Observing and Data System Design
to Support NOAA's Reference Environmental Data Records

Jeffrey L. Privette and Edward J. Kearns
Center for Weather and Climate
NOAA's National Centers for Environmental Information

Introduction

As part of its mission, NOAA provides objective data and tools to help the nation monitor, characterize, and predict changes in Earth's environment. In support of this goal, NOAA's National Centers for Environmental Information (NCEI) initiated a program in 2009 to develop and sustain satellite Climate Data Records (CDRs). The National Research Council (NRC, 2004) defines a CDR as "a time series of measurements of sufficient length, consistency, and continuity to determine climate variability and change." Most NCEI CDRs are derived from data collected by operational meteorological satellites maintained by NOAA and the Department of Defense. CDR production typically requires the reprocessing of multi-satellite time series of raw observations using baselined "climate quality" calibration and retrieval algorithms. It also includes the regular extension of records through processing of new observations forward in time. As described below, NCEI recently expanded the Program's mission to encompass Reference Environmental Data Records (Reference EDRs). Reference EDRs are also reprocessed, high quality, long-term records, but which address societal benefit areas beyond climate. Compared to low-latency weather and hazard satellite products, Reference EDRs provide longer, more consistent and seamless records with well-characterized uncertainties. Currently, NCEI sustains 30 packages of Reference EDRs and their associated collateral variables. It maintains a similar number in development or in transition to operations. Based on lessons learned, this White Paper describes Reference EDR production and use, and the resulting considerations for observing and data system design.

The Reference EDR User Community

NCEI initially focused on fundamental (calibrated base observations) and geophysical CDRs per the recommendations of the former Climate Change Science Program (2006; now the U.S. Global Change Research Program) and others. NCEI particularly pursued CDRs that could help address outstanding questions identified in Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4; 2007). The resulting suite of records has contributed to key climate discoveries, sustained climate monitoring, seminal assessments (e.g., National Climate Assessment, 2014), and increasingly underpins key climate modeling activities via the Ops4MIPS program (Teixeira et al., 2014), among others.

Importantly, however, the private sector and others are now leveraging these records for uses beyond climate research. The uses typically involve putting current weather and climate events into historical perspective such that businesses can anticipate likely market or environmental outcomes. In recent years, for example, agricultural interests have been better able to predict impacts of the Midwestern and California-based droughts using CDRs of vegetation health, precipitation and snow cover (Karl et al., 2012). Similarly, water managers are assessing likely outcomes of the current El Nino based on precipitation patterns during analog periods found in CDRs. In some cases, the private sector is leveraging CDRs to improve near-term profitability, gain competitive advantage or grow market share. For example, by merging historic climate and business records, some are predicting near-term consumer patterns based on recent past weather and climate events (Schreck et al, 2015). In other cases, businesses are providing value-added products and services based on CDRs.

The extent and growth of these alternative applications led NCEI to transition the CDR Program into the Reference EDR Program, where a Reference EDR is defined as "a time series of scientifically-based measurements of the Earth's environment with sufficient length, consistency, and continuity to provide stakeholders and decision-makers with timely, relevant and reliable information." This terminology is consistent with NOAA's use of Environmental Data Record (EDR) to describe its operational satellite products, and the use of "reference" to describe authoritative data sets (e.g., the U.S. Climate Reference Network). Accordingly, NCEI now manages the Reference EDR portfolio, inclusive of CDRs, based on needs of a broader set of stakeholders. Although many needs still emanate from the climate community, they are now complemented by those from geographic region interests, economic sectors, and other non-research communities.

Lessons Learned and Recommendations

Using guidelines from experts groups (e.g., WMO, 2003; NRC, 2004, 2005), NCEI has distilled eight non-functional requirements for Reference EDRs, specifically that they be: accessible, extensible, preserved, reproducible, sustainable, transparent, and continuously improved. The following outlines recommendations for mission and program design that would facilitate meeting these requirements effectively and affordably.

Issues for Observations

1. Observatory Continuity

The reprocessing of single mission data sets to correct for instrument, spacecraft or orbital degradation has become common, and results have grown increasingly accurate. However, multi-mission Reference EDRs introduce an additional reprocessing challenge since mission upgrades within an observation family (e.g., thermal sounding) typically impart data discontinuities. The discontinuities arise for many reasons, from changes in instrument design and sampling approach to orbital changes which effectively alter the time of observation. The problem is particularly acute for records derived from satellites from different agencies and countries (e.g., the global geostationary constellation). Regardless, discontinuities typically require extensive effort to ameliorate, and may impact downstream products and applications even after significant correction. For this reason, Reference EDR development is most straightforward and affordable when the source data come from the same or similar instrument and mission designs. Obviously, this consistency must be balanced by the advantages offered by occasional design improvements. We encourage the Decadal Survey team to consider the competing merits of continuity vs. upgrades in instrument design, and the efficiencies afforded by greater international coordination in this area, in their recommendations.

2. Calibration and Validation

Accurate instrument calibration is clearly important since calibration errors can mask or alias important environmental signals. For multi-mission products, however, it can be especially significant since cross-instrument calibration can effectively pass calibration errors across observatories covering long time periods. Clearly, missions should therefore archive and provide access to all prelaunch instrument test and calibration information as well as post-launch calibration and time-varying instrument configuration information and updates. However, to facilitate detailed retrospective analyses, missions should consider a more comprehensive set of practices and activities as suggested below.

Besides using internal instrument systems, adequate satellite calibration and validation programs include the capture, archive and ease-of-access to ancillary data collected from lunar and other stable targets, near-simultaneous observations with other remote sensing observatories, and field networks and campaigns. In the latter cases, despite the improvement in vicarious calibration sources – including in the growth and maintenance of field networks for land, ocean, ice and atmosphere – most field sites were designed primarily for purposes other than satellite validation. Where these sites are less amenable to straightforward spatial scaling, the mismatch in field and satellite instrument resolutions can add significant uncertainty to comparisons. This problem can be mitigated in some cases for satellite instruments that can be commanded to provide higher resolution or better quality data than afforded through normal operations. When this is possible, it should be employed episodically through the mission lifetimes to enable sustained "climate quality" calibration and validation.

It is important to highlight the critical need for overlap periods between successive missions to support instrument co-calibration – the tuning of one instrument to match the observations of another. A full year of overlapping data is generally recommended. The risk of missing sufficient overlap periods is obviously reduced when successive satellites are procured, launched and stored on-orbit well before the need dates. Note that Reference EDR developers often use overlap data to "calibrate backwards" in time, i.e., adjust the predecessor instrument observations to match those of the successor instrument with better calibration systems. The effectiveness of this approach means that new missions should continue to invest in and implement improved internal instrument calibration technologies.

3. Information Conventions, Standards and Access

Efficient mission and program data management can significantly reduce the costs of re-analysis and reprocessing activities. For example, all observations, products, metadata, calibration and validation data and associated information for a mission can be made accessible through a common information portal. Where field or other independent measurements and information are associated with specific mission data subsets, they can be stored together or linked virtually to facilitate discoverability and access. In all cases, the use of common data conventions, standards and formats can significantly reduce reprocessing or data integration costs. NCEI, for example, has adopted use of the NetCDF4 format, ISO19114-2 compliant metadata, with Climate and Forecast naming conventions to simplify processing, archiving and distribution. Note that, in some cases, access to key data or information is limited by proprietary, trade or intellectual property issues. We encourage the Decadal Survey team to consider opportunities that would simplify and improve access to and use of disparate data types and sources. Opportunities are clearly greatest when there is national and international cooperation in this area.

Issues for Data Processing and Integration

1. Interim Records

Although teasing out faint climate change trends in data can require extensive post-observation instrument analysis to minimize product uncertainties, other Reference EDR applications – particularly in the private sector – put product timeliness at a premium. These users are often willing to trade some accuracy for lower latency, yet still insist on the overall consistency afforded by Reference EDRs. The once-common practice of initially reprocessing data years after observation to maximize instrument knowledge and hence product accuracy does not meet their needs. This has led to "Interim" Reference EDRs, which are early estimates of Reference EDR values generated using the same climate-quality retrieval algorithms but lacking complete calibration or ancillary information (see Bates et al., 2015, and reports of the Sustained, Co-Ordinated Processing of Environmental Satellite Data for Climate Monitoring (SCOPE-CM) initiative). At NCEI, Interim EDRs are often used to forward extend Reference EDRs until all auxiliary data or other information is available, at which time they are replaced with reprocessed Reference EDRs. We encourage the Decadal Survey team to consider the unique user needs of timeliness and long-term stability for this user community.

2. Data Integration and Harmonization

To meet various user needs, NCEI is increasingly seeking to merge or harmonize satellite observations with legacy or contemporaneous other measurements, including those from in situ instruments, radar and other remote sensing sources, human-observers and proxy records (e.g., paleoclimate). Integrating satellite and meteorological station data, for example, can provide spatially and temporally continuous estimates of key variables such as surface air temperature (1.5- to 2 m height), which currently is not accurately retrieved from remote sensing alone. Merging data types can also extend record lengths from decades to centuries and put recent environmental changes in a richer historical context, enable better interpolation between satellite observations, and anchor remote observations to highly accurate field observations during processing (e.g., Reynolds et al., 2007). Because different observation approaches typically have different sampling characteristics – often measuring somewhat different environmental variables (e.g., sea surface skin vs. bulk temperatures), the integration and harmonization of these measurements is challenging. We encourage the Decadal Survey team to consider mission data system designs that would facilitate greater integration of space measurements with other observation types. Doing so may markedly reduce the cost, complexity and uncertainty in integrated products.

3. From Research to Operational Production

Algorithm development continues to be a long, labor-intensive process, centered on the principal investigators and their expertise in the science of Reference EDRs. To enable the continued production of their products into the future, NCEI has engaged in a series of transition activities from the developer to move the algorithms, documentation, and workflows necessary for Reference EDR production to an operational environment. In doing so, NCEI has approached the transition and operations issues related to the code through a variety of methods, including code wrapping, rejuventation, and re-writing. While all methods aspire to provide continued, institutionalized production at a reduced integral cost, rejuvenation and re-writing of original scientific code bases have proven to be, in NCEI's experience, very costly and a barrier to efficient institutionalized production. A priori adoption of community or NCEI code standards at the development stage is the least costly path to independently maintainable operational code and algorithms, and support for these code standards and practices by the community would pay dividends in the operational community. These same standards and practices also facilitate information preservation in the archive.

4. Improving Use

There is a major underlying dependency of the sustainment and use of Reference EDRs on the "back end" data systems, including software, networks, algorithm libraries, and archive facilities that are often overlooked during the planning of individual observing systems. The growth of standards and tools that underlie machine-to-machine access and other modern technologies has reduced infrastructure costs and data latency for many user communities. However, other users communities – particularly non-experts in IT and/or satellite data – remained challenged in discovering, accessing and using satellite products compared to other environmental products (e.g., meteorological station data). This stems in part from product volumes – which for Reference EDRs are typically measured in terabytes -- as well as the relative complexity of form and format in which these data are available. For example, some products are available only in swath-format granules or through basic access approaches (e.g., ftp, http). Although access and distribution issues are generally the domain of data centers, they are influenced or facilitated by satellite program goals and requirements. Readily achievable near-term improvements include improving server-side tools to analyze, map, subset, format and/or download data by user-specified attributes. The government's provision of end-use code through modern collaborative code repositories (e.g. GitHub) that provide examples of data access, filtering, and visualization may allow more rapid and efficient utilization of Reference EDRs by the wider data application community that are seeking to use them, particularly for broader uses outside of the geosciences, including for socioeconomic and public health applications. Users would further benefit if agencies or their industrial partners were to provide a common multi-agency discovery and access point, and tag data stores with advanced metadata schema and geohashes to enable web crawlers, to simplify and facilitate their data searches.

Sustaining NOAA's Reference EDRs

NCEI now sustains 30 satellite Reference EDR packages, with additional packages transitioning to operations in coming years (see Table 1.). These support NCEI's goal to provide fundamental records (calibrated radiances and other sensor records) and geophysical records relevant to NCEI stakeholder communities (e.g., water, energy, agriculture, and climate). Most were competitively selected based primarily on scientific merit and have undergone extensive external review. Others were selected primarily to meet emerging user requirements subject to internal NOAA review. NCEI actively manages its product portfolio in an open and transparent manner to help ensure continuing community involvement and relevance. However, even though these Reference EDRs are largely core or foundational, their sustainment depends on the continuing availability of resources and capabilities (including support of space and non-space observatory assets). We therefore encourage the Decadal Survey team to identify capabilities to support these or related Reference EDRs for the next decade.

Summary

Reprocessing and sustaining consistent and relevant multi-mission satellite records introduces new challenges and requirements on satellite programs and missions. To our knowledge, there were no operational U.S. government programs providing CDRs or Reference EDRs at the time of the last Earth Sciences Decadal Survey. Since that time, however, NOAA has instituted this capability, expanded and further diversified its user community, and learned many lessons as described above. As the value of Reference EDRs is increasingly realized by different user communities, we believe NOAA's and similar programs will continue and likely expand across the agencies, and are therefore highly appropriate for informing the Decadal Survey process.

References

Bates, John J., Jeffrey L. Privette, Edward J. Kearns, Walter Glance, Xuepeng Zhao (2015), Sustained production of multidecadal climate records - Lessons from the NOAA Climate Data Record Program, *Bull. Amer. Meteor. Soc.*, *in press*.

Climate Change Science Program, 2006, Climate Change Science Program Strategic Plan Chapter 12. Observing and Monitoring the Climate System, published by the U.S. Climate Change Science Program, Washington, DC.

Intergovernmental Panel on Climate Change, 2007, The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. See also IPCC references therein.

Karl, T. R., B. E. Gleason, M. J. Menne, J. R. McMahon, R. R. Heim, Jr., M. J. Brewer, K. E. Kunkel, D. S. Arndt, J. L. Privette, J. J. Bates, P. Y. Groisman, and D. R. Easterling, 2012, U.S. temperature and drought: Recent anomalies and trends. Eos, Transactions, American Geophysical Union, 93, 473-474, doi:10.1029/2012EO470001

National Research Council, 2004, Climate data records from Environmental Satellites, National Academies Press, Washington, DC, ISBN: 0-309-09168-3.

National Research Council, 2005, Review of NOAA's Plan for the Scientific Stewardship Program, National Academies Press, Washington, DC, ISBN: 0-309-55202-8.

National Research Council, 2007, Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond, National Academies Press, Washington, DC, ISBN-10: 0-309-10387-8.

Privette, Jeffrey, Bruce Barkstrom, John Bates, Mike Bonadonna, Kandis Boyd, DeWayne Cecil, Bryant Cramer, Gary Davis, Thomas Karl, Jack Kaye, Chet Koblinsky, Michael Tanner, and David Young (2008), Restoration of NPOESS Climate Capabilities – Climate Data Records, 4th Symposium on Future National Operational Environmental Satellite Systems, The 88th AMS Annual Meeting, New Orleans. http://ams.confex.com/ams/pdfpapers/131058.pdf

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax (2007), Daily high-resolution-blended analyses for sea surface temperature. J. Climate, 20, 5473–376 5496, doi:10.1175/2007JCLI1824.1

Schreck, C. J., III, S. Bennett, J. M. Cordeira, J. Crouch, J. Dissen, A. L. Lang, D. Margolin, A. O'Shay, J. Rennie, and M. J. Ventrice, 2015: Natural gas prices and the extreme winters of 2011/12 and 2013/14: Causes, indicators, and interactions. Bulletin of the American Meteorological Society, doi:10.1175/BAMS-D-13-00237.1. Available online at: http://dx.doi.org/10.1175/BAMS-D-13-00237.1

Teixeira, Joao, Duane Waliser, Robert Ferraro, Peter Gleckler, Tsengdar Lee, and Gerald Potter, 2014: Satellite Observations for CMIP5: The Genesis of Obs4MIPs. Bull. Amer. Meteor. Soc., 95, 1329–1334. doi: http://dx.doi.org/10.1175/BAMS-D-12-00204.1

World Meteorological Organization, 2003, The second report on the adequacy of the global observing systems for climate in support of the UNFCCC. (GCOS 82; WMO/TD no. 1143) Geneva, Switzerland: World Meteorological Organization, Global Climate Observing System

Table 1. NCEI's Reference Environmental Data Records. Program details are at www.ncdc.noaa.gov/cdr.

Reference Environmental Data Record Bundle	Start of Record	Current Space Data Sources
Atmospheric		
AVHRR Aerosol Optical Thickness	1981	AVHRR
AVHRR Cloud Properties PATMOS-x	1978	AVHRR
Mean Layer Temperature - NOAA	1978	AMSU
Mean Layer Temperature - RSS	1978	AMSU-A
Mean Layer Temperature - UAH	1978	AMSU
Mean Layer Temperature - UCAR (Lower Stratosphere)	2001	GSP RO sources, AMSU-A
Mean Layer Temperature - UCAR (Upper Trop & Lower	2001	GSP RO sources, AMSU-A
Ocean Heat Fluxes	1988	AVHRR, TMI, AMSR-E, MODIS, SSMIS, AMSU-A, AIRS, NSCAT
Ocean Near Surface Properties	1988	AVHRR, TMI, AMSR-E, MODIS, SSMIS, AMSU-A, AIRS, NSCAT
Outgoing Longwave Radiation - Daily	1979	HIRS, Global Geostationary Imagers
Outgoing Longwave Radiation - Monthly	1979	HIRS, Global Geostationary Imagers
Ozone – ESRL	1979	Not being extended
Precipitation - PERSIANN-CDR	1983	Global Geostationary Imagers, GPCP product (SSMIS, TOVS, AIRS)
Solar Spectral Irradiance	2004	Mg II index & USAF white-light sunspot regions
Total Solar Irradiance	1610	Mg II index & USAF white-light sunspot regions
Oceanic		
Sea Surface Temperature – Optimum Interpolation	1981	AVHRR, SSMIS
Sea Ice Concentration	1978	SSMIS
Sea Surface Temperature – Pathfinder	1981	AVHRR
Sea Surface Temperature - WHOI	1988	AVHRR, SSMIS
Terrestrial		
AVHRR Surface Reflectance	1981	AVHRR, TOMS
Leaf Area Index and FAPAR	1981	AVHRR, TOMS
Normalized Difference Vegetation Index	1981	AVHRR, TOMS
Snow Cover Extent (Northern Hemisphere)	1966	AVHRR, GOES Imager
Fundamental		
AMSU Brightness Temperature - NOAA	1998	AMSU
AVHRR Reflectance – PATMOS-x	1978	AVHRR
Geostationary IR Channel Brightness Temperature –	1980	Global Geostationary Imagers
HIRS Ch12 Brightness Temperature	1979	HIRS
MSU Brightness Temperature - NOAA	1978	Not being extended
SSMI(S) Brightness Temperature - CSU	1987	SSMIS