U.S. DEPARTMENT OF ENERGY | Office of Science

DOE/SC-ARM-24-012

DOE ARM Future of LASSO Workshop Report

2–3 November 2023
Boulder, Colorado

WI Gustafson Jr.
AM Vogelmann
JH Mather

June 2024

ARM

DISCLAIMER

DOE/SC-ARM-24-012

DOE ARM Future of LASSO Workshop Report

2–3 November 2023
Boulder, Colorado

Organizers
WI Gustafson Jr., Pacific Northwest National Laboratory (PNNL)
AM Vogelmann, Brookhaven National Laboratory
JH Mather, PNNL

June 2024

How to cite this document:

Gustafson, WI, AM Vogelmann, and J Mather. 2024. DOE ARM Future of LASSO Workshop Report. U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington. DOE/SC-ARM-24-012.

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) user facility held a two-day workshop at the National Science Foundation National Center for Atmospheric Research in Boulder, Colorado in November 2023 to discuss the history and future of the Large-Eddy Simulation (LES) ARM Symbiotic Simulation and Observation (LASSO) activity. LASSO is a suite of data products focused on supplementing ARM observations with high-resolution modeling. The workshop used a hybrid format with 15 invited, in-person attendees and 36 virtual attendees logging in for portions of the workshop. Many topics were covered at the workshop, with this report reflecting the ensuing discussion. Some topics discussed had clear conclusions while others require further thought, given the breadth of topics included in a short workshop.

LASSO is unique in that it provides a library of kilometer-scale and O(100 m)-scale LES simulations to the community, spanning a range of case dates, combined with ensembles for each date that are vetted with ARM observations via skill scores. These simulations and associated diagnostics are available for download from ARM or can be accessed using ARM's computing infrastructure, depending on user requirements. A primary goal is to provide plausible representations of the environments surrounding the ARM instrumentation and estimates of processes that cannot be measured. This is valuable for improving understanding of processes, such as clouds and boundary-layer turbulence, as well as for developing and evaluating coarser-scale models.

Two LASSO regime-based scenarios are currently available. The first simulates shallow convection at ARM's Southern Great Plains (SGP) atmospheric observatory in Oklahoma (LASSO-ShCu). The second simulates upscale growth of orographically driven deep convection during the Cloud, Aerosol, and Complex Terrain Interactions (CACTI) field campaign in Argentina (LASSO-CACTI). Each scenario uses a modeling approach with forcing ensembles specifically designed around the respective target cloud regimes and have a finest grid spacing of 100 m.

The shallow convection scenario has been available longest with the first data released in 2016; the full data set covers 95 case dates. These have been used by researchers to investigate cloud and boundary-layer processes, cloud-radiation and aerosol-cloud interactions, cloud and radiation parameterization, radar scan strategies, and related topics.

The deep convection scenario was released as a beta release in 2022 with the full set of simulations released in 2023. Given the short period of availability, the available corpus of work with LASSO-CACTI is still small. However, early adopters praise the ensembles of mesoscale simulations that are particularly important for understanding sensitivities of convection to the background environment, as well as the selection of highly detailed LES, which are of a magnitude that very few research groups could produce. Examples of ongoing research using LASSO-CACTI include a study of convergence patterns and how they relate to the terrain for convective initiation, and a study of mass fluxes through cloud boundaries based on tracking convection and calculating the mass transfer through the lateral cloud boundaries.

A third scenario focusing on maritime clouds is currently under development. This scenario will simulate conditions at ARM's East North Atlantic (ENA) atmospheric observatory. Science topics motivating this scenario include precipitation processes and mesoscale cloud organization.

Workshop feedback can be grouped into several categories. In summary here, it is grouped into aspects of LASSO valued by users, ways to enhance the impact of LASSO, and important topics requiring further input and thought.

Participants noted several highly valued aspects of LASSO:

- The LASSO simulation and observation data sets represent a significant effort that ARM provides to the community. This enables novel investigations that many research groups would not otherwise be able to undertake. ARM has provided the modeling and observation expertise in addition to the more tangible computational capacity to produce the detailed simulations across a substantial number of cases.
- The LASSO case library approach moves beyond the common ‘golden day’ mindset, i.e., a focus on a single well-behaved or canonical case, by providing a range of curated cases for researchers to use. This saves researchers time in terms of identifying days of interest among the longer ARM observation periods. The library also enables more statistically robust analyses by incorporating some of the day-to-day variability associated with the selected weather regime.
- LASSO scenarios consist of multiple data types such as model output, evaluation diagnostics, processed model-comparable observations, and model configuration settings. The most-used data within LASSO have been the forcing data sets, with the quality of the forcings vetted against observations. The forcings are used to drive additional simulations, as needed for particular research applications, with the LASSO simulations serving as a foundation. The ensemble of forcings is also an important aspect given the uncertainty of a single forcing without the context of additional forcing options.
- Combining the simulation library with surface and remote-sensing observations from ARM and other sources puts the simulations in the context of observed reality. This synergy helps guide new users who might not be familiar with available ARM data toward high-priority measurements related to the LASSO scenarios.

Several discussed topics involved how to enhance the impact of LASSO:

- Enhancing communication throughout the scenario development process could increase user engagement. This could take the form of more frequent status updates and continuing to use the online tools developed by the LASSO activity, including the Discourse forum and web-based documentation.
- Implementing a staged rollout of new LASSO scenarios, e.g., with intermediate products, would enable users to be more engaged in the development process and have access to useful data earlier in the process. This would enable users to begin using the scenarios sooner and would generate useful feedback before final case configurations are “baked in” and when adjustments can still be made. How best to accomplish this in a way that provides stable, reliable data needs further evaluation.
- Incorporating LASSO into the ARM concept of data epochs would bring visibility to LASSO products and help LASSO users identify complementary data sets in the ARM archive.
- Researchers often use LASSO as a foundation to build additional tools, simulations, and observation comparisons. Associating these data, workflows, and tools with LASSO would enable other researchers to leverage the community’s work, which would both increase the visibility of the original contributors and further additional research. Approaches need to be determined for how best to

connect these community-based activities with LASSO. Examples include contribution of ARM principal investigator (PI) products, adding information to the ARM online documentation, adding references in the Bundle Browser to related data sets, and using the Discourse forum.

Important topics that would require further work and community input include:

- The related topics of *representativeness* and *selection biases* were discussed extensively. It was suggested that future work be done to inform users about representativeness to assist in appropriate use of the LASSO data bundles. Selection bias was a noted concern that should be made clear to users. For example, LASSO cases are selected based on certain criteria that likely limit some of the processes/variability and how best to include "null" cases (days when the chosen phenomenon is expected to occur but does not) is unclear. However, how to meaningfully characterize the representativeness and selection bias is an open topic that has thus far evaded community consensus.
- Working closely with the Energy Exascale Earth System Model (E3SM) developers to help them integrate LASSO into their workflows was suggested as beneficial to both E3SM and ARM. Initial inroads have included the LASSO team working with E3SM developers to use LASSO data with the E3SM single-column model. Additional networking can be done with DOE Atmospheric System Research (ASR) Science Focus Areas (SFAs), such as the new Tying in High-Resolution E3SM with ARM Data (THREAD) project, which is anticipated to develop a methodology to incorporate LASSO data. In addition, E3SM capabilities and resolutions are starting to overlap with LASSO modeling scales, and there are opportunities for using the E3SM hierarchy of models from the mesoscale down to LES to parallel LASSO and assist model development.
- Accelerating scenario development would more directly link LASSO with ARM Mobile Facility (AMF) field campaign deployments. How to reduce development time is unclear, given the need to balance aspects such as data availability and quality control, case curation, and the level of complexity of the LES. Workshop discussions were wary of exclusively limiting LASSO to cheaper, coarse simulations since that would reduce the demonstrated value of the current LASSO approach.
- Scenario selection is an ongoing topic requiring detailed planning specific to each scenario. How best to balance priorities across the community for different user needs is an open topic. Additionally, a tension exists between producing "routine" simulations with less thought put into their selection and optimization of simulation accuracy versus a case-study mode where initial and boundary conditions plus model configuration details are strong foci. Example scenarios discussed during the workshop include the issue of how best to address the important processes related to aerosol-cloud interactions and how to balance the number of case dates with the level of model complexity. White papers were submitted with scenario ideas for the Bankhead National Forest (BNF) and Eastern Pacific Cloud Aerosol Precipitation Experiment (EPCAPE) field campaigns, and these were used as examples for discussing differing approaches.

Overall, workshop participants were enthusiastic about LASSO and ARM's intent to continue generating new LASSO scenarios. Useful feedback was garnered from the workshop regarding prior LASSO products and a range of ideas were presented for future development. This will be valuable for improving the ENA scenario currently under development and subsequent scenarios to be developed. Feedback from this workshop will be incorporated into LASSO planning and will inform prioritization when establishing the overall LASSO roadmap, as well as the more detailed tasks when requesting and establishing the LASSO budget.

Acronyms and Abbreviations

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
4D	four-dimensional
ACI	aerosol-cloud interaction
ADC	ARM Data Center
AERI	atmospheric emitted radiance interferometer
AERIoe	AERIoe Thermodynamic Profile and Cloud Retrieval Value-Added Product
AMF	ARM Mobile Facility
AMF1	first ARM Mobile Facility
API	application programming interface
ARM	Atmospheric Radiation Measurement user facility
ARSCL	Active Remote Sensing of Clouds Value-Added Product
ASR	Atmospheric System Research
BER	Biological and Environmental Research program (U.S. DOE)
BNF	Bankhead National Forest
BOMEX	Barbados Oceanographic and Meteorological Experiment
CACTI	Cloud, Aerosol, and Complex Terrain Interactions
CCPP	Common Community Physics Package
CF	Central Facility
CLDFRAC	Cloud Fraction
CM1	Cloud Model 1
COGS	Clouds Optically Gridded by Stereo Value-Added Product
COMBLE	Cold-Air Outbreaks in the Marine Boundary Layer Experiment
COMBLE-MIP	COMBLE-Model–Observation Intercomparison
CPU	central processing unit
CR-SIM	Cloud-Resolving Model Radar Simulator
CSAPR2	C-band Scanning ARM Precipitation Radar
DA	data assimilation
DALES	Dutch Atmospheric Large-Eddy Simulation
DEPHY	Développement et Évaluation PHYsiques des modèles de climat et prévision du temps
DMS	dimethylsulfide
DOE	U.S. Department of Energy
DOI	Digital Object Identifier
DP-SCREAM	Doubly Periodic SCREAM
DYCOMS	Dynamics of Chemistry and Marine Stratocumulus

E3SM	Energy Exascale Earth System Model
ECMWF	European Centre for Medium Range Weather Forecasts
ED2	Ecosystem Demography model v2.2
EDA	ECMWF Data Assimilation
EDMF	eddy diffusivity mass-flux
EF	Extended Facility
EMC^2	Earth Model Column Collaboratory
EPCAPE	Eastern Pacific Cloud Aerosol Precipitation Experiment
ERA5	ECMWF Reanalysis v5
FNL	Final Operational Global Analysis
FTP	File Transfer Protocol
FY	fiscal year
GCM	global climate model
GCSS	GEWEX Cloud System Study
GEFS	Global Ensemble Forecast System
GEWEX	Global Energy and Water Cycle Experiment
GFS	Global Forecast System
GMTB	Global Model Test Bed
GOES	Geostationary Operational Environmental Satellite
GPU	graphics processing unit
HPC	high-performance computing
HTTP	Hypertext Transfer Protocol
ICON	Icosahedral Nonhydrostatic
IFS	Integrated Forecasting System
IOP	intensive operational period
KAZRARSCL	Ka-band ARM Zenith Radar ARSCL Value-Added Product
KNMI	Royal Netherlands Meteorological Institute
LASSO	LES ARM Symbiotic Simulation and Observation
LASSOBLTHERMO	LASSO Middle Boundary-Layer Thermodynamics
LASSODLCBHSHCU	LASSO Doppler Lidar Cloud-Base Height for Shallow Cumulus
LASSOLWP	LASSO Liquid Water Path
LCL	lifting condensation level
LCLHEIGHT	Lifting Condensation Level Height
LES	large-eddy simulation
LLNL	Lawrence Livermore National Laboratory
MET	surface meteorological instrumentation
MOSAiC	Multidisciplinary Drifting Observatory for the Study of Arctic Climate
MSDA	Multiscale Data Assimilation

MWRRet	Microwave Radiometer Retrieval Value-Added Product
MYNN	Mellor-Yamada-Nakanishi-Niino
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
netCDF	Network Common Data Form
NOAA	National Oceanic and Atmospheric Administration
NWP	numerical weather prediction
OCLF	ORNL Leadership Computing Facility
ORNL	Oak Ridge National Laboratory
OSSE	Observing System Simulation Experiment
PDF	Portable Document Format
PI	principal investigator
PINACLES	Predicting Interactions of Aerosol and Clouds in Large-Eddy Simulation
QC	quality control
RELAMPAGO	Remote sensing of Electrification, Lightning, and Mesoscale/microscale Processes with Adaptive Ground Observations
RH	relative humidity
RRM	regionally refined model
RWPWINDCON	Radar Wind Profiler Wind Consensus
SAM	System for Atmospheric Modeling
SCM	single-column model
SCREAM	Simple Cloud-Resolving E3SM Atmosphere Model
SFA	Science Focus Area
SGP	Southern Great Plains
SOA	secondary organic aerosol
SOM	self-organizing map
SSH	Secure Shell protocol
THREAD	Tying in High-Resolution E3SM with ARM Data
TRACER	Tracking Aerosol Convection Interactions Experiment
TROPoe	Tropospheric Optimal Estimation Retrieval Value-Added Product
TSI	total sky imager
UTC	Coordinated Universal Time
VAP	value-added product
VARANAL	Constrained Variational Analysis Value-Added Product
WRF	Weather Research and Forecasting Model
WR-Chem	WRF coupled with Chemistry
WRF-Hydro	WRF-Hydro Modeling System

Contents

Figures

Tables

1.0 Introduction

The DOE ARM user facility held the *DOE ARM Future of LASSO Workshop* on November 2-3, 2023. It focused on the LASSO activity (Gustafson et al. 2019, 2020), which started in 2015 with a pilot project that grew into an ongoing ARM data product that uses high-resolution modeling to complement and add value to ARM's observations. This workshop had two primary purposes. The first was to evaluate important aspects of LASSO and how it has evolved, and the second was to discuss possible future directions for LASSO over the next five to ten years and details that could inform prioritization by DOE decision makers. This report summarizes the discussions and outcomes from the workshop.

William Gustafson, Andrew Vogelmann, and James Mather convened the workshop and developed the content. The workshop was held at the National Science Foundation National Center for Atmospheric Research in Boulder, Colorado, hosted by Hailey Shin and Hugh Morrison. It used a hybrid Zoom-based approach with 52 attendees, 15 of whom were in person. The in-person attendees were invited based on their expertise and exposure to ARM's capabilities. The goal was to have a wide range of expertise that covers many of ARM's scientific topic areas, both those currently addressed by LASSO and those that could be addressed in the future. As such, attendees included experts in clouds, aerosol-cloud interactions, weather and climate modeling, large-eddy simulation modeling, measurement and remote retrievals of clouds and aerosols, as well as the ARM Technical Director and LASSO leadership. Appendix B contains a list of attendees.

The format for the workshop consisted of presenting a series of slides informing each session's topic followed by a guided discussion, with most of the time spent in discussion. Appendix A contains the agenda. Before the workshop, white paper submissions were requested to address a range of LASSO topics that formed the final agenda. Two more white papers (Appendix C) also helped shape the agenda. The first morning of the workshop focused on examining the current LASSO products, discussing their effectiveness, and looking for insights and lessons learned that would influence future LASSO activities. The workshop then looked to the future, first taking a high-level perspective to examine the LASSO philosophy, and then narrowing focus until more detailed items were discussed by the end of the workshop. Throughout the workshop, a high priority was understanding how the LASSO data sets have been or could be used and what could be done to increase the overall scientific impact of LASSO. Participants reported the overall format was highly effective for engaging them in the discussion and inspiring their thinking.

The sessions were video recorded and notes were taken in shared Google documents to which all participants could contribute text in addition to the notes taken by two scribes, Jennifer Comstock and Scott Giangrande. The zoom chat, containing dialogues on different topics, was also saved as part of the record.

2.0 The LASSO Concept

The ARM user facility collects atmospheric, soil, and related observations at both long-term and targeted field campaign locations and has done this for over three decades (Turner and Ellingson 2016). Specifically, ARM's mission states that ARM "provides the climate research community with strategically located atmospheric observatories to improve the understanding and representation in Earth system models of clouds and aerosols and their interactions with the Earth's surface." The LASSO activity has been developed to help bridge the two foci of this mission, observations and Earth system models, as well as to provide information to understand and apply ARM observations for various research investigations.

The LASSO concept acknowledges the great value of detailed, carefully obtained observations, while at the same time addressing the difficult issues of connecting disparate observations and bridging the scale gap between localized measurements and coarse-scale models, such as Earth system models. Processes connecting high-resolution measurements can be elusive and high-resolution modeling can help clarify how ARM measurements fit together. Additionally, Earth system models commonly have grid columns in the range of 25 to 100 km wide. Comparing these to point-based observations, or even semi-volumetric observations, such as scanning cloud radars, can be hindered by different sampling methodologies, both in time and space. One must contend with the representativeness of the observations across the larger, simulated scales combined with the inability to measure process rates and other important details that connect measurements of different variables and from multiple locations.

Figure 1 schematically depicts this situation, where one might try to understand the cloud field through measurements from vertically pointing and scanning cloud and precipitation radars, vertically pointing lidars, aircraft transects, radiosonde balloons, multispectral ground-based measurements, and traditional meteorological stations. Each instrument samples different parts of clouds and the surrounding environment. Additionally, the sampling strategy of each instrument is rarely optimal for particular needs. These complications bring uncertainty from not knowing where in the cloud field each sample comes from, significantly complicating interpretation and use of the observations; for example, air entering and leaving the cloud is not symmetric across each cloud, nor is the humidity gradient extending out from cloud edges (Mallaun et al. 2019).

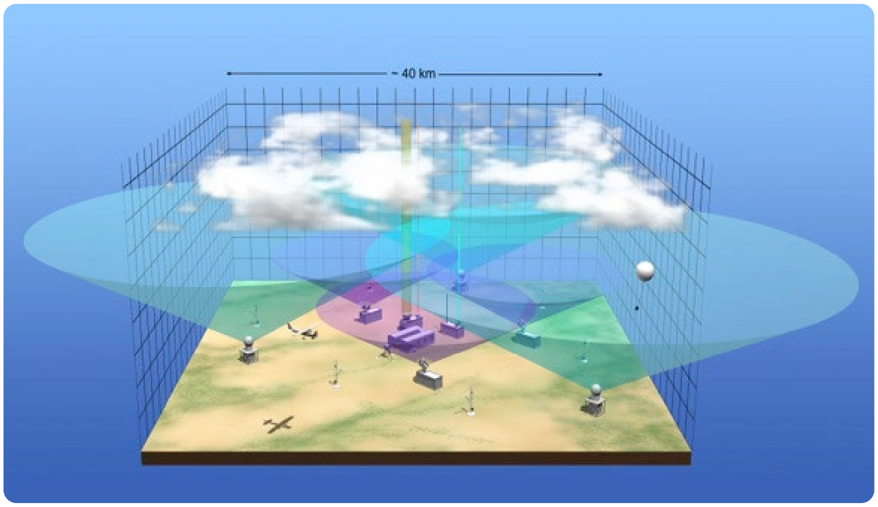


Figure 1. A schematic representation of different sampling regions from various instruments employed by ARM and how they compare to scales within Earth system models. Examples include vertically pointing and scanning cloud and precipitation radars, vertically pointing lidars, aircraft, radiosonde balloons, surface-based radiometers, and meteorological stations.The LASSO solution to the above conundrum of scales, for both understanding measurements and connecting them to large-scale models, is to use high-resolution modeling to recreate plausible representations of the region around ARM's measurements. Researchers can query

> the high-resolution, LASSO model output, which includes both process rates and basic state variables, to gain insights regarding how measurement sampling impacts their work and to make the jump from seeing state variables to understanding processes. Measurements can only get one so far; including the modeling in combination with the observations opens new avenues for interpreting and applying the observations.

LASSO products are designed with five target types of users in mind. Each has specific needs that inform decisions such as the LASSO model configuration and the types of events that LASSO simulates. LASSO developers attempt to meet as many needs as possible to make LASSO a broadly applicable data set. The five user types are:

- *Model developers:* they have applications such as model validation and a need for model forcing data sets to use in combination with their own models; these developers are characterized by the need to compare many aspects of their model with observations and benchmark models to determine model accuracy across many cases;
- *Parameterization developers*: they work on detailed pieces of models; unique aspects of these developers include the need to understand processes and fine-scale details at the process level; they need to compare their parameterizations to benchmark models and the observations for specific cases, which typically requires using LASSO forcing data to drive the parameterization and/or its host model;
- *Theoreticians:* they are more likely to use the LASSO model output directly and may or may not rerun simulations; data needs include various meteorological states;
- *Observationalists:* they approach LASSO data sets as proxies for realistic conditions at the ARM observatory; unlike the modelers, observationalists do not want to rerun simulations and often require full 3D volumes to sample from, and potentially high-frequency output to match instrument sampling strategies more closely; and
- *Educators:* LASSO is a valuable teaching tool with its readily available model output, observations, related analyses, and open-source code; it can be used for teaching contexts such as numerical modeling, observation-model comparisons, cloud and boundary-layer processes, and scientific Python coding.

In addition to attempting to meet the needs of a broad range of researchers, LASSO has several defining characteristics. They are the use of:

- *Simulation libraries* to provide more statistically robust sets of simulations,
- *Ensembles* vetted by observations to address uncertainty in the background meteorology used to force the simulations,
- *Scenarios* consisting of targeted modeling based on specific meteorological regimes and/or locations, and
- *Packaging and sharing* of LASSO data sets efficiently to simplify and increase their use.

These concepts work together to shape the LASSO data sets, including which dates are selected for simulation, which observations are used to vet the simulations, how observations and modeling get combined, configuration choices for the models, and how data is shared with users.

Use of simulation libraries consists of combining simulations from many case dates and distributing them as a set. This differentiates LASSO from most prior large-eddy simulation modeling that use a single-case mentality. The intent of libraries is to provide a more statistically robust data set of model output that spans conditions at the ARM site as they vary from day to day for the selected phenomenon. In contrast, popular single-case examples of traditional LES intercomparisons are the Barbados Oceanographic and Meteorological Experiment (BOMEX) (Siebesma et al. 2003) and the Dynamics of Chemistry and Marine Stratocumulus (DYCOMS) campaign (Stevens et al. 2005). The BOMEX and DYCOMS modeling has proved useful, but these limited number of cases cannot be used to examine issues such as sensitivity to the background conditions. These and other commonly used GEWEX (Global Energy and Water Cycle Experiment) Cloud System Study (GCSS) cases are often highly idealized, which also differentiates them from LASSO.

The use of ensembles of simulations for each case date builds upon the library concept to examine simulation sensitivities. As employed to date, the ensembles primarily reflect uncertainty in the input and boundary conditions driving the model. These are assumed to be the largest contributors to uncertainty for many of the days. As described in the scenario descriptions in Section 3, a range of available analyses and reanalyses are sampled to obtain estimates of the background meteorological state. In some cases, physics variations are also employed as additional ensemble members.

A critical aspect of using ensembles is identifying which of them are closest to observations, which are our best understanding of reality; this involves using ARM and other available observations. Unlike forecast ensembles where reality is not yet known, LASSO uses hindcast ensembles where one knows what actually happened. Each simulation is compared with observations for selected variables and scored on how closely the simulation matches. This then permits one to make a justified selection of simulations when using them for applications requiring the simulations to reflect the real world as closely as possible. This linking of observations with modeling is extremely important, and without it, the ensembles would be no more than a range of possible representations of reality. The scoring and connection to reality adds value to the hindcast ensembles, helping researchers identify the best boundary conditions to use for a given date and providing more confidence in the simulated fields.

The scenario concept focuses the simulations on specific meteorological regimes and locations. One approach for providing many cases at a particular location is to run a model continuously, which enables having model output all hours of the day and for all the different types of weather that occur at a site. An example of this approach is the Royal Netherlands Meteorological Institute (KNMI) Parameterization Testbed (Neggers, Siebesma, and Heus 2012), which is one of the modeling experiments that inspired LASSO. LASSO differs from this by using a targeted approach. For a given location with the needed critical observations, a subset of days exhibiting a selected meteorological regime are selected for simulation. This then permits use of a model configuration consistent with the regime, while not having to compromise simulation accuracy or providing “too much” modeling on days that only need a simpler configuration. For example, shallow convection does not require a large horizontal domain extent while capturing deep convection often does. Model physics parameterizations can also be tailored to the specific conditions.

The final concept, packaging and sharing of the LASSO data sets, targets making the data as easy to access and use as possible. This is in line with ARM’s user facility status, and it is assumed that ease of use will increase uptake of the products by researchers. Careful thought has gone into organizing the data within data sets, simplifying finding and downloading of the files, adding value to the basic model output through pre-computing commonly used variables and statistics, and other related details.

Overall, the LASSO activity can be summarized as the use of high-resolution modeling to provide plausible proxies for data gap filling, understanding processes, and bridging scales. This is done using a library of scenarios with multiple case dates and ensembles of simulations to address statistical robustness and clarify model sensitivities across a range of meteorological regimes. The simulations are carefully compared with observations to ground them in measurements, to reference them to observed reality, and help users select which simulations are most useful for their needs. The range of data are packaged and provided to the community to increase use of ARM observations and speed research related to the selected regimes.

3.0 Current Scenarios

As noted in the prior section, LASSO uses a regime-based, scenario approach. This involves focusing selection of case dates on the chosen meteorological regimes and locations with ARM observations. This section describes the two currently available scenarios and a new one under development.

3.1 SGP Shallow Convection Scenario

3.1.1 Description

The first LASSO scenario focused on shallow convection at the SGP. The scenario was mandated in the request for white papers for the "Development of a Framework for Routine Large-Eddy Simulations over ARM Sites" that founded the LASSO activity. This cloud regime was an excellent starting point because LES had been used for shallow convection for decades and the extensive experience could inform the overall model configuration. This permitted focusing efforts on how to perform observationally constrained modeling and integrate the model-observation products into ARM's storage, discovery, and delivery systems. Additionally, advanced instrumentation had just been added to the SGP extended facilities that would enhance the sampling of the highly variable shallow convection cloud fields.

The science drivers motivating the scenario focused on processes that drive shallow convection and the difficulty for weather and climate models to simulate them. Shallow convective clouds are small enough that even O(1-10 km)-scale atmospheric models—which are used for current state-of-the-art weather forecasts and are the aspirational scale for future climate models such as DOE's E3SM—cannot reproduce the radiative and mixing effects of these clouds at the resolved scale and, instead, must parameterize the impacts of these clouds on the resolved scales. High-resolution LASSO simulations would enable examining the processes involved, the assumptions used in parameterizations and, additionally, the resulting 3D volumes of the atmosphere would enable developing and testing retrieval methodologies for these clouds. All these topics benefit from the LASSO library approach providing a statistical view of the phenomena that goes beyond the typical LES approach of using single, finely tuned case studies.

The configuration and approach for this scenario were chosen based on a combination of science drivers and practicality. An overview of key aspects is provided in this section and further details are available in the documentation (Gustafson et al. 2019, 2020). We used the Weather Research and Forecasting Model (WRF) owing to its maturity, through heavy use and vetting within the community, and flexibility in the availability of multiple parameterizations and boundary-condition options, which enables applicability to future phenomena without changing the baseline model. The model configuration used was a traditional LES approach with doubly periodic boundaries. This means that the LES represents the average conditions around the SGP and that each model column is statistically identical such that one cannot think in terms of one-to-one comparisons to real-world point locations. A final domain size of 25 km was used with a horizontal grid spacing of 100 m, and a vertical grid spacing of 30 m up to 5 km that then stretched to 300 m at domain top. This configuration was considered the most frugal to still capture shallow convection processes and provide a good representation of the clouds.

Surface sensible and latent heat fluxes at the model's lower boundary were specified as being time-varying but spatially homogeneous based on the observationally based values from the Constrained Variational Analysis (VARANAL) Value-Added Product (VAP). This approach captured the first-order effects of the surface fluxes on the simulations, but it prohibited feedback between the land and atmosphere or the representation of the surface gradients across the region. We did not attempt computing these fluxes with an interactive land-surface model because of uncertainties in those models at that time for LES and the lack of sufficient data to initialize the soil temperature and moisture profiles to reliably yield the observed fluxes.

The LES simulations were initialized with the 12 UTC sounding from the Central Facility and the influence of changing background atmospheric conditions on the simulations were represented by profile-based, large-scale forcing applied as tendencies to the domain. A unique aspect of the LASSO approach was the use of an ensemble of large-scale forcings to generate an ensemble of LES realizations for each simulated case date, in recognition that uncertainty in the large-scale forcings is a significant contributor to accurately simulating a given day. Eight forcing methodologies were used, obtained from the VARANAL data assimilation product; three spatial scales from the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS); three spatial scales from the Multiscale Data Assimilation (MSDA) algorithm, and one without large-scale forcing.

ARM observations were compiled to vet the simulations and convey simulation quality to users via skill scores (see Table 1). Compiling the observations to accompany the LES is a critical component of LASSO, as this both highlights key ARM measurements for users and the skill scores provide the information needed for users to decide which of the ensemble members are best suited for their research application. The observations found to be of greatest value for discrimination of model fidelity were cloud macro properties observed from the Central Facility (CF): liquid water path, horizontal cloud fraction, and height-resolved cloud fraction. Other CF measurements were used to assess water vapor, relative humidity, and temperature at the surface and at the middle of the boundary layer. Example instruments used with LASSO-ShCu are shown in Figure 2. In addition to these observations from the CF, regionally representative values were obtained from the extended network of observations for cloud-base height and lifting condensation level (LCL). All these CF and network observations were used in skill scores. Additionally, data not used for skill scores were the CF sondes, used in model-observation comparison plots, and the radar wind profiler data, used in the MSDA algorithm. The ARM infrastructure dedicated a lot of support to LASSO to make available products from the newly deployed instrument networks. However, owing to some of the novel needs of LASSO and its delivery timetable, LASSO needed to undertake processing of specialized data sets when they exceeded the infrastructure's bandwidth. This resulted in the generation and distribution of the LASSO High-Frequency Observations VAP (see Table 2).

Table 1. ARM observations and VAPs used to vet SGP shallow convection simulations. Details on implementation are in Gustafson et al. (2019).

Property	Input Source	Site Location(s)
In-cloud liquid water path	Combined AERIoe and MWRRet	CF
Horizontal cloud fraction	KAZRARSCL, TSI, and COGS	CF
2D time-height cloud fraction	KAZRARSCL and COGS	CF
Surface thermodynamics (q, T, RH)	MET	CF
Mid-boundary thermodynamics	Combined AERIoe and Raman lidar	CF
Thermodynamic profiles	Balloon-borne sounding system	CF
Cloud-base height	Doppler lidar	CF and EF
Lifting condensation level	MET and Oklahoma Mesonet	CF, EF and Mesonet

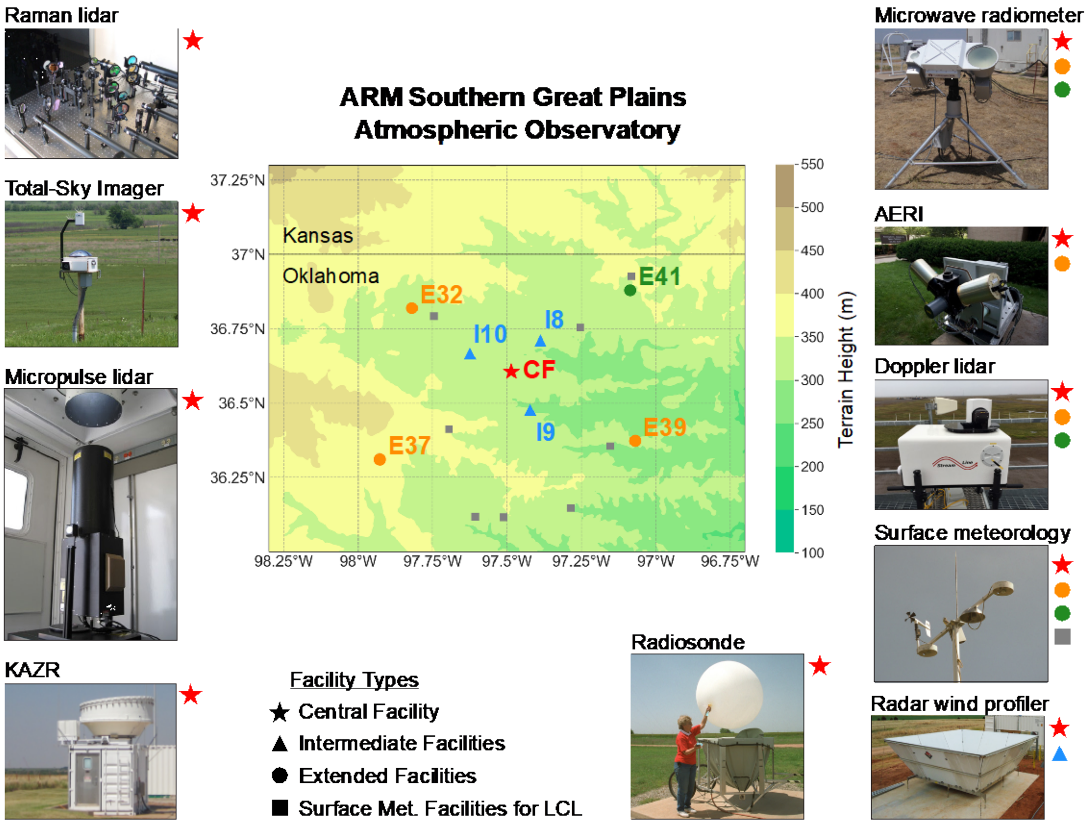


Figure 2. Some of the instruments used for the LASSO-ShCu scenario at the SGP. Figure from Gustafson et al. (2020).**Table 1.** Products contained within the LASSO High-Frequency Observations VAP (lassohighfreqobs) at native resolution, available for 2017-2019 at the SGP. Further details are given in Gustafson et al. (2019).

Name	Description
LASSO Liquid Water Path (LASSOLWP)	Liquid water path from AERIoe & MWRRet, 10-second resolution
Cloud Fraction (CLDFRAC)	Cloud cover from KAZRARSCL and TSI, 1-, 5-, 15-minute resolution
LASSO Middle Boundary-Layer Thermodynamics (LASSOBLTHERMO)	Temperature and moisture in middle of the boundary layer (500-700 m) from AERIoe and Raman lidar, 10-minute resolution
LASSO Doppler Lidar Cloud-Base Height for Shallow Cumulus (LASSODLCBHSHCU)	Cloud-base heights from the Doppler lidars, 10-minute resolution
Lifting Condensation Level Height (LCLHEIGHT)	LCL at 1-min resolution
Radar Wind Profiler Wind Consensus (RWPWINDCON)	Horizontal wind from radar wind profilers (RWPs), 10-minute resolution (2019 only)

LASSO "data bundles" organize the resulting model and observational data into a series of tar files that can be discovered and ordered through a purpose-designed "Bundle Browser."[1] These tar files contain co-gridded model and observational values and their comparison plots, skill scores computed from these values that quantify the comparison accompanied by their supporting plots, the configuration data used to drive the model, and the raw model output. Additionally, bundle files are available for the input observational data processed by the LASSO team: the LASSO High-Frequency Observations VAP and the data processed from the original Clouds Optically Gridded by Stereo (COGS) VAP. The LASSO Bundle Browser was developed by the ARM Data Center (ADC) to make it easier for users to find simulations for their particular application based on selected criteria, which is a very different approach than is available for searching the full archive of available ARM observations with the Data Discovery web interface.[2] The selection criteria include date(s), measurement used in the model comparison, type of large-scale forcing and scale, surface flux treatment, and microphysics parameterization. The choices of these criteria result in dynamic displays of quick-look plots and an interactive table with slide bars for skill score values that can be used to further winnow the results to desired specifications. The table also contains boxes for the tar files to be selected for ordering and staging the data for download via FTP or Globus.

The shallow convection scenario was run for five seasons, 2015-2019. Case days were manually selected from those exhibiting shallow convection during April-October. Other criteria included that the cloud field be relatively homogeneous across the SGP region, there should not be strong synoptical features impacting the region (e.g., fronts), and that critical instrumentation was functioning properly. The first two criteria avoided frontal/heterogeneous conditions, which were necessary to be consistent with our model configuration. The number of days with shallow convection at the SGP varied considerably across the five seasons, respectively: 5, 13, 30, 30, and 17. In total, the case library consists of 95 days and over 760 simulations.

---

[1] LASSO Bundle Browser website for shallow cumulus: https://adc.arm.gov/lassobrowser

[2] ARM Data Discovery website: https://adc.arm.gov/discovery/

3.1.2 Usage and Applications Enabled

At the time of this publication, about 30 publications and one PhD thesis have used LASSO-ShCu products. LASSO is further recognized in another 36 publications that mention the LASSO framework but do not use the data, as well as noting that international conference calls for abstracts have explicitly mentioned LASSO as a topic area of interest.

The published research enabled by LASSO covers a wide range of topics and encompasses all the anticipated user groups: model developers, parameterization developers, theoreticians, observationalists, and educators; an example of each is discussed below. Primary topics include cloud properties and processes, parameterization of boundary-layer and cloud processes, radiation, instrument retrievals, aerosol-cloud interactions, boundary-layer processes, and forcing studies. Table 3 shows the number of publications by primary topic, with the most common being cloud properties and processes (9, or 30%) and the parameterization of boundary-layer and cloud processes (5, or 17%).

Table 2. Number of publications using LASSO-ShCu categorized by primary topic.

Primary Topic	Count
Cloud properties and processes	9
Parameterization of boundary layer and cloud processes	5
Radiation	3
Instrument retrievals	3
Aerosol-cloud interactions	2
Boundary layer processes	2
Forcings	2
Other	4
Total	**30**

The frequency with which the product components are being used helps understand which parts of LASSO are most valuable to researchers. All publications presumably used the skill scores and plots to determine what cases to use. Beyond that, the provided data bundle products contain the forcing data and model input data, 3D model output, co-gridded observed and simulated values used to compute the skill scores, and LES statistical output. Table 4 contains estimates of the number of times each component was used based on the reading of the 30 manuscripts, noting that more than one component could be used in a publication. Except for the LES statistical output (not listed) for which there was no clearly stated usage, all other components were used. The most popular for shallow convection are the forcing data sets (73%) and 3D output (40%). Also provided is a count of when ARM observations external to LASSO's production were employed, indicating that LASSO potentially facilitated the use of other observations in 40% of the publications.

Table 3. Number of publications using each LASSO-ShCu data product. Note that more than one product might be used in a publication. Also included are the number of times ARM observations external to the LASSO data streams were used.

Data Product	# Uses (%)
Forcing data	22 (73)
3D output	12 (40)
Co-gridded observed and simulated values	4 (13)
LASSO-external ARM observations	12 (40)

Exploring the use cases further helps understand the breadth of research facilitated and what data components were needed to achieve their research goals. Parameterization development and testing typically uses forcing data to drive the model and the 3D output as a benchmark. A total of five parameterizations studies have been published thus far. The first example of this is Angevine et al. (2018) who tested the new Mellor-Yamada-Nakanishi-Niino (MYNN) eddy diffusivity mass-flux (EDMF) boundary-layer and shallow cloud scheme. Additionally, ingests have been generated for using LASSO-ShCu forcings to drive the E3SM single-column model (SCM) via its case library, which is used to develop and test E3SM parameterizations. LASSO also has been included in a case library included with the Common Community Physics Package (CCPP) Single-Column Model by the Development Testbed Center, a joint National Center for Atmospheric Research (NCAR)/U.S. Air Force/National Oceanic and Atmospheric Administration (NOAA) center (Heinzeller et al. 2023). The CCPP is used to test physics parameterizations for the national forecast modeling system, which includes the Unified Forecast System, as well as other models that use the CCPP framework. These are clear examples of LASSO impacting weather and climate model development, one of the key ARM objectives for LASSO.

An important LASSO objective is facilitating the use of ARM observations. This was found commonly when users used the forcings as a starting point for their own targeted simulations, which occurred in eight publications. For example, Gristey et al. (2020) used such an approach to examine the impacts of 1D versus 3D radiation that facilitated an analysis on the observed impact of 3D radiation on the observed shortwave surface fluxes.

Another key intended LASSO user group are instrumentalists, who may otherwise not have access to high-resolution simulations. Oue et al. (2016) used LASSO 3D output to determine a radar scan strategy for the cloud fields that considers the complication of the strong dependence of radar sensitivity to target distance. Finally, we note that the readily available, packaged LASSO data has classroom value that educators are becoming aware of. An excellent example was the incorporation of the readily available LASSO data in the Second ARM Training and Science Application Event (Ghate et al. 2019).

3.2 CACTI Deep Convection Scenario

3.2.1 Description

The second LASSO scenario focuses on orographically driven deep convection during the ARM CACTI field campaign (Varble et al. 2018, 2019, 2021). This scenario, referred to as LASSO-CACTI, was selected through a white paper call, workshop, and down-selection process. The *LASSO Expansion*

Workshop in 2019 used four submitted white papers as preliminary concepts to discuss potential scenario options (Gustafson et al. 2019). The foci of the four topics were arctic clouds, clear-air turbulence, deep convection, and maritime clouds. Ultimately, deep convection during CACTI was selected due to a balance of data availability, clearest vision for the modeling approach, and broad community interest. This section gives a brief description of LASSO-CACTI with a full description available in the LASSO-CACTI technical document[3] (Gustafson et al. 2024).

The CACTI campaign occurred from October 2018 through April 2019 in the Córdoba region of Argentina in coordination with the National Science Foundation's Remote sensing of Electrification, Lightning, and Mesoscale/microscale Processes with Adaptive Ground Observations (RELAMPAGO) campaign. The Sierras de Córdoba Mountain range experienced repeated storm formation along its ridgeline, which then propagated eastward over and near the first ARM Mobile Facility (AMF1). Scanning radars, radiosondes, lidars, radiometers, and other instrumentation measured a wide variety of cloud characteristics and the background atmosphere. The repeated localization of storm initiation made the AMF1 siting particularly advantageous for observing the cloud systems as they developed and passed nearby. This also made this campaign a good choice for LASSO since a single domain configuration could capture many convective days without having to move the domain to overlap with the storm trajectories, as might be necessary for simulation deep convection at many other locations.

Design of the LASSO-CACTI scenario sought to balance the science foci selected for the scenario with resource availability, where resources include computational capacity and storage, labor, and the ultimate time to completion. The primary science foci identified were convective cloud dynamics and microphysics-dynamics interactions. This drove decisions around model configuration such as the grid spacing, domain size, a multi-scale ensemble methodology, and what data to archive from the simulations. The goal was to resolve plume-like structures within the convective cores and to capture the portion of the convective life cycle from initiation through initial upscale growth.

Twenty dates were selected from the seven-month CACTI period that had convection within range of the C-band Scanning ARM Precipitation Radar (CSAPR2). A domain width of several hundred kilometers is necessary to contain the desired portion of the cloud life cycles, which is very expensive to simulate with LES grid spacings. So, a multi-scale approach was used that simulated all 20 case dates using grid spacings down to 2.5 km, from which a subset of case dates was simulated with LES grid spacings of 100 m. These two modeling scales were referred to as the "meso" runs and the "LES" runs. The meso runs provided information for a larger region to see background storm conditions and a wider variety of convective behavior, while the LES runs targeted detailed storm dynamics for selected days where the meso runs showed good behavior, and likely had accurate boundary conditions, and had initiation within the range of the CSAPR2 followed by upscale growth. These two resolutions were chosen because 2.5 km is roughly the target resolution of the high-resolution version of E3SM called SCREAM (Simple Cloud-Resolving E3SM Atmosphere Model), it is a common CRM grid spacing used for many convective studies in the past, and going finer than this encroaches on the terra incognita, which is wise to avoid due to spurious oscillations and other numerical artifacts (Wyngaard 2004, Zhou, Simon et al. 2014, Rai et al. 2019). The finer LES grid spacing of 100 m was chosen to be below the transition from plume-like to bubble-like cloud core characteristics to enable detailed studies of convective dynamics (Lebo and Morrison 2015, Varble et al. 2019).

---

[3] LASSO-CACTI Technical Document website: https://lasso-cacti-doc.arm.gov

Within each case date LASSO-CACTI uses ensembles for the meso runs based on different boundary-condition options. A total of 33 potential ensemble members were obtained from the ECMWF Reanalysis v5 (ERA5) reanalysis, the ECMWF Data Assimilation (EDA) ensemble of 10 members, the National Centers for Environmental Prediction (NCEP)'s Final Operational Global Analysis (FNL), and the Global Forecast System (GFS) Global Ensemble Forecast System (GEFS) that has 21 members. Each of these ensemble members is assumed equally probable in terms of representing reality. Overall, this LASSO-CACTI ensemble is a valuable data set for investigating convective sensitivity to background conditions, as the resulting convection varies quite substantially within each case date's ensemble.

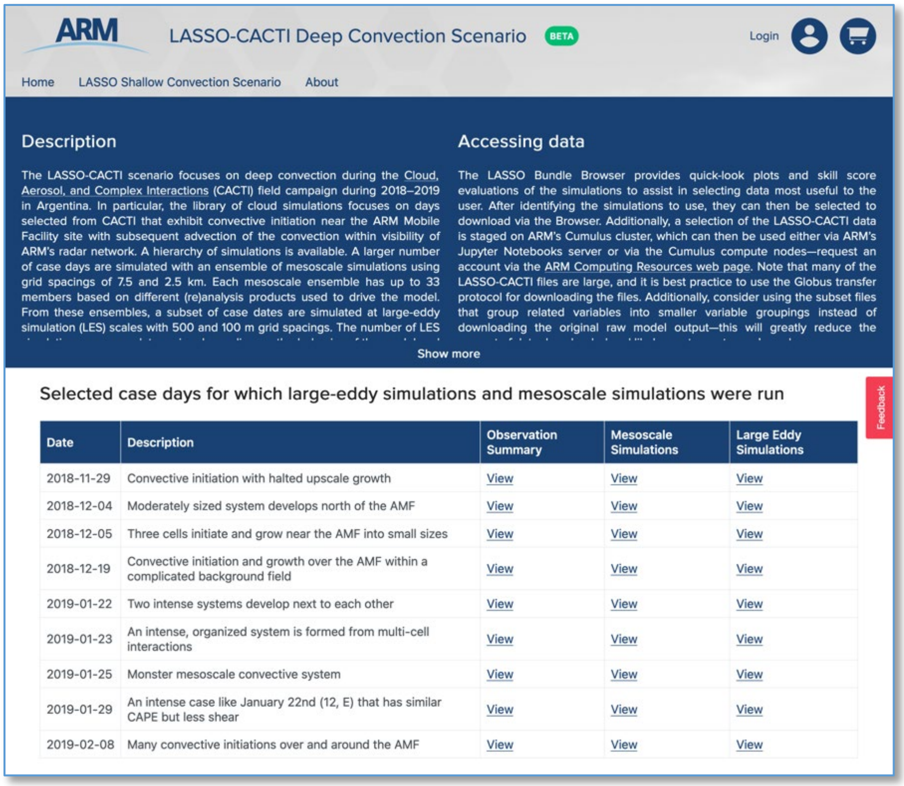
Full 33-member ensembles were deemed too costly for the LES model configuration. Instead, the meso ensembles were used to identify the best-performing boundary conditions, which were then used as a first attempt for an LES simulation on the subset of days chosen for the LES runs. Ultimately, each LES case date was run with multiple boundary-condition options. This was necessary to get a well-behaved simulation as opposed to trying to provide a sample of sensitivity, as was the goal for the meso ensembles. Additional LES ensemble members were produced using different microphysics or other parameterization settings to obtain the observed behavior.

The quality of the meso and LES runs were assessed through comparisons with several different data sets that were quantified using skill scores. Geostationary Operational Environmental Satellite (GOES-16) brightness temperatures were used to evaluate the time-dependent areal coverage of the convective cores. CSAPR2 data were used to evaluate the simulated radar echo-top heights and surface precipitation. Additionally, plots were also made for Skew-T and hodograph comparisons of the simulated profiles to radiosondes. See Gustafson et al. (2024) for more detail.

One data set need that input data sets used for the boundary conditions could not provide for initializing the LASSO-CACTI simulations was high-resolution profiles of the soil conditions. This was remedied through a continuous WRF-Hydro Modeling System (WRF-Hydro) simulation run from August 2018 through April 2019. This included about three months of spin-up prior to the first LASSO-CACTI case, which was sufficient for the LASSO-CACTI needs. The WRF-Hydro simulation was then sampled at the initialization time of each case to provide initial conditions for the soil in WRF. The WRF-Hydro simulation is being released as a sidecar product available alongside the WRF simulations. This is an example of an additional data set resulting from LASSO that will be useful for researchers, who in this case might need spatially detailed soil information beyond the available observations.

The combination of larger domains compared to the shallow convection scenario while maintaining the 100-m grid spacing increased the data storage for LASSO-CACTI by almost two orders of magnitude for each LES simulation. Domains for shallow convection were 25 km across, but for LASSO-CACTI the highest-resolution domain was 215×278 km^2. This necessitated a change in approach for the data bundling since most users could no longer easily download a full set of model output. So, a more modular approach was developed for LASSO-CACTI combined with producing a series of "subset" files, which are organized into 15 different groupings of variables by theme. Each subset grouping contains related variables and can be selected for download separately from other groupings. This way, users can request just the cloud condensate variables, radiation tendencies, meteorological state, etc., depending on their research needs. The subsets also include variables that often need to be computed in post-processing, such as de-staggered and rotated wind components, temperature, pressure, and diagnostics such as convective available potential energy, lifting condensation level, and wind shear. The raw WRF output files are also available for those that require them, along with the input and restart files. A change from the shallow-convection data bundles is that the observations are also not packaged directly with the LASSO-CACTI model output.

As for shallow convection, a specialized webpage was developed to query and download the LASSO-CACTI data set, the LASSO-CACTI Bundle Browser at https://adc.arm.gov/lasso/#/cacti, Figure 3. Lessons learned from the shallow convection browser led to using static quick-look plots to ease site maintenance. A YAML-based application programming interface (API) was also developed to aid in automating download of data across the range of case dates, ensembles, and file types. When used, this can reduce the number of clicks needed on the web page and permit automation. Overall, the LASSO-CACTI browser required a more sophisticated approach to presenting and organizing the simulations due to the multiple layers of detail within the LASSO-CACTI library. Users drill down through the data starting at the level of the case dates, followed by modeling scale (meso or LES), ensemble member, and then file type. The skill scores and various summary plots are provided for the simulation comparisons to the satellite brightness temperature, echo-top heights, and radar-retrieved surface precipitation. The Skew-T and hodograph comparisons plots are also available.



Description

The LASSO-CACTI scenario focuses on deep convection during the Cloud, Aerosol, and Complex Interactions (CACTI) field campaign during 2018–2019 in Argentina. In particular, the library of cloud simulations focuses on days selected from CACTI that exhibit convective initiation near the ARM Mobile Facility site with subsequent advection of the convection within visibility of ARM's radar network. A hierarchy of simulations is available. A larger number of case days are simulated with an ensemble of mesoscale simulations using grid spacings of 7.5 and 2.5 km. Each mesoscale ensemble has up to 33 members based on different (re)analysis products used to drive the model. From these ensembles, a subset of case dates are simulated at large-eddy simulation (LES) scales with 500 and 100 m grid spacings. The number of LES

Accessing data

The LASSO Bundle Browser provides quick-look plots and skill score evaluations of the simulations to assist in selecting data most useful to the user. After identifying the simulations to use, they can then be selected to download via the Browser. Additionally, a selection of the LASSO-CACTI data is staged on ARM's Cumulus cluster, which can then be used either via ARM's Jupyter Notebooks server or via the Cumulus compute nodes—request an account via the ARM Computing Resources web page. Note that many of the LASSO-CACTI files are large, and it is best practice to use the Globus transfer protocol for downloading the files. Additionally, consider using the subset files that group related variables into smaller variable groupings instead of downloading the original raw model output—this will greatly reduce the

Selected case days for which large-eddy simulations and mesoscale simulations were run

Date	Description	Observation Summary	Mesoscale Simulations	Large Eddy Simulations
2018-11-29	Convective initiation with halted upscale growth	View	View	View
2018-12-04	Moderately sized system develops north of the AMF	View	View	View
2018-12-05	Three cells initiate and grow near the AMF into small sizes	View	View	View
2018-12-19	Convective initiation and growth over the AMF within a complicated background field	View	View	View
2019-01-22	Two intense systems develop next to each other	View	View	View
2019-01-23	An intense, organized system is formed from multi-cell interactions	View	View	View
2019-01-25	Monster mesoscale convective system	View	View	View
2019-01-29	An intense case like January 22nd (12, E) that has similar CAPE but less shear	View	View	View
2019-02-08	Many convective initiations over and around the AMF	View	View	View

Figure 3. The LASSO-CACTI Bundle Browser home page at https://adc.arm.gov/lasso/#/cacti.

In addition to simplifying the ordering process with the Bundle Browser, ARM began offering users the capability of using the LASSO-CACTI data directly on ARM's computing hardware. Through an account request process, users can obtain access to either a Jupyter server for running Jupyter notebooks or the Cumulus high-performance computing cluster located at Oak Ridge National Laboratory (ORNL). This obviates the need to download LASSO-CACTI files to offsite locations by keeping the data within ARM resources at ORNL and the associated ORNL Leadership Computing Facility (OCLF). Analyses can also be parallelized across the Cumulus compute nodes to speed working with the large number simulations.

While it is early to know how this will work in the long term, initial feedback from early adopters is that this has greatly facilitated working with the LASSO-CACTI data set. It has also permitted doing simulation re-runs on the same platform as the original LASSO-CACTI simulations to generate supplemental data not saved for general use, such as ultra-high-frequency model output for selected variables.

3.2.2 Usage and Applications Enabled

The LASSO-CACTI data set has been formally available for less than a year, preceded by the beta release announced in May 2022. Representative usage statistics are not yet available. However, a handful of early adopters started working with the data set and provided suggestions for refinement, and examples of how researchers can use LASSO-CACTI.

Design of LASSO-CACTI was driven by the Expansion Workshop recommendation to have detailed cloud fields to study cloud motions and microphysics, inter-cloud dynamics, convective initiation, and the sensitivity of clouds to the environment (Gustafson et al. 2019). To date, work is ongoing in these areas. Zhe Feng presented preliminary examples of research on these topics to the workshop, with contributions by Adam Varble, Jim Marquis, Enoch Jo, and Zhe Feng from PNNL. Building off initial studies of the dependency of convective initiation on sounding characteristics and cloud tracking using radar data, they have extended their studies to include the LASSO-CACTI simulations (Feng et al. 2022, 2023). Cloud tracking has been applied to the ensembles of meso and LES simulations, which has been helpful to identify the environmental conditions at the specific time and location of convective initiation. This also shows resolution dependencies in cloud behavior when going from kilometer to 100-m-scale grid spacings. In a related study, they looked at regions of convergence and how this relates to convective initiation. This work made extensive use of the subsetted data provided by LASSO-CACTI, the software to generate new subsets, and the raw WRF output files.

Jo also took advantage of the frequent restart files output for the LES simulations. Using the ARM-provided simulations to identify key periods of interest, he re-ran short portions of the LES with 15-s output frequency by using the WRF restart files provided by LASSO-CACTI. Jo’s high-frequency output was then used to identify and track individual up- and downdrafts within cloud cores as well as to diagnose fluxes across cloud boundaries. Figure 4 shows an example of updrafts in relation to a cloud boundary.

An ASR request for proposals was open at the time of the Future of LASSO Workshop that included funding for CACTI research. We anticipate that this will increase the pool of researchers able to work with LASSO-CACTI over the next several years.

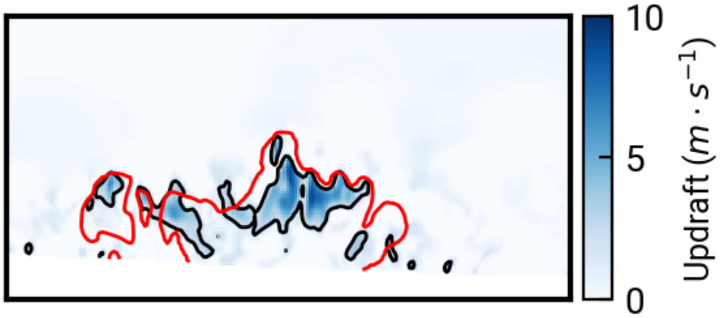


Figure 4. Vertical cross-section of a simulated cloud showing updrafts in blue. Cloud core boundaries are shown in black and the 10-dbZ reflectivity contour in red, giving an indication of the cloud location. This is preliminary work by Enoch Jo, PNNL.

ENA Maritime Cloud Scenario

The third LASSO scenario, currently in development, focuses on maritime clouds at ARM's ENA atmospheric observatory. Like LASSO-CACTI, the LASSO-ENA scenario is an outgrowth of a white paper submitted to the *LASSO Expansion Workshop* in 2019 (Gustafson et al. 2019). The discussions from that workshop, combined with additional input, have refined the vision for this scenario towards a focus on two primary aspects of the ENA clouds. The first outcome was to provide simulations that span multiple regimes of mesoscale organization. ENA regimes have traditionally been categorized as open-cell and closed-cell regimes based on cellular structures with the dominant cloud features around the rings of the cells versus the centers, respectively. The physical drivers leading to these two primary cloud modes range from changes in the background aerosol state to meteorological controls such as air-sea temperature contrasts, overall stability, and subsidence variations. A second ENA scenario driver would address ongoing scientific interest in understanding precipitation processes at ENA. As noted in the white paper, drizzle frequently forms alongside clouds at ENA, and more ubiquitously than expected by current theory.

The specific modeling approaches for LASSO-ENA are still being developed, so current scenario details were not discussed at this workshop. However, this scenario was cited multiple times throughout the workshop as an example when considering future possibilities for LASSO. For example, the types of clouds simulated for LASSO-ENA are of great interest to global-scale modelers because of biases in those models. The current ENA workplan is to produce a series of tiered simulations and to have community interactions at various stages of the release. A series of somewhat large domains, on the order of 100-km wide, capable of capturing the mesoscale cellular structures would be available earlier in the production. These simulations would compromise the resolution to some extent to enable the larger domains. These would then be followed by higher-resolution LES with smaller domains to elucidate the precipitation processes. These latter simulations could also use more detailed microphysics parameterization, such as spectral-bin microphysics (Khain et al. 2015).

4.0 Focus Topics

4.1 Ensembles and Case Selection

Addressing the challenge of obtaining reasonable simulations has involved using a forcing ensemble applied to a range of case dates that also serves to build the desired case library. The large-scale forcing used to drive LASSO simulations is based on a physical ensemble that draws its members from different sources to capture the observed behavior by at least one member. (This type of ensemble is not to be confused with the perturbation ensemble used in weather forecasting in which the ensemble average is considered the best predictor.) While all ensemble members used in LASSO are from credible sources and/or methodologies, a wide spread of simulated behaviors are typically found for a case day without any pattern emerging as to a forcing member that consistently outperforms others from case to case. Thus, a single (deterministic) forcing can have a large amount of uncertainty and the practical approach has been to run multiple forcings and determine the best member by vetting the results with observations. Similarly, predicting the case dates that will be well simulated has been challenging, requiring that the case dates selected include those that might initially seem suboptimal, for example, due to nearby weather variability.

The practice of picking the “best” simulation based on comparison to the observations essentially uses the model as an interpolator. The model provides a physically consistent picture of the atmosphere (within the ability of the model physics) that is checked as possible with available observations. Agreement is inferred to mean that dominant physical processes are reasonably simulated as are also the unverifiable dominant physical properties. Without the ability to do data assimilation at LES-scales, this is the most tractable approach to arrive at a high-resolution “4D data cube” for use in research. However, it was noted that the original concept of a data cube with which to confront models with observations (U.S.Department of Energy Climate and Environmental Sciences Division 2014a, 2014b) has become more like “data fusion,” where separation between model and data is less distinct. Choosing the best simulation based on comparison to the observations conflates different potential sources of error: modeling error, such as structural and parametric uncertainty; forcing uncertainty, which includes initial conditions; and observational uncertainty, such as measurement uncertainty and sampling representativeness of the point measurements. As a result, the source of the model-observation discrepancy is difficult if not impossible to differentiate and complicates usage for objectives such as understanding model bias.

Data assimilation (DA) was raised as a possible option for improving the forcings. LASSO has used VARANAL (Xie et al. 2004), which is a basic DA, as well as the MSDA algorithm (Li, Feng,et al. 2015, Li, McWilliams, et al. 2015). Additionally, all the forcing data sets incorporated into LASSO involve data assimilation of operational weather observations in one way or another by the operational centers. It was pointed out that DA has improved a lot in recent years so there might be better options for including DA directly in the LASSO workflow than were available in 2015 when LASSO started. Still, some fundamental challenges were noted. DA requires a dense network of observations and such a network does not exist for thermodynamic profiling. Also, even with a dense array, an analysis by Li et al. (2016) has suggested that the background error correlation length scales are large enough to prevent current DA approaches from constraining spatial scales smaller than 150 km for stream functions and 50 km for water vapor mixing ratios, thereby requiring fundamental modification for assimilating high-resolution observations into fine-resolution models O(1 km). Nevertheless, numerical weather prediction (NWP) developers have steadily improved their methods so a discussion may help determine what new products might be available for future LASSO use. Note that an additional potential advantage of using a formal DA approach could be to include the innovations resulting from the DA process in the LASSO output. The innovations would indicate how strongly the DA had to adjust the model to achieve the observation-model match. Therefore, these might be understanding model biases and behavior.

The topic of case selection involved discussions about selection bias and case representativeness. Users also noted that they appreciate and value the range of case dates that enable going beyond golden days. Having a wider selection of case dates permits increasing robustness of analyses.

LASSO has selection bias because cases were run only when the topic phenomenon was present, while, for example, days when it did not occur could also be of interest. It is also unknown whether the phase space of meteorological conditions leading to the topic phenomenon has been fully sampled by the selected cases—it is assumed to not have been due to the lack in diversity of locations experiencing the same cloud type plus the needs served by our screening process. It was pointed out that different users have different needs. Some will only want realistic simulations whereas those who are testing parameterizations will want a spread of simulations to test sensitivity of phenomena to environmental conditions. To some extent, the selection bias has been partially addressed, in a backdoor way, through the failed simulations that represent plausible environments/forcings but that did not develop properly.

Another source of selection bias that must be acknowledged, but that is difficult to avoid, is selecting cases based on periods when observations will be most representative of the larger region. This is particularly emblematic of the approach planned for the LASSO-ENA scenario, which will restrict case selection to days when the wind fetch is from roughly the northerly sector such that the instrumentation at ENA samples conditions from the open ocean with minimal contamination from the island. This excludes synoptic conditions with winds from the south, but has the advantage of avoiding waves and other features induced by flow over the main portion of the island.

Looking forward, a way to address selection bias would be to perform semi-continuous modeling, as proposed in the BNF white paper (see Appendix C.2). A discussion of considerations to improve case selection can be found in Section 5.3 “Operational Considerations.” The difficulty comes in designing a model configuration robust across the range of meteorological conditions experienced at a location combined with managing the resulting increase in data compared to only simulating a selection of tightly curated days.

Case representativeness refers to whether the selected cases span the different states that can occur within the topic phenomenon. For example, self-organizing maps (SOMs) generated for an extended period (e.g., Mechem et al. 2018) could be used to assess how completely the limited set of LASSO cases span the regime space. Taking that a step further, capturing the different states could also involve selecting cases across ARM sites for a given phenomenon to increase the variability of conditions important to climate model simulations (e.g., shallow convection occurring in the Amazon basin versus that at SGP). While generating SOMs to assess representativeness is likely beyond LASSO resources, SOMs have become more commonly used and might be available from PIs or as translator VAPs for upcoming scenarios. Expanding LASSO to simulate the full regime space is also a change in approach from how LASSO was initially envisioned. Consideration will need to be given to this topic as new scenarios are developed.

4.2 Observations

A key aspect of LASSO is making the best use of ARM observations to vet the simulations. This process helps inform the community of the critically discerning observations available for a given phenomenon, and it opens the door for further investigations that can leverage other ARM observations (e.g., Endo et al. 2019). There is a strong impetus for LASSO to adopt the most advanced instruments and retrievals available to gain from the enhanced information they offer. However, a lesson learned is that this enthusiasm should be bridled because working with new products can involve unanticipated complications while the datastreams are being stabilized. Iterating with evolving datastreams involves reprocessing the model-observation data and taxes LASSO resources. Experience has also shown that using an immature or undeveloped product leads to the LASSO team undertaking development that should be handled by other developers when it was beyond the infrastructure's bandwidth. This became necessary to prevent stalling LASSO production timelines. Development, particularly the transition of a product into production status, by its nature, can involve a long tail to maturation even for things that seem within immediate reach. A complicating factor is that LASSO's use of multi-day ensembles means that the development must work for a range of cases, which involves far more effort to accomplish than if working with a one-off case study or simulation. Thus, a lesson learned is that LASSO must work with established instruments and products. This may involve delaying production on a scenario until the needed observations have been vetted (e.g., following a field campaign). Even when working with established products, the volume of multi-day ensembles can become overwhelming to quality control the model-observation comparisons as the number (N) of observations increases; so, it is advisable to limit N to the most essential observations.

The need to restrict the number of observations to only those most critical to LASSO operations means many potentially relevant observations are not engaged in the LASSO processing. This is necessary to remain within budget and timetable when prioritizing multiple demands involved with the processing. However, the outcome does not engage the potential benefit of accessing the full breadth of ARM data to involve people with it and learn directly from the observations. An approach that could address this shortcoming would be to add LASSO cases as epochs, for which additional observations are curated by ARM for well-simulated LASSO cases. As such, attention could be drawn to data sets available outside of LASSO and could include cutting-edge techniques as they become available. For example, the Tropospheric Optimal Estimation Retrieval VAP (TROPoe) is a now-available improved version of the AERIoe algorithm that was used in the shallow convection scenario, being more accurate and with higher resolution. Having epoch data outside of the LASSO bundling would avoid the difficulty of reprocessing bundles as new data are introduced or revised, and it would simplify communicating data provenance.

Different aspects of observational uncertainty were discussed, particularly regarding how it affects the assessments of the simulations. In the LASSO processing, an attempt has been made, when possible, to make the comparison with observations in an apples-to-apples manner. Examples include using the same code to compute LCL from the model output and observations, screening the simulated liquid water path (LWP) columns to only include those above the measurement threshold, and then also using this screened field to determine horizontal cloud fraction for comparison with the total sky imager (TSI) and COGS observations (for more details, see Gustafson et al. (2019)). These approaches could be considered as poor-man's instrument simulators. As instrument simulators are developed for the ARM instrumentation, such as the Cloud-Resolving Model Radar Simulator (CR-SIM) (Oue et al. 2020) and the Earth Model Column Collaboratory (EMC^2) (Silber et al. 2022), it might be beneficial for LASSO to partner with their developers to further enhance the comparability of the measured and model-generated property. Even if the simulator is not used in LASSO, its availability to the community can be made known, for example with a Jupyter notebook. Additionally, it was noted that some variables and retrievals are readily available observationally, such as cloud optical depth and effective radius, that could be used in the evaluation matrix when appropriate assumptions are made in the model for determining the comparable observed/retrieved and simulated values.

Another important consideration is the sampling representativeness of point measurements, particularly in highly variable environments such as shallow and deep convection. This can introduce uncertainty into the observed properties used to assess the simulations and in the skill score calculations. A first step to addressing this issue has been to determine the variance computed from distributed sensor networks. Examples include LCL, computed from ARM and Mesonet sensors, and cloud-base height, determined from the Doppler lidar network. The standard deviations of these measurements are presented in the diagnostic plots for shallow convection and have been helpful for gauging simulation performance while considering the variability. However, an attempt was not made to try to propagate this information through the skill scores. One example of how this might be accomplished to first order would be to apply $\pm 1\ \sigma$ to the time series data before computing the Taylor skill score to obtain a range in the skill scores. More challenging is how to address the sampling uncertainty of single measurements, such as LWP from the SGP CF. One approach could have been to apply the frozen turbulence assumption to convert variability in the time dimension into variability in the spatial dimension and parse the LWP time series into LES-equivalent grid sizes for comparison. Still, this assumes that the true variability is captured within the limiting sampling, which may not be the case. It might be possible to leverage the variability found in the simulations to estimate the potential effects on the observations, but this assumes a level of model fidelity that might be challenging to accomplish (see earlier discussion regarding the forcings). The model configurations also neglect some processes contributing to observed heterogeneity—for example, a homogeneous surface was used for the shallow convection simulations—and thus, the simulations must be interpreted fairly for what they can offer. The question of observational uncertainty, especially sampling representativeness, has been longstanding within the community; it is clearly important, but lacks a straightforward solution.

4.3 Community Support and Engagement

Engagement and support are key to making the user community aware of LASSO products and educating them on their use. The approaches used thus far have involved online documentation, announcements through ARM newsletters and the LASSO listserv, conference presentations and breakout sessions at the ARM/ASR PI meeting, as well as targeted workshops and tutorials. While the in-person sessions have been very well attended, the team has noticed a need to regularly re-educate users on an individual basis on what exactly LASSO provides. This, along with the general desire for maximal impact of LASSO in the community, raises the question as to whether something more or different could be done to better cultivate and educate users.

Recently, an online Discourse forum was created and maintained for frequently asked questions (FAQs).[4] However, users have not been adopting it and instead have been emailing the LASSO team via

---

[4] ARM Discourse Forum: https://discourse.arm.gov

lasso@arm.gov, which does not support building the knowledge repository that may aid future users. A recommendation from the workshop was that users should be pushed to the forum by only answering questions posted there, thereby building the desired resource.

Further, online scenario documentation, until recently, was provided via static PDF reports. Such reports were difficult to update in a timely fashion, involving a formal publication review prior to their release. As such, the dissemination of information would wait until it reached maturity worthy of this process. However, evolving products like LASSO benefit from a more agile documentation approach; for CACTI, LASSO has worked with Will Provenza, Pete Eby, and Zach Price to implement a sphinx-based report format. Not only does this make it easier to evolve the central part of the documentation, but it opens up avenues to capture information dynamically as it occurs, such as suggested edits from users.

For example, the suggestion of using LASSO to define data epochs would benefit from the ease of listing the corresponding new PI data sets as they become available. Along similar lines, it was noted that sharing of use cases can stimulate research, as there is a tendency for users to cluster around particular cases as they have become demystified and known to hold value through papers published. Capturing this information within the documentation is one way to make such cases and their uses known. Other means to promote such opportunities may include news/blog articles or perhaps a research forum discussion board. For example, it was noted that the valuable restart capability for studies requiring more frequent output should be advertised. It was also noted that LASSO is advertised on the website as a modeling activity, not a modeling-observation activity. Such distinctions could be made clearer on a discussion board without becoming lost within the volume of material needed to communicate the overall scenario documentation.

It was suggested that a more involved user base might be cultivated through more extensive community interactions during the design and implementation process. LASSO has used listening sessions at different stages of development, which have provided valuable input at the granularity appropriate to that process stage. Typically, better input is received at the early stages of development when things involve more broad brushstrokes versus closer to production when information is desired on the detailed specifics of how users want the data parsed for effective access and distribution. While currently we typically provide notification of when the data are ready, to build interest LASSO could provide more regular updates on interim status. Involving early adopters during development could help ensure maximum usefulness of the product. This, for example, could include access to the early simulations—something we have been reluctant to try because of the potential complications posed by the possible changing of the simulation parameters that can occur during the design phase, which would affect their research throughput. However, there is likely a point between beginning development and fully baked where a researcher could meaningfully engage.

Finally, a recommendation was made to help inform priorities by conducting a user survey to examine what parts of LASSO are being used and what impediments might exist. For example, it is speculated that most of the user base is modelers who use LASSO’s vetted forcings as a starting point for their own simulations. Are there reasons that observationalists are not engaging the data sets more and/or impediments that could be removed? If that (or others) is not to be a prominent user group, can aspects of LASSO be simplified by addressing the needs of the dominant groups?

4.4 Computing and Data Access

Computing for LASSO has two aspects: one for the computing resources required to generate the data set and one from the user perspective. This section addresses the user perspective: Section 5.3 covers the operational aspect.

ARM's approach for accessing LASSO data has evolved with LASSO's development. The first LASSO scenario for shallow convection involved packaging data into a series of tar files that ranged from 70 MB for the observations and hourly sampled model data to 70 GB for the raw model output. Since these are relatively small in the context of a single simulation when working on compute clusters, ARM's primary goal was to simplify selection of the simulations and then stage the requested data for user download. It was assumed users would have sufficient resources to work with the LASSO shallow convection data set. This worked for most users, except those working with many simulations, which can become too large to contain on a desktop computer. The other difficulty occasionally reported was problems downloading the multi-gigabyte tar files. The download problems typically were remedied by guiding users to use Globus for the file transfers, but not all users have Globus access.

The development of LASSO-CACTI portended a dramatic increase in data set size. Single-simulation sets of files grew to 10s of terabytes, which when working with multiple simulations, stretches even one's ability to hold the data online for leadership-class high-performance computing (HPC) systems, such as the DOE Perlmutter computer commonly used by DOE-funded atmospheric scientists. In anticipation, ARM purchased additional HPC compute nodes on Cumulus and prepared a new approach to work with LASSO data. Users can still preview aspects of the simulations on the LASSO-CACTI Bundle Browser to identify simulations of interest and download files to their location of choosing, but there is now the option to keep the files within ARM's computing infrastructure and work on Cumulus. As noted in Section 3.2.1, anyone working with the ARM data can now request an ARM computing account to work on a Jupyter server or directly on Cumulus. Upon ordering the LASSO data, Globus can be used to stage the data on Cumulus if it is not already present. Also, portions of the data are currently being pre-staged to prevent users making multiple copies of the large data set and simplifying overall use.

A small number of early adopters availed themselves of the opportunity to work with both the shallow convection and deep convection scenarios on Cumulus. The feedback so far is that this is a valuable option that greatly simplifies their work. For example, the PNNL team working on CACTI, noted in Section 3.2.2, have done most of their analyses and additional WRF simulations on Cumulus. Only recently have they branched out to do more of their LASSO-CACTI work on Perlmutter due to Cumulus becoming overly busy the last quarter of 2023, preventing timely access to do work requiring real-time access, such as plotting. A team from NCAR led by Hailey Shin also has used Cumulus and had a good experience working with the shallow convection data set. The most common frustration with Cumulus is the complicated management structure to establish new accounts and support requests, which involves a mix of ARM and Oak Ridge Leadership Computing Facility (OLCF) personnel. Requests can get lost or stalled and the LASSO team has to step in and help guide users to the correct system administrators to resolve a request.

Lessons learned regarding computing and accessing LASSO data are:

- The Bundle Browser interfaces are an important aspect of accessing LASSO data sets and are critical to work around limitations in the current Data Discovery interface. Availability of the skill scores and previsualizing the data to narrow down the requested simulations combined with making the ordering process efficient are the important aspects of the browsers.

- The dynamic plot generation for the shallow convection browser is a value-added feature, but time has shown it to be problematic for system maintenance. It occasionally requires rebooting servers to refresh the database. It also can return inconsistent results when requesting many points on the plots, sometimes not showing portions of the data. Therefore, the deep-convection browser has eschewed the dynamic plots in favor of static plots in hopes of increasing the webpage uptime and consistency, thereby avoiding the additional maintenance overhead.
- Globus is a critical resource when working with LASSO. It is used both for transferring files within and outside of OLCF computers. SSH transfers are too slow for large files and HTTP transfers crash too easily.
- Providing HPC access for using LASSO data sets speeds user adoption of the data set. However, the slow processing of new accounts frustrates users.
- More detailed documentation from ARM is needed about how to run jobs, particularly for more complicated situations like running a Jupyter notebook on interactive compute nodes on Cumulus.

Several future-focused topics were also discussed during the workshop for computing details. One is the file format used by LASSO, which currently relies heavily on netCDF (Unidata 2023). This works well and can be generically read with many software libraries such as xarray (Hoyer and Hamman 2017). However, netCDF is not designed for cloud-based computing environments. Future LASSO data could benefit from using file formats such as Zarr[5]. Zarr makes accessing data easier for remote subsetting and also chunks data differently than netCDF, which can facilitate efficient parallel processing in certain computing environments.

Data compression also is important and highly relevant for LASSO due to the large file sizes. Some of the LASSO files take advantage of the compression options built into the netCDF4 file format. However, this is not used for all files and the aggressiveness of compression has not been extensively explored. This should be considered when exploring future file format options.

Another requested capability is linking user-generated data sets to LASSO to enable other users to take advantage of community contributions and clarify data provenance (e.g., see discussion of epochs in Section 4.2). Multiple derivative data sets have already been produced, and it would be beneficial to make them findable from within ARM's search tools. Some of the data sets would also be good to host directly from within ARM, such as re-runs of LASSO-CACTI simulations for high-frequency output. ARM provided the computing resources for users to generate this data, publications require the data to be archived, and doing so within ARM would enable others to easily use this data. The cloud tracking product by Feng et al. (2023) is another example that would benefit many different users. The simplest approach at this point would be to share derivative data sets as ARM PI data sets. Then, these data could be referenced both from the browser web pages and the LASSO documentation. The newly used sphinx report format for the web-based LASSO-CACTI technical report makes this particularly easy compared to the earlier fixed-PDF reporting approach. Minor changes to the report can be made at any time and uploaded to the web-based report.

---

[5] Zarr website: https://zarr.dev/

The difficulty lies in deciding which data are appropriate for ARM to archive. Hard choices had to be made to limit the number of simulations to retain and the types of data to save for each simulation when designing LASSO-CACTI. ARM does not own sufficient tape and hard-drive space to host as many simulations as desired. So, if many large derivative data sets get generated, a review process will be needed.

A suggestion was made to provide coarsened data in addition to data on the raw model grid. This feature would reduce the amount of data users would need to download while allowing them to still retain most of the benefits of the fine-resolution modeling. For example, model physics and dynamics are better resolved at 100 m, but many users do not need data with that grid spacing for their applications. They might be happy with the model output sampled to a 1-km grid spacing. Tools could be developed to request model output on various sampling grids, and the raw model output interpolated to an alternate grid when staging the data for download. Alternatively, code could be provided showing users how to do coarsening themselves, which would not shrink the download size for them, but it could make it easier for them to work with smaller data files for analysis.

4.5 Connections to Large-Scale Modelers and Institutional Modeling

A key goal for LASSO is to connect ARM observations to large-scale modeling, such as the climate models developed by DOE. Other examples are the numerical weather forecast community who continually seek to improve their forecasts (see Figure 5). Historically, this has been difficult due to the scale gap between ARM's local observations and the larger grid sizes employed by global climate models (GCMs) and NWP models. The LES modeling from LASSO helps bridge this gap by providing information on intermediary scales. Several examples were shared at the workshop of how this has happened to date, and extensive discussion occurred for how to better employ LASSO within the E3SM community.

As noted in Section 3.1.2, the shallow convection scenario has been incorporated into the CCPP SCM that is part of the Global Model Test Bed (GMTB) from NOAA's Development Testbed Center (Heinzeller et al. 2023). This is a clear example of how using the LES provides a direct connection between observations up to coarse model scales for testing physics parameterizations. The general paradigm is to use the LES and accompanying observations as a benchmark and representation of reality to compare with parameterization behavior in the SCM framework. Identical forcing information can be applied to both the LES and the SCM, which increases the ability to do direct comparisons between them. Furthermore, the LES has been compared to the observations through LASSO's diagnostic skill scores to give confidence as to a simulation's overall realism. By including LASSO in the library of forcings available, users can easily incorporate LASSO into their workflows.

A similar approach is planned for including LASSO in the E3SM SCM library of cases. To date, a selection of LASSO-ShCu simulations have been ported over to the E3SM library with plans to make additional simulations readily available. As shared during the workshop, there are also plans to use LASSO within the THREAD Project, an ASR Science Focus Area at Lawrence Livermore National Laboratory (LLNL). THREAD is tasked with diagnosing biases in high-resolution E3SM and improving performance in various convection and cloud regimes. The already available LASSO-ShCu and CACTI simulations will be used for evaluating continental shallow convection and orographically forced convection.

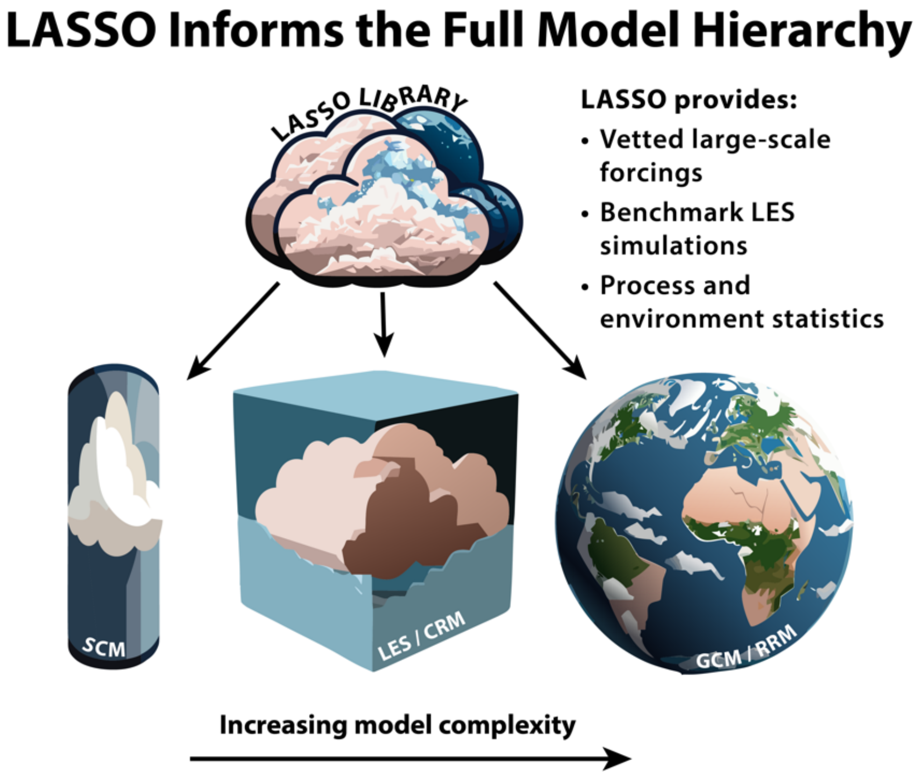


Figure 5. The LASSO library of simulations provides a range of information for use across the full range of model complexity from single-column models to cloud-resolving models and global climate models.

A long discussion considered how a LASSO scenario for the ENA atmospheric observatory and other future scenarios could be incorporated into an E3SM evaluation workflow. E3SM now has many modes of operation that include SCM, DP-SCREAM that is doubly periodic across a range of possible grid spacings from O(1-100 km), regionally refined with targeted high-resolution regions with O(1 km) grid spacing, traditional GCM with O(10-100 km) uniform grid spacing, and SCREAM as a high-resolution global model with O(1 km) uniform grid spacing. Each of these poses different evaluation needs. The range in grid spacing also causes deficiencies in parameterizations to manifest differently across grid spacings due to inadequate grid-scale awareness. Physics parameterizations must perform accurately from SCREAM's target grid spacing of 3 km up to the 100 km used by the traditional GCM for multi-century simulations. The regionally refined model (RRM) mode can include all these scales within one simulation grid.

The LASSO LES simulations can be used as benchmarks for the SCM and DP-SCREAM operational modes since the LASSO large-scale forcing can be used to make simulations that are fair to compare between LASSO and E3SM. Further, the extra detail provided by the LES is one way to inform subgrid variability potentially hidden within the coarser E3SM model. This could be useful to inform parameterization development and diagnose why parameterizations have trouble with certain scales and/or processes.

The situation changes once one moves to a global model domain for RRM, traditional, or SCREAM operational modes. The RRM and SCREAM modes have the benefit of being closer to the LES resolution and can compare to the LES results directly once considerations are made to make the comparison fair, such as coarsening the LES output to match the E3SM grid. The important difference compared to the column-based forcings used with the SCM is to constrain the synoptic scales to match between LASSO and E3SM. A straightforward way to do this is to nudge most of the globe to the same (re)analysis forcing used by LASSO for the case being compared, while leaving the local region around the LASSO domain free running. If done to constrain similar scales within E3SM as provided to the LASSO LES, this could make for a useful comparison to evaluate the physics behavior of the full global domain. One potential advantage of this approach over the SCM and DP-SCREAM modes is that it would permit more complicated situations to be simulated, such as present in the LASSO-CACTI scenario where the convection is orographically forced. Capturing orographically forced convection within a SCM or doubly periodic domain for the correct reasons is very difficult due to the inability to include realistic terrain. This limitation would not exist in the global domain. The problem then would become sufficiently resolving the domain to capture the same local terrain influences as present for LASSO.

Overall, the suggestion was made for the LASSO developers to work closely with the E3SM community to find ways to integrate LASSO into the E3SM suite of evaluation tools and test the E3SM hierarchy using some LASSO-like simulations in the future. Model testing by the E3SM developers can then provide details for ways LASSO could be enhanced to meet their needs; further, the E3SM community can be more informed as to the available data from ARM. This should be easier with funding specifically targeted for this purpose, e.g., through the LLNL THREAD SFA. It was noted that full-season SCREAM simulations should be available in a year, which could be used to identify meteorological regimes where SCREAM struggles and could benefit from a future LASSO scenario.

Another suggestion was to consider model components beyond the atmosphere. Weather and climate models contain many interacting components, of which ARM data can inform several. Specifically noted were the atmosphere, land, and aerosol portions of the models. Providing LES or other modeling approaches to inform these processes could be useful.

Providing generic large-scale forcing data for the LASSO cases is a straightforward way to make LASSO more widely applicable. As noted above in this section, this would permit running other models and sensitivity studies using the same forcing as was used for LASSO, thus permitting use of the LASSO simulations as a benchmark, or at least a first estimate of expected behavior. To date, the LASSO-ShCu scenario provides large-scale forcing and input data that works for the LASSO-modified version of WRF. This is combined with a Python script in the data bundle to convert this input into the format used by the System for Atmospheric Modeling (SAM) model (Khairoutdinov and Randall 2003). This script serves as an example of how one might adapt the provided forcings for other models as well. External to LASSO, the GMTB provides its own script to convert LASSO forcing into the format used by the CCPP SCM. It was suggested at the workshop that LASSO consider providing forcing data in the newer DEPHY-SCM[6] format embraced by the Cold-Air Outbreaks in the Marine Boundary Layer Experiment (COMBLE) (Geerts et al. 2022) Model–Observation Intercomparison (COMBLE-MIP).[7] The DEPHY-SCM name derives from the project that initiated the standard, Développement et Évaluation PHYsiques des modèles de climat et prévision du temps, funded by the France’s Centre National de Recherches Météorologiques.[8]

---

[6] DEPHY-SCM standard: https://github.com/GdR-DEPHY/DEPHY-SCM, accessed 9-Jan-2024

[7] COMBLE Model-Observation Intercomparison Project home page: https://arm-development.github.io/comble-mip/README.html, accessed 9-Jan-2024

[8] DEPHY project overview: https://www.umr-cnrm.fr/spip.php?article930&lang=en, accessed 9-Jan-2024

The DEPHY-SCM format defines a community standard for formatting 1D SCM input, which would be appropriate for the LASSO-ShCu data. While it may not be feasible to reprocess all the existing 1D LASSO-ShCu forcing data, it would be possible to provide a conversion script like the one provided for converting to the SAM format. This format could also be considered for future scenarios using 1D forcing data.

Another useful tool that LASSO could provide is software to generate column-based, 1D forcing from nested-grid simulations, such as used for LASSO-CACTI. SCM and doubly periodic domains are not appropriate for many situations, and thus LASSO has a mix of domain types depending on the scenario. We anticipate that this will continue to occur. It could be possible to provide code to run on the raw model output from nested LASSO simulations to pull 1D forcings from them for various spatial scales. These would need to be used wisely, such as over flat regions with somewhat uniform large-scale conditions. When this could be used, it would enable broadening the applicability of the LASSO nested scenarios for use with the simpler model configurations, and thus a broader swath of the overall modeling hierarchy.

5.0 Approaches for Near-Term Development Opportunities

5.1 Future Science Opportunities Not Currently Addressed by LASSO

Time was assigned to discuss future LASSO scenarios, what they might look like, and the science topics that developers could use to focus the scenario development. Three categories received the most discussion time: aerosol-cloud interactions, terrestrial and ecosystem science, and a more general topic around running scenarios across a broader range of conditions at a particular site.

Interest in integrating ARM aerosol observations with LASSO has been expressed since LASSO first began. Recent field campaigns with data appropriate for this science topic include the Tracking Aerosol Convection Interactions Experiment (TRACER), Eastern Pacific Cloud Aerosol Precipitation Experiment (EPCAPE), and the upcoming BNF campaign. The primary questions have always been: 1) how to address aerosols in a general enough way to benefit a wide swath of aerosol and/or aerosol-cloud researchers, and 2) what is the most cost-appropriate way to do high-resolution modeling of aerosol in a production setting? Researchers using aerosol model output draw from a hierarchy of aerosol model complexity, which then need to be integrated within the hierarchy of atmospheric model configurations. Aerosol models range from simple bulk formulations tied to climatological trace gas concentrations all the way to detailed particle-resolved aerosol models with fully interactive gas-aerosol components. These components are then integrated within a host atmospheric model, which could be a simple box model or a regional or global atmospheric model. The costs can vary by orders of magnitude depending on the choices made to address certain science questions.

Aerosol modeling, and aerosol-cloud interactions (ACI) in particular, bring with them unique complications absent for simpler cloud modeling. Science questions around ACI investigate subtle processes where the signal is hard to detect, and other processes correlated with changes in aerosol conditions confound interpretation. Statistically disentangling the signal requires large sample counts and the ability to conditionally sample data sets to separate impacts of changing aerosol from changing meteorological and other influences on the cloud state (McCoy et al. 2020; Glenn et al. 2020a). While LASSO aims to provide a library of case dates for researchers, to date the number of cases has been limited due to overall cost and, in the case of CACTI, by the number of days meeting the established conditions for case selection. Even with the 95 shallow convection cases, one may not be able to adequately disambiguate aerosol-cloud influences. This is partially due to the selective sampling used when selecting LASSO cases, which in the case of shallow convection, required criteria such as homogeneous synoptic conditions around the SGP CF. It is possible that this biases the sampled ACI, e.g., by reducing the occurrence of wind direction from particular quadrants, and thus biasing the pollution characteristics reaching SGP from certain emission sites. Differences in meteorological conditions might also not be sufficiently sampled to account for changes in meteorology that co-vary with aerosol.

One proposed approach to providing useful information from LASSO to ACI researchers is to use LASSO to provide baseline simulations from which the research community can build their research. This follows from how many researchers have made use of LASSO-ShCu, where they start with LASSO and then supplement it with additional simulations of their own. In the context of ACI, an example is Glenn et al. (2020b) who started with LASSO cases, but re-ran the simulations with the addition of temporo-spatially varying aerosol conditions. An aerosol field was estimated based on observations and connected with the microphysics within their model. Thus, they used a more complex model than LASSO (which assumed a constant aerosol value across all cases) but were able to glean anticipated general model behavior from the initial LASSO simulations. Depending on the ACI interactions of interest to a researcher, the added complexity for additional simulations could range from the basic approach of Glenn et al. up to the fully interactive aerosol models, such as available in the WRF coupled with chemistry (WRF-Chem) model. Future scenarios for LASSO could take a more nuanced approach to aerosol than used for LASSO-ShCu. The LASSO-CACTI scenario added a spatiotemporally varying aerosol field for the Thompson microphysics, based on aerosol information from the GEOS5 model, but still did not include a formal aerosol model within the simulations. Another potential added value that LASSO could provide is emission data for the region encompassed by LASSO simulations. Whether this would be of sufficient value to researchers needs to be ascertained.

If LASSO were to directly address aerosol-cloud processes, the choice of region becomes important to align with a strong signal. For example, aerosol is thought to be important in transitions of maritime cloud states, which makes ENA a good ARM location for this kind of scenario (Kazil et al. 2011). To adequately capture the physical processes important for the different cloud organization and aerosol states, one would need to include aerosol scavenging combined with local emission sources, such as sea salt and dimethylsulfide (DMS), to combine with an aerosol process module. LASSO could also consider the aerosol processes outside of clouds, of which TRACER could be a possible campaign for a scenario. In this case, the question about the level of complexity of the aerosol model becomes a defining characteristic of the scenario since it would determine what processes are targeted, e.g., secondary organic aerosol production, particle nucleation, and/or sulphate formation.

The EPCAPE field campaign was discussed as a possible scenario for targeting ACI. EPCAPE has maritime clouds, but its location is more strongly influenced by localized phenomena than ENA. A defining aspect of EPCAPE is its coastal location on the Scripps pier with different aerosol conditions for inflow from the open Pacific Ocean versus the Los Angeles Basin to the north, local San Diego emissions from the east, and Mexico to the south. The presence of land-sea breezes on many days, which are embedded within mesoscale circulations influenced by the nearby hills, also add complications for this location. A clear plan would be needed for an EPCAPE scenario that acknowledges these different regional influences and how LASSO would either target them or try to avoid certain conditions. The plan would need to balance providing simulations general enough to be used by a wide range of researchers with the smaller number of users that would need simulations to target understanding specific EPCAPE observations. Both uses are valid. Traditionally, ARM has attempted to make LASSO more general, but a choice could be made to use LASSO for local understanding as well. This might broaden the use of the campaign observations.

Designing a scenario around the new BNF deployment was also discussed extensively. BNF offers potential for LASSO to contribute value with observations for clouds, aerosol, and ecosystem processes planned over the next five years. The terrestrial ecosystem focus differentiates BNF from prior ARM deployments, and this topic would require a different modeling approach for LASSO. ARM could approach this either with a focus specifically on the terrestrial processes, or on interactions between the ecosystem and atmosphere. An example of a terrestrial ecosystem model that could be employed to simulate canopy emissions is the Ecosystem Demography model v2.2 (ED2; Longo et al. 2019). Alternatively, ARM could focus on airflow in and out of the forest with detailed LES modeling configured to capture turbulence and fluxes throughout the canopy. Forest emissions and their impact on secondary organic aerosol (SOA) is another area of interest. The traditional cloud foci of LASSO could also be used to motivate a BNF scenario. Many options exist, each with a unique modeling approach that would need to be designed around the science questions chosen to motivate the scenario. Various advantages exist for the options. Work on canopy fluxes would open new opportunities for partnering with the DOE Environmental System Science Program. Continuing a cloud focus would permit leveraging work from prior LASSO scenarios while creating a more varied library of cloud cases. An aerosol focus could address key questions regarding plant controls on aerosols.

A different approach to using LASSO would be to design simulations specifically to understand observations. While simulations used for instrument placement and site characterization prior to a field campaign were deemed most appropriate to be done by the team proposing the campaign as part of their planning process, simulations during or after the campaign could be useful to elucidate case-specific meteorological and land-surface features. Examples include frontal passages or other storm events, terrain-induced flows tied to different synoptic conditions, and comparisons between different cloud regimes. Along this line of thinking is running LASSO simulations in near-real time as proposed in the BNF white paper. This could focus attention on different case dates as they occur, inform processing of observations, and provide a broader range of cases than for just one meteorological regime. A staged approach was discussed that could run a cheap model quickly and save only a small subset of the model output to enable running daily. Then, cases would be selected for more detailed modeling closer to what has been done for the existing scenarios. Implications of running in near-real time are discussed in Section 5.3.

One area discussed as inappropriate for LASSO scenarios would be to use LASSO to generate a range of perturbation studies: for example, parametric sensitivity studies with the goal of defining the sensitivity. While LASSO sometimes adjusts model parameters and inputs to improve model accuracy for a given situation, this differs from intentionally trying to span a phase space for a given parameter, aerosol inputs, or other process. This type of work is more research focused and appropriate for other projects. Typically, doing perturbation studies targets specific science questions, and as a user facility, ARM focuses on providing general modeling data for the community to use, as opposed to addressing very specific science questions. In the context of LASSO, it was noted that perturbations to better match simulations to observations can be useful, while perturbations to generate hypothetical scenarios encroaches on research and moves away from a production data set.

5.2 Modeling Approaches

Discussion of modeling approaches applicable to LASSO occurred throughout the workshop, as this topic is tightly coupled to the science drivers, model selection, and computational cost tied to producing LASSO. Specifically, the primary aspects of the modeling approach of concern are the type of domain configuration to use and the specific model. This section notes key points for these aspects.

The two current scenarios demonstrate very different modeling approaches, which were chosen based on respective target cloud regimes and regions of each scenario. LASSO-ShCu uses a doubly periodic domain with a homogeneous lower boundary, while LASSO-CACTI uses a set of nested domains with topography and an interactive soil model. Choosing these approaches was primarily driven by the size and population of shallow compared to deep clouds, with shallow clouds only needing a small domain and deep convection needing a large domain to capture more of its life cycle and extent. The dominant impact of the Sierras de Córdoba on the CACTI convection also mandated inclusion of terrain within the model.

It was noted that the ability to use simultaneous, nested LES domains within a single simulation has evolved significantly since LASSO began. Nested LES domains was considered an experimental and unproven technique when first proposing the LASSO activity, and the choice was made to avoid the nested approach at the time since the production environment of ARM requires confidence in the products. Later, nested LES became more established and was adopted for LASSO-CACTI. Over this time, additional techniques have been developed to aid with nested domains that could be employed by LASSO. Physically consistent methodologies to apply inflow perturbations to the lateral boundaries to spin up turbulence can be especially useful (Muñoz-Esparza et al. 2014).

Even with the greater use of nested LES, it is an expensive methodology due to the often larger domain sizes and more sophisticated configurations used with it than for periodic domains, which precludes many researchers from taking advantage of the approach. Even those that can afford such expensive simulations may not always have the expertise needed to do nested LES well. Therefore, the ability of ARM to produce these types of simulations for the community provides an important service to researchers. Having ARM provide the initial testing of boundary conditions, work through difficulties associated with modeling ARM locations, and evaluating the simulations with available observations provides a strong foundation upon which researchers can do further modeling without having to spend time and resources on the preliminary details. Many researchers can also use the ARM simulations directly, saving them from having to run any simulations themselves.

An area recommended for the LASSO developers to investigate is providing single-column forcings from the nested LES model, or its parent domain's, output. This would enable a wider use of the nested simulations, such as for partnering with SCMs or doubly periodic LES models. As noted in Section 4.5, this could be done by providing a script to apply to the provided model output from LASSO. Within the context of nested, non-periodic LES domains, testing will be needed to identify how well the forcing works for various situations. For example, a fairly homogenous, flat region such as the ocean around ENA would be an easier location to make self-consistent forcings between the two modeling approaches, whereas CACTI with steep mountains and large cloud scales would be technically difficult, if not impossible.

A defining characteristic of the LASSO-CACTI scenario is the multi-scale approach with coarser ensembles for many days combined with LES for selected days. While one may envision an alternate approach with both mesoscale and LES simulations for all simulated days, the multi-scale approach used for the scenario permits much greater variability in the types of convection presented to users, and it serves the practical purpose of informing which days would likely be easier to simulate with LES. The inclusion of kilometer-scale simulations also simplifies more direct comparisons with models such as SCREAM, which has 3-km grid spacing. This permits E3SM developers to perform SCREAM simulations for some of the LASSO cases to facilitate comparison.

Looking to the future, another science driver motivating a multi-scale modeling approach is the strong desire for simulations capturing cloud organization for the upcoming LASSO-ENA scenario currently under development. Capturing the various cloud features of interest at ENA requires both detailed, high-resolution simulations for simulating precipitation characteristics and larger domains for simulating cloud organization that occurs on the order of 100 km. The latter requirement necessitates domains approaching or larger than 500 km across, which are too expensive to simulate with sub100-m grid spacing. Even if computing turnaround time were acceptable for this, the number of grid columns would produce too much data for ARM to archive for a reasonable number of cases. Using coarser grid spacings for large-area domains is an acceptable compromise when combined with a smaller number of LES simulations to address the cloud-scale precipitation processes. It also argues for a strategy of including multi-scale modeling in any hierarchy of LASSO simulations. Using an Earth system model, e.g., E3SM with its refined domain and nesting capabilities, could also bring along the availability of aerosol models for ACI-focused simulations.

Discussion also addressed aspects of choosing Eulerian versus Lagrangian domains. Existing LASSO scenarios use a Eulerian approach with the domain fixed over the respective ARM instrument sites. Eulerian is the traditional approach for atmospheric modeling, is straightforward to define since one does not have to decide the domain trajectory, maximizes the domain colocation time with observations, and is the easiest to use for many research applications since one does not have to consider moving grid coordinates when doing analyses and data comparisons.

Alternatively, Lagrangian domains have benefits in certain situations and are becoming more popular for real-world, non-idealized maritime cloud modeling (e.g., Kazil et al. 2021; Erfani et al. 2022). Configuring the model domain to follow a defined track enables possibilities such as having the domain follow specific features over time, allowing them to evolve within the model simulation. This is particularly useful for cloud transitions where the changing conditions can be simulated using a more cost-affordable domain size than would be required to contain the entire cloud life cycle in a large Eulerian domain. Kazil et al. (2021) note that their Lagrangian simulations of pockets of open cells capture the evolution of the cloud morphology including the observed transition from open to closed cells, and their domain is sufficiently small to afford using spectral-bin microphysics, which is a big advantage when looking at ACI and cloud microphysical processes. The COMBLE-MIP also uses a Lagrangian model configuration for capturing the transitioning cloud state of air moving over the Norwegian Sea.[9]

---

[9] https://arm-development.github.io/comble-mip

Lagrangian domains are very beneficial for transitioning states since the physical processes driving the transition are more easily contained within the domain and time-evolving mesoscale circulations can set up without having to be passed through the domain boundaries, as is done for Eulerian domains not large enough to hold the entire cloud life cycle. For similar reasons, Lagrangian domains aid in spinning up initial cloud states in models where initial conditions contain insufficient detail of the mesoscale circulations and the model must spin up the circulations at the correct spatial scales. LASSO can learn and collaborate with these earlier efforts to improve the accuracy of the LASSO simulations, particularly where this would be facilitated by projects with DOE Biological and Environmental Research (BER) program funding, such as COMBLE-MIP.

Complications arising from use of Lagrangian domains are both technical and philosophical. Technically, one must define the trajectory for the domain to follow. This is easy if the atmosphere contains no shear, which is rarely the case. One must identify the most representative level to follow, which can vary depending on the situation. Some work is ongoing within ARM to produce trajectory products that LASSO could leverage. One is for the Multidisciplinary Drifting Observatory for the Study of Arctic Climate (MOSAiC) campaign and another is for ENA. Philosophically, Lagrangian domains can limit one's ability to compare simulations to static observations. Depending on the size of the domain and its track speed, the domain may only pass over the observations for a portion of the simulated period. If the observations are not representative of the larger surrounding region, which is quite possible for transitioning periods when Lagrangian domains have the most value, observations not coincident with the domain have decreased usefulness in the modeling context. Ultimately, one must prioritize accuracy of the simulation for the selected phenomena in conjunction with the available observations.

We note that Lagrangian domains can use either doubly periodic lateral boundary conditions or nested domains. The Kazil et al. (2021) and Erfani et al. (2022) examples both use doubly periodic boundaries with the SAM model. Alternatively, an example of moving domains with a nested domain approach is hurricane simulations where an inner domain follows the hurricane (Alaka et al. 2022).

Entwined with the domain configuration is model selection. While the science drivers around a scenario dictate the domain configuration, the different capabilities between models subsequently determine which models could be used for a given domain configuration. From this subset, one must choose the model to use based on additional criteria such as required features for physics parameterizations, input options, and computational efficiency. The current generation of ARM HPC does not contain graphics processing units (GPUs), so that simplifies the decision a bit since almost all models will run on central processing unit (CPU)-only architectures. For the near future on ARM HPC, this pushes the efficiency criteria to the numerical methodologies implemented in the models, which balance cost, accuracy, and ease of maintenance. To date, WRF has been used due to its broad range of features, which permit almost any required configuration[10] (Skamarock et al. 2008). However, WRF uses compressible dynamics, which is 5-10 times more computationally expensive than the anelastic dynamics used in models like SAM. This added cost could become too much for sub-100-m grid spacing anticipated for highest-resolution domains expected for LASSO-ENA. However, the SAM model does not readily handle nested domain configurations[11] (Khairoutdinov and Randall 2003). Alternatively, the Predicting Interactions of Aerosol and Clouds in Large-Eddy Simulation (PINACLES) model uses an anelastic dycore combined with more

---

[10] WRF website: https://www2.mmm.ucar.edu/wrf/users/, accessed 12-Jan-2024

[11] SAM website: http://rossby.msrc.sunysb.edu/~marat/SAM.html, accessed 12-Jan-2024

modern numerical algorithms[12] (Pressel and Sakaguchi 2021, Dhandapani et al. 2023). This improves its overall accuracy for a cost similar to SAM, and PINACLES can use nested domains like WRF. The risk with PINACLES is that it is a new model that has not been widely tested and some features are still being implemented, such as terrain. Other possible models ARM could use for LASSO are the MicroHH model[13] (van Heerwaarden et al. 2017), which is like SAM but has the advantage of running on GPUs as does FastEddy[14] (Sauer and Muñoz-Esparza 2020). Other models include the Cloud Model 1 (CM1) model[15] (Bryan and Fritsch 2002) or possibly European LES models such as PALM[16] (Maronga et al. 2020), Dutch Atmospheric Large-Eddy Simulation (DALES)[17] (Heus et al. 2010), or Icosahedral Nonhydrostatic (ICON)[18] (Zängl et al. 2015, Dipankar et al. 2015, Heinze et al. 2017). Each model has advantages and disadvantages to be evaluated on a scenario-by-scenario basis. A parallel option to integrate closer to other DOE projects could be to use an E3SM hierarchy of regionally refined and kilometer-scale versions as well as DP-SCREAM to parallel the "traditional" LES simulations. These are also not widely tested but would be extremely valuable to DOE model development for many applications.

5.3 Operational Considerations

The structural approach used by LASSO was discussed to determine whether there may be ways for optimization or perhaps consider using a different mode of operation, particularly to speed scenario throughput. The general workflow elements are currently as follows.

- *Initial setup* for modeling primarily involves determining the source and implementation of boundary conditions and forcings, sensitivity tests to determine domain configuration for the phenomenon of interest, and implementation of any model output enhancements. For observations, work primarily focuses on finding and quality-controlling reliable datastreams that are effective at discerning the critical model behavior (assessed using the model sensitivity test runs), determining comparable model and observed properties for comparison, and the approach for skill score implementation.
- *Case selection* involves identifying days when the phenomenon occurs and critical instrumentation was functioning for a variety of events to fill out the library.
- *Production* involves running the simulations and merging the output with the observations for the generation of diagnostic plots, skill scores, and their QC review.
- *Data distribution and documentation* involves making the data available with ARM-compliant variable names, DOIs, and any variable subsetting before staging the files for the ADC plus advising on revisions to the Bundle Browser.

---

[12] PINACLES website: https://github.com/pnnl/pinacles, accessed 12-Jan-2024
[13] MicroHH website: http://microhh.org/, accessed 12-Jan-2024
[14] FastEddy website: https://ral.ucar.edu/fasteddy, accessed 12-Jan-2024
[15] CM1 website: https://www2.mmm.ucar.edu/people/bryan/cm1/, accessed 12-Jan-2024
[16] PALM website: https://palm.muk.uni-hannover.de/trac/wiki/palm, accessed 12-Jan-2024
[17] DALES website: https://github.com/dalesteam/dales, accessed 12-Jan-2024
[18] ICON website: https://www.icon-model.org/, accessed 8-Apr-2024

Troubleshooting and the need to iterate possible approaches can come into play with essentially any of these elements. As LASSO has matured, some of the challenges of arriving at ARM-compliant variable names for the model output have largely been addressed, so future releases should require less effort.

Within the current operational paradigm, one way to reduce the time-to-product availability would be to use an iterative, staged modeling approach. Mesoscale simulations would be run expediently and assessed in a simple manner and then released, providing users with preliminary fields and modelers with the information on which of the forcings initially fared best. The more time-consuming LES are then initiated and released later. With the earlier availability of the mesoscale runs, user interest may be curated as the LES are being processed. As critical observations or input data for a higher-quality simulation or evaluation might not initially be available, a supplemental or re-release may be necessary later to update to the final version. While getting the data into the hands of users more quickly, this approach would involve more overhead with the multiple, staged data releases and the users would need to contend with the iterative data release.

Another way to reduce the workload and speed scenario production would be a minimal modeling approach. This would emphasize mesoscale runs and limit the number of LES. For example, providing a mesoscale simulation for a range of forcings routinely for multiple locations would be valuable. It was noted that most LES modelers would want to run their own model but would likely use the mesoscale output as is. LES modelers can run their model knowing which are the good forcings, as they often have with LASSO, and could even use the restart files (if using the same model) with LASSO doing the lift on the spin up. LASSO could run LES for one of the best-behaved ensemble members per day to provide the high-resolution data needed by users who lack sophisticated modeling abilities. It was mentioned that mesoscale is important to compare with km-scale grid spacing of models like SCREAM; with GCMs now operating at the mesoscale, mesoscale is a bridge to them. While modelers will tend to use their own model, they need well-characterized data that would come from the data bundle. The discussion, however, also recognized the importance of the large LES 'hero runs' that were done for CACTI. They can open doors for researchers and demonstrate possibilities to the community—this leaves a legacy and has a utility that ARM does the hard things others cannot easily do. Note too that if less emphasis were placed on LES-scales, comparison with ARM observations becomes more difficult as only bulk/averaged properties could be compared to km-scale simulations that would be problematic for heterogeneous phenomenon (e.g., shallow convection).

The BNF white paper proposed a semi-continuous modeling approach. The aim is to get a quick turnaround (near-real time) so that the simulations would be available closer to when events of interest occur to increase their value, particularly with respect to field campaigns. Operating in a more routine manner would address the previously discussed issue with selection bias. Also, processing of the tower flux data would benefit from the flow field information. Implementing this approach would involve determining how to reduce the data volumes to afford storing the increased frequency of runs while maintaining the forcing ensemble that has been determined to be critical to obtaining satisfactory simulations. It was noted that using a smaller domain and/or coarser grid would introduce more noise but might be sufficiently accurate to determine promising forcings to be released to modeling users; however, this likely would not be accurate enough for other user applications and would involve rerunning selected cases with a larger domain and/or finer grid spacing.

While LASSO was initially conceived to operate semi-continuously for a phenomenon, such as the seasonal shallow convection addressed at the SGP, there was a discussion of how one might more tightly couple LASSO to field campaigns and their subsequent funding cycles. This approach has the benefit of having a vested user group who could help drive scenario design decisions as well as later hone focus on specific cases of interest. The time from a campaign to the peak number of publications is about six years (LeMone 1983). This suggests that for LASSO to have the highest impact, the products would need to be out about 2-3 years after a campaign for their inclusion in proposals and research activities. If it is known ahead of time that LASSO will be run for a campaign, LASSO could be included within the campaign design and execution phases. However, doing so would carry the expectation that the campaign successfully captured the phenomenon with critical instrumentation as expected, when in practice natural variability might shift things to something slightly different. Given campaign uncertainty, it may be wise to wait until after the campaign to assess whether LASSO should be included. A sample timeline could be as follows:

1. Within a couple of months of the campaign, assess the benefit and viability of LASSO, considering the actual conditions sampled and instrument availability.
2. If a go, prioritize the processing of critical data sets/VAPs. Simultaneously, begin obtaining forcings data and determination of the model configuration. Coordinate case selection between the campaign science and LASSO teams.
3. Run and distribute the mesoscale products followed by LES for specific cases of interest.

The topic of running LASSO pre-campaign for Observing System Simulation Experiments (OSSEs) was discussed but generally viewed as impractical, being more of a research issue and the responsibility of the campaign proposers.

6.0 Summary and Conclusions

This “Future of LASSO Workshop” explored a wide range of topics on aspects of the current LASSO implementations and future possibilities. We conclude this report with the key points as summarized by the workshop organizers. Several drivers of our workshop discussions had clear conclusions, while others were left open-ended and will require further thought. The organizers indicate community confidence in each summary point as is possible.

So far, the LASSO activity has produced two different scenarios—one focused on shallow convection using a simplified LES configuration, and one focused on deep convection using a nested, downscaling LES configuration. While each scenario necessitated different model configurations to best address respective scientific drivers, both scenarios shared common aspects that define LASSO and set LASSO activities apart from other modeling activities. Specifically, each scenario used a LES library approach spanning a range of case dates that targeted a specific weather phenomenon. This approach was combined with ensembles of forcing data to capture the range of possible large-scale conditions leading to each event’s weather. These simulations were vetted with ARM and satellite data to anchor them in reality, assisting researchers in identifying simulations most representative of the observed real-world processes and enabling connection of the simulations to other ARM observations for analyses.

Workshop participants valued several aspects of LASSO, which are concepts that should be encouraged in future scenario decisions:

- The large computational expense of the ARM-funded simulations provides a resource for the community that most researchers would not be able to produce. This increases the number of researchers who can use the ARM observations for fundamental research in topics related to LASSO.
- The LASSO case library approach is unique as it moves beyond *golden day* mindsets, i.e., a focus on a single well-behaved or canonical case, by providing a range of vetted cases for researchers. This approach was viewed as valuable because it saves researcher time, yet still focuses the community on targeted days.
- The vetted forcing data sets were also considered valuable and among the most-used aspects of the LASSO products. These forcing data are commonly used to drive additional model simulations, often selected on the basis of foundational LASSO modeling and availability of ARM-informed validation data sets. A single forcing can have large uncertainty, so an observationally vetted ensemble is necessary to obtain good simulations.
- Implicit in the selection of the case library days and vetted forcing data sets is the value in the curation of ARM observations for each target scenario.
- The synergistic nature of the points above make LASSO a groundbreaking, unique data set. The foundational modeling provided by ARM saves researchers a lot of time by completing many of the initial steps required in new modeling studies—for example, narrowing down of potential case dates, evaluation of model configuration, testing of input and forcing data sets, and identification of potential observational data to pair with the simulations.

These aspects have an underlying theme in that they center on the large effort contributed by ARM to produce the libraries of simulations and paired observations. This enables others to accelerate their research by exploiting this foundational work, and the research can potentially address larger topics through the saved effort.

Some topics were discussed that could further enhance the impact of LASSO. These topics are lower effort than those listed in the grouping below, and thus will be useful to consider for LASSO's ongoing development These topics are:

- Increasing user engagement would benefit from enhanced communication from the LASSO team throughout scenario development. Suggestions for how to accomplish this included more frequent updates to keep users informed and engaged, and to continue using online tools or other agile ways for communicating new information, e.g., the Discourse forum and web-based documentation. This builds upon efforts undertaken for LASSO-CACTI and will be important for furthering LASSO-ENA this coming year.
- Users also may become engaged earlier in the process through the release of intermediate products, such as the mesoscale simulations, rather than waiting for the completion of the full data bundle for release. The approach would need to balance time to arrive at a stable configuration with the risk of the final release differing substantially from the intermediate release. We want to avoid confusing users, but their feedback is important for finding issues and improving the overall product while it is still easy to incorporate changes before doing a substantial number of simulations and preparing the data for release. Considerations of how to stage data releases will be incorporated into future scenario planning and development. Since taking time to do releases and promote them will increase effort, this will need to be planned for when budgeting and establishing delivery timelines.

- Incorporating LASSO scenarios into ARM's concept of data epochs would bring visibility to LASSO products and help LASSO users identify complementary data sets within the ARM archive. Epochs would help users coalesce resources around and build upon cases of interest, both from the ARM infrastructure and individual investigators. Epochs can come from two directions. The first would be in consultation with the community while picking case dates to target simulations around know cases of high value. The second would be through turning selected case dates into new epochs that would be useful for researchers to investigate. The newly developed web-based documentation would record and communicate epochs. Identified epochs lower the barrier to entry for users, who can have confidence that the case is 'good' and can see the degree that analyses can be extended successfully through the use-case examples from prior studies. While this seems simple, we anticipate that it will require integration across ARM. This is both why epochs have value—they integrate knowledge from many sources—and why they are not quick to do well. Once potential epochs are identified, it will take time for the added knowledge to build around the time periods.
- Related to epochs, LASSO's value would be enhanced through incorporation of user-contributed products or workflows into LASSO community materials. This topic includes finding ways to provide the necessary links between LASSO and its users that allow improved two-way benefits between LASSO and community contributions. Example contributions include analysis software, new observations or PI products for key events, and/or example pathways for additional simulations. The best way to connect LASSO and its community to these different product types or tools will need some thought, such as a vetting process, so that there is confidence in the products being linked, suitable documentation, and the means for distribution through the ADC. Discussion with the ADC and other stakeholders will begin this coming year to investigate options and to prioritize aspects for implementation over the coming years.

Additional important topics discussed that require further work and community input include the following list. These topics are high value, but are differentiated from the above list because they would require more effort, either from substantial community interaction and/or technical aspects.

- The related topics of *representativeness* and *selection biases* were discussed extensively. The community suggested that future work be done to inform users about representativeness to assist in appropriate use of the LASSO data bundles. Selection bias was a noted concern that should be made clear to users. For example, LASSO cases are selected based on certain criteria that likely limit some of the processes/variability and how best to represent and include "null" cases (days when the chosen phenomenon is expected to occur but does not) is unclear. However, how to meaningfully characterize the representativeness and selection bias is an open topic that has thus far evaded community consensus.
- Working closely with E3SM model developers to help them integrate LASSO into their workflows was suggested as beneficial to both E3SM and ARM. Initial inroads have included the LASSO team working with E3SM developers to use LASSO data with the E3SM single-column model. Additional networking can be done with DOE laboratory ASR-SFAs, such as the new THREAD Project, which is anticipated to develop a methodology to incorporate LASSO data. In addition, E3SM capabilities and resolutions are starting to overlap with LASSO modeling scales, and there are opportunities for using the E3SM hierarchy of models from the mesoscale down to LES to parallel LASSO and assist model development. This topic is a high priority and will receive increased attention. Discussions will be necessary to align project timelines and identify the best ways to do the data integrations in a way that maximizes usefulness. Increasing communication is a potentially high-impact and minimal-effort

first step for this topic, and developing the technical software links between the projects will require high effort. The LASSO and E3SM teams will further discuss partnering over the coming year.

- Accelerating scenario development would more directly link LASSO with ARM AMF field campaign deployments. How to reduce development time is unclear, given the need to balance aspects such as data availability and quality control, case curation, and the level of complexity of the LES. Workshop discussions were wary of exclusively limiting LASSO to cheaper, coarse simulations since that would reduce the demonstrated value of the current LASSO approach. Therefore, an open question remains as to how to produce LASSO scenarios more quickly while not making unacceptable compromises. Given that this goal is at odds with the emphasis on the value of data curation, highlighted in the first group of bullets above, the pursuit of rapid scenario development is currently considered to be a low priority.

- Finally, scenario selection is an ongoing topic that requires detailed planning specific to each scenario. How best to prioritize user desires is an open topic impacting both the scenarios to produce and their details. Top-level discussion topics included:
    - Aerosol-cloud interactions is of interest for several upcoming scenario options. However, how best to implement LASSO to encourage these themes was unclear given the wide range of pathways aerosols and clouds can be simulated depending on the research needs. General thoughts pointed to using LASSO to provide foundational modeling to establish forcing options using simpler modeling approaches that others could then build upon to add further complexity, such as more detailed microphysics parameterization.
    - Opinions differed over how to approach case selection for upcoming scenarios. One group of attendees strongly favored the current approach of curating high-priority case dates and focusing the modeling effort just on these cases. Another group preferred an approach of running simulations almost continuously during a campaign. A tension exists between producing “routine” simulations with less thought put into their selection and optimization of simulation accuracy versus a case-study mode where initial and boundary conditions plus model configuration details are a strong foci. Tradeoffs between these two approaches include the level of effort that could be given toward evaluating each case date, the level of complexity permitted in the modeling due to the large difference in number of cases, and the related issue of how much data can be saved per simulation to stay within the allotted quota for a scenario.

Overall, many options exist for LASSO going forward. Users have adopted LASSO-ShCu and have been publishing valuable research with it. The LASSO-CACTI scenario has been available for a shorter period and users are beginning to work with this data set, exploiting the computationally demanding library of simulations. Decisions are being made this year regarding the LASSO-ENA scenario, which will be run over the coming year. As this is done, initial choices will be made for the subsequent scenario and operational mode for LASSO and how to balance resources among the different possibilities discussed. Feedback from this workshop will be incorporated into LASSO planning and will inform prioritization when establishing the overall LASSO roadmap and associated tasks.

7.0 References

Alaka, GJ, X Zhang, and SG Gopalakrishnan. 2022. "High-Definition Hurricanes: Improving Forecasts with Storm-Following Nests." *Bulletin of the American Meteorological Society* 103 (3): E680–E703, https://doi.org/https://doi.org/10.1175/BAMS-D-20-0134.1

Angevine, WM, J Olson, J Kenyon, WI Gustafson, S Endo, K Suselj, and DD Turner. 2018. "Shallow Cumulus in WRF Parameterizations Evaluated against LASSO Large-Eddy Simulations." *Monthly Weather Review* 146 (12): 4303–4322, https://doi.org/10.1175/mwr-d-18-0115.1

Bryan, GH, and JM Fritsch. 2002. "A Benchmark Simulation for Moist Nonhydrostatic Numerical Models." *Monthly Weather Review* 130 (12): 2917–2928, https://doi.org/10.1175/1520-0493(2002)130<2917:ABSFMN>2.0.CO;2

Dhandapani, C, CM Kaul, KG Pressel, R Wood, and G Kulkarni. 2023. "Sensitivities of Large Eddy Simulations of Aerosol Plume Transport and Cloud Response." *ESS Open Archive*, https://doi.org/10.22541/essoar.170365352.28240067/v1

Dipankar, A, B Stevens, R Heinze, C Moseley, G Zängl, M Giorgetta, and S Brdar. 2015. "Large eddy simulation using the general circulation model ICON." *Journal of Advances in Modeling Earth Systems* 7 (3): 963–986, https://doi.org/https://doi.org/10.1002/2015MS000431

Endo, S, D Zhang, AM Vogelmann, P Kollias, K Lamer, M Oue, H Xiao, WI Gustafson Jr, and DM Romps. 2019. "Reconciling Differences Between Large-Eddy Simulations and Doppler Lidar Observations of Continental Shallow Cumulus Cloud-Base Vertical Velocity." *Geophysical Research Letters* 46 (20): 11539–11547, https://doi.org/10.1029/2019gl084893

Erfani, E, P Blossey, R Wood, J Mohrmann, SJ Doherty, M Wyant, and O Kuan-Ting. 2022. "Simulating Aerosol Lifecycle Impacts on the Subtropical Stratocumulus-to-Cumulus Transition Using Large-Eddy Simulations." *Journal of Geophysical Research-Atmospheres* 127 (21): e2022JD037258, https://doi.org/10.1029/2022JD037258

Feng, Z, J Hardin, HC Barnes, JF Li, LR Leung, A Varble, and ZX Zhang. 2023. "PyFLEXTRKR: a flexible feature tracking Python software for convective cloudanalysis." *Geoscientific Model Development* 16 (10): 2753–2776, https://doi.org/10.5194/gmd-16-2753-2023

Feng, Z, A Varble, J Hardin, J Marquis, A Hunzinger, ZX Zhang, and M Thieman. 2022. "Deep Convection Initiation, Growth, and Environments in the Complex Terrain of Central Argentina during CACTI." *Monthly Weather Review* 150 (5): 1135–1155, https://doi.org/10.1175/Mwr-D-21-0237.1

Geerts, B, SE Giangrande, GM McFarquhar, L Xue, SJ Abel, JM Comstock, S Crewell, PJ DeMott, K Ebell, P Field, TCJ Hill, A Hunzinger, MP Jensen, KL Johnson, TW Juliano, P Kollias, B Kosovic, C Lackner, E Luke, C Lüpkes, AA Matthews, R Neggers, M Ovchinnikov, H Powers, MD Shupe, T Spengler, BE Swanson, M Tjernström, AK Theisen, NA Wales, Y Wang, M Wendisch, and P Wu. 2022. "The COMBLE Campaign: A Study of Marine Boundary Layer Clouds in Arctic Cold-Air Outbreaks." *Bulletin of the American Meteorological Society* 103 (5): E1371–E1389, https://doi.org/10.1175/Bams-D-21-0044.1

Ghate, VP, P Kollias, S Crewell, AM Fridlind, T Heus, U Loehnert, M Maahn, GM McFarquhar, D Moisseev, M Oue, M Wendisch, and C Williams. 2019. The Second ARM Training and Science Application Event: Training the Next Generation of Atmospheric Scientists." *Bulletin of the American Meteorological Society* 100 (1): ES5–ES9, https://doi.org/10.1175/Bams-D-18-0242.1

Glenn, IB, G Feingold, JJ Gristey, and T Yamaguchi. 2020a. "Quantification of the Radiative Effect of Aerosol-Cloud-Interactions in Shallow Continental Cumulus Clouds." *Journal of Geophysical Research Atmospheres* 77(8): 2905

Glenn, IB, G Feingold, JJ Gristey, and T Yamaguchi. 2020b. "Quantification of the Radiative Effect of Aerosol–Cloud Interactions in Shallow Continental Cumulus Clouds." *Journal of the Atmospheric Sciences* 77 (8): 2905–2920. https://doi.org/https://doi.org/10.1175/JAS-D-19-0269.1

Gristey, JJ, G Feingold, IB Glenn, KS Schmidt, and H Chen. 2020. "Surface Solar Irradiance in Continental Shallow Cumulus Fields: Observations and Large-Eddy Simulation." *Journal of the Atmospheric Sciences* 77 (3): 1065–1080, https://doi.org/10.1175/JAS-D-19-0261.1.

Gustafson, WI, AM Vogelmann, X Cheng, KK Dumas, S Endo, KL Johnson, B Krishna, Z Li, T Toto, and H Xiao. 2019. *Description of the LASSO Data Bundles Product.* Atmospheric Radiation Measurement (ARM) Research Facility. DOE/SC-ARM-TR-216, https://doi.org/10.2172/1469590. https://www.arm.gov/publications/tech_reports/doe-sc-arm-tr-216.pdf

Gustafson, WI, AM Vogelmann, MM Delgado, S Endo, EK Schuman, AC Varble, and H Xiao. 2024. *Description of the LASSO-CACTI Activity: A LASSO Scenario for Orographically Forced Deep Convection.* Atmospheric Radiation Measurement (ARM) Research Facility. DOE/SC-ARM-TR-288, https://doi.org/10.2172/1905845. https://lasso-cacti-doc.arm.gov

Gustafson, WI, AM Vogelmann, Z Li, X Cheng, KK Dumas, S Endo, KL Johnson, B Krishna, T Fairless, and H Xiao. 2020. "The Large-Eddy Simulation (LES) Atmospheric Radiation Measurement (ARM) Symbiotic Simulation and Observation (LASSO) Activity for Continental Shallow Convection." *Bulletin of the American Meteorological Society* 101 (4): E462–E479, https://doi.org/10.1175/bams-d-19-0065.1

Gustafson, WI, AM Vogelmann, and JH Mather. 2019. *Science Drivers and Proposed Modeling Approaches for Future LASSO Scenarios, Report from the LASSO Expansion Workshop.* DOE Atmospheric Radiation Measurement Facility. DOE/SC-ARM-19-023, https://doi.org/10.2172/1569273. http://www.arm.gov/publications/programdocs/doe-sc-arm-19-023.pdf

Heinze, R, A Dipankar, CC Henken, C Moseley, O Sourdeval, S Trömel, X Xie, P Adamidis, F Ament, H Baars, C Barthlott, A Behrendt, U Blahak, S Bley, S Brdar, M Brueck, S Crewell, H Deneke, P Di Girolamo, R Evaristo, J Fischer, C Frank, P Friederichs, T Göcke, K Gorges, L Hande, M Hanke, A Hansen, H-C Hege, C Hoose, T Jahns, N Kalthoff, D Klocke, S Kneifel, P Knippertz, A Kuhn, T van Laar, A Macke, V Maurer, B Mayer, CI Meyer, SK Muppa, RA J. Neggers, E Orlandi, F Pantillon, B Pospichal, N Röber, L Scheck, A Seifert, P Seifert, F Senf, P Siligam, C Simmer, S Steinke, B Stevens, K Wapler, M Weniger, V Wulfmeyer, G Zängl, D Zhang, and J Quaas. 2017. "Large-eddy simulations over Germany using ICON: a comprehensive evaluation." *Quarterly Journal of the Royal Meteorological Society* 143 (702): 69–100, https://doi.org/https://doi.org/10.1002/qj.2947

Heinzeller, D, L Bernardet, G Firl, M Zhang, X Sun, and M Ek. 2023. "The Common Community Physics Package (CCPP) Framework v6." *Geoscientific Model Development* 16 (8): 2235–2259, https://doi.org/10.5194/gmd-16-2235-2023

Heus, T, CC van Heerwaarden, H. JJ Jonker, A Pier Siebesma, S Axelsen, K van den Dries, O Geoffroy, AF Moene, D Pino, SR de Roode, and J Vilà-Guerau de Arellano. 2010. "Formulation of the Dutch Atmospheric Large-Eddy Simulation (DALES) and overview of its applications." *Geosci. Model Dev.* 3 (2): 415–444. https://doi.org/10.5194/gmd-3-415-2010

Hoyer, S, and J Hamman. 2017. xarray: N-D labeled arrays and datasets in Python. Accessed April 8, 2024, https://doi.org/10.5334/jors.148

Kazil, J, MW Christensen, SJ Abel, T Yamaguchi, and G Feingold. 2021. "Realism of Lagrangian Large Eddy Simulations Driven by Renalysis Meteorology: Tracking a Pocket of Open Cells Under a Biomass Burning Aerosol Layer." *Journal of Advances in Modeling Earth Systems* 13 (12): e2021MS002664, https://doi.org/10.1029/2021MS002664

Kazil, J, H Wang, G Feingold, AD Clarke, JR Snider, and AR Bandy. 2011. "Modeling chemical and aerosol processes in the transition from closed to open cells during VOCALS-REx." *Atmospheric Chemistry And Physics* 11 (15): 7491–7514, https://doi.org/10.5194/acp-11-7491-2011

Khain, AP, KD Beheng, A Heymsfield, A Korolev, SO Krichak, Z Levin, M Pinsky, V Phillips, T Prabhakaran, A Teller, SC van den Heever, and J-I Yano. 2015. "Representation of microphysical processes in cloud-resolving models: Spectral (bin) microphysics versus bulk parameterization: BIN VS BULK." *Reviews of Geophysics* 53: 247-322, https://doi.org/10.1002/2014RG000468

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 (4): 607–625, https://doi.org/10.1175/1520-0469(2003)060<0607:Crmota>2.0.Co;2

Lebo, ZJ, and H Morrison. 2015. "Effects of Horizontal and Vertical Grid Spacing on Mixing in Simulated Squall Lines and Implications for Convective Strength and Structure." *Monthly Weather Review* 143 (11): 4355–4375, https://doi.org/10.1175/mwr-d-15-0154.1

LeMone, MA. 1983. "The Time between a Field Experiment and its Published Results." *Bulletin of the American Meteorological Society* 64 (6): 614–615, https://doi.org/https://doi.org/10.1175/1520-0477-64.6.614

Li, Z, S Feng, Y Liu, W Lin, M Zhang, T Toto, AM Vogelmann, and S Endo. 2015. "Development of fine-resolution analyses and expanded large-scale forcing properties: 1. Methodology and evaluation." *Journal of Geophysical Research: Atmospheres* 120 (2): 654–666, https://doi.org/10.1002/2014jd022245

Li, Z. J, XP Cheng, WI Gustafson, and AM Vogelmann. 2016. "Spectral characteristics of background error covariance and multiscale data assimilation." *International Journal for Numerical Methods in Fluids* 82 (12): 1035–1048, https://doi.org/10.1002/fld.4253

Li, ZJ, JC McWilliams, K Ide, and JD Farrara. 2015. "A Multiscale Variational Data Assimilation Scheme: Formulation and Illustration." *Monthly Weather Review* 143 (9): 3804–3822, https://doi.org/10.1175/mwr-d-14-00384.1

Longo, M, RG Knox, DM Medvigy, NM Levine, MC Dietze, Y Kim, ALS Swann, K Zhang, CR Rollinson, RL Bras, SC Wofsy, and PR Moorcroft. 2019. "The biophysics, ecology, and biogeochemistry of functionally diverse, vertically and horizontally heterogeneous ecosystems: the Ecosystem Demography model, version 2.2 – Part 1: Model description." *Geoscientific Model Development* 12 (10): 4309–4346, https://doi.org/10.5194/gmd-12-4309-2019

Mallaun, C, A Giez, GJ Mayr, and MW Rotach. 2019. "Subsiding shells and the distribution of up- and downdraughts in warm cumulus clouds over land." *Atmospheric Chemistry and Physics* 19 (15): 9769-9786, https://doi.org/10.5194/acp-19-9769-2019

Maronga, B, S Banzhaf, C Burmeister, T Esch, R Forkel, D Fröhlich, V Fuka, KF Gehrke, J Geletič, S Giersch, T Gronemeier, G Groß, W Heldens, A Hellsten, F Hoffmann, A Inagaki, E Kadasch, F Kanani-Sühring, K Ketelsen, BA Khan, C Knigge, H Knoop, P Krč, M Kurppa, H Maamari, A Matzarakis, M Mauder, M Pallasch, D Pavlik, J Pfafferott, J Resler, S Rissmann, E Russo, M Salim, M Schrempf, J Schwenkel, G Seckmeyer, S Schubert, M Sühring, R von Tils, L Vollmer, S Ward, B Witha, H Wurps, J Zeidler, and S Raasch. 2020. "Overview of the PALM model system 6.0." *Geoscientific Model Development* 13 (3): 1335–1372. https://doi.org/10.5194/gmd-13-1335-2020

McCoy, DT, P Field, H Gordon, GS Elsaesser, and DP Grosvenor. 2020. "Untangling causality in midlatitude aerosol–cloud adjustments." *Atmospheric Chemistry and Physics* 20 (7): 4085–4103, https://doi.org/10.5194/acp-20-4085-2020

Mechem, DB, CS Wittman, MA Miller, SE Yuter, and SP De Szoeke. 2018. "Joint Synoptic and Cloud Variability over the Northeast Atlantic near the Azores." *Journal of Applied Meteorology and Climatology* 57 (6): 1273–1290, https://doi.org/10.1175/Jamc-D-17-0211.1

Muñoz-Esparza, D, B Kosovic, J Mirocha, and J van Beeck. 2014. "Bridging the Transition from Mesoscale to Microscale Turbulence in Numerical Weather Prediction Models." *Boundary-Layer Meteorology* 153 (3): 409–440, https://doi.org/10.1007/s10546-014-9956-9

Neggers, RAJ, AP Siebesma, and T Heus. 2012. "Continuous Single-Column Model Evaluation at a Permanent Meteorological Supersite." *Bulletin of the American Meteorological Society* 93 (9): 1389–1400, https://doi.org/10.1175/bams-d-11-00162.1

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 (20): 10,998–11,006. https://doi.org/10.1002/2016GL070776.

Oue, M, A Tatarevic, P Kollias, D Wang, K Yu, and AM Vogelmann. 2020. "The Cloud-resolving model Radar SIMulator (CR-SIM) Version 3.3: description and applications of a virtual observatory." *Geosci. Model Dev.* 13 (4): 1975–1998, https://doi.org/10.5194/gmd-13-1975-2020

Pressel, KG, and K Sakaguchi. 2021. *Developing and testing capabilities for simulating cases with heterogeneous land/water surfaces in a novel atmospheric large eddy simulation code.* Pacific Northwest National Laboratory, https://doi.org/10.2172/1869291, https://www.osti.gov/servlets/purl/1869291

Rai, RK, LK Berg, B Kosovic, SE Haupt, JD Mirocha, BL Ennis, and C Draxl. 2019. "Evaluation of the Impact of Horizontal Grid Spacing in Terra Incognita on Coupled Mesoscale-Microscale Simulations Using the WRF Framework." *Monthly Weather Review* 147 (3): 1007–1027, https://doi.org/10.1175/Mwr-D-18-0282.1

Sauer, JA, and D Muñoz-Esparza. 2020. "The FastEddy® Resident-GPU Accelerated Large-Eddy Simulation Framework: Model Formulation, Dynamical-Core Validation and Performance Benchmarks." *Journal of Advances in Modeling Earth Systems* 12 (11): e2020MS002100, https://doi.org/10.1029/2020MS002100

Siebesma, AP, CS Bretherton, A Brown, . Chlond, J Cuxart, PG Duynkerke, HL Jiang, M Khairoutdinov, D Lewellen, CH Moeng, E Sanchez, B Stevens, and DE Stevens. 2003. "A large eddy simulation intercomparison study of shallow cumulus convection." *Journal of the Atmospheric Sciences* 60 (10): 1201–1219, https://doi.org/10.1175/1520-0469(2003)60<1201:alesis>2.0.co;2

Silber, I, RC Jackson, AM Fridlind, AS Ackerman, S Collis, J Verlinde, and J Ding. 2022. "The Earth Model Column Collaboratory (EMC2) v1.1: an open-source ground-based lidar and radar instrument simulator and subcolumn generator for large-scale models." *Geoscientific Model Development* 15 (2): 901–927, https://doi.org/10.5194/gmd-15-901-2022

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, https://doi.org/10.5065/D68S4MVH. http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf

Stevens, B, CH Moeng, AS Ackerman, CS Bretherton, A Chlond, S De Roode, J Edwards, JC Golaz, HL Jiang, M Khairoutdinov, MP Kirkpatrick, DC Lewellen, A Lock, F Muller, DE Stevens, E Whelan, and P Zhu. 2005. "Evaluation of large-Eddy simulations via observations of nocturnal marine stratocumulus." *Monthly Weather Review* 133 (6): 1443–1462, https://doi.org/10.1175/mwr2930.1

Turner, DD, and RG Ellingson, eds. 2016. *The Atmospheric Radiation Measurement (ARM) Program: The First 20 Years*. Vol. 57, *Meteorological Monographs*. American Meteorological Society, https://journals.ametsoc.org/toc/amsm/57

Unidata. 2023. Network Common Data Form (NetCDF). Accessed December 8, 2023, https://doi.org/10.5065/D6H70CW6. https://www.unidata.ucar.edu/software/netcdf/

U.S Department of Energy Climate and Environmental Sciences Division. 2014a. Atmospheric Radiation Measurement Climate Research Facility - Atmospheric System Research High-Resolution Modeling Workshop. U.S. Department of Energy. DOE/SC-0169, https://doi.org/10.2172/1471544. https://science.osti.gov/~/media/ber/pdf/workshop%20reports/doe-sc-0169-low-resolution.pdf

U.S. Department of Energy. 2014b. Atmospheric Testbed Workshop. U.S. Department of Energy. DOE/SC-0163, http://science.energy.gov/~/media/ber/pdf/Brochures/CESD_Atmospheric_Testbed_Workshop_V9_LOW.pdf?id=97

van Heerwaarden, CC, BJH van Stratum, T Heus, JA Gibbs, E Fedorovich, and JP Mellado. 2017. "MicroHH 1.0: a computational fluid dynamics code for direct numerical simulation and large-eddy simulation of atmospheric boundary layer flows." *Geoscientific Model Development* 10 (8): 3145–3165, https://doi.org/10.5194/gmd-10-3145-2017

Varble, A, H Morrison, and E Zipser. 2019. "Effects of under-resolved convective dynamics on the evolution of a squall line." *Monthly Weather Review* 148 (1): 289–311, https://doi.org/10.1175/mwr-d-19-0187.1

Varble, A, S Nesbitt, P Salio, E Avila, P Borque, P DeMott, G McFarquhar, S van den Heever, E Zipser, D Gochis, JR Houze, M Jensen, P Kollias, S Kreidenweis, R Leung, K Rasmussen, D Romps, and C Williams. 2019. *Cloud, Aerosol, and Complex Terrain Interactions (CACTI) Field Campaign Report.* Atmospheric Radiation Measurement (ARM) Research Facility. DOE/SC-ARM-19-028, https://doi.org/10.2172/1574024. https://www.arm.gov/publications/programdocs/doe-sc-arm-19-028.pdf

Varble, A, S Nesbitt, P Salio, E Zipser, S van den Heever, G MdFarquhar, P Kollias, S Kreidenweis, P DeMott, M Jensen, R Houze Jr, K Rasmussen, R Leung, D Romps, D Gochis, E Avila, CR Williams, and P Borque. 2018. *Cloud, Aerosol, and Complex Terrain Interactions (CACTI) Science Plan.* Atmospheric Radiation Measurement (ARM) Research Facility. DOE/SC-ARM-17-004, https://www.arm.gov/publications/programdocs/doe-sc-arm-17-004.pdf

Varble, AC, SW Nesbitt, P Salio, JC Hardin, N Bharadwaj, P Borque, PJ DeMott, Z Feng, TCJ Hill, JN Marquis, A Matthews, F Mei, R Öktem, V Castro, L Goldberger, A Hunzinger, K. R. Barry, SM Kreidenweis, GM McFarquhar, LA McMurdie, M Pekour, H Powers, DM Romps, C Saulo, B Schmid, JM Tomlinson, SC van den Heever, A Zelenyuk, Z Zhang, and EJ Zipser. 2021. "Utilizing a Storm-Generating Hotspot to Study Convective Cloud Transitions: The CACTI Experiment." *Bulletin of the American Meteorological Society* 102 (8): E1597-E1620, https://doi.org/10.1175/bams-d-20-0030.1

Wyngaard, JC. 2004. "Toward numerical modeling in the "terra incognita"." *Journal of the Atmospheric Sciences* 61 (14): 1816–1826. https://doi.org/10.1175/1520-0469(2004)061<1816:Tnmitt>2.0.Co;2

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 – Atmospheres* 109 (D1): D01104, https://doi.org/10.1029/2003jd004045

Zängl, G, D Reinert, P Rípodas, and M Baldauf. 2015. "The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core." *Quarterly Journal of the Royal Meteorological Society* 141 (687): 563–579, https://doi.org/https://doi.org/10.1002/qj.2378

Zhou, BW, JS Simon, and FK Chow. 2014. "The Convective Boundary Layer in the Terra Incognita." *Journal of the Atmospheric Sciences* 71 (7): 2545–2563, https://doi.org/10.1175/jas-d-13-0356.1

Appendix A

Agenda

DOE ARM Future of LASSO Workshop

November 2–3, 2023
EOL Atrium, Foothills Laboratory
National Center for Atmospheric Research, Boulder, Colorado
Time zone is Mountain Daylight Time (UTC−6)

Thursday, November 2, 2023

9:00 a.m. Welcome and meeting overview

9:10 a.m. Introductions of participants

9:30 a.m. LASSO overview

9:50 a.m. Current Scenarios: Shallow convection at SGP
- Scenario overview
- User highlight by Graham Feingold
- Discussion

10:40 a.m. Morning break

10:55 a.m. Current Scenarios: Deep convection during CACTI
- Scenario overview
- User highlight by Zhe Feng
- Discussion

11:30 a.m. Evaluation of the state of LASSO
- Lessons Learned
- Discussion: What is liked? What could be improved?

12:15 p.m. Lunch

1:45 p.m. How can LASSO increase scientific impact and better address its primary goals?
- Intro comments: Research target groups
- Discussion: Target group enablement
- Intro comments: Institutional modeling outreach
- User highlight by Yunyan Zhang, PI of THREAD
- Discussion: How can LASSO effectively integrate with the larger modeling community?

3:00 p.m. Afternoon break

3:15 p.m. How might LASSO operate better going forward?
- Understanding scenario variety and throughput: The tale of 2 workflows
- Discussion: Optimization of LASSO elements for utility
- Approaches to modes of operation
- Discussion: Approaches to modes of operation

4:30 p.m. Open discussion for topics of interest

5:00 p.m. Adjourn for dinner

6:00 p.m. Dinner at Boulder Social, 1600 38th St., Boulder

Friday, November 3, 2023

9:00 a.m. Summary of what we heard yesterday & feedback

9:25 a.m. Prioritization for the next 5 years
- Optimal science foci for LASSO & User group curation
- What are the considerations for prioritizing ARM locations?
    - ENA, Arctic, BNF, EPCAPE, TRACER, COURAGE, CAPE-K, other?
    - What additional observations and/or retrievals should be developed to support LASSO?

10:35 a.m. Break

10:50 a.m. Prioritizing, cont.
- Discussion: Balancing priorities–resource constraints versus science needs
    - Talk about currently available resources and how we fit within them
    - What are acceptable compromises?
- Discussion: Summary of feedback and participants' input on needed discussions
    - Missed topics?
    - New or continued discussion topics?

12:00 p.m. Adjourn

Appendix B

Workshop Participants

Name	Institution	Specialty
Meeting Conveners		
William I. Gustafson Jr.	PNNL	LASSO PI
Andrew M. Vogelmann	BNL	LASSO Co-PI
James H. Mather	PNNL	ARM Technical Director
In-Person Invited Experts		
Christine Chiu	Colorado State University	Remote sensing of clouds and the atmosphere
Graham Feingold	NOAA Chemical Sciences Laboratory (CSL)	Aerosol and cloud processes
Andrew Gettelman	PNNL	Global-scale modeling for clouds and climate
Scott Giangrande	BNL	Radar observations and cloud processes
Timothy Juliano	NSF NCAR	Mesoscale/microscale modeling
Chongai Kuang	BNL	Aerosol processes, BNF PI
Hsi-Yen Ma	LLNL	Large-scale atmospheric model evaluation
Hugh Morrison	NSF NCAR	Modeling of cloud microphysics and convection
Shawn Serbin	NASA Goddard Space Flight Center (GSFC)	Land-atmosphere interactions, land surface modeling, and remote sensing
Hyeyum (Hailey) Shin	NSF NCAR	Boundary layer and cloud parameterization
David Turner	NOAA Global Systems Laboratory (GSL)	Remote retrievals of boundary layer and cloud state
Yunyan Zhang	LLNL	Shallow clouds and large-scale model evaluation, THREAD PI

Virtual Attendees

Yaosheng Chen	NOAA CSL
Jennifer Comstock	PNNL
Aryeh Drager	BNL
Satoshi Endo	BNL
Zhe Feng	PNNL
Virendra Ghate	ANL
Jake Gristey	NOAA CSL
Thijs Heus	Cleveland State University
Jay Hineman	Geometric Data Analytics, Inc.
Yongjie Huang	University of Oklahoma
Oye Ideki	University of Missouri
Michael Jensen	BNL
Zhongjing Jiang	BNL
Weiwei Li	NSF NCAR
Toshihisa Matsui	NASA GSFC
Isabel McCoy	NOAA CSL
Sally McFarlane	DOE ARM
David Mechem	University of Kansas
Shaima Nasiri	DOE ASR
Nathan Philippot	Meteo France
James Polly	Geometric Data Analytics, Inc.
Derek Posselt	NASA Jet Propulsion Laboratory
Prasanth Prabhakaran	NOAA CSL
John Rausch	BNL
Lynn Russell	University of California, San Diego
Tamanna Subba	BNL
Cheng Tao	LLNL
Nathan Urban	BNL
Die Wang	BNL
Mikael Witte	Naval Postgraduate School
Robert Wood	University of Washington
Heng Xiao	PNNL
Takanobu Yamaguchi	NOAA CSL
Damao Zhang	PNNL
Man Zhang	NOAA GSL
Xiaoli Zhou	NOAA CSL

Appendix C

Submitted White Papers

A call for white papers was issued prior to this workshop to gather information from the community about their experience with existing LASSO data sets, priorities, and vision for LASSO. Specifically, information was requested for the following topics of interest:

1. ***The LASSO Concept**:* Briefly, how have you used LASSO in your research? What have been the most valuable aspects of the LASSO data sets? What could be improved? How can the LASSO approach be modified for greater scientific impact?
2. ***Science Drivers**:* What science drivers and ARM observations should LASSO focus on for future scenarios? We are specifically interested in understanding what the added benefit would be for a LASSO-style implementation. For example, recent and upcoming ARM deployments that could be addressed by LASSO include:

- Multidisciplinary Drifting Observatory for the Study of Arctic Climate (MOSAiC), Arctic Ocean, October 2019–October 2020
- Cold-Air Outbreaks in the Marine Boundary Layer Experiment (COMBLE), North Atlantic, December 2019–May 2020
- Surface Atmosphere Integrated Field Laboratory (SAIL), Crested Butte, Colorado, September 2021–June 2023
- TRacking Aerosol Convection interactions ExpeRiment (TRACER), Houston, Texas, October 2021–September 2022
- Eastern Pacific Cloud Aerosol Precipitation Experiment (EPCAPE), San Diego, California, February 2023–February 2024
- Bankhead National Forest (BNF), northern Alabama, late 2023–2028
- Cloud And Precipitation Experiment at Kennaook (CAPE-K), Tasmania, Australia, April 2024–September 2025
- Coast-Urban-Rural Atmospheric Gradient Experiment (CoURAGE), Baltimore, Maryland, likely fall 2024 to fall 2025.

A LASSO scenario in early development will focus on marine clouds over ARM's Eastern North Atlantic atmospheric observatory.

Designs around alternate meteorological conditions at the SGP, CACTI, and the ENA are also options, as well as additional ideas beyond existing and upcoming data sets. For example, are there instruments that could be deployed that would be particularly helpful to pair with LASSO modeling?

Authors should be aware of the timeline for requesting new campaigns and the associated lag before LASSO would be able to work with newly acquired data.

3. ***Scenario Implementation****:* LASSO must balance benefits versus resource constraints when choosing the model resolutions and observations used and would like your feedback on the tradeoffs involved. Current efforts employ ensembles primarily to examine large-scale forcing sensitivity—is this the right approach? What modeling approaches should be used within LASSO? Example details to consider include model choice, domains used, ensemble approach, and complexity of physics parameterizations. Should we focus more on mesoscale simulations than LES?
4. ***Data Content and Delivery****:* What are preferred methods for formatting and sharing LASSO data sets? For example, are subsets of variables useful? Are there better ways to package LASSO data to ease access and simplify analysis for users?

   Jupyter notebook servers and high-performance computing have recently been made available to ease user access to the large data sets and minimize the need to download data to non-ARM computers. How do you foresee these resources meeting your needs? Would other methods be useful? (Please be mindful that computational resources are expandable but finite.)
5. ***Use of ARM Observations****:* How should observational capabilities be blended with LASSO modeling, and are there other ways observation products could be used within or specifically developed for use by LASSO? An example of past integration includes the LASSO–Clouds Optically Gridded by Stereo (COGS) stereo camera cloud mask that improves upon the original use of Active Remote Sensing of CLouds (ARSCL) cloud masks by avoiding contamination from insect clutter. Options for future scenarios might include data from the tethered balloon system and ARM's uncrewed aerial systems should they address a pressing need. Please identify the need and indicate how a blended product would advance associated research.
6. ***LASSO Model Engagement****:* How can LASSO better integrate with other modeling activities? An important driver for LASSO is using LES to bridge the scale gap from localized observations to the coarser grids of global models. Are there approaches that will facilitate this interaction and cross-community engagement? There is particular interest in finding ways to use LASSO to enable the application of ARM observations to DOE's Energy Exascale Earth System Model (E3SM), but ideas for linking with other modeling systems are also welcome.

Community members submitted two white papers in answer to the call, included in the next appendix.

C.1 — White Paper #1: ACI with EPCAPE Data

Title: Advancing ACI by Perturbing Simulations Across Regimes

Authors: Lynn Russell, Israel Silber, Markus Peters, Jim Smith, Mark Miller, Suzanne Paulson, Dan Lubin

1. *The LASSO Concept*

Constraining Aerosol-Cloud Interactions (ACI) with observations is limited by the conditions that occur in a given year at a given site, but since many variables co-vary a functional dependence signal is more likely to be faint or obscured. This proposal uses the LASSO concept with the consistent, multi-day cloudiness observations of EPCAPE to first constrain case-specific nested simulations and then to perturb the liquid water content (by modifying the temperature profile) in the simulation to shift from aerosol-limited to updraft-limited regimes (and vice versa). This approach provides an evaluation of microphysical processes that allows us to put the observations in the context of process-based regimes. Specifically,

- *Microphysical Processes:* Calculating the effects of aerosols on clouds requires knowing the processes that formed the clouds. However, while cloud measurements characterize cloud states, there are no direct measurements of the processes that induce their initial formation. For example, aerosol activation relies on the time-dependent history of supersaturation in individual parcels. Simulated supersaturations are needed for interpreting the observations to assess the effects of different aerosol compositions and concentrations, since this key quantity cannot be directly observed but provides the context for aerosol activation in cloud. LASSO can provide a plausible set of unobserved microphysical process quantities that are consistent with the observed cloud macrophysics, thereby providing the added value of realistic rather than arbitrary (and often inconsistent) conditions.
- *Process-based Regimes:* Modeling studies clearly illustrate that aerosol effects on clouds have different regimes, but identifying such regimes in observations requires robust data sets. By using ensembles of runs, LASSO can perturb the observed conditions to explore the boundaries of these regimes and provide the ability to cross from one regime to the next. Such counter-factual perturbations would facilitate the mapping of regime boundaries and uncertainties from a limited set of cases. Thus we propose that LASSO expands on the cases provided by nature during EPCAPE to create physically consistent ensembles across regimes of ACI.

We envision these LASSO results as being of interest to two user groups, observationalists and modelers. We expect that the targeting of microphysical processes and regimes will expand the existing user groups to include more aerosol-focused researchers.

2. *Science Drivers*

- Diagnosing aerosol effects on clouds requires knowledge of both cloud microphysical and macrophysical processes and properties. Since some of these process quantities (especially at the microphysical level) cannot be observed, realistic model simulations are needed to provide them.

- Half of the world's population is concentrated along coastlines, making coastal areas major pollution sources. These sources are important for causing ACI, which, in turn, affect the coastal populations by changing cloud properties. Global climate models do a poor job of representing cloudiness along coastlines, including the western coast of the U.S.
- Multiple ground-based measurement campaigns that are valuable for marine ACI assessment are located in coastal regions or islands (EPCAPE, CAPE-K, COMBLE, ENA), but models could be used to show their relevance to over-ocean conditions. Land-based measurements are often selected to study marine clouds because of their logistical support, even though they are known to be affected by non-marine orographic effects on updraft velocities, aerosol sources, and radiative fluxes. However, modeling support can bridge the gap by showing the extent to which coastal measurements are representative of marine conditions.

3. *Scenario Implementation*

We propose a "hybrid" nested/ensemble approach. Similar to past studies, we propose initializing simulations to specific cases with a comprehensive nested model, but then expanding this simulation by an ensemble of runs that cover a larger subset of EPCAPE and that support generalization. The ensemble of supersaturations and associated liquid water and updraft velocities provide a much more robust foundation for both interpreting observations and generalizing to global simulations. By using perturbations around the more costly individual nested simulations, the ensemble can be achieved at relatively low cost. This approach provides a way to expand the conditions provided by a single set of observations to more global conditions.

4. *Data Content and Delivery*

We suggest accommodating studies of comprehensive simulations by providing the "complete" model simulations in the format in which they are performed, for which Jupyter notebooks would be lovely. In addition, to accommodate other users studying cloud process variables for interpreting aerosol observations, we suggest providing a few key variables for each ensemble (e.g., supersaturation, liquid water content, updraft velocity, entrainment rate, inversion strength) in a standard but abbreviated format (NetCDF or similar). We recognize that two versions require more resources, but we suggest that it may increase data use to also provide the standard format, although perhaps a hybrid option would be to just provide the standard NetCDF output.

5. *Use of ARM Observations*

We recommend that the proposed approach will be applied to the EPCAPE data set because of the following characteristics:

- Persistent cloudiness at EPCAPE supports multi-day simulations, allowing spin-up time for simulations as well as investigation of diurnal cycles.
- The extent of Guest-PI observations of in-cloud size distributions and composition (Russell, Petters, Paulson, Smith, Liggio, Wheeler, Wentzell, Chang, Galewsky) exceeds other ARM campaigns, with over 300 hours and dozens of in-cloud events. Interpreting these measurements relies heavily on the understanding of cloud processes, making the LASSO results indispensable for the guest investigators and broad community members relying on these collected observations.

- The first 7 months of the 12-month EPCAPE measurements illustrate a variety of low-cloud conditions, with a substantial range of liquid water content, inversion strengths, and drizzle, with a nearly perfect record of instrument uptime to date.
- The consistency of northwesterly trajectories reduces the need for major differences in large-scale forcings, effectively allowing more focus on the range of microphysics by constraining the macrophysics.
- The typical light precipitation conditions renders EPCAPE an important contrast to ENA and deep convection, featuring instead the drizzling conditions captured by the laser disdrometer that are pervasive in the many marine stratocumulus decks that cover the oceans.
- The range of aerosol concentrations from similar upwind source mixtures provides a large and unique dynamic range of aerosol that likely have very similar composition, meaning that ACI processes are more likely to be statistically significant because the sources are so consistent.

6. *LASSO Model Engagement*

- Ensemble Approach: The LASSO use of ensembles of supersaturations and associated liquid water and updraft velocities would provide a robust foundation for generalization. By using perturbations around longer individual nested simulations, the ensemble can be achieved at relatively low cost. This approach provides a way to expand the conditions provided by a single set of observations to more globally relevant conditions.
- Orographic Effects: LASSO can characterize the effect of a coastal mountain region in representing low clouds and diagnosing which conditions require orography to represent ACI. The range of models and ensemble approach will make this result more robust than a single-PI effort and will augment the justification for using coastal and island sites for studying marine clouds.
- Benchmarks: By allowing direct comparisons to observations, LASSO can provide benchmarks against which to evaluate climate model representations of coastal cloudiness. This is a key step in the “scaling up” of our understanding of ACI from LES to regional/global modeling.
- Process Regimes: By investigating the range of aerosol conditions at EPCAPE for coastal trajectories with persistent cloudiness, LASSO will enable the identification of the factors that control aerosol-limited and updraft-limited regimes. This delineation of factors is key to understanding and simplifying aerosol interactions for use in global models.

C.2 — White Paper #2: AMF3-BNF LASSO Opportunity

Title: The Future of LASSO: An AMF3-BNF Site Science Team Response to the Request for White Papers

Authors: Thijs Heus, Girish N. Raghunathan, Allison L. Steiner, Chongai Kuang, Shawn P. Serbin, and Scott E. Giangrande

The upcoming Atmospheric Radiation Measurement (ARM) Mobile Facility (AMF) deployment to the southeast United States (SE U.S.) will provide unique opportunities for the study of important land-atmosphere interactions and processes. Extended ARM observations will be distributed across the diverse northern Alabama landscape to help resolve the role of local forcing in climate-relevant processes and differentiate those from climate processes driven by larger-scale circulations. A distinguishing factor for Third ARM Mobile Facility (AMF3) studies will also be its deployment within the Bankhead National Forest (BNF), promoting regionally important SE U.S. land-atmosphere two-way interactive studies *“from the canopy to the clouds.”* In partnering with the DOE ASR program on the AMF3-BNF deployment, ARM solicited a multi-agency Site Science Team (SST) to provide input and close interaction with ARM management towards a successful multi-year SE U.S campaign. One expectation for this Site Science Team is that they help promote to the ARM climate community various possibilities to bolster AMF3 campaign efforts, examples that include offering collective team knowledge of this AMF3-BNF site, its justification for its recommended instruments and operations, and as relevant to this white paper activity, guidance to complementary ARM activities such as LASSO towards their potential consideration of AMF3 scenarios. For this “Future of LASSO” solicitation, the SST attempts to address each white paper topic of interest as anchored by several key AMF3-BNF themes.

1. LASSO Concept

Previously, members of this SST have benefitted from LASSO bundles, but mostly by drawing from LASSO forcing information to perform complementary simulations for selected events. For our suggestions on future LASSO scenarios, we sought examples in “Scenario Implementation” wherein our SST may more directly benefit from LASSO model outputs that overlap topics also of interest to the wider climate community for expanding land-atmosphere interaction (LAI) studies at the AMF3-BNF.

In our opinion, the LASSO concept has been successful in providing additional value to the ARM observations by a) Providing 3D, complete, and internally consistent data sets, and b) providing a case selection and quality-assessed forcing files that may help individual researchers to drive their own simulations.

In this white paper, we propose to take the next logical step, and have LASSO simulations run on a near-daily basis, and as close to the observational date as possible. This way, the peak interest in the observations should coincide with the availability of the simulations.

2. AMF3 “Science Drivers” and LASSO

One recommendation for upcoming LASSO activities would be future scenarios that focus on key science drivers associated with the upcoming AMF3 deployment to the BNF. As representatives of AMF3-BNF’s Site Science Team, we have identified high-priority activities relevant to these LASSO discussions, including the potential for LASSO to perform *continuous, near-real-time simulations* to inform primary and guest observations and instruments (both what and measurement modes) collected at this site. Notably, many forms of continuous and observationally constrained simulations may be positioned to provide potential gap-filling capabilities for key quantities that cannot be readily observed by the AMF3. As one example, an important driver of the coupled AMF3-BNF science themes is an emphasis on turbulence, with LES capable of adding value to the observational data in those respects, or addressing the challenges of partitioning surface energy with flux tower observations alone. Overall, the AMF3-BNF deployment offers a wide range of interesting meteorological phenomena and science motivations across cloud, aerosol, and coupled aerosol-cloud land-atmosphere interaction (LAI) themes (as outlined by AMF3-BNF's recent Science Plan (Kuang et al. 2023) that would provide challenging yet achievable activities for LASSO. There are strong potential similarities and known contrasts between this AMF3-BNF and the previous SGP LASSO shallow cloud regimes, or CACTI LASSO deep convective cloud themes, which may ensure that new LES LASSO data sets potentially complement existing efforts and/or users. As a practical consideration, this campaign is a longer-term deployment that begins during the current FY and will continue for a minimum of five years. In this regard, LASSO activities may be performed and available *within* the campaign's lifespan, while scientific interest in the AMF3-BNF site and its observations are still fresh, and potential modifications to experiments/designs can be implemented in later years.

For this white paper, we emphasize the following higher-impact research topics from the AMF3-BNF campaign, where we believe LES has been (or will be) a well-proven tool such that LASSO may add significant value. All these examples would emphasize continuous or near-real-time delivery of LES outputs:

1. Land-atmosphere interactions, including interactions between the atmosphere, soil, and canopy.
2. The development of convection, including convective initiation over the forest.
3. The evolution of aerosols in the atmospheric boundary layer.

First among these examples, the inclusion of a more realistic representation of land surface properties and processes (than previous LASSO implementations) implies the surface coupling itself may be studied standalone, including the effects of small-scale heterogeneity, and the validity of Monin-Obuhkov and other surface parameterizations. Improved surface coupling implies that a longer time span of the atmospheric diurnal cycle may be simulated, without a mismatch between prescribed surface fluxes and LES response. Even if the nocturnal boundary may be challenging at the anticipated LES resolutions available, the morning and evening transitions should be within reach. With all input openly available, interested researchers may be able to rerun at improved resolution.

The onset of convection, and specifically shallow cumulus, was at the core of the first LASSO implementation over SGP. For the upcoming AMF3-BNF, we anticipate a higher frequency of cloud initiation and transition to deeper cloud modes over these sites. Importantly, a moister environment over BNF contrasts sharply in terms of Bowen ratio with the cloud conditions found over SGP. A long-term LES record of this area would enable testing for hypotheses such as the cloud/shade/surface flux feedback loop (Vila et al. 2014), or the sub-LCL cloud formation (Altaratz et al. 2021).

Our third topic is at the frontier of the field and would include aerosols in the LES simulations. This presents a tremendous opportunity (yet, challenge) for model-data integration, observing system simulations, and optimal experimental design. Under these themes, AMF3-BNF "scenarios" would provide an opportunity to engage with the terrestrial ecosystem science community in new and exciting ways via coupled land-atmosphere processes that connect terrestrial emissions, atmospheric chemistry, and the boundary layer energy budget. While the relatively idealized LES simulations that we propose are not best suited for highly complex chemistry, we can include heterogeneous emission sources from 3D modeling runs as tracers (e.g., prescribed aerosol number and size) that can be transported in the boundary layer and interact with the radiation and cloud microphysics schemes.

3. *Scenario Implementation*

As surprising to no one, LES has advanced since the initial LASSO-SGP simulations. Land-surface models are now more commonly available to simulate the interactions with the soil and the canopy, and important plant physiological processes that drive heat, moisture and CO_2 exchanges, characteristics that are absolutely crucial for AMF3-BNF. Additionally, GPU-based models such as MicroHH (van Heerwaarden et al. 2017) are now also available and can utilize the relatively underused GPUs on the ARM-HPC cluster in a highly efficient manner. Therefore, in any of the above implementations, there is now the possibility to generate reasonable data sets, comparable to the original LASSO-SGP bundles, sets in real time, and those with potentially little delay beyond the publication of initial and boundary conditions (this is around 5 days for ERA5).

Based on these potential updates, such outputs would be available for direct comparisons with observations for key complementary ARM themes such as improved ARM data quality control, in addition to the value-added context to some observational data, such as spatial variability, cloud size, and tendencies that are not easily observed.

For one proposed scenario, continuous simulations would be relatively straightforward in setup:

1. Periodic boundary conditions,
2. Interactive land, radiation, and microphysics (no aerosol),
3. ~25 km horizontal domain, 50 m grid spacing, all situated above the forest.

While for the above example we propose periodic boundary conditions, that does not necessarily imply that surface heterogeneity is ignored; fluctuations in surface properties (e.g., roughness length, LAI, etc.) can be represented in the model, even if such fluctuations are repeated on the scale of one domain length. On the current ARM-HPC hardware, which includes four Nvidia A100 GPUs, such simulations are possible in 50% of real time on a single GPU. A screening algorithm (e.g., Naud et al. 2023) could limit the simulations to days without dominating synoptic forcing. More complex simulations with open boundaries and/or complex chemistry may be performed at a later time during the project, for instance focusing on intensive operational periods (IOPs).

4. *Data Content and Delivery*

Data content should include a direct link to the version of the code that is used for generating the LES, including all preprocessing and post-processing scripts. This means that any simulation can be redone at a higher resolution, complexity, or for a different scenario. Similar to observational data sets, a quality control would evaluate all simulations, for instance in their proximity to the forcing data (ERA5) or the observations, using the similar tools as developed by the LASSO team before. To limit the amount of storage necessary for the continuous simulations, we believe it should be sufficient to store only the average profiles of standard statistics, including conditional samples over specific criteria (e.g., updrafts, cloud) and vertical and horizontal cross-sections of the data at high (60 s) temporal frequency. 3D output fields could be written on a less frequent basis, hourly or less. It may be useful to have the 3D output written out on a higher frequency, so that instrument simulators or other semi-offline calculations (e.g., direct entrainment, cloud size/tracking, for instance, using tobac (Heikenfeld et al. 2019)) can be performed, after which most of the data can be discarded).

An advantage of having the LASSO simulations performing in real time is that any IOP can request additional output, based on their specific goals.

5. *Use of ARM Observations*

The synergy of LES and observations will be particularly strong over BNF, where the detailed single-column measurements of the ARM tower observations can be complemented by the spatial variability that can be assessed by LES modeling over space and time. AMF3-BNF also has a unique set of surface measurements and configurations that provide novel vertical measurements of mass, heat, and momentum that could inform new LES model integration approaches designed to use these data sets. Combined with the use of surface remote sensing (e.g., lidar, hyperspectral, thermal, optical time series) from airborne and spaceborne platforms, it would be possible to provide critical boundary/initial conditions, drivers, and the canopy characterization needed for detailed for LES efforts designed to study processes and boundary layer development. Data simulators for AERI, PCCP/COGS, and/or cloud radar and similar ARM site capabilities anticipated for the AMF3-BNF deployment can be immediately performed over the LES output files for optimal comparison and validation; If high-temporal, 3D data sets are necessary for these simulators, the largest data sets may be deleted after generating the simulator output if required to save disk space.

References

Altaratz, O, I Koren, E Agassi, E Hirsch, Y Levi, and N Stav. 2021. "The environmental conditions behind the formation of small (subLCL) clouds." *Geophysical Research Letters* 48(23): e2021GL096242, https://doi.org/10.1029/2021GL096242

Naud, CM, JE Martin, P Ghosh, G Elsaesser, and D Posselt. 2023. "Automated identification of occluded sectors in midlatitude cyclones: Method and some climatological applications." *Quarterly Journal of the Royal Meteorological Society* 149(754): 1990–2010, https://doi.org/10.1002/qj.4491

Heikenfeld, M, PJ Marinescu, M Christensen, D Watson-Parris, F Senf, SC van den Heever, and P Stier. 2019. "tobac 1.2: towards a flexible framework for tracking and analysis of clouds in diverse datasets." *Geoscientific Model Development* 12(11): 4551–4570, https://doi.org/10.5194/gmd-12-4551-2019

Van Heerwaarden, CC, BJH van Stratum, T Heus, JA Gibbs, E Fedorovich, and JP Mellado. 2017. "MicroHH 1.0: a computational fluid dynamics code for direct numerical simulation and large-eddy simulation of atmospheric boundary layer flows." *Geoscientific Model Development* 10(8): 3145–3165, https://doi.org/10.5194/gmd-10-3145-2017

Kuang, C, S Giangrande, S Serbin, G Elsaesser, P Gentine, T Heus, M Oue, J Peters, J Smith, A Steiner, A McComiskey, M Jensen, A Sedlacek, P Kollias, A Vogelmann, H Morrison, M Petters, and D Turner. 2023. "AMF-3 Bankhead National Forest Science Plan." Atmospheric Radiation Measurement (ARM) Research Facility. DOE/SC-ARM-23-035, https://www.arm.gov/publications/programdocs/doe-sc-arm-23-035.pdf

Vilà-Guerau de Arellano, J, HG Ouwersloot, D Baldocchi, and CMJ Jacobs. 2014. "Shallow cumulus rooted in photosynthesis." *Geophysical Research Letters* 41(5): 1796–1802, https://doi.org/10.1002/2014GL059279

ARM

www.arm.gov

U.S. DEPARTMENT OF
ENERGY

Office of Science