IEAGHG workshop: Environmental Impacts of CO_2:
Bozeman, Montana, 18 July 2012

“No detectable leakage”: accuracy and sensitivity of storage monitoring methods

Anna Korre and Charles Jenkins

Imperial College London

mining and Environmental Engineering Research Group
Department of Earth Science and Engineering
Royal School of Mines
Prince Consort Road
London, SW7 2BP

Outline

- Issues related to “No detectable leakage”
- Will use two examples to illustrate how we can assess accuracy and sensitivity in CO_2 storage monitoring
- Strengths and weaknesses of different methods
  - Classical hypothesis testing
  - Comparing two models
  - Comparing multiple models
- Decision considering uncertainty – can we do it?

Accuracy and sensitivity of monitoring methods

- Important to measure the amount of CO_2 securely stored
- Required to underpin financial mechanisms (e.g. carbon credits)
- Required to demonstrate no leakage
- Part of the social license to operate a storage site (key assets are protected, no risk to public safety)

Requirement for accuracy and sensitivity of monitoring methods

- May vary across applications (e.g. saline aquifers, depleted gas fields, …)
- May be stringent (e.g. allowable leakage rate to atmosphere < 0.01% yr^{-1})
- Tolerable leakage to a potable aquifer
  - difficult to measure,
  - difficult to quantify,
  - only ‘no detectable leakage’ may be acceptable.

for well chosen and designed sites will we be measuring “nothing”?

Where do uncertainties come from ?

- “Many uncontrolled and unknown variables find their way into the data.
- The models used to interpret the data may contain uncertainties
- Data refer to open systems, affected by varying factors outside monitoring control, which may also be unknown.
- Dealing with ‘differences’ is made more acute due to real world operational difficulties.
- Assurance measurements comparing baseline with post-injection is
  - vulnerable to false alarms,
  - connected by a long chain of causation to leakage,
  - marginal detections and data of any type will need to be interpreted.

Terminology in CCS regulations

- “no detectable leakage” technique dependent concept
- “no significant deviation” subjective concept

- Imprecision may be an advantage (future-proofing regulations)
- Requires objective and quantitative ways to assess monitoring data

Monitoring data interpretation requires

the **use of statistical methods** and

to be **communicated in the language of probability**

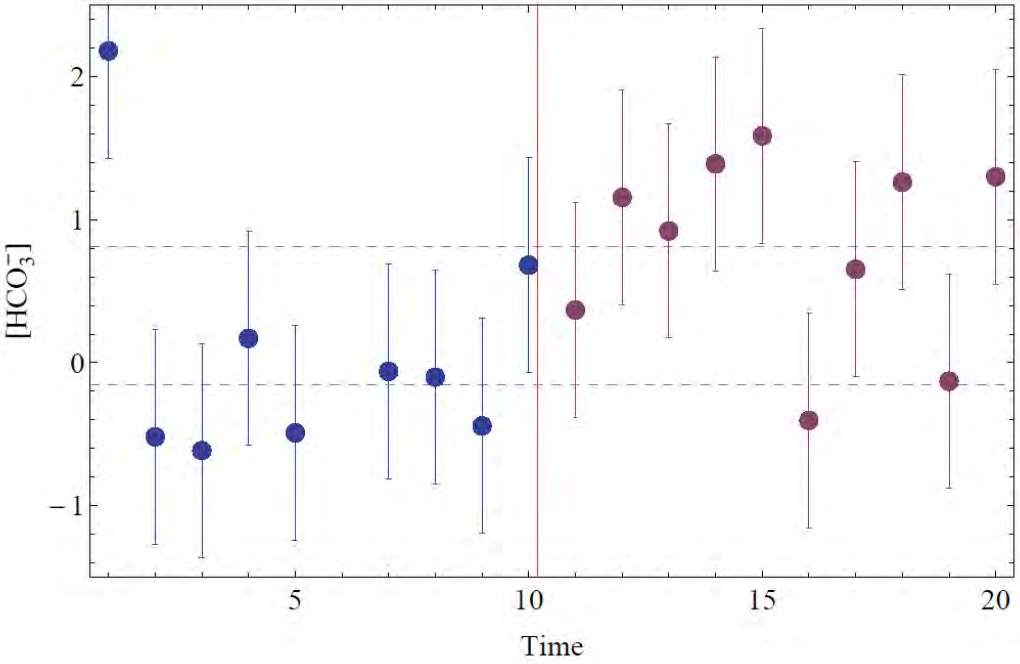
Example 1: bicarbonate concentrations in groundwater samples

Assumption :

- Little is available by way of models that link these data to CO_2 leakage

Question :

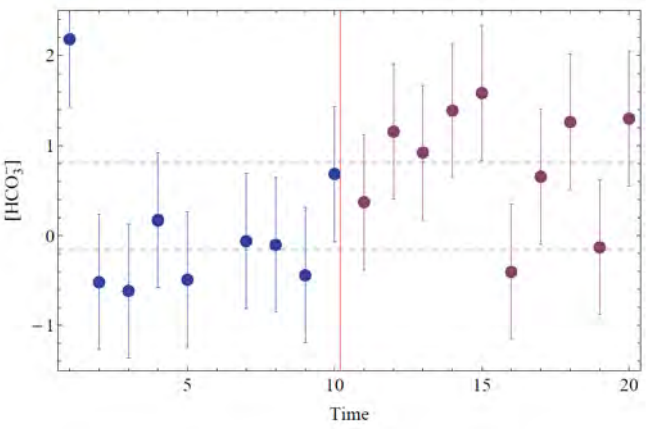
- Are data consistent with no significant change post injection?



No leakage model: The true bicarbonate level is constant.

Leakage model: The true bicarbonate level is constant before and constant after injection, but the levels are different.

Classical “hypothesis testing”



We only consider one model: **no leakage**

We calculate the probability p (significance) of obtaining the data assuming no leakage.

- If p is small, the no-leakage assumption is in doubt.
- If we set a prior threshold α in p, below which we will reject the no-leakage assumption, then α is the *false alarm rate*.

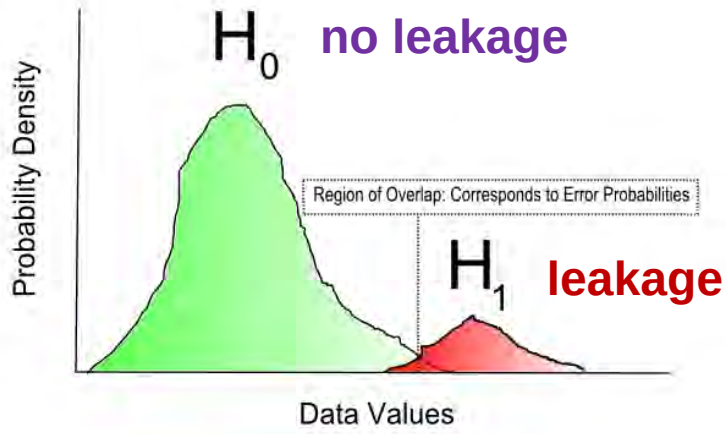
Issues:

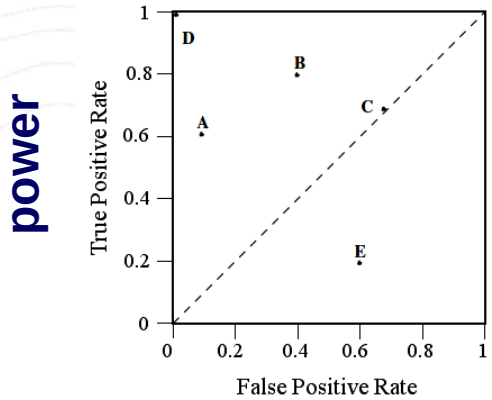
- There is a probability that a value $p < \alpha$ will arise entirely by chance!
- Rejecting **no leakage** tells nothing about “how much leakage” or what kind (we have not specified a model).
- If the data have high uncertainty **leakage** may not be statistically significant.
- Cannot provide the probability that the **no leakage** model is correct.

Comparing two models Neyman-Pearson

The focus is on setting decision rules in advance so either the **no leakage** or the **leakage** model will be accepted, **based on the data**.

It is possible to calculate the performance of a given decision rule.



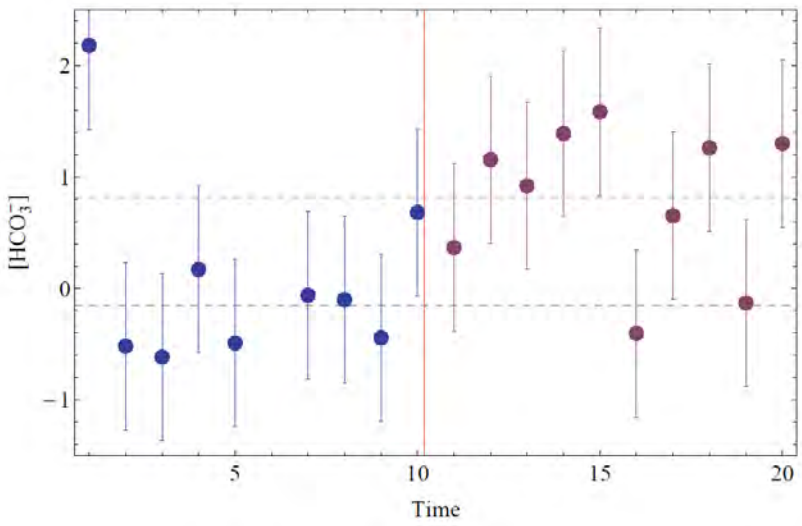


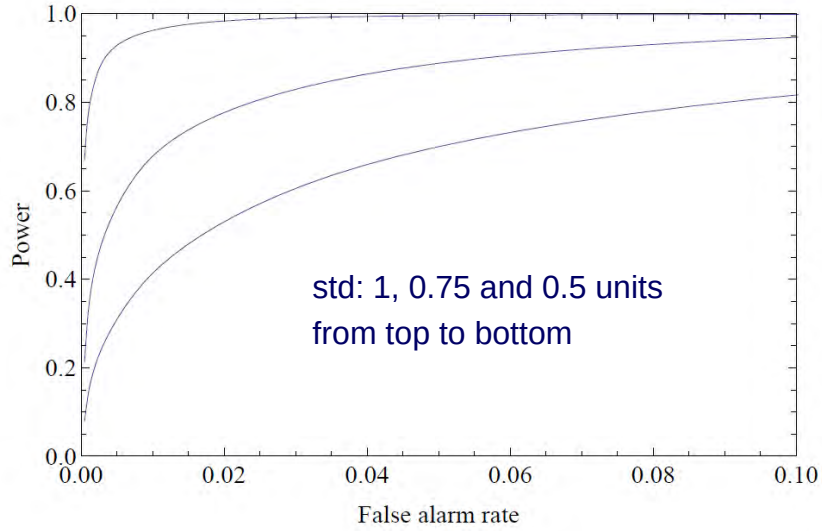
		True class	
		no leakage P	**leakage** **N**
Predicted class	**no leakage** P'	True Positive (TP) 1-α	False Positive (FP) P
	leakage N'	False Negative (FN) α	True Negative (TN) 1-P

$$\text{TP Rate} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{FP Rate} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

Comparing two models - Neyman-Pearson





- Use the *t*-statistic.
- specify the difference in means (1 unit between pre- and post-injection).
- Calculate *t*-statistic for a range of false alarm rates α, derive P corresponding to success rate.
- Can consider uncertainty in data.

Therefore leakage or no leakage classification depends on the **acceptable false alarm rate**

Comparing two models - Neyman-Pearson

Issues:

- We work with probabilities of data given the competing models.
- However, would prefer to take decisions based on the probabilities of the models.

We cannot discuss **"the probability of a leak"**

We can only discuss

"the probability of data assuming a leak exists"

- There may be several models that we wish to consider
- Can perform a series of tests but combining the information in various ROC curves is not possible

Multiple models and Bayesian methods

The conceptual difference between Bayesian methods, and the classical methods is that:

- The **Bayesian view** of probability includes the possibility **that models and parameters have probabilities**, whereas in
- the **classical view** the **models and parameters are fixed** and **probabilities can only be associated with data**.

Accepting a Bayesian approach obliges to postulate specific alternative models.

A natural consequence is that we obtain the posterior probabilities of the various models, which can be ranked and evaluated.

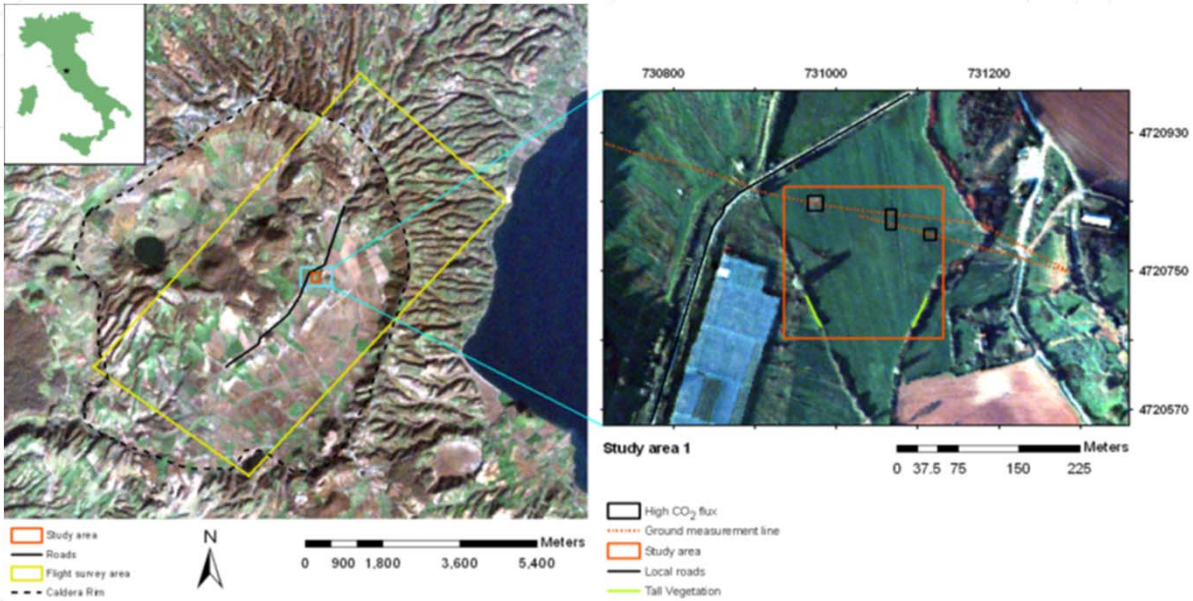
We can discuss

"the probability of leakage models, given the observed data"

An exhaustive suite of leakage models has to be specified!

Example 2: Indirect detection of CO_2 leakages using remote sensing data

Govindan, R., Korre, A., Durucan, S., Imrie, C.E., 2011. A geostatistical and probabilistic spectral image processing methodology for monitoring potential CO_2 leakages on the surface. *International Journal of Greenhouse Gas Control* 5, 589-597.



Airborne data:

Multispectral data 11 bands (Visible to TIR); spatial resolution: 2.5m. (provided by BGS, UK)

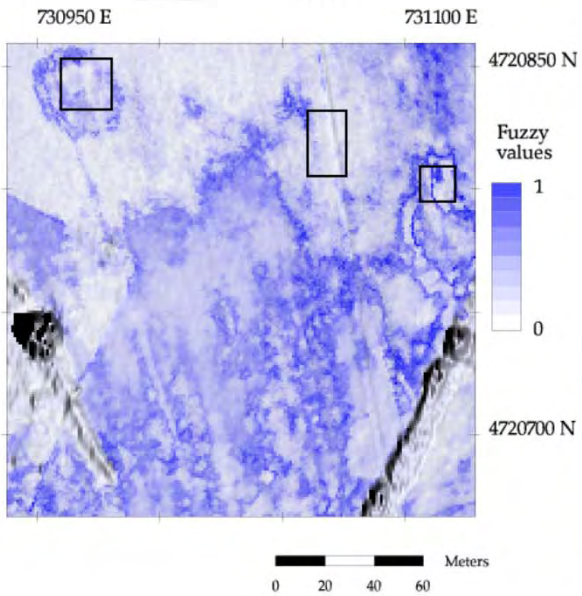
Hyperspectral data 63 bands (Visible to VNIR); spatial resolution: 2m and 1m. (two data sets provided by OGS, Italy)

Field measurement data:

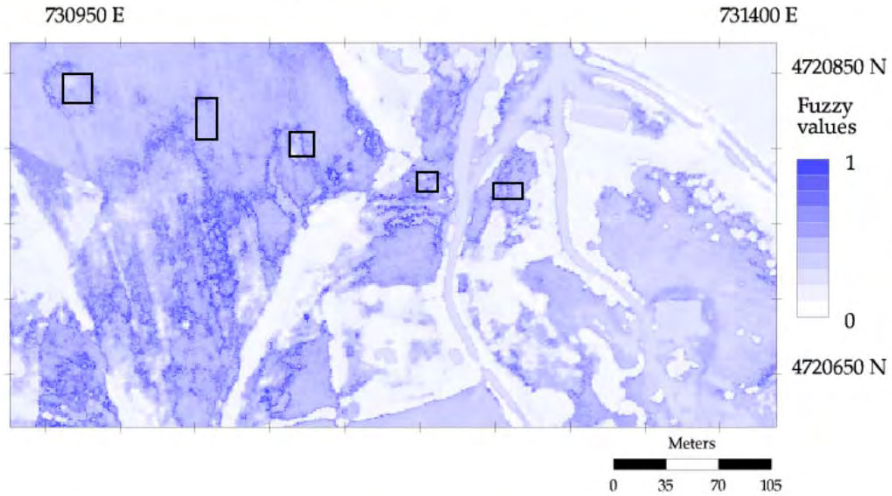
Gas flux and concentration measurements (data provided by the University of Rome 'La Sapienza').

Comparing two models - Neyman-Pearson

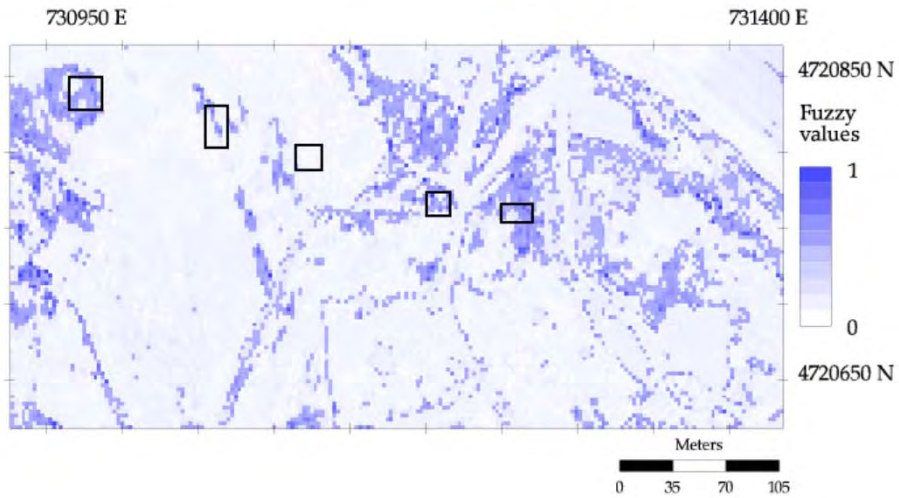
Fuzzy classification maps representing the likelihood of CO_2 leakage on the surface for Latera data (Study areas 1 and 2):



Study area 1 leakage likelihood result using hyperspectral data



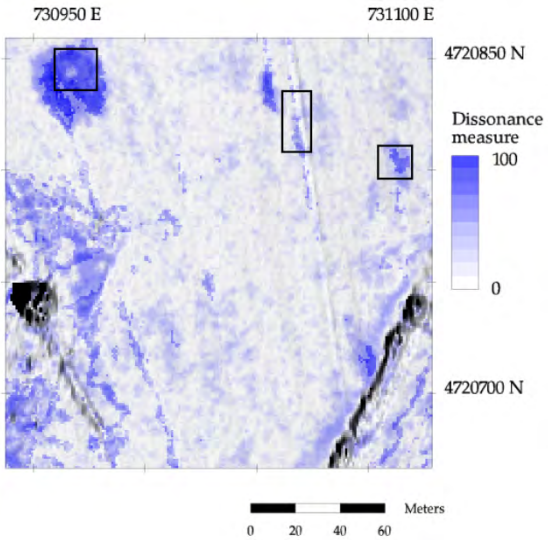
Study area 2 leakage likelihood result using hyperspectral data



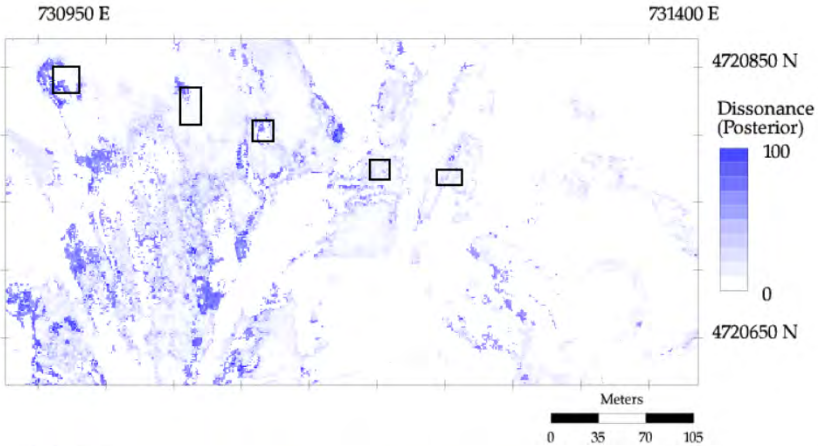
Study area 2 leakage likelihood result using multispectral data

With use of prior knowledge in a Bayesian framework we can improve results

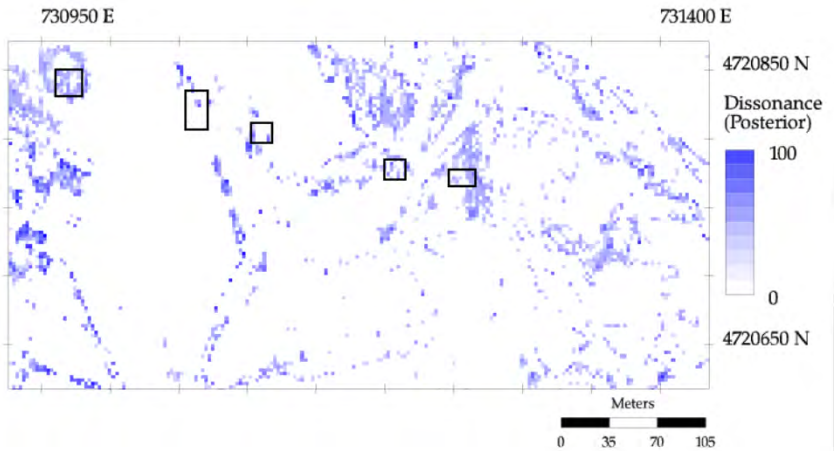
Posterior detection results after applying the DS theory of evidence combination using the prior detection and likelihood maps



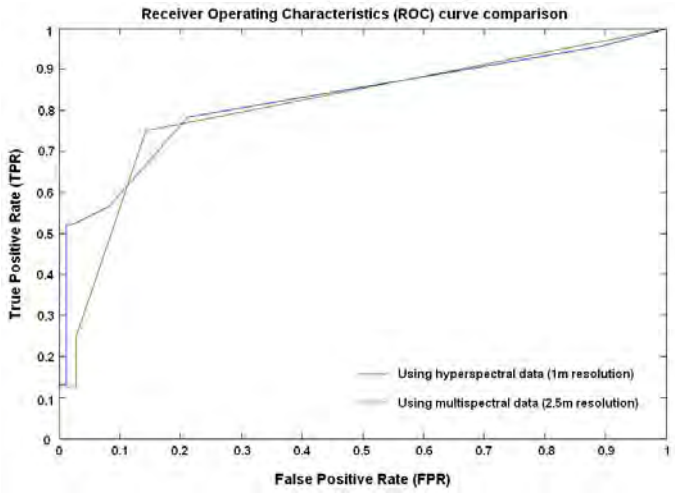
Study area 1 posterior detection result using hyperspectral data



Study area 2 posterior detection result using hyperspectral data



Study area 2 posterior detection result using multispectral data



ROC curves comparison for the methodology using Latera 'Study Area 2' hyperspectral and multispectral datasets.

Decision in the Bayesian framework

We can formulate a **decision rule** e.g. report leakage if posterior **event** occurs

event = odds in favour of leakage are better than 1 in 10

We can calculate the **probability of event assuming there is no leakage** (this is the false alarm probability for the threshold in the odds assumed for event).

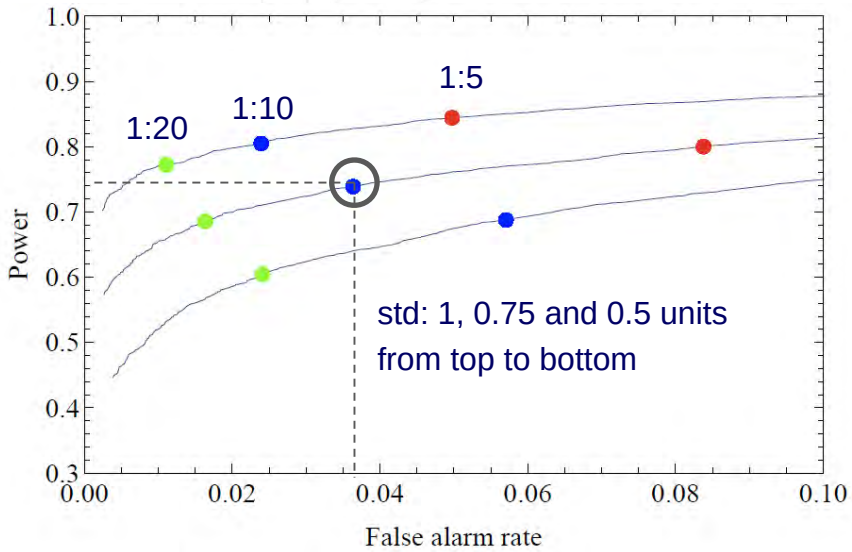
We can also calculate the **probability of event assuming there is leakage**, (which is the statistical power).

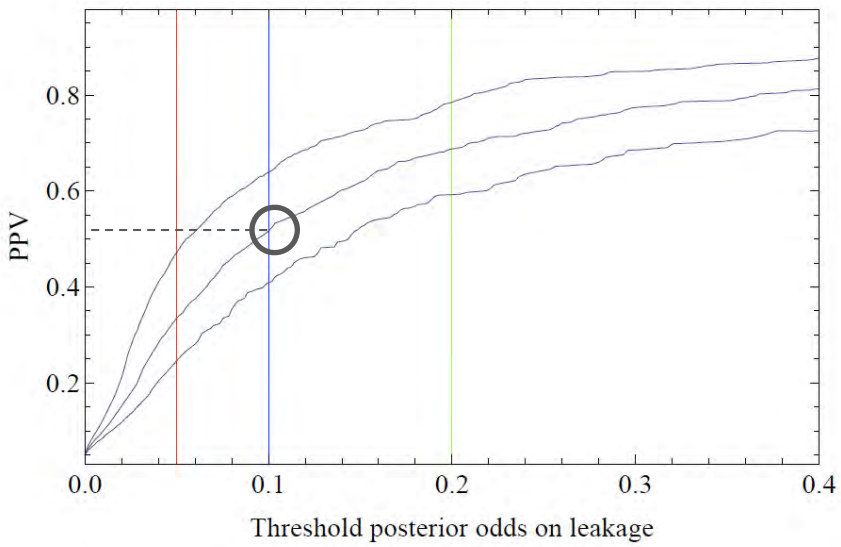
And finally we can combine the **false alarm rates** and **powers** of several **leakage** and **no leakage** models.

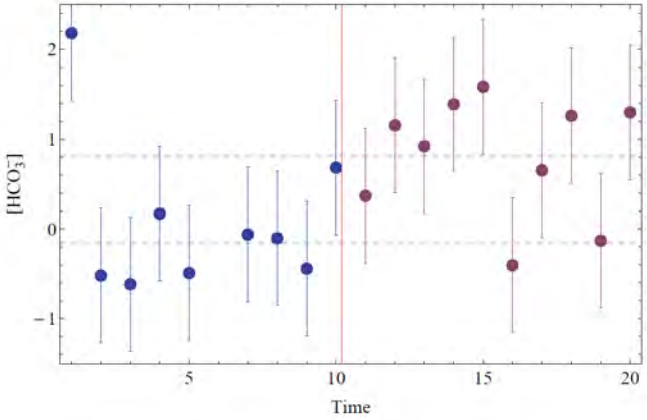
The probability of leakage given that event has occurred is the

Positive Predictive Value (PPV).

Bayesian decision outputs







We can calculate:

- Posterior probability, given the data
- Decision rule: if odds on leakage are better than 1:10, declare leakage.

3.6% false alarm rate, 74% success rate;
need better data

- PPV: if odds on leakage turn out to be better then 1:10 then the probability of leakage is about 0.52.

Conclusions

- Detection of leakage by monitoring inevitably involves **statistical analysis** (detecting the signal of leakage above the noise of various types of measurement and modelling error, dealing with small or non-existent signals in well-designed storage).
- With the **classical hypothesis tests** the only result is the **probability that difference between the pre- and post-injection data has arisen by chance** and is vulnerable to false alarms and highly dependent on data uncertainty.
- Considering a specific leakage model in **the Neyman-Pearson acceptance testing**, allows to evaluate the **probability of taking the correct decision about whether leakage exists or not**. The methodology however still only gives probabilities of the data under the leak and no-leak hypothesis, and deals in only two possibilities.

Conclusions

- Extension to **multiple leakage models**, of varying a priori probability, leads naturally into **Bayesian** terrain.

  It is possible to produce statements of the form “the probability of leakage model #A is 35%, in light of the data.

  At this point “no detectable leakage” has acquired operational meaning with both “leakage” and “detectability” rigorously quantified.

- Common to all broad categories of detection discussed is the need for detailed information in the statistical structure of the data and the uncertainties in that knowledge.
- Accumulation of enough background data may be impractical. However, many statistical techniques are available when dealing with uncertainties.
- False alarm rates are likely to be encountered with data that have very low statistical power to detect leakage.