

Article

## Defense of the Least Squares Solution to Peelle's Pertinent Puzzle

Tom Burr <sup>1,\*</sup>, Toshihiko Kawano <sup>2</sup>, Patrick Talou <sup>2</sup>, Feng Pan <sup>3</sup> and Nicolas Hengartner <sup>4</sup>

<sup>1</sup> Statistical Sciences, Los Alamos National Laboratory, Los Alamos NM, USA; E-Mail: tburr@lanl.gov

<sup>2</sup> Nuclear and Particle Physics, Los Alamos National Laboratory, Los Alamos NM, USA;  
E-Mail: talou@lanl.gov

<sup>3</sup> Decision Applications, Los Alamos National Laboratory, Los Alamos NM, USA;  
E-Mail: fpan@lanl.gov

<sup>4</sup> Information Sciences, Los Alamos National Laboratory, Los Alamos NM, USA;  
E-Mail: nickh@lanl.gov

\* Author to whom correspondence should be addressed; E-Mail: kawano@lanl.gov;  
Tel.: +1-505-664-0513; Fax: +1-505-667-1931.

Received: 6 December 2010; in revised form: 15 January 2011 / Accepted: 7 February 2011 /  
Published: 15 February 2011

---

**Abstract:** Generalized least squares (GLS) for model parameter estimation has a long and successful history dating to its development by Gauss in 1795. Alternatives can outperform GLS in some settings, and alternatives to GLS are sometimes sought when GLS exhibits curious behavior, such as in Peelle's Pertinent Puzzle (PPP). PPP was described in 1987 in the context of estimating fundamental parameters that arise in nuclear interaction experiments. In PPP, GLS estimates fell outside the range of the data, eliciting concerns that GLS was somehow flawed. These concerns have led to suggested alternatives to GLS estimators. This paper defends GLS in the PPP context, investigates when PPP can occur, illustrates when PPP can be beneficial for parameter estimation, reviews optimality properties of GLS estimators, and gives an example in which PPP does occur.

**Keywords:** Peelle's puzzle; mean squared error; measurement error modeling

---

## 1. Introduction

Generalized least squares (GLS) for parameter estimation has a long and successful history dating to its development by Gauss in 1795. In some settings, alternatives to GLS can be effective, and are sometimes sought when GLS exhibits curious behavior, such as in Peelle's Pertinent Puzzle (PPP).

PPP was introduced in 1987 in the context of estimating fundamental parameters that arise in nuclear interaction experiments [1]. PPP is described below and when it occurs, the GLS estimate of the parameter is guaranteed to be outside the range of the data, which has elicited concerns that GLS is flawed and has led to suggested alternatives to GLS estimators.

A GLS estimate lying outside the range of the data causes heartache among nuclear scientists. Therefore, PPP continues to be of theoretical and practical interest. For example, a summary report of International Evaluation of Neutron Cross Section Standards [2] discusses PPP in terms of a standard least-squares procedure. Neutron cross sections are fundamental parameters that describe the probabilities of various neutron interactions. These cross sections are typically estimated from multiple experiments so some type of weighted average estimation scheme is used. The cross section estimates from any two experiments can have shared errors arising for example from using the same measured background, which can lead to covariance structures such as those described below.

We quote the original PPP problem proposed by [1] from a report of Chiba and Smith [3], "Suppose we are required to obtain the weighted average of two experimental results for the same physical quantity. The first result is 1.5, and the second result 1.0. The full covariance matrix of those data is believed to be the sum of three components. The first component is fully correlated with standard error 20% of each respective value. The second and third components are independent of the first and of each other, and correspond to 10% random uncertainties in each experimental result.

Although this PPP statement is vague, by converting it to something more interpretable, GLS can be applied and the resulting estimate is 0.88 (with an associated standard deviation of 0.22), which is outside the range of the measurements. Zhao and Perey [4] re-interpreted PPP by introducing a third datum  $c$ , through which the common error can be explicitly specified as follows, "Suppose we have two independent measurements. One is  $m_1 = 1.5 \pm 10\%$ . Another is  $m_2 = 1.0 \pm 10\%$ . To convert this quantity into another physical quantity, we need a conversion factor  $c$ , which after intermediate steps omitted here is 1.0 with uncertainty of 20%. Now the experimental results are,  $y_1 = cm_1 = 1.5$  and  $y_2 = cm_2 = 1.0$ . We are required to obtain the weighted average of those experimental data.

In this interpretation, the common error (the "fully correlated" component) is understood to be *multiplicative*, and  $m_1 = 1.5 \pm 10\%$  is assumed to mean that the *true* standard deviation is 0.15 for  $m_1$  (and 0.10 for  $m_2$ ). Even after these interpretations, some vagueness remains. There is no convention regarding what confidence is associated with  $\pm 10\%$ . Nor is there a convention for whether the standard deviation includes all error sources, or only includes random error effects, ignoring accuracy. In addition, we show below that it can matter whether the standard deviation is expressed as a fraction of the true quantity or of the measured quantity.

One of our contributions is to make explicit assumptions and examine their implications in order to convert vague statements to statements about which it is possible to find agreement among physical scientists and statisticians regarding suitable approaches. We also defend GLS in the PPP context,

illustrate when PPP can be beneficial, briefly describe properties of GLS estimators, show that PPP cannot occur for certain measurement error models, and calculate a covariance matrix  $\Sigma$  for  $y_1$  and  $y_2$  for which PPP occurs that follows from a physical description of a realistic measurement scenario.

## 2. PPP

The vagueness of the original PPP statement is one reason there are so many interpretations of PPP [5]. PPP can occur if there is a large positive covariance between  $y_1$  and  $y_2$  and the variances of  $y_1$  and  $y_2$  are very different.

Let  $\Sigma$  be the 2-by-2 symmetric covariance matrix for  $y_1$  and  $y_2$  with diagonal entries  $\sigma_1^2$ ,  $\sigma_2^2$ , and off-diagonal entry  $\sigma_{12}$ , which denote the variance of  $y_1$ , the variance of  $y_2$ , and the covariance of  $y_1$  and  $y_2$ , respectively. Zhao and Perry [4] approximated  $\Sigma$  for their definition of PPP (using  $y_1 = cm_1$  and  $y_2 = cm_2$  as described in the Introduction) as

$$\begin{aligned} \Sigma &= \text{Cov}(y_1, y_2) \\ &= \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \\ &\approx \mu_c^2 \begin{pmatrix} \sigma_{m1}^2 & \\ & \sigma_{m2}^2 \end{pmatrix} + \sigma_c^2 \begin{pmatrix} \mu_m^2 & \mu_m^2 \\ \mu_m^2 & \mu_m^2 \end{pmatrix} \\ &\approx \begin{pmatrix} 0.1125 & 0.06 \\ 0.06 & 0.05 \end{pmatrix} \end{aligned} \tag{1}$$

where  $\sigma_{m1}$ ,  $\sigma_{m2}$ , and  $\sigma_c$  are the standard deviations in the quantities  $m_1$ ,  $m_2$ , and  $c$ , respectively. The first approximation arises because the relatively small term  $\sigma_c^2 \sigma_m^2$  was omitted (see below). The second approximation follows by approximating  $\mu_m^2$  by  $m_1 m_2$ . Equation 1 assumes that  $m$  and  $c$  are independent. And, the relative standard deviations of 10%, 10%, and 20% are assumed to be the fractions relative to the measured values of 1.5, 1.0, and 1.0, respectively by definition, so that  $\sigma_{m1} = 0.15$ ,  $\sigma_{m2} = 0.10$ , and  $\sigma_c = 0.2$ . See Section 4 for further discussion.

Readers might find it informative that those with traditional statistical education among the authors were the most willing to accept GLS estimates despite the apparent flaw of lying outside the range of the data. Statisticians will often consider alternatives to GLS, but recognize that GLS estimation is difficult to beat, at least in terms of typical performance measures such as being close on average to the true parameter value over hypothetical repeats of the pair of experiments [6]. Also, note that because  $\frac{y_1 - y_2}{\sigma_{y_1 - y_2}} = 0.5/0.21 = 2.39$ , one might consider the original PPP to be an unusual data realization. However, we show using elementary algebra and error modeling in Theorem 2 that when PPP occurs, it occurs for *all* data realizations.

Although Peelle [1] originally constructed the covariance matrix  $\Sigma$  as in Equation 1 and other authors followed, this result is not exact if the common error is multiplicative because of the omitted term  $\sigma_c^2 \sigma_m^2$  and because of the need to estimate  $\mu_m$ . If we include the  $\sigma_c^2 \sigma_m^2$  term, then

$$\text{Var}(cm) = \mu_c^2 \sigma_m^2 + \mu_m^2 \sigma_c^2 + \sigma_c^2 \sigma_m^2 \tag{2}$$

where  $\sigma_m$  is  $\sigma_{m1}$  or  $\sigma_{m2}$  in Equation 1. The covariance matrix is therefore estimated as

$$\begin{aligned} \Sigma &= \mu_c^2 \begin{pmatrix} \sigma_{m1}^2 & \\ & \sigma_{m2}^2 \end{pmatrix} + \sigma_c^2 \begin{pmatrix} \mu_m^2 + \sigma_{m1}^2 & \mu_m^2 \\ \mu_m^2 & \mu_m^2 + \sigma_{m2}^2 \end{pmatrix} \\ &\approx \begin{pmatrix} 0.1134 & 0.06 \\ 0.06 & 0.0504 \end{pmatrix} \end{aligned} \tag{3}$$

where again as in Equation 1,  $\mu_m$  is estimated by  $m_1$  or  $m_2$ .

As shown below, prior to substituting the approximation for  $\mu_m^2$ , Equations 1 and 3 satisfy conditions under which PPP cannot occur (Theorem 2 in Section 3). However, because in all published investigations we are aware of,  $\mu_m^2$  is approximated by  $m_1 m_2$  in the off-diagonal and  $\mu_m^2$  is approximated by  $m_1^2$  in the upper left entry in the second term and by  $m_2^2$  in the lower right entry in the second term,  $\Sigma$  as estimated does not have the condition referred to in Theorem 2 below. Therefore, due to estimation error in  $\Sigma$ , PPP does appear to occur, which means that the GLS estimate of  $\mu$  is outside the range of the data for  $\Sigma$  as given by Equation 1 or Equation 3.

The fact that  $\mu_m$  is approximated in the context of estimating variance and covariance in a multiplicative error model raises at least three issues: (1) the issue of how uncertainty in measurement is expressed; (2) an issue related to simulating observations from the assumed measurement error model as a way to consider likelihoods other than the Gaussian, and (3) accounting for uncertainty in  $\Sigma$ . For issue (1), we will make our measurement error model assumptions explicit throughout in order to eliminate needless ambiguities. Issues (2) and (3) are investigated in [6].

To focus on GLS behavior when PPP occurs, *this paper assumes  $\Sigma$  is known exactly without error*. However, for historical and presentation purposes, Equations 1 and 3 are presented, and both clearly involve approximations. In contrast, Theorem 3 illustrates a measurement error model for which the exact covariance can satisfy the PPP condition.

### 3. GLS

It is well known that the GLS method can be applied to  $y_1$  and  $y_2$  to obtain the best linear unbiased estimate (BLUE)  $\hat{\mu}$  of  $\mu$  [7]. Here, “best” means minimum variance and unbiased means that on average (across hypothetical or real realizations of the same experiment), the estimate  $\hat{\mu}$  will equal its true value  $\mu$ . The GLS estimate for  $\mu$  arising from the model

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \mu + e_1 \\ \mu + e_2 \end{pmatrix} \tag{4}$$

with  $\Sigma = \text{Cov}(e_1, e_2) = \text{Cov}(y_1, y_2)$  given by

$$\hat{\mu} = c G^t \Sigma^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \tag{5}$$

where the scalar  $c = (G^t \Sigma^{-1} G)^{-1}$ . And the variance  $\sigma^2$  of the GLS estimate  $\hat{\mu}$  is given by

$$\sigma^2 = (G^t \Sigma^{-1} G)^{-1} \tag{6}$$

where the matrix  $G$  is given by  $G^t = (1, 1)$ . Putting the covariance of Equation 1 into Equations 5 and 6 gives  $\hat{\mu} = c_1 y_1 + c_2 y_2 = 0.88$  and  $\sigma = 0.22$  where  $c_1 = -0.24$  and  $c_2 = 1.24$ . Notice that  $c_1 + c_2 = 1$  (so that  $\hat{\mu}$  is unbiased) but also that  $c_1 < 0$  and that  $\hat{\mu}$  is smaller than each of the two measured values,  $y_1 = 1.5$  and  $y_2 = 1.0$ .

GLS is usually introduced in the context of estimating  $\beta$  and future  $y$  values in a linear regression relating the response  $y$  to predictors  $X$  via  $y = X\beta + \epsilon$  ([7]). Therefore, in Equation 4, the mean  $\mu$  plays the role of the unknown  $\beta$ . The GLS solution in Equations 5 and 6 then follows from standard calculus or projection matrix results. For example, one can write  $\hat{\mu} = a_1 y_1 + (1 - a_1) y_2$ , note that  $\text{var}(\hat{\mu}) = a_1^2 \sigma_{y_1}^2 + (1 - a_1)^2 \sigma_{y_2}^2 + 2a_1(1 - a_1) \sigma_{y_1, y_2}^2$  and solve for  $a_1$  to minimize  $\text{var}(\hat{\mu})$  by setting the derivative of  $\text{var}(\hat{\mu})$  with respect to  $a_1$  to 0.

The GLS solution of Equations 5 and 6 with covariance  $\Sigma$  of Equation 3 is  $0.89 \pm 0.22$ . Because the last term in Equation 2 is smaller than the others, the  $\Sigma$  in Equation 3 that is slightly modified compared to the  $\Sigma$  in Equation 1 still leads to PPP. But in Section 4 we describe other modifications to  $\Sigma$  that do not lead to PPP.

GLS estimation has a long and successful history, but met with serious objection within the nuclear physics community in the context of combining estimates from multiple experiments upon observing a tendency to produce estimates that are outside the range of the data. More specifically, to date the tendency has been to produce estimates that are less than the minimum data value, so have been criticized as being “too small” [2].

GLS estimation is guaranteed to produce the BLUE even if the underlying data are not Gaussian. However, if the data is not Gaussian, then the minimum variance unbiased estimator (MVUE) is not necessarily linear in the data. Also, though unbiased estimation might sound politically correct, it is not necessarily superior to biased estimation [8]. Therefore, PPP has motivated the nuclear physics community to consider estimators other than GLS.

If the data has a Gaussian distribution, then it is well known that the GLS estimate is the same as the maximum likelihood (ML) estimate. This is because the log of the Gaussian likelihood involves a sum of squares, so choosing an estimate (the GLS estimate) that minimizes a sum of squares corresponds to choosing an estimate (the ML estimate) that maximizes the likelihood. Ordinary LS (OLS), weighted LS (WLS), and GLS are all essentially the same technique, but OLS is used if  $\Sigma$  is proportional to a unit-diagonal matrix, WLS is used if  $\Sigma$  is proportional to a diagonal matrix, and GLS is used if  $\Sigma$  is an arbitrary positive definite covariance matrix. The Gauss-Markov theorem [7] proves that the OLS estimator is the BLUE, and very similar theorems prove the same result for WLS or GLS.

The ML estimate depends on the assumed distributions for the errors. For example, if we replace the Gaussian (Normal) distributions with logNormal distributions, the ML estimate will change. In the cases considered here, ML gives the same estimate as GLS, because the data distribution is Gaussian. Because the ML approach makes strong use of the assumed error distributions, the ML estimate is sensitive to the assumed error distribution. The ML method has desirable properties, including asymptotically minimum variance as the sample size increases. However, in our example, the sample size is tiny (two), so asymptotic results for ML estimates are not relevant. It still is possible that an ML estimator will be better for nonGaussian data than GLS [6]. Typically, “better” is defined as the mean squared error (MSE) of the estimator, which is well known to satisfy  $\text{MSE} = \text{variance} + \text{bias}^2$ . In some cases, biased

estimators have lower MSE than unbiased estimators because the bias introduced is more than offset by a reduction in variance [8].

#### 4. Closer Look into PPP

The original PPP does not clearly state whether the common error is additive or multiplicative. This ambiguity was examined in [5]. In the “additive” scenario case,  $y_1 = m_1 + b$  and  $y_2 = m_2 + b$ , and the source of correlation may be a common background measurement  $b$ . And, the situation is somewhat different from PPP, because if  $\sigma_b$  is 20% of  $y_1 = 1.5$ , then it is 30% of  $y_2 = 1.0$ . The covariance matrix for this case is

$$\begin{aligned} \Sigma &= \begin{pmatrix} \sigma_{m1}^2 & \\ & \sigma_{m2}^2 \end{pmatrix} + \sigma_b^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0.1125 & 0.09 \\ 0.09 & 0.1 \end{pmatrix} \end{aligned} \tag{7}$$

and the GLS solution is  $1.15 \pm 0.31$ . With the covariance matrix of Equation 7, PPP does not occur, and the GLS solution of 1.15 is a weighted average of  $1.0 \pm 0.1$  and  $1.5 \pm 0.15$ . Although as  $\sigma_b$  changes, the standard deviation  $\sigma_{\hat{\mu}}$  is scaled accordingly, the GLS solution does not change. It is reasonable that the GLS solution does not change as  $\sigma_b$  changes, because whatever the background fluctuation is,  $y_1$  and  $y_2$  are impacted in the same manner by  $b$ .

If instead  $\sigma_b$  is 20% of  $\mu_m$ , then its value is unknown because  $\mu_m$  is unknown and must be estimated. However, regardless of the value of  $\sigma_b > 0$ , PPP cannot occur in the additive case as parameterized in Equation 7 (Theorem 1 below).

In Sivia’s [9] notation, Equation 5 can be written as

$$\hat{\mu} = \frac{a_1 y_1 + a_2 y_2}{a_1 + a_2} \tag{8}$$

with

$$a_1 = \sigma_2(\sigma_2 - \rho\sigma_1), \quad a_2 = \sigma_1(\sigma_1 - \rho\sigma_2) \tag{9}$$

where  $\rho$  is the correlation coefficient. Sivia demonstrated that PPP does not occur if the condition

$$\rho \leq \min \left( \frac{\sigma_1}{\sigma_2}, \frac{\sigma_2}{\sigma_1} \right) \tag{10}$$

is fulfilled (and does occur if the condition does not hold). It is not difficult to show that the additive case of Equation 7 satisfies this condition, which we state as Theorem 1.

**Theorem 1.** If  $\Sigma$  is given by

$$\Sigma = \begin{pmatrix} \sigma_{m1}^2 & \\ & \sigma_{m2}^2 \end{pmatrix} + \sigma_b^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \tag{11}$$

then PPP does not occur.

Proof. The covariance matrix  $\Sigma$  can be expressed as

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{12}$$

with  $\rho = \frac{\sigma_b^2}{(\sigma_1^2 + \sigma_b^2)^{1/2}(\sigma_2^2 + \sigma_b^2)^{1/2}}$ . If  $\sigma_1^2 < \sigma_2^2$  then  $\rho \leq \min\left(\frac{\sigma_1}{\sigma_2}, \frac{\sigma_2}{\sigma_1}\right)$  is equivalent to  $\sigma_b^4 \leq (\sigma_1^2 + \sigma_b^2) \times (\sigma_1^2 + \sigma_b^2)$  which obviously holds. The proof is identical if  $\sigma_1^2 > \sigma_2^2$ . And if  $\sigma_1^2 = \sigma_2^2$ , then PPP can never occur.

Notice that Equations 1 and 3 have the same form as assumed in Theorem 1. Therefore, PPP also cannot occur for the multiplicative error model assumed. However, as discussed, PPP can occur with the multiplicative error model case if  $\Sigma$  is estimated as in Equations 1 and 3.

Several authors explored different values for the covariance matrix  $\Sigma$  to understand the relationship between the covariance matrix and the estimates. Some numerical examples are in [2]. Jones *et al.* [10] and Finn *et al.* [11] also reported linear regression of strongly correlated data and emphasized that if  $\sigma_1^2 \neq \sigma_2^2$  then contrary to common belief, positive or negative correlation can be exploited to produce an unbiased GLS estimate that has lower variance than in the zero correlation case. The historical definition of PPP requires the correlation to be positive because only in that case will the GLS estimate lie outside the range of the data for certain  $\Sigma$ .

Next rewrite Equation 8 as  $\hat{\mu} = a'_1 y_1 + a'_2 y_2$  with  $a'_1 = \frac{a_1}{a_1 + a_2}$  and  $a'_2 = \frac{a_2}{a_1 + a_2}$ . A condition equivalent to Equation 10 is obtained as follows. By setting the first derivative with respect to  $a'_1$  to zero, one can solve for the value of  $a'_1$  such that the variance of  $a'_1 y_1 + a'_2 y_2$  is minimum subject to  $a'_1 + a'_2 = 1$ . The result is  $a'_1 = \frac{\sigma_2^2 - \sigma_{12}^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}^2}$ . Therefore, a condition equivalent to Equation 10 is: PPP occurs if  $a'_1 < 0$  or  $a'_1 > 1$ . Another interesting fact is that if PPP occurs (so Equation 10 does not hold), then either  $a'_1 < 0$  and  $a'_2 > 0$  or  $a'_1 > 0$  and  $a'_2 < 0$ , with  $a'_1 + a'_2 = 1$  so that  $\hat{\mu}$  is unbiased. We can now state and prove Theorem 2.

**Theorem 2.** Suppose  $a'_1$  and  $a'_2$  have opposite signs. Then either  $\hat{\mu} < \min(y_1, y_2)$  or  $\hat{\mu} > \max(y_1, y_2)$ . That is,  $\hat{\mu}$  will *always* fall outside the range of the  $(y_1, y_2)$ .

Proof. First assume  $a'_1 > 0$  and  $a'_2 < 0$ . If  $y_1 < y_2$  then  $\hat{\mu} = a'_1 y_1 + a'_2 y_2 < a'_1 y_1 + a'_2 y_1 = y_1$  because  $a'_2 < 0$ . If  $y_1 > y_2$  then  $\hat{\mu} = a'_1 y_1 + a'_2 y_2 > a'_1 y_1 + a'_2 y_1 = y_1$  because  $a'_2 < 0$ . The proof is completed by next assuming  $a'_1 < 0$  and  $a'_2 > 0$ , and following similar steps.

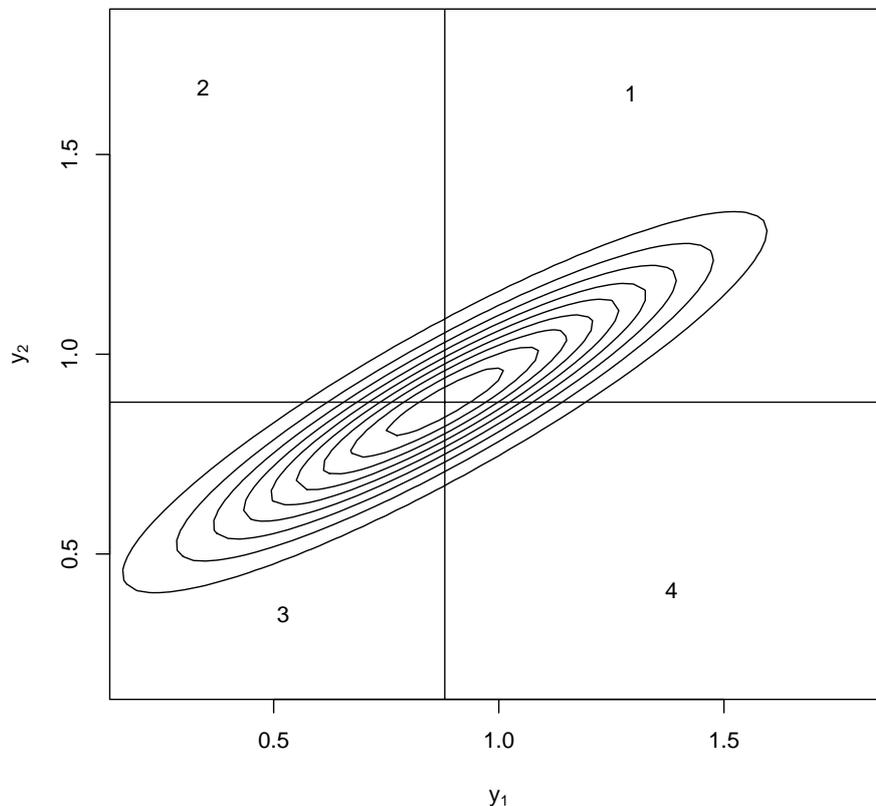
#### 4.1. Additional Support for GLS by Numerical Example

The fact that the GLS estimate is the BLUE estimate (and also the MVUE estimate if the data is Gaussian) and that it lies below the range of  $(y_1, y_2)$  suggests two features. First,  $\mu$  must be likely to fall outside the range of  $(y_1, y_2)$ . Second, there must be better than random chance capability to guess on which side of the range of  $(y_1, y_2)$  that  $\mu$  lies.

In addition to GLS's BLUE property, we can add support for GLS by numerical example to illustrate features one and two. Figure 1 plots the contours of the bivariate normal density having  $\Sigma$  given by Equation 1 and  $\mu = 0.88$  which as shown previously is the OLS estimate of the mean in the case  $y_1 = 1.5$  and  $y_2 = 1.0$ . Informally, we can integrate this density over regions 1 and 3 to see that there

is a large probability that both  $y_1$  and  $y_2$  lie either above the mean or below the mean, so indeed  $\mu$  is likely to fall outside the range of  $(y_1, y_2)$ . Integration of the bivariate normal for  $\Sigma$  given by Equation 1 indicates that with probability approximately 0.40,  $\mu$  lies below the minimum of  $(y_1, y_2)$  and with the same probability  $\mu$  lies above the maximum of  $(y_1, y_2)$ . This is a total of approximately 0.80 probability that  $\mu$  falls outside the range of  $(y_1, y_2)$ , which is an example of feature one. Having an estimate lie outside the range of the data is therefore defensible, provided (feature two) that one can guess with better than random chance performance whether  $\mu$  lies below the minimum or whether  $\mu$  lies above the maximum of  $(y_1, y_2)$ . To see that one can beat random guessing performance, suppose  $y_1 > y_2$  as in our case ( $y_1 = 1.5$  and  $y_2 = 1.0$ ). Then, because  $\sigma_{y_1}^2 = 0.1125$  is larger than  $\sigma_{y_2}^2 = 0.05$  in Equation 1,  $\mu$  is more likely to fall below  $y_2$  because if instead  $\mu > y_1$  then the distance from  $y_1$  to  $\mu$  would be smaller than the distance from  $y_2$  to  $\mu$ , contradicting the fact that  $\sigma_{y_1}^2 > \sigma_{y_2}^2$ . To confirm this line of reasoning, in 10,000 simulations (in the statistical computing language R) of  $(y_1, y_2)$  pairs having  $\mu = 0.88$ , 57% of the simulation runs for which  $y_1 > y_2$  did in fact also have  $\mu < y_2$ . On the basis of 10,000 simulations, 57% is repeatable to within  $\pm 1\%$  or less, so this is better than random chance (50%) guessing. This is not a formal proof but does suggest a direction to understand when a GLS estimate falling outside the range of the data is effective. Note that  $y_1 = 1.5$  and  $y_2 = 1.0$  in the PPP statement, and the GLS estimate is  $\hat{\mu} = 0.88 < y_2$ .

**Figure 1.** Contours of example bivariate normal density of  $y_1$  and  $y_2$  illustrating that  $\mu = 0.88$  is likely to fall outside the range of  $y_1$  and  $y_2$  because of the large probability of  $(y_1, y_2)$  falling in region 1 or 3.



## 5. Example Where PPP Occurs without Approximation

Thus far we have not demonstrated any error model which exactly (without approximation) produces a  $\Sigma$  that leads to PPP. A situation that leads to PPP without the approximations in Equation 1 is expressed in Theorem 3.

**Theorem 3.** Suppose  $m_1 = \mu + \epsilon_{R1}$  and  $m_2 = \mu + \epsilon_{R2}$  where  $\epsilon_{R1}$  is random error in  $m_1$  with variance  $\sigma_{R1}^2$  and similarly for  $\epsilon_{R2}$ . Then if  $y_1 = m_1 + \epsilon_S$  and  $y_2 = m_2 + \alpha\epsilon_S$ , where  $\epsilon_S \sim N(0, \sigma_S)$  and  $\alpha$  is any positive scale factor other than 1, the covariance matrix  $\Sigma$  of  $(y_1, y_2)$  can lie in the PPP region.

**Proof.** The proof is by demonstration. Specify any values  $\sigma_1, \sigma_2$ , and  $\sigma_{12}$  that satisfy the PPP condition ( $\rho > \min\left(\frac{\sigma_1}{\sigma_2}, \frac{\sigma_2}{\sigma_1}\right)$ ). Choose any  $\alpha > 0$ , then  $\sigma_S^2 = \sigma_{12}^2/\alpha$ ,  $\sigma_{R1}^2 = \sigma_1^2 - \sigma_S^2$ , and  $\sigma_{R2}^2 = \sigma_2^2 - \alpha^2\sigma_S^2$ . As an example, let  $\sigma_1^2 = 0.1134$ ,  $\sigma_2^2 = 0.0504$ , and  $\sigma_{12}^2 = 0.06$  as in Equation 3. Then if  $\alpha = 0.7$ , we have  $\sigma_S^2 = \sigma_{12}^2/\alpha = 0.06/0.7 = 0.0857$ ,  $\sigma_{R1}^2 = \sigma_1^2 - \sigma_S^2 = 0.1134 - 0.0857 = 0.0277$ , and  $\sigma_{R2}^2 = \sigma_2^2 - \alpha^2\sigma_S^2 = 0.06 - 0.7^2 \times 0.06/0.7 = 0.018$ .

Note:

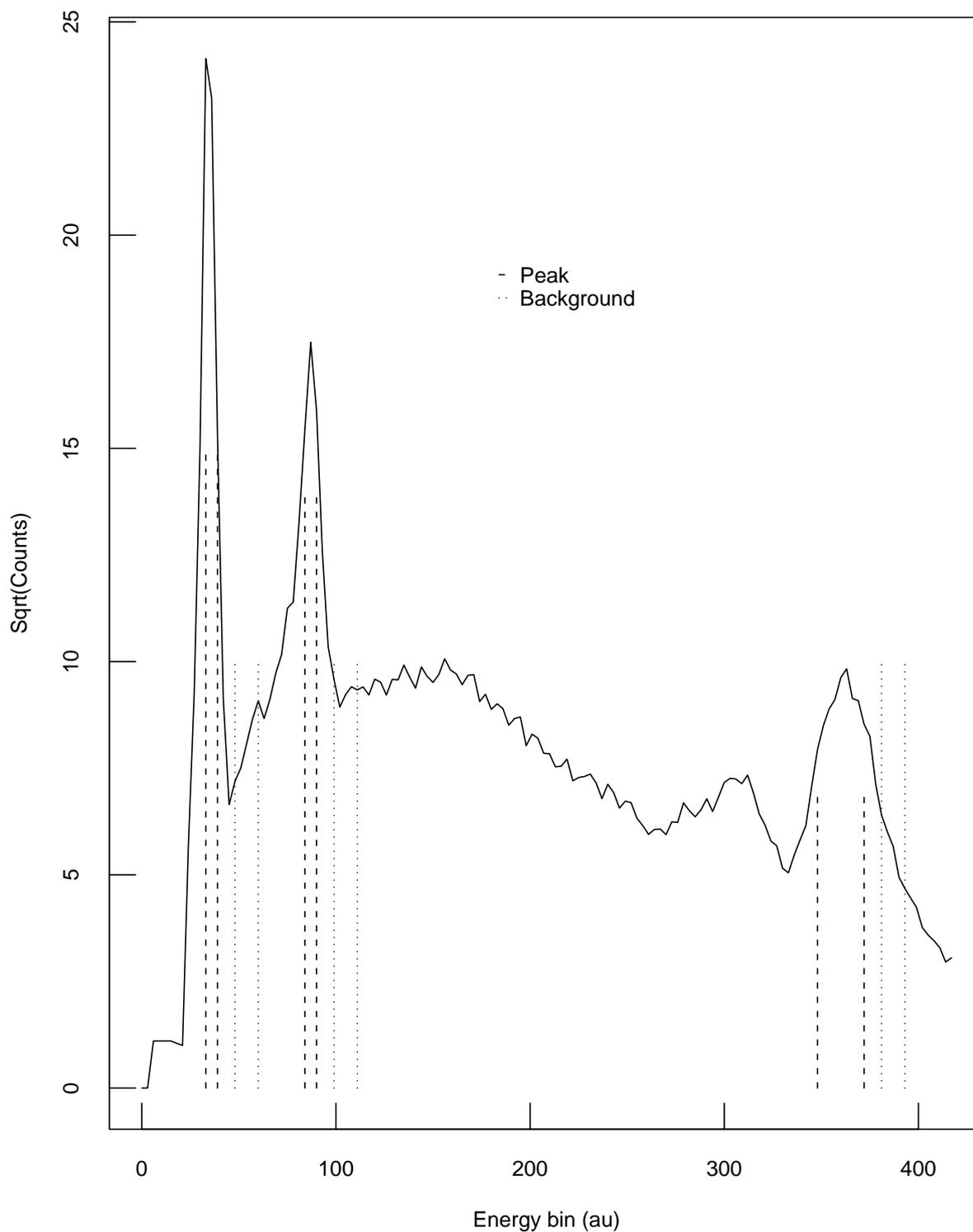
- If  $\alpha = 1$ , then  $\Sigma$  has the form of Equation 11, so PPP cannot occur ([2]).
- If  $\alpha < 0$  then PPP cannot occur. However, our applications have  $\alpha > 0$ . Jones *et al.* [10] showed that if  $\rho \rightarrow \pm 1$  and  $\sigma_1 \neq \sigma_2$ , then  $\mu$  can be estimated with surprisingly small variance. That fact plus the known BLUE property of GLS could convince us to just “live with” the PPP because it can make sense for the GLS estimate  $\hat{\mu}$  to lie outside the range of  $(y_1, y_2)$  in the case  $\alpha > 0$ , or  $\sigma_{12} > 0$ .

One example in which the assumptions of Theorem 3 hold involves subtracting a background measurement from a region of interest (ROI) measurement to get a net result. Because the background measurement often involves a different number of channels than the ROI measurement, a scale factor  $k$  is introduced to estimate the net counts as  $\text{net} = \text{peak} - k \times \text{background}$ . Figure 2 illustrates a hypothetical example where each of three peak ROIs have a corresponding background measurement in the plot of the square root of detected neutron counts *versus* neutron energy in arbitrary units (au).

Suppose each ROI and corresponding background are analyzed separately, and consider the first ROI in Figure 2. The count times could vary between the two experiments, so  $\sigma_{R1}^2 \neq \sigma_{R2}^2$ . Both experiment 1 and experiment 2 measure the ROI counts, but in many situations, the background measurement is made only by experiment 1. In that case, experiments 1 and 2 must use the background measurement made by experiment 1. Also, if the ROI is found by analyzing the shape of the curve (the “spectrum”) that describes detected count rates *versus* particle energy, then the ROI for experiments 1 and 2 could differ. We then have  $N_1 = G_1 - a_1 \times B$  and  $N_2 = G_2 - a_2 \times B$ , where  $N$  is net counts,  $G$  is gross counts,  $B$  is background,  $a_1$  is the scale factor for experiment 1, and  $a_2$  is the scale factor for experiment 2. The scale factor  $a_1$  for experiment 1 is the ratio of the number of ROI channels to the number of background channels, and similarly for the scale factor  $a_2$  for experiment 2. The channel counts have variation from repeat to repeat so the detected counts will vary around the true counts with some error. As an aside, often the channel counts have approximately a Poisson distribution which for large count rates is well

approximated by a Gaussian distribution. Regardless of which probability distribution best describes the channel counts, there are measurement errors in  $N_1$ ,  $N_2$ , and one can divide  $N_1 = G_1 - a_1 \times B$  by  $a_1$  to convert this pair of equations to those assumed in Theorem 3.

**Figure 2.** Example region of interest and corresponding background that can lead to the PPP condition.



## 6. Conclusions

Because Peelle's original statement is vague, there have been several interpretations and solutions. In our experience, there is considerable variation among experimentalists in the expression of measurement uncertainty, and a wide the range of analyses can result from vague uncertainty statements.

The three main contributions of this paper are: (1) illustrating examples when PPP cannot occur (Theorem 1); (2) providing insight when PPP is effective and appropriate (related to Theorem 2), and (3) deriving a realistic covariance matrix  $\Sigma$  for which PPP occurs according to physical descriptions of realistic measurement scenarios (Theorem 3). We also showed via numerical integration that an estimate lying outside the range of the data is sensible. This is because the unequal variances of  $x_1$  and  $x_2$  provide information regarding whether  $\mu$  is more likely to be less than the minimum or greater than the maximum of  $x_1$  and  $x_2$ .

Of course GLS provides a good estimate  $\hat{\mu}$  in general because of its well-known BLUE property (and MVUE if the data is Gaussian) and in particular for the PPP problem if the covariance  $\Sigma$  is well known.

There will almost always be estimation error in  $\hat{\Sigma}$ , and often the measurement errors are nonGaussian. Therefore, we consider the following two topics in [6]: (1) alternatives to GLS when there is estimation error in  $\hat{\Sigma}$ , and to provide estimators other than GLS that use the estimated likelihood in the case of non-Gaussian error models that make ML estimation difficult. Regarding (1), it is known that weighted estimates do not always outperform equally-weighted estimates when there is estimation error in the weights. We have already noted that estimation of  $\Sigma$  in Equation 1 introduces apparent PPP when PPP does not actually occur.

Finally, [10] showed that when the diagonal entries in  $\Sigma$  are different and the correlation is large and positive, the GLS estimate can have lower variance than if the correlation is zero. This fact does not seem to be widely known, and experimental opportunities to exploit this fact are under investigation.

## Acknowledgements

We acknowledge the next generation safeguards initiative within the U.S. Department of Energy.

## References

1. Peelle, R. *Peelle's Pertinent Puzzle*; Oak Ridge National Laboratory Memorandum: Washington DC, USA, 1987.
2. Nichols, A.L. *International Evaluation of Neutron Cross Section Standards*; International Atomic Energy Agency: Vienna, Austria, 2007.
3. Chiba, S.; Smith, D. *A Suggested Procedure for Resolving an Anomaly in Least-Squares Data Analysis Known as Peelle's Pertinent Puzzle and the General Implications for Nuclear Data Evaluation*, Report ANL/NDM-121; Argonne National Laboratory: Argonne, IL, USA, 1991.
4. Zhao, Z.; Perey, R. *The Covariance Matrix of Derived Quantities and Their Combination*, Report ORNL/TM-12106; Oak Ridge National Laboratory: Oak Ridge, TN, USA; 1992.

5. Hanson, K.; Kawano, T.; Talou, P. Probabilistic Interpretation of Peelle's Pertinent Puzzle. In Proceeding of *International Conference of Nuclear Data for Science and Technology*, 26 September–1 October 2004; Haight, R.C., Chadwick, M.B., Kawano, T., Talou, P., Eds.; AIP Conference Proceedings: Melville, NY, USA, 2005.
6. Burr, T.; Kawano, T.; Talou, P.; Hengartner, N.; Pan, P. Alternatives to the Least Squares Solution to Peelle's Pertinent Puzzle, submitted for publication, 2010.
7. Christensen, R. *Plane Answers to Complex Questions, The Theory of Linear Models*; Springer: New York, 1999; pp. 23–25.
8. Burr, T.; Frey, H. Biased Regression: The Case for Cautious Application. *Techometrics* **2005**, *47*, 284–296.
9. Sivia, D.; Data Analysis—A Dialogue With The Data. In *Advanced Mathematical and Computational Tools in Metrology VII*, Ciarlini, P., Filipe, E., Forbes, A.B., Pavese, F., Perruchet, C., Siebert, B.R.L., Eds.; World Scientific Publishing Company: Lisbon, Portugal; 2006; pp. 108–118.
10. Jones, C.; Finn, J.; Hengartner, N. Regression with Strongly Correlated Data. *J. Multivariate Anal.* **2008**, *99*, 2136–2153.
11. Finn, J.; Jones, C.; Hengartner, N. Strong Nonlinear Correlations, Conditional Entropy, and Perfect Estimation. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, Proceeding of 27th International Workshop Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Saratoga Springs, NY, USA, 2007; AIP Conference Proceedings: Melville, NY, USA, 2007; p. 954.

© 2011 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>.)