Adapting **large lagrangian datasets** into **zarr archives**, improves data engineering and analysis efficiency, **encouraging** analysis experimentation!

Clouddrift, improving analysis and manipulation of large lagrangian datasets

Kevin Santana

1 Intro

- Lagrangian data typically refers to oceanic and atmosphere information acquired by observing platforms drifting with the flow they are embedded within.

2 Datasets available

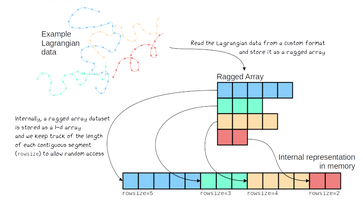
- MoSAiC, sea ice trajectories
- GDP, ocean drifter trajectories
- HURDAT2, cyclone trajectories

3 Scope And Key Features

- Working with contiguous ragged array representations of data
- Delivering functions and methods to perform scientific analysis of Lagrangian data
- Processing publicly available Lagrangian datasets into the common ragged array data structure
- Making cloud-optimized ragged array datasets easily accessible

Extra figures

Ragged Array Data Structure



- Clouddrift organizes data variables within a dataset in the Ragged Array data structure
- A Ragged Array is organized where each particle/trajectory/row associated observations are concatenated resulting in a 1-d array in order.
- A resulting metadata variable stores the rowsize of each individual particle/trajectory/row.

UNIVERSITY OF MIAMI

Download the paper →

Scan me