

Statistical Eco(-toxico)logy

Improving the Utilisation of Data for
Environmental Risk Assessment

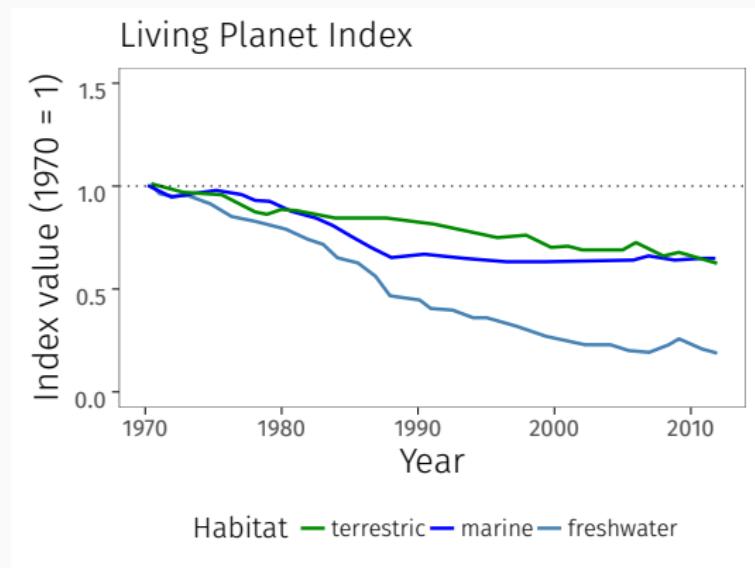
Eduard Szöcs

25th January 2017

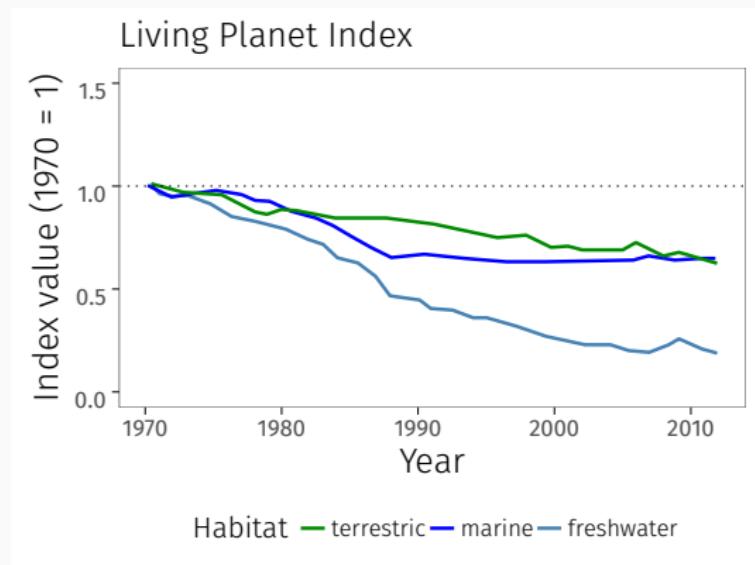
Table of contents

1. Environmental Risk Assessment (ERA) and Monitoring
2. Improving Statistics in ERA
3. Exploring Monitoring Data for ERA
4. Solutions for Data Handling in ERA

Freshwater biodiversity is strongly declining



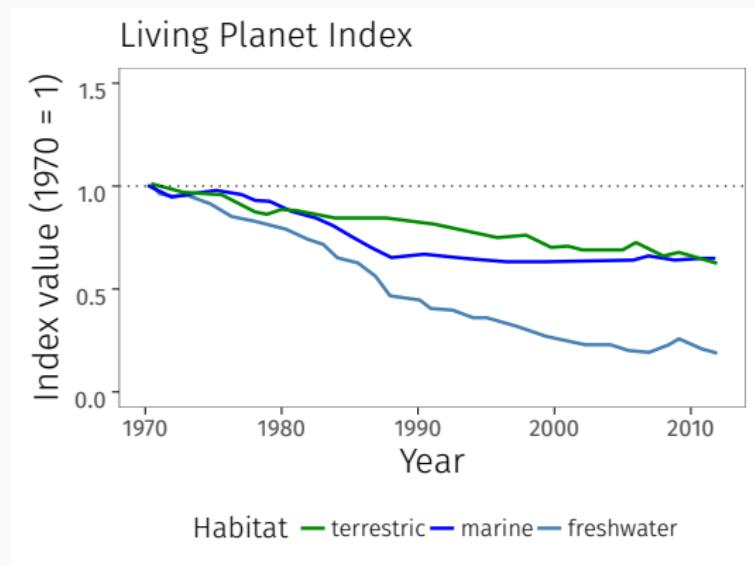
Freshwater biodiversity is strongly declining



Reasons

- Habitat loss
- Overexploitation
- Pollution
- Invasive species

Freshwater biodiversity is strongly declining



Reasons

- Habitat loss
- Overexploitation
- Pollution
- Invasive species

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

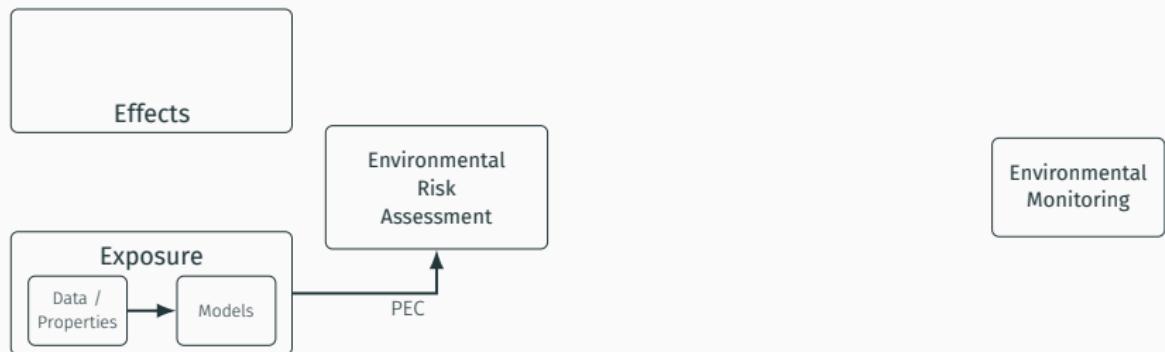
Water Framework Directive
2000/60/EC



Environmental Risk Assessment and Monitoring

Plant Protection Products
Regulation 1107/2009

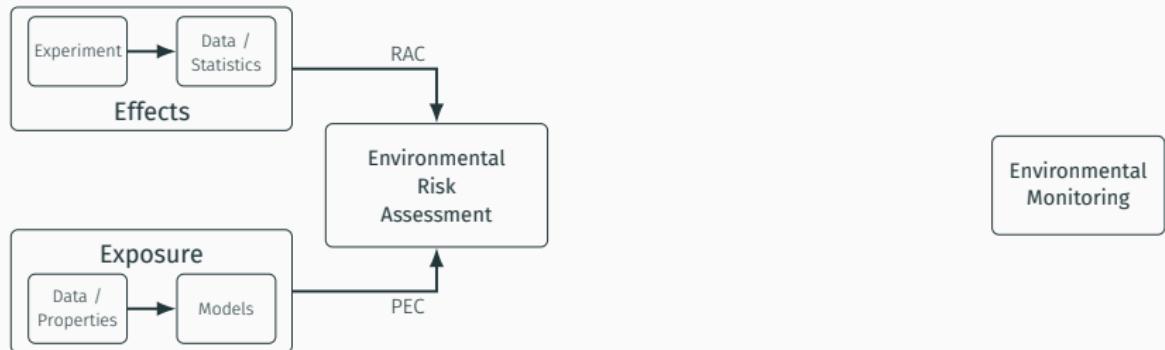
Water Framework Directive
2000/60/EC



Environmental Risk Assessment and Monitoring

Plant Protection Products
Regulation 1107/2009

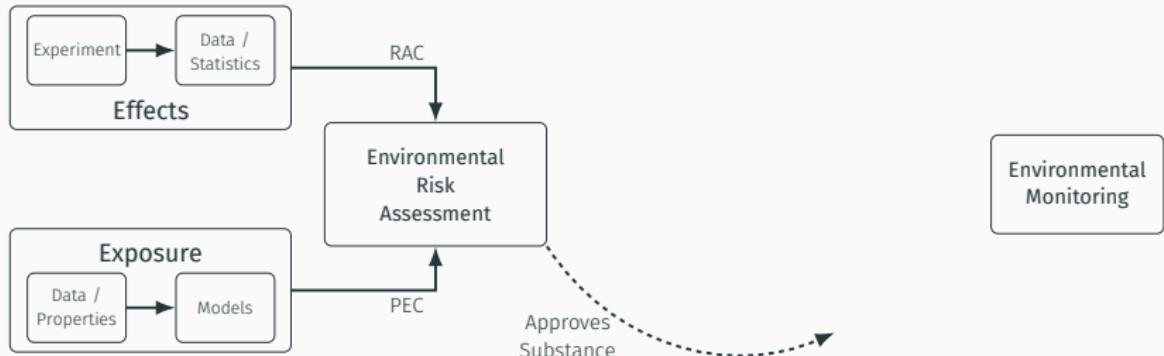
Water Framework Directive
2000/60/EC



Environmental Risk Assessment and Monitoring

Plant Protection Products
Regulation 1107/2009

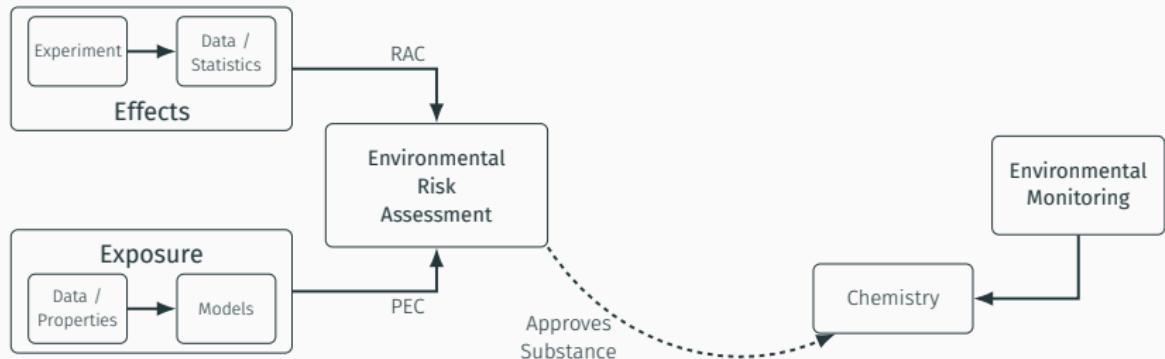
Water Framework Directive
2000/60/EC



Environmental Risk Assessment and Monitoring

Plant Protection Products
Regulation 1107/2009

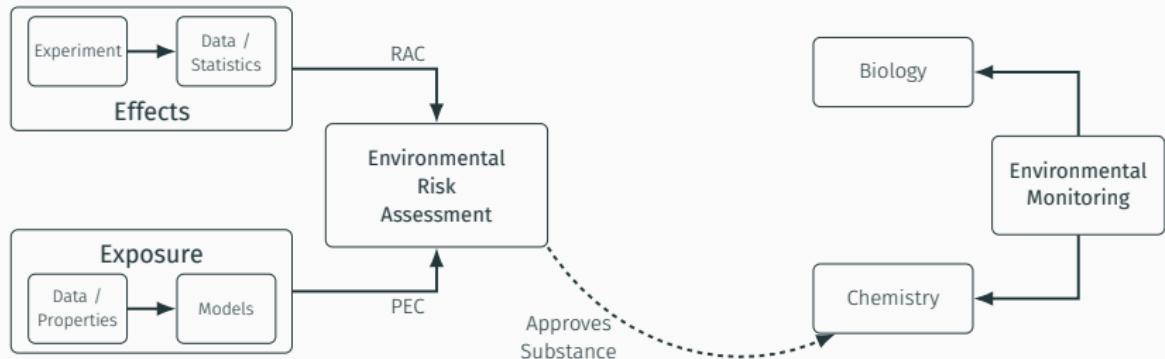
Water Framework Directive
2000/60/EC



Environmental Risk Assessment and Monitoring

Plant Protection Products
Regulation 1107/2009

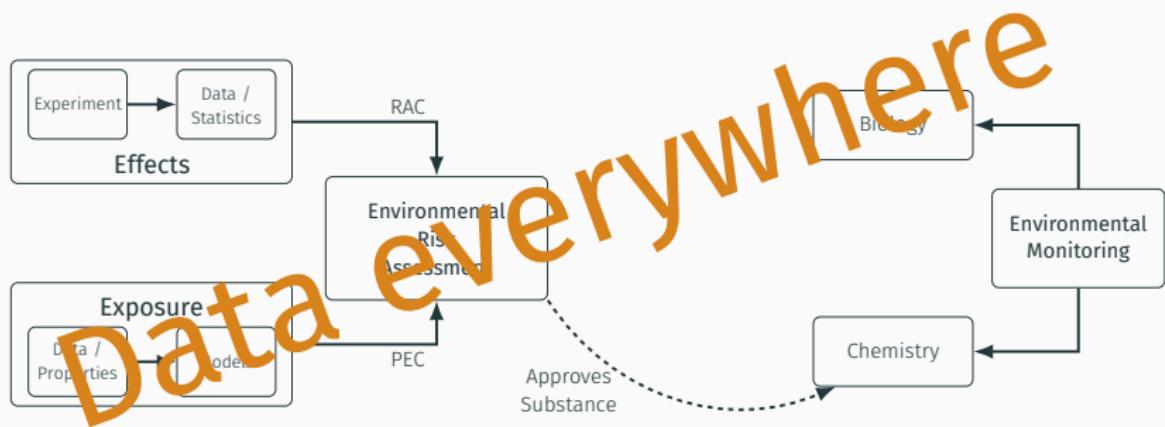
Water Framework Directive
2000/60/EC



Environmental Risk Assessment and Monitoring

Plant Protection Products
Regulation 1107/2009

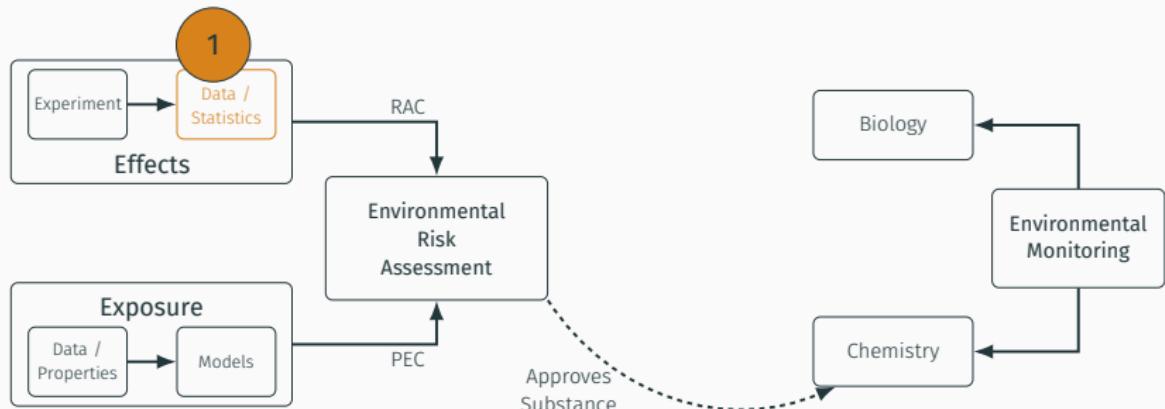
Water Framework Directive
2000/60/EC



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



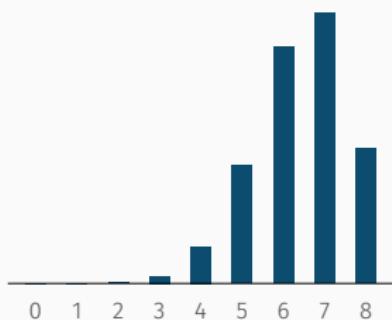
- Daphnia Test
- " x out of n survived"

Experiments in Effect Assessment

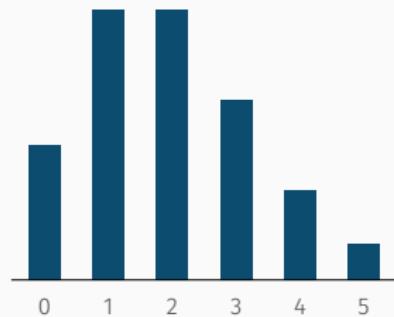


- Daphnia Test
- “*x out of n survived*”
- Mesocosm
- “*number of animals*”

Ecotoxicology is not normal

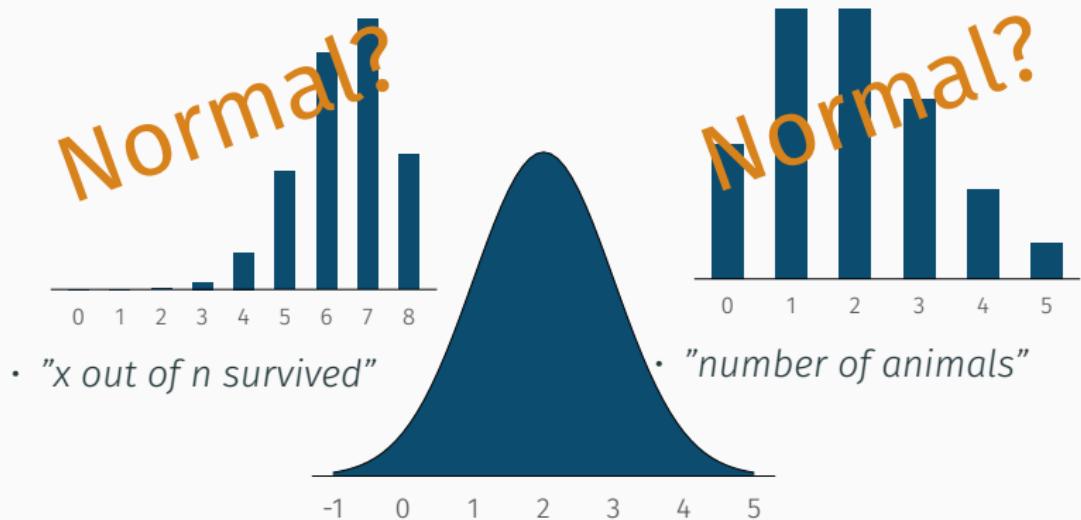


- " x out of n survived"

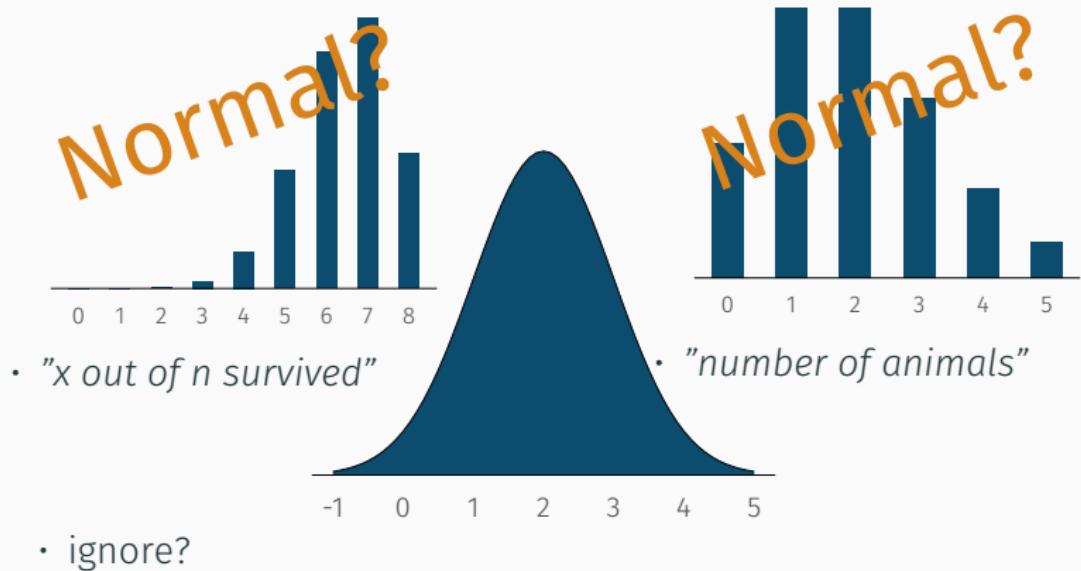


- "number of animals"

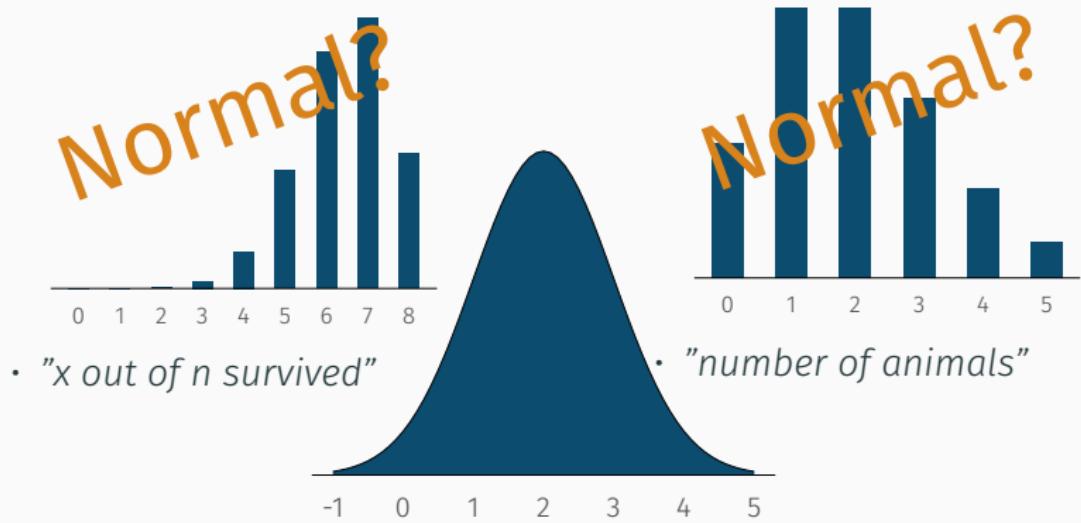
Ecotoxicology is not normal



Ecotoxicology is not normal

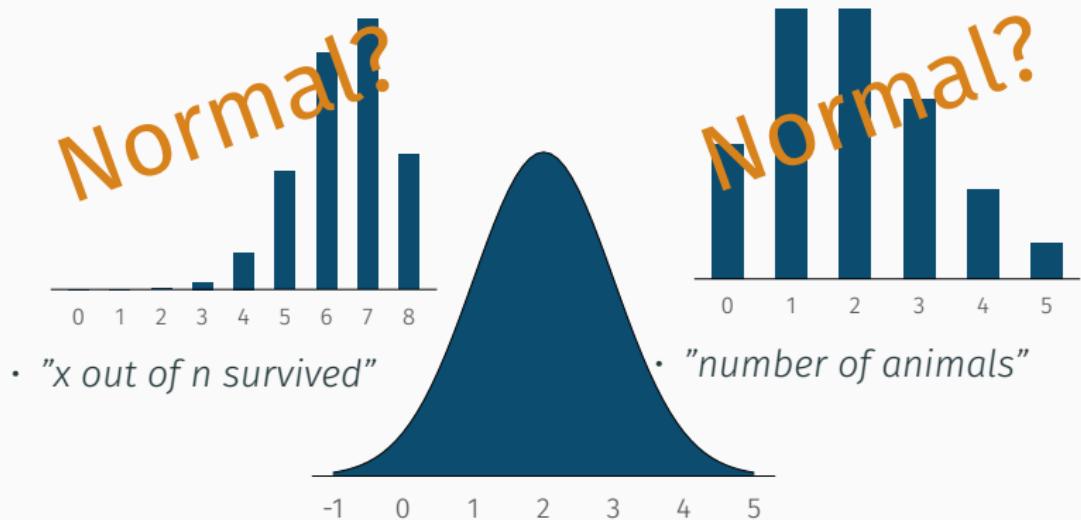


Ecotoxicology is not normal



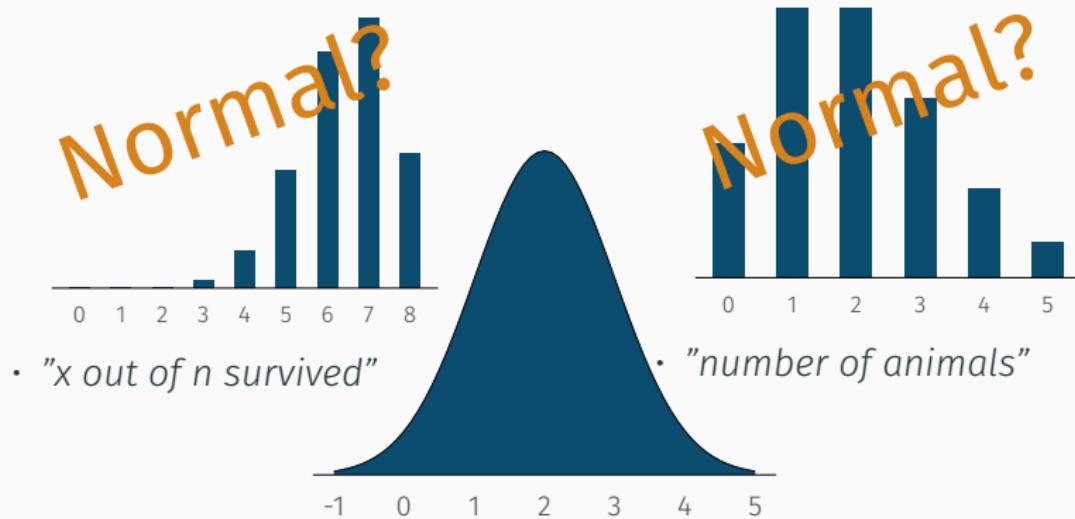
- ignore?
- transform?

Ecotoxicology is not normal



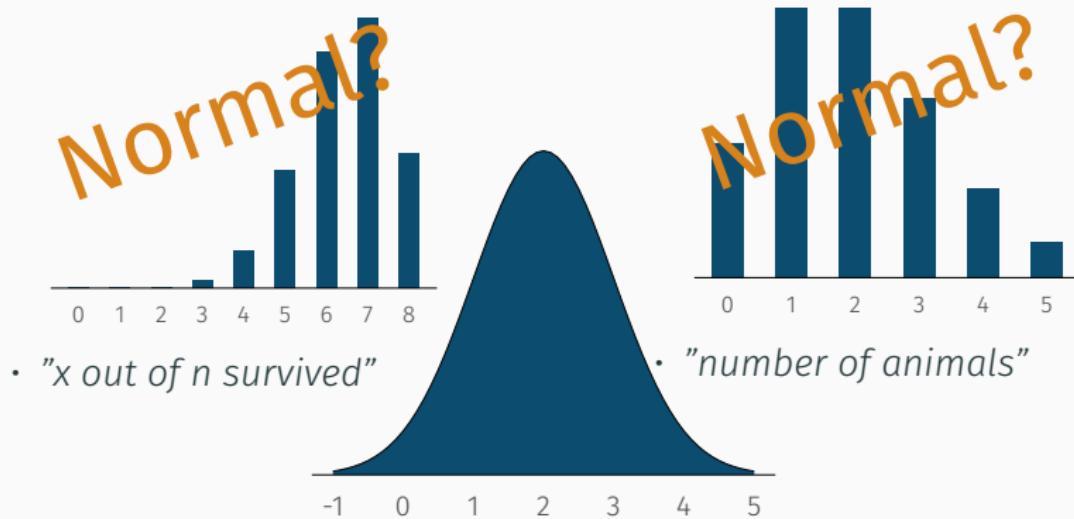
- ignore?
- transform?
- non-parametric?

Ecotoxicology is not normal



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

Ecotoxicology is not normal



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

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

370

Generalized Linear Models

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

Rothamsted Experimental Station, Harpenden, Herts



Methods in Ecology and Evolution



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 Kotze²

¹Biodiversity and Climate Research Centre, Senckenberganlage 25, D-60325 Frankfurt am Main, Germany and

²Department of Environmental Sciences, PO Box 65, University of Helsinki, Helsinki FI-00014, Finland



Methods in Ecology and Evolution



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. Wright¹ and Yi Wang^{1,2}

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



A recent history (uncomprehensive) of GLM 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 (uncomprehensive) of GLM 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 Szöcs¹ · Ralf B. Schäfer¹

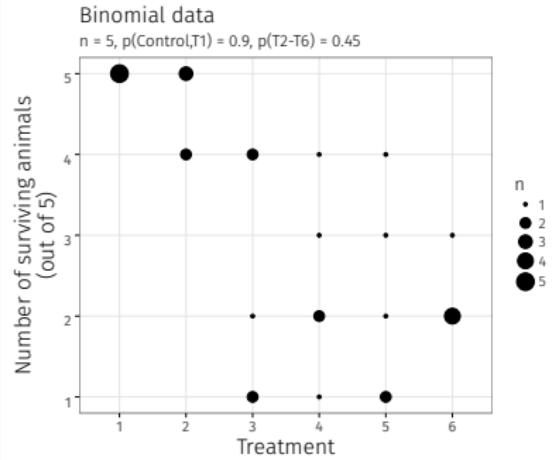


A simulation study

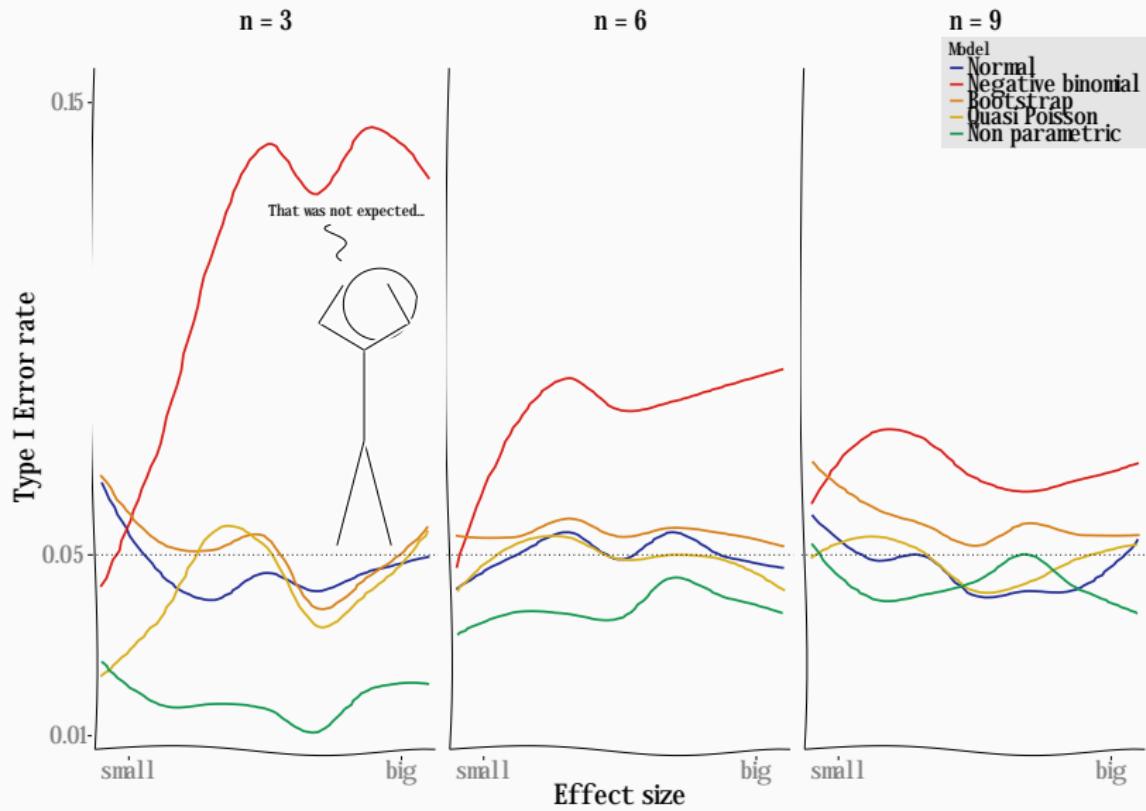


Simulation:

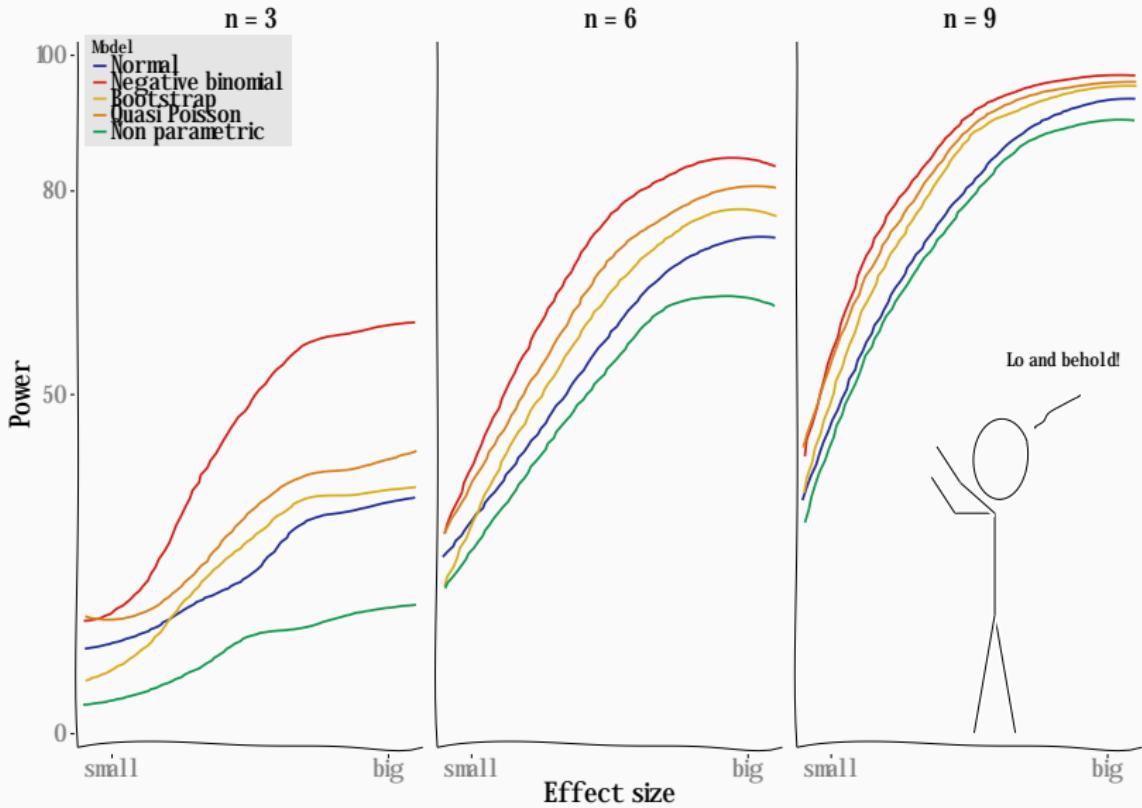
- Count & Binomial data
- Vary replicates & effect sizes
- LM, GLMs, Non-parametric
- Power & Type I Error



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. GLMs can increase this power

Where are we today?

Three days earlier...



Where are we today?

Three days earlier...

Methods in Ecology and Evolution



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?

Current state of knowledge:

Methods in Ecology and Evolution



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 Lyons², Jakub Stoklosa¹ and Anthony R. Ives³

¹School of Mathematics and Statistics and Evolution & Ecology Research Centre, University of New South Wales, NSW 2052, Australia; ²School of Biological, Earth and Environmental Sciences, University of New South Wales, NSW 2052, Australia; and

³Department of Zoology, University of Wisconsin-Madison, Madison, WI 53706, USA



Where are we today?

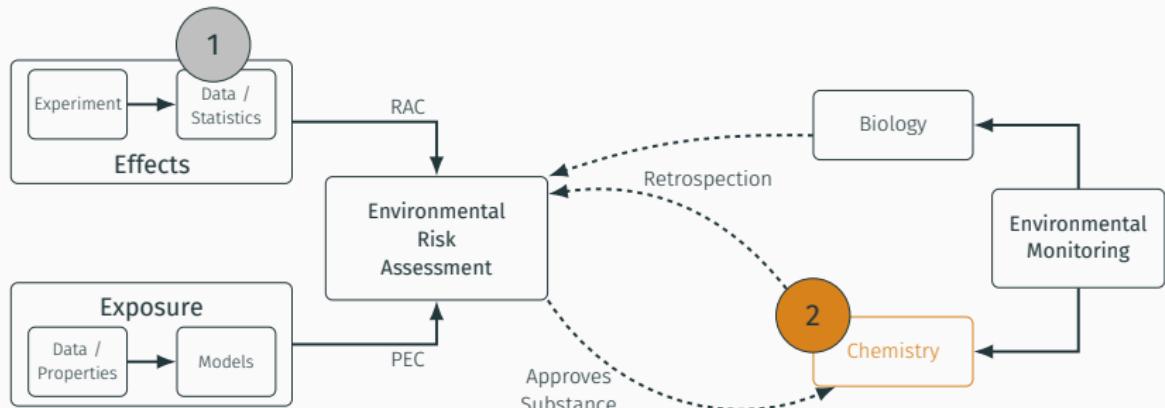
Three points to consider ...

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

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

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 with highest concentrations
3. Precipitation increases concentrations
4. Highest concentrations in summer

Analysing chemical concentrations

Analysing chemical concentrations

- Concentrations < LOQ (96% of all measurements)

Analysing chemical concentrations

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

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

Analysing chemical concentrations

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

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

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

Analysing chemical concentrations

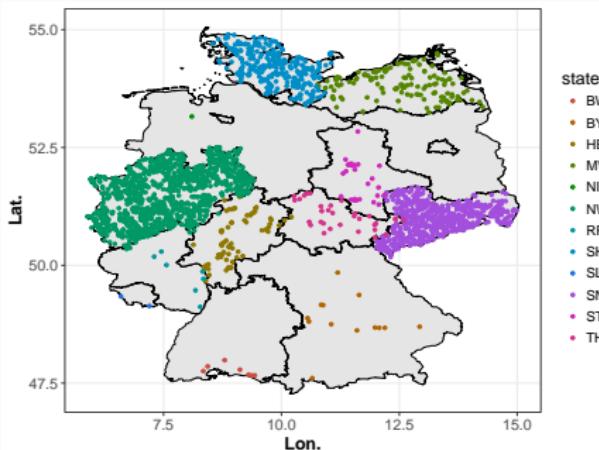
- Concentrations < LOQ (96% of all measurements)
- Hurdle-model:

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

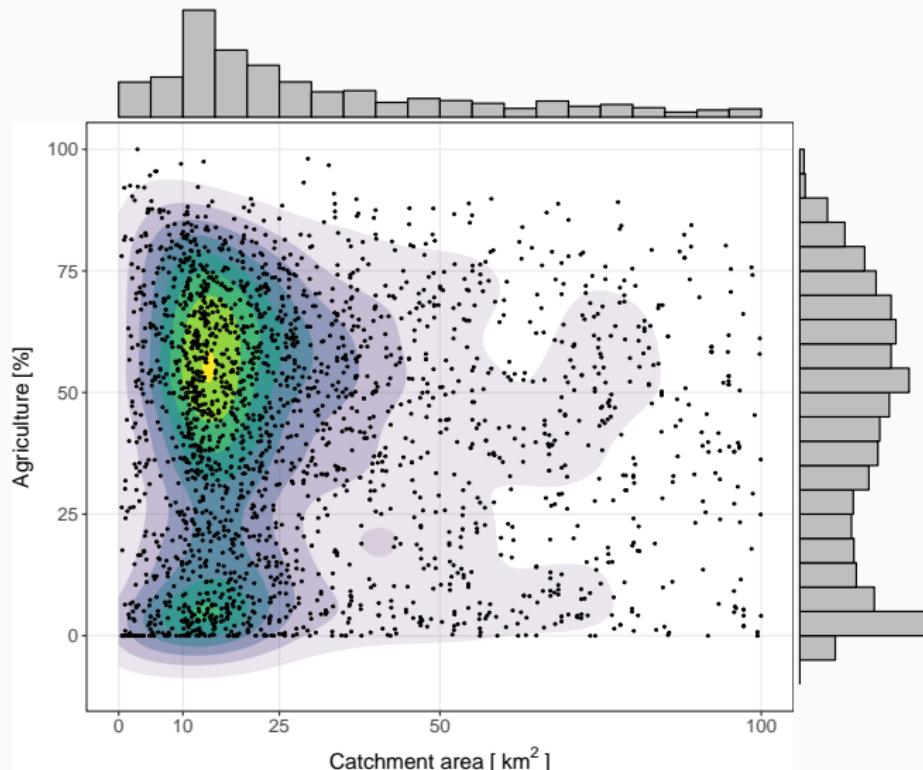
- Risk Quotient
 - $RQ = \frac{C}{RAC}$
- Predictors
 - Catchment Size
 - Agricultural land use
 - Precipitation

Compiled data: Big, but inhomogeneous

- 1,766,104 measurements
- 478 pesticides
- 24,743 samples
- 2,301 sites

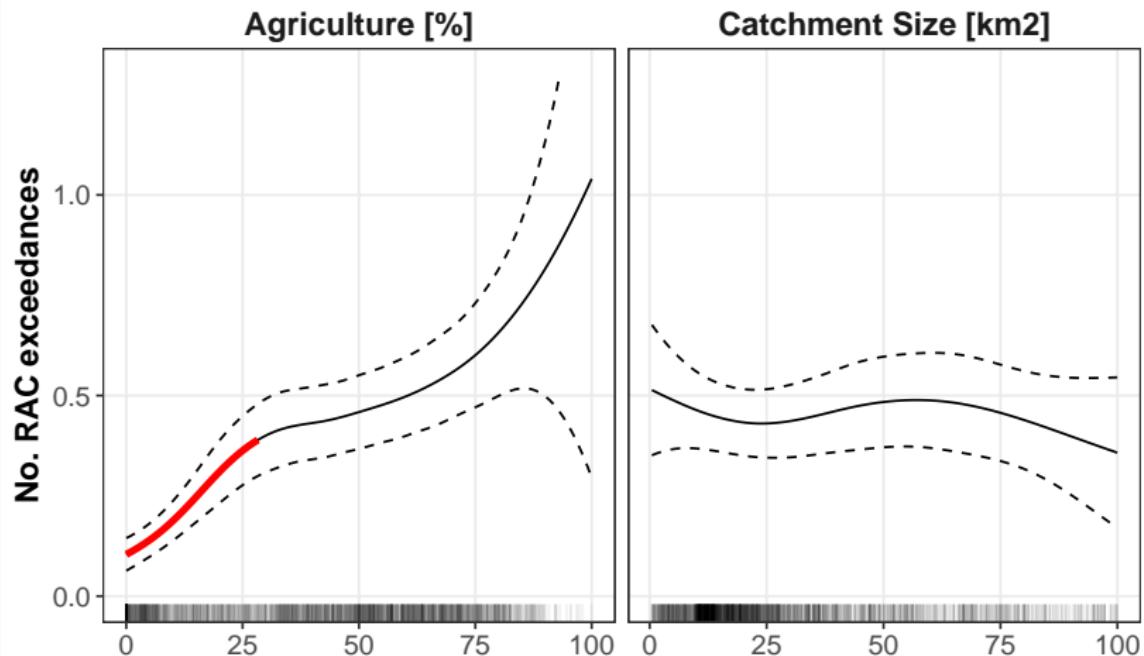


Monitoring: Small streams are underrepresented

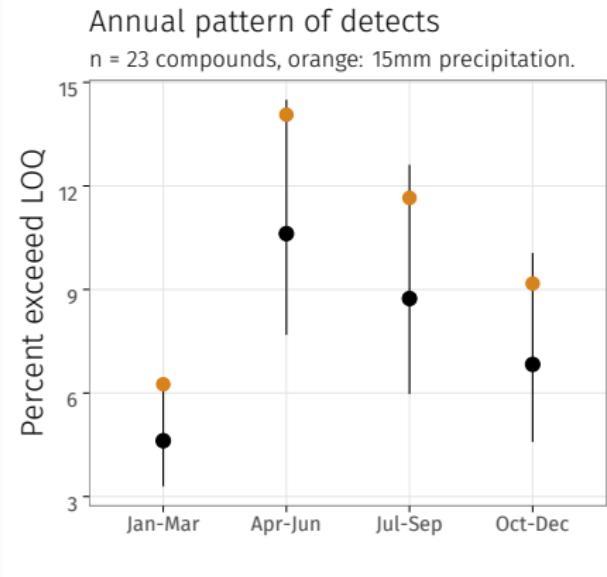


Landscape: Factors influencing risk

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

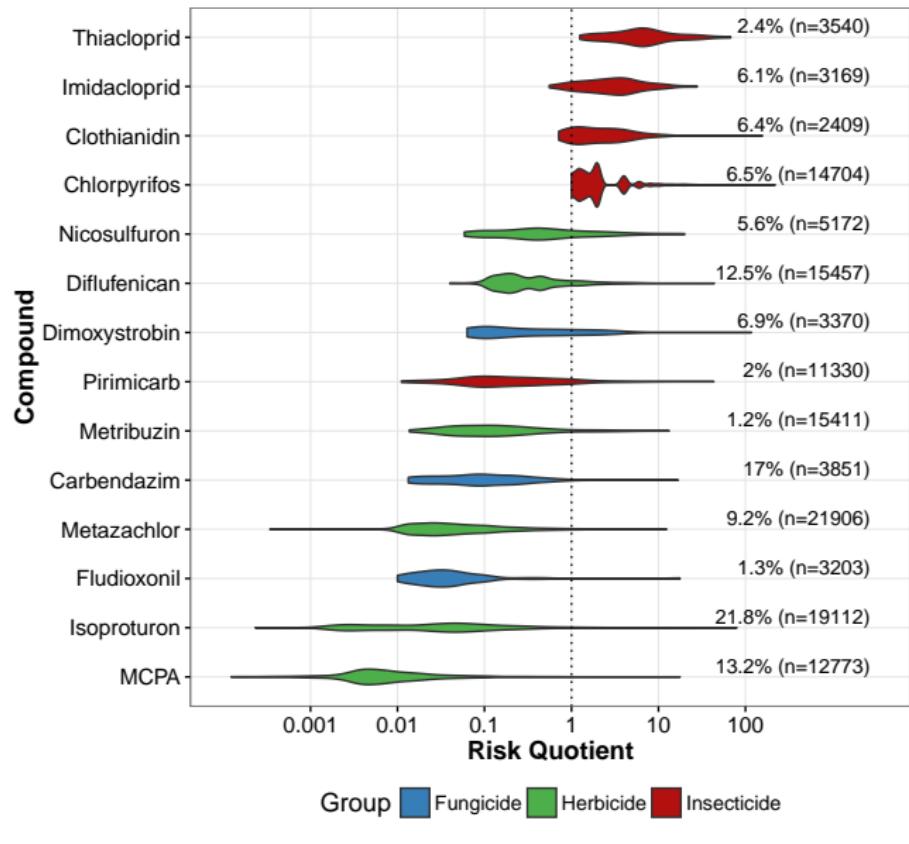


Sampling: Factors influencing risk



- Peak in summer
- Increase by **precipitation**
- absolute concentrations show much higher variability

Risks: Compounds exceeding risk thresholds

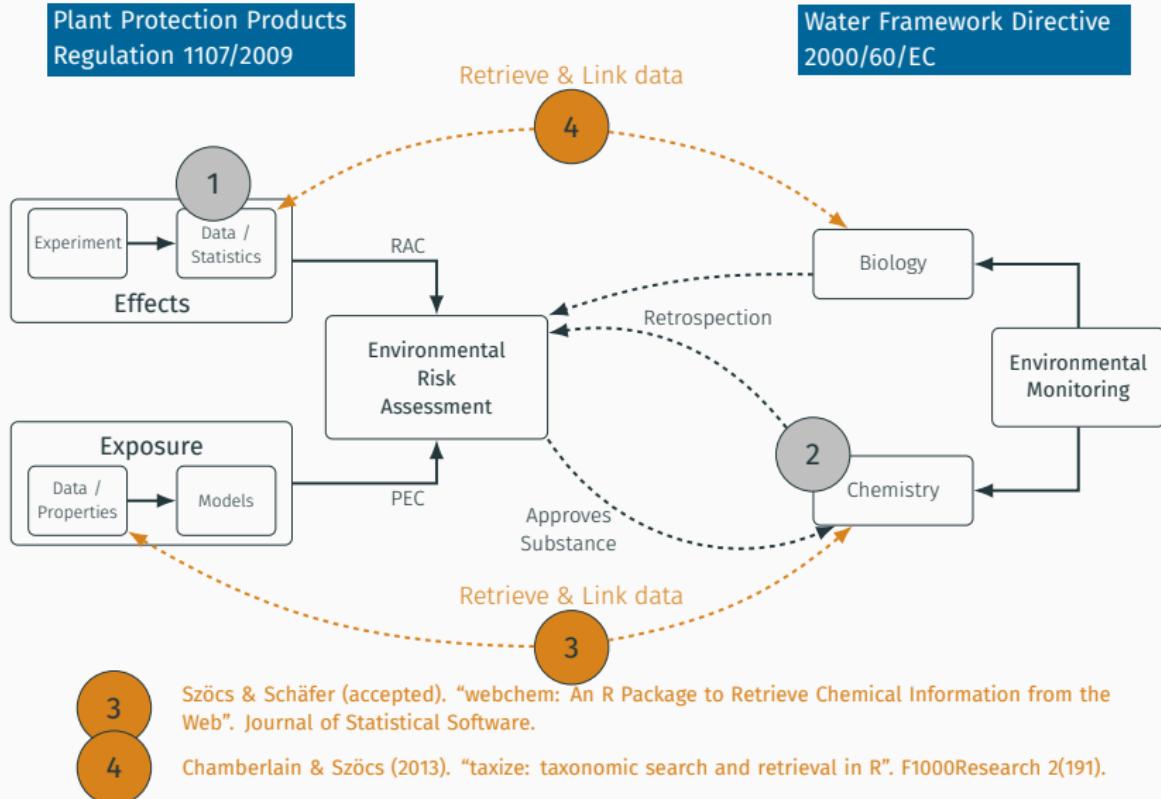


What we learned from this study

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



Biologists & Chemists face the same problems

Biologists & Chemists face the same problems

Names

Osmia rufa, Osmia bicornis, Osmia ruffa, Osmia unilandauis

Chlorpyrifos, Chlorpyriphos,
Chlorphyrifos, Chlorpyrifot

Biologists & Chemists face the same problems

Names

Osmia rufa, *Osmia bicornis*, *Osmia ruffa*, *Osmia unilandauis* Chlorpyrifos, Chlorpyriphos, Chlorphyrifos, Chlorpypifot

Hierarchies

Biologists & Chemists face the same problems

Names

Osmia rufa, *Osmia bicornis*, *Osmia ruffa*, *Osmia unilandauis* Chlorpyrifos, Chlorpyriphos, Chlorphyrifos, Chlorpypifot

Hierarchies

Traits / Properties

Wing length, Mass, Season

Biologists & Chemists face the same problems

<u>Names</u>	
<i>Osmia rufa</i> , <i>Osmia bicornis</i> , <i>Osmia ruffa</i> , <i>Osmia unilandauis</i>	Chlorpyrifos, Chlorpyriphos, Chlorphyrifos, Chlorpypifot
<u>Hierarchies</u>	
Hymenoptera/ Apoidea/ Megachilidae/ <i>Osmia/ rufa</i>	organophosphate, ester, insecticide
<u>Traits / Properties</u>	
Wing length, Mass, Season	Mass, K_{OW} , LC_{50}
<u>Identifiers</u>	
NCBI, ITIS, EOL, ...	2921-88-2, Clc1c(OP(=S)[...], InChI=1S/C9H11C[...], SBPBAQFW[...], CSID,...

Biologists & Chemists face the same problems

Names

Osmia rufa, Osmia bicornis, Osmia ruffa, Osmia unilandauis Chlorpyrifos, Chlorpyriphos, Chlorphyrifos, Chlorpyrifot

Hierarchies

Hymenoptera/ Apoidea/
Megachilidae/ *Osmia/ rufa* organophosphate, 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,...

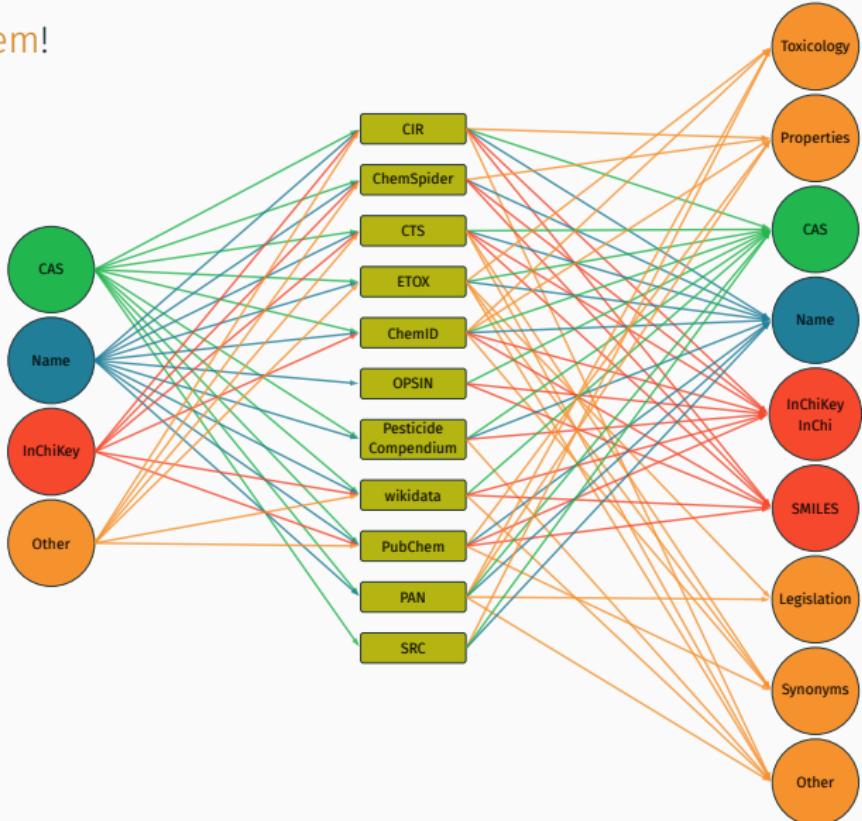
Amount of data

2993 taxa

478 pesticides

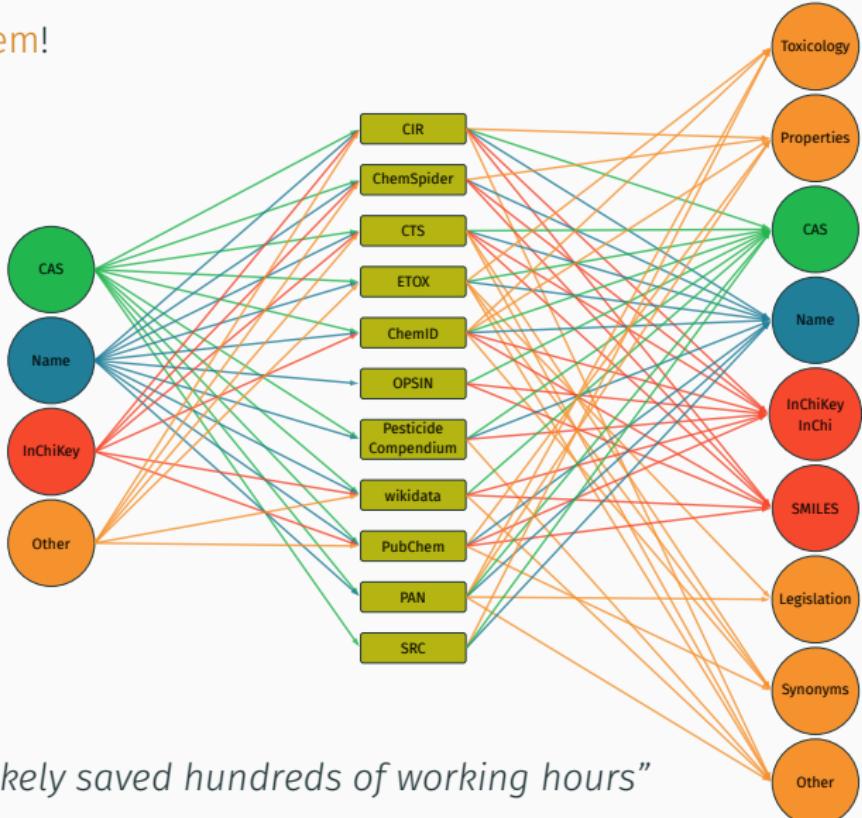
Instead of wasting time...

... use webchem!



Instead of wasting time...

... use webchem!



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

Instead of wasting time...

... use taxize!

A grid of logos for various biodiversity databases and tools:

- ITIS**: Global Invasive Species Database
- iPlant Collaborative™**
- Plantminer**
- Tropicos®**
- GBIF**
- NCBI**
- eOL Encyclopedia of Life**
- gnl**
- RED LIST**
- ubio**
- Canadensys**
- ThePlantList**

Instead of wasting time...

... use taxize!

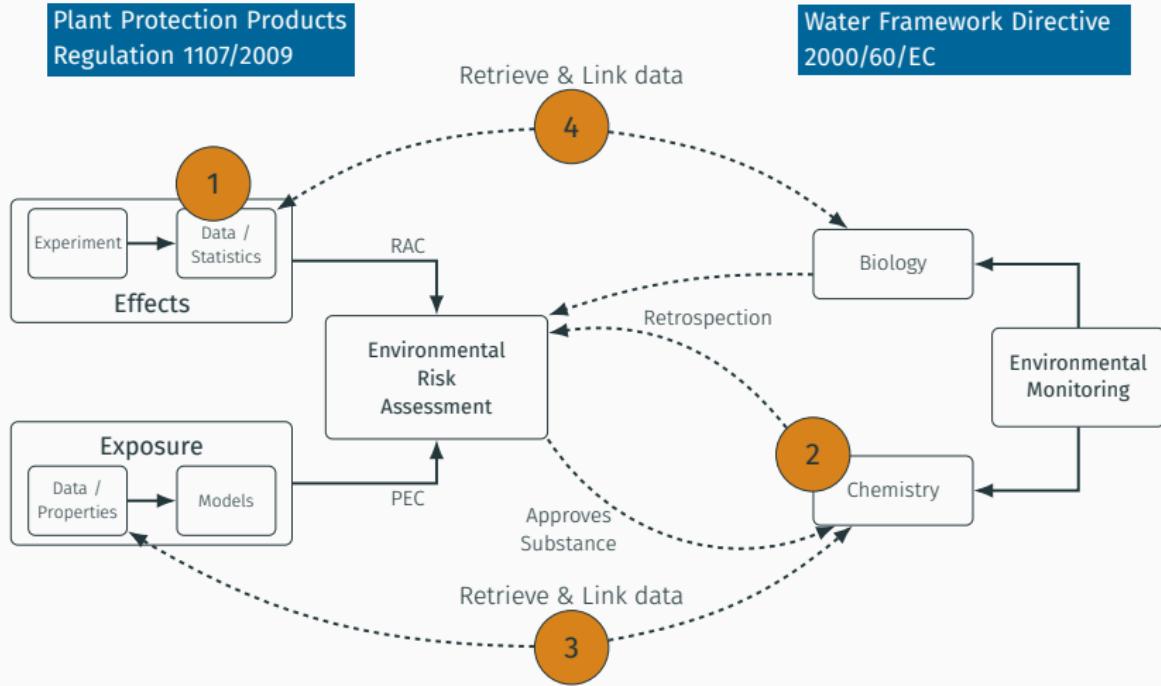
The image shows a collection of logos for different biological databases and projects, all related to taxonomy and biodiversity. The logos are arranged in a grid-like fashion within a white rectangular box.

- ITIS**: Logo features a green circular icon with a tree, a bee, and a mushroom, next to the acronym "ITIS".
- iPlant Collaborative**: Logo features a blue circular icon with a plant, next to the text "iPlant Collaborative™".
- Plantminer**: Logo features a green circular icon with dots connected by lines forming a network, next to the text "Plantminer".
- Catalogue of Life**: Logo features a horizontal bar with blue, yellow, red, and green segments, next to the text "Catalogue of Life".
- Tropicos**: Logo features a green circular icon with a plant, next to the text "Tropicos®".
- Global Invasive Species Database**: Logo features a green circular icon with a plant, next to the text "Global Invasive Species Database".
- gbif**: Logo features a green circular icon with a plant, next to the text "gbif".
- NCBI**: Logo features a blue square icon with a white stylized "N" shape, next to the text "NCBI".
- eOL Encyclopedia of Life**: Logo features a green circular icon with a plant, next to the text "eOL Encyclopedia of Life".
- gnl**: Logo features a green circular icon with colored dots, next to the text "gnl".
- RED LIST**: Logo features a red circular icon with a white plant shape, next to the text "RED LIST".
- ubio**: Logo features a green circular icon with colored dots, next to the text "ubio".
- Canadensys**: Logo features a red circular icon with a white plant shape, next to the text "Canadensys".
- ThePlantList**: Logo features a green circular icon with colored dots, next to the text "ThePlantList".

"Days of searching done during my morning coffee. Amazing. **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/>

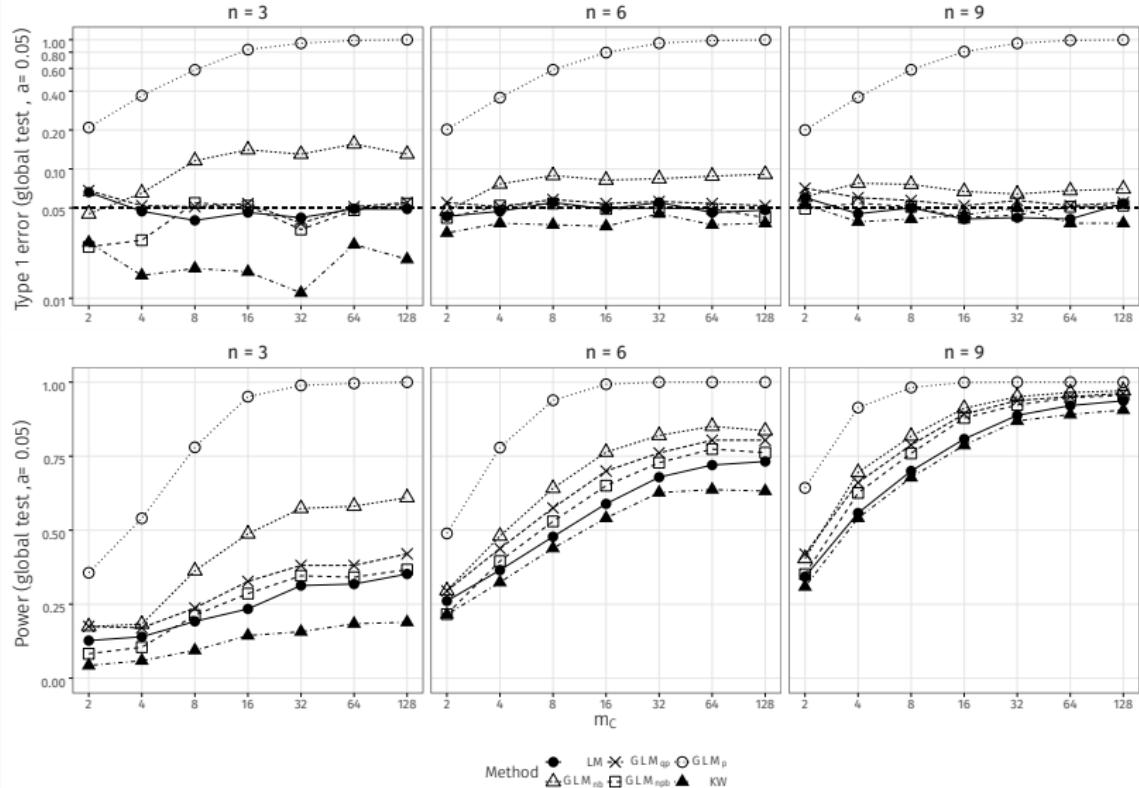
🐦 [@EduardSzoecls](https://twitter.com/EduardSzoecls)

📄 https://github.com/edild/phd_defense

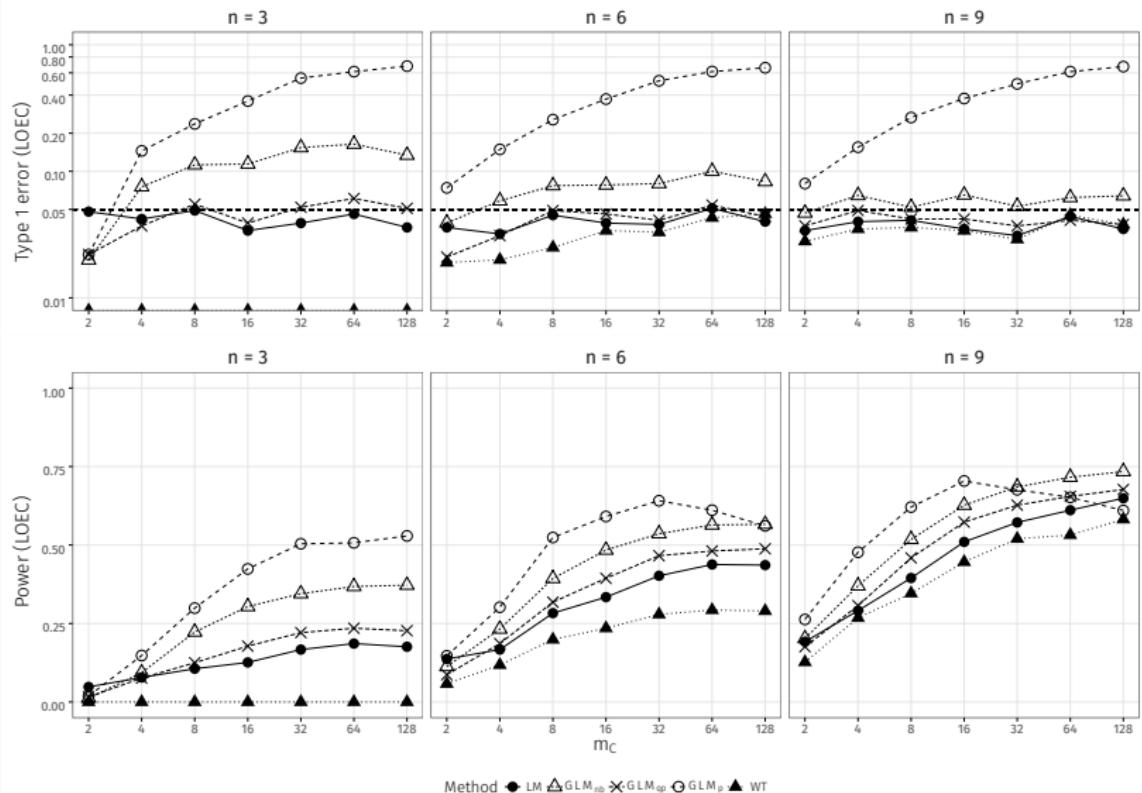
📄 https://github.com/edild/phd_thesis



Power en detail



For LOEC it is even worse



Comparison with Ives...

Szöcs (2015)

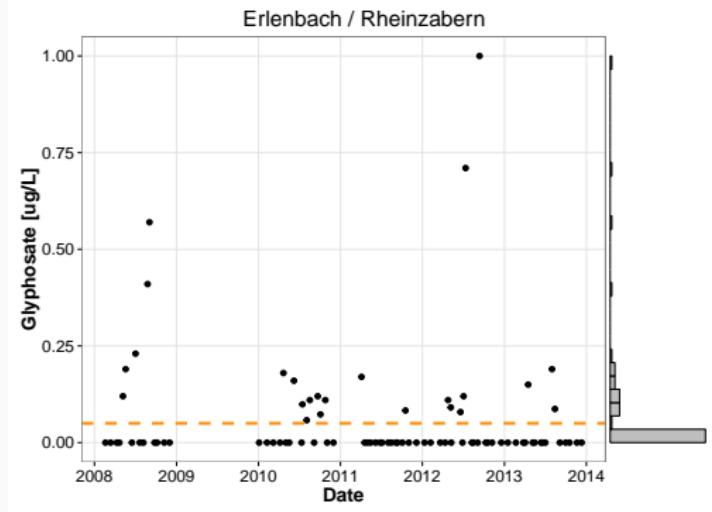
- factorial design
- one predictor
- low replicated
- LM, GLM, bootstrap
- High T1 error of NB
- Quasi-Poisson worked well
- Bootstrap fixes the problems

Ives (2015)

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

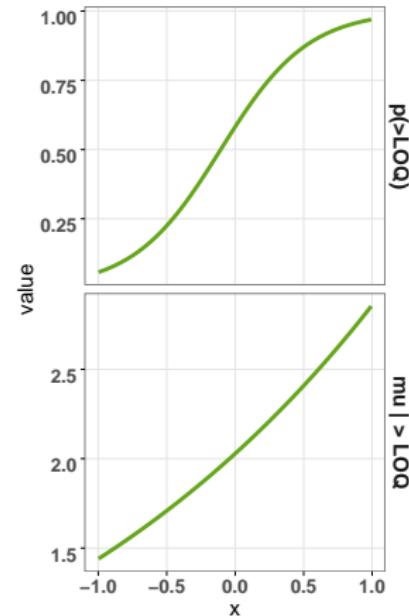
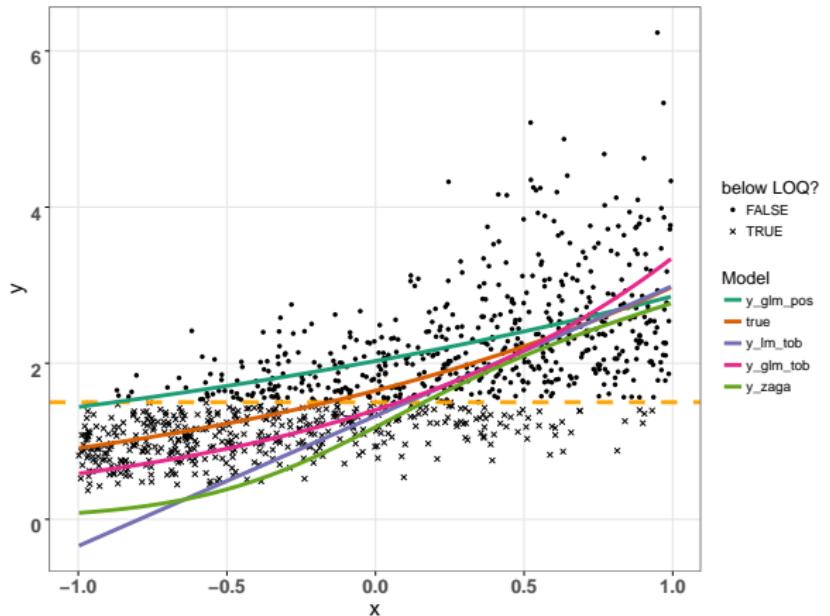
Idiosyncrasies of chemical concentrations

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

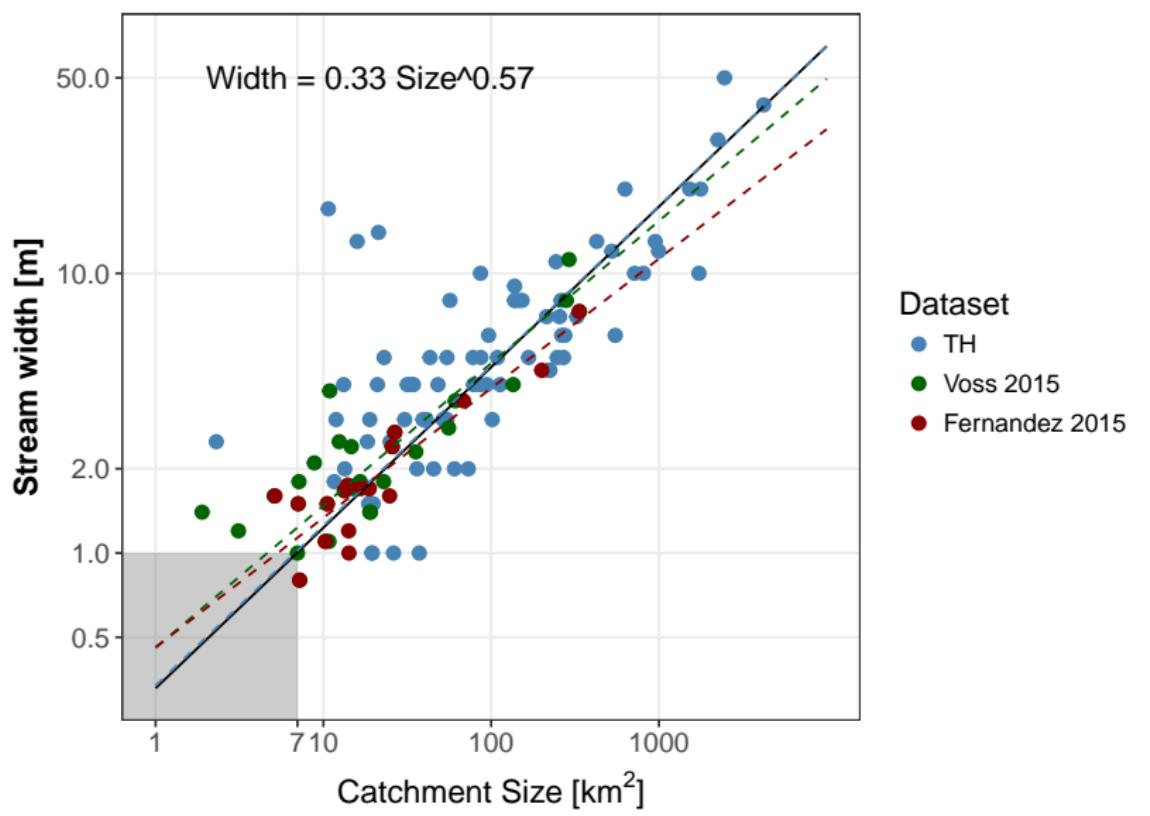


ZAGA what...?

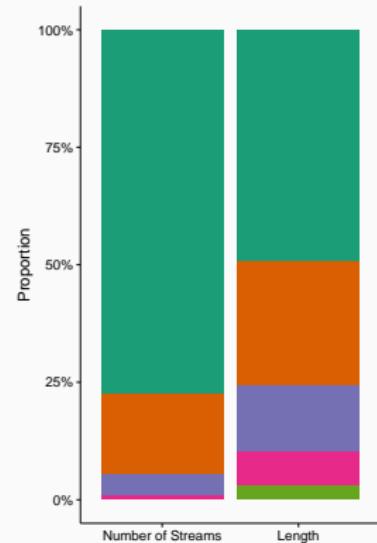
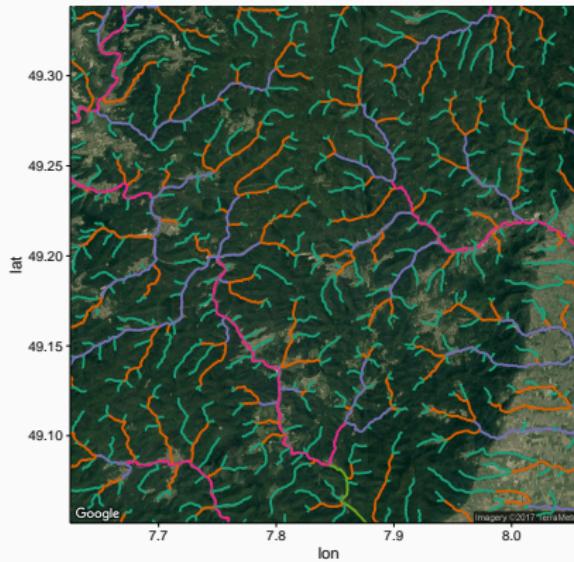
shiny app: <http://uni-ko-ld.de/g4>



Stream size / width



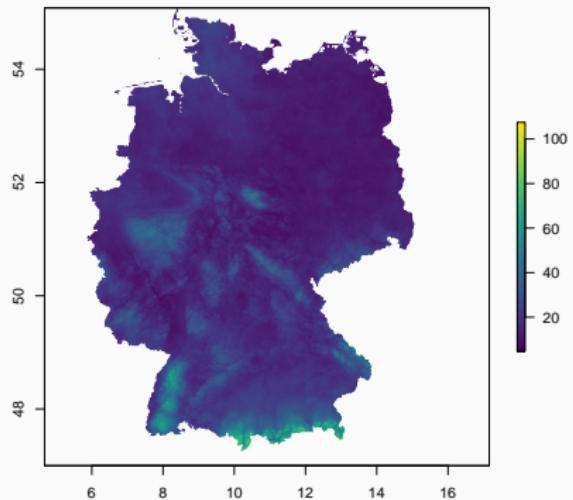
Importance of small streams



- Biodiversity
- Refuge for re-colonisation

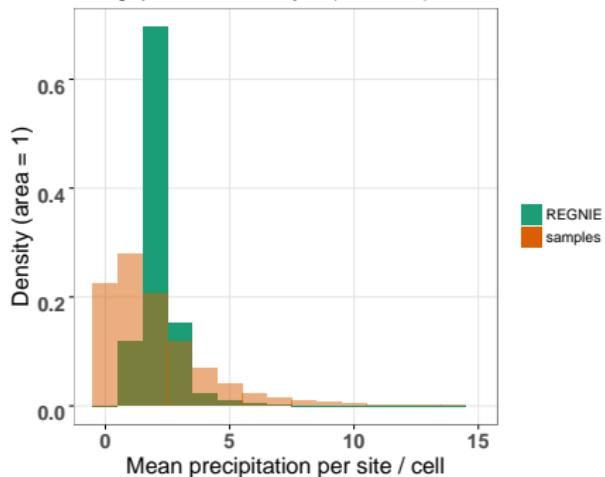
Precipitation in Germany and the samples

Mean # days > 10mm (REGNIE, 2005–2015)



Distribution of precipitation at sites / cell

Average per site / cell within year (2005–2015)



Comparison with other studies

Szöcs (2016)

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

Stehle (2015)

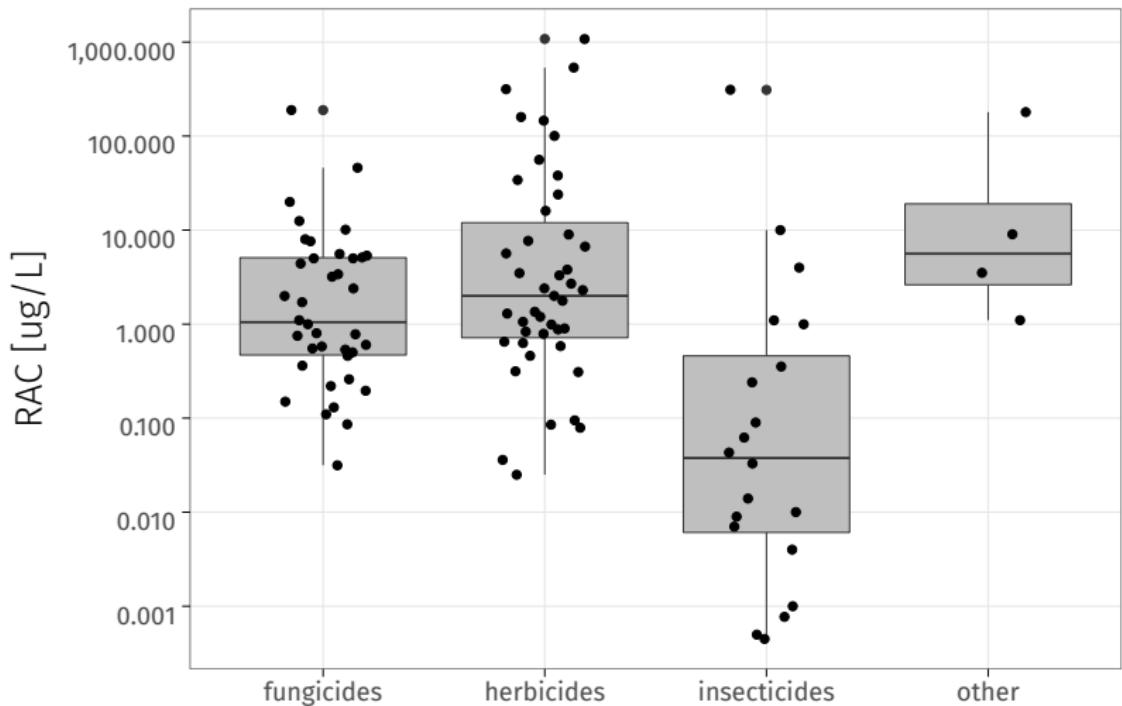
- Europe / Global
- Publications
- Grab & Event driven sampling
- Insecticides
- Organophos.+ Pyrethroids
- LM for >LOQ

Knauer (2016)

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

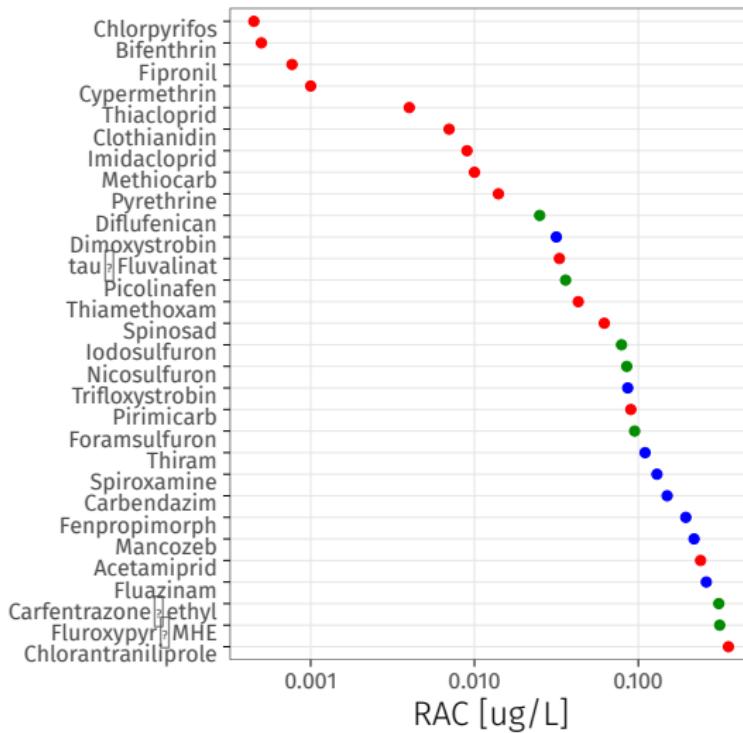
RACs by Type

105 RACs provided by UBA splitted by group



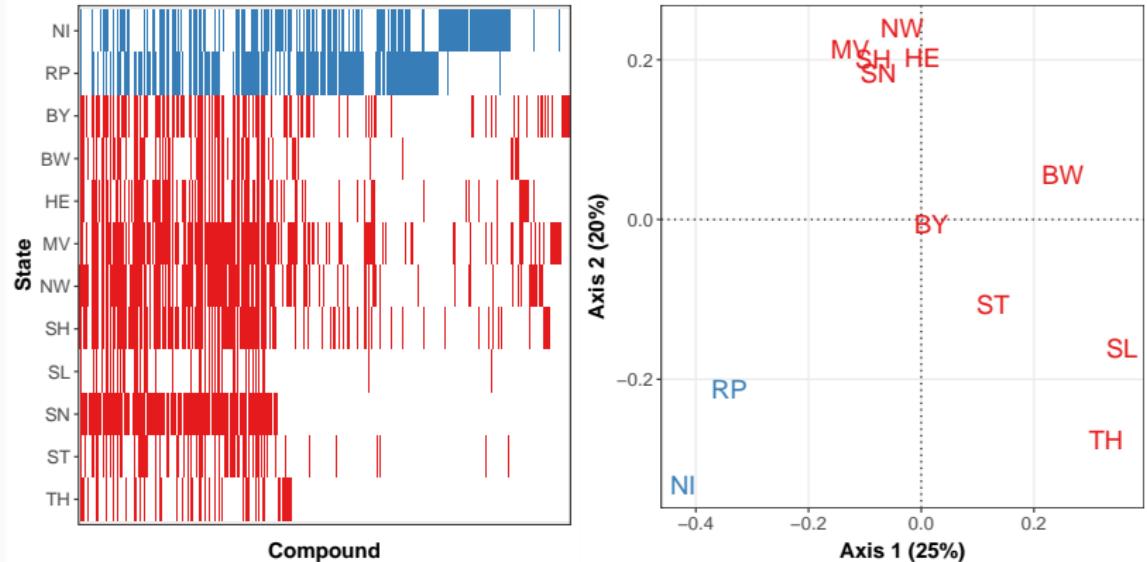
RACs by Compound

30 lowest RACs

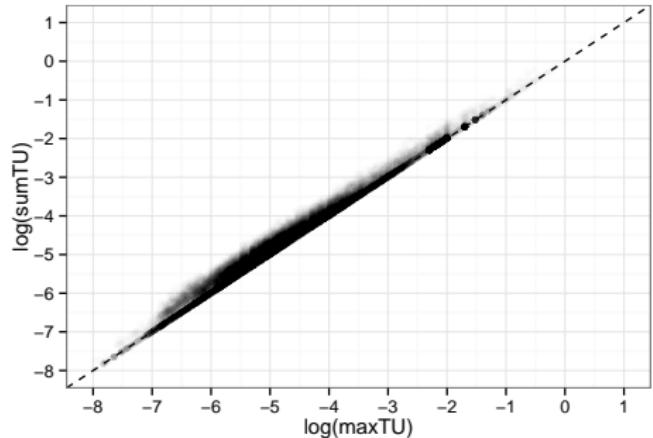


Type • fungicides ● herbicides • insecticides

Analysed compound spectra by state



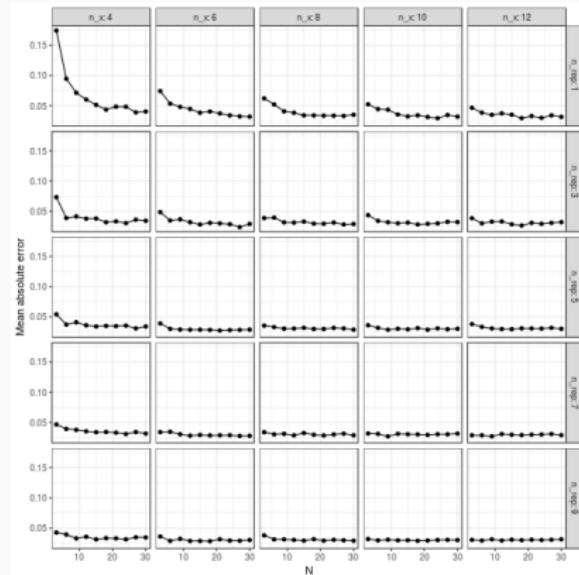
Mixtures are common, but one compound dominates the risk



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

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

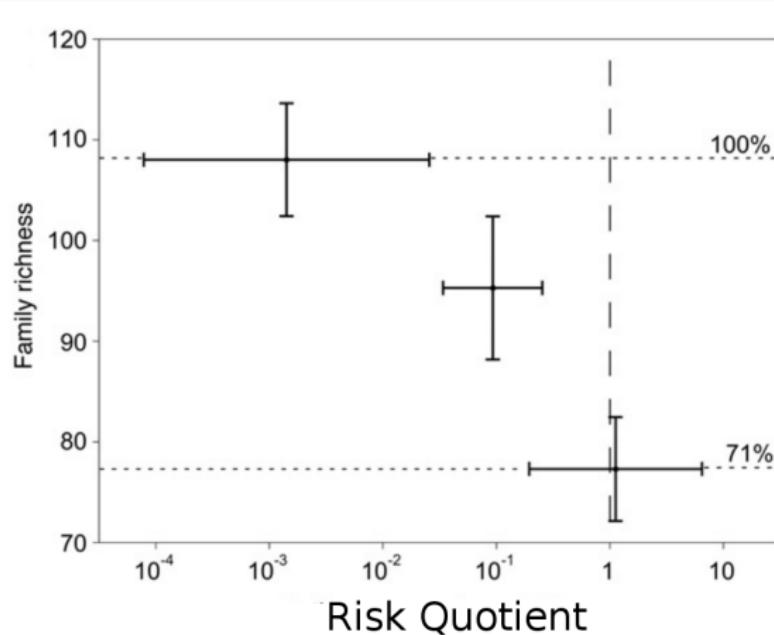
Experimental design for dose-response experiments - a simulation
http://edild.github.io/lc50_bias_sim/



GLM-Explorer: <http://uni-ko-lid.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

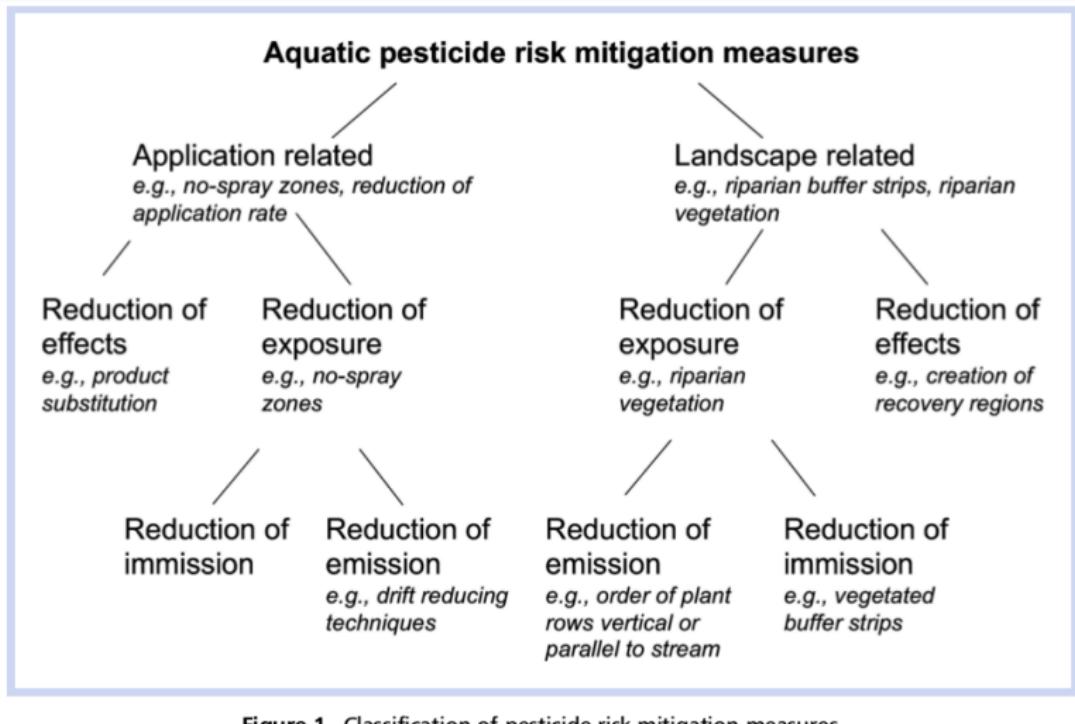
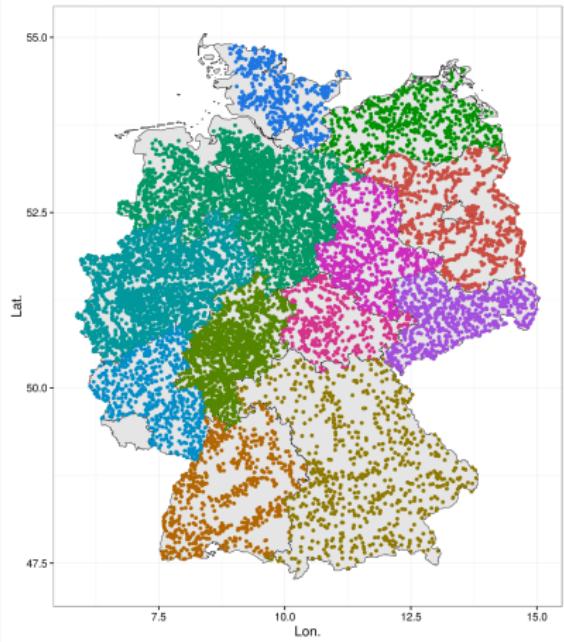


Figure 1. Classification of pesticide risk mitigation measures.

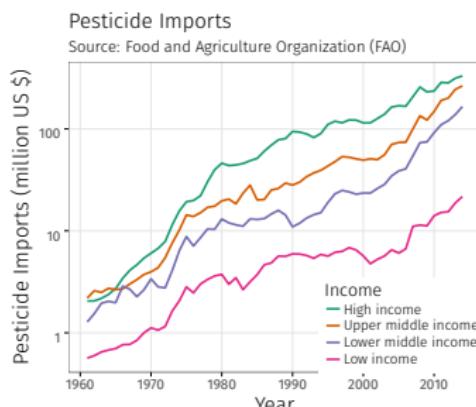
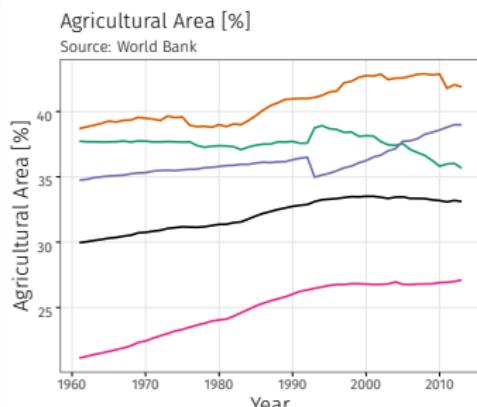
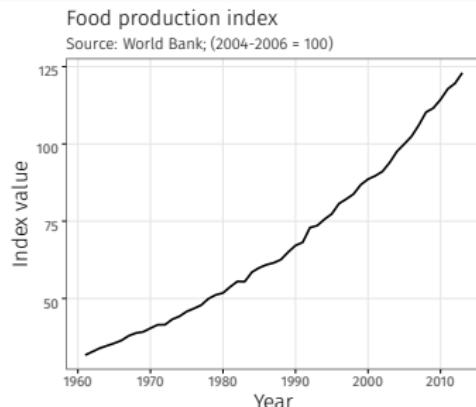
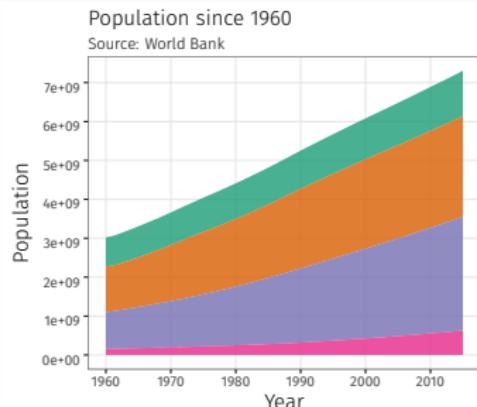
Bereswill, Streloke, Schulz (2014). Risk mitigation measures for diffuse pesticide entry into aquatic ecosystems: Proposal of a guide to identify appropriate measures on a catchment scale: Guide to Identify Pesticide Risk Mitigation Measures. IEAM 10, 286–298.

Biotic field effects



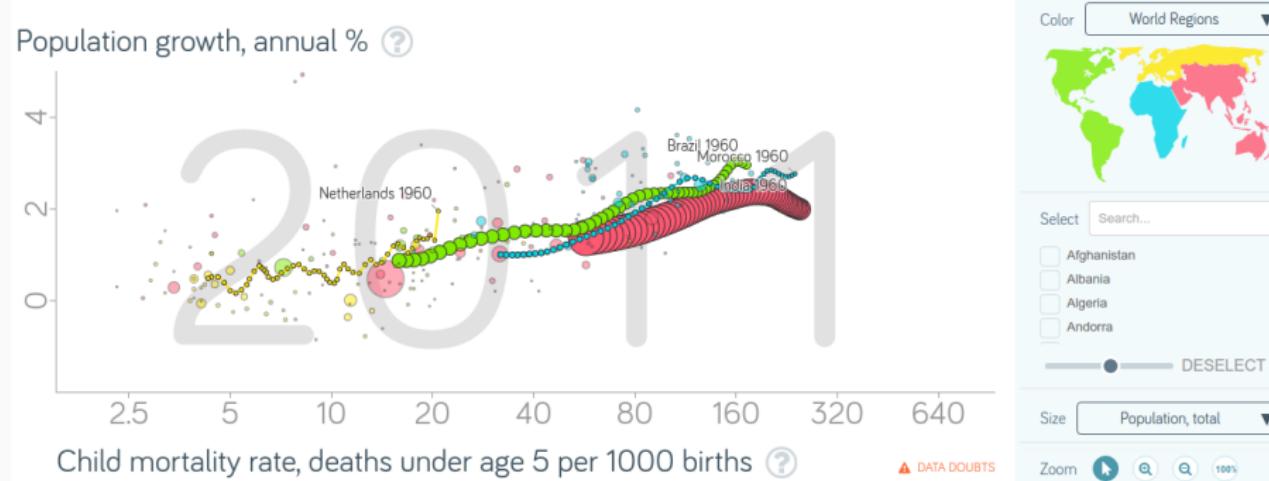
- biological data with good spatial coverage
- 60% of spatial congruence
- Large scale effects largely unknown.
- Some work left...
- Future....

A global perspective (I)



Source code to retrieve data and reproduce results: <http://uni-ko-l.de/g7>

A global perspective (II)



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)

Many Thanks To

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