
REANA
WORKFLOW
FOR DARK MATTER
SEARCHES
CERN
NSF
Princeton
Research Computing
iris
hep
Institute for Research & Innovation
in Software for High Energy Physics

MY MENTORS

Technical University of Munich

Lukas Heinrich

THE UNIVERSITY of WISCONSIN MADISON

Matthew Feickert

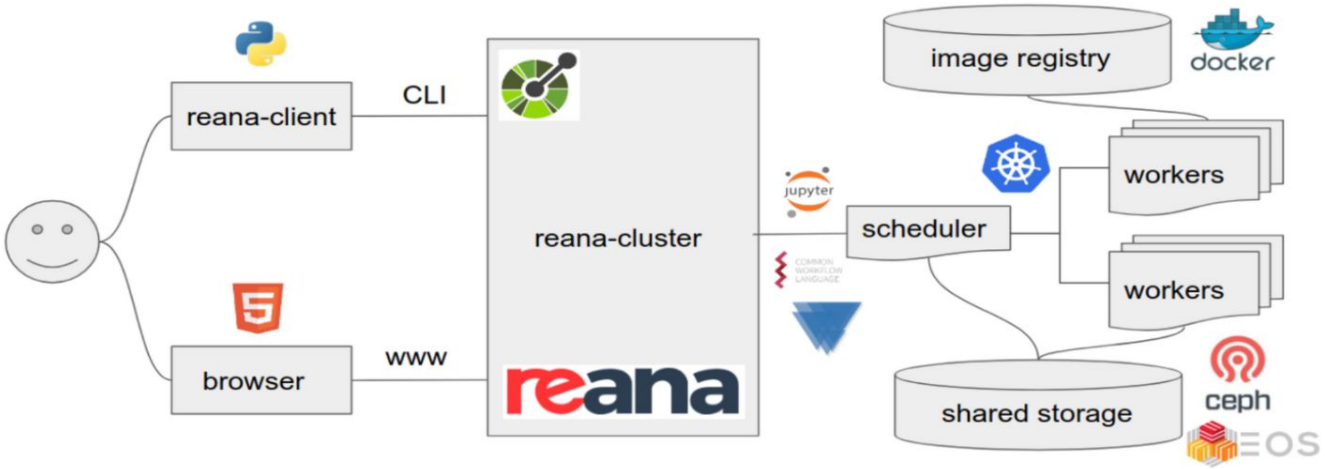
Amy Roberts

UNIVERSITY OF CALIFORNIA SANTA CRUZ

Giordon Stark

Sambridhi Deo

- Fisk University
- Major : Computer Science



[REANA](https://reana.io/) is a platform for reproducible data analysis workflows that can be run at scale. REANA has been used extensively for running containerized workflows of LHC experiments, like ATLAS, and for reinterpretation of published analyses. This project would aim to implement a REANA workflow for a galaxy rotation-curve fitting analysis (RCFM) to improve replicability and to provide a starting point for future work.

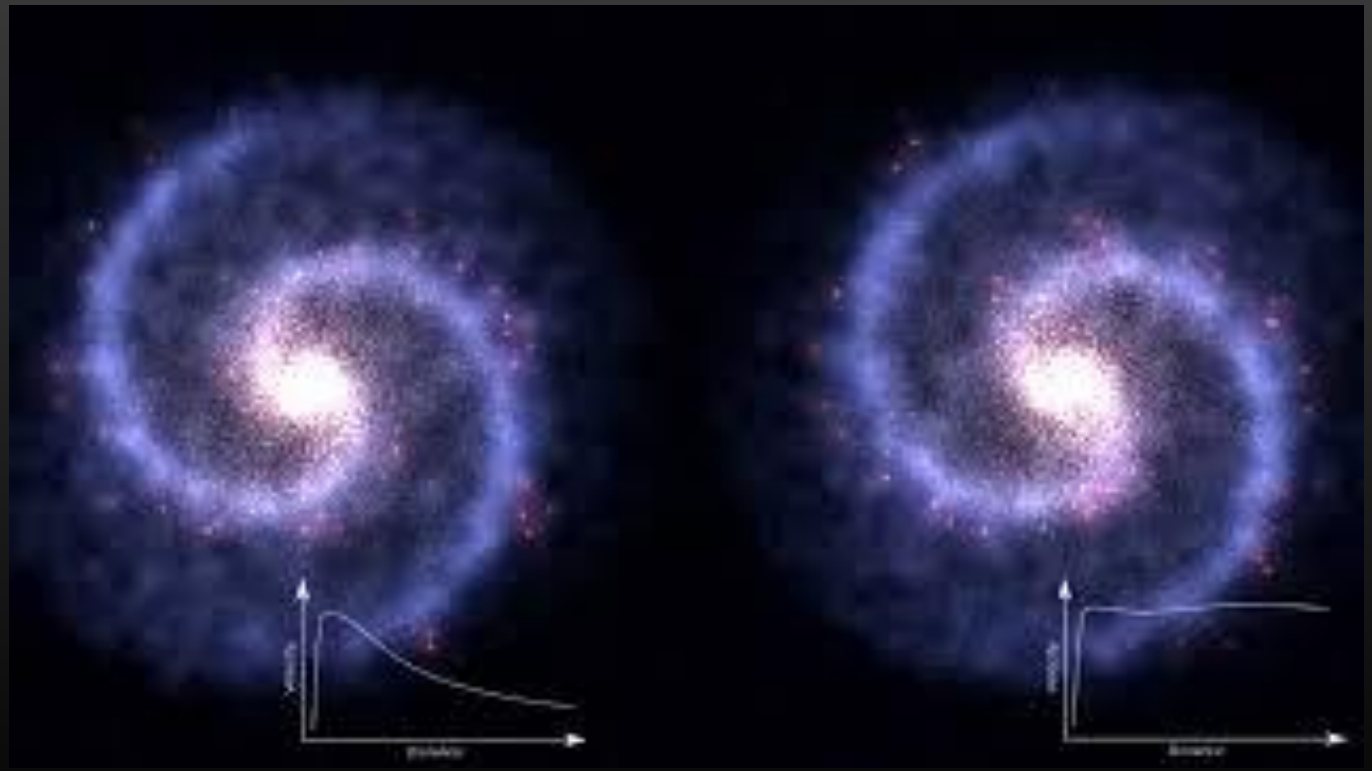
TARGET ANALYSIS

https://github.com/Cisneros-Galaxy/RCFM

Contributors:

- Sophia Cisneros and team

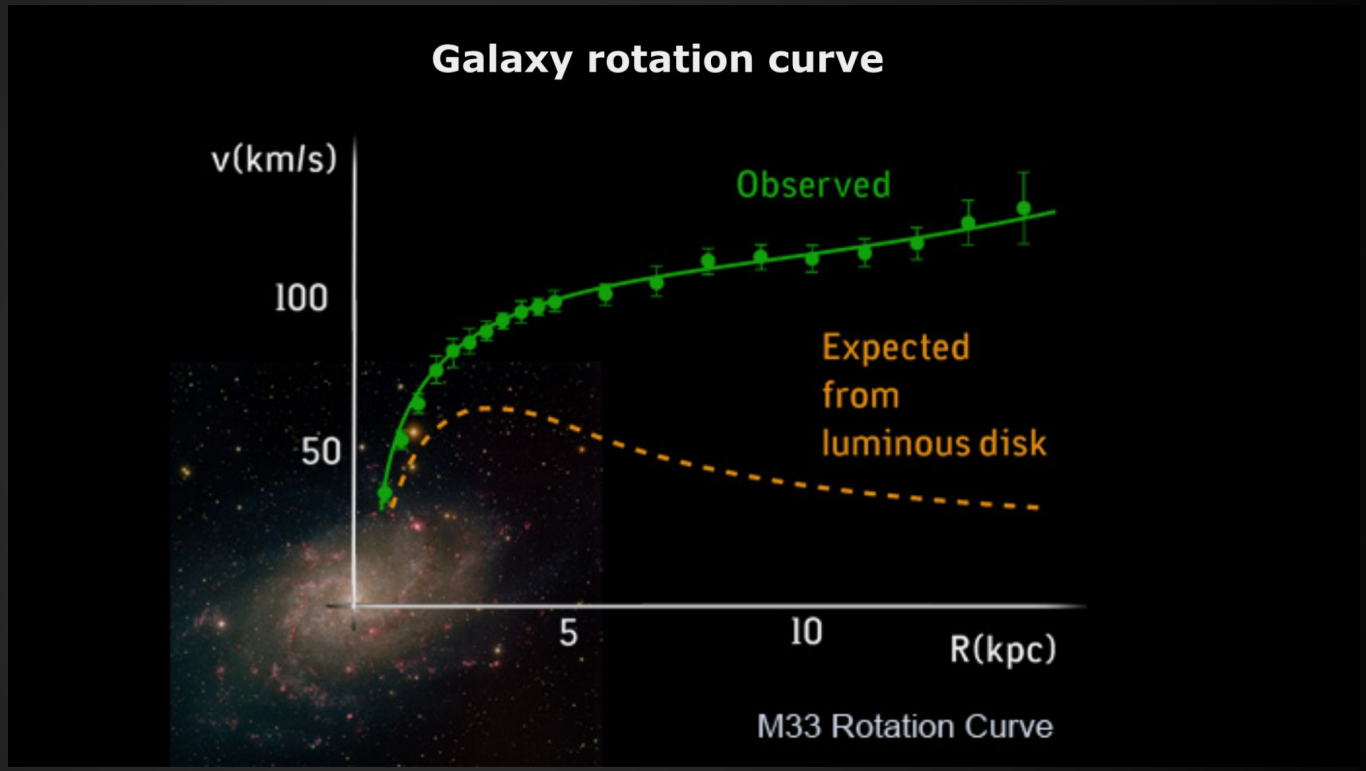
Galaxy don’t rotate the way you think !



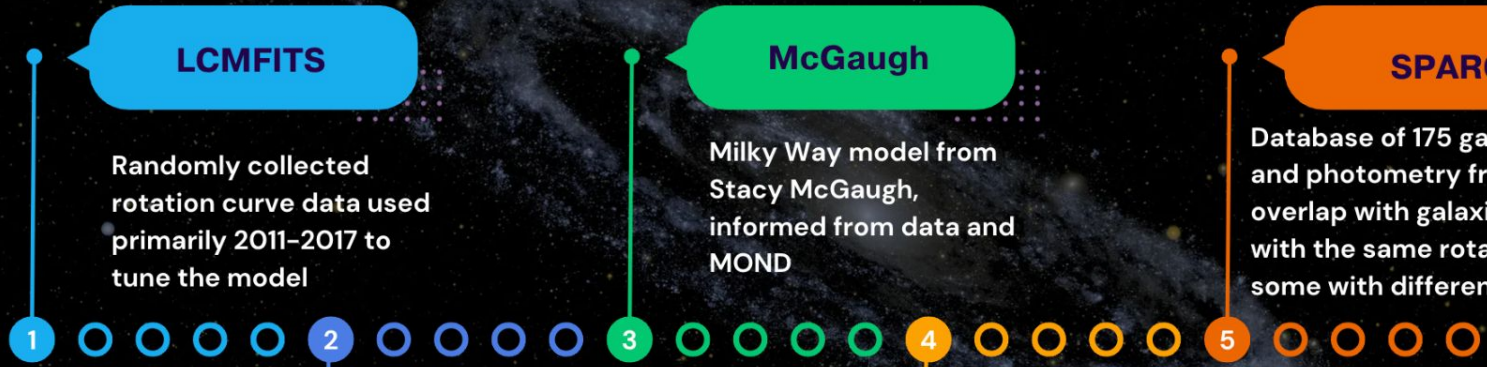
Things orbit fast in center and much more slowly when you go far

Whole galaxy rotating at roughly constant speed

Galactic rotation curves seem to suggest that each galaxy is surrounded by significant amounts of dark matter. Dark matter is composed of particles that do not absorb, reflect, or emit light, so they cannot be detected by observing electromagnetic radiation.



RCFM


LCMFITS
Randomly collected rotation curve data used primarily 2011-2017 to tune the model
McGaugh
Milky Way model from Stacy McGaugh, informed from data and MOND
SPARC
Database of 175 galaxy rotation curves and photometry from McGaugh. Some overlap with galaxies in LCM fits (some with the same rotation curve data and some with different)
1
2
3
4
5
Milky Way model made from Sofue (0-20kpc) and Xue (20-60kpc)
XueSofue
Small dwarf galaxies
little-data-things/data:

```
#Distance
#Radius Vobs    Verr    Vgas    Vdisk   Vbul    These LUminosity are just placeholders
#kpc    km/s    km/s    km/s    km/s    km/s    L/pc^2  L/pc^2
5.635   235.43  18.099  0       137.557 217.33  0.1     0.1
6.787   242.804 10.055  0       148.953 205.263 0.1     0.1
7.94    251.518 9.385   0       157.668 194.538 0.1     0.1
9.053   262.244 9.385   0       163.701 185.823 0.1     0.1
10.205  259.563 9.385   0       169.064 177.779 0.1     0.1
11.318  255.54  9.386   12.2    172.416 171.075 0.1     0.1
12.47   252.189 16.759  19.574  174.427 164.372 0.1     0.1
13.583  252.859 12.541  26.948  175.768 159.009 0.1     0.1
14.736  251.518 20.682  34.322  175.768 153.646 0.1     0.1
15.848  245.485 30.715  38.344  175.768 148.953 0.1     0.1
17.001  232.748 9.385   38.344  175.097 144.931 0.1     0.1
18.153  231.408 14.747  36.333  173.757 140.909 0.1     0.1
19.306  235.43  10.055  34.992  172.416 136.887 0.1     0.1
20.419  230.067 20.233  34.322  169.734 133.535 0.1     0.1
21.452  227.385 28.826  34.992  167.723 130.854 0.1     0.1
22.485  226.045 28.155  36.333  166.383 128.172 0.1     0.1
23.479  226.045 28.155  36.333  164.372 126.161 0.1     0.1
24.472  227.385 28.826  37.003  161.69  123.48  0.1     0.1
25.505  226.715 29.496  35.662  159.679 120.798 0.1     0.1
26.539  224.704 29.496  34.992  157.668 119.457 0.1     0.1
27.572  224.704 28.155  34.992  154.986 117.446 0.1     0.1
28.566  222.022 28.826  34.322  152.975 114.765 0.1     0.1
29.639  221.352 29.496  32.311  150.964 113.424 0.1     0.1
30.632  224.704 28.825  32.311  148.283 111.413 0.1     0.1
31.665  228.056 28.825  30.97   146.942 110.072 0.1     0.1
32.699  231.408 28.155  30.3    144.261 108.061 0.1     0.1
33.692  230.737 28.155  29.629  143.59  106.721 0.1     0.1
34.725  227.385 27.485  29.629  143.59  105.38  0.1     0.1
```

Data

Most datas are in .dat or .csv format.

RCFM Model

1. Import modules and helper functions

```
# Modules
import matplotlib.pyplot as plt
import numpy as np
from math import sqrt
from scipy.optimize import curve_fit

# Helper functions from DataAid.py and DataImport.py
import DataAid
import DataImporter

# Numerically stable class of functions from Neros_v2.py
import Neros
```

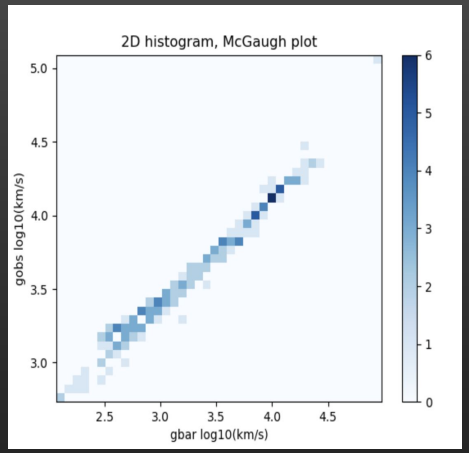
[1]

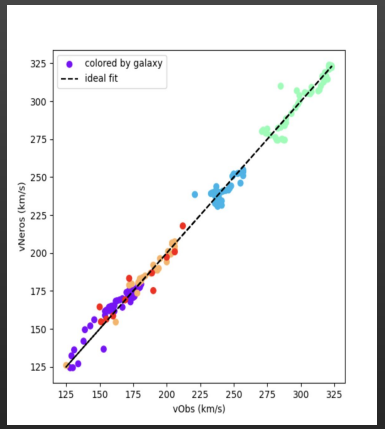
2. Load Galaxy Data

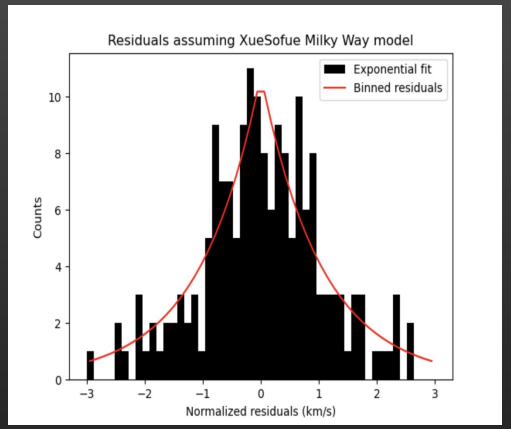
```
# Load Galaxy Data
sparcGalaxies = DataAid.GetGalaxyData("data/Sparc/Rotmod_LTG/")
sparc128Galaxies = DataAid.GetGalaxyData("data/Sparc/SparcSubset135/")
sparcTset = DataAid.GetGalaxyData("data/Sparc/TrainingSet/")
littleDataGalaxies = DataAid.GetGalaxyData("data/little-data-things/data/")
lcmGalaxies = DataAid.GetGalaxyData("data/LCMFits/data/")

# Load Milky Way Model Data
xueSofueGalaxies = DataAid.GetGalaxyData("data/XueSofue/")
mcGaughMW = DataAid.GetGalaxyData("data/McGaugh/")

# Create array of Milky Way radius and vlum tuples from model data
MWXueSofue = np.array(xueSofueGalaxies['MW_lum'])
MWMcGaugh = np.array(mcGaughMW['MW_lumMcGaugh'])
```

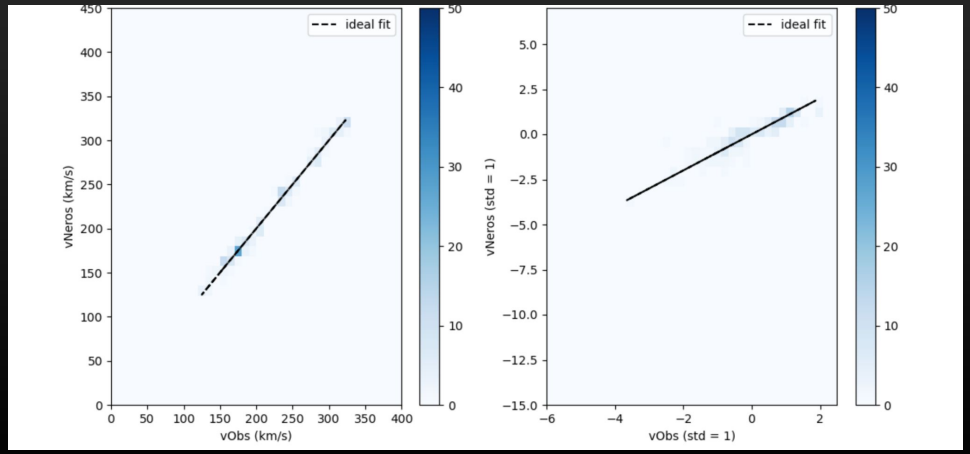
[2]

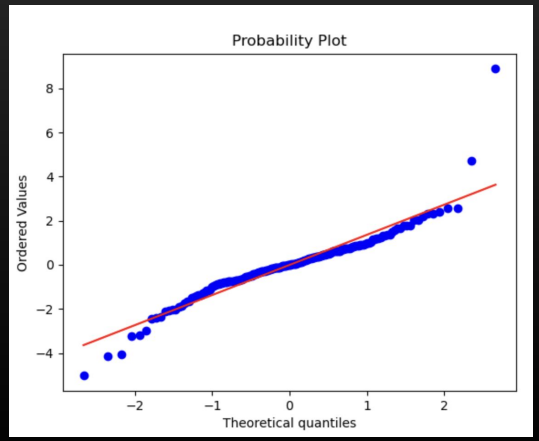
Fit-analysis plots



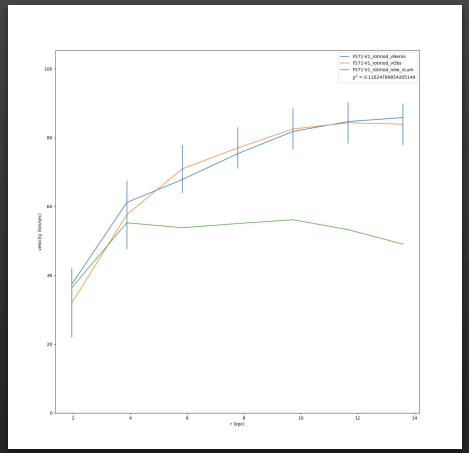


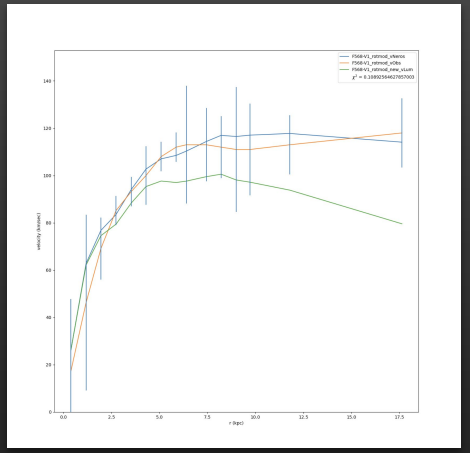


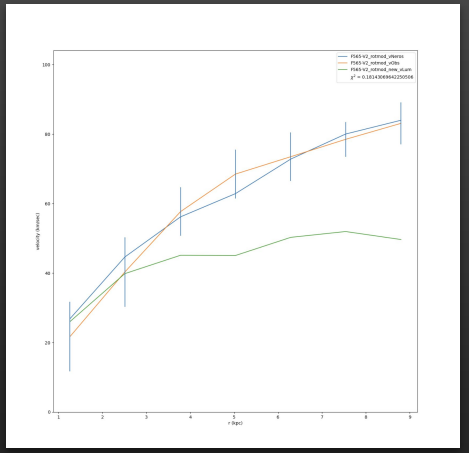


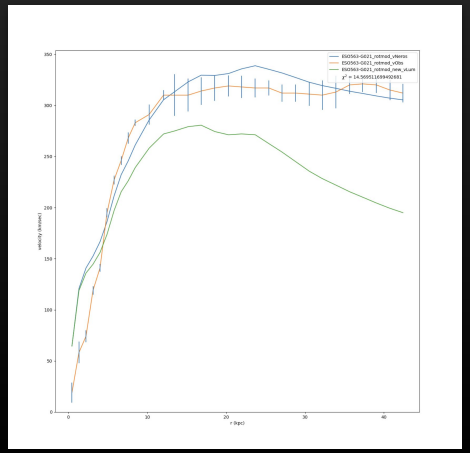


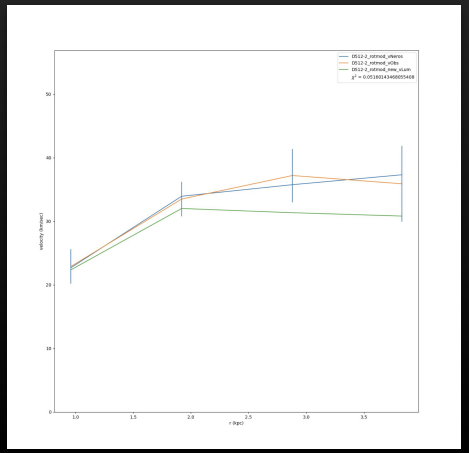
Graphs

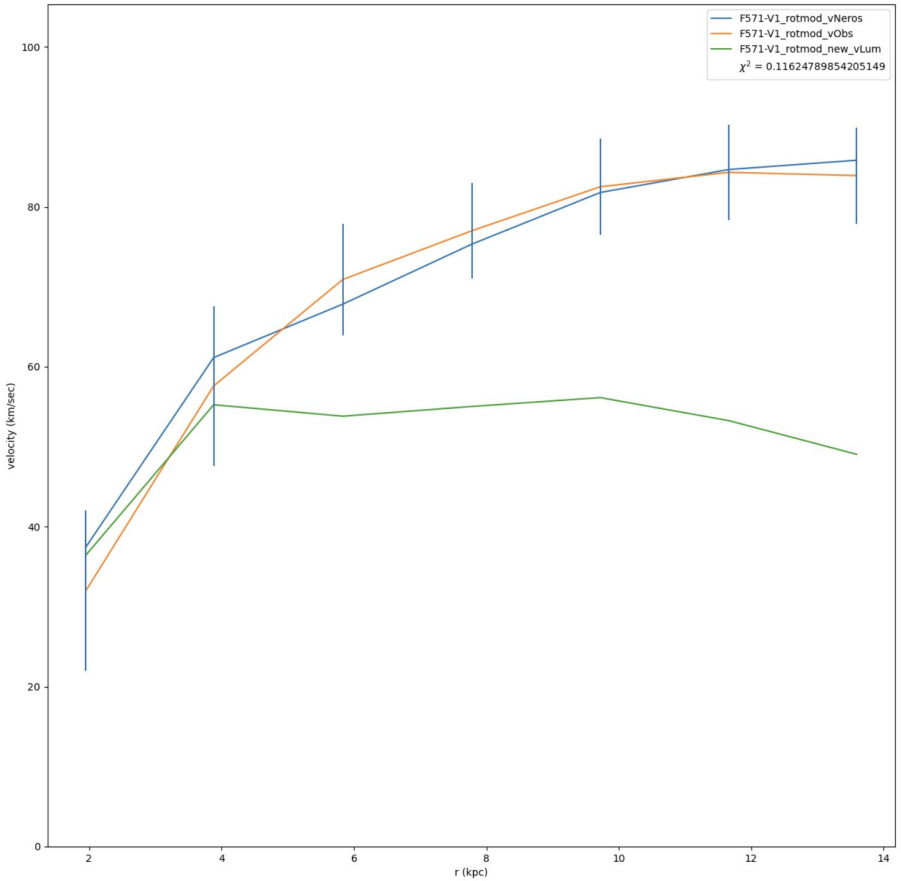

F571-V1_rotmod_vtheros
F571-V1_rotmod_vObs
F571-V1_rotmod_new_vLum
χ² = 0.11824789854201149
velocity (km/sec)
r (kpc)


F568-V1_rotmod_vtheros
F568-V1_rotmod_vObs
F568-V1_rotmod_new_vLum
χ² = 0.18892564627057003
velocity (km/sec)
r (kpc)


F565-V2_rotmod_vtheros
F565-V2_rotmod_vObs
F565-V2_rotmod_new_vLum
χ² = 0.18143089842250546
velocity (km/sec)
r (kpc)


E50563-G021_rotmod_vtheros
E50563-G021_rotmod_vObs
E50563-G021_rotmod_new_vLum
χ² = 14.569511689492691
velocity (km/sec)
r (kpc)


D512-2_rotmod_vtheros
D512-2_rotmod_vObs
D512-2_rotmod_new_vLum
χ² = 0.05180143468855408
velocity (km/sec)
r (kpc)


F571-V1_rotmod_vNeros
F571-V1_rotmod_vObs
F571-V1_rotmod_new_vLum
χ^2 = 0.11624789854205149
velocity (km/sec)
r (kpc)
0
20
40
60
80
100
2
4
6
8
10
12
14

ROADMAP

Research

Literature review on galaxy rotation-curve fitting analysis techniques, and familiarization with the REANA platform and Linux container technology (e.g. Docker). The various accounts for use of computing resources and access to data and simulation will be acquired.

Week 1-3

REANA Workflow

Get started on REANA workflow framework. Conduct initial testing to ensure functionality and compatibility. Begin writing documentation concurrently with software development.

Week 7-8

Week 4- 6

Study and contanerize

Containerize all analysis software necessary for running the selected RCFM analysis.

Week 9-10

Optimize

Refine and optimize the workflow to enhance its performance, scalability, and reproducibility. Identify and address any potential issues or improvements.

Week 11 -12

Present

Prepare final presentations and finalize the documentation with the setup instructions, workflow execution, and interpretation of results.

October 31st

DONE!

What I did?

- Leveraged Conda Environment
  - Created an Conda environment using an environment.yml file to specify the required packages and dependencies.

- Played around with REANA with example analysis:
  - Experimented with REANA using a sample analysis to understand its functionality and capabilities.

- Created reana.yaml file with the workflow specification:
  - Developed a reana.yaml file to define the workflow for dark matter searches. This file includes specifications for input data, workflow structure, environment, and commands/scripts for each step.

RCFM2 > ! reana.yaml

```yaml
version: 0.9.1
inputs:
  files:
# PYTHON FILES ⋯

# CSV FILES ⋯

# DAT FILES LCMFITS ⋯

# DAT FILES LITTLE-DATA ⋯

# McGaugh txt ⋯

# Sparc Files ⋯

# XUESOFUE ⋯

# Phi calculations txt ⋯



workflow:
  type: serial
  specification:
    environment:
      name: rcfm-envi
    steps:
      - name: data_import
        commands:
          - python DataImporter.py
        inputs:
        #imported data
        - RCFM2/imported-data/galaxy
        - RCFM2/imported-data/1data_MWXueSofue.csv
        - RCFM2/imported-data/data_McGaugh.csv
        - RCFM2/imported-data/data_MWMcGaugh.csv
        - RCFM2/imported-data/data_MWXueSofue.csv
        - RCFM2/imported-data/data_XueSofue.csv
        - RCFM2/imported-data/data.csv

      # Data
      # DAT FILES LCMFITS
        - RCFM2/data/LCMFits/data
        - RCFM2/data/LCMFits/M31_Carignan_SNC.dat
        - RCFM2/data/LCMFits/M31_Carignan_SNC2.dat
        - RCFM2/data/LCMFits/M33_Corbelli_LCM.dat
        - RCFM2/data/LCMFits/NGC891_Frat11_IsothermDM.dat
```

RCFM2 > ! reana.yaml

```yaml
    # XUESOFUE
        - RCFM2/data/XueSofue/MW_lum.dat

    # Phi calculations txt
        - RCFM2/phiCalculations/NGC5371.txt

    # Graphs
        - RCFM2/graphs

        #Fit analysis plots
        - RCFM2/fit-analysis-plots/MWXueSofue


        outputs:
          - name: imported_data

      - name: data_processing
        commands:
          - jupyter nbconvert --execute --to notebook --inplace model.ipynb
        inputs:
          - name: imported_data
          - name: DataAid.py
          - name: Neros.py
          - name: Neros_test.py
          - name: rotCurve.py
        outputs:
          - name: analysis_results

      - name: generate_plots
        commands:
          - jupyter nbconvert --execute --to notebook --inplace model.ipynb
        inputs:
          - name: analysis_results

        outputs:
          - name: png_files
            files:
            #Graphs
            - RCFM2/graphs

            #Fit analysis plots
            - RCFM2/fit-analysis-plots/MWXueSofue
```

Your workflows

Refreshed at 10:04:51 UTC

Search...

Status | Show deleted runs | Latest first

rcfm-analysis #2
36 KiB
Finished a day ago
failed after 10h 23m 10s
step 0/3

rcfm #2
1.07 MiB
Finished a day ago
failed after 11h 45m 9s
step 0/3

rcfm #1
1.07 MiB
Finished a day ago
failed after 11h 59m 37s
step 0/3

roofit #1
173 KiB
finished in 1 min 19 sec
step 2/2

Still Debugging! :)
Target Analysis

Example Analysis

ROADMAP

Research

Literature review on galaxy rotation-curve fitting analysis techniques, and familiarization with the REANA platform and Linux container technology (e.g. Docker). The various accounts for use of computing resources and access to data and simulation will be acquired.

Week 1-3

REANA Workflow

Get started on REANA workflow framework. Conduct initial testing to ensure functionality and compatibility. Begin writing documentation concurrently with software development.

Week 7-8

Week 4- 6

Study and contanerize

Containerize all analysis software necessary for running the selected RCFM analysis.

Week 9-10

Optimize

Refine and optimize the workflow to enhance its performance, scalability, and reproducibility. Identify and address any potential issues or improvements.

Week 11 -12

Present

Prepare final presentations and finalize the documentation with the setup instructions, workflow execution, and interpretation of results.

October 31st

DONE!

Sincerest Gratitude!

- IRIS-HEP Fellowship team
- Mentors
- Target analysis contributors