

LABORATÓRIO DE INSTRUMENTAÇÃO E FÍSICA EXPERIMENTAL DE PARTÍCULAS partículas e tecnologia

Data Science in Astroparticle Physics

L. Cazon (cazon@lip.pt)

Astroparticle physics

• Study of elementary particles of extraterrestrial origin

It is a multidisciplinary field

- It connects Astrophysics, Particle Physics and Cosmology
- Technological challenge to build experiments: Electronics, computing, atmospheric physics, optics, communications, geology....



History

- 1896 H. Becquerel discovered radioactivity
- 1913 V. Hess discovered that part of the radiation had extraterrestrial origin
- 1920 R. Millikan called it Cosmic Rays



AMS Alpha Magnetic Spectrometer

Looks for dark matter evidence through cosmic rays



AMS: A TeV precision, multipurpose, magnetic spectrometer



SNO+

Neutrino hunters

Very deep (2km)

Ideal conditions for background physics





LATTES

Gamma Ray astronomy

Unexplored energy region

LATTES @ ALMA site Large Array Telescope for Tracking Energetic Sources

 Joint Brazil / Italy / Portugal initiative
Interest from Czech group

Possible site:

- Atacama Large Millimeter Array site
- Chajnantor plateau
- ♦ 5200 meters altitude in north Chile
- ♦ Good position to survey the Galactic Center



LATTES array

trategies for primary discrimination



2

AUGER

Ultra High Energy Cosmic Rays

The Pierre Auger Observatory



- Malargüe. Mendoza
- Latitude 35 S Longitude 69 W
- 1400m a.s.l. X=870 g cm²
- Data taking since 2004
- Installation completed in 2008

Surface Detector (SD)

1600 Cherenkov stations spaced 1.5 km Area of 3000 km² 100% duty cycle Provides Large Statistics



+ Enhancements: AMIGA, HEAT, Radio, etc





Energy comparison

~1000 per km²/year

1.5 eV



10 000.0 eV



~1000 per km²/year

6 500 000 000 000.0 eV



1 per km²/century

UHECR. 300 000 000 000 000 000 000.0 eV



Muon Tomography

LIP already produces the technology needed

Many applications

Nuclear Smuggling

Civil engeneer applications









Muon Tomography

Image of large human and geological structures

Mining, search for cavities, aquifers, magma conduits.

L. Cazor



Extensive air shower

The complexity of data

- 1. Individual detectors are triggered
- 2. They communicate to the central station
- 3. if several triggered detectors meet a topological criteria, an event is created
- 4. The event is checked to be physically consistent
- 5. A last quality selection is performed.



The complexity of noise



The Scientific method



Monte Carlo simulations

- Some phenomena are made of collection of many single stochastic events.
- A **new phenomenum emerges at a larger scale**, and sometimes it is not possible to understand the large scale from the small scale constituents.
- For instance, the simulation of a full air ahower at 1E19 eV takes 1 year in a single CPU

1E10 particles, ~1E11 reactions, ~1TB

we use "thinning" algorithms to sample only the most relevant process (<1day <1GB per shower in a single santard CPU)







A cosmic ray enters the atmosphere





A cosmic ray enters the atmosphere



Its energy, composition, and arrival direction are the inputs to solve puzzle about their origin



Its energy, composition, and arrival direction are the inputs to solve puzzle about their origin











Electrons Photons Muons Neutrons protons

Ultra-High Energy interaction. Cascade start-up



Electrons Photons Muons Neutrons protons

2nd and 3rd generation. Leading baryons still carrying very high energy.



Electrons Photons Muons Neutrons protons

The orignal information information is being camouflaged

Hajo Drescher, Frankfurt U.

time = -100 µs



Beyond statistical uncertainty: Systematic uncertainty

This uncertainty it is introduced by an inaccuracy (involving either the observation or measurement process) inherent to the system, the theory, your detectors, or (unnknown) biasses when sampling.



Calibration to remove bias





Model	R_E	R _{had}
QII-04 p	$1.09 \pm 0.08 \pm 0.09$	$1.59 \pm 0.17 \pm 0.09$
QII-04 mixed	$1.00 \pm 0.08 \pm 0.11$	$1.61 \pm 0.18 \pm 0.11$
EPOS p	$1.04 \pm 0.08 \pm 0.08$	$1.45 \pm 0.16 \pm 0.08$
EPOS mixed	$1.00 \pm 0.07 \pm 0.08$	$1.33 \pm 0.13 \pm 0.09$

Extrapolation vs interpolation

Whereas interpolation is safe, (by means of Machine Learning or conventional methods), extrapolation towards regions with no prior data must rely on actual understanding of the underlying phenomena.



Heat Map of 1,058,383 Basketball Shots from NCAA Games Starting with the 2013-14 season. Via Sportradar data located in BigQuery





Max Woolf - minimaxir.com

reddit.com/r/dataisbeautiful

Is this just a poor simplistic analysis or is it some scientific true?

Watch out!

Distribution of ratings



The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge



Partnerships opportunities



Joint projects LIP-Industry

Education and training networks LIP/Industry

EU grants joint calls

Data analysis, data mining and knowledge building at LIP

Simulations at LIP Computing **Design of Data Acquisition** Quality control of physical processes Others

Conclusions

- Astroparticle physics is a highly multidisciplinary field, with complex data..
- The scientific method is the safest way to produce knowledge (in any field), and make excursions beyond the confort zone.
- There are direct applications of astroparticle physics
- Beyond that, physics is a paradigm for *data Science* to extract solid and unbiased knowledge on any kind of dataset.

Back up slides

Dark Matter

Annihilation of Dark Matter produces additional e⁺ which are characterized by a sharp drop off at the mass of dark matter.



Selection of the signal: The p signal is well separated from the backgrounds.







