Statistical inference of cosmic ray origins and propagation conditions

Igor Vaiman
Advisor: Prof. **Carmelo Evoli**

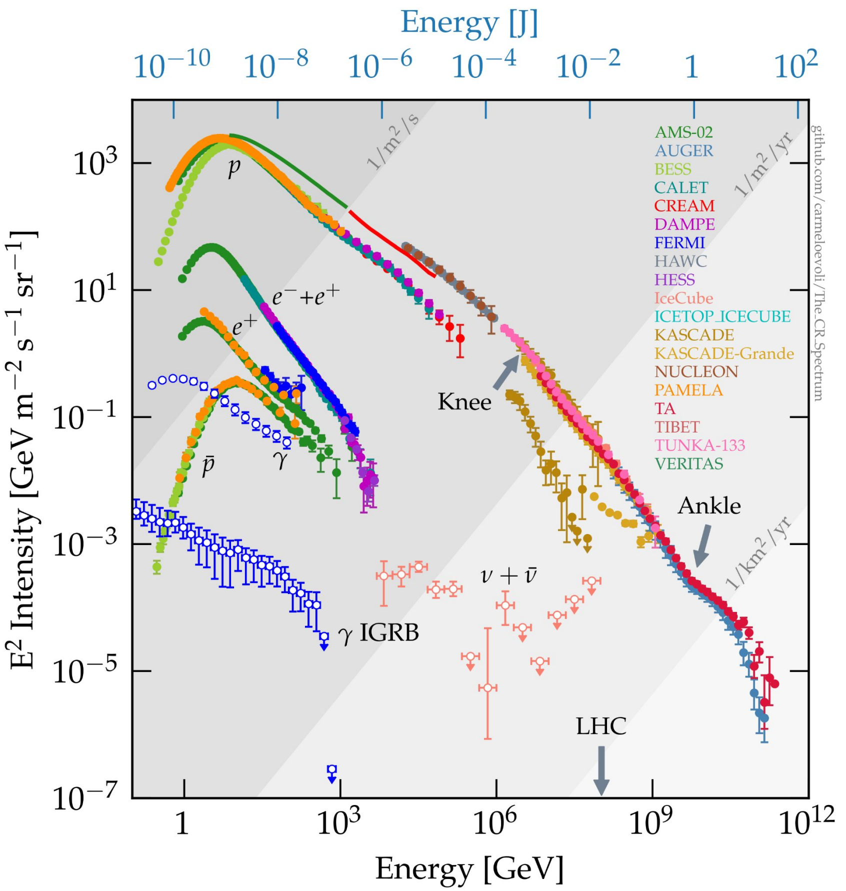
Currently @ **RWTH Aachen, Germany**

Prof. Philipp Mertsch



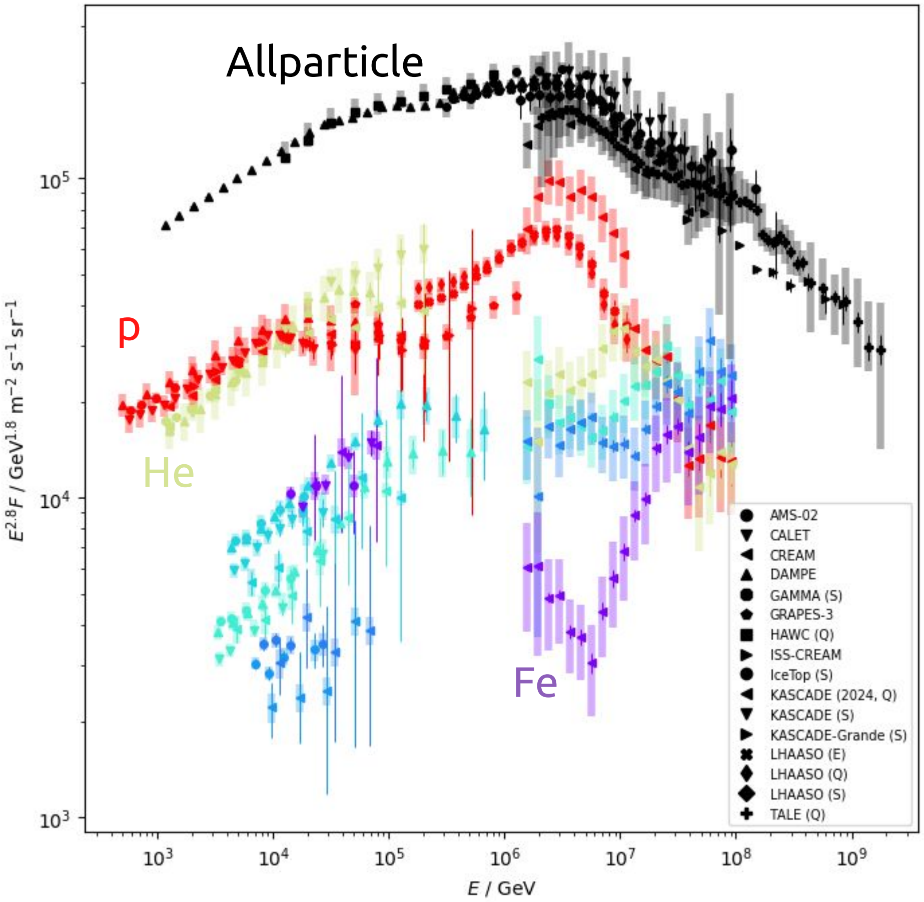

Context

- **Cosmic rays** are observed for 100+ years, yet their **sources** and **propagation** mechanisms are not fully understood
- Charged particles, they experience multiple types of losses and are deflected by magnetic fields → any inverse problem is highly degenerate
- Any CR analysis is necessarily and significantly **statistical** in nature



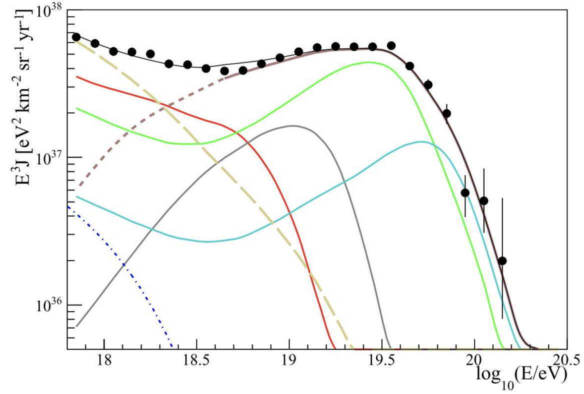
Context: the knee region

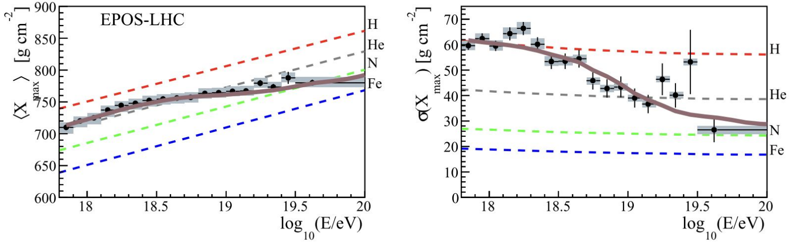
- **Supernova paradigm**: galactic CRs are accelerated at SN remnants
- Energetics are fine assuming few % energy conversion to CR
- Diffusive Shock Acceleration at SNR shocks
- Yet, reaching **PeV energies** with SN and explaining **complex spectral shapes** is difficult
- Other galactic source populations?



Context: Ultra-High Energy Cosmic Rays

- At higher energies, extragalactic sources are key
- Propagation through cosmological distances and poorly constrained environments → heavy simplifying assumptions & statistical reasoning
- CR mass composition is crucial, but hard to measure and entangled with open high energy physics problems
- Theoretical priors are very wide: one or multiple sources? nearby or far away? AGN, BN mergers, cosmic shocks? EGMF role?



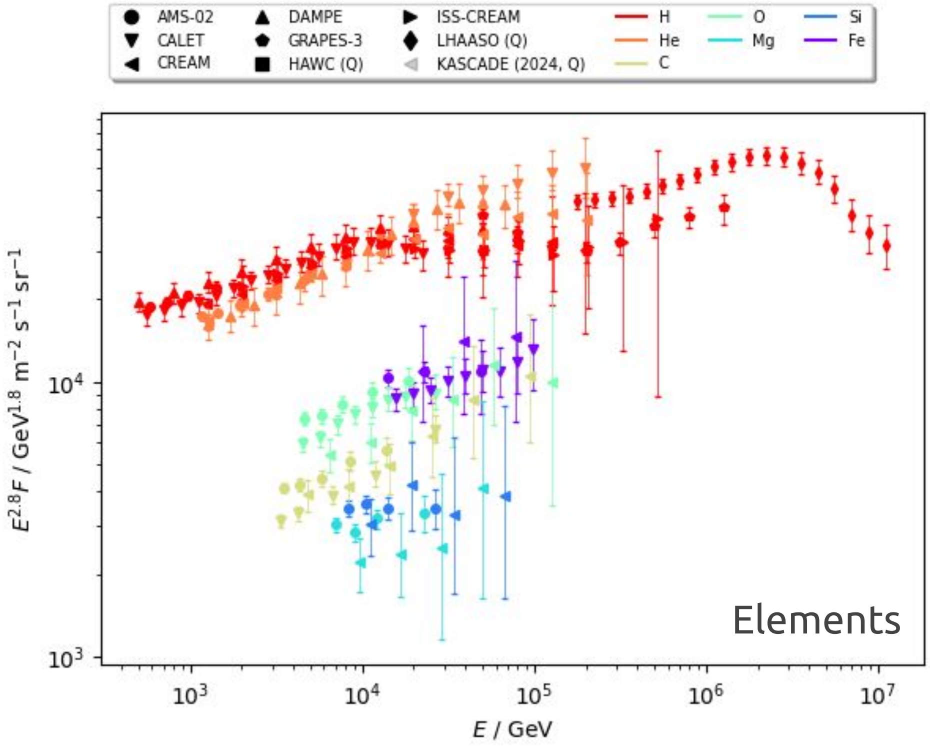


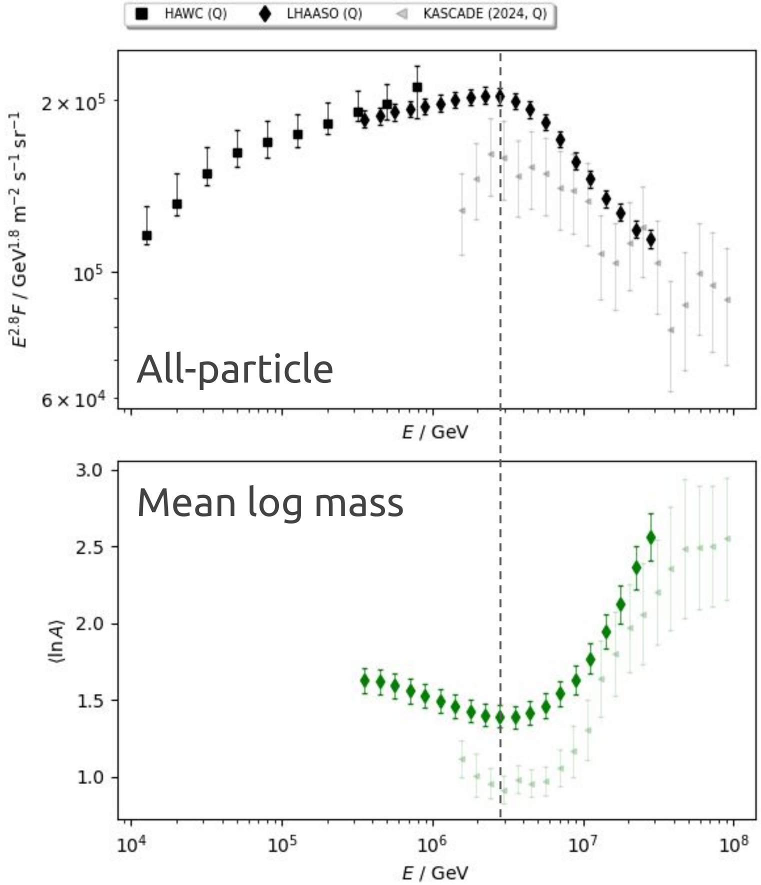
Aab, A. et al. Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory. J. Cosmol. Astropart. Phys. 2017, 038–038 (2017.

Project 1

Cosmic ray spectrum phenomenology at the knee

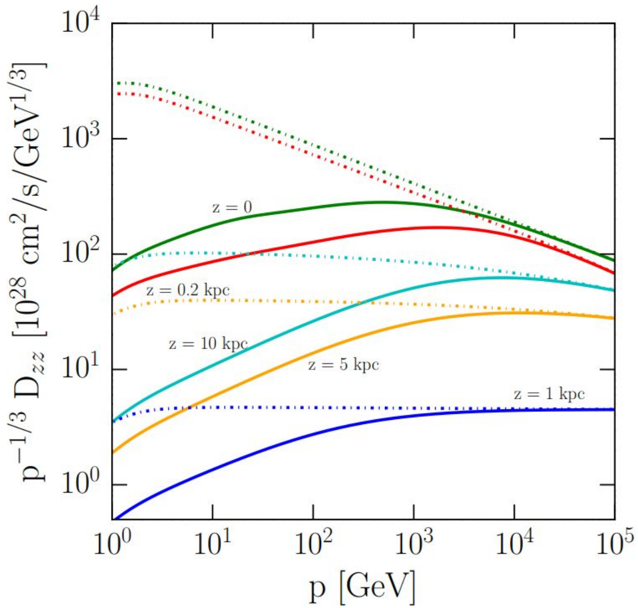
Latest measurements





Spectrum phenomenology

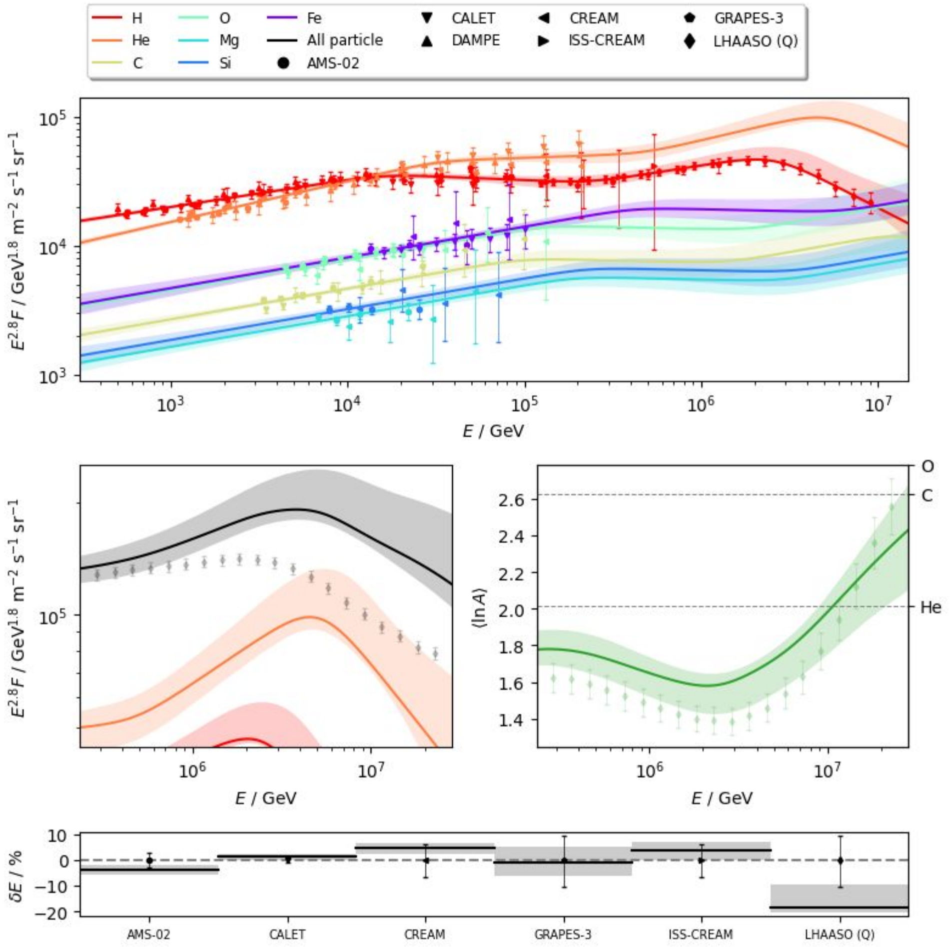
- At R ≳ 500 GV propagation is diffusion-dominated for all elements up to Fe, i.e. losses are not important → purely rigidity-dependent effects
- Phenomenologically, propagation effects likely manifest as **breaks in the R-spectrum**



Evoli, C., Aloisio, R. & Blasi, P. Galactic cosmic rays after the AMS-02 observations. Phys. Rev. D 99, (2019)

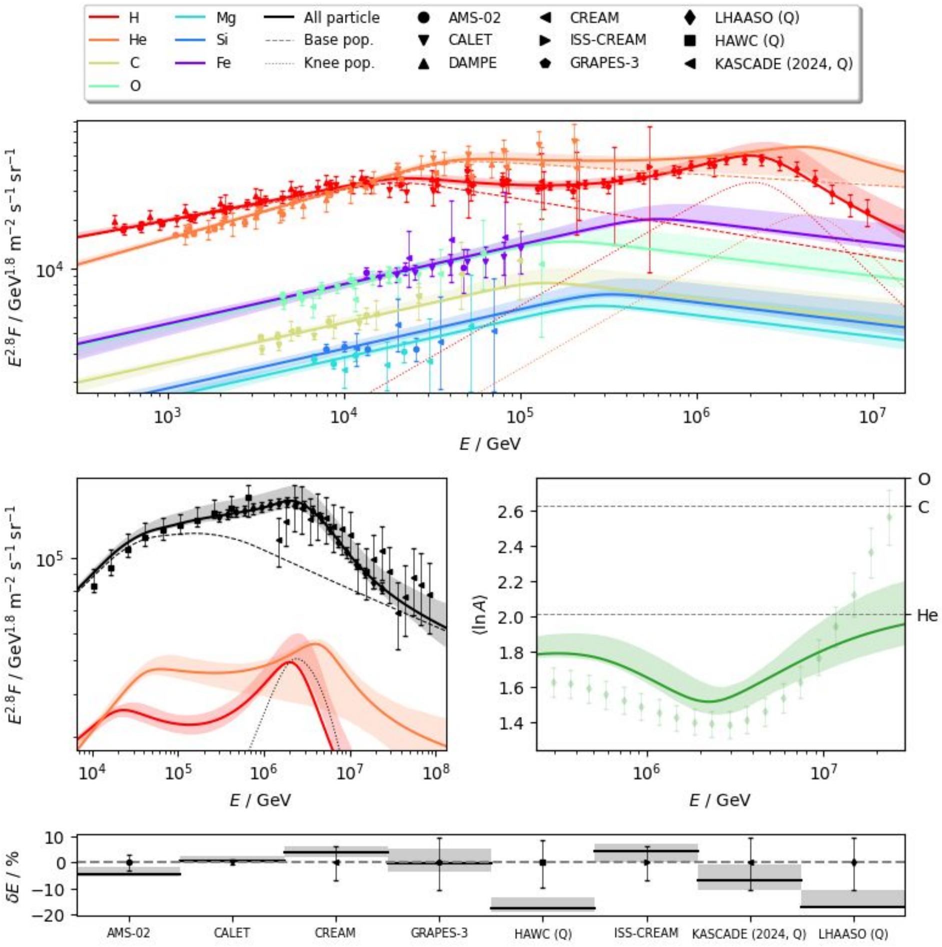
Findings: minimal model

- "Baseline" power law indices are significantly element-dependent – can be explained in non-standard DSA models or as population superposition
- Three spectral breaks are required to fit proton spectrum
- In the minimal single-population model hard helium spectrum extrapolated from direct measurements dominates and overshoots LHAASO all-particle



Findings: modified model

- Single-population model is unable to explain the data on the purely phenomenological level
- Introducing the second population contributing at the knee reduces tension
- Still, many open questions:
    - ⟨ln A⟩ and heavier elements are not well described
    - energy scale shifts at >2σ level are required
    - unresolved elements might play a role



Outlook

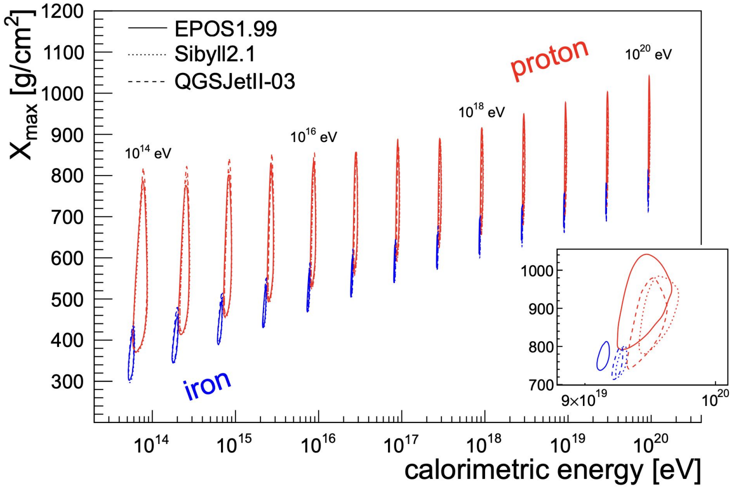
- Simple transfer of proton features to other elements in a rigidity-dependent way contradicts the data
- Multiple population must be present already at TeV-PeV range
- Work is ongoing to refine the model and include more recent observations
- More theoretical implications to investigate (local source, microquasar population, etc)
- The work has been reported at ICRC2025, publication is is preparation

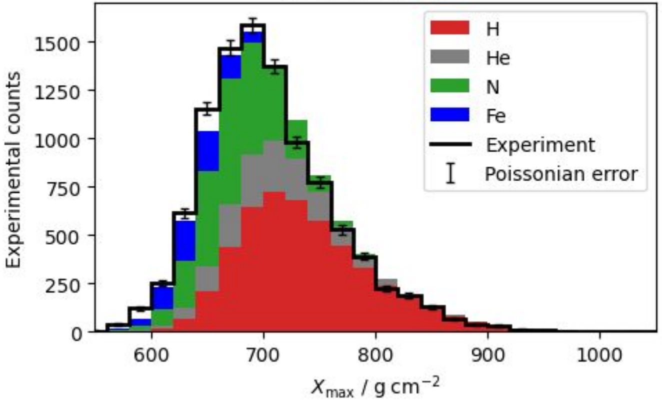
Project 2.1

UHECR mass composition at Pierre Auger Observatory

Context

- CR with energies $\gtrapprox$ PeV are observed indirectly through Extensive Air Showers
- Shower depth X_{max} measured directly by Fluorescence Detectors is sensitive to primary particle mass…
- … but EAS is intrinsically stochastic, so per-event mass determination is unfeasible!
- Composition is determined statistically for the whole sample of events by considering their X_{max} distribution

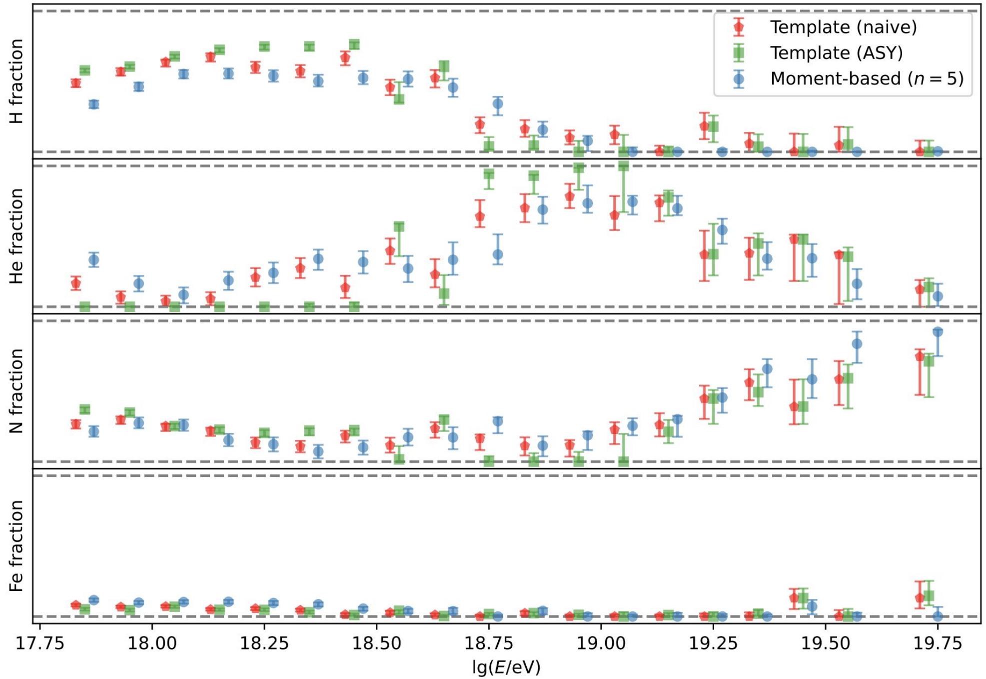




Kampert, K.-H. & Unger, M. Measurements of the cosmic ray composition with air shower experiments. Astroparticle Physics 35, 660–678 (2012)

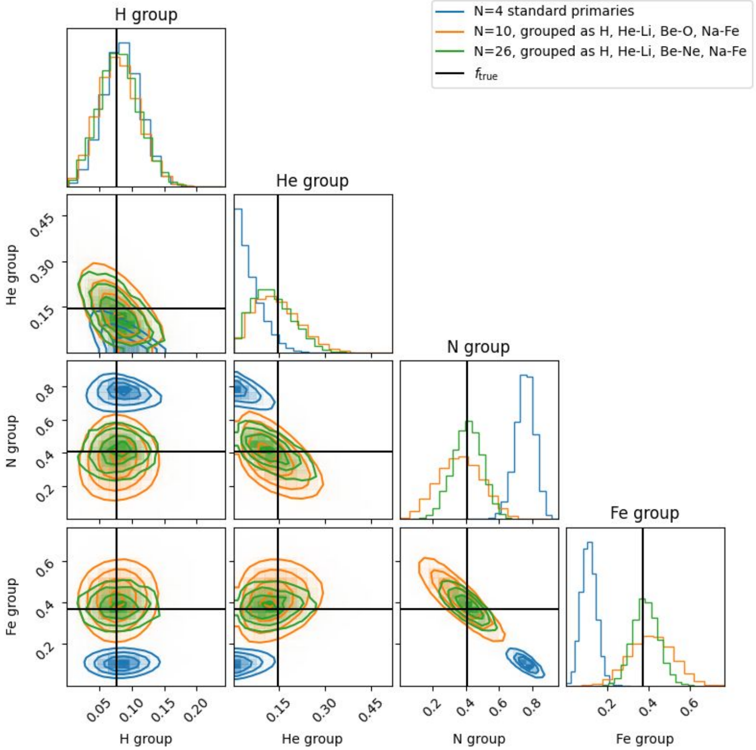
My contributions with F. Convenga & L'Aquila Auger group

- Independent cross-check of the standard result being prepared for Collaboration paper
- Extensive methodological comparisons to validate the result:
  - **Model input**: Monte-Carlo templates *vs* analytical approximation to the distribution
  - **Statistical framework**: Bayesian inference with MCMC *vs* frequentist Feldman-Cousins intervals
  - **Likelihood choice** for template analysis: naive Poissonian, fluctuations-aware, moments-based



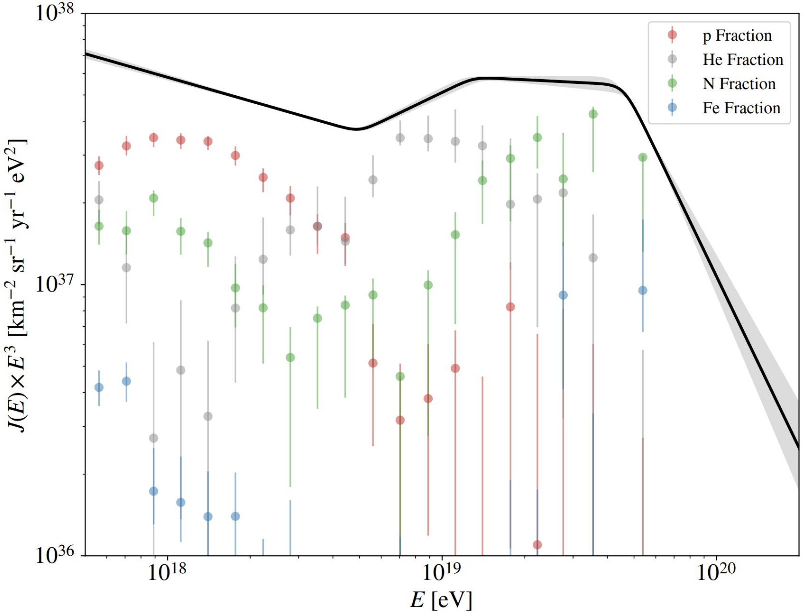
My contributions with F. Convenga & L'Aquila Auger group

- Number of primaries chosen as representatives can potentially impose bias on the fractions!



Outlook

- Our results so far have been reported in an internal Collaboration note
- Individual primary (e.g. proton) spectrum analysis to extract features
- Further study of the effect of primary number
- Multiparametric analysis using observables other than X_{max}, e.g. muon count with Auger Phase II upgrade



Project 2.2

Non-parametric UHECR spectrum modelling

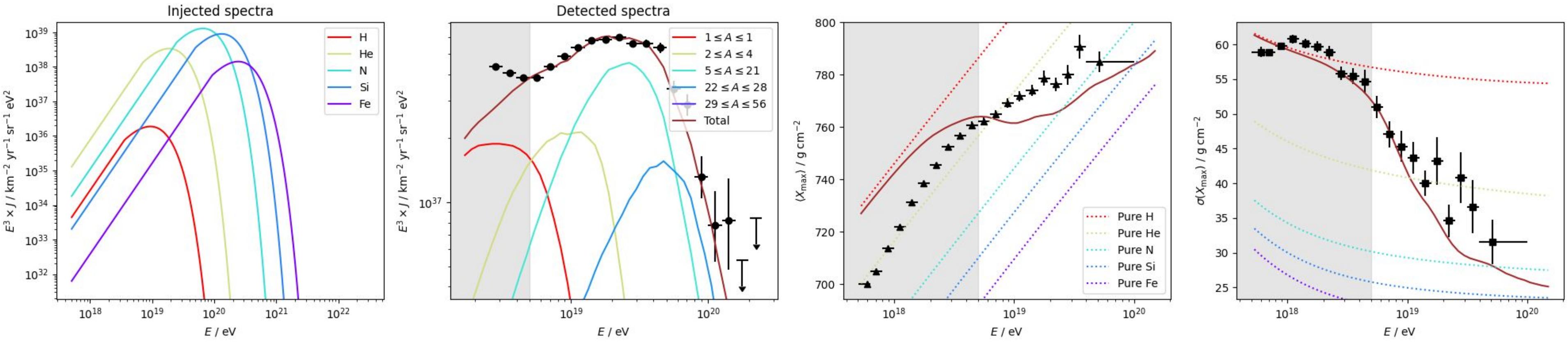
Project 2.2

Combined ***ift**** of UHECR spectrum and composition

*fit using Information Field Theory (IFT)

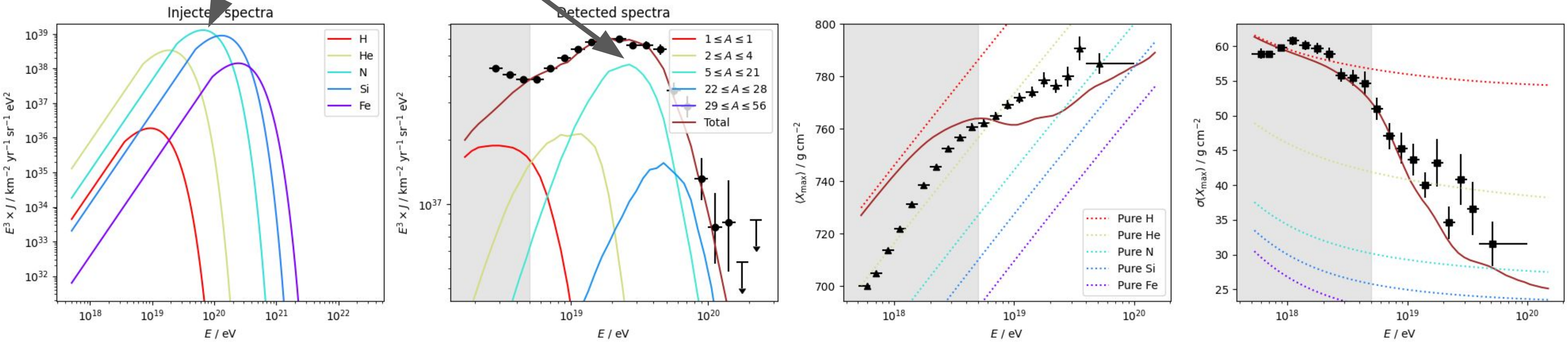
Combined fit of UHECR spectrum and composition

- Assuming cosmologically uniform injection we can propagate CR to Earth accounting for $p\gamma$ interactions
- With a simple power-law + exponential cutoff we can reproduce the observed spectrum and shower depth data *reasonably* well…



Combined fit of UHECR spectrum and composition

- … but the resulting spectra seem strange, with very hard injection and large dependence on parametric spectrum shape
- The observed spectra are shaped by a very narrow R region around the $E_{\max}$

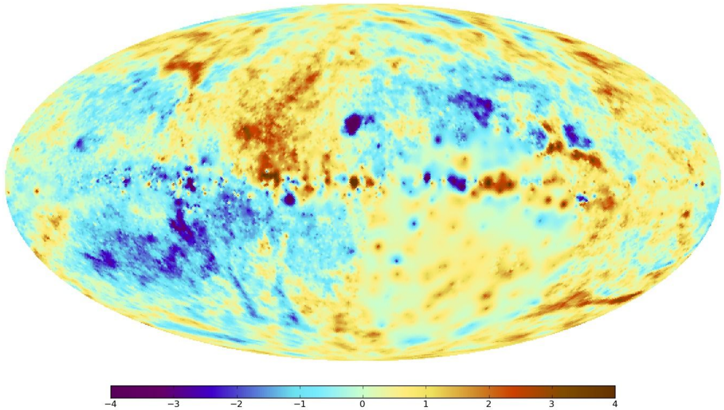


Nonparametric spectrum modelling

- What if we impose an overly restrictive model of the spectrum with the conventional PL + Exp or other parametrizations?
- Can we create a more flexible model for spectral shape while still preserving some physical interpretability?
- Let's try to look for approaches in other signal reconstruction problems…

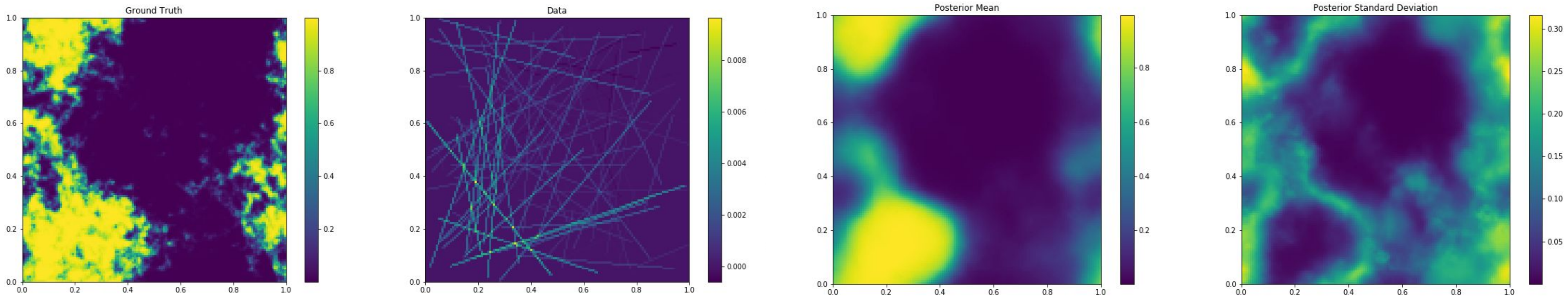
Information Field Theory (IFT)

- In theory: “An information theory for fields [joins] the measurement and prior information into probabilistic statements on field configurations.”
- In practice: Bayesian framework tailored for signal reconstruction problems
- Introduced and developed primarily by Torsten Enßlin and his group in MPA Garching
- Successfully applied to reconstruct multidimensional distributions of Galactic dust & gas density, diffuse γ-rays, magnetic field, gravitational potential, etc



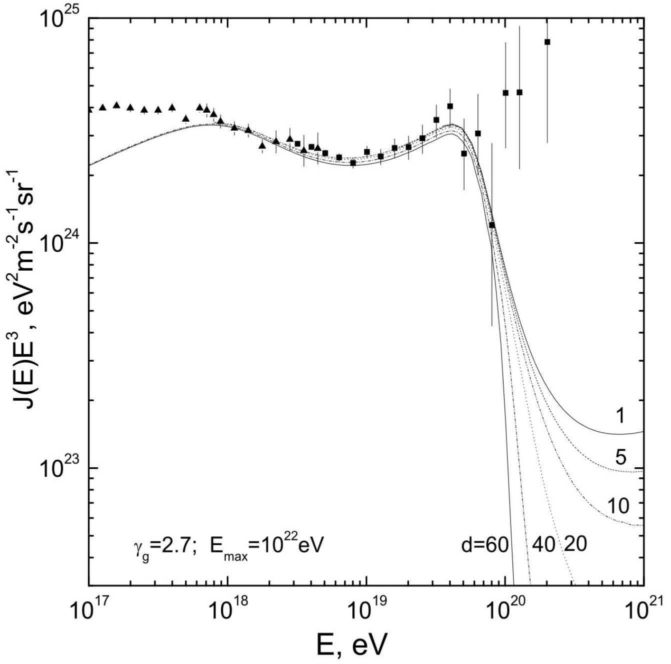
Numerical Information Field Theory

- Python implementation of IFT principles and reconstruction algorithms
- Practically, the field under investigation is reconstructed as a Gaussian process (Variational Inference, VI)



Toy problem: Berezinsky spectrum

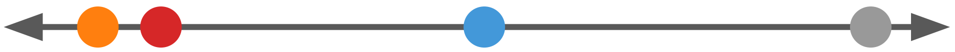
- Simplest setup: assume all UHECR are protons
- Injected PL spectrum shaped by propagation naturally produces ankle-like feature and GZK cutoff
- Latest UHECR data on spectrum and composition firmly exclude this scenario, but we can still use it as a toy problem



Berezinsky, V., Gazizov, A. & Grigorieva, S. On astrophysical solution to ultrahigh energy cosmic rays. Phys. Rev. D 74, (2006)

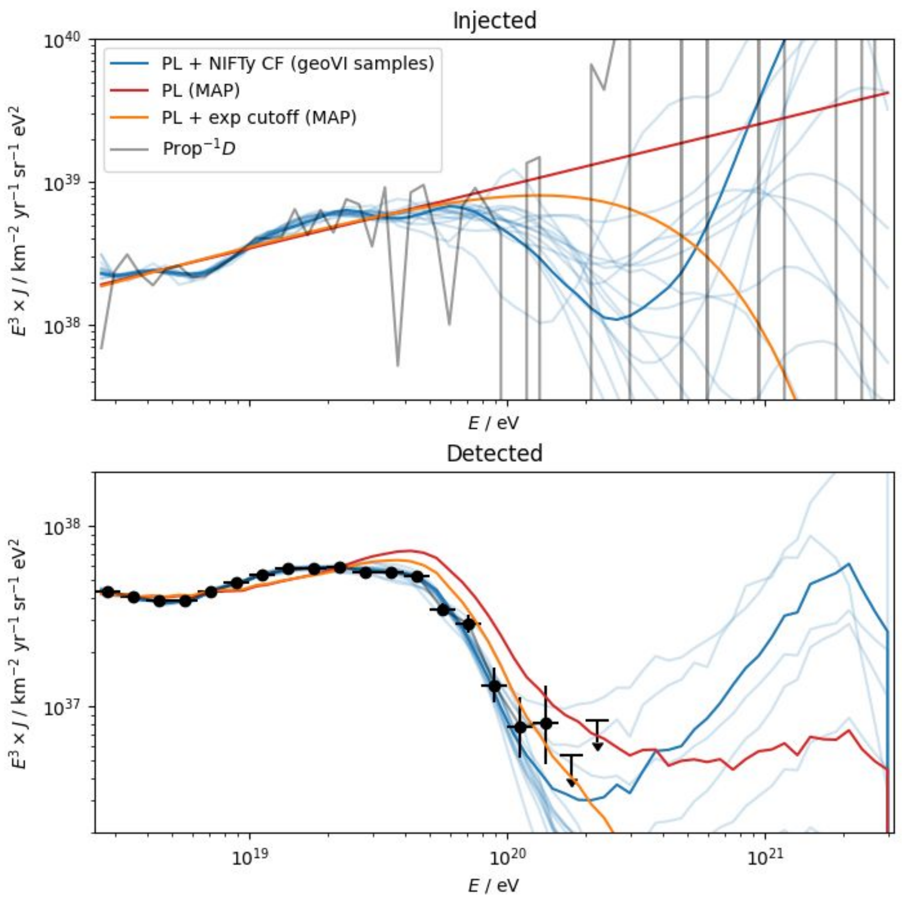
Toy problem

Parametric   Nonparametric



Too restrictive!   Unphysical

- Physical prior ensures the spectrum is not completely arbitrary
- Data constrain and bend the spectrum shape relatively freely
- VI algorithm characterizes the uncertainty of the posterior distribution



Outlook

- Active work in progress!
- Move to realistic combined fit analysis in the near future
- Extension of the nonparametric analysis to redshift source evolution, multiple populations and possible deviations from R-dependence

Activities

Conferences and collaboration meetings

- The Pierre Auger Collaboration Analysis meeting, L'Aquila, May 18–23 2025 (*Mass composition analyses by the L'Aquila group* and *Scrutinizing credible intervals on individual mass fractions*)
- International Cosmic Ray Conference (ICRC) 2025, Geneva, Switzerland, Jul 14–24 2025 (*Revisiting the Cosmic Ray knee: new insights from LHAASO*)

Schools

- International School of Nonequilibrium Phenomena, 6th Course "Astroparticle Physics, Dark Matter and High Energy Physics", Erice, Italy, August 1–7 2024

Research visit

- RWTH Aachen University, Aachen, Germany, August 4 – November 3 2025 under supervision of Prof. Philipp Mertsch, part of the PNRR-funded visiting period

Thanks for your attention!

Bayesian analysis

- Model = elemental CR spectra in R + per-experiment energy scale shifts
- Likelihood = χ^2 sum over available observables (elemental and all-particle spectra, mean log mass, etc)
- Priors = lognormal for energy scale shifts, improper flat otherwise

MCMC sampling → model prediction density, credible bounds, maximum aposteriori