Comparison of PDF combination methods: a) Compressed replicas b) meta-parametrizations

Pavel Nadolsky

with recent inputs from both groups

Combination of the PDFs into the future PDF4LHC ensemble

PDFs from several groups are combined into a PDF4LHC ensemble of error PDFs **before** the LHC observable is computed. This simplifies the computation of the PDF+α_s uncertainty and will likely cut down the number of the PDF member sets and the CPU time needed for simulations.

The same procedure is followed at NLO and NNLO. The combination was demonstrated to work for global ensembles (CT, MSTW, NNPDF). It still needs to be generalized to allow inclusion of non-global ensembles.

The PDF uncertainty at 68% c.l is computed from error PDFs at central $\alpha_s(M_Z)$.

Two additional error PDFs are provided with either PDF4LHC ensemble to compute the α_s uncertainty using $\alpha_s(M_Z) = 0.118 \pm 0.0012$ at the 68% c.l.

Progress in developing the combination procedure

We carry out detailed comparisons of two methods for the PDF4LHC combination:

1. **Compression of Monte-Carlo replicas**
(G. Watt, R. Thorne,1205.4024; R. Ball et al., 1108.1758; S. Forte, G. Watt, 1301.6754; update by S. Carrazza)

2. **Meta-parametrizations + MC replicas + Hessian data set diagonalization**
(J. Gao, J. Huston, P. Nadolsky, 1401.0013, ; update by P.N.)

Both procedures start by creating a combined ensemble of MC replicas from all input ensembles (suggested by S. Forte, G. Watt, …). They differ at the second step of reducing a large number of input MC replicas (~ 300) to a smaller number for practical applications (13-100 in the META approach; 30 in the CMC approach). The core question is how much input information to retain in the reduced replicas, and by which method.

Benchmark comparisons of two combination methods.
Work plan (from Benasque workshop)

Input MC ensemble: **NNPDF3.0+CT14+MMHT14 NNLO**, with **alphas(M_Z)=0.118**

Convert to 300 replicas in LHAPDF6 format at $Q_0 = 8$ GeV (above the bottom mass), using two independent codes (JR and JG). Cross-check that results are identical.

Done. The results from two groups agree. Mild differences are due to random variations in the generation of MC replicas.

In each approach, reduce the number of replicas to the minimal number that retains 1% or 5% accuracy in reproducing the following properties of the input ensemble:

- Means, 68%c.l. PDF uncertainties, higher moments and asymmetry (skewness), PDF-PDF correlations.

In progress. Ensembles with 60-100 META PDFs and 40 CMC replicas broadly agree.

- Predictions for the standard candle LHC observables used in the META paper: ggHiggs, ttbar, W,Z [Jun]
- Differential LHC distributions using NNPDF3.0 applgrids, supplemented with some new aMCfast grids [Juan]

Benchmark comparisons, general observations

CMC ensembles with 40 replicas and META ensembles with 60-100 replicas are compared with the full ensembles of 300-600 MC replicas.

In the (x, Q) regions covered by the data, the agreement of 68%, 95% c.l. intervals, PDF-PDF correlations is excellent.

Accuracy of both combination procedures is already competitive with the 2010 PDF4LHC procedure, can be further fine-tuned by adjusting the final number of replicas, changing the number of parameters in the META method.

Benchmark comparisons, general observations II

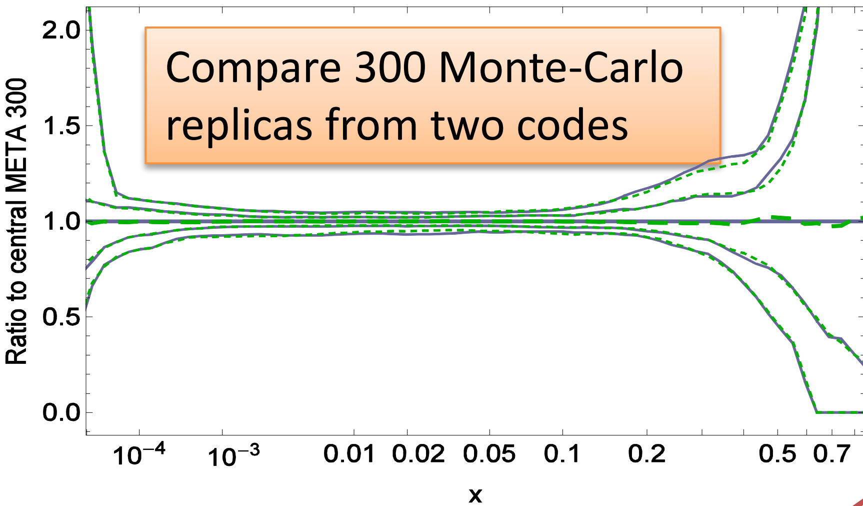
In the regions without the data, the error bands of the META ensembles are narrower as a result of more rigid interpolation than in CMC ensembles.

60-100 META PDFs appear to be necessary to compute asymmetric Hessian errors for general-purpose LHC applications. 30-50 META error are sufficient for computing the symmetric Hessian error.

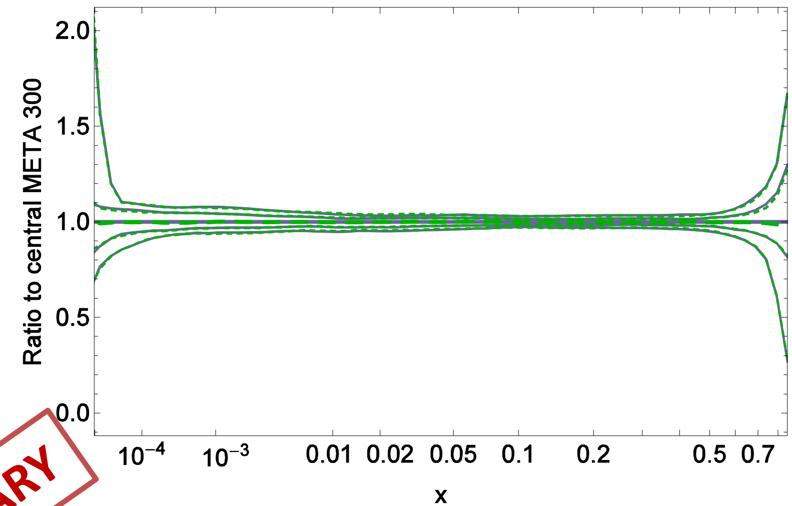
For specific classes of observables, such as various Higgs cross sections, the number of essential META PDFs can be further reduced by “data set diagonalization”, so that only of order 12 error sets are needed to reproduce PDF uncertainties and correlations for this class of observables.

Generation of MC replicas, Q=8 GeV

g (x,Q) at Q=8 GeV at 68% and 95% c.l.
META300 (solid), CMC (dashed)

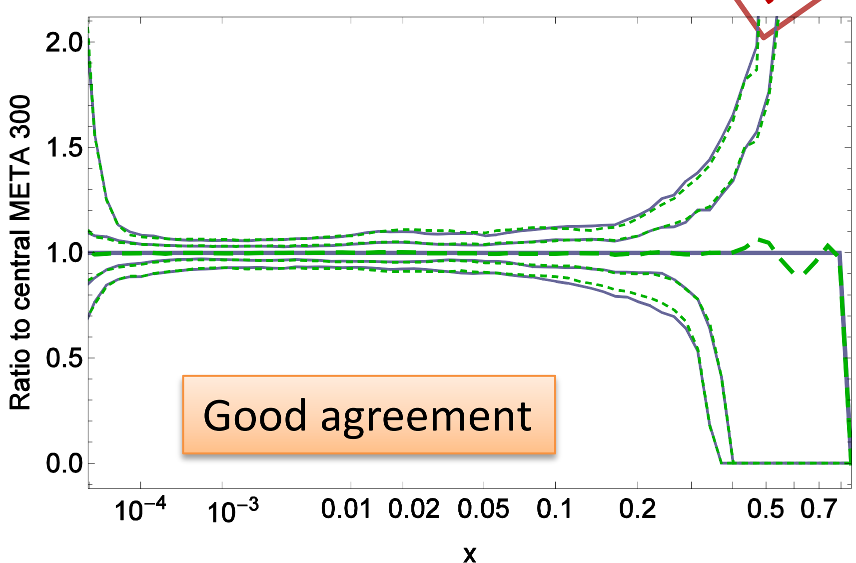


u (x,Q) at Q=8 GeV at 68% and 95% c.l.
META300 (solid), CMC (dashed)

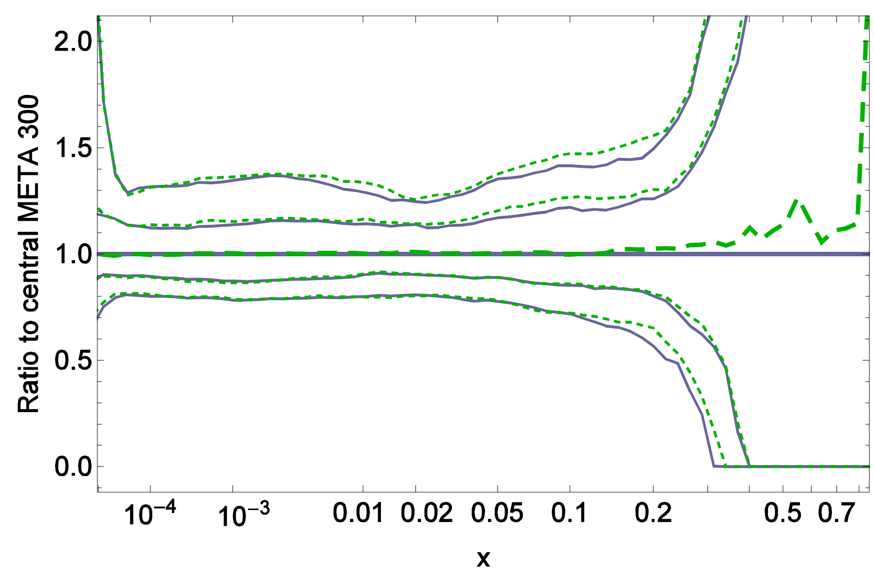


PRELIMINARY

$\bar{d}$ (x,Q) at Q=8 GeV at 68% and 95% c.l.
META300 (solid), CMC (dashed)

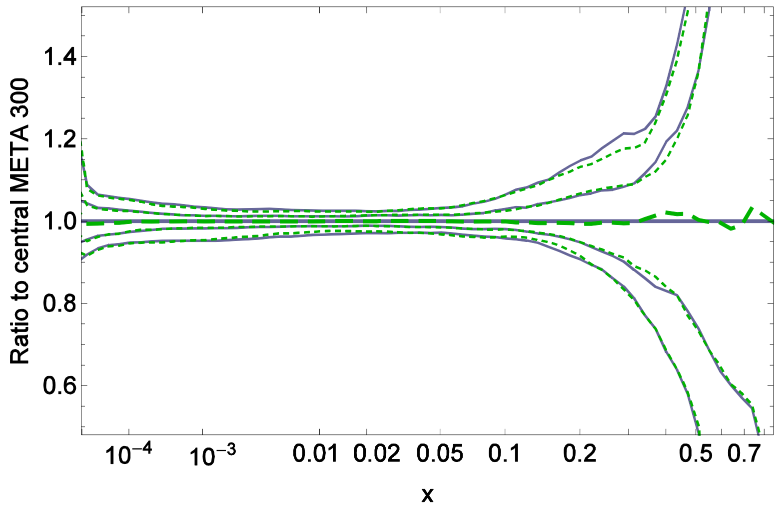


$\bar{s}$ (x,Q) at Q=8 GeV at 68% and 95% c.l.
META300 (solid), CMC (dashed)

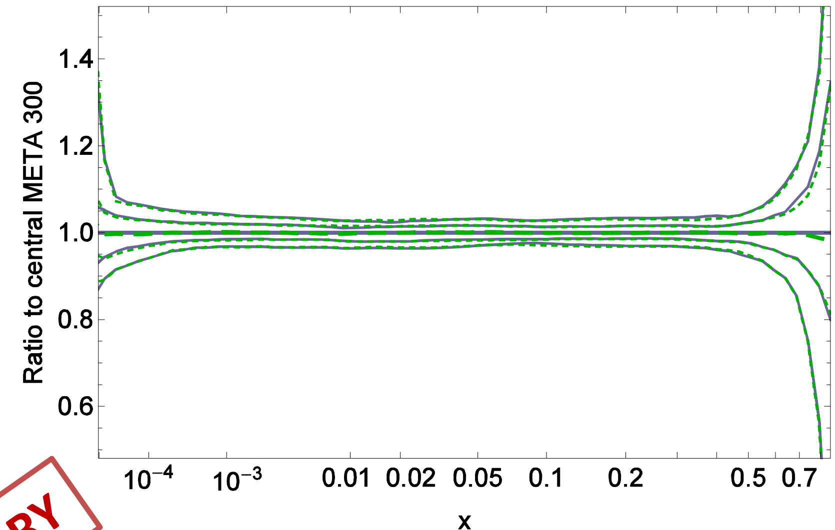


Same, Q=125 GeV

g (x,Q) at Q=125 GeV at 68% and 95% c.l.
META300 (solid), CMC (dashed)

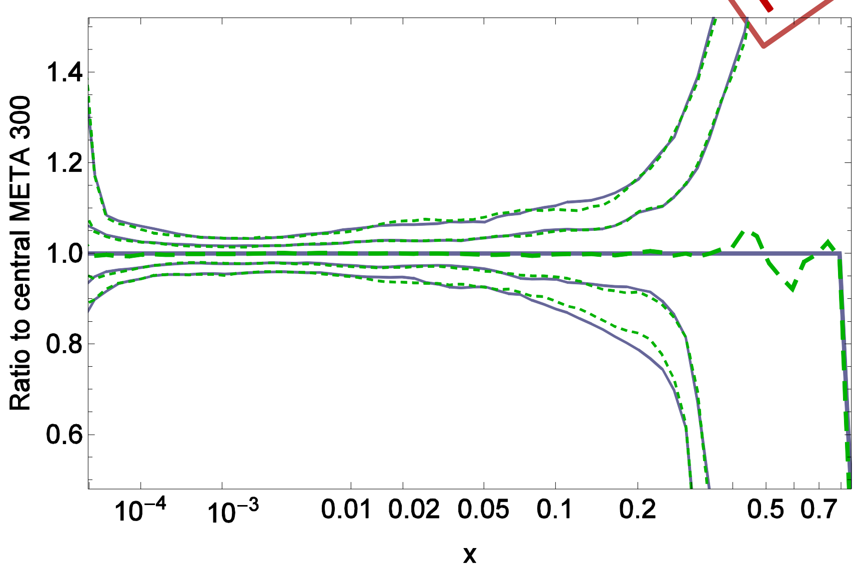


u (x,Q) at Q=125 GeV at 68% and 95% c.l.
META300 (solid), CMC (dashed)

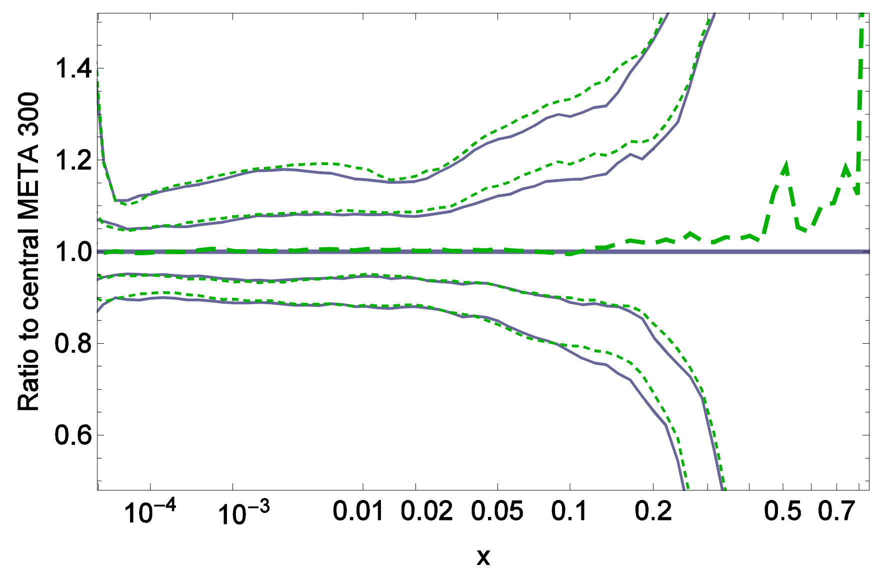


PRELIMINARY

$\bar{d}$ (x,Q) at Q=125 GeV at 68% and 95% c.l.
META300 (solid), CMC (dashed)



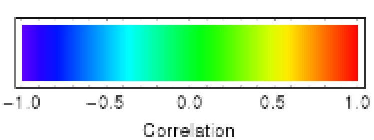
$\bar{s}$ (x,Q) at Q=125 GeV at 68% and 95% c.l.
META300 (solid), CMC (dashed)

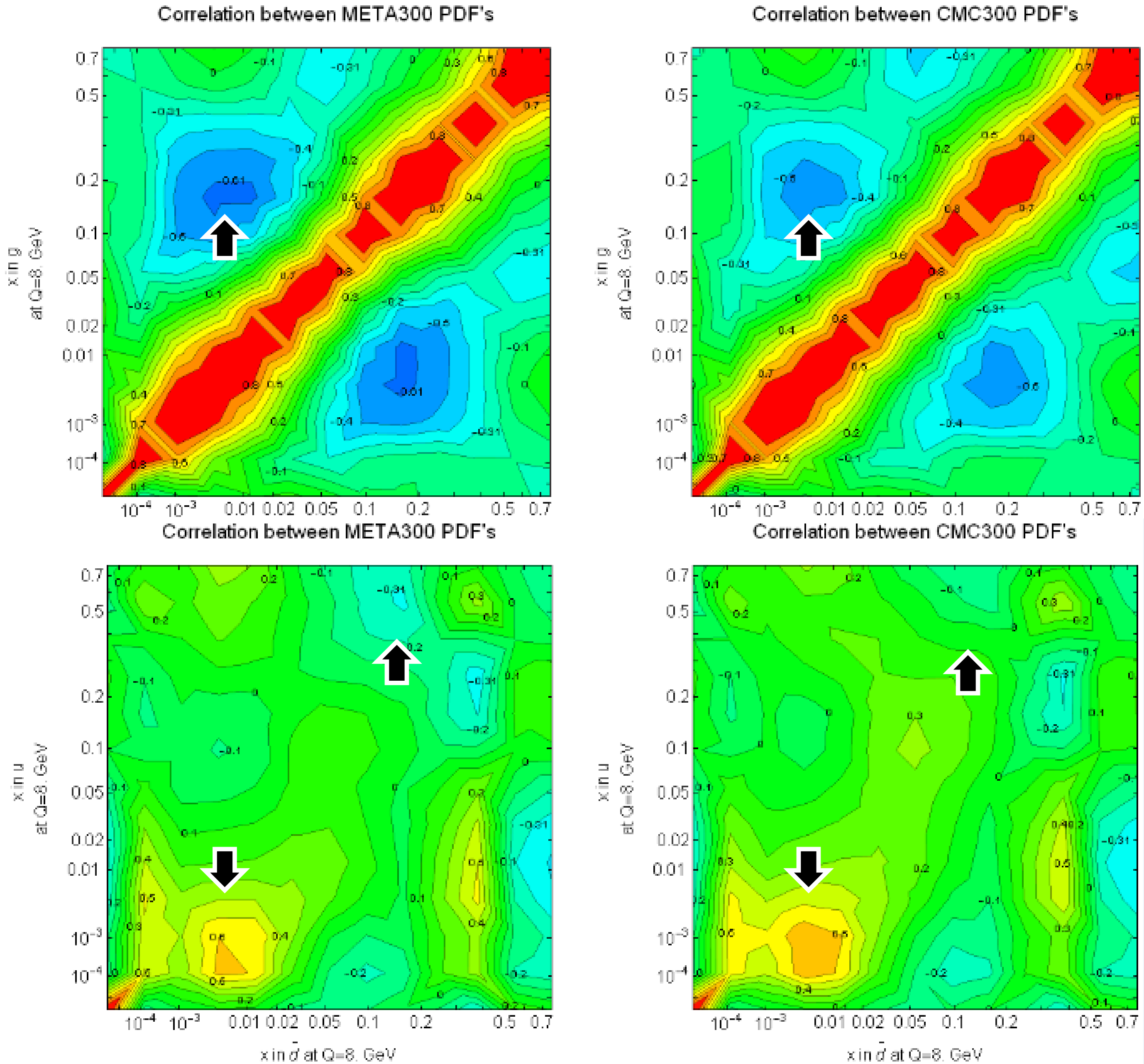


PDF-PDF correlations in 300 replica ensembles

PDF-PDF correlations are broadly consistent in two MC-300 ensembles.

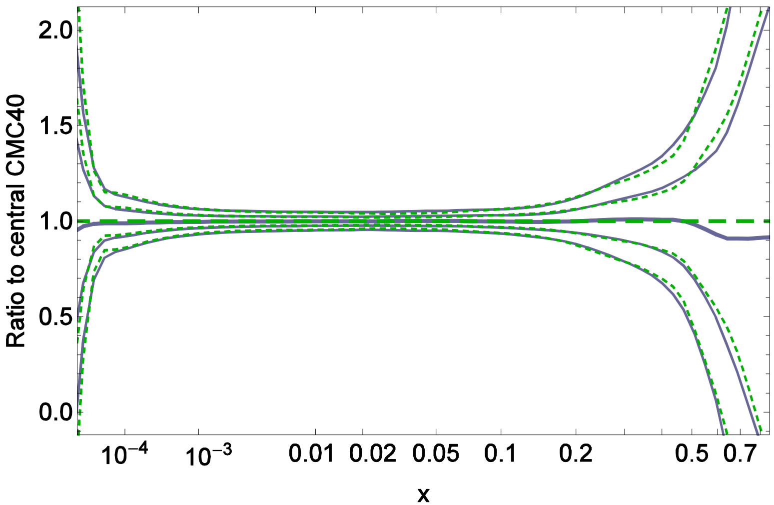
The correlation coefficient can differ by up to ± 0.25 for sea PDFs as a resulted of random variations in the MC replica generation (see the regions pointed by the arrows).



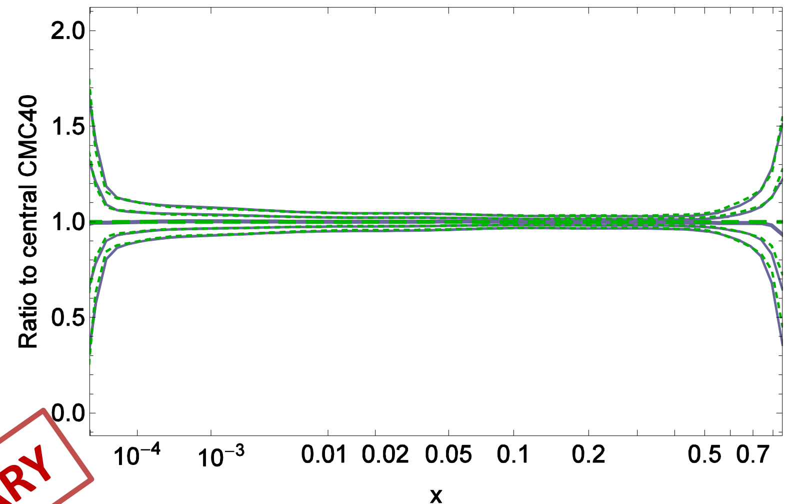


Reduction, CMC ensemble: 300 → 40 replicas

g (x,Q) at Q=8 GeV at 1σ and 2σ
CMC40 (dashed), CMC300 (solid)

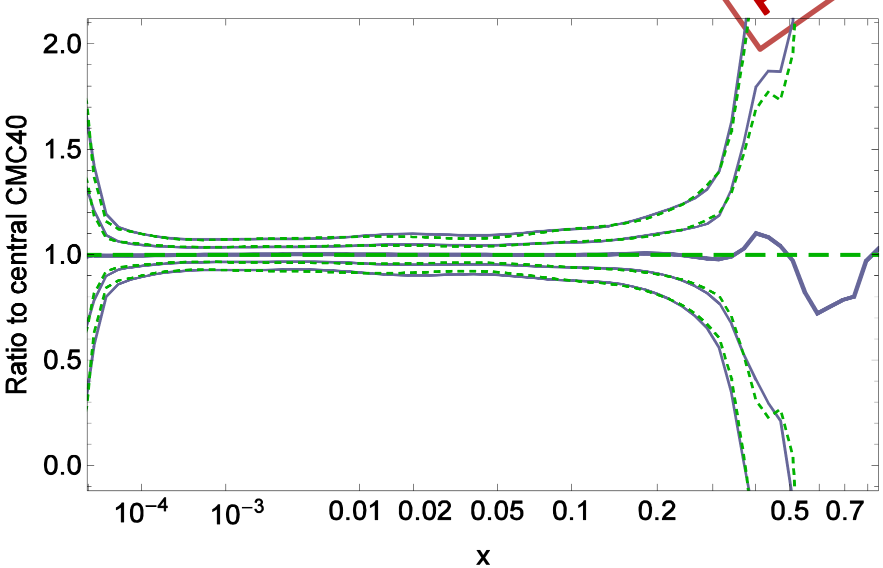


u (x,Q) at Q=8 GeV at 1σ and 2σ
CMC40 (dashed), CMC300 (solid)

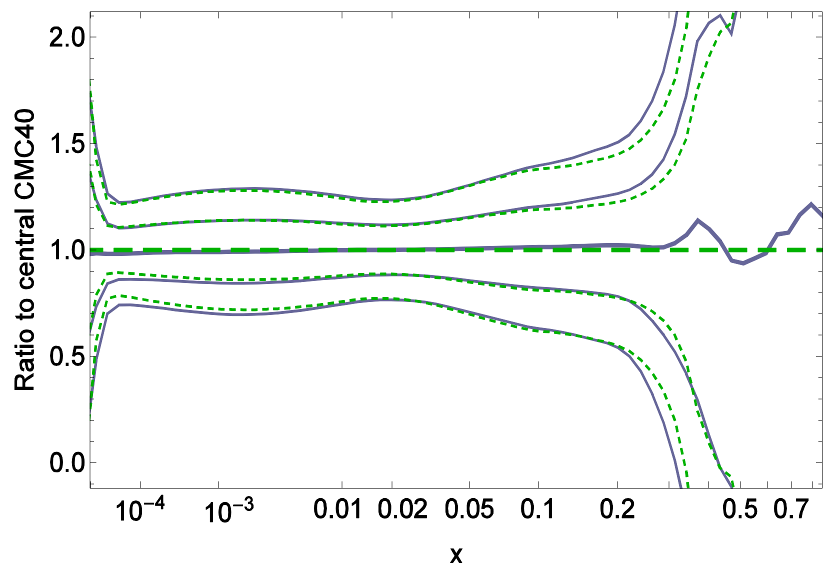


PRELIMINARY

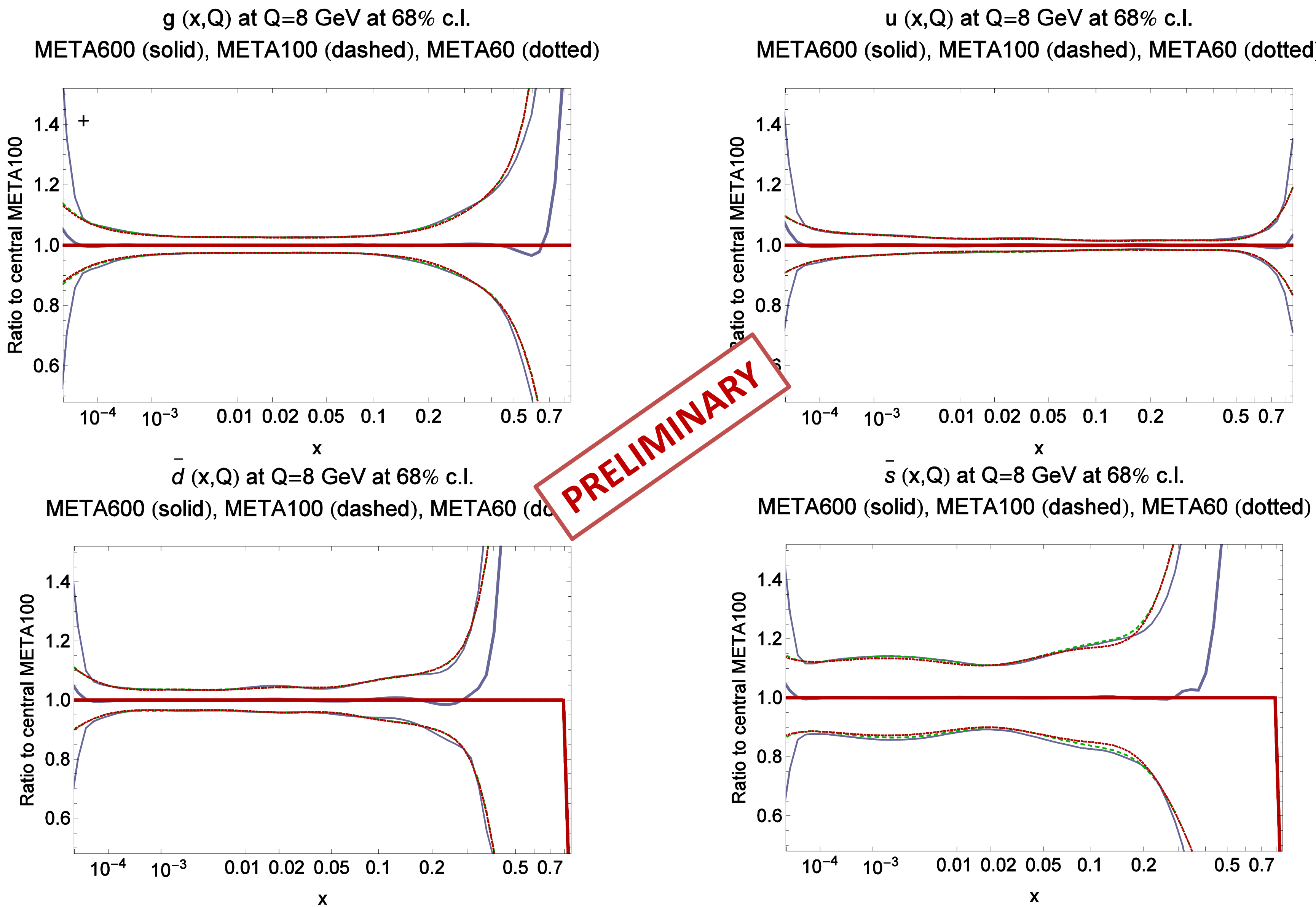
$\bar{d}$ (x,Q) at Q=8 GeV at 1σ and 2σ
CMC40 (dashed), CMC300 (solid)



$\bar{s}$ (x,Q) at Q=8 GeV at 1σ and 2σ
CMC40 (dashed), CMC300 (solid)



Reduction, META ensemble: 600 → 100 → 60 error sets

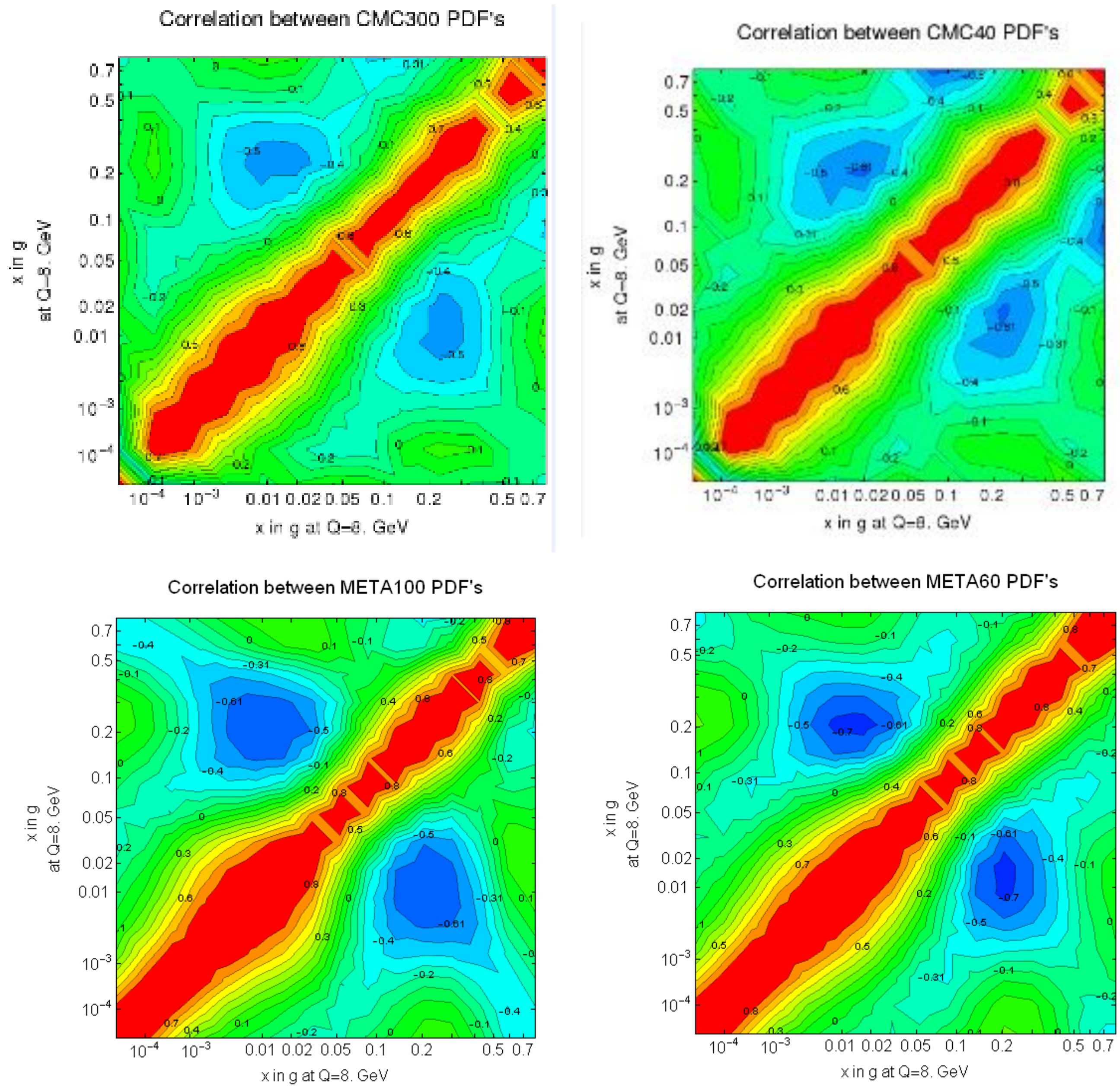


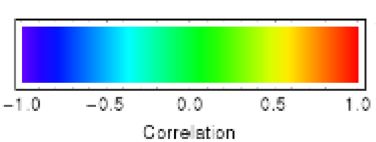
PDF-PDF correlations in reduced ensembles

g(8 GeV)-g(8 GeV)

Correlations are computed using two different formulas in the MC and Hessian ensembles, producing close, but not identical, values.

Reduction of replicas broadly preserves the patterns of correlations.

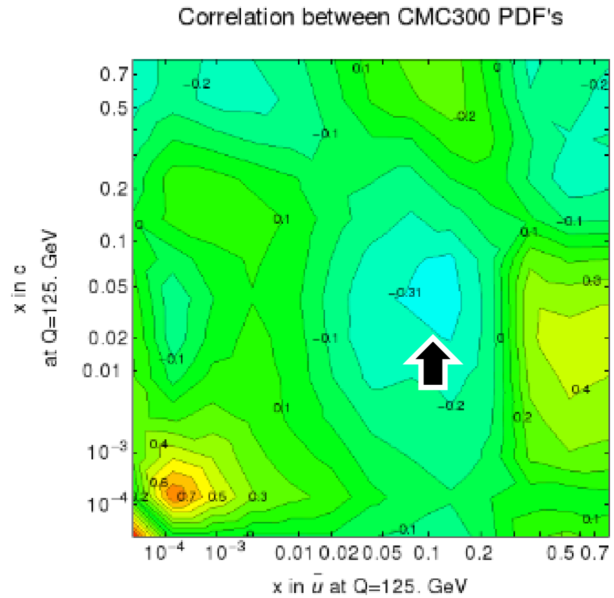


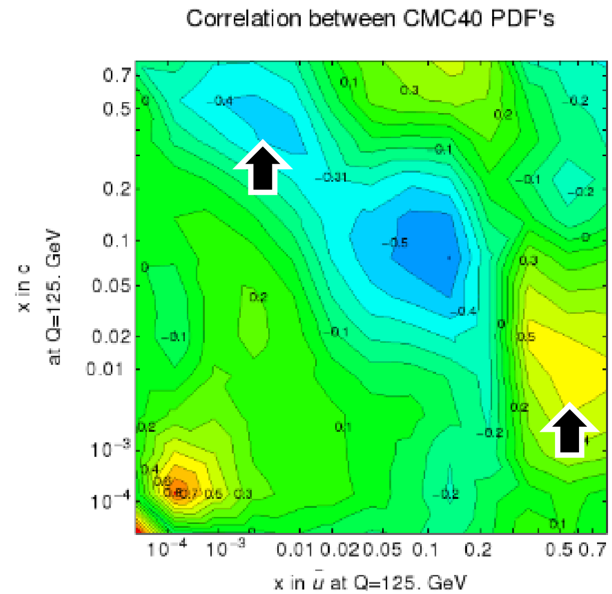


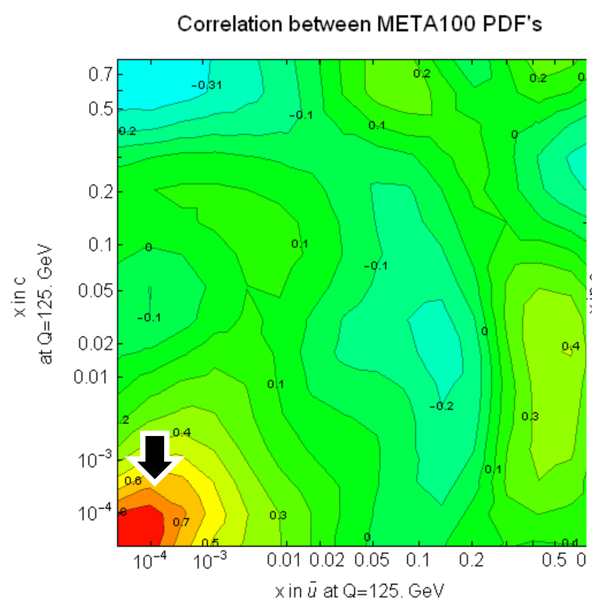
PDF-PDF correlations in reduced ensembles

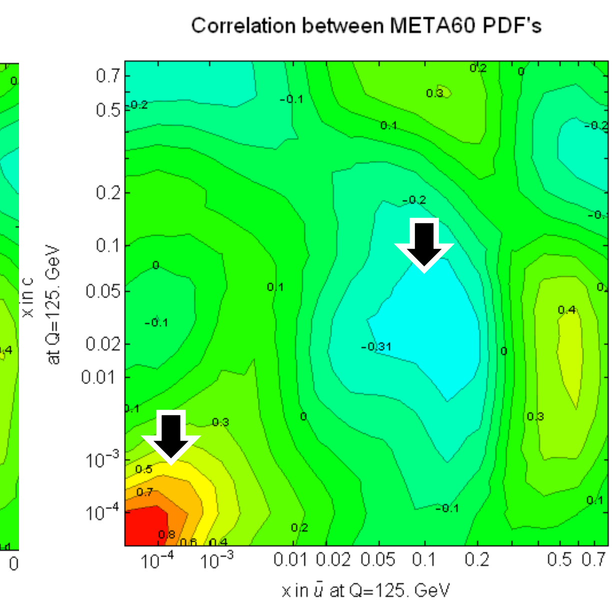
$\bar{u}$(125 GeV)-c(125 GeV)

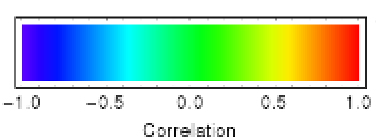
Again, we see variations in correlations among the full and reduced ensembles, but in the (x,Q) regions constrained by the data, they are of order of statistical noise.







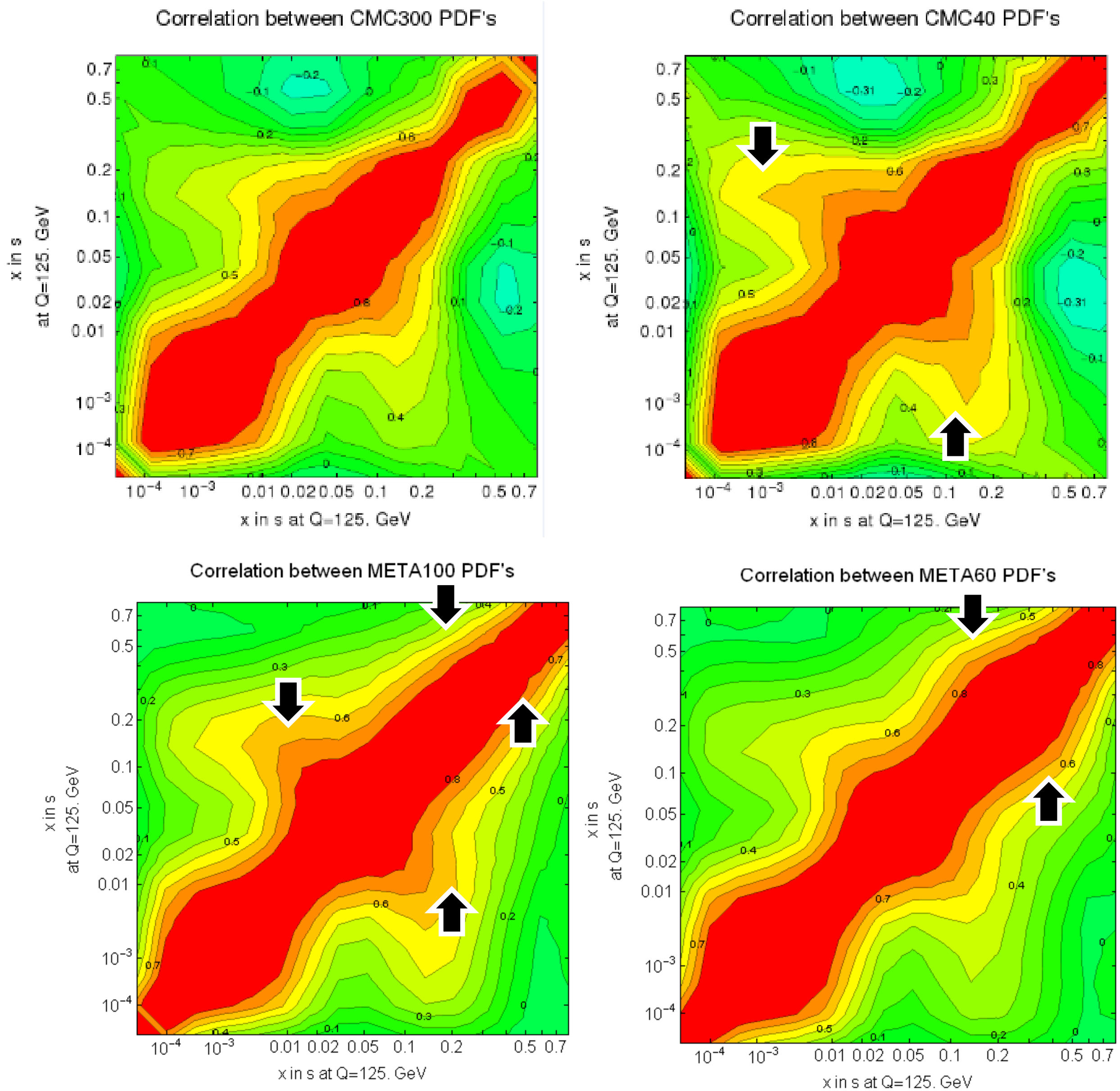


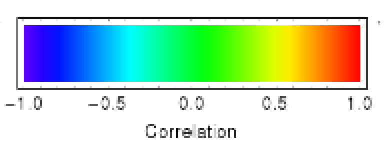


PDF-PDF correlations in reduced ensembles

s(125 GeV)-s(125 GeV)

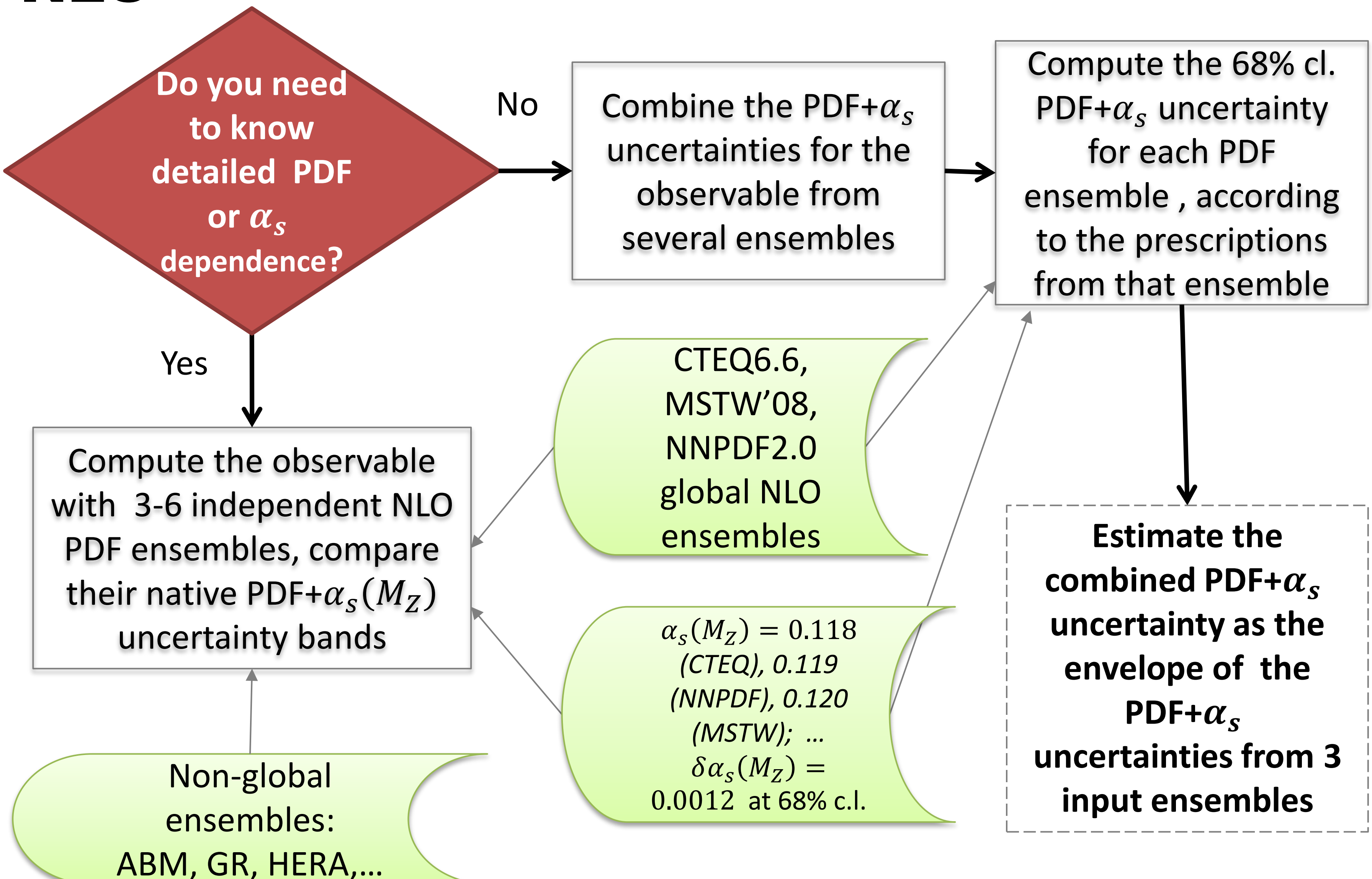
Variations in correlations are less predictable for poorly constrained PDF flavors, such as strangeness. This is not likely to cause problems in applications of CMC/META PDFs.





Backup slides

2010 PDF4LHC recommendation for an LHC observable: NLO



M. Botje et al., arXiv:1101.0538

2010 PDF4LHC recommendation for an LHC observable: NNLO

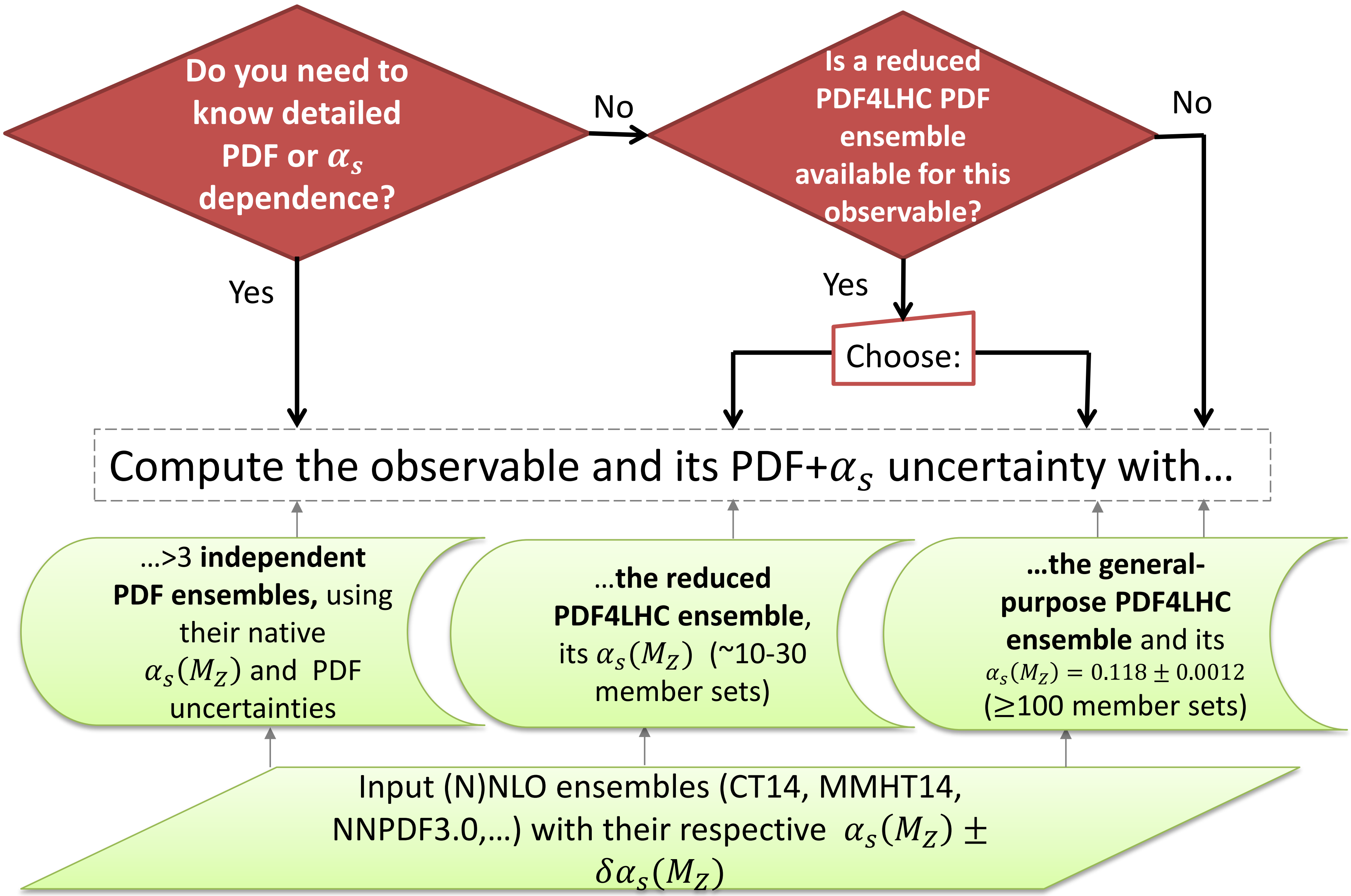
2010: M. Botje et al., arXiv:1101.0538

As a central value, use the MSTW08 prediction. As an uncertainty, take the same percentage uncertainty on this NNLO prediction as found using the NLO uncertainty prescription given above. However, if the interest of the user is in the difference of an NLO and NNLO prediction, rather than in the uncertainties, then the same PDF must be used for both evaluations. The current recommendation would be to use the MSTW08 predictions for both.

2014, Note added (see also Ball et al., arXiv:1211.5142)

Since the first documented recommendation several more up to date official versions of PDF sets became available and we therefore update the recommendation above to use these data sets, e.g., the use of CT10 instead of CTEQ6.6, NNPDF2.3 instead of NNPDF2.1. We also extend the same [NLO PDF4LHC] recommendation to use these sets for NNLO uncertainty bands.

A concept for a new PDF4LHC recommendation



This procedure applies both at NLO and NNLO

Near-term goals (on the time scale of months)

1. Spell out compatibility conditions for the combination.
   Which conditions must an input ensemble meet in order to be included into the combination?

2. Perform benchmark comparisons of two reduction methods for 3 global ensembles
   How many reduced replicas are needed in order to reproduce the PDF uncertainty of the full input MC ensemble at 1% or 5% accuracy?

3. Explore ideas for including non-global ensembles (least understood)

Volunteers are welcome to contribute to every task. For the moment, we have manpower to focus on 2 and maybe on 1.

1. Basic compatibility conditions

Theory

Order of α_s. Counting of α_s orders. Example: non-leading structure functions such as F_L.
Type of QCD calculation (fixed-order only; resummations; higher-twist corrections; …)
Heavy-flavor schemes, number of active flavors, pole/$\overline{MS}$ masses,…
Values of input parameters: $\alpha_s(M_z), m_c, m_b$ …
Compatibility of numerical DGLAP evolution
…

Experiment

compatibility of experimental data sets
with or without certain LHC measurements
…

Statistics

Definitions of the likelihood, PDF uncertainty, symmetric/asymmetric errors…
PDF parametrizations (functional form, asymptotic limits) and theoretical priors (positivity, constraints on flavor ratios, …)
Compatibility of PDF central values, confidence intervals, …
Treatment of experimental and theoretical systematic uncertainties
…