ESGF & NASA’s Modeling Projects

Tsengdar Lee
Presented at the 2016 ESGF F2F Meeting

December 6, 2016
Current NASA Earth Science Data Holding

- Observations include all products archived at EOSDIS (http://earthdata.nasa.gov/).
- Modeling products include all the high resolution climate modeling and data assimilation products at NASA Center for Climate Simulations (http://www.nccs.nasa.gov/) and NASA High-End Computing Capabilities (http://www.nas.nasa.gov/hecc/).
- Significant growth in modeling data is triggered by the availability of high resolution Earth observations and the computational resources.
# 2016 NASA Modeling Analysis & Simulation Product Plan

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>3D-Hybrid Ensemble-Var (25km)</strong>&lt;br&gt;32 ensemble members&lt;br&gt;Hydrostatic&lt;br&gt;1-Moment Cloud Microphysics&lt;br&gt;Current GEOS-5 FP system</td>
<td><strong>MERRA (50km)</strong>&lt;br&gt;Ending Feb. 2016&lt;br&gt;3D-Var&lt;br&gt;~200 TB</td>
<td><strong>M2R12K (12km)</strong>&lt;br&gt;MERRA2 downscaled to 12 km&lt;br&gt;Aerosols, CO₂, CO, SO₂, O₃&lt;br&gt;Non-Hydrostatic&lt;br&gt;1-Moment Cloud Microphysics</td>
<td><strong>G5NR (7km)</strong>&lt;br&gt;Simulated 2005-2007&lt;br&gt;Aerosols, CO₂, CO, SO₂, O₃&lt;br&gt;Non-Hydrostatic&lt;br&gt;1-Moment Cloud Microphysics&lt;br&gt;4 PB</td>
<td><strong>GEOS SFS (50km)</strong>&lt;br&gt;MERRA-2 replay&lt;br&gt;50km, 40L ocean analysis&lt;br&gt;31 members per month&lt;br&gt;Include aerosols, CO, CO₂&lt;br&gt;M2-driven EnSI ocean analysis</td>
<td><strong>GEOS CMIP (25km)</strong>&lt;br&gt;25km Atmosphere&lt;br&gt;25km 50L ocean&lt;br&gt;Include aerosols greenhouse gases&lt;br&gt;Hydrostatic&lt;br&gt;2-Moment Cloud Microphysics</td>
</tr>
<tr>
<td><strong>3D-Hybrid Ensemble-Var (12km)</strong>&lt;br&gt;32 ensemble members&lt;br&gt;Atmosphere, ocean surface&lt;br&gt;Hydrostatic&lt;br&gt;2-Moment Cloud Microphysics&lt;br&gt;Parallel FP stream in 1Q-2016</td>
<td><strong>MERRA-2 (50km)</strong>&lt;br&gt;3D-Var&lt;br&gt;Aerosols and CO₂, SO₂, O₃&lt;br&gt;1-Moment Cloud Microphysics&lt;br&gt;~400 TB</td>
<td><strong>IESA (12km)</strong>&lt;br&gt;3D-Hybrid Ensemble-Var&lt;br&gt;32 ensemble members&lt;br&gt;atmosphere, land, ocean surface&lt;br&gt;Aerosols, CO₂, CO, SO₂, O₃&lt;br&gt;Non-Hydrostatic&lt;br&gt;2-Moment Cloud Microphysics&lt;br&gt;5,000 cores; 40 simulation days/day&lt;br&gt;150 days total wallclock&lt;br&gt;~3 to 4 PB of data</td>
<td><strong>G5NR-CHEM (12km)</strong>&lt;br&gt;Simulated 2013-2014&lt;br&gt;Full Reactive Chemistry&lt;br&gt;Non-Hydrostatic&lt;br&gt;1-Moment Cloud Microphysics&lt;br&gt;1 PB of data</td>
<td><strong>GEOS SFS (25km)</strong>&lt;br&gt;Alignment with &quot;MERRA-3&quot;&lt;br&gt;25km, 50L ocean analysis&lt;br&gt;System design under review&lt;br&gt;FY2019 target</td>
<td><strong>GEOS SFS (25km)</strong>&lt;br&gt;Planning/discussion and system evaluation in progress&lt;br&gt;Will align with &quot;MERRA-3&quot; SFS&lt;br&gt;and strategic direction of ESD</td>
</tr>
<tr>
<td><strong>4D Ensemble-Var (9km)</strong>&lt;br&gt;~100 ensemble members&lt;br&gt;Atmosphere, ocean surface&lt;br&gt;Non-Hydrostatic&lt;br&gt;2-Moment Cloud Microphysics&lt;br&gt;(The first GEOS-6 system)&lt;br&gt;Parallel FP stream in 4Q-2016</td>
<td><strong>Coupled Reanalysis (&quot;MERRA-3&quot;)</strong>&lt;br&gt;IESA Downscaled to 4km&lt;br&gt;Atmosphere-land-ocean-eovosphere&lt;br&gt;(alignment with SFS and CMIP6)&lt;br&gt;FY2019 target</td>
<td><strong>IESAR4K (4km)</strong>&lt;br&gt;IESA Downscaled to 4km&lt;br&gt;downscaling evaluation for NCA&lt;br&gt;Aerosols, CO₂, CO, SO₂, O₃&lt;br&gt;Non-Hydrostatic&lt;br&gt;2-Moment Cloud Microphysics&lt;br&gt;5,000 cores; 40 simulation days/day&lt;br&gt;150 days total wallclock&lt;br&gt;~3 to 4 PB of data</td>
<td><strong>G6NR (3km)</strong>&lt;br&gt;Simulated 2015&lt;br&gt;Aerosols&lt;br&gt;CO₂, CO, SO₂, O₃, CH₄&lt;br&gt;Non-Hydrostatic&lt;br&gt;2-Moment Cloud Microphysics&lt;br&gt;~4 PB</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Core GMAO projects completed, in-progress**<br><br>**Pathfinding projects toward GMAO core efforts.**<br><br>**FY16 Projects**<br><br>**Projects undergoing GMAO discussion/evaluations**<br><br>**Planned Future Projects**
<table>
<thead>
<tr>
<th>Dataset</th>
<th>Domain/Resolution</th>
<th>Frequency</th>
<th>Variables</th>
<th>No of CMIP5 models</th>
<th>Baseline Data</th>
<th>Funding</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCP30 (Downscaled Climate Projections at 30arc sec)</td>
<td>CONUS, ~800m</td>
<td>Monthly</td>
<td>Tmax, Tmin, and Precip</td>
<td>34</td>
<td>Daly et al., 2002</td>
<td>NASA</td>
</tr>
<tr>
<td>BCCA (Bias Corrected Constructed Analogs)</td>
<td>CONUS, ~12km</td>
<td>Monthly</td>
<td>Tmax, Tmin, and Precip</td>
<td>21</td>
<td>Maurer et al. 2002</td>
<td>USBR</td>
</tr>
<tr>
<td>LOCA (Localized constructed analogs)</td>
<td>CONUS, ~6km</td>
<td>Daily</td>
<td>Tmax, Tmin, Precip; Humidity, Windspeed (in progress)</td>
<td>32</td>
<td>Livneh et al. 2013</td>
<td>USBR/CalEnergy</td>
</tr>
<tr>
<td>GDDP (Global Daily Downscaled Climate Projections)</td>
<td>Global, ~25km</td>
<td>Daily</td>
<td>Tmax, Tmin, and Precip</td>
<td>21</td>
<td>Sheffield et al. 2006</td>
<td>NASA</td>
</tr>
</tbody>
</table>
Gearing up for Climate Modeling
Data Analytics

- Traditional data center focuses on data archive, access and distribution
  - Scientists typically order and download specific data sets to a local machine to perform analysis
  - With large amount of observational and modeling data, downloading to local machine is becoming inefficient
  - Data centers are starting to provide additional services for data analysis
- NASA computing and computational science program is building “data analytics platforms” using “Climate Analytics as a Service” (CAaaS) such as NASA Earth Exchange (NEX), Regional Climate Modeling Evaluation Systm (RCMES), Climate Model Diagnostic Analyzer (CMDA) and Observation for Model Intercomparison Project (Obs4MIPs) using Earth System Grid Federation (ESGF)
  - Build on technologies
  - Enabled by a rule based data management system
  - Current research focuses on how to manage data movement from the archives to the analytical platforms
RCMES facilitates regional model evaluation efforts via open source analysis toolkit and efficient links to model output (e.g. CORDEX, CMIP) and global observations (e.g. obs4MIPs, ana4MIPs).

rcmes.jpl.nasa.gov
Regional Climate Model Evaluation System
Contributing tools, resources and training to CORDEX

CORDEX Africa
Model Evaluation
Kim et al. (2013)

RCMES Article
in WMO
Bulletin
(2012)

Assist Vietnam Gov’t
Funded Study with
Statistical
Downscaling of
Climate Projections
(K. Lee)

CORDEX S. Asia
Model Evaluation
Kim et al. (2015)

CORDEX-2016 Conference
Stockholm, Sweden
RCMES Training; ~50 attendees
A Project for identifying, documenting and disseminating observations for climate model evaluation.

Data sets accessible on the Earth System Grid Federation (ESGF) alongside the Coupled Model Intercomparison Projection (CMIP) model output, adhering to the same data conventions, greatly facilitating research.

Guided by the World Climate Research Program (WCRP) Data Advisory Council (WDAC) obs4MIPS Task Team.

Growing international partnerships.

*ESGF is partially down until March

.... and growing!
Science Cloud Architecture

- Agile, high level of support
- Storage is 90% full prior to use
- The system owns the data
- The users own their analysis
- Extensible storage; build and expand as needed
- Persistent data services built in VMs, Containers, or bare metal
- Create purpose build VMs for specific science projects
- Image management
Climate Model Diagnostic Analyzer

- Web-based tools running on Amazon cloud.
- Only requirement from a user machine is a web browser with an internet connection. No local installation needed.
- Provides datasets and analysis services.
- You can analyze the datasets using the services.
- You can download analyzed output datasets.
- You can download original input datasets.
Major Challenges Over Next 10 Years and What Can We Do Now

• Challenge: Modeling and observational data will continue to grow exponentially
  – Major challenge in data management, analysis, and collaboration
  – Tape archives will not meet big data analysis challenges
  – Network will not catch up
  – Library model will no longer work

• Actions now:
  – Build centralized data analytics systems
  – Data proximal analytic capabilities (move the analytics to the data)
  – Commoditize data storage and data analytics
  – Explore and adopt new storage technologies (e.g., object storage)

• Large scale science informatics system will be needed to solve the future data challenges