Contents

1 Scientific/Technical/Management 1-1

1.1 Introduction 1-1

1.2 Objectives and Expected Significance 1-2

1.3 Technical Approach and Methodology 1-3

1.3.1 Existing Capabilities for CMDA 1-3

1.3.2 New Capabilities for CMDA 1-4

1.3.3 Data supported by CMDA 1-4

1.3.4 Overall Architecture of CMDA 1-5

1.3.5 Data Processing System of CMDA with Interconnectivity 1-5

1.3.6 Provenance Support System of CMDA 1-7

1.3.7 Scalable Job Distribution System of CMDA 1-9

1.3.8 Scientific Use of CMDA 1-11

1.3.9 Educational Use of CMDA 1-12

1.3.10 Tool/Technology Life Cycle and Operations Concept 1-13

1.4 Perceived Impact to State of Knowledge 1-13

1.5 Relevance to Element Programs and Objectives in the NRA 1-14

1.6 Work Plan 1-15

1.6.1 Key Milestones 1-16

1.6.2 Management Structure and Team Member Contributions 1-16

1.7 Data Sharing 1-17

2 References and Citations (no page limit) 2-1

3 Acronyms 3-1

4 Biographical Sketches 4-1

4.1 Principal Investigator 4-1

4.2 Co-Investigators 4-3

4.3 Collaborator 4-6

5 Current and Pending Support 5-1

5.1 Current Awards 5-1

5.2 Pending Awards 5-2

6 Budget Justification (partially page limited) 6-1

6.1 Budget Narrative 6-1

6.1.1 Personnel and Work Effort (1 page max.) 6-2

6.1.2 Facilities and Equipment (2 pages max.) 6-2

6.1.3 Rationale and Basis of Estimate 6-3

6.2 Budget Details – Year 1 6-7

6.3 Budget Details – Year 2 6-13

6.4 Co-Investigator <Name> Budget Details 6-15

# Scientific/Technical/Management

## Introduction

***Both the National Research Council (NRC) Decadal Survey and the latest Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) stressed the need for the comprehensive and innovative evaluation of climate models with the synergistic use of the global observations****.* NASA’s Earth Observing System (EOS) provides global, near simultaneous, multi-instrument observations of the Earth’s atmosphere, ocean, and land systems. The fusion of the multiple instrument observations allows for much more comprehensive insight than any one individual instrument can provide alone. Additionally, the emergence of high-quality reanalysis datasets is instrumental in providing the large-scale dynamic and thermodynamic context of climate systems.

Despite the widely acknowledged value of the satellite observations and reanalysis datasets for model evaluations, the synergistic use of the datasets is not widespread because the datasets are heterogeneous, scattered, domain specific, and innately different from model outputs. Data from different instruments and reanalysis systems have different spatial and temporal sampling characteristics and the different formats and structures of the retrieval/assimilated products. At the same time, the datasets are scattered in different data archive centers. More challengingly, the observational datasets are intrinsically different from model outputs so that many steps of data preparation and analysis are needed to bring the two close to be comparable.

Recognizing these difficulties, NASA has recently launched a pilot project called Obs4MIPs to prepare NASA satellite observational datasets that would be directly comparable to model outputs and to facilitate the wide use of the datasets for model evaluations by archiving the specially prepared observational datasets in the same location as model outputs using the Earth System Grid. Obs4MIPs now provide over 10 parameters covering a wide range of cloud, radiation, precipitation, thermodynamic, and dynamic parameters. The Obs4MIPs datasets greatly facilitate the synergistic use of the global satellite observations for model-to-observation comparisons and analyses. ***An analysis tool that can accompany the Obs4MIPs datasets to support the model-data comparisons and analysis would further increase the benefit of the Obs4MIPs datasets to the modeling and model analysis community.***

Satellite observations have been broadly used in model-data comparisons and model evaluation studies in the past. These studies normally involve the comparison of a single parameter at a time using a global time and space average. This traditional approach can evaluate model performance and identify model systematic biases but cannot diagnose the model problem. The model diagnosis process requires physics-based, multi-variable comparisons with conditionally sampled climate regimes and phenomena. ***A model evaluation tool should have the diagnostic analysis capabilities to associate any systematic model biases with model deficiencies in physical process representations and to direct the model improvements.***

Apart from data analysis, scientists spend a large portion of research time on searching previously analyzed results and regenerating the same results when they fail to find the results. This lamentable practice is because scientists have few tools to help them keep track of the processing history of their processed datasets. Current practical solutions are organized notes, filenames long and descriptive enough to remind them the processing history of the data in the files, and hierarchical file structures to organize the processed data. ***A tool that can keep track of processing history during analysis calls, represent the summary of the processing history in a human readable way, and enable provenance-based search capabilities will greatly improve the productivity of the scientists using the analysis tool.***

## Objectives and Expected Significance

### Objectives

The proposed work is a response to the subtopic of this NRA: “*Tools that improve and expand the accessibility and usability of NASA’s Earth science observational data for the modeling and model analysis communities*,” and has three broad objectives.

***Objective 1: We will develop a provenance-empowered, distributed web-service oriented information system that enables climate model performance evaluations and diagnoses through the comprehensive, synergistic, and quality-controlled use of multiple observational data, reanalysis data, and model outputs.*** The information system, named ***Climate Model Diagnostic Analyzer (CMDA)***, will have the following core capabilities: (1) perform a multi-step analysis for model diagnosis by calling multiple individual web services in sequence with proper exchange of input and output arguments and data between the web services, (2) generate a human-readable document to describe the processing history of resulting datasets and plots, (3) search the processed datasets and plots based on the processing history, and (4) handle multiple-user simultaneous web-service calls in a load-balanced way to optimize the use of computing resources.

***Objective 2: We will engage the modeling and model analysis community to the proposed project, and promote the community awareness of CMDA and the growth of the CMDA user community.*** The target users of CMDA are modeling and model analysis communities who are engaged in the following programmatic entities: the World Climate Research Program, World Weather Research Program (WWRP), Global Energy and Water Cycle Experiment (GEWEX) Cloud System Study (GCSS), Climate Variability and Predictability (CLIVAR) research program, Working Group on Numerical Experimentation (WGNE), and the Working Group on Coupled Modeling (WGCM) for IPCC’s 5th Assessment Report. The Co-I Dr. Jiang and Collaborator Dr. Teixeira are actively involved in these modeling and model analysis programmatic entities, and are familiar with the governance of the entities through their participation and leading roles.

***Objective 3: We will demonstrate the scientific and educational use of CMDA***. The Earth science focus area of the proposed work is in understanding and evaluating clouds, convection and their radiative processes represented in numerical weather and climate models. We have identified two scientific applications that will be used as a demonstration of the scientific use of CMDR: (1) Evaluating representations of atmospheric moist process in the models and (2) Evaluating representations of atmospheric radiation process in the models. The educational use of CMDA is also identified. The Center for Climate Sciences (CCS) in JPL will use CMDA as an educational tool for its summer school that annually brings together about 30 graduate students and postdocs from around the world to engage with premier climate scientists from JPL and elsewhere.

### Expected Significance

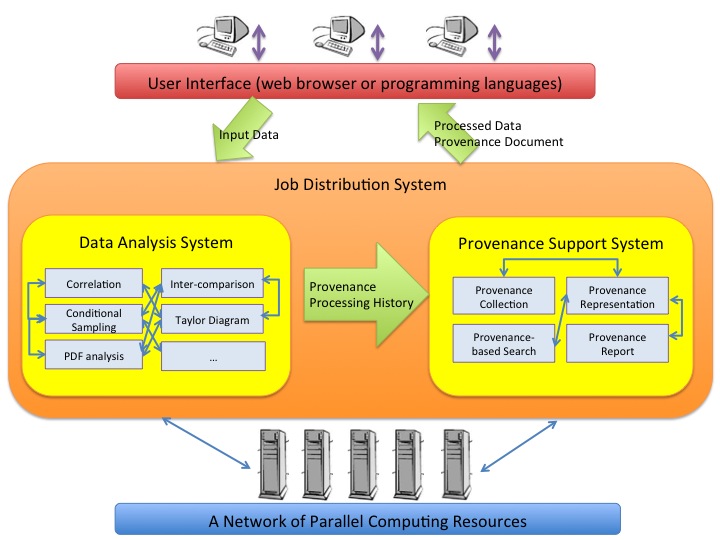
A number of national and international weather and climate programmatic efforts are calling for enhanced use and application of the observational datasets to maximize the investment put into these observational systems and also to leverage them to the greatest extent possible for improving our weather and climate simulation and prediction capabilities [e.g. NRC 2007; Randall et al. 2007]. The proposed tool CMDA will improve and expand the accessibility and usability of NASA’s Earth science observational data for the modeling and model analysis community. CMDA will increase the productivity of their research efforts by providing a comprehensive analysis tool to do model-observation comparisons, model performance verification and validation, and model diagnosis with a capability to keep track of the multi-source data analysis history. CMDA will empower scientists with capabilities to identify model biases, understand the causes of the biases, and to ultimately improve climate model performance for climate sensitivity studies and predictions. These increased research productivities of the modeling and model analysis communities will in turn increase the science return of the NASA Earth observational systems.

In addition, the proposed work will help elevate the NASA’s Obs4MIPs effort, a multi-instrument level coordinated effort to facilitate the use of NASA observational datasets for model evaluations, into a new and high level. The Obs4MIPs effort has become increasingly more visible and recognized as valuable assets by the domestic and international modeling and model analysis communities (e.g., the World Climate Research Program). CMDA will support all the Obs4MIPs datasets and all the Coupled Model Intercomparison Project Phase 5 (CMIP5) model decadal hindcast experiment outputs as input data for analyses. By supporting a diagnostic model analysis using the Obs4MIPs datasets and CMIP5 model outputs side by side, CMDA will demonstrate the utility of satellite observations in evaluating climate models on a much broader scope and deeper level than existing efforts. The capabilities provided by CMDA will be a significant boost to the NASA observations including the EOS and Obs4MIP and constitutes an important contribution to the IPCC’s climate assessment activities.

## Technical Approach and Methodology

### Overall Architecture of CMDA

The overall architecture and user interface design of CMDA are illustrated schematically in Figure 1. ***The architecture of CMDA consists of three main functional blocks: (1) data processing system, (2) provenance support system, and (3) job distribution system.*** The data processing system is composed of several processing units, individually wrapped with web-service interfaces and powered by the Cloud technology to interconnect multi-step service calls with high network communication efficiency. The provenance support system is composed of processing history collection system, processing history representation system, and processing-history based data search system. They will be also wrapped with web service interfaces. The job distribution system will handle the scalable distributions of simultaneous job calls to utilize the computing resources optimally and to reduce the user waiting time for the return of the requested process results. CMDA will utilize the network of parallel computing resources in executing the web services behind the scene.



**Figure 1. Schematic illustration of the architecture and user interface of the proposed CMDA.**

The user interface design of CMDA is tuned to meet the needs of the target users, which are the modeling and model analysis community. ***CMDA will have both a programming-language (Pyton, Matlab, IDL) interface and a web-browser interface.*** Climate model analyses are by nature exploratory which requires multiple runs with different input configurations. A programming-language interface is effective for scripting multiple runs and batching the runs. On the other hand, a web-browser interface is useful for instantaneous use without the hassle of local installation and compatibility issues. This flexible dual-interface design will not only lower the learning curve and the adoption barrier of the tool, but also increase the productivity during the period of intense usage. This will make CMDA optimal for an education tool for the JPL’s Climate Science Center Summer School, which plans to use the current version of CMDA for this summer in August 2013 and to use CMDA annually thereafter.

### Existing Capabilities for CMDA

Over the last several years, we have developed several model analysis and data processing tools for our current/prior NASA ROSES CMAC, ACCESS, COUND, MAP, MEaSURE, and AIST projects. The existing tools cover a large scope of data processing capabilities, data provenance technologies, parallel computing technologies, and web service technologies. Some of the capabilities and technologies are developed separately and not linked together for a single information system. We will leverage the existing capabilities and technologies to develop the proposed system CMDA. All of the existing technologies are Technology Readiness Level (TRL) 6 or higher. The existing capabilities accumulated from the prior/current projects that will be leveraged for CMDA are

* Model-data comparison in the common spatial grids and time series;
* Multivariate conditional sampling and correlation for model diagnosis;
* Single or joint variable probability density function analysis;
* Web service technologies to translate research analysis tools to web services;
* Web-browser interface for calling the web services;
* Provenance collection and representation technologies.

### New Capabilities for CMDA

The proposed system CMDA will be built upon the current version of CMDA, which is developed for the NASA ROSES CMAC project entitled “Parallel Web-Service Climate Model Diagnostic Analyzer” and led by Dr. Lee (PI of this proposal). The new capabilities that will be added to the current version of CMDA are

* Interconnectivity of individual analysis web services
* Provenance collection capability
* Provenance representation capability
* Provenance-based data search capability
* Scalable job distribution capability

In the current version of CMDA, individual analysis web services are implemented but the interconnectivity of the web services for multi-step analyses is not enabled yet and thus constitutes an important new capability in the proposed system. While the provenance technologies are developed as independent technologies for other applications in our prior project, incorporating them into the proposed CMDA system is a new capability. Individual analysis web services are currently optimized for the use of parallel computing resources but the distribution of multiple simultaneous service calls by multiple users is not optimized in the current system. The proposed system will have a scalable job distribution system to handle the multiple simultaneous calls in a load-balanced way to optimize the use of computing resources.

### Data supported by CMDA

The observational datasets to be supported by CMDA are the Obs4MIPs-project generated satellite observation datasets, which are specially prepared for the evaluation of climate models and model predictions. The Obs4MIPs datasets include atmosphere temperature and specific humidity retrieved from AIRS and MLS, sea surface temperature from AMSR-E, long-wave and short-wave radiation from CERES, total cloud fraction from MODIS, precipitation from TRMM, surface wind from QuickSCAT, and cloud, radiation, and atmospheric quantities from DOE’s ARM. More datasets are being prepared and expected to be available in the near future. We will include more datasets as they are available from the Obs4MIPs site.

The model outputs to be supported by CMDA are CMIP5 global circulation models (GCMs) outputs. For the IPCC 5th Assessment Report (AR5) scheduled for publication in late 2013, over 20 climate modeling groups from around the world have joined to design a new set of coordinated climate model experiments, which comprise CMIP5. The CMIP5 model outputs are a valuable and representative asset to assess the capability and limitation of the contemporary climate models. Among the CMIP5 model outputs, decadal hindcast experiment outputs are most relevant for comparisons with observations and model diagnostic studies. Therefore, we will include the decadal hindcast experiment outputs.

### Data Processing System of CMDA with Interconnectivity

The data processing system of CMDA has been partly developed under the current ROSES CMAC program. The existing system contains many individual processing units that are made to be separate web services. The new capability to be added in the data processing system is the interconnectivity between the individual processing units. This new capability will allow users to perform a multi-step analysis by calling multiple analysis web services in sequence with proper exchange of input and output arguments and automatic transfer of data between the web services.

Let us assume we have a 3D variable that a user wants to first apply a mean value calculation over the vertical dimension to bring it down to a 2D variable, and then conduct some two-dimensional time series analysis over the new 2D variable. In this scenario, our client will need to call two of our services in the order as in the following:

new\_2D\_var = <http://ws.jpl.nasa.gov/services/threeDimMeans?input1=in1&input2=in2>

time\_series\_result = http://ws.jpl.nasa.gov/services/twoDimTimeSeries?data=new\_2D\_var&input3=in3

The “equals” sign (i.e., “=”) in the above represents a web service call. This can be done from various clients, such as a python program, a shell script using curl, a web browser, or a web page running JavaScript programs. The URI (uniform resource identifier) on the right hand side of the equation is said to be in RESTful style, where the scoping information is carried in the URI and the method information is conveyed by HTTP (GET in this case). Examples of scoping information are input arguments such as spatial and temporal range, variable name, and a season of choice.

Notice that in this example two web services are “chained” with the intermediate data, namely “new\_2D\_var,” as the connection, meaning the output of the first service call becomes the input of the second. In a “normal” setting, the client file system becomes the means for intermediate data transfer. That is, the client first saves the variable array to its disk as a file after the first service call, and then uploads the file to the second server as input data for the second service call. A more efficient approach would be to have only the URI of the intermediate data, but not the data itself, be passed back and forth between the client and servers, and let the two servers negotiate between themselves how to transfer the data itself. In this way, the cost saving in data communication can be significant because the two web services usually are deployed with the network link between them much stronger than the connection from the clients, which may be wireless, even when the two servers belong to two different institutions. We will implement the URI-based interconnectivity and communication between web services.

In general, the two CMDA web services can be deployed to two different hardware servers, with no shared file system. To provide an efficient solution to allow users to perform multi-step analyses by calling multiple web services in a sequence across a LAN (local area network), a WAN (wide area network), or a mixed setting, we propose to use the Cloud technology to facilitate a direct and seamless connection among our web services. One possibility would be the Amazon AWS, but the cost of data movement on AWS is the main reason for us to look into open source Cloud solutions that are based on our local on-site storage. We propose to use ownCloud 5, which is an open source, Dropbox like cloud system with a rich set of functionalities and large user base. With ownCloud 5, a client can “drop” a file to a directory for another client to pick up and use. The files on the client are automatically synced with the server. According to the founder of ownCloud, there are more than 800,000 active users of ownCloud as of February of 2013.

The ownCloud 5 is a PHP web application running on Apache on Linux (also on Windows). This PHP application manages every aspect of ownCloud 5, namely user management, plug-in, and file sharing and storage. This is done through a database that is connected to the PHP app and serves the users and maintains control of data. The ownCloud 5 uses a built-in storage abstraction layer to leverage any storage protocol that can be mounted on the server, such as NSF, GFS2, FTP, and WebDAV. The ownCloud 5 integrates into currently prevailing file systems with features in security, monitoring, governance, back-up, activity logging. The users can store files in standard file system formats on most file systems. More importantly, the storage is physically located in our data center, allowing full data protection, secure access, and backups. The ownCloud 5 can be deployed to physical, virtual, or private cloud servers. This is good news because our CMDA web services are deployed to a VMWare virtual machine. For server side configuration, the ownCloud 5 provides a web portal as a central location for administrative control and configuration of the system and file and folder access, and for managing settings and administrative tasks such as migration and backup.

In summary, ***we propose to interconnect multi-step service calls in CMDA using the URI-based communication and the Cloud technology, namely ownCloud 5, to support multi-step analyses and improve network communication efficiency and simple and scalable application-level implementation.***

### Provenance Support System of CMDA

Another new capability to be added to CMDA is to keep track of processing history and to search results based on the processing history. A recent advance in provenance collection and representation technologies makes this capability easy to add to an existing information system with complex data processing history. The Provenance Working Group in W3C has been formed to define a language for exchanging provenance information among applications, making these provenance technologies more standardized. Following the standards set by the working group, we will make CMDA more transparent and infusible to other Earth science data information systems that would use similar provenance technologies. The benefit to the users of CMDA is very tangible. The users will have an easy-to-read processing history document attached to each dataset and plot that they create using CMDA and will be able to search their processed datasets and plots using key words in the data processing methods that they used. They will significantly ease the burden of keeping track of many files and plots that they generate during their exploratory studies.

Scientists often run into a situation where “they have a baked cake, but they don’t have the recipe.” They use our web services to analyze data, but after a while, they tend to lose track of what they did to obtain those results that they saved on their disk drives. If the results are passed on to them by their colleges, it is even harder to know exactly what algorithms and parameters were applied. Ad hoc ways, such as long, descriptive file names are often being used to attack the problem, but these are not automatic and hence a burden on the scientists. We propose to use the technology and system of provenance as the means to automatically collect the “recipe information” as the “cake” is being “baked.” Moreover, our tools will have the “recipe information” indexed and searchable.

The provenance of some data file (or, “artifact” in provenance terminology) is simply its history. Examples of the history information include: URIs (or, locations) of input and output data, upstream data source used to produce the output, source code (if open source) or executable used to produce the output, code version, ATBD (algorithm theoretical basis document) used, parameters (e.g., lat-lon bounding box, time window) used for each and every data processing step, and the user of the web service who produced the output.

The provenance mathematical model includes three roles, namely, Artifact, Process, and Agent, and their causality relationship (that is, who is the effect caused by what). In the world of data analysis, an Artifact is a data product, a Process is the execution of a data processing program, and an Agent is a human user or an institution. There are five causality relationships in the abstract provenance model: used, wasGeneratedBy, wasControlledBy, wasTriggeredBy, and wasDerivedFrom. Graphically, the causality relationships can be conveniently represented as triples (subject, predicate, and object) in RDF (Resource Description Framework), a semantic web standard model for data interchange.

In order for the provenance information from our data analysis web services to be collected and ingested into a searchable, browsable, and citable provenance database, we will leverage the provenance web services developed at JPL for the ACCESS-09 funded project “Tracking Production Legacy of a Multi-sensor Merged Climate Data Record (PI: Hook Hua).” The Co-I Dr. Pan was a core developer of the provenance web services (refer to as JPL Provenance System from here on) [Hua et al. 2012]. Using the proven technologies and existing system, we will implement the provenance support system of CMDA with the following steps:

1. Instrument our analysis web service code to generate a provenance log to collect provenance data for each session associated with a client request. The provenance collection lines are inserted into our CMDA service code wherever applicable. For example, the first line is inserted where an analysis program is starting, the second line is inserted to where input data is fed to the program, the third line is inserted to where input parameters are taken, the fourth line is inserted to where the output image is generated by the program, and finally the fifth line is put where the service is ending for the current session;
2. Call the analysis web services to upload their provenance log created in step (1);
3. The analysis service triggers the provenance service to convert the log into triples (the semantic web representation of the same information);
4. The provenance service triggers a web service to put the triples into a semantic web database called RDF Store. Jena is Apache’s Java framework for building semantic web apps;
5. The RDF Store can be queried using SPARQL (RDF Query Language) via Joseki (a SPARQL server). The provenance metadata is ingested into a relational database (Apache Solr) after being applied domain rules and classified. This enables faceted navigation of the provenance data. For example, one can constrain two facets, a user name and a time range, to find out all the data products that this user has created during the time period. Once a data product is found, one can further browse its pedigree to find out what source data products were involved in making it, what parameters were used, who created it at what time, what algorithm in which version of the code was used to generate the data, etc.

In the above steps (1) and (2), we will revisit the CMDA analysis application code, decide what level of details we want to collect as provenance on behalf of our clients, and insert code to generate provenance log. In step (5), we will decide where to store the provenance and how our clients should use it. The possible choices are the following ones: (A) Use the entire existing JPL provenance system, including its RDF Store and Solr database. The advantage is that the users get to faceted browse all analysis results and their pedigree charts; (B) Shuffle the provenance associated with a result file into its metadata section. The format can be more human readable like the log file, or more computer readable like the triples. If the metadata area is not the best location for provenance, we can create a separate text file with the same metadata, and have these metadata files be distributed and co-located along with the actual analysis results. This option requires some coding efforts; and (C) Develop a light-weight desktop browser, reusing the code that is already developed for the JPL provenance system, that gives the users the ability to search and browse their own analysis results in their local environment. The advantage of this option is that the privacy of algorithms and approaches is respected before they are openly published.

### Scalable Job Distribution System of CMDA

Currently, CMDA individual analysis web services are optimized for the use of parallel computing resources but the distribution of multiple simultaneous service calls by multiple users is not optimized. CMDA will be used as an education tool for the Summer School organized annually by the JPL Center for Climate Science (details in Section 1.3.9). A heavy and interactive use of CMDA is expected during the Summer School. And, as the user base of CMDA grows, many simultaneous web service calls will overload the web server if scalable job distribution is not supported. Therefore, a scalable job distribution capability will be a critical enhancement over the current version of CMDA. There are many ways to achieve load balancing for web pages and web services.

We realize there is no “one-size-fits-all” solution and hence propose to apply three different approaches to our CMDA web services and determine the best approach for CMDA. For fast development and testing, the initial implementation of these approaches will be done quickly to check their performance. After the performance is compared, we will determine the best approach and implement only the best one thoroughly.

**DNS Load Balancing:** The first approach uses “DNS load balancing.” The key idea is to deploy multiple identical services at different IPs and route service requests from the clients evenly to those services. To do this, we need in our DNS (Domain Name System) service multiple “Address Record” to be associated to the single host name that the clients see. An Address Record points a domain or subdomain to an IP address. For example, 137.78.73.106 is the normal IP of the CMAC web server (cmacws.jpl.nasa.gov), but we can have multiple IPs, e.g., 137.78.73.106, 137.78.73.107, 137.78.73.108, each being a server hosting identical services, all associated to this server. When the clients access our customized DNS server, a list of all the Address Records are returned, but the order of the list varies for each request, so the clients can get services from any one of these IPs. This brings us some level of load balancing and fault tolerance. The algorithms used to determine the order of the A record list are usually round robin or geographically specific. Examples of open-source DNS servers providing load balancing and high availability are HAProxy, BIND, and lbnamed.

The DNS approach to load balancing is easy to implement, but this approach has its inherent drawbacks. The task distribution is not based on current load status of the servers, and the client side IP address caching and reuse may defeat the purpose. The DNS solution depends on how different kind of clients handle (cache, reuse, etc.) multiple IPs, and a Round Robin distribution does not guarantee balanced load. Some solution, such as lbnamed, improved by polling load information, but what is polled, i.e., load averages, total users, console users, and boot time, is only a rough estimation of the actual load on a server.

**Proxy Technology:** The second approach uses “proxy technology.” The basic idea is for a proxy server to make requests to actual back-end servers behind the scenes on behalf of the client. The client only knows the existence of the proxy server but not the multiple back-end servers behind it. Conversely, there is a “reverse proxy” technology from the standpoint of the actual back-end servers, because all they know about is that they are providing services to the proxy server and they have no knowledge of the real client out there.

An example of the open-source, proxy based software package is called the Zen Load Balancer. Zen Load Balancer works with a reverse proxy technology. It forwards the connection to the back-end servers, changing the client IP with the farm’s virtual IP. In the following, we use the steps to configure Zen Load Balancer to demonstrate how the proxy technology works to achieve load balancing: (1) Install Zen Load Balancer from a VirtualBox image. The guest OS is Debian (a free Linux distribution similar to Ubuntu); (2) Create a virtual IP for the physical IP of the above installation; (3) Create a farm, assigning the above virtual IP address and a port number (e.g., 80) to it. Add the IP addresses of the back-end servers to the farm. Now the requests sent to the URL will be forwarded to the back-end servers; (4) Create a cluster using two farms, one serving as the master and the other backup. This enables “failover,” meaning if the master farm is down for any reason, the backup will be automatically turned into the new master and start serving after a very brief break.

Another open-source proxy solution is by using Apache along with the modules mod\_proxy and mod\_proxy\_balancer. It provides simple load balancing. The hardware configuration consists of a server to be used as the load balancer and two or more servers hosting the web services. The configuration of the load balancer is through Apache’s Virtual Host configuration. In the <Proxy> tag, one can name the “balancer,” i.e., the URL that the clients will see, and add as many as “BalanceMember”s as desired, which are the back-end servers. The simplest balancing algorithm that can be set is “byrequest,” which means Round Robin.

The proxy technology approach is not scalable in handling client-server communication simply because all data between all the clients and the back-end servers go through the proxy server, which can become a bottleneck for performance. Similar to the DNS approach, load balance among the back-end servers is not guaranteed in this approach since the load information polled by the balancer is only a rough estimation of the actual load on a server.

**Pull and Queue based Proxy:** The third approach involves developing our own balancer proxy based on the pull strategy built on top of a queuing mechanism. In this approach, there are queues (“first in first out” mechanism) to link the “producers” and “consumers.” The “Task Queue” is where the back-end servers compete for new tasks received from the clients and enqueued by the proxy. The results from each and every back-end server are sent back to the proxy into the queues corresponding to the individual back-end servers. In this way, the proxy is able to return the results to their corresponding clients.

There are two advantages in this design that will overcome the two drawbacks of the two open-source solutions mentioned above: (1) The back-end servers pull to get new tasks from the Task Queue. This ensures true load balancing because idling servers are the first to dequeue the Task Queue to get new tasks if any; and (2) Only meta-data is communicated through the queues. For example, if a back-end server produces a large binary data file or image, only the URIs of these large results are enqueued into its result queue. These URIs point directly to the IP of the back-end server, so the requesting client can directly get the large results from the server, by-passing the proxy altogether. This makes the design scalable.

For the queuing mechanism we propose to use RabbitMQ, an open-source message broker software package that implements the Advanced Message Queuing Protocol (AMQP). The Co-I Dr. Pan has used this to construct software pipelines for a JPL project named MIRO.

### Scientific Use of CMDA

The Earth science focus area of the proposed work is in understanding and evaluating clouds, convection and their radiative processes represented in numerical weather and climate models. We have identified two scientific applications that will be used as a demonstration of the scientific use of CMDR in the context of climate model evaluation.

**(1) Evaluating representations of atmospheric moist process in the global models**

We will apply CMDA to evaluate the moist process parameterizations of the CMIP5 GCMs. The fidelity of climate and weather models depends critically on the representation of clouds and precipitation. It is very evident that large disagreement exists in clouds and precipitation represented in GCMs, and many of the uncertainties mainly originate from inadequate representations of atmospheric moist processes, such as moist convection, precipitation formation processes, sub-grid scale dynamics, hydrometeor contents, microphysics, and the associated radiation fluxes. The processes governing convection and precipitation formation involve a wide range of temporal and spatial scales. It is therefore a great challenge to accurately represent these interactions in atmospheric models, as some of the key processes cannot be explicitly resolved at the current resolution of global models. Parameterizations need to be carefully developed and extensively evaluated to account for the sub-grid processes. The synergistic use of the global observations across multiple sensors with proper co-location and merging can provide comprehensive and detailed information for evaluating the moist process representations in GCMs.

**(2) Evaluating representations of atmospheric radiation processes in the global models**

We will apply CMDA to evaluate the radiation fields in the CMIP5 GCMs. Representing atmospheric radiation interactions remains a pressing challenge to reduce and quantify uncertainties associated with present climate and climate change projections [Randall et al., 1991; 2007; Stephens et al., 2005]. Conventional GCMs (e.g., CMIP3/CMIP5 models) exhibit persistent systematic biases such as the overestimation of downward shortwave, underestimation of downward longwave at the land and ocean surfaces, and outgoing longwave at the top of the atmosphere in the precipitating and convectively active regions (e.g., ITCZ, Warm Pool). These biases might impose strong impacts on surface wind stress, SSTs, sea surface height, land surface air temperatures, soil moisture/temperatures, soil evaporation as well as on the atmospheric and oceanic circulations in the present climate and climate projection. Co-located and conditionally sampled satellite retrievals from multiple sensors and platforms will be a rich dataset to provide a global-scale distribution and long-term variability of these highly coupled parameters and to trace the impact of the model biases in the radiative process on the other climate parameters.

### Educational Use of CMDA

The Center for Climate Sciences (CCS) in JPL is planning to use the current version of CMDA as an educational tool for its summer school to be hosted in August 2013. The summer school annually brings together about 30 graduate students and postdocs from around the world to engage with premier climate scientists from JPL and elsewhere. Functional and user interface requirements for the educational tool have been defined according to its curriculum and expected use. We plan to continue this educational and outreach use of CMDA in the ACCESS award period. We will utilize this opportunity to promote the community awareness of CMDA and cultivate its user base. CCS will provide separate funding to develop a curriculum and to customize the user interface and functionality of CMDA for the education use.

As an example of the planned use of CMDA in summer school exercises, the students will be asked to use the monthly mean CMIP5 model outputs and Obs4MIPs datasets to compute and plot the zonal mean total cloud fractions and report model systematic biases if exist. The students will be also asked to consider the observational uncertainty and limitation and their effect on the comparison with model outputs. Another example is to study the relationship and correlation between cloud water content and precipitation rate or SST in the context of dynamic and aerosol conditions. The data analysis web services in CMDA provide capabilities needed for such multi-step diagnostic analyses and visualizations.

The user interface for the education tool will be a web browser based interface so that the initial learning curve is minimal, the compatibility with any operating systems of student machines is guaranteed, and no local installation of libraries or programs is needed. These advantages are found to be very critical from the previous experience of the summer school where an IDL based tool was used and its use was challenged by compatibility issues and steep learning curves. The web-browser based interface with web services providing analysis engines solves all of the problems encountered by the previous tool.

### Tool/Technology Life Cycle and Operations Concept

We will take the following multiple complementary approaches to achieve an effective total life-cycle development, sustainable operations, and a high Reuse Readiness Level (RRL 6 or higher) for successful infusion elsewhere in Earth science data systems:

* **Proven Technologies**: We will leverage only proven technologies and open source packages to minimize the development cost. The proposed work is based on the proven technologies such as the Web Service, the Cloud, the Provenance, and the Job Distribution technologies.
* **Earth Science Projects**: The proposal team members are involved in several NASA Earth Science Research Programs including MEaSUREs, MAP, COUND, and CMAC (details in Section 5). We will collaborate with the other projects to leverage the scientific and technological experiences and to cross-promote the products of the projects.
* **Educational Use**: We will use CMDA as an education tool for the JPL CCS Summer School (details in Section 1.3.9). The actual use of the tool during the proposed work period will significantly strengthen the usefulness of CMDA. We will use this outreach opportunity to promote the tool in the community. User feedbacks we will receive from the Summer School students will guide directions to polish the tool to be more useful for the community.
* **Open-source Deployment**: We will create an open source project for CMDA early in the ACCESS award period with the current version of CMDA in order to engage the open source community to use and maintain the capabilities of CMDA. We will release a new version of CMDA each time when some key milestones scheduled for the proposed work are met. We will document installation and developer guides, API specifications, and tutorials.
* **Modeling and Model Analysis Community**: The Co-I Dr. Jiang and Collaborator Dr. Teixeira are actively involved in many modeling and model analysis programmatic entities. With the connections, we will engage the modeling and model analysis community to the proposed project, and promote the community awareness and user-base growth of CMDA.
* **Reuse Readiness Level 6**: We will design CMDA for extensibility, modularity, and portability. In order to achieve high extensibility, we will use the general interface and structure for each functional code and give special attention to avoid any ad-hoc hard-coded statements. For modularity, we will encapsulate well-defined separate functional groups. For portability, we will avoid any platform-specific technologies and use programming languages and libraries that can be compiled for various computing platforms. Our web-browser user interface code will be written and tested to support all popular web browsers.

## Perceived Impact to State of Knowledge

Information systems that facilitate individual NASA data product discovery and processing services are becoming more available (e.g. GCMD, ECHO, ESG, Giovanni). However, information systems that support the comprehensive processing of multi-source data products and the synergistic use of multiple-instrument measurements for model evaluations and diagnoses are limited in availability and insufficient in capability. Most of the existing climate data analysis tools that are available for a community use (e.g. PCMDI’s CDAT, NCAR’s CCMVal) focus on model output analyses and visualization and lack the capability to process multi-instrument observational data in combination with model outputs (e.g. co-locate and conditionally sample multiple-instrument data to one another, and use them to evaluate and diagnose model performance). Certainly, there exist many in-house research tools for such analyses but they are not designed for a community use. Those tools are highly domain specific, thus not general enough for applying to other investigations, and are characterized by low software quality and limited reuse and sharing.

The proposed system CMDA will provide the modeling and model analysis community with capabilities they need for model-data intercomparisons and model diagnostic analyses. CMDA will enable multi-source dataset processing and model performance diagnostic analysis using satellite observational datasets, reanalyses, and model outputs. The proposed system will be complementary to those climate model analysis tools currently available to the community such as CDAT and CCMVal. CMDA is becoming a community tool as it will be used as an educational tool for the summer school organized by JPL CCS this summer. With the excellent outreach opportunity, it will produce immediate and tangible benefits to the young generation of the modeling and model analysis community.

Scientifically, the multi-variate model-to-data comparisons and analyses are critical in understanding the causes of model errors and biases and improving model performance for climate sensitivity studies and predictions. A simple one-parameter comparison with a traditional model analysis tool can illustrate the symptomatic biases and errors in models but cannot trace back to the cause of the biases errors. For example, a study of the instantaneous interaction and relationship between physical variables in the context of its environmental condition using the conditional sampling method will be needed to gain new insight into possible errors in model physics and parameterizations. CMDA will provide capabilities to conduct such a study.

Technologically, CMDA will demonstrate the first implementation of an information system that utilizes parallel, distributed web services with interconnectivity, provides provenance capabilities, and delivers a scalable job distribution. The individual technologies exist but the technical challenges will be to incorporate them into one system with efficient connections and communications. Approaches we will take to integrate these components into a single system will represent a blueprint for future Earth science information systems that will have these multifaceted capabilities.

In terms of utilities, CMDA will give very tangible benefits to users in the modeling and model analysis community. We recognize that multi-sensor, multi-parameter, and multi-variable analyses are difficult; multi- source datasets are rarely designed to work together, science analyses are usually very “free form,” exploratory with many “what if” scenarios, and processing algorithms are complex and ever changing. The data analysis system with interconnectivity will enable a multi-step analysis involving sequential multiple web service calls without manual interruption for file and input and output exchanges. The provenance support system will significantly ease the burden of keeping track of the processing history of many datasets and plots that the users generate during their exploratory studies by producing provenance that is human readable, searchable, browsable, and citable. The scalable job distribution system will provide a reliable and optimal use of the computational resources that are seemingly never sufficient for all the computational needs for the users.

## Relevance to Element Programs and Objectives in the NRA

The proposed work responds to one focus area of the NRA – ***Tools that improve and expand the accessibility and usability of NASA’s Earth Science observational data for the modeling and model analysis communities***, and addresses several objectives of the NRA:

* We will *enhance, extend, and improve existing components of NASA’s distributed and heterogeneous data and information systems infrastructure* by leveraging NASA-funded proven and existing technologies and datasets (the current version of CMDA, the JPL provenance system, and Obs4MIPs datasets) and adding critical new capabilities (i.e. multi-step analysis support, provenance support, scalable job distribution).
* We will *deploy data and information capabilities that enable the freer movement of data and information within NASA’s distributed environment of providers and users* by providing web services to access and analyze Obs4MIPS datasets and CMIP5 model outputs.
* We will *address the existing needs of research and applied science communities* by providing a tool to perform multi-step analyses for model-data intercomparisons and model diagnoses, to automatically generate data processing history, and to trace and search processed datasets through the data processing history.
* We will *increase the utilization of NASA Earth science observational data by modeling and model analysis communities* by providing a tool to access and analyze Obs4MIPs datasets in direct comparison with model outputs.
* We will *reuse existing technological solutions in the support of Earth science data and information needs* by developing the proposed system on top of the existing system (CMDA) and adding new capabilities that are based on the proven and existing technologies.

## Work Plan

In Year 1, we will first develop the scalable job distribution system of CMDA and test and benchmark its performance in the scalability. This task is scheduled first in order to make CMDA reliable for its education use for the JPL CCS Summer School scheduled in August 2014. Next, we will enhance the existing data analysis system of CMDA by adding the interconnectivity capabilities provided by the Cloud technology. We will experiment several server/client configurations to achieve optimal network communication efficiency and at the same time simple application-level implementation, which will allow the extension and update of the tool in the future easy. We will also create an open source project for CMDA early in Year 1 and release the latest version of CMDA to the open source community at the end of Year 1.

In Year 2, we will develop the provenance support system of CMDA. The first task for the provenance support system is to instrument the existing analysis web service code to generate a provenance log. Next, we will adapt and apply the existing provenance system (developed for the ACCESS-2009 project: PI Hook Hua) to the proposed tool CMDA. We will adapt a provenance representation service that convert the provenance log into triples and convert the triples into a semantic web database, and adapt a provenance search system that enables faceted navigation and search of provenance data. We will release the latest version of CMDA source code to the open source community toward the end of Year 2. We will also document the technical approaches we took to develop CMDA and publish in a peer-reviewed journal and present the result to the Earth science information system development community.

During the two-year work period, we will demonstrate the education use of CMDA in two summer schools organized by JPL CCS. CCS will provide separate funding to develop a curriculum and to customize the user interface and functionality of CMDA for the education use. The work for the summer school and the proposed work will be coordinated so that the key milestones and capabilities for the proposed ACCESS work are aligned with those for the CCS summer school work.

**The Co-I Dr. Pan will participate in the Technology Infusion Working Group (TIWG), which is one of Earth Science Data System Working Groups (ESDSWG). Dr. Pan will bring his expertise in the state-of-art web service, distributed computing, and provenance technologies to TIWG and contribute to TIWG’s activities in defining processes to infuse new technologies into the evolving Earth science data systems.**



**Figure 2. Work schedule of the proposed work with key milestones**

### Key Milestones

Five key milestones are planned for the proposed work:

1. Implementation of scalable job distribution system;
2. Implementation of data analysis system with interconnectivity;
3. Implementation of provenance support system;
4. Implementation of fully integrated CMDA;
5. Demonstration of educational use of CMDA.

Figure 2 shows the workflow to achieve the key milestones (shown as diamonds).

### Management Structure and Team Member Contributions

**Management Plan:** Dr. Seungwon Lee, the PI of the proposed investigation, is responsible for the quality and direction of the proposed research and the proper use of all awarded funds. She is also responsible for all technical, management, and budget issues and is the final authority for this task. The Co-Is will provide all the contributions needed to ensure that she can effectively manage the entire task. The proposal team will host a weekly meeting to discuss task progress and coordination. The proposal team has been working together for several years and has developed an efficient working relationship from the current/prior NASA research grants.

**Dr. Seungwon Lee** of JPL, PI, will supervise the work described in this proposal. She has primary responsibility for defining the application scope of the proposed system and the requirements of its capabilities by seeking inputs from the targeted research community. By leveraging her combined experience of both the information technology development and the Earth science investigations, Dr. Lee will facilitate effective interactions between the Earth scientists and the information technologists.

**Dr. Lei Pan** of JPL, Co-I, will be the lead architect and developer of the proposed information system. He is the information technology expert in the proposal team. He has designed the architecture of the existing CMDA, and developed and implemented it. He has extensive experience in numerical methods, large-scale software for computer-aided engineering, high performance computing, distributed parallel programming methodologies, and Web Service technologies. He will be responsible for the architecture, design and development of CMDA.

**Dr. Chengxing Zhai** of JPL, Co-I will be the lead developer of the data analysis system of CMDA. He has developed many model analysis tools for observational data processing, model-data intercomparisons, which are the building blocks of the data analysis system of CMDA. He will work together with Dr. Pan to implement the interconnectivity of the data analysis system and the instrumentation of the data analysis system for provenance generation.

**Dr. Jonathan Jiang** of JPL, Co-I, will lead the effort of applying the developed CMDA to scientific investigations, validating results, and demonstrating scientific values of CMDA. He is the lead scientist of the proposal team. He has led and contributed to several important publications in using satellite datasets for model evaluations and model-data comparisons. He will leverage the expertise gained from the research projects to maximize the scientific usefulness of CMDA for our target users. He is also the instructor of the JPL CCS summer school who will use CMDA as an educational tool to demonstrate the use of satellite data for climate studies and model-data intercomparisons. He will provide the scientific inputs to customize CMDA for the summer school use.

**Dr. Joao Teixeira** of JPL, Collaborator, will serve as the main liaison between this proposal team and the climate modeling and model analysis community. He has extensive research experience in climate model parameterization of the atmospheric boundary layer and evaluation of the climate model parameterization. He is the deputy director of JPL Center for Climate Science (CCS). He also holds a leading role in several international weather and climate programmatic entities including Process Studies and Model Improvement Panel in US CLIVAR, GCSS/WGNE Pacific Cross-Section Intercomparison Working Group for WCRP, and GEWEX Cloud System Study Scientific Steering Committee. With the leadership roles, he will cultivate a close interaction between the proposal team and the climate modeling and model analysis community and promote the community awareness and the user-base growth of CMDA.

# References and Citations

Owncloud.com, “ownCloud Architecture Overview,” owncloud.com online document.

The ownCloud developers, “ownCloud User Manual, Release 5.0,” owncloud.com online document.

The ownCloud developers, “ownCloud Administrators Manual, Release 5.0,” owncloud.com online document.

Hook Hua, Brian Wilson, Gerald Manipon, Lei Pan, and Eric Fetzer, 2012: Tracking Production Legacy of a Multi-sensor Merged Climate Data Record, ACCESS-09 project final review, Oct 9, 2012.

Hook Hua, Brian Wilson, Gerald Manipon, Lei Pan, and Eric Fetzer, 2012: Improving the Understanding of Provenance and Reproducibility of a Multi-Sensor Merged Climate Data Record, Abstract submitted to IPAW 2012 - 4th International Provenance and Annotation Workshop, Santa Barbara, CA, June 19-21, 2012.

NRC, 2007: Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond, National Research Council, National Academies Press, 2007.

Randall, D. A., et al. (2007): Climate models and their evaluation, in Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by S. Solomon et al., pp. 589 – 662, Cambridge Univ. Press, Cambridge, U. K.

# Acronyms

GEWEX Global Energy and Water Exchange Project

Obs4MIPs A pilot activity to make observational products more accessible for climate model intercomparisons

ESDSWG

TIWG

# Biographical Sketches

## Principal Investigator

**Dr. Seungwon Lee**

Jet Propulsion Laboratory, M/S 168-200, 4800 Oak Grove Drive, Pasadena, CA 91109; [Seungwon.Lee@jpl.nasa.gov](mailto:Seungwon.Lee@jpl.nasa.gov); (818) 393-7720; http://hpc.jpl.nasa.gov/people/seungwon

**EDUCATION**

Ph. D. Physics Ohio State University 2002

M.S. Physics Ohio State University 2001

B.A. Physics Seoul National University 1995

**RELEVANT EXPERIENCE**

Dr. Lee has over 15 years of experience of research and technology development in computational physics, dynamic systems, statistical data analysis, optimization algorithms, information systems, and high performance computing systems for science and engineering applications. She is a PI of the NASA ROSES COUND project for evaluating CMIP5 models for cloud-climate feedbacks, and a PI of the NASA ROSES CMAC project for developing a web-service based climate model analyzer. She is a Co-I of the NASA US-Rosetta for MIRO instrument forward and retrieval model development. She led several research projects as a PI under the JPL internal Research and Technology Development program, which did innovative research on Earth science data summarization with advanced clustering algorithms, evolutionary optimization methods for spectral retrieval problems and trajectory design, parallel computing for Satellite Orbit Analysis Program. She served as a Co-I of a NEWS-data web services project (led by Mr. Hook Hua) under the NASA ROSES ACCESS program and a Co-I of EPISODE project (led by Dr. Jeffrey Jewell) under the NASA ROSES AISR program and. She also contributed to other NASA research projects including the development of a co-location tool for A-Train data and ECMWF outputs, the characterization of cloud properties using NASA’s observational data, development of cloud-related parameterizations for climate models, the development of a statistical method to quantify the relationships between passive radiometer and active radar measurements for hurricane studies, the development of radiative transfer models for water-level populations of cometary coma for the Microwave Instrument for Rosetta Orbiter, and the development of quantification of trace chemicals with a gas chromatograph and mass spectrometer system for Vehicle Cabin Atmosphere Monitor. She is the first author or co-author of over 40 refereed publications.

**PROFESSIONAL EXPERIENCE**

**2003 – present:** Researcher – High Capability Computing and Modeling Group

Science and Technology Directorate, Jet Propulsion Laboratory, Pasadena CA

**2002 – 2003**: Postdoctoral Fellow, Department of Chemistry

University of California, Berkeley CA

**SELECTED PUBLICATIONS**

**S. Lee**, J-L. Li, D. Waliser, J. Teixeira, T. Kubar, W-T. Chen, and H-Y. Ma, Evaluation of cloud liquid water content simulations in CMIP3 and CMIP5 GCMs and analysis using A-Train satellite observaitons, WCRP Workshop on CMIP5 Climate Model Analyses, Honolulu, Hawaii, 5-9 March, 2012.

S. Gulkis, S. Keihm, L. Kamp, **S. Lee**, et al., Continuum and spectroscopic observations of asteroid (21) Lutetia at millimeter and submillimeter wavelengths with the MIRO instrument on the Rosetta spacecraft, Planetary and Space Science, doi:10.1016/j.pss.2011.12.004, 2012

**S. Lee**, Paul von Allmen, Lucas Kamp, Samuel Gulkis, and Bjorn Davidsson, Non-LTE radiative transfer for sub-millimeter water lines in Comet 67P/Churyumov-Gerasimenko, Icarus, Vol 215, Issue 2, Pages 721-731, doi:10.1016/j.icarus.2011.07.007, 2011.

**S. Lee**, L. Pan, R. Morris, P. von Allmen, W.-T. Chen, T. Kubar, J.-L. Li, and D. Waliser, Co-locating A-Train observations and ECMWF analysis outputs for comprehensive understanding of the Earth’s weather and climate, A-Train Symposium, New Orleans, October 25-28, 2010.

**S. Lee,** A. Guillanme, D. Waliser, R. Ferraro, and J. Teixeira, Systematic and characteristic differences between NASA observational data and CMIP5 outputs for model evaluations, A-Train Symposium, New Orleans, October 25-28, 2010.

**S. Lee**, B. Kahn, and J. Teixeira, “Characterization of Cloud Liquid Water Content Distributions from CloudSat,” J. Geophys. Res., 115, D20203, doi:10.1029/2009JD013272, 2010.

B. Davidsson, S. Gulkis, C. Alexandre, P. von Allmen, L. Kamp, **S. Lee**, and J. Warell, “Gas kinetics and dust dynamics in low-density comet comae,” Icarus, 210, 455-471, 2010.

S. Gulkis, S. Keihm, L. Kamp, C. Backus, M. Janssen, **S. Lee,** et al., Millimeter and submillimeter measurements of asteroid (2867) Steins during the Rosetta fly-by, Planetary and Space Science, 58, 1077-1087, 2010.

**S. Lee**, L. Mandrake, B. Bornstein, and B. Bue,“Quantification of Trace Chemicals using Vehicle Cabin Atmosphere Monitor,” IEEE Aerospace Conference Proceedings, Big Sky, Montana, March, 2009

H. Hua, E. Fetzer, A. Braverman, **S. Lee**, M. Henderson, V. Dang, M. de la Torre Juarez, and S. Lewis, “Web Services for Custom Level 2 Data Subsetting and Level 3 Data Summarization of Merged A-Train Data,” IEEE Aerospace Conference Proceedings, Big Sky, Montana, March, 2009.

P. von Allmen, **S. Lee**, L. Kamp, S. Gulkis, “Molecular Excitation and Radiative Transfer Model for MIRO,” IEEE Aerospace Conference Proceedings, Big Sky, Montana, March, 2009.

**S. Lee** and B. Bornstein, "Autonomous Calibration of Vehicle Cabin Atmosphere Monitor," IEEE Aerospace Conference Proceedings, Big Sky, Montana, 2008

A. Guillaume, **S. Lee**, A. Braverman, and R. Terrile, "Entropy Constrained Clustering Algorithm Guided by Differential Evolution," IEEE Aerospace Conference Proceedings, Big Sky, Montana, 2008

R. J. Terrile, **S. Lee**, G.Tinetti, W. Fink, T. Huntsberger, and P. von Allmen, Evolutionary Computational Methods for Spectral Instruments Design," IEEE Aerospace Conference Proceedings, Big Sky, Montana, 2008

J. B. Jewell, **S. Lee**, M. Lo "A Probabilistic Approach to Trajectory Generation in the Presence of Uncertainty," NASA Science Technology Conference, Adelphi, Maryland, June 19-21, 2007

**S. Lee**, H. Hua, R. Carnright, J. Coggi, and D. Stodden, "Lessons Learned from Adapting Aerospace Engineering Tools to the Parallel and Grid Computing Environment," IEEE Aerospace Conference Proceedings, Big Sky, Montana, 2007.

A. Guillaume, **S. Lee**, H. Zheng, S. Chau, Y-W Tung, Y-F Wang, and R. J. Terrile, "Deep Space Network Scheduling Using Evolutionary Computational Methods," IEEE Aerospace Conference Proceedings, Big Sky, Montana, 2007.

**S. Lee** and P. von Allmen, "Magnetic-field dependence of valley splitting for Si quantum wells grown on tilted SiGe substrates," Phys. Rev. B 74, 245302, 2006.

**S. Lee** and Paul von Allmen, "Tight-Binding Modeling of Thermoelectric Properties of Bismuth Telluride," Applied Physics Letters, 88, 022107, 2006.

**S. Lee** and R. P. Russell, "Multi-Objective Parallel Genetic Algorithms Applied to the Primer Vector Control Law," AAS 06-196 Paper, AAS/AIAA Space Flight Mechanics Meeting, Tampa, Florida, January 22-26, 2006.

**S. Lee**, C. H. Lee, S. Kerridge, C. D. Edwards, and K-M Cheung "Orbit Design and Optimization Based on Global Telecommunication Performance Metrics," IEEE Aerospace Conference Proceedings, Big Sky, Montana, 2006.

## Co-Investigators

**Dr. Lei Pan**

Jet Propulsion Laboratory, California Institute of Technology, MS 168-200, 4800 Oak Grove Drive, Pasadena, CA 91109-8099; (818) 393-0477, [Lei.Pan@jpl.nasa.gov](mailto:Lei.Pan@jpl.nasa.gov)

**RELEVANT EXPERIENCE**

Over 20 years of experience of research and technology development in numerical methods, large-scale software for computer-aided engineering, high performance computing applications, distributed parallel programming methodologies, Web Service technologies, and applying provenance in the semantic web. He had been with The MSC Software Corp., CA for 6 years and led several projects that built software components for MSC.Patran and MSC.Nastran. He led or participated JPL R&TD projects aimed at porting legacy software packages, such as ROI PAC, IMOS, and GEOS-Chem, to modern cluster computers. He is the original developer of the concept and methodology of Navigational Programming. He architected the ATrain co-location tool, and developed its core algorithm – the linear nearest neighbor search. He is a core developer of the RESTful web services for correcting atmosphere effects in radar, a 3-year NASA ROSES AIST project, a core developer of a 2-year ACCESS Provenance project, and an architect and core developer of a 2-year CMAC project, which serves as the basis of this work.

**EDUCATION**

• Ph.D., Information & Computer Sciences, University of California, Irvine (UCI), 2005

• M.S., Information & Computer Science, University of California, Irvine (UCI), 2002

• M.S., Mechanical Engineering, Rensselaer Polytechnic Institute (RPI), Troy NY, 1996

• M.S., Mechanical Engineering, Zhejiang University (ZJU), Hangzhou China, 1989

• B.S., Mechanics, Peking University (PKU), Beijing China, 1987

**PROFESSIONAL EXPERIENCE**

2005 – present: Senior Member of Technical Staff, Jet Propulsion Laboratory

2003 – 2005: Member of Technical Staff, Jet Propulsion Laboratory

1996 – 2002: Senior Software Engineer, The MSC Software Corp., Santa Ana, CA

1992 – 93: Chief Public Relations Rep., Hangzhou Int’l Public Relations Inc., Hangzhou, China

1990 – 1992: Software Engineer, National Institute of Project Planning, Hangzhou, China

**SELECTED PUBLICATIONS**

H. Hua, B. Wilson, G. Manipon, L. Pan, and E. Fetzer, "Improving the Understanding of Provenance and Reproducibility of a Multi-Sensor Merged Climate Data Record," Abstract, IPAW 2012 - 4th International Provenance and Annotation Workshop, Santa Barbara, CA, June 19-21, 2012.

L. Pan, J. Xue, M. B. Dillencourt, and L. F. Bic, “Toward automatic data distribution for migrating computations,” International Conference on Parallel Processing (ICPP 07), Xian, China, Sept., 2007.

L. Pan, M. K. Lai, M. B. Dillencourt, and L. F. Bic, “Mobile pipelines: Parallelizing left-looking algorithms using navigational programming,” in Proceedings, 12th Int’l Conference on High Performance Computing - HiPC 2005, ser. Lecture Notes in Computer Science, Dec. 2005.

L. Pan, W. Zhang, A. Asuncion, M. K. Lai, M. B. Dillencourt, and L. Bic, “Incremental parallelization using navigational programming: A case study,” in Proceedings of the 2005 Int’l Conference on Parallel Processing (ICPP 2005), Oslo, Norway, June 2005, pp. 611–620.

L. Pan, M. K. Lai, K. Noguchi, J. J. Huseynov, L. Bic, and M. B. Dillencourt, “Distributed parallel computing using navigational programming.” Int’l Journal of Parallel Programming, vol. 32, no. 1, pp. 1–37, 2004.

**Dr. Jonathan H. Jiang**

E-mail: [Jonathan.H.Jiang@jpl.nasa.gov](mailto:joe@mls.jpl.nasa.gov) Address: Mail stop 183-701  
Phone: 818-354-7135 Jet Propulsion Laboratory  
Fax: 818-393-5065 Pasadena, CA 91109

**Education**

* Ph.D. in Atmospheric Physics, York University (1996)
* M.Sc. in Astrophysics, York University (1991)
* B.Sc., in Astrophysics, Beijing Normal University (1985)

**Professional Experience**

* **Research Scientist** (2004-), Jet Propulsion Laboratory, California Institute of Technology, U.S.A.

Major responsibilities include:

* Funded Research Projects:
* NASA Aura Science Team: Principal Investigator (2011-)
* NASA COUND Program: Principal Investigator (2011-)
* NASA MAP Program: Principal Investigator (2013-)
* NASA CloudSat/CALIPSO Science Team: Co-Investigator (2009-)
* NASA ACMAP: Co-Investigator/Principal Investigator (2002-2010)
* NASA IDS Program: Co-Investigator/Principal Investigator (2007-2010)
* Aura MLS project: Lead, MLS upper tropospheric measurements summary/inspection (2005-)
* **Scientist** (2001-2003), Jet Propulsion Laboratory, California Institute of Technology, U.S.A.

Major responsibilities: MLS cloud forward model development, cloud retrieval using satellite obs.

* **Caltech Postdoctoral Scholar** (1999-2000), California Institute of Technology, USA
* **Research Associate** (1998-1999), Université du Québec à Montréal, Canada
* **Postdoctoral Research Fellow** (1996-97), McGill University, Canada

**Selected Awards**

* **NASA Exceptional Achievement Medal** *for pioneering a new approach to quantifying the impact of air*

*pollution on clouds and climate, through combing observations from multiple NASA satellites (2010)*

* **NASA TC4 Team Achievement Award** *for outstanding contribution to NASA TC4 field experiment (2008)*
* **NASA Group Achievement Awards** *for contribution to Aura Microwave Limb Sounder Project (2005, 2006)*
* **NASA Space Act Award** *for significant contribution to MLS cloud forward model and level 2 software (2005)*

**Scientific Publications** **(89 peer-reviewed scientific publications to date; 2283 citations; H-Index 29)**

* Su, H., **J.H. Jiang**, Tropical Clouds and Circulation Changes During the 2006-07 and 2009-10 El Niños, *J. Climate*, 26, 2, doi:10.1175/JCLI-D-1200 152.1, 2013.
* **Jiang, J.H.**, et al., Evaluation of Cloud and Water Vapor Simulations in CMIP5 Climate Models Using NASA A-Train Satellite Observations, *J. Geophys. Res*. 117, 10.1029/2011JD017237, 2012.
* **Jiang, J.H.**, et al., Influence of convection and aerosol pollution on ice cloud particle effective radius, *Atmos. Chem. Phys.* 11, 457-463, doi:10.5194/acp-11-457-2011, 2011.
* **Jiang, J.H.** et al., Five-year Observations of Upper Tropospheric Water Vapor and Cloud Ice from MLS and Comparisons with GEOS-5 analyses, *J. Geophys. Res.* 115, doi:10.1029/2009JD013256, 2010.
* L'Ecuyer, T., **J.H. Jiang**, Touring the atmosphere aboard the A-Train, *Phys. Today,* 63, 7, 36-41, 2010.
* **Jiang, J.H.**, et al. Aerosol-CO relationship and aerosol effect on Ice cloud particle size: Analyses from Aura MLS and Aqua MODIS observations, *J. Geophys. Res.* 114, doi: 10.1029/2009JD012421, 2009.
* **Jiang, J.H.**, et al., Clean and polluted clouds: relationships among pollution, ice cloud and precipitation in South America, *Geophys. Res. Lett.,* 35, doi:10.1029/2008GL034631, 2008.
* **Jiang, J.H.**, et al., Connecting surface emissions, convective uplifting, and long-range transport of carbon monoxide in the upper-troposphere: New observations from the Aura MLS, *Geophys. Res. Lett*. 34, doi:10.1029/2007GL030638, 2007.

**Dr. Chengxing Zhai**

Jet Propulsion Laboratory, California Institute of Technology,

4800 Oak Grove Dr, Pasadena, CA 91109, U.S.A.

Phone: (818) 393-0758 Email: Chengxing.Zhai@jpl.nasa.gov

**Education**

**B.S.**,Physics, Peking University, China, 1989

**M.S.**,Physics, University of Washington, 1991

**Ph.D.**, Physics, University of Washington, 1994

**Work Experience**

2004-present: Technologist, Jet Propulsion Laboratory, California Institute of Technology, CA

2000–2004: Software Engineer, Jet Propulsion Laboratory, California Institute of Technology, CA

1999-2000: Software Engineer, Hughes Space and Communications Company, El Segundo, CA.

1997–1999: Software Engineer, Raytheon System Company, El Segundo, CA

**Relevant Experience**

Dr. Zhai has 20 years of experience in modeling, data processing and analysis, and algorithm development. He is specialized in estimation, statistical analysis, spectral analysis, signal processing and detection, retrieval algorithm development, and instrument calibration technology. In the past three years, he has been working on evaluating the CMIP5 General Circulation Models (GCMs) performance in simulating current climate using A-train satellite observations and analyzing model processes relevant to climate sensitivity. He has performed independent quality checks on the NASA data and technical notes for the obs4MIPs project, and thus he is intimately familiar with the requirements. He is currently a Co-I of project, Parallel Web-Service Climate Model Diagnostic Analyzer, funded by NASA ROSES/CMAC program.

**Recent Honors and Awards**

* NASA Group Achievement Award for demonstrating micro-arc-second astrometry capability using SIM Spectral Calibration Development Unit, 2011.
* NASA Group Achievement Award for studying exoplanet system detection using astrometry and radial velocity 2010.
* JPL Ranger Award for supporting SIM project, 2008.

**Selected Recent Peer-Reviewed Publications**

* 1. Su, H., J.H. Jiang, **C. Zhai**, et al., Diagnosis of Regime-dependent Cloud Simulation Errors in CMIP5 Models Using A-Train Satellite Observations, J. Geophys. Res., doi:10.1029/2012JD018575, 2013.
  2. Jiang, J.H., H. Su., **C. Zhai**, A. Del Genio, et al. Evaluation of Cloud and Water Vapor Simulations in CMIP5 Climate Models Using NASA “A-Train” Satellite Observations, J. Geosphys. Res., 118, doi:10,1029/ 2011JD017237, 2012.
  3. Jiang, J.H., H. Su, **C. Zhai**, S.T. Massie, M.R. Schoeberl, P.R. Colarco, S. Platnick, Y. Gu, and K.-N. Liou, Influence of convection and aerosol pollution on ice cloud particle effective radius, *Atmos. Chem. Phys*., 11, 457-463, 2011.
  4. **Zhai, C**., M. Shao, G. Goullioud, and B. Nemati, Micro-pixel accuracy centroid displacement estimation and detector calibration, Proc. R. Soc. A 8, vol. 467 no. 2136 3550-3569 arXiv:1102.2248v2, 2011.
  5. Small, J., J.H. Jiang, H. Su, and **C. Zhai**, Relationship between aerosol and cloud fraction over Australia, Geophys. Res. Lett. 38, L23802, doi:10.1029/ 2011GL049404, 2011.

## Collaborator

**Dr. João Teixeira**

Jet Propulsion Laboratory

4800 Oak Grove Drive, MS 169-237, Pasadena, CA 91109

[teixeira@jpl.nasa.gov](mailto:teixeira@jpl.nasa.gov)

626-395-6720

**EDUCATION**

Ph.D., Meteorology, *University of Lisbon*, Lisbon, Portugal, 2000.

Licentiate, Geophysical Sciences, *University of Lisbon*, Lisbon, Portugal, 1992

**RESEARCH AND PROFESSIONAL EXPERIENCE**

2010-present Deputy Director, JPL Center for Climate Sciences, Pasadena, CA

2011-present AIRS Science Team Leader, JPL, Pasadena, CA

2010-present Supervisor, Climate Physics Group, Jet Propulsion Laboratory, Pasadena, CA

2008-present Research, Principal Scientist, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

2005-2007 Senior Scientist, NATO Undersea Research Centre, La Spezia, Italy

2000-2005 UCAR Scientist, Naval Research Laboratory, Monterey, CA

1993-1999 Scientist, European Centre for Medium-Range Weather Forecasts, UK

**PROFESSIONAL ACTIVITIES**

Chair, ONR Science Team “Unified Parameterizations for Seasonal Prediction,” 2011-present

Member, Scientific Steering Committee, US CLIVAR, 2010-2012

Co-chair, Process Studies and Model Improvement Panel, US CLIVAR, 2010-2012

Lead Principal Investigator, NOAA Climate Process Modeling Team, 2010-present

Member, Process Studies and Model Improvement Panel, US CLIVAR, 2009-2010

Member, JIFRESSE Executive Committee, 2008-present

Chair, GCSS Pacific Cross-Section Intercomparison (GPCI) working group, 2005-2012

Member, GEWEX Cloud System Study (GCSS) Scientific Steering Committee, 2004-2010

**SELECTED RECENT PUBLICATIONS**

Teixeira and 37 co-authors, Tropical and sub-tropical cloud transitions in weather and climate prediction models: The GCSS/WGNE Pacific Cross-section Intercomparison (GPCI), *J. Clim.*, **24**, 5223-5256.

Teixeira, J., B. Stevens, C. S. Bretherton, R. Cederwall, J. D. Doyle, J. C. Golaz, A. A. M. Holtslag, S. A. Klein, J. K. Lundquist, D. A. Randall, A. P. Siebesma, & P. M. M. Soares, The parameterization of the atmospheric boundary layer: a view from just above the inversion, *Bull. Amer .Meteor. Soc.*, **89**, 453-458, 2008.

Teixeira, J., & C. A. Reynolds, Stochastic nature of physical parameterizations in ensemble prediction: a stochastic convection approach, *Mon. Weath. Rev.*, **136,** 483-496, 2008.

Teixeira, J., P. May, M. Flatau, & T. F. Hogan, On the sensitivity of the SST from a global ocean-atmosphere coupled system to the parameterization of boundary layer clouds, *J. Mar. Sys.*, **69**, 29–36, 2008.

Teixeira, J., C. Reynolds, & K. Judd, Time-step sensitivity of non-linear atmospheric models: numerical convergence, truncation error growth and ensemble design. *J. Atmos. Sci.*, **64**, 175-189, 2007.

# Current and Pending Support

## Current Awards

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PI: Dr. Seungwon Lee Current Support | | | | |
| PI Name | Award/Project Title | Program Info | Period & Total Budget | Commitment (Work Year) |
| Seungwon Lee | Evaluating CMIP5 models with satellite observations: cloud-climate feedbacks | NASA ROSES COUND program  Dr. Tsengdar J. Lee  Tsengdar.J.Lee@nasa.gov | 9/1/2011-8/31/2013  $420K | 0.20 |
| Seungwon Lee | Parallel Web-Service Climate Model Diagnostic Analyzer | NASA ROSES CMAC program  Dr. Tsengdar J. Lee  Tsengdar.J.Lee@nasa.gov | 10/1/2012-09/30/2014  $616K | 0.20 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Co-I: Dr. Jonathan Jiang Current Support | | | | |
| PI Name | Award/Project Title | Program Info | Period & Total Budget | Commitment (Work Year) |
| Jonathan Jiang | Utilizing Aura MLS and A-Train datasets to analyze and evaluate IPCC AR5 models in the upper troposphere | NASA Aura Science Team  Dr. Kenneth W. Jucks  Kenneth.W.Jucks@nasa.gov | 04/2011-03/2014  $500K | 0.15 |
| Jonathan Jiang | Utilizing NASA A-Train datasets for IPCC AR5 climate model evaluation | NASA ROSES COUND program  Dr. Tsengdar J. Lee  Tsengdar.J.Lee@nasa.gov | 10/1/2011-09/2013  $470K | 0.15 |
| Hui Su | Investigating the Influence of Asian Aerosol Pollution on Water Vapor Transport from Troposphere to Stratosphere | NASA Aura Science Team  Dr. Kenneth W. Jucks  Kenneth.W.Jucks@nasa.gov | 06/2011-05/2014  $692K | 0.10 |
| Hui Su | Using NEWS Water and Energy Cycle Products to investigate processes that control cloud feedback. | NASA ROSES NEWS program  Dr. Jared Entin  Jared.K.Entin@nasa.gov | 07/2011-08/2013  $259K | 0.10 |
| Rong Fu | Exploring the impacts of climate variations and land use on interannual changes of CO in the tropical tropopause layer using multi-year Aura and A-Train measurements | NASA Aura Science Team  Dr. Kenneth W. Jucks  [Kenneth.W.Jucks@nasa.gov](mailto:Kenneth.W.Jucks@nasa.gov) | 06/2011-05/2014  $542K | 0.05 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Co-I: Dr. Lei Pan Current Support | | | | |
| PI Name | Award/Project Title | Program Info | Period & Total Budget | Commitment (Work Year) |
| Seungwon Lee | Parallel Web-Service Climate Model Diagnostic Analyzer | NASA ROSES CMAC program  Dr. Tsengdar J. Lee  Tsengdar.J.Lee@nasa.gov | 10/1/2012-09/30/2014  $616K | 0.50 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Co-I: Dr. Chengxing Zhai Current Support | | | | |
| PI Name | Award/Project Title | Program Info | Period & Total Budget | Commitment (Work Year) |
| Seungwon Lee | Parallel Web-Service Climate Model Diagnostic Analyzer | NASA ROSES CMAC program  Dr. Tsengdar J. Lee  Tsengdar.J.Lee@nasa.gov | 10/1/2012-09/30/2014  $616K | 0.20 |

## Pending Awards

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PI: Dr. Seungwon Lee Pending Support | | | | |
| Proposer Name | Award/Project Title | Program Info | Period & Total Budget | Commitment (Work Year) |
| Jui-Lin F. Li | Characterizing and Understanding Cloud-Radiation-Dynamics using CloudSat/CALIPSO and other A-Train observations as well as Reanalysis for Model  Improvement  SECTION VI **- Team Members** | NASA ROSES CloudSat And CALIPSO Science Team Recompete | 01/01/2013-12/31/2015  $473K | 0.20 |
| Brian Mapes | A cloud-systems view of A-train nadir data: morphometrics, global impacts, and interactions in regional and local circulations | NASA ROSES  CloudSat And CALIPSO Science Team Recompete | 01/01/2013-12/31/2015  $600K | 0.10 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Co-I: Dr. Jonathan Jiang Pending Support | | | | |
| Proposer Name | Award/Project Title | Program Info | Period & Total Budget | Commitment (Work Year) |
| Hui Su | Aerosol effects on upper tropospheric ice clouds | NASA ROSES CloudSat/CALIPSO Science Team | 09/2013-08/2016 | 0.10 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Co-I: Dr. Chengxing Zhai Pending Support | | | | |
| Proposer Name | Award/Project Title | Program Info | Period & Total Budget | Commitment (Work Year) |
| Gautam Vasisht | Advanced wave front control with Palomar’s Project | NASA ROSES APRA program  Dr. Hashima Hasan  hhasan@nasa.gov | 01/01/2013-09/30/2015  $751K | 0.15 |
| Mike Shao | Enabling exo-earth direct imaging using concurrent astrometry | NASA ROSES  TDEM  Dr. Hashima Hasan  [hhasan@nasa.gov](mailto:hhasan@nasa.gov) | 01/01/2013-09/30/2015  $774K | 0.10 |

Co-I Dr. Lei Pan Pending Awards: None

# Budget Justification

## Budget Narrative

The Jet Propulsion Laboratory is a Federally Funded Research and Development Center owned by NASA and operated by the California Institute of Technology. Work performed at JPL is done under a Prime Contract with NASA through task orders that are added to the contract when work is approved. This proposal, if selected, will be executed through a task order under this Prime Contract.

The workforce requested for the proposed work is based on our prior project experience in similar work related to the development of an Earth science information system for scientific analyses using observational datasets and model outputs. Our current skill, expertise, and relevant ongoing work and preliminary results are also taken into account in estimating the workforce required to accomplish the proposed work. All team members are JPL employees and thus all of the funding is calculated using JPL cost estimation. One domestic trip per year is planned for PI to attend the annual ESDSWG meeting. No equipment purchases are planned.

### Personnel and Work Effort

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Organization | Role | Work Commitment | |
| Year 1 | Year 2 |
| Dr. Seungwon Lee | JPL | Principal Investigator | 0.20 | 0.20 |
| Dr. Lei Pan | JPL | Co-Investigator | 0.60 | 0.60 |
| Dr. Jonathan Jiang | JPL | Co-Investigator | 0.10 | 0.10 |
| Dr. Chengxing Zhai | JPL | Co-Investigator | 0.10 | 0.10 |
| Dr. Joao Teixeira | JPL | Collaborator | 0.05 | 0.05 |

Note that the work effort for Dr. Joao Teixeira as a collaborator of the proposed work will not be funded by this proposal budget. His contribution will be funded by his other research projects that are related with climate model development and evaluation.

### Facilities and Equipment

The following computing and data storage facilities are required and will be available for this work:

* A mini cluster computer containing four compute nodes with 12 cores in each node.
* A 100 TB disk disk array with a raid system

The computer and the disk array will be shared with a NASA ROSES CMAC project that this proposal PI Dr. Lee manages. The disk array contains many CMIP5 model outputs, Obs4MIPs data products, observational data products, and reanalyses products that the proposed work will use. The computer has many of the scientific tools installed, which will be used for the proposed work. This computer and raid system will be used to develop the proposed system CMDA and to operate CMDA for scientific and educational use.

### Rationale and Basis of Estimate

The cost estimate for this proposal was prepared using JPL’s Pricing System and the current internally published Cost Estimation Rates and Factors dated January 2013. The derivation of the cost estimate is a grassroots methodology based on the expert judgment from a team of experienced individuals who have performed similar work. The team provides the necessary relevant experience to develop a credible and realistic cost estimate. The cognizant individuals identify and define the products and the schedule needed to complete the tasks for each work element. Then they generate the resource estimates for labor, procurements, travel, and other direct costs for each work element. The resource estimates are aggregated and priced using JPL’s Pricing System. JPL’s process assures that lower level estimates are developed and reviewed by the performing organizations and their management who will be accountable for successfully completing the proposed work scope within their estimated cost.

### Summary Cost Estimate





## Budget Details – Year 1

Direct Labor – Year 1

* Dr. Seungwon Lee is the PI and will oversee all aspects of the proposed work. She will work on defining system requirements, designing user interfaces of the system, and testing the system performance. Her time commitment is 0.20 wy ($25,300 requested salary with $ 12,880 fringe benefits).
* Dr. Lei Pan will serve as a Co-Investigator on this effort. He will be the lead architect and developer of the proposed system CMDA. He will design the architecture of CMDA, define the technical approach, and implement the data analysis system with interconnectivity and the scalable job distribution system. His time commitment is 0.6 wy ($66,130 requested salary with $33,740 fringe benefits).
* Dr. Chengxing Zhai will serve as a Co-Investigator on this effort. He will support Dr. Pan to develop the data analysis system with interconnectivity since Dr. Zhai is the original developer of the analysis code. He will define the interfaces of individual analysis code and input/output exchange between the analysis web services. He will also support Dr. Lee for testing the CMDA data analysis system. His time commitment is 0.1 wy ($12,670 requested salary with $6,460 fringe benefits).
* Dr. Jonathan Jiang will serve as a Co-Investigator on this effort. He will lead the effort of applying the developed CMDA to scientific investigations, validating results, and demonstrating scientific values of CMDA. He is the lead scientist of the proposal team. He will also lead the effort of demonstrating the educational use of CMDA since he is an instructor of the JPL CCS summer school. His time commitment is 0.1 wy ($12,670 requested salary with $6,460 fringe benefits).

Other Direct Costs – Year 1

Subcontracts/Subawards

* Desktop Network Chargebacks (calculated at $6.06/hr.): All JPL computers are subject to a monthly subcontractor service charge that includes hardware, software, and technical support. ($11,500)

Consultants

* There are no consultants required for this task.

Equipment

* There are no major equipment purchases necessary.

Services

* No services are required for this task.

Travel

* The PI will travel to the annual ESDSWG meeting to participate in Earth science data system working group activities. Estimated cost includes economy airfare, lodging, local transportation, and per diem for 4 days ($2,000).

Other

* Multiple Program Support (MPS) $25,870.

Facilities and Administrative (F&A) Costs – Year 1

* Allocated Direct Costs (ADC) $55,510.
* Applied General ADC $29,990.

Total Estimated Costs for Year 1: $301,180

## Budget Details – Year 2

Direct Labor – Year 2

* Dr. Seungwon Lee is the PI and will oversee all aspects of the proposed work. She will work on defining system requirements, designing user interfaces of the system, and testing the system performance. Her time commitment is 0.20 wy ($26,130 requested salary with $13,320 fringe benefits).
* Dr. Lei Pan will serve as a Co-Investigator on this effort. He will be the lead architect and developer of the proposed system CMDA. He will implement the provenance support system of CMDA since he is the main developer of the existing provenance system that we will apply to CMDA. He will also integrate the subsystems of CMDA and will document installation and developer guides, API specifications, and tutorials. His time commitment is 0.6 wy ($68,270 requested salary with $34,830 fringe benefits).
* Dr. Chengxing Zhai will serve as a Co-Investigator on this effort. He will work on the instrumentation of provenance in existing data analysis code since he is the original developer of the analysis code and knows the algorithmic basis of the analysis code. He will also support Dr. Lee for testing the CMDA provenance support system. His time commitment is 0.1 wy ($13,080 requested salary with $6,650 fringe benefits).
* Dr. Jonathan Jiang will serve as a Co-Investigator on this effort. He will lead the effort of applying the developed CMDA to scientific investigations, validating results, and demonstrating scientific values of CMDA. He is the lead scientist of the proposal team. He will also lead the effort of demonstrating the educational use of CMDA since he is an instructor of the JPL CCS summer school. His time commitment is 0.1 wy ($13,080 requested salary with $6,650 fringe benefits).

Other Direct Costs – Year 2

Subcontracts/Subawards

* Desktop Network Chargebacks (calculated at $6.06/hr.): All JPL computers are subject to a monthly subcontractor service charge that includes hardware, software, and technical support. ($11,440)

Consultants

* There are no consultants required for this task.

Equipment

* There are no major equipment purchases necessary.

Services

* No services are required for this task.

Travel

* The PI will travel to the annual ESDSWG meeting to participate in Earth science data system working group activities. Estimated cost includes economy airfare, lodging, local transportation, and per diem for 4 days ($2,000).

Other

* Multiple Program Support (MPS) $27,700.

Facilities and Administrative (F&A) Costs – Year 2

* Allocated Direct Costs (ADC) $58,230.
* Applied General ADC $31,090.

Total Estimated Costs for Year 2: $312,470