Statistical Eco(-toxico)logy

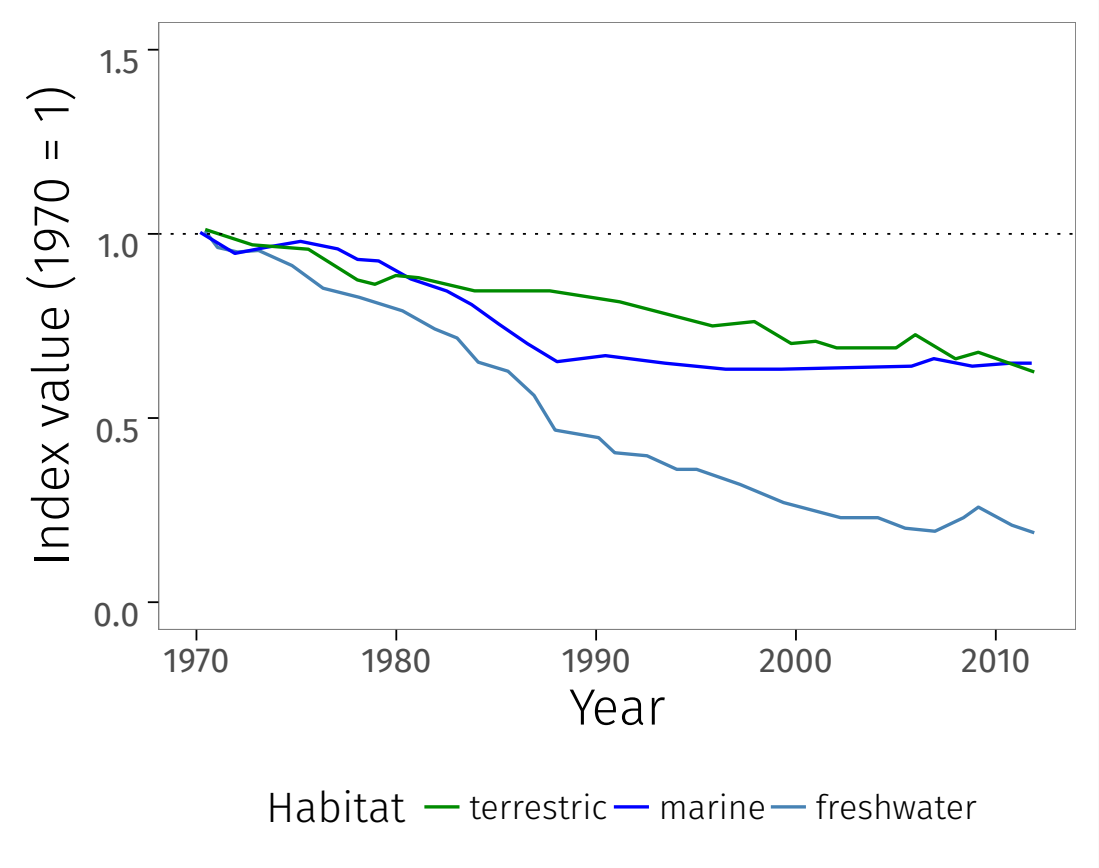
Improving the Utilisation of Data for Environmental Risk Assessment

Eduard Szöcs

25th January 2017

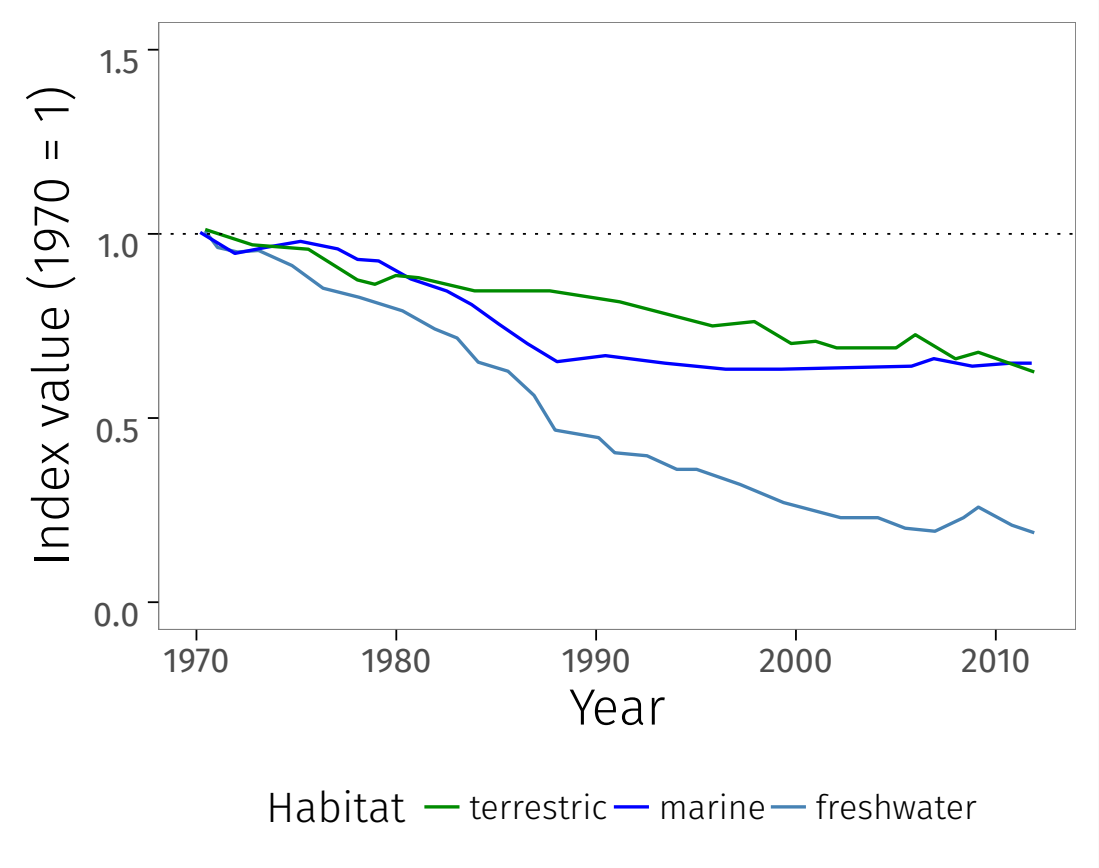
Table of contents

Freshwater biodiversity is strongly declining



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

Freshwater biodiversity is strongly declining

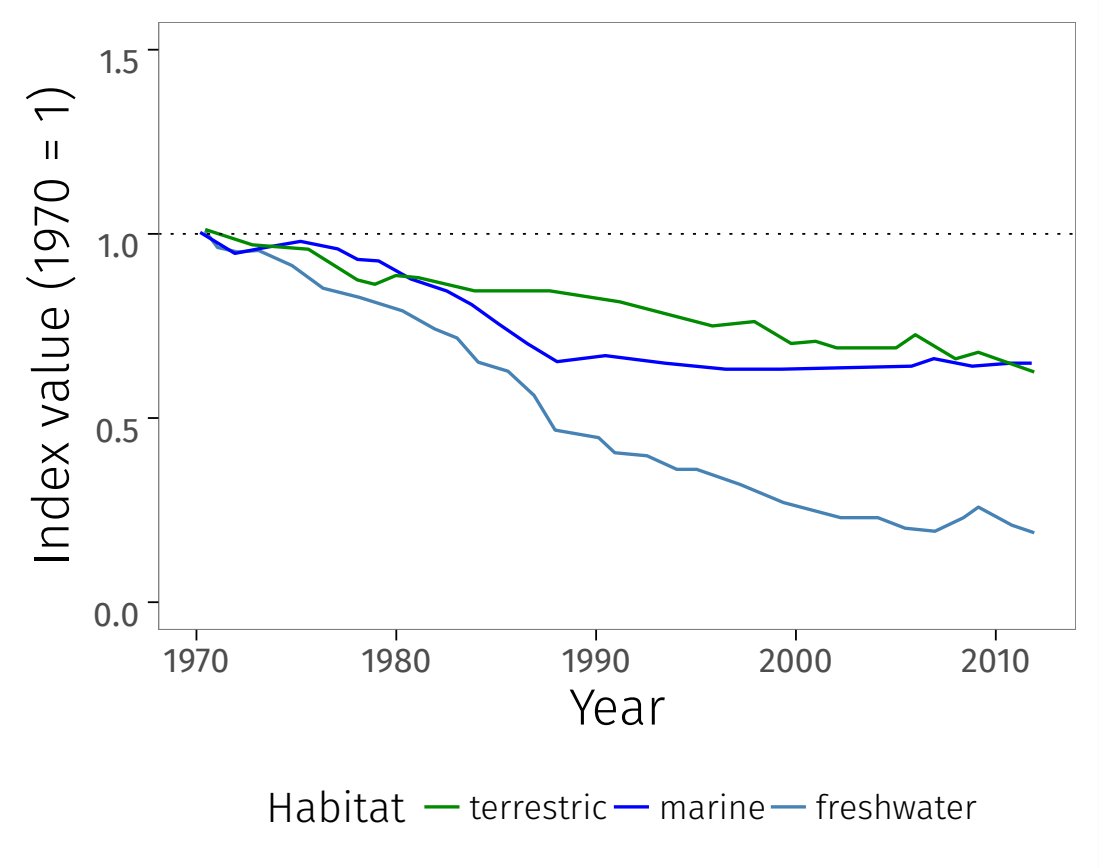


Reasons

- 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
1107/2009

Water Framework Directive
2000/60/EC

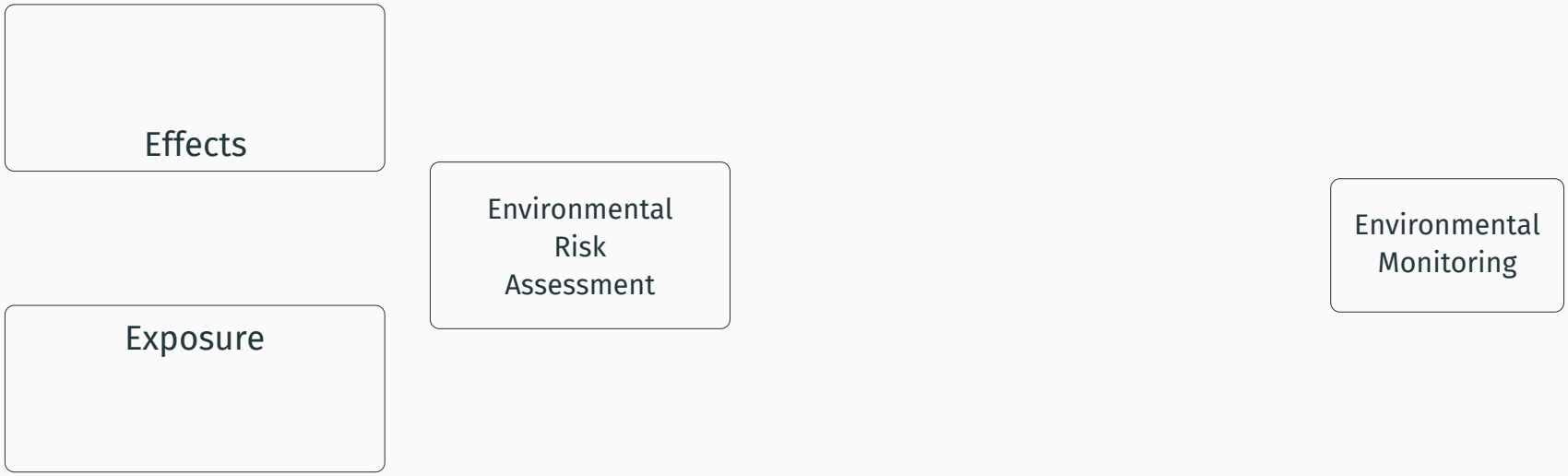
Environmental
Risk
Assessment

Environmental
Monitoring

Environmental Risk Assessment and Monitoring

Plant Protection Products
1107/2009

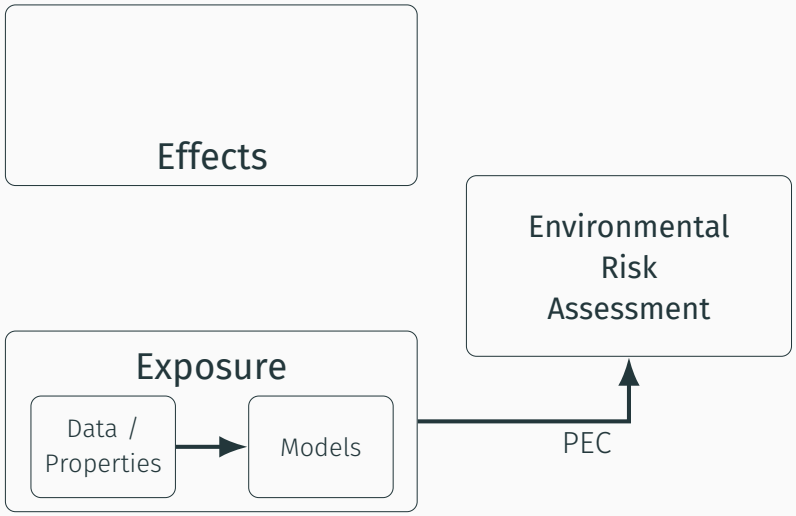
Water Framework Directive
2000/60/EC

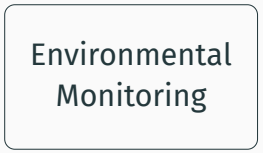


Environmental Risk Assessment and Monitoring

Plant Protection Products 1107/2009

Water Framework Directive 2000/60/EC

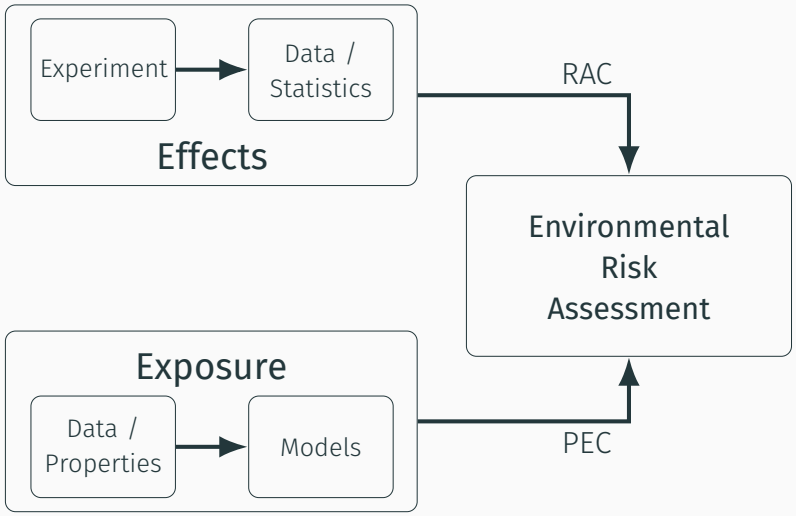




Environmental Risk Assessment and Monitoring

Plant Protection Products 1107/2009

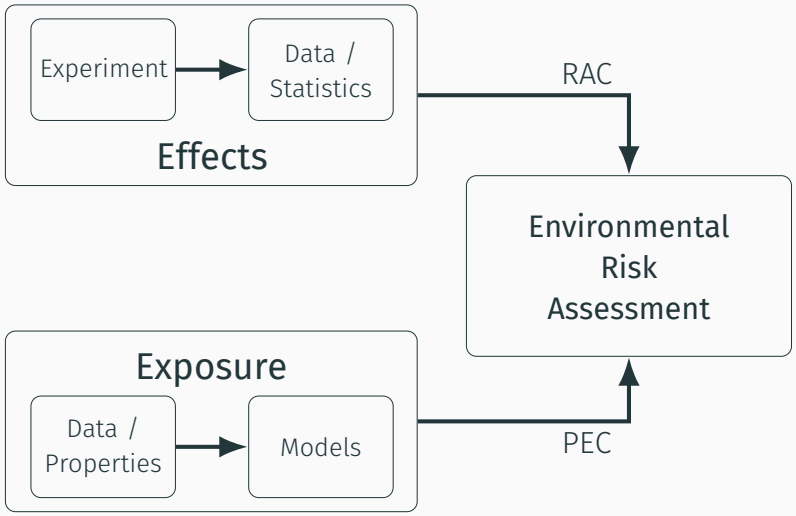
Water Framework Directive 2000/60/EC



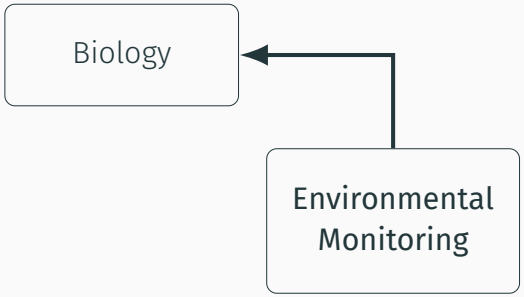
Environmental Monitoring

Environmental Risk Assessment and Monitoring

Plant Protection Products 1107/2009

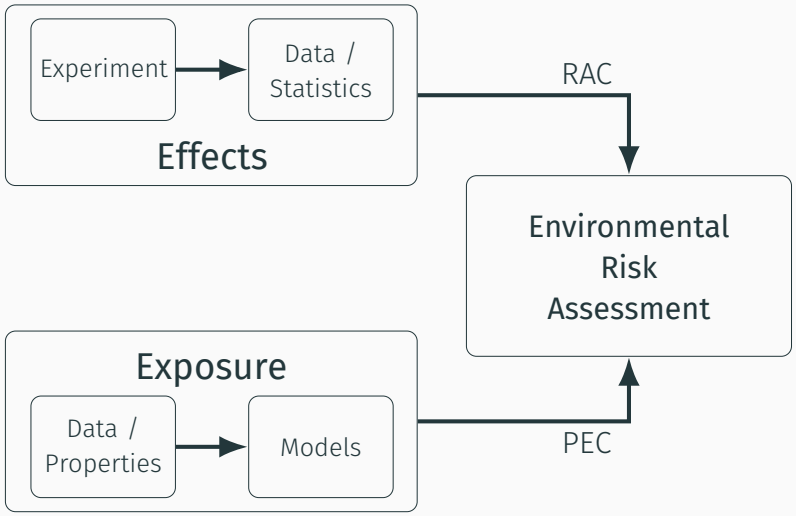


Water Framework Directive 2000/60/EC

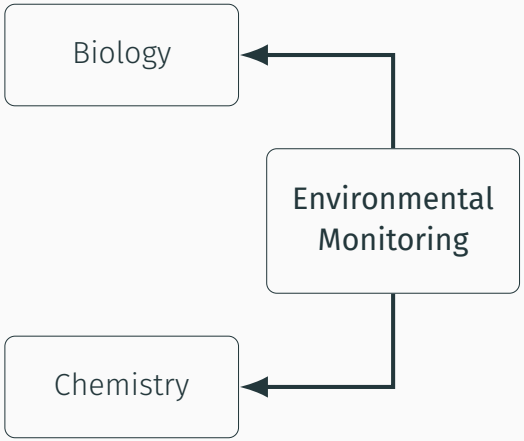


Environmental Risk Assessment and Monitoring

Plant Protection Products 1107/2009



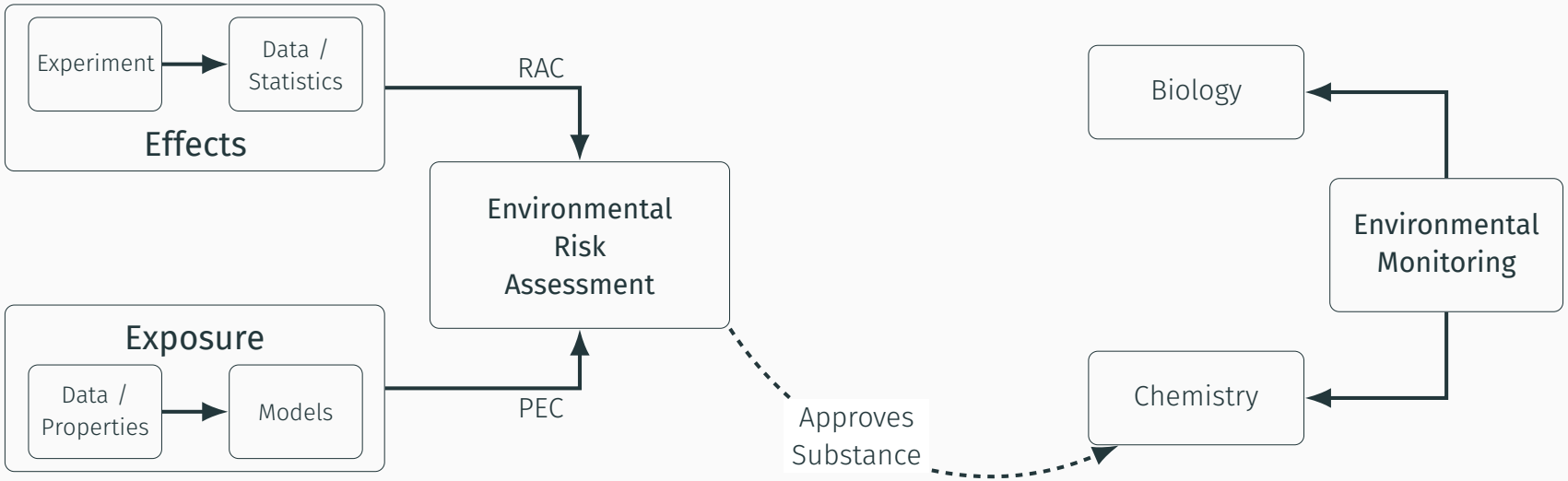
Water Framework Directive 2000/60/EC



Environmental Risk Assessment and Monitoring

Plant Protection Products 1107/2009

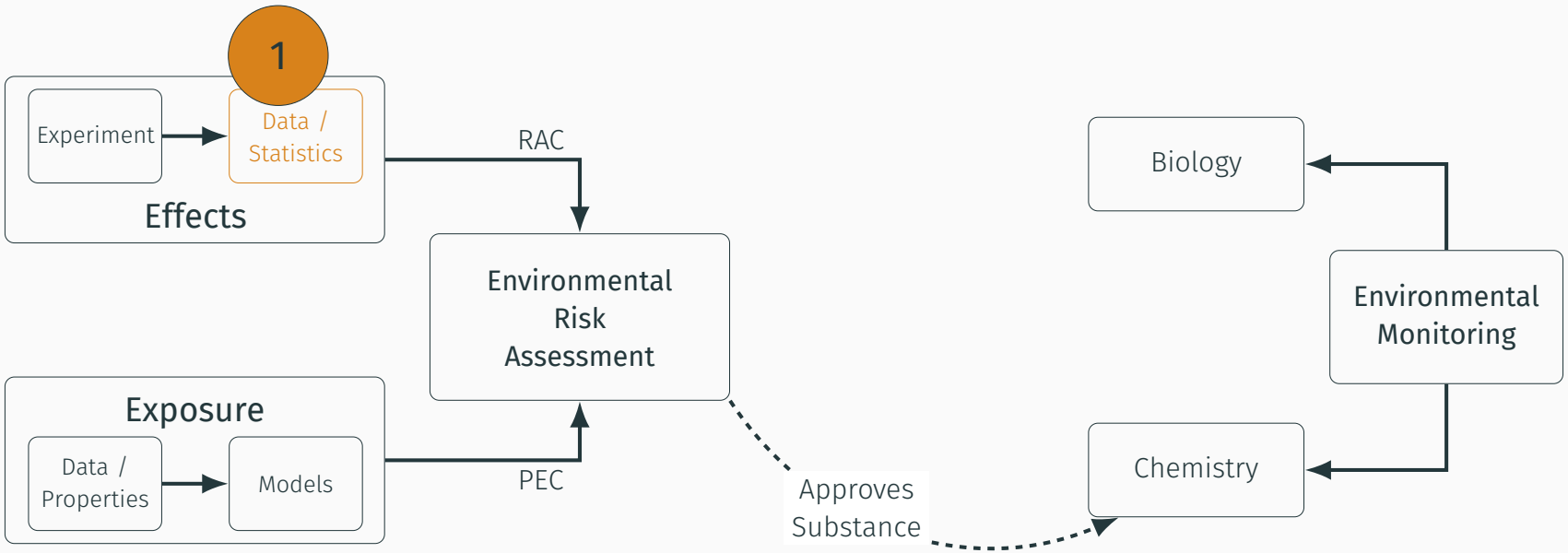
Water Framework Directive 2000/60/EC



Environmental Risk Assessment and Monitoring

Plant Protection Products 1107/2009

Water Framework Directive 2000/60/EC

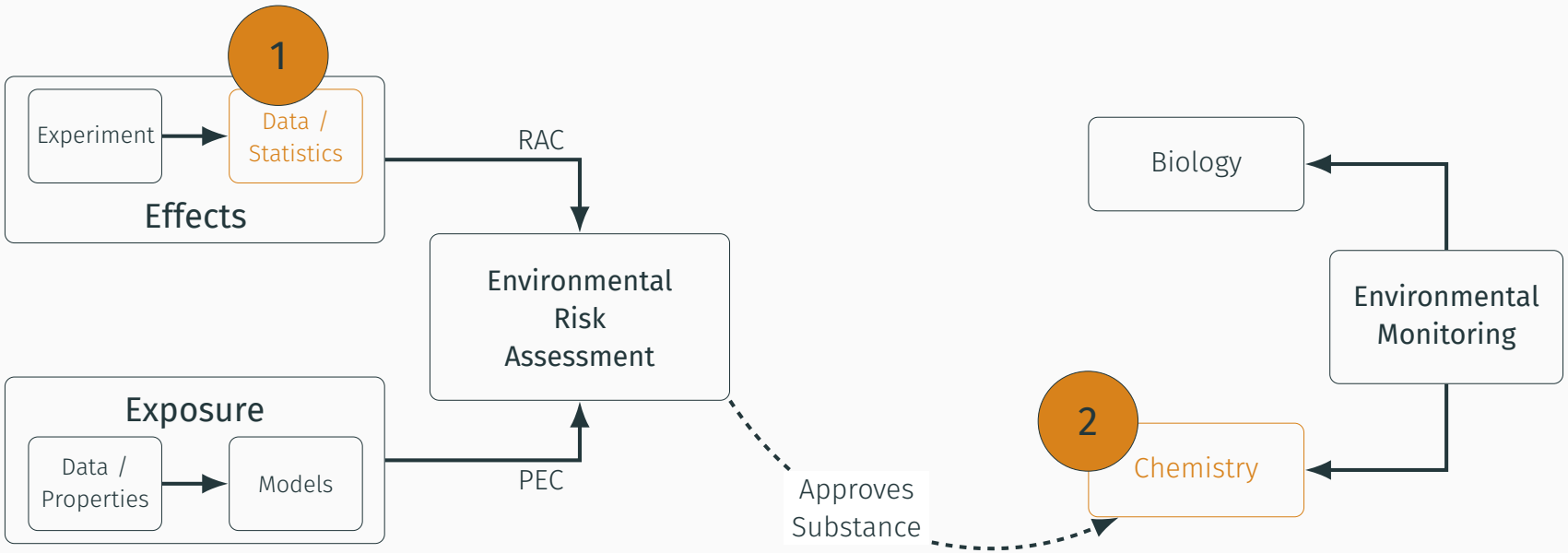


1 Szöcs & Schäfer (2015). “Ecotoxicology is not normal”. ESPR 22(18), 13990–13999.

Environmental Risk Assessment and Monitoring

Plant Protection Products 1107/2009

Water Framework Directive 2000/60/EC

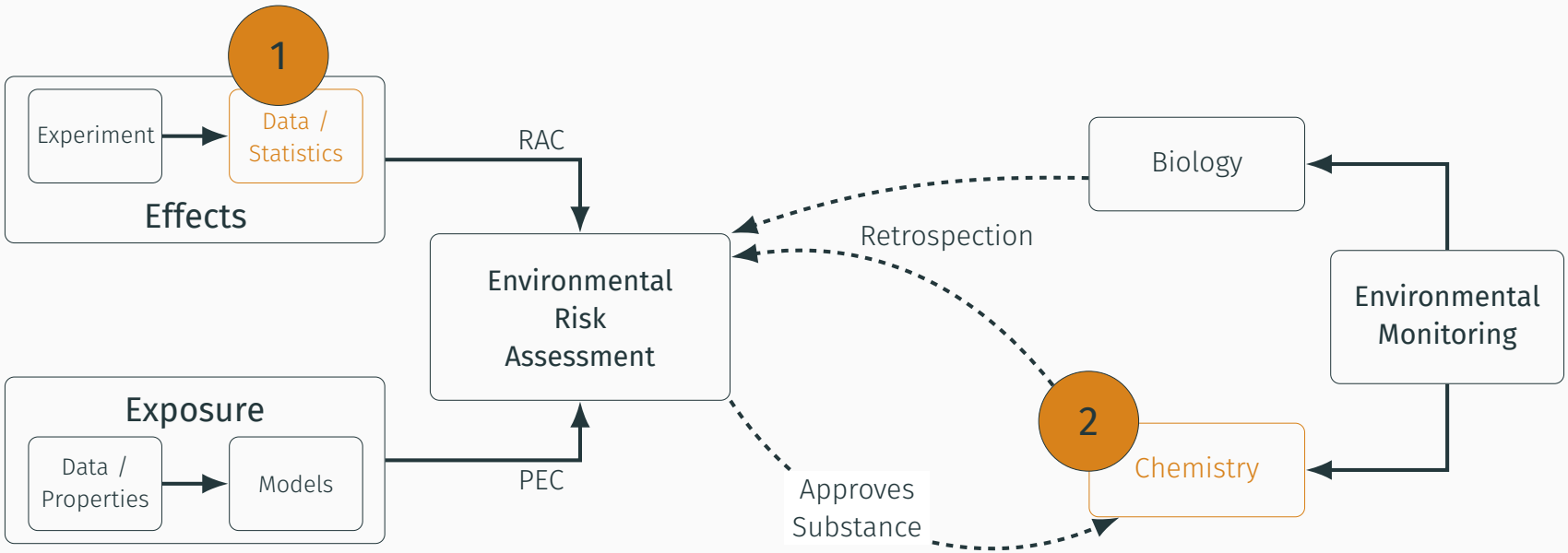


2 Szöcs, Brinke, Karaoglan & Schäfer (submitted). “Large scale risks from pesticides in small streams”. ES&T.

Environmental Risk Assessment and Monitoring

Plant Protection Products 1107/2009

Water Framework Directive 2000/60/EC

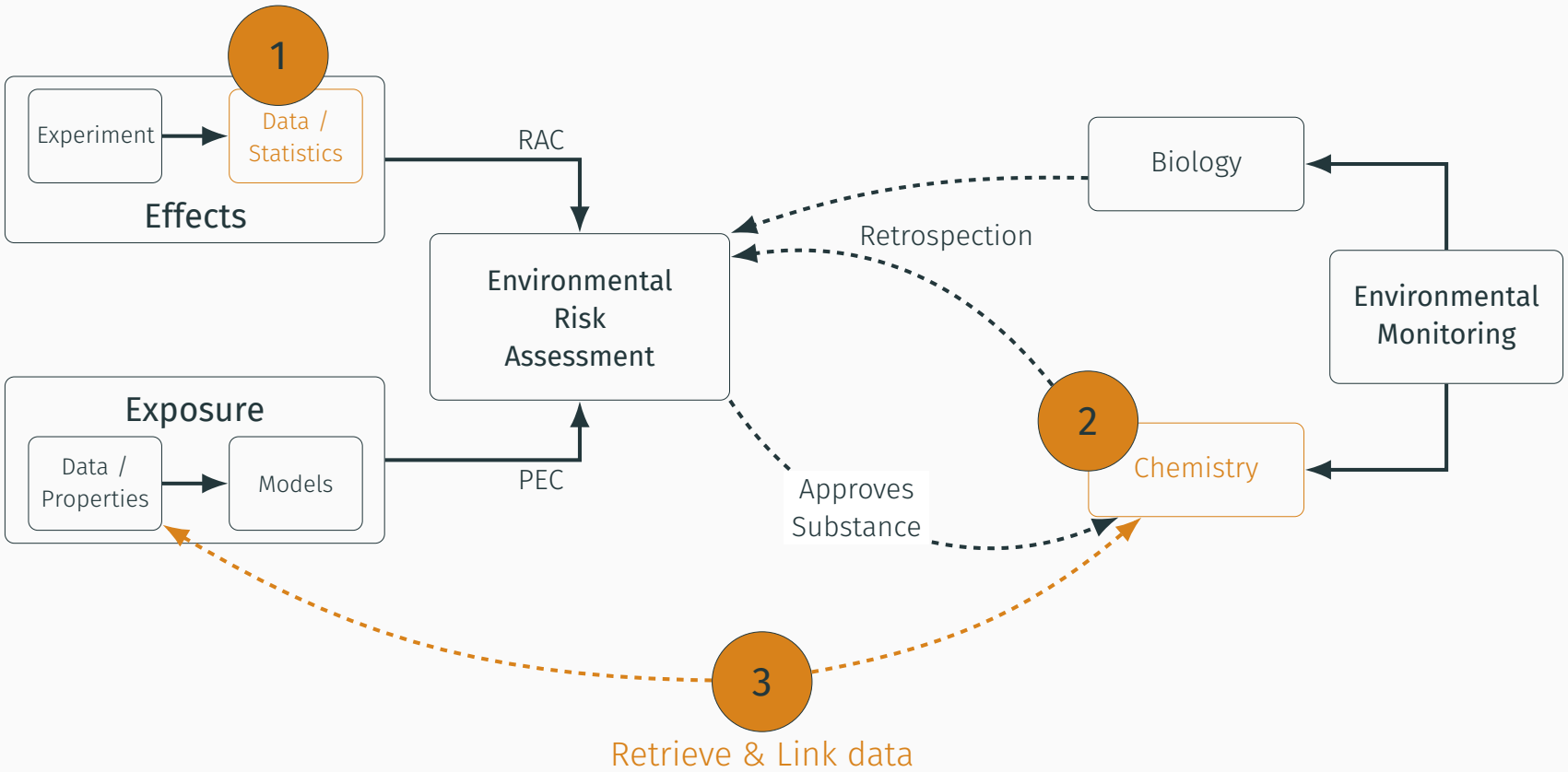


2 Szöcs, Brinke, Karaoglan & Schäfer (submitted). “Large scale risks from pesticides in small streams”. ES&T.

Environmental Risk Assessment and Monitoring

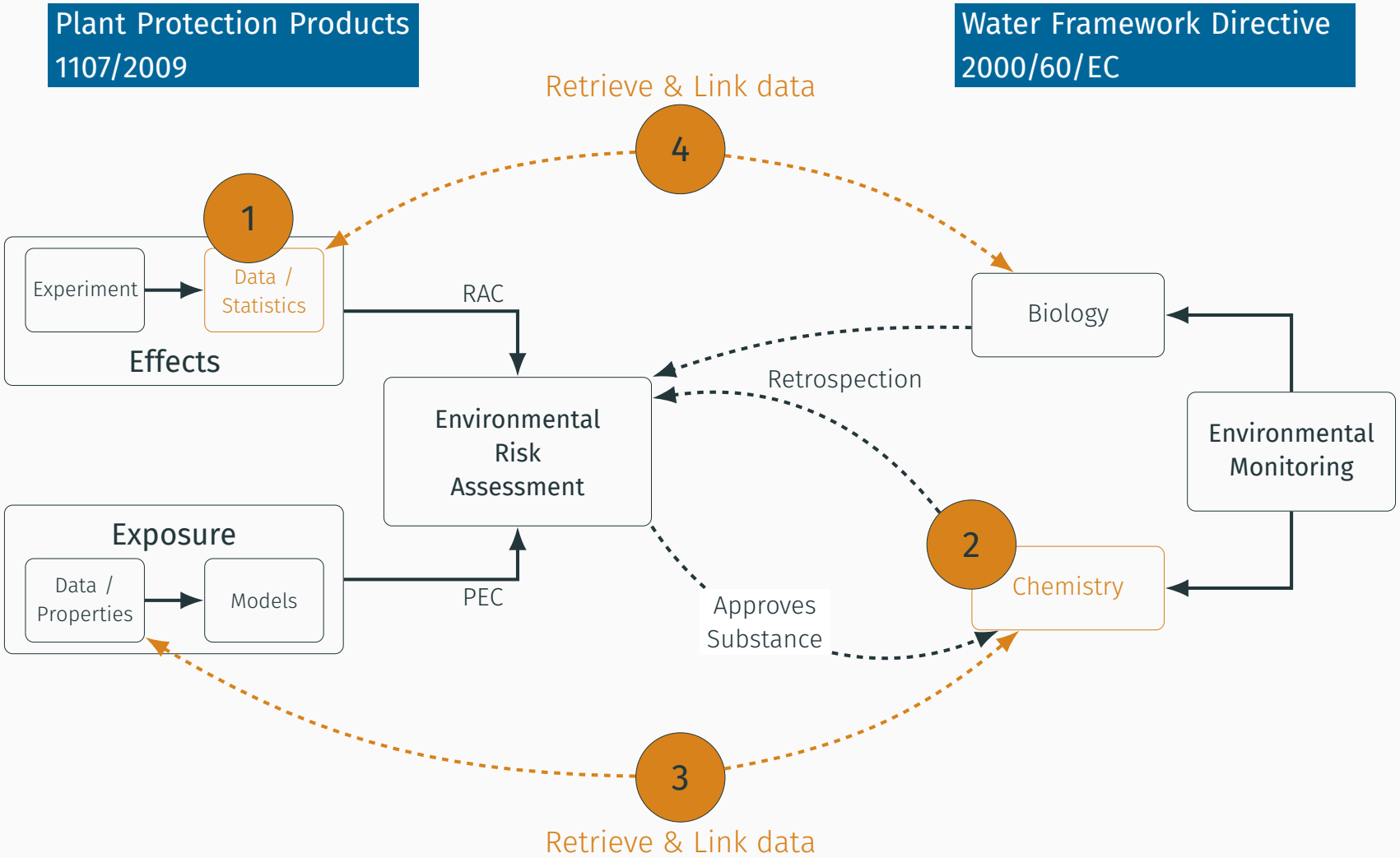
Plant Protection Products 1107/2009

Water Framework Directive 2000/60/EC



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

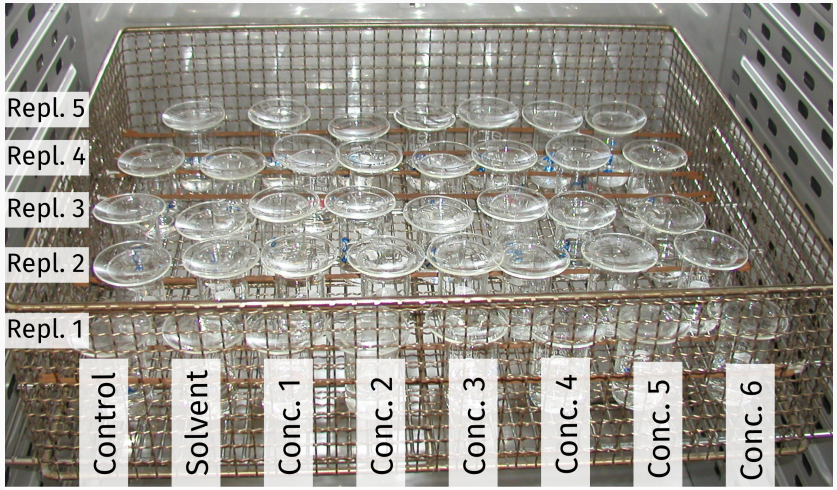
Environmental Risk Assessment and Monitoring



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

Improving Statistics in ERA

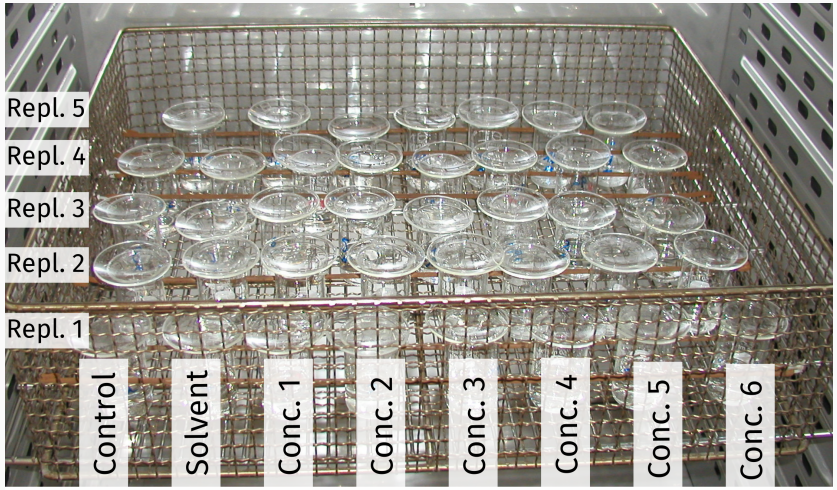
Experiments in Effect Assessment



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

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

Experiments in Effect Assessment

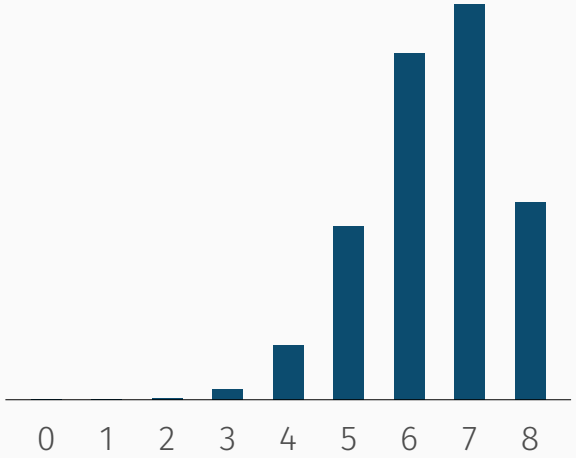


- Daphnia Test
- Lower Tier
- *”x out of n survived”*

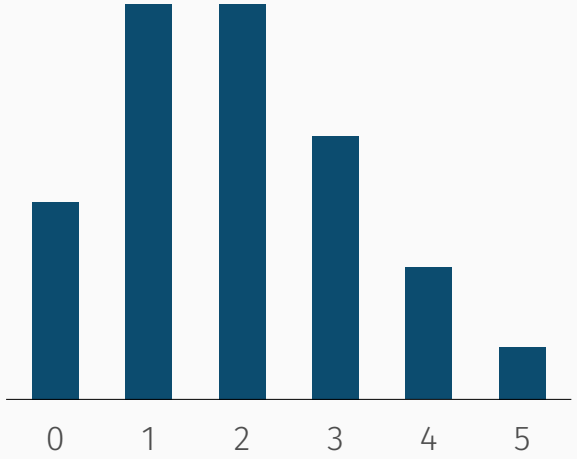
- Mesocosm
- Higher Tier
- *”number of animals”*

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

Ecotoxicology is not normal

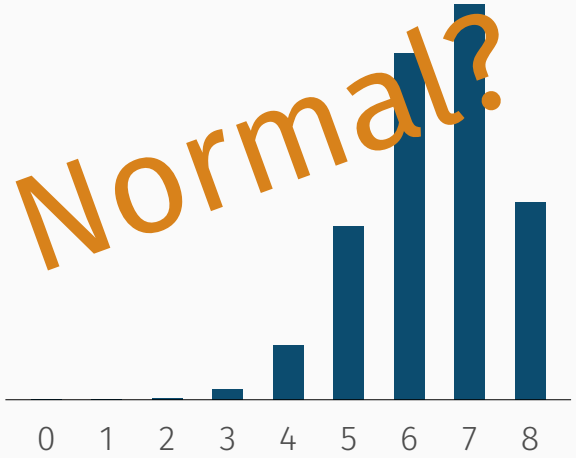


- *”x out of n survived”*

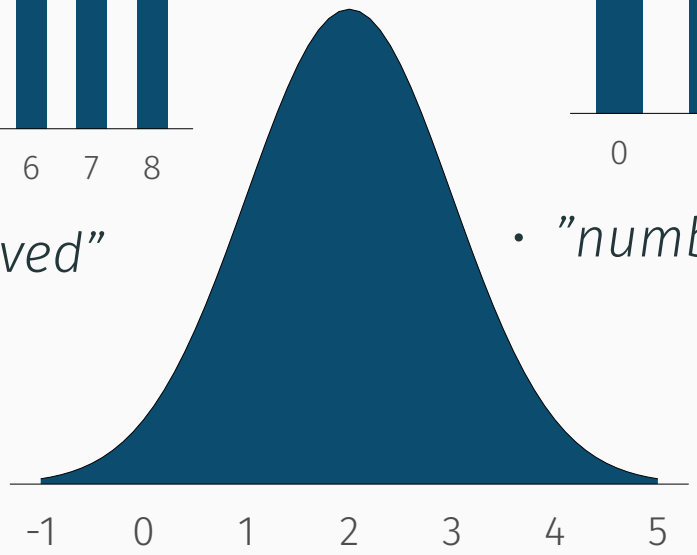


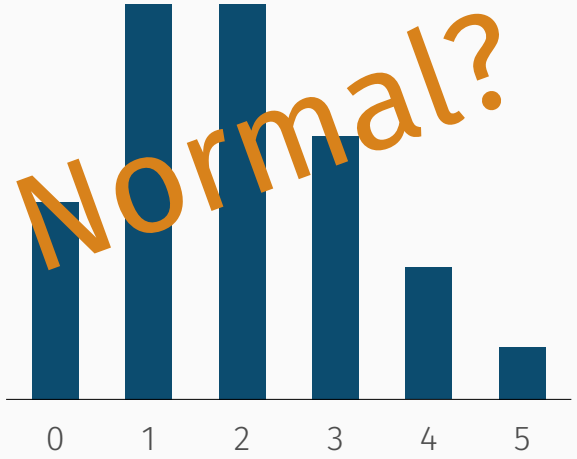
- *”number of animals”*

Ecotoxicology is not normal



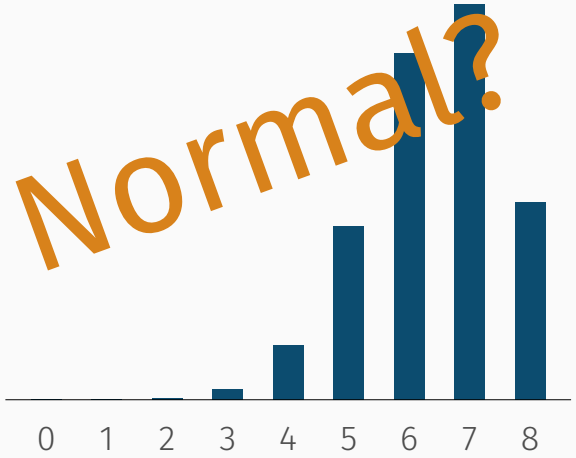
- *"x out of n survived"*



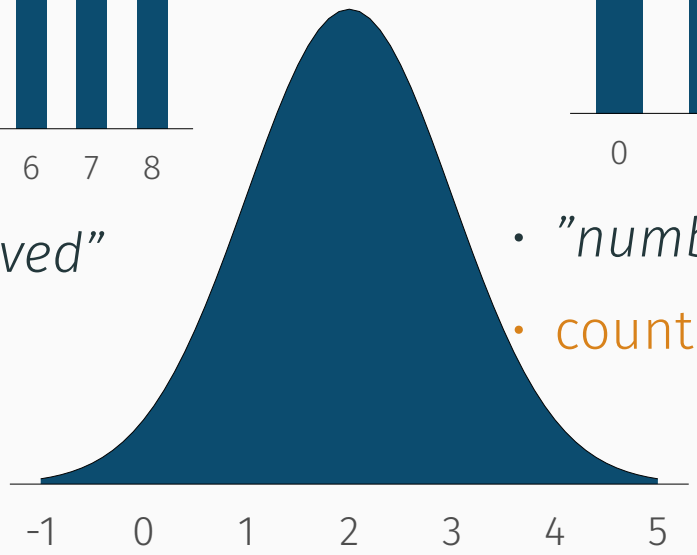


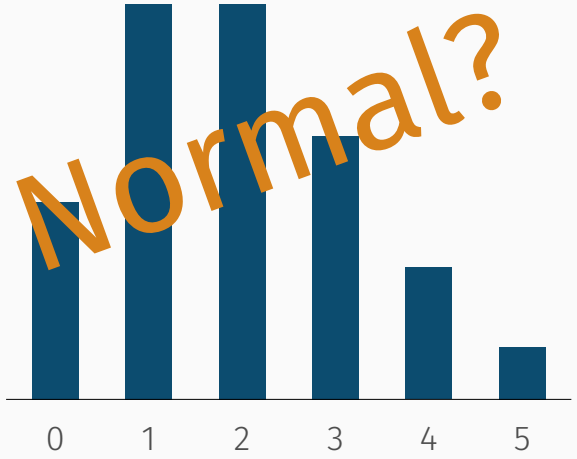
- *"number of animals"*

Ecotoxicology is not normal



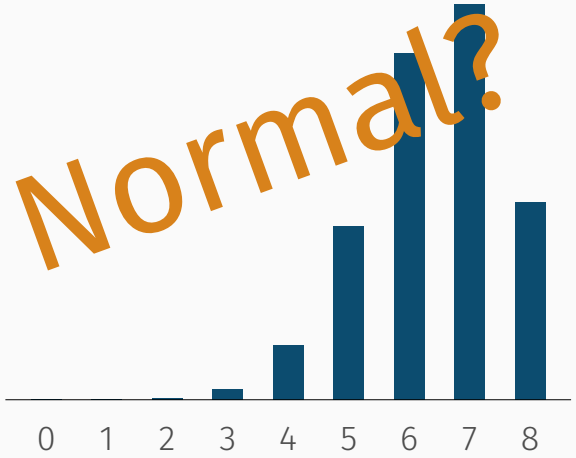
- *"x out of n survived"*
- binomial data



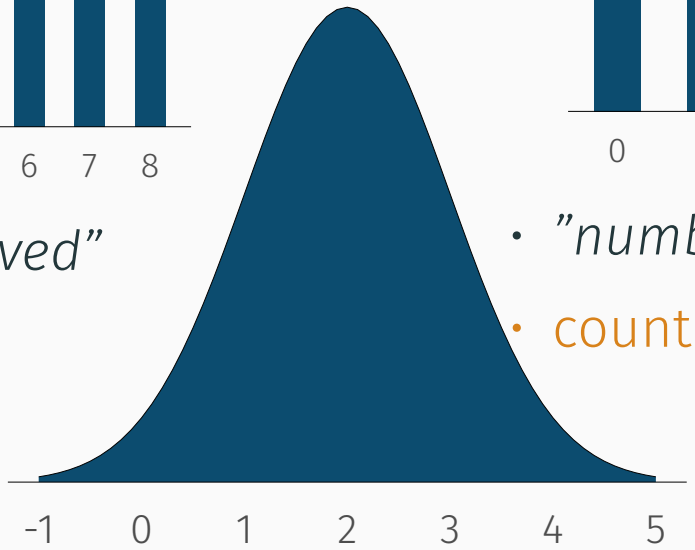


- *"number of animals"*
- count data

Ecotoxicology is not normal



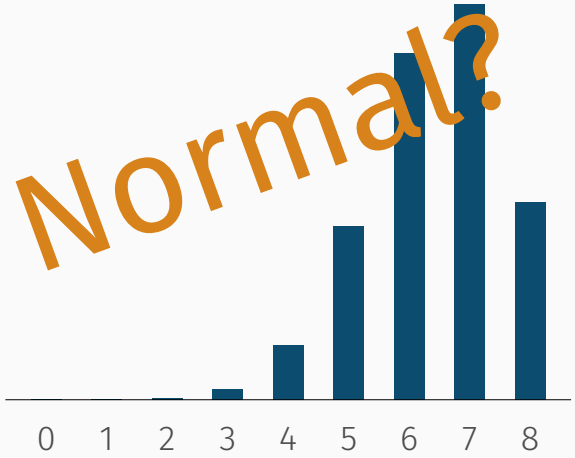
- *"x out of n survived"*
- binomial data



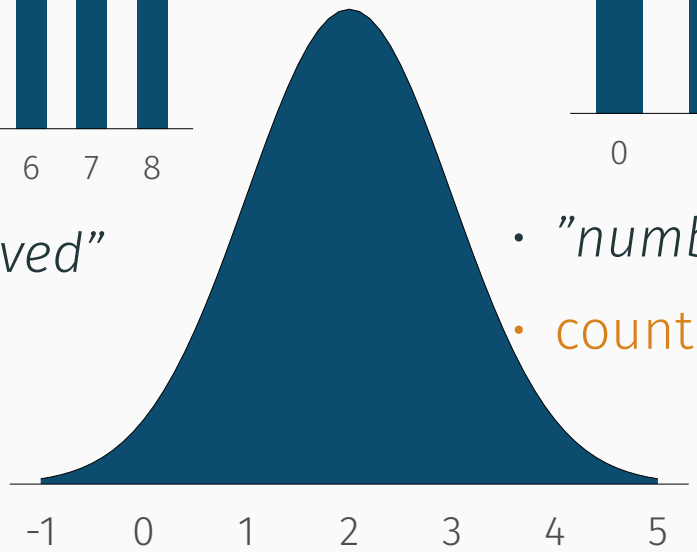
- *"number of animals"*
- count data

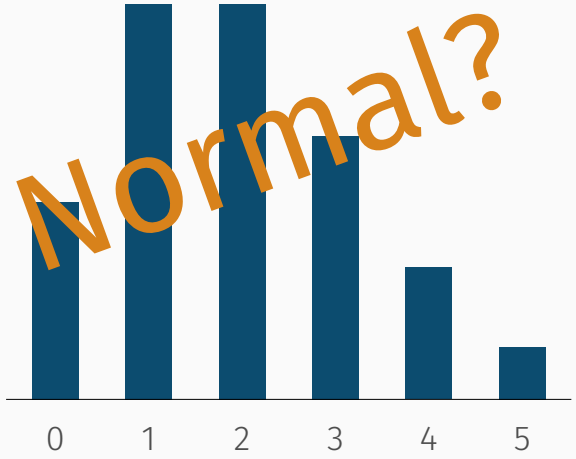
- ignore?

Ecotoxicology is not normal



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

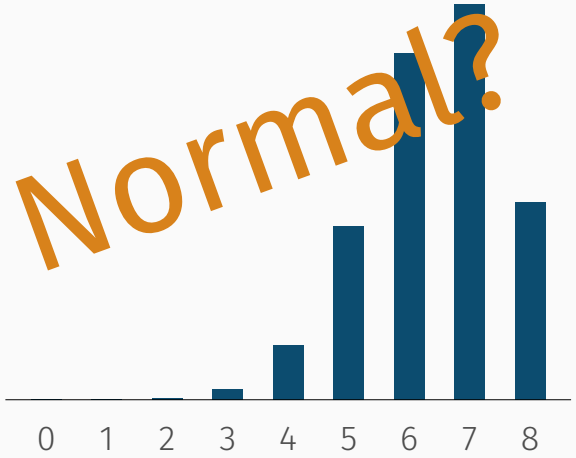




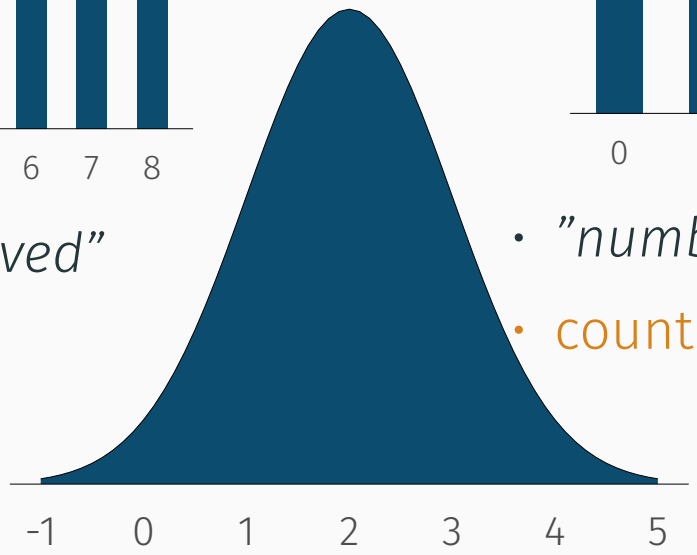
- *"number of animals"*
- count data

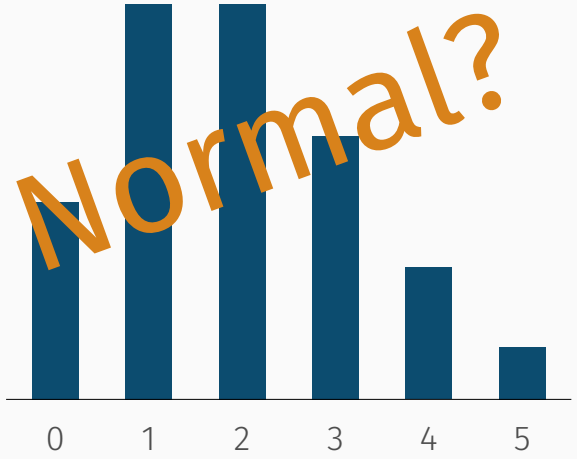
- ignore?
- transform?

Ecotoxicology is not normal



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

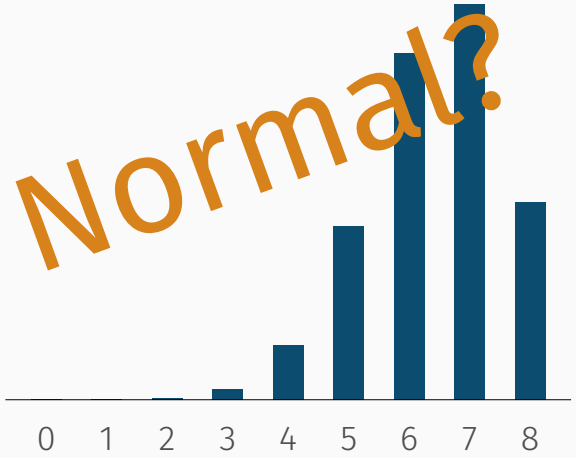




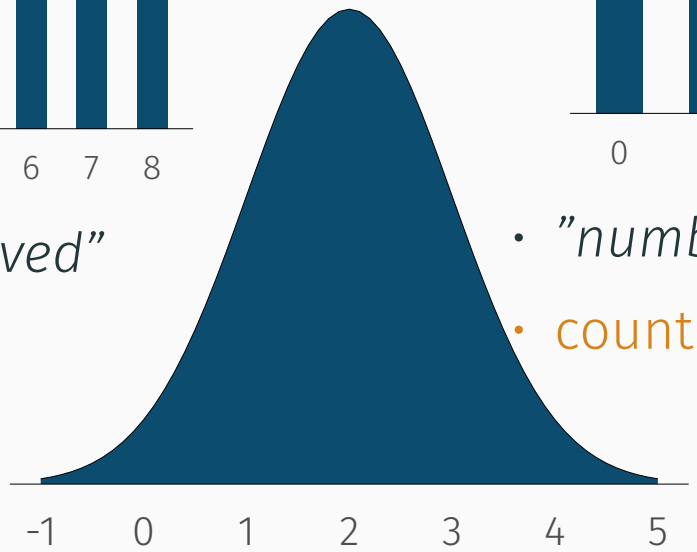
- *"number of animals"*
- count data

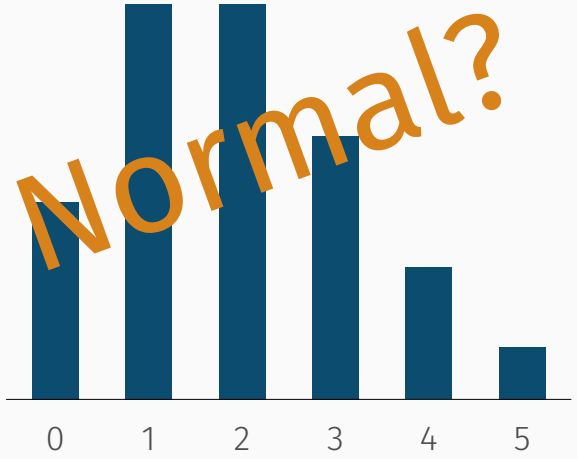
- ignore?
- transform?
- non-parametric?

Ecotoxicology is not normal



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

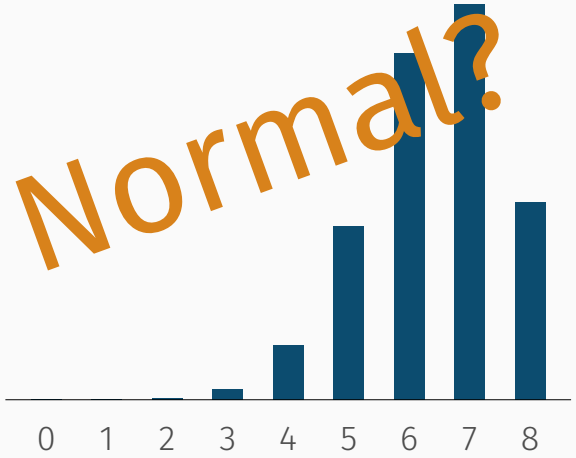




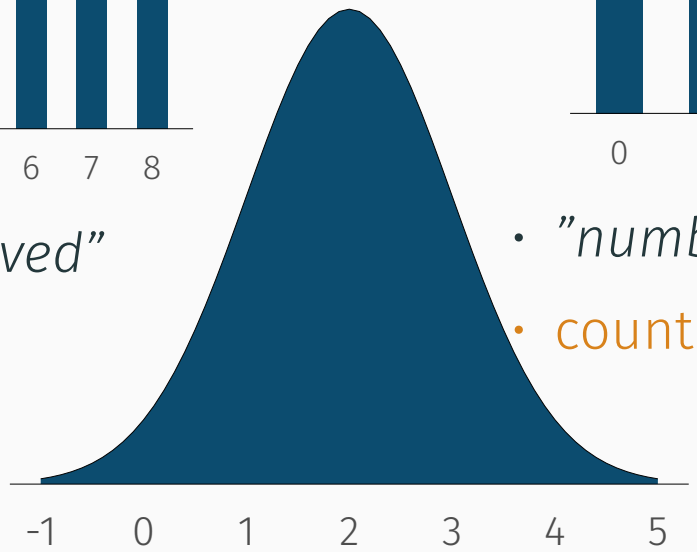
- *"number of animals"*
- count data

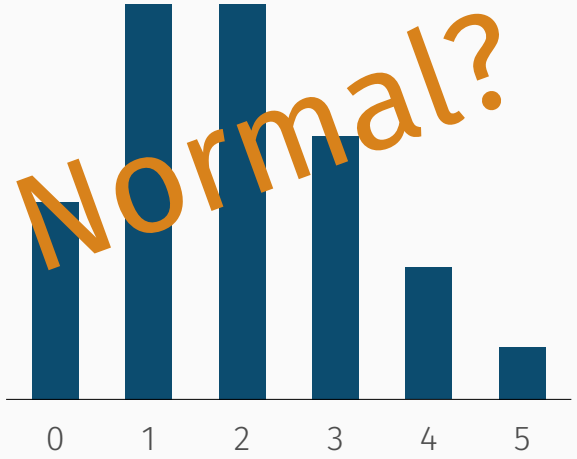
- ignore?
- transform?
- non-parametric?
- Generalized Linear Model (GLM)

Ecotoxicology is not normal



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





- *"number of animals"*
- count data

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

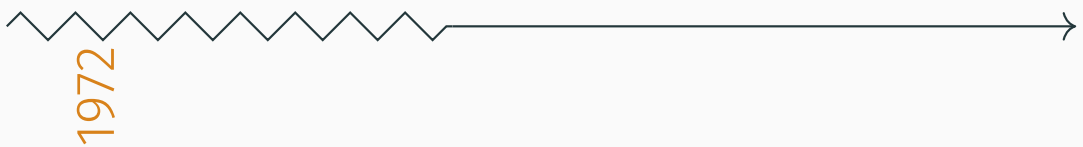
A recent history of GLM (uncomprehensive) in ecology

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

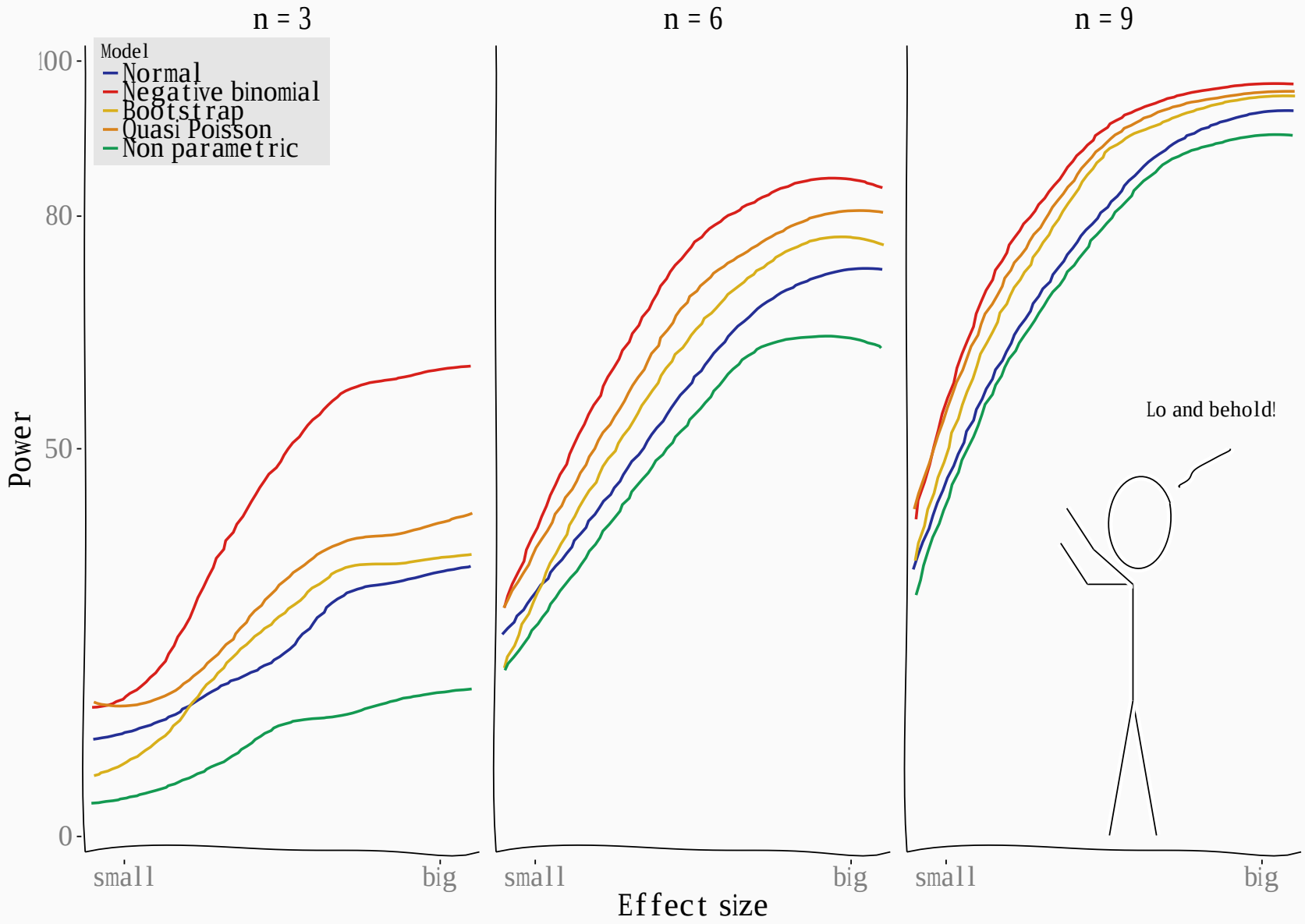
A comparison of statistical approaches for analysis of count and proportion data in ecotoxicology

Eduard Szöcs[1] · Ralf B. Schäfer[1]

A simulation study

GLMs can fail

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. Power in ecotoxicological experiments generally low
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

Environmental Monitoring

Overview on data compiled

Thresholds

Statistics with chemical measurements

Dynamics

Risks

What we learned

Solutions for Linking Data in ERA

Biologists & Chemists face the same problems

Names

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

Chlorpyrifos, Chlorpyriphos, Chlorphyrifos, Chlorpyrifos-ethyl, Chlorpypifot

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

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}

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,...

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

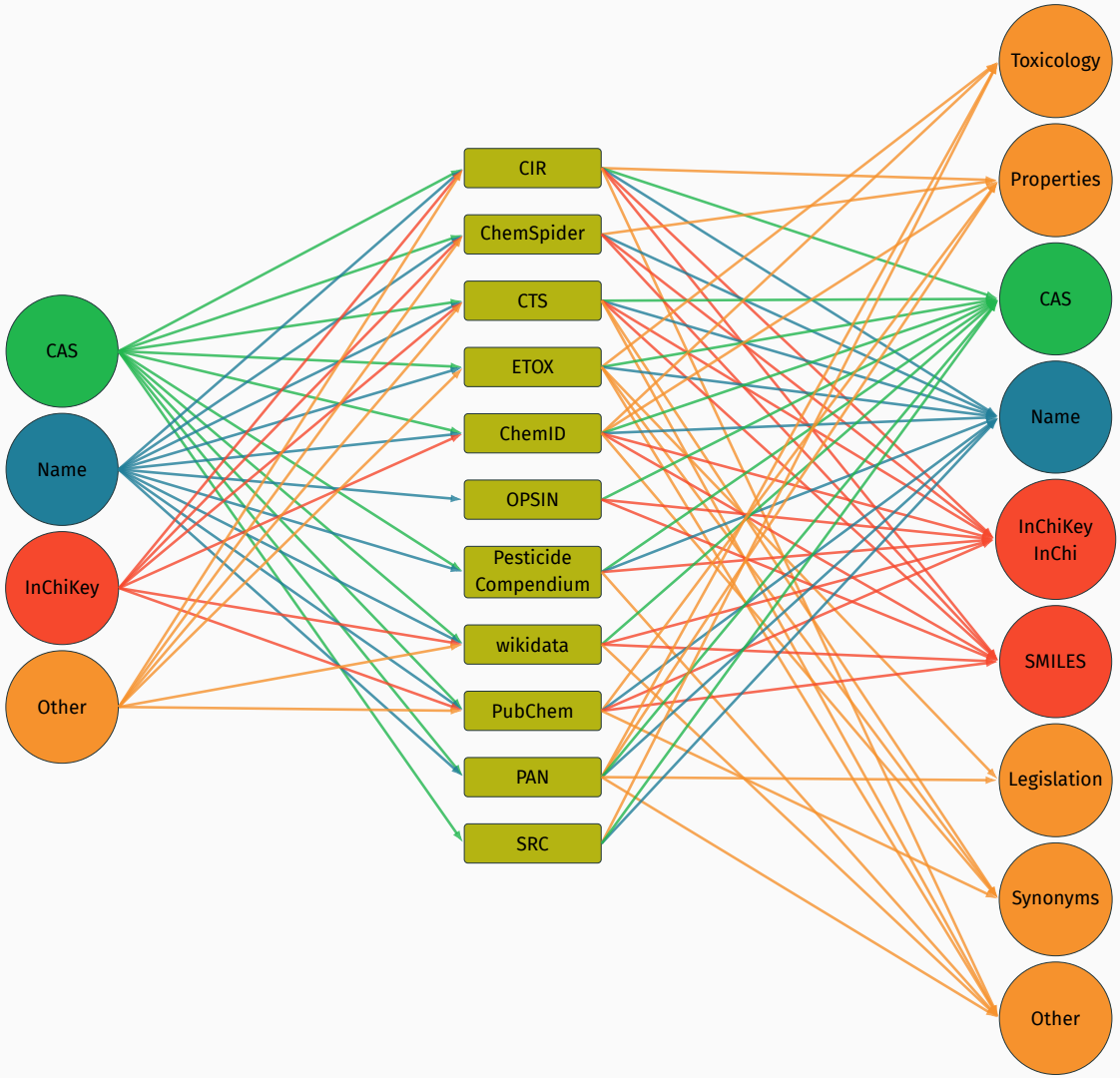
NCBI, ITIS, EOL, ...	2921-88-2, Clc1c(OP(=S)[...], InChI=1S/C9H11C[...], SBPBAQFW[...], CSID,...

Amount of data

2993 taxa	489 pesticides (+ 590 other organics)

Instead of wasting time...

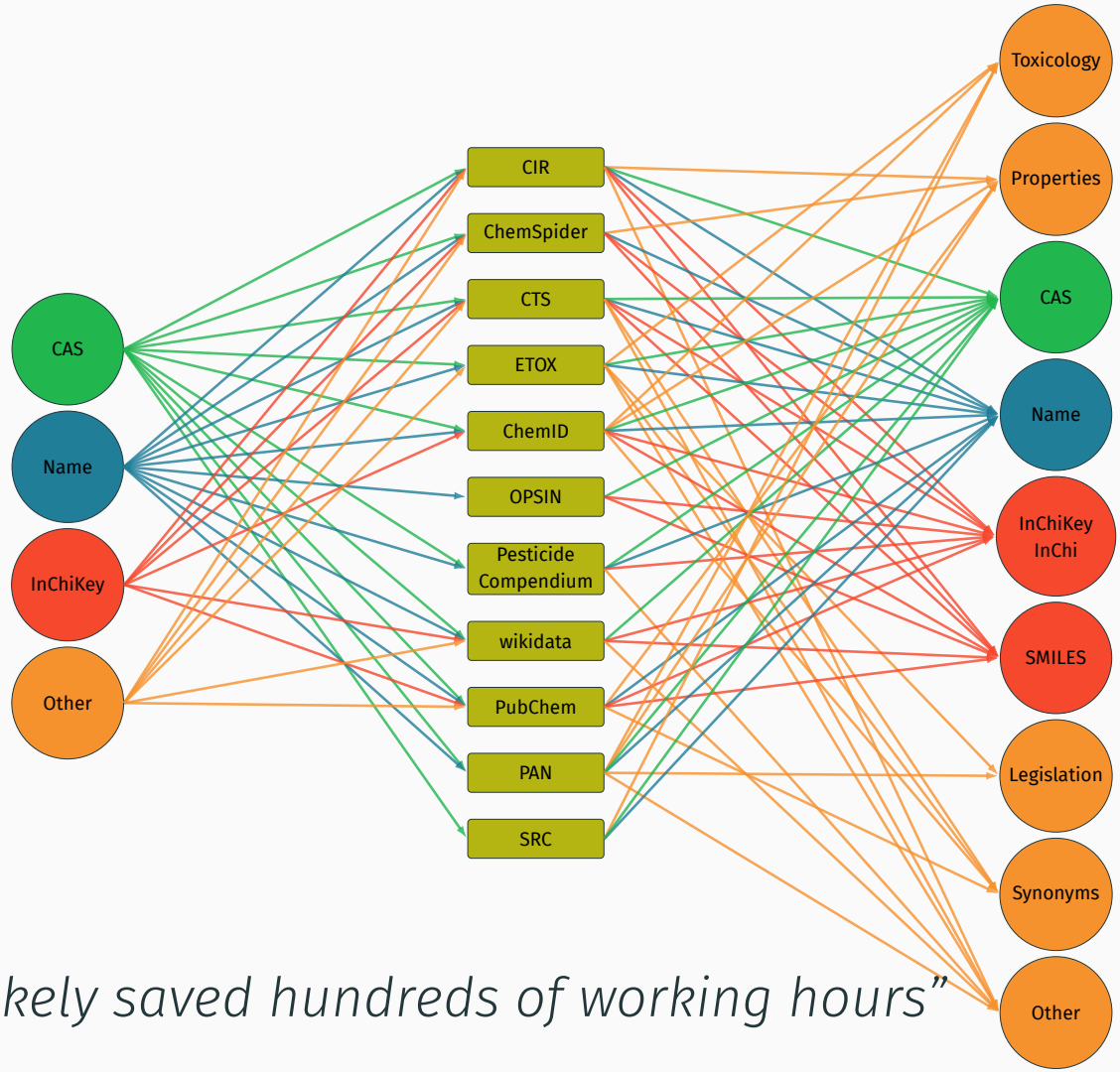
... use webchem!



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

Instead of wasting time...

... use webchem!

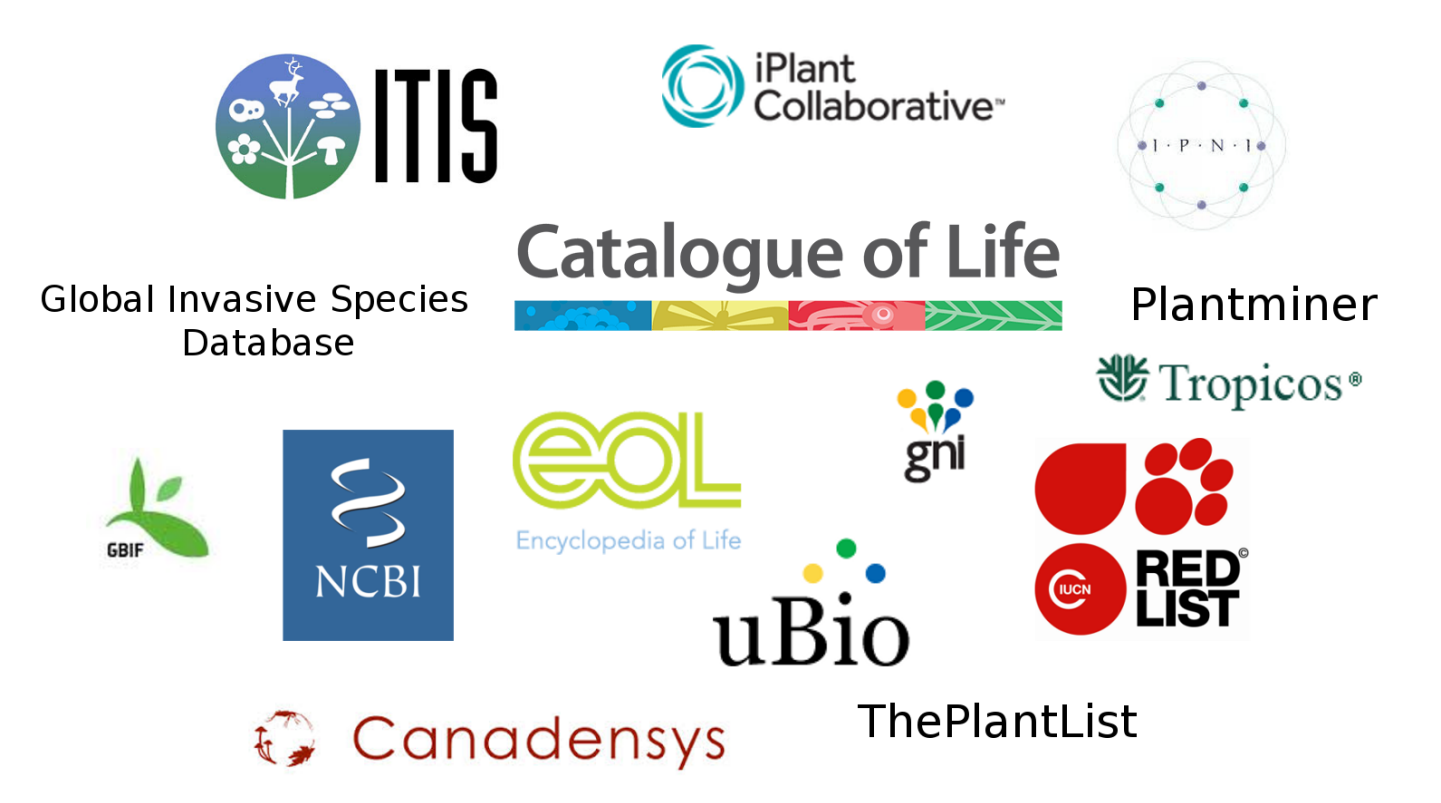


"webchem ...likely saved hundreds of working hours"

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

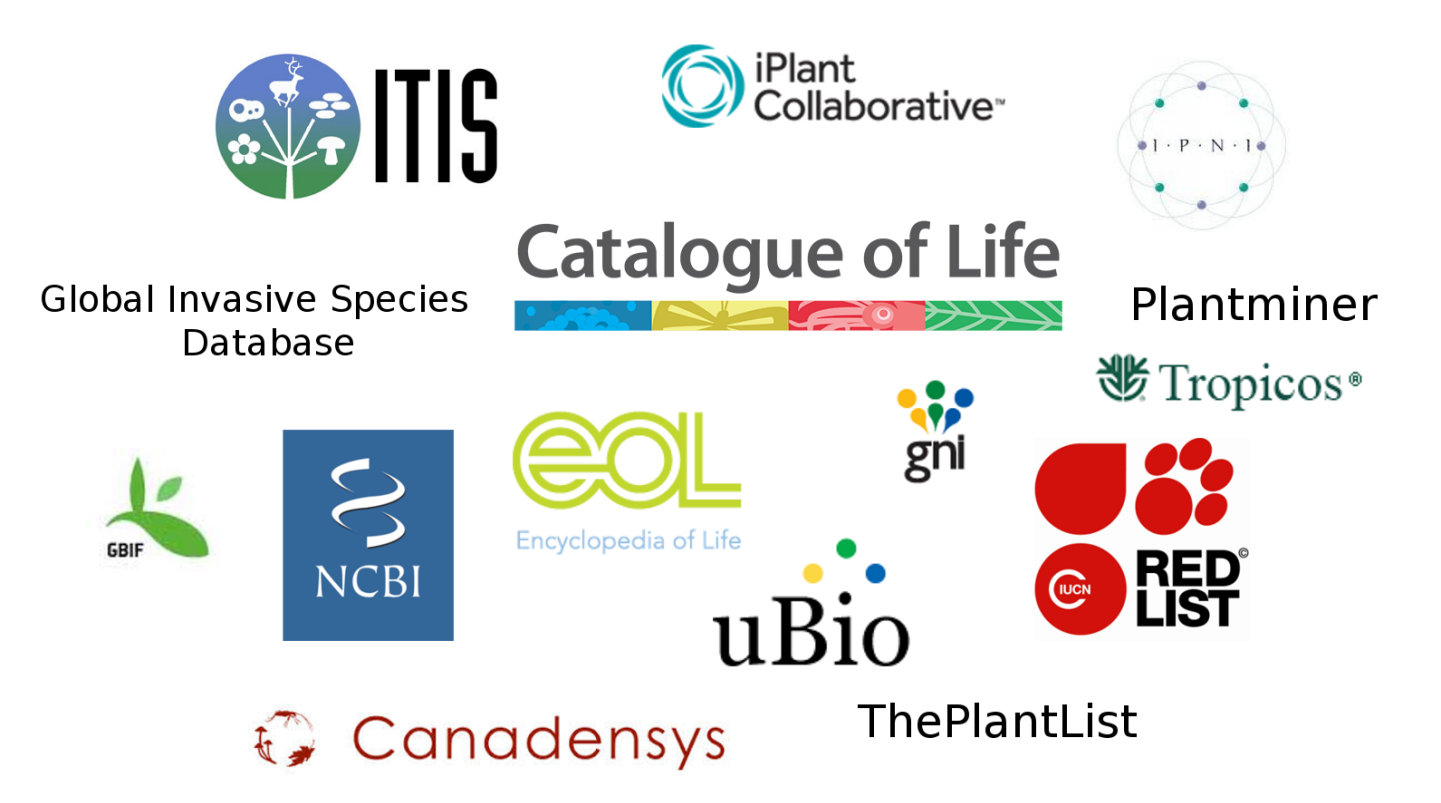
Instead of wasting time...

... 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.

Recap

What we learned

✓Improving Statistics in ERA

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

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

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

Statistical Ecotoxicology

Improving the Utilisation of Data for Environmental Risk Assessment

Eduard Szöcs

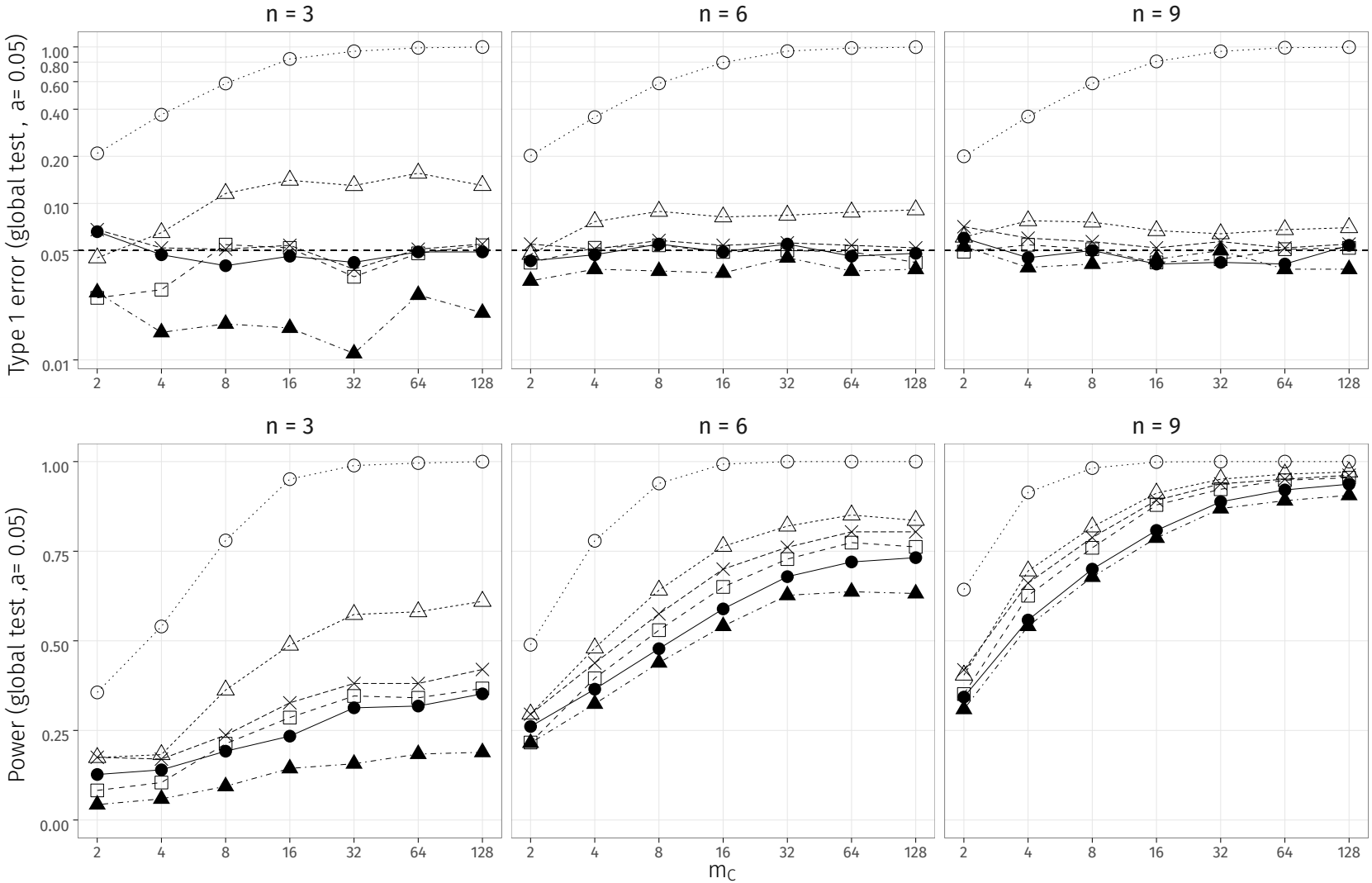
http://edild.github.io/

@EduardSzoecs

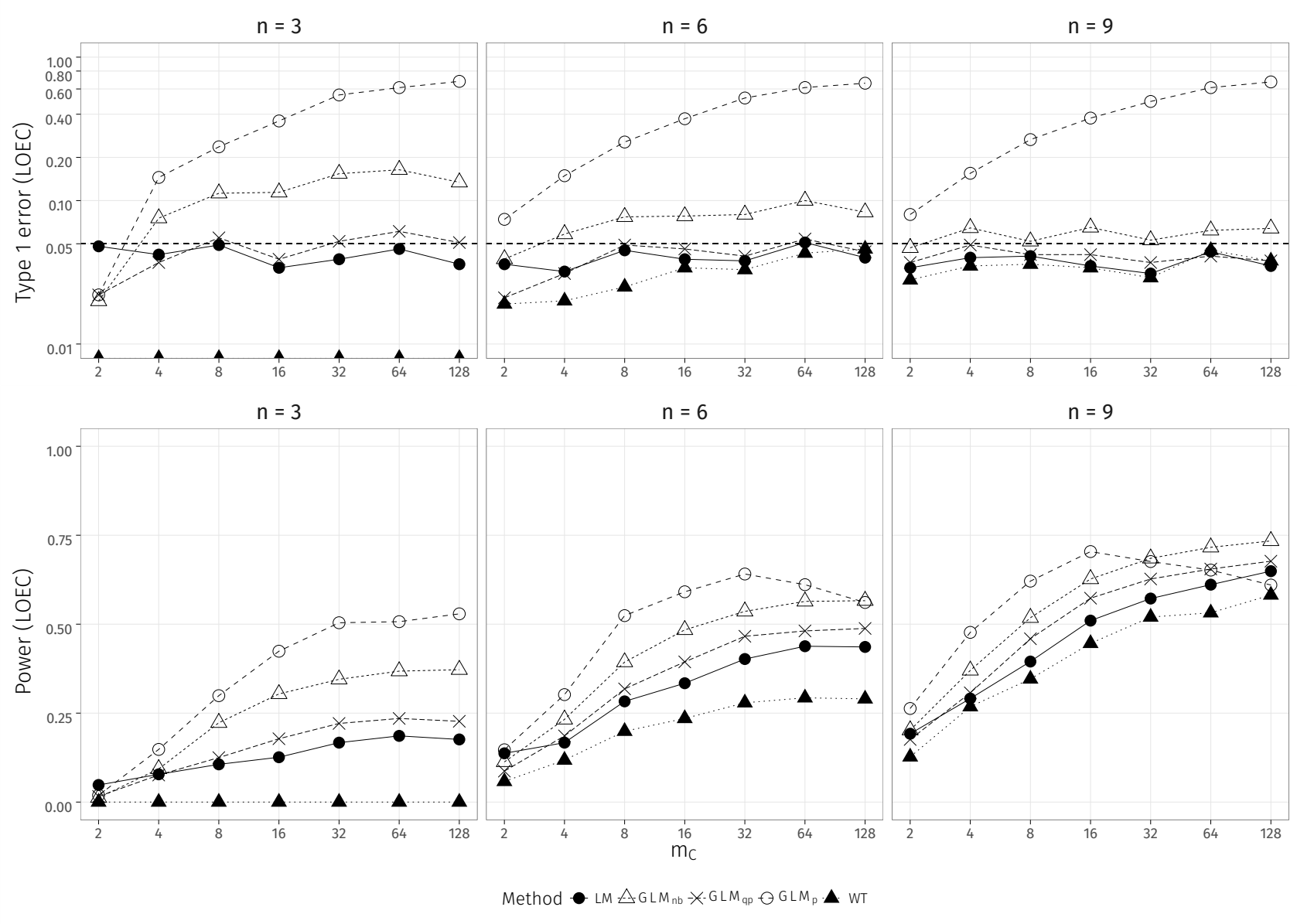
https://github.com/edild/phd_defense

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
-

ZAGA what...?

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

Comparison with recent studies (TODO!)

Szöcs (2016)

- Germany
- Monitoring
- Grab sampling
- % LOQ:

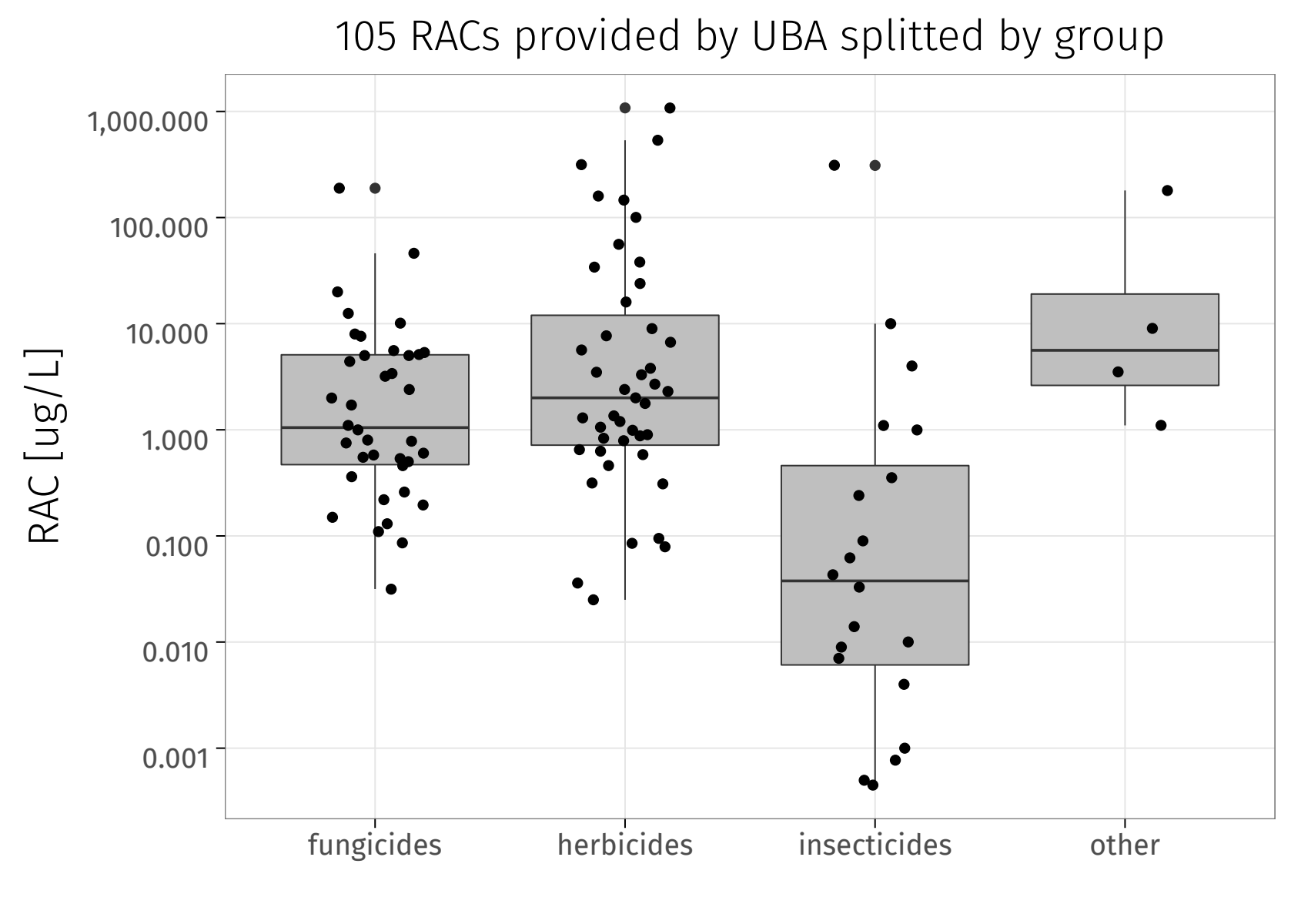
Stehle (xxx)

- Europe / Global
- Scientific Publications
- Grab & Event driven sampling
- % LOQ:

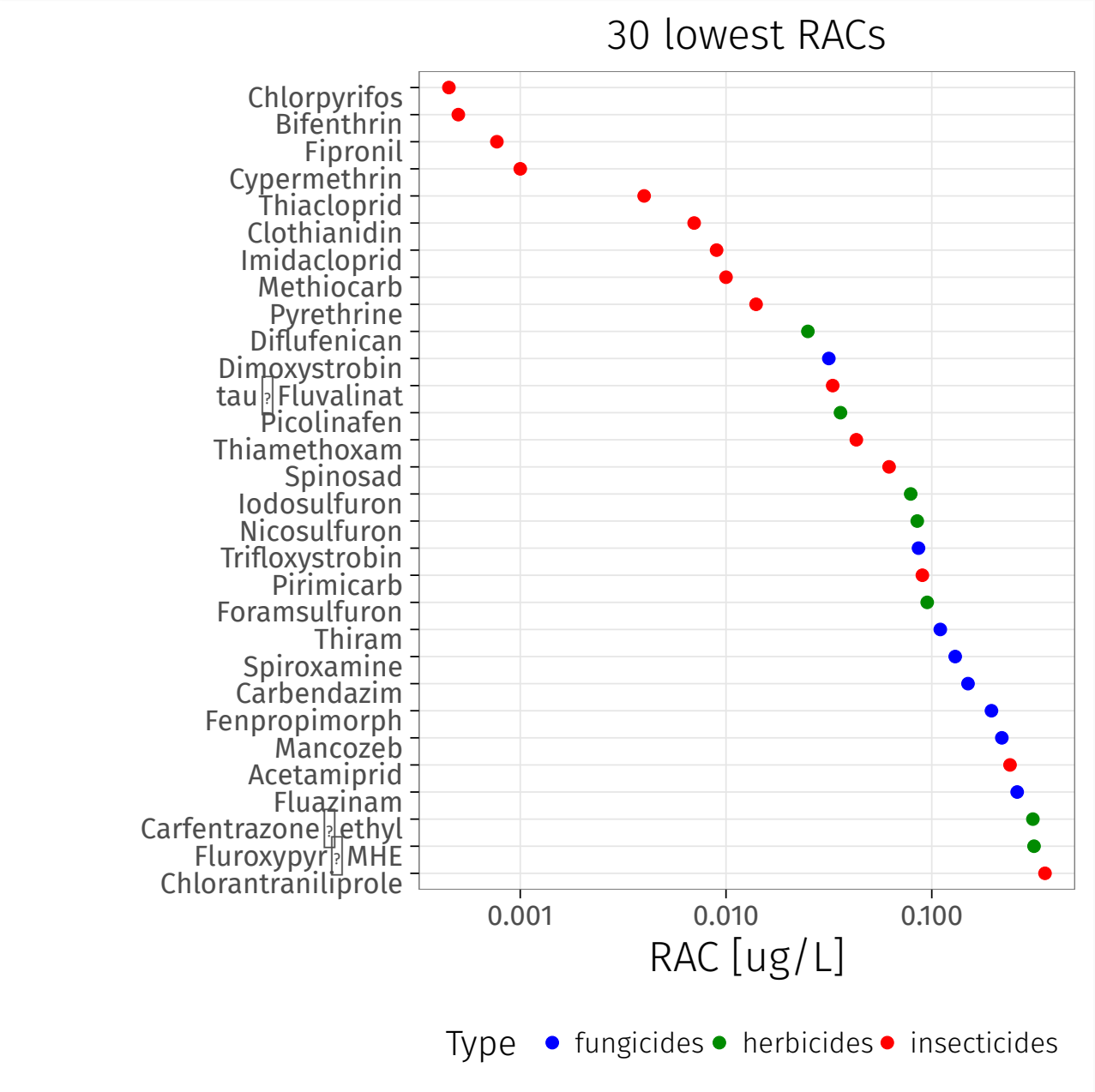
Knauer (xxx)

- Switzerland
- Monitoring
- Grab sampling
- % LOQ:

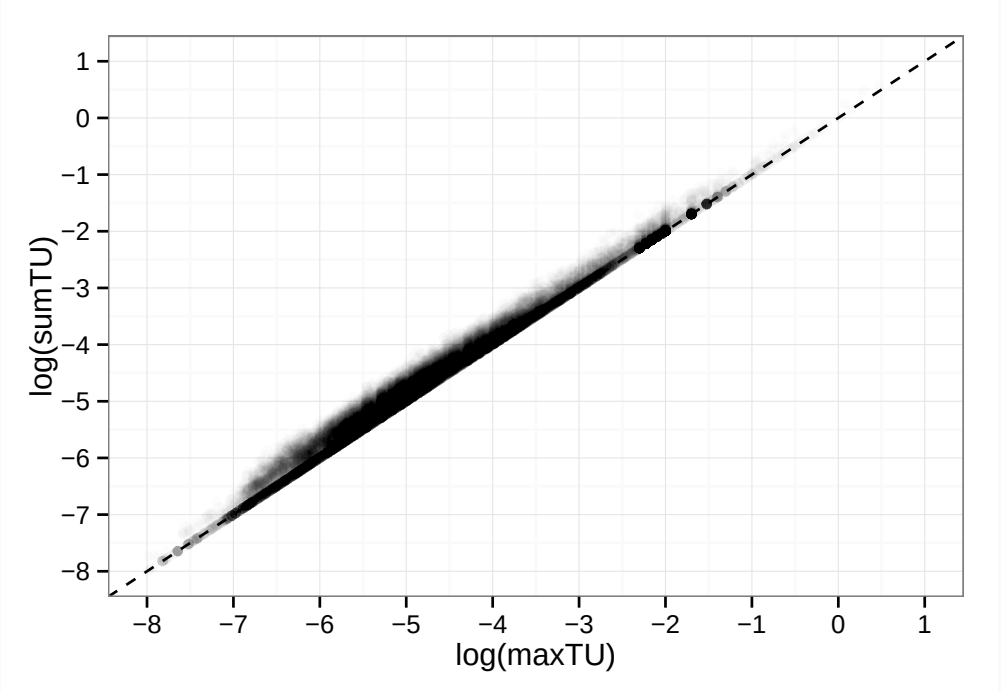
RACs by Type


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...)



Mixtures are common, but one compound dominates the risk

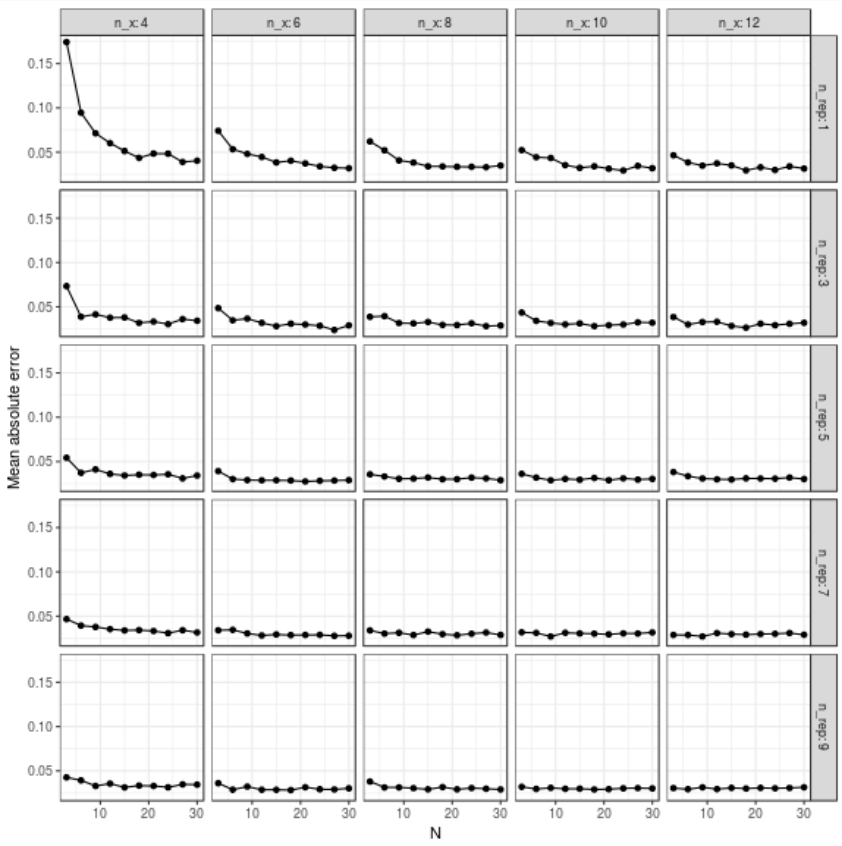


- up to 50 compounds in one sample
- high correlation
- $\sim$ 0.5 TU increase
- mainly one compound responsible for risk

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

Experimental design for dose-response experiments - a simulation

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

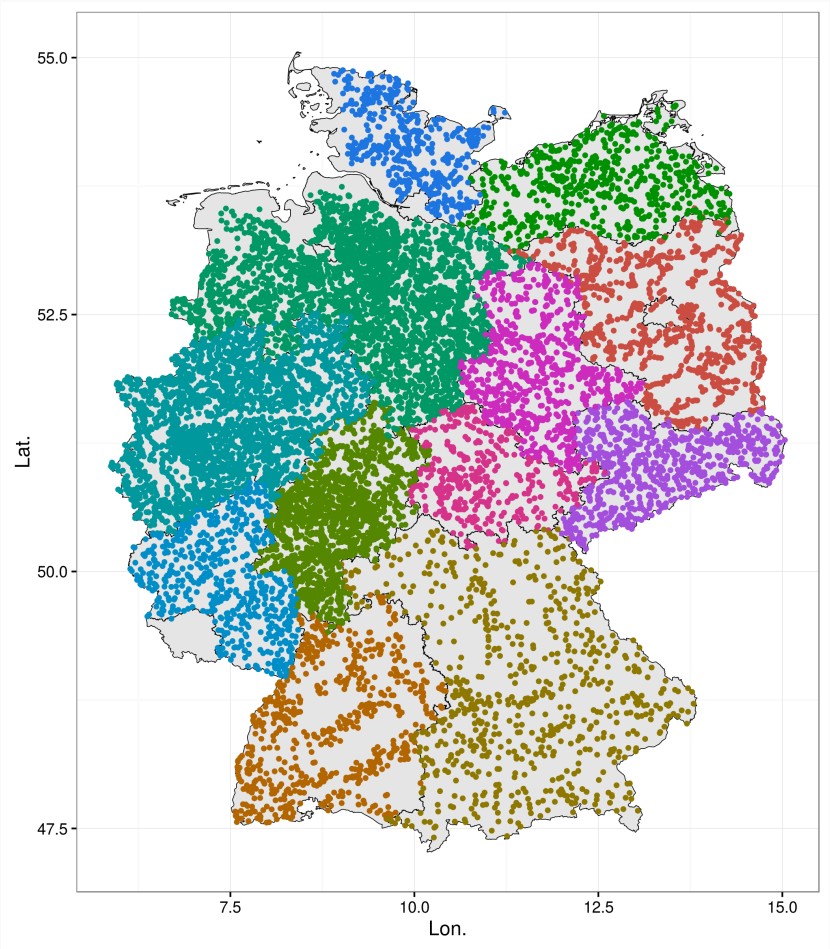


`http://uni-ko-ld.de/g3`

Reasons for observed RAC exceedances

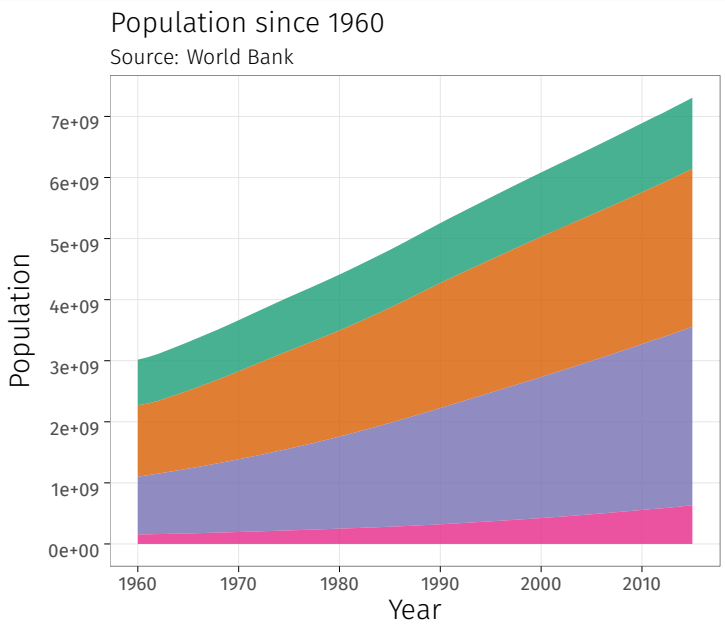
see notes on lecture on 16.12.2016

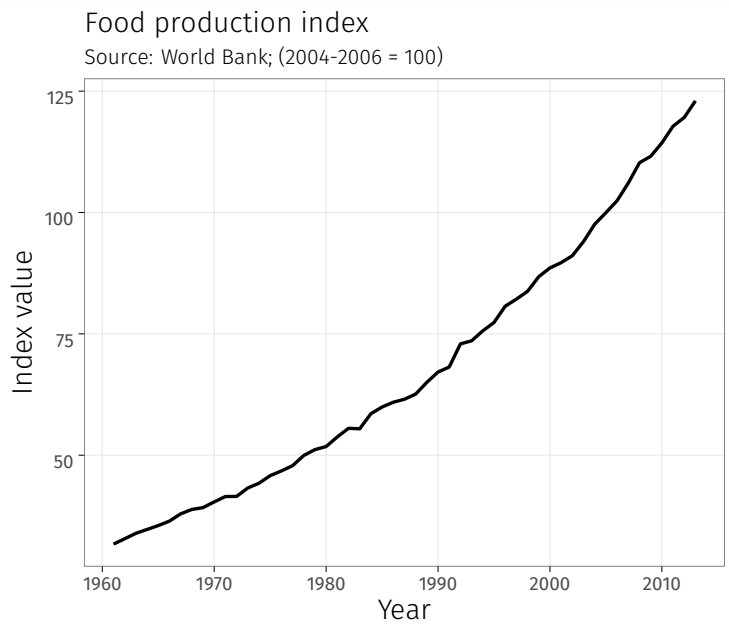
Biotic field effects

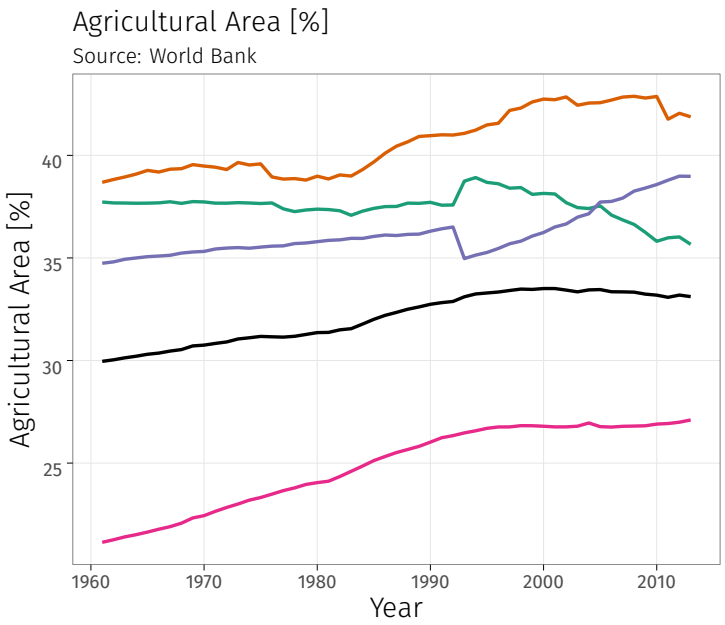


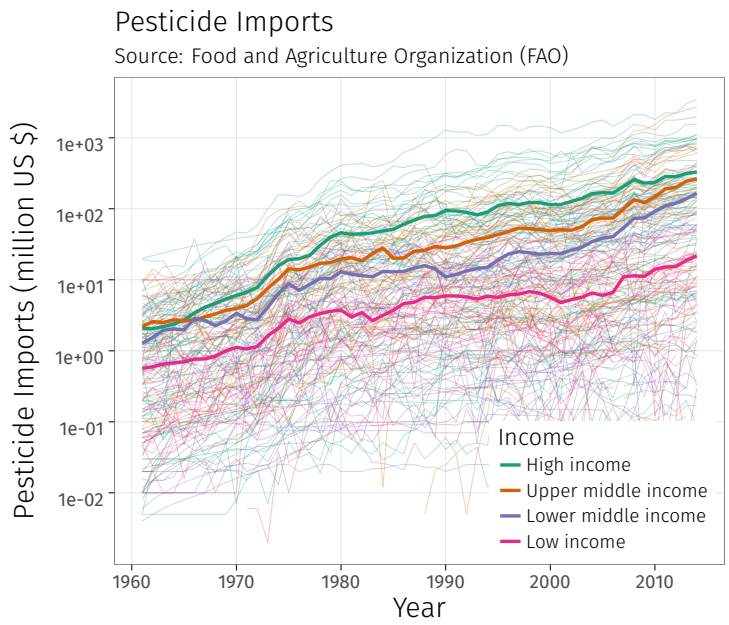
- biological data with good spatial coverage
- 60% of spatial congruence
- Large scale effects largely unknown.
- Some work left...
- Future....

A global perspective

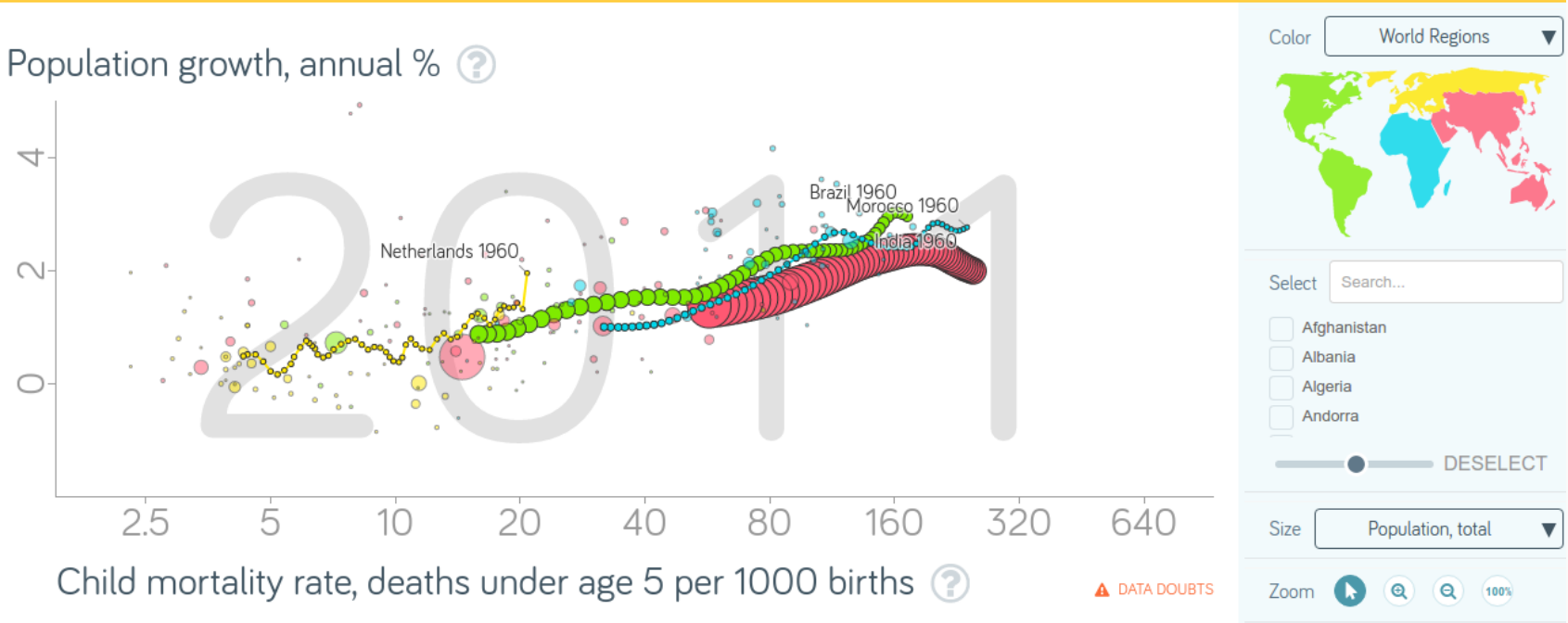








A global perspective (II)



https://www.gapminder.org/tools

Software

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

Stable versions also on CRAN...