Big Projects, Diverse Users, and Usable Data: Lessons and Examples from NARCCAP

Seth A. McGinnis, Linda O. Mearns, and Larry McDaniel

ISSE, National Center for Atmospheric Research, Boulder, CO

email: *mcginnis@ucar.edu*   website: *http://www.narccap.ucar.edu*

GC53A-0708

NCAR

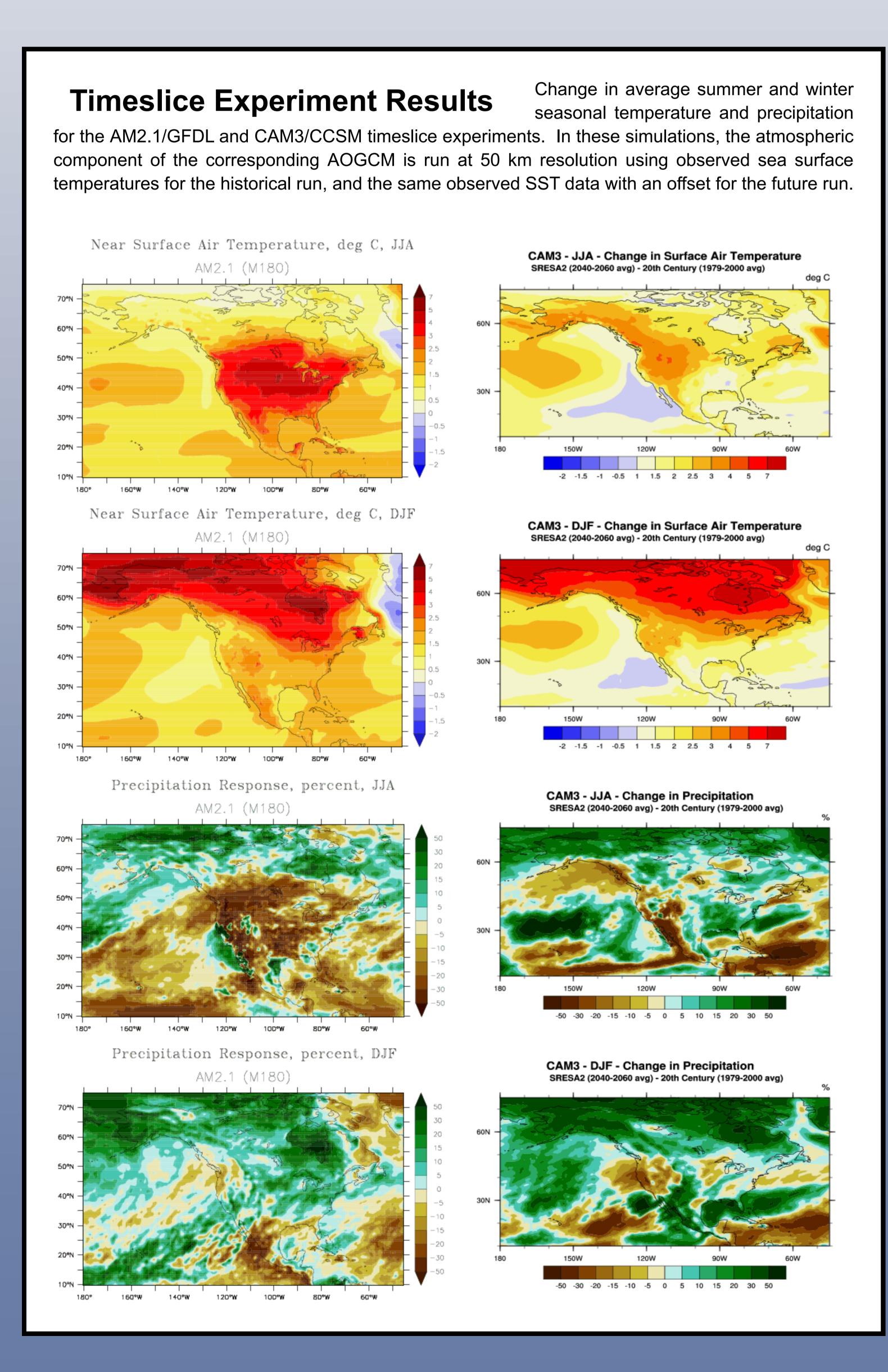**ABSTRACT**   The North American Regional Climate Change Assessment Program (NARCCAP) is an international program to produce high resolution climate change scenarios and investigate uncertainties in regional scale projections of future climate by nesting multiple regional climate models (RCMs) within multiple atmosphere-ocean general circulation models (AOGCMs) forced with the A2 SRES scenario and with historical data over a domain covering the conterminous United States and most of Canada. The resulting 60+ TB of data will be archived for distributed storage and made available to global change impacts researchers worldwide via the Earth System Grid (ESG). This presentation will describe the progress of the project and challenges we have encountered, and will discuss some useful techniques and practices for the management of large collaborative modeling projects, promotion of data usability, and building and maintenance of a diverse end user community. Examples of some preliminary results from the high-resolution GCM timeslice experiments and the NCEP-driven historical RCM runs will also be presented.

Map Projections, Sponge Zones, and GIS Metadata

Although all six RCMs have the same 50 km spatial resolution and cover the same domain, because they use different map projections they cannot cover exactly the same set of gridpoints. The models also differ in the size of the 'sponge zone' where the model data is mixed with the driving lateral boundary conditions. This results in differences in the effective domain size, and in one case required adjustment of the simulated domain to increase the area of overlap. In addition, certain of the required output variables simply do not exist for some models, or cannot be captured as single variable. These factors all show that even in an experiment designed to enable the intercomparison of different models, there are limits on how much they can be made to model "the same thing".

The impacts community uses GIS heavily in their analysis, so GIS compatibility is very important for making the data useful to impacts users. ESRI's ArcMap v.9.3 supports direct import of NetCDF data if the files follow the CF metadata standard. Ensuring CF compliance required correct specification of the map projection parameters used by each model and testing for proper GIS interoperability. This effort was significant, but paid off in transparent ingestion of NARCCAP data into GIS for the impacts community, and illustrates the importance and value of using a standard and adhering to it strictly.

NARCCAP RCM Domains

NARCCAP AT A GLANCE

- 4 different AOGCMs driving 6 different RCMs
- 50 km spatial resolution
- 3 hourly temporal resolution
- 52 output variables
- 2 high-resolution GCM timeslice experiments
- Future scenario: A2 SRES emissions

GCM RCM	Phase I	Phase II			
	NCEP	GFDL	CGCM3	HADCM3	CCSM
CRCM	DONE		1		2
ECPC	DONE	1		2	
HRM3	DONE	2		1	
MM5I	DONE			2	1
RCM3	DONE	1	2		
WRFP	DONE		2		1

Phase I: RCMs are driven by historical (1979-2003) observed (NCEP2 reanalysis) data
Phase II: Each RCM is driven by 2 GCMs for current (1968-2000) and future (2038-2070) scenarios

CURRENT STATUS: Impacts-relevant data for Phase I published via ESG for 5 of 6 RCMs. All groups have finished Phase I modeling, and 3 have finished the first Phase II GCM-driven run.

Timeslice Experiment Results

Change in average summer and winter seasonal temperature and precipitation for the AM2.1/GFDL and CAM3/CCSM timeslice experiments. In these simulations, the atmospheric component of the corresponding AOGCM is run at 50 km resolution using observed sea surface temperatures for the historical run, and the same observed SST data with an offset for the future run.

Techniques for Managing Large Distributed Projects

The biggest challenge faced by a project involving collaboration between multiple institutions is its disconnected nature. Geographical and organizational separation of the collaborators induces a kind of friction that slows interactions and introduces error. Frequent, clear, and effective communication is the best tool for counteracting this effect, but a distributed project will always proceed at a slower pace than a consolidated one. Specific techniques of value include:

- Restrict decision-making to involve the minimal set of participants needed to effectively address the issue.

- Automate everything you possibly can, including communications with users.

- Combat email fatigue by deploying wikis and other technologies that match medium to message well.

- Develop procedures for dealing with the normal flow of data and information.

- Don't hesitate to employ ad hoc substitutions for procedure when they are more effective.

- Employ higher-bandwidth communication channels when dealing with more complex issues.

- Recognize that the entire process is a balancing act, and be prepared to make changes when things go wrong, as they inevitably will. Evolve better methods.
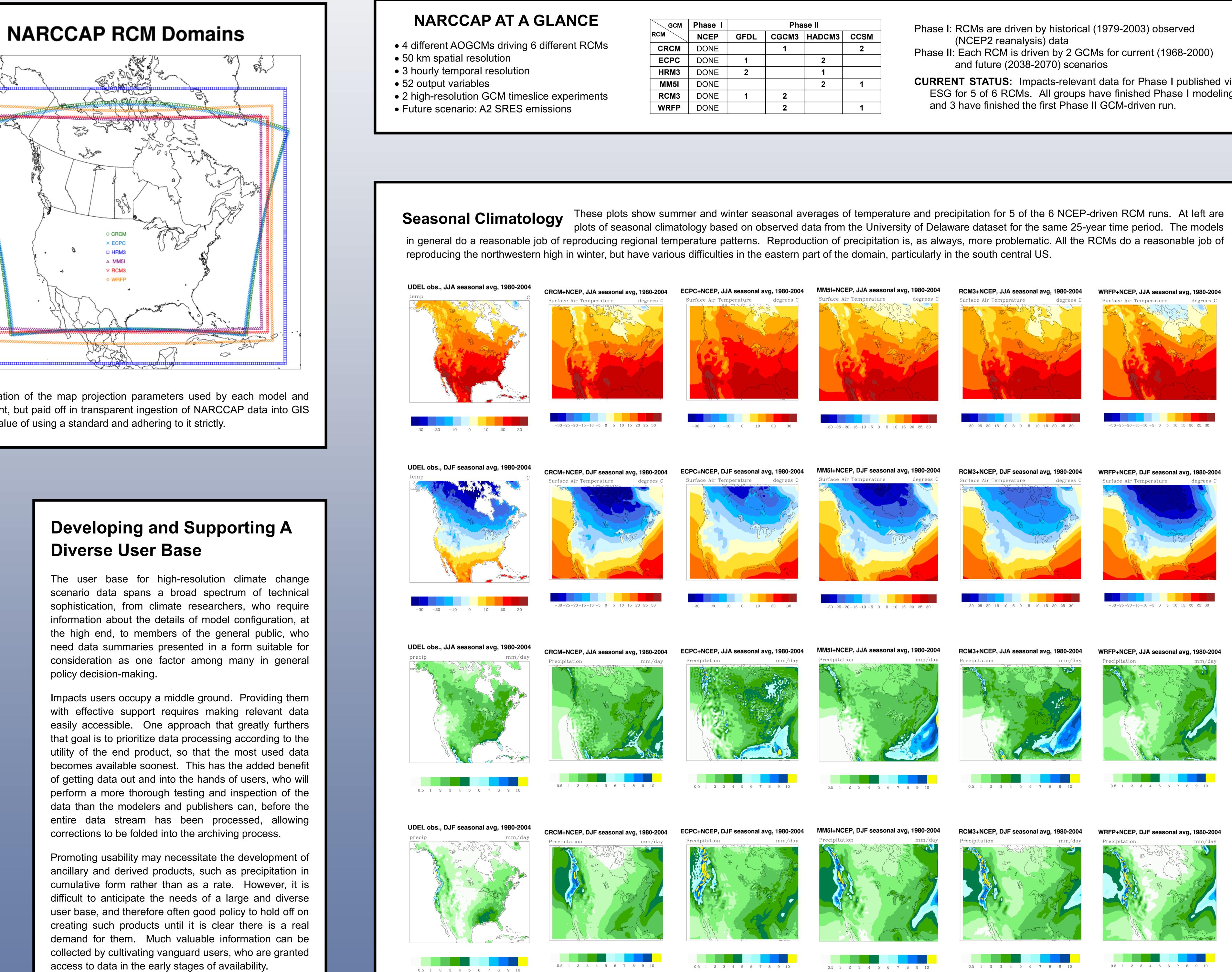
Developing and Supporting A Diverse User Base

The user base for high-resolution climate change scenario data spans a broad spectrum of technical sophistication, from climate researchers, who require information about the details of model configuration, at the high end, to members of the general public, who need data summaries presented in a form suitable for consideration as one factor among many in general policy decision-making.

Impacts users occupy a middle ground. Providing them with effective support requires making relevant data easily accessible. One approach that greatly furthers that goal is to prioritize data processing according to the utility of the end product, so that the most used data becomes available soonest. This has the added benefit of getting data out and into the hands of users, who will perform a more thorough testing and inspection of the data than the modelers and publishers can, before the entire data stream has been processed, allowing corrections to be folded into the archiving process.

Promoting usability may necessitate the development of ancillary and derived products, such as precipitation in cumulative form rather than as a rate. However, it is difficult to anticipate the needs of a large and diverse user base, and therefore often good policy to hold off on creating such products until it is clear there is a real demand for them. Much valuable information can be collected by cultivating vanguard users, who are granted access to data in the early stages of availability.

Seasonal Climatology

These plots show summer and winter seasonal averages of temperature and precipitation for 5 of the 6 NCEP-driven RCM runs. At left are plots of seasonal climatology based on observed data from the University of Delaware dataset for the same 25-year time period. The models in general do a reasonable job of reproducing regional temperature patterns. Reproduction of precipitation is, as always, more problematic. All the RCMs do a reasonable job of reproducing the northwestern high in winter, but have various difficulties in the eastern part of the domain, particularly in the south central US.