Statistical Eco(-toxico)logy

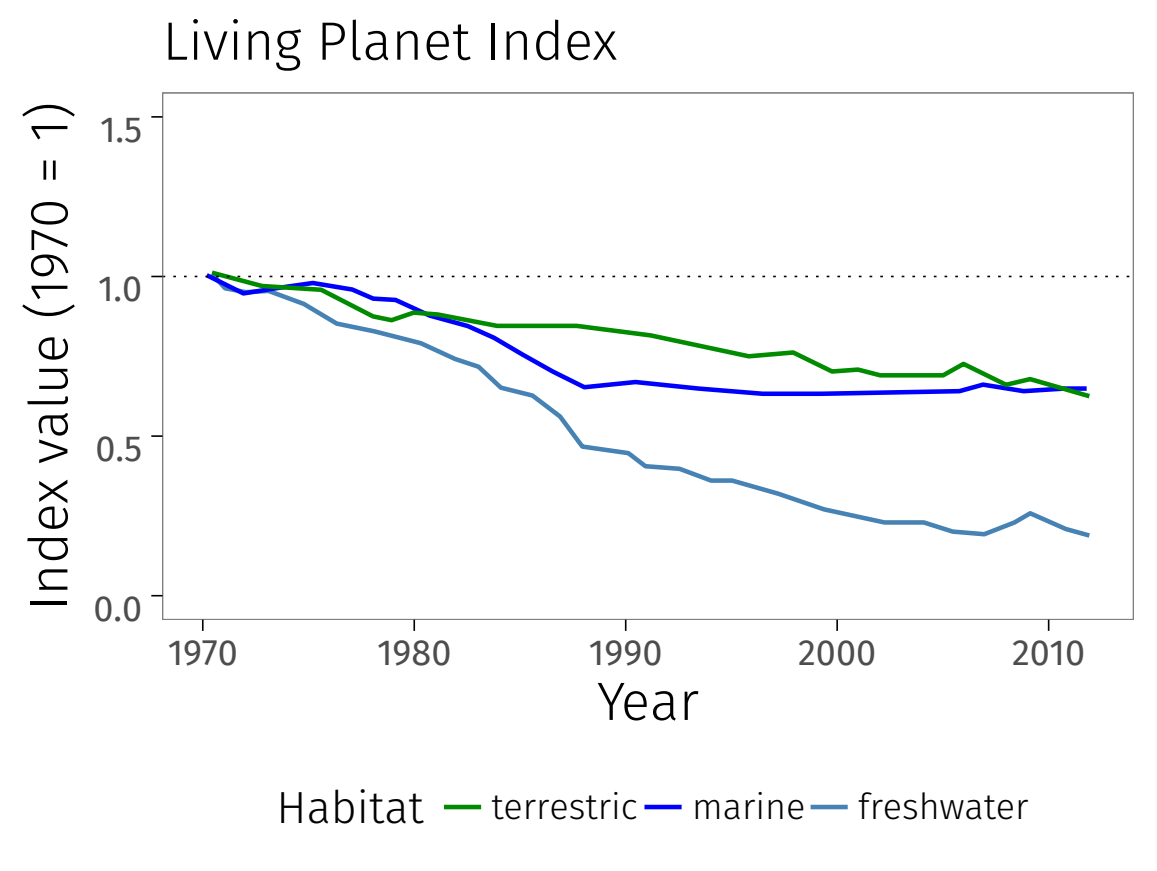
Improving the Utilisation of Data for Environmental Risk Assessment

Eduard Szöcs

25th January 2017

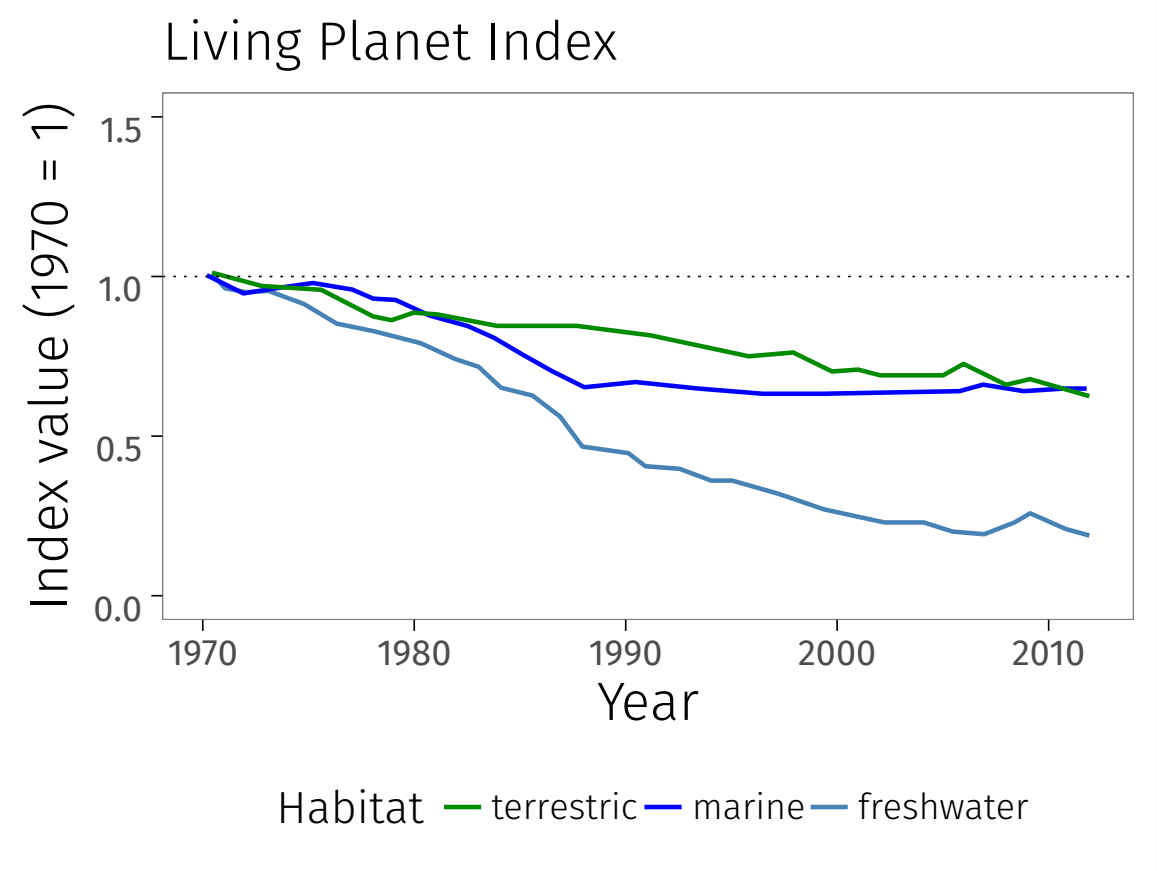
Table of contents

Freshwater biodiversity is strongly declining



WWF (2016), Living Planet Report; Vörösmarty (2010). Nature.

Freshwater biodiversity is strongly declining

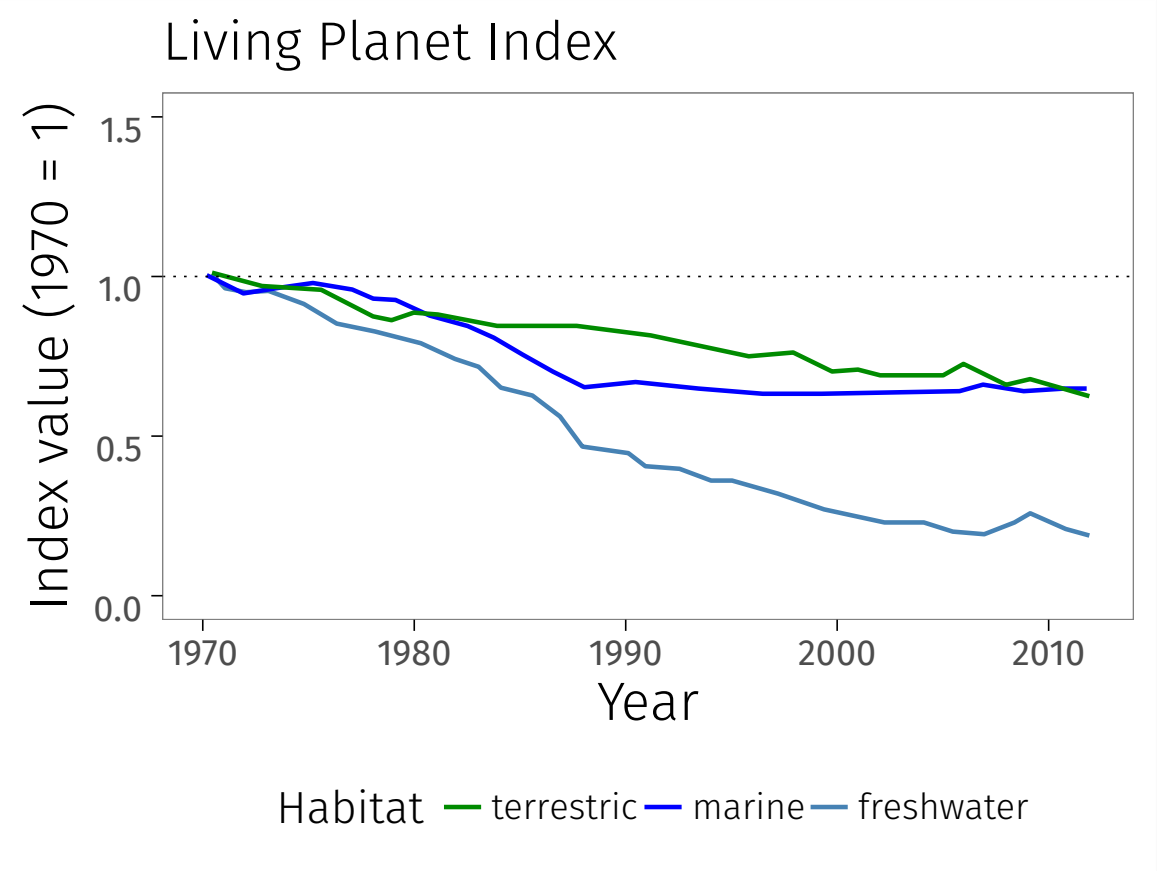


Reasons

- Habitat loss
- Overexploitation
- Pollution
- Invasive species

WWF (2016), Living Planet Report; Vörösmarty (2010). Nature.

Freshwater biodiversity is strongly declining



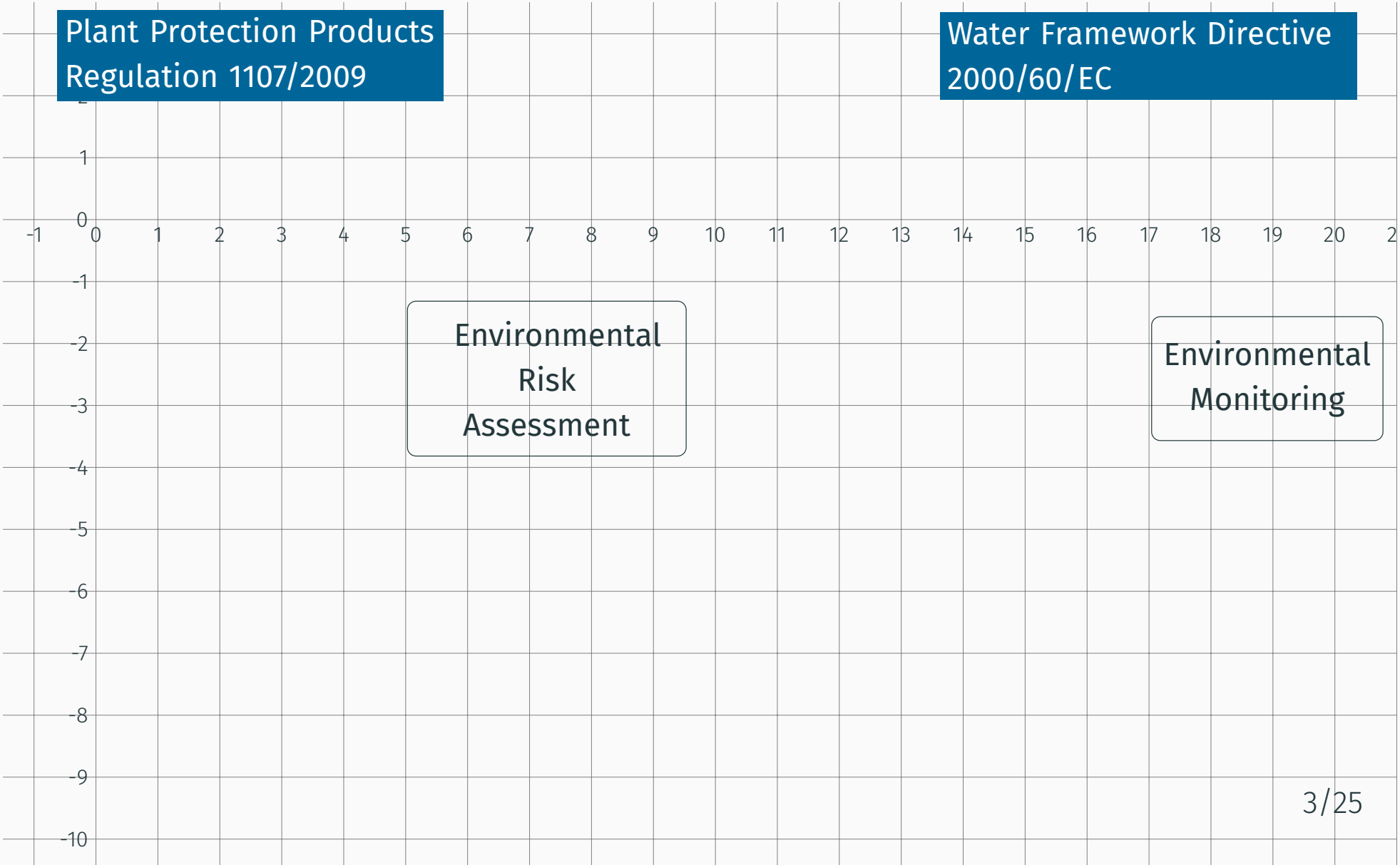
Reasons

- Habitat loss
- Overexploitation
- Pollution
- Invasive species

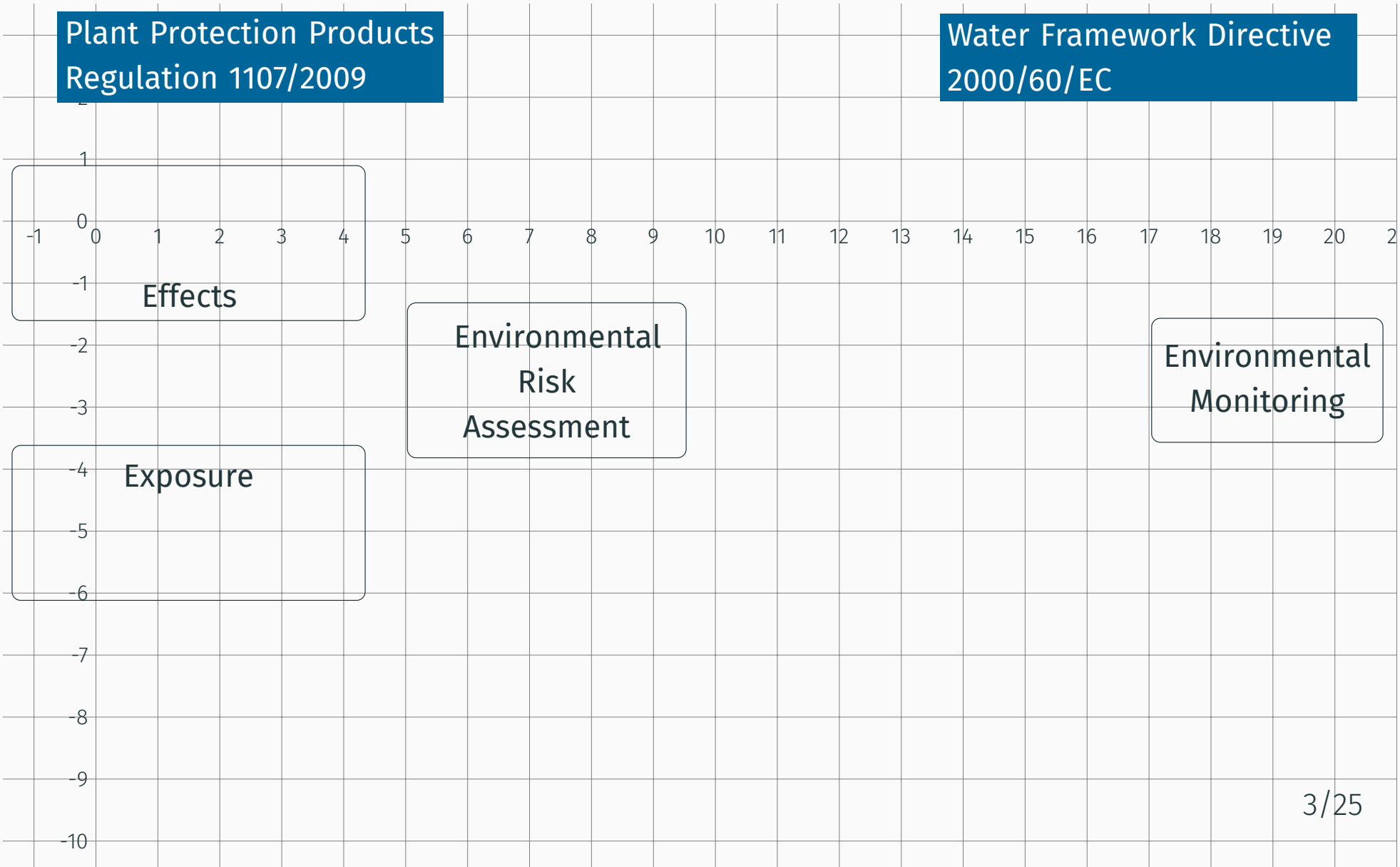
WWF (2016), Living Planet Report; Vörösmarty (2010). Nature.

Environmental Risk Assessment and Environmental Monitoring

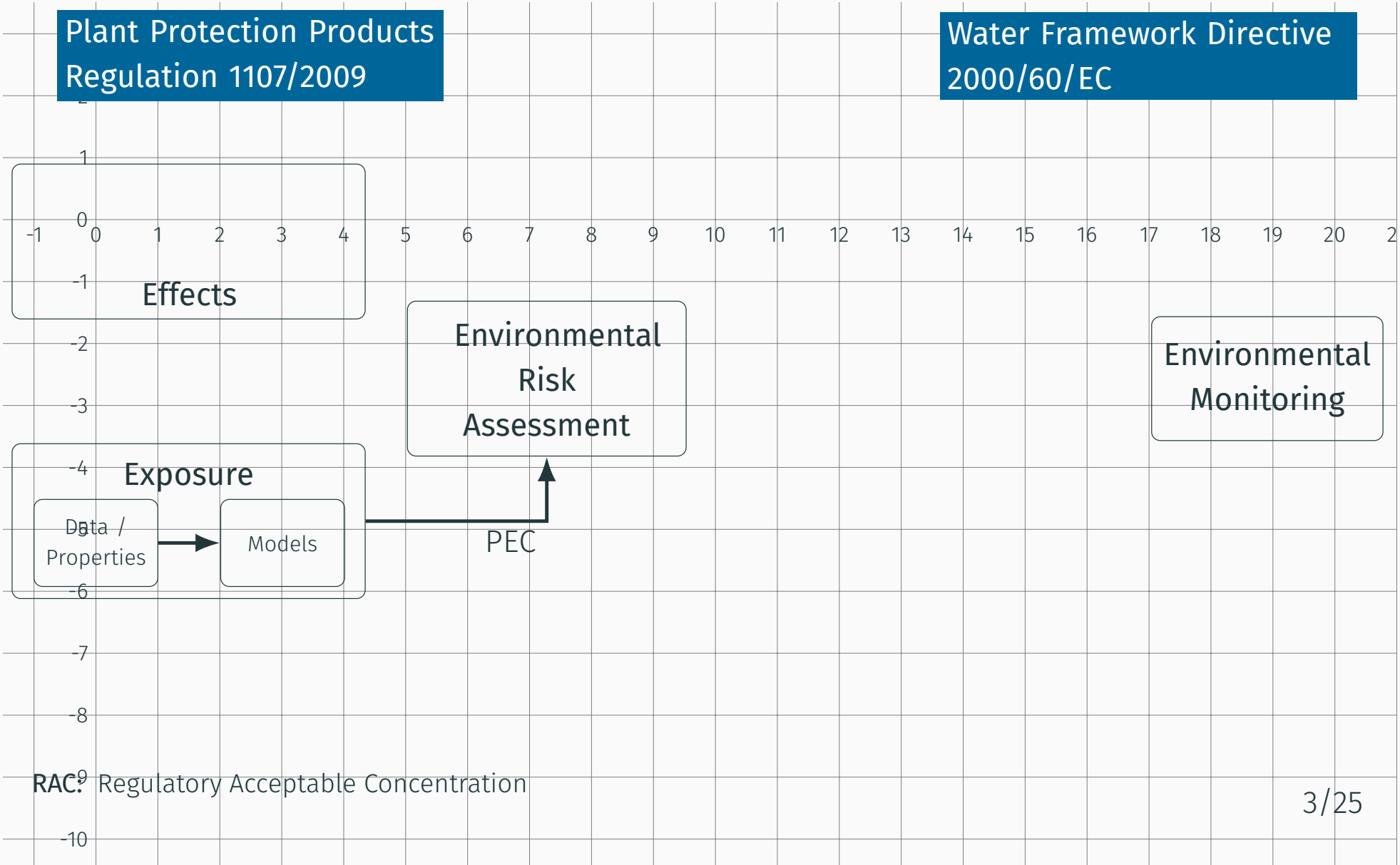
Environmental Risk Assessment and Monitoring



Environmental Risk Assessment and Monitoring



Environmental Risk Assessment and Monitoring

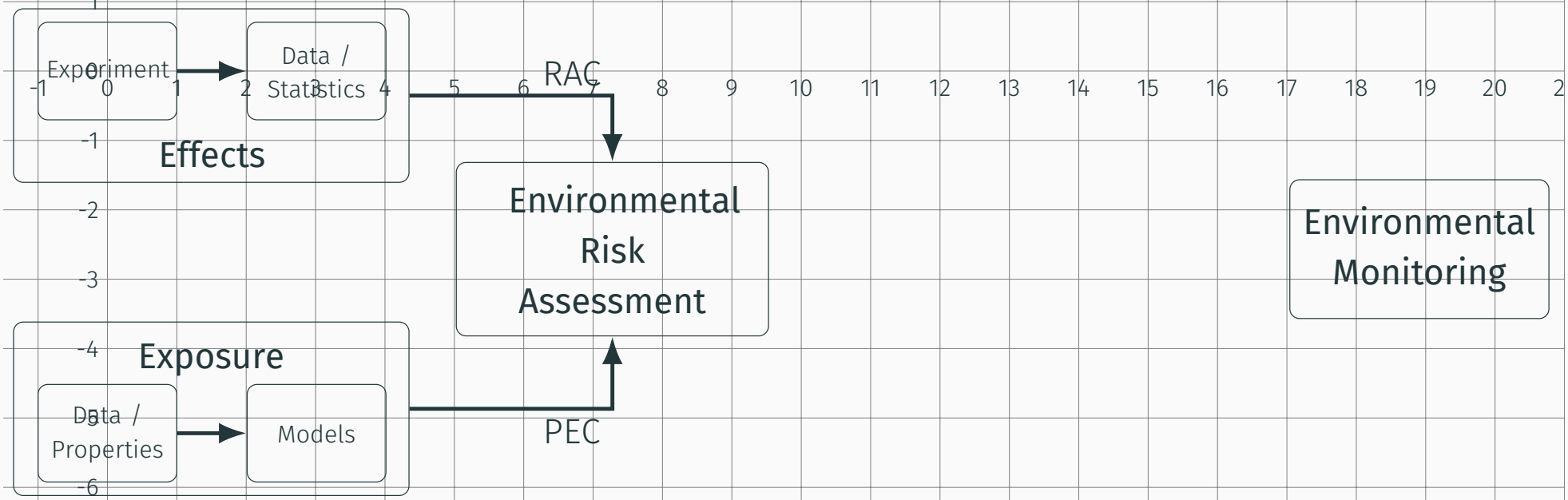


RAC: Regulatory Acceptable Concentration

Environmental Risk Assessment and Monitoring

Plant Protection Products Regulation 1107/2009

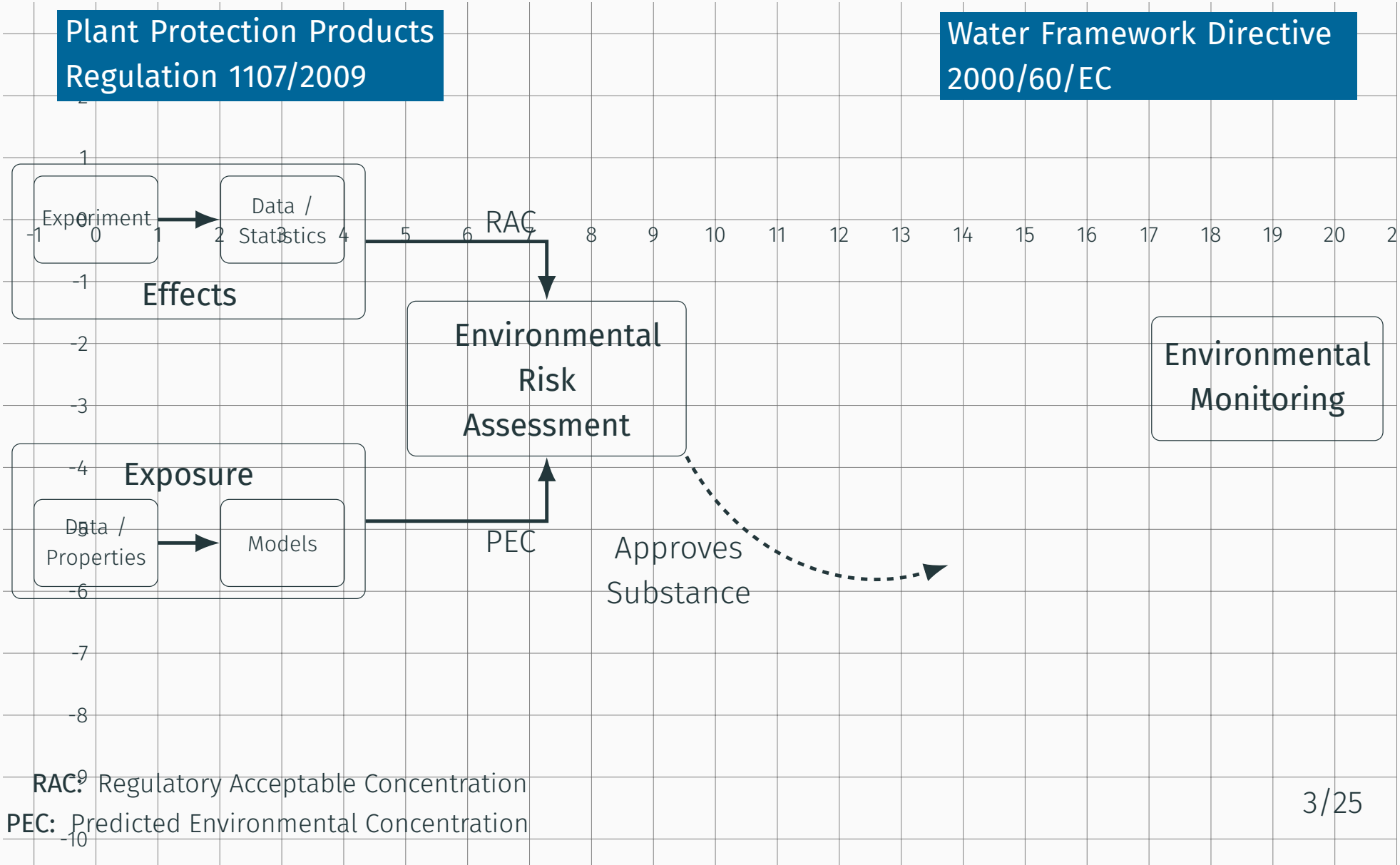
Water Framework Directive 2000/60/EC



RAC: Regulatory Acceptable Concentration

PEC: Predicted Environmental Concentration

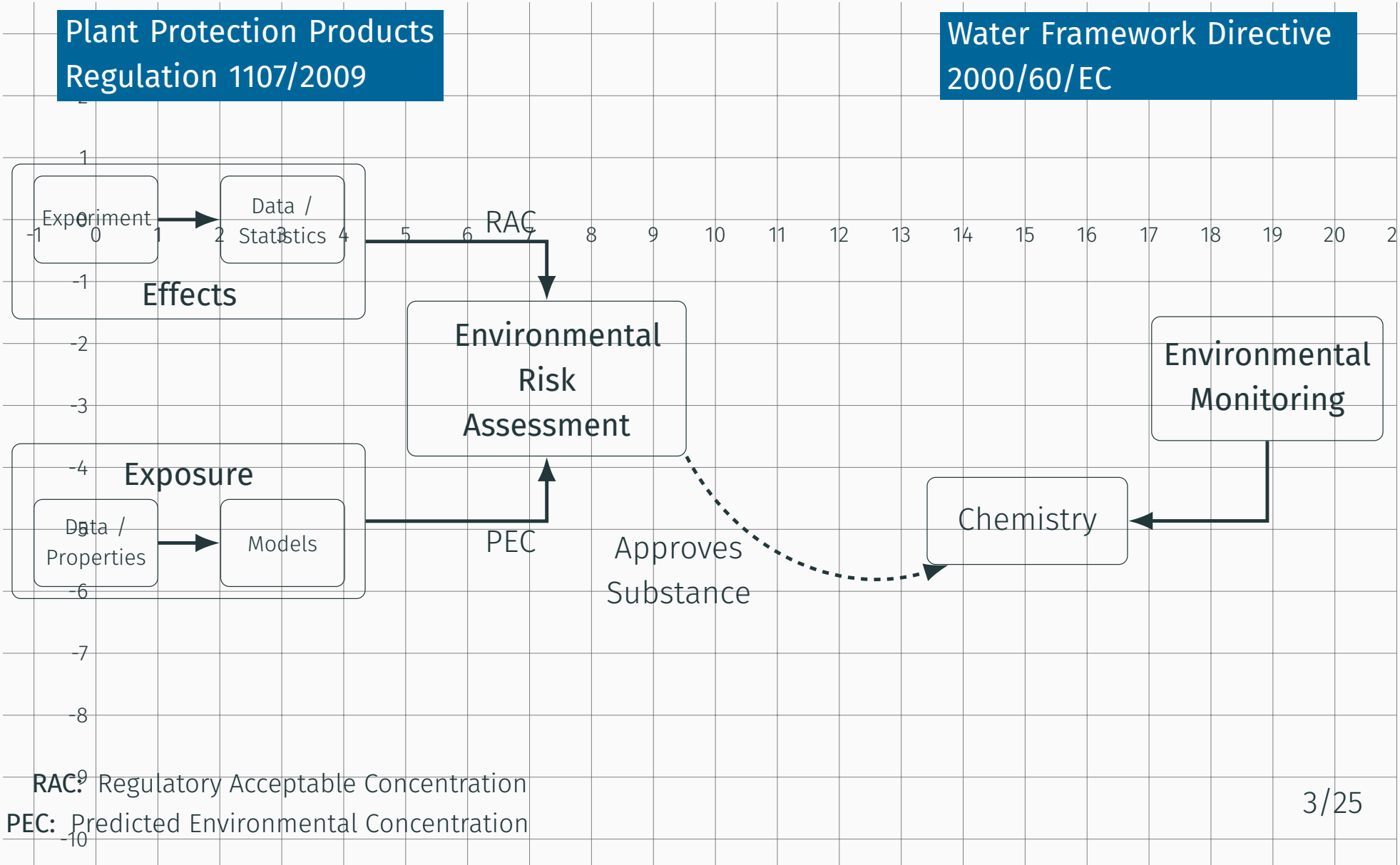
Environmental Risk Assessment and Monitoring



RAC: Regulatory Acceptable Concentration

PEC: Predicted Environmental Concentration

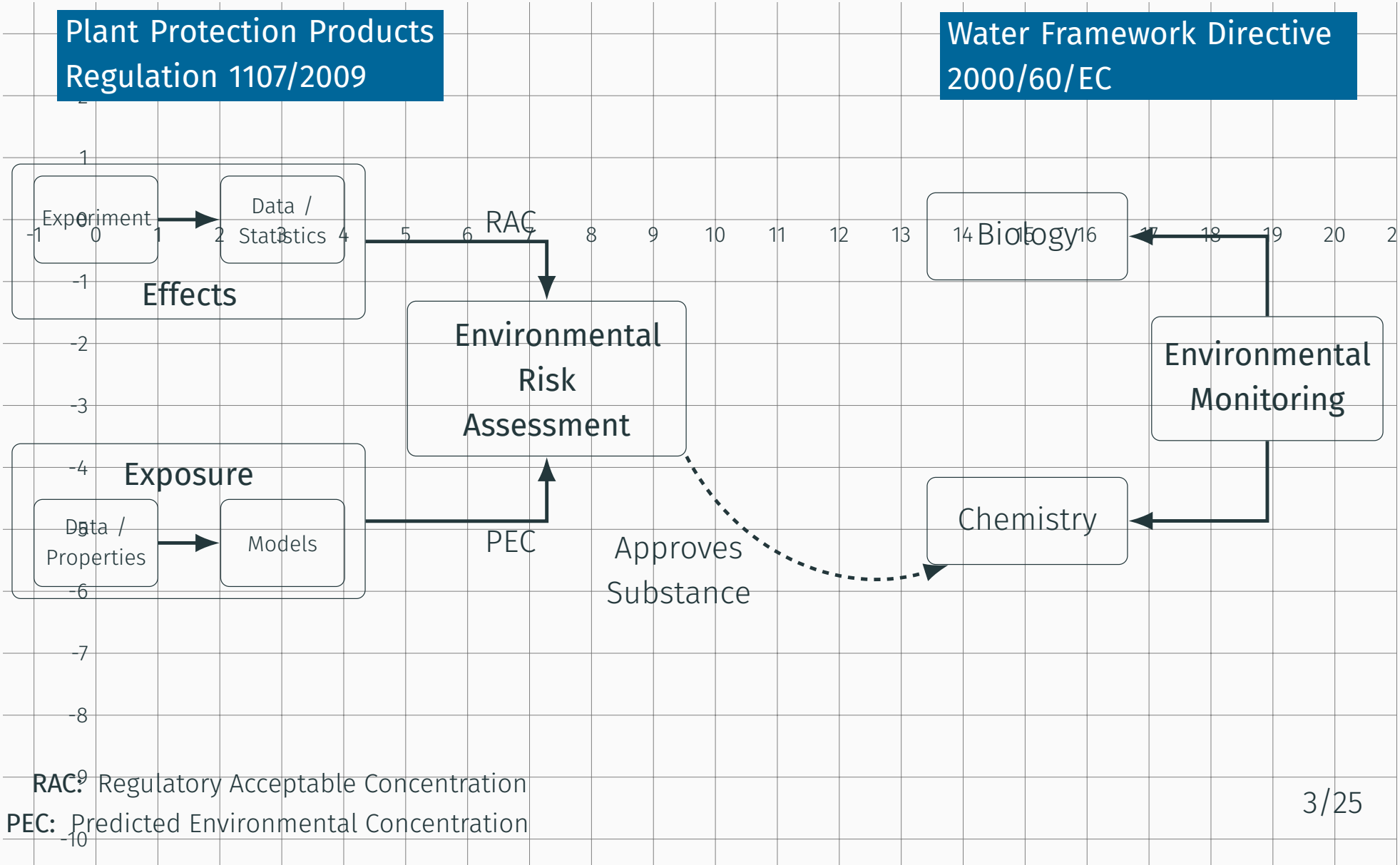
Environmental Risk Assessment and Monitoring



RAC: Regulatory Acceptable Concentration

PEC: Predicted Environmental Concentration

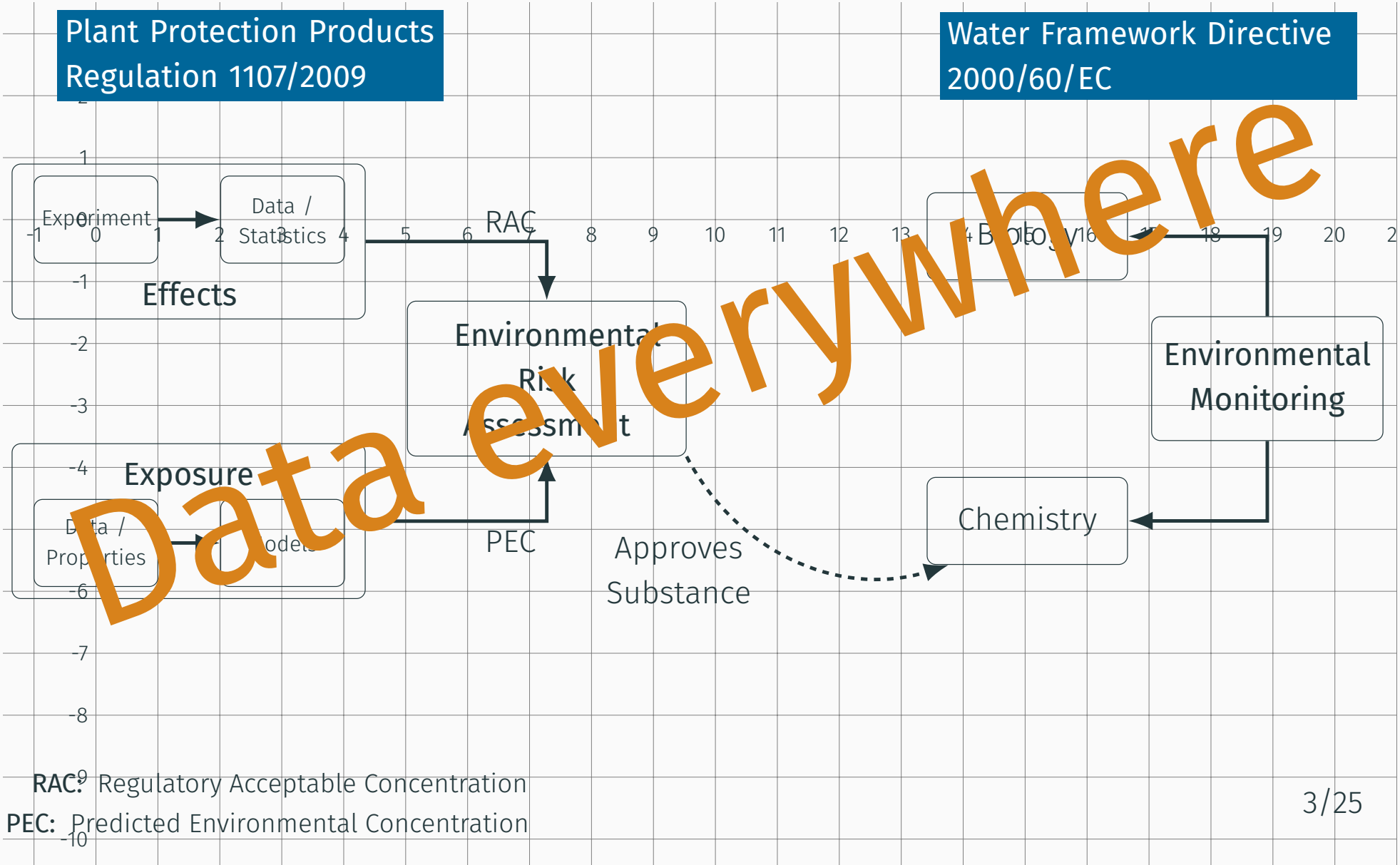
Environmental Risk Assessment and Monitoring



RAC: Regulatory Acceptable Concentration

PEC: Predicted Environmental Concentration

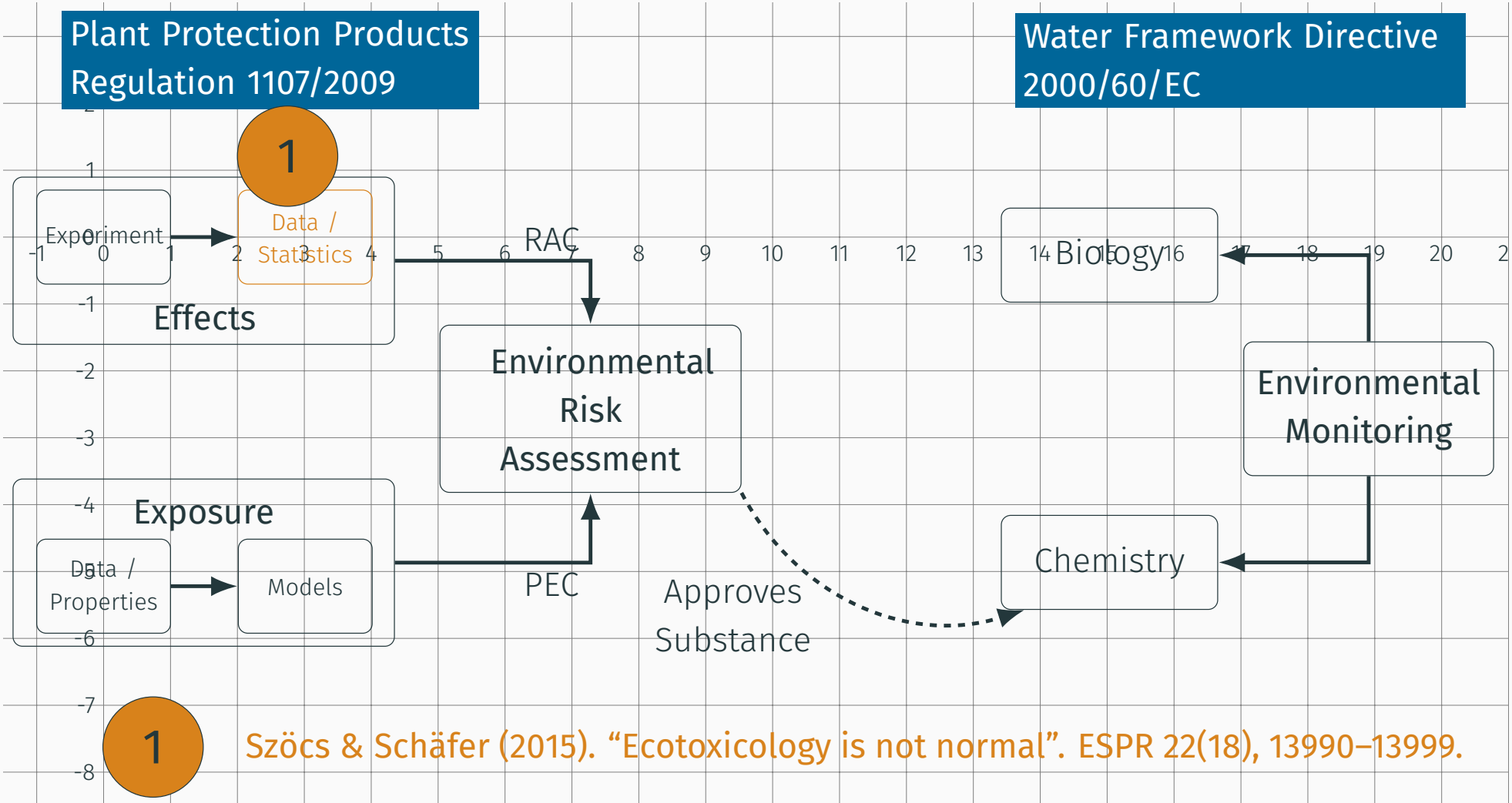
Environmental Risk Assessment and Monitoring



RAC: Regulatory Acceptable Concentration

PEC: Predicted Environmental Concentration

Improving Statistics in ERA

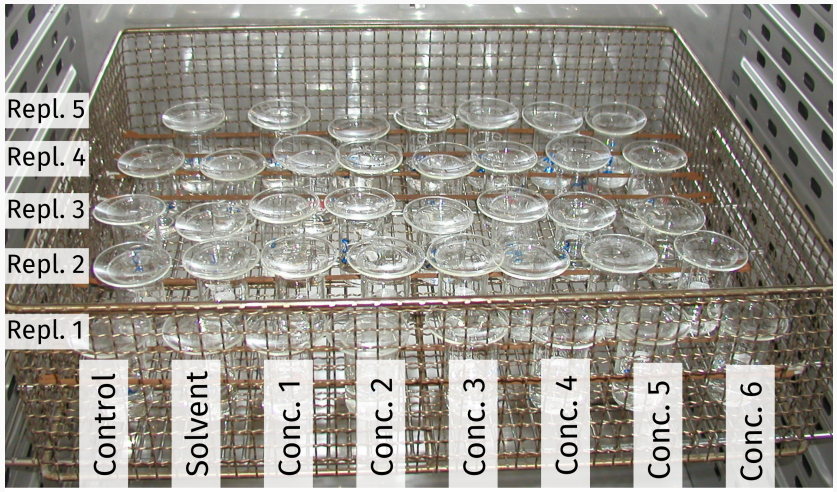


Szöcs & Schäfer (2015). “Ecotoxicology is not normal”. ESPR 22(18), 13990–13999.

RAC: Regulatory Acceptable Concentration
PEC: Predicted Environmental Concentration

Improving Statistics in ERA

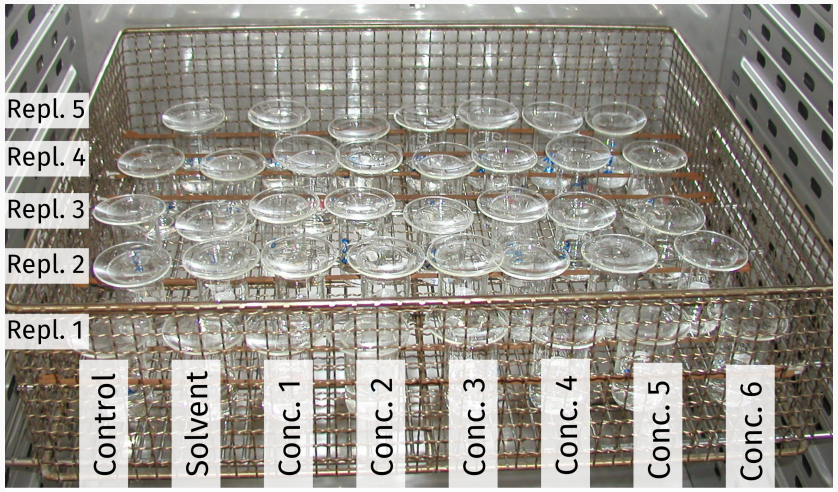
Experiments in Effect Assessment



- Daphnia Test
- *”x out of n survived”*

Source: `www.umweltbundesamt.de` (modified)

Experiments in Effect Assessment

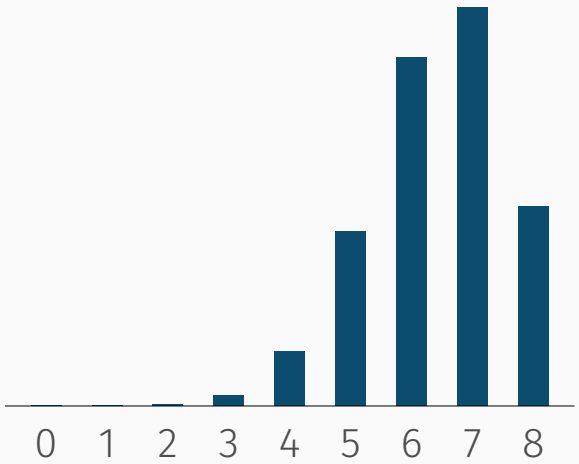


- Daphnia Test
- *"x out of n survived"*

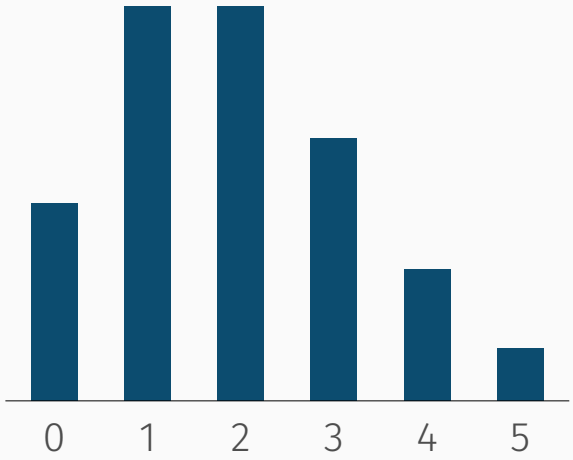
- Mesocosm
- *"number of animals"*

Source: `www.umweltbundesamt.de` (modified)

Ecotoxicology is not normal

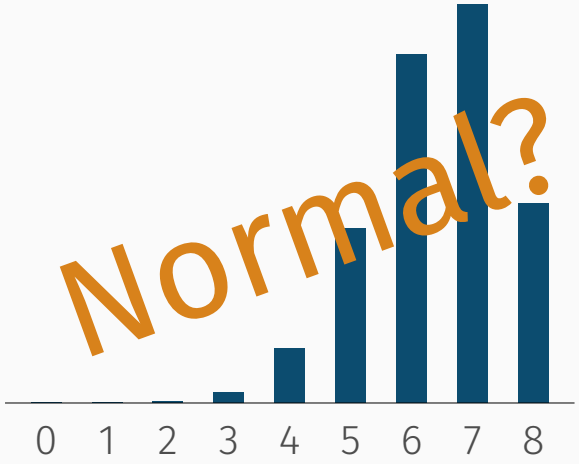


- *”x out of n survived”*

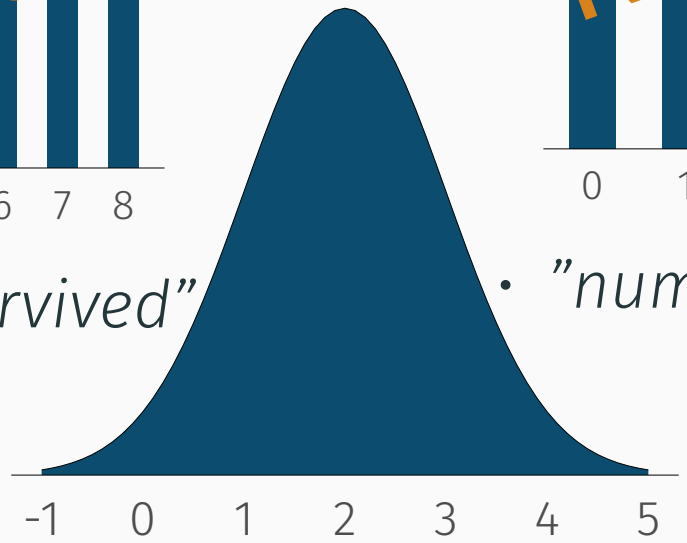


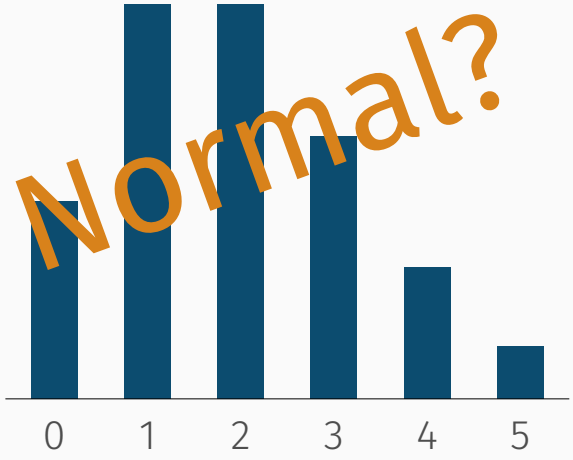
- *”number of animals”*

Ecotoxicology is not normal



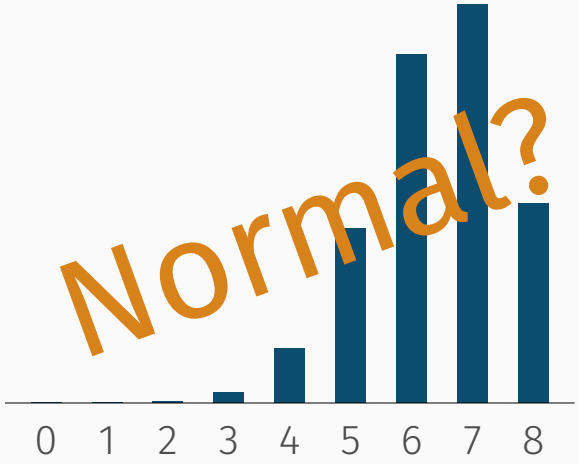
- *"x out of n survived"*



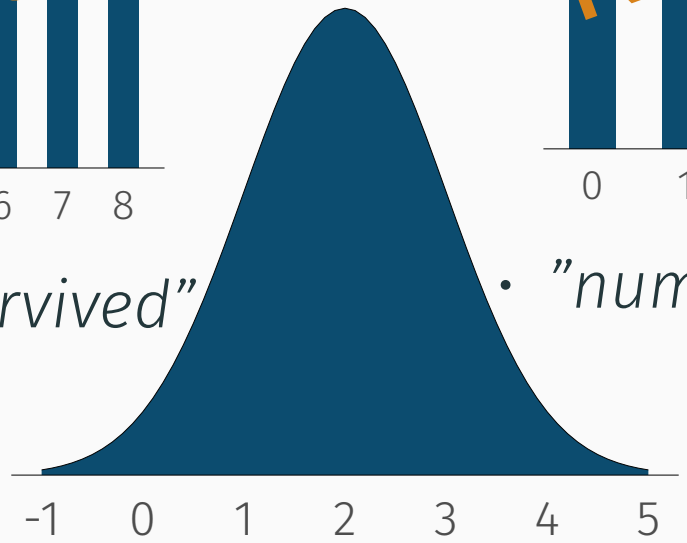


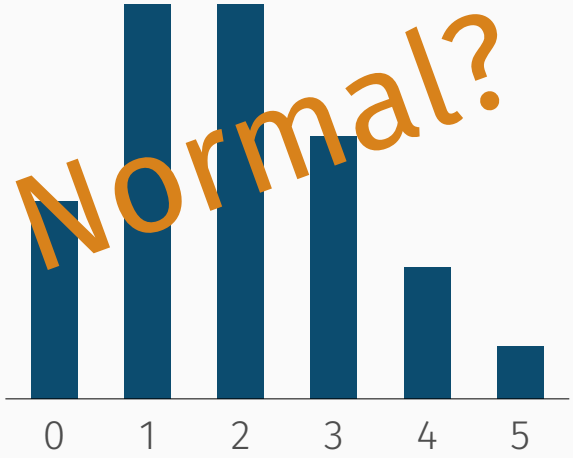
- *"number of animals"*

Ecotoxicology is not normal



- *”x out of n survived”*

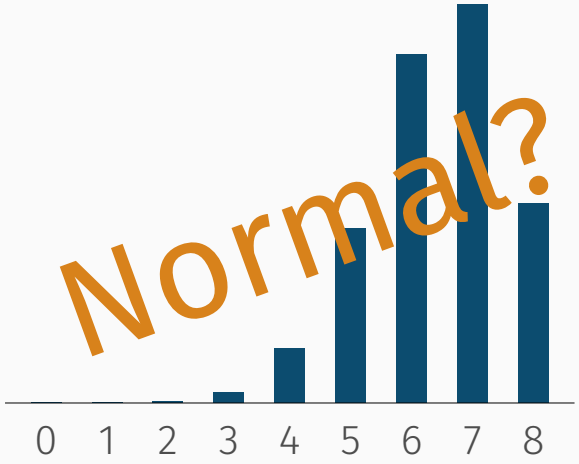




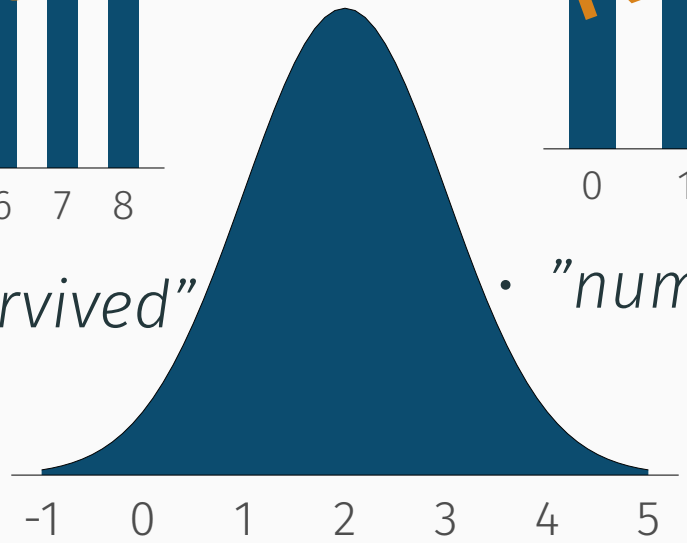
- *”number of animals”*

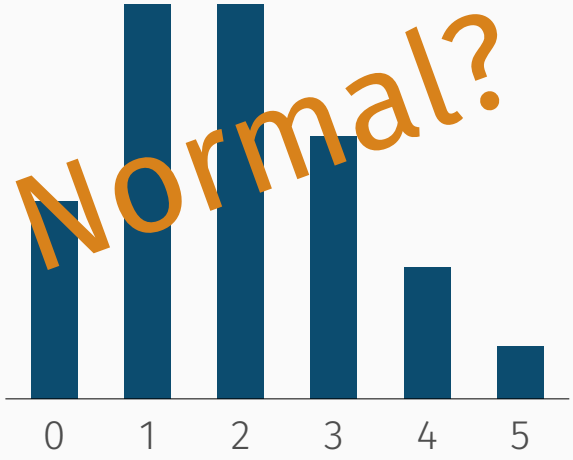
- ignore?

Ecotoxicology is not normal



- *"x out of n survived"*

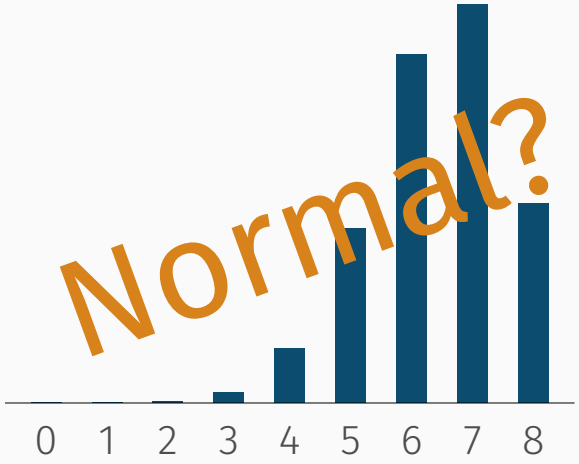




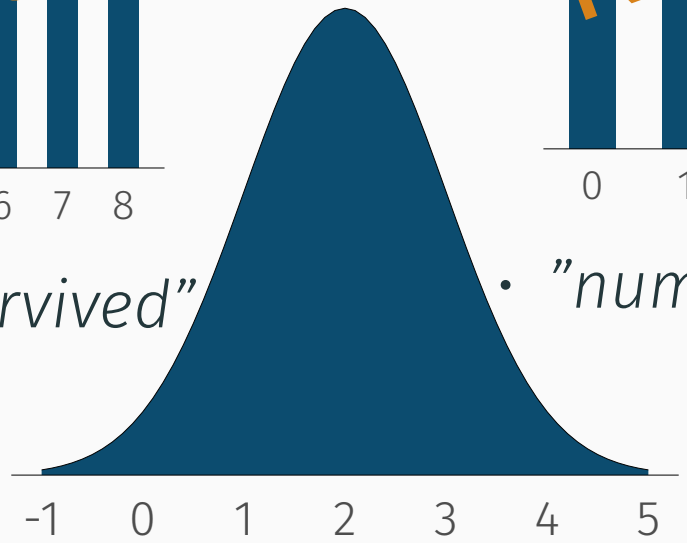
- *"number of animals"*

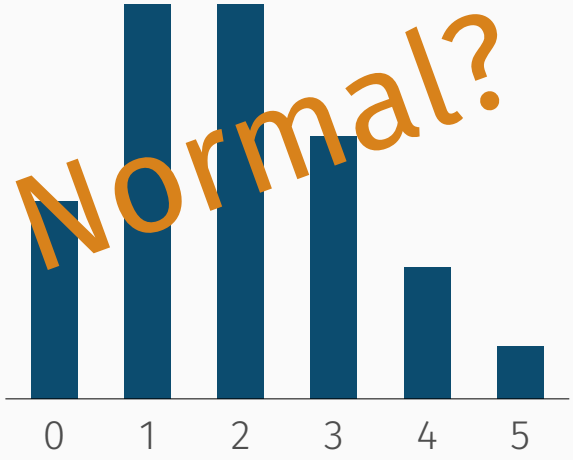
- ignore?
- transform?

Ecotoxicology is not normal



- *"x out of n survived"*

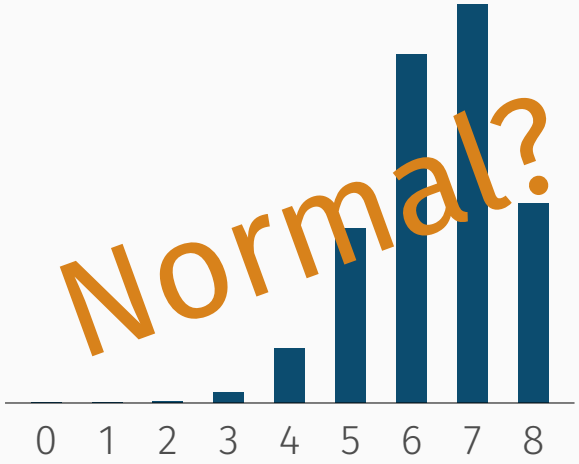




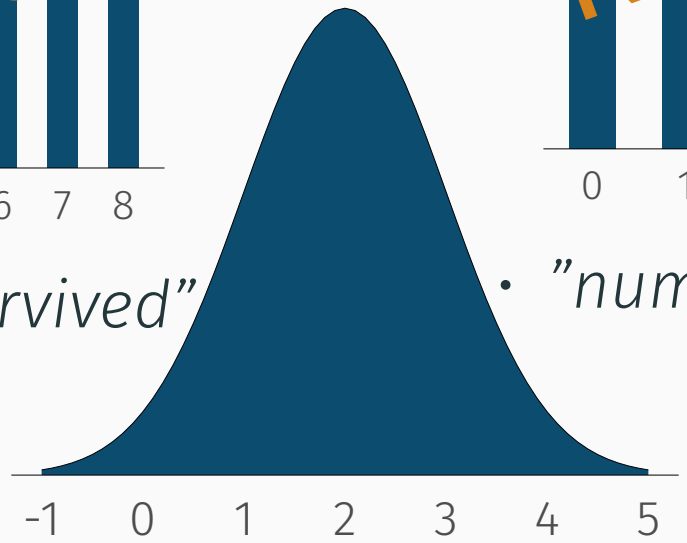
- *"number of animals"*

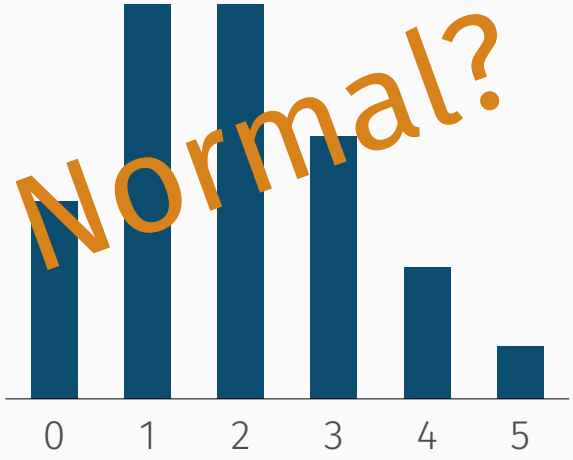
- ignore?
- transform?
- non-parametric?

Ecotoxicology is not normal



- *"x out of n survived"*

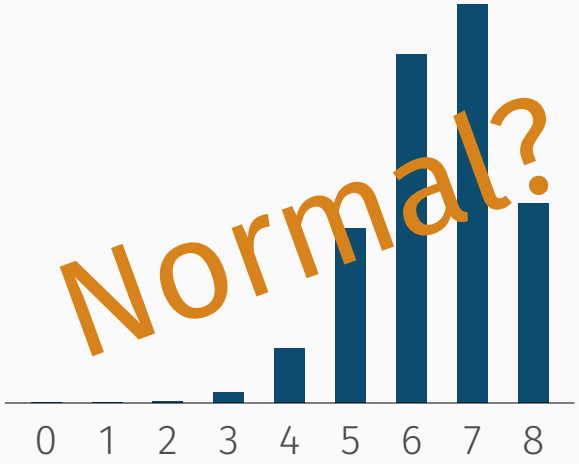




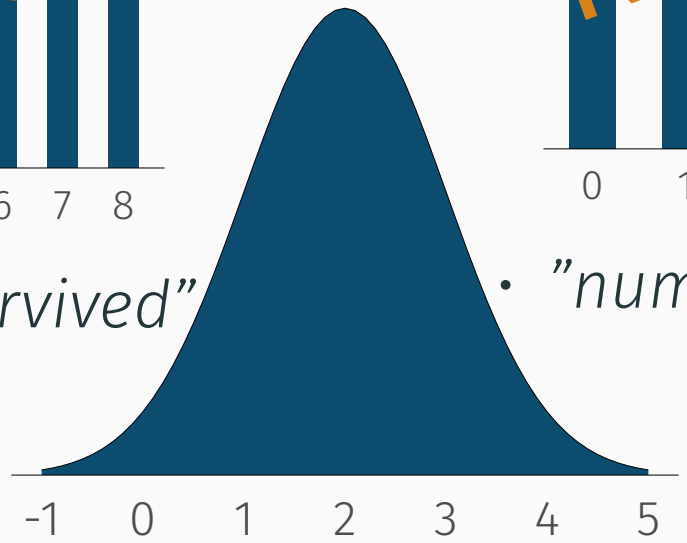
- *"number of animals"*

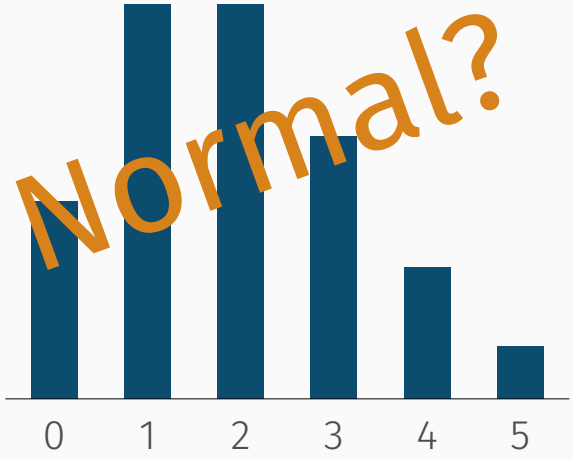
- ignore?
- transform?
- non-parametric?
- Generalised Linear Model (GLM)?

Ecotoxicology is not normal



- *"x out of n survived"*



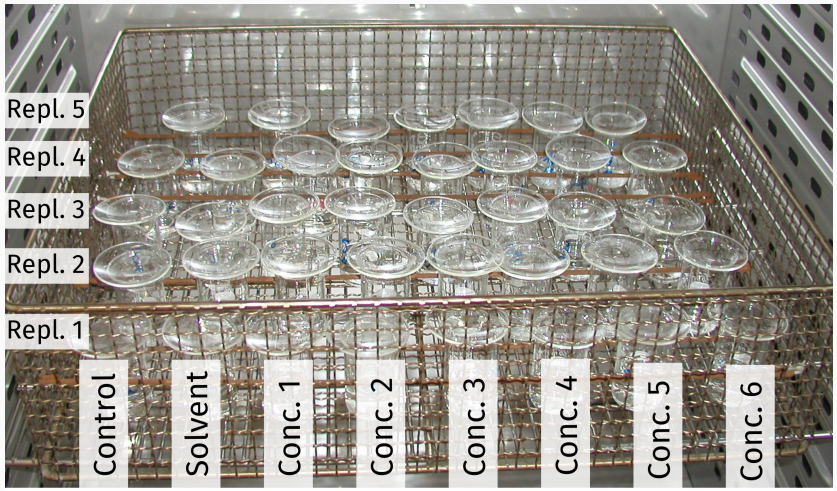


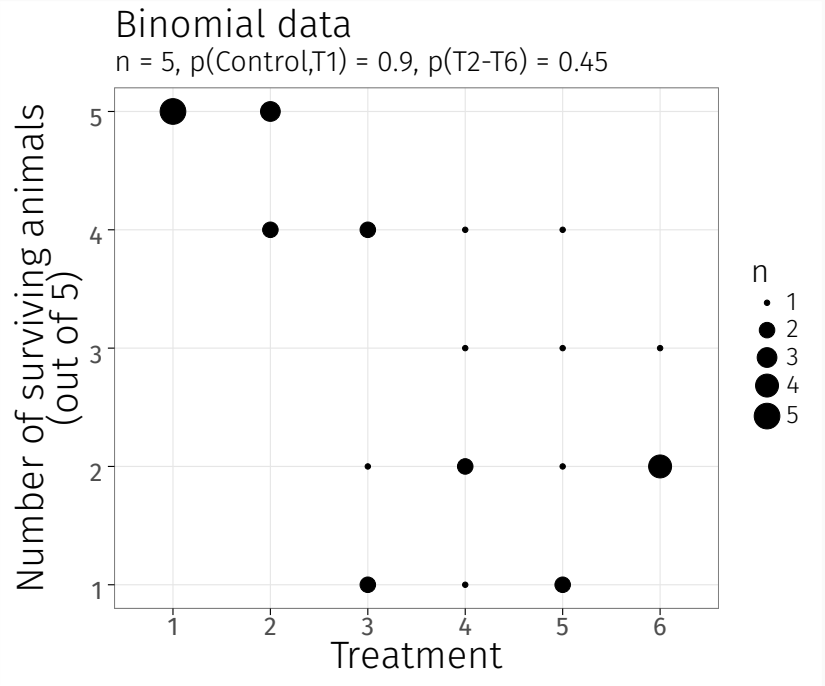
- *"number of animals"*

- ignore?
- transform?
- non-parametric?
- Generalised Linear Model (GLM)?

Discussion on GLM + GLM short intro

A simulation study

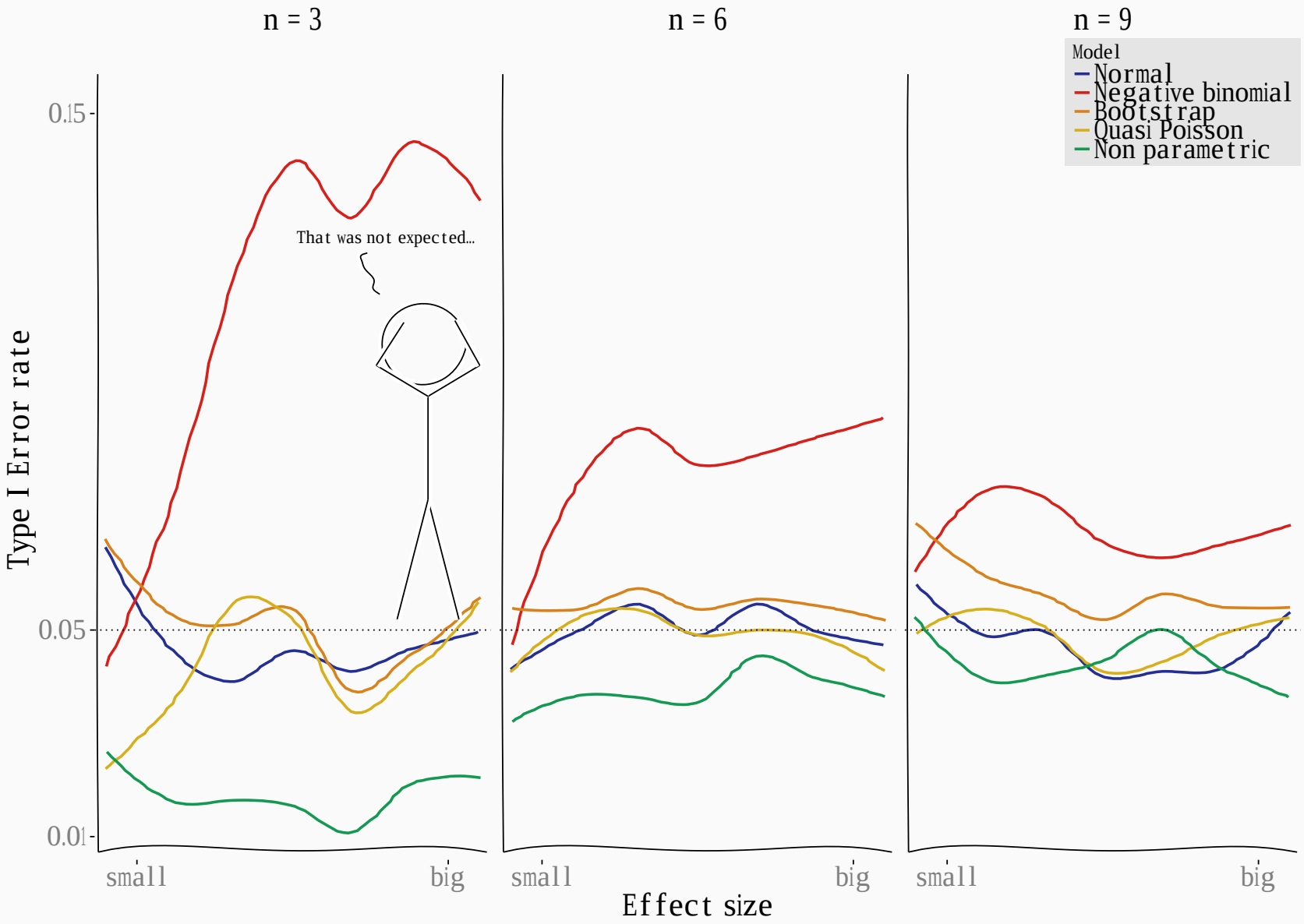




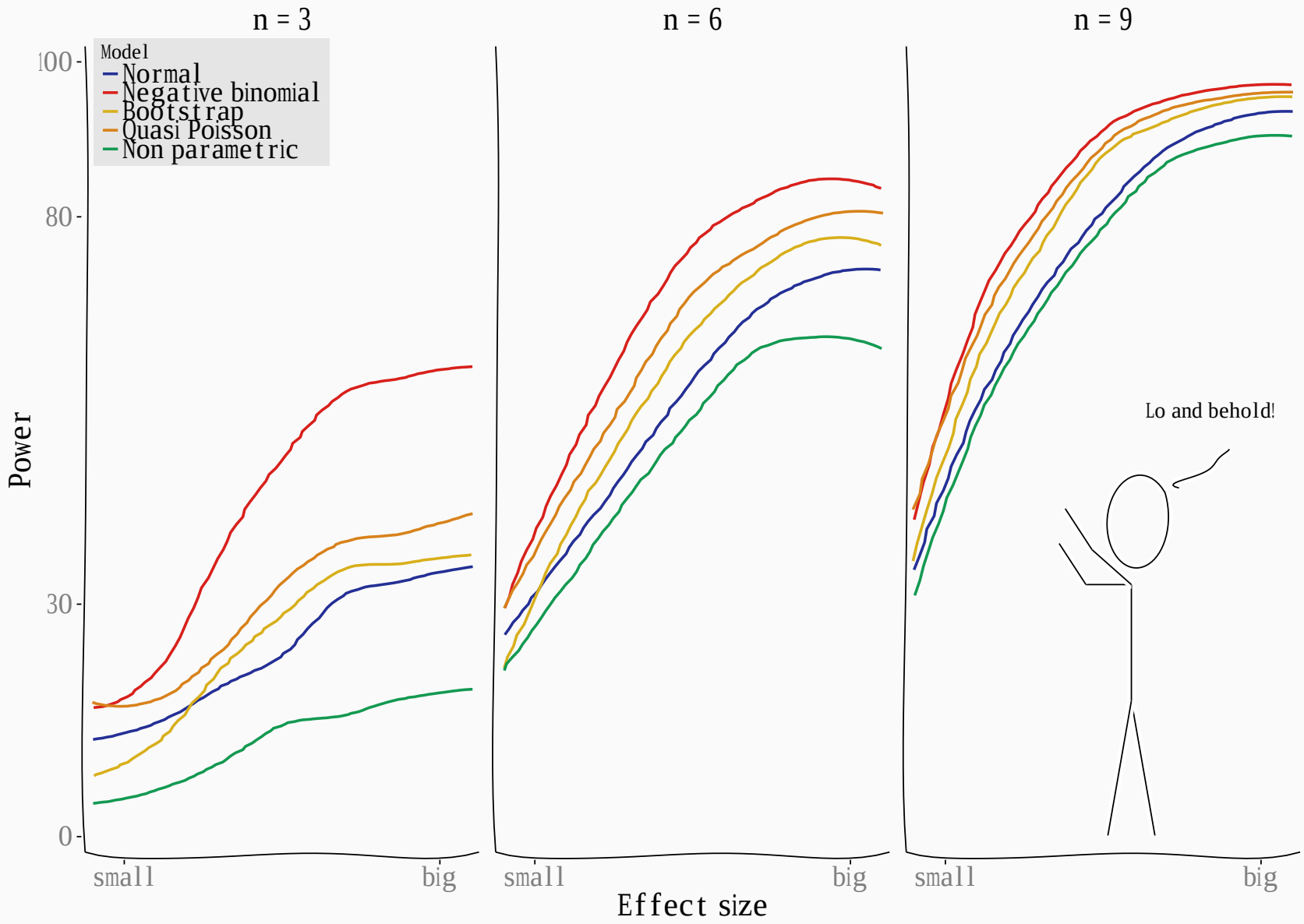
Simulation:

- Count & Binomial data
- Vary replicates & effect sizes
- LM, GLMs, Non-parametric

Type I Errors: GLMs can fail



Power: But GLMs can do also better



What we learned from this study

1. Negative-binomial GLM show increased Type I errors
2. Can be fixed via bootstrap
3. GLMs can increase this power
4. Ecotoxicological experiments commonly low power

Where are we today?

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2016, **7**, 882–890 doi: 10.1111/2041-210X.12552

FORUM

Three points to consider when choosing a LM or GLM test for count data

David I. Warton[1]*, Mitchell Lyons[2], Jakub Stoklosa[1] and Anthony R. Ives[3]

[1]*School of Mathematics and Statistics and Evolution & Ecology Research Centre, University of New South Wales, NSW 2052, Australia;* [2]*School of Biological, Earth and Environmental Sciences, University of New South Wales, NSW 2052, Australia; and* [3]*Department of Zoology, University of Wisconsin-Madison, Madison, WI 53706, USA*

State of the art

Where are we today?

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2016, **7**, 882–890 doi: 10.1111/2041-210X.12552

FORUM

Three points to consider when choosing a LM or GLM test for count data

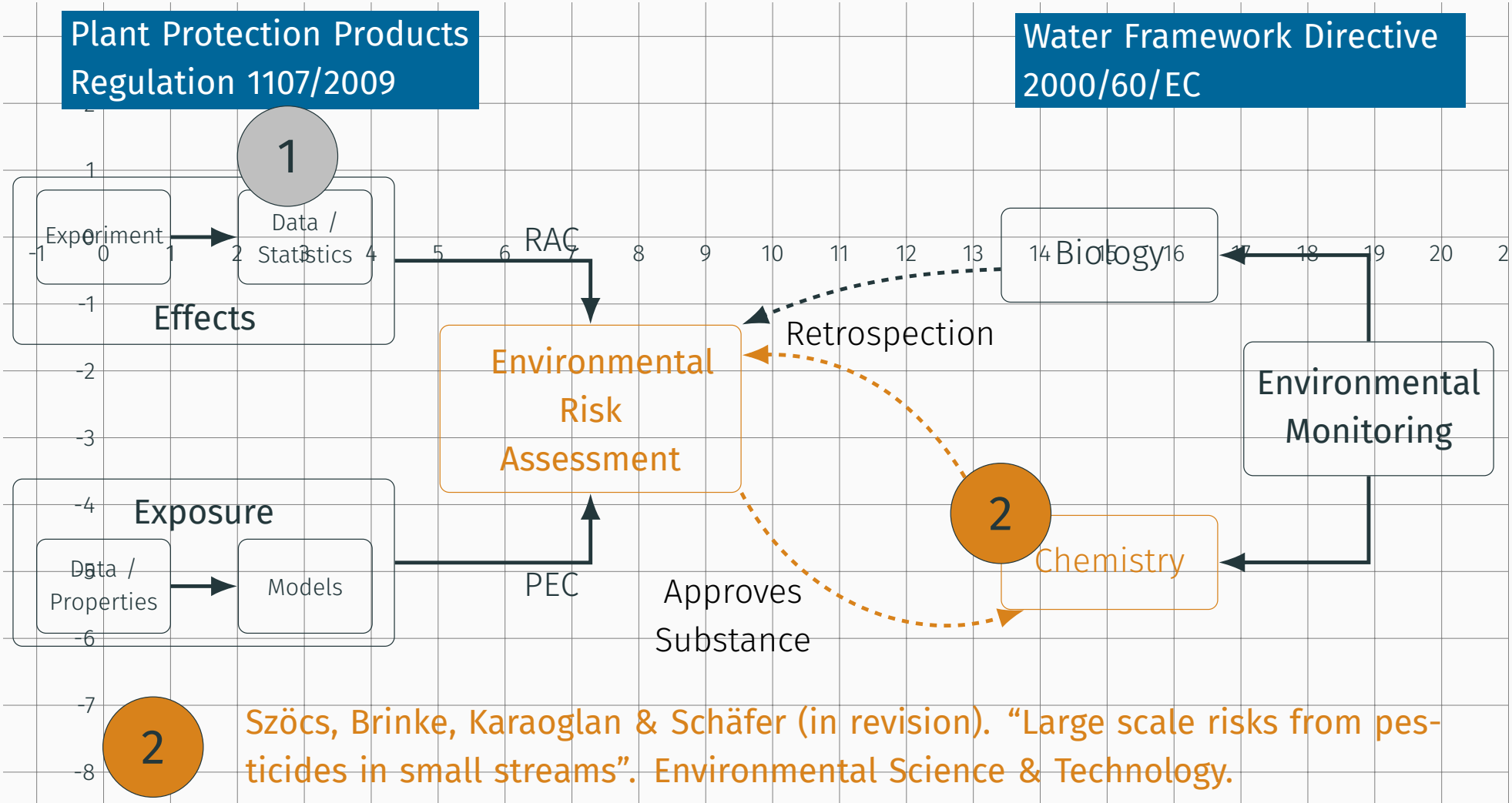
David I. Warton[1]*, Mitchell Lyons[2], Jakub Stoklosa[1] and Anthony R. Ives[3]

[1]*School of Mathematics and Statistics and Evolution & Ecology Research Centre, University of New South Wales, NSW 2052, Australia;* [2]*School of Biological, Earth and Environmental Sciences, University of New South Wales, NSW 2052, Australia; and* [3]*Department of Zoology, University of Wisconsin-Madison, Madison, WI 53706, USA*

State of the art

1. Choose your model based on data properties
2. Fix Type I errors by resampling
3. GLMs have greater power

Exploring Monitoring Data for ERA



Szöcs, Brinke, Karaoglan & Schäfer (in revision). "Large scale risks from pesticides in small streams". Environmental Science & Technology.

RAC: Regulatory Acceptable Concentration

PEC: Predicted Environmental Concentration

Exploring Monitoring Data for ERA

Goals & Hypotheses

Goal: Combine monitoring and ERA

- Compile nation-wide monitoring data
- Focus on small streams
- Identify risks & influencing factors

Goals & Hypotheses

Goal: Combine monitoring and ERA

- Compile nation-wide monitoring data
- Focus on small streams
- Identify risks & influencing factors

Hypotheses

1. Agriculture: ↑ Risk: ↑

Goals & Hypotheses

Goal: Combine monitoring and ERA

- Compile nation-wide monitoring data
- Focus on small streams
- Identify risks & influencing factors

Hypotheses

1. Agriculture: ↑ Risk: ↑
2. Stream size: ↓ Risk: ↑

Goals & Hypotheses

Goal: Combine monitoring and ERA

- Compile nation-wide monitoring data
- Focus on small streams
- Identify risks & influencing factors

Hypotheses

1. Agriculture: ↑ Risk: ↑
2. Stream size: ↓ Risk: ↑
3. Precipitation: ↑ Risk: ↑

Goals & Hypotheses

Goal: Combine monitoring and ERA

- Compile nation-wide monitoring data
- Focus on small streams
- Identify risks & influencing factors

Hypotheses

1. Agriculture: ↑ Risk: ↑
2. Stream size: ↓ Risk: ↑
3. Precipitation: ↑ Risk: ↑
4. Annual dynamics - Summer: ↑

Analysing chemical concentrations

Analysing chemical concentrations

- Concentrations < LOQ (96% of all measurements)

Analysing chemical concentrations

- Concentrations < LOQ (96% of all measurements)
- Hurdle-model:

$$y \sim ZAGA = \begin{cases} Binomial\ GLM & \text{if } y < LOQ \\ Gamma\ GLM & \text{if } y \geq LOQ \end{cases}$$

Analysing chemical concentrations

- Concentrations < LOQ (96% of all measurements)
- Hurdle-model:

$$y \sim ZAGA = \begin{cases} Binomial\ GLM & \text{if } y < LOQ \\ Gamma\ GLM & \text{if } y \geq LOQ \end{cases}$$

- Risk Quotient
  - $RQ = \frac{C}{RAC}$

Analysing chemical concentrations

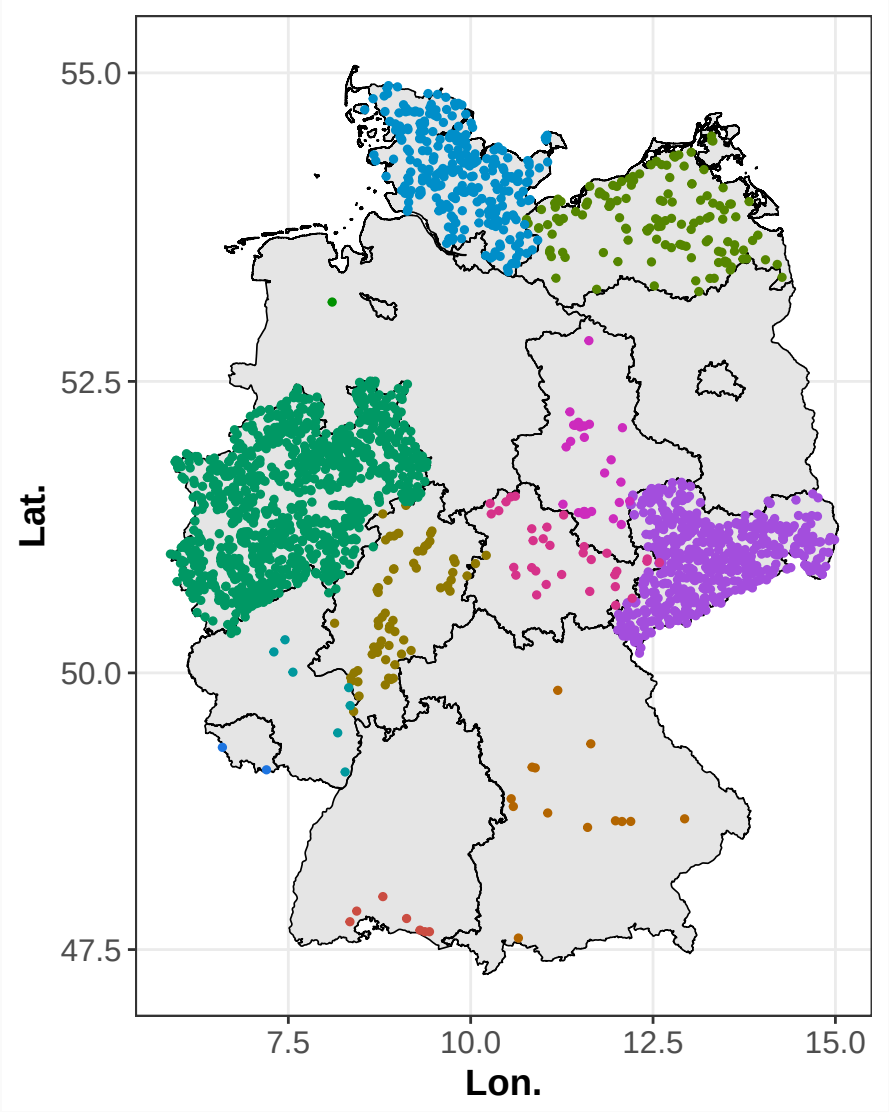
- Concentrations < LOQ (96% of all measurements)
- Hurdle-model:

$$y \sim ZAGA = \begin{cases} Binomial\ GLM & \text{if } y < LOQ \\ Gamma\ GLM & \text{if } y \geq LOQ \end{cases}$$

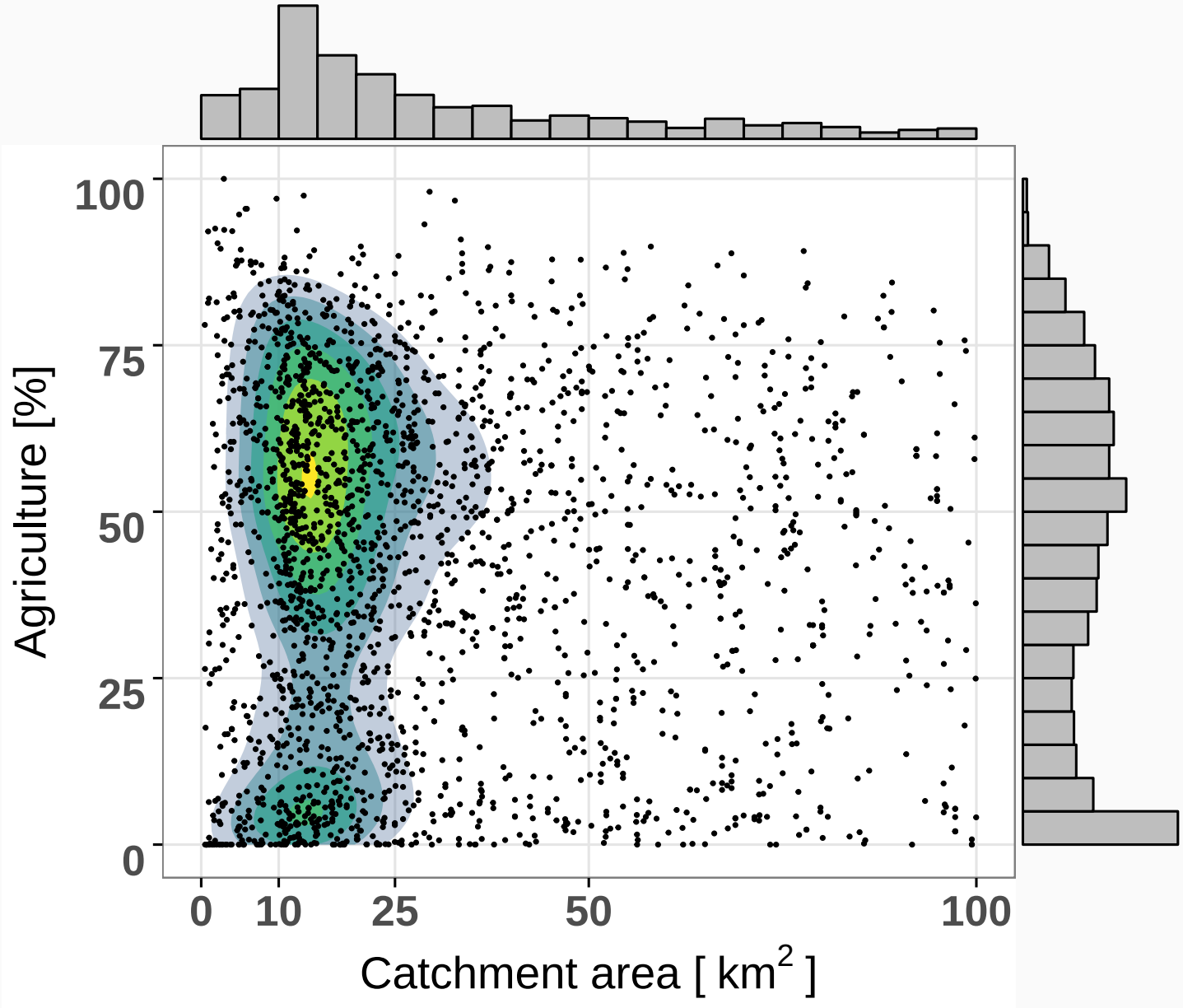
- Risk Quotient
  - $RQ = \frac{C}{RAC}$
- Predictors
  - Catchment size
  - Agricultural land use
  - Precipitation

Compiled data: Big, but inhomogeneous

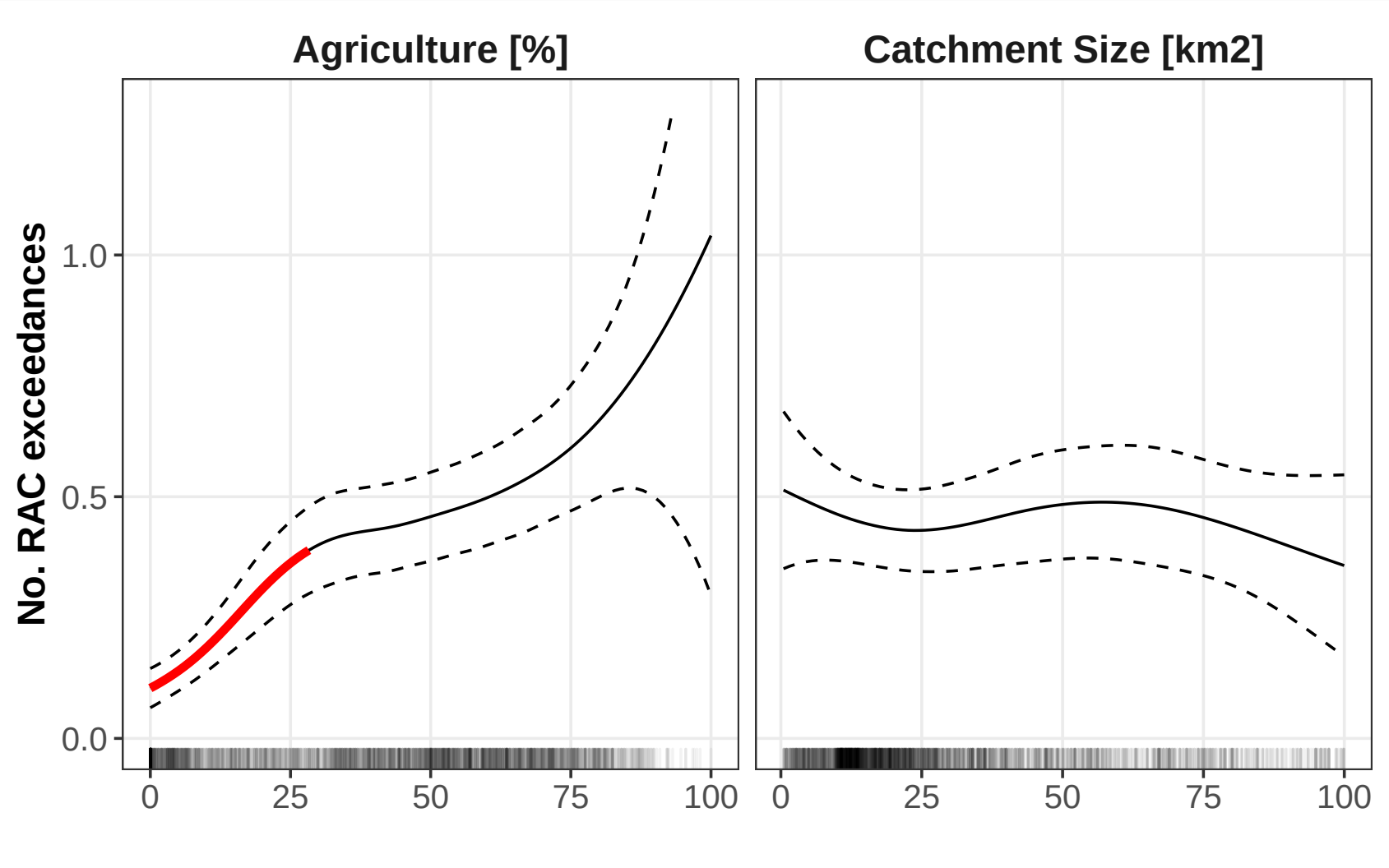
- ~ 1.8M measurements
- ~ 500 pesticides
- ~ 25,000 samples
- ~ 2,300 sites



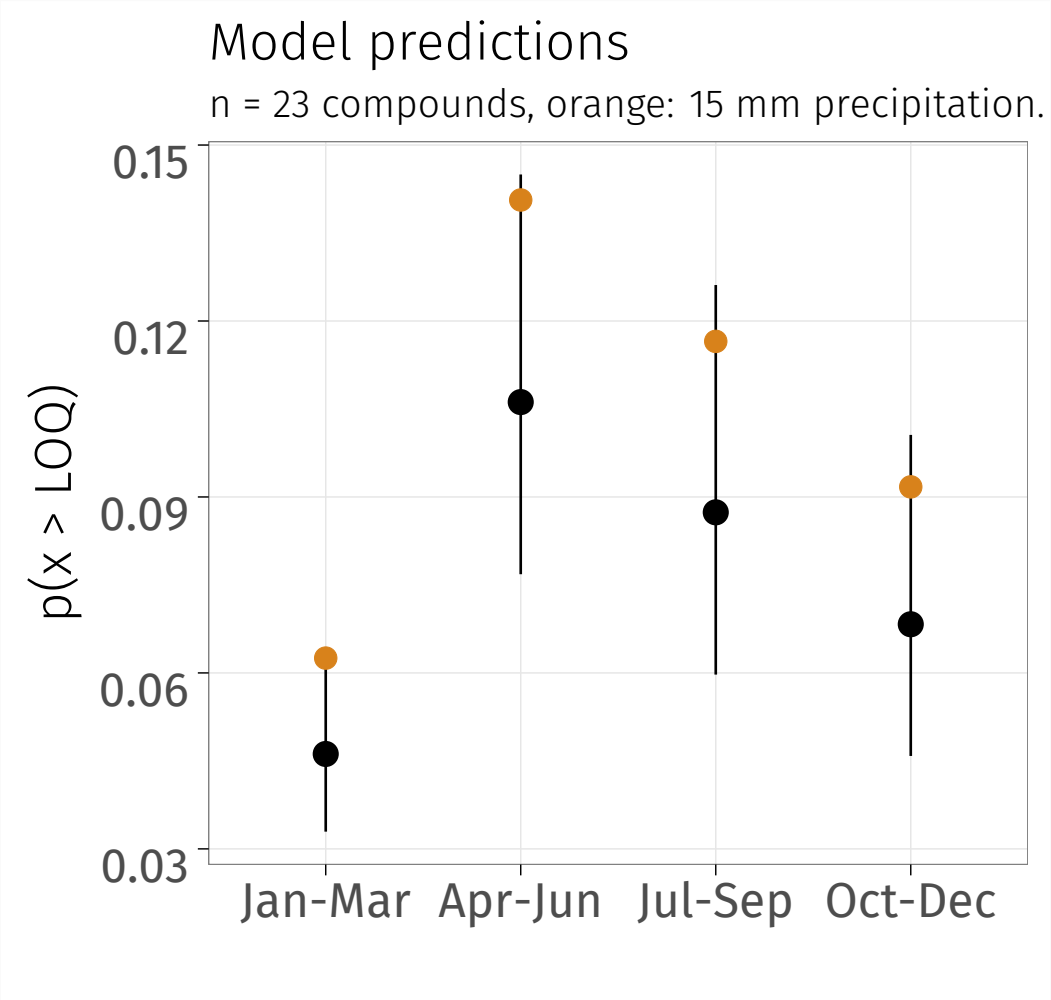
Monitoring: Small streams are underrepresented



Landscape: Factors influencing risk



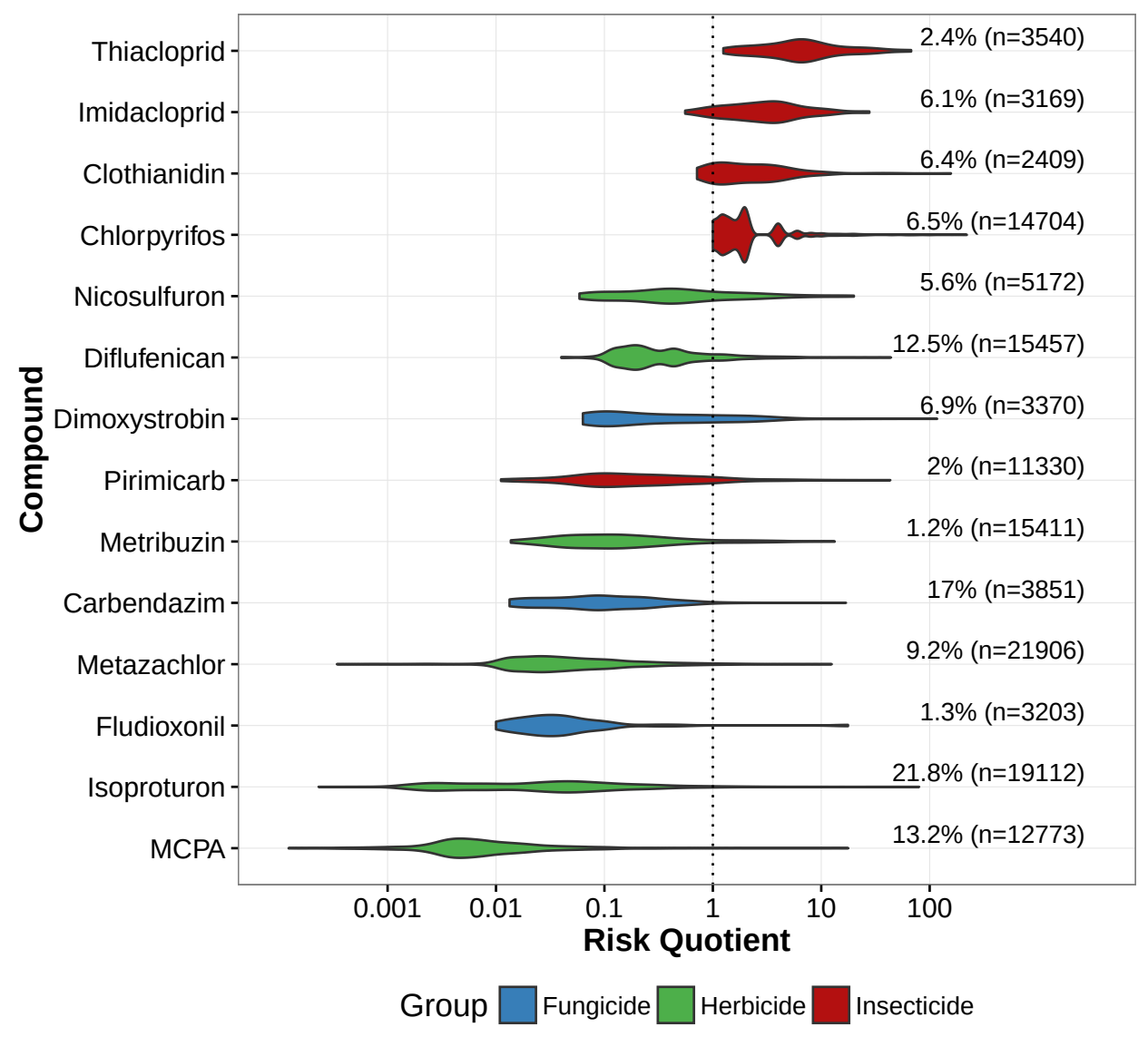
Sampling: Factors influencing risk



- Peak in summer
- Increase by precipitation
- absolute concentrations >>variability

Risks: Compounds exceeding risk thresholds

- 25% of sites with at least one concentration > RAC

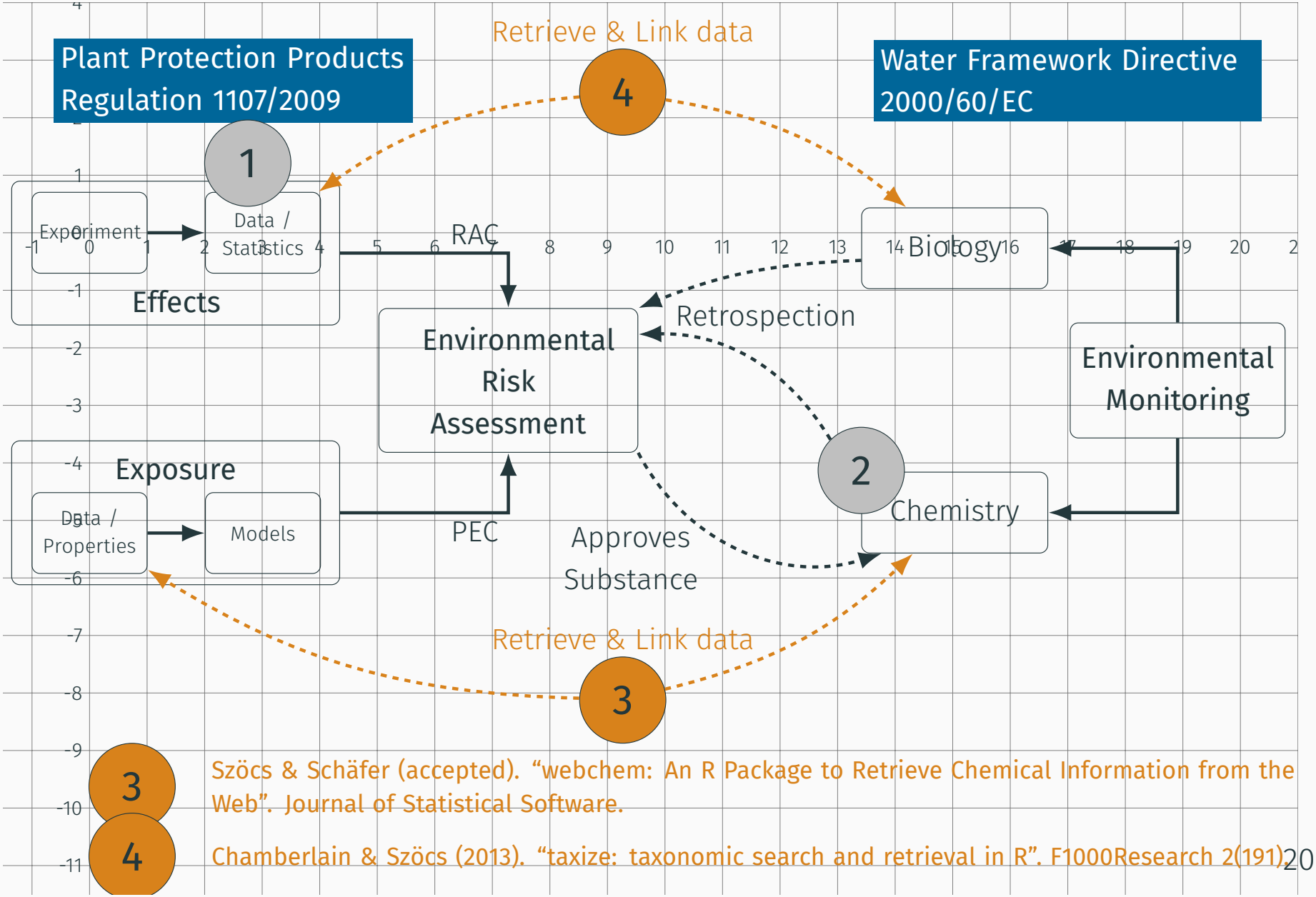


What we learned from this study

1. Differences between states
2. Small streams are underrepresented
3. Agricultural sources
4. LOQ gives additional insights
    - Annual dynamics
    - Precipitation increases concentrations
5. Neonicotinoids

Solutions for Data Handling in ERA

Solutions for Data Handling in ERA



Szöcs & Schäfer (accepted). "webchem: An R Package to Retrieve Chemical Information from the Web". Journal of Statistical Software.

Chamberlain & Szöcs (2013). "taxize: taxonomic search and retrieval in R". F1000Research 2(191).

Biologists & Chemists face the same problems

Biologists & Chemists face the same problems

Names

Osmia rufa, Osmia bicornis,
Osmia ruffa, Osmia xxxxx

Chlorpyrifos, Chlorpyriphos,
Chlorphyrifos, Chlorpypifot

Biologists & Chemists face the same problems

Names

Osmia rufa, Osmia bicornis,
Osmia ruffa, Osmia xxxxx

Chlorpyrifos, Chlorpyriphos,
Chlorphyrifos, Chlorpypifot

Hierarchies

Hymenoptera/ Apoidea/
Megachilidae/ Osmia/ rufa

organophospate, ester,
insecticide

Biologists & Chemists face the same problems

Names

Osmia rufa, Osmia bicornis,
Osmia ruffa, Osmia xxxxx

Chlorpyrifos, Chlorpyriphos,
Chlorphyrifos, Chlorpypifot

Hierarchies

Hymenoptera/ Apoidea/
Megachilidae/ Osmia/ rufa

organophospate, ester,
insecticide

Traits / Properties

Wing length, Mass, Season

Mass, K_{OW}, LC_{50}

Biologists & Chemists face the same problems

Names

Osmia rufa, Osmia bicornis,
Osmia ruffa, Osmia xxxxx

Chlorpyrifos, Chlorpyriphos,
Chlorphyrifos, Chlorpypifot

Hierarchies

Hymenoptera/ Apoidea/
Megachilidae/ Osmia/ rufa

organophospate, ester,
insecticide

Traits / Properties

Wing length, Mass, Season

Mass, K_{OW}, LC_{50}

Identifiers

NCBI, ITIS, EOL, ...

2921-88-2, SMILES, InChI, ...

Biologists & Chemists face the same problems

Names

Osmia rufa, Osmia bicornis,
Osmia ruffa, Osmia xxxxx

Chlorpyrifos, Chlorpyriphos,
Chlorphyrifos, Chlorpypifot

Hierarchies

Hymenoptera/ Apoidea/
Megachilidae/ Osmia/ rufa

organophospate, ester,
insecticide

Traits / Properties

Wing length, Mass, Season

Mass, K_{OW}, LC_{50}

Identifiers

NCBI, ITIS, EOL, ...

2921-88-2, SMILES, InChI, ...

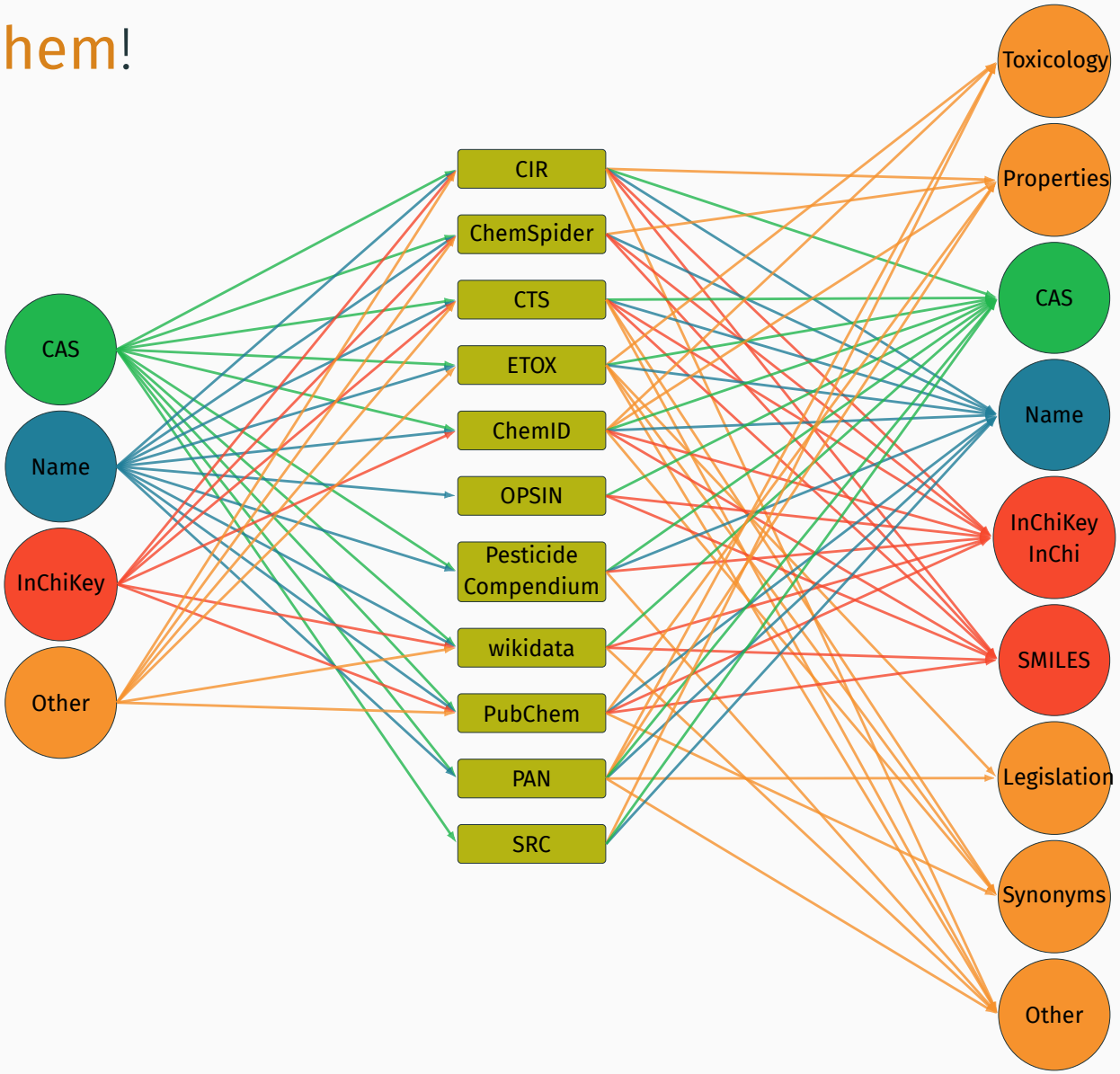
Amount of data

2993 taxa

478 pesticides

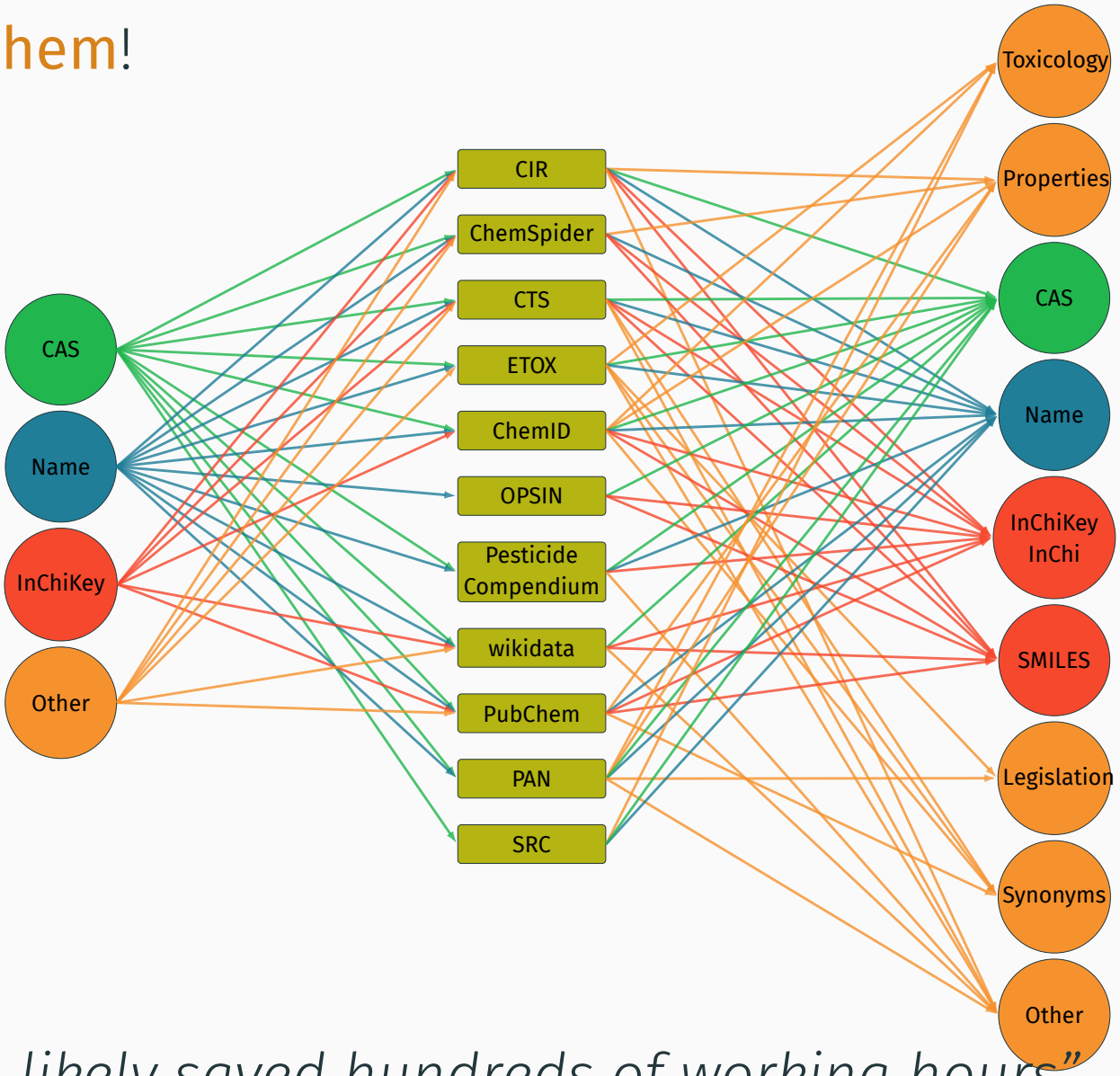
Instead of wasting time...

... use webchem!



Instead of wasting time...

... use webchem!



"webchem ...likely saved hundreds of working hours"

Instead of wasting time...

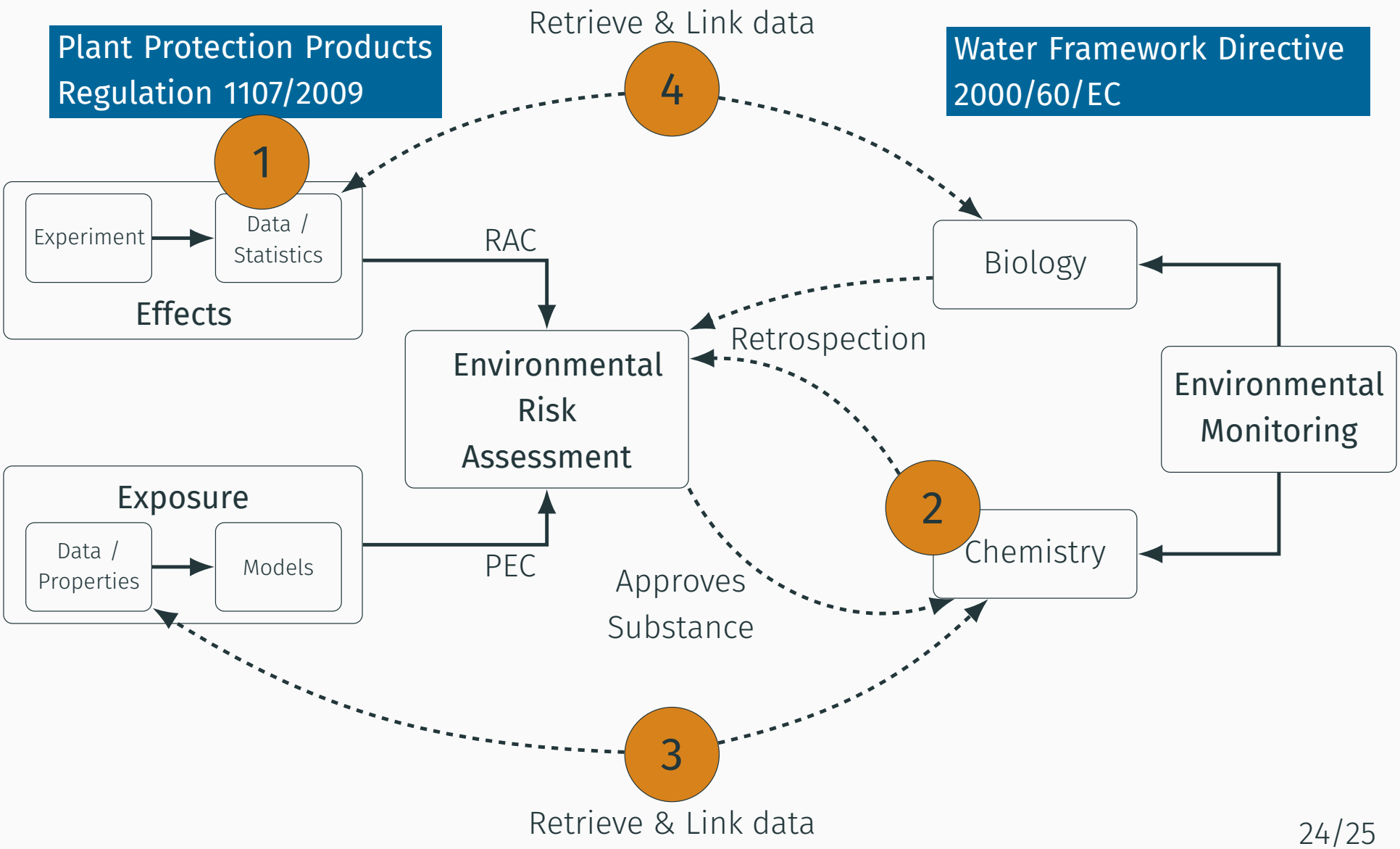
... use taxize!

Instead of wasting time...

... use taxize!

Recap

Recap: What did I look at?



Recap: What we learned from my PhD Thesis

✓Improving Statistics in ERA

- Change your model, not your data
- Take LOQ into account

Recap: What we learned from my PhD Thesis

✓Improving Statistics in ERA

- Change your model, not your data
- Take LOQ into account

✓Exploring Monitoring Data for ERA

- Risk drivers and dynamics
- Agricultural small streams neglected & at risk
- Neonicotinoids

Recap: What we learned from my PhD Thesis

✓Improving Statistics in ERA

- Change your model, not your data
- Take LOQ into account

✓Exploring Monitoring Data for ERA

- Risk drivers and dynamics
- Agricultural small streams neglected & at risk
- Neonicotinoids

✓Solutions for Linking Data in ERA

- Handling big eco(toxico-)logical data not easy
- Now easier

Statistical Ecotoxicology

Improving the Utilisation of Data for Environmental Risk Assessment

Eduard Szöcs

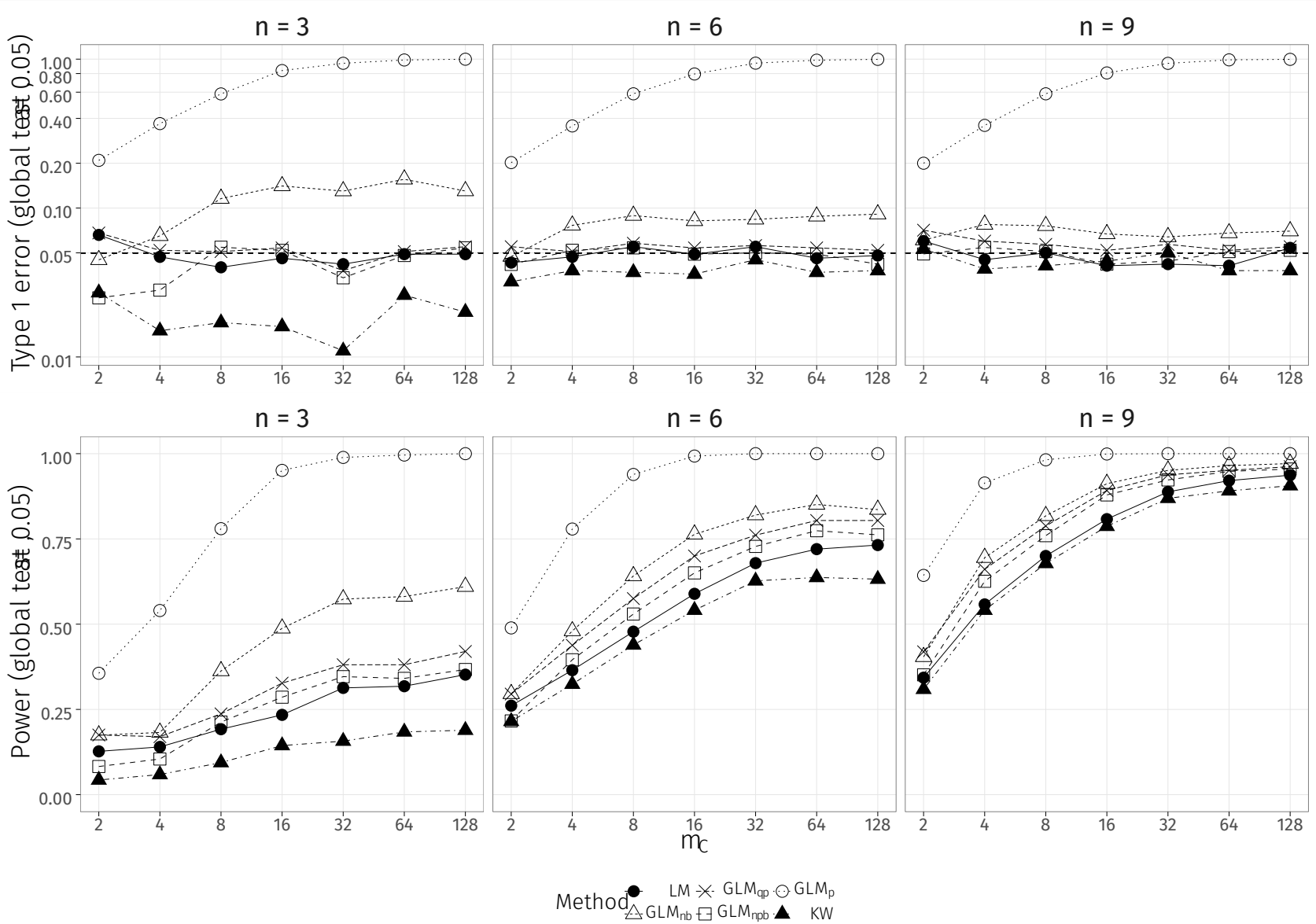
http://edild.github.io/

@EduardSzoecs

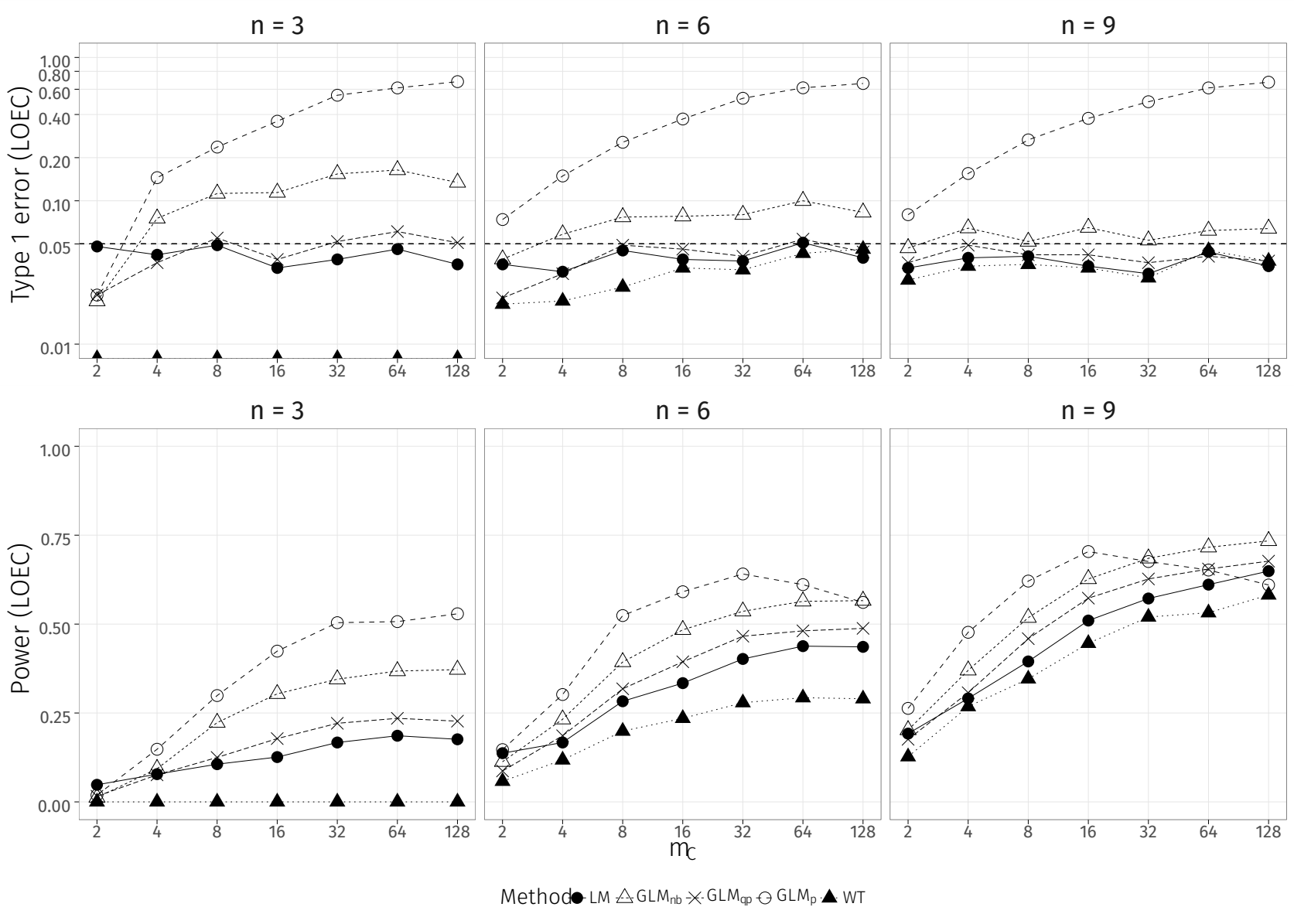
https://github.com/edild/phd_defense

https://github.com/edild/phd_thesis

Power en detail



For LOEC it is even worse



Comparison with Ives...

Szöcs (2015)

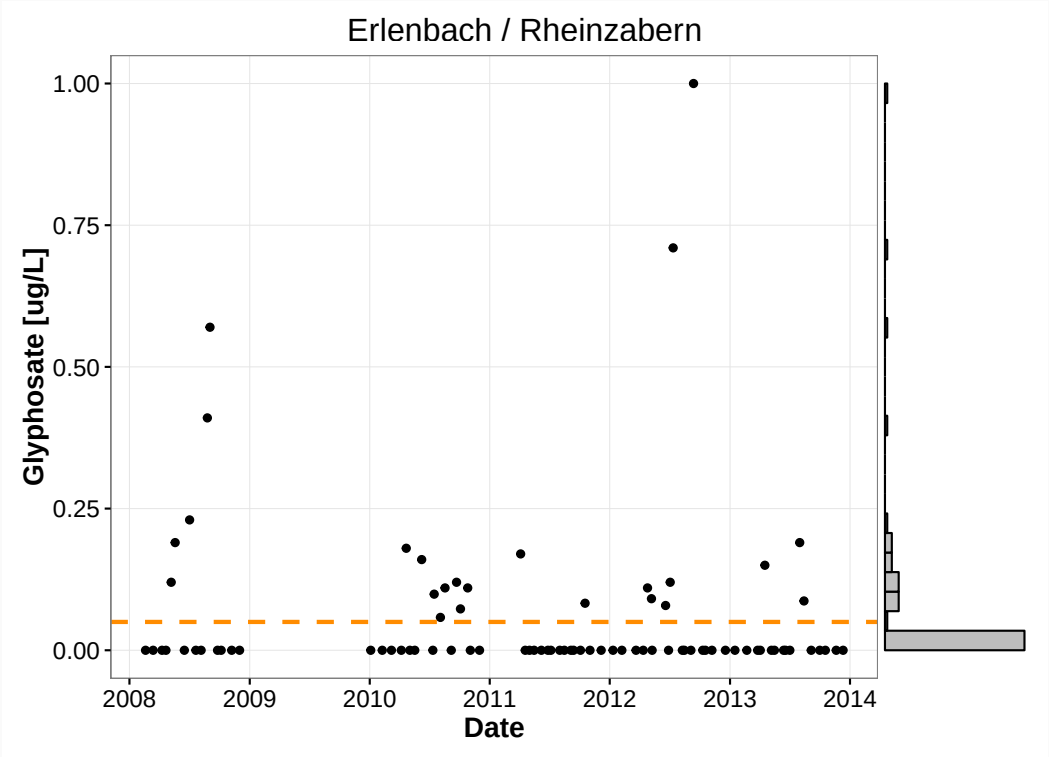
- factorial design
- one predictor
- low replicated
- LM, GLM, bootstrap
- High T1 error of NB
- Quasi-Poisson worked well
- Bootstrap fixes the problems

Ives (2015)

- continuous design
- two predictors
- well replicated
- LM, GLM
- High T1 error of NB
- Quasi-Poisson has problems with multiple predictors
-

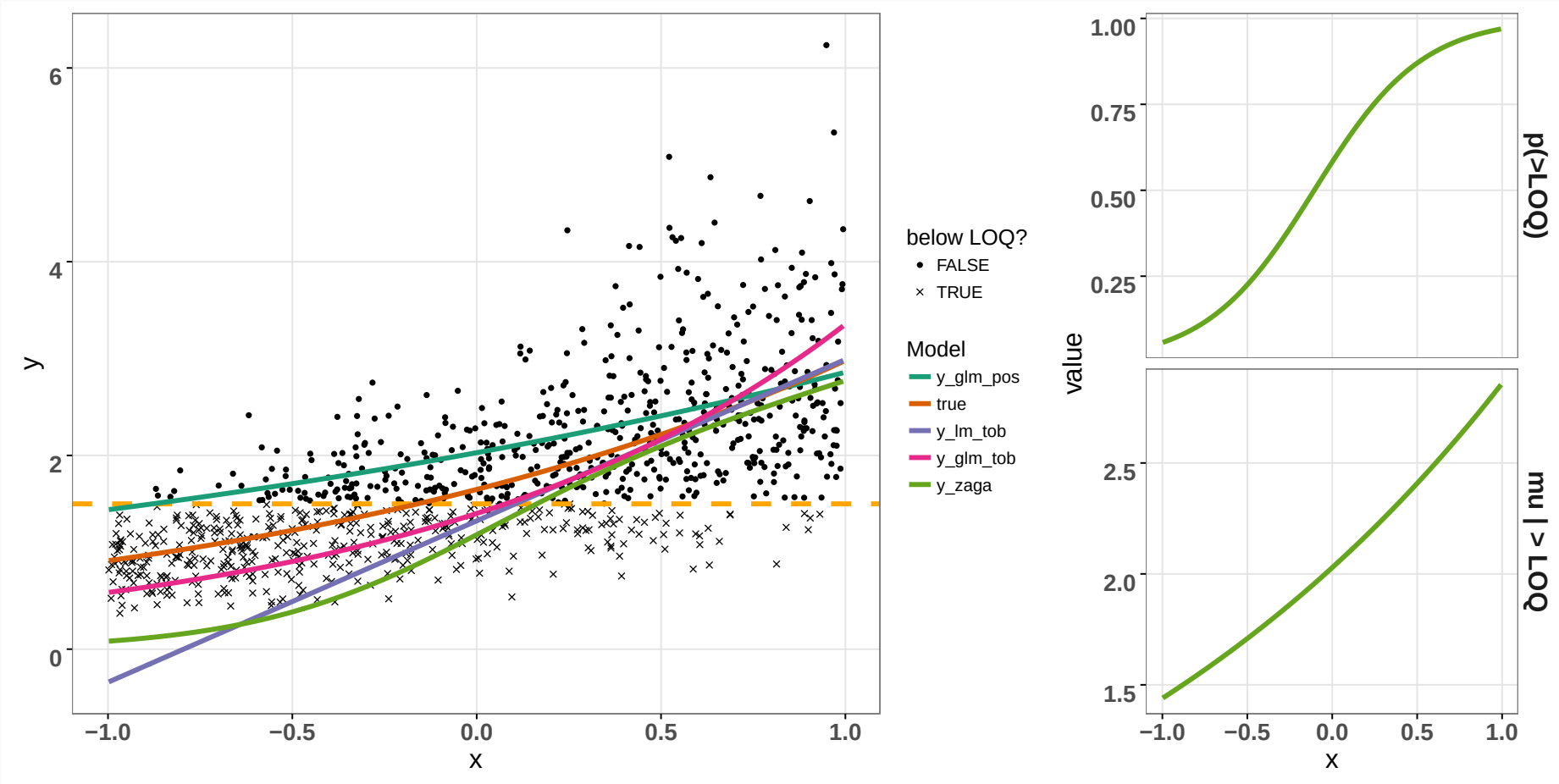
Idiosyncrasies of chemical concentrations

- continuous distribution in $\mathbb{R}_0^+$
- censoring (x <LOQ)
- non-linearity (season, trends)
- dependency (spatial, temporal)
- missing data

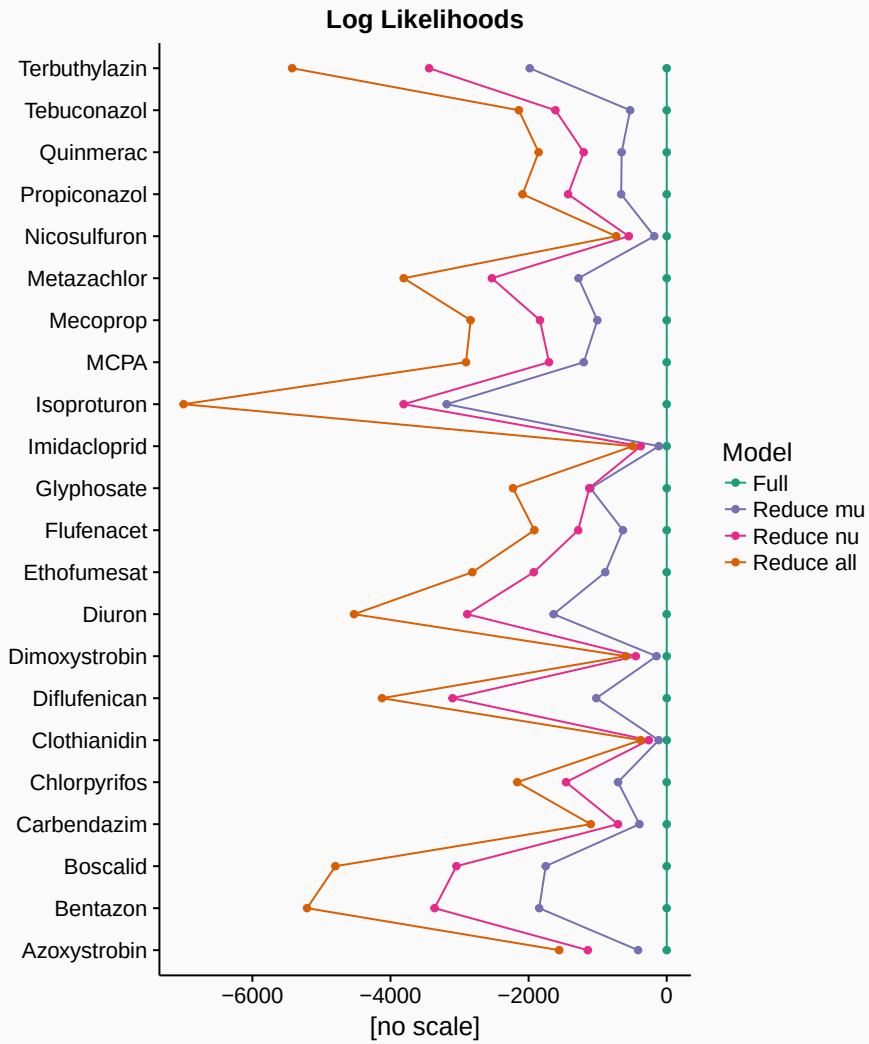


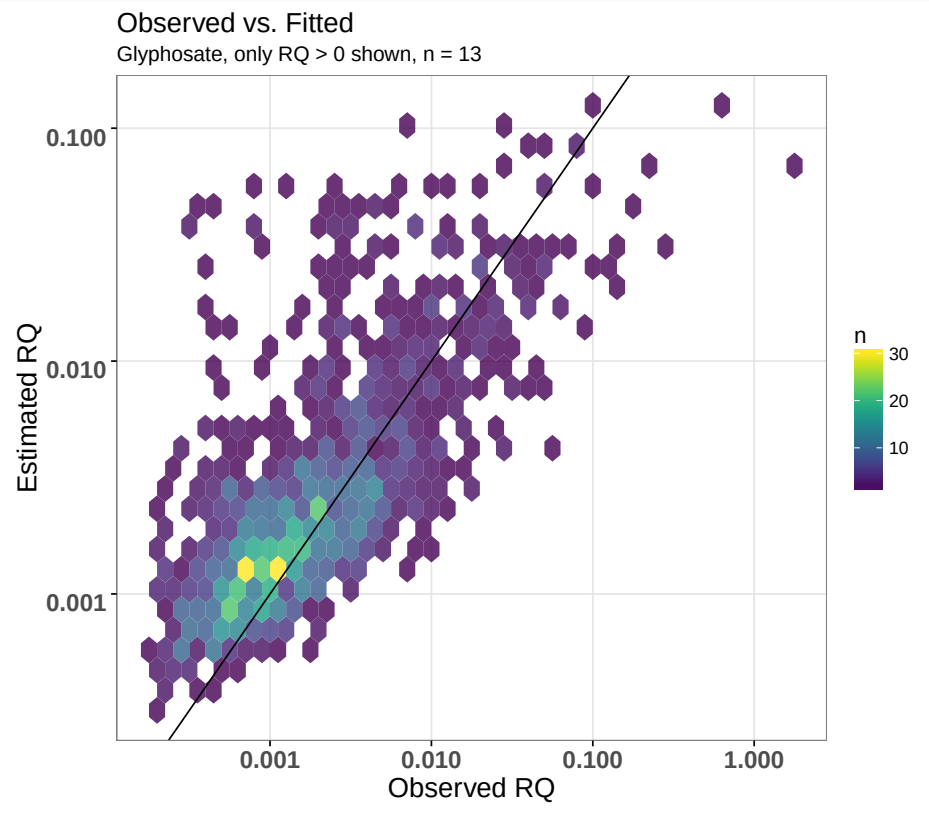
ZAGA what...?

shiny app: `http://uni-ko-ld.de/g4`

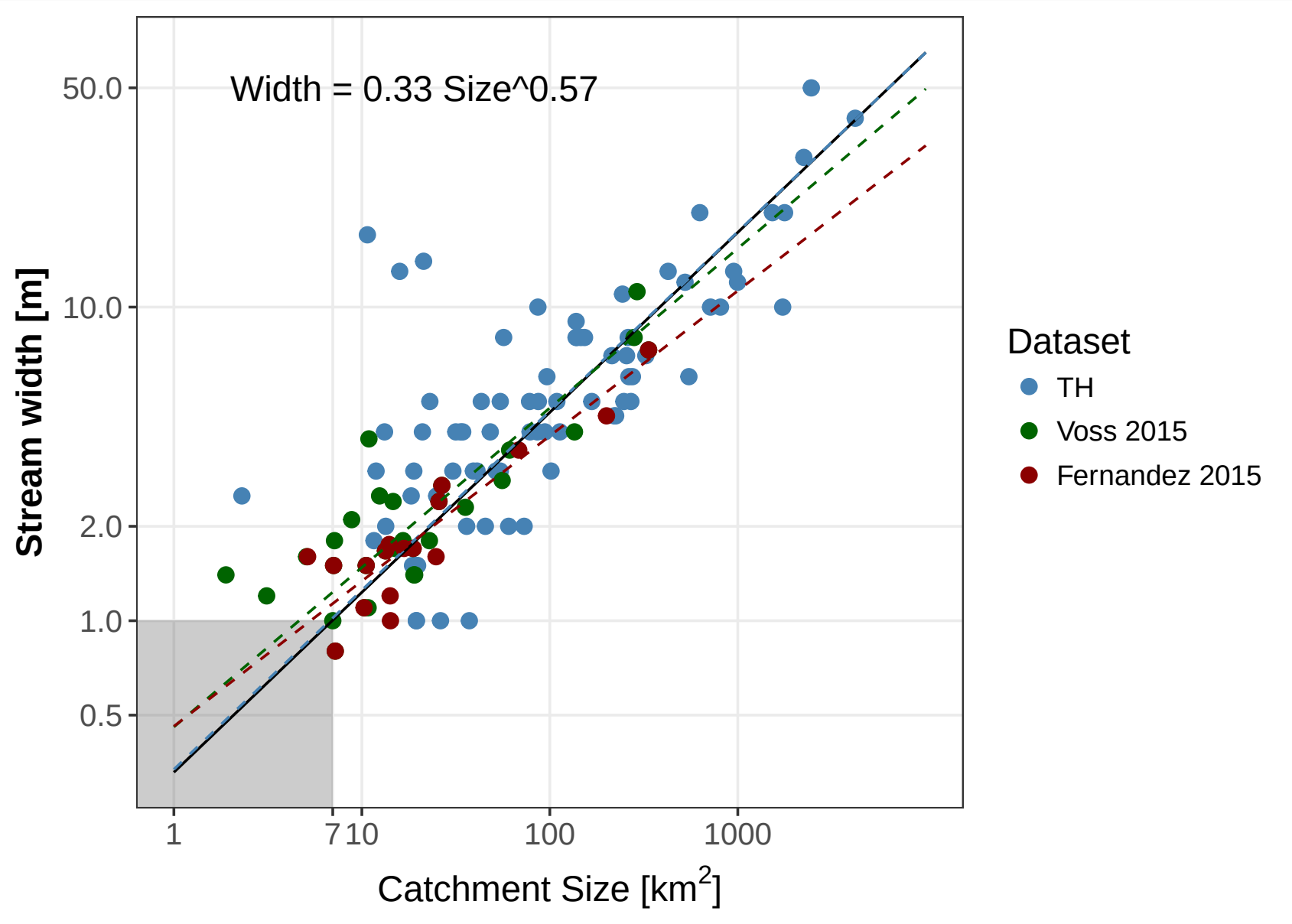


Model Performance

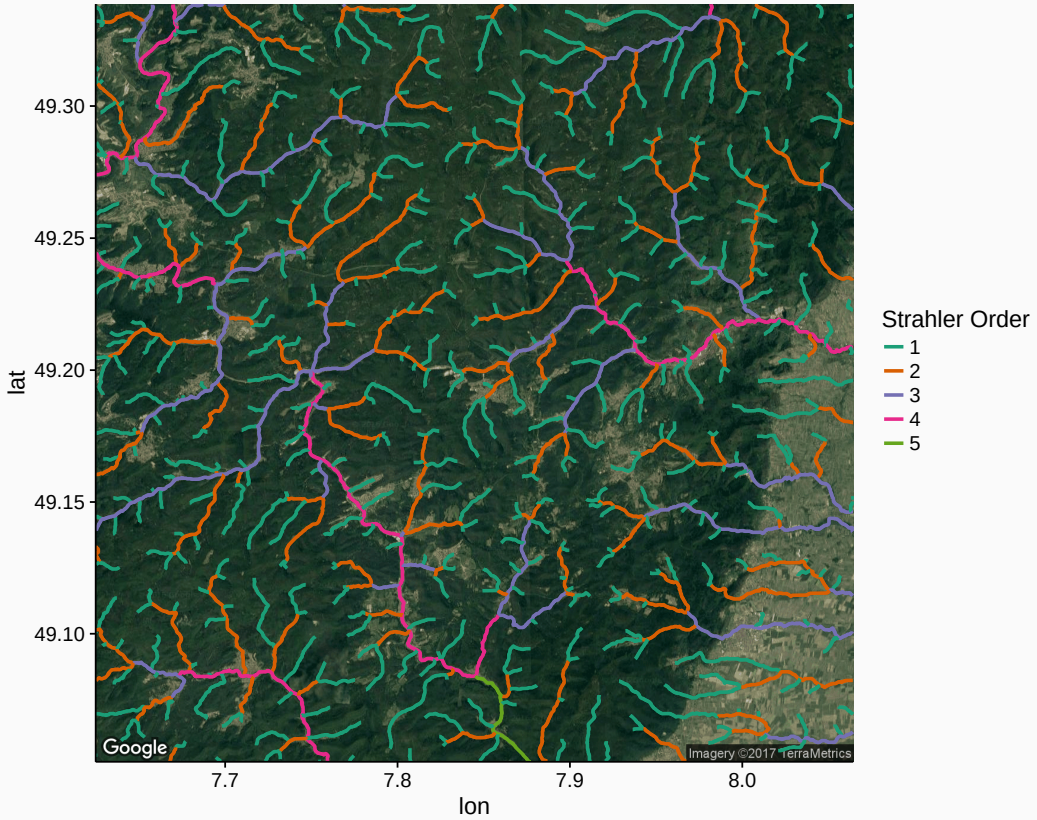


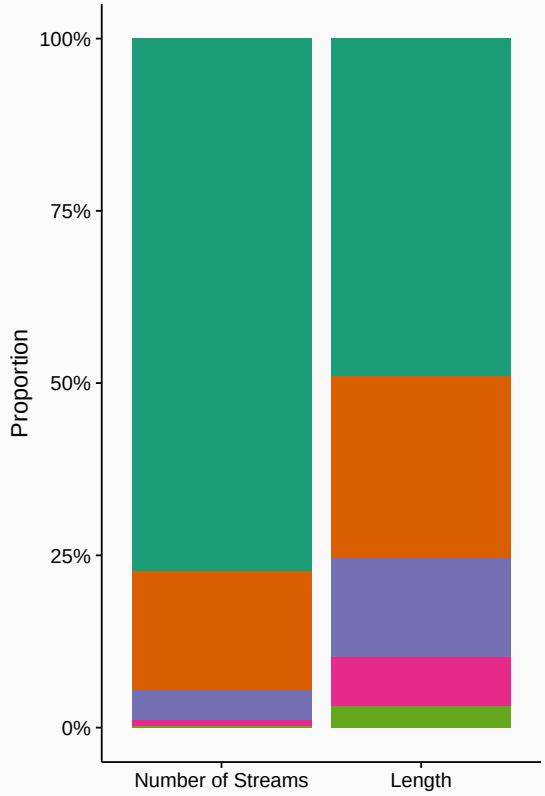


Stream size / width



Importance of small streams

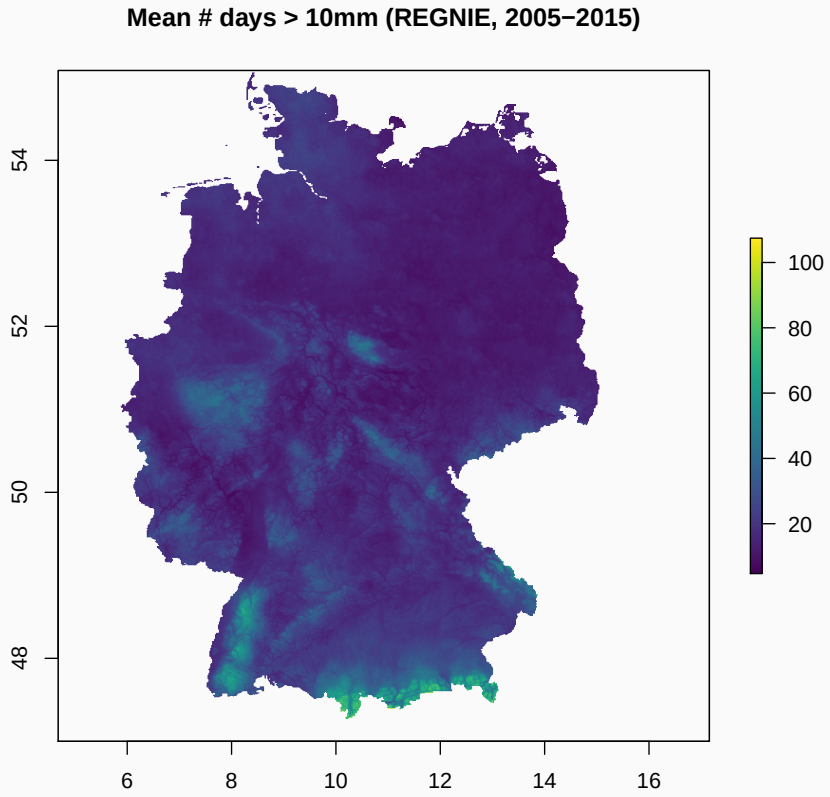


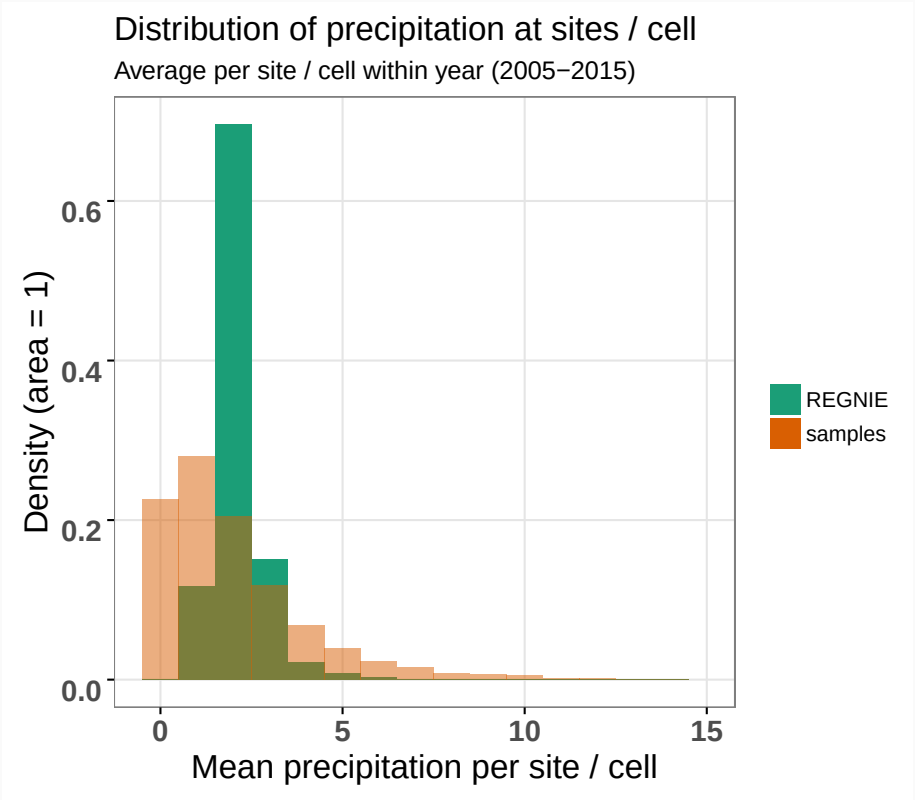


- Biodiversity
- Refuge for re-colonisation

Biggs et al. (2016). The importance of small waterbodies for biodiversity and ecosystem services: implications for policy makers. Hydrobiologia. DOI: 10.1007/s10750-016-3007-0.

Precipitation in Germany and the samples





Comparison with other studies

Szöcs (2016)

- Germany
- Monitoring
- Grab sampling
- Pesticides
- Neonics + Chlorpyrifos (OP)
- ZAGA (<LOQ)

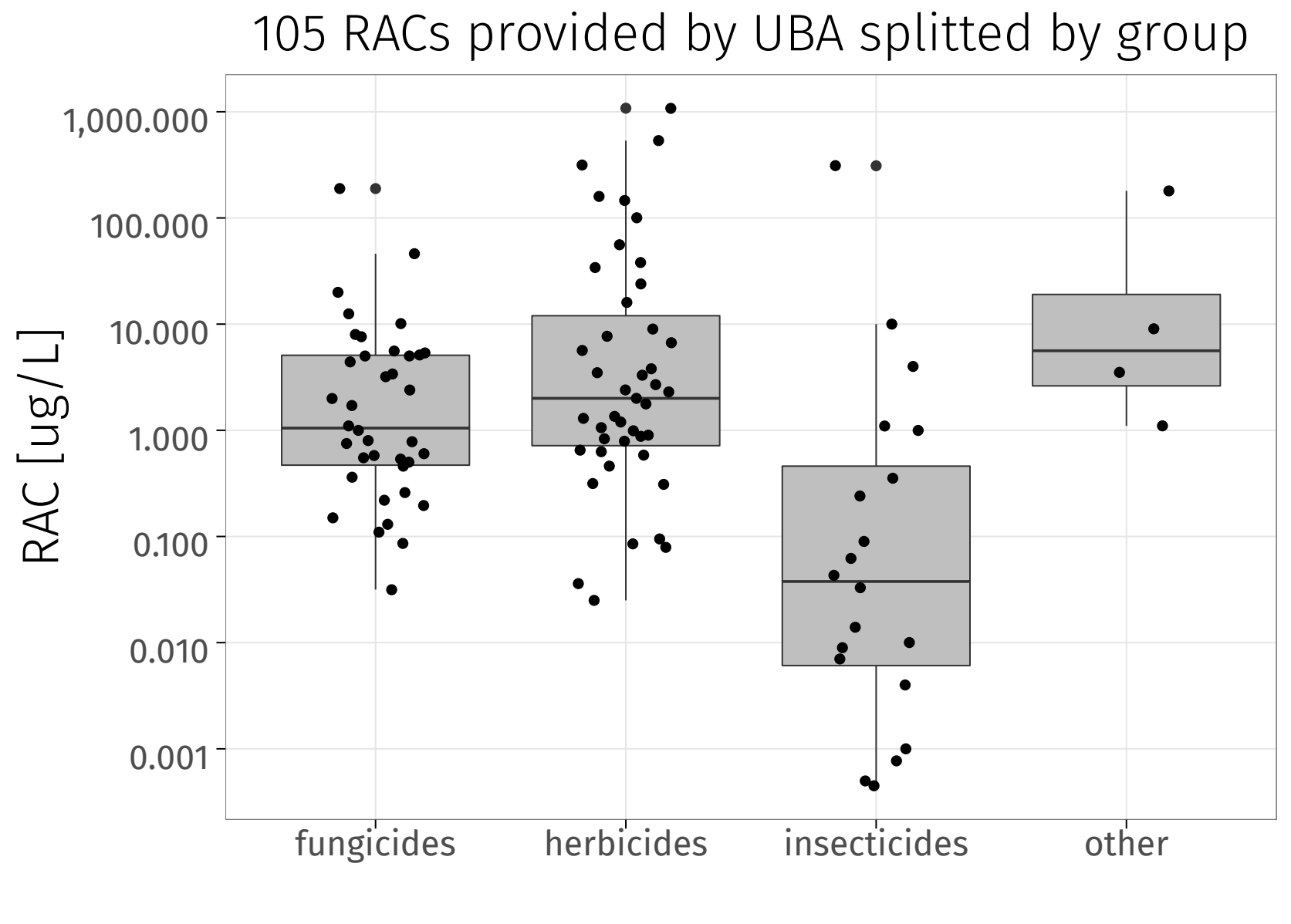
Stehle (2015)

- Europe / Global
- Publications
- Grab & Event driven sampling
- Insecticides
- Organophos.+ Pyrethroids
- LM for >LOQ

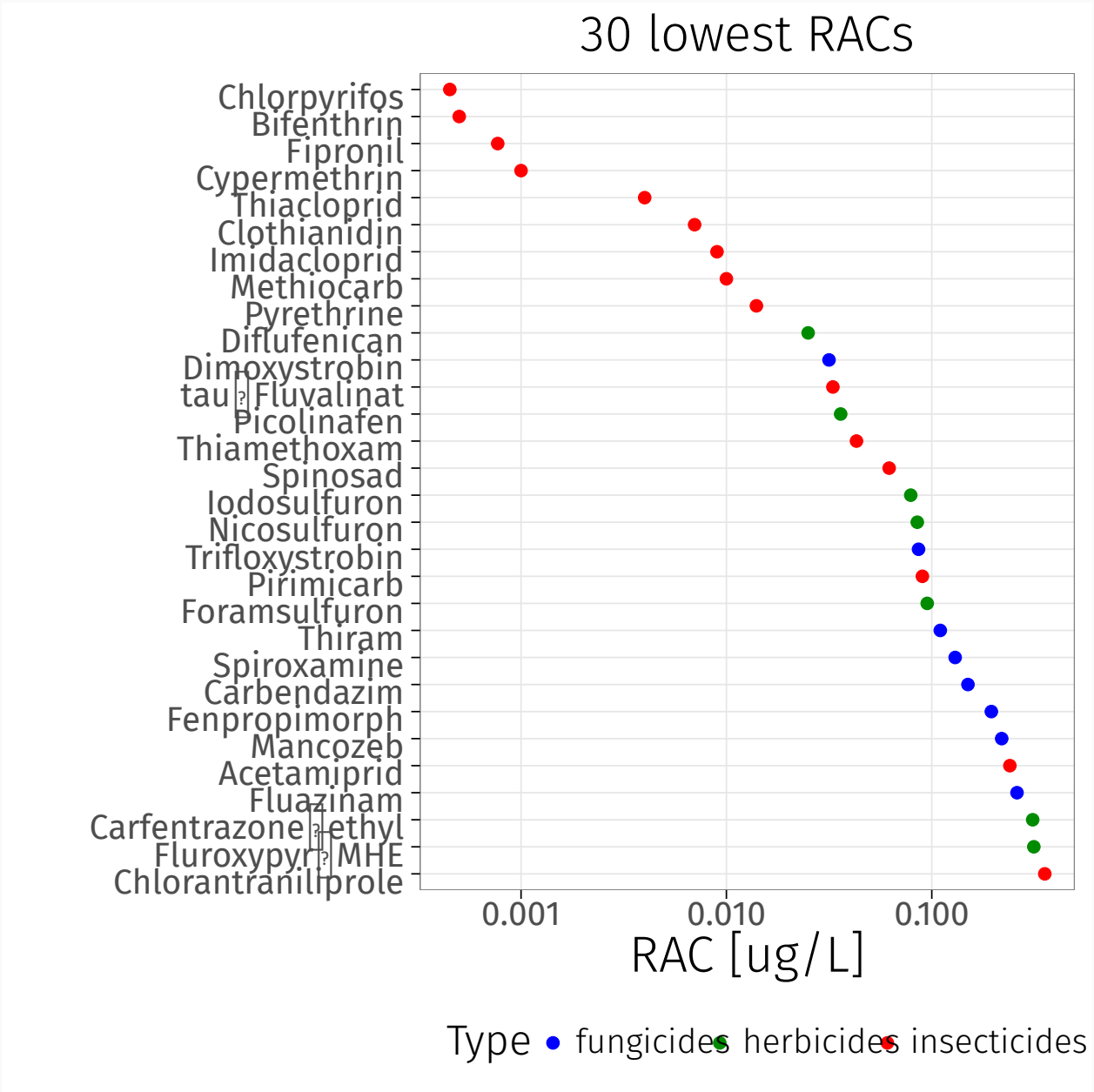
Knauer (2016)

- Switzerland
- Monitoring
- Grab sampling
- Pesticides
- Chlorpyrifos + Herb + Fung
- no model

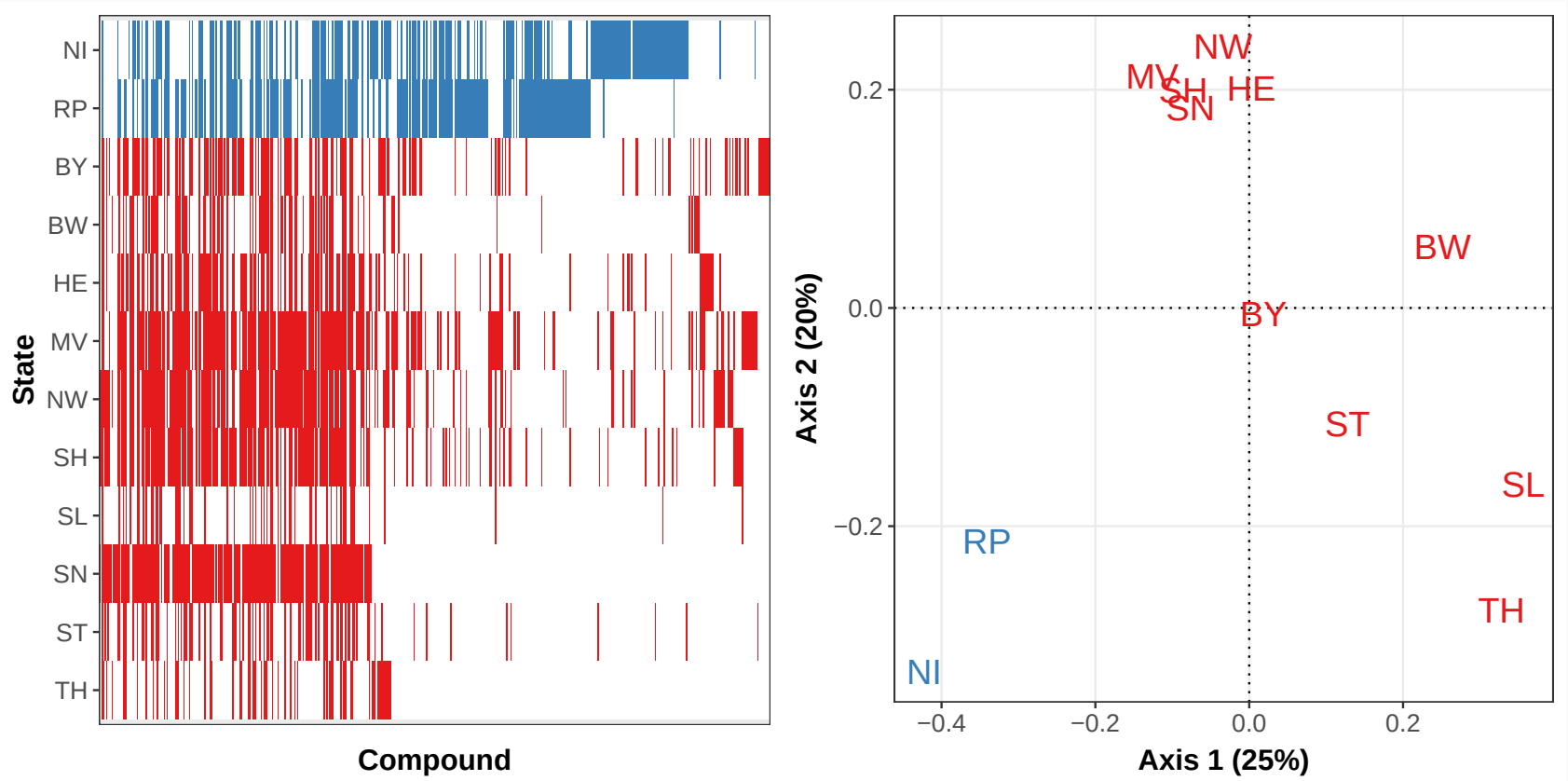
RACs by Type


105 RACs provided by UBA splitted by group
RAC [ug/L]
1,000.000
100.000
10.000
1.000
0.100
0.010
0.001
fungicides
herbicides
insecticides
other

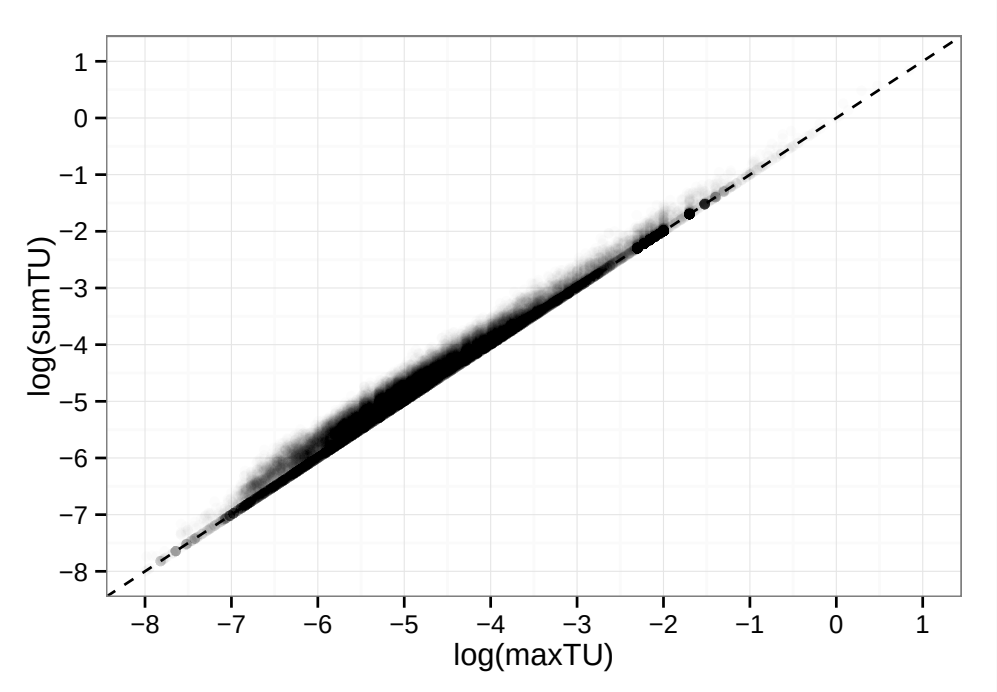
RACs by Compound



Analysed compound spectra by state



Mixtures are common, but one compound dominates the risk

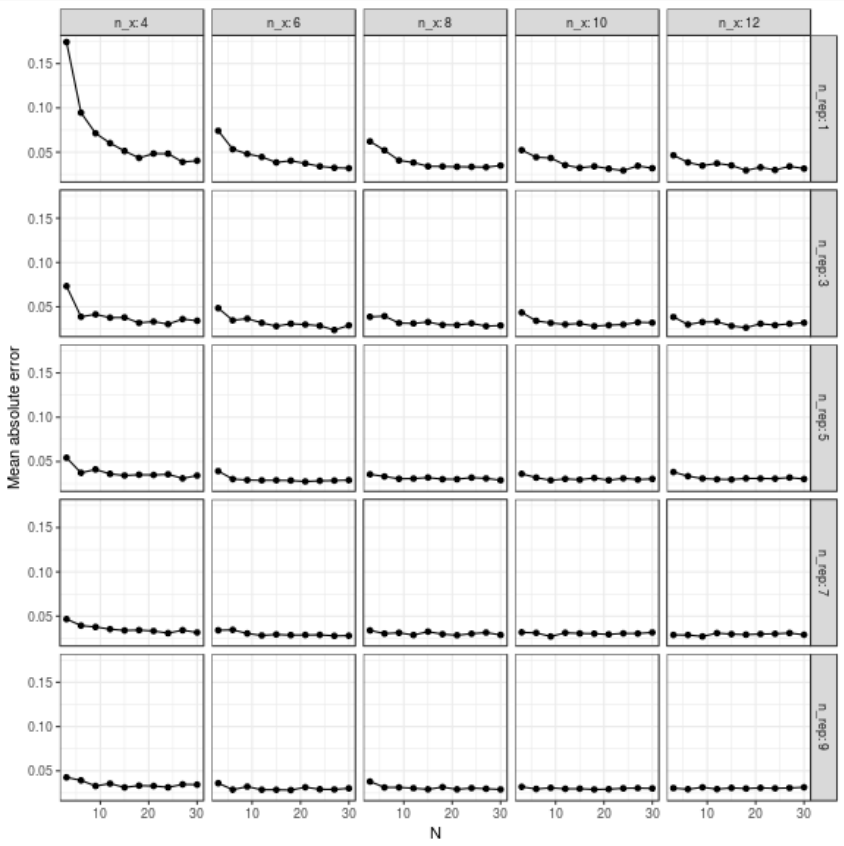


- up to 50 compounds in one sample
- high correlation
- $\sim$ 0.5 TU increase
- mainly one compound responsible for risk

Simulations are worth their work, use them *a priori*!

Experimental design for dose-response experiments - a simulation

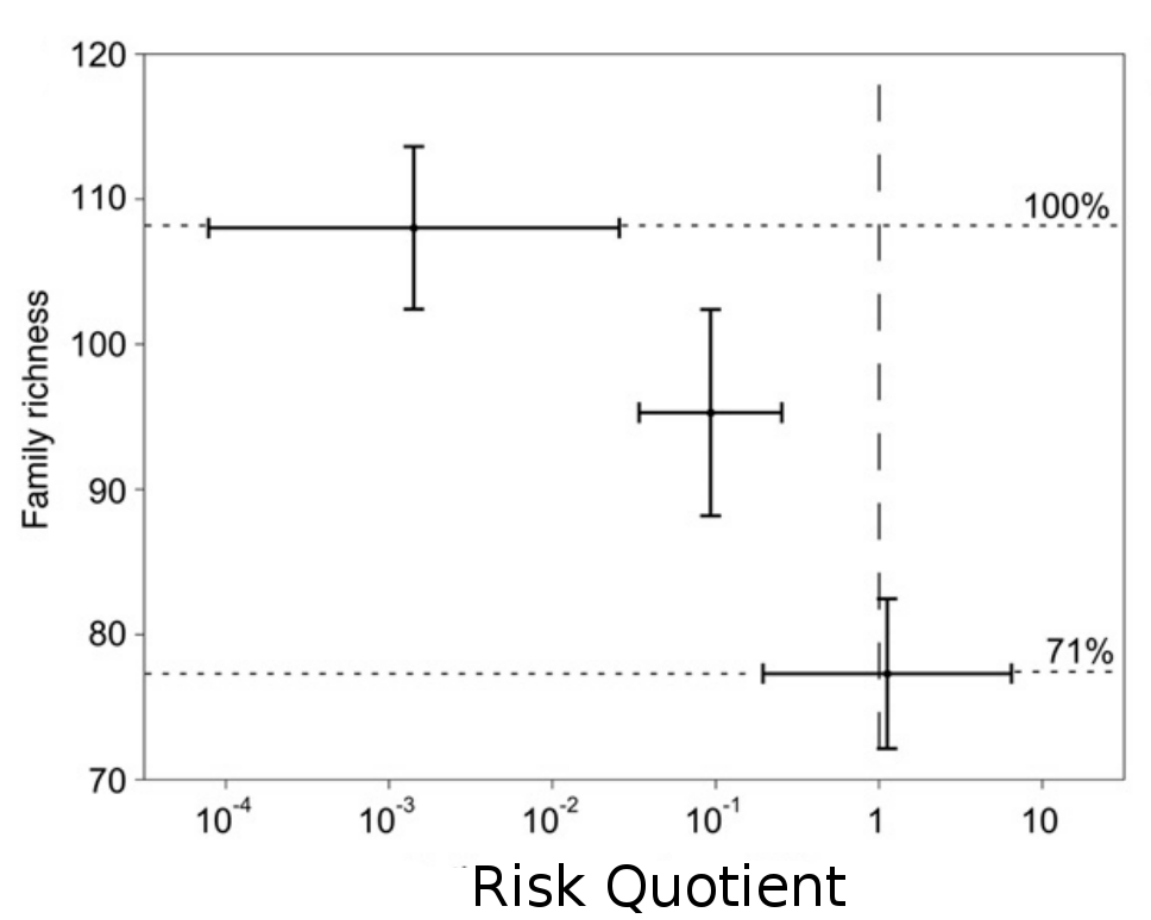
`http://edild.github.io/lc50_bias_sim/`



GLM-Explorer: `http://uni-ko-ld.de/g3`

Effects of RAC exceedances

- RACs should never be exceeded (=protection goal)
- If so, biological effects likely



Modified from: Stehle, Schulz (2015). Agricultural insecticides threaten surface waters at the global scale. PNAS 112, 5750–5755.

Reasons for observed RAC exceedances

- Risk Mitigation fails
  - Risk mitigation measures (erosion rills, wind)
  - Farmer do not adhere (GAP, no spray zones)
- Risk Assessment fails
  - Exposure Assessment
    - Models not working (Knäbel et al.)
  - Effect Assessment
    - Missed sensitive species
    - New document asks also for insects

Risk Mitigation Measures

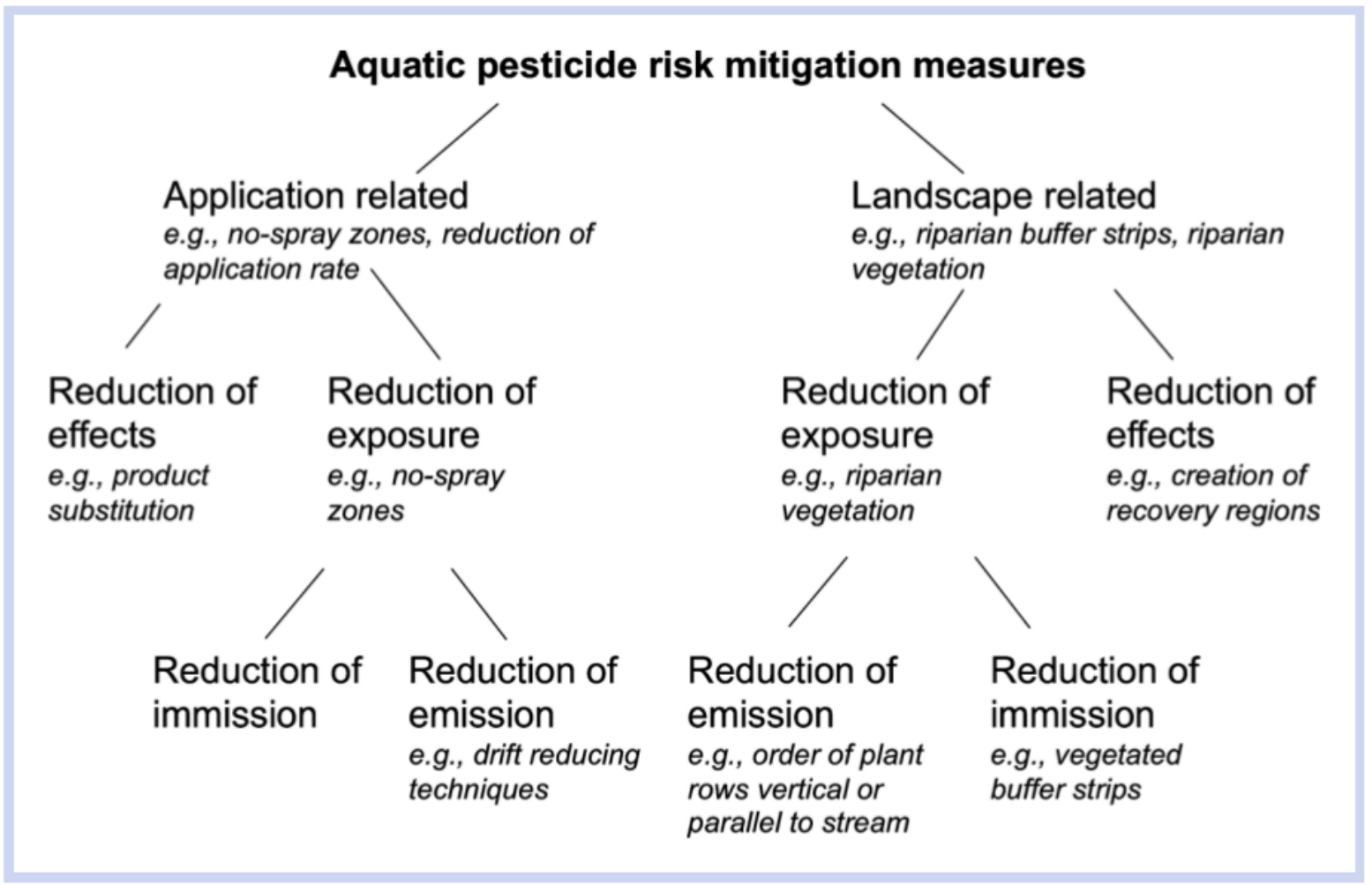
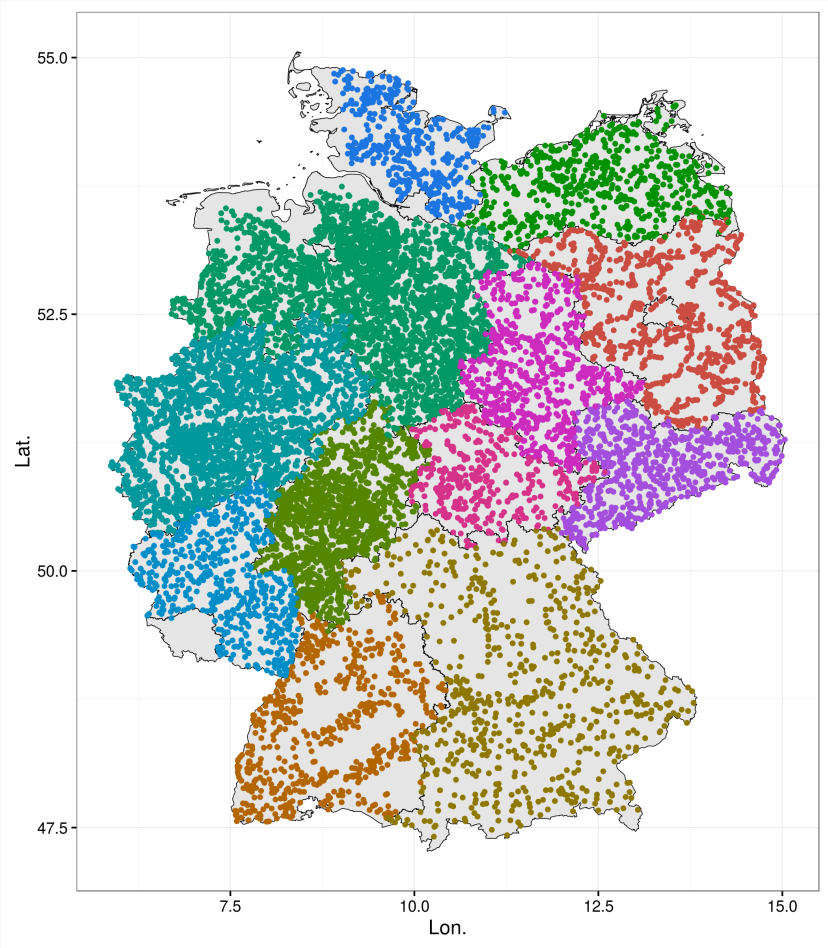


Figure 1. Classification of pesticide risk mitigation measures.

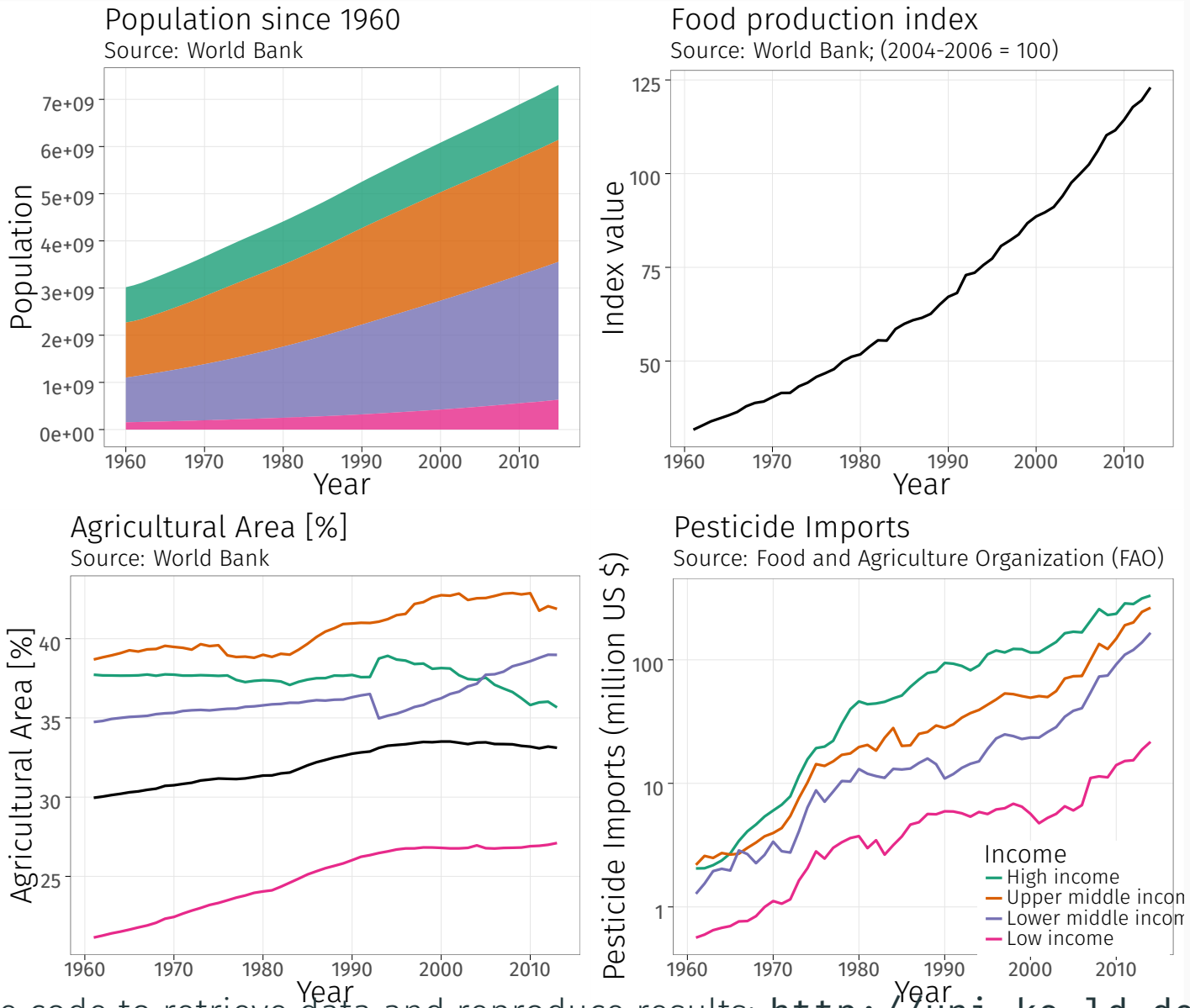
Bereswill, Streloke, Schulz (2014). Risk mitigation measures for diffuse pesticide entry into aquatic ecosystems: Proposal of a guide to identify appropriate measures on a catchment scale: Guide to Identify Pesticide Risk Mitigation Measures. IEAM 10, 286–298.

Biotic field effects



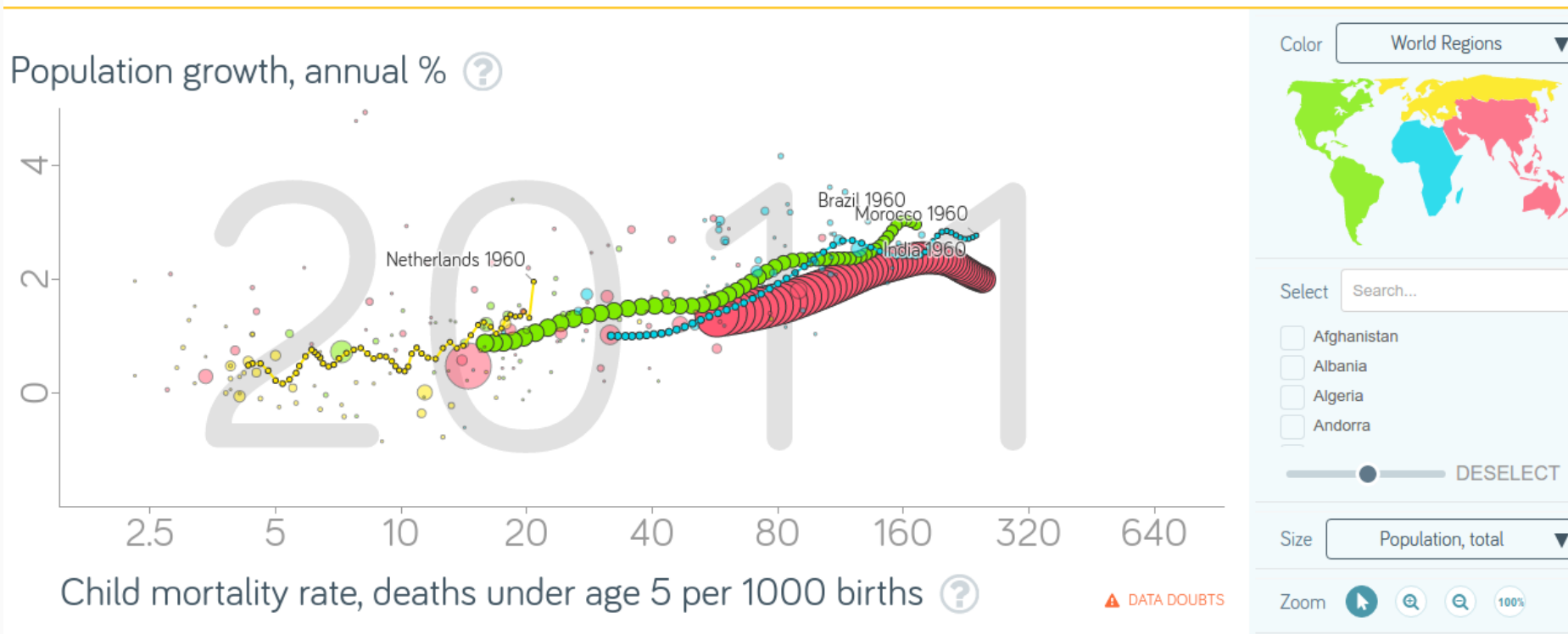
- biological data with good spatial coverage
- 60% of spatial congruence
- Large scale effects largely unknown.
- Some work left...
- Future....

A global perspective (I)



Source code to retrieve data and reproduce results: `http://uni-ko-ld.de/g7`

A global perspective (II)



Source: `https://www.gapminder.org/tools`

Software availability

Stable versions on CRAN, dev versions on github.

```
webchem github.com/ropensci/webchem
   taxize github.com/ropensci/taxize
```

Best practices for Software:

- open source (permissive MIT License)
- version control (git)
- automated tests (Travis-CI)
- in source documentation (roxygen)

Poisot, T., 2015. Best publishing practices to improve user confidence in scientific software. Ideas in Ecology and Evolution 8.

Many Thanks To

- My supervisor Prof. Dr. Ralf. B. Schäfer (for his support, openness, opportunities & discussions)
- My colleagues & collaborators (to many to list here)
- German Environment Agency (for funding & collab)
- My parents Anca & Helmut (for their support)
- My girlfriend Anja (for everything)