Geophysical Research Abstracts
Vol. 15, EGU2013-10937, 2013
EGU General Assembly 2013

NetCDF-U – Uncertainty conventions for netCDF datasets

Lorenzo Bigagli (1), Stefano Nativi (1), and Ben Domenico (2)

(1) National Research Council of Italy – IIA, Monterotondo (RM), Italy (lorenzo.bigagli@cnr.it), (2) Unidata Program Center, UCAR, Boulder (CO), USA

To facilitate the automated processing of uncertain data (e.g. uncertainty propagation in modeling applications), we have proposed a set of conventions for expressing uncertainty information within the netCDF data model and format: the NetCDF Uncertainty Conventions (NetCDF-U).

From a theoretical perspective, it can be said that no dataset is a perfect representation of the reality it purports to represent.
Inevitably, errors arise from the observation process, including the sensor system and subsequent processing, differences in scales of phenomena and the spatial support of the observation mechanism, lack of knowledge about the detailed conversion between the measured quantity and the target variable. This means that, in principle, all data should be treated as uncertain.

The most natural representation of an uncertain quantity is in terms of random variables, with a probabilistic approach. However, it must be acknowledged that almost all existing data resources are not treated in this way. Most datasets come simply as a series of values, often without any uncertainty information. If uncertainty information is present, then it is typically within the metadata, as a data quality element. This is typically a global (dataset wide) representation of uncertainty, often derived through some form of validation process. Typically, it is a statistical measure of spread, for example the standard deviation of the residuals.

The introduction of a mechanism by which such descriptions of uncertainty can be integrated into existing geospatial applications is considered a practical step towards a more accurate modeling of our uncertain understanding of any natural process.

Given the generality and flexibility of the netCDF data model, conventions on naming, syntax, and semantics have been adopted by several communities of practice, as a means of improving data interoperability. Some of the existing conventions include provisions on uncertain elements and concepts, but, to our knowledge, no general convention on the encoding of uncertainty has been proposed, to date.

In particular, the netCDF Climate and Forecast Conventions (NetCDF-CF), a de-facto standard for a large amount of data in Fluid Earth Sciences, mention the issue and provide limited support for uncertainty representation.

NetCDF-U is designed to be fully compatible with NetCDF-CF, where possible adopting the same mechanisms (e.g. using the same attributes name with compatible semantics).
The rationale for this is that a probabilistic description of scientific quantities is a crosscutting aspect, which may be modularized (note that a netCDF dataset may be compliant with more than one convention).

The scope of NetCDF-U is to extend and qualify the netCDF classic data model (also known as netCDF3), to capture the uncertainty related to geospatial information encoded in that format. In the future, a netCDF4 approach for uncertainty encoding will be investigated.

The NetCDF-U Conventions have the following rationale:
• Compatibility with netCDF-CF Conventions 1.5.
• Human-readability of conforming datasets structure.
• Minimal difference between certain/agnostic and uncertain representations of data (e.g. with respect to dataset structure).

NetCDF-U is based on a generic mechanism for annotating netCDF data variables with probability theory semantics. The Uncertainty Markup Language (UncertML) 2.0 is used as a controlled conceptual model and vocabulary for NetCDF-U annotations.
The proposed mechanism anticipates a generalized support for semantic annotations in netCDF.

NetCDF-U defines syntactical conventions for encoding samples, summary statistics, and distributions, along with mechanisms for expressing dependency relationships among variables.

The conventions were accepted as an Open Geospatial Consortium (OGC) Discussion Paper (OGC 11-163); related discussions are conducted on a public forum hosted by the OGC.

NetCDF-U may have implications for future work directed at communicating geospatial data provenance and uncertainty in contexts other than netCDF.

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n° 248488.