Tackling The Reproducibility Crisis In Ecology

ESA Annual Meeting 2024

Peter Levy, UK Centre for Ecology & Hydrology

2024-08-07

Outline

1. The Reproducibility Crisis - what is it?
2. Causes: the information architecture of science
3. Causes: high “false discovery” rates
4. Ecological examples
5. Solutions: a change in perspective in statistical thinking

The Reproducibility Crisis

Reproducibility
of results,
not methods.

Better name:
"*Credibility* crisis"?

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful. "Negative" is actually a misnomer, and

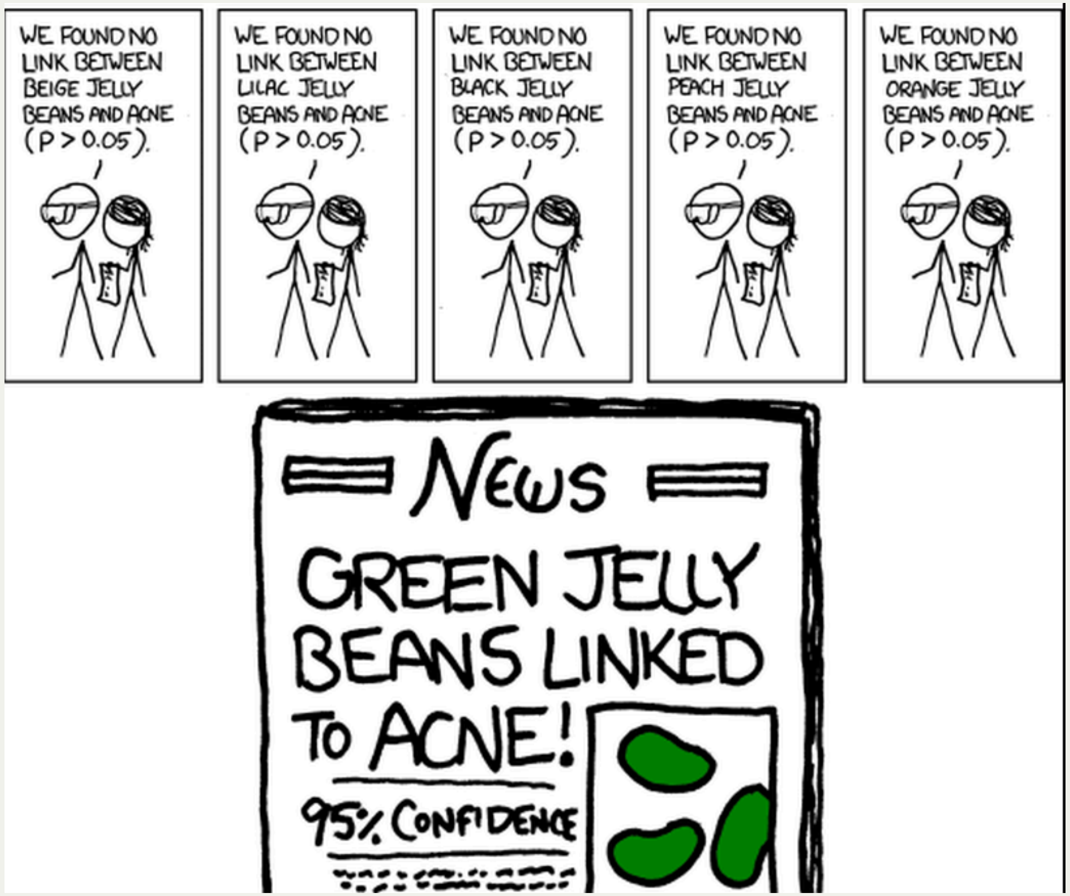
is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R+1)$. The probability of a study finding a true relationship reflects the power $1-\beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2 × 2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2 × 2 table, one gets PPV = $(1-\beta)R/(R$

Ioannidis, 2005, PLOS Medicine

The information architecture of science

“Why most published research findings are false”

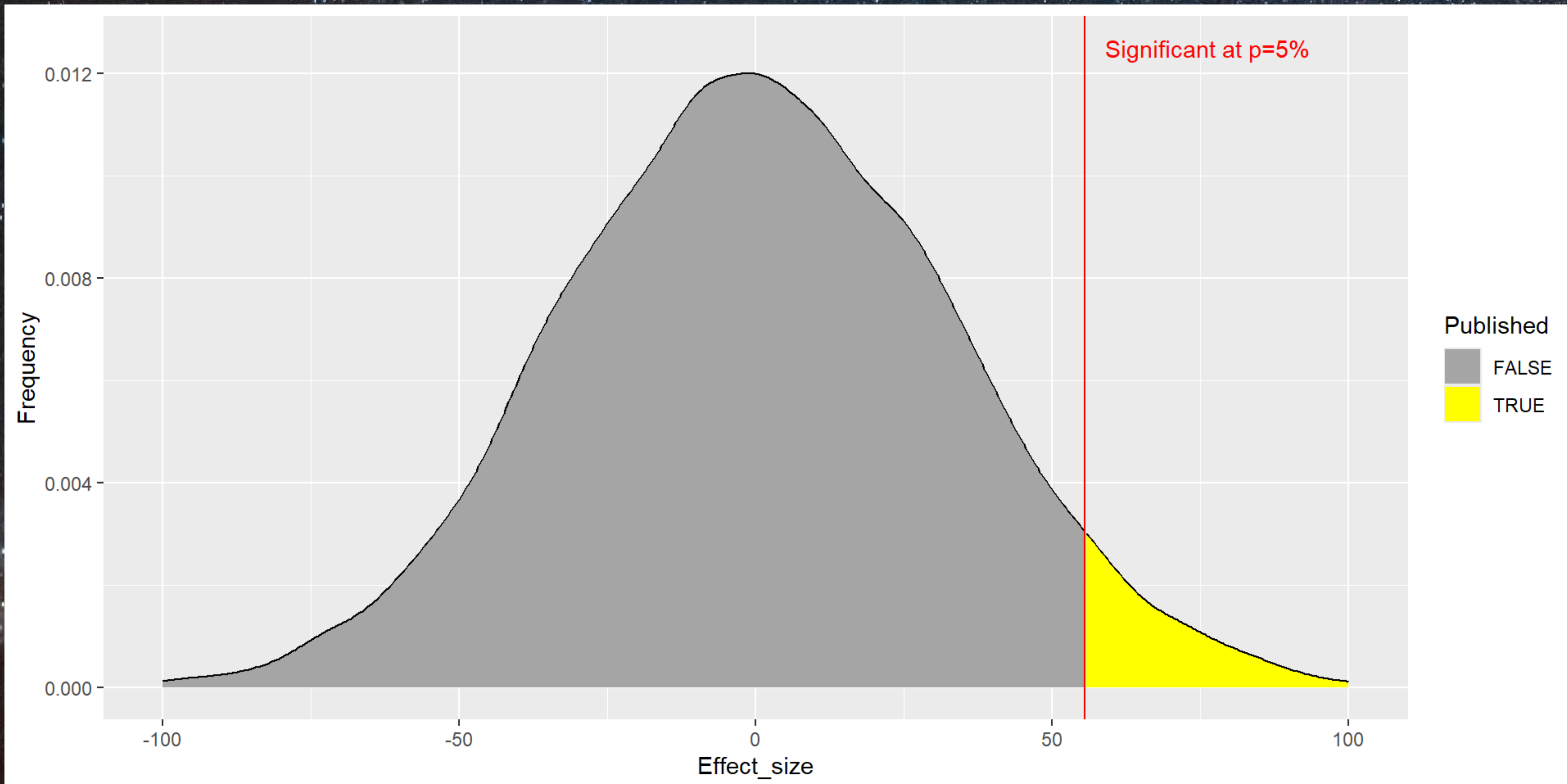
- The multiple testing problem
- What if we only publish the 5% studies with significant results?



"Why most published research findings are false"

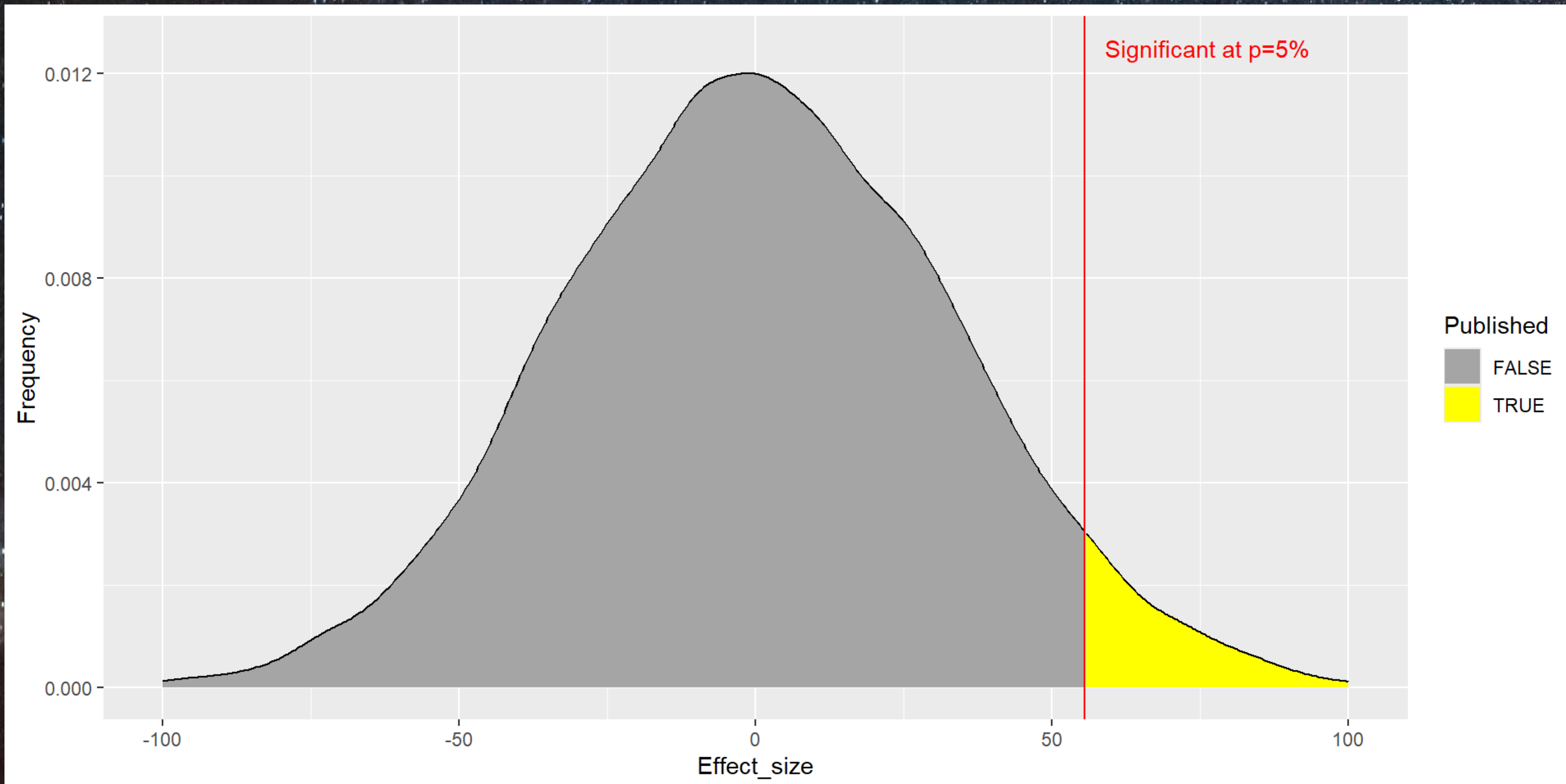
- "*The garden of forking paths*"
  - many possible choices in data collection & analysis
  - choices favour finding *interesting* patterns
- Motivated cognition
  - "the influence of motives on memory, information processing & reasoning."
- Journals reject negative or confirmation results

Unpublished work as Dark Matter



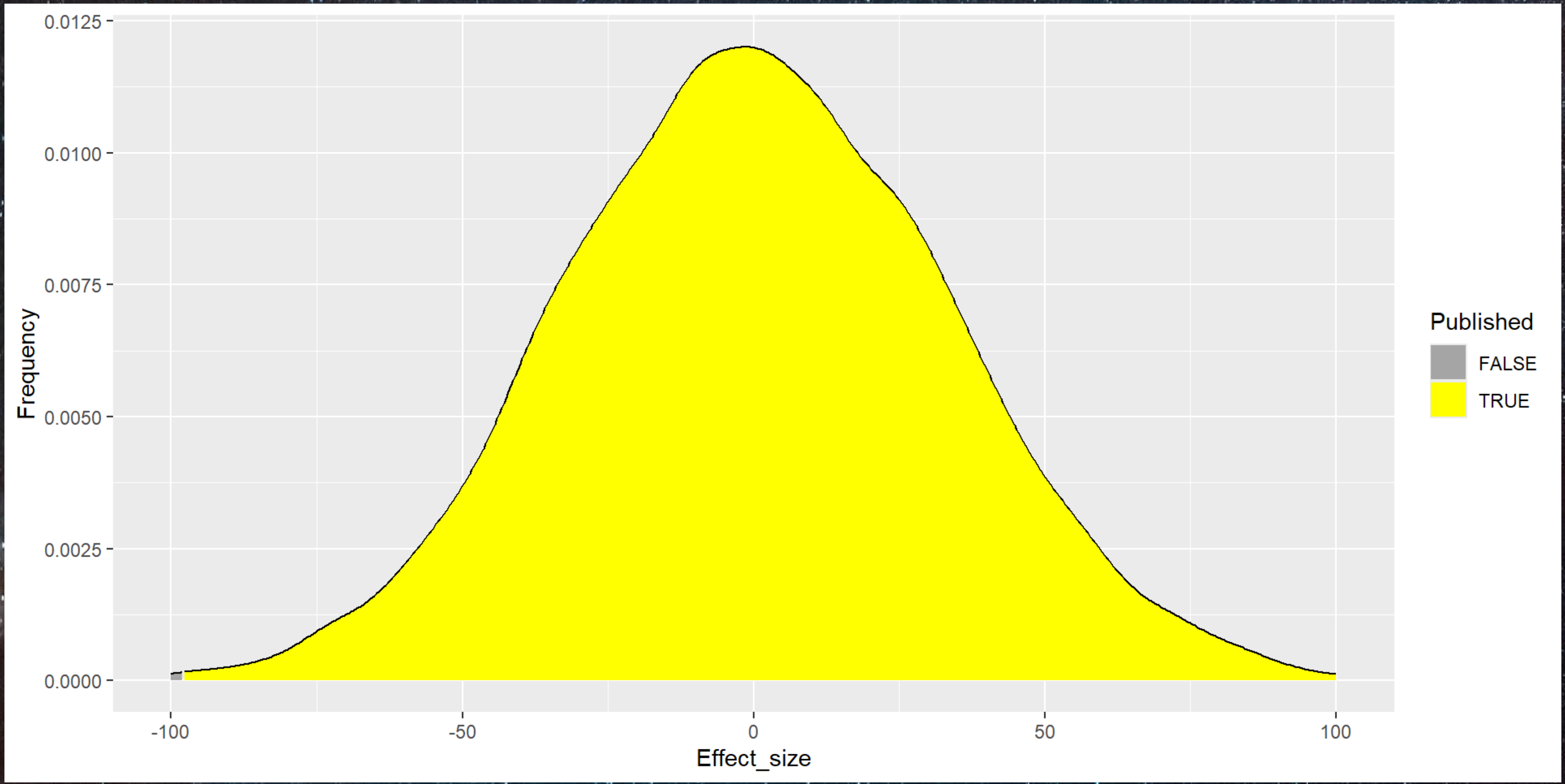
We can only see 5% of the universe.

Solutions



Don't focus on testing a null hypothesis.

Solutions



Pre-register and publish all results.

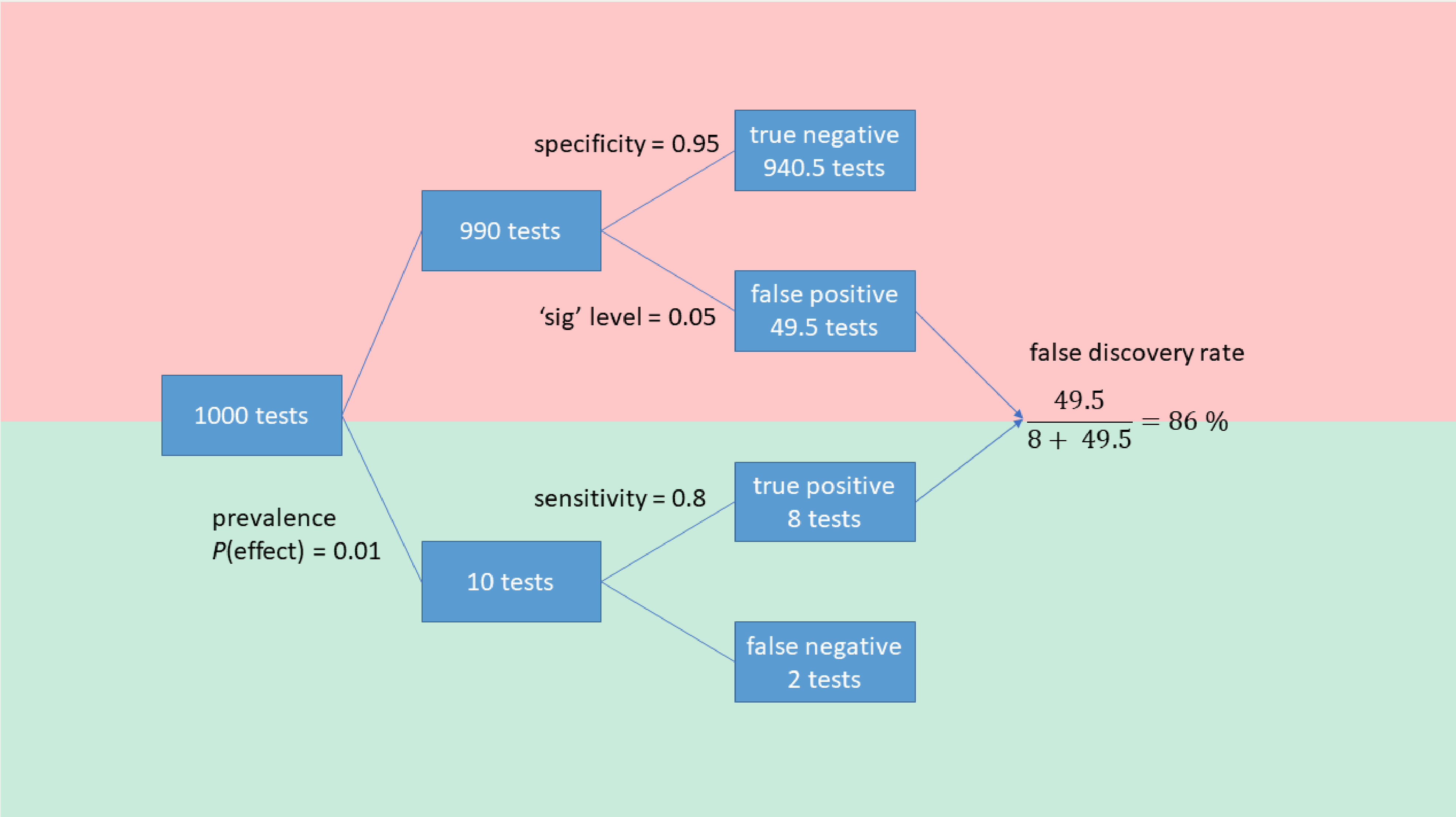
False discovery rates in ecology

“Why most published research findings are false”

1. Low prior probabilities - unlikely / rare effects
2. Low statistical power - low signal:noise
3. Bias - systematic uncertainties in observation process

-> High “false discovery rates”, often much higher than 5 %

1. Low prior probabilities



UK CEH

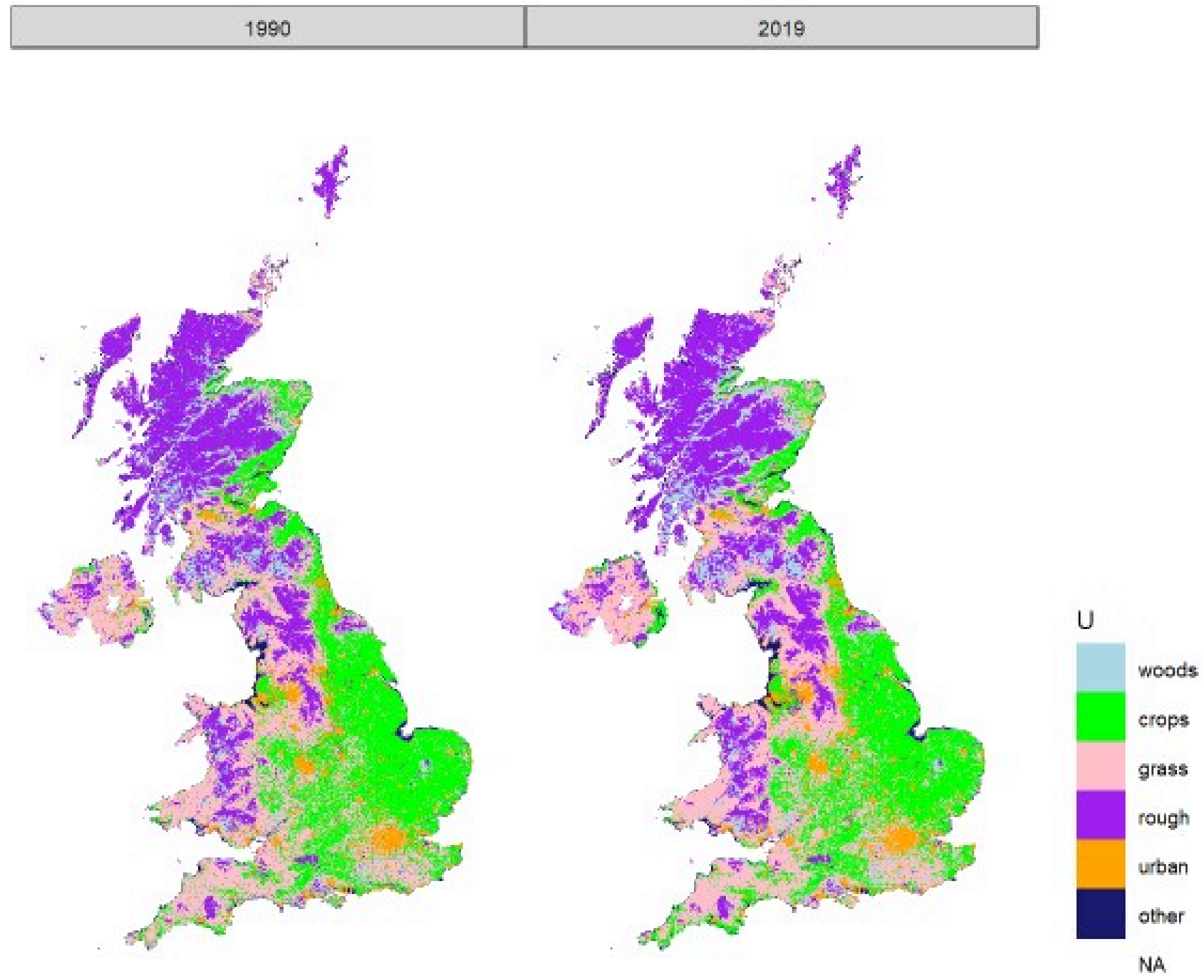
Land-Use Change: just the difference between maps at two times?

Yes, if:

- maps are accurate
- magnitude of change is large

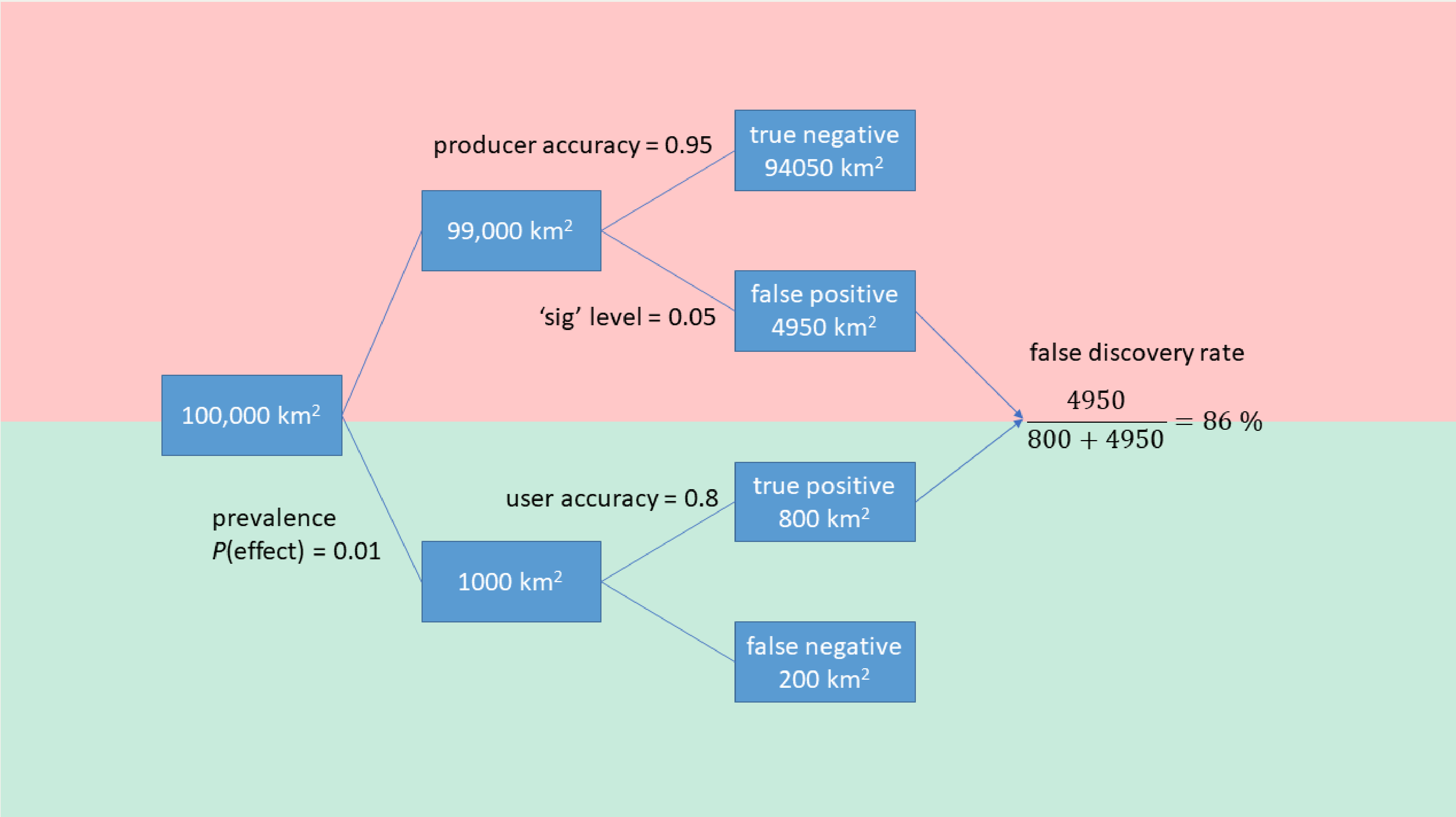
In UK:

- map classification accuracy is ~90%
- magnitude of change is small (<0.5%)



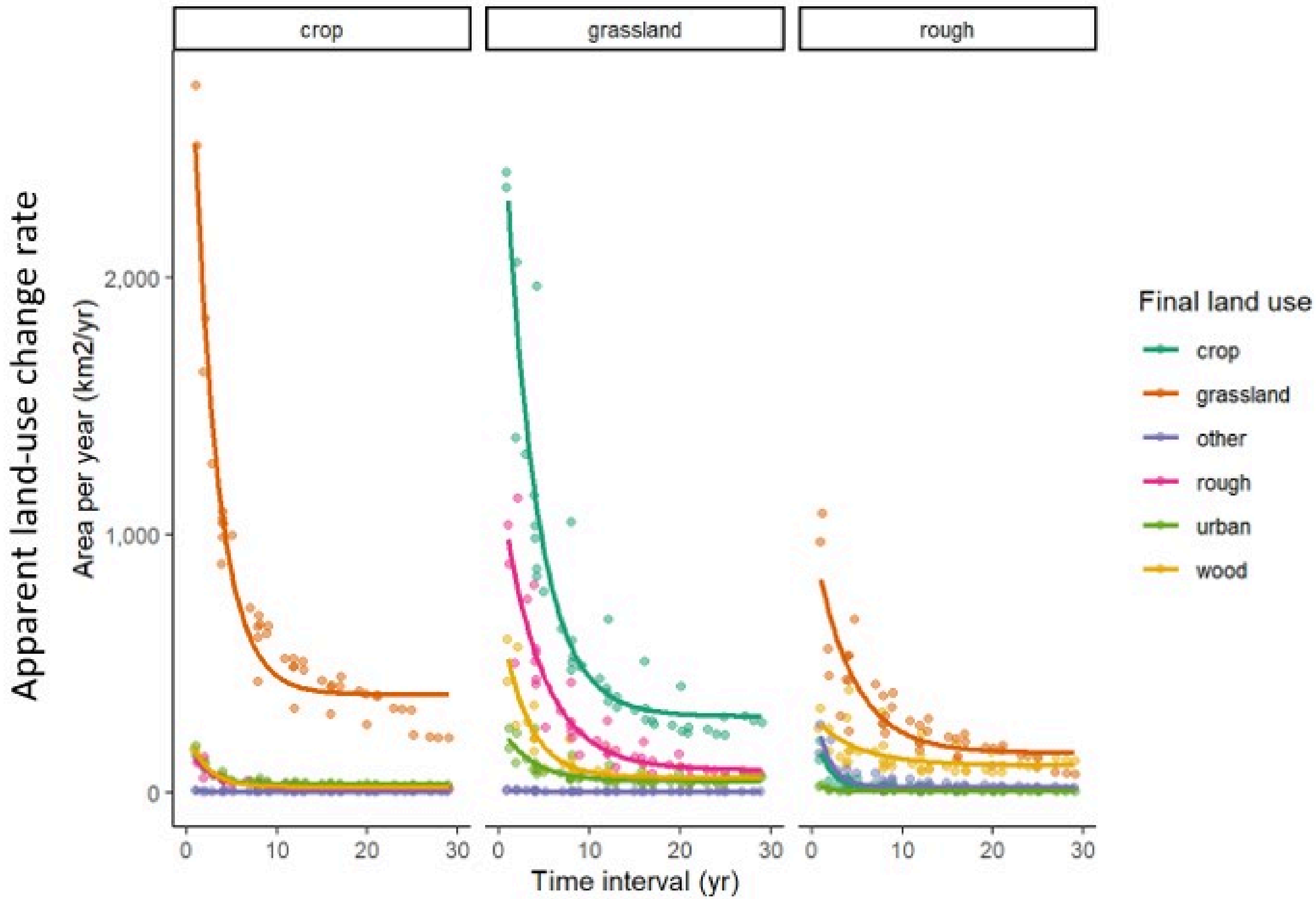
UK CEH

Example: Land-use change

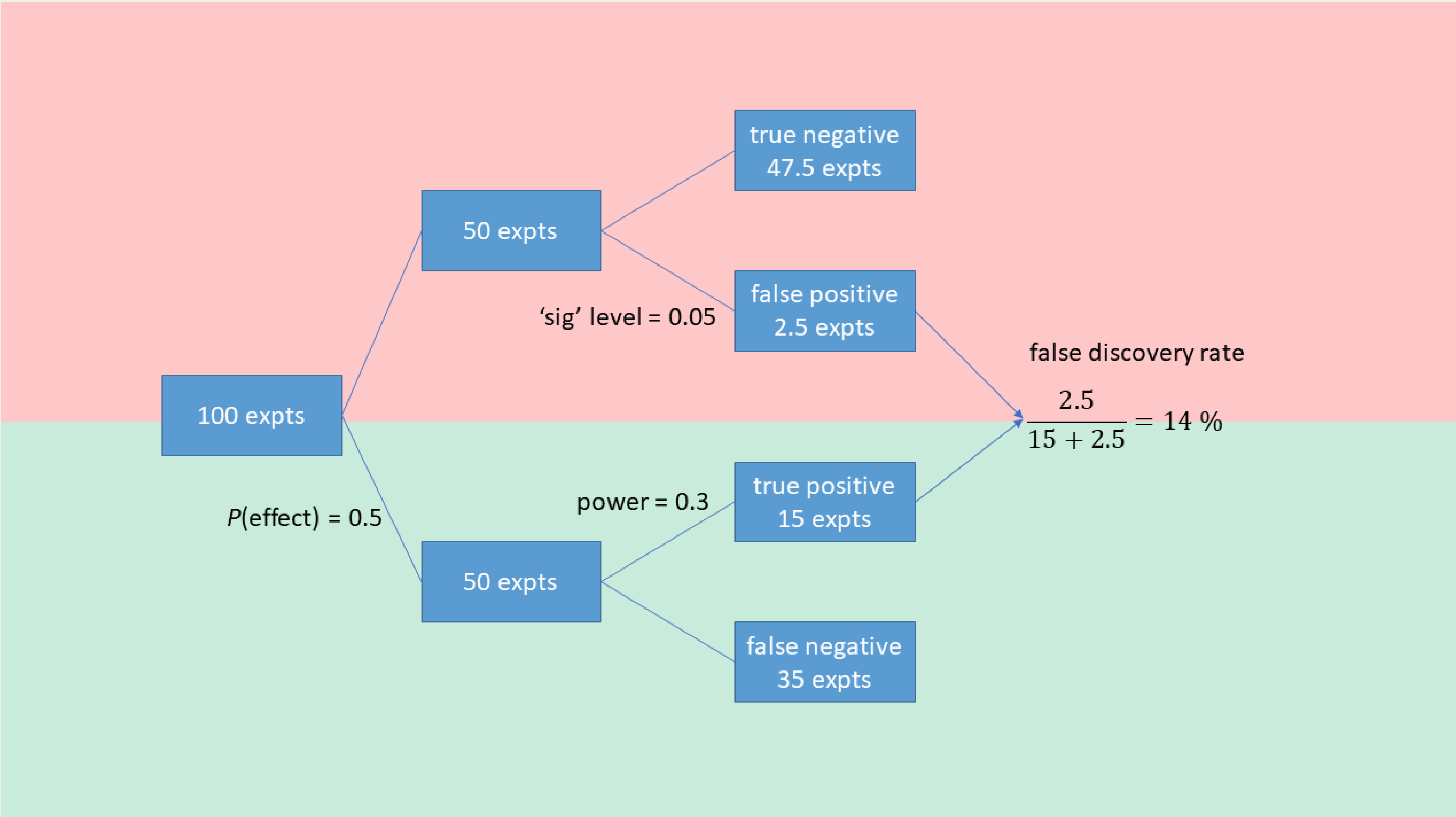


Empirical evidence from LCM analysis

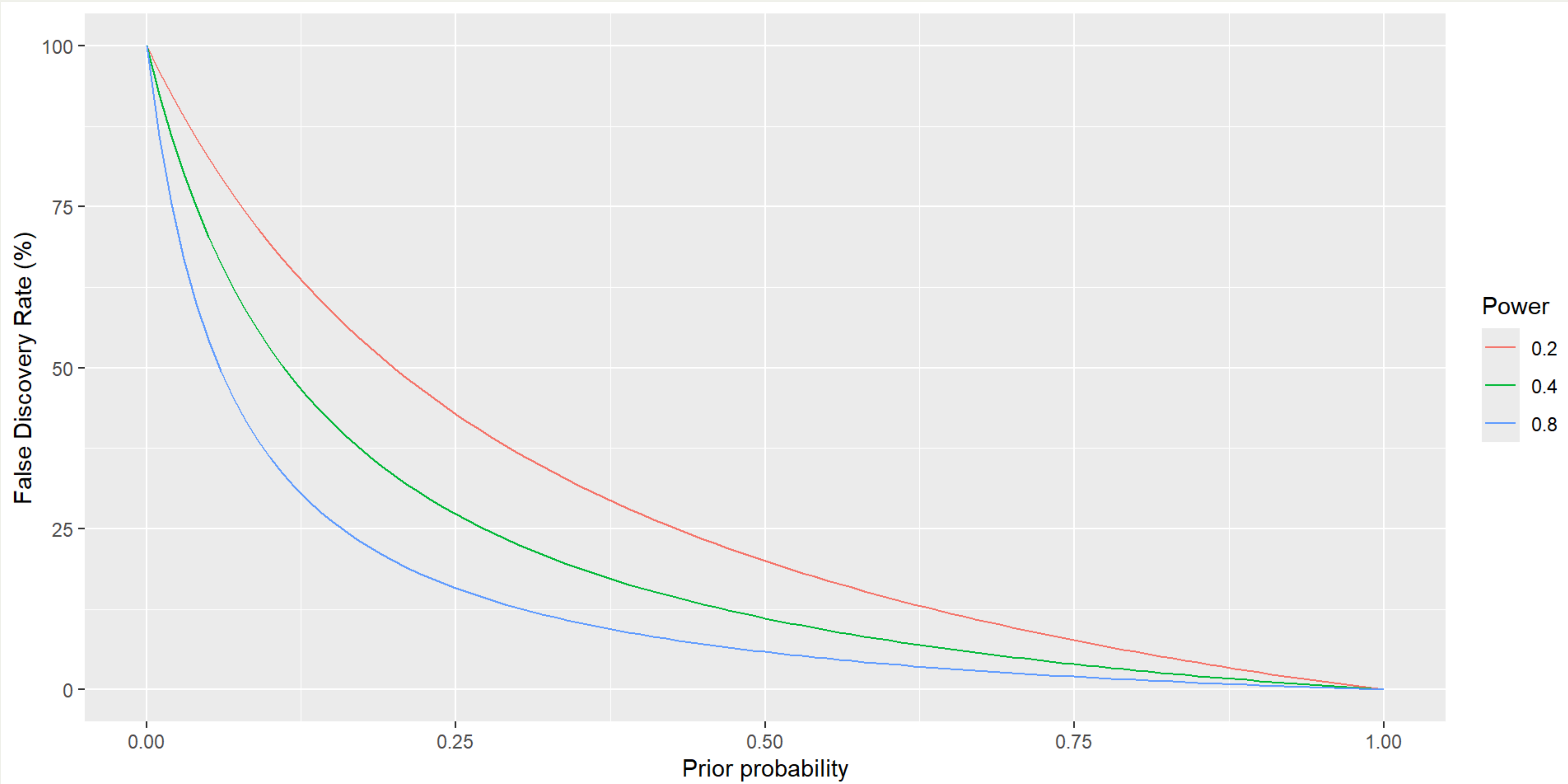
Expect horizontal line if false positive rate = 0



2. Low statistical power



2. Effect of low power



2. Low power in ecology

Measurements are a proxy for true process of interest.
Connection between the two can be:

- sample from population (sampling error)
- imperfect measurements (measurement error)
- proxy variables
  - true variable is hard to measure but can be approximated

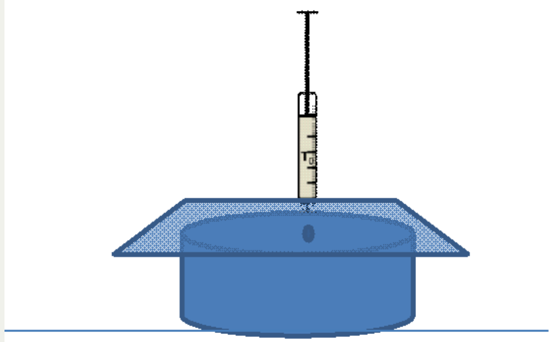
UK CEH

Example: gas emissions from soil

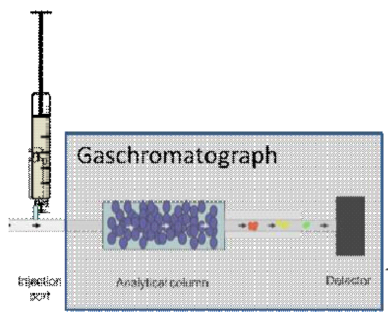
Measuring N_2O emissions

Sample collection

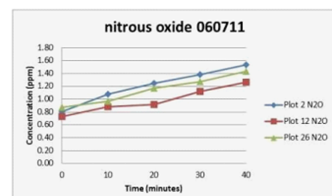
(field)



(laboratory)



Analysis
Flux calculation

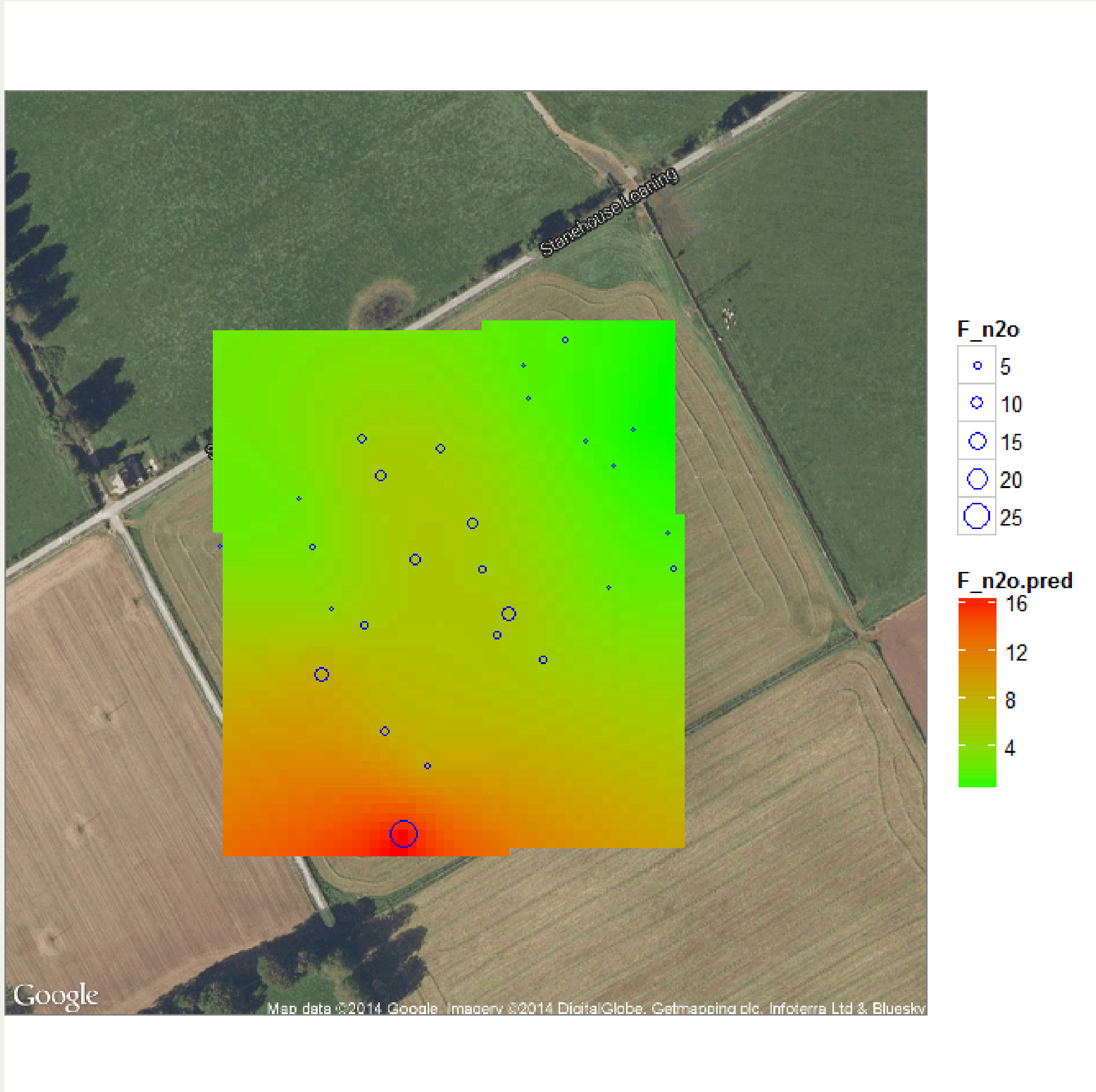


$$F = \frac{\mathrm{d}C}{\mathrm{d}t_0}\frac{\rho V}{A}$$

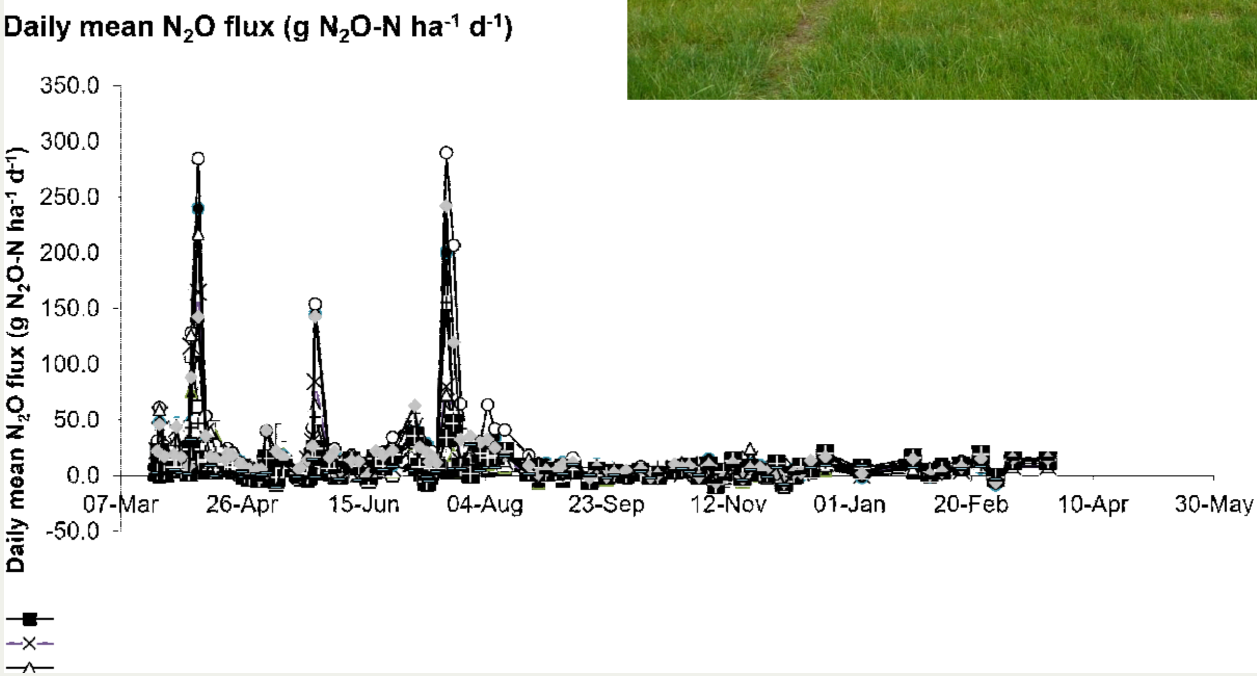
5

UK CEH

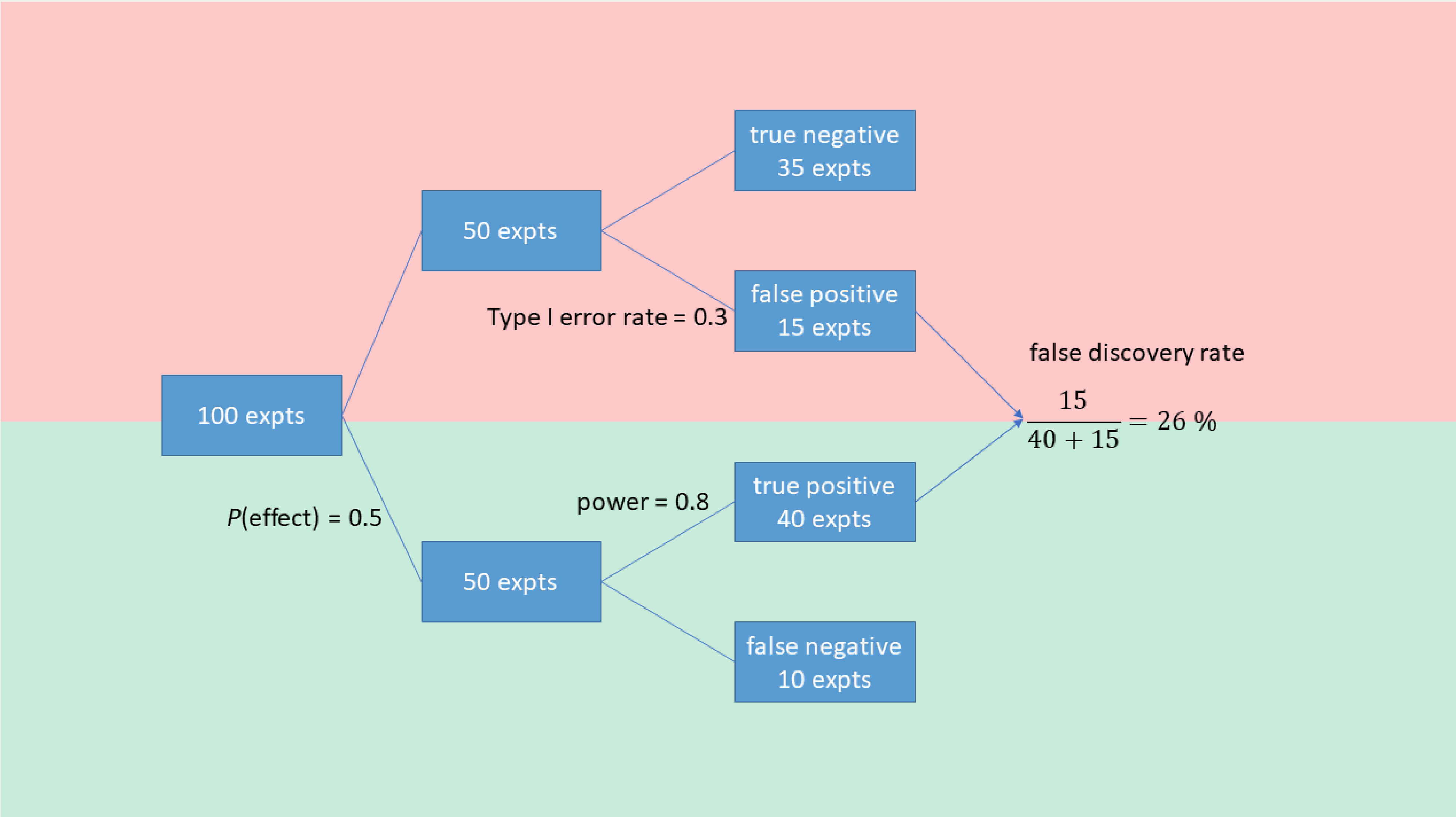
Spatial variation in gas emissions



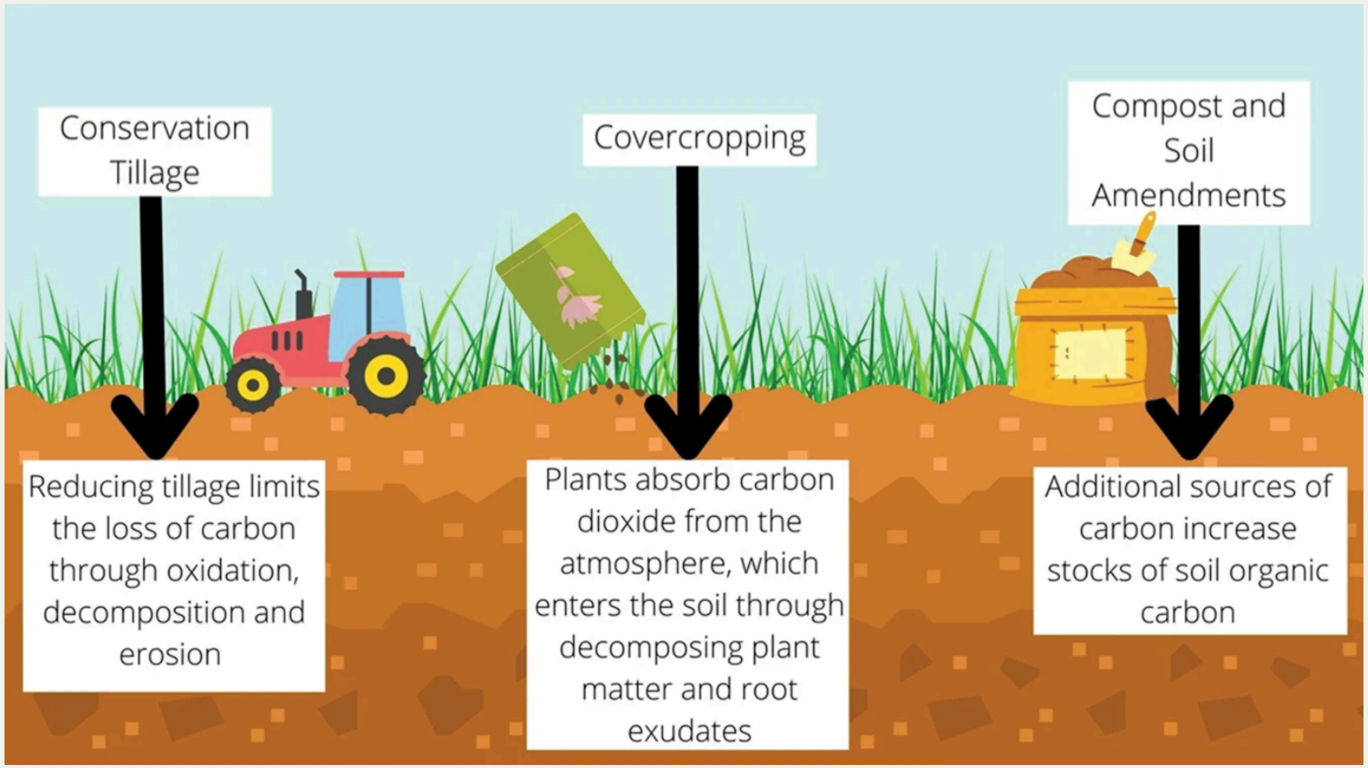
We need cumulative emissions

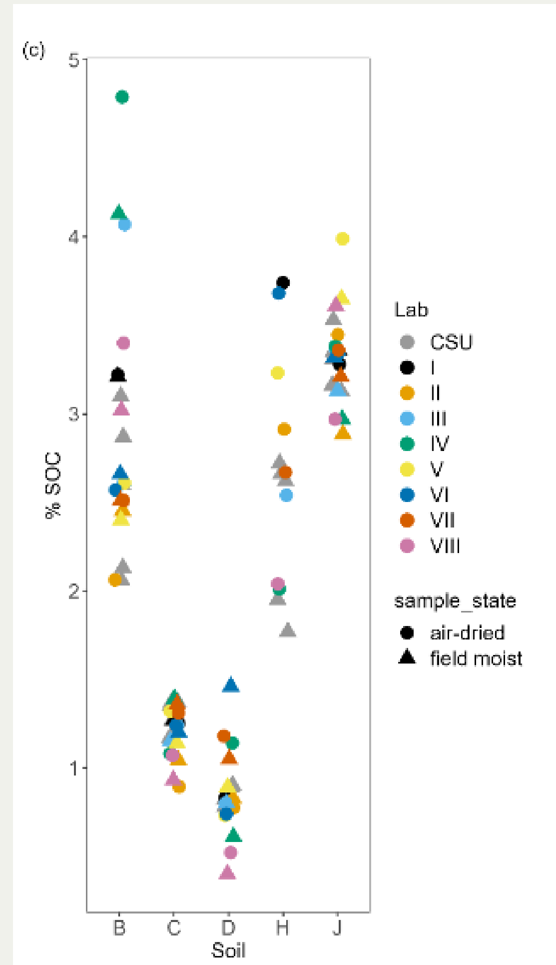


3. Bias in observation process

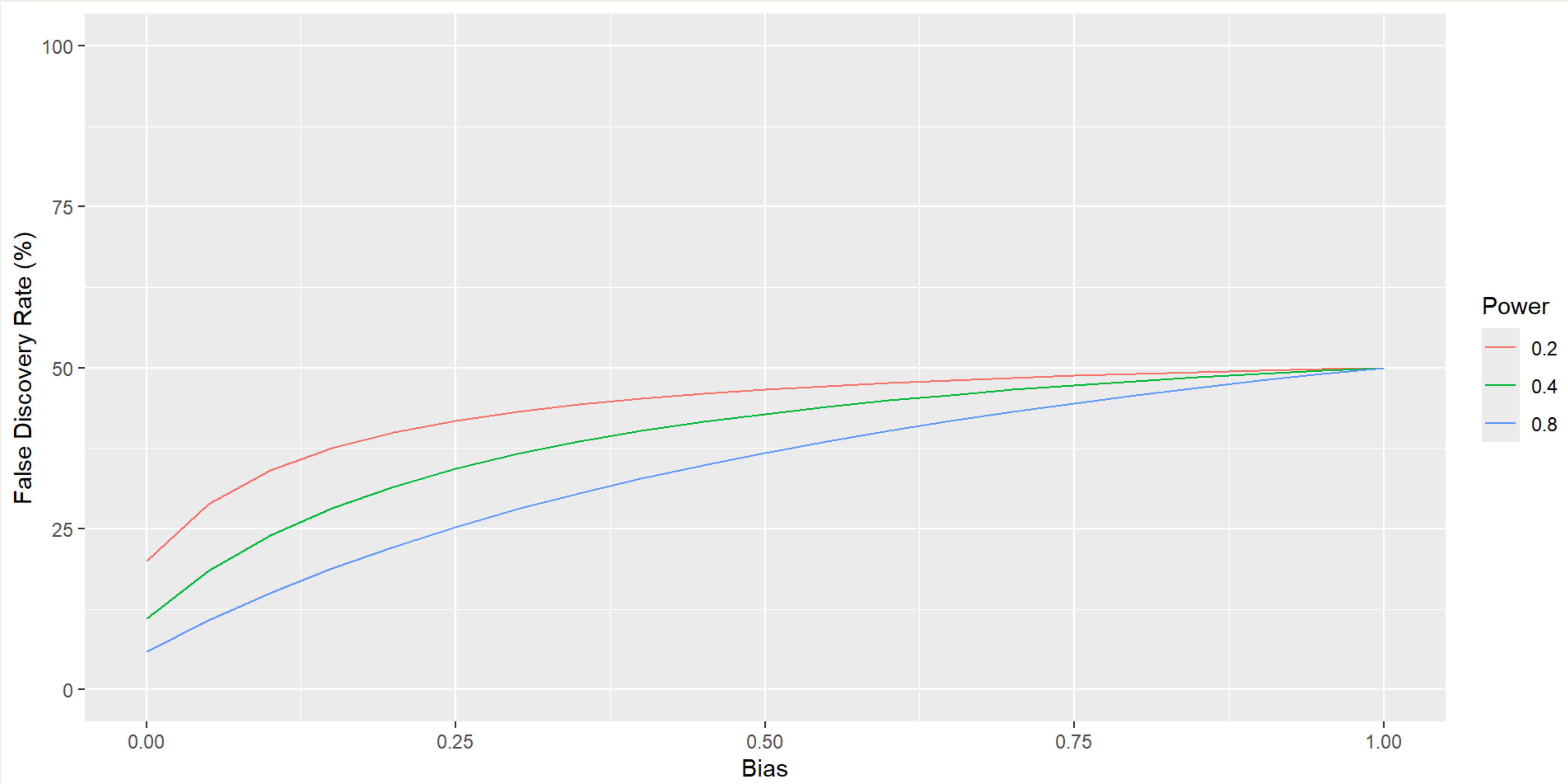


Example: soil carbon change





3. Effect of bias



The ASA statement

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: www.tandfonline.com/journals/utas20

The ASA Statement on *p*-Values: Context, Process, and Purpose

Ronald L. Wasserstein & Nicole A. Lazar

To cite this article: Ronald L. Wasserstein & Nicole A. Lazar (2016) The ASA Statement on *p*-Values: Context, Process, and Purpose, The American Statistician, 70:2, 129-133, DOI: 10.1080/00031305.2016.1154108

2016

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: www.tandfonline.com/journals/utas20

Moving to a World Beyond "$p < 0.05$"

Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar

To cite this article: Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar (2019) Moving to a World Beyond "$p < 0.05$", The American Statistician, 73:sup1, 1-19, DOI: 10.1080/00031305.2019.1583913

To link to this article: https://doi.org/10.1080/00031305.2019.1583913

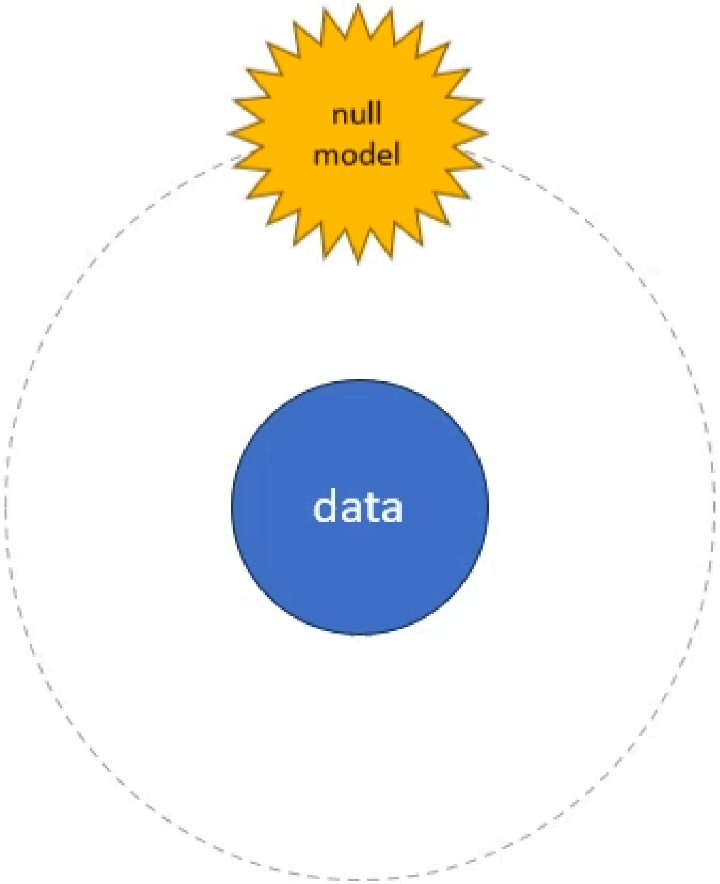
2019

Stop using *p* values and "statistical significance".

Solutions

A Copernican Revolution

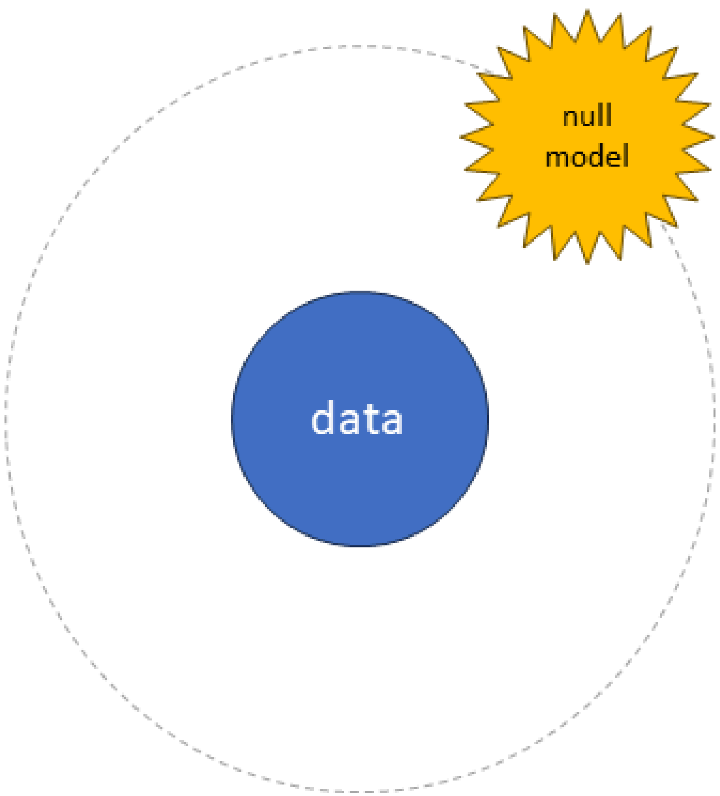
Conventional statistics



Data is fixed
Centred on P[data | null model]

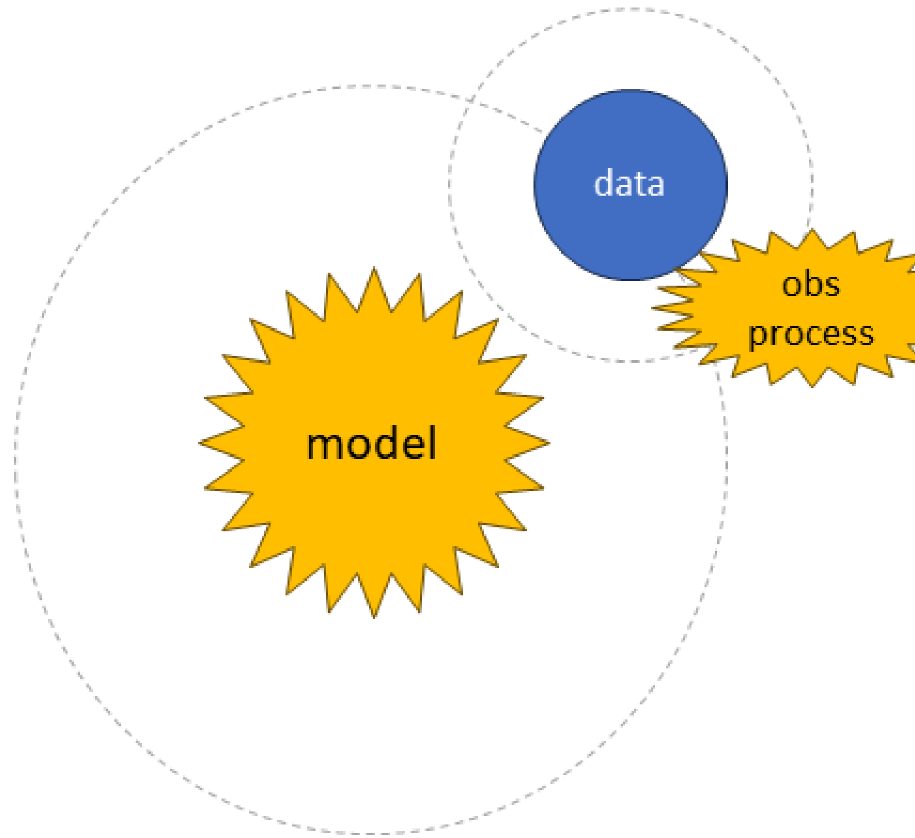
A Copernican Revolution

Conventional statistics



Data is fixed
Centred on P[data | null model]

Conditional probability (Bayes rule)



Data & observation process are uncertain
Centred on P[model, obs process | data]

In practice

- don’t take data at face value
- represent & propagate their uncertainties
- represent potential biases
- often goes beyond off-the-shelf stats model
- suits Monte Carlo or fully Bayesian approach
- examples …

Example: Land-Use Change

Solution: Hidden Semi-Markov Model

The true sequence of states is **hidden.**

We want to know the probability P(true | observed)

Observations

time

1	2	3	4	5
forest	forest	grass	grass	crop

True (hidden) state

forest	?	?	?	?

Solution: Hidden Semi-Markov Model

The true sequence of states is **hidden.**

We want to know the probability P(true | observed) given:

- the transition probability matrix

From \ To	Forest	Crop	Grass
Forest	0.97	0	0.03
Crop	0.02	0.8	0.18
Grass	0.02	0.08	0.9

Observations

time

1	2	3	4	5
forest	forest	grass	grass	crop

True (hidden) state

forest	?	?	?	?

Solution: Hidden Semi-Markov Model

The true sequence of states is **hidden.**

We want to know the probability P(true | observed) given:

- the transition probability matrix
- the observation likelihoods P(observed | true)

From \ To	Forest	Crop	Grass
Forest	0.97	0	0.03
Crop	0.02	0.8	0.18
Grass	0.02	0.08	0.9

True	Prob. of observing
Forest	0.9
Crop	0.5
Grass	0.7

Observations

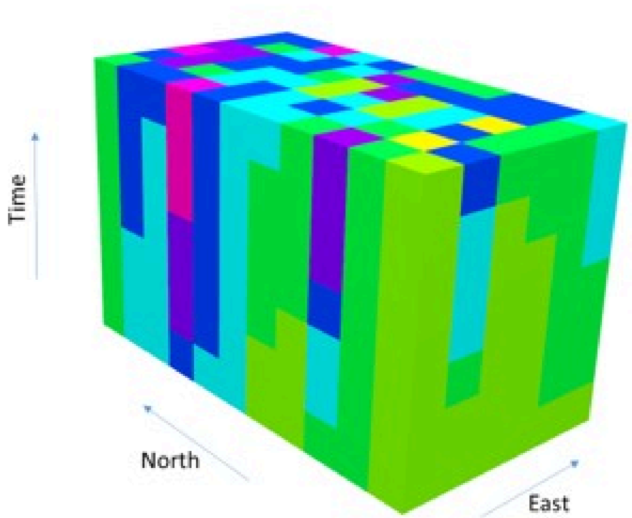
time

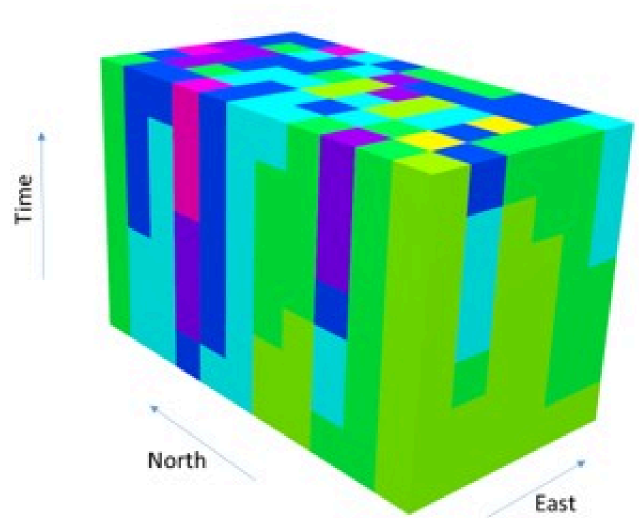
1	2	3	4	5
forest	forest	grass	grass	crop

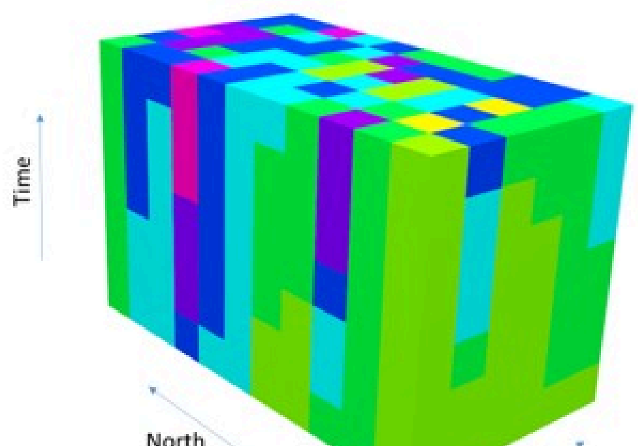
True (hidden) state

forest	?	?	?	?

Posterior distribution of LUC data cubes



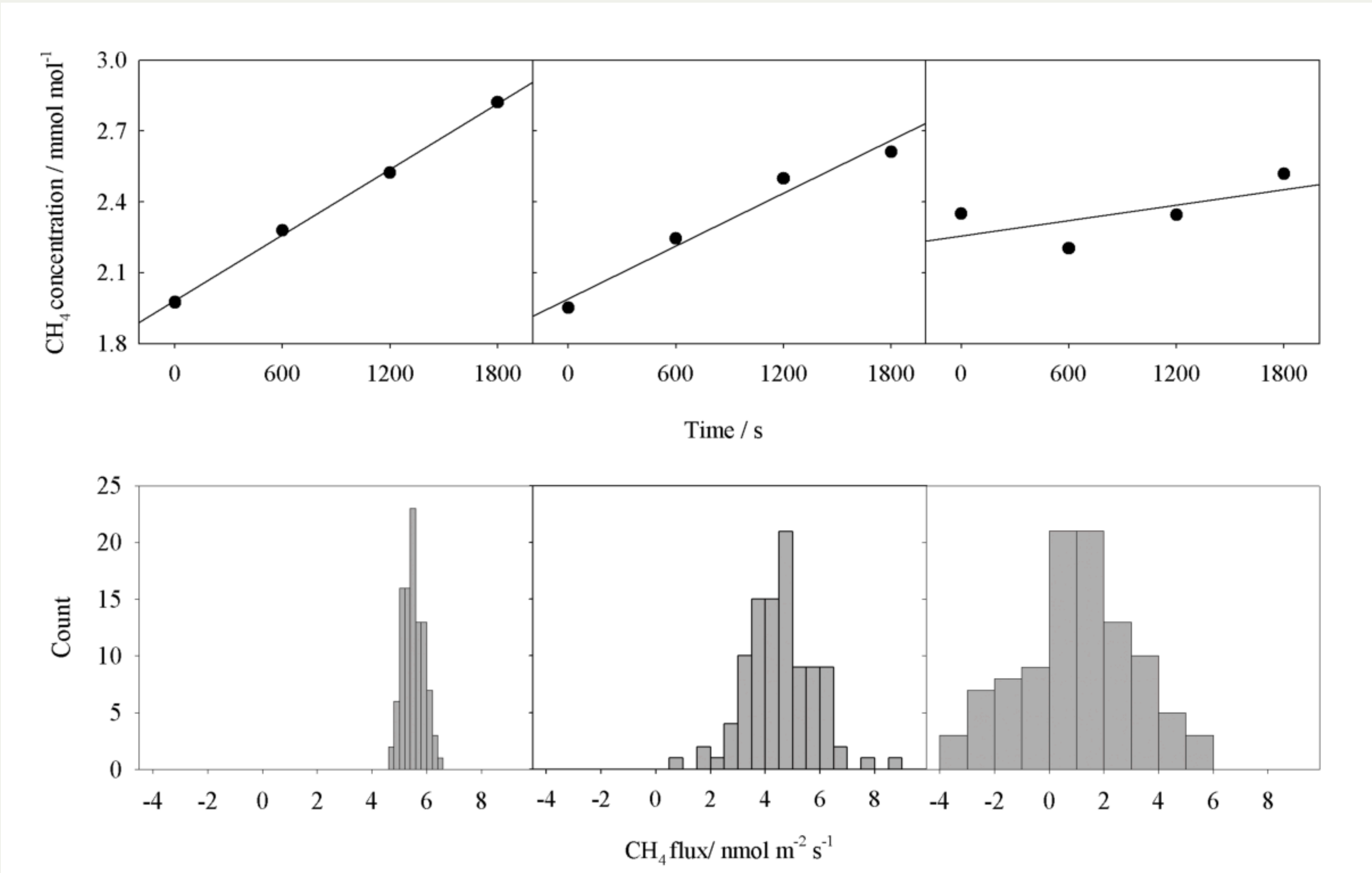




Example: soil gas emissions

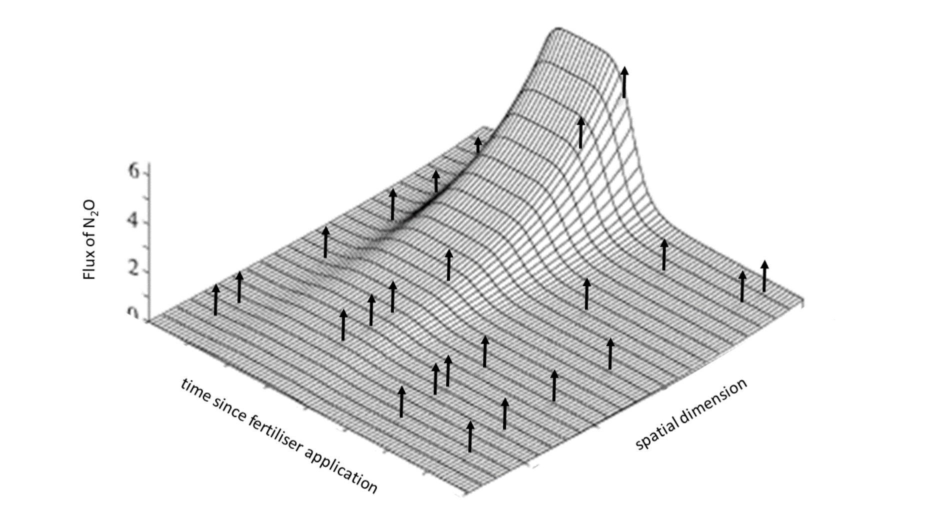
Example: soil gas emissions

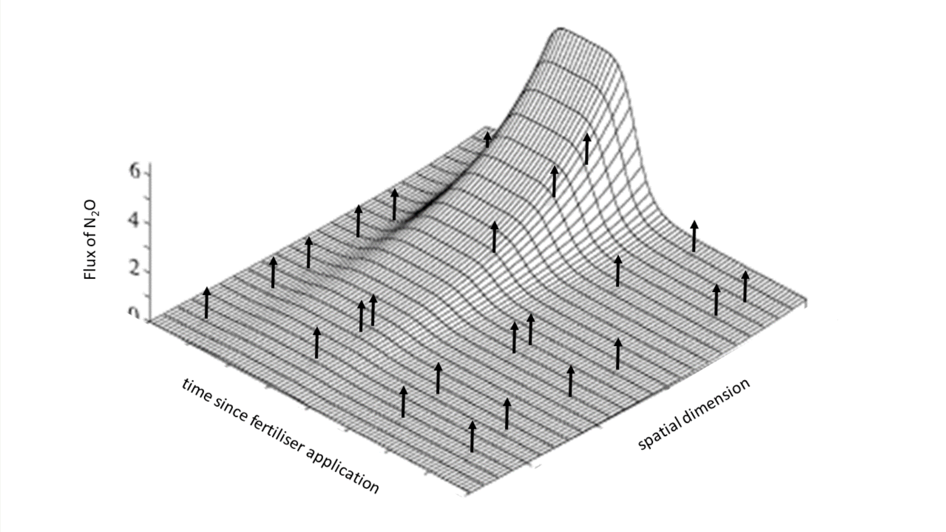
Propagate the uncertainty in each measurement …



Example: soil gas emissions

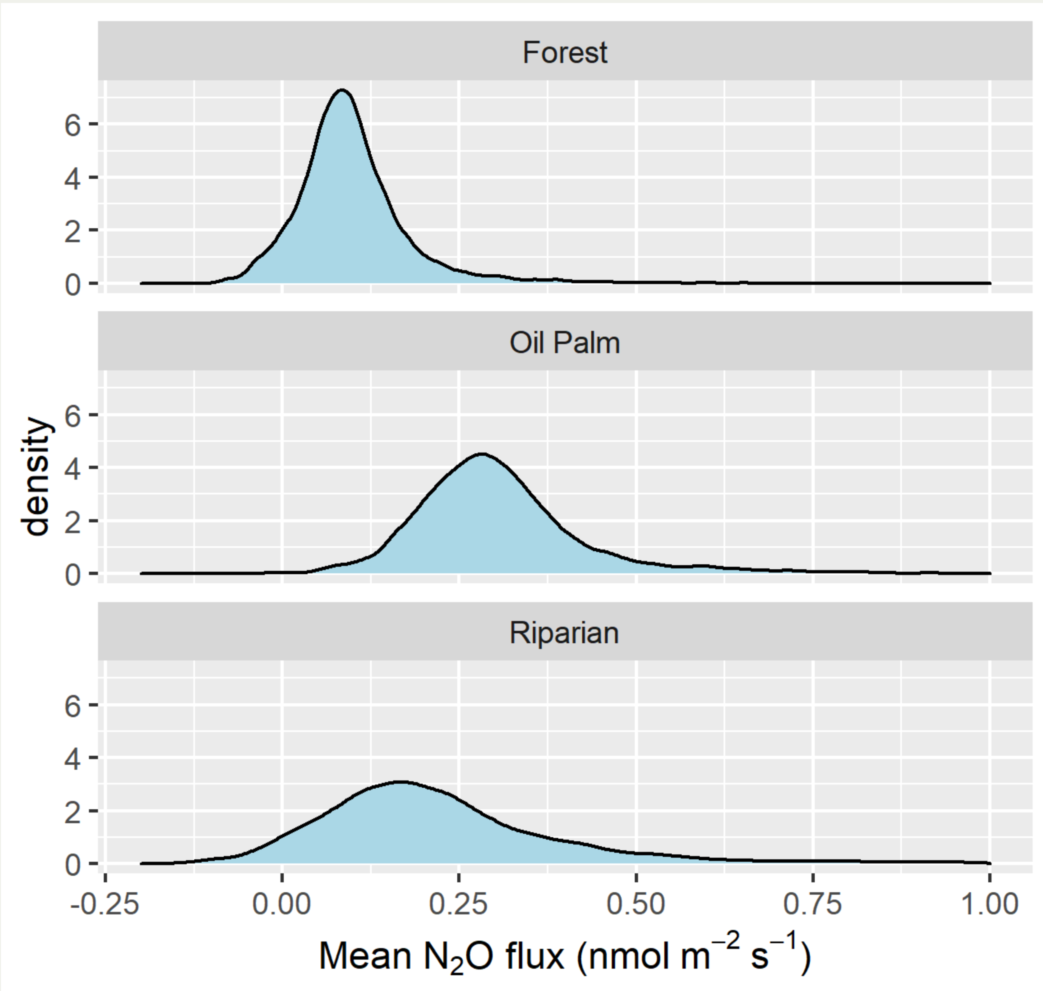
… and in the spatio-temporal extrapolation …





Example: soil gas emissions

… to give posterior uncertainty in treatment effects.



Summary

- Reproducibilty (credibility) crisis arises from:
  - information architecture i.e. filtering
  - misuse of null-hypothesis testing
- In ecology, measurement is difficult & indirect ->
  - low statistical power
  - bias in obs process
- Represent the uncertainty in data & obs process
- Suggests Bayesian approach

Final Thoughts

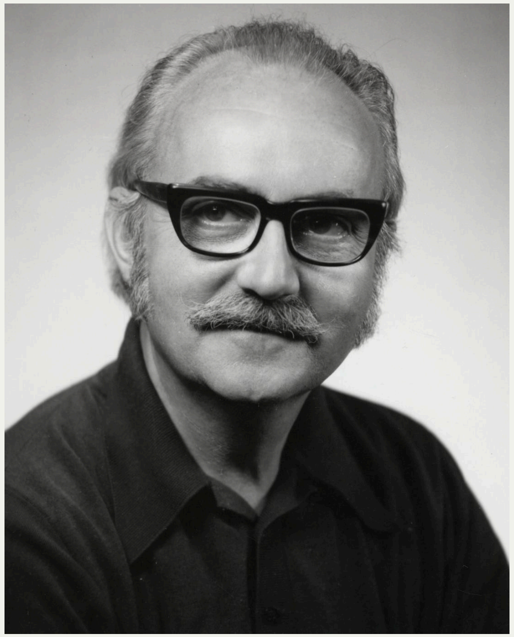
The real purpose of science

"… is to make sure Nature hasn't misled you into thinking you know something you don't actually know."

Robert Pirsig, 1974

“All models are wrong, but some are useful.”

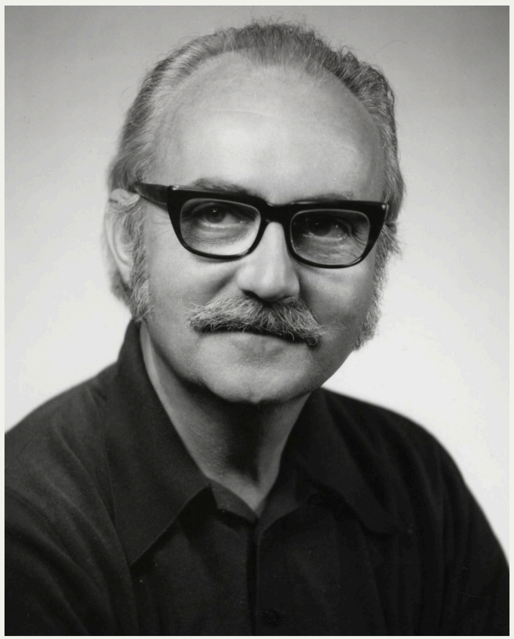
George Box, 1976.



UK CEH

"All models are wrong, but some are useful."

George Box, 1976.



"All data are wrong, but some are useful."

Patterson & Gimlin, 1967