Answers to Questions Stated in the NRC PowerPoint Presentation (Slide 5)

1. The data points that are most widely dispersed are those with very low sensitivities, and they should be given low weights if they were to be included in regression fits. Regression fits have also been demonstrated to be insensitive to the screening parameter.

2. If all assemblies in a sub-batch have the same first derivative of reactivity with burnup, the intra-batch variation of burnup will have no influence on the measured batch-averaged reactivity. Thus, by choosing the maximum variation of the second derivative and the maximum sub-batch deviation of assembly burnup from the sub-batch averaged burnup, we are maximizing the possible deviation of inferred sub-batch reactivity.

3. 12 pcm is the increase in the 95% confidence interval when the data is fitted with and without treating the intra-batch burnup variation.

4. The burnup values for a cycle/sub-batch are the sub-batch averaged burnups for each flux map - averaged over all flux-maps that exist in that particular cycle of reactor operation.

5. The weights for sub-batch point *i* are assumed to be one over the product of the Figure 3 and Figure 5 fitted standard deviation squared curves, as interpolated to the specific sub-batch sensitivity and average burnup.

$$W_i = \frac{1}{\sigma_i^2(S)} \frac{1}{\sigma_i^2(B)}$$

Questions about NRC Memo

1. For the NRC's EPRIstat analysis, it appears that the prediction interval of the raw data in Fig. 9 was used, in the analysis, rather than the more-appropriate prediction interval for the cycle-collapsed data. Why was the raw data prediction interval used?

2. The Kopp curves in the NRC document seem to only show ~0.015 Δ-k at 54 GWd/MTU, which is only half the amount that we normally expect (Δk of depletion is in range of 0.43-0.60 Δk at 60 GWd/MTU, so 5% corresponds to a range of 0.022-0.030 Δk; See 11 EPRI benchmark Tables C.1 through C.4).

3. If you plotted the real 5% Kopp curves, it would appear that all the methods satisfy the 5% criteria for all significant burnups. What is wrong with the plotted Kopp Curves in the NRC memo?

4. The magnitude of the depletion uncertainty is dependent upon several modelling approaches (depletion conditions, enrichment, presence of burnable poisons, boron, 2D versus 3D representation, etc.). A more equitable comparison would be to compare the depletion uncertainty as a percent of the reactivity decrement for the Kopp memo approach and the direct difference approach

5. Why is there a break point in the Kopp curves at ~25.0 GWd/MTU?

DESIGNATED AS ORIGINAL
BB BRIAN BENNEY

6. Why has there been no attempt to remove the uncertainty that arises from the measurement technique in any of the NRC-reviewed methods? The chemical assay uncertainties are especially high, and by not removing this uncertainty it is by default attributed (incorrectly) to the depletion code uncertainty.

7. Has anyone independently reviewed the TRITON depletions and KENO models used in NUREG/CR-7108 to make sure there is fidelity between the actual depletion conditions and geometry of the RCA samples and the models? Any modeling approximations? Any "too hard to model" issues? Any depletion history data quality issues? Fewer high quality measurements are more valuable than lots of low quality measurements.

8. RCA data is a fuel pellet level measurement, which is far smaller a region than is of interest for criticality analysis. The EPRI data is effectively at an assembly level (or axial slice of an assembly), which averages out pellet level variations appropriately for SFP criticality calculations. Use of the RCA or MC sampling results from pellet level data effectively imposes pellet level agreement on a criticality calculation that only requires assembly slice level uncertainty allowance. What has the NRC done to convert the pellet-level uncertainties associated with the RCA data to a fuel assembly basis?

9. The NRC report notes that direct difference outliers have a disproportionate effect on the overall uncertainty. Has there been an investigation of the direct difference outliers to determine whether any of the data should be rejected on experimental quality grounds?

10. The statistical approaches by nature provide physically non-sensical results at low burnup. This should be recognized and dispositioned for what it is – a statistical artifact and not a real concern.

11. Why does the NRC contend that the EPRI depletion benchmark approach is "qualitative"?

12. On the last slide, the NRC points out that the analysis is based on the "originally submitted" validation approach. Does this mean that the memos transmitted on August 8th, 2015 and September 12th, 2015 are not considered? These memos describe the statistical approach for determining the 95/95 confidence limits and has been the basis of the discussion since August 2015.

13. The statement on page 4, 3rd paragraph, is confusing and perhaps misleading:

    "**This excellent agreement is useful to keep in mind** when **comparing the EPRI benchmarking effort** for the SCALE 6.1 TRITON depletion code **with the two NUREG/CR-7108 benchmarking efforts** also performed with the SCALE 6.1 TRITON depletion code. That is, it is expected that performing a MC USM or direct difference approach to validate SCALE 6.1 TRITON will result in very similar depletion uncertainty versus burnup characterization when using the EPRI validation approach with either CASMO-5 or SCALE 6.1 TRITON ***if*** **measurement uncertainties are small relative to depletion code uncertainty or if measurement uncertainties are similar between the two validation approaches**."

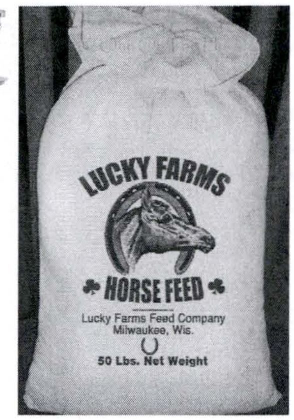
Is the NRC contending that that measurement uncertainties are similar for both approaches (Chemical Assays and Flux-Map Measurements)? Or independent of measurement uncertainty, this is all depletion code uncertainty and there is no measurement uncertainty component?

14. Comment on the following statement:

> "Like with the NUREG/CR-7108 approaches**, the major pitfall with this method is that the experimental procedure is associated with relatively high measurement uncertainties**. In this case the measurement uncertainty is based on the chosen depletion uncertainty inferencing method as EPRI notes in the 9th bullet of Section 7.1, "Interpretation of Data," of the EPRI benchmark report. Like with the two NUREG/CR-7108 approaches, these measurement uncertainties are confounded with the uncertainty that comes solely from the depletion code."

How does the NRC conclude that the flux map measurement uncertainties are as high as isotopic RCA measurements that were performed in 60s? What is the basis for this conclusion?

An Example to Focus Discussions Regarding Statistical Interpretations of Data



Federal law states that aircraft must be demonstrated to be below licensed gross weight limits (with 95/95 confidence) in order to fly passengers in the Alaska wilderness. Because of concerns about aircraft operators abusing the law, the FAA sends an inspector to Podunk Airport to make sure the law is rigorously enforced.

Upon inspection of an aircraft's weight and balance logs, the inspector determines that the aircraft has a **165-pound passenger weight limitation.** As the pilot attempts to load his passenger, the FAA inspector asks the pilot "Prove to me that your passenger meets the 165-pound weight limitation." The pilot answers, "I asked him his weight, and he told me 152 pounds." The FAA inspector responds, "everyone tends to underestimate their weight, and I cannot accept his statement as proof." The pilot responds, "Well, my aircraft manifest shows that my weight is 165 pounds, and he is smaller than I am." The FAA inspector retorts, "That is just your expert opinion - the law does not allow for that. He might be a body builder, and perhaps muscle is much heavier than fat."

So the pilot suggests, "We have a scale here, but it is not very precise. When we use it, we weigh everything 100 times to get an accurate estimate of its weight." The pilot grabs a 50-pound horse-feed sack and proceeds to weigh it 100 times while a Caltech statistics professor, waiting for the next flight, records each of the weights on his laptop. When finished, the professor reports, **"The mean weight was measured to be 51 pounds, and the weights are normally distributed with a standard deviation of 5 pounds.** While it is not a very accurate scale, it does appear to be an unbiased scale (e.g., equally likely to weigh too high or too low)." The passenger is then weighed 100 times, and the Caltech professor reports, **"His mean weight was measured to be 150 pounds, and the weights are normally distributed with a standard deviation of 15 pounds."**

So the pilot says, "We are good to go, 15 pounds under my plane's 165-pound weight limitation."

"Not so fast," says the FAA inspector, **"I see that of the hundred weight measurements, sixteen exceeded 165 pounds. To meet the 95/95 criteria that the passenger's weight is less than 165 pounds, only 2.5 of the measurements can exceed 165 pounds** (e.g., one side of the two-sided 5% criteria for 100 measurements)**. It is also clear that one measurement even showed your passenger weighing 195 pounds. Consequently, it seems apparent to me that your passenger is too heavy to meet the FAA legal gross weight requirement."**

While the pilot grinds his teeth, the Caltech statistics professor offers to help in resolving the dispute by proposing several independent assessments of the passenger's weight measurement data:

1) The data suggest that if the passenger were weighed again, **there is a sixteen percent chance that his weight will be measured to exceed 165 pounds**.

2) The data also suggest that **the passenger's weight is best estimated as 150 +/- 1.5 pounds** (e.g., +/- one sigma of the measured population divided by the square root of the number of measurements).

3) The data also suggest that **with 95% confidence, I can say that it is 95% likely that the passenger's weight is less than 154 pounds** (e.g., mean weight, plus two sigma of the measured population divided by the square root of the number of measurements, plus a small normal-to-Student's t-distribution factor).

Being diligent, the Caltech statistics professor also choses to further muddy the water by noting that:

4) One must account for the fact that perhaps the horse-feed sack did not actually weigh 50 pounds.

5) One must properly account for the fact that the demonstration of an unbiased scale was only made for 50-pound weights and not for 150-pound weights.

6) One must take into account that 100 measurements are not sufficient to strictly satisfy the central limit theorem and the implicit assumption of a normal distribution.

7) It is the professor's opinion that more formal study of alternate statistical interpretations is required in order to truly be 95% confident of meeting the FAA legal requirement, and that can only be performed after the professor returns to Palo Alto.

The FAA inspector is left in the unenviable predicament in which he must decide if the passenger's weight is best estimated by:

a) **Examining the extremes of the measured data points** (e.g., the prediction interval of the measured data).

or

b) **Using the mean measured weight plus the uncertainty of the mean** (e.g., the mean weight. plus the 95% confidence interval of the mean obtained from traditional regression analysis, plus an appropriate t-distribution factor).

This example is intended to show, for a very simple case, that purely statistical interpretation of measured data (without regard to the physics behind the measurements), can lead to very conservative uncertainty estimates, and much more useful estimates can be derived by using large sets of measured (yet uncertain) data and their corresponding regression fits – provided the measurement technique can be reasonably demonstrated to be unbiased.

We believe that the data in the EPRI/Studsvik reactivity decrement analysis study falls into the later category, and confidence intervals of WLS regression fits can reasonably be used to assess uncertainty of inferred reactivity decrements and satisfy the NRC's 95/95 criteria.