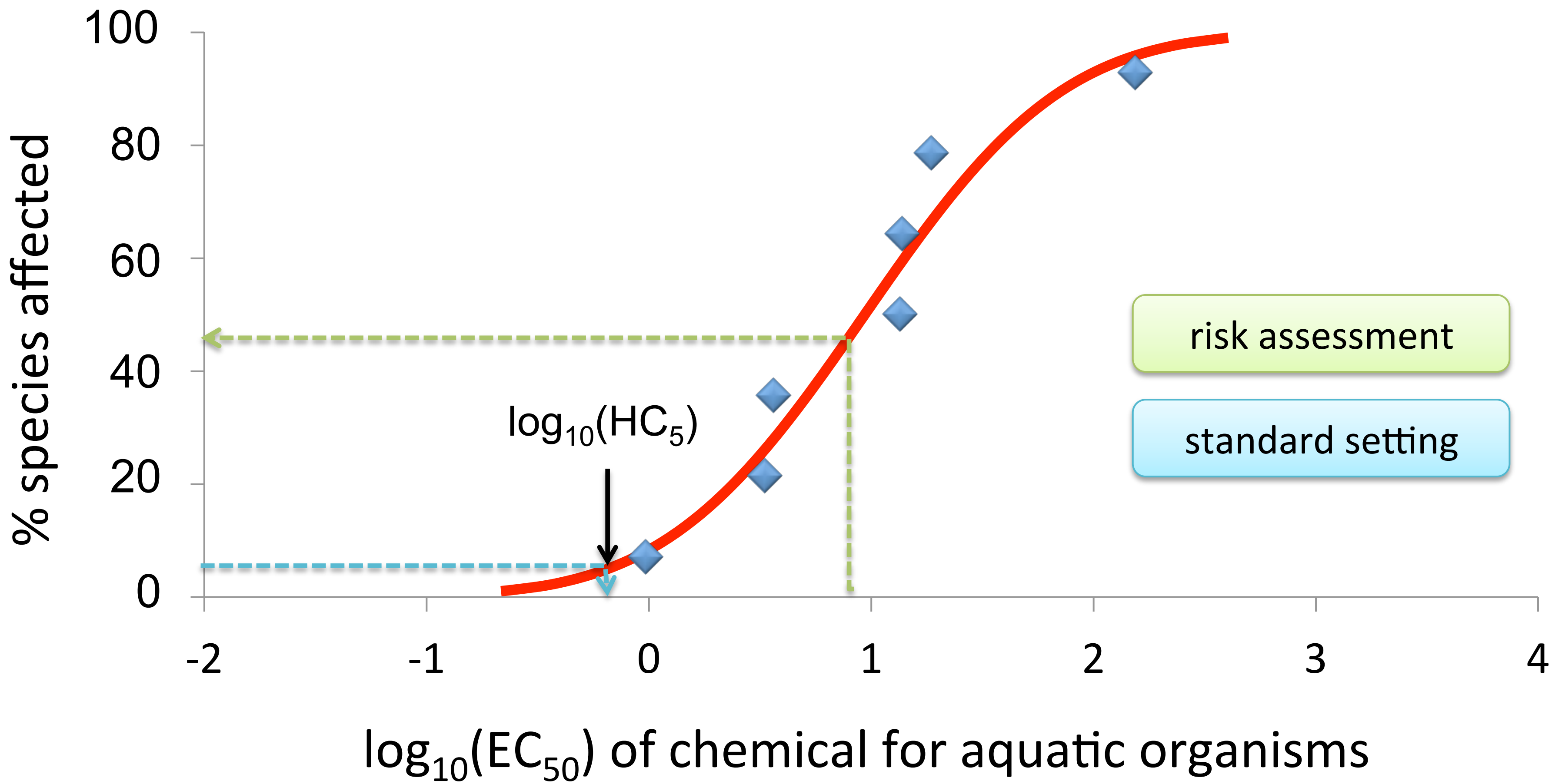
Extending the SSD Concept to Explore Some Foundational Model Limitations: A Bayesian Hierarchical Approach

***Graeme Hickey**[1], Peter Craig[1], Stuart Marshall[2], Oliver Price[2], Mathijs Smit[3], Andy Hart[4], Robert Luttik[5], Peter Chapman[2,retired], Dick de Zwart[5]*

[1]Department of Mathematical Sciences, Durham University, UK
[2]Safety and Environmental Assurance Centre, Unilever, Colworth, UK
[3]Statoil ASA, Trondheim, Norway
[4]The Food and Environment Research Agency, York, UK
[5]RIVM, Bilthoven, The Netherlands

Species Sensitivity Distributions (SSDs)



All Models Are Wrong…

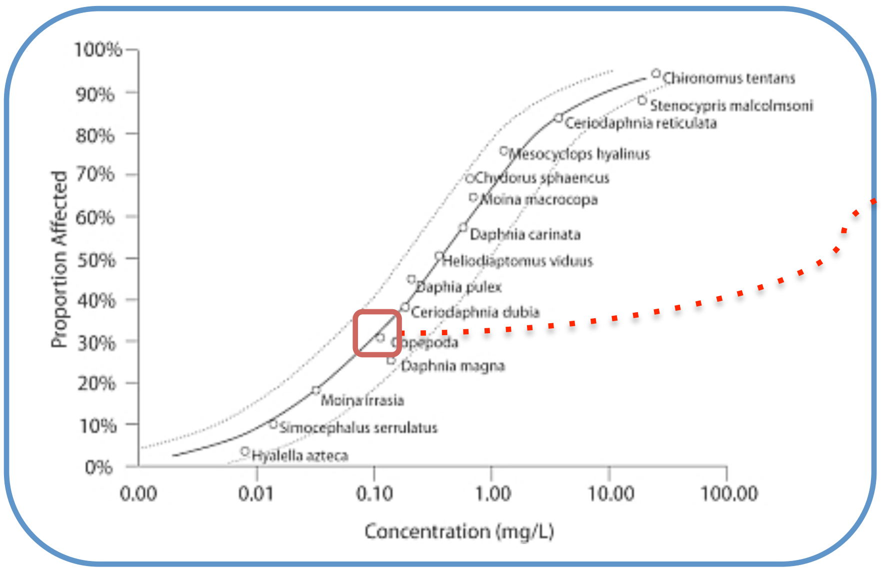
(the SSD is no exception!)… but some are better than others.

ASSUMPTION	PRACTICAL CONSEQUENCE
SSDs are independent	Information gained about each chemical risk assessment will not strengthen learning of future assessments.
Interspecies variation is attributable to chemical effects only	Observation of some species being more/less sensitive not accounted for.
Other sources of variation (e.g. inter-laboratory and intra-species variation) are ignorable *or* captured by an arbitrary assessment factor ($1 \leq AF \leq 5$)	Confounding of the HC_5 interpretation – should it just represent interspecies variation?
Representative of all ecosystems	No account of specific assemblages and differing diversities.

Hierarchical Modelling

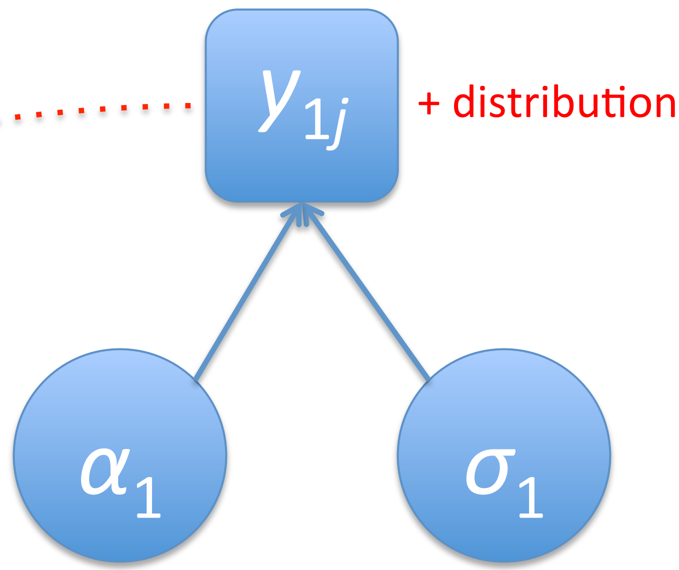
The standard SSD model can be written as a stochastic network

Chemical 1



=

log transformed toxicity datum
for chemical 1 and species *j*

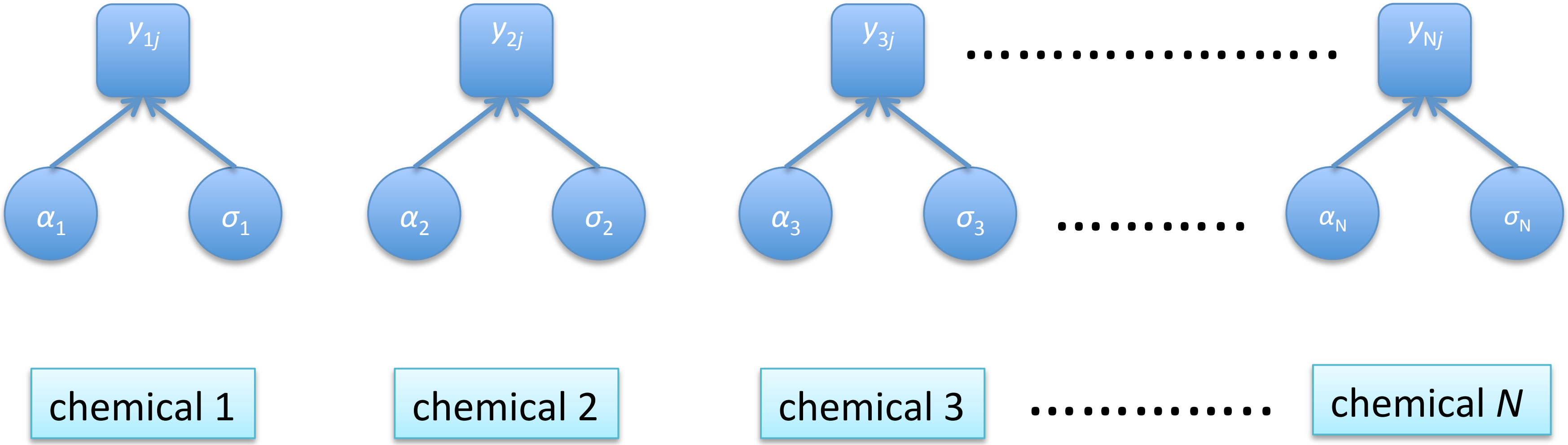


central
tendency

∝ dispersion
(interspecies
standard deviation)

Hierarchical Modelling

If we have *N* chemical risk assessments, the usual SSD model is a special (independence) case of a hierarchical model.

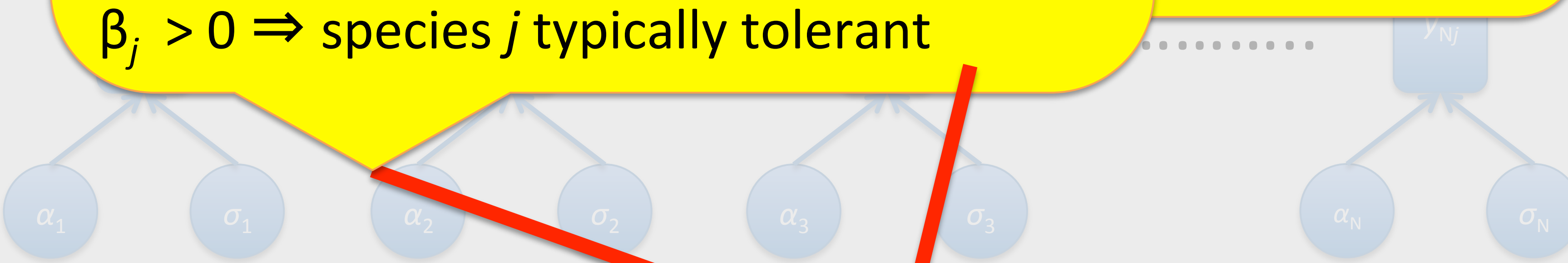


SSD interspecies variance parameters are heterogeneous *between* chemicals *i*.
Therefore β_j measures species position as number of standard deviations from mean (log-)toxicity.

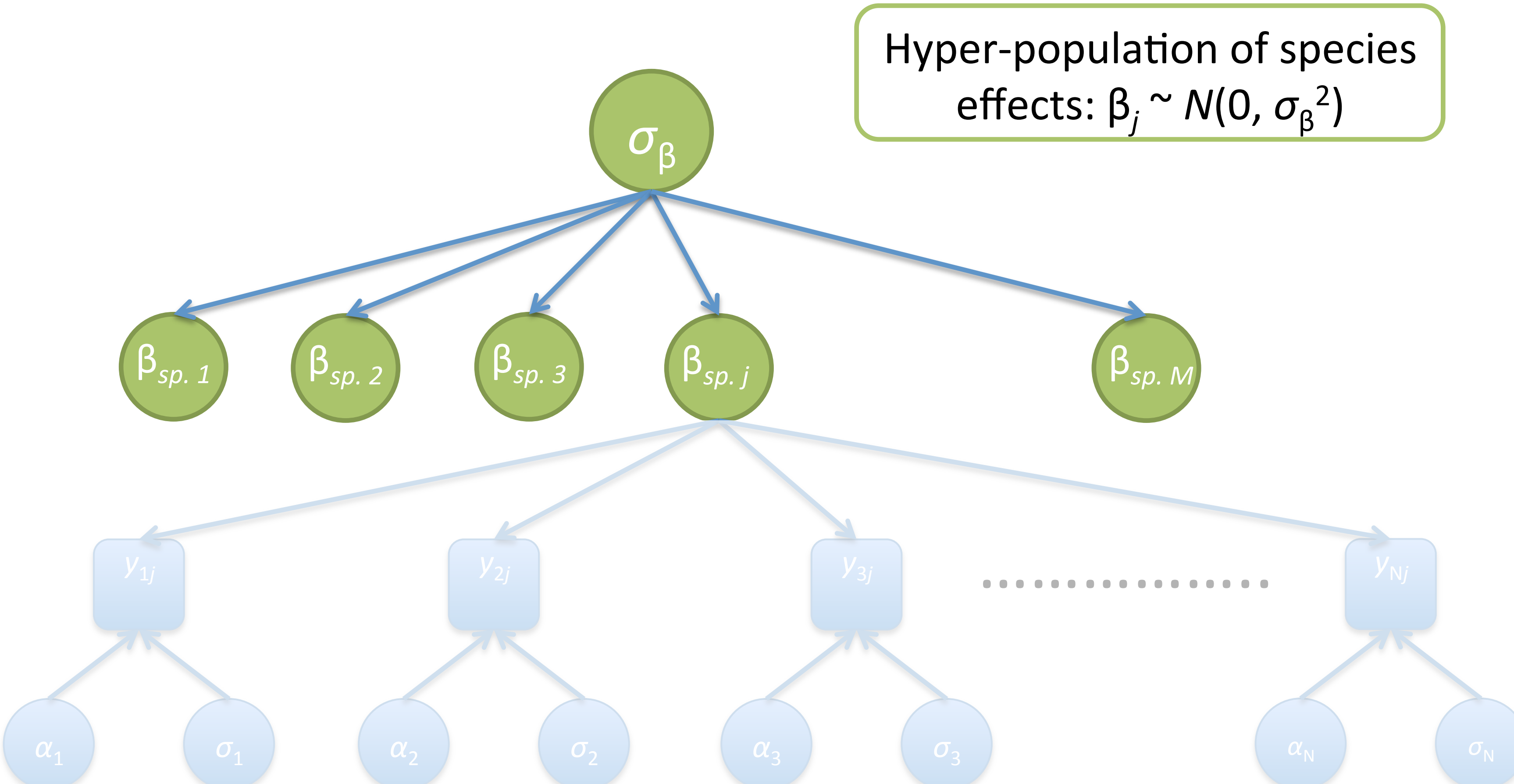
$\beta_j < 0 \Rightarrow$ species *j* typically sensitive
$\beta_j > 0 \Rightarrow$ species *j* typically tolerant

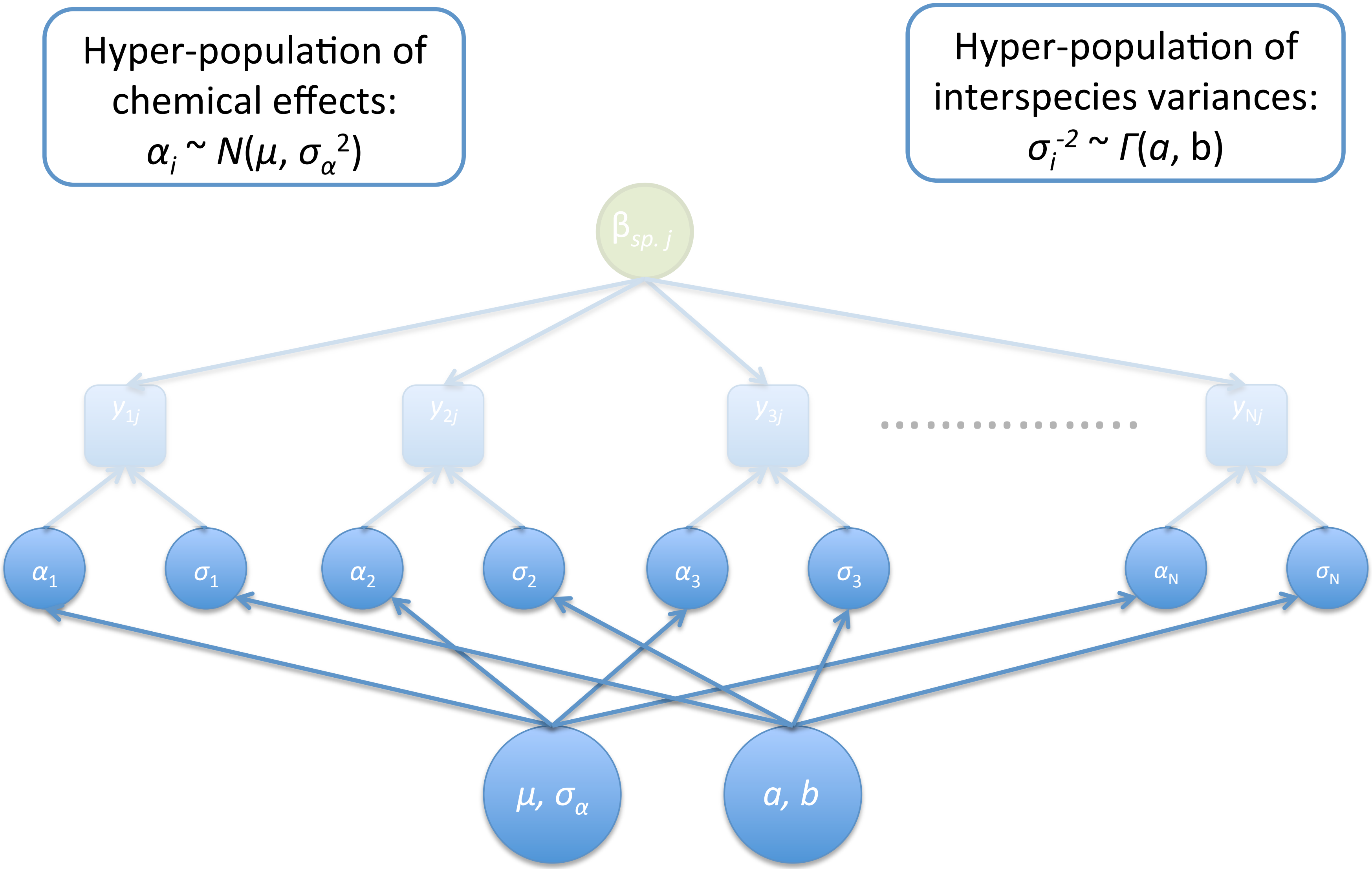
…sed into a linear sum
…and an error term

…erally accepted
…ers; although
…ed with
…suitable.



$$
\begin{aligned}
y_{ij} &= \alpha_i + \beta_j \sigma_i + \varepsilon_{ij} \\
\varepsilon_{ij} &\sim N(0, \sigma_i^2)
\end{aligned}
$$


Hyper-population of species effects: $\beta_j \sim N(0, \sigma_\beta^2)$
σ_β
$\beta_{sp.\ 1}$
$\beta_{sp.\ 2}$
$\beta_{sp.\ 3}$
$\beta_{sp.\ j}$
$\beta_{sp.\ M}$
y_{1j}
y_{2j}
y_{3j}
y_{Nj}
α_1
σ_1
α_2
σ_2
α_3
σ_3
α_N
σ_N


Hyper-population of chemical effects:
$\alpha_i \sim N(\mu, \sigma_\alpha^2)$
Hyper-population of interspecies variances:
$\sigma_i^{-2} \sim \Gamma(a, b)$
$\beta_{sp.\ j}$
y_{1j}
y_{2j}
y_{3j}
y_{Nj}
α_1
σ_1
α_2
σ_2
α_3
σ_3
α_N
σ_N
μ, σ_α
a, b

Bayesian Analysis

- Need to ensure propagation of > 1st level uncertainty.
- Update prior distributions about the hyper-parameters using observed data to retrieve posterior distributions.
- Use posterior distributions to make hazard assessment inferences for *retrospective* and *prospective* chemical assessments.

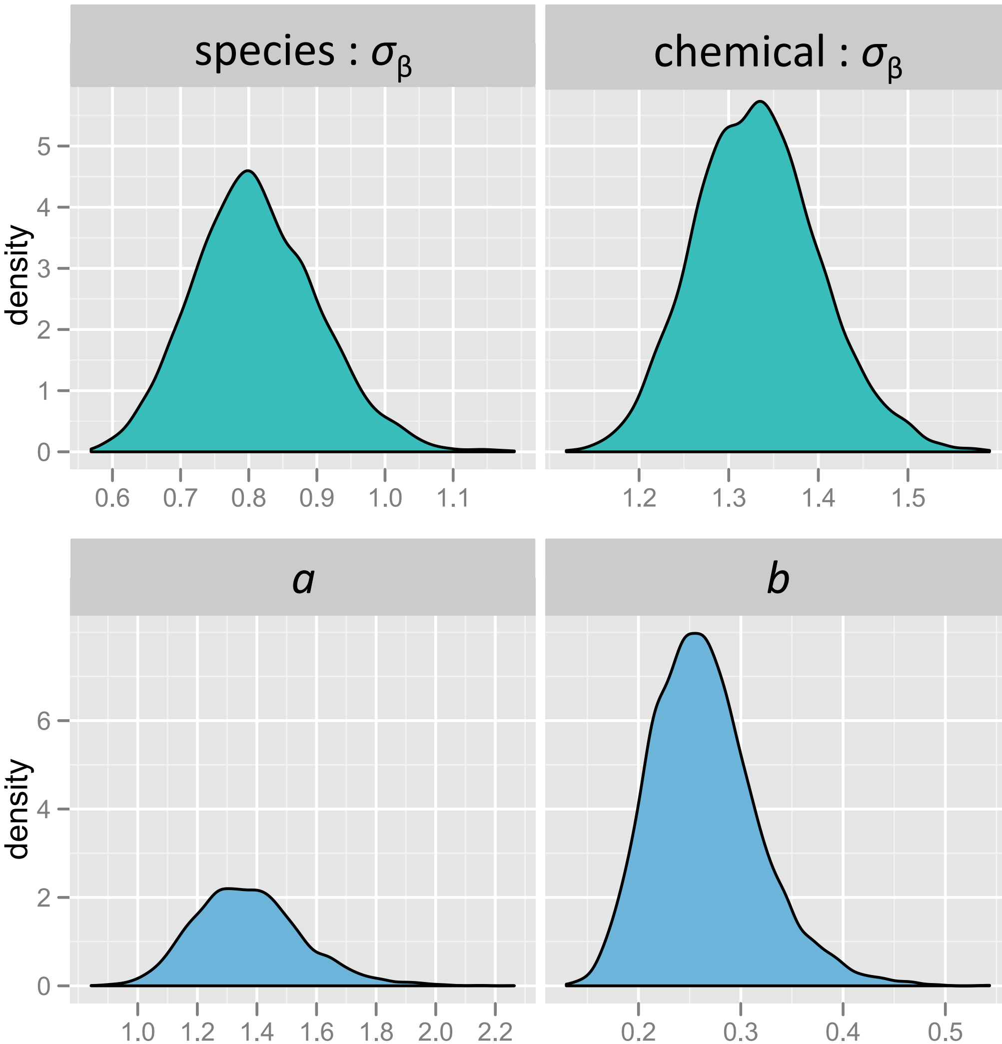
Example

- Ecotoxicity database extracted from the U.S. EPA Web-ICE database. 1600 E(L)C_{50} values (lethality or immobility) spanning 201 chemicals (each with $n_i \geq 5$) and 77 species.

  http://www.epa.gov/ceampubl/fchain/webice/

- Prior distributions chosen to closely represent 'ignorance' (so-called '*non-informative*').

Hyper-parameter Posterior Distributions



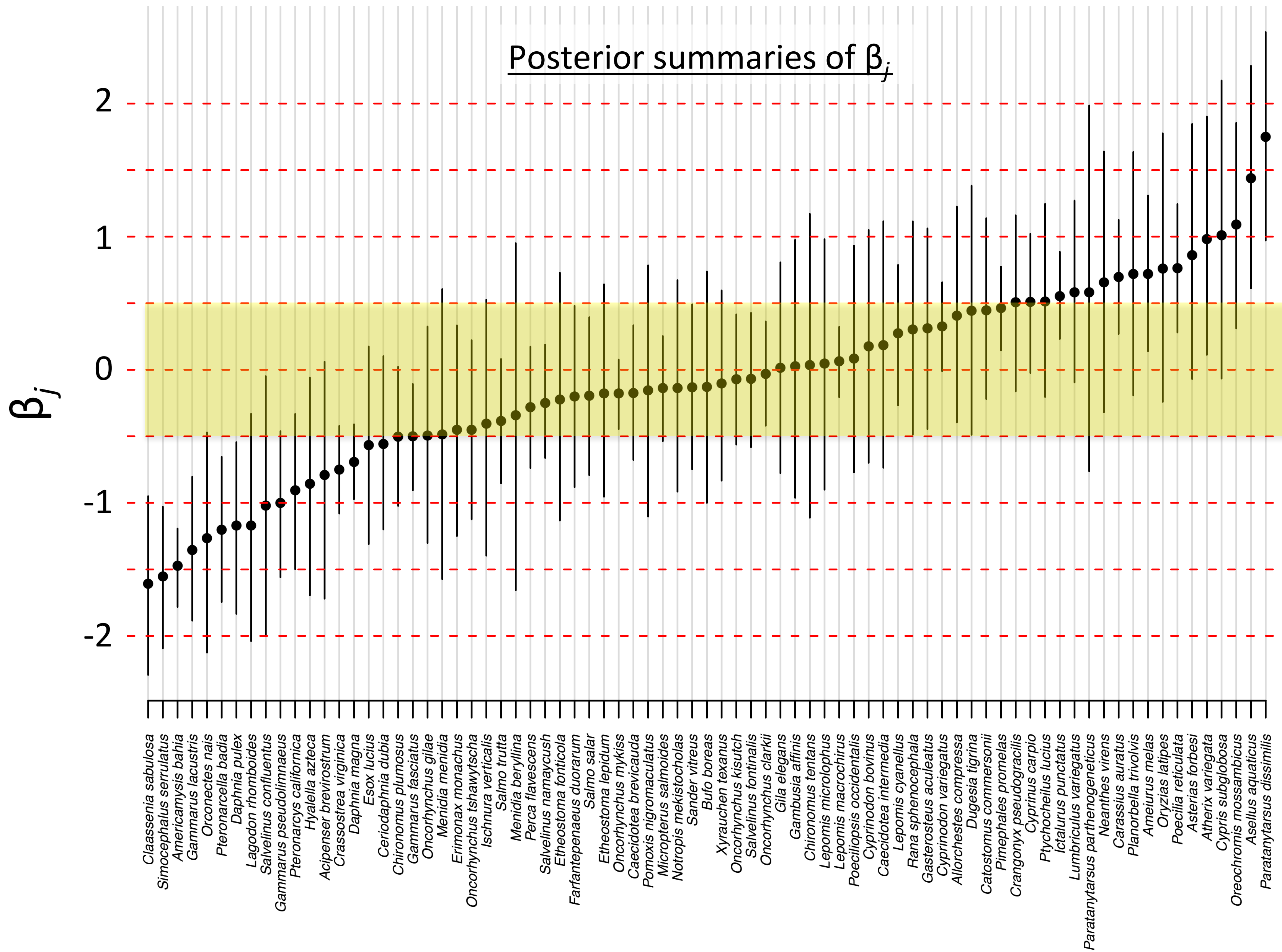
median (+ 95% credible interval)
σ_β: 0.81 (0.65, 1.01)
σ_α: 1.33 (1.21, 1.48)

⇒ more variation between chemicals

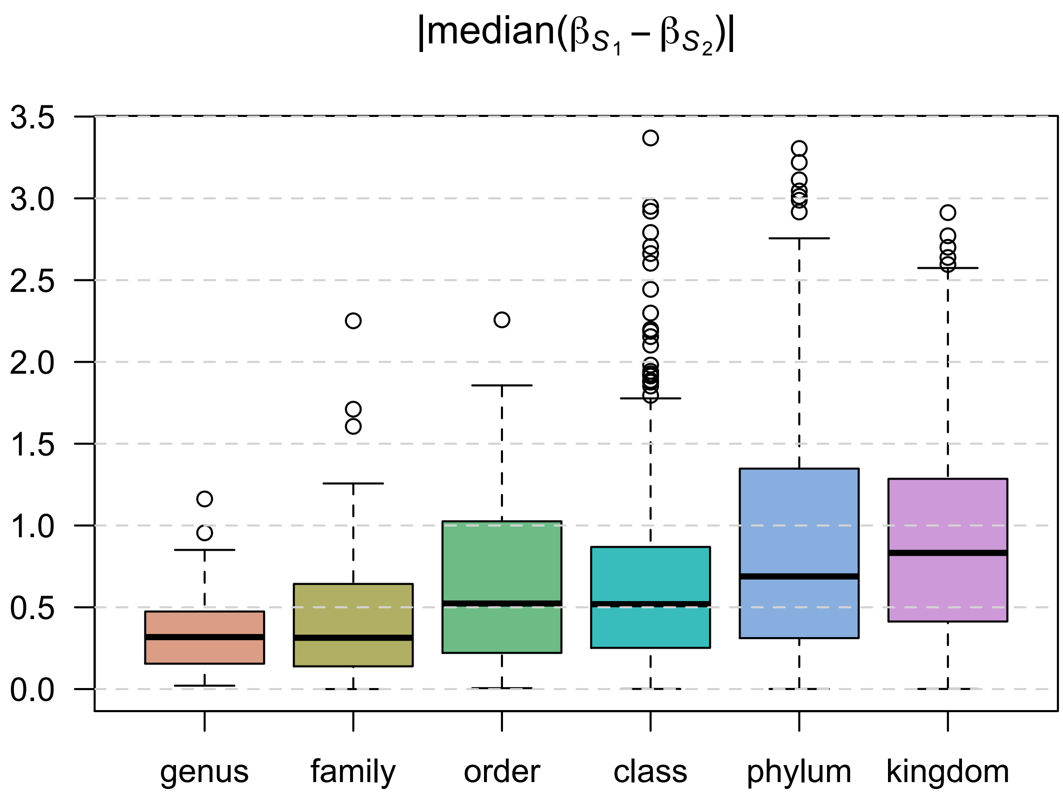
Estimates consistent with *European Food Safety Authority* (2005, *EFSA J.* **301**, pp. 1-45) report.

Equivalent to ≈ 3 additional measurements → stabilizes interspecies variance estimate.

Evidence that some species are more sensitive than others

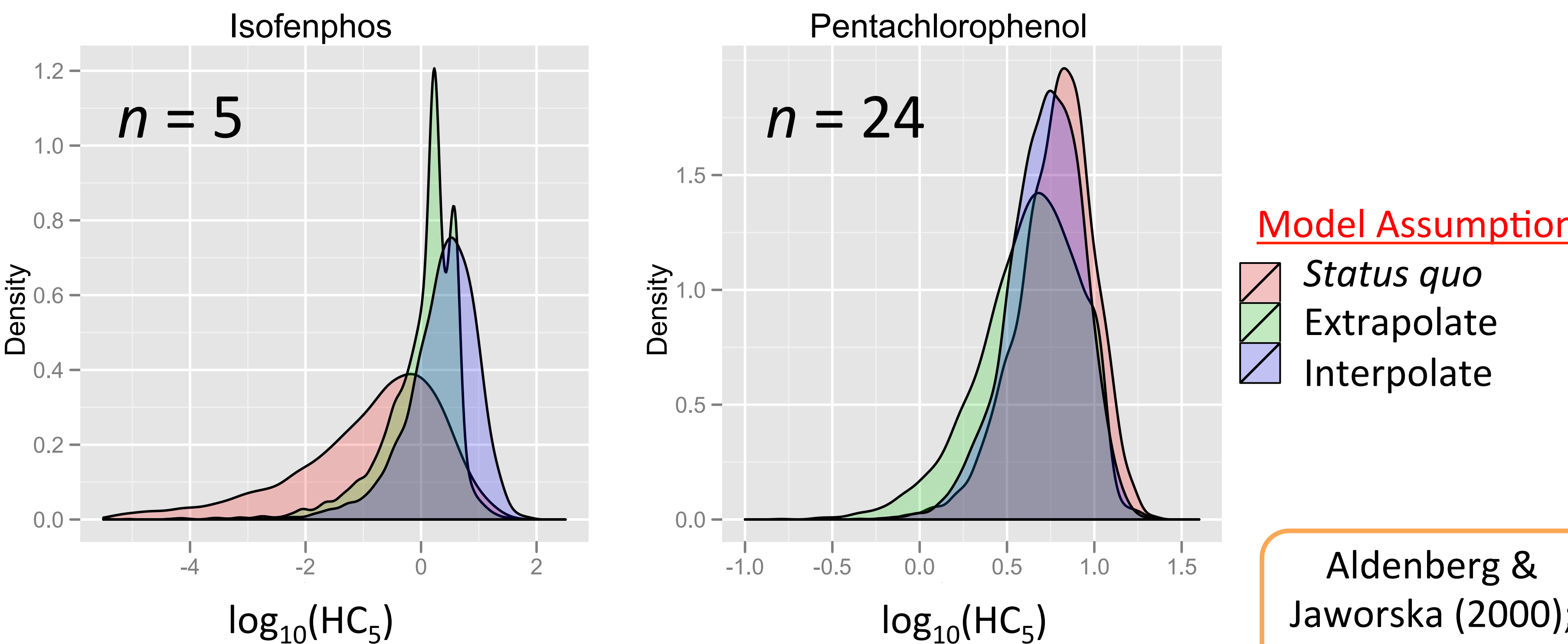


The Role of Taxonomy



The more taxonomically spread species are in an SSD, the larger the interspecies variance will be.

Posterior HC_5 Distributions



Aldenberg & Jaworska (2000); *EES*

Status quo = REACH Technical Guidance Document with log-normal SSD

Extrapolate = ecosystem is an infinitely large collection of species

Interpolate = ecosystem comprised of 77 species listed in database

} hierarchical models

Conclusions

- The SSD concept is not defunct!
- Hierarchical modelling and Bayesian statistics open up the option for 'better' modelling with transparent uncertainty propagation.
- Useable for multiple-hypothesis testing <u>and</u> risk management.

Acknowledgements

[DATA] Sandy Raimondo and Mace Barron (U.S. EPA, Gulf Ecology Division)

[DISCUSSION] Mick Hamer (Syngenta) and Malyka Galay-Burgos (ECETOC)

Existing Hierarchical Approaches

- Luttik & Aldenberg (1997), *ET&C*
- European Food Safety Authority (2005)
- Jager *et al.* (2007), *EES*
- Morton (2008), *Environmetrics*
- U.S. EPA Web-ICE Program(?)

Common theme: use data from multiple chemicals to improve future risk assessments