Achieving Comparability of Model and Observatory Datasets in YOPPsiteMIP: Challenges and Solutions

Leslie M. Hartten[1, 2], Siri Jodha S. Khalsa[3], Jareth I. Holt[4], Laura X. Huang[5], and Johanna A. K. Tjernström[6, 7]

[1] Cooperative Institute for Research in Environmental Sciences (CIRES), Univ. of Colorado Boulder   [2] NOAA/Physical Sciences Laboratory (PSL)
[3] National Snow and Ice Data Center (NSIDC), CIRES, Univ. of Colorado Boulder   [4] Department of Meteorology, Stockholm University (MISU), Sweden
[5] Environment and Climate Change Canada (ECCC), Toronto, Ontario, Canada   [6] Norwegian Meteorological Institute, Norway   [7] Swedish Meteorological and Hydrological Institute, Sweden

The Year of Polar Prediction supersite Model Intercomparison Project (YOPPsiteMIP)

YOPPsiteMIP is focused on about a dozen "supersites" in the Arctic and also in Antarctica. Each supersite has multiple suites of instruments and uses (somewhat) standardized measurement practices, data processing, and quality control.

For each site, observationalists are creating

Merged Observatory Data Files (MODFs):

- Multivariate files for each supersite
- Similar nomenclature, metadata, & structure
- High-resolution observations *directly comparable* to

Merged Model Data Files (MMDFs) for the same locations

- File Conventions (netCDF)
- Variable Naming Conventions (CMIP6* or similar)
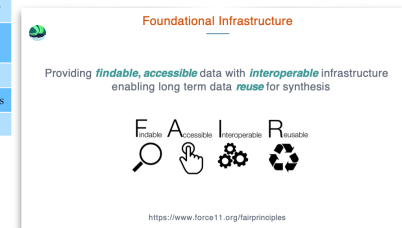- Metadata Conventions (Climate and Forecast (CF†) & Attribute Convention for Dataset Discovery (ACDD‡))
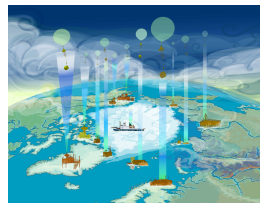
* https://cmip6dr.github.io/Data_Request_Home/ , "Spreadsheet view of the variable definitions"
† http://cfconventions.org/ ‡ http://wiki.esipfed.org/index.php/Attribute_Convention_for_Data_Discovery

Challenges of Creating MODFs (& MMDFs)
Using Open Science & Open Data Practices

Observations	Model Output
Variety of instrument suites	Variety of model formulations
Custom data collection specs, e.g. height & time	Custom space & time specs, possibly including "beams"
May duplicate variables	May produce slightly different variables than are observed
Variety of file formats	Variety of file formats
Attributable to multiple institutions	Attributable to multiple NWP centers
Multiple MODF Makers	Multiple MMDF Makers

Foundational Infrastructure

Providing *findable*, *accessible* data with *interoperable* infrastructure enabling long term data *reuse* for synthesis

F Findable A Accessible I Interoperable R Reusable

https://www.force11.org/fairprinciples

Source: Jones et al. (2019, Ch. 18; modified by overlaying Arctic Data Center logo on upper-left corner)

The Community Response: The H-K Variable SchemaTable

Overview

- Describes desired global attributes

- Defines all variables & recommended attributes

- Is published* on Zenodo with a DOI (PDF & computer-readable; currently Version 1.2)

- Enables MODF & MMDF Makers to create files using current requirements & favorite software

A Community Effort, Relying on …

- The Visionaries and the Detail-Oriented

- The Users – Observationalists & MODF Makers
  - Detailed bug reports
  - The A-M Variable & Attribute Template Table*
  - An MODF "checker"

- The Users – Modelers & MMDF Makers
  - Desire for online plotting tools drives some features
  - Model capabilities drive additions (and request for new CF standard names)

Four "Sheets" Focus on Provenance & Attribution
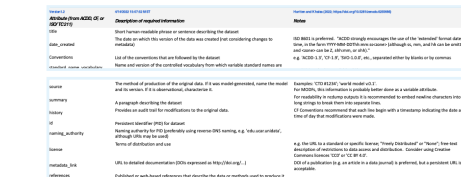
1. Instructions
2. Lineage
3. Global Attributes
4. Observatory & Model Variables (specs include minimum recommended attributes)

What's in the Near Future

- **H-K Variable SchemaTable, Version 1.3**
  - in development, expected June 2023

- **MODF Concept Paper**
  - Uttal et al. (to be submitted June/July 2023)

- **MODFs for 7 Supersites & Accomanying Data Science Journal Article**
  - Mariani, Morris, et al. (to be submitted June 2023)

- **YOPPsiteMIP Overview & Preliminary MODF–MMDF Analysis**
  - Day et al. (to be submitted Summer 2023)

Acknowledgments

References

Day, J., G. Svensson, ..., L.M. Hartten, and CoAuthors, 2023: The YOPP site Model Intercomparison Project (YOPPsiteMIP) Phase 1: Project Overview and Arctic Winter Forecast Evaluation. to be submitted to Geoscientific Model Development.

Hartten, L.M., and S.J.S. Khalsa, 2022: The H-K Variable SchemaTable developed for the YOPPsiteMIP. https://doi.org/10.5281/zenodo.6255666

Morris, S.M., and E. Akish, 2022: A-M Variable & Attribute Template Table developed for the YOPPsiteMIP. https://doi.org/10.5281/zenodo.6780400

Morris, S.M., Z. Mariani, ..., L. M. Hartten, and CoAuthors, 2023: Special Observing Period (SOP) Data for the Year of Polar Prediction site Model Intercomparison Project (YOPPsiteMIP). to be submitted to Earth System Science Data.

Uttal, T., L.M. Hartten, and CoAuthors, 2023: Merged Observatory Data Files (MODFs): An Integrated Research Data Product Supporting Process Oriented Investigations and Diagnostics. to be submitted to Geoscientific Model Development.

Contacts: hartten@colorado.edu & khalsa@colorado.edu

presented at the CIRES Rendezvous, Boulder CO USA, 16 May 2023