scientific reports

Check for updates

OPEN

Understanding Peelle's Pertinent Puzzle bias in generalized least squares regression through eigenspectrum analysis

Noah A. W. Walton[1,2✉], William N. Fritsch[2], Amanda M. Lewis[1,2], Jacob Forbes[2], Denise Neudecker[1] & Vladimir Sobes[2]

Certain correlation structures in the data covariance matrix (DCM) used for generalized least squares (GLS) regression can result in biased estimates, commonly known in the field of nuclear data evaluation as Peele's Pertinent Puzzle (PPP). This article introduces a generative, forward modeling framework within which the PPP bias is characterized through an eigenspectrum analysis of the DCM. This analysis highlights the root cause of the bias, generalizes the problem beyond the nuclear data field, and provides insight to the problem regimes where it can occur. What follows is an understanding that the bias can show up for any experimental neutron time-of-flight data for which systematic uncertainties have been quantified. Lastly, a discussion of the adaptation of cross validation approaches that require pre-whitening to incorporate the known 'fix' to the PPP bias in the GLS estimator.

Generalized least squares (GLS) is a well-established method for regression in the physical sciences and can yield maximum likelihood estimates (MLE) in the Bayesian framework with an uninformative prior and Gaussian/linearity assumptions. GLS is commonly used to infer model parameters in the physical sciences, including the analysis of experimental nuclear data. Both methods are subject to the phenomenon known in the nuclear data field as Peelle's Pertinent Puzzle (PPP), named after its discovery in the context of nuclear data by R.W. Peelle[1]. PPP describes a—sometimes extreme—bias in the GLS estimator that can occur when data have both statistical and highly correlated systematic (diagonal and non-zero off-diagonal) covariance. In statistics fields, such bias in the GLS estimator was already understood and the alternative, more robust iteratively re-weighted least squares (IRLS) estimator would have been suggested[2]; however, differences in scientific language silo-ed the nuclear data field.

In the context of nuclear data, the PPP phenomenon has been extensively explored since its discovery. Peelle's informal memorandum describing the issue[1] set off a flurry of work on the topic in the following years. In 1991, Chiba and Smith[3] presented an extensive study of the problem, the background, and a suggested solution. They first addressed the use of the least squares method in nuclear data evaluation, acknowledging that some assumptions inherent in the proofs of least squares properties are not always met (such as linearity of the model, independence of the data covariance matrix (DCM) from the solution, and that the underlying data are normally distributed). After limiting the scope to linear relationship, they conclude that least squares is a viable method for these problems from both a Probabilistic and a Bayesian perspective. The problem called PPP is then described in detail, with a full reproduction of the memorandum, the equations, and the resulting fitted mean lying outside of the range of the two data points, which they address (in emphasized text) as such[3]:

> "This is indeed quite proper if one accepts the given absolute values of the covariance matrix elements as being correct."

The remainder of the report focuses on how to avoid the PPP result by modifying the absolute values of the covariance matrix. They postulate that the underlying cause of the behavior is the extensive use of fractional errors in experimental data analysis. These come from both underlying physics (such as counting statistics) and from the analysis equations which often contain multiplicative factors and ratios. Fractional errors applied to discrepant data points (an all-too-frequent occurrence in physical measurements) lead to highly discrepant absolute errors and the condition known to produce PPP results (when the covariance between two data points is larger than the variance of one of the data points).

[1]Los Alamos National Laboratory, Los Alamos 87545, NM, USA. [2]Nuclear Engineering Department, University of Tennessee, Knoxville 37996, TN, USA. ✉email: nwalton@lanl.gov

In the nuclear data field, fractional errors are typically understood to mean that confidence in the result is not dependent on the magnitude of the result. This information is not translated into the least squares framework when fractional errors are applied to the data points to produce absolute errors used in regression. Instead, the covariance matrix indicates that the confidence in some data points is much higher than the confidence in other data points. Chiba and Smith[3] provide a workaround to the PPP result by constructing a covariance matrix that contains better represents this information. They recommend that the absolute errors should not be calculated by multiplying the fractional errors by the measured data points, but rather by multiplying them by a "reasonable a priori estimate" of the true mean. How to calculate such an estimate has been the focus of much of the work on PPP in the nuclear data field.

In their initial report, Chiba and Smith[3] proposed what is essentially the IRLS algorithm as the way to determine the a priori estimate. IRLS was later introduced as the MLE for the class of Generalized Linear Models, a generalization of ordinary least squares that allows non-linear response functions, bounded response variables, and non-normal error distributions[4]. The use of IRLS (usually not identified as such) has been the most popular PPP solution the nuclear data field, with various justifications, derivations, and extensions presented in a multitude of papers[5–15]. Justifications for this method include 'hidden variables' which create the correlation[5,6,9,10,15], that relative experimental uncertainties should be applied to the 'true' physical parameter being measured rather than the experimental result[9,11,12], and the non-linearity inherent in analysis equations utilizing ratios[7].

Another method to properly represent the confidence in relative uncertainties is to transform the variables, rather than the covariance matrix used in the fit. Relative uncertainties lead to non-constant variance, which (1) violates the assumptions of GLS, and (2) hints at an underlying non-Gaussian data distribution. Products and ratios of normally-distributed variables (the cause of the non-constant variance) are not themselves normally-distributed, even though they can be close under certain circumstances. Transforming the to log-normal distributions, the typical method for handling multiplicative errors[4], has been proposed as a method to solve PPP[16,17] by encoding the meaning of relative uncertainties in the fit.

An exhaustive survey and inter-comparison of the various causes and solutions was performed by the Standards CRP[18] in anticipation of the release of a new set of nuclear data standards in 2009[19]. The PPP effect was split into two categories: the 'mini-PPP' effect, caused by relative uncertainties, which leads to lower absolute uncertainty on lower values, and the 'maxi-PPP' effect, caused by strong positive correlations and discrepant uncertainties, which leads to fitted values outside the range of the data. The inter-comparison exercise lead to the adoption of the log-transformation for the Standards evaluation due to the impracticality of a full hidden-variables analysis[19]. This is consistent with the general recommendation in the field—use as much information as possible to avoid the hidden variables problem (something possible in only limited realistic circumstances), but if that is not possible, transform the model or the covariance matrix to correctly encode uncertainty information (i.e., relative uncertainties are meant to represent confidence that is independent of the magnitude of the measurement) into the regression equations.

PPP has been extensively explored in the nuclear data field, so, what does this article contribute? First and foremost, we introduce a new interpretation of the PPP bias using generative modeling and eigenspectrum decomposition, which generalizes the experimental analysis-focused explanations of PPP to any case of a rank-1 perturbation on a Hermitian matrix. In this work, generative modeling describes a computational approach that presumes the statistical distribution of the data itself is known and can be used to generate effectively infinite, statistically consistent samples. This approach is used to consider the inference problem from a frequentist perspective and highlight bias with respect to the known generating distribution. Secondly, we leverage the eigen-decomposition to derive an approximate regime where the PPP bias is expected to occur, thus highlighting what elements of the GLS problem exacerbate the bias. Thirdly, as far as the authors understand, the existing literature on PPP and its solution in the nuclear data field generally make two assumptions: 1) the systematic error term comes from a normalization factor which can therefore be converted to relative errors, and 2) analyses should generally not observe the PPP bias if the data at hand are not strongly correlated and/or discrepant. Through the interpretation and subsequent numerical demonstration presented here, we show that the PPP bias is not strictly limited to either assumption and make the conjecture that it can occur in any neutron time-of-flight data with proper uncertainty quantification (*i.e.*, the off diagonal elements are non-zero). Lastly, we discuss how to handle cross validation when using IRLS as an estimator, a discussion brought about by emerging methods for nuclear resonance evaluation[20]. To our knowledge, this final point of implementing cross validation with IRLS has not been addressed in the context of nuclear data or more broadly for inferential regression in the physical sciences.

Introduction of the classical PPP

In 1987 at Oak Ridge National Laboratory, R.W. Peelle, a physicist doing nuclear data evaluation, described having two observations to estimate a shared mean[1]. While found in many publications, the problem is re-formulated here in notation consistent with later sections of this article.

The two observations are 20% "fully correlated," and each have an independent uncertainty of 10%. Relative uncertainties are assumed to be "1 sigma" values.

$$y_1 = 1.0 \pm 10\% \tag{1}$$

$$y_2 = 1.5 \pm 10\% \tag{2}$$

Peelle worked out the GLS estimate of the mean, as is standard in the field. What he found was that the GLS estimate of the mean was 0.88 ± 0.25 and fell outside of both observations. We formulate the problem as follows,

$$\vec{y} = \begin{bmatrix} 1.0 \\ 1.5 \end{bmatrix} \tag{3}$$

$$\Sigma = 0.1^2 \begin{bmatrix} y_1^2 & 0 \\ 0 & y_2^2 \end{bmatrix} + \vec{y}\, 0.2^2\, \vec{y}^T \tag{4}$$

$$\hat{\sigma}^2 = (\vec{1}^T\, \Sigma^{-1}\, \vec{1})^{-1} \tag{5}$$

$$\hat{\mu} = \hat{\sigma}^2\, (\vec{1}^T\, \Sigma^{-1}\, \vec{y}) \tag{6}$$

with equation 6 giving the GLS estimate of the mean. Here, we formulate the full covariance (equation 4) as a rank-1 perturbation to a diagonal matrix. The first term on the right-hand side of equation 4 represents the independent uncertainties on both observations and is a 2x2 diagonal matrix which is always full-rank. The second term represents fully correlated uncertainty and can be expanded to a 2x2 matrix; however, that matrix will always be rank-1.

PPP succinctly summarized

From the perspective of the authors, PPP describes a false solution mode that comes about from nothing more than a violation of the assumptions in the application of GLS by using a particular *estimator* of the DCM in place of the *true* DCM, that is, an accurate description of the covariance about the true mean. The aforementioned correction for PPP (IRLS in statistics fields) gives a better estimate to the DCM based on the current model estimate. The false mode will often trend towards 0, this is seen in the classical PPP problem as the estimate is below both observations and generally manifests in the nuclear data literature as a negatively biased estimate given data that is constrained to be positive. The following explanation of this false mode will expose the general trend of the false mode toward 0 and that it is not strict for multi-dimensional, non-linear problems where, more generally, the false mode trends toward a nominal or constant signal.

A new framework for PPP
The frequentist interpretation

One challenge in discussing PPP and interpreting the result of the GLS estimator, is that Peelle suggested one set of numerical values for the measurements. So, perhaps, one could ask if the values Peelle chose are a statistical outlier and the GLS procedure is, in fact, sound but gives an "outlier result" when presented with "outlier data." Therefore, we propose a frequentist approach for generating data samples which, when used in a Bayesian estimating procedure (such as GLS), will yield false results, not just once, but for a statistically significant number of sampled data. Additionally, this approach provides a way to quantitatively validate the estimating procedure by testing the credible intervals predicted by the Bayesian posterior distribution, $p(\mu)$.

We define a "data-generating model" with a known *true* mean and covariance (μ_{true} and Σ_{true}) describing a multivariate normal distribution, from which we can draw sample data ($\vec{d}$),

$$\vec{D} \sim \mathcal{N}(\mu_{\text{true}}, \Sigma_{\text{true}}), \tag{7}$$

where μ_{true} is non-zero. We then apply the estimating procedure and empirically construct the credible intervals of the Bayesian posterior distribution,

$$P\{q_{\alpha/2} < \mu_{\text{true}} < q_{(1-\alpha)/2}\} = 1 - \alpha, \tag{8}$$

where $q_{\alpha/2}$ and $q_{(1-\alpha)/2}$ are the lower and upper α-quantiles of the Bayesian posterior distribution. Within this framework, we cannot construct Peelle's original problem exactly. The reason is that the covariance matrix to sample the data is stated in terms of the sampled data itself,

$$\Sigma = 0.1^2 \begin{bmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{bmatrix} + \vec{d}\, 0.2^2\, \vec{d}^T. \tag{9}$$

This recursive definition sheds light on the issue at hand. Herein, we provide a slight but important clarification to Peelle's original statement. The DCM provided is not the *true* DCM, but rather, is an estimator, $\hat{\Sigma}$, of the *true* DCM based on the measured data, the *assumed* structure of the DCM, and the uncertainty on the normalization parameter. This has been understood in several publications[3,9,11].

With this interpretation of PPP, we construct the data generating model with the *true* mean and DCM. The generative model follows as:

1. Draw a data sample, $\vec{d}$, from a multi-variate normal distribution, $\vec{D} \sim \mathcal{N}(\mu_{\text{true}}, \Sigma_{\text{true}})$

2. Calculate the *estimated* DCM, $\hat{\Sigma} = 0.1^2 \begin{bmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{bmatrix} + \vec{d}\, 0.2^2\, \vec{d}^T$

3. Apply the estimating procedure based on the observed data, $\vec{d}$, and the *estimated* DCM, $\hat{\Sigma}$.

Defining the generative model gives us a powerful tool in the demonstration of the PPP phenomenon, that is, since we have defined the *true* mean, μ_{true}, we can identify false solution modes with certainty.

Relating to experimental neutron time-of-flight data

Experimental neutron time-of-flight data makes up a significant portion of the observations used in nuclear data evaluation. These data will always have a statistical variance component, from radiation counting statistics, and one or more systematic components. The systematic components are related to uncertainty on one or more data reduction parameters used to mathematically transform the raw radiation counting into the quantity of interest (reaction cross section, yield, etc.). The statistical (uncorrelated) and systematic (correlated) components are seen in the DCM of the classical PPP problem, with the correlated uncertainty often interpreted as a data reduction parameter that scales/normalizes the entire spectrum. Many experimental nuclear reaction cross section data have this feature. In fact, evaluators are encouraged to approximate the uncertainty from this, and other, data reduction parameters even if it is not explicitly reported[21–24].

The PPP phenomenon is often presented with relative uncertainties, one data reduction parameter that acts as an overall scaling factor, and in the 2-dimensional form presented in Sec. Introduction of the classical PPP. Herein, we expand our study to align more closely with the features of real experimental datasets. We demonstrate that the false solution mode associated with PPP can occur for any data reduction parameter for which the sensitivity is approximately proportional to the measured data, not just an overall normalization factor. We recognize that in real neutron-time-of-flight measurements, there are often more than one data reduction parameters and consider this in our analysis. Finally, we generalize the frequentist set up of PPP to an arbitrary number of data points (similar to Ref[12]) by defining the following quantities:

- *True* mean vector, $\vec{\mu}_{\text{true}}$, of arbitrary length, N.

  - *True* DCM, Σ_{true}, of arbitrary size, $N \times N$, and constructed as: $\Sigma_{\text{true}} = \text{diag}(\vec{\delta}^2) + \vec{\mu}_{\text{true}} \, (\Delta n)^2 \, \vec{\mu}_{\text{true}}^T$ where, $\vec{\delta}$ is a vector of stochastic uncertainties on the individual data points
  - where, Δn, is the normalization uncertainty

- *Estimated* DCM, $\hat{\Sigma}$, of arbitrary size, $N \times N$, still constructed as a rank-1 perturbation to a diagonal matrix: $\hat{\Sigma} = \text{diag}(\vec{\delta}^2) + \vec{d} \, (\Delta n)^2 \, \vec{d}^T$

For both the *true* and *estimated* DCM, both $\vec{\delta}$ and Δn are known. Varying the value of Δn will allow a study of the impact of the magnitude of the normalization uncertainty and in Subsection The eigen-decomposition explanation of the false mode we will derive an approximate threshold value for Δn for the occurrence of a false mode. To give some intuition to the frequentist setup, consider the *true* mean vector, $\vec{\mu}_{\text{true}}$, to represent the underlying *true* value of the observable quantity of interest in nature.

The eigen-decomposition explanation of the false mode

Setting up the MLE problem results in finding the vector, $\hat{\vec{\mu}}$, which maximizes the likelihood function, $\mathscr{L}$:

$$\hat{\vec{\mu}} = \text{argmax} \, \mathscr{L}(\vec{\mu}) \tag{10}$$

$$\mathscr{L}(\vec{\mu}) \propto \exp\left[-\frac{1}{2}(\vec{d} - \vec{\mu})^T \Sigma^{-1}(\vec{d} - \vec{\mu})\right] \tag{11}$$

$$\chi^2(\vec{\mu}) = (\vec{d} - \vec{\mu})^T \Sigma^{-1}(\vec{d} - \vec{\mu}) \tag{12}$$

$$\hat{\vec{\mu}} = \text{argmin} \, \chi^2(\hat{\mu}) \tag{13}$$

The original contribution of this work is to consider the eigen-decomposition of, Σ, the DCM used in Equation 12:

$$\Sigma = Q\Lambda Q^T \tag{14}$$

$$\Sigma^{-1} = Q\Lambda^{-1}Q^T \tag{15}$$

$$\chi^2(\vec{\mu}) = \sum_i \lambda_i^{-1} ||Q_i^T(\vec{d} - \vec{\mu})||_2^2 \tag{16}$$

In equation 16 we see that the traditional χ^2-metric becomes a simple sum of the projection of the vector of residuals, $(\vec{d} - \vec{\mu})$, on the eigenvectors of the DCM.

Consider the diagonal DCM with only stochastic uncertainties. In this case, the matrix of eigenvectors, Q, is just the identity matrix with the eigenvalues equal to the stochastic uncertainty of each data point, represented by the vector $\vec{\delta}$. Under the assumption of similar counting statistics for all data points (a physically justifiable assumption for radiation counting experiments as the total number of counts is generally orders of magnitude larger than the local changes from the signal being measured), the eigenvalues will be clustered together with a relatively tight spread. For example, if the *true* mean, $\vec{\mu}_{\text{true}}$, has values on the order of 10^6 and we consider Poisson counting statistics, for a purely stochastic DCM estimated based on the measured data, we would expect all the values of $\vec{\delta}$ to be on the order of 10^3. Similarly, the eigenvalues of the inverse of the DCM, $\vec{\lambda}^{-1}$, would be clustered around 10^{-3}. Herein, we will refer to this cluster of the eigenvalues for the stochastic DCM as the "*bulk*."

Consider a simplified, specific example of the frequentist set up of PPP with $N = 4$ dimensions,

$$\vec{\delta} = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^T. \tag{17}$$

In this case, all the eigenvalues of the stochastic DCM are unity, and the spread of the *bulk* is zero. Next, we identify the addition of the systematic component of the DCM as a rank-1 perturbation to the stochastic data covariance,

$$\Sigma = \text{diag}(\vec{\delta^2}) + \vec{d}(\Delta n)^2 \vec{d}^T. \tag{18}$$

The eigenvalue behavior for a rank-1 perturbation to any symmetric matrix obeys interlacing[25,26], which says that all but one of the eigenvalues of the combined DCM (stochastic and systematic) will remain bound within the *bulk* and only the top eigenvalue has the *possibility* to escape above the upper range of the *bulk*. For example, consider an $N \times N$ real symmetric matrix, A, with ordered eigenvalues $\mu_1 \geq \mu_2 ... \geq \mu_N$ and $B = A + \rho u u^T$, where u is a real column vector and ρ is a scalar. The matrix B has eigenvalues $(\lambda_1 \geq \lambda_2 ... \geq \lambda_N)$ and interlacing says that $\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 ... \geq \lambda_N \geq \mu_N$. This can also be seen from the more general Cauchy interlacing theorem[27] by embedding the rank-1 perturbation into a bordered matrix C for which A is a principal submatrix,

$$C = \begin{bmatrix} A & u \\ u^{\text{T}} & -1/\rho \end{bmatrix}. \tag{19}$$

The Schur compliment of C is equal to B exactly, consequently C will share all eigenvalues of B, plus one additional given by $-1/\rho$. It follows then that Cauchy interlacing of A with C also causes interlacing of A with B.

In the simplified case here, the rank-1 perturbation forces the largest eigenvalue to separate from the bulk. The other three eigenvalues remain at unity, bound by the Cauchy interlacing theorem. The value of the largest eigenvalue can be determined by subtracting the three fixed eigenvalues from the trace of the matrix (the trace of the matrix equals the sum of the eigenvalues),

$$\lambda_{\max} = 1 + \Delta n^2 ||\vec{d}||_2^2. \tag{20}$$

For an arbitrary $\vec{\delta}$, we do not have an analytic expression for the largest eigenvalue. However, in the limit that the magnitude of the rank-1 perturbation grows large enough, while the $\vec{\delta}$ maintains its finite range, then Equation 20 becomes predictive and gives good intuition. This is justified by the regime assumption,

$$\langle \vec{\delta^2} \rangle >> (\max \vec{\delta^2} - \min \vec{\delta^2}), \tag{21}$$

where $\langle \vec{\delta^2} \rangle$ is the average stochastic uncertainty.

As the magnitude of the rank-1 perturbation grows larger, the trace of the matrix continues to increase monotonically, however the sum of all of the eigenvalues, except the largest one, is bounded by the largest value in $\vec{\delta}$ per the interlacing theorem. Therefore, the largest eigenvalue has to absorb the difference and move further away from the bulk. For arbitrary $\vec{\delta}$, the largest eigenvalue separated from the bulk has a lower limit,

$$\lambda_{\max} = \max(\vec{\delta^2}) + \Delta n^2 ||\vec{d}||_2^2. \tag{22}$$

The Bunch–Nielsen–Sorensen formula[26] gives an exact equation for the behavior of the eigenvectors due to a rank-1 perturbation of a diagonal matrix. To observe the PPP phenomenon, we need only analyze the eigenvector corresponding to the eigenvalue which has escaped the *bulk*. The elements of the eigenvector, $\vec{q}_{\max}$, which corresponds to the largest eigenvalue, $\lambda_{\max}$ are,

$$q_j = b \left(\frac{1}{\lambda_{\max} - \delta_j^2} \right) d_j, \tag{23}$$

where b is a constant to ensure that the eigenvector remains normalized. Elements of the data vector, $\vec{d}$, appear in Equation 23 only because the systematic component of uncertainty is estimated based on the observed data in the estimated DCM. This is exactly the cause of the PPP phenomenon.

In the simplified example above, where all elements of $\vec{\delta}$ are equal, the eigenvector, $\vec{q}_{\max}$, is exactly aligned with the measured data, $\vec{d}$, because the term in the parenthesis of Equation 23 is constant for all vector elements, j. Again, the regime defined by Equation 21 allows Equation 23 to become predictive for the more realistic scenario where all elements of $\vec{\delta}$ are not equal but the spread is much smaller than the maximum eigenvalue, $\lambda_{\max}$. In this case, $\vec{q}_{\max}$ is nearly-aligned with the experimentally observed data since $\vec{\delta}$ varies relatively little across its elements.

Returning to the eigenvalue decomposition of the χ^2 minimization objective in Equation 12, we can explicitly pull out the eigen-mode corresponding to the eigenvalue which has escaped the *bulk*,

$$\chi^2(\vec{\mu}) = \lambda_{\max}^{-1} \left|\left| \vec{q}_{\max}^T (\vec{d} - \vec{\mu}) \right|\right|_2^2 + \sum_{i=1}^{N-1} \lambda_i^{-1} ||Q_i^T (\vec{d} - \vec{\mu})||_2^2. \tag{24}$$

If the range of the values in the *bulk* (i.e.: $\max \vec{\delta^2} - \min \vec{\delta^2}$), is much smaller than the maximum eigenvalue, $\lambda_{\max}$, then $\vec{q}_{\max}$ is aligned with the experimentally observed data, $\vec{d}$,

$$\vec{q}_{\max} \approx \frac{\vec{d}}{||\vec{d}||_2}, \tag{25}$$

$$\chi^2(\vec{\mu}) \approx \lambda_{\max}^{-1} \left|\left| \left(\frac{\vec{d}}{||\vec{d}||_2}\right)^T (\vec{d} - \vec{\mu}) \right|\right|_2^2 + \sum_{i=1}^{N-1} \lambda_i^{-1} ||Q_i^T (\vec{d} - \vec{\mu})||_2^2, \tag{26}$$

where Equation 25 comes about because of the normalization of the eigenvector.

Now, consider the projection of $(\vec{d} - \vec{\mu})$ onto the set of eigenvectors, $\vec{Q}_i$, for everything but the eigen-mode corresponding to $\lambda_{\max}$. In this case, a minimum of Eq. 26 appears near zero. It can be seen by taking $\hat{\vec{\mu}} = 0$, then $(\vec{d} - \hat{\vec{\mu}})$ (almost) aligns with the eigen-mode corresponding to $\lambda_{\max}$ and, by definition, becomes (almost) orthogonal to all of the other eigenvectors, driving the summation term of Equation 26 towards 0. Thus,

$$\chi^2(\hat{\vec{\mu}}) \approx \lambda_{\max}^{-1} ||\vec{d}||_2^2, \tag{27}$$

and substituting Eq. 22,

$$\chi^2(\hat{\vec{\mu}}) \approx \frac{||\vec{d}||_2^2}{\max(\vec{\delta}^2) + \Delta n^2 ||\vec{d}||_2^2}. \tag{28}$$

Note that the same argument can be made for any vector $\hat{\vec{\mu}}$ with all elements equal as the vector $(\vec{d} - \hat{\vec{\mu}})$ would still be (almost) orthogonal to all Q_i^T in the summation term of Equation 26.

Equation 28 can also be derived by applying the Woodbury matrix identity[28] to the inverse of the estimated covariance matrix in the calculation of the χ^2 and plugging-in $\hat{\mu} = 0$:

$$\chi^2(\hat{\mu}) = (\vec{d} - \vec{\mu})^T \hat{\Sigma}^{-1} (\vec{d} - \hat{\mu}) \tag{29}$$

$$\chi^2(\hat{\mu} = 0) = \vec{d}^T \hat{\Sigma}^{-1} \vec{d} \tag{30}$$

$$\hat{\Sigma}^{-1} = \left(\text{diag}(\vec{\delta}^2) + \vec{d}\,(\Delta n)^2\,\vec{d}^T \right)^{-1} \tag{31}$$

$$\hat{\Sigma}^{-1} = \text{diag}(\vec{\delta}^{-2}) - \frac{\text{diag}(\vec{\delta}^{-2})\vec{d}\vec{d}^T \text{diag}(\vec{\delta}^{-2})}{\Delta n^{-2} + \vec{d}^T \text{diag}(\vec{\delta}^{-2})\vec{d}} \tag{32}$$

$$\chi^2(\hat{\mu} = 0) = \vec{d}^T \text{diag}(\vec{\delta}^{-2})\vec{d} - \frac{\left(\vec{d}^T \text{diag}(\vec{\delta}^{-2})\vec{d}\right)\left(\vec{d}^T \text{diag}(\vec{\delta}^{-2})\vec{d}\right)}{\Delta n^{-2} + \vec{d}^T \text{diag}(\vec{\delta}^{-2})\vec{d}} \tag{33}$$

$$\chi^2(\hat{\mu} = 0) = \left(\vec{d}^T \text{diag}(\vec{\delta}^{-2})\vec{d}\right) - \frac{\left(\vec{d}^T \text{diag}(\vec{\delta}^{-2})\vec{d}\right)^2}{\Delta n^{-2} + \left(\vec{d}^T \text{diag}(\vec{\delta}^{-2})\vec{d}\right)} \tag{34}$$

$$\chi^2(\hat{\mu} = 0) = \frac{\Delta n^{-2}\left(\vec{d}^T \text{diag}(\vec{\delta}^{-2})\vec{d}\right)}{\Delta n^{-2} + \left(\vec{d}^T \text{diag}(\vec{\delta}^{-2})\vec{d}\right)} \tag{35}$$

$$\chi^2(\hat{\mu} = 0) = \frac{\left(\vec{d}^T \text{diag}(\vec{\delta}^{-2})\vec{d}\right)}{1 + \Delta n^2 \left(\vec{d}^T \text{diag}(\vec{\delta}^{-2})\vec{d}\right)} \tag{36}$$

The final result in Equation 36 is exact. It is slightly different from the approximate result in Equation 28 in that the norm of the data vector $\vec{d}$ is weighted by the stochastic uncertainty, and assuming that the stochastic uncertainties are equal for all of the data points reconciles Equations 36 and Equation 28. Notice, however, that for this derivation we needed to know ahead of time that the false mode minima of $\chi^2(\hat{\mu})$ occurs at $\hat{\mu} = 0$ when all stochastic uncertainties are equivalent.

Derivation of regime where occurrence of a false mode is expected
The global minimum of $\chi^2(\hat{\mu})$ is zero, which occurs when $\hat{\mu}$ matches the observed data exactly (i.e., $\hat{\mu} = \vec{d}$); however, $\hat{\mu}$ is often constrained by some model with a desired minimum $\chi^2(\hat{\mu})$ approximately equal to N, the number of observations. More precisely, the expected value of $\chi^2(\hat{\mu})$ is the number of *independent* observations less the number of *independent* model parameters—provided that the model is a good representation of the underlying data-generating process and there are no systematic biases in the data. However, for non-linear models, the effective number of model parameters can be difficult to determine, and if this number is small compared to N, its precise value becomes less critical. Therefore, for generality, we take the expected value to be $\chi^2(\hat{\mu}) \approx N$.

Under conditions of the aligned eigenvector, we know that the false mode minima of $\chi^2(\hat{\mu})$ will occur at $\hat{\mu} = 0$ with a value given by Equation 28. Therefore, we can establish an approximate threshold where the false mode becomes the global minimum,

$$\frac{||\vec{d}||_2^2}{\max(\vec{\delta^2}) + \Delta n^2 ||\vec{d}||_2^2} < N, \tag{37}$$

where the left hand side represents the false mode minima of $\chi^2(\hat{\mu})$ and the right hand side represents the minima of $\chi^2(\hat{\mu})$ given by a model that statistically explains the data. Re-arranging, we can see how different parameters, particularly the systematic uncertainty, influence this threshold:

$$\Delta n^2 ||\vec{d}||_2^2 > \frac{||\vec{d}||_2^2}{N} - \max(\vec{\delta^2}), \tag{38}$$

$$\Delta n^2 > \frac{1}{N} - \frac{1}{||\vec{d}||_2^2/\max(\vec{\delta^2})}. \tag{39}$$

The earlier regime assumption that the spread of elements of $\vec{\delta}$ is much smaller than the maximum eigenvalue, $\lambda_{\max}$, is restated here:

$$\max(\vec{\delta^2}) + \Delta n^2 ||\vec{d}||_2^2 >> (\max \vec{\delta^2} - \min \vec{\delta^2}). \tag{40}$$

The eigenvalue decomposition of the χ^2 metric leads us to conclude that if the systematic uncertainty component of the DCM is estimated to be proportional to the measured data, then there is a value of the uncertainty on the data reduction parameter which will result in the false solution mode dominating the global objective surface. The corollary is even more striking! Notice that an increase in the number of observed data points will increase the value of $||\vec{d}||_2^2$ proportional to N. On the right hand side of Equation 39, the term on the right will be less than the term on the left so long as the elements d_i are generally larger than $\max(\vec{\delta^2})$, which is roughly equal to the data point's corresponding stochastic uncertainty, δ_i, per the assumption in Equation 40. As both terms scale inversely with N, so should the right hand side, leading to a decrease in the value of Δn^2 necessary to meet the condition of Equation 39 and for the false mode to become the global minimum. Under the same conditions, for any non-zero uncertainty on the data reduction parameter, with enough data points, the PPP phenomenon will occur and the false solution mode will emerge as the global minimum!

There is not a requirement that the right hand side of Equation 39 be positive, the regime where this could occur is if a significant number of the data points have a stochastic uncertainty more than 100%. From the perspective of the correlation coefficient in the DCM (*true* or *estimated*), the the PPP false-mode phenomenon is predicted to occur even as the correlation coefficient gets arbitrarily close to zero, given that there are enough data points, N. However, the systematic uncertainty cannot be exactly 0 because the derivation of the escaped eigenvalue and corresponding eigenvector would not hold.

Extension to multiple data reduction parameters
Additional data reduction parameters adding to the systematic uncertainty are also subject to the Cauchy interlacing theorem. Therefore, if we have M data reduction parameters, at most M eigenvalues can separate out of the *bulk*. If the data reduction parameters are correlated, then the preceding discussion can be translated to independent linear combinations (another eigenvalue decomposition) of the data reduction parameters. If we have multiple data reduction parameters, consider adding the systematic components of uncertainty to the stochastic (diagonal) DCM, one at a time, as a series of rank-1 perturbations.

As far as we understand, there is no closed-form prediction for how the eigenvectors change upon further rank-1 additions to the DCM for other data reduction parameters. That is, we cannot analytically prove that $\vec{q}_{\max}$ will align with $\vec{d}$ as in Equation 23. However, we believe that the intuition provided by the observation that the first perturbation to the diagonal DCM produces an eigenvector aligned with the data will still be valid upon subsequent perturbations. This is supported by limited empirical evidence later in Section 5.2.

Solutions
IRLS: A dynamic estimate of the DCM
One proposed resolution to the PPP phenomenon, originally in Ref.[3], that is widely accepted in the nuclear data field has been to change the estimator of the DCM from being based on the measured data to based on the current best estimator of the mean, $\hat{\vec{\mu}}$,

$$\hat{\Sigma}_{\text{fit}} = \text{diag}(\vec{\delta^2}) + \hat{\vec{\mu}} (\Delta n)^2 \hat{\vec{\mu}}^T. \tag{41}$$

Plugging this DCM into Equation 13, the MLE becomes a function of the estimator itself. This requires an algorithm known in statistics fields as IRLS which iteratively updates $\hat{\Sigma}_{\text{fit}}$ and asymptotically converges to the unbiased MLE[15,29]. With this DCM, $\vec{d}$ in Eqs. 20 and 23 (defining the separated eigenvalue/vector) is replaced with $\hat{\vec{\mu}}$. This means that Eq. 25 becomes

$$\vec{q}_{\max} \approx \frac{\hat{\vec{\mu}}}{||\hat{\vec{\mu}}||_2}, \tag{42}$$

the left and right terms on the right hand side of Eq. 26 are no longer orthogonal, and the false-minimum at $\hat{\vec{\mu}} = 0$ no longer appears.

Cross-validation and IRLS

Here, we address another challenge that PPP presents. Namely, if the mechanism to avoid the false solution mode induced by the data-based estimator of the DCM is to use IRLS, then how can one successfully do cross-validation for correlated data?

Recent work in nuclear resonance evaluation uses cross validation to determine the number of resonances in a given energy range[20]. In cross-validation, the entirety of the data is separated, ahead of time, into independent training and validation subsets. For example, 80% of the data is selected for training and 20% of the data is held back for validation. The independence of training and validation sets is vital; if the observational data are correlated, data splitting can be done along the independent principal components of the data (eigenvalue decomposition of the DCM) in a process often called pre-whitening[30]. An issue arises upon the implementation of IRLS, if the DCM is now to be estimated based on the current fit and the fit continues to change, then how could one do the initial splitting of the data into independent subsets for training and validation?

One solution is to stratify correlated data into train and validation parts in a naïve manner and correct for correlation in the cross validation score. In cross validation, it is vital that the cross validation score must be independent of the training objective function, but it does not necessarily mandate that the training and validation data are uncorrelated. This is referred to as cross-validation on non-factorized models and is a common method within Gaussian Process regression[31]. Suppose the experimental data is split into training ($\vec{d}_{\text{tr}}$) and validation ($\vec{d}_{\text{va}}$) data sets. Assuming normality, the experimental data will follow the normal distribution:

$$\begin{bmatrix} \vec{D}_{\text{tr}} \\ \vec{D}_{\text{va}} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \vec{\mu}_{\text{tr}}^{\text{true}} \\ \vec{\mu}_{\text{va}}^{\text{true}} \end{bmatrix}, \begin{bmatrix} \Sigma_{\text{tr,tr}}^{\text{true}} & \Sigma_{\text{tr,va}}^{\text{true}} \\ \Sigma_{\text{va,tr}}^{\text{true}} & \Sigma_{\text{va,va}}^{\text{true}} \end{bmatrix} \right) \tag{43}$$

The model is fit to experimental training data, d_{tr}, finding a mean of $\hat{\vec{\mu}}_{\text{tr}}$ and a DCM estimate, $\hat{\Sigma}_{\text{tr,tr}}(\hat{\vec{\mu}}_{\text{tr}})$, according to IRLS. The fit provides an estimate on the validation data, $\hat{\vec{\mu}}_{\text{va}}$. The cross validation chi-squared, χ^2_{CV}, can be calculated as follows:

$$\chi^2_{\text{CV}} = (\vec{d}_{\text{eff}} - \hat{\vec{\mu}}_{\text{va}})^T \hat{\Sigma}_{\text{eff}}^{-1} (\vec{d}_{\text{eff}} - \hat{\vec{\mu}}_{\text{va}}) \tag{44}$$

$$\vec{d}_{\text{eff}} = \vec{d}_{\text{va}} - \hat{\Sigma}_{\text{va,tr}} \hat{\Sigma}_{\text{tr,tr}}^{-1} (\vec{d}_{\text{tr}} - \hat{\vec{\mu}}_{\text{tr}}) \tag{45}$$

$$\hat{\Sigma}_{\text{eff}} = \hat{\Sigma}_{\text{va,va}} - \hat{\Sigma}_{\text{va,tr}} \hat{\Sigma}_{\text{tr,tr}}^{-1} \hat{\Sigma}_{\text{tr,va}} \tag{46}$$

Just as $\Pr(d_{\text{tr}}) \propto \exp\left(-\frac{1}{2}\chi^2_{\text{tr}}\right)$, where $\chi^2_{\text{tr}} = (\vec{d}_{\text{tr}} - \vec{\mu}_{\text{tr}})^T \hat{\Sigma}_{\text{tr,tr}}^{-1} (\vec{d}_{\text{tr}} - \vec{\mu}_{\text{tr}})$, we find $\Pr(d_{\text{va}}|d_{\text{tr}}) \propto \exp\left(-\frac{1}{2}\chi^2_{\text{CV}}\right)$. With a correctly specified covariance matrix (i.e. $\hat{\Sigma} = \Sigma^{\text{true}}$), χ^2_{CV} is uncorrelated with the training data. Consider the matrix relationship

$$\begin{bmatrix} \vec{d}_{\text{tr}} - \hat{\vec{\mu}}_{\text{tr}} \\ \vec{d}_{\text{eff}} - \hat{\vec{\mu}}_{\text{va}} \end{bmatrix} = \begin{bmatrix} I & 0 \\ -\Sigma_{\text{va,tr}}^{\text{true}} \left(\Sigma_{\text{tr,tr}}^{\text{true}}\right)^{-1} & I \end{bmatrix} \begin{bmatrix} \vec{d}_{\text{tr}} - \hat{\vec{\mu}}_{\text{tr}} \\ \vec{d}_{\text{va}} - \hat{\vec{\mu}}_{\text{va}} \end{bmatrix}. \tag{47}$$

Using error propagation on equation 43, assuming $\hat{\vec{\mu}}_{\text{tr}}$ and $\hat{\vec{\mu}}_{\text{va}}$ estimate $\vec{\mu}_{\text{tr}}^{\text{true}}$ and $\vec{\mu}_{\text{va}}^{\text{true}}$, and $\hat{\Sigma} = \Sigma^{\text{true}}$, one finds the covariance matrix

$$\text{Cov}\left(\begin{bmatrix} \vec{d}_{\text{tr}} \\ \vec{d}_{\text{eff}} \end{bmatrix} \right) = \begin{bmatrix} \hat{\Sigma}_{\text{tr,tr}} & 0 \\ 0 & \hat{\Sigma}_{\text{eff}} \end{bmatrix}. \tag{48}$$

$\vec{d}_{\text{eff}}$ is independent of the training data, $\vec{d}_{\text{tr}}$, and has a covariance of $\hat{\Sigma}_{\text{eff}}$. Therefore, the cross validation chi-square goodness of fit is calculated as written in equation 44 and is independent on the training data. In practice, Σ^{true} is unknown and estimated with $\hat{\Sigma}$, resulting in correlation on the order of the quality of the DCM estimate. Poor DCM estimates may come from improper uncertainty estimation or – in the case of IRLS – poor estimate on the fit, $\hat{\mu}$.

Numerical results
Demonstration on linear model

Consider a simplified numerical example involving a linear regression model with two parameters. We generate N–dimensional vector samples, $\vec{d}$, from $\vec{d}$ defined in Eq. 7 with

$$\vec{\mu}_{\text{true}} = X\vec{\sigma}_{\text{true}}, \tag{49}$$

$$\vec{\sigma}_{\text{true}} = \begin{bmatrix} 1.0 \\ 5.0 \end{bmatrix}, \tag{50}$$

where X is the design matrix defined as

$$X = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{bmatrix} \in \mathbb{R}^{N \times 2}, \tag{51}$$

and x_i represents the independent variable for observation i and the vector $\mathrm{x} = [x_1, x_2, \ldots, x_N]^\top \in \mathbb{R}^N$ we define to be N–linearly spaced values between 0 and 10. That is,

$$x_i = 1 + \frac{(i-1)}{N-1} \cdot (10 - 1), \quad i = 1, \ldots, N. \tag{52}$$

For the numerical demonstration, we no longer need to assume the specific case that the spread of the bulk of eigenvalues is 0 (as in Eq. 17), and instead can investigate the more realistic case where $\vec{\delta} = \delta_0 \vec{d}$. The true DCM is then given by

$$\Sigma_{\text{true}} = \delta_0^2 \text{diag}(\vec{\mu}_{\text{true}}) + \vec{\mu}_{\text{true}} \, (\Delta n)^2 \, \vec{\mu}_{\text{true}}^T, \tag{53}$$

and the estimated DCM for any one sample, $\bar{d}$, is given by

$$\hat{\Sigma} = \delta_0^2 \text{diag}(\vec{d}) + \vec{d} \, (\Delta n)^2 \, \vec{d}^T. \tag{54}$$

The frequentist interpretation allows us to draw a large number of data samples (10000) and visualize the distribution of estimated values ($\hat{\vec{\sigma}}$) using the estimated versus true DCM. This is shown in Fig. 1 for a varying number of observed data points, N, with $\delta_0 = 0.1$ and $\Delta n = 0.2$ resembling the original PPP setup.

We consider the estimates using the true DCM to be what we want to estimate for any given data sample, effectively the ground truth for validation. The bias from the estimated DCM is characteristic of the PPP phenomenon, and the increase in bias as N increases highlights the relationship derived in Sec. The eigendecomposition explanation of the false mode.

Figure 2 shows the eigenvalue spectrum and dominant eigenvector for a sample from the $N = 20$ case. We observe the separation of a single eigenvalue from the bulk in the left figure. In the right figure, we see that the corresponding, dominant eigenvector of the estimated DCM is aligned very closely with the normalized data vector while that of the true DCM seems to be aligned with the model.

Extension to neutron transmission data

Transmission data are the least likely neutron time-of-flight data to suffer from the PPP bias and bias from PPP effects are often not suspected by evaluators. A major reason is that transmission is a ratio of two measurements
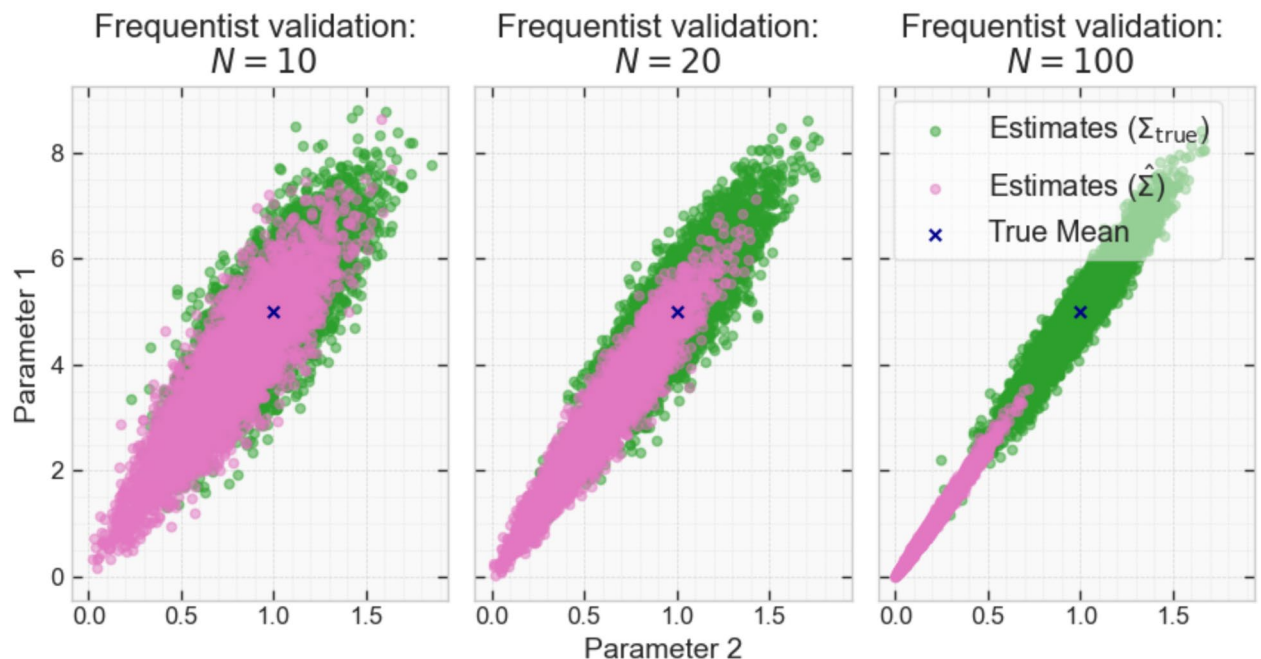
Fig. 1. Frequentist validation approach for GLS estimates using the true versus estimated DCM. The former, in green (color online), is considered the ground truth and is seen to always be centered around the true mean. The PPP bias towards zero is demonstrated by using the estimated DCM, shown in pink (color online). This bias is shown to increase with an increase in observed data, N.
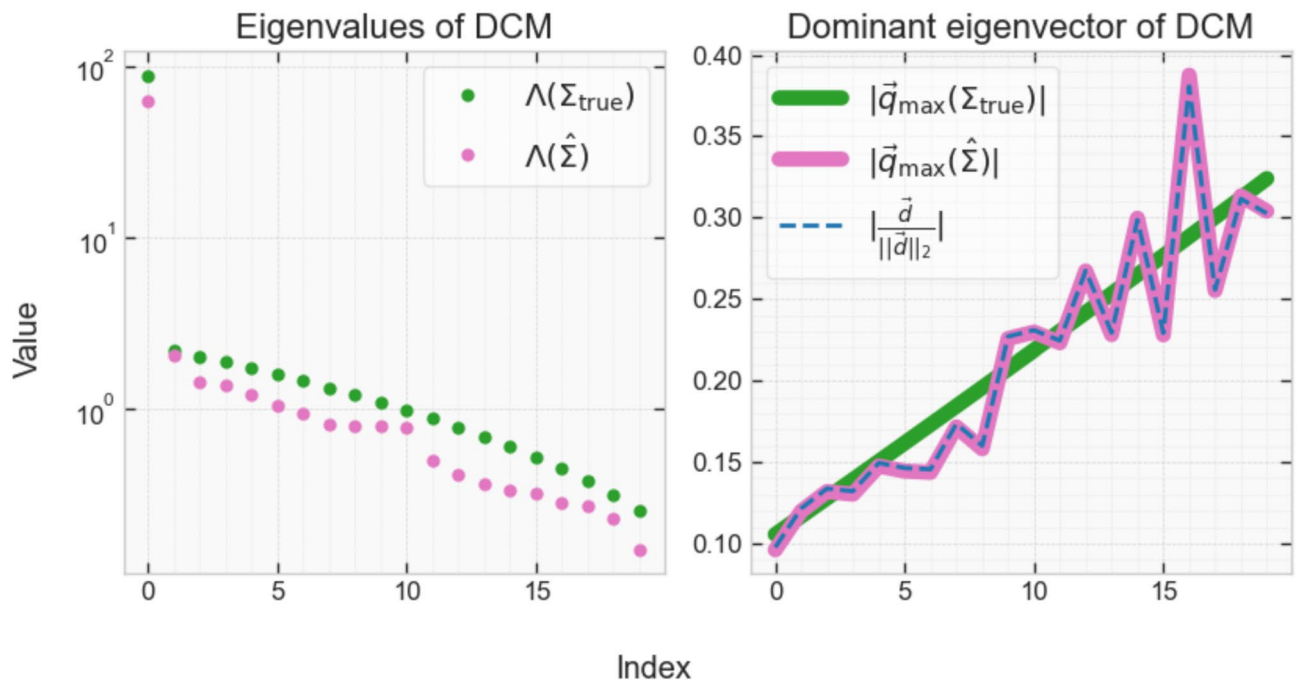
Fig. 2. Eigen-decomposition analysis of PPP phenomenon for a linear model. The left figure shows the eigenvalues of the true and estimated DCM with one separated from the bulk. The right figure shows the eigenvector corresponding to the separated/dominating eigenvalue for both the true and estimated DCM, the latter of which is shown to align closely with the normalized data vector.

and thus does not require an absolute flux normalization, the reduction parameter most often associated with the PPP bias. Instead, small differences in the flux between in-ratio measurements are corrected for with flux monitors. Additionally, the in-ratio measurements are often cycled tens of times to further minimize differences. As a result, the correlating uncertainty from this correction is generally very small (1-2% or 2-6% with or without cycling)[22].

The ^{181}Ta transmission measurements by Brown, et al.[32] would not be suspected for PPP; the data are not discrepant, the correlated uncertainties are minimized, and there is no overall normalization (i.e., $\hat{\Sigma}_{\text{sys}} \neq \vec{d}(\Delta n)^2 \vec{d}^T$). Instead $\hat{\Sigma}_{\text{sys}} = J(\Delta n)^2 J^T$ where J is a more general Jacobian describing the derivative of the reduced data with respect to the measured/raw data). A PPP false mode was discovered in this data and partially inspired the investigation in this article. Its existence—shown in the following figure/table—demonstrates that the intuitions given by the simplified analytic derivation in Sec The frequentist interpretation still hold for real data where (a) there are more than one data reduction parameters (10 in this case) and (b) the systematic uncertainty is not an overall normalization.

Figure 3 shows experimental transmission, the ENDF/B-VIII.0 evaluation[33], and two candidate models, labeled Fit A and Fit B. Table 1 shows the χ^2 objective for each model when the DCM is calculated using the data ($\hat{\Sigma}_{\text{data}}$ as in Equation 18) and using the fit ($\hat{\Sigma}_{\text{fit}}$ as in Equation 41). The latter corresponds to the converged IRLS estimate of the DCM as, for real data, the *true* DCM requires knowledge of the true mean and is therefore not accessible.

Visually, Fit A and the ENDF/B-VIII.0 evaluation follow the data well while Fit B does not. However, evaluating the models with a χ^2 objective that uses $\hat{\Sigma}_{\text{data}}$ indicates that Fit B is the best. In fact, $\chi^2 << N_{\text{data}}$ for Fit B indicates that it is too good to be true and is overfitting the data. Meanwhile, the other two models have $\chi^2 >> N_{\text{data}}$, indicating that they explain the data very poorly. If instead the models are evaluated with a χ^2 objective that uses $\hat{\Sigma}_{\text{fit}}$, the χ^2 values agree more with with what we expect: that Fit A and ENDF/B-VIII.0 explain the data much better than Fit B and none of the fits are fully or over-explaining the data as $\chi^2 > N_{\text{data}}$ for all. The false mode in the objective $\chi^2(\hat{\Sigma}_{\text{data}})$ is explained by the eigen-decomposition of $\hat{\Sigma}_{\text{data}}$ shown in Fig. 4. Two out of 10 possible eigenvalues have separated from the bulk, one of which is strongly aligned with the data while the other is less so. In this case, it is the second strongest eigenvalue that causes the false mode.

This example highlights the false mode candidate models Fit A and B lie at the global minimum of $\chi^2(\hat{\Sigma}_{\text{fit}})$ and $\chi^2(\hat{\Sigma}_{\text{data}})$ respectively. The models were produced by global optimization of the χ^2 objective for MLE (Equation 12). During the optimization, the DCM was calculated using $\hat{\Sigma}_{\text{fit}}$ for Fit A and $\hat{\Sigma}_{\text{data}}$ for Fit B. In both cases, the starting point was the ENDF/B-VIII.0 evaluation. The SAMMY resonance evaluation code[9] was used for these calculations along with global optimization methods detailed in Ref.[20]. In both cases, the optimization required iteration to traverse the non-linear objective surface. In the case of $\chi^2(\hat{\Sigma}_{\text{fit}})$, $\hat{\Sigma}_{\text{fit}}$ was updated at each iteration using the fit from the previous step. This means that the objective surface changes as the optimization proceeds and that both the model and the DCM converge simultaneously.
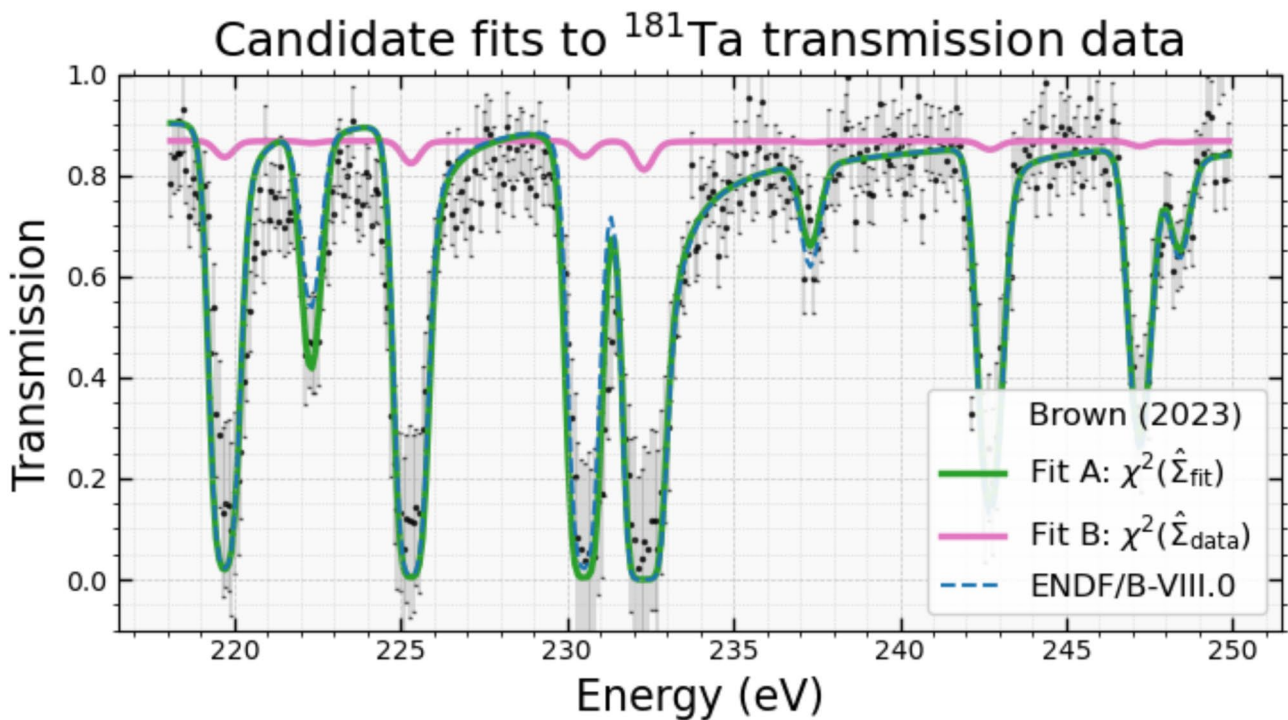
Fig. 3. Experimental transmission data from Ref.[32] along with the ENDF/B-VIII.0 evaluation and two candidate models. The two candidate models are labeled Fit A and Fit B as they represent the global minima of the objective functions $\chi^2(\hat{\Sigma}_{\text{fit}})$ and $\chi^2(\hat{\Sigma}_{\text{data}})$ respectively. The former corresponds to the IRLS estimator and the latter corresponds to standard GLS.

	χ^2	
Fit	$\hat{\Sigma}_{\text{data}}$	$\hat{\Sigma}_{\text{fit}}$
A	1125.5	502.9
B	153.6	2243.0
ENDF/B-VIII.0	1373.1	525.1

Table 1. Objective function values for different models using a DCM estimated at the data versus at the model fit. The number of experimental data points $N_{\text{data}} = 316$ and the number of model parameters $N_{\text{par}} = 24$

This is not to say that any SAMMY fit will land in the false mode if $\hat{\Sigma}_{\text{data}}$ is used; as shown in Ref.[34], a local minimum can exist around a more reasonable result. Still, relying on this local minima is not advisable as it may be shallow, it may not agree with the global minima using $\hat{\Sigma}_{\text{fit}}$, and the magnitude of the χ^2 metric can be misleading.

Conclusions

In this article, we revisit the well-studied PPP phenomenon through a new lens. The eigenspectrum analysis of the incorrect DCM is revealing of the underlying cause of the PPP bias while the generative model allows us to compare against a ground truth. The primary contribution of this article is simply a new way to look at the problem; however, we recognize a few nuanced, novel contributions.

Considering the assumptions/regimes explored in Section The frequentist interpretation gives us an intuition about how and where this can show up as well as what features influence the bias. The intuition is subsequently supported by numerical examples which lead to the conjecture that the PPP bias in GLS estimates can show up for any experimental neutron time-of-flight data, regardless of the data quality or functional form of the systematic errors. This is somewhat contrary to much of the literature where PPP is discussed in the context of relative normalization errors and strongly correlated, discrepant data. It is also commonly found in literature that the PPP bias comes about in the regime of small stochastic and large systematic errors. A nuanced understanding that follows from the eigenspectrum analysis is that the overall magnitude of the stochastic error on the data points, $\vec{\delta}$, ultimately does not influence the bias; instead, it is the range of the stochastic errors that matter. Furthermore, it was shown how stochastic errors interplay with systematic errors and dimensionality to influence the bias.
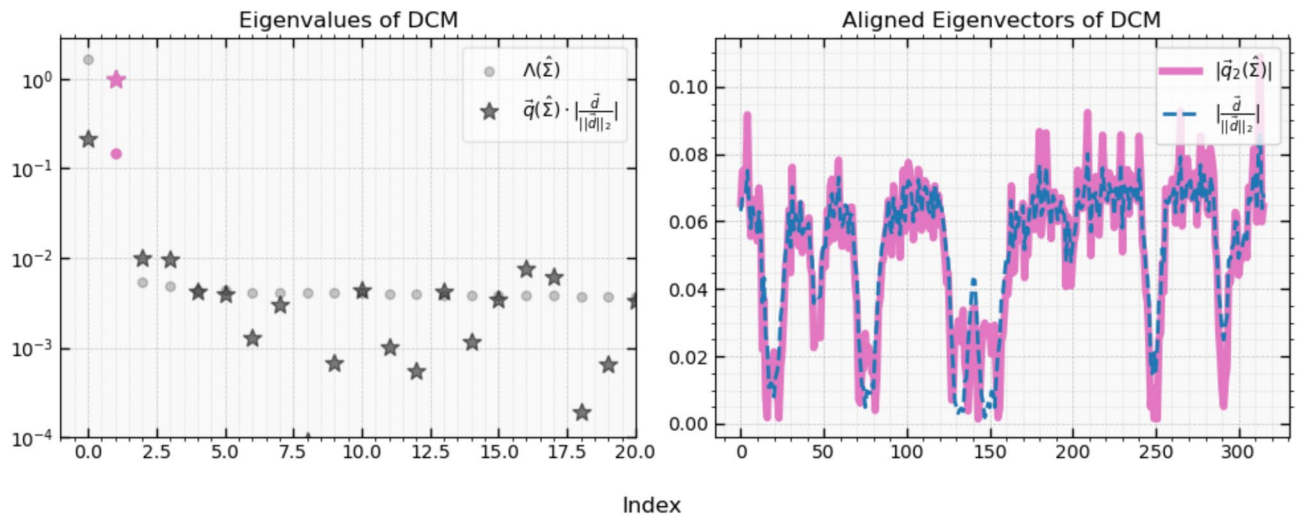
Fig. 4. Eigen-decomposition analysis of PPP phenomenon for a neutron transmission data. The left figure shows the eigenvalues of the true and estimated DCM with two separated from the bulk. The right figure shows the eigenvector corresponding to the separated/dominating eigenvalue for both the true and estimated DCM, the latter of which is shown to align closely with the normalized data vector.

We want to emphasize that, while it seems a bit extreme, the neutron transmission example given in Section Extension to neutron transmission data comes about from using the DCM as reported. This example highlights a number of misconceptions about the PPP bias in that it does not have a relative normalization uncertainty, the data are not very strongly correlated, and neither do they seem discrepant. In this case, the PPP bias comes about as a false global minimum far from the data. In fact, there is a shallow, local minimum closer to the data that a global optimization algorithm can easily escape. We expect that the false solution often goes unnoticed because of this shallow minimum, especially as analysts often start from a previous evaluation that is close to the data and likely are not performing global optimization. Additionally, the local minimum may be made more stable with the introduction of other experimental data. We note that even though local minima in the GLS estimator may exist close to the data, it still has the possibility to be biased and the IRLS DCM should always be used since it is known to give unbiased MLE estimates. The fact that the DCM as reported is almost always evaluated at the data presents another issue: IRLS can only be implemented if the systematic uncertainty is a simple normalization factor. If more complicated data reduction parameters are used, then proper implementation of IRLS requires the individual components of the DCM and the functional relationship to the observable. In many cases, this information is not published alongside experimental nuclear data.

Lastly, we present a challenge that the IRLS solution to the PPP bias presents to statistical methods, such as cross validation, that leverage independent/orthogonal subsets of data and discuss a potential solution. The challenge is that orthogonal components of systematically correlated data require linearly–independent principle components of the DCM and IRLS proposes that the DCM changes based on the current estimator. In Section Cross-Validation and IRLS we derive a correction to the cross validation score that accounts for this.

Data availability

The linear model data generated and analyzed are fully described in the methods section and can be made available upon request. The experimental data set analyzed in this article is openly available in the EXFOR database[35]. For questions or data requests, contact the corresponding author Noah A.W. Walton.

References

1. Peelle, R. *Peelle's Pertinent Puzzle, Informal memorendum dated October 13, 1987* (ORNL, USA, 1987).
2. Holland, P. W. & Welsch, R. E. Robust regression using iteratively reweighted least-squares. *Communications in Statistics - Theory and Methods* **6**, 813–827. https://doi.org/10.1080/03610927708827533 (1977).
3. Chiba, S. & Smith, D. L. 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. Tech. Rep. ANL/NDM-121, Argonne National Laboratory (1991).
4. Draper, N. R. & Smith, H. *Applied Regression Analysis.* (John Wiley & Sons, Inc, 1998), 3rd edn.
5. Zhao, Z. & Perey, F. G. The Covariance Matrix of Derived Quantities and Their Combination. Tech. Rep. ORNL/TM-12106, Oak Ridge National Laboratory (1992).
6. Chiba, S. & Smith, D. L. Impacts of data transformations on least-squares solutions and their significance in data analysis and evaluation. *Journal of Nuclear Science and Technology* **31**, 770–781. https://doi.org/10.1080/18811248.1994.9735223 (1994).
7. Frohner, F. H. Assigning Uncertainties to Scientific Data. *Nuclear Science and Engineering* **126**, 1–18, https://doi.org/10.13182/NSE97-A24453 (1997).

8. Frohner, F. H. Evaluation and Analysis of Nuclear Resonance Data. Tech. Rep. JEFF Report 18, OECD (2000).
9. Larson, N. M. Updated users' guide for SAMMY: Multilevel R-matrix fits to neutron data using Bayes' equations. Tech. Rep. ORNL/TM-9179/R8, ORNL, ORNL, Oak Ridge, TN (2008).
10. Hanson, K. M., Kawano, T. & Talou, P. Probabilistic interpretation of Peelle's pertinent puzzle and its resolution. *AIP Conference Proceedings* **769**, 304–307. https://doi.org/10.1063/1.1945011 (2005).
11. Frühwirth, R., Neudecker, D. & Leeb, H. Peelle's Pertinent Puzzle and its solution. *EPJ Web of Conferences* **27**, 1–6, https://doi.org/10.1051/epjconf/20122700008 (2012).
12. Neudecker, D., Frühwirth, R. & Leeb, H. Peelle's pertinent puzzle: A fake due to improper analysis. *Nuclear Science and Engineering* **170**, 54–60, https://doi.org/10.13182/NSE11-20 (2012).
13. Neudecker, D., Frühwirth, R., Kawano, T. & Leeb, H. Adequate Treatment of Correlated Experimental Data in Nuclear Data Evaluations Avoiding Peelle's Pertinent Puzzle. *Nuclear Data Sheets* **118**, 364–366, https://doi.org/10.1016/j.nds.2014.04.081 (2014).
14. Schnabel, G. & Leeb, H. A modified Generalized Least Squares method for large scale nuclear data evaluation. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **841**, 87–96. https://doi.org/10.1016/j.nima.2016.10.006 (2017).
15. Grosskopf, M. PPP and Chiba-Smith is not a fix but IRLS. In *Mini-CSEWG 2004* (2004).
16. Summary Report of the First Research Co-ordination Meeting on Improvement of the Standard Cross Sections for Light Elements. Tech. Rep., International Atomic Energy Agency (2002).
17. Kawano, T. et al. Simultaneous Evaluation of Fission Cross Sections of Uranium and Plutonium Isotopes for JENDL-3.3. *Journal of Nuclear Science and Technology* **37**, 327–334, https://doi.org/10.1080/18811248.2000.9714902 (2000).
18. International Evaluation of Neutron Cross-Section Standards. Tech. Rep., International Atomic Energy Agency, (2007).
19. Carlson, A. D. et al. International Evaluation of Neutron Cross Section Standards. *Nuclear Data Sheets* **110**, 3215–3324. https://doi.org/10.1016/j.nds.2009.11.001 (2009).
20. Walton, N. A. W. et al. Automated resonance fitting for nuclear data evaluation. *Nuclear Science and Engineering* 1–16. https://doi.org/10.1080/00295639.2024.2439700 (2025).
21. Neudecker, D. et al. Templates of expected measurement uncertainties. *EPJ Nuclear Sci. Technol.* **9**, 35. https://doi.org/10.1051/epjn/2023014 (2023).
22. Lewis, A. M. et al. Templates of expected measurement uncertainties for total neutron cross-section observables. *EPJ Nuclear Sci. Technol.* **9**, 34. https://doi.org/10.1051/epjn/2023018 (2023).
23. Lewis, A. M. et al. Templates of expected measurement uncertainties for neutron-induced capture and charged-particle production cross section observables. *EPJ Nuclear Sci. Technol.* **9**, 33. https://doi.org/10.1051/epjn/2023015 (2023).
24. Neudecker, D. et al. Templates of expected measurement uncertainties for prompt fission neutron spectra. *EPJ Nuclear Sci. Technol.* **9**, 32. https://doi.org/10.1051/epjn/2023013 (2023).
25. Thompson, R. The behavior of eigenvalues and singular values under perturbations of restricted rank. *Linear Algebra and its Applications* **13**, 69–78. https://doi.org/10.1016/0024-3795(76)90044-6 (1976).
26. Bunch, J., Nielsen, C. & Sorensen, D. Rank-one modification of the symmetric eigenproblem. *Numerische Mathematik* **31**, 31–48 (1978/79).
27. Hwang, S.-G. Cauchy's interlace theorem for eigenvalues of hermitian matrices. *The American Mathematical Monthly* **111**, 157–159 (2004).
28. Max, A. W. Inverting modified matrices. In *Memorandum Rept. 42, Statistical Research Group*, 4 (Princeton Univ., 1950).
29. McCullagh, P. & Nelder, J. A. *Generalized Linear Models*. Monographs on Statistics and Applied Probability (Chapman & Hall/CRC, London, 1989), 2nd edn.
30. Cao, J., Murata, N., i. Amari, S., Cichocki, A. & Takeda, T. A robust approach to independent component analysis of signals with high-level noise measurements. *IEEE Transactions on Neural Networks* **14**, 631–645, https://doi.org/10.1109/TNN.2002.806648 (2003).
31. Bürkner, P. C., Gabry, J. & Vehtari, A. Efficient leave-one-out cross-validation for bayesian non-factorized normal and student-t models. *Computational Statistics* **36**, 1243–1261, https://doi.org/10.1007/S00180-020-01045-4/FIGURES/4 (2021).
32. Brown, J. M. et al. New measurements to resolve discrepancies in evaluated model parameters of 181-Ta. *Nuclear Science and Engineering* https://doi.org/10.1080/00295639.2023.2249786 (2023).
33. Brown, D. et al. ENDF/B-VIII.0: The 8th Major Release of the Nuclear Reaction Data Library with CIELO-project Cross Sections, New Standards and Thermal Scattering Data. *Nuclear Data Sheets* **148**, 1–142, https://doi.org/10.1016/j.nds.2018.02.001 (2018).
34. Walton, N. A. W. *A Computational Framework for Automated Nuclear Resonance Evaluation and Validation*. Phd dissertation, University of Tennessee (2024).
35. Otuka, N. et al. Towards a more complete and accurate experimental nuclear reaction data library (exfor): International collaboration between nuclear reaction data centres (nrdc). *Nuclear Data Sheets* **120**, 272–276. https://doi.org/10.1016/j.nds.2014.07.065 (2014).

Acknowledgements

Author contributions

All authors contributed to the mathematical derivations. N.W. and W.F. conducted the numerical demonstrations and analyzed the results. A.L. and N.W. conducted the historical review. All authors reviewed the manuscript.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to N.A.W.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.