Cover Page for Proposal Submitted to the National Aeronautics and Space Administration

NASA Proposal Number

11-CMAC11-0008

NASA PROCEDURE FOR HANDLING PROPOSALS

This proposal shall be used and disclosed for evaluation purposes only, and a copy of this Government notice shall be applied to any reproduction or abstract thereof. Any authorized restrictive notices that the submitter places on this proposal shall also be strictly complied with. Disclosure of this proposal for any reason outside the Government evaluation purposes shall be made only to the extent authorized by the Government.

SECTION I - Proposal Information

Principal Investigator	E-mail Address	Phone Number
Seungwon Lee	**Seungwon.Lee@jpl.nasa.gov**	**818-393-7720**

Street Address (1)	Street Address (2)
4800 Oak Grove Dr	**M/S 168-200**

City	State / Province	Postal Code	Country Code
Pasadena	**CA**	**91109-8001**	**US**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

Proposed Start Date	Proposed End Date	Total Budget	Year 1 Budget	Year 2 Budget	Year 3 Budget
10 / 01 / 2012	**09 / 30 / 2014**	**615,900.00**	**302,820.00**	**313,080.00**	**0.00**

SECTION II - Application Information

NASA Program Announcement Number	NASA Program Announcement Title
NNH11ZDA001N-CMAC	**Computational Modeling Algorithms and Cyberinfrastructure**

For Consideration By NASA Organization *(the soliciting organization, or the organization to which an unsolicited proposal is submitted)*

Earth Science

Date Submitted	Submission Method	Grants.gov Application Identifier	Applicant Proposal Identifier
03 / 08 / 2012	**Electronic Submission Only**		**17871**

Type of Application	Predecessor Award Number	Other Federal Agencies to Which Proposal Has Been Submitted
New		

International Participation	Type of International Participation
No	

SECTION III - Submitting Organization Information

DUNS Number	CAGE Code	Employer Identification Number (EIN or TIN)	Organization Type
095633152	**23835**	**951643307**	**2A**

Organization Name (Standard/Legal Name)	Company Division
Jet Propulsion Laboratory	

Organization DBA Name	Division Number
JET PROPULSION LABORATORY	

Street Address (1)	Street Address (2)
4800 OAK GROVE DR	

City	State / Province	Postal Code	Country Code
PASADENA	**CA**	**91109-8001**	**USA**

SECTION IV - Proposal Point of Contact Information

Name	Email Address	Phone Number
Seungwon Lee	**Seungwon.Lee@jpl.nasa.gov**	**818-393-7720**

SECTION V - Certification and Authorization

Certification of Compliance with Applicable Executive Orders and U.S. Code

By submitting the proposal identified in the Cover Sheet/Proposal Summary in response to this Research Announcement, the Authorizing Official of the proposing organization (or the individual proposer if there is no proposing organization) as identified below:

- certifies that the statements made in this proposal are true and complete to the best of his/her knowledge;
- agrees to accept the obligations to comply with NASA award terms and conditions if an award is made as a result of this proposal; and
- confirms compliance with all provisions, rules, and stipulations set forth in the two Certifications and one Assurance contained in this NRA (namely, (i) the Assurance of Compliance with the NASA Regulations Pursuant to Nondiscrimination in Federally Assisted Programs, and (ii) Certifications, Disclosures, and Assurances Regarding Lobbying and Debarment and Suspension.

Willful provision of false information in this proposal and/or its supporting documents, or in reports required under an ensuing award, is a criminal offense (U.S. Code, Title 18, Section 1001).

Authorized Organizational Representative (AOR) Name	AOR E-mail Address	Phone Number
Karen Piggee	**Karen.R.Piggee@jpl.nasa.gov**	**818-354-9154**

AOR Signature *(Must have AOR's original signature. Do not sign "for" AOR.)*	**Date**

PI Name : **Seungwon Lee**	**NASA Proposal Number**
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**
Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**	

SECTION VI - Team Members

Team Member Role	Team Member Name	Contact Phone	E-mail Address
PI	**Seungwon Lee**	**818-393-7720**	**Seungwon.Lee@jpl.nasa.gov**
Organization/Business Relationship		Cage Code	DUNS#
Jet Propulsion Laboratory		**23835**	**095633152**
International Participation	U.S. Government Agency		Total Funds Requested
No	**Other**		**336,570.00**

Team Member Role	Team Member Name	Contact Phone	E-mail Address
Co-I	**Jonathan Jiang**	**818-354-7135**	**Jonathan.H.Jiang@jpl.nasa.gov**
Organization/Business Relationship		Cage Code	DUNS#
Jet Propulsion Laboratory		**23835**	**095633152**
International Participation	U.S. Government Agency		Total Funds Requested
No	**Other**		**37,590.00**

Team Member Role	Team Member Name	Contact Phone	E-mail Address
Co-I	**Lei Pan**	**818-393-0477**	**lei.pan@jpl.nasa.gov**
Organization/Business Relationship		Cage Code	DUNS#
Jet Propulsion Laboratory		**23835**	**095633152**
International Participation	U.S. Government Agency		Total Funds Requested
No	**Other**		**129,260.00**

Team Member Role	Team Member Name	Contact Phone	E-mail Address
Co-I	**Chengxing Zhai**	**818-393-0758**	**chengxing.zhai@jpl.nasa.gov**
Organization/Business Relationship		Cage Code	DUNS#
Jet Propulsion Laboratory		**23835**	**095633152**
International Participation	U.S. Government Agency		Total Funds Requested
No	**Other**		**112,480.00**

Team Member Role	Team Member Name	Contact Phone	E-mail Address
Collaborator	**Joao Teixeira**	**818-354-2762**	**teixeira@jpl.nasa.gov**
Organization/Business Relationship		Cage Code	DUNS#
Jet Propulsion Laboratory		**23835**	**095633152**
International Participation	U.S. Government Agency		Total Funds Requested
No	**Other**		**0.00**

PI Name : **Seungwon Lee**	NASA Proposal Number
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

SECTION VII - Project Summary

The latest Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report stressed the need for the comprehensive and innovative evaluation of climate models with newly available global observations. The traditional approach to climate model evaluation, which is the comparison of a single parameter at a time, identifies symptomatic model biases and errors but fails to diagnose the model problems. The model diagnosis process requires physics-based multi-variable comparisons, which typically involve large-volume and heterogeneous datasets, and computationally demanding and data-intensive operations. We will develop a computationally efficient information system to enable the physics-based multi-variable model performance evaluations and diagnoses through the comprehensive and synergistic use of multiple observational data, reanalysis data, and model outputs. This proposal is a response to the NRA topic of computational model and analysis algorithms. Satellite observations have been widely used in model-data intercomparisons and model evaluation studies. These studies normally involve the comparison of a single parameter at a time using a time and space average. For example, modeling cloud-related processes in global climate models requires cloud parameterizations that provide quantitative rules for expressing the location, frequency of occurrence, and intensity of the clouds in terms of multiple large-scale model-resolved parameters such as temperature, pressure, humidity, and wind. One can evaluate the performance of the cloud parameterization by comparing the cloud water content with satellite data and can identify symptomatic model biases or errors. However, in order to understand the cause of the biases and errors, one has to simultaneously investigate several parameters that are integrated in the cloud parameterization. Such studies, aimed at a multi-parameter model diagnosis, require locating, understanding, and manipulating multi-source observation datasets, model outputs, and (re)analysis outputs that are physically distributed, massive in volume, heterogeneous in format, and provide little information on data quality and production legacy. Additionally, these studies involve various data preparation and processing steps that can easily become computationally very demanding since many datasets have to be combined and processed simultaneously. It is notorious that scientists spend more than 60% of their research time on just preparing the dataset before it can be analyzed for their research. To address these challenges, we will build Parallel Web-Service Climate Model Diagnostic Analyzer (PAWS-CMDA) that will enable a streamlined and structured preparation of multiple large-volume and heterogeneous datasets, and provide a computationally efficient approach to processing the datasets for model diagnosis. We will leverage the existing information technologies and scientific tools that we developed in our current NASA ROSES COUND, MAP, and AIST projects. We will utilize the open-source Web-service technology and Parallel Python to achieve the distributed parallel service-oriented system. We will make PAWS-CMDA complementary to other climate model analysis tools currently available to the research community (e.g. PCMDI's CDAT and NCAR's CCMVal) by focusing on the missing capabilities such as co-location, conditional sampling, and probability distribution function and cluster analysis of multiple-instrument datasets. Users will be able to choose between two ways to interface with PAWS-CMDA: (1) a web browser interface for quick and easy exploratory runs and (2) programming language interfaces (e.g. Matlab, Python, IDL) for heavy duty usage and batch runs.

PI Name : **Seungwon Lee**	NASA Proposal Number
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

SECTION VIII - Other Project Information

Proprietary Information

Is proprietary/privileged information included in this application?

Yes

International Collaboration

Does this project involve activities outside the U.S. or partnership with International Collaborators?

No

Principal Investigator	Co-Investigator	Collaborator	Equipment	Facilities
No	**No**	**No**	**No**	**No**

Explanation :

NASA Civil Servant Project Personnel

Are NASA civil servant personnel participating as team members on this project (include funded and unfunded)?

No

Fiscal Year	Fiscal Year	Fiscal Year	Fiscal Year	Fiscal Year
Number of FTEs	Number of FTEs	Number of FTEs	Number of FTEs	Number of FTEs

PI Name : **Seungwon Lee**	NASA Proposal Number
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

SECTION VIII - Other Project Information

Environmental Impact

Does this project have an actual or potential impact on the environment?	Has an exemption been authorized or an environmental assessment (EA) or an environmental impact statement (EIS) been performed?
No	**No**

Environmental Impact Explanation:

Exemption/EA/EIS Explanation:

PI Name : **Seungwon Lee**	**NASA Proposal Number**
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

SECTION VIII - Other Project Information

Historical Site/Object Impact

Does this project have the potential to affect historic, archeological, or traditional cultural sites (such as Native American burial or ceremonial grounds) or historic objects (such as an historic aircraft or spacecraft)?

Explanation:

PI Name : **Seungwon Lee**	NASA Proposal Number
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

SECTION IX - Program Specific Data

Question 1 : Short Title:

Answer: Parallel Web-Service Climate Model Diagnostic Analyzer

Question 2 : Type of institution:

Answer: NASA Center (including JPL)

Question 3 : Will any funding be provided to a federal government organization including NASA Centers, JPL, other Federal agencies, government laboratories, or Federally Funded Research and Development Centers (FFRDCs)?

Answer: Yes

Question 4 : Is this Federal government organization a different organization from the proposing (PI) organization?

Answer: No

Question 5 : Does this proposal include the use of NASA-provided high end computing?

Answer: No

Question 6 : Research Category:

Answer: 2) Data analysis/data restoration/data assimilation/Earth System modeling (including Guest Observer Activities)

Question 7 : Team Members Missing From Cover Page:

Answer:

None

Question 8 : This proposal contains information and/or data that are subject to U.S. export control laws and regulations including Export Administration Regulations (EAR) and International Traffic in Arms Regulations (ITAR).

Answer: No

Question 9 : I have identified the export-controlled material in this proposal.

Answer: N/A

Question 10 : I acknowledge that the inclusion of such material in this proposal may complicate the government's ability to evaluate the proposal.

Answer: N/A

Question 11 : Does the proposed work include any involvement with collaborators in China or with Chinese organizations, or does the proposed work include activities in China?

Answer: No

Question 12 : Are you planning for undergraduate students to be involved in the conduct of the proposed investigation?

Answer: No

Question 13 : If yes, how many different undergraduate students?

Answer: N/A

Question 14 : What is the total number of student-months of involvement for all undergraduate students over the life of the proposed investigation?

Answer: N/A

Question 15 : Provide the names and current year (1,2,3,4) for any undergraduate students that have already been identified.

Answer:

N/A

Question 16 : Are you planning for graduate students to be involved in the conduct of the proposed investigation?

Answer: No

Question 17 : If yes, how many different graduate students?

Answer: N/A

Question 18 : What is the total number of student-months of involvement for all graduate students over the life of the proposed investigation?

Answer: N/A

Question 19 : Provide the names and current year (1,2,3,4, etc.) for any graduate students that have already been identified.

Answer:

N/A

Question 20 : Research sub-topic:

Answer: 1.2.1 Computational model and analysis Algorithms

Question 21 : Other Science Disciplines:

Answers :

Earth science

PI Name : **Seungwon Lee**	NASA Proposal Number
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

SECTION X - Budget

Cumulative Budget

Budget Cost Category	Funds Requested ($)			
	Year 1 ($)	Year 2 ($)	Year 3 ($)	Total Project ($)
A. Direct Labor - Key Personnel	174,910.00	179,410.00	0.00	354,320.00
B. Direct Labor - Other Personnel	0.00	0.00	0.00	0.00
Total Number Other Personnel	0	0	0	0
Total Direct Labor Costs (A+B)	174,910.00	179,410.00	0.00	354,320.00
C. Direct Costs - Equipment	10,000.00	10,000.00	0.00	20,000.00
D. Direct Costs - Travel	0.00	4,000.00	0.00	4,000.00
Domestic Travel	0.00	4,000.00	0.00	4,000.00
Foreign Travel	0.00	0.00	0.00	0.00
E. Direct Costs - Participant/Trainee Support Costs	0.00	0.00	0.00	0.00
Tuition/Fees/Health Insurance	0.00	0.00	0.00	0.00
Stipends	0.00	0.00	0.00	0.00
Travel	0.00	0.00	0.00	0.00
Subsistence	0.00	0.00	0.00	0.00
Other	0.00	0.00	0.00	0.00
Number of Participants/Trainees				0
F. Other Direct Costs	10,760.00	10,760.00	0.00	21,520.00
Materials and Supplies	0.00	0.00	0.00	0.00
Publication Costs	0.00	0.00	0.00	0.00
Consultant Services	0.00	0.00	0.00	0.00
ADP/Computer Services	0.00	0.00	0.00	0.00
Subawards/Consortium/Contractual Costs	10,760.00	10,760.00	0.00	21,520.00
Equipment or Facility Rental/User Fees	0.00	0.00	0.00	0.00
Alterations and Renovations	0.00	0.00	0.00	0.00
Other	0.00	0.00	0.00	0.00
G. Total Direct Costs (A+B+C+D+E+F)	195,670.00	204,170.00	0.00	399,840.00
H. Indirect Costs	107,150.00	108,910.00	0.00	216,060.00
I. Total Direct and Indirect Costs (G+H)	302,820.00	313,080.00	0.00	615,900.00
J. Fee	0.00	0.00	0.00	0.00
K. Total Cost (I+J)	302,820.00	313,080.00	0.00	615,900.00
Total Cumulative Budget				615,900.00

PI Name : **Seungwon Lee**	NASA Proposal Number
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

SECTION X - Budget

Start Date :	End Date :	Budget Type :	Budget Period :
10 / 01 / 2012	**09 / 30 / 2013**	**Project**	**1**

A. Direct Labor - Key Personnel

Name	Project Role	Base Salary ($)	Cal. Months	Acad. Months	Summ. Months	Requested Salary ($)	Fringe Benefits ($)	Funds Requested ($)
Lee, Seungwon	PI_TYPE	0.00				24,510.00	12,490.00	37,000.00
Jiang, Jonathan	CO-I	0.00				12,290.00	6,270.00	18,560.00
Pan, Lei	CO-I	0.00				42,280.00	21,550.00	63,830.00
Zhai, Chengxing	CO-I	0.00				36,770.00	18,750.00	55,520.00
							Total Key Personnel Costs	174,910.00

B. Direct Labor - Other Personnel

Number of Personnel	Project Role	Cal. Months	Acad. Months	Summ. Months	Requested Salary ($)	Fringe Benefits ($)	Funds Requested ($)
0	Total Number Other Personnel					Total Other Personnel Costs	0.00
						Total Direct Labor Costs (Salary, Wages, Fringe Benefits) (A+B)	174,910.00

PI Name : **Seungwon Lee**	NASA Proposal Number
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

SECTION X - Budget

Start Date :	End Date :	Budget Type :	Budget Period :
10 / 01 / 2012	**09 / 30 / 2013**	**Project**	**1**

C. Direct Costs - Equipment

Item No.	Equipment Item Description	Funds Requested ($)
1	Computer and Disk Array	10,000.00
	Total Equipment Costs	10,000.00

D. Direct Costs - Travel

	Funds Requested ($)
1. Domestic Travel (Including Canada, Mexico, and U.S. Possessions)	0.00
2. Foreign Travel	0.00
Total Travel Costs	0.00

E. Direct Costs - Participant/Trainee Support Costs

		Funds Requested ($)
1. Tuition/Fees/Health Insurance		0.00
2. Stipends		0.00
3. Travel		0.00
4. Subsistence		0.00
Number of Participants/Trainees:	**Total Participant/Trainee Support Costs**	0.00

PI Name : **Seungwon Lee**	NASA Proposal Number
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**
Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**	

SECTION X - Budget

Start Date :	End Date :	Budget Type :	Budget Period :
10 / 01 / 2012	**09 / 30 / 2013**	**Project**	**1**

F. Other Direct Costs

	Funds Requested ($)
1. Materials and Supplies	0.00
2. Publication Costs	0.00
3. Consultant Services	0.00
4. ADP/Computer Services	0.00
5. Subawards/Consortium/Contractual Costs	10,760.00
6. Equipment or Facility Rental/User Fees	0.00
7. Alterations and Renovations	0.00
Total Other Direct Costs	10,760.00

G. Total Direct Costs

	Funds Requested ($)
Total Direct Costs (A+B+C+D+E+F)	195,670.00

H. Indirect Costs

	Indirect Cost Rate (%)	Indirect Cost Base ($)	Funds Requested ($)
MPS (reported as Direct Costs per NASA Prime Contract)	0.00	0.00	23,500.00
ADC (reported as Direct Costs per NASA Prime Contract)	0.00	0.00	52,670.00
Gen. (reported as Direct Costs per NASA Prime Contract)	0.00	0.00	30,980.00
Cognizant Federal Agency: None		Total Indirect Costs	107,150.00

I. Direct and Indirect Costs

	Funds Requested ($)
Total Direct and Indirect Costs (G+H)	302,820.00

J. Fee

	Funds Requested ($)
Fee	0.00

K. Total Cost

	Funds Requested ($)
Total Cost with Fee (I+J)	302,820.00

PI Name : **Seungwon Lee**	NASA Proposal Number
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

SECTION X - Budget

Start Date :	End Date :	Budget Type :	Budget Period :
10 / 01 / 2013	**09 / 30 / 2014**	**Project**	**2**

A. Direct Labor - Key Personnel

Name	Project Role	Base Salary ($)	Cal. Months	Acad. Months	Summ. Months	Requested Salary ($)	Fringe Benefits ($)	Funds Requested ($)
Lee, Seungwon	PI_TYPE	0.00				25,250.00	12,740.00	37,990.00
Zhai, Chengxing	CO-I	0.00				37,850.00	19,110.00	56,960.00
Jiang, Jonathan	CO-I	0.00				12,660.00	6,370.00	19,030.00
Pan, Lei	CO-I	0.00				43,470.00	21,960.00	65,430.00
							Total Key Personnel Costs	179,410.00

B. Direct Labor - Other Personnel

Number of Personnel	Project Role	Cal. Months	Acad. Months	Summ. Months	Requested Salary ($)	Fringe Benefits ($)	Funds Requested ($)
0	Total Number Other Personnel					Total Other Personnel Costs	0.00
						Total Direct Labor Costs (Salary, Wages, Fringe Benefits) (A+B)	179,410.00

PI Name : **Seungwon Lee**	NASA Proposal Number
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

SECTION X - Budget

Start Date :	End Date :	Budget Type :	Budget Period :
10 / 01 / 2013	**09 / 30 / 2014**	**Project**	**2**

C. Direct Costs - Equipment

Item No.	Equipment Item Description	Funds Requested ($)
1	**Computer and Disk Array**	**10,000.00**
	Total Equipment Costs	**10,000.00**

D. Direct Costs - Travel

	Funds Requested ($)
1. Domestic Travel (Including Canada, Mexico, and U.S. Possessions)	**4,000.00**
2. Foreign Travel	**0.00**
Total Travel Costs	**4,000.00**

E. Direct Costs - Participant/Trainee Support Costs

		Funds Requested ($)
1. Tuition/Fees/Health Insurance		**0.00**
2. Stipends		**0.00**
3. Travel		**0.00**
4. Subsistence		**0.00**
Number of Participants/Trainees:	**Total Participant/Trainee Support Costs**	**0.00**

PI Name : **Seungwon Lee**	**NASA Proposal Number**
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

SECTION X - Budget

Start Date :	End Date :	Budget Type :	Budget Period :
10 / 01 / 2013	**09 / 30 / 2014**	**Project**	**2**

F. Other Direct Costs

	Funds Requested ($)
1. Materials and Supplies	0.00
2. Publication Costs	0.00
3. Consultant Services	0.00
4. ADP/Computer Services	0.00
5. Subawards/Consortium/Contractual Costs	10,760.00
6. Equipment or Facility Rental/User Fees	0.00
7. Alterations and Renovations	0.00
Total Other Direct Costs	10,760.00

G. Total Direct Costs

	Funds Requested ($)
Total Direct Costs (A+B+C+D+E+F)	204,170.00

H. Indirect Costs

	Indirect Cost Rate (%)	Indirect Cost Base ($)	Funds Requested ($)
MPS (reported as Direct Costs per NASA Prime Contract)	0.00	0.00	23,930.00
ADC (reported as Direct Costs per NASA Prime Contract)	0.00	0.00	53,560.00
Gen. (reported as Direct Costs per NASA Prime Contract)	0.00	0.00	31,420.00
Cognizant Federal Agency: None		**Total Indirect Costs**	108,910.00

I. Direct and Indirect Costs

	Funds Requested ($)
Total Direct and Indirect Costs (G+H)	313,080.00

J. Fee

	Funds Requested ($)
Fee	0.00

K. Total Cost

	Funds Requested ($)
Total Cost with Fee (I+J)	313,080.00

PI Name : **Seungwon Lee**	NASA Proposal Number
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

SECTION X - Budget

Start Date :	End Date :	Budget Type :	Budget Period :
		Project	**3**

A. Direct Labor - Key Personnel

Name	Project Role	Base Salary ($)	Cal. Months	Acad. Months	Summ. Months	Requested Salary ($)	Fringe Benefits ($)	Funds Requested ($)
Lee, Seungwon	**PI_TYPE**	**0.00**				**0.00**	**0.00**	**0.00**
							Total Key Personnel Costs	**0.00**

B. Direct Labor - Other Personnel

Number of Personnel	Project Role	Cal. Months	Acad. Months	Summ. Months	Requested Salary ($)	Fringe Benefits ($)	Funds Requested ($)
0	**Total Number Other Personnel**					**Total Other Personnel Costs**	**0.00**

Total Direct Labor Costs (Salary, Wages, Fringe Benefits) (A+B)	**0.00**

PI Name : **Seungwon Lee**	NASA Proposal Number
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

SECTION X - Budget

Start Date :	End Date :	Budget Type : **Project**	Budget Period : **3**

C. Direct Costs - Equipment

Item No.	Equipment Item Description	Funds Requested ($)
	Total Equipment Costs	**0.00**

D. Direct Costs - Travel

	Funds Requested ($)
1. Domestic Travel (Including Canada, Mexico, and U.S. Possessions)	**0.00**
2. Foreign Travel	**0.00**
Total Travel Costs	**0.00**

E. Direct Costs - Participant/Trainee Support Costs

		Funds Requested ($)
1. Tuition/Fees/Health Insurance		**0.00**
2. Stipends		**0.00**
3. Travel		**0.00**
4. Subsistence		**0.00**
Number of Participants/Trainees:	**Total Participant/Trainee Support Costs**	**0.00**

PI Name : **Seungwon Lee**	NASA Proposal Number
Organization Name : **Jet Propulsion Laboratory**	**11-CMAC11-0008**

Proposal Title : **Parallel Web-Service Climate Model Diagnostic Analyzer**

SECTION X - Budget

Start Date :	End Date :	Budget Type :	Budget Period :
		Project	**3**

F. Other Direct Costs

	Funds Requested ($)
1. Materials and Supplies	0.00
2. Publication Costs	0.00
3. Consultant Services	0.00
4. ADP/Computer Services	0.00
5. Subawards/Consortium/Contractual Costs	0.00
6. Equipment or Facility Rental/User Fees	0.00
7. Alterations and Renovations	0.00
Total Other Direct Costs	0.00

G. Total Direct Costs

	Funds Requested ($)
Total Direct Costs (A+B+C+D+E+F)	0.00

H. Indirect Costs

	Indirect Cost Rate (%)	Indirect Cost Base ($)	Funds Requested ($)
	0.00	0.00	0.00
	0.00	0.00	0.00
	0.00	0.00	0.00
	0.00	0.00	0.00
	0.00	0.00	0.00
	0.00	0.00	0.00
	0.00	0.00	0.00
	0.00	0.00	0.00
	0.00	0.00	0.00
	0.00	0.00	0.00
Cognizant Federal Agency:		**Total Indirect Costs**	0.00

I. Direct and Indirect Costs

	Funds Requested ($)
Total Direct and Indirect Costs (G+H)	0.00

J. Fee

	Funds Requested ($)
Fee	0.00

K. Total Cost

	Funds Requested ($)
Total Cost with Fee (I+J)	0.00

Contents

1 Scientific/Technical/Management

1.1 Introduction

The latest Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report stressed the need for the comprehensive and innovative evaluation of climate models with newly available global observations [Randall et al. 2007]. The traditional approach to climate model evaluation, which is the comparison of a single parameter at a time, identifies symptomatic model biases and errors but fails to diagnose the model problems. The model diagnosis process requires physics-based multi-variable comparisons, which typically involve large-volume and heterogeneous datasets, and computationally demanding and data-intensive operations. We will develop a computationally efficient information system to enable the physics-based multi-variable model performance evaluations and diagnoses through the comprehensive and synergistic use of multiple observational data, reanalysis data, and model outputs. This proposal is a response to the NRA topic of *computational model and analysis algorithms.*

Satellite observations have been widely used in model-data intercomparisons and model evaluation studies. These studies normally involve the comparison of a single parameter at a time using a time and space average. For example, modeling cloud-related processes in global climate models requires cloud parameterizations that provide quantitative rules for expressing the location, frequency of occurrence, and intensity of the clouds in terms of multiple large-scale model-resolved parameters such as temperature, pressure, humidity, and wind [e.g. Slingo 1987; Tiedtke 1993]. One can evaluate the performance of the cloud parameterization by comparing the cloud water content with satellite data and can identify symptomatic model biases or errors. However, in order to understand the cause of the biases and errors, one has to simultaneously investigate several parameters that are integrated in the cloud parameterization.

Such studies, aimed at a multi-parameter model diagnosis, require locating, understanding, and manipulating multi-source observation datasets, model outputs, and (re)analysis outputs that are physically distributed, massive in volume, heterogeneous in format, and provide little information on data quality and production legacy. Additionally, these studies involve various data preparation and processing steps that can easily become computationally very demanding since many datasets have to be combined and processed simultaneously. It is notorious that scientists spend more than 60% of their research time on just preparing the dataset before it can be analyzed for their research [see this NRA].

To address these challenges, we will build Parallel Web-Service Climate Model Diagnostic Analyzer (PAWS-CMDA) that will enable a streamlined and structured preparation of multiple large-volume and heterogeneous datasets, and provide a computationally efficient approach to processing the datasets for model diagnosis. We will leverage the existing information technologies and scientific tools that we developed in our current NASA ROSES COUND, MAP, and AIST projects. We will utilize the open-source Web-service technology and Parallel Python to achieve the distributed parallel service-oriented system. We will make PAWS-CMDA complementary to other climate model analysis tools currently available to the research community (e.g. PCMDI's CDAT and NCAR's CCMVal) by focusing on the missing capabilities such as co-location, conditional sampling, and probability distribution function and cluster analysis of multiple-instrument datasets. Users will be able to choose between two ways to interface with PAWS-CMDA: (1) a web browser interface for quick and easy exploratory runs and (2) programming language interfaces (e.g. Matlab, Python, IDL) for heavy duty usage and batch runs.

1.2 Objectives and Expected Significance

1.2.1 Objectives

The objective of the proposed project is to enable multi-aspect physics-based and phenomenon-oriented model performance evaluation and diagnosis through the comprehensive and synergistic use of multiple observational data, reanalysis data, and model outputs. We will streamline and structure long and complex steps involved in processing multi-source heterogeneous datasets, and will enhance the computational efficiency and data-volume handling capacity for the large-volume data analysis problem. To achieve this goal, we will develop a parallel, distributed web-service oriented system named PAWS-CMDA. The information system will provide the following key capabilities:

1. *Parallel web-service data preparation of observation data and model outputs for model-data intercomparisons,* supporting data subsetting, quality screening, data concatenation, horizontal and vertical coordination change, regridding, variable unit conversion, data co-location, and data-format change.
2. *Parallel web-service data analysis for model performance evaluation and diagnosis,* supporting multivariate conditional sampling, sorting, cluster analysis, statistical and mathematical operations, and analysis output data visualization

The target users of PAWS-CMDA are the climate model analysis community such as the participants in the World Climate Research Programme (WCRP). The end product of the proposed work is an information system that operates the parallel, distributed web services on a network of SMP (Symmetric Multiprocessor) clusters and provides the target users with the aforementioned data preparation and analysis capabilities through a dual-interface design: (1) a web-browser interface for quick exploratory runs and (2) a programming-language interface for long production runs.

1.2.2 Expected Significance

The expected significance of the proposed work lies in the improvement of NASA's climate and weather predictions and is twofold: (1) *Scientific advance through innovative data processing approaches;* (2) *Technological advance through computationally and data efficient approaches.*

Scientifically, our system will enable advances in model-observation comparisons, model performance verification and validation (V&V), and model diagnosis. Unlike the traditional approach, which is to compare a single parameter at a time in the temporal and spatial average domain, PAWS-CMDA will apply an innovative multivariate conditional sampling approach in a phenomenon-oriented and physics-based framework. This approach will enable the study of the instantaneous interaction and relationship between physical variables in the context of their environmental conditions. For example, PAWS-CMDA will make it possible to study the relationship between cloud water content and precipitation rate in the context of the atmospheric dynamics and aerosol condition. The new information gained from this approach will help identify possible errors in model physics and parameterizations, and, ultimately, improve the model performance in climate and weather predictions.

Technologically, PAWS-CMDA will increase the productivity of research involving multi-year, multi-source datasets of model outputs and observations for the climate model analysis community. The multi-year, multi-source dataset preparation and analysis are both computationally and data intensive. By utilizing parallel and distributed web service technology, PAWS-CMDA will significantly reduce the data preparation time and further reduce the data analysis time.

1.3 Technical Approach and Methodology

1.3.1 Overall Architecture of PAWS-CMDA

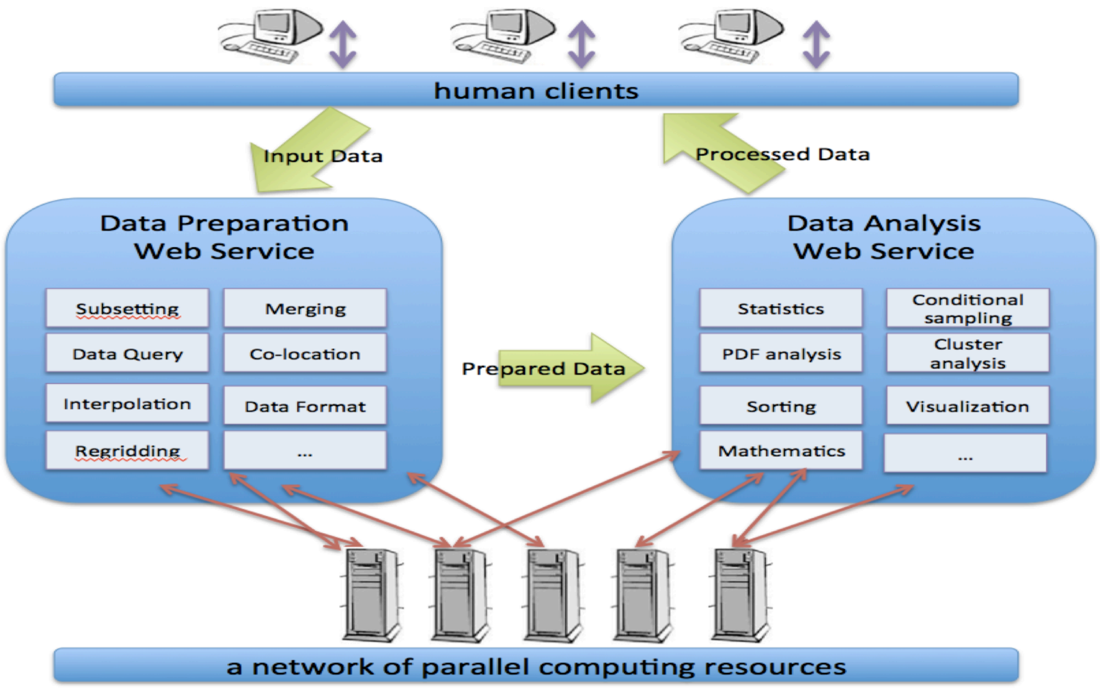

Figure 1. Architecture of the proposed system PAWS-CMDA.

The overall architecture of PAWS-CMDA consists of two main functional blocks: (1) the data preparation system and (2) the data analysis system. The functional blocks are composed of several operations, individually wrapped with Web Service interfaces, and powered by Parallel Python for computational performance enhancement from leveraging multi-core symmetric multiprocessing clusters. The system will utilize a network of parallel computing resources to execute the web services tasks behind the scene. Figure 1 illustrates the architecture of PAWS-CMDA.

PAWS-CMDA will have user interfaces using web browsers as well as various programming languages. The dual-interface design of PAWS-CMDA is tuned to meet the needs of its target users. Climate model analyses are by nature exploratory, which requires multiple runs with different input configurations (co-location conditions, data sampling conditions, etc). The programming-language interface is effective for scripting multiple runs and submitting them in batch mode. On the other hand, a web-browser interface enables 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 periods of intense usage. The final version of PAWS-CMDA will be delivered to the target users in two ways: (1) client-module open source packages that are downloadable (e.g. sample client programs in Python, IDL, and Matlab); (2) web services on remote servers that are accessible through a web browser with minimal local installation requirements (e.g., JSON).

1.3.2 Capabilities of PAWS-CMDA

Over the last several years, we have developed model analysis tools for several research projects including the NASA ROSES COUND project (entitled "Evaluating CMIP5 models with satellite observations: cloud-climate feedbacks") led by Dr. Lee (PI of this proposal), the NASA ROSES COUND project (entitled "Utilizing NASA A-Train datasets for IPCC AR5 climate model evaluation") led by Dr. Jiang (Co-I of this proposal), the NASA ROSES MAP project (entitled "Cloud transitions in the tropics and sub-tropics: improving the representation of shallow cumulus convection in coupled systems") led by Dr. Teixeira (collaborator of this proposal), the NASA ROSES MAP project (entitled "Judicious application of satellite observations to evaluate and improve cloud ice and liquid water representations in conventional and muti-scale weather and climate models") staffed by Dr. Lee and Dr. Pan (Co-I of this proposal), and the NASA ROSES AIST project (entitled "Online services for correction of atmosphere in radar") staffed by Dr. Pan. The existing tools cover a large scope of data preparation and processing capabilities, parallel computing capabilities, and web service technologies. We will build upon these existing capabilities and technologies to develop PAWS-CMDA.

The two main components of PAWS-CMDA are data preparation and data analysis. The data preparation system will enable several commonly needed operations in the data preparation step:

1) Change data format (e.g. change a HDF file to a NetCDF file);
2) Subset data conditionally by time, space, and variable (e.g. select tropical summer water vapor data);
3) Concatenate data from multiple files (e.g. collect precipitation rate from year 2000 to 2005);
4) Change horizontal and vertical coordinates (e.g. change height to pressure level);
5) Average and regrid the temporal and spatial resolution (e.g. monthly 1x1 degree averaged values);
6) Co-locate and interpolate multi-source outputs to match time and locations (e.g. co-locate MODIS footprint data with CloudSat footprint data);
7) Convert variable units (e.g. convert cloud water content from the density unit (mg/m^3) to the mass fraction unit (mg/kg)).

The data analysis system will enable the following operations that are critical in the data analysis step:

1) Apply mathematical operations (e.g. apply algebraic, logical, and calculus operations);
2) Apply statistical operations (e.g. calculate standard deviation and correlation);
3) Assess probability density function (PDF) distributions of the data (e.g. estimate the PDF of total water content in the stratocumulus regions);
4) Analyze cluster distributions of the data (e.g. identify the number of clusters for the cloud classifications or scene classifications);
5) Multivariate conditional sampling based on phenomena and physics (e.g. select cloud water content data in non-convective and non-precipitation conditions);
6) Sort data by a given variable condition (e.g. sort cloud water path data in order of precipitation rate).

All the listed operations for the data preparation and analysis systems are currently fully or partially developed by the proposal team. Figure 2 illustrates some of the capabilities covered by the existing tools. While some of the operations are commonly available in other climate model analysis tools, the proposed system provides several capabilities that are usually missing in the other tools, namely co-location, multi-variable conditional sampling, cluster analysis, and PDF analysis.

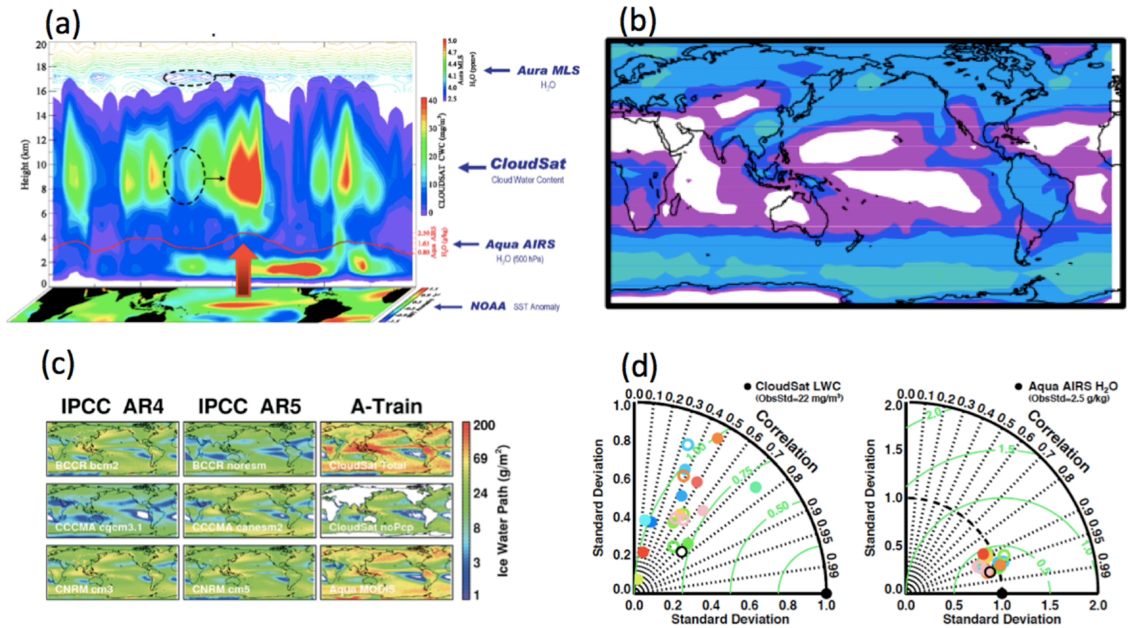

Figure 2. Capabilities of the existing tools developed by our team: (a) co-located parameters observed by multiple instruments in both space and time, (b) a multi-year averaged, regridded, and conditionally sampled parameter with parameters from multiple instruments, (c) multi-year averaged and conditionally sampled parameters for data-model intercomparisons, (d) Taylor diagram of standard deviation and correlation of model parameters for model evaluation

One of the existing tools that we developed and will integrate into PAWS-CMDA is a Level-2 data co-location tool. The tool is currently capable of co-locating source datasets such as AIRS, AMSR, CERES, CALIOP, MLS, MODIS, ECMWF ERA-interim, and ECMWF-YOTC analysis onto the footprint of the target dataset CloudSat. Using this tool, we have generated multiple CloudSat-centric co-located data products for the WCRP/WWRP THORPEX MJO Task Force and YOTC activity. The co-located data products are currently available for the YOTC research community through the web service maintained by the CloudSat Data Processing Center at http://csyotc.cira.colostate.edu. We will extend the capability of the co-location tool to co-locate any given two datasets (within some selected datasets – e.g. A-Train datasets, CMIP5 datasets, ECMWF and MERRA reanalysis datasets) with a few general co-location conditions such as finding neighbors within a given distance and a given time difference and distance-weighted averaging of the neighbor parameters.

Along the lines of evaluating the CMIP5 models using A-train satellite observations, we have built several scientific data analysis tools. An efficient multi-dimensional re-gridding tool was developed to resample the GCM model data to customized grids and to facilitate detailed pattern comparison. We have also developed techniques to support data pattern comparison for specific geographical regions, e.g. over tropical oceans or land. Our tools generate Taylor diagrams to concisely display the correlations and standard deviations. To evaluate the model physics, we have developed conditional sampling capabilities to display variables in physical parameter space, e.g. SST, up draft wind, lower troposphere stability etc. We have developed a multi-panel graphics tool that can conveniently display maps or plots with user-defined nonlinear scales. We have also developed a metric system to grade model performance by comparing the model deviations from the observations with the observation uncertainties. We have applied these tools to evaluate the cloud and water vapor simulations of climate models [Jiang et al. 2012, Su and

Jiang et al. 2011]. These tools will be incorporated into the PAWS-CMDA system to provide scientists with online high-performance analysis capabilities.

For the large-volume data analysis and summary, we have also developed a probability density function (PDF) estimation tool and a cluster analysis tool. These tools efficiently process large-size datasets and reduce the data volume by representing them as a PDF or as clusters. The summarization is useful in classifying the patterns and group types and extracting data content for statistical correlation and model parameterization. Note that the PDF estimation is different from a simple histogram representation. It involves a likelihood estimation and optimization of parameters associated with a chosen PDF function. We have applied the PDF estimation tool to study the cloud water content distribution from a multi-year CloudSat observational dataset and to relate it to the cloud parameterization in global circulation models (Lee et al. 2010).

We will describe below the details of how we propose to integrate the existing tools into PAWS-CMDA.

1.3.3 Strategies for computationally efficient and scalable system

While the data processing and analysis operations listed in Section 1.3.2 seem straightforward to implement, they can be computationally very demanding due to the large volume of data to handle. For example, a simple task we tested is to collect CloudSat-retrieved cloud water path for three years and regrid it to make a monthly averaged series. It took ~5 hours to complete the task on a single modern-speed processor (2.3 GHz Intel Core). The data involved in the simple task is ~5.6 TB stored in ~16,000 files. The file IO and network IO are the main bottleneck of this task. Another example task that we tested is to co-locate the MODIS 1km-resolution cloud product onto the CloudSat footprints for two years. This task requires a nearest-neighbor search between the two datasets that are stored with different file granularities and time and location distributions. The task involves ~180GB of CloudSat data stored in 10,000 files and ~3.4TB of MODIS data stored in ~200,000 files. It took about a month of wall-clock time to complete this task on an 8-core processor with 24 GB memory. The file IO and network speed, memory size, and co-location computation speed all contributed to the long execution time.

The key success criterion in developing model-data preparation and analysis tools like PAWS-CMDA is to address the complexity arising from the large data volume, the demanding file IOs, and computationally expensive operations. From the previous research projects, we have developed a strategy to optimize the file IO, memory usage, and computational efficiency in these operations.

The approach we will take to optimize memory usage is to selectively read a subset of the data from an input file in the order it is needed for the upcoming computational operation instead of reading the entire data/variable from a given file. This will not only minimize the memory usage but also speed up the execution time due to the fast buffer/cache loading.

To improve the computational efficiency of data processing operations, we have profiled the computational performance of the existing functions and selected the functions that are the critical bottlenecks. We found that it is best to code the bottleneck functions in C/C++ and wrap the functions in Python for the integration. In some cases, it was necessary to develop a new algorithm to make the program scalable in data size. For example, we have developed an alternative algorithm for nearest-neighbor search that scales linearly with data volume for our co-location tool. We will apply these experience-based programming principles to PAWS-CMDA development.

Because we are dealing with diverse sets of data with different formats and structures, it is very important that we develop the core functions of PAWS-CMDA to be applicable to various datasets. This means that the core functions should not assume any data-specific properties. For example, it would be more straightforward to develop a co-location function if we assumed that the datasets are all regularly gridded. However this condition is not often satisfied in Level-2 satellite datasets while it is often true in model outputs. We will pay close attention to making the core functions and algorithms to be general, and applicable to a wide range of datasets. Any data-specific information should be handled in "front-end" code where the initial data processing will be conducted in order to prepare the dataset for the upcoming data processing.

PAWS-CMDA will be developed through the following three sequential tasks.

(1) Develop a Python-driven data preparation and analysis system.

(2) Apply Parallel Python to the Python-driven data preparation and analysis system.

(3) Create Web services for the Parallel-Python data preparation and analysis system.

The details of how to accomplish the tasks are described below.

1.3.4 Python-driven data preparation and analysis system

We choose Python as the driving programming language for four reasons. First, Python is a free-license, open-source tool. Second, Python developers are one of the fastest growing programming communities and therefore provide many useful packages to users. Third, Python has strengths in OS-level interactions (e.g., networking, file I/O, process spawning and joining) and string and data/time manipulation. Finally, Python offers flexible and established ways to integrate with many other programming languages.

While Python will be the driving programming language, some of the underlying operations will be written in other languages for better computational performance and will then be wrapped in Python for integration into the system. The existing data preparation and analysis tools that our team developed are written mainly in Python, Matlab, and C/C++. The tools written in Matlab and C/C++ will be either converted to Python or wrapped in Python after computational performance and code development cost are weighed. We will use SWIG (Simplified Wrapper and Interface Generator) or Ctypes open-source packages to call C/C++ functions in Python. As for the existing Matlab code, we will use Mlabwrap package to bridge them to Python. In addition, several open-source Python packages will be utilized for the development of PAWS-CMDA. NumPy will be used for numerical operations, SciPy for more advanced algebraic operations, Scientific.IO for NetCDF file IO, pyhdf for HDF file IO, and matplotlib for visualization. The proposal team is familiar with all the packages mentioned above.

The PAWS-CMDA system will be structured into three components, a Front End, a Middle End, and a Back End. The Front End is responsible for reading a specific dataset and storing data and metadata in a commonly structured way. The Middle End hosts the operation-specific functions (e.g. co-location and conditional sampling) that are listed in Section 1.3.2. The Back End writes the output data and metadata in the format of the user's choice (e.g., NetCDF, HDF). Additionally, we will build several utility tools, such as a Python-based visualization tool to visualize output datasets (as shown in Figure 2). Most of our existing tools follow the three-component structure and therefore will be easily integrated into the PAWS-CMDA structure.

1.3.5 Parallel-Python powered data preparation and analysis system

We choose Parallel Python to achieve the goal of large-scale distributed parallelism. Parallel Python is a Python module that is open source and available at no cost. It allows users to establish a server network among the cores of a SMP (Symmetric Multiprocessor with multi-processors or multi-cores) or among the CPUs of a SMP cluster. A client application program can then "submit" sub-tasks to the server network, each defined by a reference to a function to be executed and a set of input arguments. At the end of the computation, the client program retrieves the results from these sub-tasks. Another feature of Parallel Python is the possibility to specify callback functions through additional arguments at task submission, so that a "remote" function can call the local callback function upon finishing its computation on a "remote" core. Parallel Python also features auto-discovery of computational resources on a network, fault-tolerance if any SMP nodes fail, and dynamic allocation of computational resources. The programming APIs of Parallel Python are very familiar to our proposal team, which has extensive experiences in parallel distributed programming using MPI, OpenMP, pthreads, the Python Multiprocessing package, and Navigational Programming [Lee et al. 2007; Pan et al. 2001; Pan et al. 2002; Pan et al. 2003].

Our existing co-location tool uses Python Multiprocessing to speed up the co-location process, which consists of finding nearest neighbors among pixels of granule groups. This approach leverages two facts: (1) the nature of the co-location algorithm is embarrassingly parallel; and (2) multi-core servers are affordable and a good fit for shared-memory multi-processing. Python Multiprocessing uses the spawning of child processes to parallelize a task on a single SMP (Symmetric Multiprocessor, which has multi-processors or multi-cores) computer. Its programming API is similar to that of multithreaded programming (e.g., Python Threading), but it side-steps the Python GIL (Global Interpreter Lock), which limits Python threads from fully utilizing the power of all the cores on a SMP. The advantage of Python Multiprocessing lies in its shared memory programming API and effective memory recycling at the point where child processes joint. The accompanying disadvantage is that parallel sub-tasks cannot access a distributed memory system, which prohibits scaling to hundreds of processors and beyond.

Yet our problem is large scale, involving granules or granule groups of tens of thousands. We have seen that co-location of some datasets can take months of CPU time on an 8-core SMP. At the same time, the core algorithms, like co-locating through nearest neighbor searching, and data I/O throughout the processing are embarrassingly parallel, which makes it a perfect fit for distributed parallel computing on a SMP cluster of thousands of cores. It is our goal to achieve at least one order of magnitude speedup by scaling to 10s or 100s of network-connected SMPs.

An example of co-location will illustrate our strategy in applying Parallel Python to PAWS-CMDA to achieve the highly scalable computational performance. A-Train data co-location involves two satellite data sets, one source and one target set, and the objective is to project source data values onto the target pixels using interpolation between the two pixel spaces. One module of Python code can be written to play two roles: (1) It is the entry point for the user interface, through which our end users specify the spatial and temporal constraints, along with other the parameters of their request, such as a selection of variables or a choice of nearest neighbor versus spatial/temporal averaging interpolation schemes; and (2) It is the top level driver that will direct the target granules to the specific cores, local or remote, initialize the Parallel Python server network, and submit the sub-tasks assigning the granule sub-sets to their respective cores, using the reference to a sequential function as one of the submission inputs. On each of the cores in the Python Parallel server, this sequential function will call the Front End to read the relevant source and target granules, call the Middle End to conduct interpolation, and call the Back End to write the co-location product. Finally, the top-level driver collects status results from all cores, joins all the sub-processing, and terminates the server network. In this context, the top-level driver is the client application in the Parallel Python terminology. Notice that the co-location of one particular target granule is independent of that of another one, which makes the algorithm embarrassingly parallel. No communication is needed among the sequential co-location processes running on their respective cores.

We can envision potential challenges in using Parallel Python for our purposes: (1) It is not clear to us whether disk I/O by functions submitted on remote computers is supported in Parallel

Python. In the affirmative, we will use a parallel file system (e.g., PVFS) to host the source data granules and the intermediate and final products. Parallel I/O should improve performance. If remote disk access is not possible, we will design a pre-fetching algorithm/mechanism that uses function arguments to ship source data from the client to all the servers on the network, and later collect the results for the client to write them to disk files. Pre-fetching will use Python Multithreading or Multiprocessing to partially hide the cost of networking and I/O; (2) We might encounter algorithms that, unlike co-location, are fine-grained if we take a naive parallelization approach. These algorithms may not be the core part of the entire processing, but they could become the bottleneck for performance. In such a case, we will modify the algorithms to coarsen the amount of computation in between communications. Typical solutions involve redundant computations on all the server nodes to reduce the frequency of communication.

Why do we not propose to use MPI, which is a more familiar approach in the parallel computing community? In our applications, the communication pattern is very simple: SMP nodes or multi-cores only need to know on which data granules to work, along with spatial, temporal, or variable parameters. The nodes will complete their job by independently writing the results to the disk. There is no communications, whether collective or point-to-point, during data processing of each data granule or granule groups, as shown in the co-location example above. Consequently, we decided to avoid the more powerful but more complex MPI approach in favor of the simple distributed computing scripting environment enabled by Parallel Python to fully leverage our already existing Python-driven tools and packages.

1.3.6 Web-Service wrapped data preparation and analysis system

We choose to build PAWS-CMDA using the Web Service technology to benefit from its simple and flexible environment with a rich set of open source packages for distributed programming. Another important reason to build web services instead of standalone offline equivalents is the heavy requirement of PAWS-CMDA in terms of the large volume of input datasets, which are typically needed for long-term data-model intercomparisons. For example, A-Train datasets for the entire mission lifetime, CMIP5 model outputs, ECMWF ERA-interim and MERRA datasets for over 20 years of the modern era can easily reach a petabyte. A standalone offline tool would require users to have the large source data volume on their server, which is impractical and unnecessary. Ideally, the data should reside on the server where data will be first processed and down-selected so that only the final dataset needed by the users should be sent. A web service enables the processing of data on the remote server as if it resided on the user's machine, and then the retrieval of the final product from the remote server.

Web services have a dual access mode: human users can directly access them through web browsers, or other programs on the Web can use them. The proposed PAWS-CMDA services are meant to facilitate both. To be served, a client tells a service on what data to operate, and what to do with the data. The former is referred to as "scoping information," and the latter as "method information." Scoping information can be placed in a URI (Uniform Resource Identifier), while method information is conveyed by a HTTP (the Hypertext Transfer Protocol) method. If we follow the above practice, we are building REST-ful (Representational State Transfer) services using the Resource-Oriented Architecture (ROA) [Fielding and Taylor, 2002; Richardson and Ruby, 2007].

Why do we choose REST and ROA? Among several reasons, one important consideration is ease of use in order to meet the requirements of our system. PAWS-CMDA should support two types of clients: (1) Human clients using web browsers and (2) Program clients such as Matlab, IDL, and python programs. These two types of clients should access our services in the same way, by using URIs to address resources, and by using HTTP methods to assign a task to the service. In contrast, the RPC (Remote Procedure Call)-style services may use SOAP (Simple Object Access Protocol) envelopes to carry both scoping and method information, and use only HTTP GET to create, read, update, and delete resources. This type of services is hard to use, if possible at all, for human clients using browsers.

Besides being addressable by exposing URIs to resources, RESTful and ROA services are stateless. This means that a service does not store any application state, and hence treats each client request in isolation. In other words, a service does not save session information for its clients. Being stateless is simple to implement, and it makes it trivial to scale out the services. When workload is too heavy, more workstations can be deployed, with no need to synchronize among them. A client can request services from any of these servers and get the same result. Stateless services and clients are also reliable; when a client request times out, it can be resent without worrying about what session state the client is in.

Dr. Pan (the Co-I of the proposal team) has developed several web services for the NASA ROSES AIST project entitled "Online services for correction of atmosphere in radar". We will utilize the framework and technologies developed within this project and leverage open source packages including Apache HTTPD, mod_wsgi, JSON (JavaScript Object Notation) and Ajax (Asynchronous JavaScript and XML) utilities. Following standard software engineering principles, we will first modularize our existing science applications clearly identifiable as operators with well-defined input and output streams. Then, we will turn them into a web service which simply regulates how the operators behave on the web through well-defined APIs and conform to a set of predefined input/output conventions.

1.3.7 Data discovery, access, and sharing in PAWS-CMDA

Information technology for efficient discovery, access, and sharing of NASA's Earth science datasets is critical for increasing science productivity. We recognize that the development of such technologies constitute a separate project by itself and is outside the scope of the proposed work. However, we will address in this section some of our basic strategies to optimize the performance of data discovery, access and sharing within our PAWS-CMDA system.

PAWS-CMDA is both a consumer and a producer of data; it reads in pre-processed data as input and produces processed data products as output. Both input and output data can be large in volume, which calls for a strategy for efficient I/O. We will leverage hard disks that are locally mounted to the server where PAWS-CMDA resides in order to boost I/O performance. Source data granules from a remote data center (e.g., DAAC or ESG) will be cached to this disk, in two scenarios: on demand or pre-fetched. On-demand downloading is needed when data granules that PAWS-CMDA requires immediately do not exist on the local disk, in which instance a download program is kicked off. Pre-fetching ahead of time downloads the data granules that "may be" needed next, to reduce the overall amortized delay, with the assumption that our data processing operations exhibit locality of access. Pre-fetching usually piggybacks on-demand downloading. If the local disk space is large and if pre-fetching is done aggressively, less on-demand downloading from a DAAC is needed, which reduces the overall delay. But on-demand downloading cannot be completely avoided simply because new data streams regularly flow into the remote data centers.

Since the local disk space is smaller than the total amount of source data that PAWS-CMDA will support, we need a data purge/cleanup policy. We will purge the data granules that are least recently used. For output data, PAWS-CMDA returns to the client a new URI pointing to where the new co-location product resides. The purging of the output data also uses the "least recently used" policy. This purging policy assumes that PAWS-CMDA user requests exhibit temporal locality of access. Our caching mechanism and strategy are similar to those of Operating System (OS) paging, if one realizes that our local disk is equivalent to the fast but small media in OS – the main memory - and the remote DAAC sites are equivalent to the slow but large media in OS – the hard disk.

The data stored in our server will be indexed and stored with time, variable, and space information (i.e. meta-data). The indexed meta-data will be used to discover existing datasets in the server for PAWS-CMDA operations. For example, the file name and the start and end times of CloudSat files are stored in terms of CloudSat granule number which is a unique number for every CloudSat product and sorted in time order. For our existing tools, we have developed an algorithm and data structure for indexing and searching the file granules in the time domain, which takes place before the data processing step.

1.3.8 Scientific applications of PAWS-CMDA

We have identified two scientific applications that will drive the development and performance evaluation of PAWS-CMDR in the context of climate model evaluation.

Evaluating representations of atmospheric moist process in the global models

We will apply PAWS-CMDA to process the A-Train datasets, ECMWF/MERRA outputs, and CMIP5 model outputs to evaluate the moist process parameterizations in the general circulation models (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 as 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 synergistic use of the global observations across multiple sensors with proper co-location and merging can provide comprehensive and detailed information for evaluating the radiative and moist process representations in GCMs.

Figure 3a illustrates relevant measurements from A-Train multiple instruments and ECMWF/MERRA (re)analysis outputs that are complementary and can be co-located to develop and constrain model representations of clouds, convection and their processes in GCMs. For example, CALIPSO is sensitive to thin cirrus with optical thickness $\tau < \sim 0.05$, whereas CloudSat observes thicker clouds and MLS's sensitivity lies between CALIPSO and CloudSat. These three instruments also have different vertical ranges. Using PAWS-CMDR, we can collocate and combine cloud measurements made from the above three satellite sensors and thus create combined total cloud profiles as shown in Figure 3b. This combined total cloud profiles allow us to compute more accurate cloud radiative heating rates and estimate cloud forcing and feedbacks to climate change. The proposal team has recently demonstrated the use and scientific value of the co-located datasets in this context (Lee et al. 2012; Jiang et al. 2012).

Investigating the relationships between convective precipitation formation and the ambient thermodynamic and dynamical controlling factors

We will apply PAWS-CMDA to collocate and conditionally sample A-Train satellite datasets and CMIP5 model outputs for the evaluation of the convection and precipitation parameterizations in the models. 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. Observational resources for the development and evaluation of convection and precipitation parameterizations in global models mostly rely on *in situ* measurements from ground-based sounding stations and field campaigns over specific regions in limited time periods. Co-located and conditionally sampled satellite retrievals from multiple sensors and platforms will be as rich a dataset as in situ measurements or coordinated field campaigns, but available globally over a longer temporal span and will hence provide a global-scale distribution and long-term variability.

For example, Aqua AIRS water vapor (H_2O) measurements can be sampled onto the MLS orbits and collocated with MLS cloud samples as shown in Figure 3c. Figure 3d further shows H_2O at six different altitudes, observed by Aqua AIRS and Aura MLS, scattered against Aura MLS measured upper tropospheric ice water content (IWC), an index for deep convection. The positive correlation of H_2O with IWC persists throughout the troposphere, except near the tropopause at 100 hPa, where increasing cirrus clouds are associated with decreasing water vapor. The observed height-dependent sensitivity of water vapor to clouds can also be examined in the model. With PAWS-CMOR, we can sample the model output onto the satellite orbit in both space and time and thus make collocated comparisons with the remote sensing observations. Figure 3e shows an example of NCAR CAM model simulated IWC sampled on MLS orbits.

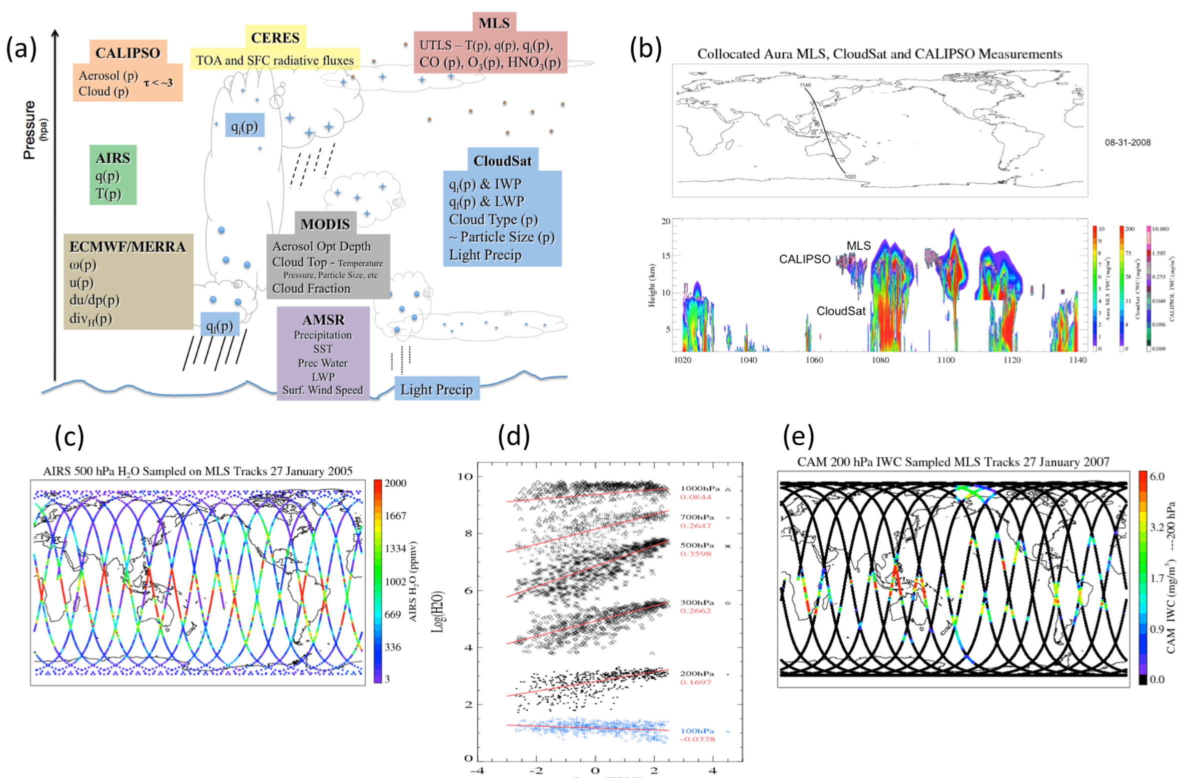

Figure 3. (a) Measurements from A-Train multiple instruments and ECMWF/MERRA (re)analysis outputs that can be co-located to constrain GCM model representations of the moist and convection processes; (b) total cloud profiles combined with measurements made from three satellite sensors (MLS, CloudSat, and CALIPSO); (c) AIRS water vapor measurements collocated with MLS cloud samples; (d) Correlations of the co-located AIRS water vapor with the MLS ice water content (IWC); (e) NCAR CAM model IWC sampled on MLS orbits and co-located with MLS IWC for model-data intercomparisons.

1.4 Perceived Impact to the State of Practice

Earth science research routinely requires discovering, accessing, understanding, and processing Earth science data such as satellite remote sensing data sets, model outputs, and (re)analysis outputs that are physically distributed, massive in volume, and heterogeneous in format. While information systems that facilitate individual NASA data product discovery and processing services are becoming more widely available (e.g. GCMD, ECHO, ESG, Giovanni), information systems that support the comprehensive processing of multi-source data products and the synergistic use of co-located multiple-instrument measurements for model evaluations and diagnoses are limited. Furthermore, most of the existing climate data analysis tools (e.g. PCMDI's CDAT and NCAR's CCMVal, and other in-house research tools) 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 multiple-instrument data with one another and use them to evaluate and diagnose model performance). PAWS-CMDA will enable multi-source dataset processing and model performance diagnostic analysis based on multivariate physics-based, phenomenon-oriented comparisons and analyses using satellite observational datasets. Therefore, PAWS-CMDA will be complementary to the climate model analysis tools currently available to the community such as CDAT and CCMVal.

Scientifically, the multivariate model-to-data comparisons and analyses are critical to understand the causes of model errors and biases and to improve climate model performance for 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 and errors. A study of the instantaneous interaction and relationship between physical variables in the context of its environmental condition, using PAWS-CMDA will provide new insight into possible errors in the model physics and parameterizations.

Technologically, the proposed system will embody the prototype of a model analysis tool that utilizes a parallel and distributed web service, and is scalable with data volume. In the current practice, the large overhead of preparing datasets (over 60% of the research time) is due to the *ad hoc* approach taken in handling the large-volume and heterogeneous datasets. PAWS-CMDA will provide a streamlined and structured approach to process the data and standardize the output format and content, and provide a computationally efficient way to analyze the large-volume datasets, using data-volume scalable algorithms and parallel distributed computing. The dual interface of PAWS-CMDA (i.e. a programming-language interface and a web browser interface) will give users flexibility in preforming their analysis within a familiar programming environment as well as within a web browser without the hassle of local installation.

1.5 Relevance to Element Programs and Objectives in the NRA

The proposed work responds to the "*Computational model and analysis algorithms*" area in this NRA. In particular, the proposed work is a direct response to the development of a model output analysis system based on an open source parallel Python and in a distributed parallel computing environment. The proposed work will make the existing model and data analysis tools more efficient in a distributed parallel environment and make the computational efficient tool available to the climate model analysis community as web services.

1.6 Work Plan

1.6.1 Task Schedules and Key Milestones

◆ = milestone, # = task, ● = deliverable	Year 1				Year 2			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Parallel distributed web service for data processing					◆ M1			
Develop a python-driven data processing system.	#							
Apply Parallel Python to the python-driven system.		#	● D1					
Develop a Web Service for the data processing system.			#					
Demonstrate the scientific use of the processing system.				#	● D2			
Parallel distributed web service for data analysis								◆ M2
Develop a python-driven data analysis system.					#			
Apply Parallel Python to the python-driven system.						#	● D3	
Develop a Web Service for the data analysis system.							#	
Demonstrate the sceintific use of the analysis system								#
Integrated PAWS-CMDA								◆ M3
Integrate the two systems to build PAWS-CMDA.								# ● D4,5

The proposed work will be accomplished with one Full Time Employee (FTE) work level combined by the proposal team members' contributions for two years. The following tasks, deliverables, and milestones are planned to achieve the objectives of the proposed work. The time line of the tasks, deliverables, and milestones are shown in the table above.

<u>Year 1</u>

Task 1: Develop a python-driven data processing system.
Task 2: Apply Parallel Python to the python-driven data processing system.
Deliverable 1: Semi-annual project report.
Task 3: Develop a Web Service for the Parallel-Python driven data processing system.
Task 4: Demonstrate the scientific use of the data processing system.
Deliverable 2: Semi-annual project report.
Milestone 1: Implementation of a parallel distributed web service for data processing system.

<u>Year 2</u>

Task 4: Develop a python-drive data analysis system.
Task 5: Apply Parallel Python to the python-driven data analysis system.
Deliverable 3: Semi-annual project report.
Task 6: Develop a Web Service for the Parallel-Python driven data analysis system.
Task 7: Demonstrate the scientific use of the data analysis system.
Milestone 2: Implementation of a parallel distributed web service for data analysis system.
Task 8: Integrate the data processing system and data analysis system to build PAWS-CMDA.
Deliverable 4: Semi-annual project report.
Deliverable 5: PAWS-CMDA client-module packages and web services on remote servers.
Milestone 3: Implementation of the integrated PAWS-CMDA.

1.6.2 Management Plan 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 PAWS-CMDA system and the requirements of its capabilities by seeking inputs from the targeted research community. She is a core developer of a co-location tool for A-Train data and ECMWF analysis outputs for a NASA ROSES MAP project, and is a PI of a NASA ROSES COUND project, conducting research on evaluating CMIP5 models with satellite observations for cloud-climate feedbacks. By leveraging the combined experience of both the information technology development and the Earth science investigations, Dr. Lee will facilitate effective interactions between the Earth scientists of the proposal team (Dr. Jiang and Dr. Teixeira) and the information technologists of the team (Dr. Pan and Dr. Zhai).

Dr. Lei Pan of JPL, Co-I, will be the lead architect and developer of the proposed PAWS-CMDA system. He is the information technology expert in the proposal team. He has designed the architecture of the existing co-location tool, and developed and implemented a scalable co-location algorithm. He has also developed web services for a NASA ROSES AIST project. He has extensive experience in numerical methods, large-scale software for computer-aided engineering, high performance computing applications, distributed parallel programming methodologies, and Web Service technologies. He will be responsible for the architecture, design and development of PAWS-CMDA. He will apply the Parallel Python and Web Service technology to our Python-driven data preparation and analysis system.

Dr. Chengxing Zhai of JPL, Co-I will be the lead developer of the data analysis components of PAWS-CMDA system. He has developed the existing model analysis tool for the A-Train data processing, model-output processing, and model-data intercomparison from the NASA ROSES COUND project led by Dr. Jiang (a Co-I of the proposal team). He will convert the existing tool to a Python-driven tool.

Dr. Jonathan Jiang of JPL, Co-I, will lead the effort of applying the PAWS-CMDA system to scientific investigations, validating results, and demonstrating the scientific value of PAWS-CMDA. He is the PI of the current NASA ROSES COUND project entitled "Utilizing NASA A-Train datasets for IPCC AR5 climate model evaluation". He has extensive experience with A-Train data analysis, especially in sampling various global model output onto satellite measurement "footprints" [e.g. Jiang et al. 2004, 2008, 2010; Li et al. 2007; Su et al. 2006b, 2008a, 2008b, 2011]. He has led or/and contributed to several important publications about using satellite datasets for model evaluations and model-data comparisons [e.g. Jiang et al. 2004, 2010; Li et al. 2005a, 2005b, 2007; Su et al. 2006b, 2010b; Waliser et al. 2009]. He will leverage the expertise gained from the research projects for the scientific applications of PAWS-CMDA.

Dr. Joao Teixeira of JPL, Collaborator, will serve as the main liaison between this proposal team and the climate modeling community. He has extensive research experience in climate model parameterization of the atmospheric boundary layer and evaluation of the climate model parameterization. He is currently leading a project to provide NASA satellite observational data and technical documents to the CMIP5 community for CMIP5 model evaluations. He is also the PI of the NASA ROSES MAP project entitled "Cloud Transitions in the Tropics and Sub-Tropics: Improving the Representation of Shallow Cumulus Convection in Coupled Systems". He is also currently leading the Pacific Cross-Section Intercomparison working group of Global Energy and Water Cycle Experiment Cloud System Study. Using these leadership roles, he will provide scientific inputs from the climate model community for the scientific requirements of the PAWS-CMDA development.

2 References and Citations

Fielding, Roy T. and Taylor, Richard N. (2002-05): "Principled Design of the Modern Web Architecture" (PDF), *ACM Transactions on Internet Technology (TOIT)* (New York: Association for Computing Machinery) **2** (2): 115150.

Jiang, J.H., S.D. Eckermann, D.L. Wu, J. Ma (2004): A Search for Mountain Waves in MLS Stratospheric Limb Radiances from the Winter Northern Hemisphere: Data Analysis and Global Mountain Wave Modeling, *J. Geophys. Res.* 109, D03107, doi:10.1029/2003 JD003974.

Jiang, J.H., H. Su, S. Pawson, H.C. Liu et al. (2010): Five-year (2004-2009) Observations of Upper Tropospheric Water Vapor and Cloud Ice from MLS and Comparisons with GEOS-5 analyses, *J. Geophys. Res.*, doi:10.1029/2009JD013256.

Jiang, J.H et al (2012): Evaluation of cloud and water vapor simulations in IPCC AR5 climate models using NASA A-Train satellite observaionts, WCRP Workshop on CMIP5 Climate Model Analyses, Honolulu, Hawaii, 5-9 March, 2012.

Jiang, J.H., H. Su, C. Zhai, et al. (2012): Evaluation of Cloud and Water Vapor Simulations in IPCC AR5 Climate Models Using NASA 'A-Train' Satellite Observations, *J. Geophys. Res.* in review.

Lee, S., L. Pan, R. Morris, P. von Allmen, W.-T. Chen, T. Kubar, J.-L. Li, and D. Waliser (2010): 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.

Lee, S. et al (2012): 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.

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

Lee, S., Hook Hua, Robert Carnright, John Coggi, and David Stodden (2007): Lessons Learned from Adapting Aerospace Engineering Tools to the Parallel and Grid Computing Environment, IEEE Aerospace Conference Proceedings, Big Sky, Montana, March 2007.

Li, J-L., J.H. Jiang, D.E. Waliser, and A.M. Tompkins (2007): Assessing Consistency between EOS MLS and ECMWF Analyzed and Forecast Estimates of Cloud Ice, *Geophys. Res. Lett.* 34, L08701, doi:10.1029/2006GL029022.

Li, J-L., D.E. Waliser, J.H. Jiang et al. (2005a): Comparisons of EOS MLS Cloud Ice Measurements with ECMWF analyses and GCM Simulations: Initial Results, *Geophys. Res. Lett.* 32, L18710, doi:10.1029/2005GL023788.

Li, Q.B., J.H. Jiang, et al. (2005b): Convective outflow of South Asian pollution: A global CTM simulation compared with EOS MLS observations, *Geophys. Res. Lett.*, 32, L14826, doi:10.1029/ 2005GL022762.

Li, J-L., J.H. Jiang, D.E. Waliser, and A.M. Tompkins (2007): Assessing Consistency between EOS MLS and ECMWF Analyzed and Forecast Estimates of Cloud Ice, *Geophys. Res. Lett.* 34, L08701, doi:10.1029/2006GL029022.

Pan, L., L. F. Bic, and M. B. Dillencourt (2001): "Distributed sequential computing using mobile code: Moving computation to data," in Proceedings of the 2001 International Conference on Parallel Processing (ICPP 2001), L. M. Ni and M. Valero, Eds. Los Alamitos, Calif.: IEEE Computer Society, Sept. 2001, pp. 77–84.

Pan, L., L. F. Bic, and M. B. Dillencourt (2002): "Shared variable programming beyond shared memory: Bridging distributed memory with mobile agents," in Proceedings of the 6th International Conference on Integrated Design & Process Technology (IDPT-2002), H. Ehrig, B. Kramer, and A. Ertas, Eds. Grandview, Texas: Society for Design & Process Science, June 2002.

Pan, L., L. F. Bic, M. B. Dillencourt, and M. K. Lai (2007): "NavP versus SPMD: Two views of distributed computation," in Proceedings of the Fifteenth IASTED International Conference on Parallel and Distributed Computing and Systems, T. Gonzalez, Ed., vol. 2, Algorithms. Anaheim, Calif.: ACTA Press, Nov. 2003, pp. 666–673.

Richardson, Leonard and Ruby, Sam (2007-05): *RESTful Web Services*, O'Reilly, ISBN 978-0-596-52926-0.

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.

Slingo, J. M., (1987): The development and verification of a cloud prediction scheme for the ECMWF model, Q.J.R. Meteorol. Soc., 113, 899-927.

Slingo, A., (1989): A GCM parameterization for the shortwave radiative properties of water clouds, J. Atmos. Sci., 46, 1419-1427.

Su, H., D.E. Waliser, J.H. Jiang, J-L. Li, W.G. Read, J.W. Waters, and A.M. Tompkins (2006), Relationships of upper tropospheric water vapor, clouds and SST: MLS observations, ECMWF analyses and GCM simulations, *Geophys. Res. Lett.* 33, L22802, doi:10.1029/2006GL027582.

Su, H., J.H. Jiang, D.G. Vane, and G.L. Stephens (2008a), Observed Vertical Structure of Tropical Oceanic Clouds Sorted in Large-scale Regimes , Geophys. Res. Lett. 35, doi:10.1029/2008GL035888.

Su, H., J.H. Jiang, G.L. Stephens, D.G. Vane (2008b), Variations of tropical upper tropospheric clouds with sea surface temperature and implications for radiative effects, *J. Geophys. Res.* 113, D10211, doi:10.1029/2007JD009624.

Su, H., J.H. Jiang, et al. (2011): Comparison of Regime-Sorted Tropical Cloud Profiles Observed by CloudSat with GEOS5 Analyses and Two General Circulation Model Simulations, *J. Geophys. Res.*, 116, D0910, doi:10.1029/2010JD014971, 2011.

Tiedtke, M.m (1993): Representation of clouds in large-scale models, Mon. Wea. Rev., 121, 3040-3061.

Waliser, D.E. et al. (2009): Cloud ice: A climate model challenge with signs and expectations of progress, *J. Geophys. Res.* 114, D00A21, doi:10.1029/2008JD010015 , 2009.

3 Biographical Sketch

3.1 Principal Investigator

Dr. Seungwon Lee

Jet Propulsion Laboratory, M/S 168-200, 4800 Oak Grove Drive, Pasadena, CA 91109; 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

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 aimed at evaluations of CMIP5 models for cloud-climate feedbacks. 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
1998 – 2002:	Teaching and Research Assistant, Department of Physics Ohio State University, Columbus OH

PROFESSIONAL ACTIVITIES AND SELECTED AWARDS

- Conference Session Organizer for High Capability Computing and Modeling, IEEE Aerospace Conference.
- NSF and JPL internal R&TD/DRDF Proposal Review Panelist.
- Journal Referee for Journal of Climate; Nanotechnology, Institute of Physics; Journal of Systems and Software; Transactions on Nanotechnology, IEEE; Journal of Physics: Condensed Matter, Institute of Physics; Physics Review Letter; Phys. Review B.
- Member, American Geophysical Union; American Physical Society; Material Research Society.
- Over 50 NASA Software/Tech Brief/Board Awards for developing new technology and software.
- Presidential Fellowship, Ohio State University, 2001-2002.

SELECTED PUBLICATIONS

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. 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.

S. Lee, R. P. Russell, W. Fink, R. J. Terrile, A. E. Petropoulos, and P. von Allmen, "Low-Thrust Mission Trade Studies with Parallel Evolutionary Computing", IEEE Aerospace Conference Proceedings, Big Sky, Montana, 2006.

S. Lee, W. Fink, R. P. Russell, R. J. Terrile, P. von Allmen, and A. E. Petropoulos, "Evolutionary Computing for Low-Thrust Navigation", AIAA 2005-6835 Paper, AIAA Space Conference, Long Beach, 2005.

S. Lee, A.E. Petropoulos, and P. von Allmen, "Low-Thrust Orbit Transfers Optimization with Refined Q-Law and Multi-Objective Genetic Algorithms", AAS 05-393 Paper, AAS/AIAA Astrodynamics Specialist Conference, Lake Tahoe, 2005

3.2 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

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, which serves as a basis of this proposed work, 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.

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

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
Research Scientist, Jet Propulsion Laboratory,
California Institute of Technology, 4800 Oak Grove Dr.
Pasadena, CA 91109, U.S.A.
Phone:(818)354-7135; Fax: (818)393-5065
Email: Jonathan.H.Jiang@jpl.nasa.gov

Education

B.Sc. (with Honors), Astrophysics, Beijing Normal University (1985);
M.Sc. Astrophysics, York University (1991);
Ph.D. Atmospheric Physics, York University (1996)

Recent Awards (selected)

NASA Exceptional Achievement Medal (2010)
NASA Space Act Board Award for Significant Contribution to National Space Program (2005)

Professional Experience

Scientist/Research Scientist (2001-present), Jet Propulsion Laboratory (JPL), Caltech, U.S.A.
Research Associate (1998-1999), Université du Québec à Montréal, Canada
Postdoctoral Research Fellow (1996-97), McGill University, Canada
Research Assistant (1992-95), York University, Canada
Assistant Astronomer (1989-1991) Institute for Space & Terrestrial Science, Canada

Relevant Experience

Dr. Jiang has been a member of the JPL's Aura MLS team since September 1999 and was the original co-author of MLS cloud forward model. He has extensive experience with A-Train data analysis, especially in collocating multiple satellite datasets and sampling various global model output onto satellite measurement "footprints". He has been the Principal Investigator for two NASA programs (ROSES AST and COUND), which focus on the evaluation of CMIP5 climate model simulations of clouds and water vapor using satellite observations. Dr. Jiang is a recipient of the NASA Exceptional Achievement Medal for his leadership in innovative applications of multiple satellite data for climate research. Dr. Jiang has authored and co-authored 81 research publications.

Selected Recent Publications

Jiang, J.H., H. Su, C. Zhai, et al., Evaluation of Cloud and Water Vapor Simulations in IPCC AR5 Climate Models Using NASA 'A-Train' Satellite Observations, *J. Geophys. Res.,* in review.

Su, H., **J.H. Jiang**, et al. Comparison of Regime-Sorted Tropical Cloud Profiles Observed by CloudSat with GEOS5 Analyses and Two General Circulation Model Simulations, *J. Geophys. Res.*, 116, D0910, doi:10.1029/2010JD014971, 2011.

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.

Su, H., **J.H. Jiang**, et al. Observed Increase of TTL Temperature and Water Vapor in Polluted Clouds over Asia, J. Climate 24, 11, 2728-2736, 2011.

Jiang, J.H. et al. Five-year climatology of upper tropospheric water vapor and cloud Ice from Aura MLS and GEOS-5, *J. Geosphys. Res.*, doi:10,1029/2009JD13256, 2010.

L'Ecuyer, T., and **J.H. Jiang**, Touring the atmosphere aboard the A-Train, an invited review article, *Physics 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, D20207, doi:10.1029/2009JD012421, 2009.

Jiang, J.H., Clean and polluted clouds: relationships among pollution, ice cloud and precipitation in South America, *Geophys. Res. Lett*. 35, L14804, 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 Microwave Limb Sounder, *Geophys. Res. Lett.* 34, L18812, doi:10.1029/2007GL030638, 2007

Dr. Chengxing Zhai
Jet Propulsion Laboratory, MS321-520 4800 Oak Grove Dr, Pasadena, CA 91109
818-434-9721(cell) 818-393-0758(work)

Technical background

20 years of experience in modeling, data analysis, and algorithm development with education in physics. Specialized in estimation, statistical analysis, spectral analysis, signal processing and detection, retrieval algorithm development, and instrument calibration technology. Experted in modeling systems using physical laws and empirical relations. Proficient in Matlab, C/C++, Fortran.

Education

- Ph.D., Physics, University of Washington, 1994
- M.S., Physics, University of Washington, 1991
- B.S., Physics, Peking University, China, 1989

Experience

9/2004 – present **Technologist** – Develop data analysis technologies for evaluating GCM models using NASA A-Train satellite observations; Optical modeling and algorithm development for SIM project, JPL.

1/2004 – 9/2004 **Software engineer** – Developed framework and adaptation software for rover real time control for the Mission Data System project, JPL.

10/2000–12/2003 **Software Engineer** – Developed framework and adaptation software for rover real time control for the Mission Data System project, QSS Group, Inc (TSEP for JPL) .

1/1999 –10/2000 **Software Engineer** – Developed flight software for GOES satellite, Hughes Space and Communications Company, El Segundo, CA.

1/1997 – 1/1999 **Software Engineer** – Developed satellite simulator, Raytheon System Company, El Segundo, CA

Honor

- NASA Group Achievement Award for demonstrating micro-arcsecond 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 Publications

[1] J. H. Jiang, 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).

[2] C. Zhai, M. Shao, G. Goullioud, and B. Nemati, *Micro-pixel accuracy centroid displacement estimation and detector calibration,* arXiv:1102.2248, accepted for publication in the Proceedings of Royal Society, A. (2011).

[3] J. D. Small, J. H. Jiang, H. Su, C. Zhai, *Relationship between aerosol and cloud fraction over Australia,* to accepted for publication in Geophys. Res. Lett. (2011).

[4] J. H. Jiang, H. Su, C. Zhai, et al, *Evaluation of Cloud and Water Vapor Simulations in IPCC AR5 Climate Models Using NASA "A-Train" Satellite Observations, submitted.(2011)*

4 Current and Pending Support

4.1 Current Awards

Principal Investigator Dr. Seungwon Lee				
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 tsengdar.j.lee@nasa.gov	9/1/2011-8/31/2013 $420K	0. 5

Co-Investigator Dr. Jonathan Jiang				
PI Name	**Award/Project Title**	**Program Info**	**Period & Total Budget**	**Commitment (Work Year)**
Jonathan Jiang	Utilizing NASA A-Train datasets for IPCC AR5 climate model evaluation	NASA ROSES COUND program tsengdar.j.lee@nasa.gov	9/1/2011-8/31/2013 $470K	0. 25
Jonathan Jiang	Utilizing Aura MLS and A-Train datasets to analyze and evaluate IPCC AR5 models in the upper troposphere	NASA ROSES AST program kenneth.jucks@nasa.gov	2/1/2011-1/31/2014 $499.5K	0.25
Steve Massie	Aerosol Effects on Cloud Heights and Precipitation	NASA CloudSat/CALIPSO science team recomplete	1/24/2010-1/23/2013	0.05
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 ROSES AST program kenneth.jucks@nasa.gov	2/1/2011-1/31/2014	0.05
Hui Su	Investigating the Influence of Asian Aerosol Pollution on the Water Vapor Transport from the Troposphere to the Stratosphere	NASA ROSES AST program kenneth.jucks@nasa.gov	2/1/2011-1/31/2014	0.05

Co-Investigator Dr. Lei Pan				
PI Name	**Award/Project Title**	**Program Info**	**Period & Total Budget**	**Commitment (Work Year)**
Paul von Allmen	Online Services for Correction of Atmosphere in Radar	NASA ROSES AIST program michael.s.seablom@nasa.gov	5/1/2009-5/1/2012 $1030K	0. 25

No current awards for Co-I Dr. Chengxing Zhai

4.2 Pending Awards

None for PI Dr. Seungwon Lee
None for Co-Is Dr. Lei Pan, Dr. Chengxing Zhai, and Dr. Jonathan Jiang

5 Budget Justification

5.1 Budget Narrative

5.1.1 Personnel and Work Effort

Name	Organization	Role	Work Commitment (FTE)	
			Year 1	Year 2
Dr. Seungwon Lee	JPL	Principal Investigator	0.20	0.20
Dr. Lei Pan	JPL	Co-Investigator	0.40	0.40
Dr. Chengxing Zhai	JPL	Co-Investigator	0.30	0.30
Dr. Jonathan Jiang	JPL	Co-Investigator	0.10	0.10
Dr. Joao Teixeira	JPL	Collaborator	0.10	0.10

5.1.2 Facilities and Equipment

A dedicated machine and disk array are required for the development, test, and deployment of the proposed information system. In the first year, we will purchase a mini cluster computer containing two compute nodes and a 100 TB disk array with a raid system to develop the proposed PAWS-CMDA system. NASA observational data and (re)analysis outputs will be available through DAACs, the JPL ESG node, and JPL local Science Computing Facilities (SCF) such as AIRS, MLS, and TES data processing servers. The mini cluster computer for the proposed work will have a direct access to the JPL ESG node and JPL local SCFs. In the second year, we will extend the computational resource of the mini cluster computer and the disk storage of the disk array by buying two additional compute nodes and a 100 TB disk array with a raid system. Since the cost of the computing resource and disk array reduces significantly each year, it is advantageous to build the mini-computing facility in two stages. The reason for acquiring multiple compute nodes is to develop the tool with scalable mapping of services to a network of servers. This capability is necessary to support multiple user requests simultaneously and to handle production run use case scenarios, which involve computation and data intensive processes.

We will also utilize the existing computing and data resources that we own and maintain, while we are building the dedicated machine and disk array for the proposed work. The current resource consists of five compute nodes (including one head node) with 8 cores in each node and a 100 TB disk array. The resource is currently used for operating our existing offline analysis tool and conducting research. Because of its heavy use, it would not be possible to use the existing system as a main computing resource for the proposed work. We are therefore proposing to purchase a dedicated machine and disk array to support the proposed work. The computational power and disk size needed for the proposed work was estimated based on our experience with the existing offline analysis tool.

5.1.3 Budget Summary

PAWS-CMDA: Parallel Web-Service Climate Model Diagnostic Analyzer

ROSES CMAC
Timephased Cost Estimate Sheet
Dollars (Does not include Gov't Co-I's)

	Oct 2012 - Sep 2013		Oct 2013 - Sep 2014		Total Program		
Hours / (FTEs)							
Seungwon Lee (PI)	377	(0.21 FTE)	376	(0.21 FTE)	753	(0.42 FTE)	Hours / (FTEs)
Lei Pan (Co-I)	753	(0.43 FTE)	753	(0.43 FTE)	1,506	(0.86 FTE)	Hours / (FTEs)
Chengxing Zhai (Co-I)	565	(0.32 FTE)	565	(0.32 FTE)	1,130	(0.64 FTE)	Hours / (FTEs)
Jonathan Jiang (Co-I)	188	(0.11 FTE)	188	(0.11 FTE)	377	(0.22 FTE)	Hours / (FTEs)
Total Hours:	1,884	(1.07 FTE)	1,882	(1.07 FTE)	3,766	(2.14 FTE)	Subtotal
Amount		$115,850		$119,230		$235,080	JPL Direct Labor Cost w/o Fringe
Fringe		$59,060		$60,180		$119,240	Fringe
Category A		$0		$0		$0	Cat A Direct Labor Cost
Total Direct Compensation (includes Employee Benefits)		$174,910		$179,410		$354,320	Subtotal
Travel		$0		$4,000		$4,000	Direct Travel Cost
JPL Services		$0		$0		$0	Direct Services Cost
Procurements							
Chargebacks		$10,760		$10,760		$21,520	Direct Chargebacks cost
Subcontracts		$0		$0		$0	Direct PS cost
Procurement RSA		$0		$0		$0	Direct RSA cost
Purchase Orders		$10,000		$10,000		$20,000	Direct PM cost
Caltech Transfers		$0		$0		$0	Direct CT cost
Multi-Program Support		$23,500		$23,930		$47,430	Direct MPS cost
Total Direct Costs		**$219,170**		**$228,100**		**$447,270**	Subtotal
Allocated Direct Charge		$52,670		$53,560		$106,230	Total ADC
General & Admin		$30,980		$31,420		$62,400	Total G&A
Reserves (Burdened)		$0		$0		$0	
Total JPL Costs		**$302,820**		**$313,080**		**$615,900**	Subtotal
Government Co-I's Not in JPL's Costs		**$0**		**$0**		**$0**	Bypass
Total Costs		**$302,820**		**$313,080**		**$615,900**	Subtotal

5.1.4 Rationale and Basis of Estimate

The cost of the proposed work was prepared using JPL's Pricing System and the current internally published Cost Estimation Rates and Factors dated October 2010. 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. They generate the resource estimates for labor, procurements, travel, and other direct costs for each work element of the proposal. The resource estimates are aggregated and priced using JPL's Pricing System. JPL's process ensures that 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.

A discussion of the JPL cost accumulation system follows:

JPL Cost Accumulation System

Introduction

All costs incurred at the Laboratory, including JPL applied burdens, are billed to the Government as direct charges at the rates in effect at the time the work is accomplished.

Allocated Direct Costs

Allocated Direct Cost (ADC) rates contain cost elements benefiting multiple work efforts, including Project Direct, MPS, and Support and Services activities. Rate applications for cost estimates are specific to the given category as stated below:

1) Engineering and Science (E&S)
2) Procurement: Purchase Order, Subcontract, Research Support Agreement (RSA)
3) General and Administrative (G&A): Basic, RSA
4) Specialized G&A applications: Remote Site

The accounting process fully distributes these costs to the respective project/task(s).

Multiple Program Support

The Multiple Program Support (MPS) rate applies costs for program management and technical infrastructure. Cost estimates and system application tools will apply the composite rate to all project direct hours charged to projects managed by JPL.

Employee Benefits

All costs of employee benefits are collected in a single intermediate cost pool, which is then redistributed to all cost objectives as a percentage of JPL labor costs, including both straight-time and overtime. Functions and activities covered by this rate include paid leave, vacations, and other benefits including retirement plans, group insurance plans, and tuition reimbursements.

For this proposal the estimated costs have been derived in the same manner as stated above. However, presentation of the estimated costs in the required tables has been adapted in the following ways:

1. The costs for Employee Benefits are included in the Direct Labor costs stated in this proposal.
2. Engineering and Science ADC and Procurement ADC along with MPS costs are displayed in the “Other” category in the Other Direct Costs section.
3. G&A is shown in the Facilities and Administrative Costs section.
4. JPL’s forecasted labor rates equal an hourly laboratory-wide average for each job family and are further broken down by career level within the job family. Labor cost estimates apply the family average or family average career level rate to the estimated work hours. An actual individual’s labor is considered discrete and confidential information and is only released on an exception basis and only if a statement of work identifies that specific individual as the only one able to perform a task. The use of family average or family average career level rates in consistent with the JPL CAS disclosure statement and the Cost Estimating Rates and Factors CDRL published in response to a requirement in NASA prime contract NAS7-03001 I-10 (d) (1).

The proposed budget of the NRA proposal also covers labor costs for serving on NASA peer-review panels and advisory committee at the request of NASA discipline scientists or program managers.

5.3 Budget Details – Year 1

Direct Labor Costs

- Dr. Seungwon Lee is the PI and will oversee all aspects of the proposed work. Her time commitment is 0.2 WY with $24,510 requested salary and $12,490 fringe benefits.
- Dr. Lei Pan is a Co-I on this effort and will be the lead architect and developer of the proposed system. His time commitment is 0.4 WY with $42,280 requested salary and $21,550 fringe benefits.
- Dr. Chengxing Zhai is a Co-I on this effort and will develop the data analysis components. His time commitment is 0.3 WY with $36,770 requested salary and $18,750 fringe benefits.
- Dr. Jonathan Jiang is a Co-I on this effort and will lead the effort of applying and validating the proposed tool with his specific scientific investigations. His time commitment is 0.1 WY with $12,290 requested salary and $6,270 fringe benefits.

Other Direct Costs

Subcontracts/Subawards

- Desktop Network Chargebacks (calculated at $5.70/hr.): All JPL computers are subject to a monthly service charge that includes hardware, software, and technical support. ($10,760)

Equipment

- A server with two compute nodes and a 100TB raid system will be purchased for the development of the proposed tool ($10,000).

Indirect Costs

- Multiple Program Support (MPS): $23,500
- Allocated Direct Costs (ADC): $52,670
- Applied General ADC: $30,980

Total Estimated Costs for Year 1: $302,820

5.4 Budget Details – Year 2

Direct Labor Costs

- Dr. Seungwon Lee is the PI and will oversee all aspects of the proposed work. Her time commitment is 0.2 WY with $25,250 requested salary and $12,740 fringe benefits.
- Dr. Lei Pan is a Co-I on this effort and will be the lead architect and developer of the proposed system. His time commitment is 0.4 WY with $43,470 requested salary and $21,960 fringe benefits.
- Dr. Chengxing Zhai is a Co-I on this effort and will develop the data analysis components. His time commitment is 0.3 WY with $37,850 requested salary and $19,110 fringe benefits.
- Dr. Jonathan Jiang is a Co-I on this effort and will lead the effort of applying and validating the proposed tool with his specific scientific investigations. His time commitment is 0.1 WY with $12,660 requested salary and $6,370 fringe benefits.

Other Direct Costs

Subcontracts/Subawards

- Desktop Network Chargebacks (calculated at $5.70/hr.): All JPL computers are subject to a monthly service charge that includes hardware, software, and technical support. ($10,760)

Equipment

- A server with two compute nodes and a 100TB raid system will be purchased for the development of the proposed tool ($10,000).

Travel

- The PI Dr. Seungwon Lee will travel to San Francisco, CA to present the scientific applications of the proposed PAWS-CMDA system at the annual American Geophysical Union Fall meeting. ($2,000 total – air fair $400, 4 night hotel lodging $800, registration fee $500, per diem and ground transportation $300).
- The Co-I Dr. Lei Pan will travel to San Francisco, CA to present the technical approach and development of the proposed PAWS-CMDA system at the annual American Geophysical Union Fall meeting. ($2,000 total – air fair $400, 4 night hotel lodging $800, registration fee $500, per diem and ground transportation $300).

Indirect Costs

- Multiple Program Support (MPS): $23,930
- Allocated Direct Costs (ADC): $53,560
- Applied General ADC: $31,420

Total Estimated Costs for Year 2: $313,080