Parton Distributions in the Higgs Boson Era

Juan Rojo
CERN, PH Division, TH Unit

LAPTH Annecy-les-Vieux
15/11/2012

Parton Distributions and LHC phenomenology

Proton-Proton collisions at the LHC

Our ability to **exploit the LHC potential** depends on the understanding of the **various processes** that take place in proton proton collisions

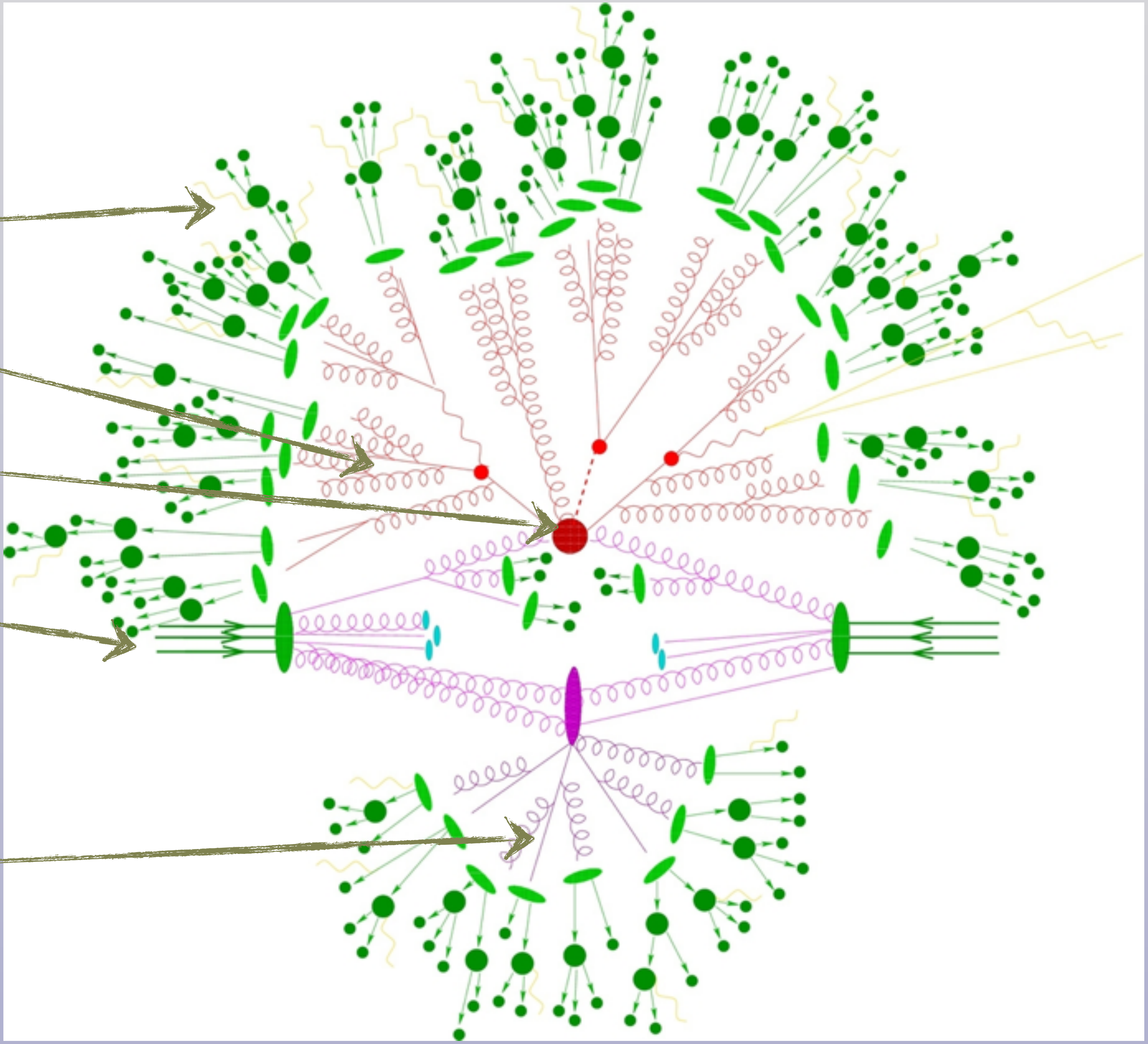
Hadronization: Modeling + Tunes to Data

Parton Showering and Matching: pQCD + Modeling

Hard-Scattering Matrix Elements: perturbative QCD (pQCD) + EW theory

Parton Distribution Functions: pQCD + Data + Methodology

Multiple Interactions, Pile-Up: Modeling



Our abi
potentia
understa
processe
proton c

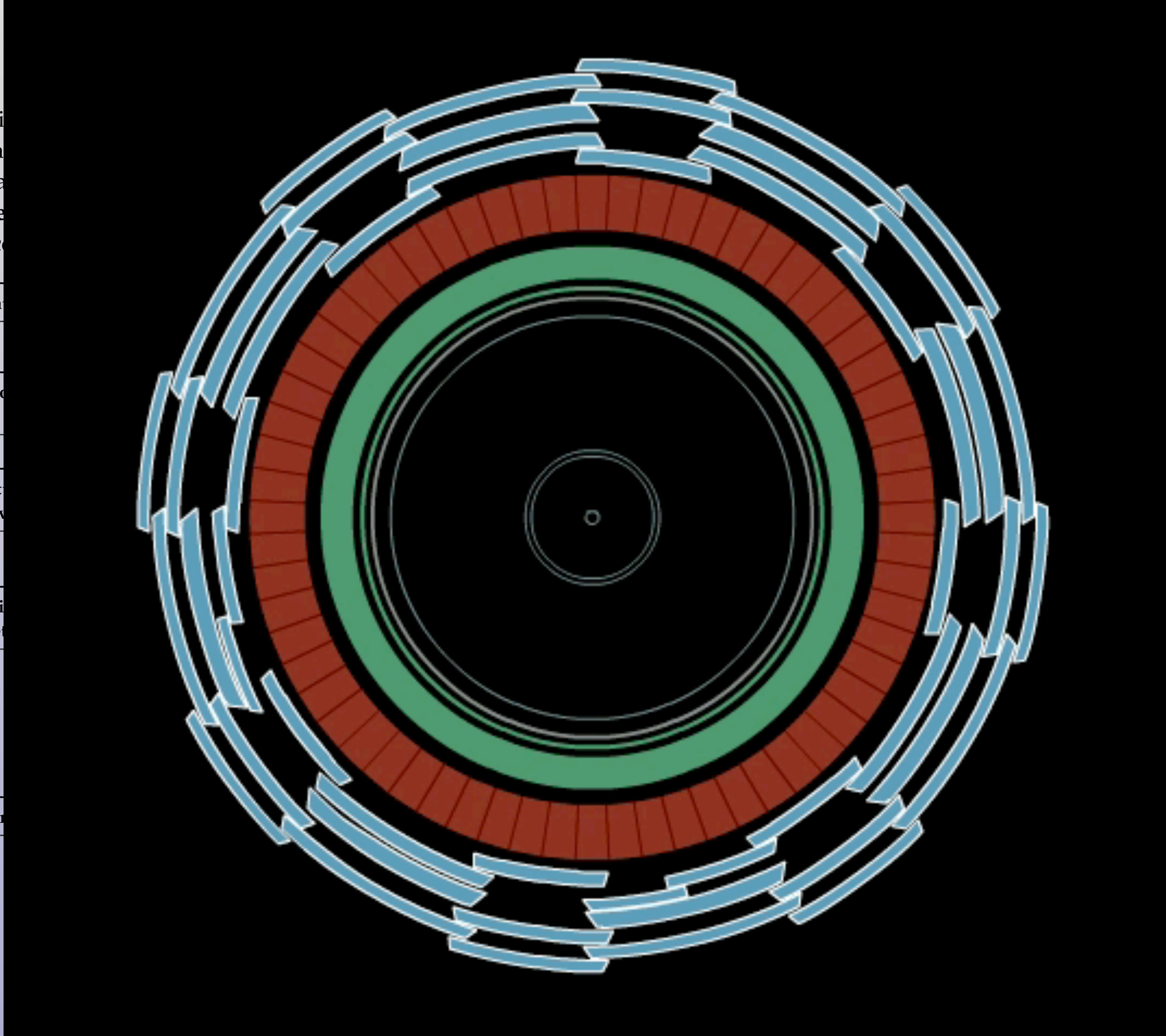
Hadroniza

Parton Sho
Modeling

Hard-Scat
perturbativ

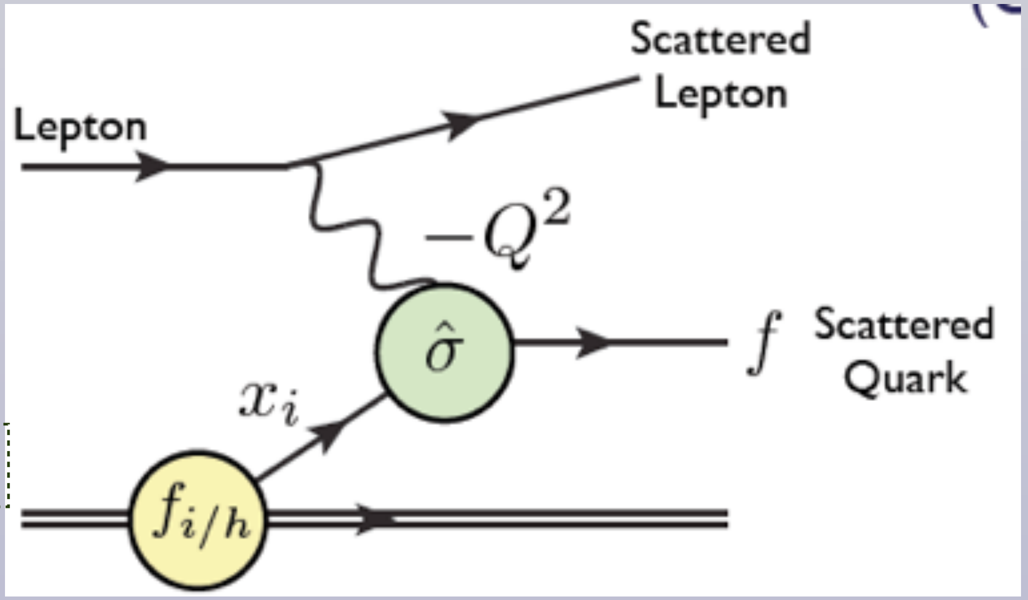
Parton Di
Data + Met

Multiple I



QCD Factorization

- Deep-inelastic **lepton-proton scattering**: First evidence for **proton structure** (70s)
- QCD Factorization allows to separate the hadronic cross section into a **perturbative, process dependent partonic cross section** and **non-perturbative, process independent Parton Distributions**



$$F_i(x,Q^2) = x \sum_i \int_x^1 \frac{dz}{z} C_i\left(\frac{x}{z}, \alpha_s(Q^2)\right) f_i(z,Q^2).$$

Partonic xsec | Parton Distribution

- The same factorization allows to use the same **universal PDFs** to predict proton-proton collisions at the LHC:

$$\sigma_X(s, M_X^2) \quad = \quad \sum_{a,b} \int_{x_{\min}}^1 dx_1\, dx_2\, f_{a/h_1}(x_1, M_X^2) f_{b/h_2}(x_2, M_X^2) \hat{\sigma}_{ab\to X}\left(x_1 x_2 s, M_X^2\right)$$

x-Bjorken: momentum fraction carried by **parton q**
Q^2 = Resolution scale at which proton is being probed

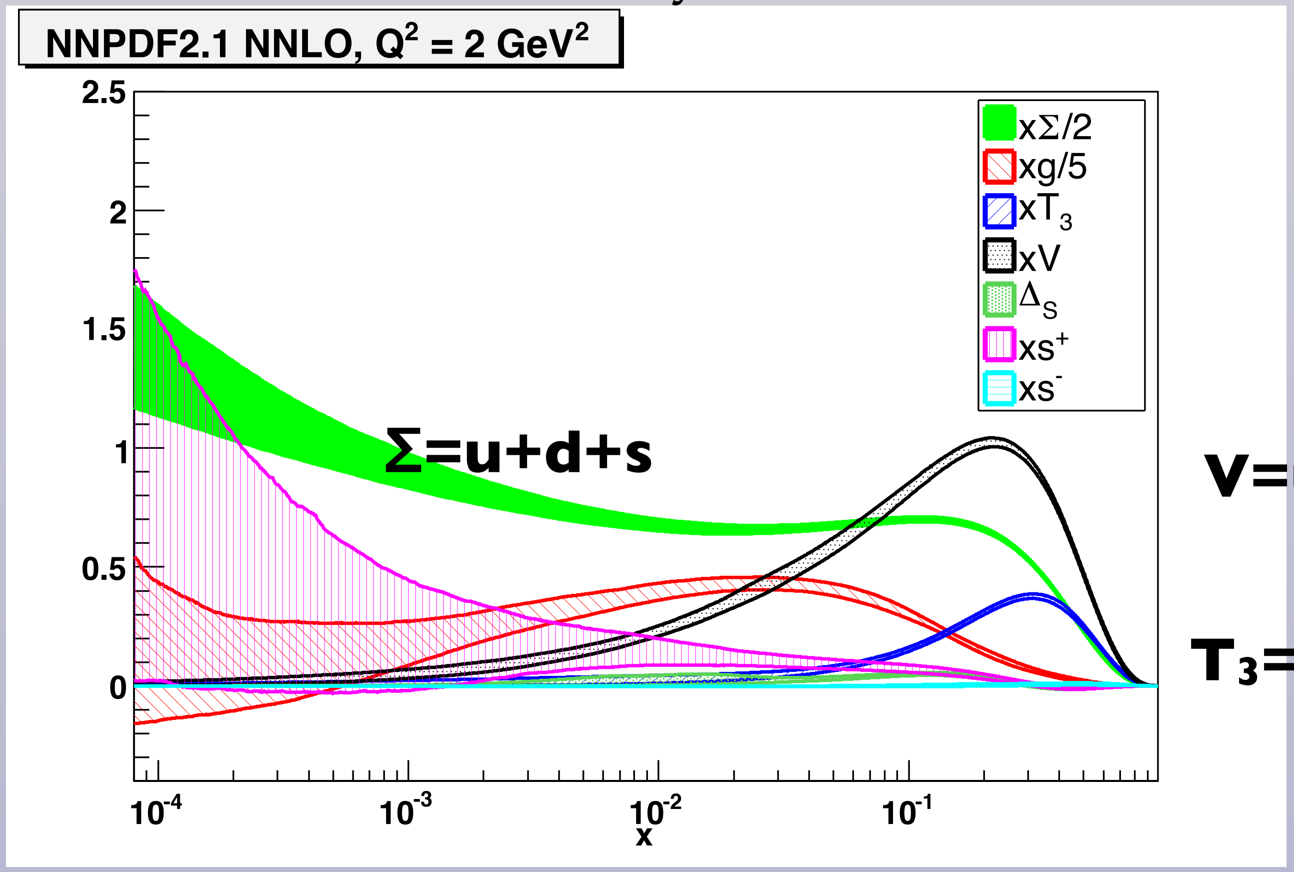
Parton Distributions | Partonic xsec

Parton Distributions

- One independent PDF for each parton in the proton: **u(x,Q²), d(x,Q²), g(x,Q²), ... 13 PDFs**

- At Leading Order PDFs understood as the **probability of finding a parton of a given flavor that carries a fraction x** of the total proton's momentum

- **Shape and normalization** of PDFs are very different for each flavor



V=u-ubar+d-dbar

T_3=u+ubar-d+dbar

- PDFs scheme-dependent, but **valence** and **momentum** sum rules **valid to all orders**

$$\int_0^1 dx\, x\left[\Sigma(x)+g(x)\right]=1$$

$$\int_0^1 dx\, \left(u(x)-\bar{u}(x)\right)=2\ , \quad \int_0^1 dx\, \left(d(x)-\bar{d}(x)\right)=1$$

DGLAP evolution

- The dependence of PDFs on **Bjorken-x is non perturbative**, but the scale (resolution) dependence is dictated by the integro-differential **DGLAP evolution equations**

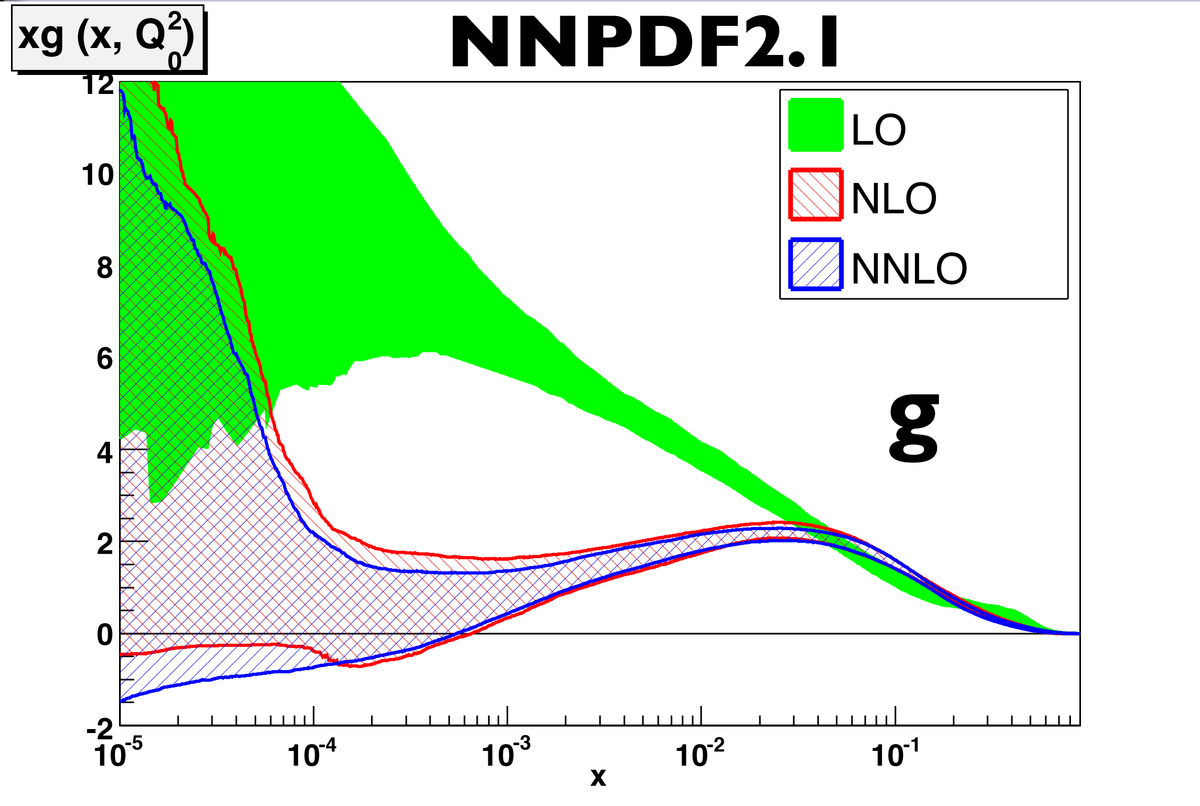
$$\frac{\partial q_i(x,Q^2)}{\partial \ln Q^2} = \frac{\alpha_s\left(Q^2\right)}{2\pi}\int_x^1 \frac{dz}{z} P_{ij}\left(z,\alpha_s\left(Q^2\right)\right) q_j\left(\frac{x}{z},Q^2\right)$$

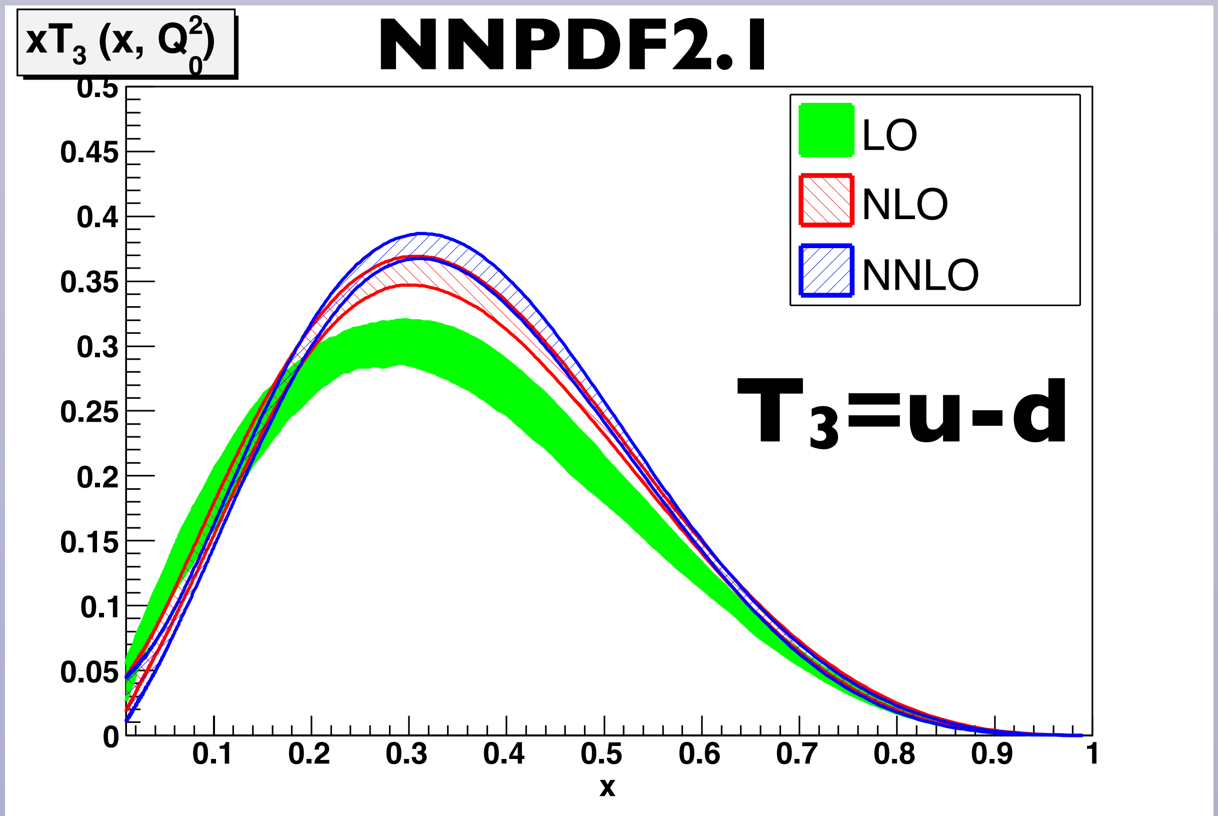
- x-dependence $\mathbf{q(x,Q^2_0)}$ extracted from data, pQCD determines PDFs at other scales $\mathbf{q(x,Q^2)}$

- **Evolution kernels** have been computed up to next-to-next-to-leading order (NNLO):

$$P(z,\alpha_s\left(Q^2\right)) = P^{(0)}(z) + \frac{\alpha_s\left(Q^2\right)}{2\pi}P^{(1)}(z) + \left(\frac{\alpha_s\left(Q^2\right)}{2\pi}\right)^2 P^{(2)}(z)$$

- Reasonable **convergence of the perturbative expansion** of PDFs up to NNLO





PDF determination

PDF determination is based on a **global analysis of hard scattering data** to extract, thanks to the factorization theorem, **universal PDFs for LHC predictions**

Experimental data → **Lepton-proton structure functions**, heavy quark production, jet production, Drell-Yan pair production, electroweak bosons, isolated photons, ...

↓

QCD Theory → **DGLAP evolution, hard scattering cross sections**, heavy quark treatment, strong coupling, electroweak effects, hadronic production, PDF flavor separation,...

↓

Methodology → *Ansatze* **for the x-dependence of PDFs $q(x,Q_0)$**, propagation of **experimental errors from data to PDFs**, model uncertainties, minimization, definition of the **figure of merit χ^2**

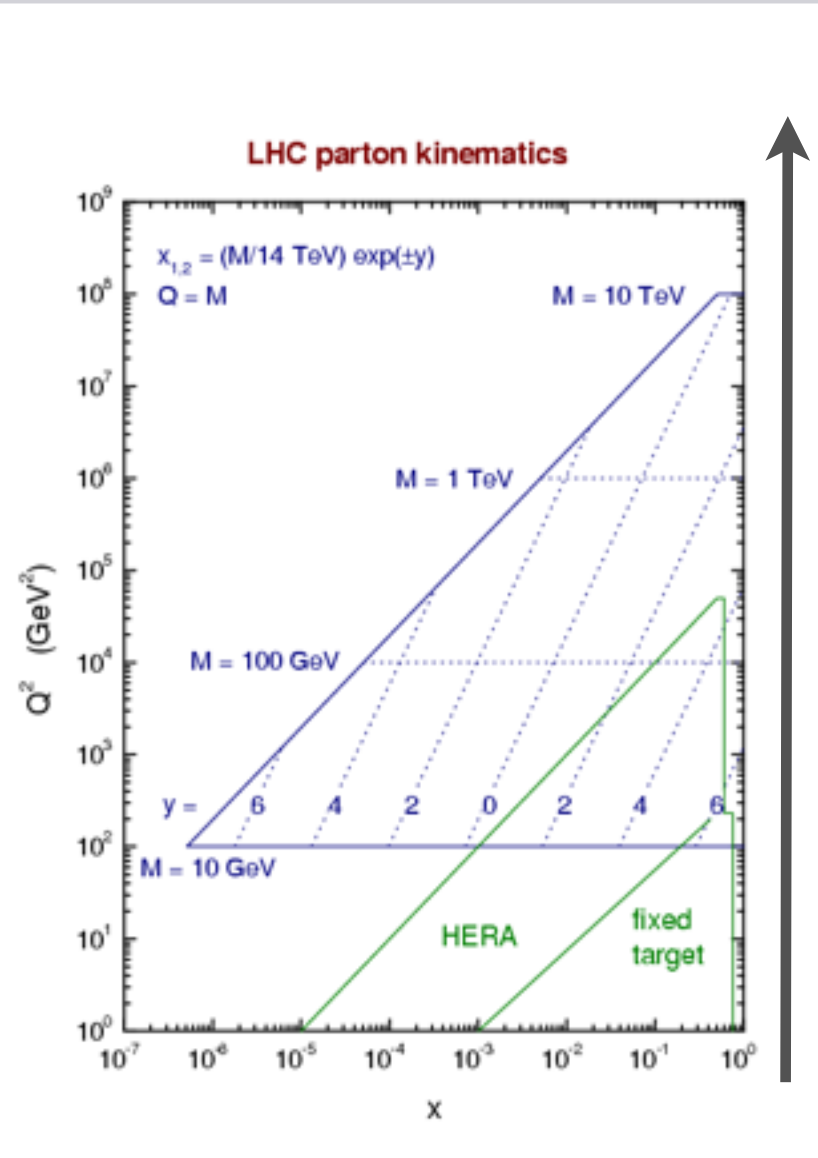
↓

Parton Distributions → **LHC phenomenology**, predictions for LHC processes, PDF and Higgs production, ...

All modern PDF sets available from the **LHAPDF** library

Experimental data in global PDF fits

Q² dependence of PDFs: determined by pQCD



x dependence of PDFs: determined from data

- A **global dataset** covering a wide set of hard-scattering observables is required to constrain **all possible PDF combinations** in the **whole range of Bjorken-x**

- For example, **inclusive jets** are sensitive to the **large-x gluon**, while **HERA neutral current** data pins down the **small-x quarks**

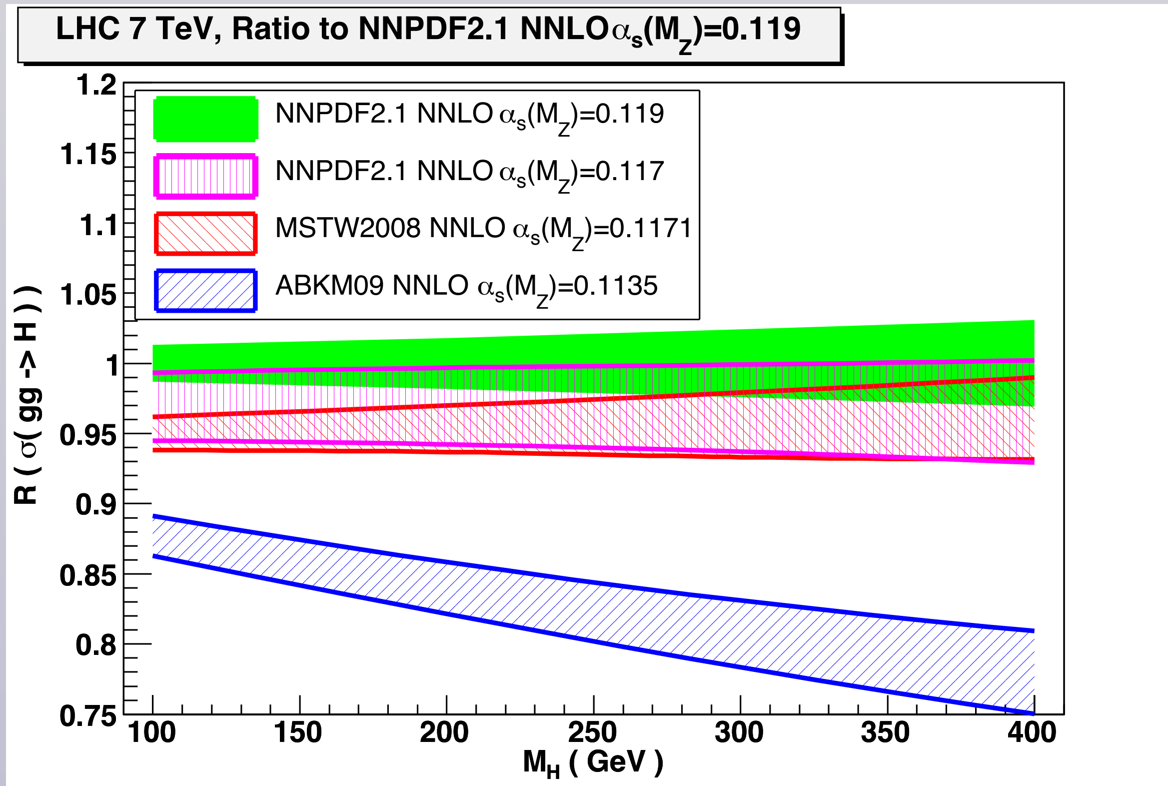
- **LHC data** is introducing completely new observables to be used for PDF constraints

Process	Subprocess	Partons	x range
$\ell^\pm \{p,n\} \to \ell^\pm X$	$\gamma^* q \to q$	$q, \bar{q}, g$	$x \gtrsim 0.01$
$\ell^\pm\ n/p \to \ell^\pm X$	$\gamma^*\ d/u \to d/u$	d/u	$x \gtrsim 0.01$
$pp \to \mu^+\mu^-\ X$	$u\bar{u}, d\bar{d} \to \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
$pn/pp \to \mu^+\mu^-\ X$	$(u\bar{d})/(u\bar{u}) \to \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
$\nu(\bar{\nu})\, N \to \mu^-(\mu^+)\, X$	$W^* q \to q'$	$q, \bar{q}$	$0.01 \lesssim x \lesssim 0.5$
$\nu\, N \to \mu^-\mu^+\ X$	$W^* s \to c$	s	$0.01 \lesssim x \lesssim 0.2$
$\bar{\nu}\, N \to \mu^+\mu^-\ X$	$W^* \bar{s} \to \bar{c}$	$\bar{s}$	$0.01 \lesssim x \lesssim 0.2$
$e^\pm\, p \to e^\pm\, X$	$\gamma^* q \to q$	$g, q, \bar{q}$	$0.0001 \lesssim x \lesssim 0.1$
$e^+\, p \to \bar{\nu}\, X$	$W^+ \{d,s\} \to \{u,c\}$	d, s	$x \gtrsim 0.01$
$e^\pm p \to e^\pm\, c\bar{c}\, X$	$\gamma^* c \to c,\ \gamma^* g \to c\bar{c}$	c, g	$0.0001 \lesssim x \lesssim 0.01$
$e^\pm p \to \text{jet} + X$	$\gamma^* g \to q\bar{q}$	g	$0.01 \lesssim x \lesssim 0.1$
$p\bar{p} \to \text{jet} + X$	$gg, qg, qq \to 2j$	g, q	$0.01 \lesssim x \lesssim 0.5$
$p\bar{p} \to (W^\pm \to \ell^\pm \nu)\, X$	$ud \to W, \bar{u}\bar{d} \to W$	$u, d, \bar{u}, \bar{d}$	$x \gtrsim 0.05$
$p\bar{p} \to (Z \to \ell^+\ell^-)\, X$	$uu, dd \to Z$	d	$x \gtrsim 0.05$

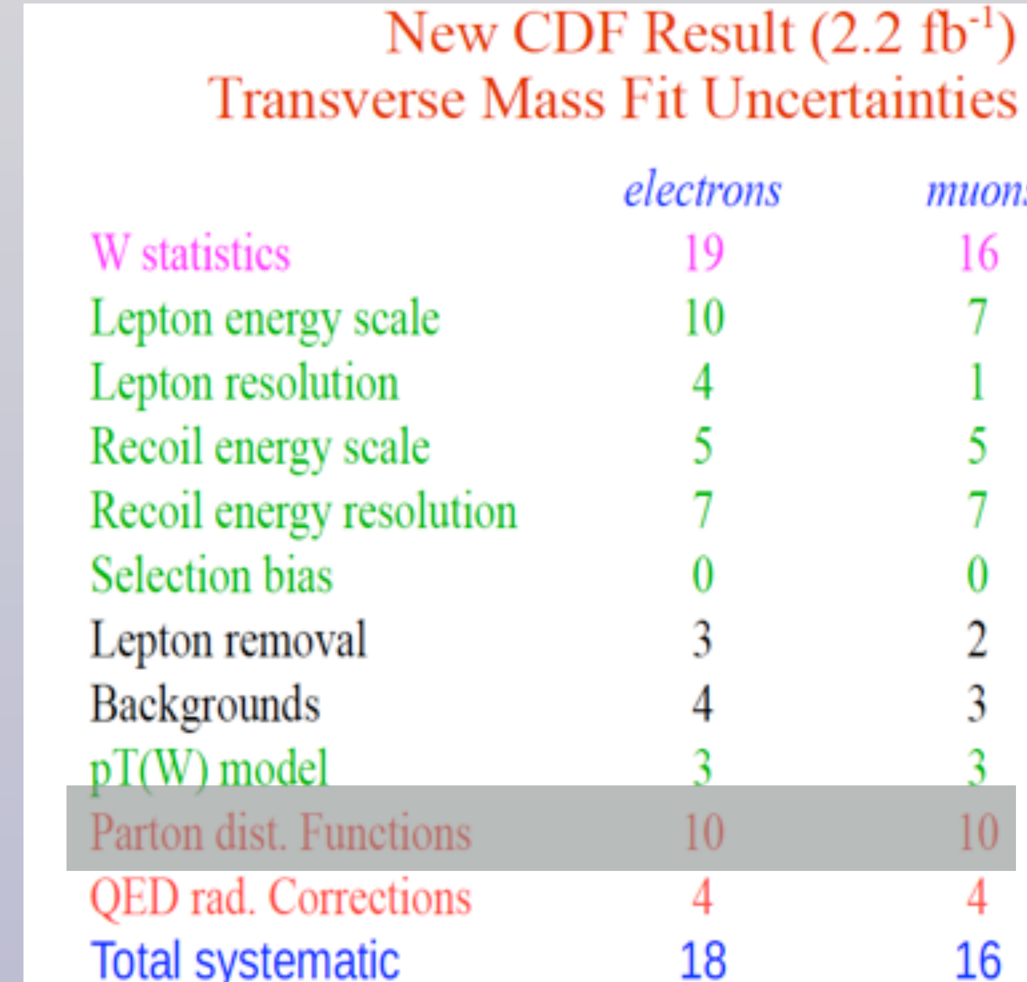
MSTW08, arXiv:0901.0002

Parton Distributions at the LHC

Parton Distributions, and their associated **theoretical and experimental uncertainties** play a crucial for **hadron collider phenomenology**:



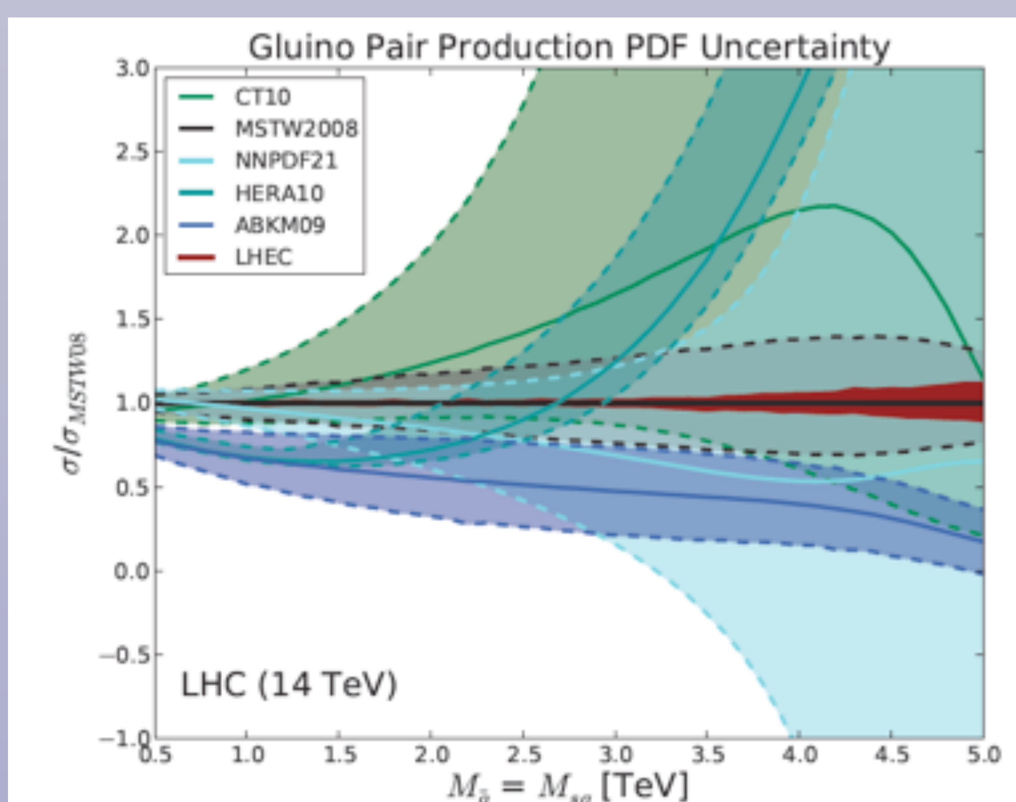
PDFs one of main TH uncertainties in Higgs production: limit coupling extraction, ...



New CDF Result (2.2 fb^{-1})
Transverse Mass Fit Uncertainties

	electrons	muons
W statistics	19	16
Lepton energy scale	10	7
Lepton resolution	4	1
Recoil energy scale	5	5
Recoil energy resolution	7	7
Selection bias	0	0
Lepton removal	3	2
Backgrounds	4	3
pT(W) model	3	3
Parton dist. Functions	10	10
QED rad. Corrections	4	4
Total systematic	18	16

PDFs are dominant systematic in the very precise W mass @ Tevatron (even more at LHC), which indirectly constraints Higgs mass



PDF uncertainties affect substantially theory predictions for BSM high mass production (SUSY, Z', KK)

And many other cases: alphas determination, CKM elements, effective lepton mixing angle, cross section rations between CM energies, neutrino astrophysics ....

Parton Distributions at the LHC

Various collaborations provide regular updates of their **PDF determinations**:

Collaboration	Authors	arXiv
ABM	S. Alekhin, J. Blümlein, S. Moch	1105.5349, 1101.5261, 1107.3657, 0908.3128, 0908.2766, ...
CTEQ/TEA	M. Guzzi J. Huston, H.-L. Lai, P. Nadolsky, J. Pumplin, D. Stump, C.-P. Yuan	1108.5112, 1101.0561, 1007.2241, 1004.4624, 0910.4183, 0904.2424, 0802.0007, ...
GJR	M. Glück, P. Jimenez-Delgado, E. Reya	1003.3168, 0909.1711, 0810.4274, ...
HERAPDF	H1 and ZEUS Collaborations	1107.4193, 1006.4471, 0906.1108, ...
MSTW	A. Martin, J. Stirling, R. Thorne, G. Watt	1107.2624, 1006.2753, 0905.3531, 0901.0002, ...
NNPDF	R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, AG, N. P. Hartland, J. I. Latorre, J. Rojo, M. Ubiali	1110.2483, 1108.2758, 1107.2652, 1103.2369, 1102.3182, 1101.1300, 1005.0397, 1002.4407, 0912.2276, 0906.1958, ...

Parton Distributions at the LHC

PDF sets differ by **choice of dataset, QCD treatment, methodology, ...**

	DATASET	PERT. ORDER	HQ TREATMENT	α_s	PARAM.	UNCERT.
ABM11	DIS Drell-Yan	NLO NNLO	FFN (BMSN)	Fit (multiple values available)	6 indep. PDFs Polynomial (25 param.)	Hessian ($\Delta\chi^2$=1)
CT10	Global	LO NLO NNLO	GM-VFNS (S-ACOT)	External (multiple values available)	6 indep. PDFs Polynomial (26 param.)	Hessian ($\Delta\chi^2$=100)
JR09	DIS Drell-Yan Jets	NLO NNLO	FFN VFN	Fit	5 indep. PDFs Polynomial (15 param.)	Hessian ($\Delta\chi^2$=1)
HERAPDF1.5	DIS (HERA)	NLO NNLO	GM-VFNS (TR)	External (multiple values available)	5 indep. PDFs Polynomial (14 param.)	Hessian ($\Delta\chi^2$=1)
MSTW08	Global	LO NLO NNLO	GM-VFNS (TR)	Fit (multiple values available)	7 indep. PDFs Polynomial (20 param.)	Hessian ($\Delta\chi^2$~25)
NNPDF2.1/2.3	Global	LO NLO NNLO	GM-VFNS (FONLL)	External (multiple values available)	7 indep. PDFs Neural Nets (259 param.)	Monte Carlo

The Neural Network Approach to Parton Distributions

(Neural Network) PDF determination

The **NNPDF approach** aims to improve on the **shortcomings of standard PDF determinations**, with the use of a **modern robust statistical methodology** coupled to the **most updated theoretical information** and **all the relevant hard scattering data,** including **LHC data**

Experimental data → **NNPDF2.3:** First and only PDF set which includes **all relevant LHC data on jet and W/Z production** (on top of HERA, fixed target, Tevatron, ...)

↓

QCD Theory → **NLO and NNLO PDFs**, Heavy quark effects with the **FONLL scheme, exact NLO** (no K-factors) for all hadronic data + EW effects and resummations ....

↓

Methodology → **Artificial Neural Networks for PDF q(x,Q_0) parametrization**, **Monte Carlo methods** for the propagation of **experimental errors, Genetic Algorithms** for ANN learning, **Bayesian Inference, ..**

↓

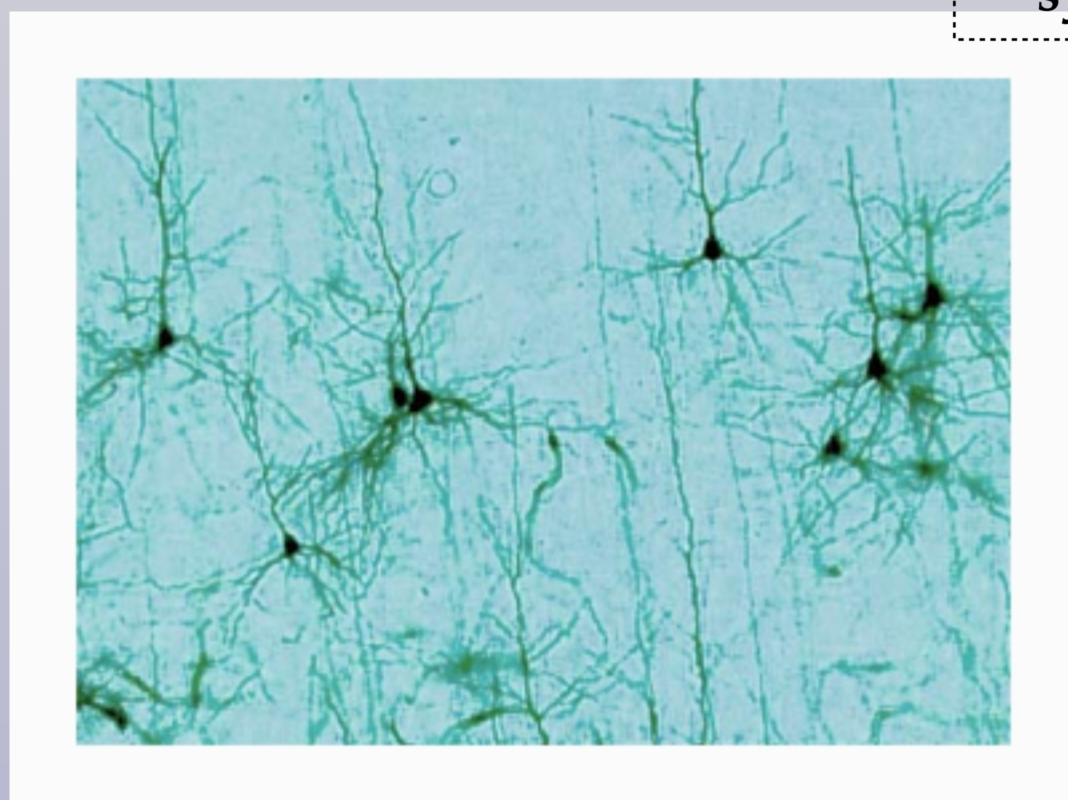
Parton Distributions → PDFs for Higgs production, SUSY cross sections, determination of α_S

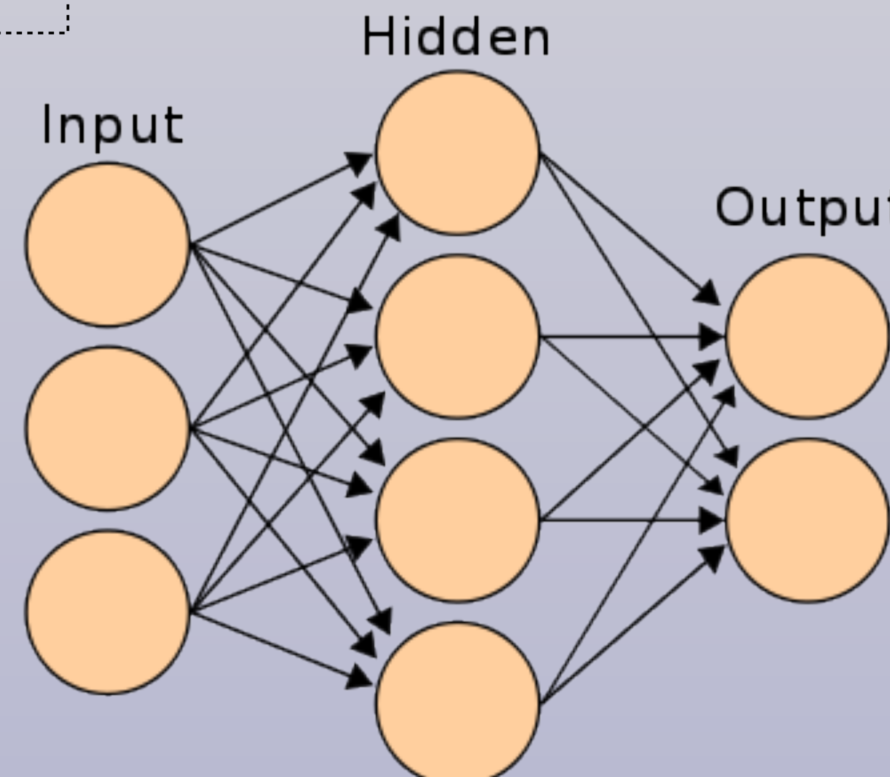
All NNPDF sets available from the **LHAPDF** library and our HepForge website:
http://nnpdf.hepforge.org/

Artificial Neural Networks

- Inspired by **biological brain models**, **Artificial Neural Networks** (ANNs) are **mathematical algorithms** widely used in a wide range of applications, from **high energy physics** to **targeted marketing** and **finance forecasting**

Biological NeuralNets → Neurons, axions, synapses, ... ← **Artificial NeuralNets**





- Artificial neural networks aimed to excel in the same domains as their biological counterparts: **pattern recognition, forecasting, classification**, ... where our **evolution-driven biology** outperforms traditional algorithms

Artificial Neural Networks

Example 1: **Pattern recognition.** During the Yugoslavian wars, the NATO used ANNs to recognize hidden military vehicles

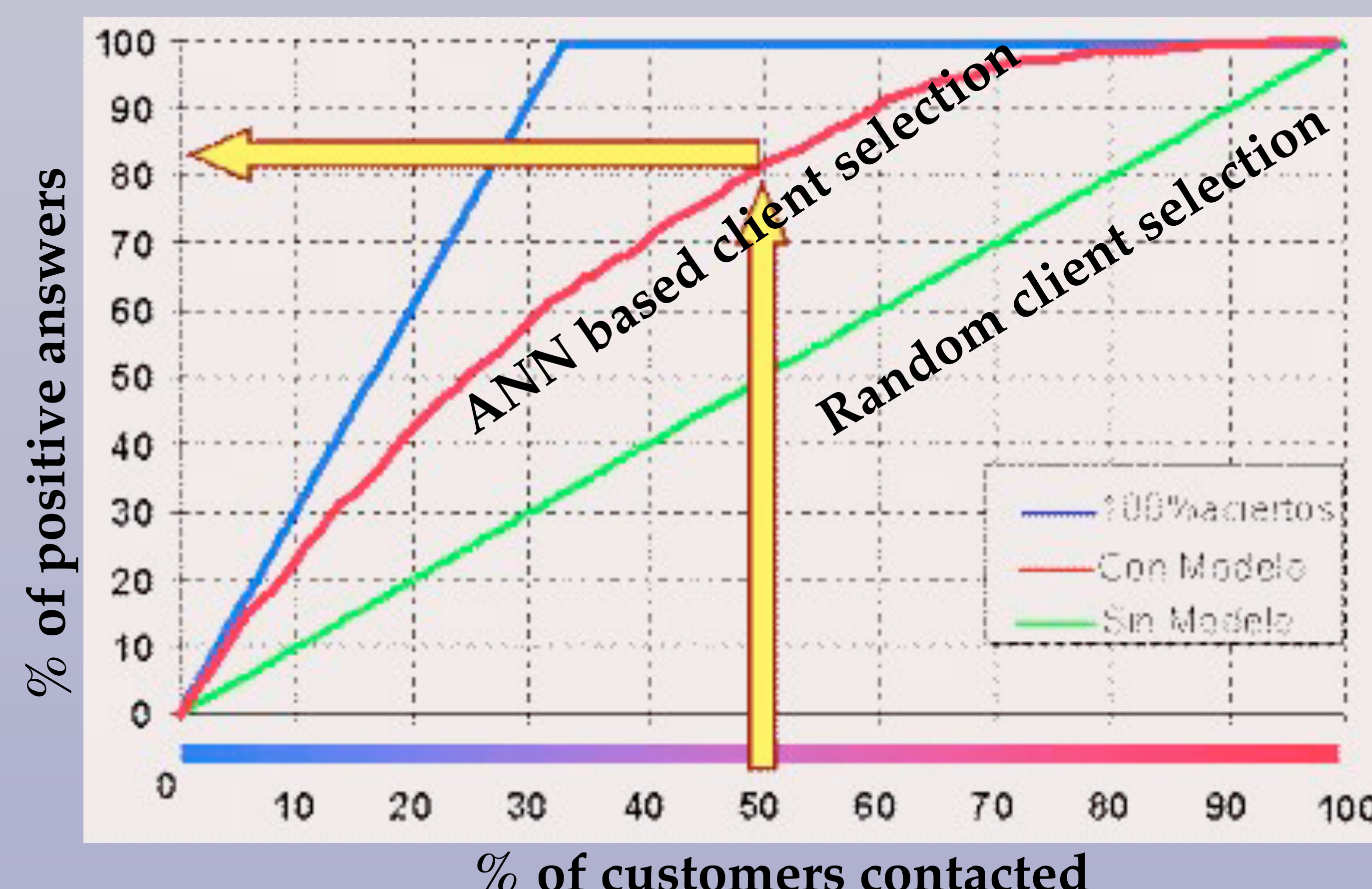
A military aircraft is identified, despite being hidden below a commercial plane.

Many other applications of ANN in **pattern recognition**: OCR software, hand writing recognition, automated anti-plagiarism software, ...

Example 2: **Marketing.** A bank wants to offer a new credit card to their clients. Two possible strategies:

- **Contact all customers**: slow and costly
- Contact 5% of the customers, **train a ANN with their input** (sex, income, loans) and **their ourput** (yes/no) and use the information to contact only clients likely to accepy the offer

Cost-effective method to improve marketing performance

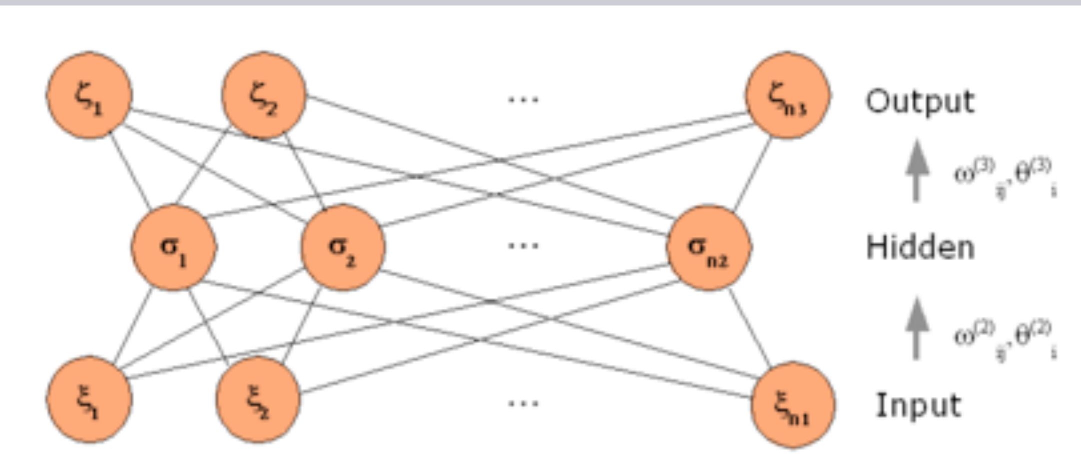


Artificial Neural Networks

- **Artificial Neural Networks** (ANNs) provide **universal unbiased interpolants** to parametrize PDFs at low input scales

$$\begin{aligned}\Sigma(x,Q_0^2) &= (1-x)^{m_\Sigma}x^{-n_\Sigma}\mathrm{NN}_\Sigma(x)\\ g(x,Q_0^2) &= A_g(1-x)^{m_\Sigma}x^{-n_\Sigma}\mathrm{NN}_g(x)\end{aligned}$$

- The ANN class that we adopt are **feed-forward multilayer neural networks** (perceptrons)



$$\xi_i^{(l)} = g\left(h_i^{(l)}\right)\ , \quad i=1,\ldots,n_l\ , \qquad l=2,\ldots,L$$

$$h_i^{(l)} = \sum_{j=1}^{n_{l-1}} \omega_{ij}^{(l)}\xi_j^{(l-1)} - \theta_i$$

- In traditional PDF determinations, the input *ansatz* is a simple **polynomial**

$$\begin{aligned}\Sigma(x,Q_0^2) &= (1-x)^{m_\Sigma}x^{-n_\Sigma}\left(1+a_\Sigma\sqrt{x}+b_\Sigma x+\ldots\right)\ ,\\ g(x,Q_0^2) &= A_g(1-x)^{m_\Sigma}x^{-n_\Sigma}\left(1+a_g\sqrt{x}+b_g x+\ldots\right)\end{aligned}$$

- The use of Neural Networks allows:
  - **No theory bias** introduced in the PDF determination by the choice of *ad-hoc* functional forms
  - The use of very flexible parametrizations for all PDFs - regardless of the dataset used. The NNPDF analysis allow for **O(400) free parameters**, to be compared with **O(10-20) in traditional PDFs**
  - **Faithful extrapolation**: PDF uncertainties **blow up** in regions with scarce experimental data

PDF Uncertainties: The Monte Carlo Method

- Generate a **large number of Monte Carlo replicas of the experimental data** with the same underlying probability distribution

sys errors stat error

$$F_{I,p}^{(\mathrm{art})(k)} = S_{p,N}^{(k)} F_{I,p}^{(\mathrm{exp})} \left(1 + \sum_{l=1}^{N_c} r_{p,l}^{(k)} \sigma_{p,l} + r_p^{(k)} \sigma_{p,s} \right) \quad , \quad k = 1, \ldots, N_{\mathrm{rep}}$$ **>>1**

lumi error random numbers

- Perform a **PDF determination** on each of these MC replicas

- The set of PDF replicas form a **representation of the probability density in the space of parton distribution functions**

- PDF uncertainties can be propagated to physical cross sections using **textbook statistics**, no need of linear/gaussian assumptions

Central PDF prediction = Expectation Value of MC sample

$$\langle \mathcal{O} \rangle = \int \mathcal{O}[f]\, \mathcal{P}(f)\, Df = \frac{1}{N} \sum_{k=1}^{N} \mathcal{O}[f_k]$$

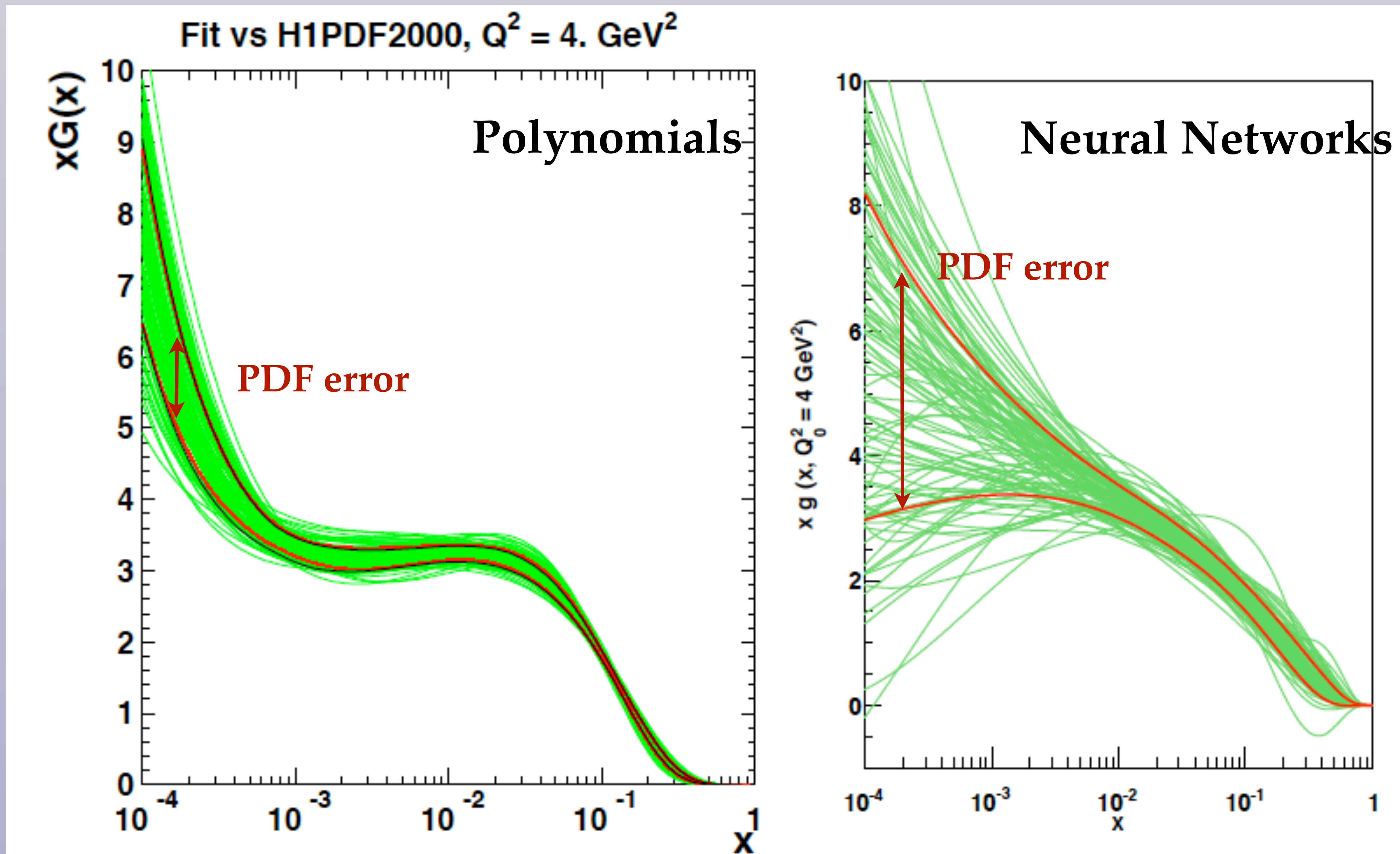
PDF Uncertainty = Standard Deviation of MC sample

$$\Delta f = \sqrt{\frac{1}{N} \sum_{k=1}^{N} f_k^2 - \left(\frac{1}{N} \sum_{k=1}^{N} f_k \right)^2}$$

Artificial Neural Networks vs. Polynomials

- Compare a **benchmark PDF analysis** (HERALHC workshop) where **the same dataset** is fitted with **Artificial Neural Networks** and with **standard polynomials** (everything else identical)

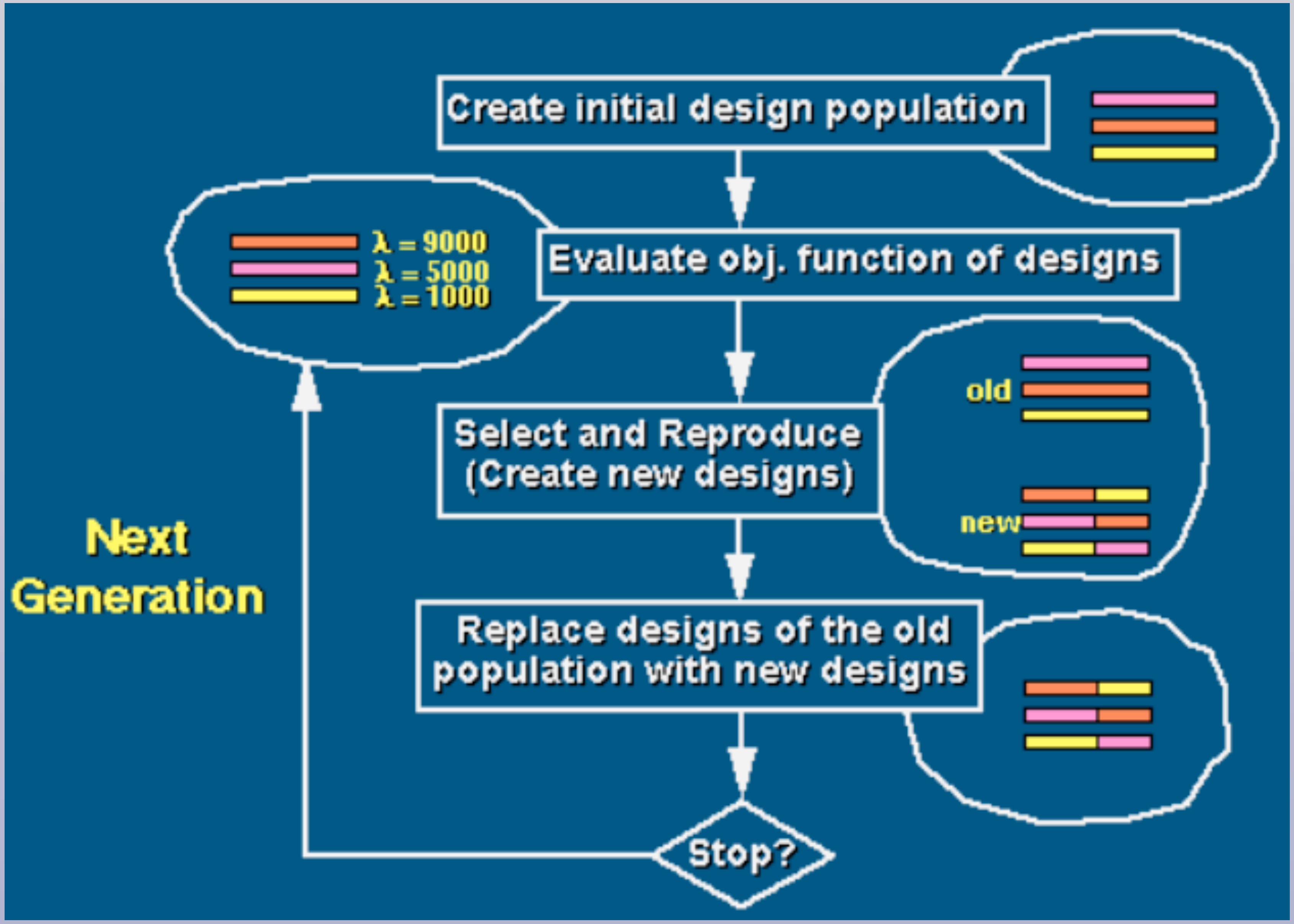
- **ANN** avoid biasing the PDFs, **faithful extrapolation at small-x** (very few data, thus error blow up)



HERALHC proceedings, arXiv:1201.2504

PDF Learning: Genetic Algorithms

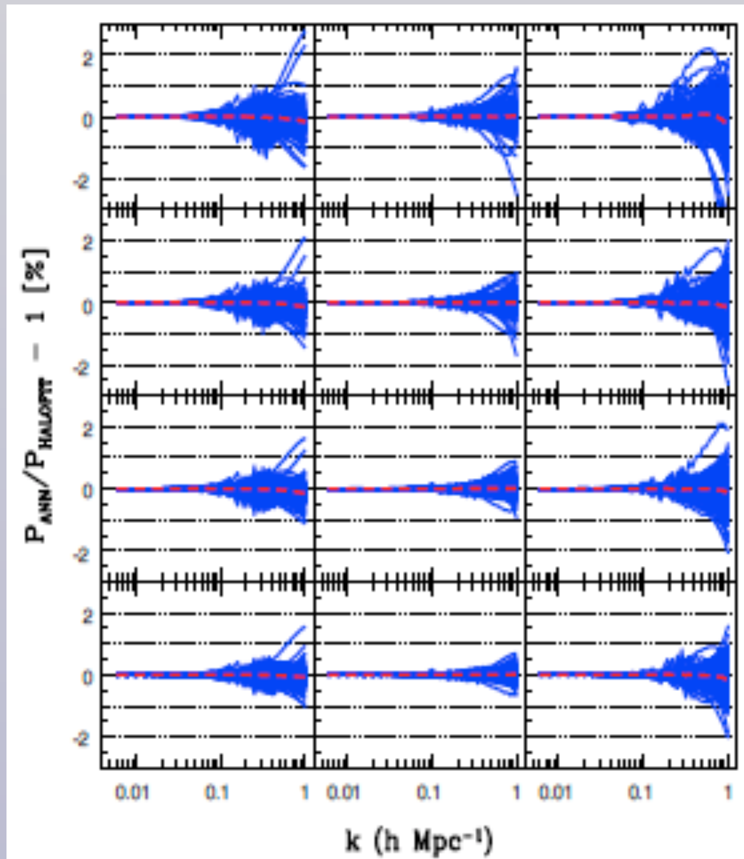
- Traditional minimization algorithms (*ex* MINUIT) are not suitable to **explore huge minima space**
- **Genetic Algorithms** provide a **combination of stochastic elements applied under deterministic rules** which improve optimization efficiency in problems with many extrema



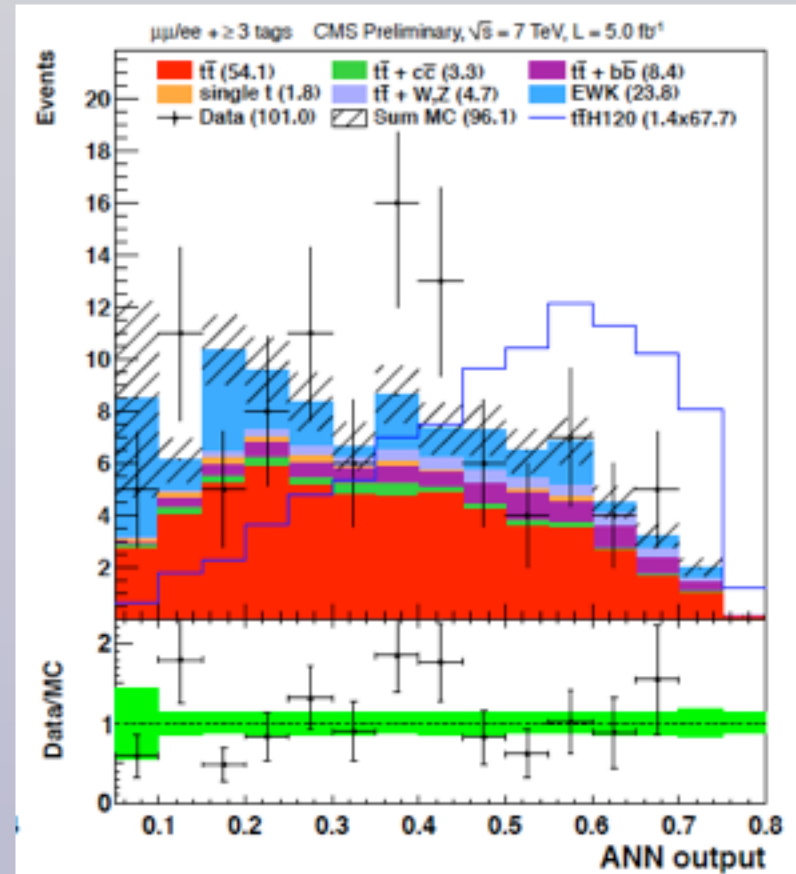
- A first random set of possible solutions is encoded into **chromosomes**
- This initial population undergoes a series of **mutations** and **crossings,** breeding a next generation of individuals
- The **fitness** for each individual is evaluated, and according to that a **selection** process follows
- The process is **iterated** until some convergence criterion is satisfied
- Closely inspired in **Darwinian evolution**

Artificial Intelligence in High Energy Physics

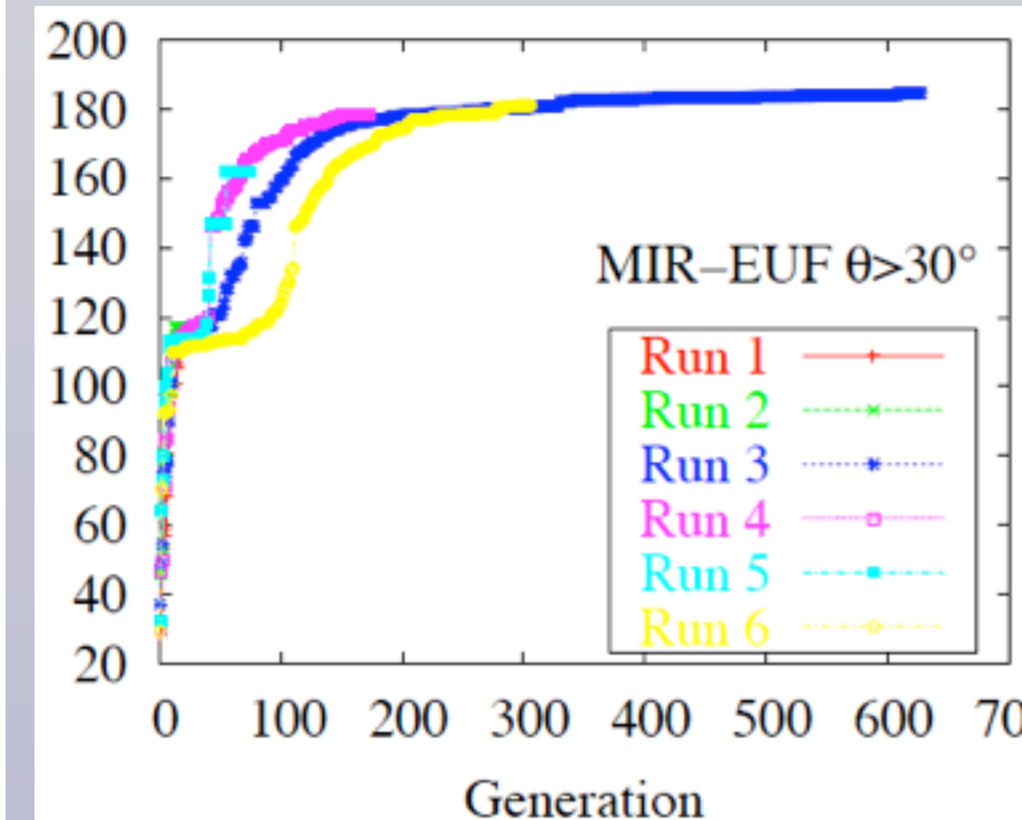
- AI methods are now **ubiquitous in High Energy Physics** and related areas, like Cosmology
- Artificial Neural Networks, Boosted Decision Trees and other **multivariate techniques** are a cornerstone of the LHC data analysis, including **Higgs searches and characterization**



ANNs for Power Spectrum analysis in Cosmology
arxiv:1203.1695

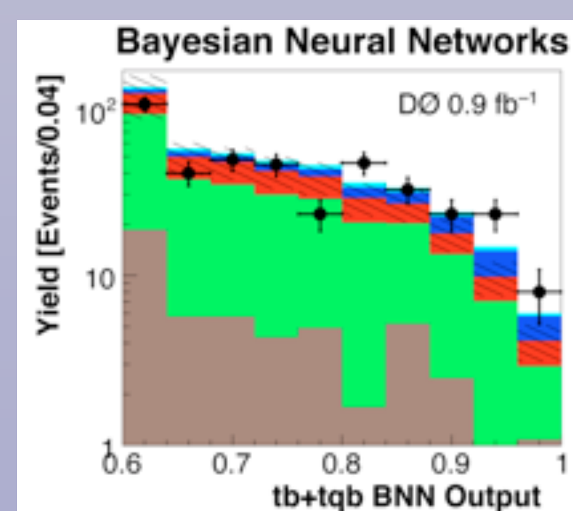


ANNs for Higgs searches in the Htt channel
CMS PAS-HIG-12-025



Genetic Algorithms for the experimental discrimination of SUSY models
hep-ph/0406277

First evidence for single top at the Tevatron from ANNs



Filtering Noise: Cross-Validation Stopping

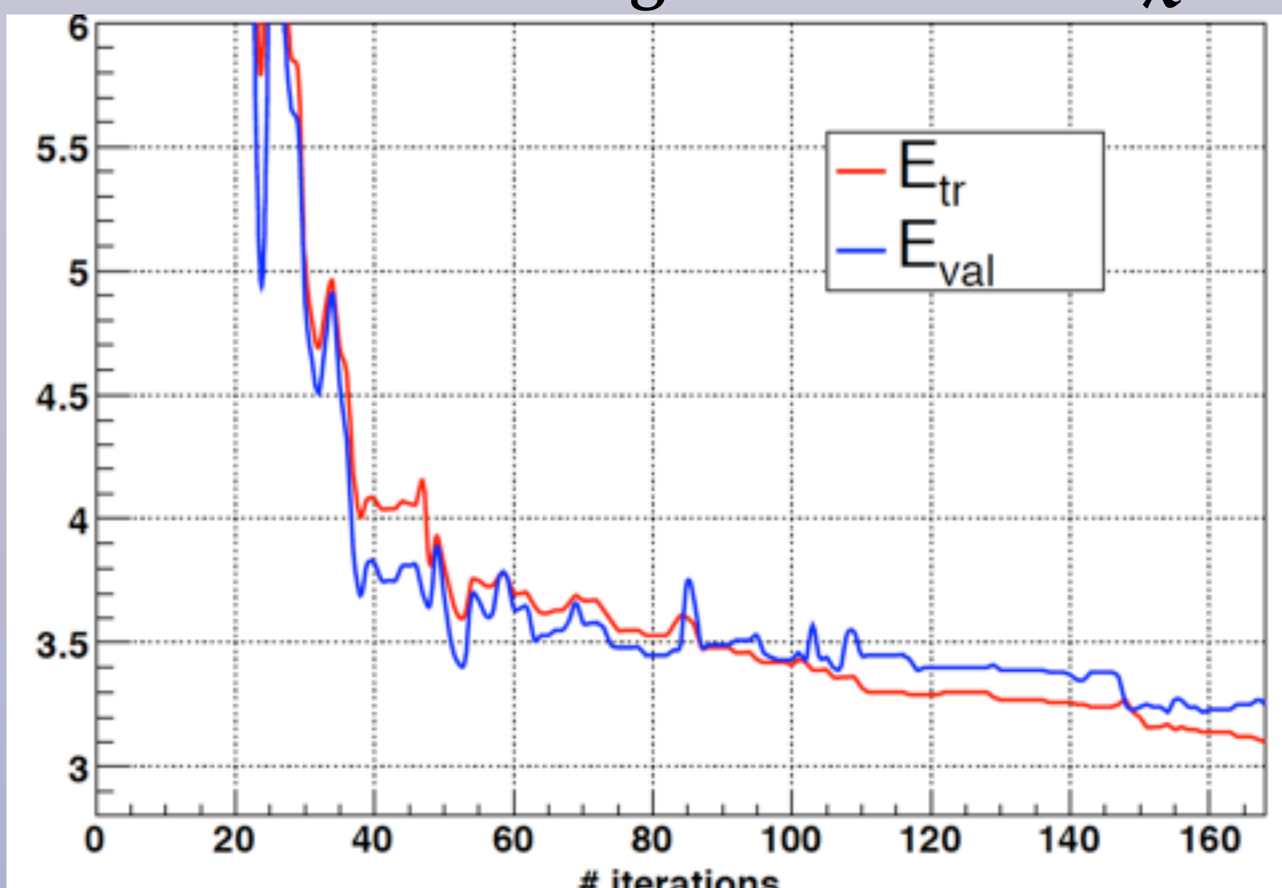
With a **flexible PDF parametrization** as ANNs one can reach the point of fitting **the statistical fluctuations** on the data - on top of the **underlying physical law**

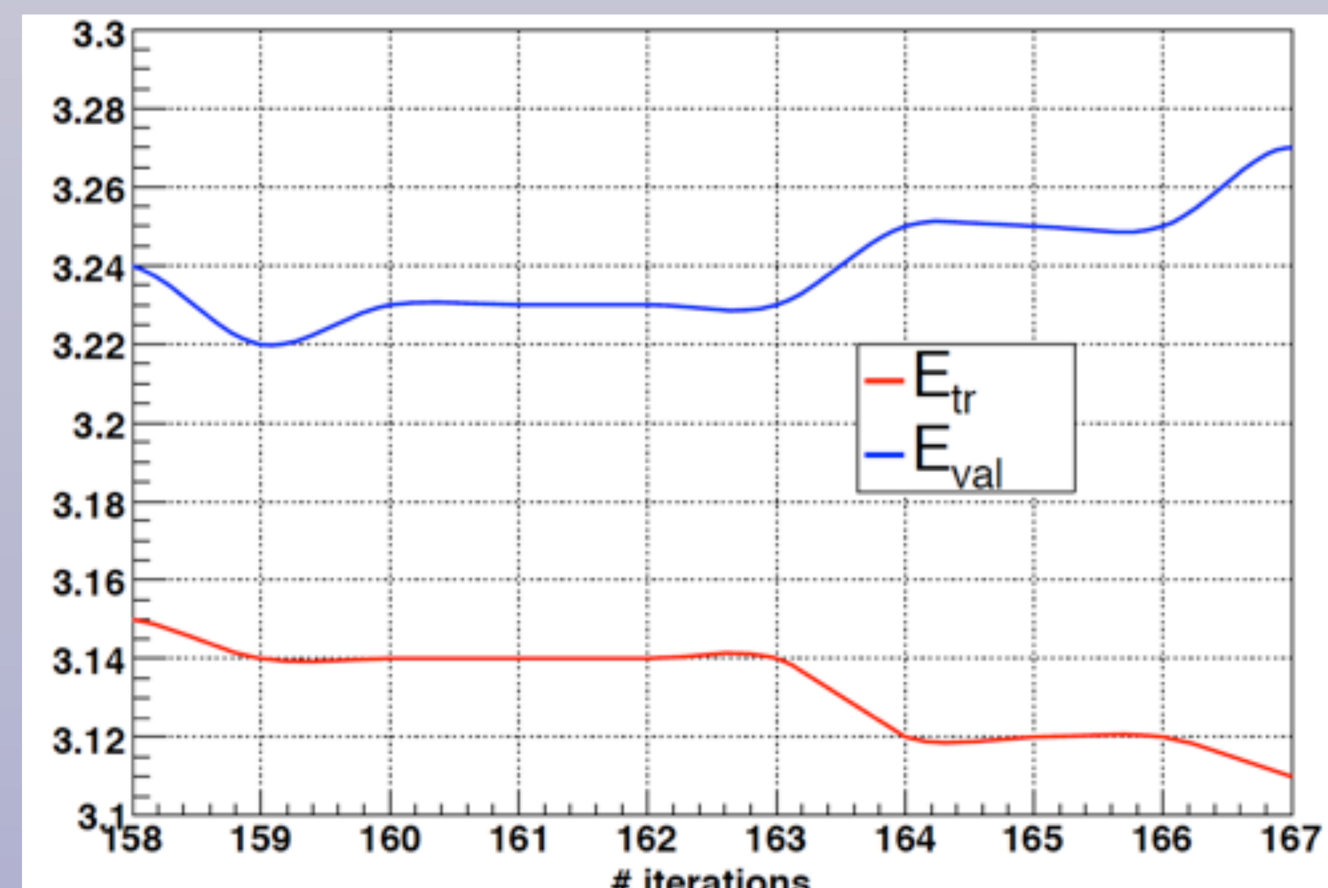
To avoid this, we use the **cross-validation method:** separate data into two disjoint sets

- The **training** set, which is used in the minimization for the neural networks
- The **validation** set, which is only monitored but not used in the fit

The **optimal stopping point** is the one where the fit **quality to the validation set stops improving**: this implies one is fitting the **training set statistical fluctuations**

Training and validation χ^2 as a function of # of GA iterations





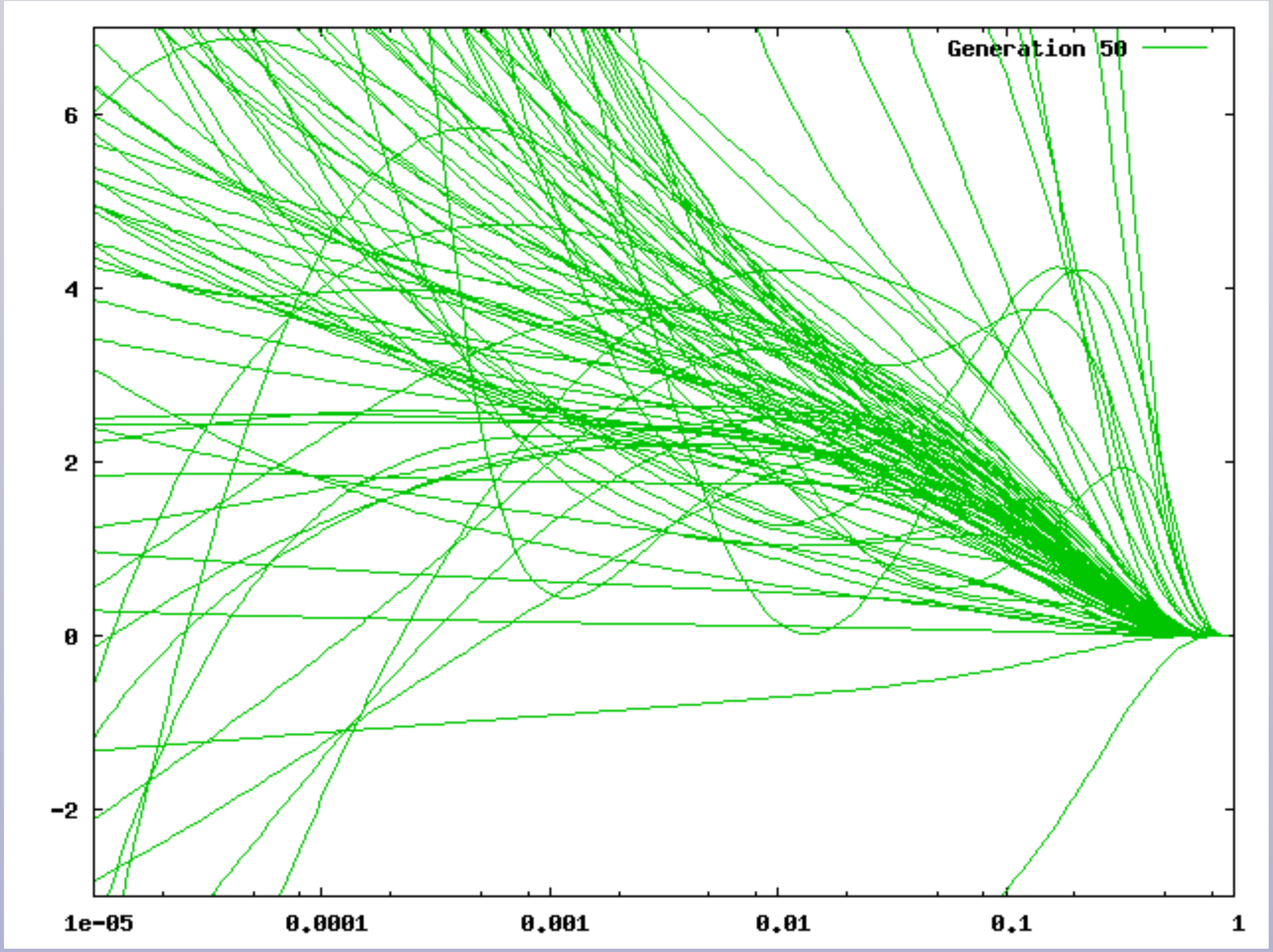
PDF Replica Neural Network Learning

- Now we can combine all the NNPDF methodology together:
  - **Artificial Neural Networks** as unbiased interpolants,
  - **Monte Carlo PDF replicas** for error estimation and propagation,
  - **Genetic Algorithms** for neural network learning,
  - Dynamical **Cross-Validation** Stopping ....

- .... and see how the NNPDF determination works **live** ....

PDF Replica Neural Network Learning

Each **green curve** corresponds to a **gluon PDF Monte Carlo** replica



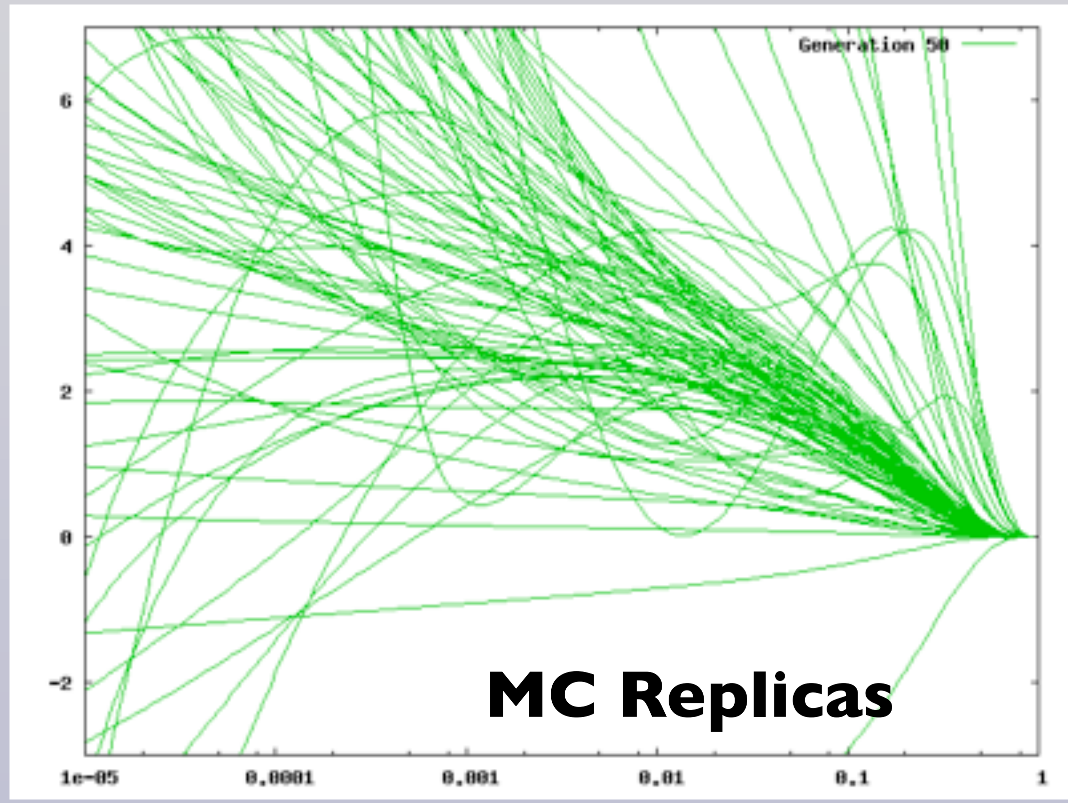
x

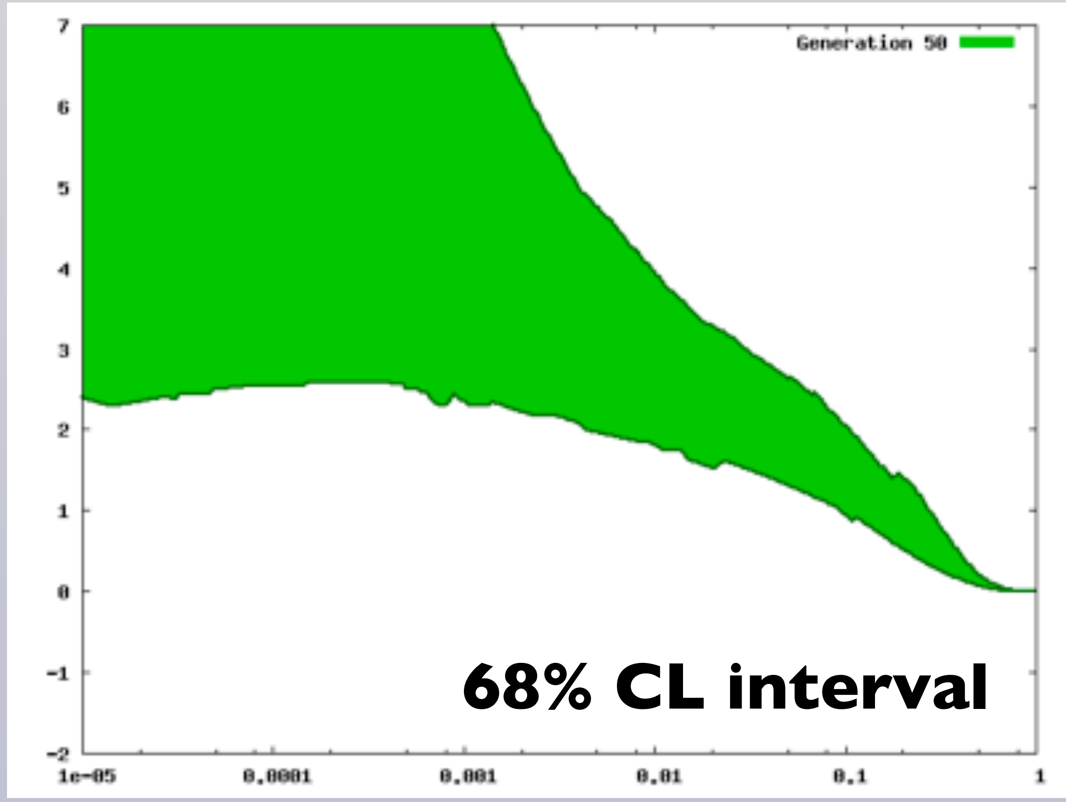
x g(x, Q^2 = 2 GeV^2)

PDF Replica Neural Network Learning

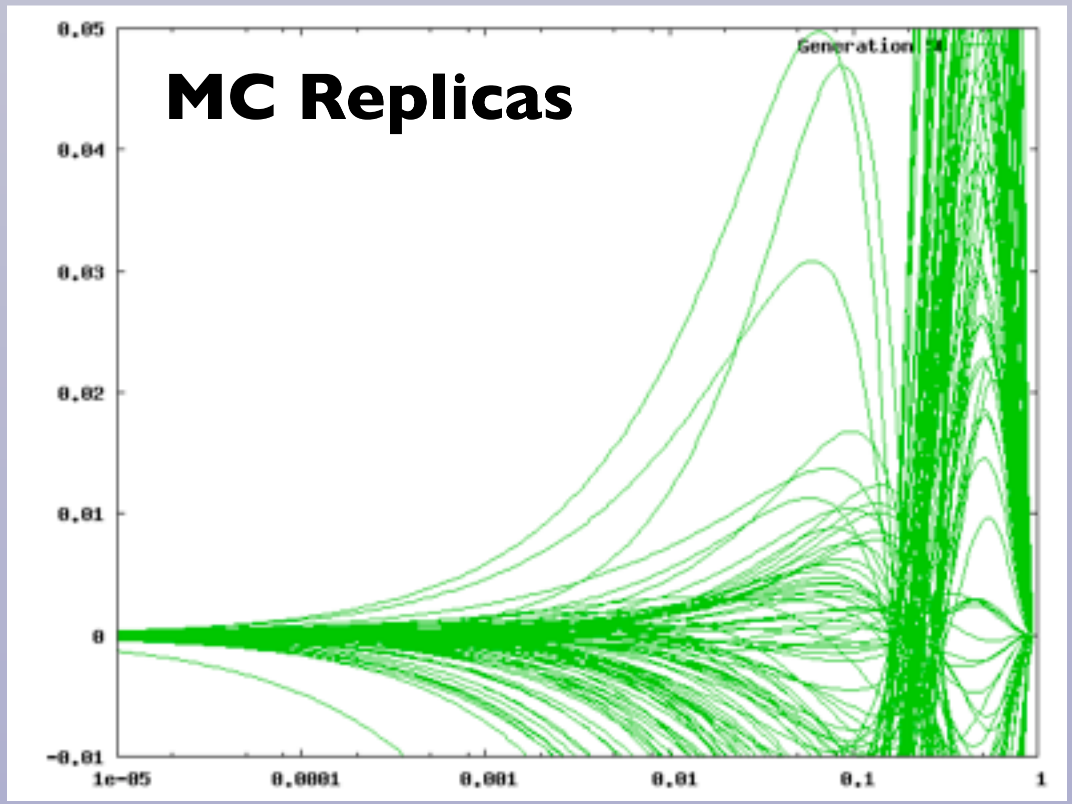
PDF uncertainty band defined as **68% Confidence Level** over Monte Carlo replica sample

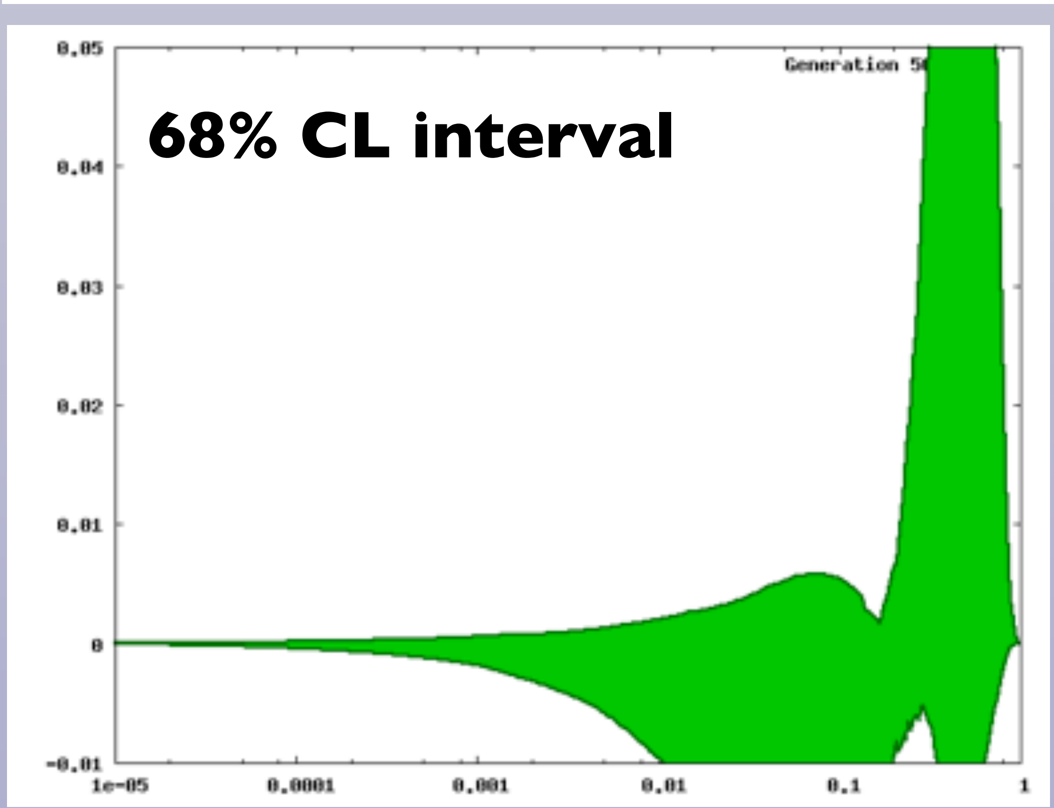
$x\,g(x, Q^2 = 2\ \mathrm{GeV}^2)$





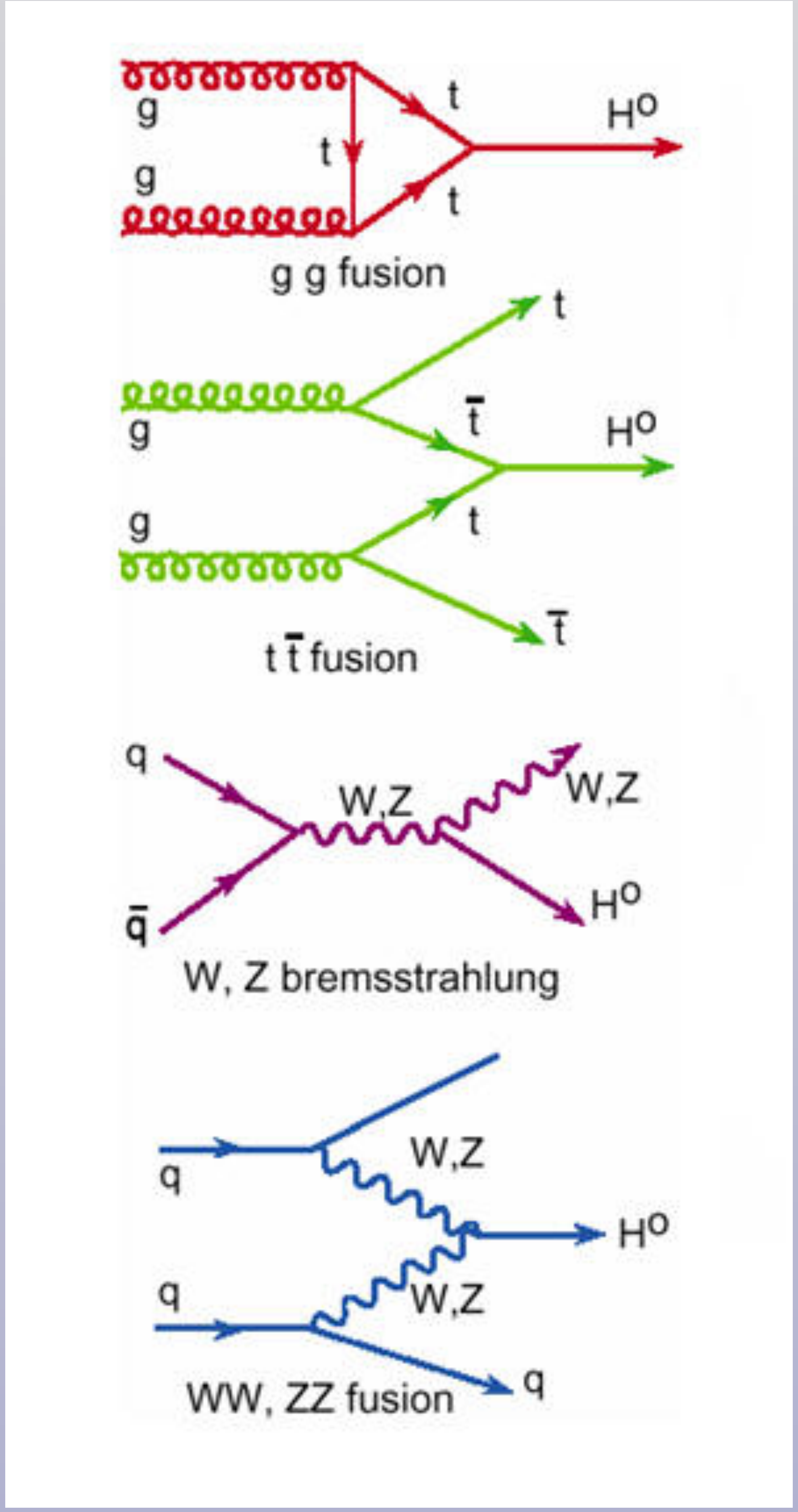
$x\,\Delta s(x, Q^2 = 2\ \mathrm{GeV}^2)$



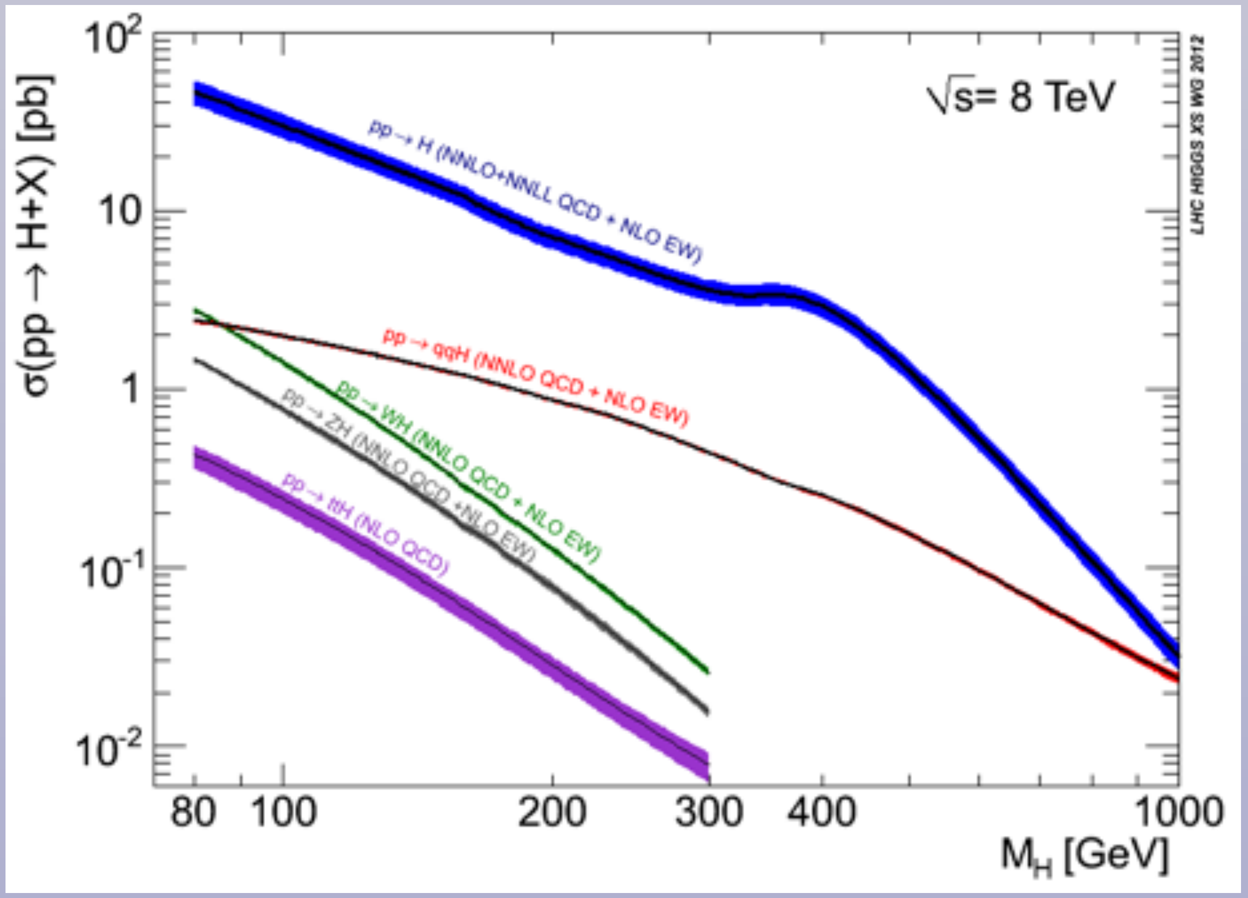


Implications for SM and BSM phenomenology

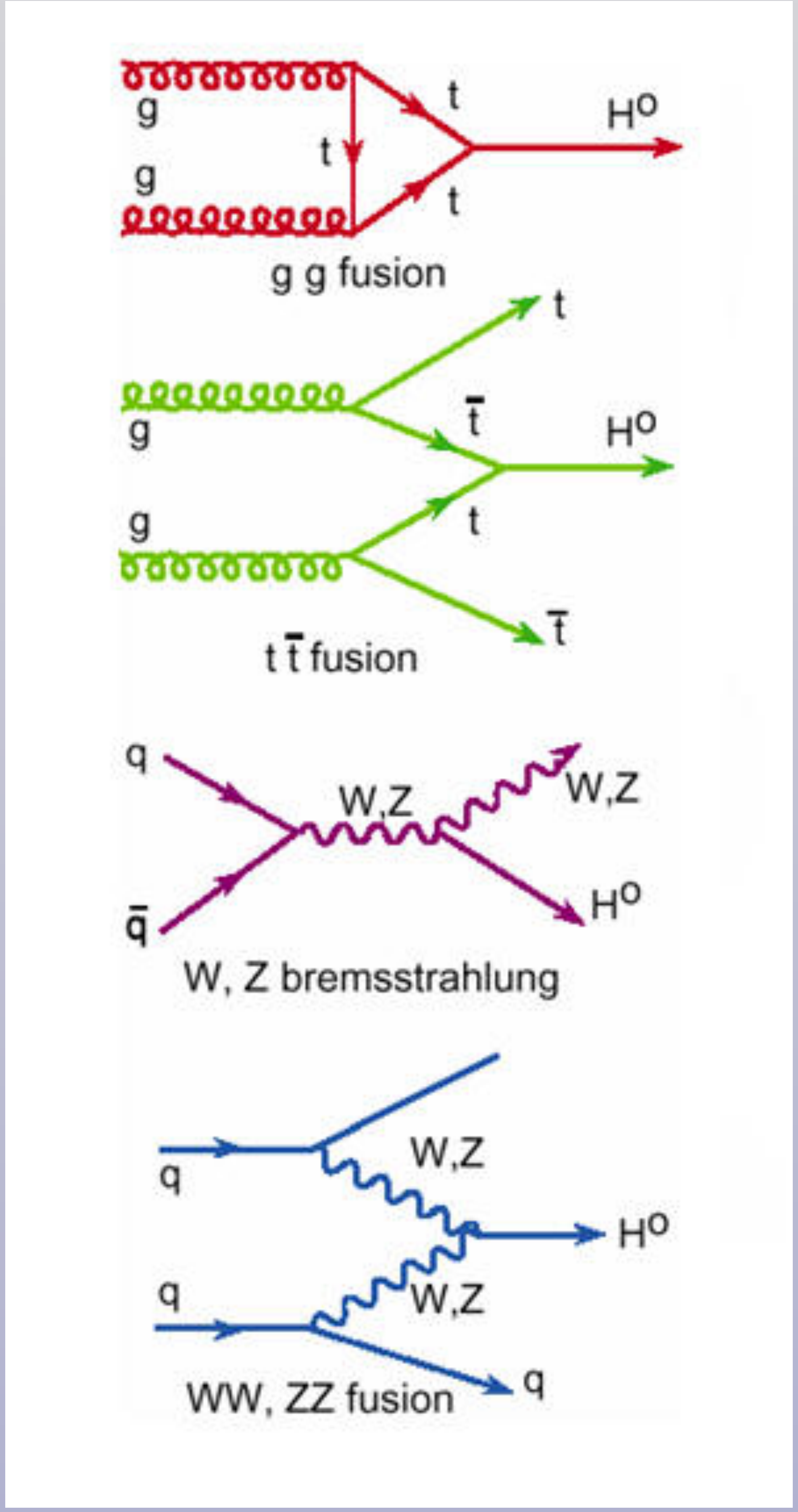
Higgs Boson Production



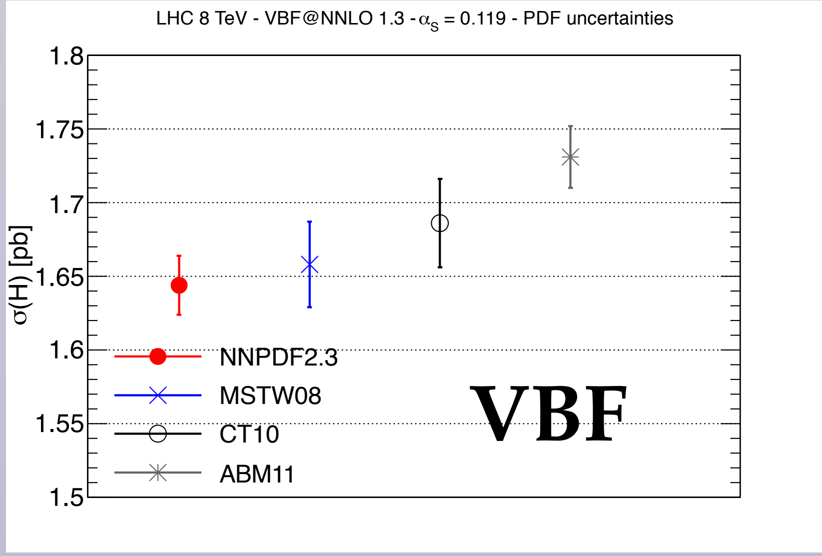
- The study of the Higgs boson properties is a cornerstone of the LHC program. **All production cross sections** require accurate knowledge of different PDF combinations
  - **gg fusion, ttH**: gluon luminosity
  - **vector-boson fusion**: quark-quark luminosity
  - **associated production with W/Z**: quark-antiquark luminosity
- The *Higgs Cross Section Working Group* prescription, used in the ATLAS and CMS analysis, adopts the **envelope of NNPDF2.1, CT10 and MSTW08** sets to **estimate PDF uncertainty**
-

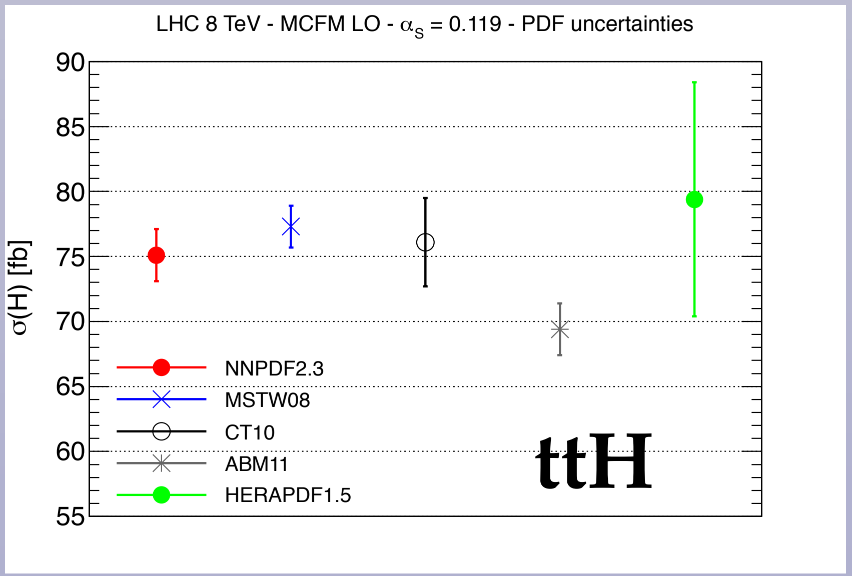


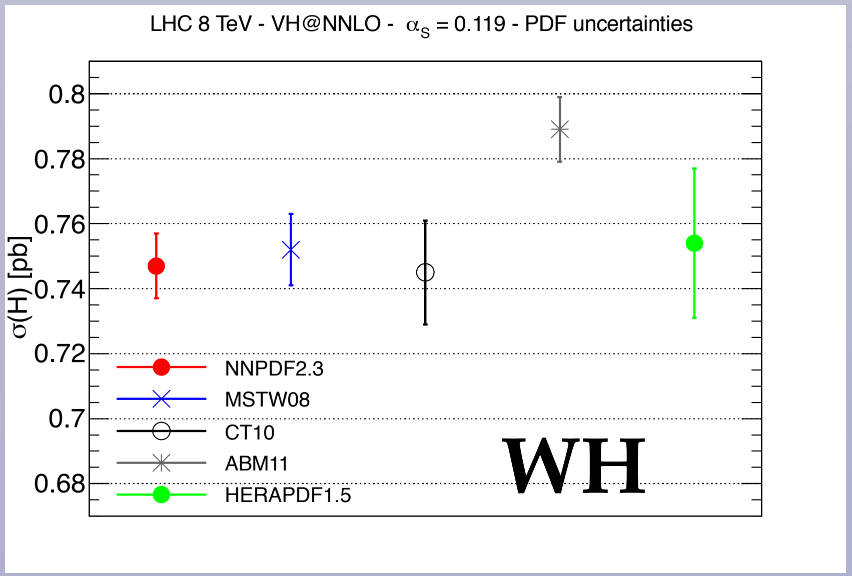
Higgs Boson Production



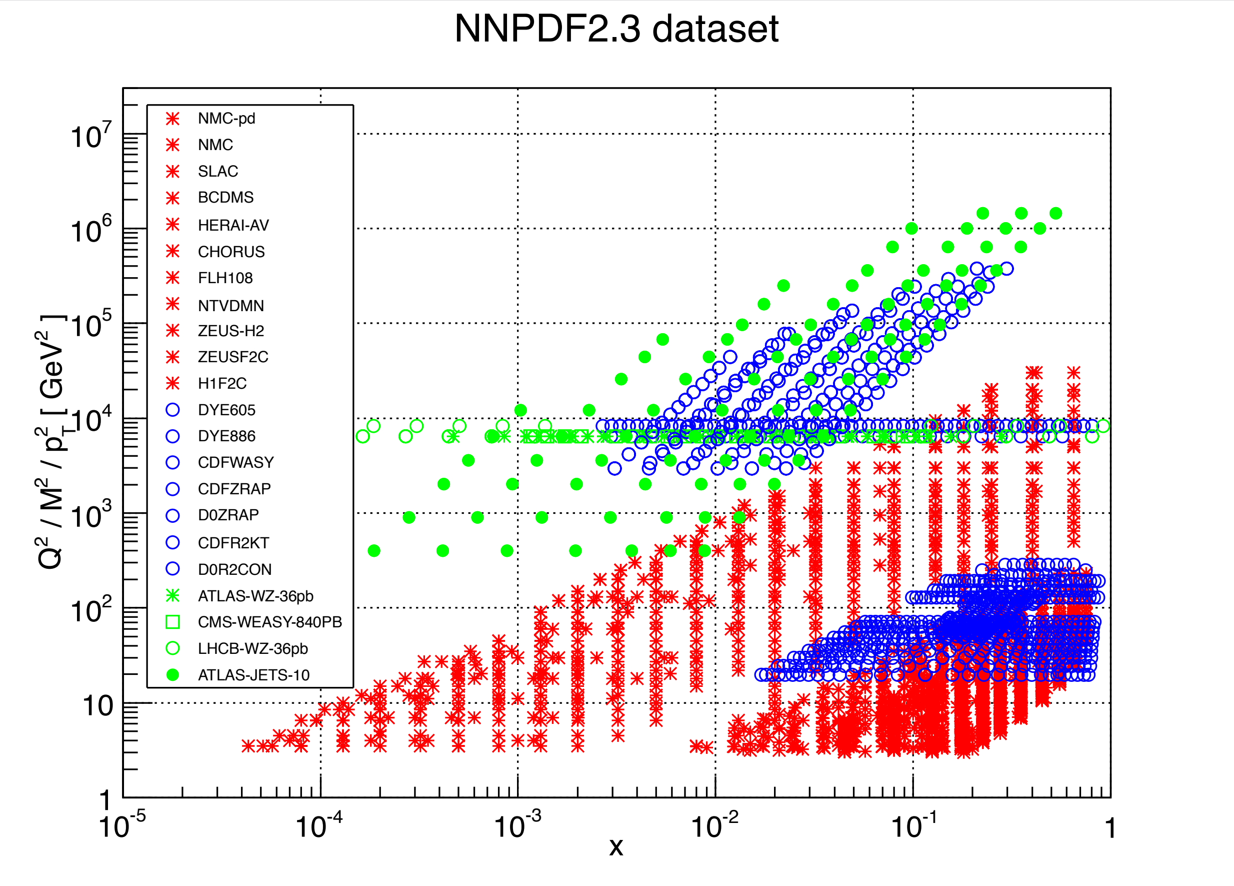
- **PDF uncertainties** in Higgs production are comparable to other theory uncertainties (like missing higher orders), **larger in some cases**
- Thus **improving the precision of PDF determination** is an important ingredient of the Higgs characterization program
- Differences between PDF sets often larger than nominal uncertainty







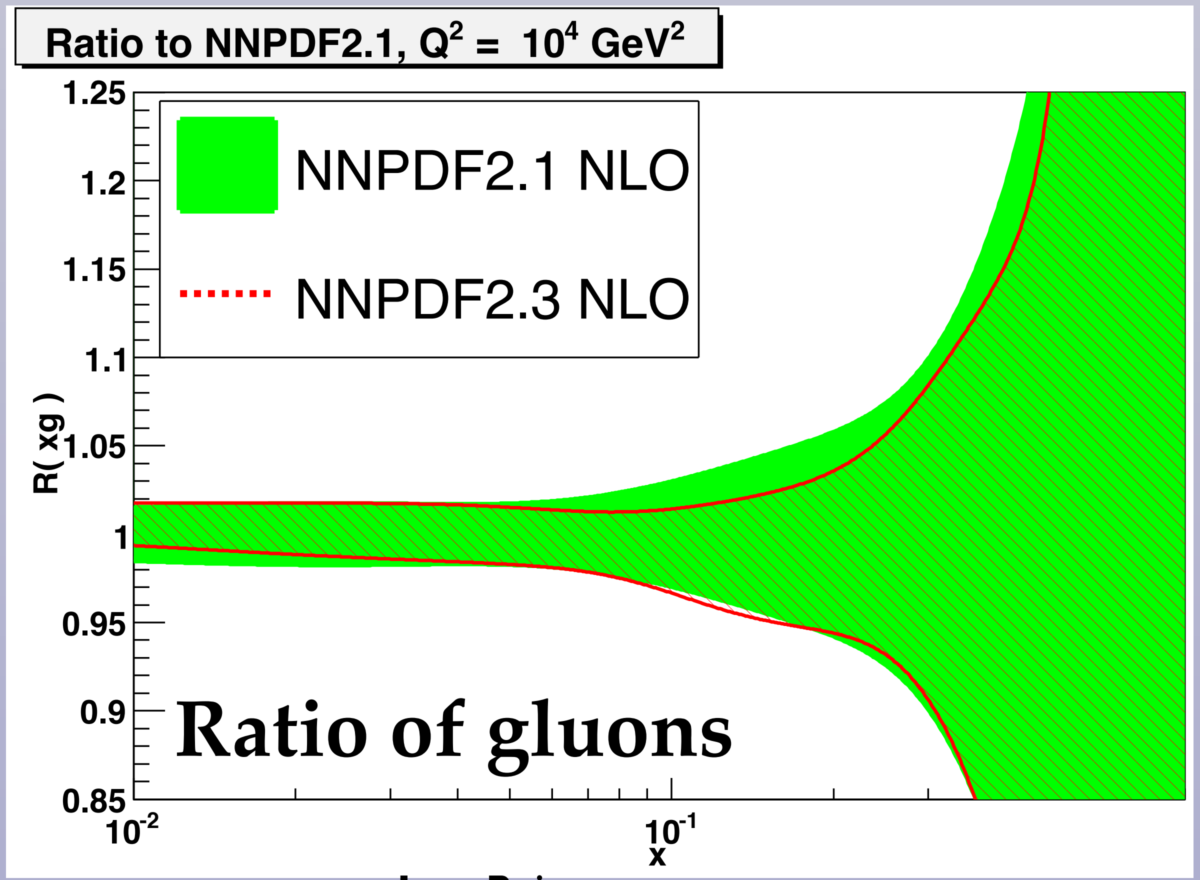
LHC data and PDF analysis

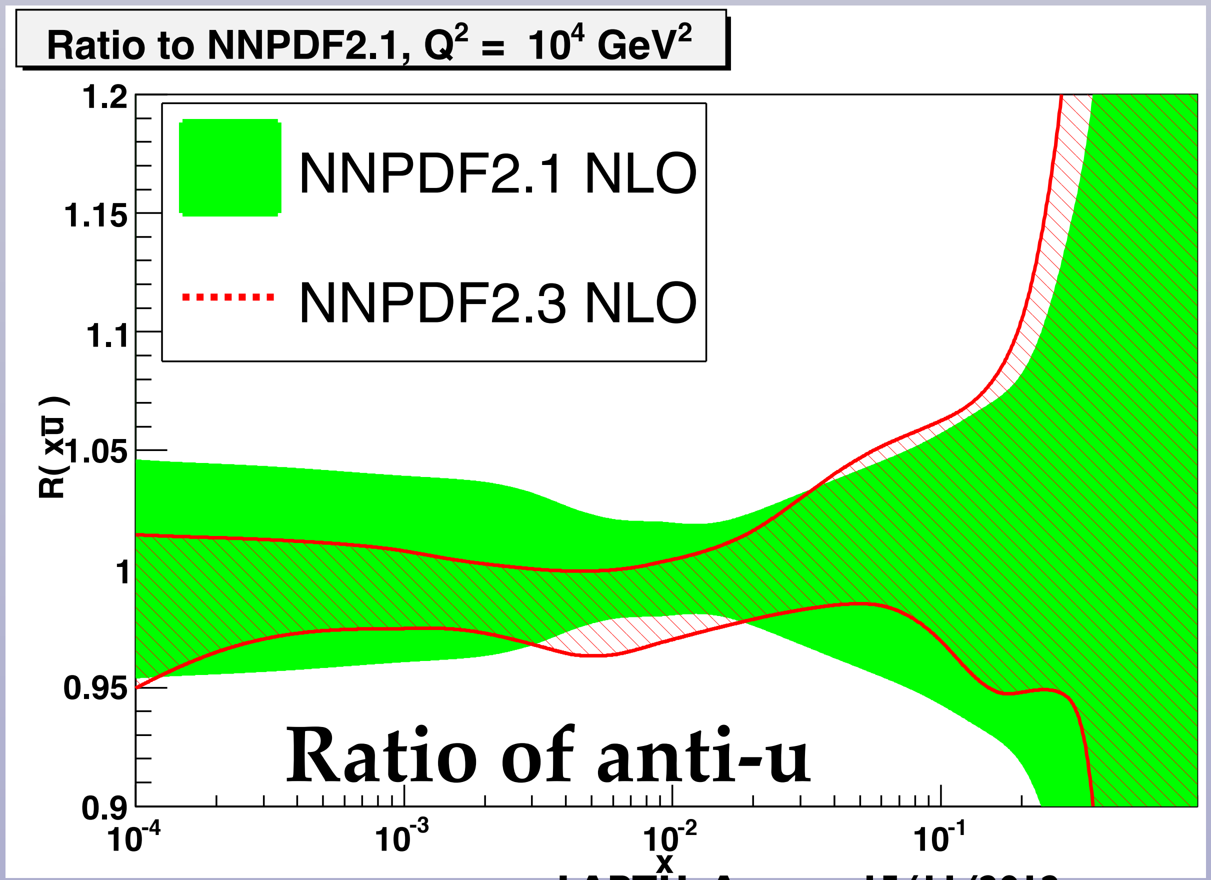


LHC data already part of global PDF analysis, *ie.* the recent **NNPDF2.3** sets

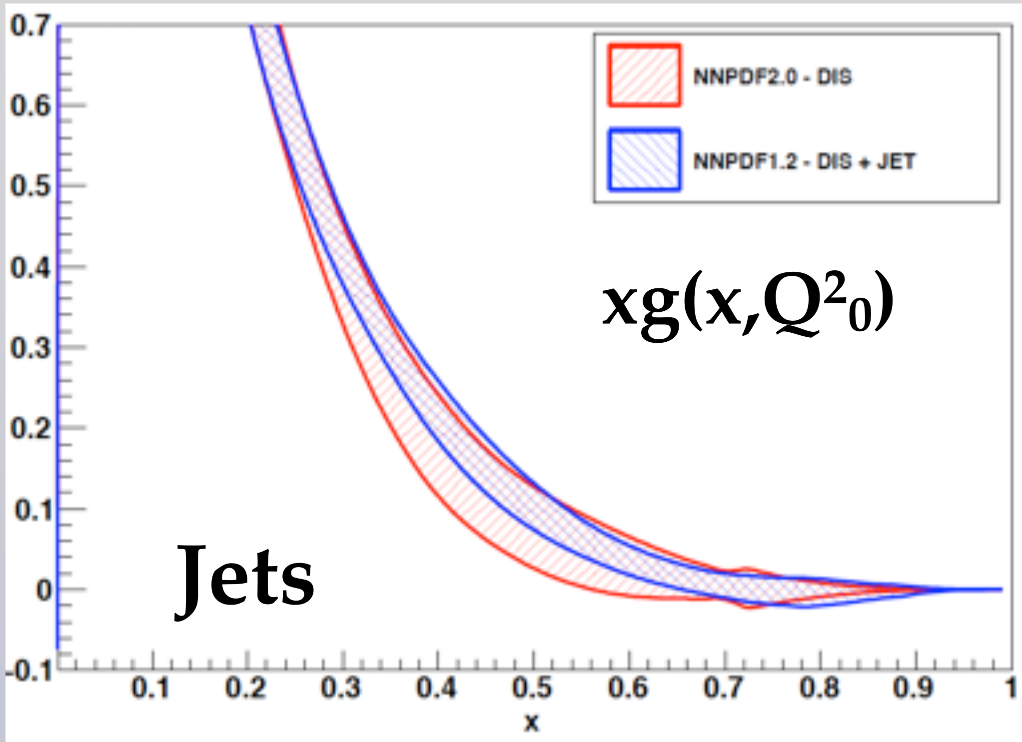
The **inclusive jet data** constrains large-x gluon

The **W and Z production** data from CMS, ATLAS and LHCb constrain medium-x antiquarks

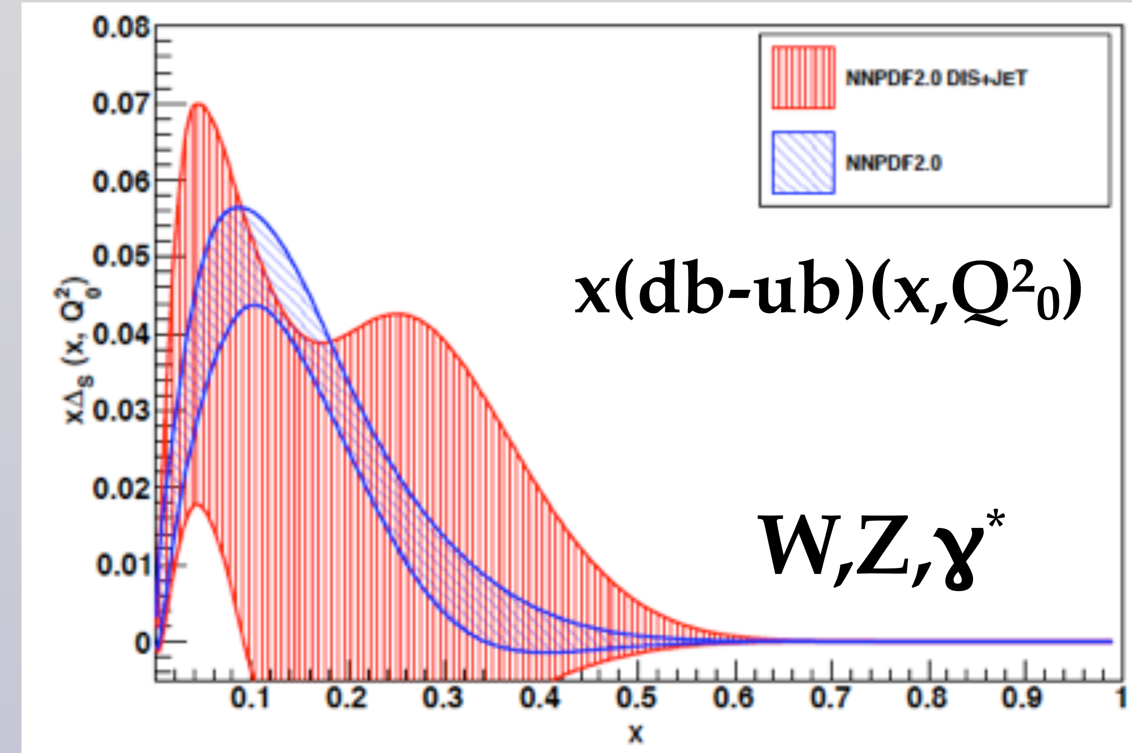




Impact of Tevatron and LHC data

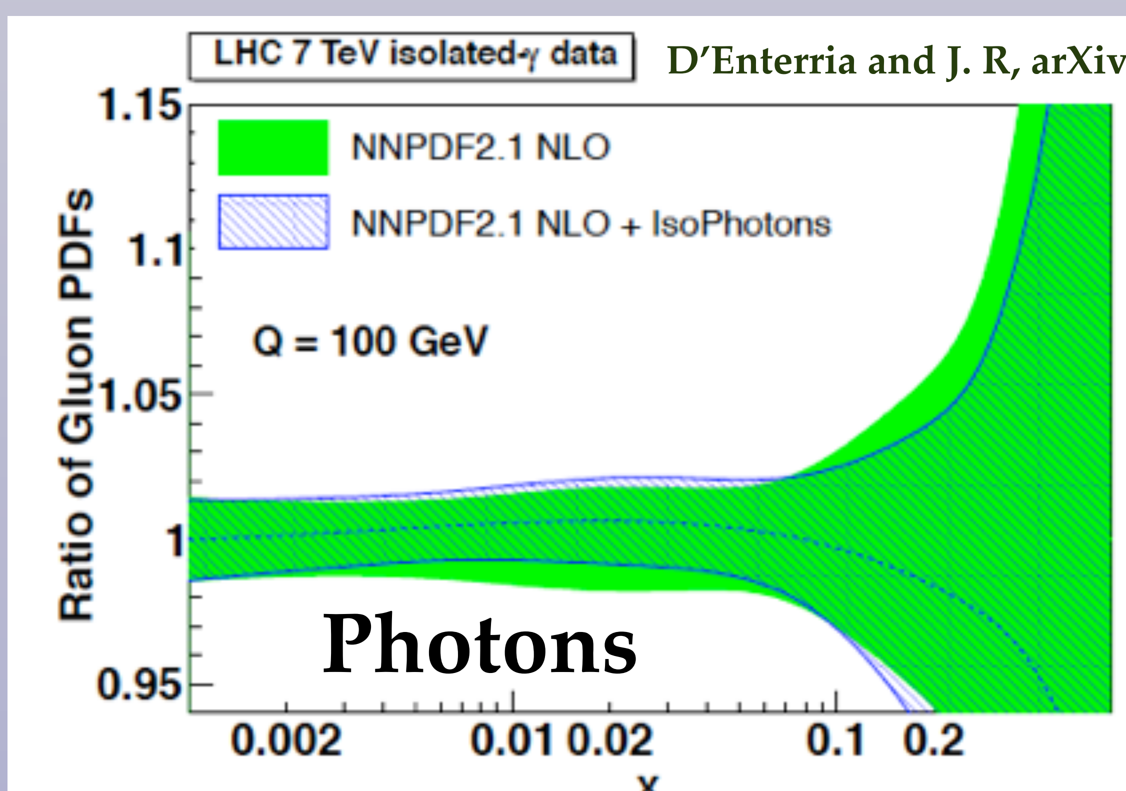


Inclusive **jets** pin down **large-x gluon**

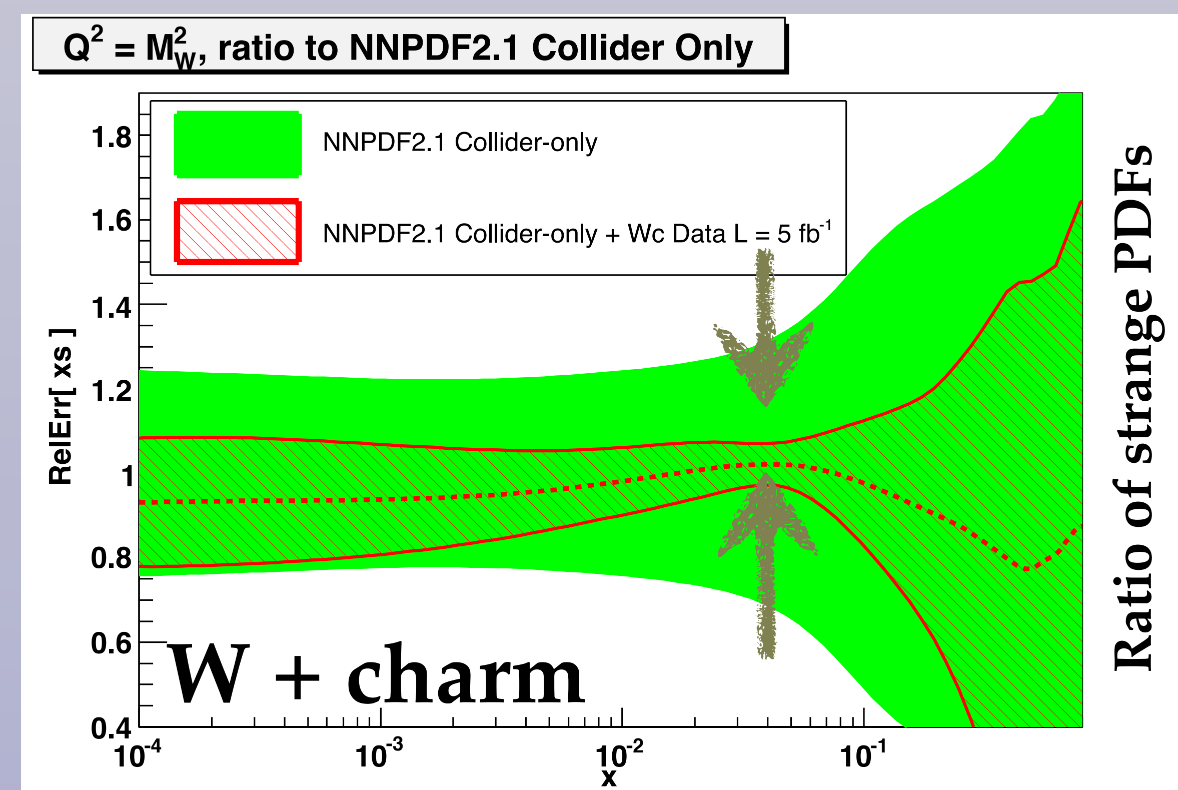


Drell-Yand and W,Z data determine **quark flavor separation**

D'Enterria and J. R, arXiv:1202.1762



Isolated photon LHC data constraints **gluons at medium-x**: relevant for **Higgs production in gluon fusion**



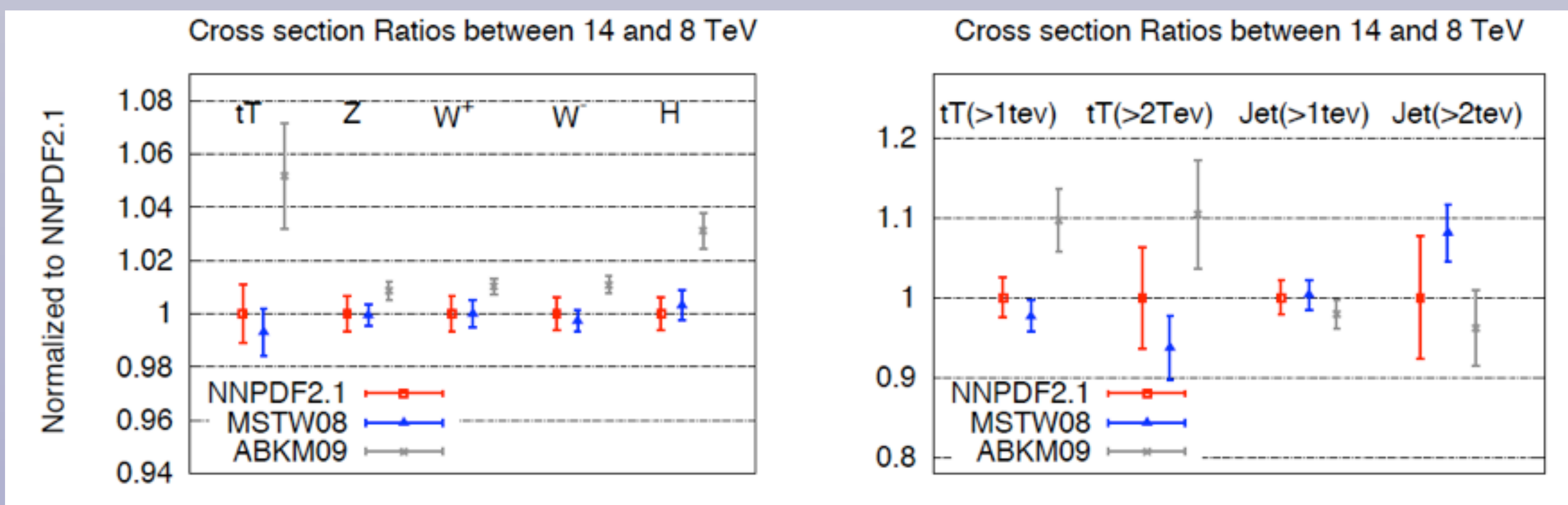
W production in association with charm quarks provides direct access to the proton strangeness

Cross section Ratios between 7, 8 and 14 TeV

- The **staged increase of the LHC beam energy** provides a new class of interesting observables: **cross section ratios** for different beam energies

$$R_{E_2/E_1}(X) \equiv \frac{\sigma(X, E_2)}{\sigma(X, E_1)} \qquad R_{E_2/E_1}(X, Y) \equiv \frac{\sigma(X, E_2)/\sigma(Y, E_2)}{\sigma(X, E_1)/\sigma(Y, E_1)}$$

- These ratios can be computed with **very high precision** due to the large degree of **correlation of theoretical uncertainties** at different energies
- **Experimentally** these ratios can also be measured accurately since many systematics, like luminosity or jet energy scale, **cancel partially in the ratios**
- These ratios allow **stringent precision tests of the SM**, like **PDF discrimination**



M. Mangano, J. Rojo, arXiv:1206.3557

Cross section Ratios between 7, 8 and 14 TeV

- If SM theory systematics under control, cross section ratios can show an improved sensitivity to New Physics than absolute cross sections

$$\sigma(pp \to X) = \sigma^{SM}(pp \to X) + \sigma^{BSM}(pp \to X)$$

- The **visibility of a BSM contribution** in the **evolution with energy** of the cross section requires that it evolves **differently from the SM contribution**

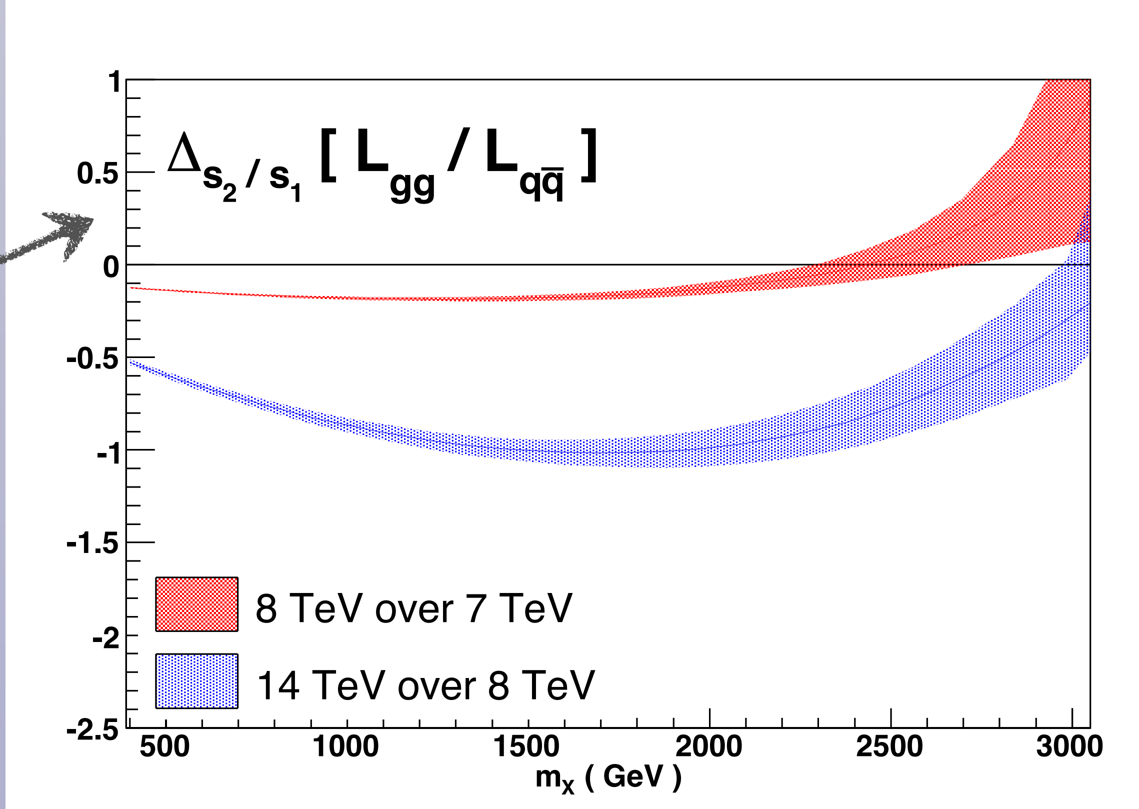
$$R^X_{E_1/E_2} \sim \frac{\sigma_X^{SM}(E_1)}{\sigma_X^{SM}(E_2)} \times \left\{ 1 + \frac{\sigma_X^{BSM}(E_1)}{\sigma_X^{SM}(E_1)} \Delta_{E_1/E_2}\left[\frac{\sigma_X^{BSM}}{\sigma_X^{SM}}\right]\right\}$$

$$\Delta_{E_1/E_2}(A) = 1 - \frac{A(E_2)}{A(E_1)}$$

Example: a **gluon-gluon initiated BSM** contribution to **high-mass Z production**. The cross section ratio enhanced by:

$$\frac{\sigma_Z^{\rm BSM}(m_X)}{\sigma_Z^{\rm SM}(m_X)} \Delta_{E_1/E_2}\left[\frac{\mathcal{L}_{gg}(m_X)}{\mathcal{L}_{q\bar{q}}(m_X)}\right]$$

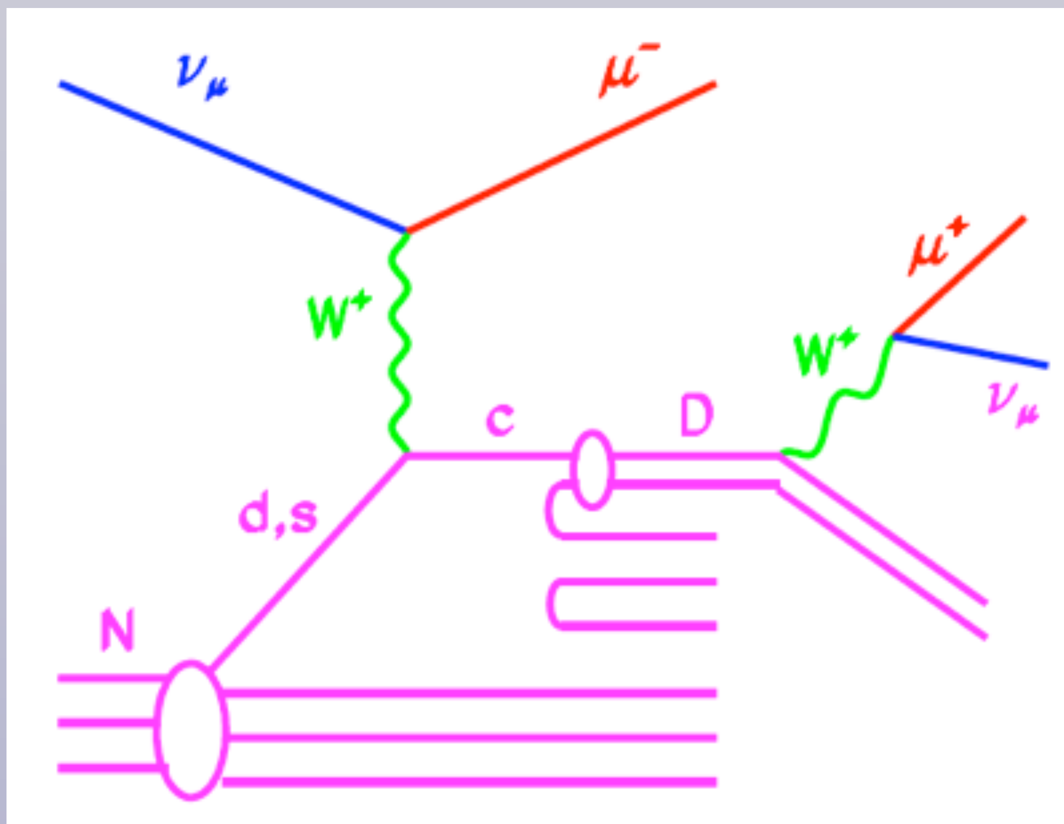
With greatly reduced experimental and theoretical uncertainties



But **theory systematics, mostly PDFs**, need to be known accurately for this new approach to show its **full potential**

Determination of Standard Model parameters

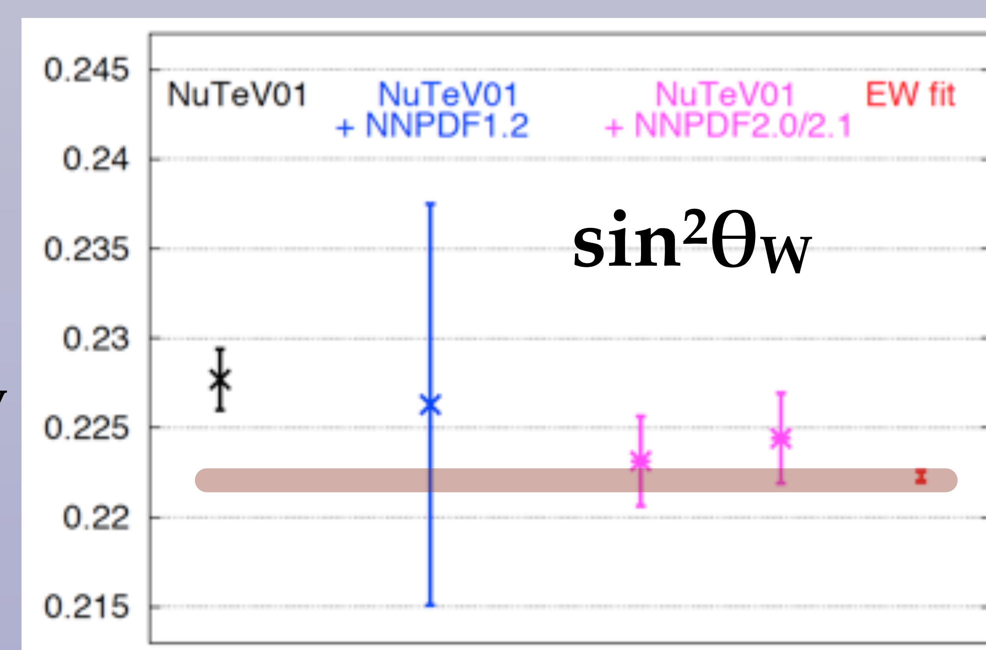
- **Accurate PDFs** are required for **precision determination of fundamental Standard Model parameters** in processes involving initial state hadrons
- These include, among many others, the **strong coupling constant** α_S, the **W boson mass**, the effective **lepton mixing angle**, **CKM** matrix elements, ...
- The **unbiased** nature of the NNPDF approach approach to **faithfully disentangle** PDF uncertainties from other parametric uncertainties. One example in neutrino DIS:



CKM matrix element V_{cs} can be determined from **neutrino DIS data** - but large uncertainties from **strange PDF**

NNPDF analysis manages to obtain the **most accurate ever determination** of V_{cs} from a single process:

V_{cs} = 1.04 ±0.06 (PDG average)
V_{cs} = 0.96 ±0.07 (NNPDF from NuTeV data)



The same analysis shows that the **strangeness asymmetry** in the proton has just the right size to **cancel the NuteV anomaly**

$$
\begin{aligned}
R_{PW} &= \frac{\sigma(\nu\mathcal{N}\to\nu X)-\sigma(\bar\nu\mathcal{N}\to\bar\nu X)}{\sigma(\nu\mathcal{N}\to\ell X)-\sigma(\bar\nu\mathcal{N}\to\bar\ell X)} \\
&= \frac{1}{2}-\sin^2\theta_W+\left[\frac{([U^-]-[D^-])+([C^-]-[S^-])}{[Q^-]}\frac{1}{6}\left(3-7\sin^2\theta_W\right)\right]
\end{aligned}
$$

NNPDF, arXiv:0906.1958

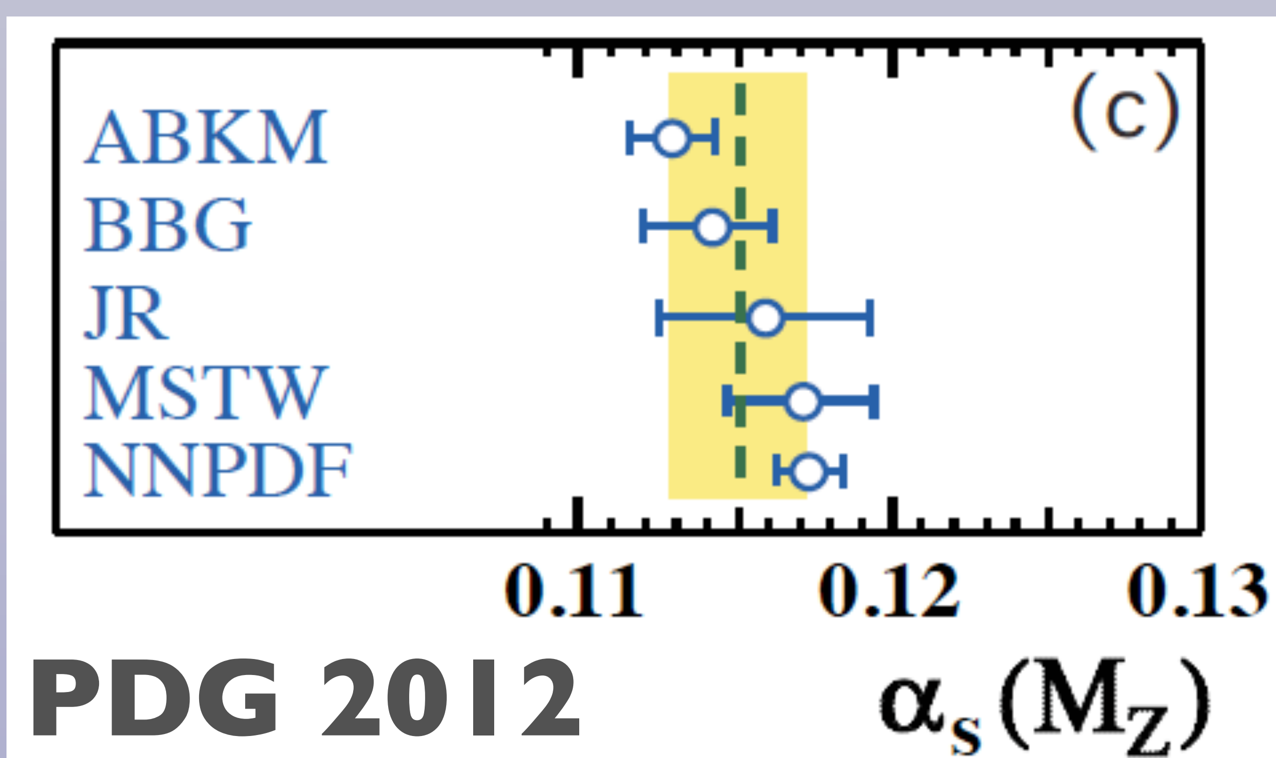
Determination of Standard Model parameters

- **Accurate PDFs** are required for **precision determination of fundamental Standard Model parameters** in processes involving initial state hadrons
- The **strong coupling constant** α_S can be determined from a global PDF analysis, mostly from scaling violations in Deep-Inelastic Scattering and in inclusive jet production
- The NNPDF result is the **most accurate determination** of α_S from a QCD global fit, and nicely consistent with the latest PDG average, to which is one of the dominant contributions
- In the pipeline: α_S determinations from LHC data at the **higher scale**s ever probed

PDG 2012 average

$$\alpha_s(M_Z^2) = 0.1184 \pm 0.0007$$

PDG average from PDF fits

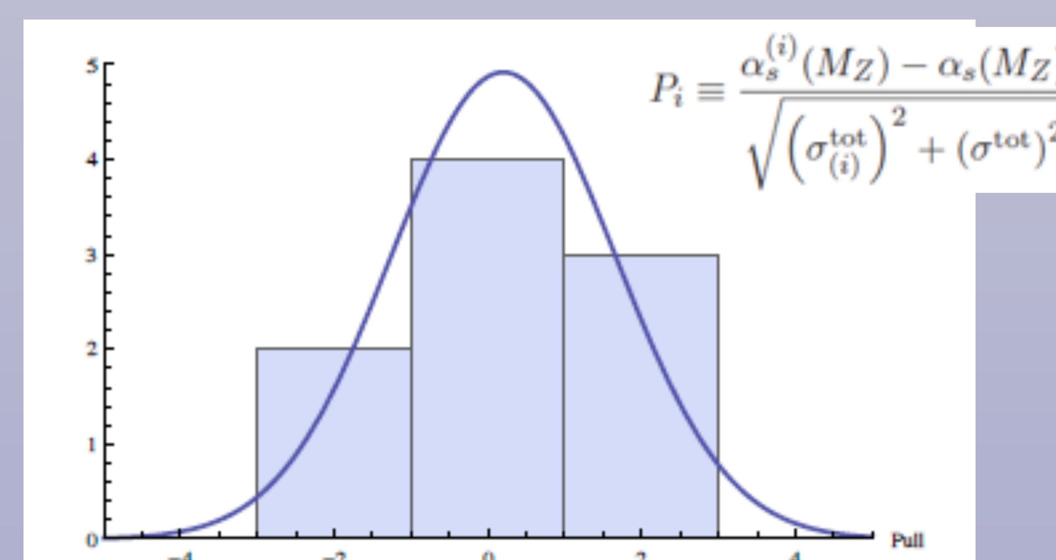


NNPDF2.1 NNLO

$$\alpha_s^{\rm NNLO}(M_Z) = 0.1173 \pm 0.0007^{\rm stat} \pm 0.0001^{\rm proc}$$

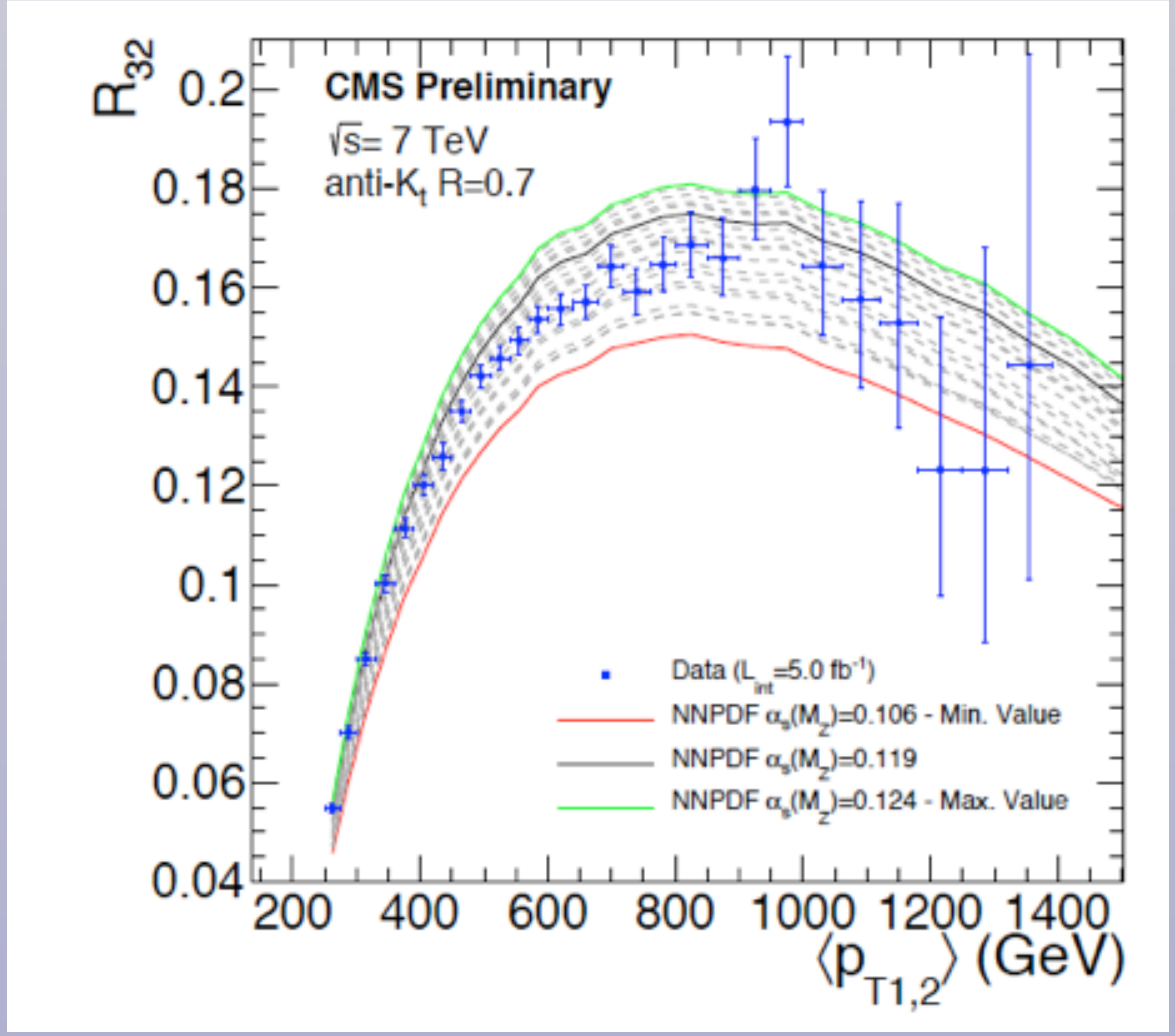
NNPDF, arXiv:1110.2483

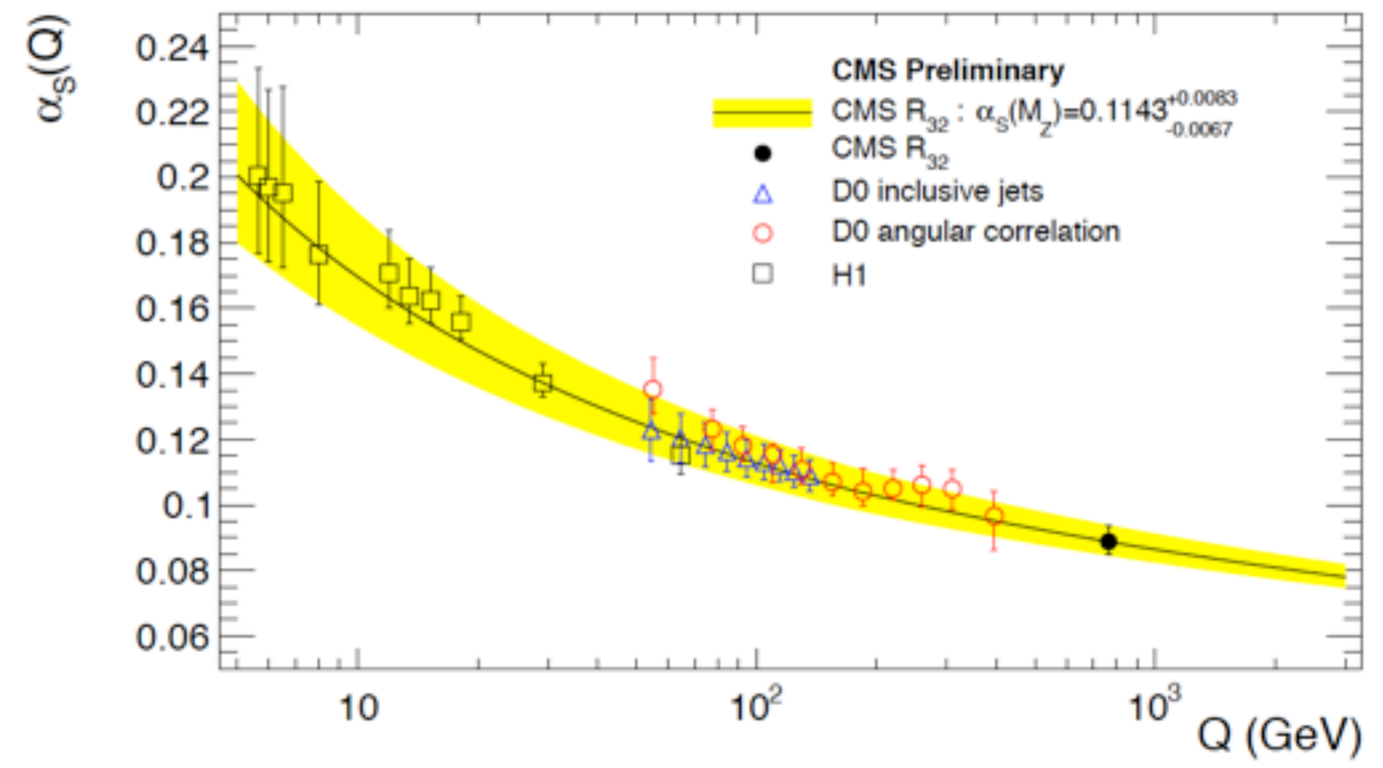
- **Consistency check of the global PDF framework**: the distributions of pulls for α_S fitted to **individual experiments** follows a Gaussian distribution



Determination of Standard Model parameters

- **Accurate PDFs** are required for **precision determination of fundamental Standard Model parameters** in processes involving initial state hadrons
- The **strong coupling constant** α_S can be determined from a global PDF analysis, mostly from scaling violations in Deep-Inelastic Scattering and in inclusive jet production
- CMS has recently determined α_S from the ratio of 3-jet to 2-jet cross sections at the LHC, providing the determination of the strong coupling at the highest scales ever probed, using **NNPDF2.1** as input



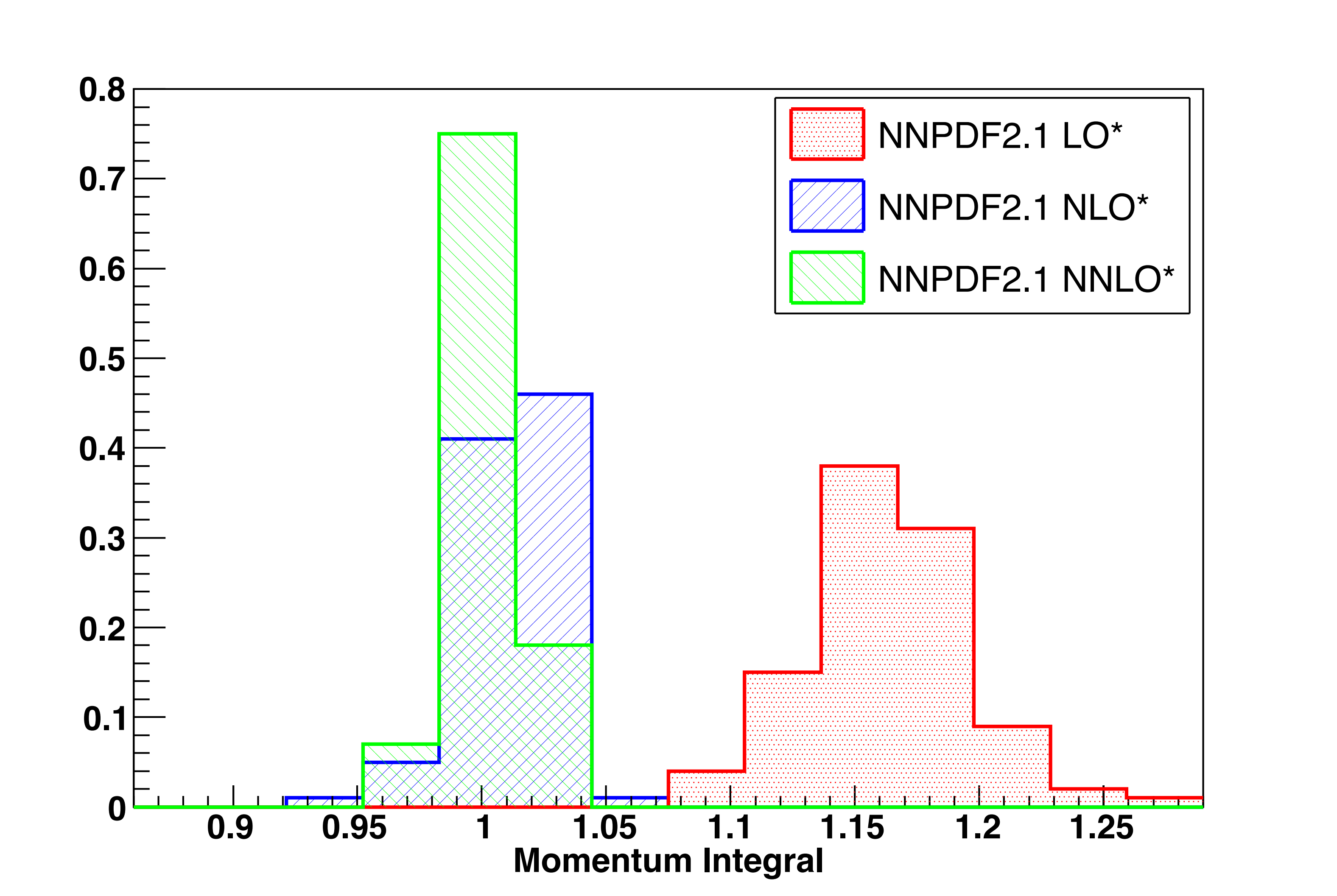


Precision tests of the Factorization Theorem

- Perturbative QCD requires that the **momentum integral** should be unity to all orders

$$[M]\left(Q^2\right) \equiv \int_0^1 dx \left(xg\left(x, Q^2\right) + x\Sigma\left(x, Q^2\right)\right)$$

- Is it possible to **determine** the value of the momentum integral from the global PDF analysis, rather than **imposing it?** Check in LO*, NLO* and NNLO* fits **without setting M=1**



$$[M]_{\mathrm{LO}} = 1.161 \pm 0.032\,,$$
$$[M]_{\mathrm{NLO}} = 1.011 \pm 0.018\,,$$
$$[M]_{\mathrm{NNLO}} = 1.002 \pm 0.014\,.$$

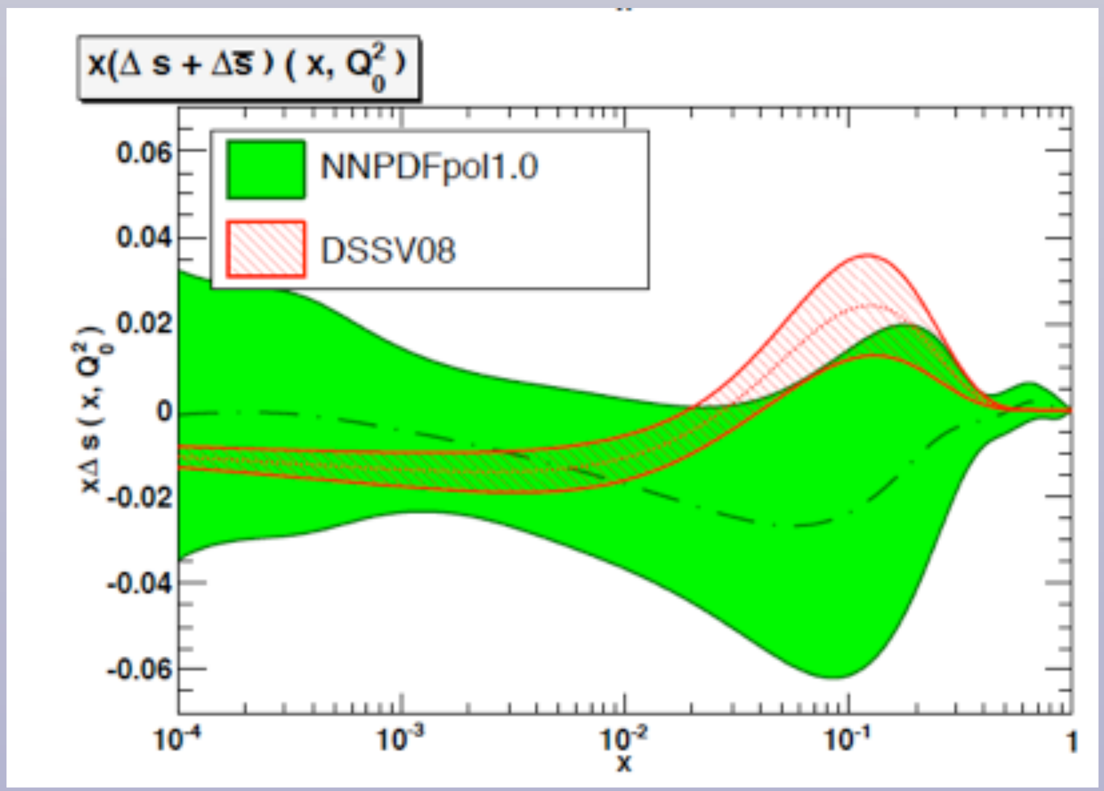
- Experimental data beautifully **confirms the pQCD expectation**
- **Extremely non trivial test** of the global analysis framework and the **factorization hypotheses**
- Very good convergence of the QCD perturbative expansion

PDF prospects at the LHC

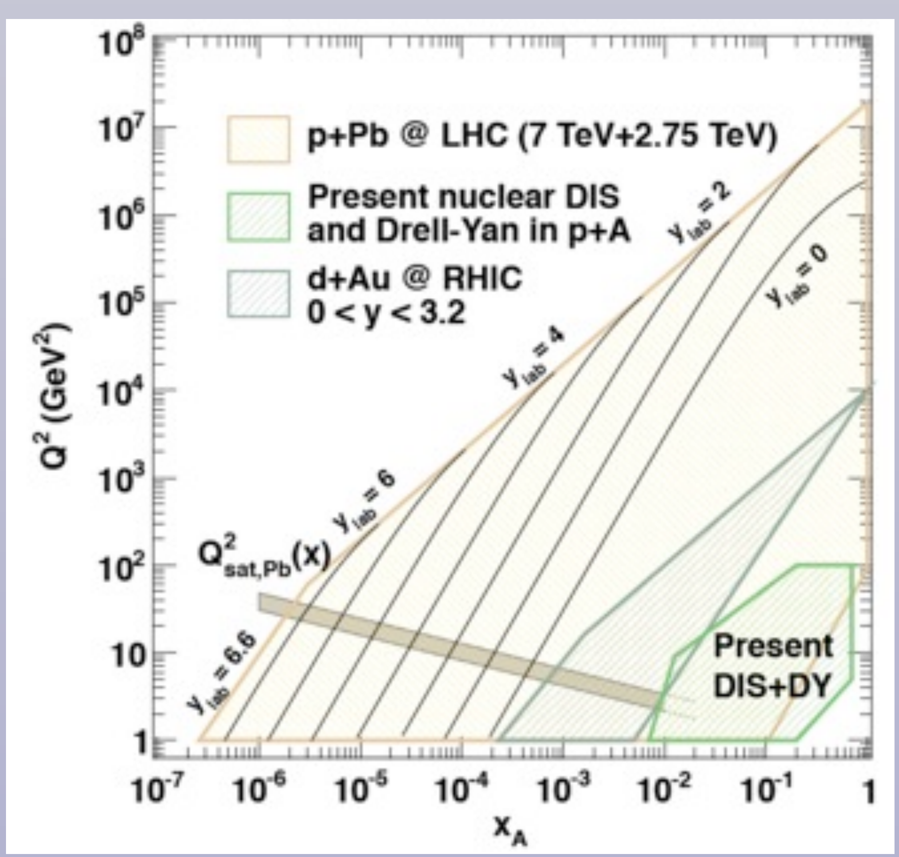
- From the **experimental data point of view,** all the **current and future needs of the LHC in terms of PDFs** can be addressed by a **specific PDF program at the LHC,** without the need of new facilities
- There is a long list of measurements to be pursued, that will provide all required information on PDFs:
    - Inclusive jets and dijets, central and forward: **large-x quarks and gluons**
    - Isolated photons: **medium-x gluons**
    - Inclusive W and Z production and asymmetries: **quark flavor separation, strangeness**
    - W production with charm quarks: **direct handle on strangeness**
    - W production with jets: **medium small-x gluon**
    - Off resonance DY and W production at small and high mass: **quarks at very small and very large-x**
    - Top quark distributions: **large-x gluon**
    - Z+charm: **intrinsic charm PDF**
    - Single top production: **gluon and bottom PDFs**
    - Charmonium production: **small-x gluon**
- Some of these have/are being already carried out, and LHC data is already being used in PDF fits like **NNPDF2.3.** Constraints are expected to be larger with the **full 8 TeV dataset** and with **13/14 TeV** data
- To maximize the **LHC data impact on PDFs**, it is crucial to **coordinate a detailed PDF program** between the LHC experiments and the Theory community

Beyond unpolarized PDFs

- The **NNPDF methodology** can be applied to many other closely related problems
  - **Polarized** parton distributions: The **spin content of the proton**
  - **Nuclear** parton distributions: Initial Conditions for **Quark-Gluon Plasma studies** at the LHC
  - Hadron f**ragmentation** functions
  - Transverse momentum dependent PDFs, Generalized PDFs, ...
- NNPDF is already working on **polarized PDFs** (paper to appear next week) and **nuclear PDFs**. Other groups use **NNPDF-like technology** in their QCD analysis



NNPDFpol1.0: unbiased determination of the spin content of the proton
Substantial error underestimation in the standard polarized approach



N3PDFs: unbiased determination of nuclear PDFs from Proton-Lead LHC data: crucial input for QGP characterization

Summary

- **Parton Distributions** are an essential ingredient for LHC phenomenology
- Accurate PDFs are required for **precision SM measurements, Higgs characterization** and many **New Physics searches**
- The determination of **fundamental SM parameters** like the **W mass** or α_S **from LHC data** also greatly benefit from improved PDFs
- The NNPDF approach provides parton distributions based on a **robust, unbiased methodology,** the most updated **theoretical information** and all the relevant hard scattering data **including LHC data**
- Near future developments in NNPDF:
  - **Inclusion of more LHC data**: 7 and 8 TeV W, Z, dijets, top distributions, photons, W +charm, W,Z+jets, high mass off resonance W, ...
  - Inclusion of the **complete HERA-II** inclusive and charm dataset
  - PDFs for **NLO Monte Carlo event generators** at the LHC
  - **PDFs with QED** and electroweak effects, and PDFs with **Intrinsic Charm**
  - PDFs with **threshold and high energy resummation**

Extra Material

Genetic Algorithms: Example

Maximisation of $f(x) = x^2$ on the interval $x \in [0, 31]$

1. **Encode** our problem parameter x into a string, the *chromosome*, on which the GA can then operate. Possibility: binary encoding, $x = 1$ codes as 00001 and $x = 31$ as 11111.
2. Create at random the initial population with fixed number of individuals $i = 1, \ldots, N$. We take $N = 4$ for illustration. Fitness calculated with the function to maximise: $f(x) = x^2$

i	Genotype	Phenotype x_i	Fitness $f_i = f(x_i)$	$f_i / \sum f_i$
1	01101	13	169	0.14
2	11000	24	576	0.49
3	01000	8	64	0.06
4	10011	19	361	0.31

The first child generation after selection and crossover:

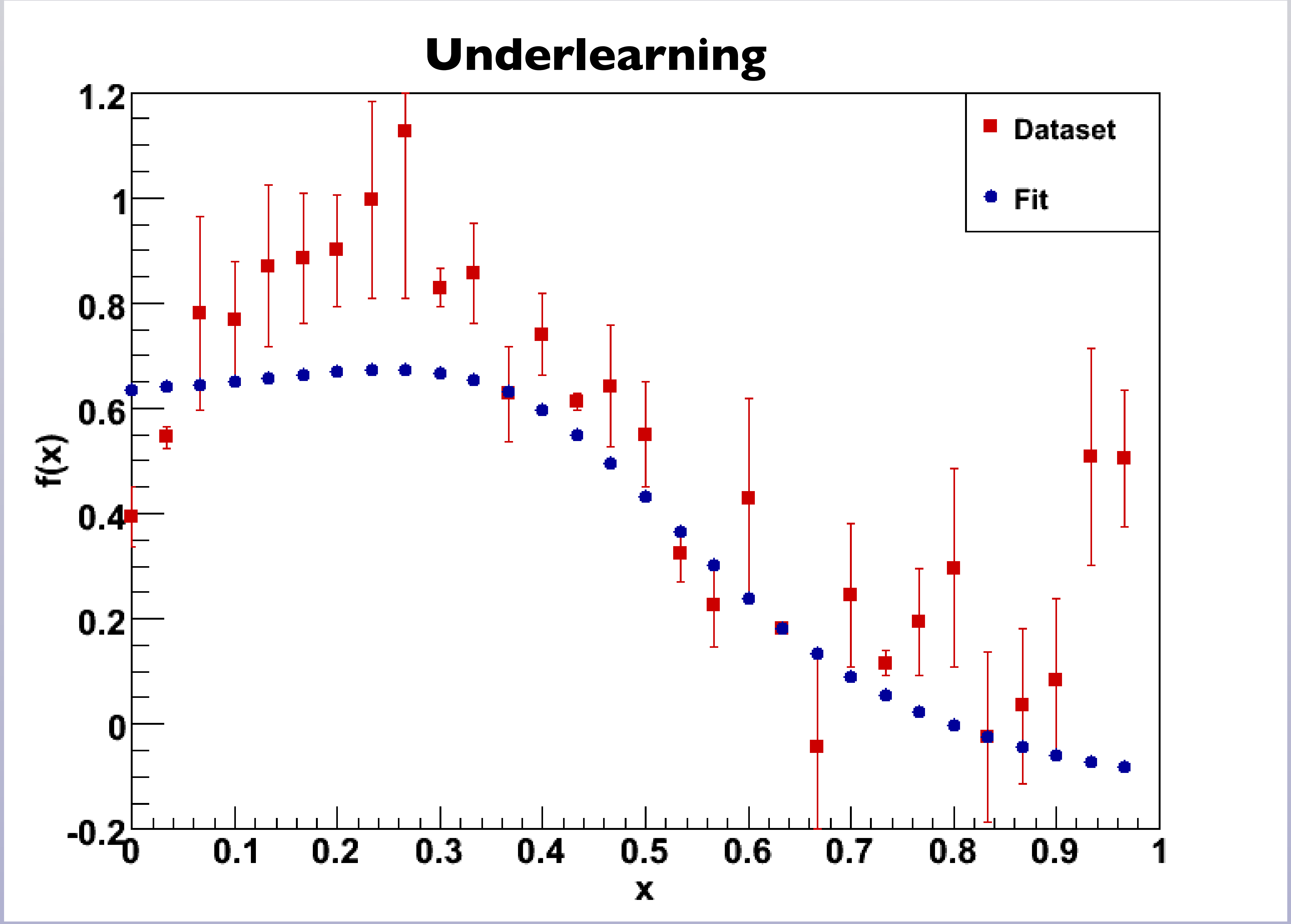
i	Genotype	Phenotype x_i	Fitness $f_i = f(x_i)$	$f_i / \sum f_i$
5	01100	12	144	0.08
6	11001	25	625	0.36
7	11011	27	729	0.42
8	10000	16	256	0.14

The rapid increase of fitness over the very first few generations is a common feature of GAs.

Example from B. Allanach et al, hep-ph/0406277

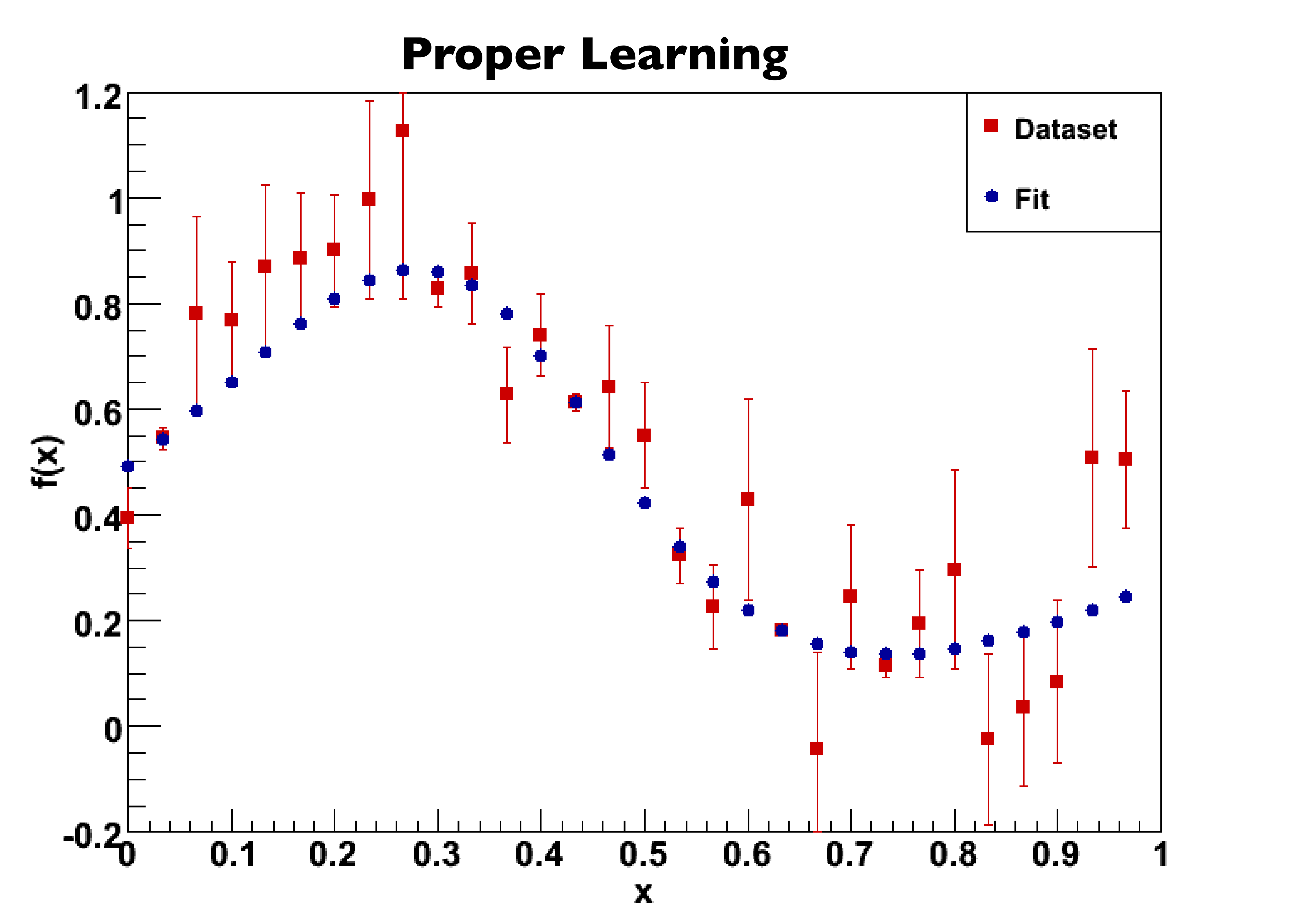
Filtering Noise: Cross-Validation Stopping

With a **flexible PDF parametrization** as ANNs one can reach the point of fitting **the statistical fluctuations** on the data - on top of the **underlying physical law**



Filtering Noise: Cross-Validation Stopping

With a **flexible PDF parametrization** as ANNs one can reach the point of fitting **the statistical fluctuations** on the data - on top of the **underlying physical law**



Filtering Noise: Cross-Validation Stopping

With a **flexible PDF parametrization** as ANNs one can reach the point of fitting **the statistical fluctuations** on the data - on top of the **underlying physical law**