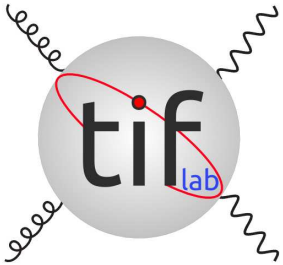


CMC & MC2Hessian:

UPDATES, COMPARISONS, IMPROVEMENTS

Stefano Forte

Università di Milano & INFN

in collaboration with S. Carrazza, Z, Kassabov and J. Rojo

UNIVERSITÀ DEGLI STUDI
DI MILANO

PDF4LHC

CERN, June 26, 2015

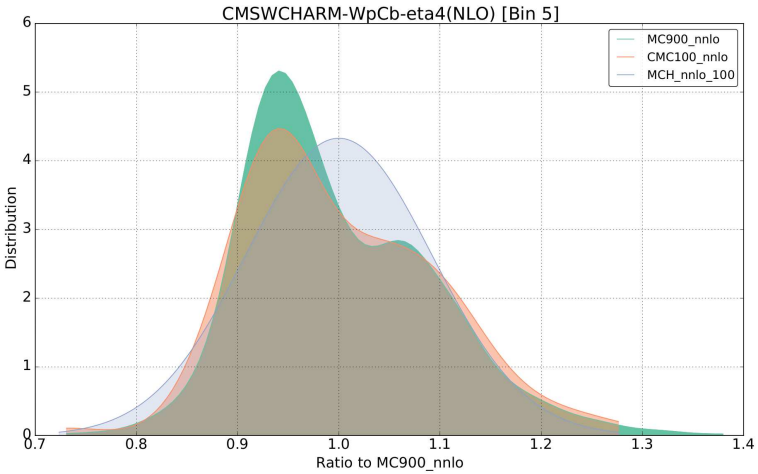
CMC PDFs AND NONGAUSSIAN BEHAVIOUR

DEVIATION FROM GAUSSIANITY OBSERVED

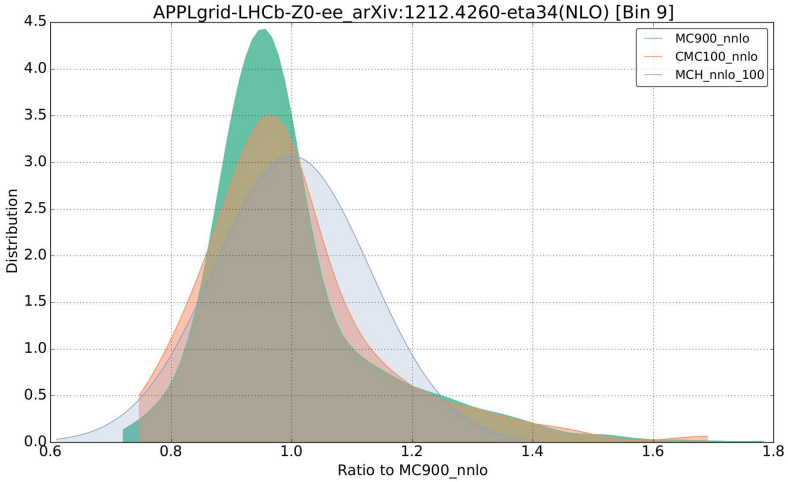
E.G. AT LARGE x, DUE TO LARGE UNCERTAINTY & POSITIVITY BOUNDS (MAY BE RELEVANT FOR SEARCHES)

MONTE CARLO COMPARED TO HESSIAN

CMS $W + c$ production



LHCb electrons



- MONTE CARLO PDFs REPRODUCE NONGAUSSIAN, HESSIAN FAIL
- PROBLEM DOES NOT ARISE FOR BULK OF DATA $\Rightarrow$ HESSIAN ADEQUATE!

NOTE: ALL PLOTS PRODUCED USING FINAL COMBINED CT14-MMHT-NNPDF3.0 MC900 REPLICA SET & META300 PRIOR AS COMMUNICATED BY JUN GAO

SEE LAST SLIDE FOR LINK TO FULL CATALOGUE OF PLOTS

THE MC2H IDEA:

USE REPLICAS AS BASIS FUNCTIONS (CONTINUITY, SUM RULES, DGLAP AUTOMATICALLY IMPLEMENTED)

MC2H-GA

(as described at previous meeting, & `arXiv:1505.06736`)

- SINGLE OUT GAUSSIAN REGION (ONE SIGMA = 68% C.L.)
- OPTIMIZE UNCERTAINTIES VIA GENETIC ALGORITHM

CRITICISM

(talk by Pavel at previous meeting)

- CHOICE OF SAMPLING AND GA INTRODUCE BIAS
- NONGAUSSIANITIES IRRELEVANT: REPRODUCE COVARIANCE MATRIX AS ACCURATELY AS POSSIBLE

THE MC2H IDEA:

USE REPLICAS AS BASIS FUNCTIONS (CONTINUITY, SUM RULES, DGLAP AUTOMATICALLY IMPLEMENTED)

MC2H-GA

(as described at previous meeting, & `arXiv:1505.06736`)

- SINGLE OUT GAUSSIAN REGION (ONE SIGMA = 68% C.L.)
- OPTIMIZE UNCERTAINTIES VIA GENETIC ALGORITHM

CRITICISM

(talk by Pavel at previous meeting)

- CHOICE OF SAMPLING AND GA INTRODUCE BIAS
- NONGAUSSIANITIES IRRELEVANT: REPRODUCE COVARIANCE MATRIX AS ACCURATELY AS POSSIBLE

REPLY: MC2H-PCA

(as presented in `arXiv:1505.06736`)

- REPRODUCE COVARIANCE MATRIX ON VERY FINE SET OF GRID POINTS
- REPRESENT IT BY SINGULAR-VALUE DECOMPOSITION ON REPLICAS
- PICK LARGEST CONTRIBUTIONS (PRINCIPAL COMPONENT ANALYSIS)

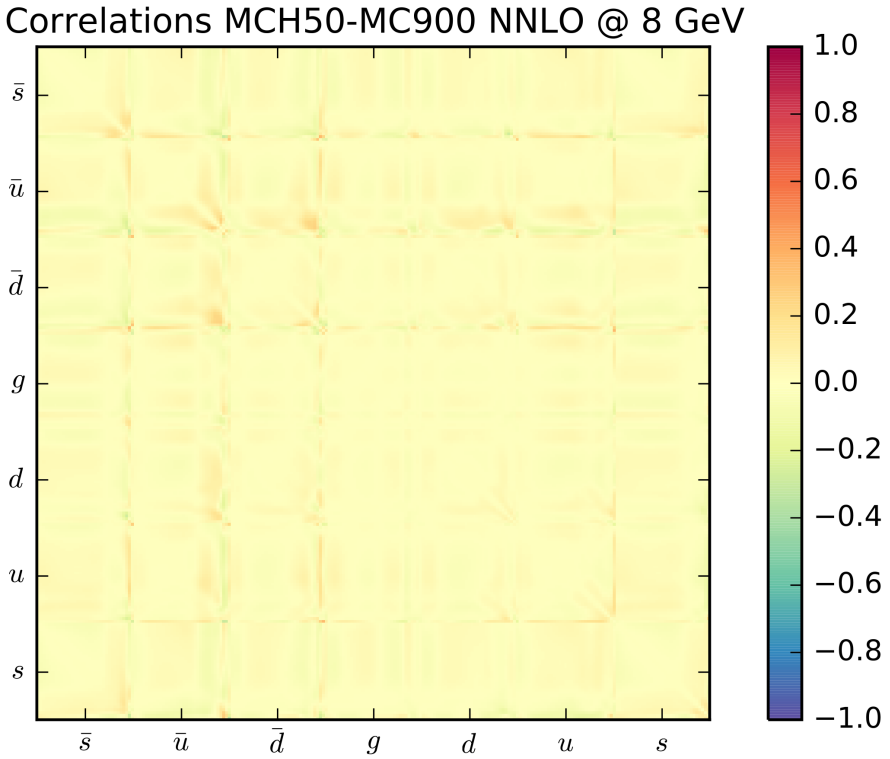
$\Rightarrow$ COVARIANCE MATRIX PERFECTLY REPRODUCED; NO CHOICES: NO BIAS

MC2H-PCA vs Meta-PDFs

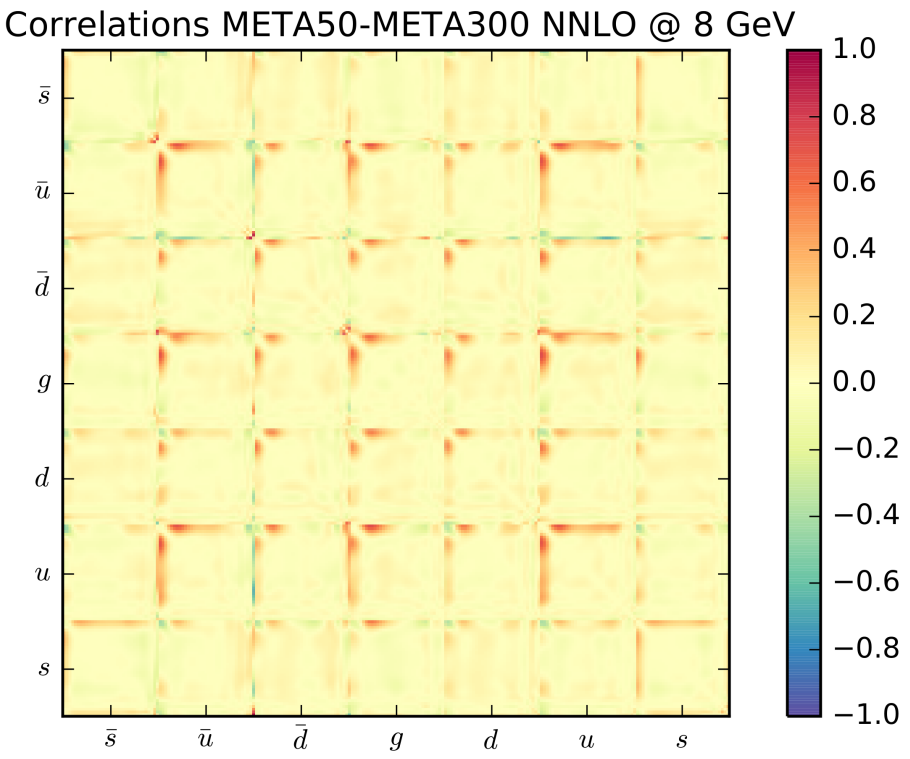
- prior combination of CT14, MMHT, NNPDF3.0: MC900 or META300
- MC2H-PCA: sets of 100 or 120 Hessian eigenvectors
- Meta-PDFs: v2 sets of 2*30 and 2*50 eigenvectors

THE CORRELATION MATRIX:
PRIOR-FINAL (LOW RESOLUTION)

MC2H



META



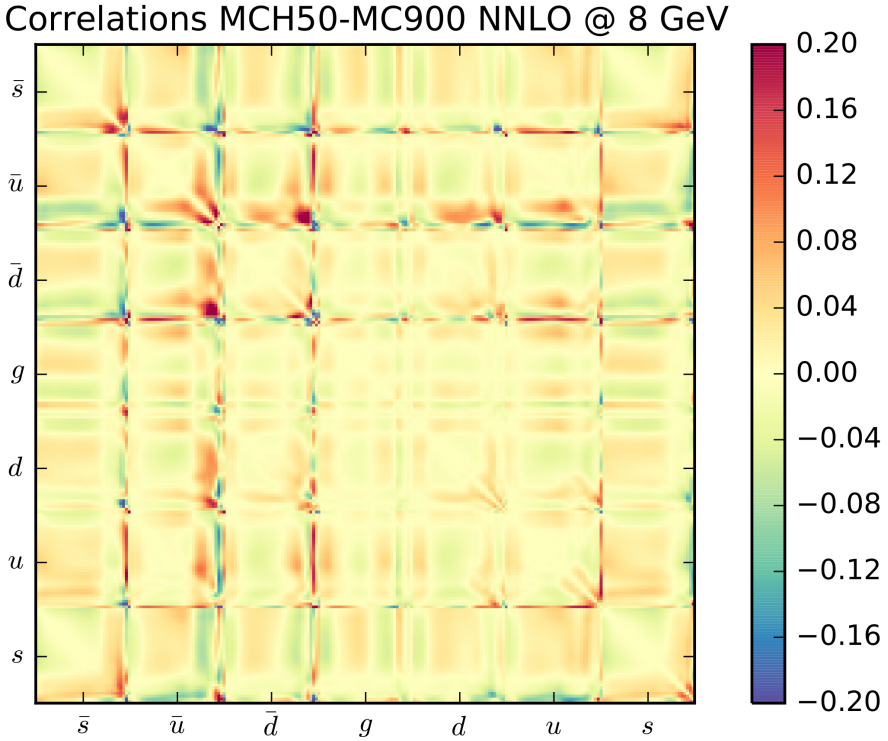
MC2H-PCA vs Meta-PDFs

- prior combination of CT14, MMHT, NNPDF3.0: MC900 or META300
- MC2H=PCA: sets of 100 or 120 Hessian eigenvectors
- Meta-PDFs: v2 sets of 2*30 and 2*50 eigenvectors

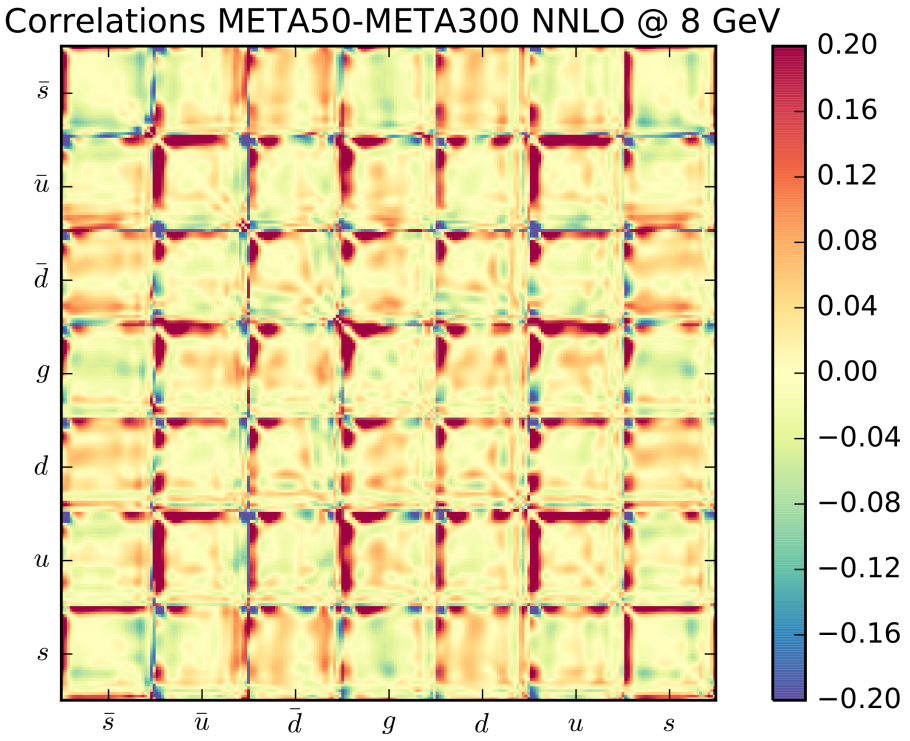
THE CORRELATION MATRIX:

PRIOR-FINAL (MEDIUM RESOLUTION)

MC2H



META

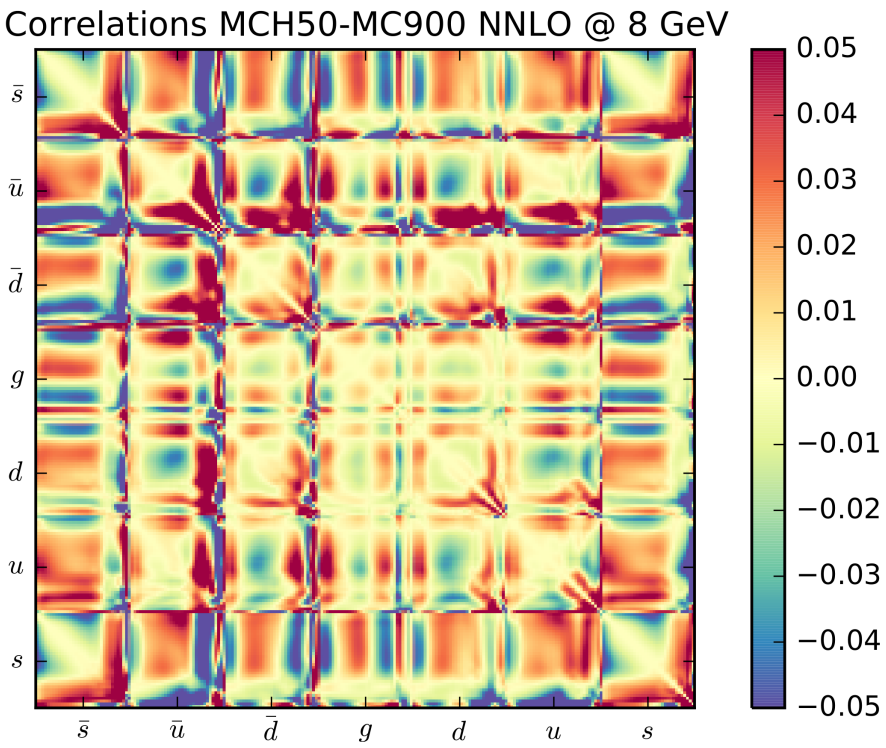


MC2H-PCA vs Meta-PDFs

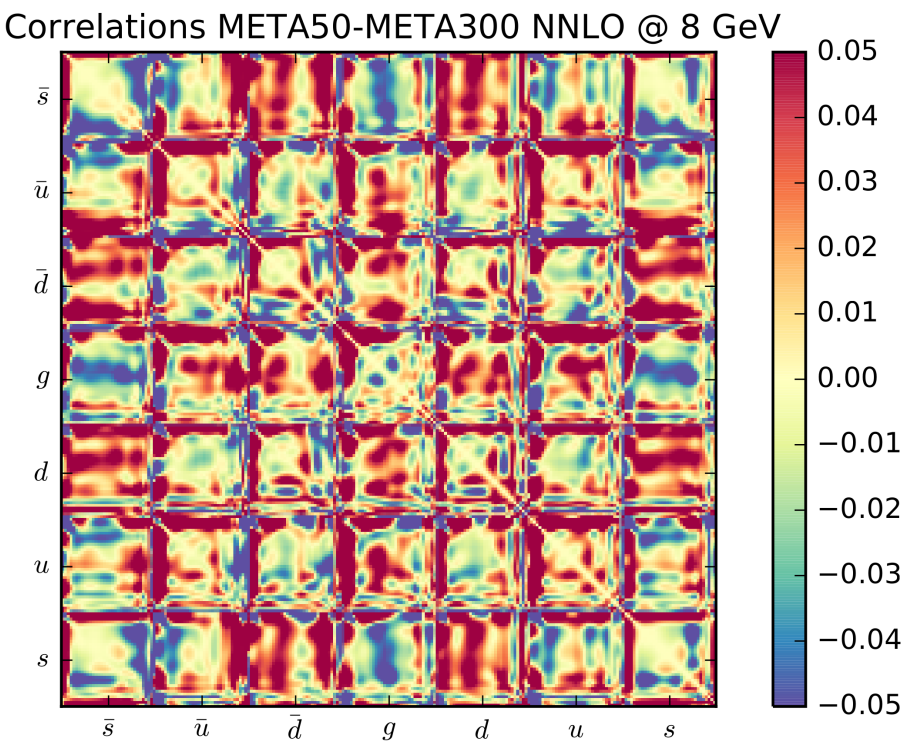
- prior combination of CT14, MMHT, NNPDF3.0: MC900 or META300
- MC2H=PCA: sets of 100 or 120 Hessian eigenvectors
- Meta-PDFs: v2 sets of 2*30 and 2*50 eigenvectors

THE CORRELATION MATRIX:
PRIOR-FINAL (HIGH RESOLUTION)

MC2H



META



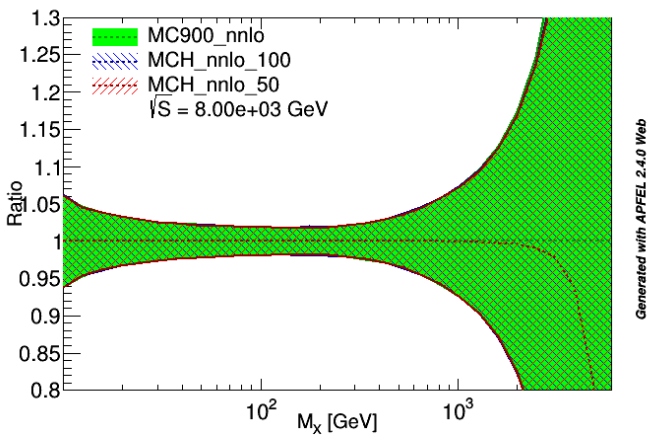
MC2H achieves percent accuracy

MC2H-PCA vs META-PDFs

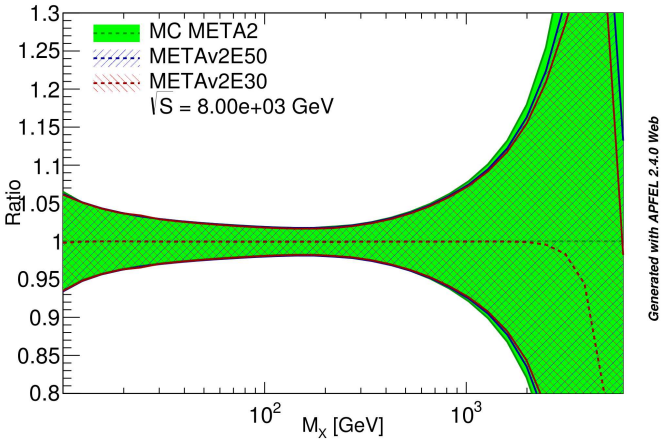
LUMINOSITIES

GLUON-GLUON

MC2H

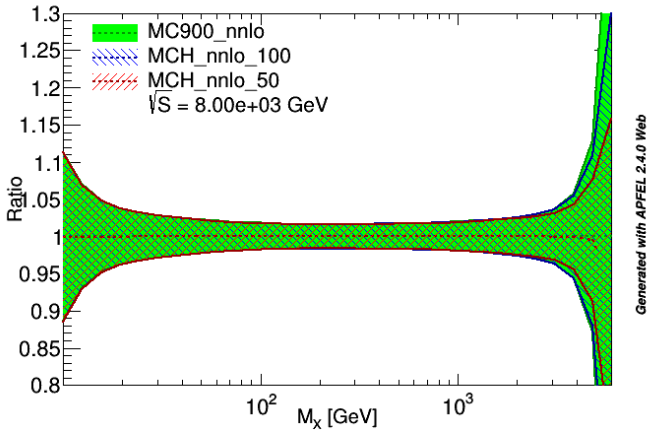


META

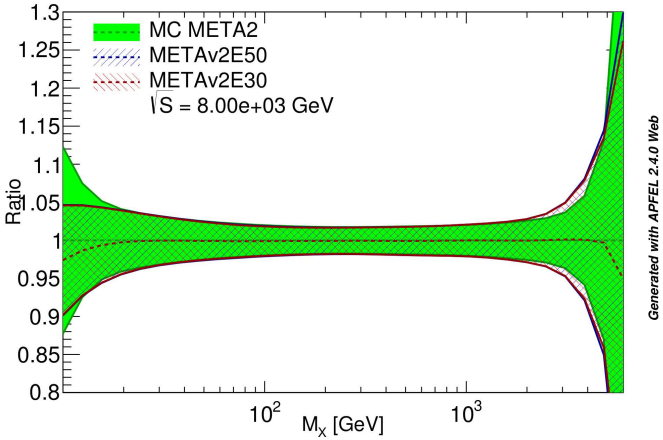


QUARK-QUARK

MC2H



META



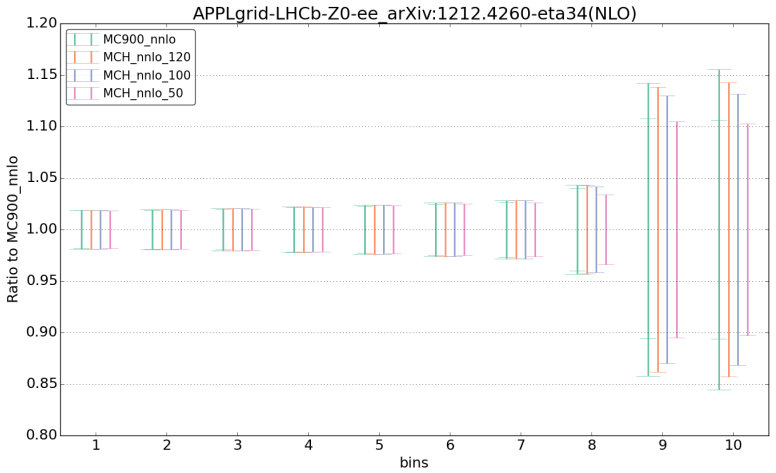
META: DISCREPANCIES OBSERVED AT LARGE/SMALL x

MC2H-PCA vs META-PDFs

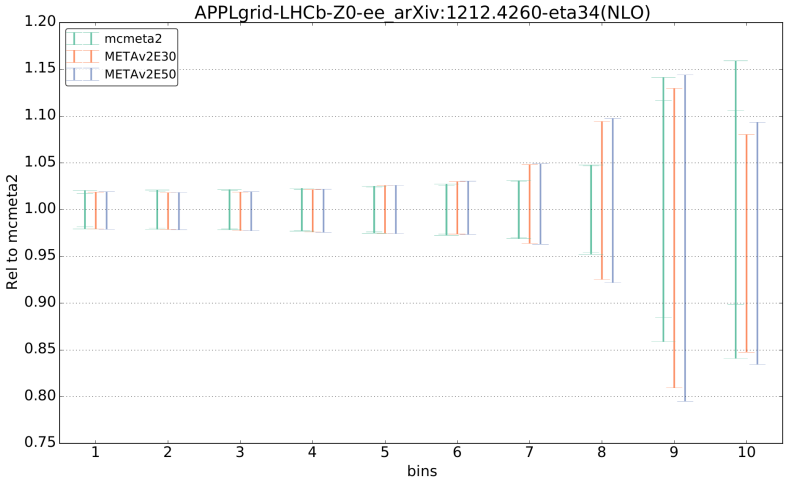
OBSERVABLES

LHCb μ DISTRIBUTION

MC2H

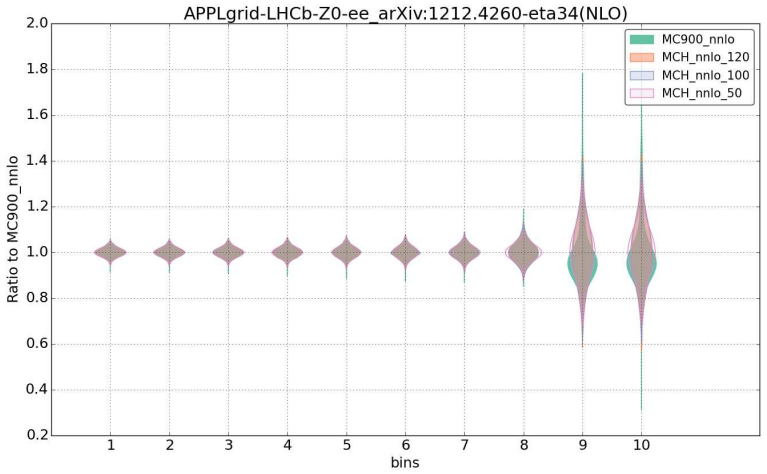


META

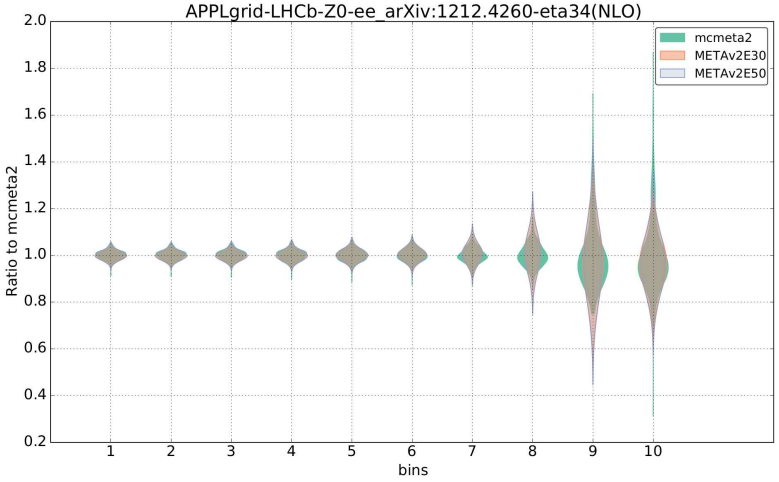


probability distribution of results

MC2H



META



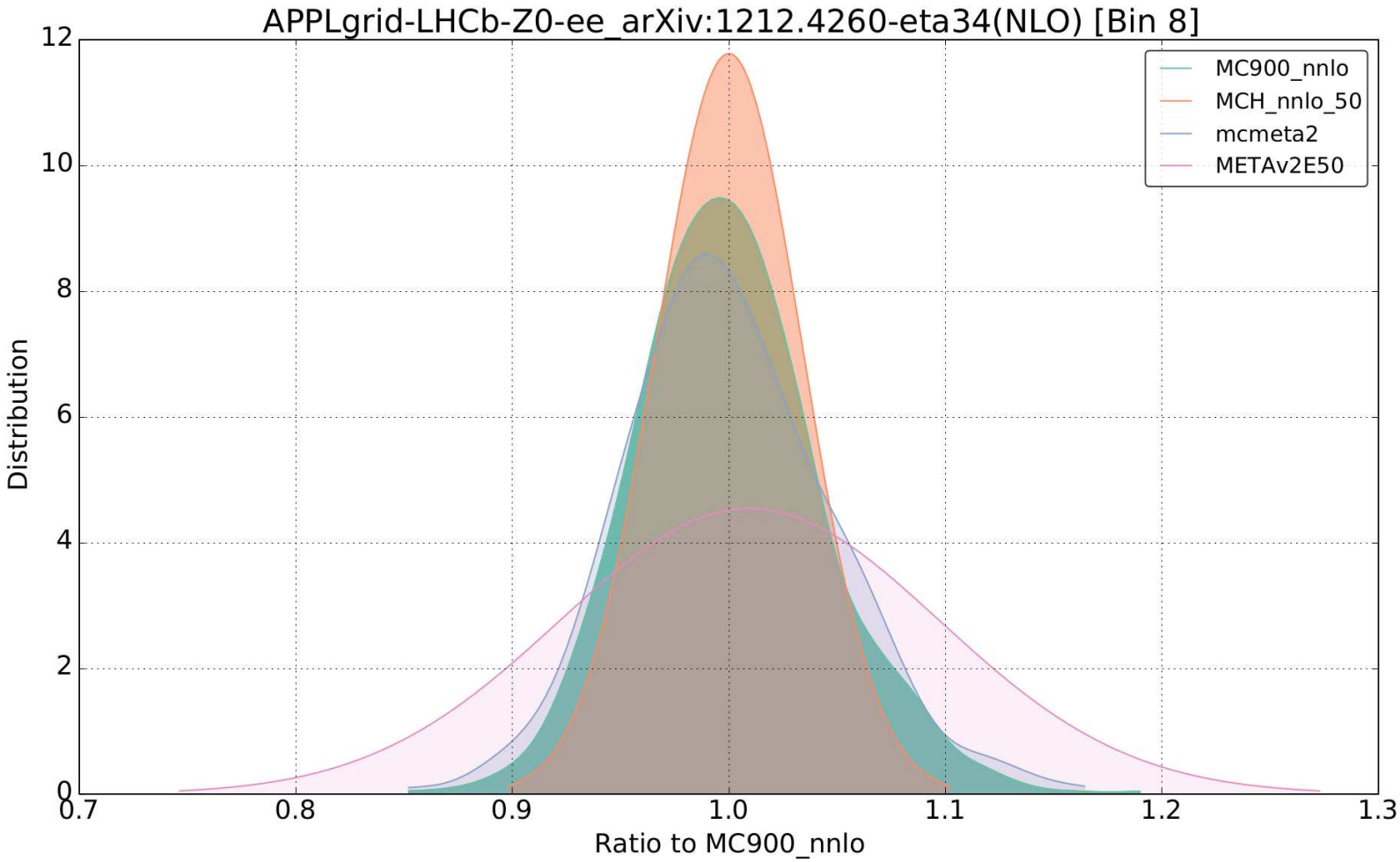
META: DISCREPANCIES OBSERVED IN OUTER KINEMATIC REGIONS

MC2H-PCA vs META-PDFs

OBSERVABLES

LHCb μ DISTRIBUTION

ZOOM ON BIN 8:



DISCREPANCIES (META) NOT RELATED TO NON-GAUSSIANITY

PARAMETRIZATION BIAS?

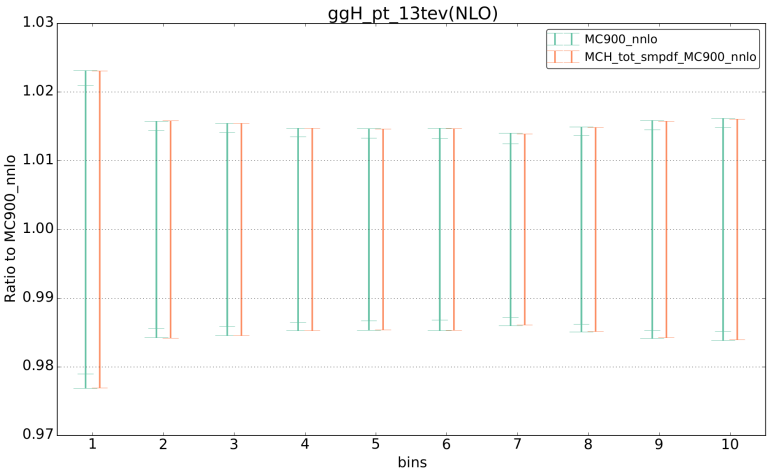
MINIMAL SETS: SM-PDFS

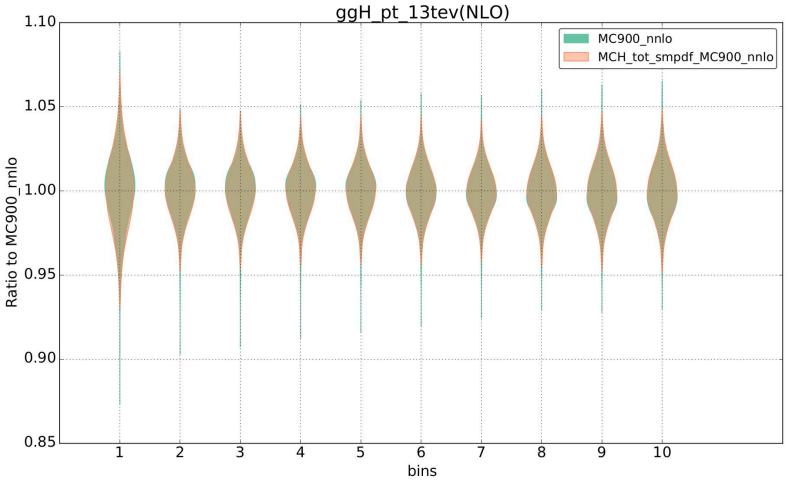
- GIVEN ONE OR MORE PROCESSS ("SIGNAL"), SELECT EIGENVECTORS WHICH PROVIDE THE DOMINANT CONTRIBUTION:
  - DETERMINE SUBSET OF GRID POINTS WHICH HAVE THE HIGHEST CORRELATION TO THE PROCESS
  - PERFORM PCA ANALYSIS ON THE CORRESPONDING SUBMATRIX
  - NOTE USAGE OF SUBGRID GUARANTEES STABILITY: SIMILAR PROCESSES WILL ALSO BE WELL REPRODUCED (EXAMPLE: TOP VS HIGGS)
- .
- .

HIGGS

13 eigenvectors

HIGGS p_T DISTRIBUTION IN GLUON FUSION





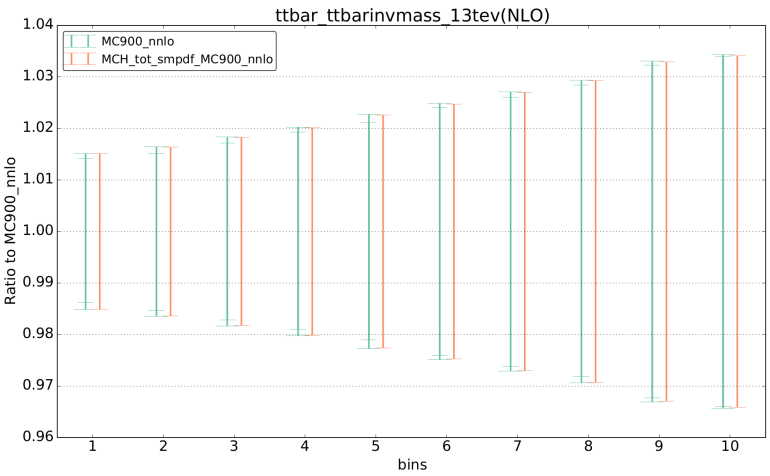
MINIMAL SETS: SM-PDFS

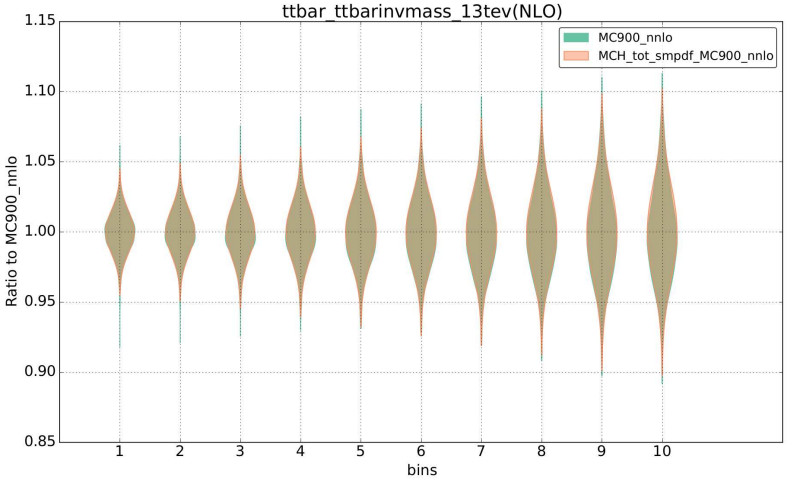
- GIVEN ONE OR MORE PROCESSS ("SIGNAL"), SELECT EIGENVECTORS WHICH PROVIDE THE DOMINANT CONTRIBUTION:
  - DETERMINE SUBSET OF GRID POINTS WHICH HAVE THE HIGHEST CORRELATION TO THE PROCESS
  - PERFORM PCA ANALYSIS ON THE CORRESPONDING SUBMATRIX
  - NOTE USAGE OF SUBGRID GUARANTEES STABILITY: SIMILAR PROCESSES WILL ALSO BE WELL REPRODUCED (EXAMPLE: TOP VS HIGGS)
- ADD ONE OR MORE PROCESSES ("BACKGROUND"): REPEAT ANALYSIS & PICK FURTHER EIGENVECTORS
- .

HIGGS+TOP

14 eigenvectors

TOP INVARIANT MASS DISTRIBUTION





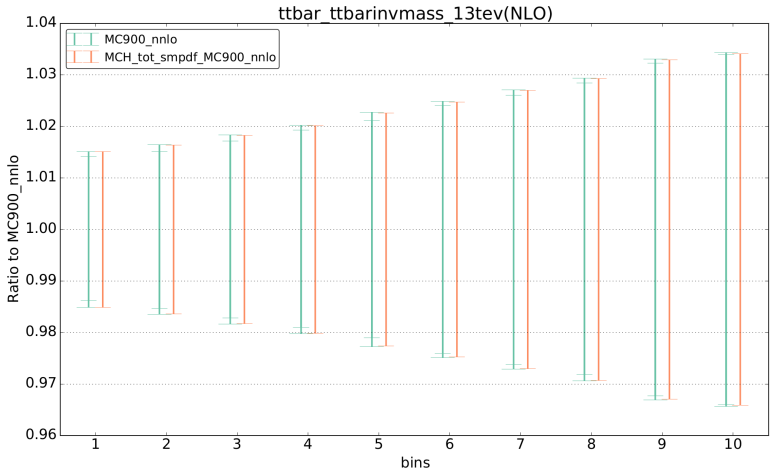
MINIMAL SETS: SM-PDFS

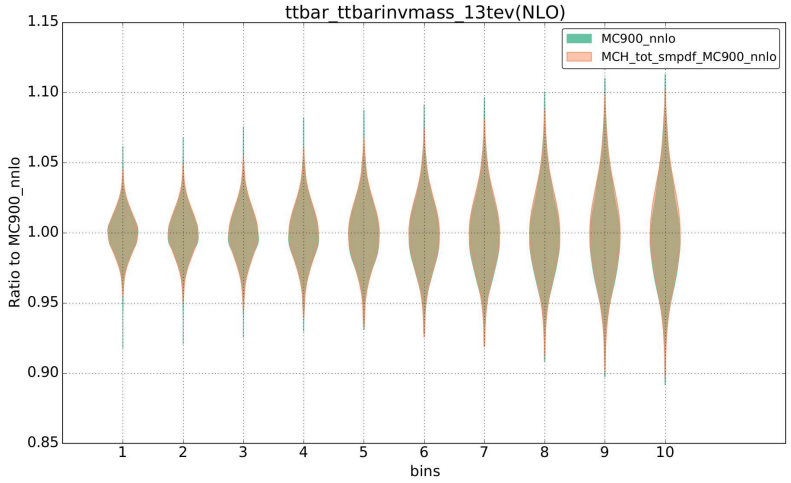
- GIVEN ONE OR MORE PROCESSS ("SIGNAL"), SELECT EIGENVECTORS WHICH PROVIDE THE DOMINANT CONTRIBUTION:
  - DETERMINE SUBSET OF GRID POINTS WHICH HAVE THE HIGHEST CORRELATION TO THE PROCESS
  - PERFORM PCA ANALYSIS ON THE CORRESPONDING SUBMATRIX
  - NOTE USAGE OF SUBGRID GUARANTEES STABILITY: SIMILAR PROCESSES WILL ALSO BE WELL REPRODUCED (EXAMPLE: TOP VS HIGGS)
- ADD ONE OR MORE PROCESSES ("BACKGROUND"): REPEAT ANALYSIS & PICK FURTHER EIGENVECTORS
- CAN ITERATE AT WILL $\Rightarrow$ EVENTUALLY RECOVER FULL EIGENVECTOR SET

HIGGS+TOP+Z+W

15 eigenvectors

W p_T DISTRIBUTION





DELIVERY

- FULL PHENO ANALYSIS AVAILABLE (more than 1000 pheno plots) FROM
  `http://pcteserver.mi.infn.it/ñnpdf/mc2h-gallery/website`
- MC2H (PCA) PYTHON CODE PUBLICLY AVAILABLE FROM
  `https://github.com/scarrazza/mc2hessian`
- CUSTOMIZABLE SM-PDF PUBLICLY AVAILABLE FROM
  `https://github.com/scarrazza/smpdf`
- EFFICIENT AUTOMATING SCRIPTING FOR PHENO APPLICATIONS
- MC2H & SM-PDFs TO BE INCLUDED IN `APFEL Web`, WITH CUSTOMIZABLE CHOICE OF SIGNAL AND BACKGROUND PROCESSES

MC2H: ADVANTAGES

- NO FUNCTIONAL FORM & NO FITTING REQUIRED
- ARBITRARILY HIGH ACCURACY ON FULL COVARIANCE MATRIX;
  PERCENT LEVEL WITH 100 EIGENVECTORS
- SM MINIMAL SETS CAN BE PRODUCED WITH RECOVERABLE ACCURACY
  AND ARBITRARY CHOICE OF PROCESSES
- PYTHON CODE ALREADY PUBLICLY AVAILABLE
  WITH EFFICIENT SCRIPTING FOR PHENO APPLICATIONS
- APFEL WEB INTERFACE SOON TO BE AVAILABLE

"Meta-PDFS and MC2H (...) both realize variations of a generic meta-parametrization method" (P. Nadolsky, PDF4LHC June 4 meeting)