Volume Title
*ASP Conference Series, Vol. **Volume Number***
Author

The Science of Calibration

Stephen M. Kent[1]

[1]*Fermilab/MS 127, P.O. Box 500, Batavia, IL 60510*

Abstract. This paper presents a broad overview of the many issues involved in calibrating astronomical data, covering the full electromagnetic spectrum from radio waves to gamma rays, and considering both ground-based and space-based missions. These issues include the science drivers for absolute and relative calibration, the physics behind calibration and the mechanisms used to transfer it from the laboratory to an astronomical source, the need for networks of calibrated astronomical standards, and some of the challenges faced by large surveys and missions.

1. The Hyperspace of all data

One can think of electromagnetic radiation as being characterized by four dimensions (or five if one allows that a solid angle on the sky is two-dimensional.) Furthermore, each dimension can be characterized by a range in both absolute value and in precision. It is this space that one seeks to calibrate. These dimensions and their ranges can be described as follows:

- Flux (measured, e.g., in erg cm^{-2} s^{-1} hz^{-1}) is usually specified at a specific frequency, wavelength or energy.
  - The objects that we are interested in calibrating span a range in total flux of approximately 23 decades (sun to faintest LSST objects).
  - A high precision experiment such as the Kepler Mission (Borucki et al. 2006) attains a precision of order 10 micromags, or a range of 5 decades, in its planetary occultation search program.
- Wavelength is measured for two purposes. One is to specify the bandpass in which a flux is measured. The other is to measure quantities such as redshift.
  - The range over which we might possibly obtain a signal is of order 25 decades, although the extremes are not yet explored. At the low frequency end, the limit is ultimately set by the plasma frequency of the interstellar medium, which is of order 3 khz. At the high energy end, a cutoff exists for photons above about 100 TeV from scattering off the microwave background. (Galactic sources, if any exist, presumably could be detected to still higher energies.)
  - The highest precision measurements at present are probably those of planet searches utilizing radial velocity measurements - roughly 1 meter/sec, or a range of 10 decades.

- Astrometric measurement include both the position on the sky (2 coordinates) and the angular velocity.
  - For position, the range is of no interest (2π radians) but for angular velocity, we are interested in a range of perhaps 11 decades, from the rotation of the Earth to the rotation of the Milky Way.
  - The highest precision astrometry will likely be that of the GAIA mission (Lindegren et al. 2008) - 10 decades, or of order 20 microarcsec.
- Time.
  - Absolute timing is of interest over perhaps 18 decades - from the arrival times of pulsar pulses (100 nanoseconds) to the time kept by the Julian calendar (6000 years).
  - The highest relative timing precision is perhaps that of lunar laser ranging - 12 decades, or 1 mm out of 600,000 km.

In addition to these primary dimensions, one also has the "extra dimensions" corresponding to the three additional Stokes parameters, which one might use if measuring, say, polarization of the microwave background. Additionally, one might want to consider "extra particles" of astronomical origin, such as cosmic rays, neutrinos, axions, neutralinos, gravitions, and other more exotic species. Thus, the space of astronomical data in need of calibration is potentially vast.

2. Flavors of Calibration

Calibrations of measurements come in many varieties or "flavors", and it is useful to distinguish them explicitly, since different flavors impose different requirements on a calibration.

2.1. Absolute Calibration

Such calibrations are expressed in physical "SI" units, such as a flux in erg cm^{-2} s^{-1} hz^{-1}. In the case of a flux, one is usually interested in just the shape of a spectrum (spectrophotometric calibration) and only occassionally in the absolute zero-point. The reason is that we usually don't know the intrinsic luminosity or distance to an object with sufficient precision to make an accurate measure of the absolute flux of much use. For the time dimension, one might choose the big bang as an abosolute zero point, but again, we do not know it with enough precision for it to be of much use.

2.2. Relative Calibration

Often one makes measurements of one object relative to another in a way that does not make direct use of SI units and often with an accuracy that is much better than the SI values are known - e.g., one can measure the color excess of F type stars relative to a nominal main sequence and obtain a measure of the metallicity of a star. Such comparisons are often made by combining data from different instruments and different telescopes and rely on using celestial objects to perform "intercalibration" of the two data sets; much of classical broadband (e.g., *UBV*) photometry would fall in this category.

2.3. Differential Calibration

The most precise measurements are typically time-series observations of objects that rely on all data being collected by a single instrument. For example, the Kepler Mission, which carries out precision differential photometry of stars, utilitized out-of-focus CCD detectors that are unfiltered, and thus the fluxes are of no particular use beyond the Mission itself.

3. Science Drivers

Are we "science-limited" or "calibration-limited"? In other words, is there a science payoff to be had if we keep pushing harder on improving calibrations? The following are examples of some areas where the science is (or will soon be) limited by calibration, although the list of all possible examples is large.

3.1. Absolute (or spectrophotometric) flux

Type Ia supernovae are one of four principal methods for probing the expansion history of the universe. These supernovae are considered to be standardizable candles: one can derive the rest-frame B-band luminosity of a supernova that is the same on average, with a scatter of 15% for a single object. By observing hundreds of these objects, one can hope to distinguish cosmological models that differ in their predictions by as little as 2%. Since the rest-frame *B* band falls in different observational band passes, depending on its redshift, the relative zero-points of all bands from 0.35 microns to 1.7 must must be cross-calibrated to an accuracy of better than 1% across this wavelength range (Kent et al. 2009).

White dwarfs models are thought to be good enough that one can predict the absolute fluxes and spectra of individual objects to an accuracy of 1-2% (Holberg et al. 2008). GAIA will be able to measure the absolute distances to a number of nearby white dwarfs and thus provide a direct test on these models. At present white dwarf models are thought to provide the best spectrophotometric calibrations, and independent verification is essential.

3.2. Absolute calibration - Astrometry and Time

The absolute rotation rate of the Milky Way, about 5 milliarcsec per year, is important for determining fundamental parameters of the Galaxy and measuring the kinematics of stars. Measuring this rate accurately requires establishing an intertial reference frame with high precision.

Pulsars measure elapsed time, while atomic clocks measure frequency. The stability of pulsars is not currently as good as the best atomic clocks over periods of a few years, but they may begin to rival the accuracy of such clocks on time periods greater than about a decade (Hartnett & Luiten 2011).

3.3. Relative Calibration - Flux

Globular clusters were originally thought to be homogenous systems with all stars having the same metallicity and age. High precision photometry, however, shows that the lower main sequence in some clusters can be split into two or even three distinct branches, such as for NGC 2808 (Piotto et al. 2007). The separation can be just a few hundredths of a magnitude, requiring statistical errors to be smaller than this in order to make the separation visible.

4. Relative calibration - Astrometry

The orientation of the velocity ellipsoid for stars above and below the plane of the Milky Way is sensitive to the shape of the gravitational potential of the Galaxy. The ellipsoid is measured by combining radial velocities with relative proper motions of large numbers of stars. Increasing the precision with which proper motions can be measured increases the distance out to which the ellipsoid can be determined.

4.1. Differential Calibration - Astrometry

Parallaxes are the most reliable way to determine the distances to stars, but up to now the precision has only allowed distances to be measured out to of order 100 pc. GAIA will obtain parallaxes with a precision of order 20 microarcseconds, allowing distances to be measured out to of order 5 kpc.

5. The Physics of Calibration

In almost all cases, one can measure physical quantities in the laboratory to a higher accuracy than can be achieved at the telescope. Improving the situation requires ”bringing physics to the data” with as little loss in accuracy as possible.

5.1. Flux

A flux calibration requires the ability to generate photons at a known rate with a known energy. At radio through optical wavelengths, one often still relies on black bodies, which have a known spectrum controlled by a single parameter, the temperature (Findlay 1966). [Amusingly, a bank thermometer, which always seems to be in error by a few degrees, can still measure an absoute temperature to an accuracy of better than 1%.] At optical wavelengths, where small errors in temperature can lead to big errors in the computed flux, an alternate technique is now preferred, in which tunable lasers beams illuminate a thermally-controlled cavity, which acts as a calorimeter - the power of the beam is determined by comparing with the equivalent power from an electrical current (Livigini 2003). At higher energies, one can use synchrotron accelerators to generate a beam of light with a known spectrum - the controlling parameters are the energy of electrons and the radius of curvature (Arp et al. 2007). At high energies one relies either on accelerator technology, for which particle of a known energy generate photons by bremsstrahlung (Baldini et al. 2007).

5.2. Wavelength

In the radio regime, frequency synthesizers provide more accuracy than is generally needed. In the optical one relies on emission line lamps using atomic species such as thorium-argon to provide an absolute reference. Frequency combs are now being developed that provide exquisite precision in the lab, although tranferring that precision to the telescope is still problematic (Wilken et al. 2012). At still higher energies, one relies on the same bremsstrahlung mechanism described above or radioactive decays of heavy elements, which produce gamma rays with a well-defined energy.

5.3. Astrometry

A radian has no SI units and thus no laboratory reference is needed. For measuring angular motions, however, one still needs to define an intertial reference system. Laboratory devices such as a Sagnac inteferometer can measure rotations as low as 10^3 mas per year (Stedman et al. 2003), which is much too coarse an accuracy to be useful for measuring the rotation of the Galaxy. Gravity Probe B was able to measure general-relativistic "frame-dragging" in an orbiting satellite with a precision of about 7 milliarcsec per year, which is still too coarse to measure Galaxy rotation (Everitt et al. 2011). Thus, one must resort to "Mach's Principle" and use the universe at large to define an intertial system, e.g., with a system such as the ICRS (Arias et al. 1995).

5.4. Time

For time intervals that reach to centuries or longer, the best "timepiece" is still the calendar. There does exist one project, the "10,000 year clock", that seeks to implement a clock that will operate continuously for such durations (Hillis et al. 2011). For more pedestrian purposes, one can make use either of atomic clocks or GPS.

6. The Experimental Apparatus

Having determined the science and the physics, we proceed next to the experimental apparatus itself. In nearly all cases, independent of wavelength, we can break down the appartus into detector/instrument, telescope, and (for ground-based experiments), the atmosphere.

Ideally we would observe our physics source the same way we observe our science object. In reality, such a situation is generally not possible, so we are left calibrating our apparatus either in pieces or by observing calibration sources in a way that is representative of science observations.

In some cases one can switch the instrument/detector between a calibrated source and a science target - such a procedure is particularly beneficial when operating in a power mode where the gain or bandpass of the instrument varies with time. Examples of space missions employing this technique include Mariner 6/7, which carried radiometers along with calibrated black body loads (Chase 1969) and the MODIS instrument on board the Tera and Aqua missions (Xiong & Barnes 2006), which carried a set of calibration devices including a black body and a monochromator.

Array detectors such as CCDs are now prevalent over a wide range of wavelengths, from the millimeter for QUIET (Buder 2010) and the South Pole Telescope (Ruhl et al. 2004) to the X-ray for Chandra (Weisskopf et al. 2002) and XMM-Newton (Struder et al. 2001). A common need is to intercalibrate all the pixels or channels of the detector, since reference sources such as standard stars cannot be conveniently observed everywhere in the focal plane. The wavelength response across the focal plane may also be variable. New ways to create "flatfield" screens are needed.

In some cases, the telescope may be calibrated along with the instrument as part of the calibration process, but in other cases, such as the radio, one utilizes simple designs such as horn antennae for which the gain pattern can be calculated from first principles.

Any ground-based observation must contend with the atmosphere at some level. Water vapor is particularly pernicious because it is time variable and affects nearly all wavelengths, modulating either the opacity, bandpass, or timing (for interferometric observations). A variety of methods are now being utilized to monitor the water vapor content of the atmosphere (Cooper et al. 2003; Blake & Shaw 2011).

7. The ”Chain of Calibration”

When it is not possible to bring “physics to the data” directly (particularly for flux calibration), one often resorts to intermediate calibration sources, such as calibrated lamps, standard stars, or other references. The calibration process then becomes one of measuring a succession of ratios or differences in going from physics to science.

7.1. Flux

7.1.1. Fundamental flux standards

Very few astronomical sources can be compared directly against physics references, since one is often limited to the brightest objects in the sky. A list of typical objects includes the following:

- Cas A (radio)
- Mars (thermal IR)
- Vega (optical/IR)
- HST White Dwarfs (both physics and astronomical objects)
- Sun (optical/IR)
- Crab Nebula (X-ray)

Rather interestingly, three of these sources are noticeably variable, so the variability must either be calibrated itself or compensated in some other way.

7.1.2. Standard Source Networks

The fundamental flux standards are often too bright to be compared to science objects directly, so one relies on extensive networks of secondary standards, which extend both the areal coverage across the sky and the intensity range to fainter objects.

In the optical and near-IR, one additionally makes use of extensive standard star networks that are set up to perform relative photometry, such as $UBVRI$. While such networks are tied to physics sources, the relative accuracy within the network is much higher than the absolute tie to physics.

7.2. Astrometric Standards

Astrometric calibration benefits from having an extensive network of reference sources across a wide dynamic range, just as for flux calibration. The ICRS system utilizes distant radio-loud QSOs around the sky for its absolute references, but such a network is too sparse and does not reach to all wavelengths to suffice by itself. Thus, we rely on catalogs of stars such as UCAC (Zacharias et al. 2000) and NOMAD (Zacharias et al. 2004) to provide extra coverage, even though one must deal with the annoying problem of stars that move.

7.3. Error Creep and Cures

Each step in a multistep calibration process inevitably introduces errors, and quantifying the errors at each step is often nontrivial. As an example, the COBE FIRAS instrument carried thermometers calibrated to an accuracy of 1 mK, but unexpected problems led ot the final measurement of the CMB temperature to have an uncertainty of 5 mK. (Mather et al. 1999). Subtle sources of error in the calibration process are often unanticipated and must be addressed using data in the field, often with novel, unplanned procedures.

The cure for error creep is to have multiple independent, parallel methods at each step or across a combination of steps in the calibration process. As an example, the MSX mission relied on three calibration methods (Fig. 1): laboratory characterization, external standard stars, and onboard black body references that were ejected from the spacecraft and observed with the satellite (Price et al. 2004).

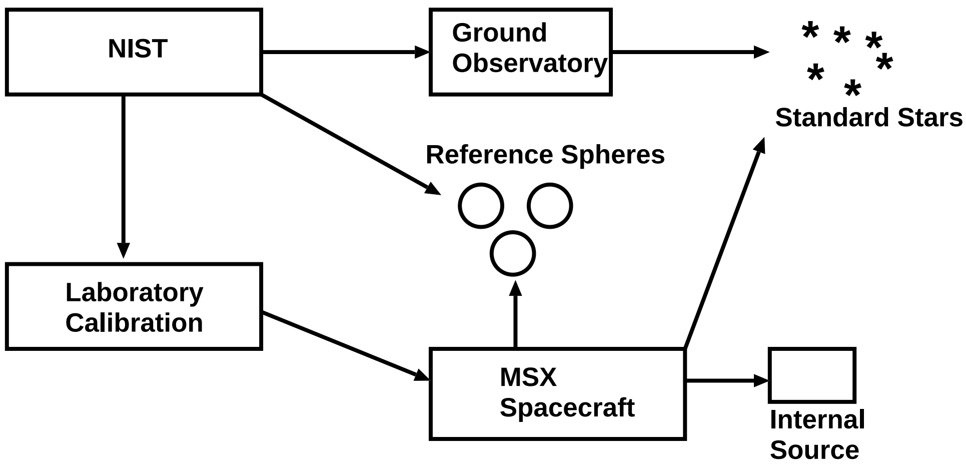

Figure 1. The three legs of calibration for the MSX experiment, showing how each is traceable back to NIST. (Adapted from Egan et al. 1999.)

8. Challenges for Large Surveys And Missions

For space missions, the good news, if you are an astronomer, is that the atmosphere is eliminated. The bad news, if are an Earth scientist, is that the atmosphere has been introduced. The ugly, no matter what your discipline, is that onboard calibration hardware costs money. There is competition between scientists, who want the best calibration possible, and program managers, who want the least budget that will meet the mission needs. The difficulty is that one cannot necessarily anticipate all the calibration needs in advance or demonstrate, with any reliability, that a particular calibration scheme will, indeed, meet mission requirements.

The need for new flatfielding techniques, including in new regimes such as the millimeter, has already been mentioned.

Requirements are becoming more demanding than in the past. LSST wants to achieve a precision of 1% with a single observation (Ivezic et al. 2008) in order to detect low-amplitude variable objects. In the past, one relied on making multiple observations on different nights to meet that goal.

In the radio, traditional flux standards no longer work as receivers and and telescopes are pushed to millimeter and submillimeter wavelengths.

One is in continuous needs of addition astrometric standards that reach to ever fainter limits.

Surveys need to be realistic. Often the question is not "what is the best that I can achieve" but rather "what is the worst that I will consider acceptable."

How do we report data for surveys where the bandpass varies across the field of view? Homogenizing the data means having some knowledge of the spectral energy distribution. Color terms that were adequate in the past may no longer be acceptable as we include nonstellar objects (such as emission line galaxies) with non-smooth spectra when we want to achieve better than 1% accuracy.

9. Conclusions

It is fair to say that we are "calibration-limited" rather than "science-limited" for many problems.

In closing, the space of astronomical data is large, and while one's individual calibration needs might seem disjoint from one another, there is actually considerable commonality.

Acknowledgments. Fermilab is operated by Fermi Research Alliance, LLC under Contract No. De-AC02-07CH11359 with the United States Department of Energy.

References

Arias, E. F., et al. 1995, AA, 303, 604
Arp, U., et al. 2007, Appl. Opt., 46, 25
Baldini, L., et al. 2007, AIP Conf. Proc., 921, 190
Blake, C. H., & Shaw, M. M. 2011, PASP, 123, 1302
Borucki, W., et al. 2006, ISSIR, 6, 2007
Buder, I. 2010, Proc. SPIE, 7741, 77441
Chase, S. C. 1969, Appl. Opt., 8, 639
Cooper, R., et al. 2003, Proc. SPIE, 4857, 11
Egan, M. P., et al. 1999, AFRL-VS-TR-1999-1522, Tech. rep., AFRL
Everitt, C. W. F., et al. 2011, PRL, 106, 221101
Findlay, A. 1966, ARAA, 4, 77
Hartnett, J. G., & Luiten, A. 2011, Rev. Mod. Phys., 83, 1
Hillis, D., et al. 2011, arXiv:1112.3004
Holberg, J. B., et al. 2008, AJ, 135, 1239
Ivezic, Z., et al. 2008, arXiv:0805.2366
Kent, S. M., et al. 2009, arXiv:0903.2799
Lindegren, L., et al. 2008, Proc. IAU Symp. 248, 217
Livigini, D. 2003, Special Publication 250-62, Tech. rep., NIST
Mather, J., et al. 1999, ApJ, 512, 511
Piotto, G., et al. 2007, ApJ, 661, 53

Price, S., et al. 2004, AJ, 128, 889
Ruhl, J. E., et al. 2004, Proc. SPIE, 5498, 11
Stedman, G. E., Schreiber, K. U., & Bilger, H. R. 2003, Class. Quantum Grav., 20, 2527
Struder, L., et al. 2001, AA, 365, 18
Weisskopf, M. C., et al. 2002, PASP, 114, 1
Wilken, T., et al. 2012, Nature, 485, 611
Xiong, X., & Barnes, W. 2006, Adv. Atmos. Sci., 23, 69
Zacharias, N., et al. 2000, AJ, 120, 2131
— 2004, AAS, 205, 4815