LEARNING FROM EXPERIENCE – PESTICIDE MONITORING IN ENGLISH GROUNDWATER

Fred Worrall

Department of Earth Sciences, Durham University.

Durham University

Earth Sciences

Acknowledgements

- Miles Wilson
- Tim Besien
- ReFINE consortium
  - Funding from Shell, Chevron, Total, GDF-Suez, INEOS, & NERC
  - RSC has supported ReFINE

ReFINE
RESEARCHING FRACKING

Environment Agency

Hypothesis

- The occurrence of pesticides in groundwater can predicted by generalised linear modelling

$$G = f(site) + f(compound) + f(site \times compound) + error$$

What endpoint to predict to?

The vulnerability

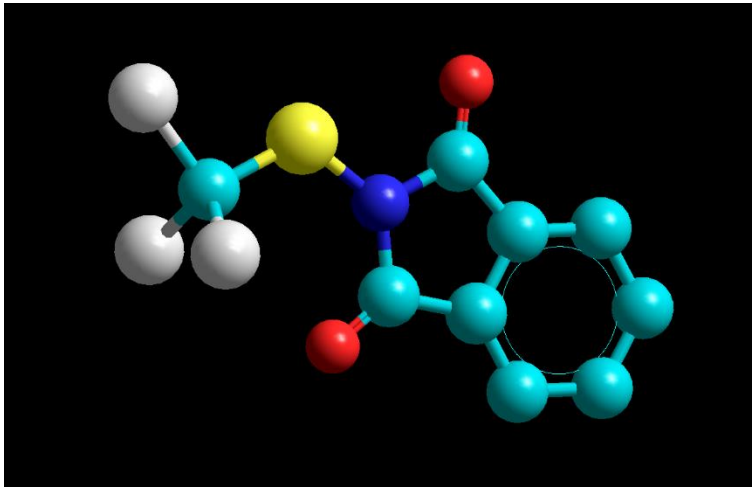
The hazard

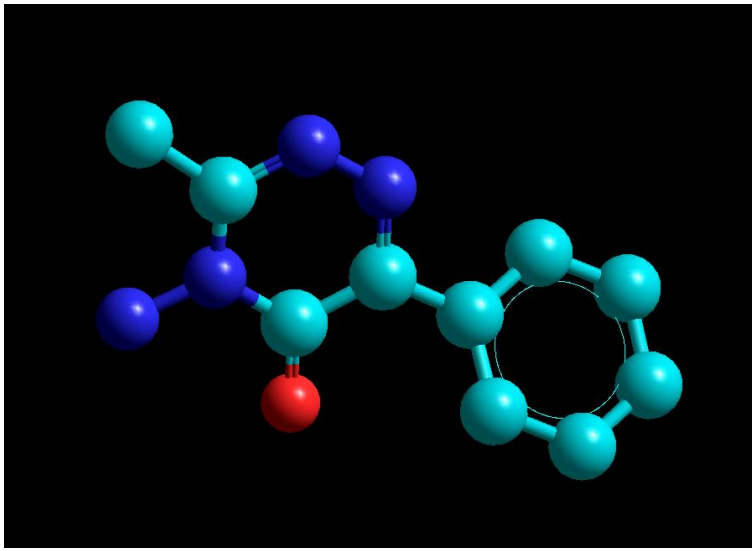
Interaction – some sites are less vulnerable to some compounds than others

What distribution?

The Hazard – f(compound)

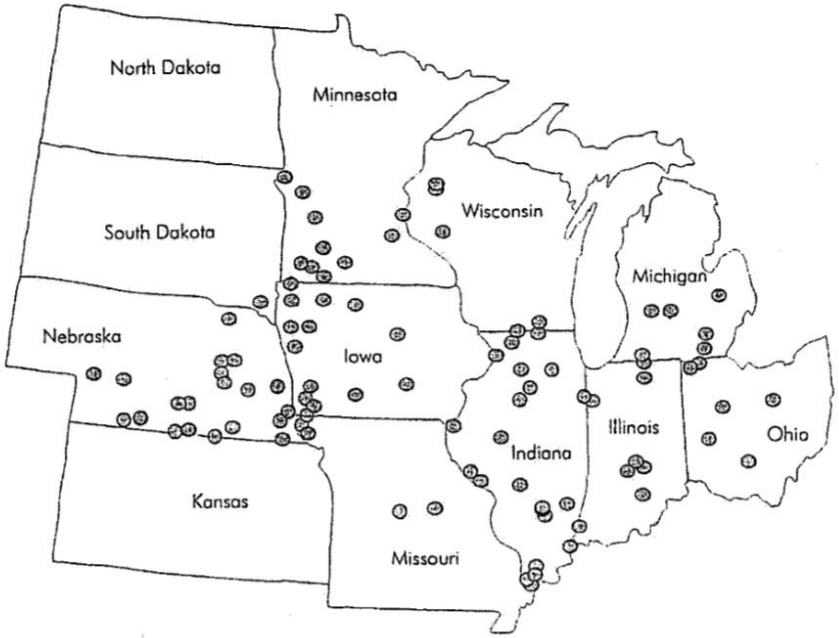
- Use properties only dependent upon the molecule
  - K_{oc} and $t_{1/2}$ are no good
- Model using molecular descriptors
  - Connectivity
  - Atom and group counts
  - Molecular orbitals
  - Hydration energy
  - Dipole moment
  - Refractivity



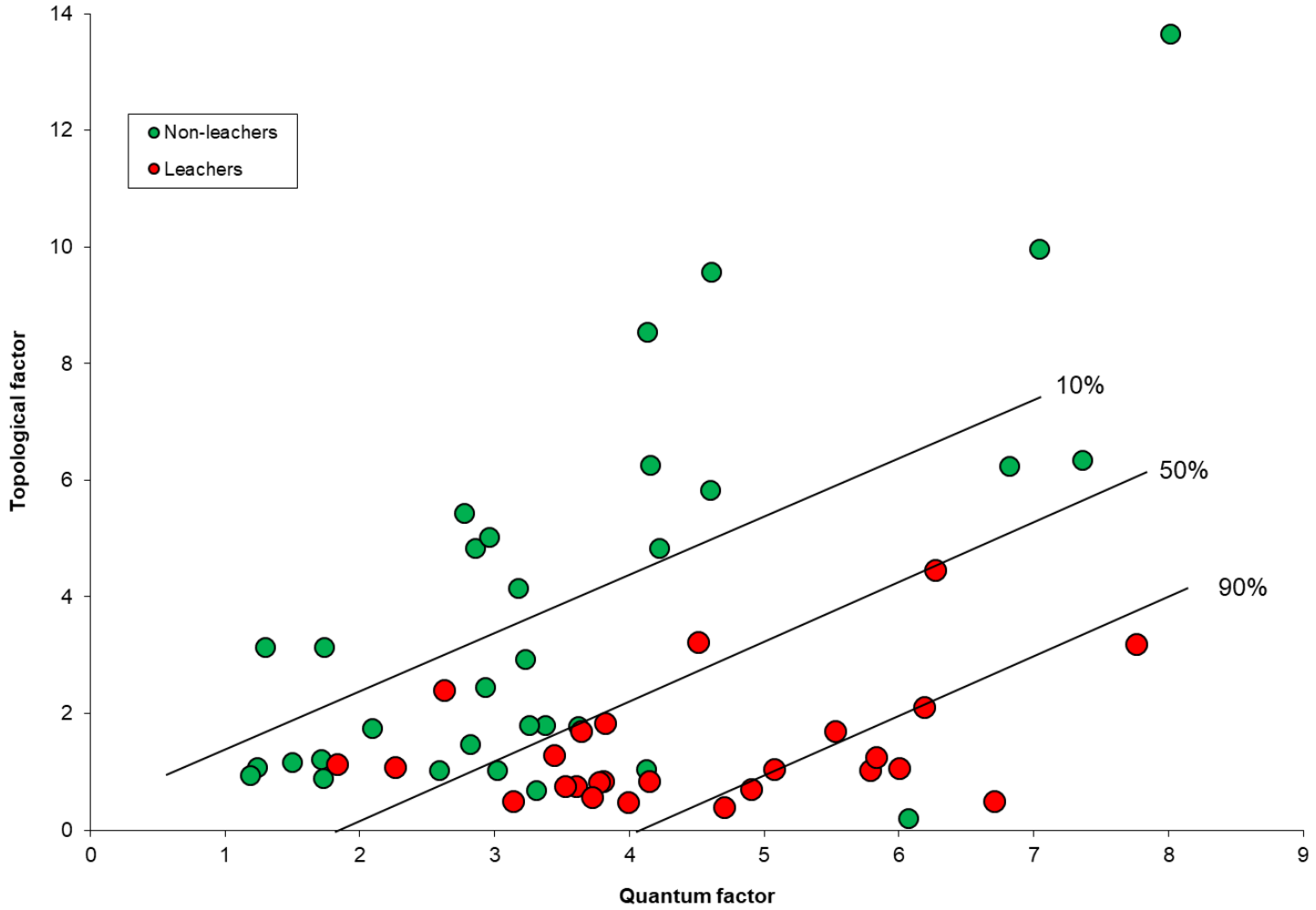


The Hazard – f(compound) – US data

- Great Valley, California
  - 40 compounds
- Kolpin et al. (1997)
  - 56 compounds
  - 303 boreholes
  - 12 US states
- 61 compounds, 27 compounds in common
- Logistic model
  - Bernoulli trial
  - Modelling detected vs. non-detected



The Hazard – f(compound) – molecular model



The Hazard – f(compound) – molecular model

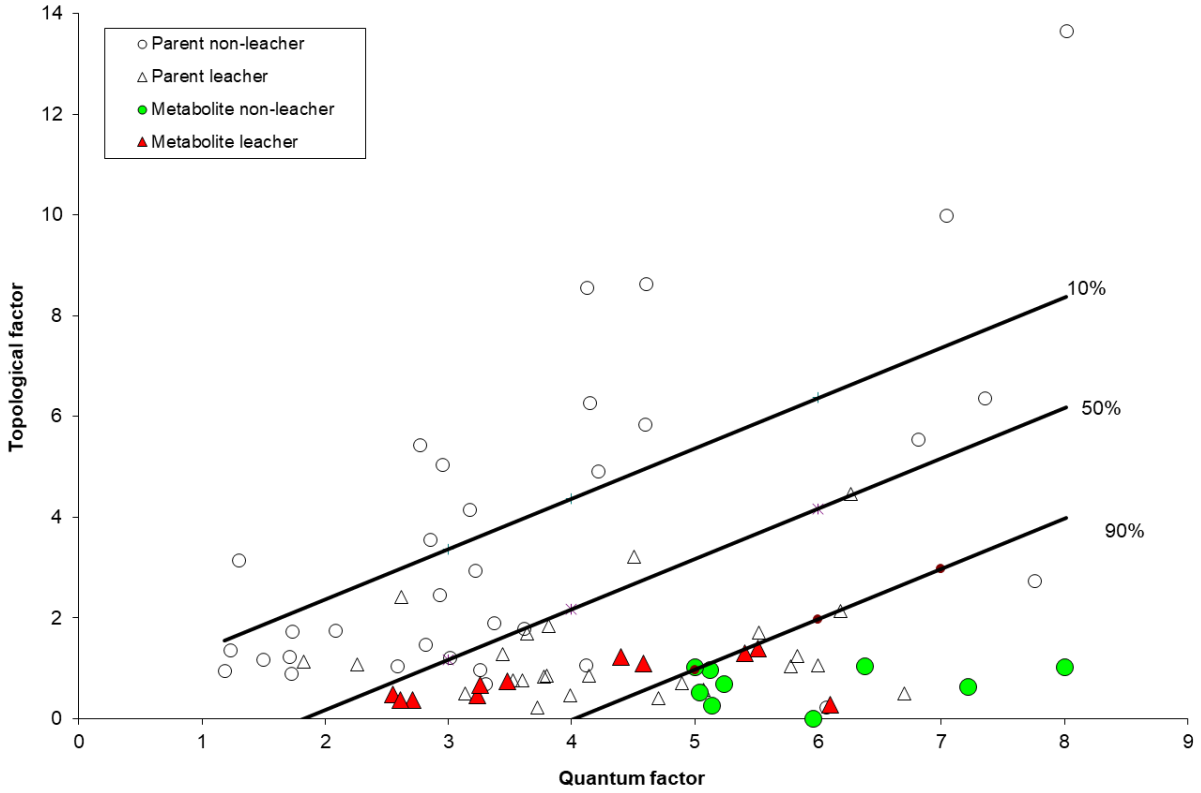
- The probability of a compound being detected (θ)

$$\log\left(\frac{\theta}{1-\theta}\right) = \underbrace{0.77\mu - 0.185\,\Delta H_{hyd}}_{\text{Quantum factor - Solubility}} \underbrace{- 2.215\,{}^{6}\chi_{p}^{v} - 53\,{}^{7}\chi_{pc}^{v}}_{\text{Topological factor - degradation?}} - 1.27$$

- Correctly classifies 91% of compounds
- 50% probability of being detected if:

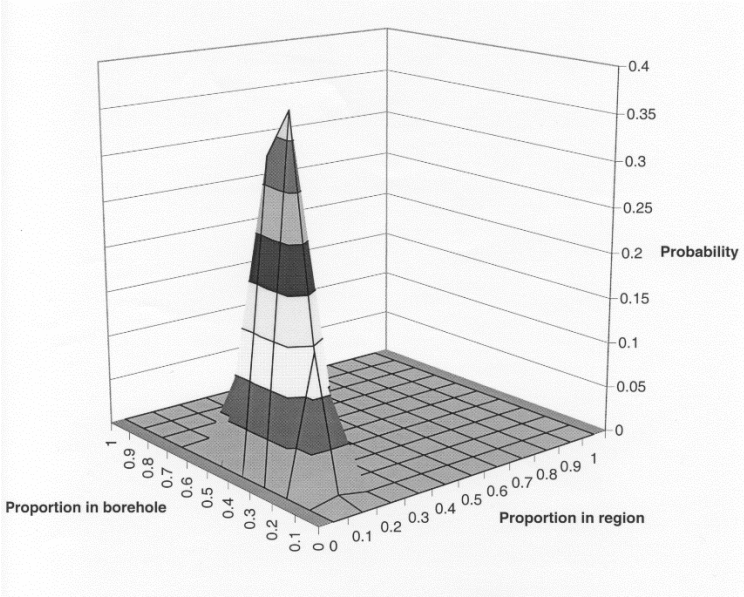
$$0.28\,\mu < {}^{6}\chi_{p}$$

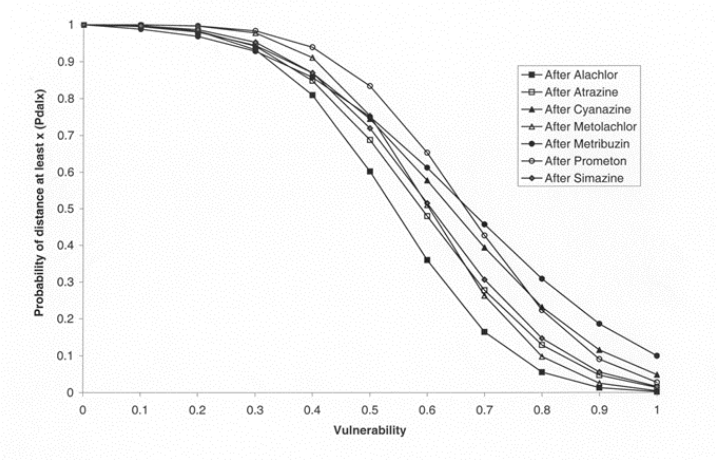
The Hazard – f(compound) – metabolites



- The mobile daughter of immobile parent or the immobile daughter of a mobile parent

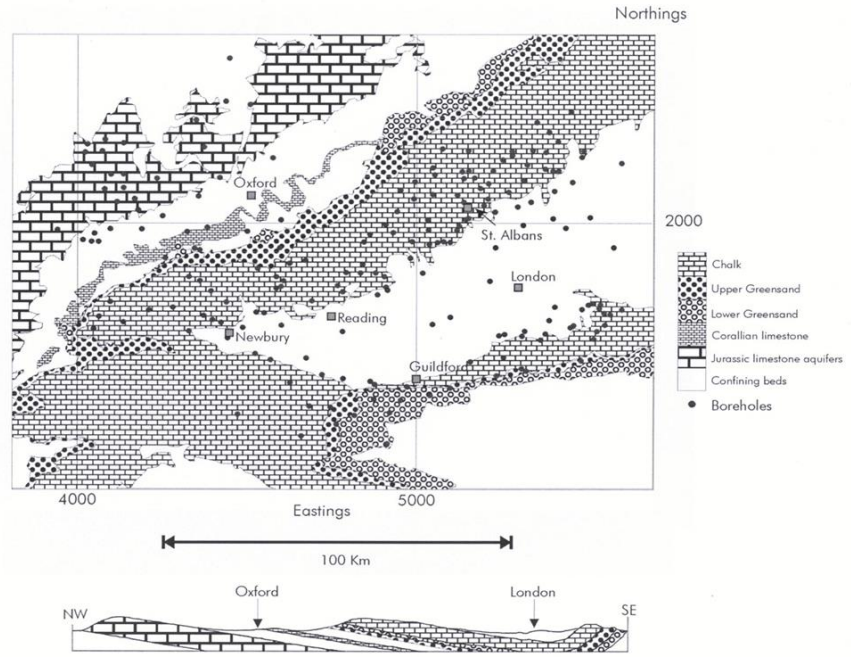
Vulnerability – f(site)



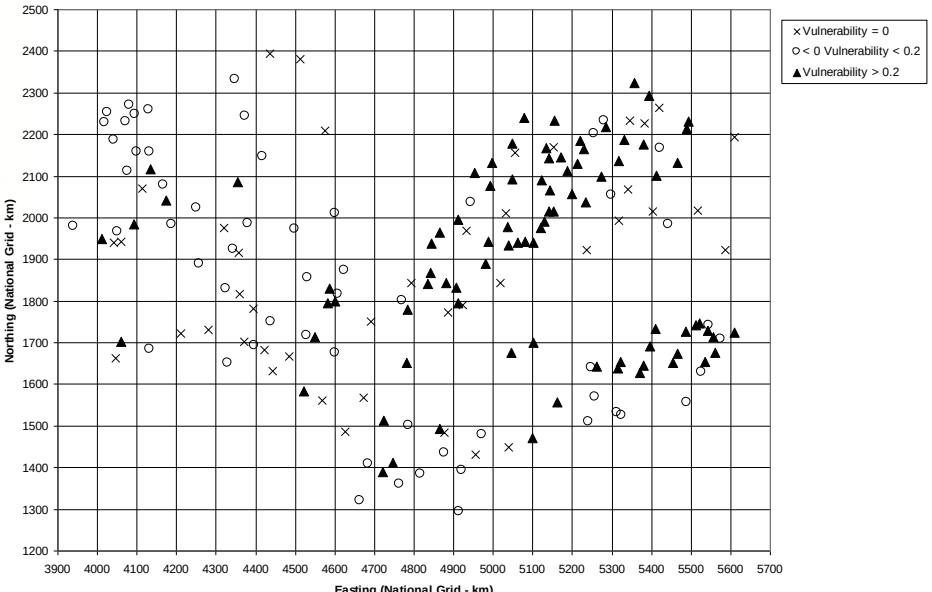


- A vulnerable site is one which shows more detects of pesticides than would be expected
- Bayesian analysis of proportions
  - Prior distribution is the proportion of detects in the region
  - Pesticide analysis seen as a Bernoulli trial
  - Updates with each compound
  - Independent of compound
  - Probability of detection of the next compound
  - Comes with uncertainty

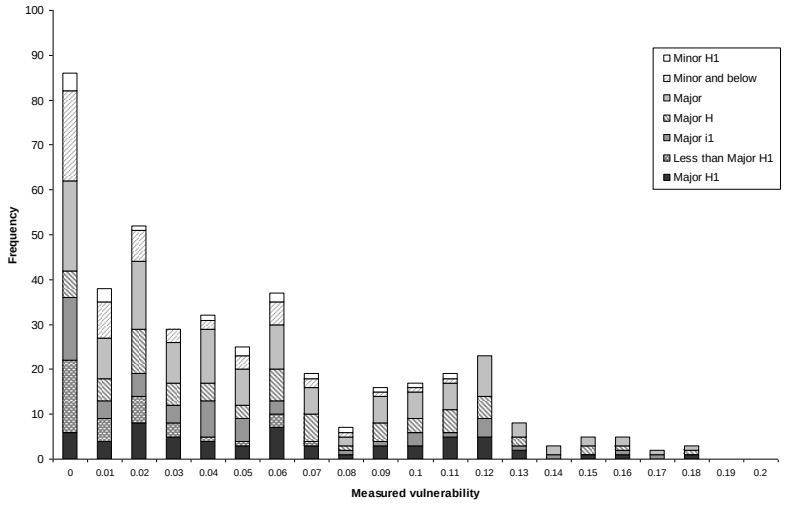
Vulnerability – f(site) – Thames basin



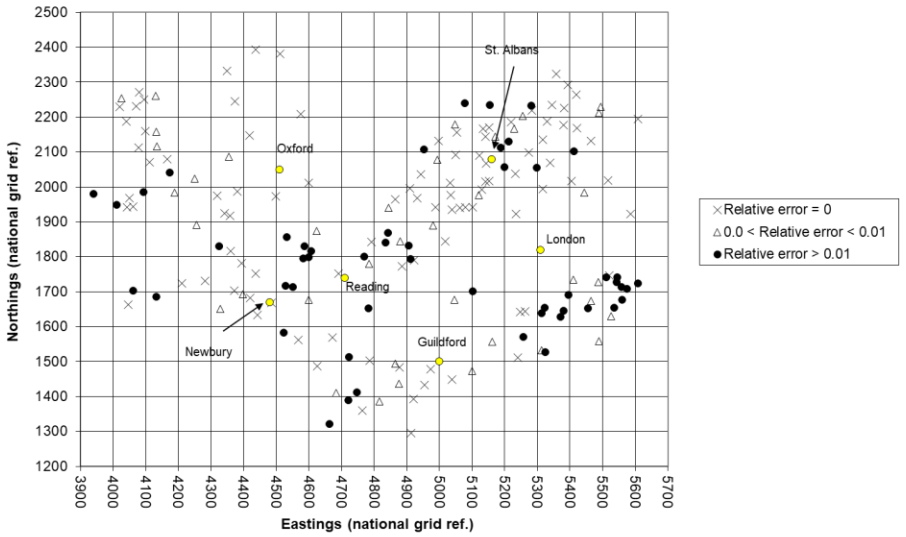
- Thames region EA has best pesticide dataset in UK
  - 359 boreholes
  - 27 compounds



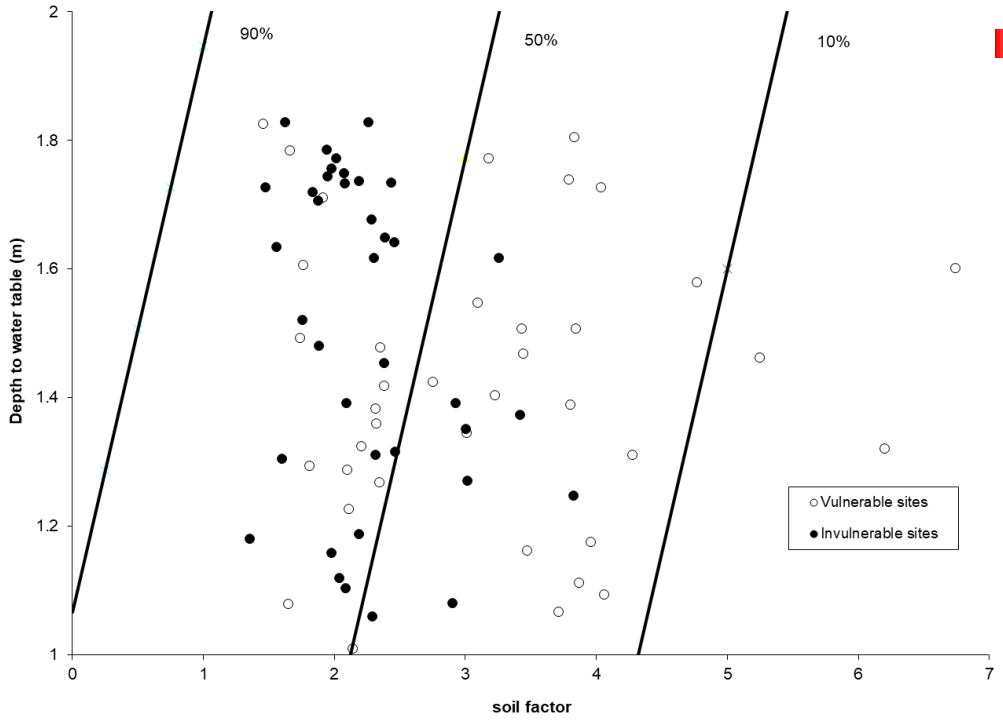
Vulnerability – f(site)



- Can be compared to existing groundwater vulnerability
  - They do not match well
- The uncertainty can be plotted



Vulnerability – f(site) – US dataset



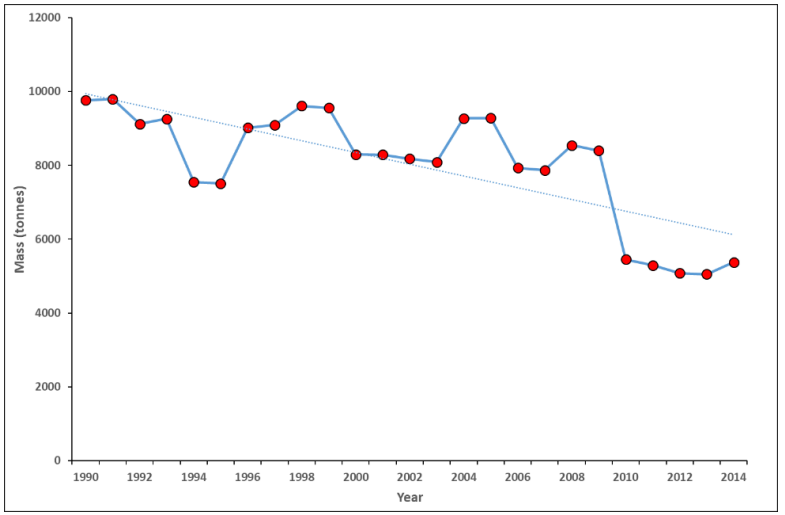
- What controls vulnerability?
  - Compare detection to catchment properties

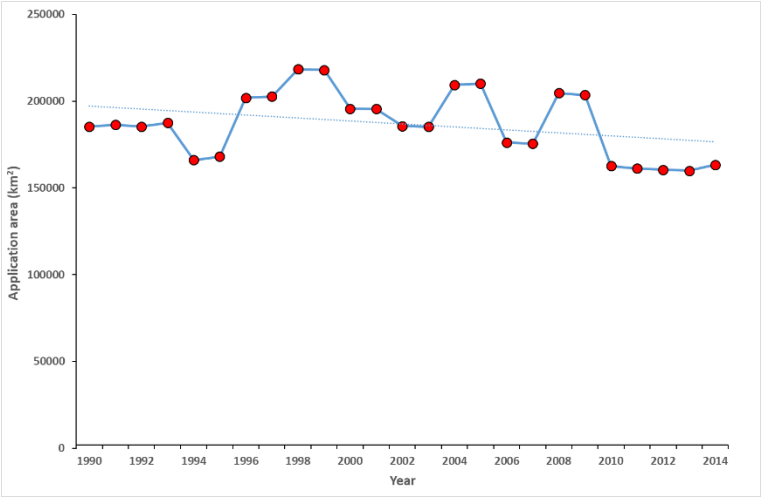
- Probability of detecting a compound (θ)

$$\ln\left(\frac{\theta}{1-\theta}\right) = 0.985 - 0.695\%OM - 0.03\%sand + 1.135wtdepth$$

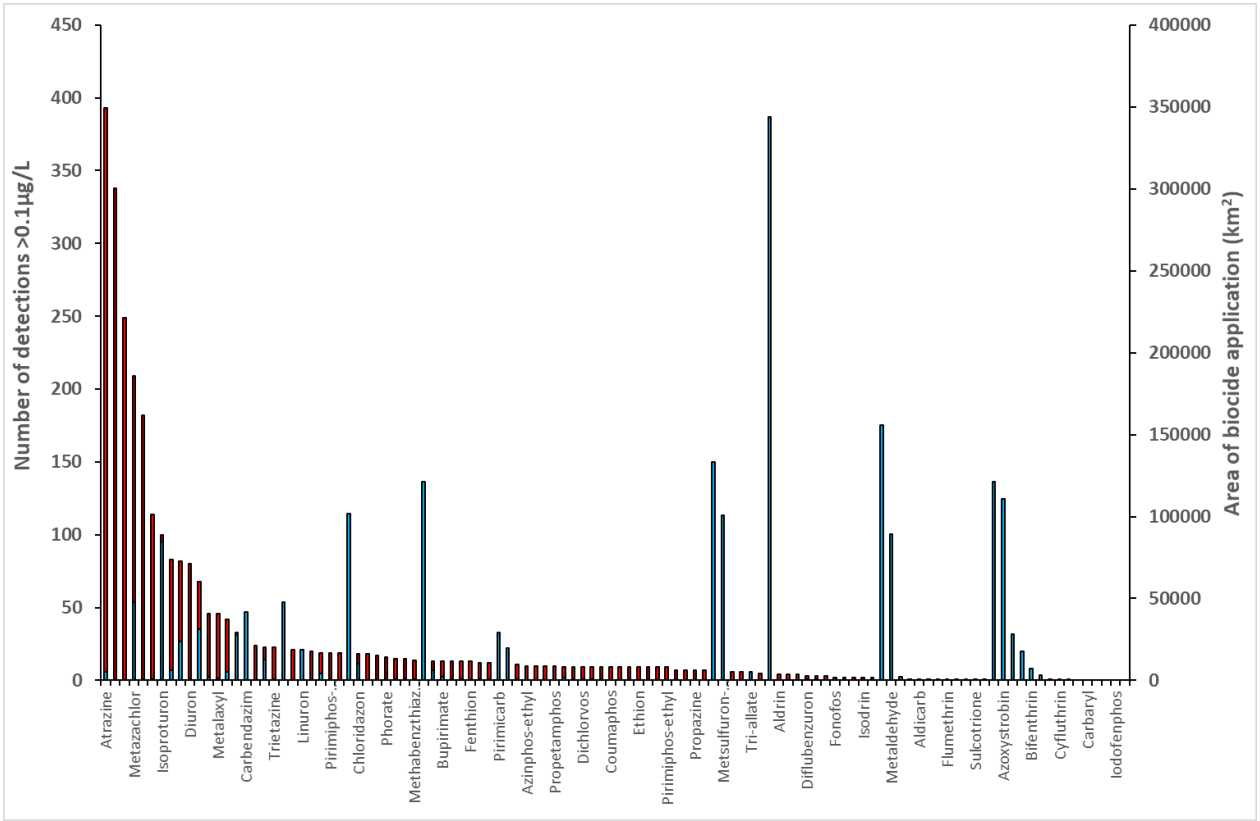
English groundwater

- EA groundwater data for all English groundwater
  - Data from 2005 to 2017
  - 113 compounds
  - 3357 unique borehole locations
  - > 1.5 million observations
- Pesticide usage statistics available
  - Pesticide area and amount
  - 68300 tonnes applied over study period
  - Significant decline in total amount, total area and average application rate



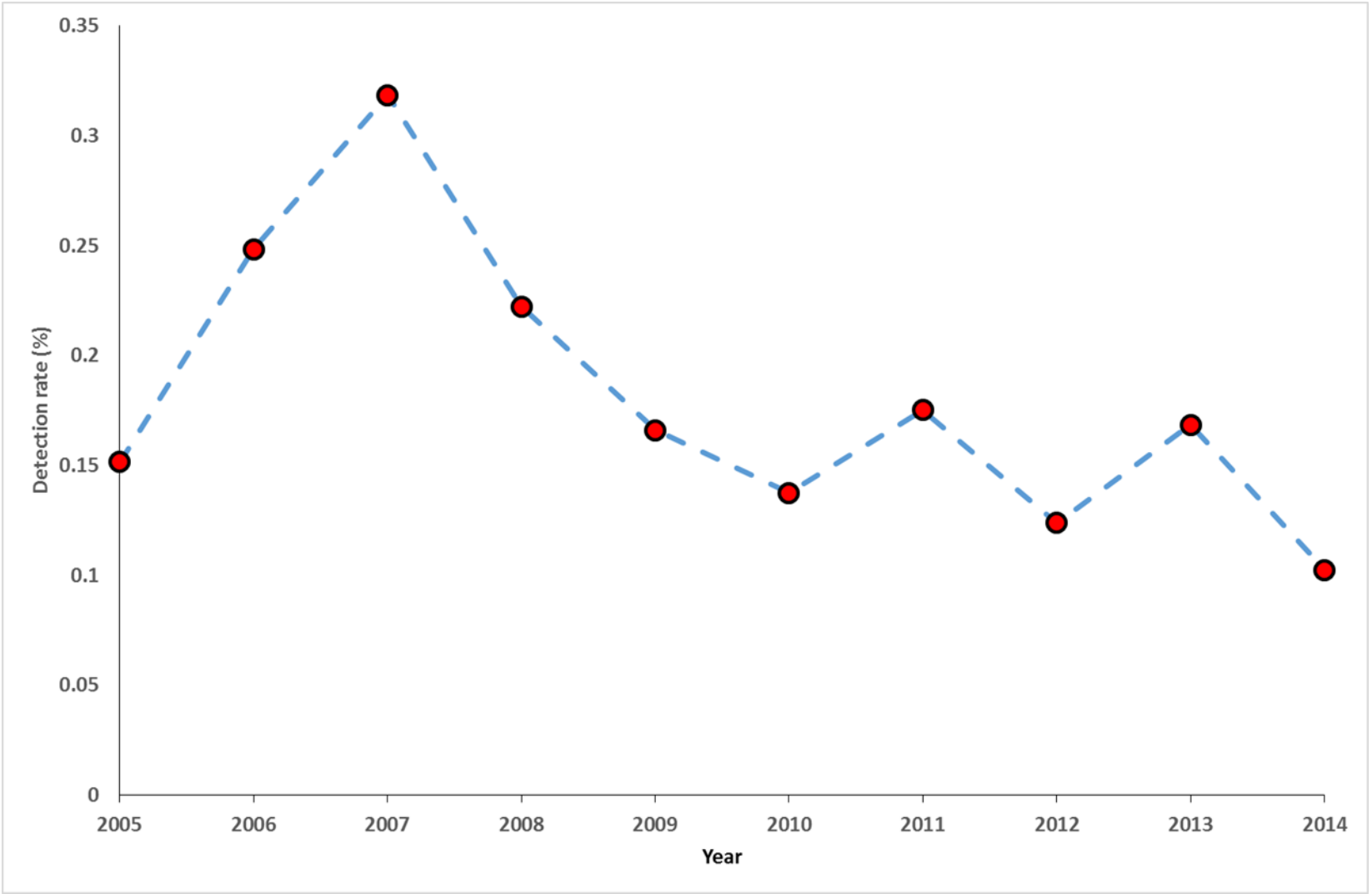


English groundwater – Detection rate

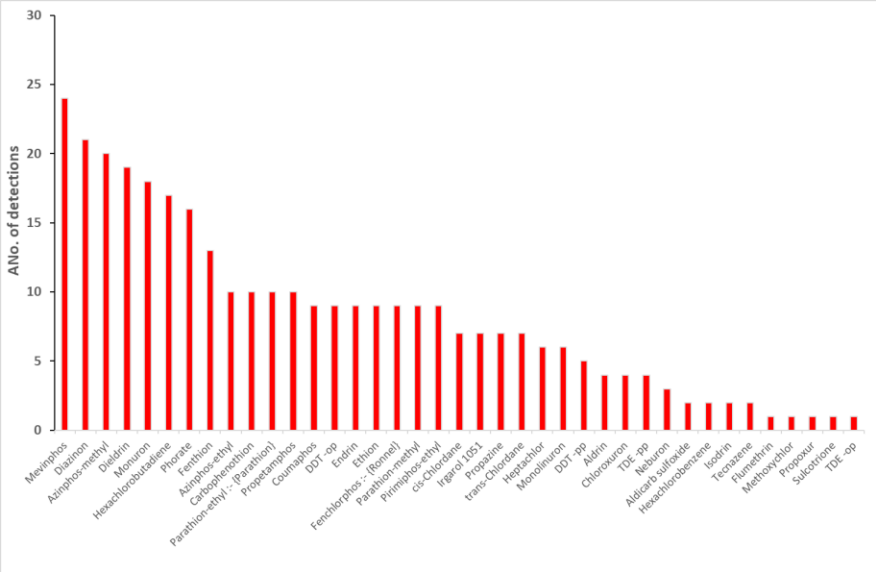


- 98 out 113 compunds detected above 0.1 µg/l
- Average detection rate was 0.19%
  - Detection rate declined over the period

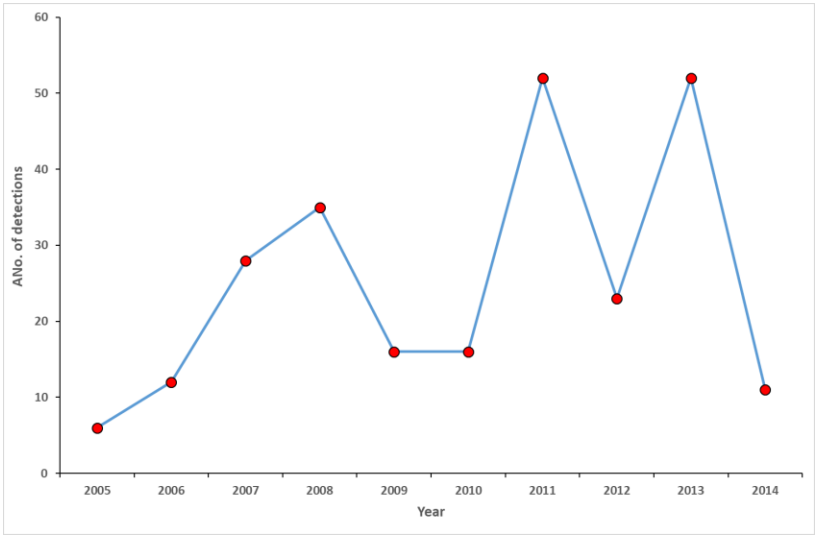
English groundwater – Detection rate



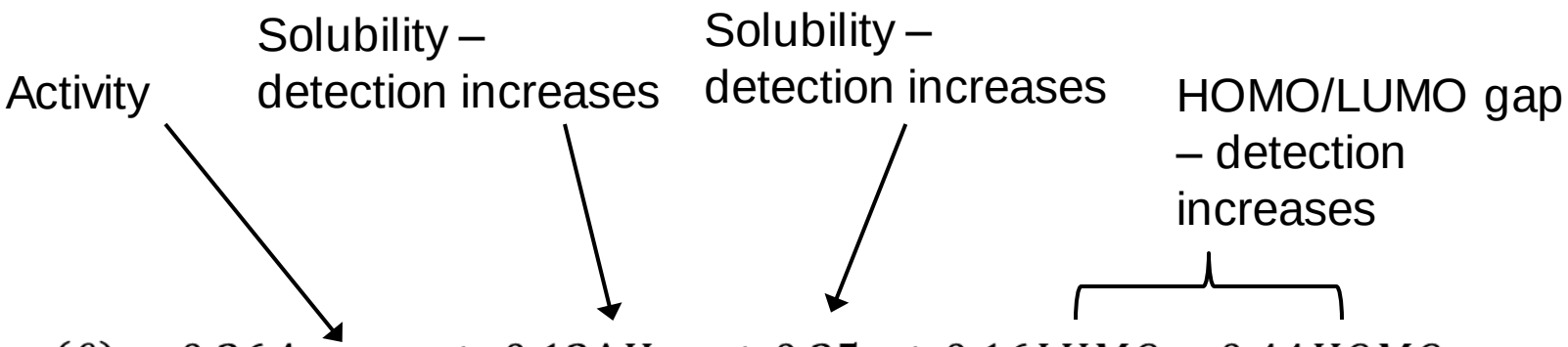
English groundwater – legacy detection rate



- 38 of the 98 detected were never applied during the period
  - 11.4% of all detections were these legacy pesticides
  - The detection rat efor these did not change



English groundwater – Detection rate

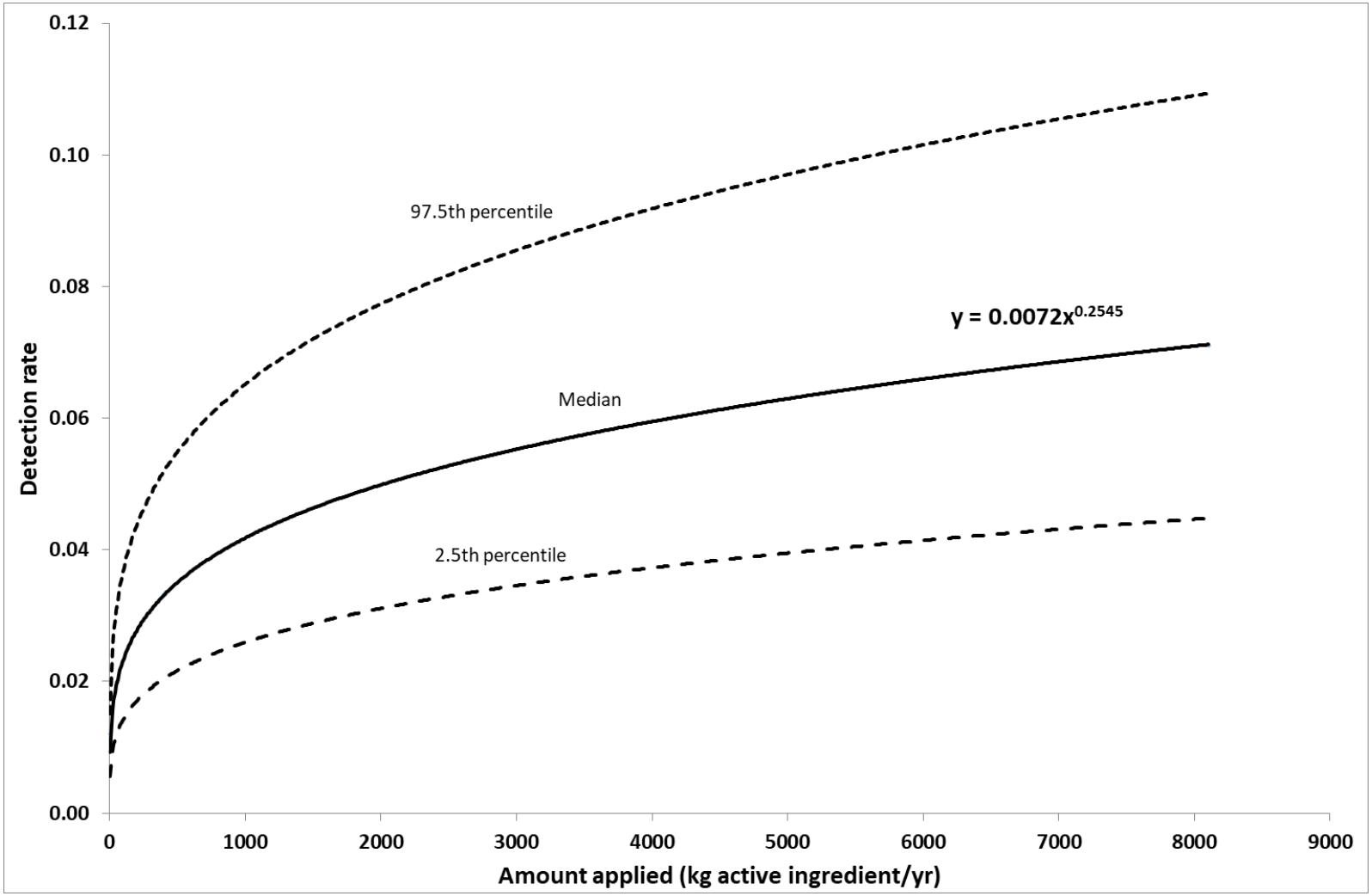


$$g(\theta) = 0.26Amount + 0.12\Delta H_{hyd} + 0.25\eta + 0.16LUMO - 0.44HOMO$$

$$rate\ of\ detection = bin(\theta, no.\ of\ observations)$$

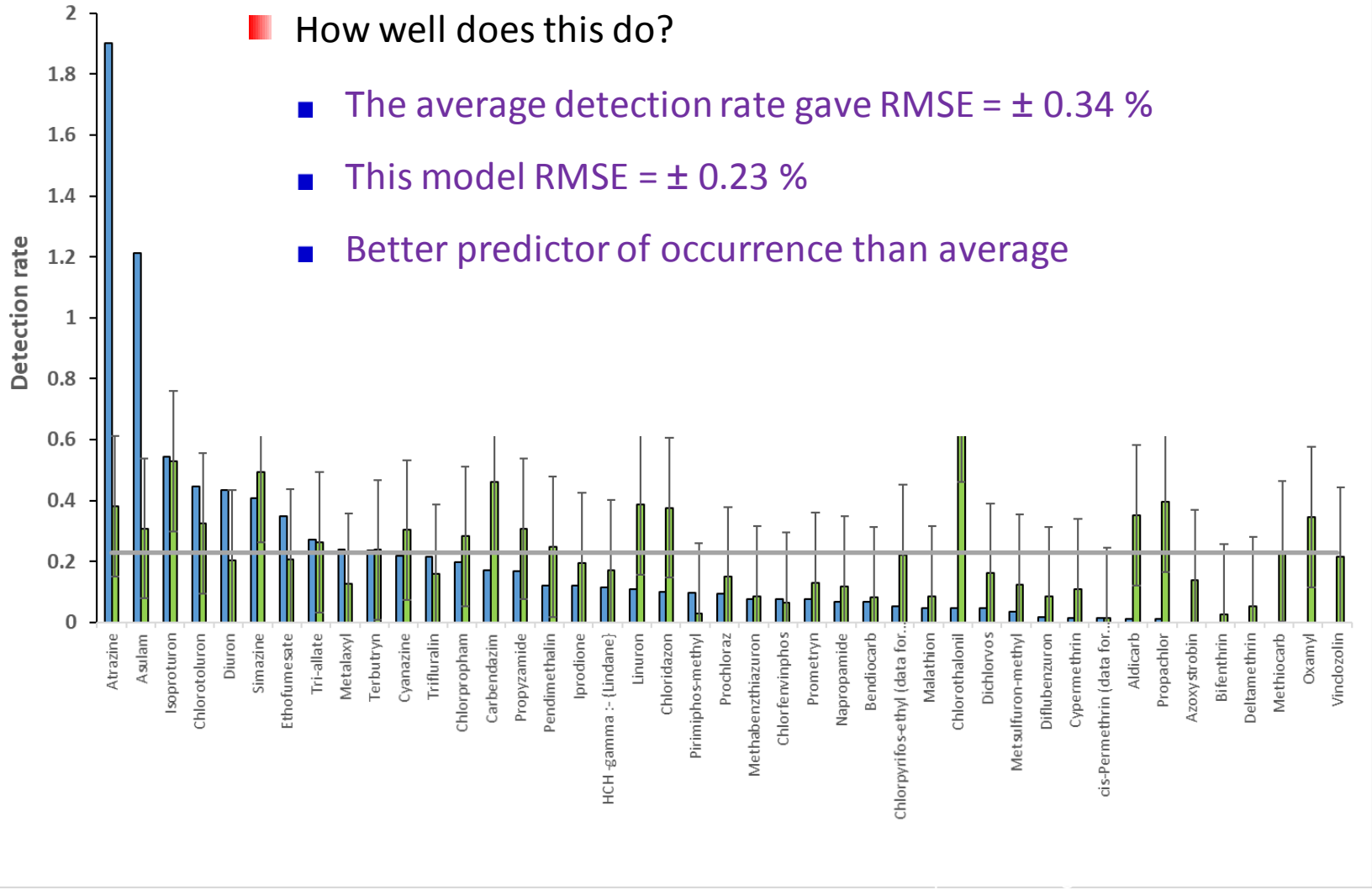
- What controls detection rate?
  - Detection rate not detection
  - Binomial regression
  - Loading
  - Solubility
  - degradability

English groundwater – Detection rate



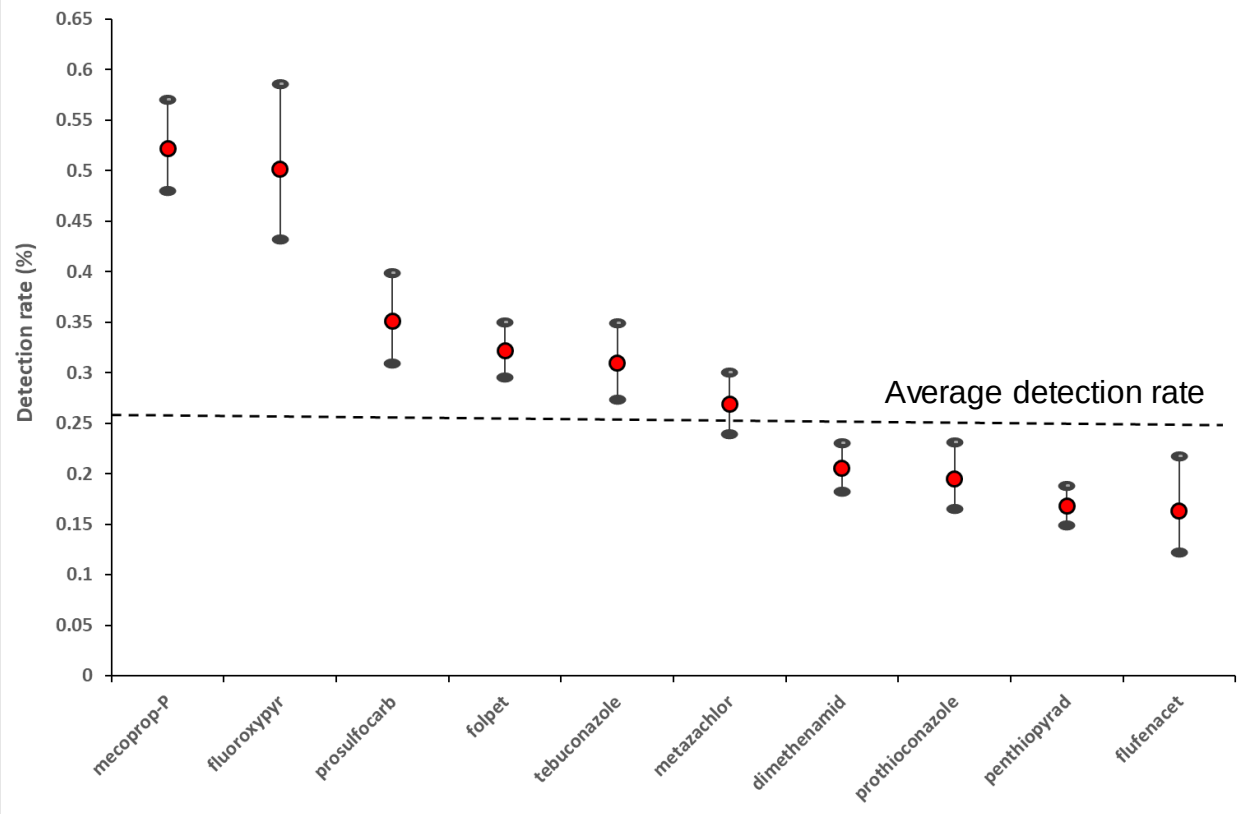
English groundwater – model performance

- How well does this do?
  - The average detection rate gave RMSE = ± 0.34 %
  - This model RMSE = ± 0.23 %
  - Better predictor of occurrence than average

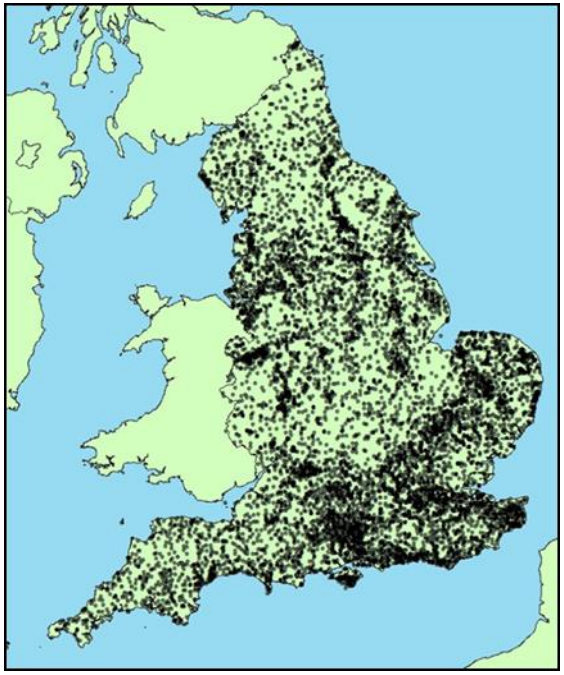


English groundwater – application

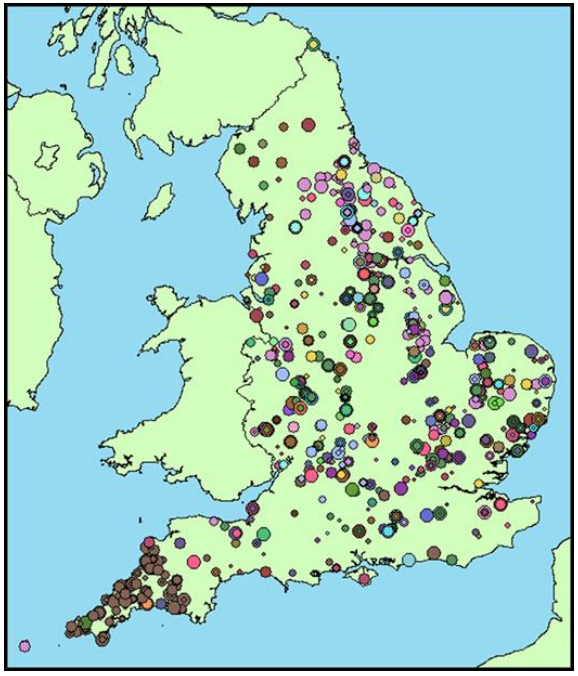
- Apply the model to the most used compounds never analysed for
  - More modern compounds are not necessarily less polluting



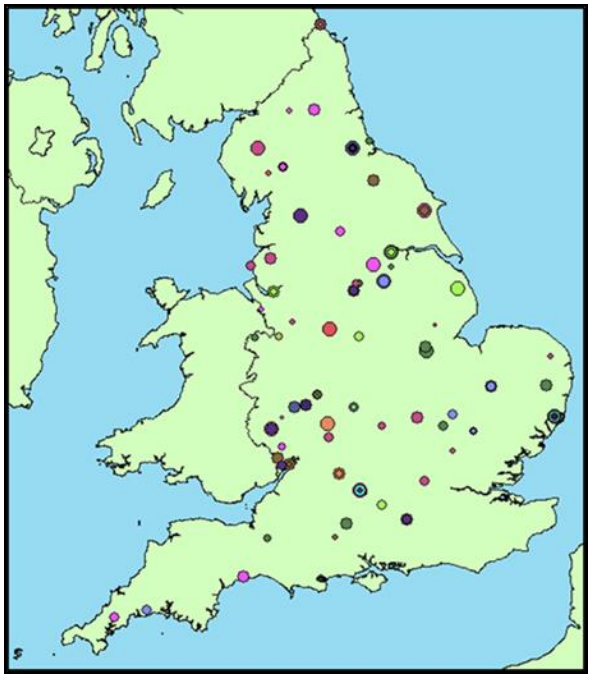
English groundwater - Vulnerability - f(site)



Locations of sampled boreholes



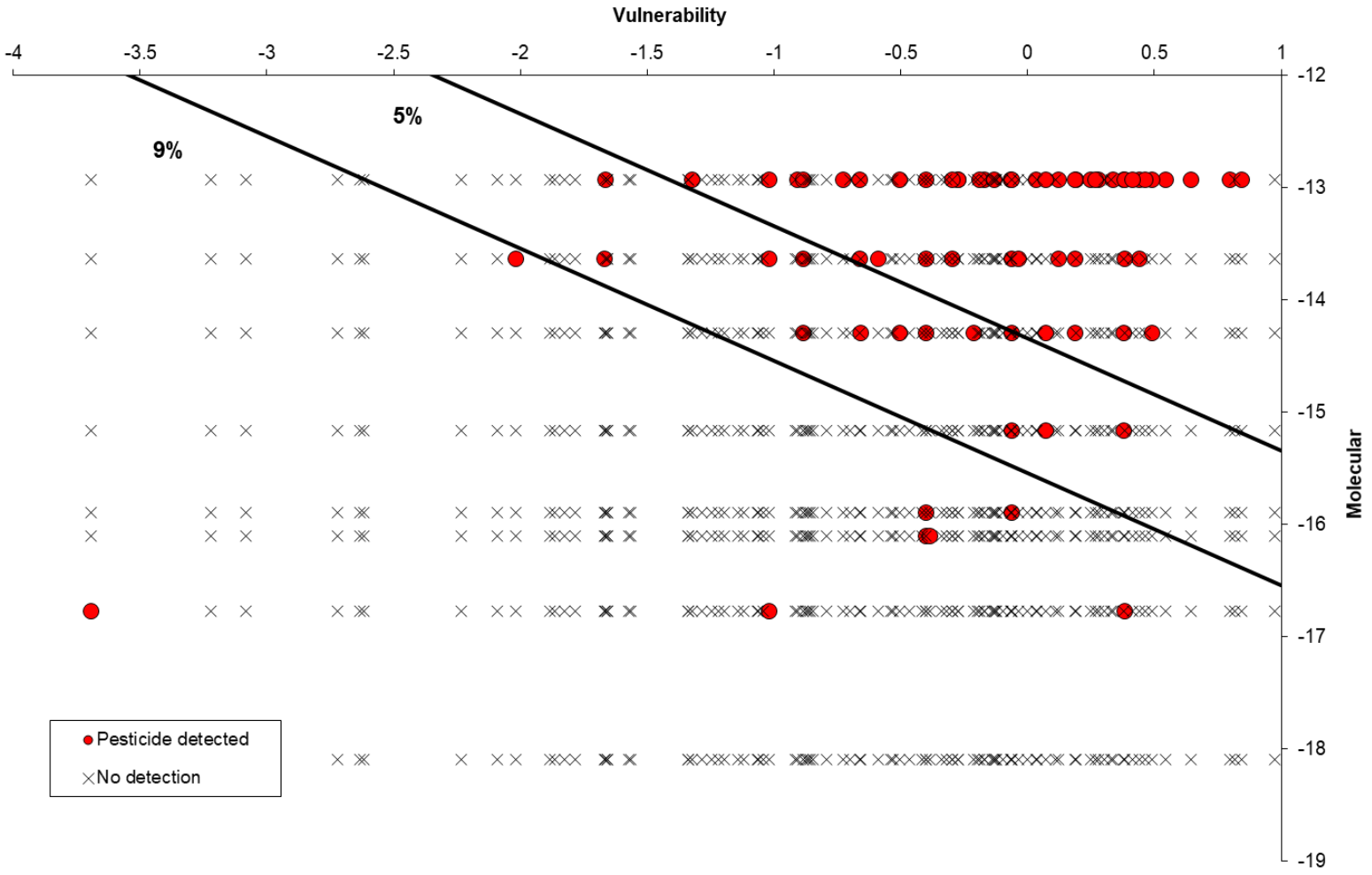
Locations of boreholes with detections



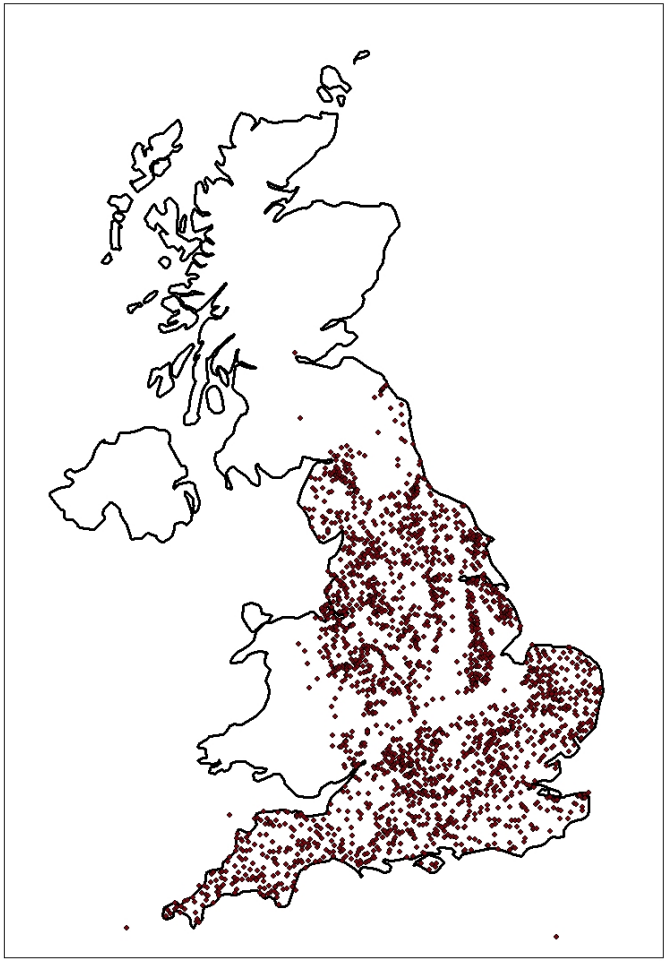
Locations of boreholes with legacy detections

- Not got to this yet.
  - Average 456 observation per borehole (1 to 3030)

Interaction – f(site x compound)



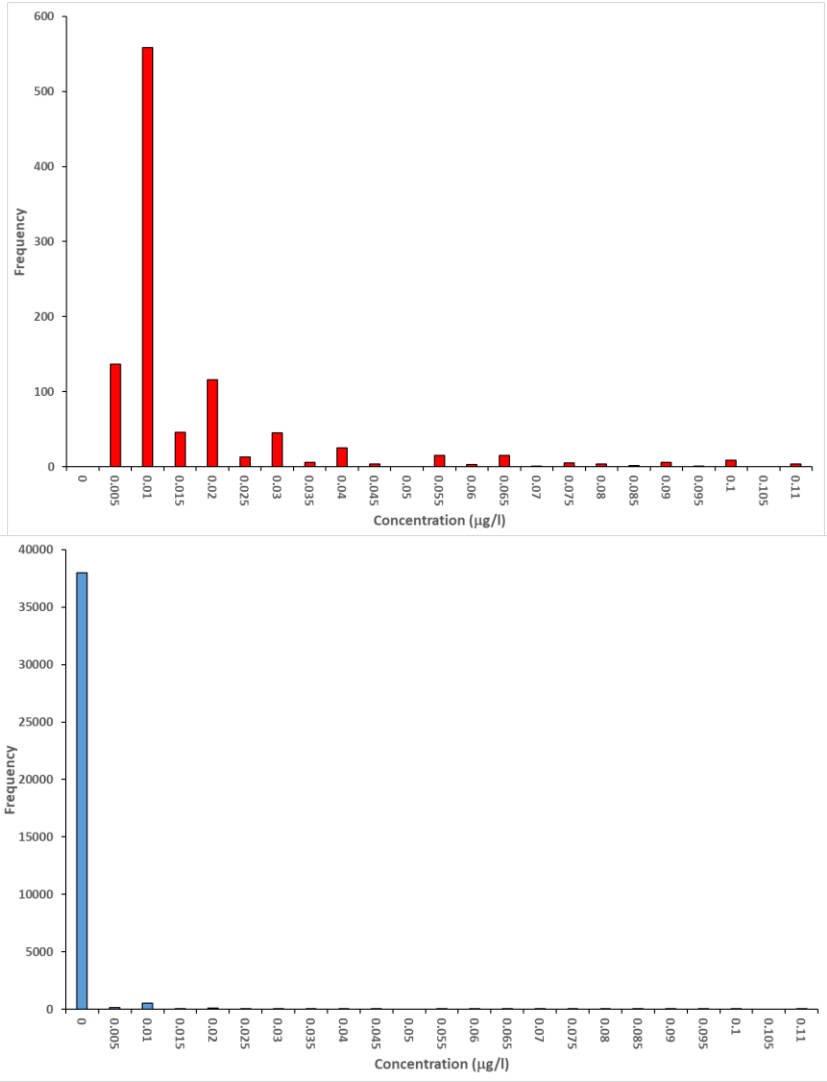
English groundwater – validation and next steps



- Independent dataset
- Very low detection
- Screens for very wide range of compounds
- EA has been using LC-MS on groundwater since 2009
  - 2415 locations
  - 39055 observations

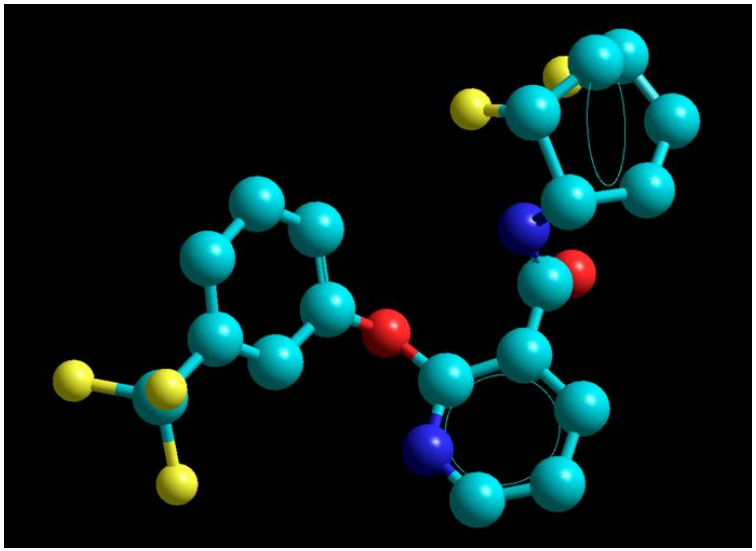
Future directions

- Tracking legacy compounds
  - Real persistence in the environment
- We want to predict concentration
- Distribution is dominated by zeros
  - Even allowing for truncated and censored data
  - But zero is a reasonable result
- Two part model
  - Logistic model
  - Weibull model



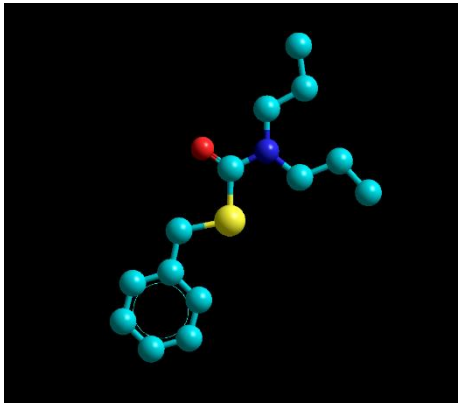
Conclusions

- All achieved on free data
- The developing hierarchy of generalised linear models
  - Logistic models
    - Molecular model
    - Vulnerability model
    - Interaction model
  - Binomial regression
    - Molecular model
- Models have physical interpretation
  - Activity
  - Solubility
  - degradation

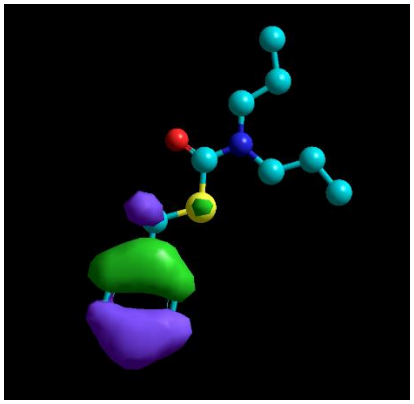


Could we do this?

New compound



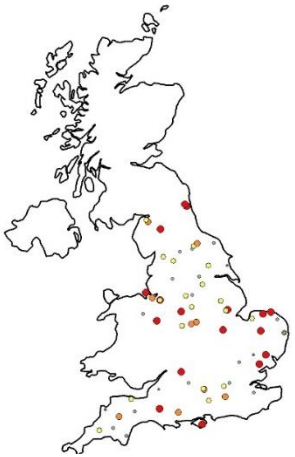
Molecular model



Screened out

Screening locations

Vulnerability model



What concentration? Where?