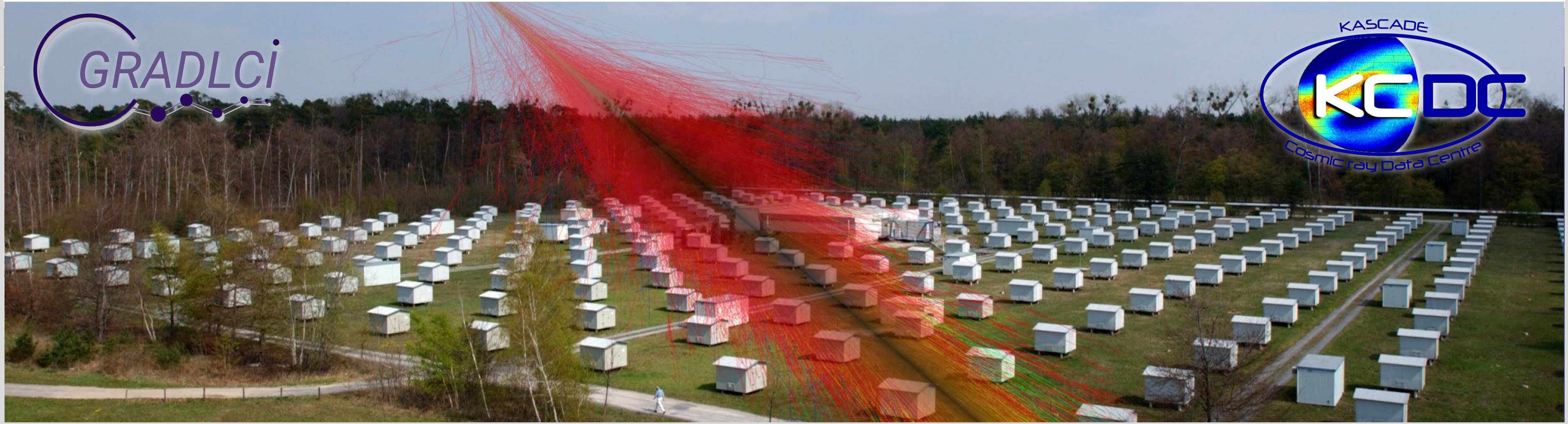
Astroparticle physics online masterclass built on the KASCADE Cosmic Ray Data Centre

Karlsruher Institut für Technologie

V.Tokareva*, K. Link, A. Haungs, J. Wochele, D.Wochele, F. Polgart, D. Kang, O. Tkachenko, P. Kundal

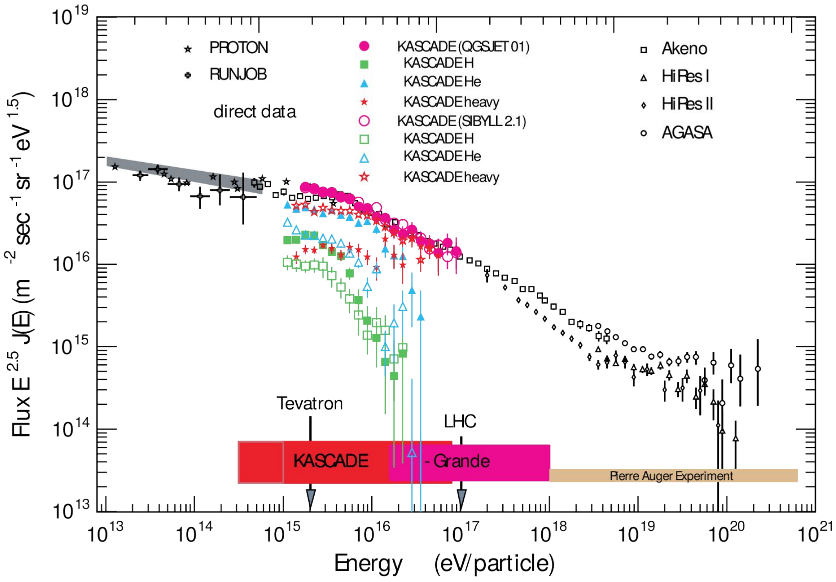
INSTITUT FÜR ASTROTEILCHENPHYSIK (IAP)

18 March 2021, DPG Spring meeting



Motivation

- Online education: bringing an equal access to education worldwide
- Corona situation
- GRADLCI project and IAP BigData Infrastructure



Primary cosmic ray flux and primary energy range covered by KASCADE-Grande and KASCADE experiments

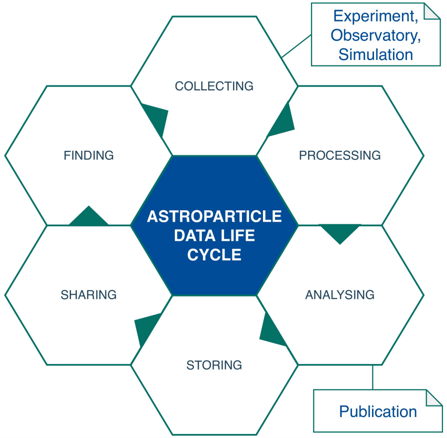
IAP BigData Infrastructure timeline

1989 - 2013: Data engineering for KASCADE and other experiments
2013 - now: KASCADE Cosmic-Ray Data Center (KCDC)

- only open-source technologies
- all-in-one: data center, archive, information and educational platform
- http://kcdc.ikp.kit.edu

2018 - now: German-Russian Astroparticle Data Life Cycle Initiative (GRADLCI):

- KCDC update
- online analysis server
- machine learning methods
- outreach and education
- https://gradlc-dc.ikp.kit.edu/



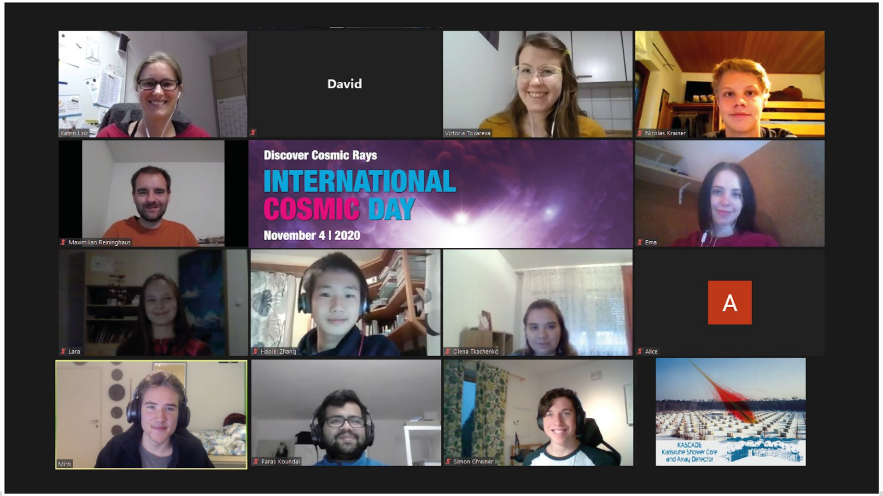
Educational program

Karlsruher Institut für Technologie

Hour	Session	Location	Time
14:00	International Cosmic Day: Welcome	Online	14:00 - 14:05
	Astroteilchenphysik	Online	14:05 - 15:00
15:00	Luftschauerphysik	Online	15:00 - 15:30
	Pause	Online	15:30 - 15:45
	KASCADE Grande	Online	15:45 - 16:00
16:00			
	KCDC Introduction	Online	16:15 - 16:30
	Break	Online	16:30 - 16:45
	Python /Data Analysis Introduction	Online	16:45 - 17:05
17:00	Data analysis	Online	17:05 - 17:55
18:00	Video call	Online	17:55 - 18:15

Lecturers and tutors:

- Maximilian Reininghaus
- Paras Koundal
- Victoria Tokareva
- Katrin Link
- Olena Tkachenko
- Miro Joensuu



Requirements to the tutorial environment

- Identical workspace for all participants and tutors
- Save as much time as possible on installation and setup of the working environment
- Fast processing of rather significant amount of data
- Data format, well-known for students
- Interactivity
- Minimal programming knowledge required

Requirements to the tutorial environment

Karlsruher Institut für Technologie

- Identical workspace for all participants and tutors
- Save as much time as possible on installation and setup of the working environment
- Fast processing of rather significant amount of data
- Data format, well-known for students
- Interactivity
- Minimal programming knowledge required

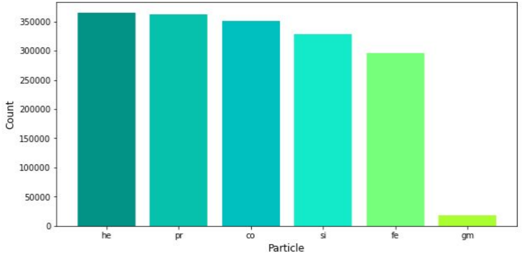
Solution: JupyterHub, Python3, Pandas+matplotlib, *.csv, *.txt.

Data samples

Karlsruher Institut für Technologie

- Sample 1:
  - KASCADE data shortened
  - .txt
  - 4102999 events

	E	Ne	Nmu	Ze
0	14.9335	4.55575	4.02881	18.2060
1	15.1466	4.83307	4.15094	23.3173
2	14.8551	4.56024	3.86447	21.4443
3	14.6315	4.22959	3.68777	21.1421
4	14.5349	4.54745	3.37969	18.2577



- Sample 2:
  - KASCADE simulations
  - QGS-Jet4 + FLUKA
  - .csv
  - 1720319 events

	Particle	lgE	X	Y	CoreDist	Ze	Az	lgNe	lgNmu
0	co	14.1173	59.4186	-39.391200	71.2898	17.5974	63.79230	3.53951	3.26399
1	co	14.1492	-11.8334	55.367400	56.6179	33.5999	205.23600	3.43951	3.08027
2	co	14.3944	-28.4504	-0.732422	28.4599	15.7591	60.74060	3.99656	3.49747
3	co	14.2358	61.4100	-28.061300	67.5176	26.3883	282.38000	2.98808	3.49411
4	co	14.4836	-45.8457	-35.637500	58.0677	25.4860	6.27402	4.13502	3.43070

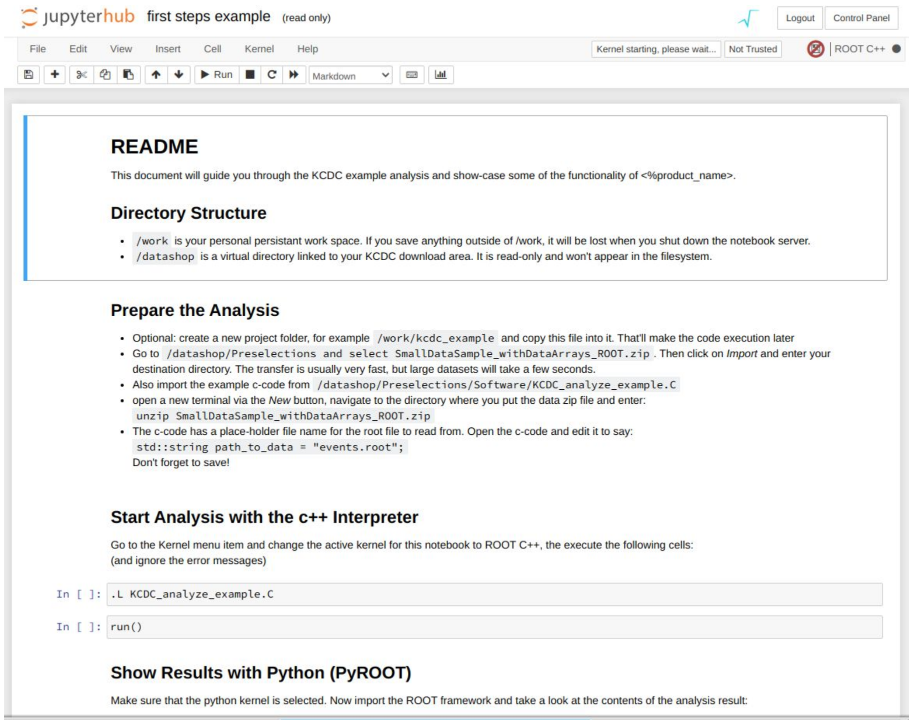
- Sample 3:
  - KASCADE data
  - Sept - Oct 1999
  - .csv
  - 1500317 events

	datetime	particle_id	energy	zenith	azimuth	core_x	core_y	electron_number	muon_number	shower_age	temperature	pressure
0	1998-05-11 07:54:03	0	14.619700	17.664600	117.292000	-22.430800	13.229600	4.145430	3.748440	1.041510	18.420200	1003.600000
1	1998-05-09 06:15:11	0	14.775900	27.404800	78.212900	68.574100	-23.827000	4.432480	3.697480	1.113110	19.080500	1011.330000
2	1998-05-14 07:50:46	0	14.730100	8.570720	159.879000	-65.301000	-60.631700	4.487310	3.829930	1.010680	13.820400	1007.990000
3	1998-05-16 20:59:44	0	14.576800	24.471400	260.102000	-1.075740	-59.533100	4.400930	3.442650	1.139100	17.870000	1005.960000
4	1998-05-19 01:52:24	0	14.946300	45.756200	3.883100	3.685060	-68.299400	3.947760	3.574800	0.854729	17.809900	1012.410000

Jupyter Interactive environment

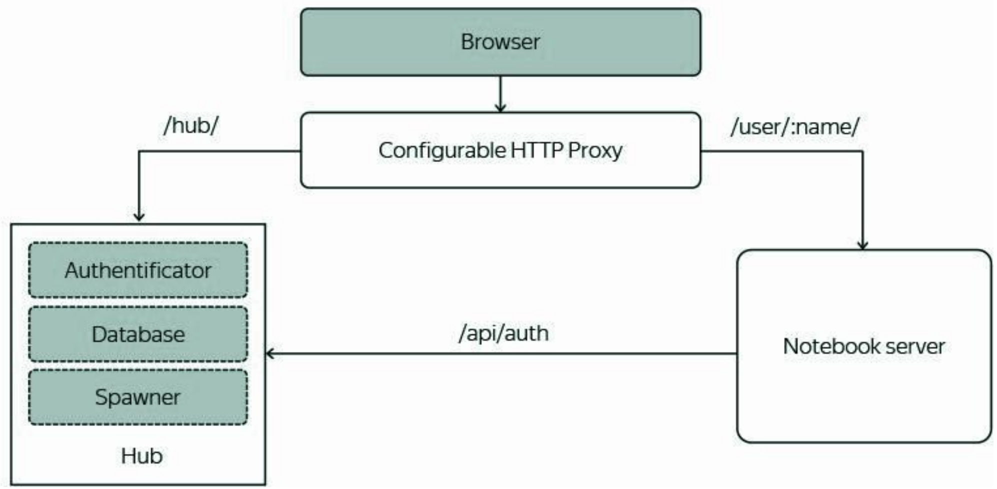
Karlsruher Institut für Technologie

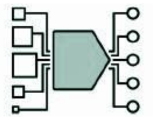
- Well-known
- Different kernels, pyRoot and C++ ROOT support
- Interactive code, easy to test and modify
- Visualizations, arbitrary output



JupyterHub @ IAP

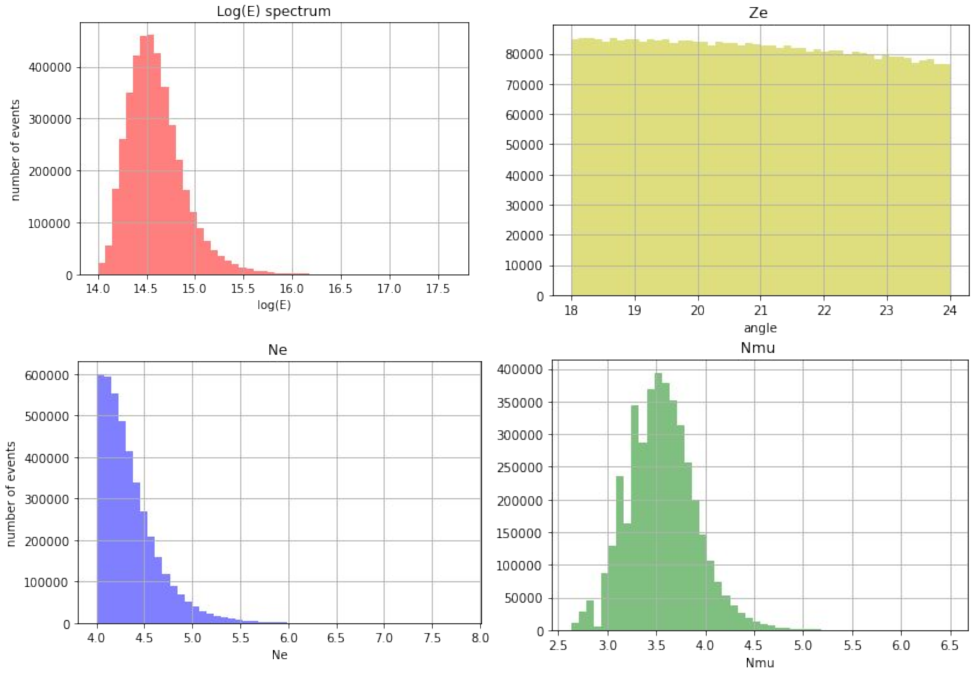
- Login via KCDC credentials
- Administration using Docker Swarm
- Up to 24 connections at the same time

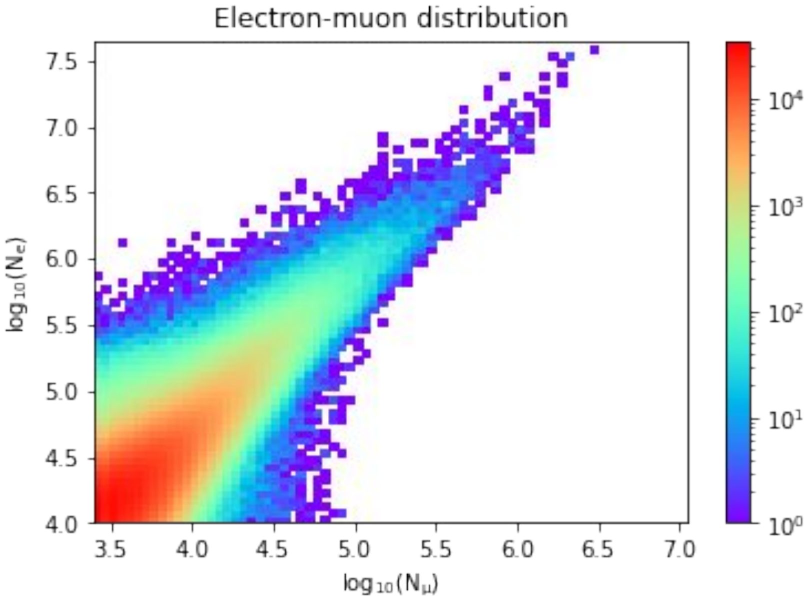


Authorization 	Run individual Jupyter servers 
Proxy to Jupyter servers 	System and database management 

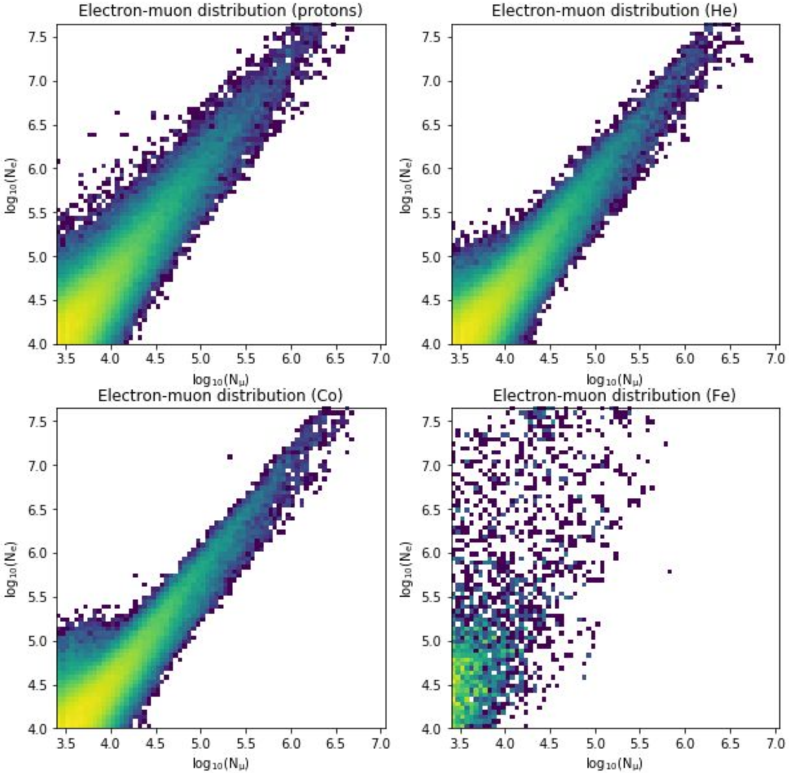
Karlsruher Institut für Technologie

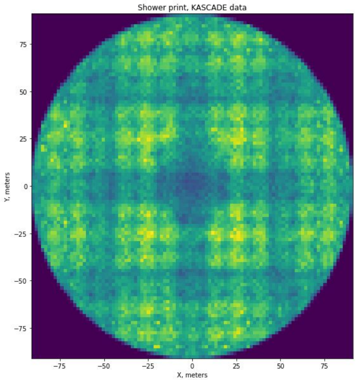
Particle mass tutorial

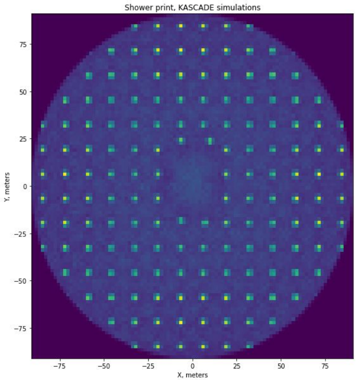


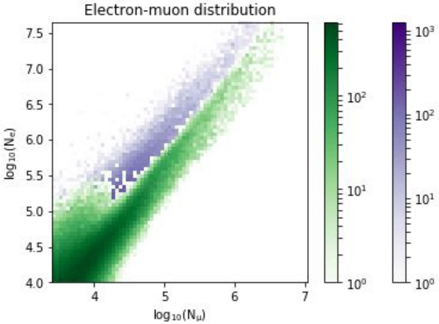


Particle mass tutorial

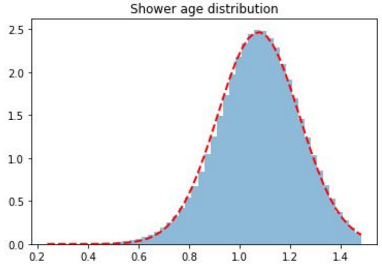









```
mu =  1.0742623777506533 , sigma =  0.16167346289339393
```



ICD booklet

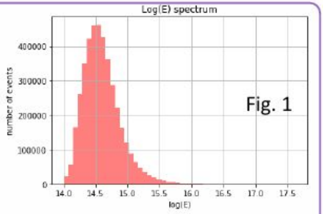
Karlsruher Institut für Technologie

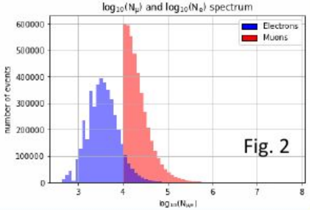
What have you done? Part II

During the data analysis we split into smaller groups of two or three, each working together with one tutor from the KIT and went through problems using Jupyter Notebook. First, some easier examples were done, such as how to import and use libraries, as well as plotting simple histograms. Then data from the KCDC library of both real and simulated events was imported. In the last step, plots of muon to electron ratios were made and compared with the results from simulated events.

What did you find out? Part I

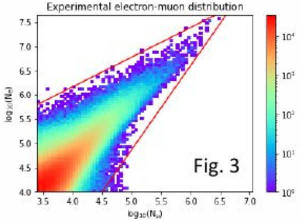
The number of events decreases overall with energy, because higher energy events creating these types of particles are rarer (fig.1).
Another interesting observation was that the electron and muon distributions are shifted to one another (fig.2).
Cosmic ray events create electrons and muons unequally. In general, more electrons are created in total, than muons, but the former will scatter away, while the latter can reach the ground more easily.

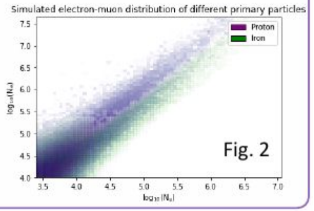




What did you find out? Part II

Events with heavier primary particles, such as iron cores, which tend to interact with the atmosphere earlier have more muons, than light primary particles, which tend to interact later and have an abundance of electrons. This can be seen in the 2d histograms from simulated events, which use different initial particles (as seen in fig. 3), such as iron or protons. These 2d histograms show the number of events with a specific number of electron and muon combinations through colour intensity. The results from the simulations can be used in experimental data to differentiate between different primary particles. Particles leading to events closer to the bottom to figure 4 are more likely to have heavy initial particles, while events closer to the top probably have light primary particles.





What's your take-home message?

Apart from the extremely interesting theoretical input regarding cosmic rays, the event gave us an insight into how data analysis works and how meaningful conclusions can be drawn from a large amount of data. This enabled us a sneak peek into how science is conducted in real life, which we greatly appreciated.

Analysis of cosmic ray data from KASCADE. In ICD booklet, 2020.

Outlook

- The BigData Infrastructure of IAP KIT and experience of our team in organisation of outreach events enabled us to organize an online data analysis masterclass
- In order to do so, we implemented an educational program and a masterclass program, organized preliminary data selection and preprocessing, instructed our students about work with KCDC and Jupyter Notebook in JupyterHub environment
- The master class was attended by 9 students aged 14-19 from Villach (Austria). Good beta testing for our outreach methodology and analysis environment at the same time
- Students acquired new knowledge and skills while working with open data and presented their results to other participants of ICD from Italy and Great Britain. Interesting discussion!
- Future plans include expanding our available materials with masterclasses that include machine learning, enriching our methodology with automated code validation capabilities, and further development of our BigData infrastructure

Thank you for your attention!

Contact: Victoria.Tokareva@kit.edu