Integrating data across different scales and extents: challenges and opportunities

Dr Pete Henrys
UK Centre for Ecology & Hydrology

UKEOF Conference
27th February 2020

UK-SCAPE

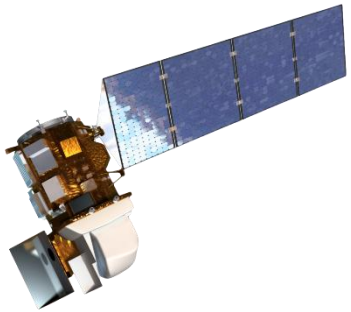
Monitoring Design Options

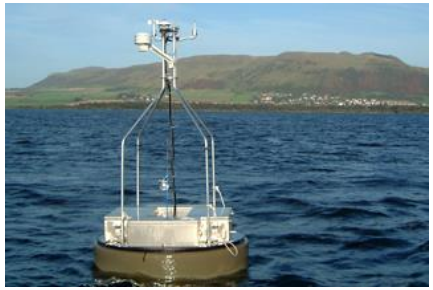
- Random (simple, stratified, cluster, systematic)
- Unbiassed

- Targeted (inc. spatial and temporal structure)
- Bias known

- Opportunistic
- Bias unknown

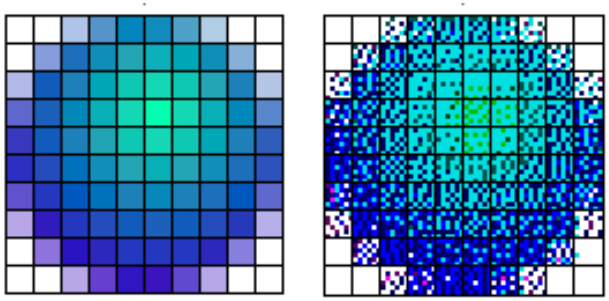
Not just traditional survey



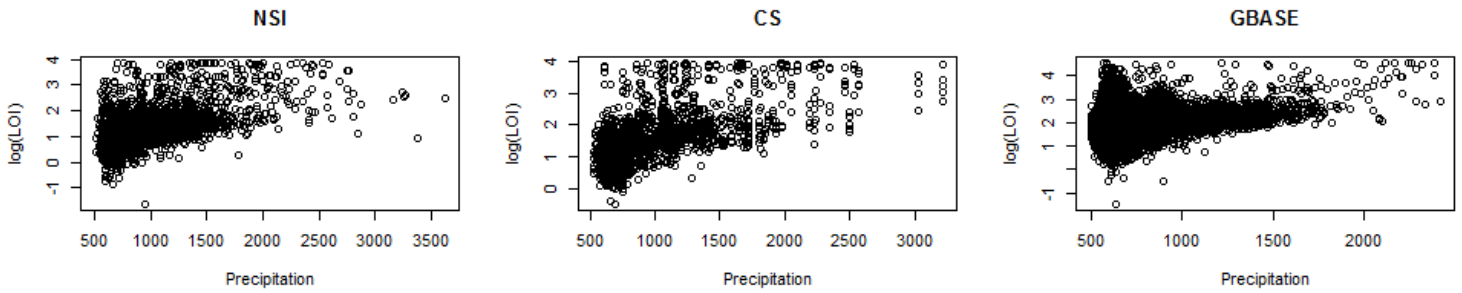


Why Integrate data?

- Consistency
- Increased resolution



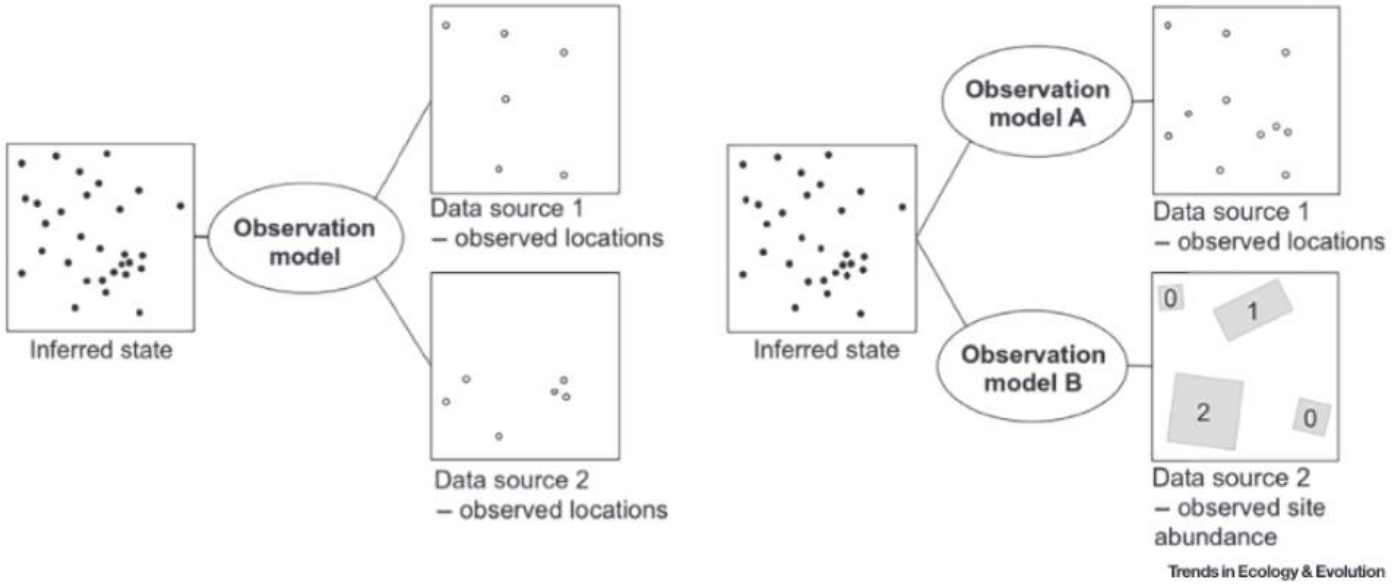
- Increased precision
- Increased covariate gradient
  - Increased power to detect effects



Methodologies for combining data

Simplest is to pool data without explicitly accounting for different observation processes

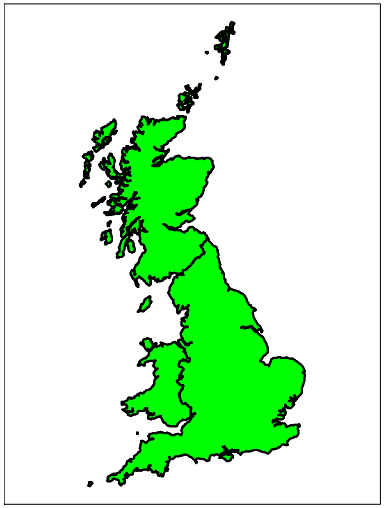
Assume an unobserved true state which is observed by all datasets but with different observation processes

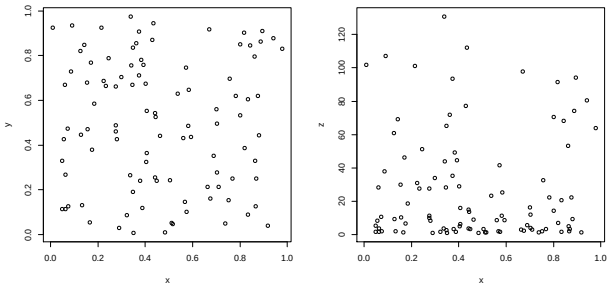


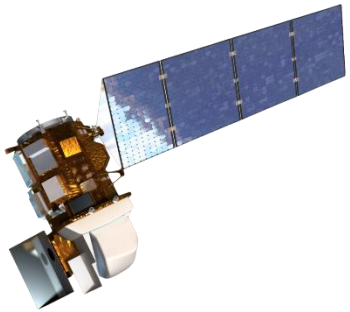
Use joint likelihood, use one dataset as a covariate or use one dataset as prior information

Differences between sources

- Scientific rationale
- Scale
- Representative
- Protocols
- Uncertainty



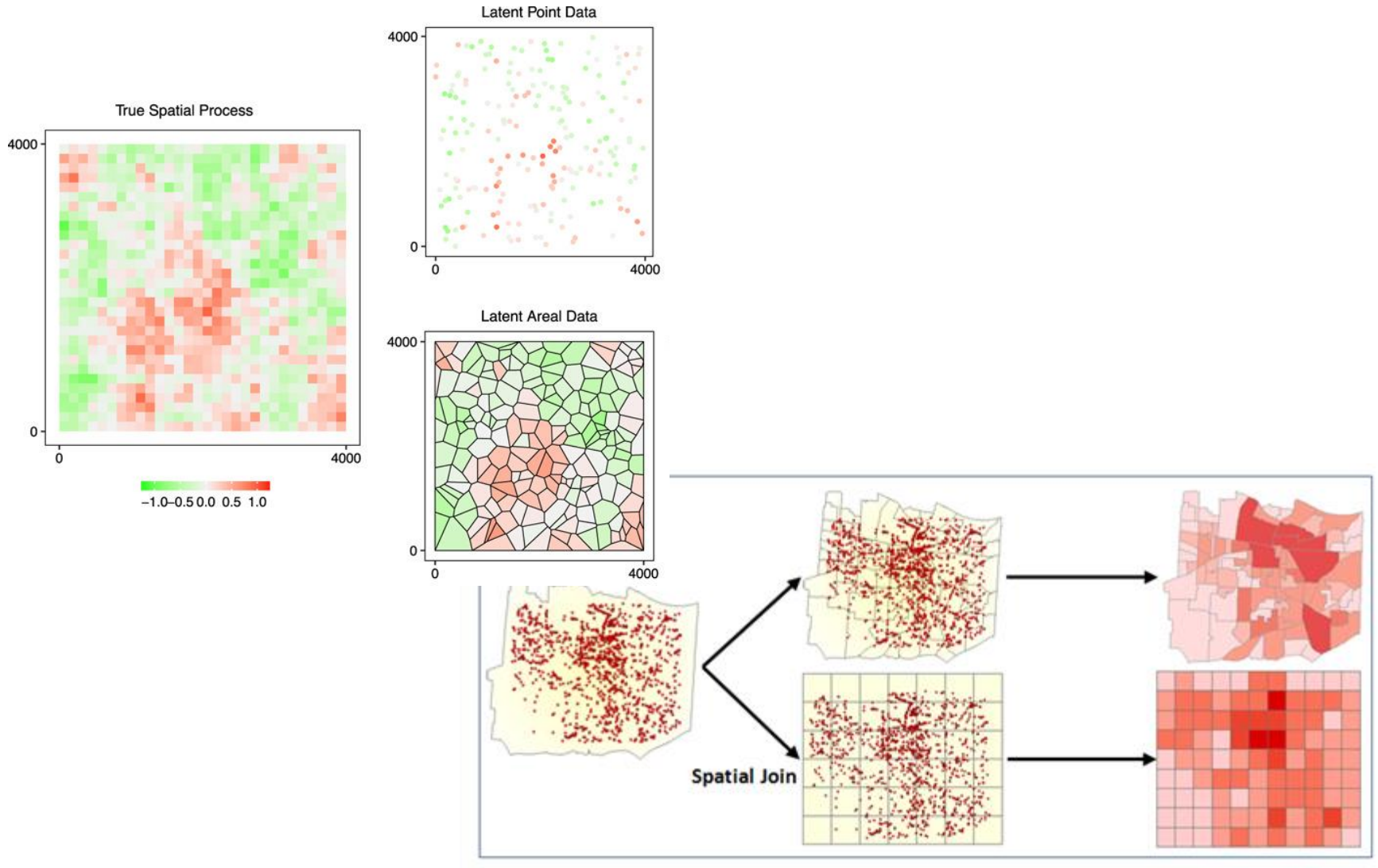




UK Centre for Ecology & Hydrology

UK-SCAPE

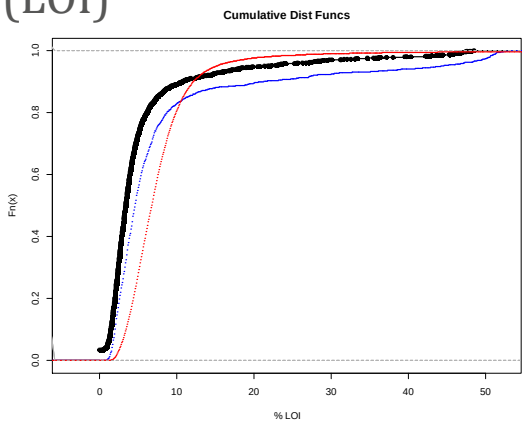
The change of support problem

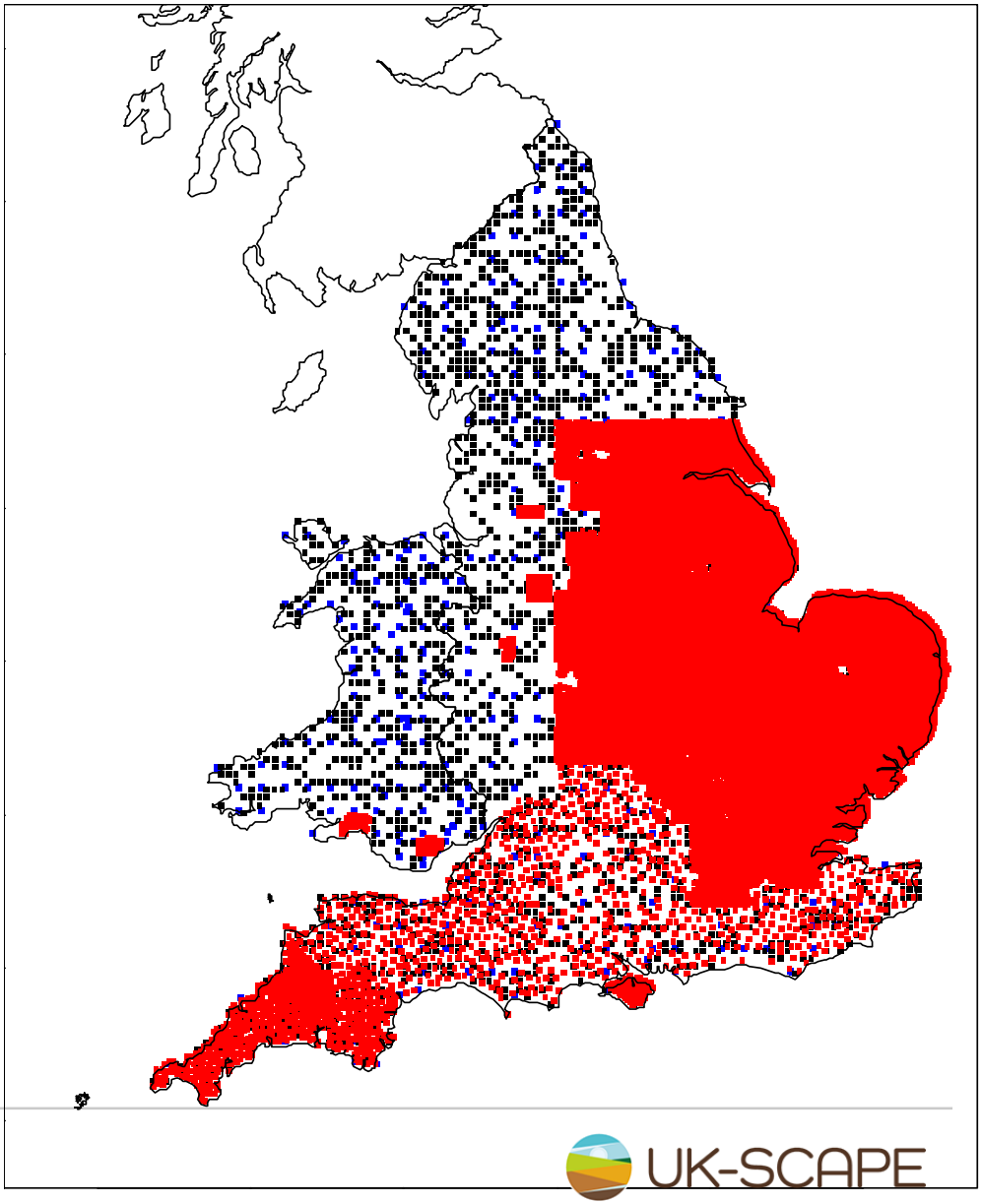


Example: Soil Carbon

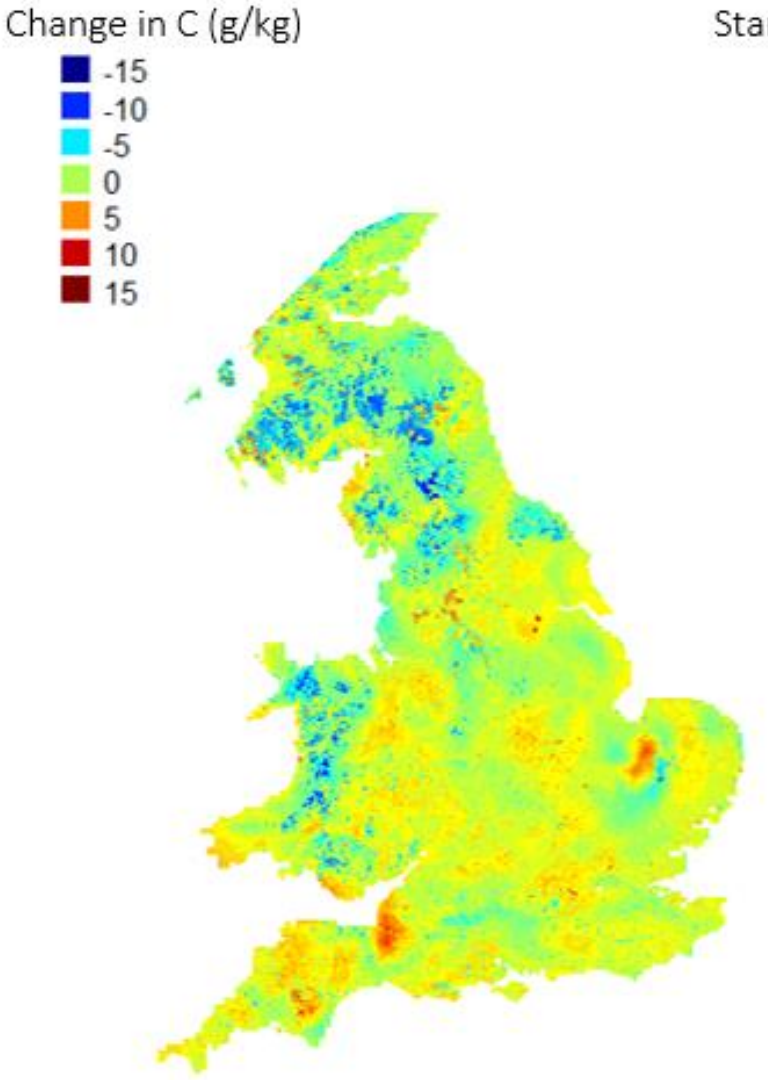
- Countryside Survey
- GBASE
- NSI

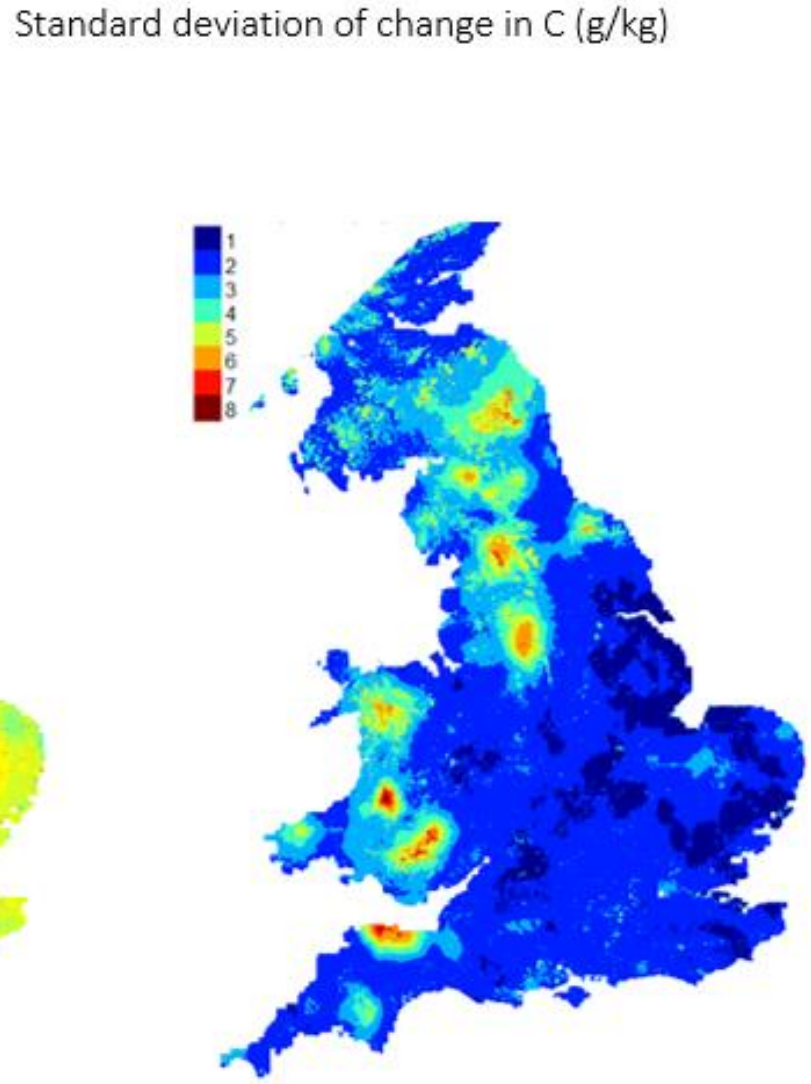
Protocols differ slightly but all measuring same endpoint (LOI)





Example: Soil Carbon





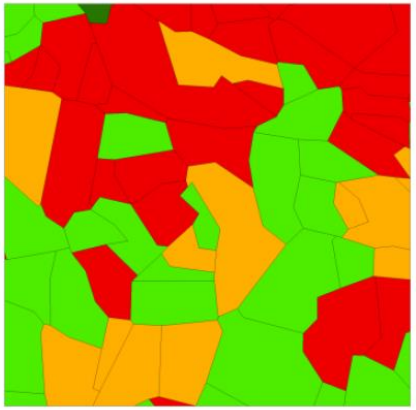
Example: Habitat Extent

Proportion Broadleaved woodland per 1km

Countryside Survey 2007

Field Survey

Land Cover Map 2007



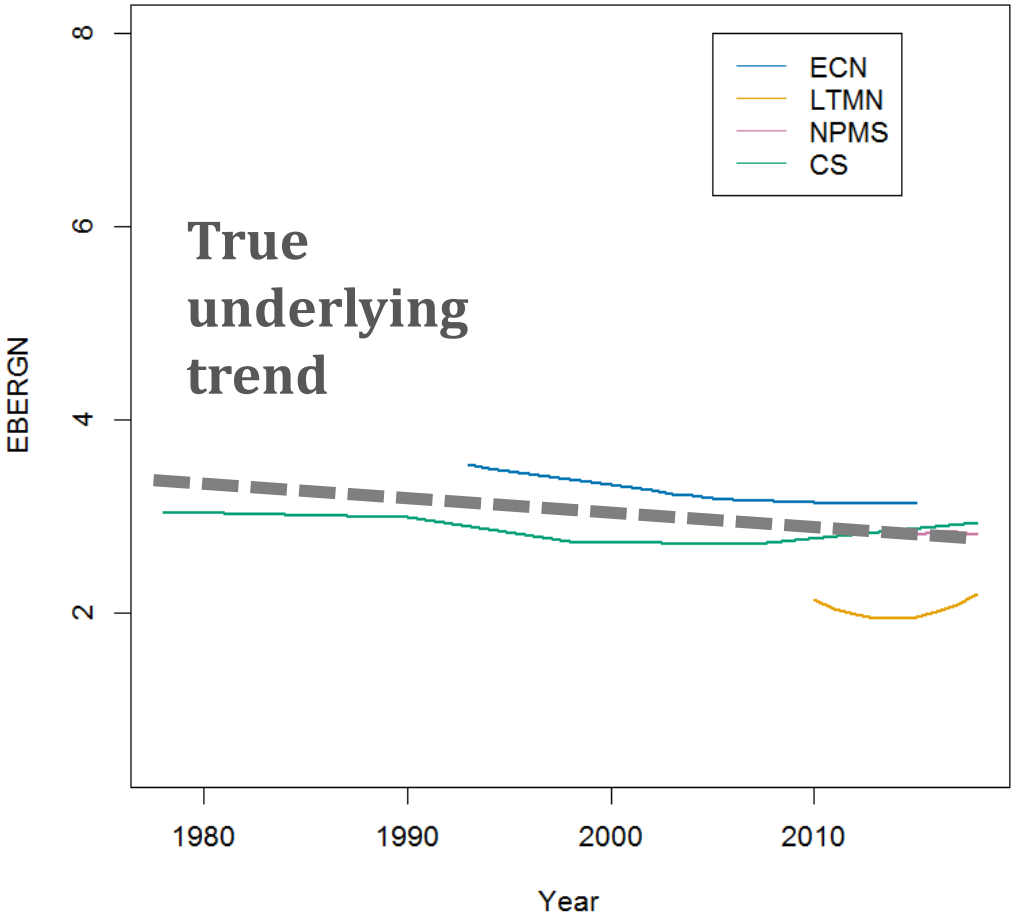
Remote Sensing

Biased Classification error

Truth 2007

Random Measurement error

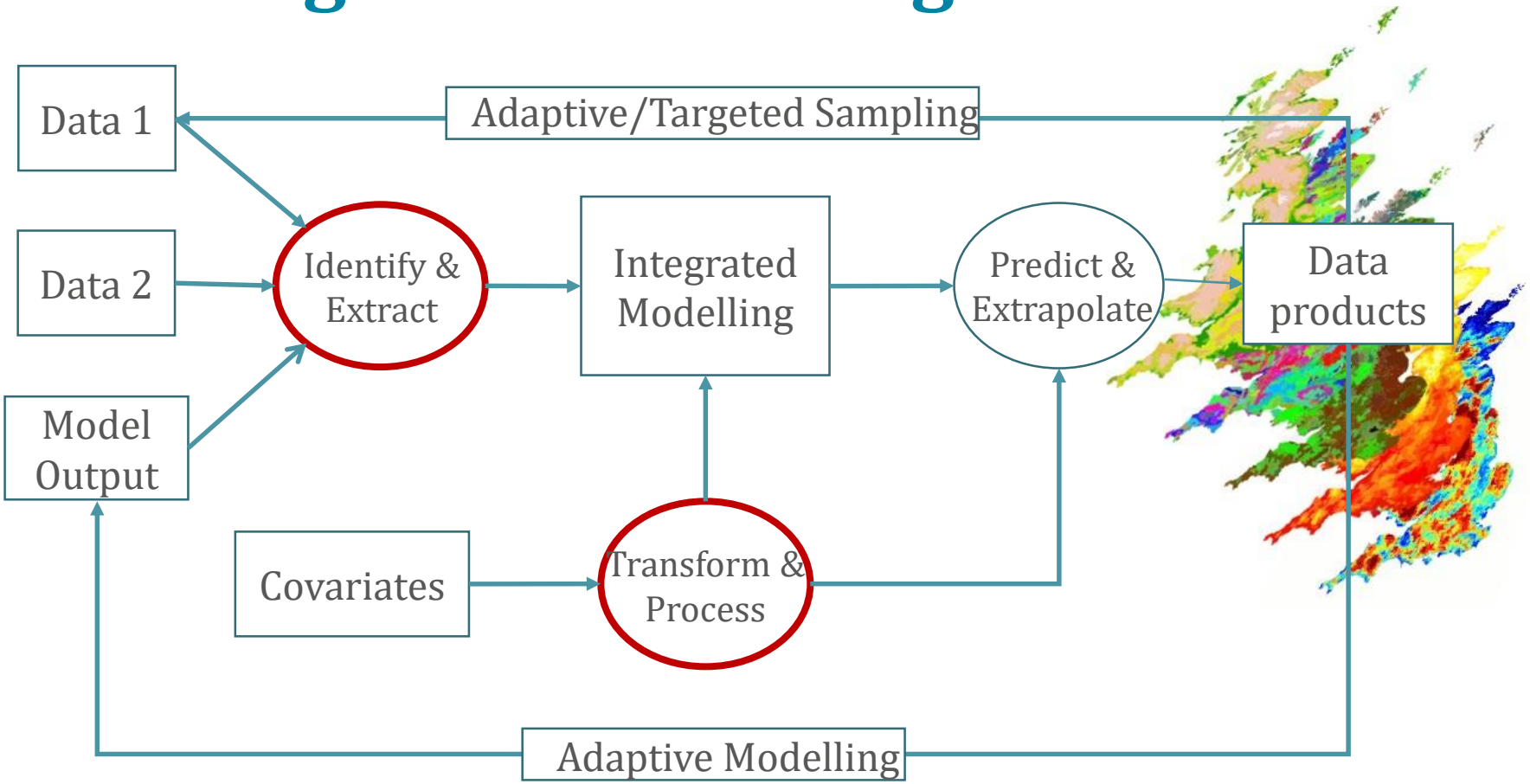
Example: Vegetation Indicators



Estimate the underlying trend while accounting for different:

- Quadrat sizes
- Cover estimation
- Visit frequencies
- Temporal extents
- Etc…

An integrated modelling framework



Summary

- There are numerous methods for combining data from different sources, of different quality and of different resolution
- Need to know about different observation processes to inform the most appropriate model
- Important to consider if it is sensible to combine data
  - Are they measuring the same thing?
  - Are the driver relationships the same?
  - What are the assumptions?