EGU22-10799, updated on 11 Aug 2022
https://doi.org/10.5194/egusphere-egu22-10799
EGU General Assembly 2022

Design Considerations for the 3rd Spatial Dimension of the Spatiotemporal Adaptive Resolution Encoding (STARE)

Michael Rilee[1] and **Kwo-Sen Kuo**[2]
[1]Rilee Systems Technologies LLC, Derwood, United States of America (mike@rilee.net)
[2]Bayesics LLC, Greenbelt, United States of America (kuo@bayesics.com)

The real world does not live on a regular grid. The observations with the best spatiotemporal resolution are generally irregularly distributed over space and time, even though as data they are generally stored in arrays in files. Storing the diverse data types of Earth science, including grid, swath, and point based spatiotemporal distributions, in separate files leads to computer-native array layouts on disk or working memory having little or no connection with the spatiotemporal layout of the observations themselves. For integrative analysis, data must be co-aligned both spatiotemporally and in computer memory, a process called data harmonization. For data harmonization to be scalable in both diversity and volume, data movement must be minimized. The SpatioTemporal Adaptive Resolution Encoding (STARE) is a hierarchical, recursively subdivided indexing scheme for harmonizing diverse data at scale.

STARE indices are integers embedded with spatiotemporal attributes key to efficient spatiotemporal analysis. As a more computationally efficient alternative to conventional floating-point spatiotemporal references, STARE indices apply uniformly to all spatiotemporal data regardless of their geometric layouts. Through this unified reference, STARE harmonizes diverse data in their native states to enable integrative analysis without requiring homogenization of the data by interpolating them to a common grid first.

The current implementation of STARE supports solid angle indexing, i.e. longitude-latitude, and time. To fully support Earth science applications, STARE must be extended to indexing the radial dimension for a full 4D spatiotemporal indexing. As STARE's scalability is based on having a universal encoding scheme mapping spatiotemporal volumes to integers, the variety of existing approaches to encoding the radial dimension arising in Earth science raises complex design issues for applying STARE's principles. For example, the radial dimension can be usefully expressed via length (altitude) or pressure coordinates. Both length and pressure raise the question as to what reference surface should be used. As STARE's goal is to harmonize different kinds of data, we must determine whether it is better to have separate radial scale encodings for length and pressure, or should we have a single radial encoding, for which we provide tools for translating between various (radial) coordinate systems. The questions become more complex when we consider the wide range of Earth science data and applications, including, for example, model simulation output, lidar point clouds, spacecraft swath data, aircraft in-situ measurements, vertical or oblique

parameter retrievals, and earthquake-induced movement detection.

In this work, we will review STARE's unifying principle and the unique nature of the radial dimension. We will discuss the challenges of enabling scalable Earth science data harmonization in both diversity and volume, particularly in the context of detection, cataloging, and statistical study of fully 4D hierarchical phenomena events such as extratropical cyclones. With the twin challenges of exascale computing and increasing model simulation resolutions opening new views into physical processes, scalable methods for bringing best-resolution observations and simulations together, like STARE, are becoming increasingly important.