

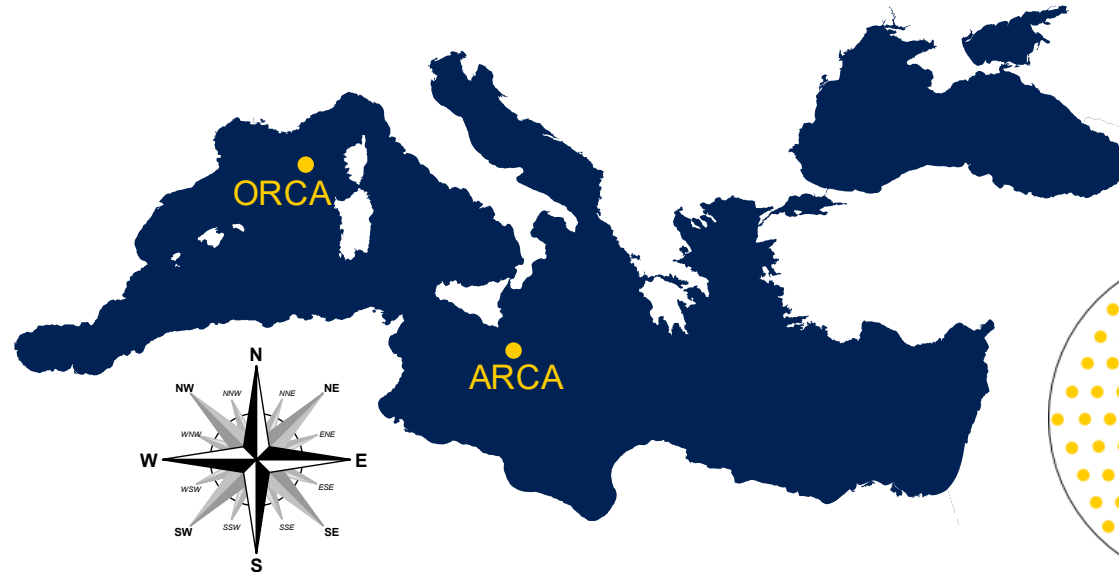
Study of cosmic ray muons with KM3NeT

Piotr Kalaczyński

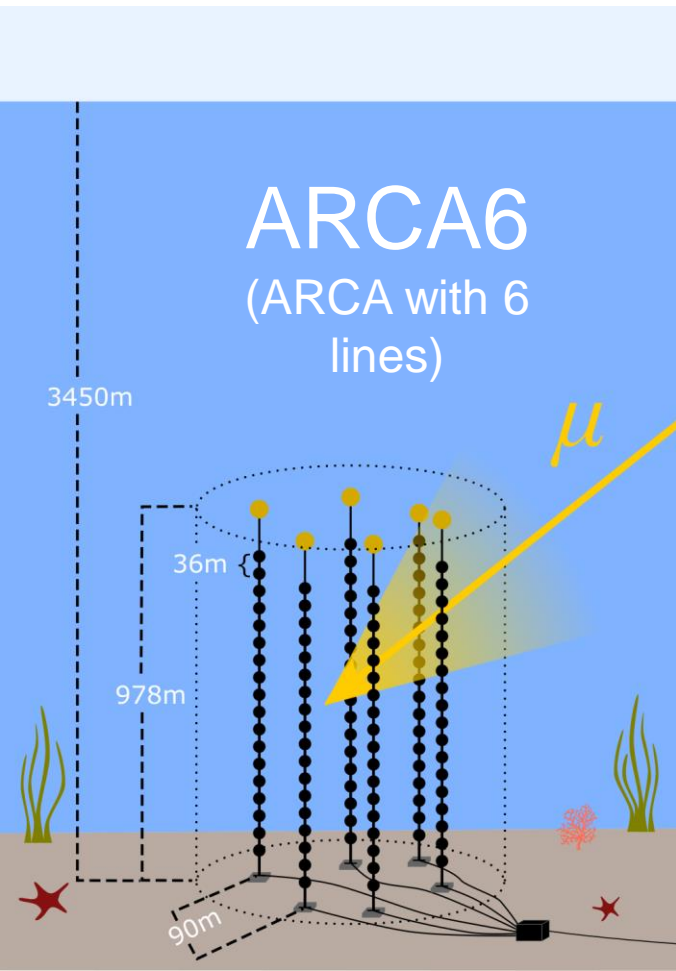
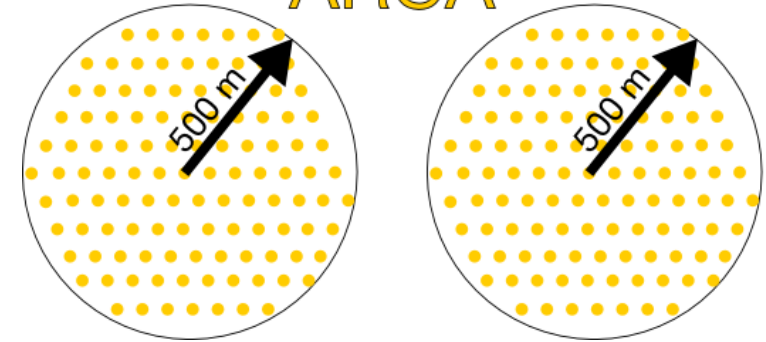
Work supported by:



ORCA



ARCA

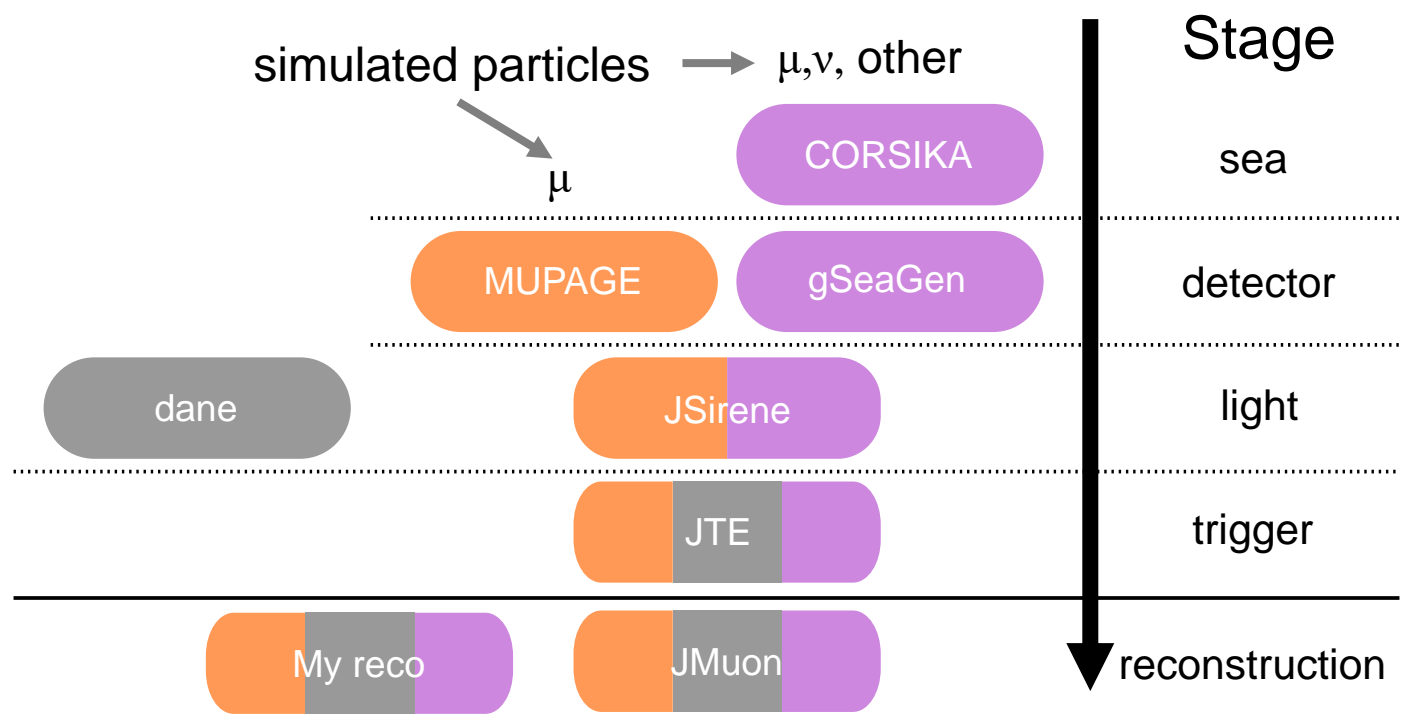


ARCA6 (ARCA with 6 lines)

Detector	ARCA	ORCA
Depth	3.5 km	2.45 km
Volume	1 km ³ (1Gton)	0.007 km ³ (7Mton)
# lines	28 / 2x115	16 / 115
Topic	Astroparticle RCA*	Oscillation RCA*
Goal	ν_{astro}	m_ν hierarchy

*RCA : Research with Cosmics in the Abyss

KM3NeT CR muon simulation chain:



My work:

- ❖ Running **CORSIKA**:
 - Test productions
 - Mini-prods for systematic studies
 - First mass prod
- ❖ **gSeaGen** development: git.km3net.de/opensource/gseagen
 - Implementation of **CORSIKA** processing
 - Improvements in the propagation routine:
 - Speed-up: for EeV 1w -> 40min
 - Earth's curvature
 - Statistics optimization: muon range tolerance margin, possibility to retry propagation etc.
 - More in the publication [under review]: "gSeaGen code by KM3NeT: an efficient tool to propagate muons simulated with CORSIKA" (to be submitted to: Comp. Phys. Comm.)
- ❖ Reconstruction of new observables

Here:

sea – sea surface above the KM3NeT detectors

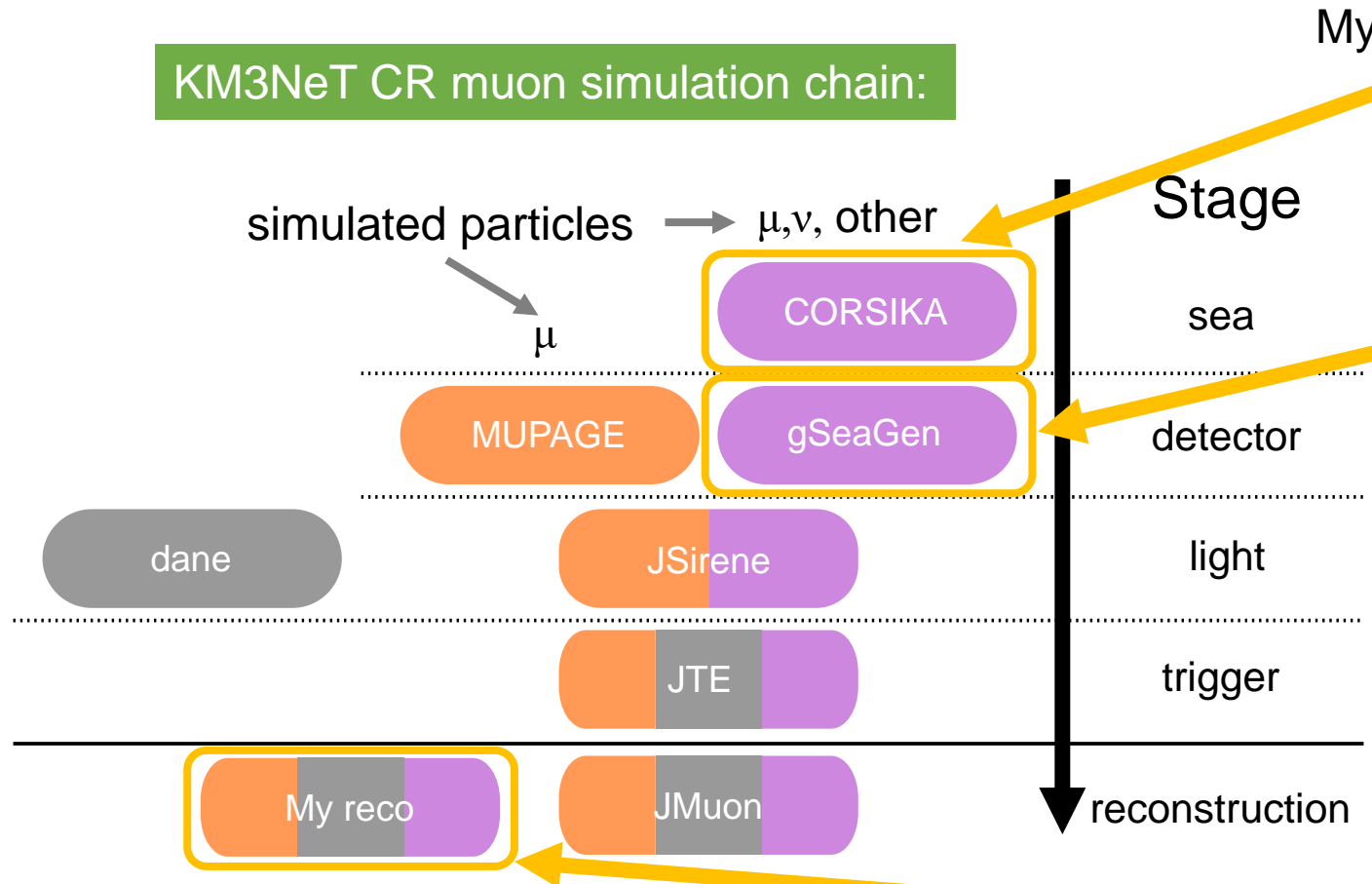
can – cylindrical volume around the detector

light – simulation of the γ emission inside the can (including the environmental bgd)

trigger – preselection of interesting events by applying trigger conditions

reconstruction – reconstruction of observables, like energy, direction etc.

KM3NeT CR muon simulation chain:



My work:

- ❖ Running **CORSIKA**:
 - Test productions
 - Mini-prods for systematic studies
 - First mass prod
- ❖ **gSeaGen** development: git.km3net.de/opensource/gseagen
 - Implementation of **CORSIKA** processing
 - Improvements in the propagation routine:
 - Speed-up: for EeV 1w -> 40min
 - Earth's curvature
 - Statistics optimization: muon range tolerance margin, possibility to retry propagation etc.
 - More in the publication [under review]: "gSeaGen code by KM3NeT: an efficient tool to propagate muons simulated with CORSIKA" (to be submitted to: Comp. Phys. Comm.)
- ❖ Reconstruction of new observables

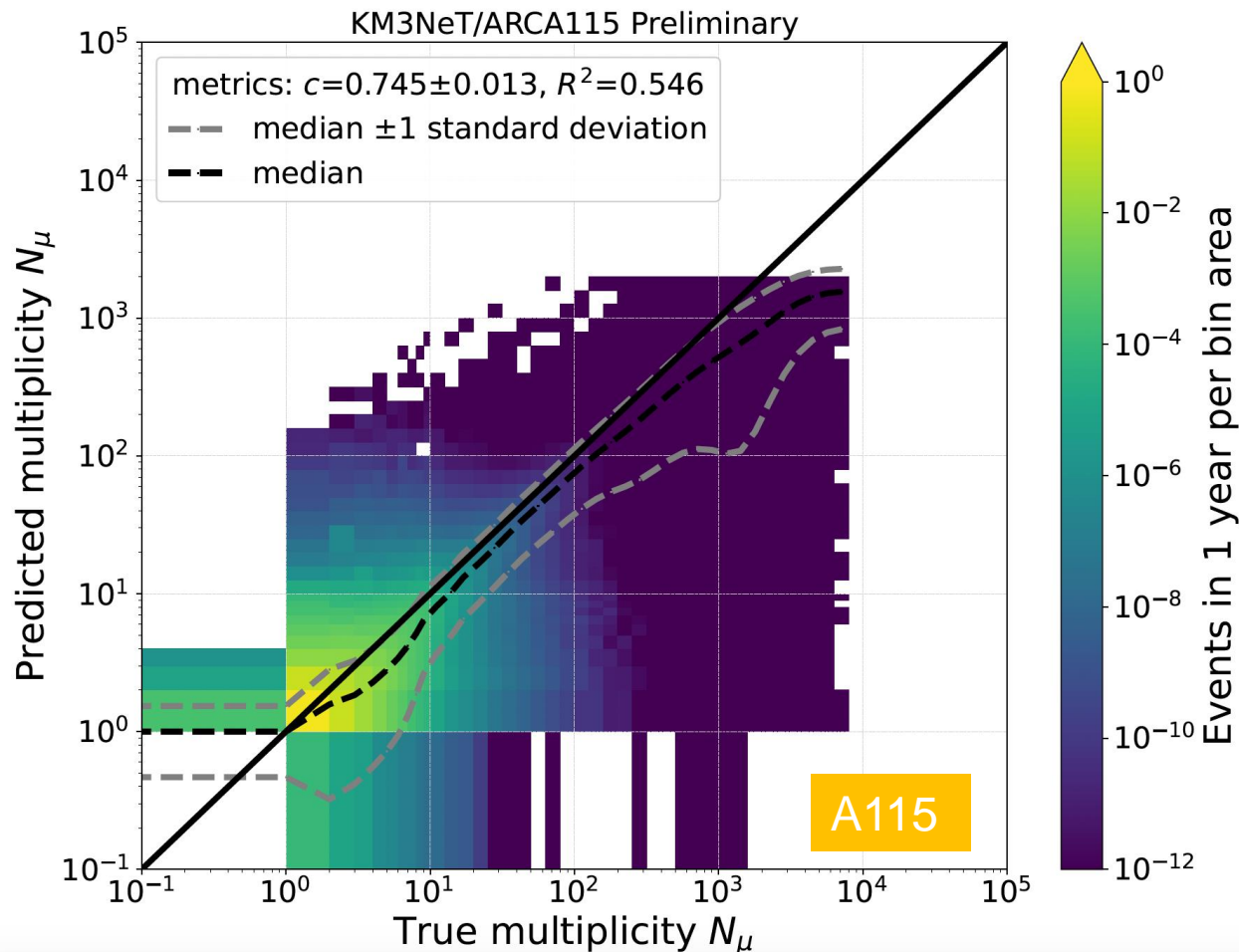
Here:

- sea** – sea surface above the KM3NeT detectors
- can** – cylindrical volume around the detector
- light** – simulation of the γ emission inside the can (including the environmental bgd)
- trigger** – preselection of interesting events by applying trigger conditions
- reconstruction** – reconstruction of observables, like energy, direction etc.

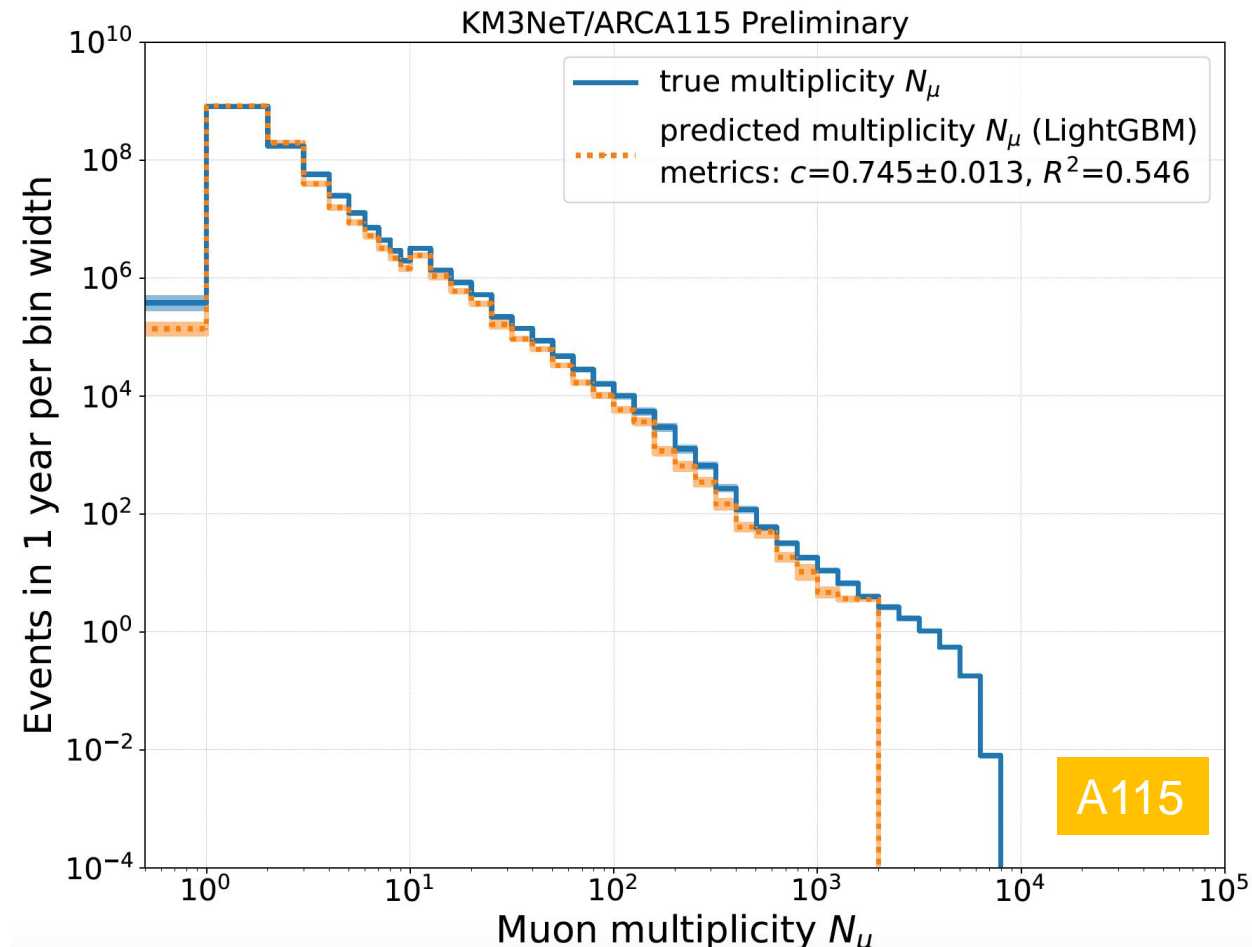
Example of reco results: muon multiplicity for ARCA115

Analogical results for ARCA6,
ORCA115 i ORCA6
(and for 2 other observables)

2D: reco vs true



1D histograms

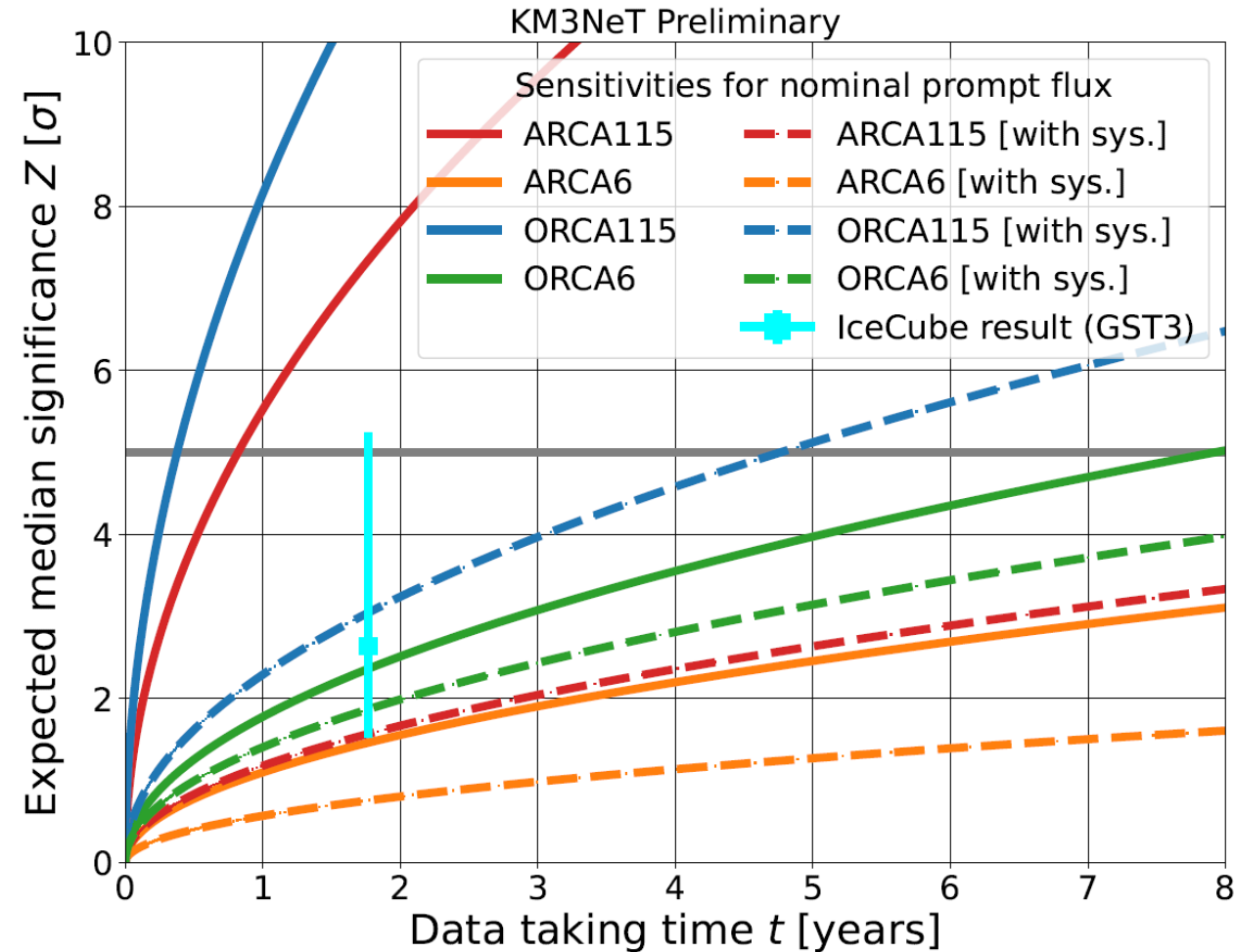
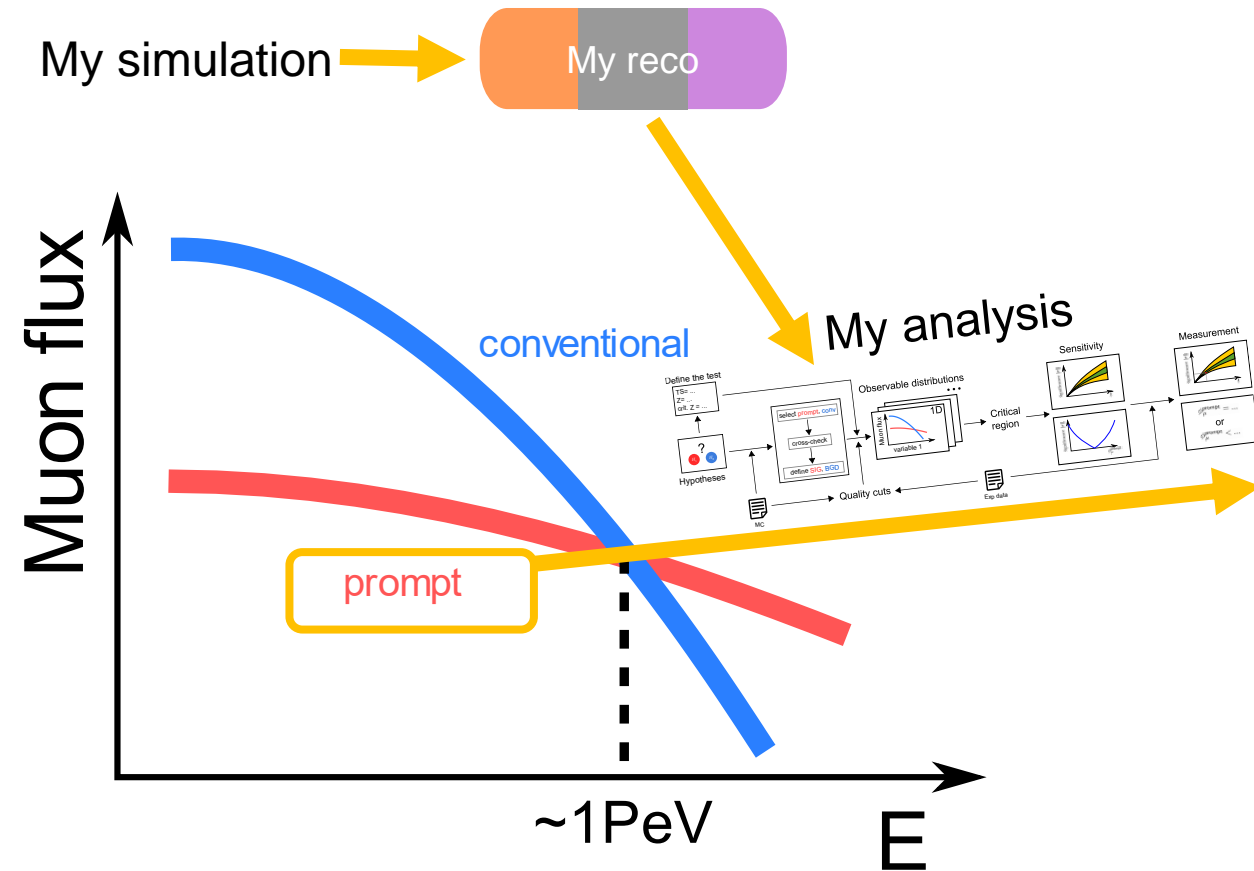


Sensitivities vs time

Comments:

- ❖ Sensitivity limited by systematics
- ❖ Should improve by the time we complete the construction
- ❖ IceCube uncertainty: various CR models

full analysis covered in my PhD thesis:
<https://bip.ncbj.gov.pl/attachments/download/258>





TL&DR:

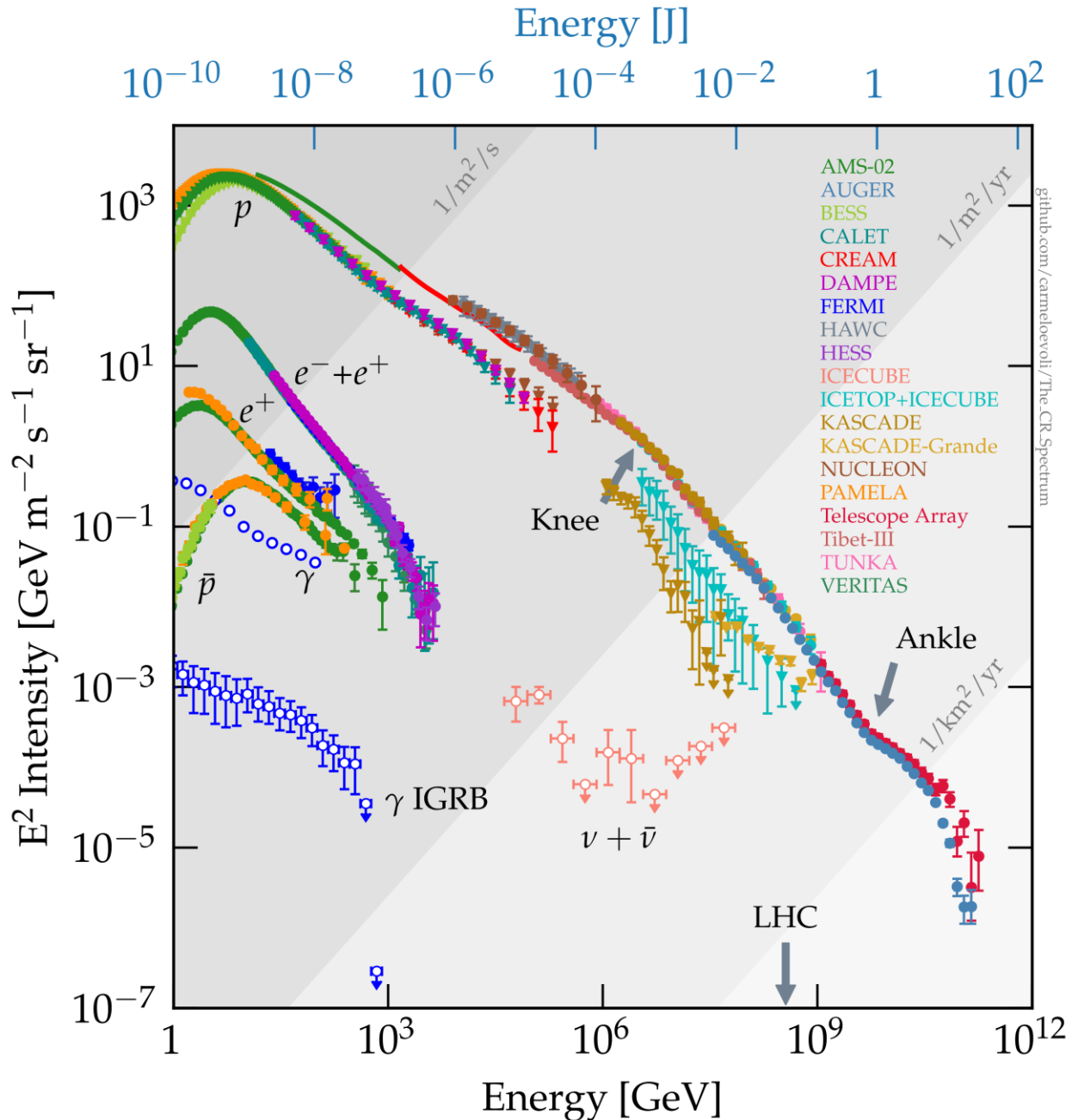
- ❖ KM3NeT detectors under construction, already collect valuable data
- ❖ Reliable muon simulations necessary
- ❖ Successful reco of N_μ , E_{bundle} , E_{prim}
- ❖ Complete ARCA & ORCA will confirm prompt muon flux in first few years of operation
- ❖ Stay tuned for more results! 😊



Thank you for your attention!



Backup



Promieniowanie kosmiczne (CR):

Wysokoenergetyczne cząstki i jądra atomowe docierające do Ziemi z kosmosu

Odkryte przez Victora Hessa w 1912
(nagroda Nobla w 1936)

Od tego czasu pomiarów było całkiem sporo ...

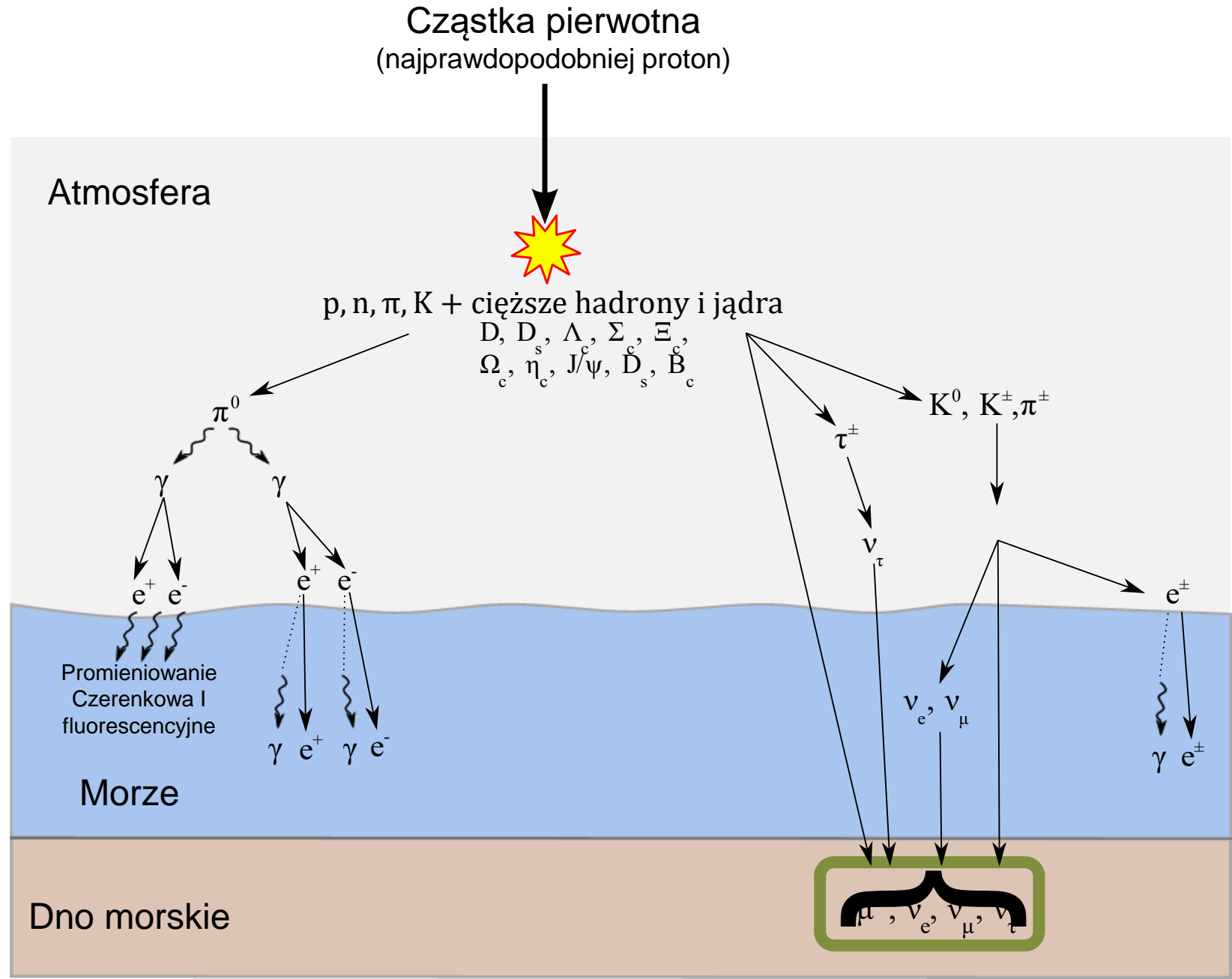
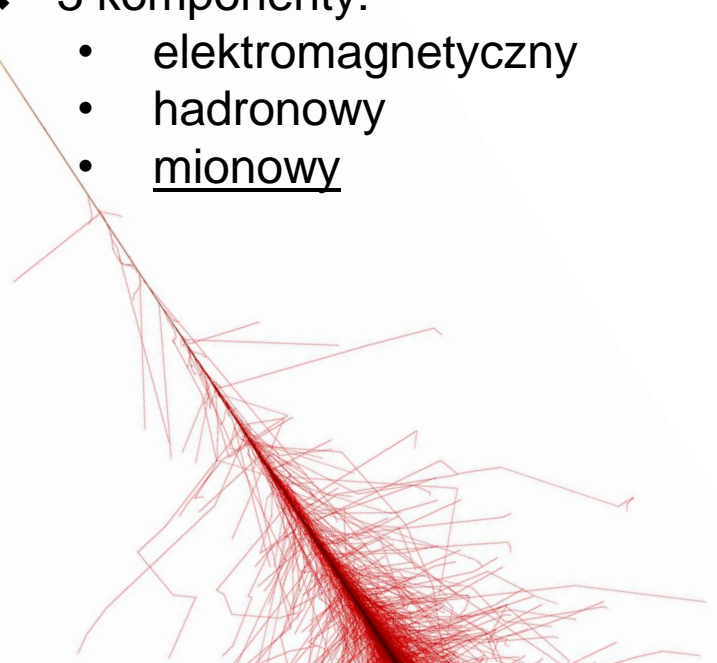
Ale zaraz, na tym wykresie nie ma mionów ... !?

Miony nie są częścią pierwotnego promieniowania kosmicznego!

Są cząstkami wtórnymi produkowanymi w ...

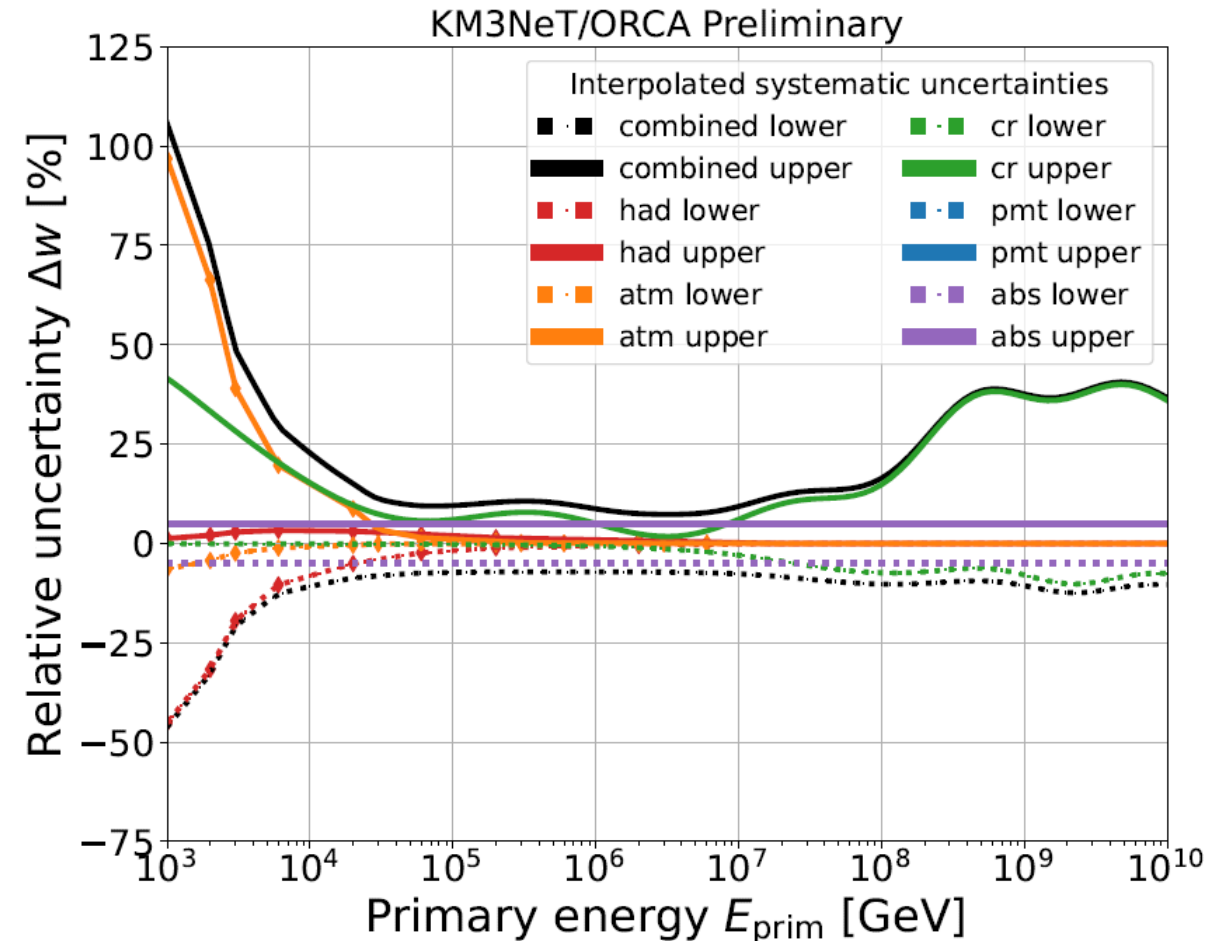
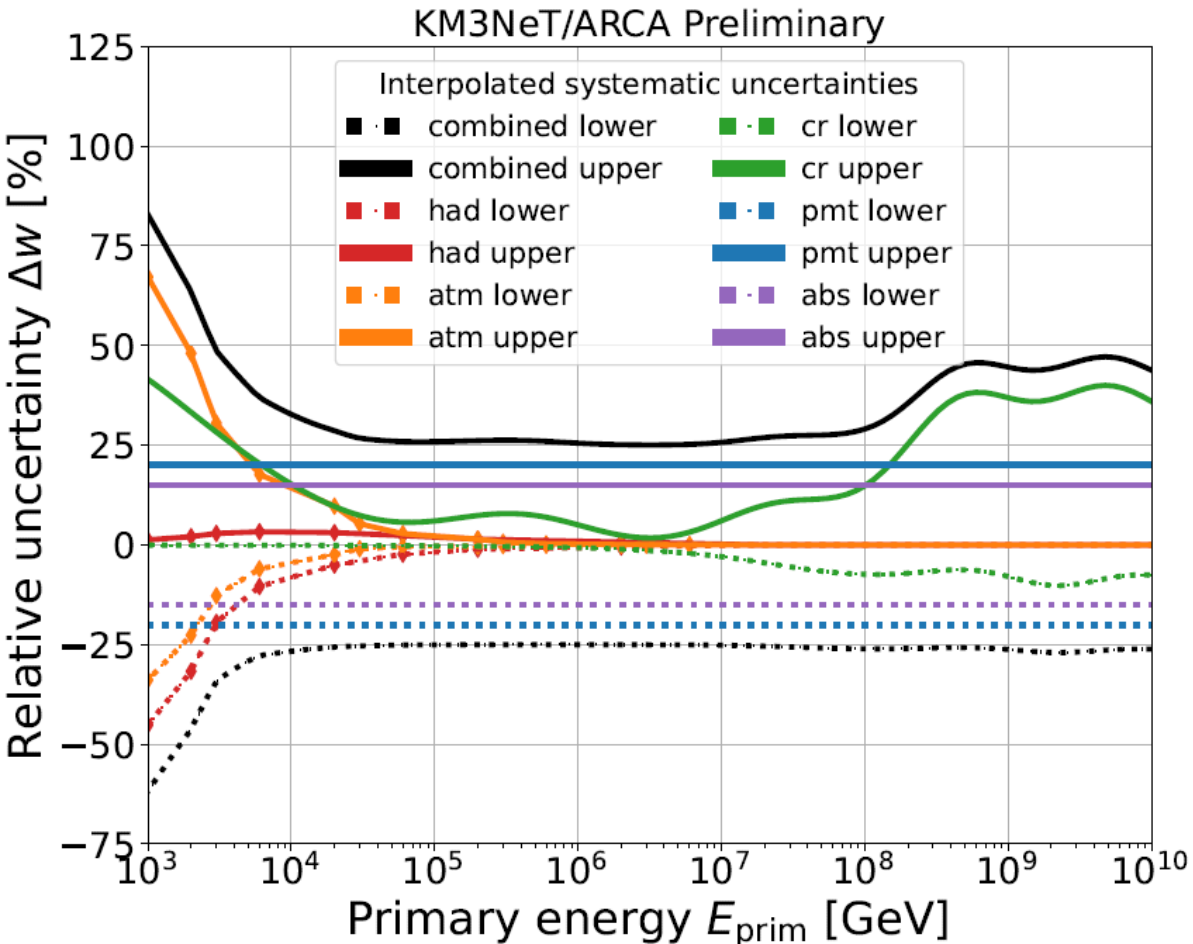
Wielkich pękach atm. (EAS):

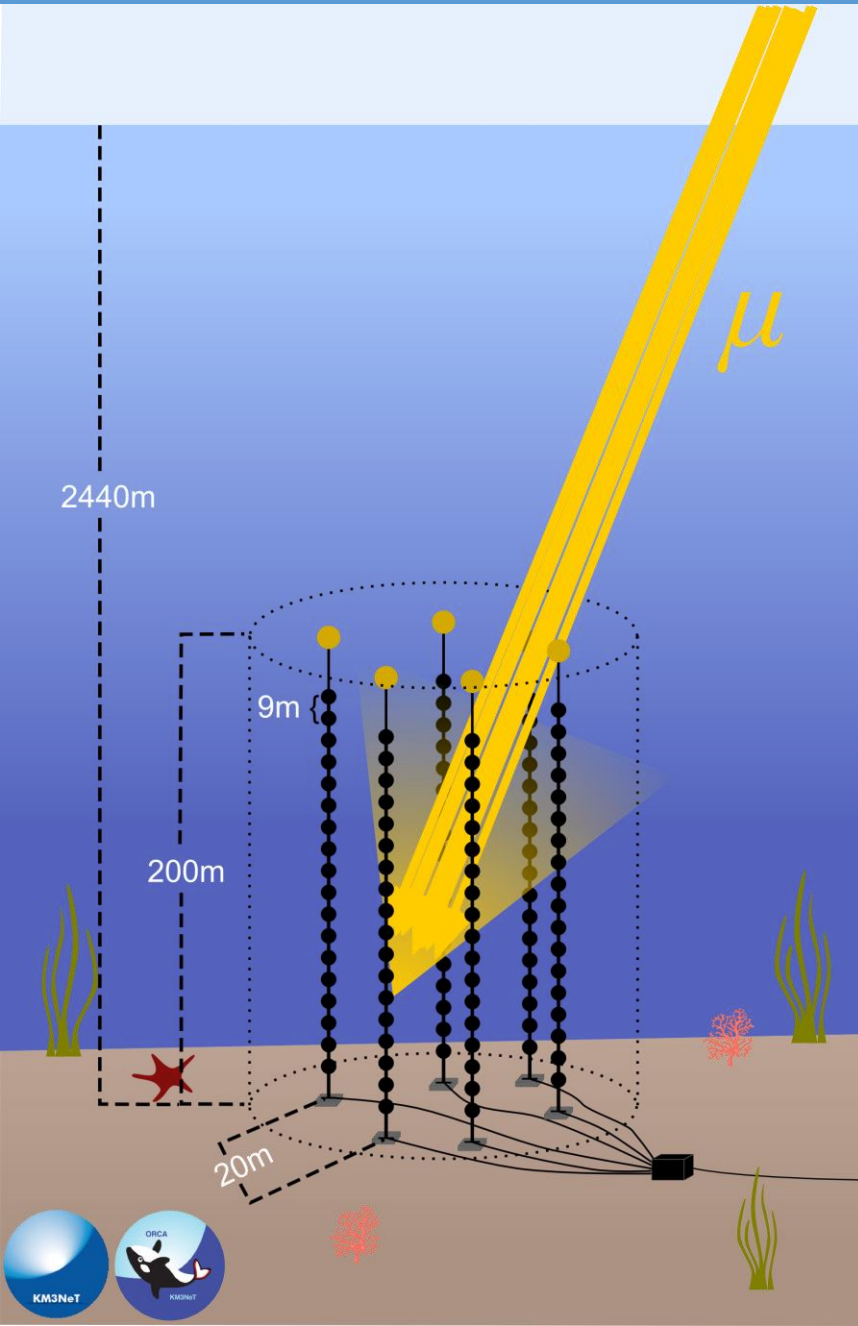
- ❖ Powodują je cząstki pierwotne CR
- ❖ Większość zaczyna się na wysokości $h \sim 30 - 40 \text{ km}$
- ❖ 3 komponenty:
 - elektromagnetyczny
 - hadronowy
 - mionowy



Indywidualne i połączone wkłady do niepewności systematycznej:

- ❖ Jeżeli są punkty wzdłuż linii: interpolacja
- ❖ Zakres energii odpowiada naszej symulacji CORSIKA





Pęk mionów: **Miony pochodzące z tego samego EAS**

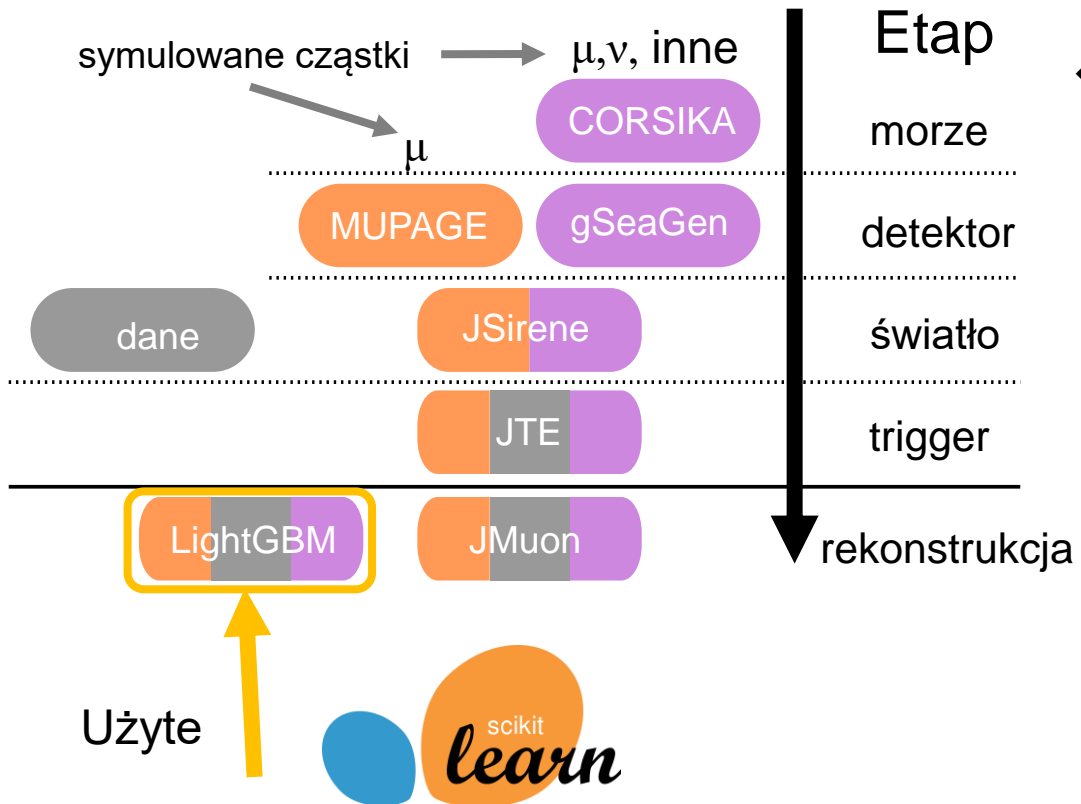
Kilka istotnych własności:

Zmienna	Opis	Standardowa rekonstrukcja?
$\cos \theta_{\text{zenith}} = \frac{\sum \cos \theta_{\text{zenith}}}{N_{\mu}}$	Kierunek (zenit)	dobra
$E_{\text{bundle}} = \sum E_{\mu}$	Energia pęku	raczej zła (skupia się tylko na przypadkach z 1 mionem)
E_{primary}	Energia cząstki pierwotnej	brak
N_{μ}	Krotność mionów w pęku	brak

wymagały nowych narzędzi do rekonstrukcji



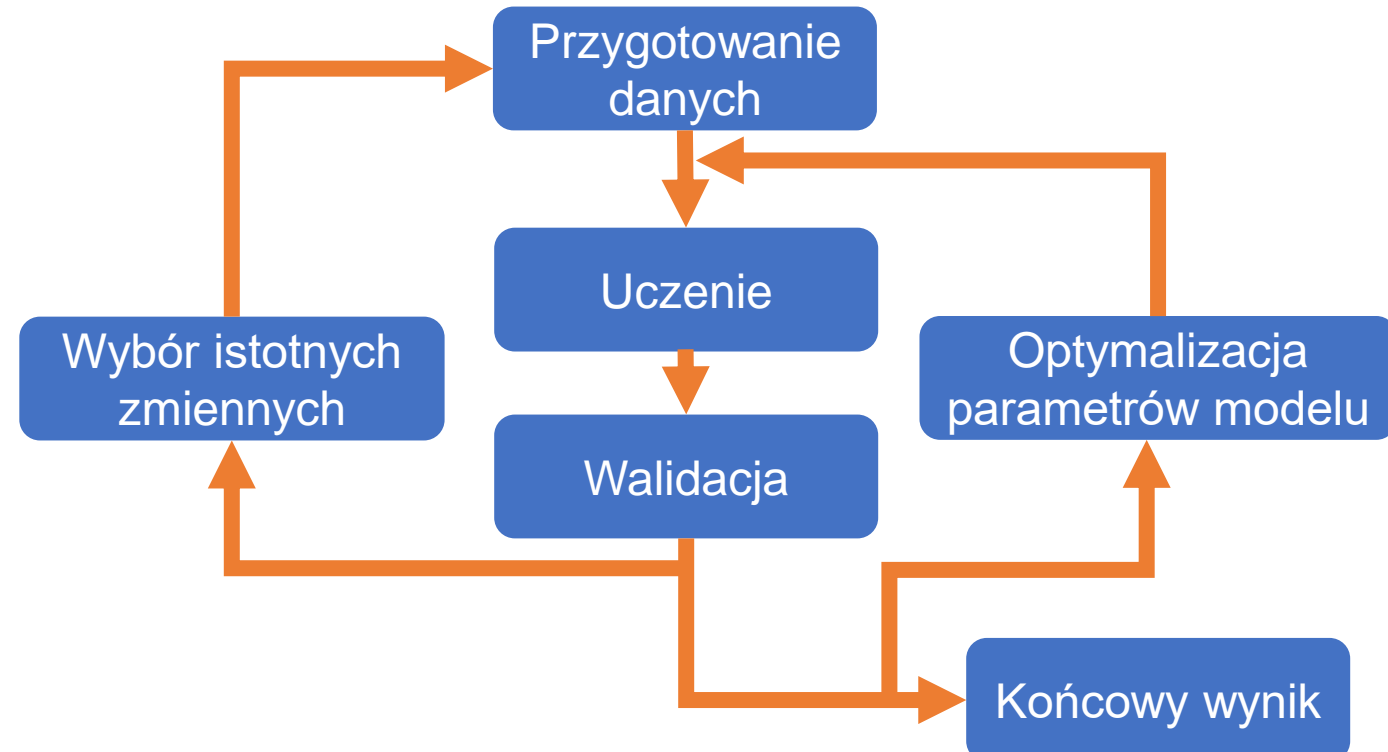
Schemat symulacji mionowych



Moja rekonstrukcja:

- ❖ Dane uczące: moja symulacja CORSIKA (3-15M przypadków w zbiorze treningowym)
- ❖ Użyty model: LightGBM (gradient boosting machine Microsoft-u) wybrany spośród wielu rozważanych (więcej w drugiej prezentacji)

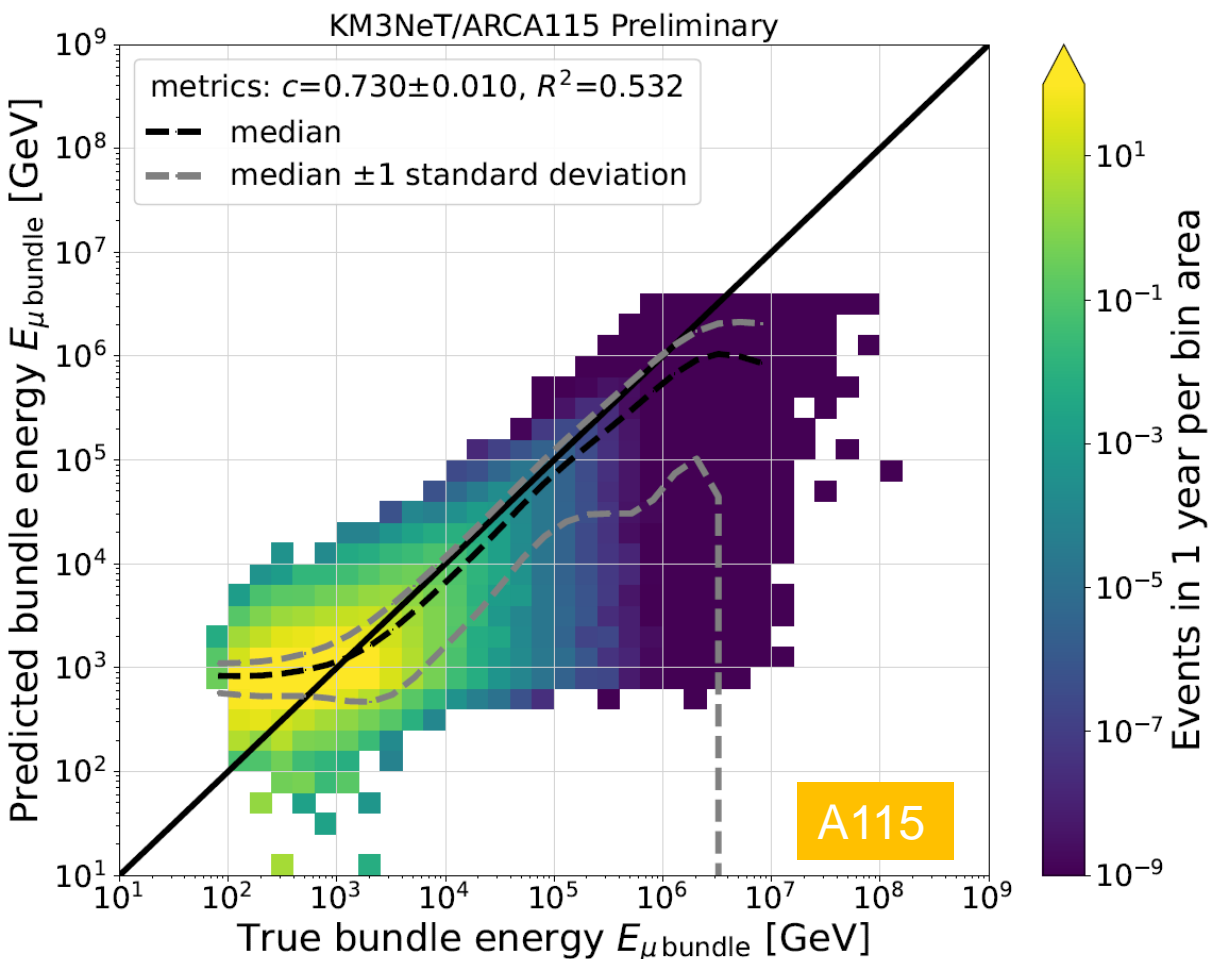
Procedura przygotowania modelu:



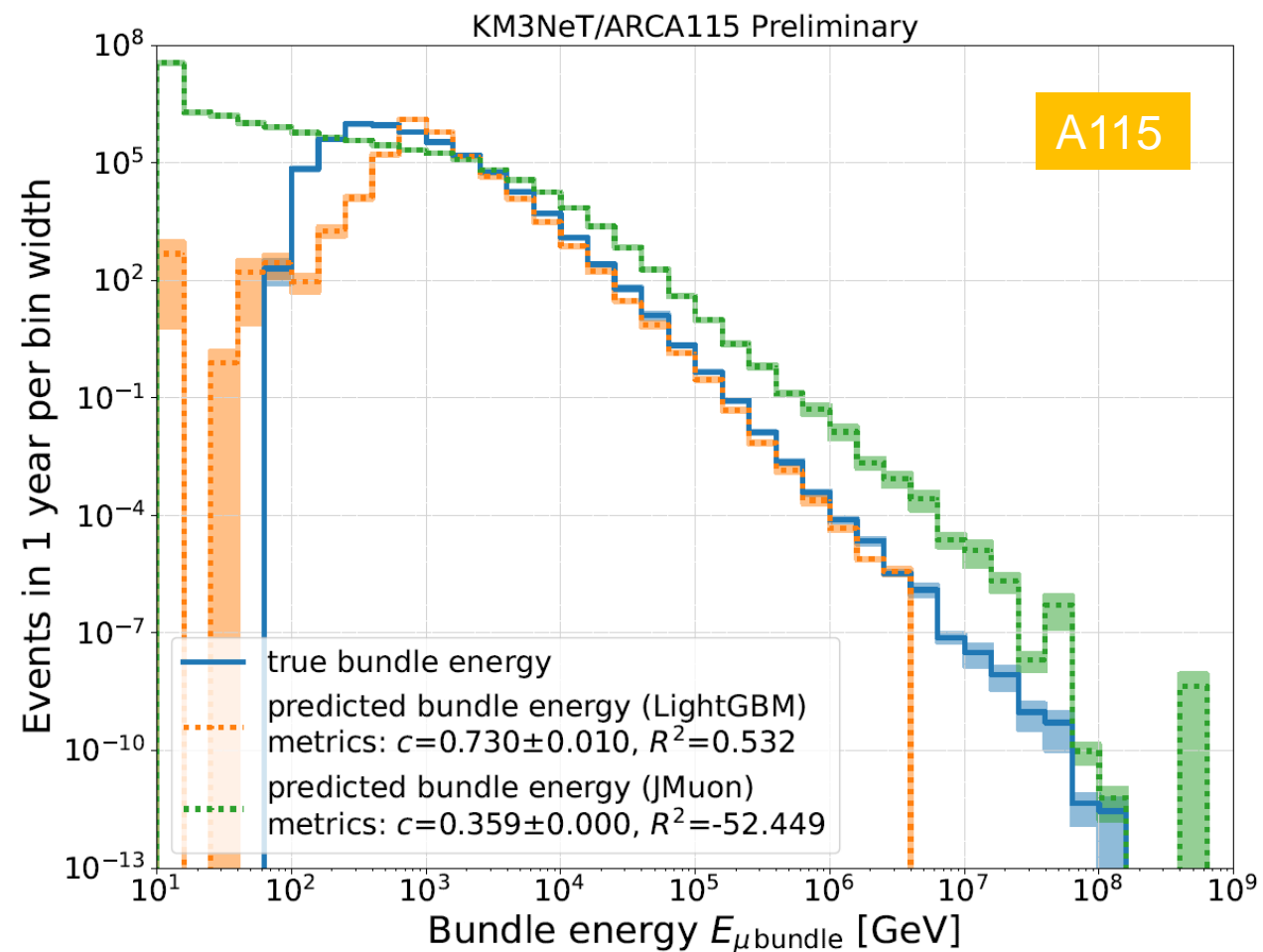
Analogiczne wyniki dla ARCA6, ORCA115 i ORCA6

Przykład wyników dla ARCA115:

2D: rekonstrukcja vs prawda



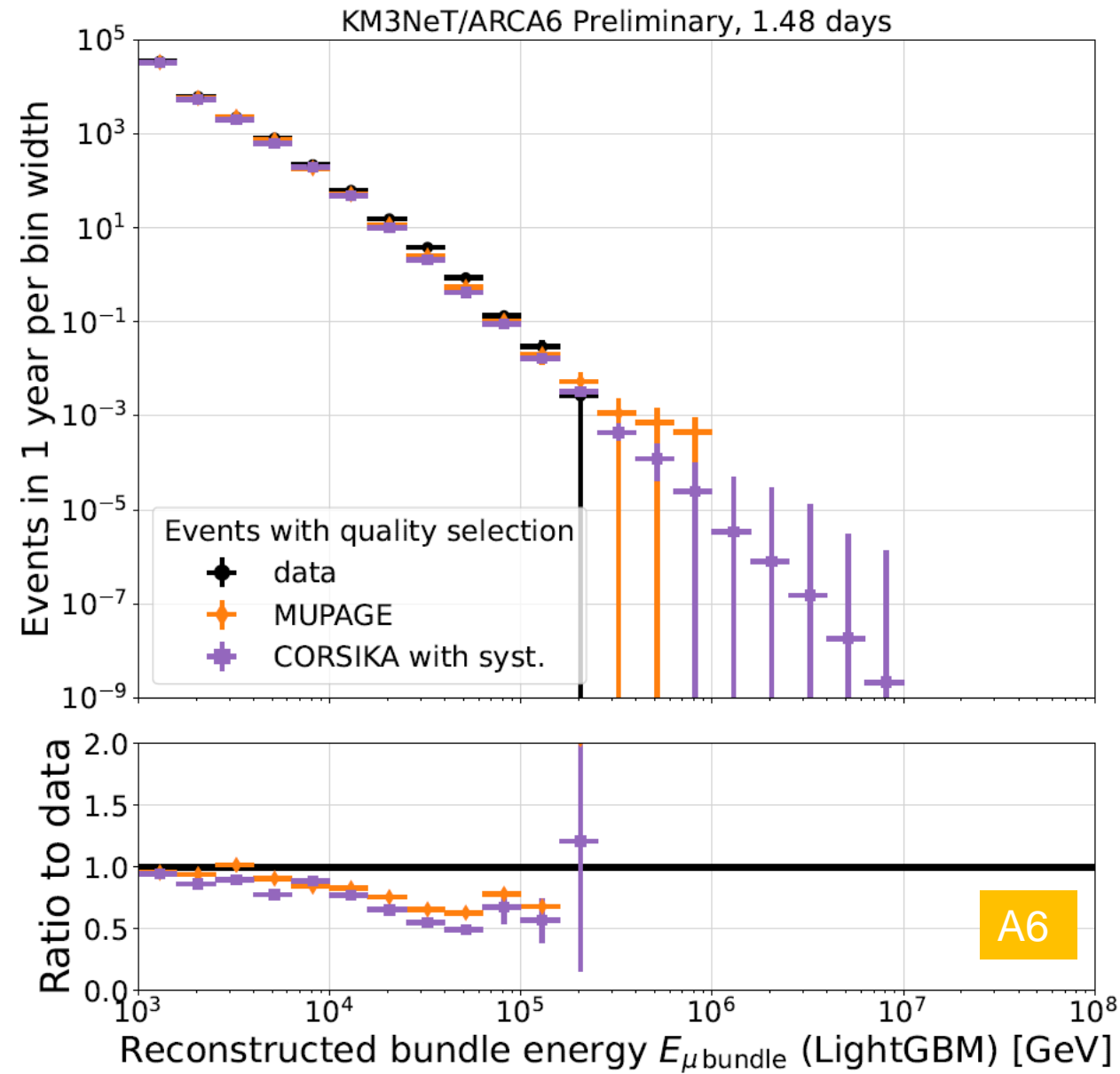
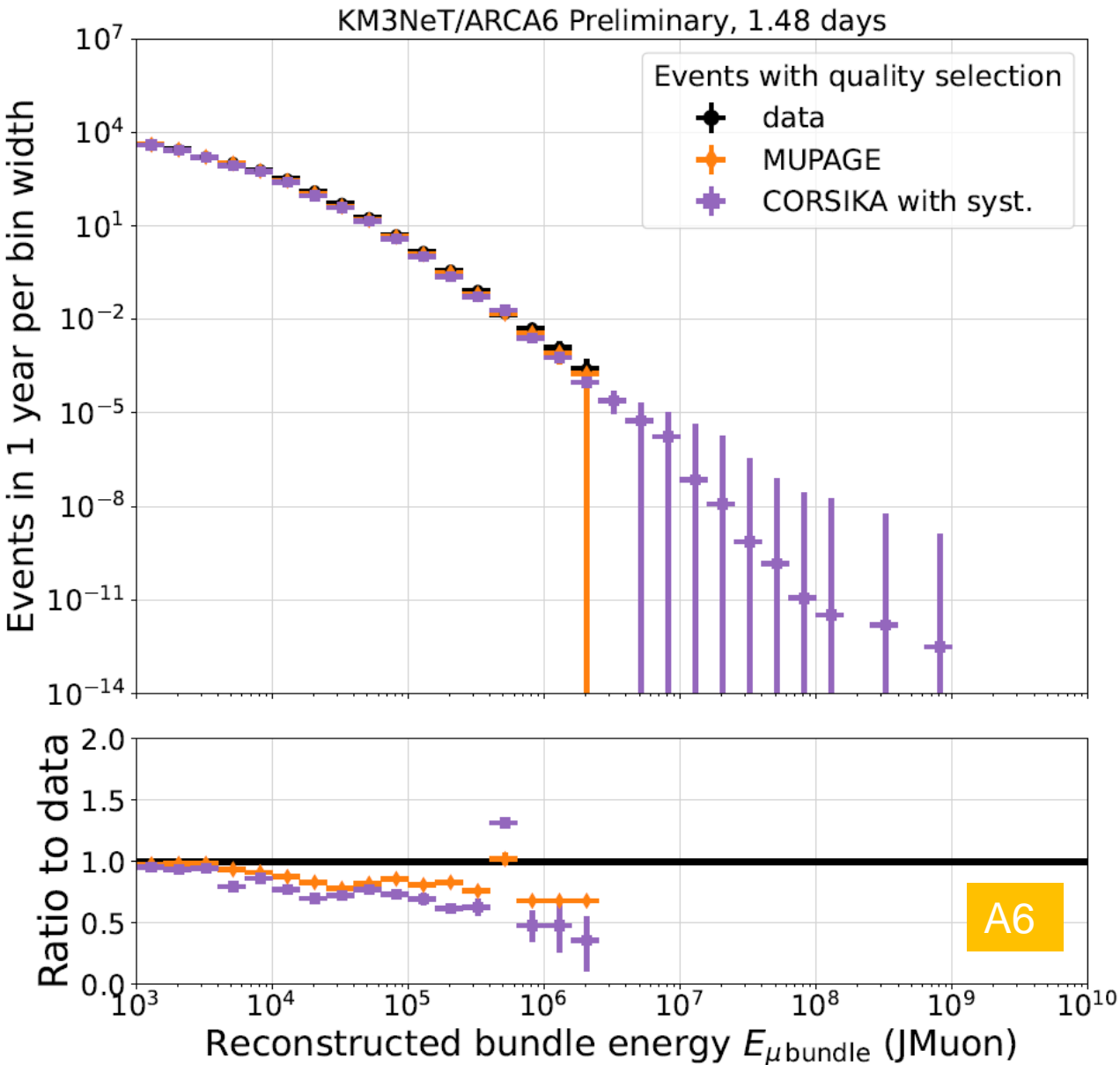
histogramy 1D



Jmuon (standardowa reko)

Analogiczne wyniki dla ORCA6

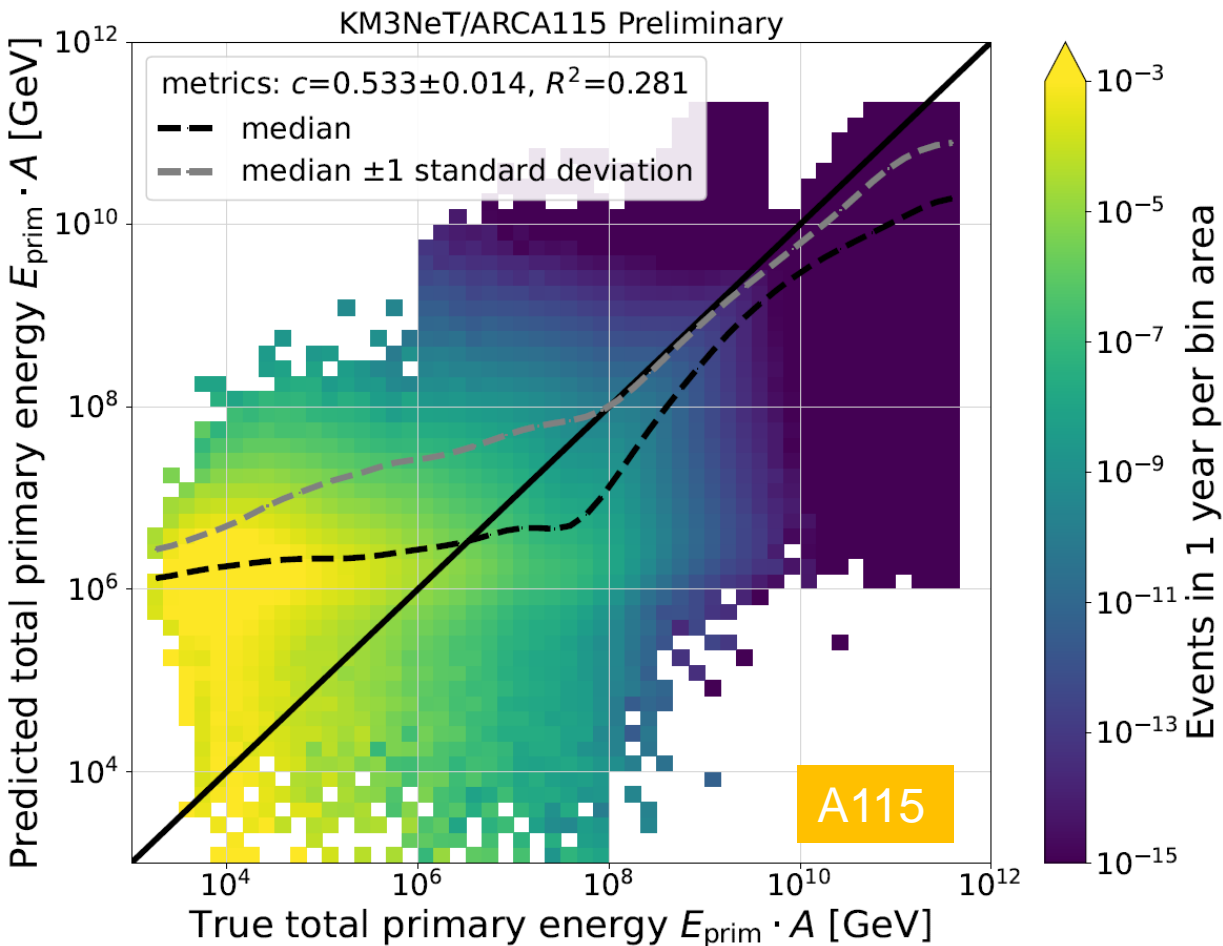
LightGBM



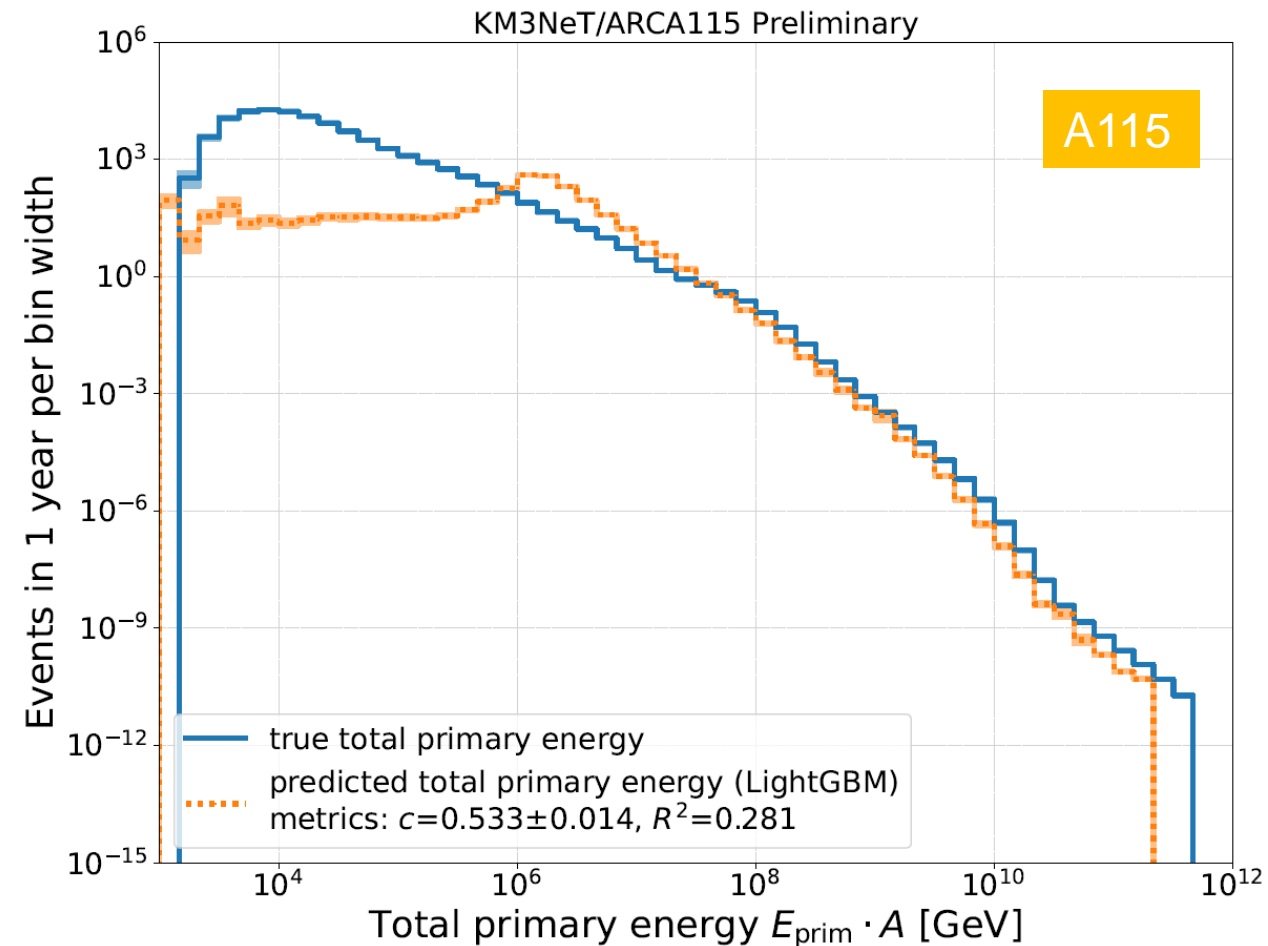
Analogiczne wyniki dla ARCA6, ORCA115 i ORCA6

Przykład wyników dla ARCA115:

2D: rekonstrukcja vs prawda

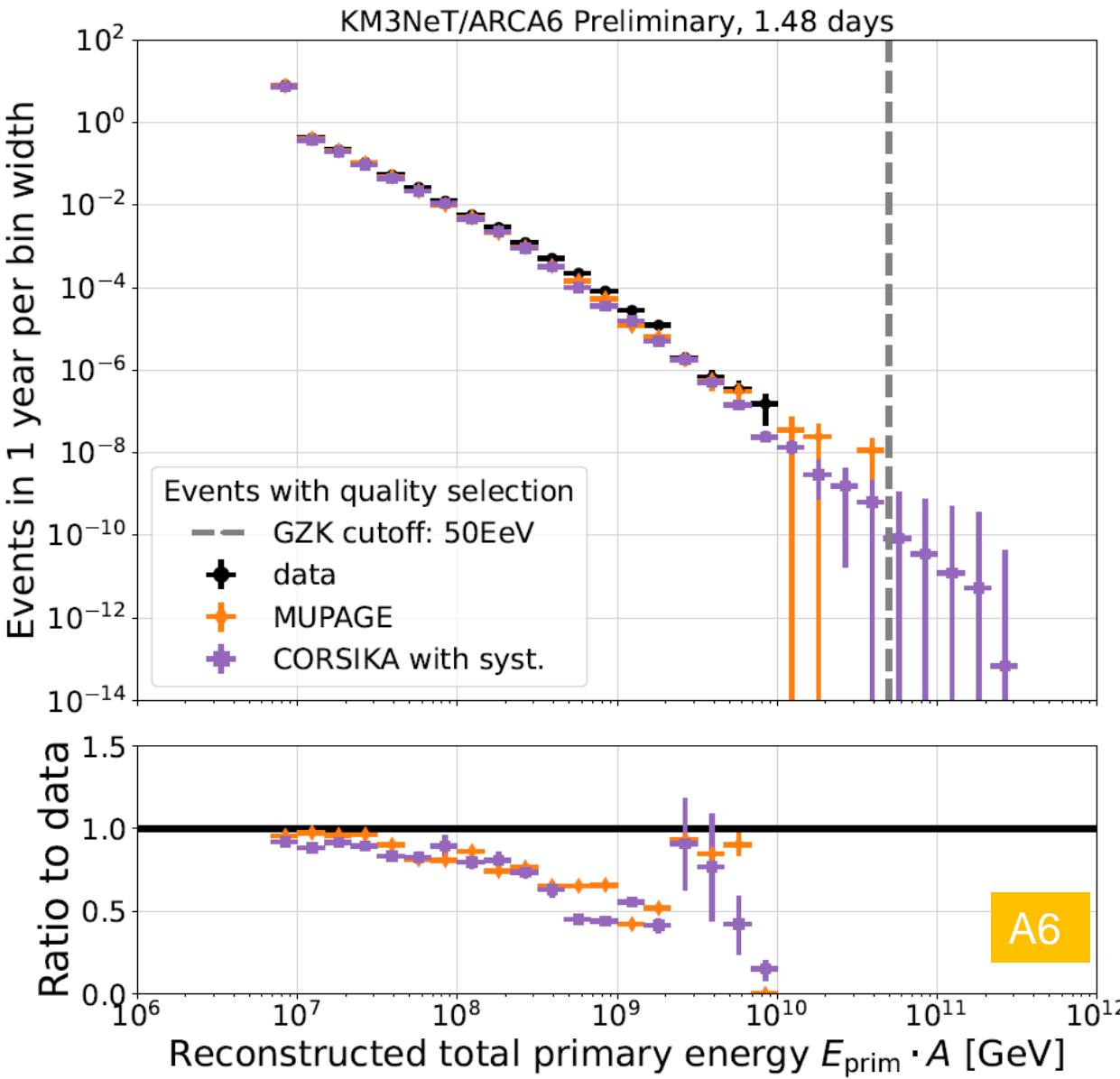


histogramy 1D



LightGBM

Analogiczne wyniki dla ORCA6



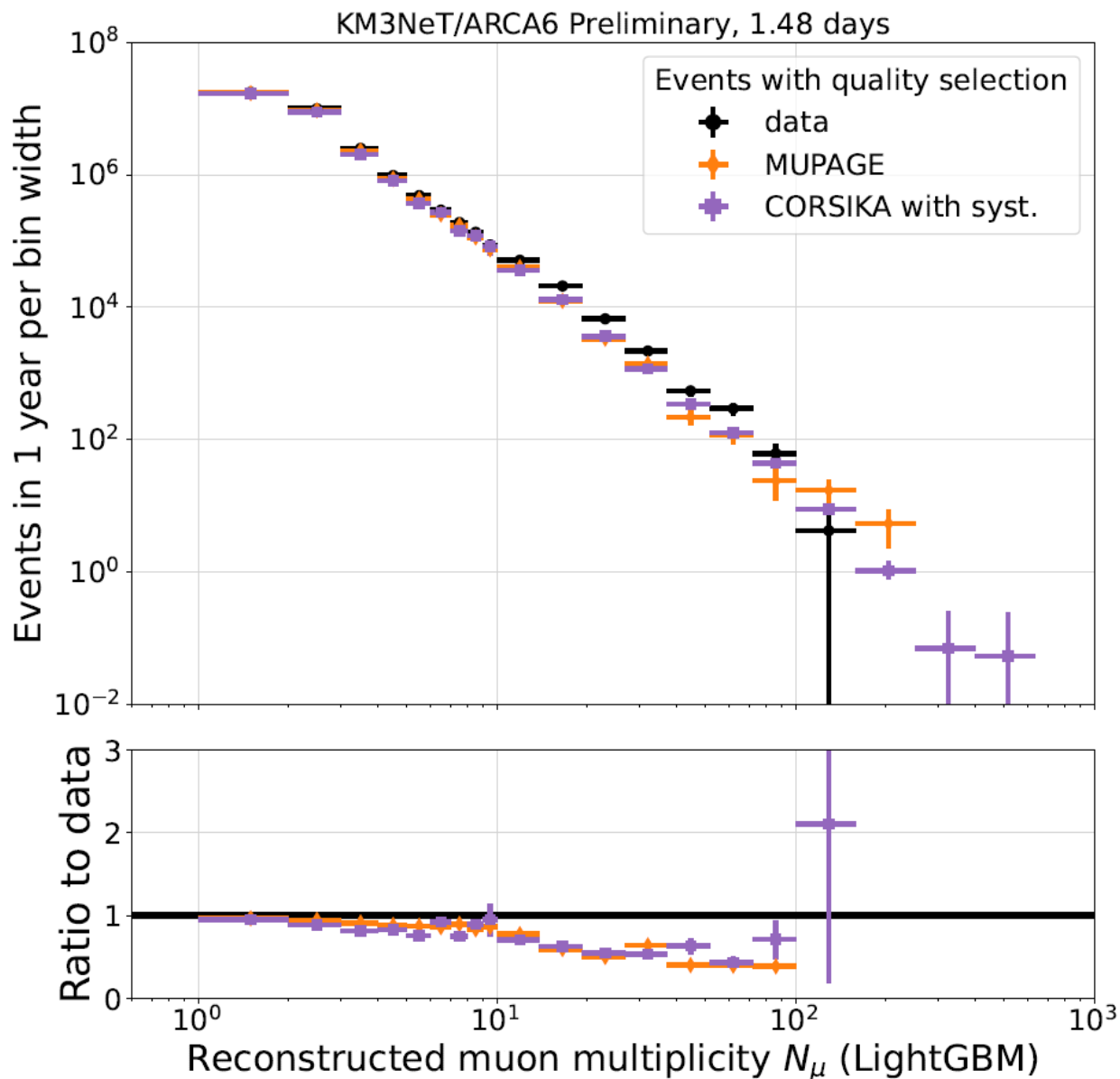
Wyniki uzyskane z danych ARCA6/ORCA6:

- ✦ Pierwszy pomiar energii cząstki pierwotnej detektorami KM3NeT
- ✦ Wkład wysokich energii niedoszacowany przez symulacje (analogicznie jak przy energii pęku, czego należało się spodziewać)
- ✦ Na tą chwilę nie można nic stwierdzić o granicy Greisena–Zatsepina–Kuzmina (GZK)

*granica GZK – teoretyczne górne ograniczenie na możliwe energie promieniowania kosmicznego, związane z oddziaływaniami z kosmicznym mikrofalowym promieniowaniem tła (CMB)

LightGBM

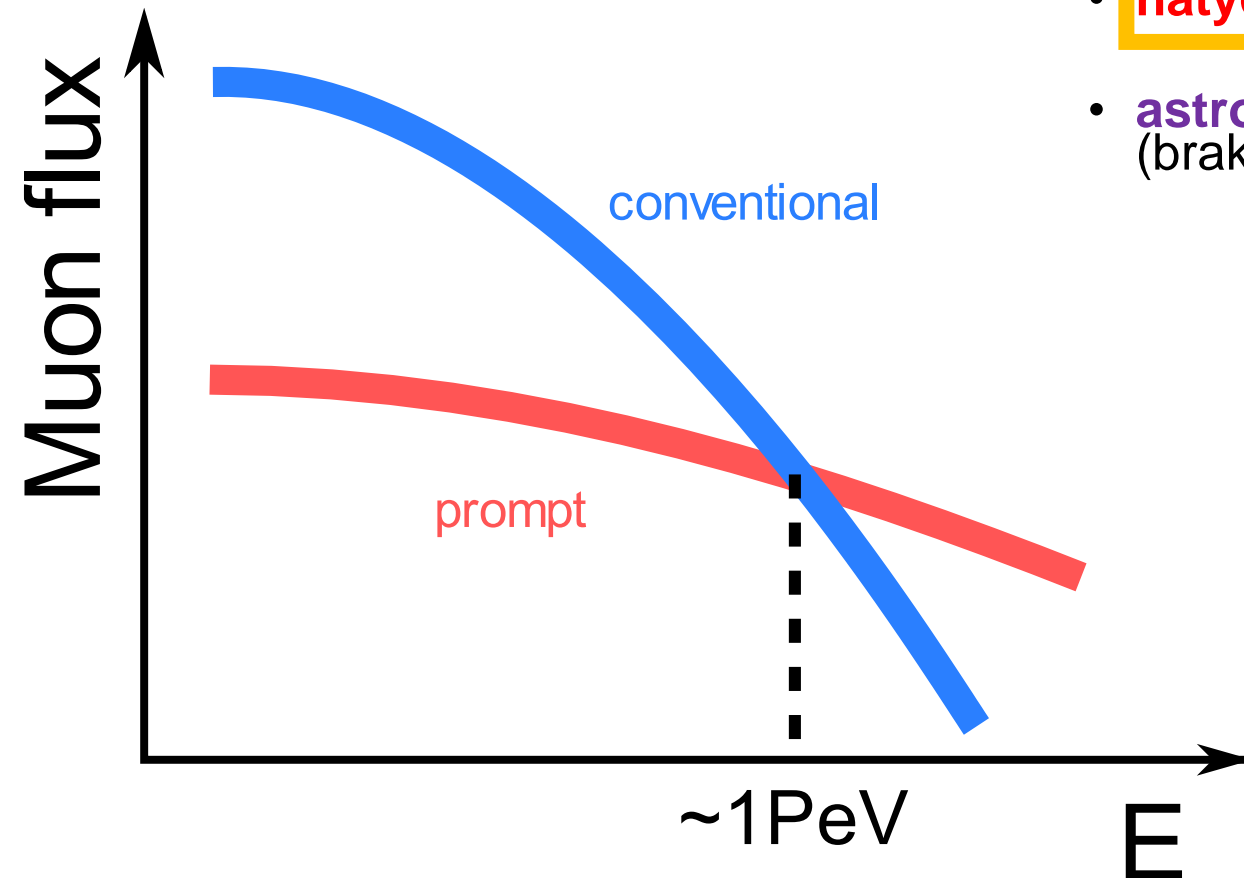
Analogiczne wyniki dla ORCA6



- Wyniki uzyskane z danych ARCA6/ORCA6:
- ❖ Pierwszy pomiar krotności mionów detektorami KM3NeT:
 - Ex aequo ze S. Reckiem używającym danych ORCA4
 - Pierwszy w ogóle dla ARKI
 - ❖ Wysokie krotności również niedoszacowane w symulacji

Kategorie strumienia cząstek używane powszechnie przez teleskopy neutrinowe:

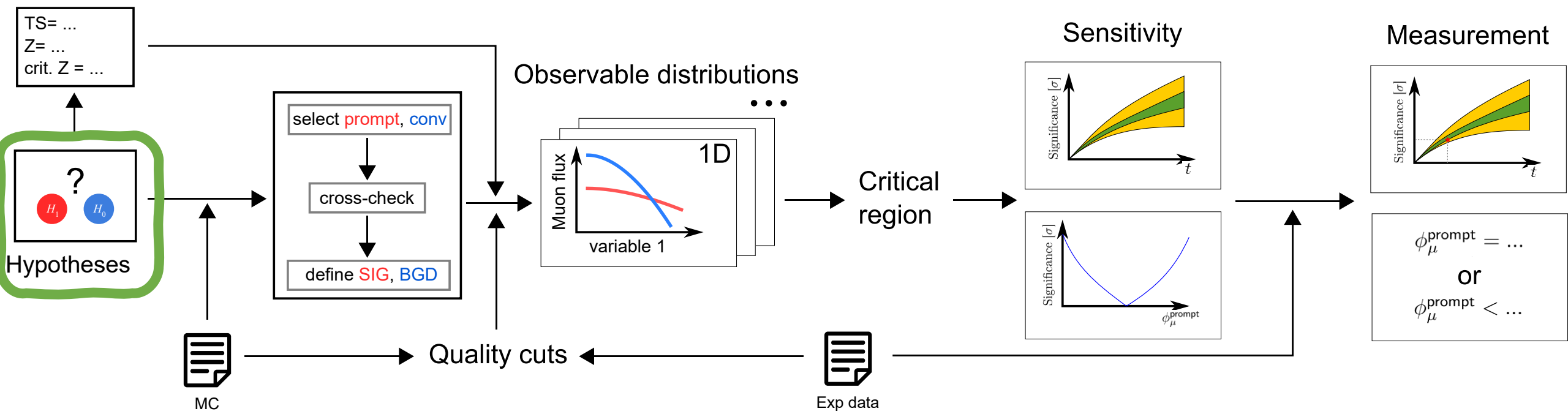
- **konwencjonalny** – μ , ν głównie z rozpadów π i K
- **natychmiastowy (prompt)** – μ , ν głównie z rozpadów ciężkich hadronów
- **astrofizyczny** – ν z AGN-ów, SN-y, itd. (brak astro μ , **promptowe** μ , ν są tłem)



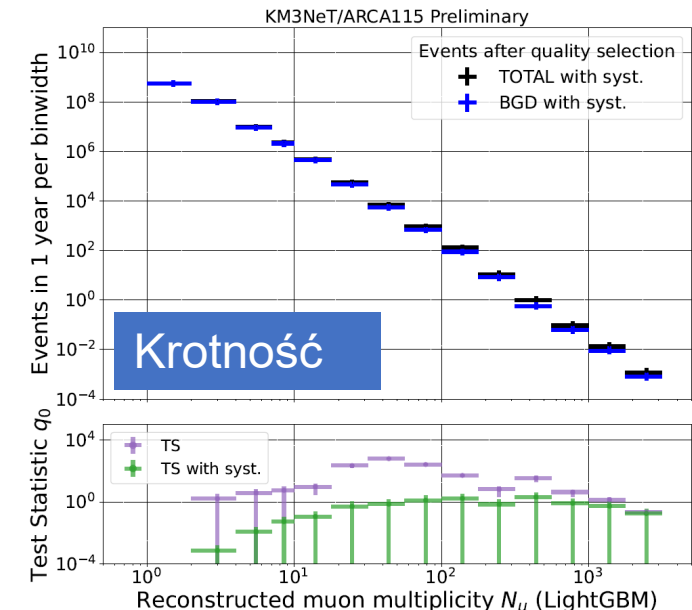
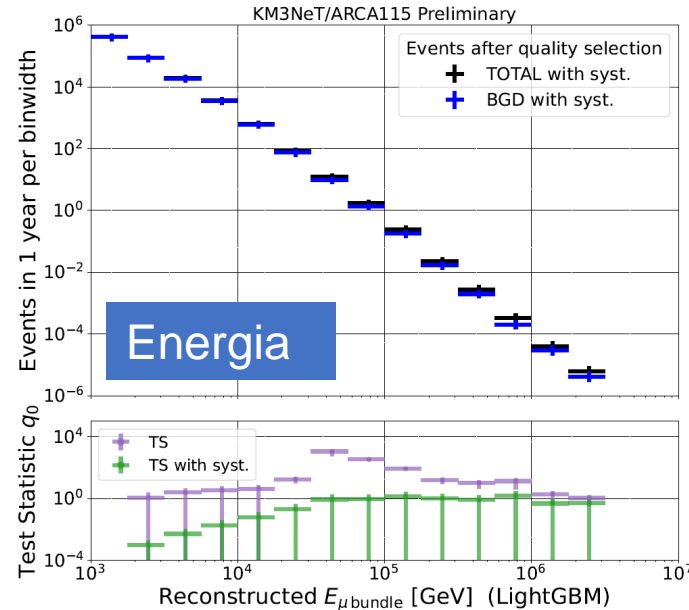
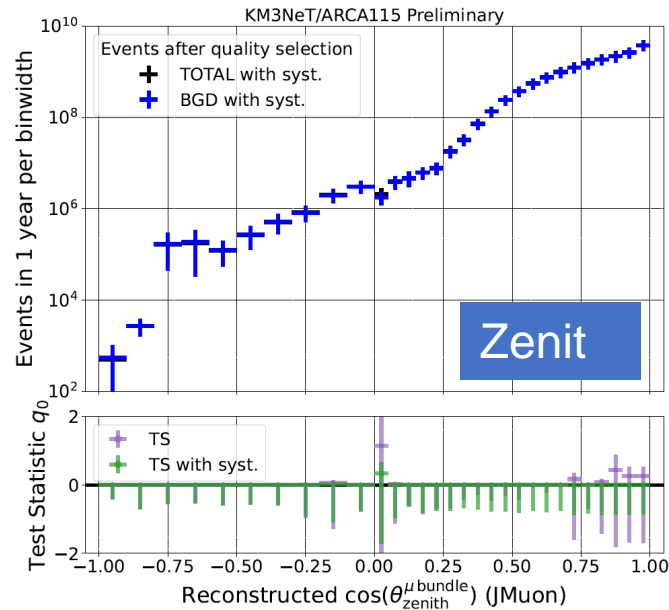
Do tej pory niepotwierdzony!

- H_0
Strumień mionów składa się w całości z mionów **konwencjonalnych**.
→
BGD
- H_1
Strumień mionów zawiera zarówno wkład **promptowy** jak i **konwencjonalny**.
→
TOTAL

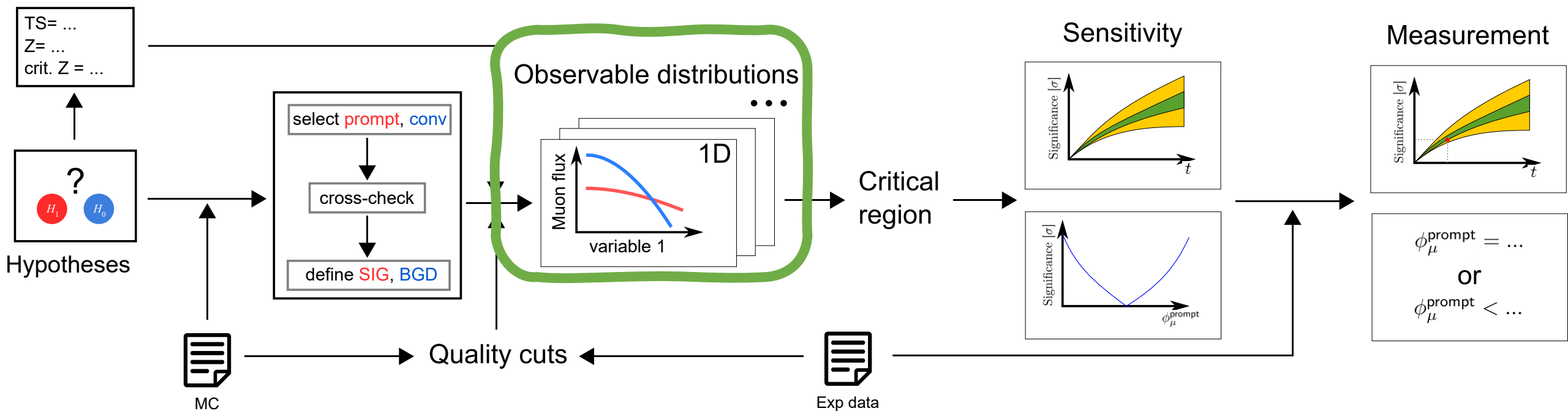
Define the test

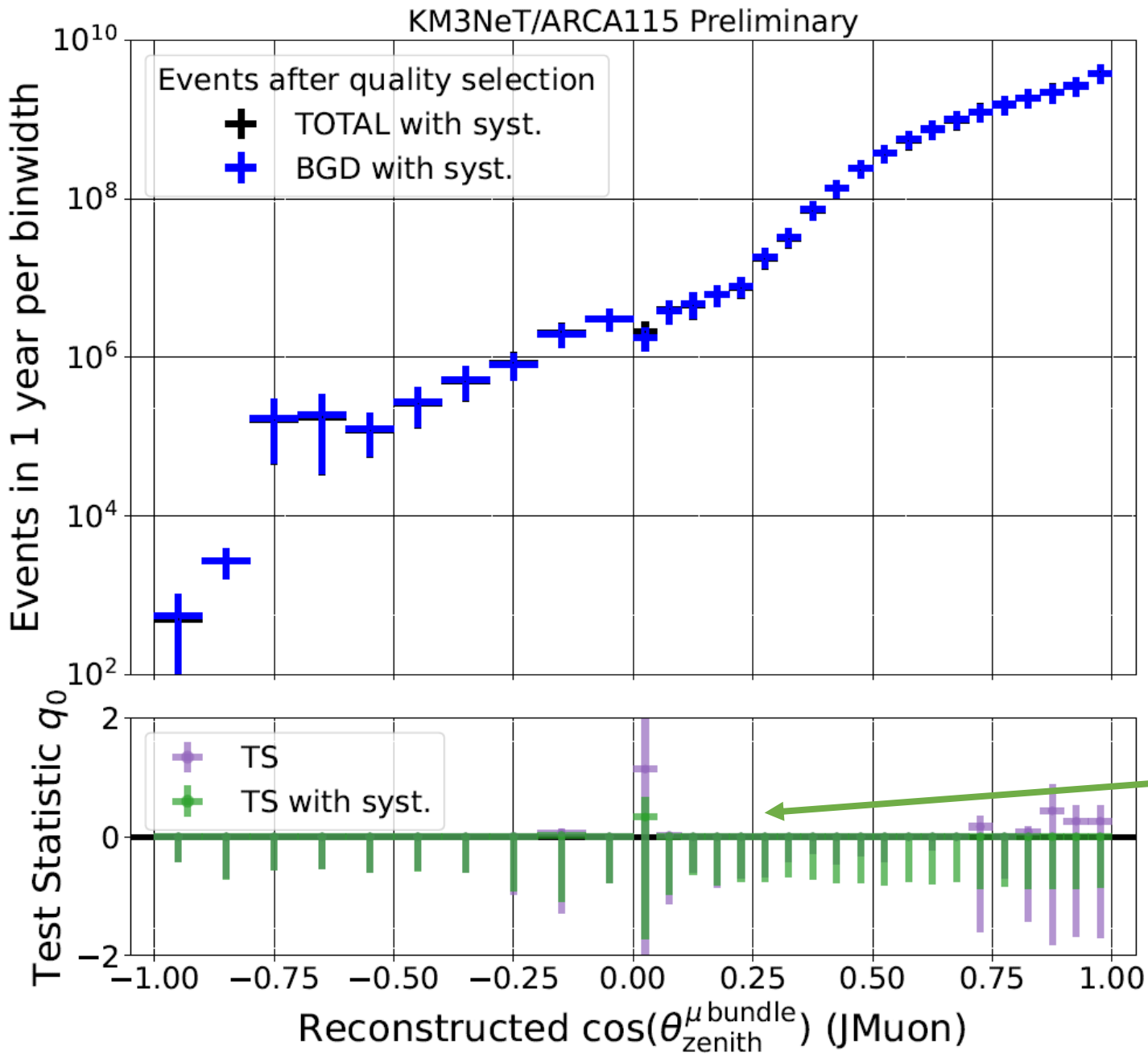


Rozkłady obserwabli

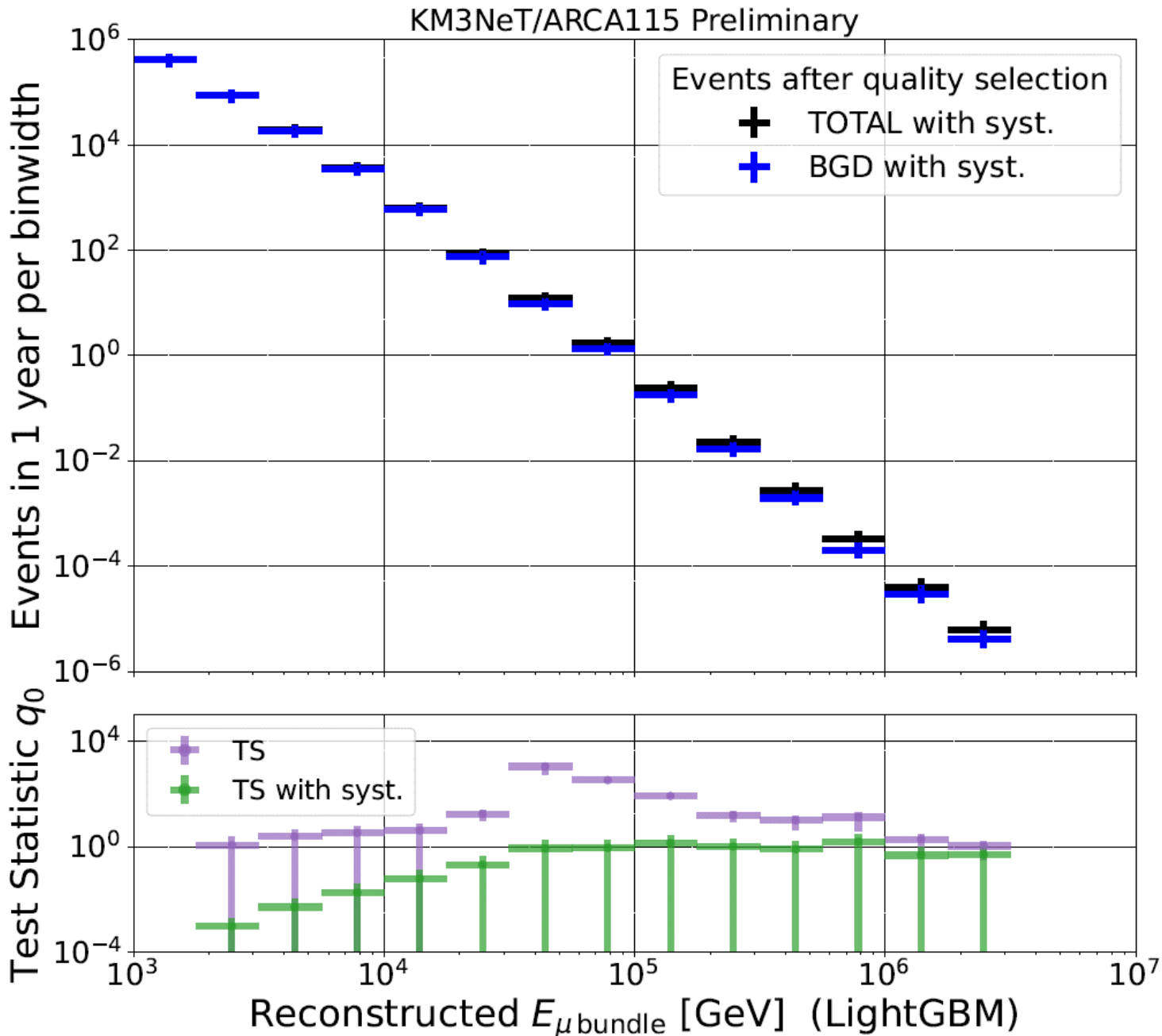


Define the test



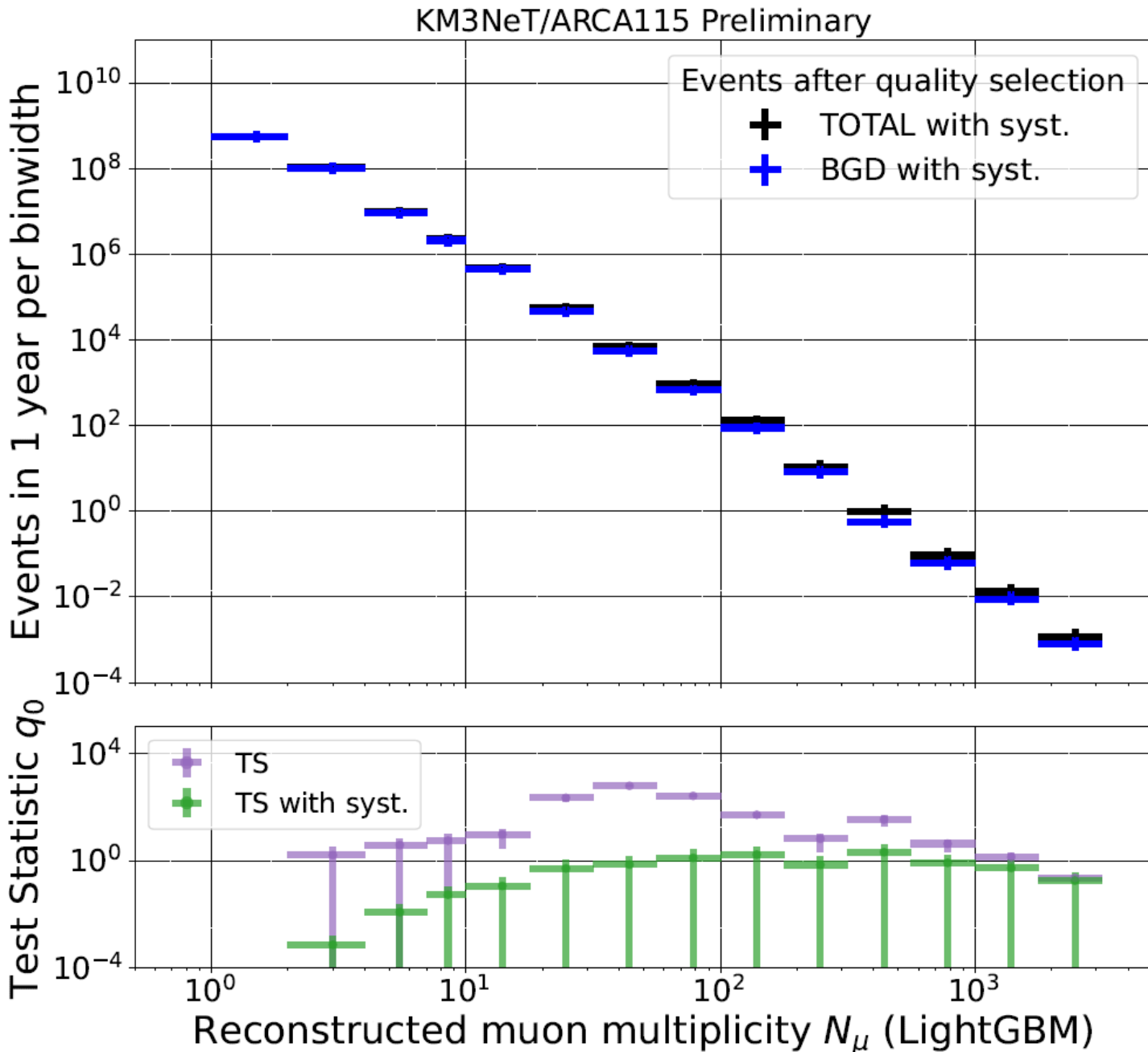


Zenit okazał się nieprzydatny (TS ≈ 0) dla tej analizy, pokazuję jedynie dla kompletności



Komentarze:

- ✦ Analiza jest zdecydowanie czuła na energię pęku
- ✦ Wygląda na to, że nawet przy energiach rzędu TeV jest różnica między hipotezami

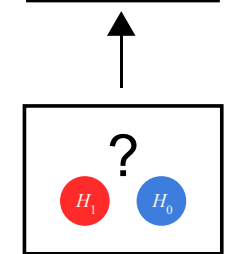


Komentarze:

- ❖ Krotność mionów również jest użyteczna
- ❖ Najlepsza czułość dla pęków składających się średnio z ~ 40 mionów

Define the test

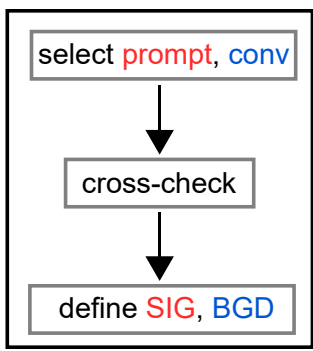
TS= ...
Z= ...
crit. Z = ...



Hypotheses

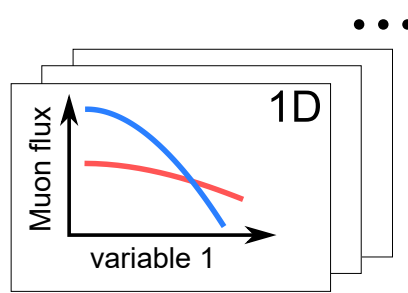


MC



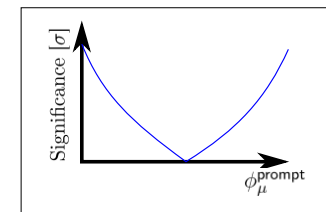
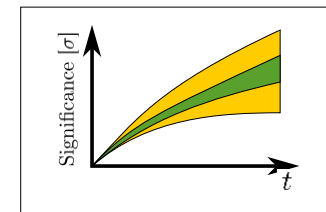
Quality cuts

Observable distributions

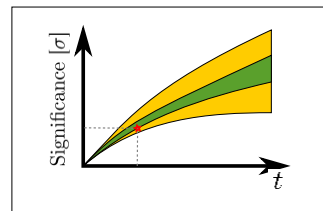


Critical region

Sensitivity



Measurement



$\phi_\mu^{\text{prompt}} = \dots$
or
 $\phi_\mu^{\text{prompt}} < \dots$

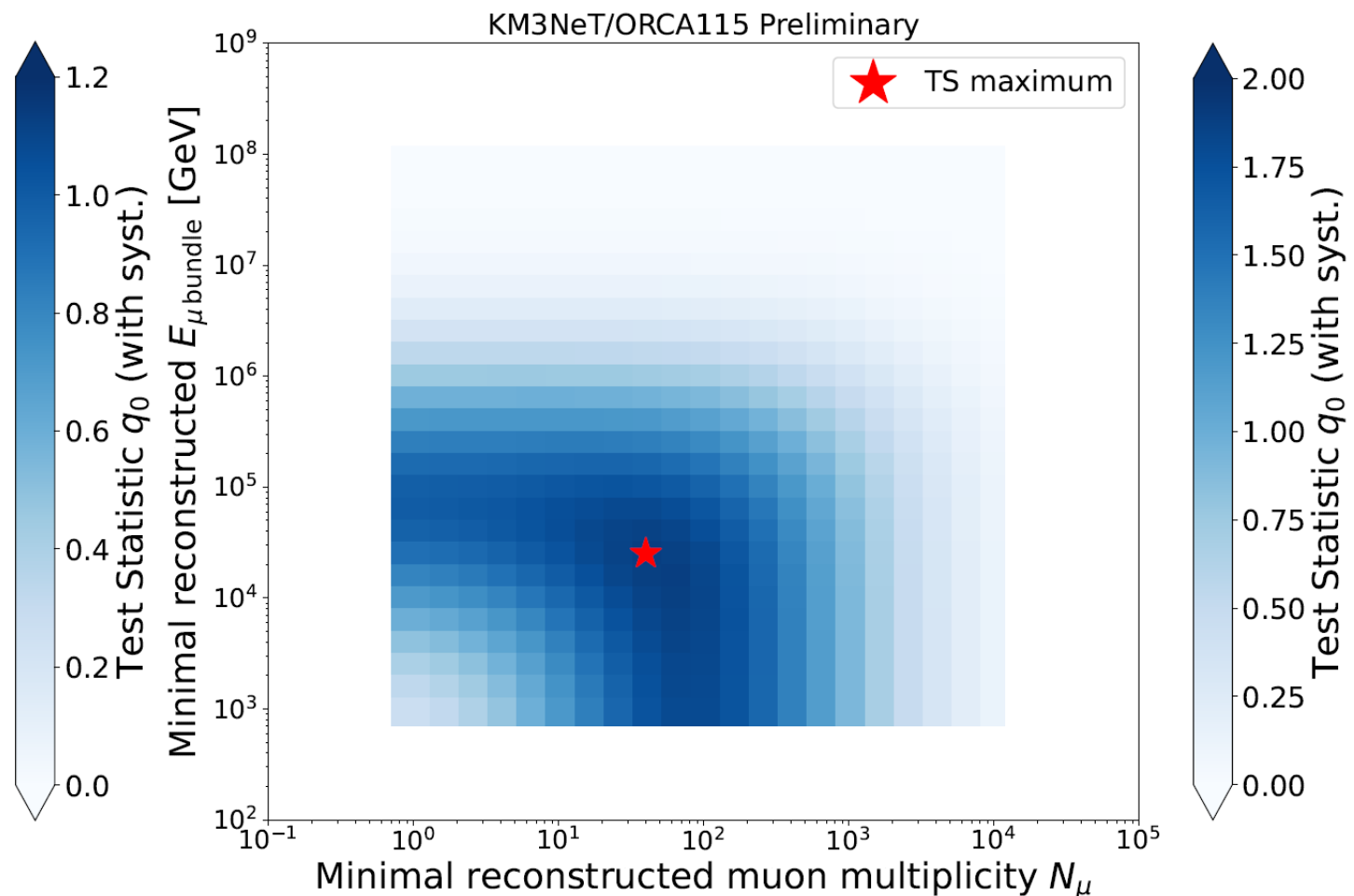
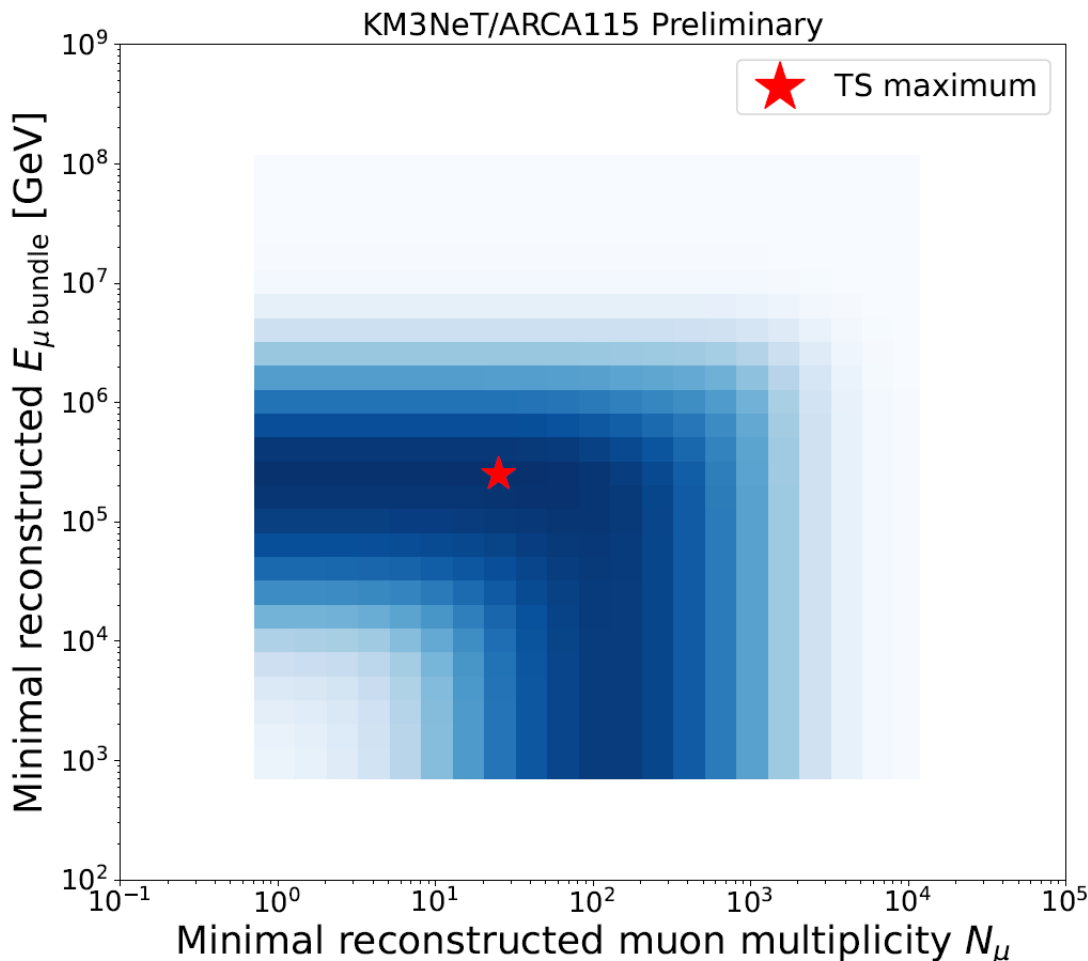


Exp data

Detektor	Min E_{bundle} [TeV]	Min N_{μ}
ARCA115	251	40
ARCA6	251	40
ORCA115	16	40
ORCA6	16	40

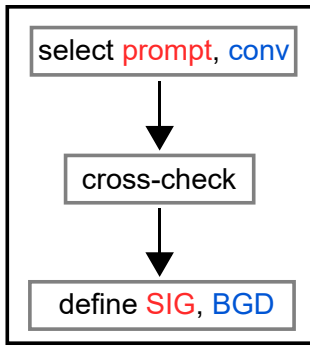
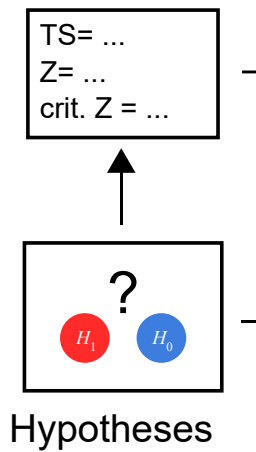
Dla mniejszych konfiguracji detektorów użyłem tych samych wartości jak dla pełnych dla lepszej porównywalności.

Optymalizuję tylko dolne granice, ponieważ ograniczenie od góry nie poprawia czułości testu (dla wysokich energii i krotności sygnału jest tylko więcej).

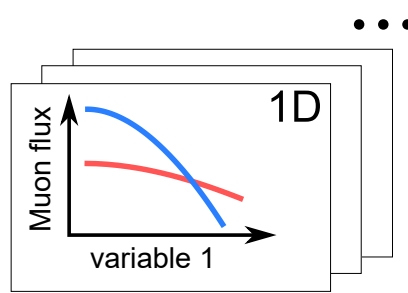


więcej szczegółów w slajdach zapasowych ...

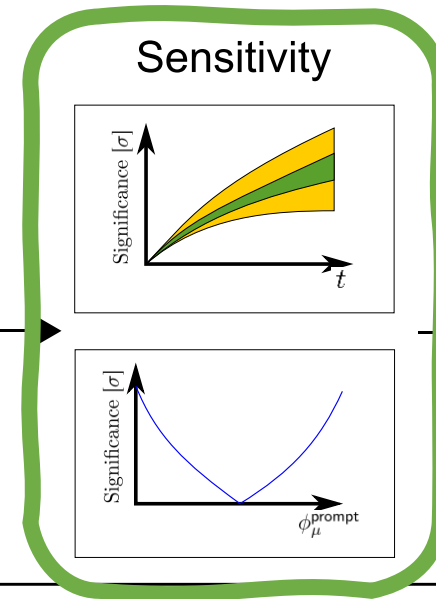
Define the test



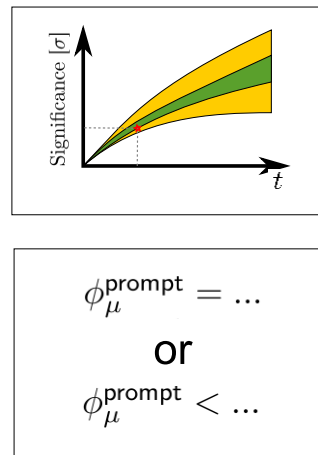
Observable distributions



Critical region



Measurement

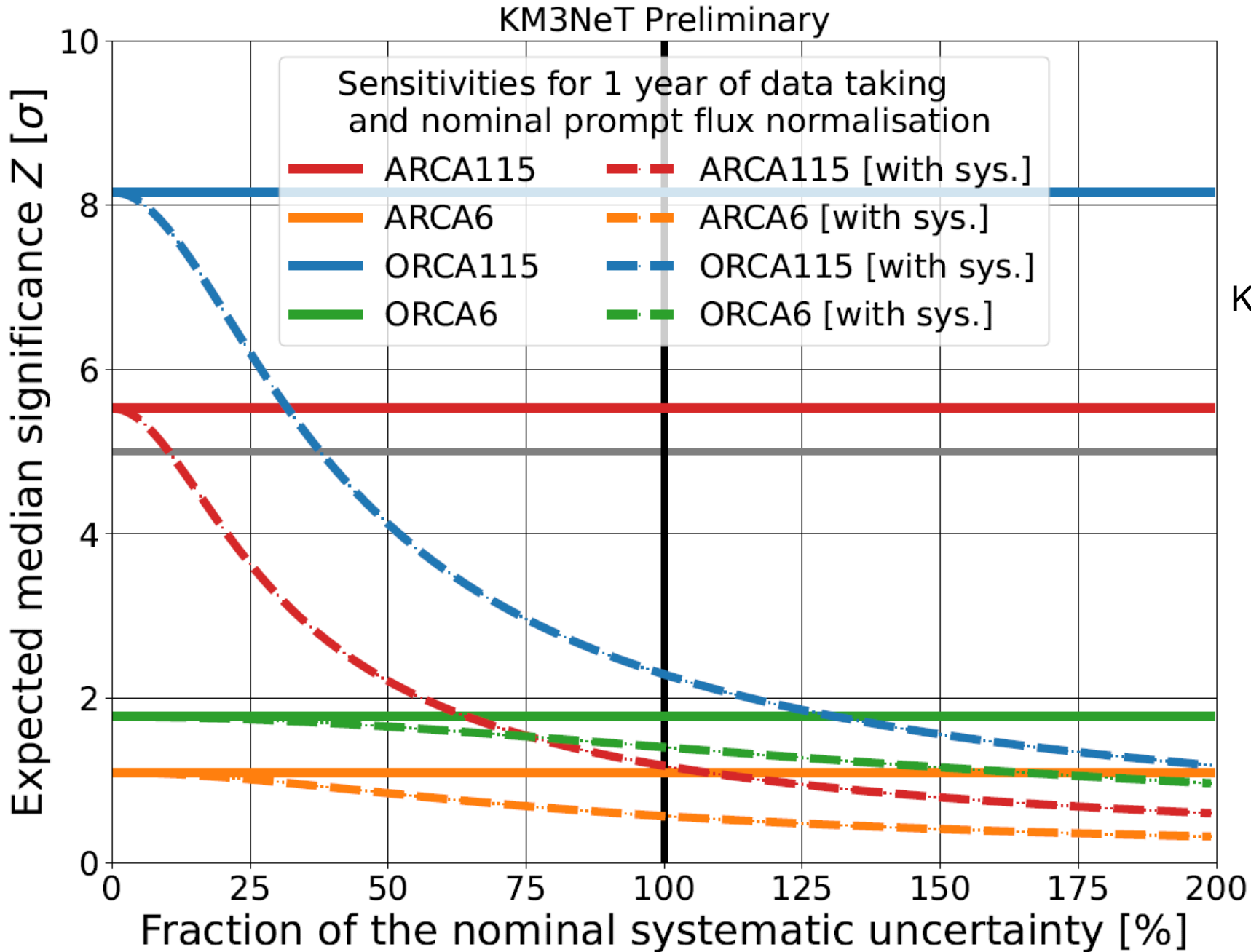


MC

Quality cuts



Exp data



Komentarze:

- ❖ systematyka wpływa silniej na większe konfiguracje
- ❖ jej redukcja może dramatycznie polepszyć czułość
- ❖ będzie to wymagało kosztownych dedykowanych symulacji

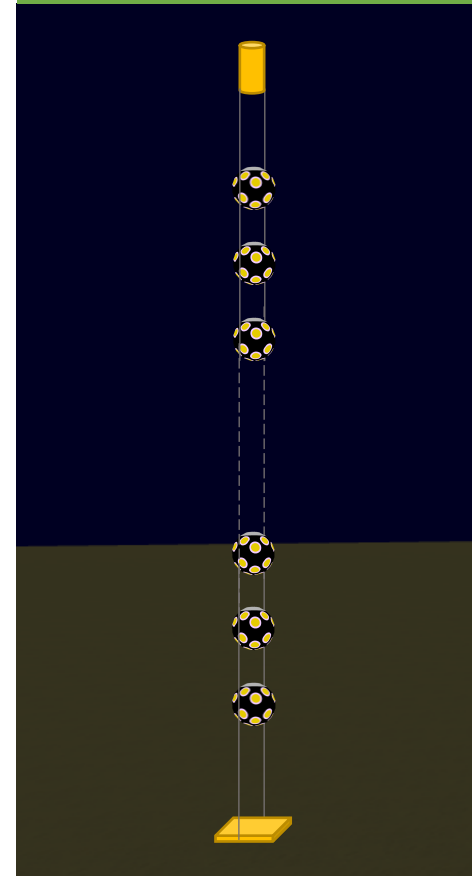
DOM:
71 unique components



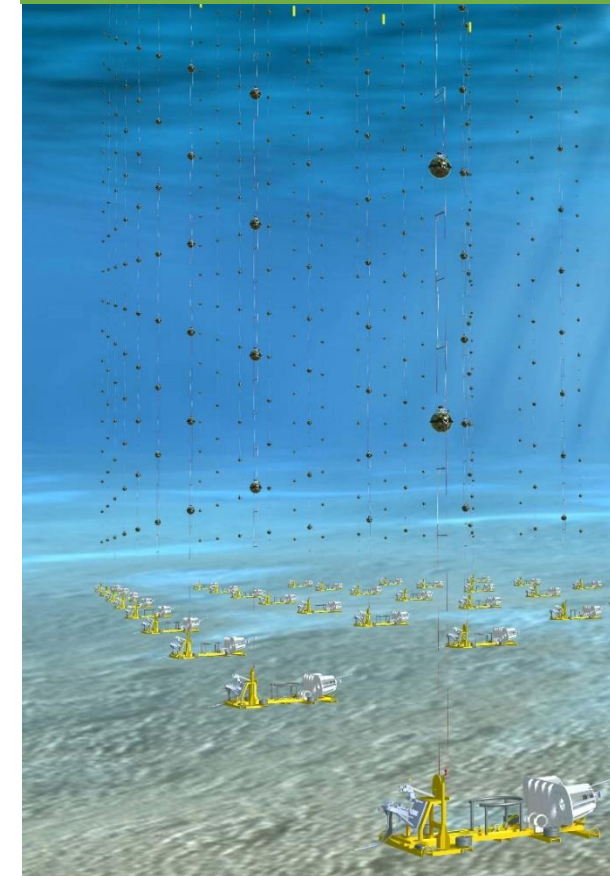
1 DOM:
31 PMTs



1 string (DU):
18 DOMs



1 building block:
115 DUs



[DOM production: \(@Nikhef\)](#)



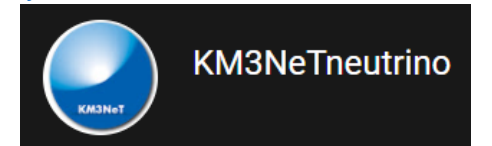
[Preparation for deployment:](#)

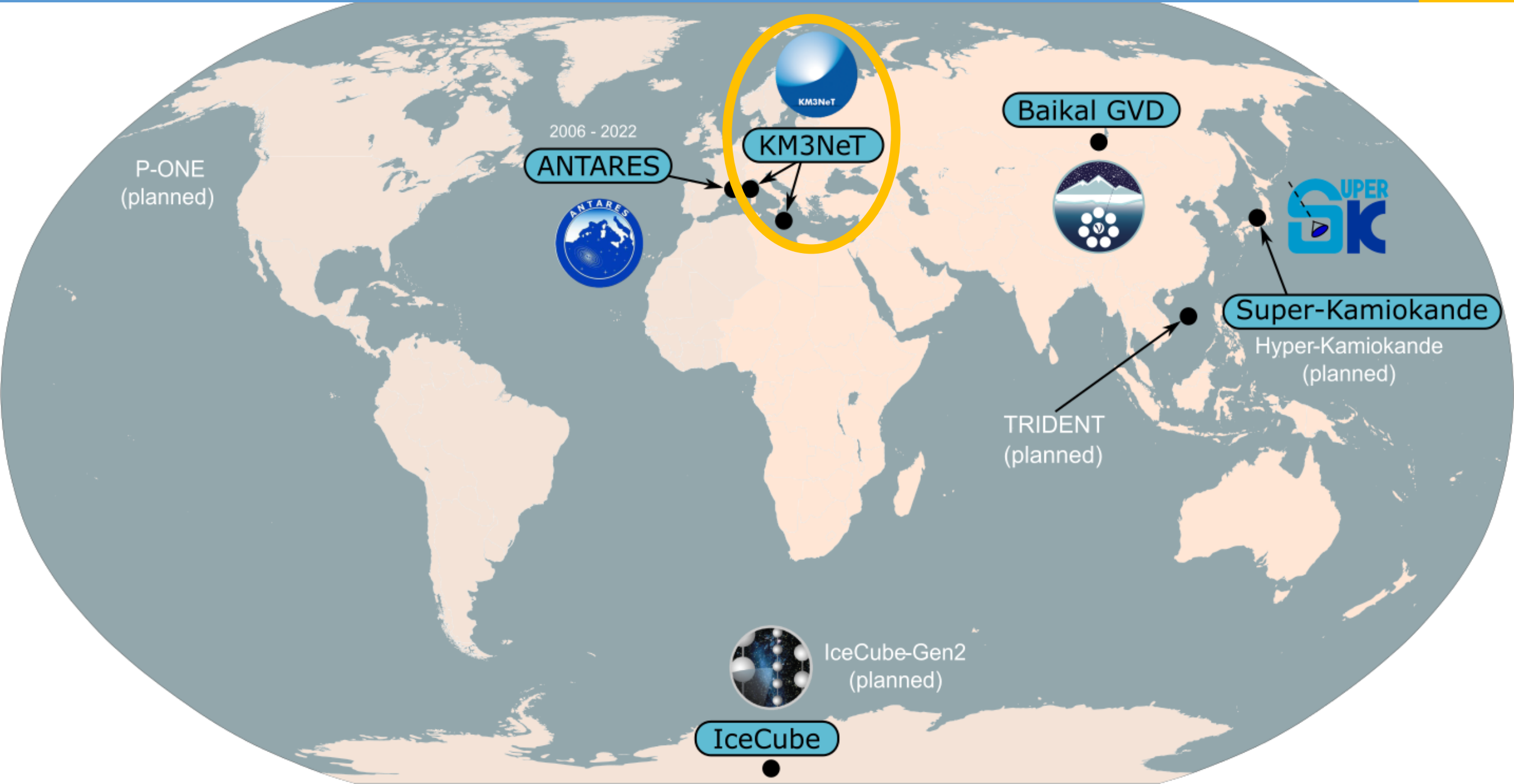


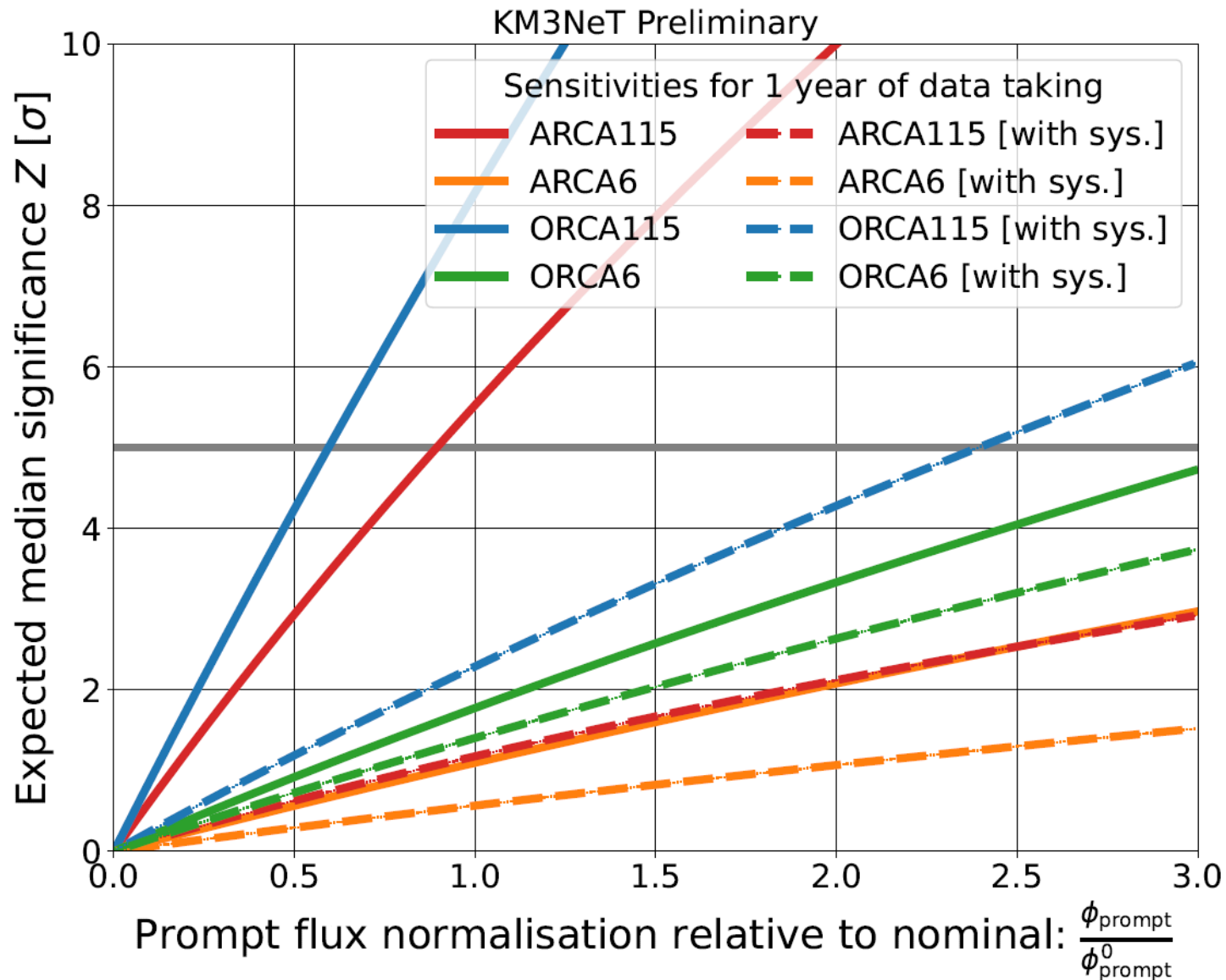
[String deployment:](#)



More at:
youtube.com/KM3NeTneutrino





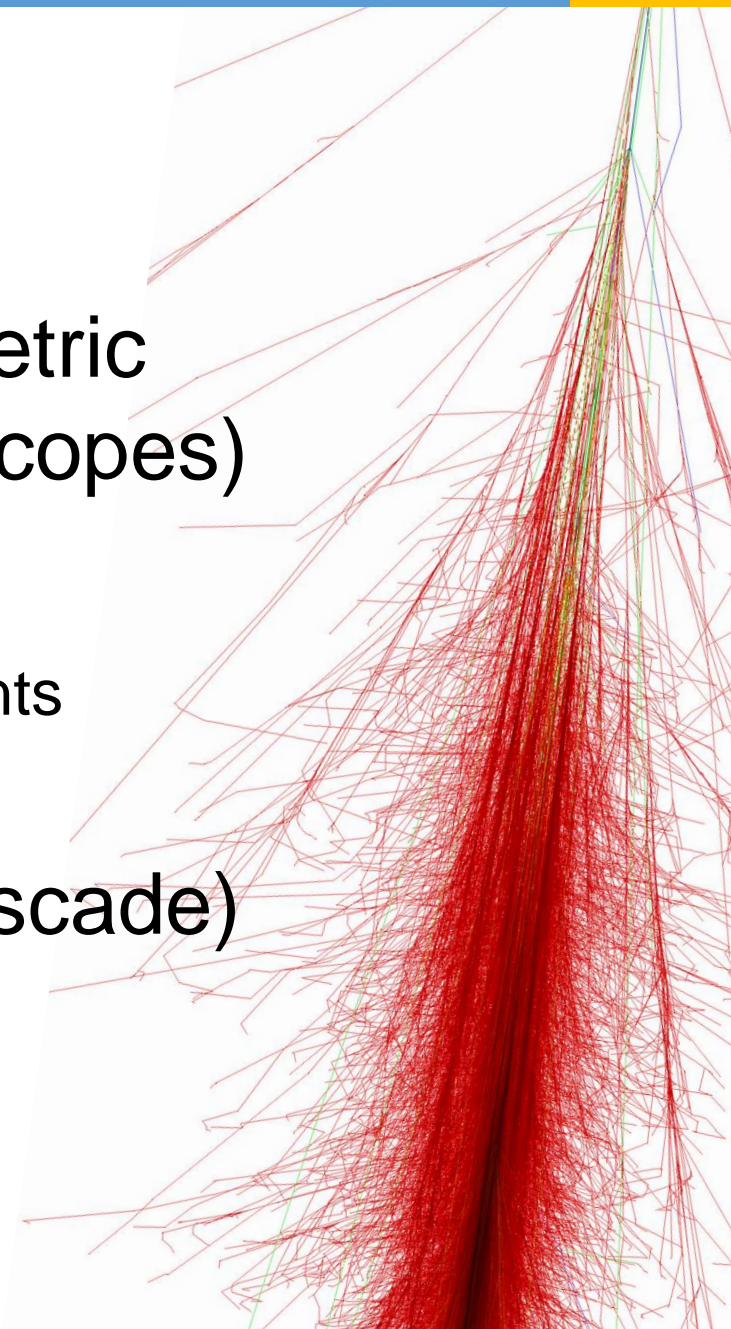


Comments:

- ❖ prompt flux normalisation has a linear effect on sensitivity
- ❖ still, systematics are the dominant issue

We have 2 options:

1. [MUPAGE](#) (atmospheric **MU**ons from **PA**rametric formulas: a fast **GE**nerator for neutrino telescopes)
 - developed for ANTARES
 - fast muon MC generator
 - based on parametric formulas and MACRO measurements
 - parameters can be freely tuned
2. [CORSIKA](#) (**CO**smic **R**ay **SI**mulations for **KA**scade)
 - developed for KASCADE
 - full simulation of air showers
 - customizable (models, primaries, etc.)



Metoda:

- ❖ Niepewności liczone indywidualnie dla każdego przypadku, tzn. oszacowuję $\Delta w_{\text{event}}(E_{\text{prim}})$
- ❖ Jeżeli to możliwe, osobne oszacowania wkładów dla detektorów ARCA and ORCA
- ❖ Każda niepewność liczona jako: $\frac{\text{max-domyślne}}{\text{domyślne}}$ albo $\frac{\text{domyślne-min}}{\text{domyślne}}$
- ❖ 3 podejścia:
 - [2,3] Dedykowane mini-produkcje CORSIKI (tylko pionowe pęki, dla ustalonych wartości E_{prim} , $1.11 \cdot 10^7$ pęków/rodzaj cząstki pierwotnej)
 - [1] Pakiet crflux (bezpośrednie oszacowanie strumienia CR)
 - [4,5] Dedykowane mini-produkcje MUPAGE'a ([wygenerowane przez Andreya Romanova](#))

Uwzględnione źródła niepewności:

- | | |
|---|---|
| 1. Modele strumienia cząstek pierwotnych CR | domyślnie: GST3 |
| 2. Wysoko-energetyczne modele oddziaływań hadronowych | domyślnie : SIBYLL 2.3d |
| 3. Sezonowe różnice w profilu gęstości atmosfery | domyślnie : dopasowanie do atmosfery MSIS2.0 uśrednionej po porach roku i lokalizacji (ARCA i ORCA) |
| 4. Efektywność fotopowielaczy | domyślnie : wartość nominalna (A. Romanov) |
| 5. Długości absorpcji światła w wodzie | domyślnie : wartość nominalna (A. Romanov) |

Digital Optical Module (DOM)

acrylic glass sphere with:

- 31 3" PMTs,
- readout electronics,
- pressure gauge,
- acoustic sensors,
- ...

2022 JINST 17 P0703

JATIS 7(1), 016001 (2021)

Photomultiplier Tube (PMT)

converts light into electric signal

JINST13 (2018) P05035



Detection Unit (DU):
vertical string with 18 DOMs

Eur. Phys. J. C 76 (2016) 76:54

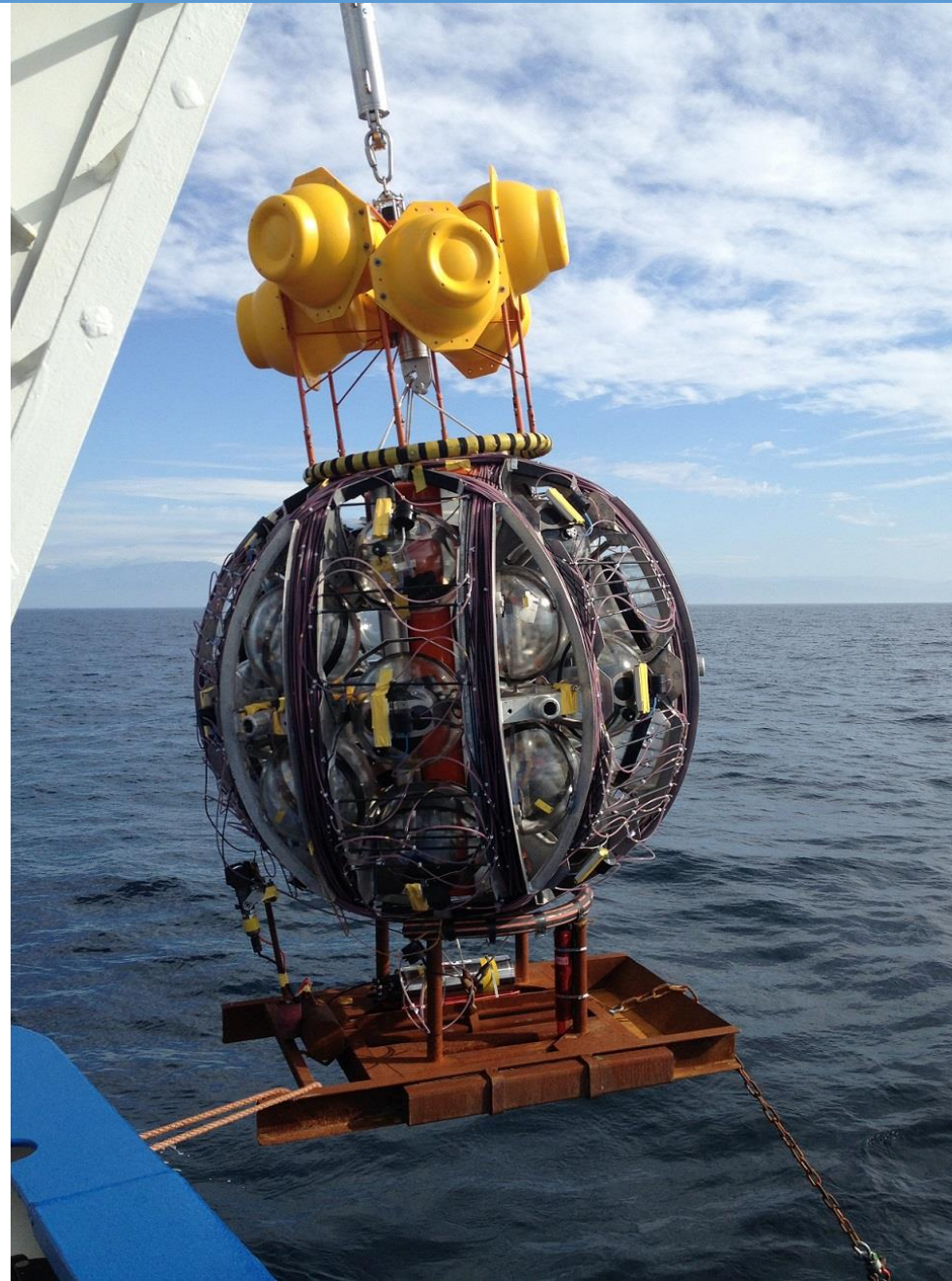
Naming:

ORCA6 ↔ ORCA with 6 strings

ARCA2 ↔ ARCA with 2 strings

etc.

2020 JINST 15 P11027



Detection Unit (DU):
vertical string with 18 DOMs

Eur. Phys. J. C 76 (2016) 76:54

Naming:

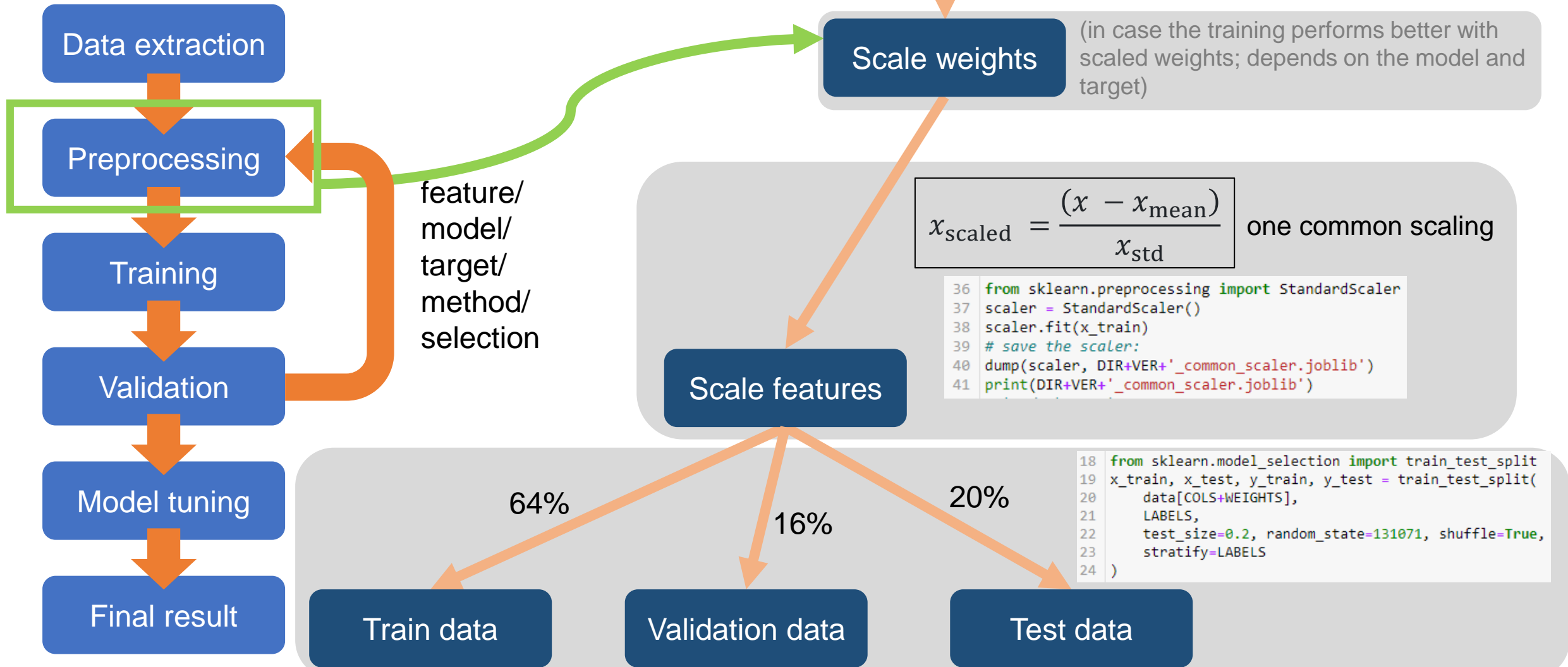
ORCA6 ↔ ORCA with 6 strings

ARCA2 ↔ ARCA with 2 strings
etc.

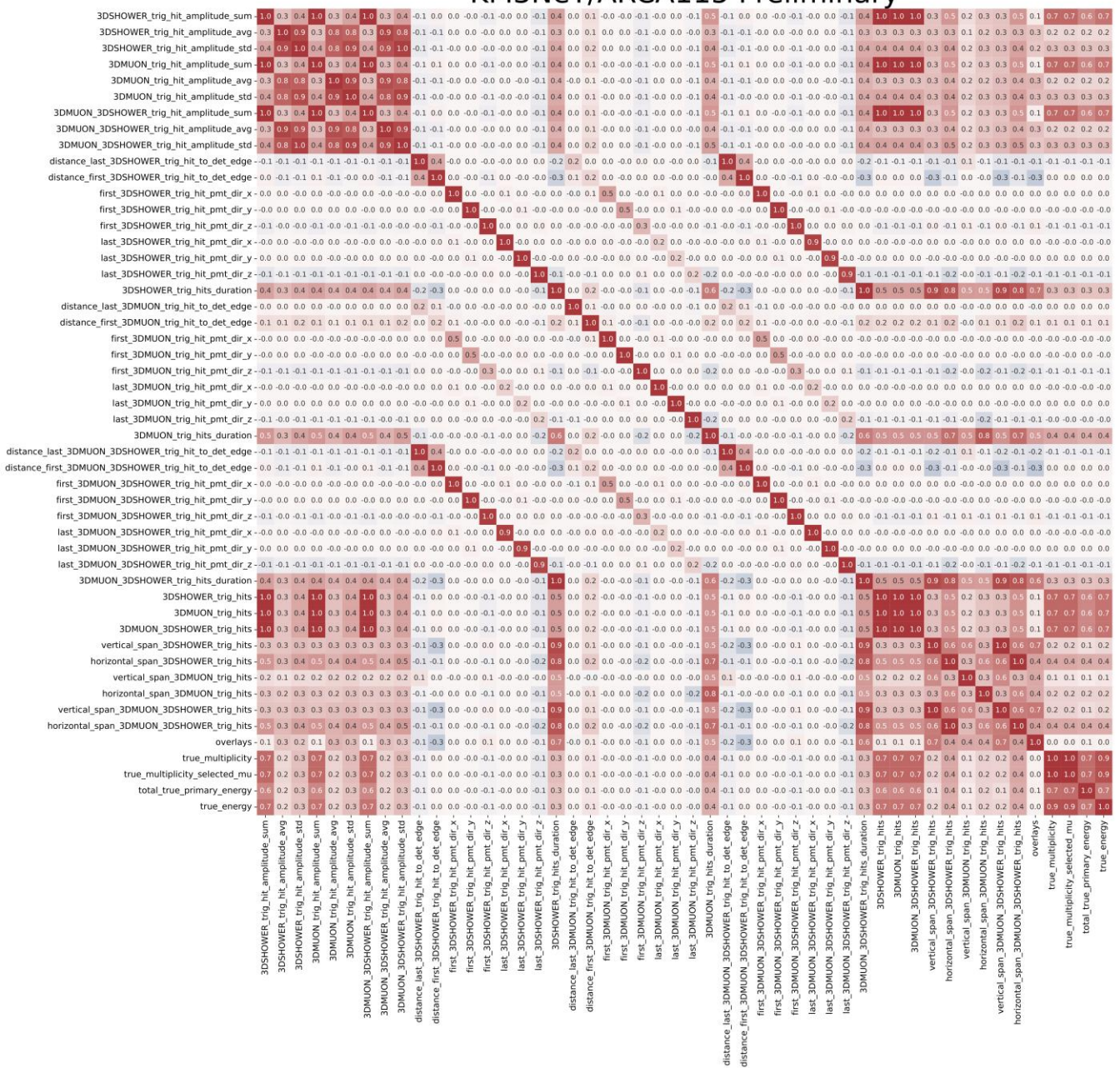
2020 JINST 15 P11027



Workflow of the reconstruction:



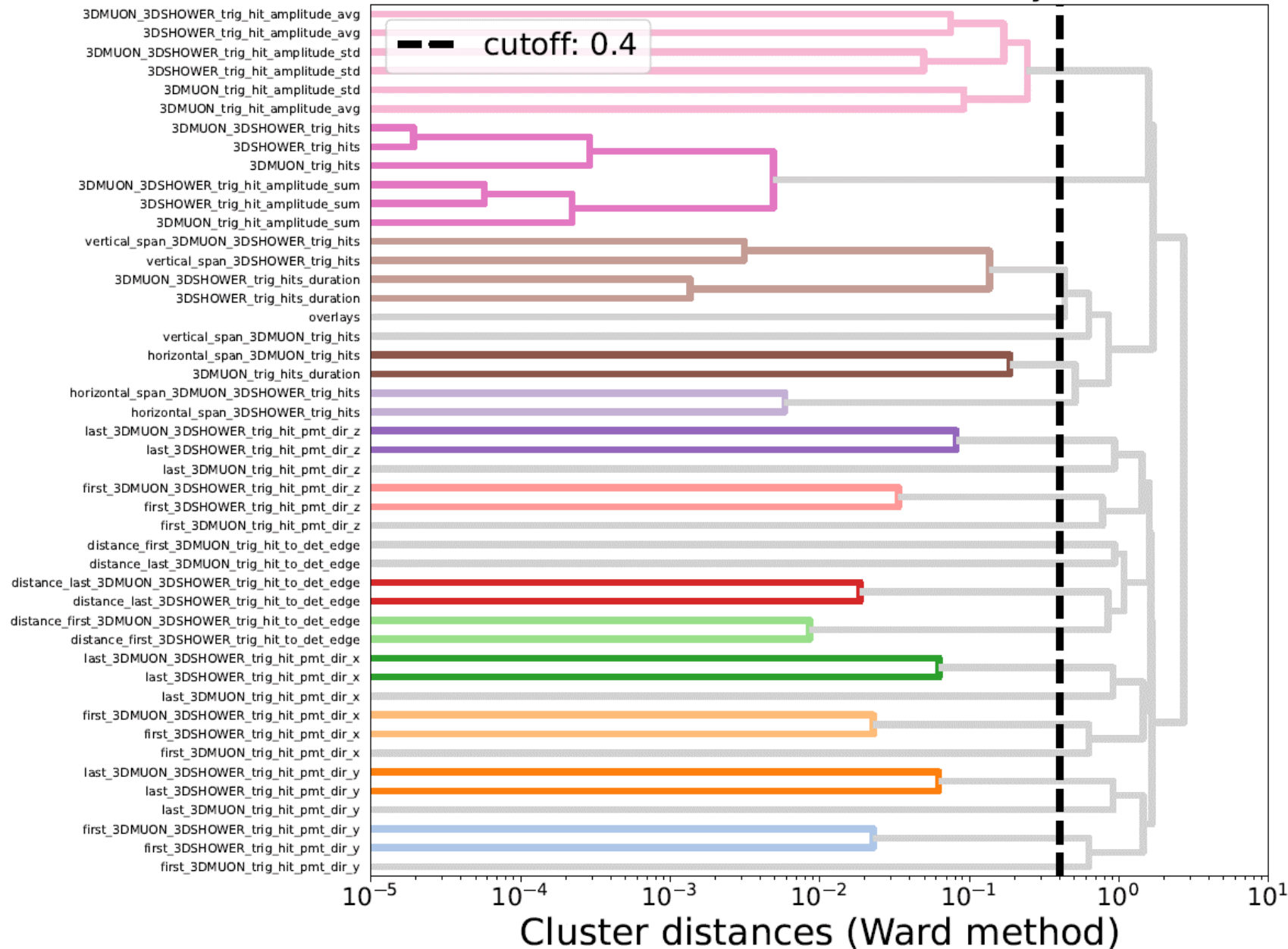
KM3NeT/ARCA115 Preliminary



In total: 46 features (+4 targets)

Example for ARCA115 (the same was done for ARCA6, ORCA115 and ORCA6)

KM3NeT/ARCA115 Preliminary

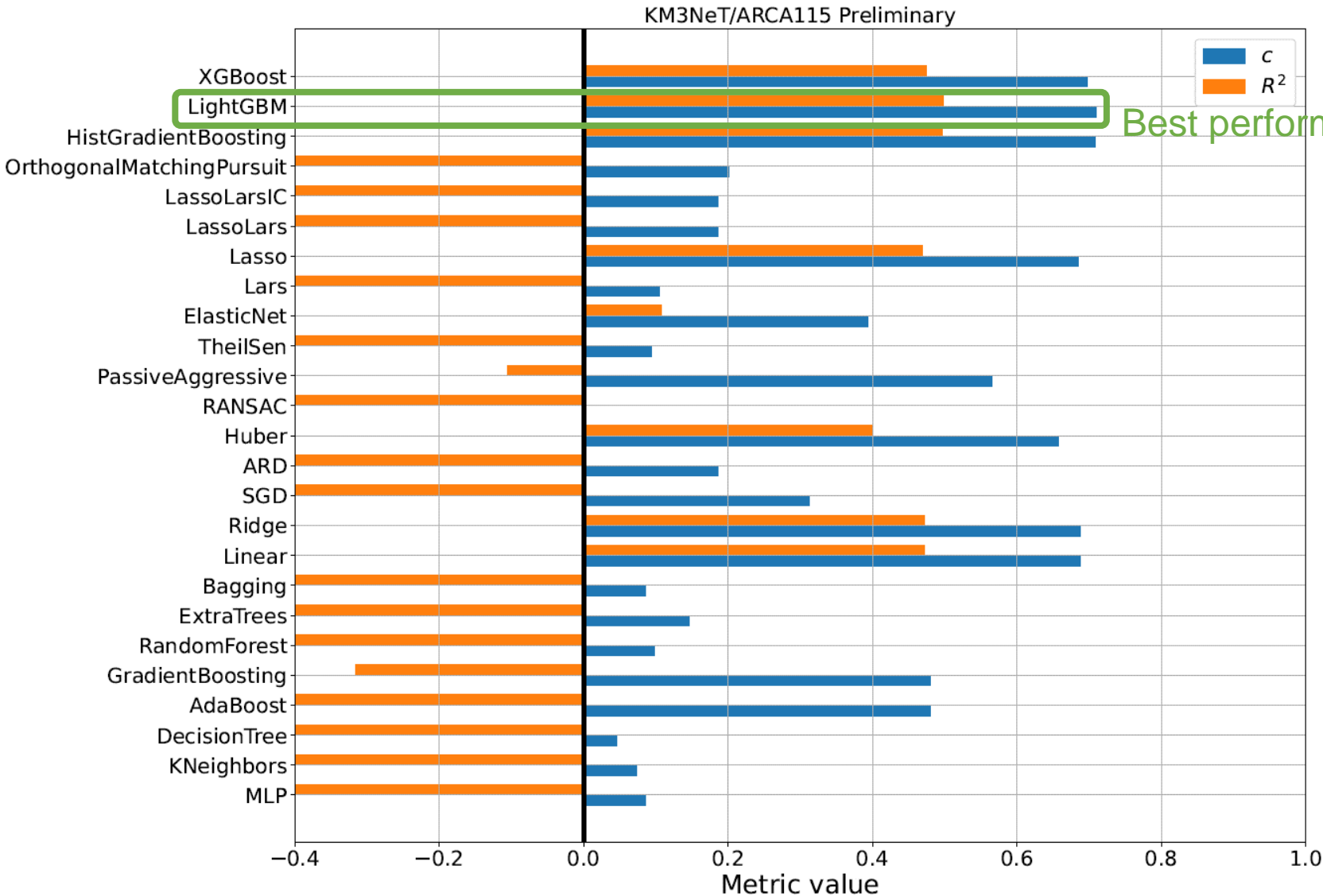


Cluster distance cutoff is arbitrary

Clusters are marked by different colors

Example for ARCA115
(the same was done for ARCA6, ORCA115 and ORCA6)

Performance comparison on a fraction (50k events) of the training dataset:



Best performance

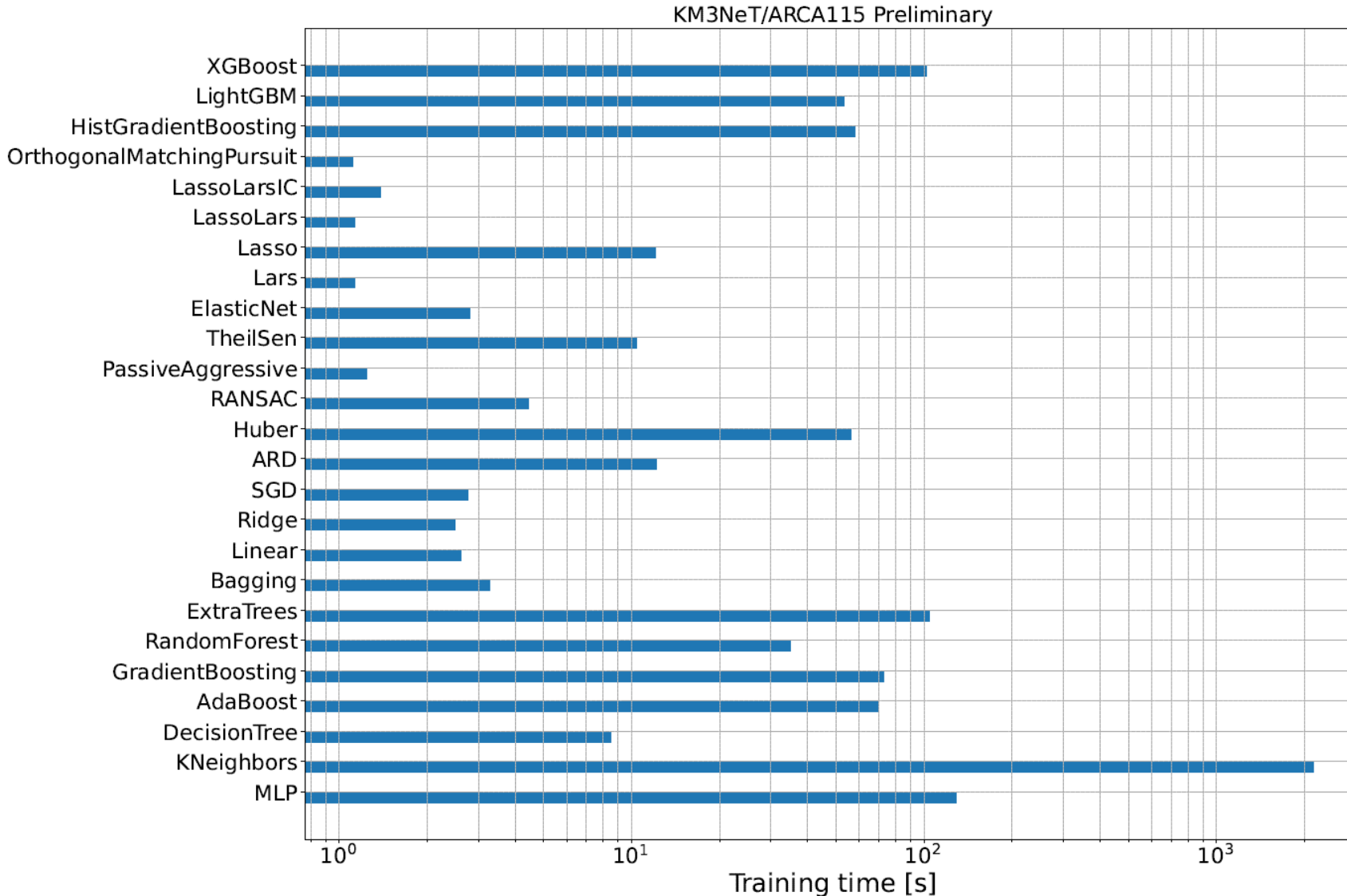
The (weighted) Pearson correlation coefficient:

$$c(x, y) = \frac{\sum_i w_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i w_i (x_i - \bar{x})^2 \sum_i w_i (y_i - \bar{y})^2}}$$

The (weighted) R2-score, called the coefficient of determination:

$$R^2(y_{\text{true}}, y_{\text{pred}}) = 1 - \frac{\sum_i w_i (y_{\text{true}} - y_{\text{pred}})^2}{\sum_i w_i (y_{\text{true}} - \bar{y})^2}$$

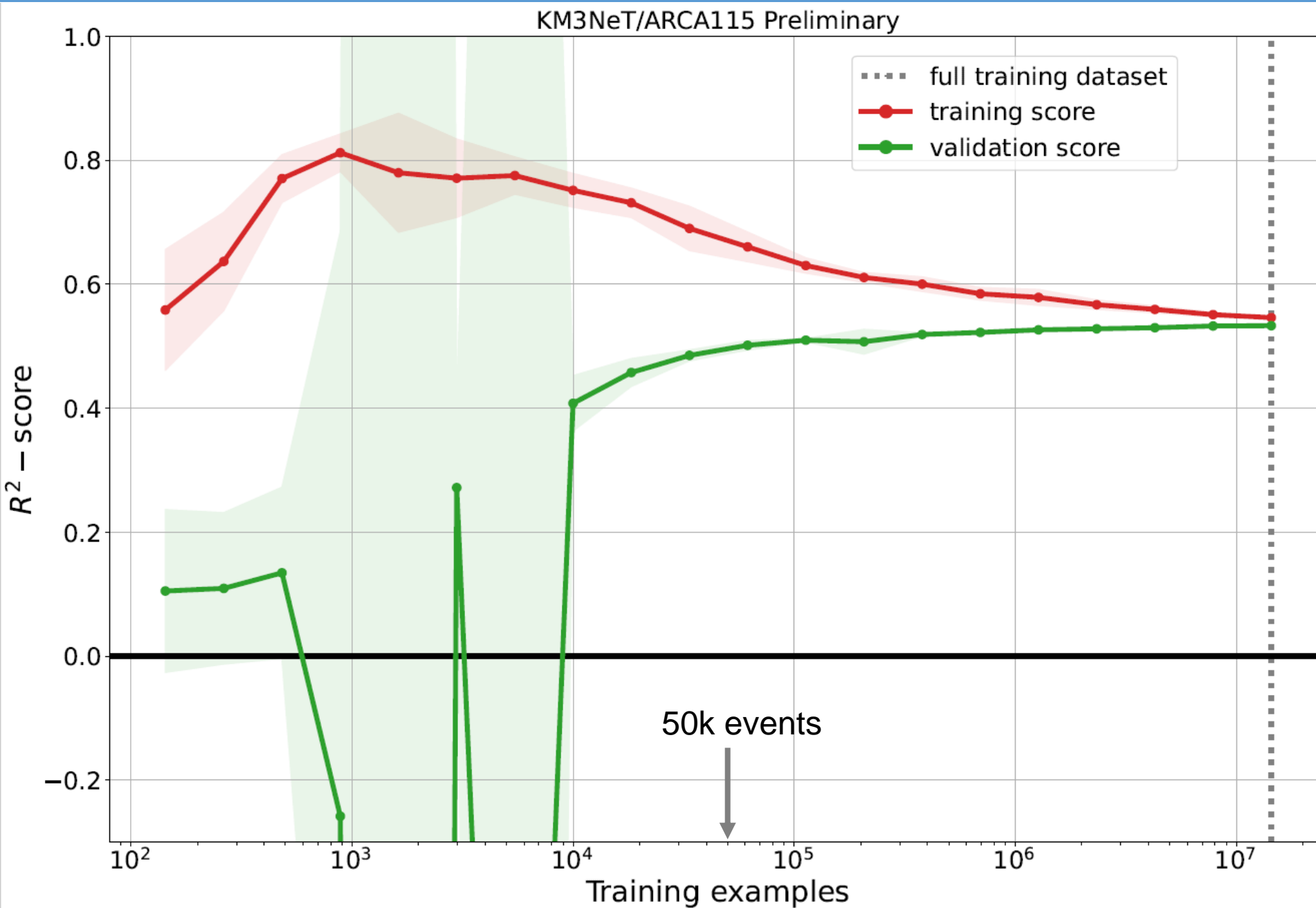
Speed comparison on a fraction (50k events) of the training dataset:



LightGBM:

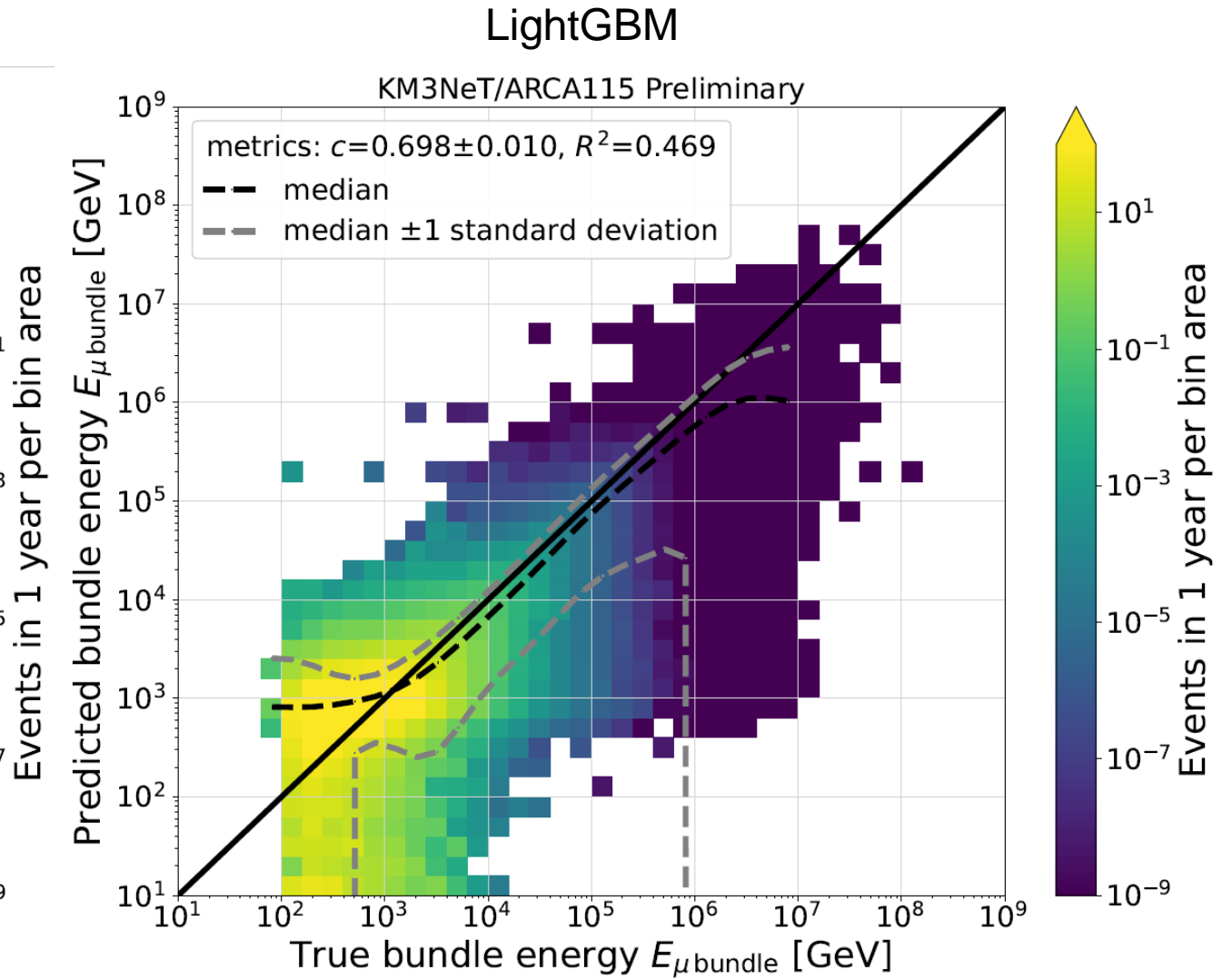
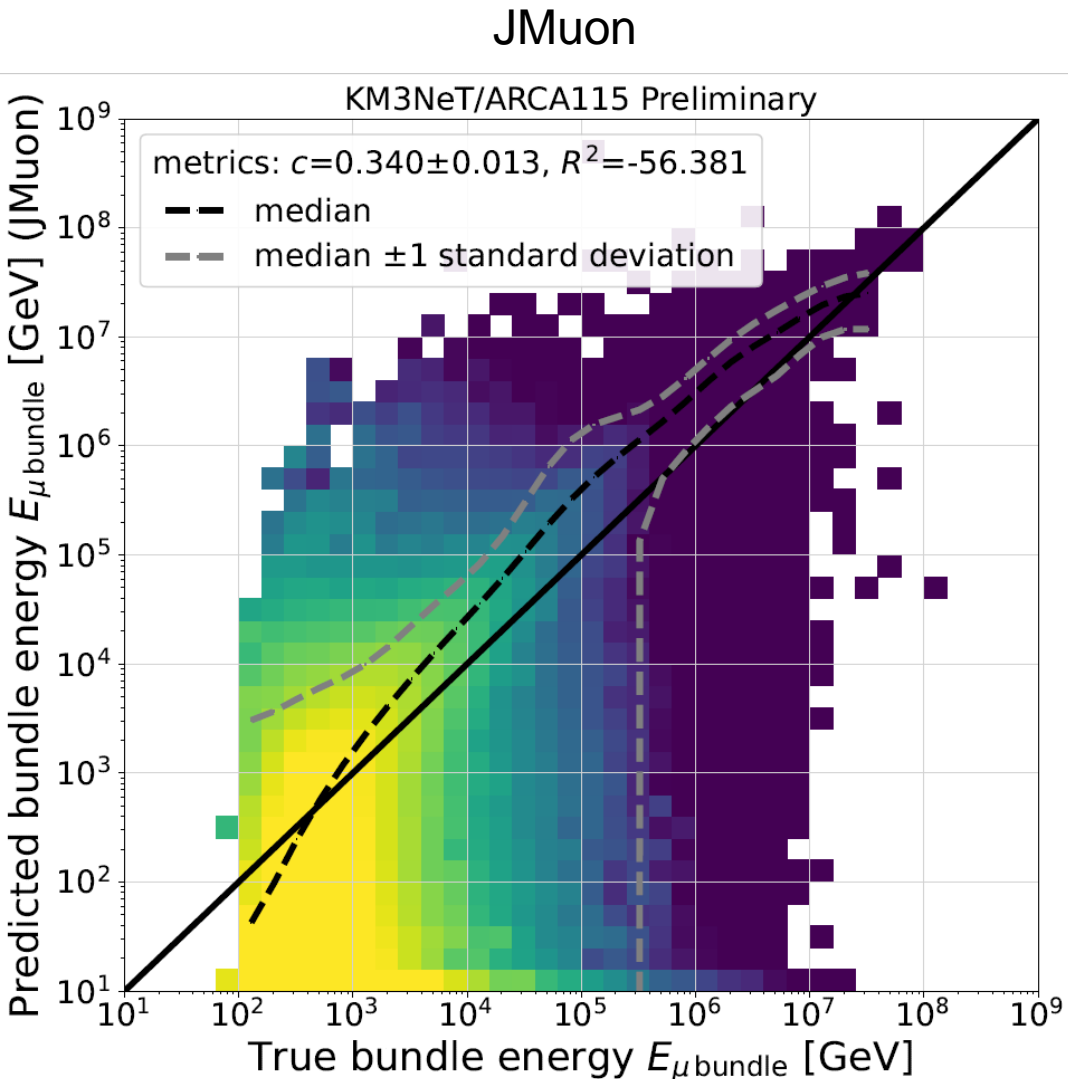
- ❖ not the fastest, but still very decent
- ❖ + it turned out to scale up very well (entire dataset is orders of magnitude larger)

These times were obtained running with 20 CPU cores in parallel

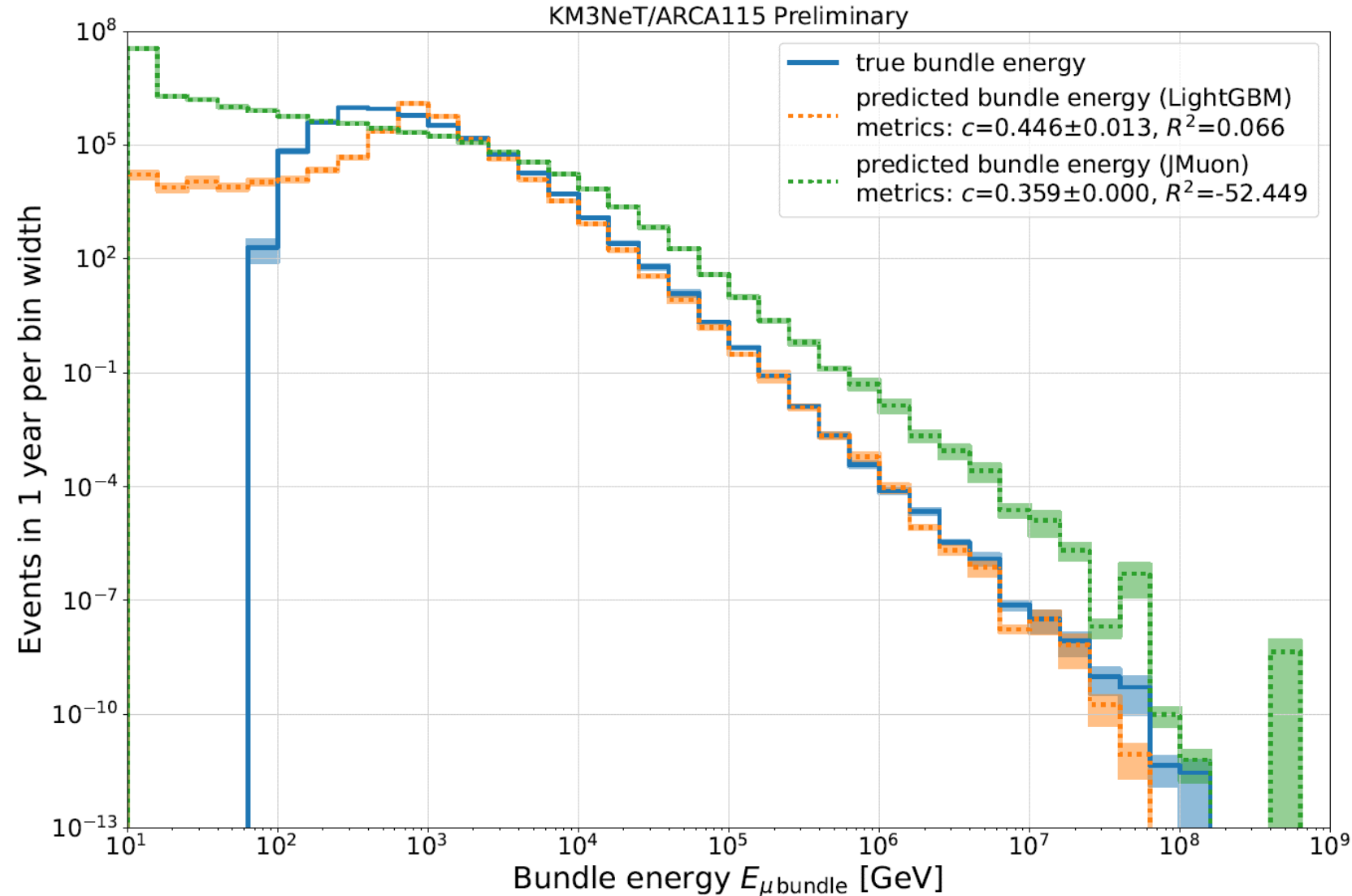


Here we see why 50k events were fine for testing (but e.g. 5k would not be)

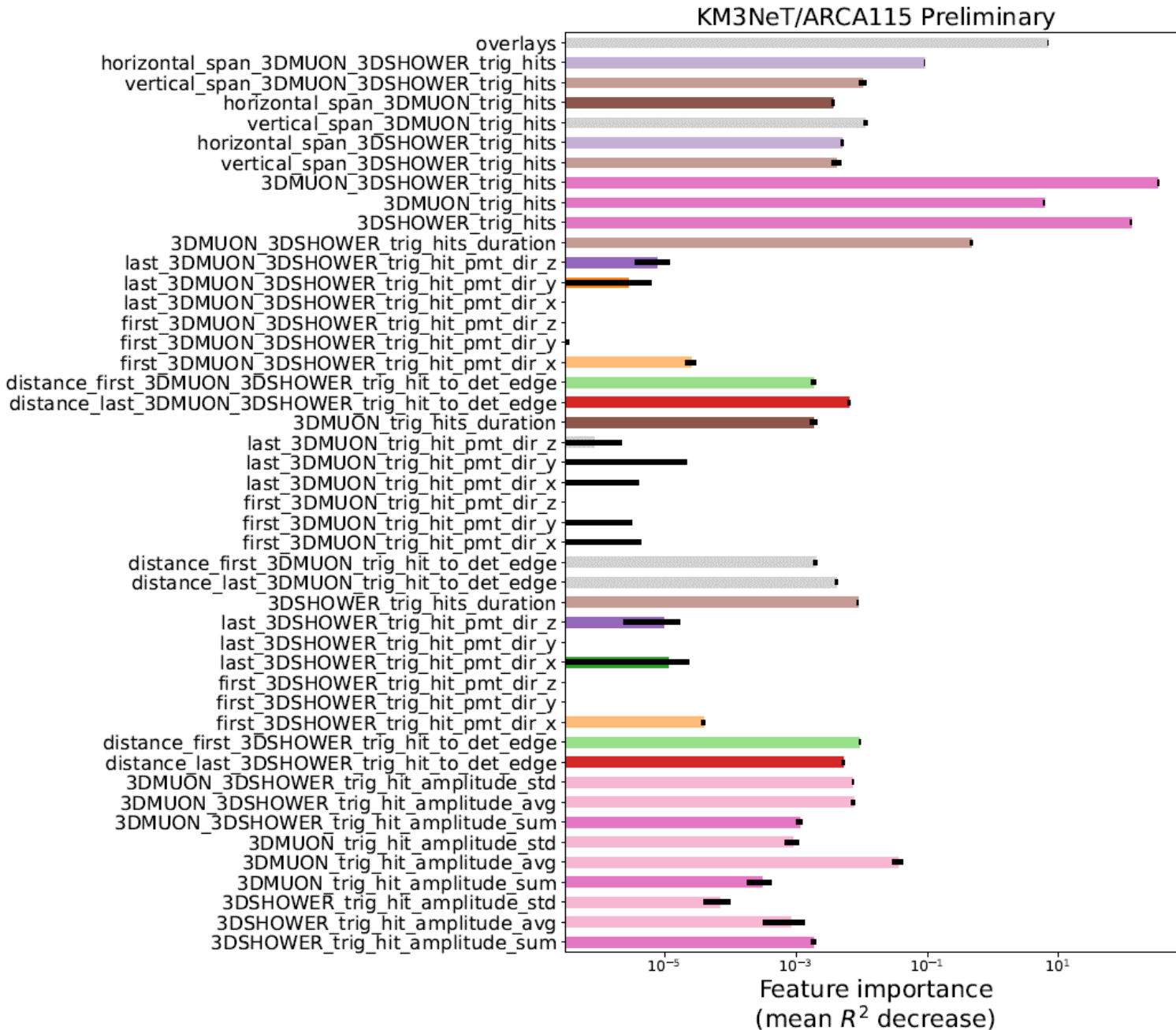
Here I just compare LightGBM (no tuning whatsoever) and JMuon reco
(non-ML reco)



Comparison in 1D:



Clearly even untuned ML approach reproduces the distribution much closer



Colors here are not random!

They match the feature clustering

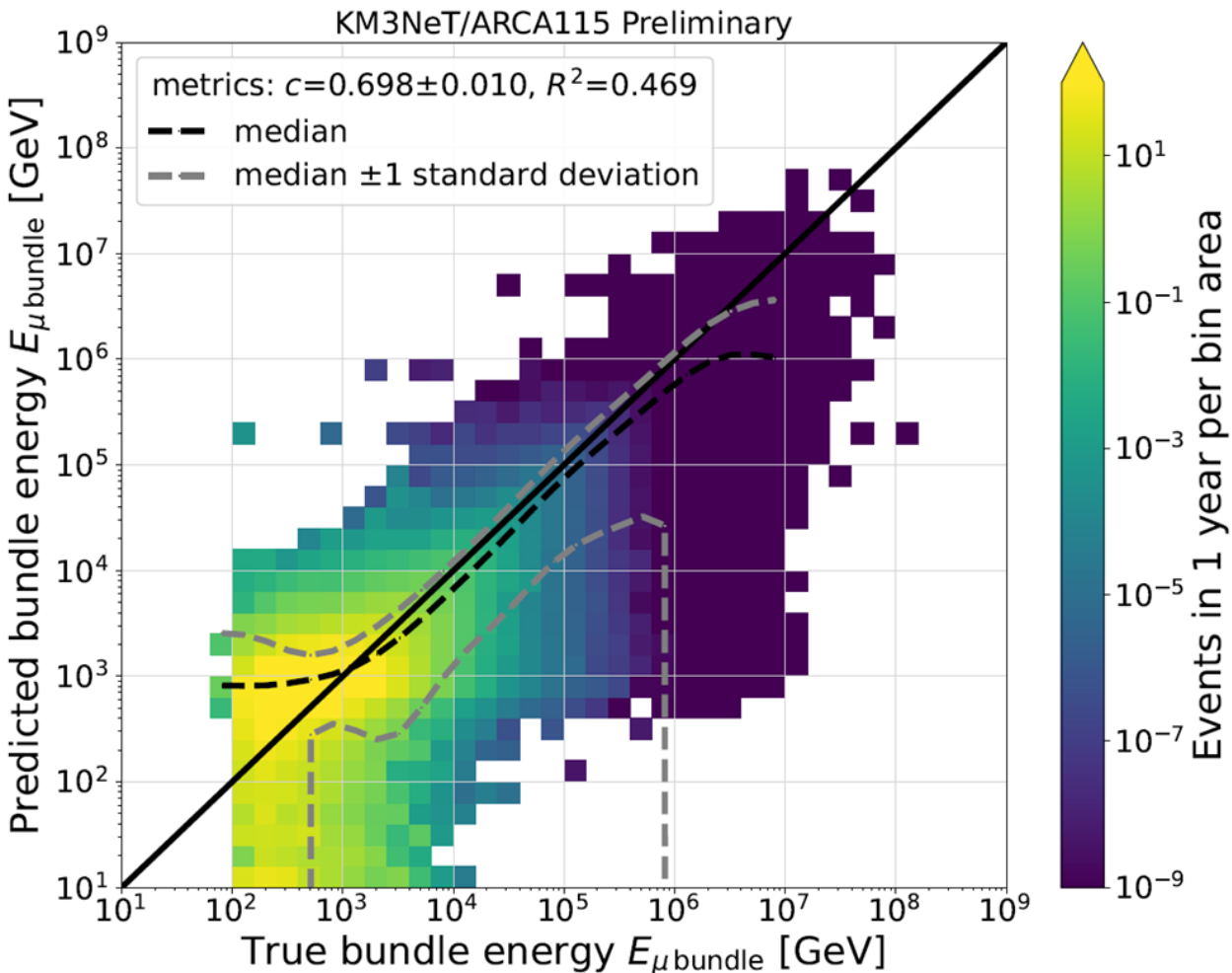
The idea:

Try to select only the most important feature in each cluster

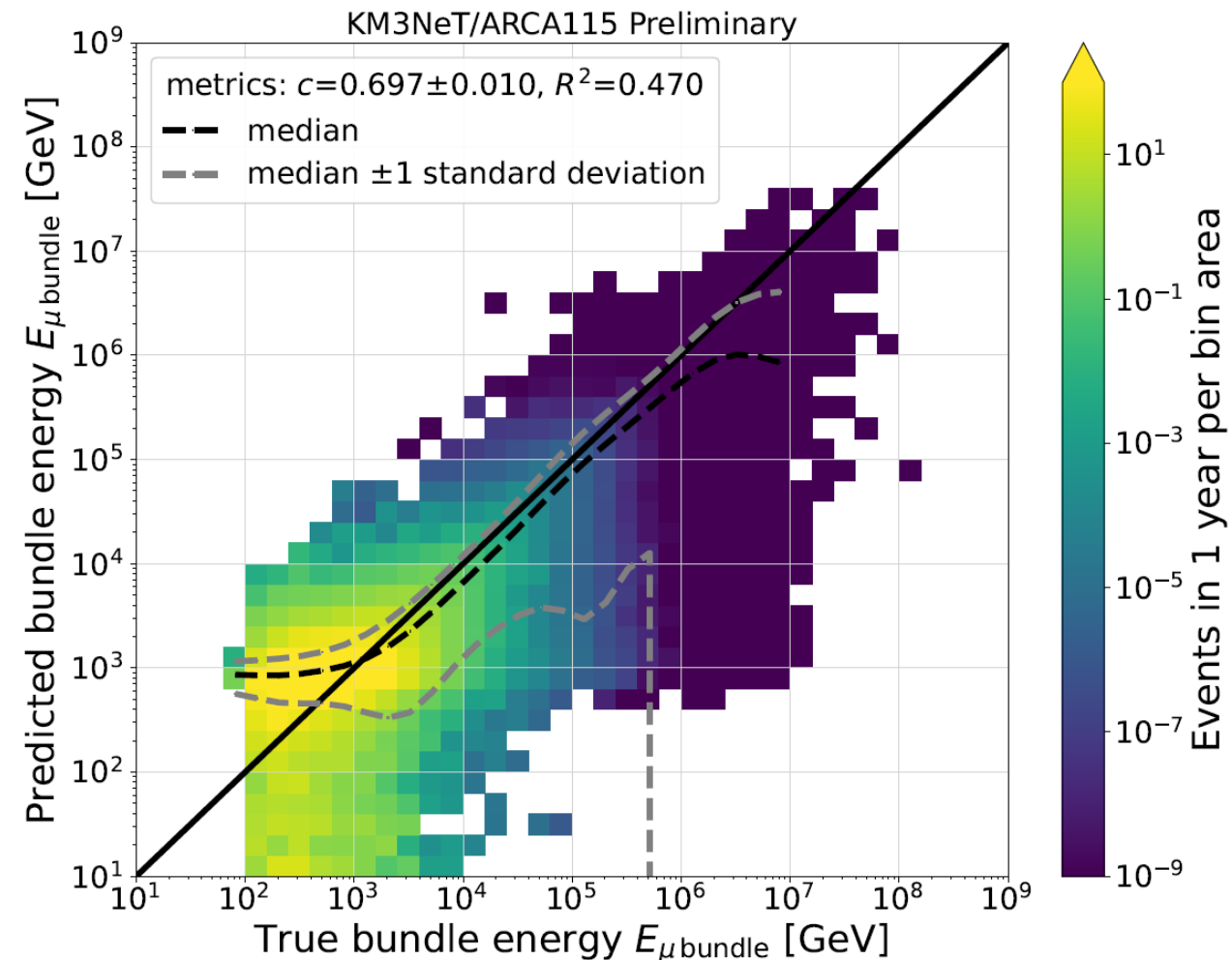
I considered 4 options:

1. All features
2. Features with importance>0 & only the most important
3. The most important feature only
4. Features with importance>0

1. All features



2. importance>0 & clustering

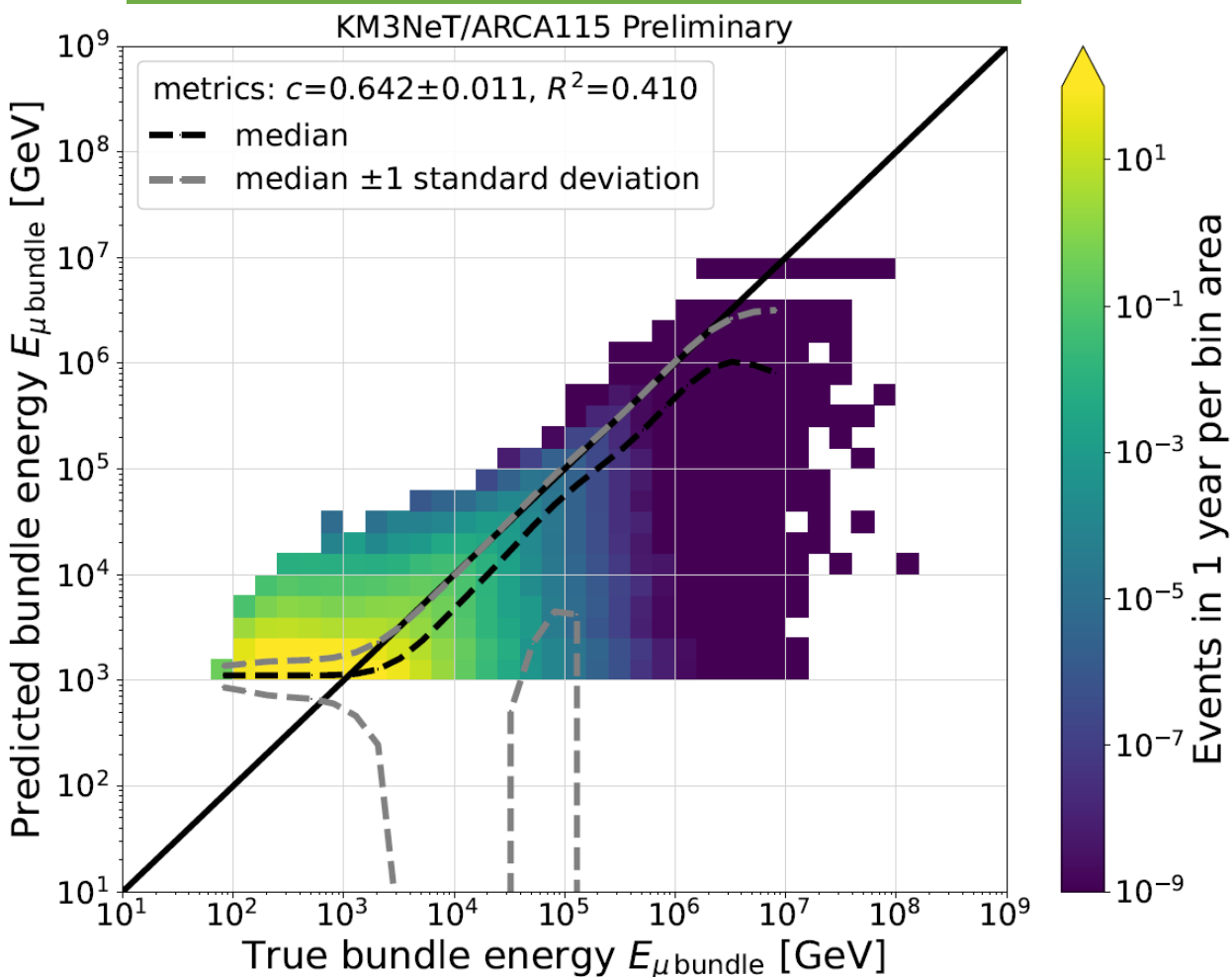


Bundle energy reco: feature selection

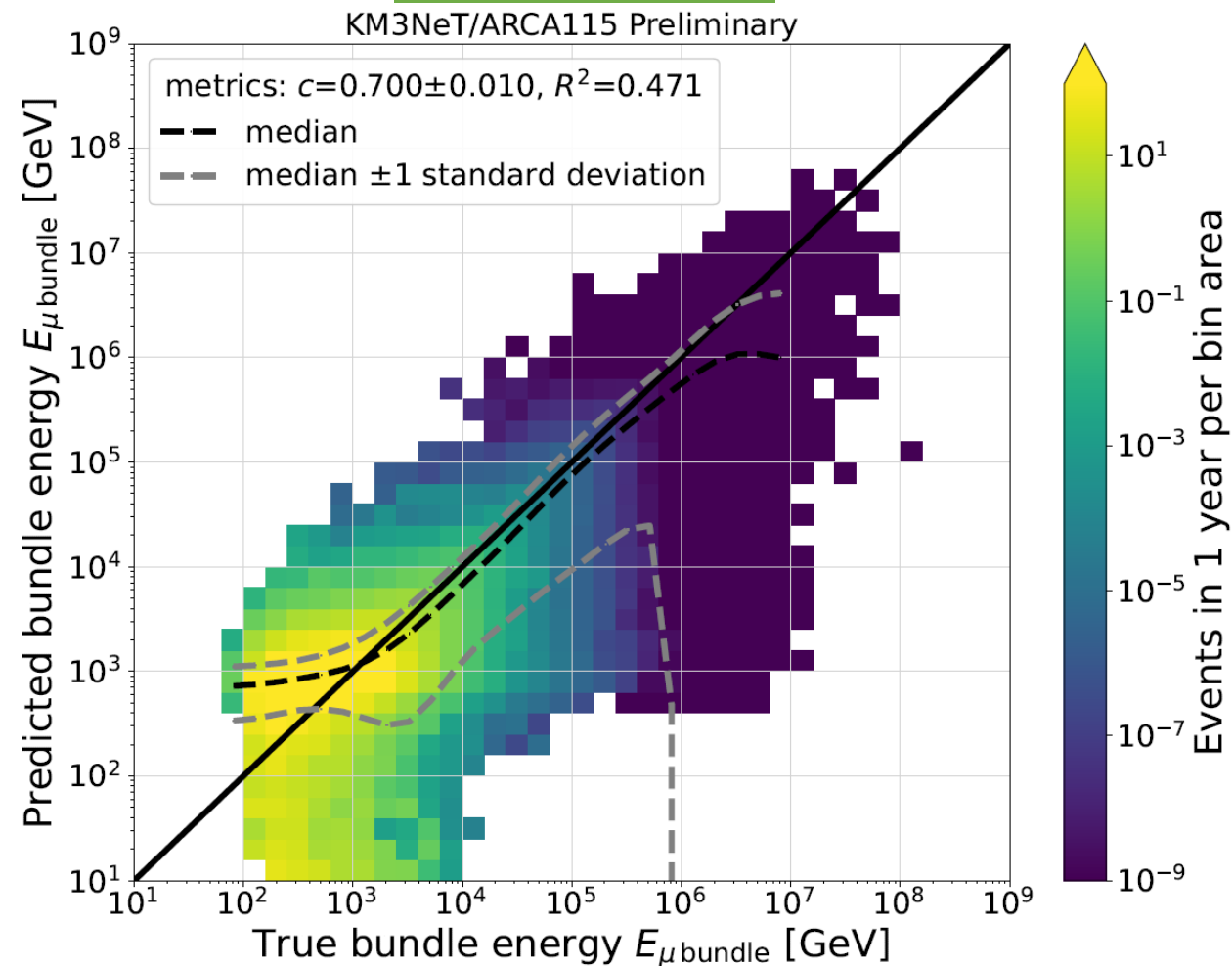
I considered 4 options:

1. All features
2. Features with importance>0 & only the most important
3. The most important feature only
4. Features with importance>0

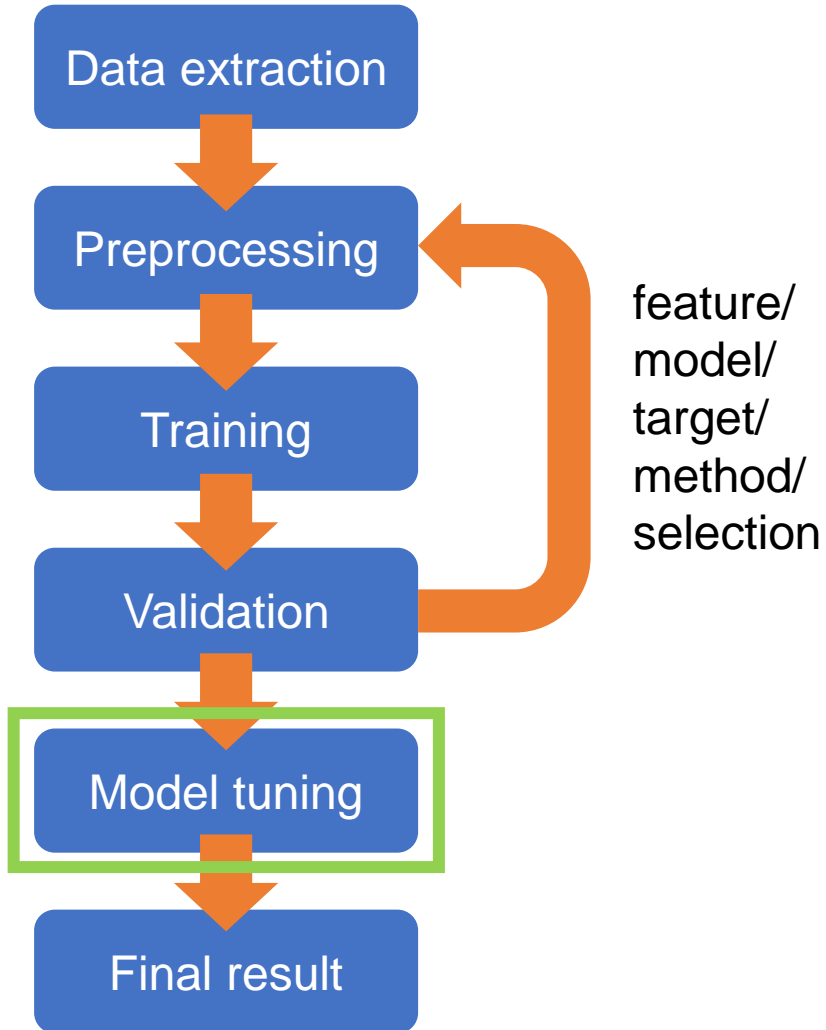
3. 3DMUON_3DSHOWER_trig_hits only



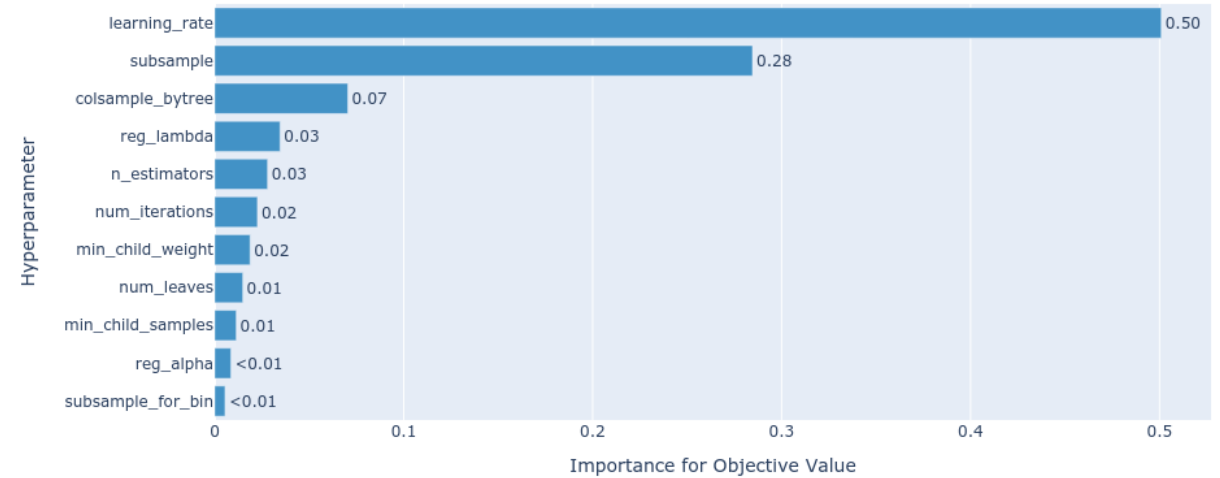
4. importance>0



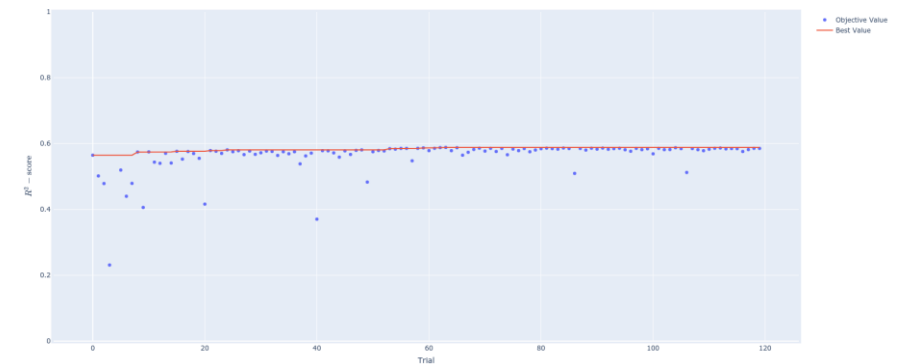
Workflow of the reconstruction:



Hyperparameter Importances

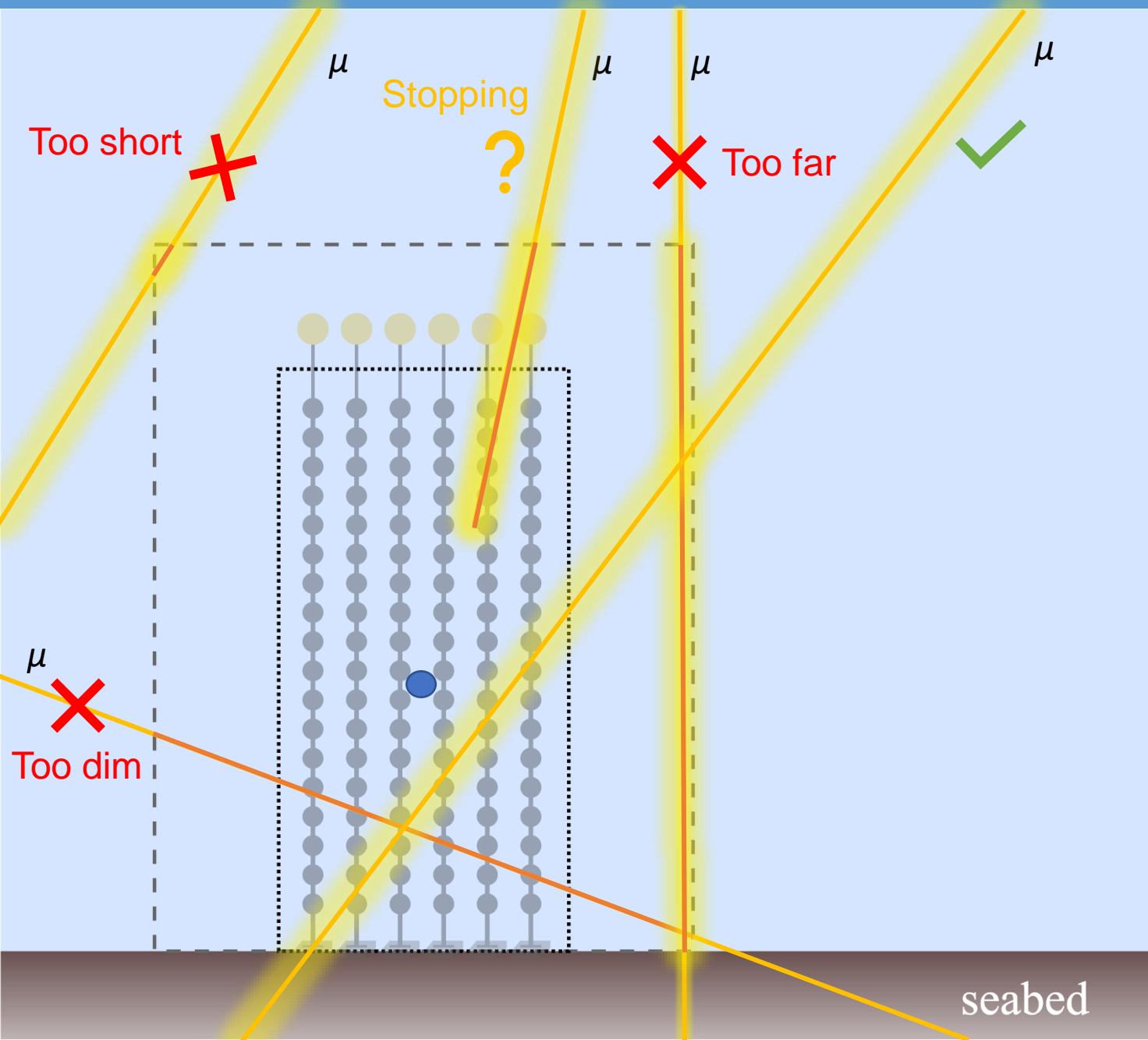


Optimization History Plot



Slice Plot





We want to exclude muons, which:

- ❖ Are too far from the detector
- ❖ Have too short pathlength inside the volume of interest
- ❖ Emit too faint light (have too low Energy)
- ❖ Basically are not visible or would be poorly reconstructed

How?

- ❖ Check the JMMuon* likelihood \mathcal{L} for single muon events against:
 - distance of muon from the DET center (●) for vertical muons \rightarrow pick an optimal volume by shrinking the can by x as:

$$r_{\text{can}} - x, h_{\text{can}} - x$$
 - muon pathlength L but for shrunked can
 - muon energy $\rightarrow E$ cut

Summary of the selection:

Detector	Minimal E_μ [GeV]	d_{\max} [m]	minimal L_μ [m]
ARCA115	120	-	-
ARCA6	120	269.4	240
ORCA115	1	-	-
ORCA6	1	-	-

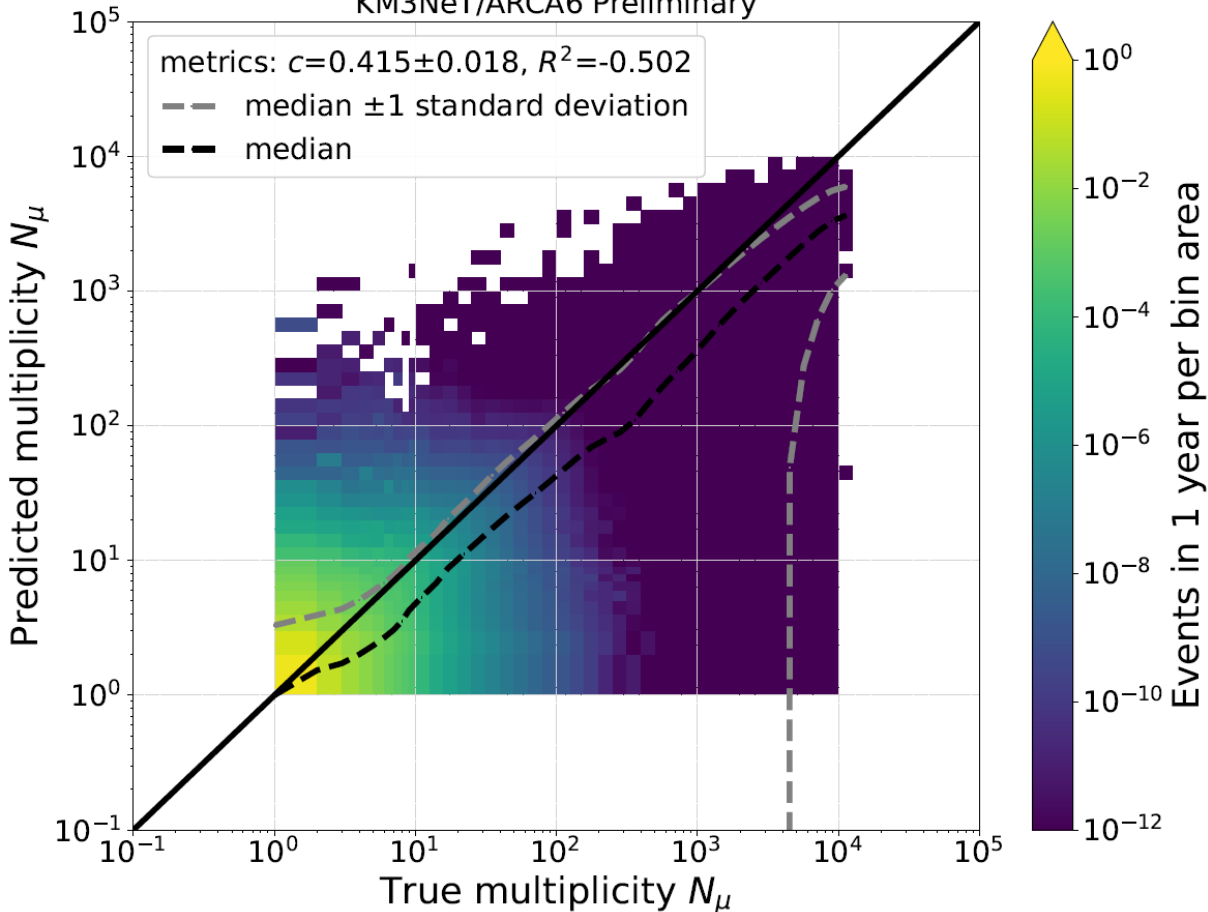
(plots in the backup)

This selection is used for further multiplicity results

Example of ARCA6, for which the effect is the most pronounced

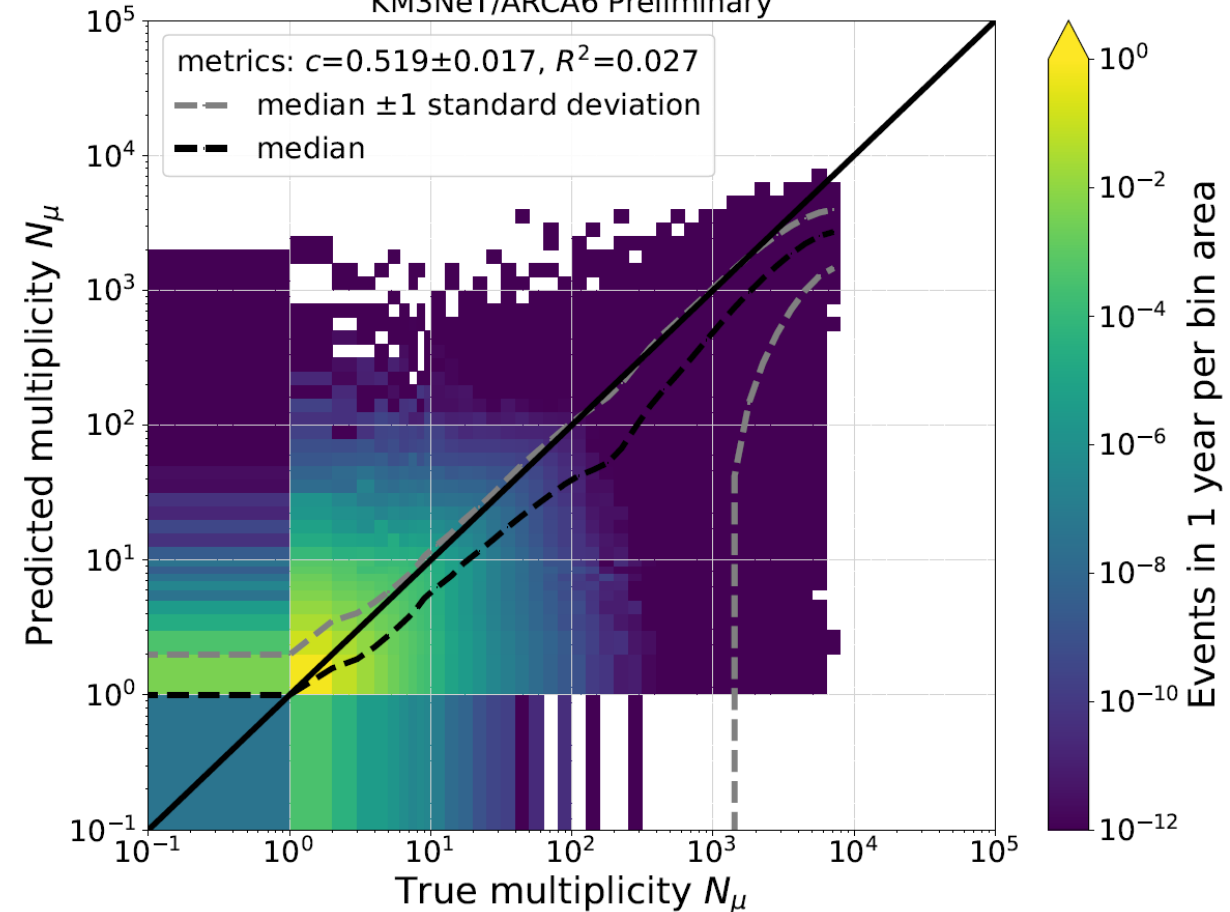
All muons

KM3NeT/ARCA6 Preliminary



Selected muons

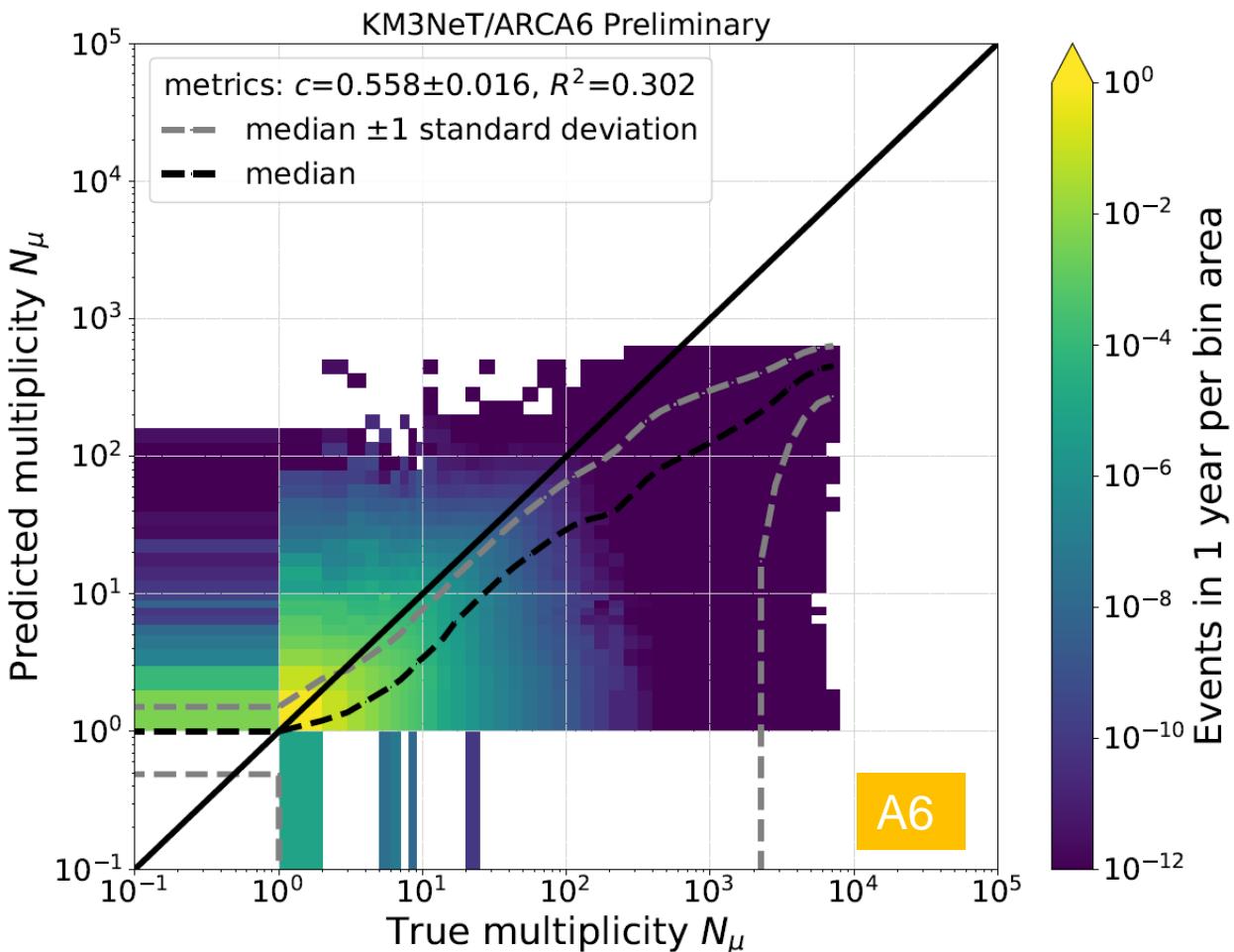
KM3NeT/ARCA6 Preliminary



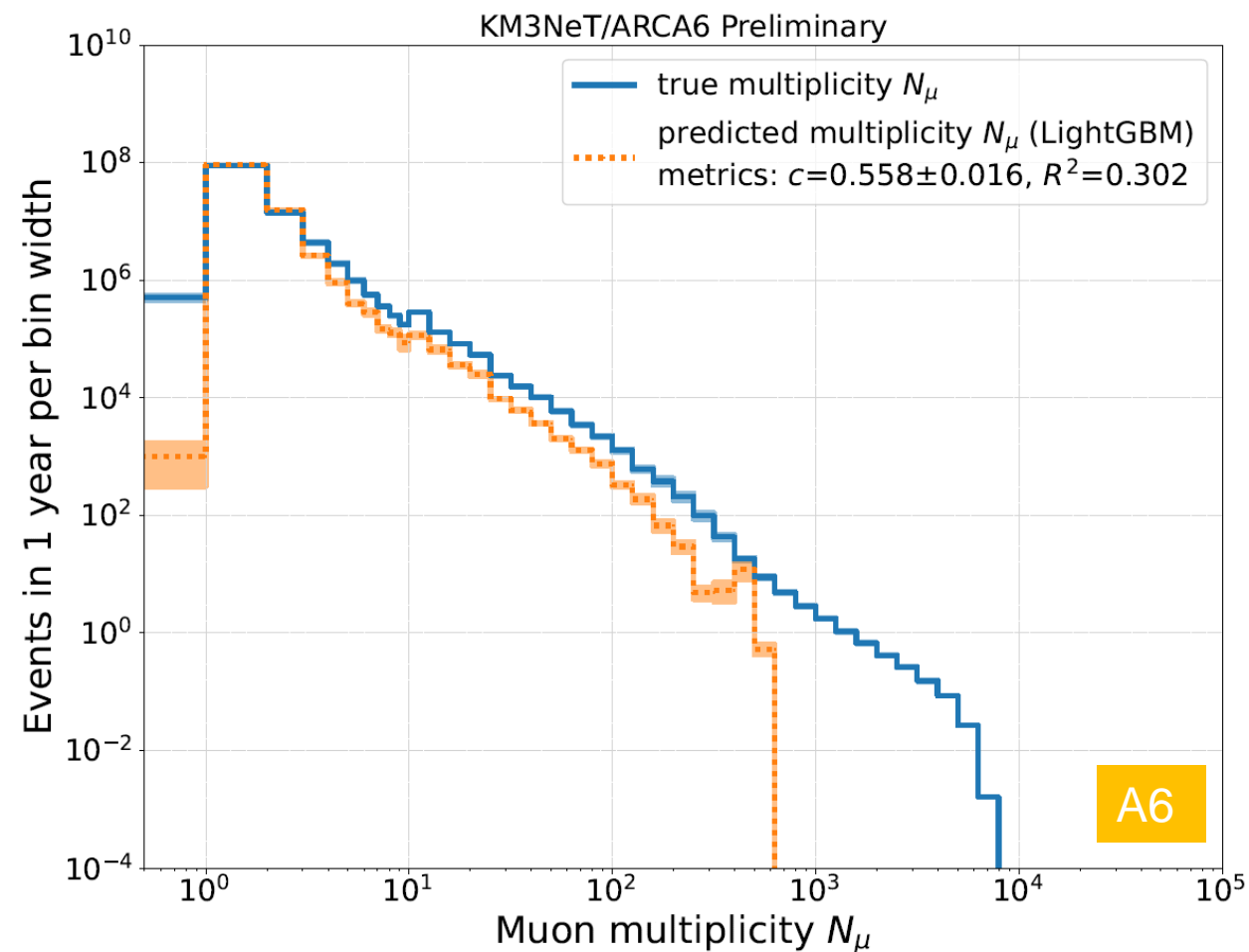
Analogical results obtained for ARCA115, ORCA115 and ORCA6

Example of the results for ARCA6:

2D: pred vs true



1D histograms



Definition of the test

Poisson formula for BGD with non-negligible uncertainty:

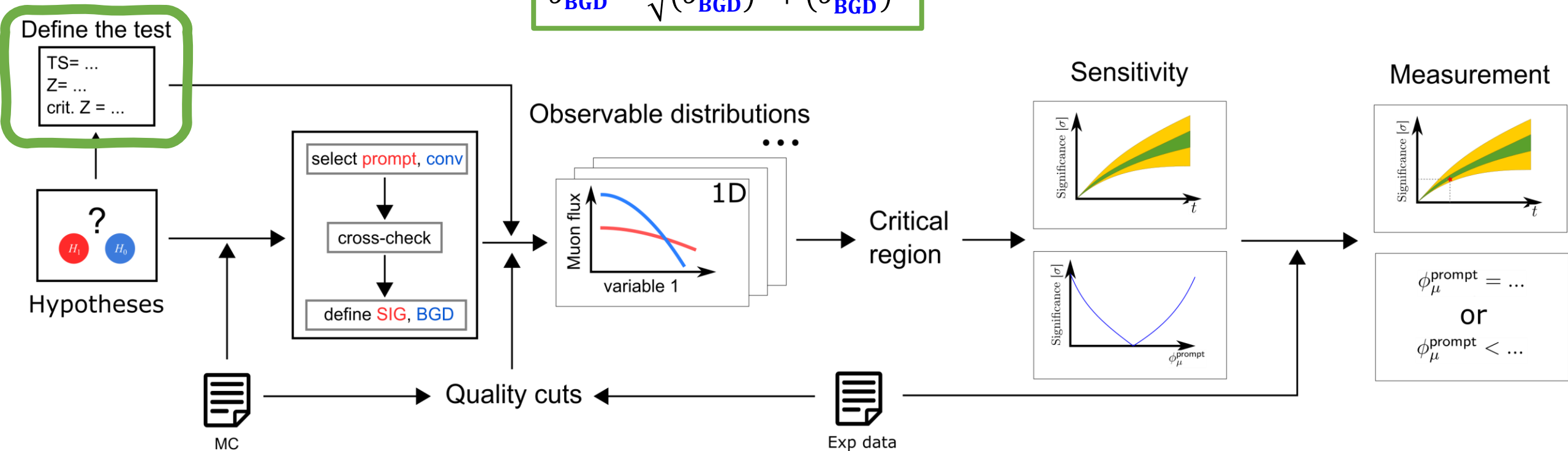
[ref1](#) [ref2](#) [ref3](#)

$$TS: q_0 = \begin{cases} 2 \cdot \left[N_{TOTAL} \cdot \ln \left(\frac{N_{TOTAL} \cdot (N_{BGD} + \sigma_{BGD}^2)}{N_{BGD}^2 + N_{TOTAL} \cdot \sigma_{BGD}^2} \right) - \frac{N_{TOTAL}^2}{\sigma_{BGD}^2} \cdot \ln \left(1 + \frac{\sigma_{BGD}^2 \cdot (N_{TOTAL} - N_{BGD})}{N_{BGD} \cdot (N_{BGD} + \sigma_{BGD}^2)} \right) \right] & \text{for } N_{TOTAL} \geq N_{BGD} \\ 0 & \text{for } N_{TOTAL} < N_{BGD} \end{cases}$$

Significance: $Z = \sqrt{q_0}$
 Critical Z: 5σ

systematic uncertainties are included:

$$\sigma_{BGD} = \sqrt{(\sigma_{BGD}^{stat})^2 + (\sigma_{BGD}^{syst})^2}$$



Definitions:

- ❖ **prompt**: μ „directly” from 1st interaction:
 - Possibly few parent particles
 - Parents with lifetime $\tau < \tau_{K_S^0}$
- ❖ **conventional**: all other μ

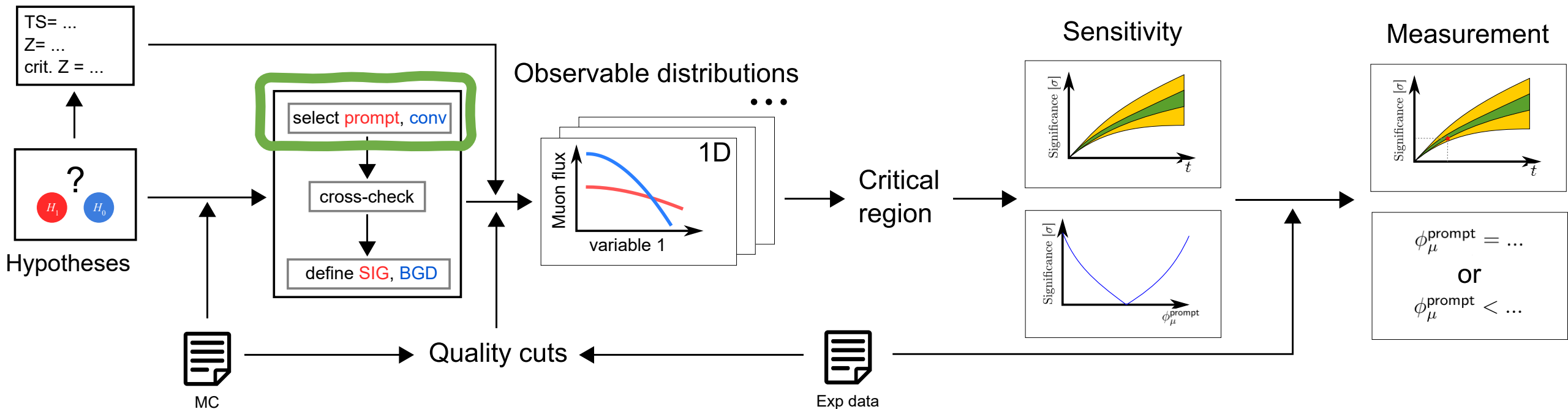
MC
prompt μ :

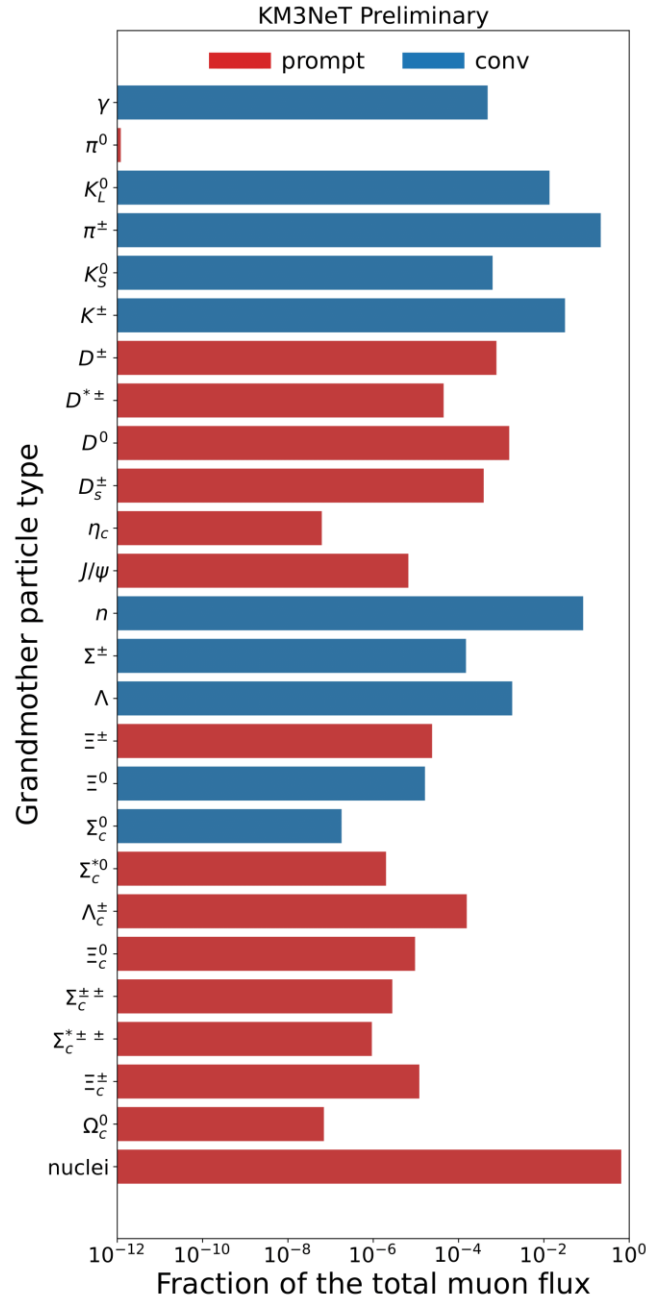
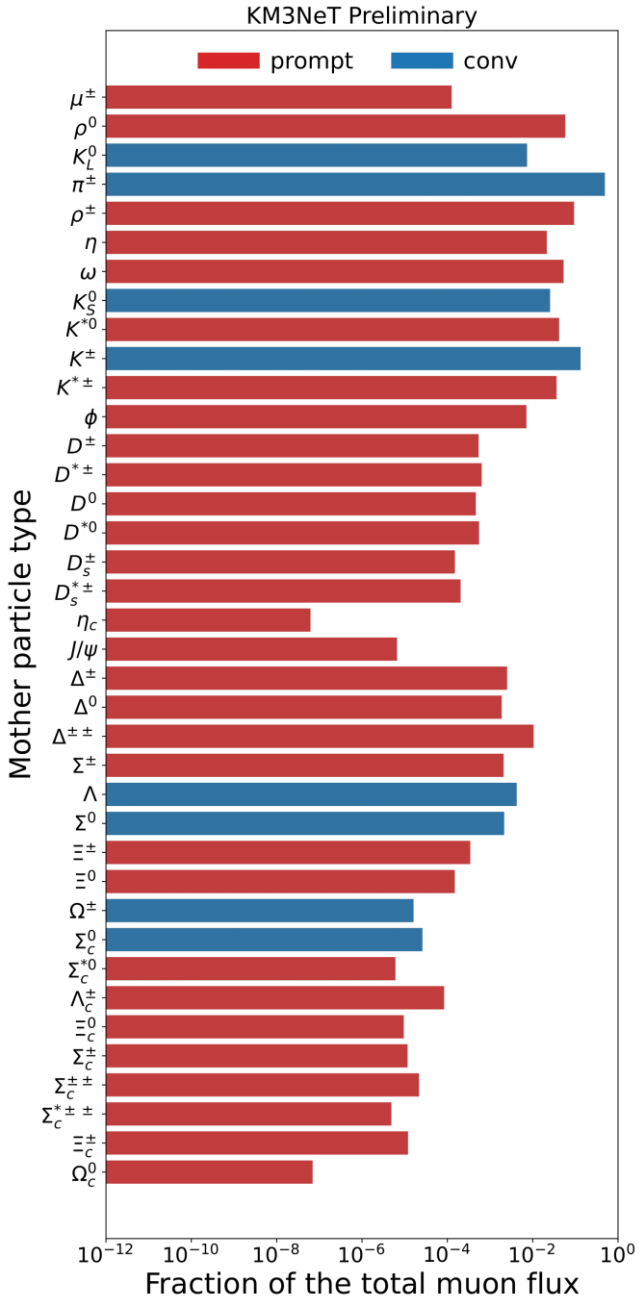
(hadronic counters)

gmom	mom	h_{mom}	h_{μ}	EM counters
prim	muon	1	1	0
prompt	muon	any	any	0
prim	prompt	≤ 2	≤ 2	0
prompt	prompt	2	32	0
		3	33	0

$\tau_{K_S^0} = 89.5\text{ps}$

Define the test





Note: 1 parent **conventional** \rightarrow the muon is **conventional**.

The colours here only tell you if particles have short or long lifetimes (if applicable).

Most muons originate from π^\pm and K^\pm , as expected.

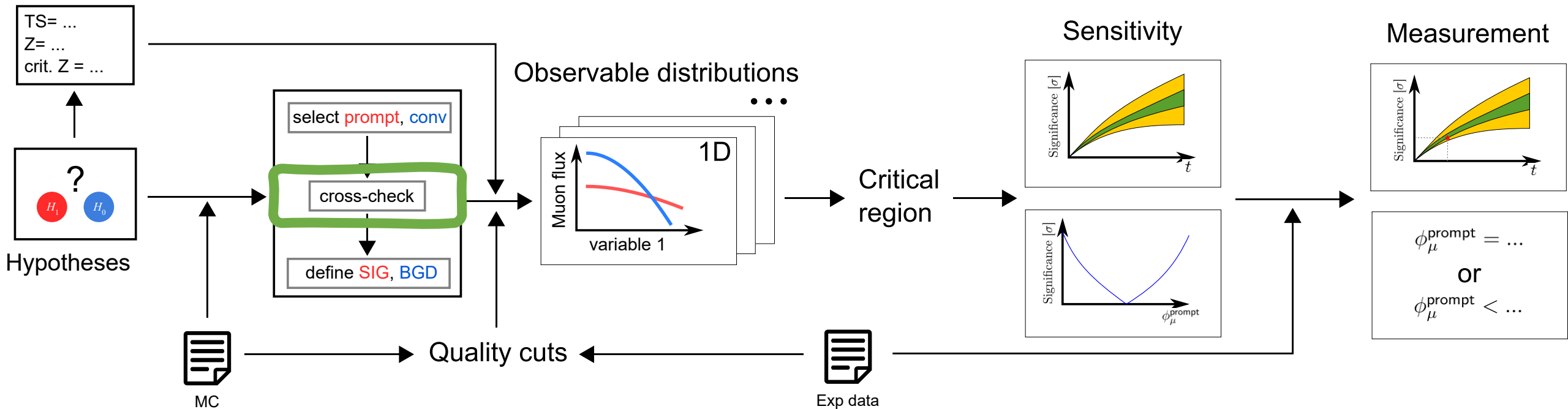
The most important **prompt** mother particles for muons are light vector mesons (η, ρ, ω), not D mesons (also expected).

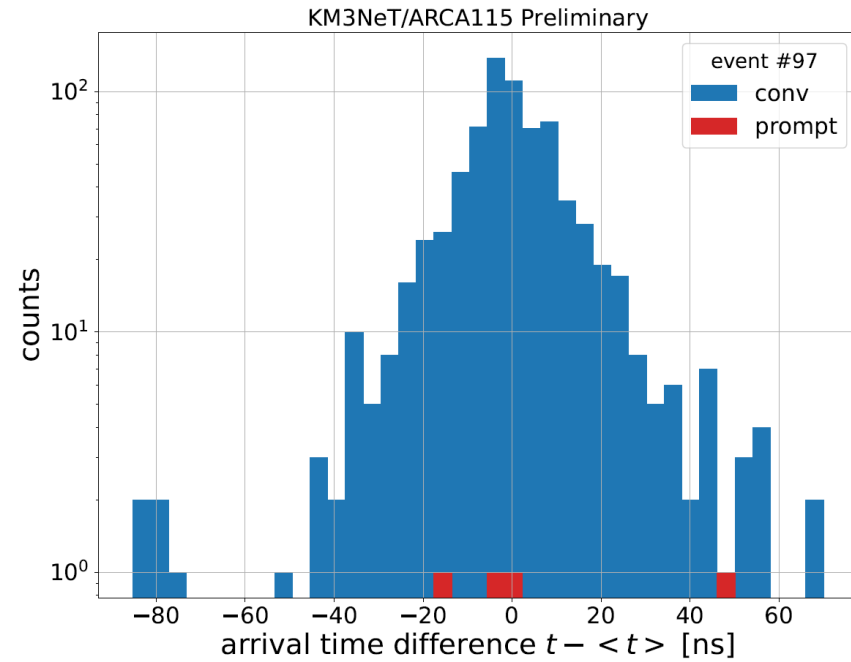
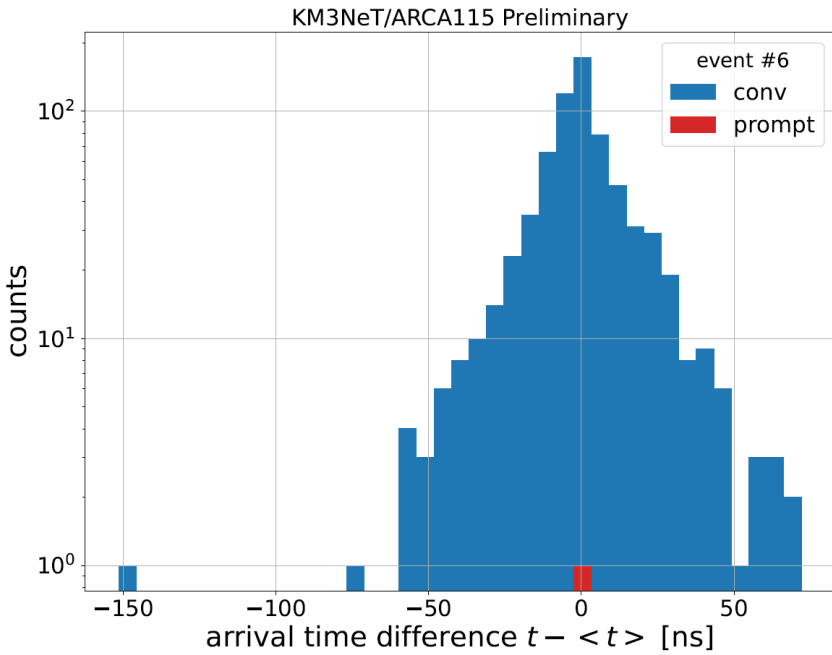
If mother is a muon or grandmother is the same nucleus as the primary, it means that there were just less interactions between shower start and muon creation.

NB: particles & antiparticles are counted together! (and so are all nuclei, including hydrogen)

- I look at 3 things:
1. Muon arrival time
 2. Muon energy share
 3. Muon production point

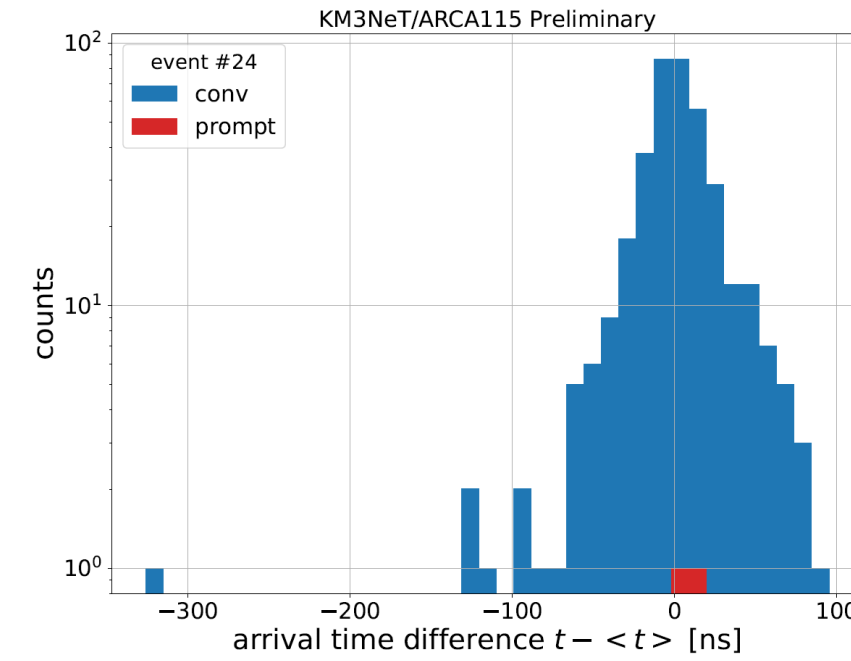
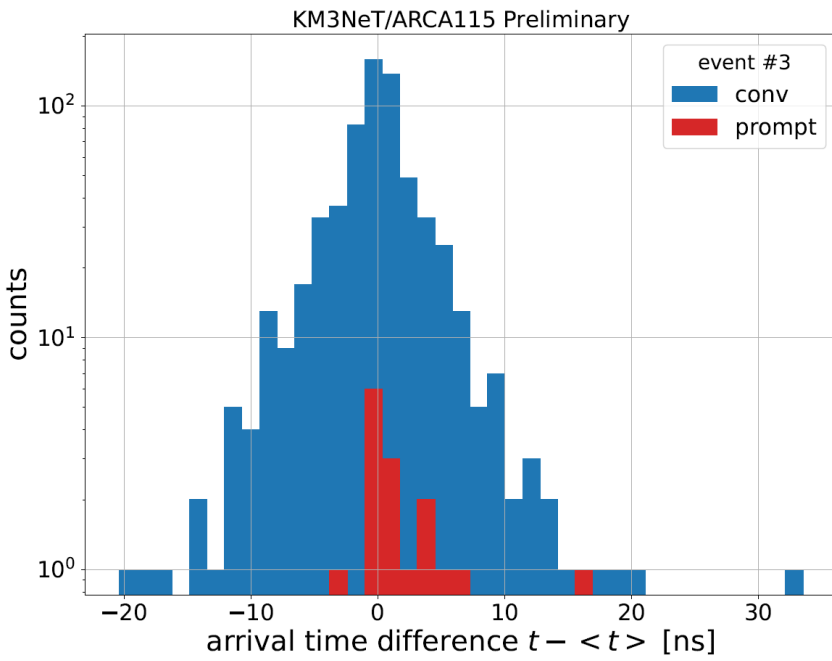
Define the test





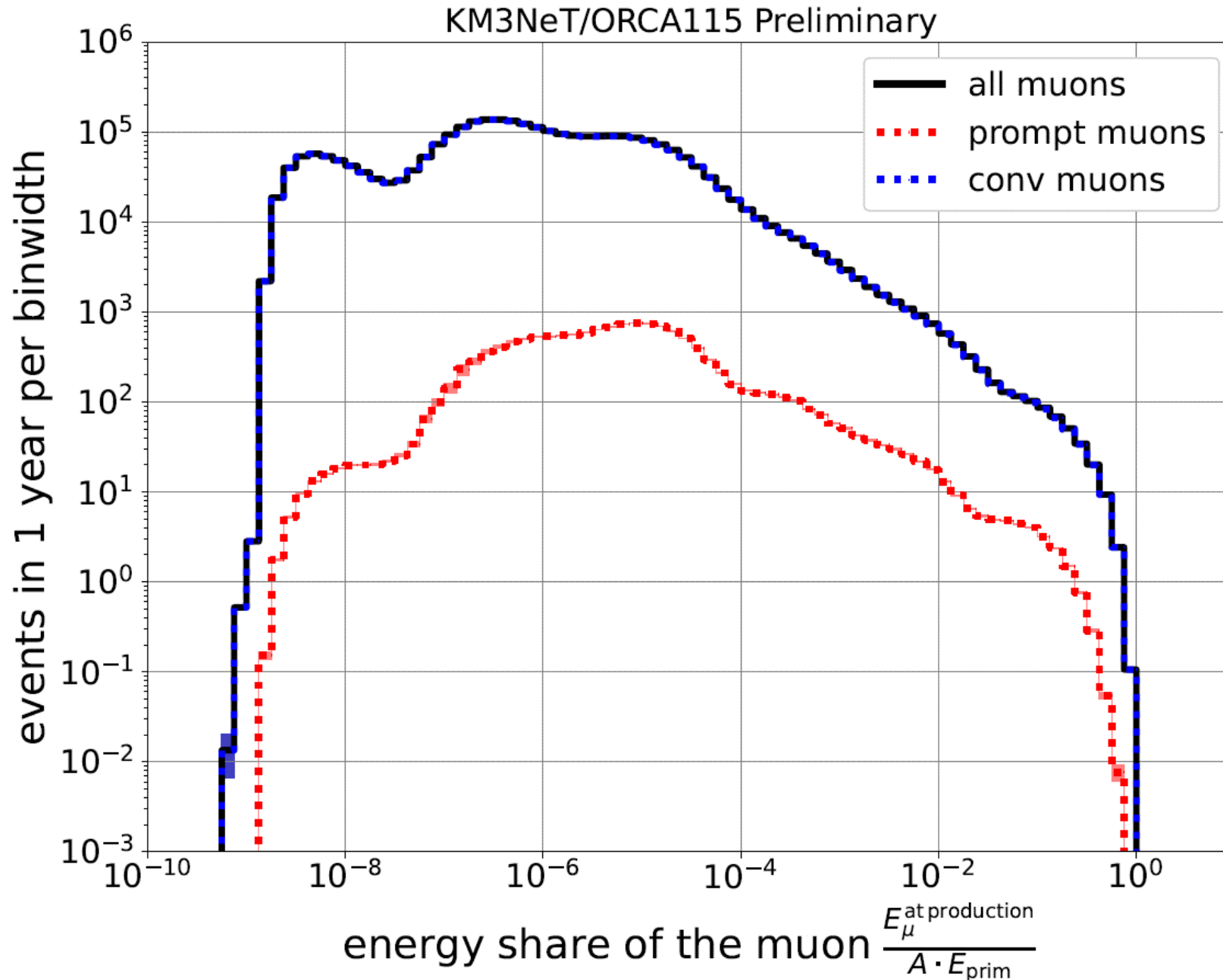
Proton events

arrival time: time between the first interaction of the primary and the muon crossing the can boundary



Iron events

Conclusion here is that **prompt** is not really evident from arrival times on event-by-event basis (which is a bummer, because this could have been measurable)



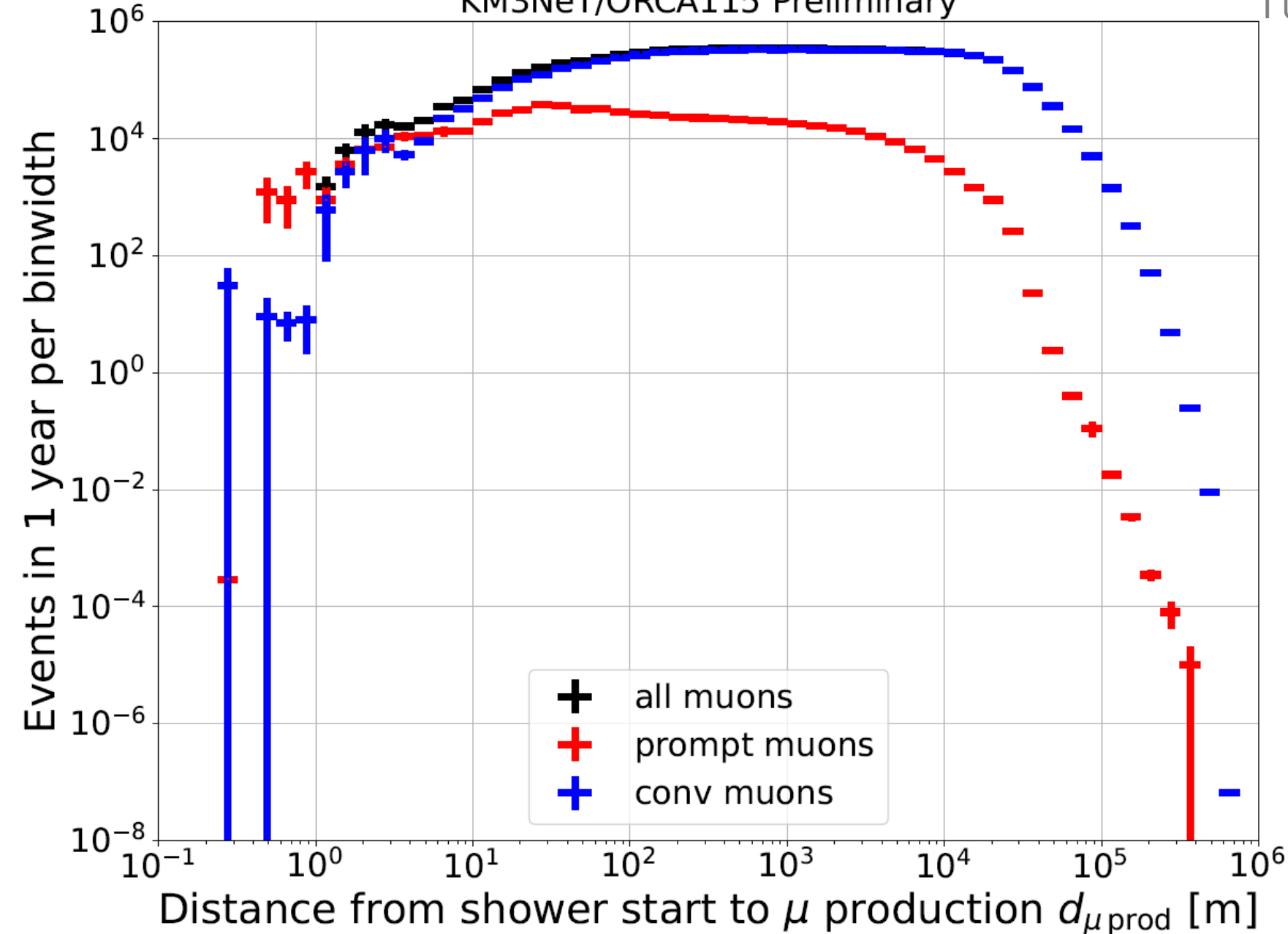
I use ORCA115 to boost the statistics

Prompt muons indeed tend to carry a larger portion of the total primary energy

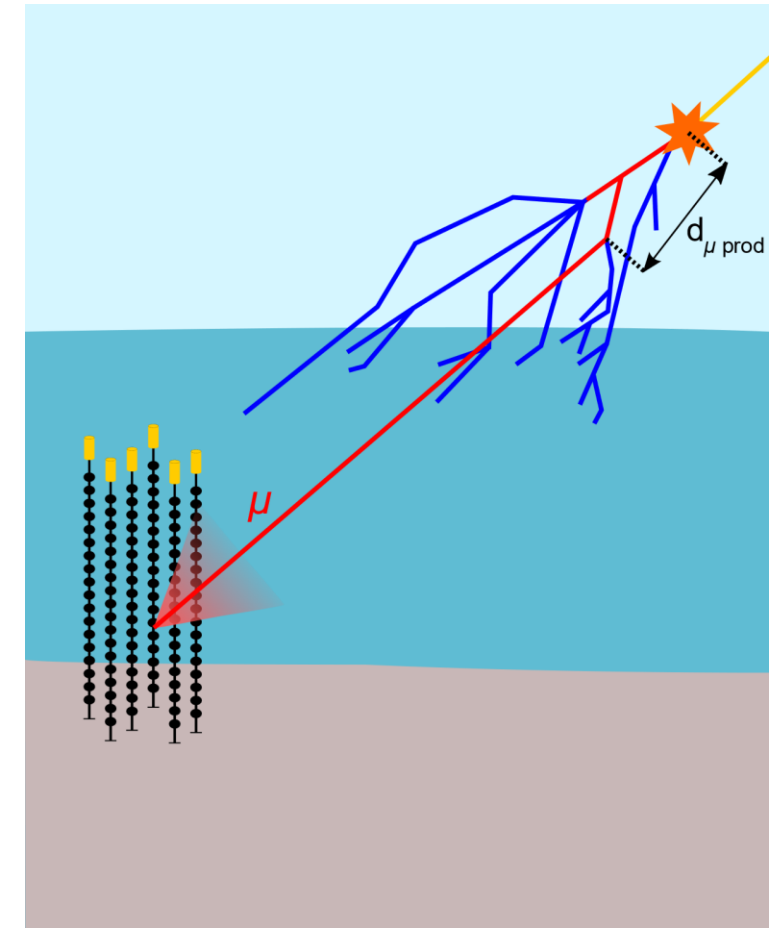
The wiggles are coming from the contributions of different primaries

KM3NeT/ORCA115 Preliminary

I use ORCA115 to boost the statistics



Prompt muons indeed are more often produced close to the 1st interaction

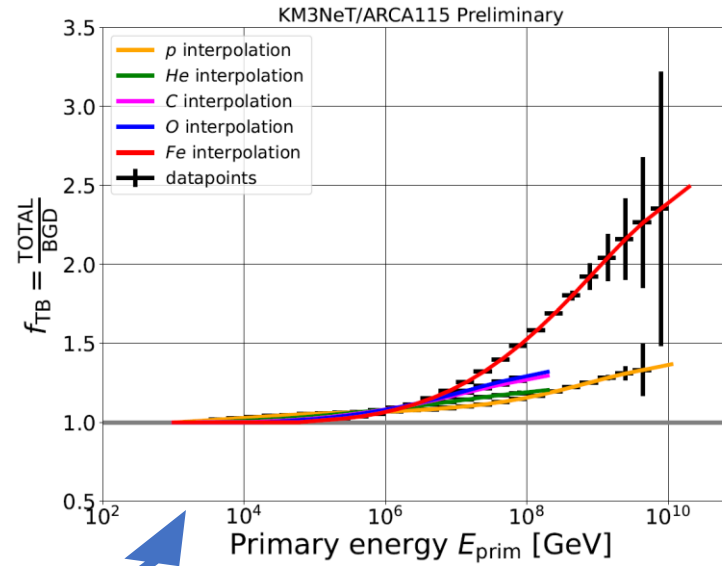


Definitions:

- ❖ **SIG**: ≥ 1 prompt μ in bundle
- ❖ **BGD**: 0 prompt μ in bundle

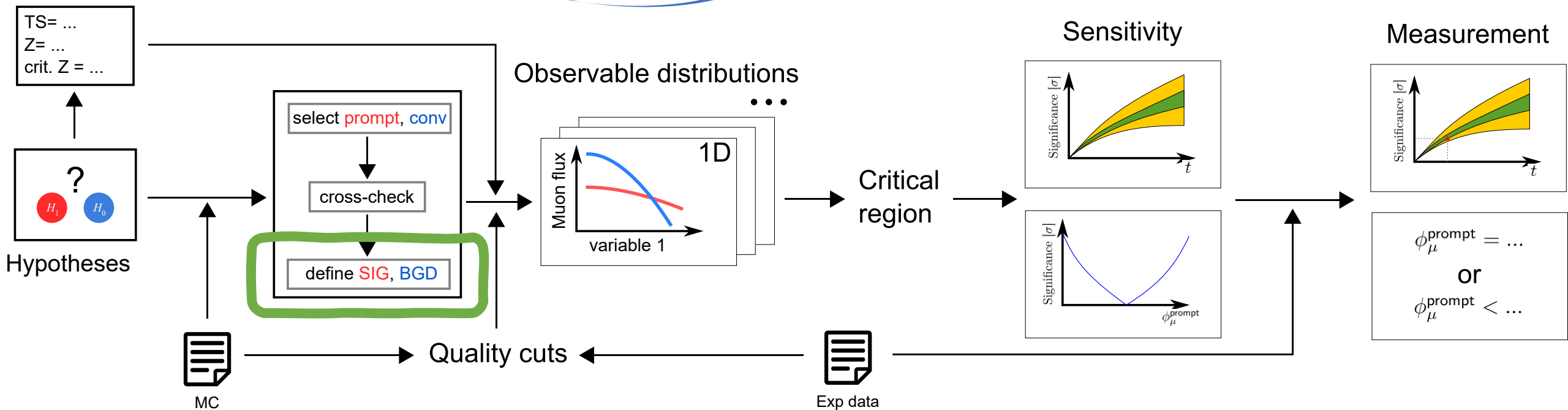
BGD has to be reweighted:

$$w_{\text{BGD}} = w_{\text{event}}(\text{prim}, E_{\text{prim}}) \cdot f_{\text{TB}}(\text{prim}, E_{\text{prim}})$$



Note:
Ideally, f_{TB} should be evaluated at generation, but that's not possible (or we could use 2 separate MC's, but it's not possible yet)

Define the test



The quality cuts:

❖ Likelihood cut:

ARCA: $\mathcal{L} > 50$

ORCA: $\mathcal{L} > 280$

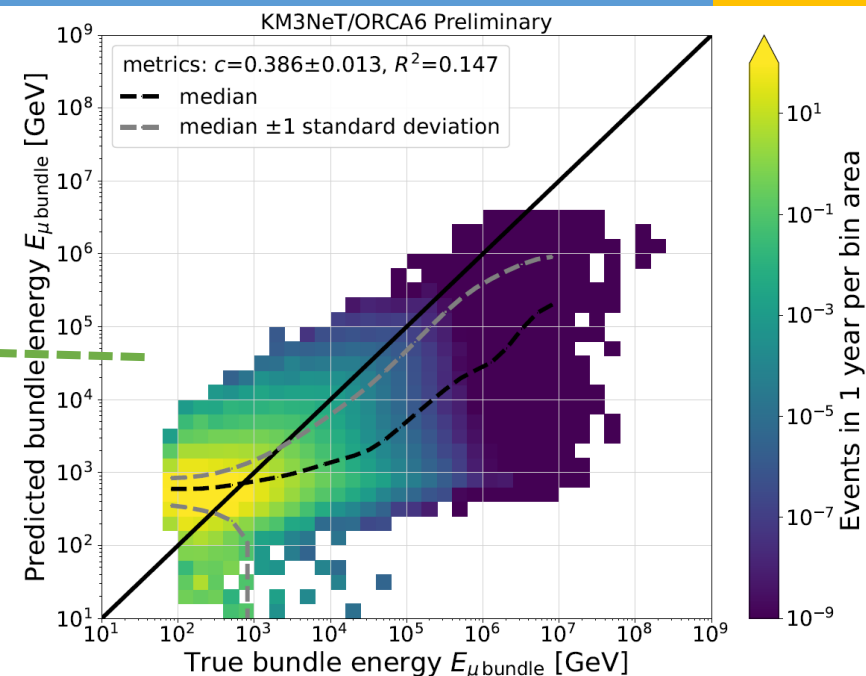
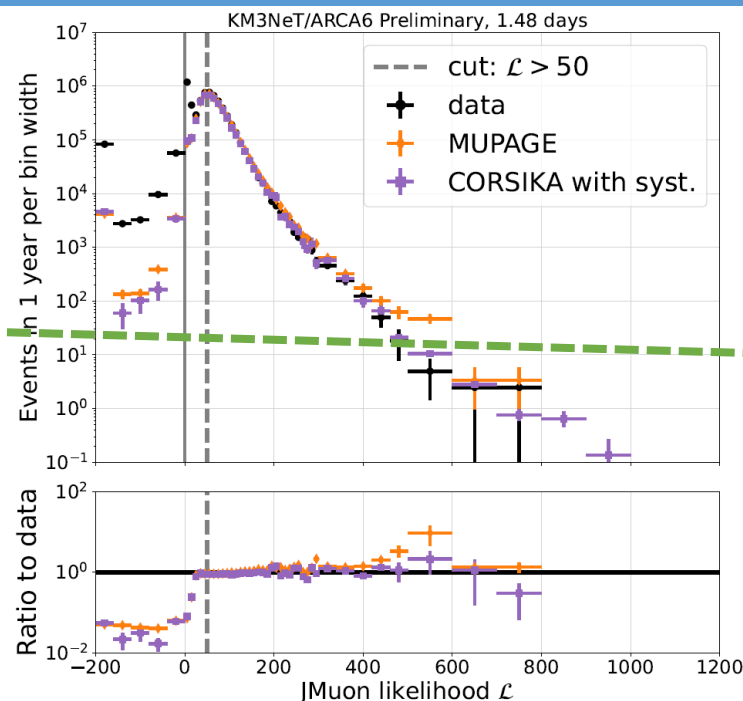
❖ ML reco reliability cut:

ARCA115: $E_{\text{bundle}} > 3\text{TeV}, N_{\mu} > 0$

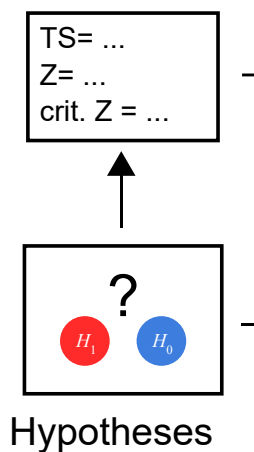
ARCA6: $E_{\text{bundle}} > 3\text{TeV}, N_{\mu} > 0$

ORCA115: $E_{\text{bundle}} > 300\text{GeV}, N_{\mu} > 0$

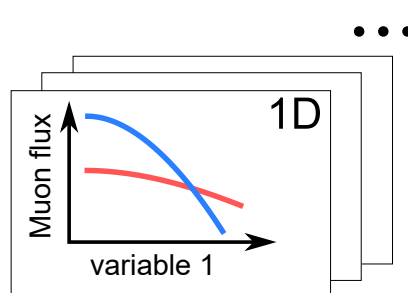
ORCA6: $E_{\text{bundle}} > 800\text{GeV}, N_{\mu} > 0$



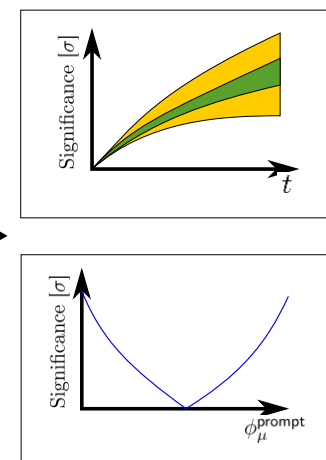
Define the test



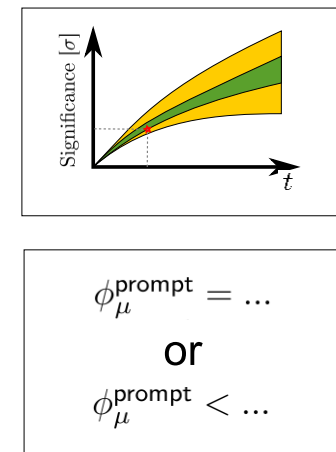
Observable distributions



Critical region



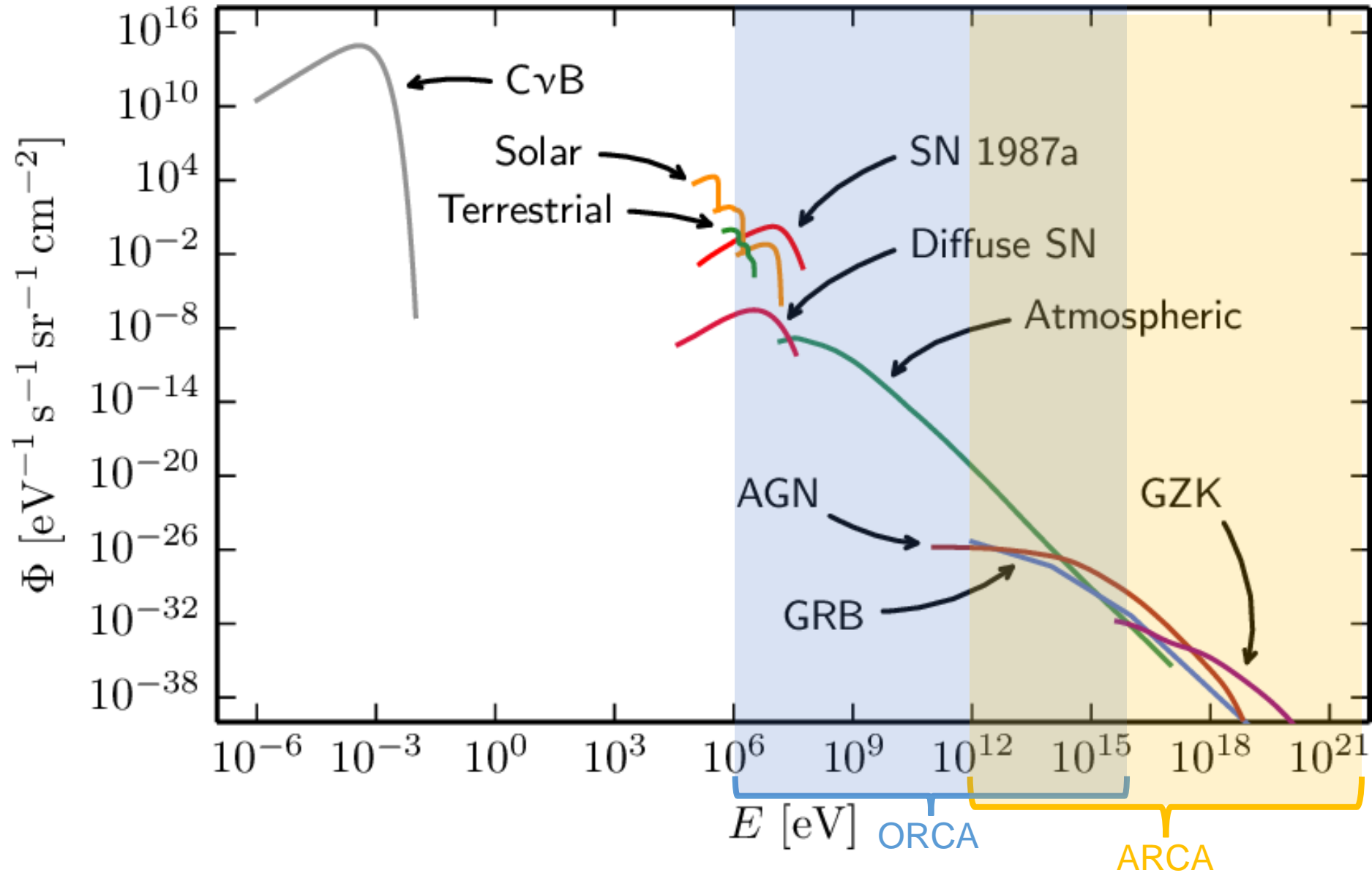
Measurement



MC

Quality cuts

Exp data

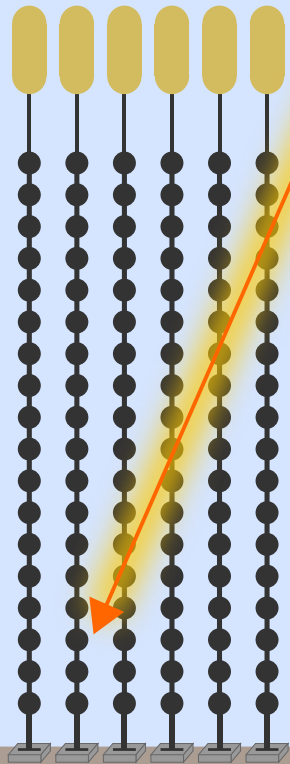


Examples of basic event topologies:

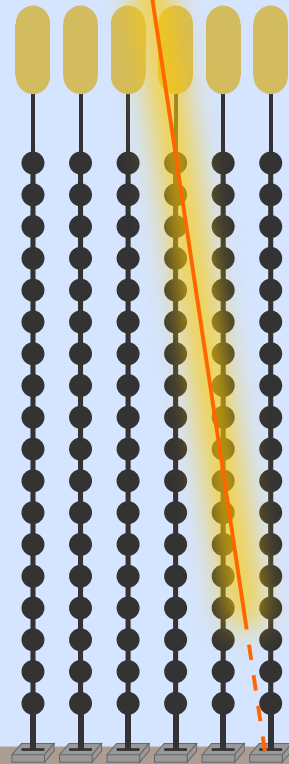
Classes based on combinations of:

- ❖ Direction
- ❖ Shape

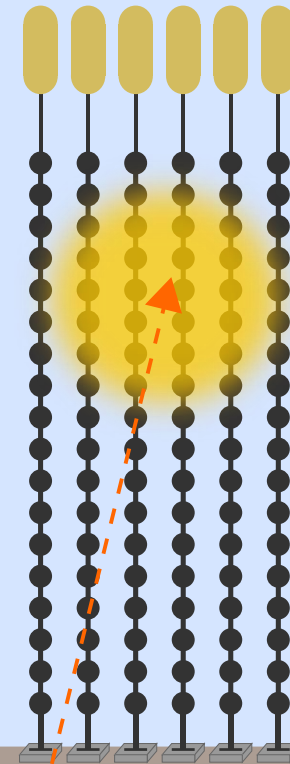
down-going track
(typically atm. μ)



up-going track
(typically atm. ν_μ)



single cascade
(typically atm. ν_e/ν_τ)



double cascade
(typically atm. ν_τ)

