CHFP Data Producer Guide
V 2.0 15 March 2013

1. Variables to be produced

See the CHFP proposal for details. The following tables include a summary of information for the requested variables. Variable and CF standard names are left blank where we do not yet have any example data.

Table 1: CHFP atmospheric surface fields

Variable	Var. Name	CF Standard Name	Output
Surface temperature (SST+land)	ts	surface_temperature	24h inst
2m T daily max	tasmax	air_temperature	24h inst
2m T daily min	tasmin	air_temperature	24h inst
Mean sea level pressure	psl	air_pressure_at_sea_level	24h inst
Snow depth	snld	lwe_thickness_of_surface_snow_amount	24h inst
Total soil moisture	mrsov	volume_fraction_of_water_in_soil	24h inst
Total precipitation	prlr	lwe_thickness_of_precipitation_amount	24h acc
Downward surface solar	rsds	surface_downwelling_shortwave_flux_in_air	24h acc
Downward surface longwave	rlds	surface_downwelling_longwave_flux_in_air	24h acc
Net surface solar	rss	surface_net_downward_shortwave_flux	24h acc
Net surface longwave	rls	surface_net_downward_longwave_flux	24h acc
Top net solar	rst	toa_net_downward_shortwave_flux	24h acc
Top net longwave	rlt	toa_net_downward_longwave_flux	24h acc
Surface latent flux	hflsd	surface_downward_latent_heat_flux	24h acc
Surface sensible flux	hfssd	surface_downward_sensible_heat_flux	24h acc
Surface stress (x)	**tauu**	**surface_downward_eastward_stress**	24h acc
Surface stress (v)	**tauv**	**surface_downward_northward_stress**	24h acc
2m temperature	tas	air_temperature	6h inst
2m dewpoint temperature	tdps	dew_point_temperature	6h inst
Total cloud cover	clt	cloud_area_fraction	6h inst
10m wind (u)	uas	eastward_wind	6h inst
10m wind (v)	vas	northward_wind	6h inst

Table 2: CHFP atmospheric pressure level fields

Variable	Var. Name	CF Standard Name	Output
Geopotential	g	geopotential	24h inst
Temperature	ta	air_temperature	24h inst
Zonal velocity	ua	eastward_wind	24h inst
Meridional velocity	va	northward_wind	24h inst
Specific humidity	hus	specific_humidity	24h inst

Requested pressure levels are 850, 500, 200, 100, 50 and 10 hPa. The highest level(s) will only make sense for models that have an appropriate representation of the stratosphere. As well as the "high frequency" output, monthly means of all of the fields must also be calculated and provided.

If, for logistical reasons, high frequency fields are provided at a lower frequency than requested in the CHFP protocol (for example, daily fields instead of 6 hourly), then the monthly means should still be calculated using data **at the originally requested frequency.** Note that it is important that two "constant fields" are also supplied, namely the land-sea mask and the model orography.

Table 3: CHFP ocean fields

Variable	Var. Name	CF Standard Name	Output
Potential temperature	thetao	sea_water_potential_temperature	monthly avg
Salinity	so	sea_water_salinity	monthly avg
Zonal velocity	uo	sea_water_x_velocity	monthly avg
Meridional velocity	vo	sea_water_y_velocity	monthly avg
Vertical velocity	wo	upward_sea_water_velocity	monthly avg
Sea level	zoh	sea_surface_height_above_geoid	monthly avg
Mixed layer depth	zmlo	ocean_mixed_layer_thickness	6h inst
SST	**sst**	**sea_surface_temperature**	6h inst
Net surface heat flux	**hfns**	**surface_downward_heat_flux_in_air**	6h acc
Net surface solar flux	**rss**	**surface_downward_shortwave_flux**	6h acc
Zonal surface stress	**tauu**	**surface_downward_eastward_stress**	6h acc
Meridional surface stress	**tauv**	**surface_downward_northward_stress**	6h acc
Fresh water flux	**fwf**	**water_flux_into_sea_water** or **surface_downward_water_flux**	6h acc

The high frequency ocean data should also be supplied as monthly means. The CHFP protocol also asks for zonal sections of T, S and currents to be archived at 24h intervals at the

equator and 2N/2S. These should use the same standard names as above, but may need different variable names.

2. Interpolating data to the standard grid

To allow model comparisons and multi-model calculations to be made easily, it is crucial that all of the model data are made available on a **common grid**. This grid is a 2.5 by 2.5 deg grid, ranging from 90N,0E to 90S,357.5E, ie 144*73 grid points.

To minimize degradation of information when interpolating, it is recommended to use some form of area averaging. That is, the point 10N, 20E should represent the area average of 8.75-11.25N, 18.75-21.25E. This is particularly important when interpolating from higher resolution grids.

Ocean data are to be provided on a common 1 by 1 deg grid, ranging from 89S,0E to 89N, 259E, ie 360*179 grid points. Note this is shifted by 0.5 degrees from the standard Levitus grid. It is recommended to use the ENACT common land-sea mask. This is available as a 3-D netCDF file from http://www.ecmwf.int/research/EU_projects/ENSEMBLES/data/ocean_archiving_interpol.html. For the three dimensional fields, standard Levitus levels are to be used, ie 33 levels with depths of 0.0, 10.0, 20.0, 30.0, 50.0, 75.0, 100.0, 125.0, 150.0, 200.0, 250.0, 300.0, 400.0, 500.0, 600.0, 700.0, 800.0, 900.0, 1000.0, 1100.0, 1200.0, 1300.0, 1400.0, 1500.0, 1750.0, 2000.0, 2500.0, 3000.0, 3500.0, 4000.0, 4500.0, 5000.0, 5500.0 metres.

Of course, for some types of analysis of individual models it would be better to have the data on the original model grid. Data producers are thus encouraged to make data available on the original grid where they have the resources to do so. It is recommended that they are made available in netCDF. The “original grid” datasets are at the present time not centrally coordinated, and the “central” dataservers do not have the resources to handle them. If you are able to make available data on the original grid, please let us know about it so that we can give the web link.

3. Creating netCDF data files

The data should be in CF-compliant netCDF files. (Note: we are still waiting to hear whether our extensions to allow “ensemble dimension” will be accepted as part of the official CF standard. We will use them in any case). Variables should be labeled with the “standard names” given above.

It will be easiest for the dataserver if the data files from the individual models are supplied with metadata that match the requirements of CHFP. It is thus recommended to use the variable names from the table above, eg “ts” for surface temperature. These follow PCMDI conventions where possible, and are used in the European ENSEMBLES project for seasonal, decadal and climate change integrations. See the accompanying documentation on CHFP netCDF specification for full details.

4. Finding a Data Server

Once data has been interpolated to the standard grid and converted to netCDF, arrangements

need to be made to serve the data via THREDDS. Data producers may choose to do this themselves - it gives flexible control over the data, potentially allows large data volumes to be made available, and avoids data transfer issues. It does require some resources, though, both in terms of hardware and human effort, and for some groups it is likely to be more appropriate to work with another organization who have the resources in place to serve CHFP data.So far three “central” Data Servers have been identified. These are the ENSEMBLES dataserver at ECMWF, which serves the data from the European ENSEMBLES project; a dataserver at APCC in Korea; and a dataserver at CIMA in Argentina. Web links to these data centres and any others serving CHFP data are available from the CHFP web pages. Data producers are free either to contact an appropriate Data Server directly, or to ask for advice from WGSIP. It is possible that we will duplicate some of the most popular datasets across several servers.

Once a Data Server has been found, arrangements will need to be made to transfer the data to them. In some situations ftp might be possible, but for large datasets a faster and more cost-effective method might be physical transfer of the data on one or more USB hard disks.