cdo
Data Processing (and Production)

Luis Kornblueh, Uwe Schulzweida, Deike Kleberg, Thomas Jahns, Irina Fast

Max-Planck-Institut für Meteorologie, DKRZ

September 24, 2014

MAX-PLANCK-GESELLSCHAFT

Data format standards

Pushing forward ......

- World Meteorological Organization (WMO): grib and bufr, (*since 1980, data converted back to 1900*)
- (NetCDF) Climate and Forecast (CF) Metadata Convention, (*since 1999*)
- World Climate Research Program (WCRP): CMOR (*for CMIP5 and onwards*)

Missing common semantics and vocabulary!

What is CDO ?

CDO is a collection of tools to process and analyze data from climate and NWP models.

- (File) format conversion: GRIB ⇔ netCDF
- Interpolation between different grid types and resolution
- Portability (ANSI C99 with some POSIX extentions)
- Performance (fast processing of large datasets, muti-threaded)
- Modular design and easily extendable with new operators
- UNIX command line interface, tested on many UNIX/Linux systems, Cygwin, and MacOS-X

And what is CDI ?

Data I/O Interface

CDI, used by CDO, is the I/O interface shared by all major MPI-M models. GRIB support includes highly efficient, fast compression algorithms.

- GRIB1 via CGRIBEX (MPI-M)
- GRIB2 via GRIB_API (ECMWF)
- netCDF, CF-convention (UNIDATA)
- SERVICE, EXTRA, IEG (MPI-M legacy binary formats)
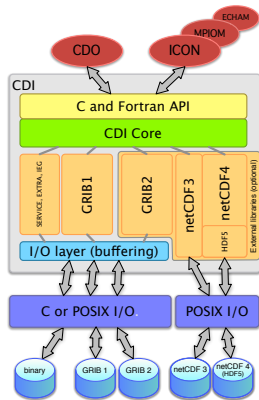
Available Operators

CDO provides more than 400 operators which can be pipelined on thread level. CPU time intensive operators are OpenMP parallelized.

| Main categories	Description
File information	Print information about datasets
File operations	Copy, split and merge datasets
Selection	Select parts of a dataset
Comparision	Compare datasets
Modification	Modify data and metadata
Arithmetic	Arithmeticly process datasets
Statistical values	Ensemble, field, vertical and time statistic
Interpolation	Horizontal, vertical and time interpolation
Import/Export	HDF5, binary, ASCII
Climate indices	ECA Indices

Supported Grids

A large set of grids is supported including spectral- and Fourier-coefficients. Gaussian grids, regular and rotated lat-lon grids, conformal mapped quadrilateral grids, and finally general unstructured grids.
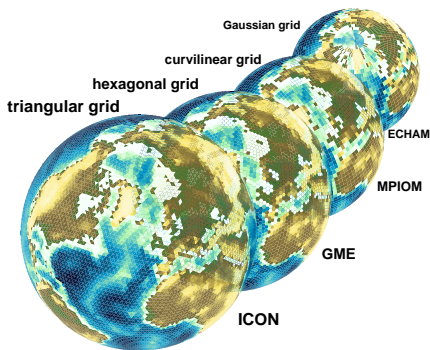
A lot of models world wide are supported:
COSMOS, CLM/COSMO, ECHAM, GME, HIRLAM, ICON, IFS, MPIOM, NEMO, REMO, and ...
only to mention a few

Satellite-data Support

EUMETSAT's Climate Monitoring Satellite Application Facility provides satellite-derived geophysical parameter for climate monitoring. Data sets contain several cloud parameters, surface albedo, radiation fluxes, temperatur and humidity profiles. These products are stored in HDF5. DWD has funded an CDO import operator import_cmsaf.

toa radiation

cloud cover

surface radiation

humidity
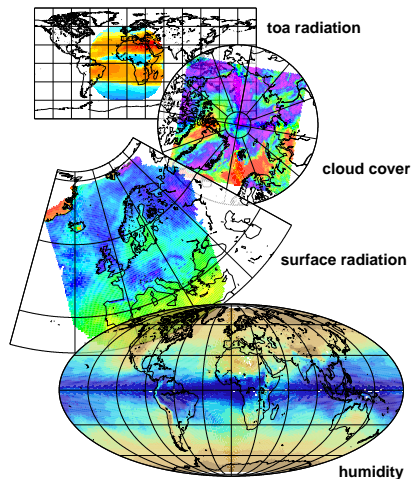
Community Support

The rapidly increasing number of CDO installations and users create a very high demand of support. A fully featured development platform is available to support the community. The CDO community page was funded by the European Commission infrastructure project IS-ENES.

- User wiki
- Documentation
- Bug tracking system
- User forums
- Download area
- Repository access
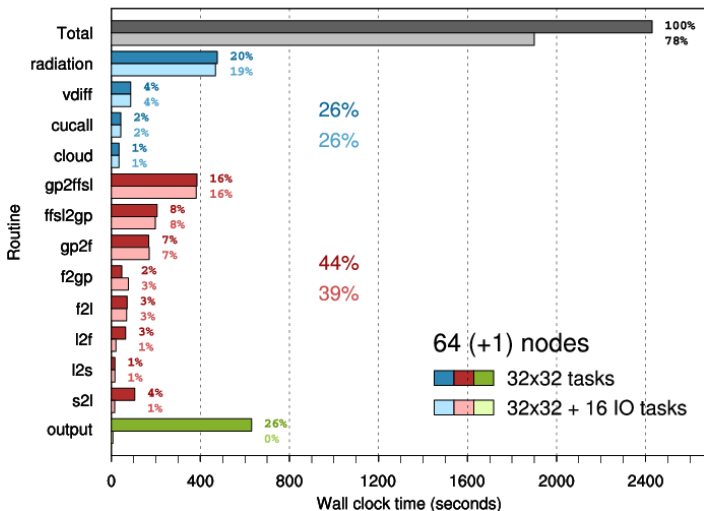
http://code.zmaw.de/projects/cdo

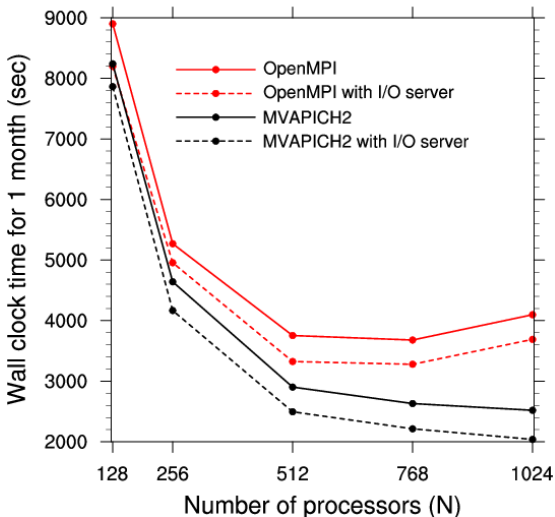# Parallel Output Design
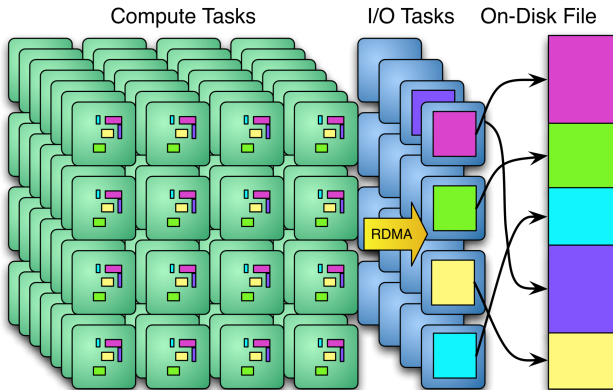
Parallel Output Implementation

Top routines in ECHAM6 T127L95

Parallel Output scaling

ECHAM6 T127L95 OpenMPI vs. MVAPICH2

Upcoming features and future development

cdo/cdi needs to be able to handle 10^8 grid points per level (800 GB, double). A few developments necessary:

- very fast addon compression for grib2 to be validated with WMO members (interface for libaec now in grib_api)
- change from ANSI-C to C++ (mac create minor portability problems)
- change to an master-slave scheduling model to achieve maximum parallelization inside nodes
- add cmorizing capabilities
- add full single precision data flow (only grib_api is missing)

- do not develop a MPI parallelized version, as I/O is the major bottleneck.