

*Supplement of*

## **Overview of experiment design and comparison of models participating in phase 1 of the SPARC Quasi-Biennial Oscillation initiative (QBOi)**

N. Butchart et al.

*Correspondence to:* Neal Butchart (neal.butchart@metoffice.gov.uk)

### **Data format and storage**

It is very strongly recommended that groups should upload their QBOi model output data in the CF-compliant netCDF format, which is the standard CMIP-required format. Storage for the QBOi multi-model dataset is provided by the Centre for Environmental Data Analysis (CEDA). The CEDA data storage and processing service is called JASMIN, and the storage area it provides for the data is referred to as a "group workspace". JASMIN also offers the option to process data locally, for which a range of standard software packages is provided on the JASMIN machine (Python, R, IDL, etc). Further information on JASMIN resources is available at:

<http://www.ceda.ac.uk/services/analysis-environments/>

More detail on data format and JASMIN is given in the subsections below. Please bring any errors, ambiguities, or suggested improvements in this information to the attention of the QBOi coordinators as soon as possible.

#### **S1 Data format: CF-compliant netCDF**

Guidelines for preparing the data in CF-compliant netCDF format are available at:

[https://badc.nerc.ac.uk/help/formats/netcdf/index\\_cf.html](https://badc.nerc.ac.uk/help/formats/netcdf/index_cf.html)

This webpage gives an overview of the netCDF file structure and metadata requirements. The CMIP conventions should be followed, so that if a modelling centre already has a process in place to produce CF-compliant netCDF files (e.g., as a result of having contributed to CMIP5) then it is sufficient to use this process for QBOi model output files. It is recommended that the CMIP5 version of the conventions be used. It is also acceptable if filenames and directory structure use the CMIP6 standard, but at the time of this writing (Oct 2016) the CMIP6 standard has not yet been finalized. For this reason the CMIP5 standard is preferred. The CMIP5 conventions are described at:

[http://cmip-pcmdi.llnl.gov/cmip5/data\\_description.html](http://cmip-pcmdi.llnl.gov/cmip5/data_description.html)

and by following the links on that page. The filename and directory structure should conform to the CMIP5 standard as described by:

[http://cmip-pcmdi.llnl.gov/cmip5/docs/cmip5\\_data\\_reference\\_syntax.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/cmip5_data_reference_syntax.pdf)

which for the directory structure is (following Sec. 3.1 of the above document):

<institute>/<model>/<experiment>/<frequency>/<modeling realm>/<variable name>/<ensemble member>

(It is not necessary to include the "<activity>/<product>" directories at the start of the path. Also note that the "<model>" directory is useful if a group contributes QBOi runs using more than one version of their model.) The corresponding standard filename structure is (following Sec. 3.3 of the above document):

<variable name>\_<MIP table>\_<model>\_<experiment>\_<ensemble member>[\_<temporal subset>][\_<geographical info>].nc

For example, one netCDF file of CCCma's model output (monthly-mean zonal wind for QBOi Experiment 1) would reside on JASMIN as:

```
/group_workspaces/jasmin2/qboi/CCCma/CMAM/QBOiExp1/mon/atmos/ua/rlilp1/ua_Amon_CMAM_QBOiExp1_rlilp1_197901-200912.nc
```

See **Sec. S2**, below, for information on standard values of labels that should be used (e.g., "QBOiExp1"). See **Sec. S3**, below, for an example of how this file would be uploaded to CEDA. Online tools are available to check that a netCDF file is CF-compliant:

<http://puma.nerc.ac.uk/cgi-bin/cf-checker.pl>  
<http://titania.badc.rl.ac.uk/cgi-bin/cf-checker.pl>

For reference, the CF conventions are described at:

<http://cfconventions.org/>

A useful section of this website gives a concise listing of the requirements and recommendations for any given CF version:

<http://cfconventions.org/requirements-and-recommendations.html>

For the above example of a CCCma netCDF file, CF version 1.4 was used since this was standard at the time of CMIP5. However, note that CF versions are designed to be backward compatible with previous versions. The CMOR software, which can be used to produce CF-compliant netCDF files, is described at:

<http://cmor.llnl.gov/>

## **S2 QBOi-specific labels**

Standard labels help to avoid unnecessary variation of terminology. For CMIP this is referred to as a Control Vocabulary (CV); for example,

[http://cmor.llnl.gov/mydoc\\_cmor3\\_CV/](http://cmor.llnl.gov/mydoc_cmor3_CV/)

describes the CMIP6 standards, including standard experiment labels ("historical", "piControl", etc).

Here we recommend some standard usage for labels used in QBOi filenames and directories.

The "<model>" label is to be decided by each modelling centre, but it should uniquely identify the model that was used. For example, if CCCma uploads runs from two versions of the model, both model versions cannot be called "CMAM". It is understood that this label is taken to be unique only within the QBOi project (e.g. a different "CMAM" would have contributed to the CCMI project). The label does not need to be descriptive; it is simply required that the model configuration corresponding to the label can be unambiguously identified.

For the "<experiment>" label, the following values are recommended:

QBOi experiment	<experiment> label
Experiment 1 (AMIP)	QBOiExp1
Experiment 2 (1xCO2)	QBOiExp2
Experiment 3 (2xCO2, +2K SST)	QBOiExp3
Experiment 4 (4xCO2, +4K SST)	QBOiExp4
Experiment 5 (hindcast)	QBOiExp5 (see note below)

For Experiment 5 the initialization date should be indicated by a "<sub-experiment>" label that is added to the "<ensemble member>" label:

```
May1993-r1i1p1  
Nov1993-r1i1p1
```

and so forth. This is to avoid ambiguity in the filename between the initialization date and the time range of the data (which is indicated by the "<temporal subset>" label). For example, a filename ending in

```
..._QBOiExp5_r1i1p1_199311-199410.nc
```

might indicate one year of output for a hindcast initialized in Nov 1993, but if the hindcast experiments are run for longer than a year (as could be done to examine decadal predictability) then this filename might also indicate the Nov 1993 to Oct 1994 period of a hindcast that was initialized earlier. The ambiguity is avoided by incorporating the "<sub-experiment>" label into the "<ensemble member>" part of the filename:

```
ua_Amon_CMAM_QBOiExp5_Nov1993-r1i1p1_199311-199410.nc
```

The full path of the above file then will then appear appear as:

```
/group_workspaces/jasmin2/qboi/CCCma/CMAM/QBOiExp5/mon/atmos/ua/Nov1993-  
r1i1p1/ua_Amon_CMAM_QBOiExp5_Nov1993-r1i1p1_199311-199410.nc
```

### S3 Data storage: BADC (JASMIN) access

Registration is required to access JASMIN at CEDA. An overview of the required steps is given at:

<http://jasmin.ac.uk/workflow/>

Users may also find the following summary of the required steps to be useful:

<http://s-rip.ees.hokudai.ac.jp/resources/data.html>

under the section "How to Access JASMIN". This summary was prepared by the SPARC S-RIP Activity in response to user difficulties with the CEDA registration process. While the CEDA's own information should take precedence, the S-RIP page may help to clarify the process. Also on the S-RIP page, the "What You Can Do With JASMIN" section may help you to get started once you have access.

In particular, note that data can be uploaded using the Linux rsync command:

```
rsync -tput [local filenames or dirs] [username]@jasmin-xfer1.ceda.ac.uk:/group_workspaces/jasmin2/qboi/[your modelling centre]/[your model]/[path to data]
```

For example, to upload one netCDF file of CCCma's model output (monthly-mean zonal wind for QBOi Experiment 1):

```
rsync -tput ua_Amon_CMAM_QBOiExp1_r1i1p1_197901-200912.nc janstey@jasmin-xfer1.ceda.ac.uk:/group_workspaces/jasmin2/qboi/CCCma/CMAM/QBOiExp1/mon/atmos/ua/r1i1p1/
```

The software packages installed on JASMIN are listed at

[https://github.com/cedadev/jasmin\\_scivm/wiki/Packages](https://github.com/cedadev/jasmin_scivm/wiki/Packages)

Note that IDL is not listed, but can be accessed by typing "module add idl", after which then the "idl" command will invoke IDL. The CEDA asks that users who wish to run more than one instance of idl simultaneously please use run-time licences in order to be economical with the more limited pool of development licences. For information on this, see

<http://proj.badc.rl.ac.uk/cedaservices/wiki/JASMIN/AnalysisServers>

under "Making efficient use of IDL development licences" and "Example usage of IDL Runtime Licences".