Data Quality for ICARUS Neutrino Detector

Alexandra Lehan
Dr. Minerba Betancourt
Dr. Harry Hausner

Objectives

Discuss metrics of Data Quality
To understand the importance of data quality measuring

Observe Profiles from Beams
To analyze aggregate data

Interpret CI Builds
To analyze individual runs

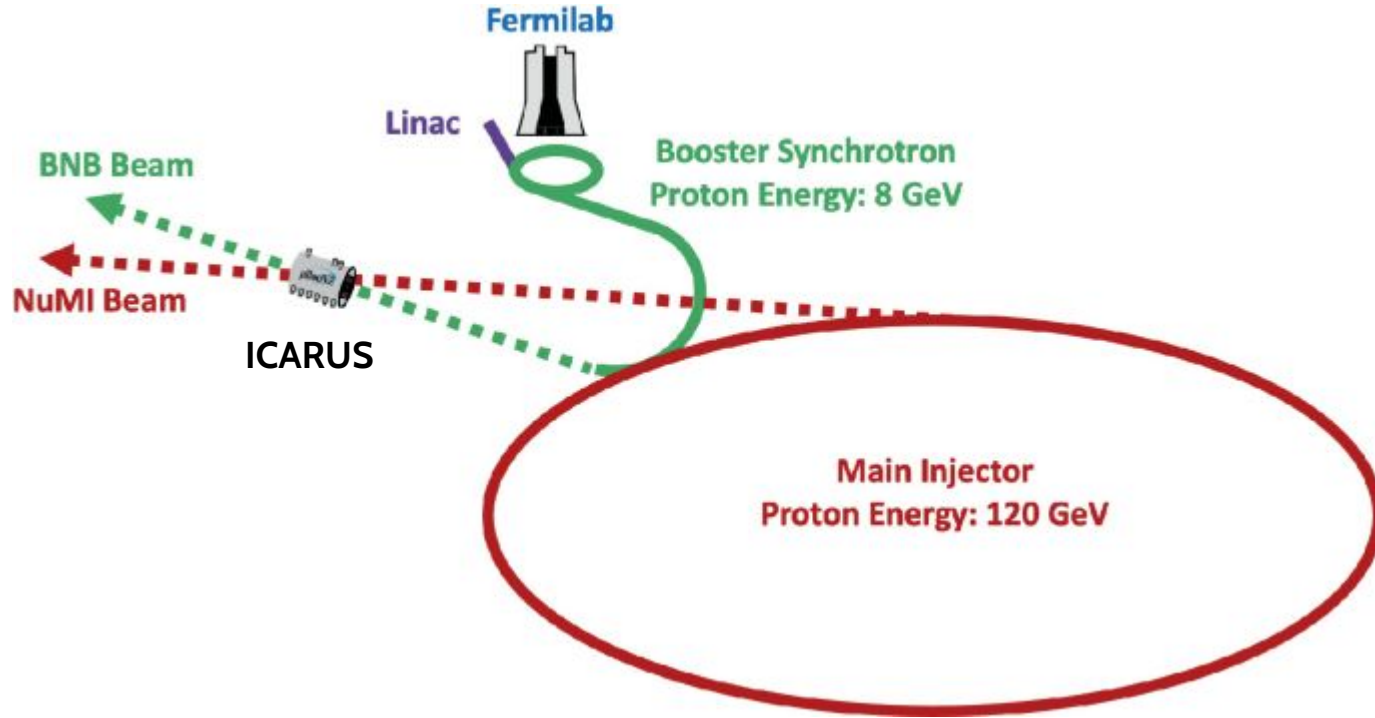**Classify Data Quality and Analyze Results**
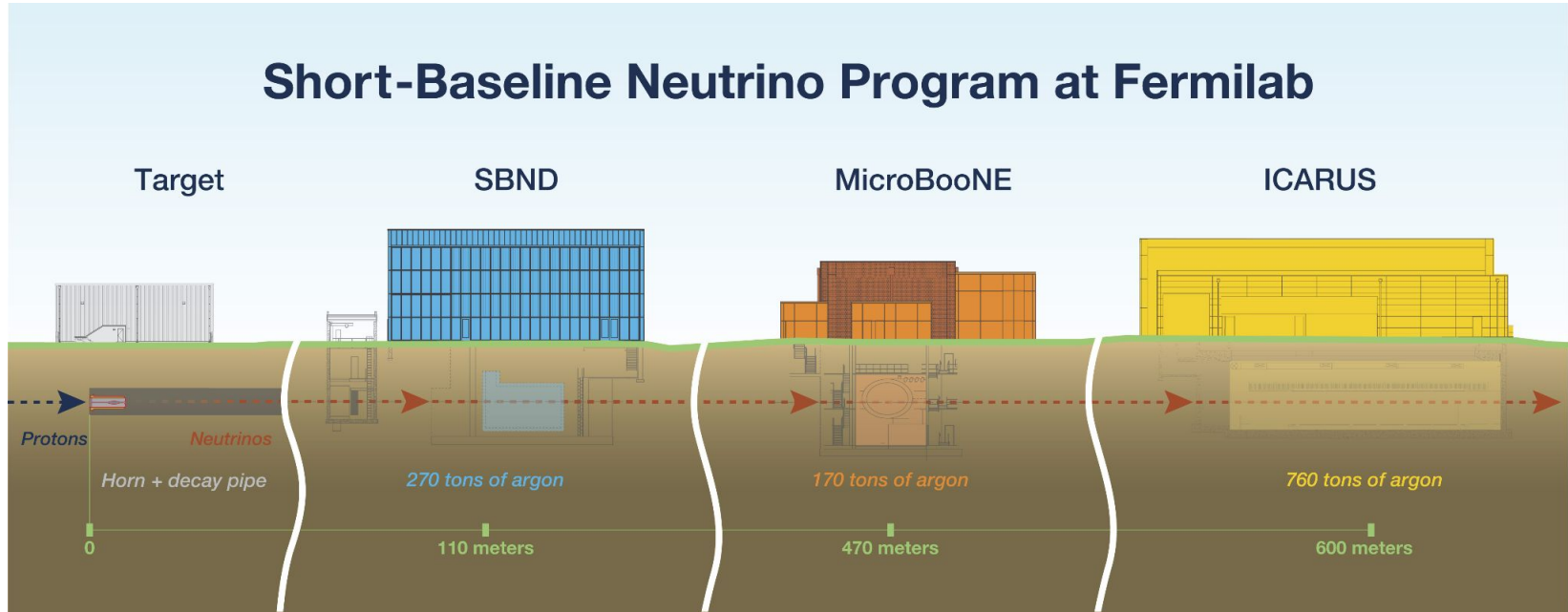To glean information from the profiles and individual builds

Why Data Quality?

Completeness
To determine if there is any missing data

Timeliness
Data should be from the desired timeframe.

Validity
To ensure that data is in the correct format prior to analysis

Integrity
To ensure that the data is reliable and trustworthy

Uniqueness
Data should not be duplicated and should represent distinct information

Consistency
To determine that the data observes the same metrics and standards

3

NuMI and BNB

The Far Detector

Short-Baseline Neutrino Program at Fermilab

Target	SBND	MicroBooNE	ICARUS

Protons

Neutrinos

Horn + decay pipe

270 tons of argon

170 tons of argon

760 tons of argon

0

110 meters

470 meters

600 meters

What is the detector?

TPC

1 T600 module
© 2016-2018 CERN
Wire planes (anode)
Cathode
Field cage
PMTs
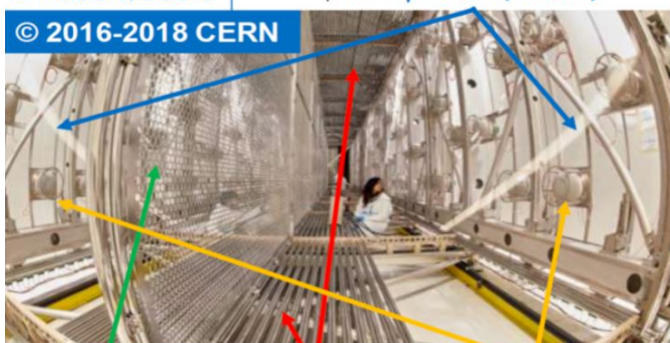
PMT

side CRT

Top CRT
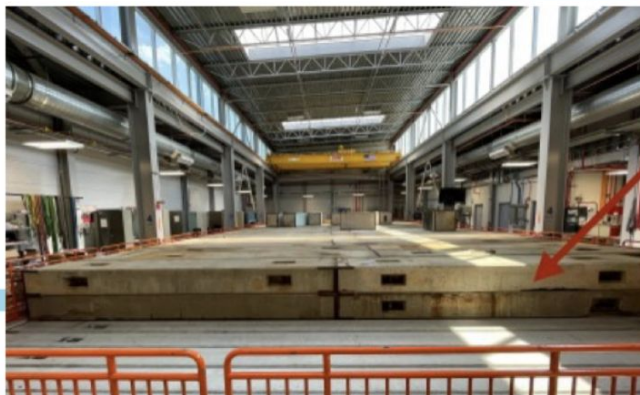
3m concrete overburden

Fermilab

Two Methods to Analyze Data

Profiles
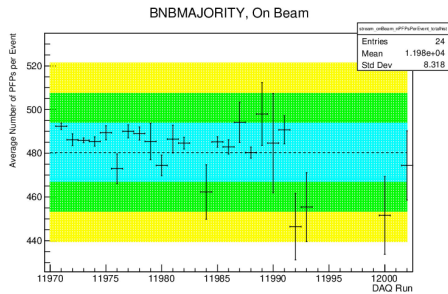
Broad View
Shows the consistency of the data as a function of time

Outliers
Helps identify runs that have issues

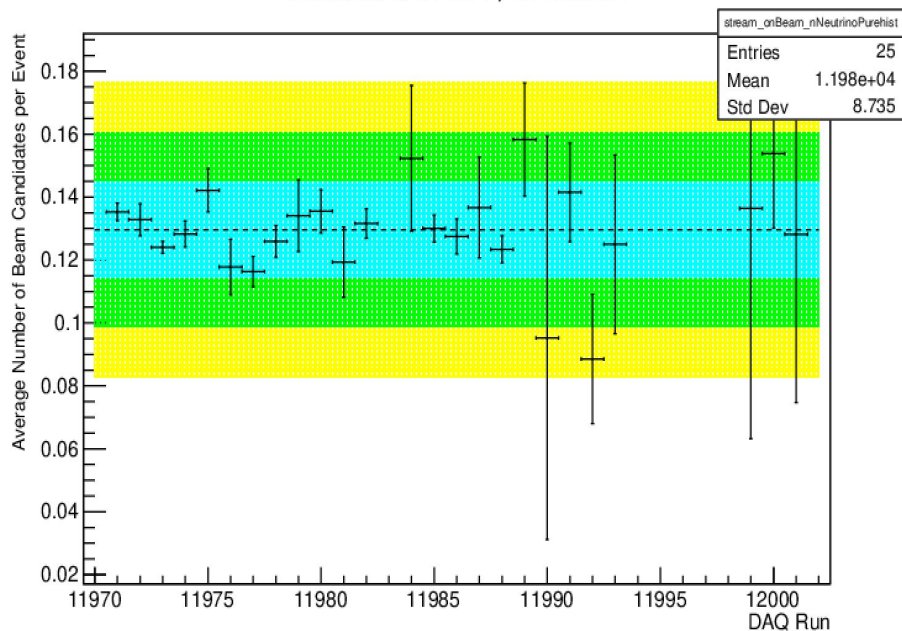## CI Validations
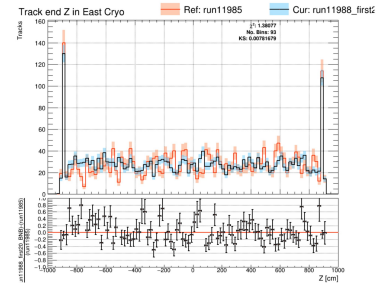
Individual View
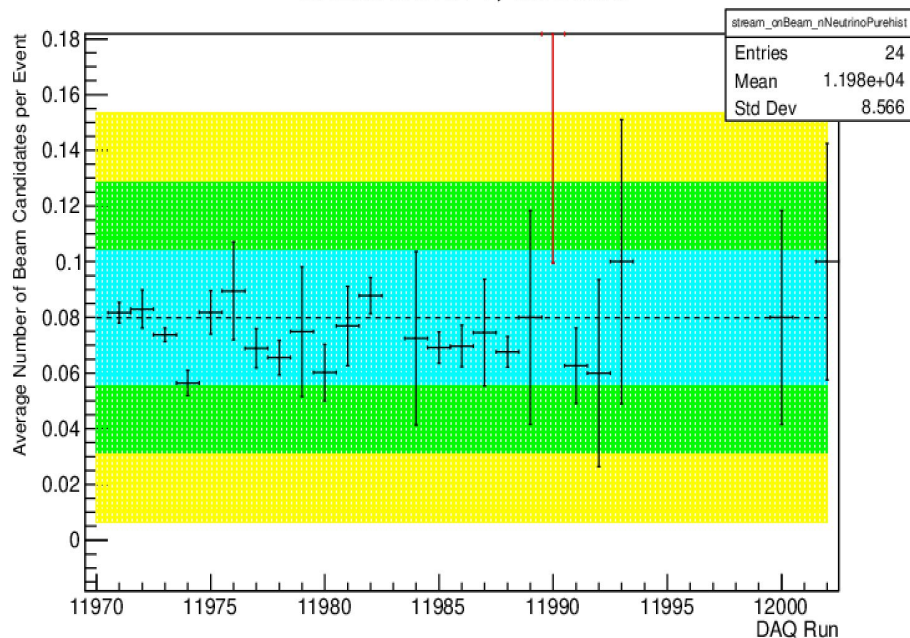Each run can be analyzed individually

In-Depth View
Looks at many qualities from the detector
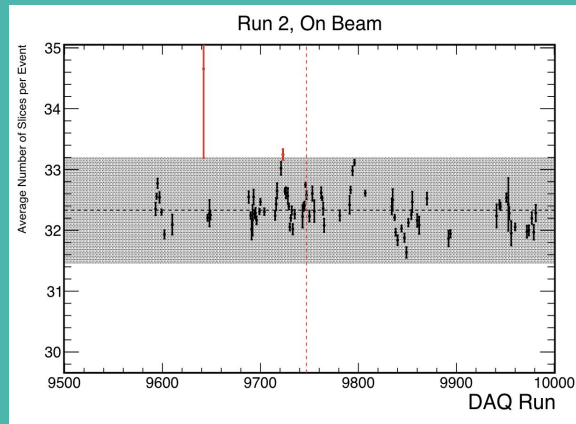
Average Number of Beam Candidates Per Event

Comparing the Different Runs
(NuMI Average Number of Slices)
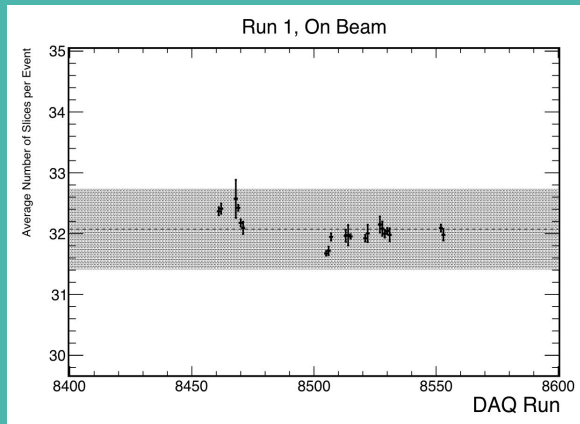
Run 1 (2022)   Run 2 (2023)   Run 3 (2024)
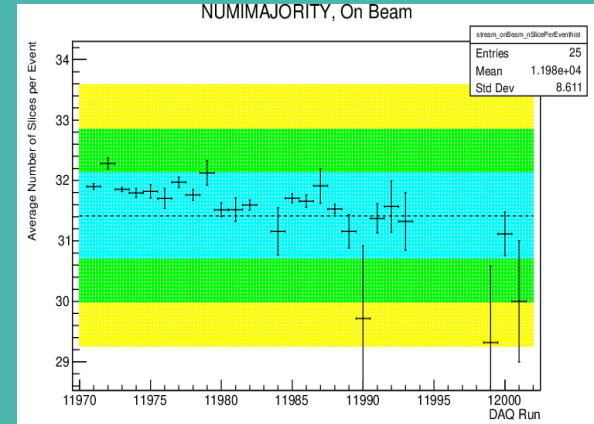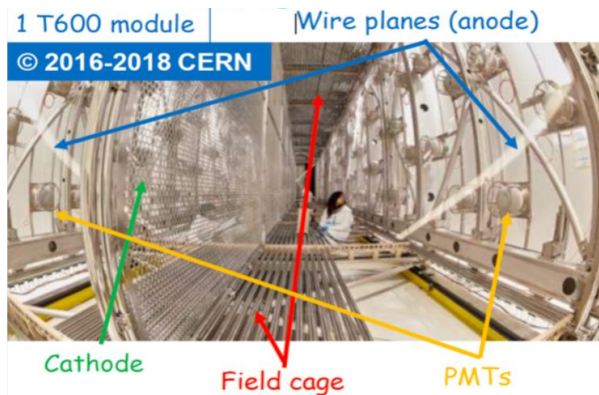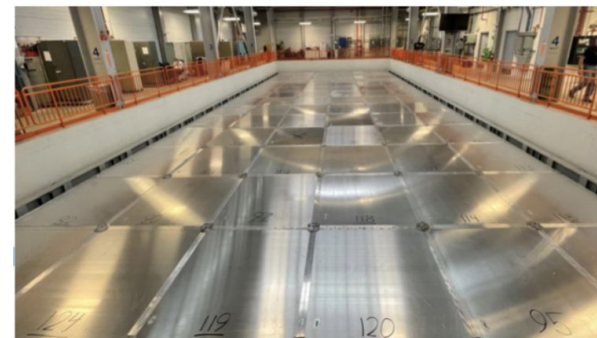
A closer look…

More Vocabulary

TPC - liquid-argon <u>time projection chamber</u>
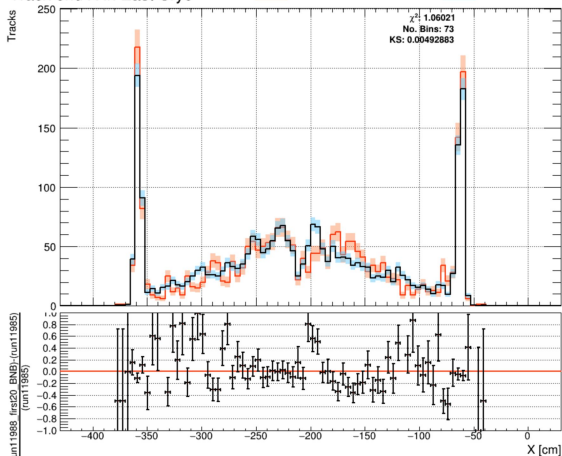

PMT

CRT - Cosmic Ray Tagger

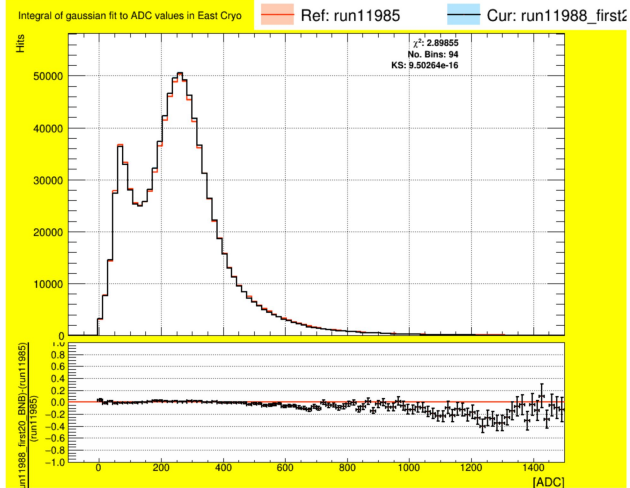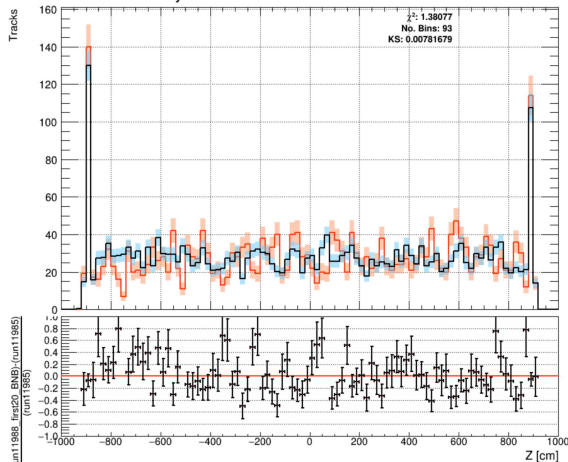# Run 11988 - TPC - BNB

Run 11988 - PMT - BNB
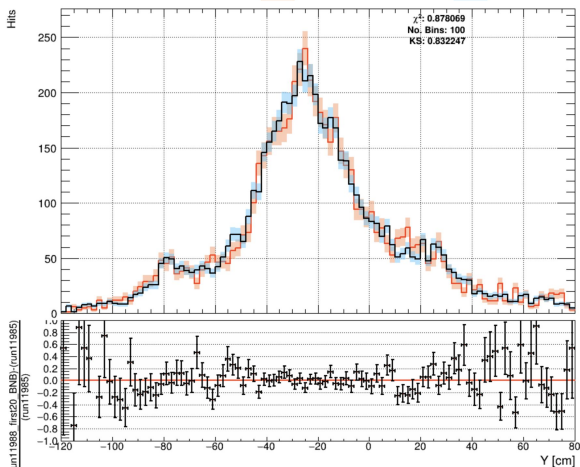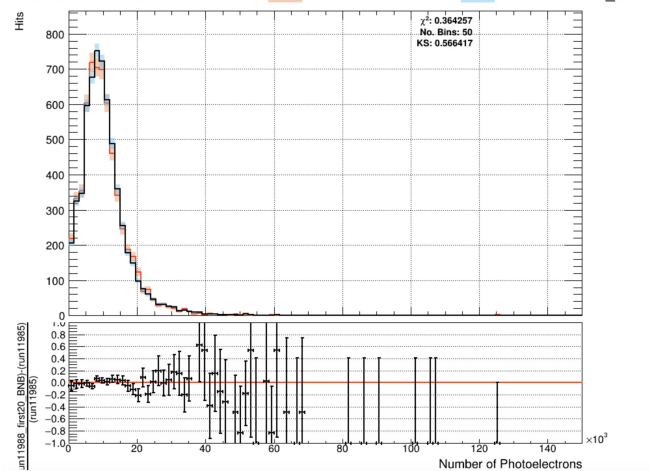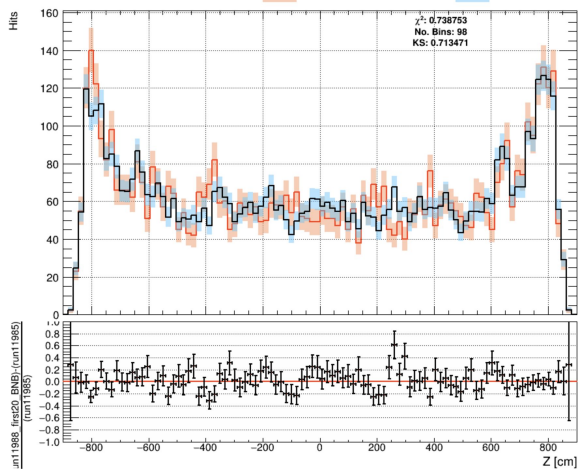
Run 11988 - CRT - BNB
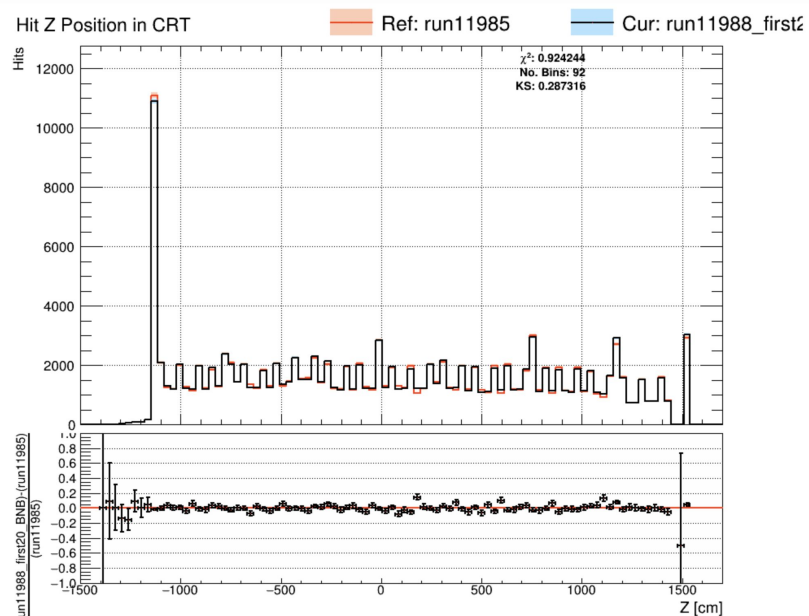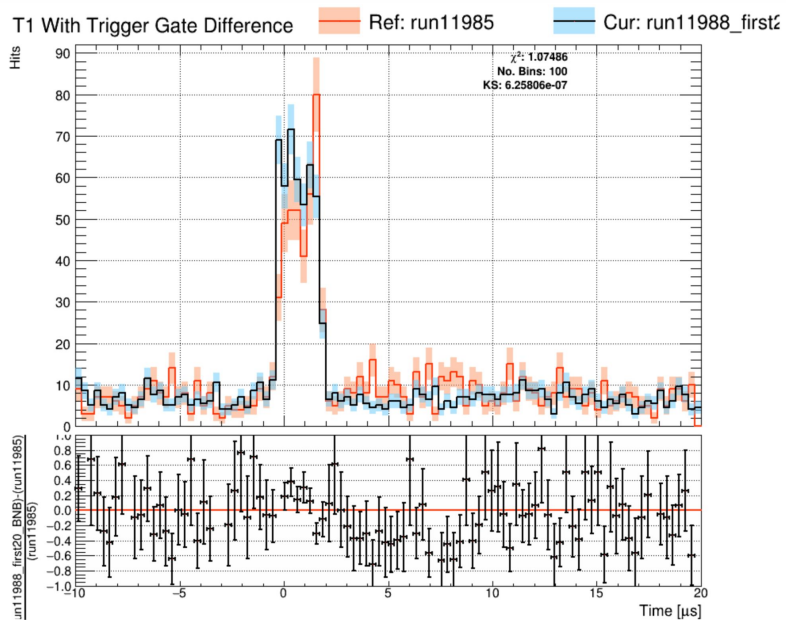
Conclusion / Summary

➢ Two frameworks used for data quality: CI Validation Framework and Profiles
➢ Data from Run 3 looks great!
➢ The data from Run 1, Run 2, and Run 3 is consistent over time
➢ CI Validation is very useful to study details of data from current runs
➢ Running CI Validations with available data in SAMWEB is very manual and time-consuming
  ○ Ran over files that did not have all information and we had to redo work
  ○ Still working on recovering all the data from Run 3
  ○ Automation could be efficient
➢ Next steps → framework will be used for data quality on a daily basis and may be used by the ICARUS collaborators