data are

> For example: durability studies

How to model the food pathway?
Model in the case study

Bacillus cereus in a cooked, pasteurized and chilled food



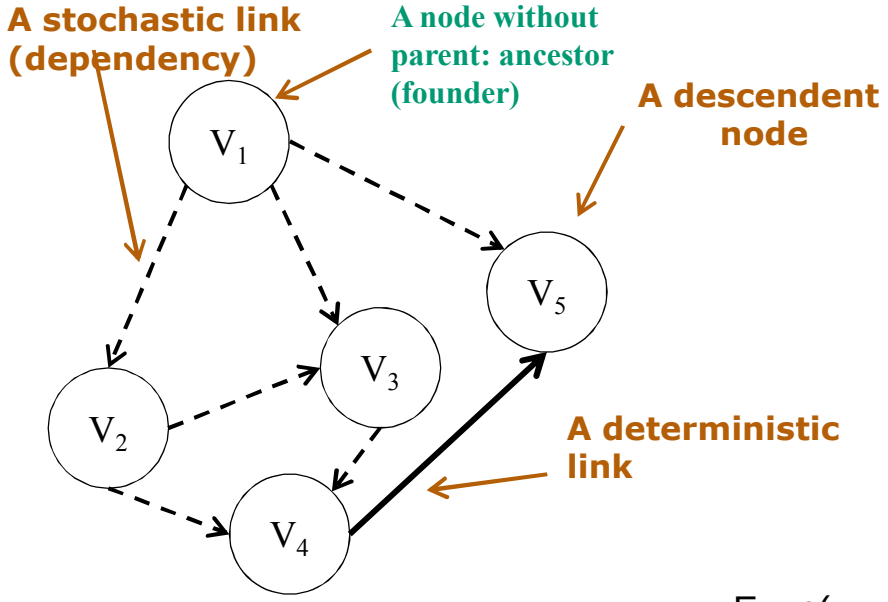
Impacting mechanisms	Step in the food pathway	Model or distribution
Contamination	1, 4	BetaPert distributions
Inactivation	2, 3, 7	Primary model: Weibull model (Mafart et al., 2002) Secondary model: Bigelow model (Couvert et al., 2005)
Partitioning	5	Multinomial distribution
Germination	6, 8	Empirical model (see Afchain et al., 2008)
Growth	8, 9, 10, 11, 12	Primary model: Logistic model with delay (Rosso et al., 1995) Secondary model: Cardinal Temperature model with inflection point (Rosso et al., 1993) Competition model (see Afchain et al., 2008)

How to model the food pathway?

Model

Construction of a Bayesian network with probabilistic or deterministic links between the nodes

Helpful tool: Bayesian Networks (BN)



A probabilistic graphical model

- G=(V, E) a finite Directed Acyclic Graph
- $V=(V_i)_{i\in V}$ a set of random variables (nodes)

$V=(V_i)_{i\in V}$ is a Bayesian Network w.r.t the Graph G if for all i in V :

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

desc(i): descendants of node i, any node on a directed path starting from i
pa(i): parents of node i, any node with an arrow emanating from it pointing to i

$V=(V_i)_{i\in V}$ is a Bayesian Network w.r.t the graph G if:

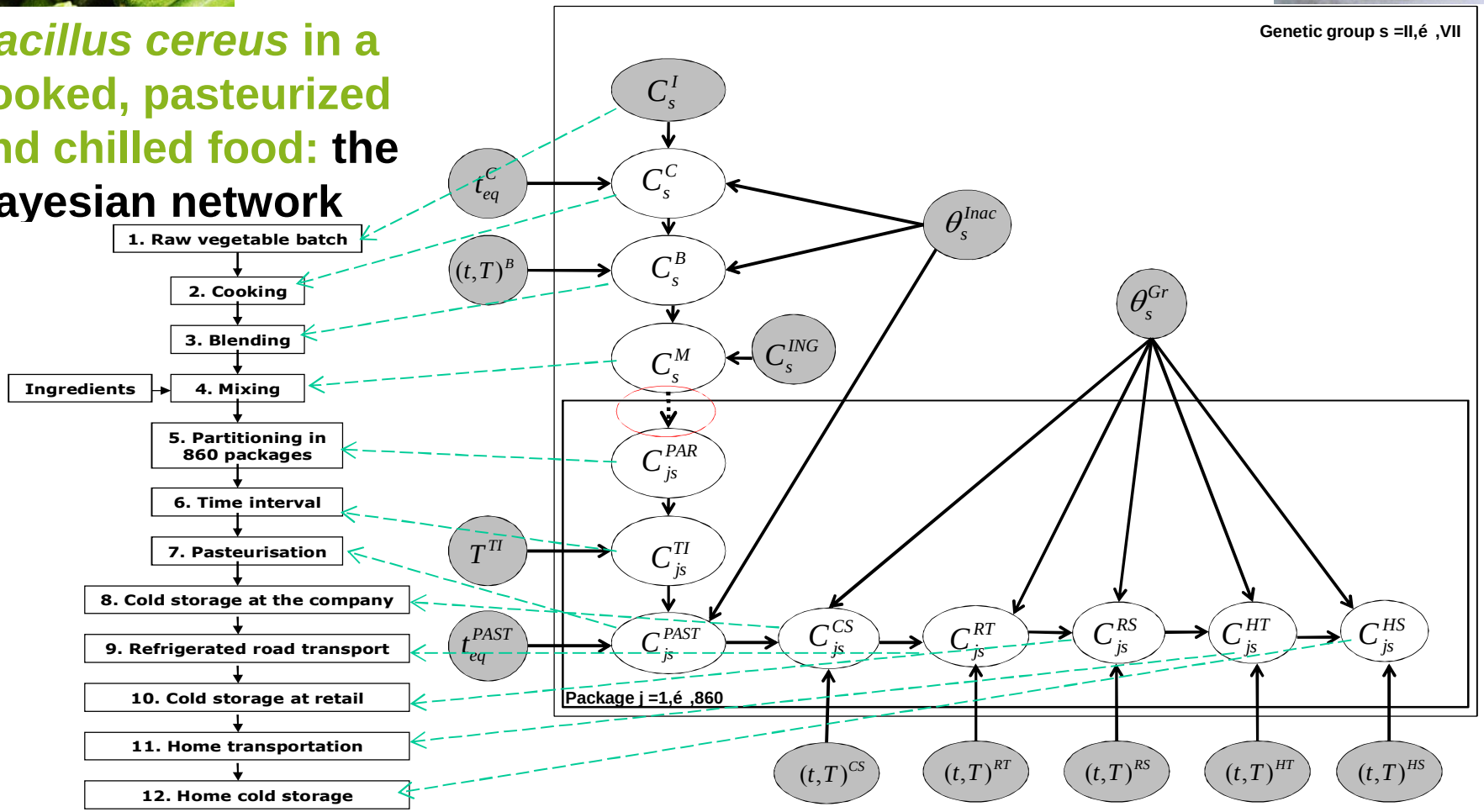
$$p(V) = \prod_{i\in\Gamma} p(V_i \mid V_{pa(i)})$$

Ex: $p(v_1,v_2,v_3,v_4,v_5) = p(v_1)p(v_2|v_1)p(v_3|v_1,v_2)p(v_4|v_2,v_3)p(v_5|v_1,v_4)$

How to model the food pathway?

Model in the case study

***Bacillus cereus* in a cooked, pasteurized and chilled food: the Bayesian network**



How to model the food pathway?

Model

The functional or deterministic links

➢ **The heat inactivation model**: Weibull model (Peleg and Cole, 1998) and Biglow model (Biglow et al., 1920). *t*: duration of the thermal process

$$\log_{10}\left(\frac{N_{t,s}}{N_{0,s}}\right) = -\left(\frac{t_{eq}}{\delta_s^{'}(T)}\right)^{p_s} \qquad with\ \delta_s^{'}(T) = \delta_s\, 10^{(T_r - T)/z}$$

$$\Rightarrow \quad \theta_{inac} = (\delta_s, p_s, z)$$

➢ **The growth model:** logistic model with delay (Rosso et al., 1995)

$$N_{t,s} = \frac{N_{max}}{1 + \left(\frac{N_{max}}{N_{0,s}} - 1\right) e^{-\mu^s_{max}(T)\,t}} \qquad \forall\ t \geq 0$$

$$\mu^s_{\max}(T) = \begin{cases} \dfrac{\mu_{opt}(T - T^s_{\max})(T - T^s_{\min})^2}{(T^s_{opt} - T^s_{\min})\left[(T^s_{opt} - T^s_{\min})(T - T^s_{opt}) - (T^s_{opt} - T^s_{\max})(T^s_{opt} + T^s_{\min} - 2T)\right]} & T \in \left[T^s_{\min}, T^s_{\max}\right] \\ 0 & T \notin \left[T^s_{\min}, T^s_{\max}\right] \end{cases}$$

$$\Rightarrow \quad \theta_{gr} = (N_{max}, \mu_{opt}, T^s_{min}, T^s_{max}, T^s_{opt})$$

+ a competition model to force:

$$\sum_s N_{t,s} \leq N_{max}$$

How to model the food pathway?

Model

Where do these models come from?

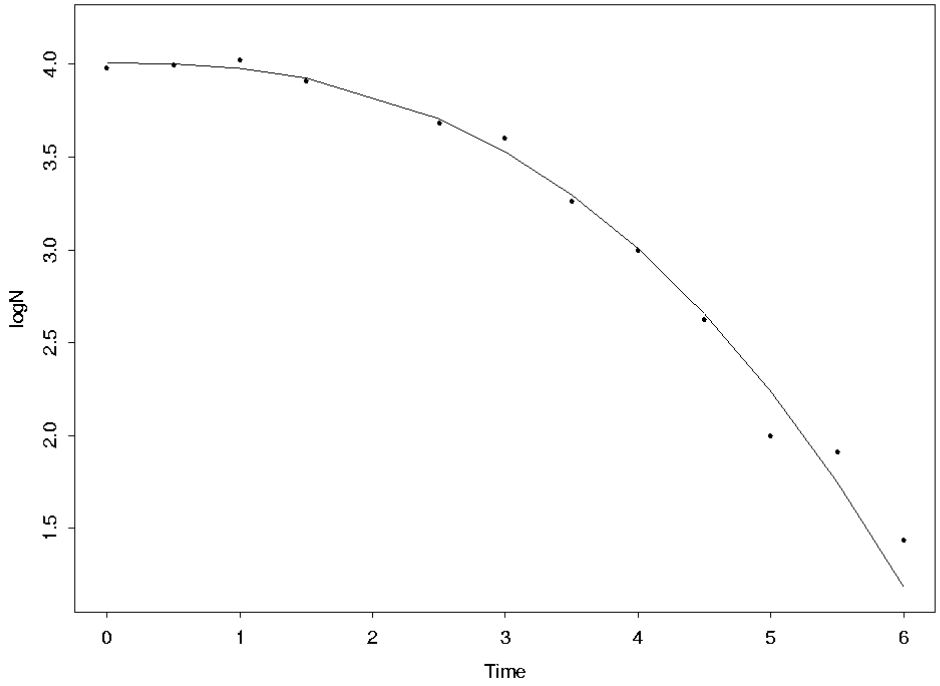
> Predictive microbiology

> For example:

Inactivation models

Primary models (dynamics in constant environment)

Fitting a model to experimental data



Inactivation model

The model:

$$\gamma(t) = (N_0 - N_{res})10^{\left(-\left(\frac{t}{\delta}\right)\right)^p} + N_{res}$$

I. Albert and P. Mafart. A modified Weibull model for Bacterial inactivation. *International Journal of Food Microbiology,* 100 : 197-211, 2005.

- t is time; $N_0 > 0$, is the true unknown value of the initial number of bacteria at time $t = 0$; $N_{res} > 0$, is the true unknown value of the residual number of bacteria (**observation** of a stop in the population decrease **at the end of the heat treatment when sigmoidal curve**)

Å $\delta\ (>0)$ represents the time of the first decimal reduction concentration for the non-resistant bacteria: at $t = \delta$,

$$\frac{\gamma(\delta) - N_{res}}{N_0 - N_{res}} = \frac{1}{10}$$

Inactivation model

The statistical regression model:

Expectation model:

$$N(t) = \gamma(t) + \varepsilon(t)$$

For $0 \le t \le t_n$,

$\varepsilon(t)$ are the discrepancies between the observations $N(t)$ and their expectations $\gamma(t)$ (because of the biological variability of the phenomenon)

$\varepsilon(t)$ are centred random variables $\left(E[\varepsilon(t)] = 0\right)$, assumed to be independent

Variance model:

$$Var[\varepsilon(t)] = \sigma_t^2 = \sigma^2 \gamma(t)^q$$

The variability of the response depends on its level

$q = 2$ corresponds to the logarithmic transformation (which can be tested)

One additional parameter to estimate but it allows to take into account a variance heterogeneity much more adapted to the observations

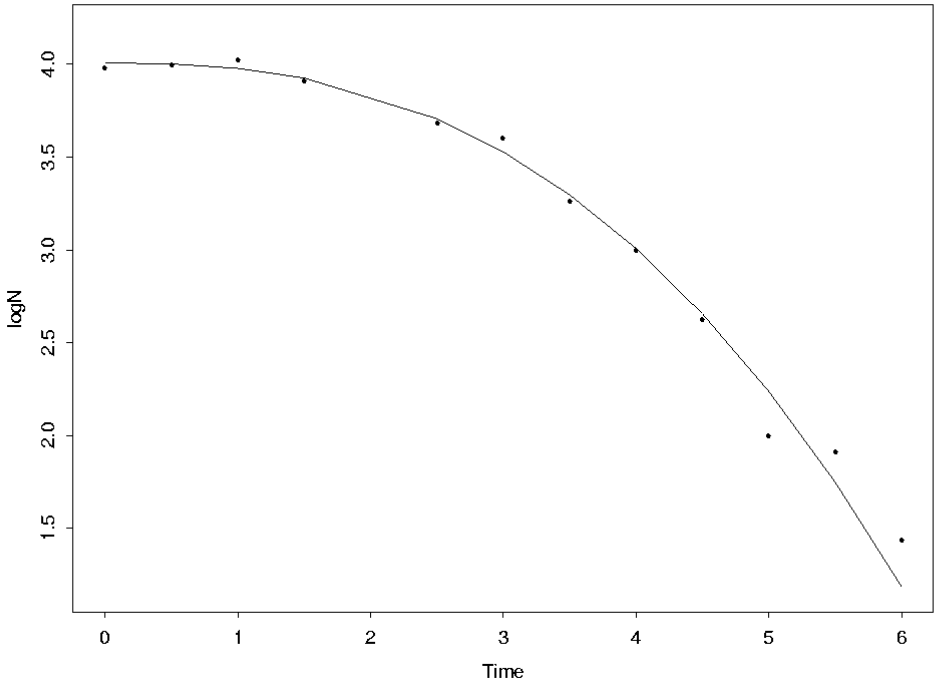
Inactivation model

Model fit to experimental data: nonlinear model with unequal variance errors (see function gnls in **R**)

Type A (N_{res}= 0)

Example: *Bacillus pumilus* at 98°C

Parameter	Mean	Standard deviation	Gaussian CI (95%)
δ	0.04	0.01	$[0.02; 0.06]$
p	0.32	0.02	$[0.29; 0.35]$
N_0	6×10^8	1×10^8	$[4\times10^8; 8\times10^8]$
q	1.89	0.03	$[1.83; 1.95]$
σ^2	0.33		



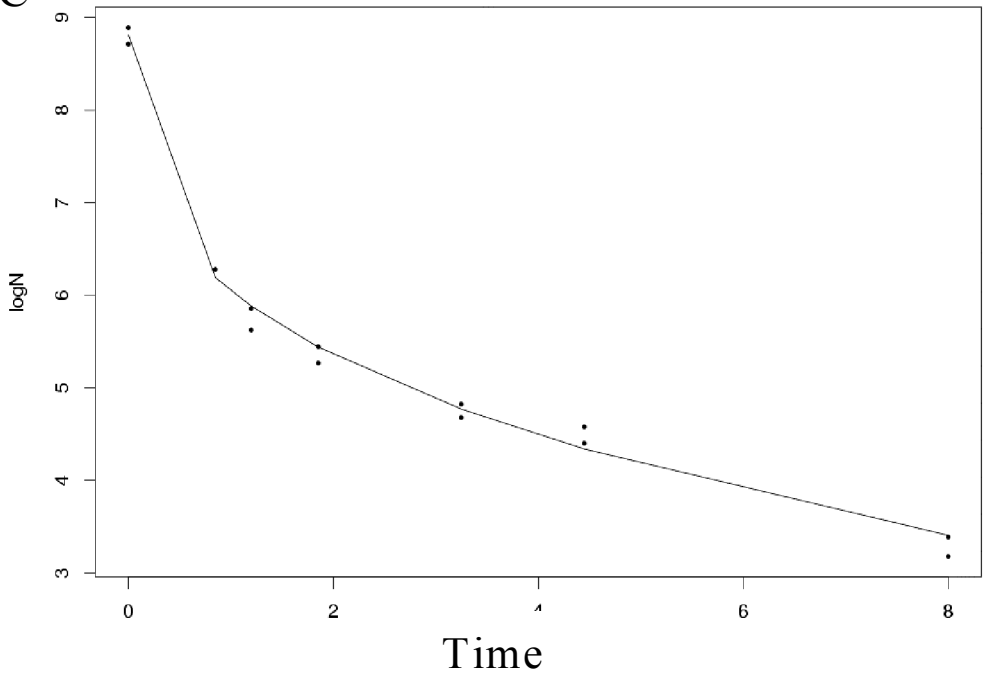
Inactivation model

Model fit to experimental data:

Type B (N_{res}= 0)

Example: *Clostridium botulinum* at 111°C

Parameter	Mean	Standard deviation	Gaussian CI (95%)
δ	0.04	0.01	$[0.02; 0.06]$
p	0.32	0.02	$[0.29; 0.35]$
N_0	6×10^8	1×10^8	$[4\times10^8; 8\times10^8]$
q	1.89	0.03	$[1.83; 1.95]$
σ^2	0.33		



How to model the food pathway?

Model

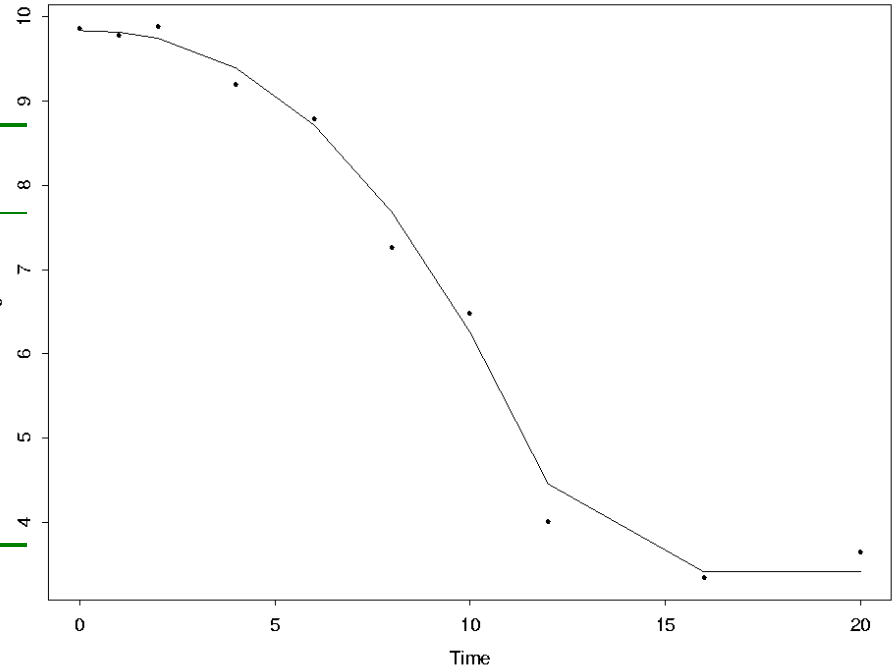
Inactivation model

Model fit to experimental data:

Type C

Example: *Listeria innocua* at 55°C

Parameter	Mean	Standard deviation	Gaussian CI (95%)
δ	5.7	0.27	$[5.17; 6.24]$
p	2.27	0.16	$[1.96; 2.58]$
N_0	6.9×10^9	1.1×10^9	$[4.6\times10^9; 9.1\times10^9]$
N_{res}	2 559	950	$[6\times10^2; 44\times10^2]$
q	1.87	0.03	$[1.82; 1.93]$
σ^2	1.45		

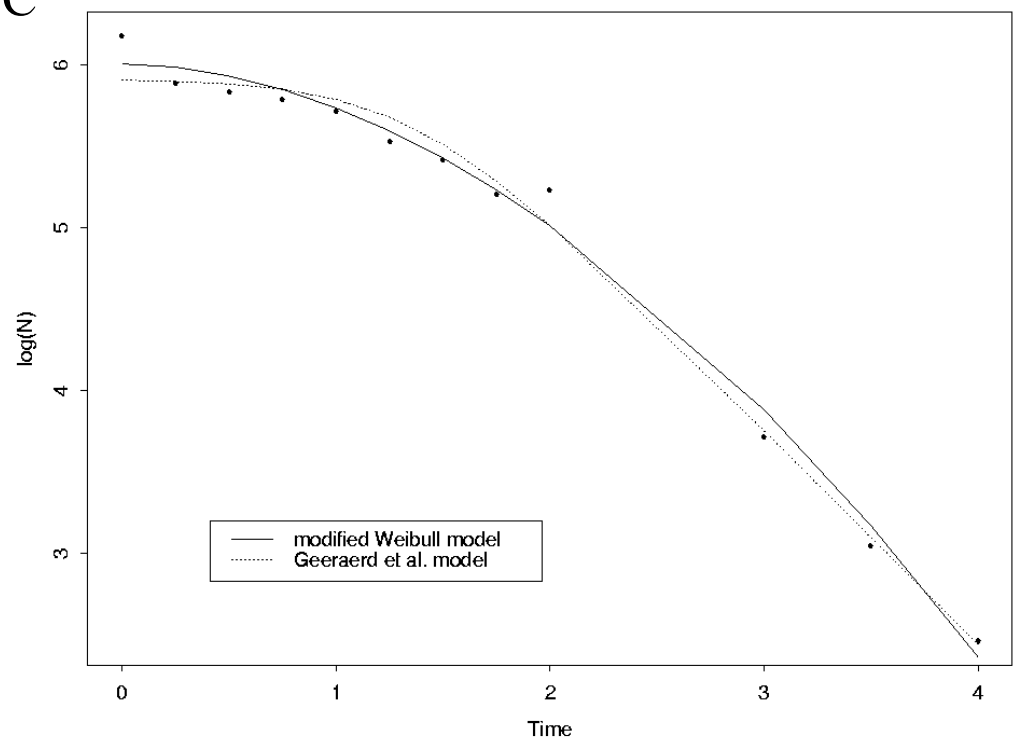


Inactivation model

Comparison to the model proposed by Geeraerd et al. (2000)

Example: *Bacillus cereus* at 96°C

Model	AIC value
Modified Weibull	291.53
Geeraerd et al.	290.91

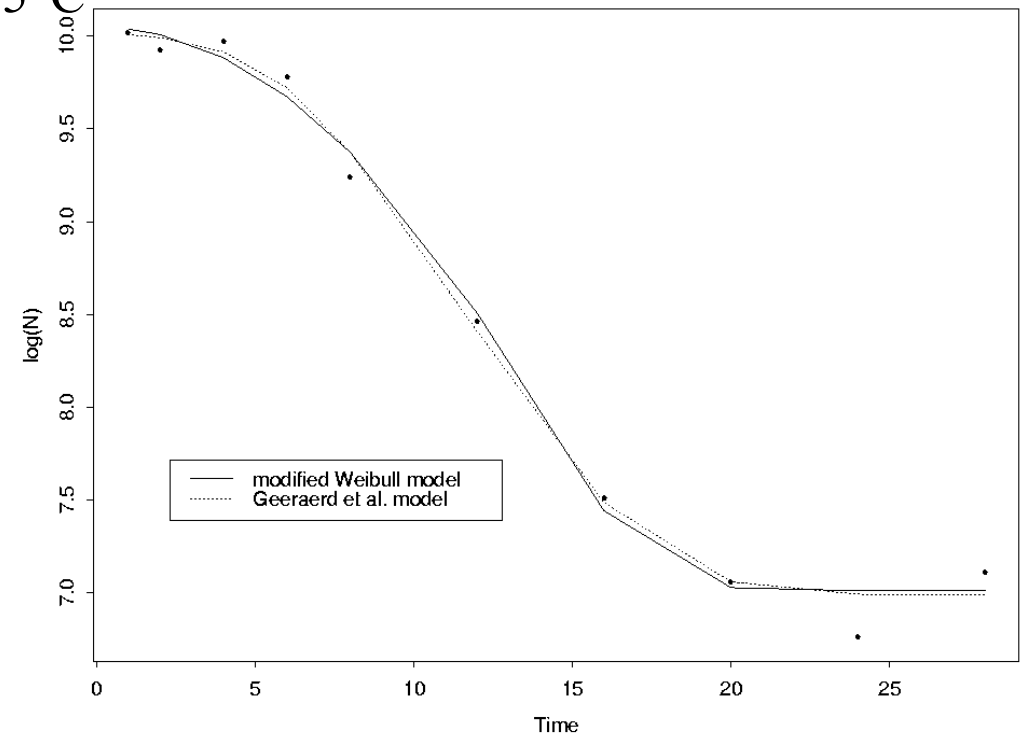


Inactivation model

Comparison to the model proposed by Geeraerd et al. (2000)

Example: *Listeria innocua* at 55°C

Model	AIC value
Modified Weibull	364.67
Geeraerd et al.	365.25



How to model the food pathway?

Model

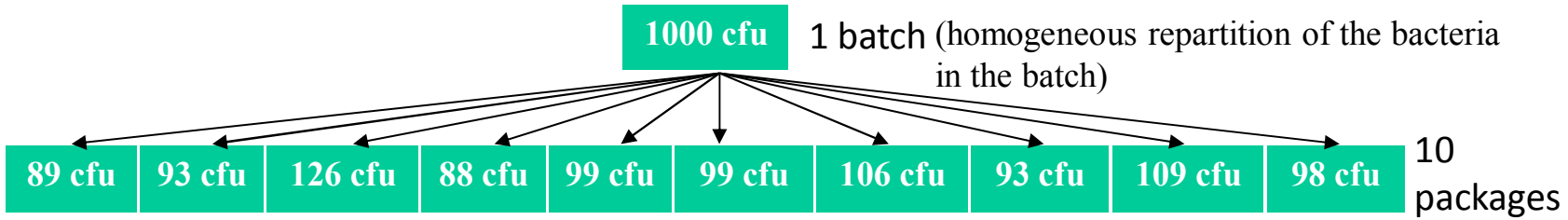
The stochastic links

➢ **Any stochastic link is possible but one needs an idea about the parameters of the assumed distribution**: **ex:** normal distribution on the logit of the probability of a contaminated product

$$\text{logit}\left(p_f\right) = N\left(m_f, s_f\right) \qquad m_f \sim N(0, 0.22) \qquad s_f \sim U(0, 0.2)$$

I. Albert, E. Grenier, J.-B. Denis, J. Rousseau. Quantitative Risk Assessment from Farm to Fork and Beyond: a global Bayesian Approach concerning food-borne diseases, *Risk Analysis*, 28(2) : 557-571, 2008.

➢ **For random impacting mechanisms in the food chain: ex:** partitioning (homogeneous contamination)



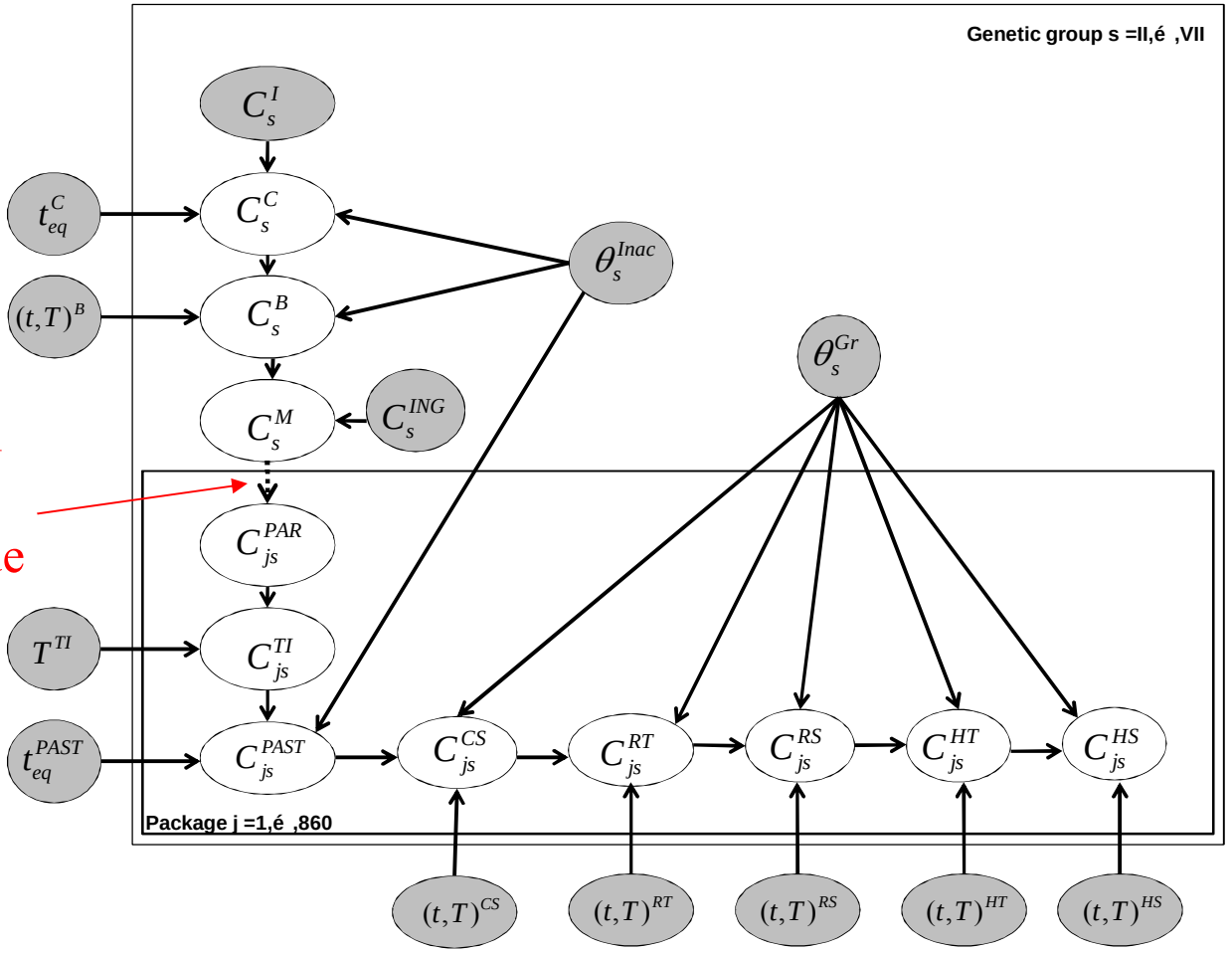
Multinomial distribution: n bacteria in a batch with probability $p_i = 1/10$ to be in the package i

Mult(n, (p_1,..,p_k)), avec p_1,..,$p_k = 1$ and $\Sigma_i n_i = n$

How to model the food pathway?
Model in the case study

***Bacillus cereus* in a cooked, pasteurized and chilled food: the Bayesian network**

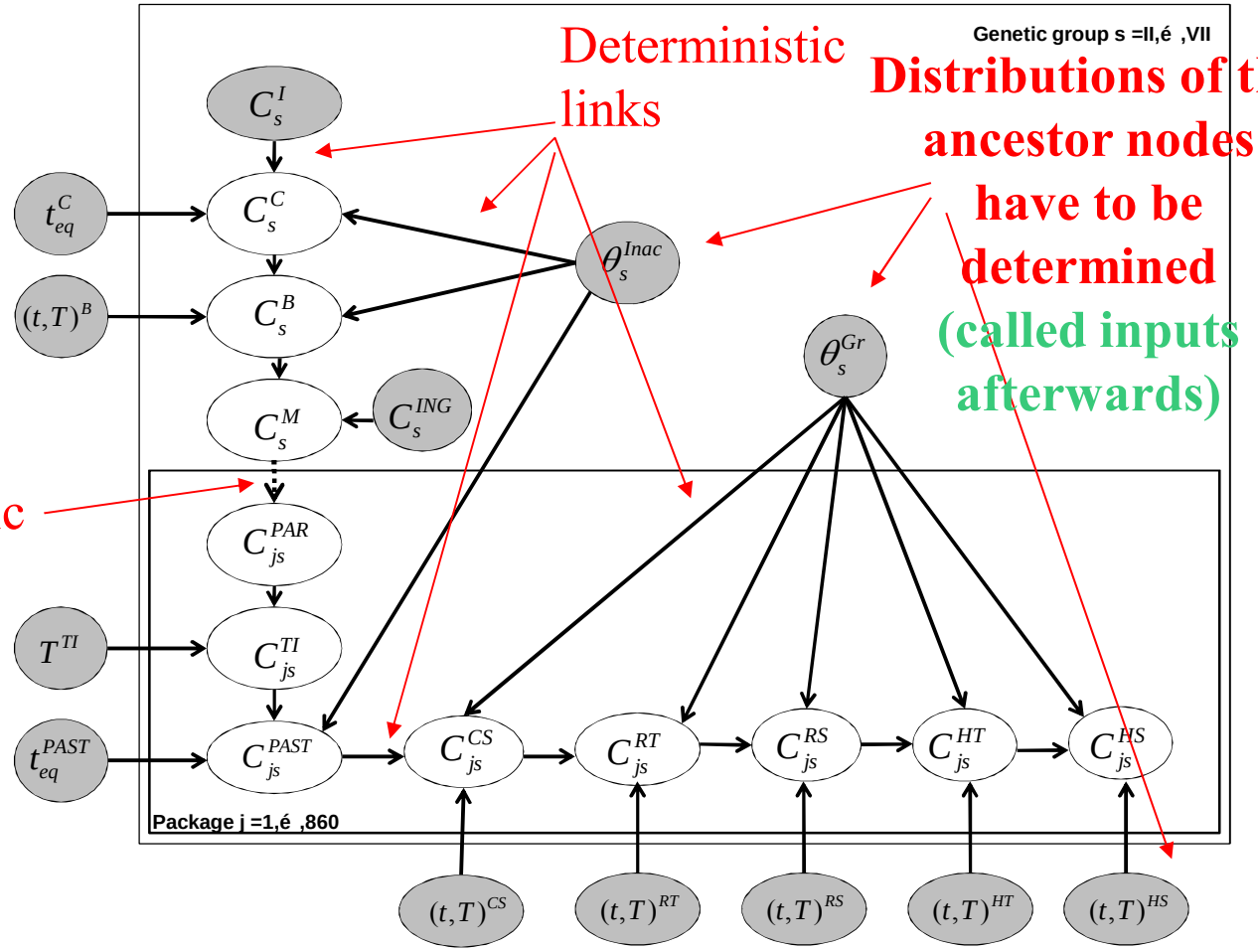
Stochastic link: multinomial distribution describing the partioning of the bacteria due to the portioning of a 344kg batch into 860 packages of 400g each



How to model the food pathway?

Model in the case study

Bacillus cereus **in a cooked, pasteurized and chilled food: the Bayesian network**



Part II: Which distributions for the model inputs?

Why probability distributions?

> Because of uncertainty and variability in all parameters and because these quantities are of interest

> No constant, always imperfect knowledge: Bayesian point of view

Which distributions?

- > Marginal approach:
  - > Classical approach:
    - > To fit a distribution to raw data
    - > To fit a distribution to expert opinion
  - > To sample directly into the data sets
- > Global approach: to keep all raw data and perform a global inference
  - > Evidence synthesis

Which distributions for the model inputs?
Marginal approach

Fitting distributions to data

> Estimating the parameters of the chosen distributions from the dataset

> Bayesian inference with WinBugs or OpenBugs

R. Pouillot, I. Albert, M. Cornu et J.-B. Denis. Estimation of uncertainty and variability in bacterial growth using Bayesian inference. Application to Listeria monocytogenes. IJFM, 81 : 87-104, 2003.

> Frequentist methods (ex: Max. Likelihood)

> R package ñfitdistrplusò (for continuous or discrete data)

R. Pouillot, ML. Delignette-Muller. Evaluating variability and uncertainty separately in microbial assessment using two R, IJFM, 142: 330-340, 2010.

Fitting distributions to data

> Assessing the goodness-of-fit

> Ex: Kolmogorov-Smirnov statistics:

$$KS = \sup_x \left| F_n(x) - F(x) \right|$$

Empirical cumulative distribution function

Cumulative distribution function of the fitted distribution

> A more sophisticate statistics: Anderson-Darling statistics

> Anderson-Darling test

> In Bayesian framework : posterior predictive model checks

> Goodness-of-fit Bayesian p-value

B.P. Carlin, T.A. Louis. Bayesian Methods for Data Analysis, CRC Press, Chapman & Hall, Third Edition, **page 88**

Fitting distributions to data

> Comparing different fitted distributions (models)

> Frequentist approaches:

> *Examples*: AIC, BIC for non nested models

> In a Bayesian framework:

> *Examples*: Bayesian Factor or DIC

(not always univariate distributions)

Which distributions for the model inputs?

Marginal approach in the case study

Bacillus cereus in a cooked, pasteurized and chilled food

Time-temperature profiles of raw vegetable cooking

18	17	13.6	14.9	9.5	15.1	13.7
45.8	17.1	18.6	30.8	28.8	37.6	42.7
62	46	33.4	39.3	32.6	57	40.6
71.3	61.6	59	52.1	32.3	67.6	38.9
79.7	75.1	78.9	63.9	32	76.4	37.3
86.9	86.9	78.9	74	31.7	80.5	64.2
93.1	95.4	85.9	83.4	31.6	87	75.2
98.1	99.2	90.8	89.8	31.4	92	83.4

Equivalent heat treatment times at 90°C

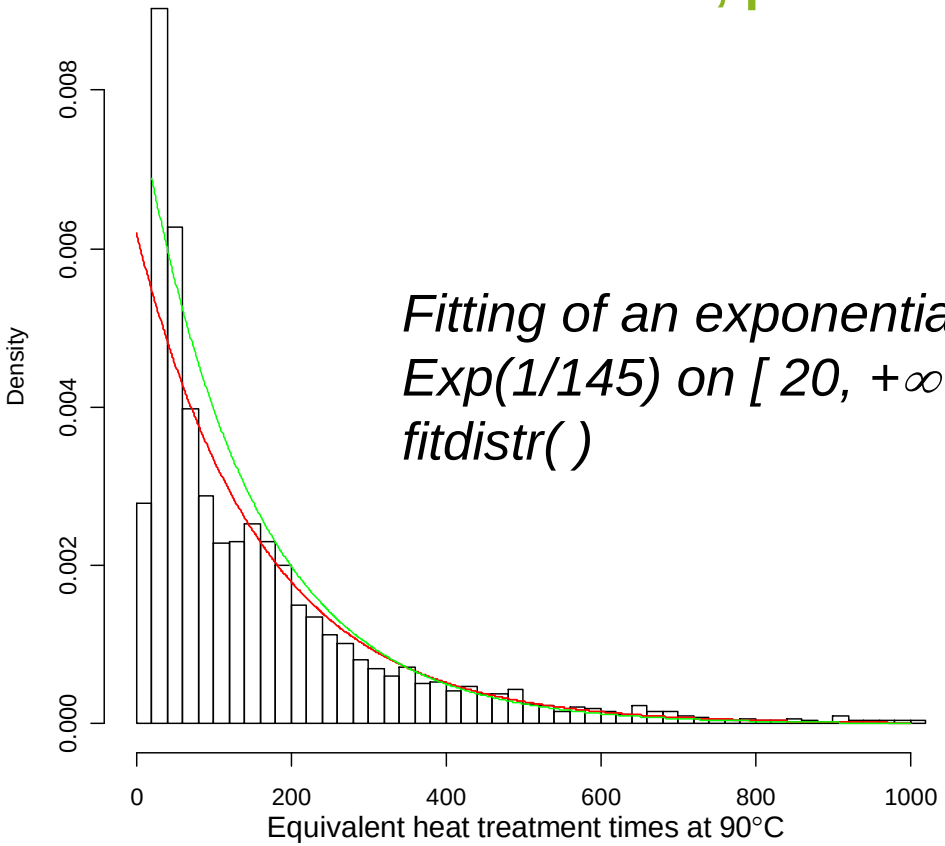
$$t_{eq} = \sum_{k=1}^{n} (t_k - t_{k-1}) 10^{1/z\left(\frac{T_k + T_{k-1}}{2} - 90\right)}$$

where

- $t_k - t_{k-1}$ (min) is the time interval between two temperature measurements,
- T_k (°C) is the k-th observed temperature in sample (T_0, …, T_n) constituted by n+1 temperatures T_0, …, T_n, where T_n is measured at the end of the thermal process.
- z (°C) is the increase in temperature resulting in a 10-fold reduction in the time to first decimal reduction

Which distributions for the model inputs?
Marginal approach in the case study

Bacillus cereus in a cooked, pasteurized and chilled food

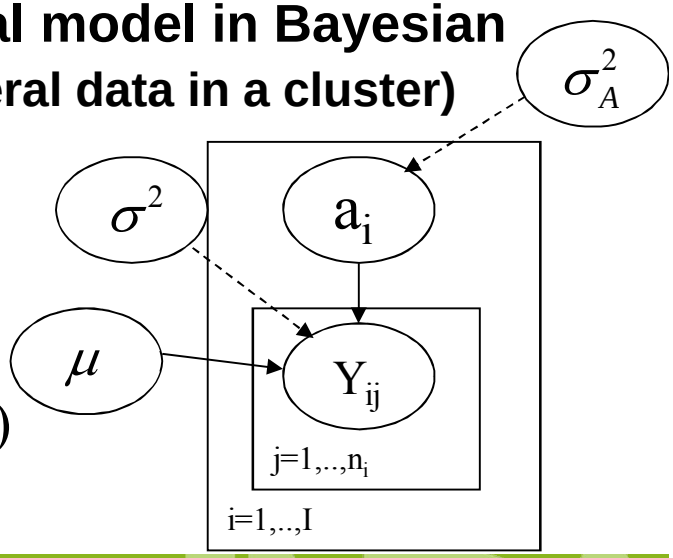


Fitting of an exponential distribution: Exp(1/145) on [20, +∞ [using the R function fitdistr( )

Modelling the variability between clusters into the dataset

> Variability between batches? strains? (clusters): a variance decomposition: variance between and within-clusters of data

> ANOVA with a random effect or hierarchical model in Bayesian statistics: for $i=1,..,I$ clusters and $j=1,$é n_i (several data in a cluster)



(Gaussian model)

> Then, overdispersion of the outputs

> To act accordingly for further sampling plans for example

$$Y_{ij} = \mu + a_i + \varepsilon_{ij}$$

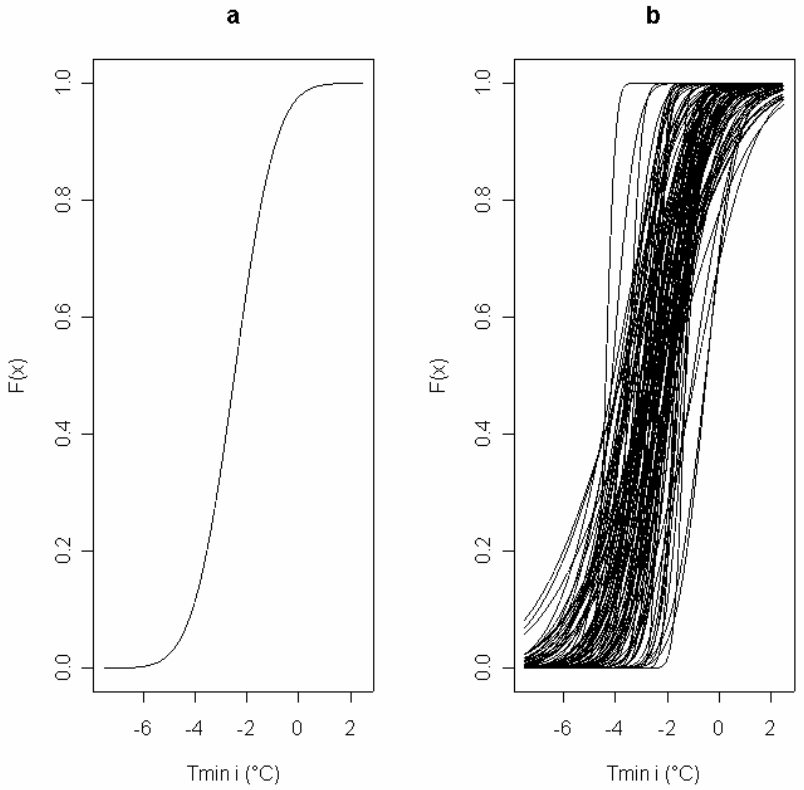
$$a_i \sim N(0, \sigma_A^2)$$

$$Y_{ij} \,|\, \mu, a_i \sim N(0, \sigma^2)$$

Which distributions for the model inputs?
Marginal approach

Second-order Monte Carlo simulation

> In order to take into account the uncertainty around the parameters of the fitted distribution. Example: Figure : (a) Modelling of $T_{min,i}$ parameter variability amongst strains expressed by the cumulative density function of N(-2.47, 1.26); (b) Modelling of $T_{min,i}$ parameter variability and uncertainty using a second order simulation: 100 M^*_{Tmin} values are randomly selected from a N(-2.47, 0.690) distribution and 100 S^*_{Tmin} values are randomly selected from a W(2.21, 1.43) distribution. The 100 respective cumulative density function $T_{min,i} \sim N(M^*_{Tmin}, S^*_{Tmin})$ are presented.



R. Pouillot, I. ALBERT, M. Cornu et J.-B. Denis. Estimation of uncertainty and variability in bacterial growth using Bayesian inference. Application to Listeria monocytogenes. International Journal of Food Microbiology, 81 : 87-104, 2003.

Which distributions for the model inputs?
Marginal approach

Fitting distributions to expert opinion

> Elicitation: acquisition of information from a person or group in a manner that does not disclose the intent of the interview

> Interview on tangible parameters for the expert

Example: Exponential dose-response model

$p(d) = 1 - \exp(-\alpha d)$, where $p(d)$ = Prob to die with d

α = ***shape parameter, not concrete interpretation !***

Questions on X_d: nb of dead mice when n mice received a same dose d

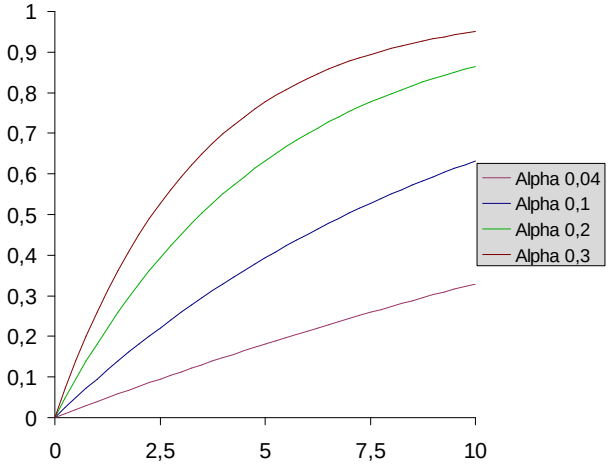
$X_d \mid p(d) \sim \text{Bin}(n, p(d))$, where $p(d)$ = Prob to die with d

Example of question in terms of proportion:

« An experiment concerns 10 mice injected with a dose d. If you have done 100 experiments, how many experiments lead to less than 2 dead mice? »

$P(X_d \leq 2) \Rightarrow P(p(d) \leq 0.2) = P(\log(\alpha) < g(0.2))$ where g is the inverse function: $\log(\alpha) \rightarrow p(d)$

P(d) exponential



Which distributions for the model inputs?
Marginal approach

Fitting distributions to expert opinion

> To try to capture the expertôs confidence in his/her response and introduce it in the modelling

> *Example*: considering elicited data as data. In the likelihood of these data, one incorporates larger variances for uncertain responses

> The knowledge of the expert is partial, so parametric distributions makes sense

> Expert feedback

> Ask for *p*-probabilities or *q*-quantiles of the quantity X, let him/her choose the level *p* or *q* if possible (because they depends on his/her knowledge)

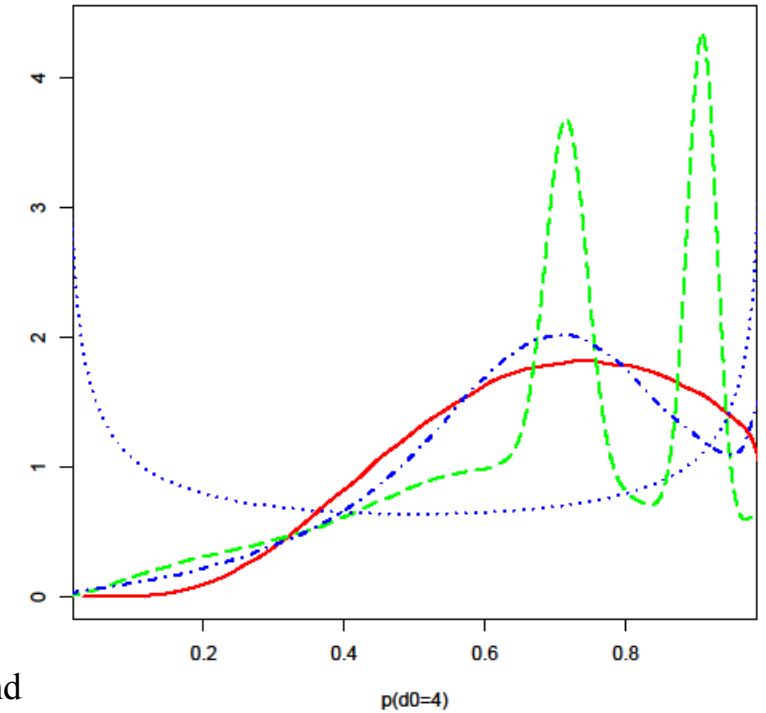
> Expert training: ex: dice throw

Which distributions for the model inputs?
Marginal approach

Fitting distributions to expert opinion

> Combining opinions from different experts

> The unique distribution of the parameter of interest comes from: pooling estimates, a mixture of the different expertsôdistributions or a hierarchical approach taking variation between and within expertsôgroups into account



p(d_0=4) by mixture (--), pooling estimates (solid line), hierarchical (-·) approaches. The density of beta(.5,.5) (···) is an example of non informative p(*d*)

I. Albert, S. Donnet, C. Guihenneuc, S. L. Choy, K. Mengersen and J. Rousseau. Combining expert opinions in prior elicitation, submitted, 2011.

Which distributions for the model inputs?

To sample directly into the data sets

> No fitted distributions

> If a large representative data set

> Sampling into the simulated values obtained after a Bayesian analysis from a MCMC algorithm

- **> The simulated values belong to the parameter posterior distribution**
- **> The simulated values belong to the predictive distribution of** $y \sim g(y \mid \theta, x)$

> Precision depends on the MC error

Which distributions for the model inputs?
Global approach

Bayesian evidence synthesis

> 1) Specifying the marginal distributions of the inputs from expert opinion or vague information

> 2) Running the model to check that all variables behave sensibly

> 3) Incorporating data within a Bayesian paradigm

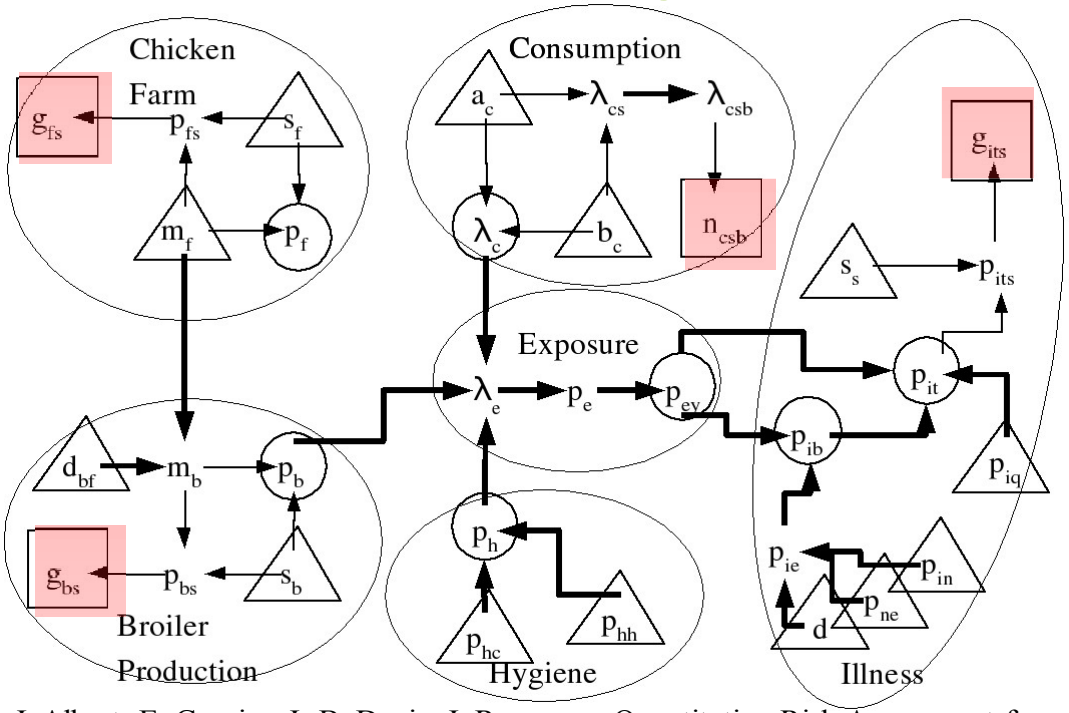
> Data are represented by new random va. in the Bayesian networks whose distributions depend on the core model va.

> complementary va. and new inputs (parameters requiring new prior distributions) could be necessary

> in the Bayesian setting, this substep corresponds to the definition of the likelihood

Which distributions for the model inputs?
Global approach in a case study

Assessing the prevalence of campylobacteriosis in France in a modelling from chicken farms to illness



$g_{fs} \sim \text{Bin}(n_{fs}, p_{fs})$
g_{fs}: number of chicken flocks contaminated out of n_{fs} in a sample (s=1,…,16)
$\text{logit}(p_{fs}) \sim N(m_f, s_f)$

$gbs \sim \text{Bin}(n_{bs}, p_{bs})$
g_{fs}: number of chicken carcasses contaminated out of n_{bs} in a sample (s=1,…14)
$\text{logit}(p_{bs}) \sim N(m_b, s_b)$
…..

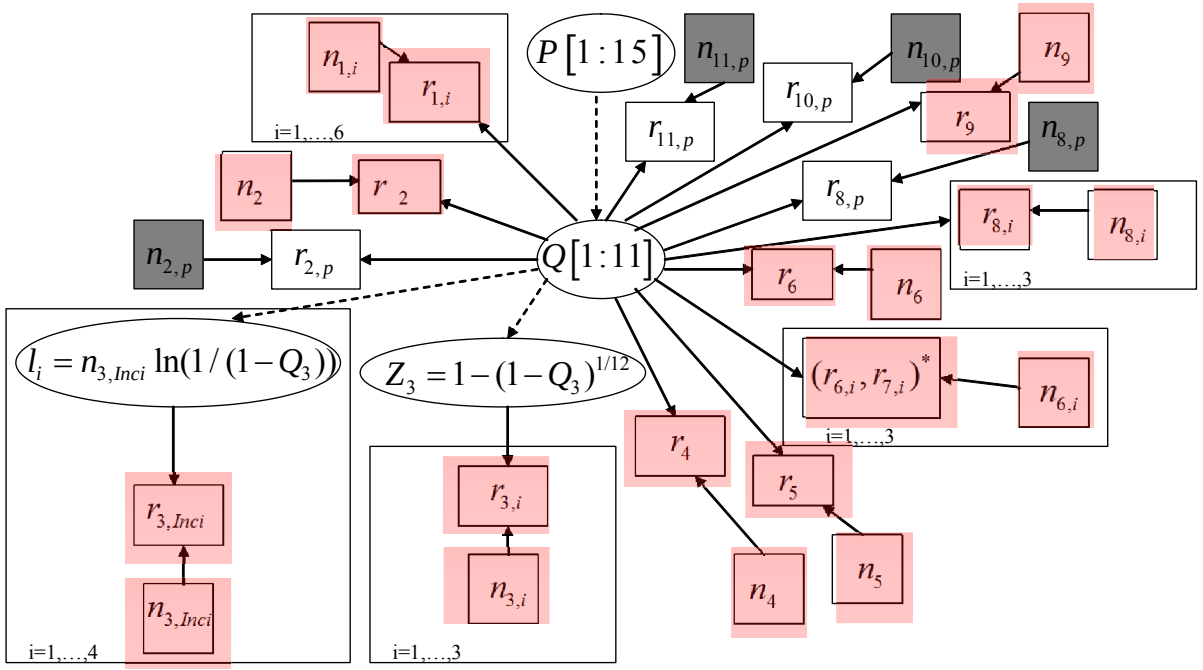
I. Albert, E. Grenier, J.-B. Denis, J. Rousseau. Quantitative Risk Assessment from Farm to Fork and Beyond: a global Bayesian Approach concerning food-borne diseases, *Risk Analysis*, 28(2) : 557-571, 2008.

Which distributions for the model inputs?

Global approach in a case study

Assessing the prevalence of campylobacteriosis in France from related human studies of the disease

(from active surveillance systems, laboratory surveys, physician surveys, epidemiological surveys,é)



$r_{1,i} \sim Bin(n_{1,i}, Q_1)$

$r_{1,i}$: number of stool cultures (SC) done out of $n_{1,i}$ people ($i=1$,**é** ,6)

Q_1: probability of having a SC

Q_2: probability of having a campylobacteriosis

Q_3: probability of having an acute gastroenteritis, and so on

…

A model relates all the Qs parameters

I. Albert, E. Espié, H. de Valk, J.-B. Denis. A Bayesian evidence synthesis for estimating campylobacteriosis prevalence, *Risk Analysis,* 31(7): 1141-1155, 2011.

Which distributions for the model inputs?
Global approach

Bayesian evidence synthesis

> Kind of meta-analysis encompassing diverse results from diverse sources that inform indirectly the parameters of interest through parameters that are functions of them

> In a Bayesian setting, each piece of evidence contributes to the likelihood through its likelihood function. The likelihood of the model is the product of the likelihood functions

> The posterior distribution of the parameters of interest is obtained via Bayesô theorem:

$$P(\theta \mid Data) \propto P\left(Data \middle| \theta\right) P\left(\theta\right)$$

Which distributions for the model inputs?
Global approach

Bayesian evidence synthesis

> Compared with marginal approach

> A better propagation of the uncertainty associated to the data sets

> The sample sizes define the precision of the estimates

> All sparse data are pooled in the likelihood and reduce the global uncertainty

> As in the marginal approach, a distribution is chosen. Here it is the likelihood function

Part III: How to validate the model?

How to validate the model?
Introduction

Check the model behaviour for a prediction goal

> Check that all variables behave sensibly

> Using plots, scenarios,é

> Validation step based on the global behaviour of the locally constructed system

> Using ñback-calculationò (Bayesian inference from data obtained ñdownstreamò in the food chain). Comparing posterior inference to substantive knowledge

> Predictive check

> Posterior predictive check and discrepancy measures

> Cross-validation

> One validation sample

How to validate the model?
Check that all variates behave sensibly

Using plots, scenarios,é

> Check the model behaviour in extreme contexts

> For example, verify if a very large contamination in input (out of the real variability) induces a contamination downstream in accordance with what it is known about the microbial dynamics

> Plot the marginal and bivariate distributions of variables or pairs of variables of interest known by the experts

> Produce some conditional distributions corresponding to specific cases well known by the experts

How to validate the model?
Check that all variates behave sensibly

Using Bayesian inference in a subjective context

> Are the posterior distributions of the variables in accordance with what it is known?

> Expert feedback

> if agreement because reduced uncertainty or new expert opinion

> if no agreement, questioning of the model construction (links between the variables, prior distributions)

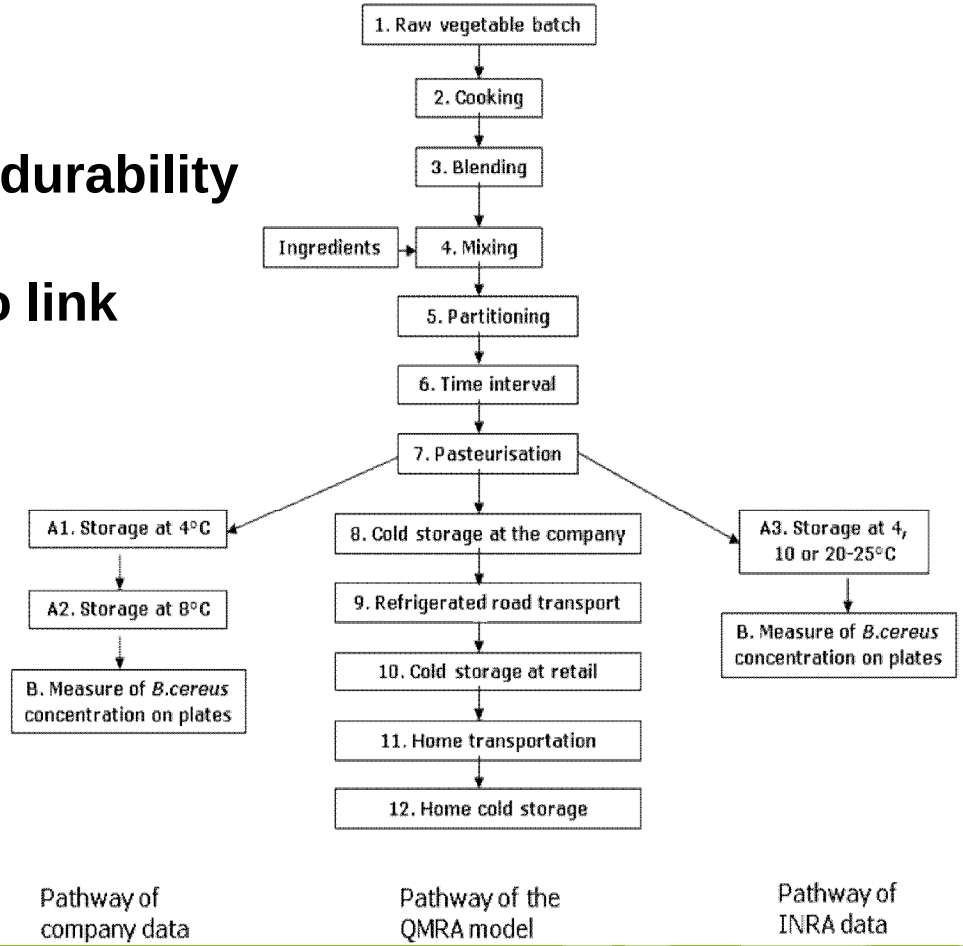
> Posterior distributions in accordance with data

> But they also depend on the modelling

How to validate the model?
Model validation in the case study

***Bacillus cereus* in a cooked, pasteurized and chilled food: using data from durability studies to improve the model**

> To augment the model to link the data sets



How to validate the model?
Model validation in the case study

***Bacillus cereus* in a cooked, pasteurized and chilled food: using data from durability studies**

> 68% of left-censored data but some quantifications and specifications of strain types

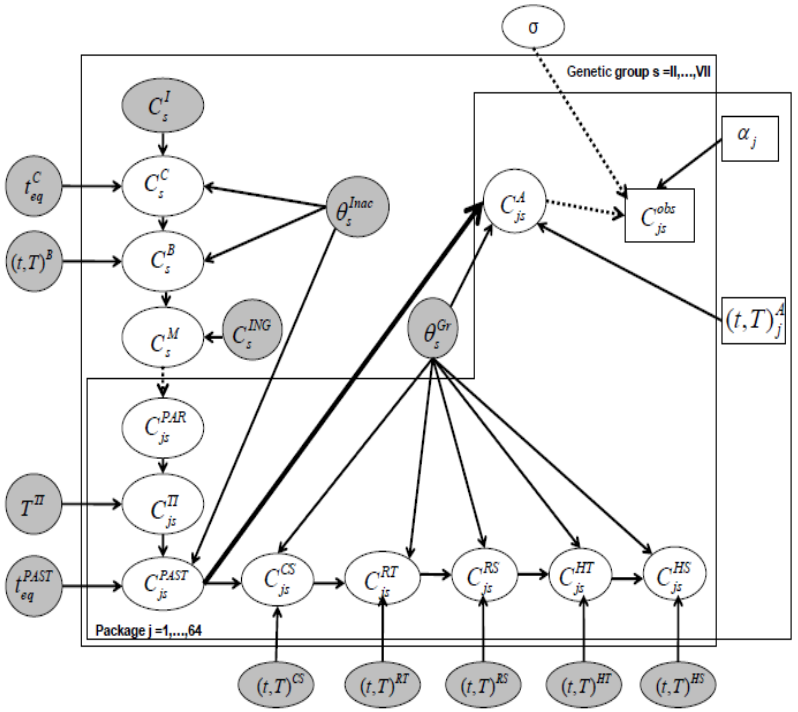
Data (source)	Product	Conditions	Contamination in log CFU/g (detected genetic groups*)
15 packages coming from the same batch (INRA)	5 packages of 800g	21 days at 4°C	< 1,7
	5 packages of 800g	21 days at 10°C	5,5 (VI) 3 (VI) 7,5 (VI) 3,8 (II) 3,1 (II)
	5 packages of 800g	21 days at 20-25°C	6 (IV) 7,2 (II and IV) 6,4 (IV) 6,3 (IV) 6,3 (IV)
55 packages coming from different batchs (COMPANY)	28 packages of 400g	20 days at 4°C then 10 days at 8°C	3,7 (II) 2,7 (II) 3,7 (II) 3,9 (VI) 5 (II) < 2 for the 23 other packages
	27 packages of 400g	10 days at 4°C then 20 days at 8°C	5,2 (VI) 4,7 (VI) 5,4 (II) 4,4 (VI) 5,4 (VI) 4,7 (VI) 3,9 (II) < 2 for the 20 other packages

How to validate the model?
Model validation in the case study

***Bacillus cereus* in a cooked, pasteurized and chilled food: using data from durability studies**

> *Augmented* Bayesian network:

A new intermediate variable (C^A, concentration after storage in durability studies) and a new parameter (σ, experimental error)



Likelihood:

$$\log_{10}(C_{js}^{obs}) \sim \begin{cases} N(\log_{10}(C_{js}^{A}),\sigma^2) \quad if\ \log_{10}(C_{ij}^{obs}) > \alpha_j \\ N(\log_{10}(C_{js}^{A}),\sigma^2)C(;\alpha_j)\ if \log_{10}(C_{ij}^{obs}) \le \alpha_j \end{cases}$$

j: package, *s*: genetic group, α: limit of detection (inference with vague prior on σ)

How to validate the model?
Model validation in the case study

Bacillus cereus in a cooked, pasteurized and chilled food:

Update of the B. cereus concentrations through the food pathway

Increase for groups II and VI in accordance to data

Genetic Group	Initial	After cooking	Milk (ingredient)	Starch (ingredient)	After mixing with ingredients and partitioning	After pasteurization	After home cold storage
MC prevalence (%)							
II	100	60	100	100	28	13	13
III	100	48	100	100	80	40	40
IV	100	45	100	100	100	84	84
V	100	24	100	100	5	0.1	0.1
VI	100	45	100	100	6	0.03	0.03
VII	100	64	100	100	10	1	1
Total	100	83	100	100	100	88	88
MCMC prevalence (%)							
II	100	100	100	100	80	59	59
III	100	87	100	100	65	1	1
IV	100	86	100	100	100	72	72
V	100	82	100	100	6	0.00	0.00
VI	100	93	100	100	50	15	15
VII	100	99	100	100	19	0.05	0.05
Total	100	100	100	100	100	90	90

How to validate the model?
Model validation in the case study

Bacillus cereus in a cooked, pasteurized and chilled food:

Update of the B. cereus concentrations through the food pathway

Lower prediction

Genetic Group	Initial	After Cooking	Milk (ingredient)	Starch (ingredient)	After mixing with ingredients and partitioning	After pasteurization	After home cold storage
			MC 95% CI				
II	[-1.6; 0.1]	[-5.3; -1]	[-5.1; -2.6]	[-3.8; -1.7]	[-2.9; -1]	[-2.7; -1]	[-2.5; 4.7]
III	[-4; -2.4]	[-5.4; -2.9]	[-1.6; 0.9]	[-3.8; -1.8]	[-2.9; -0.9]	[-2.7; -1]	[-2.7; 0.2]
IV	[-4; -2.4]	[-5.4; -3.2]	[-1.6; 0.8]	[-0.8; 1.2]	[-2.2; -0.3]	[-2.6; -0.5]	[-2.6; 0.1]
V	[-4; -2.4]	[-5.4; -3.7]	[-5.1; -2.6]	[-3.8; -1.7]	[-2.9; -2.3]	[-2.9; -2.3]	[-2.8; 1.4]
VI	[-1.6; 0]	[-5.4; -2.3]	[-5.1; -2.6]	[-3.8; -1.7]	[-2.9; -2.3]	[-2.9; -2.4]	[-2.8; 8.3]
VII	[-4; -2.4]	[-5.4; -2.9]	[-5.1; -2.6]	[-3.8; -1.8]	[-2.9; -2.3]	[-2.7; -2.1]	[-2.7; -2.1]
Total	[-1.1; 0.2]	[-5.3; -1.1]	[-0.9; 1]	[-0.8; 1.2]	[-1.9; -0.3]	[-2.6; -0.5]	[-2.5; 1.8]
			MCMC 95% CI				
II	[-1.7; 0.1]	[-2.8; -0.7]	[-5.1; -2.6]	[-3.8; -1.7]	[-2.9; -0.7]	[-2.6; -0.9]	[-2.5; 1.1]
III	[-4.1; -2.4]	[-5.1; -2.7]	[-1.7; -0.1]	[-3.8; -1.8]	[-2.9; -1.8]	[-2.9; -2.1]	[-2.9; -1.7]
IV	[-4; -2.3]	[-5.3; -2.9]	[-1.6; 0.7]	[-0.8; 0.8]	[-2; -0.7]	[-2.8; -1.5]	[-2.8; -1.3]
V	[-4.1; -2.4]	[-5.4; -3.5]	[-5.1; -2.6]	[-3.8; -1.8]	[-2.9; -2.3]	[-2.7; -2.3]	[-2.6;-1.9]
VI	[-1.3; 0.2]	[-4.8; -1.5]	[-5.1; -2.6]	[-3.7; -1.4]	[-2.9; -2.1]	[-2.9; -2.3]	[-2.7; 3.7]
VII	[-4.1; -2.4]	[-4.7; -2.7]	[-5.1; -2.6]	[-3.8; -1.7]	[-2.9; -2.3]	[-2.6; -2.3]	[-2.6; -2.3]
Total	[-0.8; 0.3]	[-2.6; -0.6]	[-1; 0.7]	[-0.7; 0.8]	[-1.8; -0.5]	[-2.6; -0.9]	[-2.6; 1.4]

How to validate the model?
Model validation in the case study

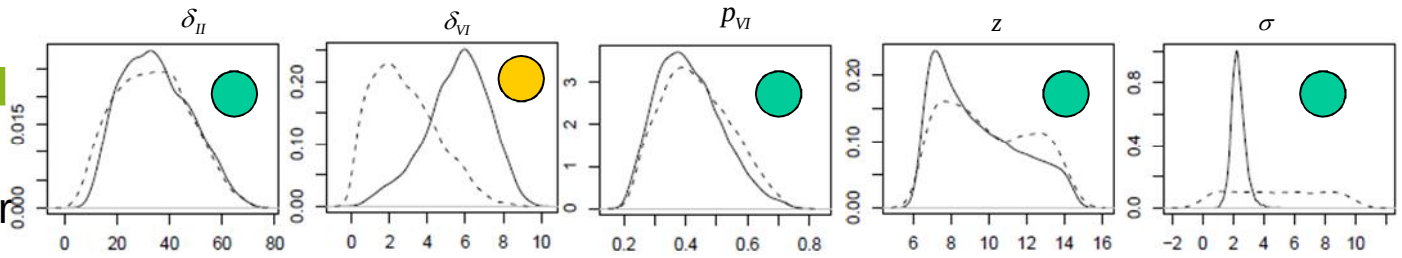
Bacillus cereus in a cooked, pasteurized and chilled food:

***Prior beliefs update*:** Prior (dashed) vs posterior (solid) distributions of the **most updated** inputs over 66

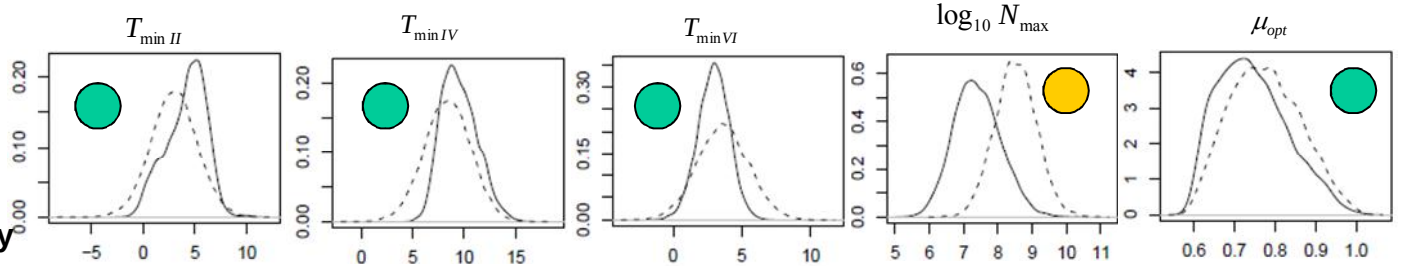
Expert Feedback:

- : **reduced uncertainty**
- : **expert agreement, prior belief change**
- : **question the modeling**

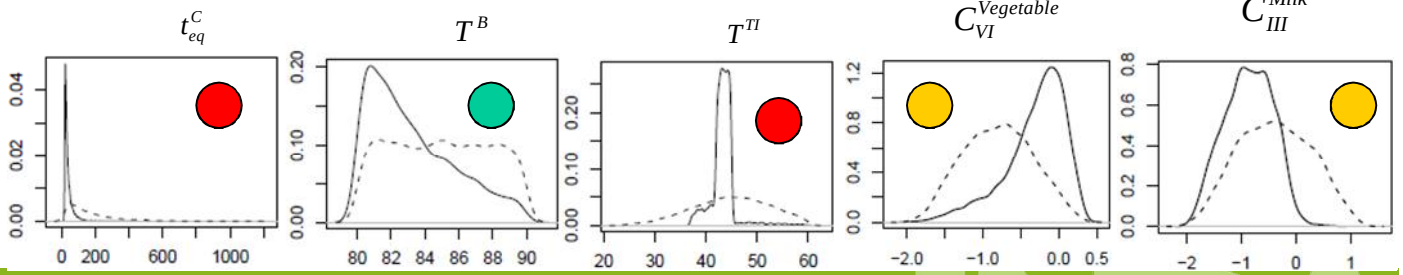
(a) Inactivation biological parameters and σ



(b) Growth biological parameters



(c) Inactivation process parameters and two initial concentrations



How to validate the model?
Model validation in the case study

Bacillus cereus in a cooked, pasteurized and chilled food:

Bayesian inference

> to link the data and the QMRA model if it exists yet

> Some difficulties in complex models

- Implementation of the MCMC algorithms

- Look at the algorithm convergence because very large auto-correlation in a chain of simulated values (strong correlation between consecutive values of a parameter because high correlation between the parameters through the used models)

- so very low convergence

In this case study, *jags* software used and 2 Markov chains of 1 million iterations in burn-in and 1 million for posterior distributions per chain (thin of 1000 iterations for the posterior distributions keeping 1000 iterations per chain)

> But some updates can be obtained with not many data because very subjective priors (shifts)

How to validate the model?
Predictive check

Posterior Predictive Check

> Comparing measurements (data) to the posterior predictive distribution

$$p(y^{rep} \mid y) = \int p(y^{rep} \mid \theta)\, p(\theta \mid y) d\theta$$

Sampling distribution ← → **Parameter Posterior distribution**

y^{rep}**: replicated data that could have been observed with the same model**

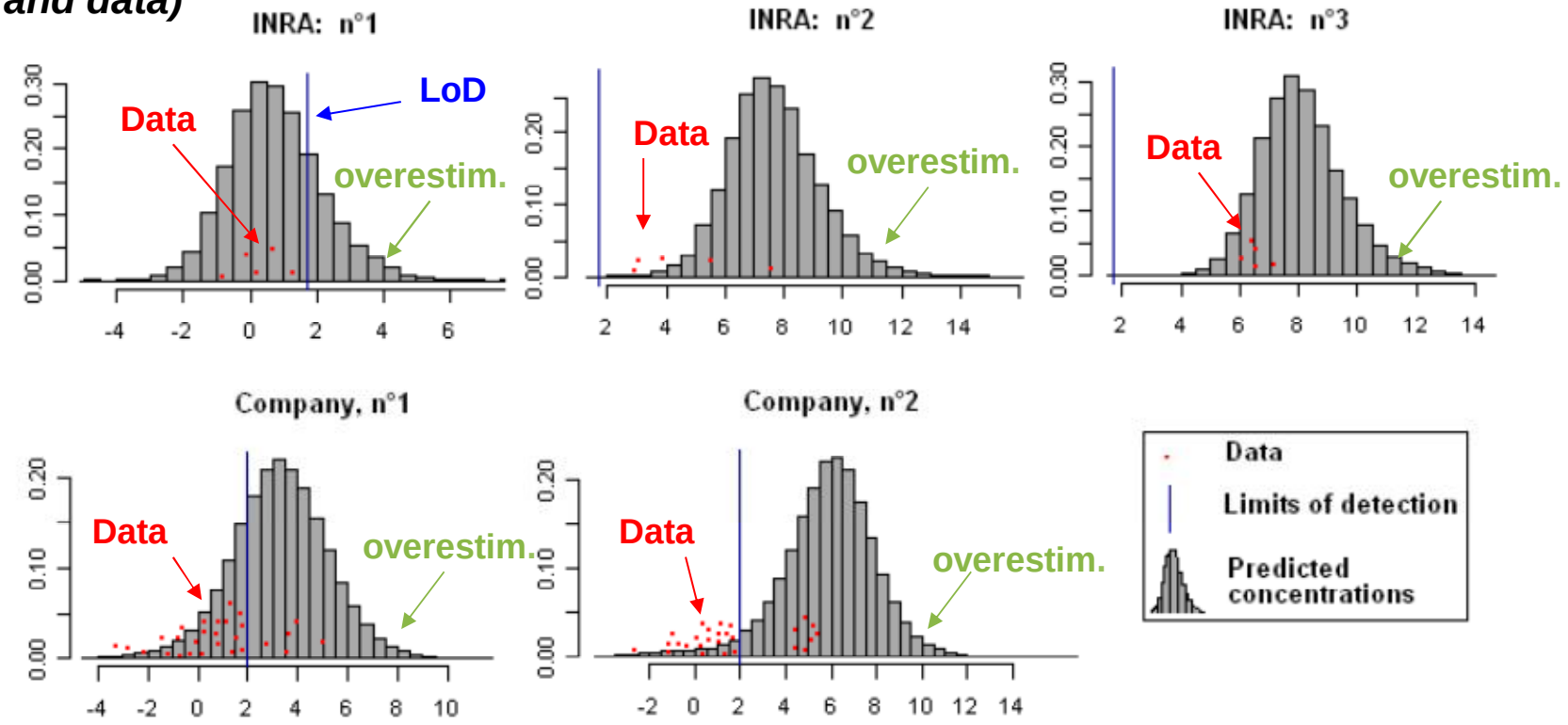
> Discrepancy measures

> Choose a measure that will be used to compare data to predictive simulations

> ex: largest value, smallest value, average, standard deviation, omnibus goodness-of-fit measure (χ^2)

How to validate the model?
Model validation in the case study

Bacillus cereus in a cooked, pasteurized and chilled food: *Posterior predictive check (mixture of prior knowledge and data)*



How to validate the model?
Predictive check

Cross-validation

> Rotation estimation and validation

> Randomly splitting the data set into K subsamples. Of the K subsamples, a single subsample is retained as the validation data set for testing the model, and the remaining K ī 1 subsamples are used for the parameter estimation

> Same goodness-of-fit criteria than before

> Then, combine or average the K results

> Leave-one-out validation

> Computational intensive

How to validate the model?
Predictive check

One validation sample

> Another data set for validation

> But less data for estimation

> A good way to act?:

> To compare the data obtained ñdownstreamò in the food chain to prior predictive results (simulated values obtained through Monte Carlo simulations) using goodness-of-fit criteria (depending to objectives) and then if accordance to proceed Bayesian inference to update the prior knowledge and to look at the posterior distributions (coherence or not)

Conclusion

Conclusion

Do not forget your objectives

> Well-define the objectives: simplification is possible driven by objectives

> Modelling at the level where you have the most knowledge

> To gather a maximum of data, new experiments if possible

> Do not forget that your results depend on your modelling, your prior opinion if you are Bayesian

> Bayesian networks and DAG: helpful tools to construct a complex modelling

> Bayesian inference to update the prior knowledge and to validate

> All QMRA is incomplete, uncertain, just one help for risk management

Conclusion

Perspectives

> A better modelling taking account of an acute objective

> ex: loss function, utility function, cost function

> Better elicitation of the expert opinions

> In so complex models, which are the parameters that have to be better informed a priori?

> A simultaneous modelling for a risk and a benefice

> ABC (Approximate Bayesian Computation) techniques to perform Bayesian inference

> To use a QMRA farm-to-fork modelling to reach a Food Safety Objective (FSO)

References

References of the presented examples

> A.L. Afchain, F.Carlin, C. Nguyen-the, I. Albert. Improving quantitative exposure assessment by considering genetic diversity of B. cereus in cooked, pasteurised and chilled foods, *International Journal of Food Microbiology*, 128 : 165-173, 2008.

> I. Albert, R. Pouillot et J.-B. Denis. Stochastically modeling Listeria monocytogenes growth in farm tank milk, Risk Analysis, 25 : 1171-1185, 2005.

> I. Albert and P. Mafart. A modified Weilbull model for Bacterial inactivation. International Journal of Food Microbiology, 100 : 197-211, 2005.

> I. Albert, E. Grenier, J.-B. Denis, J. Rousseau. Quantitative Risk Assessment from Farm to Fork and Beyond: a global Bayesian Approach concerning food-borne diseases, Risk Analysis, 28(2) : 557-571,2008.

> R. Pouillot, I. Albert, M. Cornu et J.-B. Denis. Estimation of uncertainty and variability in bacterial growth using Bayesian inference. Application to Listeria monocytogenes. International Journal of Food Microbiology, 81 : 87-104, 2003.

> I. Albert, S. Donnet, C. Guihenneuc, S. L. Choy, K. Mengersen and J. Rousseau. Combining expert opinions in prior elicitation, submitted, 2011.

References of the presented examples

> I. Albert, E. Espié, H. de Valk, J.-B. Denis. A Bayesian evidence synthesis for estimating campylobacteriosis prevalence, Risk Analysis, 31(7): 1141-1155, 2011.

> C. Rigaux, S. Ancelet, F. Carlin, C. Nguyen-thé, I. Albert. Using an augmented Bayesian network to improve a complex quantitative microbial risk assessment model from durability studies: Application to Bacillus cereus in courgette purée, submitted, 2011.

Thank you very much for your attention

Kiitos!