IPCC Coupled Model Output For Working Group 1

Karl Taylor

Program for Climate Model Diagnosis and Intercomparison
Lawrence Livermore National Laboratory

Workshop on Analyses of Climate Model Simulations for the IPCC AR4

Honolulu, Hawaii

1-4 March 2005

We are lucky!

- We are the beneficiaries of a remarkable international effort to coordinate the distribution of output from climate simulations of relevance to the IPCC assessment.
- At present the model output archive contains 14.3 Tbytes
- It may grow over the next several months to 30 Tbytes or more.

14 terabytes: How much data is that?

- A single latitude/longitude map at typical climate model resolution represents about ~40 kilobytes.
- If you wanted to look at all 14 Tbytes in the form of these latitude/longitude plots, and if
  - → every 10 seconds you displayed another map, and if
  - → you worked 24 hours a day 365 days each year,
- You could complete the task in about 100 years.
- If we divided up the task among the scientists in this room (working 50-hour weeks), each of us would have to look at a new plot every 10 seconds for 3 years.

Purpose of this talk

- Provide overview of preparation of the database.
- Give credit where it is due (with thanks and appreciation, and apologies to those I inadvertently slight or fail to mention).
- Summarize present state of data archive.
- Discuss what yet needs to be done.

How did the IPCC model output database come about?

- The JSC/CLIVAR Working Group on Coupled Modeling (WGCM) assumed responsibility for coordinating model simulations in support of the IPCC 4th Assessment.
- The WGCM and its “Climate Simulation Panel” agreed on a set of simulations that would be performed.
- In the September 2003, the WGCM asked PCMDI to volunteer to host the database in support of IPCC Working Group I.
- The WGCM provided strong support for PCMDI’s request that data be prepared in such a way that it would not have to be rewritten by PCMDI, and that the output would conform to strict structural and metadata standards.

How did the IPCC model output database come about? (continued)

- By spring, 2004, PCMDI, with input and oversight by the WGCM's Climate Simulation Panel, drew up the list of standard output fields and established the format requirements for model output.
- By summer, 2004, PCMDI made available to modeling groups a FORTRAN library (CMOR), which facilitated conformance with the WGCM's requirements.
- Modeling groups ran their models, rewrote their model output and sent it to PCMDI.
- PCMDI organized the output and made it available via ftp and a "data portal".

Credit: The modeling centers

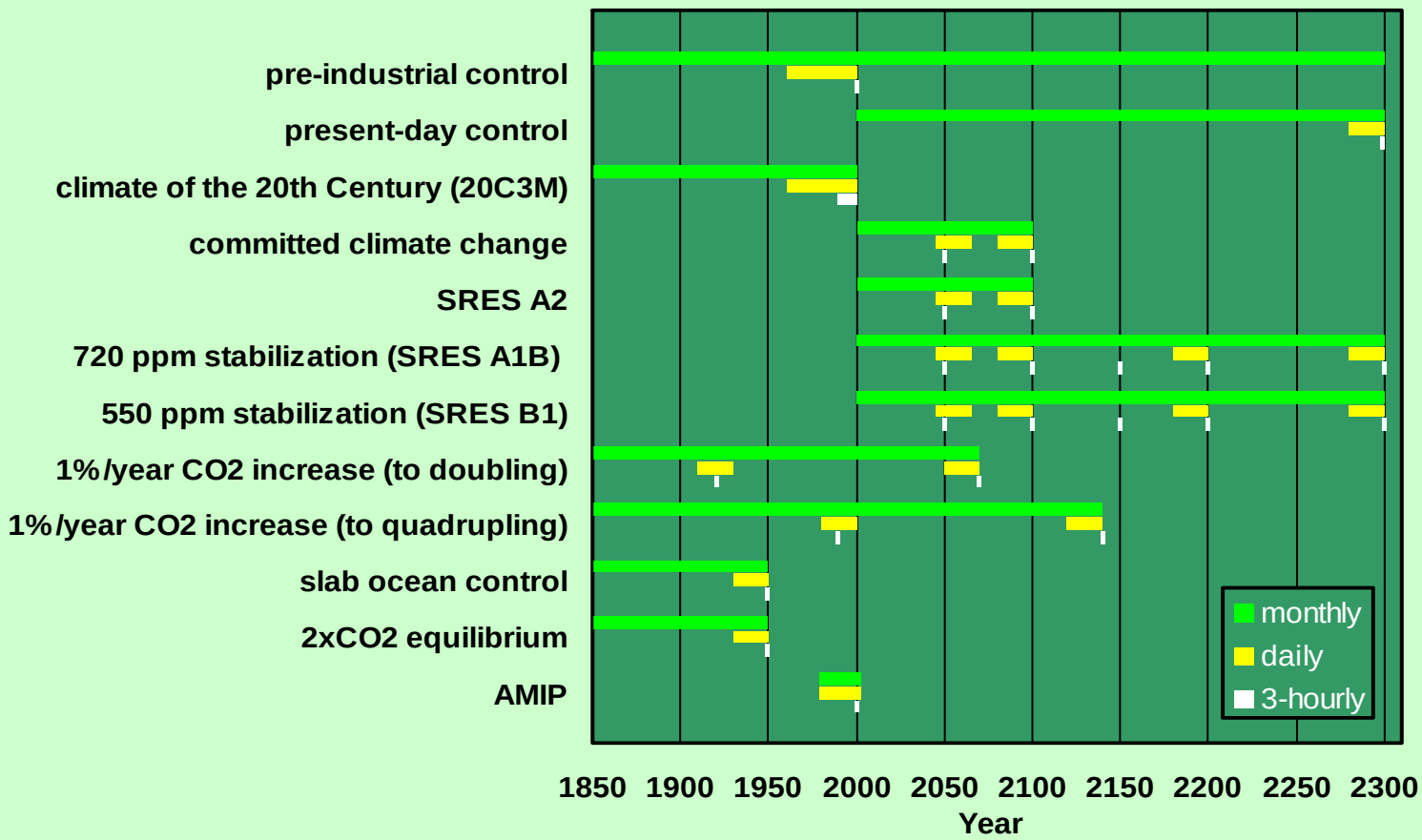
- Many modeling groups have only recently completed an intensive model development cycle.
- Much personal scientific research has been put on hold in order to complete IPCC simulations for analysis by the community.
- Also considerable resources were devoted to rewriting model output for the IPCC archive.

Credit: The Climate Simulation Panel

- Jerry Meehl (chair), John Mitchell, Bryant McAvaney, Curt Covey, Mojib Latif, and Ron Stouffer.

- Provided oversight and input to many of the decisions made concerning:
  - → Simulations performed
  - → Model output fields collected
  - → Time periods focused on
  - → Rules for data access
  - → Etc.

Reporting periods for requested model output



http://www-pcmdi.llnl.gov/ipcc/standard_output.html#Experiments

Credit: developers of the netCDF CF-conventions for metadata

- The CF-conventions allows for creation of self-describing files with information typically needed to perform analysis (substantially extending the COARDS standard).

  (See http://www.cgd.ucar.edu/cms/eaton/cf-metadata/)

- Developers: Jonathan Gregory, Brian Eaton, Bob Drach, Karl Taylor, Steve Hankin.

- For the IPCC model output, additional requirements were imposed to ensure that specific metadata were included. (Karl Taylor, Jonathan Gregory)

  (See http://www-pcmdi.llnl.gov/ipcc/IPCC_output_requirements.htm)

The CF-conventions specify the “syntax” of self-describing files, but the IPCC requirements are stricter.

- Each file must contain only a single output field.
- For data that are a function of longitude and latitude, only grids representable as a Cartesian product of longitude and latitude axes are allowed.
- Most atmospheric fields that are functions of the vertical coordinate must be interpolated to standard pressure levels
- The names of output netCDF files must begin with the 'output variable name' listed in the IPCC Standard Output list of variables, followed by an underscore and then the table number (e.g., tas_O1, psl_A1, tas_A4, etc.).
- The units required for the output fields are given in the IPCC Standard Output tables.
- The positive direction of vertical fluxes must be consistent with that specified in the IPCC Standard Output table "CF standard name".
- The order of array dimensions must be: time, level, latitude, longitude.
- If longitude is a coordinate, data must be stored west to east (in degrees), starting with the first grid point greater than or equal to 0 degrees east.
- If latitude is a coordinate, data must be stored south to north (in degrees).
- If there is a vertical coordinate, data must be stored starting with the grid point nearest the surface.

http://www-pcmdi.llnl.gov/ipcc/IPCC_output_requirements.htm

Credit: Software developers of CMOR

- Climate Model Output Rewriter (CMOR) is a FORTRAN code that rewrites model output to comply with the IPCC requirements.
- Authors: Karl Taylor, Charles Doutriaux, Jean-Yves Peterschmitt (LSCE, France)
- http://www-pcmdi.llnl.gov/software/cmor/cmor_users_guide.pdf

Credit: Development of a list of standard output

- Started from lists developed for AMIP and CMIP.
- Considerable input from PCMDI staff, including Peter Gleckler, Karl Taylor, Ken Sperber, Krishna AchutaRao, Jim Boyle, Curt Covey.
- Considerable input from others, including Jonathan Gregory, Ron Stouffer, Jerry Meehl.
- Variables defined in explicit detail.

(See http://www-pcmdi.llnl.gov/ipcc/standard_output.html)

Sample of standard output table

Table A1a: **Monthly-mean 2-d atmosphere or land surface data (longitude, latitude, time:month).**

	CF standard_name	output variable name	units	notes
1	air_pressure_at_sea_level	psl	Pa	
8	surface_snow_thickness	snd	m	this thickness when multiplied by the average area of the grid cell covered by snow yields the time-mean snow volume. Thus, for time means, compute as the weighted sum of thickness (averaged over the snow-covered portion of the grid cell) divided by the sum of the weights, with the weights equal to the area covered by snow. report as 0.0 in snow-free regions.
15	surface_temperature	ts	K	"skin" temperature (i.e., SST for open ocean)
16	surface_air_pressure	ps	Pa	*not* mean sea-level pressure
19	atmosphere_water_vapor_content	prw	kg m^{-2}	vertically integrated through the atmospheric column
21	surface_runoff_flux	mrros	kg m^{-2} s^{-1}	compute as the total surface runoff leaving the land portion of the grid cell divided by the land area in the grid cell; report as "missing" or 0.0 where the land fraction is 0.
22	runoff_flux	mrro	kg m^{-2} s^{-1}	compute as the total runoff (including "drainage" through the base of the soil model) leaving the land portion of the grid cell divided by the land area in the grid cell; report as "missing" or 0.0 where the land fraction is 0.

http://www-pcmdi.llnl.gov/ipcc/standard_output.html

Credit: Dozens of scientists who spent months preparing IPCC model output for your analysis.

- CCCma, Canada: Daniel Y. Robitaille, Larry Solheim, Warren Lee
- CCSR/NIES/FRCGC (hi-res), Japan: Teruyuki Nishimura, Takashi Sakamoto, Koji Ogochi
- CCSR/NIES/FRCGC (med-res), Japan: Toru Nozawa, Naosuke Okada.
- CNRM, France: Sophie Tyteca, David Salas-Melia
- CSIRO, Australia: Mark Collier, Martin Dix
- GFDL, USA: Keith Dixon, Eric Marshall, Brett DiFrischia, Zhi Liang, Hans Vahlenkamp, V. Balaji
- GISS (C4x3): Gary Russell
- GISS (model E-H, E-R) , USA: Ken Lo
- IAP, China: Weipeng Zheng, Ningfang Zhou, Yongqiang Yu
- INM, Russia: Evgeny Volodin
- IPSL, France: Sébastien Denvil, Laurent Fairhead, Jean-Yves Peterschmitt
- MPI, Germany: Joerg Wegner
- MRI, Japan: Takao Uchiyama, Seiji Yukimoto, Shoji Kusunoki
- NCAR, USA: Lawrence Buja, Gary Strand, Julie Arblaster, Haiyan Teng, Aixue Hu, Diane Feddema
- NCC, China: Ying Xu, Zong-ci Zhao
- UKMO, UK: Simon Gosling, Mark Webb, Jason Lowe

Credit: Transfer, “publication”, and back-up of archive

- 1 Tbyte hard disks used to transfer data (Tony Hoang)
- Sample data checked for compliance (Charles Doutriaux, Karl Taylor)
- Data organized and “published” (Bob Drach, Bob Drach, Bob Drach, …)
- Data back-up (Dean Williams)
- Coordination, oversight, worry (Dave Bader, Karl Taylor, Dean Williams)

Credit: Hardware and systems support

- 35 Tbyte server
- 15 1-Tbyte portable disks
- Jenny Aquilino and Tony Hoang

Credit: those enabling browsing access to output

- Earth System Grid (data portal and search capability): Luca Cinquini, Rob Markel, Bob Drach, Dean Williams

  (See https://esg.llnl.gov:8443/index.jsp and http://www.earthsystemgrid.org)

- CDAT analysis and display software: Dean Williams, Bob Drach, Charles Doutriaux.

  (See http://www-pcmdi.llnl.gov/software/about_software.php)

- Web pages: Steve Davis, Karl Taylor, Dean Williams, Bob Drach.

- Registration of users: Curt Covey, Bob Drach

Credit: others who have contributed in many ways

- Model documentation
  - → Soon to be posted responses to a questionnaire put together by Ron Stouffer, Karl Taylor, Tom Phillips, and various AR4 authors.
  - → Thanks to scientists at each modeling center.
- Advice on a wide variety of scientific and technical issues: Jonathan Gregory and Jim Boyle

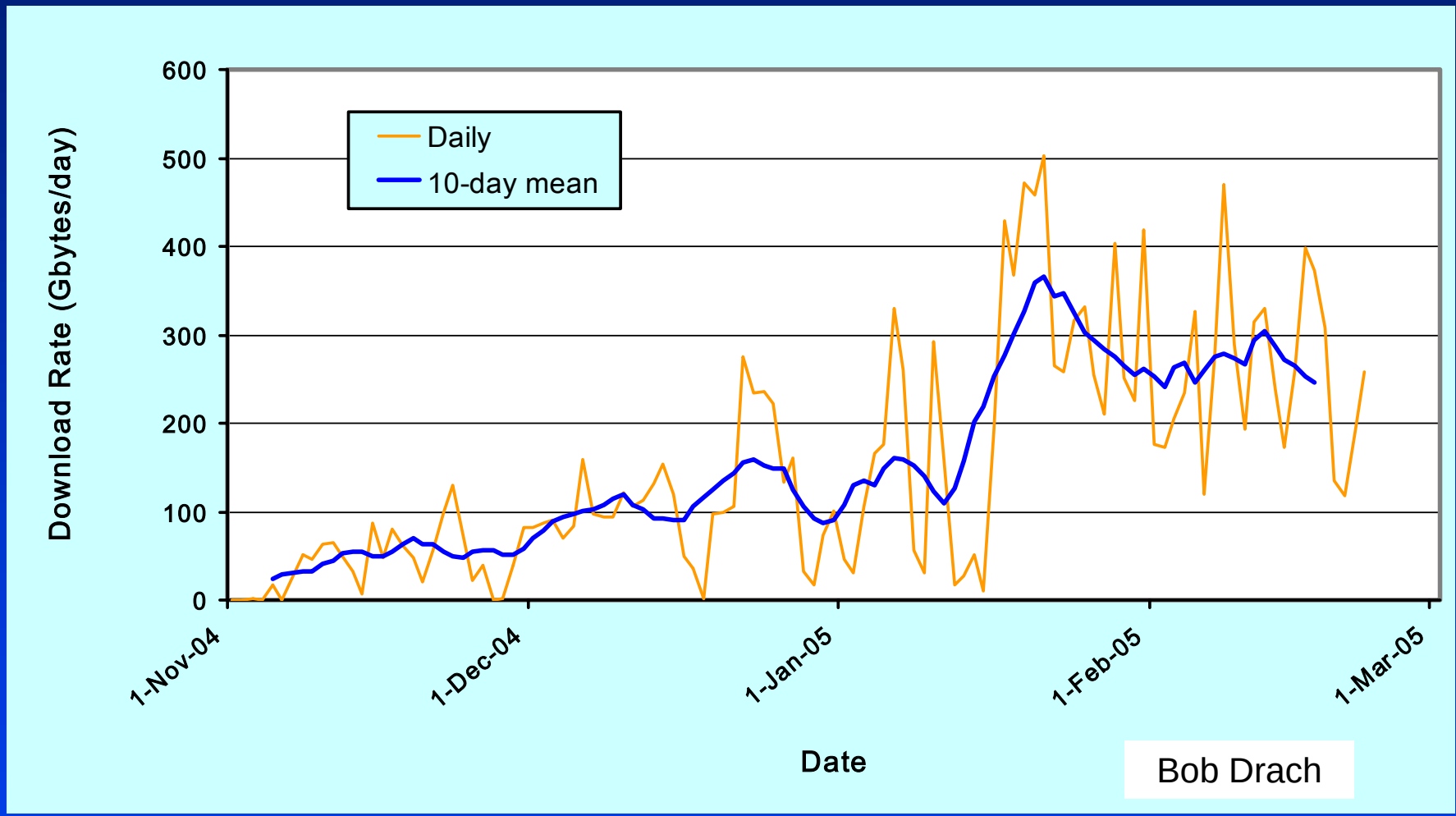
Our obligation

- Appreciate the generous contributions of all these individuals
- Reap the benefits
- Publish great science

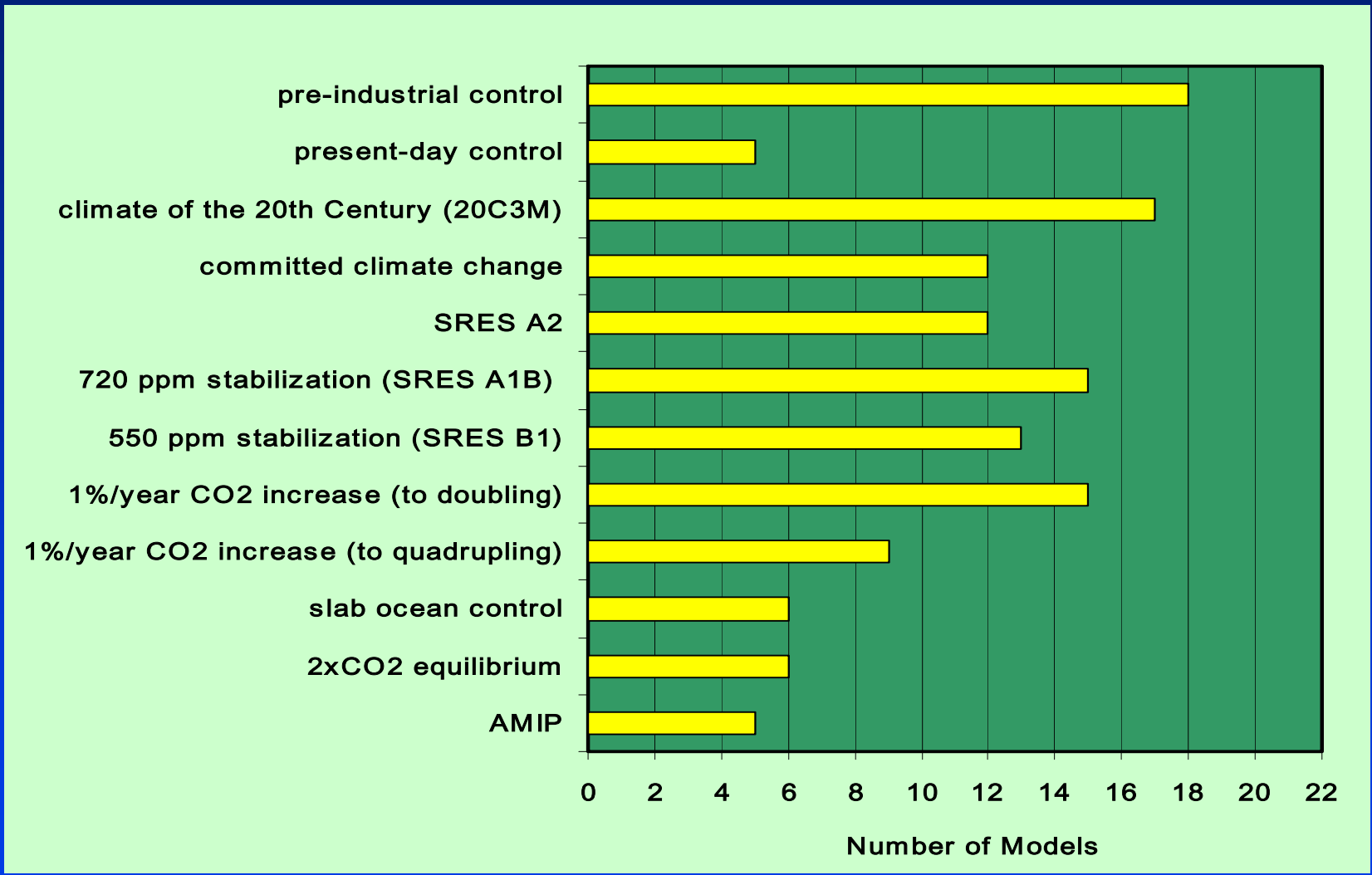
Current archive statistics (2/22/05)

- Archive size: 14.3 Tbytes & 35,000 files
- Data distributed to users: 17.4 Tbytes & 108,000 files
- Registered users: 266
- Download rate now averaging nearly 300 Gbytes/day

Download rate



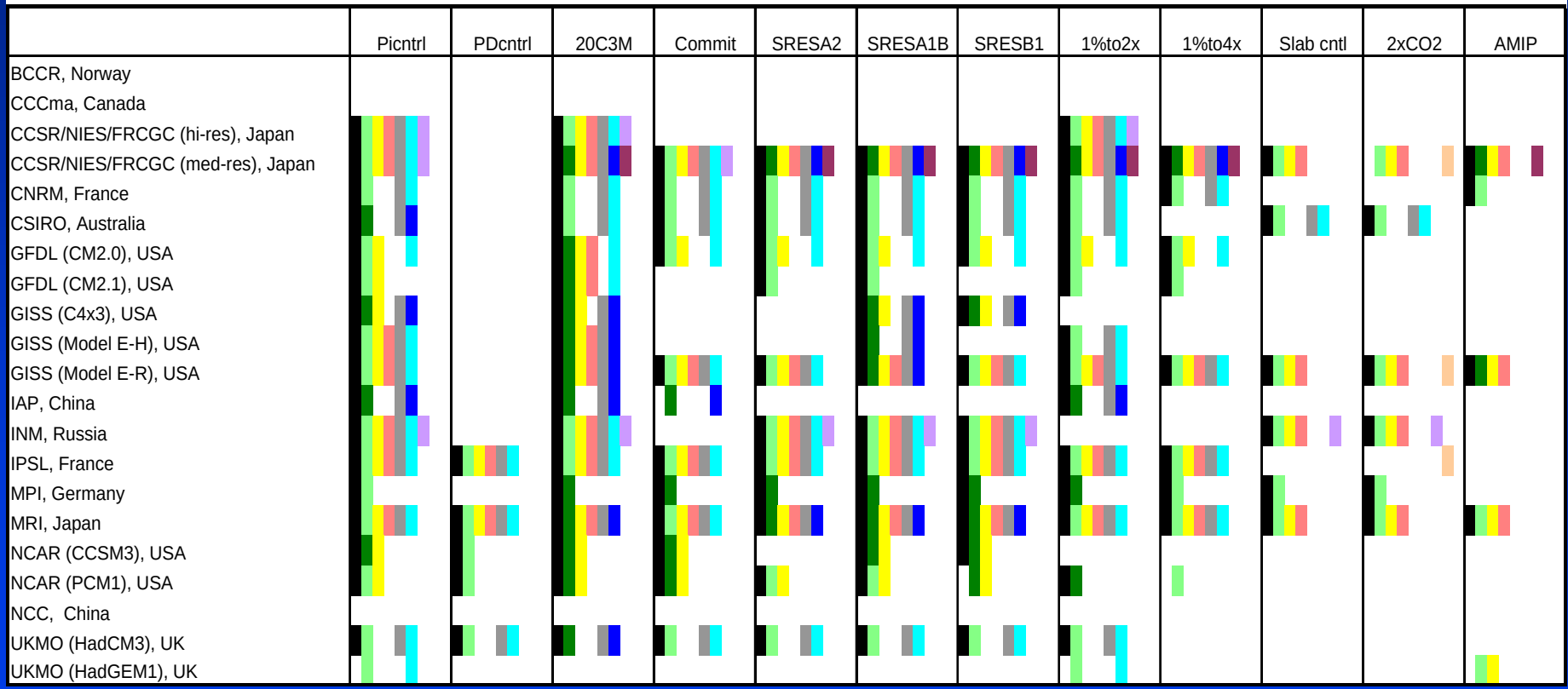
Number of models contributing data for each experiment



Data availability summary (as of 23 February 2005)

shaded area indicates that at least some but not necessarily all fields are available for data type indicated

time-independent land surface
monthly-mean atmosphere
daily-mean atmosphere
3-hourly atmosphere
time-independent ocean
monthly-mean ocean
Extreme Indices
Forcing
ISCCP Simulator



http://www-pcmdi.llnl.gov/ipcc/data_status_tables.htm

Coming soon:

- Expect the archive to double in size.
- Model documentation summaries will be accessible from the web site.
- Table indicating times in control run from which individual perturbation experiments were spawned.

What can we do to improve the archive and help the user?

- Unique and uniformly constructed file names indicating variable, experiment, and model.
- A searchable errata page (suggestions?)
- For those choosing not to use ftp, improvements in the method of transferring data via the “portal” interface.
- Provide list of “official” model names
- Suggestions from you, the user community?

What could be done to make archive easier to use?

Modeling Center	Model
BCCR, Norway	BCM 2.0
CCCma, Canada	CGCM3.1
CCSR/NIES/FRCGC (hi-res), Japan	MIROC3.2 (hi-res)
CCSR/NIES/FRCGC (med-res), Japan	MIROC3.2 (med-res)
CNRM, France	CNRM-CM3
CSIRO, Australia	CSIRO Mk3.0
GFDL (CM2.0), USA	GFDL_CM2.0
GFDL (CM2.1), USA	GFDL_CM2.1
GISS (C4x3), USA	C4x3
GISS (Model E-H), USA	Model E-HYCOM
GISS (Model E-R), USA	Model E-Russell

Modeling Center	Model
IAP, China	FGOALS1.0_g
INM, Russia	INMCM3.0
IPSL, France	IPSL-CM4
MPI, Germany	ECHAM 5 / MPI-OM
MRI, Japan	MRI-CGCM2.3.2a
NCAR (CCSM3), USA	CCSM3.0
NCAR (PCM1), USA	PCM1
NCC, China	CSM T63 (Temporal)
UKMO (HadCM3), UK	HadCM3
UKMO (HadGEM1), UK	HadGEM1

Warning: the archive is not static

- Withdrawal and replacement of various files is ongoing, as problems are uncovered.
- Visit the summary errata page: https://esg.llnl.gov:8443/about/errata.do
- Visit modeling center websites.

Sample errata entries (selected from 34)

Date	Model	Files	Description	Status
12/09/04	giss_model_e_r	All SRESA2 monthly files	The start time should be 2004-1-1, not 1901-1-1. The approximate time range is 2004 to 2100.	1/3/05: Updated files available.
1/3/05	miroc3_2_hires	/ipcc/1pctto2x/atm/yr/gsl/miroc3_2_hires/run1/gsl_A4.nc /ipcc/20c3m/atm/yr/gsl/miroc3_2_hires/run1/gsl_A4.nc /ipcc/picntrl/atm/yr/gsl/miroc3_2_hires/run1/gsl_A4.nc	Growing season length is wrong.	1/10/05: New files available.
1/3/05	mri_cgcm2_3_2a	/ipcc/*/atm/mo/<var>/mri_cgcm2_3_2a/*/<var>_A1*.nc where <var> = cl, clt, rldscs, rlutcs, rsdscs, rsuscs, rsutcs	Data withdrawn.	Files withdrawn. Awaiting replacement.
2/9/05	cnrm_cm3	All scenarios. Variables: tos, sic, sit, usi, vsi, wfo, stfbarot, zobt, so, thetao, rhopoto, uo, vo, wo, zmlo, sbl, hfsib, sltfsib	Land-sea mask is wrong (land area is too large).	Files withdrawn. Awaiting replacement
2/9/05	ukmo_hadcm3	Variable: ts	Data provided only on the ocean grid (N144).	2/9/05: Replacement data available on the atmosphere grid over land and ocean.
2/23/05	gfdl_cm2_0	Variable: sftlf (land area fraction)	Ocean cells had missing data flag (1.e20) instead of 0 values.	Replacement data available.
2/23/05	miroc3_2_medres	Variables: zosga, zostoga	Units should be '100m' ,not 'm'.	Files withdrawn. Awaiting replacement.

https://esg.llnl.gov:8443/about/errata.do

Bob Drach

Summary

- We owe a debt of appreciation to those who have made our research possible.
- We have a tremendous opportunity to “mine” this database:
  - → for shiny nuggets that we can contribute to the IPCC AR4, and
  - → for fruitful veins that will continue to enrich our understanding as we dig deeper in coming months and years.
- Reminders:
  - → Don’t forget to check the errata page!
  - → Please include proper acknowledgement (Jerry’s email)


PCMDI