Statistical Eco(-toxico)logy

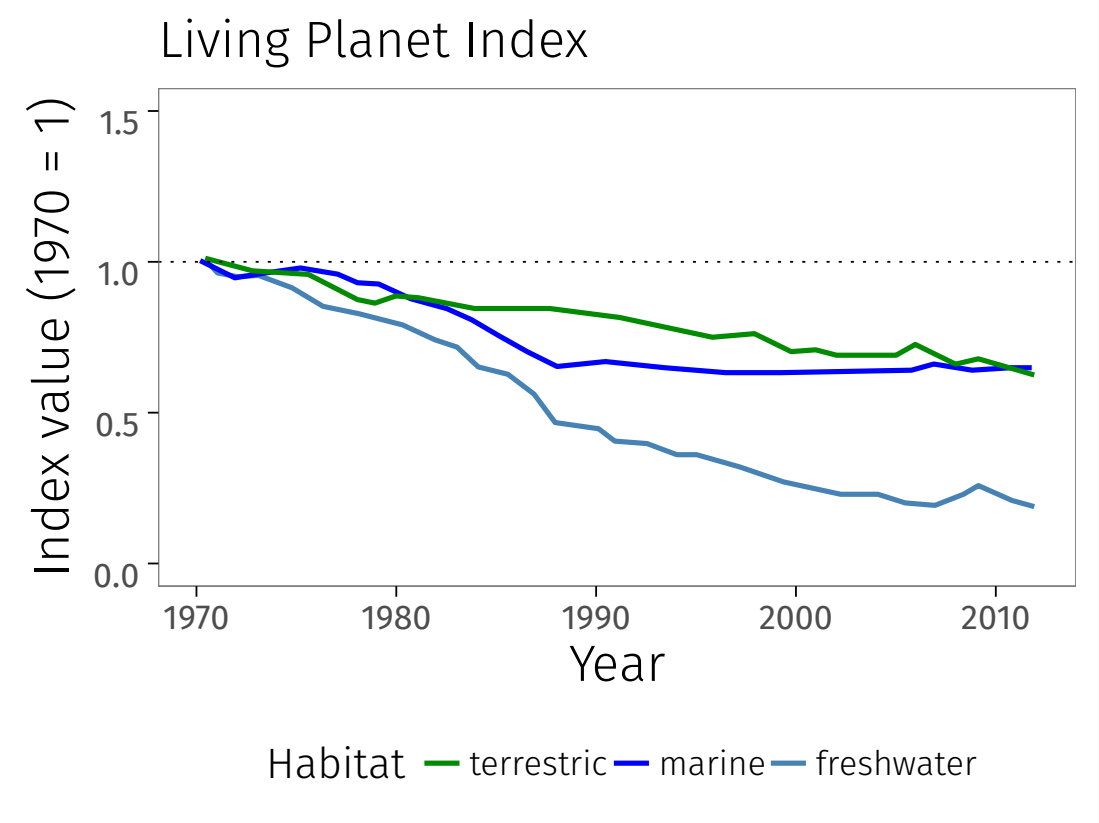
Improving the Utilisation of Data for Environmental Risk Assessment

Eduard Szöcs

25th January 2017

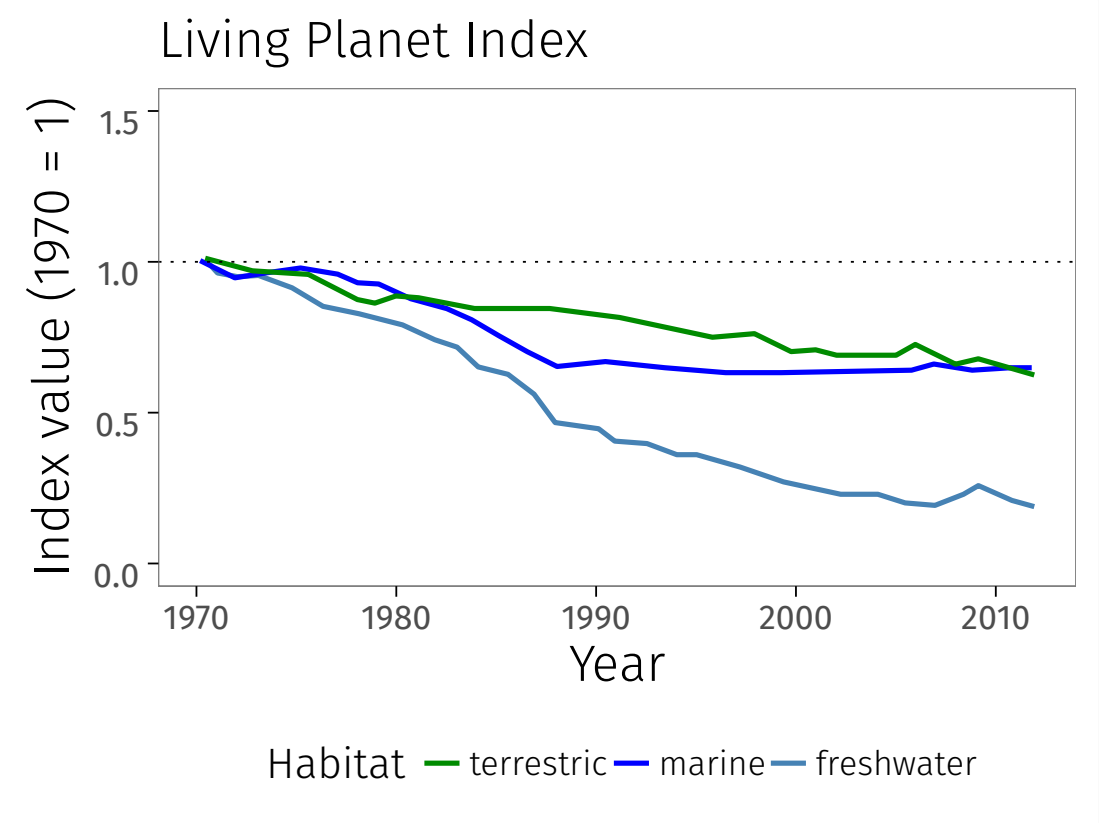
Table of contents

Freshwater biodiversity is strongly declining



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

Freshwater biodiversity is strongly declining

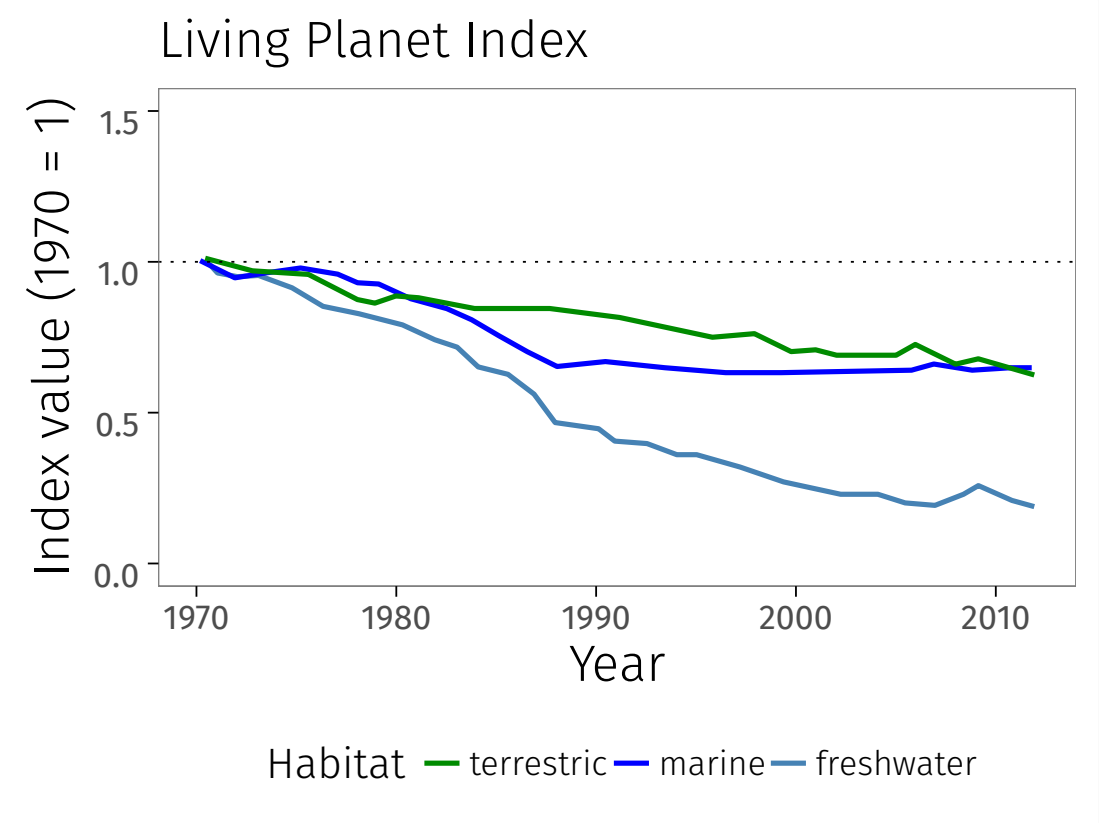


Reasons

- 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

Plant Protection Products Regulation 1107/2009

Water Framework Directive 2000/60/EC

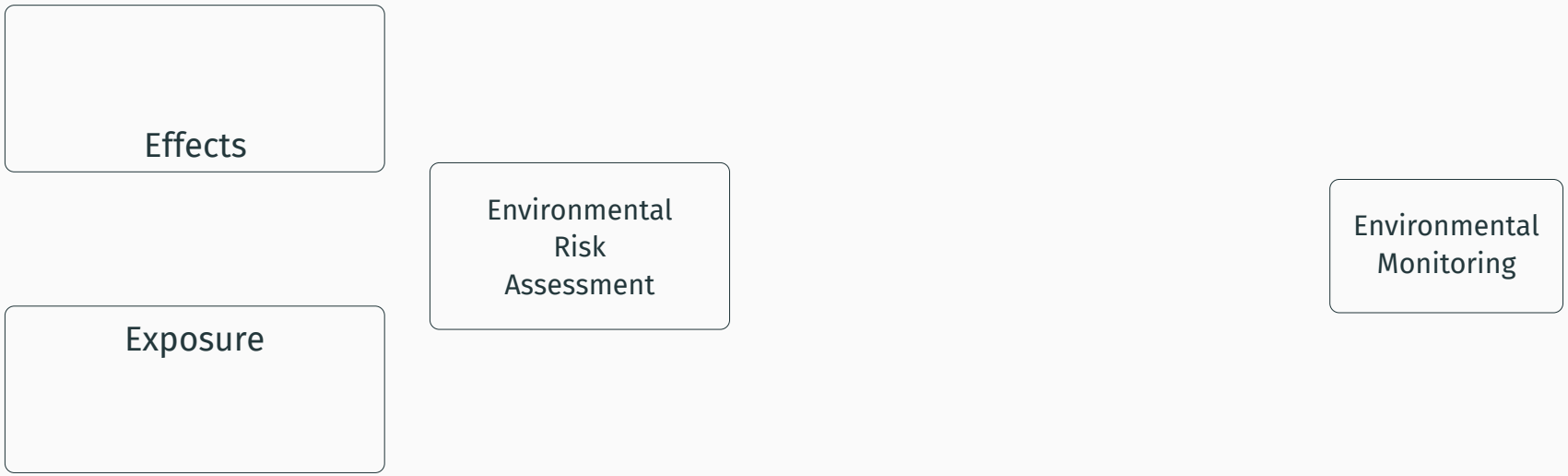
Environmental Risk Assessment

Environmental Monitoring

Environmental Risk Assessment and Monitoring

Plant Protection Products Regulation 1107/2009

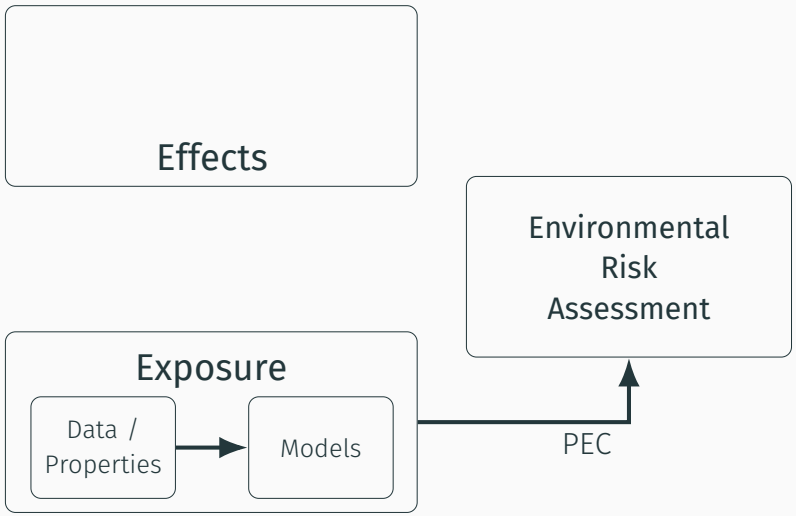
Water Framework Directive 2000/60/EC



Environmental Risk Assessment and Monitoring

Plant Protection Products Regulation 1107/2009

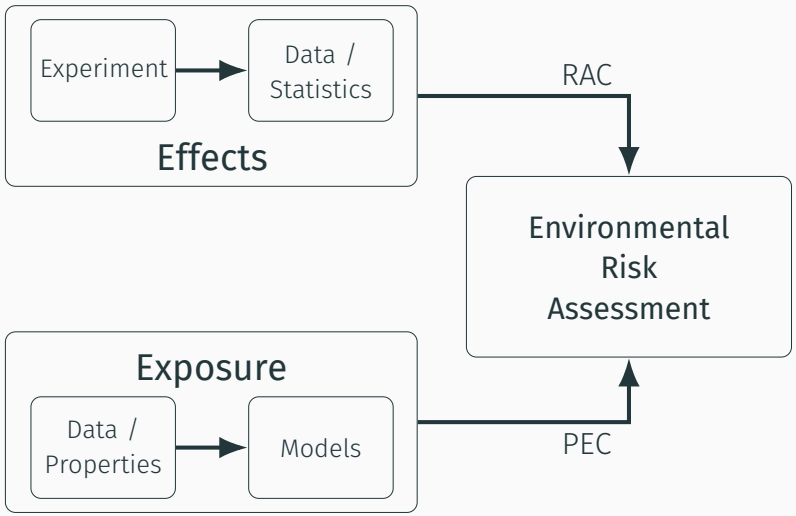
Water Framework Directive 2000/60/EC



Environmental Monitoring

Environmental Risk Assessment and Monitoring

Plant Protection Products
Regulation 1107/2009



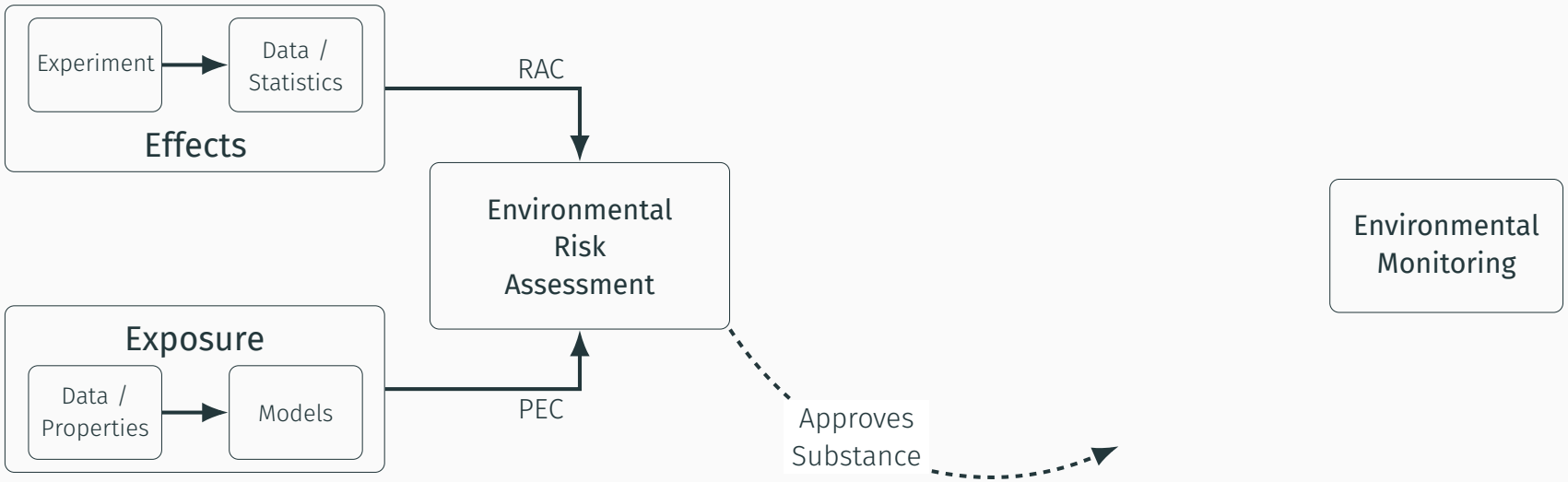
Water Framework Directive
2000/60/EC

Environmental Monitoring

Environmental Risk Assessment and Monitoring

Plant Protection Products Regulation 1107/2009

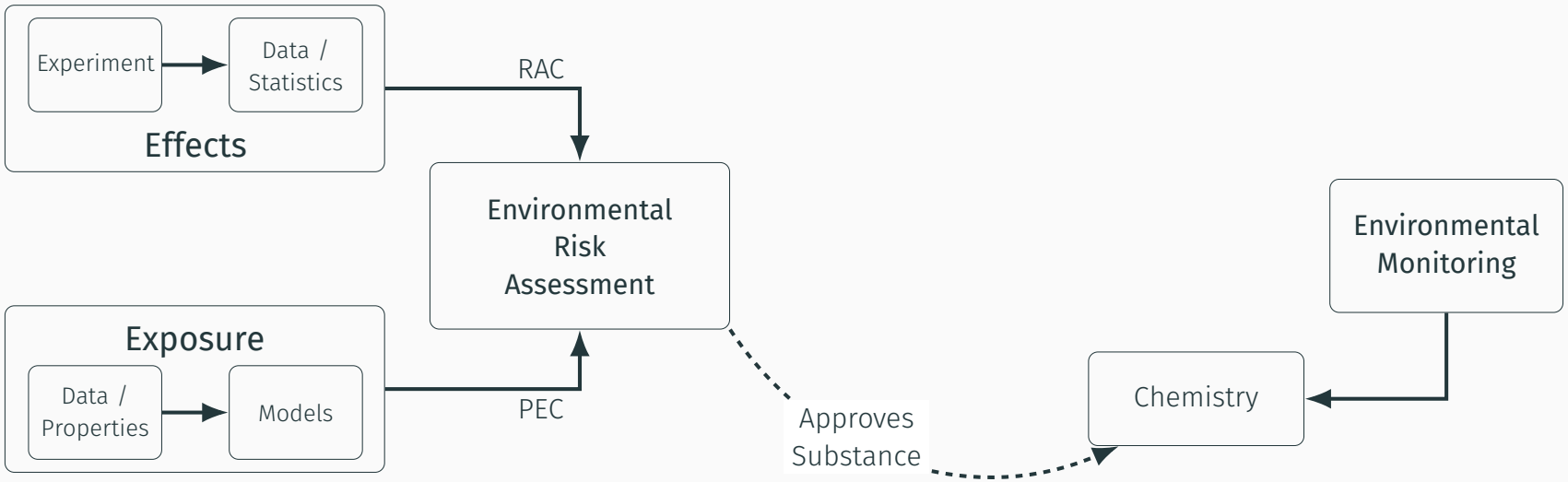
Water Framework Directive 2000/60/EC



Environmental Risk Assessment and Monitoring

Plant Protection Products Regulation 1107/2009

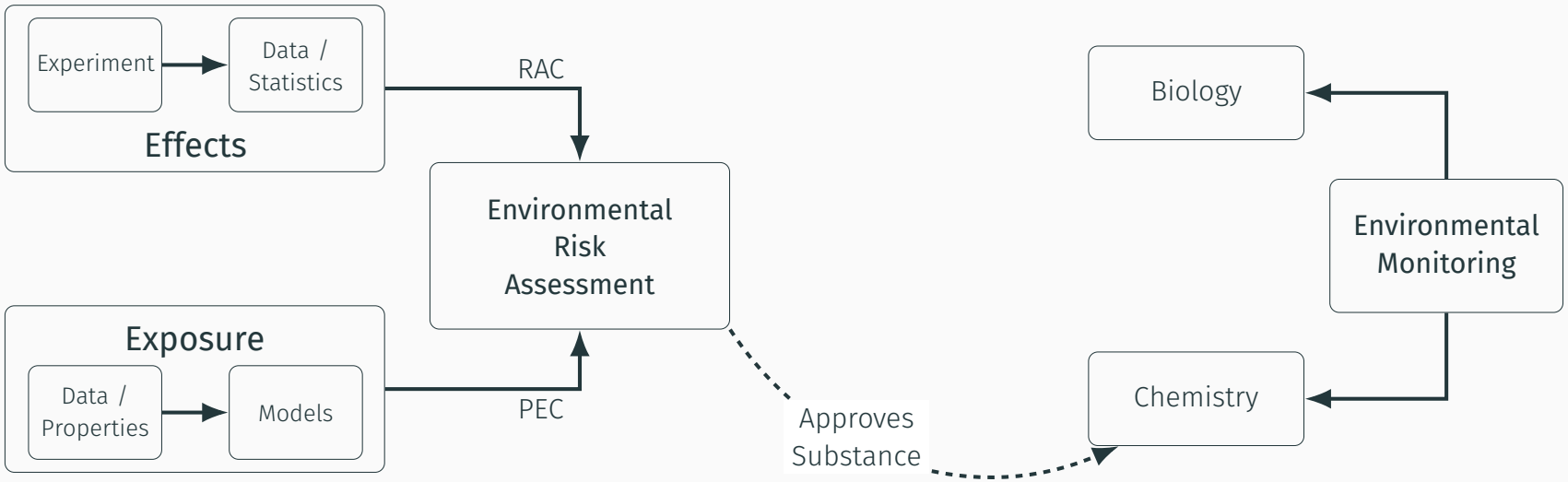
Water Framework Directive 2000/60/EC



Environmental Risk Assessment and Monitoring

Plant Protection Products Regulation 1107/2009

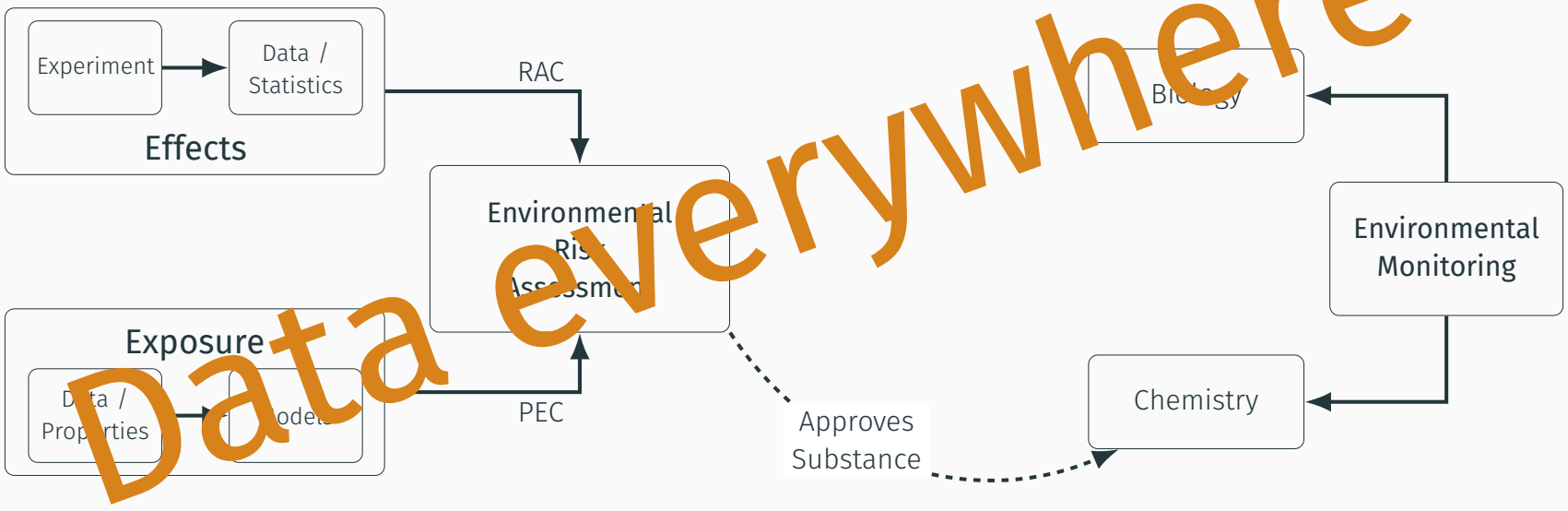
Water Framework Directive 2000/60/EC



Environmental Risk Assessment and Monitoring

Plant Protection Products Regulation 1107/2009

Water Framework Directive 2000/60/EC

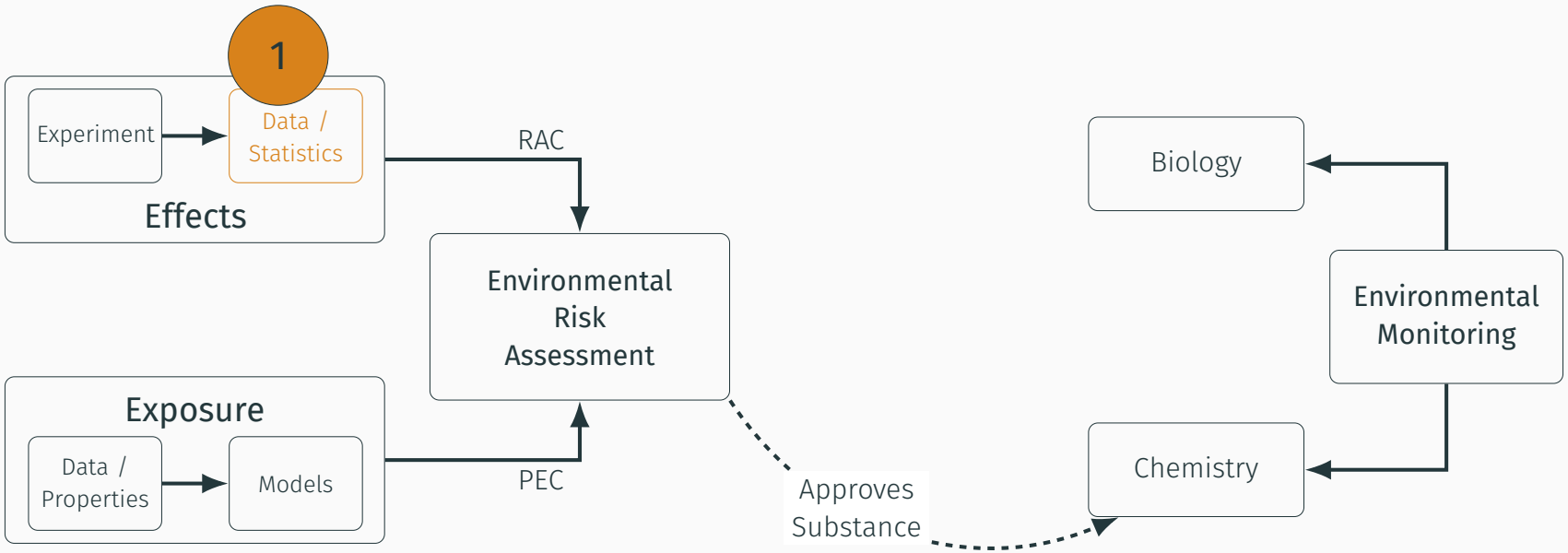


Data everywhere

Improving Statistics in ERA

Plant Protection Products Regulation 1107/2009

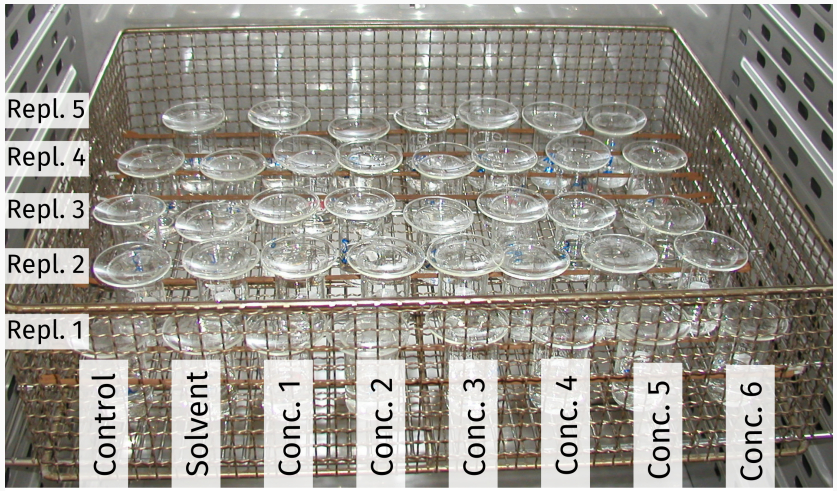
Water Framework Directive 2000/60/EC



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

Improving Statistics in ERA

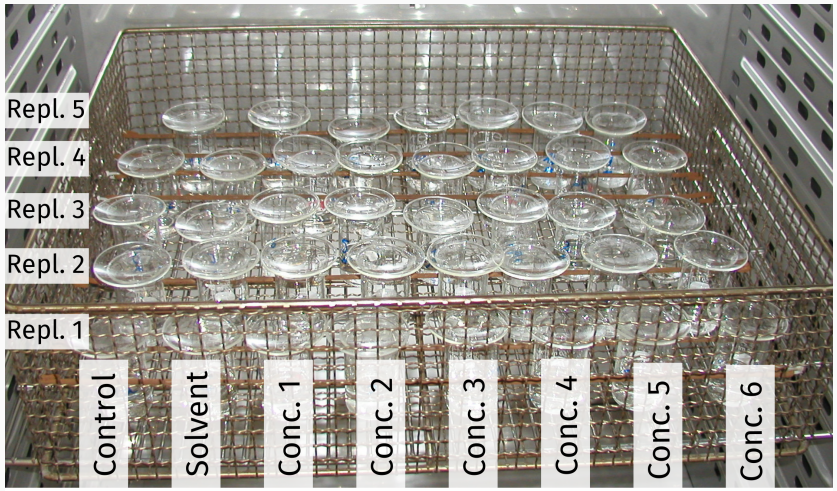
Experiments in Effect Assessment



- Daphnia Test
- Lower Tier
- *"x out of n survived"*
- EC_{50} / NOEC

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

Experiments in Effect Assessment

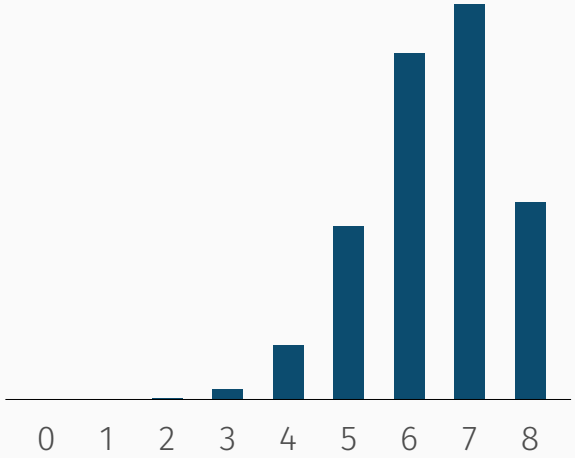


- Daphnia Test
- Lower Tier
- *"x out of n survived"*
- EC_{50} / NOEC

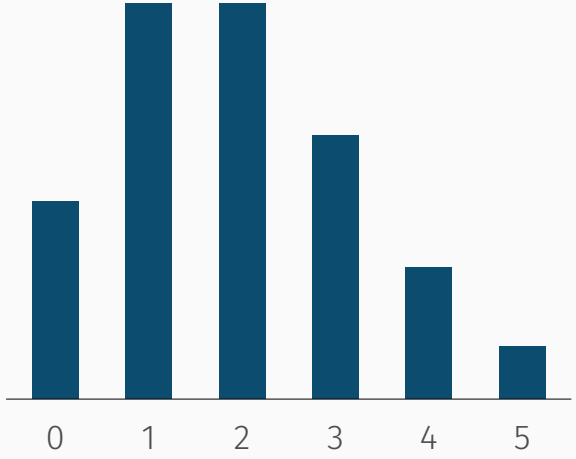
- Mesocosm
- Higher Tier
- *"number of animals"*
- NOEC

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

Ecotoxicology is not normal

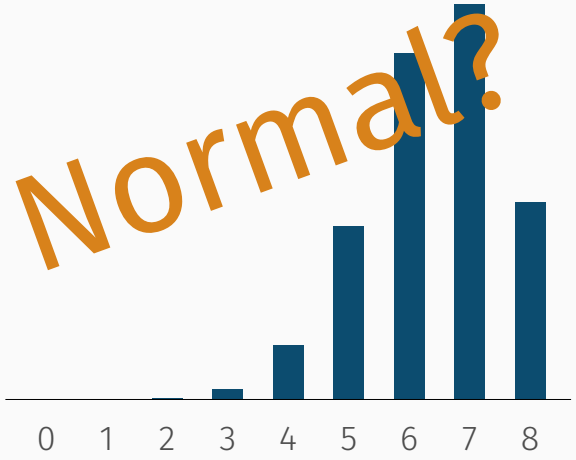


- *”x out of n survived”*

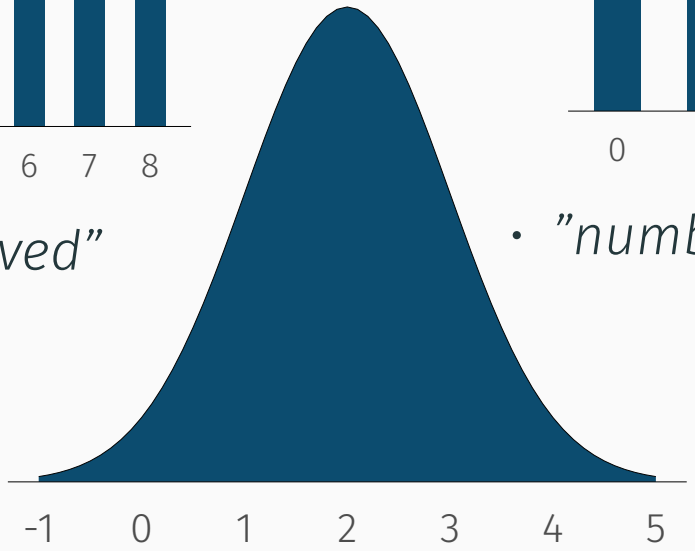


- *”number of animals”*

Ecotoxicology is not normal

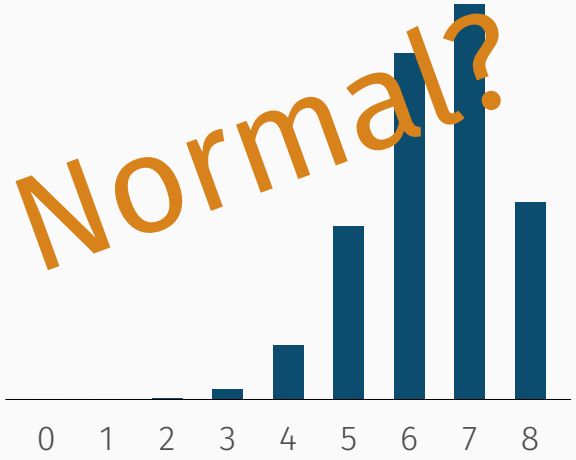


- *"x out of n survived"*

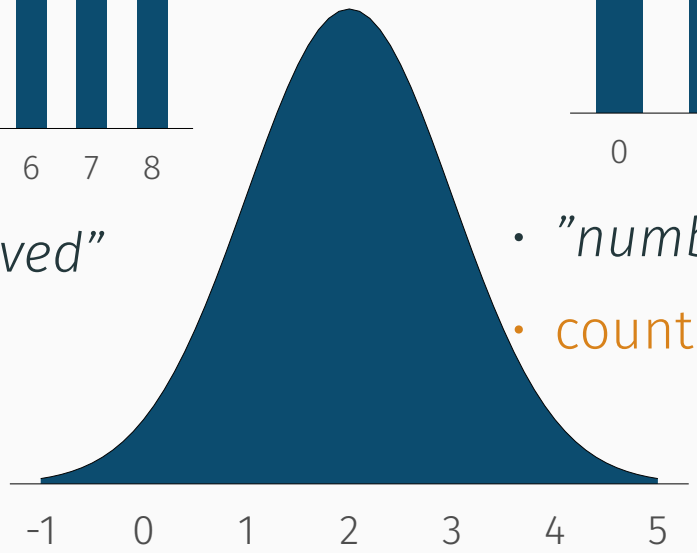


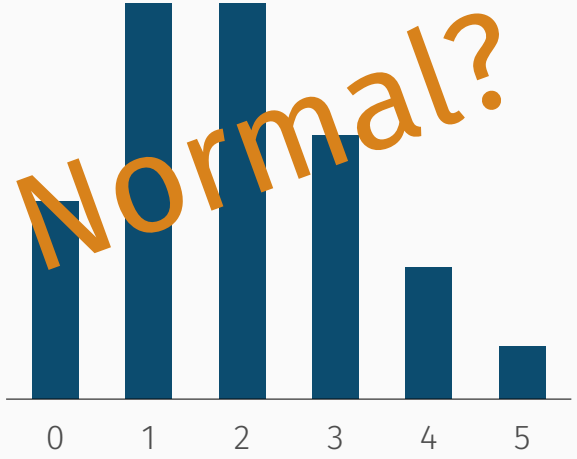
- *"number of animals"*

Ecotoxicology is not normal



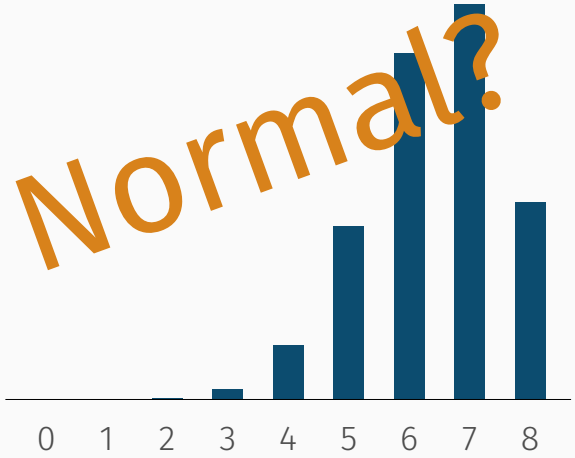
- *"x out of n survived"*
- binomial data



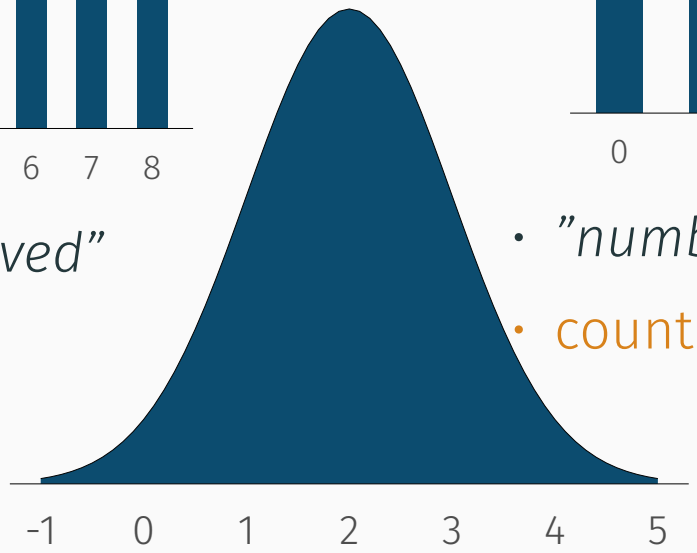


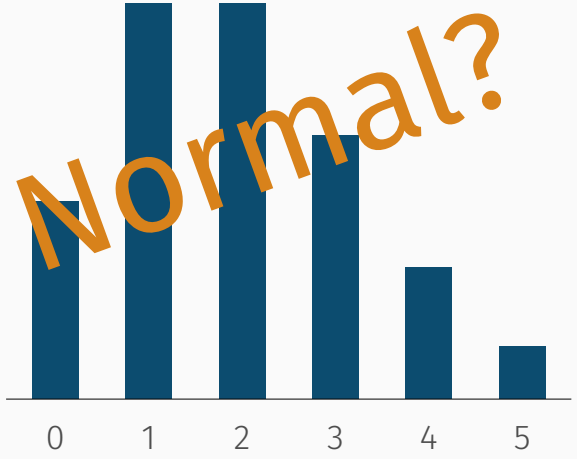
- *"number of animals"*
- count data

Ecotoxicology is not normal



- *"x out of n survived"*
- binomial data

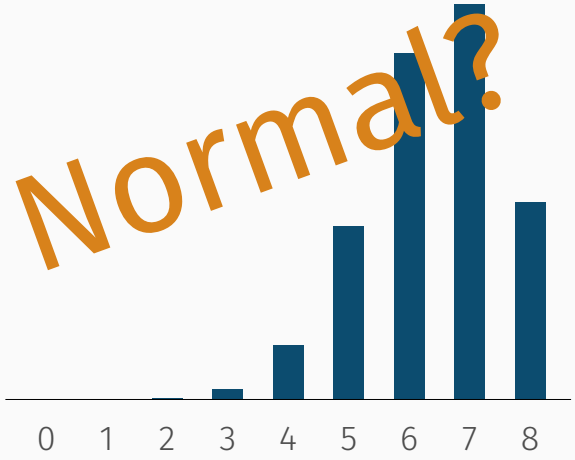




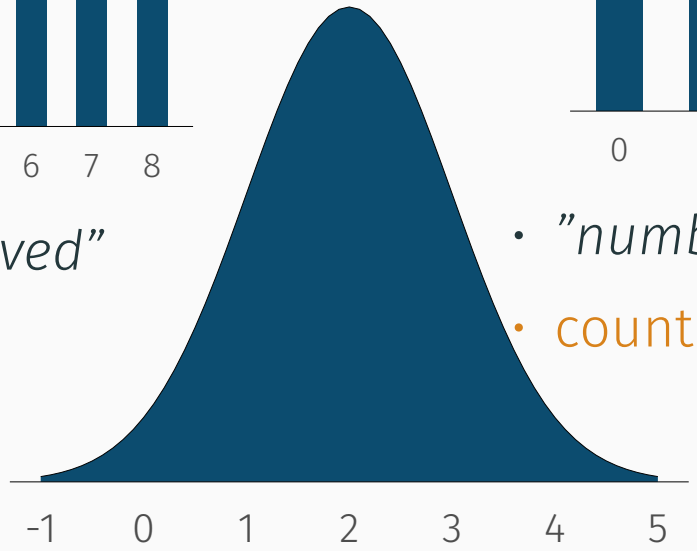
- *"number of animals"*
- count data

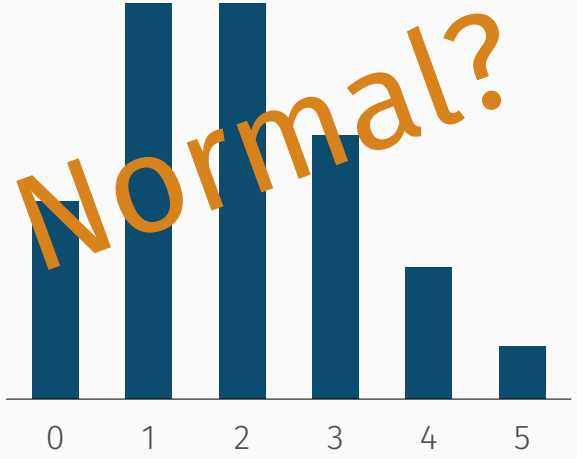
- ignore?

Ecotoxicology is not normal



- *"x out of n survived"*
- binomial data

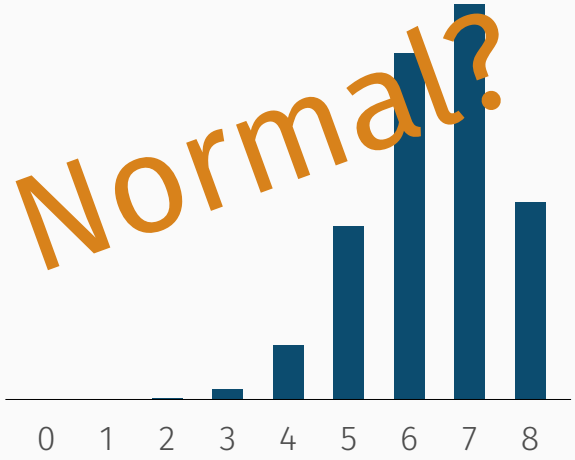




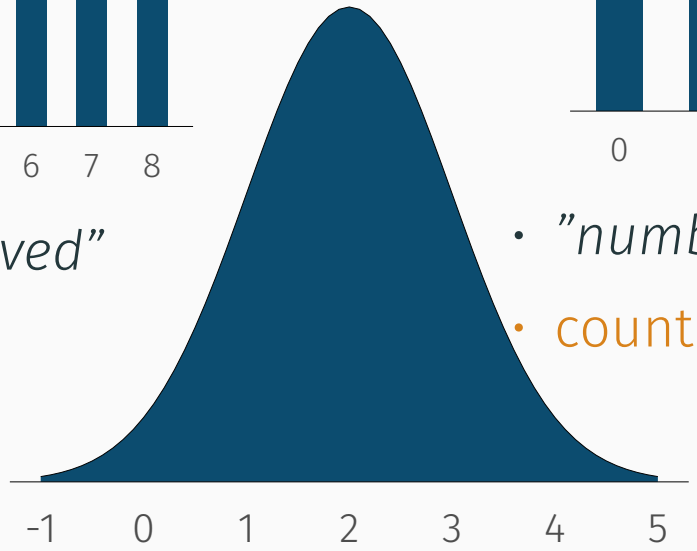
- *"number of animals"*
- count data

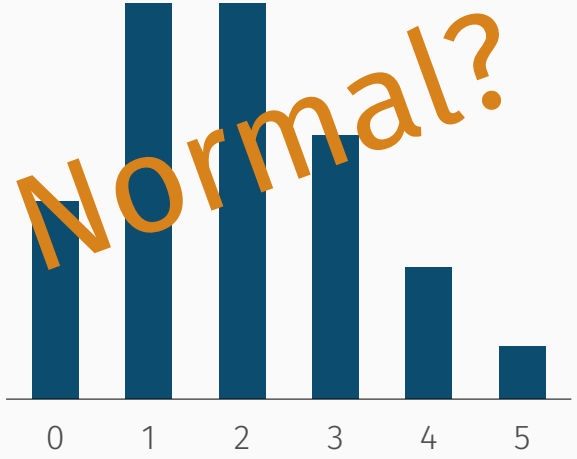
- ignore?
- transform?

Ecotoxicology is not normal



- *"x out of n survived"*
- binomial data

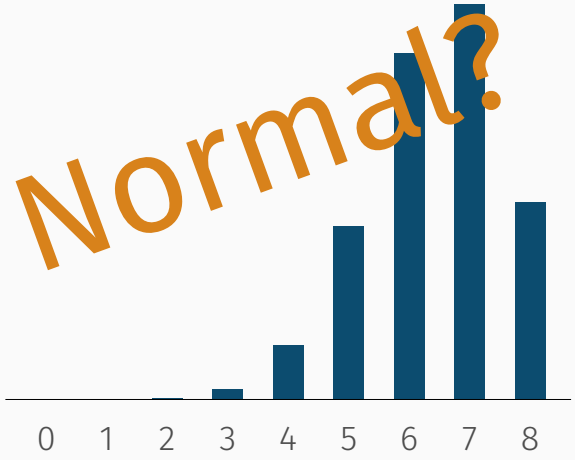




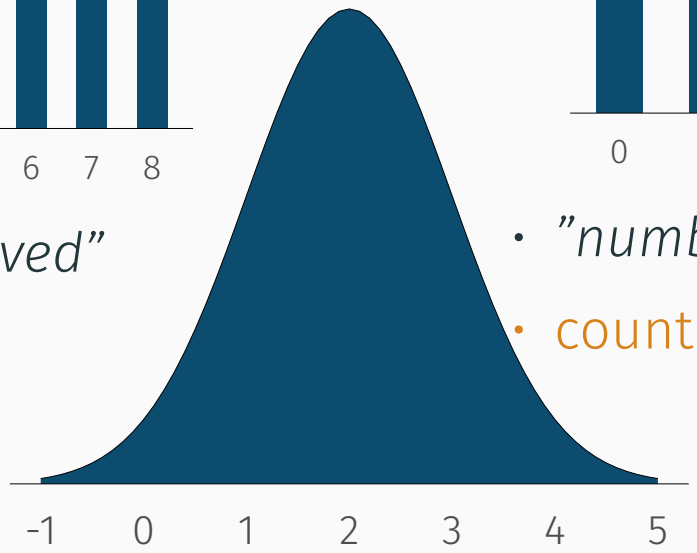
- *"number of animals"*
- count data

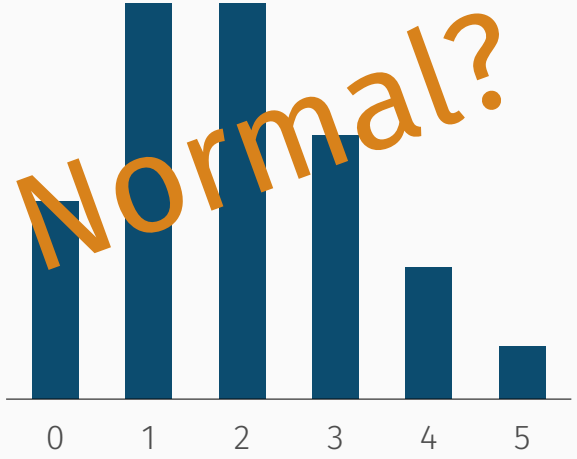
- ignore?
- transform?
- non-parametric?

Ecotoxicology is not normal



- *"x out of n survived"*
- binomial data

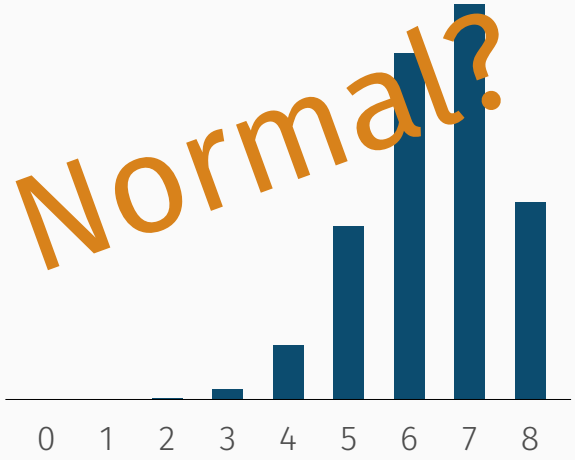




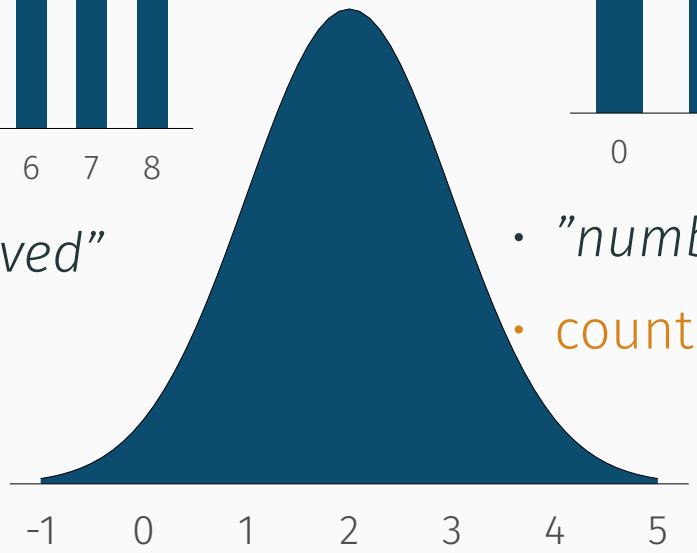
- *"number of animals"*
- count data

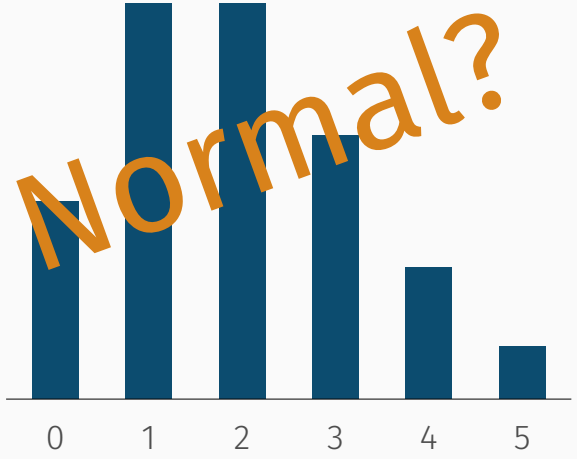
- ignore?
- transform?
- non-parametric?
- Generalized Linear Model (GLM)

Ecotoxicology is not normal



- *"x out of n survived"*
- binomial data





- *"number of animals"*
- count data

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

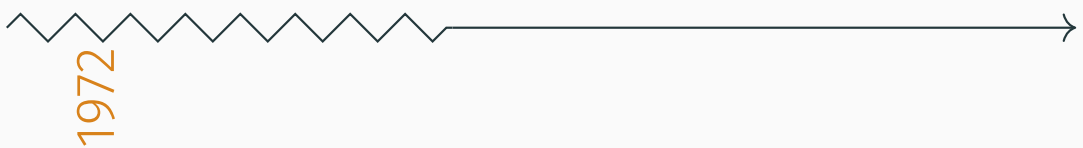
A recent history of GLM (uncomprehensive) in ecology

J. R. Statist. Soc. A,
(1972), **135**, *Part* 3, p. 370

Generalized Linear Models

By J. A. NELDER and R. W. M. WEDDERBURN

Rothamsted Experimental Station, Harpenden, Herts



A recent history of GLM (uncomprehensive) in ecology

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology & Evolution

doi: 10.1111/j.2041-210X.2010.00021.x

Do not log-transform count data

Robert B. O'Hara[1]* and D. Johan Kotze[2]

[1]*Biodiversity and Climate Research Centre, Senckenberganlage 25, D-60325 Frankfurt am Main, Germany and*
[2]*Department of Environmental Sciences, PO Box 65, University of Helsinki, Helsinki FI-00014, Finland*

A recent history of GLM (uncomprehensive) in ecology

Ecology, 92(1), 2011, pp. 3–10

The arcsine is asinine: the analysis of proportions in ecology

DAVID I. WARTON[1,2,3] AND FRANCIS K. C. HUI[1]

[1]*School of Mathematics and Statistics, The University of New South Wales, Sydney, NSW 2052 Australia*
[2]*Evolution and Ecology Research Centre, The University of New South Wales, Sydney, NSW 2052 Australia*

A recent history of GLM (uncomprehensive) in ecology

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution doi: 10.1111/j.2041-210X.2011.00127.x

Distance-based multivariate analyses confound location and dispersion effects

David I. Warton[1]*, Stephen T. Wright[1] and Yi Wang[1,2]

[1]*School of Mathematics and Statistics and Evolution & Ecology Research Centre; and* [2]*School of Computer Science and Engineering, The University of New South Wales, NSW 2052, Australia*

A recent history of GLM (uncomprehensive) in ecology

Ecotoxicology
DOI 10.1007/s10646-015-1421-0

Analysing chemical-induced changes in macroinvertebrate communities in aquatic mesocosm experiments: a comparison of methods

Eduard Szöcs · Paul J. Van den Brink · Laurent Lagadic · Thierry Caquet · Marc Roucaute · Arnaud Auber · Yannick Bayona · Matthias Liess · Peter Ebke · Alessio Ippolito · Cajo J. F. ter Braak · Theo C. M. Brock · Ralf B. Schäfer

A recent history of GLM (uncomprehensive) in ecology

Environ Sci Pollut Res
DOI 10.1007/s11356-015-4579-3

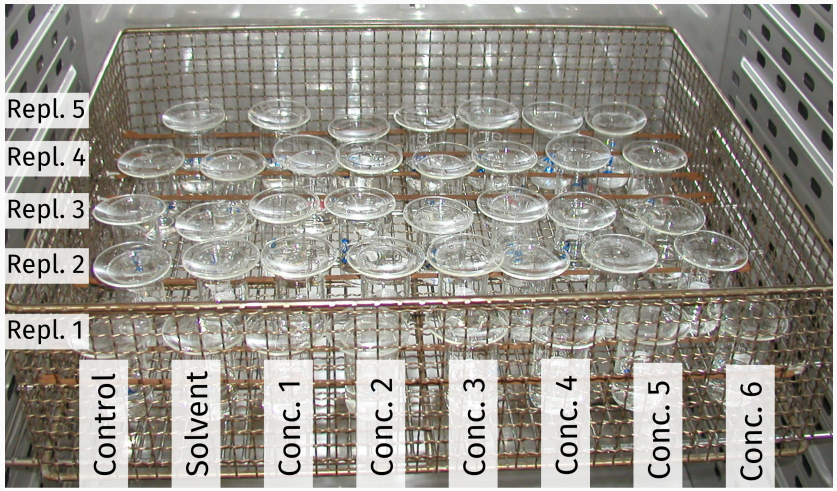
RESEARCH ARTICLE

Ecotoxicology is not normal

A comparison of statistical approaches for analysis of count and proportion data in ecotoxicology

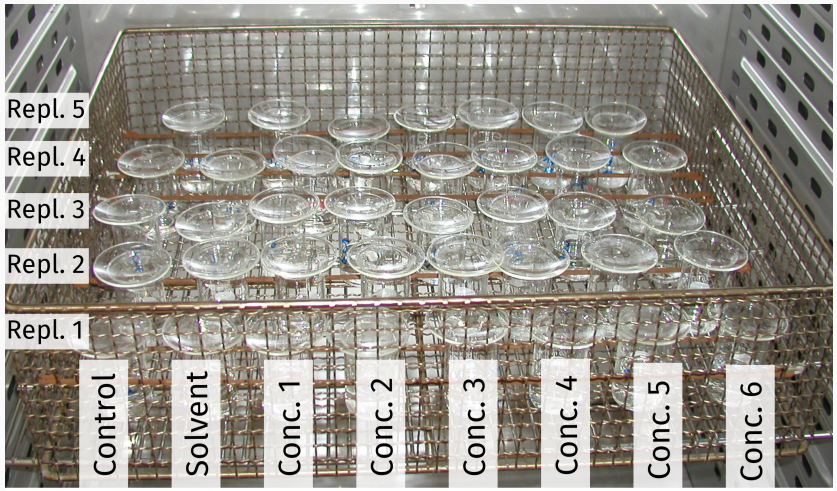
Eduard Szöcs[1] · Ralf B. Schäfer[1]

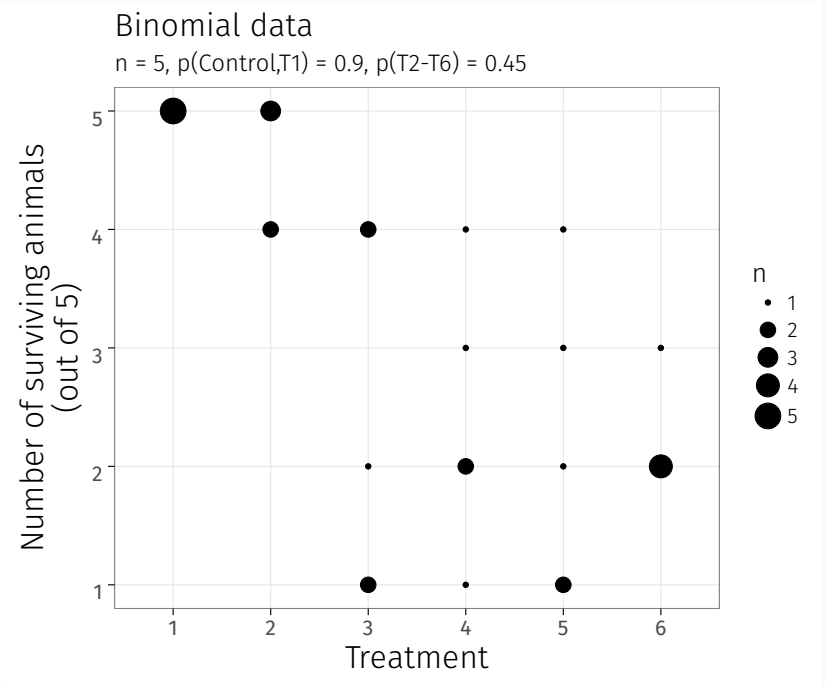
A simulation study



- Count & Binomial data
- Vary replicates & Effect size
- Global test & LOEC

A simulation study

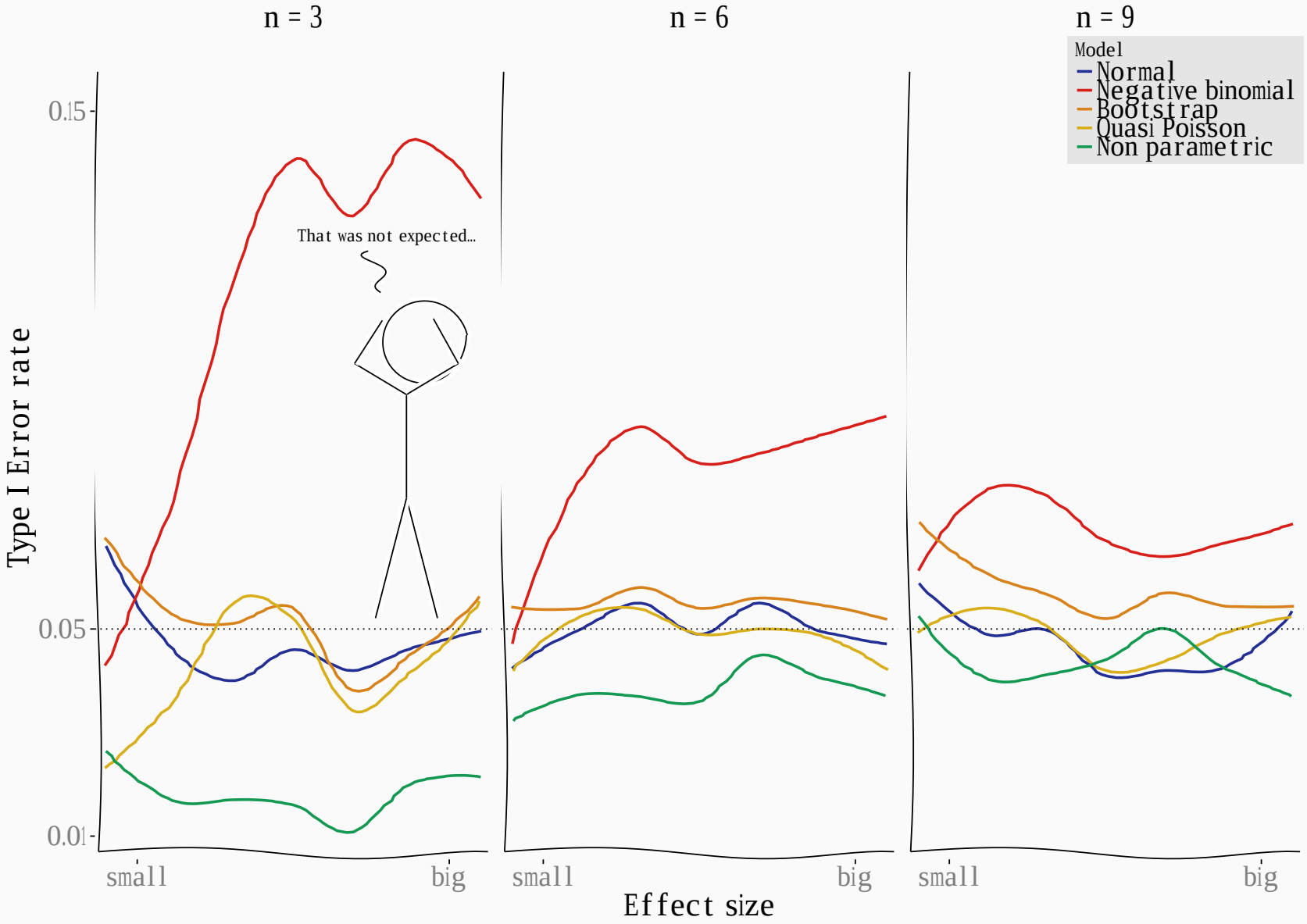




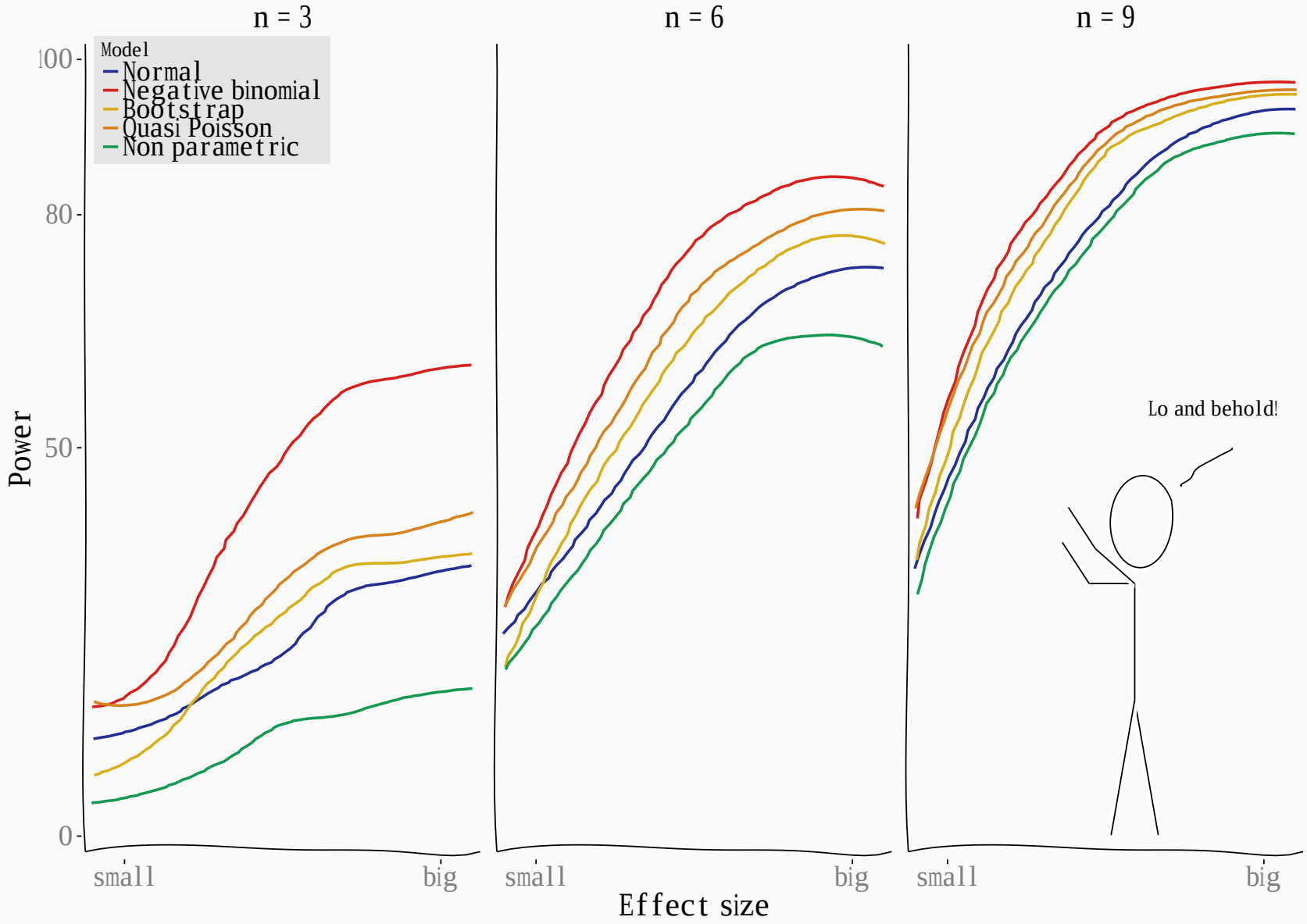
- Count & Binomial data
- Vary replicates & Effect size
- Global test & LOEC

- 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. Ecotoxicological experiments commonly low power
4. NOECs are not reliable
5. GLMs can increase this power

Where are we today?

Three days earlier...

Where are we today?

Three days earlier...

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2015, **6**, 828–835 doi: 10.1111/2041-210X.12386

For testing the significance of regression coefficients, go ahead and log-transform count data

Anthony R. Ives*

Department of Zoology, University of Wisconsin-Madison, Madison, WI 53706, USA

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*

Where are we today?

Three points to consider ...

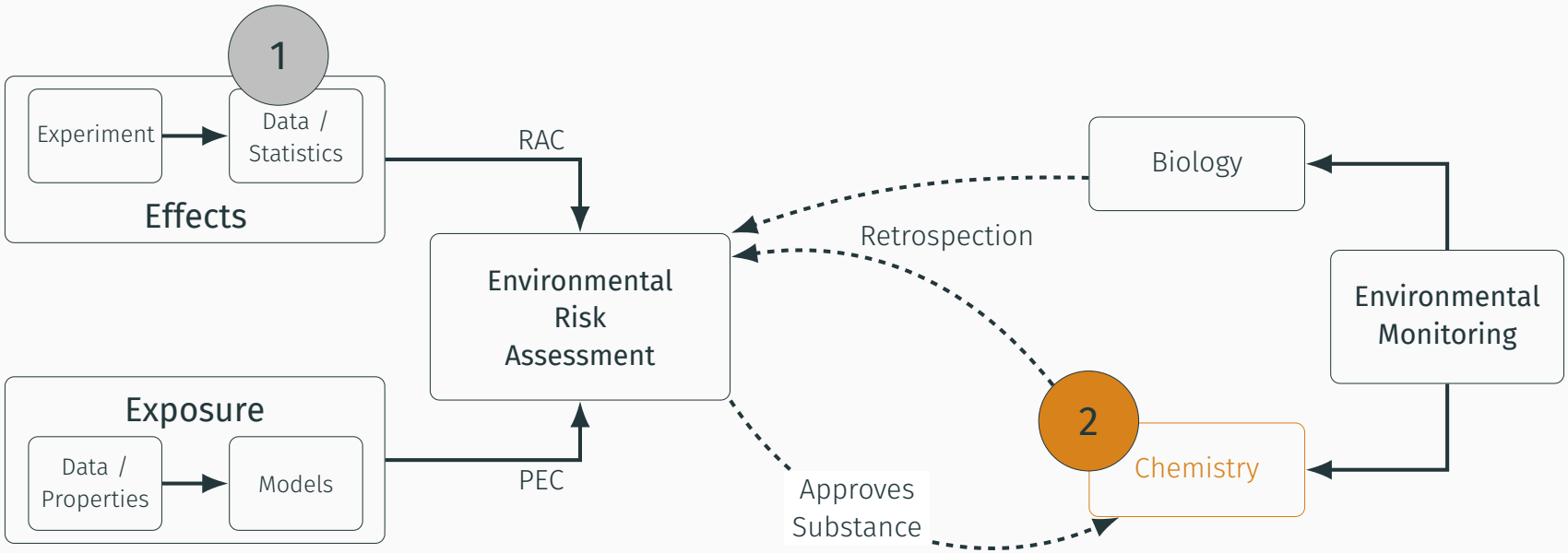
1. Choose your model based on data properties
2. Fix Type I errors by resampling
3. Models that better fit the data have better power properties

Warton, D.I., Lyons, M., Stoklosa, J., Ives, A.R., 2016. Three points to consider when choosing a LM or GLM test for count data. Methods in Ecology and Evolution 7, 882–890.

Exploring Monitoring Data for ERA

Plant Protection Products Regulation 1107/2009

Water Framework Directive 2000/60/EC



2 Szöcs, Brinke, Karaoglan & Schäfer (submitted). “Large scale risks from pesticides in small streams”. Environmental Science & Technology.

Exploring Monitoring Data for ERA

Exposure Assessment & Environmental Monitoring

Risk Assessment

- prospective

Chemical Monitoring

- retrospective

Brock et al. (2006). Aquatic risks of pesticides, ecological protection goals and common aims in EU legislation. IEAM 2, e20–e46.

Exposure Assessment & Environmental Monitoring

Risk Assessment

- prospective
- Plant Protection Products

Chemical Monitoring

- retrospective
- All chemicals

Brock et al. (2006). Aquatic risks of pesticides, ecological protection goals and common aims in EU legislation. IEAM 2, e20–e46.

Exposure Assessment & Environmental Monitoring

Risk Assessment

- prospective
- Plant Protection Products
- small waters
  (1m width, 30cm depth)

Chemical Monitoring

- retrospective
- All chemicals
- large waters
  (mainly >10km^2)

Brock et al. (2006). Aquatic risks of pesticides, ecological protection goals and common aims in EU legislation. IEAM 2, e20–e46.

Exposure Assessment & Environmental Monitoring

Risk Assessment

- prospective
- Plant Protection Products
- small waters
  (1m width, 30cm depth)
- Regulatory Acceptable concentrations (RAC)

Chemical Monitoring

- retrospective
- All chemicals
- large waters
  (mainly >10km^2)
- Environmental Quality Standards (EQS)

Brock et al. (2006). Aquatic risks of pesticides, ecological protection goals and common aims in EU legislation. IEAM 2, e20–e46.

Goals & Hypotheses

Goal: Can monitoring inform ERA?

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

Goals & Hypotheses

Goal: Can monitoring inform ERA?

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

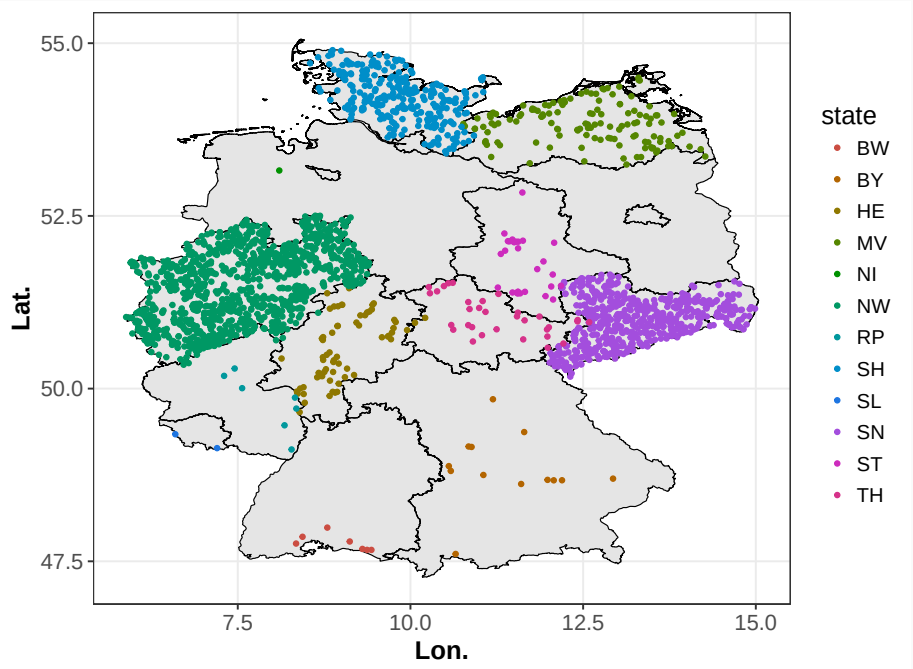
Hypotheses

1. Agricultural streams show highest concentrations
2. Small streams show highest concentrations
3. Precipitation at/before sampling increases concentrations
4. Highest concentrations in summer

Compiled data: Big, but inhomogeneous

- Federalism
- Compilation & Homogenisation

- 1,766,104 measurements
- 478 pesticides (including metabolites)
- 24,743 samples
- 2,301 sites



Analysing chemical concentrations

Analysing chemical concentrations

- concentrations < LOQ (96% of all measurements)

Analysing chemical concentrations

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

$$RQ_i \sim ZAGA(\mu_i, \sigma, \nu_i) = \begin{cases} (1 - \nu_i) & \text{if } y < LOQ \\ \nu_i \times f_{Gamma}(\mu_i, \sigma) & \text{if } y \geq LOQ \end{cases} \quad (1)$$

Analysing chemical concentrations

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

$$RQ_i \sim ZAGA(\mu_i, \sigma, \nu_i) = \begin{cases} (1 - \nu_i) & \text{if } y < LOQ \\ \nu_i \times f_{Gamma}(\mu_i, \sigma) & \text{if } y \geq LOQ \end{cases} \quad (1)$$

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

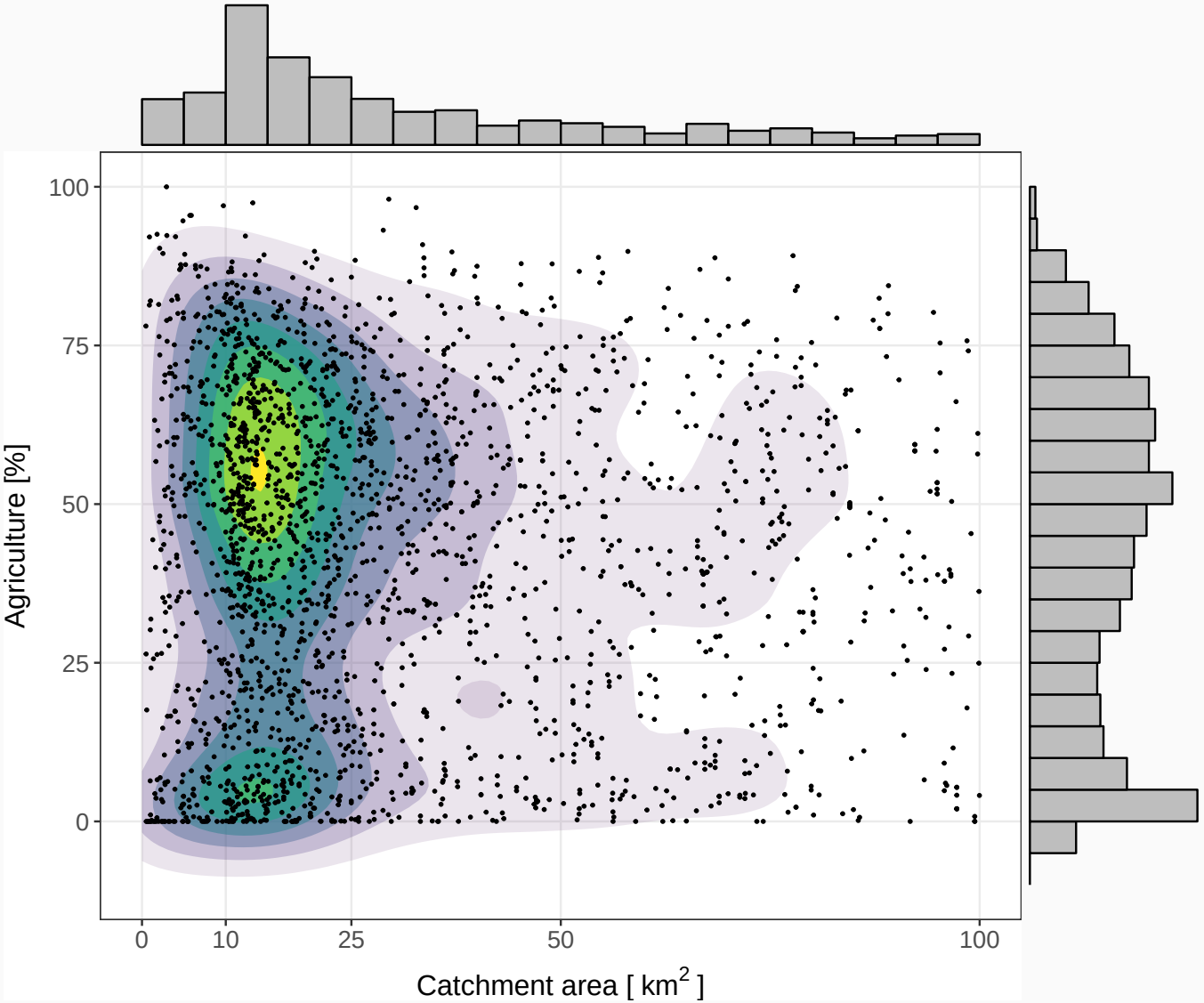
Analysing chemical concentrations

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

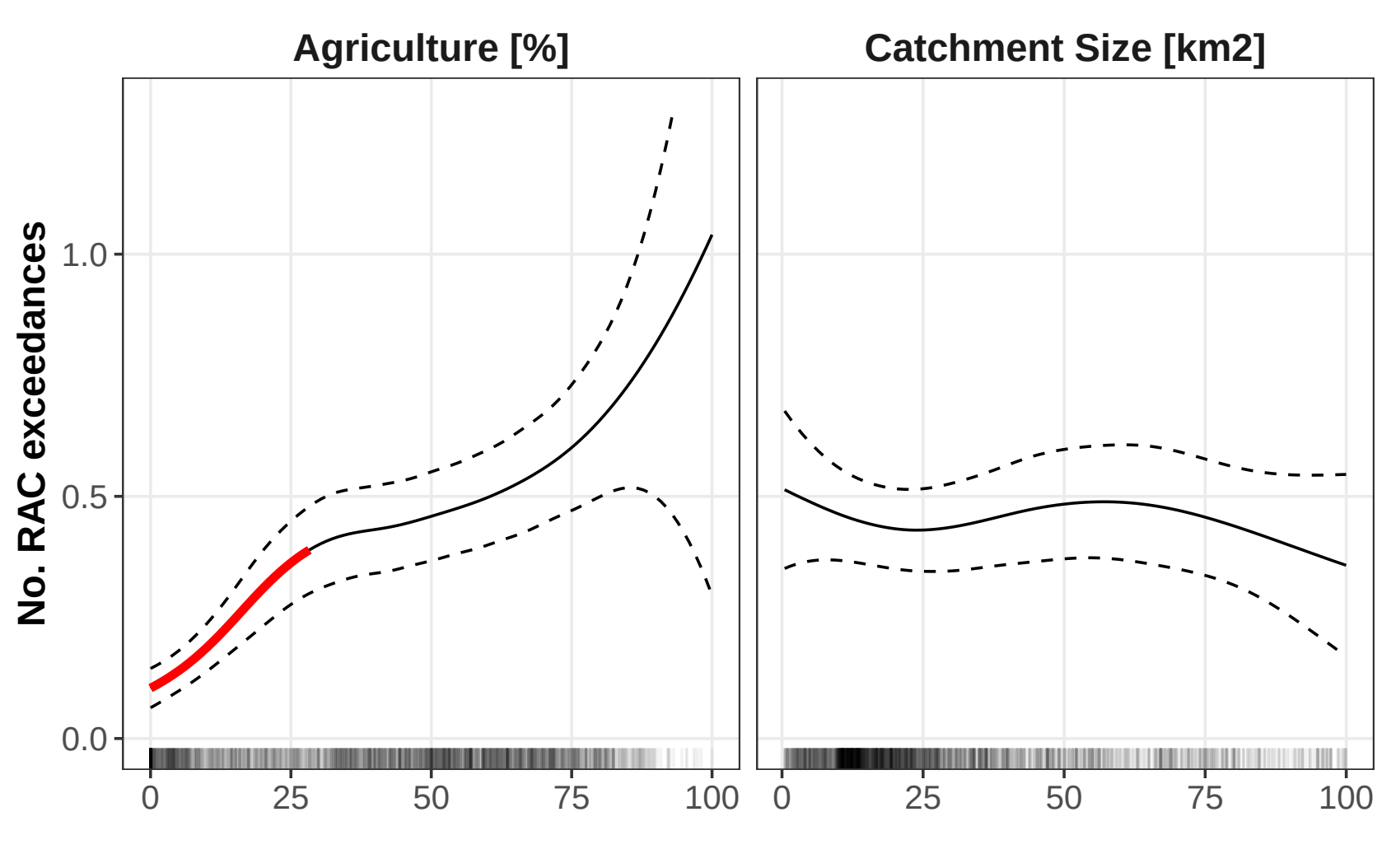
$$RQ_i \sim ZAGA(\mu_i, \sigma, \nu_i) = \begin{cases} (1 - \nu_i) & \text{if } y < LOQ \\ \nu_i \times f_{Gamma}(\mu_i, \sigma) & \text{if } y \geq LOQ \end{cases} \quad (1)$$

- Risk Quotient: $RQ = \frac{C}{RAC}$
- Catchment Size (DEM + Authorities)
- Agriculture (ATKIS)
- Weather (DWD)

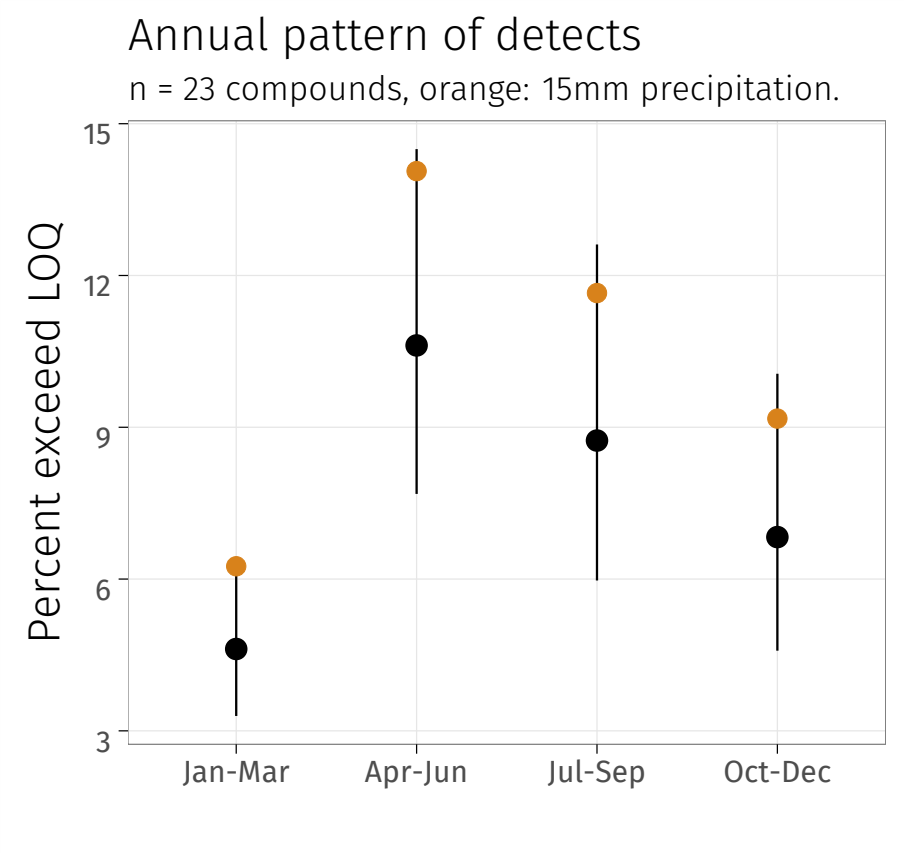
Monitoring: Small streams are underrepresented



Landscape: Factors influencing risk

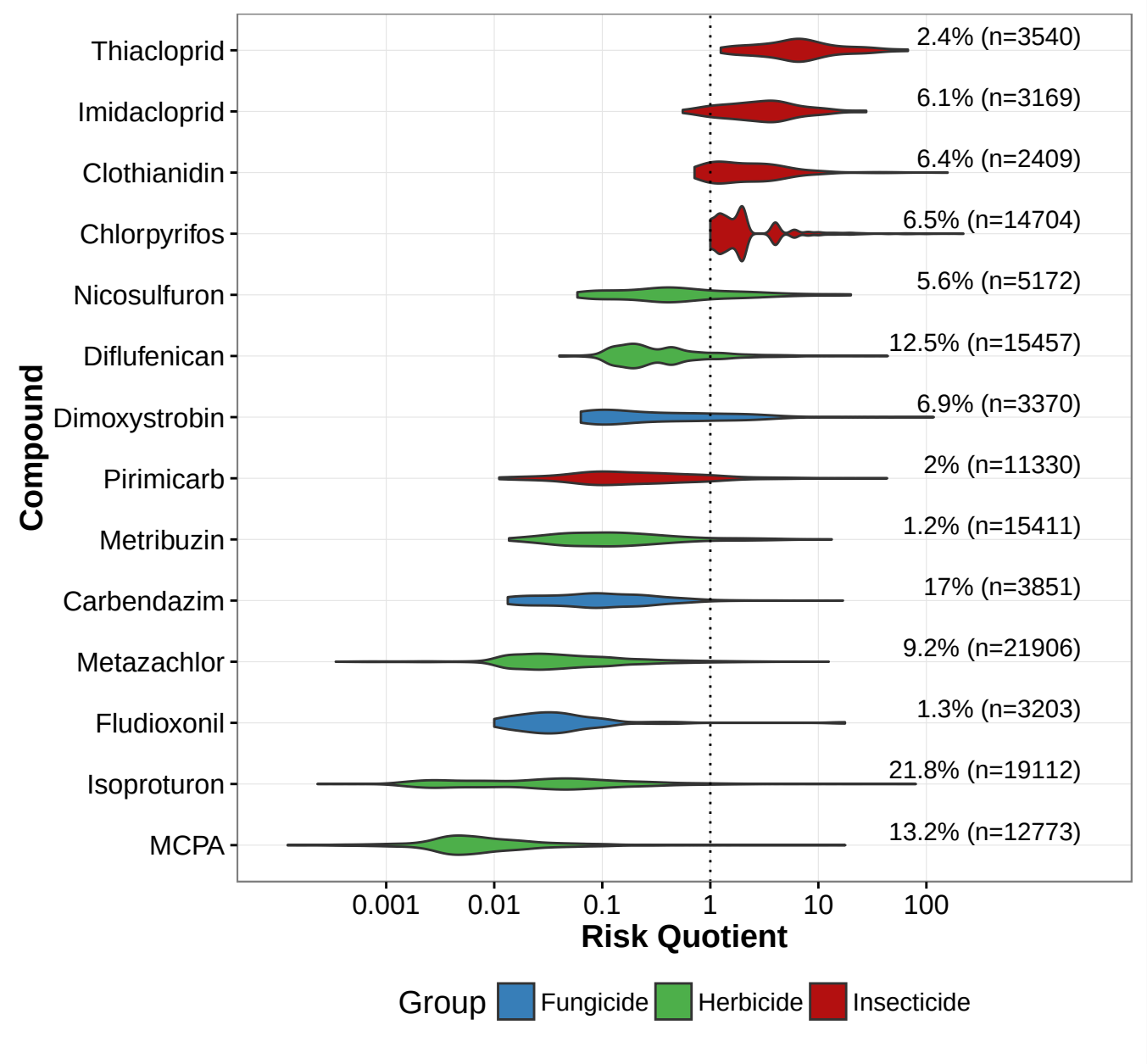


Sampling: Factors influencing risk



- Peak in summer
- Increase by precipitation
  - day before sampling
  - not at day of sampling
- compound specific
- absolute concentrations show much higher variability

Risks: Compounds exceeding risk thresholds

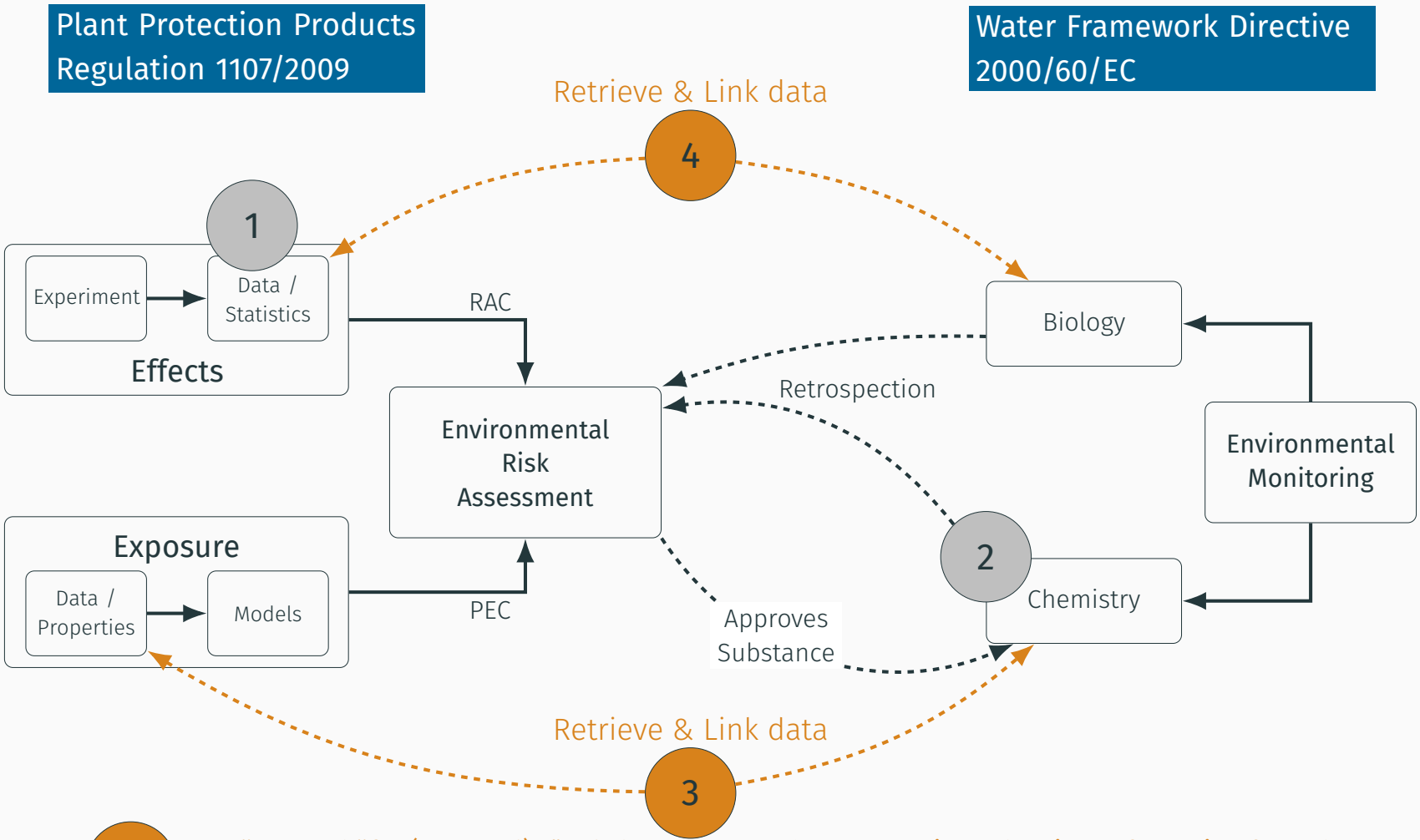


What we learned

1. Huge differences between states
2. Small streams currently underrepresented
3. Agricultural sources
4. LOQ gives additional insights
    - Annual dynamics
    - Precipitation increases concentrations
5. Risk underestimated
6. Neonicotinoids + Chlorpyrifos + others

Solutions for Data Handling in ERA

Solutions for Data Handling in ERA



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

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

Biologists & Chemists face the same problems

Ziemann, M., Eren, Y., El-Osta, A., 2016. Gene name errors are widespread in the scientific literature. Genome Biology 17.

Biologists & Chemists face the same problems

Names

Osmia rufa, Osmia bicornis, Osmia ruffa, Osmia unilandauis, Osmia spec.

Chlorpyrifos, Chlorpyriphos, Chlorphyrifos, Chlorpyrifos-ethyl, Chlorpypifot

Ziemann, M., Eren, Y., El-Osta, A., 2016. Gene name errors are widespread in the scientific literature. Genome Biology 17.

Biologists & Chemists face the same problems

Names

Osmia rufa, Osmia bicornis, Osmia ruffa, Osmia unilandauis, Osmia spec.

Chlorpyrifos, Chlorpyriphos, Chlorphyrifos, Chlorpyrifos-ethyl, Chlorpypifot

Hierarchies

Hymenoptera/ Apoidea/ Megachilidae/ Osmia/ rufa

organophospate, ester, insecticide

Ziemann, M., Eren, Y., El-Osta, A., 2016. Gene name errors are widespread in the scientific literature. Genome Biology 17.

Biologists & Chemists face the same problems

Names

Osmia rufa, Osmia bicornis, Osmia ruffa, Osmia unilandauis, Osmia spec.

Chlorpyrifos, Chlorpyriphos, Chlorphyrifos, Chlorpyrifos-ethyl, Chlorpypifot

Hierarchies

Hymenoptera/ Apoidea/ Megachilidae/ Osmia/ rufa

organophospate, ester, insecticide

Traits / Properties

Wing length, Mass, Season

Mass, K_{OW}, LC_{50}

Ziemann, M., Eren, Y., El-Osta, A., 2016. Gene name errors are widespread in the scientific literature. Genome Biology 17.

Biologists & Chemists face the same problems

Names

Osmia rufa, Osmia bicornis, Osmia ruffa, Osmia unilandauis, Osmia spec.

Chlorpyrifos, Chlorpyriphos, Chlorphyrifos, Chlorpyrifos-ethyl, 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, Clc1c(OP(=S)[...], InChI=1S/C9H11C[...], SBPBAQFW[...], CSID,...

Ziemann, M., Eren, Y., El-Osta, A., 2016. Gene name errors are widespread in the scientific literature. Genome Biology 17.

Biologists & Chemists face the same problems

Names

Osmia rufa, Osmia bicornis, Osmia ruffa, Osmia unilandauis, Osmia spec.

Chlorpyrifos, Chlorpyriphos, Chlorphyrifos, Chlorpyrifos-ethyl, 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, Clc1c(OP(=S)[...], InChI=1S/C9H11C[...], SBPBAQFW[...], CSID,...

Amount of data

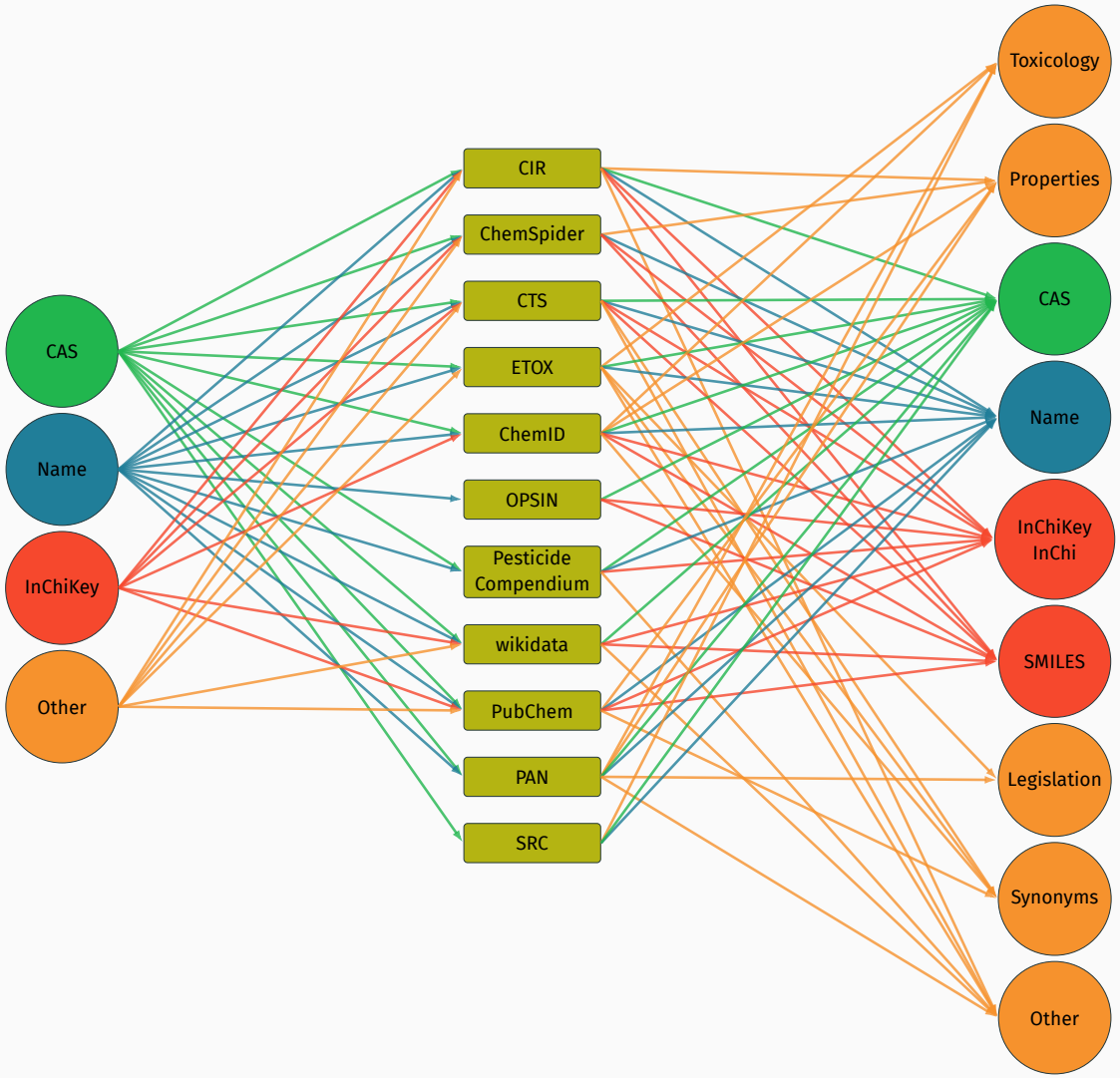
2993 taxa

489 pesticides (+ 590 other organics)

Ziemann, M., Eren, Y., El-Osta, A., 2016. Gene name errors are widespread in the scientific literature. Genome Biology 17.

Instead of wasting time...

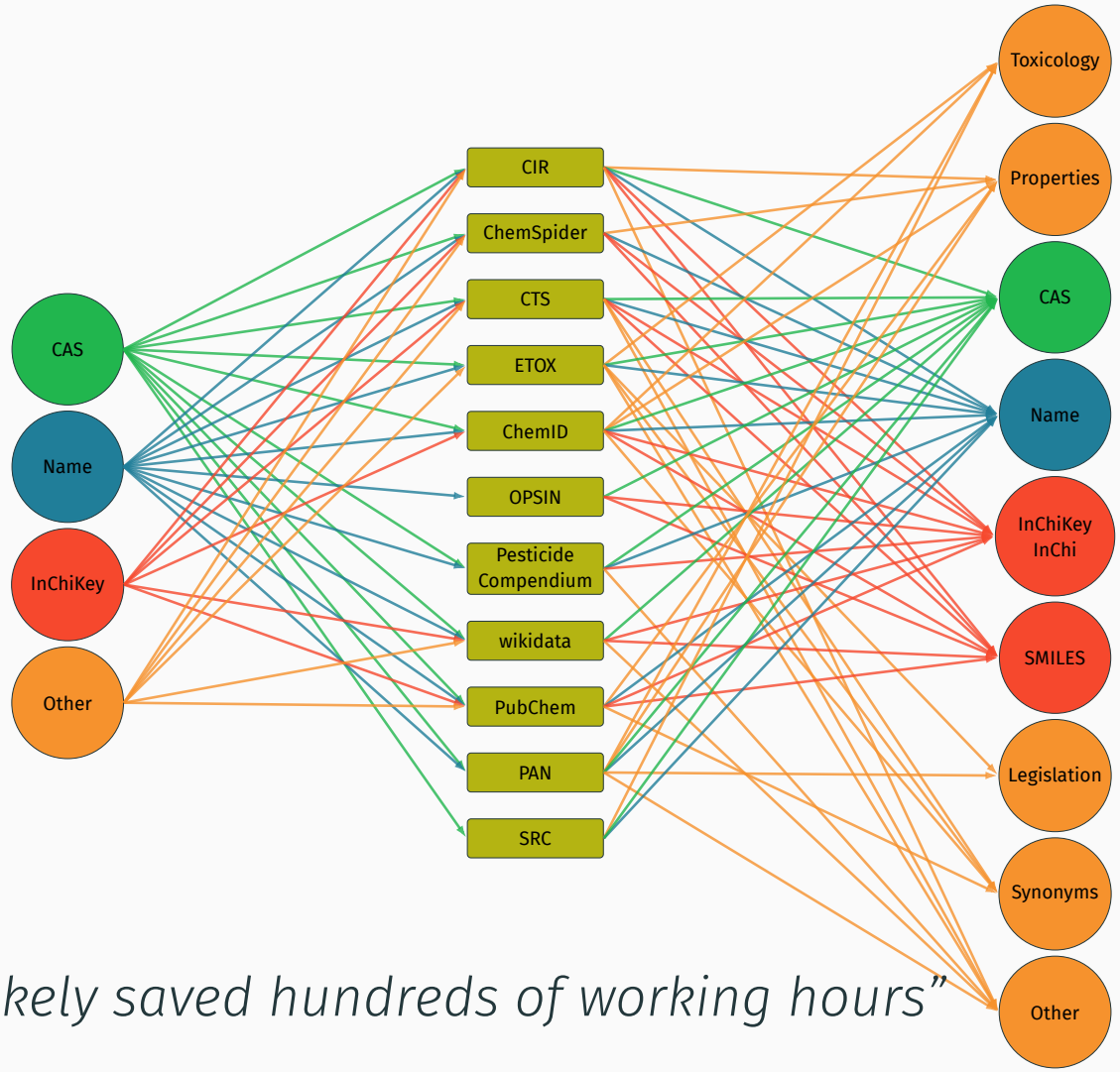
... use webchem!



Münch et al. (2016). DoOR 2.0 - Comprehensive Mapping of Drosophila melanogaster Odorant Responses. Scientific Reports 6, 21841

Instead of wasting time...

... use webchem!

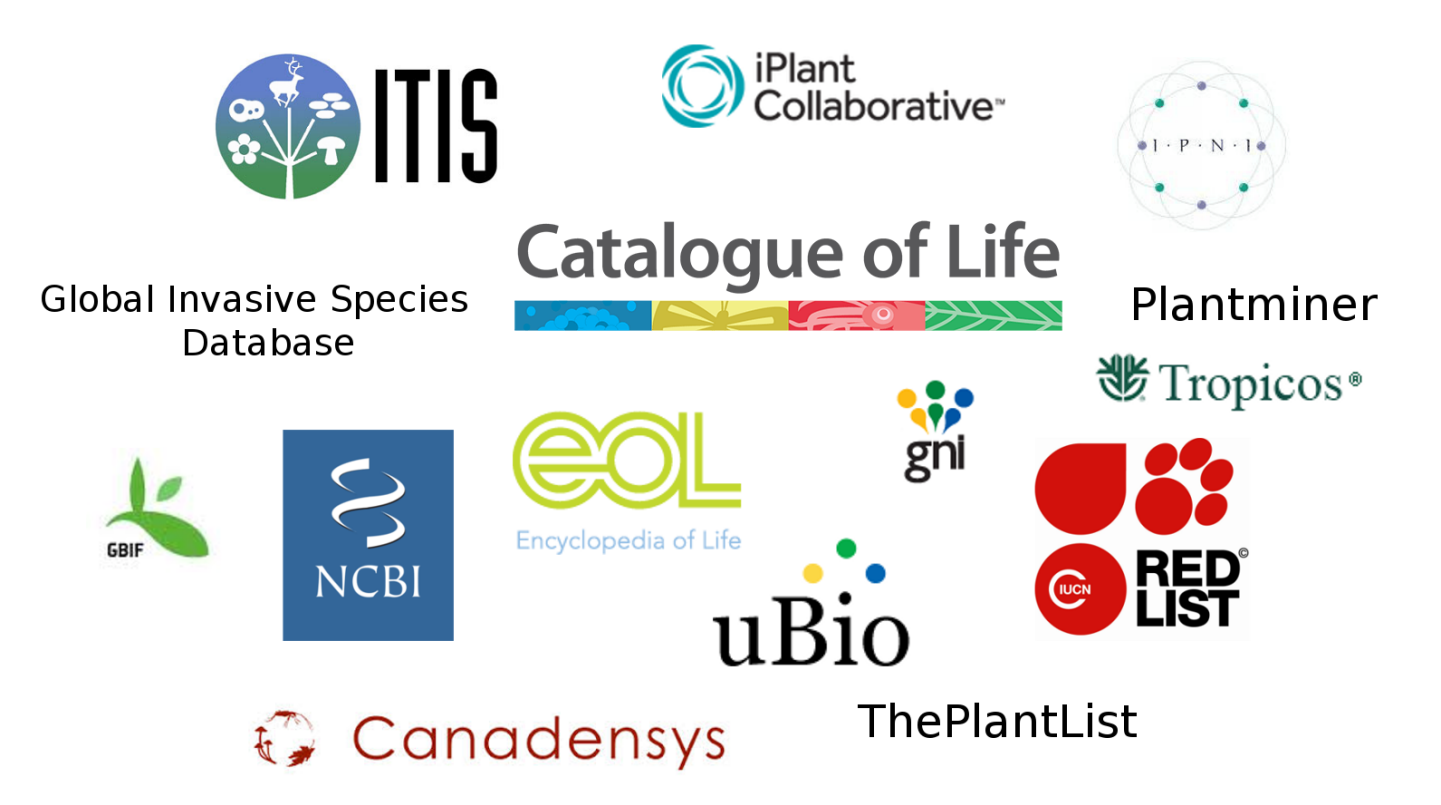


"webchem ...likely saved hundreds of working hours"

Münch et al. (2016). DoOR 2.0 - Comprehensive Mapping of Drosophila melanogaster Odorant Responses. Scientific Reports 6, 21841

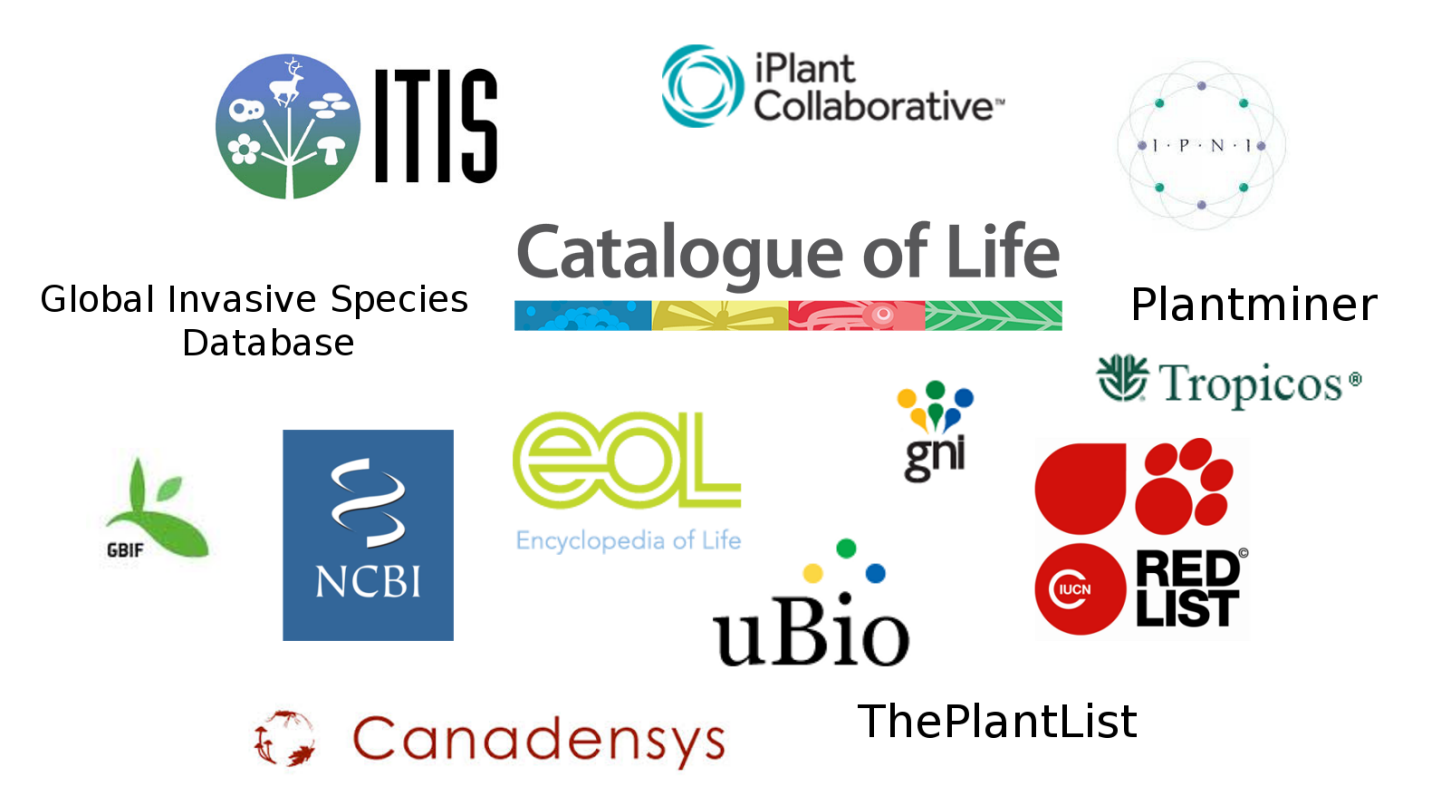
Instead of wasting time...

... use taxize!



Instead of wasting time...

... use taxize!



"Days of searching done during my morning coffee. Amazing. taxize."

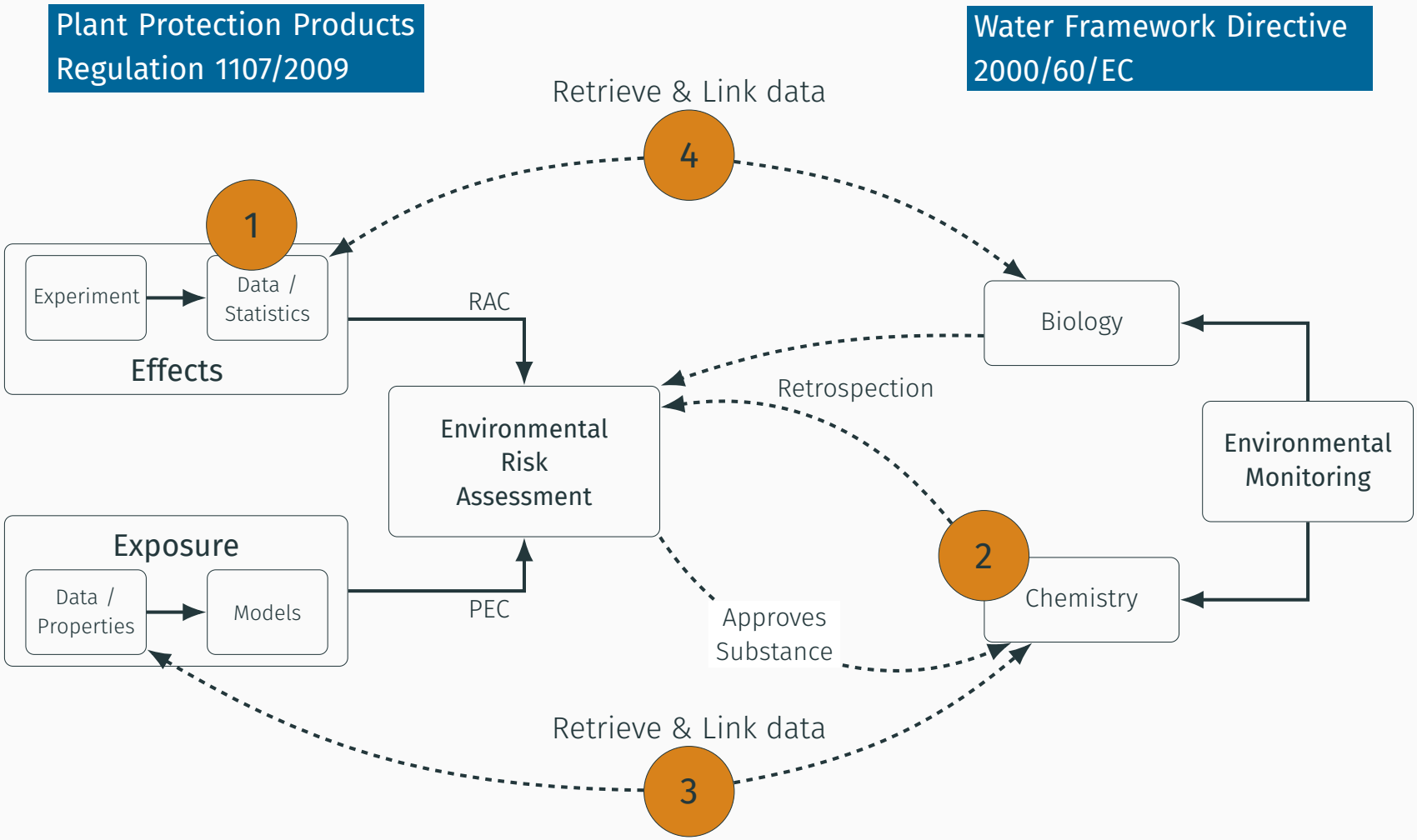
Dr. Susan E. Johnston, University of Edinburgh. On twitter.

taxize & webchem: Advances for you (researchers)

1. saving of time
2. less errors
3. reproducibility
4. join different data sets
5. easy aggregations

Recap

Recap: What did I look at?



Recap: What we learned

✓Improving Statistics in ERA

- Change your model, not your data
- Ultimately ban NOEC
- Take LOQ into account

Recap: What we learned

✓Improving Statistics in ERA

- Change your model, not your data
- Ultimately ban NOEC
- Take LOQ into account

✓Exploring Monitoring Data for ERA

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

Recap: What we learned

✓Improving Statistics in ERA

- Change your model, not your data
- Ultimately ban NOEC
- Take LOQ into account

✓Exploring Monitoring Data for ERA

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

✓Solutions for Linking Data in ERA

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

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)

Statistical Ecotoxicology

Improving the Utilisation of Data for Environmental Risk Assessment

Eduard Szöcs

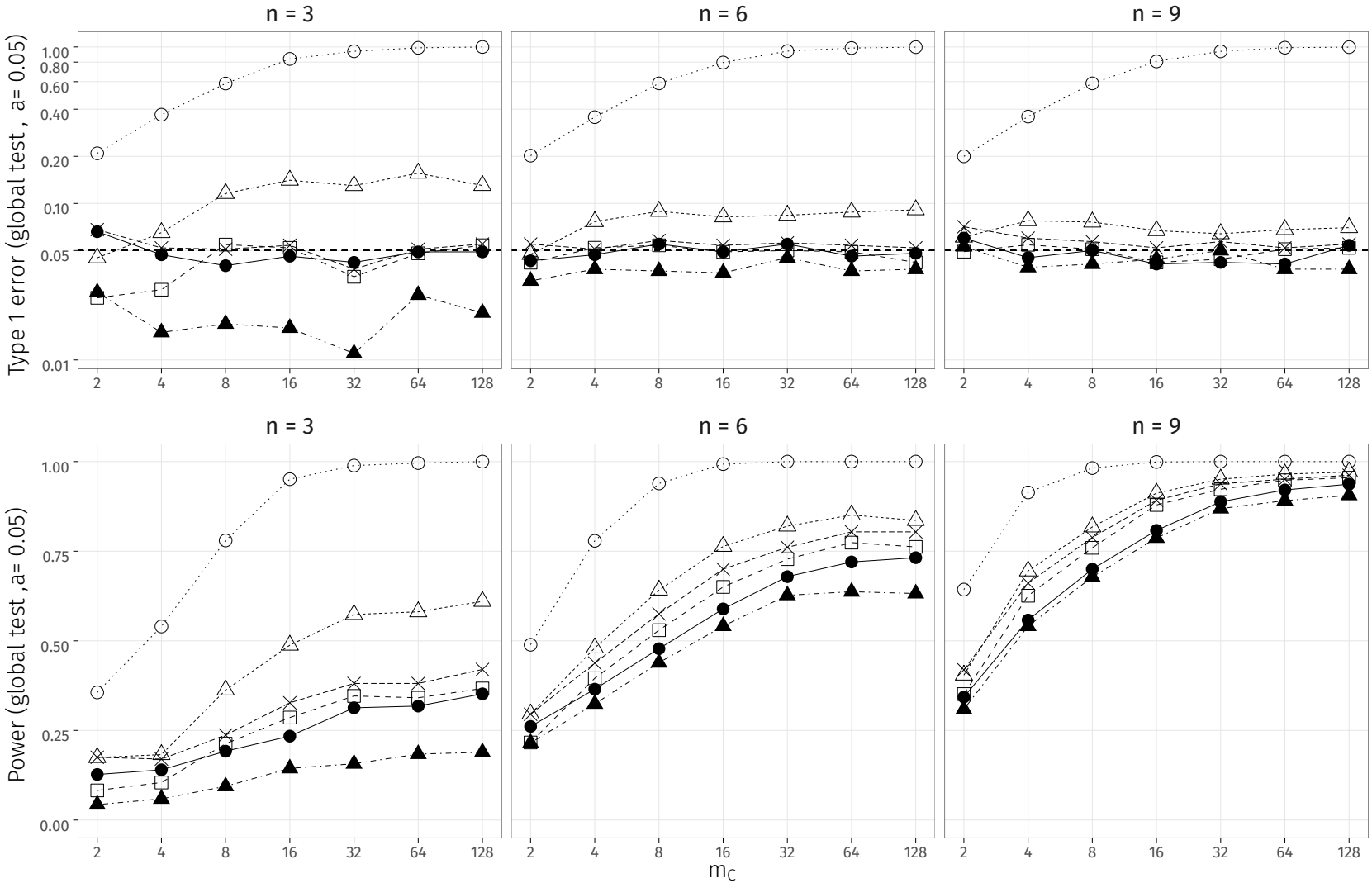
http://edild.github.io/

@EduardSzoecs

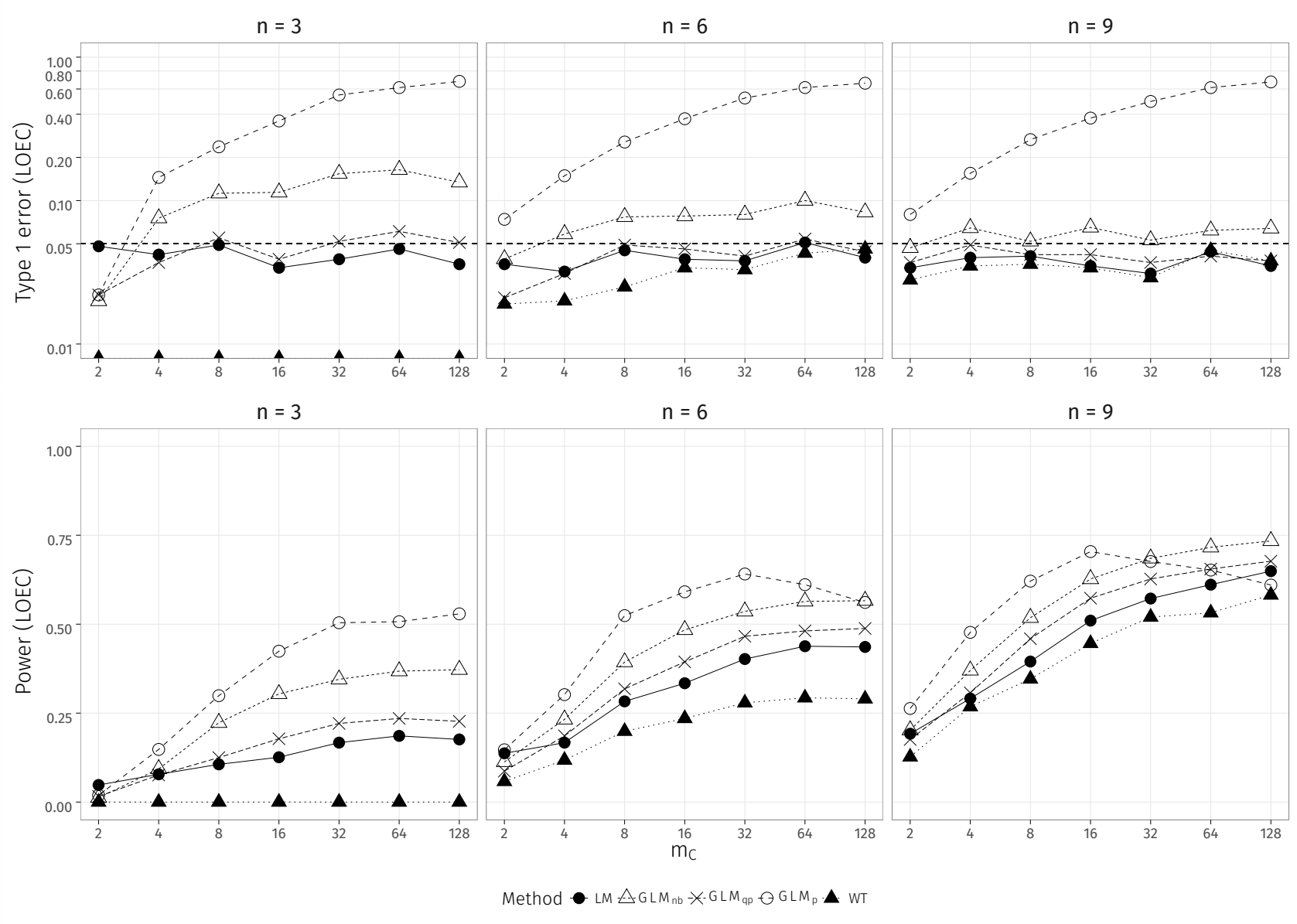
https://github.com/edild/phd_defense

https://github.com/edild/phd_thesis

Power en detail



For LOEC it is even worse



Comparison with Ives...

Szöcs (2015)

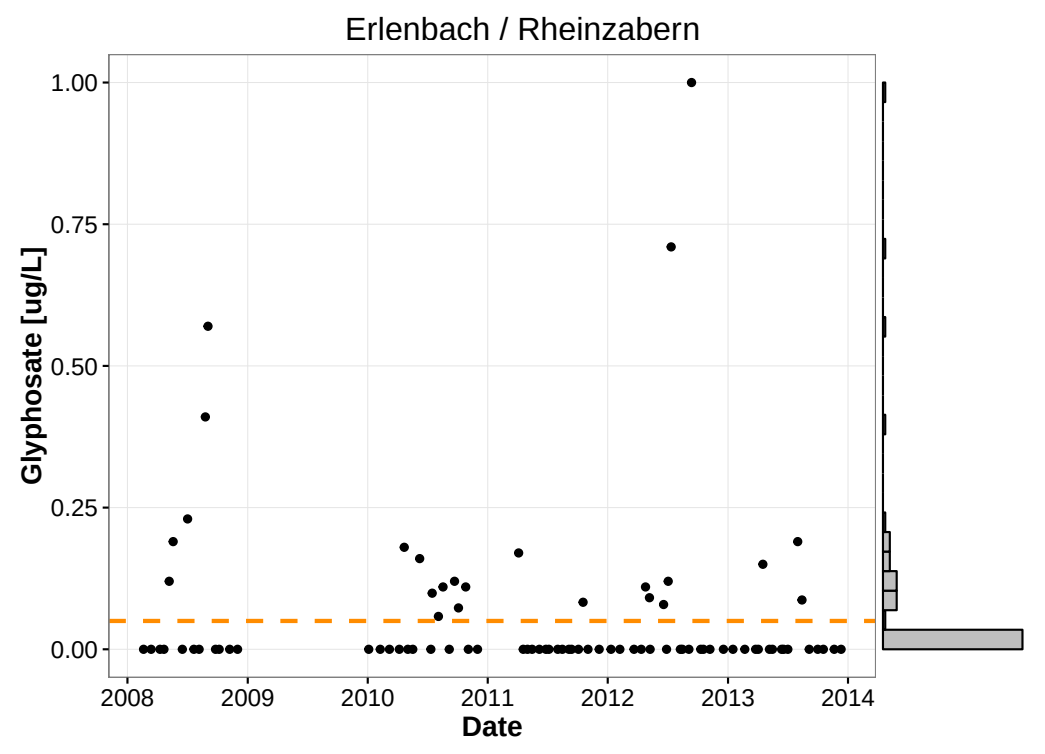
- factorial design
- one predictor
- low replicated
- LM, GLM, bootstrap
- High T1 error of NB
- Quasi-Poisson worked well
- Bootstrap fixes the problems

Ives (2015)

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

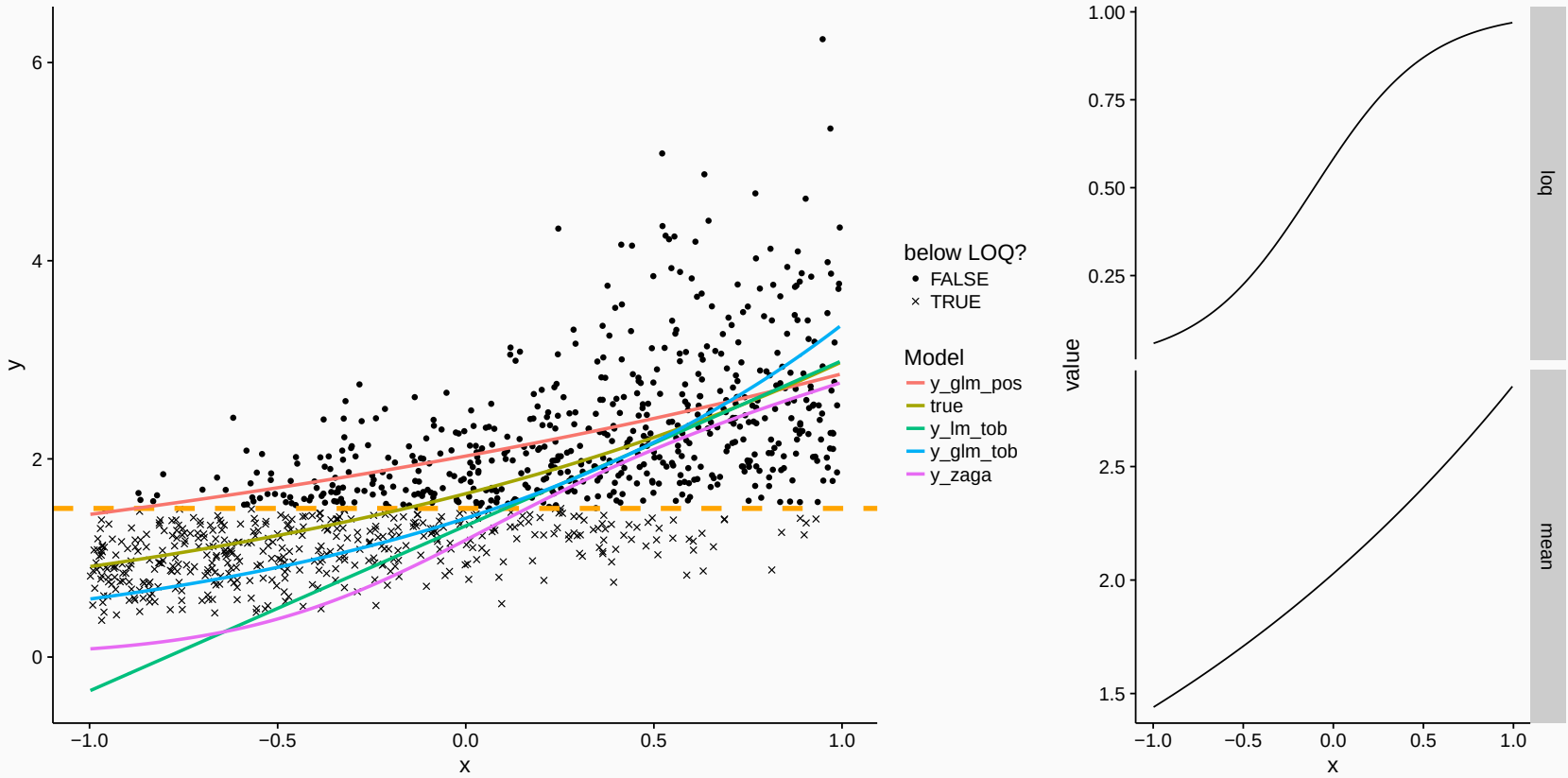
Idiosyncrasies of chemical concentrations

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

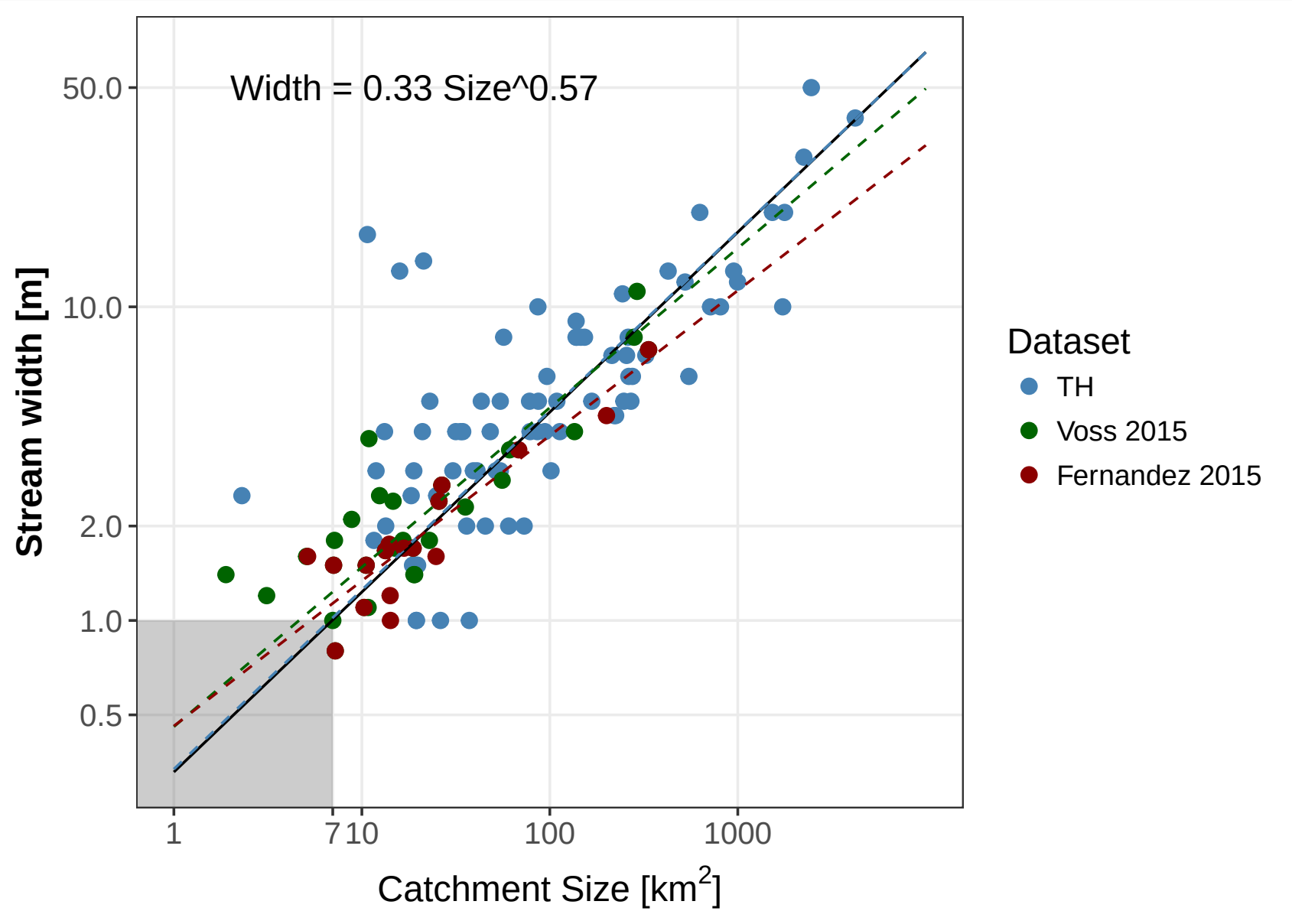


ZAGA what...?

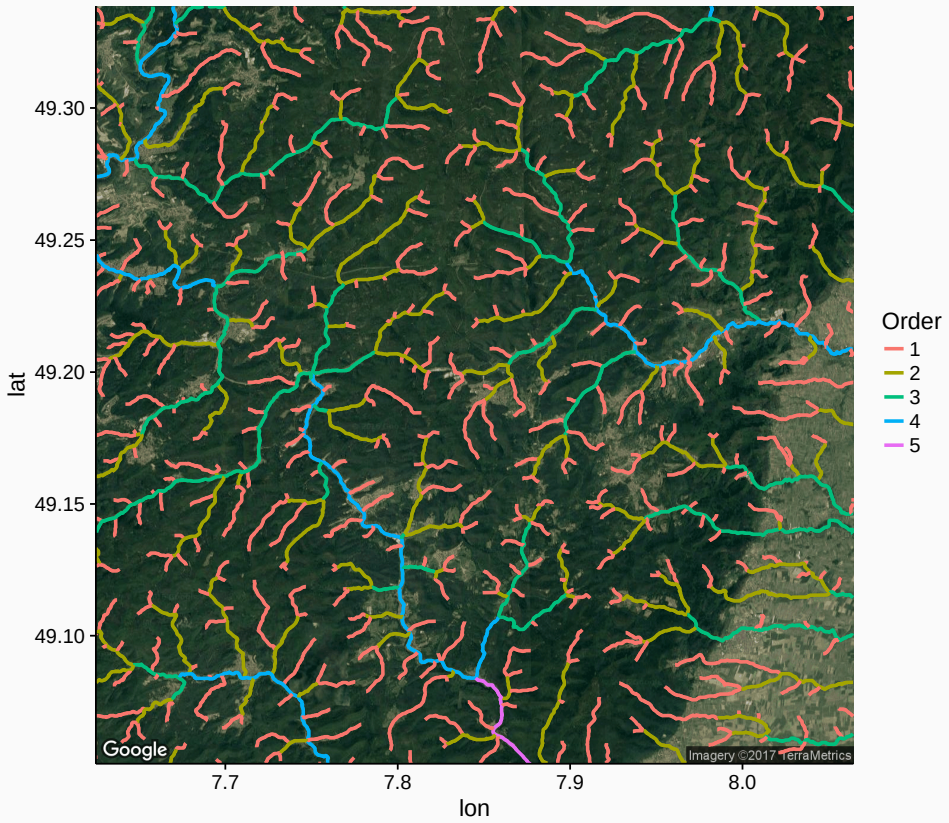
shiny app: `http://uni-ko-ld.de/g4`

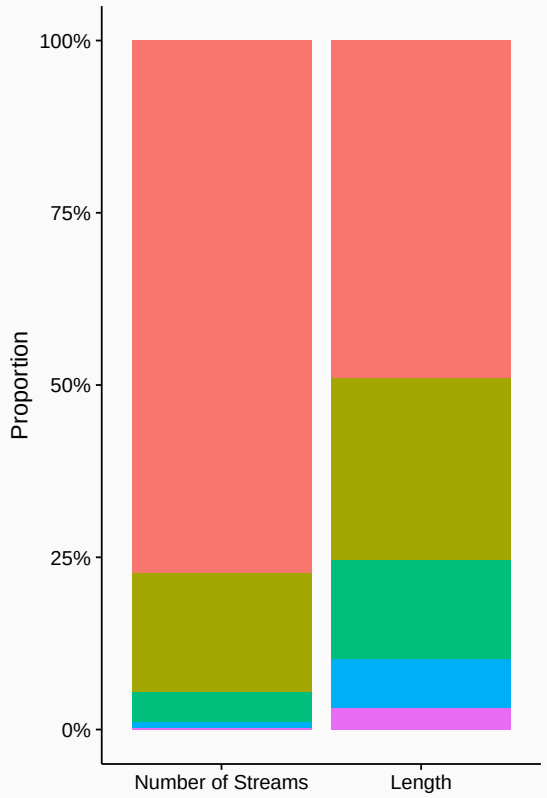


Stream size / width



Importance of small streams





- Biodiversity
- Refuge for re-colonisation

Comparison with other studies

Szöcs (2016)

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

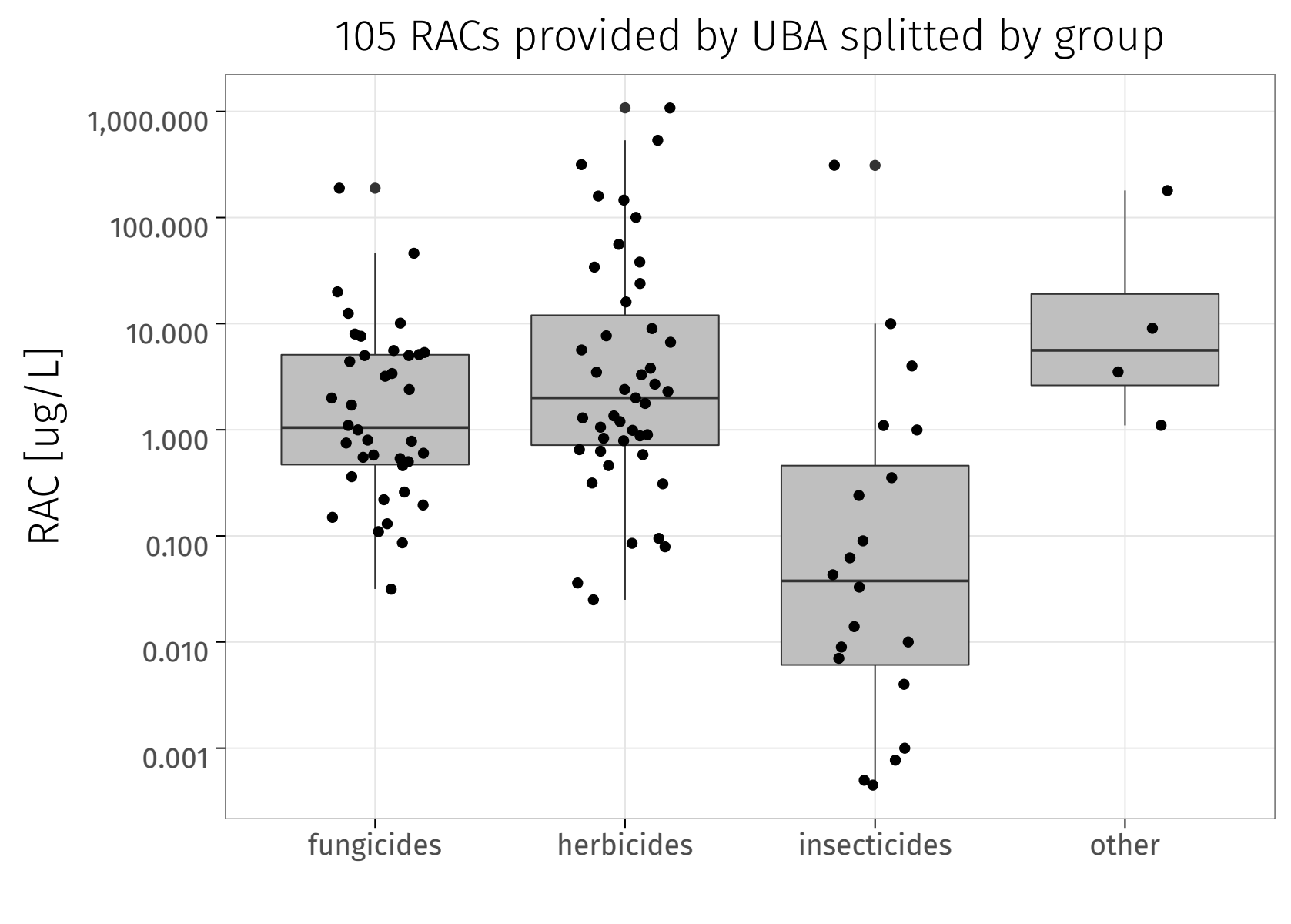
Stehle (2015)

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

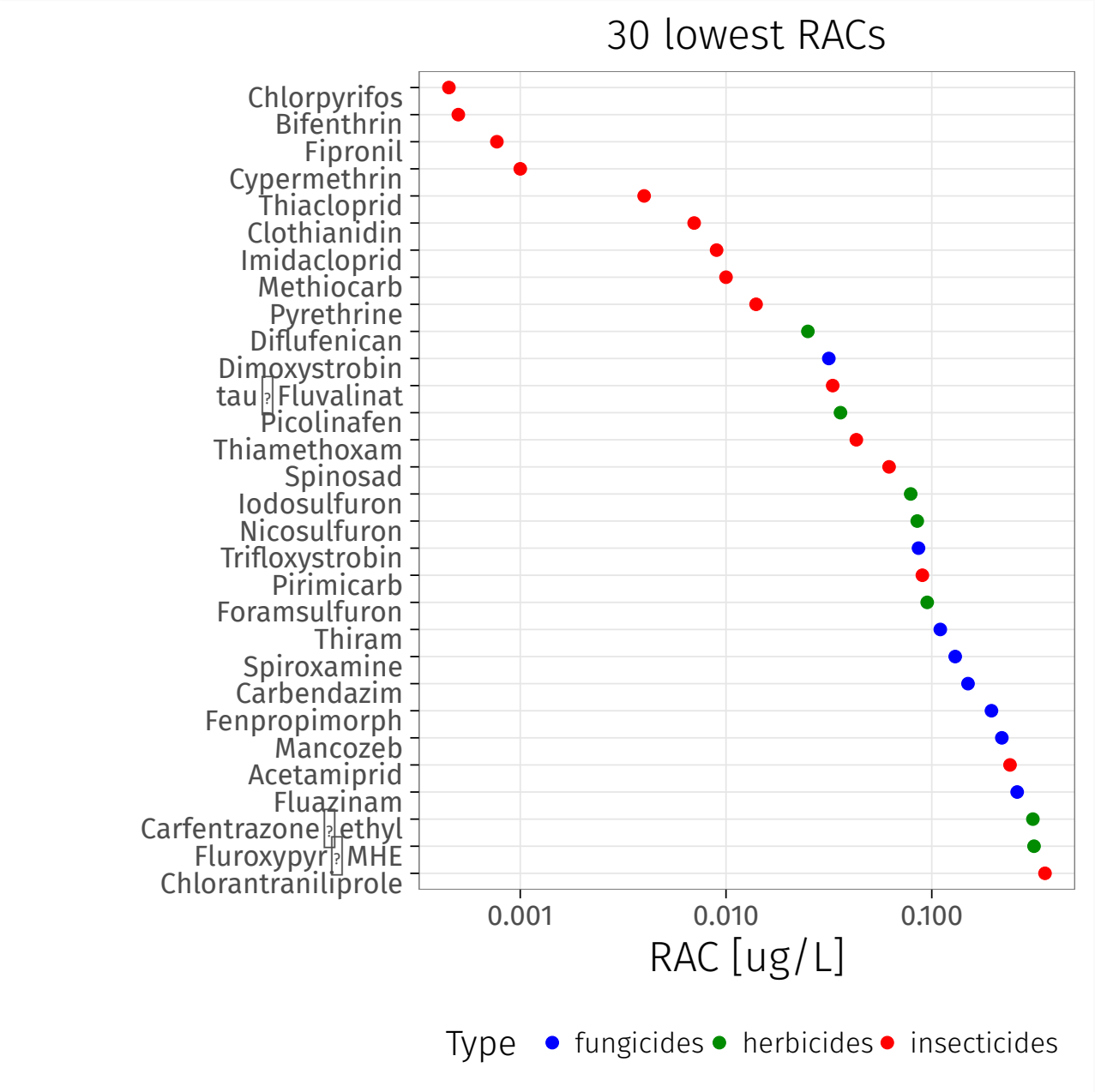
Knauer (2016)

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

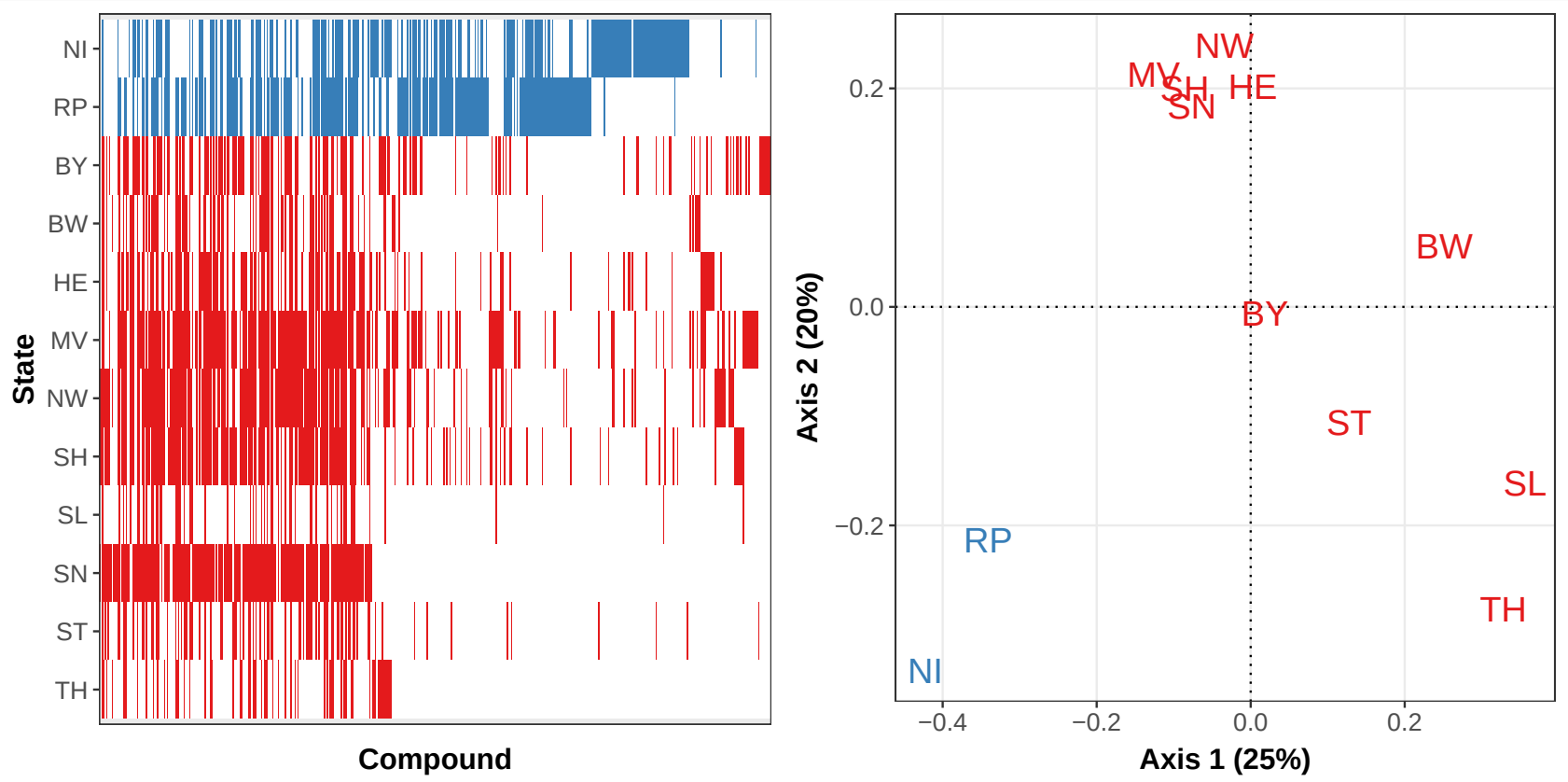
RACs by Type


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

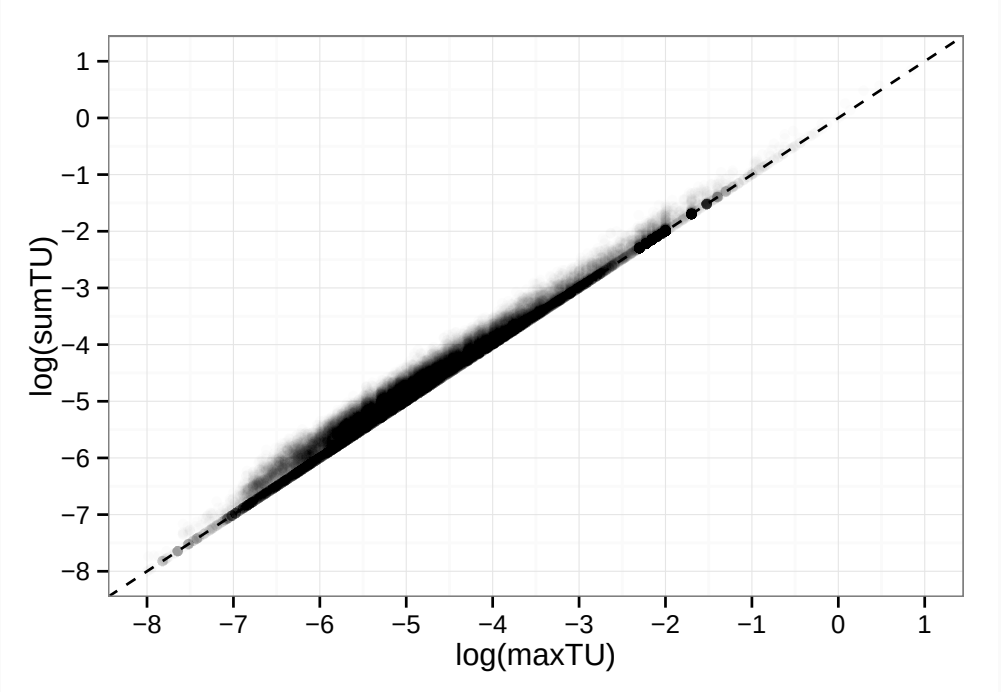
RACs by Compound (Add Stehle to plot...)



Analysed compound spectra by state



Mixtures are common, but one compound dominates the risk

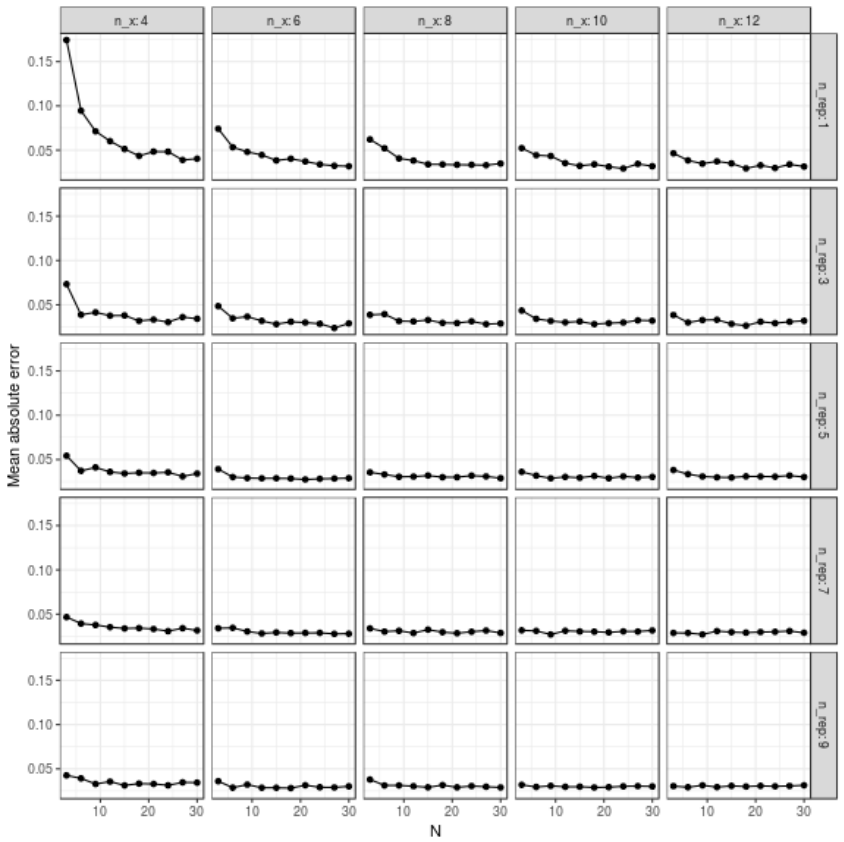


- up to 50 compounds in one sample
- high correlation
- $\sim$ 0.5 TU increase
- mainly one compound responsible for risk

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`

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

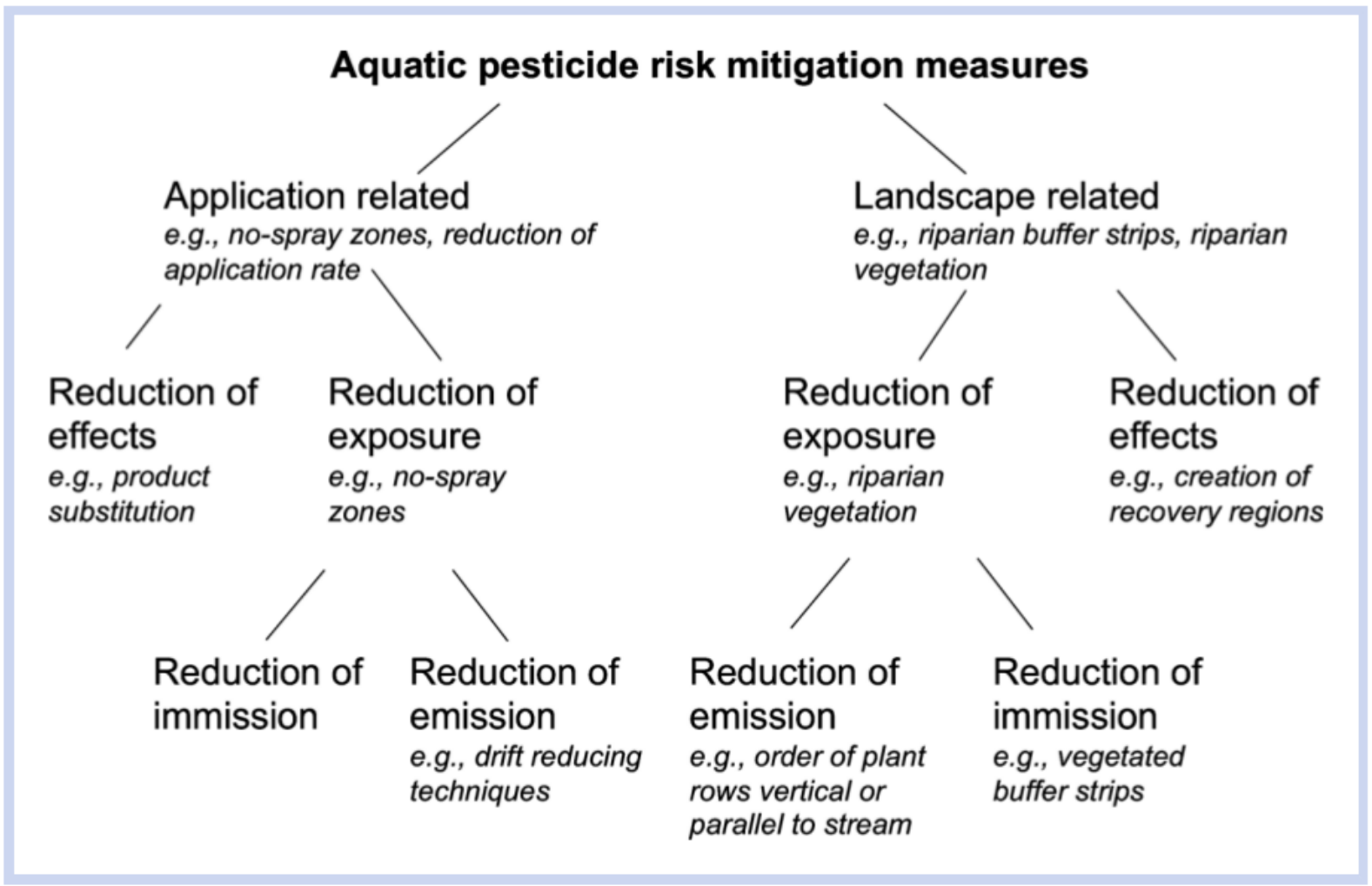
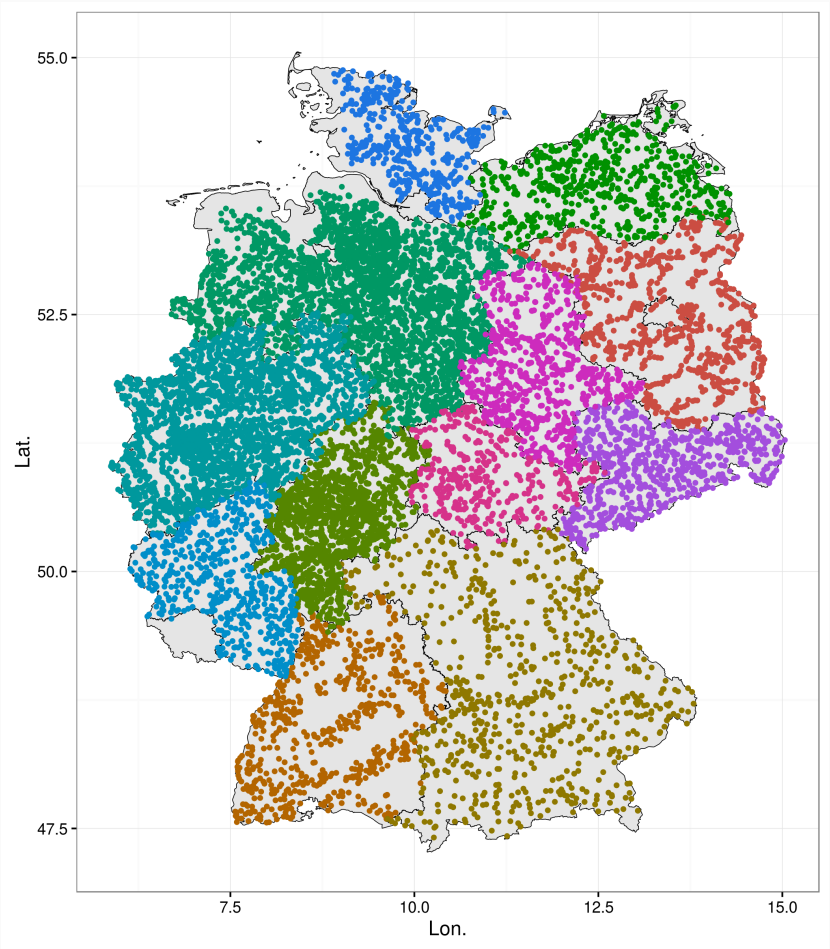


Figure 1. Classification of pesticide risk mitigation measures.

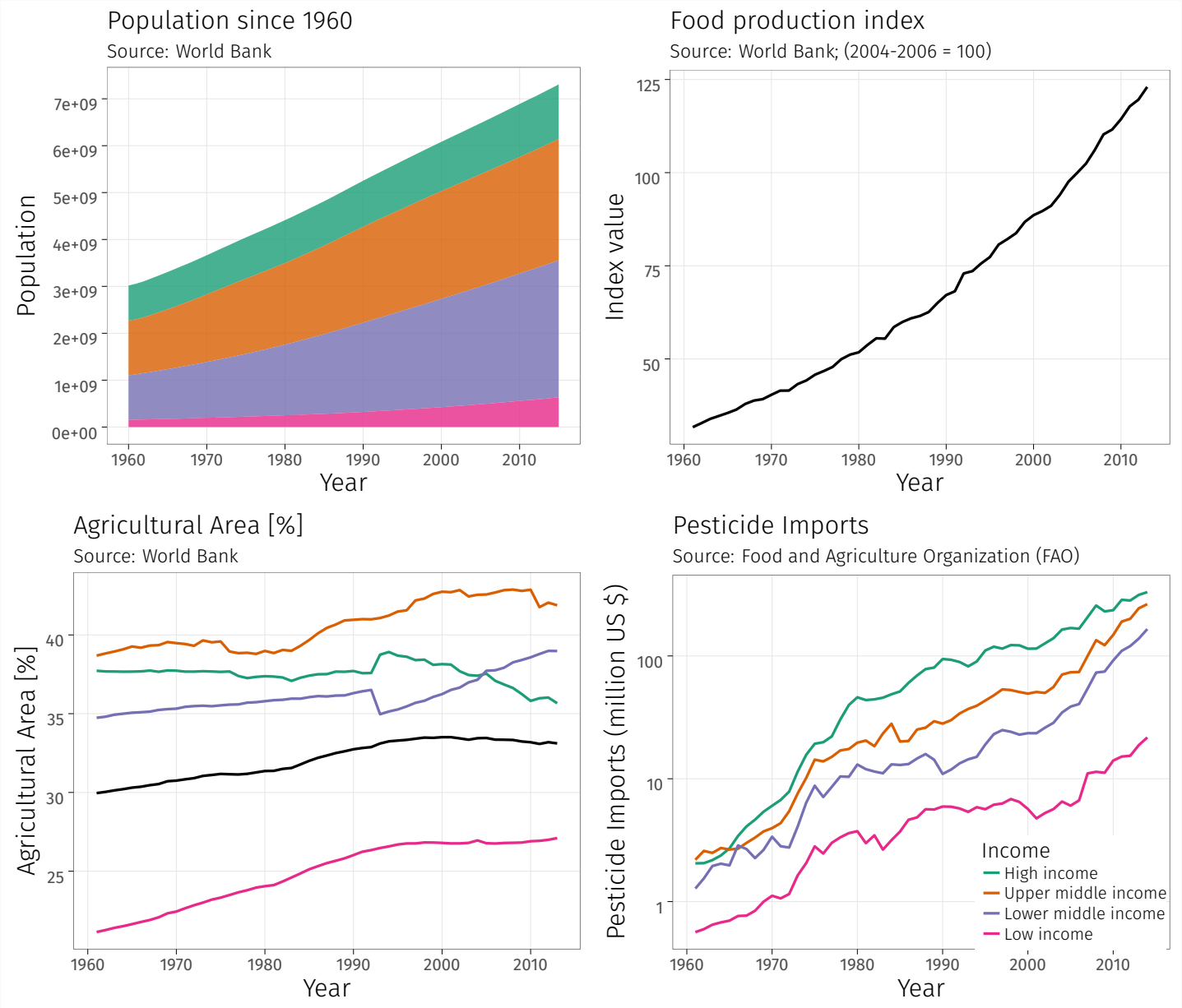
Bereswill, Streloke, Schulz (2014). Risk mitigation measures for diffuse pesticide entry into aquatic ecosystems: Proposal of a guide to identify appropriate measures on a catchment scale: Guide to Identify Pesticide Risk Mitigation Measures. IEAM 10, 286–298.

Biotic field effects



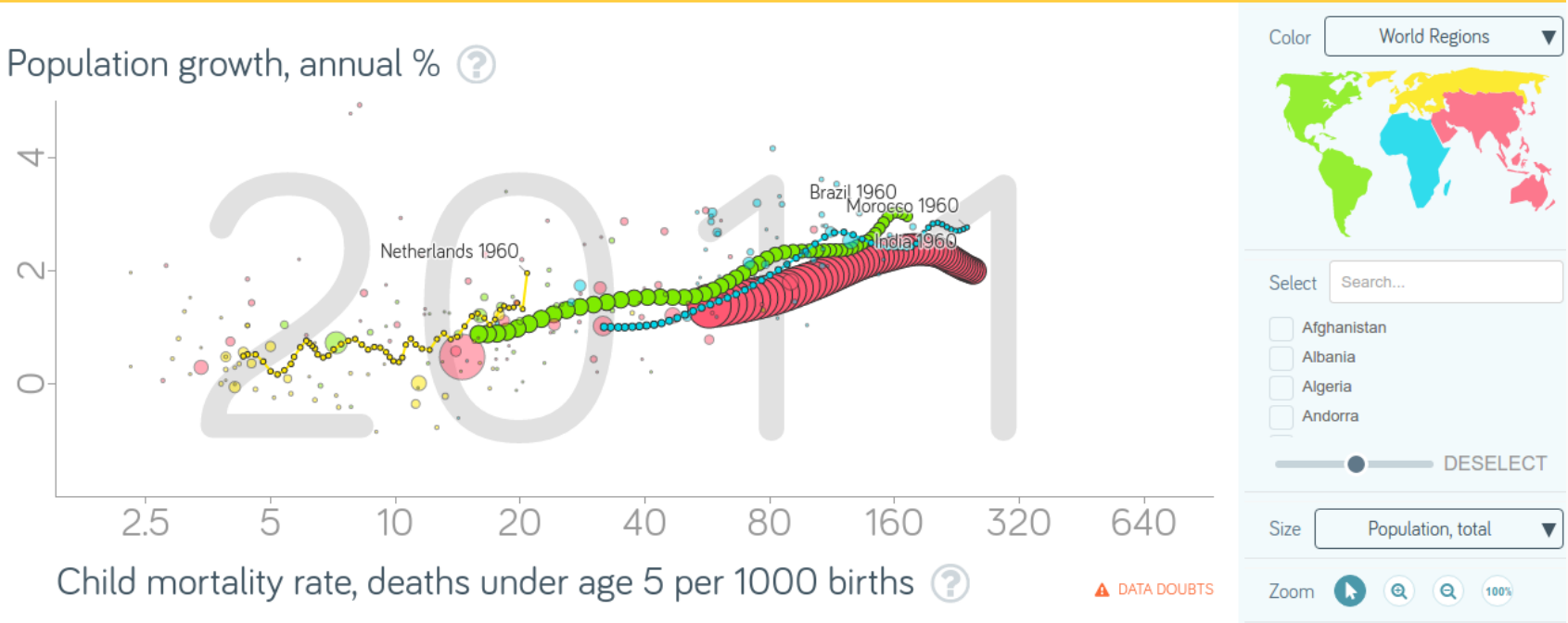
- biological data with good spatial coverage
- 60% of spatial congruence
- Large scale effects largely unknown.
- Some work left...
- Future....

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.