Efficient Access to Climate Products using ACIS Web Services

BY ARTHUR T. DEGAETANO, WILLIAM NOON, AND KEITH L. EGGLESTON

Despite the proliferation of in situ, modeled, and remotely sensed climate data over the last decade, users of these datasets outside of the research community cite numerous hurdles to the effective use of these data in decision making (e.g., Overpeck et al. 2011; Beniston, et al. 2012). These reviews describe many climate datasets as large, covering geographic domains of little interest to users having a need for only a small subset of these data. The studies mention a range of data formats that complicate the use of data from different networks. Gridded datasets, as well, are stored in a variety of data formats and projections. Based on decades of experience, the Regional Climate Center (RCC) Program has found that end users are often frustrated by the onerous tasks of downloading and parsing large volumes of irrelevant data, reformatting files, and recreating commonly used analyses. These steps introduce potential errors to the data stream and are inefficient uses of human and financial resources. The World Bank recently labeled data that are "hard to use" as a barrier, stating, "Developers tend to use data they can easily access and understand—which may not be the most current, comprehensive or robust data available" (http://data.worldbank.org/news/climate-data-meeting-discussion).

In many real-time applications (e.g., the monitoring of extremes), climate datasets are dynamic in that they are constantly evolving as either new data are added or earlier observations are updated to replace missing values or reflect more rigorous quality control. This can lead to suboptimal decisions based on data that, although complete and of the highest quality when downloaded, no longer reflect the best-available data source. This is especially true in static graphical and tabular data products that, once produced, are not refreshed to reflect new or amended input data. For instance, the number of degree days accumulated during February 2013 in Philadelphia, Pennsylvania, was queried on websites operated by private, state, and federal entities as well as a site using the Applied Climate Information System (ACIS). Of the four values (740, 828, 830, and 830), only those from ACIS and the National Weather Service (NWS) agreed, as the other sources omitted daily observations that were initially missing or used nonstandard conventions for summing fractional degree-day values. Although any original data values failing quality control are retained in the ACIS database, they are marked as invalid and by default excluded from data products.

In a 2004 *BAMS* article, Hubbard et al. describe ACIS, an Internet-based system designed to facilitate the generation and dissemination of climate data products to users. Much more than a standard climate database, this original version of ACIS linked historical and recent in situ daily climate data from a variety of networks to standardized analyses that parsed and/or summarized these data to meet specific user requests for climate information. In its initial version, user access to ACIS was possible via three types of interfaces: 1) high-level web-based, 2) mid-level EXtensible Markup Language-Remote Procedure Call (XML-RPC), and 3) low-level Common Object Request Broker Architecture (CORBA). Current users are perhaps most familiar with the ACIS web interfaces that provide free access to the general public via NOAA Online Weather Data (NOWData) (e.g., www.nws.noaa.gov/climate/xmacis.php?wfo=box) and climate product access sites maintained by the Regional Climate Centers (e.g., http://scacis.rcc-acis.org) for specific users such as the NWS.

The potential of this version of ACIS was never fully realized, as the mid- and low-level interfaces

AFFILIATIONS: DEGAETANO, NOON, AND EGGLESTON—Northeast Regional Climate Center, Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York
CORRESPONDING AUTHOR: Dr. Art DeGaetano, 1119 Bradfield Hall, Cornell University, Ithaca, NY 14853
E-mail: atd2@cornell.edu

that were intended to allow users to develop customized decision tools and climate products based on the ACIS database and suite of analysis tools proved to be too cumbersome and inefficient to all but the most sophisticated users. Heightened interest in and use of climate data and information in recent years by a variety of federal, state, and private decision makers has necessitated the development of a replacement for these ACIS interfaces.

ACIS VERSION 2. With ACIS Version 2 (http://rcc-acis.org), users can access ACIS data products free of charge using web services (computer-to-computer communication over a network). This new interface is intended to significantly reduce the hurdles previously associated with integrating ACIS into user-specific decision tools, customized websites, and client-developed computer code. Through the web services interface, ACIS can now serve as a core resource for developing climate services at national, regional, and state levels. ACIS merges data and its corresponding metadata from a variety of different networks and sources, each with potentially different attributes such as format, temporal resolution, and for gridded data, projection (Fig. 1). Currently the datasets included in ACIS reflect primarily once-a-day observations of maximum, minimum, and observation time air temperature, precipitation, snowfall, and snow depth. For some sites, variables such as daily wind run and pan evaporation are also available. In addition, direct access to daily and monthly climate normals and derived variables such as degree days is available. Access to variables observed at subdaily time intervals is expected later in 2015.

Users can select from a set of five web service calls. *StnData, MultiStnData,* and *GridData* provide access to ACIS data holdings. *StnMeta* returns metadata for a station or stations, while the *General* call provides metadata related to the geographic area descriptors supported by ACIS. Complete descriptions of these calls can be found in the online documentation at http://data.rcc-acis.org. In addition, a tool available at http://builder.rcc-acis.org/ provides a web-based means for experimenting with and testing the capabilities of ACIS web services.

For each call, users specify a set of parameters. These parameters describe the data being requested (e.g., station, date range, and variable); the type of summarization, which is called a reduction in ACIS syntax (e.g., average); the temporal or spatial reduction interval (e.g., monthly or county); and the format of
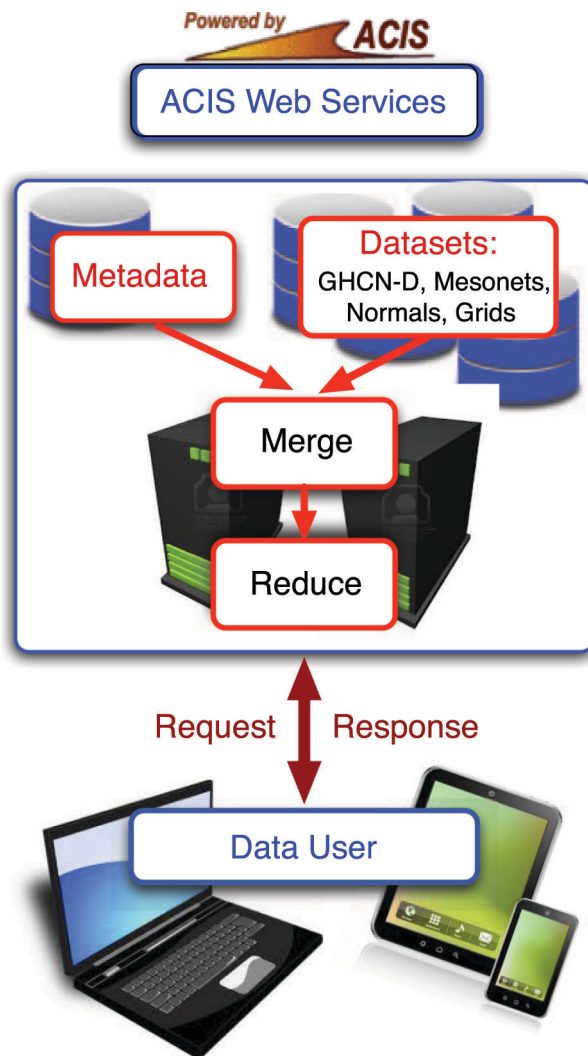
Fig. 1. Schematic of ACIS web services.

the output. ACIS web services pass these parameters to a server that requests the necessary data from the ACIS database. These data are returned to the server, summarized, and formatted. A climate data product is then returned via web services to the user (Fig. 1).

The reduction of requested data by ACIS differs from most traditional climate data management systems, which typically return subsets of data files to users who must then subsequently develop code to transform these data into the required climate information. Although ACIS could support data delivery using methods that are more familiar to the meteorological community [such as Open-source Project for a Network Data Access Protocol (OPeNDAP) or Thematic Real-time Environmental Distributed Data Services (THREDDS)], its primary users have

guided development toward other access methods and conventions that are better suited to the needs and competencies of users and decision-makers outside of the traditional scientific research community. Nonetheless, the datasets that form the foundation of the climate information products available from ACIS [e.g., the Global Historical Climatology Network—Daily (GHCN-Daily)] are compliant with existing data standards.

Alternatively, conventional systems may store key commonly requested climate products as static data files. In ACIS, such analyses are conducted "on the fly," assuring the requested products represent the most recent and highest quality data available. These real-time analyses have been optimized for responsiveness with the latency of most simple requests under 10 ms and large multiple station requests limited by the speed of the connection. Access to the data reported by all U.S. stations during a given month (approximately 40MB of data) averages 90 s.

Given the primary role of ACIS as a platform for serving the day-to-day needs of climate-dependent decision-makers, the provision of the most recent, highest quality dataset versions is paramount. As a result (and to facilitate use outside of the traditional research community), extensive efforts to archive multiple dataset versions or implement capabilities to regenerate earlier dataset states have been avoided. Arguably, this has been at the expense of research applications requiring strict reproducibility. This does not preclude the use of ACIS as a source of data for research applications, as many federal data access systems such as the National Climatic Data Center's Climate Data Online also limit access to the most recent dataset version. In terms of ACIS software, the

distributed revision control repository Github has served as a code repository during the ACIS development process. Provided researchers are aware of these caveats, ACIS also offers efficient data access and manipulation capabilities to the research communities within and outside the atmospheric sciences.

THE STNDATA CALL. Years of experience in answering user data requests have led the Regional Climate Center Design Team to identify a base set of 11 reductions that essentially form the foundation of nearly all climate information requests (Table 1). These reductions, when combined with parameters specifying the interval (time step of the results) and duration (period over which the reduction occurs), allow a wide range of climate analyses. For example, a user interested in the lowest temperature recorded during each month would specify the interval as monthly (*mly*), the reduction as minimum (*min*), and the variable as minimum temperature (*mint*). When the request is made from the URL line, the term element (*elems*) is used to describe this combination of three parameters (*elems = mly_min_mint*) in Fig. 2. Also specified in the URL is the address of the ACIS server, the type of web services call (*StnData*), and parameters representing the desired station (*sid*), starting date (*sdate*), and ending date (*edate*). The duration parameter is not required in this simple request.

The climate information is returned in JavaScript Object Notation (JSON), although a limited number of products support an option that provides comma-separated value (CSV) output. JSON is a data-interchange standard (www.json.org/) that is not only easy for humans to read and write, but is also easy for computers to parse and generate. Although JSON is completely language-independent, it uses structures that are similar to those used in many coding languages, including C, C++, Java, JavaScript, Perl, and Python.

In Fig. 2, the JSON results object reports the metadata for station 304174, including its longitude and latitude (*ll*), elevation (*elev*), station ids (*sids*), state, and common name. As the Ithaca station reports data as a part of three networks, its *sids* are given for each network. The key *"uid"* is an internal unique ACIS identifier. The JSON object also returns the requested data as an array of arrays, with each monthly array containing two elements: the month and the lowest daily minimum temperature reported during that month. English units (°F for temperature-based and inches for precipitation-based variables) are used in all ACIS data products. This decision was again driven


```
http://data.rcc-acis.org/StnData?sid=304174&
sdate=2012-01&edate=2012-12&elems=mly_min_mint
```

```
{"meta":{"uid": 18667, "ll": [-76.44917, 42.44917],
"sids": ["304174 2", "USC00304174 6", "ITHN6 7"],
"state": "NY", "elev": 960.0, "name": "ITHACA
CORNELL UNIV"},
"data":[["2012-01","-5"],
["2012-02","13"],
["2012-03","11"],
["2012-04","20"],
["2012-05","34"],
["2012-06","43"],
["2012-07","49"],
["2012-08","44"],
["2012-09","38"],
["2012-10","23"],
["2012-11","20"],
["2012-12","15"]]}
```

FIG. 2. Example of web services request from the URL line and corresponding output.

GRIDDATA

Since its inception, the primary focus of ACIS has been on station data. However users of climate information are increasingly requesting gridded ACIS-like climate data products. The ACIS Version 2 web services also include support for gridded data.

The *GridData* call provides the same temporal reduction capabilities as *StnData* and *MultiStnData* (Table 1). However, the GridData call also provides spatial reduction capabilities. For instance, the values for all grids within some user-defined area (e.g., a state, county, or climate division) can be averaged to yield an areal average, or these values can be queried to return an extreme value (areal maximum or minimum) for the region (Fig. SB1). Any combination of spatial and areal reductions is possible, allowing spatial statistics such as the areal percentage of a county with total monthly rainfall ≤ 0.50 in. to be generated.

Currently, *GridData* provides users with access to several gridded datasets, including daily observed temperature and precipitation fields generated using the methods of DeGaetano and Belcher (2007) for temperature and DeGaetano and Wilks (2009) for precipitation. Natural-neighbor–interpolated daily temperature and precipitation data fields are also available, allowing map-based products analogous to the current suite of ACIS map products (www.hprcc.unl.edu/maps/current/) to be generated. Recently, daily and monthly gridded datasets produced using the Parameter-elevation Regressions on Independent Slopes Model (PRISM) have been added to ACIS. Later in 2015, access to a monthly gridded dataset based on a methodology developed by the National Climatic Data Center (NCDC) is anticipated. The *GridData* call includes an output option that returns images in png format (Fig. SB1).

GridData also offers access to the dynamically downscaled suite of North American Regional Climate Change Assessment Program (NARCCAP) global climate model projections. In addition to providing the same temporal and spatial reduction capabilities, *GridData* facilitates the projection of these gridded datasets to a common Lambert conformal conic projection. In future modifications of ACIS, we anticipate the provision of an option allowing user-specified output projections.

TABLE 1. ACIS web services reduction codes.

Reduction Code	Description
max	Maximum value
min	Minimum value
sum	Sum of values
mean	Average of values
Stdev	Standard Deviation of values (GridData only^)
cnt_xx_yyy	Count of values passing a threshold *
pct_xx_yyy	Percent of values passing a threshold *
fct_xx_yyy	Fraction of values passing a threshold *
first_xx_yyy	First occurrence of passing a threshold *
last_xx_yyy	Last occurrence of passing a threshold *
run_xx_yyy	Consecutive values passing a threshold *

^ Standard Deviation was a "special" request for the National Climate Assessment that is still implemented only as part of GridData.

* xx is an operator (e.g., ge, gt, le, lt, eq, ne) and yyy is a numeric threshold. For example, specifying a reduction of cnt_ge_0.01 with precipitation as the variable would return the number of days receiving greater than or equal to 0.01 inches of precipitation during the specified range of dates.

JSON parameters object requesting monthly (August, 2011) county-average precipitation for all counties in Nebraska using the NRCC high resolution precipitation grid

```
{"state":"ne","grid":"3","elems":
[{"name":"pcpn","interval":"mly","duration":
"mly","reduce":"mean","area_reduce":"county_
mean"}],"date":"2011-08"}
```

JSON results object (abridged) from above request showing county FIPS id (e.g. " 31119") and corresponding county-average rainfall (e.g. 0.112)

```
{"data":[["2011-08",{"31119":0.11208746582269669,
"31113":0.044888101518154144,"31111"
0.08875981718301773,
...,"31039":0.11941716820001602}]]}
```

JSON parameters object requesting monthly (August, 2011) average precipitation for grids encompassing Nebraska specifying image as the output type

```
{"state":"ne","grid":"3","output":"image","elems":
[{"name":"pcpn","interval":"mly","duration":"mly","
reduce":"mean"}],"image":{"proj":"lcc",
"overlays":"state","interp":"cspline","cmap":"Blues
","width":350},"date":"2011-08"}
```

Image (png) returned by the above request with JSON object specifying colors (cmap), range of data values, contour levels and image size. Each of these parameters can be set in the parameters object

```
{"cmap":
["#f7fbff","#c6dbef","#6baed6","#2171b5","#08306b"],
"range":[0.004448307678103447,0.3454032242298126],
"levels":[0.10000000149011612,0.20000000298023224,
0.30000001192092896,0.4000000059604645],"size":
[350,161]}
```

FIG. SB1. Example of GridData request to obtain (top) numeric county-average precipitation and (bottom) an image of monthly precipitation. These samples were generated from http://builder.rcc-acis.org/. **The map is obtained by entering "image" as the output and checking the Generate map box. The JSON object is generated by a separate call in which the Generate map box is checked, but the output field is JSON. The data field has been omitted from the object given in the figure for brevity.**

by the overwhelming consensus of user needs. In keeping with conventional standards, data are stored in their native units and these units recorded as a part of the metadata. Any necessary conversion to default English units occurs as a part of the data request. Future versions of ACIS web services will accommodate requests for user-specified units and thus also return unit information as a part of the metadata.

ACIS currently allows products to be generated based on data from 11 observation networks including the NWS Cooperative Observer Network, the Community Collaborative Rain, Hail and Snow (CoCoRaHS) Network, the USDA Natural Resources Conservation Service Snow Telemetry (SNOTEL) network, and ThreadEx, a specialized long-term dataset of concatenated data series from major metropolitan regions (http://climatedataguide.ucar.edu/guidance/noaa-threadex-long-term-station-extremes-america). Canadian and Mexican data from GHCN-Daily can also be accessed. Network membership can be requested, allowing information to be parsed by network if desired.

Users may also access ACIS web services through calls embedded in computer code. In these cases (and also more complex calls), the calls are initiated using a JSON parameters object. Figure 3 shows an example using Python of a request to obtain the date of the last occurrence of a temperature ≤28°F during each spring from 1991 to 2000. Following the import of the necessary Python modules, the input parameters are specified and assigned to the variable "params" that maps values to key names in a structure known as a Python dictionary. In this example, Minneapolis's 4-character International Civil Aviation Organization (ICAO) id is specified as the sid. The elements array sets the variable to minimum temperature ("name": "mint") and the duration to season-to-date ("duration": "std"). This duration requires an additional parameter that specifies the starting date of the "season," July 1 ("season_start": "07–01") in this case. In this more complex request, the interval is specified as an array of ones and zeros ("interval":[1,0,0]). Here, the number of values in the interval array (3) indicates that the returned values will have a temporal precision of days, and the 1 in the first position indicates that one value will be returned each year. Using this syntax, which is explained in more detail in the documentation (http://data.rcc-acis.org/), the array can have a length of 1, 2, or 3. A length of 1 indicates annual precision, 2 monthly precision, and 3 daily precision. The position of the value 1 within the array signifies

```
## Python standard library for making http requests
import urllib2

## Python standard library for parsing JSON formatted data
import json

## Parameters to get the last "spring" (July 1-June 30) temperature <= 28 degrees
## for the years 1991-2000.
## Also report what the min temperature was on that date:

params = {
    "sid":"kmsp",  "sdate":"1991-06-30",  "edate":"2000-06-30",
    "elems":[{"name":"mint",  "interval":[1,0,0],
    "duration":"std",  "season_start":"07-01",
    "reduce": {"reduce":"last_le_28",  "add":"value"}}
            ]
    }

## Send JSON encoded params to ACIS StnData method.
## Request return values as JSON object

request = urllib2.Request(
    'http://data.rcc-acis.org/StnData',
    json.dumps(params),
    {'Content-Type':'application/json','Accept':'application/json'})
response = urllib2.urlopen(request)
print 'Results as JSON'
print response.read()
```

Results as JSON

```
{"meta":{"uid": 10410, "ll": [-93.22889, 44.88306], "sids": ["14922 1", "215435 2",
"MSP 3", "72658 4", "KMSP 5", "USW00014922 6", "MRFM5 7", "MSP 7"], "state":
"MN", "elev": 872.0, "name": "MINNEAPOLIS/ST PAUL AP"},
"data":[["1991-06-30",["1991-04-20","28"]],
["1992-06-30",["1992-04-12","17"]],
["1993-06-30",["1993-04-21","27"]],
["1994-06-30",["1994-04-06","18"]],
["1995-06-30",["1995-04-05","18"]],
["1996-06-30",["1996-04-23","24"]],
["1997-06-30",["1997-05-01","28"]],
["1998-06-30",["1998-04-10","28"]],
["1999-06-30",["1999-03-26","23"]],
["2000-06-30",["2000-04-21","27"]]]}
```

FIG. 3. Example of a more complex web-services request from within a Python program and corresponding output.

the frequency of the values returned (first position, annually; middle position, monthly; last position, daily). Finally, the key of "reduce" specifies the type of reduction that is desired. Here, the desired reduction is the date of the last occurrence of 28°F or less ("reduce":"last_le_28"). Since the temperature that occurred on this date is also desired, the parameters object also includes "add": "value." In other requests, the "add" key can also be used to return two additional data values: 1) the number of missing data values in the analysis period or within a run of consecutive days, and 2) the date associated with a value, such as the lowest minimum temperature in a year.

The requested data are also returned as a JSON object (Fig. 3). In this case, the results object contains two keys. Station metadata follow the "meta" key, while the last frost dates and corresponding minimum temperatures follow the "data" key.

In addition, a tool available at http://builder.rcc-acis .org/ provides a web-based means of formulating JSON parameters objects to facilitate the incorpora-

tion of ACIS web services calls in user-developed software. Upon selecting the desired web services call from the bar at the top of the ACIS Query Builder, input boxes prompt users for both the required and optional fields necessary for each call. Clicking in an input box launches a table that outlines and describes the allowable entries. Clicking the "submit" button displays the JSON formatted parameters object and also the JSON-encoded results object containing the output. This is a useful means for both constructing a parameters object that can be inserted into code and determining if the web service call returns the desired information.

OTHER ACIS CALLS. The documentation also describes the other types of ACIS web services calls in more detail. Briefly, the *MultiStnData* call returns data for multiple stations within some user-specified geographical area. Currently the area can be specified as a state, county, climate division, National Weather Service County Warning Area, river basin (eight-digit hydrologic unit code), or latitude-longitude bounding box. The call also allows a list of station ids to be specified in lieu of an area. The same reduction options that are available in the *StnData* call can be used to summarize the data from these multiple stations.

The *StnMeta* call uses the same geographic areas as the *MultiStnData* call, but returns metadata describing the stations within the designated regions (e.g., a list of all available stations in a county). A useful feature of the *StnMeta* call is the ability to obtain the range of dates over which a particular station collected data. This allows users to determine, for example, which stations in a state (or other area) are currently active or which stations have the longest records.

The *General* call is a useful tool for obtaining information about the geographic areas supported by ACIS. For example, obtaining data for stations within a particular county in Kansas requires specification of that county's Federal Information Processing Standard (FIPS) code in the *MultiStnData* call. These codes can be obtained using the *General* call. In a browser, this call is invoked using http://data .rcc-acis.org/General/county?state=KS. As an option, the *General* call can also return a GeoJSON object that defines the collection of points that describe the boundary of a geographic region.

A new feature of ACIS Version 2 is the ability to access and summarize gridded data. This feature makes use of the *GridData* call that is highlighted in the sidebar.

SUMMARY. ACIS Version 2 facilitates the integration of ACIS climate data access and reduction capabilities into user-specific decision tools, customized dynamic websites (such as those highlighted at http://rcc-acis.org/examples.html), and client-developed computer code. Through effective use of ACIS, climate service providers are relieved of the burden and inefficiencies of managing and updating their own climate databases. Rather, the onus is placed upon the RCC Program, as a part of its stated mission is to facilitate the dissemination and use of climate information by decision makers from a wide range of climate-sensitive sectors. This assures that users (and providers) of climate data products receive the most accurate and up-to-date climate information possible, regardless of source. Likewise, standardized ACIS reduction protocols assure consistent handling of subtleties such as rounding, accumulations, missing data, etc.

The ACIS web services interface currently serves as a core resource for climate service activities within the Regional Climate Centers, National Weather Service, Natural Resources Conservation Service, and several private industry partners. It provides efficient, responsive, and standardized access to summarized climate data products that are consistent with data from the NCDC archive. This allows public and private climate service providers to focus on developing tools and disseminating information to meet customer needs.

ACKNOWLEDGMENTS. ACIS Version 2 software is a joint effort of the Regional Climate Center Program. We thank our fellow RCC directors for their guidance and encouragement, the team of developers at each RCC for their insights and technical expertise, and the ACIS design team for their input and knowledge of climatology and user services.

FOR FURTHER READING

Beniston, M., M. Stoffel, R. Harding, M. Kernan, R. Ludwig, E. Moors, P. Samuels, and K. Tockner, 2012: Obstacles to data access for research related to climate and water: Implications for science and EU policy-making. *Environ. Sci. Policy,* **17,** 41–48, doi:10.1016/j.envsci.2011.12.002.

Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain.

J. Appl. Meteor., **33,** 140–158, doi:10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2.

DeGaetano, A. T., and B. N. Belcher, 2007: Spatial interpolation of daily maximum and minimum air temperature based on meteorological model analyses and independent observations. *J. Appl. Meteor. Climatol.,* **46,** 1981–1992, doi:10.1175/2007JAMC1536.1.

——, and D. S. Wilks, 2009: Radar-guided interpolation of climatological precipitation data. *Int. J. Climatol.,* **29,** 185–196, doi:10.1002/joc.1714.

——, T. J. Brown, S. D. Hilberg, K. Redmond, K. Robbins, P. Robinson, M. Shulski, and M. McGuirk, 2010: Toward regional climate services: The role of NOAA's regional climate centers. *Bull. Amer. Meteor. Soc.,* **91,** 1633–1644, doi:10.1175/2010BAMS2936.1.

Hubbard, K. G., A. T. DeGaetano, and K. D. Robbins, 2004: SERVICES: A modern applied climate information system. *Bull. Amer. Meteor. Soc.,* **85,** 811–812, doi:10.1175/BAMS-85-6-811.

Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily database. *J. Atmos. Oceanic Technol.,* **29,** 897–910, doi:10.1175/JTECH-D-11-00103.1.

Overpeck, J. T., G. A. Meehl, S. Bony, and D. R. Easterling, 2011: Climate data challenges in the 21st century. *Science,* **331,** 700–702, doi:10.1126/science.1197869.

Vose, R. S., and Coauthors, 2014: Improved historical temperature and precipitation time series for U.S. climate divisions, *J. of Appl. Meteorol. and Climatol.,* **53,** 1232–1251, doi:10.1175/JAMC-D-13-0248.1.