Neural Data Compression of Plasma Distributions for Space Weather

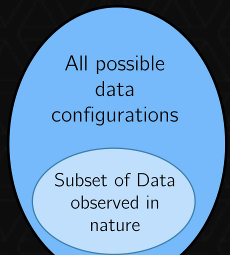
Daniel da Silva[123], Christopher Bard[1]

1 – NASA Goddard Space Flight Center; 2 – Laboratory for Atmospheric and Space Physics, University of Colorado Boulder; 3 – University of Maryland, Baltimore County

Introduction

Physics has long history of theory-driven dimensionality reduction (Fourier Transform, Spherical Harmonics)

We present a technique for data-driven dimensionality reduction on ion counts distributions (detections as a function of velocity).

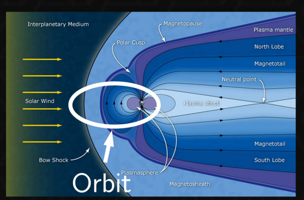


Relevance to:
- Feeding other ML algorithms
- Data compression for instruments

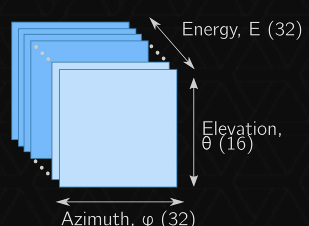
The algorithm presented here is lossy, and we present a novel validation technique:
- Calculate and compare macroscopic fluid variables between original/reconstructed data
- **Argument:** Preserving fluid moments would in turn preserve fluid-level physics, and in turn a degree of scientific validity

Data

Direct Measurements of Space Plasmas near Earth
- Magnetospheric Multiscale Mission (MMS) from NASA
- Studies plasmas within the context of Earth's magnetic field and solar wind
- High-Earth Orbit (up to 11 Earth Radii away in data used)



Instrument Measures Ion Particles
- **Ion Count Detection Rate** $C(\phi,\theta,E)$
- ϕ – *Azimuth*
- θ – *Elevation*
- E – *Eenergy*
- 3D array: shape = (32, 16, 32)

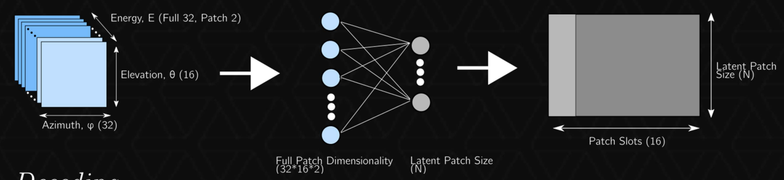


Training Set
- 90/10 Test/Train Split
- Data from Phase 4B of mission: 11/29/2018 – 4/13/2019

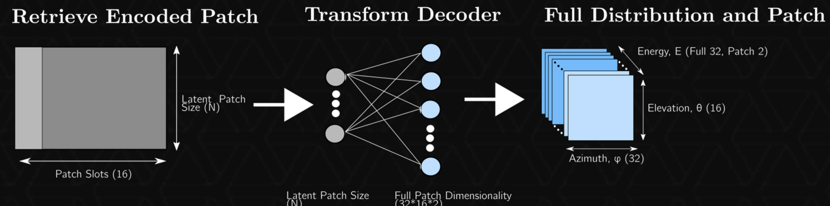
Data Available under Creative Commons at:
https://lasp.colorado.edu/mms/sdc/public/

AutoEncoder Network Architecture

Encoding



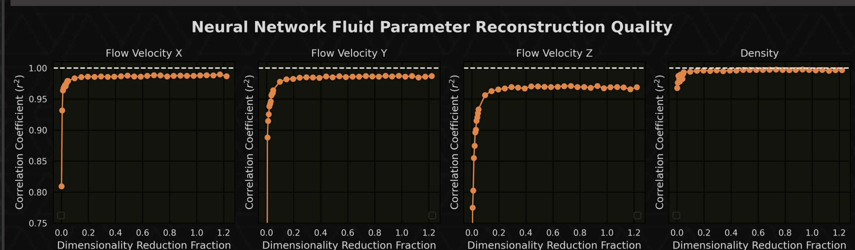
Decoding



- Separate MLPs encode each patches of two energies (da Silva et al., 2020)
- Loss: RMS in Counts between Original/Reconstructed
- Patches:
  - Covers all azimuths/elevations and two energies
  - Mean of Energy Shell forced equal after decoding

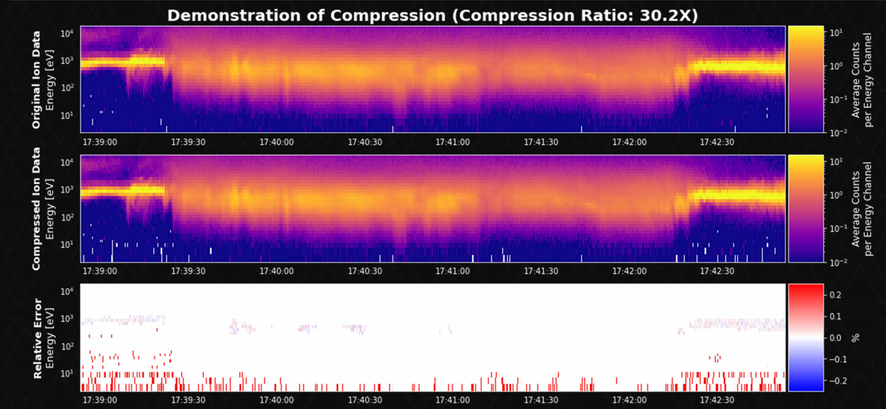
$$\vec{C}^{E}_{recon,adjusted} = \vec{C}^{E}_{recon}\left(\frac{\overline{C}^{E}_{orig}}{\overline{C}^{E}_{recon}}\right)$$

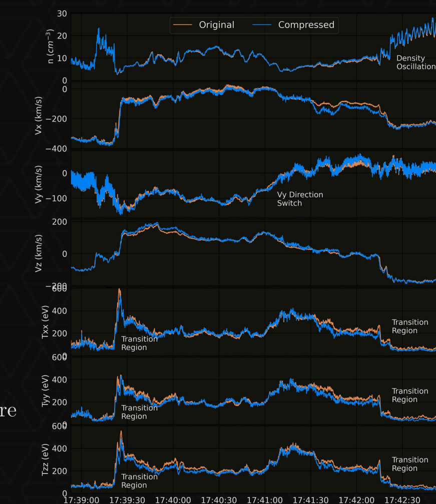
Validation and Fluid Variable Information Bottleneck



- Correlation coefficient r^2 between the original and reconstructed plasma data as a function of the dimensionality reduction fraction (DRF)
- DRF: defined as the size of the latent representation relative to the original dimensionality (1.0 → no dimensional reduction)
- We observe that an information bottleneck, which prevents accurate reconstructions, occurs for dimensional reductions between 10-20x (DRF ≈ 0.05-0.10).

Applications to Data Compression





Potential use case for reducing satellite transmission budget for observed data, in turn (i) reducing spaceflight costs and (ii) relieving over-congested telemetry networks.

Compression Specifications
- Auto-encoder is main transform method
- Quantizer is trimming 6 bits off fractional part of IEEE 16-bit floating point
- Entropy Coding is GZIP/DEFLATE
- Packaged in custom data format

Qualitative Overview
- Observe figure shows strong ability to capture
  - Energy Spread (bimodal, skewed)
  - Transition Regions (cold to hot plasma)
  - Plasma Waves

Acknowledgements

We thank the NASA/GSFC Center for HelioAnalytics for seed funding and community support, as well as the AWS Heliocloud resources provided by the NASA/GSFC Heliophysics Sciences Division.

Source Code

Source code available at: *https://github.com/ddasilva/mms-compression-neurips-2022*