US National Multi-Model Data Management Plan

1. Estimate of Data Requirements

To calculate the data requirement, we first estimate the data volume associated with daily (and higher frequency) fields and then adjust for additional monthly mean output.

a) There are 33 years of retrospective forecasts plus real-time forecasts proposed
b) There are 12 initial condition months for each year
c) There are 365 days of lead-time for each forecast (i.e., daily output)
d) We assume that there are 100 ensemble members in the NMME phase 2 project

Parameters a)-d) imply that there are 14454 x 10^3 (33*12*365*100) "data saves" required.

Each data save includes 119 atmospheric and land fields (see tables 1 and 2 below) and 70 ocean and sea-ice fields (see table 3 below). Therefore each data save has 189 fields.

Each field for both the atmosphere and the ocean will be interpolated to a 1x1 degree field (see section 3 below). Depending on the amount of metadata included and on the compression, we estimate that on average each 1x1 degree field (359x180 grid points) will require 500 KB.

Therefore, each data save generates (189 x 500 KB) 94.5 MB giving a grand total of (94.5 MB x 14454 x 10^3 data saves) 1.3659 PB for the daily and higher frequency data. Assuming the monthly mean data are a factor of 30 smaller, the total data volume for both daily and monthly is 1.41 PB. NetCDF4, which all the partners have agreed to use, give a data compression of between 50-70%. This suggests that the total data requirements are 0.705-0.985 PB. Allowing for some flexibility we have requested 1PB of data storage.

2. Variables to be produced

The following tables include a summary of information for the requested variables. Data names conform to CF standards.

Table 1: NMME atmospheric and land surface fields

Variable	Var. Name	CF Standard Name	Output	Number of fields per data save
Surface temperature (SST+land)	Ts	surface_temperature	24h acc	1
2m T daily max	Tasmax	air_temperature	24h inst	1
2m T daily min	Tasmin	air_temperature	24h inst	1
Mean sea level pressure	Psl	air_pressure_at_sea_level	24h acc	1
Water equivalent snow depth	snowhlnd	Water equivalent snow depth	24h acc	1
Total soil moisture	Mrsov	volume_fraction_of_water_in_soil	24h acc	1
Convective precipitation	precc	Convective precipitation rate	3h acc	8
Large scale precipitation	Precl	Large scale (stable) precipitation	3h acc	8
Downward surface solar	Rsds	surface_downwelling_shortwave_flux_in_air	3h acc	8
Downward surface longwave	Rlds	surface_downwelling_longwave_flux_in_air	3h acc	8

Net surface solar	Rss	surface_net_downward_shortwave_flux	24h acc	1
Net surface longwave	Rls	surface_net_downward_longwave_flux	24h acc	1
Top net solar	Rst	toa_net_downward_shortwave_flux	24h acc	1
Top net longwave	Rlt	toa_net_downward_longwave_flux	24h acc	1
Surface latent flux	Hflsd	surface_downward_latent_heat_flux	24h acc	1
Surface sensible flux	Hfssd	surface_downward_sensible_heat_flux	24h acc	1
Surface stress (x)	Stx	surface zonal stress positive to the west	24h acc	1
Surface stress (y)	Sty	surface meridional stress positive to the west	24h acc	1
Precipitable water	Tqm	Total column vertically integrated water	24h acc	1
Vertical integrated moisture flux convergence		Vertically integrated moisture flux convergence	24h acc	1
Total runoff		Total runoff	24h acc	1
Ground heat flux		Ground heat flux	24h acc	1
2m temperature	Tas	air_temperature	3h inst	8
2m dewpoint temperature	Tdps	dew_point_temperature	3h inst	8
Total cloud cover	Clt	cloud_area_fraction	3h inst	8
10m wind (u)	Uas	eastward_wind	3h inst	8
10m wind (v)	Vas	northward_wind	3h inst	8

Table 2: NMME atmospheric pressure level fields

Variable	Var. Name	CF Standard Name	Output	Number of Fields per data save
Geopotential	G	geopotential	24h acc	5
Temperature	Ta	air_temperature	24h acc	5
Zonal velocity	ua	eastward_wind	24h acc	5
Meridional velocity	uv	northward_wind	24h acc	5
Specific humidity	hus	specific_humidity	24h acc	5
Velocity potential 850 hPa		Velocity potential 850 hpa	24h acc	1
Velocity potential 200 hPa		Velocity potential 850 hpa	24h acc	1
Stream function 850 hPa		Stream function 850 hPa	24h acc	1
Stream function 200 hPa		Stream function 200 hPa	24h acc	1

The six pressure levels requested are 850, 500, 200, 100, 10 hPa.

As well as the “high frequency” output, monthly means of all of the fields must also be calculated and provided as included in the data volume estimate. Note that it is important that two “constant fields” are also supplied, namely the land-sea mask and the model orography.

Table 3: NMME ocean and sea-ice fields

Variable	Var. Name	CF Standard Name	Output	Number of fields per save
Potential temperature	thetao	sea_water_potential_temperature	24h acc	13
Salinity	so	sea_water_salinity	24h acc	13
Zonal velocity	uo	sea_water_x_velocity	24h acc	13
Meridional velocity	vo	sea_water_y_velocity	24h acc	13
Vertical velocity	wo	upward_sea_water_velocity	24h acc	13
Sea level	zoh	sea_surface_height_above_geoid	24h acc	1

Mixed layer depth	zmlo	ocean_mixed_layer_thickness	3h acc	1
Fresh water flux	fwf	fresh_water_flux	24h acc	1
Sea-ice extent			24h acc	1
Sea-ice thickness			24h acc	1

The 13 requested ocean levels are 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, and 400.0 m.

The high frequency ocean data will also be supplied as monthly means.

3. 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 1x1 deg grid, ranging from 90N,0E to 90S,359E, ie 360*180 grid points.

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 "centralized" data service does not have the resources to handle them.

4. Creating netCDF data files

These data should be in CF-compliant netCDF files. Variables should be labeled with the "standard names" given above. Data files from the individual models will be provided with metadata that match the NMME requirements. The metadata will also include sufficient details regarding the model configuration and forcing protocol. It is thus recommended to use the variable names from the table above (e.g., "ts" for surface temperature). These follow PCMDI conventions where possible, and are used in the CMIP5 project for decadal and climate change integrations.

5. Additional Implementation Issues

The NMME experiment is a two-year project, and it is anticipated that the Phase-II data will generated over time. This phasing of the Phase-II data is important in balancing how much data is on tape and how much data is on spinning disk. Our current implementation strategy, developed in consultation with NCAR management, calls for half a petabyte (PB) of storage in each year of the project for a total of 1-PB. Eighty percent of the data storage will be on tape (i.e., HPSS) and twenty percent will be on NCAR's GLADE (Globally Accessible Data Environment, rotating storage) system. This allows for all of the monthly mean and selected high frequency data to be on disk. Additional details will be addressed as the project evolves.