

Data Extraction

Matthew Leonawicz

December 24, 2014

1 Introduction

This project organizes various aspects of QA/QC on data extraction operations at SNAP. Currently, there are three projects under this umbrella:

- * An evaluation of data extraction routines often applied to large SNAP data sets.
- * Simulation and application of efficient data distribution estimation methods to SNAP data.
- * A priori compiling of vectorized (spatial polygons) data in the form of raster indexing for more efficient subsequent use.

1.1 Motivation

Although these projects are not directly dependent upon one another, they all relate in terms of the efforts put toward efficient data processing at SNAP. The goals are to improve efficiency while maintaining, and in some cases also improving, statistical accuracy and sensible probabilistic bounds on uncertainty.