Issues Related to the Dissemination of Climate Model Output

Karl E. Taylor

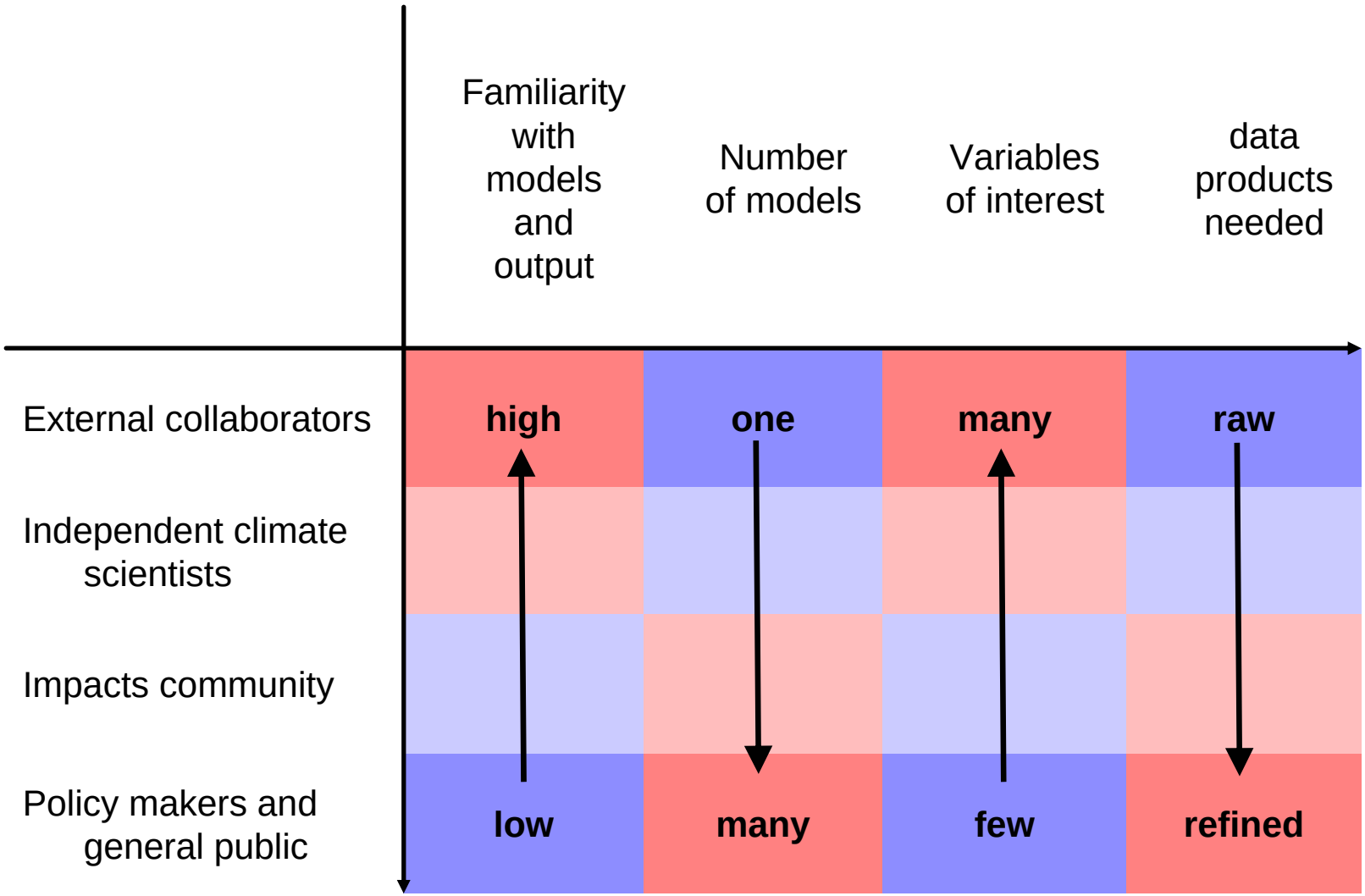
Program for Climate Model Diagnosis and Intercomparison
Lawrence Livermore National Laboratory

First Session of the WCRP Modelling Panel (WMP)

Met Office, Exeter, UK

6-7 October 2005

Who wants model output?



What *common* needs are there?

- Model documentation
- Description of experiment design
- Metadata describing field(s) stored
  - → Precise definition of what physical quantity is represented by the data
  - → Units (and calendar for time)
  - → Space-time location of data, as well as definition of other “independent” variables
  - → Other grid information (e.g., grid-cell area)
  - → Processing applied (interpolation, climatology, zonal averaging, etc.)
- Uniform representation of all of the above, across models and experiments.
- Data “discovery” tools to help users locate desired model output.

What is current situation?

- Model documentation: uneven and not easy to find – exception IPCC model documentation

  (see http://www-pcmdi.llnl.gov/ipcc/info_for_analysts.php)

- Description of experiment design: see above

- Metadata describing field(s) stored: often insufficient

- Uniform representation of all of the above, across models and experiments – only for a few well-coordinated model intercomparison projects (e.g., IPCC, AMIP, PMIP)

- Data “discovery” tools to help users locate desired model output – available only for limited sets of model experiments.

What about the mechanics of serving the data?

- For coordinated projects serving a wide variety of users (like the recent IPCC modeling exercise), data have been sent to a central repository.
  - → Output accepted only if rigid metadata requirements are met.
  - → Quality control and management is relatively easy.
  - → Serves users well with today's technology/software.
  - → Several disadvantages also.
- Various web interfaces are available – e.g., ftp, GDS, ESG, LAS – with various evolving capabilities
  - → All can: transfer complete files from a single site.
  - → Some can deny access except to approved users.
  - → Some can extract subsets of data and perform simple server- side calculations (e.g., obtain a single pressure level, a climatological mean, a zonal mean)
  - → Some can perform more complex server-side calculations
  - → Soon (now?) some can transfer data from disperse sites, but make it look like a single site.

What can we do to move forward?

- Move toward distributed model output archives.
  - → Avoids bandwidth and logistical problems transmitting data to a single repository.
  - → Allows groups to immediately correct their output when errors are found.
  - → Minimizes single point of failure issues.
  - → May not completely supplant a central repository since some groups can't serve data.
- Support use of CF-conventions for netCDF files.
  - → What is CF?
    - Specifications governing creation of fully self-describing netCDF files.
    - An extension of the earlier, more limited "COARDS" standard.
    - It encourages storage of metadata that can be useful in model diagnosis.
    - The conventions enable development of common software that can understand model output from diverse sources.
  - → CF white paper would benefit from endorsement.

What can we do to move forward? (cont.)

- For coordinated modeling experiments, establish IPCC-like requirements on model output.
  - → The CF-conventions specify the "syntax" of self-describing files, but the IPCC requirements specify what information must be included.
  - → For example, IPCC required certain attributes identifying precisely the
    - modeling group
    - experiment performed
    - variable
    - Units
    - calendar assumed
    - area of grid cells, as well as their location
    - a host of other information useful to analysts
    - See http://www-pcmdi.llnl.gov/ipcc/IPCC_output_requirements.htm
  - → IPCC output conforms to the CF conventions

What can we do to move forward? (cont.)

- Capitalize on modeling groups' familiarity with CMOR
  - → Climate Model Output Rewrite (CMOR) facilitated writing data in conformance with IPCC requirements.
  - → CMOR "input tables" provided metadata information and QC specifications, so that CMOR trapped mistakes in model output before it was released for analysis.
  - → CMOR input tables can be tailored to the needs of future data-sharing activities.
  - → See http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php
- Encourage GO-ESSP (Global Organization for Earth System Science Portal)
  - → Collaboration to develop a software infrastructure that will provide distributed access to weather and climate data.
  - → Weave together various frameworks designed for data discovery, access, and analysis.
  - → Mainly involves European & American individuals and institutions.

The IPCC exercise: What was the scientific payoff?

- Output from 21 different models is being widely analyzed
- More than 400 users are registered to examine the output
- More than 25 Tb (60,000 files) collected and 60 Tb (280,000 files) disseminated to analysts
- More than 200 manuscripts have been written, based (at least in part) on the IPCC database.
- The database will likely attract continued scientific interest for several years.
  - → Continue to document current generation models and serve as a reference for assessing the skill of future models.
  - → Additional scientific discoveries.
- The standard output tables and the software developed in support of the IPCC database can be used as is or easily modified to meet the needs of other model intercomparison projects.


PCMDI