

Statistical Eco(-toxico)logy

Improving the Utilisation of Data for
Environmental Risk Assessment

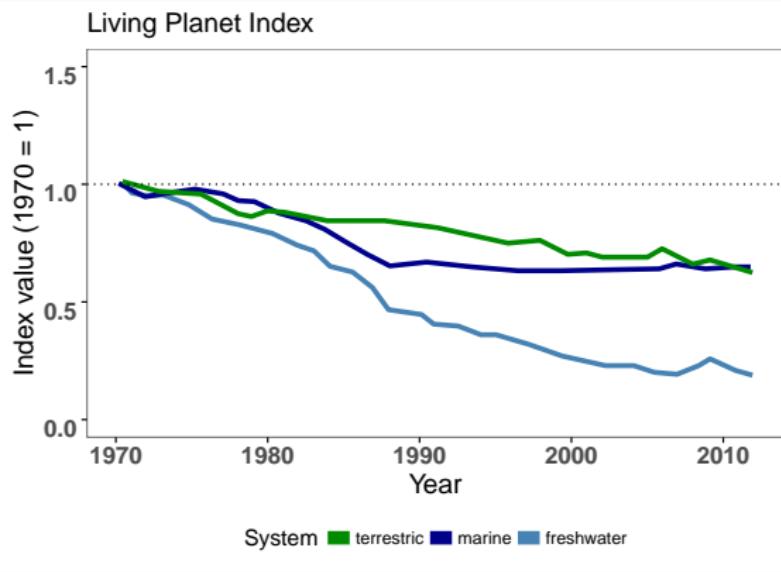
Eduard Szöcs

25th January 2017

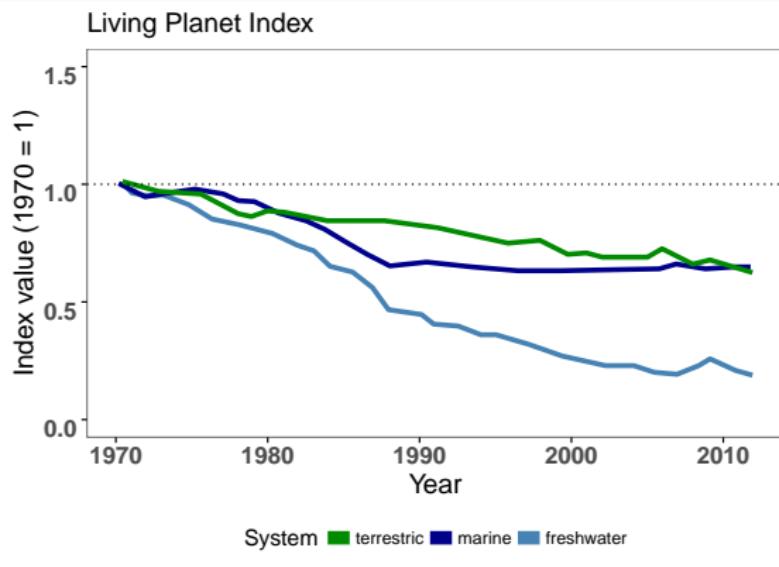
Table of contents

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

Freshwater biodiversity is strongly declining



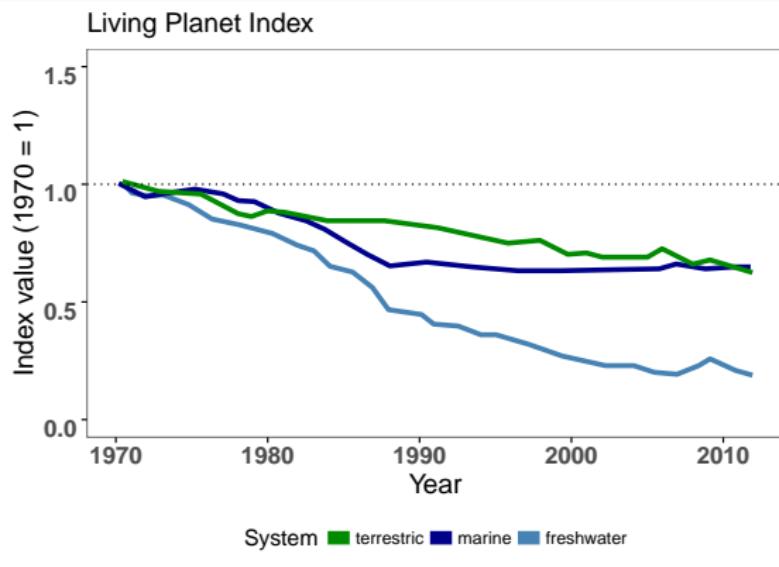
Freshwater biodiversity is strongly declining



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

Effects

Environmental
Risk
Assessment

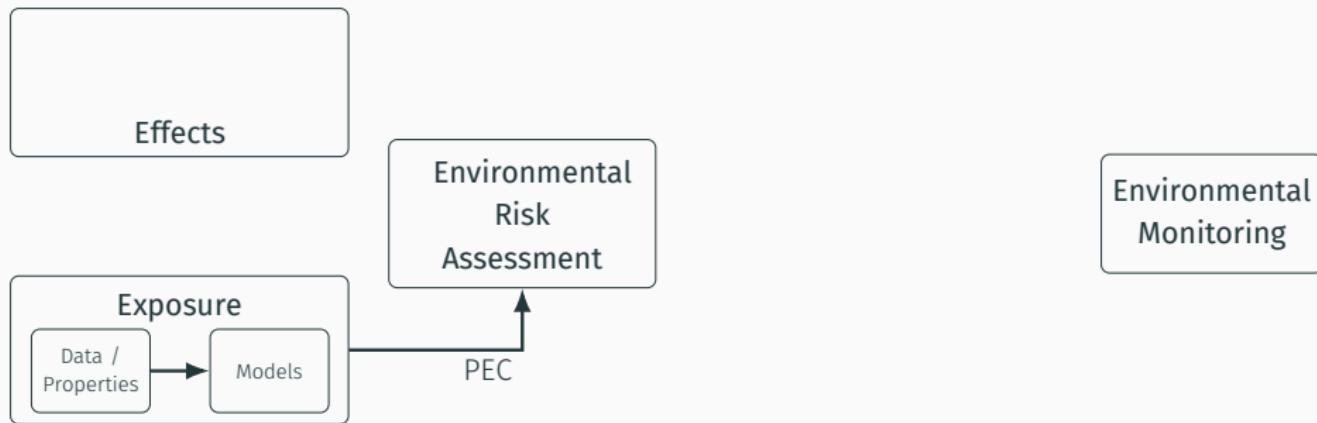
Exposure

Environmental
Monitoring

Environmental Risk Assessment and Monitoring

Plant Protection Products
Regulation 1107/2009

Water Framework Directive
2000/60/EC

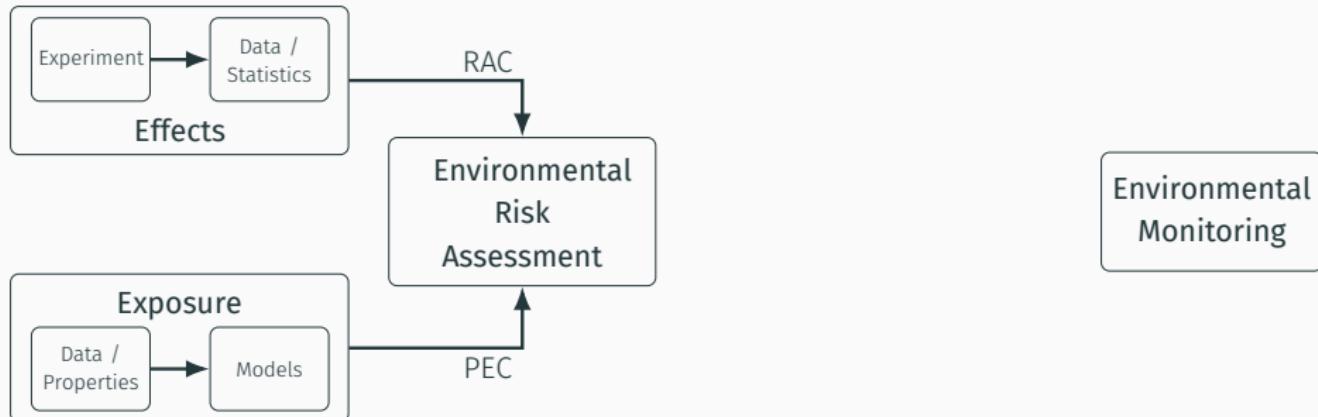


PEC: Predicted Environmental Concentration

Environmental Risk Assessment and Monitoring

Plant Protection Products
Regulation 1107/2009

Water Framework Directive
2000/60/EC



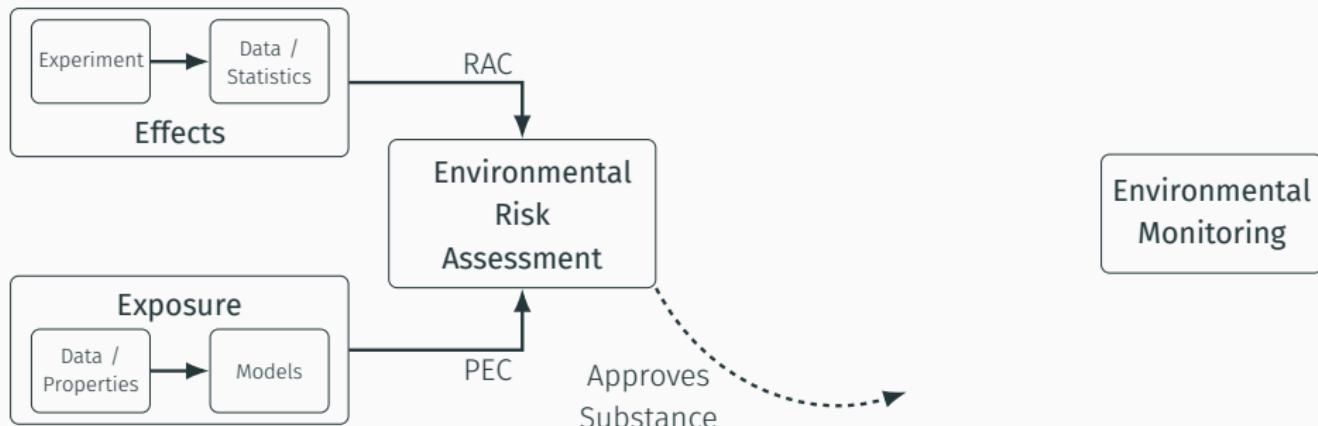
RAC: Regulatory Acceptable Concentration

PEC: Predicted Environmental Concentration

Environmental Risk Assessment and Monitoring

Plant Protection Products
Regulation 1107/2009

Water Framework Directive
2000/60/EC



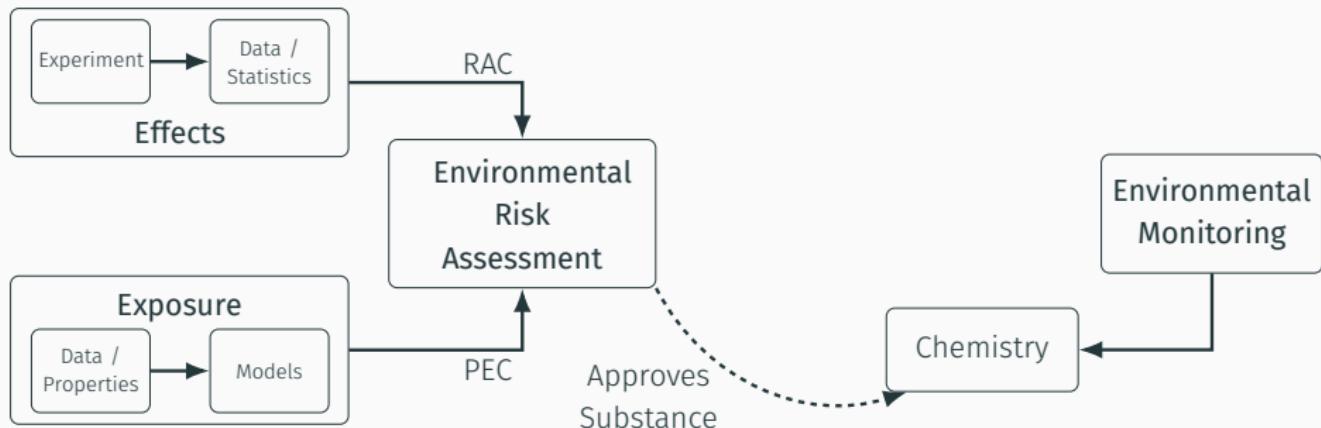
RAC: Regulatory Acceptable Concentration

PEC: Predicted Environmental Concentration

Environmental Risk Assessment and Monitoring

Plant Protection Products
Regulation 1107/2009

Water Framework Directive
2000/60/EC



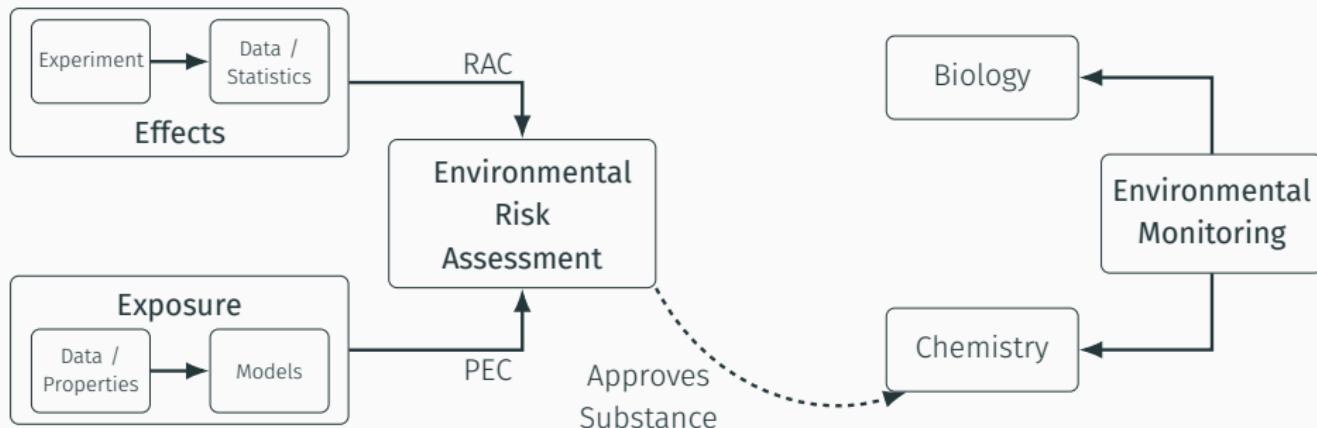
RAC: Regulatory Acceptable Concentration

PEC: Predicted Environmental Concentration

Environmental Risk Assessment and Monitoring

Plant Protection Products
Regulation 1107/2009

Water Framework Directive
2000/60/EC



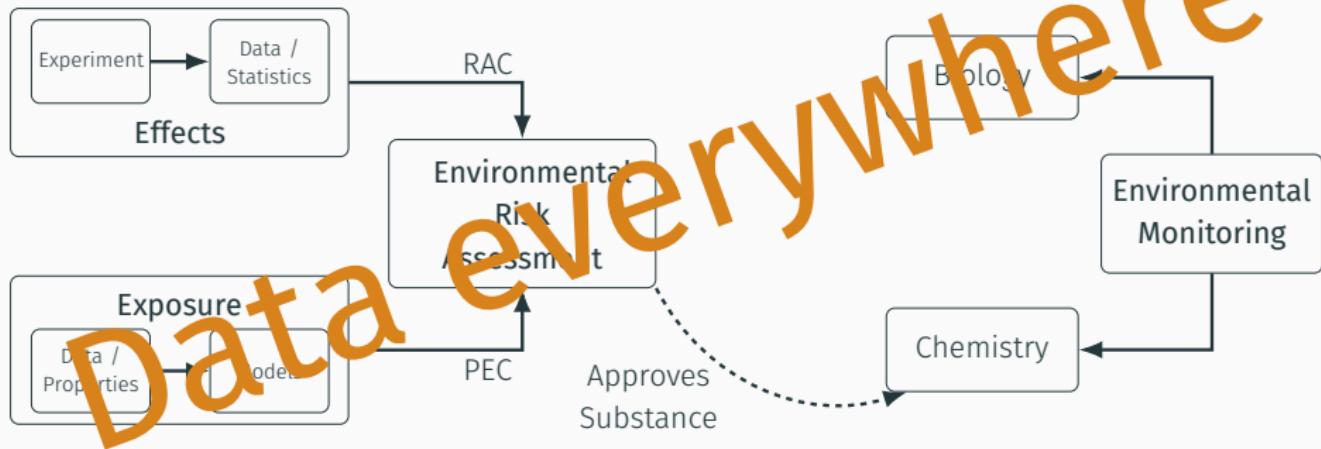
RAC: Regulatory Acceptable Concentration

PEC: Predicted Environmental Concentration

Environmental Risk Assessment and Monitoring

Plant Protection Products
Regulation 1107/2009

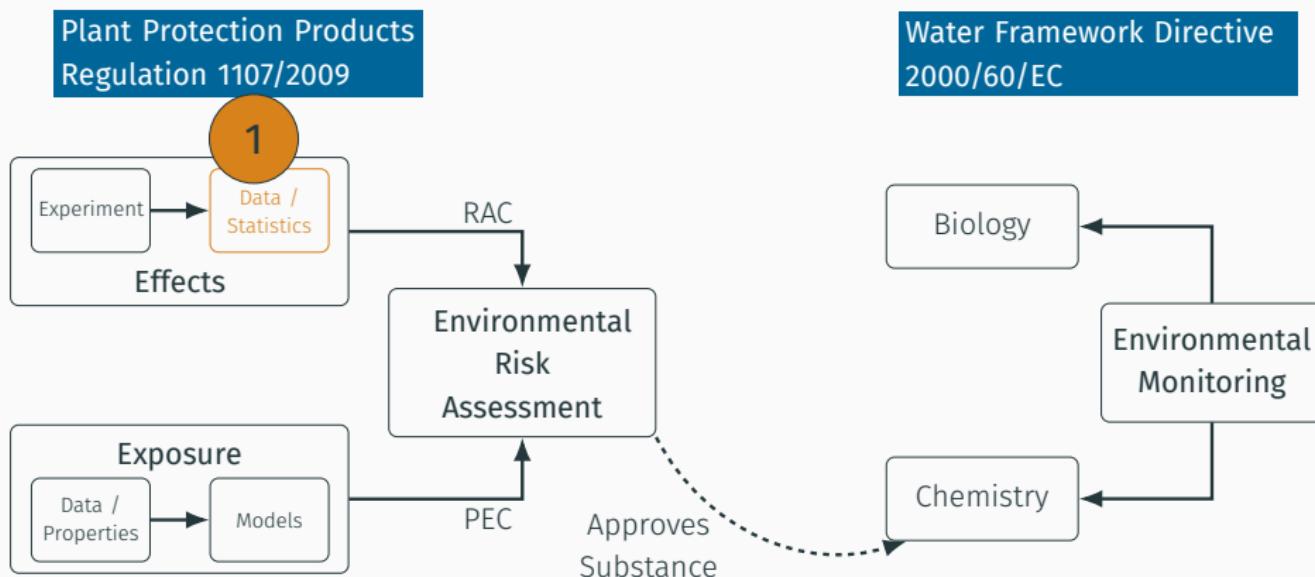
Water Framework Directive
2000/60/EC



RAC: Regulatory Acceptable Concentration

PEC: Predicted Environmental Concentration

Improving Statistics in ERA



1

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



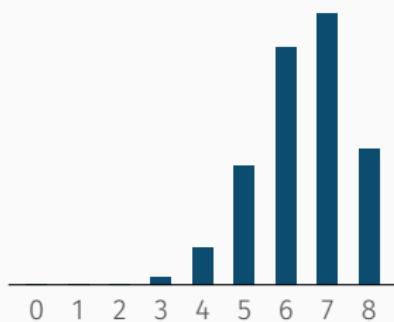
- Daphnia Test
- " x out of n survived"

Experiments in Effect Assessment

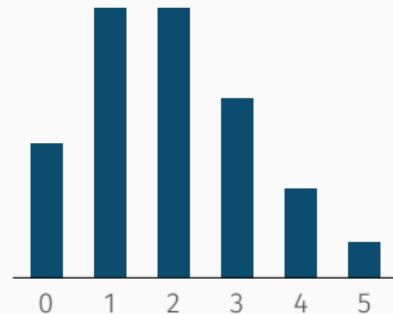


- Daphnia Test
- “*x out of n survived*”
- Mesocosm
- “*number of animals*”

Ecotoxicology is not normal

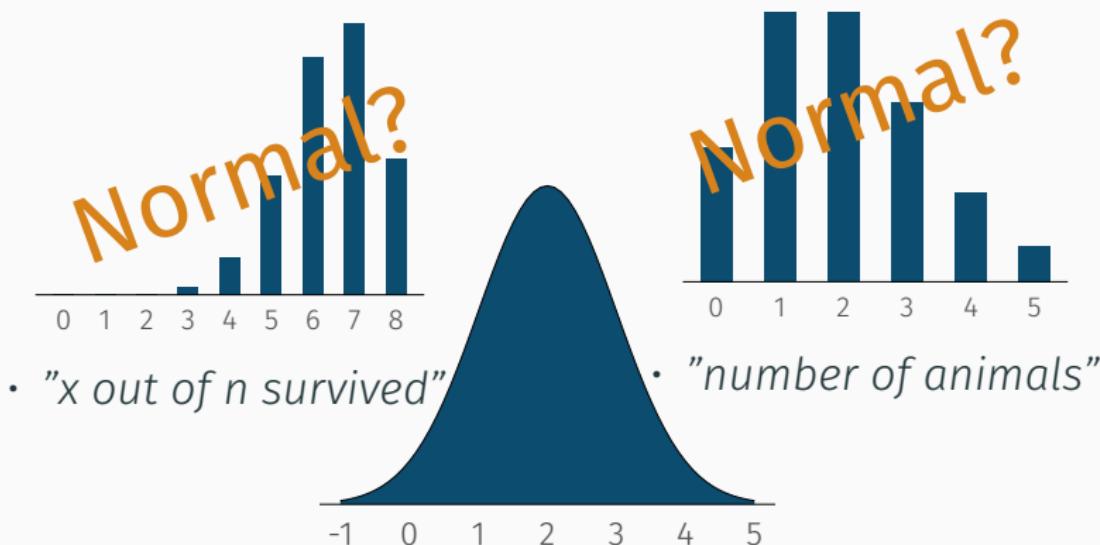


- " x out of n survived"



- "number of animals"

Ecotoxicology is not normal



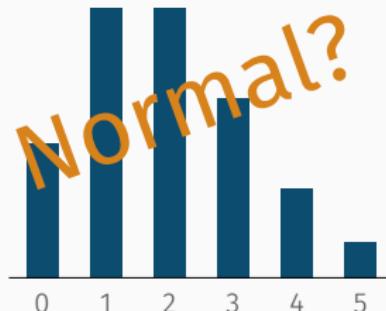
Ecotoxicology is not normal

Normal?

0 1 2 3 4 5 6 7 8

- "x out of n survived"

-1 0 1 2 3 4 5



- "number of animals"

- ignore?

Ecotoxicology is not normal

Normal?

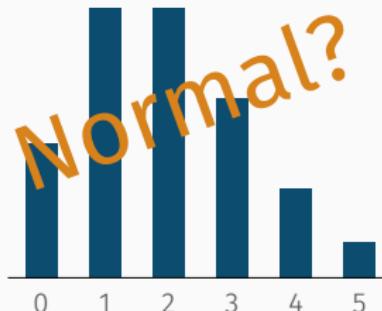
0 1 2 3 4 5 6 7 8

- "x out of n survived"

-1 0 1 2 3 4 5

ignore?

transform?



- "number of animals"

Ecotoxicology is not normal

Normal?

0 1 2 3 4 5 6 7 8

- "x out of n survived"

-1 0 1 2 3 4 5

- ignore?
- transform?
- non-parametric?

Normal?

0 1 2 3 4 5

- "number of animals"

Ecotoxicology is not normal

Normal?

0 1 2 3 4 5 6 7 8

- " x out of n survived"



- "number of animals"

-1 0 1 2 3 4 5

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

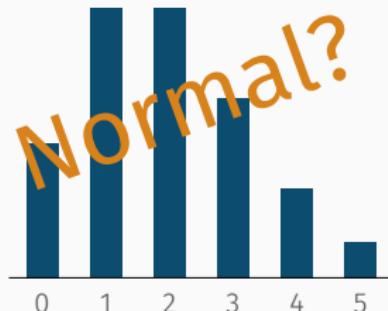
Ecotoxicology is not normal

Normal?

0 1 2 3 4 5 6 7 8

- "x out of n survived"

-1 0 1 2 3 4 5



- "number of animals"

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

GLMs may have better properties

Model	Observations	Predictors	Link
LM	$Y \sim N(\mu, R)$	$\eta = X\beta$	$\eta = \mu$

GLMs may have better properties

Model	Observations	Predictors	Link
LM	$Y \sim N(\mu, R)$	$\eta = X\beta$	$\eta = \mu$
GLM	$Y \sim G(\mu, R)$	$\eta = X\beta$	$\eta = g(\mu)$

GLMs may have better properties

Model	Observations	Predictors	Link
LM	$Y \sim N(\mu, R)$	$\eta = X\beta$	$\eta = \mu$
GLM	$Y \sim G(\mu, R)$	$\eta = X\beta$	$\eta = g(\mu)$

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

Ecology, 92(1), 2011, pp. 3–10
© 2011 by the Ecological Society of America

The arcsine is asinine: the analysis of proportions in ecology

DAVID I. WARTON^{1,2,3} AND FRANCIS K. C. HUI¹

GLMs in Ecotoxicology?

Unclassified

ENV/JM/MONO(2006)18



Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

09-May-2006

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND
THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

ENV/JM/MONO(2006)18
Unclassified

OECD SERIES ON TESTING AND ASSESSMENT
Number 54

**CURRENT APPROACHES IN THE STATISTICAL ANALYSIS OF ECOTOXICITY DATA: A
GUIDANCE TO APPLICATION**

GLMs in Ecotoxicology?



Unclassified

ENV/JM/MONO(2006)18

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development

09-May-2006

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND
THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

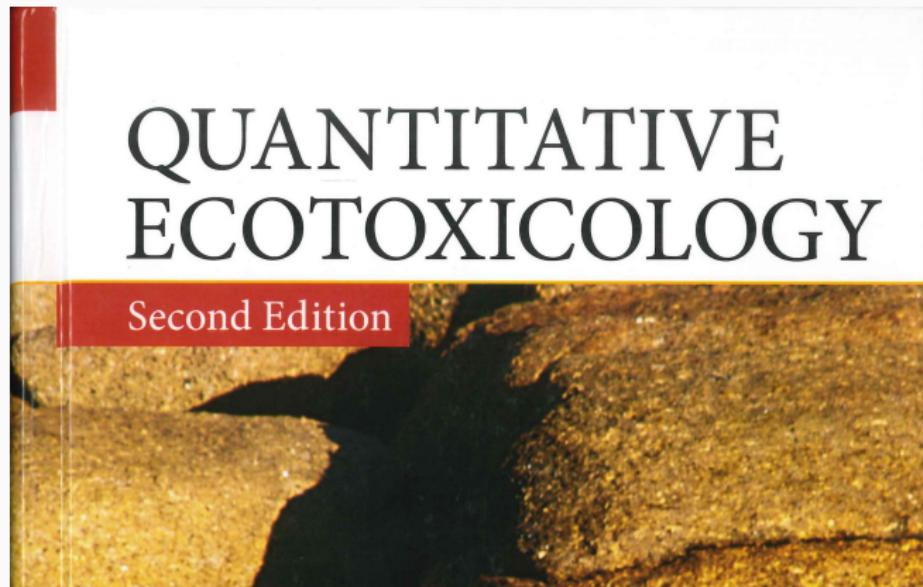
**"GLMs are not discussed
in this document."**

ENV/JM/MONO(2006)18
Unclassified

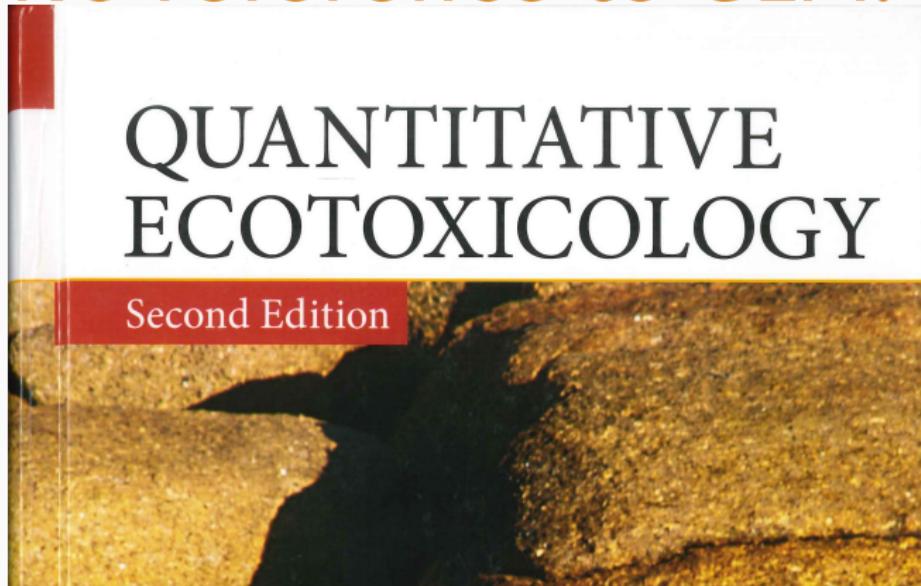
OECD SERIES ON TESTING AND ASSESSMENT
Number 54

CURRENT APPROACHES IN THE STATISTICAL ANALYSIS OF ECOTOXICITY DATA: A
GUIDANCE TO APPLICATION

GLMs in Ecotoxicology?



No reference to GLM.



Method: Simulations

Scenarios	Simulations (n = 1000)	Methods	Statistics	Performance
-----------	---------------------------	---------	------------	-------------

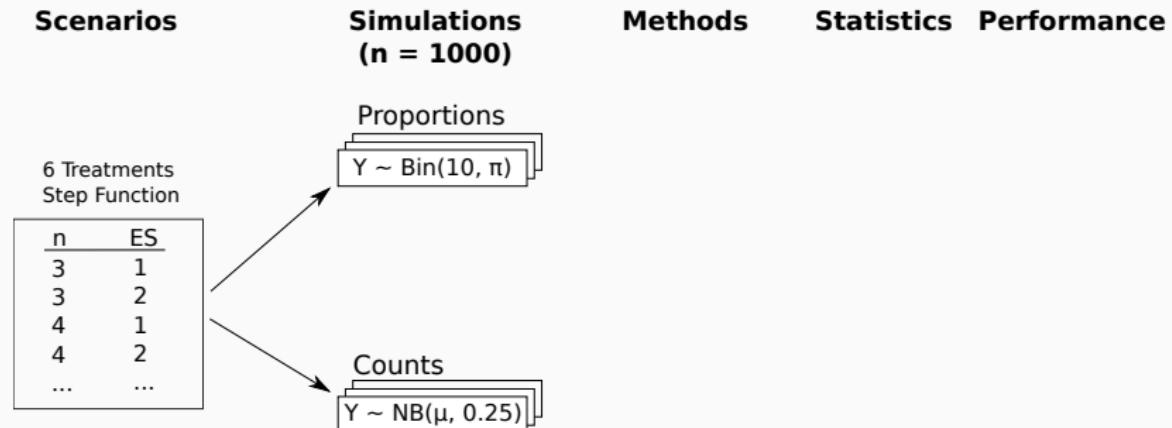
Method: Simulations

Scenarios **Simulations
(n = 1000)** **Methods** **Statistics** **Performance**

6 Treatments
Step Function

n	ES
3	1
3	2
4	1
4	2
...	...

Method: Simulations



Method: Simulations

Scenarios

6 Treatments
Step Function

Simulations (n = 1000)

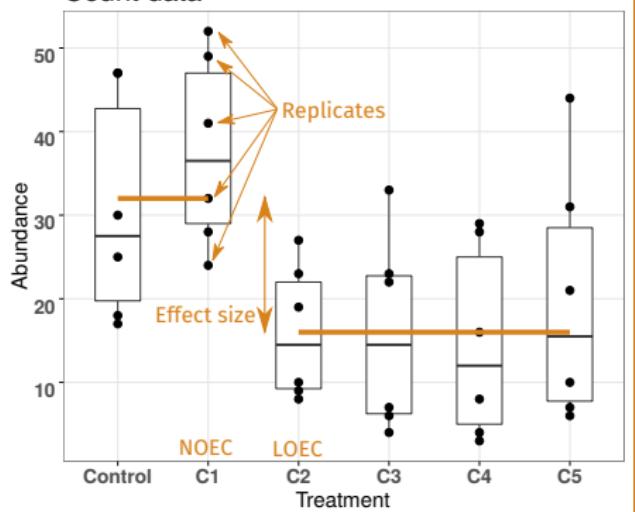
Proportions

$$Y \sim \text{Bin}(10, \pi)$$

Methods

Statistics Performance

Count data



Method: Simulations

Scenarios

6 Treatments
Step Function

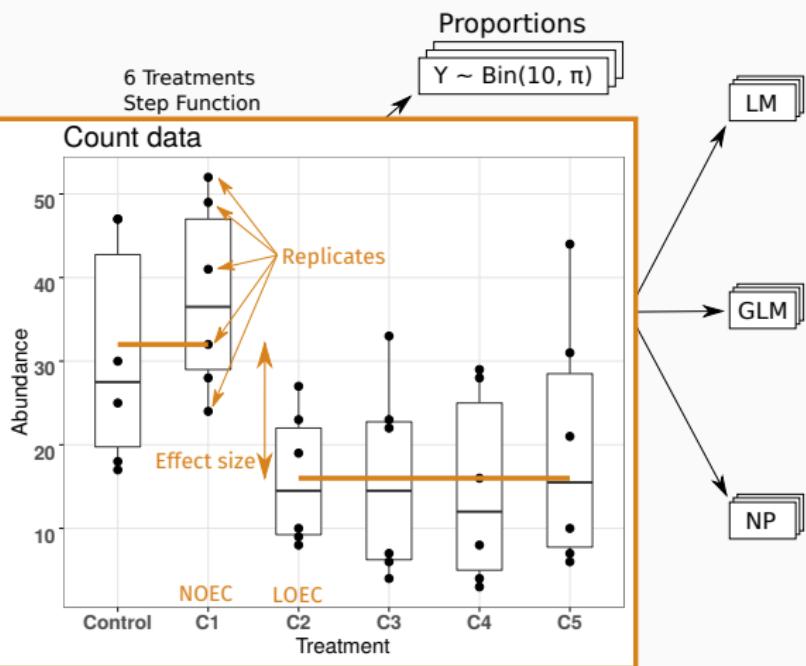
Simulations (n = 1000)

Proportions

$$Y \sim \text{Bin}(10, \pi)$$

Methods

Statistics Performance



LM

GLM

NP

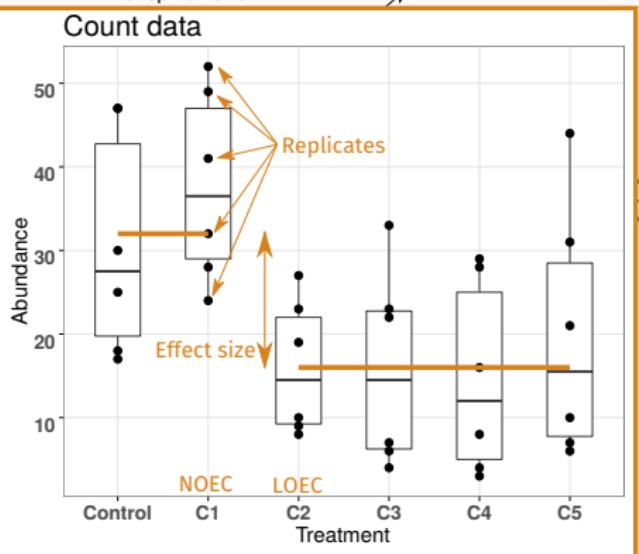
Method: Simulations

Scenarios

6 Treatments
Step Function

Simulations (n = 1000)

Proportions
 $Y \sim \text{Bin}(10, \pi)$



Methods

LM

Statistics

global

LOEC

GLM

global

LOEC

NP

global

LOEC

Method: Simulations

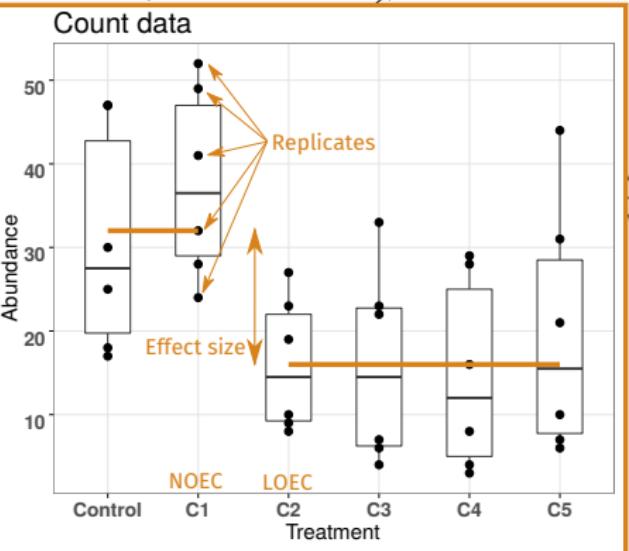
Scenarios

6 Treatments
Step Function

Simulations (n = 1000)

Proportions

$$Y \sim \text{Bin}(10, \pi)$$



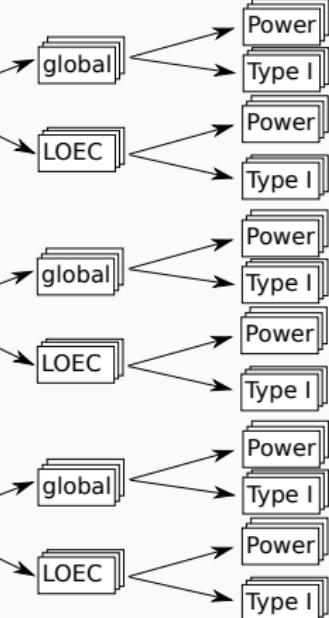
Methods

LM

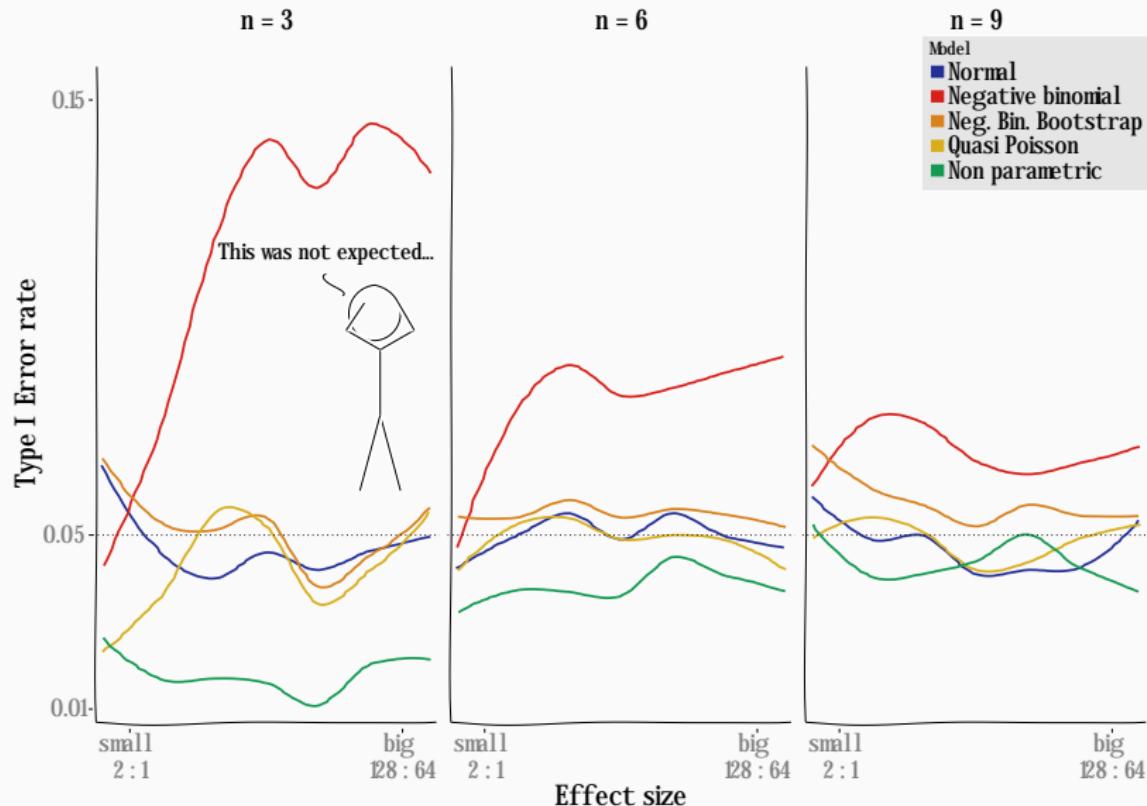
GLM

NP

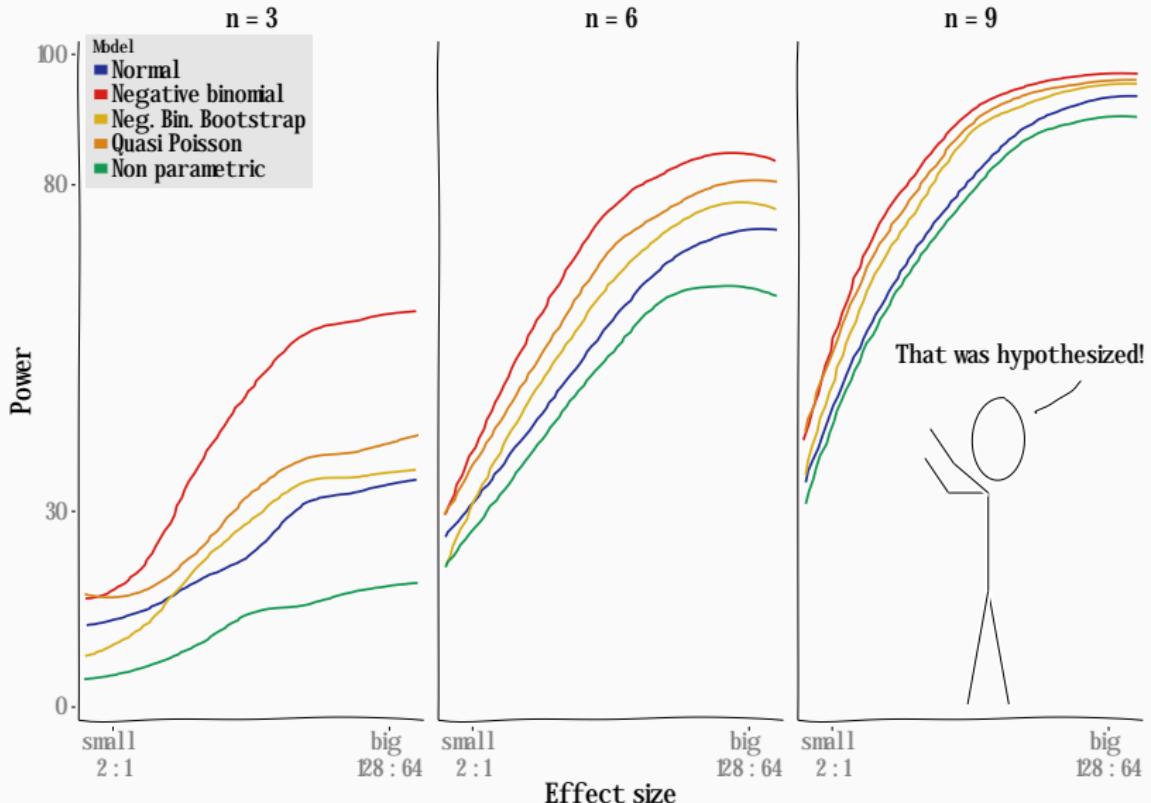
Statistics Performance



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

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?

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;

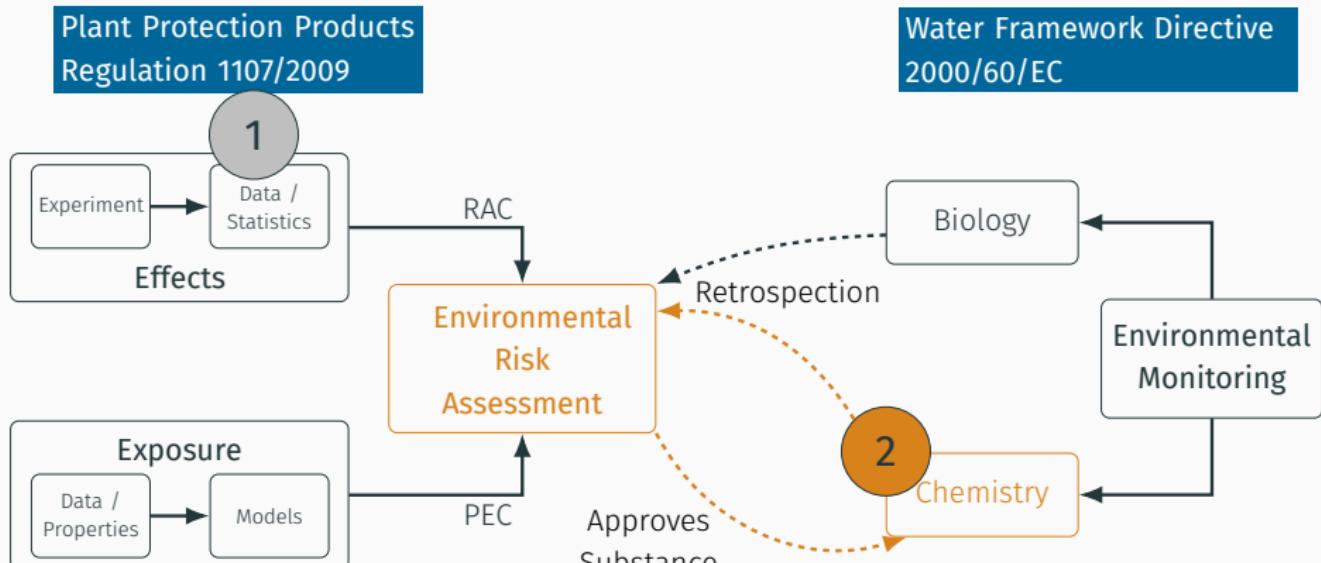
²School of Biological, Earth and Environmental Sciences, University of New South Wales, NSW 2052, Australia; and

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

State of the art

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

Identifying Risks using Monitoring Data



2

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

Identifying Risks using Monitoring Data

Goals & Hypotheses

Goal: Combine monitoring and ERA

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

Goals & Hypotheses

Goal: Combine monitoring and ERA

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

Hypotheses

1. Agriculture: \uparrow Risk: \uparrow

Goals & Hypotheses

Goal: Combine monitoring and ERA

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

Hypotheses

1. Agriculture: \uparrow Risk: \uparrow
2. Stream size: \downarrow Risk: \uparrow

Goals & Hypotheses

Goal: Combine monitoring and ERA

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

Hypotheses

1. Agriculture: \uparrow Risk: \uparrow
2. Stream size: \downarrow Risk: \uparrow
3. Precipitation: \uparrow Risk: \uparrow

Goals & Hypotheses

Goal: Combine monitoring and ERA

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

Hypotheses

1. Agriculture: \uparrow Risk: \uparrow
2. Stream size: \downarrow Risk: \uparrow
3. Precipitation: \uparrow Risk: \uparrow
4. Annual dynamics - Summer: \uparrow

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

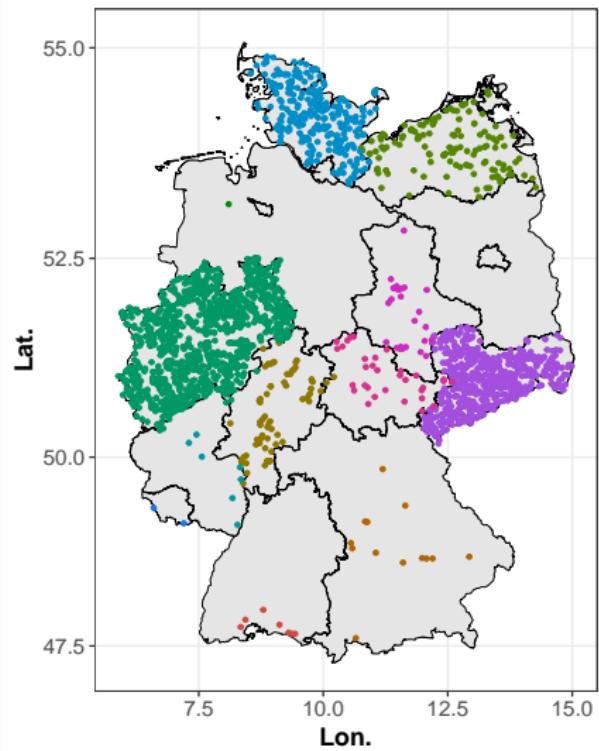
- Concentrations < LOQ (96% of all measurements)
- Hurdle-model:

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

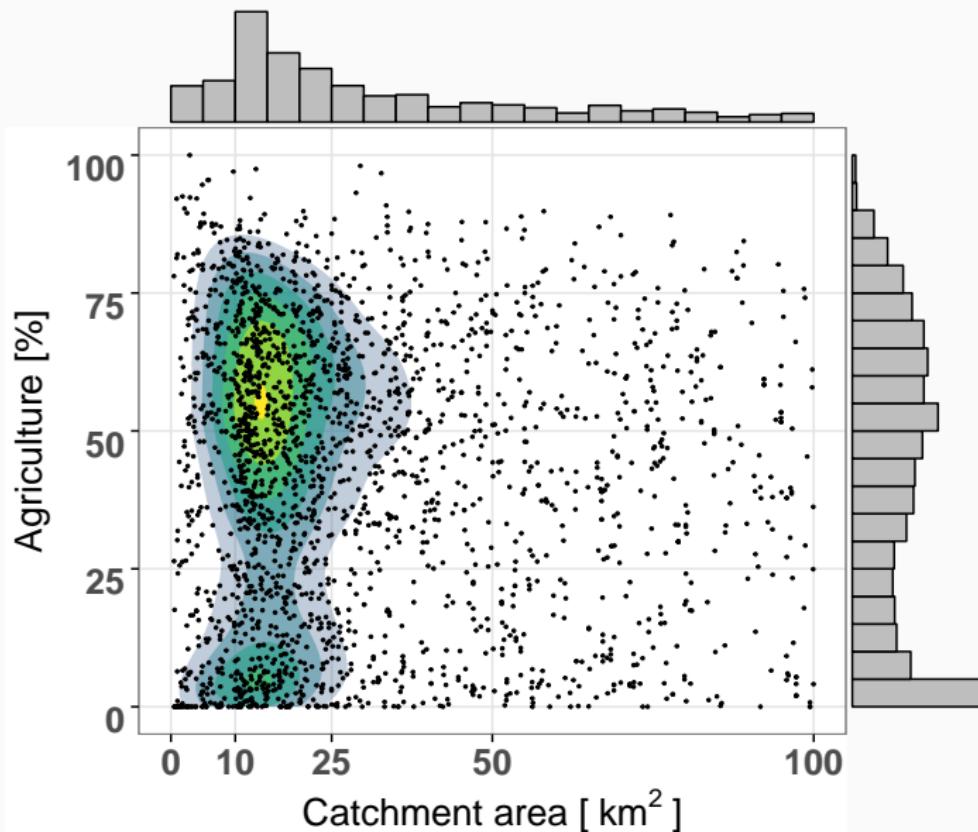
- Risk Quotient
 - $RQ = \frac{c}{RAC}$
- Predictors
 - Catchment size
 - Agricultural land use
 - Precipitation

Compiled data: Big, but inhomogeneous

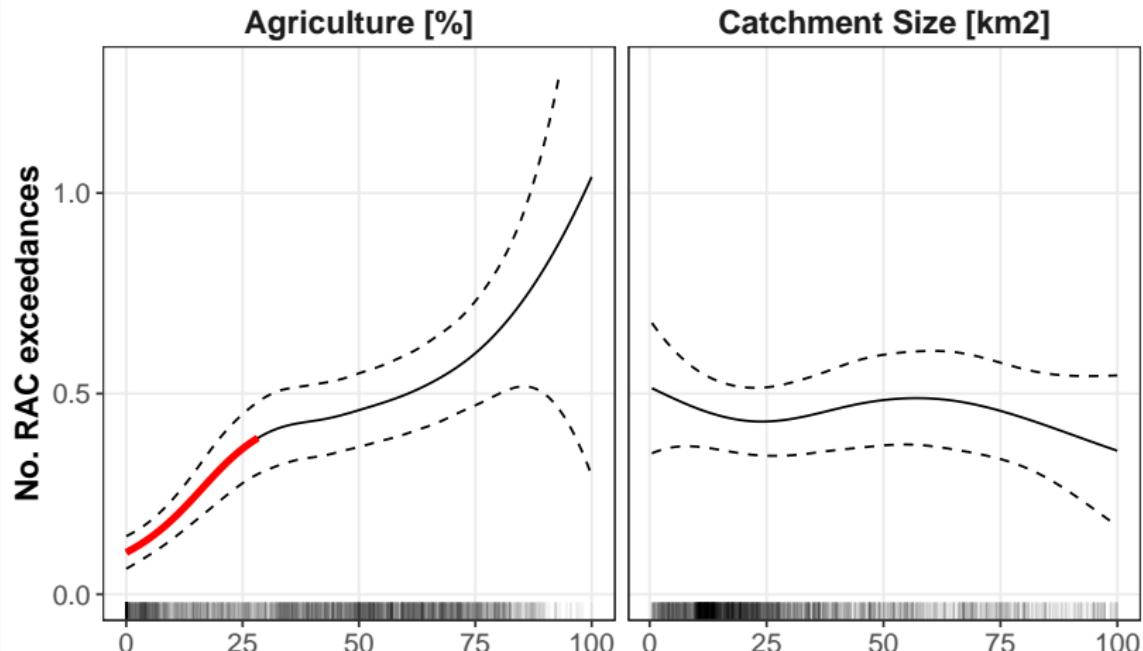
- ~ 1.8M measurements
- ~ 500 pesticides
- ~ 25,000 samples
- ~ 2,300 sites



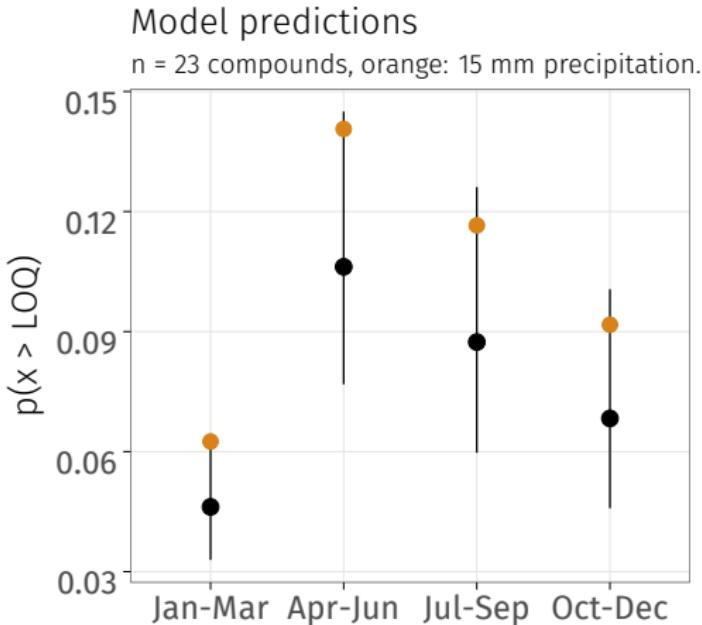
Monitoring: Small streams are underrepresented



Landscape: Factors influencing risk



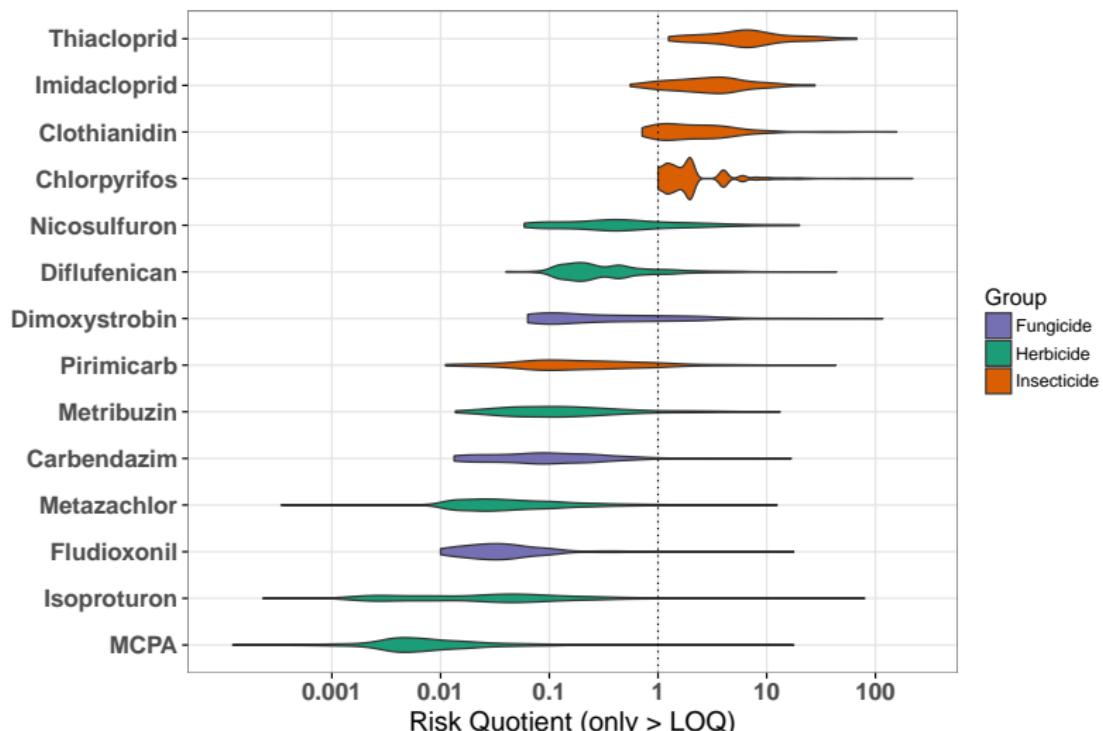
Sampling: Factors influencing risk



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

Risks: Compounds exceeding risk thresholds

- 25% of sites with at least one RQ > 1

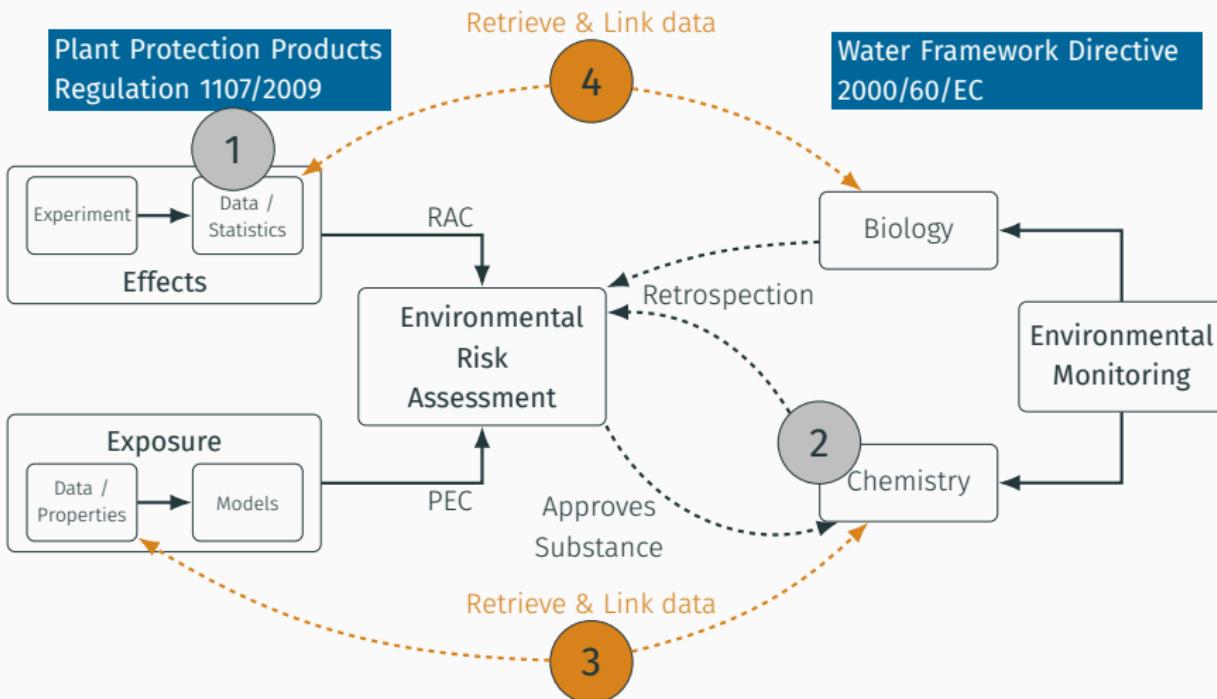


What we learned from this study

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

Solutions for Data Handling

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) 21/26

3

4

Biologists & Chemists face the same problems

Biologists & Chemists face the same problems

Names

Osmia rufa, *Osmia bicornis*,
Osmia ruffa, *Osmiaxxxx*

Chlorpyrifos, Chlorpyriphos,
Chlorphyrifos, Chlorpypifot

Biologists & Chemists face the same problems

Names

<i>Osmia rufa</i> , <i>Osmia bicornis</i> ,	Chlorpyrifos, Chlorpyriphos,
<i>Osmia ruffa</i> , <i>Osmiaxxxx</i>	Chlorphyrifos, Chlorpypifot

Hierarchies

Hymenoptera / Apoidea /	organophosphate, ester,
Megachilidae / Osmia / rufa	insecticide

Biologists & Chemists face the same problems

Names

<i>Osmia rufa</i> , <i>Osmia bicornis</i> ,	Chlorpyrifos, Chlorpyriphos,
<i>Osmia ruffa</i> , <i>Osmiaxxxx</i>	Chlorphyrifos, Chlorpypifot

Hierarchies

Hymenoptera / Apoidea /	organophosphate, ester,
Megachilidae / Osmia / rufa	insecticide

Traits / Properties

Wing length, Mass, Season	Mass, K_{OW} , LC_{50}
---------------------------	----------------------------

Biologists & Chemists face the same problems

Names

<i>Osmia rufa</i> , <i>Osmia bicornis</i> ,	Chlorpyrifos, Chlorpyriphos,
<i>Osmia ruffa</i> , <i>Osmiaxxxx</i>	Chlorphyrifos, Chlorpypifot

Hierarchies

Hymenoptera / Apoidea /	organophosphate, ester,
Megachilidae / Osmia / rufa	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

<i>Osmia rufa</i> , <i>Osmia bicornis</i> ,	Chlorpyrifos, Chlorpyriphos,
<i>Osmia ruffa</i> , <i>Osmia</i> xxxx	Chlorphyrifos, Chlorpypifot

Hierarchies

Hymenoptera / Apoidea /	organophosphate, ester,
Megachilidae / Osmia / rufa	insecticide

Traits / Properties

Wing length, Mass, Season	Mass, K_{OW} , LC_{50}
---------------------------	----------------------------

Identifiers

NCBI, ITIS, EOL, ...	2921-88-2, SMILES, InChI, ...
----------------------	-------------------------------

Amount of data

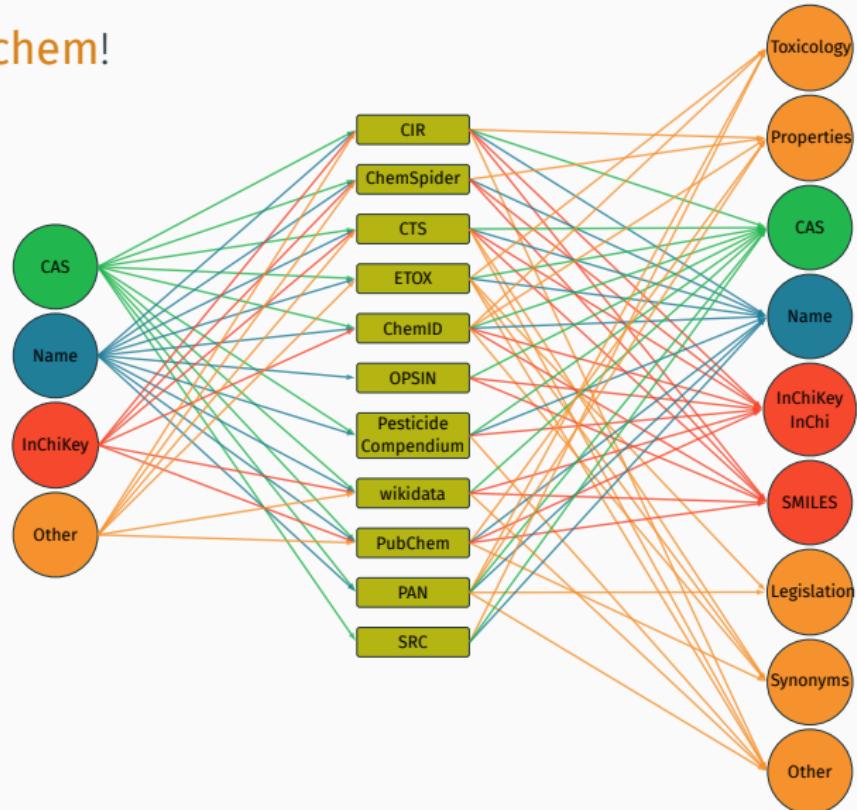
~3000 taxa

~500 pesticides

22/26

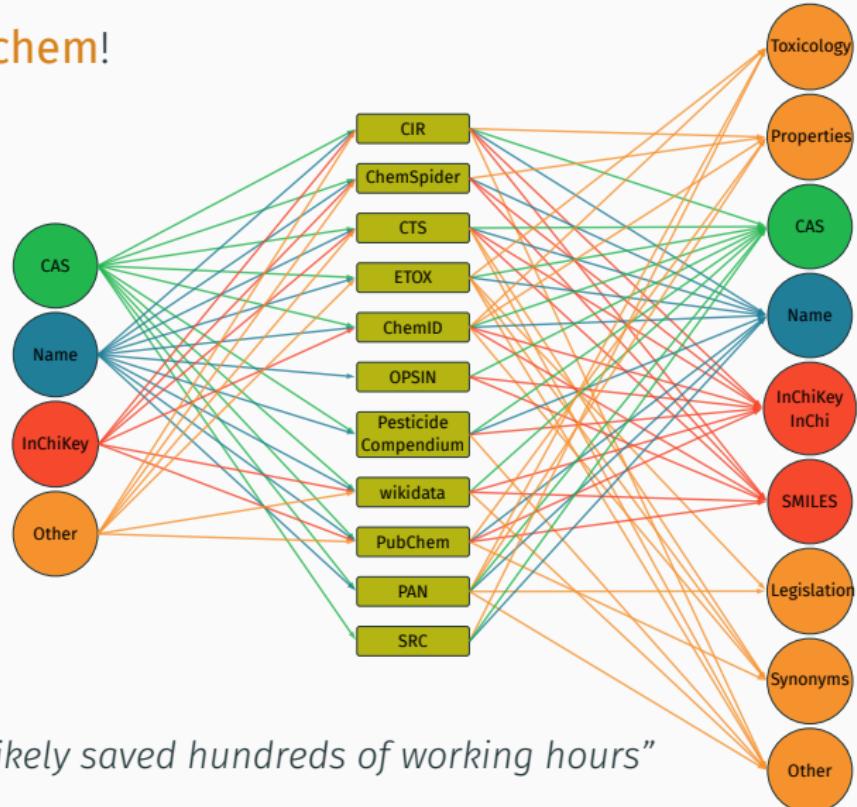
Instead of wasting time...

... use webchem!



Instead of wasting time...

... use webchem!



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

Instead of wasting time...

... use **taxize!**



Global Invasive Species
Database



Catalogue of Life

Plantminer



uBio



Canadensys

ThePlantList

Instead of wasting time...

... use **taxize!**



Global Invasive Species Database

Catalogue of Life

Plantminer

Tropicos®

eol

gni

RED LIST

Susan Johnston
@SuseJohnston

Folgen

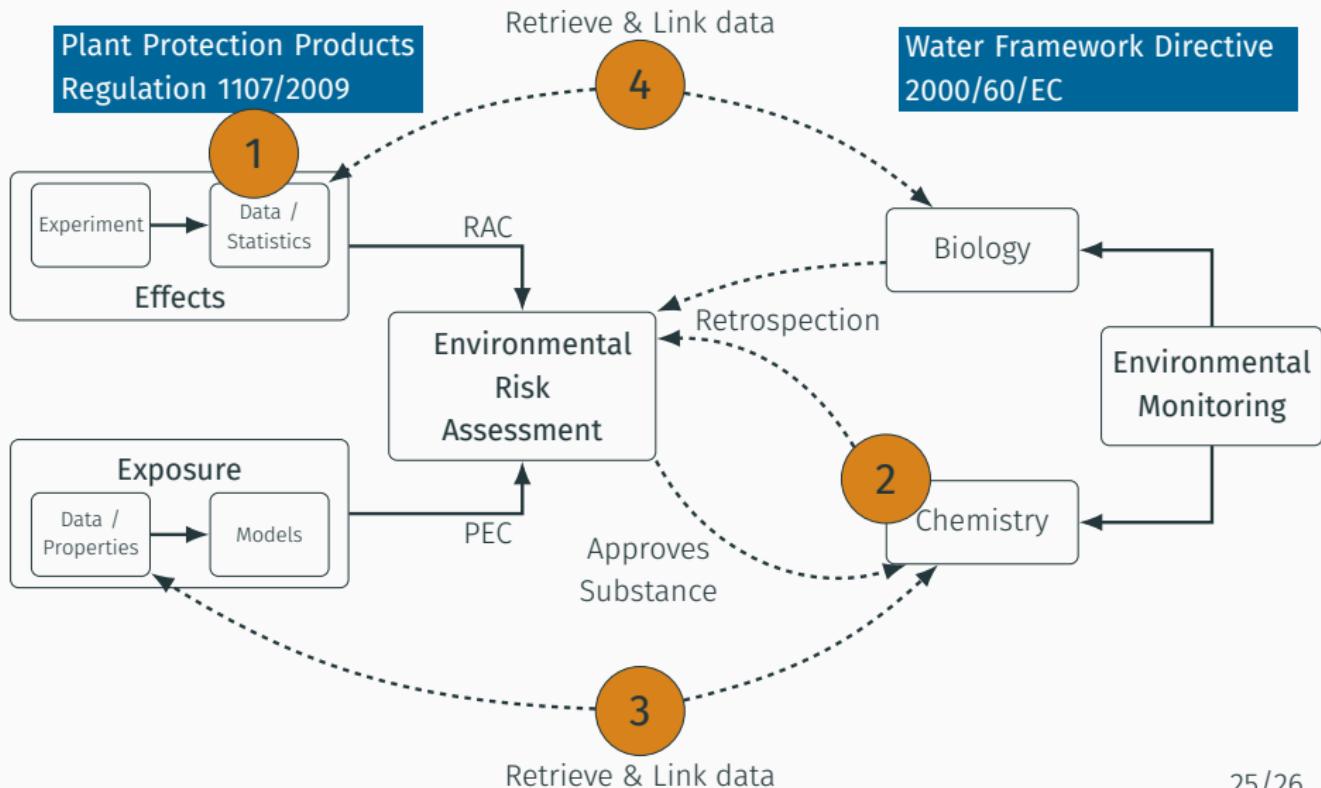
Discovered taxize taxonomic toolbelt for R on @rOpenSci. Days of searching done during my morning coffee. Amazing. ropensci.org/tutorials/taxi ...

Übersetzung anzeigen

RETWEETS	GEFÄLLT
8	17

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

✓ Identifying Risks using Monitoring data

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

Recap: What we learned from my PhD Thesis

✓ Improving Statistics in ERA

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

✓ Identifying Risks using Monitoring data

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

✓ Solutions for Data Handling

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

Statistical Ecotoxicology

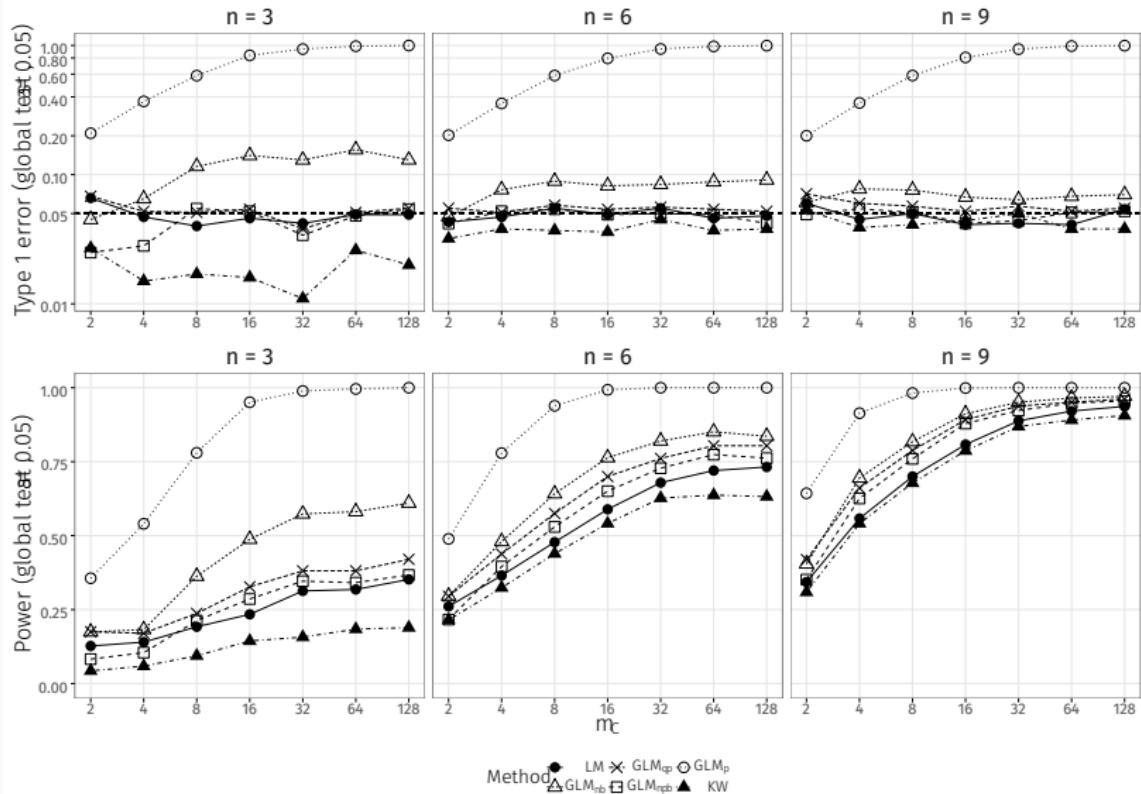
Improving the Utilisation of Data for Environmental Risk Assessment

Eduard Szöcs

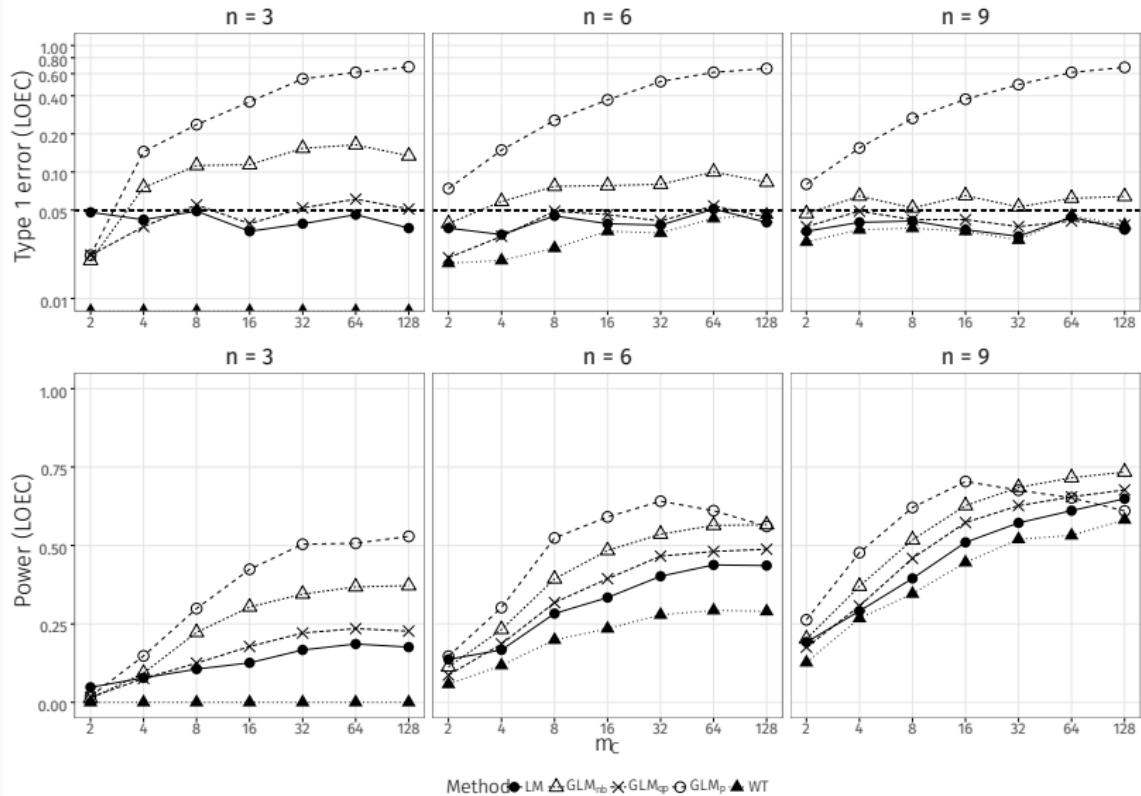
- ❑ <http://edild.github.io/>
- 🐦 [@EduardSzoebs](https://twitter.com/EduardSzoebs)
- 📄 https://github.com/edild/phd_defense
- 📝 https://github.com/edild/phd_thesis



Power en detail



For LOEC it is even worse

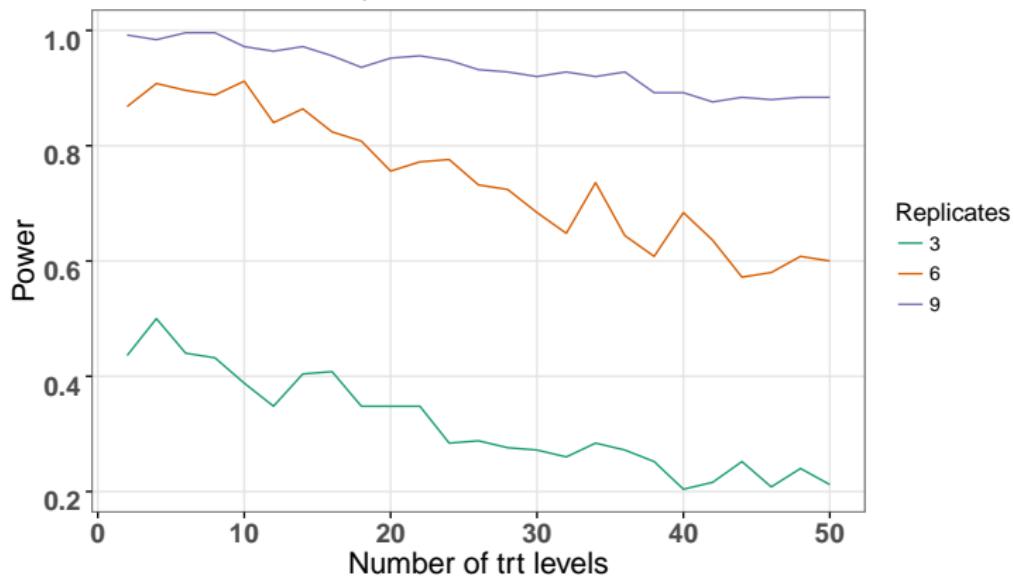


Effect of number of treatments

Power analysis for different numbers of treatment levels.

Treatment mean = 100, last treatment was reduced to 80.

Distribution: Norman; Link: Identity; Sigma_res = 10;
250 simulation runs per scenario



Source code: <http://uni-ko-ld.de/gz>

Comparison with Ives...

Szöcs (2015)

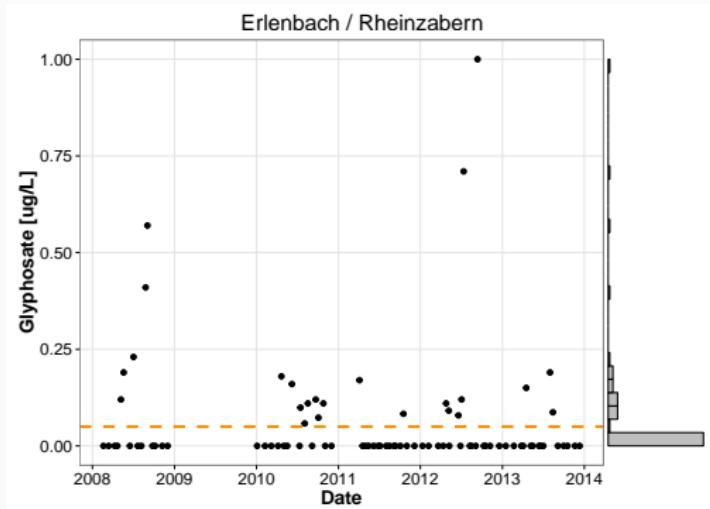
- factorial design
- one predictor
- low replicated
- LM, GLM, bootstrap
- High T1 error of NB
- Quasi-Poisson worked well
- Bootstrap fixes the problems

Ives (2015)

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

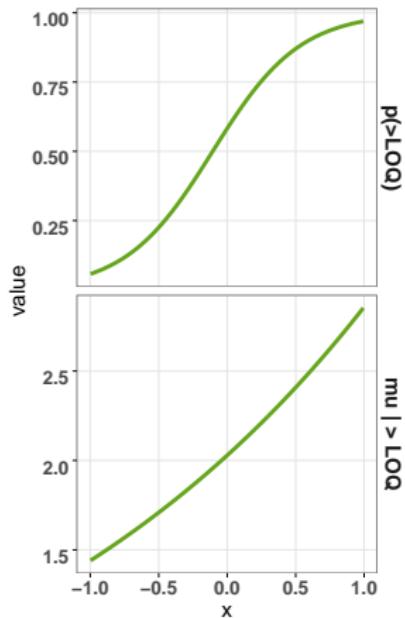
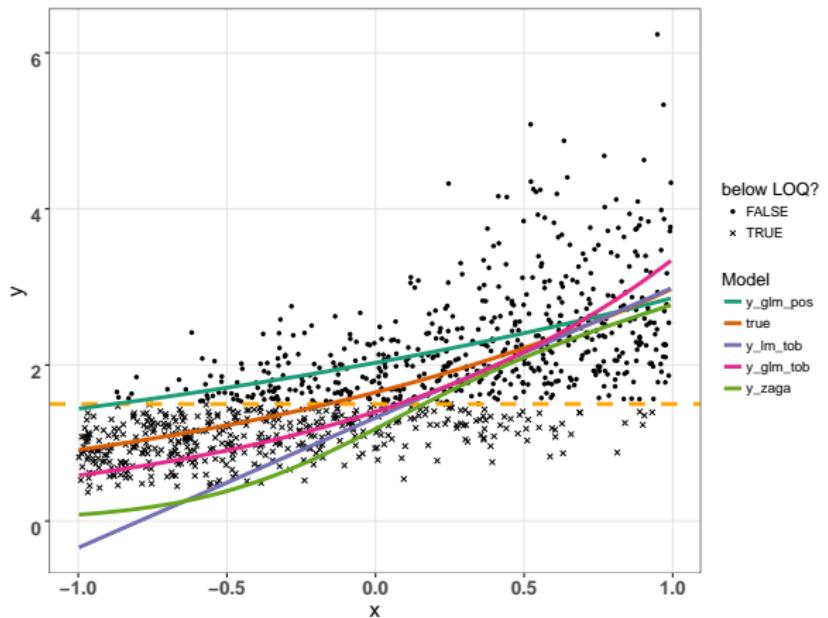
Idiosyncrasies of chemical concentrations

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

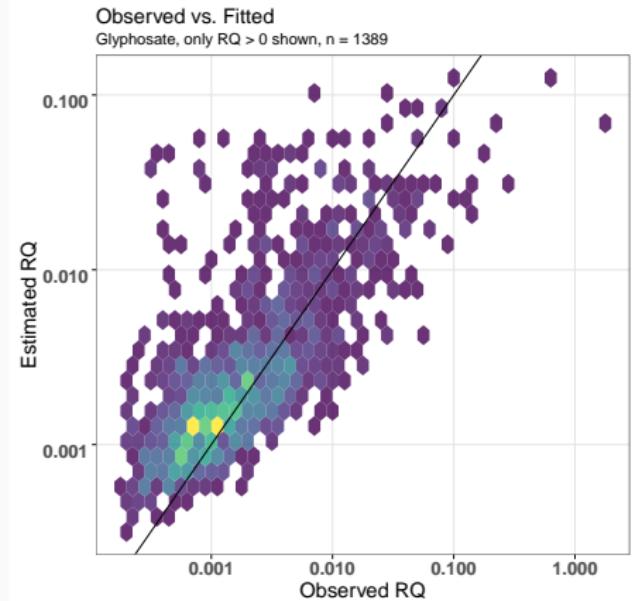
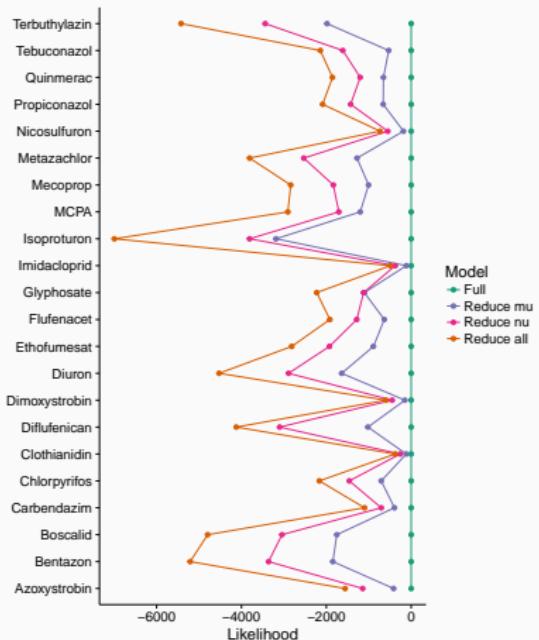


ZAGA what...?

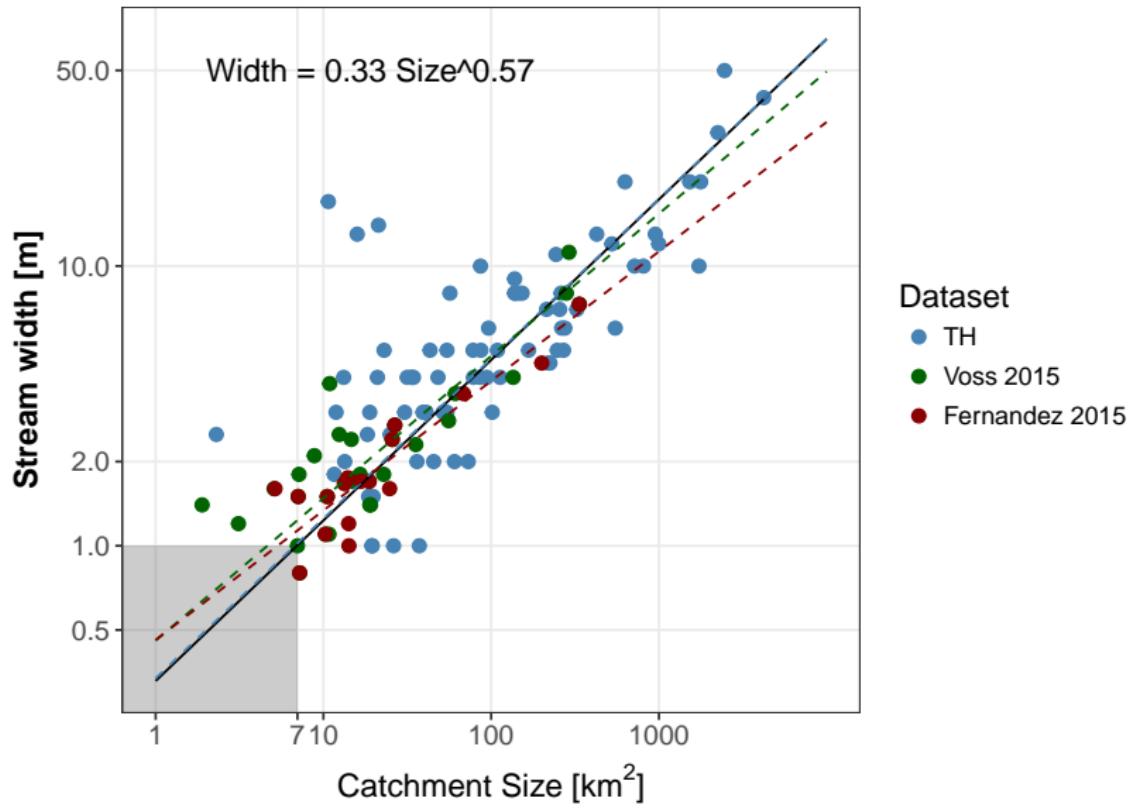
shiny app: <http://uni-ko-ld.de/g4>



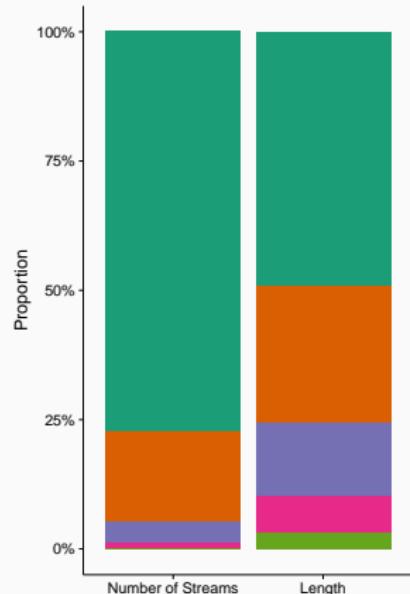
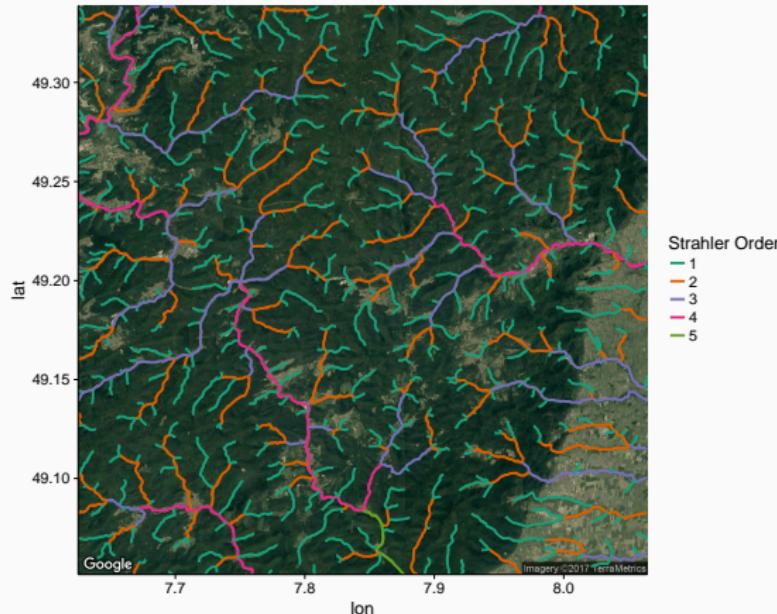
Model Performance



Stream size - width relationship



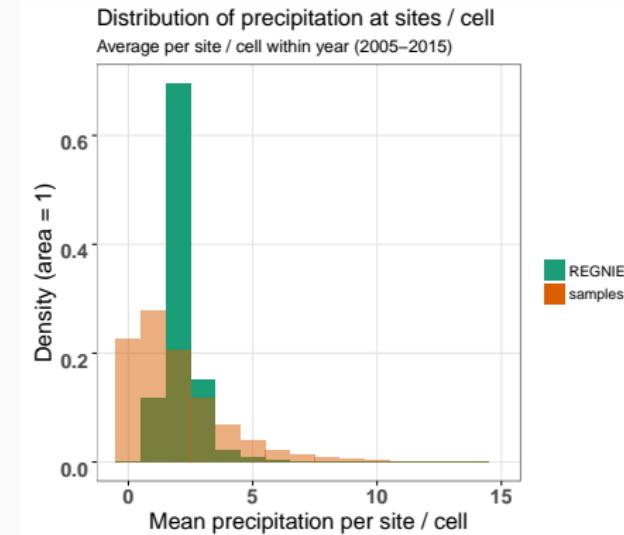
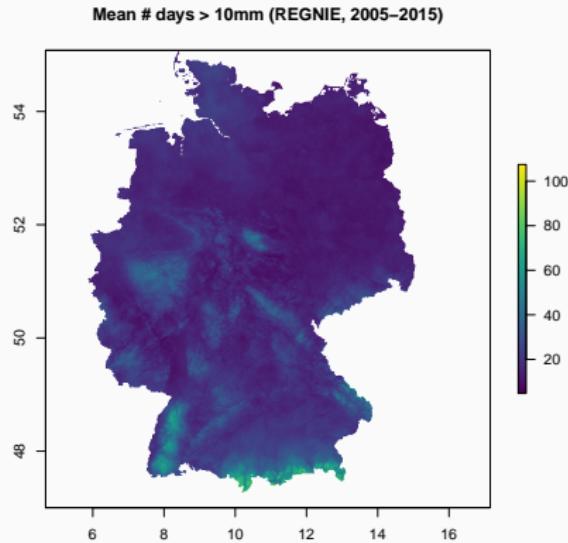
Small streams in the landscape



- 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)

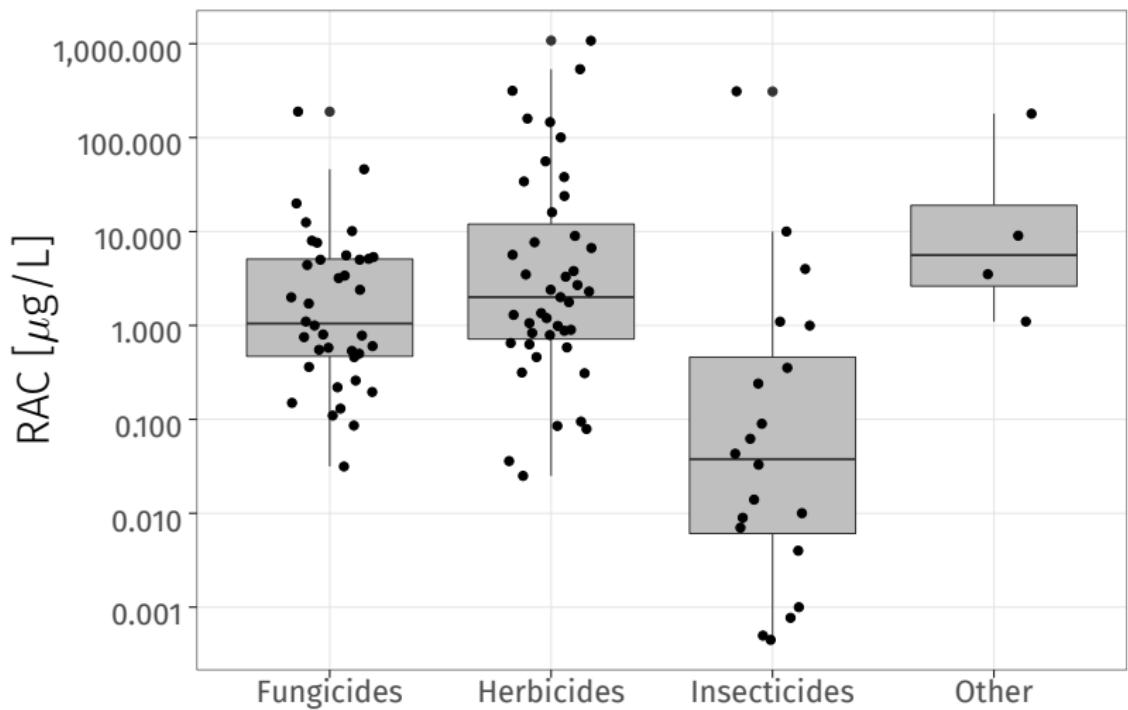
- Europe / Global
- Publications
- Grab & Event driven sampling
- Insecticides
- Organophos.+ Pyrethroids
- LM for >LOQ

Knauer (2016)

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

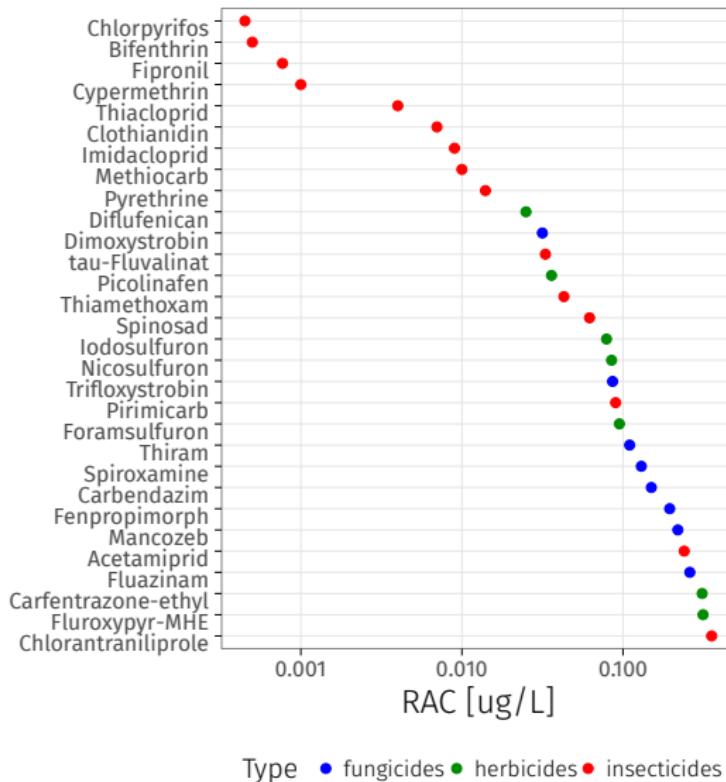
RACs by Type

105 RACs provided by UBA splitted by group

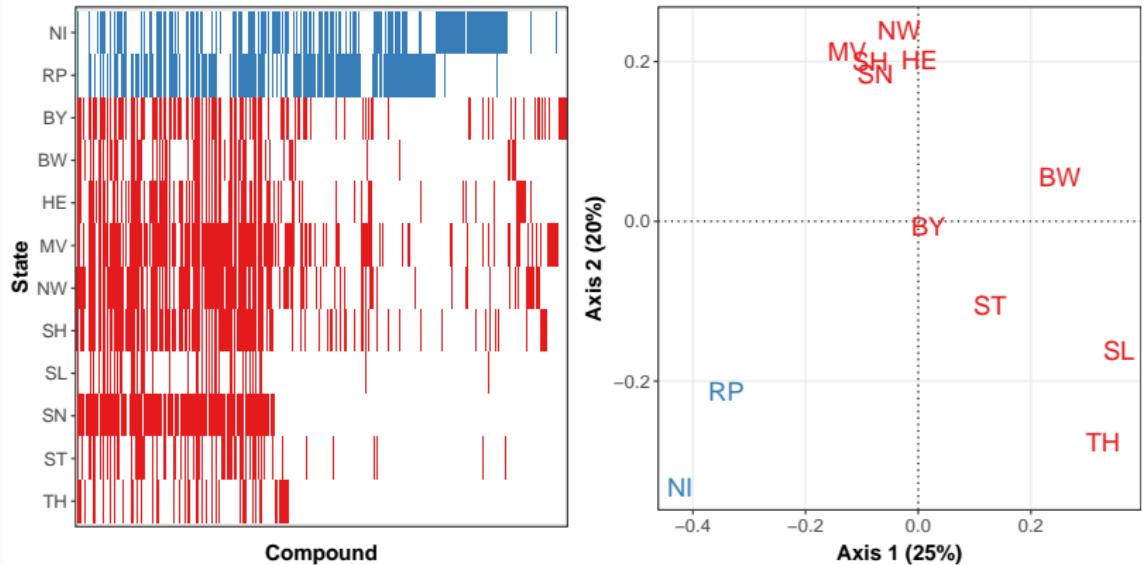


RACs by Compound

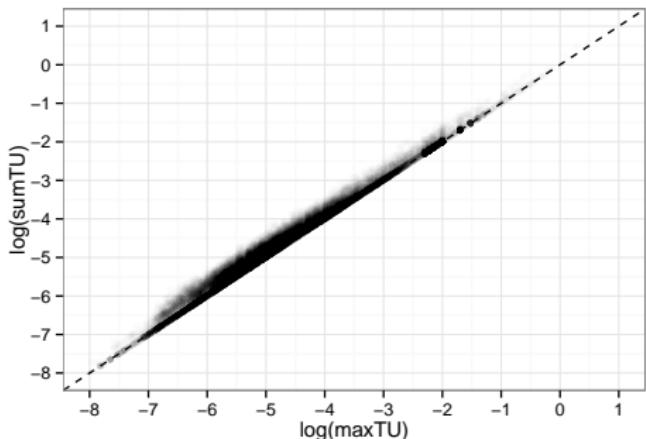
30 lowest RACs



Analysed compound spectra by state



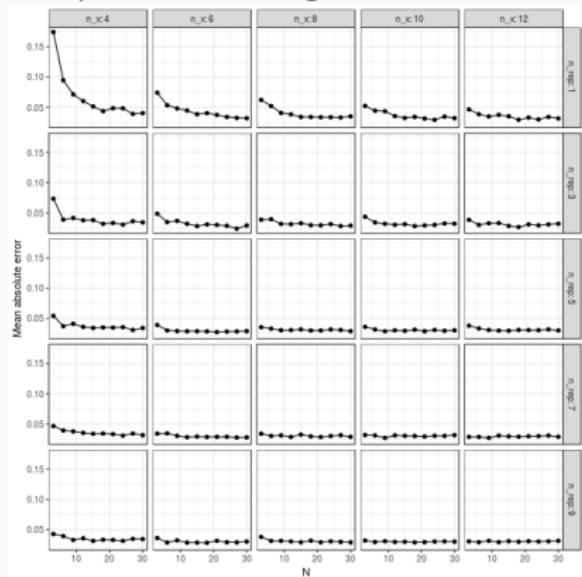
Mixtures are common, but one compound dominates the risk



- 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*!

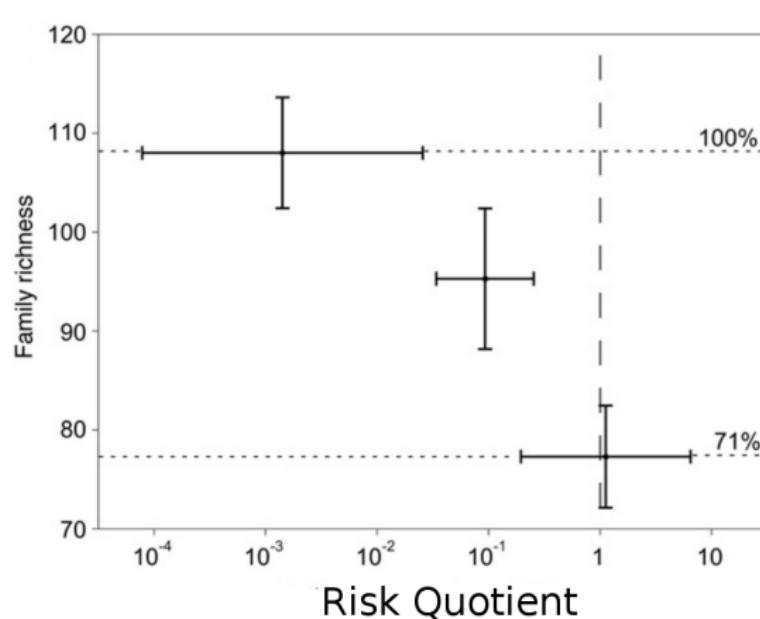
Experimental design for dose-response experiments - a simulation
http://edild.github.io/lc50_bias_sim/



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

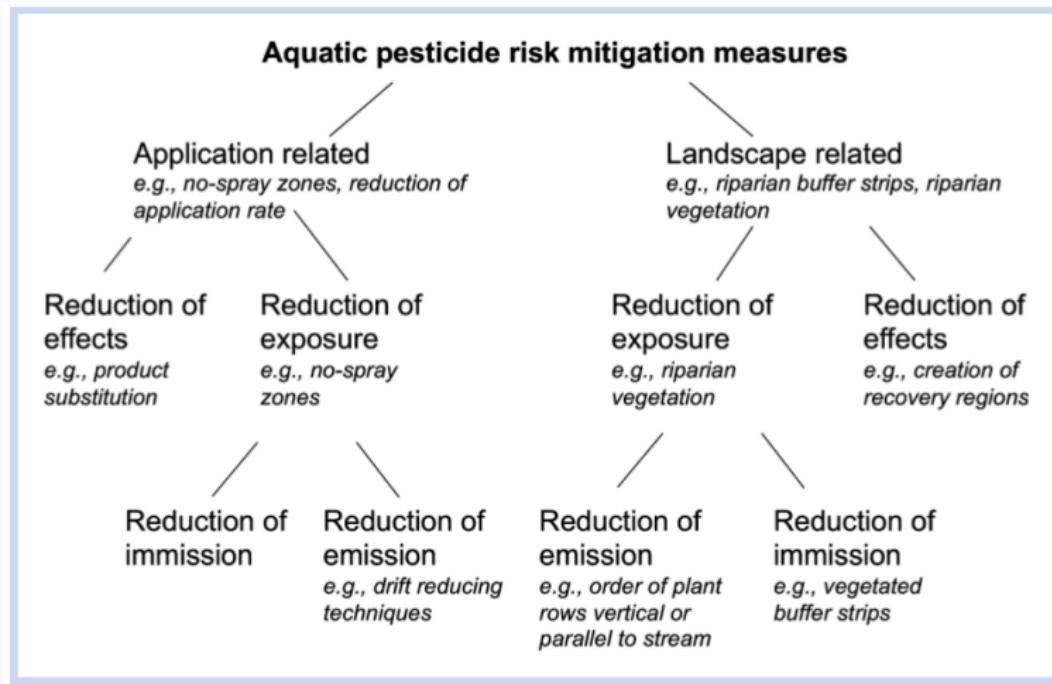
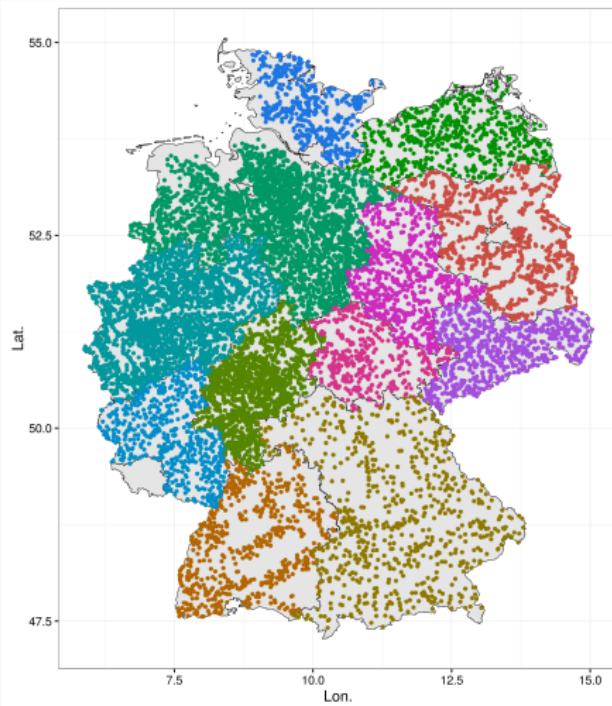


Figure 1. Classification of pesticide risk mitigation measures.

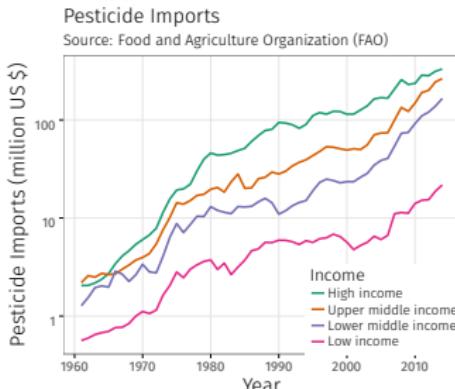
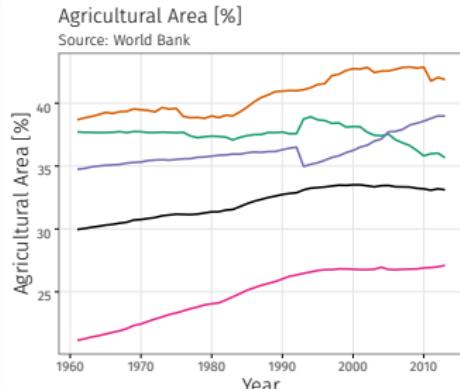
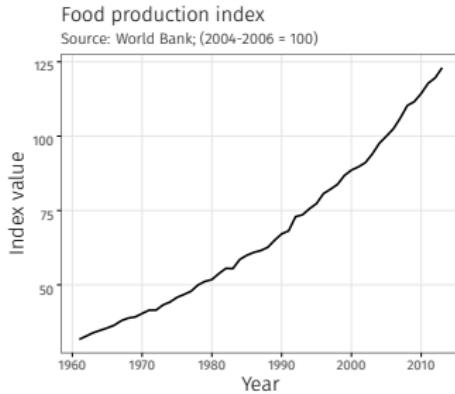
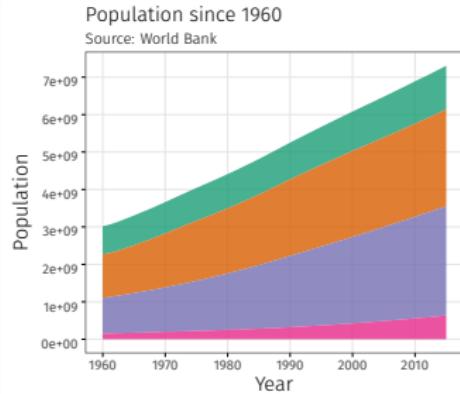
Bereswill, Strelcok, Schulz (2014). Risk mitigation measures for diffuse pesticide entry into aquatic ecosystems: Proposal of a guide to identify appropriate measures on a catchment scale: Guide to Identify Pesticide Risk Mitigation Measures. *IEAM* 10, 286–298.

Biotic field effects



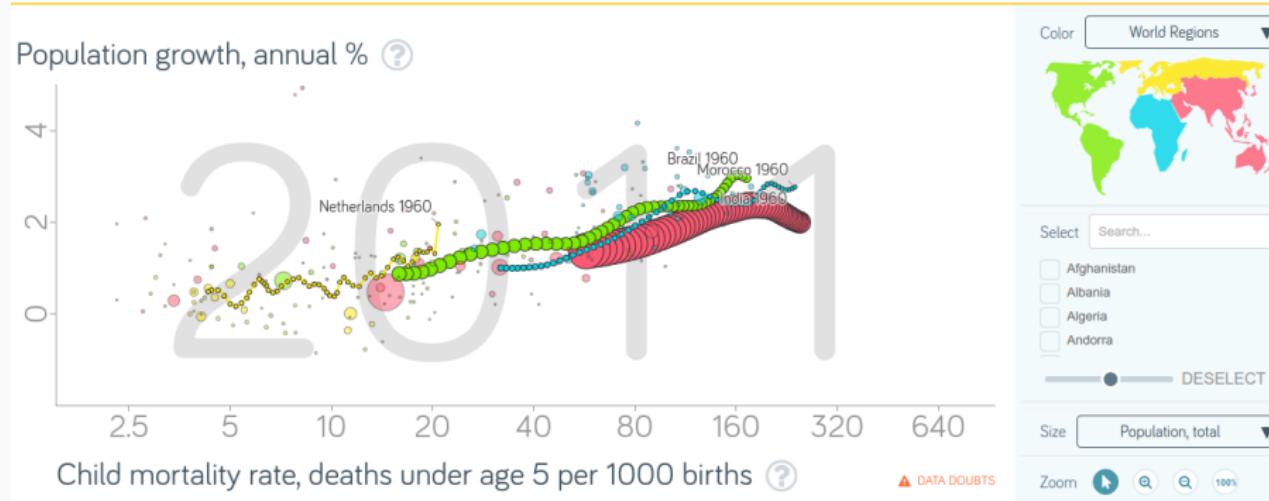
- biological data with good spatial coverage
- 60% of spatial congruence
- Large scale effects largely unknown.
- Some work left...
- Future....

A global perspective (I)



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

A global perspective (II)



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)

Many Thanks To

- My supervisor Prof. Dr. Ralf. B. Schäfer (for 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)