URS

Washington TRU Solutions LLC

A new statistical analysis technique for air monitoring

Robert Hayes, PhD, CHP, PE

AMUG 2012, Las Vegas, NV

Introduction (historical methods)

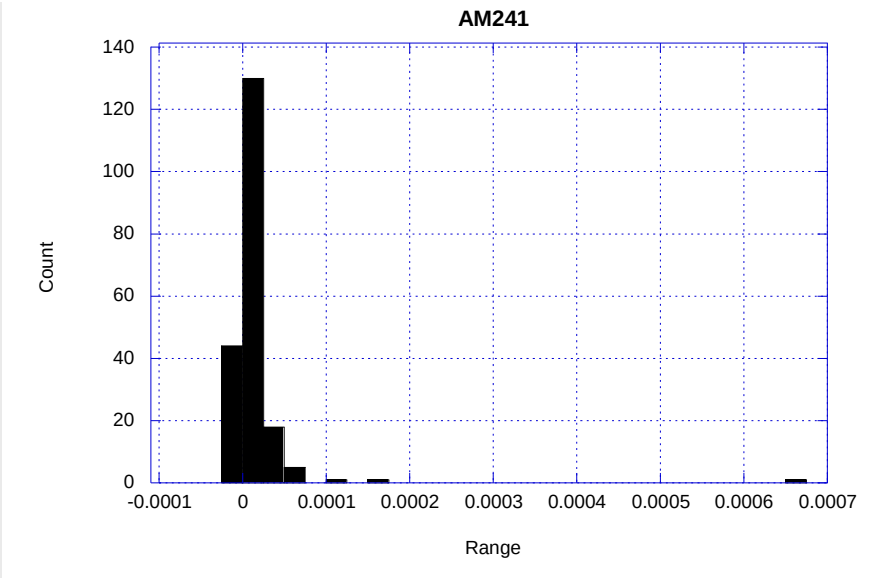
URS
Washington TRU Solutions LLC

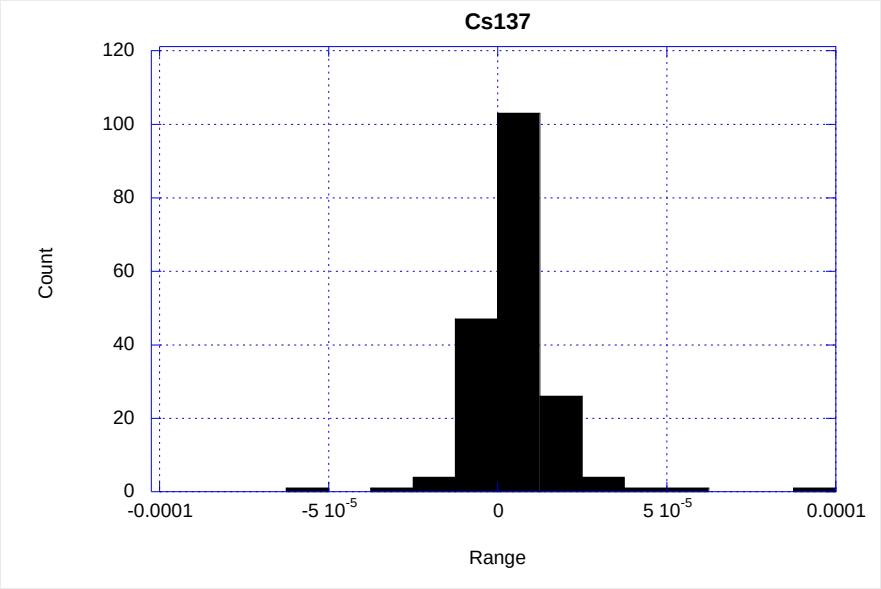
- Histogram (traditional)
  - Dependent on bin sizes and frequencies
- Time series plot
  - Density of occurrences may warrant averaging of some kind
- Analysis of Variance (ANOVA)
  - Can be tedious to identify failing groups and their statistical significance
- Linear fitting (least squares, weighted least squares)
- Cumulative distribution plots
- Probability distribution plots
- Chi-squared distribution tests
- Chebyshev (a method previously used for WIPP data sets)

Histogram (traditional)

Washington TRU Solutions LLC

- Am241 (top)
- Cs137 (bottom with one apparent outlier removed)
- Data represent actual off-site assay results
- Distributions do not consider any individual assay uncertainties,
  - This quality is assumed inherent to the distribution only

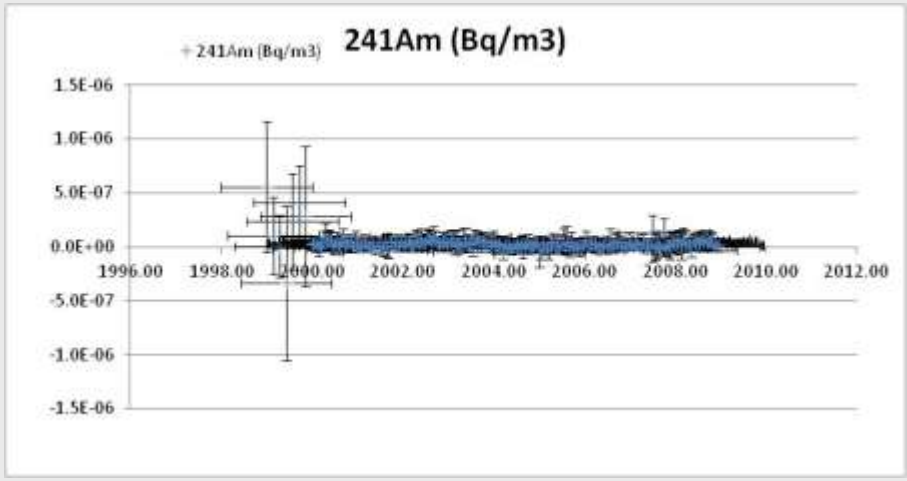


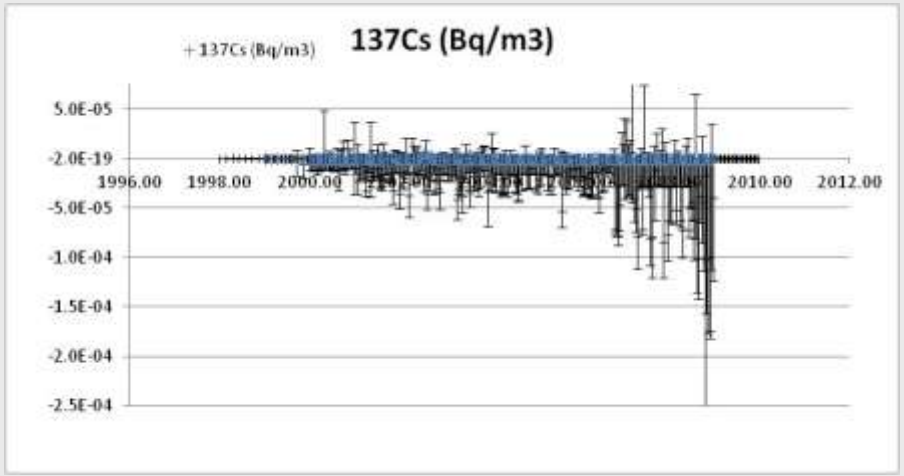


Time series plot

Washington TRU Solutions LLC

- This gives an indication of any temporal trends but more importantly shows measurement error
- Bias and skew can be seen visually if they are sufficiently large
- Not necessarily quantitative in and of itself without additional statistical analysis

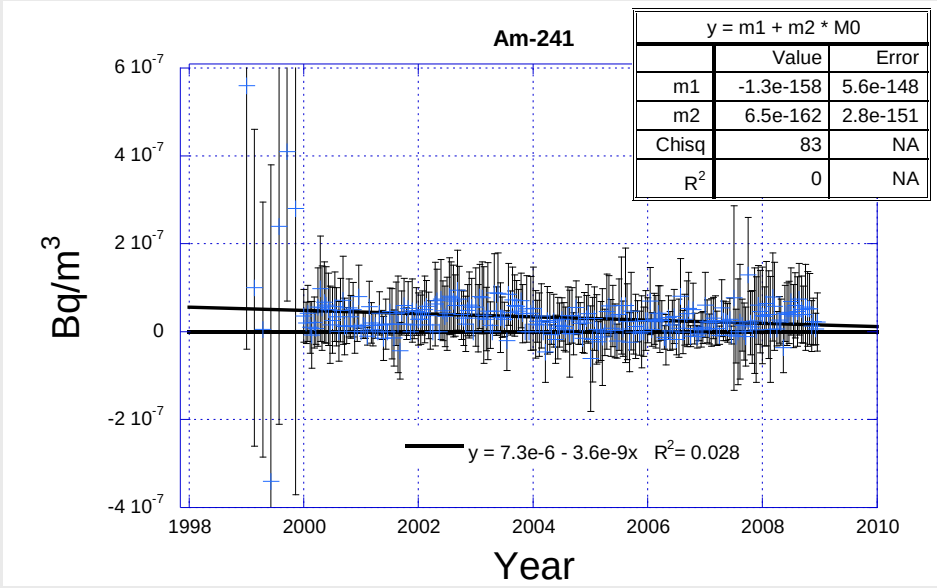


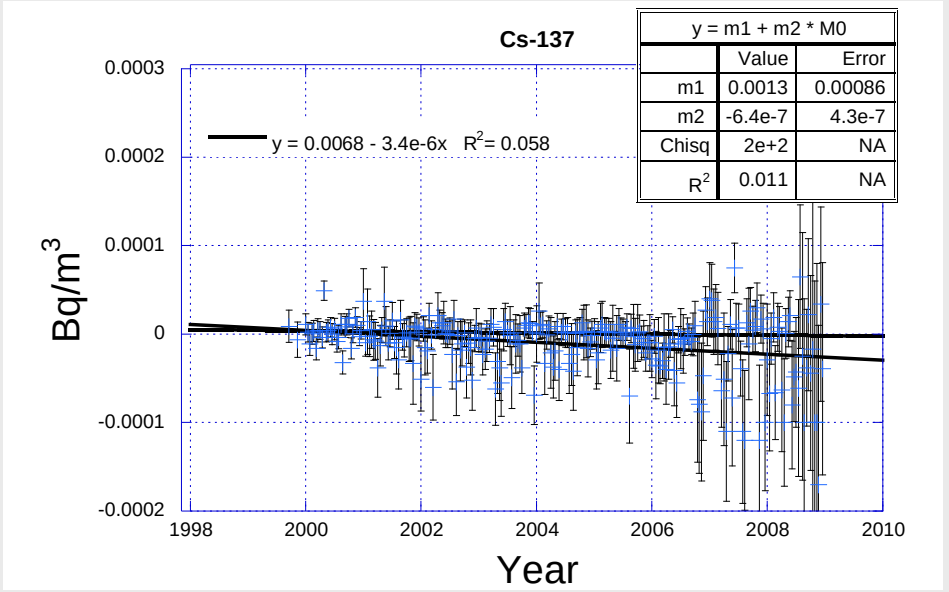


Linear fitting (least squares, weighted least squares)

URS
Washington TRU Solutions LLC

- A more robust type of time series with quantitative test criteria which can use individual measurement errors
- Provides predictive trend data if linear
- Cannot predict distribution type or limits

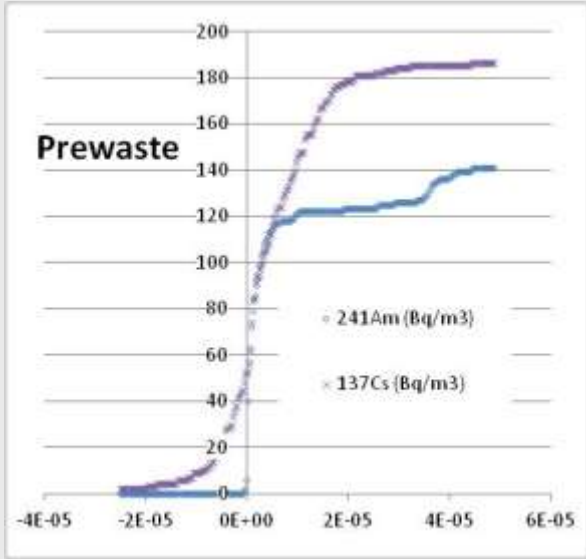


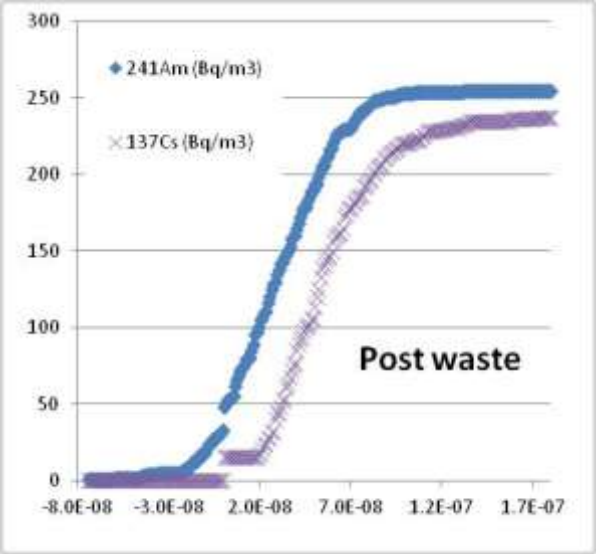


Cumulative distribution plots

URS
Washington TRU Solutions LLC

- Distribution bias can be inferred but not quantified
  - Advantage to histogram is that both bunched and sparse data regions are easily seen
- Does not account for individual measurement error
- Large concentrations of points at a given value or infrequent but large deviations from the mean can hinder interpretation

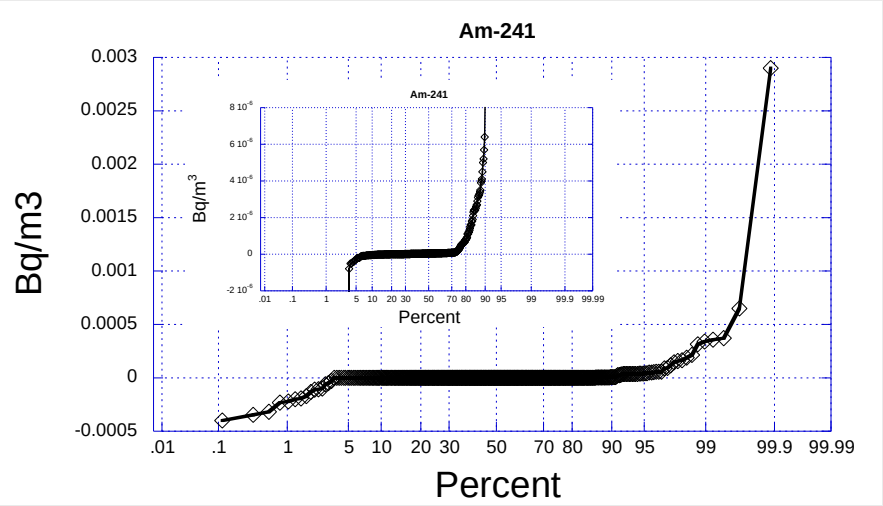


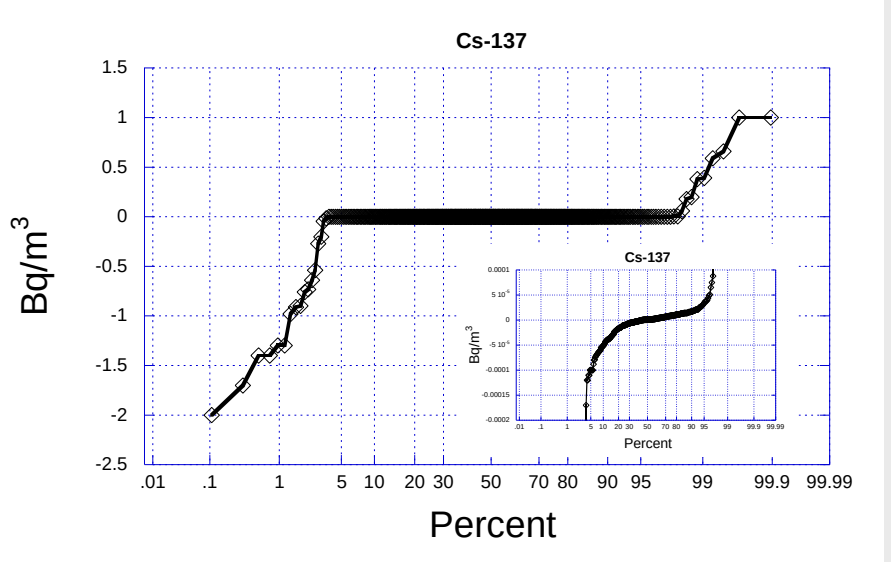


Probability distribution plots

Washington TRU Solutions LLC

- The effect of having data with mixed uncertainty sizes is seen in these graphs
- The small uncertainty data is bunched near the zero
  - How useful is this data, only the large uncertainty data is clear
- Data should be linear for normal distribution





Analysis of Variance (ANOVA)

URS
Washington TRU Solutions LLC

- Does not have a well known method that accounts for individual measurement errors
- Cannot determine distribution limits/types
- Does not a-priori provide a method for which subset was the reason for a failure of the null hypothesis if the result was a failure
- Assumes normal dist
  - Should pass χ^2 test

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	287	0.0026	9.2E-06	3.4E-08
Column 2	149	0.0020	1.4E-05	3.2E-09

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.9E-09	1	1.9E-09	0.079	0.778	3.863
Within Groups	1.0E-05	434	2.3E-08			
Total	1.0E-05	435				

Anova: Single Factor
Cs-137 pre vs. post

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	282	-10.61	-0.038	0.078
Column 2	189	0.0031	0.0000	0.0000

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.16	1	0.16	3.45	0.06	3.86
Within Groups	21.80	469	0.05			
Total	21.96	470				

Chi-squared distribution tests

URS
Washington TRU Solutions LLC

- Tests for a normal distribution using F-test
- Tests for equivalent distribution among separate distributions (such as pre and post waste)
- Not a well known method for testing data with known error values
- If failure of this test, ANOVA assumptions of normal distribution is invalid

F-Test Two-Sample for Variances		
	Am241	
	Variable 1	*Variable 2*
Mean	9.17E-06	1.35E-05
Variance	3.38E-08	3.17E-09
Observations	287	149
df	286	148
F	10.7	
P(F<=f) one-tail	2.71E-43	
F Critical one-tail	1.3	

F-Test Two-Sample for Variances		
	Cs137	
	Variable 1	*Variable 2*
Mean	-3.76E-02	1.64E-05
Variance	7.76E-02	2.80E-08
Observations	282	189
df	281	188
F	2.77E+06	
P(F<=f) one-tail	0.00E+00	
F Critical one-tail	1.2	

Chebyshev's theorem

URS
Washington TRU Solutions LLC

- The proportion of data in any set must lie within *k* standard deviations σ of the expression $1-k^{-2}$
  - σ is calculated in the normal way here
- This expression is fundamentally bounding on the dispersion inherent to any distribution and applies to all numerical data sets when counting the number of σ's
- More values can lie in the interval but no less.
  - This means *p*-values are only bounding lower limits
- Results analyzed up to 2005 showed no WIPP releases using this method
  - Updates to this are in progress and include the present work

Chebyshev's theorem

URS
Washington TRU Solutions LLC

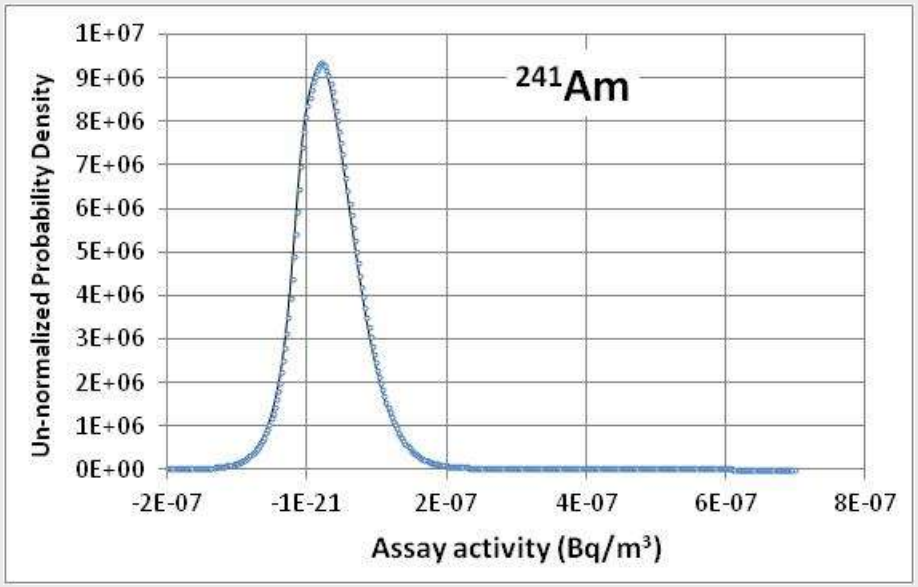
- Only method discussed here having the power to answer the question so far
- Does not however take into consideration individual measurement errors
  - Is this really needed?
  - Doesn't the distribution already contain this information?
- Data sets having large variations in measurement error could be a problem
- It would be nice to have a method which considered individual errors.
- The proportion that must lie within k sigma's is given by $1-1/k^2$

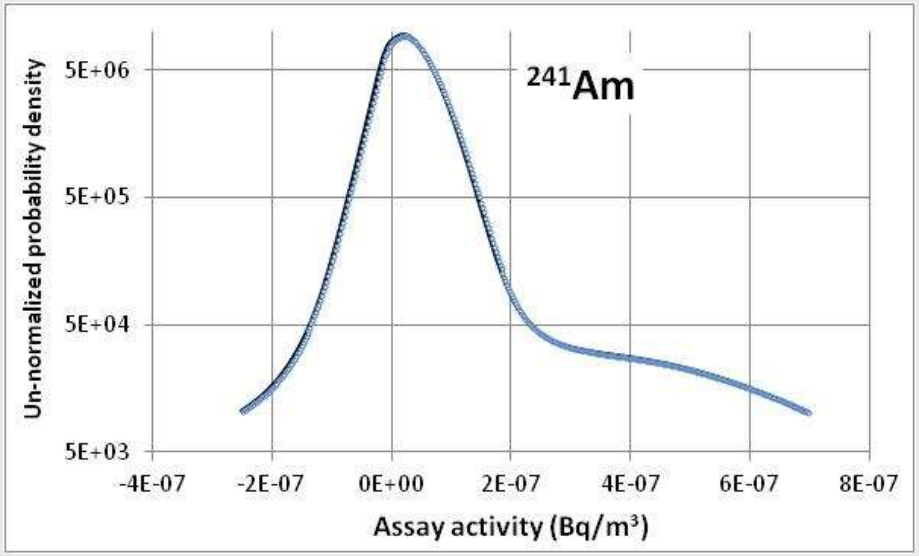
Data falling within plus or minus one standard deviation		
Number of standard deviations	Normal distribution	Chebyshev's Limit
1.01	68.8%	2%
1.1	72.9%	17%
1.5	86.6%	56%
2	95.4%	75%
3	99.7%	89%
5	99.99994%	96%
7	99.9999999997%	98%

Now for the new stuff

URS
Washington TRU Solutions LLC

- The method is a superposition analysis of normalized Gaussian's (SANG).
- Turn each measurement and its error into a normalized (unit area) Gaussian.
- Histogram the resultant normalized Gaussians to obtain a spectrum
  - Accounts for all measurement info





Interpretation Details

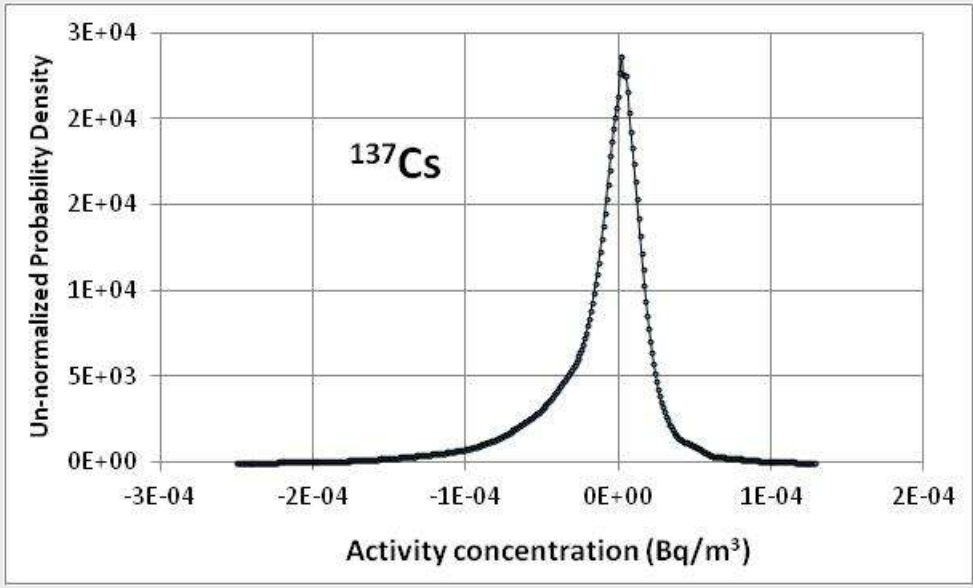
URS
Washington TRU Solutions LLC

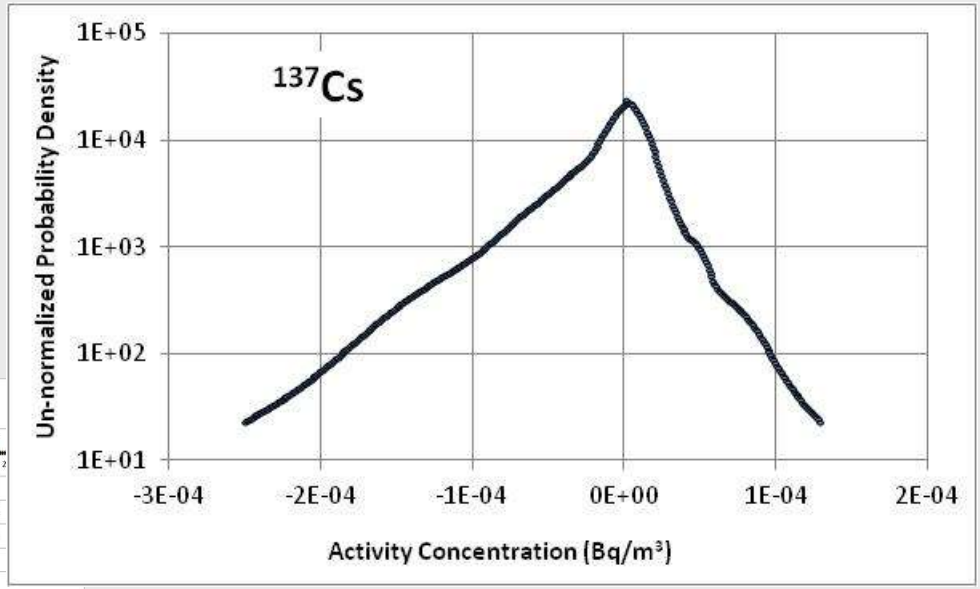
- If you integrate from/to ±∞, you get back the number of occurrences
  - So the abscissa is not a normalized probability density.
  - The distributions are akin to a frequency distribution in this sense
- The area under the curve in any region will give the expected number of occurrences in that region from the data set evaluated.
- To make the distribution a true probability density function, you have to divide each entry of the distribution by the number of events making up the SANG graph.
  - This will give an integrated area under the whole curve of unity
- If you have a normalized SANG, then the area under the curve in any region is the probability of getting a single event in that region

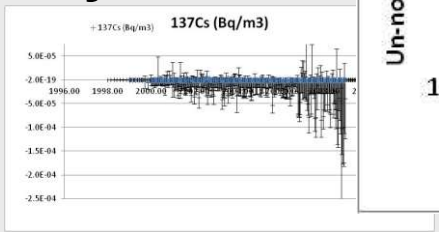
Provides more information than ANOVA or Chebyshev methods

URS
Washington TRU Solutions LLC

- All measurement information is utilized
- Current data is displayed assuming normal errors
- Method requires functionally known error distribution for each measurement
- Tail information is weighted properly



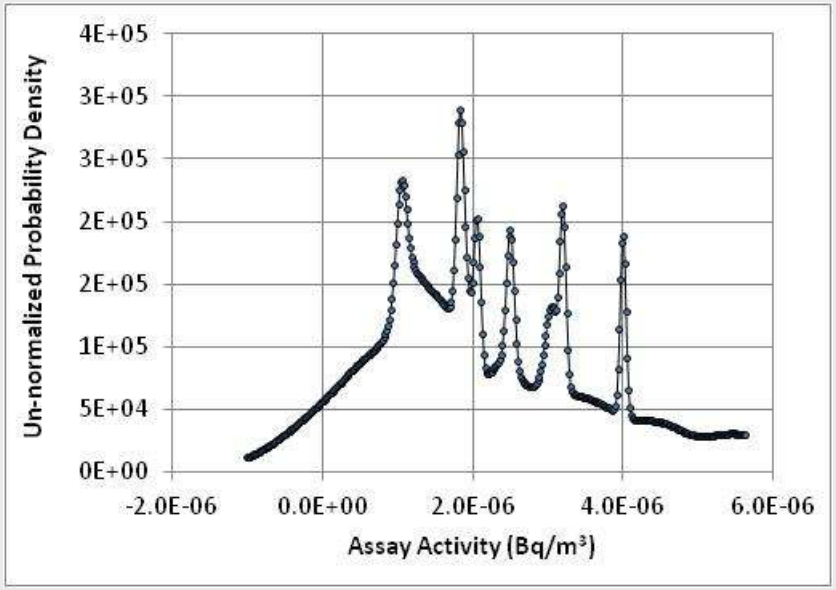




Spectral analysis techniques

URS
Washington TRU Solutions LLC

- Quantitative methods which can be used now include
  - FWHM
  - Peak fitting algorithms of all kinds including those from gamma or alpha spectrometry
  - Probability limits such as *p*-values
  - Outlier detection or uncertainty mistakes
- An example is visual detection of outliers or uncertainty underestimation (see below).



Conclusion

URS
Washington TRU Solutions LLC

- The SANG method was originally utilized for a different application and reported in the ANS transactions but the method was thought up by me (RH)
  - Rhoden W. G. and Hayes R. B. Alternate approach to the upper subcritical limit determination for MCNP. *Trans. Amer. Nucl. Soc.* **89**, 109-110, 2003.
- The technique coupled with Chebyshev's method appear to be adequate for addressing the question at hand.
- Other methods to be used include Cs/Pu ratios for identifying environmental radionuclide's vs WIPP inventory
- All data to date do not contradict the expectation that all effluent Pu measured to date originated from the environment and not the WIPP repository