

Characterization of Neutron Detection Systems with Confidence via Statistical Parameter Correction

Phillip Boone
NuSAFE, Inc.
Oak Ridge, USA
pboone@nuSAFE.com

Abstract—Neutron detection systems utilize statistical alarm techniques where a measured false alarm rate (FAR) can vary drastically from the FAR predicted by a theoretical model. The ability to set an alarm threshold that results in a practically controlled FAR is crucial to characterize detector sensitivity with both accuracy and precision. A generalized and automated method is presented to statistically evaluate FAR performance by assuming that the FAR itself is not deterministic, but a normal stochastic process over a specific parameter to be corrected that will hereafter be referred to as the *correction*. In this manner, a specific correction results in not only a point estimate of FAR, but also a confidence interval. The central objective is focused exclusively on characterization assuming that experiments are executed in a tightly controlled environment so that an accurate comparison is enabled across detectors. Once a correction is calculated, the estimated FAR is only assumed accurate in a similar environment for sensitivity evaluation. Initially, the calculated correction factor was used to compare FARs across various distributions including normal, corrected normal, Poisson, and a simplified normal distribution. Later verification data sets were used to empirically demonstrate the rate of containment of measured confidence coefficients using two detectors of different technology. A second application uses the correction method to improve the signal-to-noise ratio metric to agree more with dynamic sensitivity results. Finally, a third application studies the effect of altering the duration of background acquisition on FAR performance.

I. INTRODUCTION

BY partitioning a notably large set of background data into potentially thousands of mutually exclusive periods, a distinct alarm threshold can be calculated using a specific correction and statistics on an individual background period. By applying this distinct threshold, a single alarm rate sample can be calculated over the remaining background data. Each subsequent background period is used likewise to create additional alarm rate samples for the entire data set. A maximum likelihood problem is formulated in section II-D1 assuming that a detector's false alarm rate (FAR) is a normal random process of correction. Using numerical optimization in section II-E, a first-order search is utilized to determine a correction to meet a target FAR.

False alarm performance of the corrected model is compared to three statistical models in section III-C. In section IV, additional background data is gathered to verify that the estimated correction results in the desired FAR with statistical confidence.

In section V three neutron detection systems demonstrate that the signal-to-noise ratio (SNR) metric may result in a theoretical sensitivity that contradicts sensitivity measured with a more rigorous dynamic test [1] according to American National Standards Institute (ANSI). Section V-A2 introduces a correction to SNR and section V-C demonstrates that the correction results in a sensitivity that agrees more with the rigorous ANSI test.

Section VI studies the effect of reducing the duration of background acquisition from 5 minutes down to a minute. As a result, system sensitivity decreased via the increase in critical value required to meet a target FAR. In addition, the overall precision of FAR drastically decreased via the estimated increase in standard deviation of FAR by a factor of 28.

In the context of the experiments performed, FAR is simply a scaled version of significance level, critical value indicates a specific alarm threshold, and standard critical value represents the number of standard deviations above the mean for an alarm threshold. FAR is used in place of significance level for convenience when presenting measured false alarm rates.

II. STATISTICAL MODEL

A. Data Collection Periods

1) *Large Experimental Set of Data*: Assume that count rate samples for a neutron radiation detector under natural ambient background conditions follow a normal distribution with random variable $D \sim \mathcal{N}(\mu_D, \sigma_D^2)$ and a realization of N_D random samples is the vector $\vec{D} \triangleq [D_1, \dots, D_{N_D}]^T$.

2) *Background Partitions of Data*: These samples are partitioned into N_{PB} non-overlapping background periods of length N_B with the i -th period defined as $\vec{B}_i \triangleq [D_{N_B(i-1)+1}, \dots, D_{N_B(i-1)+N_B}]^T \triangleq [B_{i,1}, \dots, B_{i,N_B}]^T$ with $i \in \{1, 2, 3, \dots, N_{PB}\}$.

For each background period i the statistic that is a minimum variance unbiased estimator (MVUE) of mean μ_D is

$$\hat{\mu}_{B,i} \triangleq \frac{1}{N_B} \sum_{j=1}^{N_B} B_{i,j}$$

For each background period i the statistic that is a MVUE of variance σ_D^2 is

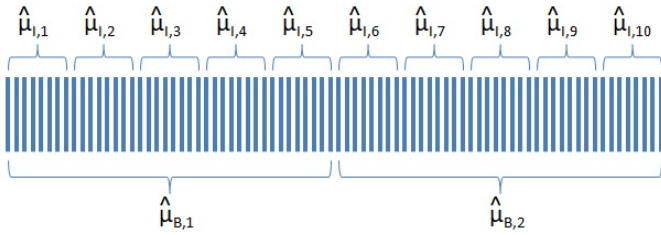


Fig. 1. Example Data Partitions : $N_D = 80$ Samples Partitioned into $N_{PB} = 2$ Background Periods of $N_B = 40$ Samples and $N_{PI} = 10$ Integration Periods of $N_I = 8$ Samples

$$\hat{\sigma}_i^2 \triangleq \frac{1}{N_B - 1} \sum_{j=1}^{N_B} (B_{i,j} - \mu_D)^2$$

The number of samples N_B should be chosen as large as possible within system initialization in a known environment without artificial sources.

3) *Integration Partitions of Data:* The samples in vector \vec{D} are partitioned again into $N_{PI} \ll N_{PB}$ non-overlapping integration periods of length N_I with the i -th period defined as $\vec{I}_i \triangleq [D_{N_I(i-1)+1}, \dots, D_{N_I(i-1)+N_I}]^T \triangleq [I_{i,1}, \dots, I_{i,N_I}]^T$ with $i \in \{1, 2, \dots, N_{PI}\}$.

For each period i the statistic that's a MVUE of mean μ_D is

$$\hat{\mu}_{I,i} \triangleq \frac{1}{N_I} \sum_{j=1}^{N_I} I_{i,j}$$

The number of samples N_I is typically chosen to optimize system detection sensitivity for a given source strength and speed.

B. Determine the Theoretical Parameters of a Hypothesis Test

Since it can be shown that the test statistic, a linear combination of normal samples, also follows a normal distribution according to $\hat{\mu}_{I,i} \sim \mathcal{N}(\mu_D, \frac{\sigma_D^2}{N_I})$, if a one-sided hypothesis test is used then either one of the following choices are possible:

- 1) **Null Hypothesis H_0 (radiation is background)**
 - a) $\hat{\mu}_{I,i} = \mu_D$
- 2) **Alternate Hypothesis H_1 (radiation is significantly higher than background)**
 - a) $\hat{\mu}_{I,i} > \mu_D$

The critical region of statistic $\hat{\mu}_{I,i}$ is the region where background radiation is determined to erroneously contain an artificial source (H_0 is rejected) and is defined for some constant $\bar{Z}_\alpha > 0$ such that

$$C_i \triangleq \left\{ \hat{\mu}_{I,i} : \hat{\mu}_{I,i} > \mu_D + (\bar{Z}_\alpha) \frac{\sigma_D}{\sqrt{N_I}} \right\} \quad (1)$$

A desired significance level $\alpha \in [0, 1]$ together with the probability P_{H_0} of the error region of H_0 constrain a specific solution for constant \bar{Z}_α according to the equation

$$P_{H_0}(C_i) = \alpha \quad (2)$$

1) *Determining a Value for the Standard Critical Value Constant:* The standardized test statistic for the i -th period of an integration partition is

$$Z_i \triangleq \frac{\hat{\mu}_{I,i} - \mu_D}{\left(\frac{\sigma_D}{\sqrt{N_I}} \right)}$$

and the critical region C_i can be rewritten in terms of the probability of a standard normal distribution (or the cumulative distribution function $\Phi(Z)$) as

$$C_i \triangleq \{Z_i : P(Z_i > \bar{Z}_\alpha) = \alpha\} = \{Z_i : 1 - \Phi(Z_i) = \alpha\}$$

where the threshold \bar{Z}_α , referred to as the standard critical value of the hypothesis test, and corresponding significance level α can be looked up in a standard normal Z table.

2) *Critical Value of the Hypothesis Test:* Once the standard critical value \bar{Z}_α of the standard normal distribution is found, the corresponding critical value $\bar{\mu}_{I,i}$ for the statistic $\hat{\mu}_{I,i}$ is

$$\bar{\mu}_{I,i} = \mu_D + (\bar{Z}_\alpha) \frac{\sigma_D}{\sqrt{N_I}} \quad (3)$$

3) *Verification of Significance Level:* In order to facilitate calculating the value of the error probability in (2), an indicator function $1_\mu(\mu) : \mathbb{R} \rightarrow \{0, 1\}$ is defined as

$$1_\mu(\mu) = \begin{cases} 1 & \text{when } \mu > \mu_D + (\bar{Z}_\alpha) \frac{\sigma_D}{\sqrt{N_I}} \\ 0 & \text{when } \mu \leq \mu_D + (\bar{Z}_\alpha) \frac{\sigma_D}{\sqrt{N_I}} \end{cases}$$

where the expected value

$$E[1_\mu(\hat{\mu}_{I,i})] = P_{H_0}(C_i) = \alpha$$

can be estimated using a sample mean of $1_\mu(\hat{\mu}_{I,i})$ where

$$\hat{\alpha} = \frac{1}{N_{PI}} \sum_{i=1}^{N_{PI}} 1_\mu(\hat{\mu}_{I,i}) \quad (4)$$

C. Determine the Practical Significance Level of a Hypothesis Test

Note that in section II-B the population mean μ_D and variance σ_D^2 of a detector are not known exactly. Therefore, assume that a background period of sufficient length N_B is used so that $\hat{\mu}_{B,i} \approx \mu_D$ and $\hat{\sigma}_i^2 \approx \sigma_D^2$ are substituted in the formulas of the previous section.

1) *Execution of a Very Large Number of Experiments via Background Partitions:* Typically, when a user performs a false alarm test using a radiation detector the user programs a standard critical value, a single background period is acquired, statistics are performed resulting in an alarm threshold, and this threshold is applied using a live system for a long duration in order to verify a specified FAR.

Instead, if a large number N_D of data samples are recorded and if these samples are partitioned as in II-A, then thresholds can be calculated and a false alarm experiment can be performed in post-processing for every possible background period in the entire data set. Also, if the theoretical standard

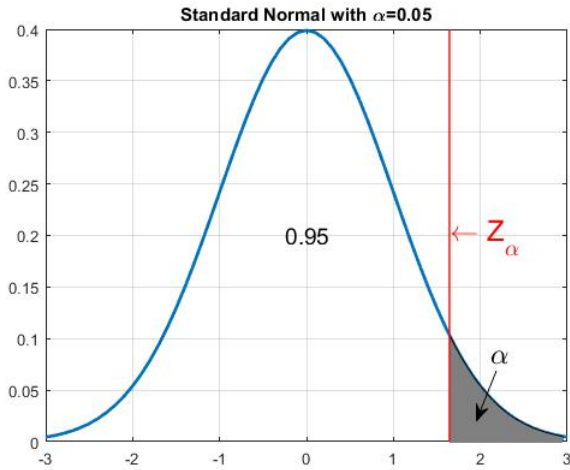


Fig. 2. Standard Critical Value

critical value \bar{Z}_α results in a measured significance level that doesn't agree with the theoretical significance α within acceptable tolerance, then adjusted standard critical values can be used to iterate through the whole process again.

Furthermore, significance level estimates $\hat{\alpha}$ can be computed for each background and statistics can be calculated on the resulting set to obtain realistic false alarm performance metrics across backgrounds.

2) *Verification of Significance Level using Empirical Data:* The indicator function defined in II-B3 is redefined using the j -th actual background parameter estimates as $1_{\mu,j}(\mu) : \mathbb{R} \rightarrow \{0, 1\}$ according to

$$1_{\mu,j}(\mu) = \begin{cases} 1 & \text{when } \mu > \hat{\mu}_{B,j} + (\bar{Z}_\alpha) \frac{\hat{\sigma}_j}{\sqrt{N_I}} \\ 0 & \text{when } \mu \leq \hat{\mu}_{B,j} + (\bar{Z}_\alpha) \frac{\hat{\sigma}_j}{\sqrt{N_I}} \end{cases}$$

where the corresponding estimated significance level for the j -th background period is

$$\hat{\alpha}_j = \frac{1}{N_{PI}} \sum_{i=1}^{N_{PI}} 1_{\mu,j}(\hat{\mu}_{I,i}) \quad (5)$$

Theorem 1 (Unique Significance): For a one sided upper hypothesis test, if significance level α is written as a function of standard critical value Z_α in the form $\alpha(Z_\alpha) : \mathbb{R} \rightarrow [0, 1]$ then the function is strictly decreasing over the domain of the function as illustrated in Fig.2. Therefore, given a unique significance level $\alpha_0 \in [0, 1]$ there exists a corresponding unique standard critical value Z_{α_0} .

3) *Experimental Error:* Assume that the theoretical standard critical value \bar{Z}_α does not result exactly in the desired significance level α due to experimental data not fitting the model exactly, but instead \bar{Z}_α results in the significance value $\tilde{\alpha}$. The significance level experimental error is then defined as

$$\epsilon_\alpha \triangleq \tilde{\alpha} - \alpha$$

Assume that if experimental data approximately retain the property of Theorem 1, then there exists a unique experimental standard critical value \hat{Z}_α that results in the theoretical significance level α within tolerable precision. The standard critical value experimental error is then defined as

$$\epsilon_Z \triangleq \bar{Z}_\alpha - \hat{Z}_\alpha$$

4) *Random Variable of an Experimental Standard Critical Value:* Assume that the significance level estimates $\{\hat{\alpha}_j : j \in \{1, \dots, N_{PB}\}\}$ are a realization of N_{PB} random samples from a random variable with a standard critical value \hat{Z}_α that results in the ideal significance level α . Samples are assumed to approximately follow a normal distribution [2][pp.216] for sufficiently large N_{PB} according to

$$\hat{\alpha}_j \sim \mathcal{N}(\alpha, \sigma_\alpha^2) \quad (6)$$

5) *Practical Significance Level Estimate:* Therefore, the sample mean of $\hat{\alpha}_j$ is the MVUE of significance level

$$\hat{\mu}_\alpha \triangleq \frac{1}{N_{PB}} \sum_{j=1}^{N_{PB}} \hat{\alpha}_j \quad (7)$$

The sample variance of $\hat{\alpha}_j$ is the MVUE of significance level variance

$$\hat{\sigma}_\alpha^2 \triangleq \frac{1}{N_{PB} - 1} \sum_{j=1}^{N_{PB}} (\hat{\alpha}_j - \hat{\mu}_\alpha)^2 \quad (8)$$

D. *Determine the Practical Standard Critical Value of a Hypothesis Test*

Since samples of $\hat{\alpha}_j$ are assumed to approximately follow a normal distribution with mean α and variance σ_α^2 , the probability density function (PDF) is

$$f(\hat{\alpha}_j) = \frac{1}{\sigma_\alpha \sqrt{2\pi}} \exp \left\{ -\frac{(\hat{\alpha}_j - \alpha)^2}{2\sigma_\alpha^2} \right\} \quad (9)$$

1) *Maximum Likelihood Estimate of Standard Critical Value:* The only quantity in (9) that is a function of standard critical value Z_α is $\hat{\alpha}_j$, which can be thought of as either a continuous-time stochastic process [3] over parameter Z_α or a random variable over the family of PDFs [2] indexed on the parameter Z_α written as $\hat{\alpha}_j(Z_\alpha)$ when used in the likelihood functions.

The joint probability of N_{PB} random samples taken from the PDF in (9) results in the likelihood function

$$\begin{aligned} L(Z_\alpha) &= \prod_{j=1}^{N_{PB}} f(\hat{\alpha}_j(Z_\alpha)) \\ &= \frac{1}{\sigma_\alpha^{N_{PB}} (2\pi)^{\frac{N_{PB}}{2}}} \exp \left\{ -\frac{\sum_{j=1}^{N_{PB}} (\hat{\alpha}_j(Z_\alpha) - \alpha)^2}{2\sigma_\alpha^2} \right\} \end{aligned}$$

and the corresponding log likelihood function is

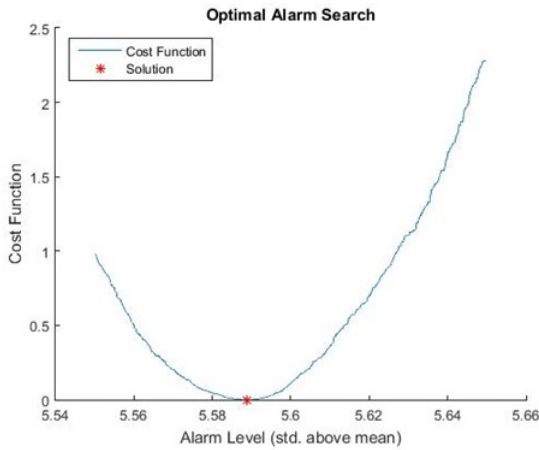


Fig. 3. Solution to the Cost Function

$$l(Z_\alpha) = -\frac{N_{PB}}{2} \left[\log(\sigma_\alpha^2) + \log(2\pi) + \frac{1}{N_{PB}\sigma_\alpha^2} \sum_{j=1}^{N_{PB}} (\hat{\alpha}_j(Z_\alpha) - \alpha)^2 \right]$$

Therefore the maximum likelihood estimate of standard critical value is

$$\hat{Z}_\alpha = \arg \max_{Z_\alpha} \{l(Z_\alpha)\} \quad (10)$$

E. Optimization Problem Formulation

1) *Cost Function*: In order to facilitate computation of the likelihood function in (10), all constants independent of Z_α are eliminated and the negative sign is removed resulting in the following minimization cost function

$$C(Z_\alpha) = \sum_{j=1}^{N_{PB}} (\hat{\alpha}_j(Z_\alpha) - \alpha)^2 \quad (11)$$

Therefore, the maximum likelihood estimate of (10) is equivalent to minimizing the mean squared error of significance level α over standard critical value.

2) *Optimization Problem*: Using the cost function from (11) in a minimization problem results in the following optimization problem formulation:

$$\begin{aligned} \hat{Z}_\alpha &= \arg \min_{Z_\alpha} \{C(Z_\alpha)\} \\ &= \arg \min_{Z_\alpha} \left\{ \sum_{j=1}^{N_{PB}} (\hat{\alpha}_j(Z_\alpha) - \alpha)^2 \right\} \end{aligned} \quad (12)$$

3) *Optimization Problem Solver*: The optimization problem formulated by (12) can be solved using various first order numerical methods, but care must be taken to use a method that is tolerant of noise as illustrated by the solution in Fig.3. In this case, the problem was solved via simulated annealing [4].

F. Generalization of Parameter Correction

The upper hypothesis test utilized normal samples $\{\hat{\mu}_{I,i} : i \in \{1, \dots, N_{PI}\}\}$ while adjusting the parameter Z_α to yield a target theoretical significance α . However, the model can be made even more flexible by slightly altering definitions including the indicator function in (5). Instead let $\mu \in \mathbb{R}$ be a measurable random variable that follows all axioms of probability with the indicator function $1_{\mu,j}(\mu, \beta_j, \theta_\alpha) : \mathbb{R} \rightarrow \{0, 1\}$ where β_j is a set of statistics on background period j and the parameter $\theta_\alpha \in \mathbb{R}$ is the target parameter to correct. The indicator function must also be a simple inequality for an upper hypothesis test, $\hat{\alpha}_j$ must be approximately normal according to (6), and experimental samples must approximately adhere to Theorem 1. In this case, a much more flexible framework results allowing modification to other system factors such as distribution, correction parameter θ_α , sampling techniques as in section III-B, or another system parameter as in section VI. As factors are adjusted the impact on measured FAR performance values $\hat{\mu}_\alpha$ and $\hat{\sigma}_\alpha^2$ could be evaluated.

III. ERROR ANALYSIS USING A NEUTRON DETECTOR

A. Theoretical Probability Distribution vs. Detector Probability Histogram

In order to qualitatively illustrate the disparity between model and detector data, in Fig.4 a histogram was created for neutron data and scaled theoretical PDFs were overlaid for the normal models. Curie's criteria was used where variance is assumed equal to the sample mean of data [5][pp.75]. Matlab's histogram function was used with the "Normalization" parameter set to "probability" in order to create a probability mass function (PMF). Both PDFs were scaled to have the same maximum value as the PMF. Not only is the histogram data highly discretized but the overall shape does not match either normal model at lower count rates.

B. Correlated Sampling

During problem formulation in section II-A the term "random samples" was used to describe the N_D number of data samples acquired. This is a fundamental assumption often referred to as a set of independent, identically-distributed samples of data. The integration samples as defined in section II-A3 follow this convention of sampling. However, the neutron detector data acquired are not sampled independently with non-overlapping segments, but in a temporally correlated and overlapping manner with a slightly different definition of $\vec{I}_i \triangleq [D_{(i-N_I)+1}, \dots, D_{(i-N_I)+N_I}]^T \triangleq [I_{i,1}, \dots, I_{i,N_I}]^T$ with $i \in \{N_I, N_I + 1, \dots, N_D\}$. In other words, every detector sample in \vec{D} starting with number N_I is processed as an average over a rolling buffer where at each iteration the oldest

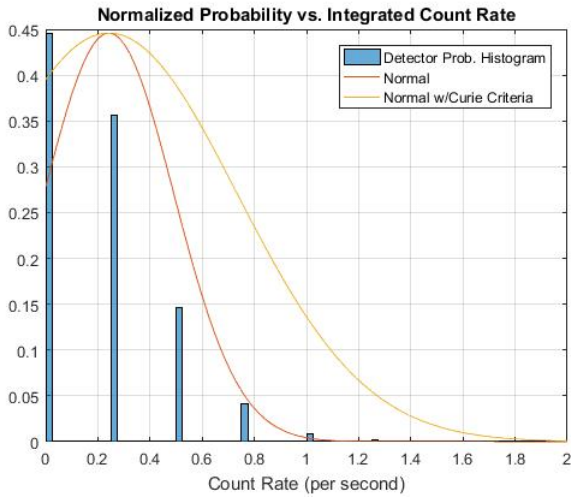


Fig. 4. Qualitative Probability Comparison

sample is discarded from the beginning and a new sample is inserted to the end. Using this method of sampling is likely a significant contributing factor to the discrepancy between the theoretical standard critical value \bar{Z}_α and the practical value \hat{Z}_α .

If independent sampling were performed as originally defined the sample rate would be decimated by a factor of N_I . This would result in under sampling where a neutron detector could miss a moving source at sufficient speed that would have been otherwise detected using correlated sampling. Correlated sampling allows the increased sensitivity of averaging periods while retaining the sample rate.

1) *Alternate Alarm Definition to Compensate for Correlation:* A single outlier D_O in \bar{D} will persist in the $N_I - 1$ correlated samples before and $N_I - 1$ correlated samples after D_O . Correlated samples $\hat{\mu}_{I,O-N_I+1}$ through $\hat{\mu}_{I,O+N_I+1}$ all contain the outlier D_O . When processing false alarms for a set of data this is evident when false alarms themselves are observed as temporally correlated.

Consequently, when counting alarms all alarms that occur within a distance of $N_I - 1$ samples of another alarm are counted as a single alarm event. The simple heuristic is employed where all correlated samples are first compared as normal to the threshold $\bar{\mu}_{I,j}$ for the j -th background period and each alarm index is appended to a list. Alarms are processed in sequential order. Assuming that the i -th sample is an alarm then if any alarm exists after the i -th sample in the interval $[\hat{\mu}_{I,i+1}, \dots, \hat{\mu}_{I,i+N_I-1}]$ then the i -th alarm is removed from the alarm set. The same heuristic continues at the subsequent alarm. The whole process is repeated for all background periods.

The alarm heuristic described is incorporated into a class that is used for all presented false alarm measurements of correlated samples.

TABLE I
FALSE ALARM RATE (per hour)

Distribution	Critical Value	Target FAR	Measured FAR
Normal	1.2459	0.5	5.5195
Poisson ^a	10	0.5	0.4952
Normal w/Curie Criteria	2.2258	0.5	0
Corrected Normal	1.6673	0.5	0.1311

C. Theoretical Model Significance vs. Measured False Alarm Rates

In order to demonstrate the discrepancy between theoretical and measured significance levels, several theoretical models were evaluated. For convenience estimated significance levels were scaled to false alarm rates measured per hour. An 8 segment neutron detector was used to collect 68 hours of data with a sample period of 0.2 seconds, an integration period of 4 seconds, and a background period using the first 5 minutes of data. The requisite statistics for each model were computed on a single background period. The resulting critical values were employed on integrated samples using a similar hypothesis test of section II-B and compensated according to section III-B1. Each distribution was configured with the theoretical critical value to meet a target FAR of 0.5 alarms per hour.

Refer to the results compiled in Table I according to distribution. Since a Poisson model is a discrete distribution of natural numbers, measurements of count rates included in the correlated sampling of III-B are not possible. In this case the unprocessed integer counts were used at the 0.2 second sampling interval. Again, results include a normal model assuming the Curie criteria [5][pp.75]. The last entry labeled "Corrected Normal" uses a corrected standard critical value that was estimated using the theory developed in this paper. The value $\bar{Z}_\alpha = 4.0309$ that results in a false alarm every 2 hours was corrected to $\hat{Z}_\alpha = 5.7235$.

The results in Table I indicate that the statistical model with the closest FAR to the desired rate is actually a Poisson distribution. However, justification for correlated sampling becomes apparent when the resulting thresholds are all placed on the same scale in the critical value column. The Poisson threshold is approximately six times that of the corrected normal rendering the Poisson model relatively insensitive.

IV. STATISTICAL VERIFICATION OF FALSE ALARM RATE USING EMPIRICAL DATA

In this section the statistical model of section II is applied to hundreds of background periods using a single data set referred to as the training data. A statistical inference is made during training concerning FAR performance and the training data is set aside. Verification of FAR utilizes a large number of individual false alarm tests [1][pp.17] each of duration 10 hours where the verification library consists of over 1000 hours

^aThe critical value was actually 2 counts but was scaled up to count rate per second for comparison purposes.

of background data per scintillating detector for both a ${}^6\text{Li}$ optical fiber detector and a Li^6 coated waveguide detector. Using the statistics estimated during training, confidence intervals were constructed at various coefficients and applied across all verification false alarm tests.

In this section several common experimental procedures are followed. Data is acquired with a sample period of 0.2 seconds, an integration period of 4 seconds, and a background period of 5 minutes. Again, the standard critical value is tuned to meet a target FAR of 0.5 alarms per hour.

A. Confidence Interval using Training Data

Each training data set is acquired such that the number of samples N_D results in about 70 hours of data to estimate a corrected \hat{Z}_α in tandem with the FAR performance statistics $\hat{\mu}_\alpha$ and $\hat{\sigma}_\alpha$ of section II-C5.

Instead of using the training statistic $\hat{\mu}_\alpha$ over N_{PB} background periods, the target statistic $\hat{\alpha}_j$ over a single background is used in verification to estimate a confidence interval that models an individual ANSI false alarm experiment [1][pp.17]. The standardized statistic for $\hat{\alpha}_j$ is $Z_j \triangleq \frac{\hat{\alpha}_j - \hat{\mu}_\alpha}{\hat{\sigma}_\alpha}$ where the corresponding standard critical value is $\bar{Z}_{\alpha/2}$, confidence coefficient is $(1 - \alpha)$ for $\alpha \in [0, 1]$, and probability of containment is

$$1 - \alpha = P\left(-\bar{Z}_{\alpha/2} < \frac{\hat{\alpha}_j - \hat{\mu}_\alpha}{\hat{\sigma}_\alpha} < \bar{Z}_{\alpha/2}\right) \quad (13)$$

Note that the critical value $\bar{Z}_{\alpha/2}$ can be looked up in a standard normal Z table accounting for a two-sided test while \bar{Z}_α in (2) uses a single-sided hypothesis test. A realization of a confidence interval [2][pp.216] of FAR for a specific training set is

$$(\hat{\alpha}^l, \hat{\alpha}^u) \triangleq (\hat{\mu}_\alpha - (\bar{Z}_{\alpha/2}) \hat{\sigma}_\alpha, \hat{\mu}_\alpha + (\bar{Z}_{\alpha/2}) \hat{\sigma}_\alpha) \quad (14)$$

and is tested for accuracy via the verification data set described next.

B. Containment of Verification Data

Verification data is segmented into background periods and processed like training data as in section II except that \hat{Z}_α , $\hat{\mu}_\alpha$, and $\hat{\sigma}_\alpha$ were already calculated via training data. Each verification data set is acquired such that the number of samples N_D results in exactly 10 hours of data. For a given verification data set and j-th background period, $\hat{\alpha}_j$ is calculated according to (5) using \hat{Z}_α from training while the selected verification data set is used to calculate both background statistics $\hat{\mu}_{B,j}$ and $\hat{\sigma}_j$. The value $\hat{\alpha}_j$ is tested to determine if contained within the interval in (14). All verification data sets are processed similarly for the j-th background period. In this manner a measured confidence coefficient is calculated as a simple decimal fraction of successful containment. In order to account for background variation, a confidence coefficient is calculated for every possible background period in addition to the j-th and the result is averaged. $\bar{\mu}_\alpha$ is the sample mean and $\bar{\sigma}_\alpha^2$ is the sample variance of $\hat{\alpha}_j$ over every background period and every verification data set.

TABLE II
COATED DETECTOR TRAINING DATA SETS

Hours	68.7	71.5	71.8	71.9	71.4	72.4	72.1	95.1
$\hat{\mu}_D$	0.24	0.22	0.23	0.22	0.23	0.22	0.22	0.24
$\hat{\sigma}_D$	0.25	0.24	0.24	0.24	0.24	0.24	0.24	0.25
\hat{Z}_α	5.70	5.74	5.83	5.77	5.75	5.68	5.83	5.67
$\hat{\mu}_\alpha$	0.5	0.49	0.49	0.51	0.51	0.5	0.5	0.5
$\hat{\sigma}_\alpha$	0.58	0.55	0.57	0.62	0.62	0.58	0.54	0.61
$\hat{\alpha}^l$	-0.63	-0.58	-0.62	-0.71	-0.70	-0.64	-0.56	-0.69
$\hat{\alpha}^u$	1.63	1.56	1.61	1.73	1.72	1.63	1.56	1.69

TABLE III
COATED DETECTOR VERIFICATION RESULTS FOR EACH TRAINING DATA SET

Ver. Sets	$\bar{\mu}_\alpha$	$\bar{\sigma}_\alpha$	85%	90%	92.5%	95%	97.5%	99%
128	0.52	0.63	87.6	89.8	91.8	94.0	95.1	96.3
127	0.48	0.60	86.5	89.2	90.2	92.3	94.8	95.8
127	0.40	0.54	92.6	93.6	94.9	95.6	96.4	97.9
127	0.46	0.58	91.8	93.5	94.6	95.5	97.5	98.6
127	0.48	0.61	88.8	92.2	93.0	94.6	95.9	98
127	0.53	0.66	86.7	89.4	91.2	92.5	94.1	96.3
127	0.42	0.56	89.5	91.2	92.8	94.2	96.2	96.8
125	0.54	0.66	86.6	88.4	91.2	92.3	95.8	97.6

TABLE IV
COATED DETECTOR VERIFICATION RESULTS OVER ALL TRAINING DATA SETS

Statistic	$\hat{\alpha}_j$	85%	90%	92.5%	95%	97.5%	99%
Sample Mean	0.48	88.8	90.9	92.5	93.9	95.7	97.2
Sample Std.	0.61	3.34	2.90	2.53	2.24	1.90	1.64

TABLE V
FIBER DETECTOR TRAINING DATA SETS

Hours	67.8	69.7	71.5	71.8	71.9	71.4	72.4	72.1
$\hat{\mu}_D$	2.28	2.30	2.26	2.28	2.27	2.26	2.25	2.24
$\hat{\sigma}_D$	0.81	0.82	0.81	0.81	0.81	0.81	0.80	0.80
\hat{Z}_α	4.66	4.80	4.71	4.59	4.66	4.59	4.63	4.58
$\hat{\mu}_\alpha$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
$\hat{\sigma}_\alpha$	0.35	0.30	0.31	0.41	0.36	0.44	0.41	0.39
$\hat{\alpha}^l$	-0.18	-0.09	-0.10	-0.30	-0.20	-0.36	-0.30	-0.27
$\hat{\alpha}^u$	1.18	1.10	1.10	1.29	1.21	1.36	1.30	1.26

TABLE VI
FIBER DETECTOR VERIFICATION RESULTS FOR EACH TRAINING DATA SET

Ver. Sets	$\bar{\mu}_\alpha$	$\bar{\sigma}_\alpha$	85%	90%	92.5%	95%	97.5%	99%
125	0.49	0.41	87.6	90.8	92.8	92.8	94.7	96.1
125	0.35	0.30	85.7	85.7	97.0	97.0	97.3	98.1
125	0.44	0.37	84.3	93.3	93.3	95.1	95.1	96.3
125	0.58	0.46	86.5	88.3	91.2	91.2	95.4	96.2
125	0.49	0.40	90.9	90.9	92.7	94.4	96.3	97.2
125	0.59	0.46	88.2	91.6	91.6	93.8	95.5	96.9
125	0.53	0.42	89.0	91.4	93.8	95.6	96.8	97.4
125	0.60	0.47	84.9	87.8	87.8	90.8	93.4	95.9

C. Confidence Testing Results

A total of 8 training files were used per detector with all results listed in each column of Tables II and V. $\hat{\mu}_D$ is the

TABLE VII
FIBER DETECTOR VERIFICATION RESULTS OVER ALL TRAINING DATA SETS

Statistic	$\hat{\alpha}_j$	85%	90%	92.5%	95%	97.5%	99%
Sample Mean	0.51	87.1	90.0	92.5	93.9	95.6	96.7
Sample Std.	0.42	3.43	3.33	3.23	2.81	2.04	1.63

sample mean of all samples measured in counts per second. $\hat{\sigma}_D$ is the sample standard deviation of all samples measured in counts per second. \hat{Z}_α is an estimate of the standard critical value that will achieve the desired false alarm rate measured in the number of standard deviations above the mean. $\hat{\mu}_\alpha$ is the estimated significance level of threshold \hat{Z}_α scaled to FAR measured in alarms per hour. $\hat{\sigma}_\alpha$ is the estimated standard deviation of $\hat{\mu}_\alpha$. $\hat{\alpha}^l$ and $\hat{\alpha}^u$ are the lower and upper bounds of FAR, respectively, for a 95% confidence interval.

The results of testing each training file with verification data is summarized in Tables III, IV, VI, and VII. Each training data set column number of Table II corresponds to the verification row number result listed in Table III. Each training data set column number of Table V corresponds to the verification row number result listed in Table VI. Measured confidence is in percent and is a function of not only $\hat{\alpha}^l$ and $\hat{\alpha}^u$, but also \hat{Z}_α .

V. CORRECTED SIGNAL-TO-NOISE RATIO

A. Corrected Signal-to-Noise Ratio Definition

1) *Signal to Noise Ratio Definition:* With a neutron detector setup with ANSI moderation [1], a sample mean of count rate measurements denoted μ is acquired using a background period of 5 minutes. Using a ^{252}Cf source with an emission rate of 20,000 n/s placed at 1.5m from the detector [1], a sample mean of count rate measurements notated ρ is acquired using a period of 5 minutes. It is assumed in [5][pp.94] that data follow a normal distribution where $\sigma^2 \approx \mu$.

SNR is then defined [5][pp.94] as

$$\text{SNR} \triangleq \frac{S^2}{B} = \frac{(\rho - \mu)^2}{\mu} \quad (15)$$

2) *Sigma Correction Factor:* When utilizing a neutron detector to alarm using a specific threshold, the threshold is set using a critical value that assumes that the detector will result in a specific FAR. Section III-C demonstrates how a theoretical standard critical value may result in a measured error rate that is drastically different than the rate that the model predicts.

When comparing two detectors a simple metric often used is SNR. The noise value encompassed by the denominator parameter σ represents a certain spread of data from the mean that occurs for a specific percent of samples. In other terms, this spread in data can be viewed as occurring at a specific rate for a real-time system. Let's define a correction factor γ^2 to adjust the measured variance σ^2 to account for the disparity between theoretical and measured rates. Using the alarm threshold developed in (3) with the additional correction

factor and notation of this section results in a new alarm threshold

$$T_C = \mu + (\bar{Z}_\alpha) \frac{\gamma\sigma}{\sqrt{N_I}} \quad (16)$$

Using the notation of this section and the practical standard critical value of section II-D, the practical threshold results in

$$T_P = \mu + (\hat{Z}_\alpha) \frac{\sigma}{\sqrt{N_I}} \quad (17)$$

To correct the theoretical critical value according to (16) this threshold should be equal to the practical or measured threshold in (17) resulting in

$$\mu + (\bar{Z}_\alpha) \frac{\gamma\sigma}{\sqrt{N_I}} = \mu + (\hat{Z}_\alpha) \frac{\sigma}{\sqrt{N_I}}$$

Therefore, if using the theoretical threshold \bar{Z}_α then σ could be corrected using the factor

$$\gamma = \frac{\hat{Z}_\alpha}{\bar{Z}_\alpha} \quad (18)$$

3) *Corrected Signal to Noise Ratio:* Using the sigma correction factor in (18) and the approximation $\sigma^2 \approx \mu$ [5][pp.94], SNR of section V-A1 is adjusted to

$$\text{SNR}_\gamma = \frac{(\rho - \mu)^2}{\gamma^2 \mu} \quad (19)$$

B. Corrected Signal-to-Noise Ratio Application

1) *Backpack Test Overview:* With respect to false alarms, the system should tolerate no more than five alarms due to natural background in a 10 hour period [1]. In addition, once the system is configured with an alarm threshold to meet the mentioned error rate, a source that is moving at a specific rate is required to be detected successfully in 96 out of 100 trials [1]. Note that an alarm threshold equal to or greater than the threshold to meet the specified FAR must be used in the moving source test to meet all system requirements.

In the subsequent sections several detector technologies are evaluated using the same parameters. The sampling periods described in section III were used so that the value $\bar{Z}_\alpha = 4.0309$ results in a false alarm every 2 hours.

2) *Fiber Detector Version 1 Test Results:* The first fiber detector tested labeled D1 required modification due to an inherent flaw. A standard critical value of $\hat{Z}_\alpha = 5.87$ resulted in a measured FAR of 0.5 alarms per hour. Using the corresponding threshold in a moving source test yielded a detection rate of 90%. Detector D1 measured a background mean of $\mu = 4.94$ according to section V-A1. When measuring the ^{252}Cf source a flaw with D1 occurred where the measurement slowly increased in time. Consequently, the initial value of count rate $\rho = 20$ was used instead of a 5 minute average. A sigma correction factor of $\gamma = 1.46$ resulted.

TABLE VIII
SIGNAL-TO-NOISE RATIO RESULTS

Detection System	Detection %	SNR	SNR $_{\gamma}$
Fiber D1	90	45.9	21.6
Fiber D2	95	22.8	17
Coated D3	97	43.8	22.6

3) *Fiber Detector Version 2 Test Results:* An improved detector tested was labeled D2. A standard critical value of $\hat{Z}_{\alpha} = 4.67$ resulted in a measured FAR of 0.5 alarms per hour. Using the corresponding threshold in a moving source test yielded a detection rate of 95%. Detector D2 measured a background mean of $\mu = 1.75$ according to section V-A1. When measuring the ^{252}Cf source the detector resulted in a count rate of $\rho = 8.07$ after a 5 minute average. A sigma correction factor of $\gamma = 1.16$ resulted.

4) *^6Li Coated Detector Test Results:* A third detector consisting of a ^6Li coated waveguide was labeled D3. A standard critical value of $\hat{Z}_{\alpha} = 5.61$ resulted in a measured FAR of 0.5 alarms per hour. Using the corresponding threshold in a moving source test yielded a detection rate of 97%. The detector measured a background mean of $\mu = 0.26$ according to section V-A1. When measuring the ^{252}Cf source the detector resulted in a count rate of $\rho = 3.63$ after a 5 minute average. A sigma correction factor of $\gamma = 1.39$ resulted.

C. Signal-to-Noise Ratio Test Results

The conventional method resulted in a SNR value for D1 about *twice* that of the value for D2 despite sensitivity testing [1] indicating D2 is marginally more sensitive than D1. Likewise, the conventional method resulted in a SNR value for D1 that was higher than the value for D3 despite sensitivity testing [1] indicating the contrary. The conventional method also indicated a SNR value for D3 that was substantially higher than D2 despite similar detection sensitivity.

Notice that in Table VIII in both cases the relative spread of SNR $_{\gamma}$ values more accurately reflected the corresponding spread in measured detection percent compared to the conventional SNR metric. However, the value of SNR $_{\gamma}$ for detector D1 remained higher than the value for detector D2 despite D1 proving less sensitive in a dynamic test.

VI. IMPLICATIONS OF SYSTEM PARAMETER MODIFICATION

Utilizing the method developed to process many background periods has additional advantages with respect to the analysis of practical system parameters. Parameters such as sample period, background duration, or even alarm methods discussed in section II-F could be adjusted in order to measure the effects on system response such as \hat{Z}_{α} in (10) or the observed FAR in (6).

For instance, the background period was altered from 1 minute to 5 minutes in 1 minute increments using the same detector, configuration, and data set as in section III-C. The

TABLE IX
BACKGROUND PERIOD (minutes) VS. FAR (per hour)

Period	\hat{Z}_{α}	Estimated FAR	Estimated FAR Std. Dev.
5	5.72	0.51	0.55
4	5.8	0.50	0.69
3	5.91	0.50	0.83
2	6.41	0.46	4.14
1	8.55	0.44	15.31

resulting maximum likelihood estimate of \hat{Z}_{α} , mean estimate $\hat{\mu}_{\alpha}$, and variance estimate $\hat{\sigma}_{\alpha}^2$ are listed in Table IX where significance levels are scaled to FAR per hour for convenience.

With FAR approximately constant, notice as background period decreases system sensitivity decreases via the estimated standard critical value \hat{Z}_{α} increase. Each estimated FAR data point remains approximately equal to the target rate but slightly lower when the background period falls below 3 minutes. However, note the drastic increase in the standard deviation of FAR when the background falls below 3 minutes.

VII. CONCLUSION

When utilizing all practical aspects of a radiation detection system to recognize neutron sources including the background period, integration period, correlated sampling, and the sparse nature of acquired data, the measured noise performs at a FAR significantly different than the FAR that the theoretical model predicts. This practical error performance of a neutron detection system must be normalized to some known reference in order to properly characterize practical detection sensitivity.

Modifications to detection algorithms, manufacturing process, or physical composition of neutron detectors could result in a deviation in the necessary \hat{Z}_{α} to meet a required FAR. A method was introduced to adjust theoretical alarm parameters to more accurately quantify and meet a desired FAR using a static measurement and an automated algorithm.

Performing an additional static measurement with a source while applying the correction factor γ results in a SNR metric that better agrees with more involved dynamic detection tests.

In conclusion, the estimate of standard critical value in (10) and the sigma correction factor in (18) should be considered as methods to characterize how the measured error rate and detection sensitivity are affected by differences in neutron detection systems.

REFERENCES

- [1] "American national standard performance criteria for backpack-based radiation-detection systems used for homeland security," *ANSI N42.53-2013*, pp. 1–62, Aug 2013.
- [2] R. Hogg, J. McKean, and A. Craig, *Introduction to Mathematical Statistics*, ser. Always learning. Pearson, 2013. [Online]. Available: <https://books.google.com/books?id=xFIVYAAACAAJ>
- [3] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, ser. Communications and signal processing. McGraw-Hill, 1991. [Online]. Available: <https://books.google.com/books?id=4IwQAQAIAAJ>
- [4] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *SCIENCE*, vol. 220, no. 4598, pp. 671–680, 1983.
- [5] G. F. Knoll, *Radiation Detection and Measurement; 3rd ed.* New York, NY: Wiley, 2000. [Online]. Available: <https://cds.cern.ch/record/441925>