The influence of data availability on parameter estimation and prediction uncertainty

A case study from Lanna (FOCUS D1)

Mats Larsbo and Nick Jarvis

Department of Soil Sciences, SLU

Box 7014, 750 07 Uppsala, Sweden

Mats.Larsbo@mv.slu.se

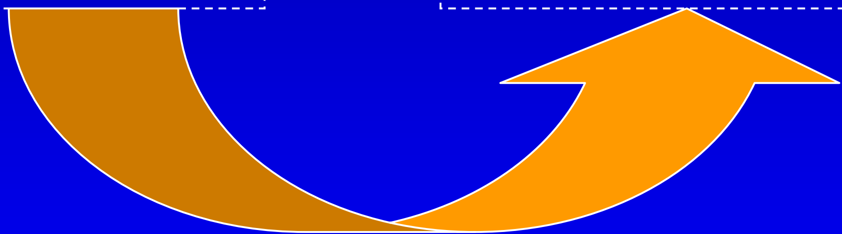
Predictive 'higher-tier' modelling

Lower-tier information

Laboratory
K_{oc}, DT50

Higher-tier information

Lysimeter and field experiments



1. Calibration (inverse modelling)

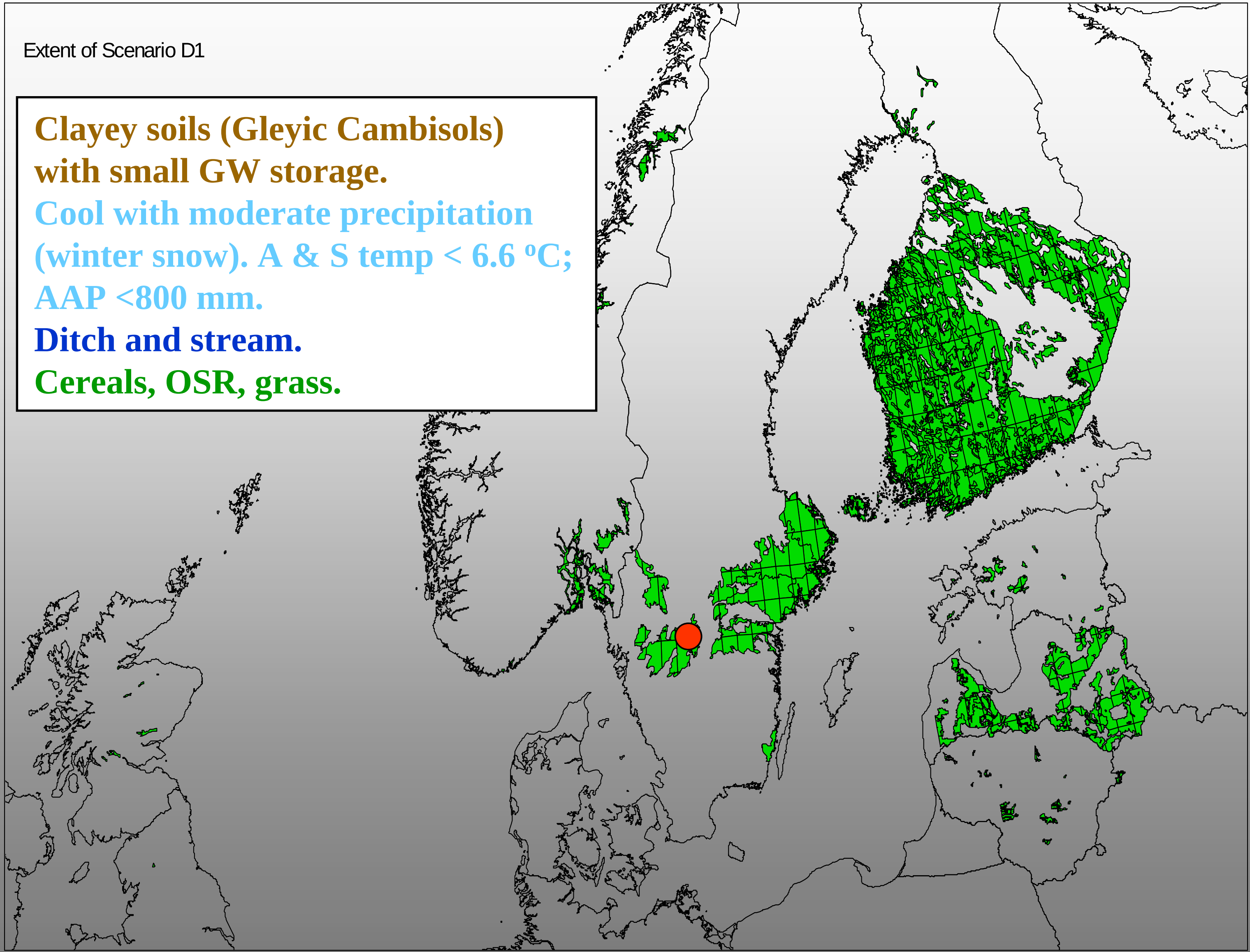
2. Prediction (extrapolation)

Predictive higher-tier modelling

- What methodology should be used?
- What is the uncertainty in parameter estimates and predictions?
- How does data quantity and quality influence the outcome?

Case study for FOCUS Surface Water D1

Extent of Scenario D1



Clayey soils (Gleyic Cambisols)
with small GW storage.
Cool with moderate precipitation
(winter snow). A & S temp < 6.6 ºC;
AAP <800 mm.
Ditch and stream.
Cereals, OSR, grass.

FOCUS D1 (Lanna)

Soil: well-structured, 46% clay, 2% OC in topsoil, tile drains at 1 m depth and 14 m spacing

Available data at Lanna

- DT50 determined in the laboratory
  - Topsoil: 12.6 d (k = 0.055 d^{-1})
  - Subsoil: no detectable degradation in 50 d incubation
- Measured hydraulic properties
  - Soil water retention, saturated hydraulic conductivity
  - Saturated matrix hydraulic conductvity (tension infiltrometer)

- Field experiment on the leaching of bromide and bentazone to drains during 1 year

Field data (Lanna)

Observations	Resolution	Number	Method
Soil water content	10 cm intervals to 90 cm depth	5 profiles	Soil cores, 8 to 11 replicates
Resident concentration	0-10 cm, 20 cm intervals to 90 cm depth	5 profiles for bromide, 3 for bentazone	Soil cores, 8 to 11 replicates
Drain flow	Continuous	415	Float activated pump
Flux concentration	Every 1.5 mm of drain flow	103	Automatic flow proportional sampling

GLUE (Generalized Likelihood Uncertainty Estimation)

- 'Equifinality': multiple parameterisations will give equally acceptable fits to data
  - Rejects the idea of 'optimal' parameter values

- GLUE methodology:
  - Parameters defined by uncertainty ranges
  - Monte Carlo simulations (random or LHS sampling)
  - Goal (objective) function as measure of 'likelihood'
  - Cut-off value defines 'acceptable' parameter sets
  - 'Acceptable' parameterisations used for prediction (weighted by their 'likelihood')

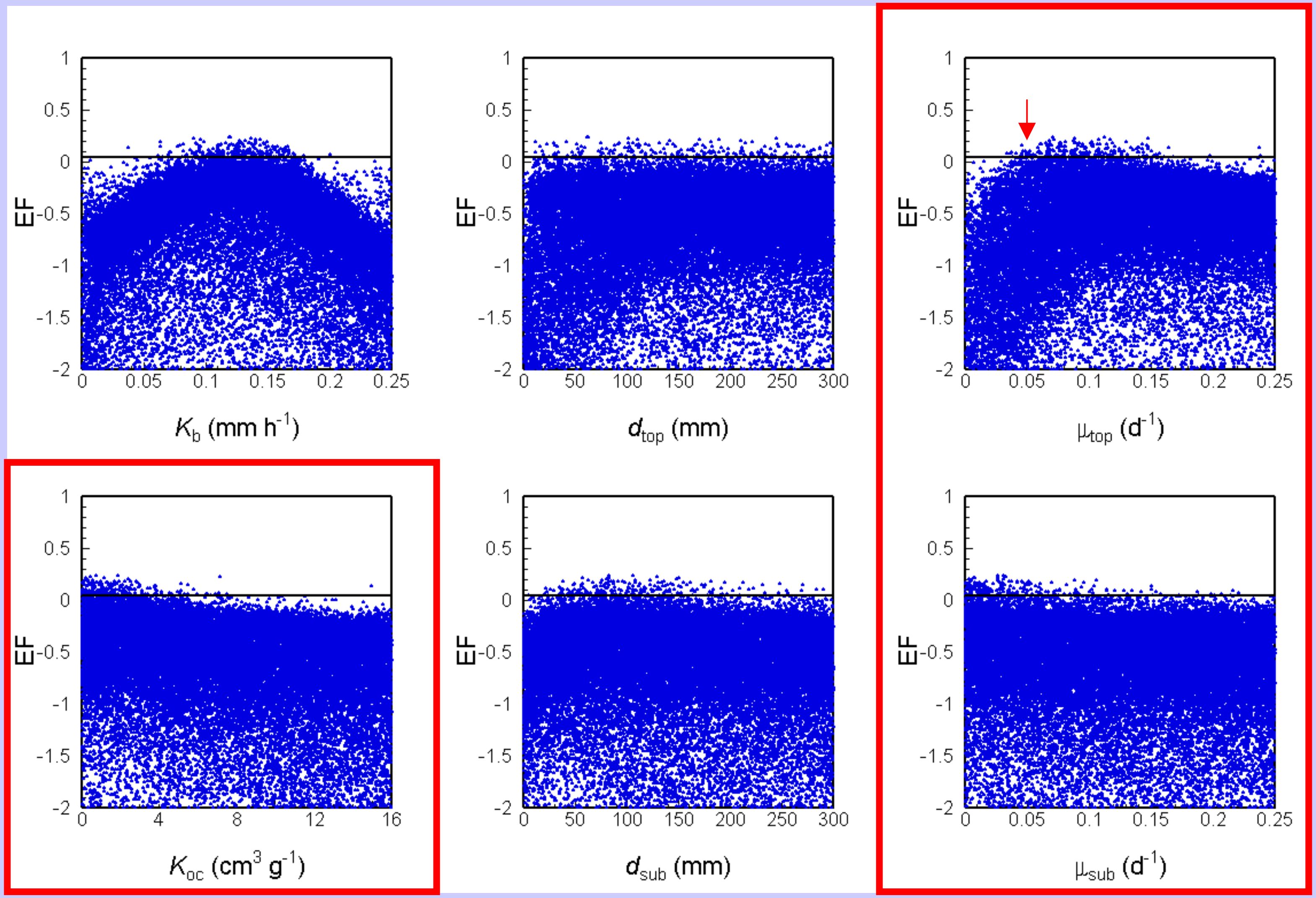
GLUE'ing MACRO at Lanna

- Four sensitive parameters:
  - Degradation rate coefficient (μ, topsoil/subsoil)
  - Sorption K_{oc} value
  - Diffusion pathlength (*d*, topsoil/subsoil)
  - Saturated matrix hydraulic conductivity, K_b
- 30,000 simulations sampled by LHS
- Model efficiency (EF) used as 'likelihood' measure
- $EF > EF_{max} - 0.2$ considered 'acceptable'
- Data grouped in four different ways:
  - All data
  - Water contents and resident concentrations
  - Drainflow and flux concentrations
  - No tracer data

Prediction uncertainty

- Lanna experiment was definitely not 'reasonable worst-case'
  - Double-dose, autumn application to bare soil !
- Acceptable parameter sets used to make predictions for GAP scenario in Sweden
  - Recommended dose, spring application to peas
  - Same 2-year weather data set
  - Two target output variables
    - Maximum daily concentration in drainflow
    - Total loading in drainflow

GLUE 'Dotty plots' (all data)

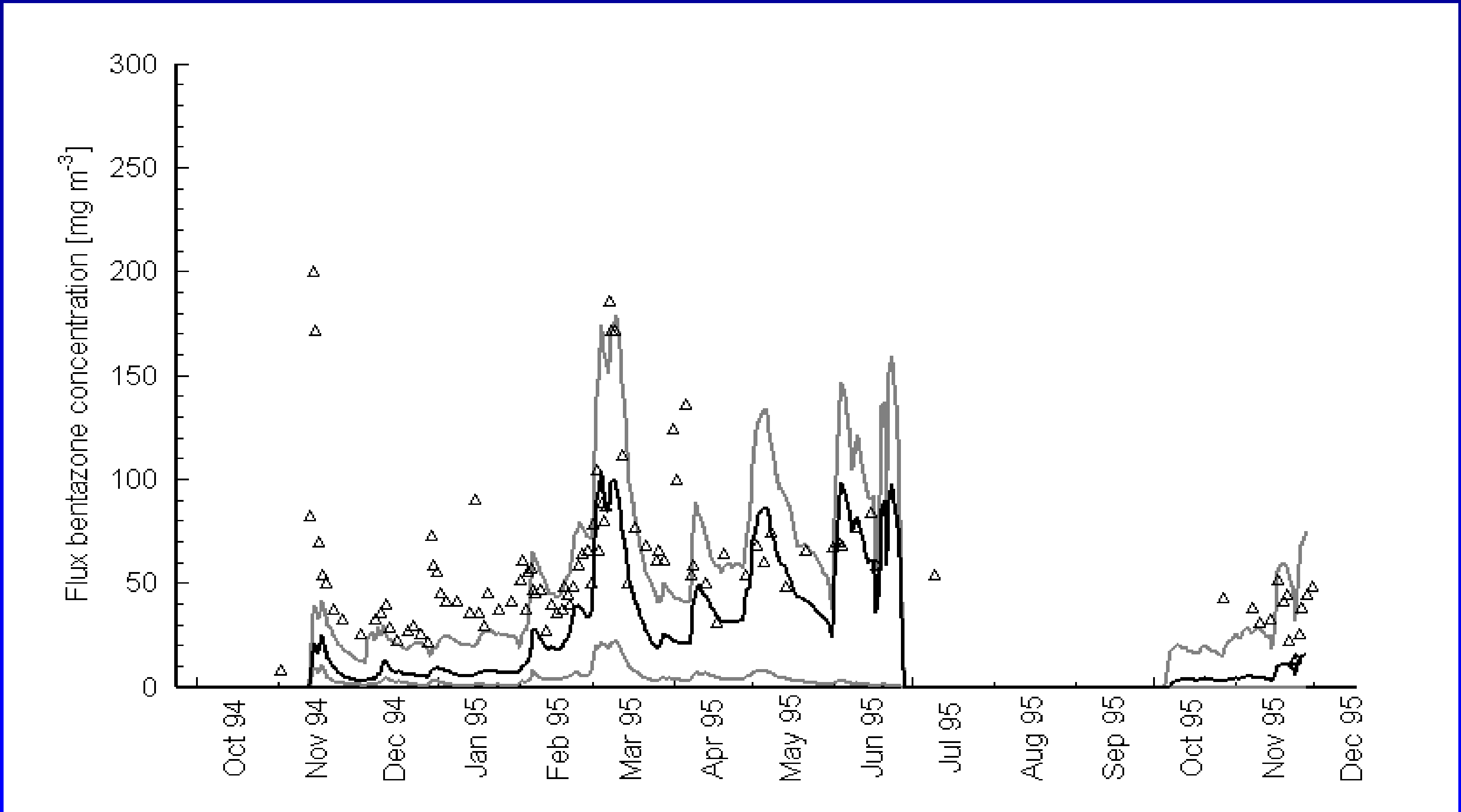


235 acceptable simulations (of 30,000)

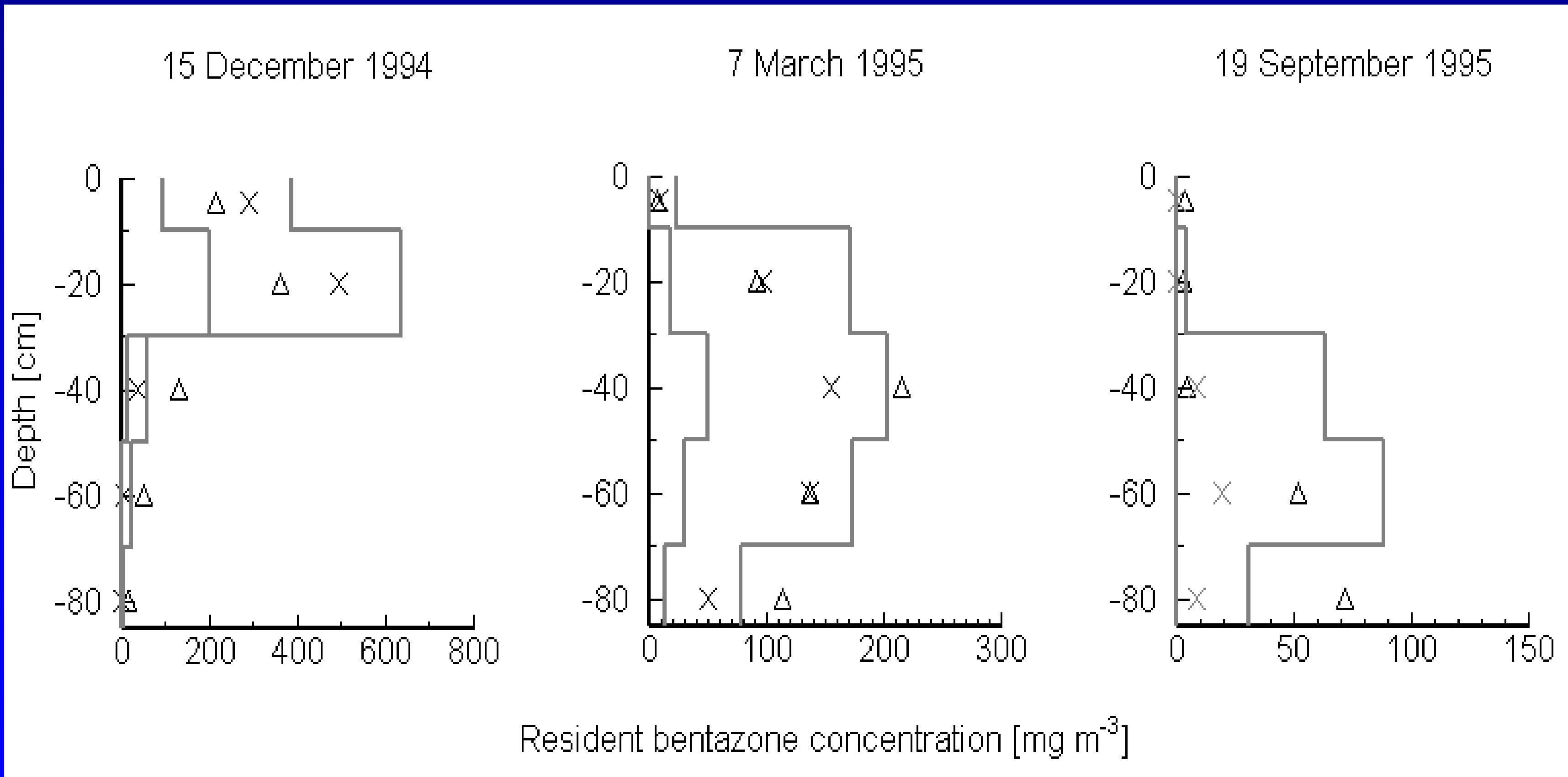
Model efficiencies

Data group	'Overall'	'Optimal'
Water content	0.51	0.54
Drain flow	0.30	0.44
Bromide resident concentrations	0.34	0.80
Bromide flux concentrations	-0.20	0.33
Bentazone resident concentrations	0.69	0.82
Bentazone flux concentrations	-0.19	0.31
All data	0.24	-

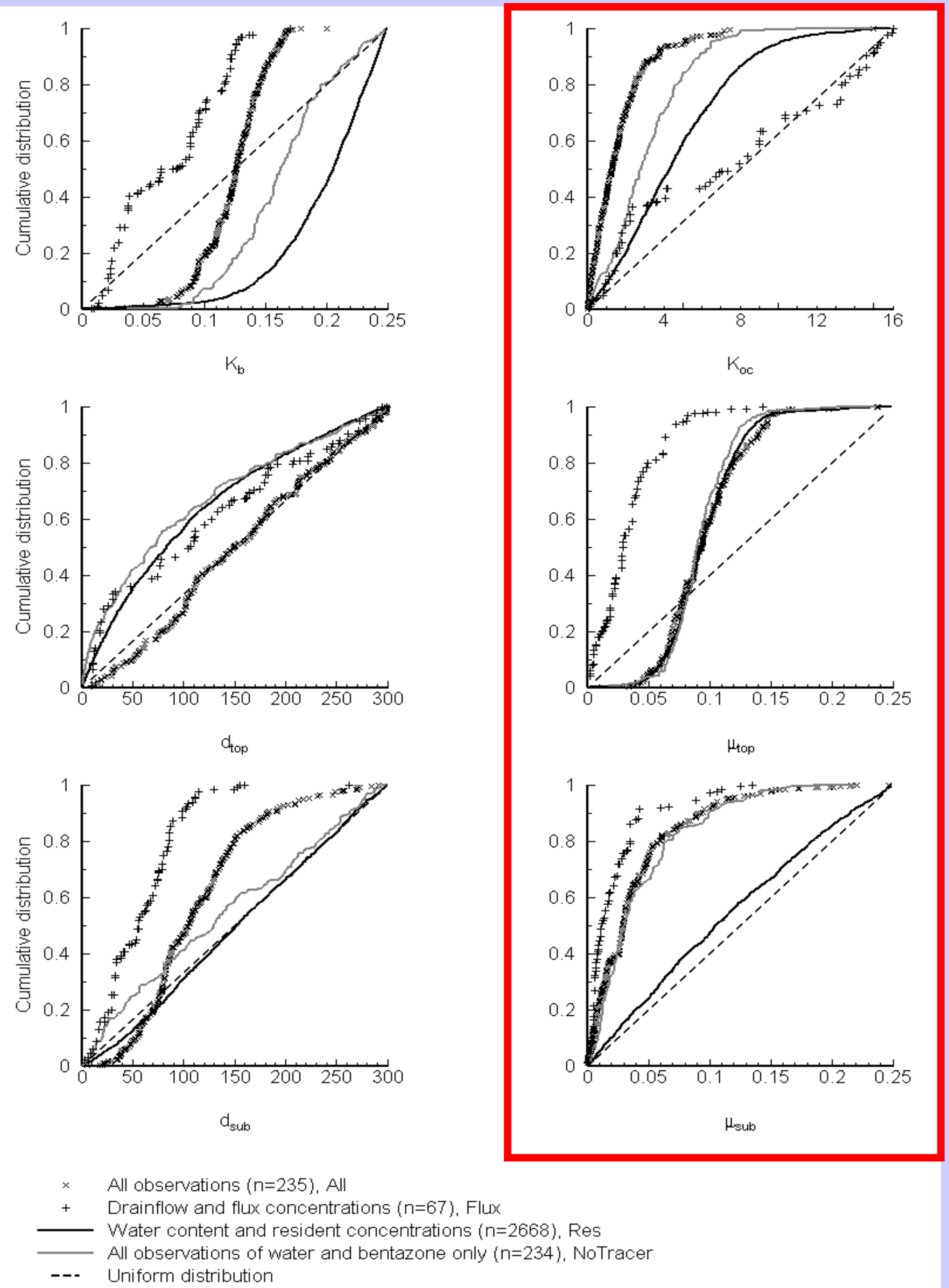
Best overall GLUE simulation, and 5th and 95th percentile estimation intervals



Best overall GLUE simulation, and 5th and 95th percentile estimation intervals



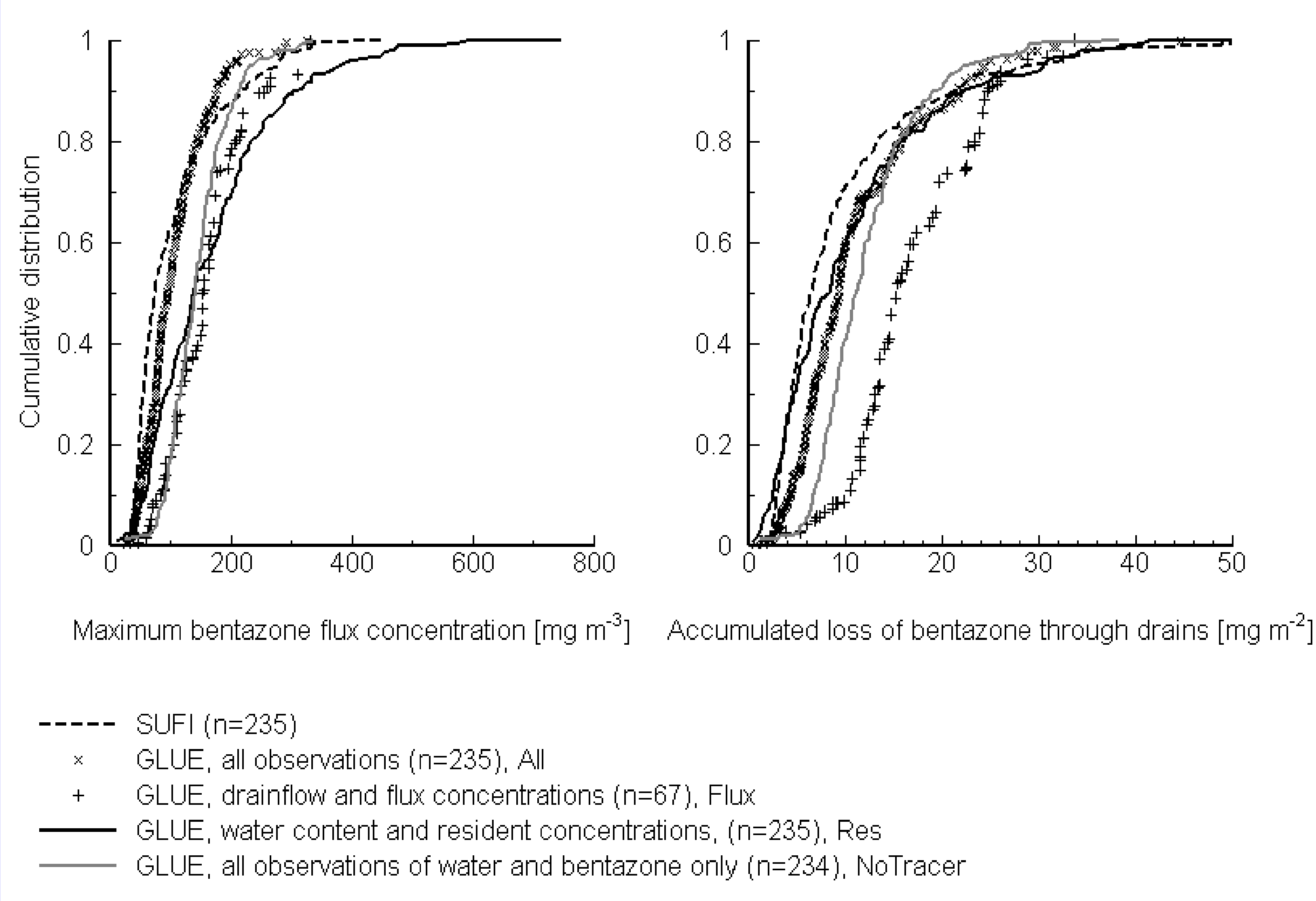
Parameter conditioning and data availability



<u>Both</u> flux and resident pesticide concentrations are needed to obtain well-conditioned and unbiased estimates of the degradation rate coefficient

Tracer data also improves the estimation of K_{oc}

Prediction uncertainty



Prediction uncertainty: maximum concentration in drainflow

Data group	[a]Bias (%)	[b]Precision (mg m^{-3})
All	-	157
Resident	+46	338
Flux	+63	250
No tracer	+49	157

[a] defined as the deviation from 50^{th} percentile value for 'All'

[b] defined as 95^{th} percentile value minus 5^{th} percentile value

Conclusions

- Inverse methods
  - Efficient, quantitative (supplies estimates of uncertainty) and objective
  - But…. non-unique solutions inevitable?
  - Work with range of acceptable parameters sets?
- Comprehensive datasets reduce uncertainty in parameter estimates and model predictions
  - Flux and resident concentrations
  - Tracer
- Significant uncertainty in predictions despite comprehensive data and efficient conditioning

For more information : Larsbo, M. & Jarvis, N. 2005. Simulating solute transport in a structured field soil: uncertainty in parameter identification and predictions. Journal of Environmental Quality, 34, 621-634