Statistical Methods for Estimating Sources of Variation in the Cygnus Radiographic Imaging System

K. Joyce, J. Bardsley

Department of Mathematical Sciences

University of Montana, Missoula MT, 59812-0864 USA

A. Luttman, M. Howard

National Securities Technologies, 2621 Lossee Rd

North Las Vegas, NV 89030 USA

E. Ormond

Sandia National Laboratory, PO Box 238, Mail Stop 944

Mercury, NV 89023 USA

December 8, 2013

Abstract

In this paper, we develop a statistical model for estimating the mean and standard deviation of the dose at one meter (DAOM) diagnostic for the Cygnus Radiographic imaging system. These statistics are estimated from an extensive record of thermoluminescent dosimetry (TLD) measurements made over a large portion of the shot history of the machine. The methods presented decouple TLD measurement error from signal variation and account for x-ray attenuation due to a stacked TLD measurement configuration.

1 Introduction

An extensive data record of thermoluminescent dosimeter (TLD) readings has been kept as a shot performance diagnostic for the Cygnus radiographic imaging system since August 2009. From the reported dose measurements, an estimate of the radiation dose at one meter (DAOM) from the imaged package is extrapolated based on a model of inverse square radiation decay. The data, however, present certain challenges that make a rigorous statistical analysis somewhat non-standard due to the presence of significant measurement error and non-identical measurement replication.

The primary objective of this paper is to quantify the variation of the DAOM of each Cygnus radiation source. Obtaining a statistical estimate for this requires modeling the uncertainty of TLD measurement and how it interacts with the variation in the DOAM. In both cases, the statistic of interest is the relative standard deviation (RSD) which is the population standard deviation divided by its mean.

2 TLD Data

Dose data were taken from three batches of 100 high precision Harshaw TLD-100 ribbons ($0.125'' \times 0.125'' \times 0.035''$ thick) with corresponding labels A,B, and C manufactured by Thermo Scientific. Within each batch, each TLD ribbon is numbered 1-100. Readings are taken from TLD ribbons one batch at a time, sequentially within each batch. For each Cygnus shot, the primary dose reading is obtained from three TLD ribbons that are aligned in a stacked configuration in one inch thick aluminum shielding and placed outside of the bulkhead equidistant each of the Cygnus radiation sources.

After x-ray exposure from one of the radiation sources, the chip is exposed to a controlled and calibrated heat source. This causes visible light to be emitted dependent upon the amount of ionizing radiation absorbed by the LiF crystal. Each TLD is calibrated yearly by exposure to a NIST traceable Cesium-137 source from which a linear coefficient is derived for calibrating the visible light reading to implied dose. After each measurement and calibration, the TLD is annealed and cooled. This process is reported by the manufacturer to give the dose measurements with less than 2 % standard deviation based on 10 controlled sequential measurements.

The TLD chips are stacked facing the radiation source so that the absorbed ionizing radiation is attenuated by consecutive chips. That is, chips stacked behind other chips absorb less radiation and under-report dose. See Figure 1. Hence, a model describing statistics about each Cygnus machine such as average DAOM or percent standard deviation must include must account for the effect of attenuation in the TLD measurements.

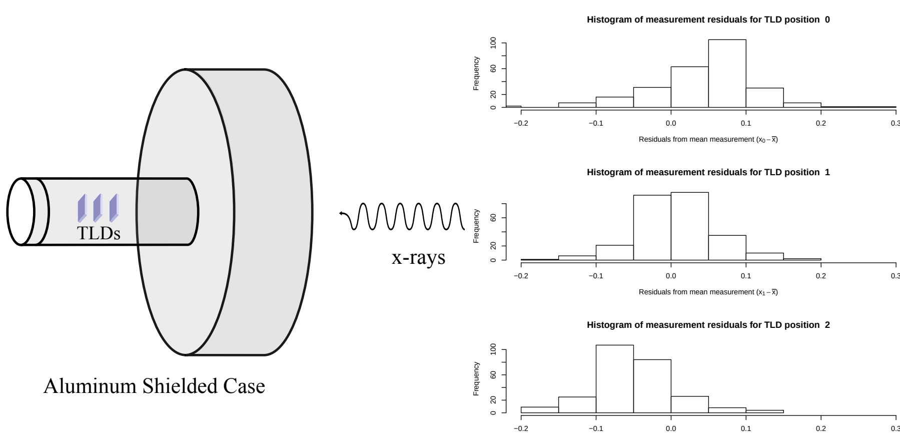

Figure 1: Three-sample-residual histograms of each TLD position over the relevant Cygnus shots provide evidence of an effect of attenuation.

3 Statistical Model

We selected $n = 100$ shots from from each Cygnus machine in the recent TLD record history. The shots were choosen so that each machine had a consistent configuration shot to shot. For this analysis, the TLD data are indexed over three categories - the shot sample index $i = 1, \ldots, n$ (note that this differs from the shot record index); the position of the TLD, $j = 0, 1, 2$; and the Cygnus radiation source $k = 1, 2$. Each calibrated measurement is denoted $\widetilde{y}_{ijk}$. Hence there are $6n$ records for this analysis.

To account for the TLD attenuation, we model the TLD measurements as proportional to the integrated energy intensity of the x-rays absorbed by the TLD. We assume this follows exponential attenuation. Additionally, the model simultaneously accounts for variation of the DAOM due to machine shot-to-shot variance and the random measurement error of the TLDs. The form of the model is

$$\widetilde{y}_{ijk} = \exp\left(\beta_{0k} + \beta_1 x_{i1} + \beta_2 x_{i2} + \eta_{ik} + \epsilon_{ijk}\right). \tag{1}$$

where the fixed coefficient β_{0k} represents the unattenuated mean log DOAM for Cygnus machine k; β_1 and β_2 are coefficients of attenuation through the first and second TLD chips;

$$x_{i1} = \begin{cases} 1 & \text{if measured from the second TLD,} \\ 0 & \text{otherwise,} \end{cases}$$

and

$$x_{i2} = \begin{cases} 1 & \text{if measured from the third TLD,} \\ 0 & \text{otherwise;} \end{cases}$$

$\epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$ is the random error due to TLD measurement; and $\eta_{ik} \sim N(0, \sigma_{\eta k}^2)$ is the random variation in each machine source. We assume that within each fixed machine factor k, the η_{ik} are identically distributed but that $\sigma_{\eta 1}$ may not necessarily $\sigma_{\eta 2}$. Note that we also assume that the attenuation through the TLD chips does not depend on the radiation source index k, and that η_{ik} does not depend on the shot TLD placement index j.

A log transformation of (1) results in a linear model in the parameters with two random components. Thus, we can write the model in the matrix-vector form

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{P\eta} + \boldsymbol{\epsilon}, \tag{2}$$

where $\boldsymbol{y}$ as the $(6n) \times 1$ vector of log transformed data points $\log \widetilde{y}_{ijk}$ (indexing first by i, then j, then k); $\boldsymbol{X}$ is a $6n \times 4$ matrix

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{1_{3n\times 1}} & & \uparrow & \uparrow \\ & & \boldsymbol{x_1} & \boldsymbol{x_2} \\ & \boldsymbol{1_{3n\times 1}} & \downarrow & \downarrow \end{bmatrix};$$

where $\boldsymbol{x}_j$ are vectors whose entires are the values of the indicator variables x_{ij}; $\boldsymbol{\beta} = (\beta_{01}, \beta_{02}, \beta_1, \beta_2)$; $\boldsymbol{P}$ is a $6n \times 2n$ block matrix with 3×1 blocks of ones along the diagonal; and both $\boldsymbol{\eta}$ and $\boldsymbol{\epsilon}$ are the vectors whose coefficents are the random variables described above.

If we consider both random pieces as one random vector $\boldsymbol{\xi} = \boldsymbol{\epsilon} + \boldsymbol{P\eta}$, then by properties of sums of independent normal random vectors [3], the model simplifies further to

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\xi} \tag{3}$$

where $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ is a $6n \times 6n$ block matrix

$$\boldsymbol{\Sigma} = \left[\begin{array}{cc|cc} \boldsymbol{B_1} & & & \\ & \ddots & & \\ \hline & & \boldsymbol{B_2} & \\ & & & \ddots \end{array}\right],$$

where the blocks are of the form

$$\boldsymbol{B}_k = \begin{bmatrix} \sigma^2_{\eta k} + \sigma^2_{\epsilon} & \sigma^2_{\eta k} & \sigma^2_{\eta k} \\ \sigma^2_{\eta k} & \sigma^2_{\eta k} + \sigma^2_{\epsilon} & \sigma^2_{\eta k} \\ \sigma^2_{\eta k} & \sigma^2_{\eta k} & \sigma^2_{\eta k} + \sigma^2_{\epsilon} \end{bmatrix}. \tag{4}$$

Here, note that there are two perspectives to view the log DAOM attenuation model. From (2), we view $\boldsymbol{y}$ as the attenuation corrected measurements from each machine source with a random effect, $\boldsymbol{P\eta}$, from the variation of each the machine shot-to-shot. From (3), we view $\boldsymbol{y}$ as *non*-independent normally distributed data where the measurements grouped by shot are correlated with $\mathrm{cov}(y_{i0k}, y_{i1k}) = \mathrm{cov}(y_{i0k}, y_{i2k}) = \mathrm{cov}(y_{i1k}, y_{i2k}) = \sigma^2_{\eta k}$. In the following section, we exploit both viewpoints to obtain estimates for $\beta_{01}, \beta_{02}, \beta_1, \beta_2, \sigma^2_{\epsilon}, \sigma^2_{\eta 1}$ and $\sigma^2_{\eta 2}$.

4 Parameter Estimation

When a statistical linear model has a non-standard error structure as in (3), the standard technique for analysis and parameter estimation is generalized least squares (GLS) [1] [3].

The GLS solution of (3) is

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{y}. \tag{5}$$

The estimate $\widehat{\boldsymbol{\beta}}$ has the same optimal characteristics as those in standard least squares [3].

In our case, $\boldsymbol{\Sigma}$ involves the unknown parameters $\sigma^2_{\eta 1}, \sigma^2_{\eta 2}$ and σ^2_{ϵ}, so we cannot apply (5) directly. We present a simple iterative algorithm that exploits (2) to intermediately estimate $\boldsymbol{\xi}$ to estimate the parameters of $\boldsymbol{\Sigma}$.

Estimating Model Parameters

---

1. Initially estimate $\widehat{\boldsymbol{\beta}}_{\mathbf{0}} = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{y}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{0}} = \mathrm{Var}^* \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_0\right) \boldsymbol{I}$. Set $k = 0$.

2. Estimate $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{k}}$.

   (i) $\widehat{\boldsymbol{\xi}}_k = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{k-1}$.

   (ii) $\widehat{\boldsymbol{\eta}}_k = (\boldsymbol{P}^T \boldsymbol{P})^{-1} \boldsymbol{P}^T \widehat{\boldsymbol{\xi}}_k$.

   (iii) $\widehat{\sigma}_{\epsilon} = \mathrm{Var}^* \left(\widehat{\boldsymbol{\xi}}_k - \boldsymbol{P}\widehat{\boldsymbol{\eta}}_k\right)$,

   $\widehat{\sigma}_{\eta\, 1} = \mathrm{Var}^* \left([\widehat{\boldsymbol{\eta}}_k]^n_{j=1}\right)$,

   $\widehat{\sigma}_{\eta\, 2} = \mathrm{Var}^* \left([\widehat{\boldsymbol{\eta}}_k]^{2n}_{j=n+1}\right)$.

(iv) Update $\widehat{\boldsymbol{\Sigma}}_k$ via (4).

3. Estimate $\widehat{\boldsymbol{\beta}}_{\boldsymbol{k}}$.

   (i) $\widehat{\boldsymbol{\beta}}_k = \left(\boldsymbol{X}^T \widehat{\boldsymbol{\Sigma}}_k^{-1} \boldsymbol{X}\right)^{-1} \widehat{\boldsymbol{\Sigma}}_k^{-1} \boldsymbol{X}^T \boldsymbol{y}$.

4. If convergence in $\widehat{\boldsymbol{\xi}}_k^T \widehat{\boldsymbol{\xi}}_k$ is reached, stop, else set $k = k + 1$ and continue from step 2.

In the above algorithm, Var* denotes the unbiased variance estimator given by dividing by the appropriate degrees of freedom in each estimation. Note that estimation of $\boldsymbol{\eta}$ is given by the regular least squares solution of $\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{P}\boldsymbol{\eta} + \boldsymbol{\epsilon}$.

The results are given in the following summary.

5 Summary

The results of the analysis are outlined in Table 1.

TLD measurements	
Estimate of % STD DEV	%
Estimate of Atten. Coef. 1	
Estimate of Atten. Coef. 2	

Cygnus - 1 DAOM		
Estimate of Mean	rad	95% CI: ( ) rad
Estimate of % STD DEV	%	95% CI: ( ) %

Cygnus - 2 DAOM		
Estimate of Mean	rad	95% CI: ( ) rad
Estimate of % STD DEV	%	95% CI: ( ) %

Table 1: Model estimates.

References

[1] J. Faraway (2005) *Linear Models with R.* Chapter 6, pages 90-92.

[2] A. Mood, F. Graybill, D. Boes, (1974) *Introduction to the Theory of Statistics* (3rd Edition). Chapters IV-V, pages 189-202, 239-249.

[3] A. C. Rencher, G. B. Schaalje, (2008) *Linear Models in Statistics* (2nd Edition). Chapter 12, pages 295-339.

[4] G. Corrow, et.al. Cygnus Performance in Subcritical Experiments. No. DOE/NV/25946–294. National Security Technologies, LLC, 2008.