A general methodology for updating PDF sets with LHC data

Francesco de Lorenzi*, Ronan McNulty
(University College Dublin)
* now at Iowa State (ATLAS)

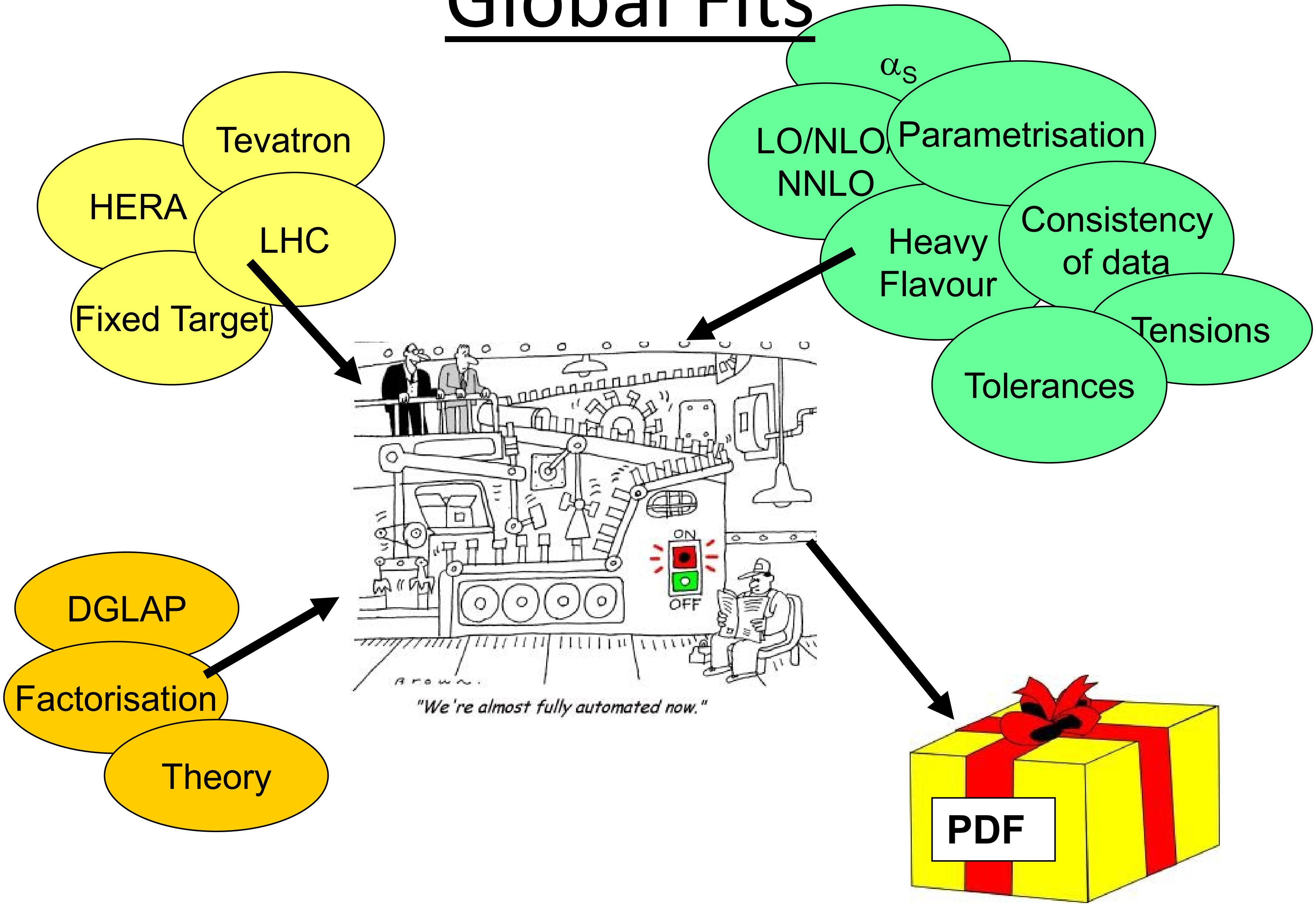
Concept

- It would be nice to know the impact that your experimental measurement has on the PDFs.
- Since each PDF set defines a probability density function (either multinomial distribution or sampling), in principle one can update this using your measurement.

- NB: This technique assumes that the PDFs can be described using standard statistical technique.
- NB: This is often not true (issue of tolerances).

- Consequently this methodology is NOT a replacement for Global Fits, but is indicative of the likely improvements your data will bring.

Global Fits



αS
LO/NLO/
NNLO
Parametrisation
Heavy
Flavour
Consistency
of data
Tensions
Tolerances
Tevatron
HERA
LHC
Fixed Target
DGLAP
Factorisation
Theory
ON
OFF
"We're almost fully automated now."
PDF

What do you get for your money?

(An end-user perspective)

A set of parameters: $(\lambda_1, \lambda_2, \lambda_3 \ldots \lambda_n)$

A covariance matrix:

$$\begin{pmatrix} \delta_1^2 & 0 & 0 & 0 & 0 \\ 0 & \delta_2^2 & 0 & 0 & 0 \\ 0 & 0 & \delta_3^2 & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & \delta_n^2 \end{pmatrix}$$

These correspond to distances along a set of orthonormal vectors defined at the minimum of the global chisquared.

For any function, *F*, depending on λ, the best value is:

$$F(\lambda_1, \lambda_2, \lambda_3 \ldots \lambda_n)$$

and the uncertainty is

$$\delta F^2 = \Sigma_i \left(\frac{dF}{d\lambda_i} \right)^2 \delta_i^2$$

What do you get for your money?

(An end-user perspective)

PDFs are of form (e.g.): $xf(x, Q_0^2) = (1-x)^\eta(1+\epsilon x^{0.5}+\gamma x)x^\delta$

with $\eta(\lambda_i)$ $\varepsilon(\lambda_i)$ $\gamma(\lambda_i)$

So $u_v = u_v(\lambda_1, \lambda_2, \ldots \lambda_n)$, $d_v = d_v(\lambda_1, \lambda_2 \ldots \lambda_n)$, $g = g(\lambda_1, \lambda_2, \ldots \lambda_n)$

And $\delta_{u_v} = \sqrt{\sum_i \left(\frac{du_v}{d\lambda_i}\right)^2 \delta_i^2}$ $\delta_{d_v} = \sqrt{\sum_i \left(\frac{dd_v}{d\lambda_i}\right)^2 \delta_i^2}$ $\delta_g = \ldots$ etc.

Thus PDFs and uncertainties defined through λ_i

But also observables e.g. $\sigma_Z(\lambda_1, \lambda_2 \ldots \lambda_n)$ with

$$\delta_\sigma = \sqrt{\sum_i \left(\frac{d\sigma}{d\lambda_i}\right)^2 \delta_i^2}$$

or differential cross-sections, or ratios of cross-sections.

How does new data affect PDF?

The global PDF fit gives us $\lambda_1^0 +- \delta_1^0$, $\lambda_2^0 +- \delta_2^0 \ldots \lambda_n^0 +- \delta_n^0$ and predicts $\sigma_Z^{th}(\lambda_1^0, \lambda_2^0, \lambda_3^0 \ldots \lambda_n^0)$ and now we measure σ_Z +- δ_Z .

We can fit for new values of λ_i by minimising:

$$\chi^2(\lambda_1, \lambda_2, \ldots\lambda_n) = \sum_i \left(\frac{\sigma_Z - \sigma_Z^{th}(\lambda_1, \lambda_2, \ldots\lambda_n)}{\delta_Z} \right)^2 + \sum_i \left(\frac{\lambda_i - \lambda_i^0}{\delta_i^0} \right)^2$$

measurement

prediction

constraint

And taking second derivative of χ2 we get a covariance matrix:

$$\begin{pmatrix} \delta_1^2 & \neq 0 & \neq 0 & \neq 0 & \neq 0 \\ \neq 0 & \delta_2^2 & \neq 0 & \neq 0 & \neq 0 \\ \neq 0 & \neq 0 & \delta_3^2 & \neq 0 & \neq 0 \\ \neq 0 & \neq 0 & \neq 0 & \ldots & \neq 0 \\ \neq 0 & \neq 0 & \neq 0 & \neq 0 & \delta_n^2 \end{pmatrix}$$

where $\delta_i \leq \delta_i^0$
and in general off-diagonal terms are non-zero.

So improved values for λ_i with smaller (though correlated) errors

The old function relationship still holds (though the basis is no longer orthonormal)

PDFs are of form (e.g.): $xf(x, Q_0^2) = (1-x)^\eta(1+\epsilon x^{0.5}+\gamma x)x^\delta$

with $\eta(\lambda_i)$ $\varepsilon(\lambda_i)$ $\gamma(\lambda_i)$

So $u_v = u_v(\lambda_1, \lambda_2, \ldots \lambda_n)$, $d_v = d_v(\lambda_1, \lambda_2 \ldots \lambda_n)$, $g = (\lambda_1, \lambda_2, \ldots \lambda_n)$

And $\delta_{u_v} = \sqrt{\sum_{ij} \frac{du_v}{d\lambda_i} V_{ij} \frac{du_v}{d\lambda_j}}$ δ_{dv}=…. etc.

Thus PDFs and uncertainties defined through λ_i

While observables like $\sigma_Z(\lambda_1, \lambda_2, \lambda_3 \ldots \lambda_n)$ with $\delta_\sigma = \sqrt{\sum_{ij} \frac{d\sigma}{d\lambda_i} V_{ij} \frac{d\sigma}{d\lambda_j}}$

or differential cross-sections, or ratios of cross-sections, are predicted. (Compare to two slides back)

Constraining the PDFs

Before the fit

$$\delta F = \sqrt{\sum_i \left(\frac{\partial F}{\partial \lambda_i}\right)^2 \delta_i^2}$$

Uncertainties obtained adding in quadrature eigenvector deviation from central value

After the fit

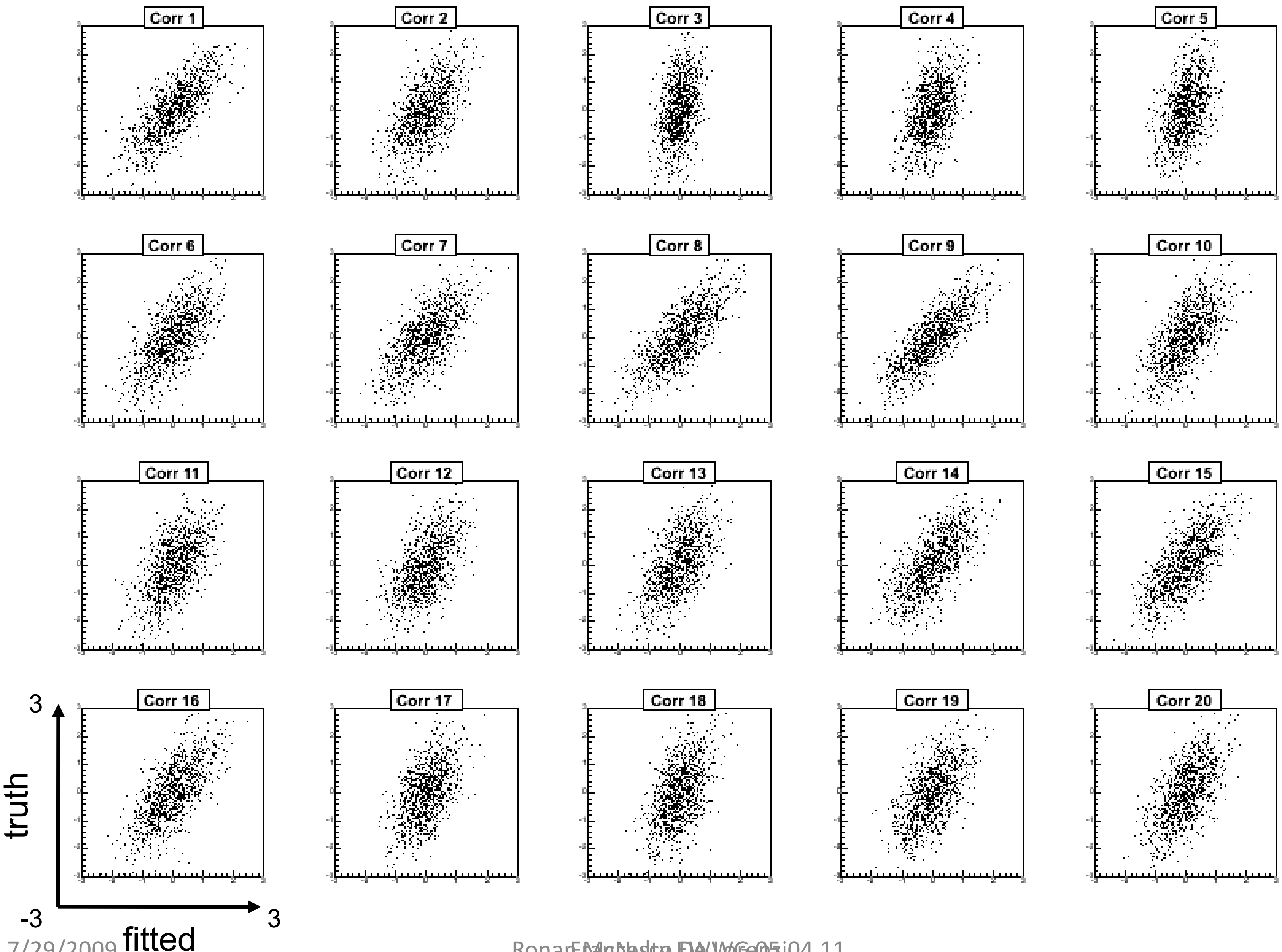
$$\delta F = \sqrt{\sum_{ij} \frac{\partial F}{\partial \lambda_i} V_{ij} \frac{\partial F}{\partial \lambda_j}}$$

V is the new covariance matrix of the fit

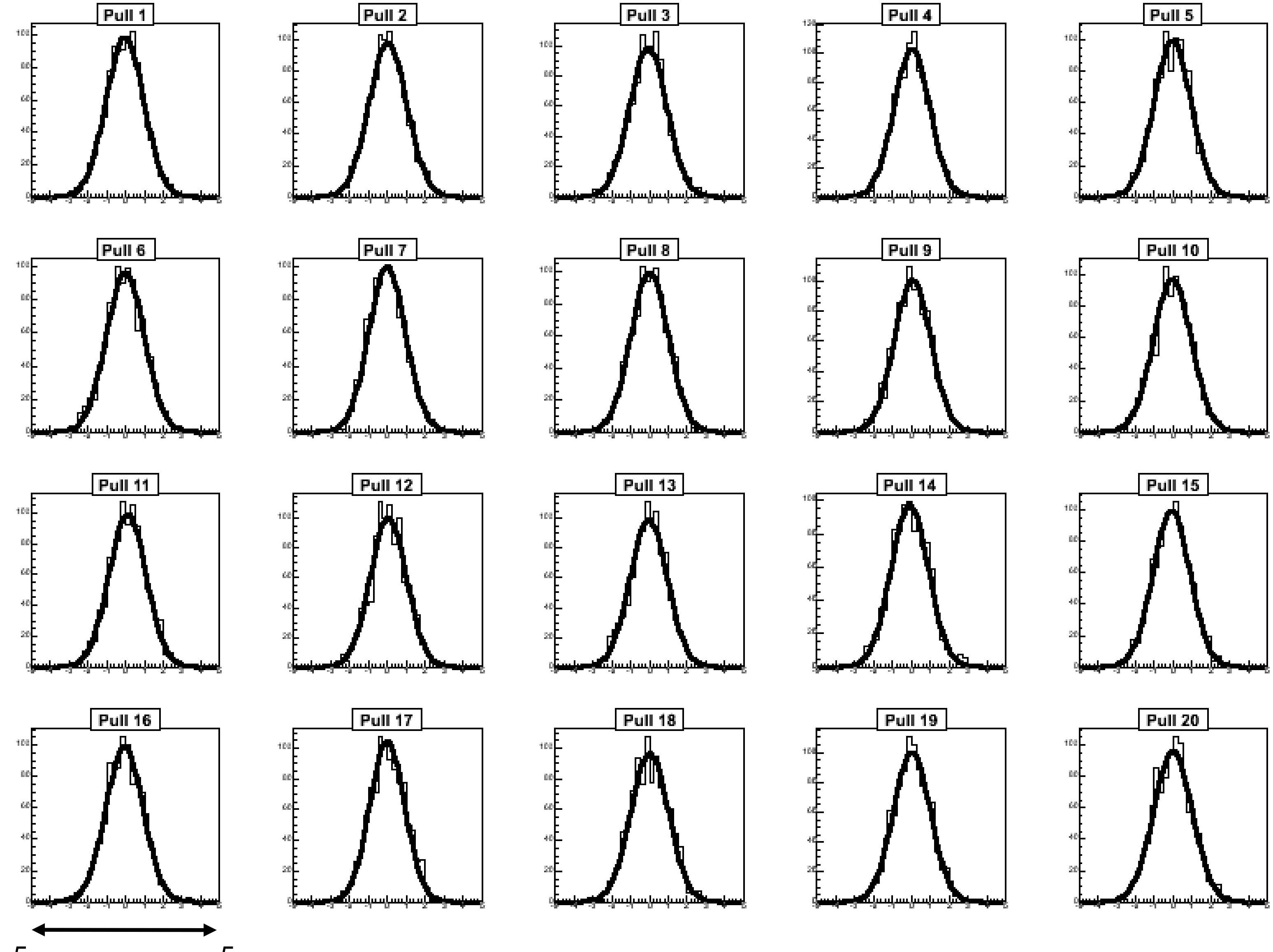
- F can be σ, u_V, d_V, g

Note: Before taking any data, it's possible to estimate the improvement in the PDFs with a given luminosity. The central value though, will depend on what you measure.

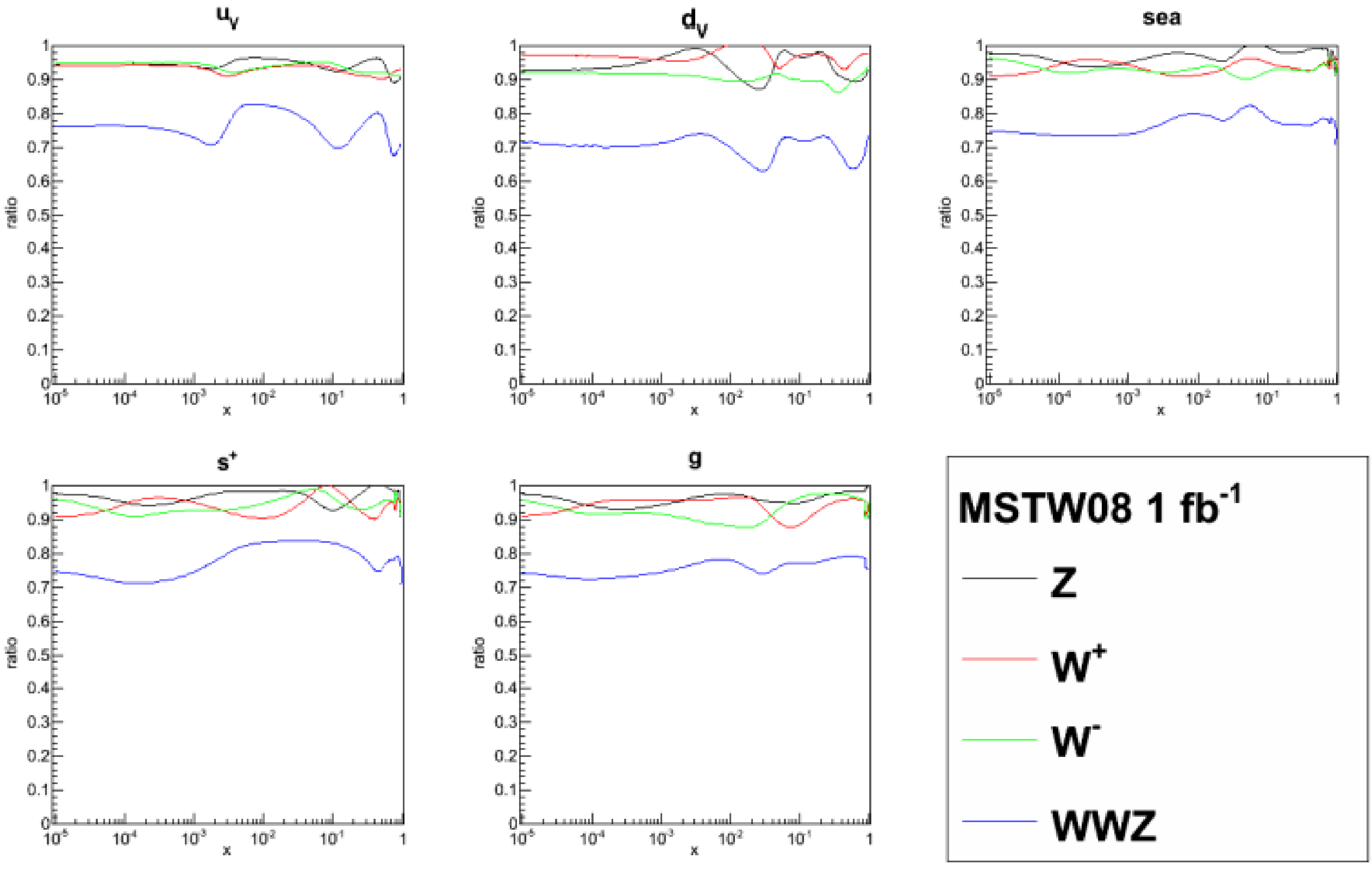
truth VS fitted (each eigenvalues)

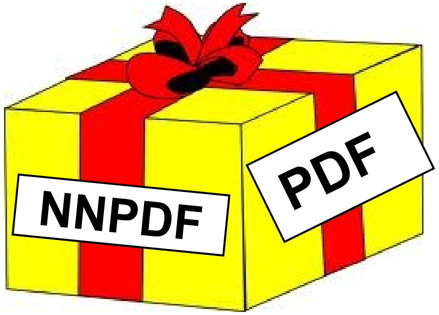


Pull = (truth – fitted)/error_fitted



Expected improvement of MSTW08 PDFs with 1fb-1 of LHCb data





What do you get for your money?

(An end-user perspective)

N equal probablity replicas: u_v^i, d_v^i, g^i, s^i etc. (i=1,N)
The ensemble defines the probability phase space.

For any function, *F*, depending on u,d,s…g, there will be N possible values for F.

The 'best value' of F from mean:

$$< F >= \frac{1}{N}\sum_{i}^{N} F(u_i, d_i, s_i \ldots g_i)$$

The uncertainty from RMS.

$$\delta_F = \sqrt{\frac{1}{N}\sum_{i}^{N}\Big(F(u_i, d_i, s_i \ldots g_i) - < F >\Big)^2}$$

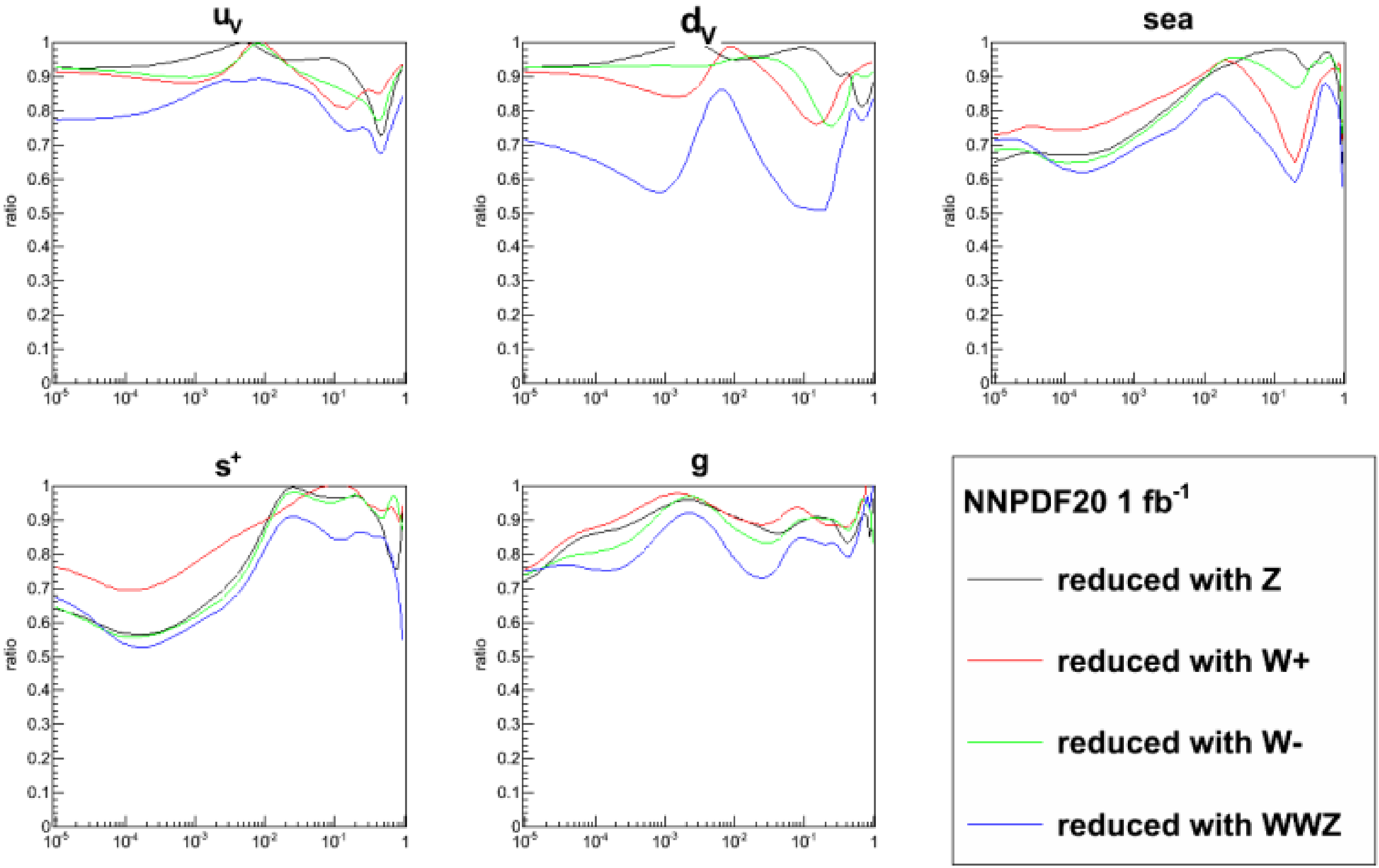
Again, F can be a PDF, or an observable.

How does new data affect PDF?

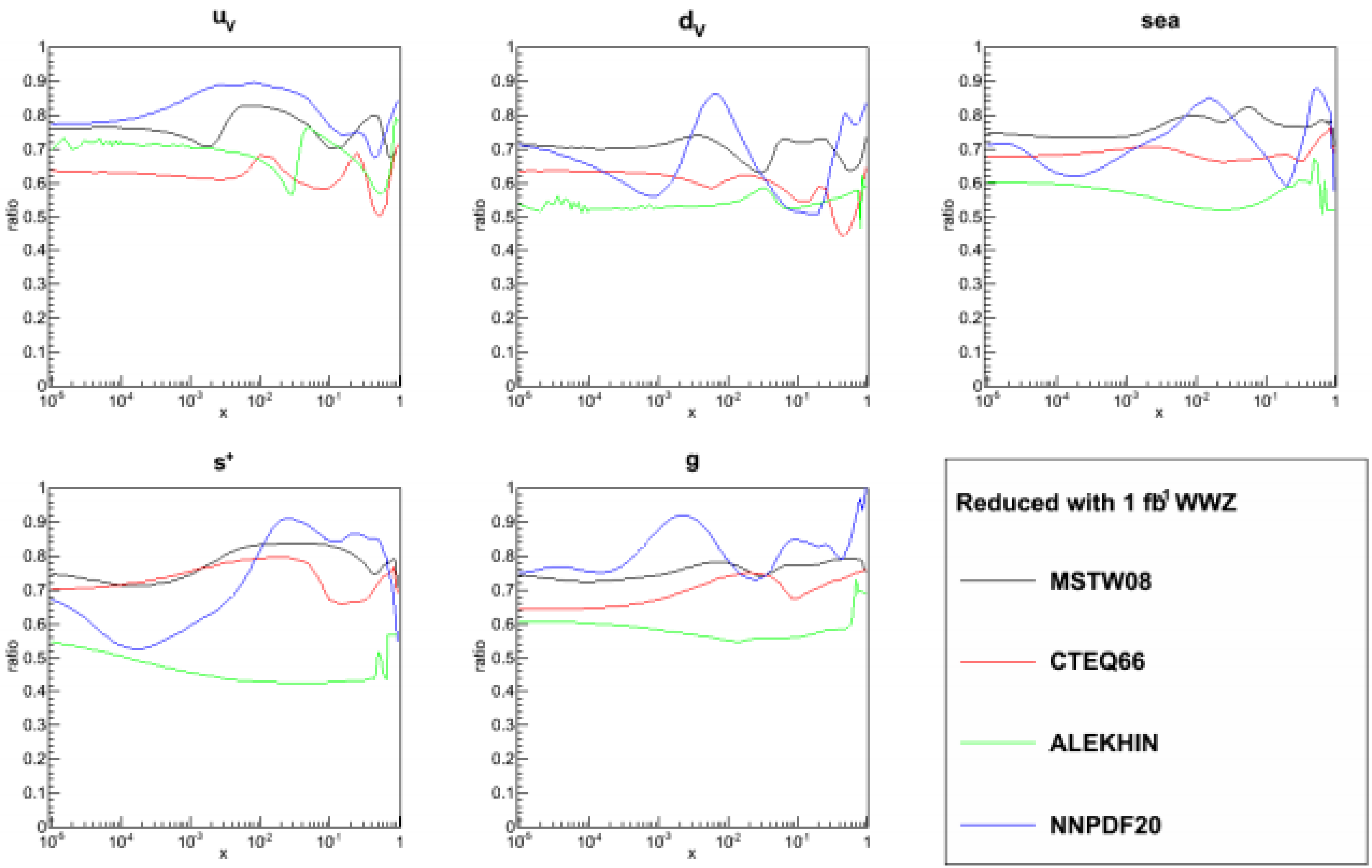
Simplistic prescription:

- The data will be compatible with some replicas, and incompatible with others.
- Evaluate with $\chi^2 = \left(\frac{\sigma - \sigma_Z^{th}(u_i, d_i, ...g_i)}{\delta_Z} \right)^2$
- Remove incompatible replicas. (Prob<1%)
- The spread of replicas reduces; the uncertainty is smaller; the mean may shift.

Expected improvement of NNPDF20 PDFs with 1fb-1 of LHCb data



Improvements for various PDF sets with 1fb-1 of LHCb electroweak data



$$ratio = \frac{\delta_{after}}{\delta_{before}}$$

How does new data affect PDF?

Sophisticated prescription (NNPDF):

- Take each PDF replica as a Bayesian Prior.
- Evaluate Chi2 compared to this replica
- Probability is $\mathcal{P}(\chi^2|f) \propto \chi^2(x,f)^{n/2-1} e^{-\frac{1}{2}\chi^2(x,f)}$
- Each replica is weighted by this probability
- Take mean and variance of weighted replicas.

Summary

- The NNPDF (sophisticated) prescription has been used to show affect of LHCb, ATLAS and CMS data. (See Maria's talk)
- Similar simple techniques can also be used to update the Hessian methods (shown here).

CAVEAT: You are assuming that the global fits obey standard error propagation.
Past experience has shown this is not always the case.