NASA Science Mission Directorate
Research Opportunities in Space and Earth Sciences
NNH17ZDA001N-CMAC
A.45 Computational Modeling Algorithms and Cyberinfrastructure

NASA's Science Mission Directorate, NASA Headquarters, Washington, DC, has selected proposals for the Computational Modeling Algorithms and Cyberinfrastructure (CMAC) Program in support of the Earth Science Division (ESD). This program element invests in research and development activities to optimize the products and services at high-end computing (HEC) centers to increase the productivity of the users who use HEC to produce modeling products and the users who need to analyze the modeling results using the HEC resources. CMAC builds advanced modeling infrastructure used at NASA computing centers to support Earth system science investigations while fundamentally utilizing both models and data.

This solicitation focuses on the information technologies required to enable and facilitate NASA's Earth system modeling efforts. The research and development themes specific to this solicitation includes the enhancement of existing high-performance computing capabilities to build a comprehensive and coherent portfolio of Earth science data analytic services. In addition, the solicitation calls for the development of a strategy to restructure NASA's GEOS atmospheric model to distributed heterogeneous multicore computing architecture. The ESD has selected 5 out of a total of 13 proposals received in response to this solicitation. The total funding for these investigations, over a period of two years, is approximately $2 million.

Huikyo Lee/Jet Propulsion Laboratory
Multi-Resolution Investigation of Climate Models Using NASA's High-End Computing Resources: A Parallel Version of Regional Climate Model Evaluation System Enhanced by HEALPix
17-CMAC17-0002

We propose to advance the NASA Earth eXchange (NEX) platform by deploying the enhanced version of the Jet Propulsion Laboratory's Regional Climate Model Evaluation System (RCMES) as a web-based service for evaluating climate models at various spatial resolutions. RCMES is a comprehensive suite of open-source software resources to standardize and streamline the process of interacting with observational data and climate model output to conduct climate model evaluations with special emphasis on climate changes at regional and local scales. We intend to enhance RCMES by employing the Hierarchical Equal Area isoLatitude Pixelization (HEALPix). HEALPix is an open-source library for fast and robust multi-resolution analysis of observational and model datasets regridded into HEALPix pixels, which have been widely used by astronomers and planetary scientists. RCMES enhanced by HEALPix will also have parallel processing capabilities by leveraging SciSpark. The unique capabilities of SciSpark include the handling of massive datasets using parallel computing. Therefore, a parallel version of RCMES enhanced by HEALPix will provide a high-performance, service-based cyberinfrastructure to support large-scale Earth science analytics that can increase

utility of the datasets and high-end computing (HEC) resources available from the NASA Earth eXchange (NEX). Leveraging NASA's HEC resources is essential for the proposed web service to broaden the user base of RCMES and NEX.

The principal science objective of this project is to provide quantitative and robust evaluations of spatial patterns simulated by climate models across multiple scales: comparison of spatial features at coarse (e.g. 100 km) and fine scales (e.g. 1-10 km) separately between observations and models. Our specific objectives are to:

I. Reengineer HEALPix for RCMES to carry out fast and accurate multi-resolution evaluation of the climate models participating in the Coupled Model Intercomparison Project (CMIP),

II. Streamline the data processing of RCMES by applying SciSpark and generate data-intensive examples of regridding and comparing observed and simulated spatial patterns at different resolutions,

III. Refactor and integrate RCMES graphical user interface (GUI) as a web service that can be executed using NASA's HEC resources.

Both RCMES and NEX are ongoing investments by NASA to advance infrastructure for the U.S. National Climate Assessment (NCA). Our RCMES service through NEX will use a variety of observational datasets from the Observations for Model Intercomparison Project (Obs4MIPs) and NASA satellites, and climate simulations from CMIP5 and NEX-Global Daily Downscaled Projections (NEX-GDDP) to complement NASA's support to advance NCA-relevant science.

Seungwon Lee/Jet Propulsion Laboratory
Climate Model Diagnostic Analyzer Services on NASA Earth Exchange Platform
17-CMAC17-0001

What will be done? We propose to integrate Climate Model Diagnostic Analyzer (CMDA) services into the NASA Earth Exchange (NEX) platform as participating services. CMDA is a collection of analysis web services for multi-aspect physics-based and phenomenon-oriented climate model performance evaluation and diagnosis through the comprehensive and synergistic use of multiple observational data, reanalysis data, and model outputs. CMDA has been developed under the investment of NASA ROSES CMAC and AIST programs, and has established a real use case through Jet Propulsion Laboratory (JPL) Center for Climate Sciences (CCS) Summer School since 2014. NEX is a platform for scientific collaboration, knowledge sharing and research for the Earth science community, maintained by the High-End Computing Capability (HECC) project at NASA Ames Research Center. OpenNEX is the public cloud version of NEX, and will be the integration target for the proposed work. NEX is one of the two cyberinfrastructure facilities that this CMAC Funding Announcement (FA) has selected to build centralized large-scale Earth Science analytic services cyberinfrastructure. Therefore, this proposal responds to this FA element 2.1 Earth Science Analytic Services Cyberinfrastructure.

Why is it needed? Understanding climate change at various temporal and spatial scales and related human influences becomes ever more important as the public’s concerns

about climate changes have steadily grown over the past years. However, achieving this goal has remained a tremendous challenge to the Earth Science community, manifested by the persistence of the large spreads in the model projections from sub-grid scale processes (e.g. cumulus parameterization) to global-scale variability (e.g. mean global surface air temperature increase). These uncertainties in the model predictions are due to many complex climate feedback processes that are either poorly represented or not yet included in the models. A better understanding of the processes and in return a better representation of the processes in the models are needed to improve the model fidelity for the current climate representation and future climate prediction. Process-level model evaluations with high-resolution simulations require unprecedentedly high data-volume throughput and high computational efficiency. These challenges are too big for individual researchers to tackle, and will require the whole community to share services, data, knowledge, computing capabilities, and data storage. The vision of this CMAC FA is to build a comprehensive and coherent portfolio of Earth science data analytic services and the vision is achievable only if we build the cyberinfrastructure together by combining existing mature and various services.

What is the value added? The proposed work will contribute to the portfolio of Earth science data analytic services by providing mature and diverse analytic services for Earth science data. While its focus application area is climate model evaluation and model-data inter-comparison, the CMDA services covers a wide range of analytic services from a single-variable analysis to multi-variable analysis, from temporal analysis to spatial analysis, and from low-level data manipulation to high-level data generation. The CMDA services are implemented as web services and have been proven to improve the productivity of scientists to prepare and analyze data by several years of applications to the real-world use cases in the JPL CCS Summer School. Thanks to its maturity and diversity, the CMDA services are one of the best suited candidates to incorporate into the existing cyberinfrastructure to build a coherent and comprehensive portfolio of Earth science data analytic services.

Jianwu Wang/University of Maryland Baltimore County
Efficient and Flexible Aggregation and Distribution of MODIS Atmospheric Products Based on Climate Analytics-As-A-Service Framework
17-CMAC17-0004

Remote sensing products of atmospheric clouds and aerosols from MODIS are one of the most widely used data sets for study Earth-Atmosphere systems. Support is requested to refactor the current MODIS atmosphere Level-3 product aggregation and distribution system to drastically reduce data generation and processing time, and to facilitate user-defined customization, using the framework of Advanced Data Analytics Platform (ADAPT) and Data Analytics and Storage System (DASS) provided by the NASA Center for Climate Simulation (NCCS). We believe the Climate Analytics-as-a-Service (CAaaS) framework provided by ADAPT and DASS is well suited for the proposed project. Specifically, the major objectives are to:

1. From a system developer perspective, revolutionize the efficiency for generating the current MODIS Level-3 products. The objective can be achieved through scalable, generation of gridded (Level-3) MODIS products from pixel-level (Level-2) products based on the Analytic Service design of CAaaS and the Big Data analytics capability of ADAPT/DASS.
2. From a system customer perspective, enable flexible, on-demand MODIS data access and analytics capability, and thereby shift the Level-3 data paradigm from static product to science-oriented service. The objective can be achieved through customizable generation and science-oriented data distribution of Level-3 MODIS data using the Analytic and Persistent Service design of CAaaS and the Big Data analytics and storage capability of ADAPT/DASS.

The following engineering and research activities are proposed to be undertaken to achieve the above objectives:
1. Refactoring existing MODIS product storage and distribution
a. Storing MODIS Level-2 and Level-3 product on DASS/ADAPT.
b. Supporting customized data discovery/access services based on user defined time, geographical location and variables.
2. Refactoring existing MODIS product generation and analytics
a. Parallelizing existing MODIS Level-2 to Level-3 aggregation programs by following the MapReduce programing model and the microservice model.
b. Supporting user-defined Level-2 data aggregation by integrating flexible data access and analytics.
3. Evaluation
a. Execution Performance Evaluation.
b. Usability Evaluation.
c. Extensibility Evaluation.

The deliverables of the proposed project include: 1) open source software implemented based on the activities; 2) openly available documents on software design, deployment, testing; 3) openly available evaluation reports on performance, usability and extensibility of the developed and deployed system on ADAPT/DASS.

Brian Wilson/Jet Propulsion Laboratory
Notebook Analytics: SciSpark Parallel Analytics and the Open Climate Workbench for Model Evaluation
17-CMAC17-0012

Understanding climate change at various temporal and spatial scales and related human influences becomes ever more important as the public's concerns about climate changes have steadily grown over the past years. Model simulations are achieving greater fidelity by increasing spatial grid resolutions (7 km or better) and generating output variables at hourly, 30-minute, or even 5-minute temporal cadences for process and mission formulation studies. For example, the NCCS GEOS-5 Nature run (G5NR) at 7 km resolution covers a two-year period from June 2005 through June 2007 with 30-minute

outputs. The full G5NR output with many variables is ~4 Petabytes, and so it is not practical to move it all across the Internet backbone. Thus, there is a need for a parallel analytics facility for model evaluation within the NCCS infrastructure.

To meet this challenge, we propose to deploy two cutting-edge capabilities, developed under prior NASA funding, into the ADAPT infrastructure at NCCS: Model Evaluation services using the RCMES/Open Climate Workbench library and Parallel Climate Analytics using the SciSpark Map-Reduce cluster computing technology. We will provide both remote access to the services using standard Web Services API's (HTTP GET/POST) and a direct iPython Notebook environment in which climate scientists can compare gridded satellite data to high-resolution model runs at high temporal cadence (e.g. 7 km and 30 minutes). By providing parallel algorithms that can scale to high-resolution comparisons and process-level or phenomena studies, new science and mission formulation activities will be enabled.

Thus, the primary objective is the creation of an eScience Environment in which collaborative science is done on-line where applications and/or web service calls can be assembled into custom workflows and published in shareable eNotebooks. The resulting open-source environment will include:

1. Parallel Analytics -- Integration of SciSpark and custom analytics capable of running highly optimized parallel analytics on extremely large climate data sets through electronic notebooks.

2. Multi-Model Evaluations -- Multiple-objective optimization library for climate model evaluations to enable multi-model, weighted ensembles with uncertainty quantification. In addition, the full suite of RCMES/OCW capabilities will be integrated, including data readers, reformatting, spatial & temporal regridding, metrics calculation, and custom plot types (e.g. bias & variance plots, pattern correlation, Taylor diagram, etc.)

3. Productive "Live Code" eNotebooks -- All analytics services will be callable from web services API's, python scripts, and "live" iPython Notebooks that invoke parallel operations on the SciSpark cluster.

The Notebook Analytics system will be exercised on numerous model evaluation scenarios, some drawn from prior RCMES and CORDEX/CMIP work, and also new scenarios including a detailed aerosol science analysis using the G5NR and G5NR-Chem Nature runs and the latest AOD retrievals from MODIS and MISR.

This project responds to objective 2.1, "Earth Science Analytic Services Cyberinfrastructure", of the CMAC solicitation. We are bringing multiple capabilities to NCCS ADAPT where they are collocated with the model runs, including: parallel model evaluation algorithms callable as web services, easy to use "live code" eNotebooks, convenient access to Obs4MIPS & ana4MIPS observations, and parallel access to resident model grids (e.g. G5NR and CMIP5).

Jia Zhang/Carnegie Mellon University
NEX App Store
17-CMAC17-0008

NASA Earth Exchange (NEX), and her public cloud version OpenNEX, have become platforms supporting scientific collaboration, knowledge sharing and research for the entire Earth science community. To date, a number of custom tools and capabilities have been integrated into the platforms. However, such integration has to undergo a case-by-case manual process thus lacks scalability. With an explosive growth of datasets and tools, the community is in need of infrastructure tools to support the organization of their work and sharing their knowledge on the (Open)NEX.

Therefore, we propose this timely CMAC project aiming to build an App Store onto OpenNEX as a building block. Our vision is that the climate data analytics tools/programs can be easily uploaded, shared, organized, searched, and recommended like photos and videos on the YouTube. We intend to build a community-driven app store with a full understanding of the needs of Earth science researchers, and make it easy for the next-generation Earth scientists to contribute to a pool of data analytics methods, build communities, and form collaborative relationships.
The key functionalities of the proposed repository are four-fold:

-CMDA service publication: Define and design a technique that either absorbs user-generated HTML pages or automatically generates browser-neutral frontend HTML pages to CMDA services to be shared by the community.
-CMDA service containerization: Develop DevOps technique that allows registered CMDA services to be executed on Amazon cloud as microservices.
-CMDA service execution provenance management: Keep track of service execution history with user profile, execution configuration, and execution context such as the purpose of the study, comments from users, and ratings from other users. Such information as service provenance will support reproducibility of scientific experiments.
-Personalized CMDA service search and recommendation: Provide semantic search and historical usage-supported recommendation for CMDA web services, leveraging machine learning and deep learning techniques.

As a concrete show case, we plan to study how to publish, share, and execute Climate Model Diagnostic Analyzer (CMDA) services onto the OpenNEX. Scalability, sustainability, extensibility, usability, adaptability, security and privacy will be considered.