Deep learning based Jet substructure at H1

Phys.Lett. B, 844, 138101 (2023)

Radek Žlebčík
on behalf of the H1 Collaboration

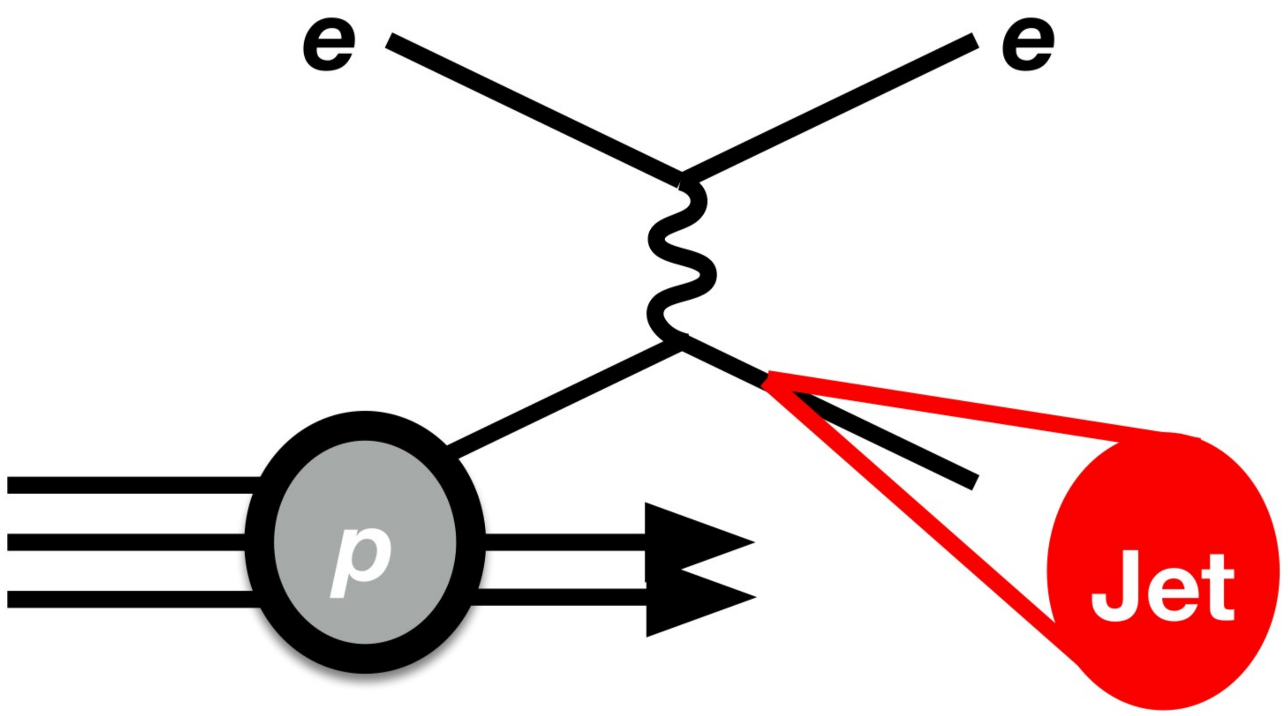
August 22, 2023
ISMD2023
Gyöngyös, Hungary

CHARLES
UNIVERSITY

Egerszalok's stunning salt terraces

H1 @ HERA

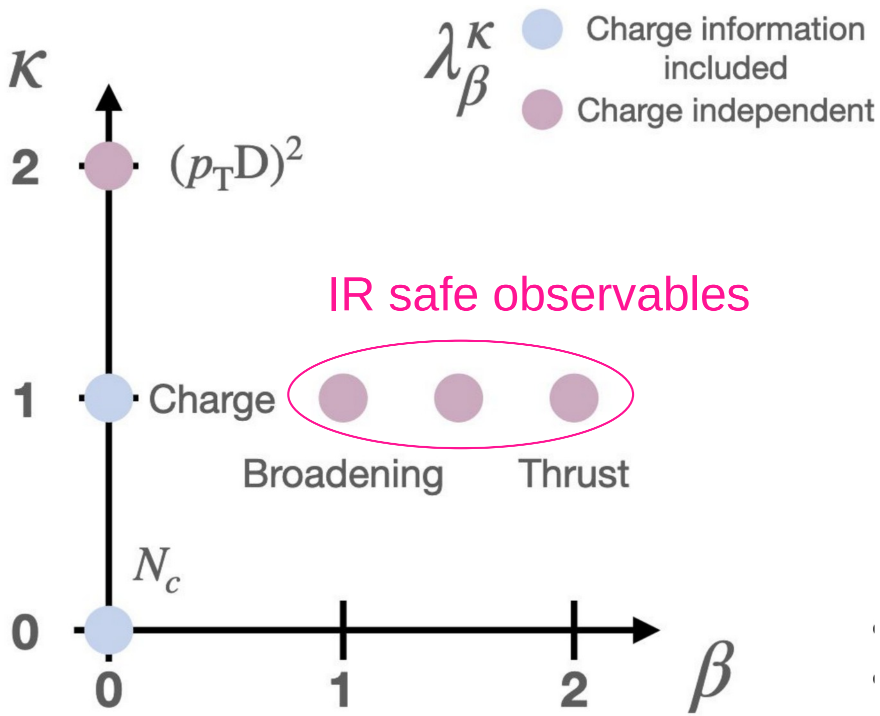
- Deep inelastic scattering: e p → e + Jet + X
  → unique environment without Underlying Event
- Crossroad between LEP (e^+e^-) and LHC (pp)
  → important input for example for MC tuning

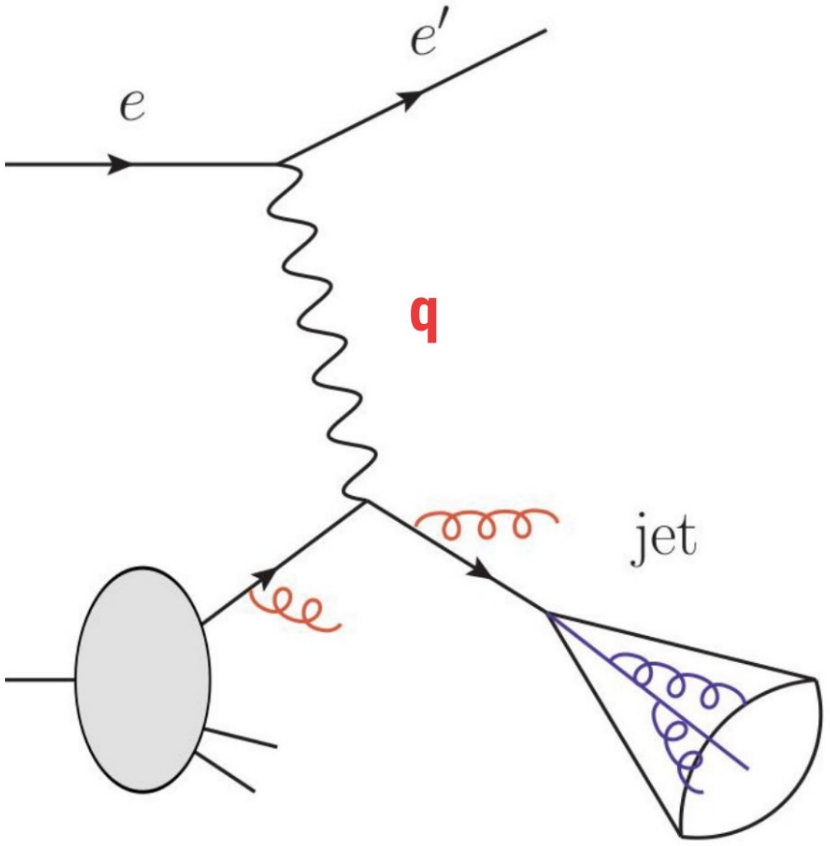


Jet angularities

- Jet constituents used to calculate several observables

$$\lambda_\beta^\kappa = \sum_{i\in\text{jet}} z_i^\kappa \left(\frac{R_i}{R_0}\right)^\beta \qquad z_i = p_{\mathrm{T},i}/p_{\mathrm{T}}^{\mathrm{jet}}$$

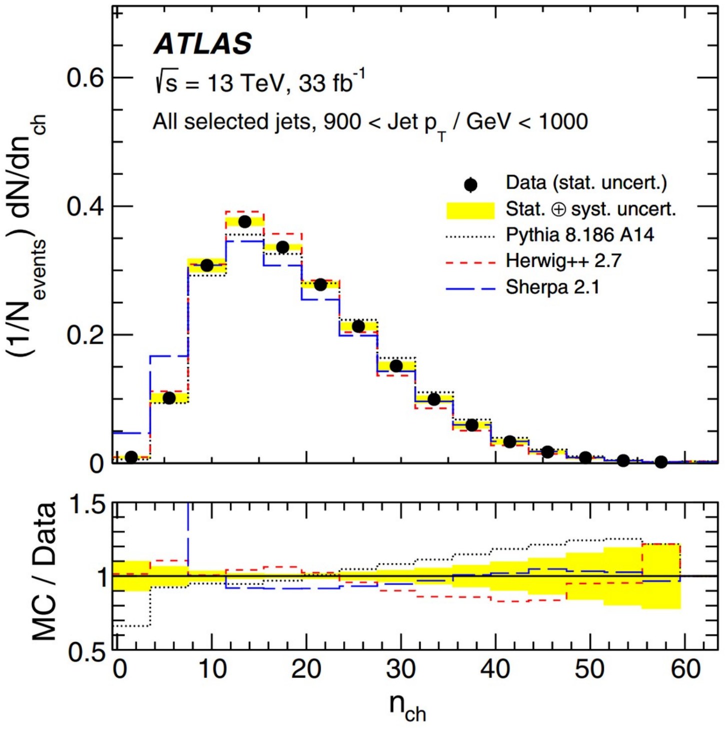


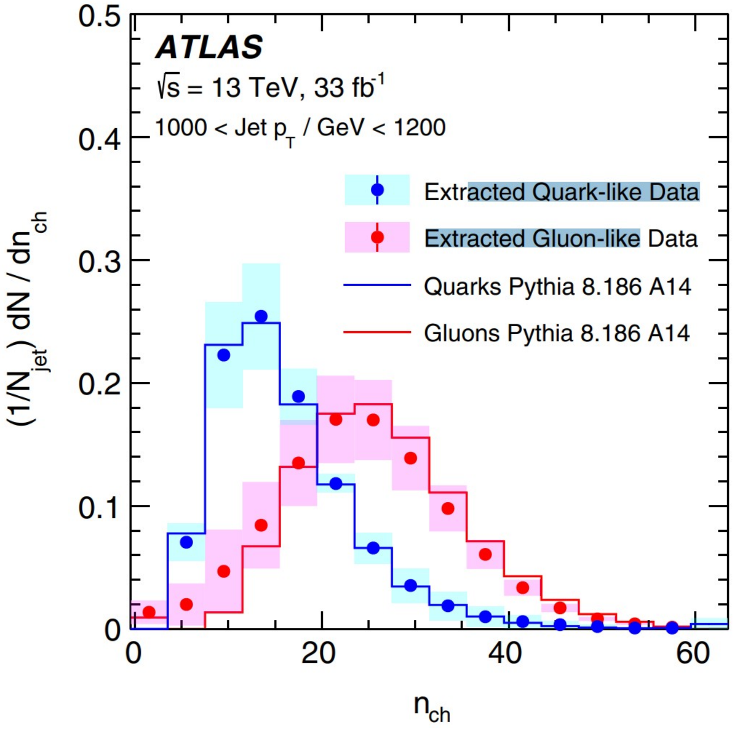


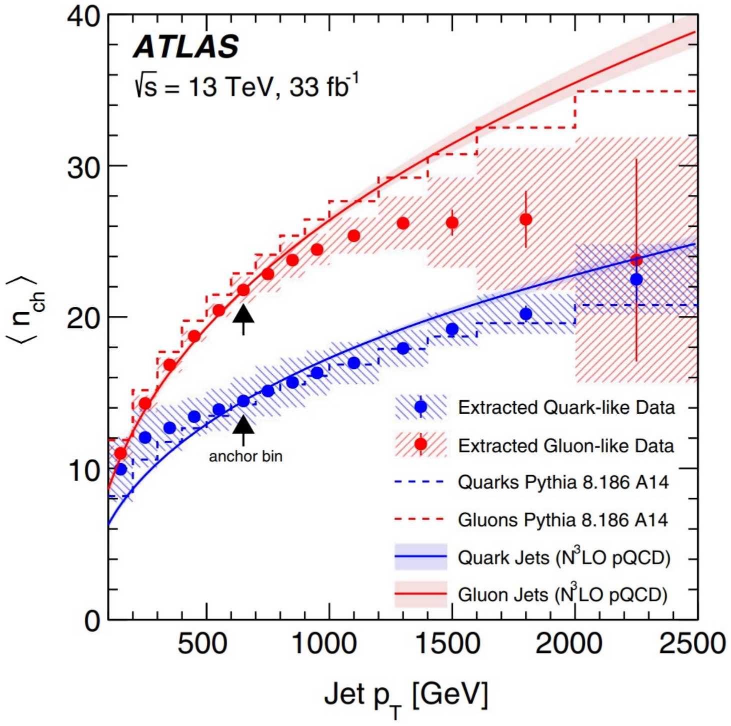
Evolution studied as a function of $Q^2 = -q^2$

Why?

- Charged particle multiplicity is sensitive to quark/gluon tagging





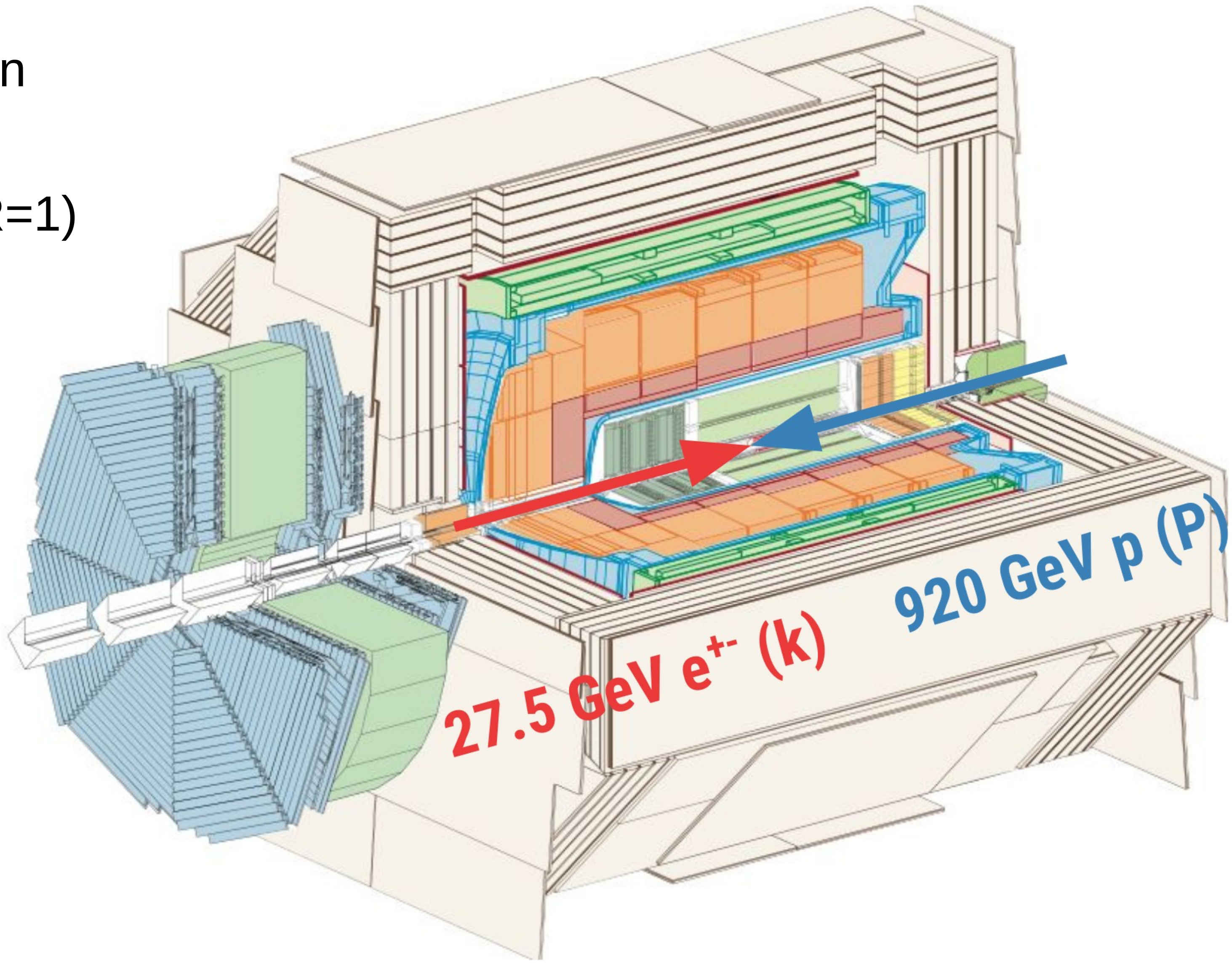


ATLAS, Phys. Rev. D 100, 052011 (2019)

HERA – the world only e$^{\pm}$p collider

Experimental setup

- 228 pb^{-1} of data collected by H1 in 2006-2007 at $\sqrt{s}$ = 319 GeV
- Phase space definition (k_T jets, R=1)
  $0.2 < y < 0.7$, $Q^2 > 150$ GeV^2
  Jet $p_T > 10$ GeV, $-1 < \eta_{lab} < 2.5$

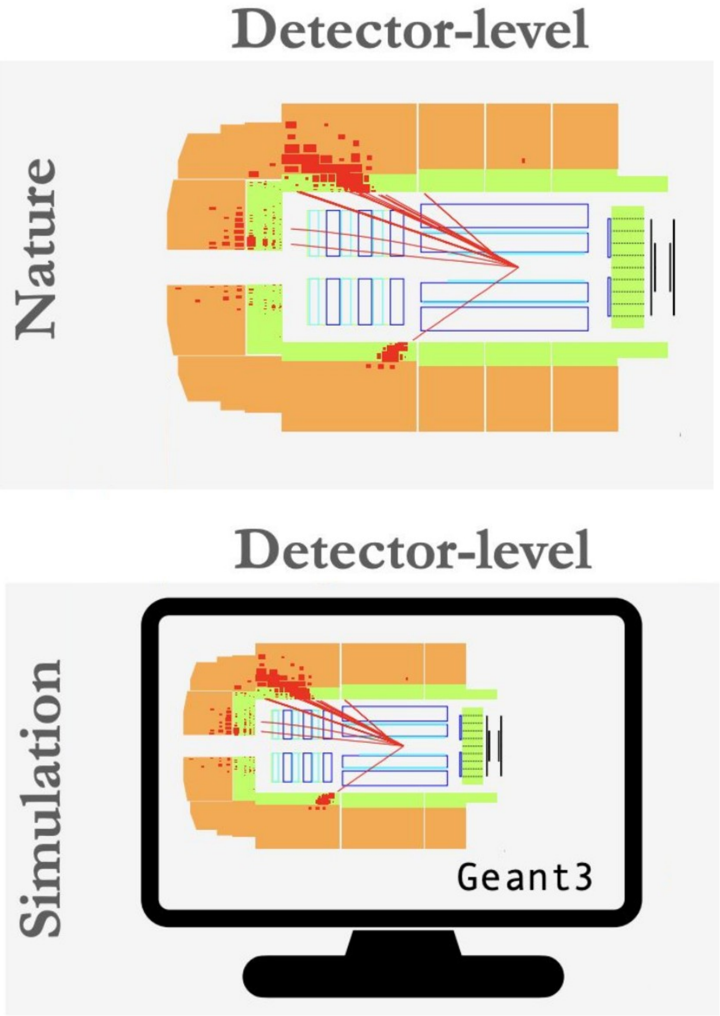


$$Q^2 = -q^2$$
$$y = Pq / pk$$

P: incoming proton 4-vector
k: incoming electron 4-vector
q=k-k' : 4-momentum transfer

Omnifold: Step 1 - Detector level reweighting

- A **Binary Classifier** is trained to distinguished **Data** from **Simulation**



$$t \to d$$

OmniFold: A Method to Simultaneously Unfold All Observables, Phys.Rev.Lett. 124 (2020) 18, 182001

Binary Classifier

$$\omega = \frac{p^{\mathrm{Data}}(d)}{p^{\mathrm{MC}}(d)}$$

MC reweighted with ω weights

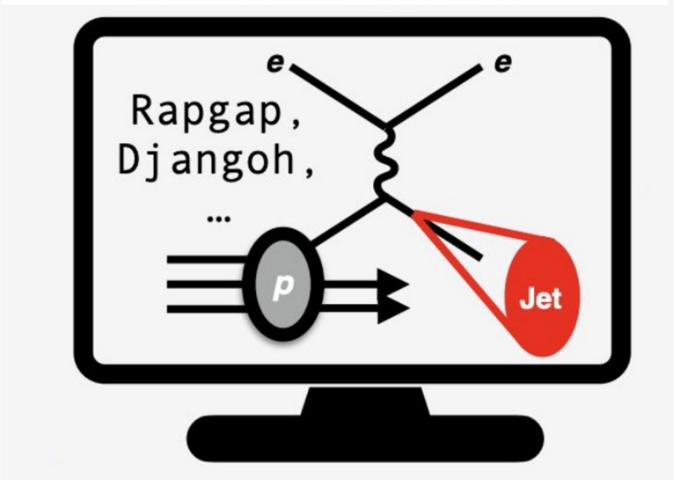
Omnifold: Step 2 - Particle level reweighting

- A **Binary Classifier** is trained to distinguished **Reweighted Simulation** from **Simulation**

Particle-level

MC reweighted
with ω weights

Particle-level



Binary Classifier

$$\nu = \frac{p^{\mathrm{MC}'}(t)}{p^{\mathrm{MC}}(t)}$$

MC reweighted
with ν weights

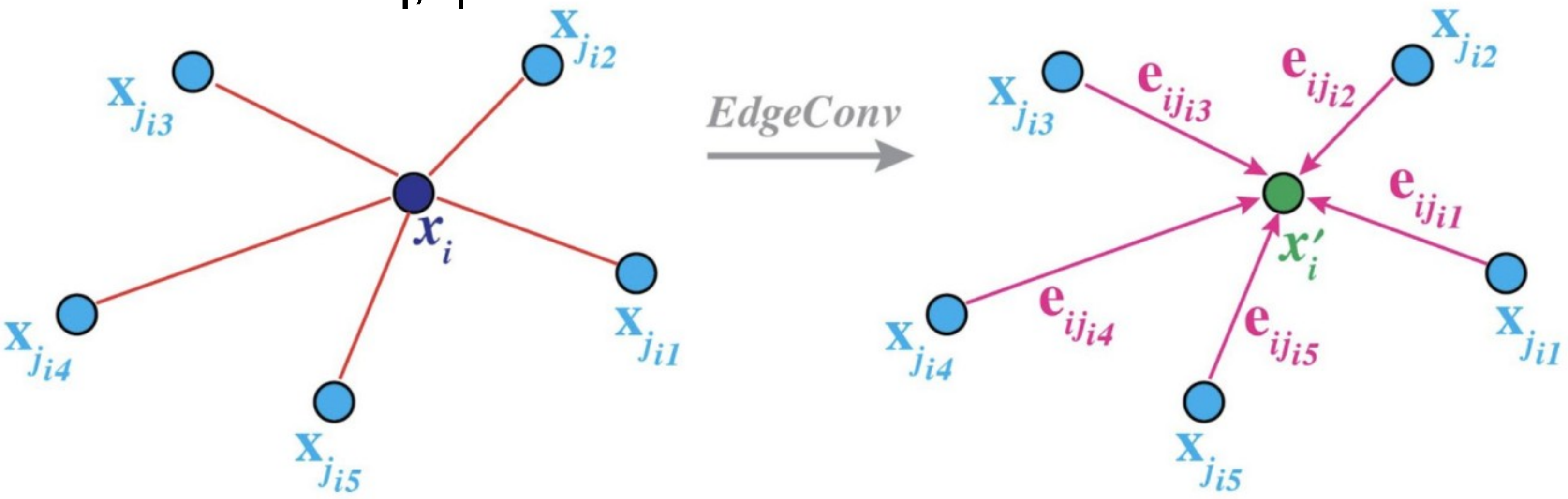
Omnifold: Detector level Classifier

- All jet particles used as input to the NN Model
  → (p_T, η, φ, charge) for each particle + (p_T^{jet}, η^{jet}, φ^{jet}) + Q^2
  up to (30x4 + 3 + 1) features at detector level

- Point Cloud Transformer architecture
  → exploiting success of Transformers in LLM

V. Mikuni and F. Canelli 2021 Mach. Learn.: Sci. Technol. 2 035027

η, φ



Trained at 50M jets using 128 GPUs of Perlmutter supercomputer

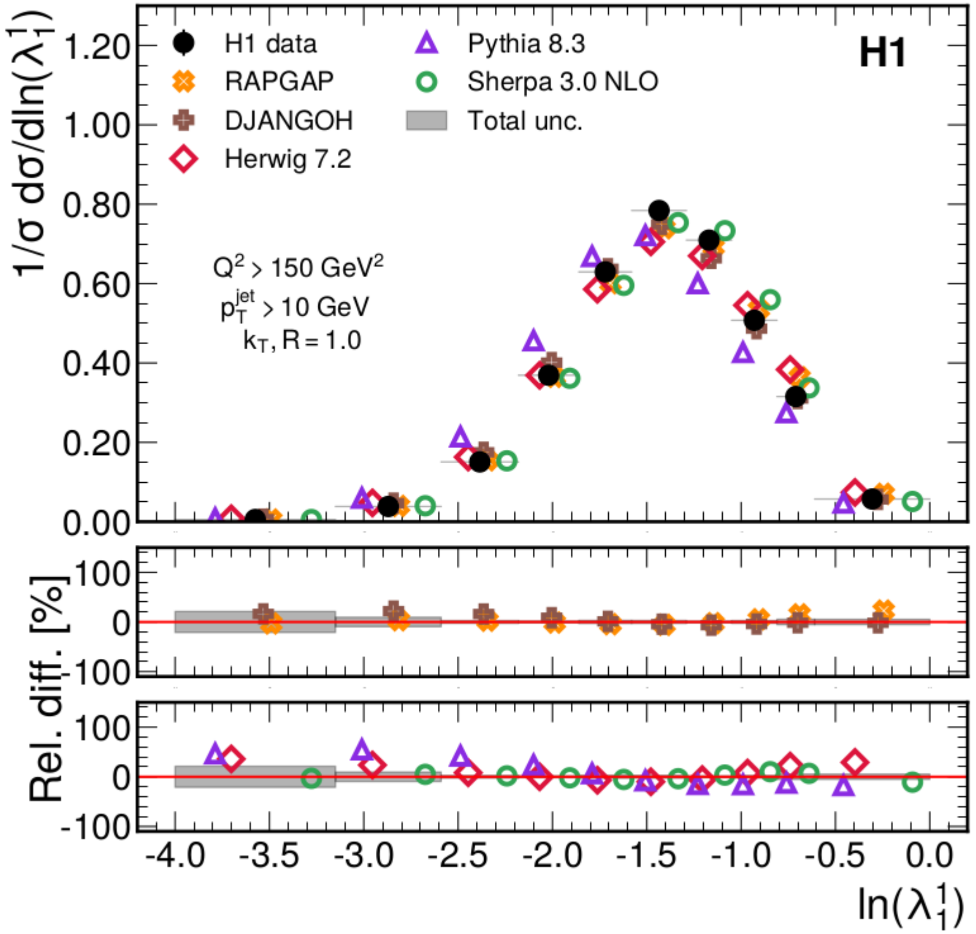
MC Models

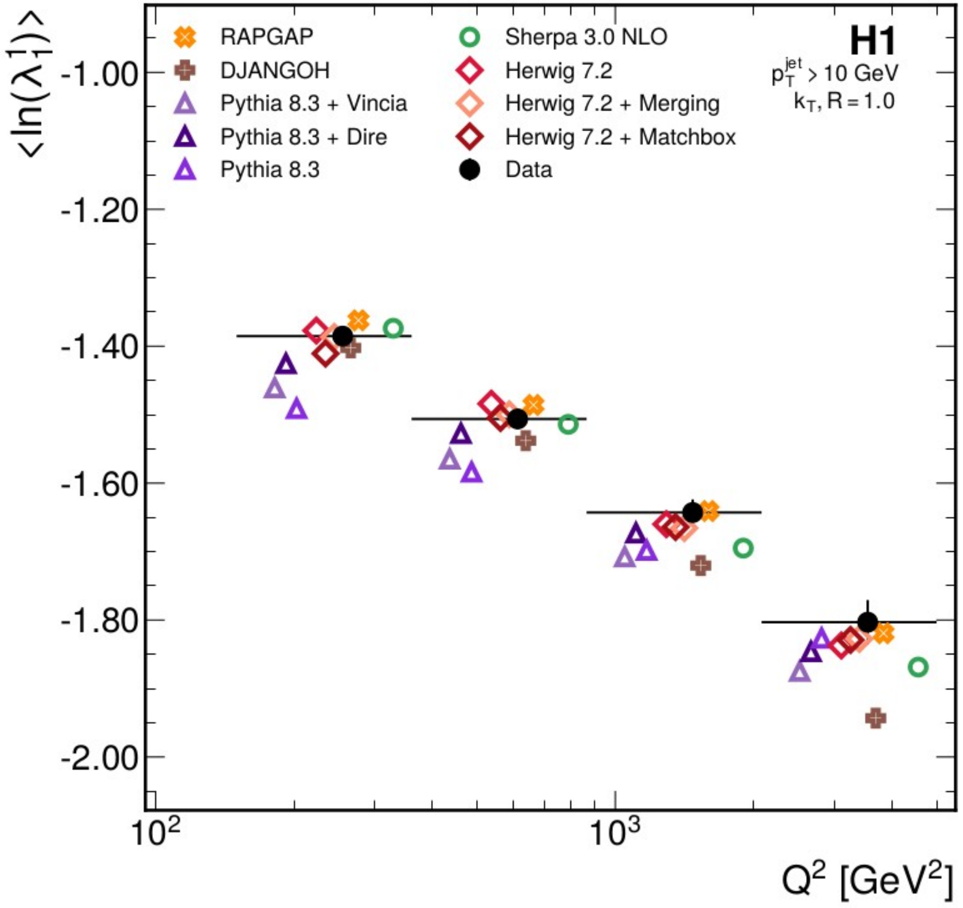
- **Legacy ep models** (Lund string hadr., CTEQ6L PDF set)
  → Rapgap: Py6-like shower
  → Djangoh: Dipole shower from Ariadne
- **Pythia 8.3**: default NNPDF3.1 PDF
  → Vincia: p_T ordered antenna showering
  → Dire: NLL dipole model, MMHT14 PDF
- **Herwig 7.2**: Shower with angular ordering,
  cluster hadronization, CT14 PDF
- **Sherpa 3.0**: NLO pQCD for 2 jets,
  cluster hadronization

Jet broadening

$$\lambda_1^1 = \sum_{i \in \text{jet}} z_i \frac{R_i}{R_0}$$

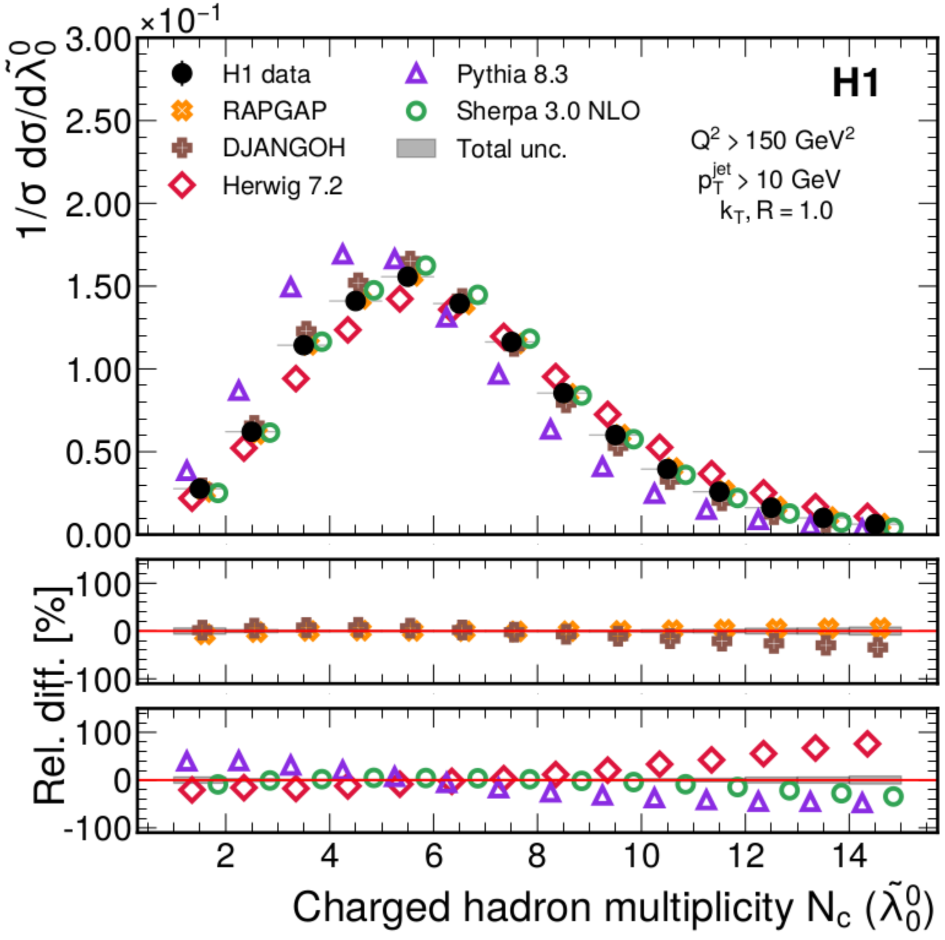
- **Herwig** describes best the Q^2 dependence
- **Pythia** underestimates the broadening

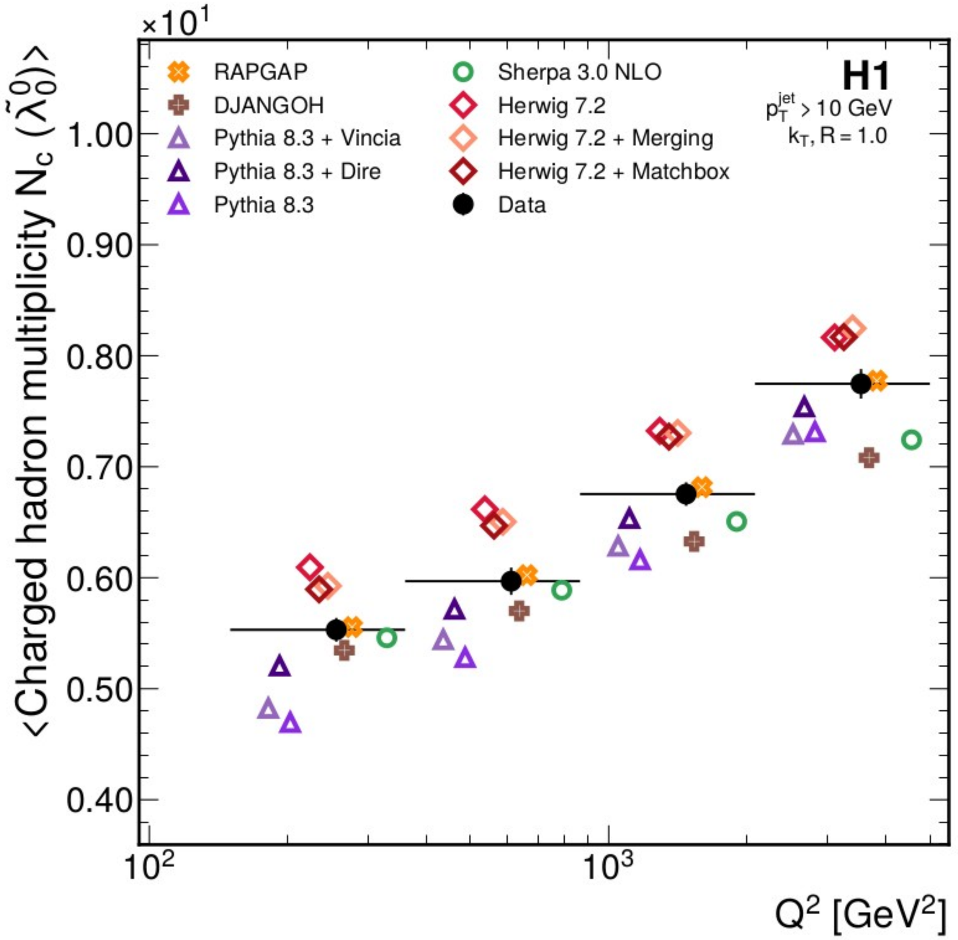




Charged hadron multiplicity in the jet

- **Sherpa** does good job in describing the shape
- **Pythia** predicts lower multiplicities

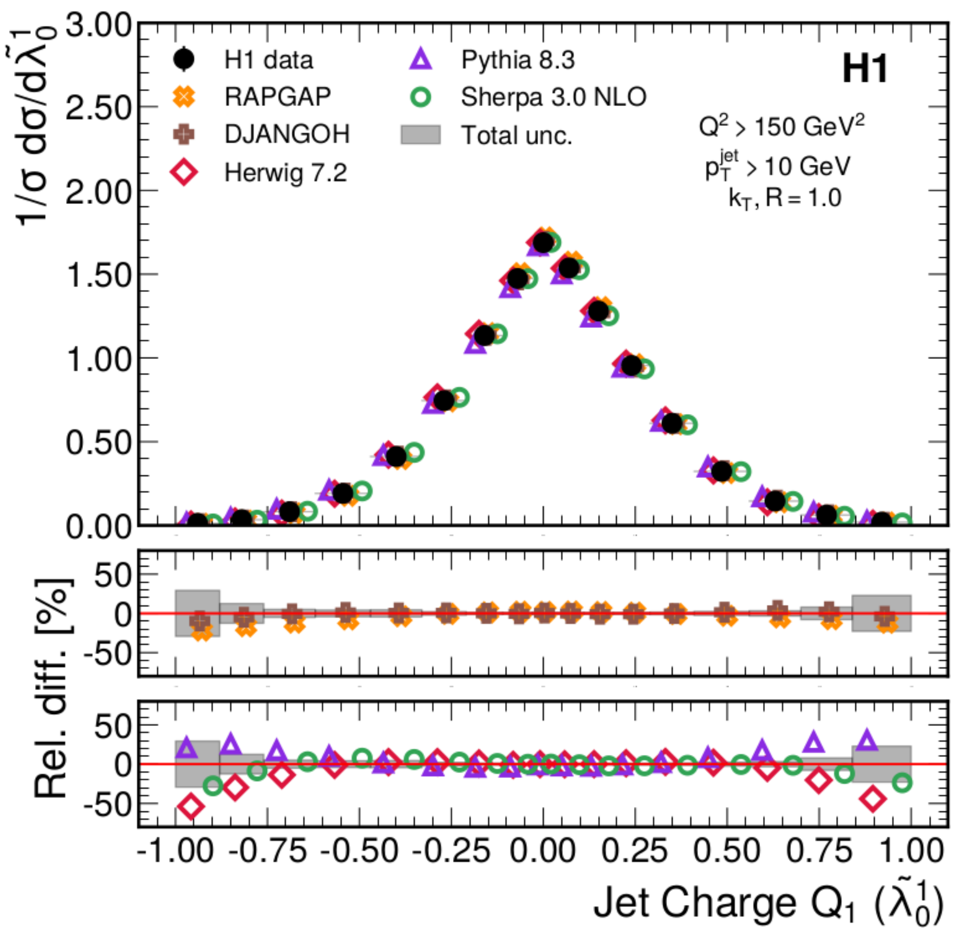


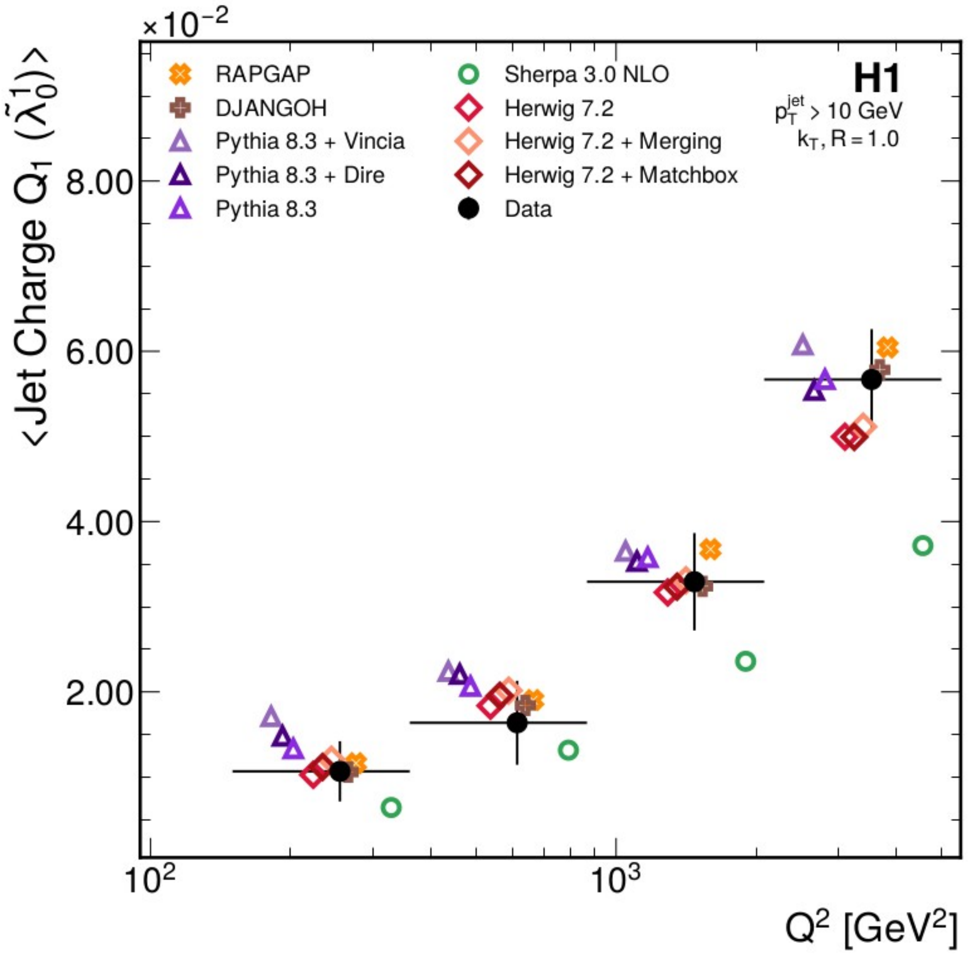


Jet charge

- **Sherpa** underestimates jet charge at higher Q2
- **Pythia** & **Herwig** do comparable job

High Q2 jets are more quark-like





Conclusions

- Revival of H1 data using state-of the art statistical methods
- Substructure of jets in ep can help to disentangle underlying event contribution and better understand of the QCD dynamics

Backup

Systematic uncertainties currently considered

- **HFS energy scale:** +- 1%
- **HFS azimuthal angle:** +- 20 mrad
- **Lepton energy:** +- 0.5% (mainly affects Q^2)
- **Lepton azimuthal angle:** +- 1 mrad (mainly affects Q^2)
- **Model uncertainty:** differences in unfolded results between Djangoh and Rapgap
- **Non-closure uncertainty:** Differences between the expected and obtained values of the closure test
- **QED uncertainty**: Use the variation of measured quantities when radiation is turned off in the simulation
- **Statistical uncertainty:** Standard deviation of 100 bootstrap samples with replacement

Backup

- Example of results with uncertainties

Q^2 [GeV^2]	$<\tilde{\lambda}_0^0>$	Stat.	Tot.	HFS(jet)	HFS(other)	HFS(ϕ)	Lepton(E)	Lepton(ϕ)	Model	Closure
[150.00, 360.42]	5.531	0.019	0.019	0.010	0.020	0.047	0.006	0.041	0.012	0.081
[360.42, 866.03]	5.969	0.024	0.021	0.021	0.023	0.020	0.020	0.036	0.070	0.081
[866.03, 2080.90]	6.753	0.035	0.016	0.012	0.028	0.032	0.048	0.006	0.037	0.066
[2080.90, 5000.00]	7.747	0.056	0.017	0.053	0.056	0.050	0.050	0.028	0.019	0.050
Q^2 [GeV^2]	$<\tilde{\lambda}_0^1>$	Stat.	Tot.	HFS(jet)	HFS(other)	HFS(ϕ)	Lepton(E)	Lepton(ϕ)	Model	Closure
[150.00, 360.42]	0.011	0.001	0.332	0.001	0.001	0.001	0.001	0.001	0.001	0.001
[360.42, 866.03]	0.016	0.001	0.302	0.001	0.001	0.002	0.003	0.001	0.001	0.001
[866.03, 2080.90]	0.033	0.002	0.174	0.001	0.001	0.003	0.004	0.001	0.001	0.001
[2080.90, 5000.00]	0.057	0.003	0.104	0.002	0.003	0.002	0.001	0.001	0.002	0.002
Q^2 [GeV^2]	$<\sqrt{\lambda_0^2}>$	Stat.	Tot.	HFS(jet)	HFS(other)	HFS(ϕ)	Lepton(E)	Lepton(ϕ)	Model	Closure
[150.00, 360.42]	0.449	0.001	0.009	0.001	0.001	0.002	0.001	0.002	0.001	0.001
[360.42, 866.03]	0.441	0.001	0.011	0.001	0.001	0.002	0.001	0.003	0.001	0.001
[866.03, 2080.90]	0.429	0.001	0.009	0.001	0.001	0.001	0.001	0.003	0.001	0.001
[2080.90, 5000.00]	0.419	0.002	0.010	0.002	0.002	0.002	0.001	0.001	0.002	0.001
Q^2 [GeV^2]	$<\ln(\lambda_1^1)>$	Stat.	Tot.	HFS(jet)	HFS(other)	HFS(ϕ)	Lepton(E)	Lepton(ϕ)	Model	Closure
[150.00, 360.42]	-1.385	0.004	0.007	0.004	0.004	0.004	0.002	0.002	0.002	0.002
[360.42, 866.03]	-1.506	0.005	0.009	0.003	0.000	0.004	0.002	0.004	0.007	0.009
[866.03, 2080.90]	-1.643	0.006	0.012	0.006	0.005	0.000	0.014	0.005	0.004	0.005
[2080.90, 5000.00]	-1.803	0.010	0.018	0.008	0.009	0.010	0.020	0.008	0.009	0.014