Wavelet Enabled Progressive Data Access and Storage Protocol (WASP) Metadata Conventions

Version 1, August 12, 2015

John Clyne, NCAR
Larry Frank, UCSD
Thomas Lesperance, UCSD
Alan Norton, NCAR

Abstract

This document describes the WASP conventions for the storage of progressively accessible data in NetCDF files using wavelets. Like the widely used NetCDF Climate and Forecast (CF) Metadata Conventions, the WASP conventions define metadata attributes that facilitate the creation of self-describing NetCDF datasets. The WASP conventions support the description of NetCDF arrays transformed into wavelet space, and stored as a discrete set of wavelet coefficients. The resulting coefficients undergo an inverse transform to reconstruct the original array. The advantage of storing arrays as a wavelet expansion is that coarsened approximations of the original data may be reconstructed by using only a subset of the available wavelet coefficients, resulting in reduced I/O, or by limiting the number of passes of the multi-resolution inverse wavelet transform, resulting in a reduced resolution grid, or both.

WASP Attributes

The documentation below describes NetCDF global and variable attributes defined by the

WASP convention.

Global Attributes

WASP: An Boolean flag indicating whether the file adheres to the WASP conventions. This attribute must be present in a WASP NetCDF file. If it is not present and does not evaluate to true no other WASP attributes will be recognized.

WASP.Version: An integer indicating the version number of the WASP conventions.

WASP.NumFiles: To facilitate management of large data sets wavelet representations of arrays may be stored in multiple files: the *primary* file containing conventional NetCDF arrays and attributes, as well as the most significant wavelet coefficients for any arrays represented as wavelet expansions (see the WASP.CRatios attribute); and zero or more *secondary* files each containing less significant wavelet coefficients. Each successive secondary file contains the wavelet coefficients that, when combined with the coefficients stored in the primary and any preceding secondary files, are necessary to reconstruct the array at one of the refinement levels indicated by the **WASP.CRatios** variable attribute. Successive secondary files are named by appending an unformatted sequence number, starting with "1", to the primary file path name. Thus if the primary file is named "file.nc", successive secondary files would be named "file.nc1", "file.nc2", and so on. Valid values of this attribute range from 1 to the number of elements in the **WASP.CRatios** variable attribute. If the value is less than the number of elements in the **WASP.CRatios** variable attribute then the remaining least significant coefficients are stored in the last secondary file in the sequence.

Variable Attributes

WASP: An Boolean flag indicating whether the variable is stored as a wavelet expansion. The remaining WASP variable attributes are only honored if the value of this attribute is true.

WASP.DimNames: This attribute contains a space-delimited string with an ordered list of the NetCDF dimension names of the time/space domain dimensions of the variable. These dimensions must be defined in the primary file.

WASP.Wavelet: The name of the wavelet used to transform the variable. Recognized wavelet names are: bior1.1, bior1.3, bior1.5, bior2.2, bior2.4, bior2.6, bior2.8, bior3.1, bior3.3, bior3.5, bior3.7, bior3.9, and bior4.4.

WASP.Encoding: This attribute contains a string describing the encoding algorithm used for the storage of the wavelet coefficients. Currently recognized values are "SignficanceMap", "SPECK", and "SPIHT". The latter two are widely described in wavelet

compression literature. The first, “SignficanceMap”, indicates that a bitmap will be appended to each wavelet block. If a bit at location *n* is set than the wavelet coefficient corresponding to that location is present in the wavelet block, otherwise the coefficient is absent and its value is taken to be zero.

WASP.Decomposition: The wavelet decomposition strategy. Two string values are supported: “standard”, and “nonstandard”.

WASP.BlockSize: Prior to undergoing the forward wavelet transform arrays are decomposed into equal-sized rectangular regions (blocks). The dimension of each block, which may be equal to that of the original array, is given as an ordered list of integers by this attribute. The block dimensions need not be an integral factor of the array dimensions, in which case boundary blocks will be padded. The dimensionality of the blocks also determines the number of dimensions for which the multidimensional wavelet transform will be applied. Only the n fastest-varying dimensions of the array, where n is the value of this attribute, will undergo wavelet transformation. The order of transformations for the forward transform is from the fastest varying dimension to the slowest.

WASP.CRatios: An ordered list of unique integers indicating the requested compression ratio for each available approximation. For example, the values 100, 10, and 1 would indicate that approximations of the array be constructed using $1/100^{th}$ of the available coefficients, $1/10^{th}$, and all of the coefficients, respectively. The coefficients for each available approximation may be stored in separate files as described by the global WASP.NumFiles attribute documentation.