Some Thoughts on the Data Analysis Process

Nancy Larson

Second Research Co-ordination Meeting on Improvement of the Standard Cross Sections for Light Elements

13-17 October 2003
NIST, Gaithersburg, MD, USA

OAK RIDGE NATIONAL LABORATORY
U. S. DEPARTMENT OF ENERGY

Topics to be ~~covered~~ touched on

- **Peelle's Pertinent Puzzle/Paradox**
- **Implicit data covariance (IDC) methodology**
- **Transformation of variables**

What is an appropriate goal of data analysis?

- **If we make the (rash) assumption that the measured quantities[1] (the raw data) obey Gaussian statistics, then fitting to the measured quantities should give the correct result ("Truth").**
- **The goal of an analysis which fits to derived quantities[2] is then to obtain the same value of "Truth" (or as close to it as possible) .**

[1]In a time-of-flight (tof) measurement this might be counts per time channel.

[2]Cross section per energy, in a tof measurement.

References for PPP

- **Zhao and Perey**
  - **"The Covariance Matrix of Derived Quantities and Their Combination"**
  - **ORNL/TM-12106 (1992)**

- **Chiba and Smith**
  - **"Some comments on Peelle's Pertinent Puzzle"**
  - **JAERI-M 94-068 page 5-12 (1994)**

Solution to PPP ?

- **Least-squares equations* are a linear expansion of a non-linear problem**
  - **Everywhere throughout the equations, the expansion must be made with respect to the same estimate of the value of any given parameter**
  - **Data covariance matrix is generally derived assuming different estimates for the same parameter** ➡ ⬅

*** or Bayes' Equations**

A trivial example

- **Suppose the function *Y* is given by the product of two other (nonlinear) functions, *f* and *g***

$$Y(x) = f(x)\, g(x)$$

- **Suppose further that, when *f* was measured, the value of *x* was known to be approximately *a*. Therefore, for $x \approx a$, *f* can be expanded in a Taylor series to give**

$$f(x) \approx f(a) + f_x\,(x-a)$$

A trivial example, cont. ($Y = f\,g$)

- **Suppose also that, when *g* was measured, the value of *x* was known to be ≈*b*. Hence, for *x* close to *b*, *g* can be expanded as**

$$g(x) \approx g(b) + g_x(x-b)$$

- **Therefore *Y* might be written**

$$Y(x) = f(x)\,g(x) \approx [f(a) + f_x(x-a)\,]\,[g(b) + g_x\,(x-b)\,]$$

- **Where is this equation valid?**
  - **For $x \approx a$?**
  - **For $x \approx b$?**

No where! unless $a \approx b$

How does this apply to PPP ?

- **Let d_1 and d_2 represent two (uncorrelated) measurements of the same quantity, with uncertainties Δd_1 and Δd_2 respectively.**
- **Let n represent the normalization (identical for the two measurements), and Δn its uncertainty.**
- **Let P represent the parameter of interest; P is related to the measured quantity via $P = d/n$.**

PPP, continued

- **To fit to the raw data, one way to formulate Bayes' Equations is**

$$\begin{bmatrix} P' \\ n' \end{bmatrix} = \begin{bmatrix} P \\ n \end{bmatrix} - M' \boxed{G}^{t} \boxed{V}^{-1} \begin{bmatrix} nP - d_1 \\ nP - d_2 \end{bmatrix}$$

where the partial derivative matrix *G* and data covariance matrix *V* are given by

$$\boxed{G} = \begin{bmatrix} n & P \\ n & P \end{bmatrix} \qquad \boxed{V} = \begin{bmatrix} \Delta^2 d_1 & 0 \\ 0 & \Delta^2 d_2 \end{bmatrix}$$

PPP, continued (fit raw data)

The (final) parameter covariance matrix *M'* is found from

$$M' = (M^{-1} + W)^{-1}$$

with *W* defined by $W = G^t V^{-1} G$

and the (initial) parameter covariance matrix *M* by

$$M = \begin{bmatrix} \Delta^2 P & 0 \\ 0 & \Delta^2 n \end{bmatrix} = \begin{bmatrix} \infty & 0 \\ 0 & \Delta^2 n \end{bmatrix} \qquad so \quad M^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \Delta^{-2} n \end{bmatrix}$$

$$\begin{bmatrix} P' \\ n' \end{bmatrix} = \begin{bmatrix} P \\ n \end{bmatrix} - M'Y \qquad Y = G^t V^{-1} \begin{bmatrix} nP - d_1 \\ nP - d_2 \end{bmatrix} \qquad M^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & \Delta^{-2} n \end{bmatrix} \qquad G = \begin{bmatrix} n & P \\ n & P \end{bmatrix} \qquad V = \begin{bmatrix} \Delta^2 d_1 & 0 \\ 0 & \Delta^2 d_2 \end{bmatrix}$$

$$G^t V^{-1} = \begin{bmatrix} n & n \\ P & P \end{bmatrix} \begin{bmatrix} \Delta^{-2} d_1 & 0 \\ 0 & \Delta^{-2} d_2 \end{bmatrix} = \begin{bmatrix} n\,\Delta^{-2} d_1 & n\,\Delta^{-2} d_2 \\ P\,\Delta^{-2} d_1 & P\,\Delta^{-2} d_2 \end{bmatrix}$$

$$Y = G^t V^{-1} \begin{bmatrix} nP - d_1 \\ nP - d_2 \end{bmatrix} = \begin{bmatrix} n\,\Delta^{-2} d_1 & n\,\Delta^{-2} d_2 \\ P\,\Delta^{-2} d_1 & P\,\Delta^{-2} d_2 \end{bmatrix} \begin{bmatrix} nP - d_1 \\ nP - d_2 \end{bmatrix} = \begin{bmatrix} n^2 P\,(\Delta^{-2} d_1 + \Delta^{-2} d_2) - n\,(d_1 \Delta^{-2} d_1 + d_2 \Delta^{-2} d_2) \\ nP^2\,(\Delta^{-2} d_1 + \Delta^{-2} d_2) - P\,(d_1 \Delta^{-2} d_1 + d_2 \Delta^{-2} d_2) \end{bmatrix} = \begin{bmatrix} n\,(nP\,Q - R) \\ P\,(nPQ - R) \end{bmatrix}$$

$$W = G^t V^{-1} G = \begin{bmatrix} n\,\Delta^{-2} d_1 & n\,\Delta^{-2} d_2 \\ P\,\Delta^{-2} d_1 & P\,\Delta^{-2} d_2 \end{bmatrix} \begin{bmatrix} n & P \\ n & P \end{bmatrix} = \begin{bmatrix} n^2\,(\Delta^{-2} d_1 + \Delta^{-2} d_2) & nP\,(\Delta^{-2} d_1 + \Delta^{-2} d_2) \\ nP\,(\Delta^{-2} d_1 + \Delta^{-2} d_2) & P^2\,(\Delta^{-2} d_1 + \Delta^{-2} d_2) \end{bmatrix}$$

$$M^{-1} + W = \begin{bmatrix} n^2\,(\Delta^{-2} d_1 + \Delta^{-2} d_2) & nP\,(\Delta^{-2} d_1 + \Delta^{-2} d_2) \\ nP\,(\Delta^{-2} d_1 + \Delta^{-2} d_2) & P^2\,(\Delta^{-2} d_1 + \Delta^{-2} d_2) + \Delta^{-2} n \end{bmatrix} = \begin{bmatrix} n^2 Q & nPQ \\ nPQ & P^2 Q + \Delta^{-2} n \end{bmatrix}$$

$$\boxed{\begin{aligned} Q &= \Delta^{-2} d_1 + \Delta^{-2} d_2 \\ R &= d_1 \Delta^{-2} d_1 + d_2 \Delta^{-2} d_2 \end{aligned}}$$

$$M' = (M^{-1} + W)^{-1} = \left(n^2 Q\,\Delta^{-2} n\right)^{-1} \begin{bmatrix} P^2 Q + \Delta^{-2} n & -nPQ \\ -nPQ & n^2 Q \end{bmatrix} = \begin{bmatrix} \dfrac{P^2}{n^2} \Delta^2 n + \dfrac{1}{n^2 Q} & -\dfrac{P}{n} \Delta^2 n \\ -\dfrac{P}{n} \Delta^2 n & \Delta^2 n \end{bmatrix}$$

$$\begin{bmatrix} P' \\ n' \end{bmatrix} = \begin{bmatrix} P \\ n \end{bmatrix} - \left(n^2 Q\,\Delta^{-2} n\right)^{-1} \begin{bmatrix} P^2 Q + \Delta^{-2} n & -nPQ \\ -nPQ & n^2 Q \end{bmatrix} \begin{bmatrix} n(nPQ - R) \\ P(nPQ - R) \end{bmatrix} =$$

$$\left(n^2 Q\,\Delta^{-2} n\right)^{-1} \begin{bmatrix} n^2 PQ\,\Delta^{-2} n - nP^2 Q(nPQ - R) - n\Delta^{-2} n\,(nPQ - R) + nP^2 Q(nPQ - R) \\ n^3 Q\,\Delta^{-2} n + n^2 PQ(nPQ - R) - n^2 PQ(nPQ - R) \end{bmatrix} = \begin{bmatrix} R/(n^2 Q) \\ n \end{bmatrix}$$

PPP, continued (fit raw data)

- **After a bit of algebra, results are**

$$P' = \frac{\left(\frac{d_1}{\Delta^2 d_1} + \frac{d_2}{\Delta^2 d_2}\right)}{n\left(\frac{1}{\Delta^2 d_1} + \frac{1}{\Delta^2 d_2}\right)}$$

(as expected)

and

$$n' = n$$

PPP, continued (fit raw data)

$\boxed{\Delta^2 P'}$

- **New parameter covariance matrix is**

$$M' = \begin{bmatrix} \boxed{\frac{1}{n^2}\left(\frac{1}{\Delta^2 d_1}+\frac{1}{\Delta^2 d_2}\right)^{-1} + \frac{\boxed{P^2}}{n^2}\Delta^2 n} & -\frac{P}{n}\Delta^2 n \\ -\frac{P}{n}\Delta^2 n & \Delta^2 n \end{bmatrix}$$

- **Note that the updated parameter value *P'* does not depend on starting value *P*, but uncertainty does depend on $\boxed{P}$.**

Therefore we would iterate, and the result would be to write Δ*P'* in terms of *P'* rather than *P*

- **(Both value and uncertainty depend upon *n*.)**

Results for fitting to raw data –

$$P' = \frac{\left(\frac{d_1}{\Delta^2 d_1} + \frac{d_2}{\Delta^2 d_2}\right)}{n\left(\frac{1}{\Delta^2 d_1} + \frac{1}{\Delta^2 d_2}\right)}$$

"Truth"

$$\Delta^2 P' = \frac{1}{n^2}\left(\frac{1}{\Delta^2 d_1} + \frac{1}{\Delta^2 d_2}\right)^{-1} + \frac{P'^2}{n^2}\Delta^2 n$$

PPP, continued

- **So much for fitting to raw data**
- **Usually we must fit to reduced data**
  - **So how do the equations change ?**

PPP, continued

- **To fit to the reduced data, set $D_i = d_i / n$.**
- **Bayes' Equations then take the form**

$$[P'] = [P] - M' G^t V^{-1} \begin{bmatrix} P - D_1 \\ P - D_2 \end{bmatrix}$$

where the usual expressions for *V* and *G* are

$$V = \begin{bmatrix} \frac{\Delta^2 d_1}{n^2} + \frac{d_1^{\,2}}{n^2}\frac{\Delta^2 n}{n^2} & \frac{d_1 d_2}{n^2}\frac{\Delta^2 n}{n^2} \\ \frac{d_1 d_2}{n^2}\frac{\Delta^2 n}{n^2} & \frac{\Delta^2 d_2}{n^2} + \frac{d_2^{\,2}}{n^2}\frac{\Delta^2 n}{n^2} \end{bmatrix}$$

statistical

$$G = [1 \;\; 1]$$

systematic (common)

PPP, continued (fit reduced data, usual method)

The (final) parameter covariance matrix *M'* is found from

$$M' = (M^{-1} + W)^{-1}$$

with *W* defined by $W = G^t V^{-1} G$

and the (initial) parameter covariance matrix *M* by

$$M = \left[\Delta^2 P\right] = \left[\infty\right] \qquad so \quad M^{-1} = \left[0\right]$$

$$[P'] = [P] - M'Y \qquad Y = G^t V^{-1}\begin{bmatrix} P - d_1/n \\ P - d_2/n \end{bmatrix} \qquad G = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad V = \begin{bmatrix} a^2 + x^2 q & xyq \\ xyq & b^2 + y^2 q \end{bmatrix} \qquad q = \Delta^2 n / n^2$$

$$\boxed{\begin{aligned} a^2 &= \Delta^2 d_1 / n^2 \\ x &= d_1 / n \\ b^2 &= \Delta^2 d_2 / n^2 \\ y &= d_2 / n \end{aligned}}$$

$$G^t V^{-1} = [1 \;\; 1]\begin{bmatrix} b^2 + y^2 q & -xyq \\ -xyq & a^2 + x^2 q \end{bmatrix} Z = \left[(b^2 + y^2 q - xyq) \quad (a^2 + x^2 q - xyq)\right] Z \qquad Z = \left(a^2 b^2 + (a^2 y^2 + b^2 x^2) q\right)^{-1}$$

$$Y = G^t V^{-1}\begin{bmatrix} P - d_1/n \\ P - d_2/n \end{bmatrix} = Z\left[b^2 + y^2 q - xyq \quad a^2 + x^2 q + xyq\right]\begin{bmatrix} P - x \\ P - y \end{bmatrix} = Z\begin{bmatrix} P(a^2 + b^2 + (x^2 - 2xy + y^2)q) \\ \ldots - x(b^2 + (y^2 - xy)q) - y(a^2 + (x^2 - xy)q) \end{bmatrix}$$

$$W = G^t V^{-1} G = \left[a^2 + b^2 + (x - y)^2 q\right] Z$$

$$M' = W^{-1} = \left[a^2 + b^2 + (x - y)^2 q\right]^{-1} Z^{-1} = \frac{a^2 b^2 + (a^2 y^2 + b^2 x^2) q}{a^2 + b^2 + (x - y)^2 q} = \frac{1 + \left(\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2}\right) q}{\dfrac{1}{a^2} + \dfrac{1}{b^2} + \dfrac{(x - y)^2}{a^2 b^2} q}$$

$$\begin{aligned} P' = P - M'Y &= P - \{a^2 + b^2 + (x^2 - 2xy + y^2)q\}^{-1} \; Z^{-1} Z \; \{P(a^2 + b^2 + (x^2 - 2xy + y^2)q) - x(b^2 + (y^2 - xy)q) - y(a^2 + (x^2 - xy)q)\} \\ &= \{a^2 + b^2 + (x^2 - 2xy + y^2)q\}^{-1} \quad \big(P\{[a^2 + b^2 + (x^2 - 2xy + y^2)q] - [a^2 + b^2 + (x^2 - 2xy + y^2)q]\} \\ &\qquad + xb^2 + ya^2 + xq\{y^2 - x^2 y + x^2 y - xy^2\}\big) \\ &= \{a^2 + b^2 + (x^2 - 2xy + y^2)q\}^{-1} \quad (x/a^2 + y/b^2) a^2 b^2 \\ &= \{\Delta^2 d_1 + \Delta^2 d_2 + (d_1 - d_2)^2 \Delta^2 n / n^2\}^{-1} (d_1/\Delta^2 d_1 + d_2/\Delta^2 d_2) \Delta^2 d_1 \Delta^2 d_2 / n \\ &= \{Q + (d_1 - d_2)^2 \Delta^2 n / (n^2 \Delta^2 d_1 \Delta^2 d_2)\}^{-1} \; R/n \end{aligned}$$

$$\boxed{\begin{aligned} Q &= \Delta^{-2} d_1 + \Delta^{-2} d_2 \\ R &= d_1 \Delta^{-2} d_1 + d_2 \Delta^{-2} d_2 \end{aligned}}$$

PPP, continued (fit reduced data, usual method)

- **After doing all the algebra, we find the result for *P' & Δ²P'***

$$P' = \frac{\left(\frac{d_1}{\Delta^2 d_1} + \frac{d_2}{\Delta^2 d_2}\right)}{n\left(\frac{1}{\Delta^2 d_1} + \frac{1}{\Delta^2 d_2}\right) + n\,\frac{(d_2 - d_1)^2}{\Delta^2 d_1\,\Delta^2 d_2}\,\Delta^2 n}$$

$$\Delta^2 P' = \frac{\left\{\frac{1}{n^2}\left(\frac{1}{\Delta^2 d_1} + \frac{1}{\Delta^2 d_2}\right)^{-1} + \frac{1}{n^4}\left(\frac{d_1^{\,2}}{\Delta^2 d_1} + \frac{d_2^{\,2}}{\Delta^2 d_2}\right)\left(\frac{1}{\Delta^2 d_1} + \frac{1}{\Delta^2 d_2}\right)^{-1}\Delta^2 n\right\}}{\left\{1 + \frac{(d_2 - d_1)^2}{\Delta^2 d_1\,\Delta^2 d_2}\left(\frac{1}{\Delta^2 d_1} + \frac{1}{\Delta^2 d_2}\right)^{-1}\frac{\Delta^2 n}{n^2}\right\}}$$

PPP, continued (fit reduced data, usual method)

- **Reproduce PPP's original numbers?**
  - **Set d_1 = 1.5, Δd_1 = 0.15, d_2 = 1.0, Δd_2 = 0.1, n = 1.0, Δn = 0.2**

$$V = \begin{bmatrix} 0.15^2 + 1.5^2 0.2^2 & (1.5)(1.0)0.2^2 \\ (1.5)(1.0)0.2^2 & 0.10^2 + 1.0^2 0.2^2 \end{bmatrix} = \begin{bmatrix} 0.1125 & 0.06 \\ 0.06 & 0.05 \end{bmatrix}$$

$$P' = \frac{\left(\frac{1.5}{0.15^2} + \frac{1.0}{0.10^2}\right)}{1.0\left(\frac{1}{0.15^2} + \frac{1}{0.10^2}\right) + \frac{(1.5-1.0)^2}{0.15^2\,0.10^2}0.20^2} = \frac{15}{17} = 0.8823529...$$

$$\Delta^2 P' = \frac{\left\{\frac{1}{1.0^2}\left(\frac{1}{0.15^2} + \frac{1}{0.10^2}\right)^{-1} + \frac{1}{1.0^4}\left(\frac{1.5^2}{0.15^2} + \frac{1.0^2}{0.10^2}\right)\left(\frac{1}{0.15^2} + \frac{1}{0.10^2}\right)^{-1} 0.2^2\right\}}{\left\{1 + \frac{(1.5-1.0)^2}{0.15^2\,0.10^2}\left(\frac{1}{0.15^2} + \frac{1}{0.10^2}\right)^{-1}\frac{0.2^2}{1.0^2}\right\}} = (0.21828206...)^2$$

PPP, continued (fit reduced data, usual method)

- **Result for *P'* is the same result as in the fit-to-raw-data case only if $d_2 \rightarrow d_1$ or $\Delta^2 n \rightarrow 0$.**

$$P' = \frac{\left(\dfrac{d_1}{\Delta^2 d_1} + \dfrac{d_2}{\Delta^2 d_2}\right)}{n\left(\dfrac{1}{\Delta^2 d_1} + \dfrac{1}{\Delta^2 d_2}\right) + \cancel{n\dfrac{(d_2 - d_1)^2}{\Delta^2 d_1 \Delta^2 d_2}\Delta^2 n}}$$

PPP, continued (fit reduced data, usual method)

Result for *M'* ($\Delta P'$)^2 is equal to Truth only in the limit as $d_2 \rightarrow d_1$ or $\Delta^2 n \rightarrow 0$…

$$\Delta^2 P' = \frac{\left\{ \frac{1}{n^2}\left(\frac{1}{\Delta^2 d_1}+\frac{1}{\Delta^2 d_2}\right)^{-1} + \frac{1}{n^4}\left(\frac{d_1{}^2}{\Delta^2 d_1}+\frac{d_2{}^2}{\Delta^2 d_2}\right)\left(\frac{1}{\Delta^2 d_1}+\frac{1}{\Delta^2 d_2}\right)^{-1}\Delta^2 n \right\}}{\left\{ 1 + \frac{(d_2-d_1)^2}{\Delta^2 d_1\,\Delta^2 d_2}\left(\frac{1}{\Delta^2 d_1}+\frac{1}{\Delta^2 d_2}\right)^{-1}\frac{\Delta^2 n}{n^2} \right\}}$$

$$= \frac{a+b}{1+c} \Rightarrow a+(b-ac)$$

$$M' = \frac{1 + \left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right)q}{\frac{1}{a^2} + \frac{1}{b^2} + \frac{(x-y)^2}{a^2 b^2} q} \quad \rightarrow \quad what?$$

$$M' = \frac{F+G}{1+H} \rightarrow F + G - FH$$

$$F = \frac{1}{\frac{1}{a^2} + \frac{1}{b^2}} \qquad G = \left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right)qF \qquad H = \frac{(x-y)^2}{a^2 b^2} qF$$

$$\Rightarrow G - FH = Fq\left\{\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{(x-y)^2}{a^2 b^2}\frac{a^2 b^2}{a^2 + b^2}\right\} = Fq\left\{x^2 b^2 (a^2 + b^2) + y^2 a^2 (a^2 + b^2) - (x-y)^2 a^2 b^2\right\}\frac{1}{(a^2 + b^2) a^2 b^2}$$

$$= Fq\left\{x^2 (a^2 b^2 + b^4 - a^2 b^2) + y^2 (a^4 + a^2 b^2 - a^2 b^2) + 2xya^2 b^2\right\}\frac{1}{(a^2 + b^2) a^2 b^2}$$

$$= Fq\left\{x^2 b^4 + y^2 a^4 + 2xya^2 b^2\right\}\frac{1}{(a^2 + b^2) a^2 b^2}$$

$$= Fq\left\{\frac{x}{a^2} + \frac{y}{b^2}\right\}^2 \frac{a^2 b^2}{(a^2 + b^2)} = F^2 \left\{\frac{x}{a^2} + \frac{y}{b^2}\right\}^2 q = \frac{P'^2}{n^2}\Delta^2 n$$

PPP, continued (fit reduced data, usual method)

In the limit…

$$\Delta^2 P' \rightarrow \frac{1}{n^2}\left(\frac{1}{\Delta^2 d_1}+\frac{1}{\Delta^2 d_2}\right)^{-1} + \boxed{\frac{1}{n^4}\left(\frac{d_1}{\Delta^2 d_1}+\frac{d_2}{\Delta^2 d_2}\right)^{2}\left(\frac{1}{\Delta^2 d_1}+\frac{1}{\Delta^2 d_2}\right)^{-2}\Delta^2 n}$$

$$= \frac{1}{n^2}\left(\frac{1}{\Delta^2 d_1}+\frac{1}{\Delta^2 d_2}\right)^{-1} + \boxed{\frac{P'^2}{n^2}\Delta^2 n}$$

PPP, continued (fit reduced data)

- **So what's causing the problem? It's in the definition of *V* –**

$$V = \begin{bmatrix} \frac{\Delta^2 d_1}{n^2} + \frac{d_1^{\,2}}{n^2}\frac{\Delta^2 n}{n^2} & \frac{d_1 d_2}{n^2}\frac{\Delta^2 n}{n^2} \\ \frac{d_1 d_2}{n^2}\frac{\Delta^2 n}{n^2} & \frac{\Delta^2 d_2}{n^2} + \frac{d_2^{\,2}}{n^2}\frac{\Delta^2 n}{n^2} \end{bmatrix}$$

Comes from linear expansion valid only near $D = d_1/n$

Comes from linear expansion valid only near $D = d_2/n$

PPP, continued (fit reduced data)

- **Solution? Expand around *D* = *P*.**

$$V = \begin{bmatrix} \frac{\Delta^2 d_1}{n^2} + \boxed{P^2}\frac{\Delta^2 n}{n^2} & \boxed{P^2}\frac{\Delta^2 n}{n^2} \\ \boxed{P^2}\frac{\Delta^2 n}{n^2} & \frac{\Delta^2 d_2}{n^2} + \boxed{P^2}\frac{\Delta^2 n}{n^2} \end{bmatrix}$$

- **In this case, results for *P'* and Δ*P'* are the same in the fit-to-raw-data case as in the fit-to-reduced-data**

$$[P'] = [P] - M'Y \qquad Y = G^t V^{-1} \begin{bmatrix} P - d_1/n \\ P - d_2/n \end{bmatrix} \qquad G = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad V = \begin{bmatrix} a^2 + P^2 q & P\ q \\ P^2 q & b^2 + P^2 q \end{bmatrix} \qquad q = \Delta^2 n / n^2$$

$$\boxed{\begin{aligned} a^2 &= \Delta^2 d_1 / n^2 \\ x &= d_1 / n \\ b^2 &= \Delta^2 d_2 / n^2 \\ y &= d_2 / n \end{aligned}}$$

$$G^t V^{-1} = [1\ \ 1] \begin{bmatrix} b^2 + P^2 q & -P^2 q \\ -P^2 q & a^2 + P^2 q \end{bmatrix} Z = \begin{bmatrix} b^2 & a^2 \end{bmatrix} Z \qquad Z = \left(a^2 b^2 + (a^2 + b^2) P^2 q\right)^{-1}$$

$$Y = G^t V^{-1} \begin{bmatrix} P - d_1/n \\ P - d_2/n \end{bmatrix} = Z \begin{bmatrix} b^2 & a^2 \end{bmatrix} \begin{bmatrix} P - x \\ P - y \end{bmatrix} = Z\left[P(a^2 + b^2) - xb^2 - ya^2\right]$$

$$W = G^t V^{-1} G = \left[a^2 + b^2\right] Z$$

$$M' = W^{-1} = \left[a^2 + b^2\right]^{-1} Z^{-1} = \frac{a^2 b^2 + (a^2 + b^2) P^2 q}{a^2 + b^2} = \frac{a^2 b^2}{a^2 + b^2} + P^2 q = \frac{1}{n^2}\left(\frac{1}{\Delta^2 d_1} + \frac{1}{\Delta^2 d_2}\right)^{-1} + P^2 \frac{\Delta^2 n}{n^2}$$

$$P' = P - M'Y = P - \{a^2 + b^2\}^{-1} Z^{-1} Z \{P(a^2 + b^2) - xb^2 - ya^2\} = P - P + \frac{xb^2 + ya^2}{(a^2 + b^2)} = \frac{\dfrac{x}{a^2} + \dfrac{y}{b^2}}{\dfrac{1}{a^2} + \dfrac{1}{b^2}} = \frac{1}{n} \frac{\dfrac{d_1}{\Delta^2 d_1} + \dfrac{d_2}{\Delta^2 d_2}}{\dfrac{1}{\Delta^2 d_1} + \dfrac{1}{\Delta^2 d_2}}$$

Results for fitting to reduced data with the correct data covariance matrix–

$$P' = \frac{\left(\frac{d_1}{\Delta^2 d_1} + \frac{d_2}{\Delta^2 d_2}\right)}{n\left(\frac{1}{\Delta^2 d_1} + \frac{1}{\Delta^2 d_2}\right)} = \frac{\left(\frac{1.5}{0.15^2} + \frac{1.0}{0.10^2}\right)}{1.0\left(\frac{1}{0.15^2} + \frac{1}{0.10^2}\right)} = \frac{15}{13} = 1.153846...$$

$$\Delta^2 P' = \frac{1}{n^2}\left(\frac{1}{\Delta^2 d_1} + \frac{1}{\Delta^2 d_2}\right)^{-1} + \frac{P'^2}{n^2}\Delta^2 n$$

$$= \frac{1}{1.0}\left(\frac{1}{1.5^2} + \frac{1}{1.5^2}\right)^{-1} + \frac{\left(\frac{15}{13}\right)^2}{1.0^2}(0.2)^2 = (0.245311057...)^2$$

Practical application?

- **Years ago I added capability to use "implicit data covariance matrix" in SAMMY**
  - **First, using wrong *V* (the one derived from data *D*)**
    - **Runs kept misbehaving, numerical problems etc.**
  - **So switched from using *D* (measured data) in generating *V*, to using *T* (theoretical values)**
    - **Numerical problems disappeared.**

- **I got it right by accident!**

Observation and Question concerning this example

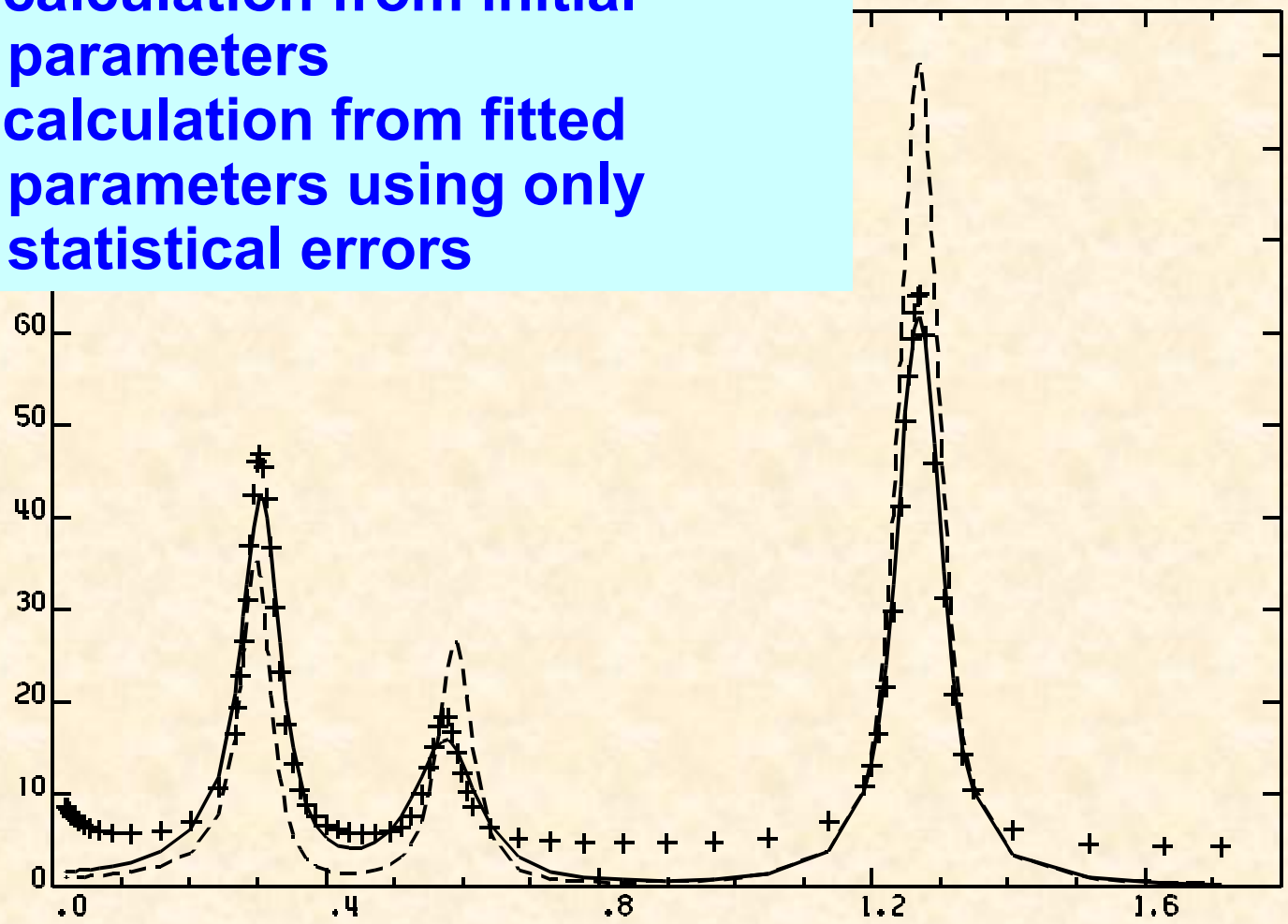
- **Observation**
  - **When fitting reduced data, if systematic uncertainties are completely ignored, then**
    - **Parameter value *P'* is correct**
    - **Uncertainty Δ*P'* can be found by adding systematic uncertainty in quadrature**
- **Question**
  - **Can this be generalized?**
    - **Multivariate?**
    - **Non-linear?**

O&Q

- **My expectation is "Not easily"**
  - **Example: R-matrix analysis of fission cross section with obviously-wrong background produces obviously-wrong resonance parameters …**

^{241}Am fission cross section

cross = "experiment"
dash = calculation from initial parameters
solid = calculation from fitted parameters using only statistical errors



(end of PPP discussion)

Topics to be ~~covered~~ touched on

- **Peelle's Pertinent Puzzle/Paradox**
- **Implicit data covariance (IDC) methodology**
- **Transformation of variables**

Implicit Data Covariance Method

Complete data covariance matrix is

$$V = v + X\,Q\,X^{t}$$

where

- ***V*** **is the data covariance**
- ***v*** **represents the statistical uncertainties**
- ***X*** **is the sensitivity matrix (partial derivative of data with respect to data-reduction parameters) [evaluated at current values of fitting parameters to avoid PPP-type difficulties]**
- ***Q*** **is the covariance matrix for the data-reduction parameters**

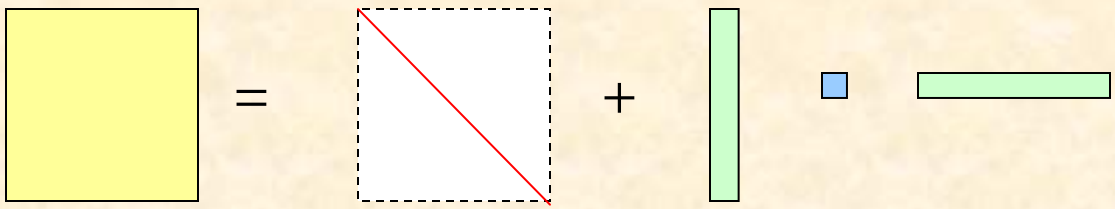
(what are these data reduction parameters?)

- **used for describing experimental conditions**
  - **normalization, background**
  - **burst width**
  - **isotopic abundance**
  - **etc.**

- **can be used in two ways**
  - **used to generate data covariance matrix**
  - **included as varied parameters in fitting procedure**
    - **mathematically equivalent to using data-covariance matrix**
    - **numerically more stable**
    - **bonus: values of data-reduction parameters are updated**

Data Covariance Matrix, symbolically

$$V = v + X\ Q\ X^t$$



where

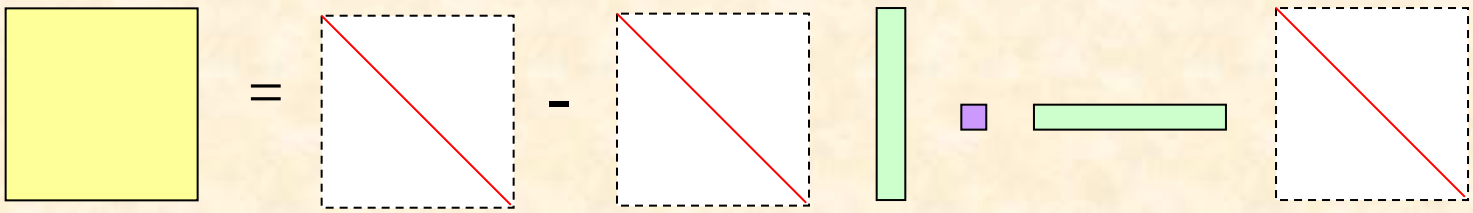
- **size of box may be considered to be logarithmic**
  - **large ~ thousands (or 10 K or 100 K)**
  - **small ~ very few (5? 10?)**
- **dashed box with diagonal line indicates diagonal matrix**
- **solid box indicates non-diagonal matrix**

Inverse of Data Covariance Matrix, symbolically

$$V^{-1} = (v + X\ Q\ X^t)^{-1}$$

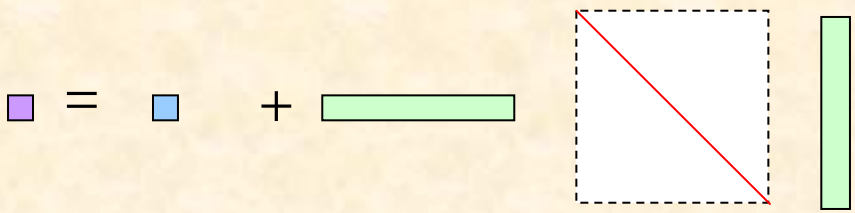
$$= v^{-1} - v^{-1} X (Q^{-1} + X^t v^{-1} X)^{-1} X^t v^{-1}$$

$$= v^{-1} - v^{-1} X\ Z^{-1}\ X^t\ v^{-1}$$



where

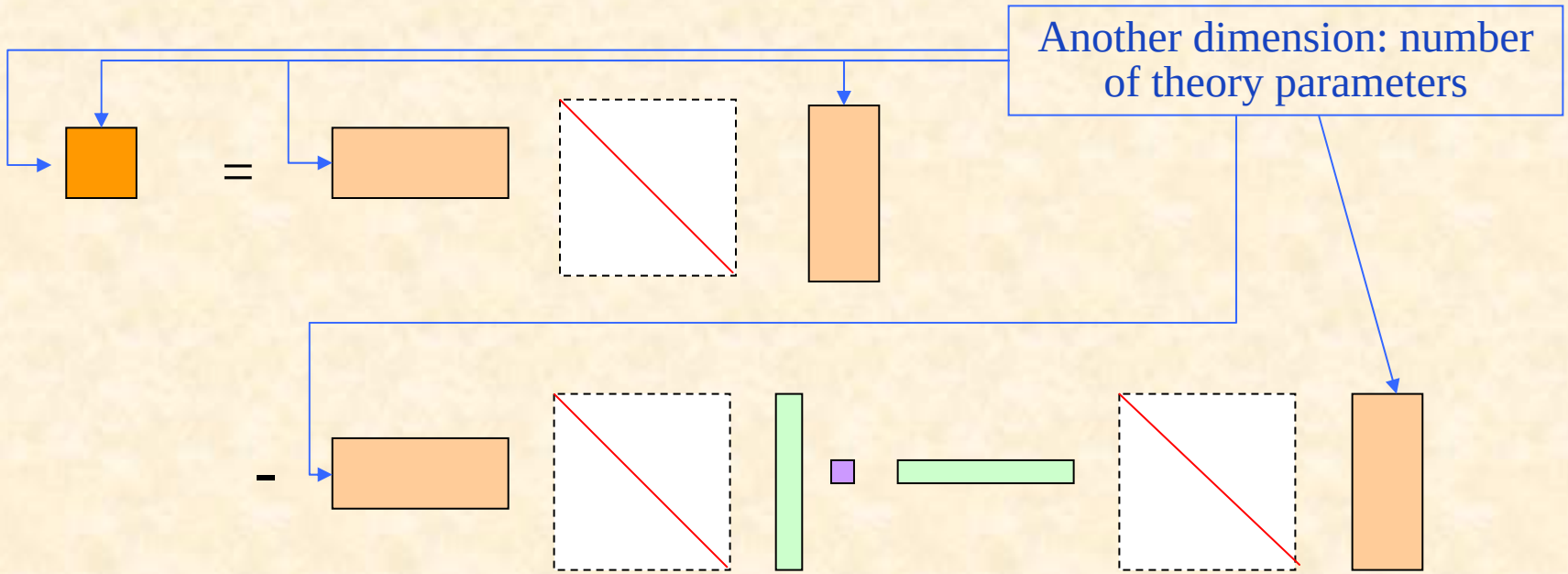
$$Z = Q^{-1} + X^t\ v^{-1}\ X$$



Quantities needed in Bayes' equations: W

$$W = G^t V^{-1} G$$

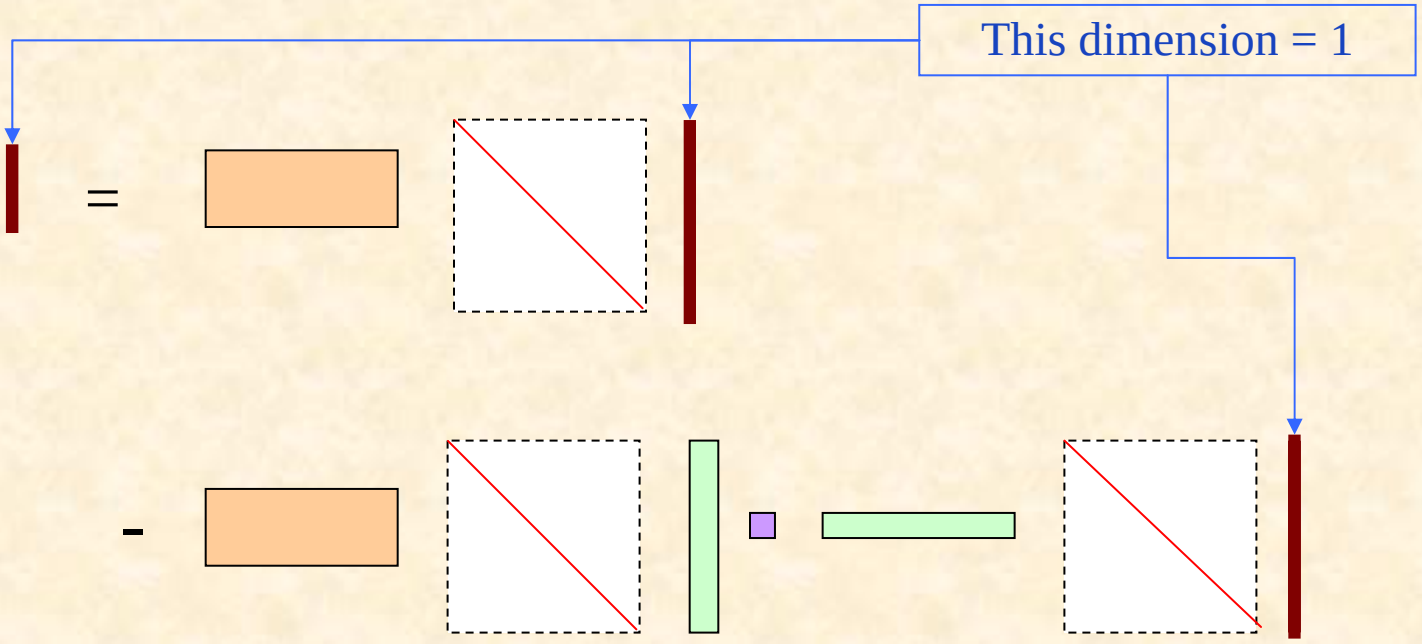
$$= G^t v^{-1} G - G^t v^{-1} X Z^{-1} X^t v^{-1} G$$



Quantities needed in Bayes' equations: Y

$$Y = G^t V^{-1} (D - T)$$

$$= G^t v^{-1} (D - T) - G^t v^{-1} X \; Z^{-1} X^t v^{-1} (D - T)$$



Why bother with all these arrays?

Matrices in *W* and *Y* are easier to invert than *V* –

- ***v* is large but diagonal**
  - **(size ~ thousands of data points)**
- ***Q* is small and often diagonal**
  - **(size ~ tens of data-reduction parameters)**
- ***Z* is off-diagonal but small**
  - **(size ~ tens)**

– which leads to savings in

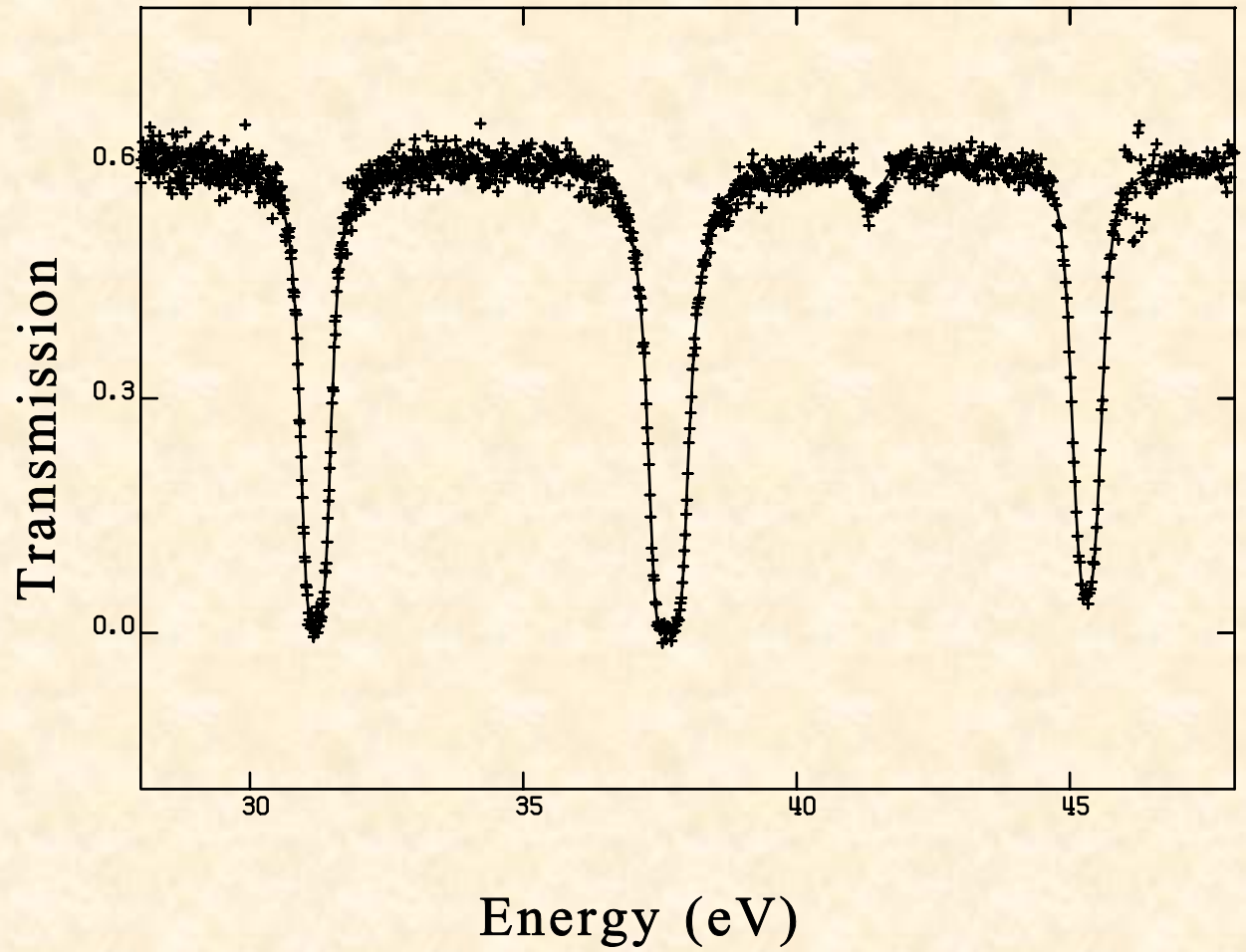
- **computation time (never calculate *V* or V^{-1})**
- **computer memory (never store *V* or V^{-1})**
- **numerical accuracy and stability (fewer round-off problems)**

In SAMMY, **implicit data covariance (IDC) matrices** can be used for

- **normalization**
- **background correction factors**
- **user-supplied implicit data covariance** new
  - **external code can be used to generate pieces (*X* and *Q*)**

Example from SAMMY Test Case tr140: ^{129}I transmission data

Geel data provided by Gilles Noguere, Cadarache



Example, continued

1245 experimental data points;
[full data set has 32660 data points]
9 data-reduction parameters;
655 resonances; 9 varied parameters

	Description of data covariance treatment for this run	Cpu time for Bayes solver (sec)	Total cpu time (sec)	Array size
a	only statistical errors	0.03	14	254 K
b	statistical plus systematic, only on diagonal	0.03	14	254 K
c	explicit data cov matrix	16.46	59	1800 K
d	IDC matrix	0.06	14	267 K

Note: c & d give essentially the same results

For more on covariance matrices and IDC –

"Practical Alternatives to Explicitly Generating and Inverting Data Covariance Matrices"

N. M. Larson

Nuclear Mathematical and Computational Sciences: A Century in Review, A Century Anew

Gatlinburg, Tennessee, April 6-11, 2003

On CD-ROM
American Nuclear Society, LaGrange Park, IL (2003)

Suppose we want to make use of uncertainties for non-varied parameters ?

(e.g., resolution-function parameters)

Add a third option for parameter flag in SAMMY:

- **0 = do not vary**
- **1 = vary**
- **3 = calculate partial derivatives, propagate uncertainties**
  - **to be treated in similar fashion to IDC**

current SAMMY options

to be implemented soon

(end of IDC discussion)

Topics to be ~~covered~~ touched on

- **Peelle's Pertinent Puzzle/Paradox**
- **Implicit data covariance (IDC) methodology**
- **Transformation of variables**

Transformation of variables

- **Least-squares equations (and/or Bayes' equations) are based on the implicit assumption that the parameters obey Gaussian statistics.**
- **Transformation (via *log*, or *sqrt*, or whatever) will introduce a new variable which obeys another statistic.**
- **Results obtained with the new variable will therefore be different from results obtained with the original.**
- **Which result is "correct" depends at least in part on which variable obeys the appropriate (Gaussian) statistic.**

Transformation of variables

- **Least-squares equations (and/or Bayes' equations) are based on the implicit assumption that the parameters obey Gaussian statistics.**
- **Transformation (via *log*, or *sqrt*, or whatever) will introduce a new variable which obeys another statistic.**
- **Results obtained with the new variable will therefore be different from results obtained with the original.**
- **Which result is "correct" depends at least in part on which variable obeys the appropriate (Gaussian) statistic.**

THE END

The End