Long-term Data Sets – Key to Understanding Past and Future of Solar Activity

A. Pevtsov (NSO), S. Gibson (HAO), D. Webb (Boston College), M. Dikpati (HAO), J. Burkepile (HAO), L. Bertello (NSO)

Summary statement: We live next to a variable star, and the only way to learn about its long-term behavior, and ultimately to be able to predict it, is to guarantee the survival and continuity of long-term synoptic observations.

The heliophysics research community is pushing hard to explore the Sun and its cosmic neighborhood with new instruments in new wavelength bands, with the highest-possible spatial resolution and the fastest time cadence. We justify this by the need to better understand the physics of solar activity and the societal "mandate" to develop a reliable space weather forecast. However, our understanding of solar activity will be incomplete and even inaccurate, if we do not know how present activity compares with the past and what changes may had been occurring in the properties of solar phenomena. This necessitates long-term monitoring of solar activity. The success of synoptic observations requires long-term sustainable funding, but unfortunately, we are witnessing an alarming decline in funding levels for these programs. The uniformity of synoptic data also requires careful planning for on-time replacement of aging instruments and their cross-calibration. Equally important is long-term preservation of existing data, but in absence of proper planning, we may already be witnessing irreversible loss of some important historical datasets.

Long-term heliophysics observing programs and datasets are essential to the following key capabilities:

Benchmarking: ***Long-term datasets provide reference for*** the ***"typical***" or "***normal" states of natural systems such as the Sun and heliosphere***. For example, several recent studies have suggested that solar cycle 24 activity may have been unusually low. However, most of the observations that we have to compare to this past cycle are from the past 50 years. How typical has this period of time actually been? By maintaining long-term datasets, we enable future researchers to obtain a progressively deeper understanding of the expected variance in solar activity from cycle to cycle.

Systemic change: ***Long-term datasets provide information about evolutionary (time scales longer than solar cycle) and non-evolutionary/transient (e.g., flare/CME) changes in these natural systems.*** Without long-term datasets we may not be able to identify how the system (sun, heliosphere) changes in course of its natural evolution. One recent example is the claim that the activity of recent solar minima is persistently getting lower. Does this indicate a major random shift in solar activity, or could it be an indication of a natural gradual (evolutionary) change?

Long-term datasets are thus required for establishing benchmarks and evaluating systemic change. However, we note the datasets alone may not be sufficient, as a strong modeling effort is necessary to properly evaluate the long-term observations in respect to typical/atypical, evolutionary/non-evolutionary changes.

Finally, long-term datasets serve as the basis for future discoveries of presently unknown phenomena. Concepts such as the impact of solar-driven space weather on technologies or the existential threat of global climate change were completely unknown when Galileo, Schwabe, and Wolf started to patiently record dark sunspots centuries ago, but their efforts were vital to our current understanding of these issues. If we do not maintain the long-term observational record, scientists 15–20 years from now will perhaps wonder with disapproval why we did not continue recording the long-term of observations of some specific phenomena potentially critical to future societal needs.

Thus, our final key point:

<u>Exploration:</u> Long-term datasets of solar activity feed future research to solve issues that may not be identified at the time when the data are acquired.

One of the current major issues that must be addressed is the lack of a global long-term plan for heliophysics data preservation. NASA has some plans in place for the preservation of data from its past and future missions, but the issue goes beyond NASA, encompassing all heliophysics data, both spaceborne and ground-based, national and international. Current modus operandi often involves a host institute "taking care of its own data", which runs the risk of loss of data at the end of project, outdated file formats, and difficulties/obsolescence of capabilities for data access. Loss of science funding has placed older datasets in jeopardy and some data that was once available online is no longer readily accessible. ***There is a lack of clear understanding of who is responsible for the long-term preservation of historical datasets.***

What can we do? This may be broken into two sub-questions dealing with 1) long-term observational programs and 2) long-term data preservation.

1) With regards to long-term observational programs, the following actions could be envisioned:

- Critical evaluation of synoptic observations (which data are critical to continue, and what could be discontinued)
- Long-range planning for the maintenance of critical datasets (e.g., the creation of a network of automated stations and the development of a plan for routine instrument upgrade/replacement, as well as a plan for maintaining the uniformity of datasets despite changing instrumentation);
- The establishment of new funding models (longer funding periods for synoptic programs, inter-agencies funding)
- The promotion of reproducibility of data reduction (e.g., common formats, unified data reduction codes, clear documentation of data and reduction, the reproducibility of data reduction codes).

2) With regards to long-term data preservation, the action items should include:

- Creation of a comprehensive list of the data to be preserved, and identification of dedicated funding sources for their preservation and curation.
- Development of a clear path for preserving synoptic data and identification of the responsible parties (NOAA, NSF, NASA, ??) as well as their respective roles.
- The establishment of requirements for long-term data preservation for all future long-term observing programs in Heliophysics.

Finally, all aspects of long-term datasets – including their creation, preservation, curation, and scientific exploration) – will benefit greatly from close international collaboration. By coordinating such efforts globally, it should be possible to share expenses and responsibilities, to prevent unnecessary duplication of efforts, and to develop data sharing policies and international agreements for data use. Steps in this direction have been taken through the International Astronomical Union (IAU), with the passage of IAU Resolution B3: "on preservation, digitization and scientific exploration of historical astronomical data", which was accepted by the XXX General Assembly in 2018. As a follow up of this work, the IAU Working Group on the Coordination of Synoptic Observations of the Sun instituted a Union-wide Survey of historical astronomical data, which received information about 115 endangered historical datasets across all fields of astronomy.

These efforts represent a start, but much remains to be done if we are to deepen our understanding of long-term solar science between now and 2050.