Machine Learning in the JAM Framework

Wally Melnitchouk

— Jefferson Lab —

JLab Angular Momentum (JAM) Collaboration https://www.jlab.org/jam

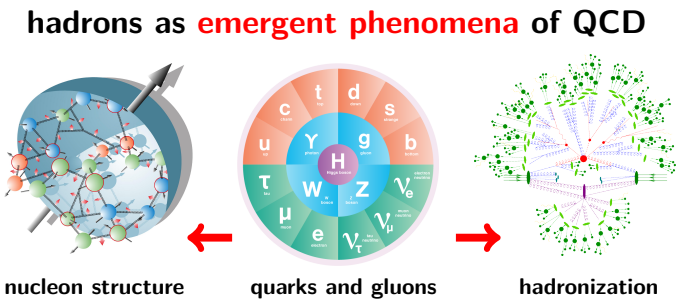
- Machine learning: "the study of computer algorithms that improve automatically through experience"

- Supervised learning: computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs
  - χ^2 minimization is an example of ML
  - JAM (and all other global QCD analyses) = supervised learning

- Unsupervised learning: allow computer to find structure in its input on its own
  - examples include *k*-means clustering, GANs, auto-encoders, …
  - slowly making its way into global QCD analysis…

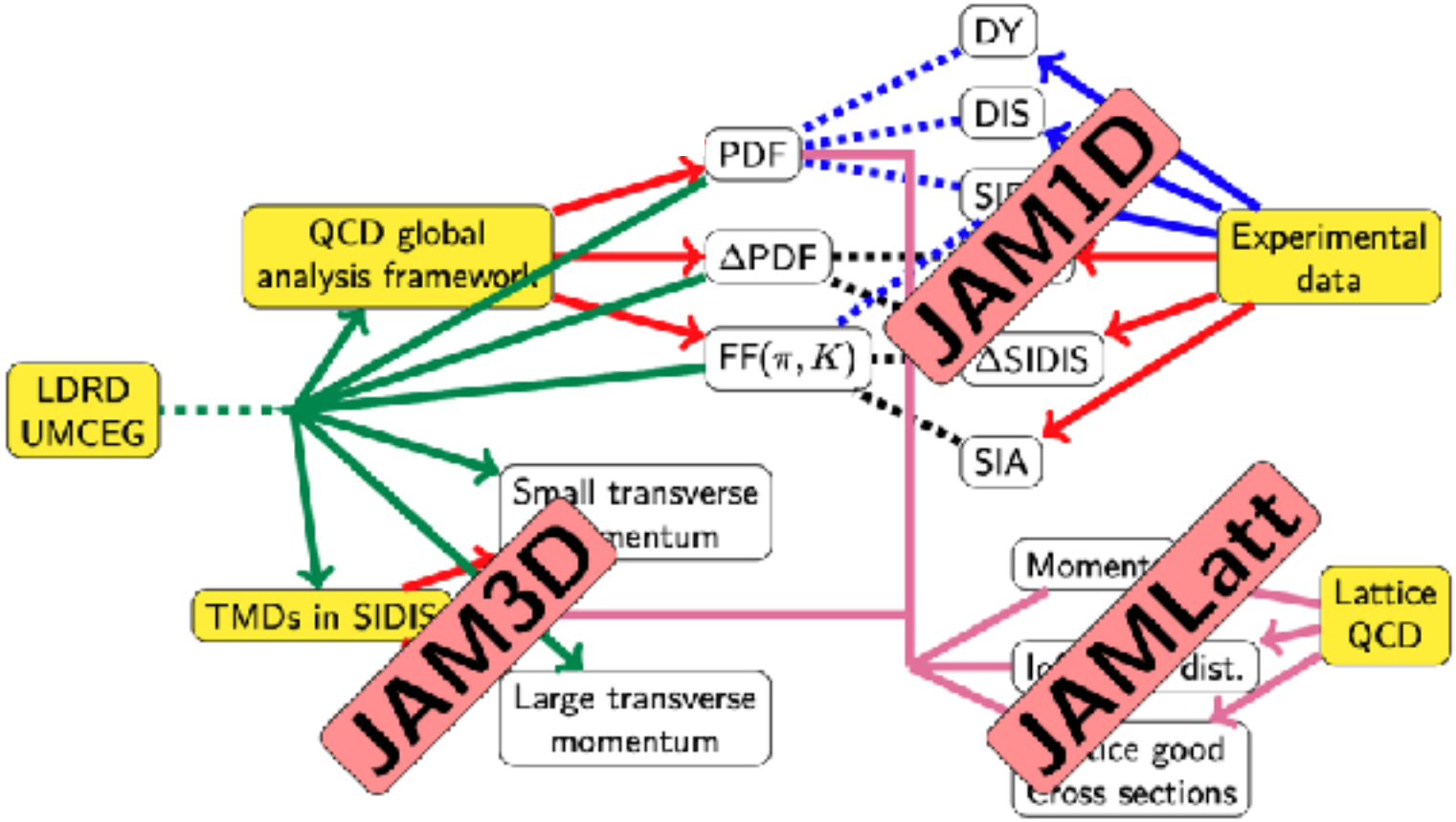
- Goal of JAM (Jefferson Lab Angular Momentum) Collaboration
  — learn about the nucleon's internal 3-D quark and gluon structure — "femtography"

- Quarks & gluons confined inside hadrons, never observed directly
  — must infer information about structure indirectly



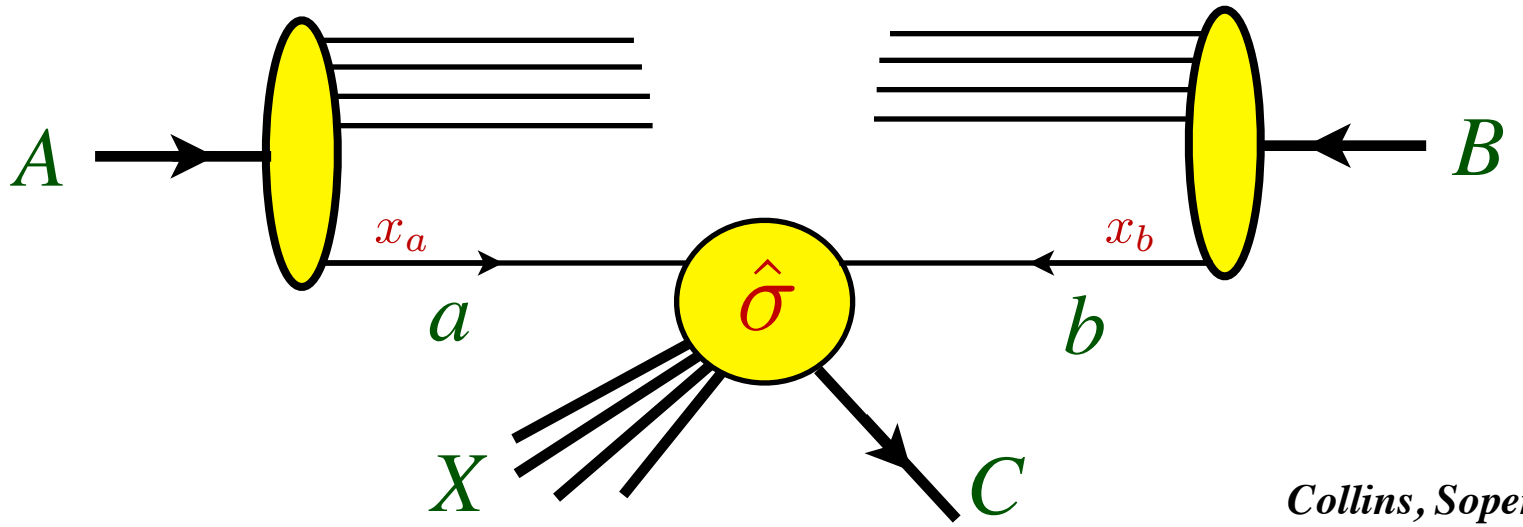
— structure information parameterized in terms of <u>quantum correlation functions</u> (QCFs),
e.g. parton distribution functions (PDFs), fragmentation functions (FFs), transverse momentum dependent distributions (TMDs), generalized parton distributions (GPDs), …

- Goal of JAM (Jefferson Lab Angular Momentum) Collaboration — learn about the nucleon's internal 3-D quark and gluon structure — "femtography"



- Problem requires theory (QCD factorization) and global QCD analysis tools to provide map between observables and QCFs

Factorization of inclusive particle-production cross sections



Collins, Soper, Sterman (1980s)

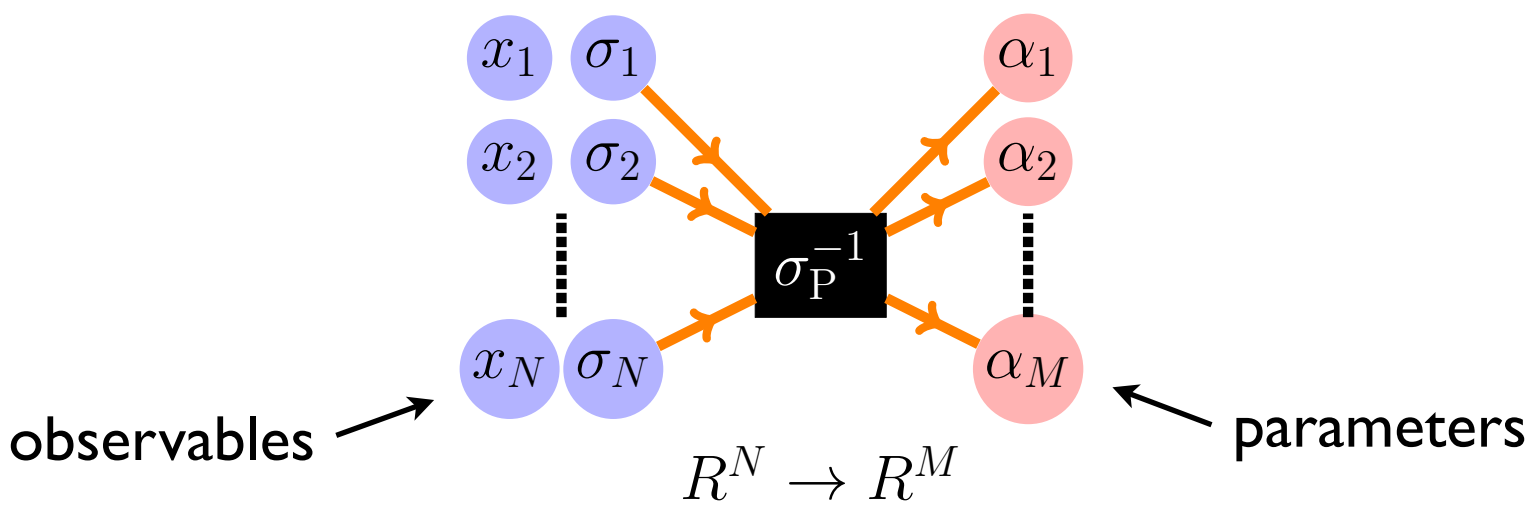
→ QCD factorization: separation of hard (perturbative, calculable) from soft (nonperturbative, parameterized) physics

$$\sigma_{AB\to CX}(p_A,p_B) = \sum_{a,b}\int dx_a\, dx_b\, \boxed{f_{a/A}(x_a,\mu)}\, f_{b/B}(x_b,\mu) \times \sum_n \alpha_s^n(\mu)\, \hat{\sigma}^{(n)}_{ab\to CX}\,(x_a p_A, x_b p_B, Q/\mu)$$

→ process-independent parton distribution functions $f_{a/A}$ characterizing structure of bound state A

- Supervised learning is on *cross sections* σ, <u>not</u> on QCFs
  — goal is not to mimic cross sections, but to *infer QCFs*
  — QCFs are not supervisable $\longrightarrow$ <u>inverse problem</u>

- In practice, analytical solutions for the inverse function for QCFs <u>do not</u> exist
  — can we use machine learning tools to find approximate numerical solutions for the inverse function?

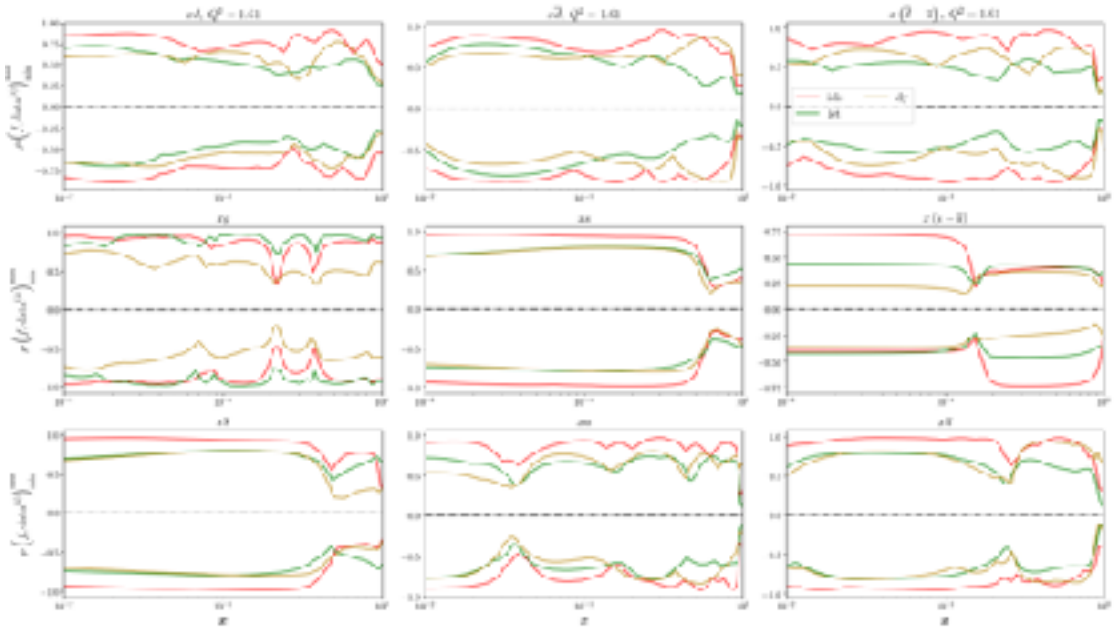


- Shortcomings of existing global QCD analysis paradigms

— need to understand quantitatively correlations between QCFs and cross sections across a range of kinematics

— what is impact of specific data sets (current & future) on particular QCFs?

— requires changing training data to explore all possibilities… but combinations grow factorially — not practical!



Yiyu Zhou (2020)

Global QCD analysis

- Universality of PDFs allows data from many different processes (DIS, SIDIS, weak boson/jet production in pp, Drell-Yan …) to be analyzed simultaneously

$\longrightarrow$ $\ell\, N \to \ell'\, X$ $\qquad d\sigma^{\rm DIS} \sim \sum_q e_q^2\; q(x)$

$\longrightarrow$ $\ell\, N \to \ell'\, h\, X$ $\qquad d\sigma^{\rm SIDIS} \sim \sum_q e_q^2\; q(x)\, D_q^h(z)$

$\longrightarrow$ $\ell\, \bar{\ell} \to h\, X$ $\qquad d\sigma^{\rm SIA} \sim \sum_q e_q^2\; D_q^h(z)$

- Extraction of PDFs is challenging because usually there exist multiple solutions

$\longrightarrow$ need to avoid fit getting stuck in a local minimum, or parameter ranges too restrictive

Global QCD analysis

- Analysis of data requires estimating expectation values E and variances V of "observables" $\mathcal{O}$ (functions of PDFs) which are functions of parameters

$$E[\mathcal{O}] = \int d^n a \, \mathcal{P}(\vec{a}|\text{data}) \, \mathcal{O}(\vec{a})$$

$$V[\mathcal{O}] = \int d^n a \, \mathcal{P}(\vec{a}|\text{data}) \, \left[\mathcal{O}(\vec{a}) - E[\mathcal{O}]\right]^2$$

"Bayesian master formulas"

- Using Bayes' theorem, probability distribution $\mathcal{P}$ given by

$$\mathcal{P}(\vec{a}|\text{data}) = \frac{1}{Z} \, \mathcal{L}(\text{data}|\vec{a}) \, \pi(\vec{a})$$

in terms of the likelihood function $\mathcal{L}$ and priors π

Global QCD analysis

- Likelihood function

$$\mathcal{L}(\text{data}|\vec{a}) = \exp\left(-\frac{1}{2}\chi^2(\vec{a})\right)$$

is a Gaussian form in the data, with χ^2 function

$$\chi^2(\vec{a}) = \sum_i \left(\frac{\text{data}_i - \text{theory}_i(\vec{a})}{\delta(\text{data})}\right)^2$$

with priors $\pi(\vec{a})$ and evidence Z

$$Z = \int d^n a\, \mathcal{L}(\text{data}|\vec{a})\, \pi(\vec{a})$$

$\longrightarrow$ Z tests if $e.g.$ an n-parameter fit is statistically different from $(n{+}1)$-parameter fit

Global QCD analysis

- Standard method for evaluating E, V via maximum likelihood

  $\longrightarrow$ maximize probability distribution

$$\mathcal{P}(\vec{a}|\text{data}) \;\rightarrow\; \vec{a}_0$$

  $\longrightarrow$ if $\mathcal{O}$ is linear in parameters, and if probability is symmetric in all parameters

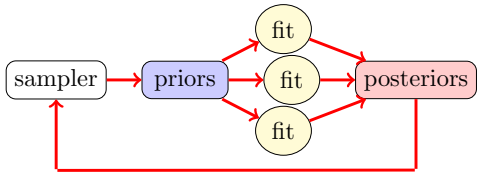
$$E[\mathcal{O}(\vec{a})] = \mathcal{O}(\vec{a}_0)\,, \quad V[\mathcal{O}(\vec{a})] \rightarrow \text{Hessian} \qquad H_{ij} = \frac{1}{2}\frac{\partial\chi^2(\vec{a})}{\partial a_i \partial a_j}\Big|_{\vec{a}=\vec{a}_0}$$

- In practice, since in general $E[f(\vec{a})] \neq f(E[\vec{a}])$, maximum likelihood method sometimes fails

  $\longrightarrow$ need more robust (Monte Carlo) approach

$$E[\mathcal{O}] \;\approx\; \frac{1}{N}\sum_k \mathcal{O}(\vec{a}_k)\,, \qquad V[\mathcal{O}] \;\approx\; \frac{1}{N}\sum_k \left[\mathcal{O}(\vec{a}_k) - E[\mathcal{O}]\right]^2$$

Global QCD analysis



- JAM — iterative, multi-step Monte Carlo

→ traditional functional form for distributions

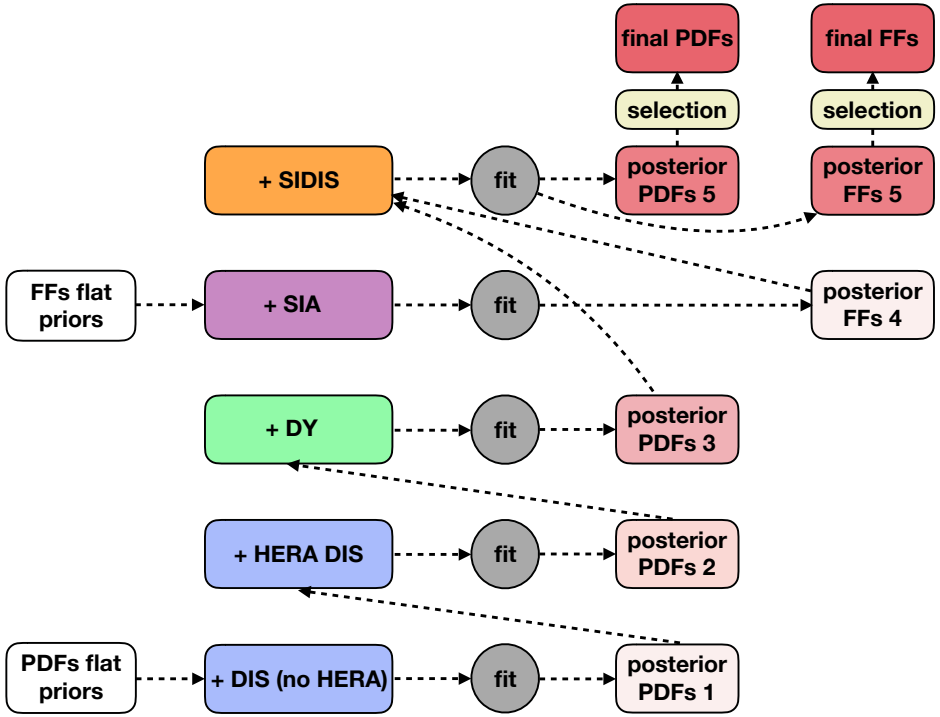
$$f(x) = N\, x^{\alpha} (1-x)^{\beta}\, P(x)$$

polynomial, neural net*, …

but sample large parameter space

→ sequentially include new types of observables to bootstrap way into relevant corner of parameter space

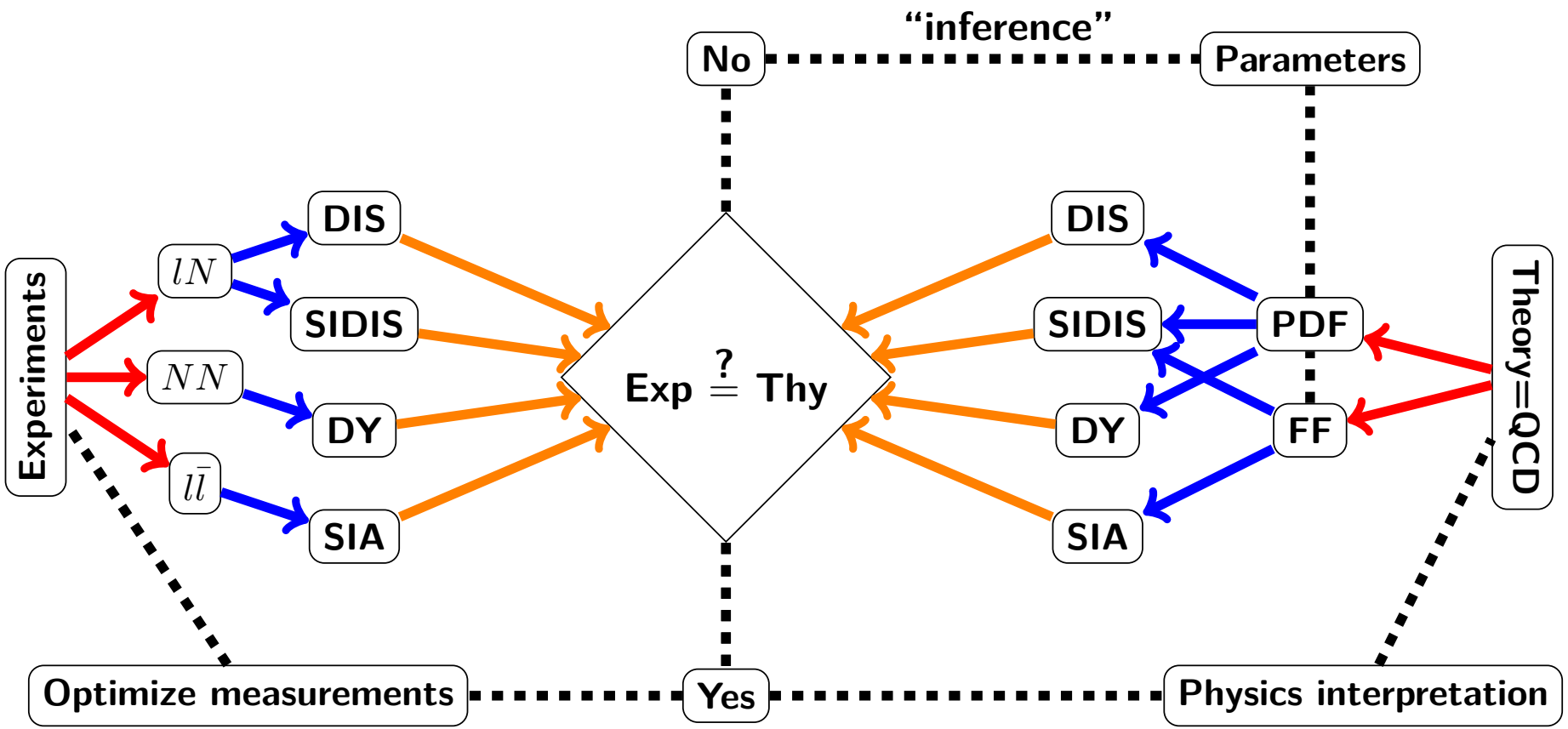
→ robust determination of PDF uncertainties



* neural net $\neq$ unbiased fit

Global QCD analysis

- Current state-of-the-art paradigm (in a nutshell)



JAM 2019 analysis

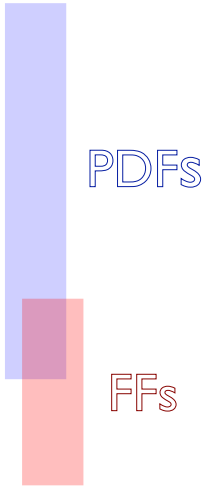
- First global analysis that fits both unpolarized PDFs and FFs <u>simultaneously</u>

  → fixed-target DIS on p, d (SLAC, BCDMS, NMC),
  HERA collider data (runs I & II)

  → Drell-Yan (Fermilab E866), jet production (CDF, D0)

  → SIDIS pion & kaon multiplicities for deuteron (COMPASS)

  → e^+e^- annihilation (DESY, LEP/CERN, SLAC, KEK)

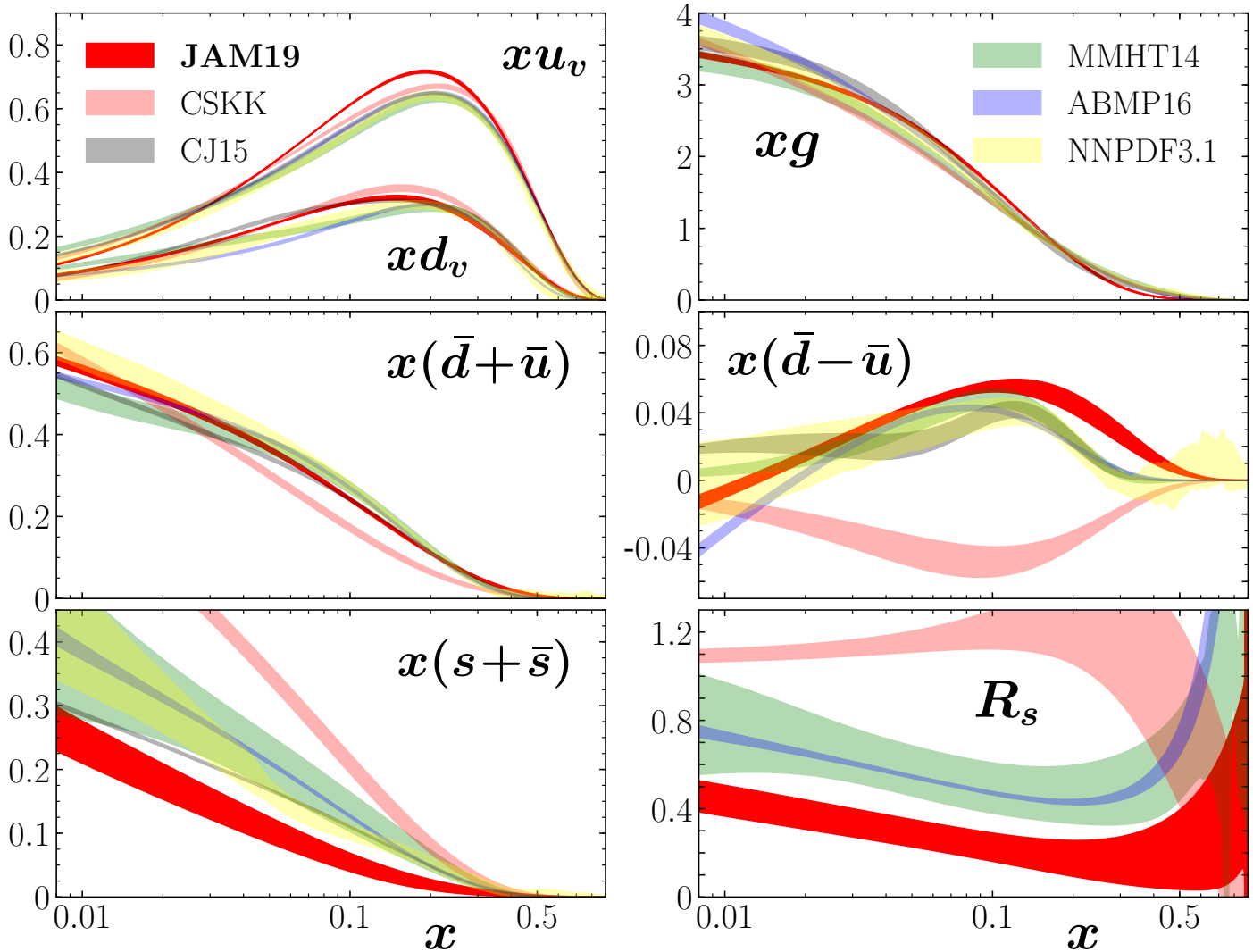


- 52 shape parameters + 41 "nuisance" parameters for systematic uncertainties (data normalizations)

- 953 fits to 4366 data points (2680 DIS, 992 SIDIS, 250 DY, 444 SIA)

  → such an analysis never attempted before…

JAM 2019 analysis



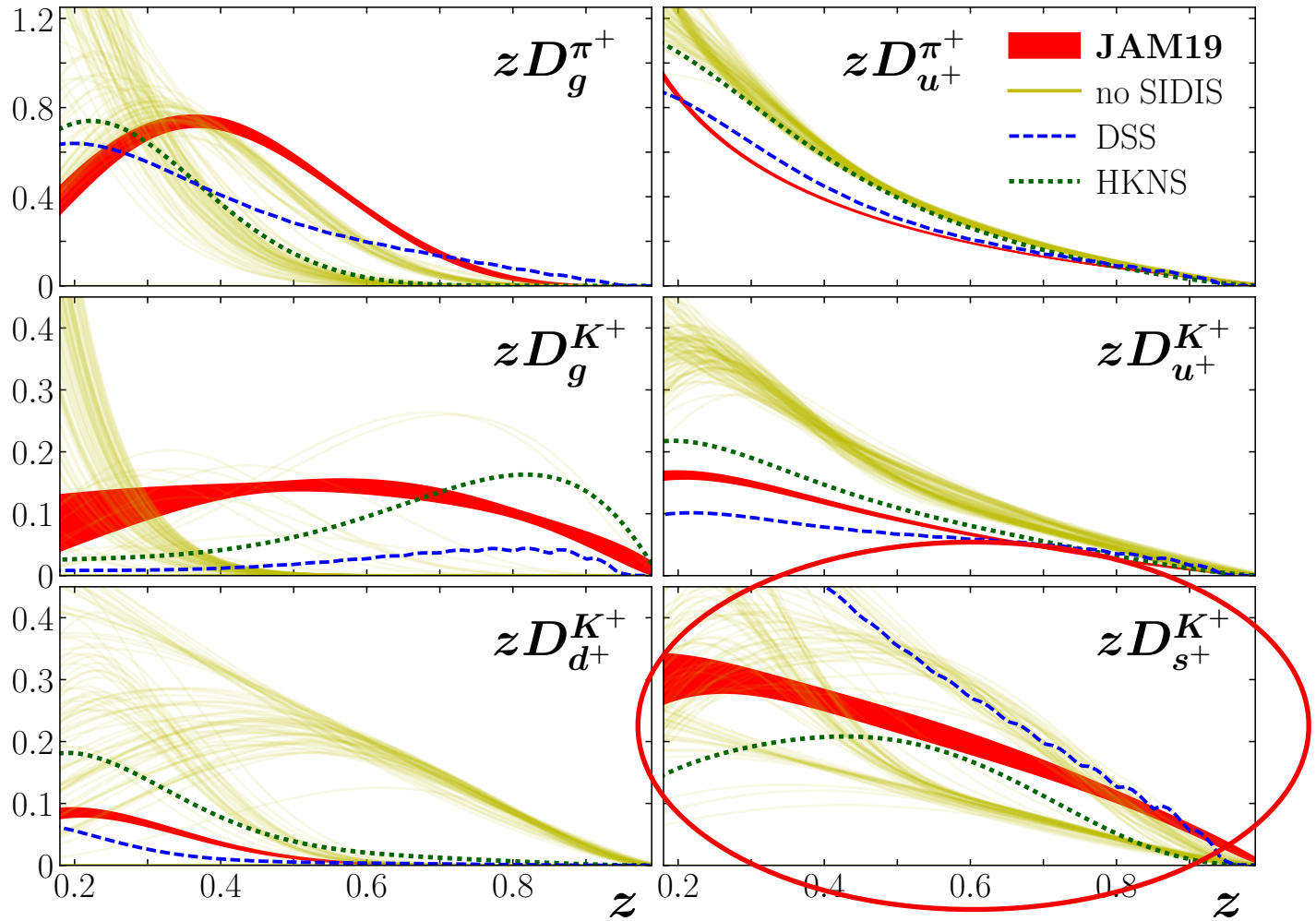
mean reduced
$\chi^2 = 1.3$
for all data

Sato, Andres, Ethier, WM (2019)

→ valence & light sea quark broadly in agreement with other groups

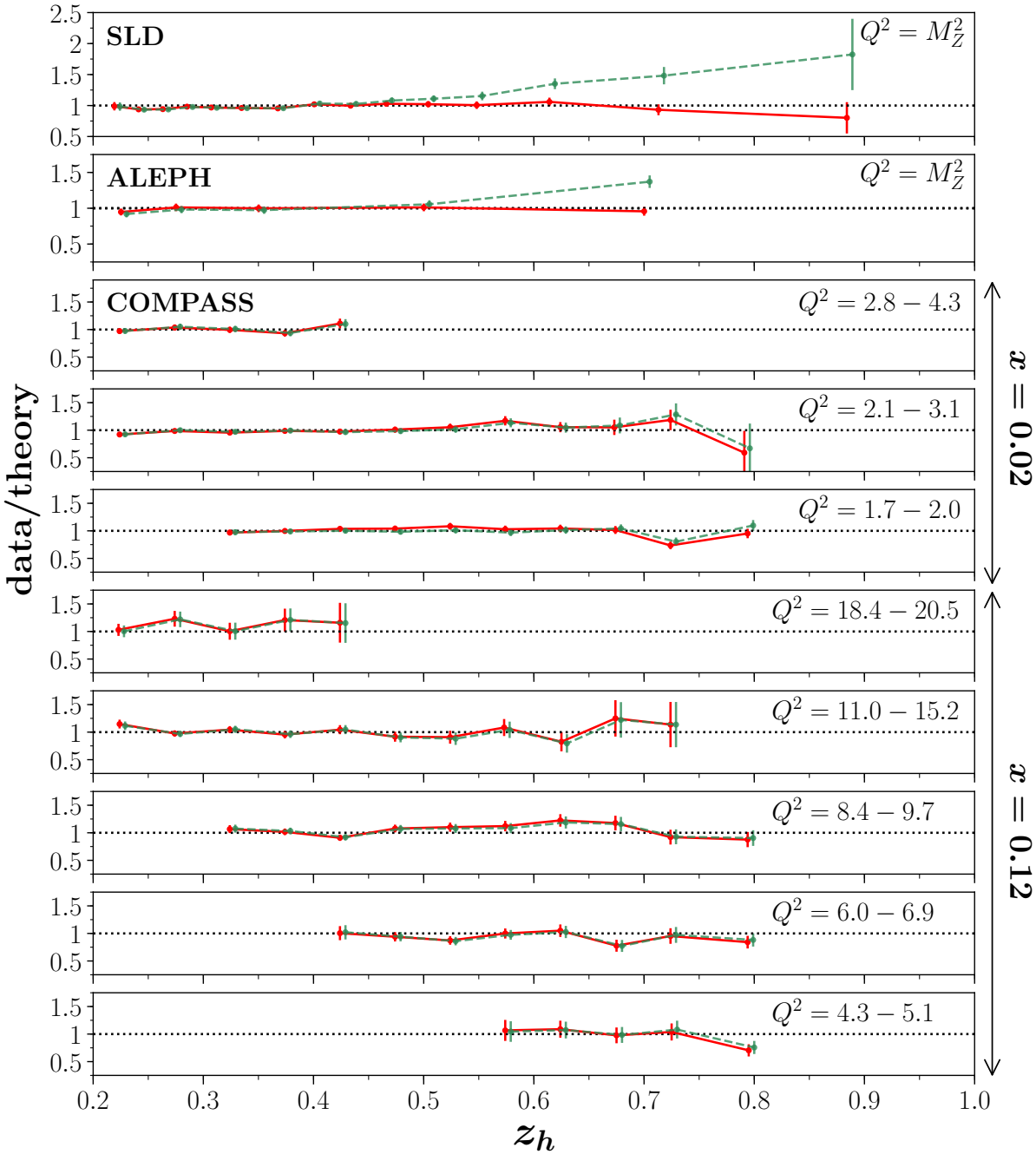
→ <u>suppression of strange</u> PDF compared to other extraction

JAM 2019 analysis



→ SIDIS + SIA data force strange to kaon FF to be larger

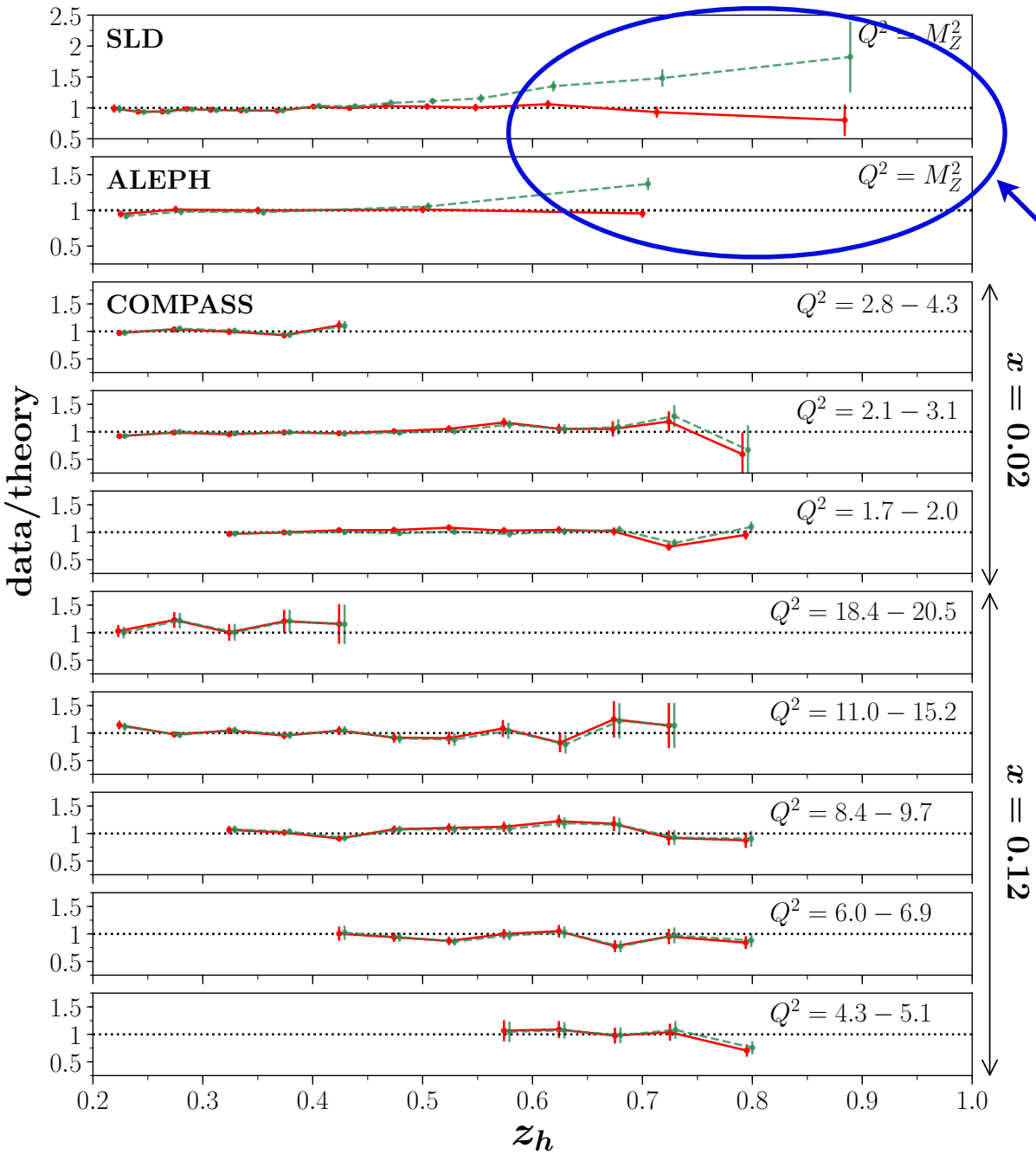
JAM 2019 analysis



solutions with large $s(x)$

fully constrained solutions

JAM 2019 analysis

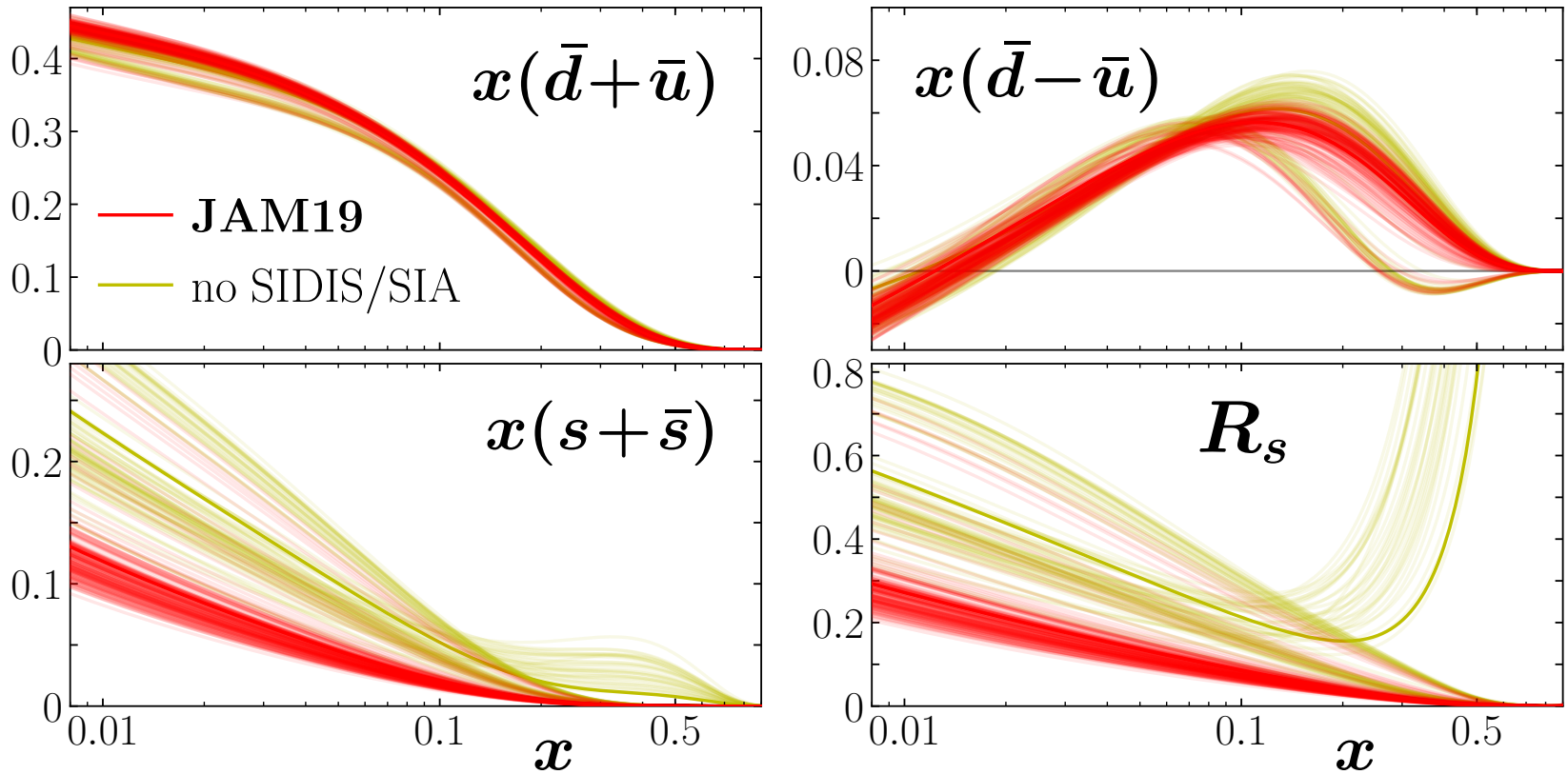


- - - - - solutions with large $s(x)$

——— fully constrained solutions

SIA data at large z strongly <u>disfavor</u> <u>small strange → K FF</u>

JAM 2019 analysis

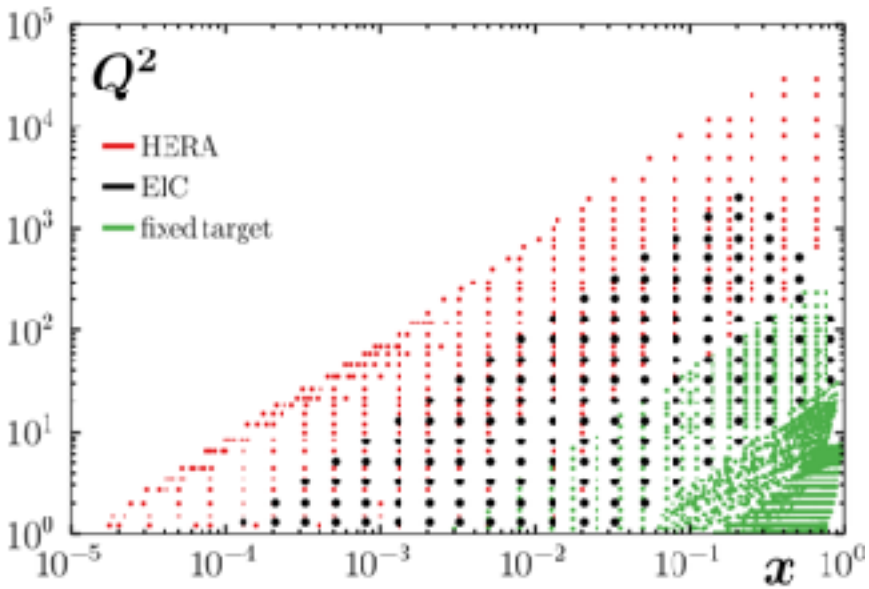


Sato, Andres, Ethier, WM (2019)

→ vital role played by SIDIS + SIA data in constraining strange PDF

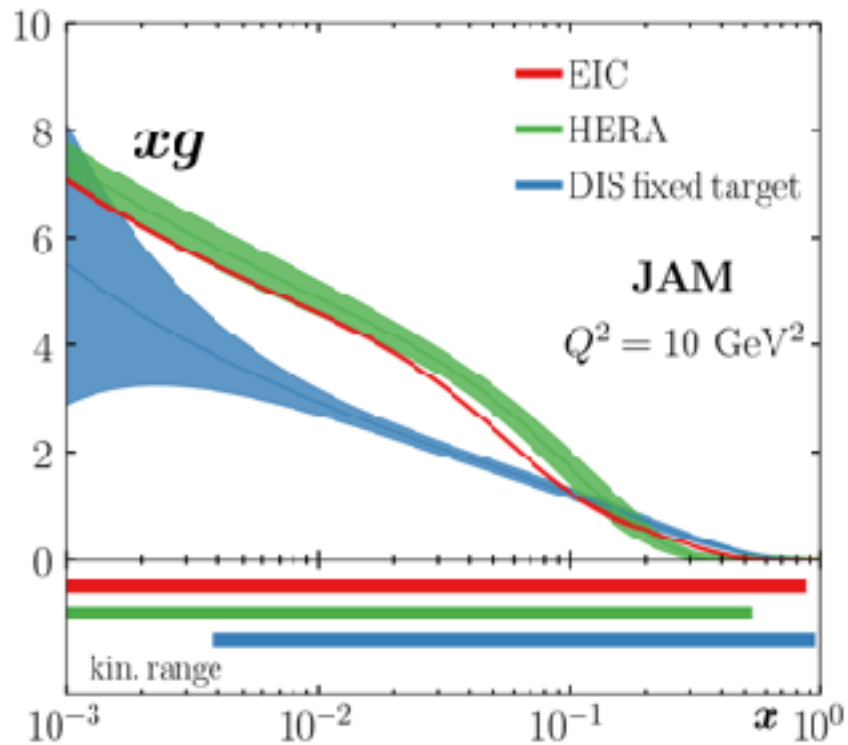
JAM – EIC analysis

- Current interest in impact of future EIC data on PDFs and their uncertainties (EIC Yellow Report, early 2021)



→ can we understand tension in gluon PDF between NC fixed-target & HERA DIS data?

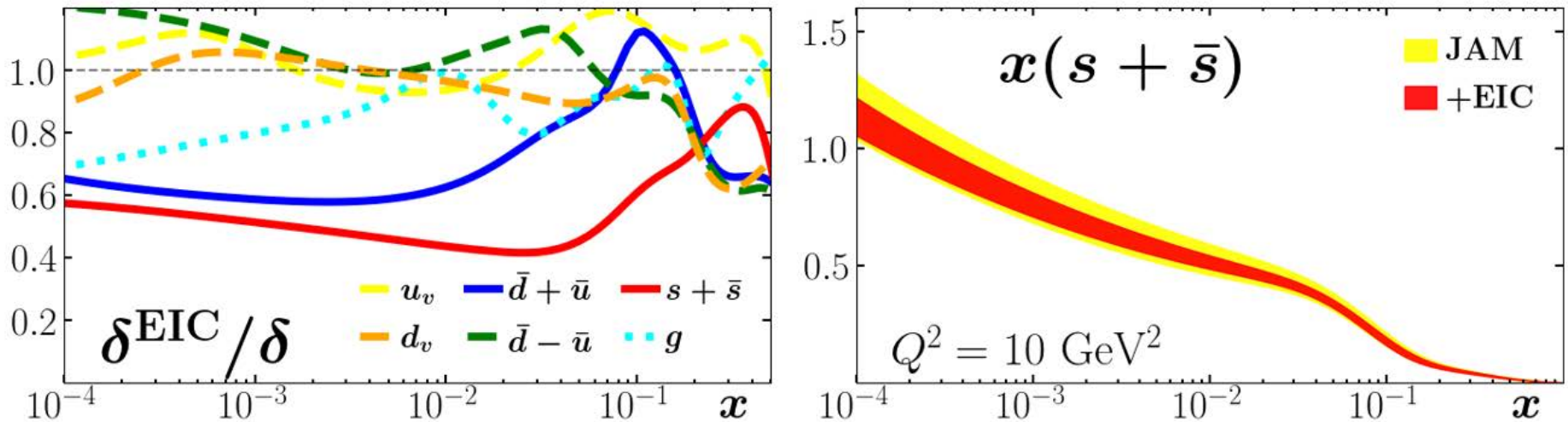
→ what is the impact of high-precision EIC data for various observables on specific PDFs?



Delcarro et al. (2020)

JAM – EIC analysis

- Current interest in impact of future EIC data on PDFs and their uncertainties (EIC Yellow Report, early 2021)



Cocuzza et al. (2020)

→ impact of parity-violating DIS data (γZ interference)

$$F_2^{\gamma Z,p} = \left(\frac{1}{3} - \frac{8}{9}\sin^2\theta_W\right) x(u+\bar{u}) + \left(\frac{1}{6} - \frac{2}{9}\sin^2\theta_W\right)(d+\bar{d}+s+\bar{s}) + \cdots$$

$$\approx \frac{1}{9}x(u+\bar{u}+d+\bar{d}+s+\bar{s}) + \cdots \quad \text{for } \sin^2\theta_W \approx 1/4$$

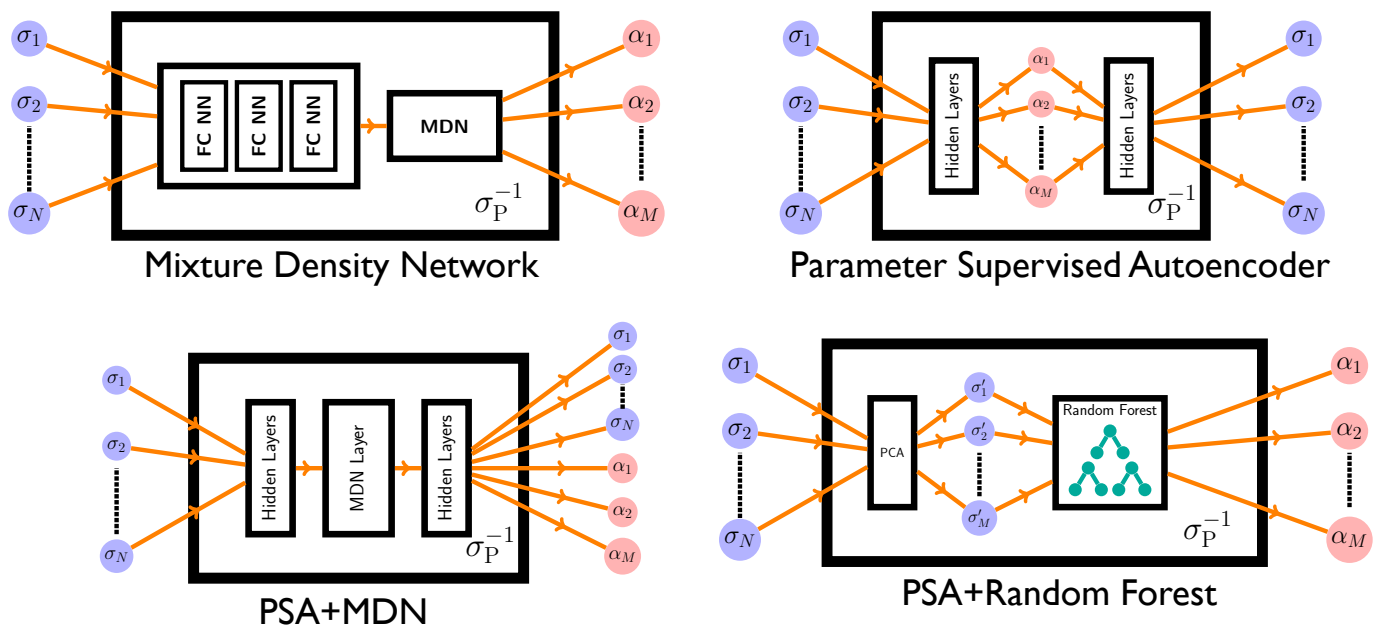
→ enhanced sensitivity to strange quark PDF

New AI tools

- Shortcomings of existing global QCD analysis paradigms
  - — what is impact of specific data sets (current & future) on particular QCFs?
  - — would like to be able to answers these questions "on the fly"

- Need "inverse mapping" from observables ⟶ parameter space

inverse mapper prototypes



Manal Almaeen, Yaohang Li, Eleni Tsitinidi, Michelle Kuchera, et al. (2020)

New AI tools

- Collaboration between nuclear theorists and computer scientists

  - M. Almaeen (ODU)
  - Y. Awadh Alanazi (ODU)
  - M. Houck (Davidson College)
  - M. P. Kuchera (Davidson College)
  - Y. Li (ODU)
  - W. Melnitchouk (JLab)
  - R. Ramanujan (Davidson College)
  - N. Sato (JLab)
  - E. Tsitinidi (Davidson College)

ODU

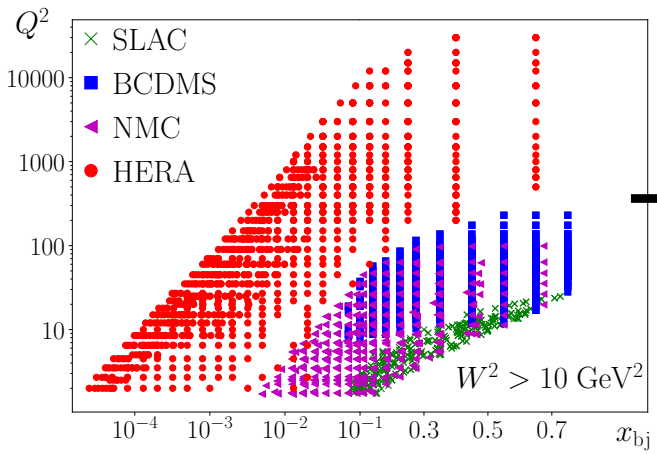
DAVIDSON

Jefferson Lab

CENTER for NUCLEAR FEMTOGRAPHY

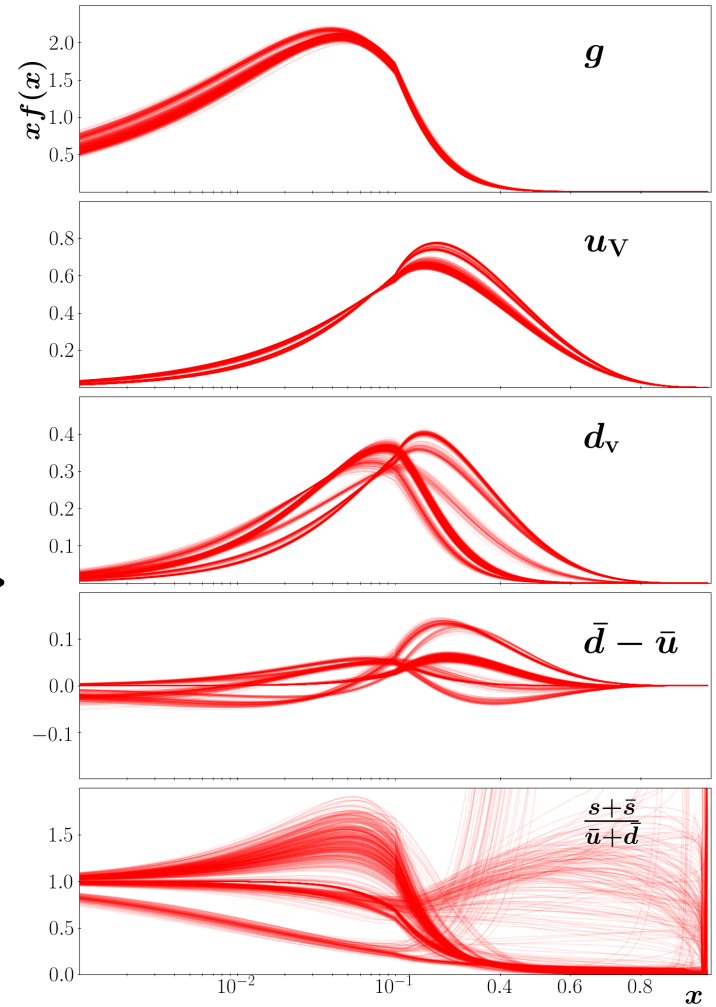
New AI tools

- Inverse mapper prototypes tested on toy models
- First application to real global data analysis — inclusive DIS



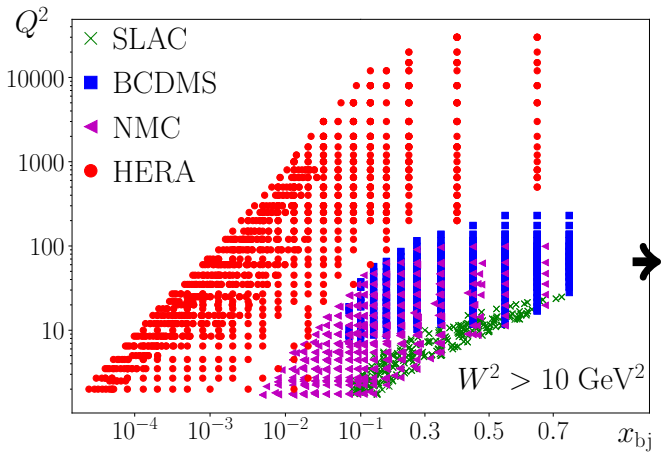
→

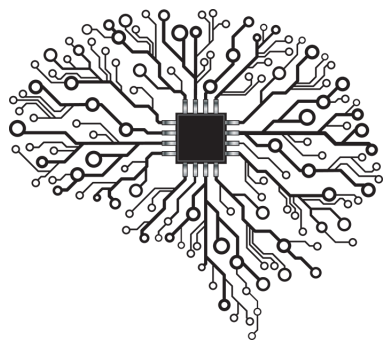
→

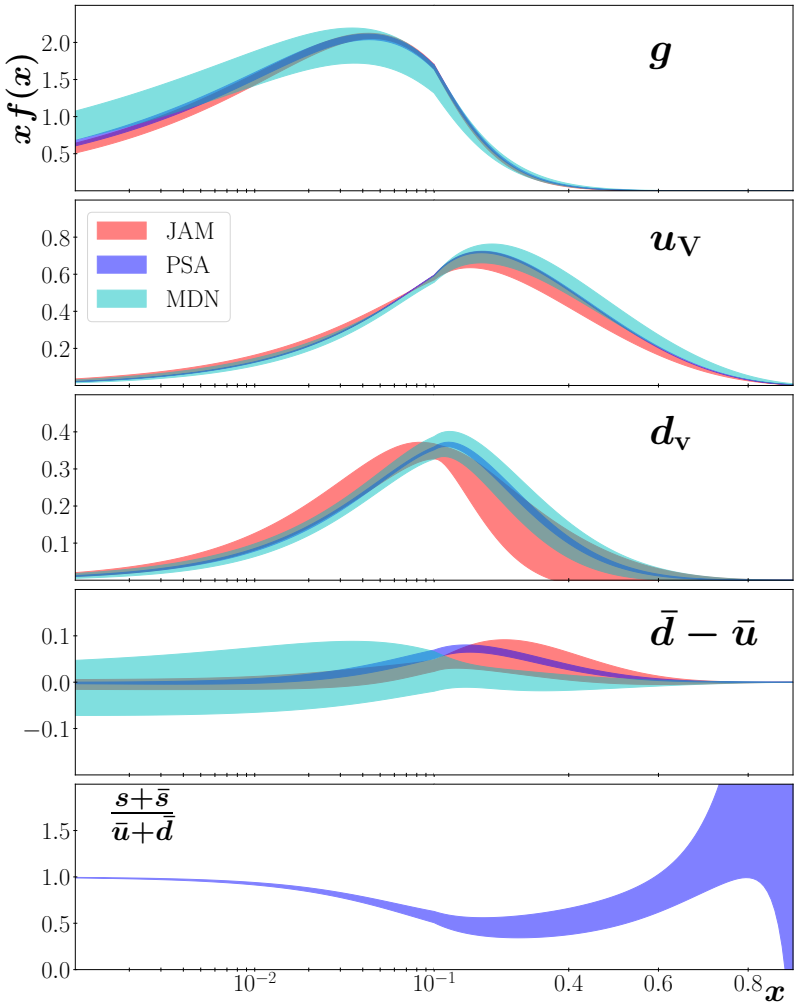


New AI tools

- Inverse mapper prototypes tested on toy models
- First application to real global data analysis — inclusive DIS







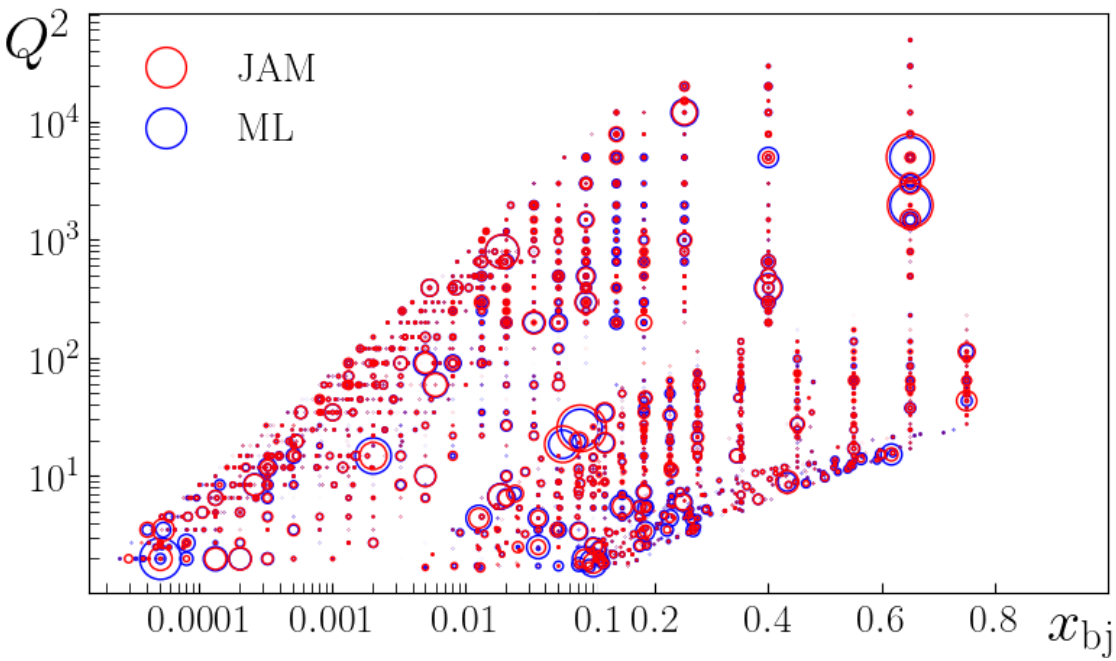
→ how to make inverse mapper even closer to JAM (with multiple solutions)?

→ explore different ML architectures *e.g.* variational auto-encoders

Manal Almaeen, Yaohang Li, Eleni Tsitinidi, Michelle Kuchera

New AI tools

- Inverse mapper prototypes tested on toy models
- First application to real global data analysis — inclusive DIS

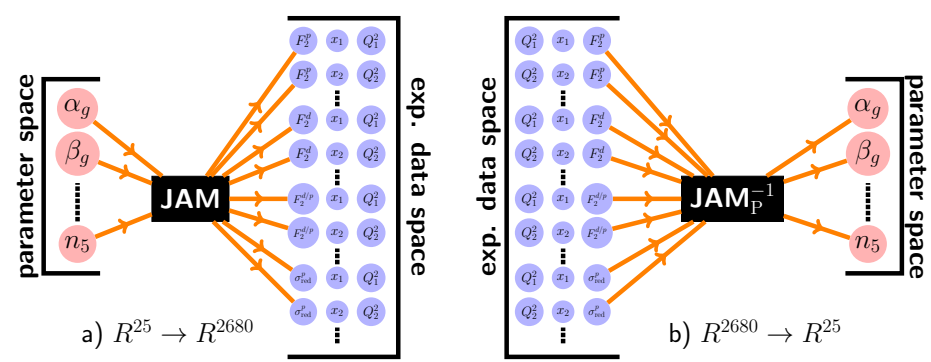


size of circles $\propto \chi^2$

$$\frac{\chi^2_{\rm JAM}}{N_{\rm dat}} = 1.25$$

$$\frac{\chi^2_{\rm ML}}{N_{\rm dat}} = 1.36$$

→ demonstration that it is indeed possible to build an inverse mapper using ML!



Open questions & challenges

- Better definitions for likelihood, $e.g.$ non-Gaussian distributions, for incompatible data sets?

- How to efficiently perform MC sampling in high dimensions
  (~ 25 for unpolarized DIS, ~ 100 for all collinear PDF/FFs, ~ 200 for TMDs)

  $\longrightarrow$ data resampling, nested sampling, HMC MC, Gaussian optimization, …

- In the near future will have more data, more parameters

  $\longrightarrow$ need for more computer resources
  (χ^2 per point in parameter space takes ~ 1 minute)

Thank you!

Co-organizers:

Martha Constantinou (Temple University)

Christopher Monahan (William & Mary/Jefferson Lab)

Alexander Rothkopf (Stavanger University)

AMHERST CENTER FOR FUNDAMENTAL INTERACTIONS

Physics at the interface: Energy, Intensity, and Cosmic frontiers

University of Massachusetts Amherst

UMASS AMHERST

U.S. DEPARTMENT OF ENERGY | Office of Science

NSF