PARTICLES AND COSMOLOGY

17th Baksan School
on Astroparticle Physics

Modern Statistical Methods and Tools

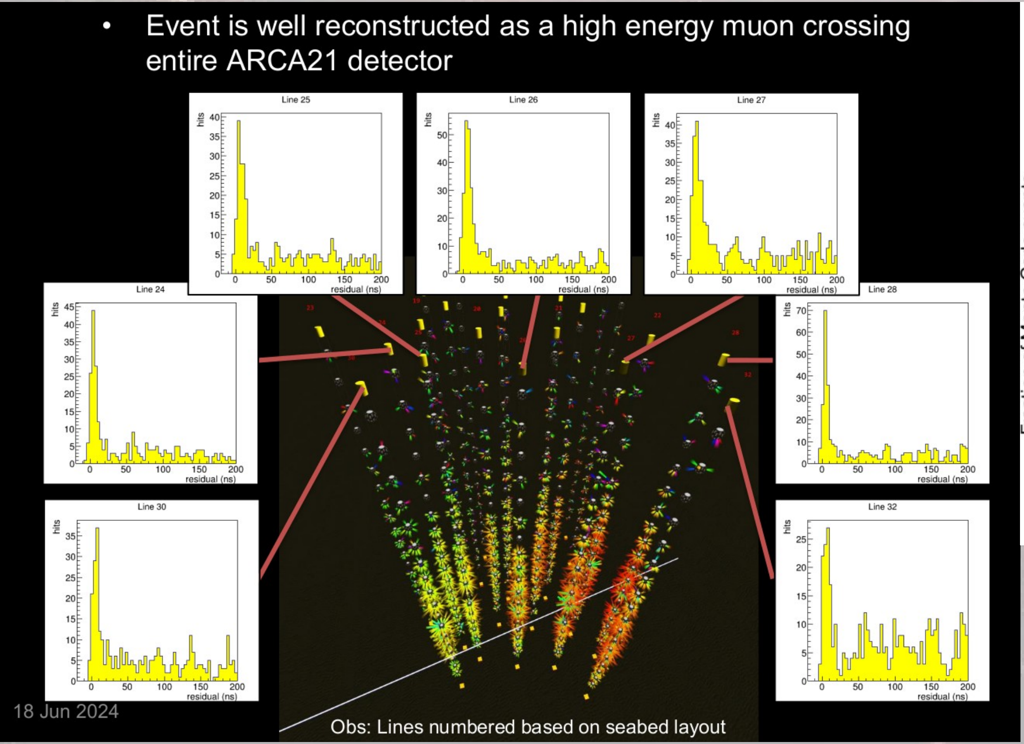
Grigory I. Rubtsov
Institute for Nuclear Research of the
Russian Academy of Sciences

Terskol, Kabardino-Balkarian Republic, Russia
April 4-11, 2025

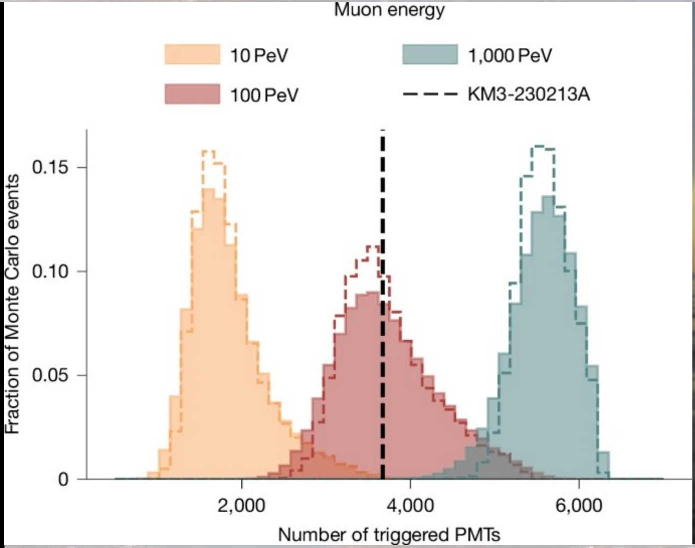
Why these lectures?

- We live in a random world
  - in several senses, will discuss in which
- We rely on big data for discoveries
  - sometimes big means **one** event
- To build physics it is necessary to draw reliable conclusions from observations
  - no matter how many events

Motivation example 1. KM3NeT ’fantastic’ particle

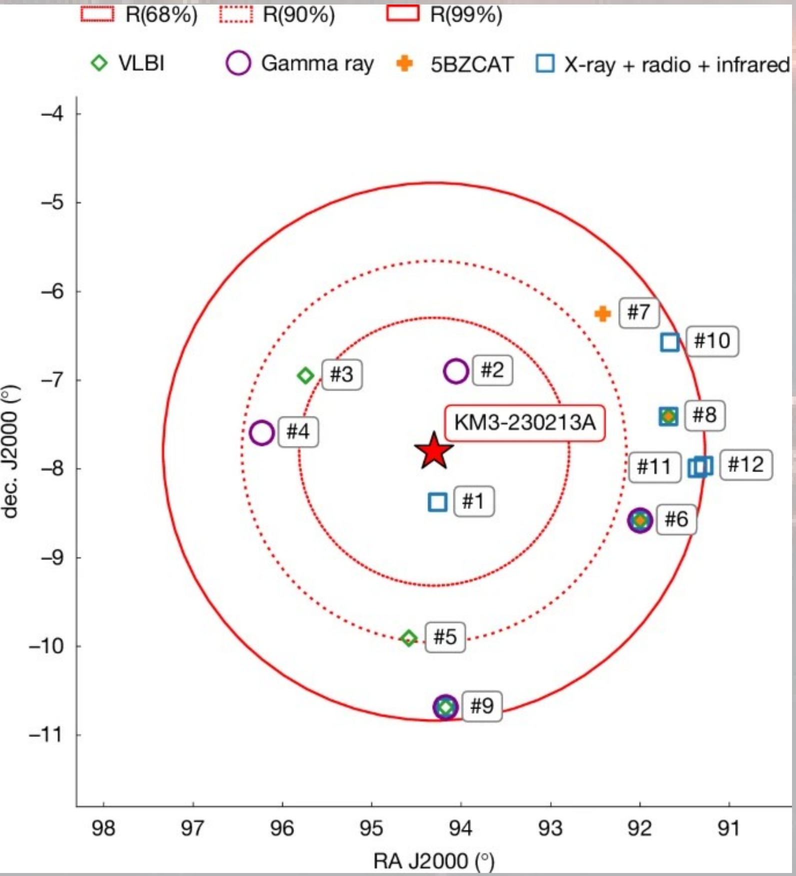


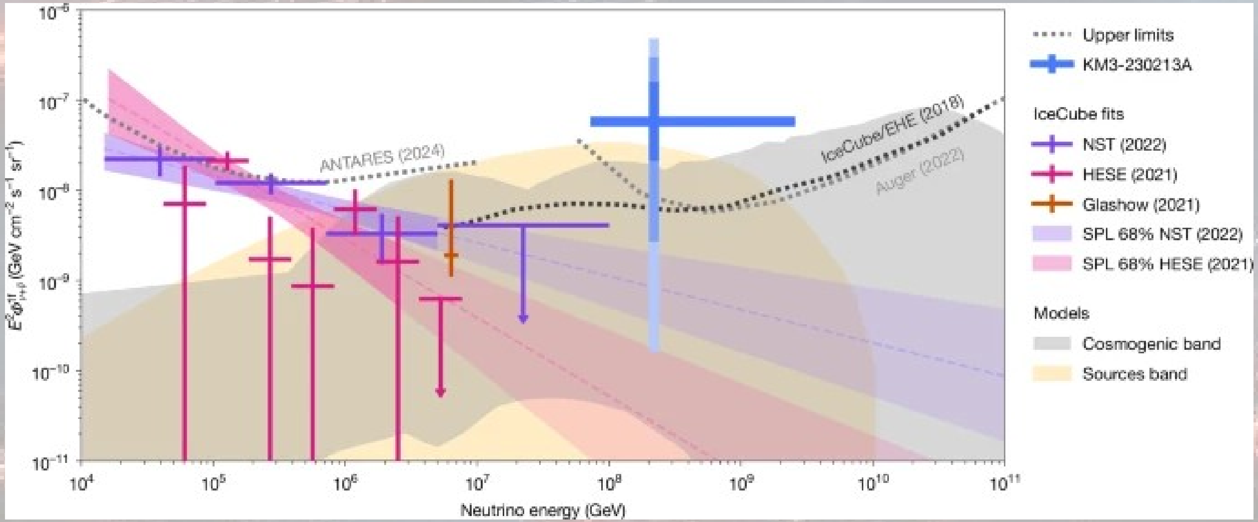
Ultra-high-energy event KM3-230213A



- 220 PeV energy!
  - Previous record — 6.05 PeV (IceCube)
- Can we call one event big data?

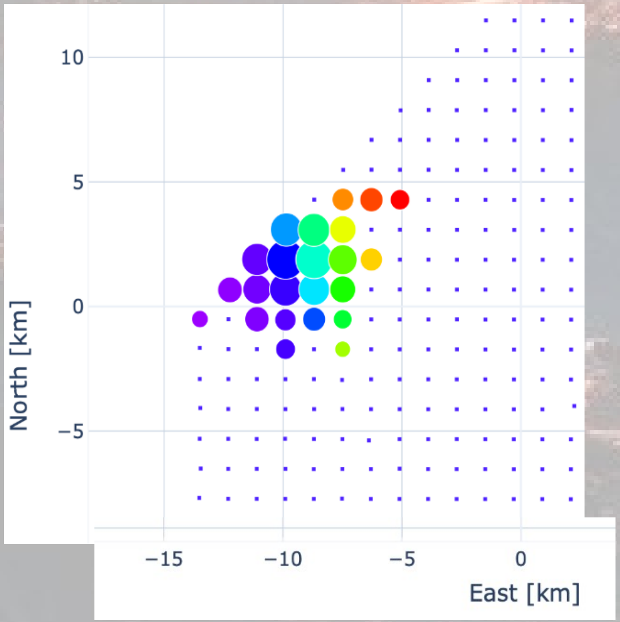
Motivation example 1. KM3NeT 'fantastic' particle

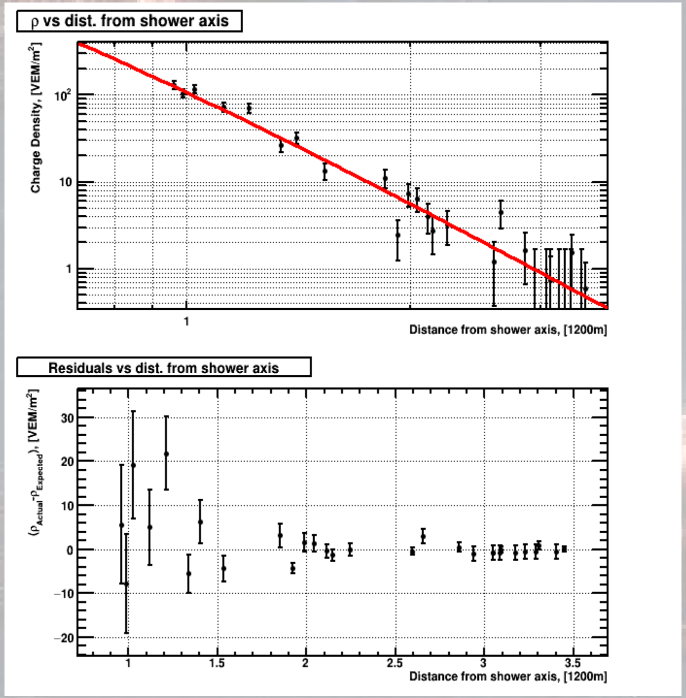




- What is the source of KM3-230213A?
  - arrival direction error is dominated by systematics
- Why not observed by IceCube & Baikal-GVD?

Motivation example 2. ”Amaterasu” particle by TA



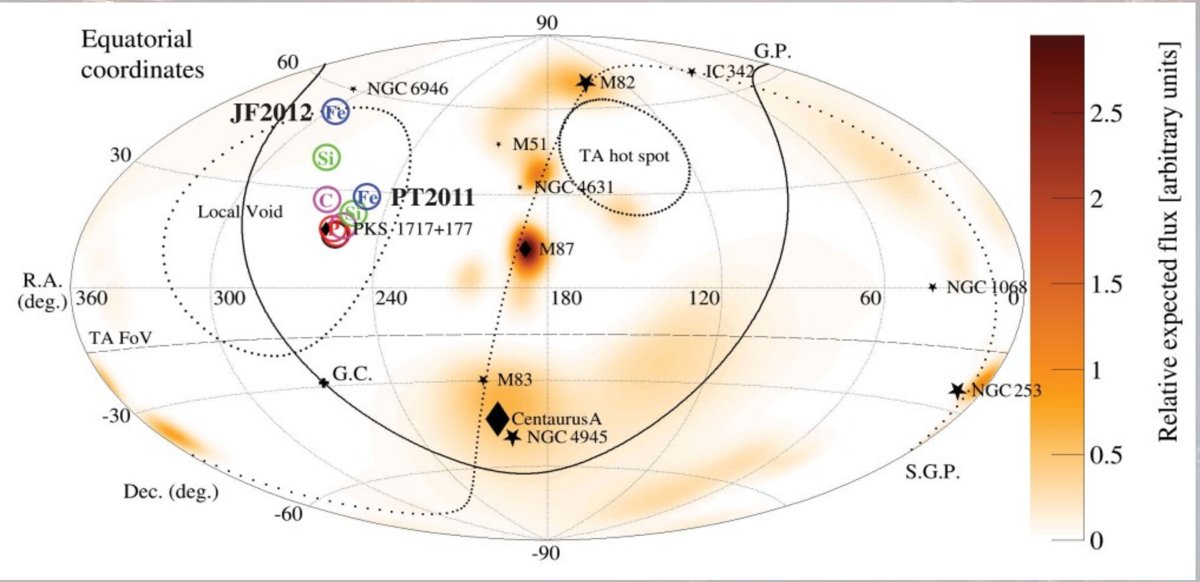


- Observed with TA SD at 10:35:56 on 27 May 2021 (UTC). No FD observation

Science 382, 903–907 (2023).

- E = **244 ± 29(stat.) ± 51(syst.) EeV**, zenith angle $\theta = 38.6°$

Motivation example 2. ”Amaterasu” particle by TA

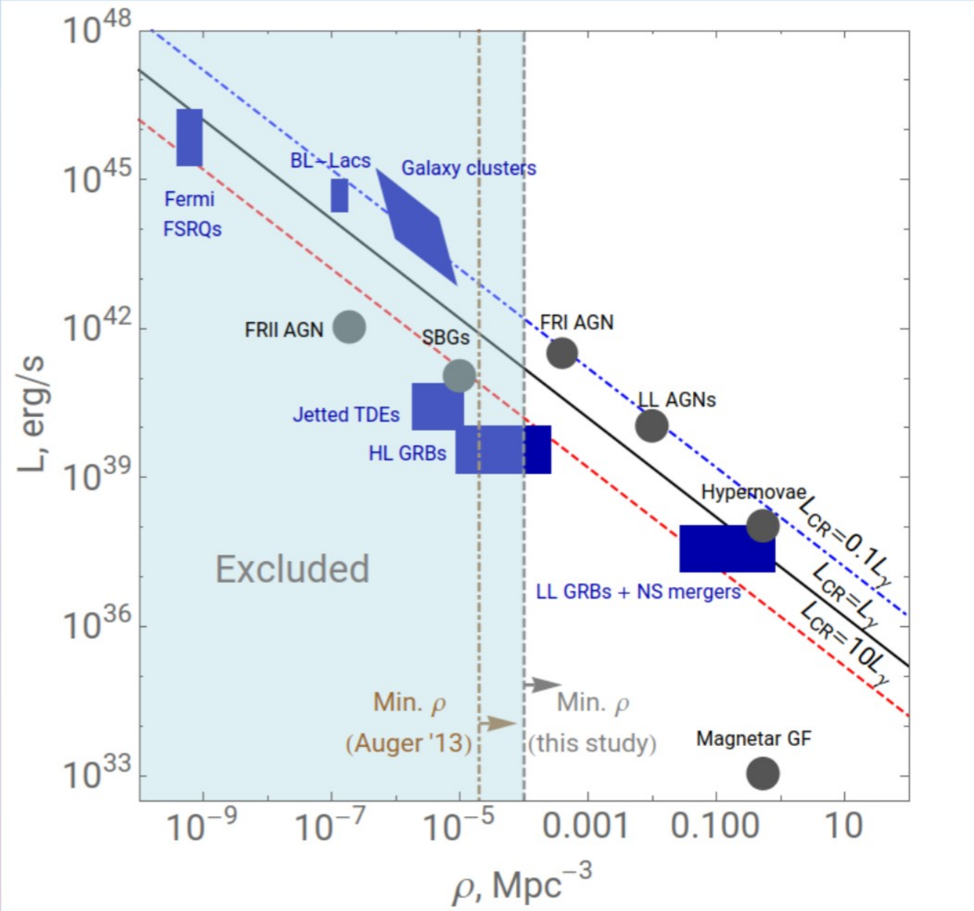


- E = 2.44×10^{20} эВ
- Event is coming from cosmic void
- Not a gamma-ray
- Primary particle should be a heavy nuclei
- The source is closer than 5 Mpc

Telescope Array Collaboration, Science 382, 903–907 (2023).
M. Kuznetsov, JCAP 04 (2024) 042

A number of conclusions are drawn from the observation of just one event!

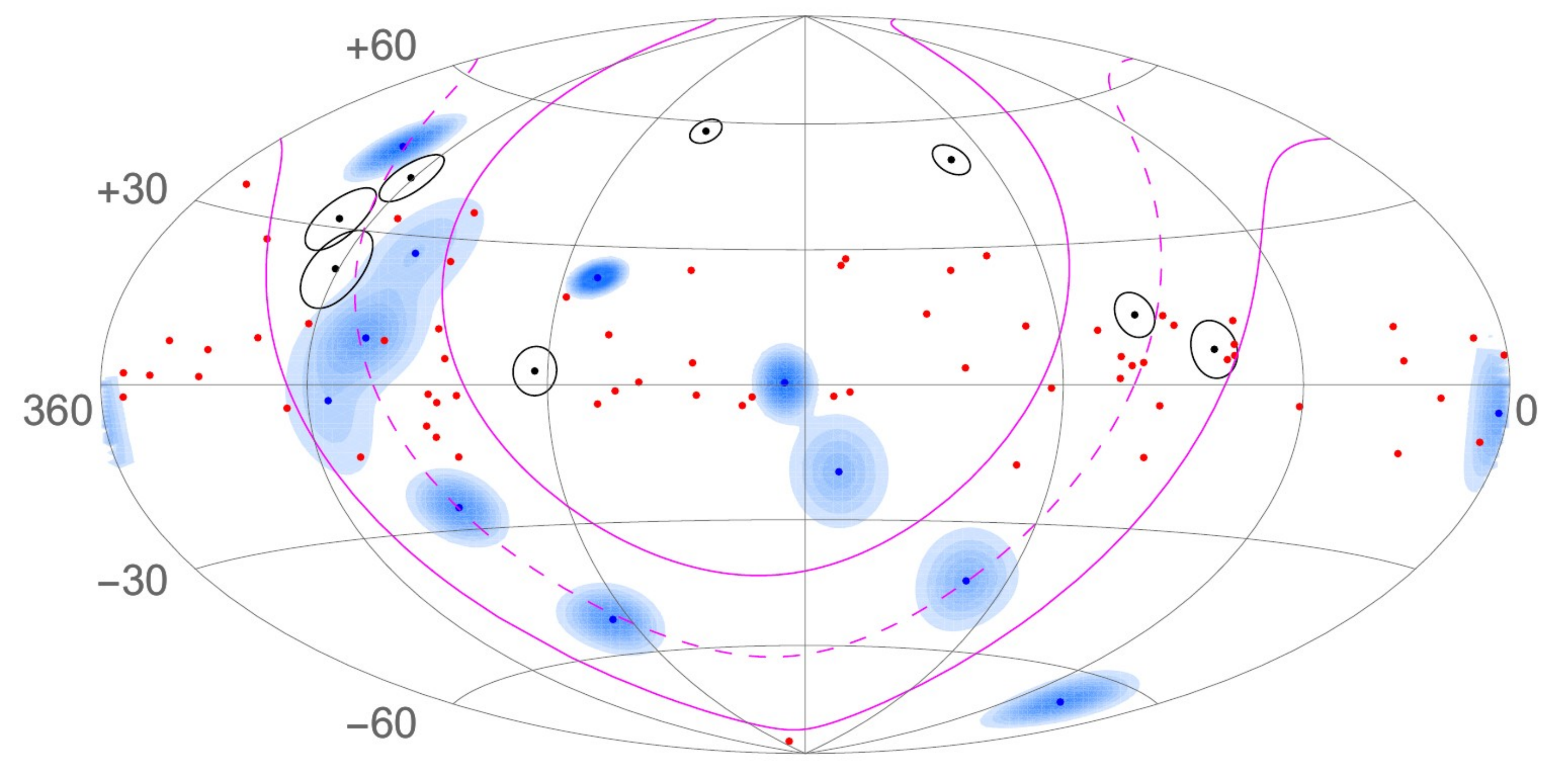
Motivation example 2. ”Amaterasu” particle by TA



- First composition independent constraint on the number density of UHECR sources

M. Kuznetsov, JCAP 04 (2024) 042

Motivation example 3.
Galactic neutrinos in Baikal-GVD



Baikal-GVD collaboration, Astrophys.J. 982 (2025) 2

Motivation example 3.
Galactic neutrinos in Baikal-GVD

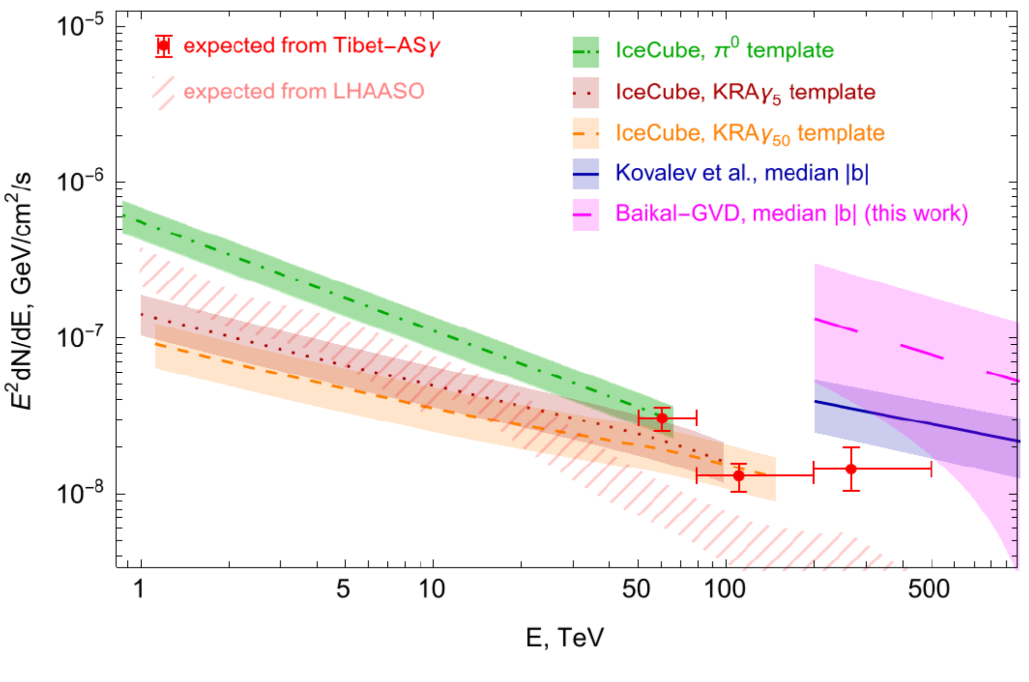
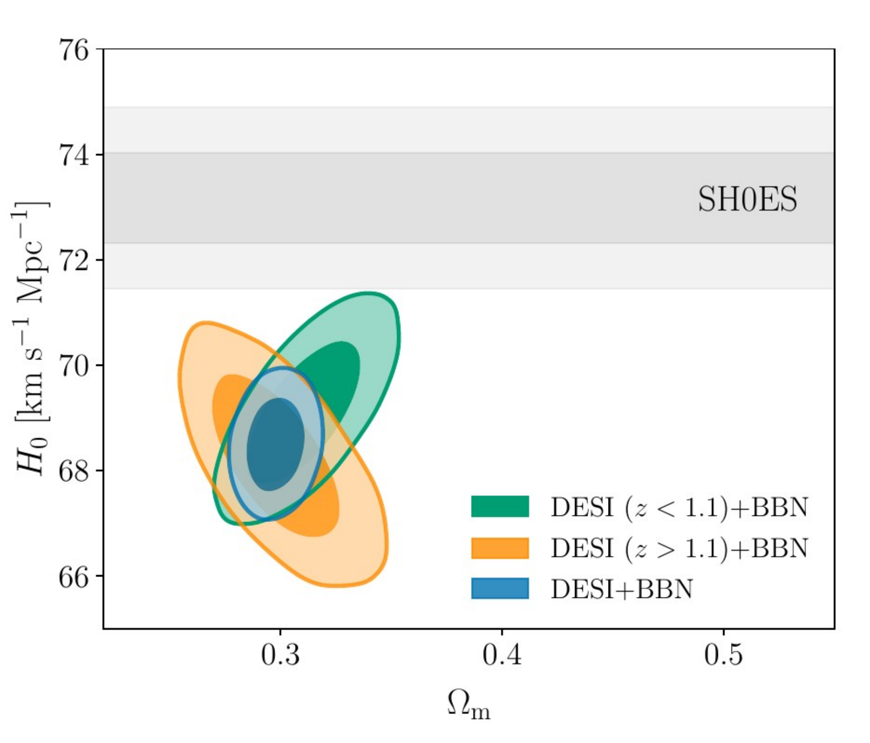
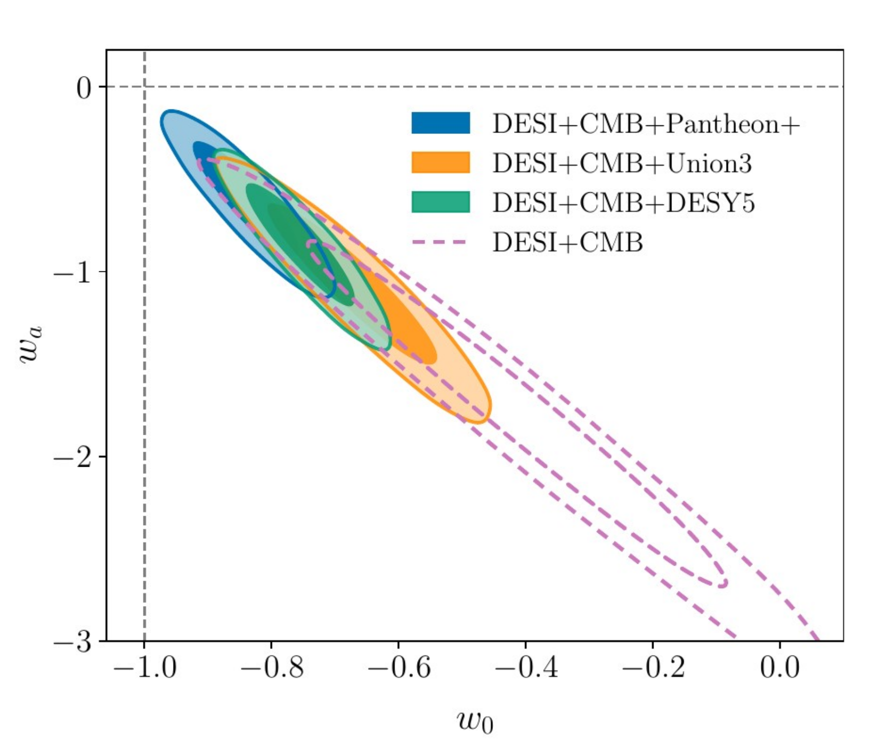


Figure 4. Estimated full-sky spectra of Galactic neutrinos

Baikal-GVD collaboration, Astrophys.J. 982 (2025) 2
See lectures by G. Safronov, this conference

Motivation example 4. DESI results on cosmology





DESI Collaboration, arXiv:2503.14738
(based on observation of 14 million galaxies)

What do I learn from these lecture?

- In brief:
  - Understand the nature of the randomness of the world
  - Understand the confidence level of the discovery or observation (3σ, 4σ, 5σ, ..)
  - Learn how to formulate hypotheses and how to exclude them
  - Learn how to estimate the parameters of the model, how to study parameter space for multi-parametric models
  - Hands on realistic data sets

What do I learn from these lecture?

- In detail:
  - Testing hypotheses: the probability and the likelihood
  - Frequentist and Bayesian approach to probability space
  - Markov Chain Monte Carlo method
    - ergodic assertion
    - warm up, convergence to stationary distribution
    - marginal distributions
  - Realistic conditions of modern data analysis
    - exposure and resolution
    - systematic errors
    - look elsewhere effect
  - Code references, snippets and more

Takeout 1

- Big data may be just one event
- Reliable scientific results are statistically correct conclusions from observations
- Our goal is to learn how to make these conclusions

What to start with?
The probability

Trinity: I know why you’re here, Neo. It's the question that brought you here. You know the question, just as I did.
Neo: What is Matrix?

The Matrix (1999)

- The question is:
  - What is the probability?

What to start with?
The probability

Trinity: I know why you’re here, Neo. It's the question that brought you here. You know the question, just as I did.
Neo: What is Matrix?

The Matrix (1999)

- The question is:
  - What is the probability?
- Answer: It is a measure $\mathbb{P}$ on the space Ω of elementary outcomes

What to start with? The probability

Trinity: I know why you’re here, Neo. It's the question that brought you here. You know the question, just as I did.
Neo: What is Matrix?

The Matrix (1999)

- The question is:
  - What is the probability?
- Answer: It is a measure $\mathbb{P}$ on the space Ω of elementary outcomes
- Practical meaning:
  - One may not speak about probability without the definition of probability space $(\Omega, \mathbb{P})$ // a common mistake
  - Other way round: if the space Ω is properly defined, it is easy to understand what is the probability of a particular outcome

What to start with?
The probability

Trinity: I know why you're here, Neo. It's the question that brought you here. You know the question, just as I did.
Neo: What is Matrix?

The Matrix (1999)

- The question is:
  - What is the probability?
- Answer: It is a measure $\mathbb{P}$ on the space Ω of elementary outcomes
- Practical meaning:
  - One may not speak about probability without the definition of probability space $(\Omega, \mathbb{P})$ // a common mistake
  - Other way round: if the space Ω is properly defined, it is easy to understand what is the probability of a particular outcome
- One real world event may belong to several probability spaces

Probability spaces? Too complex, give me the key

I offer you two popular choices of the day:

- I. Frequentist probability space
  - the model of the world (hypothesis) is fixed
  - the events are random
  - we calculate the probability of the event with the model

Probability spaces? Too complex, give me the key

I offer you two popular choices of the day:

- I. Frequentist probability space
  - the model of the world (hypothesis) is fixed
  - the events are random
  - we calculate the probability of the event with the model
- II. Bayesian probability space
  - both the model of the world and the events are random
  - we infer the probability of the models based on the observed events

Frequentist vs Bayesian

The future is not set.

There is no fate but what we make for ourselves.

The past, present and future are not set.

The fate is a random hypothesis.

Bayes' theorem

- Since both model (M) and event (obs) are random, joint and conditional probabilities may be defined

$$P(M, obs) = P(M|obs)P(obs) = P(obs|M)P(M)$$

$$P(M|obs) = \frac{P(obs|M)P(M)}{P(obs)}$$

- Here is a trick
  - since we observe the event P(obs) = 1
  - P(M) is a called a prior (a priori knowledge of model p.d.f.)

$$P(M|obs) = P(obs|M)P(M)$$

- We end up with calculating good old "frequentist" probability of the observation when the model is fixed (hence the confusion)

Probability spaces: small practice

- Provide an example of elementary outcomes belonging to the Frequentist probability space

Probability spaces: small practice

- Provide an example of elementary outcomes belonging to the Frequentist probability space

- *what is the probability that there is no rain tomorrow?*
- *what is the probability to observe neutrino today at Baikal-GVD assuming the neutrino flux published by KM3NeT*

Probability spaces: small practice

- Provide an example of elementary outcomes belonging to the Bayesian probability space

Probability spaces: small practice

- Provide an example of elementary outcomes belonging to the Bayesian probability space

- *the first step (joint probability)*
  - *there is no rain tomorrow and the equation of state of dark energy is evolving*
  - *the neutrino flux is equal to one published by KM3NET and there is a neutrino event today at Baikal-GVD*

Probability spaces: small practice

- Provide an example of elementary outcomes belonging to the Bayesian probability space

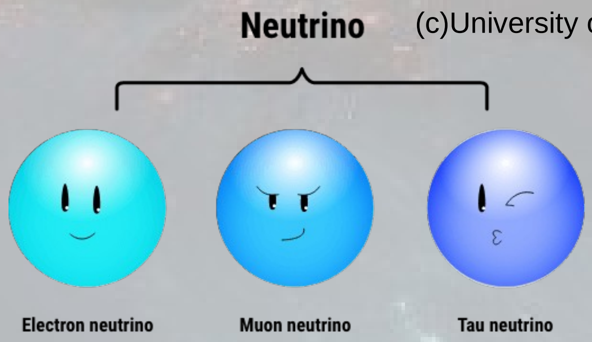
- *the second step (Bayesian approach)*
  - *the equation of state of dark energy is evolving given that there was rain yesterday*
  - *the neutrino flux is equal to one published by KM3NET given that there was a neutrino event yesterday at Baikal-GVD*

Next step: Building a hypothesis

- The hypothesis must be constructed in a mathematically complete way: it should be completely clear what is predicts
  - Include all parameters of the model
  - Include all parameters of the detector
  - Include conditions of event registration:
    - observation time
    - conditions for the start and end of observation
      - e.g. if obs. started by alert from another experiment the flux may not be considered as a stationary flux // common mistake
- The P(obs|M) is defined in unambiguous way

Model example: discrete case

- Let us have and experiment which has observed $n_e=1$ ν_e , $n_\mu=5$ ν_μ, and $n_\tau=0$ ν_τ (obs).
- What can we say about neutrino flux from this observation?
- Start constructing model (M):
  - suppose the neutrino fluxes are f_e, f_μ, and f_τ
  - the fractions are $\varepsilon_e=f_e/(f_e+f_\mu+f_\tau)$, ε_μ and ε_τ, $\varepsilon_e+\varepsilon_\mu+\varepsilon_\tau=1$
- What is a P(obs|M)?

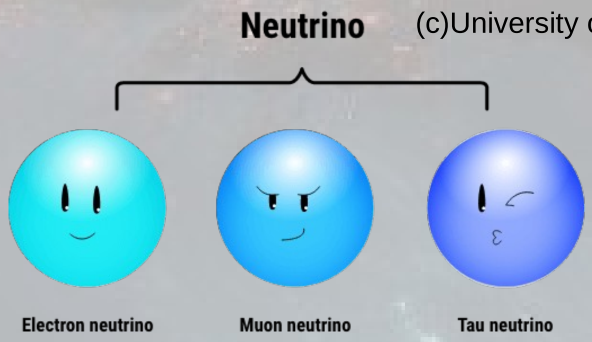


(c)University of Tokyo

Model example: discrete case

- Let us have an experiment which has observed $n_e=1$ ν_e , $n_\mu=5$ ν_μ, and $n_\tau=0$ ν_τ (obs).
- What can we say about neutrino flux from this observation?
- Start constructing model (M):
  - suppose the neutrino fluxes are f_e, f_μ, and f_τ
  - the fractions are $\varepsilon_e=f_e/(f_e+f_\mu+f_\tau)$, ε_μ and ε_τ, $\varepsilon_e+\varepsilon_\mu+\varepsilon_\tau=1$
- What is a P(obs|M)?
- Guess: multinomial distribution

$$P(obs|M)=\frac{6!}{5!1!0!}\varepsilon_e^1\varepsilon_\mu^5$$



(c)University of Tokyo

Model example: discrete case

- Guess: multinomial distribution

$$P(obs|M) = \frac{6!}{5!1!0!}\varepsilon_e^1 \varepsilon_\mu^5$$

- Is it correct?

Model example: discrete case

- Guess: multinomial distribution

$$P(obs|M)=\frac{6!}{5!1!0!}\varepsilon_e^1\varepsilon_\mu^5$$

- Is it correct?
- It may be correct, may be not
- Has experiment stopped after observing exactly 6 events?
- Is detector equivalently efficient for all type on neutrino?
- If both answers are yes, then the model with the fixed total number of events is appropriate
  - Find the best fit model by maximizing $P(\varepsilon_e,\varepsilon_\mu,\varepsilon_\tau)$ with the constraint $\varepsilon_e+\varepsilon_\mu+\varepsilon_\tau=1$

Model example: discrete case

- The neutrino fluxes are f_e, f_μ, and f_τ, in units $cm^{-2}\ s^{-1}$
- The effective area of detector is E_e,E_μ,E_τ for each ν type
- The observation time is T

Model example: discrete case

- The neutrino fluxes are f_e, f_μ, and f_τ, in units $cm^{-2}\ s^{-1}$
- The effective area of detector is E_e,E_μ,E_τ for each ν type
- The observation time is T
- The probability is given by a product of Poisson distributions

$$P(obs|M)=W(f_e E_e T, n_e)W(f_\mu E_\mu T, n_\mu)W(f_\tau E_\tau T, n_\tau)$$

$$W(\bar{n}, n)=\frac{\bar{n}^n}{n!}\exp(-\bar{n})$$

- Optimal fluxes f_e, f_μ, and f_τ may be found by maximizing P(obs|M), now without constraints

Model example: discrete case

$$P(obs|M) = W(f_e E_e T, 1) W(f_\mu E_\mu T, 5) W(f_\tau E_\tau T, 0)$$

- Is this a correct model?

Model example: discrete case

$$P(obs|M) = W(f_e E_e T, 1) W(f_\mu E_\mu T, 5) W(f_\tau E_\tau T, 0)$$

- Is this a correct model?
- Yes, if the source is nearby, so neutrino oscillations may be neglected

$$\begin{pmatrix} f_e \\ f_\mu \\ f_\tau \end{pmatrix} = M \begin{pmatrix} f_e^{src} \\ f_\mu^{src} \\ f_\tau^{src} \end{pmatrix}$$

- M – oscillation matrix (depends on the distance to source and energy spectrum)
- Now we optimize over source fluxes f_e^{src}, f_μ^{src}, and f_τ^{src}

Takeout 2

- The random world may refer to either
  - the world of random events
  - the world of random hypotheses
  - the world of random both
- One need to define the probability space before speaking about probability
- One needs to define a model in a mathematically and physically complete way

Final step: testing hypotheses

- Q: Can an experiment prove that the hypothesis is true?

Final step: testing hypotheses

- Q: Can an experiment prove that the hypothesis is true?
- A: No. BUT: one can exclude the hypothesis by experiment

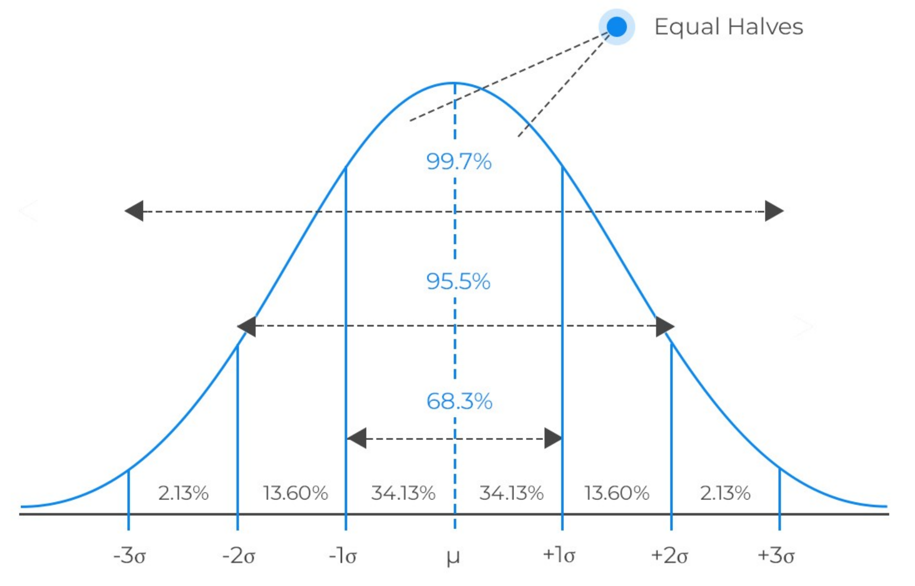
Final step: testing hypotheses

- Q: Can an experiment prove that the hypothesis is true?
- A: No. BUT: one can exclude the hypothesis by experiment
- Q: What is a discovery in this case?

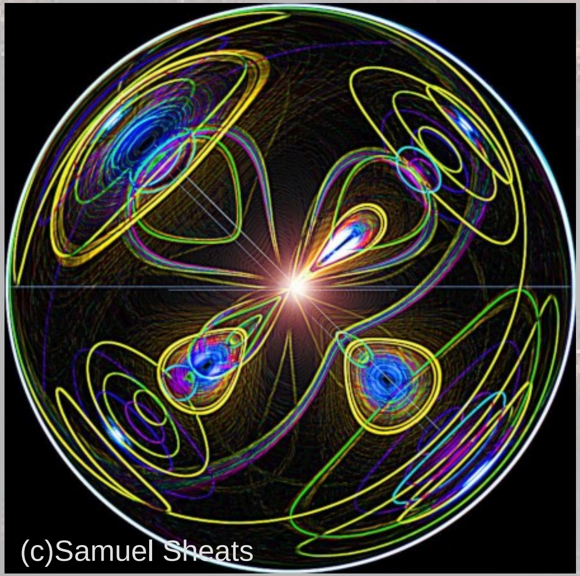
Final step: testing hypotheses

- Q: Can an experiment prove that the hypothesis is true?
- A: No. BUT: one can exclude the hypothesis by experiment
- Q: What is a discovery in this case?
- A: Exclusion of the currently accepted model (null hypothesis) at high confidence level, 5σ

N	p
1σ	32%
2σ	5%
3σ	0.27%
4σ	1 of 16 thousands
5σ	1 of 2 millions
6σ	1 of 500 millions

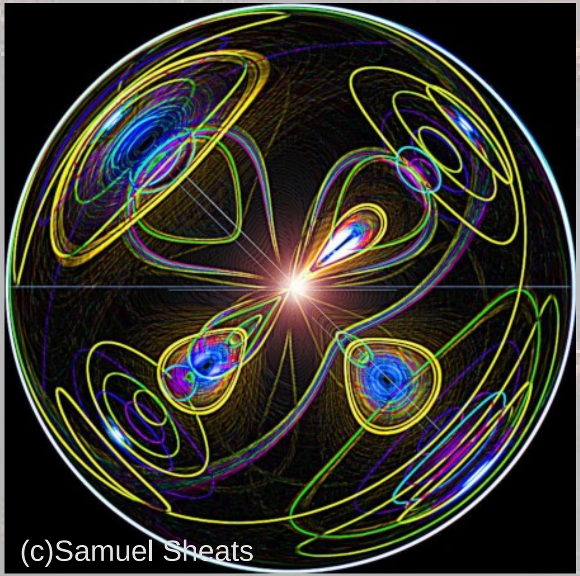


One can't prove the hypothesis. Example: Higgs boson



- Q: Have the Higgs boson been discovered at the LHC?

One can't prove the hypothesis. Example: Higgs boson



- Q: Have the Higgs boson been discovered at the LHC?
- A: not quite
- It was shown that the Standard Model without Higgs is excluded with 5σ significance and the SM with Higgs agrees to the data
- This doesn't mean that the SM with Higgs is a true model
- Other models are also possible. E.g. two scalar bosons with close masses and coupling constants equal to half of Higgs couplings

Testing hypotheses: Frequentist approach

1) Define the model M (or a set of models)
2) Define the null hypothesis, model M_0
3) Define the probability function P(obs|M)
4) Given the observed data, calculate P(obs|M) for null hypothesis M_0 and hypotheses M under the test
5) Consider the models with $P<0.05$ excluded with 2σ significance (it is also called 95% confidence level)
6) If the null hypothesis M_0 is excluded with 5σ significance and some of the models from M are not excluded, claim a discovery!

How many events are needed?

- Q: How many events one needs for discovery?

How many events are needed?

- Q: How many events one needs for discovery?
- A: It depends on the value of P
  - Even one event may be enough in certain cases
  - On the contrary, the billion events may be not enough

Takeout 3

- One can’t prove that the hypothesis is true
- A discovery is the exclusion of the null hypothesis
- The Higgs boson in nature does not necessarily exactly match the model representations

Conclusions

- The scientists live in a random world. Both the observable events and the parameters of the world are random
- Frequentist's approach doesn't respect the randomness of the parameters of the world
- A model must be defined in a mathematically and physically complete way before it may be tested
- Big data may be just one event

For the next lecture

- Bring your laptop to the next lecture
- Install python
  - `pip install notebook`
  - `pip install numpy,scipy,matplotlib,healpy`

Task for self-check

- Consider the first model of neutrino detections
- Following Frequentist approach find the answer to the questions:
  - Which models are excluded at 2σ level in $(\varepsilon_e, \varepsilon_\mu)$ plane?
  - How many events one needs to register in order to separate the models $(\varepsilon_e, \varepsilon_\mu, \varepsilon_\tau) = (1,0,0)$ and $(0,1,0)$ at 5σ confidence level?
- (*) Repeat the analysis with the account of neutrino oscillations for distant extragalactic source

Thank you!

Backup slides