HIGH BIAS MACHINE LEARNING FOR SENSITIVITY LIMITS OF ANTINEUTRINO-BASED SAFEGUARDS

[1]Matthew Dunbrack, [1]Anna Erickson, [2]Chris Stewart
[1]Georgia Institute of Technology, Atlanta, Georgia, USA
[2]University of California–Berkeley, Berkeley, California, USA

ABSTRACT
Antineutrino detectors have potential for use as independent, tamper-proof tools for reactor monitoring using the antineutrino flux from reactor cores. The antineutrino count rate needed for reasonable verification relies heavily on the strength of the statistical inference. In this work, we test the statistical limitations of a simulated Advanced Fast Reactor-100 (AFR-100) antineutrino-based safeguards system to detect special nuclear material being removed from the reactor core. After simulating 12 different diversion scenarios, expected antineutrino spectra along with the corresponding variances were created for 3-month collection periods. Support vector machine models were then applied to these spectra for diversion analysis. The diversion predicting powers, or safeguards powers, were inaccurate on a case-by-case basis. However, by generating a large enough sample size, the models converged to a robust, high bias, lower limit safeguards power that would have previously been considered zero using statistical analysis methods.

INTRODUCTION
While the International Atomic Energy Agency (IAEA) has many methods for safeguarding nuclear reactors, remote monitoring techniques are effective for reducing inspection efforts and maximizing equipment utility [1]. Antineutrino detection systems could provide remote monitoring capabilities based on a continuous and isolated stream of antineutrino data. The measured antineutrino flux should then reflect simulated reactors with an identical isotopic core composition. If special nuclear material (SNM) was removed from the core, however, the spectra would not match the simulated data. This system of verification is referred to as the Reactor Evaluation Through Inspection of Near-field Antineutrinos (RETINA) system. Similar to how a retinal scan compares unique blood vessels within an eye to a stored copy, the RETINA system collects an antineutrino spectrum and relates that spectrum to a stored copy from a simulated database [2].

The proposed RETINA system for this study was designed to safeguard the Advanced Fast Reactor (AFR)-100 reactor design. The AFR-100 is a small, fast reactor with a power rating of $250MW_{th}$ [3]. Compared to a previous RETINA study involving the Ultra-long-life Core Fast Reactor (UCFR)-1000, the AFR-100 has a much lower antineutrino source strength. The smaller core volume of the AFR-100 allows the detection system to be closer to the core (17m) compared to the distance for the UCFR-1000 (25m) [4].

ANTINEUTRINO SPECTRA GENERATION
In this work, we applied a simulated RETINA system to the AFR-100 design. Using the method outlined in Stewart et al [4], REBUS-3 [5] software was used for general neutronic computation with MC^2-3 [6] periodically updating the effective microscopic cross sections. After the reactor design was simulated under regular operating conditions, three diversion scenarios locations were selected (Figure 1) based on attractive locations and burnup steps in which a total of one significant quantity

(SQ) of plutonium (8kg) was reasonably accessible. The diversion scenario simulations also involved replacing the removed plutonium with an alternate fuel. The replacement fuels were either natural uranium or low-enriched uranium. The final diversion scenario modification was power manipulation. Since removing special nuclear material would alter the number of antineutrinos collected, a reactor operator could adjust the power of the reactor and burn more or less of the fuel to near-match the expected spectrum. For this study, the scenarios are labeled based on location (1,2,3), fuel replacement ("a" for natural uranium and "b" for low-enriched uranium), and power manipulation ("nominal" for no power alteration, "manipulated" for optimal power alteration).

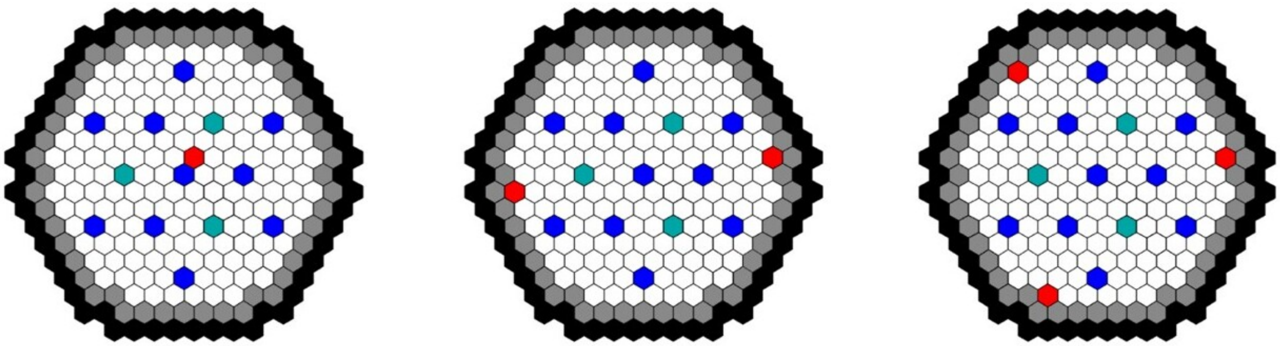

Figure 1: Diverted locations (red) for scenarios 1 (left), 2 (middle), and 3 (right). Figures from Stewart [7].

These scenarios were simulated to determined isotopic fission rates. Based on the antineutrino-fission yield libraries, provided by [8], these isotopic fission rates were translated into antineutrino yields. From these antineutrino spectra, the expected detector event rates were calculated with characteristics matching the AD-1 detector used for the PROSPECT program [9].

STATISTICAL ANALYSIS

After calculating the final antineutrino counts and energies for each scenario, a chi-square goodness of fit test was applied to quantify the diversion scenario deviation from the normal core data. Mathematically, the chi-square test summed the relative deviation for every antineutrino energy bin. As previously described, however, a reactor operator can minimize this deviation with power manipulation. Since greatly increasing or decreasing the power of a reactor also increases the chance of setting off a red flag for the IAEA, a penalty term was added to the chi-square function. The final chi-square statistic, Equation 1, included x, the power manipulation parameter, n_b, the number of antineutrinos counted in bin b over a given count integration period, n'_b, the expected number of antineutrinos counted in bin b, and σ_{norm}, the propagated uncertainty throughout the process up until this point. The power manipulation goodness-of-fit variations are illustrated in 2 for a count integration period of 3 months.

$$\chi^2 = (\sum_b \frac{(n_b - (1+x)n'_b)^2}{n_b}) + (\frac{x}{\sigma_{norm}})^2 \tag{1}$$

Once the chi-square values were calculated for each diversion scenario, they were applied to a distribution, as seen in Equation 2, previously determined by Blennow et al [10] where T_0 is the χ^2

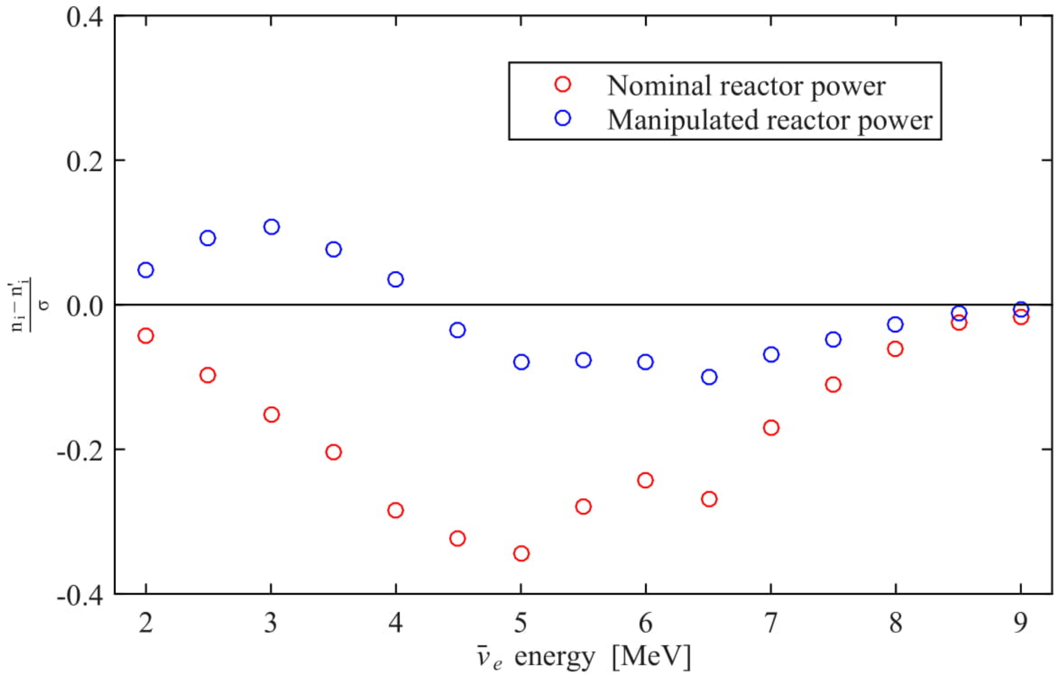

Figure 2: Effect of goodness-of-fit function minimization for diversion 1a. Figure from Stewart [7].

value with an associated distribution, T. The final safeguards power can then be calculated using Equation 3, in which Φ is a cumulative distribution function and T^{α}_{crit} is a critical threshold set for the false positive rate of 0.05. Following IAEA safeguard limits, the safeguards power needed to exceed 0.2 to reasonably detect low-probability diversion scenarios [1].

$$T \sim N(T_0, 2\sqrt{T_0}) \tag{2}$$

$$Power = 1 - \Phi(\frac{T^{\alpha}_{crit} - T_0}{2\sqrt{T_0}}) \tag{3}$$

MACHINE LEARNING ANALYSIS

An alternative method for determining the safeguards power of these diversion scenarios was applied through a machine learning approach. Since supervised machine learning methods require labeled datasets, the antineutrino spectra distributions were sampled and normalized for data training and testing. While an ideal system is trained to safeguard against all possible diversion scenarios, it is impossible to simulate every possible scenario for processing. If the system is robust, however, it could reasonably safeguard against unexpected scenarios. To test the impact of model generalizing, two training/test variations were applied to model generation. The first training/test grouping involved training and testing 12 models, in which each individual model was trained and tested for a specific diversion scenario. These models represent the RETINA system being fully prepared to identify a specific diversion scenario. The second training/test group bins 11 scenarios together for training and the excluded scenario for testing. These models represent the RETINA system being

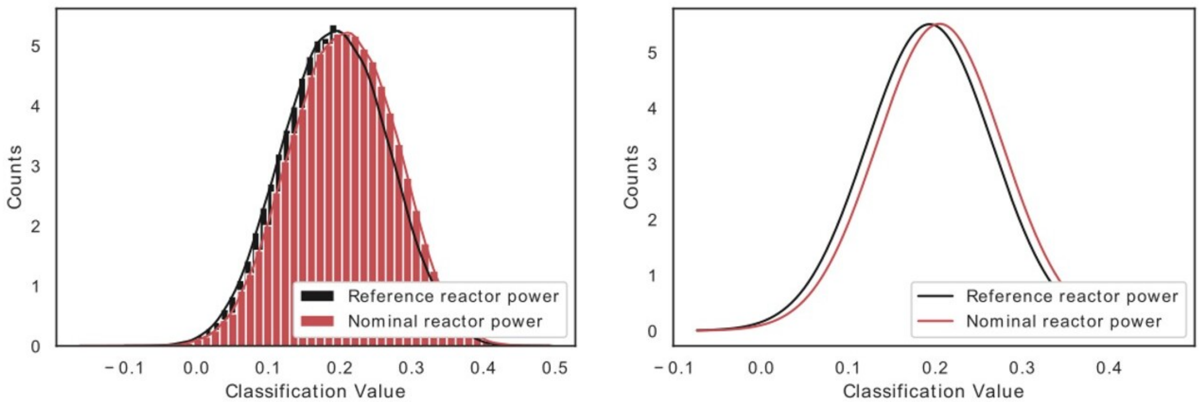

Figure 3: Histogram (left) and normal fit (right) of SVM classification values from individualized model diversion 1a sampling.

prepared for a variety of cases, but encounters a scenario that was never expected. The 12 individually trained and tested models are referred to as “Individualized Models” while the 12 group trained and individually tested models are referred to as “Unseen Models”.

The support vector machine (SVM) model was chosen for diversion classification based on its good soft-margin generalization properties and popularity in many fields [11]. As seen in Figure 3, classification value samplings for the diversion scenarios are similar to the reference case and follow a normal distribution. Based on this fitted distribution with known mean, μ_{cv}, and known standard deviation, σ_{cv}, a similar method as described in the chi-square goodness of fit process was used to determine the final safeguards power. However, while the chi-square distribution accounted for both the diverted and reference distributions, they remained separated through this machine learning approach. This difference was translated by determining the excess safeguards power for the diversion scenario compared to the safeguards power for expected operation (Equation 4).

$$Power = \Phi(\frac{T^{\alpha}_{cv,crit} - \mu'_{cv}}{\sigma'_{cv}}) - \Phi(\frac{T^{\alpha}_{cv,crit} - \mu_{cv}}{\sigma_{cv}}) \quad (4)$$

RESULTS

After applying the chi-square statistical method to the antineutrino count distributions, the results were placed in Table 1. Safeguards powers lower than 10^{-16} were approximated to be zero due calculation floating-point precision limits. For the SVM models, the mean results after generating 10^6 sample points per model with an 80/20 train-test split for 100 iterations can be seen in Table 1. While the chi-square approach had a minimum value of 0 for low safeguards power scenarios, it was possible for the SVM model approach to lead to negative safeguards powers. For cases in which the diversion scenario distribution is virtually identical to the reference distribution, the stochastic nature of the sampling method could lead to a larger reference safeguards power than the diversion scenario safeguards power. Since a negative safeguards power is not realistic and it implies that the safeguards power is lower than our model precision, 0 was set as a lower limit for the safeguards power results. The standard deviation for these models are in Table 2.

Table 1: Mean Safeguards Powers

	Chi-Square Analysis	Mean Safeguards Powers for Individualized Models	Mean Safeguards Powers for Unseen Models
1a Nominal	2.21 x 10^{-2}	4.28 x 10^{-2}	3.84 x 10^{-2}
1a Manipulated	2.52 x 10^{-13}	2.08 x 10^{-2}	1.34 x 10^{-2}
1b Nominal	8.56 x 10^{-4}	3.63 x 10^{-2}	3.55 x 10^{-2}
1b Manipulated	6.34 x 10^{-12}	2.28 x 10^{-2}	1.40 x 10^{-2}
2a Nominal	5.63 x 10^{-12}	1.26 x 10^{-2}	1.18 x 10^{-2}
2a Manipulated	0	3.82 x 10^{-3}	3.27 x 10^{-3}
2b Nominal	0	1.01 x 10^{-2}	0
2b Manipulated	0	2.70 x 10^{-3}	0
3a Nominal	4.69 x 10^{-12}	8.93 x 10^{-3}	7.91 x 10^{-3}
3a Manipulated	0	3.46 x 10^{-3}	2.54 x 10^{-3}
3b Nominal	0	6.65 x 10^{-3}	0
3b Manipulated	0	1.47 x 10^{-3}	0

CONCLUSIONS

While the chi-square approach leads to an exact solution, half of the results are lower than the 10^{-16} precision of the software and therefore considered zero. Most of these low-safeguards powers are for the manipulated power scenarios, in which the power is shifted to minimize the overall antineutrino count deviation from expectations. The benefit for this method, however, is that these safeguards powers are exact solutions. Also, this method does not require any simulated diversion scenarios for system development. Each case is compared directly to the reference scenario for deviation. Regardless of the benefits of statistical analysis, all of the safeguards powers associated with this method are extremely low compared to the 0.2 IAEA low-probability limit.

For all scenarios, the individualized models are the best at predicting diversion scenarios. Not only are all of the correlating safeguards powers higher than previously determined, but they are near-exact after 100 iterations. While the best SVM-derived power almost doubles the best chi square-derived power, significant results are seen for the manipulated scenarios in which the power was previously considered zero. With all relative powers known with little error, we conclude that the RETINA system is better prepared for scenarios in which (1) plutonium is pulled from fewer assemblies and closer towards the center of the core, (2) plutonium is replaced with natural uranium rather than LEU, and (3) the power of the reactor is not manipulated.

The unseen models, while lower than the individualized models, led to significantly higher safeguards powers compared to the safeguards powers from statistical analysis. This method, however, led to a few zero safeguards powers. Assuming the true safeguards powers for this method are slightly lower than the relative values from the individualized models, it can be assumed that the precision limit for this model is around the order of 10^{-4}. So while a generalized model can

Table 2: Standard Deviations for Safeguards Powers

	Standard Deviation for Individualized Models	Standard Deviation for Unseen Models
1a Nominal	6.94×10^{-18}	6.10×10^{-4}
1a Manipulated	1.04×10^{-17}	1.26×10^{-3}
1b Nominal	6.94×10^{-18}	6.41×10^{-4}
1b Manipulated	0	1.15×10^{-3}
2a Nominal	5.20×10^{-18}	1.32×10^{-3}
2a Manipulated	0	1.31×10^{-3}
2b Nominal	0	1.77×10^{-3}
2b Manipulated	0	1.38×10^{-3}
3a Nominal	0	1.33×10^{-3}
3a Manipulated	0	1.33×10^{-3}
3b Nominal	0	1.75×10^{-3}
3b Manipulated	0	1.51×10^{-3}

safeguard relatively well against most diversion scenarios, some cases have safeguards powers that fall below a precision limit.

For future work, this methodology should be further optimized and generalized to fit different reactor core designs and diversion scenarios. While none of the safeguards powers in this study exceeded the 0.2 IAEA low-probability limit, different reactor-detector systems could improve the sensitivity of the proposed RETINA system. Other machine learning methods, such as deep neural networks, can also be applied to further increase the safeguards powers.

ACKNOWLEDGMENTS

This work was funded by the Consortium for Monitoring, Technology, and Verification under Department of Energy National Nuclear Security Administration award number DE-NA0003920.

REFERENCES

[1] "Iaea safeguards glossary: 2001 edition," Tech. Rep. 3, International Atomic Energy Agency, Vienna, 2002.

[2] C. Stewart and A. Erickson, "Antineutrino analysis for continuous monitoring of nuclear reactors: Sensitivity study," *J. Appl. Phys.*, vol. 118, no. 164902, 2015.

[3] C. Grandy, J. Sienicki, A. Moisseytsev, L. Krajtl, M. Farmer, and T. K. Kim, "Advanced fast reactor - 100 (afr-100) report for the technical review panel," Tech. Rep. ANL-ARC-288, Argonne National Laboratory, Nuclear Engineering Division, Argonne, Illinois, June 2014.

[4] C. Stewart, A. Abou-Jaoude, and A. Erickson, "Employing antineutrino detectors to safeguard future nuclear reactors from diversions," *Nat Commun*, vol. 10, no. 3527, 2019.

[5] B. Toppel Tech. Rep. 83-2, Argonne National Laboratory, 1983.

[6] H. Henryson, B. Toppel, and M. Stenberg Tech. Rep. 8144, Argonne National Laboratory, 1976.

[7] C. Stewart, *Antineutrino-Based Safeguards for Ultra-High Burnup Fast Reactors*. Ph.D. thesis, Georgia Institute of Technology, 2016.

[8] P. Huber, "Determination of antineutrino spectra from nuclear reactors," *Phys. Rev. C*, vol. 84, no. 024617, 2011.

[9] J. Ashenfelter *et al.*, "The prospect physics program," *J. Phys. G: Nucl. Part. Phys.*, vol. 43, oct 2016.

[10] M. Blennow, P. Coloma, P. Huber, and T. Schwetz, "Quantifying the sensitivity of oscillation experiments to the neutrino mass ordering," *J. High Energ. Phys.*, no. 28, 2014.

[11] X. Qi, S. Silvestrov, and T. Nazir, "Data classification with support vector machine and generalized support vector machine," *AIP Conference Proceedings*, vol. 1798, no. 1, p. 020126, 2017.