LA-UR-15-22042
Approved for public release; distribution is unlimited.

Title: Data Mining Techniques to Estimate Plutonium, Initial Enrichment, Burnup, and Cooling Time in Spent Fuel Assemblies

Author(s): Trellue, Holly Renee
Fugate, Michael Lynn
Tobin, Stephen Joseph

Intended for: Report

Issued: 2015-03-20

Data Mining Techniques to Estimate Plutonium, Initial Enrichment, Burnup, and Cooling Time in Spent Fuel Assemblies

by Holly R. Trellue, Michael L. Fugate, and Stephen J. Tobin

Los Alamos National Laboratory

1.0 Introduction

The Next Generation Safeguards Initiative (NGSI), Office of Nonproliferation and Arms Control (NPAC), National Nuclear Security Administration (NNSA) of the U.S. Department of Energy (DOE) has sponsored a multi-laboratory, university, international partner collaboration to (1) detect replaced or missing pins from spent fuel assemblies (SFA) to confirm item integrity and deter diversion, (2) determine plutonium mass and related plutonium and uranium fissile mass parameters in SFAs, and (3) verify initial enrichment (IE), burnup (BU), and cooling time (CT) of facility declaration for SFAs. A wide variety of nondestructive assay (NDA) techniques were researched to achieve these goals [Veal, 2010 and Humphrey, 2012]. In addition, the project includes two related activities with facility-specific benefits: (1) determination of heat content and (2) determination of reactivity (multiplication). In this research, a subset of 11 integrated NDA techniques was researched using data mining solutions at Los Alamos National Laboratory (LANL) for their ability to achieve the above goals.

In FY13 a database of signals was created of the following Non Destructive Assay (NDA) passive and neutron generator-based active techniques through simulations, and data mining/integration studies were performed [Trellue, 2013].

- Passive techniques:
  - Total Neutron (TN)
  - Gross and spectral resolved Passive Gamma (PG)
- Active techniques:
  - Differential Die Away (DDA)
  - Delayed Neutron (DN)
  - Delayed Gamma (DG)

The logic for including TN and PG signals with the other techniques is that (1) both signals provide unique information about the content of the assembly and (2) both are relatively easily to deploy. The work in FY14 builds off the FY13 research, expanding the number of cases examined to include detector responses from other NDA instruments examined in the NGSI-Spent Fuel (SF) project: Californium Interrogation with Prompt Neutron (CIPN), Self-Interrogation Neutron Resonance Densitometry (SINRD), Passive Neutron Albedo Reactivity (PNAR), and Differential Die-away Self-Interrogation (DDSI) – see Figure 1. Additional work performed in FY14 and early FY15 is described in this document.

Table 1. Eleven integrated systems; five examined during FY13 and six during FY14. The number of predictors indicates the number of signals or quantities determined from signals for integration analyses.

Integrated NDA Systems	Number of Predictors
1. TN + PG (gross)	2
2. TN + PG (spectral)	6
3. TN + PG (spectral) + DDA	8
4. TN + PG (spectral) + DDA + DN	(FY13)
5. TN + PG (spectral) + DDA + DN + DG	(FY13)
6. TN + PG (spectral) + DDSI	8
7. TN + PG (gross) + active CIPN	3
8. TN + PG (spectral) + active CIPN	7
9. TN + PG (spectral) + DDA + active CIPN	9
10. TN + PG (spectral) + SINRD	12
11. TN + PG (spectral) + DDSI + DDA	10

2.0 Creation of Integration Spreadsheet Used for Analysis Formulation

To provide the necessary NDA signal responses from multiple detectors for the purpose of integrating NDA techniques together, a database was created using the second spent fuel library (SFL2a) created for the NGSI-SF Project [Galloway, 2012 and Trellue, 2011]. The database lists the NDA signal response for the neutron generator NDA instruments plus passive gamma and neutron signals determined by simulating either the neutron or gamma emission from SFL2a. It exists in the form of an EXCEL spreadsheet with separate tabs for 11 different combinations of IE and BU. All NDA systems were simulated for the following unique set of IE and BU:

- 2 wt% - 15 and 30 GWd/MTU,
- 3 wt% - 15 and 30 GWd/MTU,
- 4 wt% - 15, 30, and 45 GWd/MTU, and
- 5 wt% - 15, 30, 45, and 60 GWd/MTU.

Multiple cooling times are available for each of the 11 IE and BU values. Yet, the cooling times simulated among the NDA systems varied.

Information from the database was compacted into a single spreadsheet for the integration analyses in FY13 (i.e., the first five subsets), and the spreadsheet was expanded to include NDA signals from additional instruments in FY14. Table 2 lists the CT values for which simulation data is available for each subset, quantifying the number of assemblies used in the integration research (i.e. Systems 1 and 2 contain data for 55 assemblies whereas Systems 9 and 11 only have data for 33 assemblies because that represented the cooling time common to all techniques in the subset).

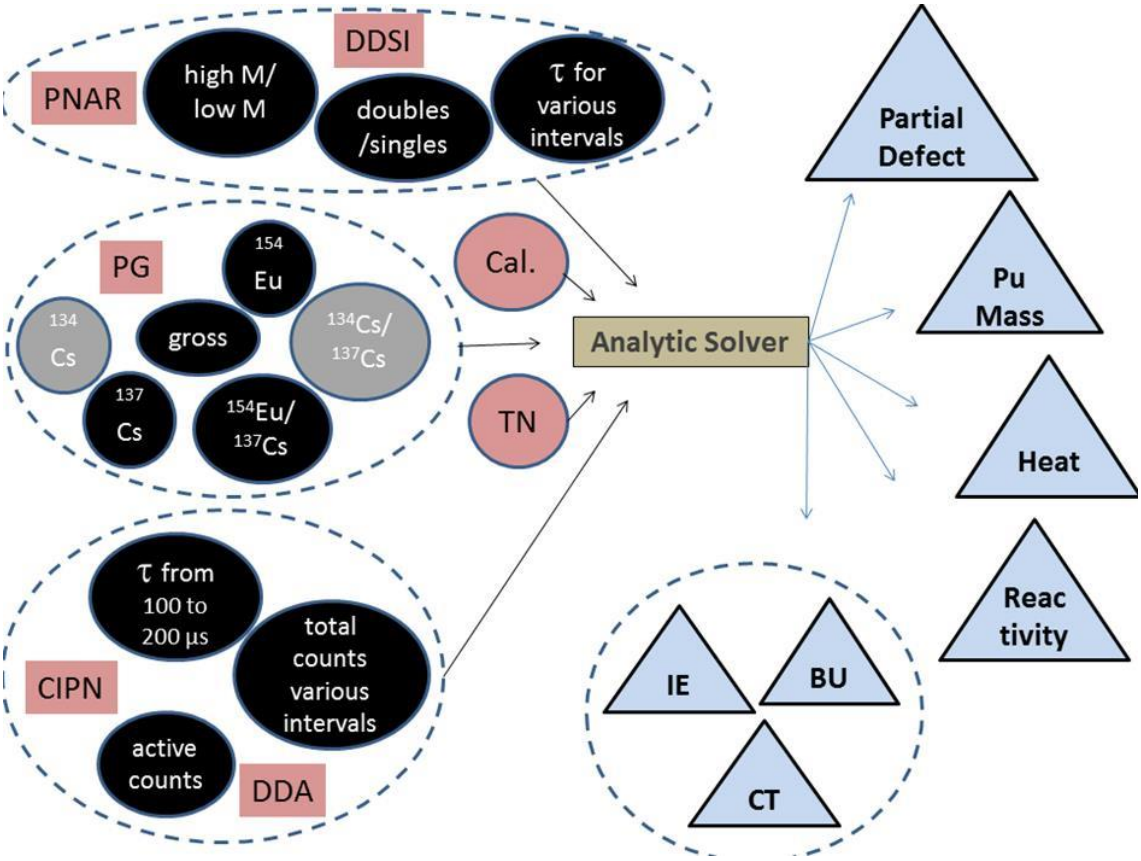

Figure 1. Flow Chart for Data Mining

Table 2. Eleven integrated systems, five analyzed in FY13 and six in FY14. The number of IE, BU, and CTs that are in common for all the instruments involved in the integrated NDA system are listed in the final two columns.

Integrated NDA Systems	Initial Enrichment and Burnup Combinations	Cooling Times (years)
1. TN + gross-PG	11	1, 5, 20, 40, 80
2. TN + spectral resolved-PG	11	1, 5, 20, 40, 80
3. TN + spectral resolved-PG + DDA	11	1, 5, 20, 80
4. TN + spectral resolved-PG + DDA + DN	11	1, 5, 20, 80
5. TN + spectral resolved-PG + DDA + DN + DG	11	5, 20
6. TN + PG (spectral) + DDSI	11	5, 20, 40, 80
7. TN + PG (gross) + active CIPN	11	5, 20, 40, 80
8. TN + PG (spectral) + active CIPN	11	5, 20, 40, 80
9. TN + PG (spectral) + DDA + active CIPN	11	5, 20, 80
10. TN + PG (spectral) + SINRD	11	5, 20, 40, 80
11. TN + PG (spectral) + DDSI + DDA	11	5, 20, 80

3.0 Predictors (inputs) Used in Creation of Algorithms

Greater detail on the content of the database created for the neutron generator techniques has been described previously [Trellue, 2013]. Table 3 lists the signals or derived quantities obtained for each NDA technique used in the integration research. The following list is consistent with the number of predictors listed in Section 1.0 for the integrated systems. Note that in several cases, the signatures measured were greater than the final derived quantities used in the integration.

The data base also contains isotopic mass information about each assembly (up to 14 Pu and U isotopes, a wide range of fission fragments, etc.), and even calculated figure-of-merit (FOM) values (in particular, the ratio of macroscopic fission cross section to macroscopic capture cross section).

Table 3. Descriptions of Predictors

NDA Instrument	Number of Predictors and Predictor Description
PG	**1 to 5 predictors**, 5 predictors for spectral resolved PG (3 peak areas {one peak per isotope}: ^{134}Cs, ^{137}Cs, and ^{154}Eu, 2 ratios of those peaks), or 1 predictor for gross (total) PG
TN	**1 predictor**, total passive count rate
DDA	**2 to 8 predictors**, total counts from 100-200 and 200-500 μs time windows in both front and back detectors, 100-200 and 500-1000 μs time windows in all detectors, and die away time (λ) of neutrons in back detectors from 100-200 μs and all detectors from 500-1000 μs [Henzl, 2014]. Note that in the FY14 analysis only the last 4 predictors were used although all eight were examined. In the FY14 analysis, only two predictors were used: counts from all detectors in the 500-1000 μs time window, and the die away time in the back detectors during the 100-200 μs time window. Only two were used because the others were found to have correlations to each other and thus offered no additional information to the data mining process (see Table 4).
DN	**1 predictor**, total count rate during delayed neutron time window
DG	**4 predictors**, 4 derived quantities estimating contribution of ^{239}Pu, ^{241}Pu, ^{235}U, and ^{238}U
DDSI	**2 to 10 predictors,** magnitudes of fast and slow components of the Rossi-alpha distribution, die-away constants of exponential fits to those components, fast and slow area integrals under the Rossi-alpha distribution in the early (5-50) and late (100-200) time domains respectively, the correlated multiplication in the assembly from the early die-away time, the total number of neutrons detected per second (singles), and the slow area divided by the singles rate [Kaplan, 2014]. However, only early, fast, and slow die away times plus slow area influence the results significantly as shown in Table 5.
CIPN	**1 predictor,** Active CIPN counts
SINRD	**6 predictors,** FFM/Base, FFM/(Gd-Cd), FFM/Gd, FFM/Cd, Bare/Gd, Bare (Gd-Cd)

Table 4. Correlations for TN, CIPN, and DDA Signals. Note that the absolute value of the correlation indicates strength of the linear relationship between parameters . For correlations near 1.0, the two parameters can predict each other and are thus redundant and only one is needed. In fact, including both can lead to unstable solutions in empirical models.

Correlations	TN	CIPN	100-200 µs F	200-500 µs F	100-200 µs B	200-500 µs B	100-200 µs T	500-1000 µs T	l 100-200 µs -B	l 500-1000 µs T
TN	1	-0.38	-0.37	-0.39	-0.35	-0.37	-0.36	-0.35	-0.45	-0.37
CIPN	-0.38	1	0.99	0.97	0.98	0.97	0.99	0.94	0.74	0.97
DDA 100-200 µs - front	-0.37	0.99	1	0.99	1	0.99	1	0.97	0.76	0.99
DDA 200-500 µs - front	-0.39	0.97	0.99	1	0.99	1	0.99	0.99	0.8	0.98
DDA 100-200 µs - back	-0.35	0.98	1	0.99	1	0.99	1	0.98	0.72	0.99
DDA 200-500 µs - back	-0.37	0.97	0.99	1	0.99	1	0.99	0.99	0.78	0.98
DDA 100-200 µs - total	-0.36	0.99	1	0.99	1	0.99	1	0.97	0.74	0.99
DDA 500-1000 µs - total	-0.35	0.94	0.97	0.99	0.98	0.99	0.97	1	0.74	0.96
DDA die away time 100-200 µs - back	-0.45	0.74	0.76	0.8	0.72	0.78	0.74	0.74	1	0.73
DDA die away time 500-1000 µs - total	-0.37	0.97	0.99	0.98	0.99	0.98	0.99	0.96	0.73	1

Table 5. Correlations for TN, CIPN, and DDA Signals. (see Note associated with Table 4 for description of correlation values).

	TN	Mult	Fast_A	Fast_Tau	Slow_A	Slow_Tau	Early_Tau	Fast_Area	Slow_Area	Singles	D/S
TN	1	-0.34	0.9	-0.24	0.86	-0.45	-0.36	0.85	0.84	0.9	-0.34
Mult	-0.34	1	-0.32	0.93	-0.24	0.88	1	-0.27	-0.21	-0.32	0.99
Fast_A	0.9	-0.32	1	-0.23	0.98	-0.42	-0.34	0.99	0.97	1	-0.32
Fast_Tau	-0.24	0.93	-0.23	1	-0.15	0.66	0.91	-0.18	-0.12	-0.23	0.91
Slow_A	0.86	-0.24	0.98	-0.15	1	-0.36	-0.25	1	0.99	0.98	-0.26
Slow_Tau	-0.45	0.88	-0.42	0.66	-0.36	1	0.89	-0.37	-0.32	-0.42	0.88
Early_Tau	-0.36	1	-0.34	0.91	-0.25	0.89	1	-0.28	-0.22	-0.33	0.98
Fast_Area	0.85	-0.27	0.99	-0.18	1	-0.37	-0.28	1	0.99	0.99	-0.28
Slow_Area	0.84	-0.21	0.97	-0.12	0.99	-0.32	-0.22	0.99	1	0.97	-0.23
Singles	0.9	-0.32	1	-0.23	0.98	-0.42	-0.33	0.99	0.97	1	-0.32
D/S	-0.34	0.99	-0.32	0.91	-0.26	0.88	0.98	-0.28	-0.23	-0.32	1

4.0 Methodology

One of the most common goals in "data mining," or "exploratory data analysis" is to estimate the function *f* in Eq. (1). For historical reasons, if *f*() and *y* are continuous-valued functions, the goal is often called "regression." If the response *y* is categorical, the goal is called "classification" or "pattern recognition." Our focus here is regression where the response *y* can be ^{239}Pu or any other Pu isotope, or elemental Pu, and the *x's* are NDA measurements or perhaps combinations of NDA instruments. For example, principal components regression fits *y* to predictors that are linear combinations of other predictors, or x values.

$$y = f(x_1, x_2, \ldots, x_p), \qquad (1)$$

where *p* is the number of predictors, ranging from 3 to 12 in Table 1, and $x_1, x_2, \ldots, x_p$ are the measured values of the NDA predictors.

The training data is a random subset of the available examples of $\{y,x_1,x_2,\ldots, x_p\}$ values, and for each random selection of training data, the remaining example $\{y,x_1,x_2,\ldots, x_p\}$ values are used for testing. The main concept is that the same data is not used for both training and testing. The ability of data mining techniques to predict the y value is given by the root mean squared error $\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n_{test}} (y_i - f_i)^2 / n_{\mathrm{test}}}$, which can go to zero as you add more predictors for the training data. However, if a portion of the testing

data is used for testing instead of training (i.e. hold-out test data), RMSE will not go to zero. For example, a polynomial function can provide arbitrarily small RMSE as the polynomial order is increased in the training data. Experience suggests that, although there are exceptions, wiggly functions such as those fit using high-order polynomials tend to "chase noise" in the training data, and do poorly in testing data. Cross validation is a technique that splits the data into separate subsets for training and testing (data can also be randomly "held out" of a set for testing). Cross validation is a well-established method to mitigate the tendency of data mining to be overly optimistic about performance on future data. Some readers might be familiar with other resample-the-data strategies such as the bootstrap. The bootstrap resamples the observations $\{y,x_1,x_2,\ldots, x_p\}$with replacement, again creating a separate training and test set.

To actually perform the empirical modeling, we are using the R statistical programming language, which provides many data mining functions and a programming environment so cross validation, adding errors ("jittering the predictors"), and evaluating several options can be evaluated, such as whether to log transform the response or predictors or whether to rescale the predictors.

We evaluated the predictors using three regression options, all of which are available as R functions by installing add-on packages for R: Linear Models (LM), Bayesian additive regression trees (BART), and Generalized Additive Models (GAM). All options can use weights (most suitable if the variance of the measured *y* value is a function of the true *y* value) or not (most suitable if the variance of the measured *y* value is not a function of the true *y* value).

Linear models express y as a linear combination of predictors or some transformation of the predictors such as $y_{true} = f(x) = \beta_0 + \sum_{j=1}^{r}(\beta_j x_j)$.

Bayesian additive regression trees (BART) are a nonparametric Bayesian regression approach that uses dimensionally adaptive random basis elements. BART uses an ensemble of trees, so it is motivated by boosting algorithms with each tree constrained to be a weak predictor (Burr, 2012-2014). A prior and a likelihood are required, so the approach enables full posterior inference including point and interval estimates of the unknown regression function as well as the marginal effects of potential predictors.

Generalized Additive Models (GAM) combine generalized linear models with additive models, using $g(E(y)) = \beta_0 + \sum_{j=1}^{p} f_j(x_j)$, where g is a link (transform) function and E(y) is the expected value of y (Burr, 2012-2014). The functions $f_i(x_i)$ can be fit either parametrically or nonparametrically, providing the potential for better fits to data than other methods. The method is quite versatile because GAM can use a scatterplot smoothing function such as a locally weighted mean for $f_1(x_1)$, and then use a factor model for $f_2(x_2)$. By allowing nonparametric fits, well designed GAMs allow good fits to the training data with relaxed assumptions on the actual relationship. As with most flexible models, over-fitting can be an issue, and the method is tested appropriately. The number of smoothing parameters can be specified, and this number should be reasonably small, much smaller than the degrees of freedom. The GAM technique was used during the FY13 analysis but failed when a large number of predictors were given with a relatively small number of observations for each subset and was thus not included in the results presented in this paper (Burr, 2012-2014).

In the FY13 research the Relative Root Mean Squared Error was used to numerically represent the ability of techniques to fit the data, but in this paper, only the RMSE is reported. Appendix A describes some of the reasoning for this change.

5.0 Results

The results in Table 6 and Figures 2 and 3 show the results for both testing and training on the same set of data (SFL2a) using the bootstrap technique to estimate hold-out RMSEs. Note that in some cases, all predictors are used whereas in others some predictors were found to be too closely correlated to each other (see values close to 1.0 in Tables 4 and 5), and only a few were used. For DDSI, the main useful predictors were found to be early, slow, and fast die away times plus slow area (anywhere from two to four predictors were used depending on the dependent variable desired), and for DDA, the most useful predictors were: counts from 500-1000 μs in all detectors and die-away time from 100-200 μs in the back detectors.

The combination of techniques PG, TN, DDSI, plus DDA appears to produce the lowest RMSEs, but the integrated systems PG, TN, and DDA and PG, TN, and DDSI perform adequately too. It is important to recall that one of the key purposes of safeguards measurements is to discourage diversion due to the fear of detection. In light of this, there are benefits to integrated signals such as PG with a neutron technique in that the combined signals would require a "would be proliferator" to alter both the gamma and neutron signals in a specific manner; hence, it is noted that the RMSE value alone is not the only indicator of the value of an integrated system. The applicability of data mining to predict IE, BU, CT, Pu and/or missing fuel rods in an assembly is an important aspect of research for the NGSI-SF project. Thus, another test performed as part of this research involved determining how well using the training data in SFL2a to develop an empirical solution to determine IE, BU, CT, and Pu mass in a mystery spent fuel library, SFL4. SFL4 was run in a similar fashion to SFL2a but contains non-integer values of IE, BU, and CT as well as cases with missing pins (partial defect) and 4-sided asymmetry instead of 3-sided (i.e. in SFL2a, the top and bottoms halves are symmetrically identical but not the left and right). Table 7 gives the parameters used to generate SFL4.

Table 6. The RMSE is listed for each case below for the prediction of elemental Pu, followed by IE, BU, and CT using 33 observation cases using the BART empirical solver. The value in parentheses in the first row indicates the RMSE if the sample mean is taken as a predicted result. Using more knowledge (i.e. signals from instruments) should reduce the RMSE from this sample mean, which it does.

	T-Pu (833)	BU (14.5)	CT (33)	IE (1.1)
1. TN + PG Gross	392	9.1	11.8	1.2
2. TN + PG Spectral	422	9.9	2.4	1.2
3. TN + PG Spectral + DDA	250	5.6	7.2	0.9
6. TN + PG Spectral + DDSI	279	5.6	7.1	0.7
8. TN + PG Spectral + CIPN	437	9.9	2.9	1.1
9. TN + PG Spectral + CIPN + DDA	260	5.7	6.9	0.9
10. TN + PG Spectral + SINRD	302	5.3	6.9	0.3
11. TN + PG Spectral + DDSI + DDA	277	5.7	9.5	0.7

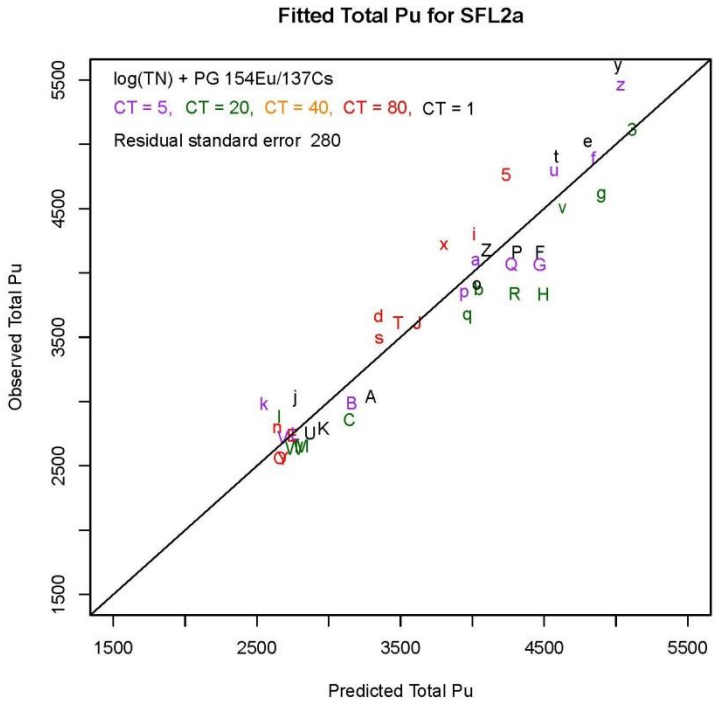

Figure 2. Passive Gamma (PG) and Total Neutron (TN) Predictors to Predict Pu. Note that uppercase letters are primarily 2 and 3 wt%, lowercase letters are 4 and 5 wt%, and numbers are 5 wt% cases.

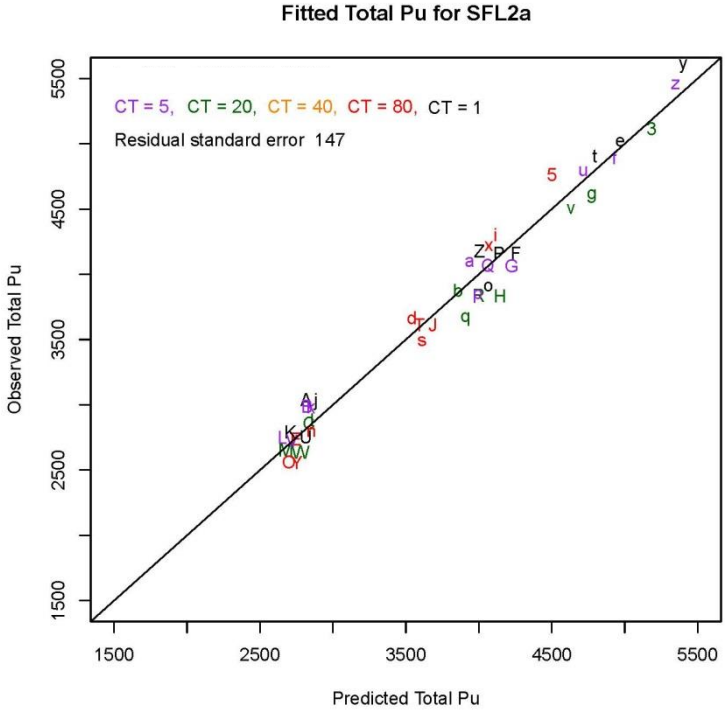

Figure 3. PG, TN, and DDSI Predictors to Predict Pu

Table 7. Operating/initial parameters for cases in SFL4.

SFL4 case	SFL4 identifier	IE	CT	BU	Downtime between cycles (days)	Notes
1	A	4	20	21.5	1125.72	
2	B	4	1	40	1125.72	
3	C	2.6	5	25.3	30	
4	D	2.6	80	10.9	30	
5	E	2.8	50	16.8	30	
6	F	2.8	1	27.4	30	
7	G	3.4	20	22.3	30	
8	H	3.4	50	34.2	30	
9	I	3.6	5	19.6	30	
10	J	3.6	80	36.1	30	
11	K	3.8	20	14.1	30	
12	L	3.8	5	37.9	30	
13	M	4	80	21.3	30	Control rods present in cycle 1
14	N	4	50	37.8	30	Control rods present in cycle 1
15	O	4	1	21.3	30	
16	P	4	80	39.9	30	
17	Q	4.2	5	22.3	30	
18	R	4.2	1	42.8	30	
19	S	3.2	20	19.7	30	
20	T	3.2	50	31.7	30	
21	U	4	5	21.4	562.86	
22	V	4	80	39.9	562.86	
23	W	4	50	21.5	2190	1434.5 d irr, 2190 CT, 1 yr irr.
24	X	4	1	40.3	2190	1434.5 d irr, 2190 CT, 1 yr irr.
25	Y	4	5	18.4		1/8 core simulation of all rods
26	Z	4	20	35.6		1/8 core simulation of all rods
27	a	4	1	21.3	30	40 void rods in center
28	b	4	1	21.3	30	40 DU rods in center
29	c	4	1	21.3	30	40 NU rods in center
30	d	4	1	21.3	30	20 NU rods equally distributed
31	e	4	1	21.3	30	20 DU rods equally distributed
32	f	4	1	21.3	30	40 NU rods equally distributed
33	g	4	1	21.3	30	40 DU rods equally distributed
34	h	4	1	21.3	30	40 void rods equally distributed
						Partial Defect Cases in Red

The results obtained from using SFL2a as training data and SFL4 as testing are shown in Table 8 using the BART empirical solver. Two different values are reported: predictions obtained for testing of assemblies without missing pins (cases 1-26 then 35-36 from Table 7), and then those for testing of assemblies including missing pins (RMSEs in parentheses include testing of cases 27-34). We expect predicted values of plutonium mass for the partial defect cases to be smaller than the declared masses, assuming that the missing pins would not be declared. Thus, the RMSEs should be larger when partial defect cases are included, which they are for most subsets. DDA combined with at least PG and TN produces the smallest RMSEs when missing pin cases are not included and a significant change when missing pin cases are added, showing its sensitivity for estimating plutonium. SINRD produces the largest difference in RMSEs on average. Figures 4-23 show representations of which cases are associated with the largest differences between predicted and declared values for Pu, BU, IE, and CT respectively. Note that a tailored linear model can produce better results than a complex BART one but is more time-consuming to implement; thus, it has currently only been analyzed for a few cases (see Figures 4 and 5). Also note that training data for a cooling time of 50 years was not included, so test cases with a cooling time of 50 years tend to be associated with larger RMSEs than those where cooling times in the test and training data were identical. Similarly, burnup appears to be over-predicted in almost all cases in SFL4. This is probably because the burnup data in the training set was only created for predefined values (15, 30, 45, and 60 GWd/MTU) whereas the test data contained a wide range of burnups (such as 10.9 GWd/MTU). Empirical solutions were not able to predict all burnups equally well.

Note that for the DDA and DDSI cases in particular, the blue values (representing cases with missing pins) are located significantly to the left of the black "predicted = observed" line. The shift occurs because the "observed" values are based on the fact that the operator would not declare any missing pins but a full assembly instead. However, the predicted values are lower than the declared because less plutonium exists. Thus, the more significantly the blue points are shifted to the left, the better the ability of the empirical solutions to detect missing pins is.

Table 8. RMSEs for SFL4 cases Using SFL2a Training Data

	T-Pu (673)	BU (9)	CT (30)	IE (0.47)
TN + PG Gross	390 (316)	6	12	0.45
TN + PG Spectral	267 (224)	6	4	0.5
TN + PG Spectral + DDA	142 (264)	5	5	0.37
TN + PG Spectral + DDSI	195 (169)	6	4	0.35
TN + PG Spectral + CIPN	266 (260)	7	6	0.45
TN + PG Spectral + CIPN + DDA	193 (234)	4	5	0.38
TN + PG Spectral + SINRD	281 (363)	11	7	0.9
TN + PG Spectral + DDSI + DDA	198 (330)	4	4	0.49

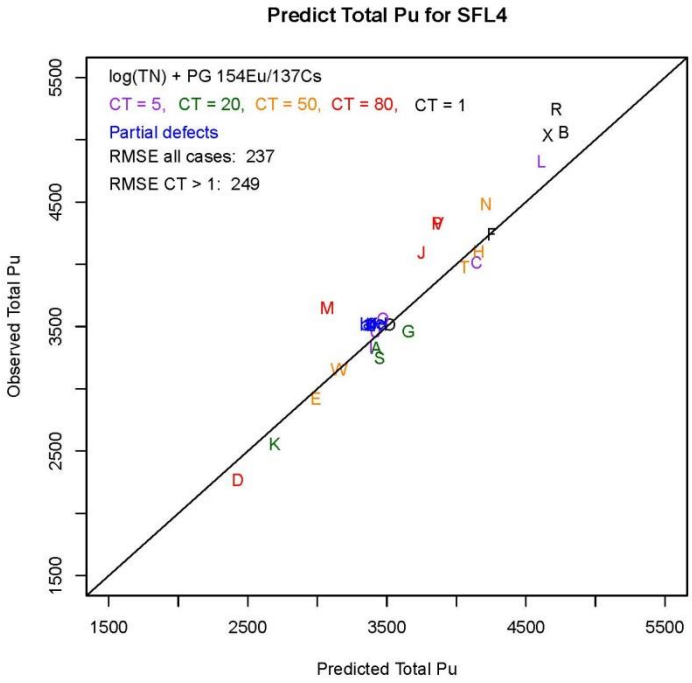

Figure 4. Linear Model Fits for PG and TN to Predict Pu. Note that Cases M and N had control rods, which changes the energy spectrum and thus build-in of plutonium and thus prediction of Pu.

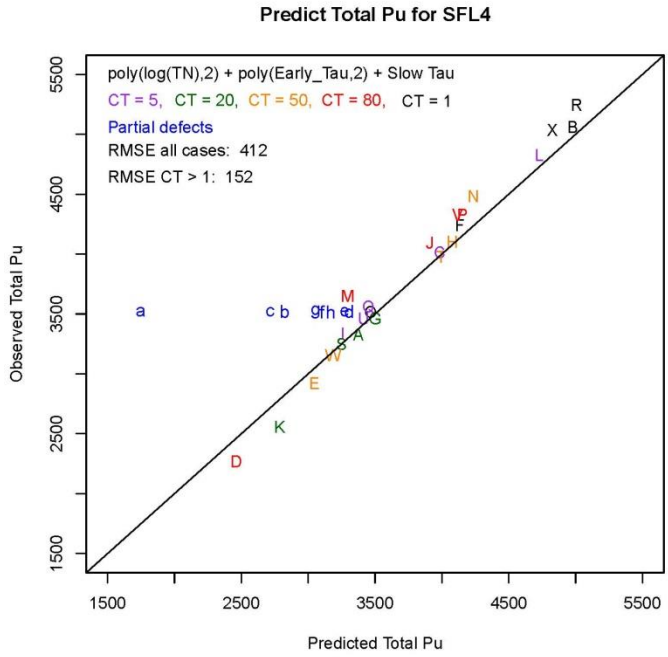

Figure 5. Linear Model Fits for TN+DDSI to Predict Pu. Note that DDSI can help predict Pu. Pu in partial defect cases should be under-predicted but the data for a cooling time of 1 year is not all yet complete and inaccuracies exist.

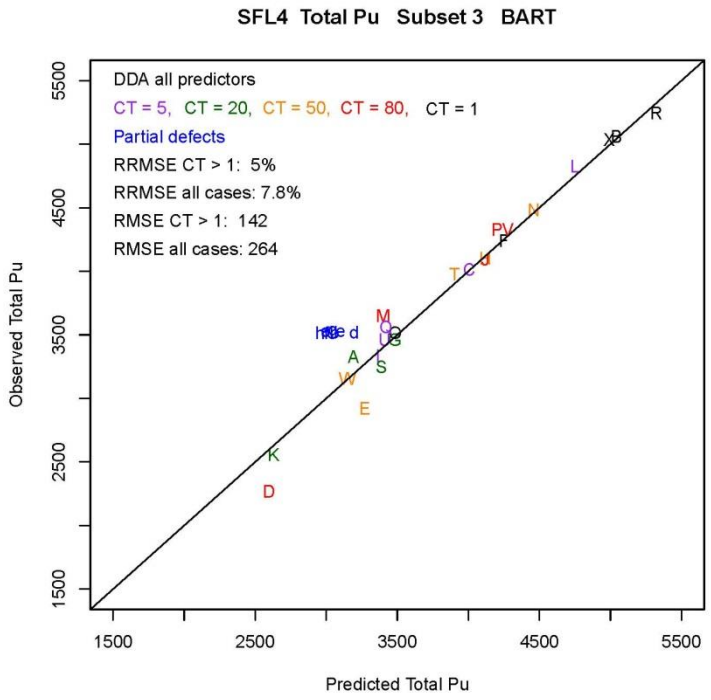

Figure 6. BART Models to Predict Pu with PG+TN+DDA. This figure shows that more complex relationships can also be used than the linear ones in Fig. 5 and 6, and DDA performs well with BART. Note that Case D with a short BU (~11 GWd/MTU) outside the training parameters is the most obvious outlier beside those with partial defects, and Case O is "base" for the other partial defect cases in SFL4.

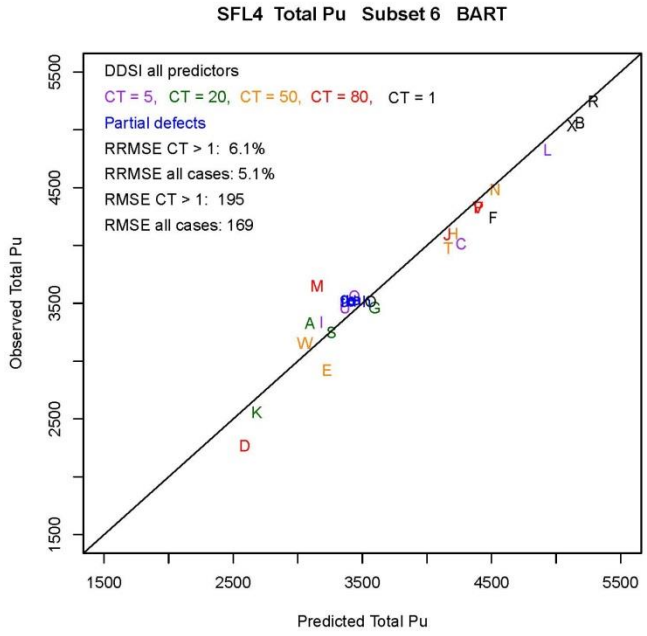

Figure 7. BART Models to Predict Pu with PG + TN + DDSI

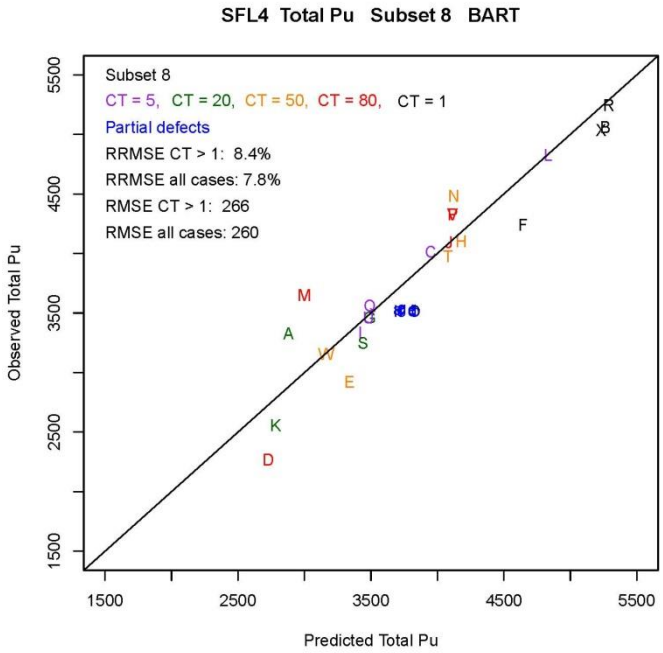

Figure 8. BART Models to Predict Pu with PG+ TN + CIPN

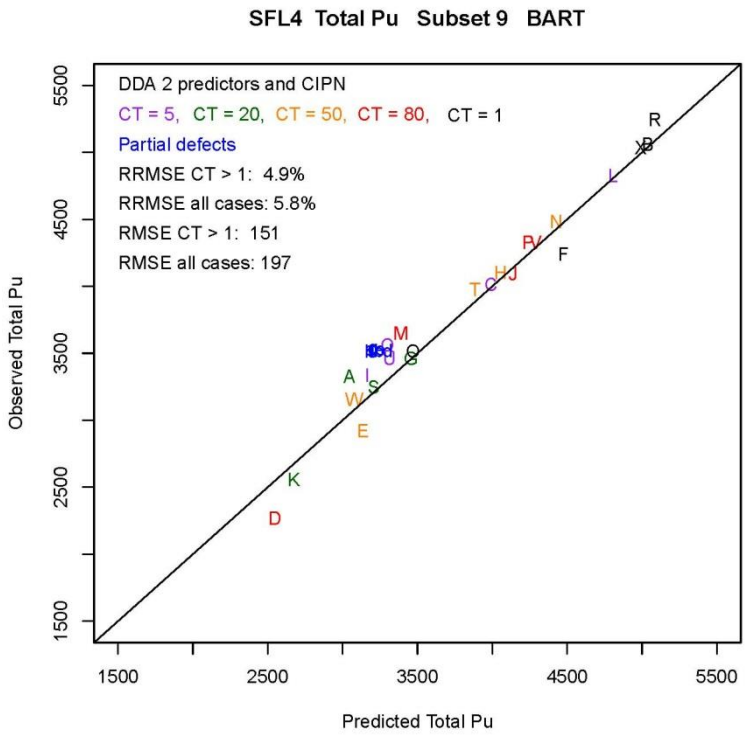

Figure 9. BART Models to Predict Pu with PG + TN + CIPN + DDA

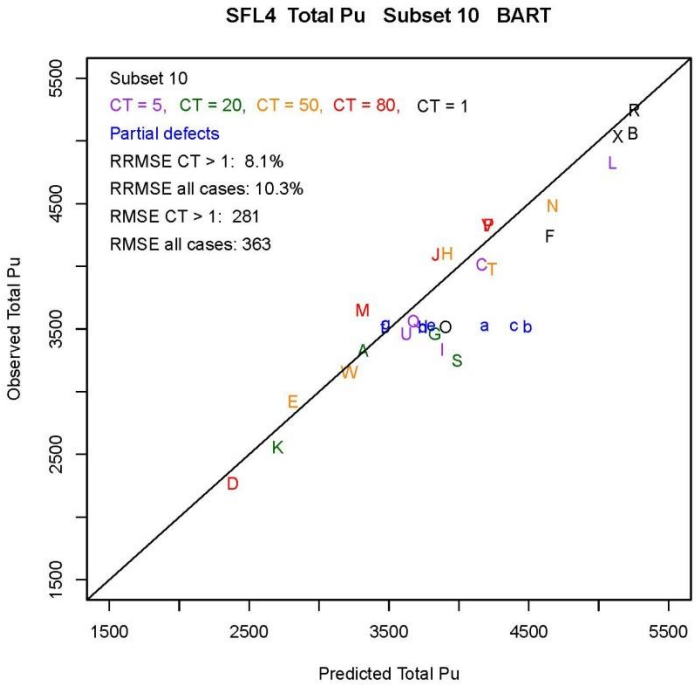

Figure 10. BART Models to Predict Pu with PG + TN + SINRD

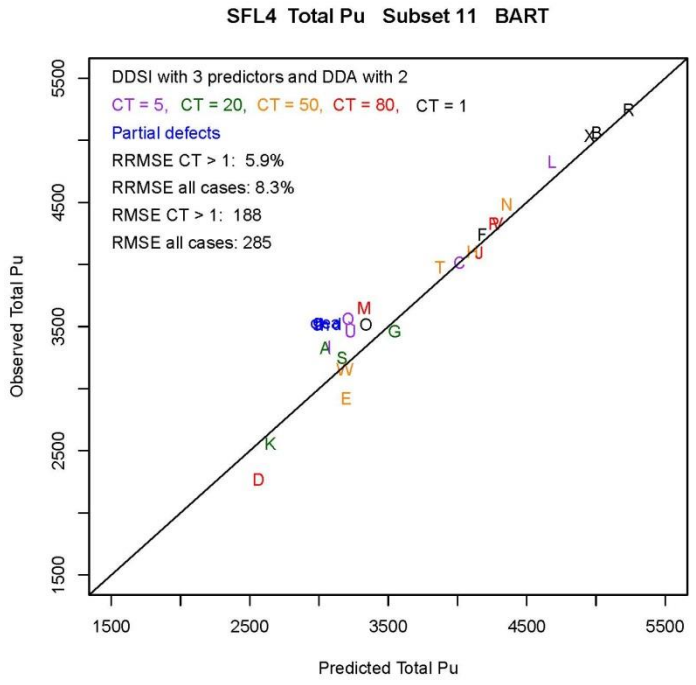

Figure 11. BART Models to Predict Pu with PG + TN + DDSI + DDA. Note that only 3 DDSI (die away time) and 2 DDA predictors were used because they obtained better results than with using all predictors.

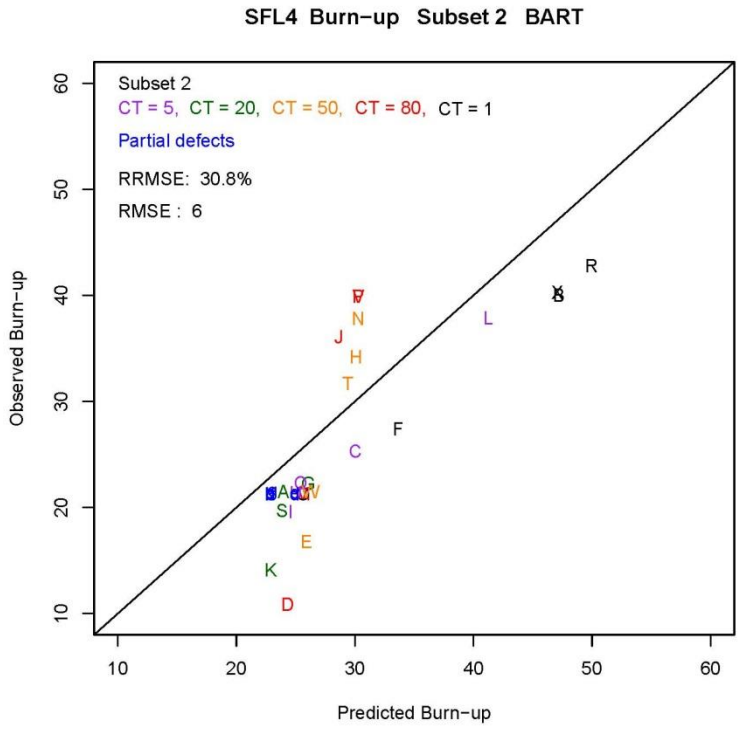

Figure 12. BART Models to Predict BU with PG + TN

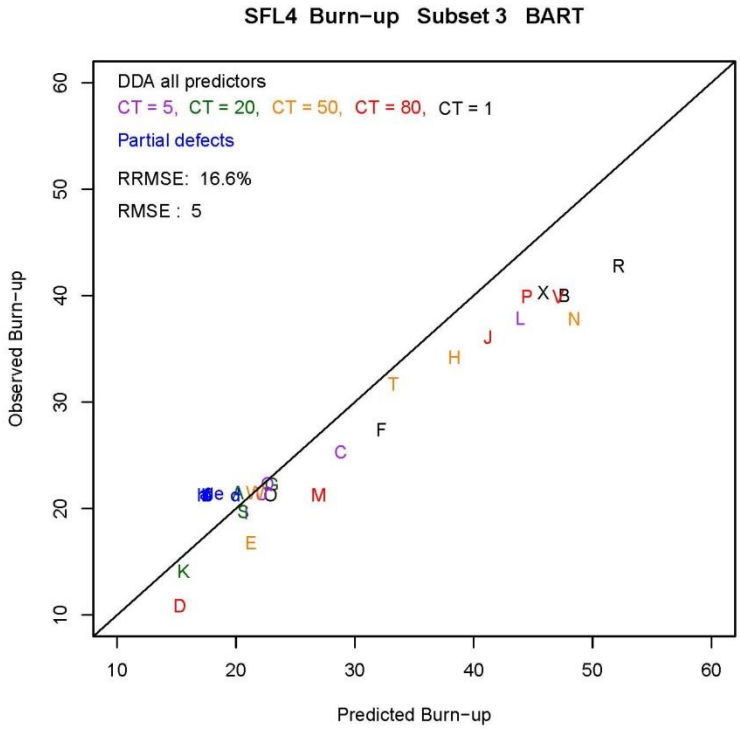

Figure 13. BART Models to Predict BU with PG + TN + DDA

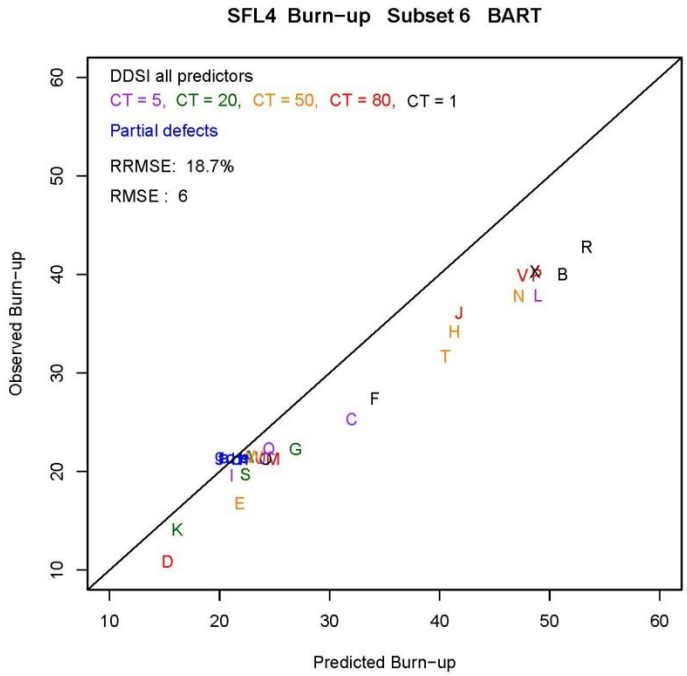

Figure 14. BART Models to Predict BU with PG + TN + DDSI

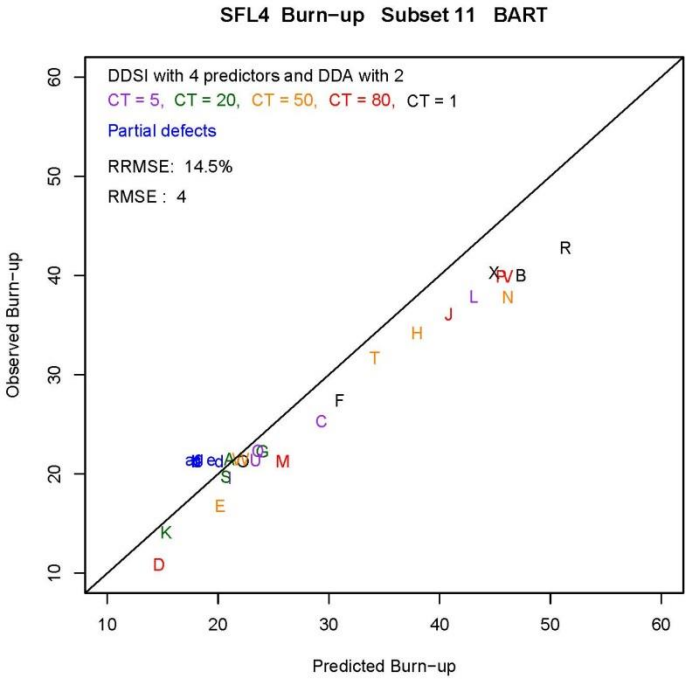

Figure 15. BART Models to Predict BU with PG + TN + DDSI + DDA. The four DDSI predictors include early, fast, and slow die away times plus slow area, and the two DDA predictors are counts from 500-1000 μs in all detectors and die-away time from 100-200 μs in the back detectors.

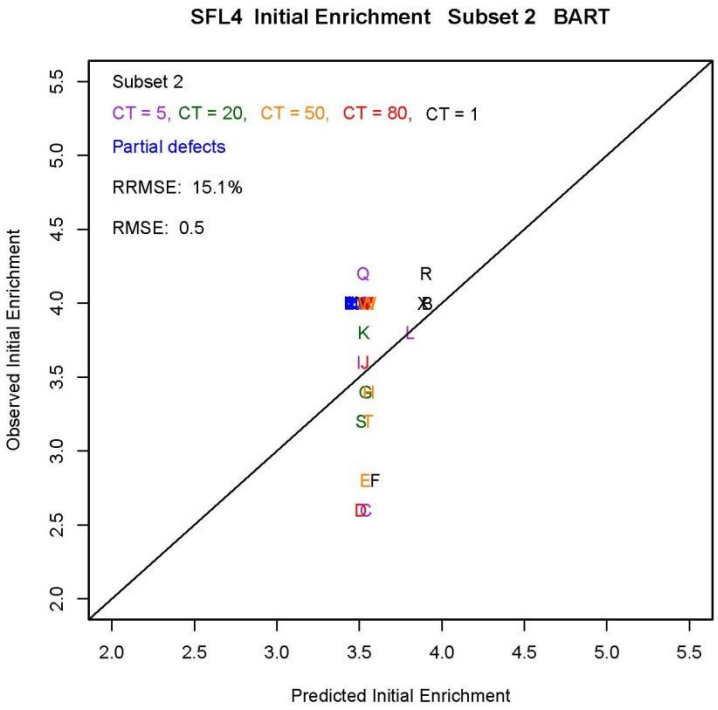

Figure 16. BART Models to Predict IE with PG + TN

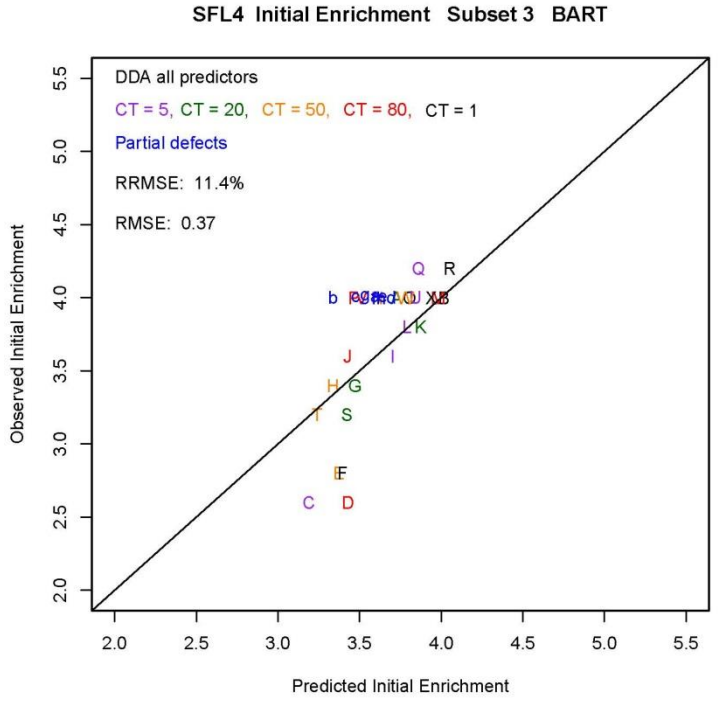

Figure 17. BART Models to Predict IE with PG + TN + DDA

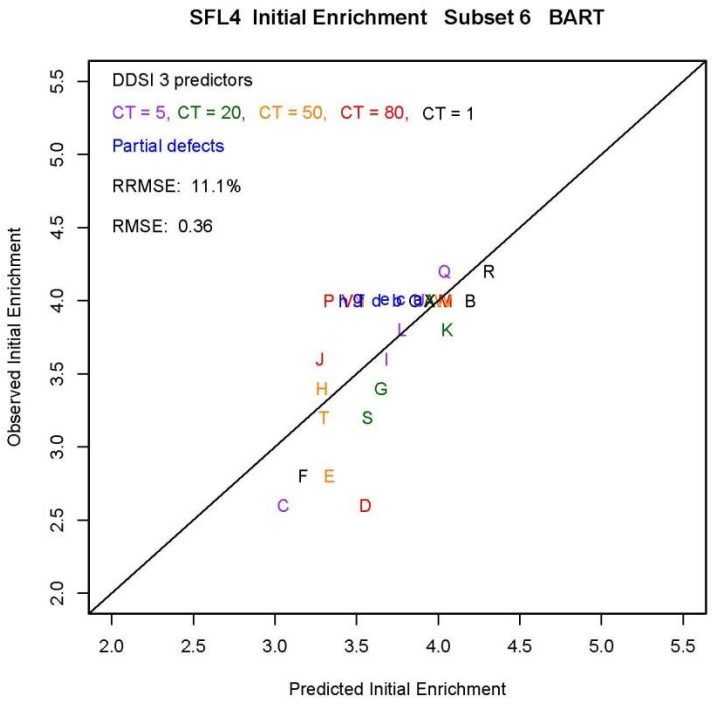

Figure 18. BART Models to Predict IE with PG + TN + DDSI

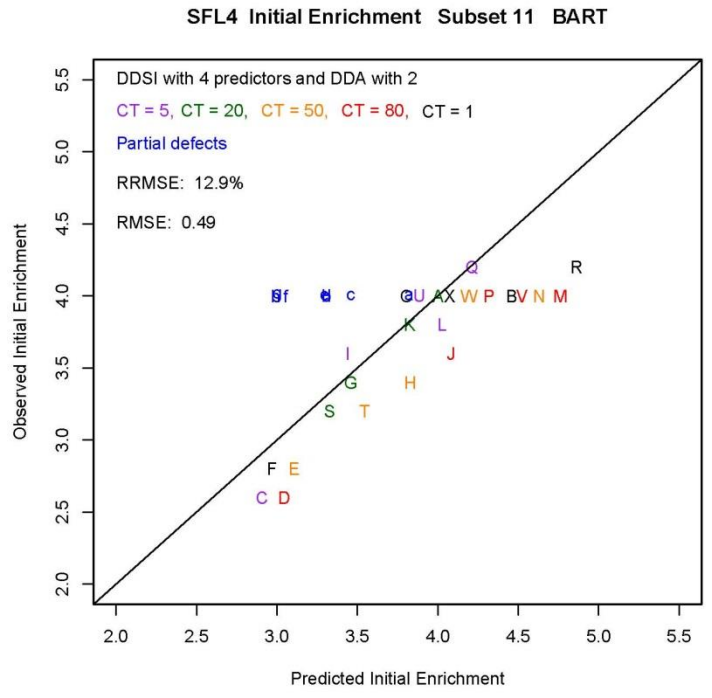

Figure 19. BART Models to Predict IE with PG + TN + DDSI + DDA

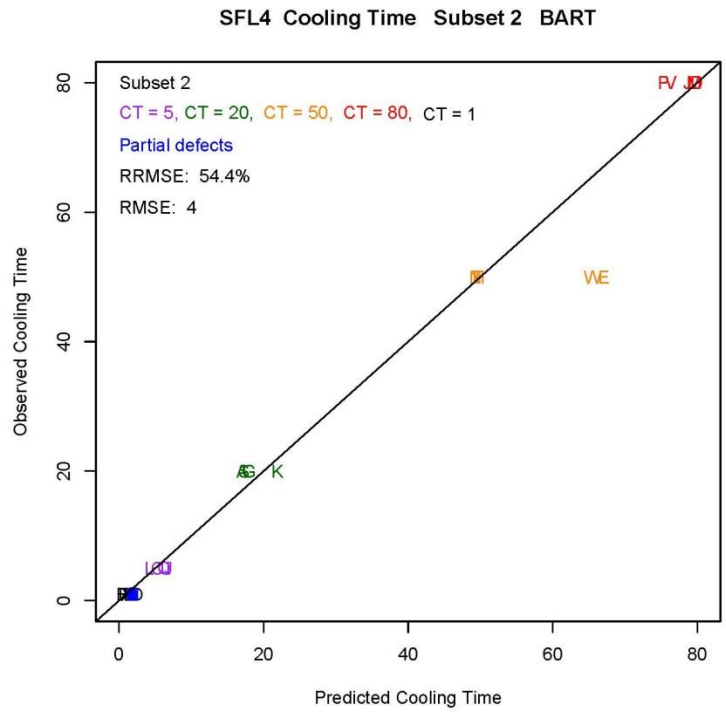

Figure 20. BART Models to Predict CT with PG + TN

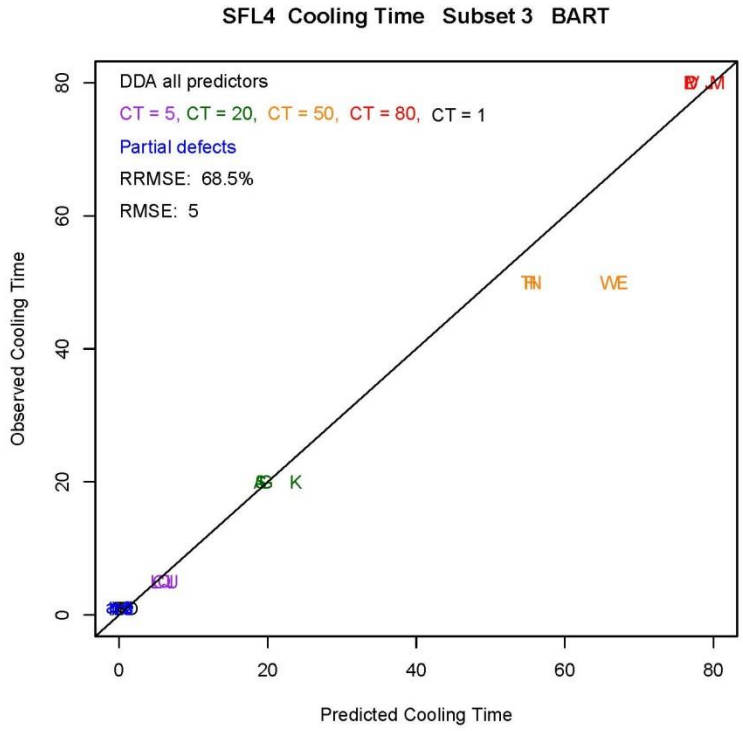

Figure 21. BART Models to Predict CT with PG + TN + DDA

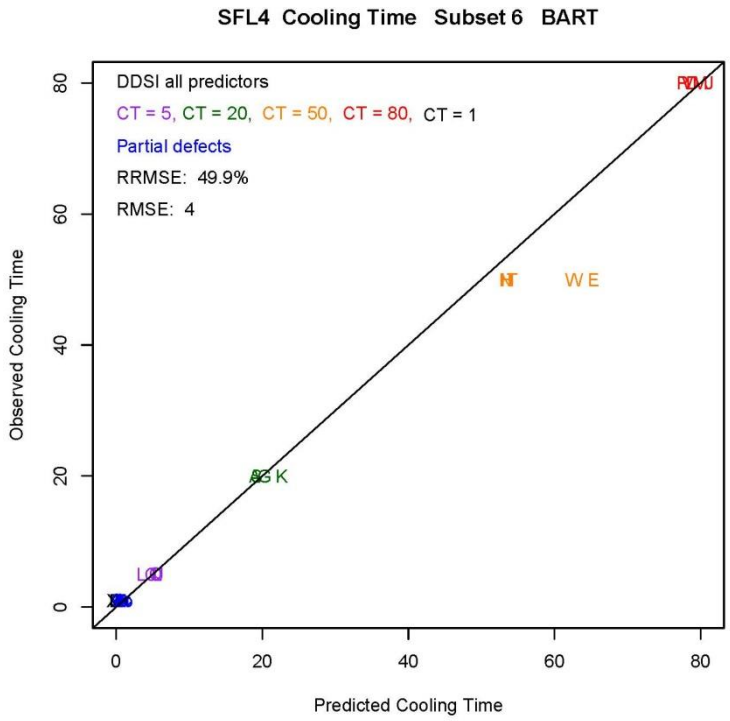

Figure 22. BART Models to Predict CT with PG + TN + DDSI

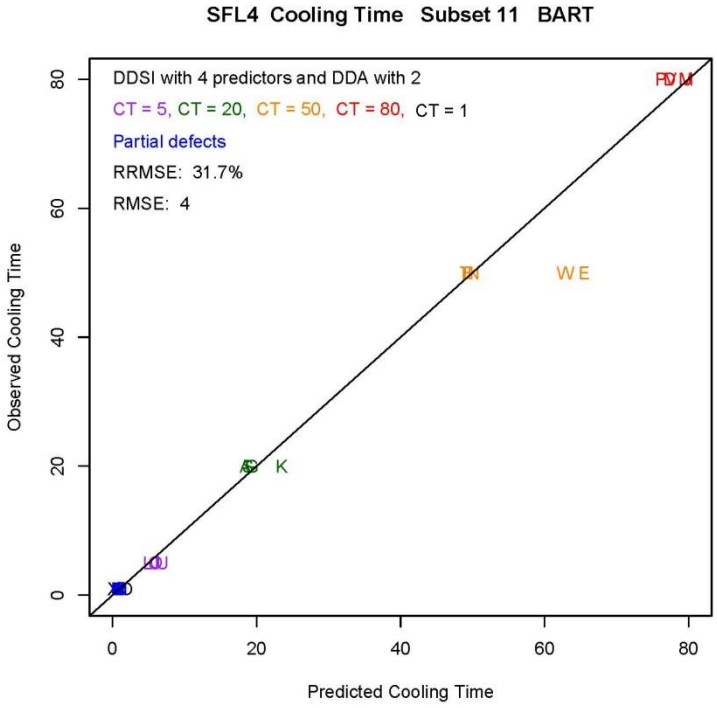

Figure 23. BART Models to Predict CT with PG + TN + DDSI + DDA. Note that the best results were obtained using 4 DDSI and 2 DDA predictors as described in Figure 15.

6.0 Summary

In conclusion, data mining offers a promising technique for taking data from measurements of spent fuel assemblies and estimating plutonium mass, IE, BU, and CT. Data mining for the NGSI-SF project has evolved from uncertainty analyses in which signals were simply put into an analytical solver to obtain predicted values and errors to a better understanding of why the predictions behave as they do. The data in the database used to generate this report were from individual researchers who focused on one NDA technique and the fundamental physics behind it, but data mining brought multiple techniques together to evaluate how combined systems meet the goals of the project. Although all instruments have the capability of helping provide information about Pu, BU, IE, and CT, passive gamma spectral measurements were the best at predicting cooling time, DDA measurements predicted total plutonium content the best, and DDSI (SFL4 predictions) and/or SINRD (SFL2a predictions) helped the most with initial enrichment determinations.

In the future, we will focus data mining efforts on determining how measured data from spent nuclear fuel assemblies meets the goals of the project. We will add multiplication as a predictor to help with partial defect detection and add more "observable cases" to help reduce error estimations (i.e. including training data for CT=50 years would have improved testing predictions at 50 years). We will also have input options for known CT, IE, and/or BU that will help reduce RMSEs in cases where that information is declared. We may also work to create a more sophisticated empirical modeling tool for general use that uses information from the researched physics relationship of each technique.

References:

Burr, T., Conlin, J., Galloway, J., Henzl, V., Hu, J., Menlove, H., Swinhoe, M., Tobin, S., Trellue, H., Ulrich, T., Uncertainty Quantification for New Approaches to Spent Fuel Assay, Nuclear Science and Engineering 172, 180-192, 2012.

Burr, T., Tobin, S., Trellue, H., Measuring the Effects of Data Mining on Inference, LA-UR-13-27096, submitted to Encyclopedia on Information Science and Technology, 2013a.

Burr, T., Trellue, H., Tobin, S,. Favalli, A., Dowell, J., Henzl, V., Mozin, V., Integrated Nondestructive Assay Systems to Estimate Plutonium in Spent Fuel Assemblies Nuclear Science and Engineering 2014.

Galloway, J.D., Trellue, H.R., Fensin, M. L., Broadhead, B., Design and Description of the NGSI Spent Fuel Library with an Emphasis on Passive Gamma Signal, Journal of Nuclear Materials Management, 40(3), 25, 2012.

Henzl, V., "Evaluation of Differential Die-Away Technique Potential in Context of Non-Destructive Assay of Spent Nuclear Fuel," Los Alamos National Laboratory report LA-UR-14-29224, November 2014.

Humphrey, M A., Tobin, S. J., Veal, K. D., The Next Generation Safeguards Initiative's Spent Fuel Nondestructive Assay Project, Journal of Nuclear Materials Management 40 (3), 6-11, 2012.

Kaplan, A. C., Henzl, V., Menlove, H. O., Swinhoe, M. T., Belian, A. P., Flaska, M., Pozzi, S. A., "Determination of Total Plutonium Content in Spent Nuclear Fuel Assemblies with the Differential Die-Away Self-Interrogation Instrument," *Nuclear Instruments and Methods in Physics Research A*, 764 (2014) pp. 347–351.

Trellue, H. R., Tobin, S. J., Burr, T. L., Dowell, L. J., and Weldon, R., "Pu Mass Determination with Integrated Differential Die-Away, Delayed Neutron, Delayed Gamma, Passive Neutron and Passive Gamma Nondestructive Assay Techniques," Los Alamos National Laboratory report LA-UR-13-27744 (2013).

Trellue, H. R., Fensin, M. L., Richard, J. R., Galloway, J. D., and Conlin, J. L., Description of the Spent Nuclear Fuel Used in the Next Generation Safeguards Initiative to Determine Plutonium Mass in Spent Fuel, Los Alamos National Laboratory report LA-UR-11-00300 (2011).

Trellue, H. R., Tobin, S. J., Henzl, V., Mozin, V., Favalli, A., and Fischer, N., Development of Database of SFL2a Signals for Neutron Generator Techniques Proposed for Integration, Los Alamos National Laboratory report LA-UR-13-27714 (2013).

Veal, K.D., LaMontagne, S.A., Tobin, S.J., and Smith, L.E., 2010. NGSI Program to Investigate Techniques for the Direct Measurement of Plutonium in Spent LWR Fuels by Nondestructive Assay, Institute of Nuclear Materials Management 51st Annual Meeting, Baltimore, MD, July 11–16, 2010.

A Relative Root Mean Squared Error for Model Selection

Here we address some issues with using the relative root mean squared error (RRMSE) for model selection instead of the root mean squared error (RMSE). The terms that make up the relative and root mean squared error share the same numerator but the relative root mean squared error divides the numerator by a random variable. Normalizing by a random variable in this context makes little sense to us.

Suppose the variable of interest y_* can be modeled as

$$y_* = f(x_*) + \varepsilon$$

where x_* is a collection of known predictor variables and f is an unknown function that needs to be estimated (f may depend on unknown parameters). ε is an unknown random error term with mean zero and unknown variance σ^2 (possibly constant, possibly not).

Assume that we have a set of training data, (x_*, y_*) pairs, available to estimate f and that an estimate of f is $\hat{f}$. We also have an independent test set, (x, y) pairs, to evaluate how well the estimated function predicts data that was not used in the training phase.

For a given observation x a model prediction of y is $\hat{y} = \hat{f}(x)$. For the ith observation in the test set the residual between the observed and predicted value is $e_i = y_i - \hat{y}_i$. The relative, or scaled residuals, are then $r_i = e_i/y_i$. Writing the scaled residuals in a little more detail we have

$$\begin{aligned} r_i &= \frac{y_i - \hat{y}_i}{y_i} \\ &= \frac{y_i - \hat{y}_i}{f(x_i) + \varepsilon_i} \end{aligned}$$

which shows that the scaling depends on the random error ε_i. It is this dependence on the random error that we find hard to defend.

The root mean squared error computed over the test set is

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

and the relative root mean squared error is

$$\text{RRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2}.$$

A useful property for residuals to have is that on average they are zero when the fitted model is correct, i.e. when the estimate $\hat{f}$ is correct . Using "E" to denote expectation, or average, we have

$$\begin{aligned} E(e_i) &= E(y_i - \hat{y}_i) \\ &= E(y_i) - E(\hat{y}_i) \\ &= f(x_i) - f(x_i) \\ &= 0. \end{aligned}$$

The next to last equality follows when the fitted model is correct. In contrast, even when the fitted model is correct the relative residuals do not have mean zero. The mean of the relative residuals is

$$\begin{aligned} E(r_i) &= E\left(\frac{y_i - \hat{y}_i}{y_i}\right) \\ &= E\left(1 - \frac{\hat{y}_i}{y_i}\right) \\ &= 1 - E\left(\frac{\hat{y}_i}{y_i}\right) \\ &\neq 1 - \frac{E(\hat{y}_i)}{E(y_i)}. \end{aligned}$$

In general the relative residuals will not have mean zero except in very special circumstances. Often when plotting residuals we draw a horizontal at zero to aid the eye in detecting departures from the mean. However, because the scaled residuals do not have a mean of zero, plotting the relative residuals and drawing a

horizontal line at zero is a bit ambiguous. A related issue is finding the variance of the relative residuals, which is useful for identifying statistically unusual observations. In general this will also be difficult.

Looking at the form of the relative residuals, if y_i is small the relative residual will be large unless $y_i - \hat{y}_i$ is very small. Thus, models with small RRMSE will be models that fit the small values of the test data very well.

When the RRMSE is large one suggested approach is to use the median of the absolute value of the relative residuals as the measure of model performance instead of the mean. The justification for this is that relative residuals are skewed which can lead to a large RRMSE. If the goal is to estimate the center of a skewed distribution then using the median is a reasonable option. But evaluating model performance is a different problem. We want to evaluate how well each candidate model fits all of the test data so that the model we ultimately choose predicts future data well. Using the median effectively ignores observations that the model does not predict well.

If the median of the absolute relative residuals is used to choose between models instead of the RRMSE, or preferably the RMSE, the results could be very misleading. As an example, assume we have 5 test set observations and we want to predict cooling time. The five cooling times are: 1, 1, 1, 80 and 80 years. Suppose one of our candidate models predicts all 5 cooling times to be 1 year. The relative residuals have absolute values of: 0, 0, 0, 79/80 and 79/80 with a median relative error of 0%. Because 0% relative error is the smallest possible error this model would be chosen. For comparison, the RRMSE of this model is 62% and the RMSE is 50 years. An estimate of 0% relative error for future data is almost certainly very optimistic.

Note that in this example we could replace the two 80 year cooling times with say 200 year cooling times and yet we would not change the median relative error in spite of the fact that our model is doing even worse at predicting the 200 year cooling times. In contrast the RRMSE and the RMSE would change substantially.

The RRMSE is a poor measure for model selection and using the median of the absolute residuals makes the measure even worse.

Figure 1 show predictions from two models. Clearly, the model corresponding to the blue dots does a much better job of predicting the observed test data even though the RRMSE for the blue dot model is substantially larger than for the red dot model.

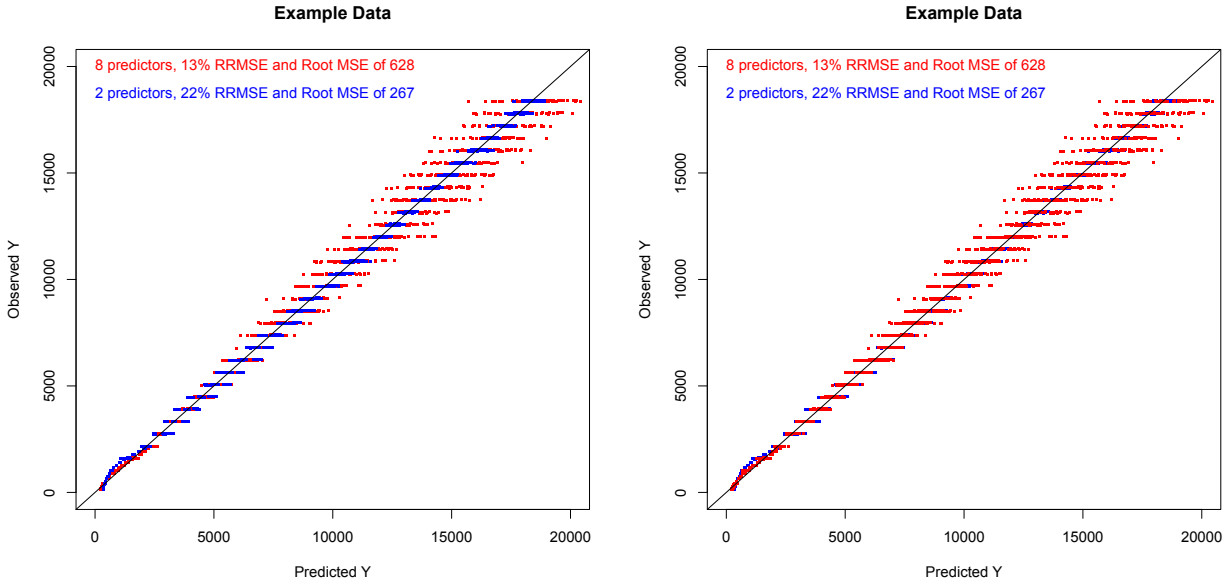

Figure 1: Predictions from two models, one chosen based on RRMSE and the other based on RMSE. The data in the two plots is exactly the same. In the left panel the blue points are plotted on top of the red and in the right panel the red are plotted on top.

We mention that sometimes a weighted RMSE is used to choose between models. This is often done in classification problems where the costs of making a mistake are not equal; e.g. classifying a cancerous tumor as benign is usually a more serious mistake than classifying a benign tumor as cancer. In this context RRMSE can be viewed as a weighted form of RMSE but now the weights are random variables, which is not a very desirable feature of the weights. Typically, if a weighted RMSE is going to be used to choose

between models we would also use a weighted loss function to train the fitted models.

Typically the model that is picked will be used to predict a future observable, e.g. cooling time for some set of predictors. It makes sense then to choose models based on how well they actually predict the test observations. Of course, once a model has been chosen based on root mean squared error the relative root mean squared error can be computed and reported if desired.