Time series modeling of plant protection products in aquatic systems in R

Analysis of governmental monitoring data

Andreas Scharmüller
Mira Kattwinkel, Ralf Schäfer

Quantitative Landscape Ecology
University Koblenz-Landau

2018/05/16
Quantitative Landscape Ecology

- R and other open source software
- Ecotoxicology
- Effects of Plant Protection Products (PPP) / pesticides on the environment
- Aquatic systems
Introduction

Why study pesticides?

- Highly used in modern agriculture, gardens
- Environmental concern
- Glyphosate, Neonicotinoids, ...
- Germany (2016):
  - 753 pesticides
  - 270 substances
- Groups:
  - fungicides
  - herbicides
  - insecticides
Data
Data

- federal monitoring program
- 3,116 sampling sites
- 3,246,690 substance detections
- 495 substances
- stored in a PostgreSQL database:
Data

```r
require(RPostgreSQL)
require(data.table)

# load data
drv = dbDriver("PostgreSQL")
con = dbConnect(...)

g = "SELECT * FROM schema.tab"
q = g
dt = dbGetQuery(con, query = q)
setDT(dt)

dbDisconnect(con)
dbUnloadDriver(drv)
```
Data
Data

- Left skewed environmental data
  - LOQ: Limit of quantification
  - Excess of 0s
- Heterogenous data set
  - Sampling frequency
  - LOQ can change over time
  - Measured compounds
- Seasonal variability
Comparability between substances?

- 10µg of substance A as toxic as 10µg od substance B?
Comparability between substances?

- 10µg of substance A as toxic as 10µg od substance B?

  It is only the dose which makes a thing poison.

  — Paracelsus
Comparability between substances?

- 10µg of substance A as toxic as 10µg od substance B?
  - It is only the dose which makes a thing poison.
    — Paracelsus

- Ecotoxicological tests
  - Effect Concentrations - EC50

![Graph showing log[Dose] vs. Response]
Comparability between substances?

- 10µg of substance A as toxic as 10µg of substance B?

  It is only the dose which makes a thing poison.
  — Paracelsus

- Ecotoxicological tests
  - Effect Concentrations - EC50

- EPA ECOTOX data base
Toxic Unit (TU)

in-stream concentrations ...

\[
\text{dt$\text{value}[1:3]} \quad \# \quad \text{concentrations in } \mu\text{g/L}
\]

## [1] 0.120 0.018 0.000

... relate to effects

\[
TU_{algae} = \log_{10}\left(\frac{\text{concentration}}{EC50_{algae}}\right)
\]
Research questions
Research questions

Are there months of increased in-stream occurrence of pesticides?

• Occurrence model:
  ◦ Binary data: concentration > LOQ: 1, concentration < LOQ: 0
  ◦ pa ~ month + year + site

How are different organism groups (Algae, Invertebrates, Fish) effected by pesticide concentrations throughout the year?

• Effect/TU-Model:
  ◦ Continuous data
  ◦ TU ~ month + site
Data preparation
Filter data

dt = dt[state == 'SN']
dt = dt[pest_type %in% c('fungicide', 'herbicide', 'insecticide')]
Filter data

\[
dt = dt[state == 'SN']
dt = dt[pest_type %in% c('fungicide', 'herbicide', 'insecticide')]
\]

\[
\text{uniqueN(dt$site)}
\]

## [1] 413

\[
\text{dt[ i = value > 0, }
\]
\[
j = .N,
\]
\[
\text{by = pest_type]}
\]

## pest_type N
## 1: fungicide 2455
## 2: herbicide 10890
## 3: insecticide 875
Filter data

Substances quantification-ratio > 5%

```r
subst_fin = dt[ ,
    .(perc = .SD[ value > 0, .N ] / .N),
    subst_name ][perc > 0.05][order(-perc)]
subst_fin[ , perc := round(perc,2)]
head(subst_fin)
```

```r
## subst_name perc
## 1: Boscalid 0.39
## 2: Bentazon 0.38
## 3: Isoproturon 0.37
## 4: Quinmerac 0.36
## 5: Glyphosate 0.29
## 6: Azoxystrobin 0.27
```

```r
nrow(subst_fin)
```

```r
## [1] 31
```
Occurrence model
Occurrence model

fit the model for each substance individually

```r
mdt[, pa := as.numeric(as.logical(value)) ]
mdt[, time := as.numeric(date) / 1000 ]

require(mgcv)
for (i in seq_along(substances)) { # for 31 pesticides
  mdt = dt[ subst == substances[i] ]
  mod_pa = gam(pa ~
               s(month, bs = 'cc', k = 12) +
               s(time, k = 20) +
               s(year, bs = 're') +
               s(site, bs = 're'),
               data = mdt,
               family = binomial(link = 'logit'),
               method = 'REML')
  # ...
}
```
Occurrence model - Herbicides
Occurrence model - Herbicides
Occurrence model - Herbicides

Seasonal change in the occurrence of herbicides
pre-emergence herbicides
Occurrence model - Fungicides
Effect model
Effect model

Effect model

dt[ , TU_algae := log10(value / EC50_algae) ]
dt[ , TU_inv := log10(value / EC50_inv) ]
dt[ , TU_fish := log10(value / EC50_fish) ]

Maximum per site & month

dt_agg = dt[ ,
   .(maxTU_al = max(TU_algae),
   maxTU_inv = max(TU_inv),
   maxTU_fi = max(TU_fish)),
   .(site, month) ]
Effect model

maximum: TU-Algae, TU-Invertebrates, TU-Fish

```r
require(mgcv)

for (i in seq_along(todo)) { # for 3 TUs
  # ...
  mod_al = gam(maxTU_al ~
    s(month, bs = 'cc', k = 12) +
    s(site, bs = 're'),
    family = gaussian(),
    data = mdt_agg,
    method = 'REML')
  # ...
}
```
Effect model

All organism groups (Algae, Fish, Invertebrates)
Conclusions

- Occurrence model
  - identify peaks in occurrence (for well measured substances)

- Effect model
  - underestimation of effects
  - sampling effort
  - different physical chemical properties of substances

- Improve model
  - include interactions
  - refine selection of EC50 values for TU calculations
  - other covariates:
    - percentage of agriculture in catchments
    - precipitation on/before sampling date
R packages + tools

- data storage + preparation
  
  ```
  require(RPostgreSQL)
  require(data.table)
  ```

- modeling
  
  ```
  require(mgcv)
  ```

- visualization
  
  ```
  require(ggplot2)
  require(sf)
  ```

- slides
  
  ```
  require(rmarkdown)
  require(knitr)
  require(xaringan)
  ```
Time series modeling of plant protection products in aquatic systems in R

Analysis of governmental monitoring data

Thank you for your attention!

Andreas Scharmüller
Mira Kattwinkel, Ralf Schäfer
Quantitative Landscape Ecology
University Koblenz-Landau

@andschar
scharmuelle@uni-landau.de