U.S. DEPARTMENT OF ENERGY | Office of Science

DOE/SC-ARM-17-031

Recommendations for Implementation of the LASSO Workflow

WI Gustafson Jr.
X Cheng
B Krishna
T Toto
AM Vogelmann
S Endo
Z Li
H Xiao

November 2017

ARM CLIMATE RESEARCH FACILITY

DISCLAIMER

DOE/SC-ARM-17-031

Recommendations for Implementation of the LASSO Workflow

doi:10.2172/1406259

WI Gustafson Jr, Pacific Northwest National Laboratory
Principal Investigator

AM Vogelmann, Brookhaven National Laboratory
Co-Principal Investigator

X Cheng, National University of Defense Technology, China
S Endo, Brookhaven National Laboratory
B Krishna, Oak Ridge National Laboratory
Z Li, University of California, Los Angeles, and Jet Propulsion Laboratory
T Toto, Brookhaven National Laboratory
H Xiao, Pacific Northwest National Laboratory
Co-Investigators

November 2017

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

Executive Summary

The U. S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Climate Research Facility began a pilot project in May 2015 to design a routine, high-resolution modeling capability to complement ARM's extensive suite of measurements. This modeling capability, envisioned in the *ARM Decadal Vision* (U.S. Department of Energy 2014), subsequently has been named the Large-Eddy Simulation (LES) ARM Symbiotic Simulation and Observation (LASSO) project, with an initial focus of shallow convection at the ARM Southern Great Plains (SGP) atmospheric observatory. This report documents the recommendations resulting from the pilot project to be considered by ARM for implementation into routine operations. These recommendations are based on the findings of the pilot study with feedback from the ARM Atmospheric Modeling Advisory Group and from the broader research community, particularly beta users and members of the LASSO email list.

The goal of the LASSO pilot project has been to design a routine, high-resolution modeling capability to assist scientists in bridging the gap from point measurements to larger scales relevant to common atmospheric modeling. This is done by combining measurements with LES modeling to provide a self-consistent representation of the atmosphere around the ARM site that can be used to better understand the measurements and the processes leading to the observed conditions. The LES must be interpreted within its representativeness, which is a statistical representation of the atmospheric state around SGP. The availability of LES simulations with concurrent observations serves many purposes.

- LES helps bridge the scale gap between ARM observations and coarse atmospheric models.
- The use of routine LES adds value to observations by providing a self-consistent representation of the atmosphere and a dynamical context for the observations.
- LES provides representations of unobservable processes and properties.

A key feature of LASSO is the generation of a simulation library for researchers that enables statistical approaches beyond a single-case mentality. Such a statistical library of cases is essential to addressing atmospheric variability for process-level understanding, as well as parameterization evaluation and improvement.

LASSO is designed to address a versatile range of science applications, and it has been envisioned that this framework could be used by, but not limited to, three primary user categories characterized as observationalists, theoreticians, and modelers. Observationalists may not have the expertise to generate their own model simulations and will be able to take advantage of the readily available LASSO simulations to act as synthetic, known data to improve the understanding of their retrievals and to help improve their design. Similarly, theoreticians can benefit from having readily available, vetted simulations to test physical relationships as well as obtain values that are either difficult or impossible to measure, such as the spatial variability of fluxes and their co-variabilities. Modelers can benefit from LASSO by having readily available control simulations combined with forcing data sets that are already validated against ARM observations to determine which days have valid forcings that they can use for further modeling studies. The modelers will also have the tools from LASSO to quickly compare their model simulations with the observations using the processed observations and skill scores developed for LASSO. This will simplify parameterization development for the boundary layer and clouds. Future

LASSO enhancements could also target the contributing role of land-atmosphere interactions through the incorporation of soil models and interactive feedbacks between the land and atmosphere.

The pilot project has provided prototype software and a set of recommendations for an operational system capable of meeting ARM's goal of simulating shallow convection at the SGP. This includes details such as model selection and configuration, criteria for days to simulate, observations to include, skill scores and an interactive web interface for evaluating the simulations, and estimates of computational cost.

The initial LASSO operations are envisioned to routinely produce ensembles of LES at the SGP observatory for roughly 30 case dates per year, based on a climatological analysis of days exhibiting shallow convection. We recommend using the Weather Research and Forecasting model for LASSO with a combination of forcing data sets to drive the LES model. The LES domain should use 100 m horizontal grid spacing, have a horizontal domain extent of 25 km, vertical grid spacing of 30 m near the surface, a model top near the tropopause, and use doubly periodic lateral boundary conditions. The LES output will be bundled with a selection of coincident observations, and simulation performance skill scores and diagnostics, all of which will be made available to users between three and six months after the calendar date of cloud occurrence. The forcings should represent a range of spatial scales and input data sources to account for the large uncertainty in forcing. At the SGP, appropriate forcings include the Variational Analysis value-added product produced by ARM, forcing derived from the European Centre for Medium-Range Weather Forecast Integrated Forecast System, and a high-resolution data assimilation methodology called Multiscale Data Assimilation that directly incorporates ARM measurements.

Overall, this report contains 28 specific recommendations broken into categories of modeling, data bundles, operations, and future development. Each of the recommendations is described with accompanying reasoning and descriptions of the intended meaning of the recommendation, as appropriate. Taken as a whole, the recommendations lay out the important details necessary for implementing LASSO into a formal ARM datastream that would be run operationally. They also form a foundation for what LASSO can do for shallow convection, which can later be expanded beyond shallow convection at the ARM Facility's SGP site to other phenomena or ARM Facility sites.

Acknowledgments

The Large-Eddy Simulation (LES) Atmospheric Radiation Measurement (ARM) Symbiotic Simulation and Observation (LASSO) Pilot Project team consists of William I. Gustafson Jr. (principal investigator, PI), Andrew M. Vogelmann (co-principal investigator, Co-PI), Xiaoping Cheng, Satoshi Endo, Zhijin Li, Tami Toto, and Heng Xiao. The LASSO Bundle Browser has been developed by Bhargavi Krishna. Funding was provided by the U.S. Department of Energy Office of Science Biological and Environmental Research via the ARM Climate Research Facility. We acknowledge the advice from external members of the ARM Atmospheric Modeling Advisory Group: Maike Ahlgrimm (European Centre for Medium-Range Weather Forecasts), Chris Bretherton (University of Washington), Graham Feingold (National Oceanic and Atmospheric Administration Earth System Research Laboratory, NOAA ESRL), Chris Golaz (Lawrence Livermore National Laboratory), David Turner (NOAA ESRL), Minghua Zhang (Stony Brook University), and Jim Mather (ARM Technical Director).

We acknowledge numerous members of the ARM infrastructure team for their coordination, installation, maintenance, processing, and advice given on the data products used and for LASSO product management as listed in Section 3. Portions of the work were performed at 1) the Pacific Northwest National Laboratory (PNNL)—Battelle Memorial Institute operates PNNL under contract DEAC05-76RL01830, 2) Brookhaven National Laboratory, and 3) the Jet Propulsion Laboratory and the University of California, Los Angeles, via a subcontract through PNNL. Computational resources have been provided by 1) the ARM Data Center Computing Facility, 2) the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725; 3) the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231; and 4) PNNL Institutional Computing.

Recommended Citation

Gustafson, WI, AM Vogelmann, X Cheng, S Endo, B Krishna, Z Li, T Toto, and H Xiao. 2017. Recommendations for Implementation of the LASSO Workflow. Ed. by R. Stafford, DOE Atmospheric Radiation Measurement Climate Research Facility. DOE/SC-ARM-17-031, doi:10.2172/1406259.

Acronyms and Abbreviations

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
3DVar	three-dimensional variational (analysis data assimilation)
4DVar	four-dimensional variational (analysis data assimilation)
ADC	ARM Data Center
ADI	ARM Data Integrator
AERI	Atmospheric Emitted Radiance Interferometer
AERIoe	Optimal Estimation AERI (value-added product)
ARM	Atmospheric Radiation Measurement Climate Research Facility
ARSCL	Active Remote Sensing of Clouds
ASCII	American standard code for information interchange
ASR	Atmospheric System Research (Program)
BF	boundary facilities
BNL	Brookhaven National Laboratory
CF	Central Facility
CF	Climate and Forecast (metadata convention)
CFTSI	TSI-based cloud fraction
CLUBB	Cloud Layers Unified by Binormals (parameterization)
CMDV	Climate Model Development and Validation
CMDV-CM4	CMDV Coupling Mechanistically the Convective Motions and Cloud Macrophysics in a Climate Model (project)
Co-PI	co-principal investigator
CR-SIM	CRM Radar Simulator
CRM	cloud resolving model
DDH	Diagnostics in the Horizontal Domains system
DOE	U.S. Department of Energy
E3SM	Energy Exascale Earth System Model
EBBR	Energy Balance Bowen Ratio (system)
ECMWF	European Centre for Medium-Range Weather Forecasts
ECOR	Eddy Correlation Flux Measurement (system)
EF	extended facilities
EPS	encapsulated postscript
ESRL	Earth System Research Laboratory
FASTER	DOE FAst-physics System TEstbed & Research (project)

FNL	Final Operational Global Analysis from the Global Data Assimilation System
GB	gigabyte
GOES	Geostationary Operational Environmental Satellite
GSI	Gridpoint Statistical Interpolation (system)
h	hour
HRRR	High-Resolution Rapid Refresh
hybrid-EnKF	hybrid ensemble-Kalman filter
I/O	input/output
IDL	Integrated Data Language
IF	intermediate facilities
IFS	Integrated Forecast System
KAZRARSCL	ARSCL from the Ka-band ARM zenith-pointing radar
km	kilometer
LASSO	LES ARM Symbiotic Simulation and Observation (project)
LCL	lifting condensation level
LES	large-eddy simulation
LST	local standard time
LWP	liquid water path
m	meter
MB	megabyte
MERRA	Modern-Era Retrospective Reanalysis for Research and Applications
MSDA	multiscale data assimilation
MWR	microwave radiometer
MWRRet	microwave radiometer retrieval (value-added product)
NCAR	National Center for Atmospheric Research
NERSC	National Energy Research Scientific Computing Facility
NEXRAD	Next-Generation Weather Radar
NOAA	National Oceanic and Atmospheric Administration
OLCF	Oakridge Leadership Computing Facility
P3	Predicted Particle Properties
PI	principal investigator
PNG	portable network graphics
QC	quality control
RAP	Rapid Refresh model
RRTMG	Rapid Radiation Transfer Model for Global Climate Models
RWP	radar wind profiler
s	second
SAM	System for Atmospheric Modeling

SGP	Southern Great Plains
TSI	Total-Sky Imager
U.S.	United States
VAP	value-added product
VARANAL	Variational Analysis (forcing product)
WRF	Weather Research and Forecasting model

Contents

Figures

Tables

1.0 Introduction

The U. S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Climate Research Facility began a pilot project in May 2015 to design a routine, high-resolution modeling capability to complement ARM's extensive suite of measurements. This modeling capability, envisioned in the *ARM Decadal Vision* (U.S. Department of Energy 2014), subsequently has been named the Large-Eddy Simulation (LES) ARM Symbiotic Simulation and Observation (LASSO) project, and it has an initial focus of shallow convection at the ARM Southern Great Plains (SGP) atmospheric observatory. This report documents the recommendations resulting from the pilot project to be considered by ARM for implementation into routine operations.

The goal of the LASSO pilot project has been to design a routine, high-resolution modeling capability to complement ARM observations. During the pilot, LASSO has evolved from the initial vision outlined in the pilot project white paper (Gustafson and Vogelmann 2015) to what is recommended in this report for establishing an operational LASSO capability. Further details on the overall LASSO project are available at https://www.arm.gov/capabilities/modeling/lasso. Feedback regarding LASSO and the recommendations in this report can be directed to William Gustafson, the project principal investigator (PI), and Andrew Vogelmann, the co-principal investigator (Co-PI), via lasso@arm.gov.

1.1 Scope and Goals

The LASSO concept is to routinely produce LES simulations that complement ARM observations. The initial scope includes producing ensembles of LES at the SGP observatory for roughly 30 case dates per year that have shallow convection. The LES output will be bundled with a selection of coincident observations, and simulation performance skill scores and diagnostics, all of which will be made available to users between three and six months after the calendar date of cloud occurrence. Specific details are provided via the list of recommendations in Section 2.

The motivation for LASSO is to assist scientists in bridging the gap from point measurements to larger scales relevant to common atmospheric modeling. This is done by combining the measurements with LES modeling to provide a self-consistent representation of the atmosphere around the ARM site that can be used to better understand the measurements and the processes leading to the observed conditions. The LES must be interpreted within its representativeness, which is a statistical representation of the atmospheric state around SGP.

The availability of LES simulations with concurrent observations serves many purposes:

- LES helps bridge the scale gap between ARM observations and coarse atmospheric models.
- The use of routine LES adds value to observations by providing a self-consistent representation of the atmosphere and a dynamical context for the observations.
- LES provides representations of unobservable processes and properties.

LASSO is designed to address a versatile range of science applications, and it has been envisioned that this framework could be used by, but not limited to, three primary user categories characterized as observationalists, theoreticians, and modelers. Observationalists may not have the expertise to generate their own model simulations and will be able to take advantage of the readily available LASSO simulations to act as synthetic, known data to improve the understanding of their retrievals and to help improve their design. For example, observationalists could use the volumetric model output to test radar scan strategies and better understand how to sample the spatial variability (Oue et al. 2016). Similarly, theoreticians can benefit from having readily available, vetted simulations to test physical relationships as well as obtain values that are either difficult or impossible to measure, such as the spatial variability of fluxes and their co-variabilities. Modelers can benefit from LASSO by having readily available control simulations combined with forcing data sets that are already validated against ARM observations to determine which days have valid forcings that can be used for further modeling studies. The modelers will also have the tools from LASSO to quickly compare their model simulations with the observations using the processed observations and skill scores developed for LASSO. This will simplify parameterization development for the boundary layer and clouds. Future LASSO enhancements could also target land-atmosphere interactions through the incorporation of soil models and interactive feedbacks between the land and atmosphere.

A key feature of LASSO is the generation of a simulation library for researchers that enables statistical approaches beyond a single-case mentality. Such a statistical library of cases is essential to addressing atmospheric variability for process-level understanding, as well as parameterization evaluation and improvement. One envisioned example of how LASSO could be used is to assist with parameterization development by providing forcing data for a single-column model, the output of which can then be compared with the LASSO LES and bundled observation data for a large number of cases. Another example includes using the LES as a proxy of realistic conditions that can assist with developing remote retrieval methodologies and understanding instrument-sampling statistics. Having the library of cases permits developing an understanding of uncertainty due to variable conditions.

As requested by the initial DOE call for LASSO white papers, the initial focus of the LASSO pilot is on shallow convection at the SGP site. Thus, the recommendations in this report specifically target this application. However, the larger vision for LASSO (U.S. Department of Energy 2014) includes expansion to additional locations and meteorological regimes, and thus, the pilot has considered these expansion possibilities when choosing the overall modeling design.

The pilot project has provided prototype software and a set of recommendations for an operational system capable of meeting ARM's goal of simulating shallow convection at the SGP. This includes details such as model selection and configuration, criteria for days to simulate, observations to include, skill scores and an interactive web interface for evaluating the simulations, and estimates of computational cost.

1.2 Pilot Project Outcomes

LASSO has been successful in delivering a system with the requested capabilities, detailed in this report, which successfully demonstrates that routine LES simulations of shallow convection can be done by ARM. Key accomplishments are:

- An ensemble of forcing data sets has been assembled that increases the ability to reproduce the observed cloud field and environmental properties in LES simulations.
- A foundational set of observations, model diagnostics, and skill scores have been assembled for evaluating the simulations.
- The modeling and evaluation system has been tested on two summer seasons (May to August) at the SGP for 2015 and 2016 and the data released in the form of preliminary data sets referred to as Alpha 1 (ARM 2016, Gustafson et al. 2016) and Alpha 2 (ARM 2017, Gustafson et al. 2017), respectively. These are available for user feedback and beta users to apply to research questions.
- An interactive web-page interface has been developed to query for simulations of interest and cross-compare simulation behavior against selected ARM observations.
- A targeted communications strategy, partially outlined in Appendix B, has increased community awareness of LASSO. This has resulted in multiple early adopters on a range of topics.

1.3 Metrics of Success for LASSO Operations

A request for the LASSO pilot is to provide input on metrics of success for LASSO operations once LASSO is fully implemented. The proposed metrics categories, below, include a number of dimensions in assessing its utility and impact. Metrics developed for the following four categories would provide a broad evaluation of how LASSO is impacting the community as well as indicate areas where LASSO could be improved. Tracking can begin during the first year of operations and will be useful for evaluating model behavior, while metrics related to community use will not show meaningful results for several years.

1. *LES Skill:* Many applications require LASSO to include simulations representative of the atmospheric conditions it simulates. Thus, comparison of the LES with available cloud and boundary-layer observations will be done. A subset of the skill scores built into the data bundles are appropriate for this use.
2. *Productivity:* Interest in and adoption of LASSO is anticipated to increase as early users demonstrate LASSO's benefits to their peers in accomplishing their research goals. Tracking use of LASSO in professional meeting presentations and publications (a lag indicator) will provide a clear sign that LASSO is being used productively. The research usage statistics should be analyzed to identify the categories of applications using LASSO to better target future LASSO development.
3. *Community Use:* Utility within and outside of DOE is sought. Within DOE, we expect that researchers from the Atmospheric System Research (ASR) Program, Climate Model Development and Validation (CMDV) projects, and Energy Exascale Earth System Model (E3SM) project will use LASSO. Use by researchers funded by other U.S. and international agencies will be a clear indication of penetration into the broader research sphere.
4. *Data Usage:* Considerable resources will be expended to generate and archive the LASSO simulations. An assessment should be made of cases or simulation types that go unused to inform criteria used to select simulations. This will avoid generating future simulations that are of minimal interest. This metric will require time to accumulate representative statistics, so it should not be

employed during the initial five years of operations. Otherwise, premature stifling of datastream components could occur.

1.4 Plans for Extending LASSO

The *ARM Decadal Vision* (U.S. Department of Energy 2014) calls for expanding ARM's high-resolution modeling to multiple sites and meteorological regimes. A plan to assess options for extending LASSO to phenomena beyond shallow convection at the SGP will be developed during the fall of 2017 with a deliverable of the plan for how to proceed with the decision process prepared by the end of February 2018. The plan is expected to include a workshop involving specialists who work with modeling and observations for other potential phenomena and sites. The critical factors for deciding the next target phenomena will include: 1) likely scientific impact within the user community and relevance to DOE objectives, 2) likely success of the LASSO modeling framework to reproduce the target phenomena, 3) availability or ability to make available critical observations, and 4) cost.

2.0 Recommendations

The primary deliverables from the LASSO pilot project are the recommendations for what should be implemented for routine operations by ARM along with prototype software. The recommendations fill out the details for the elements of the proposed LASSO workflow, shown in Figure 1, where the large blue box contains all of the LASSO components and each of the blue bubbles represents a major workflow element. The general process begins by ingesting a large amount of data, which is then used for generating large-scale forcing data, initializing the LES, and producing the diagnostics and skill scores within the data bundles. The resulting bundles are then made available to users.

The two alpha-release datastreams (ARM 2016, 2017) generated during the pilot are examples of potential operational products and have served as a means to gain community feedback regarding what aspects of the LASSO evaluation datastreams are the most valuable and ways that they could be improved. Combined, these datastreams consist of 736 data bundles, each of which contains output from a unique LES configuration from a selection of 18 days with shallow convection at the SGP site. The particular LES configurations are designed to compare a selection of different large-scale forcing data sets, model dependency, model domain choices, and physics parameterization sensitivities. Packaged with the LES output is a selection of ARM observations relevant to evaluating simulated shallow convection plus skill scores and diagnostics indicating how the LES compares with the observations.

Based on experience gained during the pilot combined with feedback from the ARM Atmospheric Modeling Advisory Group and community members at large, the following recommendations are made by the pilot project team. The recommendations are summarized in Table 1 followed by details supporting the choices provided in the remainder of this section. Many of the recommendations are specifically requested in the call for white papers that resulted in the pilot project, and other recommendations have evolved during the pilot phase.

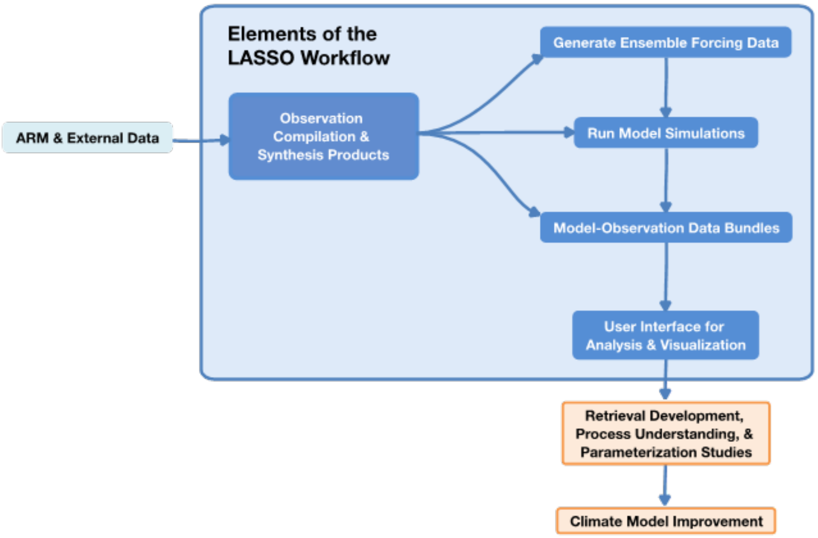


Figure 1. Schematic of the LASSO workflow elements and connection to community use.

Table 1. Overview of LASSO recommendations. "M," "D," "O," and "F" number prefixes indicate modeling, data bundle, operations, and future development recommendations, respectively.

Number	***Recommendation***[*]
M1	LASSO should use the WRF model as its LES model.
M2	The LES domain should use 100 m horizontal grid spacing, have a horizontal domain extent of 25 km, vertical grid spacing of 30 m near the surface, a model top near the tropopause, and use doubly periodic lateral boundary conditions.
M3	The WRF dynamics and physics parameters for operations should follow the attached namelist file.
M4	Forcings for the LES should be based on multiple sources including the ARM VARANAL datastream, the ECMWF IFS model, and the multiscale data assimilation methodology (MSDA; see Section 2.1.4). Multiple forcing spatial-scales should be used for the ECMWF and MSDA sources, about 75, 150, and 300 km.
M5	An ensemble of eight LES should be used for each case day based on the forcings in M4. This would consist of VARANAL at 300 km; ECMWF at 16, 114, and 413 km; MSDA at 75, 150, and 300 km; and an additional simulation with no large-scale forcing.
M6	The LASSO data bundles should contain model output every 10 minutes consisting of 3D instantaneous snapshots combined with summary LES statistics, e.g., domain average vertical fluxes.
D1	The LASSO data bundles should contain a suite of observations in a form comparable to the model output. Table 3 lists the observations by category.

---

[*] Note that acronyms are defined in the acronym list and upon first use in the main text.

Number	***Recommendation******
D2	The LASSO data bundles should contain a suite of diagnostics and skill scores selected to assist users in choosing appropriate simulations for their needs. Table 4 lists diagnostics and skill scores specifically targeting shallow convection.
D3	The data bundles should use a netCDF file format. The model output should be in the standard format output by WRF. Observations and skill scores should be collected into summary files to ease intercomparison.
D4	The files within each data bundle should be packaged into a series of tar files to ease download and storage within the ARM repository. One tar should contain the model configuration and information required to reproduce the simulations combined with the skill scores, diagnostics, and quicklook plots. A second tar file should contain the LES statistics from the model output. A third tar file should contain the instantaneous model snapshots, which would be much larger than the other two tar files.
O1	The initial implementation of LASSO should follow the initial plan to simulate days at SGP with non-precipitating shallow convection.
O2	A minimum of 10 million core hours per year should be devoted to running the LES for typical shallow convection days. Additional resources should be devoted for ongoing LASSO development and user interaction with the data bundles.
O3	The observations listed in Table 6 are required for running the LES and producing the data bundles. These observations include ARM cloud and meteorological measurements combined with data acquired from external sources, e.g., satellite data for data assimilation purposes.
O4	Data requirements and simulation run times indicate that users will need to wait a minimum of 3.5 months after an occurrence of shallow convection before the associated data bundle will be available.
O5	User support is of primary importance for building a community around the LASSO product and sustaining support for its generation. ARM should budget for sufficient labor to support user interactions during the first year and expect the need to increase as the user base increases.
O6	A priority should be placed on first developing robust code for operations, which should be done with the external reproducibility in mind when writing modular workflow software. Software necessary to reproduce the LES results, diagnostics, and skill scores should ultimately be made available via a publically facing repository with an open source copyright.
O7	An interactive web-based tool should be provided to users to query data bundles based on simulation metadata and skill cores that compare the LESs to observations.
O8	ARM should provide user access to an ARM computer(s) to ease data analysis of the very large data set. A portion of the computing should also be made available for users to do additional sensitivity simulations to compare with LASSO simulations. However, this should be carefully tracked and large requests should be diverted to traditional DOE computing facilities, e.g., NERSC.
O9	ARM should budget sufficient resources to permit ongoing development of LASSO. This will entail maintaining the model code, adding LASSO features as new observations become

Number	Recommendation*
	available, enhancing the model output, diagnostics, and skill scores to meet new user scenarios, improving the data assimilation, etc.
F1	A nested LES domain configuration should be tested at SGP to better capture spatial heterogeneity and as a preliminary step toward simulating alternate cloud types and ARM sites where the periodic configuration could be limiting.
F2	Further development of the MSDA methodology should be pursued. Particularly, it should be moved from a 3DVar to an ensemble hybrid-Kalman filter methodology. Also, additional improvements could be achieved with further fine-tuning of the MSDA grid and use of observations.
F3	A blended forcing product could be produced combining the gridded MSDA approach that incorporates ARM profile observations with the VARANAL approach that best utilizes ARM flux measurements.
F4	The pilot project had resources to design the data bundles around high-priority variables. Additional resources should be devoted to more fully including available observations, e.g., Doppler lidar profiles and photogrammetric cloud masks.
F5	LASSO data assimilation would benefit greatly from hourly thermodynamic and wind measurements near the top of the boundary layer at multiple locations in the region. In particular, high-vertical-resolution observations of the inversion are needed.
F6	Complete implementation of an ARSCL simulator based on CR-SIM and use this to replace the model cloud masks when comparing with ARSCL in the data bundles.
F7	A quality control protocol is needed for the data bundle generation.
F8	Usage of the Bundle Browser should be closely tracked and resources allocated to continually improve the Bundle Browser toward better adaptation to user needs.
F9	A methodology must be developed for communicating data provenance within the data bundles that aligns with ARM procedures and is easy for users to understand.

The specific recommendations addressed by each of the following sections are indicated by the numbers in the table above.

2.1 Modeling Recommendations

The first broad category of recommendations relates to the model selection and configuration for LASSO. This set of recommendations is essentially distinct from those related to the handling of observations, evaluation of the model, and day-to-day operations.

2.1.1 Model Selection

Recommendation M1:
LASSO should use the Weather Research and Forecasting (WRF) model as its LES model.

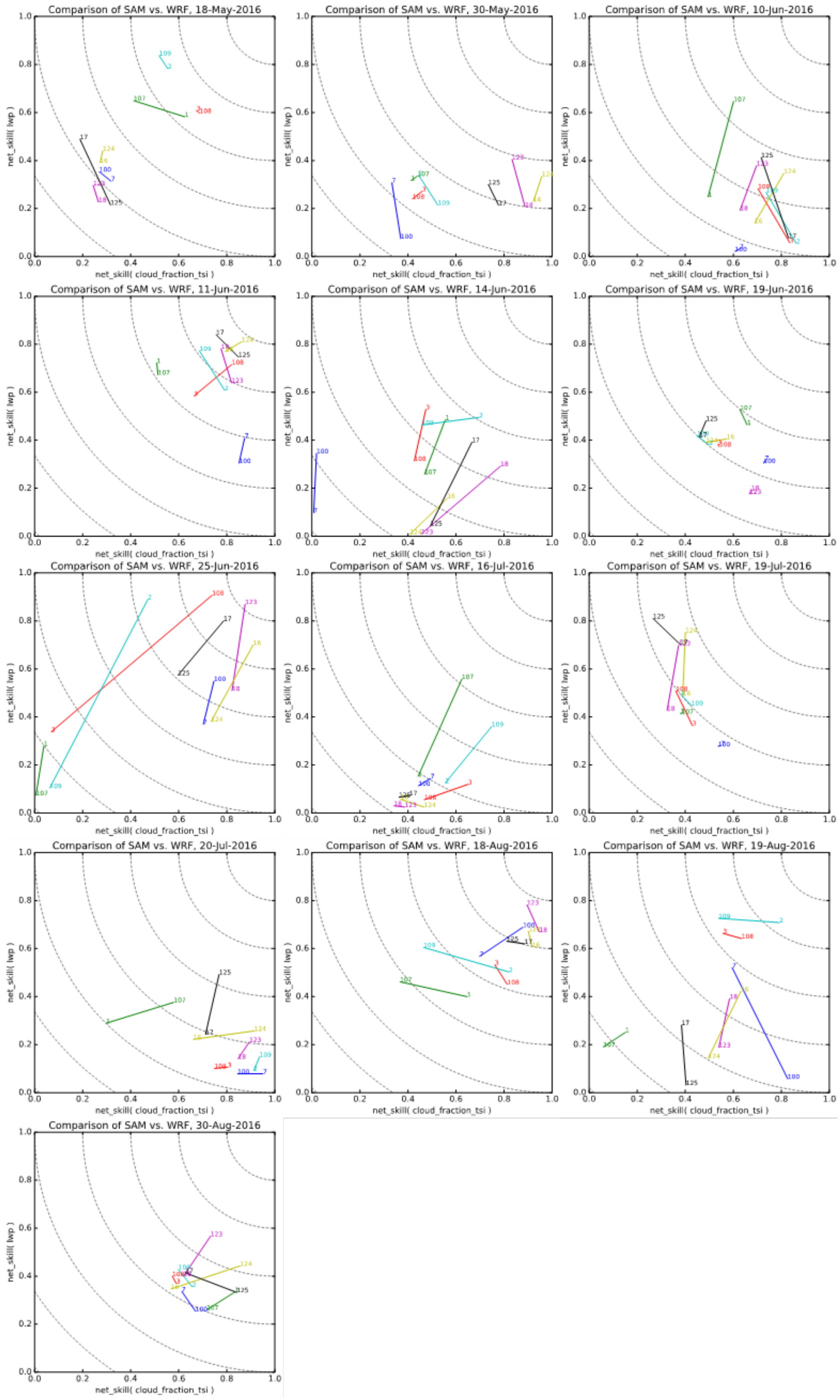
The LASSO pilot tested the System for Atmospheric Modeling (SAM) (Khairoutdinov and Randall 2003) and WRF (Skamarock et al. 2008) models for producing the LES runs. We recommend that only one of these be used for operations to reduce the overhead cost of automating and maintaining multiple models. While the two models do not always provide the same results, neither model is decisively better, as one produces better results for some situations while the other model does for other situations. Deciding which model to use operationally involves a range of criteria and an overall compromise that best meets ARM requirements. The following table summarized the criteria. We note that the ARM simulations can be used as a starting point with the model provided; then, users desiring output from alterative models can use the forcing data sets within the LASSO data bundles to generate additional simulations.

Table 2. Considerations for model selection.

Criteria	***SAM***	***WRF***
Community-based code	No, but freely available upon request from the developer; occasional updates from the developer	Yes, available via Github at https://github.com/NCAR/WRFV3; twice-annual version updates
User community	Small, but has been used in important modeling projects such as the multiscale modeling framework and development of CLUBB	Large, 39,180 registered users as of June 2017 with 20,900 new users in the previous 5 years, ~8,000 active subscribers to the WRF News listserv (Klemp 2017)
Configuration options	Traditional LES and CRM setups with periodic lateral boundaries and a flat bottom Morrison microphysics Release version of code does not include an interactive land model, but a version has been developed by a user that includes the Noah model	Traditional LES and CRM setups with periodic boundaries plus the ability to do nested simulations with realistic terrain and time-dependent, spatially varying lateral boundaries Multiple microphysics options in the standard, released code including Morrison, Thompson, and P3 Multiple land model options to permit land-atmosphere interactions and dynamically calculated surface fluxes
Accuracy	The probability distribution function of vertical velocities differs compared to WRF—it is unclear which model is more accurate in this respect Comparisons using the LASSO skill scores do not reveal a clearly better model. Both can give comparable results, and which one is better	The governing equations do not make the anelastic assumption, and therefore are potentially more accurate for some situations The availability of the Thompson microphysics permits a configuration with less abundant cirrus cloud, which

Criteria	***SAM***	***WRF***
	depends on the case (see discussion in the text)	both models tend to over-predict using the Morrison scheme
Computational cost	Can use a time step about 2 times longer than WRF A LASSO configuration with 14.4 km-wide domain, 100 m grid spacing, a 0.5 s time step, 15 h model integration, and single precision takes 3 hours to complete using 144 cores on Eos at OLCF	More easily adapts to large processor counts for various I/O methods Additional computing time could be saved by using parallel I/O, which was not tested during the pilot due to the new ARM cluster not being available until the end of the pilot A comparable configuration to the SAM run (also with dt=0.5 s) takes 8 hours to complete on Eos using 256 cores
Data assimilation	SAM is not part of any data assimilation system	Several data assimilation packages are available that work with WRF, of which, LASSO uses the GSI package to generate the MSDA forcing data set

Comparing model behavior is somewhat subjective and how it is done depends upon the particular needs of the user. An extensive comparison is beyond the scope of this report. However, the LASSO net skill scores for liquid water path (LWP) and total-sky imager (TSI)-based cloud fraction (CF_{TSI}) are helpful to show that on many days the models give very similar results, and when they differ there is no clear bias toward one model. The skill scores are a normalized score with range [0,1] where 1 is best, as described in (Gustafson et al. 2017). The net skill score represents a combination of the relative mean bias and the Taylor skill score. It captures both the magnitude of the model value versus observations as well as the time dependence. Figure 2 shows pairs of net skill scores for SAM and WRF, where each color-coded pair contains a value <100, which is the WRF simulation, and a value ≥100, which is the SAM simulation. The pairs for simulations are as directly comparable as possible except for the model used. Each simulation uses identical forcing, initial conditions, Morrison microphysics, and domain configuration. Multiple pairs exist for each case day in the Alpha 2 set of simulations due to the use of multiple large-scale forcing data sets. On some days, such as June 19, 2016, the models produce almost identical skill score values. On other days, they differ more, with the largest difference on June 25, 2016. On this day, two SAM simulations do well as do two WRF simulations, but for different forcings. The model differences leading to the large skill score differences result from tenuous clouds in the simulations that do not develop nearly as deeply as they should in the lower-scoring simulation. From these comparisons, WRF and SAM have comparable skill scores overall, although one may outperform the other in a given case and vice versa.


Comparison of SAM vs. WRF, 18-May-2016

Comparison of SAM vs. WRF, 30-May-2016

Comparison of SAM vs. WRF, 10-Jun-2016

Comparison of SAM vs. WRF, 11-Jun-2016

Comparison of SAM vs. WRF, 14-Jun-2016

Comparison of SAM vs. WRF, 19-Jun-2016

Comparison of SAM vs. WRF, 25-Jun-2016

Comparison of SAM vs. WRF, 16-Jul-2016

Comparison of SAM vs. WRF, 19-Jul-2016

Comparison of SAM vs. WRF, 20-Jul-2016

Comparison of SAM vs. WRF, 18-Aug-2016

Comparison of SAM vs. WRF, 19-Aug-2016

Comparison of SAM vs. WRF, 30-Aug-2016

Figure 2. Pairings of SAM and WRF simulations for net skill scores of liquid water path (LWP) and total-sky-imager–based cloud fraction (CF_{TSI}). WRF simulations are indicated by simulation ID labels <100. SAM simulations have values ≥100. Colors indicate different large-scale forcing data sets used to drive the models: 300 km VARANAL, 16 km ECMWF, 114 km ECMWF, 413 km ECMWF, 75 km MSDA, 150 km MSDA, and 300 km MSDA. Values closest to (1,1) indicate better performance, with range rings provided as dashed lines to show equal distances from a perfect score. See text for a full description.

The cost of WRF is more than SAM by a factor of about 5, which could be a limiting factor if computing resources were limited. However, the overall cost of the recommended model configuration, outlined later in this report, is cheap enough to be able to run an ensemble of simulations for each case with WRF and still leave a significant amount of computing time on Cumulus, the ARM cluster for LASSO. This available time can be used for further LASSO development needs, user interaction with data bundles, and future LASSO expansion of operations.

Another factor in choosing WRF for the LES model is that WRF is also the model used to generate the multiscale data assimilation (MSDA) large-scale forcing. Using the same model for both components of LASSO will save money by leveraging model updates and related code maintenance across the weather hindcasts for MSDA and the LES simulations. WRF will also permit future nested LES domain configurations, which could be embedded directly in the MSDA grid to have a seamless, integrated system across model scales from the mesoscale to the cloud scale. Alternatively, SAM cannot be nested and can only use region-averaged large-scale forcings. While this could be sufficient for the initial LASSO implantation for shallow convection that uses only the profile-based forcing derived from MSDA, the inability to form a nested system that incorporates data assimilation could limit future expansion options.

One subtlety in the WRF selection is which version of the model to use. Yamaguchi and Feingold (2012) show a strong time-step dependency in version 3.3.1 of WRF, which Xiao et al. (2015) show is due to a poor assumption regarding moisture advection in the presence of strong moisture gradients. They propose a fix to WRF consisting of using the moist potential temperature in the prognostic equation instead of the potential temperature, which results in WRF having a similar time step sensitivity to other LES models. This fix has been implemented in version 3.7 and subsequently improved in version 3.8. Thus, a recent version of WRF should be used by LASSO.

We also recommend that the LES additions to WRF from the DOE FAst-physics System TEstbed & Research (FASTER) project (Endo et al. 2015) be used with WRF. This is one of three LES packages available for WRF, the others being one built into the off-the-shelf model plus one described in Yamaguchi and Feingold (2012). The built-in LES capabilities are limited, e.g., they do not include the ability to output domain-averaged LES statistics, so they are not an option for LASSO. The package by Yamaguchi and Feingold provides the capabilities needed by LASSO but it is essentially a port of the SAM statistics package into WRF and it is not implemented in an easily maintainable way. The FASTER package provides very similar statistics and forcing capabilities, yet is implemented in a "WRF-like" way, which enables easy addition of new variables and simplifies overall code maintenance.

2.1.2 Domain Configuration

Recommendation M2:
The LES domain should use 100 m horizontal grid spacing, have a horizontal domain extent of 25 km, vertical grid spacing of 30 m near the surface, a model top near the tropopause, and use doubly periodic lateral boundary conditions.

To assess an optimal all-purpose configuration to be used, tests have been done with SAM comparing grid spacing and domain size for a selection of four shallow convection cases in 2016. The VARANAL large-scale forcing drives each of the simulations and the overall goal is to understand the convergence of the results for macro cloud properties. This analysis informs the current phase of LASSO, but users should note that other cloud conditions could have different results. For example, radiatively driven stratus would require higher resolution and deep convection could be adequately simulated with lower resolution. In addition, the shallow convection around SGP is assumed to be statistically homogeneous, which enables the use of periodic boundaries. More complicated scenarios, such as shallow-to-deep transitions of convection, would require a different setup, including domain size. Similarly, while the optimized configuration sought here is intended to be sufficient to support multiple types of studies, higher resolutions might be needed by some users. They could use the knowledge of which LASSO forcings perform best to target doing their own additional higher-resolution simulations.

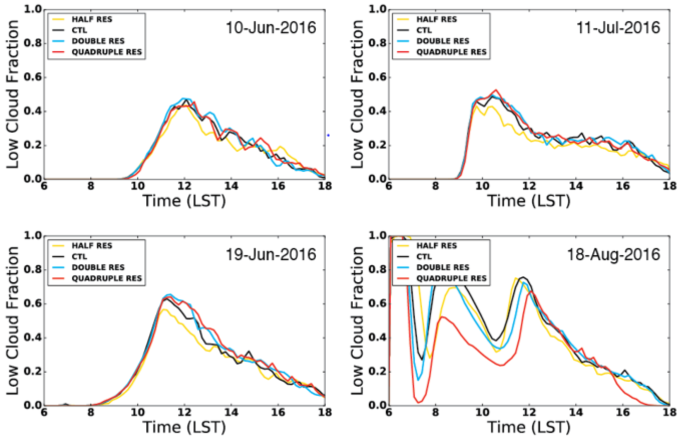


Figure 3. Simulated low-cloud fraction for grid spacing comparisons from Alpha 2. The control (CTL) grid spacing is 100 m, and the resolutions are relative to CTL.

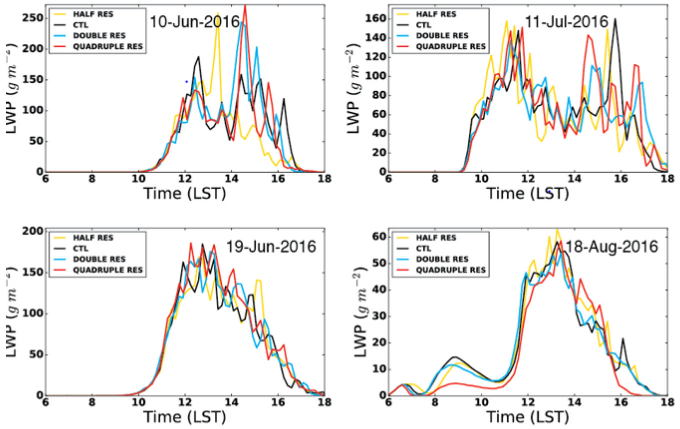


Figure 4. Simulated LWP for grid spacing comparisons from Alpha 2. The control (CTL) grid spacing is 100 m, and the other resolutions are relative to CTL.

Horizontal resolution: Comparison of 25, 50, 100, and 200 m grid spacings generally indicates that a large grid spacing of 200 m results in a lower domain-average cloud fraction and the results from 25 to 100 m grid spacing are more consistent with each other, indicating a modicum of convergence for cloud fraction. This is true for three of the four cases shown in Figure 3. The exception is August 18, 2016 when the 25 m grid spacing (quadruple resolution) deviates from the other grid spacings for much of the day. An analysis of LWP (Figure 4) also shows that 200 m grid spacing (half resolution) often is the most different from the highest-resolution simulation. This can be seen on June 10, 2016 when the half resolution simulation has higher LWP during early afternoon and much lower LWP after 14 LST. The LWP results are noisier than the cloud fraction and are therefore more difficult to interpret for consistent biases.

Overall, some cases are marginally better using 50 m grid spacing instead of 100 m, but 25 m does not noticeably improve over 50 m. To better capture every case, a 50 m grid spacing could be adopted. However, 100 m grid spacing is recommended. In most cases, this will produce good results, it is eight times cheaper computationally to use than 50 m, and the output files are smaller by a factor of 3.7, which eases their use and storage needs.

While a 100 m grid spacing is recommended as sufficient for a broad range of users while keeping the file sizes as manageable as possible, a subset of users could require finer grid spacings to meet their research needs. For this subset, they could use the LASSO simulations as a starting point to determine which simulation configuration they want to use since the overall cloud behavior is very similar between grid spacings of 25, 50, and 100 m. Then, they could use the LASSO software along with the input and configuration data available with the data bundles to generate simulations at the desired grid spacing. Ultimately, the choice for ARM relates to how many of the different research possibilities it should address directly versus providing the starting point for value-added work.

Domain size: Domain widths of 7.2, 14.4, 28.8, and 57.6 km have been tested using the same four cases as used for the grid spacing comparisons. The intermittency and spatial variation of shallow clouds in the selected cases leads to sampling noise when the domain becomes too small to contain a statistically stable sample of clouds. This results in noisy time series for domain-averaged values, and the 7.2 km domain width (half resolution) is a clear outlier for both low-cloud fraction (Figure 5) and LWP (Figure 6). A small amount of noise is still present in the 14.4 km domain, particularly for the LWP, and the two largest domains show a general convergence, particularly for cloud fraction. Thus, while the Alpha 2 simulations use a standard domain size of 14.4 km, we recommend using a ~25 km domain for operations. This would increase the computational cost over the 14.4 km domain by a factor of ~4, which is still affordable.

Vertical grid configuration: All simulations in the Alpha 2 release have 226 levels that extend from the surface to 14.7 km. Vertical grid spacing is 30 m up to 5 km and then stretches to 300 m near the model top. This configuration is determined from a series of sensitivity studies using Alpha 1 cases, and it is found to work well for Alpha 2 and would be a good choice for the operational configuration.

Model top: We recommend that the domain top be placed near the tropopause. This is motivated by the desire to simulate cirrus clouds that could be present above the shallow convection. Presently, the lower boundary fluxes are prescribed from ARM observations so cirrus do not impact the simulation surface conditions at the ground. However, researchers applying LASSO for upward-looking purposes, such as with simulating zenith-pointing instruments, benefit from having the model top extend beyond the mid-troposphere.

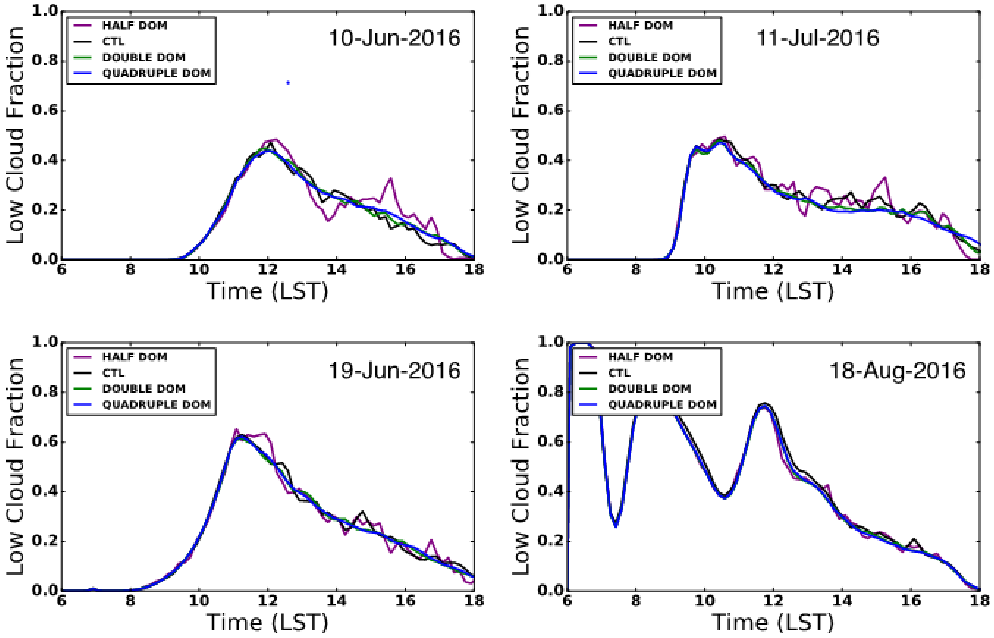


Figure 5. Simulated low-cloud fraction for domain width comparisons from Alpha 2. The control (CTL) simulation uses a 14.4-km-wide domain.

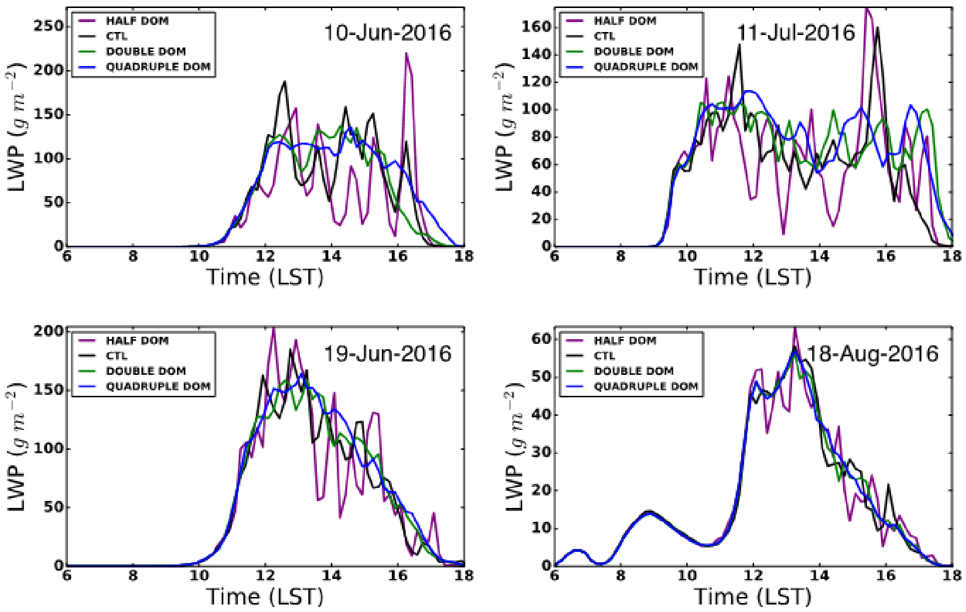


Figure 6. Simulated LWP for domain width comparisons from Alpha 2. The control (CTL) simulation uses a 14.4-km-wide domain.

Lateral boundary conditions: Only periodic lateral boundary conditions have been tested within the LASSO pilot. This choice was made because there is currently no obvious way to use ARM observations to initialize the soil conditions if one were to use an interactive land model with homogeneous soil characteristics. Periodic boundaries imply a homogeneous lower boundary, and a representative soil type could be chosen for use in the model for the SGP region. However, the larger difficulty would be to use discrete ARM soil profiles to determine a repeatable and consistent way to initialize the soil temperature and moisture profiles accurately enough to permit faithful simulations.

Also, nested LES is a relatively new research area and it is not yet known how well it will hold up over many cases. That said, research on a related DOE ASR project has shown promising results for nested LES that can better capture changing synoptic conditions throughout the day as well as capture the west-east gradient in cloud characteristics at SGP. Thus, we recommend that LASSO initially implement periodic lateral boundaries, and that resources be devoted to designing and testing a nested LASSO approach in the near future. This would involve determining the best way to nest into the MSDA forcing simulations (and possibly three-dimensional [3D] VARANAL forcings when available), running a series of cases to compare with periodic boundaries, and revising the skill scores in the data bundle since domain averages would no longer be appropriate. Re-evaluation of the LES statistical output from the model should also be done since domain-wide averages would not work with the varying terrain at the lower boundary, as well as the need to exclude the near-boundary locations from any averaged statistics. These tests can be used to determine whether the benefits to simulation accuracy justify the additional complexity.

2.1.3 Model Version and Parameter Settings

Recommendation M3:
The WRF dynamics and physics parameters for operations should follow the namelist file in Appendix A.

The recommended WRF parameter settings derive from standard choices used for LES-style modeling and are shown in Appendix A. In particular, the LES setup includes the subgrid-scale parameterization based on the 1.5-order turbulent kinetic energy approach of (Deardorff 1980) (*km_opt*=2) combined with the use of moist potential temperature in the acoustic sub-steps (Xiao et al. 2015) (*use_theta_m*=1). Physics choices include Rapid Radiation Transfer Model for Global Climate Models (RRTMG) shortwave and longwave radiation (Clough et al. 2005; Iacono et al. 2008; Mlawer et al. 1997), no boundary-layer or convection schemes, and Thompson microphysics (Thompson et al. 2004, Thompson et al. 2008).

Testing in the Alpha 1 and 2 releases included comparison of results from the Morrison (Morrison et al. 2005, Morrison et al. 2009) and Thompson microphysics. The two schemes provide similar results for the shallow clouds as shown by plots of paired simulations where the only difference is the microphysics choice. Figure 7 shows an example for June 11, 2016 in which one can see that the differences are small. Other days show similar or smaller variations. Where Thompson sometimes has the advantage is with cirrus, in which it has a tendency to not overestimate the cirrus as much as Morrison.

Some work has been done with Alpha 1 to identify a data source to routinely provide aerosol information to specify with use of two-moment microphysics. However, this has not been fully pursued at this time.

Variability in the simulated clouds due to the choice of large-scale forcing is larger than what would occur due to typical variations in aerosol. So, a constant default background aerosol number is used.

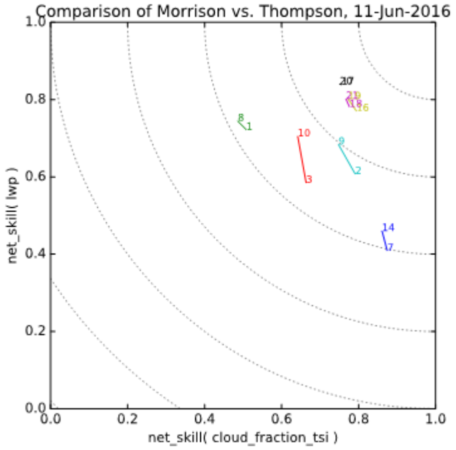


Figure 7. Pairings of WRF simulations for microphysics showing net skill scores for LWP and CF_{TSI}, similar to Figure 2. Lower-numbered simulation IDs within the pairs use Morrison, and higher-numbered IDs use Thompson microphysics.

2.1.4 Forcing Data

Recommendation M4:

Forcings for the LES should be based on multiple sources including the ARM VARANAL datastream, the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecast System (IFS) model, and the multiscale data assimilation methodology. Multiple forcing spatial-scales should be used for the ECMWF and MSDA sources, about 75, 150, and 300 km.

The LASSO pilot has tested large-scale forcings based on the ARM Variational Analysis (VARANAL) (Xie et al. 2004), ECMWF's IFS forecasts, and incorporation of ARM-specific data in a data assimilation product using the Multiscale Data Assimilation approach (MSDA) (Li et al. 2015a; Li et al. 2015b; Li et al. 2016). All three base their methodology on weather forecasts with adjustments made in different ways. Each of these is an equally plausible representation of the atmosphere around SGP and represents different methodologies for obtaining this representation. We recommend that all three be maintained for the operational implementation of LASSO.

The VARANAL approach is a standard ARM value-added product that has been available for the SGP site for many years. It performs best when used with a ring of five or more soundings that are used to determine the fluxes in and out of the ring through doing a line integral around the perimeter of the soundings. This is then combined with other ARM observations, such as the surface fluxes and precipitation, to constrain the estimate of the atmospheric conditions. When run outside of a field campaign when extra radiosondes are unavailable, VARANAL uses a gridded background field to estimate the overall flow and associated horizontal fluxes. In the case of LASSO, VARANAL uses the National Oceanic and Atmospheric Administration (NOAA) Rapid Refresh (RAP) forecast model for this background field, which is available retroactively.

ECMWF's IFS model is a direct weather forecast that uses a four-dimensional variational (4DVar) data assimilation approach. Of the three methods, this one uses the least amount of ARM observations. At most, it uses the ARM radiosonde soundings in its data assimilation process when they are available. However, the data assimilation used in IFS is arguably the most sophisticated of the three methodologies, and thus best uses the operationally available observations. The forcing used for LASSO is calculated from the IFS model using the Diagnostics in the Horizontal Domains (DDH) system, which uses physical and dynamical tendencies directly from the model to calculate closed budget terms. This method can accurately close the moisture and energy budgets, plus it can be used over multiple spatial extents to generate forcings representing different horizontal scales. At this time, we sample one deterministic IFS simulation and do not attempt sampling the spread produced by the ensemble generated by the IFS system.

The MSDA approach uses the WRF GSI code to do a multiscale data assimilation approach, which incorporates observations at multiple scales of influence to generate a high-resolution analysis that best uses available data. For LASSO, the analyses have been generated at 2 km grid spacing and incorporate satellite measurements, observations from the operational meteorological observing network, and a subset of ARM observations that includes output from surface meteorology stations and radiosondes. In addition, an analysis has examined the impact of ARM wind profiles from the four radar wind profilers at SGP. In general, the impact is small, and some cases show improvement, while others have degraded cloud forecasts when incorporating the ARM wind profiles. Further testing is needed to understand the full potential ARM observations can play in improving the data assimilation process. Thermodynamic profiles from remote-sensing instruments, such as Atmospheric Emitted Radiance Interferometer (AERI) and the Raman lidar, could be included once a methodology is developed to better handle calibration issues and provide reasonable quality controls.

In addition to the different methodologies used to generate the forcings, the ECMWF and MSDA methods can produce forcings representing multiple spatial scales by using differing numbers of grid columns when calculating the forcing. This can be useful for handling different situations when the cloud field is not homogeneous around SGP. VARANAL is produced for a 300 km region, so this scale was also selected for ECMWF and MSDA. Then, subsequent smaller scales were also chosen. Because of the spectral nature of the IFS, identical sizes were not possible, but the general goal is about half the size and a quarter of the size of the full-scale forcing. During the pilot, tests were done for spatial scales with diameters of 16 (a single column), 114, and 413 km for ECMWF, and 75, 150, and 300 km for MSDA.

The colors for each pair in Figure 2 indicate each forcing type, and the figure provides a general sense of the range of results generated by each forcing selection. No consistent bias exists for any of the forcing types or scales, i.e., no one forcing type or scale is superior to the others. Each can perform well for a given case, and the best performer cannot be predicted a priori. Thus, we recommend including all of these forcing options when implementing LASSO.

In addition to large-scale forcings, LES models with doubly periodic boundaries require surface forcing data. ARM regularly measures surface fluxes of sensible and latent heat at multiple locations around the SGP site. Therefore, we recommend that these be used as the lower boundary forcing for the LES. The most convenient way to do this is through the regionally averaged sensible and latent heat fluxes contained in the VARANAL product.

We recognize a need for the forcings to be used in other models such as SAM, which has a significant user base, and the single-column model for the DOE E3SM. To support usage in these two models prominent within DOE circles, we recommend that code be provided for users to convert the LASSO forcing files into the alternative models' input formats for use. Provisions should be made for open source contributions from the user base for conversion into other model input formats to broaden the use of LASSO forcings within other models.

2.1.5 Use of an LES Ensemble

Recommendation M5:
An ensemble of eight LES should be used for each case day based on the forcings in M4. This would consist of VARANAL at 300 km; ECMWF at 16, 114, and 413 km; MSDA at 75, 150, and 300 km; and an additional simulation with no large-scale forcing.

As noted in Section 2.1.4, the LES results are very sensitive to the forcing selection. It is impossible to identify a priori which forcing type or scale can provide the best results for a given day, and running all of them increases the odds of achieving a simulation close to observed conditions. It is recommended that ensembles of LES be used based on multiple forcing data sets. Comparisons between microphysics reveal much less sensitivity than between forcing choices for these non-precipitating clouds; therefore, a value is not seen in generating an ensemble based on differing model physics at this point. An example of this can be seen in Figure 7 where changing the microphysics only slightly alters the results in each simulation pairing, while changing the forcing leads to larger changes in the skill score values. This is true for all days in the Alpha 2 release.

In addition to the combination of seven forcing types and scales, we recommend that an eighth ensemble member be run using no large-scale forcing. This additional ensemble member serves to identify how strongly the forcing impacts the result each day, and in some cases, performs on par with the other forcings.

The desired result of using a forcing ensemble is to encompass the true forcing for a given day. However, results from the pilot show that the available forcings do not always attain this. Therefore, additional work is needed to improve the forcing data sets used for LASSO. Increasing the number of ensemble members could improve the odds. However, new methodologies will be needed that are more likely to generate more accurate forcing. Otherwise, the ensemble will become too expensive to maintain if additional members are added that do not regularly contribute to being closer to the true state.

2.1.6 Model Output

Recommendation M6:
The LASSO data bundles should contain model output every 10 minutes consisting of 3D instantaneous snapshots combined with summary LES statistics, e.g., domain average vertical fluxes.

Which variables and how frequently to archive them is a function of balancing needs of users with the capacity of the ARM Data Center (ADC) to handle the data. There is also the issue of maintaining manageable file sizes for users when they download the data bundles. Output has been done every

10 model minutes for most of the pilot phase simulations, and this has been acceptable for users to date. This results in output of about 30 GB per simulation for WRF and 8 GB for SAM when using the netCDF4 format with internal compression. A list of variables output for the Alpha 2 release can be found in the Alpha 2 documentation (Gustafson et al. 2017). We recommend that these variable lists form the basis for output during operations with some minor modifications for variables that are currently output but that have little value (e.g., the skin temperature, which is not calculated because LASSO uses prescribed surface fluxes). The WRF output is larger due to outputting additional variables since the domain sizes are identical between models. There are also some inefficiencies in that WRF outputs base and perturbation values for some 3D variables, whereas SAM outputs a single value. The WRF output size could be reduced by mimicking a more SAM-like approach, but this would then complicate use of some WRF post-processing software, such as PyWRF. Also, additional time-averaged statistics are output in WRF, described below.

The model output is done as instantaneous values, such as 3D snapshots of the meteorological state, as well as statistical averages across the domain. Both should be output at the same frequency. They differ in that the statistics are done based on 1-minute sampling of the model over the 10-minute averaging window, while the instantaneous snapshot only represents the given output time. The statistics variables are output in different ways to suit a variety of needs. The traditional way to do LES statistics is for horizontal domain averages, which reduces 3D volumes to an average profile for the domain. In addition, column-specific statistics have been implemented in WRF that are time-averaged but do not include horizontal averaging. This permits comparison of how the averaging impacts comparisons to observations, which are often done using the ergodic assumption to swap spatial and time averages.

2.2 Data Bundle Recommendations

Data bundle recommendations relate to decisions around what observations to use within LASSO, how to package LASSO data for users, and how the users interact with LASSO datastreams.

2.2.1 Observations in the Data Bundles

Recommendation D1:
The LASSO data bundles should contain a suite of observations in a form comparable to the model output. Table 3 lists the observations by category.

Given the goal of simulating shallow convection during the first implementation of LASSO, we recommend that the data bundles include observations that can be used for evaluating the simulated boundary-layer structure and cloud characteristics. A key tenant of LASSO is reproducibility by the user, so the observations in the data bundles should be extensive enough to reproduce all the diagnostics and skill scores. This will enable users to generate comparisons using their own simulations. Observations necessary to reproduce the forcings do not need to be included in the data bundles, but should be included as separate datastreams.

The observations listed in Table 3 are those recommended to provide to users via the data bundles. These observations are essentially a subset of the observations listed in Table 6 in Section 2.3.3, where the latter includes additional observations required for the LASSO workflow, such as the satellite and conventional observations used for data assimilation. Some of the values listed are not yet possible, but would be scientifically valuable if they were to become available and implemented within the data bundles, e.g., hourly measurements of inversion strength.

Observations in the data bundles should be processed to hourly intervals for comparison with the simulations, noting that many of the observations are point or column-based measurements that must be processed to understand the regional state represented in the simulations. Thus, the data in the data bundles is rarely a simple copy of the raw observation datastreams.

The data bundles will also contain the forcing data and initial conditions used to drive the LES model. These forcings are derived partly from observations to represent the best estimate of the meteorology during the simulation period based on various assumptions and retrieval processes to merge all the available observations.

Many of the listed observations are readily available from ARM datastreams. However, others have been recently developed or modified, sometimes specifically for LASSO. An example is the multi-location lifting condensation level values, in which code has been developed within the LASSO pilot to generate the values using available observations. Table 3 categorizes the observations in terms of their readiness for inclusion in the data bundles: a value of 1 implies the observation is understood and code is available to use it, 2 implies that it is understood how to use the observation within the data bundles but additional work is necessary to test or bring the code and/or associated instrument up to a stable state, and 3 implies that the observation is aspirational, i.e., it is desired due to its scientific relevancy but work is needed either with the measurement or understanding how to use it that precludes near-term use within the data bundles. Some of these measurements are discussed further in the future development recommendations given in Section 2.4.2, particularly F4.

Table 3. Recommended observations for inclusion in LASSO data bundles. Readiness values of 1 indicate available observations implemented in the data bundles, 2 indicates partially implemented observations, and 3 implies aspirational observations. Locations are the Central Facility (CF), Intermediate Facilities (IF), Boundary Facilities (BF), Extended Facilities (EF), and the Oklahoma Mesonet (Meso).

Physical Process Category	*Hourly Observation*	*Readiness at Locations*				
		CF	*IF*	*BF*	*EF*	*Meso*
Boundary-layer state	Surface temperature	1	3	3	3	3
	Surface water vapor mixing ratio	1	3	3	3	3
	Surface relative humidity	1	3	3	3	3
	Radiosonde soundings (4x daily)	1				
	Mid-boundary-layer temperature	2			3	
	Mid-boundary-layer mixing ratio	2			3	
	Mid-boundary-layer relative humidity	2			3	
	Full boundary-layer thermodynamic profile	3				
	Lifting condensation level	1	1	1	1	1
	Planetary boundary-layer height	3	3			

Physical Process Category	*Hourly Observation*	*Readiness at Locations*				
		CF	*IF*	*BF*	*EF*	*Meso*
	Boundary-layer vertical velocity	3			3	
	Inversion strength	3				
	Inversion wind shear	3			3	
Cloud characteristics	Low-cloud fraction from ARSCL	1				
	Time-height cloud frequency from ARSCL	1				
	Cloud fraction from TSI	1				
	Regional cloud fraction from Doppler lidar	3			3	
	Liquid water path	2			2	
	Cloud-base height	1			1	
Meteorological forcing	ARM Variational Analysis with sensible and latent heat fluxes	1	Spatial scales: 300 km			
	ECMWF forcing for multiple spatial scales	1	16, 114, 413 km			
	MSDA forcing for multiple spatial scales	2&3	75, 150, 300 km			

2.2.2 Diagnostics and Skill Scores to Be Included in the Data Bundles

Recommendation D2:

The LASSO data bundles should contain a suite of diagnostics and skill scores selected to assist users in choosing appropriate simulations for their needs. Table 4 lists diagnostics and skill scores specifically targeting shallow convection.

A suite of diagnostics and skill scores has been developed during the pilot phase. The goal of the suite is to provide users with quick summaries of simulation behavior versus observations, which enables selection of simulations for particular research needs. Quicklook plots of model-observation comparisons should be accessible for viewing and skill scores should be able to be queried using the online Bundle Browser (http://www.archive.arm.gov/lassobrowser) to narrow the list of simulations users would need to consider. The recommended suite is based on evaluating conditions leading to shallow convective clouds and the resulting cloud field. The suite also is designed to use the broad range of available ARM observations suited for this purpose. A full description of the plots, diagnostics, and skill scores listed in Table 4 can be found in *Description of the LASSO Alpha 2 Release* (Gustafson et al. 2017).

Table 4. Recommended diagnostics, plots, and skill scores in the LASSO data bundles.

Category	*Diagnostic Plots and Skill Scores*
Simulation-specific quicklook plots	Thermodynamic profile plots comparing each simulation with observations Time series for each evaluation variable in Table 3 Taylor diagrams for each evaluation variable in Table 3 Scatter plots of simulated versus observed values for each evaluation variable in Table 3 Scatter plots of LWP versus TSI cloud fraction Time-height plots of cloud fraction/frequency for simulation versus ARSCL

Multi-simulation quicklook plots	Heat maps for evaluation variables in Table 3 Scatter plots of relative mean skill versus Taylor skill for LWP and cloud fraction Scatter plots of frequency bias skill versus equitable threat score skill for time-height cloud masks Scatter plots of "net skill scores" from LASSO combining multiple skill scores
Quantitative values and skill scores	Mean and RMS difference of simulation versus observation for each evaluation variable in Table 3 Taylor skill score for each evaluation variable in Table 3 Relative mean skill for each evaluation variable in Table 3 Frequency bias skill for time-height cloud masks from ARSCL Equitable threat score (ETS) for time-height cloud masks from ARSCL Net skill score for each evaluation variable in Table 3 and also the combined ARSCL frequency bias and ETS skills Multivariate 1D cloud skill from combining the net skill scores for LWP and 1D cloud fraction Multivariate total cloud skill score from combining the net skill scores for LWP, 1D cloud fraction, and 2D cloud fraction

2.2.3 File Formats

Recommendation D3:

The data bundles should use a netCDF file format. The model output should be in the standard format output by WRF. Observations and skill scores should be collected into summary files to ease intercomparison.

The data bundles consist of multiple components that include observations, model input and output, skill scores, and plots. Because of this, a variety of file formats is required. Subsetted observations and post-processed model output should be formatted into an ARM-compliant netCDF file. Model input and output files are most easily left in the format used by the LES model, since many users already have routines designed to work with these files. For WRF, the input and output files are netCDF but are not Climate and Forecast (CF) compliant. ARM could consider investing in archiving a second version of the model output that meets CF compliance, but this is neither necessary nor recommended, as it would incur a nonstandard version of WRF output inconsistent with its user community and internal documentation. (A similar situation exists for SAM, except that its input files are American standard code for information interchange [ASCII] and its output is post-processed from binary into its variant of netCDF.) Plots within the data bundles should be either portable network graphics (PNG) or encapsulated postscript (EPS).

Examples of the recommended file formats can be found in the Alpha 2 release. Files in the Alpha 2 data bundles are very similar to what is being recommended. The primary exception is the recommended addition of a few additional observations.

We recommend that the netCDF files in the data bundles use internal compression to minimize file sizes. Based on testing during the pilot, the netCDF4 compression can reduce the model output files by a factor of 3 to 5, depending on the meteorological conditions. The netCDF4 file format has been available since 2008 (http://www.unidata.ucar.edu/software/netcdf/docs/RELEASE_NOTES.html, accessed October 13, 2017) and is now routinely supported by third-party software.

2.2.4 Data Bundle Packaging

Recommendation D4

The files within each data bundle should be packaged into a series of tar files to ease download and storage within the ARM repository. One tar should contain the model configuration and information required to reproduce the simulations combined with the skill scores, diagnostics, and quicklook plots. A second tar file should contain the LES statistics from the model output. A third tar file should contain the instantaneous model snapshots, which would be much larger than the other two tar files.

The files described in the previous recommendation are what users would interact with, but storing and retrieving the data bundles from the ADC requires appropriate packaging of the files. For storage at the ADC and for easily serving the data bundles to users, the Alpha 2 release packaged the data bundle files into two tar files per bundle. A similar method is recommended for the initial LASSO implementation, but with three tar files to increase granularity, described below. Using the tar files makes the data bundle concept clearer to users by grouping associated files together during the data-ordering process. The downside is users will not be able to select specific variables or a portion of the time period within a case during the ordering process. However, this is an acceptable restriction that permits the LASSO data to be handled by the ADC.

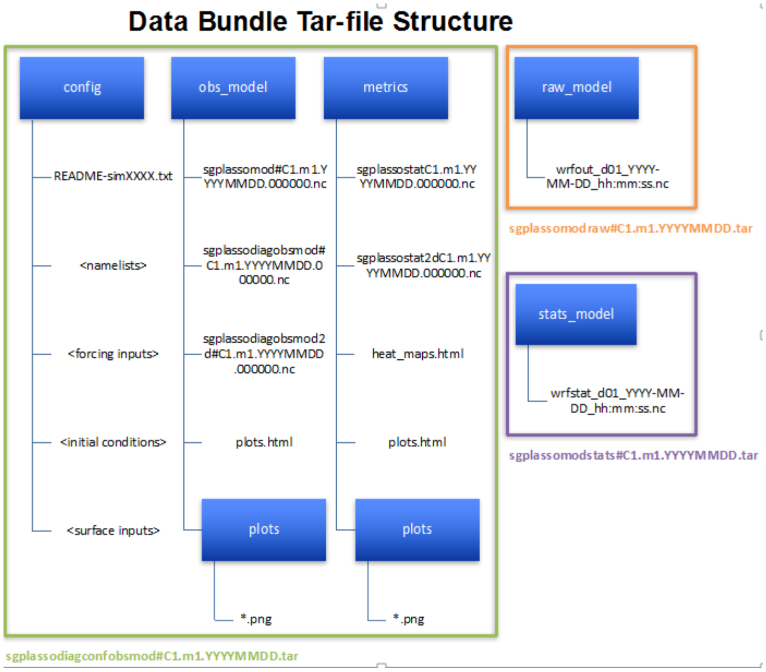


Figure 8. Data bundle tar-file structure.

Figure 8 shows the recommended structure of the files within each tar file. The smaller of the two tar files, “diagconfobsmod,” contains the model configuration and input files, the diagnostics and skill scores, the subsetted observations and model output for calculating the skill scores, and the quicklook plots for the diagnostics and skill scores specific to the particular simulation in the bundle. Its size is about 33 MB. The second tar file, “sgplassomodraw,” contains the raw model output files and thus is significantly bigger at about 22 GB for WRF. The third tar file, “sgplassomodstats,” contains the LES statistics calculated online by the model. Its size is about 7 GB for WRF. Having the three tars permits users to tailor their download selections to best meet their needs. For example, some users will not need the raw model output and could save a significant amount of time by only downloading the first tar file. Note that the provided tar file sizes are based on Alpha 2 simulations. Using a larger domain, as recommended in M2, will increase the size of the two tars containing direct model output.

2.3 Operations Recommendations

2.3.1 Case Selection

Recommendation O1:
The initial implementation of LASSO should follow the initial plan to simulate days at SGP with non-precipitating shallow convection.

We recommend that the initial implementation of LASSO follow the plan in the *ARM Decadal Vision* (U.S. Department of Energy 2014) to simulate predominately non-precipitating shallow convection at SGP. Experience from the pilot has shown a broad range of interest in simulating a wider range of conditions, but a sufficient user base is developing that justifies the initial, more minimal foray. Multiple beta users have begun interacting with LASSO for each of the three primary target audiences of observationalists, theoreticians, and model developers. The initial implementation will be valuable for developing the infrastructure necessary for ARM to routinely generate the LASSO data bundles. Plus, the experience gained during the development process will inform what will be possible for expanding LASSO beyond shallow convection at SGP.

During the pilot, the particular days to be simulated were manually selected by the PI and Co-PI based on a range of criteria. This process resulted in five case days from 2015 and 13 case days from 2016. A climatological analysis of shallow convection at SGP using the newly developed cloud classification value-added product (VAP) (cldtype) (Flynn et al. 2017) identifies that the number of cases in a year can vary from a couple to several dozen based on the criteria used to define shallow convective days. Thus, for years with abundant days to choose from, we recommend that a prioritized list be made with the most classic days receiving the highest priority followed by the more complicated days. Unless the resources (both computational and personnel) become exhausted, it should be possible to simulate all candidate days for a given year. For years with only a small number of shallow convection days, it is recommended that the selection criteria be relaxed, resulting in a broader range of conditions being simulated, with at least a dozen cases targeted per year.

The initial criteria for selecting cases, prior to making adjustments for too many or too few cases in a given year, are provided in Table 5.

Table 5. Criteria for identifying shallow convective days.

Number	Criteria	Potential Relaxation?
1	Cloud classification VAP identifies shallow convection during the day.	This VAP is a first guess and can be spoofed by the limited field of view of ARSCL. It sometimes misidentifies the cloud conditions, and thus has been designed to use a liberal interpretation of shallow convection to minimize excluded days when used as a first-brush selector criterion.
2	TSI movies and opaque cloud fraction confirm shallow convection.	The range of cloud fraction will ideally be 20 to 70%, but the number of hours throughout the day within this range can be variable.
3	No widespread precipitation during the day.	Shallow convection at the SGP generally does not precipitate. However, days that transition from shallow to deep convection can be simulated if computing resources permit. These days may not perform as well due to the forcing and domain configuration choices.
4	The mode of in-cloud liquid water path is >40 g m^{-2}.	Conditions leading to very thin clouds are difficult for the forcings to reproduce and the results are not expected to be as good for very small LWP values. This criterion could be relaxed if resources permit.
5	Shallow convection is present for at least 3 h.	Generally, clouds present less than this threshold do not justify simulating the case.
6	The cloud-base height generally follows the lifting condensation level (LCL).	Buoyancy-driven shallow convection should have the cloud-base height and LCL aligned.
7	Clouds at SGP are regionally representative.	The large-scale forcings represent a regional average. Thus, the cloud field within the forcing region should be relatively homogeneous. Homogeneity over larger regions should be given higher priority.

2.3.2 Computer Requirements

Recommendation O2:

A minimum of 10.5 million core hours per year should be devoted to running the MSDA forcing an LES for typical shallow convection days. Additional resources should be devoted for ongoing LASSO development and user interaction with the data bundles.

The computational cost to generate a set of simulations and the accompanying data bundles depends heavily on a range of choices described elsewhere in this document, such as the domain size, resolution, number of ensemble members, and number of days simulated per year. Using an assumption of a domain with 25 km extent and 100 m grid spacing, the cost is approximately 2,000 core hours per simulated model hour with the WRF model. Assuming 30 case dates per year (roughly the climatological average), 15 simulated hours per simulation, and 8 ensemble members, the resulting cost to run the LES model is approximately 7.2 million core hours per year. For reference, this is approximately 20% of the current capacity of ARM's Cumulus cluster. Additional resources will be needed to run the MSDA, which requires about 2,250 core hours per case, but that cost is much lower than the LES. LASSO will also need computing resources for model development throughout the year, e.g., testing model updates, adding new capabilities, and possibly needing to reprocess cases. The recommendations in this report leave room to expand LASSO to include additional locations and meteorological regimes while still remaining within the currently available computing capacity.

It should be noted that the computing needs for LASSO will be nonlinear throughout the year because shallow convection is more prominent during spring and summer at SGP. This will result in a greater usage of Cumulus during this surge of cases. As described in Section 2.3.4, the time lag between cases occurring and the LES being run will be at least 3 to 4 months, so the heaviest computer usage will occur between August and December. If Cumulus becomes overloaded during this period, it could slow LASSO production. Therefore, external usage of Cumulus during heavy LASSO usage may need to be given a lower priority.

In addition to production work, it is desirable to reserve a portion of the resource for user access so users can work with the large amounts of LASSO and ARM data without having to download it to non-ARM computers, as described in Section 2.3.8. Some users will lack sufficient resources to download large numbers of simulations and, even if they did, such downloads might strain ADC resources; thus, users would not get the full benefit of the LASSO library unless they could use it via ARM's computers. As the initial priority is to get LASSO running and producing scientifically useful simulations, user access should be permitted during the first year of operations, but should be given a lower queue priority until more is known about how much time is needed for LASSO production work. Once LASSO is established, it will become clearer what portion of the currently available 4,032 cores on Cumulus can be allocated to research-based users, and additional nodes purchased if too few resources would be available.

2.3.3 ARM Measurements and External Data Needs

Recommendation O3:
The observations listed in Table 6 are required for running the LES and producing the data bundles. These observations include ARM cloud and meteorological measurements combined with data acquired from external sources, e.g., satellite data for data assimilation purposes.

Generating LASSO data bundles requires a large number of input data sets with a wide range of complexities. Many of the data sets are routine ARM measurements, some are specialized products that are not yet routinely produced, and others will need to be acquired from external sources. External data are needed for data assimilation with the MSDA hindcasts. Table 6 lists the various data sets currently used within the Alpha 2 release. Additional data that could be considered are discussed in Section 2.4.2.

Table 6. Observations currently required for producing LASSO data bundles.

Observation Category	*Observation*
Cloud measurements	Cloud classification VAP and shallow convection VAP for case selection Ka-band ARM Zenith-Pointing Radar–ARSCL (KAZRARSCL) cloud mask TSI cloud fraction Optimal estimation AERI (AERIoe) liquid water path (either with v2.4 for the full LWP range by including microwave radiometer (MWR) data, or with v2.2 for LWP values <40 g m-2 as a hybrid product with MWRRet v1) Microwave radiometer retrieval (MWRRet) for evaluating AERIoe v2.4 LWP for values >30 g m-2 Doppler lidar for cloud-base height
Meteorology measurements	Radiosonde soundings for thermodynamics and wind Raman lidar thermodynamic profiles AERIoe thermodynamic profiles ARM surface meteorology stations Oklahoma Mesonet Surface sensible and latent heat fluxes from Energy Balance Bowen Ratio System (EBBR) and Eddy Correlation Flux Measurement System (ECOR) (as part of VARANAL) Radar wind profiler for winds Variational analysis for averaged surface fluxes and forcing profiles
External data	ECMWF forcing inputs Geostationary Operational Environmental Satellite (GOES) visible satellite images and loops GOES top-of-atmosphere radiances as inputs to VARANAL RAP (for AERIoe and VARANAL) Final Operational Global Analysis (FNL) gridded data for data assimilation Conventional meteorological observations for data assimilation Satellite radiances for data assimilation GPS radio occultations for data assimilation Next-Generation Weather Radar (NEXRAD) network for data assimilation

2.3.4 Production Timeline

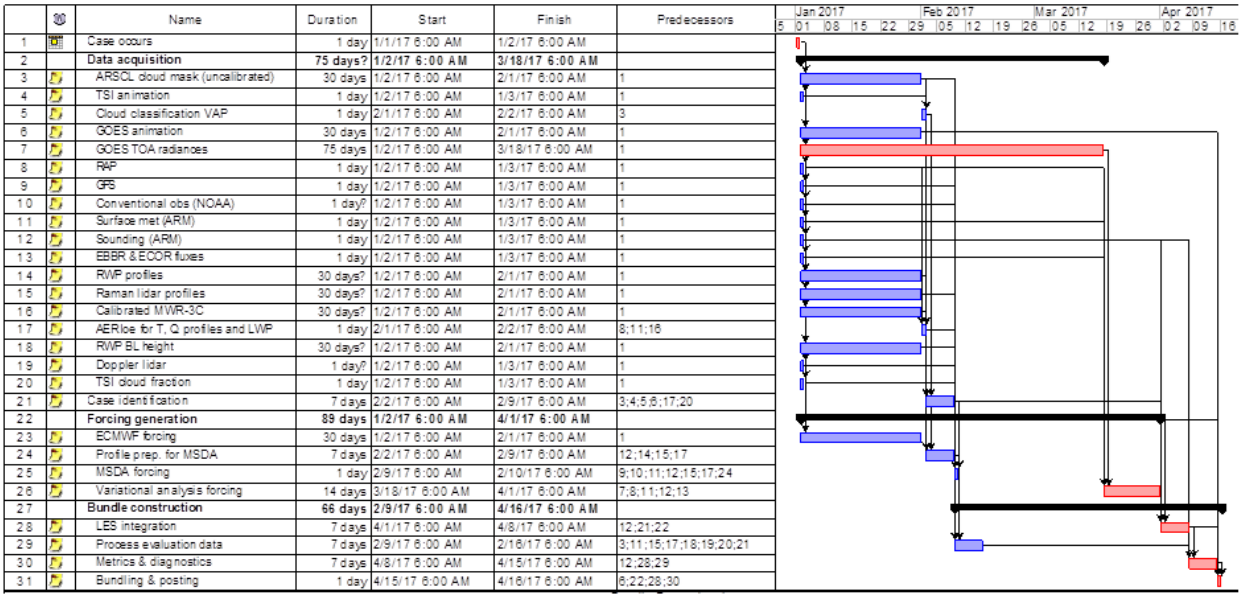
Recommendation O4:

Data requirements and simulation run times indicate that users will need to wait a minimum of 3.5 months after an occurrence of shallow convection before the associated data bundle will be available.

The complexity of the input datastreams precludes running LASSO in real time. Based on estimating the time necessary to procure the required observations, processing them to generate forcings, running a LES ensemble, and post-processing the results, the minimum, best-case scenario time required to generate LASSO data bundles will be about 3.5 months.

The Gantt chart in Figure 9 shows the dependencies and time needed for each step of data-bundle generation. The critical path, shown in red, is dominated by obtaining the Geostationary Operational Environmental Satellite (GOES) top-of-atmosphere radiances for the VARANAL forcing data set. As currently formulated, calculation of the radiances requires waiting for the Modern-Era Retrospective Analysis for Research and Applications (MERRA) reanalysis product to be generated and thus cannot be sped up unless the constraint methodology is modified. The timeline would be halved to about 2 months without the dependency from the GOES top-of-atmosphere radiances.

The estimated 3.5 month timeline is the minimum time to produce a data bundle and the expected production time likely will be longer. Practically, we recommend that production be done in batches to optimize labor usage for case selection, data acquisition, and simulation processing. Multiple steps in the process require manual intervention, precluding full automation of LASSO production. A reasonable approach would be to establish periodic points when evaluation of potential cases is done and the associated production work is done. This should be either monthly or quarterly, depending on how quickly the data bundles are deemed necessary.



		Name	Duration	Start	Finish	Predecessors
1		Case occurs	1 day	1/1/17 6:00 AM	1/2/17 6:00 AM	
2		**Data acquisition**	**75 days?**	**1/2/17 6:00 AM**	**3/18/17 6:00 AM**	
3		ARSCL cloud mask (uncalibrated)	30 days	1/2/17 6:00 AM	2/1/17 6:00 AM	1
4		TSI animation	1 day	1/2/17 6:00 AM	1/3/17 6:00 AM	1
5		Cloud classification VAP	1 day	2/1/17 6:00 AM	2/2/17 6:00 AM	3
6		GOES animation	30 days	1/2/17 6:00 AM	2/1/17 6:00 AM	1
7		GOES TOA radiances	75 days	1/2/17 6:00 AM	3/18/17 6:00 AM	1
8		RAP	1 day	1/2/17 6:00 AM	1/3/17 6:00 AM	1
9		GFS	1 day	1/2/17 6:00 AM	1/3/17 6:00 AM	1
10		Conventional obs (NOAA)	1 day?	1/2/17 6:00 AM	1/3/17 6:00 AM	1
11		Surface met (ARM)	1 day	1/2/17 6:00 AM	1/3/17 6:00 AM	1
12		Sounding (ARM)	1 day	1/2/17 6:00 AM	1/3/17 6:00 AM	1
13		EBBR & ECOR fluxes	1 day	1/2/17 6:00 AM	1/3/17 6:00 AM	1
14		RWP profiles	30 days?	1/2/17 6:00 AM	2/1/17 6:00 AM	1
15		Raman lidar profiles	30 days?	1/2/17 6:00 AM	2/1/17 6:00 AM	1
16		Calibrated MWR-3C	30 days?	1/2/17 6:00 AM	2/1/17 6:00 AM	1
17		AERIoe for T, Q profiles and LWP	1 day	2/1/17 6:00 AM	2/2/17 6:00 AM	8;11;16
18		RWP BL height	30 days?	1/2/17 6:00 AM	2/1/17 6:00 AM	1
19		Doppler lidar	1 day?	1/2/17 6:00 AM	1/3/17 6:00 AM	1
20		TSI cloud fraction	1 day	1/2/17 6:00 AM	1/3/17 6:00 AM	1
21		Case identification	7 days	2/2/17 6:00 AM	2/9/17 6:00 AM	3;4;5;6;17;20
22		**Forcing generation**	**89 days**	**1/2/17 6:00 AM**	**4/1/17 6:00 AM**	
23		ECMWF forcing	30 days	1/2/17 6:00 AM	2/1/17 6:00 AM	1
24		Profile prep. for MSDA	7 days	2/2/17 6:00 AM	2/9/17 6:00 AM	12;14;15;17
25		MSDA forcing	1 day	2/9/17 6:00 AM	2/10/17 6:00 AM	9;10;11;12;15;17;24
26		Variational analysis forcing	14 days	3/18/17 6:00 AM	4/1/17 6:00 AM	7;8;11;12;13
27		**Bundle construction**	**66 days**	**2/9/17 6:00 AM**	**4/16/17 6:00 AM**	
28		LES integration	7 days	4/1/17 6:00 AM	4/8/17 6:00 AM	12;21;22
29		Process evaluation data	7 days	2/9/17 6:00 AM	2/16/17 6:00 AM	3;11;15;17;18;19;20;21
30		Metrics & diagnostics	7 days	4/8/17 6:00 AM	4/15/17 6:00 AM	12;28;29
31		Bundling & posting	1 day	4/15/17 6:00 AM	4/16/17 6:00 AM	6;22;28;30

Figure 9. Gantt chart showing a minimum LASSO production timeline.

2.3.5 User Support

Recommendation O5:
User support is of primary importance for building a community around the LASSO product and sustaining support for its generation. ARM should budget sufficient labor to support user interactions during the first year and expect the need to increase as the user base increases.

LASSO generates a complex data product that may require some guidance until users are educated. Users are therefore encouraged to contact the LASSO team to discuss how the product is used and to ensure that the subtleties of the data are properly accounted for. During the pilot, many of the known users required several hours each of interaction with the LASSO team. Going forward, ARM should allocate sufficient labor for user support at the scientific level. This needs to be spread across multiple people, with the LASSO lead acting to triage questions and redirect them to other team members to cover the wide range of details from observations to modeling. After the first year, the level of support can be reevaluated to determine an appropriate amount. We anticipate that user support needs will grow as LASSO datastreams become increasingly known and used. The types of questions asked should also be tracked and a frequently asked questions web page be developed to cover common questions.

In addition to the scientific support, additional time will be needed to support users on the compute cluster. The level of effort needed for this will depend on how many users are permitted research time. User support will also be needed for assisting users in using the LASSO software. During the first year, this may not be a lot of time since the software will not yet be formally released. However, this will need to be taken into account for subsequent years.

2.3.6 LASSO Software

Recommendation O6:
A priority should be placed on first developing robust code for operations, which should be done with the external reproducibility in mind when writing modular workflow software. Software necessary to reproduce the LES results, diagnostics, and skill scores should ultimately be made available via a publically facing repository with an open source copyright.

LASSO consists of a series of workflow components, each with semi-independent software stacks. One approach to organizing the code for development units would be as follows:

- Case selection and an associated dashboard
- MSDA workflow and model code
- LES workflow and model code
- Data bundle workflow and post-processing code.

All of these code groups will share the need for handling data transfers (to a lesser amount for case selection) and thus could have shared libraries for this purpose. This would be for staging data from the ADC to where the calculations are done. The internal data transfers are a high priority and will be needed during the first year of LASSO when developing the overall workflow software. The Cumulus cluster resides in a different “security enclave” at Oak Ridge National Laboratory, which requires specialized scripting to transfer data from the primary ADC archive to where the LASSO computing will be done. The data transfer software for external users can be developed at a later time as funding permits. However, this future development should be considered when developing the internal code so that choke points are not built into the workflow software that would prevent its use outside of ARM.

A “dashboard” should be developed to assist in LASSO case selection and tracking of the production cycle. This dashboard should:

- Capture relevant observation quicklooks needed to determine if a given day is to be simulated,

- Show the status of required input datastreams so the LASSO team can see which are ready for use and if any need follow-up intervention, and
- Contain checklists to indicate where in the production process each case lies.

The MSDA hindcasts should be developed as an independent datastream that is then used within the LASSO data bundles. The primary software for MSDA consists of the WRF model combined with the GSI data assimilation program. Additional software includes data management tools and code to convert 3D WRF output into forcing files similar to the format of VARANAL.

The LES code consists of the chosen LES model combined with data management and automation software for simplifying case setup. If WRF becomes the chosen model, ARM should maintain its version of WRF as a fork from the National Center for Atmospheric Research (NCAR) WRF repository on Github, https://github.com/NCAR/WRFV3, which would simplify code updates, copyright issues, and user support.

The data bundle software developed during the pilot consists of analysis and plotting software, primarily written in Integrated Data Language (IDL), with code to package the results written in Python. Data staging for this portion of the workflow will need to be carefully examined to determine what can be automated and what will require ongoing manual intervention. This will be a function of the status of specialized data sets combined with current approaches for handling quality-control issues. The large number of datastreams needed for LASSO would mean that implementing the prototype using the ARM Data Integrator (ADI) would involve a number of separate processes that represent processing of many observational datastreams, processing of simulations, production of output that includes observations and simulations, and production of metrics. As such, a full ADI undertaking would be similar to, or larger than, automation of VARANAL and the ARM Best Estimate datastreams. It is also possible that ADI could have trouble reading in a non-ARM-standard file like the WRF output, which might mean that part of the code is not put into ADI. Given these factors, we recommend examining the cost-benefit of converting the different components to ADI before proceeding.

Since the goal is user reproducibility of LASSO results, we recommend that the LASSO software be released using an open-source copyright. This will enable user improvements to the code and the possibility of building a user repository of tools for working with LASSO data. Easing user interaction with LASSO data through code sharing should improve overall user experiences by shortening their spin-up time and saving them time from developing tools others have already shared via the repository. The one possible exception to the copyright choice would be the WRF model, which is owned by NCAR, and thus would need to follow NCAR's copyright requirements.

2.3.7 Web-Based User Interactions

Recommendation O7:
An interactive web-based tool should be provided to users to query data bundles based on simulation metadata and skill scores that compare the LESs to observations.

Over time, LASSO will provide a growing library of cases spread over multiple days with an ensemble of simulations available for each day. During the pilot, this has resulted in hundreds of simulations per year available to users. This will continue to grow during routine operations. Thus, a method is needed for users to distill the myriad simulations to those relevant to their research needs. The pilot project developed a prototype, web-based, visualization and data searching tool to meet these needs, which is called the LASSO Bundle Browser, http://www.archive.arm.gov/lassobrowser. The browser allows selection of simulations by date, metadata tags, and skill score values. Plots are then presented showing a selection of skill scores and model values versus observations. Importantly, users can make download requests for the relevant data bundles directly from within the browser. We recommend adopting the Bundle Browser, or an equivalent approach, as part of LASSO moving forward. User input will be needed to shape new features in the future, such as a user being able to search on case summary data values.

2.3.8 ARM Computing for LASSO Users

Recommendation O8:

ARM should provide user access to an ARM computer(s) to ease data analysis of the very large data sets. A portion of the computing should also be made available for users to do additional sensitivity simulations to compare with LASSO simulations. However, this should be carefully tracked and large requests should be diverted to traditional DOE computing facilities, e.g., the National Energy Research Scientific Computing Facility (NERSC).

Researchers requiring large numbers of 3D simulation output could be hampered by the data volume of LASSO data bundles. ARM can assuage this issue by providing computing resources within the ADC where data can be staged on an ARM server without having to be moved offsite. Users should be provided sufficient resources to do typical analysis tasks of the data bundles. During the first year of operations, the resource usage required to produce the LASSO data bundles should be tracked to determine how much of the Cumulus cluster can be made available to users. A good starting point is to make 30% of the computing time available, with the understanding that operational needs take priority and users may not be able to use all of their allocation.

Some users will desire the ability to generate their own simulations to compare with the LASSO LES output, e.g., this could be useful for parameterization development. Initially, this usage should not be discouraged on Cumulus as there should be a place for “sandbox” activities. However, the computing time should be tracked to determine how much to allocate in subsequent years. If multiple large requests are made, then users will need to be deferred to other DOE computing facilities, such as NERSC, which are specifically designed for providing production computing capacity. If many users need it, agreements could be pursued with NERSC to stage a copy of some or all of the LASSO library on their tape storage system, which would ease use by reducing the time for data transfers. This staging of data should be a low priority during the first year(s) of LASSO, and could be pursued as need dictates.

Because LASSO only simulates shallow convection in its current configuration, there will be certain times of the year when operations will need a lower percentage of Cumulus. External users could be encouraged to do their computationally expensive work during the less busy times of year. Basic analyses that only require a small number of cores can be done year round.

2.3.9 LASSO Maintenance and Development

Recommendation O9:
ARM should budget sufficient resources to permit ongoing development of LASSO. This will entail maintaining the model code, adding LASSO features as new observations become available, enhancing the model output, diagnostics, and skill scores to meet new user scenarios, improving the data assimilation, etc.

LASSO should be viewed as an evolving product that adapts to community needs, newly available observations, and modeling improvements. LASSO brings with it an ongoing obligation for maintaining it and keeping it up to date. As explained above, we recommend implementation of a typical LES model configuration for shallow convection alongside a data assimilation system for incorporating ARM observations into the LES forcings. A level of effort needs to be planned for maintaining the LASSO software stacks. For example, the LES model will have annual updates to improve the model. Some of the changes will be bug patches that should be straightforward to implement. Other changes could be new features, such as the hybrid vertical coordinate system introduced in WRF v3.9. Large changes, such as this, require extensive testing before use in an operational setting. Staff time will be needed to do this testing, which could involve duplicating a series of data bundle results to document and understand differences between LASSO versions.

Whether or not previously generated data bundles should be re-generated and re-released due to code changes will need to be determined annually based upon how the modifications impact the results. As the library grows, it will become increasingly expensive to do this sort of reprocessing, and at some point, will become impossible. Metadata is included in the data bundles indicating the model version used and users will need to be made cognizant of potential changes in model behavior due to changing model versions.

Efforts have been made to best use available ARM data when building the LASSO data bundles, but the LASSO pilot has been the first user of many of the new boundary facility instruments and has been a beta user of new remote retrieval methodologies to obtain the data necessary for the initial implementation. Because of this, many data issues have been discovered and questions remain as to how to address some of the issues. We recommend that a substantial effort be maintained beyond basic LASSO operations to continue testing and improving the use of ARM data and MSDA data assimilation methodology with ARM data. Also, new observations, such as photogrammetric-based 3D cloud masks, are being developed that could benefit LASSO. Resources should be made available to incorporate these new observations into the data bundle skill scores when the product is ready.

Additional detail and specific needs are listed in Section 2.4, Further Development Recommendations.

2.4 Further Development Recommendations

The LASSO pilot has developed a robust suite of observations, evaluation and discovery tools, and model scenarios for the initial implementation. This critical mass permits moving forward with the first round of implementation. However, many additional areas of improvements can be made that will result in a better product. The following recommendations for future development comprise a combination of high- and lower-priority recommendations, with the high-priority recommendations specifically called out with formal numbers. Additional recommendations are described in the text. These recommendations are within the context of shallow convection at SGP and do not include topics related to expansion of LASSO, since the latter are outside the scope of this report and will be the topic of the expansion planning in Section 1.4.

2.4.1 Further Developments for Modeling

Recommendation F1:
A nested LES domain configuration should be tested at SGP to better capture spatial heterogeneity and as a preliminary step toward simulating alternate cloud types and ARM sites where the periodic configuration could be limiting.

Recommendation F2:
Further development of the MSDA methodology should be pursued. Particularly, it should be moved from a 3D variational analysis (3DVar) to an ensemble hybrid-Kalman filter methodology. Also, additional improvements could be achieved with further fine tuning of the MSDA grid and use of observations.

Recommendation F3:
A blended forcing product could be produced combining the gridded MSDA approach that incorporates ARM profile observations with the VARANAL approach that best utilizes ARM flux measurements.

Three high-priority recommendations are made for advancing the modeling aspect of LASSO. The first, Recommendation F1, is to test an additional one or more ensemble members using a nested LES approach. This would most likely involve nesting the LES domain directly inside the grid used to generate the MSDA large-scale forcing or possibly the 3D VARANAL when it is available. This would permit time-dependent and spatially varying lateral boundaries to better capture variable conditions around SGP, such as for roll clouds and rapidly changing synoptic situations. This would increase the usability of LASSO where the horizontally homogeneous statistics implied by the current periodic lateral boundaries are a limitation. For example, the typical west-east gradient observed in cloud properties cannot be captured with the current LASSO approach.

The nested approach was not attempted within the LASSO pilot because two years ago, when the pilot began, this approach was in its infancy and unproven. Since then, Heng Xiao has worked on the technique via a DOE ASR project and has shown it works well in many situations for shallow convection. While implementing the LES configuration would be somewhat straightforward for nesting, interpreting the results in the context of the doubly periodic LES results will require more thought. For example, variable terrain altitude within a nested LES will make spatial averaging more difficult. The transition region around the domain edge where turbulence spins up at the resolved scale must also be excluded from the statistics, yet this region is somewhat amorphous. There is a benefit of testing nesting at SGP in a side-by-side comparison with doubly periodic boundaries for shallow convection. This will inform the ability to determine what types of conditions can be attempted in future LASSO versions.

Another potential advantage of the nested approach is the ability to use an interactive land model. While properly initializing the land is an open question at LES scales, a first approach would be to use the coarser soil information from the MSDA simulation or from the High-Resolution Rapid Refresh (HRRR) forecast model. Having the soil surface react to the presence of overlying clouds would bring an additional layer of realism to the simulations, which could open research areas in land-atmosphere interactions.

The second future modeling Recommendation F2 is to pursue improvement of the MSDA approach used within LASSO. When the pilot was proposed, the hybrid ensemble-Kalman (hybrid-EnKF) filter method was still in the research state and not regularly used operationally. Thus, it was not used within LASSO. Since then, the GSI system, used for WRF data assimilation, has been updated to include the ability to use the hybrid-EnKF method. This feature is now routinely used operationally as part of the RAP and HRRR forecast system, where it has resulted in improved forecasts (Benjamin et al. 2016). Thus, it is likely to improve the MSDA results. We recommend development so that MSDA can use the hybrid-EnKF approach, which would then be tested for large-scale forcing generation to see if it surpasses the current 3DVar approach when incorporating ARM observations.

Additional improvements with MSDA should also be pursued to fine tune the MSDA grid. Basic tests have been performed during the pilot, but extensive testing has not been done due to the lack of ARM-based retrieved thermodynamic profiles until near the end of the pilot.

Recommendation F3 builds upon improving MSDA by recommending a research investment to blend the MSDA and VARANAL methodologies into a new forcing product. Each methodology uses a different approach for blending observations to develop an estimate of the most likely environmental conditions. Because of their differences, each method optimally uses different types of observations. MSDA is best able to incorporate profile-based observations, such as wind, temperature, and water vapor. However, MSDA cannot directly use surface fluxes. In contrast, VARANAL uses surface fluxes to close the overall moisture and energy budgets and thus is well positioned to better use these data than MSDA. An opportunity exists to blend the two methods because VARANAL requires a background field to provide an initial estimate of the meteorology. Currently, the RAP model is used for this purpose, but another 3D model could be substituted. We propose to use the MSDA output with 2 km grid spacing, which is constrained by the ARM thermodynamic profiles and other relevant observations, as the first-guess background field for VARANAL, which would then take the next step of refining the environmental estimate using the flux data. Developing this hybrid forcing product will require testing optimized vertical grid spacings so that, ideally, the output from MSDA has similar vertical levels as VARANAL. The MSDA grid may also need enlargement to aid the blending process. When the MSDA fine tuning is done for Recommendation F2, the needs of F3 should be considered simultaneously.

2.4.2 Further Development for Observations

Recommendation F4:
The pilot project had resources to design the data bundles around high-priority variables. Additional resources should be devoted to more fully include available observations, e.g., Doppler lidar profiles and photogrammetric cloud masks.

Recommendation F5:

LASSO data assimilation would benefit greatly from hourly thermodynamic and wind measurements near the top of the boundary layer at multiple locations in the region. In particular, high-vertical-resolution observations of the inversion are needed.

Recommendation F6:

Complete implementation of an Active Remote Sensing of Clouds (ARSCL) simulator based on the Cloud Resolving Model (CRM) Radar Simulator (CR-SIM) and use this to replace the model cloud masks when comparing with ARSCL in the data bundles.

Recommendation F7:

A quality control (QC) protocol is needed for the data bundle generation.

The alpha releases during the pilot demonstrate many of the possible variables that could be included in the LASSO data bundles. High-priority variables were included that provide a first-order evaluation of the simulated shallow convection and some of the environmental conditions impacting the clouds formation. Recommendation F4 is for additional resources to be allocated for including additional observations based on community interest and the readiness of the associated products. Described below are three products on the horizon that are highly applicable to shallow convection. However, these are not the only observations requiring additional development resources. Observations with readiness levels of 2 and 3 in Table 3 also need development. An example is designing an appropriate methodology for comparing the regional surface meteorological measurements with the LES. A robust method for normalizing terrain height differences needs to be incorporated to make the values comparable. For comparison, another example is the need for improved calibration of the 3-channel microwave radiometers to enable multi-location liquid water path measurements for a regional estimate instead of the single measurement from the Central Facility.

The first high-priority need is measurements of planetary boundary-layer top from the radar wind profilers (RWPs) located at the Central Facility and 15 km away at three intermediate facilities. Preliminary work is encouraging but more effort is needed to extract reliable boundary-layer heights from the RWPs and develop robust code to extract a comparable estimate routinely from LES output.

The second is vertical velocity statistics and fluxes from Doppler lidar measurements. This is now a regularly produced ARM data product and a recent paper by Berg et al. (2017) demonstrates the type of vertical velocity statistics that can be measured with the Doppler lidar and that are important for shallow clouds development. Minor adjustments would be necessary to adapt their methodology for use in shallow cloud conditions. In addition, combining the Doppler lidar with a co-located Raman lidar could extend the approach to potentially retrieve profiles of moisture fluxes.

A third upcoming capability is 3D cloud mask estimates based on photogrammetry (Öktem et al. 2014; Romps and Öktem 2015). ARM recently installed multiple camera pairs at the SGP Central Facility to apply this technique and the CMDV Coupling Mechanistically the Convective Motions and Cloud Macrophysics in a Climate Model (CMDV-CM4) project is working to develop and test the approach at SGP. At this point it is premature to implement the photogrammetry into LASSO. However, in a year or two, when the product is routinely produced, it should be incorporated into the data bundles. A simple approach would be to blindly compare the 3D observed cloud mask with an LES domain-averaged cloud fraction profile. A better approach would be to develop an instrument simulator that applies the photogrammetry within the LES grid to more closely mimic the retrieval methodology.

Regarding Recommendation F5, LASSO data assimilation could potentially benefit greatly from hourly thermodynamic and wind measurements near the top of the boundary layer at multiple locations in the region. Winds are available from the RWPs but they must undergo additional processing to address clutter and spurious values that are not addressed during the initial data processing. The high-vertical-resolution thermodynamic observations of the inversion are currently only available at the Central Facility through Raman lidar observations. New approaches are needed to provide such values at sites distributed across the domain. Whether or not these would improve the MSDA forcing could be examined by running sensitivity tests with artificial data injected into the MSDA input data as well as by using high-frequency radiosonde releases during a field campaign.

Regarding Recommendation F6, the first published application using LASSO data is an analysis of cloud radar scan strategies to determine the optimal scan strategy for identifying shallow-cloud fraction (Oue et al. 2016). This work demonstrates the value of the 3D LASSO model output for evaluating retrieval methodologies. In the context of improving LASSO, this work also demonstrates the value of using radar simulators within LASSO to compare the model output to ARM's cloud radar measurements. The pilot included an initial effort along these lines, but it was discovered that the Cloud Resolving Model Radar Simulator (CR-SIM) (Tatarevic et al. 2017; http://radarscience.weebly.com/radar-simulators.html) was not yet computationally ready for efficient use. In the last year, the Brookhaven National Laboratory (BNL) Computational Science Laboratory optimized the code and it is now fast enough to routinely use with LASSO. We recommend that CR-SIM be further developed to reproduce the ARSCL retrieval methodology, and the resulting software be developed to convert the LES model output into a directly comparable ARSCL-like cloud mask. This will reduce the uncertainty surrounding interpretation of the model results when comparing with the observation-based ARSCL product.

A QC protocol, Recommendation F7, is needed for the data bundles. This involves adequate QC before data are passed into the LASSO workflow (not that the workflow acts as QC), as well as after the diagnostics and skill scores are generated. Diagnostics are generated for many variables multiplied by the number of ensemble members and number of days, yielding a large number of values and plots that must be checked for irregularities.

2.4.3 Further Development for Data Discovery

Recommendation F8:

Usage of the Bundle Browser should be closely tracked and resources allocated to continually improve the Bundle Browser to further adapt it to user needs.

Recommendation F9:

A methodology must be developed for communicating data provenance within the data bundles that aligns with ARM procedures and is easy for users to understand.

The LASSO Bundle Browser serves as the first impression of LASSO data bundles for many users. Thus, it is important that it be easy to use and have the features expected by users. One of the big improvements in the Bundle Browser between Alpha 1 and Alpha 2 is the added feature of being able to compare simulations from multiple days. This empowers users to search across the entire LASSO library of case dates to identify simulations meeting selected criteria. An addition recently raised would be for a user to be able to search on the values of key variables. It is anticipated that users will desire more complicated search criteria as the number of case days within the library grows, and more complex querying capabilities in the Bundle Browser would enable users to target their selections based on their particular needs instead of only being able to search based on the pre-calculated metrics. However, we caution loading too many features into the Bundle Browser to avoid making the interface overly complicated, resulting in confused and discouraged novice users. As capabilities are added, a “power interface” tab may be added that separates out the more complex search functions to keep the initial interface as simple as possible for new users. Usage of the Bundle Browser should be closely tracked during the next year and users should be polled to determine what new features would most benefit them. Ensuring a good Bundle Browser experience will go a long way in building a good reputation for LASSO.

The last Recommendation F9 is to develop a method to communicate data provenance within the data bundles so that users understand what they receive in the bundles. Provenance tracking has not been pursued within the pilot due to the many complexities involved in designing the bundles. Now that the bundle format has been more or less determined, effort should be invested in communicating the provenance of the bundles. This is non-trivial due to the multiple input datastreams, model configuration, and forcing choice. Each of these can evolve over time, resulting in changes to the bundle contents if it were to be reprocessed.

At a minimum, the provenance information should include the version and processing date of each observation, the model version and associated repository hash value, and all the associated configuration choices for each bundle. This should be easily accessible to users and it needs to be conveyed in a way they can easily digest. Importantly, the provenance tracking also needs to be consistent with methodologies that span ARM products. Thus, this is a discussion to be had that influences much of ARM and will be an evolving issue.

On a related note, if any of the underlying data used to generate the bundle changes, choices will also need to be made regarding when the bundle should be reprocessed to make it consistent with the most up-to-date version of the data available in the archive.

3.0 Closing

This report outlines a series of recommendations for the initial implementation of the LASSO workflow and associated datastreams for MSDA forcing, ECMWF forcing, and the LASSO data bundles. Following through on these recommendations will result in a valuable new capability for ARM combining LES modeling with observations to open new avenues of research for atmospheric scientists.

Many of the recommendations can be implemented during the first year of operations, while others will require more time to come to fruition. There are also many open areas of research and development that could be pursued to improve LASSO, but which are not required for the initial implementation. Priority should be given to implementing a semi-stable datastream for users to begin using over the near term. By finalizing file structures and contents early in the process, users will be insulated from backhouse development that will evolve. Many of the processes during the first operational year will need to be done manually, such as staging data, extensive quality control, and kicking off the hundreds of LES simulations. However, many of these tasks will become automated over time, which will ultimately reduce the cost of generating LASSO data bundles. The pilot can be viewed as a model for this process. The two alpha releases have evolved the bundle structure from loose files in Alpha 1 to a format using two tar files in Alpha 2, with the ultimate recommendation now being to use three tar files for operations. Additional variables have also been added in Alpha 2 as they have become available, e.g., cloud-base height from the new Doppler lidars at the boundary facilities. Now, a viable format is available for operations. Many of the prototype software tools used for the pilot can also be used to begin operations in 2018. These will be used as the basis for modular elements to be connected by automation scripts.

In closing, the LASSO team thanks the many people that have helped to make the LASSO pilot a success:

LASSO Pilot Team
William I. Gustafson Jr., PI, PNNL
Andrew M. Vogelmann, Co-PI, BNL
Xiaoping Cheng, UCLA/National U. of Defense Technology, China
Satoshi Endo, BNL
Zhijin Li, UCLA/JPL
Tami Toto, BNL
Heng Xiao, PNNL

ARM Atmospheric Modeling Advisory Group
Maike Ahlgrimm, ECMWF
Christopher Bretherton, U. Washington
Jennifer Comstock, PNNL
Graham Feingold, NOAA ESRL
Chris Golaz, LLNL
William I. Gustafson Jr., chair, PNNL
James H. Mather, PNNL
David D. Turner, NOAA ESRL
Andrew M. Vogelmann, BNL
Minghua Zhang, Stony Brook U.

Data Product Generation
Maike Ahlgrimm, ECMWF
Maria Cadeddu, ANL
Donna Flynn, PNNL
Virendra Ghate, ANL
Jonathon Helmus, ANL
Karen Johnson, BNL
Jinwon Kim, UCLA
Kyo-Sun Lim, PNNL
Robert K. Newsom, PNNL
Laura Riihimaki, PNNL
Yan Shi, PNNL
Timothy R. Shippert, PNNL
Shuaiqi Tang, LLNL
David D. Turner, NOAA ESRL
Shaocheng Xie, LLNL

Data Processing and Hosting
Anthony Clodfelter, ORNL
Ranjeet Devarkonda, ORNL
Kyle K. Dumas, ORNL
Kenneth E. Kehoe, Oklahoma U.
Bhargavi Krishna, ORNL
Jitendra Kumar, ORNL
Giri Prakash, ORNL
Harold A. Shanafield
Chitra Sivaraman, PNNL
Robert J. Records, ORNL

Communications
Hanna Goss, PNNL
Rolanda Jundt, PNNL
Robert Stafford, PNNL

ARM Management and Support
Jennifer Comstock, PNNL
Sarah J. Higley, PNNL
James H. Mather, PNNL
Sally McFarlane, DOE
Lucille A. Walker, PNNL

4.0 References

ARM. 2016. LASSO Alpha 1 data bundles. 36° 36’ 18.0” N, 97° 29’ 6.0” W: Southern Great Plains Central Facility (C1). Compiled by Gustafson, WI, Vogelmann, AM, Cheng, X, Endo, S, Krishna, B, Li, Z, Toto, T, Xiao, H. ARM Data Archive: Oak Ridge, Tennessee, USA, doi:10.5439/1256454.

ARM. 2017. LASSO Alpha 2 data bundles. 36° 36’ 18.0” N, 97° 29’ 6.0” W: Southern Great Plains Central Facility (C1). Compiled by Gustafson, WI, Vogelmann, AM, Cheng, X, Endo, S Krishna, B, Li, Z, Toto, T, Xiao, H. ARM Data Archive: Oak Ridge, Tennessee, USA, doi:10.5439/1342961.

Benjamin, SG, SS Weygandt, JM Brown, M Hu, CR Alexander, TG Smirnova, JB Olson, EP James, D C. Dowell, G. A. Grell, H. D. Lin, SE Peckham, TL Smith, WR Moninger, JS Kenyon, and GS Manikin. 2016. "A North American hourly assimilation and model forecast cycle: The rapid refresh." *Monthly Weather Review* 144: 1669–1694, doi:10.1175/mwr-d-15-0242.1.

Berg, LK, RK Newsom, and DD Turner. 2017. "Year-long vertical velocity statistics derived from Doppler lidar data for the continental convective boundary layer." *Journal of Applied Meteorology and Climatology* 56: 2441–2454, doi:10.1175/jamc-d-16-0359.1.

Clough, SA, MW Shephard, E Mlawer, JS Delamere, M Iacono, K Cady-Pereira, S Boukabara, and PD Brown. 2005. "Atmospheric radiative transfer modeling: A summary of the AER codes." *Journal of Quantative Spectroscopy and Radiative Transfer* 91: 233–244, doi:10.1016/j.jqsrt.2004.05.058.

Deardorff, JW. 1980. Stratocumulus-capped mixed layers derived from a 3-dimensional model. *Bound.-Layer Meteorology* 18: 495–527, doi:10.1007/Bf00119502.

Endo, S, AM Fridlind, W Lin, AM Vogelmann, T Toto, AS Ackerman, GM McFarquhar, RC Jackson, H. H Jonsson, and Y Liu. 2015. "RACORO continental boundary layer cloud investigations: 2. Large-eddy simulations of cumulus clouds and evaluation with in situ and ground-based observations." *Journal of Geophysical Research* 120: 5993–6014, doi:10.1002/2014jd022525.

Flynn, D, Y Shi, K-S Lim, and L Riihimaki. 2017. Cloud type classification (cldtype) value-added product. ARM Climate Research Facility DOE/SC-ARM-TR-200, 16 pp, https://www.arm.gov/publications/tech_reports/doe-sc-arm-tr-200.pdf.

Gustafson, WI, and AM Vogelmann. 2015. LES ARM Symbiotic Simulation and Observation (LASSO) implementation strategy. US Department of Energy Atmospheric Radiation Measurement Climate Research Facility DOE/SC-ARM-15-039, 31 pp, http://www.arm.gov/publications/programdocs/doe-sc-arm-15-039.pdf.

Gustafson, WI, AM Vogelmann, X Cheng, S Endo, B Krishna, Z Li, T Toto, and H Xiao. 2016. Description of the LASSO Alpha 1 release. Atmospheric Radiation Measurement (ARM) Climate Research Facility DOE/SC-ARM-TR-194, 163 pp, https://www.arm.gov/publications/tech_reports/doe-sc-arm-tr-194.pdf, doi:10.2172/1373564.

Gustafson, WI, AM Vogelmann, X Cheng, S Endo, B Krishna, Z Li, T Toto, and H Xiao. 2017. Description of the LASSO Alpha 2 release. Atmospheric Radiation Measurement (ARM) Climate Research Facility DOE/SC-ARM-TR-199, 209 pp, https://www.arm.gov/publications/tech_reports/doe-sc-arm-tr-199.pdf, doi:10.2172/1376727.

Iacono, MJ, JS Delamere, EJ Mlawer, MW Shephard, SA Clough, and WD Collins. 2008. "Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models." *Journal of Geophysical Research* 113: D13103, doi:10.1029/2008jd009944.

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.

Klemp, J. 2017. 18th WRF users workshop welcome. *18th WRF Users Workshop*, 13 June 2017, Boulder, CO, http://www2.mmm.ucar.edu/wrf/users/workshops/WS2017/oral_presentations/Welcome.pdf.

Li, Z, S Feng, Y Liu, W Lin, M Zhang, T Toto, AM Vogelmann, and S Endo. 2015a. "Development of fine-resolution analyses and expanded large-scale forcing properties: 1. Methodology and evaluation." *Journal of Geophysical Research* 120: 654–666, doi:10.1002/2014jd022245.

Li, ZJ, JC McWilliams, K Ide, and JD Farrara. 2015b. "A multiscale variational data assimilation scheme: Formulation and illustration." *Monthly Weather Review* 143: 3804–3822, doi:10.1175/mwr-d-14-00384.1.

Li, ZJ, XP Cheng, WI Gustafson, and AM Vogelmann. 2016. "Spectral characteristics of background error covariance and multiscale data assimilation." *International Journal of Numerical Methods in. Fluids* 82: 1035–1048, doi:10.1002/fld.4253.

Mlawer, EJ, SJ Taubman, PD Brown, MJ Iacono, and SA Clough. 1997. "Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave." *Journal of Geophysical Research* 102: 16663–16682, doi:10.1029/97jd00237.

Morrison, H, JA Curry, and VI Khvorostyanov. 2005. "A new double-moment microphysics parameterization for application in cloud and climate models. Part I: Description." *Journal of Atmospheric Science* 62: 1665–1677, doi:10.1175/jas3446.1.

Morrison, H, G Thompson, and V Tatarskii. 2009. "Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes." *Monthly Weather Review*, 137: 991–1007, doi:10.1175/2008mwr2556.1.

Öktem, R, Prabhat, J Lee, A Thomas, P Zuidema, and DM Romps. 2014. "Stereophotogrammetry of oceanic clouds." *Journal of Atmospheric and Oceanic Technology* 31: 1482–1501, doi:10.1175/jtech-d-13-00224.1.

Oue, M, P Kollias, KW North, A Tatarevic, S Endo, AM Vogelmann, and WI Gustafson. 2016. "Estimation of cloud fraction profile in shallow convection using a scanning cloud radar." *Geophysical Research Letters* 43: 10,998–11,006, doi:10.1002/2016GL070776.

Romps, DM, and R Öktem. 2015. "Stereo photogrammetry reveals substantial drag on cloud thermals." *Geophysical Research Letters* 42: 5051–5057, doi:10.1002/2015gl064009.

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. NCAR Technical Note, NCAR/TN-475+STR, National Center for Atmospheric Research, 113 pp, http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf, doi:10.5065/D68S4MVH.

Tatarevic, A, P Kollias, M Oue, and D Wang. 2017. User's guide CR-SIM software v3.0. 79 pp, http://radarscience.weebly.com/uploads/4/4/8/6/44864317/crsim-userguide-v3.0.pdf.

Thompson, G, RM Rasmussen, and K Manning, 2004. "Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis." *Monthly Weather Review* 132: 519–542, doi:10.1175/1520-0493(2004)132<0519:EFOWPU>2.0.CO;2.

Thompson, G, PR Field, RM Rasmussen, and WD Hall. 2008. "Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization." *Monthly Weather Review* 136: 5095–5115, doi:10.1175/2008MWR2387.1.

U.S. Department of Energy. 2014. Atmospheric Radiation Measurement Climate Research Facility decadal vision. DOE/SC-ARM-14-029, 30 pp, https://www.arm.gov/publications/programdocs/doe-sc-arm-14-029.pdf.

Xiao, H, S Endo, M Wong, WC Skamarock, JB Klemp, JD Fast, WI Gustafson, AM Vogelmann, HL Wang, YG Liu, and WY Lin. 2015. "Modifications to wrf's dynamical core to improve the treatment of moisture for large-eddy simulations." *Journal of Advances in Modeling Earth Systems* 7: 1627–1642, doi:10.1002/2015ms000532.

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.

Appendix A

WRF Namelist

The following WRF namelist.input settings are recommended for use with LASSO for shallow convection cases. These are in conjunction with defaults defined in WRF v3.9. The *crm* namelist block is an addition to WRF from the FASTER project.

```
&time_control
 run_days                       = 0,
 run_hours                      = 15,
 run_minutes                    = 00,
 run_seconds                    = 00,
 start_year                     = 2016,
 start_month                    = 06,
 start_day                      = 10,
 start_hour                     = 12,
 start_minute                   = 00,
 start_second                   = 00,
 end_year                       = 2016,
 end_month                      = 06,
 end_day                        = 11,
 end_hour                       = 03,
 end_minute                     = 00,
 end_second                     = 00,
 history_interval_m             = 10,
 history_interval_s             = 00,
 frames_per_outfile             = 6,
 restart                        = .false.,
 restart_interval_h             = 15,
 io_form_history                = 2,
 io_form_restart                = 2,
 io_form_input                  = 2,
 io_form_boundary               = 2,
 debug_level                    = 0,
 io_form_auxinput6              = 2,
 auxinput6_inname               =
"input_ls_forcing.nc",
 auxinput6_interval_m           = 30,
 io_form_auxinput7              = 2,
 auxinput7_inname               =
"input_sfc_forcing.nc",
 auxinput7_interval_m           = 30,
 io_form_auxinput8              = 2,
 auxinput8_inname               =
"input_aer.nc",
 auxinput8_interval_m           = 5,
 auxhist9_outname               =
"wrfstat_d<domain>_<date>",
 auxhist9_interval_m            = 10,
```

```
auxhist9_interval_s            = 0,
io_form_auxhist9               = 2,
/

&domains
time_step                      = 0,
time_step_fract_num            = 1,
time_step_fract_den            = 2,
max_dom                        = 1,
s_we                           = 1,
e_we                           = 250,
s_sn                           = 1,
e_sn                           = 250,
s_vert                         = 1,
e_vert                         = 227,
dx                             = 100,
dy                             = 100,
ztop                           = 14800,
grid_id                        = 1,
parent_id                      = 0,
i_parent_start                 = 0,
j_parent_start                 = 0,
parent_grid_ratio              = 1,
parent_time_step_ratio         = 1,
feedback                       = 0,
smooth_option                  = 0
/

&physics
mp_physics                     = 8,
ra_lw_physics                  = 4,
ra_sw_physics                  = 4,
radt                           = 1,
sf_sfclay_physics              = 1,
sf_surface_physics             = 1,
bl_pbl_physics                 = 0,
bldt                           = 0,
cu_physics                     = 0,
cudt                           = 0,
isfflx                         = 11,
ifsnow                         = 0,
icloud                         = 1,
```

```
num_soil_layers                = 5,
mp_zero_out                    = 0,
/

&fdda
/

&crm
crm_zsfc                       = 0.0,
crm_lat                        = 36.6,
crm_lon                        = -97.5,
crm_stretch                    = 3,
crm_num_pert_layers            = 33,
crm_pert_amp                   = 0.1,
crm_init_ccn                   = 100,
crm_lupar_opt                  = 1,
crm_znt                        = 0.04,
crm_emiss                      = 1.0,
crm_thc                        = 3.,
crm_mavail                     = 0.30,
crm_force_opt                  = 1,
crm_th_adv_opt                 = 1,
crm_qv_adv_opt                 = 1,
crm_th_rlx_opt                 = 0,
crm_qv_rlx_opt                 = 0,
crm_uv_rlx_opt                 = 0,
crm_vert_adv_opt               = 1,
crm_wcpa_opt                   = 0,
crm_num_force_layers           = 751,
crm_tau_s                      = 43200,
crm_tau_m                      = 3600,
crm_flx_opt                    = 2,
crm_sh_flx                     = -16,
crm_lh_flx                     = -93,
crm_albedo_opt                 = 1,
crm_albedo                     = 0.2,
crm_tsk_opt                    = 2,
crm_tsk                        = 400,
crm_ust_opt                    = 0,
crm_ust                        = 0.25,
crm_init_tke_opt               = 1,
crm_init_tke                   = 1.0,
crm_morr_act_opt               = 2,
crm_morr_hygro_opt             = 2,
crm_morr_hygro                 = 0.12,
crm_mp_nc                      = 100.,
crm_num_aer_layers             = 401,
crm_stat_opt                   = 1,
crm_stat_sample_interval_s     = 30.,
/

&dynamics
rk_ord                         = 3,
diff_opt                       = 2,
km_opt                         = 2,
diff_6th_opt                   = 0,
diff_6th_factor                = 0.12,
damp_opt                       = 3,
zdamp                          = 2000.,
dampcoef                       = 0.2,
khdif                          = 1.,
kvdif                          = 1.,
c_s                            = 0.18
c_k                            = 0.10
mix_isotropic                  = 1
smdiv                          = 0.1,
emdiv                          = 0.01,
epssm                          = 0.1,
tke_heat_flux                  = 0.24,
time_step_sound                = 6,
h_mom_adv_order                = 5,
v_mom_adv_order                = 3,
h_sca_adv_order                = 5,
v_sca_adv_order                = 3,
moist_adv_opt                  = 2,
chem_adv_opt                   = 2,
scalar_adv_opt                 = 2,
tke_adv_opt                    = 2,
tracer_adv_opt                 = 2,
mix_full_fields                = .true.,
non_hydrostatic                = .true.,
pert_coriolis                  = .true.,
m_opt                          = 1,
use_theta_m                    = 1,
/

&bdy_control
periodic_x                     = .true.,
symmetric_xs                   = .false.,
symmetric_xe                   = .false.,
open_xs                        = .false.,
open_xe                        = .false.,
periodic_y                     = .true.,
symmetric_ys                   = .false.,
symmetric_ye                   = .false.,
open_ys                        = .false.,
open_ye                        = .false.,
nested                         = .false.,
/

&grib2
/

&namelist_quilt
nio_tasks_per_group            = 0,
nio_groups                     = 1,
```

Appendix B

Papers, Products, and Presentations during Pilot

Conveying the LASSO vision to the atmospheric science community has been a critical part of the LASSO pilot. This has been done with the goal of increasing overall awareness of ARM's plans so that LASSO is readily used when it becomes operational combined with seeking input on how to best design and implement LASSO.

A strategy has been pursued that combines sample data products, ample community presentations, hosting of sessions during professional meetings, fostering beta users, an e-mail distribution list (http://us11.campaign-archive1.com/home/?u=74cd5b8a5435b8eca383fc18c&id=38f02e1568), and a website (https://www.arm.gov/capabilities/modeling/lasso). The following lists document these efforts.

B.1 Journal Articles

Oue, M, P Kollias, KW North, A Tatarevic, S Endo, AM Vogelmann, and WI Gustafson. 2016. "Estimation of cloud fraction profile in shallow convection using a scanning cloud radar." *Geophysical Reearch Letters* 43: 10,998–11,006, doi:10.1002/2016GL070776.

Li, Z, X Cheng, WI Gustafson, and AM Vogelmann. 2016. "Spectral characteristics of background error covariance and multiscale data assimilation." *International Journal of Numerical Methods Fluids*, 82: 1035–1048, doi:10.1002/fld.4253.

B.2 Data Sets

Atmospheric Radiation Measurement (ARM) Climate Research Facility. 2016. LASSO Alpha 1 data bundles. 36° 36' 18.0" N, 97° 29' 6.0" W: Southern Great Plains Central Facility (C1). Compiled by Gustafson, W. I., Vogelmann, A. M., Cheng, X., Endo, S., Krishna, B., Li, Z., Toto, T., Xiao, H. ARM Data Archive: Oak Ridge, Tennessee, USA, doi:10.5439/1256454.

Atmospheric Radiation Measurement (ARM) Climate Research Facility. 2017. LASSO Alpha 2 data bundles. 36° 36' 18.0" N, 97° 29' 6.0" W: Southern Great Plains Central Facility (C1). Compiled by Gustafson, WI, Vogelmann, AM, Cheng, X, Endo, S, Krishna, B, Li, Z, Toto, T, Xiao, H. ARM Data Archive: Oak Ridge, Tennessee, USA, doi:10.5439/1342961.

B.3 Technical Reports

Gustafson, WI, and AM Vogelmann. 2015. LES ARM Symbiotic Simulation and Observation (LASSO) implementation strategy. U.S. Department of Energy Atmospheric Radiation Measurement Climate Research Facility DOE/SC-ARM-15-039, 31 pp, http://www.arm.gov/publications/programdocs/doe-sc-arm-15-039.pdf.

Gustafson, WI, AM Vogelmann, X Cheng, S Endo, B Krishna, Z Li, T Toto, and H Xiao. 2016. Description of the LASSO Alpha 1 release. Atmospheric Radiation Measurement (ARM) Climate Research Facility DOE/SC-ARM-TR-194, 163 pp, https://www.arm.gov/publications/tech_reports/doe-sc-arm-tr-194.pdf, doi:10.2172/1373564.

Gustafson, WI, AM Vogelmann, X Cheng, S Endo, B Krishna, Z Li, T Toto, and H Xiao. 2017. Description of the LASSO Alpha 2 release. Atmospheric Radiation Measurement (ARM) Climate Research Facility DOE/SC-ARM-TR-199, 209 pp, https://www.arm.gov/publications/tech_reports/doe-sc-arm-tr-199.pdf, doi:10.2172/1376727.

B.4 Sessions at Professional Meetings Organized by LASSO

Gustafson, WI, AM Vogelmann, and R Neggers. 2015. Large-eddy and high-resolution simulations for improved understanding and parameterization of clouds and boundary layer processes. Oral and poster sessions at *2015 American Geophysical Union Fall Meeting*, San Francisco, CA, 14–15 December 2015.

Gustafson, WI, and AM Vogelmann. 2016. LASSO, year 1. Breakout session at *2016 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting*, Vienna, VA, 4 May 2016.

Gustafson, WI, and AM Vogelmann. 2017. LASSO. Breakout session at *2017 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting*, Vienna, VA, 13 March 2017.

Gustafson, WI, AM Vogelmann, and JH Mather. 2017. Observationally driven routine large-eddy simulations: enhancing community research through the DOE LASSO project. Town hall session at *2017 American Geophysical Union Fall Meeting*, New Orleans, LA, to be held 12 December 2017.

B.5 Presentations

B.5.1 2015

Gustafson, WI, and AM Vogelmann. 2015. “Overview of the LES ARM Symbiotic Simulation and Observation (LASSO) Workflow.” Invited talk at DOE *Atmospheric Radiation Measurement Developers Meeting*, Norman, OK, 10 July 2015.

Vogelmann, AM. 2015. “Linking small-scale cloud process observations to models.” Invited talk at *DOE ARM-ASR-ACME Coordination Meeting*, Germantown, MD, 21 October 2015.

Li, Z, X Cheng, WI Gustafson, H Xiao, AM Vogelmann, S Endo, and T Toto. 2015. “The DOE Atmospheric Radiation Measurement Program’s LES ARM Symbiotic Simulation and Observation (LASSO) workflow initialization, forcing and multiscale data assimilation.” Invited talk at *2015 American Geophysical Union Fall Meeting*, San Francisco, CA, 14 December 2015.

Gustafson WI, AM Vogelmann, H Xiao, S Endo, Z Li, X Cheng, and T Toto. 2015. “Modeling workflow for the DOE Atmospheric Radiation Measurement Facility’s LES ARM Symbiotic Simulation and Observation (LASSO) workflow.” Poster at *2015 American Geophysical Union Fall Meeting*, San Francisco, CA, 15 December 2015.

Xiao, H, S Endo, M Wong, WC Skamarock, J Klemp, JD Fast, WI Gustafson, AM Vogelmann, H Wang, Y Liu, and W Lin. 2015. “Modifications to WRF’s dynamical core to improve the treatment of moisture for large-eddy simulations.” Poster at *2015 AGU Fall Meeting*, San Francisco, CA, 15 December 2015.

Vogelmann, AM, WI Gustafson, T Toto, S Endo, X Cheng, Z Li, and H Xiao. 2015. “Model-observation “data cubes” for the DOE Atmospheric Radiation Measurement Program’s LES ARM Symbiotic Simulation and Observation (LASSO) workflow.” Poster at *2015 American Geophysical Union Fall Meeting*, San Francisco, CA, 15 December 2015.

B.5.2 2016

Vogelmann, AM, WI Gustafson, Z Li, X Cheng, S Endo, T Toto, and H Xiao. 2016. “Routine large-eddy simulations of continental shallow convection—workflow development.” Poster at HD(CP)2 Understanding Clouds and Precipitation, Berlin, Germany, 18 February 2016.

Vogelmann, AM, WI Gustafson, Z Li, X Cheng, S Endo, T Toto, and H Xiao. 2016. “LASSO—science requirements.” Invited talk at DOE Atmospheric Radiation Measurement Radar Workshop, Miami, FL, 25 February 2016.

Gustafson, WI, AM Vogelmann, Z Li, X Cheng, S Endo, T Toto, H Xiao, and JM Comstock. 2016. “LASSO modeling and measurements update.” Invited talk at 2016 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 3 May 2016.

Gustafson, WI, AM Vogelmann, Z Li, X Cheng, S Endo, T Toto, H Xiao, B Krishna, KS Lim, LD Riihimaki, J Kleiss, LK Berg, Y Zhang, and Y Shi. 2016. “Breakout session: LASSO, year 1.” Talk at 2016 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 4 May 2016.

Gustafson, WI, AM Vogelmann, H Xiao, S Endo, Z Li, X Cheng, and T Toto. 2016. “The LES ARM Symbiotic Simulation and Observation (LASSO) workflow pilot project.” Poster at 2016 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 4 May 2016.

Endo, S, Z Li, X Cheng, H Xiao, WI Gustafson, AM Vogelmann, T Toto, M Ahlgrimm, S Xie, and T Shuaiqi. 2016. “LES ARM Symbiotic Simulation and Observation (LASSO) workflow: ensemble forcings and LES sensitivity.” Poster at 2016 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 4 May 2016.

Vogelmann, AM, WI Gustafson, T Toto, S Endo, H Xiao, Z Li, and X Cheng. 2016. "LES ARM Symbiotic Simulation and Observation (LASSO) workflow: model-observation data bundles." Poster at 2016 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 4 May 2016.

Krishna, B, WI Gustafson, AM Vogelmann, T Toto, R Devarakonda, and G Prakash. 2016. "Exploring large-scale data analysis and visualization for ARM using NoSQL technologies." Poster at 2016 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 4 May 2016.

Comstock, JM, LD Riihimaki, WI Gustafson, AM Vogelmann, D Turner, M Cadeddu, RK Newsom, and JH Mather. 2016. "Boundary layer profiling modules: instruments and data products to support the ARM megasite at Southern Great Plains." Poster at 2016 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 4 May 2016.

Comstock, JM, JH Mather, WI Gustafson, and AM Vogelmann. 2016. "ARM Climate Research Facility megasite and high-resolution modeling—supporting model development and evaluation." Poster at 21st Annual CESM Workshop, Breckenridge, CO, 21 June 2016.

Gustafson, WI, AM Vogelmann, H Xiao, S Endo, T Toto, Z Li, X Cheng, and B Krishna. 2016. "Routine large-eddy simulations by the DOE Atmospheric Radiation Measurement Facility: the LES ARM Symbiotic Simulation and Observation (LASSO) project." Talk at 17th Annual WRF Users' Workshop, Boulder, CO, 28 June 2016.

Gustafson, WI, AM Vogelmann, H Xiao, S Endo, T Toto, Z Li, X Cheng, B Krishna. 2016. "Update on the LES ARM Symbiotic Simulation and Observation (LASSO) Project." Invited plenary talk at 2016 ARM Developer's Meeting, Oak Ridge, TN, 27 September 2016.

Gustafson, W. I., A. M. Vogelmann, H. Xiao, S. Endo, T. Toto, Z. Li, X. Cheng, B. Krishna. 2016: "LASSO Discussion Topics for Breakouts i.e., the LASSO Laundry List." Invited talk at 2016 ARM Developer's Meeting, Oak Ridge, TN, 28 September 2016.

Gustafson, W. I., A. M. Vogelmann, H. Xiao, S. Endo, T. Toto, Z. Li, X. Cheng, and B. Krishna. 2016: "Pondering radar usage within the LES ARM Symbiotic Simulations and Observation (LASSO) workflow." Invited talk at 5th Annual ARM Radar Workshop, Richland, WA, 14 November 2016.

Krishna, B, WI Gustafson, AM Vogelmann, T Toto, R Devarakonda, and G Prakash. 2016. "Exploring large-scale data analysis and visualization for ARM using NoSQL technologies." Poster at 2016 American Geophysical Union Fall Meeting, San Francisco, CA, 12 December 2016.

B.5.3 2017

Vogelmann, AM, WI Gustafson, Z Li, X Cheng, S Endo, B Krishna, T Toto, and H Xiao. 2017. "Routine large-eddy simulations of continental shallow convection—workflow development." Talk at 9th Symposium on Aerosol-Cloud-Climate Interactions, 97th American Meteorological Society Annual Meeting, Seattle, WA, 24 January 2017.

Riihimaki, LD, KSS Lim, Y Shi, JM Kleiss, LK Berg, WI Gustafson, Y Zhang, DM Flynn, and KL Johnson. 2017. “Identifying cloud types for more effective use of ARM observations in model validation and statistical analysis.” Talk at 9th Symposium on Aerosol-Cloud-Climate Interactions, 97th American Meteorological Society Annual Meeting, Seattle, WA, 24 January 2017.

Gustafson, WI, AM Vogelmann, Z Li, X Cheng, S Endo, J Kim, B Krishna, T Toto, and H Xiao. 2017. “LASSO: large-eddy simulation for the masses.” Talk at 2017 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 13 March 2017.

Vogelmann, AM, B Krishna, WI Gustafson, T Toto, Z Li, X Cheng, S Endo, J Kim, and H Xiao. 2017. “LASSO data bundle development and discovery.” Talk at 2017 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 13 March 2017.

Li, Z, X Cheng, WI Gustafson, H Xiao, AM Vogelmann, S Endo, T Toto, and J Kim. 2017. “Multiscale data assimilation forcing for LASSO.” Talk at 2017 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 13 March 2017.

Vogelmann, AM, WI Gustafson, T Toto, Z Li, X Cheng, S Endo, J Kim, B Krishna, and H Xiao. 2017. “Lidar and profiling use in LASSO.” Talk at 2017 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 13 March 2017.

Vogelmann, AM, WI Gustafson, B Krishna, T Toto, Z Li, X Cheng, S Endo, J Kim, and H Xiao. 2017. “LASSO data bundles and CMDV-CM4 related data products.” Talk at 2017 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 13 March 2017.

Vogelmann, AM, WI Gustafson, and T Toto. 2017. “Simulator prospects for LASSO.” Talk at 2017 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 13 March 2017.

Gustafson, WI, AM Vogelmann, Z Li, X Cheng, S Endo, J Kim, B Krishna, T Toto, and H Xiao. 2017. “Large-eddy simulation for the masses: LASSO’s going into production.” Poster at 2017 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 14 March 2017.

Vogelmann, AM, WI Gustafson, T Toto, S Endo, H Xiao, Z Li, X Cheng, B Krishna, and J Kim. 2017. “LASSO data bundles for consumption.” Poster at 2017 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 14 March 2017.

Li, Z, X Cheng, WI Gustafson, H Xiao, AM Vogelmann, S Endo, T Toto, and J Kim. 2017. “Multiscale data assimilation forcing for LASSO.” Poster at 2017 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 14 March 2017.

Krishna, B, K Dumas, WI Gustafson, AM Vogelmann, T Toto, and G Prakash. 2017. “Large-scale data analysis and visualization for ARM using NoSQL technologies.” Poster at 2017 Atmospheric Radiation

Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 14 March 2017.

Oue, M, P Kollias, A Tatarevic, T Toto, AM Vogelmann, S Endo, and WI Gustafson. 2017. “An ARSCL simulator: generation and application to LASSO case studies.” Poster at 2017 Atmospheric Radiation Measurement/Atmospheric System Research Principal Investigator Meeting, Vienna, VA, 14 March 2017.

Gustafson, WI, AM Vogelmann, H Xiao, S Endo, T Toto, Z Li, X Cheng, and J Kim. 2017. Ensemble large-eddy simulation for the DOE Atmospheric Radiation Measurement Facility’s LES ARM Symbiotic Simulation and Observation (LASSO) project. Talk at 18th Annual WRF Users’ Workshop, Boulder, CO, 13 June 2017.

Vogelmann, AM, WI Gustafson, Z Li, T Toto, S Endo, H Xiao, X Cheng, B Krishna, and J Kim. 2017. “Routine large-eddy simulations of continental shallow convection and megasite observations at the ARM Southern Great Plains Facility.” Poster at The Future of Cumulus Parameterization, Delft University, Delft, Netherlands, 12 July 2017.

Vogelmann, AM, WI Gustafson, Z Li, X Cheng, S Endo, B Krishna, T Toto, and H Xiao. 2017. “Using large-eddy simulations (LES)—the LASSO experience.” Invited talk at Summer Work-shop on High-Resolution Modeling and Visualization, University of Cologne, Cologne, Germany, 17 July 2017.

Gustafson, WI, H Xiao, AM Vogelmann, S Endo, T Toto, Z Li, X Cheng, and J Kim. 2017. “Transitioning LASSO from pilot to routine operations.” Invited talk at 2017 Atmospheric Radiation Measurement Developers’ Meeting, Richland, WA, 18 July 2017.

ARM

CLIMATE RESEARCH FACILITY

www.arm.gov

U.S. DEPARTMENT OF
ENERGY

Office of Science