U.S. DEPARTMENT OF ENERGY | Office of Science

DOE/SC-ARM/10-015

ARM User Survey Report

LR Roeder

June 2010

ARM

CLIMATE RESEARCH FACILITY

DISCLAIMER

DOE/SC-ARM/10-015

ARM User Survey Report

LR Roeder

June 2010

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

Summary

The objective of this survey was to obtain user feedback to, among other things, determine how to organize the exponentially growing data within the Atmospheric Radiation Measurement (ARM) Climate Research Facility, and identify users' preferred data analysis system. The survey findings appear to have met this objective, having received approximately 300 responses that give insight into the type of work users perform, usage of the data, percentage of data analysis users might perform on an ARM-hosted computing resource, downloading volume level where users begin having reservations, opinion about usage if given more powerful computing resources (including ability to manipulate data), types of tools that would be most beneficial to them, preferred programming language and data analysis system, level of importance for certain types of capabilities, and finally, level of interest in participating in a code-sharing community.

Not surprisingly, nearly two-thirds of respondents' work is related to climate research, and there is a fairly consistent timeframe of use of the data, as depicted from the responses. At present, nearly 36 percent stated they currently spend less than one week working with it, followed by 24 percent spending one week to one month, and approximately 16 percent spending one to three months. However, there was a noticeable shift in planned usage going forward. For example, those spending less than one week decreased to 16 percent from 36 percent, and interestingly, the one-to-three month choice increased from 16 percent to 23 percent.

Nearly 60 percent of respondents stated they might perform less than 10 percent of their data analysis on ARM-hosted computing resources. Further, while only half of respondents had reservations about downloading ARM data at ranges from up-to-one gigabyte (GB) to 10 GB, their reservations jump to nearly 80 percent at downloading up to 1 terabyte (TB). Although respondents did not feel strongly one way or another about the use of larger data sets if they were given access to more powerful computing resources, nearly all respondents *are* interested in using tools for manipulating and/or analyzing ARM data within an ARM computing center. Seventy percent of respondents stated that they would spend up to one month performing intensive data runs.

Respondents were highly interested in all tools listed as possibilities for manipulating and/or analyzing ARM data within an ARM computing center, save one. Of the other tools suggested, analytic/statistic, graphic, and data format tools received multiple votes. Two respondents recommend a tool that would allow time intervals to be specified (i.e., data recorded in 1-minute intervals, but specify 15-minute intervals). For preferred programming languages, respondents chose Fortran. Matlab, IDL, and C are the next-preferred languages, respectively. Windows is the preferred data analysis system, receiving greater than three-quarters of the responses.

Regarding capabilities, all respondents stated that computational resources/power, software development tools, data manipulation and visualization tools, as well as easy and rapid access to data, were highly important to them. In fact, easy and rapid access to data received nearly 70 percent of the responses.

Lastly, nearly three-quarters of respondents would be interested in participating in a code-sharing community similar to that of SourceForge.

Contents

1.0 Background

This survey was administered in an effort to determine optimal software and analysis tools for ARM data through feedback from existing users of the data. The Market and Competitive Analysis group at Pacific Northwest National Laboratory worked with web administrators to develop a landing page from which users could access the survey. The landing page allowed ARM to advertise the survey on its website, Facebook page, and on the Environmental Molecular Sciences Laboratory (EMSL) Facebook page. Additionally, ARM utilized its own "armall" distribution list and email introduction to further share the availability of the survey. The survey was open from September 30, 2009 through October 13, 2009.

2.0 Methodology

Due to the fact this survey was broadly accessible, no response rate was calculated; therefore, although the findings are valid, they should be considered qualitative in nature. The total number of responses equaled 301.[1]

Each section is organized according to responses to the question in Section 3.3, "What is your work related to?" For each question, answers are presented for (1) the total group of respondents, (2) respondents whose work is related to climate research, (3) respondents whose work is related to education, and (4) respondents whose work is related to other research.

For questions that included a field for users to type their own response, full answers are provided in the appendixes.

3.0 Findings

3.1 How much time per year do you currently spend working with ARM data?

Nearly 60 percent of respondents spend one month or less working with ARM data.

p	Chart	Frequency	Count
less than one week		**35.7%**	**104**
one week to one month		24.4%	71
one to three months		15.5%	45
three to six months		8.2%	24
more than six months		16.2%	47
Not Answered			10
	Valid Responses		**291**
(Respondents could only choose a single response)	**Total Responses**		**301**

[1] The difference between actual number of survey responses and valid responses is likely due to "drop-outs" or those who did not utilize the software as designed (e.g., tried to choose more than one "1st choice" rather than rank their order of preference as requested).

3.1.1 Climate Research

More than half of the climate research respondents currently use ARM data less than one month per year, and 19 percent work with the data more than six months per year.

Response	Chart	Frequency	Count
less than one week		**29.2%**	**54**
one week to one month		23.2%	43
one to three months		17.8%	33
three to six months		10.8%	20
more than six months		18.9%	35
	Valid Responses		**185**
(Respondents could only choose a single response)	**Total Responses**		**185**

3.1.2 Education

Nearly all education respondents currently spend one month or less working with ARM data, with the majority spending less than one week.

Response	Chart	Frequency	Count
less than one week		**61.9%**	**13**
one week to one month		23.8%	5
one to three months		4.8%	1
three to six months		4.8%	1
more than six months		4.8%	1
	Valid Responses		**21**
(Respondents could only choose a single response)	**Total Responses**		**21**

3.1.3 Other Research

Half of all other research respondents currently spend less than one week working with ARM data; in fact, approximately 78 percent stated they spend less than one month.

Response	Chart	Frequency	Count
less than one week		**51.1%**	**23**
one week to one month		26.7%	12
one to three months		15.6%	7
three to six months		2.2%	1
more than six months		4.4%	2
	Valid Responses		**45**
(Respondents could only choose a single response)	**Total Responses**		**45**

3.2 How much time per year do you plan to spend working with ARM data in the future?

In the future, the majority of respondents are planning on spending one week to one month working with ARM data. This is followed closely by respondents spending one to three months and more than six months.

Compared with current usage data, respondents plan to increase the time they spend working with ARM data. For example, those that currently spend less than one week working with the data plan to spend one week to one month working with ARM data in the future.

Response	Chart	Frequency	Count
less than one week		16.1%	47
one week to one month		**29.5%**	**86**
one to three months		22.9%	67
three to six months		10.3%	30
more than six months		21.2%	62
Not Answered			9
	Valid Responses		**292**
(Respondents could only choose a single response)	**Total Responses**		**301**

3.2.1 Climate Research

Based on the previous question where more than half of climate research respondents stated they currently use ARM data less than one month, it's interesting to note a shift in the time they plan to spend working with it in the future. For example, over this same time period, the percentage drops to less than 40 percent in the future and increases to approximately 52 percent over the one week to three months time frame. Moreover, there is an approximately 5 percent increase when comparing the 19 percent of respondents currently working with it more than six months to the 24 percent who plan to work on it in the future.

Response	Chart	Frequency	Count
less than one week		10.3%	19
one week to one month		**27.0%**	**50**
one to three months		24.9%	46
three to six months		13.5%	25
more than six months		24.3%	45
	Valid Responses		**185**
(Respondents could only choose a single response)	**Total Responses**		**185**

3.2.2 Education

Half of the education respondents plan to spend one week to one month working with ARM data in the future, twice as many as currently work with it over this time frame. Contrarily, only half as many education respondents plan to work with the data less than one week in the future, versus those that currently working with it over this time (30% vs. approximately 62%, respectively). An additional two respondents stated they would increase their time currently spent working with the data more than six months, approximately a 10 percent increase over this time frame.

Response	Chart	Frequency	Count
less than one week		30.0%	6
one week to one month		**50.0%**	**10**
one to three months		5.0%	1
three to six months		0.0%	0
more than six months		15.0%	3
Not Answered			1
	Valid Responses		**20**
(Respondents could only choose a single response)	**Total Responses**		**21**

3.2.3 Other Research

Eighty-seven percent of other research respondents plan to spend three months or less working with ARM data in the future, with the majority spending one week to one month as opposed to the current timeframe of less than one week. An additional two respondents stated they would increase their time currently spent working with the data more than six months, doubling the current number.

Response	Chart	Frequency	Count
less than one week		28.9%	13
one week to one month		**31.1%**	**14**
one to three months		26.7%	12
three to six months		4.4%	2
more than six months		8.9%	4
	Valid Responses		**45**
(Respondents could only choose a single response)	**Total Responses**		**45**

3.3 What is your work related to?

Nearly two-thirds of respondents' work is related to climate research. Only 7 percent of the respondents' work is in the educational field. Solar power, infrastructure, and radiation measurement were the most common statements when "Other" was selected. Respondents also mentioned clouds and data comparisons. For a full list of "Other" responses, see Appendix A.

Response	Chart	Frequency	Count
Climate research		**63.1%**	**185**
Other research		15.4%	45
Education		7.2%	21
Other. Please state:		14.3%	42
Not Answered			8
	Valid Responses		**293**
(Respondents could only choose a single response)	**Total Responses**		**301**

3.4 What percentage of your data analysis might you perform on ARM-hosted computing resources?

Nearly 60 percent of respondents stated they would perform less than 10 percent of their data analysis on ARM-hosted computing resources.

Response	Chart	Frequency	Count
Less than 10 percent		**59.5%**	**160**
11–25 percent		17.5%	47
26–50 percent		13.0%	35
51–75 percent		4.5%	12
76–100 percent		5.6%	15
Not Answered			14
	Valid Responses		**269**
(Respondents could only choose a single response)	**Total Responses**		**283**

3.4.1 Climate Research

According to survey findings, greater than half of climate researchers state they plan to perform less than 10 percent of their data analysis on ARM-hosted computing resources.

Response	Chart	Frequency	Count
Less than 10 percent		**55.7%**	**97**
11–25 percent		22.4%	39
26–50 percent		11.5%	20
51–75 percent		4.0%	7
76–100 percent		6.3%	11
Not Answered			2
		Valid Responses	**174**
(Respondents could only choose a single response)		**Total Responses**	**176**

3.4.2 Education

The majority of education respondents plan on performing less than 10 percent of data analyses on ARM-hosted computing resources.

Response	Chart	Frequency	Count
Less than 10 percent		**63.2%**	**12**
11–25 percent		21.1%	4
26–50 percent		10.5%	2
51–75 percent		5.3%	1
76–100 percent		0.0%	0
Not Answered			1
		Valid Responses	**19**
(Respondents could only choose a single response)		**Total Responses**	**20**

3.4.3 Other Research

Three-quarters of other research respondents plan on performing less than 10 percent of data analyses on ARM-hosted computing resources.

Response	Chart	Frequency	Count
Less than 10 percent		**75.7%**	**28**
11–25 percent		5.4%	2
26–50 percent		10.8%	4
51–75 percent		8.1%	3
76–100 percent		0.0%	0
Not Answered			1
	Valid Responses		**37**
(Respondents could only choose a single response)	**Total Responses**		**38**

3.5 At what volume would you have reservations about downloading a set of ARM data to your computer?

Greater than half of respondents (approximately 51 percent) have reservations about downloading ARM data at ranges from up to 1 GB to 10 GB. Their reservations jump to nearly 80 percent at rates up to 1 TB. Eleven percent of respondents have no reservations whatsoever regarding downloading ARM data sets.

Response	Chart	Frequency	Count
Up to 1 GB		25.3%	68
1–10 GB		**26.4%**	**71**
10–100 GB		17.5%	47
100–GB-1 TB		11.2%	30
1–10 TB		6.3%	17
10–100 TB		2.2%	6
No reservations		11.2%	30
Not Answered			14
	Valid Responses		**269**
(Respondents could only choose a single response)	**Total Responses**		**283**

3.5.1 Climate Research

While 27 percent of climate research respondents develop reservations about downloading 1–10 GB of ARM data, approximately 42 percent are reluctant to download *more than* 10GB. Ten percent of climate research respondents don't have any reservations.

Response	Chart	Frequency	Count
Up to 1 GB		20.9%	36
1–10 GB		**27.3%**	**47**
10–100 GB		18.6%	32
100 GB–1 TB		12.2%	21
1–10 TB		8.1%	14
10–100 TB		2.9%	5
No reservations		9.9%	17
Not Answered			4
	Valid Responses		**172**
(Respondents could only choose a single response)	**Total Responses**		**176**

3.5.2 Education

Between 1 and 10 GB is where most education respondents begin to have reservations about downloading ARM data sets.

Response	Chart	Frequency	Count
Up to 1 GB		15.8%	3
1–10 GB		**47.4%**	**9**
10–100 GB		15.8%	3
100 GB–1 TB		5.3%	1
1–10 TB		5.3%	1
10–100 TB		0.0%	0
No reservations		10.5%	2
Not Answered			1
	Valid Responses		**19**
(Respondents could only choose a single response)	**Total Responses**		**20**

3.5.3 Other Research

The majority of reservations about downloading ARM data range from up to 1 GB to 100 GB, with nearly 38 percent of the 37 responses claiming to have reservations beginning at up to 1 GB.

Response	Chart	Frequency	Count
Up to 1 GB		**37.8%**	**14**
1–10 GB		18.9%	7
10–100 GB		18.9%	7
100–GB-1 TB		8.1%	3
1–10 TB		5.4%	2
10–100 TB		0.0%	0
No reservations		10.8%	4
Not Answered			1
		Valid Responses	**37**
(Respondents could only choose a single response)		**Total Responses**	**38**

3.6 Are there larger, more complex data sets you would use given access to more powerful computing resources?

Respondents did not feel strongly either way regarding the use of larger data sets if given access to more powerful computing resources. Only 54 percent of respondents confirmed they would access larger data sets.

Response	Chart	Frequency	Count
Yes		**53.7%**	**145**
No		46.3%	125
Not Answered			13
		Valid Responses	**270**
(Respondents could only choose a single response)		**Total Responses**	**283**

3.6.1 Climate Research

Approximately 58 percent of climate researchers would access larger, more complex data sets if given access to more powerful computing resources.

Response	Chart	Frequency	Count
Yes		**57.5%**	**100**
No		42.5%	74
Not Answered			2
		Valid Responses	**174**
(Respondents could only choose a single response)		**Total Responses**	**176**

3.6.2 Education

Education respondents didn't feel strongly either way regarding using larger, more complex data sets if given more powerful computing resources.

Response	Chart	Frequency	Count
Yes		47.4%	9
No		**52.6%**	**10**
Not Answered			1
		Valid Responses	**19**
(Respondents could only choose a single response)		**Total Responses**	**20**

3.6.3 Other Research

Nearly 60 percent of other research respondents stated there were not more complex data sets they would use if given access to more powerful computing resources.

Response	Chart	Frequency	Count
Yes		41.7%	15
No		**58.3%**	**21**
Not Answered			2
		Valid Responses	**36**
(Respondents could only choose a single response)		**Total Responses**	**38**

3.7 If an ARM computing center provided tools for manipulating and/or analyzing ARM data, would you be interested in using such a center?

Nearly all respondents are interested in using tools for manipulating and/or analyzing ARM data within an ARM computing center.

Response	Chart	Frequency	Count
Yes		**90.0%**	**224**
No		10.0%	25
Not Answered			11
		Valid Responses	**249**
(Respondents could only choose a single response)		**Total Responses**	**260**

3.7.1 Climate Research

Nearly all climate research respondents would be interested in using tools for manipulating and/or analyzing ARM data provided by the ARM computing center.

Response	Chart	Frequency	Count
Yes		**92.6%**	**151**
No		7.4%	12
Not Answered			6
		Valid Responses	**163**
(Respondents could only choose a single response)		**Total Responses**	**169**

3.7.2 Education

If an ARM computing center provided tools for manipulating and/or analyzing ARM data, three-quarters of education respondents would be interested in using it.

Response	Chart	Frequency	Count
Yes		**75.0%**	**12**
No		25.0%	4
Not Answered			2
		Valid Responses	**16**
(Respondents could only choose a single response)		**Total Responses**	**18**

3.7.3 Other Research

If an ARM computing center provided tools for manipulating and/or analyzing ARM data, nearly 90 percent of other research respondents would be interested in using it.

Response	Chart	Frequency	Count
Yes		**88.2%**	**30**
No		11.8%	4
Not Answered			1
		Valid Responses	**34**
(Respondents could only choose a single response)		**Total Responses**	**35**

3.8 Initially, what types of tools in an ARM computing center would be most beneficial to you?

Respondents were asked to rate their overall level of interest on a scale from 5 to 1 where 5=highly interested and 1=highly uninterested.

Respondents are highly interested in all tools listed, save the single-column model (only moderately interested). Of the other tools suggested, analytic/statistic, graphic, and data format tools received multiple votes. Two respondents recommend a tool that would allow time intervals to be specified. "A tool where you can chose what time step you wish the data to be downloaded in. For example, I am in no need of radiation data in 1 minute intervals but would appreciate if there was a tool that could convert the 1 min intervals into hourly averages. That would reduce the size of the download as well." For a full list of "Other" responses, see Appendix A.

		5	4	3	2	1	Total
Data input/output	Count	**137**	64	26	5	14	**246**
	% by Row	**55.7%**	26.0%	10.6%	2.0%	5.7%	**100.0%**
Geophysical functions (like humidity conversion, solar zenith angle calculations)	Count	**106**	56	45	21	20	**248**
	% by Row	**42.7%**	22.6%	18.1%	8.5%	8.1%	**100.0%**
Radiative transfer code	Count	**99**	56	49	19	24	**247**
	% by Row	**40.1%**	22.7%	19.8%	7.7%	9.7%	**100.0%**
Single column model	Count	62	**64**	53	29	33	**241**
	% by Row	25.7%	**26.6%**	22.0%	12.0%	13.7%	**100.0%**
Model analysis tools (e.g., instrument simulator)	Count	**86**	61	40	28	25	**240**
	% by Row	**35.8%**	25.4%	16.7%	11.7%	10.4%	**100.0%**
Merging multiple parameters to a common time/height/space grid	Count	**125**	70	32	9	13	**249**
	% by Row	**50.2%**	28.1%	12.9%	3.6%	5.2%	**100.0%**
Plot/visualization	Count	**110**	63	43	17	16	**249**
	% by Row	**44.2%**	25.3%	17.3%	6.8%	6.4%	**100.0%**
Data quality	Count	**121**	59	42	13	10	**245**
	% by Row	**49.4%**	24.1%	17.1%	5.3%	4.1%	**100.0%**
Data comparison (including model/data comparisons)	Count	**103**	77	36	16	14	**246**
	% by Row	**41.9%**	31.3%	14.6%	6.5%	5.7%	**100.0%**
Other. Please state below.	Count	26	8	14	4	**33**	**85**
	% by Row	30.6%	9.4%	16.5%	4.7%	**38.8%**	**100.0%**
Total	**Count**	**975**	**578**	**380**	**161**	**202**	**2296**
	% by Row	**42.5%**	**25.2%**	**16.6%**	**7.0%**	**8.8%**	**100.0%**

3.8.1 Climate Research

The majority of climate research respondents are highly interested in all tools suggested. For a table of detailed answers, see Appendix B.

3.8.2 Education

Education respondents were interested in all types of tools listed with significant interest in data input/output, geophysical functions, model analysis, plot/visualization, and data quality tools. For a table of detailed answers, see Appendix B.

3.8.3 Other Research

Other research respondents are interested in all types of tools listed with significant interest in data input/output, geophysical functions, model analysis, plot/visualization, and data quality tools. For a table of detailed answers, see Appendix B.

3.9 What programming languages do you prefer?

We would like the core tools to be compatible with a variety of widely used scientific programming languages. Respondents were asked to choose only two.

The respondents' preferred programming language is Fortran. Matlab, IDL, and C are the next-preferred languages, respectively. Of the respondents who selected other, NCL is mentioned most frequently. For a full list of "Other" responses, see Appendix A.

Response	Chart	Frequency	Count
C		32.3%	84
Fortran		**61.9%**	**161**
Matlab		43.1%	112
IDL		33.8%	88
Python		12.3%	32
Other. Please state:		15.8%	41
	Valid Responses		**260**
(Respondents were allowed to choose multiple responses)	**Total Responses**		**260**

3.9.1 Climate Research

Fortran, Matlab, and IDL are the most preferred programming languages by climate researchers. There is also interest in C, but relatively little in Python.

Response	Chart	Frequency	Count
C		27.2%	46
Fortran		**68.6%**	**116**
Matlab		40.8%	69
IDL		38.5%	65
Python		11.2%	19
Other. Please state:		12.4%	21
		Valid Responses	**169**
(Respondents were allowed to choose multiple responses)		**Total Responses**	**169**

3.9.2 Education

Matlab, Fortran and C are the programming languages preferred by education respondents. When "other" was selected, VB.NET was suggested once.

Response	Chart	Frequency	Count
C		44.4%	8
Fortran		50.0%	9
Matlab		**55.6%**	**10**
IDL		27.8%	5
Python		11.1%	2
Other. Please state:		11.1%	2
		Valid Responses	**18**
(Respondents were allowed to choose multiple responses)		**Total Responses**	**18**

3.9.3 Other Research

Fortran, Matlab, and C are the preferred programming languages, respectively, according to Other research respondents. However, there is interest in all suggested languages. When "other" was selected, NCL, Perl, Mathematica, Excel, and Java were cited.

Response	Chart	Frequency	Count
C		34.3%	12
Fortran		**51.4%**	**18**
Matlab		48.6%	17
IDL		22.9%	8
Python		25.7%	9
Other. Please state:		20.0%	7
		Valid Responses	**35**
(Respondents were allowed to choose multiple responses)		**Total Responses**	**35**

3.10 Would you be interested in participating in a code-sharing community?

Nearly three-quarters of respondents would be interested in participating in a code-sharing community in which geophysical processing codes were contributed to an online repository along the lines of SourceForge.

Response	Chart	Frequency	Count
Yes		**74.0%**	**188**
No		26.0%	66
Not Answered			6
		Valid Responses	**254**
(Respondents could only choose a single response)		**Total Responses**	**260**

3.10.1 Climate Research

The majority of climate research respondents are in favor of participating in a code-sharing community.

Response	Chart	Frequency	Count
Yes		**78.2%**	**129**
No		21.8%	36
Not Answered			4
		Valid Responses	**165**
(Respondents could only choose a single response)		**Total Responses**	**169**

3.10.2 Education

Two-thirds of education respondents would be interested in participating in a code-sharing community.

Response	Chart	Frequency	Count
Yes		**66.7%**	**12**
No		33.3%	6
	Valid Responses		**18**
(Respondents could only choose a single response)	**Total Responses**		**18**

3.10.3 Other Research

Nearly three-fourths of other research respondents would be interested in participating in a code-sharing community.

Response	Chart	Frequency	Count
Yes		**70.6%**	**24**
No		29.4%	10
Not Answered			1
	Valid Responses		**34**
(Respondents could only choose a single response)	**Total Responses**		**35**

3.11 On an ARM-provided computing center, how much time per year would you spend performing intensive data runs?

Seventy percent of respondents would spend less than one week but up to one month performing intensive data runs.

Response	Chart	Frequency	Count
less than one week		**34.8%**	**87**
one week to one month		**34.8%**	**87**
one to three months		18.0%	45
three to six months		7.6%	19
more than six months		4.8%	12
Not Answered			5
	Valid Responses		**250**
(Respondents could only choose a single response)	**Total Responses**		**255**

3.11.1 Climate Research

The majority of climate researchers plan on spending up to three months performing data intensive runs with approximately 36 percent of respondents spending from one week to one month.

Response	Chart	Frequency	Count
less than one week		28.8%	47
one week to one month		**35.6%**	**58**
one to three months		21.5%	35
three to six months		9.8%	16
more than six months		4.3%	7
Not Answered			4
	Valid Responses		**163**
(Respondents could only choose a single response)	**Total Responses**		**167**

3.11.2 Education

Over half of education respondents are planning on spending less than one week performing intensive data runs, with approximately 84 percent spending up to one month.

Response	Chart	Frequency	Count
less than one week		**55.6%**	**10**
one week to one month		27.8%	5
one to three months		11.1%	2
three to six months		0.0%	0
more than six months		5.6%	1
	Valid Responses		**18**
(Respondents could only choose a single response)	**Total Responses**		**18**

3.11.3 Other Research

Forty-three percent of other research respondents are planning on spending less than one week performing intensive data runs, with 80 percent spending up to one month.

Response	Chart	Frequency	Count
less than one week		**42.9%**	**15**
one week to one month		37.1%	13
one to three months		14.3%	5
three to six months		2.9%	1
more than six months		2.9%	1
	Valid Responses		**35**
(Respondents could only choose a single response)	**Total Responses**		**35**

3.12 What is your preferred data analysis system?

For your data analysis work, the working assumption is that a Unix or Linux system will meet your needs. If this assumption is not true, please indicate your preferred system below.

Windows is the preferred data analysis system, receiving greater than three-quarters of the responses. For a full list of "Other" responses, see Appendix A.

Response	Chart	Frequency	Count
Windows		**77.1%**	**111**
MacOS		15.3%	22
Other. Please state:		7.6%	11
Not Answered			111
	Valid Responses		**144**
(Respondents could only choose a single response)	**Total Responses**		**255**

3.12.1 Climate Research

When Linux or Unix will not meet the climate research respondents' needs, Windows is the preferred data analysis system, receiving greater than three-quarters of the responses.

Response	Chart	Frequency	Count
Windows		**77.4%**	**65**
MacOS		14.3%	12
Other. Please state:		8.3%	7
Not Answered			83
	Valid Responses		**84**
(Respondents could only choose a single response)	**Total Responses**		**167**

3.12.2 Education

Windows is the preferred system by all education respondents when Unix or Linux system will not meet their needs for data analysis work.

Response	Chart	Frequency	Count
Windows		**93.8%**	**15**
MacOS		0.0%	0
Other. Please state:		6.3%	1
Not Answered			2
	Valid Responses		**16**
(Respondents could only choose a single response)	**Total Responses**		**18**

3.12.3 Other Research

Windows is the preferred system by all other research respondents when Unix or Linux system will not meet their needs for data analysis work.

Response	Chart	Frequency	Count
Windows		**68.2%**	**15**
MacOS		31.8%	7
Other. Please state:		0.0%	0
Not Answered			13
	Valid Responses		**22**
(Respondents could only choose a single response)	**Total Responses**		**35**

3.13 Please rate the level of importance of the following capabilities.

Respondents were asked to rate their level of importance on a scale from 5 to 1 where 5=highly important and 1=highly unimportant.

All respondents stated that computational resources/power, software development tools, data manipulation and visualization tools, and easy and rapid access to data are highly important to them. In fact, easy and rapid access to data received nearly 70 percent of the responses. The level of importance of software development tools is equal for neutral, important, and highly important with 29 percent in each level. Only 4 percent of respondents feel the capabilities were highly unimportant.

		5	4	3	2	1	Total
Computational resources/power	Count	**95**	73	50	19	13	**250**
	% by Row	**38.0%**	29.2%	20.0%	7.6%	5.2%	**100.0%**
Software development tools	Count	**73**	**73**	**73**	17	13	**249**
	% by Row	**29.3%**	**29.3%**	**29.3%**	6.8%	5.2%	**100.0%**
Easy and rapid access to data	Count	**175**	60	9	1	7	**252**
	% by Row	**69.4%**	23.8%	3.6%	0.4%	2.8%	**100.0%**
Data manipulation and visualization tools	Count	**119**	86	31	8	7	**251**
	% by Row	**47.4%**	34.3%	12.4%	3.2%	2.8%	**100.0%**
Total	**Count**	**462**	**292**	**163**	**45**	**40**	**1002**
	% by Row	**46.1%**	**29.1%**	**16.3%**	**4.5%**	**4.0%**	**100.0%**

3.13.1 Climate Research

Climate research respondents feel all the listed capabilities were highly important.

(Level of Importance (5=high, 1=low))		5	4	3	2	1	Total
Computational resources/power	Count	**64**	45	33	15	7	**164**
	% by Row	**39.0%**	27.4%	20.1%	9.1%	4.3%	**100.0%**
Software development tools	Count	**52**	46	45	11	9	**163**
	% by Row	**31.9%**	28.2%	27.6%	6.7%	5.5%	**100.0%**
Easy and rapid access to data	Count	**117**	35	6	1	6	**165**
	% by Row	**70.9%**	21.2%	3.6%	0.6%	3.6%	**100.0%**
Data manipulation and visualization tools	Count	**77**	53	24	5	5	**164**
	% by Row	**47.0%**	32.3%	14.6%	3.0%	3.0%	**100.0%**
Total	**Count**	**310**	**179**	**108**	**32**	**27**	**656**
	% by Row	**47.3%**	**27.3%**	**16.5%**	**4.9%**	**4.1%**	**100.0%**

3.13.2 Education

Education respondents feel all the listed capabilities are important with easy and rapid access to data, data manipulation and visualization tools, and computational resources/power being highly important.

		5	4	3	2	1	Total
Computational resources/power	Count	**6**	5	4	2	1	**18**
	% by Row	**33.3%**	27.8%	22.2%	11.1%	5.6%	**100.0%**
Software development tools	Count	4	**9**	2	2	1	**18**
	% by Row	22.2%	**50.0%**	11.1%	11.1%	5.6%	**100.0%**
Easy and rapid access to data	Count	**10**	8	0	0	0	**18**
	% by Row	**55.6%**	44.4%	0.0%	0.0%	0.0%	**100.0%**
Data manipulation and visualization tools	Count	**9**	8	1	0	0	**18**
	% by Row	**50.0%**	44.4%	5.6%	0.0%	0.0%	**100.0%**
Total	**Count**	**29**	**30**	**7**	**4**	**2**	**72**
	% by Row	**40.3%**	**41.7%**	**9.7%**	**5.6%**	**2.8%**	**100.0%**

3.13.3 Other Research

Other research respondents feel all capabilities are important with easy and rapid access to data, data manipulation and visualization tools, and computational resources/power (respectively) as having the highest levels of importance. Software development tools are the least important to other research respondents.

		5	4	3	2	1	Total
Computational resources/power	Count	9	**15**	8	0	3	**35**
	% by Row	25.7%	**42.9%**	22.9%	0.0%	8.6%	**100.0%**
Software development tools	Count	8	10	**13**	2	2	**35**
	% by Row	22.9%	28.6%	**37.1%**	5.7%	5.7%	**100.0%**
Easy and rapid access to data	Count	**25**	8	2	0	0	**35**
	% by Row	**71.4%**	22.9%	5.7%	0.0%	0.0%	**100.0%**
Data manipulation and visualization tools	Count	**18**	11	2	3	1	**35**
	% by Row	**51.4%**	31.4%	5.7%	8.6%	2.9%	**100.0%**
Total	**Count**	**60**	**44**	**25**	**5**	**6**	**140**
	% by Row	**42.9%**	**31.4%**	**17.9%**	**3.6%**	**4.3%**	**100.0%**

4.0 Appendix A

4.1 What is your work related to? ("Other" Comments)

Response
sun radiation research for photovoltaic power
Infrastructure
DMF
solar power
ocean research
Hydrology and Water Resources
ARM Infrastructure
Staff at AMF site
infrastructure-mm
cloud and radiation
Data comparison to Simulation made by Surface Temperature Equilibrium Energy Balance Model
Design of Mooring Systems
solar engineering software http://www.drbaumresearch.com/
Pipeline Safety
Climate Implications on Solar Power & Vice Versa
ACRF Infrastructure
Weather Forecasting
mitigation wetland design and monitoring
sensor calibration
Renewable Energy Resources
climate & weather research and education
operations checks
data quality
publishing climate graphics
remote sensing cloud
Radiation Models
data management
Financial Administrative
ARM Archive
OSS Event and Component Inventory
RS validation
Geodesy
Evapotranspiration
Solar energy
Radiation Measurement
Aviation
aerosol growth processes
solar development
Insurance
remote sensing
atmospheric ozone
ARM Infrastructure

4.2 Initially, what types of tools in an ARM computing center would be most beneficial to you? ("Other" Comments)

If "other" was chosen above, please specify the type of tool you are interested in using in an ARM computing center.

Response
direct sunlight and diffusive sunlight
I am primarily interested in quick download configured in a format that is easily understood by scientists, e.g., text, database, or spreadsheet formats
A tool where you can chose what time step you wish the data to be downloaded in. For example, I am in no need of radiation data in 1 minute intervals but would appreciate if there was a tool that could convert the 1 min intervals into hourly averages. That would reduce the size of the download as well.
data size can be selectable
NCO
We already have excess computing capacity.
Primarily concerned w/skyrad60 data from multiple latitudes within 30 degrees lon.
Estimate of data uncertainties available directly with all observational data.
Access to the archive using OPeNDAP or similar technologies that are currently built into many of the leading analysis and visualization packages (e.g,. Ferret, NCL)
Our needs are for 15 minute, 1 hour and daily averages for simulation boundary limits and to compare to our simulations.
The interpretation of data in a form of any simple solar radiation model like Bird Clear Sky Model: Bird, R. E., and R. L. Hulstrom, "Simplified Clear Sky Model for Direct and Diffuse Insulation on Horizontal Surfaces", Technical Report No. SERI/TR-642-761, Golden, CO: Solar Energy Research Institute, 1981 http://www.nrel.gov/rredc/pdfs/761.pdf
statistical analysis, parameter estimation
Aggregating climate parameters out of the high frequency measurements taken at ARM facilities
Climatic data
NCL graphing
Assuming ncl is getting popular with plotting netcdf, grib files, and with all the resources built up, building interfaces to read and analyze acrf data in ncl would benefit all modelers!
Running a GCM or CRM.
Statistical analyses.
I have been studying statistical analysis for climate and atmospherical data such as spectral analysis... and GIS application.
It could be important to be able to easily download the results of statistical analyses of larger data sets---that is, print out the statistical properties desired in a small data file.
search data tool
Radiation model Software
data format
DQ HandsPlotbrowser, DS View, DQPR, EWO, ECO, DQR
soil properties, LAI, surface temperature, soil moisture, surface fluxes, etc.
data analysis
dispersion model
1) working on the archive data remotely with my own code 2) reliable notifications on reprocessed/new data of the kind that a user downloaded before 3) automated way of releasing data to archive, with automated generation of quick look plots, statistics and difference with prior data release
batch processing of large radiance databases using new physics-based retrievals algorithms uploaded to ACRF computers (after testing on representative test data downloaded to home-institution computers)

4.3 What programming languages do you prefer? ("Other" Comments)

Mathematica, Java, NCL (3 votes), C++, Cobol, Perl, and a spreadsheet are all listed in the "other" comments.

4.4 What is your preferred data analysis system? ("Other" Comments)

Other data analysis system comments include Linux (four responses) and MS DOS.

5.0 Appendix B

5.1 Initially, what types of tools in an ARM computing center would be most beneficial to you? (Climate Research Responses)

Respondents were asked to rate their overall level of interest on a scale from 5 to 1 where 5=highly interested and 1=highly uninterested.

(Level of Interest (5=high, 1=low))		5	4	3	2	1	Total
Data input/output	Count	**87**	47	13	2	8	**157**
	% by Row	**55.4%**	29.9%	8.3%	1.3%	5.1%	**100.0%**
Geophysical functions (like humidity conversion, solar zenith angle calculations)	Count	**72**	40	28	8	14	**162**
	% by Row	**44.4%**	24.7%	17.3%	4.9%	8.6%	**100.0%**
Radiative transfer code	Count	**65**	39	29	14	13	**160**
	% by Row	**40.6%**	24.4%	18.1%	8.8%	8.1%	**100.0%**
Single column model	Count	**49**	41	29	20	18	**157**
	% by Row	**31.2%**	26.1%	18.5%	12.7%	11.5%	**100.0%**
Model analysis tools (e.g., instrument simulator)	Count	**60**	41	23	17	15	**156**
	% by Row	**38.5%**	26.3%	14.7%	10.9%	9.6%	**100.0%**
Merging multiple parameters to a common time/height/space grid	Count	**92**	40	20	4	7	**163**
	% by Row	**56.4%**	24.5%	12.3%	2.5%	4.3%	**100.0%**
Plot/visualization	Count	**74**	32	33	11	12	**162**
	% by Row	**45.7%**	19.8%	20.4%	6.8%	7.4%	**100.0%**
Data quality	Count	**82**	35	28	6	7	**158**
	% by Row	**51.9%**	22.2%	17.7%	3.8%	4.4%	**100.0%**
Data comparison (including model/data comparisons)	Count	**73**	51	21	8	7	**160**
	% by Row	**45.6%**	31.9%	13.1%	5.0%	4.4%	**100.0%**
Other. Please state below.	Count	19	5	6	0	**20**	**50**
	% by Row	38.0%	10.0%	12.0%	0.0%	**40.0%**	**100.0%**
Total	**Count**	**673**	**371**	**230**	**90**	**121**	**1485**
	% by Row	**45.3%**	**25.0%**	**15.5%**	**6.1%**	**8.1%**	**100.0%**

Other suggested tools include suggestions for variable data formats ("a format that is easily understood by scientists; e.g., text, database, or spreadsheet formats"), a tool that would allow time intervals to be specified, and the ability to perform analyses of the data. One respondent also suggested: "1) working on the archive data remotely with my own code, 2) reliable notifications on reprocessed/new data of the kind that a user downloaded before, [and] 3) automated way of releasing data to archive, with automated generation of quick look plots, statistics and difference with prior data release."

Response
I am primarily interested in quick download configured in a format that is easily understood by scientists, eg. text, database, or spreadsheet formats
A tool where you can chose what time step you wish the data to be downloaded in. For example, I am in no need of radiation data in 1 minute intervals but would appreciate if there was a tool that could convert the 1 min intervals into hourly averages. That would reduce the size of the download as well.
data size can be selectable
NCO
We already have excess computing capacity.
Estimate of data uncertainties available directly with all observational data.
Access to the archive using OPeNDAP or similar technologies that are currently built into many of the leading analysis and visualization packages (e.g., Ferret, NCL)
statistical analysis, parameter estimation
Aggregating climate parameters out of the high frequency measurements taken at ARM facilities
Climatic data
NCL graphing
Running a GCM or CRM.
Statistical analyses.
I have been studying statistical analysis for climate and atmospherical data such as spectral analysis .. and GIS application.
It could be important to be able to easily download the results of statistical analyses of larger data sets---that is, print out the statistical properties desired in a small data file.
data format
soil properties, LAI, surface temperature, soil moisture, surface fluxes, etc.
dispersion model
1) working on the archive data remotely with my own code 2) reliable notifications on reprocessed/new data of the kind that a user downloaded before 3) automated way of releasing data to archive, with automated generation of quick look plots, statistics and difference with prior data release

5.2 Initially, what types of tools in an ARM computing center would be most beneficial to you? (Education Responses)

Respondents were asked to rate their overall level of interest on a scale from 5 to 1 where 5=highly interested and 1=highly uninterested.

		5	4	3	2	1	Total
Data input/output	Count	**13**	2	1	0	2	**18**
	% by Row	**72.2%**	11.1%	5.6%	0.0%	11.1%	**100.0%**
Geophysical functions (like humidity conversion, solar zenith angle calculations)	Count	**8**	3	2	5	0	**18**
	% by Row	**44.4%**	16.7%	11.1%	27.8%	0.0%	**100.0%**
Radiative transfer code	Count	**7**	3	5	1	2	**18**
	% by Row	**38.9%**	16.7%	27.8%	5.6%	11.1%	**100.0%**
Single column model	Count	1	**7**	5	2	3	**18**
	% by Row	5.6%	**38.9%**	27.8%	11.1%	16.7%	**100.0%**
Model analysis tools (e.g., instrument simulator)	Count	**7**	4	2	3	2	**18**
	% by Row	**38.9%**	22.2%	11.1%	16.7%	11.1%	**100.0%**
Merging multiple parameters to a common time/height/space grid	Count	6	**7**	3	2	0	**18**
	% by Row	33.3%	**38.9%**	16.7%	11.1%	0.0%	**100.0%**
Plot/visualization	Count	**9**	5	1	1	1	**17**
	% by Row	**52.9%**	29.4%	5.9%	5.9%	5.9%	**100.0%**
Data quality	Count	**8**	6	1	2	0	**17**
	% by Row	**47.1%**	35.3%	5.9%	11.8%	0.0%	**100.0%**
Data comparison (including model/data comparisons)	Count	5	**7**	2	2	1	**17**
	% by Row	29.4%	**41.2%**	11.8%	11.8%	5.9%	**100.0%**
Other. Please state below.	Count	0	0	1	1	**4**	**6**
	% by Row	0.0%	0.0%	16.7%	16.7%	**66.7%**	**100.0%**
Total	**Count**	**64**	**44**	**23**	**19**	**15**	**165**
	% by Row	**38.8%**	**26.7%**	**13.9%**	**11.5%**	**9.1%**	**100.0%**

When “other” was suggested, one education respondent is “primarily concerned with skyrad60 data from multiple latitudes within 30 degrees longitude."

5.3 Initially, what types of tools in an ARM computing center would be most beneficial to you? (Other Research Responses)

Respondents were asked to rate their overall level of interest on a scale from 5 to 1 where 5=highly interested and 1=highly uninterested.

		5	4	3	2	1	Total
Data input/output	Count	**18**	7	5	3	2	**35**
	% by Row	**51.4%**	20.0%	14.3%	8.6%	5.7%	**100.0%**
Geophysical functions (like humidity conversion, solar zenith angle calculations)	Count	**15**	5	7	5	2	**34**
	% by Row	**44.1%**	14.7%	20.6%	14.7%	5.9%	**100.0%**
Radiative transfer code	Count	**14**	5	10	3	3	**35**
	% by Row	**40.0%**	14.3%	28.6%	8.6%	8.6%	**100.0%**
Single column model	Count	5	9	**12**	5	4	**35**
	% by Row	14.3%	25.7%	**34.3%**	14.3%	11.4%	**100.0%**
Model analysis tools (e.g., instrument simulator)	Count	**11**	7	6	7	3	**34**
	% by Row	**32.4%**	20.6%	17.6%	20.6%	8.8%	**100.0%**
Merging multiple parameters to a common time/height/space grid	Count	**16**	10	3	3	3	**35**
	% by Row	**45.7%**	28.6%	8.6%	8.6%	8.6%	**100.0%**
Plot/visualization	Count	**15**	13	3	3	1	**35**
	% by Row	**42.9%**	37.1%	8.6%	8.6%	2.9%	**100.0%**
Data quality	Count	**16**	8	7	3	1	**35**
	% by Row	**45.7%**	22.9%	20.0%	8.6%	2.9%	**100.0%**
Data comparison (including model/data comparisons)	Count	**12**	10	8	3	2	**35**
	% by Row	**34.3%**	28.6%	22.9%	8.6%	5.7%	**100.0%**
Other. Please state below.	Count	**4**	0	3	3	**4**	**14**
	% by Row	**28.6%**	0.0%	21.4%	21.4%	**28.6%**	**100.0%**
Total	**Count**	**126**	**74**	**64**	**38**	**25**	**327**
	% by Row	**38.5%**	**22.6%**	**19.6%**	**11.6%**	**7.6%**	**100.0%**

When other was suggested, one other research respondent is “primarily concerned with skyrad60 data from multiple latitudes within 30 degrees longitude."

www.arm.gov

U.S. DEPARTMENT OF
ENERGY

Office of Science