U.S. DEPARTMENT OF ENERGY | Office of Science

DOE/SC-ARM-15-039

LES ARM Symbiotic Simulation and Observation (LASSO) Implementation Strategy

WI Gustafson Jr.
AM Vogelmann

September 2015

ARM
CLIMATE RESEARCH FACILITY

DISCLAIMER

DOE/SC-ARM-15-039

LES ARM Symbiotic Simulation and Observation (LASSO) Implementation Strategy

WI Gustafson Jr.
AM Vogelmann

September 2015

Work supported by the U.S. Department of Energy,
Office of Science, Office of Biological and Environmental Research

Executive Summary

This document illustrates the design of the Large-Eddy Simulation (LES) ARM Symbiotic Simulation and Observation (LASSO) workflow to provide a routine, high-resolution modeling capability to augment the U.S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Climate Research Facility's high-density observations. LASSO will create a powerful new capability for furthering ARM's mission to advance understanding of cloud, radiation, aerosol, and land-surface processes. The combined observational and modeling elements will enable a new level of scientific inquiry by connecting processes and context to observations and providing needed statistics for details that cannot be measured. The result will be improved process understanding that facilitates concomitant improvements in climate model parameterizations. The initial LASSO implementation will be for ARM's Southern Great Plains site in Oklahoma and will focus on shallow convection, which is poorly simulated by climate models due in part to clouds' typically small spatial scale compared to model grid spacing, and because the convection involves complicated interactions of microphysical and boundary layer processes.

In this document, we present the prototype LASSO workflow and provide implementation strategies that, through working with ARM software developers, will result in a functioning workflow at the end of the two-year project. The plan outlines the following:

- Implementation of high-resolution modeling constrained by ARM observations
- Construction of "data cubes" that combine observations, model output, and metrics into a unified package
- Recommendations for analysis and visualization tools for efficient user access to LASSO data cubes.

The design and recommendations will be formulated based on extensive testing and prioritized cost-benefit analyses of options that will be vetted with ARM leadership and staff, and the user community to ensure that best practices are used to obtain a result that will be of immediate use to the community. Our development will be guided by community-based science drivers, will use a tiered implementation approach, will aim for flexibility and extensibility, and will seek robustness and reproducibility. This will ensure the final product meets ARM's current requirements and future evolution in program needs, such as new observations, additional science foci and cloud types, and implementation at other ARM sites. Ultimately, this new modeling ability will play a pivotal role in realizing the full potential of ARM observations to accelerate the DOE Climate and Environmental Sciences Division's mission that includes developing new process-level understanding and using that knowledge to improve climate models.

Acronyms and Abbreviations

3DVar	three-dimensional variational-analysis data assimilation
AERI	Atmospheric Emitted Radiance Interferometer
AERIoe	AERI Optimal Estimation
AGU	American Geophysical Union
ARM	DOE Atmospheric Radiation Measurement
ARSCL	ARM Remotely-Sensed Cloud Locations VAP
ASR	DOE Atmospheric System Research Program
BNL	Brookhaven National Laboratory
CAM5	Community Atmosphere Model version 5
CARES	Carbonaceous Aerosols and Radiative Effects Study
CESD	DOE Climate and Environmental Sciences Division
CHAPS	Cumulus Humilis Aerosol Processing Study
CLASIC	Cloud and Land Surface Interaction Campaign
CLUBB	Cloud Layers Unified By Binormals parameterization
Co-PI	co-principal investigator
CPU	central processing unit
CuP	Cumulus Potential shallow cloud parameterization
DOE	U.S. Department of Energy
ECMWF	European Centre for Medium-Range Weather Forecasts
ESM	DOE Earth System Modeling Program
EnKF	ensemble Kalman filter data assimilation technique
FASTER	Fast-physics System Testbed and Research project
FDR	Fourteen Data Rate (14 Gb s−1)
G&A	General and Administrative
GCM	global climate model
GISS	NASA Goddard Institute of Space Studies
GSI	Gridpoint Statistical Interpolation (data assimilation software)
IDL	Interactive Data Language
IMADA-AVER	Investigacion sobre Materia Particulada y Deterioro Atmosferico-Aerosol and Visibility Research Campaign
KPT	KNMI Parameterization Testbed
LASSO	LES ARM Symbiotic Simulation and Observation
LES	large-eddy simulation
MILAGRO	Megacity Initiative: Local and Global Research Observations Campaign
MS-DA	Multi-scale Data Assimilation framework
NERSC	DOE National Energy Research Scientific Computing Center

NWS	National Weather Service
NASA	National Aeronautic and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Center for Environmental Research
NERSC	DOE National Energy Research Scientific Computing Center
NFS	Network File System
OME	ARM Data Product Registration Tool
PDF	probability distribution function
PI	principal investigator
PNNL	Pacific Northwest National Laboratory
Py-ART	Python ARM Radar Toolkit
RACORO	Routine ARM Aerial Facility Clouds with Low Optical Water Depths Optical Radiative Observations
Reff	effective radius
RGCM	DOE Regional and Global Climate Modeling Program
SAM	System for Atmospheric Modeling
SCM	single-column model
SGP	Southern Great Plains site
TCAP	Two Column Aerosol Project
TOPSE	Tropospheric Ozone Production about the Spring Equinox Campaign
UCLA	University of California, Los Angeles
UNICON	Unified Convection parameterization
UV-CDAT	Ultrascale Visualization Climate Data Analysis Tools
VAP	value-added product
VARANAL	Variational Analysis VAP
VTMX	Vertical Transport and Mixing Campaign
WRF	Weather Research and Forecasting model

Contents

Figures

Tables

1.0 Introduction

1.1 Context and Science Motivation

The United States (U.S.) Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Climate Research Facility is entering an exciting new era where ARM observing systems will be augmented by routine, high-resolution modeling to enable better understanding of cloud, radiation, aerosol, and land-surface processes, ultimately furthering the facility's mission "to improve the understanding and representation, in climate and Earth system models, of clouds and aerosols as well as their interactions and coupling with the Earth's surface" [*DOE ARM Climate Research Facility*, 2013]. The new modeling capability will fit within the greater ARM observational capabilities and is a natural extension of the "megasite" concept for high-density observations, whose implementation has already begun. Measurements routinely made by ARM have high spatial and temporal resolutions that provide a unique capability for detailed study of climatically important small- and mesoscale processes. For example, shallow cumuli at the ARM Southern Great Plains site (SGP) in Oklahoma have an average shortwave radiative forcing of $-45.5\ \mathrm{W\ m^{-2}}$ [*Berg et al.*, 2011] and mean spatial scale around 1.0 km [*Berg and Kassianov*, 2008], which means they are an important part of the radiation budget, yet are sub-grid scale for all but the highest resolution models and must be represented by parameterizations. ARM observational capabilities have been significantly enhanced over the decades. However, realizing the full potential of ARM observations, developing new process-level understanding, and using that knowledge to accelerate DOE Climate and Environmental Sciences Division's (CESD's) mission of improving global climate models (GCMs) requires a commensurately high-resolution, routine modeling component that provides a mechanism for synthesizing numerous types of atmospheric observations.

The initial focus of ARM modeling at SGP will be on shallow convection, which is poorly simulated by climate models due in part to clouds' typically small spatial scale compared to model grid spacing, and because the convection involves complicated interactions of microphysical and boundary layer processes. The recent high-resolution workshop report [*DOE CESD*, 2014a] highlights several scientific challenges related to shallow convection that will be addressable by ARM's new modeling paradigm, such as:

- *Biases in GCMs over the Great Plains*. A warm temperature bias has been suggested to result from an underestimate of clouds and precipitation in the region, leading to a high-biased surface shortwave flux and a deficit in soil moisture; both amplify the temperature bias [*Klein et al.*, 2006; *Ma et al.*, 2014]. The high-resolution modeling capability would enable new scientific insights associated with the role played by various processes, such as cumulus cloud-top detrainment and its effect on the diurnal cycle of continental convection [*Guichard et al.*, 2004], needed to eliminate this bias.
- *Representation of heterogeneities in land-surface effects*. Shallow convection is linked to variability in radiation, land-use, and soil moisture through surface sensible and latent heat fluxes. Agriculture around SGP will allow study of the impacts of spatial and seasonal variability from changes in evapotranspiration related to irrigation and the growth-harvest cycle, which are not represented in all GCMs, yet are important for improving simulation of agricultural regions [*Adegoke et al.*, 2007; *Qian et al.*, 2013].

- *New approaches to parameterizing convective clouds*. The routine nature of the simulations will provide the statistics necessary to develop the next generation of cloud parameterizations for GCMs, particularly those requiring turbulence statistics that cannot adequately be measured. The simulations will also enable differentiating statistically significant signals from background variability for connecting land-surface conditions and clouds, and understanding aerosol impacts on clouds.

Routine, high-resolution modeling integrated with ARM's extensive measurements will provide information needed to develop methods for capturing effects such as these in GCMs.

This document outlines the design and development of a robust and efficient workflow for producing routine large-eddy simulation (LES) model output at the SGP site to support key DOE CESD science priorities. We have developed our overall approach in the context of needs and timelines presented in ARM's Decadal Vision [*DOE ARM Climate Research Facility*, 2014], the recent high-resolution modeling workshop report [*DOE CESD*, 2014a], and discussions with ARM staff. The workflow will encompass data gathering, steps necessary to run the simulations, and assembling of observations and model output into a package for easy analysis and visualization. The resulting capability will be critical for addressing a wide range of science questions of interest to users, including the science themes described above, and the workflow will be flexible and extensible to accommodate future evolution in program needs.

Through this project we will work closely with the ARM team to provide the expertise needed to design the workflow for deploying the high-resolution simulations. Our workflow recommendations will be guided by science needs and feedback from the CESD community for concepts including model configurations, operating scenarios, input data and forcing data sets, integration of model output with observations, and post-simulation analyses and visualization tools needed to extract the maximum value from the simulations. We will develop a list of options and recommendations for the ARM leadership team regarding components of the workflow and staged implementation strategies. Based on the original call for white papers and discussions with the ARM Technical Director, we will work with ARM software developers outside this project who will contribute the effort necessary to convert our recommendations and prototype workflow components into an operational workflow within the context of ARM's overall priorities, with the goal of a functioning workflow at the conclusion of the two-year project.

1.2 Designing the LASSO Workflow

Because the value of the new ARM modeling paradigm lies in both its routine operation and "tying together" of observations and modeling, we call our approach the *LES ARM Symbiotic Simulation and Observations* (LASSO) workflow. The LASSO workflow will be designed for extensibility, and its development will be closely coupled to the ARM Data Archive infrastructure so that, as LASSO matures, it will permit ready extension from simulating shallow clouds at the SGP site to include land-atmosphere coupling, deep convection, cloud-aerosol interactions, other science foci, and implementation at other ARM sites.

The LASSO workflow will run high-resolution simulations incorporating megasite observations; provide a synthesis of simulations and observations to obtain the best multi-dimensional characterization of the atmosphere, referred to as a "4-D data cube" in the Atmospheric Testbed Workshop report [*DOE CESD*, 2014b] and hereafter as a "data cube" for generality; and it will include tools for users to efficiently analyze and visualize the data cube for their research objectives. Figure 1 shows our proposed flow of LASSO, within the large blue rectangle, combined with how LASSO fits into the larger ARM and community context. ARM and external data will feed into LASSO, and the resulting output will be used to develop improved process-level understanding and climate models. Subsequent user feedback will help refine LASSO and the observation network.

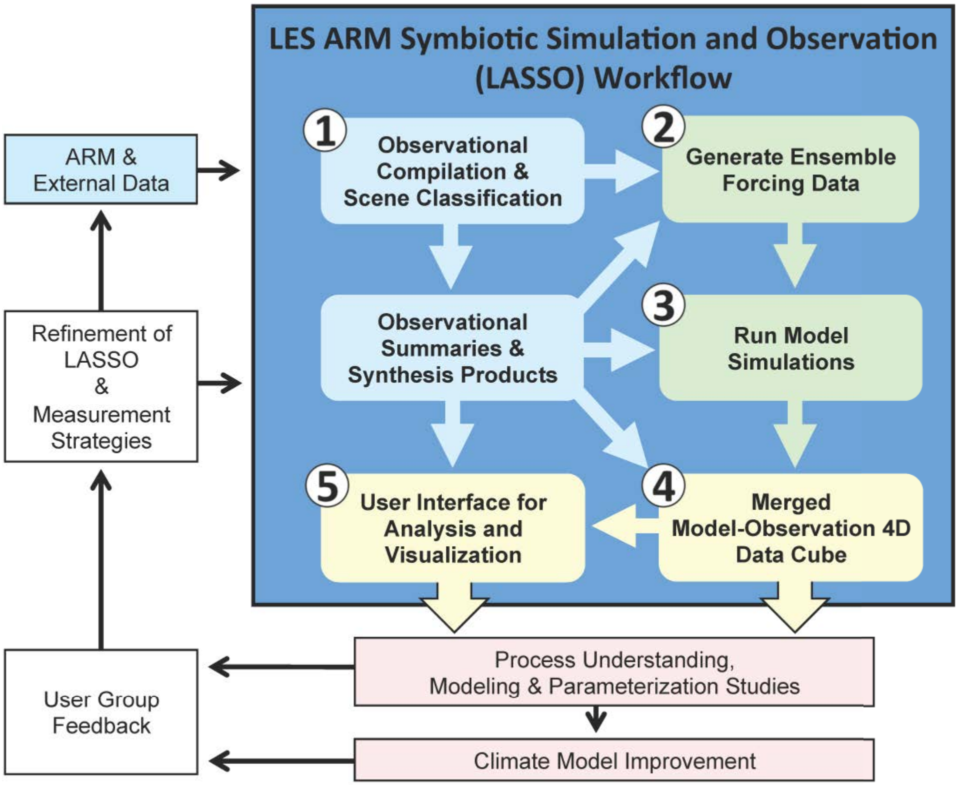

Figure 1. LASSO workflow (blue box) and its context.

LASSO will be designed based on key guiding paradigms to ensure it meets expectations:

- *Emphasize science drivers.* Decisions on workflow implementation, including prioritization and configuration choices, will be based on how well they address the underlying scientific research needs.
- *Tiered implementation approach*. High-priority components will be implemented first to create an initial, basic modeling approach. Then, the workflow will be enhanced over time (including after the two-year initial development project ends), as resources allow, to add features that will increase simulation accuracy, broaden the scientific foci, and increase user interaction features.
- *Flexibility and extensibility.* The workflow will consist of extensible, componentized software that can easily adapt as new observations, forcing data sets, modeling scenarios, and analysis tools become available; allow for different observations, model domains, and model configurations to be used interchangeably; and accommodate anticipated expansion to include a wider range of cloud types, additional sites, and additional scientific goals. This flexibility is vital for supporting the long-term scientific vision of ARM's measurement strategy and CESD's science mission.

- *Robustness and reproducibility*. To facilitate user enhancements to ARM's modeling and support the peer-reviewed research process, our workflow will emphasize transparency and reproducibility, including the data flow, metadata, provenance, code availability, and scripting methodologies.

The last bullet deserves additional context as it involves how users will interact with LASSO. We envision two primary classes of LASSO users. The first user class will be those whose science questions require a data mining approach to access a large number of simulated periods; these scientists will approach the simulations as a statistical representation of observed conditions. The second user class will address science questions requiring additional simulations to complement those from LASSO; in this case, LASSO simulations would serve as the control and users would add sensitivity experiments. Examples include testing a new parameterization to see whether it improves the simulation, or seeking to better understand the effect of entrainment on cloud formation by modifying assumptions in the microphysics related to heterogeneous and homogeneous mixing. To make these additional simulations tractable for such a user, the entire workflow needs to be reproducible so they can supplement what ARM has already done. These users' needs could range from simply making the forcing data sets readily available to providing the LASSO software so they can automate their own modeling workflow.

A critical aspect of this project is communicating workflow recommendations to the CESD community (DOE ASR, ESM, and RGCM Programs) and adjusting the workflow based on their feedback and needs. Community interaction will be essential to ensure the workflow meets the widest range of scientific needs because many decisions will require balancing competing interests while staying within ARM's allotted resources. We are planning for ARM to form an advisory team (possibly a subset of the User Executive Committee) that reflects the broad community base we will serve to facilitate feedback regarding our priorities and recommendations. We will also interact regularly with science translators, observationalists, and other modeling experts to get targeted feedback and input.

The following five sections describe the proposed design process for the five main elements of the LASSO workflow (see numbered boxes in Figure 1): model choice and configuration, forcing data set generation, proposed operating scenarios, data cube generation, and the user interface.

1.2.1 Model Selection and Configuration

Decisions related to model selection and configuration for shallow clouds are drawing on the history of LES modeling for this cloud type [e.g., *Siebesma and Holtslag*, 1996; *Siebesma et al.*, 2003]. Best practices will guide initial choices, which will then be tested and optimized for the SGP location and science drivers (such as moisture transport from the soil through the boundary layer and into the clouds, entrainment effects on the cloud lifecycle, and heterogeneity of land-surface interactions). Unlike traditional case study approaches, ARM needs the best generalized model configuration for a wide range of conditions, rather than a configuration optimized for a particular case.

Two commonly used LES models will be evaluated for inclusion in the LASSO workflow. Each has advantages for certain criteria, as outlined in the high-resolution modeling workshop report *[DOE CESD*, 2014a], and our team has expertise with both models. The models we will evaluate are:

1. The Weather Research and Forecasting (WRF) model [Skamarock et al., 2008] includes a selection of state-of-the-art land-surface, cloud, radiation, and boundary layer parameterizations, as well as aerosol-cloud-radiation interactions [Fast et al., 2006; Grell et al., 2005]. Advantages include its

ability to use nested/open domains in addition to periodic boundaries, and the availability of accompanying, operationally used data assimilation software that we have extended to generate forcing data sets by combining ARM observations with larger scale conditions [Li et al., 2015]. WRF enhancements by team member Endo [Blossey et al., 2013] enable multiple LES-style forcing configurations, and WRF benefits from continual improvements from its wide user community.

2. The System for Atmospheric Modeling (SAM) [Khairoutdinov and Randall, 2003] has been used for many shallow cloud applications [e.g., Ovchinnikov et al., 2014]. The primary advantage of SAM over WRF is its anelastic approach to the equations of motion, which allows for more efficient numerics and lower computational costs. However, it can only use periodic boundaries, limiting the range of forcing conditions that can be accommodated and thus the shallow clouds types that can be simulated.

The decision regarding which model to ultimately use within LASSO will require balancing accuracy, cost, and breadth of capability. At this point, we favor using WRF because of the advantages cited above. In addition, its ability to use time-dependent boundaries will simplify the evolution of LASSO to simulate deep convection at a later date. A critical concern that will be examined, with input from the user community, is whether WRF's broader overall capabilities are worth its higher computational cost.

The best choices for model settings and domain selection depend strongly on the science questions to be addressed and the available computational resources. The domain is one of the more fungible choices with respect to computational cost because tradeoffs exist between grid spacing and horizontal and vertical extent. The shallow cloud focus will permit a smaller domain than would be needed for deep convection, but will require finer grid spacing. We will present a range of implementation options for ARM leadership to consider along with the science questions that will be enabled at each level of cost. Then, we will work with ARM leadership to identify the appropriate configuration for the initial LASSO implementation, with the intent to allow for other options to be considered based on community feedback after the initial implementation is functional. We anticipate the recommended domain will be around 30 km across to take advantage of ARM's scanning cloud radars and nearby locations of the new SGP profiling sites, which will include thermodynamic profiles retrieved at high frequency by the Atmospheric Emitted Radiance Interferometers (AERI) Optimal Estimation (AERIoe) algorithm [*Turner and Löhnert*, 2014].

We will test multiple model configurations to determine the optimal one(s) for reproducing shallow clouds across multiple seasons. We have a starting point for the WRF configuration based on our experience during our WRF Feasibility Study investigating numerical issues raised by *Yamaguchi and Feingold* [2012], which involved testing many shallow cloud cases at SGP. A similar configuration can be mimicked in SAM. Given that configuration testing will occur early in the project and automation through LASSO will not yet be available, we anticipate using manually staged input data combined with weather forecast automation scripts developed by Gustafson [*Fast et al.*, 2012] to simplify conducting a large number of test simulations. For this testing, our primary metrics for evaluation will be the ability to reproduce the observed shallow clouds, focusing particularly on simulated cloud timing, fraction, height, depth, and liquid water path in comparison to observed values. Radiation and turbulent fluxes will also be considered. By using the same forcing for WRF and SAM, we will be able to isolate some of the issues related to bad forcing information when it occurs, as well as identify any consistent biases between the two models or conditions when their solutions tend to diverge.

We will also evaluate two different types of boundary conditions. Days with relatively constant large-scale forcing could be simulated using doubly periodic horizontal boundaries, which both WRF and SAM can accommodate. We will also test time-dependent boundaries with WRF (this feature is not available for SAM) using a nested/open configuration, which will enable the simulation to “feel” changes in large-scale forcings more realistically. This could provide simulations over a wider range of weather conditions, but it would also entail additional computational cost because a larger domain will be necessary to ensure that turbulence and clouds spin up on the inflow boundaries before reaching the region of interest. The use of specified boundaries with LES is a more recent methodology, and we will compare its accuracy with that of the periodic domain in the context of shallow cloud simulations.

Finally, we will perform a cost-benefit analysis (in relation to relevant science questions) for including routine spectral bin microphysics as part of LASSO, which would involve an order of magnitude cost increase over the standard bulk microphysics. While bulk microphysics is routinely used in regional and global models, it has limitations, especially for dynamically vigorous shallow cumuli that exhibit a range of cloud properties [*Kogan*, 2013]. Bin microphysics simulations will be valuable when more detailed microphysical information is critical to the user, such as when they use fine-scale radar simulators. Having both bin and bulk microphysics in LASSO will also enable users to evaluate and further develop the bulk method. Both WRF [*Fan et al.*, 2012] and SAM [*Fan et al.*, 2009] include bin microphysics options that we have used, so implementation and testing will be straightforward.

1.2.2 Forcing Data Set Generation

One of the more challenging decisions for LASSO is identification of the appropriate forcing data sets to use. We anticipate robust debate from the community regarding forcings because their accuracy is critical to simulations, they can be generated from many different methods, and they involve various levels of complexity. We believe that an ensemble of forcings is needed to reflect forcing uncertainty, which could be used to help flag days with suspect forcing, thus assisting users searching for appropriate periods. Without a method to differentiate failed model simulations due to incorrect forcings versus internal model issues, users will be inclined to limit use of LASSO for case study analyses where they could manually evaluate cases to be confident in their behavior. This would defeat the potential value that LASSO could provide to support tasks, such as parameterization development, requiring long-term compilations of reliable simulations to perform robust climatological analyses.

As a starting point, we propose exploring three forcing data set methodologies that reflect a range of complexity and ability to incorporate ARM observations:

1. The ARM continuous forcing data [Xie et al., 2004] is based on a constrained variational-analysis approach [Zhang and Lin, 1997; Zhang et al., 2001] that combines National Weather Service (NWS) analyses with ARM observations. It is a standard ARM value-added product (VAP) data set.
2. Forcings derived from European Centre for Medium-Range Weather Forecasts (ECMWF) forecasts, which incorporate ARM radiosonde data. The model and data assimilation techniques of ECMWF differ from NWS’s and therefore constitute a somewhat independent data source. While previous work has questioned the accuracy of ECMWF-derived forcing for strongly precipitating periods [Xie et al., 2003], it is much better for non-precipitating periods, which would encompass shallow clouds.

3. The Multi-Scale Data Assimilation (MS-DA) methodology developed by Li et al. [2013] is capable of generating forcings down to O(1 km) scale and thus would be ideal for incorporating the range of ARM observations. MS-DA uses the community-based Gridpoint Statistical Interpolation (GSI) software in conjunction with a scale separation algorithm to combine observations representing coarse and fine scales to accurately reflect the atmospheric state. 3D variational-analysis data assimilation (3DVar) is used operationally at NWS and other institutions and is known to be robust. For the initial LASSO testing, we propose using 3DVar-based MS-DA that has already been largely automated by Li. In the second year of the project, we will conduct basic testing of the more advanced hybrid ensemble Kalman (EnKF) filter MS-DA methodology to assess potential improvements in accuracy.

For our initial model configuration tests, we anticipate using the ARM continuous forcing data because it will be the most readily available. Once other forcings become available for a sufficient number of days, we will generate additional model simulations and analyze the relative skill of each methodology compared to the others and identify any consistent biases or differences in variability using the same metrics as for the initial configuration tests. We will be cognizant of issues related to the same data being used for evaluation as was used as input to the models. For our testing, should independent data not be available, we will consider withholding certain data from the inputs to aid evaluation purposes.

1.2.3 Operating Scenario Determination

Decisions regarding near-term operating scenarios for LASSO are dependent on science questions related to process-level understanding of shallow clouds and related parameterization development. By “operating scenario” we refer to the overall method of applying the selected model configurations, such as how frequently to run the model and the use of ensembles.

We anticipate LASSO will produce simulations routinely; that is, it will be run regularly for selected cases but not in real time. This will permit LASSO production to occur on a regular schedule, for example once per month, after the relevant observations have been processed into VAPs. Using automated scene-selection routines (Figure 1, box 1), days during the period will be examined for relevant weather conditions (e.g., the presence of shallow clouds) and the availability of observations, and decisions will be made regarding which days to simulate. We will build upon existing shallow cloud climatology methodologies at SGP [*Berg and Kassianov*, 2008; *Chandra et al.*, 2013; *Zhang and Klein*, 2013] combined with one developed during our WRF Feasibility Study to estimate the typical number of simulations possible each year. In contrast to these conservative climatologies (e.g., removing days impacted by mesoscale and synoptic features [*Zhang and Klein*, 2013]), our criteria will be somewhat broader to test the limits of the LES approach, and therefore we will exceed the 10–20% of May–August days identified in these climatologies. To assist making design decisions and to allow for investigating a broader range of science questions, we will also sub-classify different shallow cloud periods based on conditions impacting configuration choices, e.g., days when shallow clouds later develop into deep convection that would require a high-altitude model top and increased computational cost. Such considerations significantly impact the cost-benefit analyses we will perform for estimating resource requirements.

Once a set of days has been selected, the next task is defining the simulations that will be performed for each period, where we anticipate treating each day as an independent period. We propose running an ensemble of simulations using bulk microphysics plus a single deterministic simulation with bin microphysics to represent forcing and input data uncertainty while providing at least one simulation per selected day with detailed (but computationally expensive) cloud microphysics information. We will develop criteria to select the best set of forcing and initial conditions from the bulk microphysics ensemble members, which will then be used for the single bin microphysics simulation. The choice of appropriate ensemble members is a research question we will address and vet with the community. Our proposed approach is to base the ensemble on perturbing the initial conditions combined with multiple forcing data sets for the boundary conditions. This will balance the uncertainty in each, with the initial conditions mostly impacting the first 6 hours of the simulation and the different forcings mostly impacting subsequent hours [*Vogelmann et al.*, 2015]. We will estimate uncertainty in the thermodynamic profiles used for initial conditions based on temporal and spatial variability among the AERIoe profiles to perturb the initial conditions. The multiple forcing data sets, described in the previous subsection, naturally produce a selection of boundary conditions, which we will supplement with multiple forcings from MS-DA.

Uncertainty in surface conditions is another factor that impacts shallow cloud formation. For the initial workflow testing, we plan to use prescribed surface fluxes from the available measurements. Later in the project, we will use a land-surface model, such as the Community Land Model implemented in WRF [*Lu and Kueppers*, 2012], that can be used to address science questions related to surface heterogeneity and temperature biases. We will test which available land schemes add value at LES scales.

1.2.4 Merging Observations and Simulations into a Data Cube

Ultimately, the value of LASSO for informing process studies, developing improved climate model parameterizations, and increasing scientific understanding will only be realized if the simulation output, corresponding observations, related statistics, estimated uncertainties, and metadata are presented in an accessible and easily used form. Combining the model output and ARM’s wide variety of observations and VAPs into a unified package—the “data cube” concept (Figure 1, box 4)—will significantly enhance the ability of scientists to access and use both the LES simulation output and ARM observational products. For example, ARM observations of surface fluxes, moisture profiles, and cloud liquid water can be combined with model-derived estimates of the horizontal variability of turbulence characteristics (which is difficult to measure) to inform studies of moisture transport from the surface into the clouds. This type of integrated information will be invaluable for answering questions, such as those related to cloud initiation. We also propose that the capability be developed to place observations and simulations on a common spatial grid and sampling interval, where appropriate, to simplify subsequent analyses (the ARM Data Integrator software could be applicable). For example, researchers comparing model output with observations could average the observed vertical resolution of the Raman lidar retrievals of water vapor content and temperature to match the model levels to avoid an inconsistent scale comparison.

Community input is critical to define the data cube contents and ensure its broadest applicability. Therefore, our work for this portion of the LASSO workflow will follow a more iterative, fact-finding and demonstration approach than the model configuration testing described in the previous three subsections. We will manually build prototype data cubes for selected periods to demonstrate candidate contents and vet options with the CESD community (see Section 1.2) to clarify preferences and help us develop recommendations for the data cube content, structure, and user interface. After ARM leadership reviews the recommended options and selects a plan for initial deployment, ARM developers will subsequently code the software needed to automate generation of the cube as part of the regular LASSO workflow.

To ensure the applicability of the resulting data cubes to a broad range of scientific questions, we anticipate that the cubes should include both raw and processed simulation information. For example, LES models traditionally save domain averaged profiles and statistics for meteorological state (e.g., temperature, moisture, winds), turbulence statistics (e.g., first and second order moments), cloud fields (e.g., base, height, water path), and diagnostics (e.g., boundary layer height). In addition to these diagnostic profiles, the data cube should contain full-domain (3D) fields for selected variables (e.g., meteorological state, cloud properties, and radiation fluxes) that users could then use to calculate their own statistics as needed. Saving full-domain fields every time step would imply excessive data storage demands, but we suspect saving a snapshot every 20 minutes would likely be adequate for many uses, such as quantifying the spatial variability of cloud characteristics. Which diagnostics and 3D variables to be saved and their frequencies are topics that we will vet with the CESD community.

In terms of the observational content of the data cube, we will establish with observational scientists feasible recommendations for evaluation data sets and metrics for evaluating the model. The recommendations will be discussed with ARM leadership to form priorities for ARM developers. This process is particularly important because many of the new, valuable retrievals from the megasites are still under development and have not yet been implemented as automated VAPs. A large variety of observations will be important; the general categories particularly relevant to the study of shallow cloud processes include lidar and cloud radar products, such as column vertical velocity retrievals from the surface through the cloud [*Luke and Kollias*, 2013]; turbulent kinetic energy in and below cloud; and 3D cloud fraction, cloud base statistics, and liquid-water contents from the scanning radars.

Complementing the observational and simulated data cube content, we will determine what pre-computed summary statistics, metrics, and “quick-look” plots (pre-plotted fields of commonly used variables) to include. Such characterizations can be used to classify the weather conditions and will be critical for helping users determine whether the LES output and its accuracy during the simulation periods meets their needs. The accuracy of each simulation and of the ensemble spread would be presented within the context of observation uncertainty and natural variability using metrics, such as root mean square error, bias, and Taylor diagrams. These metrics would compare each ensemble member to critical observations and identify the best ensemble member. Critical parameter values would be included as metadata to make the LASSO library searchable to locate simulations of interest; for example, users could search for simulations by cloud type, cloud mass flux magnitude, or model-observation agreement.

1.2.5 Designing Data Cube Interfaces for Analysis and Visualization

The long-term success of ARM’s high-resolution modeling endeavor, as well as of the megasite concept, depends heavily on the accessibility, usability, and scientific relevance of the data that is generated. We therefore will invest a significant portion of the second year of the project designing and working with ARM software developers to implement a state-of-the-art visualization and data-interface capability that is synergistic with existing ARM capabilities. As a general guiding principle for this design, we anticipate that LASSO users will value a robust search functionality that provides the ability to find the simulation periods most useful to them, data access methodologies that provide quick and reliable access to the data cube contents, and a flexible interface allowing them to use the data in various ways, such as compositing, visualizing, performing custom post-simulation diagnostics, and reproducing the LES simulations for sensitivity experiments. Figure 2 presents an example of how we envision users might interact with the data cubes generated by LASSO. From the database of simulations, they will need a methodology to find the periods meeting their criteria, the ability to visualize the conditions on each of those days, and then methods to quickly compare variables across the periods by using a range of analysis techniques that would draw on both observations and simulation output. We will work with the CESD community to identify needs and with ARM developers to devise practical solutions to these requirements, which are the final component of the LASSO workflow (Figure 1, box 5).

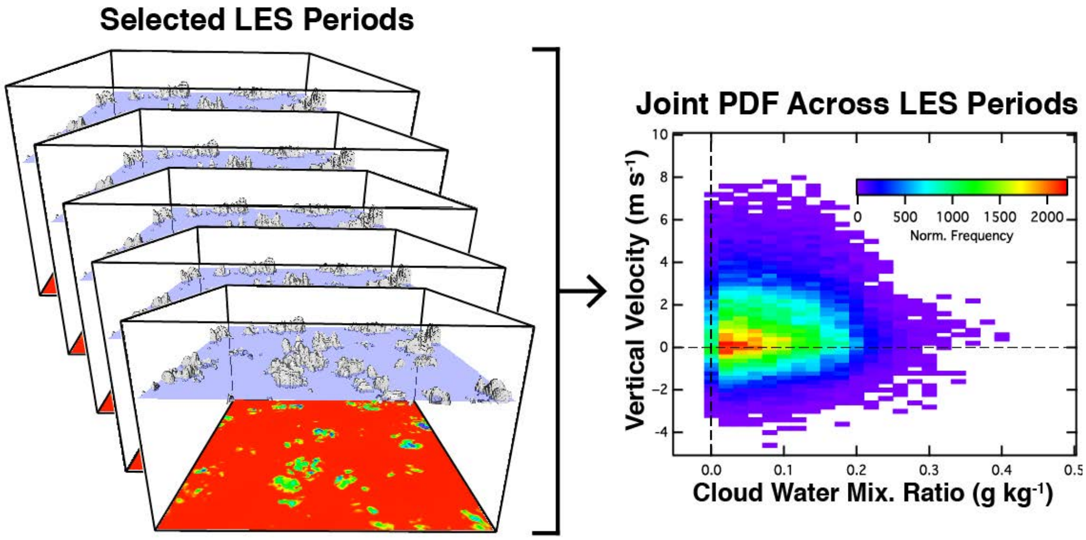

Figure 2. Example visualization and subsequent analysis of user-selected periods based on a given criteria, and a joint probability distribution function generated from the data cube contents. Users would be able to analyze both observations and model output.

The ARM website will play a key role as the first-level user interaction with LASSO. To facilitate data discovery, we will develop recommendations for evolving the ARM Data Discovery Tool to include metadata searches based on model-specific parameters and summary statistics such as weather conditions, cloud fraction statistics, and simulation accuracy metrics. These search capabilities will allow users to quickly find simulations of interest. Similarly, recommendations will be developed for accompanying quick-look visualizations of observations versus model output, which will allow users to easily select the variables of interest to download. We will also work with ARM software developers to investigate options to parse, select, package, regrid, and order observations and simulations from the data cube.

The second level of user interaction with LASSO would be via the ARM Data Visualization Cluster, where we will recommend a prioritized list of capabilities, tool developments, and publicly available packages that would facilitate scientific discovery. We see a number of options for these tools:

- *Free-form interactive user capabilities for analysis and visualization.* Providing computing resources with direct access to the LASSO data cubes for analyses within the same computer network as the archive would enable users to apply their own tools or use available packages. We anticipate that many users would be satisfied with the ability to run their own programs based on Python, Fortran, IDL, Matlab, etc., on the LASSO data cubes without having to pre-order them via the website, significantly quickening their inquiries. The benefits of working on the ARM cluster will be demonstrated using retrievals of entrainment profiles [*Jensen et al.*, 2014] as an example user-contributed data product that would be compared against simulated profiles in the data cubes. This prototype example will help clarify data flow needs and development of protocols to facilitate such interactions. We will also develop recommendations for publicly available analysis packages to include on the cluster, such as the Ultrascale Visualization Climate Data Analysis Tools (UV-CDAT), which could prove useful for parallel access and processing (e.g., for simultaneous analysis of ensemble members). The UV-CDAT is tightly coupled with Python, which will also enable efficient use of other more specialized Python modules such as the Python ARM Radar Toolkit (Py-ART), which is used for analyzing radar data. Over time, we foresee a library of specialized Python modules developed and contributed by users for interacting with LASSO data cubes.

- *Instrument simulators.* We will investigate and provide recommendations on instrument simulators to include, which compute synthetic observations from model output for direct comparison to observed values. For demonstration and concept testing, we plan to use the interactive Radar Doppler Spectra Visualization and Analysis Java Toolkit that team member Luke has developed. The toolkit contains a real-time interactive simulator enabling dynamic exploration of synthetic and observed radar Doppler spectra in tandem, and supports ancillary data from lidars, balloon soundings, microwave radiometers, and irradiance sensors to provide interpretive context. Extending this to simultaneously receive LES output is straightforward and will be a valuable capability for assessing the simulations and their treatment of cloud microphysics. Hosting sophisticated simulators such as this on ARM’s cluster will allow specialists to pre-configure them to use LASSO contents correctly, e.g., ensuring that the radar simulator’s forward model is appropriately coupled with the microphysics assumptions in the LES model. More importantly, hosting simulators on ARM’s cluster will also have the benefit of providing efficient access to the data cube information to reduce latency and speed discovery.
- *Complementary models for parameterization development.* Considerable value would be added by running complementary, computationally efficient models on the cluster for analysis with LASSO data cubes. We will develop and recommend software protocols that would be needed to efficiently access the data cubes for this purpose. They would be similar to the needs for the pre-configured tools, such as the radar simulator example, but would need to be more general to handle a wider range of user data needs because the variables needed would not always be known beforehand. An already developed tool that would be easy for us to demonstrate and use for concept testing on the ARM cluster is the Fast-physics System Testbed and Research (FASTER) testbed software for running single-column model (SCM) simulations using LASSO forcing data. By hosting this tool on the cluster, the data flow for inputs could be optimized and the SCM output easily compared to the data cube. This would provide capabilities similar to the Royal Netherlands Meteorological Institute Parameterization Testbed (KPT) [*Neggers et al.*, 2012]. This capability would be very useful for parameterization developers to quickly test their code. For example, shallow cloud parameterizations proposed for the next generation of climate models, such as Cloud Layers Unified By Binormals (CLUBB) [*Larson et al.*, 2012] and Unified Convection (UNICON) [*Park*, 2014], as well as other new parameterizations, such as the Cumulus Potential (CuP) scheme [*Berg et al.*, 2013], could be quickly compared and validated within this framework over a large number of days, making it much quicker to analyze and improve these parameterizations for a range of meteorological conditions.

The third level of user interaction will be external access to LASSO data cubes from computers outside of ARM’s infrastructure. Many scientists will prefer transferring data cubes to their own computers for analysis. Simplifying external data cube access will allow leveraging other computing infrastructure, such as NERSC, for users to do expensive analyses such as performing a range of LES runs for a multi-year period to compare against the LASSO simulations. In this case, scientists would need to easily acquire the forcings used in LASSO, the data sets needed for comparison, and the LASSO software for replicating the workflow. We will recommend the tool development that would allow automation of this data flow chore. Possible methods include tools built around Globus, due to its efficiency, and the Earth System Grid, which is familiar to the climate modeling community. Issues that will need addressing are ARM data-use accounting and turn-around time for data accessibility, which is impacted by data on tape. To make LASSO more efficient, we will work with ARM staff to develop needed features, seek optimized storage methods for retrieving data from tape, and establish priorities to streamline data access.

1.3 Deliverables and Outcomes

The key deliverables we will provide can be summarized as follows:

1. A prototype workflow with all the components and requirements defined, which will be automated by ARM software developers in coordination with the project principal investigators (PIs). Our team will develop and test the model configuration, domain choice, forcing data sets, and operating scenarios.
2. Sets of simulations consisting of multiple model configurations that will be suitable for implementing observation-model integration software; these will span shallow cloud days over multiple seasons.
3. Extensive recommendations surrounding the many facets of developing a useful workflow including:
    a. Implementation of high-resolution modeling constrained by ARM observations
    b. Construction of the data cubes to combine observations, model output, and metrics
    c. Provision of analysis and visualization tools, including instrument simulators
    d. Efficient user access to LASSO data cubes via ARM infrastructure and externally

The deliverables will provide a roadmap for ARM to follow for implementing LASSO and will lead directly to a functioning modeling workflow for scientific discovery at the end of the project. Our goal is to complete LASSO in stages with an initial prototype of the model input generation and simulation capabilities that can be run manually. It then will be expanded with the assistance of ARM developers to run operationally (Figure 1, boxes 1–3). Recommendations for user interactions will be developed that ARM will implement within the context of its overall priorities (Figure 1, boxes 4–5). Our decisions will be made in coordination with the ARM Technical Director and the ARM Program Manager. Science drivers from the CESD community will guide the direction of the workflow development with practicalities providing bounds for expectations.

To make LASSO operational at the end of the two-year project, many decisions will need to be anticipated early in the project. We expect initial computational cost estimates will need to be made during the first 6–9 months, and recommendations regarding appropriate computing architecture and resources will need to be made before LASSO is fully defined. Using the LES portion of the workflow as an example, based on prior experience, an LES domain 25.6 km across requires approximately 1.6 GB of storage per output snapshot. A full day's simulation with data saved every 20 minutes would require 115 GB per ensemble member. The storage estimates will be combined with the required computational time, the expected number of simulations, and more precise benchmarking tests to recommend how much storage and computing resources will be needed.

Because many of the decisions will involve tradeoffs, we will present our recommendations with a cost-benefit mentality (computational and personnel costs vs. scientific capabilities, ease of use, and future expansion), targeting several levels of complexity and associated cost. For example, the lowest cost could be to implement a single deterministic simulation for single-layered shallow convection. This would address the practical need of starting operations quickly; however, it would only address the smallest number of science applications. A mid-level recommendation could be adding either ensembles or bin microphysics, each option having roughly the same overall computational cost within a factor of 2-3. This would add either the ability to estimate forcing uncertainty, or to have more detailed microphysics information for use in radar simulators and evaluating bulk microphysics simulations. The most expensive option might be performing ensembles with bulk microphysics combined with a single deterministic bin microphysics simulation. Early estimates using WRF (neglecting the cost of forcing data set generation) suggest the options would need between ~13,000 to more than 300,000 core-hours per simulated day (a 23× difference), depending on chosen features, domain extent, and resolution.

Ultimately, this new modeling ability will play a pivotal role in realizing the full potential of ARM observations for advancing CESD's mission that includes developing new process-level understanding and accelerating improvement of climate models. Our efforts described in this implementation strategy will lay the groundwork for enhancements to be implemented in subsequent years: the initial focus on shallow clouds will expand to include all cloud types, land-atmosphere interactions, and aerosols; the methodology developed at SGP will be extended to include additional ARM sites; and powerful new access methods to ARM data will be developed to broaden the user base benefiting from ARM's investments.

2.0 Project Management Plan

2.1 Team Coordination

A core team of scientists is supported as needed by an extensive team of in-house specialists. The core team, their expertise, and roles are summarized below. Key unfunded laboratory support team members are given in Table 1. We will also reach out to relevant ARM translators and observationalists at other institutions. Having ARM's Data Management Facility housed at PNNL will facilitate coordination of the many technical details necessary to successfully have LASSO function well with the ARM computing environment. The following list gives the project's core team, where an "*" indicates those who will be mostly responsible for conducting tasks:

*Gustafson, William (PNNL) — PI and PNNL lab lead. Expertise: Cloud and climate modeling, workflow automation and software development, high-performance computing. Will manage the overall project, direct model configuration and testing efforts, coordinate with ARM managers and developers, disseminate the results, and interact with the observation and modeling communities to vet plans and solicit input.

*Vogelmann, Andrew (BNL) — Co-principal investigator (co-PI) and BNL lab lead. Expertise: Cloud observations and modeling. Will coordinate tasks with PI Gustafson and BNL team members, disseminate results, and interact with the observation and modeling communities to vet plans and solicit input.

Berg, Larry (PNNL) — Expertise: Shallow cloud parameterization and modeling, boundary layer and cloud observations. Will guide processes related to shallow cloud lifecycle, generation of the SGP climatology, and selection of cases and scenarios for testing.

*Endo, Satoshi (BNL) — Expertise: LES cloud modeling, model automation. Will perform tasks associated with WRF LES modeling and interfacing the simulations with observations.

Fast, Jerome (PNNL) — Expertise: Cloud and aerosol modeling and observations. Will guide decisions at multiple junctures within the workflow development, including operational configurations, and selection of cases and scenarios for testing and benching.

*Li, Zhijin (UCLA) — Expertise: Data assimilation. Will interface the MS-DA algorithm with ARM data for use in this project, lead hybrid EnKF tests, and provide MS-DA algorithms.

Luke, Edward (BNL) — Expertise: Radar observations, radar simulators, visualization. Will guide use of retrieved cloud dynamics and microphysics properties for the evaluation of LESs, and will interface his Radar Doppler Spectra Visualization and Analysis Java Toolkit with data cubes for demonstration and assessment.

Ovchinnikov, Mikhail (PNNL) — Expertise: LES cloud modeling, SAM model, and spectral bin microphysics. Will guide tasks associated with configuration and simulations using SAM and spectral bin microphysics.

*Toto, Tami (BNL) — Expertise: ARM data analysis and visualization. Will perform data analysis and information gathering for creation of data sets for model initialization and evaluation, and for demonstrating sample applications using the prototype LASSO interfaces.

*Xiao, Heng (PNNL) — Expertise: LES cloud and climate modeling, and cloud parameterization. Will perform tasks associated with LES modeling and testing.

Software Engineer Consultants (PNNL) — Software engineers from PNNL’s National Security Directorate will provide guidance and options that complement the expertise available from ARM software engineers (see letter of support from Ranata Johnson). This National Security Directorate team has provided invaluable assistance to other computing projects within PNNL’s Atmospheric Sciences and Global Change Division, where their expert opinions and state-of-the-art computing techniques have saved significant time, e.g., through software approaches developed for PNNL’s Platform for Regional Integrated Modeling and Analysis Laboratory Directed Research and Development project.

Table 1. Unfunded support team.

Support Team	Institution	Expertise
Beus, Sherman	PNNL	ARM data flow, software automation, Globus
Cialella, Alice	BNL	Metadata provenance, Manager ARM External Data Center
Comstock, Jennifer	PNNL	Cloud observations, instrument simulators, ARM infrastructure
Feng, Zhe	PNNL	Radar and cloud observations
Giangrande, Scott	BNL	Wind profiler and scanning radar products/analyses
Jensen, Mike	BNL	ARM Lead Translator, Cloud Lifecycle Working Group Translator
Johnson, Karen	BNL	Cloud radar data mentor and developer
Kollias, Pavlos	McGill/BNL	ARM Radar Science Group Leader
Lin, Wuyin	BNL	ECMWF forcing generation, FASTER SCM Testbed developer
Liu, Yangang	BNL	Multi-scale model methodologies
Newsom, Rob	PNNL	Lidar observations, ARM mentor
Riihimaki, Laura	PNNL	Cloud and meteorological observations, ARM translator
Sivaraman, Chitra	PNNL	ARM Data Management Facility Value-Added Products Manager

Gustafson will manage the overall project ensuring deadlines are met and deliverables are of high quality, and he will also serve as the PNNL laboratory lead. Vogelmann is the Co-PI who will serve as the BNL laboratory lead, and Li will be the UCLA lead. Gustafson will be responsible for the overall assignments in coordination with Vogelmann and Li. Regular teleconferences will be used to address necessary business, coordinate work, and work through relevant issues.

Project tasks are organized by theme and their fit within the two-year deadline. Careful thought has gone into determining which aspects of the project can advance independently of other parts. For example, the final domain configurations are not needed to develop large-scale forcings, so team members can work these areas simultaneously. Task leads will be assigned based on their expertise and in consideration of ensuring deliverables will be available on time. A laboratory will take lead on a given task, to optimize production by facilitating face-to-face interactions on a day-to-day level, with at least one participant from the other laboratory for seamless data exchange among the institutions. The technical nature of much of the work will require more frequent interactions among team members to ensure compatibility between software developed by different individuals, exchange technical details related to data products, and work through model configuration decisions. Technical discussions will be held via phone and e-mail exchanges with the frequency determined by the stage of the project. A Wiki will be set up to share project materials among team members and select stakeholders, serving as a record of progress towards project objectives, and a repository for documentation, ultimately used in the final report of recommendations.

Some aspects of LASSO development will be more efficient in person. We have budgeted travel for relevant team members to meet both with each other and with ARM management and staff. This will be particularly important during the formation stages of the project when protocols are developed that will guide the project. Working through decisions with ARM translators, developers, and instrument mentors will also benefit from in-face meetings. We anticipate facilitating one-on-one discussions by leveraging travel to meetings that are attended by numerous team members. However, many of these meetings do not allow sufficient time for significant offline discussions, so we anticipate using an extra day on one side of meetings to permit extended discussions. Team members leading specific task elements will develop ongoing interactions with relevant ARM staff, e.g., with the support of the ARM Technical Director.

Feedback from the community of LASSO users will play a large role in this project. We plan to host forums at relevant meetings, where we can gain input to inform our decisions and make presentations to promote the new capabilities. In addition to the ASR PI and Working Group meetings, we anticipate presenting at the WRF Users Workshop, the Annual Community Earth System Model Workshop, and will seek to hold a Town Hall at the American Geophysical Union (AGU) Fall Meeting in the second year of the project to communicate the capabilities to a wide audience. We will gain valuable feedback by recruiting beta users to test the data cube products, which will initially be based on our manually created cubes, and ultimately on automatically generated cubes once the software is available for this capability. Also, an advisory team will give credence to our recommendations and to help clarify priorities when we receive contradictory feedback.

2.2 Data Management

We will adhere to the protocols used by ARM in their normal distribution of data. The project’s goal is to provide the expertise needed to develop a fully reproducible workflow, so we will recommend that ARM make the LASSO software developed available to the community via a public code repository, such as Github, including the software to automate LASSO, the LES code, and code used to transform observations for use in LASSO. Li agrees to make his MS-DA data assimilation code available to ARM and the community through this project. Software we develop within this project will be released with a copyright in coordination with ARM’s policies, with our preference of an open source copyright. WRF is a community model, and changes we make to it will most likely be of use to others. In addition to making our version accessible via a code repository, we will forward useful modifications to NCAR for inclusion in subsequent WRF releases. Likewise, we will contribute relevant modifications to SAM.

Data used to generate the simulations will come primarily through the ARM Data Archive and, therefore, will be accessible to the public. Some developmental observation products, such as the prototype data cubes, may not be archived by ARM due to their preliminary nature, and, therefore, will be made available upon request. We will consult with the translator team for the most appropriate way to share these developmental products from within ARM. Once operational, LASSO output will ultimately be stored within the ARM Data Archive and be publically available. Our test simulations used to validate the final model configurations will be archived within the ARM’s archive, as appropriate.

Data formats within the project will primarily be netCDF, and will typically follow either the Climate and Forecast convention or the ARM convention within the netCDF files. Metadata will play a central role in our data cube development. Developing adequate metadata and provenance requirements to support data cube use and retrieval will be a deliverable from this project. We plan to make use of the ARM Data Product Registration Tool (OME) for documenting the key metadata for data generated within the project. Full functionality of the LASSO data cubes, as we envision them, will require minor modifications to OME to include data value ranges and other parameter values, as discussed in the narrative. Such parameters include model-specific parameters and summary statistics such as weather conditions, cloud fraction statistics, and simulation accuracy metrics. These would then be used in conjunction with the ARM Data Discovery Tool for users to select periods of interest.

3.0 Deliverables

The deliverables from this project, identified in Table 2, derive from the objectives in the DOE white paper call submitted in January 2014 and accepted in May 2015. Those white paper deliverables are articulated in Table 3. We have established tentative timelines to produce these deliverables based on the anticipated workflow. A natural progression occurs throughout the project, with the pieces of LASSO building as work progresses, while, at the same time, there is the sufficient ability to overlap different aspects of the work to concurrently engage multiple team members and achieve a faster throughput.

We expect to work closely with ARM staff outside the project during the two-year period. Achieving an automated workflow will require close coordination between these staff and us, taking into consideration their availability and other priorities. Therefore, we present the milestone and task timelines as an example of what will be needed to reach an operational state in two years. When the project starts, we will need to reevaluate the timeline and specific tasks with input from the ARM leadership.

Table 2. Deliverables. The “objective(s) addressed” column refers to the DOE white paper call elements.

Number	Objective(s) Addressed	Deliverables
D1	1	Presentations to the community and feedback from them regarding recommendations
D2	2	Recommendations for model to use for ongoing routine simulations
D3	4	Recommended whether to use bulk or bulk+bin microphysics for ongoing simulations
D4	5	Recommendation whether to use bulk or bulk+bin microphysics for ongoing simulations
D5	3	Identify pros and cons of different forcing data set methodologies and recommend method(s) to sue for ongoing simulations
D6	5	Recommend operation scenario(s) for simulations, including whether and how to include ensembles
D7	6	Prioritized list of model variables to be saved and the format (e.g., 3D, statistical summaries)
D8	11	Prioritized list of feasible observations for evaluation data sets and metrics
D9	6,11	Recommended format for data cube and methods to link observations and model output

Number	Objective(s) Addressed	Deliverables
F10	8	Prototype LASSO workflow to be automated by ARM with each necessary component defined
D11	4,5,6	Sets of simulations for tested model configuration spanning shallow clouds over multiple seasons
D12	12,13	Recommendations for website interface to search and access LASSO data cubes
D13	12,13	Recommendations for software tools and development needed to access LASSO data cubes from ARM analysis cluster
D14	12,13	Recommendations for visualization tools, instrument simulators, and computationally efficient models in the analysis framework
D15	13	Recommendations regarding external LASSO data distribution (e.g., Globus, ESG, and related portals)
D16	9,10	Estimated computational and storage costs to operate the workflow
D17	7	Recommended computing infrastructure for ongoing LASSO operation
D18	All	Identified high-priority research areas for follow-on work to improve simulations and workflow
D19	1	Publication documenting LASSO workflow
D20	All	Final report

Table 3. Sections addressing objectives in original white paper call.

Objective	White Paper Section
1) Actively engage with the Climate and Environmental Sciences Division (CESD) science community to ensure the model framework is suitable to address core goals of CESD	1.2 2.1
2) Select a model (or possibly a set of models) taking into account factors such as computational efficiency and flexibility for evaluating diverse physical processes and parameterizations	1.2.1
3) Identify model forcing methodologies including pros and cons for options presented	1.2.2
4) Recommend appropriate model domain(s) for the Southern Great Plains. These are likely to differ by application	1.2.1
5) Recommend model operation scenarios (e.g. frequency of model simulations, use of ensembles, etc.)	1.2.3

Objective	White Paper Section
6) Indicate what model output should be saved. If full 3D output is to be saved, what parameters should be saved at what time interval? If statistical summaries, what is the nature of that statistical output?	1.2.4
7) Identify what computational infrastructure would be appropriate to carry out routine operations	1.3
8) Develop a workflow that includes all phases of the model operation including creation of forcing data sets necessary for model runs, operation of the model, and output of the appropriate simulation data	1.2
9) Perform benchmark tests of the model workflow on the target machine—or a comparable proxy—to provide estimates for the cost of routine operational scenarios	1.3
10) Assess the computational requirements in terms of central processing unit (CPU) and storage for routine operation scenarios	1.3
11) Work with observational scientists to establish feasible recommendations for evaluation data sets and metrics for evaluating the model, including both parameters and format	1.2 1.2.4
12) Develop recommendations for analysis tools including tools for visualizing model output and co-analysis of model output and observational data	1.2.5
13) Develop recommendations for the analysis framework including the need for remote work space vs. data delivery requirements	1.2.5

4.0 References

Adegoke, J O, R Pielke, and AM Carleton. 2007. "Observational and modeling studies of the impacts of agriculture-related land use change on planetary boundary layer processes in the central US," *Agricultural and Forest Meteorology*, *142*, 203–215, doi:10.1016/j.agrformet.2006.07.013.

Berg, L K, and EI Kassianov. 2008. "Temporal variability of fair-weather cumulus statistics at the ACRF SGP site," *Journal of Climate*, *21*, 3344–3358, doi:10.1175/2007jcli2266.1.

Berg, L K, EI Kassianov, CN Long, and DL Mills. 2011. "Surface summertime radiative forcing by shallow cumuli at the Atmospheric Radiation Measurement southern great plains site," *Journal of Geophysical Research*, *116*, D01202, doi:10.1029/2010jd014593.

Berg, LK, WI Gustafson Jr, EI Kassianov, and L Deng. 2013. "Evaluation of a modified scheme for shallow convection: Implementation of CuP and case studies," *Monthly Weather Review*, *141*, 134–147, doi:10.1175/mwr-d-12-00136.1.

Blossey, PN, CS Bretherton, MH Zhang, AN Cheng, S Endo, T Heus, YG Liu, AP Lock, SR de Roode, and KM Xu. 2013. "Marine low cloud sensitivity to an idealized climate change: The CGILS LES intercomparison," *Journal of Advances in Modeling Earth Systems*, *5*, 234–258, doi:10.1002/jame.20025.

Chandra, AS, P Kollias, and BA Albrecht. 2013. "Multiyear summertime observations of daytime fair-weather cumuli at the ARM Southern Great Plains Facility," *Journal of Climate*, *26*, 10031–10050, doi:10.1175/jcli-d-12-00223.1.

DOE ARM Climate Research Facility. 2013. *ARM Climate Reserach Facility Management Plan*, 26 pp, DOE/SC-ARM-13-022, http://www.arm.gov/publications/programdocs/doe-sc-arm-13-022.pdf?id=3, accessed 3-Nov-2014.

DOE ARM Climate Research Facility. 2014. *Atmospheric Radiation Measurement Climate Research Facility Decadal Vision*, 27 pp, DOE/SC-ARM-14-029.

DOE Climate and Environmental Sciences Division. 2014a. *Atmospheric Radiation Measurement Climate Research Facility - Atmospheric System Research High-Resolution Modeling Workshop,* 47 pp, DOE/SC-0169, U.S. Department of Energy, http://science.energy.gov/~/media/ber/pdf/workshop%20reports/doe-sc-0169-low-resolution.pdf?id=34, accessed 30-Dec-2014.

DOE Climate and Environmental Sciences Division. 2014b. *Atmospheric Testbed Workshop*, 42 pp, DOE/SC-0163, U.S. Department of Energy, http://science.energy.gov/~/media/ber/pdf/Brochures/CESD_Atmospheric_Testbed_Workshop_V9_LOW.pdf?id=97, accessed 30-Dec-2014.

Fan, JW, M Ovtchinnikov, JM Comstock, SA McFarlane, and A Khain. 2009. "Ice formation in Arctic mixed-phase clouds: Insights from a 3D cloud-resolving model with size-resolved aerosol and cloud microphysics," *Journal of Geophysical Research*, *114*, D04205, doi:10.1029/2008jd010782.

Fan, JW, LR Leung, ZQ Li, H Morrison, HB Chen, YQ Zhou, Y Qian, and Y Wang. 2012. "Aerosol impacts on clouds and precipitation in eastern China: Results from bin and bulk microphysics," *Journal of Geophysical Research*, *117*, D00K36, doi:10.1029/2011jd016537.

Fast, JD, WI Gustafson, RC Easter, RA Zaveri, JC Barnard, EG Chapman, GA Grell, and SE Peckham. 2006. "Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model," *Journal of Geophysical Research*, *111*, D21305, doi:10.1029/2005JD006721.

Fast, JD, WI Gustafson, LK Berg, WJ Shaw, M Pekour, M Shrivastava, JC Barnard, RA Ferrare, CA Hostetler, JA Hair, M Erickson, BT Jobson, B Flowers, MK Dubey, S Springston, RB Pierce, L Dolislager, J Pederson, and RA Zaveri. 2012. "Transport and mixing patterns over central California during the Carbonaceous Aerosol and Radiative Effects Study (CARES)," *Atmospheric Chemistry and Physics*, *12*, 1759–1783, doi:10.5194/acp-12-1759-2012.

Grell, GA, SE Peckham, R Schmitz, SA McKeen, G Frost, WC Skamarock, and B Eder. 2005. "Fully coupled 'online' chemistry within the WRF model," *Atmospheric Environment*, *39*, 6957–6975, doi:10.1016/j.atmosenv.2005.04.027.

Guichard, F, JC Petch, JL Redelsperger, P Bechtold, JP Chaboureau, S Cheinet, W Grabowski, H Grenier, CG Jones, M Köhler, JM Piriou, R Tailleux, and M Tomasini. 2004. "Modelling the diurnal cycle of deep precipitating convection over land with cloud-resolving models and single-column models," *Quarterly Journal of the Royal Meteorological Society*, *130*, 3139–3172, doi:10.1256/qj.03.145.

Jensen, MP, T Toto, P Kollias, C Lu, and T Wagner. 2014. "Entrainment rates in non-precipitating cumulus from cloud radar observations, presented at *2014 DOE Atmospheric System Research Principal Investigator Meeting*," Potomoc, MD, 13 March 2014, http://asr.science.energy.gov/meetings/stm/2014/presentations/ASR2014-Jensen-Entrainment.pdf.

Khairoutdinov, MF, and DA Randall. 2003. "Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities," *Journal of the Atmospheric Sciences*, *60*, 607–625, doi:10.1175/1520-0469(2003)060<0607:Crmota>2.0.Co;2.

Klein, SA, XN Jiang, J Boyle, S Malyshev, and SC Xie. 2006. "Diagnosis of the summertime warm and dry bias over the U.S. Southern great plains in the GFDL climate model using a weather forecasting approach," *Geophysical Research Letters*, *33*, L18805, doi:10.1029/2006gl027567.

Kogan, Y. 2013. "A cumulus cloud microphysics parameterization for cloud-resolving models," *Journal of the Atmospheric Sciences*, *70*, 1423–1436, doi:10.1175/jas-d-12-0183.1.

Larson, VE, DP Schanen, MH Wang, M Ovchinnikov, and S Ghan. 2012. "PDF parameterization of boundary layer clouds in models with horizontal grid spacings from 2 to 16 km," *Monthly Weather Review*, *140*, 285–306, doi:10.1175/mwr-d-10-05059.1.

Li, Z, Y Chao, JD Farrara, and JC McWilliams. 2013. "Impacts of distinct observations during the 2009 Prince William Sound field experiment: A data assimilation study," *Continental Shelf Research*, *63*, S209–S222, doi:10.1016/j.csr.2012.06.018.

Li, Z, S Feng, Y Liu, W Lin, M Zhang, T Toto, AM Vogelmann, and S Endo. 2015. "Development of fine-resolution analyses and expanded large-scale forcing properties. Part I: Methodology and evaluation," *Journal of Geophysical Research*, in press, doi:10.1002/2014JD022245.

Lu, YQ, and LM Kueppers. 2012. "Surface energy partitioning over four dominant vegetation types across the United States in a coupled regional climate model (Weather Research and Forecasting Model 3-Community Land Model 3.5)," *Journal of Geophysical Research*, *117*, doi:10.1029/2011jd016991.

Luke, EP, and P Kollias. 2013. "Seamless measurements of vertical air velocity associated with warm clouds through the integration of multiple retrieval techniques," presented at *American Meteorological Society 36th Conference on Radar Meteorology*, Brekenridge, CO, 19-Sep-2013, https://ams.confex.com/ams/36Radar/webprogram/Paper229088.html.

Ma, HY, S Xie, SA Klein, KD Williams, JS Boyle, S Bony, H Douville, S Fermepin, B Medeiros, S Tyteca, M Watanabe, and D Williamson. 2014. "On the correspondence between mean forecast errors and climate errors in CMIP5 models," *Journal of Climate*, *27*, 1781–1798, doi:10.1175/jcli-d-13-00474.1.

Neggers, RAJ, AP Siebesma, and T Heus. 2012. "Continuous single-column model evaluation at a permanent meteorological supersite," *Bulletin of the American Meteorlogical Society*, *93*, 1389–1400, doi:10.1175/bams-d-11-00162.1.

Ovchinnikov, M, AS Ackerman, A Avramov, AN Cheng, JW Fan, AM Fridlind, S Ghan, J Harrington, C Hoose, A Korolev, GM McFarquhar, H Morrison, M Paukert, J Savre, BJ Shipway, MD Shupe, A Solomon, and K Sulia. 2014. "Intercomparison of large-eddy simulations of Arctic mixed-phase clouds: Importance of ice size distribution assumptions," *Journal of Advances in Modeling Earth Systems*, *6*, 223–248, doi:10.1002/2013ms000282.

Park, S. 2014. "A unified convection scheme (UNICON). Part I: Formulation," *Journal of Atmospheric Sciences*, *71*, 3902–3930, doi:10.1175/jas-d-13-0233.1.

Qian, Y, MY Huang, B Yang, and LK Berg. 2013. "A modeling study of irrigation effects on surface fluxes and land-air-cloud interactions in the southern great plains," *Journal of Hydrometeorology*, *14*, 700–721, doi:10.1175/jhm-d-12-0134.1.

Siebesma, AP, and AAM Holtslag. 1996. "Model impacts of entrainment and detrainment rates in shallow cumulus convection," *Journal of Atmosphereic Sciences*, *53*, 2354–2364. doi:10.1175/1520-0469(1996)053<2354:MIOEAD>2.0.CO;2.

Siebesma, AP, CS Bretherton, A Brown, A Chlond, J Cuxart, PG Duynkerke, HL Jiang, M Khairoutdinov, D Lewellen, CH Moeng, E Sanchez, B Stevens, and DE Stevens. 2003. "A large eddy simulation intercomparison study of shallow cumulus convection," *Journal of Atmospheric Sciences,* 60, 1201–1219, doi:10.1175/1520-0469(2003)60<1201:alesis>2.0.co;2.

Skamarock, WC, JB Klemp, J Dudhia, DO Gill, DM Barker, MG Duda, W Wang, and JG Powers. 2008. "A description of the advanced research WRF version 3," 113 pp, NCAR Technical Note, NCAR/TN-475+STR, National Center for Atmospheric Research, http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf.

Turner, DD, and U Löhnert. 2014. "Information content and uncertainties in thermodynamic profiles and liquid cloud properties retrieved from the ground-based atmospheric emitted radiance interferometer (AERI)," *Journal of Applied Meteorology and Climatology*, 53, 752–771, doi:10.1175/JAMC-D-13-0126.1.

Vogelmann, AM, AM Fridlind, T Toto, S Endo, W Lin, J Wang, S Feng, Y Zhang, DD Turner, Y Liu, Z Li, S Xie, AS Ackerman, M Zhang, and M Khairoutdinov. 2015. "RACORO continental boundary layer cloud investigations. Part I: Case study development and ensemble large-scale forcings," *Journal of Geophysical Research*, 120, 5962-5992, doi:10.1002/2014JD022713.

Xie, SC, RT Cederwall, MH Zhang, and JJ Yio. 2003. "Comparison of SCM and CSRM forcing data derived from the ECMWF model and from objective analysis at the ARM SGP site," *Journal of Geophysical Research*, 108, 4499, doi:10.1029/2003jd003541.

Xie, SC, RT Cederwall, and MH Zhang. 2004. "Developing long-term single-column model/cloud system-resolving model forcing data using numerical weather prediction products constrained by surface and top of the atmosphere observations," *Journal of Geophysical Research,* 109, D01104, doi:10.1029/2003jd004045.

Yamaguchi, T, and G Feingold. 2012. "Technical note: Large-eddy simulation of cloudy boundary layer with the advanced research WRF model," *Journal of Advances in Modeling Earth Systems*, 4, M09003, doi:10.1029/2012ms000164.

Zhang, MH and JL Lin. 1997. "Constrained variational analysis of sounding data based on column-integrated budgets of mass, heat, moisture, and momentum: Approach and application to ARM measurements," *Journal of the Atmospheric Sciences*, 54, 1503–1524, doi:10.1175/1520-0469(1997)054<1503:Cvaosd>2.0.Co;2.

Zhang, MH, JL Lin, RT Cederwall, JJ Yio, and SC Xie. 2001. "Objective Analysis of ARM IOP data: Method and Sensitivity," *Monthly Weather Review*, 129, 295–311, doi:10.1175/1520-0493(2001)129<0295:Oaoaid>2.0.Co;2.

Zhang, YY, and SA Klein. 2013. "Factors controlling the vertical extent of fair-weather shallow cumulus clouds over land: Investigation of diurnal-cycle observations collected at the ARM Southern Great Plains site," *Journal of the Atmospheric Sciences*, 70, 1297–1315, doi:10.1175/jas-d-12-0131.1.

ARM

CLIMATE RESEARCH FACILITY

www.arm.gov

U.S. DEPARTMENT OF ENERGY

Office of Science