Non-Equilibrium Modeling Approaches for Velocity Distribution Functions with Machine Learning Methods

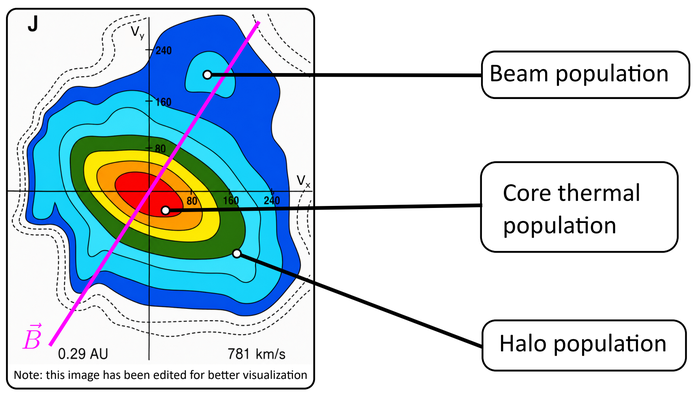
P. T. Ho[1,2], A. R. Azari[1,2,3], J. Newby[1], N. Hegde[1,2,3]

1. University of Alberta, Edmonton, Alberta, Canada
2. Alberta Machine Intelligence Institute, Edmonton, Alberta, Canada
3. Canada CIFAR AI Chair

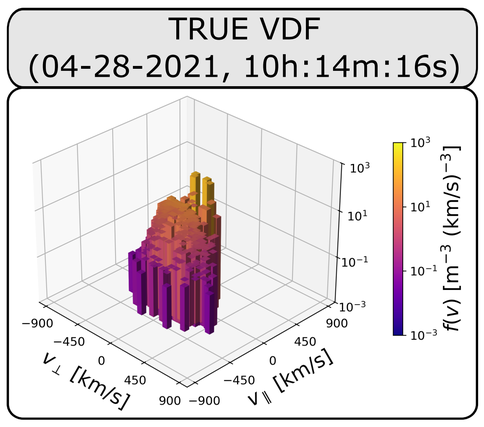
Background & Motivation

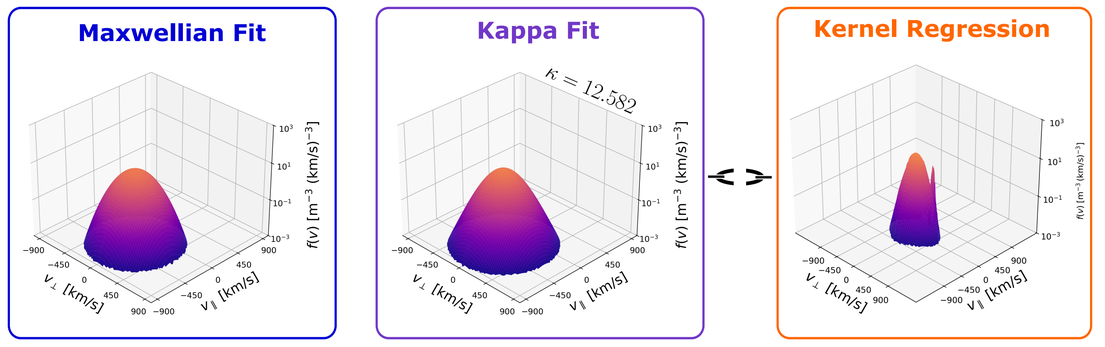
1. The velocity distribution function (VDF) describes the velocity-space structure of solar wind ions, from which macroscopic plasma properties such as density, bulk velocity, and temperature are derived [1].
2. Beyond these bulk moments, non-equilibrium features within ion VDFs serve as critical reservoirs of free energy that drive wave-particle interactions and kinetic microinstabilities associated with plasma heating and solar wind acceleration [1], [2].
3. VDF modeling methods has progressed from Maxwellian and Kappa fits [3] to Slepian basis reconstructions [4]; however, these approaches still impose assumptions on the underlying distribution shape, limiting their ability to capture non-thermal features and asymmetries in the solar wind [5].
4. More recently, nonparametric methods such as Gaussian Processes have been proposed to estimate VDFs directly from data without restrictive distributional assumptions. Here, we establish a Nadaraya–Watson kernel regression baseline for comparison against such methods [6], [7].
5. Ion VDF measurements acquired by the SPAN-I instrument aboard the Parker Solar Probe (PSP) are employed throughout this work as the observational dataset for evaluating nonparametric reconstruction methodologies in the solar wind [8].



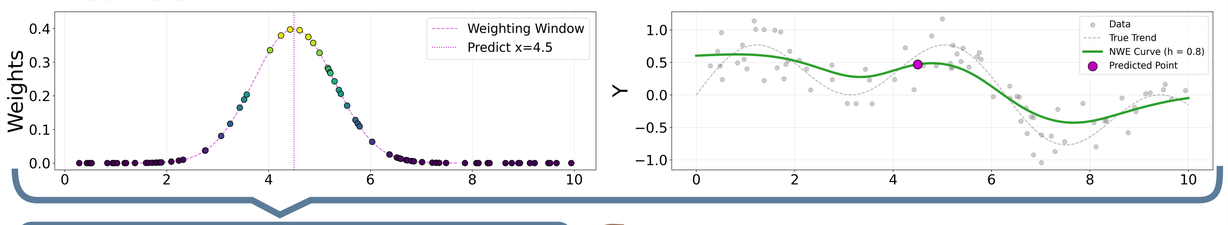


E. Marsch, 2006





Method



Nadaraya–Watson Estimator (NWE)

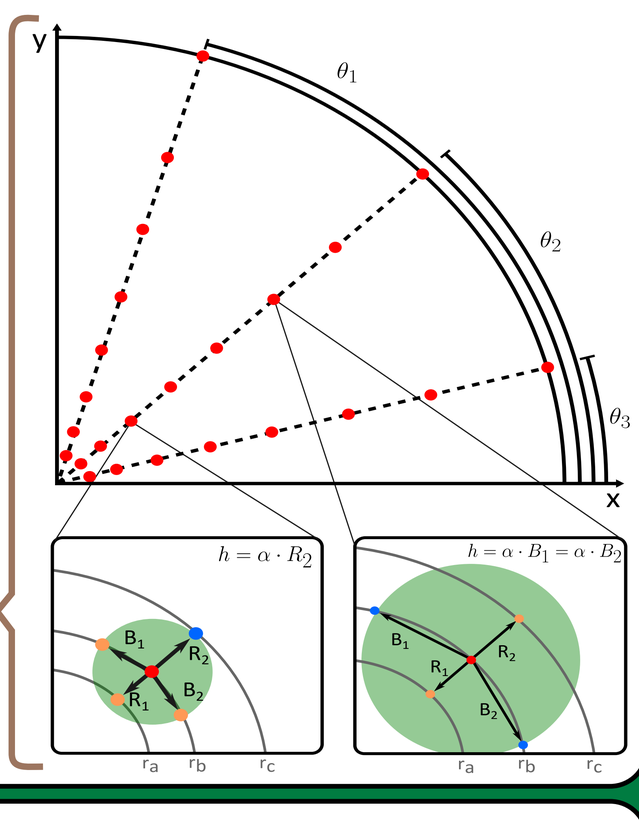
$$\hat{f}(\mathbf{x}) = \frac{\sum_{i=1}^{n} \frac{1}{h_i^{d/2}} \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}_i\|^2}{2h_i}\right) y_i}{\sum_{j=1}^{n} \frac{1}{h_j^{d/2}} \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}_j\|^2}{2h_j}\right)}$$

Function estimator via locally weighted average, with weights determined by a kernel function.

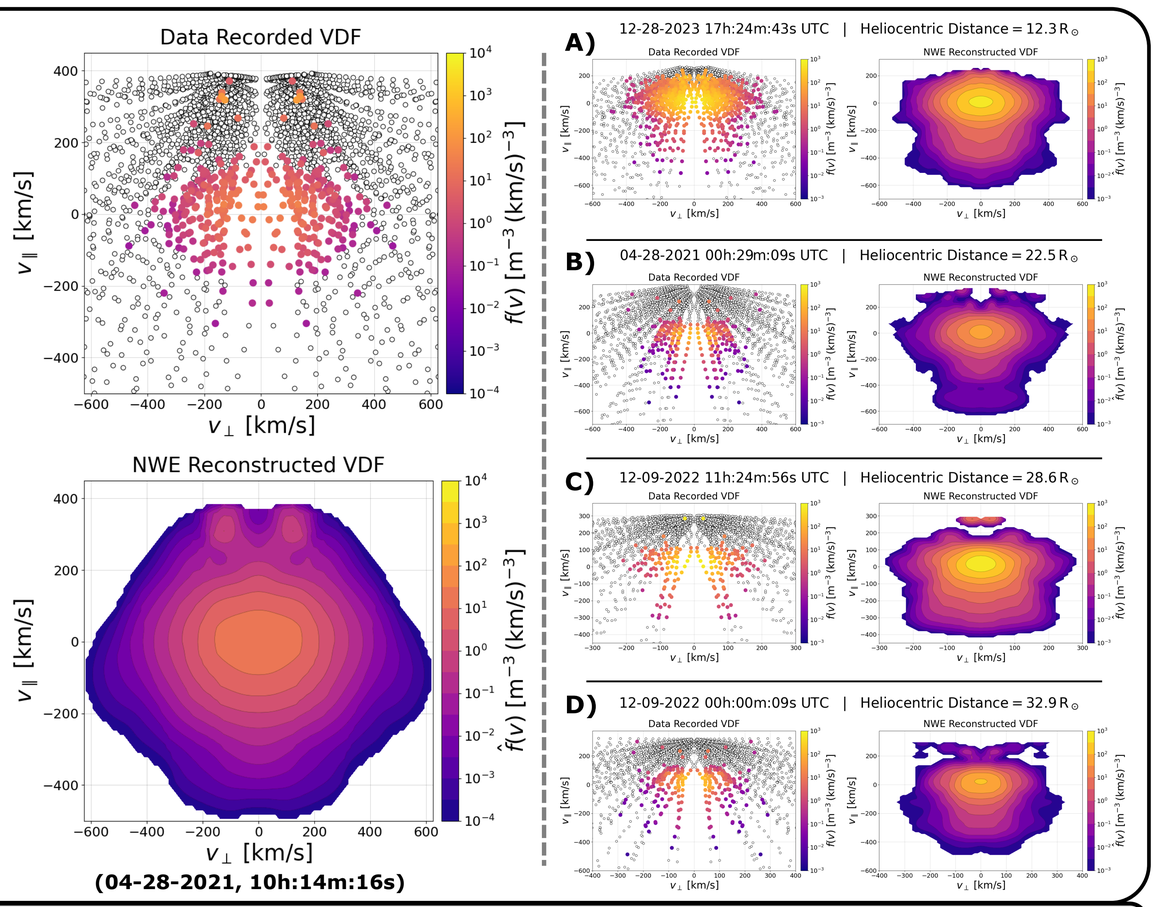
Geometric Bandwidth, h_i

$$h_i = \alpha \cdot D_i$$

Bandwidth (h_i) controls the fit's smoothness by defining kernel width for each data point. We propose a variable bandwidth to the constraint of the instrument and scaled by a constant determine through cross-validation.



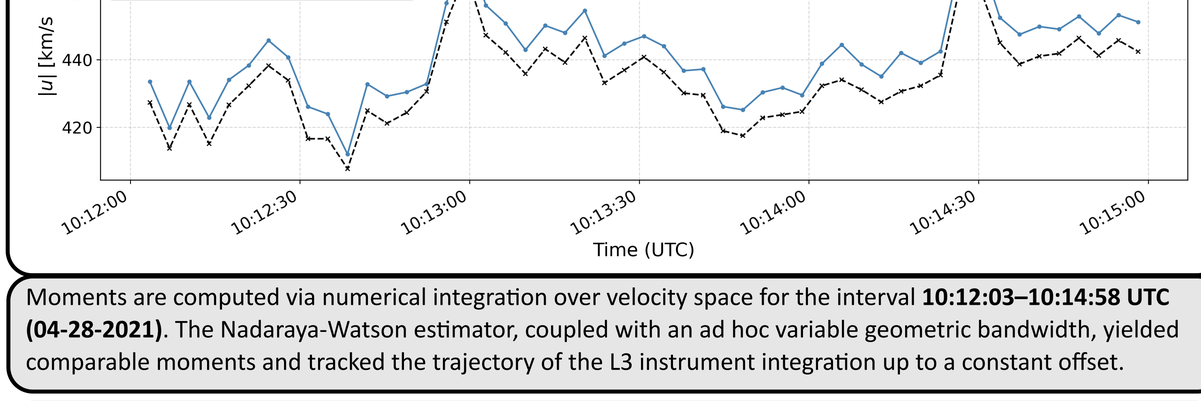
NWE VDF Reconstructions



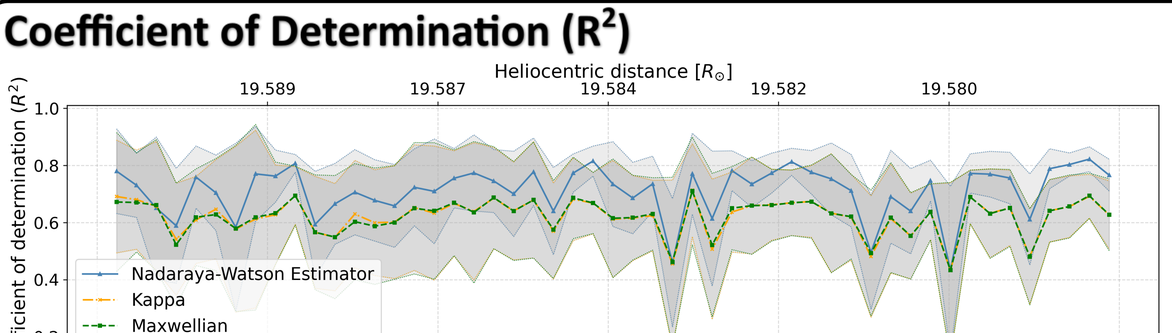
Figures shown were generated under the assumption of gyrotropy to enable two dimensional visualization of the originally three dimensional distribution.

Results

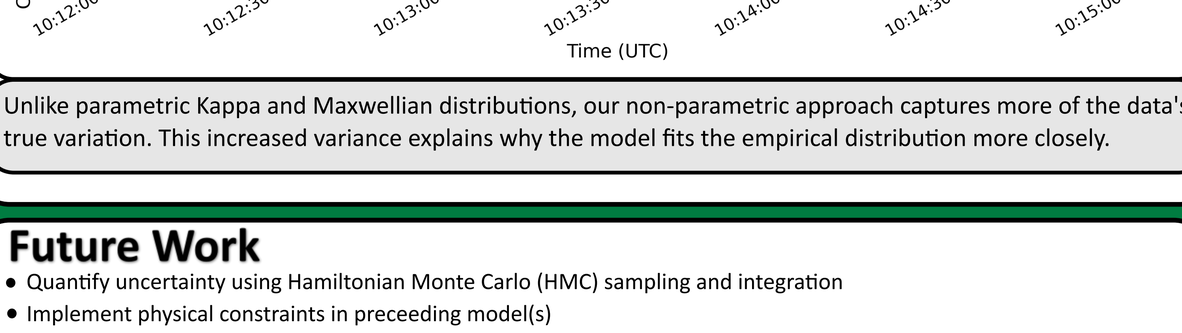
0th Moment - Density [cm^{-3}]



1st Moment - Bulk Flow Components [km/s]



1st Moment - Bulk Flow Norm [km/s]



Moments are computed via numerical integration over velocity space for the interval **10:12:03–10:14:58 UTC (04-28-2021)**. The Nadaraya-Watson estimator, coupled with an ad hoc variable geometric bandwidth, yielded comparable moments and tracked the trajectory of the L3 instrument integration up to a constant offset.

Coefficient of Determination (R^2)

Unlike parametric Kappa and Maxwellian distributions, our non-parametric approach captures more of the data's true variation. This increased variance explains why the model fits the empirical distribution more closely.

References

[1] D. Verscharen, K. G. Klein, B. A. Maruca, Living Rev. Sol. Phys., 16, 5 (2019).
[2] K. G. Klein et al 2021 ApJ 909 7
[3] G. Livadiotis, J. Geophys. Res. Space Phys., 120, 880–903 (2015).
[4] Srijan Bharati Das and Michael Terres 2025 ApJ 982 96
[5] E. Marsch, Living Rev. Sol. Phys., 2006.
[6] H. J. Bierens, Topics in Advanced Econometrics, Cambridge Univ. Press (1994).
[7] A. R. Azari, Non-Parametric Estimation of Solar Wind Proton Velocity Distribution Functions from Parker Solar Probe using Gaussian Processes, in prep.
[8] R. Livi et al., Astrophys. J., 938, 138 (2022).

Future Work

- Quantify uncertainty using Hamiltonian Monte Carlo (HMC) sampling and integration
- Implement physical constraints in preceeding model(s)
- Evaluate model performance through error metrics and distributional diagnostics
- Investigate the moment ratio between data limits and the entire space