General Unified Threshold model of Survival (GUTS) under time variable chemical exposure:

Comparison of Bayesian implementation with integration of ODEs in R using JAGS or Stan.

Virgile BAUDROT & Sandrine CHARLES

Journées AppliBUGS – 14 december 2017

cnrs dépasser les frontières
LBBE BIOMETRIE ET BIOLOGIE EVOLUTIVE
UCB Lyon 1

Environmental Risk Assessment

OECD

efsa
European Food Safety Authority

Environmental Risk Assessment

OECD

efsa
European Food Safety Authority

5. Review risk assessment

1. Hazard identification

2. **Hazard characterization**

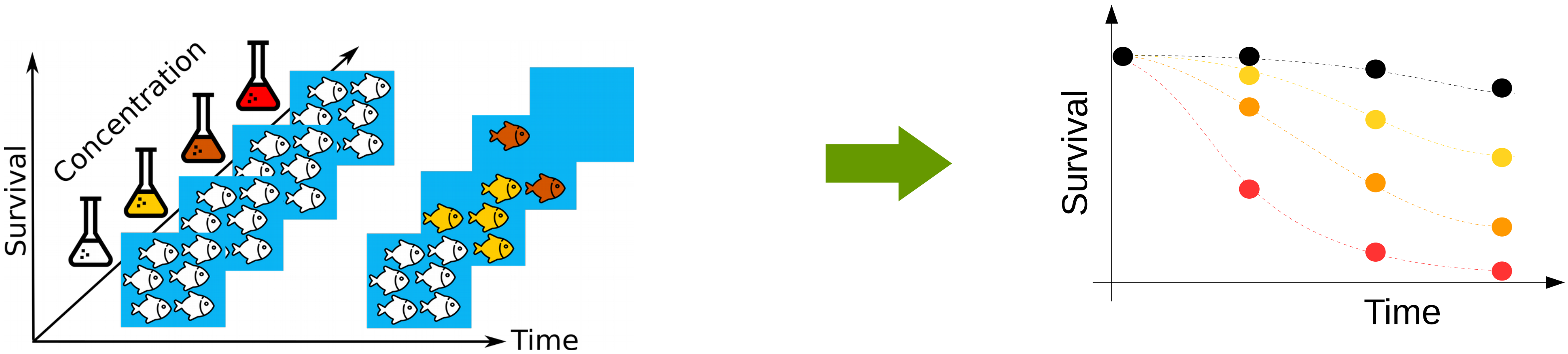
Dose-response modelling

3. Exposure assessment

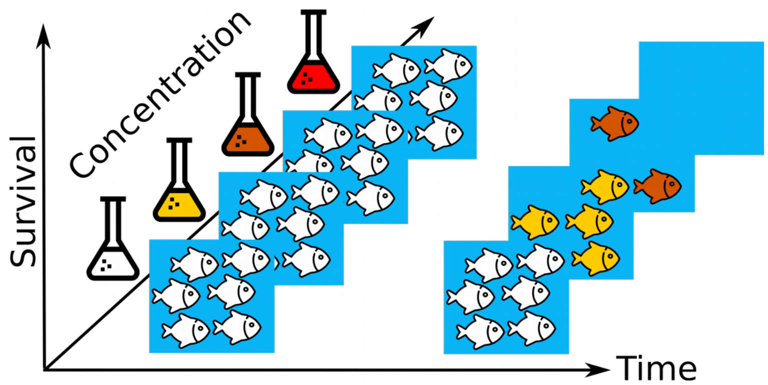
4. **Risk characterisation**

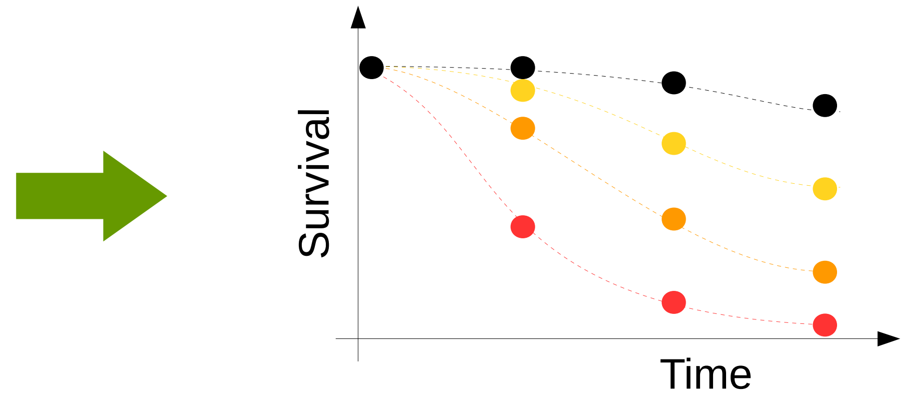
Prediction

Experiment

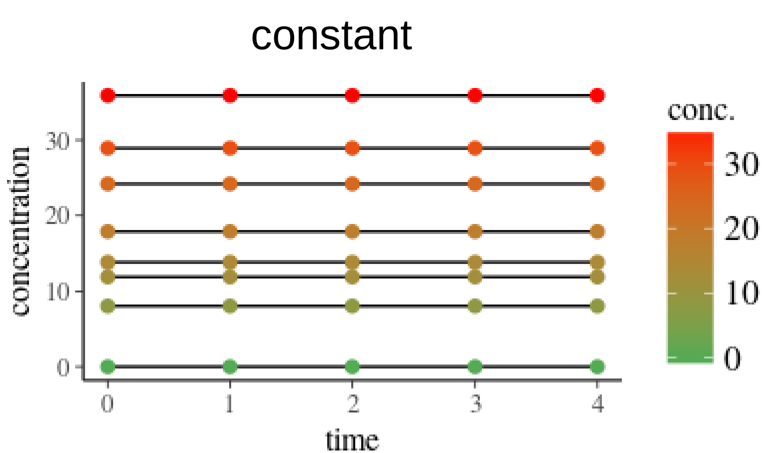


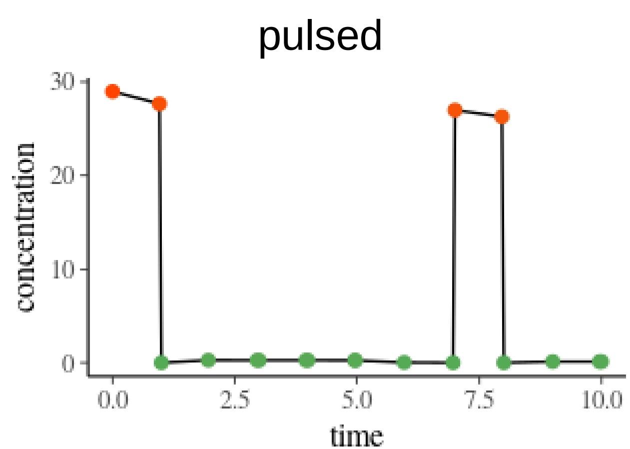
Experiment

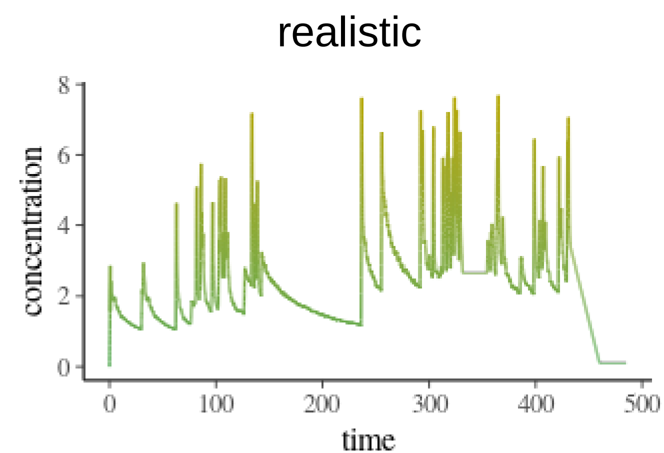




Exposure profiles:

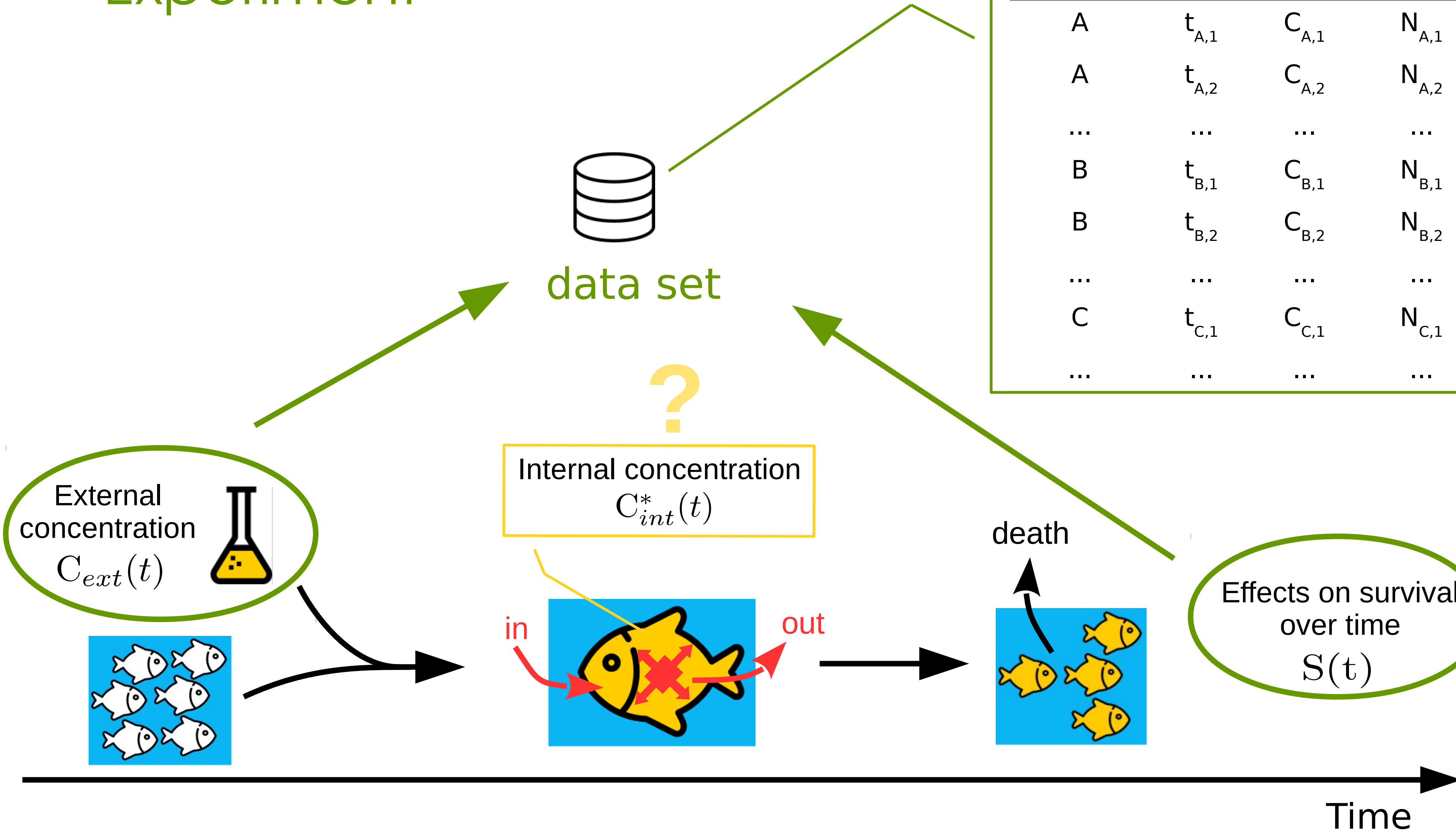






Experiment

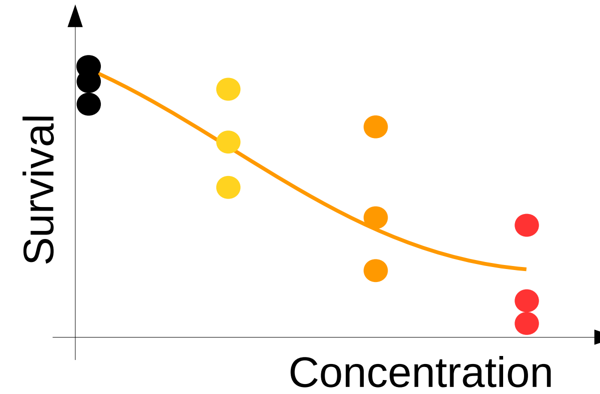
replicate	time	conc	survival
A	$t_{A,1}$	$C_{A,1}$	$N_{A,1}$
A	$t_{A,2}$	$C_{A,2}$	$N_{A,2}$
...	...	...	...
B	$t_{B,1}$	$C_{B,1}$	$N_{B,1}$
B	$t_{B,2}$	$C_{B,2}$	$N_{B,2}$
...	...	...	...
C	$t_{C,1}$	$C_{C,1}$	$N_{C,1}$
...	...	...	...

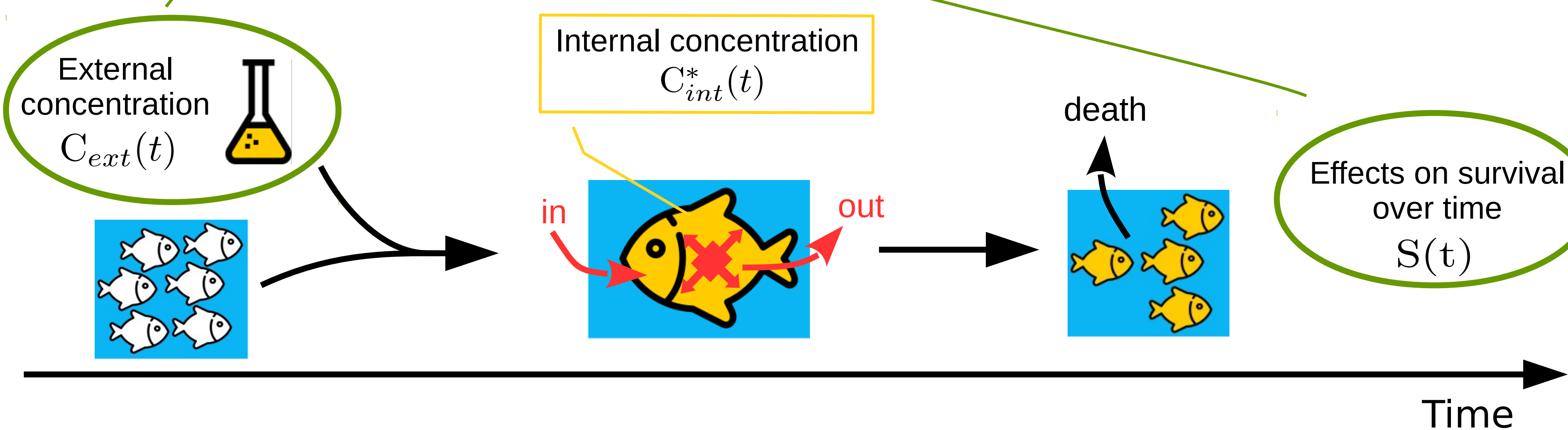


Model: Target Time (TT)

$$S = \frac{d}{1 + (C_{ext}/LC_{50})^b}$$

- Only constant exposure profile
- At given Target Time

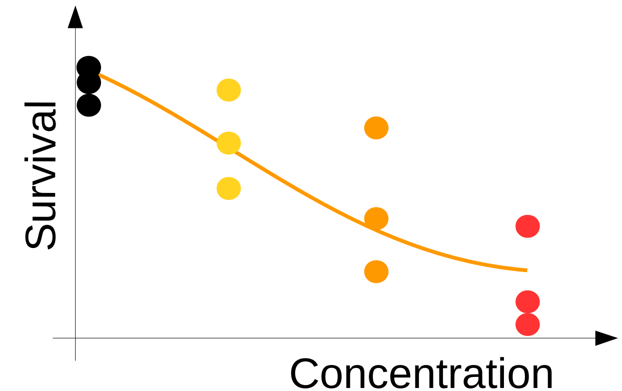




Model: Target Time (TT)

$$S = \frac{d}{1 + (C_{ext}/LC_{50})^b}$$

- Only constant exposure profile
- At given Target Time



Least-squares estimates

function `drm` from `drc` package (Ritz et al., 2015)

Bayesian inference

+ JAGS
Just Another Gibbs Sampler

function `survFitTT` from `morse` package

(version 2.2.0 - Delignette-Muller *et al.*, 2016 ;
(version 3.0.0 - Baudrot *et al.*, 2017)

web interface
`http://pbil.univ-lyon1.fr/software/mosaic/`

MOSAIC
modeling and statistical tools for ecotoxicology

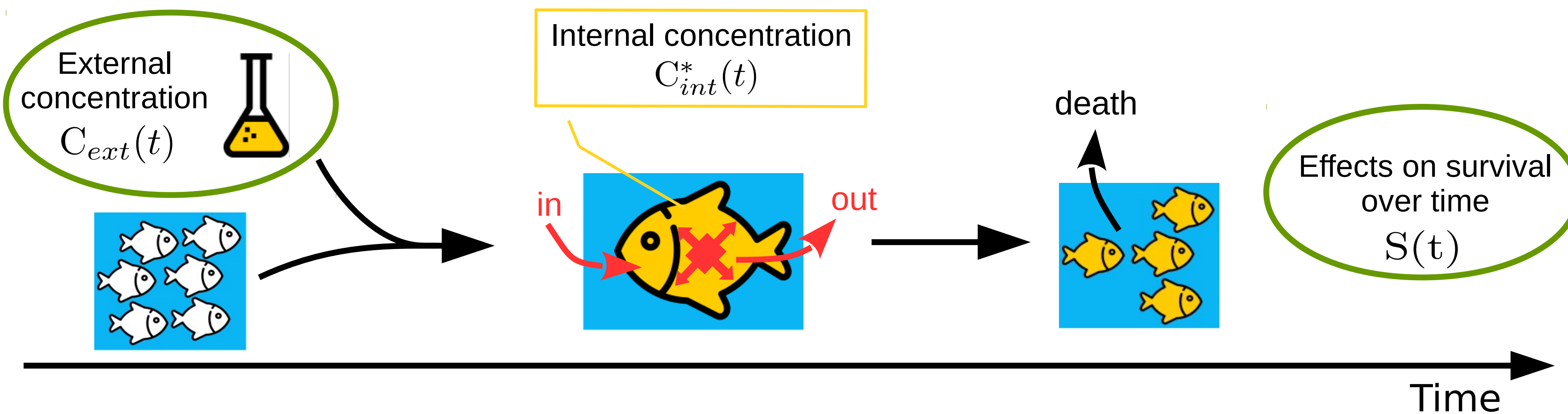
Model: TKTD

- Bioaccumulation
- Diffusion
- Biotransformation
- Elimination

Toxicokinetics (TK) →

- Toxic mechanism
- Biochemical effects
- Physiological effects
- Mortal., Repro., Behav.

Toxicodynamics (TD) →



Model: TKTD

Theoretical & Applied ecotoxicology

- **mechanistic** rather than descrive (e.g., Grimm et *al.*, 2009 ; Forbes et *al.*, 2009)
- **temporal aspects** of **exposure** & **toxicity** (non-constant profiles) (e.g., Ashauer et *al.*, 2007)
- consider **all data-point** of the experiment (no time-specific)
- derive time-independent **toxicity parameters** (e.g., Jager et *al.*, 2011)
- make **predictions** for untested situations (e.g., Ashauer et *al.*, 2007)
- effects of **mixture** over-time (Jager et *al.*, 2010)

Policies

- TKTD model are **promoted by OECD** (Series on testing and assessment N°54, 2006)
- part of **ERA of PPPs** (Plant Protection Products) (EFSA PPR Panel, 2013)
- can be used for **REACH dossier**

Model: TKTD

Time

Toxicokinetics (TK)

Toxicodynamics (TD)

$$\begin{cases} \dfrac{dC^*_{int}(t)}{dt} = k_d(C_{ext}(t) - C^*_{int}(t)) \\ C^*_{int}(t=0) = 0 \end{cases}$$

$$\Rightarrow \quad C^*_{int}(t) = k_d e^{-k_d t} \int_0^t e^{k_d \tau} C_{ext}(\tau) d\tau$$

Model: TKTD

Time

Toxicokinetics (TK)

$$\begin{cases} \frac{dC^*_{int}(t)}{dt} = k_d(C_{ext}(t) - C^*_{int}(t)) \\ C^*_{int}(t=0) = 0 \end{cases}$$

$$\Rightarrow \quad C^*_{int}(t) = k_d e^{-k_d t} \int_0^t e^{k_d \tau} C_{ext}(\tau) d\tau$$

Toxicodynamics (TD)

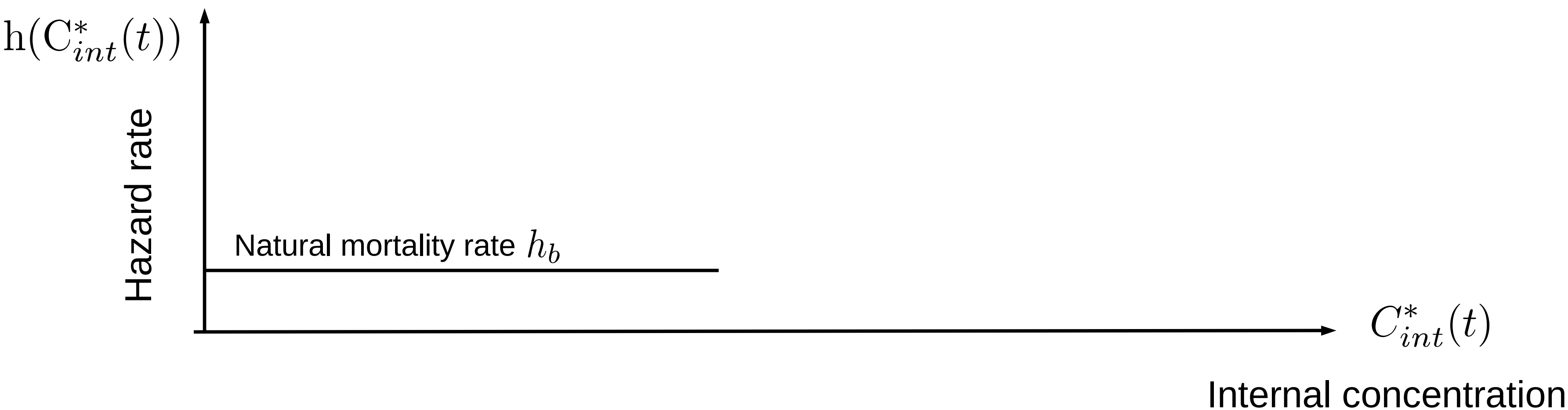
General Unified Threshold model of Survival (GUTS)

(Jager et *al.*, 2011 ; Ashauer et al., 2013)

Two complementary death mechanisms:

1. **Stochastic Death (SD)**
2. **Individual Tolerance (IT)**

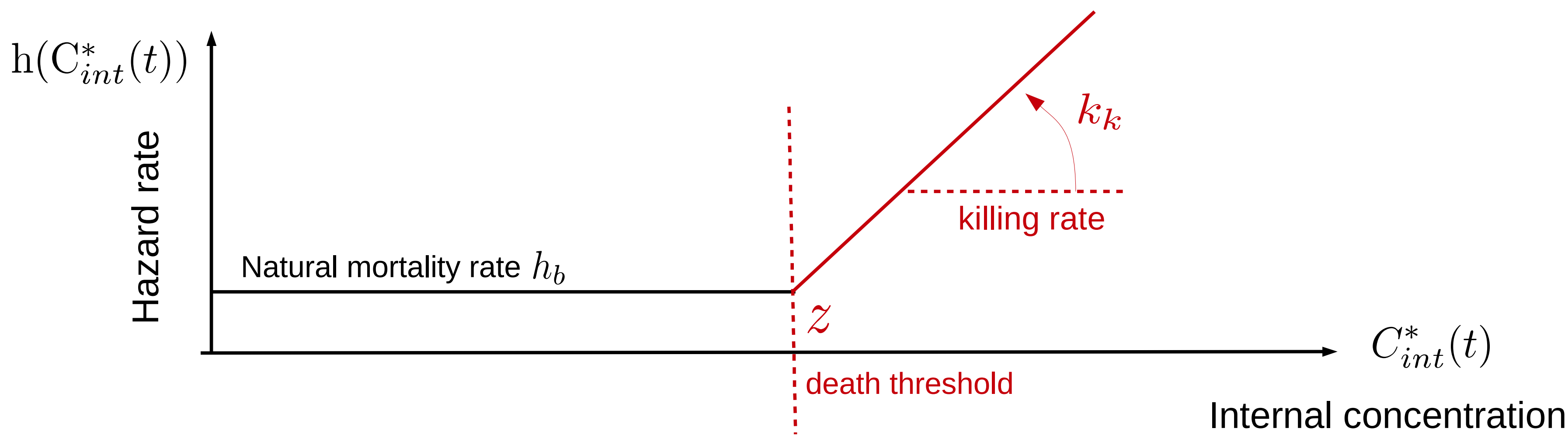
Model: TKTD



(Jager *et al.*, 2011)

Model: TKTD

Stochastic Death (SD)

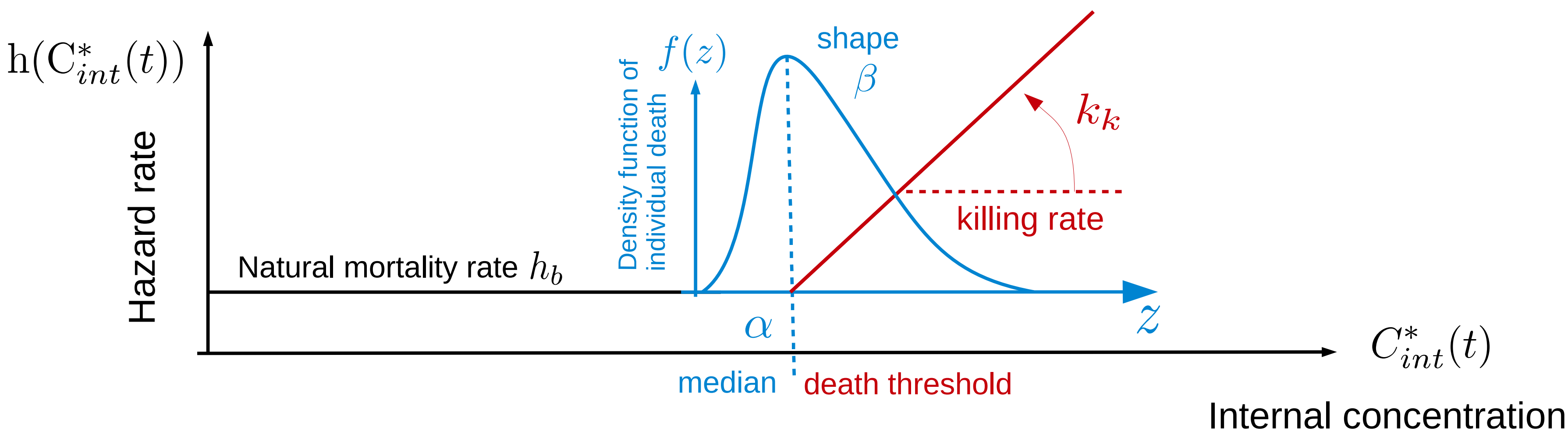


(Jager *et al.*, 2011)

Model: TKTD

Stochastic Death (SD) Individual Tolerance (IT)

'proper'



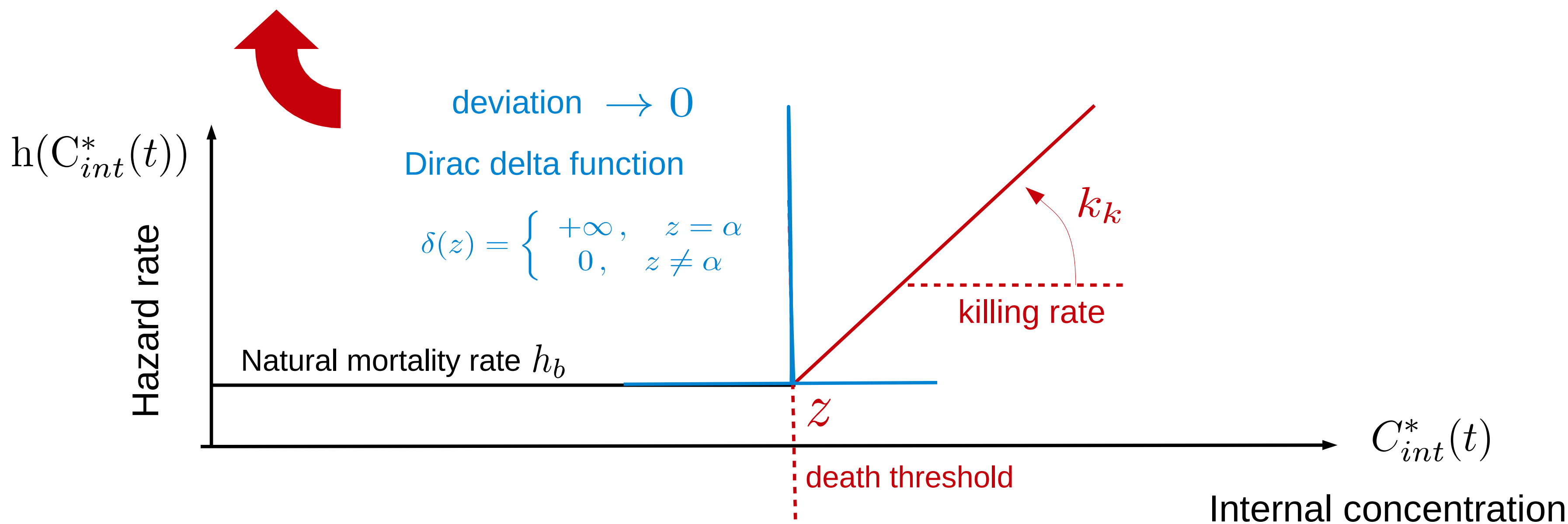
(Jager *et al.*, 2011)

Model: TKTD

Stochastic Death (SD)

$$S_{SD}(t) = \exp\left(-\int_0^t \overbrace{k_k \max(C^*_{int}(\tau) - z, 0) + h_b}^{\mathrm{h}(\mathrm{C}^*_{int}(\tau))} \, d\tau\right)$$

- All individual are identical
- $\mathrm{C}^*_{int}(t)$ increases death probability



(Jager *et al.*, 2011)

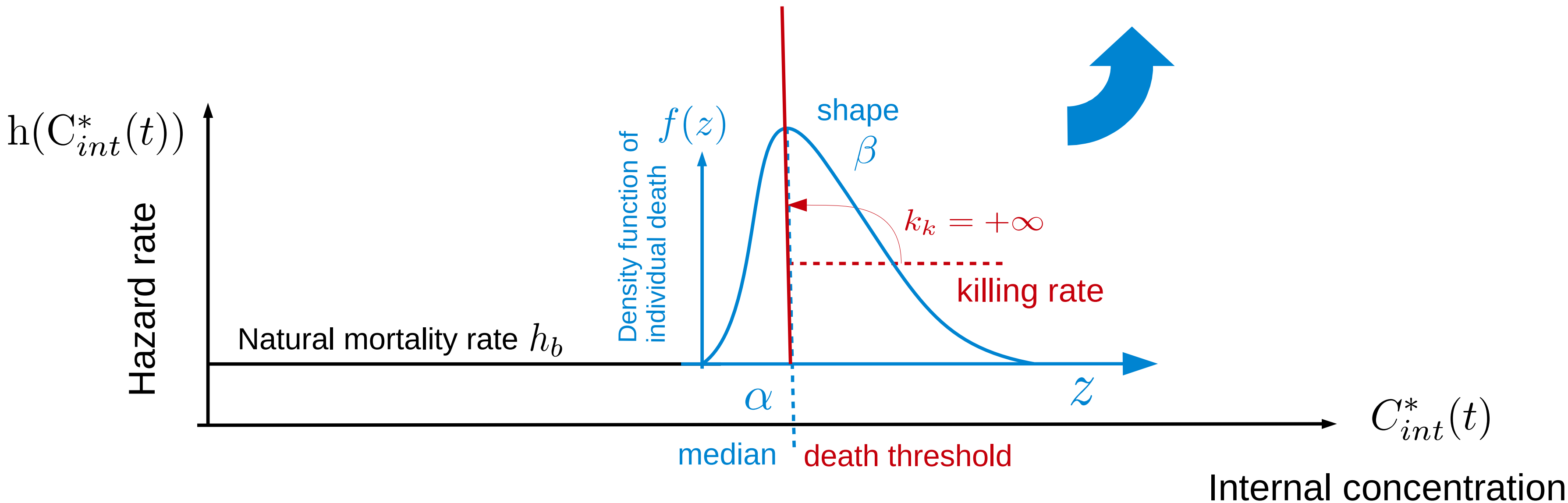
Model: TKTD

- Death is immediate if $C^*_{int} > z$
- Individuals differ in z

Individual Tolerance (IT)

$$S_{IT}(t) = e^{-h_b t} \int_{\max_{0 \le \tau \le t} C^*_{int}(\tau)}^{+\infty} f(z)dz$$

$$= e^{-h_b t} \left(1 - F\left(\max_{0<\tau<t} C^*_{int}(\tau)\right)\right)$$



(Jager *et al.*, 2011)

Model: TKTD

Toxicokinetics (TK)

$$\begin{cases} \dfrac{dC^*_{int}(t)}{dt} = k_d(C_{ext}(t) - C^*_{int}(t)) \\ C^*_{int}(t=0) = 0 \end{cases}$$

Toxicodynamics (TD) - GUTS

Stochastic Death (SD)

$$S_{SD}(t) = \exp\left(-\int_0^t \overbrace{k_k \max(C^*_{int}(\tau) - z, 0) + h_b}^{\mathrm{h}(\mathrm{C}^*_{int}(\tau))}\, d\tau\right)$$

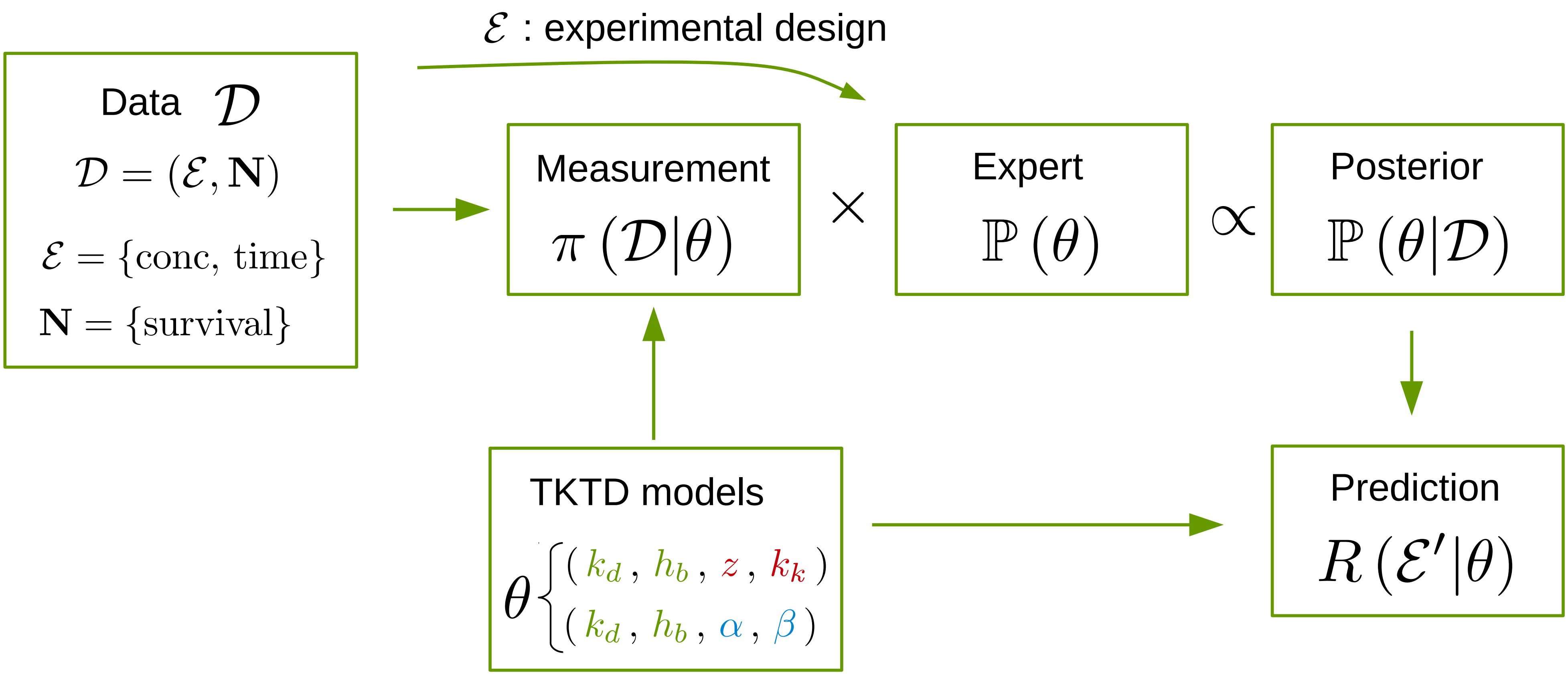
Individual Tolerance (IT)

$$S_{IT}(t) = e^{-h_b t} \int_{\max\limits_{0\leq\tau\leq t} C^*_{int}(\tau)}^{+\infty} f(z)dz$$

$$= e^{-h_b t}\left(1 - F\left(\max_{0<\tau<t} C^*_{int}(\tau)\right)\right)$$

θ : (k_d, h_b, z, k_k) $(k_d, h_b, \alpha, \beta)$

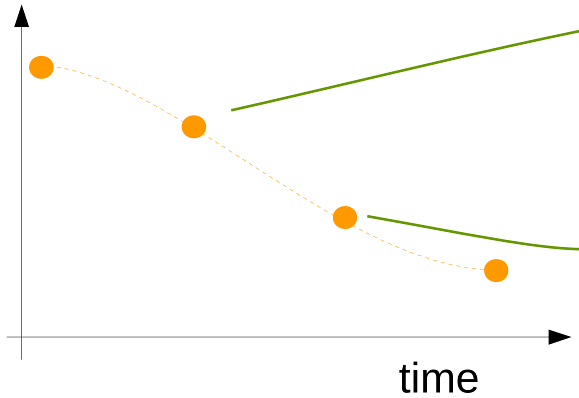
Linking data & model: Bayesian workflow



Likelihood – Multinomial (= conditional binomial)

(Jager *et al.*, 2011 ; Kon Kam King *et al.*, 2015)

$\mathbf{N} = \{\text{survival}\}$



SD or IT on $\mathcal{E} = \{\text{conc, time}\}$

$$N(t_i) \sim \mathcal{B}\left(\text{size} = N(t_{i-1}), \text{prob} = \frac{S(t_i, c(t_i))}{S(t_{i-1}, c(t_{i-1}))}\right)$$

$$\pi(\mathbf{N}|\theta) = N_0! \prod_{i=1}^{n+1} \frac{(S(t_{i-1}) - S(t_i))^{N_{i-1}-N_i}}{(N_{i-1} - N_i)!}$$

Priors : from the experimental design

(Delignette-Muller *et al.*, 2017)

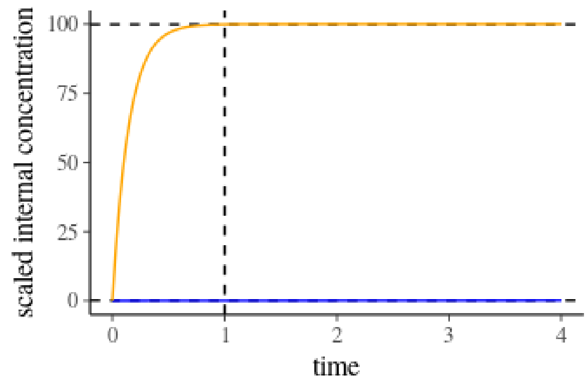
$$\log_{10}(\theta) \sim \mathcal{N}\left(\frac{\log_{10}(\theta_{min}) + \log_{10}(\theta_{max})}{2}, \frac{\log_{10}(\theta_{max}) - \log_{10}(\theta_{min})}{4}\right) \Rightarrow [\theta_{min}, \theta_{max}] = [Q2.5, Q97.5]$$

Toxicokinetics (TK)

$$\frac{dC^*_{int}(t)}{dt} = k_d(C_{ext}(t) - C^*_{int}(t))$$

Dominant exposure/excretion rate

$$\forall t\,,\; C_{ext}(t) = C_{ext} \quad \Rightarrow \quad C^*_{int}(t) = C_{ext}\left(1 - e^{-k_d t}\right)$$



$$k_d : \quad [C^*_{int} \text{ is } 99.9\%\, C_{ext} \text{ at } t_{min}\,,\; C^*_{int} \text{ is } 0.1\%\, C_{ext} \text{ at } t_{max}] = \left[\frac{-\ln(0,999)}{t_{max}}, \frac{-\ln(0,001)}{t_{min}}\right]$$

Priors : from the experimental design

(Delignette-Muller *et al.*, 2017)

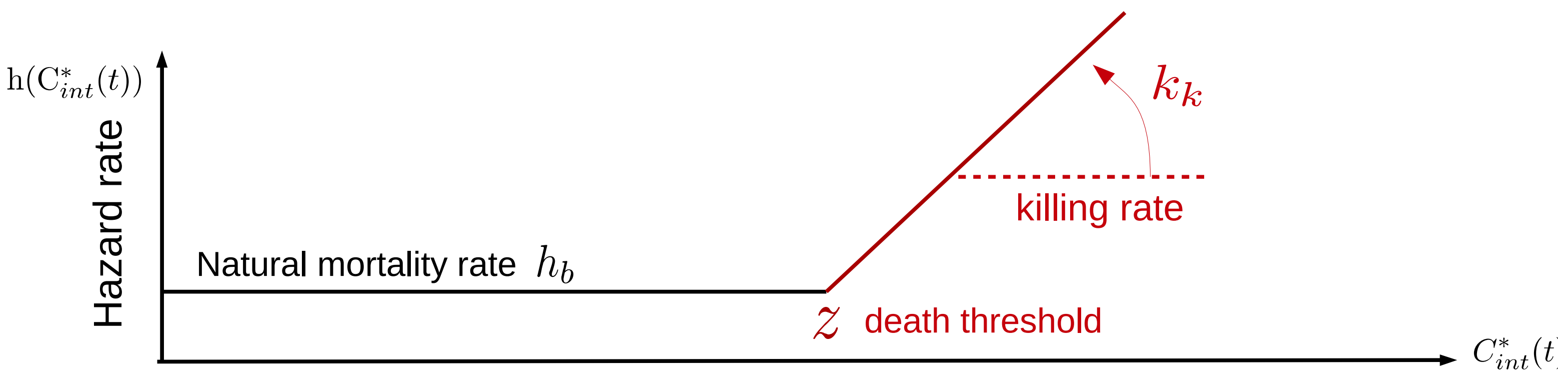
$$\log_{10}(\theta) \sim \mathcal{N}\left(\frac{\log_{10}(\theta_{min}) + \log_{10}(\theta_{max})}{2}, \frac{\log_{10}(\theta_{max}) - \log_{10}(\theta_{min})}{4}\right) \Rightarrow [\theta_{min}, \theta_{max}] = [Q2.5, Q97.5]$$

Toxicodynamics (TD) - SD

$$h_b : \quad \left[\text{ lifetime } \overset{\text{avg}}{=} 1000 \times t_{max} \text{ , } 50\% \text{ death } t_{min}\right] = \left[\frac{-\ln(0,999)}{t_{max}}, \frac{-\ln(0,5)}{t_{min}}\right]$$

$$z : \quad [C_{min} \text{ , } C_{max}]$$

$$k_k : \quad \left[99.9\% \text{ survive at } (C_{max}, t_{max}, z \approx C_{min}) \text{ , } 99.9\% \text{ death } \Delta^C_{min}\right] = \left[\frac{-\ln(0,999)}{t_{max}(C_{max} - C_{min})}, \frac{-\ln(0,001)}{t_{min}\Delta^C_{min}}\right]$$



Priors : from the experimental design

$$\log_{10}(\theta) \sim \mathcal{N}\left(\frac{\log_{10}(\theta_{min}) + \log_{10}(\theta_{max})}{2}, \frac{\log_{10}(\theta_{max}) - \log_{10}(\theta_{min})}{4}\right) \Rightarrow [\theta_{min}, \theta_{max}] = [Q2.5, Q97.5]$$

Toxicodynamics (TD) – IT log-logistic

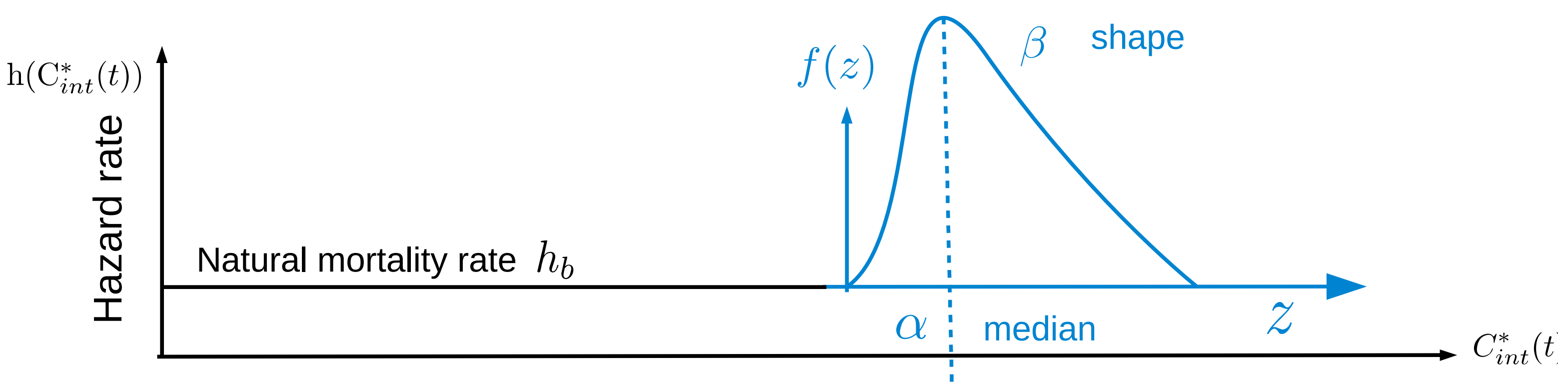
h_b : $$\left[\text{lifetime} \overset{\text{avg}}{=} 1000 \times t_{max}\,,\ 50\%\ \text{death}\ t_{min}\right] = \left[\frac{-\ln(0,999)}{t_{max}}\,,\ \frac{-\ln(0,5)}{t_{min}}\right]$$

α : $[C_{min}\,,\ C_{max}]$

β : $\log_{10}(\beta) \sim \mathcal{U}\,[-2\,,\ 2]$ or

$$\beta = \ln(39)/\ln(F_s) \quad \Rightarrow \quad 95\%\ \text{interval} = [\alpha/F_s\,,\ \alpha \times F_s]$$

$$\ln(F_s) \sim \mathcal{U}[0\,;\ C_{max}]$$



Application: Environmental Risk Assessment

1. **Easy** to use
2. **Minimize user influence** on calibration of parameters
3. Allow **prediction with uncertainty**

replicate	time	conc	survival
A	$t_{A,1}$	$C_{A,1}$	$N_{A,1}$
A	$t_{A,2}$	$C_{A,2}$	$N_{A,2}$
...	...	...	...
B	$t_{B,1}$	$C_{B,1}$	$N_{B,1}$
B	$t_{B,2}$	$C_{B,2}$	$N_{B,2}$
...	...	...	...
C	$t_{C,1}$	$C_{C,1}$	$N_{C,1}$
...	...	...	...

R + Stan

'rstanTKTD'

R + JAGS

'morse'

(version 2.2.0 - Delignette-Muller *et al.*, 2016 ;
(version 3.0.0 - Baudrot *et al.*, 2017)

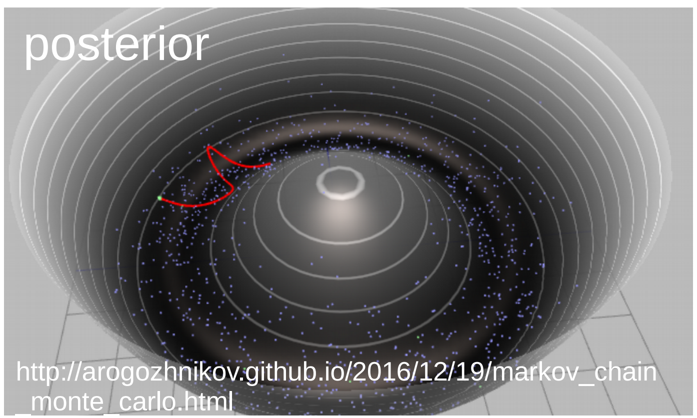
Bayesian Software : JAGS & Stan

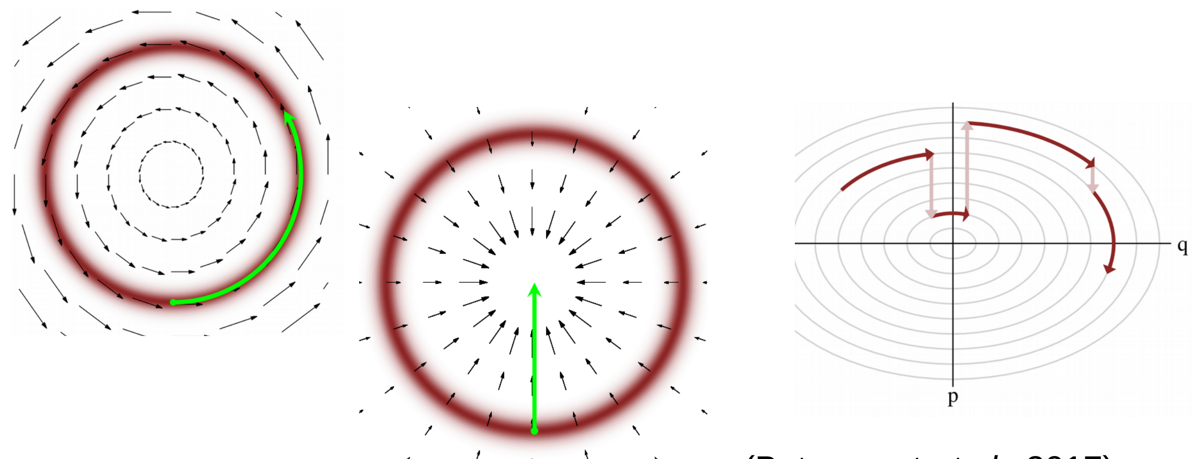
JAGS: Gibbs sampler

- Accept/reject on ‘energy’ of current position to next position

Stan: Hamiltonian Monte Carlo

- Accept/reject on ‘potential energy’ and move with ‘kinetic energy’





(Betancourt *et al.*, 2017)

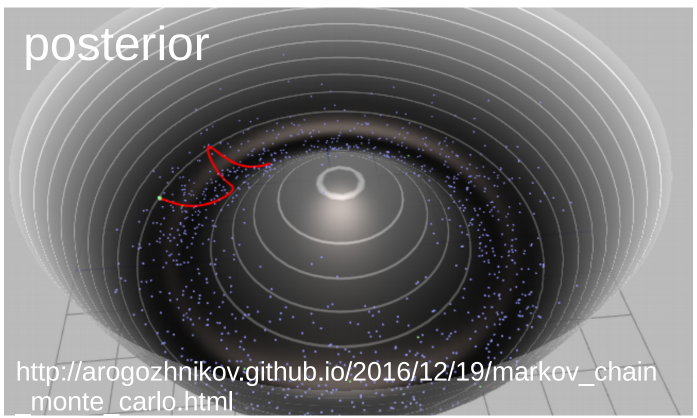
Bayesian Software : JAGS & Stan

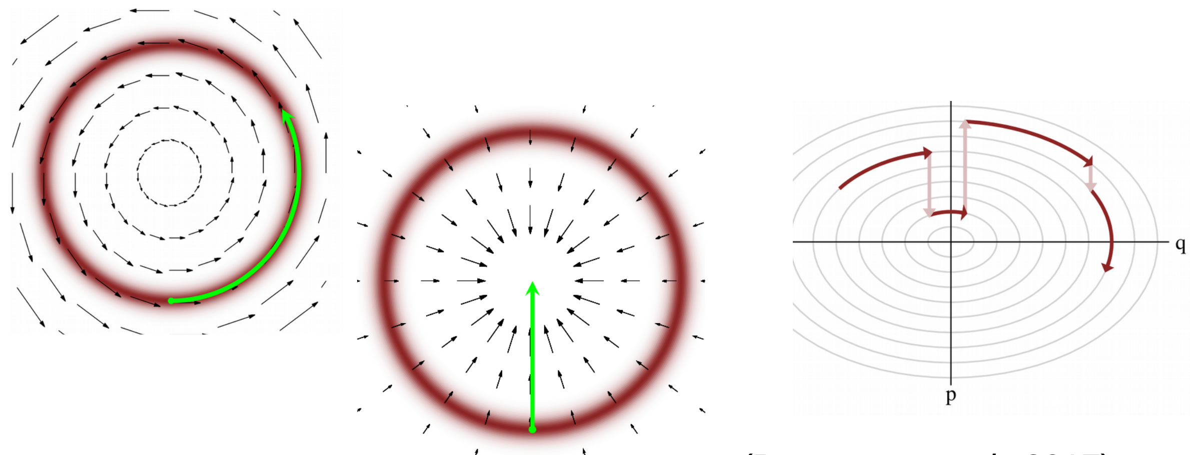
JAGS: Gibbs sampler

- Accept/reject on ‘energy’ of current position to next position

Stan: Hamiltonian Monte Carlo

- Accept/reject on ‘potential energy’ and move with ‘kinetic energy’





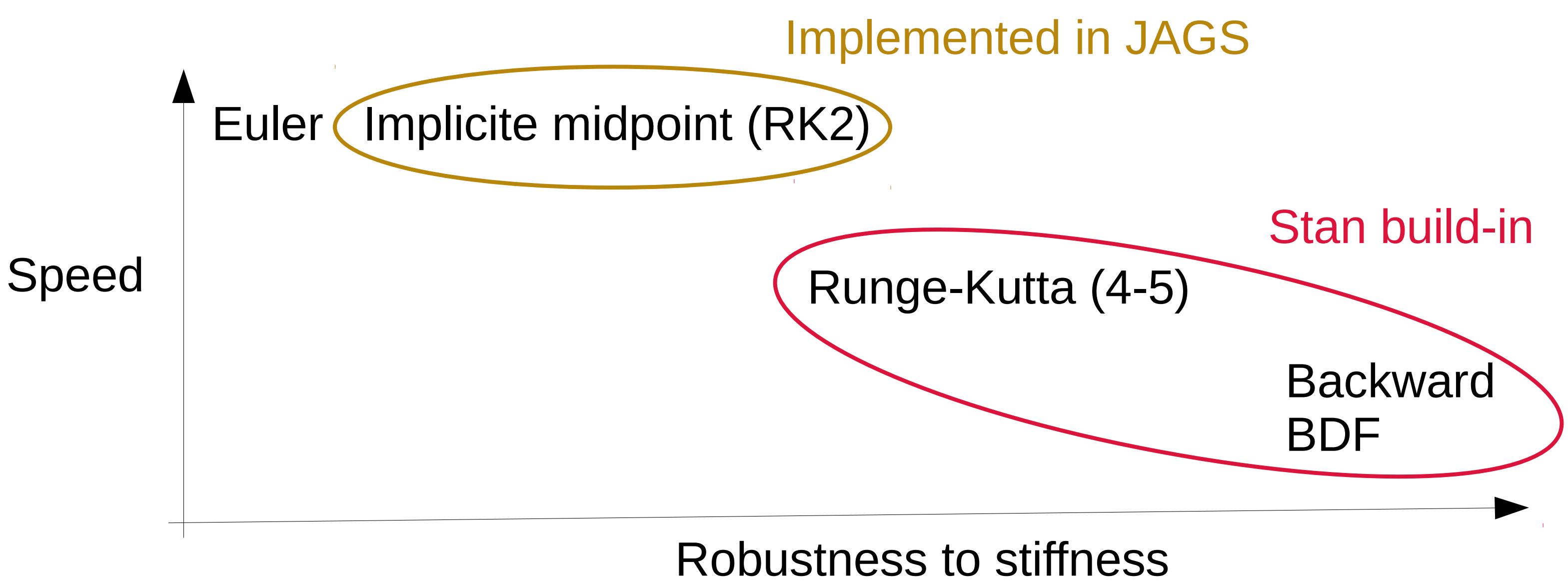
(Betancourt *et al.*, 2017)

- JAGS: Some difficulties when variables in hierarchical models are highly correlated (M. Plummer blog post)
- Stan: Iterations have a higher cost BUT significantly less iterations
- Stan is delicate to tune (Betancourt *et al.*, 2017)

ODEs integrators for variable concentrations

Potential **stiff problems** with integration of 2 time-scales:

- TK: time-scale of physiological mechanisms
- TD: time-scale of individual effect dynamics



Implementation

JAGS

Just Another Gibbs Sampler

```
"model {
  #--- priors
  kd_log10 ~ dnorm(kd_meanlog10, kd_taulog10)
  ...
  #--- parameter
  kd <- 10**kd_log10
  ...
  #--- ODEs Integration
  for( i in 1:n_data){
    ##--- Implicit midpoint method
    H_int[replicate_ID_long[i], time_ID_long[i]]
    ...
    ...
  }
  for( i in 1:n_data_red){
    ##--- Survival rate
    H[replicate_ID[i], time_ID_red[i]]  <- H_int[...]
    psurv[i] = exp( - H[replicate_ID[i], time_ID_red[i]])
    ##--- Likelihood
    Nsurv[i] ~ dbin(psurv[i]/psurv[i_prec[i]] , Nprec[i])
}"
```

Stan

Turing complete language

```
functions {
  int find_interval_elem(...) {
    ...
  }
  real linearInterp(...){
    ...
  }
  real[] TKTD_varSD(...) {
    ...
    //--- Model :
    dy_dt[1] =  kd * ( conc_linInterp - y[1]);
    dy_dt[2] = kk * max(0, y[1] - z) + hb;
  }
  matrix solve_TKTD_varSD(...){
    integrate_ode_rk45(TKTD_varSD, y0, t0, ts, theta, ...);
    ...
  }
}
data {
  ...
}
parameters {
  ...
}
transformed parameters {
  ...
}
model {
  Nsurv[...] ~ binomial( Nprec[...], Conditional_psurv[...]);
}
```

Examples

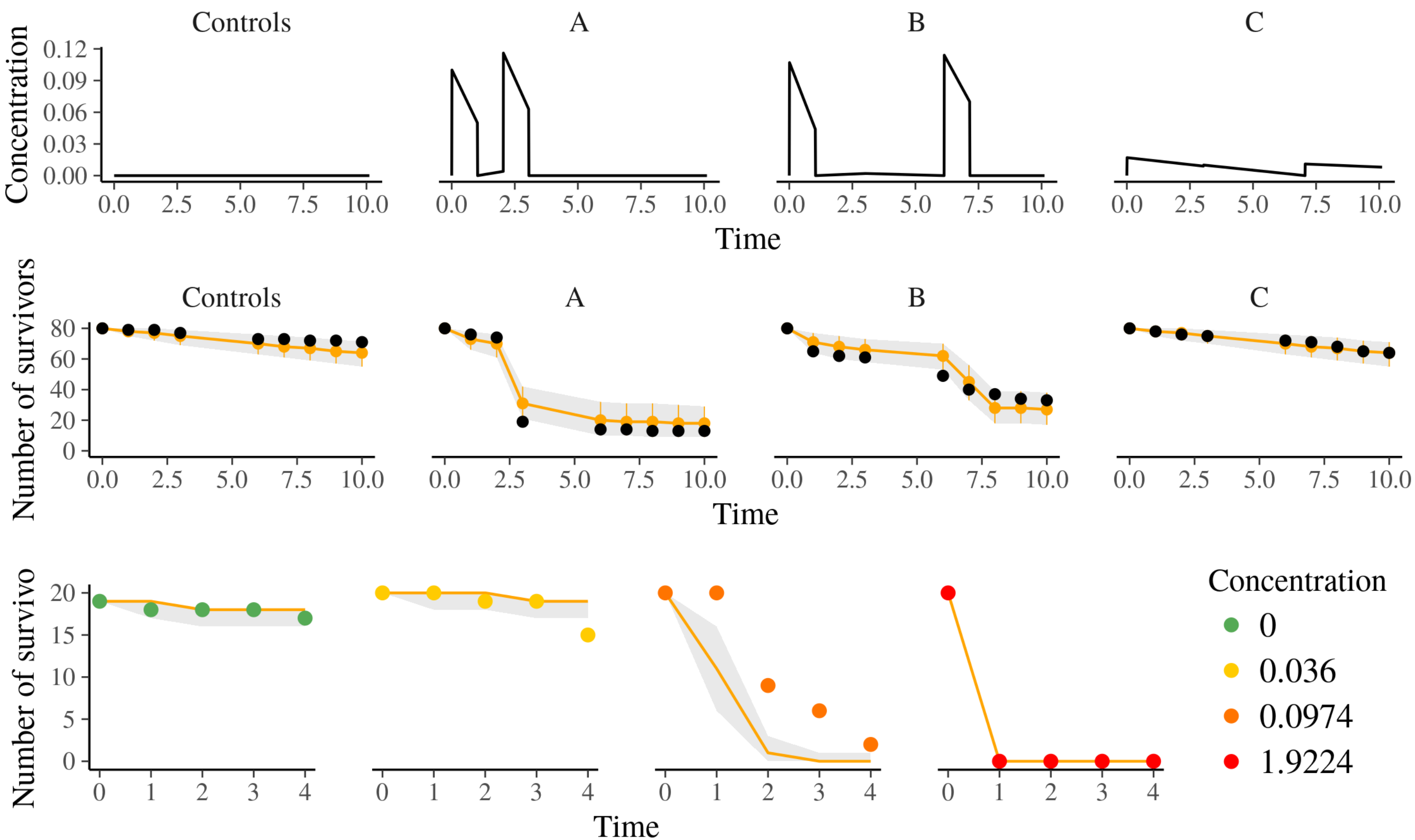
(Ashauer *et al.*, 2016 ; Nyman *et al.*, 2012)

- Survival of *Gammarus pulex*
- 5 pesticides: malathion, dimethoate, **cypermethrin**, carbendazim and propiconazole

Examples

(Ashauer *et al.*, 2016 ; Nyman *et al.*, 2012)

- Survival of *Gammarus pulex*
- 5 pesticides: malathion, dimethoate, **cypermethrin**, carbendazim and propiconazole



$$\mathcal{D} = (\mathcal{E}, \mathbf{N})$$

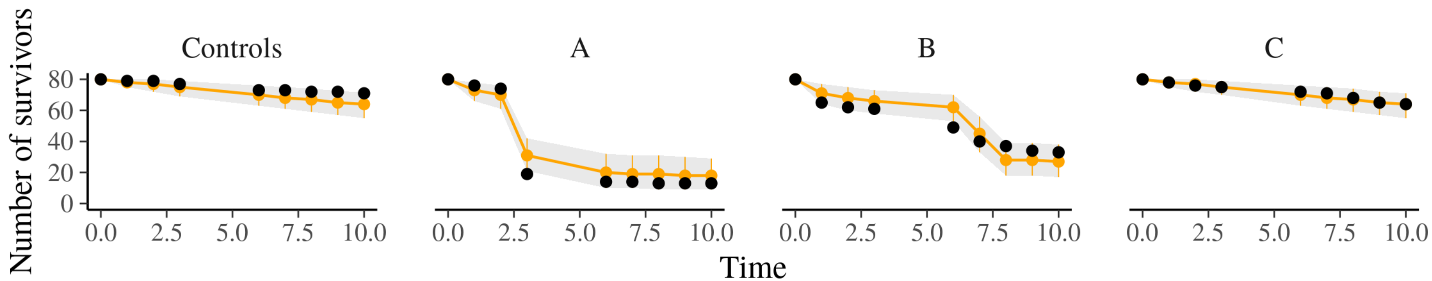
estimation

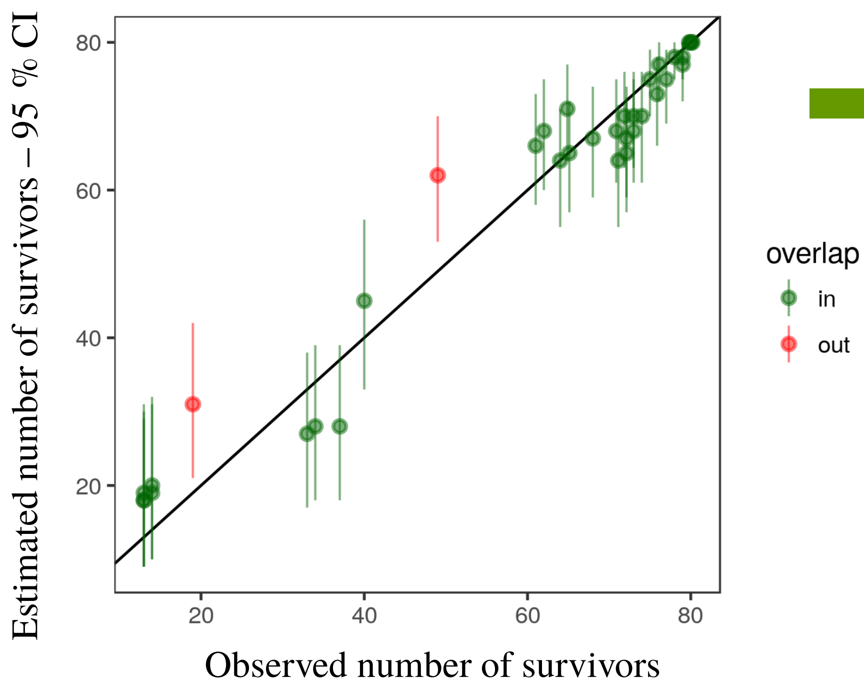
$$\mathbb{P}\left(\theta | \mathcal{D}\right)$$

prediction

$$R\left(\mathcal{E}' | \theta\right)$$

Posterior Predictive Check





94.4% of observations in 95 % Credible Intervals

WAIC & LOO

(Watanabe, 2010 ; Vehtari *et al.*, 2016)

- Log pointwise predictive density: $lpd = \log \prod_{i=1}^{n} \mathbb{P}_{post}(y_i) \Rightarrow \widehat{\text{lpd}} = \sum_{i=1}^{n} \log \left(\frac{1}{S} \sum_{s=1}^{S} \mathbb{P}(y_i|\theta^s) \right)$

- Variance of the log pointwise predictive density: $p_{WAIC} = \sum_{i=1}^{n} \mathbb{V}_{s=1}^{S} \mathbb{P}\left(y_i|\theta^s\right)$

➡ WAIC (Widely applicable information criterion): $WAIC = -2(\widehat{\text{lpd}} - p_{WAIC})$

WAIC & LOO

(Watanabe, 2010 ; Vehtari *et al.*, 2016)

- Log pointwise predictive density: $lpd = \log \prod_{i=1}^{n} \mathbb{P}_{post}(y_i) \Rightarrow \widehat{\text{lpd}} = \sum_{i=1}^{n} \log \left(\frac{1}{S} \sum_{s=1}^{S} \mathbb{P}(y_i|\theta^s) \right)$

- Variance of the log pointwise predictive density: $p_{WAIC} = \sum_{i=1}^{n} \mathbb{V}_{s=1}^{S} \mathbb{P}\left(y_i|\theta^s\right)$

➡ WAIC (Widely applicable information criterion): $WAIC = -2(\widehat{\text{lpd}} - p_{WAIC})$

- Log pointwise predictive density on sub-sample: $\widehat{\text{lpd}}_{-i} = \sum_{i=1}^{n} \log \left(\frac{1}{S} \sum_{s=1}^{S} \mathbb{P}(y_i|y_{-i};\theta^{is}) \right)$

- First order bias correction: $\overline{\text{lpd}}_{-i} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \log \left(\frac{1}{S} \sum_{s=1}^{S} \mathbb{P}(y_j|y_{-i};\theta^{is}) \right)$

➡ LOO (Leave-one-out cross-validation): $LOO = -2(\widehat{\text{lpd}_{-i}} + \widehat{\text{lpd}} - \overline{\text{lpd}}_{-i})$

Results – Feedback on:

- JAGS vs. Stan
- IT vs. SD
- Constant vs. Variable

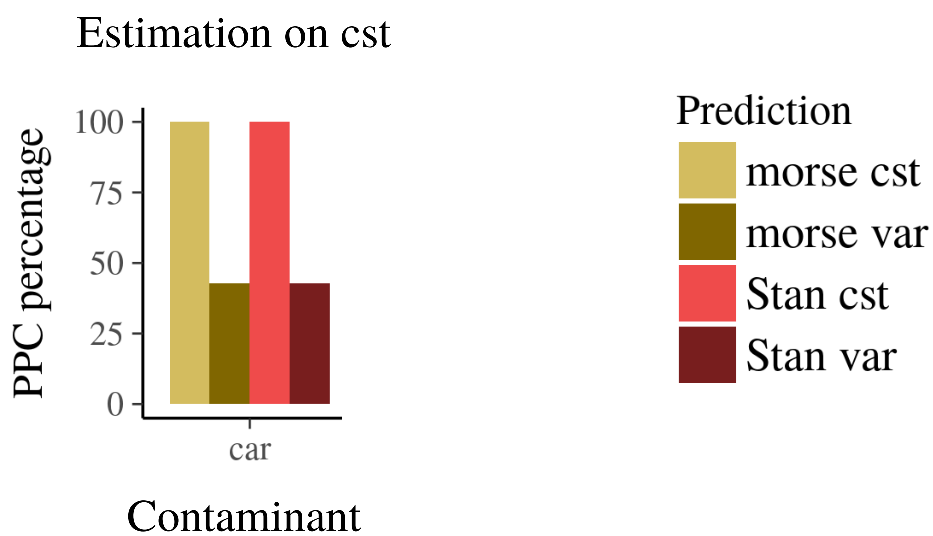
X

- Estimation
- Prediction

X

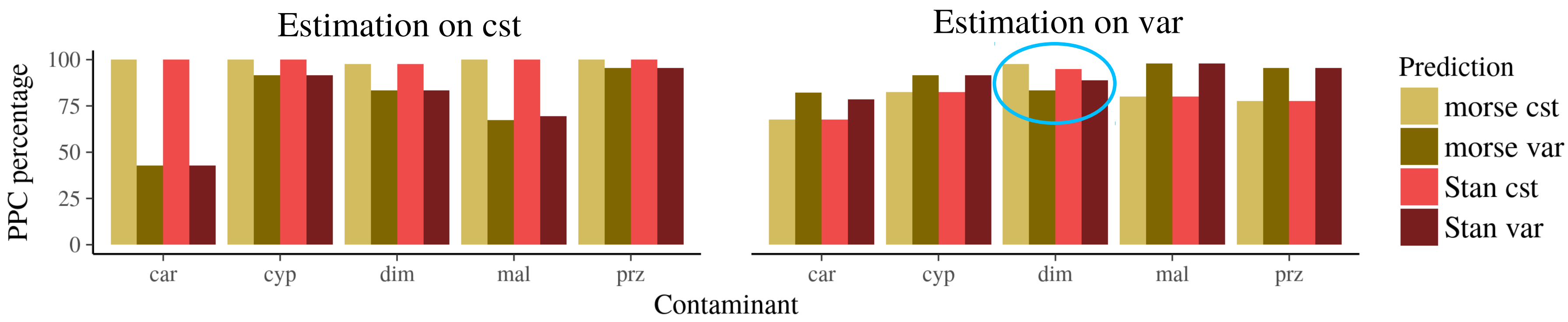
- Goodness-of-fit
- Time to compute

Results – SD



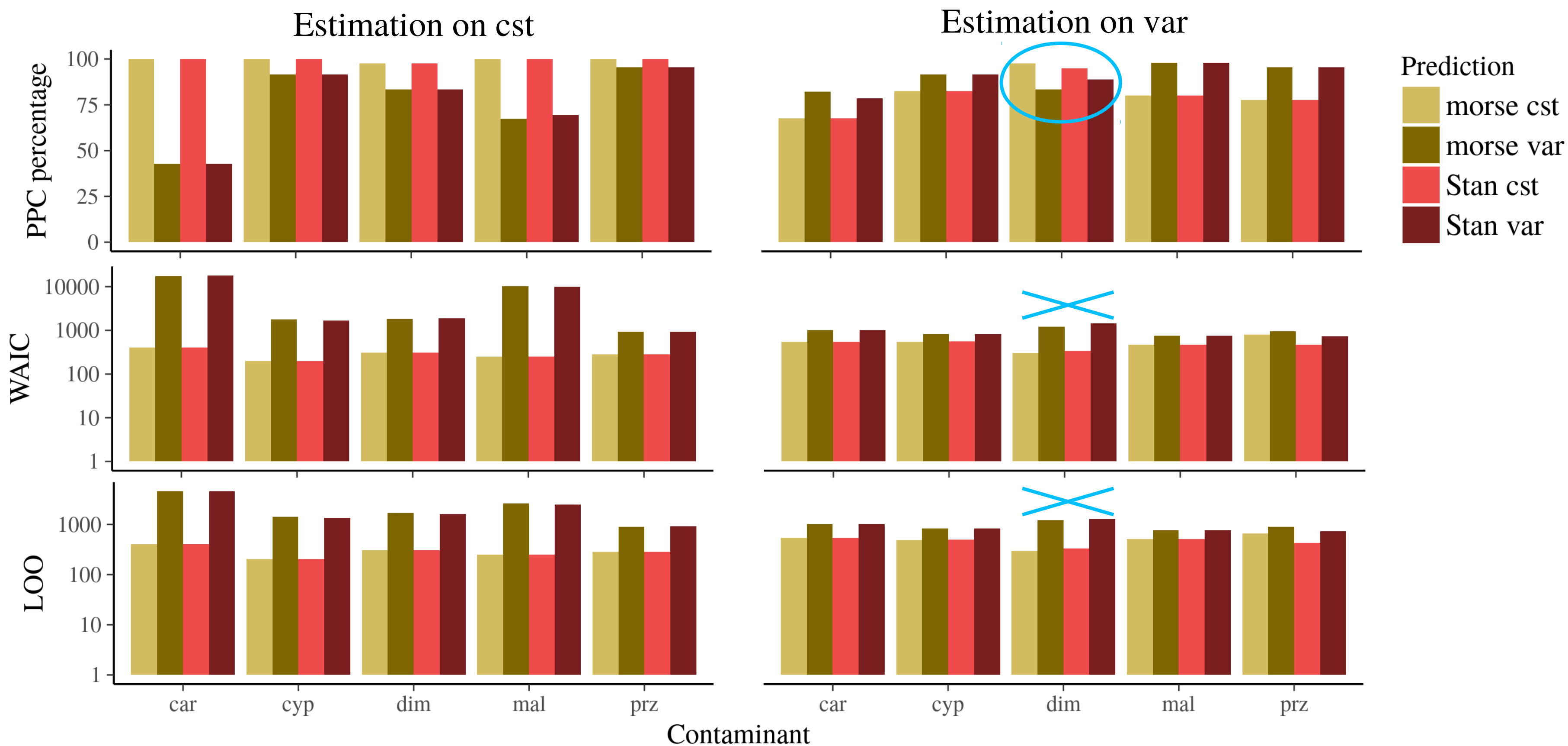
Estimation on cst
PPC percentage
100
75
50
25
0
car
Contaminant
Prediction
morse cst
morse var
Stan cst
Stan var

Results – SD



Estimation on cst
Estimation on var
PPC percentage
100
75
50
25
0
car
cyp
dim
mal
prz
Contaminant
Prediction
morse cst
morse var
Stan cst
Stan var

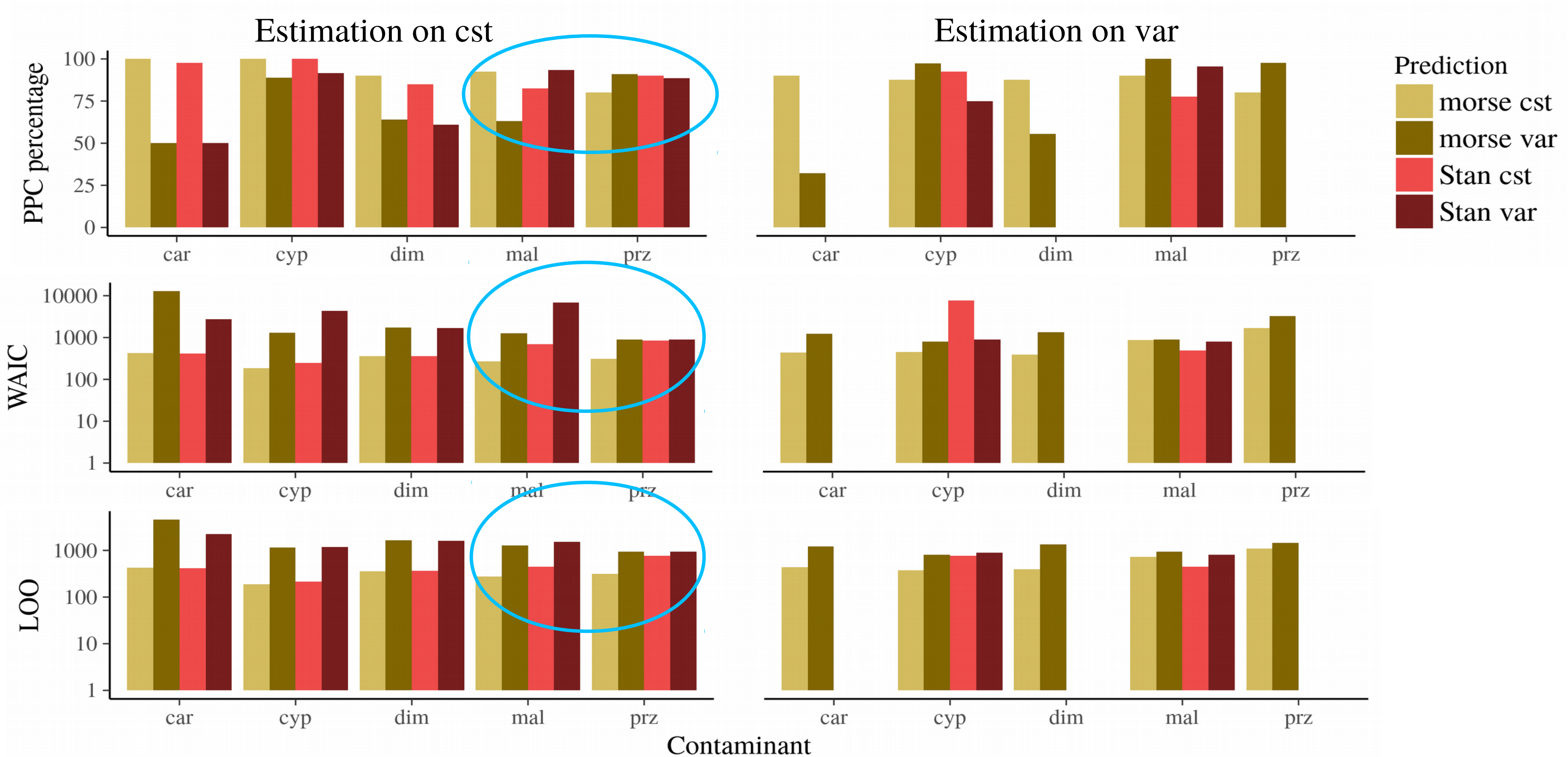
Results – SD



- Stan & JAGS: Same estimates
- Same message from WAIC & LOO
- PPC != WAIC & LOO for ‘variable dimethoate’: #data + estimation variability

Results – IT

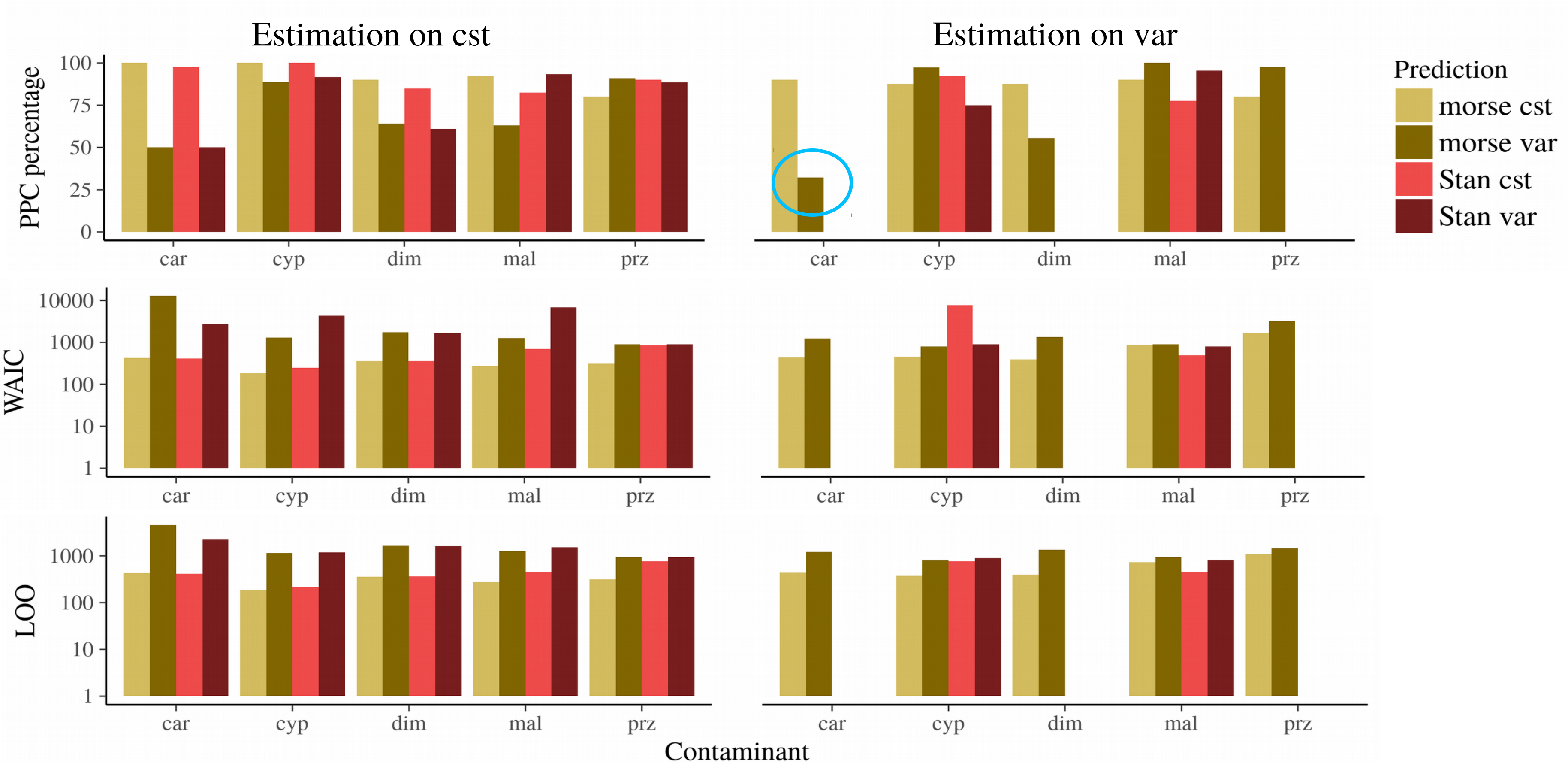
- No convergence in Stan for variable carbendazim, dimethoate and propiconazole



- Different estimates for ‘constant malathion and propiconazole’
- IT models seems harder to fit (high variability for Stan)

Results – IT

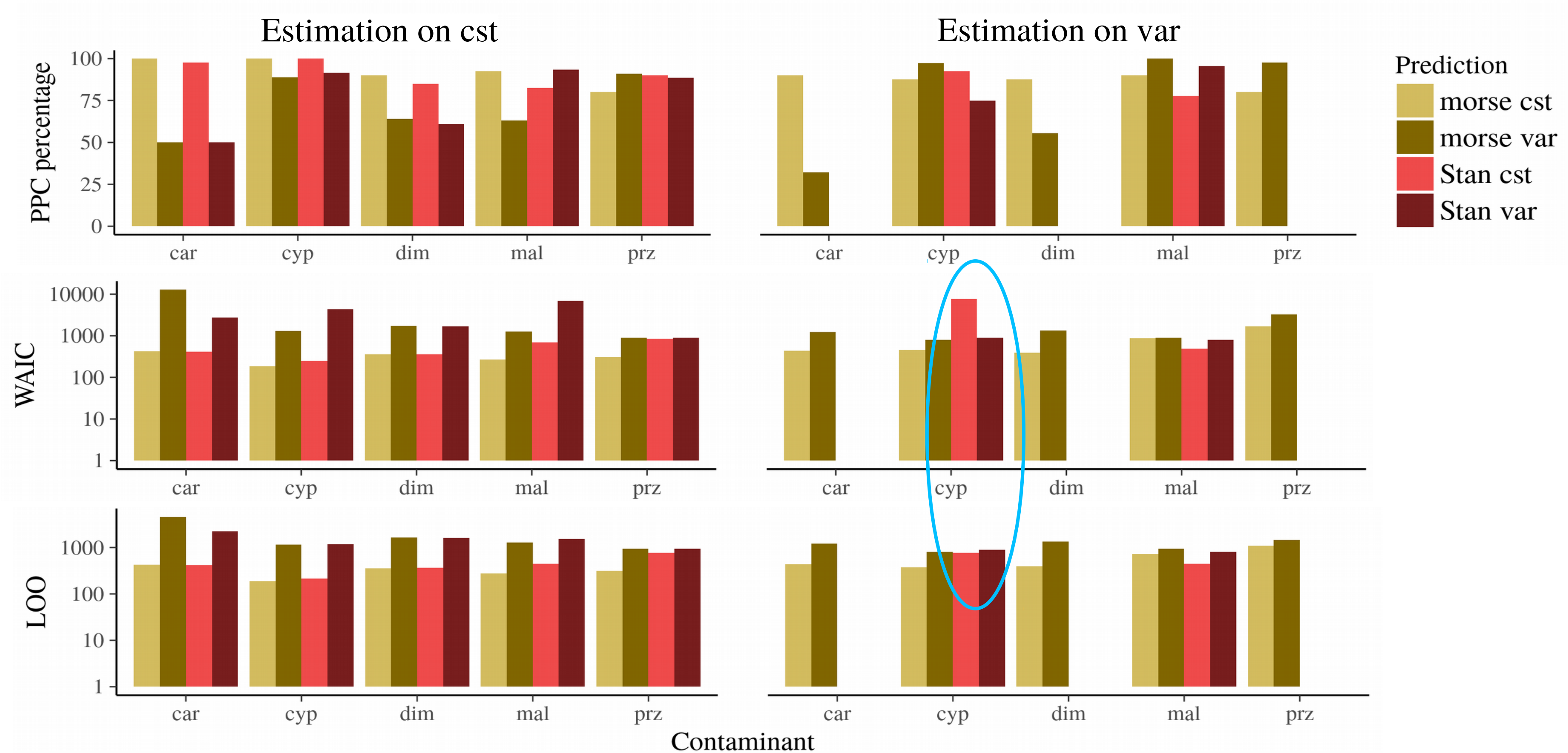
- No convergence in Stan for variable carbendazim, dimethoate and propiconazole



- Different estimates for 'constant malathion and propiconazole'
- IT models seems harder to fit (high variability for Stan)

Results – IT

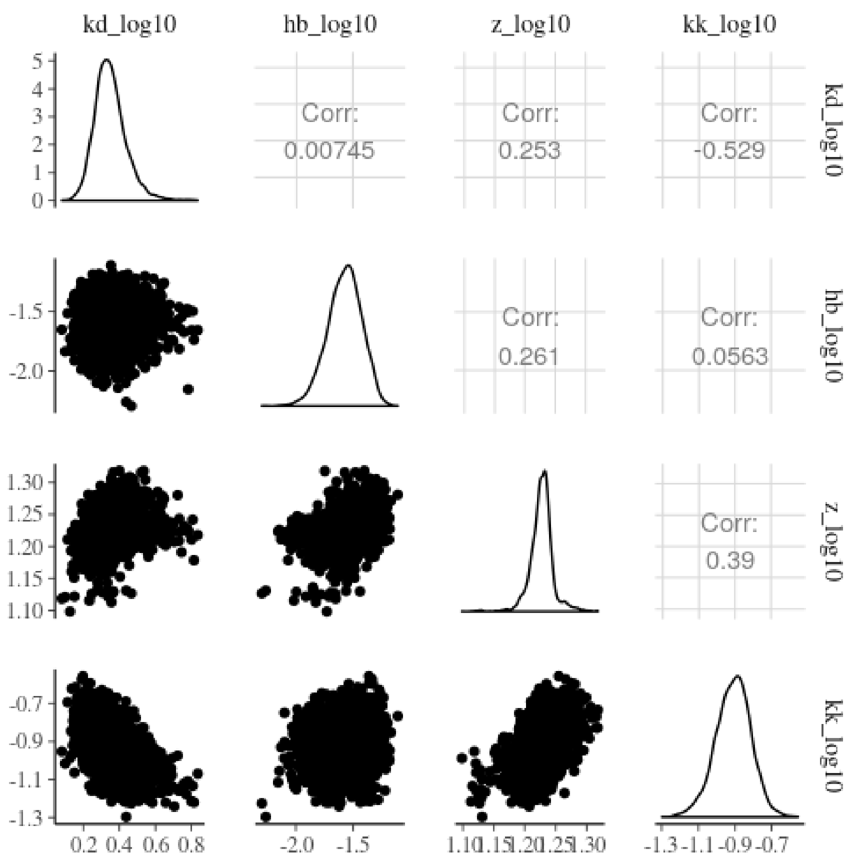
- No convergence in Stan for variable carbendazim, dimethoate and propiconazole



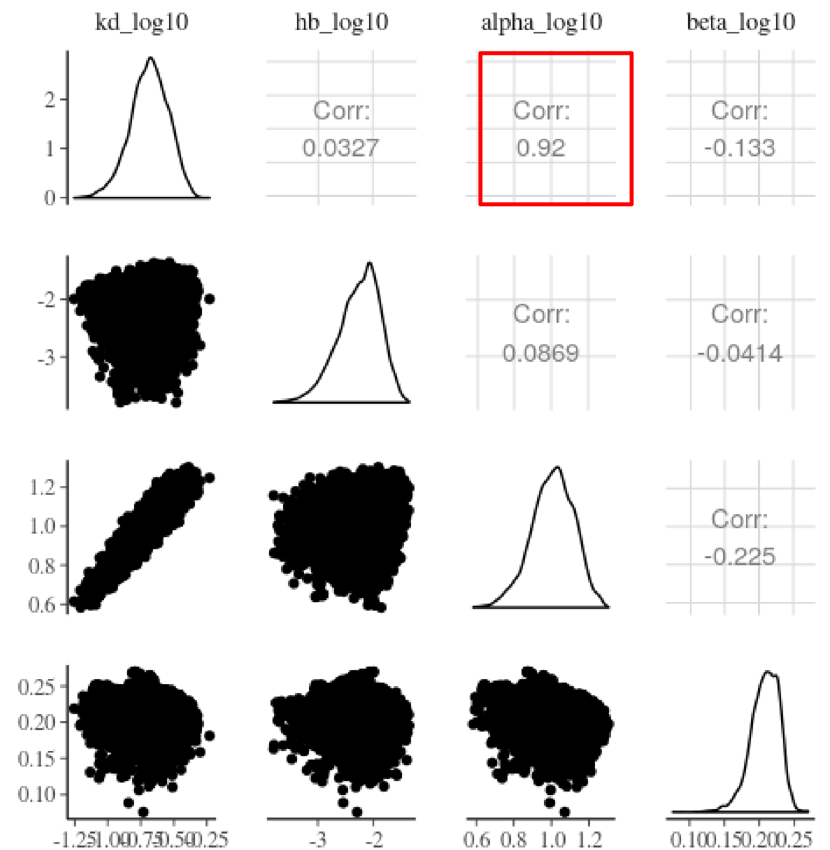
- Different estimates for ‘constant malathion and propiconazole’
- IT models seems harder to fit (high variability for Stan)

Results – SD vs IT

Propiconazole - GUST-SD



Propiconazole - GUST-IT



Results & Perspectives

Time to compute:

- **IT models faster than SD**. For ‘variable’ 2 ODEs in SD, 1 ODE in IT
- ‘variable’: Stan (~5-10min) faster than JAGS (~ 15-20min)

SD or IT:

- **SD robust compared to IT**: SD always converge + Same results JAGS & Stan
- Hard to fit IT model with Stan ; especially ‘variable’

- **JAGS always converge** with default parameters
- Stan is hard to tune (adapt_delta, tolerance + length ODEs)

‘Constant’ or ‘Variable’ exposure profile:

- Experimentation: easier for ‘constant exposure profile’
- In tested dataset: more replicates and/or profiles in constant scenarios
- Higher credible interval for ‘variable exposure profiles’
- ‘Variable’ profile are longer to fit: But 30min fit compared to 10 days of experiment

Goodness-of-fit:

- PCC, WAIC & LOO similares for estimations
- WAIC and LOO provides very similar results. WAIC faster to compute (in seconds)

Ongoing questions

Is there another way to choose between SD and IT?

- Biological system: chemical compound, species
- Structure of data: number, range of concentration, time-points, …
- Purpose: estimation, prediction
- What about the general model 'SD+IT' ?

R packages 'morse' and 'rstanTKTD'

→ **How the user can tune the analysis?**

- MCMC: Number chains, Iteration, level accept/reject, ...
- ODEs integrator (type, step, tolerance)
- IT distribution: log-logistic, log-normal, ...

Thank you for your attention

Sciencehood

$$\mathbb{P}(\text{Thanks}|\text{Attention}) = \frac{\mathbb{P}(\text{Attention}|\text{Thanks}) \times \mathbb{P}(\text{Thanks})}{\mathbb{P}(\text{Attention})}$$

Goodness of fitting interesting question