Some potential use cases of IPCC AR5 datasets

Third ESC Workshop, Princeton, NJ

V. Balaji

Princeton University

NOAA/GFDL

18 October 2007

Talk outline...

1. From IPCC AR4 to AR5
   - Convergence of models and data
2. AR5 use cases
   - Querying model characteristics
   - Regridding
   - Offline models
   - Branch runs
3. Summary

Talk outline . . .

1. From IPCC AR4 to AR5
   - Convergence of models and data
2. AR5 use cases
   - Querying model characteristics
   - Regridding
   - Offline models
   - Branch runs
3. Summary

Linking model and data frameworks

Community data frameworks (e.g ESG, an ESC partner) are under development, at various institutions, informally linked by the GO-ESSP. For model output data to be scientifically useful, the researcher must have some knowledge of how the data was produced. Model data requires a *model's eye view* description of the data, another layer of metadata, which might include:

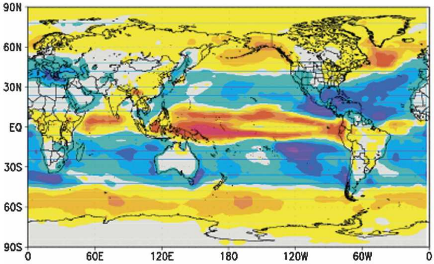
- Description of model components: e.g GEOS-5 atmosphere, land and sea ice coupled to MIT ocean.
- Description of grid configurations and resolutions.
- Choice of physics packages and input parameters.
- Model state and its fields.

ESMF and PRISM are emerging standards that allow the development of the model metadata layer, based on the state data structures and its base classes. (Think `State`, `Grid`, `LocStream`, `...`)

Semantic vs. syntactic, discovery vs. use

Descriptive metadata can be succinct, and can be used to discover certain aspects of the data. But almost any serious use requires deeper knowledge. The boundary between *discovery* and *use*, *semantic* and *syntactic*, is blurred by the use of controlled vocabularies and ontologies.

Graphics such as this from Held and Soden (2006) are so routinely produced from the IPCC AR4 database that we've ceased to marvel at it. This is a composite of output from 20 models worldwide, run with minimal coordination.



Talk outline . . .

1. From IPCC AR4 to AR5
   - Convergence of models and data
2. AR5 use cases
   - Querying model characteristics
   - Regridding
   - Offline models
   - Branch runs
3. Summary

Can the database answer these questions?

- What's the difference between the NASA GISS-EH and GISS-ER models? *(Answer: the ocean component).* (Russell et al 2006).
- Which runs from the GFDL CM2.1 model would I compare to isolate the effects of volcanoes on 20th century climate? (Stenchikov et al 2006).
- Do volcano runs from GFDL CM2.1 and CCSM use the same forcing dataset?
- Which runs in the database include the *indirect effect of aerosols*?
- Retrieve "high cloud amount" from multiple models.

Dynamically derived data catalogues

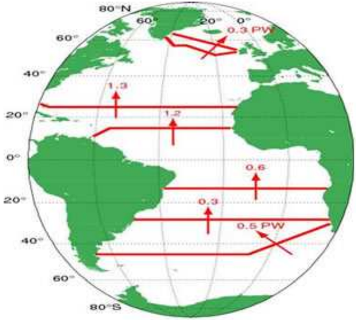
Table 1 The models used in the present study, including, configurations (near the equator) and number of years of simulations

Model	Institution	Atmosphere resolution	Ocean resolution	Length picntrl	Length 1pctto2x	Length 1pctto4x
CCSM3	NCAR (USA)	T85L26	1.125°×0.27°L40	230	150	n/a
CGCM3.1(T47)	CCCMA (Canada)	T47L31	1.85°×1.85°L29	500	150	150
CNRM-CM3	Meteo-France/CNRM (France)	T63L45	2°×0.5°L31	390	100	110
CSIRO-Mk3.0	CSIRO (Australia)	T63L18	1.875°×0.84°L31	380	10	n/a
ECHAM5/MPI-OM	MPI-M (Germany)	T63L31	1.5°×0.5°L40	332	100	81
FGOALS-g1.0	LASG/IAP (China)	T42L26	1°×1°L33	150	80	n/a
GFDL-CM2.0	GFDL (USA)	2.5°×2°L24	1°×0.33°L50	500	100	160
GFDL-CM2.1	GFDL (USA)	2.5°×2°L24	1°×0.33°L50	500	150	160
GISS-AOM	NASA/GISS (USA)	4°×3°L12	4°×3°L16	251	n/a	n/a
GISS-EH	NASA/GISS (USA)	5°×4°L20	2°×2°L16	500	80	150
GISS-ER	NASA/GISS (USA)	5°×4°L20	5°×4°L13	400	100	n/a
INM-CM3	INM (Russia)	5°×4°L21	2.5°×2°L33	330	n/a	n/a
IPSL-CM4	IPSL (France)	2.5°×3.75°L19	2°×0.5°L31	230	80	n/a
MIROC3.2(hires)	CCSR/NIES/FRCGC (Japan)	T106L56	0.28°×0.1875°L47	100	10	n/a
MIROC3.2(medres)	CCSR/NIES/FRCGC (Japan)	T42L20	1.4°×0.5°L43	500	100	150
MRI-CGM2.3.2	MRI (Japan)	T42L30	2.5°×0.5°L23	350	150	150
PCM	NCAR (USA)	T42L18	0.66°×0.5°L32	350	96	90
UKMO-HadCM3	HadleyCentre (UK)	3.75°×2.5°L19	1.25°×1.25°L20	341	10	n/a
UKMO-HadGEM1	HadleyCentre (UK)	1.875°×1.25°L38	1°×0.33°L40	80	10	n/a
SINTEX T30	IPSL/INGV (France,Italy)	T30L19	2°×0.5°L31	200	n/a	n/a
SINTEX T106	INGV/IPSL (Italy,France)	T106L19	2°×0.5°L31	100	n/a	n/a
SINTEX T106mod	IPSL/INGV (France,Italy)	T106L19	2°×0.5°L31	100	n/a	n/a
HadOPA	CGAM/IPSL (UK,France)	3.75°×2.5°L19	2°×0.5°L31	100	n/a	n/a

From Guilyardi (2006): for the most part, tables such as this are laboriously filled by hand. Currently deployed data frameworks (PCMDI, GFDL, DDC) are becoming capable of dynamically generating these tables.

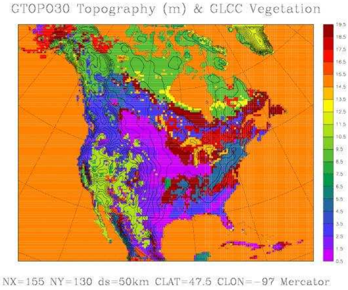
Horizontal regridding: poleward heat transport

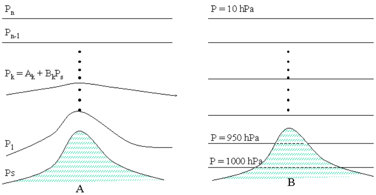
- Atmospheric data:
  - v, T, q, $\overline{v'T'}$, $\overline{v'q'}$
  - $F^{\uparrow}_{\text{sfc}}$, $F^{\uparrow}_{\text{TOA}}$
  - p_s
- Ocean data:
  - v, T, $\overline{v'T'}_{\text{total,gyre,eddy},\ldots}$: total and per basin.
  - meridional mass overturning circulation: total and per basin



`http://www-pcmdi.llnl.gov/ipcc/project_detail.php?ipcc_-subproject_id=174`

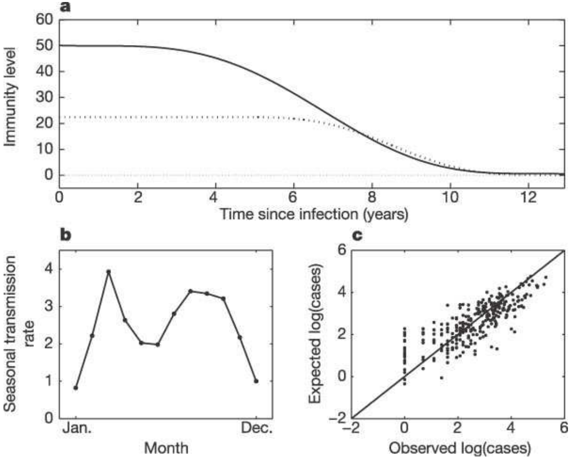
Vertical regridding: NARCCAP





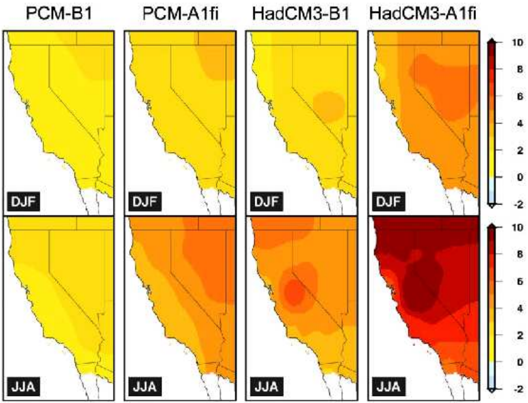
- The NARCCAP experiment is a MIP aimed at the "development of multiple high resolution regional climate scenarios for use in impacts assessments."
- High-resolution models requires forcing data from global models and analysis in specified resolution, projection, and vertical levels.
- Data volumes are high: GFDL has chosen to supply data on its native grid (24 levels) instead of the required 40; in conjunction with a program for converting data from σ-hybrid to pressure.

Disease vectors in a changing climate



Koelle et al, *Nature*, 2005: *Refractory periods and climate forcing in cholera dynamics.* Requires monthly forcing data, no feedback. This usage is typical of IPCC WG2 users.

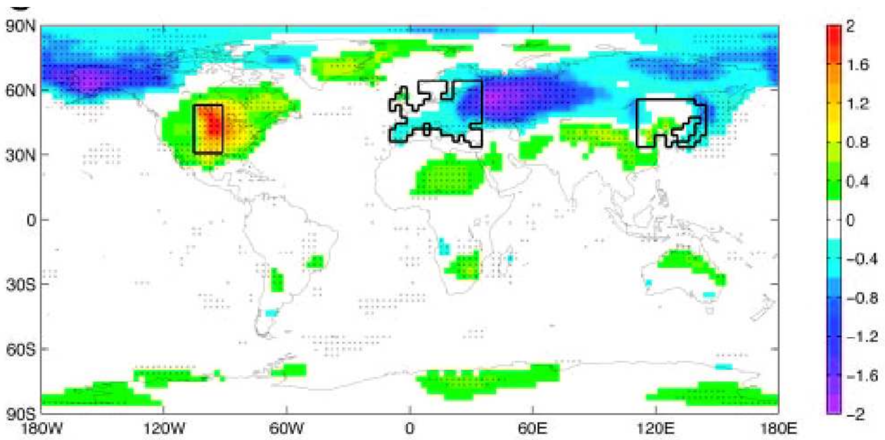
Statistical downscaling of climate change projections



Hayhoe et al, *PNAS*, 2004: *Emissions pathways, climate change, and impacts on California.* Uses daily data for "heat degree days" and other derived quantities.

What if it requires data beyond that provided by IPCC AR4 SOPs (1960-2000)?

Alternate energy sources



Keith et al, *PNAS*, 2005: *The influence of large-scale wind power on global climate.*
Feedback on atmospheric timescales: but does not require model to be retuned.

Talk outline . . .

1. From IPCC AR4 to AR5
   - Convergence of models and data
2. AR5 use cases
   - Querying model characteristics
   - Regridding
   - Offline models
   - Branch runs
3. Summary

Summary

- The boundary between discovery and use metadata is fuzzy: many relatively simple applications require more metadata than is currently available.
- Data stored on the native grid allows the derivation of auxiliary variables, and may also reduce data volume.
- A relatively short, focused effort, led by PCMDI for example, could identify and define the required metadata. The “use case” approach could be used to provide a clear boundary between what is and isn’t going to be possible.