

PROPOSAL FOR S-RIP DATA SHARING: SUGGESTED VARIABLES AND FORMATS

Working Draft (June 24, 2014)

Edwin P. Gerber (gerber@cims.nyu.edu)

Analysis of stratospheric variability on daily to sub-daily time scales across multiple reanalyses requires a tremendous amount of data transfer, disk space, and brute force computation. Many important diagnostics (e.g. Eliassen-Palm fluxes and annular modes) however, can be efficiently computed from zonal mean variables and fluxes. We propose to use S-RIP's BADC disk space to share zonal mean reanalysis data with high time resolution. This sharing will only be possible with a sufficient number of volunteers to share the initial work.

I believe that one can compute many diagnostics of interest (e.g., Eliassen-Palm fluxes, the residual circulation, downward control diagnostics, annular modes, and kinetic energy), with the zonal mean quantities listed in Table 1. I've also suggested some file structures and variable names. It would really help if we applied the same strategy to each reanalysis, including the procedure for handling missing data. Please comment on these four issues:

1. Would these variables would be helpful for your analysis? Is something critical missing?
2. Will this format (i.e. name of variables, files, etc.) be convenient?
3. How fine scale does the output need to be in time? Is daily mean (or one daily snap shot) sufficient? Do we need as high resolution in time as possible?
4. In terms of missing data, should we compute the zonal mean based on the values present (i.e. if half the data is there, the zonal mean is based on just that half). If no values are present at all, the zonal mean will be marked as missing.

The suggested variables and their names are listed in Table 1. I propose assembling the data in yearly chunks. Thus, for a given reanalysis, you'd see a directory in BADC with a name describing the reanalysis and the time sampling, say:

```
merra_zonal_mean_diagnostics_3hr/
```

with (in this case) 34 files:

1979.nc, 1980.nc ... 2012.nc.

This is a rather minimalist naming convention. Another convention would be to have the file name describe everything, say:

```
merra_zonal_mean_diagnostics_3hr_1979.nc
merra_zonal_mean_diagnostics_3hr_1980.nc
...
merra_zonal_mean_diagnostics_3hr_2012.nc
```

I'm writing some simple shell scripts that use the NCO utilities to prepare these data files from the MERRA reanalysis. I believe that they could be applied to other reanalyses, too. We could also share the necessary postprocessing scripts to convert these zonal mean quantities into the residual circulation, E-P fluxes, etc..

	variable	units	description
1	U	m s^{-1}	zonal mean zonal wind, \bar{u}
2	V	m s^{-1}	zonal mean meridional wind, \bar{v}
3	OMEGA	Pa s^{-1}	zonal mean pressure velocity, $\bar{\omega}$
4	T	K	zonal mean temperature, \bar{T}
5	H	m	zonal mean geopotential height, \bar{Z}
6	uu	m^2s^{-2}	zonal mean squared eddy zonal velocity, $\overline{u'u'}$
7	vv	m^2s^{-2}	zonal mean squared eddy meridional velocity, $\overline{v'v'}$
8	uv	m^2s^{-2}	zonal mean meridional flux of zonal momentum, $\overline{u'v'}$
9	uomega	m Pa s^{-2}	zonal mean vertical flux of zonal momentum, $\overline{u'\omega'}$
10	vt	m K s^{-1}	zonal mean meridional heat flux, $\overline{v'T'}$

Table 1: Proposed zonal mean quantities to share. The second column lists (a suggestion) for the actual name of the variables, as you would find them in the netcdf file. In terms of the descriptions, u , v , ω , T and Z are the zonal and meridional winds, pressure velocity, temperature, and geopotential height, respectively. An overline indicates a zonal mean, and a prime, a deviation from the zonal mean.