Climate Modeling Research in the Era of MIPs and PCMDI

Peter J. Gleckler and Karl E. Taylor
Program for Climate Model Diagnosis and Intercomparison (PCMDI)
LLNL

I. A brief history of PCMDI and model intercomparison
II. The Coupled Model Intercomparison Project (CMIP3 & CMIP5)
III. Emergence of climate model performance metrics
IV. Expanding the use of NASA products for climate model evaluation

CERES Science Team Meeting
October 4-6, 2011

LLNL/PCMDI's dual mission in climate research

- **Research**
  - Climate model evaluation
  - Cloud process research
  - Uncertainty quantification
  - Climate change detection and attribution
  - Atmospheric chemistry, aerosols, and earth system modeling

- **Enable & facilitate research by others**
  - Coordinate climate modeling activities worldwide
  - Make available model output from simulations of high interest
  - Provide summaries of model results relied on by the IPCC
  - Provide a multi-model perspective on model projections
  - Diagnostic and data access capabilities

History: Before the dawn of the MIP's

- In the 1970s and 1980s, the evaluation of climate models was largely a qualitative endeavor (and mostly done by a small group of modelers)
- Often involved purely visual comparison of selected "maps" from a model simulation and observations, with similarities and discrepancies noted.
- No standard benchmark experiments
- Little community involvement in model diagnosis
- Difficult to track changes in model performance over time

History: Establishment of the first MIP

- 1980's: MIP precursors – FANGIO, radiation code intercomparison

- ca. 1991: The Atmospheric Model Intercomparison Project (AMIP), following inception of PCMDI
  - Championed by PCMDI and encouraged/endorsed by the WCRP's Working Group on Numerical Modelling
  - Modeling groups were initially reluctant to share results
  - Roughly 30 modeling groups from 10 different countries
  - Community involvement for the first time in experimental design (10 yrs of prescribed SST and sea-ice) and diagnosis

- ca. 1995: AMIP2 – tighter experimental protocol, more extensive diagnostics

History: From atmosphere-only to coupled models

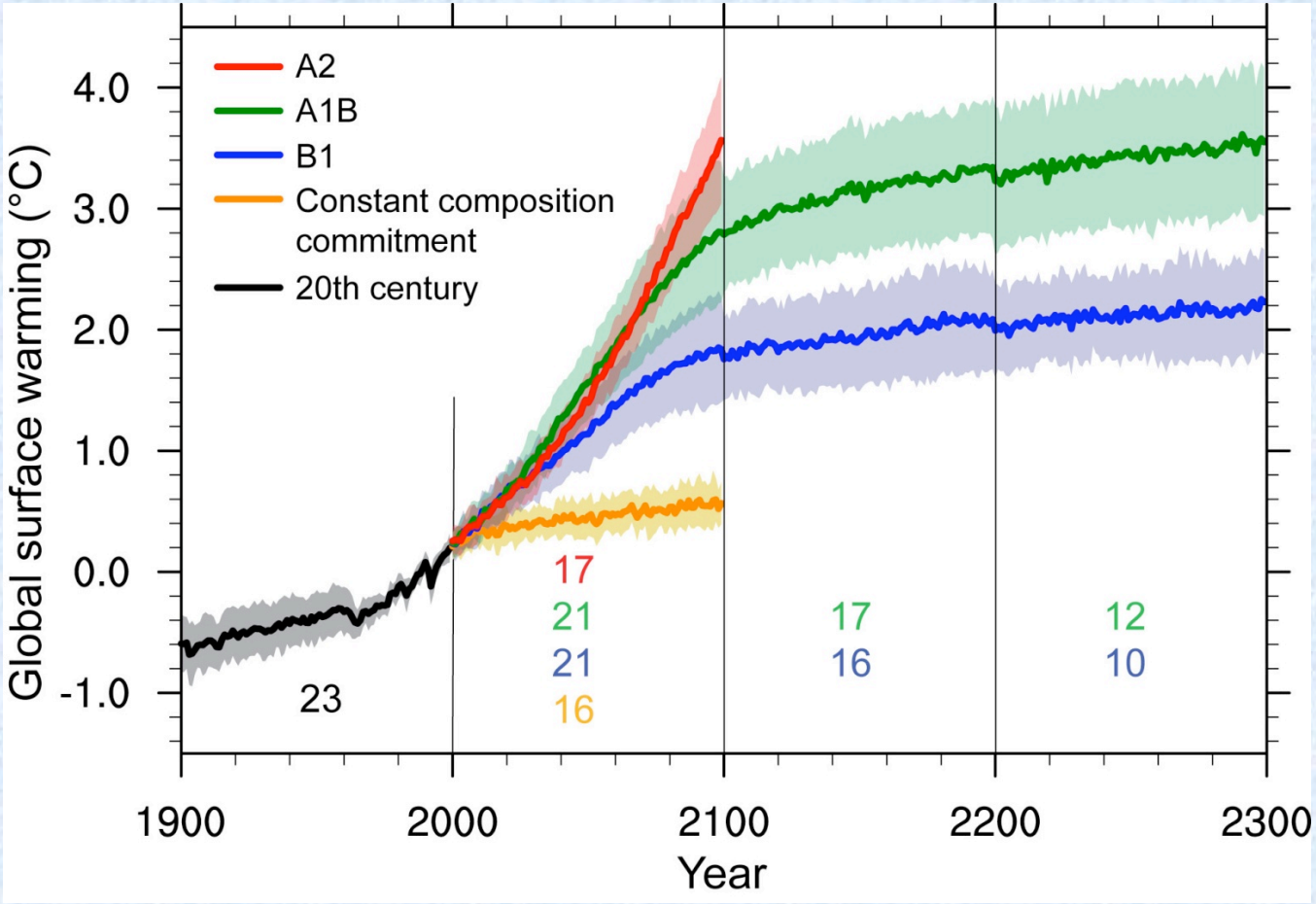
- CMIP1 (ca. 1995): control run
- CMIP2 (ca. 1997): 1%/year CO_2 increase (idealized climate change) ~Gigabytes
- **CMIP3** (2003 – ca. 2013): ~30 Terabytes
  - Expts: control, idealized, historical, and SRES (future scenario) runs
  - Output largely available by 2005
- [CMIP4 (ca. 2007): "single forcing" experiments for detection/attribution studies]
- **CMIP5** (2006 – beyond 2016; ongoing and revisited) ~3 Petabytes (estimated)
  - An ambitious variety of "realistic" and diagnostic experiments
  - Output largely available by 2012

Nearly all the new, model-based conclusions in the IPCC AR4 rely upon analyses of CMIP3

- ~75% of the more than 100 figures in AR4 Chapters 8-11 are based on CMIP3
- 4/7 “Summary for Policy Makers” figures are based on CMIP3
- AR4 conclusions are more robust because they are based on
  - the CMIP3 multi-model ensemble

  rather than on
  - “anecdotal” conclusions from individual modeling studies

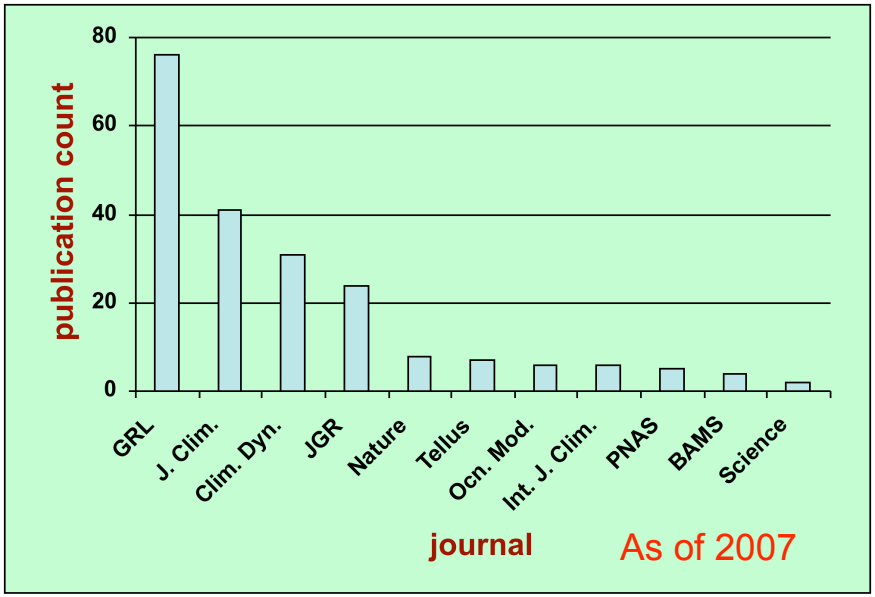
AR4 Chapter 10 (global projections): Future scenario simulations from CMIP3 provide a range of projections of climate change



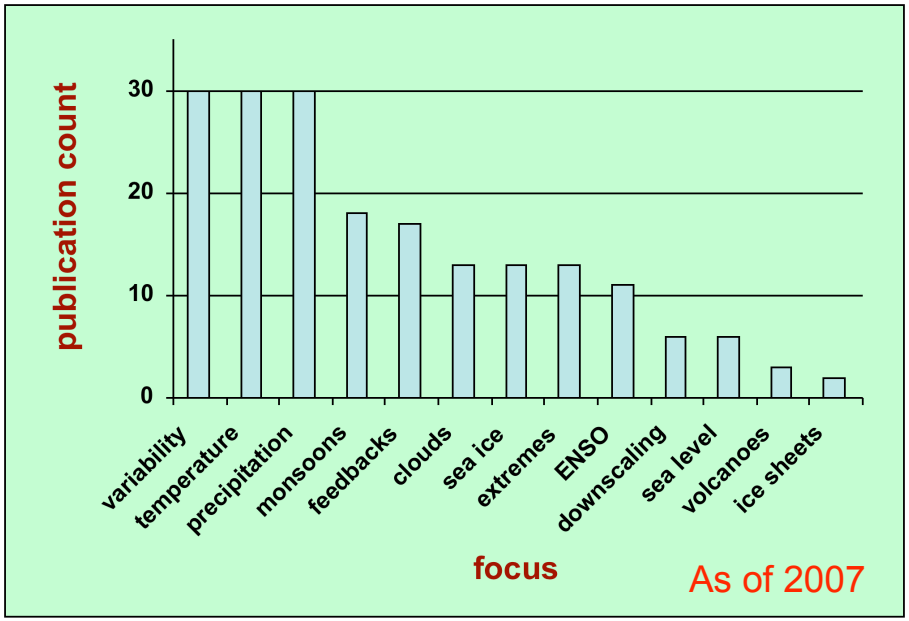
[From Summary for Policy Makers]

CMIP3 Research: a few statistics

published in
a variety of journals
> 500 publications



covering
a wide range of topics



What made the difference in CMIP3?

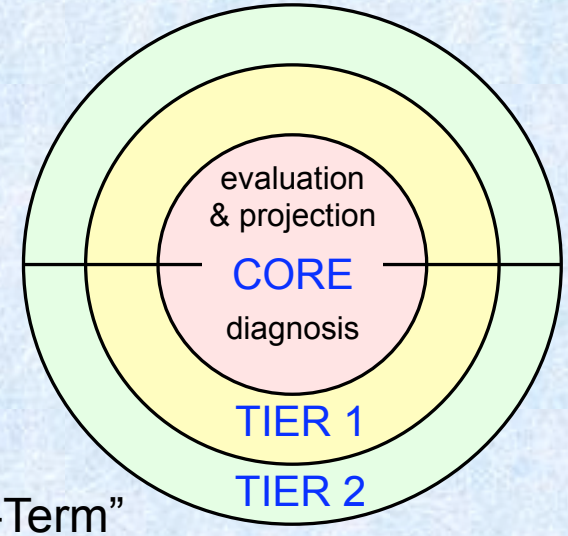
An investment in experimental design, infrastructure and development of standards

- Community-developed metadata conventions
  - The “Climate-Forecast” metadata convention (CF)
- Software to ensure data complies to conventions
  - The Climate Model Output Writer (CMOR)
- State-of-the-art data delivery methods
  - The Earth System Grid (ESG)

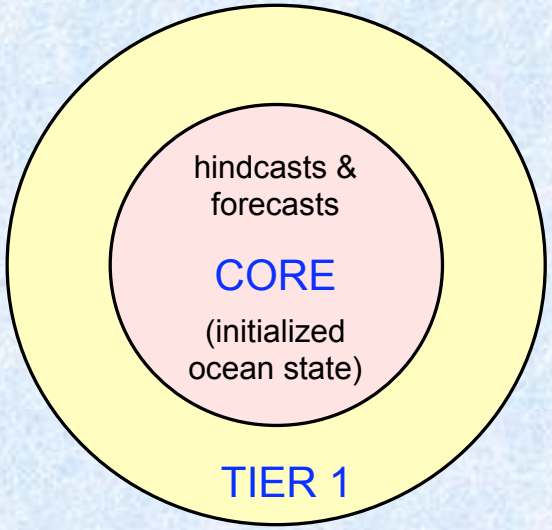
Where does CMIP come from?

- CMIP is overseen by the Working Group on Coupled Modeling (WGCM) which is jointly sponsored by the WCRP and CLIVAR
- WGCM members include leads from the world's major climate modeling centers
- PCMDI works closely with the WGCM in the design and implementation of CMIP
- The WGCM, PCMDI, and many others have been preparing for CMIP5
    - 2006-2009: Experimental design
    - Ongoing: Modeling centers performing/submitting simulations
    - Research: Just beginning…

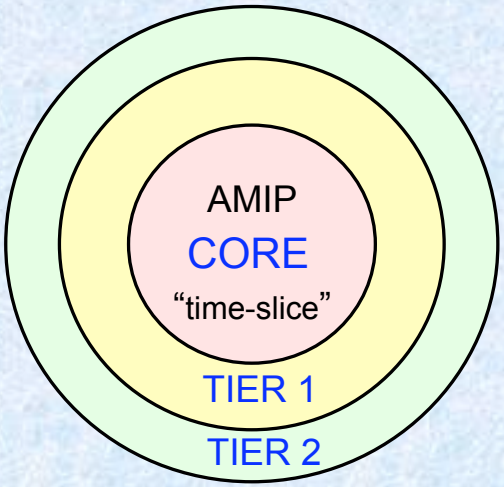
CMIP5 is organized around three types of simulations



"Long-Term" Projections (century & longer)



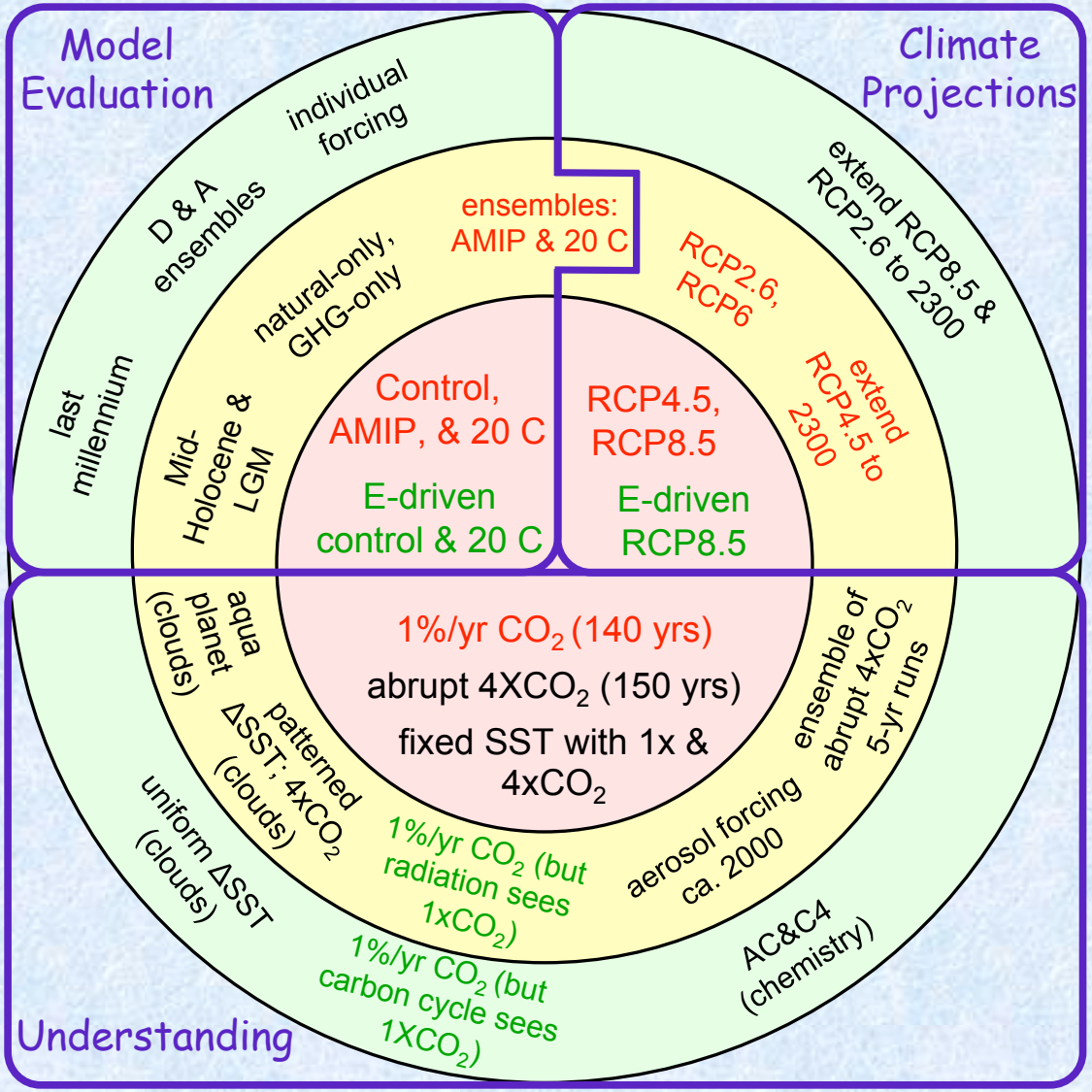
Decadal Climate Prediction



Atmosphere-Only Simulations

(for computationally demanding and NWP models)

The CMIP5 design provides opportunities for evaluation and understanding model behavior, as well as producing projections



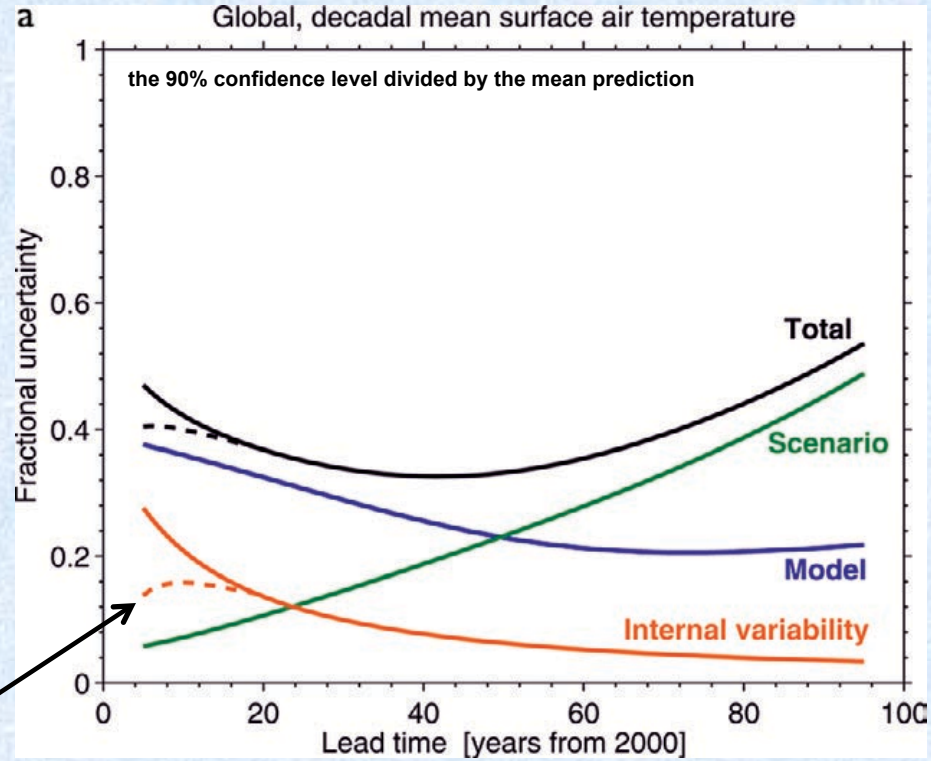
Red subset matches the entire CMIP3 experimental suite

Green subset is for coupled carbon-cycle climate models only

Taylor et al., *BAMS* 2011

CMIP5 will also include models initialized with the observed state (in particular of the upper ocean)

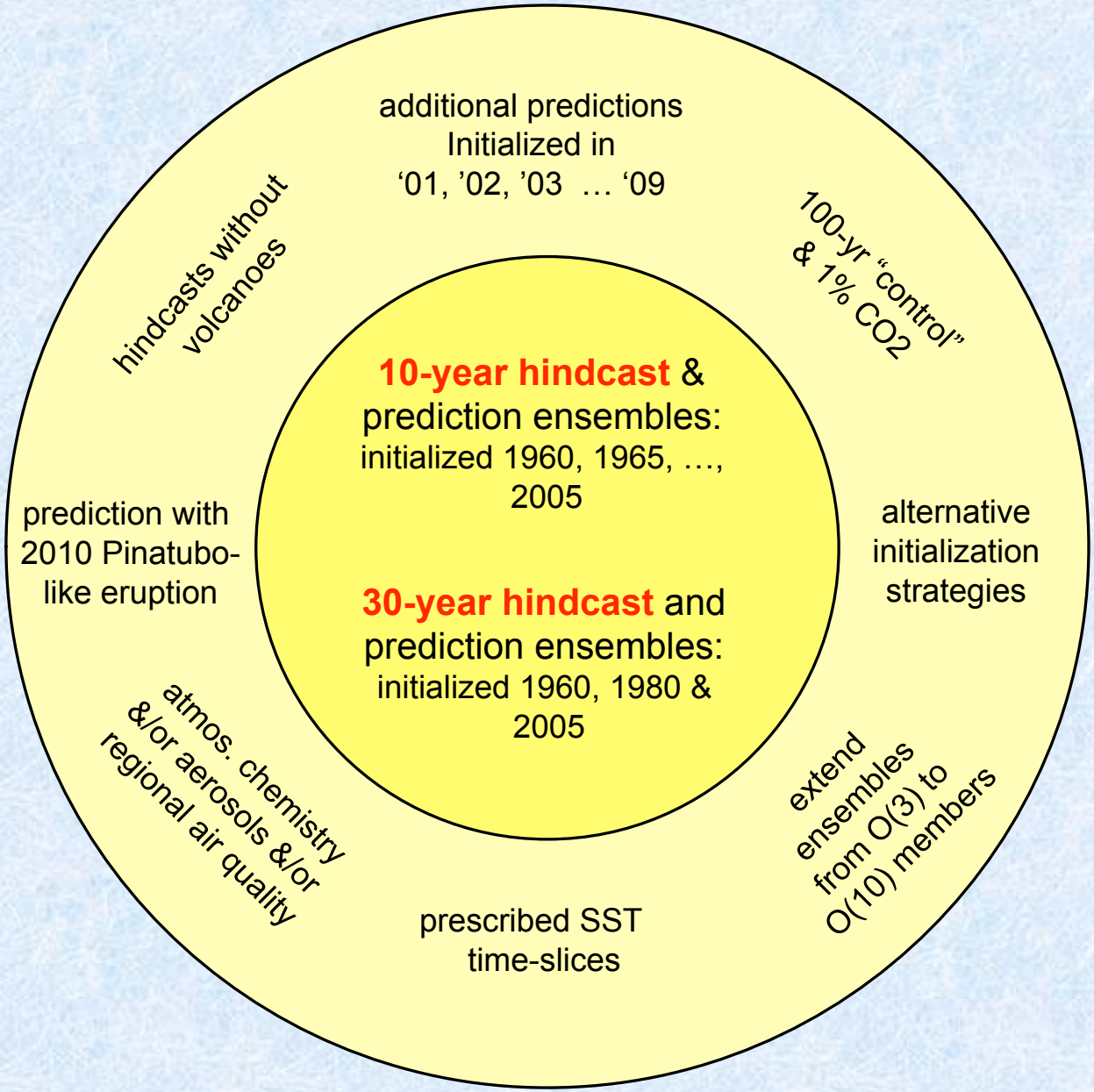
- The hope is that through initialization the models will be able to predict the actual trajectory of “unforced” climate variations.
- The hypothesis is that some longer time-scale natural variability is predictable if the initial state of the system is known



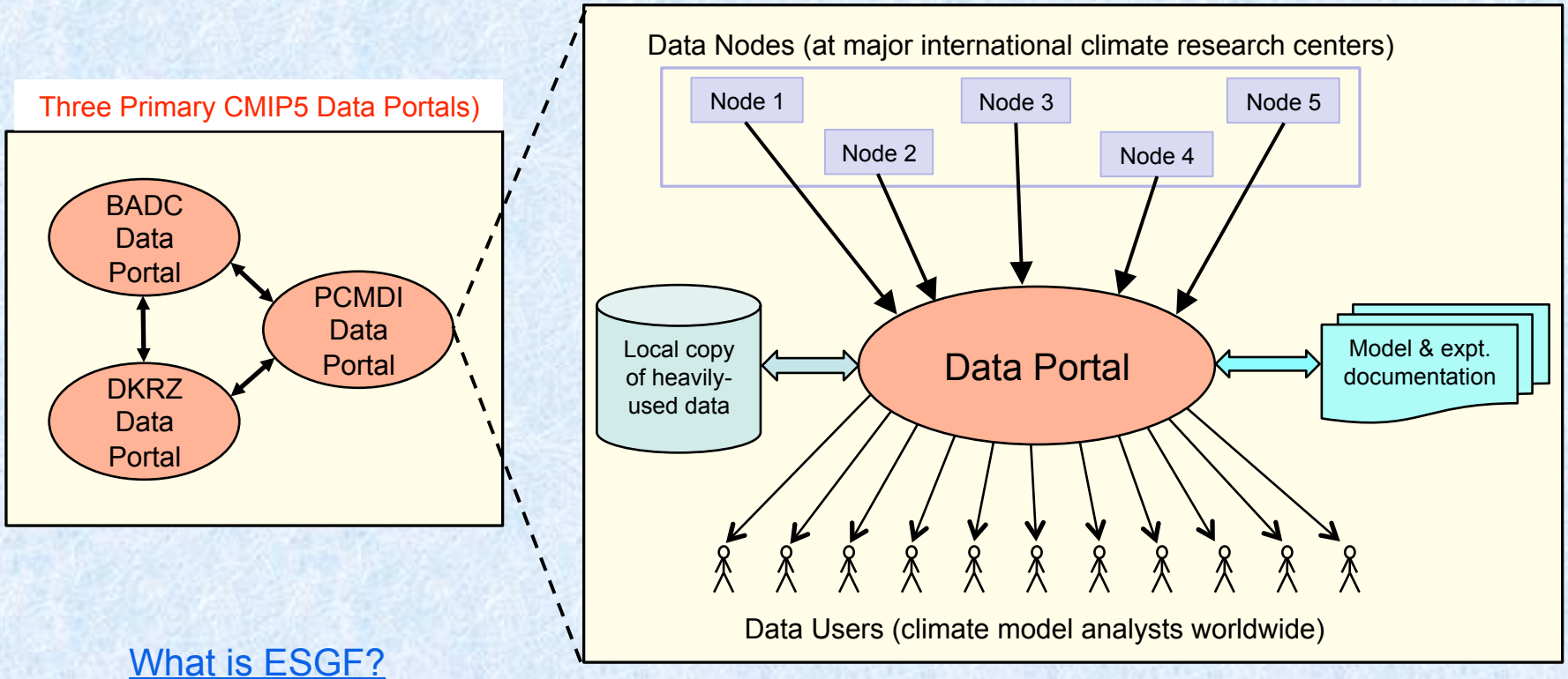
The deviation from observations caused by unforced variability can potentially be reduced through initialization.

Hawkins & Sutton, 2009

CMIP5 Decadal Predictability/Prediction Experiments



LLNL-led Earth System Grid Federation (ESGF) serves climate simulation output to analysts worldwide



What is ESGF?

- ESGF links together all major climate centers and provides access to climate simulations
- Currently expanding from 10's to 1000's of Tbytes
- Serves 1000's of researchers

CMIP5 output fields requested (goes well beyond what was available from CMIP3)

- Domains (number of monthly variables*):
  - Atmosphere (60)
  - Aerosols (77)
  - Ocean (69)
  - Ocean biogechemistry (74)
  - Land surface & carbon cycle (58)
  - Sea ice (38)
  - Land ice (14)
  - CFMIP output (~100)
- Temporal sampling (number of variables*)
  - Climatology (22)
  - Annual (57)
  - Monthly (390)
  - Daily (53)
  - 6-hourly (6)
  - 3-hourly (23)

*Not all variables will be saved for all experiments and time-periods

http://cmip-pcmdi.llnl.gov/cmip5/output_req.html

CMIP5 participating groups (23 groups; 50+ models; 18 Sept 2011: 15 models available from 10 centers)

Primary Group	Country	Model
CAWCR	Australia	ACCESS
BCC	China	BCC-CSM1.1
GCESS	China	BNU-ESM
CCCMA	Canada	CanESM2, CanCM4, CanAM4
CCSM	USA	CESM1, CCSM4
RSMAS	USA	CCSM4(RSMAS)
CMCC	Italy	CMCC- CESM, CM, & CMS
CNRM/CERFACS	France	CNRM-CM5
CSIRO/QCCCE	Australia	CSIRO-Mk3.6
EC-EARTH	Europe	EC-EARTH
LASG, IAP	China	FGOALS- G2.0, S2.0 & gl
FIO	China	FIO-ESM
NASA/GMAO	USA	GEOS-5
GFDL	USA	GFDL- HIRAM-C360, HIRAM-C180, CM2.1, CM3, ESM2G, ESM2M
NASA/GISS	USA	GISS- E2-H, E2-H-CC, E2-R, E2-R-CC, E2CS-H, E2CS-R
MOHC	UK	Had CM3, CM3Q, GEM2-ES, GEM2-A, GEM2-CC
NMR/KMA	Korea / UK	HadGEM2-AO
INM	Russia	INM-CM4
IPSL	France	IPSL- CM5A-LR, CM5A-MR, CM5B
MIROC	Japan	MIROC 5, 4m, 4h, MIROC- ESM, ESM-CHEM
MPI-M	Germany	MPI-ESM- HR, LR
MRI	Japan	MRI- AGCM3.2H, AGCM3.2S, CGCM3, ESM1
NCC	Norway	NorESM1-M, NorESM-ME, NorESM1-L

“Long-term” experiments: planned contributions

** Core simulations* *(# available as of 18 Sept 2011)*

Experiment(s)	# of models
* Control & historical	35 (10)
* AMIP	26 (8)
* RCP4.5 & 8.5	29 (9)
RCP2.6	18 (6)
RCP6	13 (6)
RCP’s to year 2300	10 (?)
* 1% CO2 increase	28 (7)
* Fixed SST CO2 forcing diagnosis	16 (4)
* Abrupt 4XCO2 diagnostic	22 (7)

Experiment(s)	# of models
Fast adjustment diagnostic	9 (?)
Aerosol forcing	9 (2)
*ESM control, historical & RCP8.5	18 (3)
Carbon cycle feedback isolation	9 (2)
Mid-Holocene & LGM	11 (2)
Millenium	7 (0)
CFMIP runs	7-9 (1-4)
D & A runs	15 (6)

Timelines: CMIP5 and the IPCC AR5

- **Present:** Model output database rapidly expanding
- **July 31, 2012:** Papers must be submitted for publication to be eligible for assessment by WG1,
- **March 15, 2013**: Papers cited by WG1 must be published or accepted.
- The IPCC's AR5 is scheduled to be published in **September 2013**.

<u>Like CMIP3/AR4, we expect the bulk of the CMIP5 science will be post AR5</u>

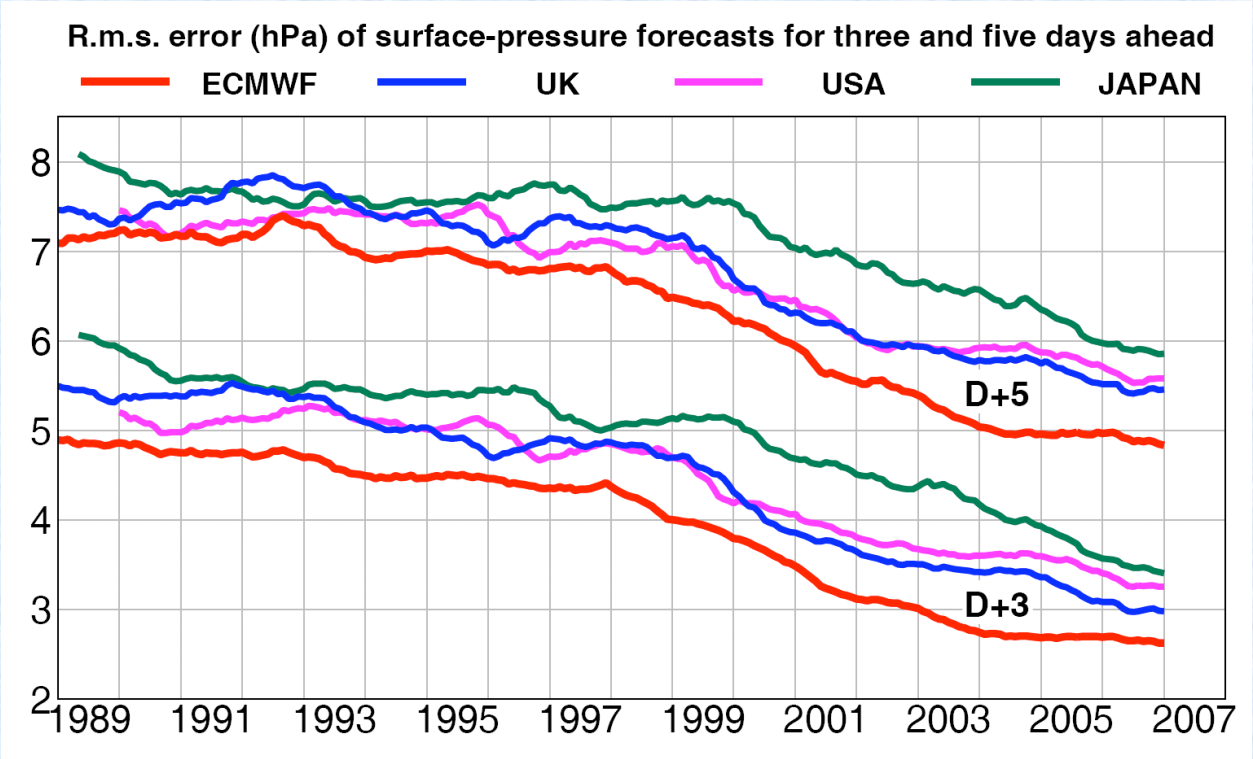
Climate Model Performance Metrics

CMIP establishes some benchmark experiments that allow us to gauge changes in model performance

- AMIP runs (prescribed SST's and sea-ice)
- CMIP control runs (variability characteristics)
- Historical runs (1850 – present)
- Idealized 1%/yr CO_2 increases (determine climate sensitivity)

Monitoring evolution of model performance: Example from Numerical Weather Prediction

- WGNE routinely reviews skill of daily forecasts
- Improvements and deficiencies in the systems identified



The climate modeling community does not yet have well-established benchmarks

Courtesy M.Miller, ECMWF

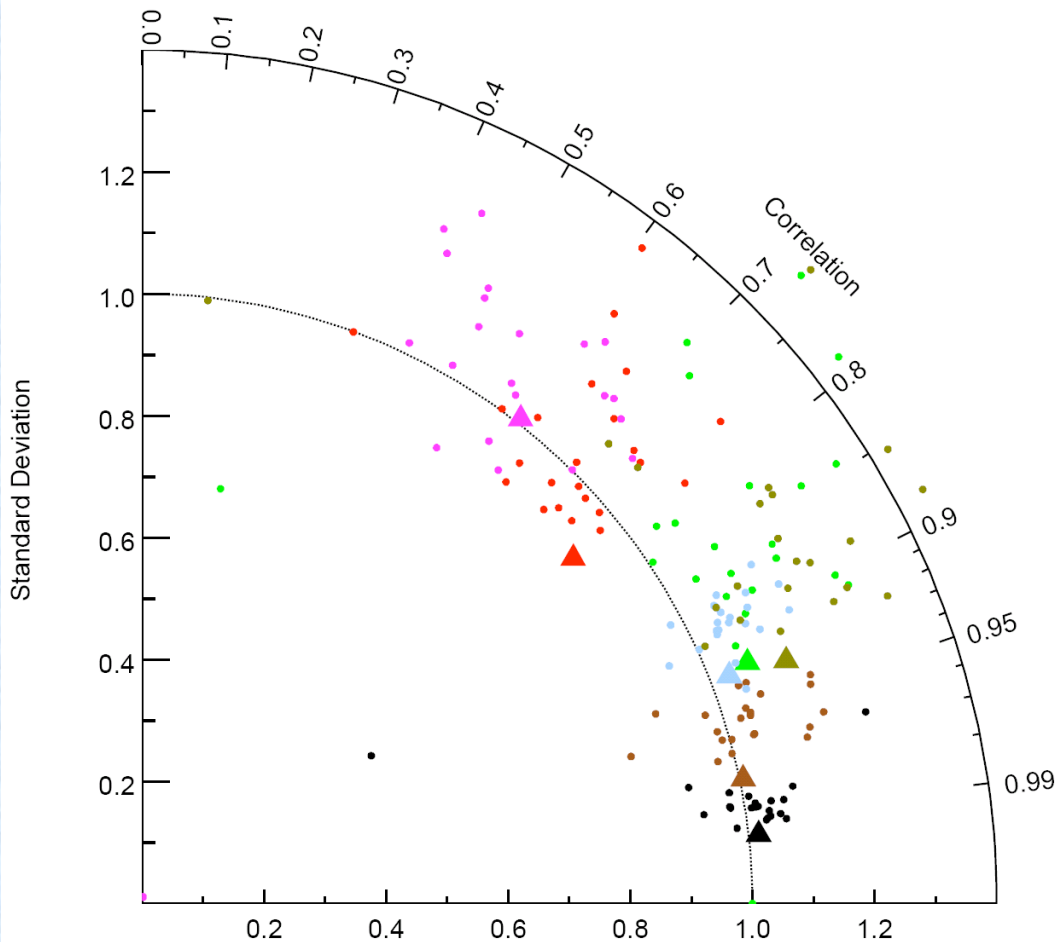
What do we mean by "metrics"?

- Metrics, for our purposes, are scalar quantities that objectively measure the quality of a model simulation, e.g.,
  - Skill in simulating things we have observed ("performance metrics")
  - Model reliability for applications (e.g., "projection reliability metrics")
    - How accurate are model projections of climate change?
    - Extremely valuable… and… extremely difficult
- Quantify errors, but usually *not* designed to diagnose reasons for model errors

What opportunities are there to construct climate model performance metrics?

- Model's externally "forced" responses on a range of time-scales:
  - Diurnal cycle
  - Annual cycle
  - Volcanic eruptions, changes in solar irradiance, …
- Model's "unforced" behavior (weather, MJO, ENSO, NAO, PDO …)
- Evaluate model representation of individual processes and co-variability relationships
- Test model ability to solve the "initial value" problem

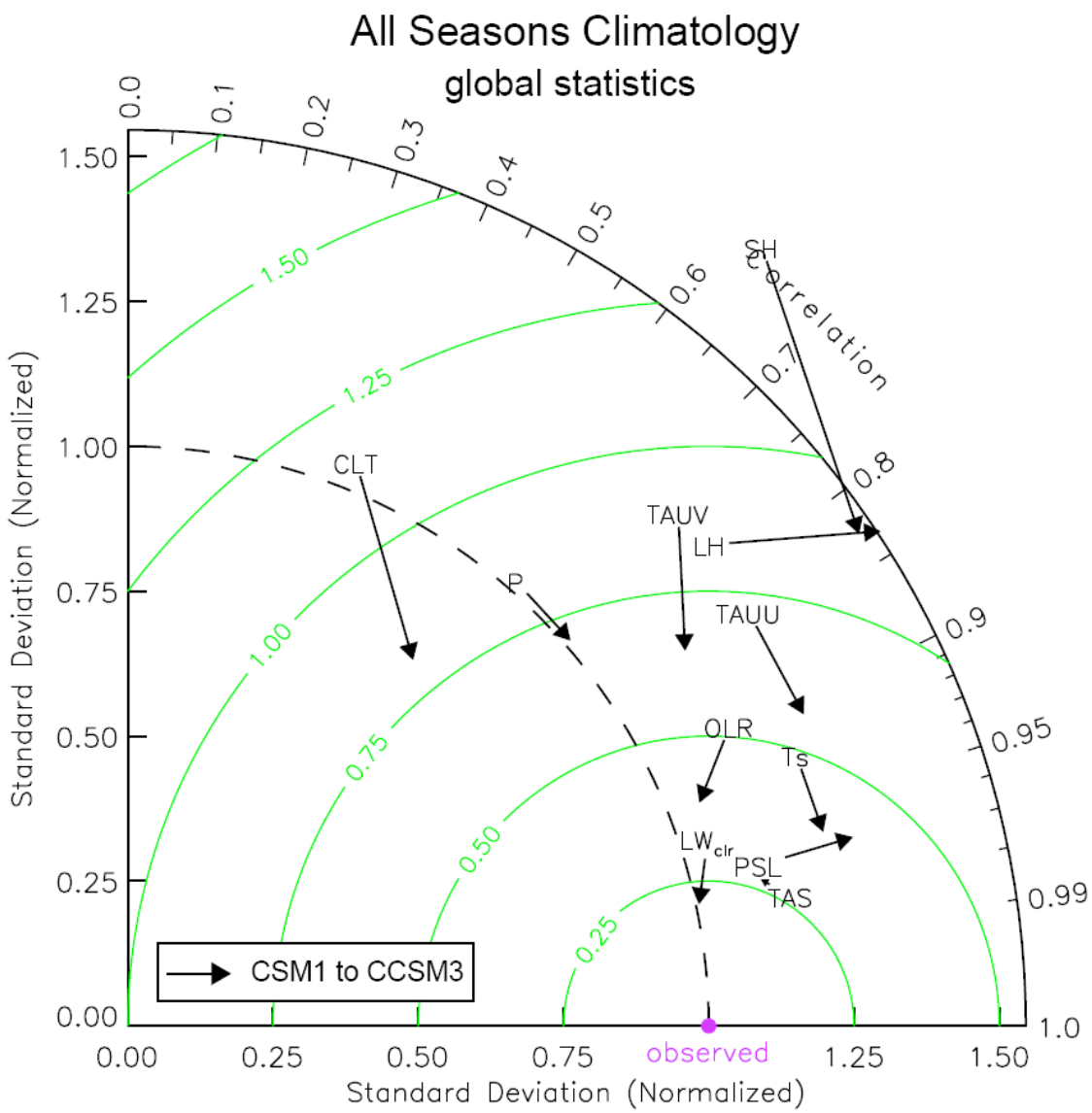
Taylor diagram for CMIP3 annual cycle global climatology (1980-1999)



Sea Level Pressure: ERA40 reference
Total precipitation rate: CMAP reference
Total Cloud Cover: ISCCP reference
LW radiation TOA (OLR): CERES reference
Reflected TOA Shortwave: ERBE reference
Air Temperature (850 hPa): ERA40 reference
Zonal Wind (850 hPa): ERA40 reference

- Variable dependent skill
- Multi-model mean "superiority"

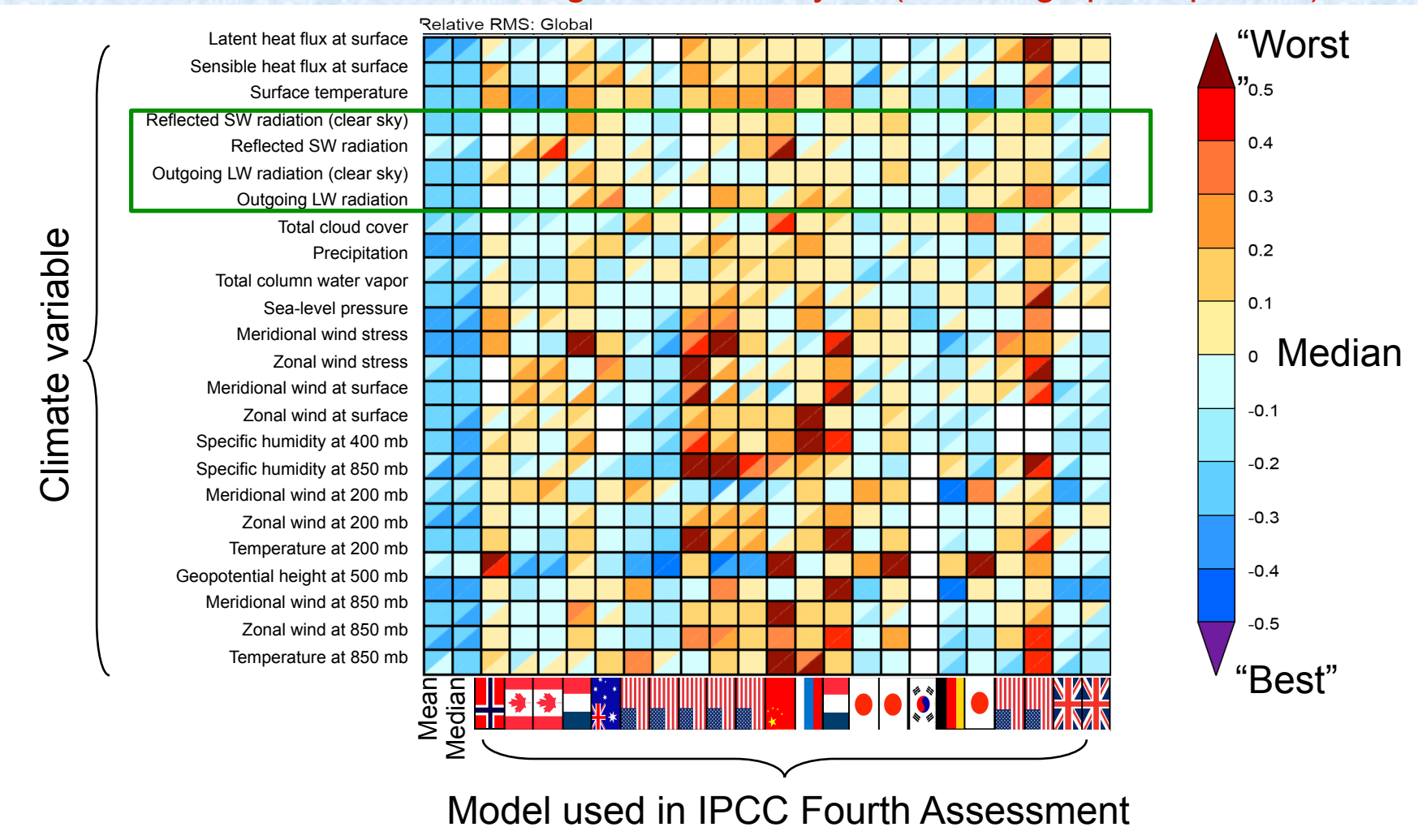
Tracking model performance in the development process



Comparing different model versions

Evaluating how well climate models simulate the annual cycle: A "Performance Portrait" of relative errors

Relative RMSE in Climatological Annual Cycle (including spatial pattern)



Gleckler, P, K. Taylor and C. Doutriaux, *J.Geophys.Res.* (2008)

An update of the
WGNE/WGCM* Climate Model Metrics Panel

Members appointed based on relevant and diverse areas of expertise, and potential to liaison with key WCRP activities:

Beth Ebert (BMRC) – JWGV/WWRP, WMO forecast metrics

Veronika Eyring (DLR Germany) – WGCM/SPARC, stratosphere

Pierre Friedlingstein (U. Exeter) – IGBP, carbon cycle

Peter Gleckler (PCMDI), chair – WGNE, atmosphere

Robert Pincus (NOAA) – GEWEX/GCSS, clouds/radiation

Karl Taylor (PCMDI) – WGCM, CMIP5

Helene Hewitt (U.K. Met Office) – WGOMD, ocean and sea-ice

* WGNE – Working Group on Numerical Experimentation
WGCM – Working Group on Coupled Modeling

Questions motivating routine benchmarks for climate models

- Of direct concern to the WGNE/WGCM metrics panel:
  - Are models improving?
  - Are some models more realistic than others?
  - What do models simulate robustly, and what not?
- Related research drivers, but not (currently) the panel's focus:
  - How does skill in simulating observed climate relate to projection credibility?
  - Can we justify weighting model projections based on metrics of skill?

Metrics panel terms of reference

- **Identify a limited set of basic climate model performance metrics**
  - based on comparison with observations
  - well established in literature
  - easy to calculate, reproduce and interpret
  - covering a diverse suite of climate characteristics
    - large- to global-scale mean climate and variability
    - atmosphere, oceans, land surface, and sea-ice

Metrics panel terms of reference

- **Identify a limited set of basic climate model performance metrics**
  - based on comparison with observations
  - well established in literature
  - easy to calculate, reproduce and interpret
  - covering a diverse suite of climate characteristics
    - large- to global-scale mean climate and variability
    - atmosphere, oceans, land surface, and sea-ice

- **Justify and promote these basic metrics in an attempt to**
  - establish routine community benchmarks
  - facilitate further research of increasingly targeted metrics

Metrics panel terms of reference

- **Identify a limited set of basic climate model performance metrics**
  - based on comparison with observations
  - well established in literature
  - easy to calculate, reproduce and interpret
  - covering a diverse suite of climate characteristics
    - large- to global-scale mean climate and variability
    - atmosphere, oceans, land surface, and sea-ice

- **Justify and promote these basic metrics in an attempt to**
  - establish routine community benchmarks
  - facilitate further research of increasingly targeted metrics

- **Coordinate with other WCRP/CLIVAR working groups**
  - Identify metrics for more focused evaluation (e.g. modes of variability)
  - Striving towards a community-based activity by engaging expertise from a spectrum of perspectives

Metrics panel terms of reference

- **Identify a limited set of basic climate model performance metrics**
  - based on comparison with observations
  - well established in literature
  - easy to calculate, reproduce and interpret
  - covering a diverse suite of climate characteristics
    - large- to global-scale mean climate and variability
    - atmosphere, oceans, land surface, and sea-ice

- **Justify and promote these basic metrics in an attempt to**
  - establish routine community benchmarks
  - facilitate further research of increasingly targeted metrics

- **Coordinate with other WCRP/CLIVAR working groups**
  - Identify metrics for more focused evaluation (e.g. modes of variability)
  - Striving towards a community-based activity by engaging expertise from a spectrum of perspectives

- **Ensure that these metrics are applied in CMIP5 and widely available**

Current status: Focus is on a limited set of metrics to be periodically reviewed and augmented

Climatological annual cycle:

- 15-20 large- to global- scale statistical or “broad-brush” metrics
- Domains: Global, tropical, NH/SH extra-tropics
- 20 year climatologies: Annual mean, 4 seasons
- Routine metrics: bias, centered RMS, MAE, correlation, standard deviation
- Field examples: OLR, T850, q, SST, SSH, sea-ice extent
- Observations: multiple for most cases

Extended set of metrics, coordinating (in progress) with other working groups:

- ENSO (CLIVAR Pacific Panel)
- MJO (YOTC Task force)
- Monsoons (CLIVAR AAMP)
- Carbon cycle in emission-driven ESMs (ILAMB)
- Coordination with other working groups is planned…

  (e.g., GCSS/CFMIP and WGOMD)

IV.
Expanding the use of NASA products for climate model evaluation

Revisiting the PCMDI/NASA October 2010 (same viewgraph)

- NASA data products are invaluable for climate model evaluation/research
- NASA DAACs provides a wealth of information and data
- Many potential non-expert model evaluation users find this resource overwhelming, are unsure how to proceed, and potentially bypass using the data – many “don’t have the time” to invest
- CMIP5 is going to be a very visible and heavily utilized resource for at least the next 5 years
- Many of us here believe that there is a ripe opportunity to coordinate relevant NASA data products with CMIP5, and that this could greatly enhance the use and usefulness of these products for climate model evaluation/research

Like CMIP5, “Obs4MIPs” is a now an ESG Project

- Conceived at PCMDI/NASA October 2010 meeting
- NASA and PCMDI are taking the lead on improving how observations are made available specifically for the purpose of climate model evaluation
- Obs4MIPs is limited to data that can be quantitatively compared to model output
- Once a dataset has been chosen for Obs4MIPs, the following is needed:
  - Expert judgment – selecting a version for model evaluation (with alternates)
  - Technical alignment with CMIP5 (via conventions/format/ESG, quality control)
  - Documentation tailored for model evaluation/research, highlighting:
    - Measurement origins
    - Sampling and uncertainty characteristics
    - Traceability of any data product updates

Obs4MIPs to be limited to products that can be directly compared to CMIP5 model output

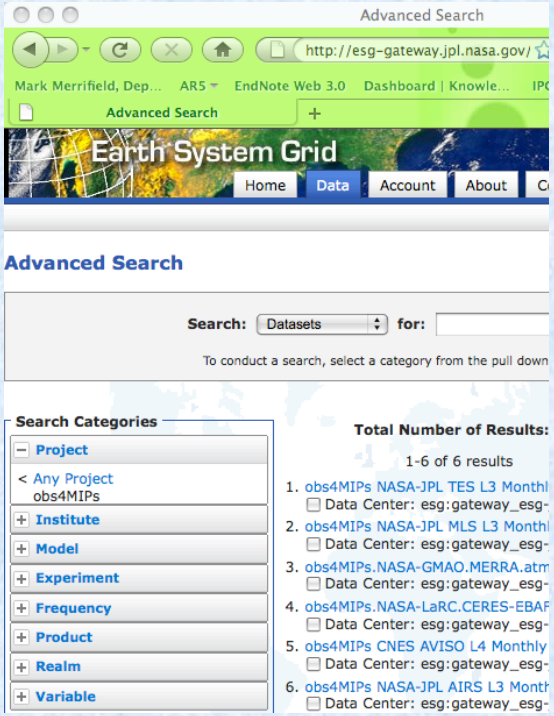
Some baseline model output examples used for performance metrics:

- Temperature (200,850hPa)
- Zonal and meridional wind (200,850 hPa)
- Specific humidity (200, 850 hPa)
- Surface (10m) zonal and meridional wind
- Ocean surface zonal and meridional wind stress
- Sea surface temperature
- TOA reflected shortwave radiation and OLR
- TOA longwave and shortwave TOA clear-sky fluxes
- Total precipitation
- Cloud cover
- Precipitable water
- Sea surface height
- Sea ice

We are excited about the collaboration!

CERES EBAF Ed2.6 is now accessible via ESG

Thank you!



Other NASA datasets now available on ESG:

- AIRS (temperature, specific humidity)
- AVISO (sea surface height)
- MLS (temperature)
- TES (ozone)
- MERRA (reanalysis)