Measuring Galactic Dark Matter through Unsupervised Machine Learning

Sung Hak Lim

Rutgers University

RUTGERS

Phenomenology 2023,
University of Pittsburgh
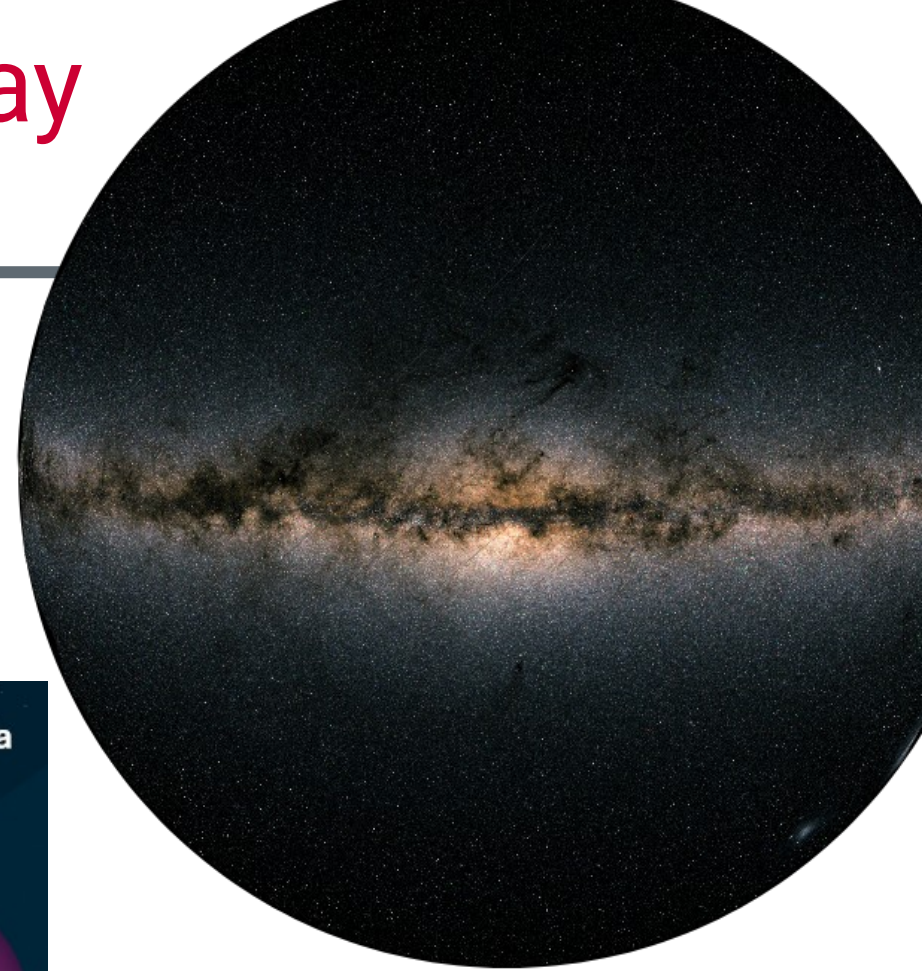
May. 2023

A Snapshot of Milky Way from Gaia

Recently, Gaia mission released a new catalog containing very detailed measurement of stars in the Milky Way that can be used for various physics analysis.
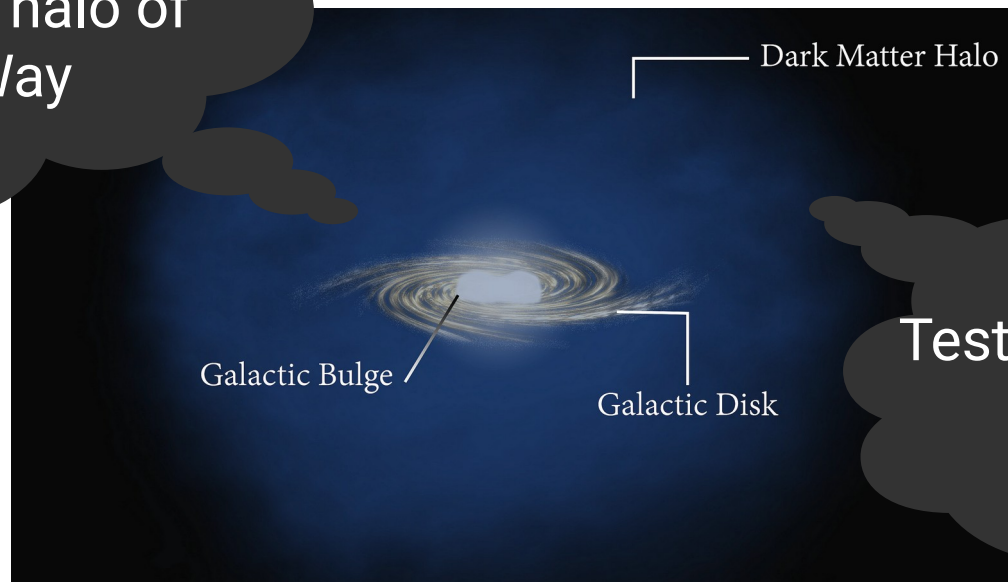

All the stars are under the gravitational field of the Milky way.
We could use this dataset to understand the **dark matter distribution** of the **Milky Way**

Image from https://gea.esac.esa.int/archive/visualization/

Why understanding galactic dark matter is important?

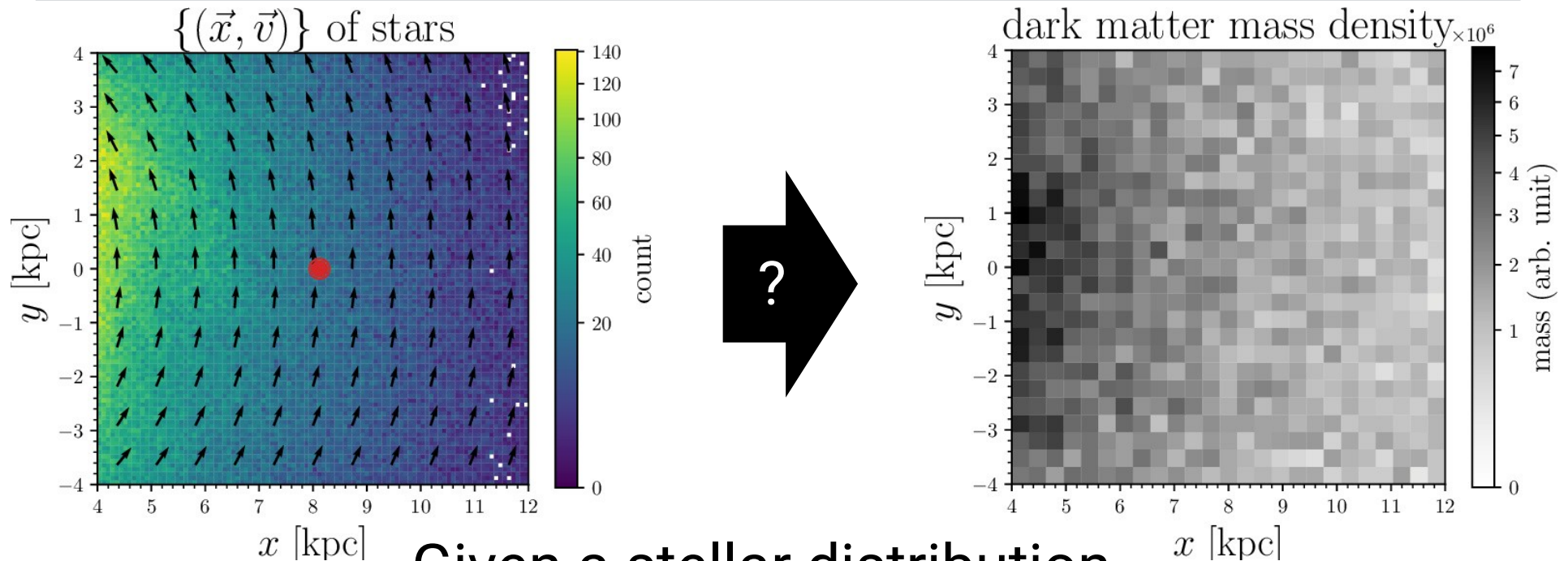Inputs to Direct Detection experiments
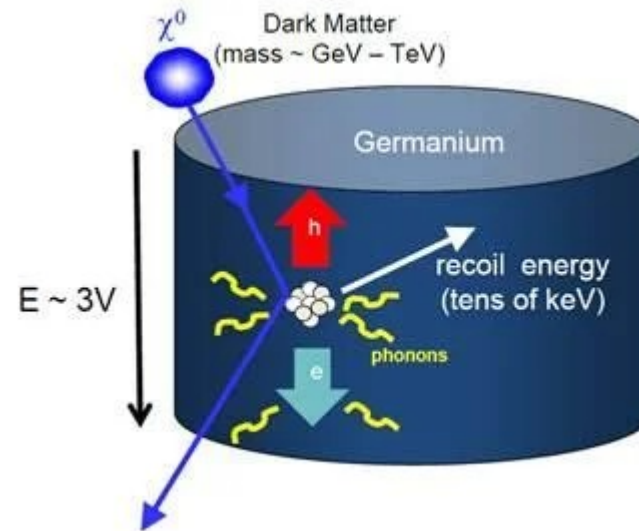
Understanding the dark matter halo of the Milky Way

Testing modified gravity solutions?

Main Question:

Given a stellar distribution,
Infer the **local galactic dark matter density**
from observed position and velocities of stars.

+ Now we have a huge dataset. Can we do the estimation
without assuming **symmetries** and **models**?

Conventional methods assumes symmetries and
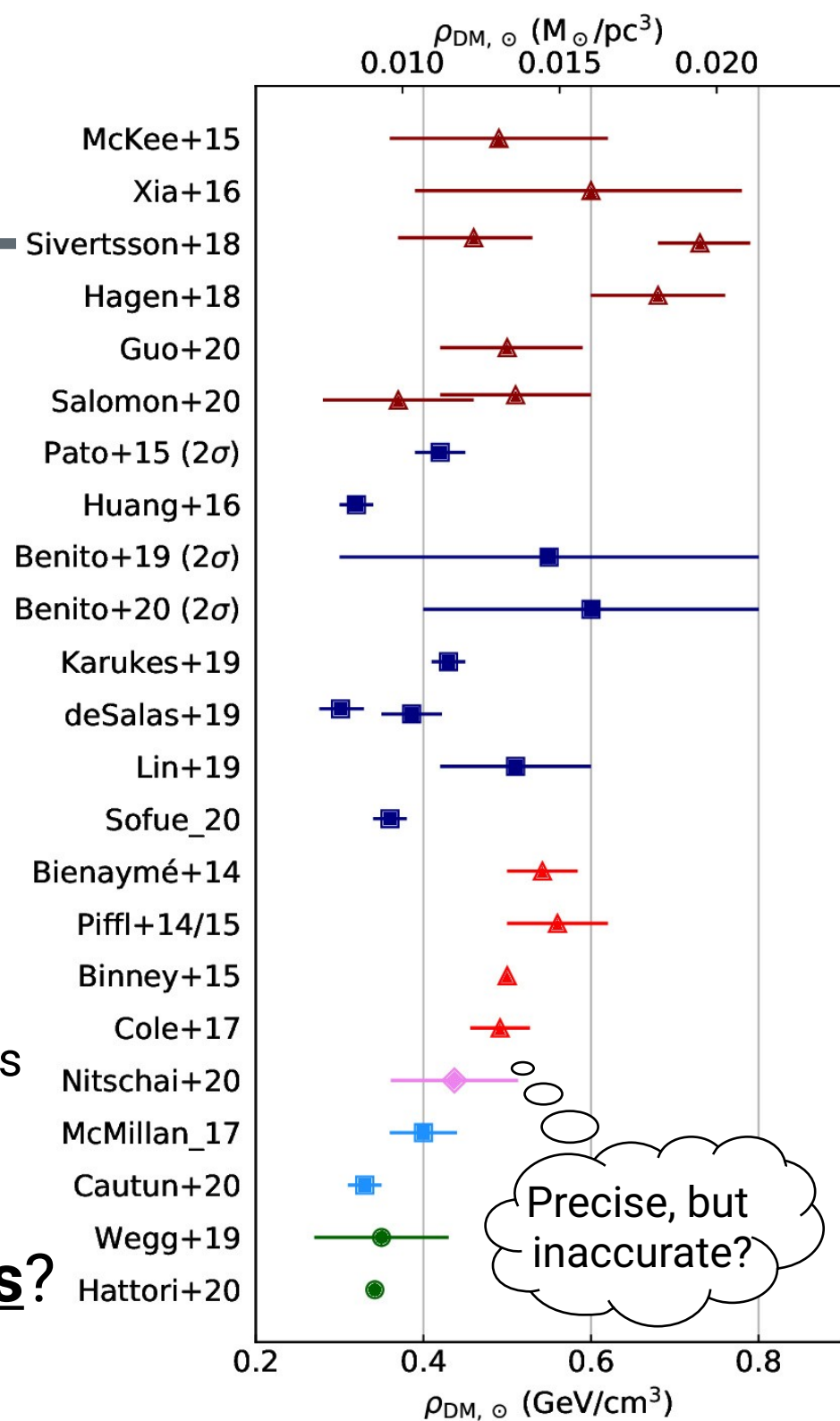physics-based density models to reduce complexity of modeling 6D density.

Accuracy Matters

Thanks to recent progress in observing stars in the Milky Way, we can measure **the dark matter density in the Solar neighborhood** in very high precision using model-based analysis.

Accuracy

Precision
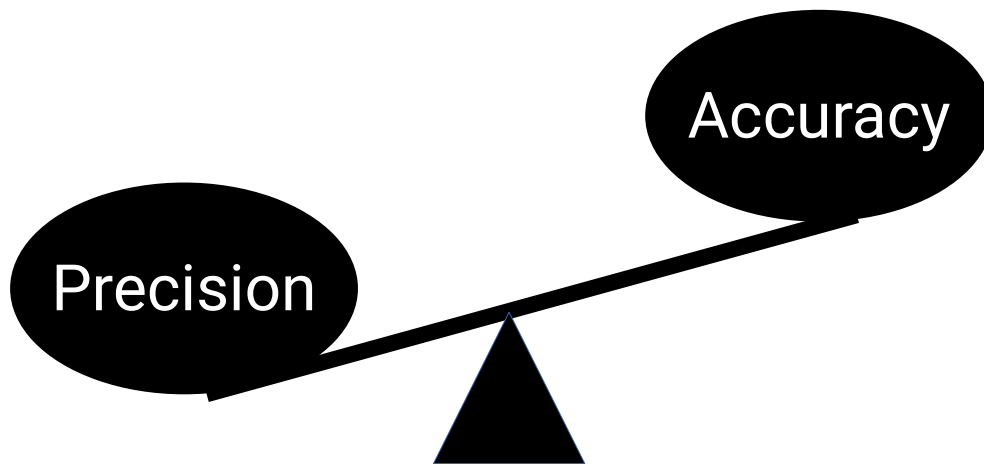
When sufficient number of data are available, using overconstrained models may results in inaccurate results.

Need of analysis without assumed **symmetries** and **models**?

Figure from P. F. de Salas, and A. Widmark, arXiv:2012.11477

$\rho_{DM, \odot}$ $(M_\odot/pc^3)$

Precise, but inaccurate?

$\rho_{DM, \odot}$ (GeV/cm^3)

Solution: solve equation of motion of the stars in the galaxy
$$f(\vec{x}, \vec{v})d\vec{x}d\vec{v}$$

How do we infer the dark matter density from the stellar distribution?
→ Solve equation of motion of stellar phase-space density
   in order to get gravitational acceleration and mass density.

$$\left[\frac{\partial}{\partial t} + \vec{v} \cdot \frac{\partial}{\partial \vec{x}} + \vec{a} \cdot \frac{\partial}{\partial \vec{v}} \right] f(\vec{x}, \vec{v}) = 0, \quad \vec{a} = -\frac{d\Phi(\vec{x})}{d\vec{x}}$$

Assuming that the galaxy is in dynamic equilibrium (∂f/∂t = 0),
we could estimate the acceleration field a(x) from
the Milky Way snapshot at the current time.

In order to solve this equation, we first have to estimate the
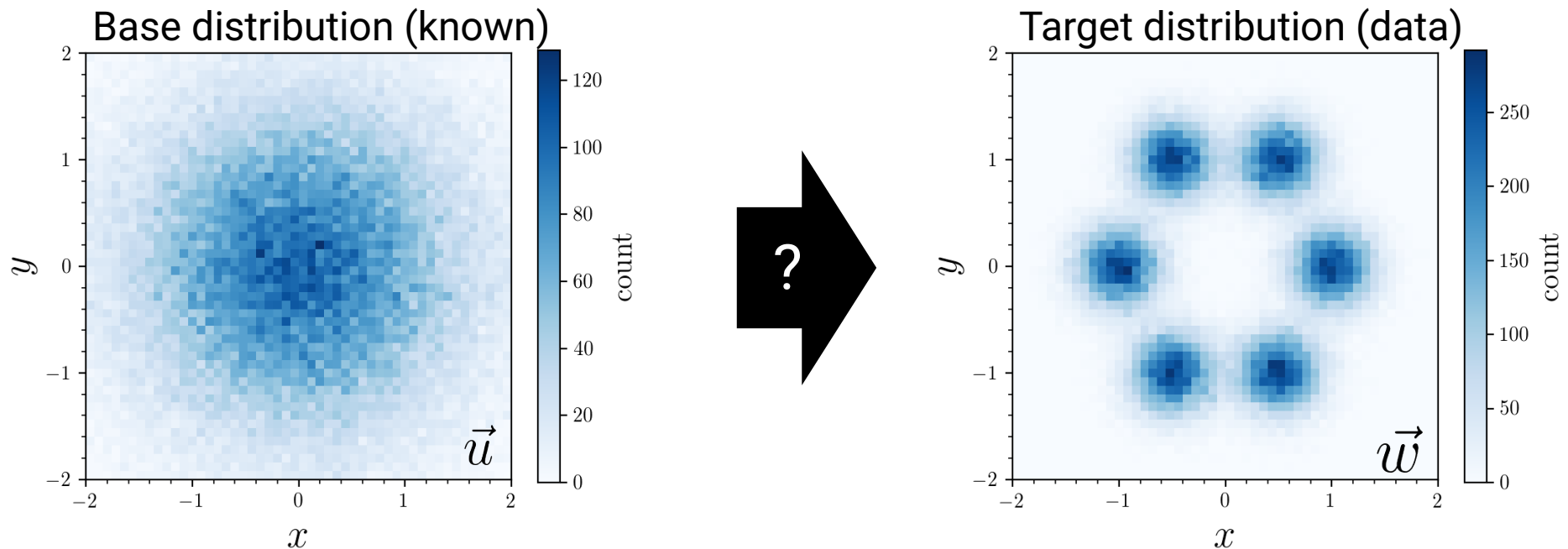6D phase space density very precisely.

$$\{(\vec{x}, \vec{v})\} \rightarrow f(\vec{x}, \vec{v})$$

Neural network-based density estimation technique:
<u>Normalizing Flows</u>

Normalizing Flows:
Neural Network learning a Transformation

Normalizing Flows (NFs) is an artificial neural network that learns a transformation of random variables.
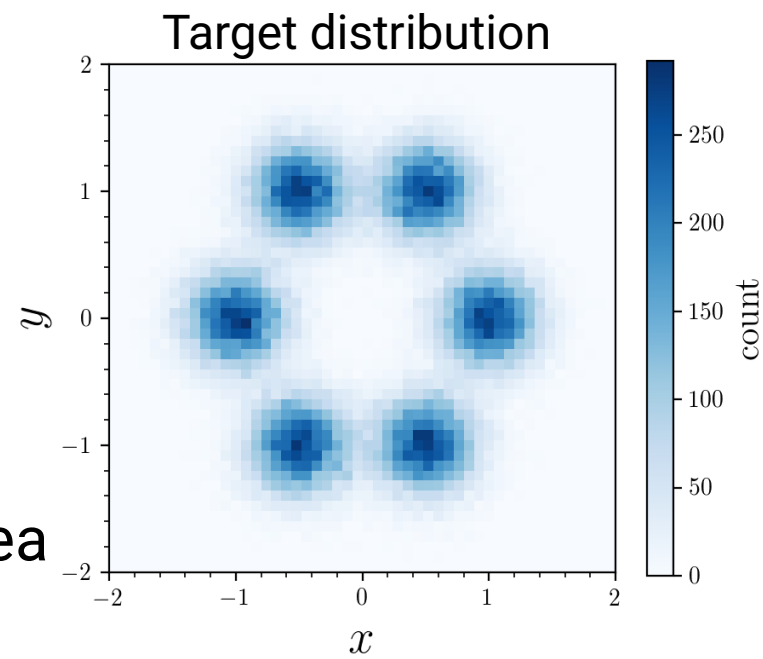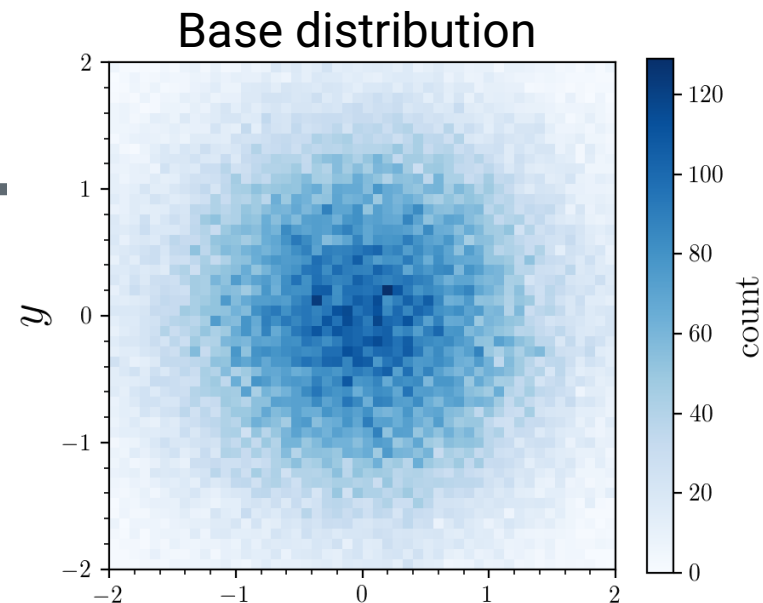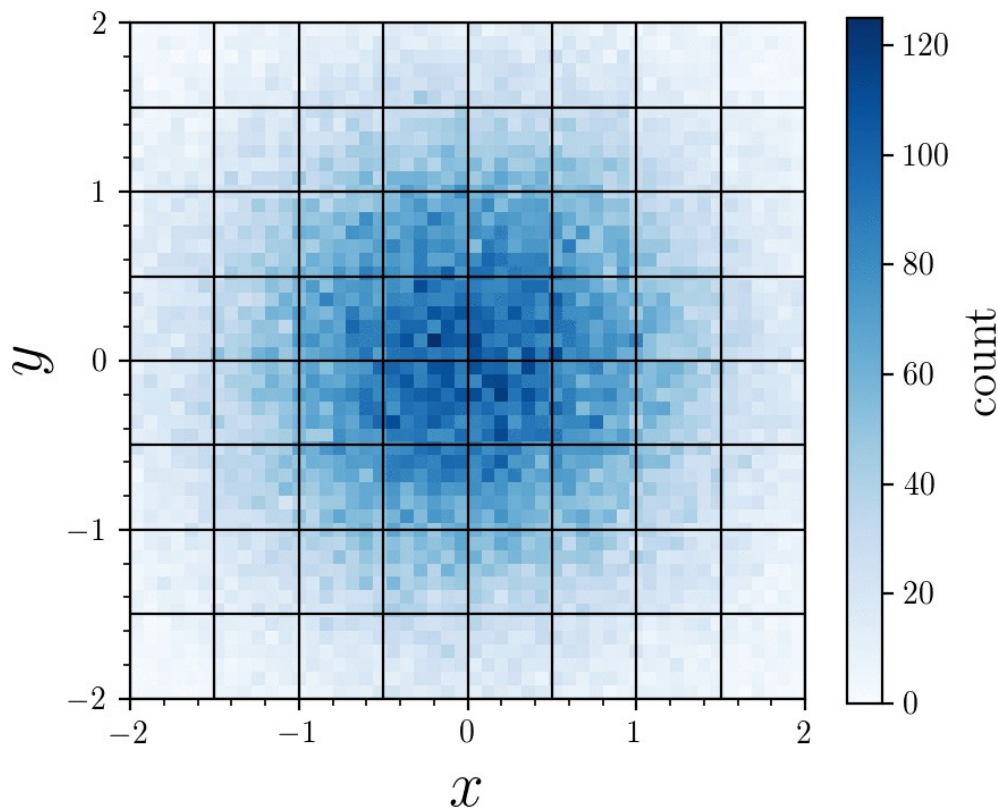

Main idea: if we could find out such transformation, we can use the transformation formula for the density estimation:

$$p_W(\vec{w}) = p_U(\vec{u}) \cdot \left| \frac{d\vec{u}}{d\vec{w}} \right|$$

We will use this model for estimating the phase space density f(x,v) from the data.

Normalizing Flows: How it works?

Base distribution

Target distribution

As a proof-of-concept, we will test this idea on an N-body simulated galaxy.
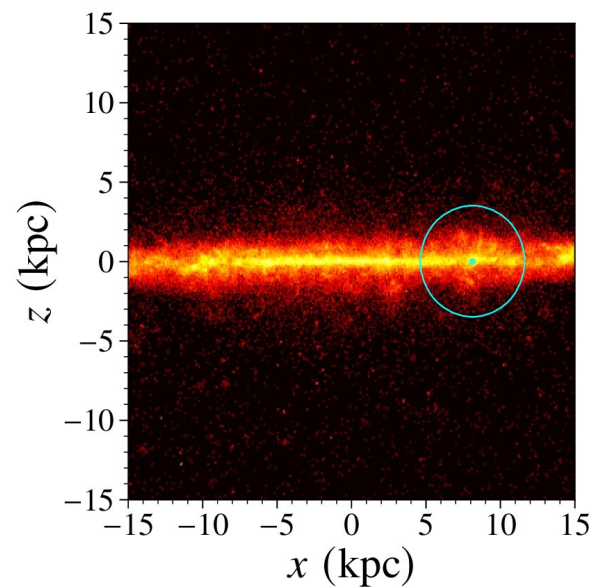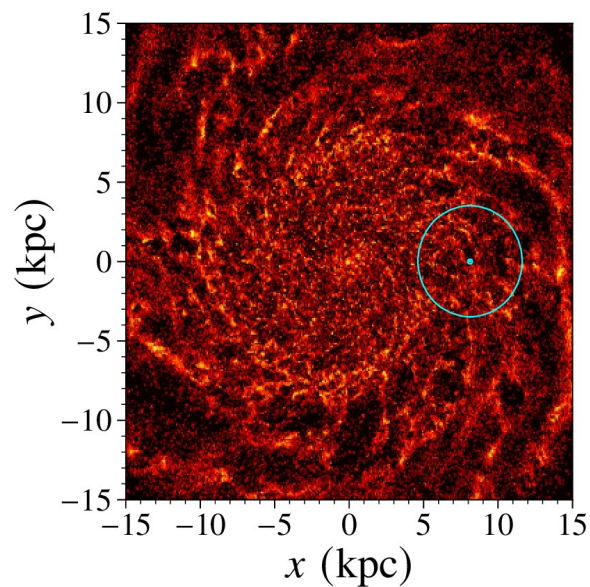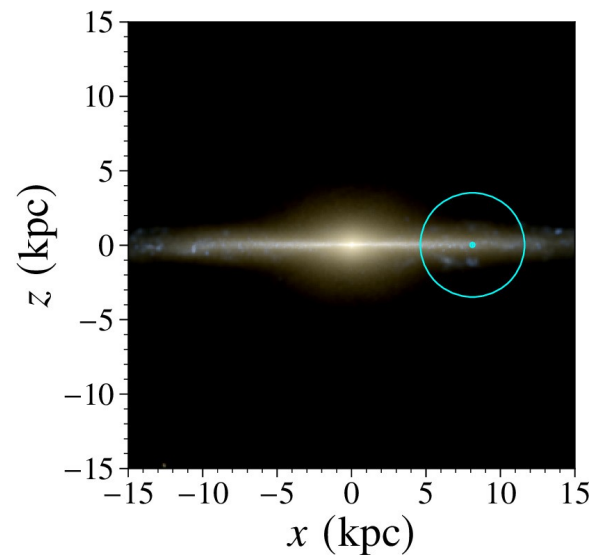
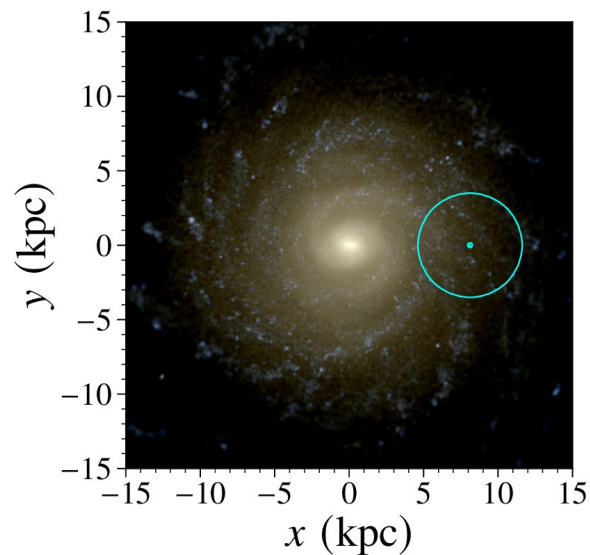* result of a continuous normalizing flow learning infinitesimal transformations

Training Dataset: Introducing h277
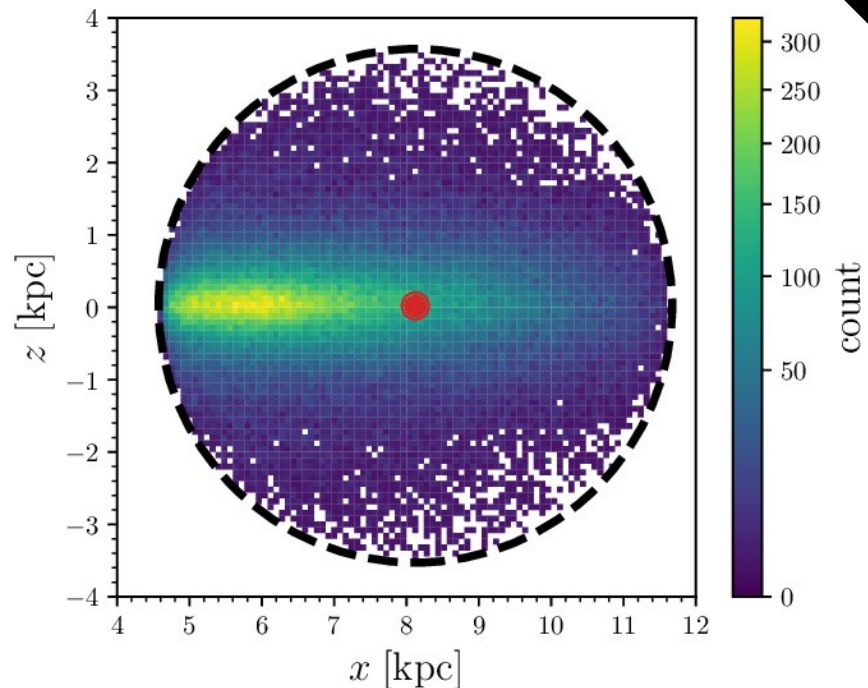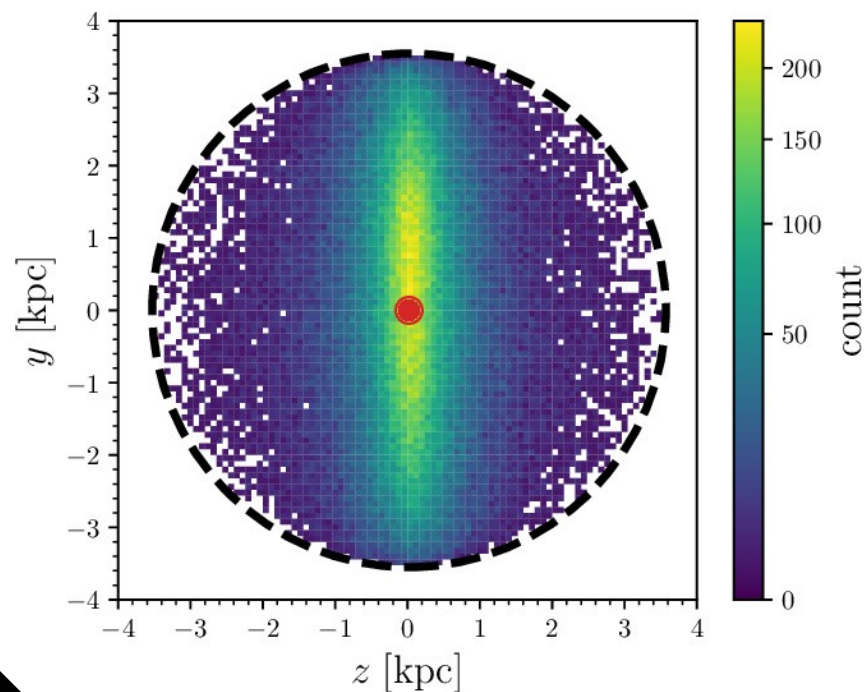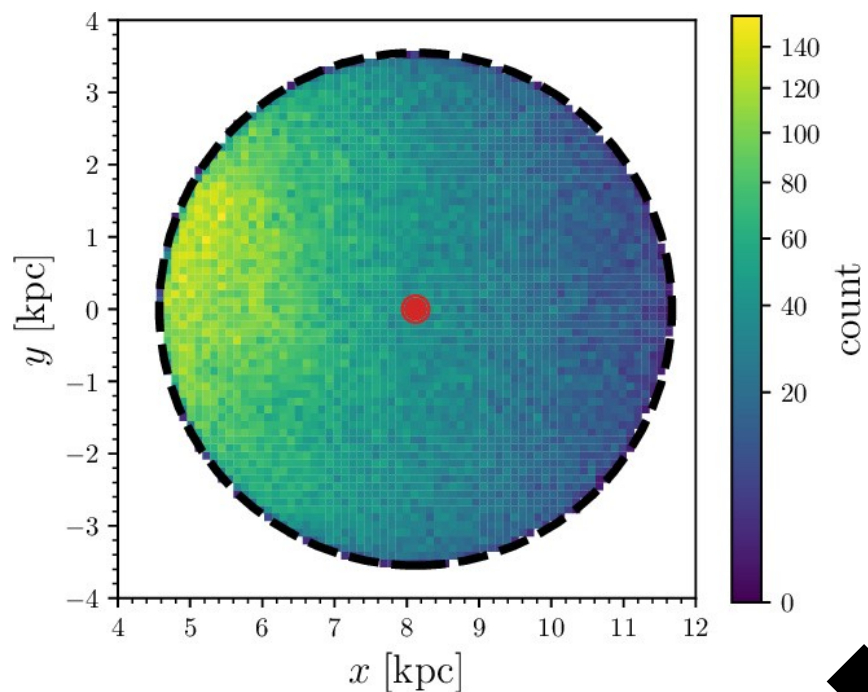
Training Dataset:
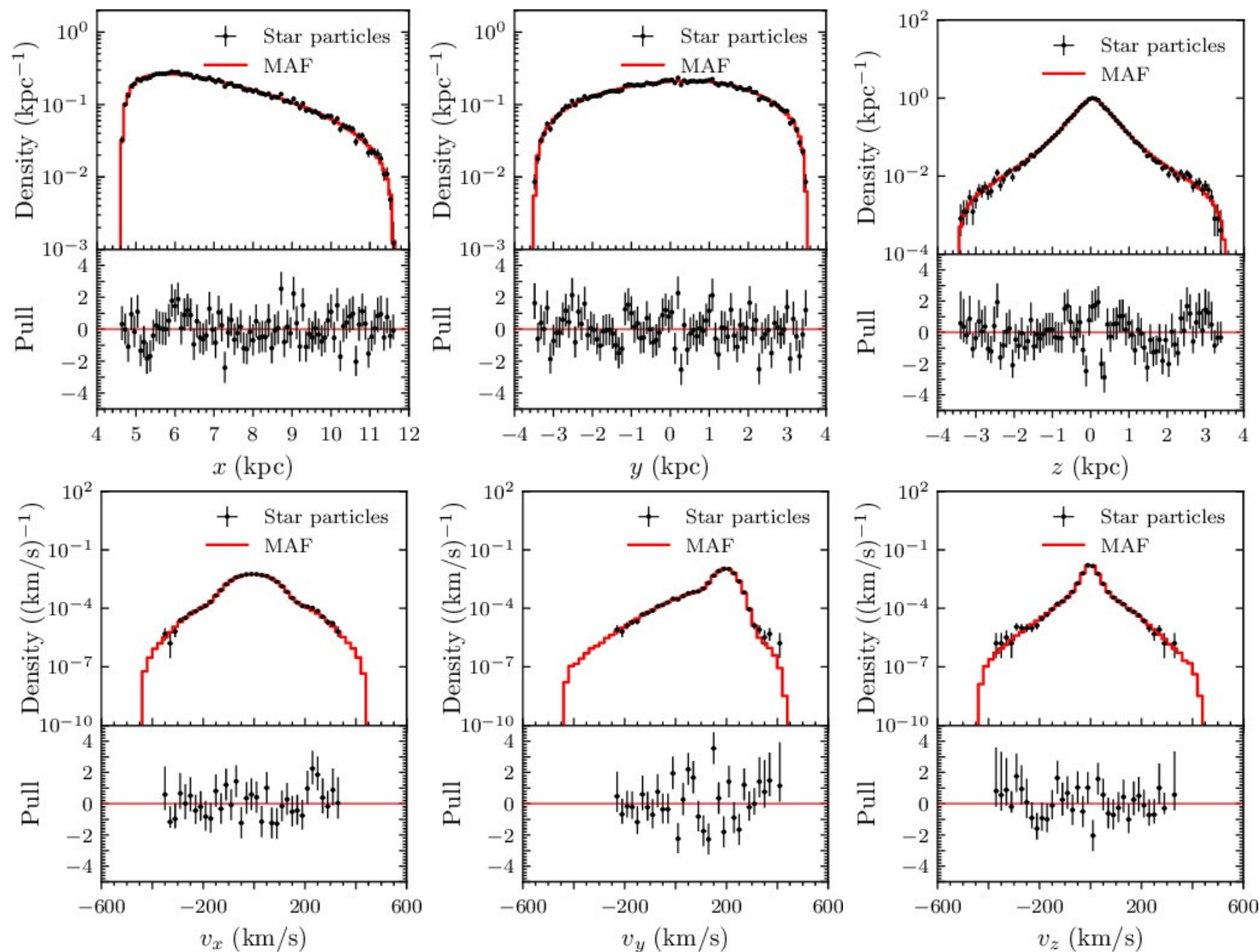# h277 at present

Machine-learned phase-space density

$$f(\vec{x}, \vec{v})$$

Neural Networks for Density Estimation:
Normalizing Flows

$$\vec{u}_0 \to \vec{u}_1 \to \cdots \to \vec{u}_n = (\vec{x}, \vec{v})$$

Results:
Phase-space density estimation

Acceleration Estimation: solving EOM by minimizing mean squared error

Now we have the estimated phase-space density estimation on our hand. Let's try to solve the Boltzmann equation.

$$\left[\vec{v} \cdot \frac{\partial}{\partial \vec{x}} + \vec{a} \cdot \frac{\partial}{\partial \vec{v}} \right] f(\vec{x}, \vec{v}) = 0, \quad \vec{a} = -\frac{d\Phi(\vec{x})}{d\vec{x}}$$

- Underdetermined in a point of view at each star position.
- Overdetermined in a point of view of phase-space density.

Given the fact that we could resample velocities at given position multiplie times, we can solve the overdetermined system using least square minimization.

$$\mathcal{L}(\vec{x}) = \frac{1}{N} \sum_{\alpha=1}^{N} \left| \left[\vec{v}^{\alpha} \cdot \frac{\partial}{\partial \vec{x}} + \vec{a} \cdot \frac{\partial}{\partial \vec{v}} \right] f(\vec{x}, \vec{v}^{\alpha}) \right|^2$$

We draw 10,000 samples per position to reduce the QMC integration error below the statistical and measurement errors.
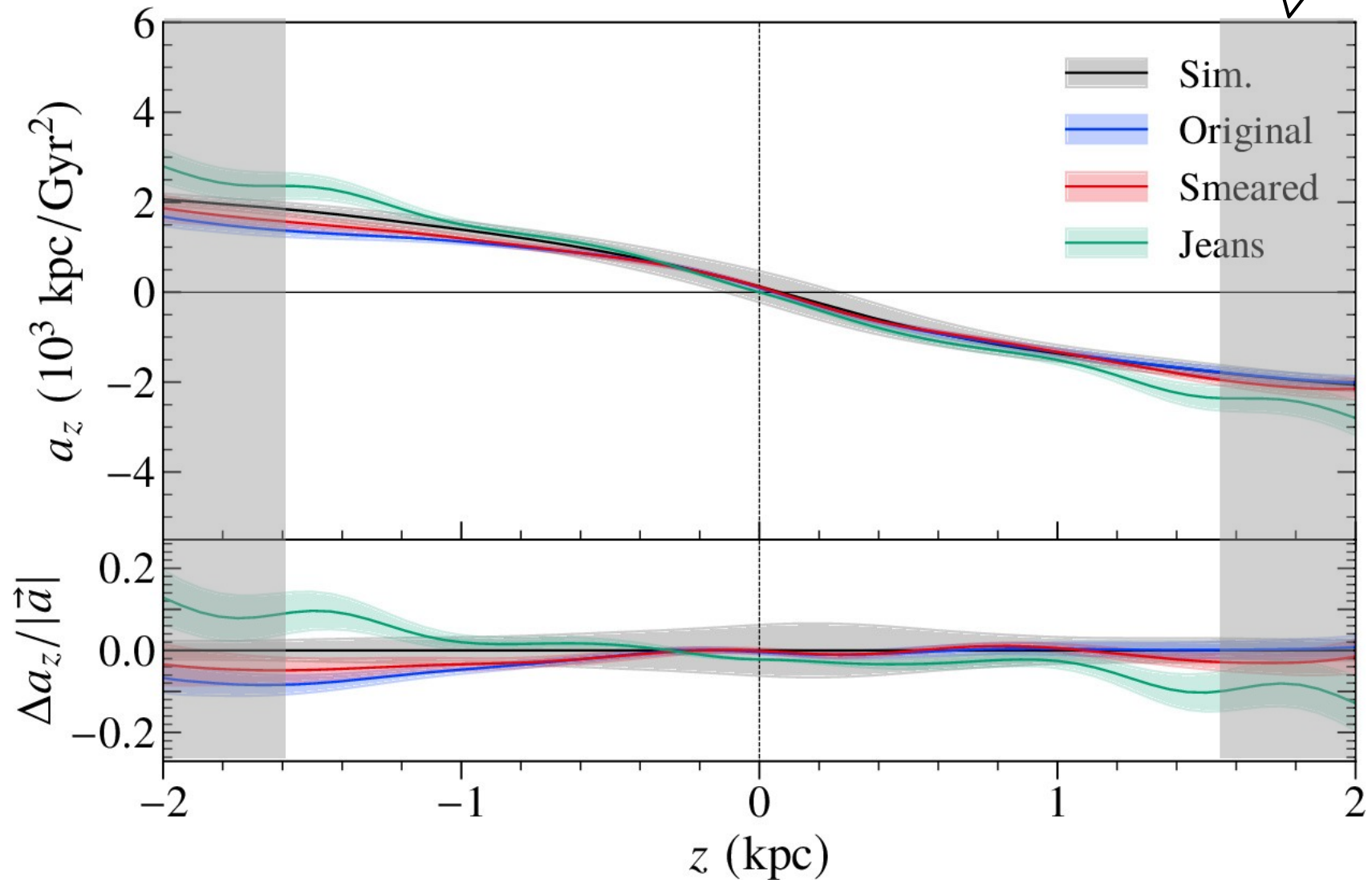
Acceleration along x-axis
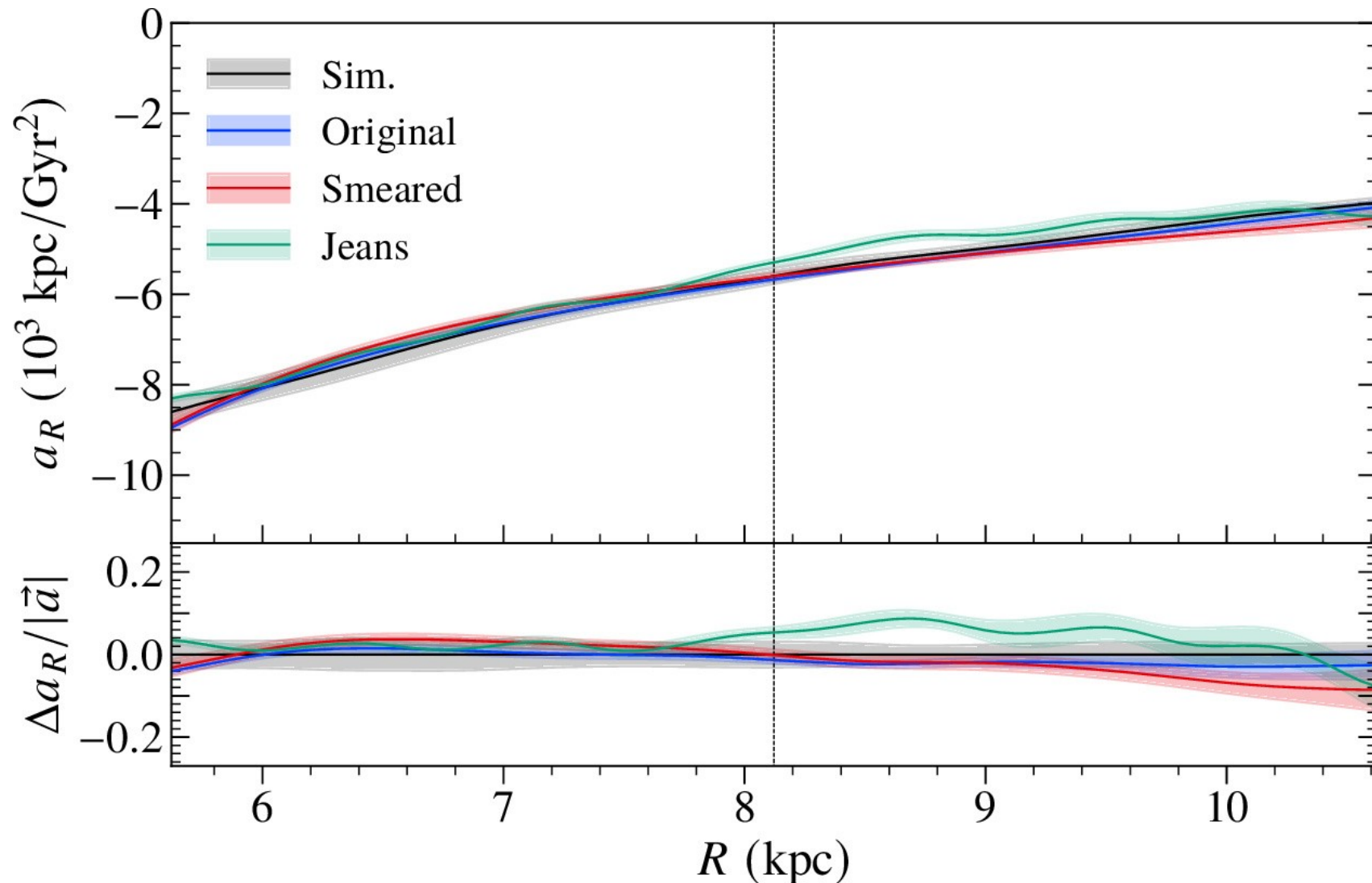
Our method can find out the acceleration within 5% accuracy!

Acceleration along z-axis

Our method can find out the acceleration within 5% accuracy!

Mass Density Estimation: Solving Gauss's Equation

Mass density can be estimated by solving smoothed Gauss's Equation:

$$-4\pi G \rho * K_h = (\nabla \cdot \vec{a}) * K_h$$

(Smoothed)
mass density function:
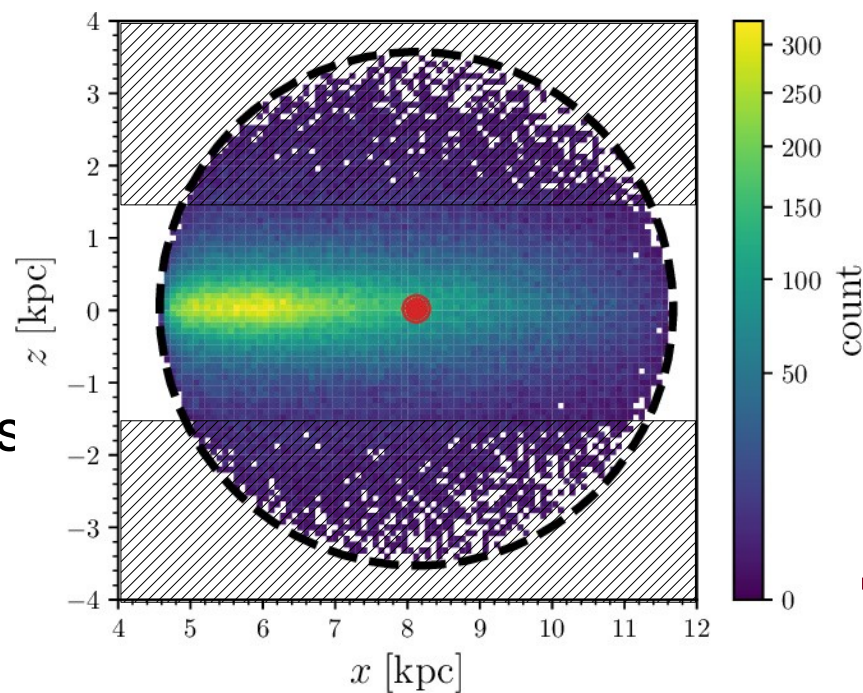at kernel bandwidth scale

Compatible!

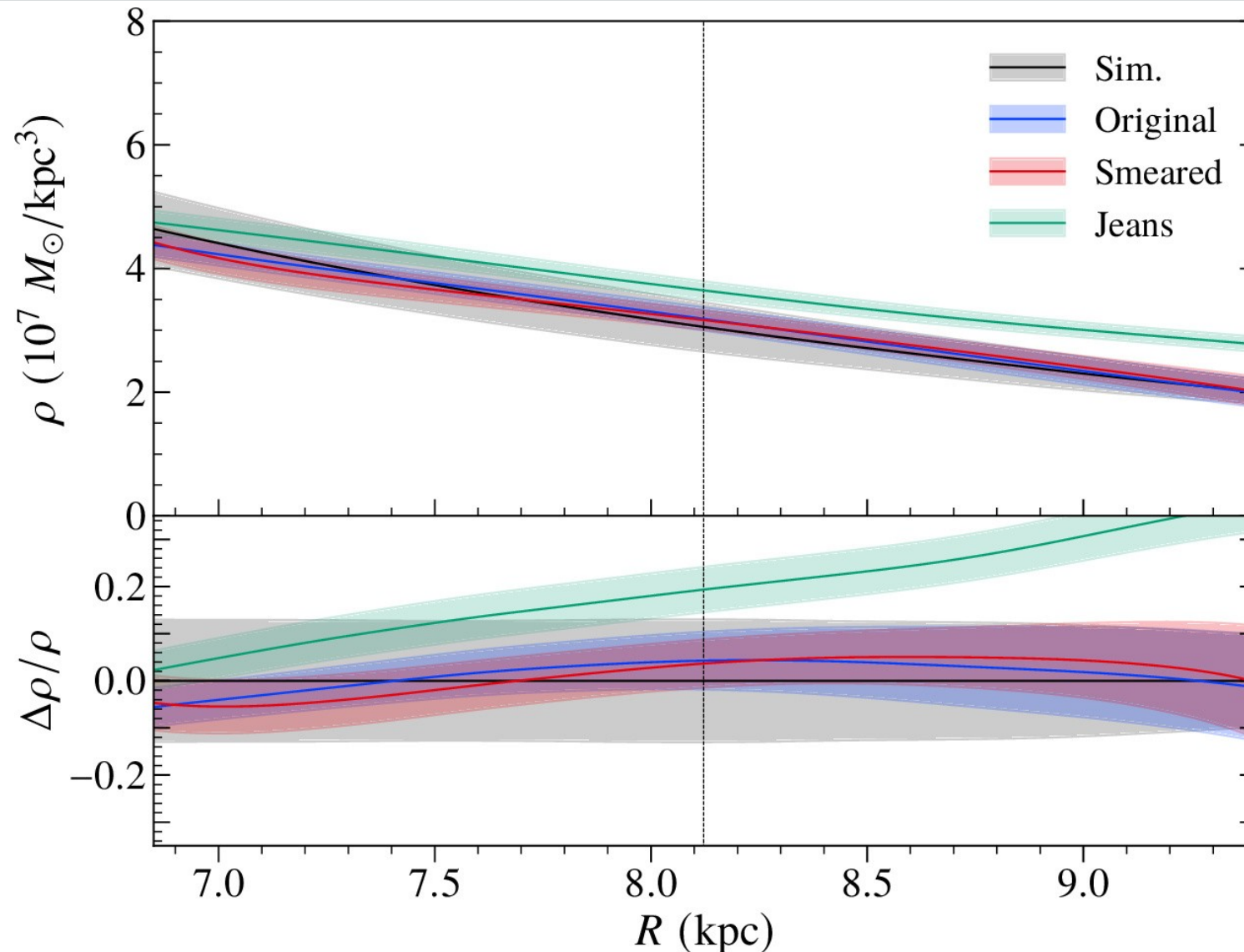(Smoothed)
estimated accelerations:
at kernel bandwidth scale

For kernels, bandwidths larger than the simulation resolution (0.173 kpc) is ideal for our purpose.

But not much stars are available at high-|z|, so we use the following bandwidths
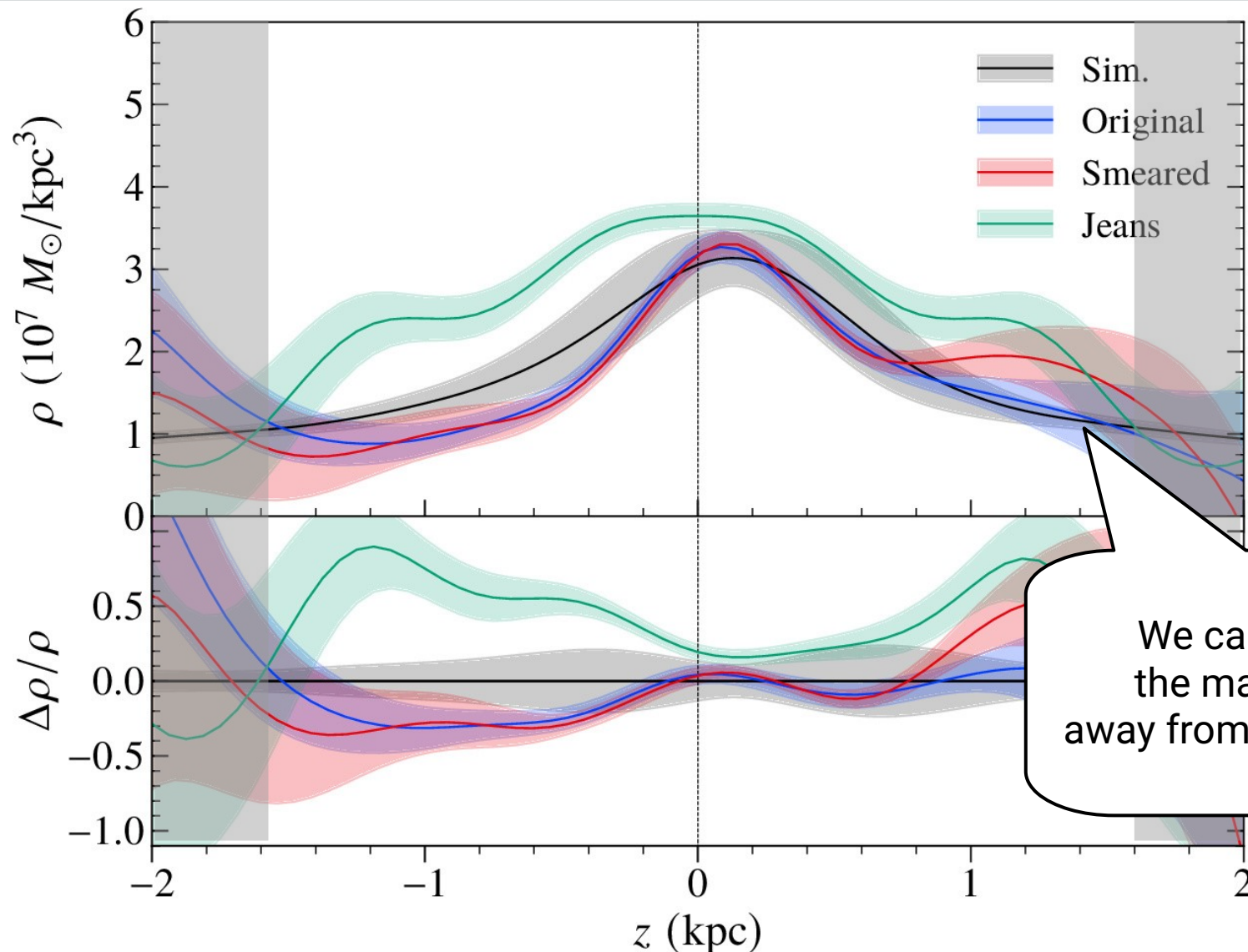
$(h_x, h_y, h_z) = (1.0, 1.0, 0.2)$ kpc

Mass Density along x-axis

Our method can find out the mass density within 10~20% accuracy!

Mass Density along z-axis

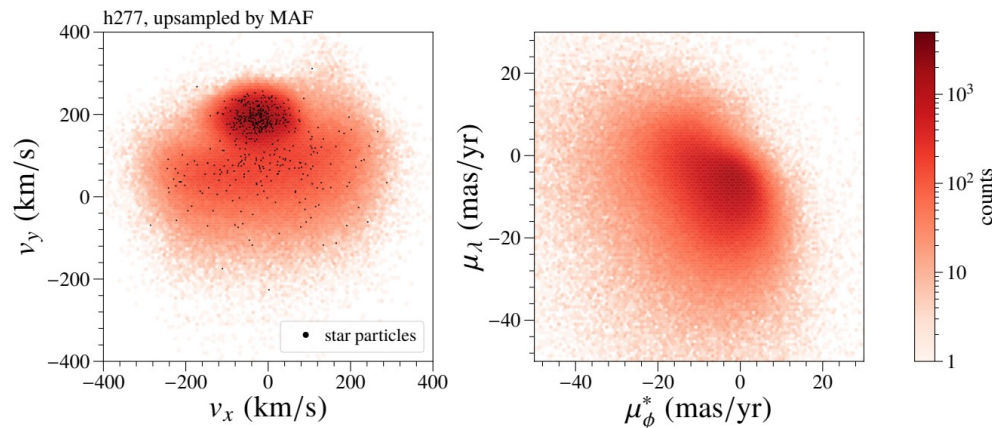Our method can find out the mass density within 10~20% accuracy!

Other ongoing projects!

Unsupervised DM density estimation of Milky Way using Gaia DR3 dataset!
with M. Buckley, E.Putney, D. Shih (Rutgers)
(to be appeared on arXiv soon!)

Unsupervised DM density estimation of simulated dwarf spheroidal galaxies
with K. Hayashi (Ichinoseki U.), M. N. Nojiri (KEK)
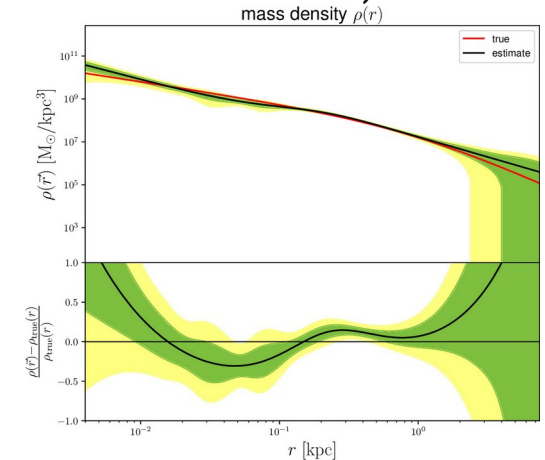(to be appeared on arXiv soon!)

See **Eric Putney**'s talk
DM II session, 2:15 PM,
tomorrow!

Upsampling hydrodynamic simulation of a galaxy
with K. Raman (Bekerly Lab), M. Buckley, D. Shih (Rutgers)
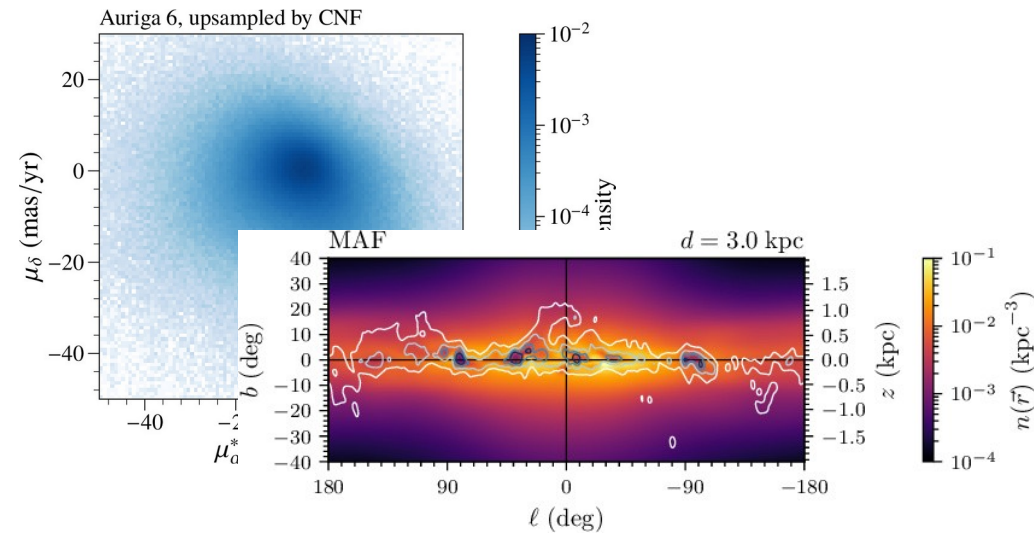(arXiv: 2211.11765)

Others...
- improving density estimation performance
- measurement bias corrections
- playing with interstellar dust

Summary of Strategy

Mock data
$\{(\vec{x}, \vec{v})\}$

Grab stars near the Sun in
an N-body simulated galaxy

Phase space density
$f(\vec{x}, \vec{v})$

Neural Networks for Density Estimation:
Normalizing Flows
$$\vec{u}_0 \rightarrow \vec{u}_1 \rightarrow \cdots \rightarrow \vec{u}_n = (\vec{x}, \vec{v})$$

Gravitational accel.
$\vec{a}(\vec{x})$

Solving Boltzmann Equation
$$\left[\vec{v} \cdot \frac{\partial}{\partial \vec{x}} + \vec{a} \cdot \frac{\partial}{\partial \vec{v}} \right] f(\vec{x}, \vec{v}) = 0$$

Mass density
$\rho(\vec{x})$

Solving Gauss's Equation
$$-4\pi G \rho = \nabla \cdot \vec{a}$$