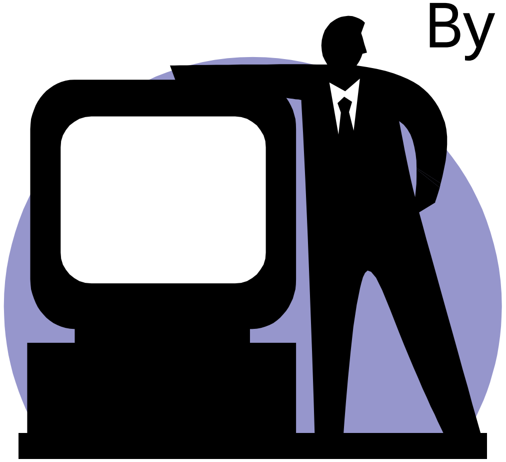
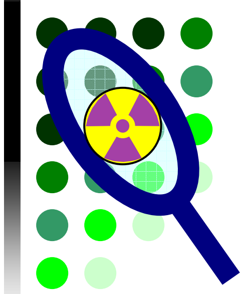
Spectral Analysis Options

Radio Nuclide Identification for Mindless Automatons

By Karl Einar Nelson, PhD





UCRL-PRES-226143

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

Goals

- Present the challenges in automated nuclide identification
- Define common terms required to discuss identification algorithms
- Demonstrate the need for Metrics and Test Benches

Problem Statement

What challenges we face.

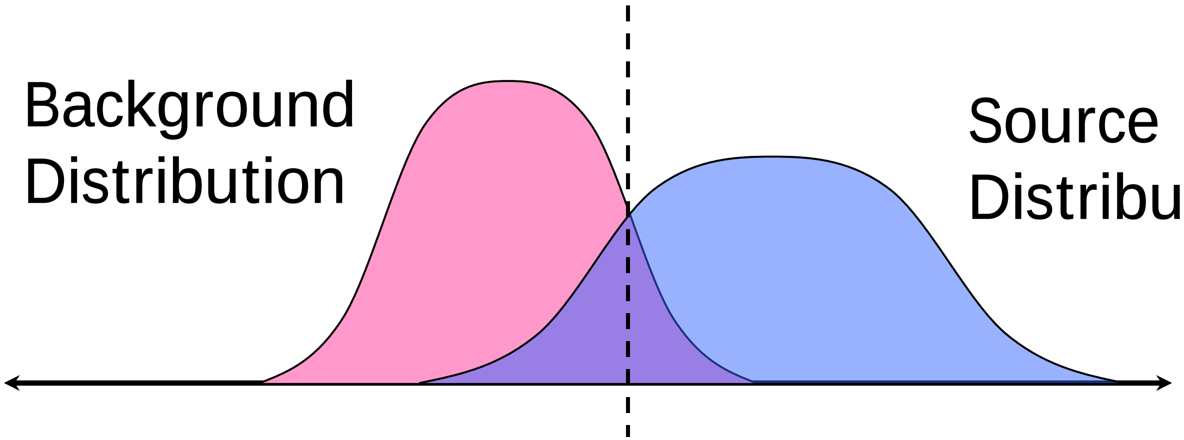
What is the value of nuclide identification?

- The goal of any protection system is to recommend a course of action to maximize the probability that a threat will be caught.
  - Minimize unnecessary searches of non-threatening material
  - Raising response appropriately on potentially threatening sources

- The limiting resource is the number of secondary inspections
- The brightest sources are either RDD potentials or medical sources
- Nuclide Identification is simply a means to achieve our ends

Statistical Definition

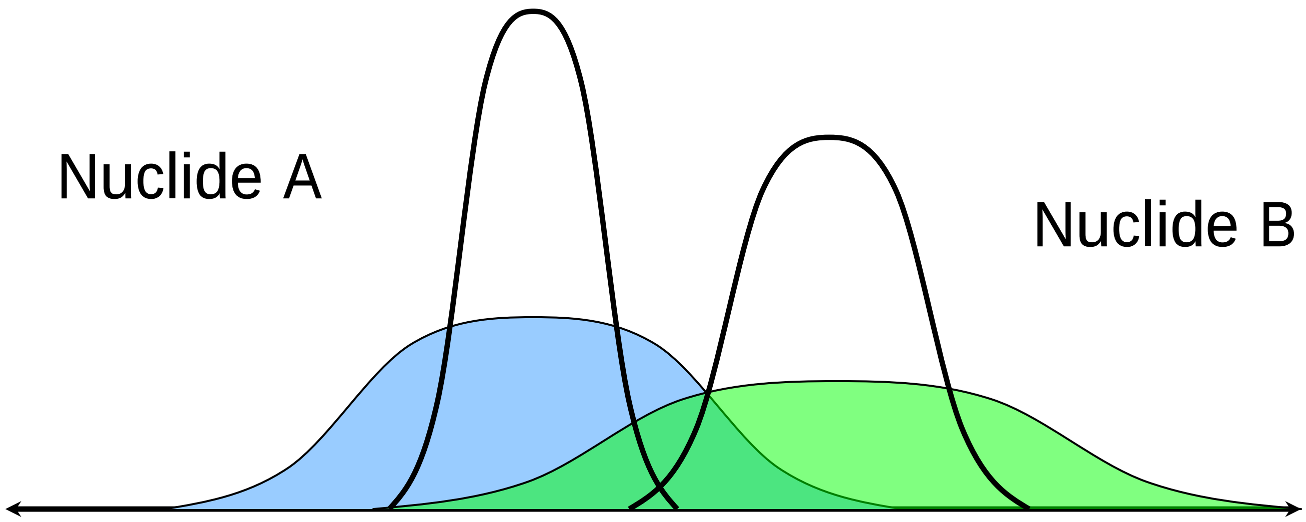
Errors



- False Negative
  - Incorrectly omitting a nuclide that is present
  - Allows a threat to pass
- False Positive
  - Incorrectly reporting a nuclide that is not present
  - Results in an unnecessary search

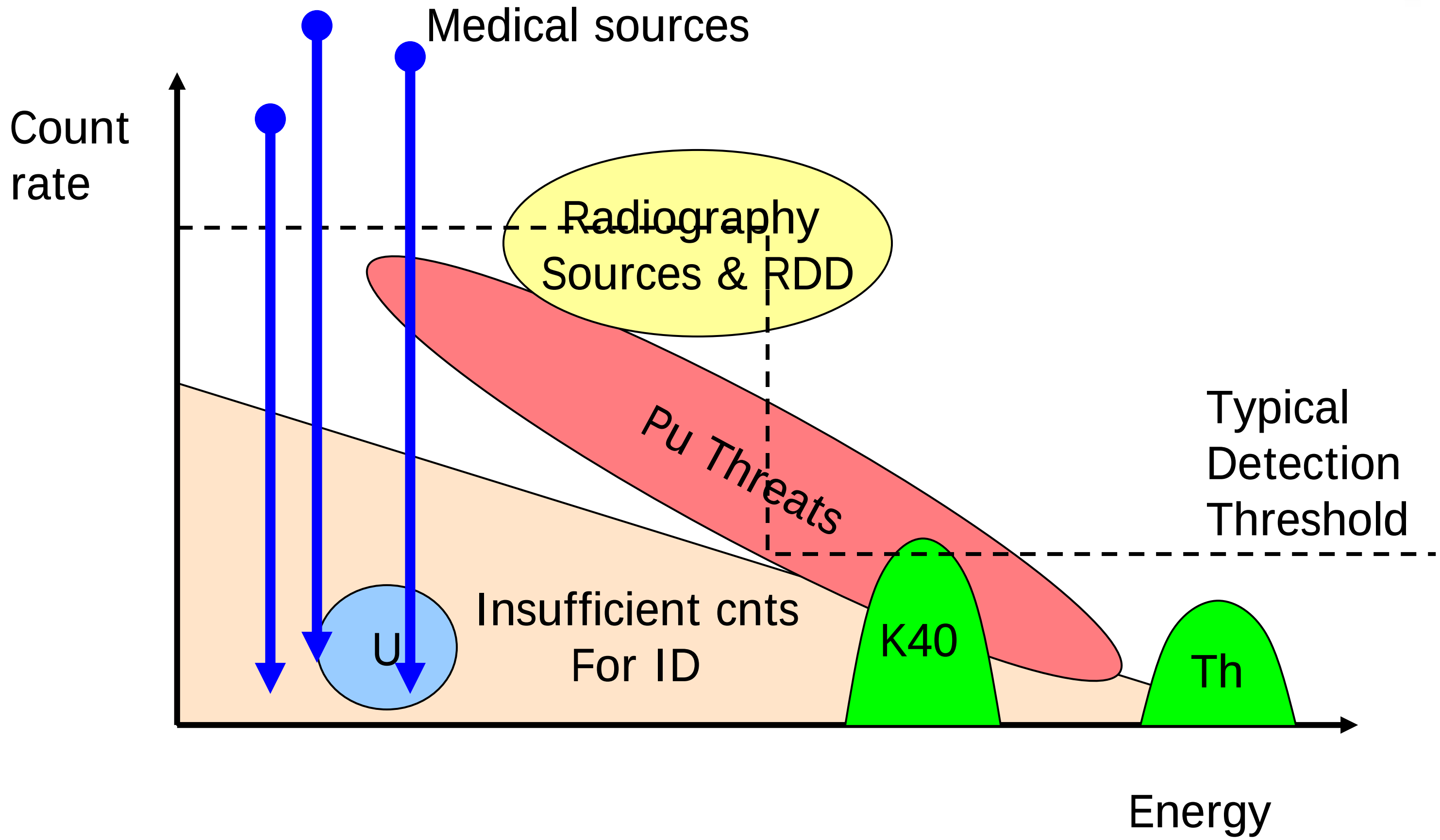
- Errors
  - Result from overlap in statistical distributions
  - Can be minimized but never eliminated

Ambiguity



- Ambiguity – an observation can be reasonably interpreted in more than one way.
  - As the source counts are reduced distributions grow wider
  - Nuclides become more difficult to separate
  - Numbers of errors increase

Domain
Source Sample Space



Risk Categories

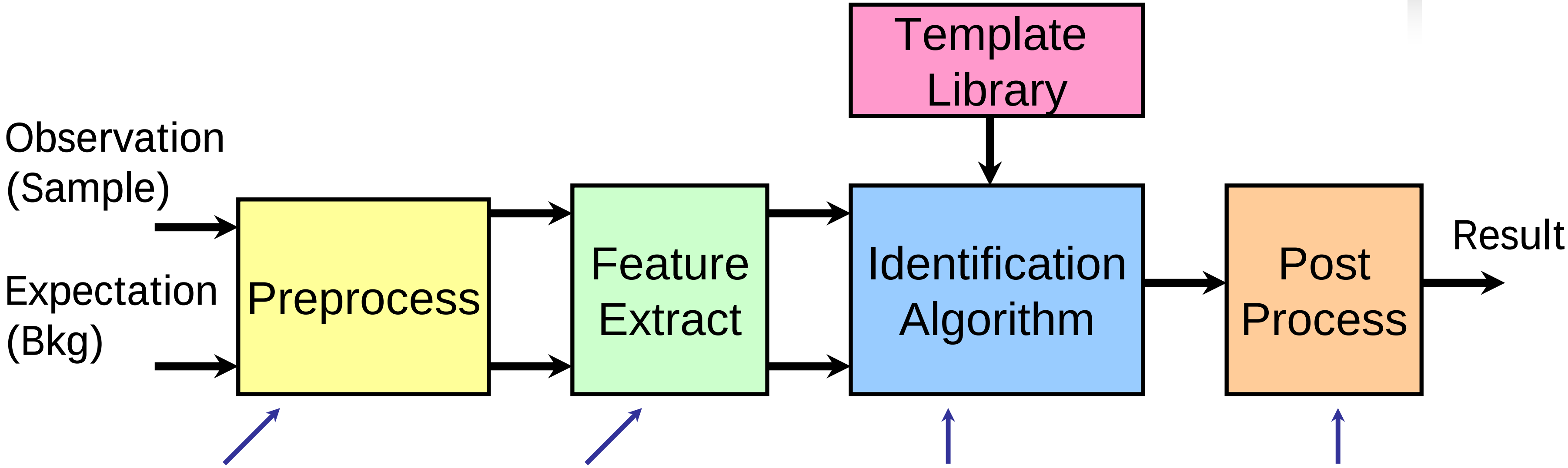
HIGH Must do secondary inspection	**MEDIUM**	**LOW** May avoid secondary inspection
•High confidence threatening ID •Highly shielded ID (RDD) •High count rate, no ID •Suspect Mixture	•Ambiguous ID with more counts than expected from background	•Apparent Background •High confidence non-threatening sources

Improvements in spectral id => Fewer errors + Minimize ambiguous cases => Reduced number of unnecessary secondary inspections

Identification Components

What makes mindless automatons tick?

Anatomy of Nuclide Identifier



- Gain estimation
- Minimum rate assertion

- Channel decimation
- PCA
- Peak extractor

- Classifier
  - Distance metric
  - Decision function
- Estimator
  - Objective function
  - Iterative Solver
- Heuristic

- Significance thresholding
- Confidence computation
- Alarm Categorization

Template Algorithm Types

CLASSIFIERS

- Asks

  What template best fits the observation?
- Properties
  - Produces a list of templates ranked by fitness to the observation
  - Template library must fully span input domain, including naturally occurring mixtures

ESTIMATORS

- Asks

  What weighted sum of templates best represents the observation?
- Properties
  - Produces a weighting vector
  - Mixture can span into novel and unanticipated situations
  - Best solution may not represent the true solution (No ranking)

CLASSIFIER
- Has Ranking
- Limited Solutions

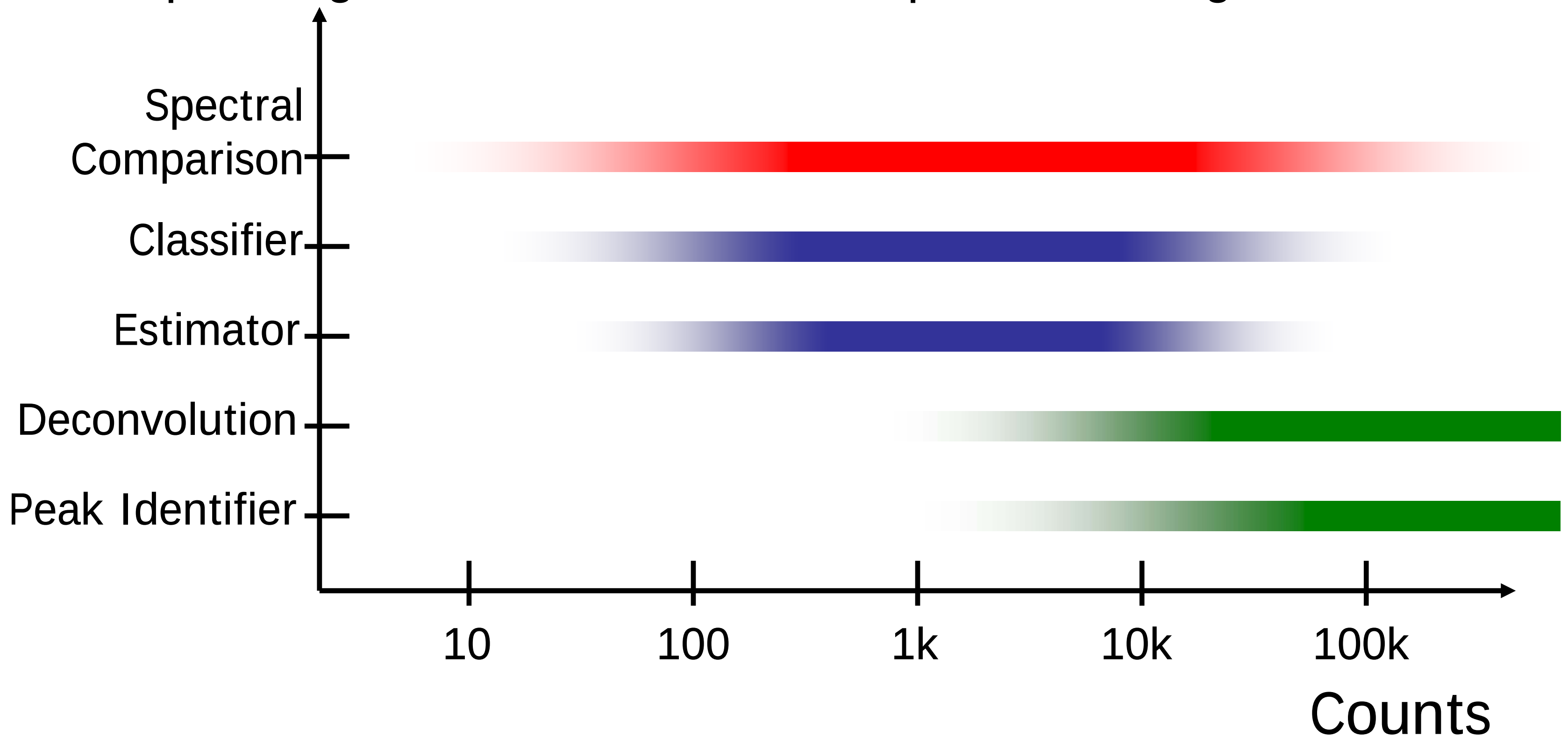
ESTIMATOR
- Lacks Ranking
- Unlimited Solutions

Identification Methods

- Feature Extractors
  - **Channel Decimation**
  - **PCA**
  - Peak Extractor
- Feature enhancers
  - **Deconvolution**
- Other
  - **Multi-detector Spectral Comparison**
  - **Asymmetric Detector Nuclide Identification / Deconvolution**
- Identification Algorithms
  - Classifiers
    - Neural Networks
    - Bayes Classifier
    - **Nearest Neighbor**
  - Estimators
    - Multiple linear regression (Fittodb)
    - **Gauss-Newton**
    - **Expectation Maximization**
  - Heuristics
    - Energy Banding
    - Peak identification
  - Other
    - Shielding Estimators (MBS)

Domain of Nuclide Identifiers

- As number of counts increase, un-modeled effects become visible
  - Ground Bounce, Scattering in Environment, Age, Trace isotopes
- Template algorithms fail due to inadequate modeling



Performance Testing

How well do automatons work under pressure?

Characteristics of methods

	GADRAS	PCA Classifier	MLG
Type	Estimator	Classifier	Hybrid
Algorithm	MLR	PCA/MSE (dot prod)	Gauss-Newton ML/LBF
Metric	Chi Squared	Poisson Likelihood	AIC
Groups/Trials	5	528	25
Nuclide Output	Composition	1 Library Element	1 Nuclide Group
Reports "No Detection"	Yes	Yes	Yes
Reports "Unknown"	No	Yes	Yes
Additional Output	SNM catagory	Sorted Library List	Fitness of each Isotope Group
Run Speed	1 – 3 seconds	0.5 seconds	0.25 seconds

Evaluation methodology

- Evaluated on two sets
  - Verification – library elements as inputs to algorithm
  - Test – set of similar inputs representing real world inputs
- Ran 100 trials for each using Poisson random draws
- Varied the expected input signal from 50 to 1000 counts for test set
- Background added to sample was Poisson random draw with expected 400 counts
- Reference background was Poisson random draw with expected 120000 counts
- Produced
  - Score based on exact matches or inclusion
  - Cross Correlation Matrix of outputs produced by sample
  - Threat class based assignment

Verification results

- When given library elements and asked figure out which one, we expect nearly perfect performance.
- We didn't get it.

Algorithm	False Negative	False Positive
GADRAS	29%	34%
PCA	10%	10%
MLG	7%	6%

(at 300 cps)

Verification results

- PCA and MLG both performed significantly better than GADRAS (1/3 total number of errors)
- Some items failed to identify or were labeled as mixtures by all methods.

^{232}**Th** ⇔ ^{228}**Th**
^{201}**Tl** ⇔ ^{139}**Ce**
^{226}**Ra** ⇨ ^{252}**Cf**

^{192}**Ir** ⇨ ^{237}**Np**
^{67}**Ga** ⇨ ^{237}**Np**
^{133}**Ba** ⇔ ^{131}**I**

Implies library contains indistinguishable nuclides that cannot be resolved with NaI resolution.

Choice of test cases

- Sources
  - Most commonly detected medical isotopes
    - ^{99m}Tc
    - ^{131}I
    - ^{201}Tl
    - ^{67} Ga
  - Common industrial and potential RDD, ^{137}Cs
  - Natural nuisance sources
    - Bananas
    - Wood
    - Fertilizer w/ potash
  - Weapon surrogates with and without lead shielding
    - HEU
    - Virgin HEU
    - Weapons grade Pu
    - Reactor grade Pu

Correlation Result for Tests

Outputs are 90% identification point with noted errors.
Best is overall performance at all count levels.

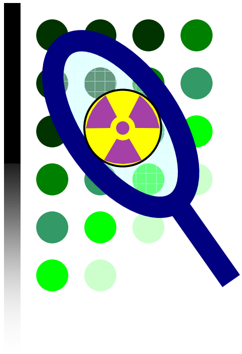
Isotope	GADRAS	PCA	MLG	Best
^{99m}Tc	100	100	100	PCA
^{131}I	1000 ^{133}Ba	1000 ^{133}Ba	1000 ^{133}Ba	PCA
^{201}Tl	^{139}Ce	300	200	MLG
^{67}Ga	300	^{237}Np 32%	300	MLG
^{137}Cs	200	100	100	MLG
Bananas	300	500	300	GAD
Wood	300	500	500	GAD
Potash	500	^{152}Eu 70%	^{152}Eu 30%	GAD
HEU	300	200	200	GAD
Virgin HEU	300	200	300	GAD
RG Pu	1000 ^{137}Cs	^{40}K 30%	^{137}Cs 12%	MLG
WG Pu	500	300	500	PCA
HEU 1” Pb	^{232}U 84%	^{232}U 55%	^{232}U 55%	none
Virgin HEU 1” Pb	^{238}U 100%	^{238}U 100%	^{238}U 100%	none
RG Pu 1” Pb	^{232}Th 79%	^{232}Th 79%	^{252}Cf 40%	none
WG Pu 1” Pb	^{137}Cs 70%	^{252}Cf 33%	^{252}Cf 90%	none

Conclusions

The bottom line

Conclusion

- What are the problems in nuclide identification
  - Ambiguity, Nuisance sources, Inadequate resolution
- What are the problems in a Classifier and an Estimator
  - Classifier – Restricted to ranked list
  - Estimator – Restricted to one solution
- How well do current technologies work
  - Not so swift
- What is to be gained by defining a test bench
  - We can optimize our methods to improve their performance



Learning from our mistakes

Fool me once shame on you,

Fool me twice shame on me.

LLNL's Template Based Isotopic Identifier

- Produces
  - Optimum fit for single isotope or specified mixtures with arbitrary ratios
  - An "unknown" result in response to novel samples or unexpected mixtures
  - Multiple solutions where more than one possible
  - Significance and Confidence levels for each solution
- Designed for low count spectra (<1000 cts)
- Works with large (2x4x16) NaI detectors
- Uses non-linear Gauss-Newton solver

Verification Result

- Verified by probing algorithm 100 random samples for each template.
  - Samples were drawn to have ~300 counts of source with 334 counts of background
  - Reference 5 minute background provided
- Scored based on false positive and false negative

Un-weighted Results:

False Negative: 134/14300 trials less than 1%.
False Positive: 4217/14300 trials ~30%

- Require real world frequency data to compute weighted score

LLNL

Feature Extraction

Getting the most out of an observation

Channel Decimation

- Method
  - Adds groups of neighboring channels
- Properties
  - Maintains statistical properties
  - Resulting information loss can cause ambiguity
  - Increases statistical significance of each channel
  - Must strike a balance between information loss and improved significance
  - Radiation spectrum has natural grouping by the square root of the channel energy

LLNL

Principle Component Analysis

- Method
  - Transform input along the basis vector representing the greatest variance (most representative)
- Mathematics

$R = PP^T = U^T \Lambda V$	Eigenvalue decomposition
$W = [U_1 \cdots U_k]$	Select K largest eigenvalues
$X = W^T O$	Transform on this basis

- Properties
  - Does not preserve the statistics of the observation
  - Good to reduce high dimensional data sets with minimal loss
  - Can reduce noise if it can not be represented in the feature space

Classifiers

What is the best match?

Nearest Neighbor Classifier
Dot Product

- Purpose
  - *Compute the nearest template using L2 norm to the observation.*
- Mathematics

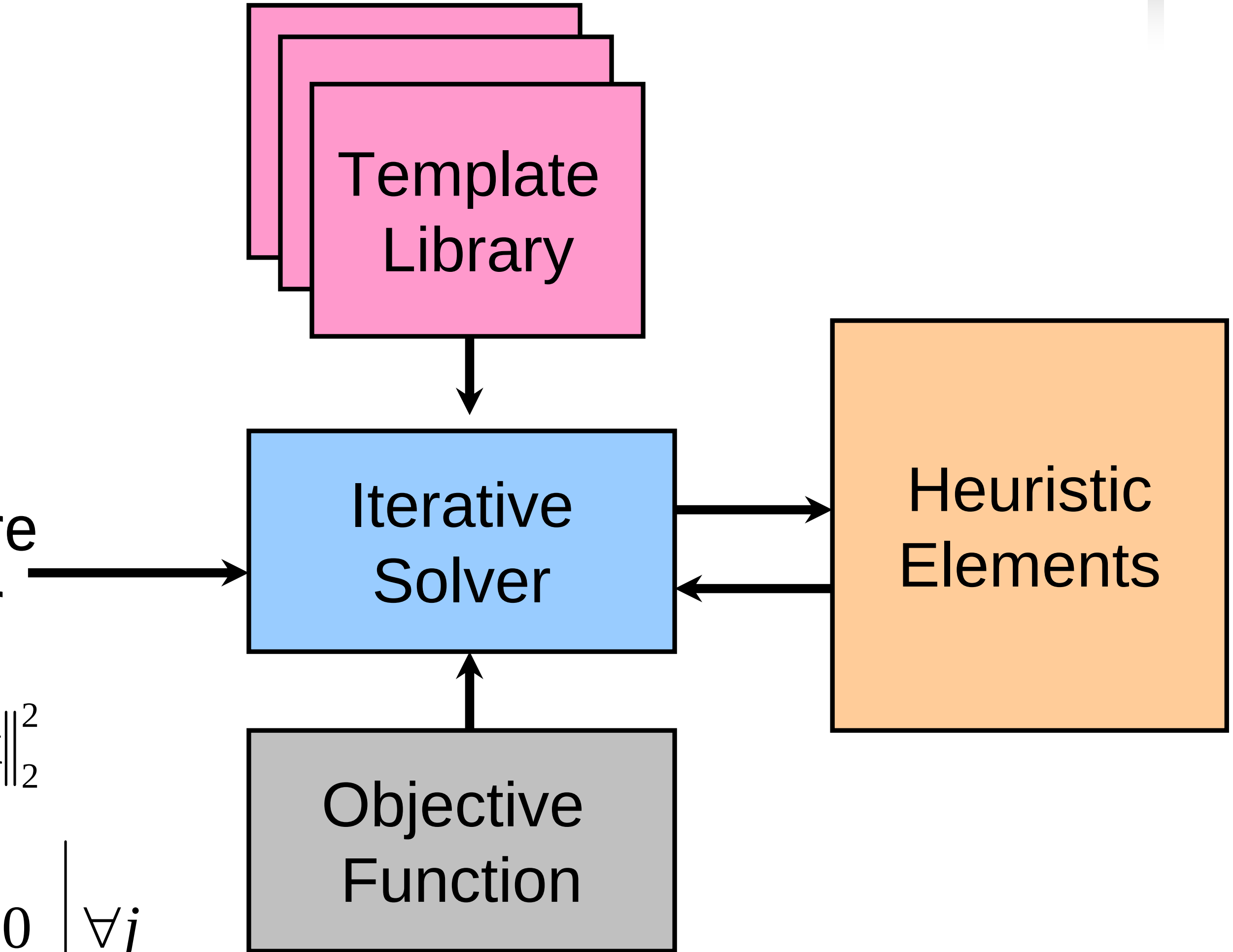
$$\underset{\text{minimize}}{\varepsilon = \left\| O - E - kT_i \right\|_2^2} \longrightarrow \underset{\text{maximize}}{\varepsilon = \left(O - E \right) \cdot \frac{T_i}{\left\| T_i \right\|}}$$

- Properties
  - Shaped based spectral analysis frequently paired with PCA.
  - Computationally inexpensive for each template.
  - Requires many templates to fully span input domain.
  - Does not account for statistical nature of noise.
  - Does not allow for fluxations in background conditions.

Estimators

How much of each?

Anatomy of An Estimator



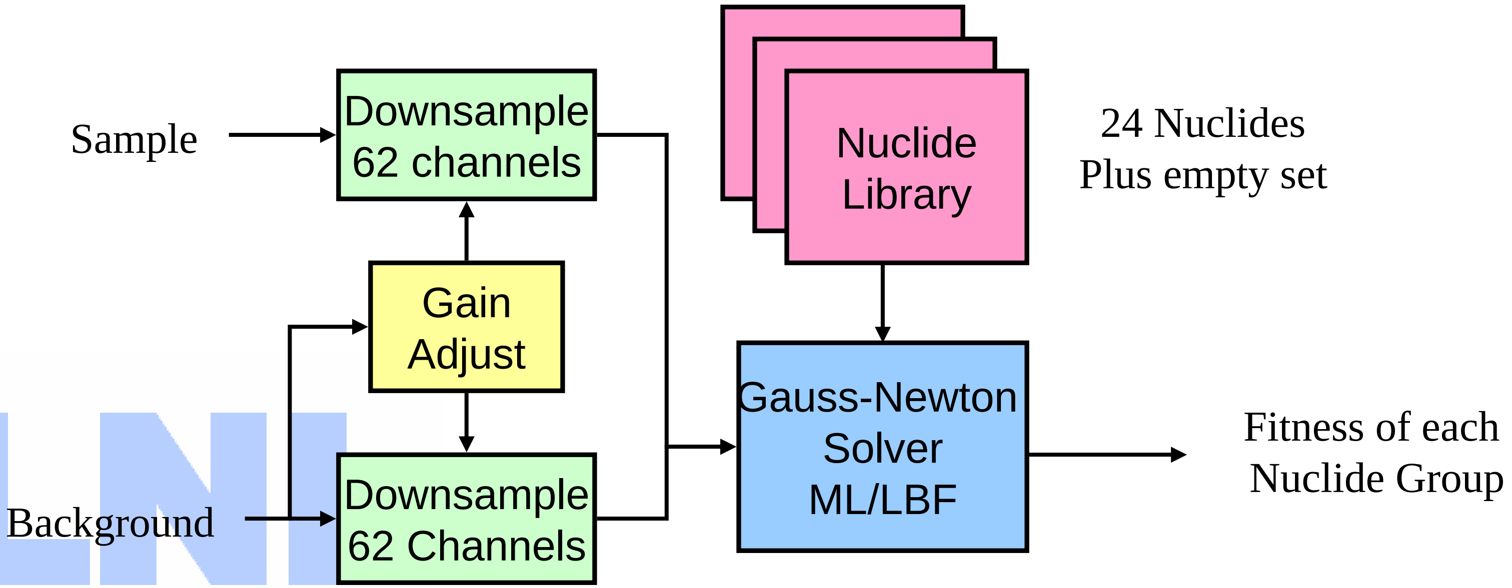
MSE

$$\varepsilon = \frac{1}{2}\left\| O - \mathbf{A}\mathbf{x} \right\|_2^2$$

$$f(\mathbf{x}) = \frac{\partial \varepsilon}{\partial x_j} = 0 \quad \Big| \quad \forall j$$

Maximum Likelihood of Subgroups

- Estimator using Poisson Statistics and Maximum Likelihood Nonlinear Solver
- Requires Logarithmic Barrier Function (LBF) to impose fully positive constraints
- Computes Likelihood of best fit for Background plus each set of nuclides (all shieldings)

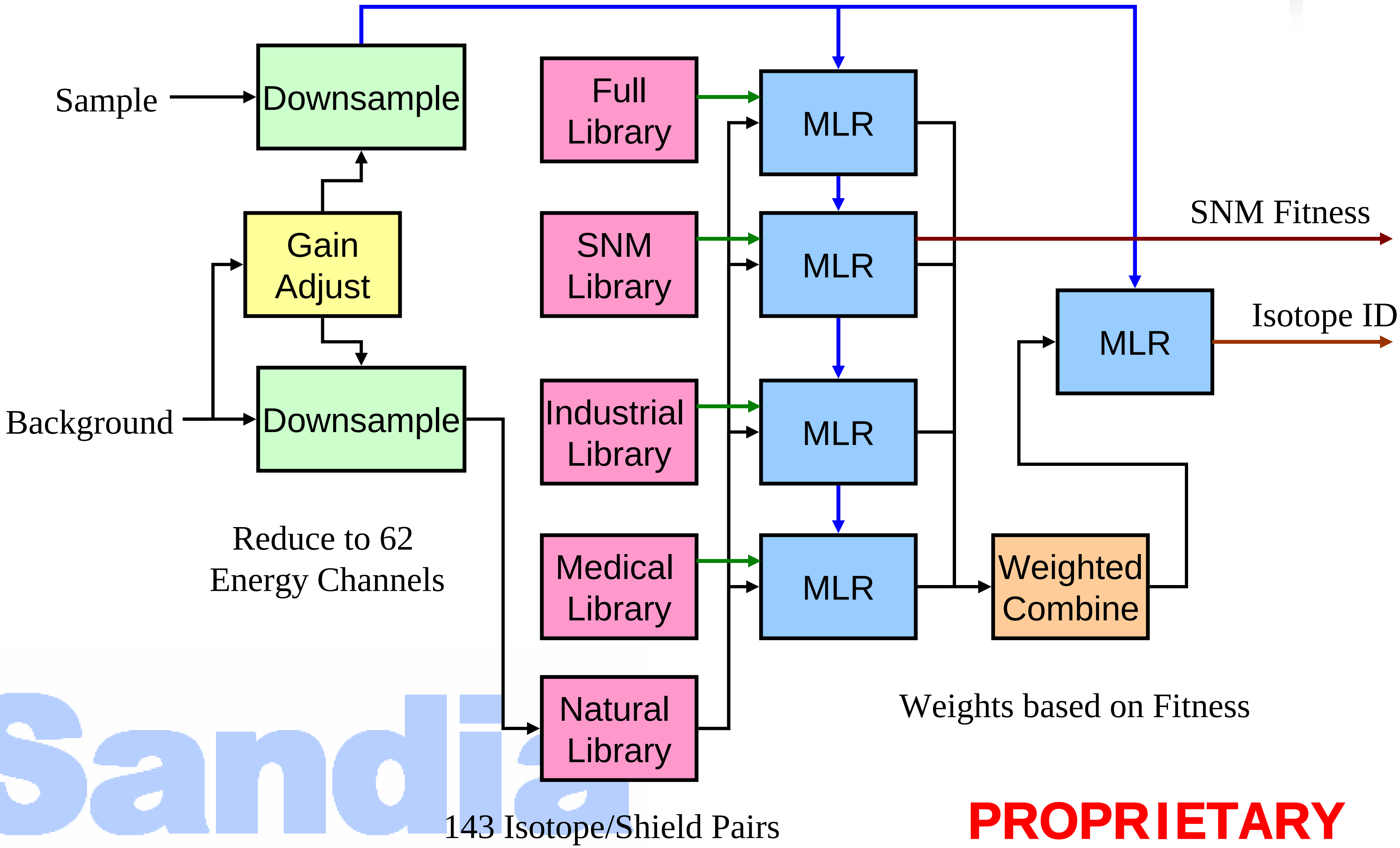


Anatomy of fittodb (GADRAS)

- Core is Heuristic guided MLR (multiple linear regression)
- Objective Function is χ^2 based on 7 variance estimations
- Breaks the library into 4 classes for subgroup analysis
  - Natural
  - Medical
  - Industrial
  - SNM
- Produces
  - Fitness of SNM
  - Nuclide Identification (only reports “significant contributors”)
  - Significance (reduced χ^2)

Estimator Algorithm

Anatomy of fittodb (GADRAS)



PROPRIETARY

Other Techniques

One of these things are not like the others.

Shielding Estimator

Materials Basis Set

- Purpose
  - Estimate the shielding located between a known nuclide and the detector.
- Mathematically

$$b_{z,i} = -\ln\left({}^{c_{z,i}} \middle/ {}_{c_{z,ref}} \right) \qquad U = \begin{vmatrix} b0,ref & b0,z1 & \cdots & 1 \\ b1,ref & & & 1 \\ \vdots & & \ddots & 1 \\ bn,ref & & \cdots & 1 \end{vmatrix} \qquad b = U \cdot \rho$$

- Properties
  - Uses Beer's Law ($A=\varepsilon lc$) estimate shielding thickness
  - Works in a log space
  - Assumes that increased shielding decreases the mean counts in a channel

LANL

Materials Basis Set

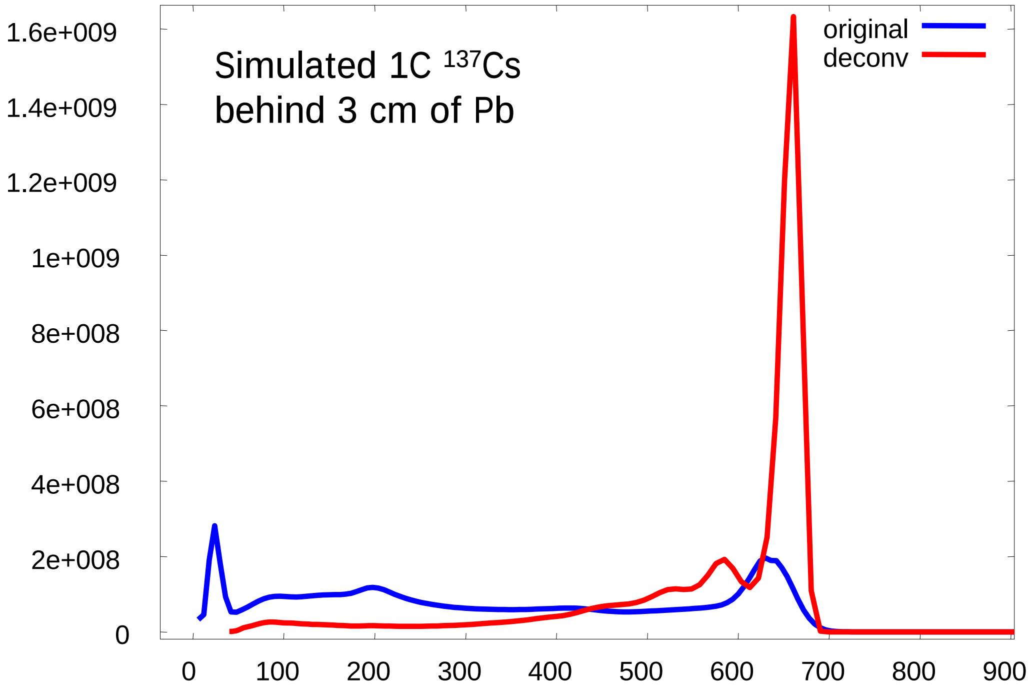
- Primary functions on the photo-peaks
  - Compton scatter breaks assumptions thus is biasing the solution
- Working in a logarithmic space makes mixture analysis difficult.
  - Nuclide signatures add linearly
  - Complexity of estimator will increase as N^2 as additional nuclides are included in the mixture
  - Perturbations from Compton Scattering likely to bias mixture solution
- Assumes χ^2 statistics on reconstructed spectra

Deconvolution

- Purpose
  - Compute the energy flux at the surface of the detector from the observation
- Properties
  - Requires relatively large numbers of counts to produce a quality result (10k)
  - Requires extremely well tuned detector model
  - Removes virtually all detector effects
    - Compton Scatter
    - Backscatter peaks
  - Decreases the FWHM of the detector by a factor of 4
  - Allows isolation of peaks within ½ of the FWHM

NaI Deconvolution

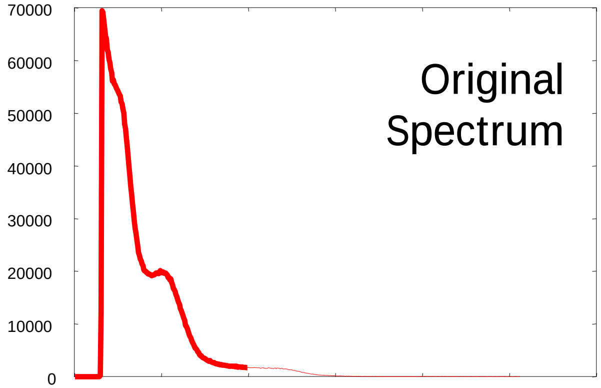
- Deconvolution using EM algorithm
- Allows direct evaluation of shielding from Compton scattering or branching ratios
- Removed backscatter peak
- Recovered Compton Scatter in the detector to the photo-peak



LLNL

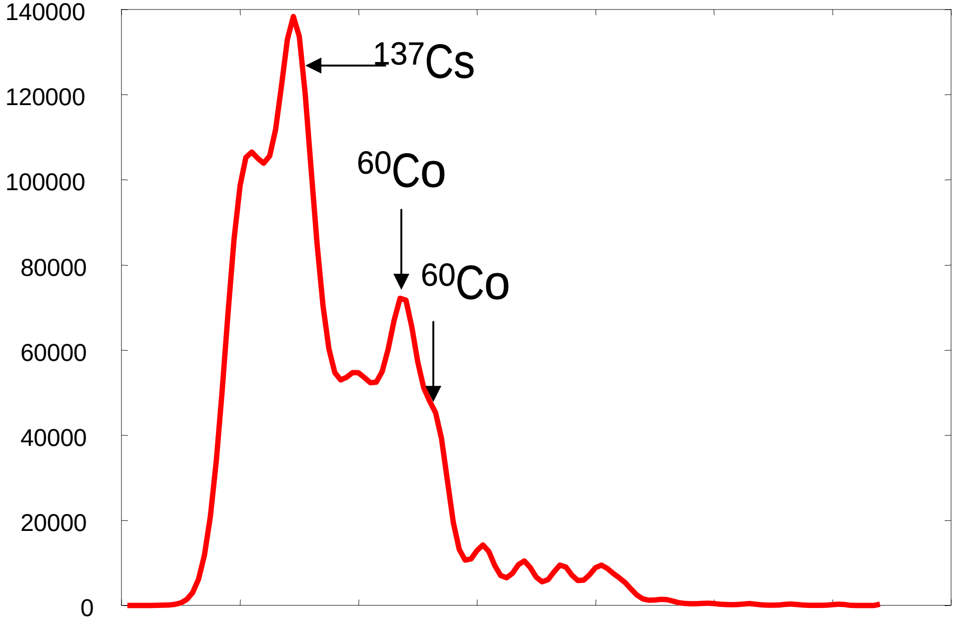
Spectral Enhancement

Plastic Deconvolution



- Deconvolution using EM algorithm
  - 128 bin energy square root energy
  - Gaussian basis

- Provides
  - Peaks otherwise not visible can be identified
  - Allows calibration of energy scale



LLNL

Multi-detector Nuclide Comparison

- Purpose
  - Uses multiple detectors to improve identification or match spectra between previous encounters
- Properties
  - Simultaneously solve detectors by information sharing or summing
  - Depending on the number of energy channels used can operate with less counts at a particular encounter than possible otherwise
  - Demonstrated in DTS to improve nuclide identification for high speed traffic by using detections from multiple locations

LLNL

Asymmetric Detector Nuclide Identification/Deconvolution

- Purpose
  - Using two detectors with different resolutions and efficiencies to improve identification.
- Properties
  - Simultaneously solves identification in both detectors by information sharing
  - Complexity increases at number of detectors squared
  - Pairs small high resolution (HPGe, SiLi, CZT) detector with larger low resolution detector (NaI, Plastic)
  - Creates hybrid detector with resolution and efficiency somewhere between the two

LLNL

Non-negative Iterative Solvers

- NNLS - Lewson and Hanson
- Logarithmic Barrier Function – method of generalized inequality constraints
- Expectation Maximization – uses probability density function that are positive by design

Statistics

What do these silly numbers all mean?

Model Confidence (Trust)

- Purpose
  - *Indicates how frequently a specified model is a correct identification.*
- Mathematics

$$C_{M_i} = \frac{P(O \mid M_i)P(M_i)}{\sum_{M_j} P(O \mid M_j)P(M_j)}$$

- Properties
  - Depends on immeasurable quantity P(M)
  - Requires the sum of all possible solutions that may not be computable in finite time.
  - At best we can compute a empirical estimate of confidence or psuedo-confidence

Nuclide Confidence

- Purpose
  - *Indicate how frequently a specified nuclide is a correct identification.*
- Mathematics

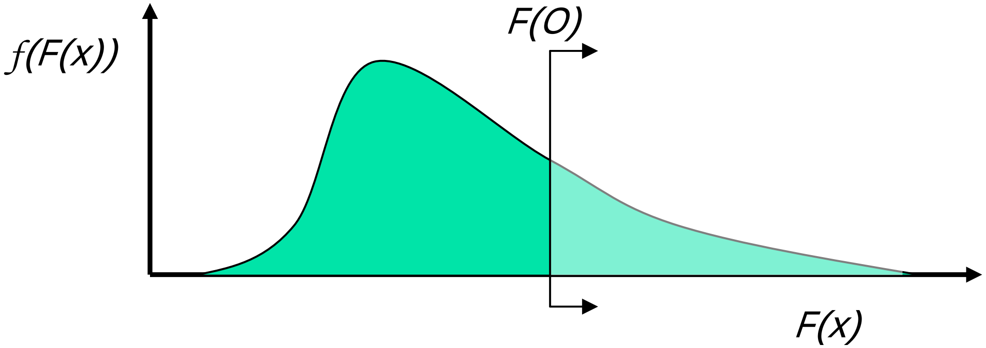
$$C_X = \frac{\sum_{X \subset M_i} P(O \mid M_i) P(M_i)}{\sum_{M_j} P(O \mid M_j) P(M_j)}$$

- Properties
  - As with Model Confidence, depends on immeasurable quantities. Thus we can at best estimate it.

Significance

- Purpose
  - *Indicates how frequently a similar observation may be produced from a model.*
- Mathematics

$$P(F(x) > F(O) \mid M)$$



- Properties
  - Evaluated by either a null hypothesis test (χ^2) or by likelihood ratios.
  - Used as the objective function in estimators.
  - Well defined for radio-nuclide identification problem.
  - Controls the rate of Type II errors.