KM3NeT Data processing and Monte Carlo simulations

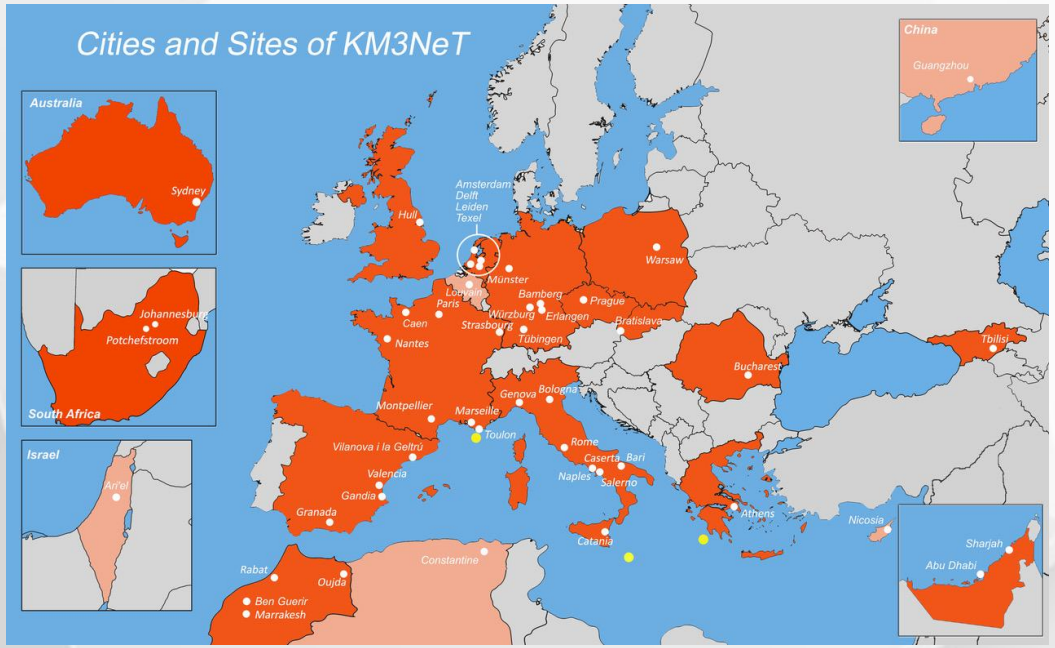
The path to cosmic neutrino event identification

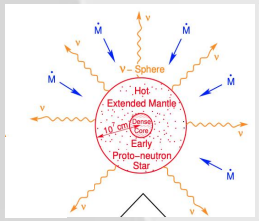
A.Sinopoulou
anna.sinopoulou@ct.infn.it

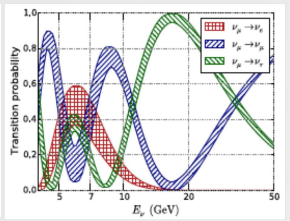
Espresso talk, 22.11.2023

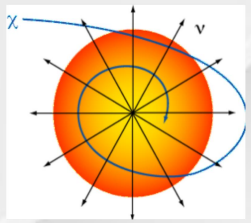
KM3NeT

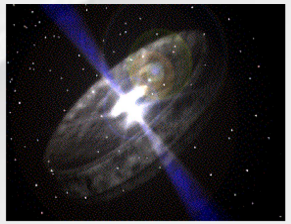
KM3NeT is **1 collaboration** contructing **2 neutrino detectors** in the Mediterranean Sea, based on **1 technology**.











Physics studies in a wide energy range
from MeV to PeV

KM3NeT/ARCA:
observation of high energy neutrinos (GeV - TeV)

KM3NeT/ORCA:
determination of the neutrino mass hierarchy (MeV)

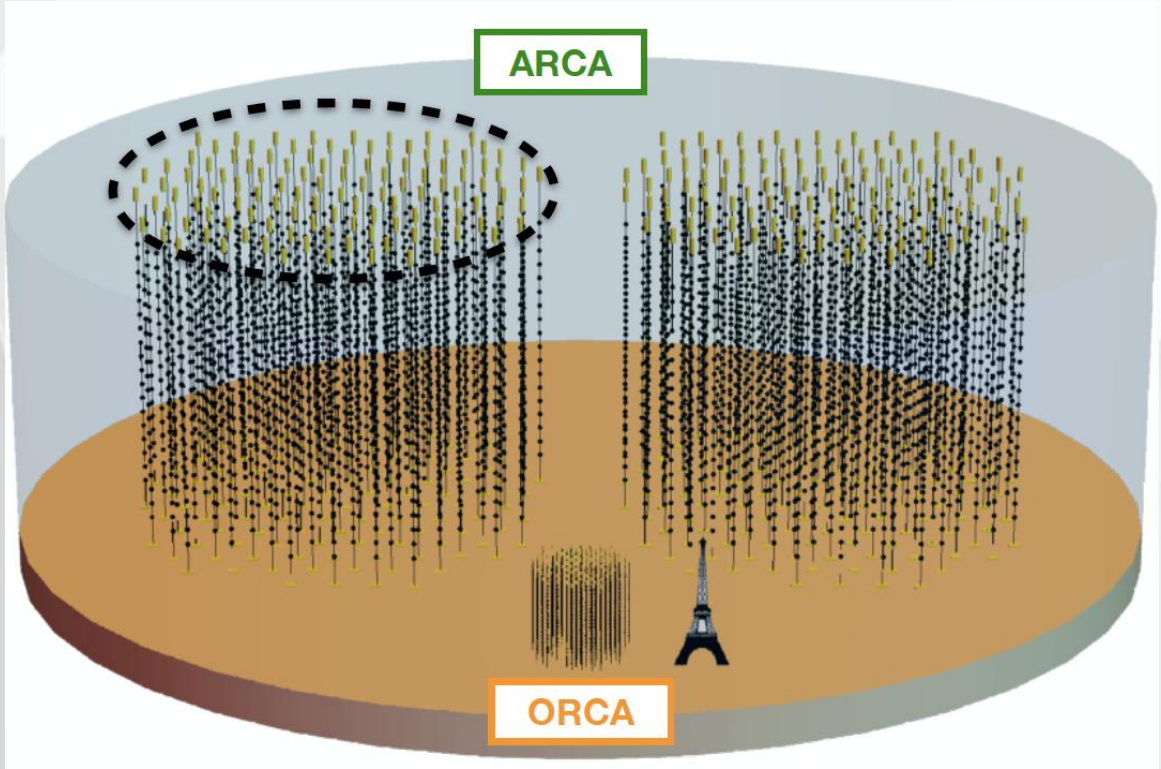
KM3NeT innovative element
Digital **O**ptical **M**odule (DOM)

Each DOM is a small detector instrumenting 31 3'' PhotonMultiplier tubes (PMTs)

☺4π signal coverage
☺excellent angular resolution

KM3NeT detector configurations

Different objectives -> different geometries
BUT same technology !!!



*115 DUs per building block

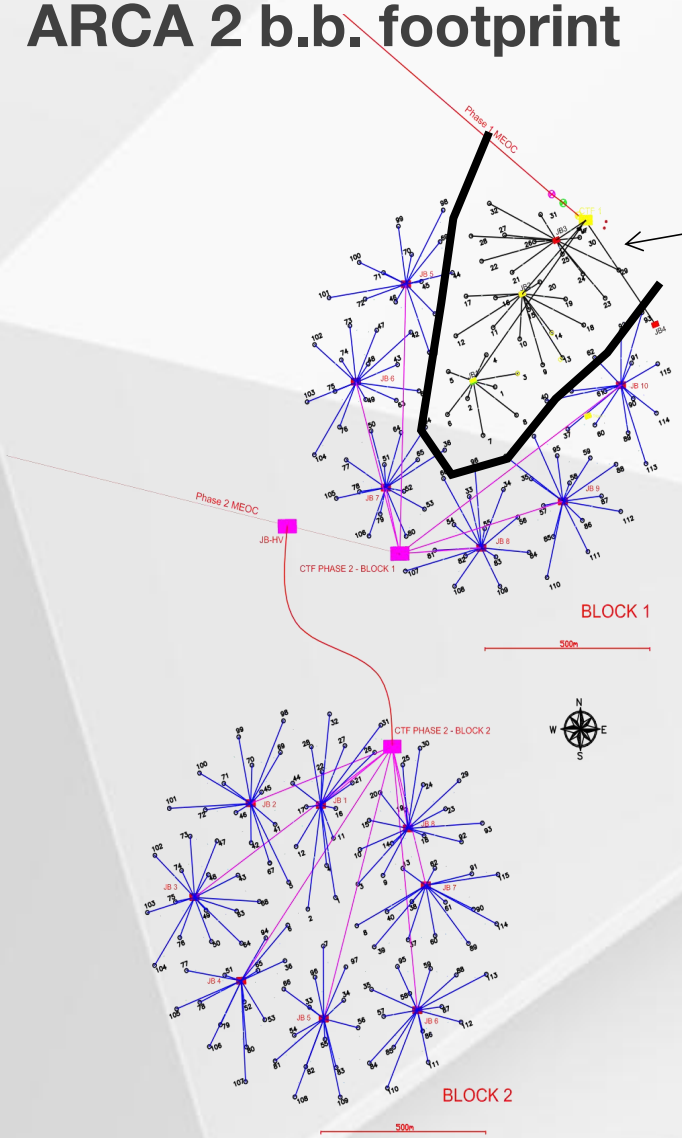
Detection Unit (DU)

DOM

	ORCA (denser)	ARCA (larger)
Eff. Mass	~ 7 Mt	~ 1 Gt
Line length	200 m	650 m
Interline distance	20 m	90 m
Vertical spacing	9 m	36 m
Depth	~ 2500 m	~ 3500 m

KM3NeT detector status

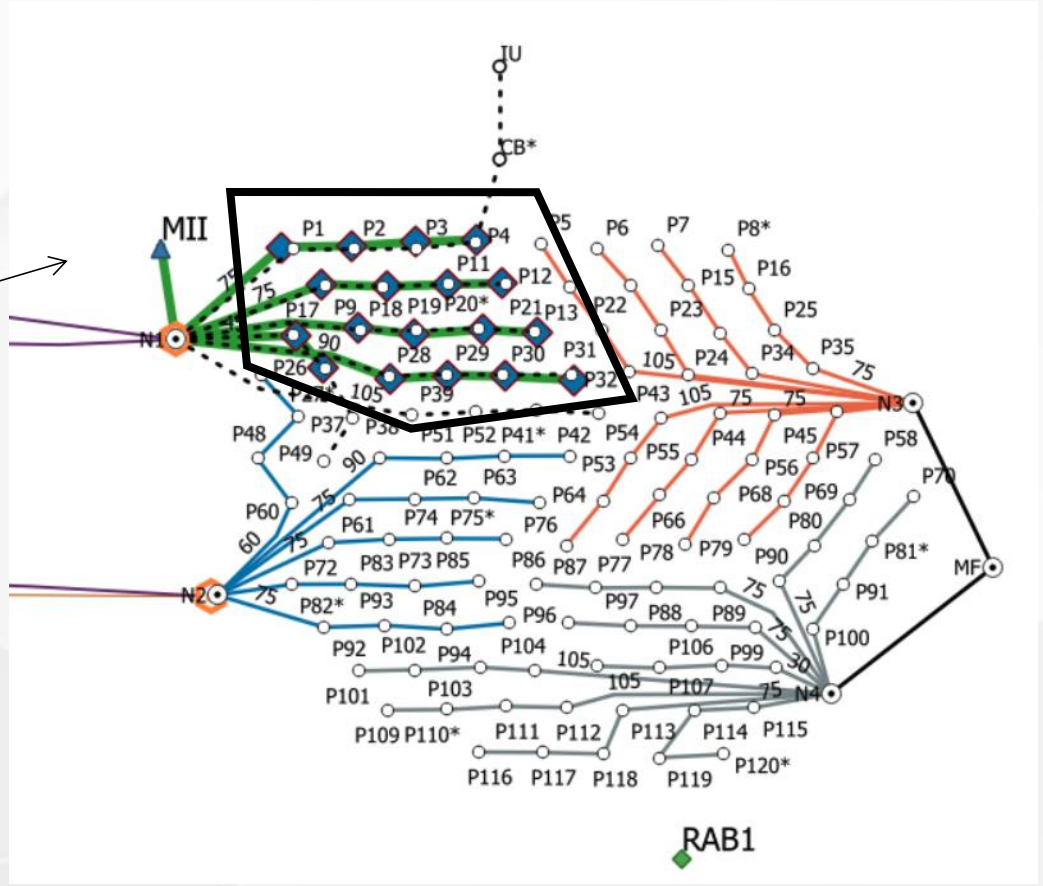
ARCA 2 b.b. footprint



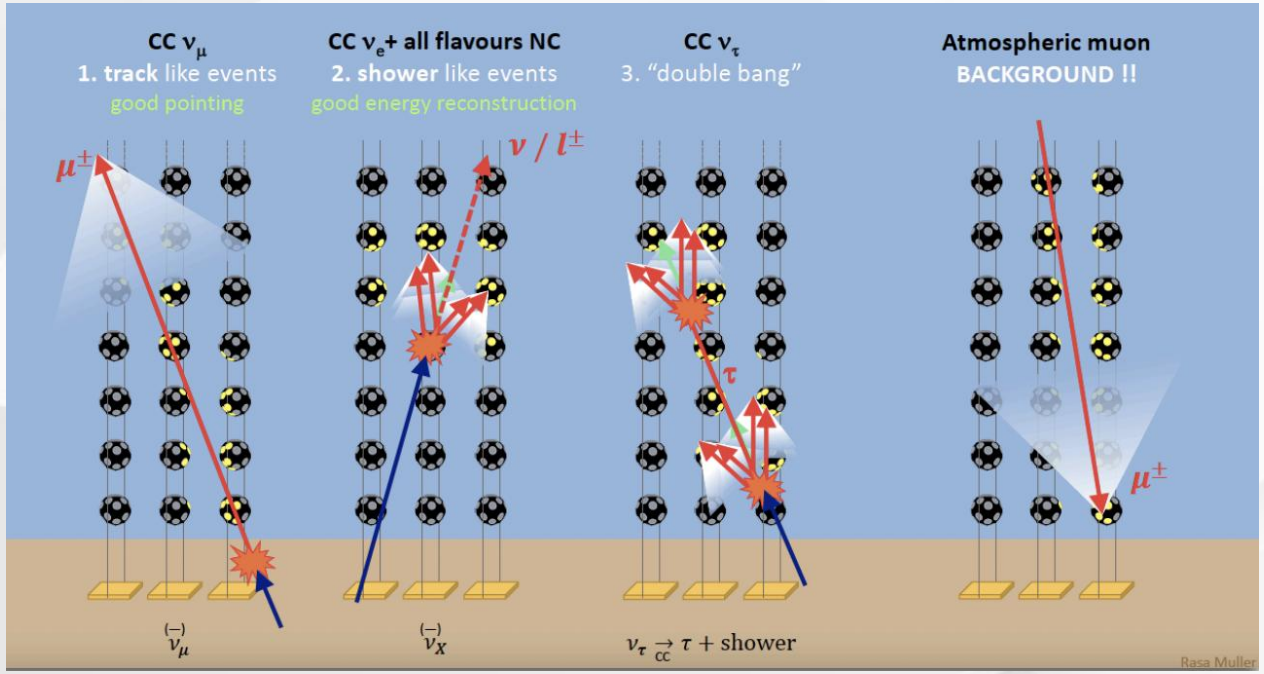
Current data-taking ARCA configuration
28 DUs deployed

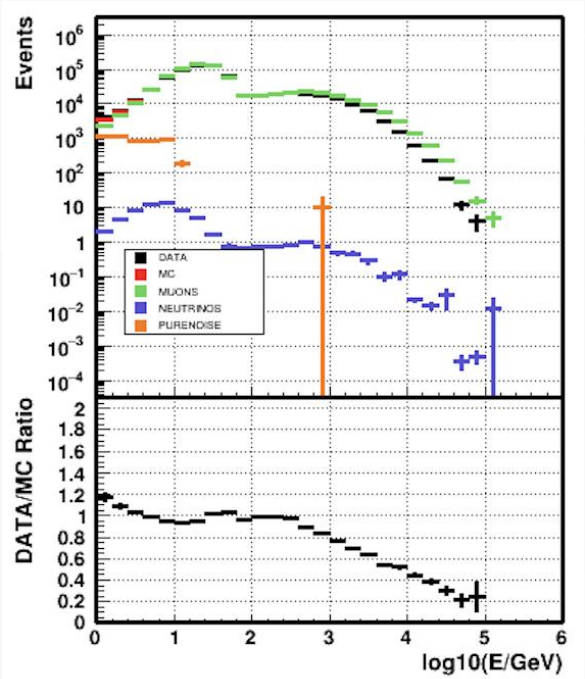
Current data-taking ORCA configuration
18 DUs deployed

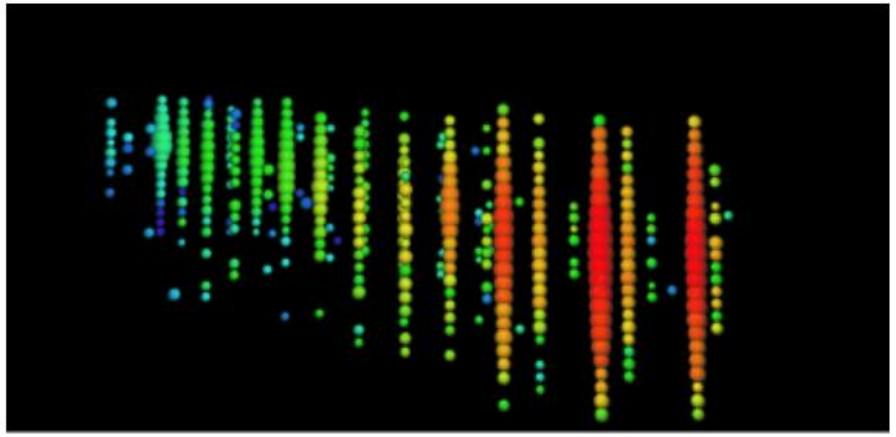
ORCA b.b. footprint



Data from detection to discovery...

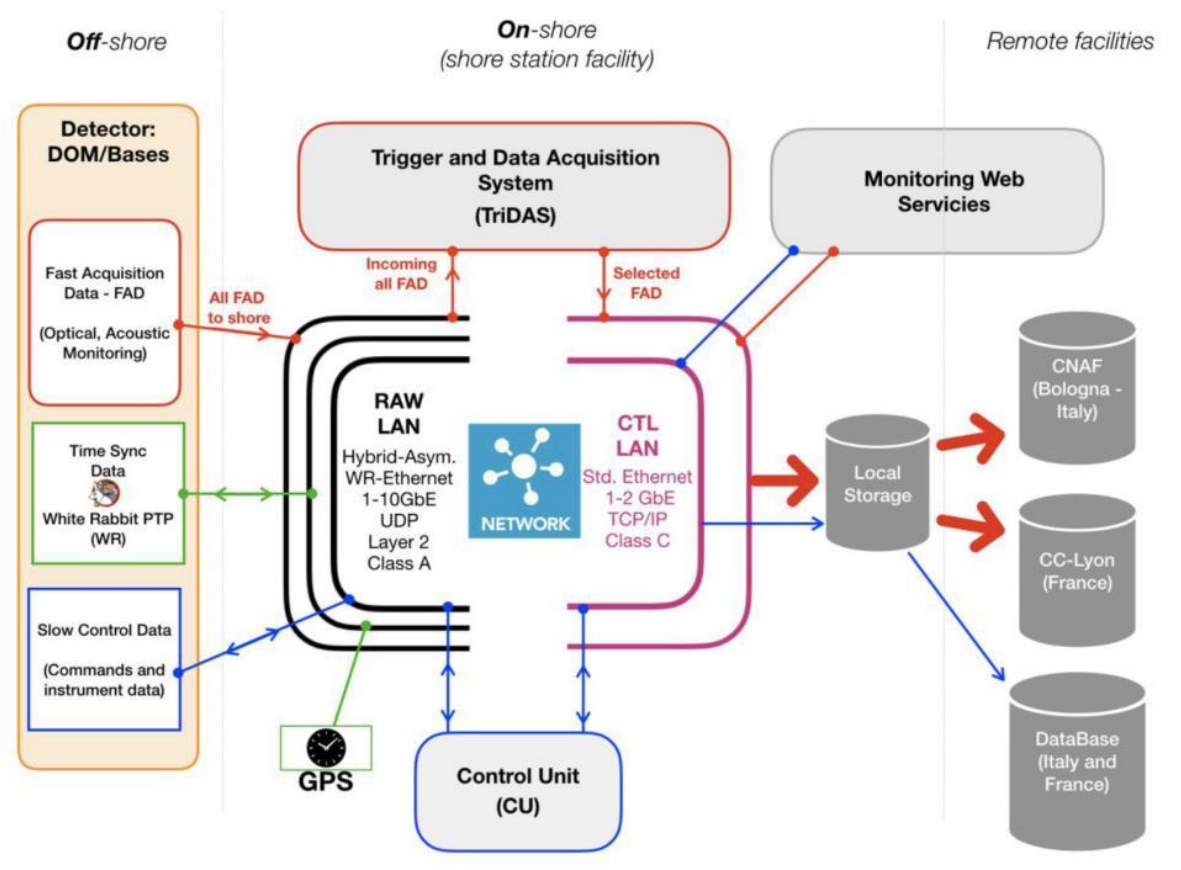






IceCube real-time alert IC170922A was a muon neutrino with an estimated energy of 290 TeV. It pointed within 0.06 degrees at the active galaxy TXS0506+06 located four billion light-years from Earth.

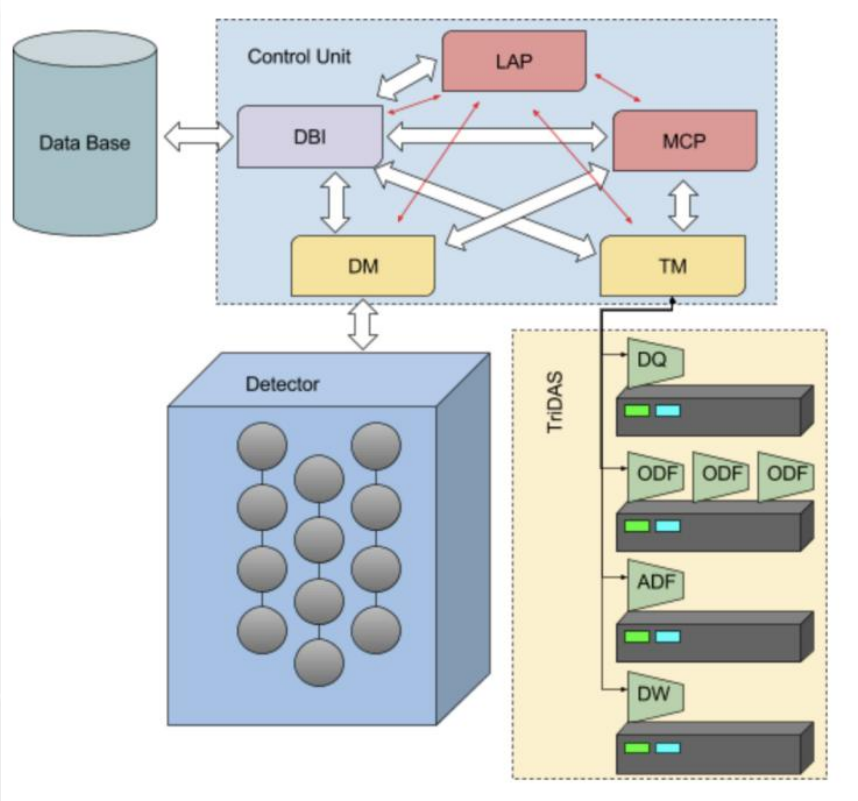
DAQ - how are data recorded and distributed to researchers ?



signal-to-noise ratio extremely disfavoured :
muon rate (atmospheric dominating) : O(100) Hz/km^3
40K decays (~constant) : O(10) kHz/PMT
Bioluminescence (occasional) : O(100) kHz/PMT

- complex DAQ structures in extreme conditions
  (mandatory: minimal underwater complexity)
- different data streams: optical, acoustic, etc.

“All-data-to-shore” principle
(a.k.a. trigger-less streaming readout)



Calibrations - essential info to “correctly” reconstruct the events

In order to achieve the physics goals, we need to ensure:

- **Recording and registering very fast events, events happening in <ms.**
- Event reconstruction relies on **good time-position relative accuracy** within the detector.

GEOMETRY: **PMT positions**

Nominal positions: “Ideal” detector positions

ROV updated positions: “Actual” detector positions after deployment (GPS)

Compass corrections: Detector orientation

Acoustic Positioning System correction: based on hydrophones and emitters

Compass + Acoustic Positioning System: Fit of the DU position

Muon crosschecks: track quality etc...

e⁻

TIME (T0): PMT time constants + efficiencies

Dark Room:

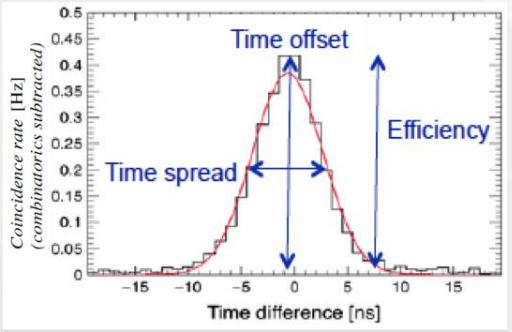
- intra-DOM, efficiencies (K40)
- inter-DOM crosschecks/tests with nanobeacons, muons (in-situ)

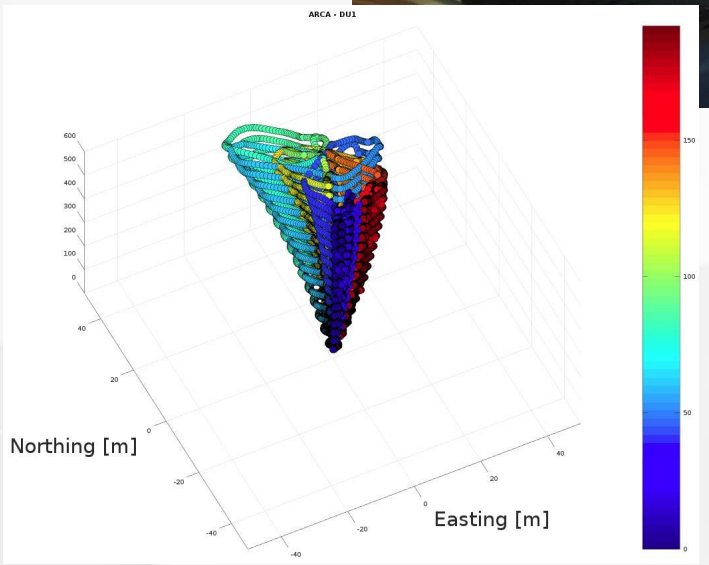
Dark Room to Sea:

- inter-DU

HV-tuning:

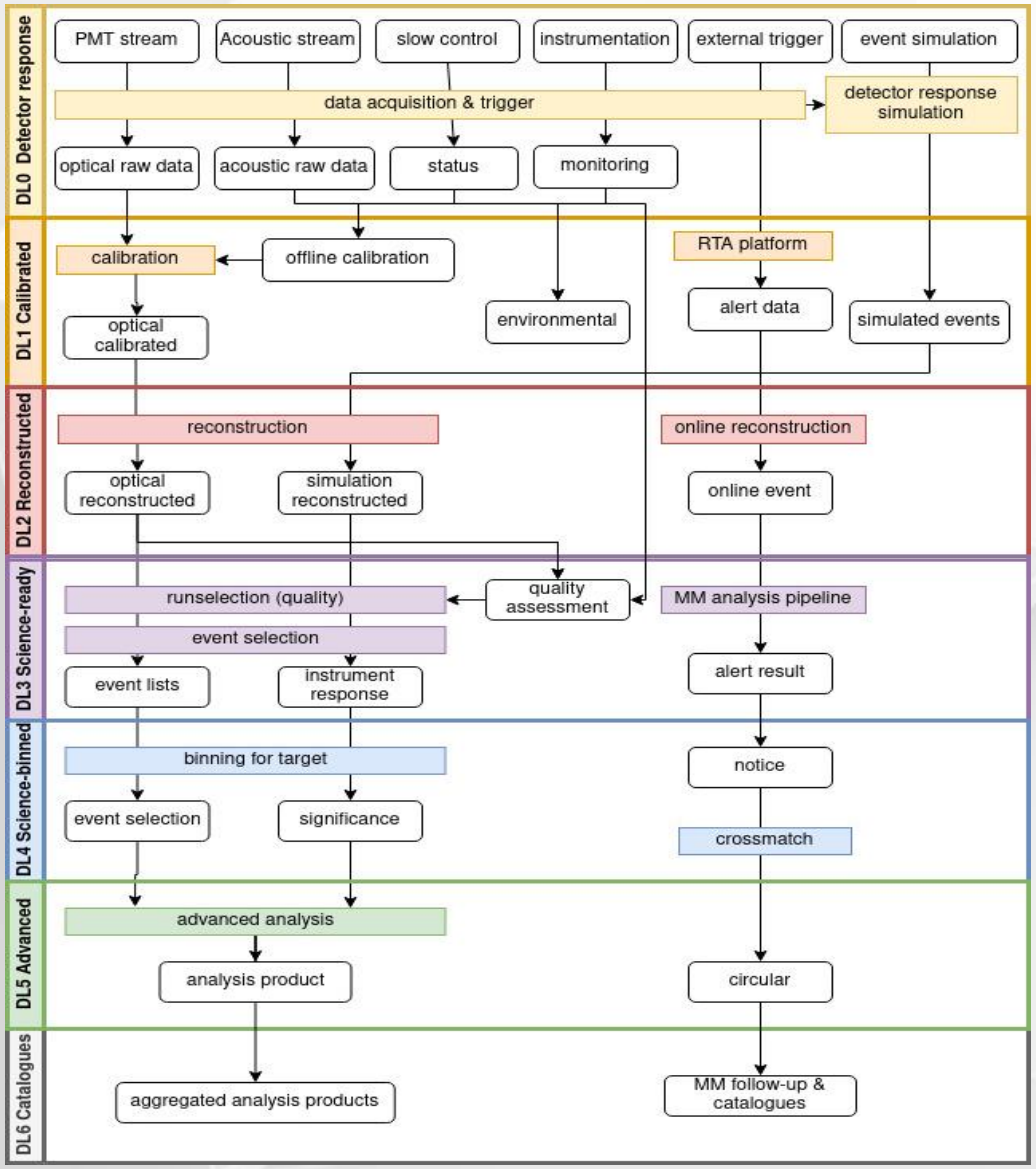
- PMT gain equalization





KM3NeT data management

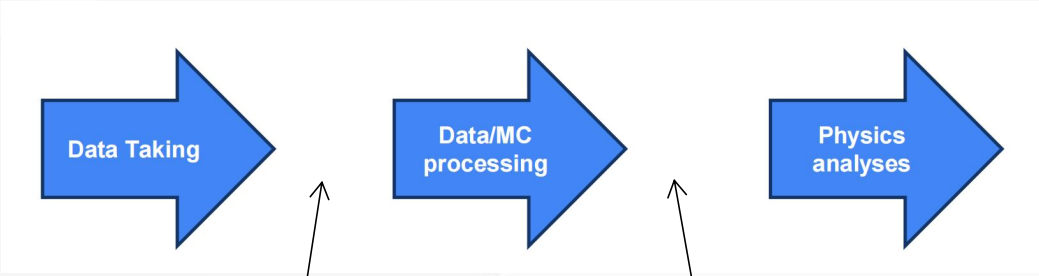
Data management distributed in levels.



Data Processing and Data Quality WG

The **Data Processing & Data Quality** working group (main) responsibilities are:

- To ensure the data taking quality by performing studies on quality assessments of the KM3NeT.
- To regularly process the data and to produce MC simulations ensuring that the input for data analyses are the best obtainable at a given moment.



data quality (level 1)
Investigate data "health" for issues, potentially not reproducible, in simulations

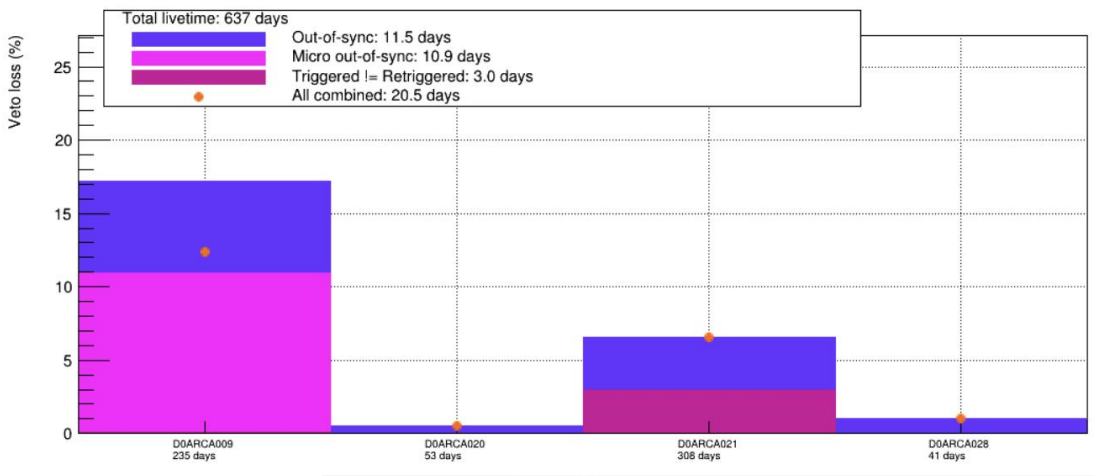
data quality (level 2)
Provide data/simulations runlists to be used for physics analyses

Data quality

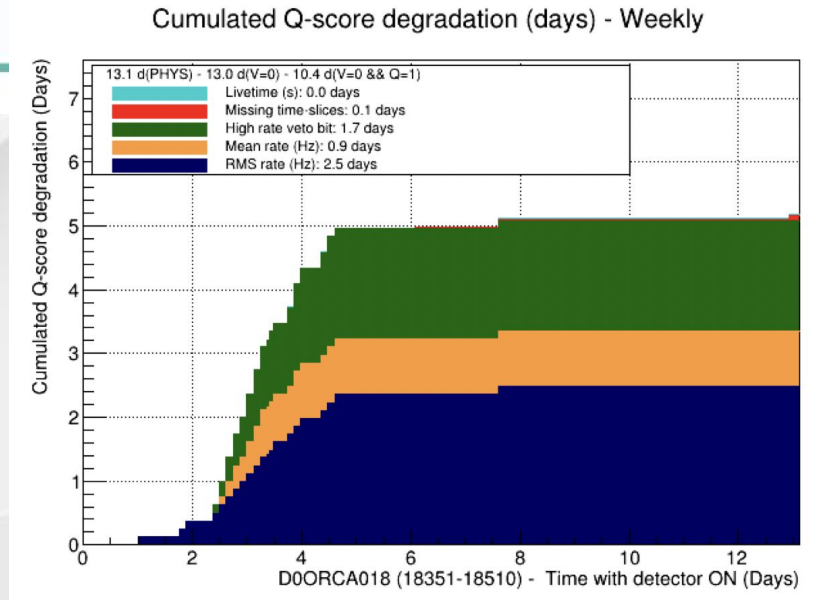
- Investigate data “health” - probable issues that may not be reproducible in simulations
- Cross check with database info and shift reports for potential problems or interruptions in the data taking

Access quality score (Q) and veto score (V) based on important observables during data taking

```
JDataQuality \
    -D "$DETECTOR_ID"                                                                  \
    -R "${RANGE%%-*} ${RANGE##*-}"                                                     \
    -f  $QUALITY_TXT                                                                   \
    -Q "livetime_s;                                                    1200     60000" \
    -Q "(UTCMax_s - UTCMin_s) - livetime_s;                           -100     +100"  \
    -Q "JDAQEvent / livetime_s;                                        0.05     50.0"  \
    -Q "HRV;                                                           0.00     0.2"   \
    -Q "DAQ;                                                           0.95     1.05"  \
    -Q "WR;                                                            0.95     1.05"  \
    -Q "FIFO;                                                          0.00     0.01"  \
    -Q "MEAN_Rate_Hz;                                                  4500     15000" \
    -Q "RMS_Rate_Hz;                                                   200      5000"  \
    -Q "Acoustics / livetime_s;                                        0.0      100.0" \
    -Q "AHRS      / livetime_s;                                        0.0      100.0" \
    -V "(JDAQEvent - JTriggerReprocessor) * 1.0 / (JDAQEvent + 1.0e-10);  -5.0e-3  +5.0e-3" \
    -V "out_sync;                                                      0        0"     \
    -V "out_usync;                                                     0              \
    -o  $QUALITY_ROOT
    -d  $DEBUG --!
```

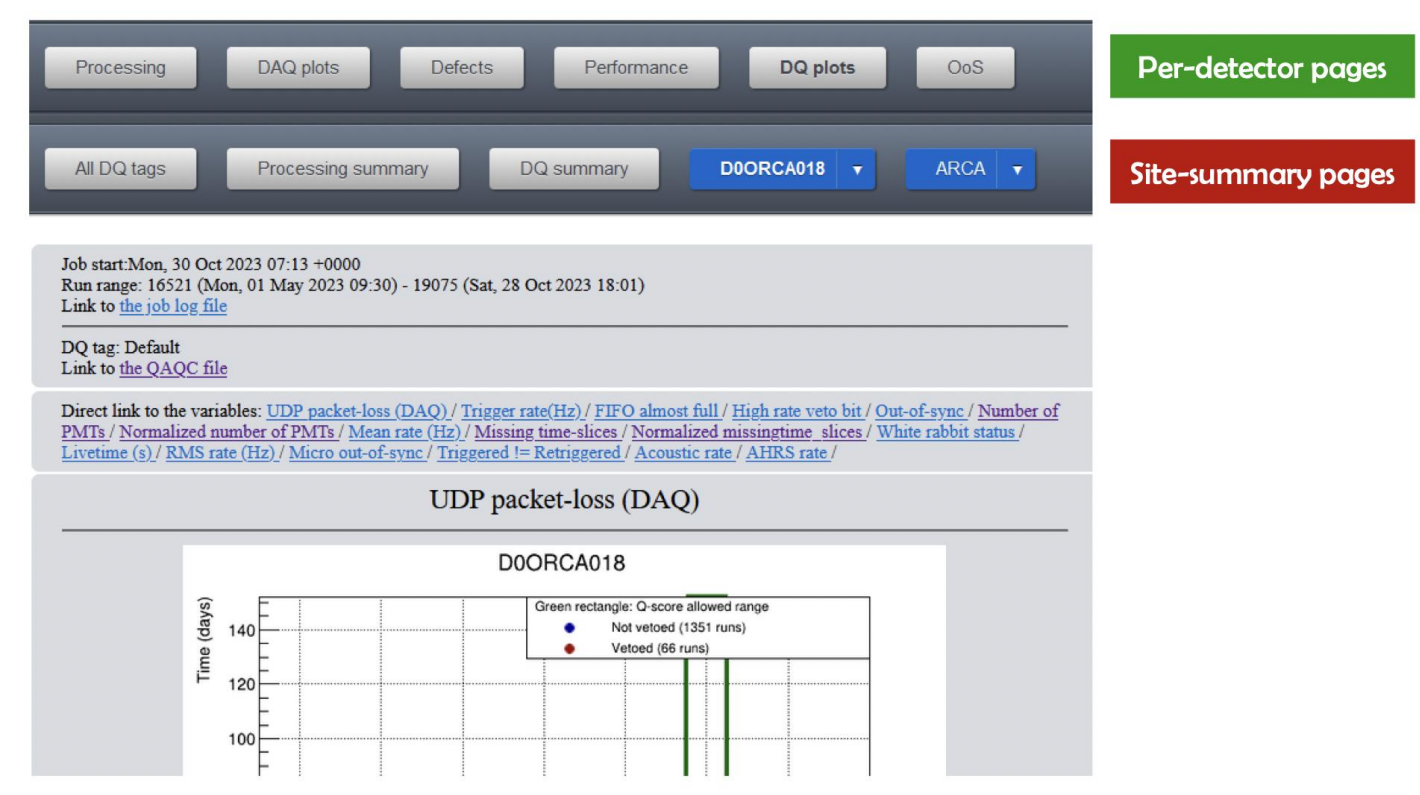


ARCA with active DUs: Veto-score monitoring



ORCA with active 18 DUs: issues that affect the Q-score

Data quality - WG developments & activities



Per-detector pages

Site-summary pages

1. **Web page** to monitor data quality for each site.

2. Implementation of ***defects*** page:
for each author, runs, DP-DQ-DAQ, issue "severity", possible recovery.

3. **Out-of-Sync issue investigation online.**

4. **Cross-checking** of defects with DQ observables and data quality health.

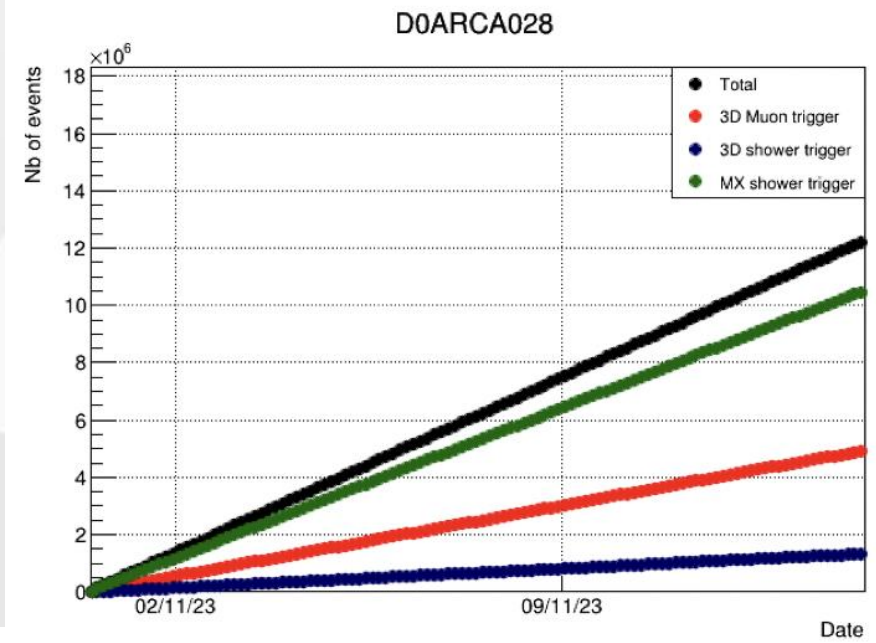
5. **Quality score degradation investigation.**

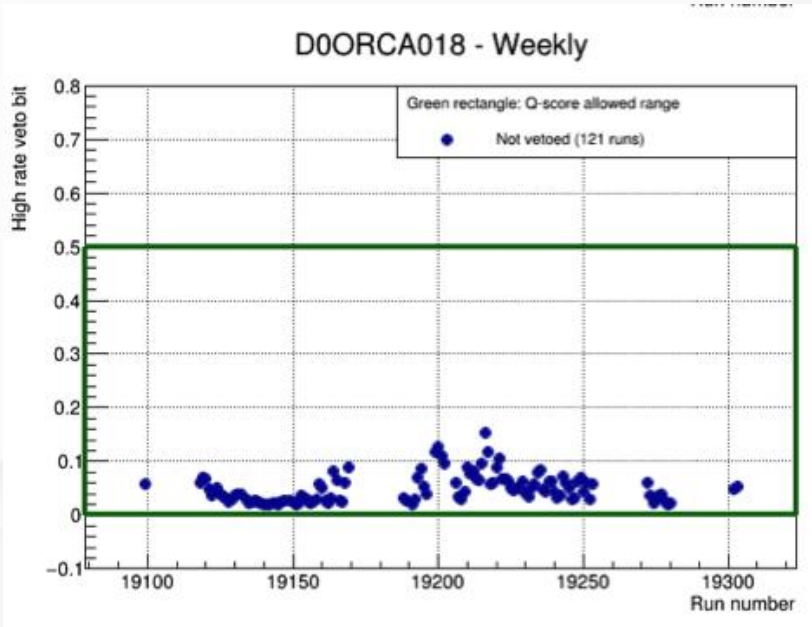
6. Optimization of the **veto** score conditions.

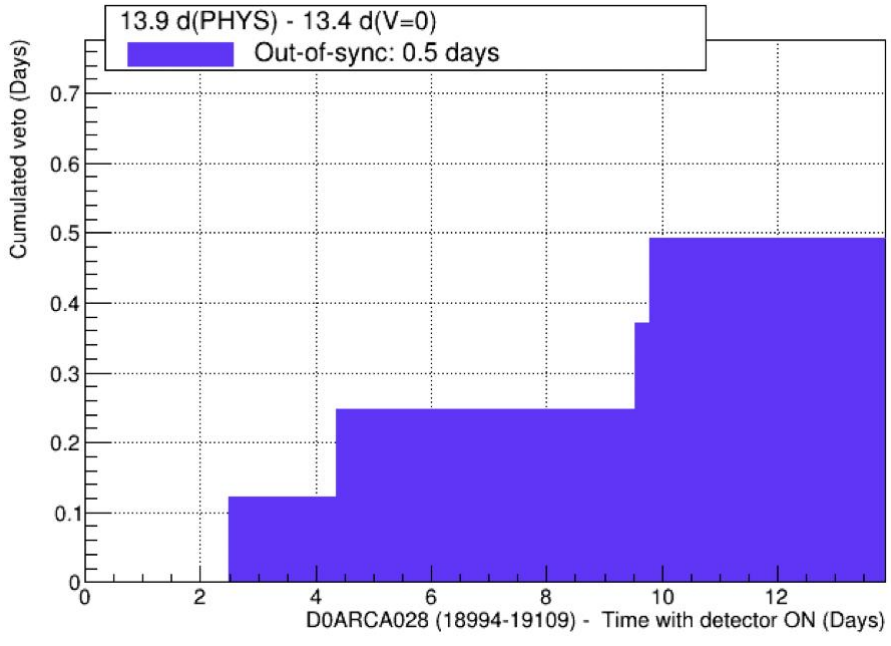
Biweekly report on tha data-taking quality status in DPDQ group meetings.

Discussion on on-going issues.

Data quality - Data taking monitoring (a few examples)

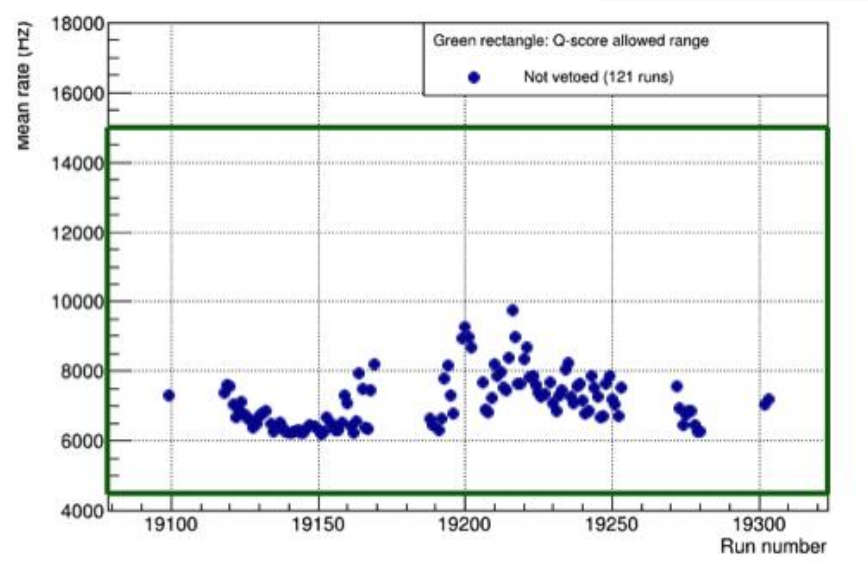




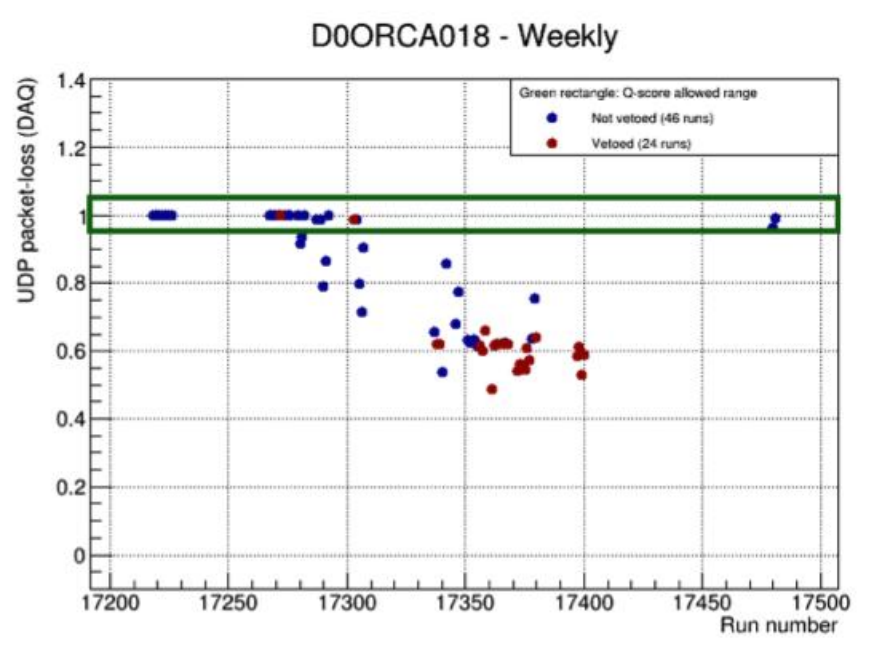


ARCA detector:
stable data taking
a few Out-of-Sync cases

ORCA detector:
hiccups in the data taking
(understood, not warning)
moderate bioluminescence
activity

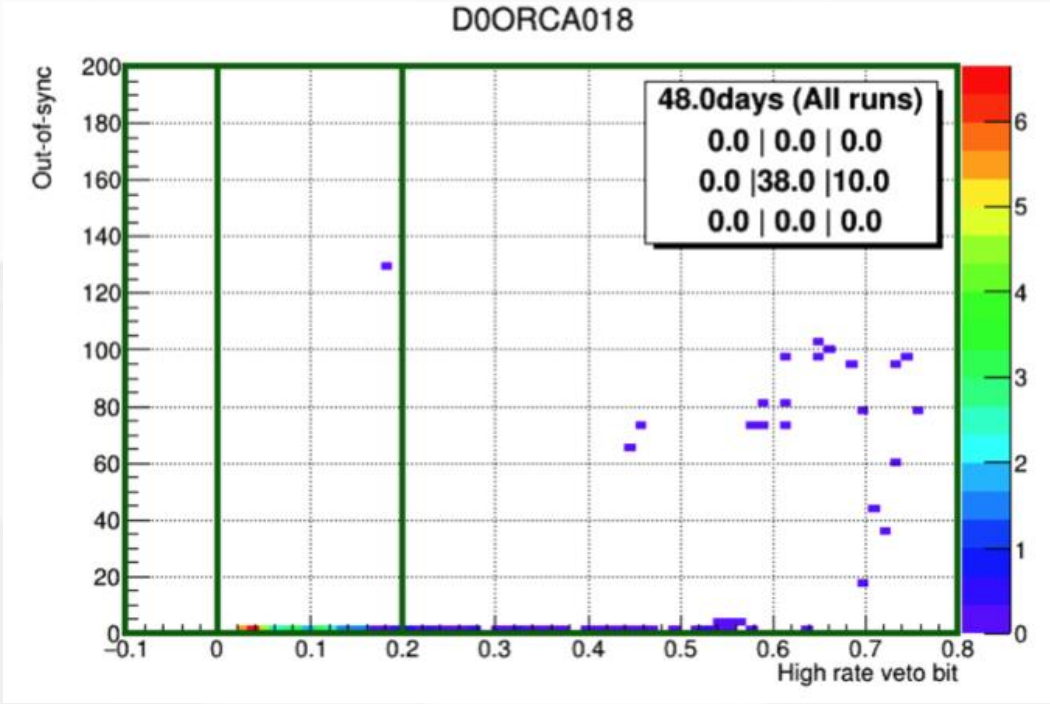


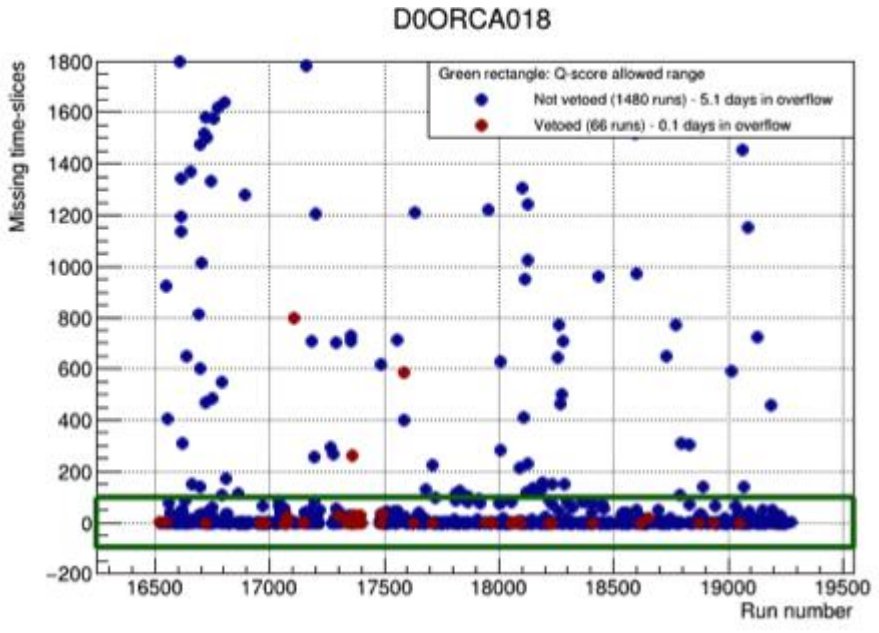
Data quality - Developments/Investigations (a few exapmples)



DAQ plots added in the web page

Correlate issues with Out-of-Sync cases





Missing timeslices investigation

d = | (Run time stop – run time start) – run livetime |

Data processing

- Organize & schedule mass and test processings.
- Implement mass processing tool/workflow.
- Performance investigations (collab. with software WG and with simulations WG).
- Monitor the computing systems usage.
- Maintain documentation.

Data-taking periods to be processed [edit]

ARCA data-taking periods

Detector (DB naming)	Detctor ID (DB number)	Run range	Dates
D0ARCA020	116	12288 - 13197	July - September 2022
D0ARCA021	133	13198 - 17664	September 2022 - 11 September 2023

ORCA data-taking periods

Detector (DB naming)	Detctor ID (DB number)	Run range	Dates
D_ORCA006	49	7221 - 11296	February 2020 - November 2021
D0ORCA010	100	11323 - 13212	December 2021 - May 2022
D0ORCA007	110	13213 - 13703	May - June 2022
D1ORCA013	117	13704 - 13961	July 2022
D0ORCA011	123	13964 - 14318	August - September 2022
D1ORCA011	132	14356 - 15146	October - December 2022
D0ORCA015	138	15158 - 16339	December 2022 - April 2023
D1ORCA015	146	16341 - 16516	5 April 2022 - 29 April 2023
D0ORCA018	148	16517 - 18298	May 2023 - 31st August 2023

Computing Centers & Taskforces [edit]

Computing center	Detctor	Taskforce
CC Lyon	D_ORCA006, D0ORCA018	Santiago, Walid
Nikhef	D0ORCA011, D1ORCA011, D0ORCA015, D1ORCA015	Jhilik, Vincent, Mieke
CNAF	D0ARCA020, D0ARCA021	Vittorio, Anna, Mieke
ECAP	D0ORCA010 (KM3NeT_00000100), D0ORCA007 (KM3NeT_00000110), D1ORCA013 (KM3NeT_00000117)	Kay

ORCA

Week	19/6	26/6	3/7	10/7	17/7	24/7	August	4/9	11/9	18/9	25/9	2/10	9/10	Collab.M	23/10	30/10	6/11	13/11	20/11	27/11	4/12	11/12	18/12	25/12	1/1	8/1	15/1	22/1
																D_0ORCA	D_0ORC	D_0ORC	D_0ORCA006									
Preparatory Phase																		D0ORCA	D0ORCA010									
WG requests																			D1ORCA011									
Software tests																												
Version validation																												
Version freezing																												
Input Preparation																												
Calibration preparation																												
Data Quality preparation																												
Mass production																												
Pre-mass production																												
Validation																												
Mass production																												
DPDQ WG Report																												

Processing status Runlists to be processed: here — Last update 21/11/23

D_ORCA006 (KM3NeT_00000049)	Priority runlist - 1805 runs			Secondary runlist - 1279 runs		
***CCA IN2P3**	light	reco	dst	light	reco	dst
Data	-	1786	1786			
Pure_noise	-	1738	1738			
MUPAGE	1805	1805	1805			
Neutrinos	1805	1805	1805			

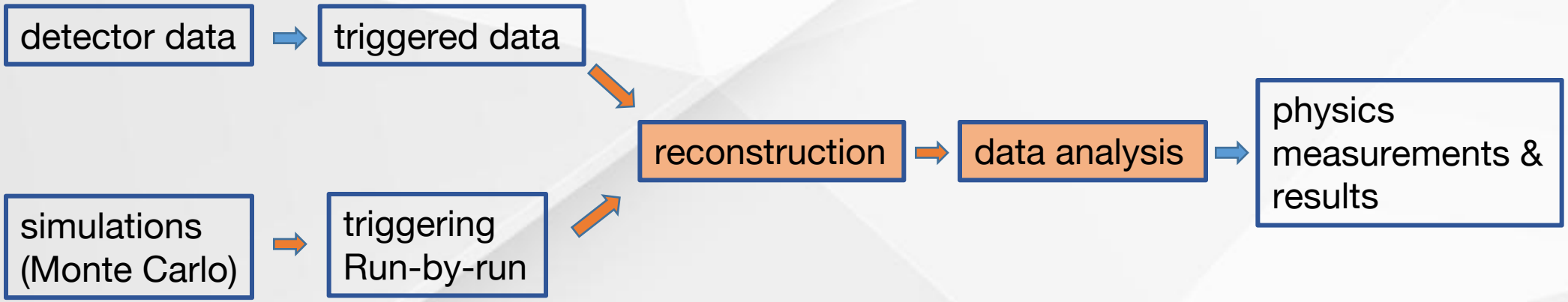
Git issue with details: here — Data/MC comparisons: here

Elog with details: here

Data processing - General idea

In data processing, **recorded data are reconstructed**, **simulations are produced**, **triggered** according the detector condition by the time of the data taking **(Run-by-run)** and **reconstructed.**
Part of the group task is to verify the “correctness” of the processing and the input by performing a primary **data analysis** (data/MC comparisons etc.).

**Run-by-run approach: using the detector conditions, from a given run for: hit rates (including HRV), detector geometry, PMT efficiency from K40 analysis etc, -> reproducing the run conditions as good as possible.*



Simulation chain:
Event generation (atm. muons, neutrinos) -> Light simulation ->
Detector response simulation (data conditions) -> Event reconstruction (track, shower)

Data processing - New Workflow - Currently used for the 1st time

Using containerized software:

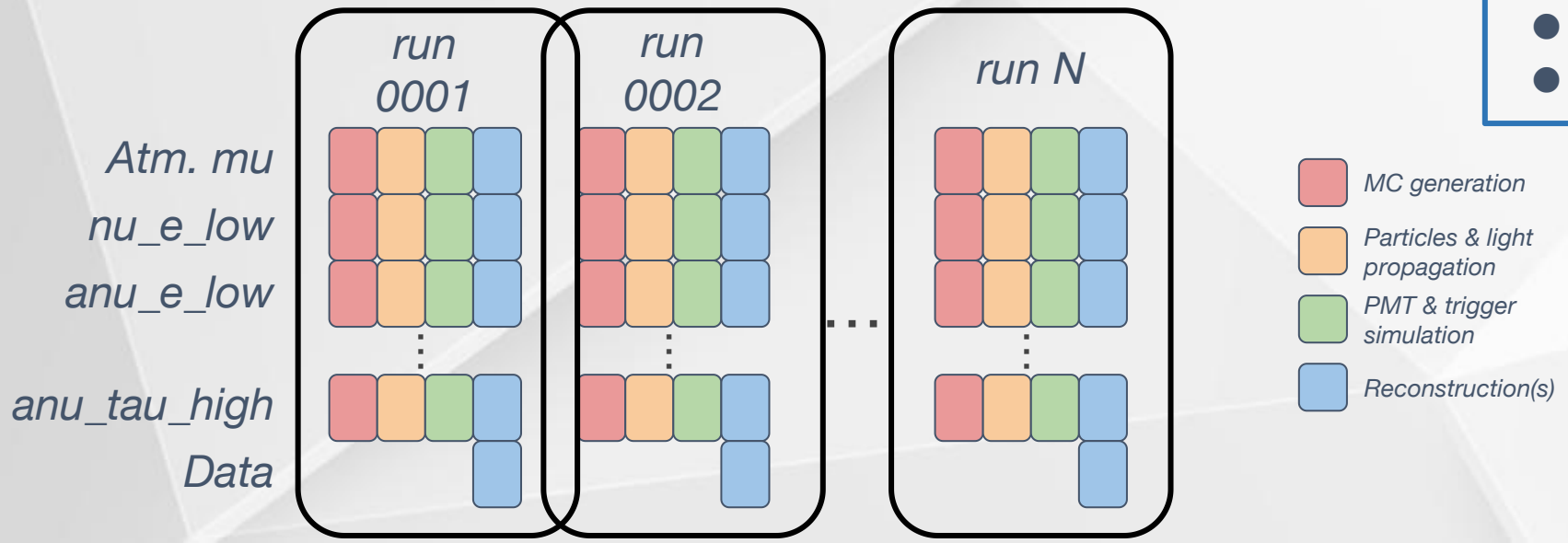
- Ensuring software versions
- No doubt about other dependencies (e.g. which ROOT version)
- Can also include processing tools used in the path

Using a workflow management system:

- Optimize execution, monitoring performances
- Decouple job submission from workflow implementation
- Take care of logging
- Mutualizes resources, allow for early file merging
- Optimizes disk-space used during processing

Workflow steps:

- **Input collection:**
  - Raw data (iRods), online detx (DB), offline detx (git)
- **Generation:**
  - gSeaGen, MUPAGE
- **Light propagation:**
  - Km3sim, JSirene
- **Trigger:**
  - JTERBR, pure_noise MC
- **Reconstruction**
- **DST files production**



Data processing - How to

Simple use: Making a .yaml file containing software needs, input, versions and desired output!

List of files to produce. Contains all information about the processing steps:

```
to_produce:
  - "{detid}/{run}/reco/{version}.data.jorcarec.jshower.aanet.{detid}.{run}.root"
  - "{detid}/{run}/reco/{version}.mc.mupage_std.sirene.jterbr.jorcarec.jshower.aanet.{detid}.{run}.root"
  - "{detid}/{run}/reco/{version}.mc.gsg_neutrinos.merged.jterbr.jorcarec.jshower.aanet.merged.{detid}.{run}.root"
  - "{detid}/{run}/reco/{version}.mc.pure_noise.jterbr.jorcarec.jshower.aanet.{detid}.{run}.root"
```

Merging rules to define how to produce combined files

```
light:
  "{detid}/{run}/light/{version}.mc.gsg_nue_CC.merged.{detid}.{run}.root":
    - "{detid}/{run}/light/{version}.mc.gsg_nue_low_CC.km3sim.{detid}.{run}.root"
    - "{detid}/{run}/light/{version}.mc.gsg_anue_low_CC.km3sim.{detid}.{run}.root"
    - "{detid}/{run}/light/{version}.mc.gsg_nue_high_CC.sirene.{detid}.{run}.root"
    - "{detid}/{run}/light/{version}.mc.gsg_anue_high_CC.sirene.{detid}.{run}.root"
    - "{detid}/{run}/light/{version}.mc.gsg_nue_higher_CC.sirene.{detid}.{run}.root"
    - "{detid}/{run}/light/{version}.mc.gsg_anue_higher_CC.sirene.{detid}.{run}.root"
```

Monte-Carlo generator setup

```
gseagen:
  types:
    nue_low_CC:    {flavour:  12, nev: 6e4, emin_gev: 1, emax_gev: 100, can_margin_m:  80, gspec: 3.0, part_int: CC, tune: "G18_02a_00_000"}
    anue_low_CC:   {flavour: -12, nev: 6e4, emin_gev: 1, emax_gev: 100, can_margin_m:  80, gspec: 3.0, part_int: CC, tune: "G18_02a_00_000"}
```

Software container version and generic options.

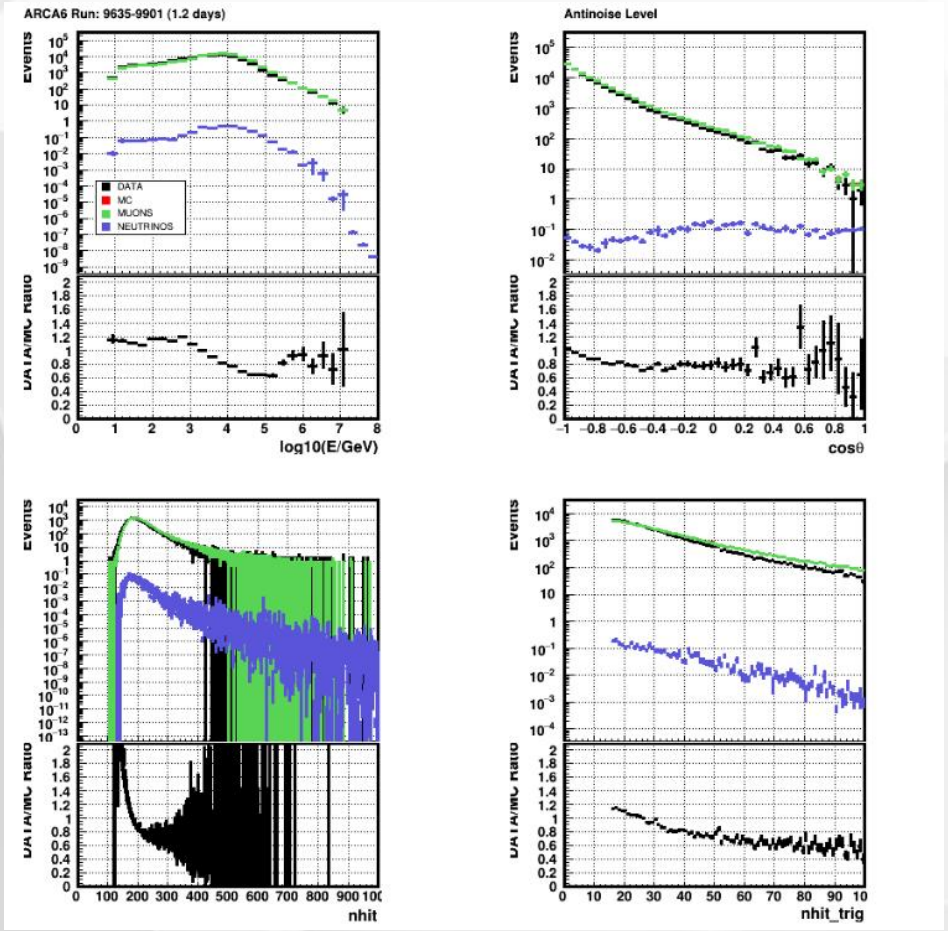
```
00000049/00007794
├── calibration
├── detector
│   └── logs
├── generator
│   ├── benchmarks
│   ├── logs
│   └── parameters
├── light
│   ├── benchmarks
│   ├── logs
│   └── parameters
├── reco
│   ├── benchmarks
│   └── logs
└── trigger
    ├── benchmarks
    └── logs
```

Main advantage: Compatible with multiple Computing Centers -> Optimizes the mass processing procedures.

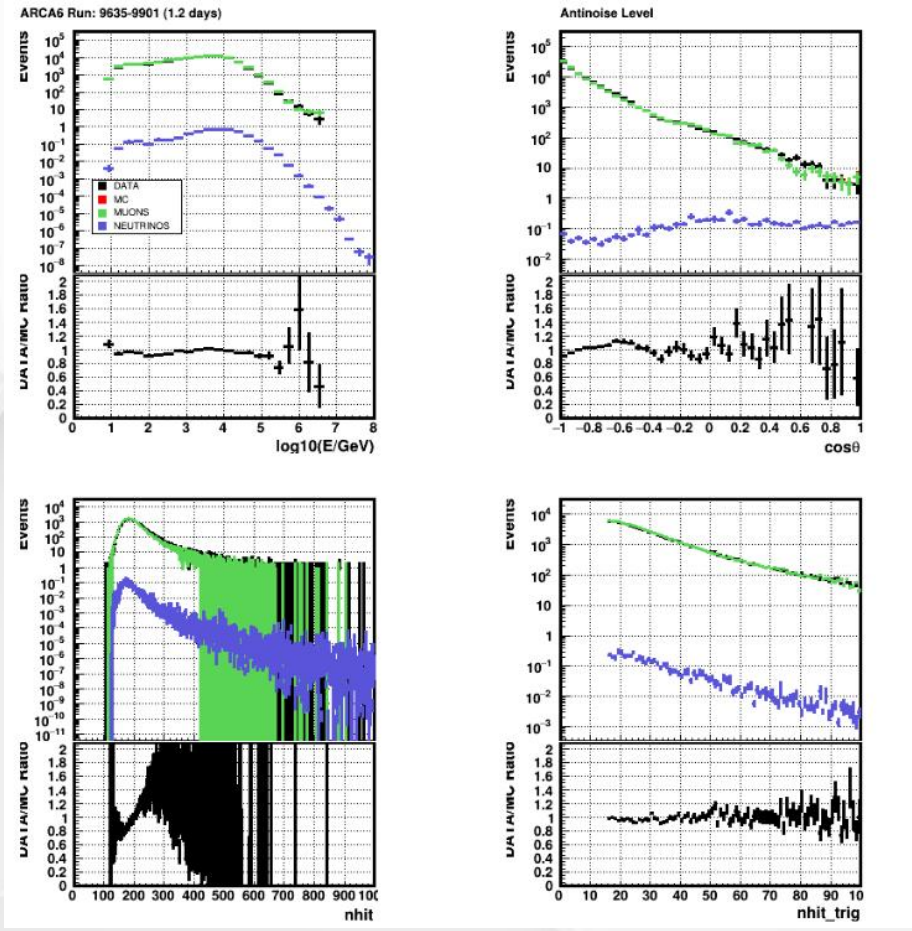
Processing checks & investigations - optimization of simulations input

- Comparing distributions of “basic” variables for data and MC.
- Investigated the MUPAGE (atm. muon simulation) parametrization tunings.

MUPAGE tuning 1



MUPAGE tuning 2



Important variables compared in the first checks:
LogEnergy
CosTheta
N. of reconstructed hits
N. of triggered hits etc.

Mass processing checking tool by C. Distefano

Processing checks & investigations - optimization of simulations input

- Comparing distributions of “basic” variables for data and MC.
- Investigated the MUPAGE (atm. muon simulation) parametrization tunings.

MUPAGE tuning 1

Livetime

Units	Data	Muons	Neutrinos
sec	104692	210930	15396.6
days	1.21172	2.44132	0.178201

Number of events

Level	Data	Muons	Neutrinos
PreCut	151791	294049	0.69816
Antinoise	97269	213512	0.518314
Upgoing	914	2366	0.307737
Select	13	57	0.0717727

Event rates

Level	Data	Muons	Neutrinos	DATA/MC
PreCut	1.44988	1.39406	4.53451e-05	1.04004
Antinoise	0.929093	1.01224	3.36642e-05	0.917853
Upgoing	0.00873034	0.011217	1.99873e-05	0.777978
Select	0.000124173	0.000270232	4.6616e-06	0.451413

Events in DATA livetime

Level	Data	Muons	Neutrinos
PreCut	151791	145947	4.6904
Antinoise	97269	105974	3.47925
Upgoing	914	1174.34	2.06152
Select	13	28.2886	0.483032

Run list 6
Run processed 6

MUPAGE tuning 2

Livetime

Units	Data	Muons	Neutrinos
sec	104692	115830	104692
days	1.21172	1.34062	1.21172

Number of events

Level	Data	Muons	Neutrinos
PreCut	154735	163144	7.021
Antinoise	100043	112953	5.08005
Upgoing	864	853	3.23718
Select	5	12	0.453275

Event rates

Level	Data	Muons	Neutrinos	DATA/MC
PreCut	1.478	1.40848	6.70631e-05	1.04935
Antinoise	0.95559	0.975162	4.85236e-05	0.979925
Upgoing	0.00825275	0.00736424	3.09209e-05	1.11992
Select	4.7759e-05	0.0001036	4.32959e-06	0.440653

Events in DATA livetime

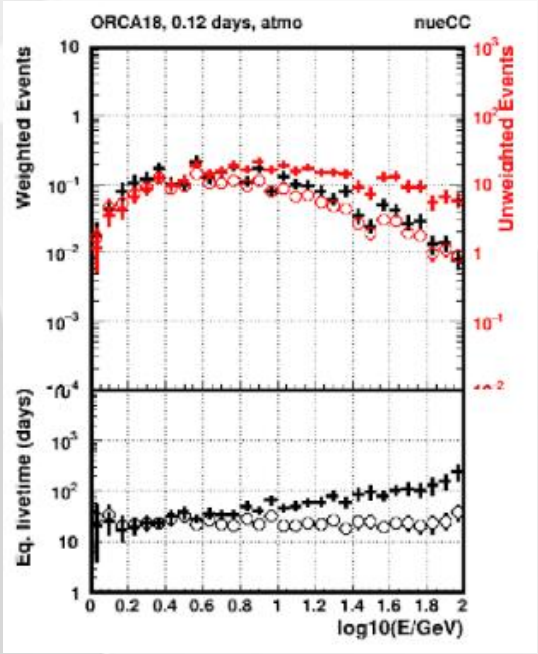
Level	Data	Muons	Neutrinos
PreCut	154735	148260	7.021
Antinoise	100043	102688	5.08005
Upgoing	864	771.342	3.23718
Select	5	10.966	0.453275

Run list 6
Run processed 6

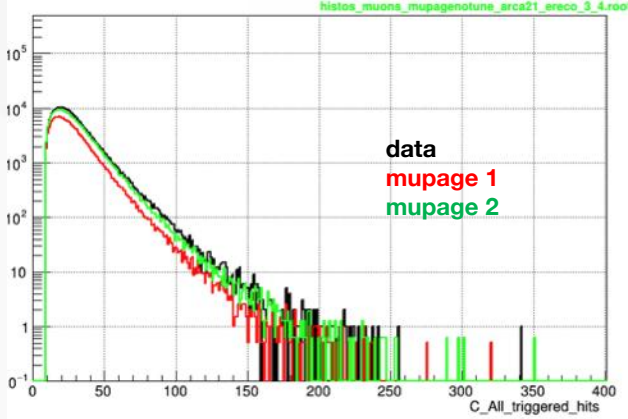
Important variables compared in the first checks:
Data/MC comparisons in different analysis levels.
Numbers extrapolated to data livetime.
Trigger level is also reported.

Processing checks & investigations - data/MC comparisons

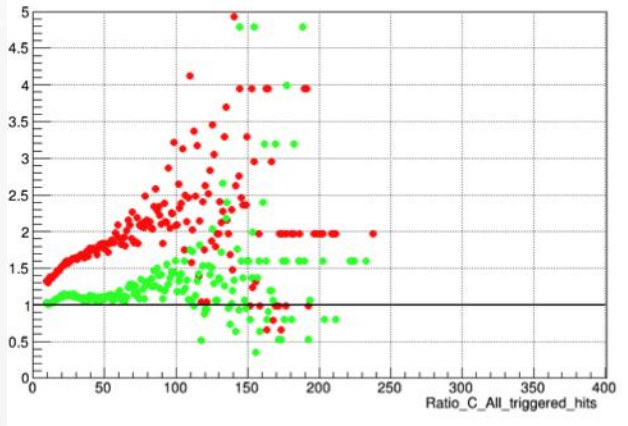
- Detailed investigations of the processings to be used for physics analyses
- Investigations in different levels (generation, light simulation, reconstruction)
- Plans to include benchmark checks per processing step



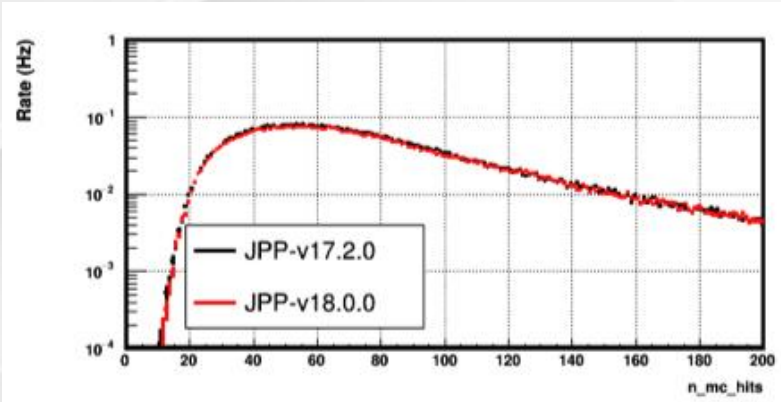
Generation level:
Unweighted/weighted events investigation

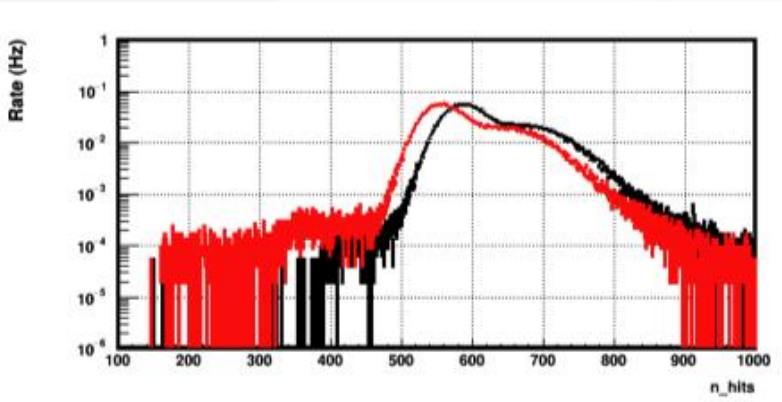


Reconstruction level:
Comparison of data with 2 different atmospheric muon productions (tunings)



Trigger level:
Comparisons of productions with different software versions

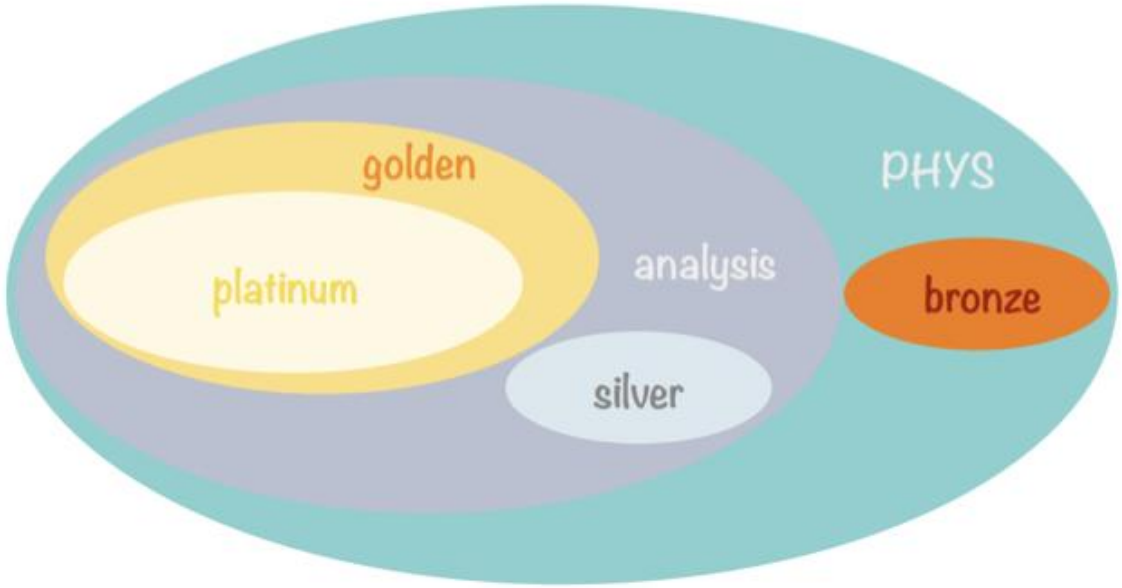




Data quality - “safe” data/MC to be used for physics analyses

- Optimize quality criteria to classify runs that can be used in the analyses – physics runlists.
- Data/data comparisons (investigate differences between data-taking periods)

Physics runlists classification scheme
golden, silver etc. lists

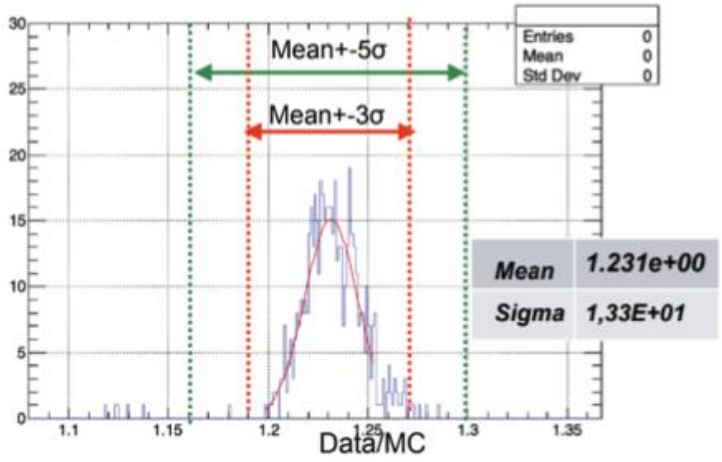


Classification according to data/MC degradation
What is acceptable ?

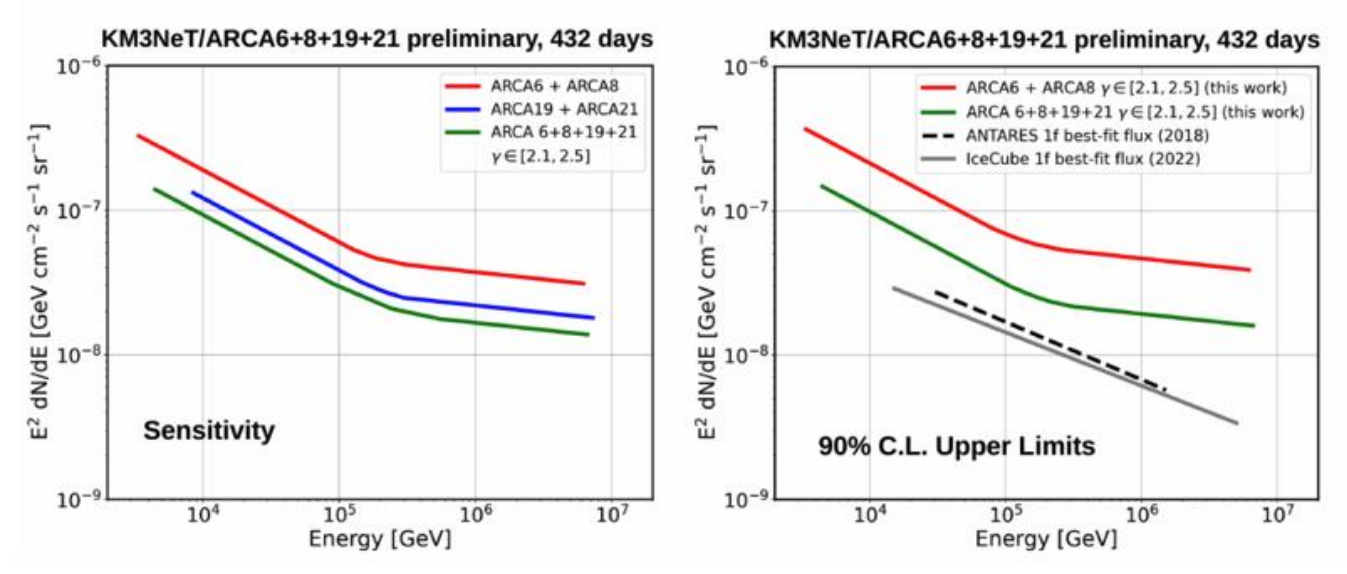
- Golden Run 👉 V=0 && Q=1 && data/MC = Mean +- 3σ

By R.Coniglione

- Silver run 👉
  - OOS=1
  - Missing Time Slice <500
  - Live Time < 1200s
  - 3σ < Data/MC < 5σ
- Bronze runs 👉 all the remaining

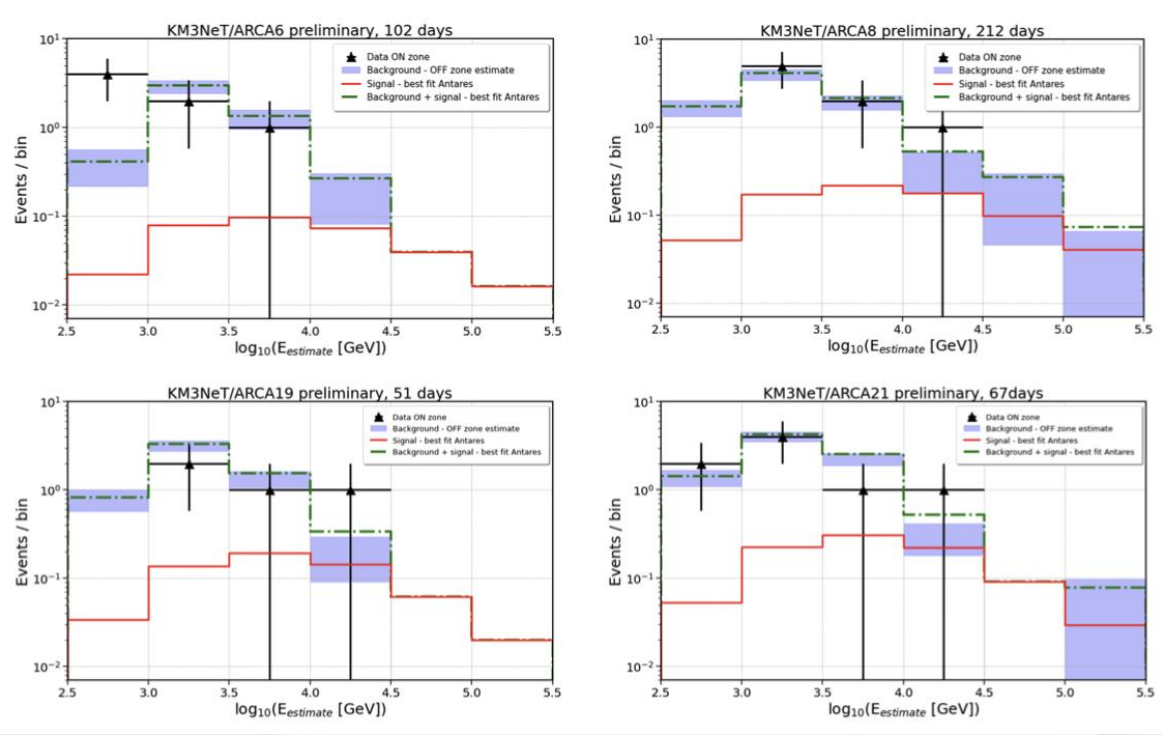


Physics data analyses - Astrophysical analyses performed with KM3NeT data

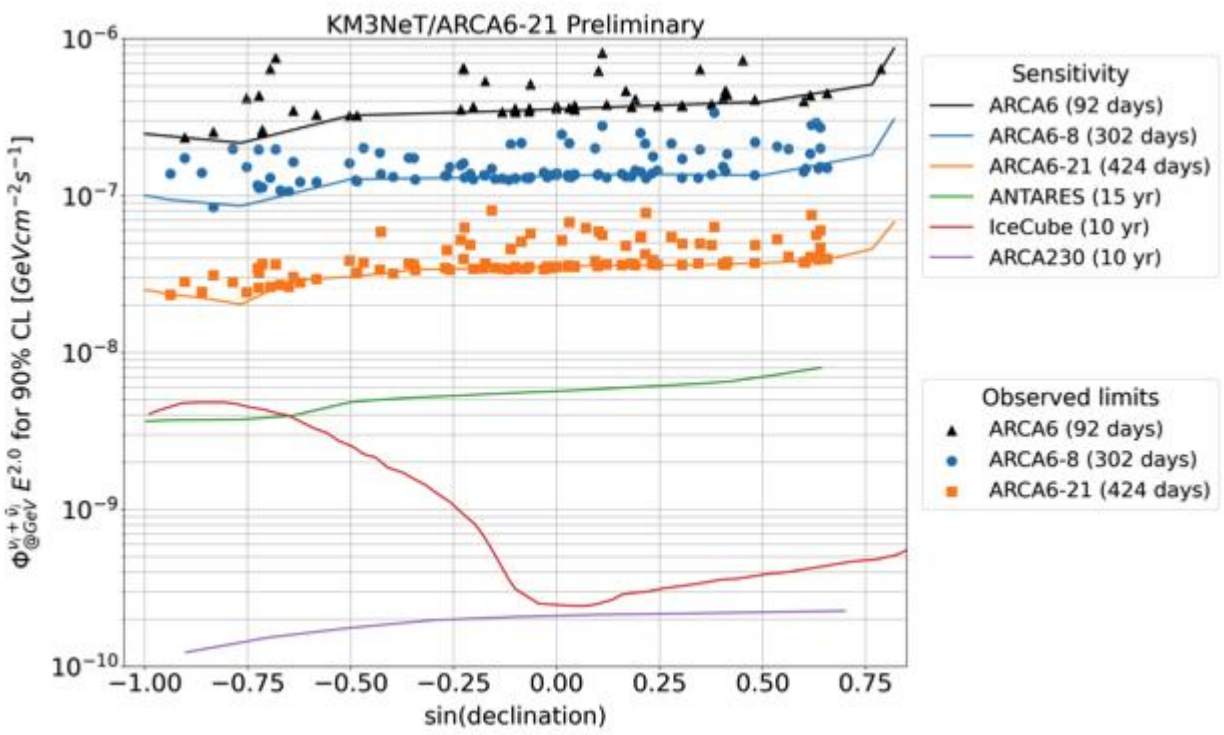


Search for a diffuse astrophysical neutrino flux with KM3NeT/ARCA
https://pos.sissa.it/444/1195/pdf

Search for a diffuse astrophysical neutrino flux from the Galactic Ridge using KM3NeT/ARCA data
https://pos.sissa.it/444/1190/pdf



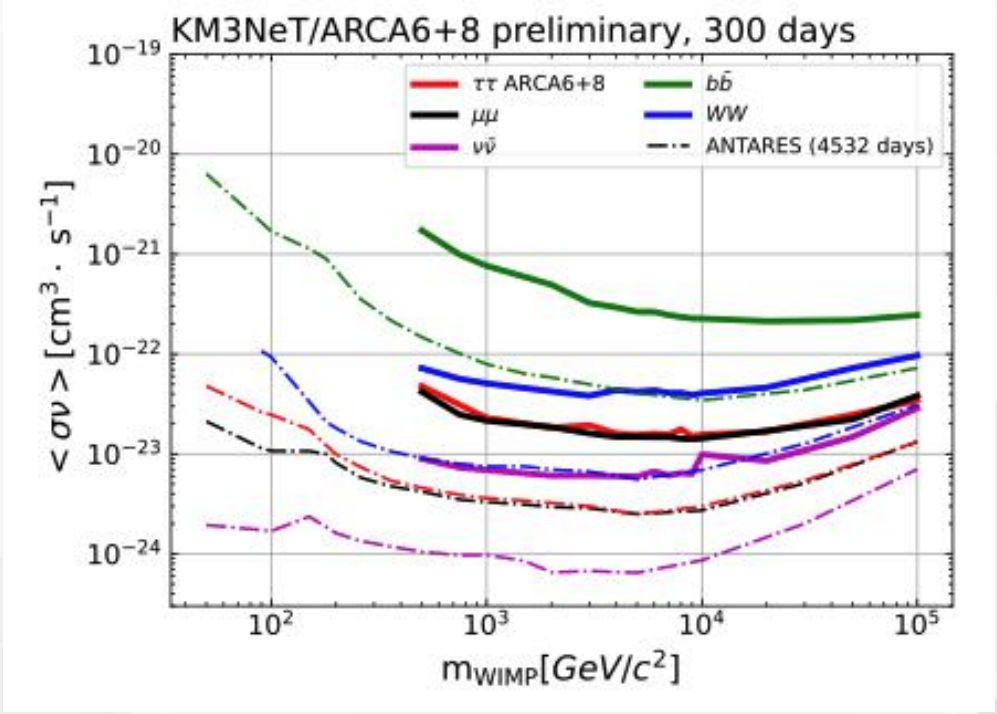
Physics data analyses - Astrophysical analyses performed with KM3NeT data



Search for cosmic neutrino point sources and extended sources with 6-21 lines of KM3NeT/ARCA
https://pos.sissa.it/444/1018/pdf

KM3NeT upper limits quickly reaching the ANTARES 15yr limits

Indirect Search for Dark Matter with the KM3NeT Neutrino Telescope
https://pos.sissa.it/444/1377/pdf



and many more to come soon...

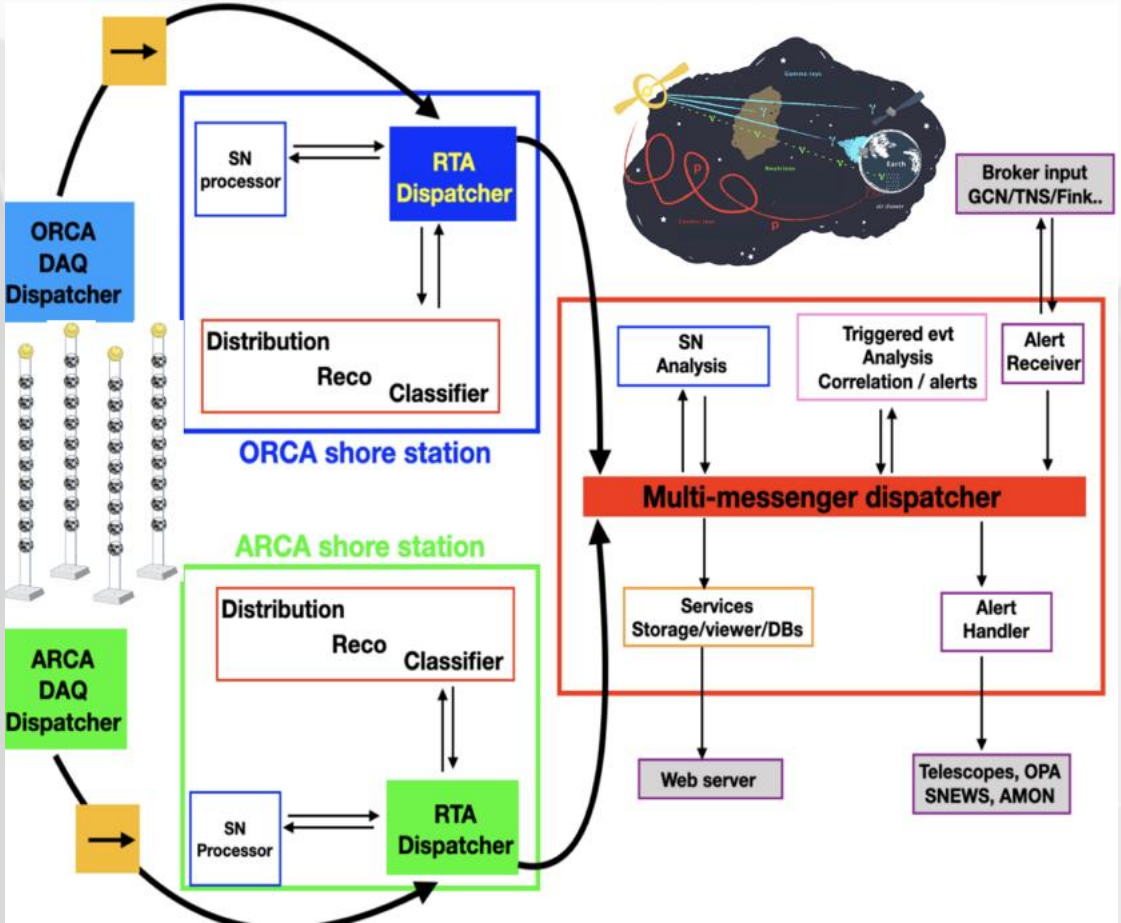
Physics data analyses - Multimessenger program

With the Real-Time Analysis platform, we perform:
- Auto-correlation searches
- Follow-up studies

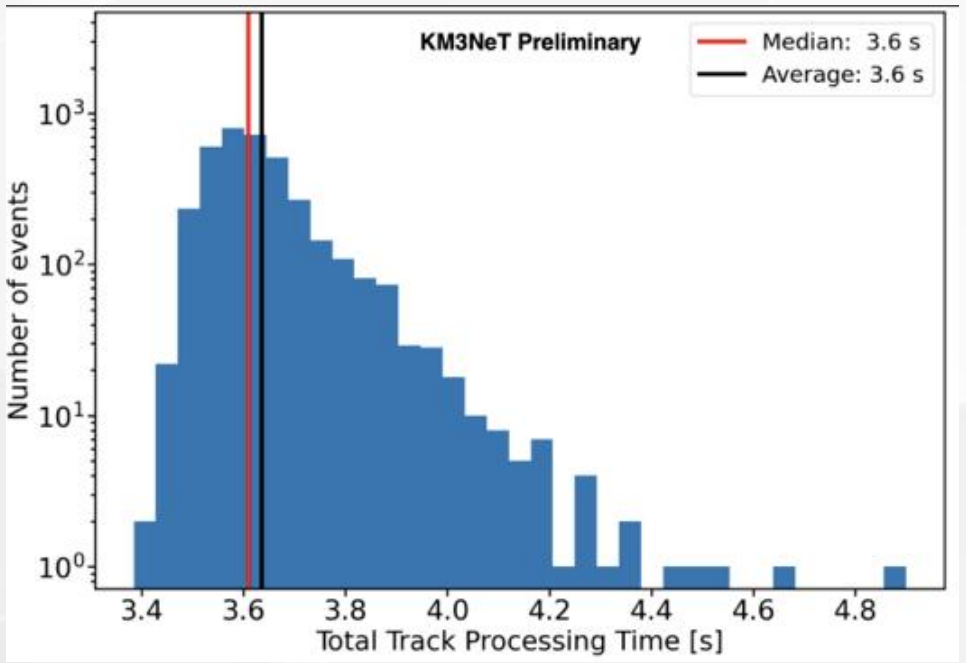
starting automatically whenever an interesting external alert is received

By 2024, high-energy neutrino alerts will be sent in real-time!!!

The RTA platform of KM3NeT



In ARCA21 **a median delay of 4 s is obtained from data filtering to classification,** including event buffering, dispatching and reconstruction times.



https://pos.sissa.it/444/1125/pdf

Thank you for your attention!

A.Sinopoulou
anna.sinopoulou@ct.infn.it

Stay tuned:
km3net.org Instagram Twitter Facebook