

### Climate Model Diagnostic Analyzer

Seungwon Lee, Lei Pan, Chengxing Zhai, Benyang Tang, Jonathan Jiang, Terry Kubar, Juilin Li, Joao Teixeira, Zhangfan Xing, and Hook Hua

at

Jet Propulsion Laboratory, Caltech

Jia Zhang and her team at Carnegie Mellon University

Presented at 2015 Earth Science Technology Forum, June 23-25, Pasadena CA





- Develop a novel methodology to diagnose model biases in contemporary climate models, to identify the physical processes responsible for creating model biases, and to incorporate the understanding into new model presentation that reduce the model biases.
- Implement the methodology as a web-service based, cloud-enabled, provenance-supported climate-model evaluation system named Climate Model Diagnostic Analyzer (CMDA) for the Earth science modeling and model analysis community.





## CMDA

- Enables multi-aspect, physics-based climate data analyses.
- Facilitates comprehensive and synergistic use of observational data, reanalysis data, and model outputs.
- Is a web-service oriented system.
- Does not require local software/library installation.
- Provides all the input data needed for analysis.
- Runs on the Amazon cloud system.
- Supports interactive visualization.
- Collects provenance and supports provenance-based search.





### CMDA Data Sets

### **Model Outputs from CMIP5 project**

- Experiments:
  - Historical, AMIPs, Forecast
- Models:
  - CCCMA/canesm2, GFDL/esm2g, GISS/e2-h, GISS/ e2-r, NCAR/cam5, NCC/noresm, UKMO/hadgem2es, CCCMA/canam4, CSIRO/mk3.6, GFDL/cm3, IPSL/cm5a-lr, MIROC/miroc5, UKMO/hadgem



**Reanalysis Data from ECMWF and Merra** 

- Vertical Wind
- Relative Humidity

### **Observation Data from Obs4MIPs**

- AMSR-E surface temperature
- AIRS and MLS air temperature & water vapor content

To be Evaluated

References

- MODIS total cloud fraction, leaf area index
- GPCP and TRMM precipitation
- AVISO sea surface height
- CERES radiation fluxes



### CMDA Analysis Tools







### Analysis Tool Screenshots

### http://cmacws2.jpl.nasa.gov:8080/cmac/

		5	iervice: 2-D Vari	able Map	
This servi Select a data sou	ce generates a map rce (model or obser	of a 2-c rvation)	limensional varial , a variable name,	ble with time averaging a time range, and a spa	and spatial subsetting. tial range (lat-lon box) below
data source:	NASA/MODIS	0	variable name:	Total Cloud Fraction	
start year-month:	2004-01		end year-month:	2004-12	
select months:	select all				
⊮ Jan ⊮ Feb ⊮ Mar	⊯ Apr ⊮ May ⊮ Jun		⊯ Jul ⊯ Aug ⊮ Sep	<ul> <li>Øct</li> <li>Nov</li> <li>Øcc</li> </ul>	
start lat (deg):	-90		end lat (deg):	90	
start lon (deg):	0		end lon (deg):	360	
color scale:		thmic			



### Service: 2-D Variable Zonal Mean This service generates a graph of a 2-dimensional variable's zonal mean with time averaing



### This service generates a graph of a 2-dimensional variable's time series with monthly averaged value Select a data source (model or observation), a variable name, a time range, and a spatial range below data source: NASA/Qu start year-month: 2000-01 start lat (deg): -90 NASA/QuikSCAT variable name: Eastward Near-Surface Wind end year-month: 2012-01 end lat (deg): start lon (deg): end lon (deg): variable scale: 🛞 linear 🛛 loga uas, average value over lon(0,360)deg, lat(-90,90)deg, (m s-1) -0.2 -1.2 -1.4 LL. 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009

Service: 2-D Variable Time Series

### Service: 3-D Variable 2-D Slice This service generates a slice map of a 3-dimensional variable at a selected pressure level. Select a data source, a variable name, a time range, a spatial range, and a pressure level below variable name: Vertical Wind Velocity data source: ECMWF/interin end year-month 2004-12 atmosphere pressure level (hPa) or ocean pressure level (dbar) elect months: select all May Nov Aug Jun Ser start lat (deg): -90 start lon (deg): end lon (deg color scale



### Service: 3-D Variable Zonal Mean

This service generates a contour plot of zonal-mean vertical profiles of a 3-dimensional variable. Select a data source, a variable name, a time range, and a pressure range below.

P	•	variable name:	Cloud Ice Water Content
2004-01		end year-month:	2004-12
select all	•	atmosphere pressure range (hPa) or ocean pressure range (dbar):	100,1000
☑ Apr		🕑 Jul	☑ Oct
May		☑ Aug	Nov
Jun		✓ Sep	☑ Dec
-90		end lat (deg):	90
○ linear    logarithn	nic	color scale:	
	2004-01 select all	2004-01 select all ♥ ♥ Apr ♥ May ♥ Jun -90 ● linear ⊛ logarithmic	2004-01         [end year-month].           telect all         2)           for occan pressure range (dbar):         #           # Apr         # Aug           # Jun         # Aug           # Jun         # Aug           # On         [end lat (deg):           -50         [inear + b]ogarithmic

### Mass Fraction of Cloud Ice, 2004/01-2004/12 zonal mean map climatology (Kg/Kg), Annual



### Service: Scatter and Histogram Plot of Two Variables

variables can be either a two-dimensional variable or a slice of a three-dimensional variable at a specific pressure level. The number of samples used for this analysis should

	data 1	data 2		
source:	NASA/CERES \$	source:	CCCMA/CANESM2 \$	
variable name:	Surface Downwelling Shortwave Radiation \$	variable name:	Surface Downwelling Shortwave Radiation \$	
atmosphere pressure level (hPa) or ocean pressure level (dbar)	N/A	atmosphere pressure level (hPa) or ocean pressure level (dbar)	N/A	
start year-month:	2004-01	end year-month:	2004-12	
start lat (deg):	-90	end lat (deg):	90	
start lon (deg):	0	end lon (deg):	360	
number of samples:	10000	]		
Get Plot		Download Data		



### Physical Variable (sampled variable stal Variable (sampling variable UKMO/HadGEM2-ES \$ ECMWF/interim variable name: Precipitation Flux variable name: osphere pressure range (hPa N/A 500 or ocean pressure range (dbar) or ocean pressure level (dbar): 2000-0 start year-month end year-month select all elect month ⊯ Apr ⊮ May ≷ Jun 🛛 Jan e Jul Oct e Feb e Mar Nov ≥ Nov ≥ Dec Aug Sep start lat (deg) end lat (deg) start lon (deg): end lon (deg K-axis (sampling variable) scale: Inear logarithm lisplay option Y-axis (sampled variable or pressure) scale: $\odot$ linear $\circledast$ logarit Z-axis (color) scale: inear logarithmic default: @ pling variable binning specification Max: N/A number of bins: N//

Service: Conditional Sampling

a Jan

Feb
 Mar





This service generates a scatter plot between two specified variables and the histograms of the two variables, and calculates the correlation of the two variables. The two

be specified.

This service sorts one variable by the values of another variable (environmental condition, e.g. SST) and displays the averaged value of the first variable as a function of the bin value of the second the first variable is a two-dimensional variable, the plot will be a X-Y plot. If the first variable is a three-dimensional variable, the plot will be a colored profile image.



- **CMDA was used as an educational tool for the Summer School** organized by JPL's Center for Climate Science in September 22-26, 2014.
- The theme of the Summer School in 2014 was "Using Satellite Observations to Advance Climate Models".
- **23 students attended** the summer school: graduate students or postdocs from around the world.
- Five research topics were prepared by the JPL Earth scientists.
- The students worked on their research topic for total 5 hours of group research time.
- Each student was assigned to a virtual machine in Amazon Cloud to use CMDA.
- An one-hour session of the CMDA introduction was given and the students were able to use the tool immediately to do their research.
- The students presented their result on the last day of the school.





### 2014 JPL Summer School





### Student Feedback

- A tool survey was given to students, and students comments were collected.
- Comments from students for appreciation:
  - "I found the tool to be very user-friendly."
  - "It is really straightforward to use."
  - "It is not so difficult to use, seems intuitive."
  - "It was very useful for the exploratory analysis"
  - "The web tool was pretty instrumental in the completion of the research project."
  - "I could see this tool bing easily used in classroom settings."

### • Comments from students for improvement:

- "Provide a comprehensive description of the datasets available."
- "Provide the time ranges over which the variables are available."
- "Provide a tool to download data with a user specified condition (time, space, grid resolution for more detailed analyses in other programs."
- "Provide interactive visualization capabilities: interactive space selection, color range selection, etc."
- "Provide an exemplary picture of each analysis tool's output"





## Diagnostic Methodology

- Conditional Sampling
- Random Forest Feature Ranking
- Conditional Probability Density
- Time Lagged Correlation





## **Conditional Sampling**

Displays a physical quantity  $\mathbf{X}$  according to the values of another physical quantity  $\mathbf{E}$ , which is related to  $\mathbf{X}$  via physical processes.







## **Conditional Sampling**



# Random Forest Feature Ranking

Measures how important each input variable  $X_i$  is in predicting the regression target variable Y without assuming a particular relationship. Build an ensemble of decision trees.







### Random Forest



# Conditional Probability Density

Displays a probability density function of the sampled parameterX for a given sampling parameter bin E. Carries considerablymore information about the relationship between parameter X andE than the mean and variance of the parameter X for a given E.





0.05 0.07 0.09 0.11 0.13 0.15 0.17 0.19





### Provenance System

- A pilot study with Carnegie Mellon University Prof Jia Zhang's Group.
- Building a provenance-driven recommendation engine for CMDA

http://einstein.sv.cmu.edu:9019/climate

- Service configuration/execution provenance
- Data usage provenance
- Service usage provenance
- Provenance-based search
- Data recommendation
- Service recommendation
- Bug report
- Web service publication and interface design





## Summary

- CMDA will provide the climate modeling and model analysis community with a diagnostic tool to evaluate climate models.
- CMDA will help the scientists identify the physical processes responsible for creating model biases.
- CMDA will facilitate community-wide use and relatively effortless adoption of the novel diagnostic methodology through web-service and cloud technology.
- CMDA will collect processing history and allow provenancebased search.

