DOE/SC-ARM-14-029

ATMOSPHERIC RADIATION MEASUREMENT
CLIMATE RESEARCH FACILITY

Decadal Vision

October 2014

ARM CLIMATE RESEARCH FACILITY

U.S. DEPARTMENT OF ENERGY | Office of Science

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

DISCLAIMER

DOE/SC-ARM-14-029

Atmospheric Radiation Measurement Climate Research Facility Decadal Vision

October 2014

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

Acronyms and Abbreviations

AAF	ARM Aerial Facility
ADI	ARM Data Integrator
AMF	ARM Mobile Facilities
API	Application Programming Interface
ARM	Atmospheric Radiation Measurement Climate Research Facility
ARRA	American Recovery and Reinvestment Act
ASR	Atmospheric System Research
BER	Biological and Environmental Research
CESD	Climate and Environmental Sciences Division
CF	Climate and Forecast
CRM	cloud-resolving model
DOE	U.S. Department of Energy
DOI	Digital Object Identifiers
DQO	Data Quality Office
ENA	Eastern North Atlantic
ESGF	Earth System Grid Federation
FGDC	Federal Geospatial Data Committee
GCM	Global Circulation Model
IASOA	International Arctic Systems for Observing the Atmosphere
IDL	interactive data language
ISO	International Organization for Standardization
LES	large-eddy simulation
NGEE	Next-Generation Ecosystem Experiment
NSA	North Slope of Alaska
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OGC	Open Geospatial Consortium
OME	Online Metadata Editor
ORNL	Oak Ridge National Laboratory
OSSE	Observing System Simulation Experiments
PCMDI	Program for Climate Model Diagnosis and Intercomparison
PI	principal investigator
SGP	Southern Great Plains
THREDDS	Thematic Real-time Environmental Distributed Data Services
UAS	unmanned aerial systems

UEC	User Executive Committee
VAP	value-added products
VLI	Virtual Laboratory Initiative

Contents

Figures

Tables

1.0 Introduction

The Atmospheric Radiation Measurement (ARM) Climate Research Facility was established in 1989 within the Department of Energy (DOE) Office of Biological and Environmental Research (BER) to provide an observational basis for studying the Earth's climate. This work is becoming increasingly important as the evidence for climate change becomes clearer (IPCC 2014). Climate change has been identified as a challenge to U.S. energy security (US DOE 2014). ARM is managed within the BER Climate and Environmental Science Division (CESD), whose mission (US DOE 2012a) is:

> *To advance a robust predictive understanding of Earth's climate and environmental systems and to inform the development of sustainable solutions to the Nation's energy and environmental challenges.*

The ARM Facility supports this mission by providing comprehensive sets of observations in diverse climatic regimes and providing associated data infrastructure to support the research community.

Designated as a DOE Office of Science user facility in 2003, the ARM Facility includes a network of long-term, fixed-location observation sites and mobile observation facilities that are typically deployed for about a year at a time. Both the fixed and mobile sites are equipped with an extensive array of instruments. The ARM Facility also includes an aerial component to augment these ground-based measurements and infrastructure to collect, process, and deliver data to the research community (U.S. DOE 1990; Stokes and Schwartz 1994; Ackerman and Stokes 2003; Mather and Voyles 2013). The ARM mission is:

> *(To provide) the climate research community with strategically located in situ and remote-sensing observatories designed to improve the understanding and representation, in climate and Earth system models, of clouds and aerosols as well as their interactions and coupling with the Earth's surface.*

The ARM Facility is currently undergoing a reconfiguration that is designed to accelerate the application of ARM observations and data processing for the understanding of key atmospheric processes and the representation of these processes in global climate models. This enhanced impact on the research community will be achieved by:

1. Enhancing ARM observations and measurement strategies to enable the routine operation of high-resolution models and to optimize the use of ARM data for the evaluation of these models.
2. Producing routine high-resolution model simulations over domains coincident with ARM sites.
3. Developing data products and analysis tools that enable the evaluation of models using ARM data.

The combination of high-resolution model simulations with ARM observations will provide the most detailed possible description of the region around the ARM site. This reconfiguration of the ARM Facility does not alter the ARM mission; however, it does involve a number of changes to align all aspects of the Facility with the new strategy. Figure 1 illustrates this integrated vision through which ARM observations will be used to advance climate models. This vision also includes developing and improving interactions between the ARM Facility and the research community to make full use of this next-generation strategy.

2.0 Strategic Goals for the Next-Generation ARM Facility

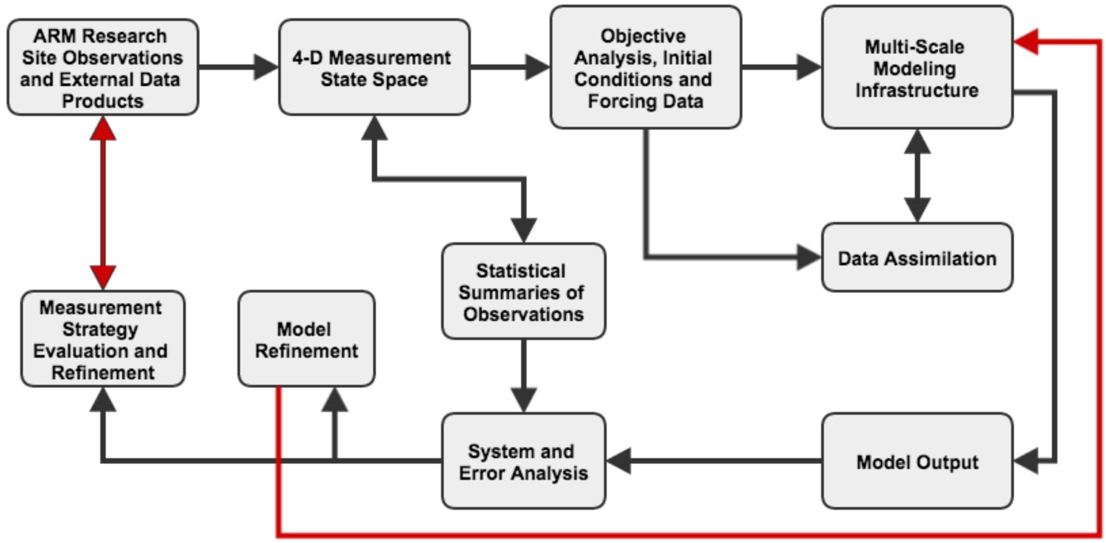

Figure 1. An integrated strategy to apply ARM data to model improvement.

The vision for the ARM Facility for the next 5-10 years centers on developing an integrated observation-modeling-analysis framework (Figure 1). The integrated observation/model data set will provide the basis for determining detailed aerosol and cloud properties and for exploring detailed atmospheric processes that control the life cycle and radiative impacts of aerosols and clouds, including the role of aerosols in cloud formation. This vision includes strengthening existing areas of expertise and working to accelerate the application of ARM data to modeling activities. To achieve this vision, the ARM Facility has established five high-level goals:

- Establish observation-modeling megasites at the Southern Great Plains (SGP) and North Slope of Alaska (NSA)
- Continued focus on ARM measurement excellence to support DOE climate science research
- Producing routine high-resolution model simulations over domains coincident with ARM sites
- Enhance data products and processes
- Strengthen interactions with the user community.

ARM provides comprehensive measurements at high spatial and temporal resolution and, therefore, is ideally suited to study atmospheric processes in key regions, which are generally exhibited through many atmospheric parameters. However, there is a scale mismatch between high-resolution ARM observations and global-scale climate models (horizontal resolution on the order of 100 km). An effective mechanism to bridge this scale gap is to run a high-resolution (horizontal resolution on the order of 1 km or less) model over an ARM site, use ARM data to evaluate parameters and processes in the high-resolution model, and then use the high-resolution model in conjunction with ARM data to develop representations of physical processes in the climate model. Historically, this type of analysis has only been done for short-term intensive field campaigns due, in large part, to the lack of data to initialize and constrain the model on an ongoing basis. This episodic approach does not take advantage of the continuous nature of ARM observations.

The clear choices for implementing an integrated observation-modeling framework are SGP and NSA. The SGP site has always served as the proving ground for new ARM activities because of the excellent infrastructure support available there as well as the wide range of meteorological conditions found in that region. Meanwhile, NSA is a region experiencing dramatic climate change and complex physical phenomena, such as the formation and maintenance of mixed-phase clouds and the interaction between clouds and sea ice require significant study to understand the impacts of future climate change in that region.

Coupled with the strategy to implement this observation-modeling framework at the SGP and NSA sites are plans to optimize the use of the other components of the ARM observation network and development of data products and processes to accelerate the application of ARM observations to model evaluation and development. This development also requires careful attention to setting priorities that are consistent with the needs of the ARM user community to ensure that the new observations, data sets, and modeling framework will be of broad use to the community, so there will be an increased focus on soliciting input from the user community.

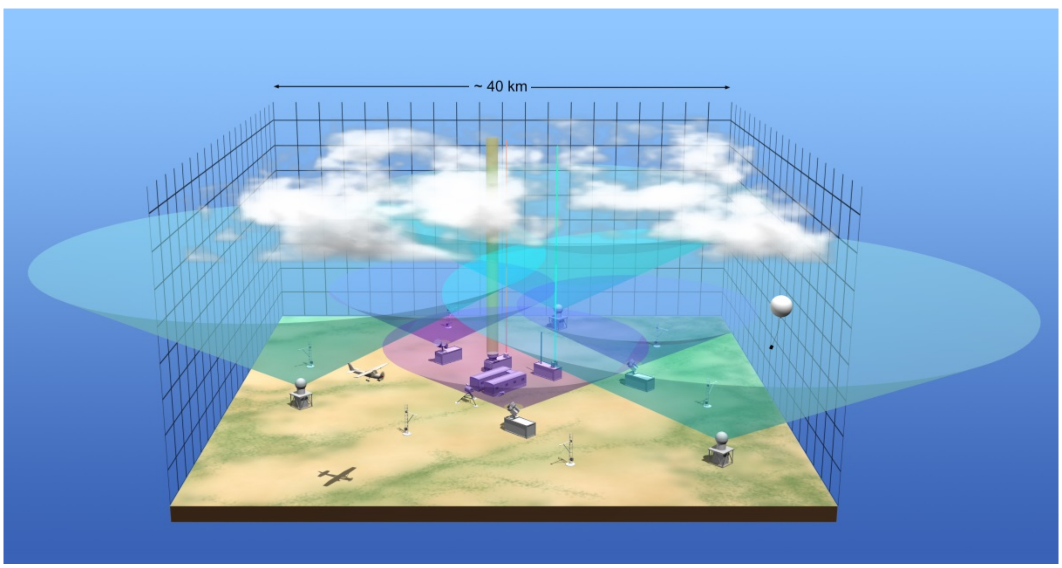

Figure 2. Distributed megasite incorporating a network of instruments to support model development and evaluation.

2.1 Establish Observation-Modeling Megasites

2.1.1 Southern Great Plains

Figure 3. The ARM SGP site exhibits frequent convection during the spring and summer.

The SGP site is the first and most extensive of the ARM sites. It includes the SGP Central Facility, which is comparable to other ARM sites, as well as a network of extended and intermediate facilities that provide spatial information over a domain that is intended to represent a global circulation model grid box. Looking forward, the goal for SGP is to build on the significant measurement capabilities available there to develop a unique integration of high-density measurements with routine high-resolution models. Work over the next few years will be geared toward improving and better integrating the SGP instrument network to address a range of science themes and develop the observation-model testbed methodology.

To obtain input from the user community on strategies for optimizing the application of observations and high-resolution models at the SGP on DOE climate science goals, a workshop was held in May of 2014. The workshop was designed to address three questions:

- What science questions can be addressed using an augmented SGP site
- How can ARM measurements be best applied to address these science issues
- What modeling strategies can be applied to support addressing these questions

Participants in the workshop, led by Graham Feingold and James Mather, identified five main science themes (U.S. DOE 2014a) particularly relevant to the SGP locale. These science themes provide guidance for what measurement enhancements need to be made at the SGP and are being used to organize discussion about future modeling strategies. The five core science themes identified for the SGP are:

- shallow convection
- deep convection

- aerosol
- radiation
- land surface and the carbon cycle.

Discussion was organized around the topics of shallow and deep convection at the joint Atmospheric System Research (ASR) program and ARM Facility SGP high-resolution modeling workshop, which then encompassed the other science themes (U.S. DOE 2014a). Looking forward, these topics provide effective means for organizing observation and modeling activities at the SGP site. The two topics represent different spatial scales and levels of complexity, suggesting short- and long-term goals for an integrated measurement/modeling strategy.

In the short term, the focus of the integrated observation-modeling system will be on shallow convection, which is often dominated by local forcing and can be modeled on a relatively small-scale horizontal domain on the order of 30-50 km. To provide a more continuous forcing data set for the shallow convection case, the facility plans to establish a network of thermodynamic and dynamic profiling instruments spanning a domain appropriate for large-eddy simulation (LES)/cloud-resolving model (CRM) modeling. These instruments will characterize the lateral boundaries of the model domain. There will also be an emphasis on characterizing the land-atmosphere interface. The land surface is heterogeneous in the SGP domain and this heterogeneity may be important for the distribution of shallow clouds over the region. The ARM Facility will work to improve measurements of the land-atmosphere interface, beginning with an evaluation of soil moisture and developing a representation of the heterogeneity across the SGP domain using a combination of satellite imagery and in situ surveys.

Deep convection involves inherently larger scales than shallow convection and, at the SGP, deep convection is often influenced by upstream phenomena. To satisfy the demands of both the shallow and deep convection cases, the atmospheric profiling systems will be developed to be portable, so the network of forcing sites can be expanded or contracted as the need demands.

2.1.2 North Slope of Alaska

ARM currently maintains two facilities on the NSA, the original Barrow site, which has operated since 1998, and the Oliktok deployment of the third ARM Mobile Facility, which began operation in 2013. The NSA is an important region of study both because of the significant climate changes currently underway there and the range of complex science issues that challenge understanding of these changes. As with the SGP megasite, the goal for the NSA is to develop an integrated observation-modeling testbed, but because of the complexity of the NSA region, the initial focus in the NSA will be on enhancing measurement capabilities to accelerate scientific understanding of critical Arctic issues.

A second reconfiguration workshop, led by Johannes Verlinde, Mark Ivey, and Robert Ellingson, was held in 2014 to explore key science issues associated with the NSA sites and strategies for addressing those issues (U.S. DOE 2014b). The science issues were segregated into four categories:

- clouds
- aerosols

- vertical structure
- large-scale phenomena.

Specific topics included exploring what processes drive the distribution and properties of aerosols and clouds, including the roles of the heterogeneous lower boundary layer and large-scale forcing. To support the study of these themes, ARM will adopt a three-pronged approach with enhancements to measurements at the Barrow and Oliktok locations; development of additional ground sites, potentially including the former Atqasuk ARM site; and the application of manned and unmanned aerial systems (UAS). Enhancements at the Barrow site and the Oliktok deployment will support the characterization of the atmosphere structure and composition over those locations, while the addition of auxiliary sites combined with aerial measurements will provide valuable information about spatial heterogeneity.

There is a particular need to characterize the heterogeneity of the land-atmosphere interface in the Arctic. This is a major issue at the NSA, where the land surface is a mosaic of tundra and melt ponds during the summer months, and the nearby ocean often has a complex distribution of open water and sea ice. Efforts to characterize the land surface will be done in collaboration with the DOE Next-Generation Ecosystem Experiment (NGEE), which is studying soil processes (e.g., permafrost degradation) in the Arctic tundra. Observations over open water and sea ice are also important because of the sharp contrast in energy fluxes between these surfaces.

The short-term focus for the NSA region will be on enhancing the measurement capabilities there, although the vision for the NSA megasite also includes modeling. It is likely that regional-scale modeling will be emphasized on the NSA because of complex geography including the Arctic Ocean to the immediate north and a mountain range to the south. It is also expected that long-range transport is likely to be important for aerosol and cloud processes.

2.2 Continued Focus on Measurement Excellence to Advance DOE Climate Science Research

2.2.1 Continue to Build Excellence in Ground-Based Measurements

The ARM Facility has long been a leader in providing long-term observations from advanced instruments. The American Recovery and Reinvestment Act of 2011 significantly expanded ARM's measurement capabilities and provided the means to explore the spatial distribution of clouds and precipitation as never before. The Recovery Act also provided the means to obtain information about aerosol composition. Over the next decade, ARM will continue to work closely with the user community to determine how to best apply these measurement capabilities to address pressing science issues. This work is expected to include the application of multiple sensors to better characterize the physical properties of clouds and aerosols.

At the SGP, scanning radars are being redeployed from the former tropical ARM sites to provide greater sampling flexibility and higher-density clouds sampling near the Central Facility close to Lamont, Oklahoma. Improvements in the characterization of aerosols will require future investments but should include instruments to better characterize aerosol composition and measurements to characterize dry and wet deposition. Dry deposition may be accomplished through the measurement of near-surface aerosol fluxes while rain sampling would support analysis of wet deposition.

The NSA already includes extensive instrumentation for measuring cloud and precipitation properties with scanning radars, lidars, radiometers, and instruments particularly designed for cold environments such as a high-resolution snow camera. However, there are several improvements that ARM is likely to implement in the coming several years including an emphasis on in situ and remote-sensing measurements of frozen precipitation, which is central to many Arctic processes. In terms of aerosols, important areas of development on the NSA will include advances in ice nuclei measurements and techniques for determining the vertical distribution. Promising methods for measuring aerosol profiles include routine in situ sampling using UAS and application of a recently developed technique to derive aerosol optical properties using multi-frequency lidar.

At all the ARM sites, there will be an emphasis on using multiple instruments to probe atmospheric properties. Each ARM site now includes multi-frequency scanning radars that are designed to provide new information about hydrometeor properties. Exploiting this information will be a focus of attention. There has been a great deal of interest recently in applying radar and lidar data to derive vertical velocity fields. Vertical velocity is central to many atmospheric processes but until recently was considered to be out of reach. But significant advances in this area have occurred recently and developing vertical velocity fields that span a wide range of conditions will also be a focus of attention.

In addition, the aerosol observing systems are undergoing a standardization to facilitate the use of data from these systems by the research community. Meanwhile, ARM is working with the research community to determine the optimum measurement sets at each of these sites.

2.2.2 Direct the Use of the ARM Mobile Facilities to DOE High-Priority Science

Climate models exhibit problems in conjunction with specific meteorological regimes and associated geographical regions. ARM mobile facilities (AMFs) provide the means to extend ARM measurements to diverse climate regimes and areas of key atmospheric phenomena in order to advance scientific understanding and improve models in these regimes. Some settings that have not yet been explored include the Southern Ocean, arid regions, and regions influenced by significant orography or urban development. While AMF deployments are typically only for a single year, such deployments provide much more information about local variability than a typical field campaign, which would often last only a few weeks. AMFs already have been used to study diverse phenomena including maritime boundary layer and deep convective clouds, orographic clouds, the African monsoon, cloud-aerosol interactions, and aerosol processing.

Figure 4. The second ARM Mobile Facility deployed on the *Horizon Spirit* for a marine boundary layer cloud experiment in the Eastern Pacific.

Looking ahead, we anticipate that the ARM Mobile Facilities will continue to be used to explore diverse climate regimes around the world, although we are also considering possible refinements to the deployment strategy. Now that a wide variety of environments have been sampled, it would be desirable to begin to constrain where the AMFs are deployed. The selection of these locales could draw on previous workshops that sought to identify high-priority measurement areas (e.g., US DOE 2007) or further assessment could be done to identify additional regions of model uncertainty. As the climate research community identifies regions where measurements are most needed, focused calls for deployment to those regions could be constructed. As the megasites mature, these calls could specifically request that an AMF be used in conjunction with one of the megasites. In this way, the community would have flexibility to design some details of a deployment but there would be a clear mechanism to meet high-priority climate science measurement needs.

2.2.3 Apply the Eastern North Atlantic Site to Study Maritime Stratocumulus

In 2009–2010, the first AMF was deployed to Graciosa Island, part of the Azores island chain in the Eastern North Atlantic (ENA). This is one of several regions in the world dominated by marine stratocumulus, which climate models have a notoriously difficult time simulating accurately. The AMF campaign provided an unprecedented opportunity to study a full annual cycle but because of the importance of the region, support was provided to make the site a long-term facility. Operations with many of the ARM instruments at the Azores site began in at the end of FY13. Most of the remaining instruments will be installed by the end of calendar year 2014.

An important question at the Azores locale is the role of aerosols in affecting stratocumulus cloud properties and associated drizzle. The Azores site can be subject to a range of aerosol concentrations depending on the trajectory of the air mass at a given time. Sorting out the effects of aerosols will require examining many cases to separate aerosol effects from other meteorological factors. The long-term nature of the ENA deployment will provide the perfect opportunity to explore the role of these various factors.

In addition to the extended nature of the deployment, the ENA site also incorporates important measurement advantages over the original AMF campaign. Key instrument additions include a Raman lidar, a Doppler lidar, and a scanning precipitation-sensitive radar. The Raman lidar provides continuous measurements of water vapor and temperature profiles while the Doppler lidar provides profiles of vertical velocity. Together, these two lidars give important information about the boundary-layer structure that will be critical for understanding the marine stratocumulus. The precipitation radar will provide information about the spatial and temporal distribution of drizzle. Drizzle is a very important part of the life cycle in these marine clouds, so this information will also be valuable in unraveling the factors governing their life cycle under various conditions.

2.2.4 Adapt the ARM Aerial Facility to Support the Next-Generation ARM Facility

Airborne observations enhance the surface-based ARM measurements by providing *i)* vertical and horizontal context, *ii)* evaluation of remote-sensing measurements made from space or the surface, *iii)* data for development of model parameterizations, and *iv)* information necessary for process studies that is not available from surface or space-based remote-sensing methods. ARM has primarily used manned aircraft to address these aerial measurement needs, though there have been a few experiments with UAS. In most of the United States, current regulations make flying UAS very difficult. However, access to special restricted four-nautical-mile-diameter air space over the Oliktok site is providing the opportunity to undertake UAS operations over the NSA. UAS operations will be particularly valuable in this region for routine sampling of both vertical structure and horizontal heterogeneity. ARM is now assessing how both unmanned and manned flights can best serve the vision of integrated observations and simulations over the next decade.

Numerous UAS-compatible sensors measuring platform and atmospheric state, gases, radiation, and aerosol properties are currently available and many more are in development. Through a series of workshops, ARM has obtained feedback from the scientific community on what it should try to achieve scientifically with UAS. ARM does not currently possess in-house UAS capabilities. However, we envision acquiring a range of UAS platforms and to procure and develop compatible, well-characterized sensors to be maintained and operated by the ARM Aerial Facility (AAF). Missions to be executed with the instrumented UAS would be selected using the existing ARM Facility proposal process. The decision of which UAS platform(s) and sensors to use would be driven by scientific requirements. Guest sensors could be proposed to complement AAF-owned instrumentation. We envision AAF offering the same level of service to the scientific community in the UAS arena currently offered with the piloted platforms. This will include data processing and archival.

UAS would be operated in conjunction with an ARM fixed or mobile site and site operations would support AAF with logistics and personnel. Due to the special situation at the Oliktok deployment, UAS operations will start at that location but later expanded to the Barrow site, SGP, ENA, and other AMF deployment sites, including ships. In addition to proposal-driven missions, we envision ARM will establish routine UAS observations at one or several of these sites to build up statistically robust observational databases.

The payload capacity of UAS is orders of magnitude smaller than those of piloted aircraft. Many of the currently required instruments cannot be miniaturized to be compatible with UAS in the foreseeable future. Furthermore, new, cutting-edge instruments often require hands-on interaction from an onboard operator before they can be deployed in an autonomous fashion. Finally, UAS currently have no de-icing capabilities (prototype systems are in development), precluding flying them into mixed-phase clouds.

Characterization of mixed-phase clouds in the Arctic has been particularly difficult. There have been several ARM campaigns focusing on mixed-phase clouds but a long-term database is extremely important. A long-term characterization with a piloted aircraft is the only way to build confidence in ground-based retrievals and model simulations. Therefore, we envision ARM to carry out year-round routine flights at NSA with a well-instrumented aircraft (clouds, aerosol, radiation, gases). Instruments deployed on tethered balloons can augment this activity.

The AAF currently has two dedicated aircraft (a single-engine Cessna 206 and a Gulfstream G-159, or G-1). Neither of these is suited for year-round observation at NSA (cold temperature operating limits). We also know the G-1 will need to be replaced within the next 6–7 years and expect that a dedicated aircraft will continue to be in high demand by the research community. In 2015, ARM will participate in a workshop with other Climate and Environmental Sciences Division (CESD) programs to determine the requirements for future aerial measurements. A replacement of the G-1, with a carefully selected newer aircraft (e.g., DeHavilland Dash-8-200), could serve multiple purposes (e.g., routine-flight campaigns on the NSA and specific principle investigator-awarded campaigns for ARM and other programs).

2.3 Implement Routine High-Resolution Modeling with a Focus on US ARM Megasites

2.3.1 Modeling Goals

Enhancements to the SGP and NSA sites are designed to support high-resolution modeling at these sites. By running a model more frequently than during occasional field campaigns, the continuous nature of ARM observations will be more fully utilized. The concentration of measurement resources will provide the observational basis necessary to support routine modeling including the implementation of a continuous dynamical forcing data set that provides the means to initialize and constrain model simulations.

Routine operation of high-resolution models at ARM sites will begin at the SGP site. High-resolution simulations provide the means of detailed process studies and detailed comparisons with ARM observations. In the short term, there will be an emphasis on the shallow convection case, taking advantage of the planned enhanced measurements at the SGP and associated dynamical forcing data set.

While developing the local-scale framework, we will also explore strategies for studying deep convection. The question will be how to optimally organize ARM resources to support the modeling of large-scale systems. A valuable tool for this work will be observing system simulation experiments (OSSEs). These numerical experiments make use of instrument simulators embedded in model simulations and can be used to assess the influence of diverse measurement configurations on simulations. Carrying out a series of OSSEs in parallel with early local-scale work will provide the means to develop strategies for more complex large-scale work. Similar studies could also be applied to supplementing measurements on the NSA.

It is envisioned that routine modeling at the SGP site—and eventually at the NSA site—will serve several communities. First, the availability of routine model simulations will be valuable for the non-modeling community to combine with observations in order to provide a more complete data set for process studies. The merging of observations from a dense network with high-resolution simulations has the potential to provide the best possible description of the atmospheric volume for process studies and validation of other models. Meanwhile, the dynamical forcing data set will be required to perform these routine simulations and will be valuable to the modeling community as it will provide the means to run a variety of models under controlled conditions.

Routine model simulations are also expected to be of great value for evaluating processes in global models over a wide range of conditions. In this way, the operation of high-resolution models at the ARM megasites will provide a bridge to global-scale models.

A key point is that, at present, there has been little discussion about applying the observation-model framework to an AMF site—or to the fixed-location site in the ENA. It would be quite desirable to do so, although a precondition would be to develop appropriate forcing data sets for these locales. An expected outcome of operating the framework at the SGP and NSA will be to develop a better understanding of developing forcing data sets. A desirable outcome of that work will be to ultimately apply the high-resolution modeling framework to additional sites to better engage the modeling community in AMF deployments.

2.3.2 Developing the Modeling Framework

There are many aspects of the modeling work to consider including what model to use, over what domain, with what resolution, what output to save, and what computing infrastructure will be required. To effectively address these issues, the first phase of this modeling work will be to undertake a modeling pilot study over the next one to two years that focuses on developing a detailed modeling strategy and addressing these issues. The main outcome of this study will be a modeling framework that will then be implemented first for the SGP site on a routine basis.

The computational demands of running these models and the output from these simulations will be considerable. The data volume is likely to rival that of high-output instruments such as the scanning radars. Because of these known demands, a focus of the pilot study will be to recommend an achievable balance between simulation parameters and computational resources. The additional data volume also points to a need for community analysis resources, which has already been initiated in conjunction with the ARM Data Archive as a result of the growth in measurement data volume following the Recovery Act.

A significant contributor to quantifying the data volume produced by routine simulations will be the choice in what output is saved. At the high temporal resolution with which LES and CRMs are run, it will not be possible to save all output. Most likely, we will be saving complete output snapshots, which can be used to re-initialize a model run, along with detailed statistical output. Many modeling systems include packages that provide detailed summaries of data. But even here, the output from these systems is highly configurable, so there will need to be a careful analysis of what output is most useful for routine analysis with ARM observations.

2.4 Enhancing Data Products and Processes

2.4.1 Data Processing to Bridge Observations and Models

The third component of the overarching ARM strategy described in section 2.1 is to develop new data products and tools to bridge observations and models. Factors that complicate the use of observational data for model analyses include differences in spatial-temporal resolution and differences in reported parameters. To bridge the spatial-temporal gap between observations and models, ARM will explore a range of data-processing avenues, including synthesis data products, tools for merging and gridding data, statistical summaries of observations, instrument simulators, and support for a community software development space.

Synthesis data products merge observations from multiple instruments to a common grid. A current example is the ARM Best Estimate product, which has facilitated the use of ARM data in numerous cloud-modeling studies. Expanding the parameters available through such merged data sets, and providing tools for custom merging, would facilitate consistency checks among variables, multi-parameter analyses, and assimilation by models.

Data assimilation offers several powerful outcomes depending on what ARM data are assimilated. Assimilating the background atmospheric state (temperature, humidity, and winds) provides the means of developing a continuous model forcing data set that can then be used by a wide variety of models. If, instead, cloud and aerosol measurements are also assimilated, the product would represent the convergence of all that is known about the atmospheric volume and would represent the best possible description of that volume for process studies and other analyses.

Facilitating the co-analysis of multiple parameters, combined with the long-term, continuous nature of ARM data, provides the basis for developing statistical summaries of observations and parametric relationships. These summaries could be evaluated for a range of time scales and provide a valuable means of bridging the temporal and spatial gap between ARM observations and models. This sort of statistical analysis is particularly valuable for this high-resolution analysis where there is not an expectation that the model will exactly simulate the spatial-temporal distribution of atmospheric parameters. Instead, the expectation is that the model will reproduce representative statistical distributions.

There has been an emphasis within ARM on creating value-added products (VAPs), which typically involve deriving high-order geophysical parameters from measured parameters—such as deriving cloud ice water content from a radar, which naturally measures reflected power. Another approach to bring measurements and models closer together is to embed instrument simulators within models. These simulators derive the equivalent of instrument measurement fields, such as radar reflected power, from model output variables. Comparing a model and a measurement through an instrument simulator involves many of the same assumptions and uncertainties as a comparison through a VAP, but applying a simulator may be more efficient to implement.

In addition to facilitating the comparison of measurements and models, instrument simulators also enable OSSEs. In an OSSE, a virtual instrument is deployed in a model. The virtual instrument can be inexpensively operated in a variety of configurations and locations to determine the optimum deployment for a particular application. In this way, an OSSE can be used to inexpensively design real measurement experiments. It is anticipated that OSSEs will be used to guide future instrument deployments in conjunction with the SGP and NSA megasites.

Of course, the development of traditional VAPs will continue to be an important component of ARM data activities. There will continue to be a need to derive these high-order products from the raw measurements. ARM is working to improve the processes for developing these products by implementing community standards and developing tools like the ARM Data Integrator (ADI) that increase efficiency through a higher level of automation. ADI is an open-source software framework that automates data retrieval, merges diverse data sets, and creates ARM data products that conform to ARM standards. Algorithm development within the ADI framework is currently supported in C, Python, and interactive data language (IDL) (Gaustad et al. 2014). ARM tools like ADI already lend automation and transparency to many ARM data products, and these tools will be expanded further to facilitate greater automation across ARM data holdings.

Another way to accelerate product development is to engage the ARM user community. Open-source software development facilitates community engagement and collaboration. ARM will use modern software engineering practices (such as modular frameworks, common data models, unit testing, and continuous integration) and collaborative platforms (such as GitHub). The Python-ARM Radar Toolkit (Py-ART; http://github.com/ARM-DOE/pyart) is successfully using open-source practices. Py-ART has hundreds of users worldwide and has attracted many contributions from inside and outside of DOE. Working with an open-source community coding paradigm, ARM will greatly accelerate and ease the processes of getting state-of-the-art, user-driven codes into operation.

The ARM data life cycle infrastructure is also moving toward providing on-the-fly VAPs as a way to engage the user community. A workflow and provenance environment is needed to easily reproduce these products for anyone requesting the results of work entered into the ARM data infrastructure. A standards-based workflow will be adopted to capture the provenance information. Remote services, including analysis and visualization as part of the BER Virtual Laboratory Infrastructure (BERAC 2013), will be connected through a provenance application programming interface (API) to automatically capture meaningful history and workflow.

2.4.2 Improve the Discoverability of ARM Data

The number and variety of ARM data sets are constantly growing. The challenge for the user community is to identify the data set that best meets their need. ARM recently deployed a powerful data discovery tool to enable users to search and order 4000 different ARM data products. The ARM Data Discovery Browser (http://www.archive.arm.gov/discovery/#v/home/s/) is built on a service-oriented architecture, which will allow the ARM Data Center to scale the tool for future data expansion. The current architecture allows any future visualization and web technologies to integrate with the data discovery tool. ARM anticipates building new web tools to support data discovery. These tools may include on-demand and interactive visualization capabilities, data extraction, and sub-setting capabilities. These services can also be accessed using a set of standards-based APIs, which will allow the BER Virtual Laboratory services to directly query and access the ARM data.

Currently, the data discovery tool is focused on standard operational data products, but work is underway to include non-standard data, including field campaign data and data products that are under evaluation. Because of the specialized nature of these classes of data, they are often the preferred source of information for a given parameter for limited periods. With so many data products, communicating the preferred source of a given parameter for a particular application is a significant challenge. ARM has developed a list of recommended datastreams that have been integrated into the current ARM Data Discovery Browser. We will continue to define and refine this list to match user needs with data sources.

Ultimately, the path to enhanced ARM data discovery will be through accurate and clear metadata collection practices. Over the next 10 years, the ARM Facility will fund additional staff and focus resources to improve the metadata process. ARM metadata categories will be mapped to community standards, such as the Climate and Forecast (CF) Conventions, whenever possible, increasing ARM data discoverability by outside agencies and programs. The Facility will work strongly toward more timely metadata assignments for all ARM data, including baseline, VAPs, principal investigator (PI), and field campaign data. The ARM Data Product Registration and Submission Tool will continue to be used and improved upon to capture metadata from the PIs directly, as part of the effort to streamline the metadata collection and assignment process. The complete process will be documented and accessible via the ARM website.

2.4.3 Improve the Characterization and Communication of Data Quality and Uncertainty

A key criterion in determining the utility of scientific data is quality. It is important that researchers using ARM data have a clear understanding of the general state of data they are using in their analysis (is the instrument functioning properly, was it recently calibrated, etc.) and of the quantitative uncertainty of a measurement.

The ARM Facility has implemented a dual approach for assessing general data quality. The first level of characterization uses automated quality checks to quickly highlight substandard data using analysis methods well understood at the time of file creation. The second level occurs following the creation of a data file and allows for evolution of data quality characterization disconnected from the processing of the data itself. The disconnected format simplifies the characterization of data quality after data have been processed.

Currently, much of the second-level quality assessment occurs through technical evaluations by the ARM Data Quality Office (DQO) and technical leads for a given instrument or data product. However, ARM is currently implementing an improved data format standard based on the community-developed CF standard to facilitate hands-off bulk processing. These data standards and processing developments will also make it easier to implement more complex automated quality control procedures. By standardizing the processing environment, new or established open-source quality control procedures can be implemented, providing users access to well-honed methods of removing substandard data from the analysis and allowing other users to benefit from the quality control work of the community.

In addition to the general quality of data, the data product uncertainty must be determined and documented. This uncertainty includes the basic measurement characteristics, influences from the environment and instrument state, and where appropriate, assumptions made through data processing, such as a complex parameter retrieval. For some basic measurements, the determination of uncertainty is relatively straightforward, though even in cases where uncertainties are prescribed by the vendor, field conditions must be taken into account. The propagation of uncertainties must be quantified and communicated to the scientific data user. ARM is actively applying standard approaches to quantify the uncertainties across the wide variety of measurements. However, for many complex measurements, this characterization will constitute a research activity that will be carried out in partnership with the research community.

Achieving the optimum level of reporting for both quality status and uncertainty will require close communication with the user community. These are complex issues and could consume a large portion of ARM development activities. Close communication with the user community in assessing and prioritizing ARM activities related to data quality will ensure that the appropriate balance is achieved.

2.4.4 Use Digital Object Identifiers to Link Data to Citations and Background Information

The ARM Facility assigns Digital Object Identifiers (DOIs) to regular ARM datastreams, and for PI and field campaign products. DOIs allow users to cite ARM data in research articles and, conversely, allow publishers to link journal articles to the cited data. DOIs also provide the means to give credit to ARM staff responsible for a given data set. This feature, in turn, provides the means to point users to the appropriate source of information and contact reference for a data set in a dynamic way. DOIs also provide a means to assess the usefulness of ARM data. In the future, the ARM Facility's data generation process will consider adding DOIs as part of the header, which will allow users to easily access the DOI.

The ARM Facility is currently setting up an industry standard for linking data with publications using a new data citation structure. ARM staff are working with major scientific publishers (e.g., Elsevier, Thomson Reuters, etc.) to directly link articles that cite and refer to ARM data to that data. These efforts will continue and expand over the coming decade and will enable the ARM website to display relevant publications for a given ARM measurement.

2.4.5 Ensure the Security and Stability of ARM Data and Software

Currently, the ARM Facility uses regular backups to ensure proper recovery from failures. Oak Ridge National Laboratory (ORNL) systems rotate back-up tapes offsite. Currently the offsite tape back-up is in a building on the ORNL campus. The ARM Facility is establishing a secondary back-up repository at a second DOE laboratory. Measurement facilities also generate back-up copies to a local disk. As bandwidth increases and the cost for secure cloud storage decreases, ARM will use these cloud services to maintain online back-ups wherever possible. This will allow ARM systems to remain online in the event of a systems/network or more catastrophic failure (e.g., natural disaster) as the systems can be replicated offsite. In the event that the cost does not become reasonable, then this offsite replication can also occur between other national laboratories and will provide the same protection. Bandwidth between laboratories is inexpensive. Research will ensure the most cost-effective solution is chosen. Detailed plans will also be written to outline recovery options or processes.

In addition to preparing against catastrophic loss, we must also plan for change. Every piece of software ARM uses to accomplish its mission will need to adapt multiple times to changing systems and environments over the years to come. This creates a challenge to maintain an operational system with software that has dependencies on legacy components. To mitigate this, ARM will continue to proactively invest in plans to migrate software to the updated technology. In addition, continued investment in building and maintaining shared components maximizes code reuse and reduces the complexity of migrations. There will be work to automate testing and reviews, especially for data generators. Wherever possible, open source, or other well-supported software components, will be considered for use.

As the ARM infrastructure continues to evolve and new services come online in support of the ARM mission, the cyber threat landscape will continue to increase. ARM staff will become adept at mitigating these risks in all phases of the software development life cycle. Developers will continue to learn secure coding practices. Administrators will continue to learn how to effectively secure ARM infrastructure. Cyber professionals will continue to learn new methods to protect the infrastructure. Failure to continue to grow in these areas will increase risk to the ARM infrastructure. The goal will be to better automate each of these processes.

2.4.6 Integrate ARM Data with other BER Climate Measurements and Simulations

The ARM Facility is currently establishing a flexible adaptive architecture to meet future data management challenges. This new architecture will allow individual cyber-infrastructure components (Figure 5) to function independently but communicate with other components to support the overall data management. These individual components can be scaled independently without affecting the overall workflow (Figure 6). As an example, databases can be modernized to take advantage of new technologies while maintaining the communication layer between tools and applications using standards-based web services and API calls. Similarly, analytical and visualization tools used with ARM data could be upgraded independently while using the same database components. In support of our user community, this architecture will provide an effective framework for the use of existing tools such as those developed through the Program for Climate Model Diagnosis and Intercomparison (PCMDI; http://www-pcmdi.llnl.gov). This approach also provides a mechanism for the ARM Facility to contribute new tools and components to broader communities.

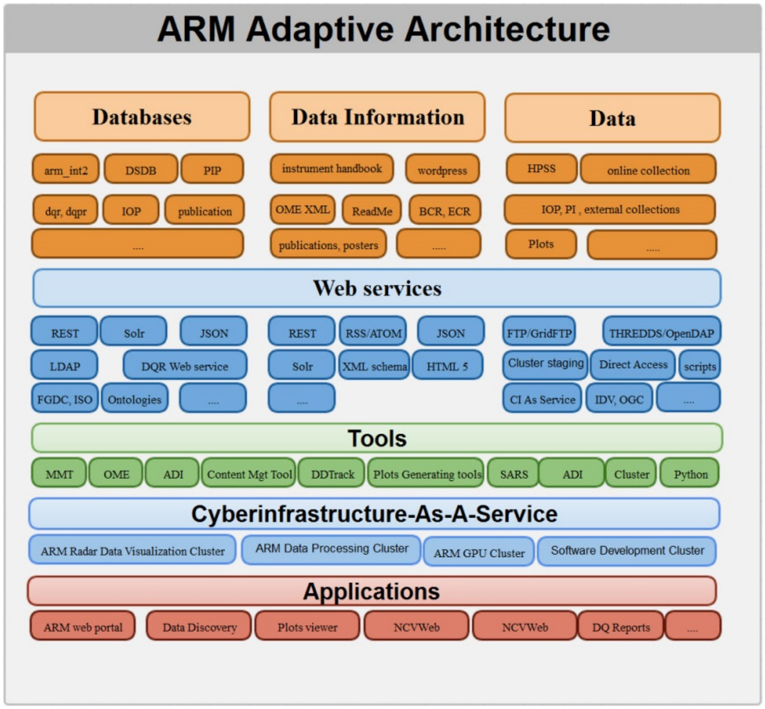

Figure 5. Support for open data architecture and integration with other community portals.

This adaptive architecture will also allow the ARM Data Center to integrate and serve as one of the data center nodes in the Williams et al. (2014) proposed BER Virtual Laboratory Initiative (VLI). For each BER CESD scientific domain included in the BER VLI, both data and metadata will be archived and accessed from existing data centers. Participating data centers will be part of one or more virtual scientific focus groups, allowing for sharing of data and metadata services with other data centers in the same scientific domain. A software stack will co-evolve to share data, metadata, data quality information, ontologies, visualization, and analysis services between VLI data centers. In order to support a powerful and flexible access model, each service hosted on a data center will be exposed through a simple and well-documented service API (layered with security when appropriate). Through this API, clients can easily execute invocations and possibly chain requests in complex scientific workflows.

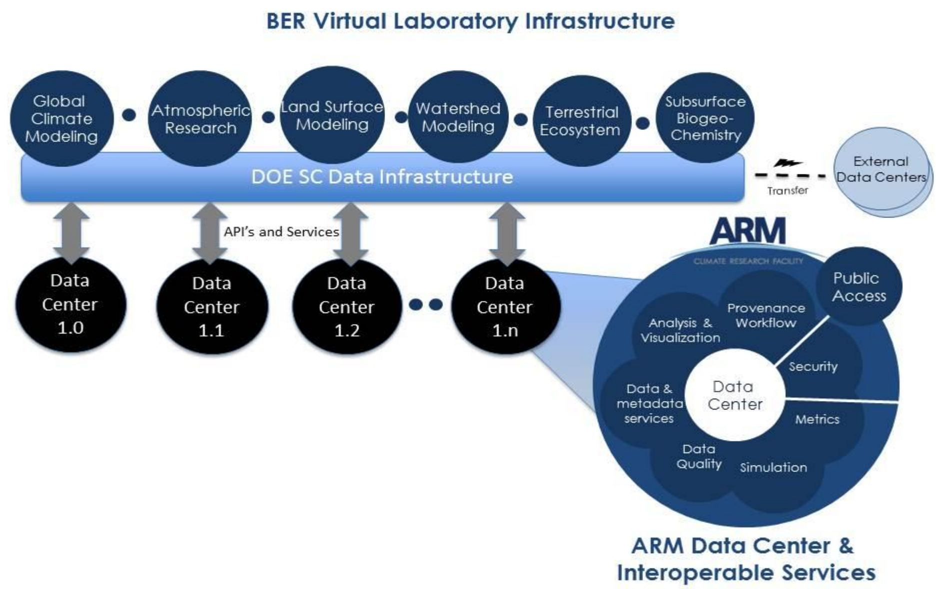

Figure 6. Connecting ARM with the Office of Science Data Infrastructure.

The BER VLI proposed architecture would have an exemplary system for sharing data and metadata records based on community-developed standards such as International Organization for Standardization (ISO) 19115, Federal Geospatial Data Committee (FGDC), OAI-PMH, Thematic Real-time Environmental Distributed Data Services (THREDDS) and Open Geospatial Consortium (OGC). These standards are already used across the ARM Facility. The architecture will reuse some of the metadata creation tools such as the Online Metadata Editor (OME), which is currently used by many DOE projects, including ARM. Using OME will not only allow users to register their data sets, but also to use consistent keywords using standards such as the CF and Global Change Master Directory-controlled vocabularies. The BER VLI architecture will have a common resource registry, allowing ARM to register their resources—such as tools, web services, and domain expertise—using common protocols and standards already used by the ARM Adaptive Architecture.

The ARM Data Center is currently publishing data in other community portals such as the Earth System Grid Federation (ESGF), International Arctic Systems for Observing the Atmosphere (IASOA), and NGEE Arctic. ARM is enabling this service by using community-developed standards and protocols such as CF, OpenDAP, THREDDS, FGDC, and ISO 19115. ARM's Adaptive Data Architecture will continue adding ARM data in other relevant portals in the future.

2.5 Strengthen and Expand Interactions with the User Community to Establish Priorities and Maximize the Science Impact of ARM

The ARM Facility has continuously maintained a close relationship with its user community, particularly with the ASR program, which is comprised of the science component from the original ARM program and the DOE Atmospheric Sciences Program, or ASP. High-level communications have been managed through a steering committee that combines leaders of the ASR working groups and the ARM management team. However, as a user facility, ARM serves a broad, international science community. To better facilitate communication with the broader user community, the ARM Facility will institute a User Executive Committee (UEC) that will be chartered to serve as a liaison between the Facility and the user community and to provide feedback. With the many changes underway in the ARM Facility, effective communication with the user community will be important to ensure that the direction of the Facility changes are coordinated with the needs of the climate research community.

The UEC will provide a formal mechanism to solicit feedback, but ARM strives to engage with the research community in a variety of ways. To help users and potential users of the ARM Facility more easily use its capabilities, ARM conducts introductory tutorials at ASR meetings on the access and use of ARM data. In the future, we will explore other potential venues for these tutorials, such as summer workshops for graduate students or professional conferences. We will also explore expanding the material presented through these tutorials.

Climate research is very much an international activity and ARM Facility users currently represent more than 30 countries. To engage with this international community, ARM participates in activities associated with the World Climate Research Program. Recently, ARM co-organized a workshop with colleagues from European observation programs, designed to find ways to collaborate to benefit all participants. The ARM Facility plans to continue to seek opportunities to collaborate with the global community and build on the European collaborations established over the past few years to maximize the global impact of ARM observations.

3.0 Five-Year Plan

The ARM Facility is embarking on a major development activity that is designed to bring to fruition long-standing goals to link observations and models. This document has briefly outlined many facets of this plan. Table 1 captures key milestones over the next few years associated with its implementation.

Table 1. Key Milestones.

	Milestone	Date	Owner
1	Upgrade the ARM Data Discovery Browser to include field campaign and principle investigator-contributed data	12/1/14	Chief Operating Officer
2	Form the ARM User Executive Committee	12/31/14	Technical Director
3	Initiate modeling pilot study	3/31/15	Technical Director
4	Initiate science flights using UAS and Tethered Balloons at Oliktok	6/30/15	NSA Site Manager
5	Begin offsite tape back-up of the ARM Data Archive	9/30/15	Data Services Manager
6	Begin enhanced operations at the SGP	9/30/16	SGP Site Manager
7	Implement analysis environment and tools for combined model-observation data	12/31/16	Data Services Manager
8	Begin routine modeling at the SGP	4/1/17	Technical Director
9	Implement enhanced measurements at NSA to support routine modeling	9/30/18	NSA Site Manager
10	Begin routine modeling at the NSA	9/30/19	Technical Director

4.0 Bibliography

Ackerman, TP and GM Stokes. 2003. "The Atmospheric Radiation Measurement Program." *Physics Today* 56(1): 38–44. DOI:10.1063/1.1554135.

Biological and Environmental Research Advisory Committee (BERAC). 2013. *BER Virtual Laboratory: Innovative Framework for Biological and Environmental Grand Challenges; A Report from the Biological and Environmental Research Advisory Committee*. DOE/SC-0156. http://science.energy.gov/~/media/ber/berac/pdf/20130221/BERACVirtualLaboratory_Feb-18-2013.pdf.

IPCC. 2014. *Climate Change 2014 Synthesis Report: Summary for Policymakers*, https://www.ipcc.ch/pdf/assessment-report/ar5/syr/AR5_SYR_FINAL_SPM.pdf.

Mather, JH and JW Voyles. 2013. "The ARM Climate Research Facility: A Review of Structure and Capabilities." *Bulletin of the American Meteorological Society* 94(3): 377–392. DOI:10.1175/BAMS-D-11-00218.1.

Gaustad, KL, TR Shippert, BD Ermold, SJ Beus, A Borsholm, and KM Fox. 2014. "A scientific data processing framework for time series netCDF data." *Environmental Modelling and Software* 60: 241–249. DOI:10.1016/j.envsoft.2014.06.005.

Stokes, GM and SE Schwartz. 1994. "The Atmospheric Radiation Measurement (ARM) Program: Programmatic background and design of the cloud and radiation testbed." *Bulletin of the American Meteorological Society* 75(7): 1201–1221. DOI:10.1175/1520-0477(1994)075<1201:TARMPP>2.0.CO;2.

U.S. Department of Energy. 1990. *Atmospheric Radiation Measurement Program Plan.* DOE/ER-0441, http://www.arm.gov/publications/archive.

U.S. Department of Energy. 2007. *Report on the ARM Climate Research Facility Expansion Workshop.* DOE/SC-ARM-0707, https://www.arm.gov/publications/programdocs/doe-sc-arm-0707.pdf.

U.S. Department of Energy. 2008a. *Atmospheric Radiation Measurement Program: Accomplishments from the Science Program and User Facility.* DOE/SC-ARM/P-08-017, https://www.arm.gov/publications/brochures/docs/doe-sc-arm-p-08-017.pdf.

U.S. Department of Energy. 2008b. *ARM Climate Research Facility Workshop Report.* DOE/SC-ARM-0804, http://science.energy.gov/~/media/ber/pdf/Doe_sc_arm_0804.pdf.

U.S. Department of Energy. 2010. *Atmospheric System Research (ASR) Science and Program Plan.* DOE/SC-ASR-10-001, http://science.energy.gov/~/media/ber/pdf/Atmospheric_system_research_science_plan.pdf.

U.S. Department of Energy. 2012a. *Climate and Environmental Sciences Division Strategic Plan.* DOE/SC-0151, http://science.energy.gov/~/media/ber/pdf/CESD-StratPlan-2012.pdf.

U.S. Department of Energy. 2012b. *Biological and Environmental Research Climate and Environmental Science Division U.S./European Workshop on Climate Change Challenges and Observations*. DOE/SC-0154, http://science.energy.gov/~/media/ber/pdf/CESD_EUworkshop_report.pdf.

U.S. Department of Energy. 2013a. *Atmospheric Radiation Measurement Program: Accomplishments in Atmospheric Science 2008-2013.* DOE/SC-0159 http://science.energy.gov/~/media/ber/pdf/Brochures/CESD_accomplishments2_v2.pdf,.

U.S. Department of Energy 2013b. *Biological and Environmental Research Climate and Environmental Science Division Atmospheric Testbed Workshop*. DOE/SC-0163, http://science.energy.gov/~/media/ber/pdf/Brochures/CESD_Atmospheric_Testbed_Workshop_V9.pdf.

U.S. Department of Energy. 2014a. *U.S. Department of Energy Strategic Plan 2014-2018.* http://www.energy.gov/budget-performance.

U.S. Department of Energy. 2014b. *ARM/ASR High-Resolution Modeling Workshop*. DOE/SC-0169, http://science.energy.gov/~/media/ber/pdf/workshop%20reports/doe-sc-0169-low-resolution.pdf.

U.S. Department of Energy. 2014c. *ARM/ASR North Slope of Alaska Priorities Workshop*. DOE/SC-0176, http://science.energy.gov/~/media/ber/pdf/workshop%20reports/NSA_WorkshopReport_2015_WEB.pdf.

Williams, DN, G Palanisamy, G Shipman, TA Boden, and JW Voyles. 2014. "U.S. Department of Energy Strategic Roadmap for Earth System Science Data Integration." *Institute of Electrical and Electronic Engineers 2014 Big Data Conference* (October 2014, Washington, DC). http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7004304.