The Impact on Quality and Uncertainty of Regridding Diverse Earth Science Data for Integrative Analysis

Michael Rilee[1] and Kwo-Sen Kuo[2]

[1]Rilee Systems Technologies LLC
[2]Bayesics LLC

November 23, 2022

Abstract

Understanding and communicating the impact of uncertainty on scientific understanding is a critical unmet need in Earth Science. Challenging as uncertainty determination is for "low level" data, the impact of further processing must be understood, particularly when integrating diverse data types. For example, a fundamental source of diversity is data's spatiotemporal distribution, for which Point, Grid, and Swath are the most important overarching geographical types. To bring different kinds of data together or bring observational data and model simulations together, observation and simulation values must be "regridded" onto the same grid, i.e. onto a comparable spatiotemporal representation. At the finest level, this requires a detailed understanding of instrumental fields of view and sensitivities. But regridding itself affects uncertainty, especially in situations where significant or irregular interpolation or even extrapolation are required. Furthermore, care must be taken combining diverse data lest the integrated product inherit the worst uncertainty characteristics of each. We have been developing capacities for indexing, regridding, and integrating observation and model simulation that scale to the size and diversity of Earth Science data. In this presentation, we review the fundamental problems associated with combining big, diverse Earth Science data for integrative analysis and how to quantitatively assess and propagate the uncertainties introduced. In particular, uncertainties associated with regridding diverse data for various popular grids and regridding schemes will be assessed.

The Impact on Quality and Uncertainty of Regridding Diverse Earth Science Data for Integrative Analysis

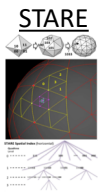
Michael Lee Rilee[1,3] and Kwo-Sen Kuo[2,3] [1]Rilee Systems Technologies LLC; [2]Bayesics LLC; [3]NASA Goddard Space Flight Center

Abstract

- Understanding the impact of uncertainty is critical for Earth Science.
- While determining low-level uncertainty is challenging, its propagation through data processing steps compounds the challenge.
- Integrative analysis requires regridding diverse data types (Point, Grid, Swath, observations and models/simulation) onto the same grid.
- Regridding affects uncertainty propagation, which must be characterized.
- We have started bringing the tools together empirically assess the effect of uncertainty propagation on integrated analysis of diverse Earth Science data.

Motivation

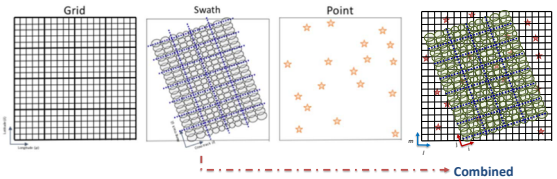
The new SpatioTemporal Adaptive Resolution Encoding (STARE) will make it easier to identify and bring together overlapping, diverse data, which must then be regridded for integrated analysis.



- How do low level uncertainties affect higher level products?
- How do uncertainties from different kinds of data combine?
- What role does processing itself play on integrated products?
- What role do different regridding schemes play?
- How does grid choice affect analysis results?
- Can we construct and propagate useful uncertainty information?
- Can techniques, like Kriging, for addressing data problems help?

Earth Science Data is Diverse – Even in logical model

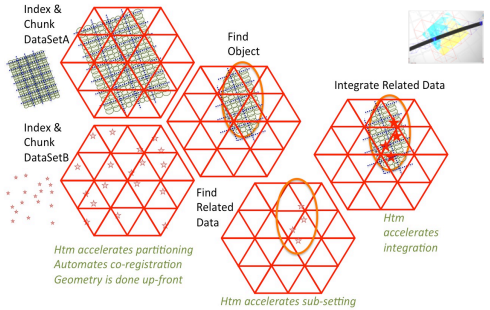
Common Earth Science Data Models



STARE Indexing, beyond the Hierarchical Triangular Mesh

Versatile, efficient, scalable, parallel – one index for diverse data

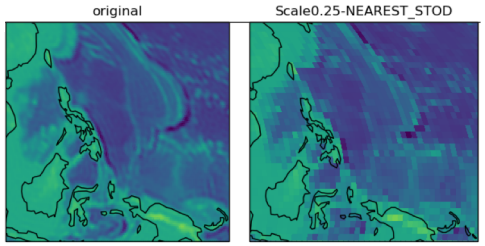
STARE indexing accelerates data registration, sub-setting, fusion

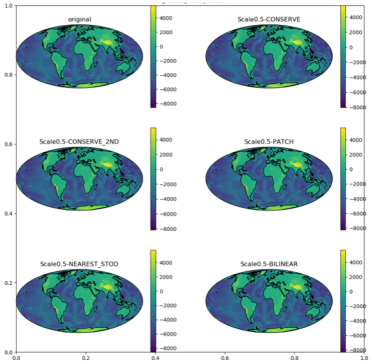


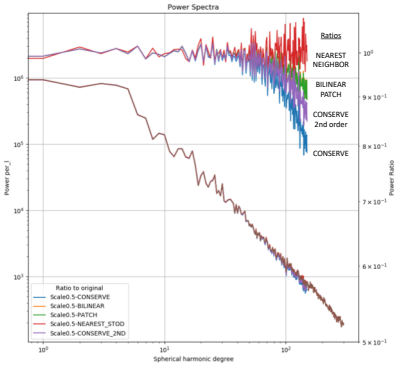
Preliminaries

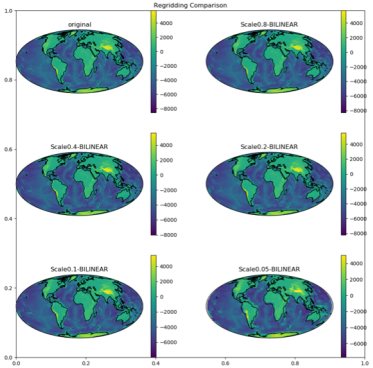
We are constructing an analysis framework for studying uncertainty and regridding.

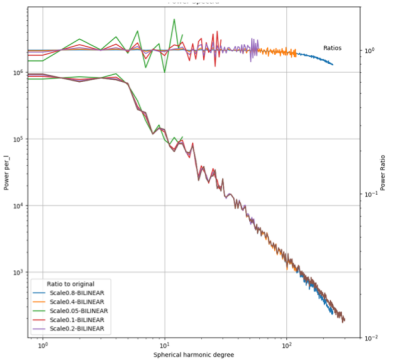
- Initial forays involve
  - ESMPy and ESMF for regridding
  - PyShtools for Spherical Harmonic Analysis











Example Data & Calculation

The examples show expected, basic features required for more analysis.

- Topographic data from PyShtools (a 1204x602 data set) was chosen for convenience, its use in the toolkit, and its power-law distribution of scales.
- Resolution scales in regridding examples reduce both dimensions by the roughly the stated factor, e.g. Scale~0.5 => 604x302.
- The regridding shown acts as a low pass filter, and regridding schemes of different orders show the expected effect on the spectral amplitudes.

Future plans

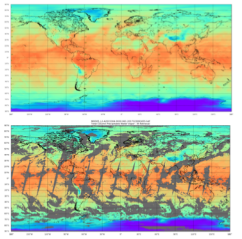
Key capabilities can be built on the existing analyses.

- Add support for experimenting with and characterizing errors and uncertainties, including spatial correlated variations.
- Ensemble calculation will help us understand how regridding propagates variability and uncertainty distributions.
- Kriging can interpolate data gaps and process lower level data, but depends on the calculation and modeling of variograms.
- Add support for more data types and the spectral analysis of more grids.
- Use STARE to integrate diverse data and calculate uncertainties.
- The NASA Open Access Geo-Gridding Infrastructure, NOGGIn, now an internal development project at NASA/LAADSWEB, provides a way to perform some of these transformations as a service "in the cloud," easing some data processing chores, including regridding, Kriging, data access, and uncertainty estimation.

An example – integrating Level 2 Data

NOGGIn developed a Kriging capability that can be used to construct custom data products from lower level data.

- As a statistical interpolation technique, Kriging can provide uncertainty estimates aiding integration and characterization.



Summary

This work in progress shows a foundation for more interesting analyses.

- The results show a basic capability to load and process data, performing a variety of regridding functions and spectral analysis.
- We can observe differences in how different regridding schemes affect a natural spatial structure through spherical harmonic analysis.
- This work provides a good foundation for further experimentation with uncertainty propagation and regridding, with natural connections to previous work in Kriging and future work in STARE-based diverse data integration.

Acknowledgement
Funding for this research is provided by NASA Advancing Collaborative Connections for Earth System Science (ACCESS) Program, Grant No. 80NSSC18M0118. Support for NOGGIn, T. Clune, PI, NASA/GSFC, was provided by NASA Earth Science Technology Office (ESTO) through the Advanced Information Systems Technology (AIST) program and continued collaboration with NASA/LAADS DAAC, for which we are very grateful. The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of NASA.