A REDEFINITION OF THE SDSS PHOTOMETRIC SYSTEM I: PRINCIPLES AND DEFINING EQUATIONS

James E. Gunn, David Hogg, Doug Finkbeiner, David Schlegel

12/08/2001

I. INTRODUCTION

The current *definition* of the SDSS photometric system is seriously flawed in several aspects, the worst of which is that it is currently most emphatically *not* the natural system of the 2.5-meter telescope. When the system was originally designed, an attempt was made to make the filters for the PT, the USNO telescope, and the 2.5-meter as nearly identical as possible, but, as is now well known, those filters are modified by the different environmental effects present in the three telescopes sufficiently that they are very far from defining a uniform system. Some time ago the photometric group agreed that we would at some point migrate to describing our photometry on *two* systems, the standard system set up by the standard precepts by the 40-inch Navy telescope (hereinafter the *USNO* system, designated by u', g', r', i', and z'), and a system defined by the average responses of the SDSS camera *at a standard airmass*, to be called simply u, g, r, i, and z. We refer to this system as the *SDSS system* in the text below.

We have been using now for some years the designation $u*$, $g*$, $r*$, $i*$, $z*$ awaiting a proper definition and establishment of stable zero-points for the 2.5-meter. This document is the first of two which attempt to define the new system. It will be followed by one which spells out in detail the mathematical approach and those of the defining parameters which can be calculated *a priori*.

Though we have discussed the reason for this distinction and change fairly extensively at many meetings, let us review the reasons briefly here.

1. We need the USNO system. It was carefully developed, is reasonably self-consistent internally, and would require an unacceptably long time to reproduce on a system closer to the SDSS one — even if that could be done, which would not be easy. The USNO system is sufficiently close to the SDSS one that we can reasonably expect linear transformations to carry us from one to the other *for ordinary stars over a restricted color range*, which is sufficient for our purposes. Indeed, theoretical transformations have been calculated for spectrophotometric atlas stars, and they are well behaved.

2. If we chose to attempt to transform the survey magnitudes to the USNO system, we would be in a position in which

a. The magnitudes in a given band would not be simply related to the measurements on a frame taken in that band. To measure an i' magnitude, say, one would have to measure images on at least two frames in different bands and calculate a (perhaps zenith-angle dependent) linear combination.

b. The colors of objects with unusual spectra (*i.e.* those which differ markedly from the spectra of ordinary stars), which account for a very large number of our targets and a completely disproportionate amount of astrophysical interest, are not represented at all accurately by such standard linear transformations; in fact, such transformations often yield very much *less* accurate results than ignoring the transformations altogether. In order to understand their colors of, *e.g.* emission line objects, one *must* work on the natural system of the camera.

This is not a satisfactory situation, and suggests that we must publish magnitudes for the survey on the natural camera system, *i.e.* the SDSS system.

3. We need the USNO system only to standardize the SDSS system, which can be done entirely using ordinary stars with a restricted color range, so long as one knows the transformations between the two systems for these prosaic objects.

In order for this scheme to work, we must have several things in place:

1. We must know the transformations (the "b-terms") which connect the USNO and SDSS systems, and must also know the transformations which connect the PT (which is sufficiently different from the Navy system that it represents yet *another* natural system) to USNO. The proposal is never to acknowledge the existence of the PT system as an independent entity, but to do photometry with the PT on the USNO system, using standard photometric transformation techniques. Thus the PT system exists as an intermediate product, but magnitudes are *never* reported on it. Indeed, there is very strong evidence that it changes, at least prior to July when new (hopefully) stable filters were installed, and that the PT-USNO b terms are highly variable: for g', the band for which they are best determined, over a range of nearly 0.07 mag/mag. These transformations have (we think we now understand why) been unexpectedly difficult to come by, but in the process of producing them we have come to understand a number of error sources in the photometry which would have plagued us had the effort not been made. As of this writing, convergence seems nigh.

2. We must rewrite the photometric code to take explicit account of the two systems. It would make sense for the MTPipe/Excal/Kali sequence, which outputs the photometric system and calibrated magnitudes, to do the latter on both the USNO and the SDSS systems but pass to Photo and NFCALIB the SDSS-transformed versions of those magnitudes for patch stars using the structures already in place. Thus all magnitudes in use anywhere in the system past the PT processing are SDSS magnitudes, and the data model changes required will be minimal. This also concentrates all knowledge of the photometric equations and transformations in one place.

II. THE PHOTOMETRIC EQUATIONS

The changes to the photometric code are not trivial–both because of the paradigm change and because of a few errors of principle in the way photometry is being done now. This section presents the new photometric equations, introduces the necessary conventions, and indicates the necessary flow of processing.

The fundamental equations connecting the count rates C_{up}, C_{gp}, etc, for the PT and the magnitudes are the usual set:

$$\begin{aligned}
u' &= -2.5\log C_{up} - a_{up} - k_u(t)X - [c_u(X - X_0) + b_{up}]\,[(u' - g') - Z_{up}] \\
g' &= -2.5\log C_{gp} - a_{gp} - k_g(t)X - [c_g(X - X_0) + b_{gp}]\,[(g' - r') - Z_{gp}] \\
r' &= -2.5\log C_{rp} - a_{rp} - k_r(t)X - [c_r(X - X_0) + b_{rp}]\,[(r' - i') - Z_{rp}] \\
i' &= -2.5\log C_{ip} - a_{ip} - k_i(t)X - [c_i(X - X_0) + b_{ip}]\,[(r' - i') - Z_{ip}] \\
z' &= -2.5\log C_{zp} - a_{zp} - k_z(t)X - [c_z(X - X_0) + b_{zp}]\,[(i' - z') - Z_{zp}]
\end{aligned} \tag{1}$$

Here X is the airmass, X_0 the median airmass for which the system is defined, 1.3 in our case, and Z_{up}, Z_{gp}, etc, are the *crossing colors*, suitably defined average colors for which the natural *magnitudes* of the measuring telescope are defined to agree with stars on the standard system. These need to be near the mean color of the standards, but *must be defined once and for all.* We discuss their determination below. The colors used in the b-terms we will call *transforming colors.* Note that both r' and i' use $r' - i'$ as their transforming color, but in general the crossing color appropriate for the r' equation need not be exactly the same as the crossing color for the i' equation. Thus in the system of equations above, a star of the crossing color and canonical airmass has no color-airmass correction at all, and no b-term correction, which corresponds to the transformation from the natural PT system to the USNO system. It is useful to think of the equations in two steps, without the b terms, which define a system which we might call u_p, g_p, etc, and then transform that to the USNO system with equations like

$$u_p = u' + b_u\,[(u' - g') - (u' - g')_0]\,. \tag{2}$$

One can do this decomposition more formally, and determine the intermediate PT natural magnitudes, but the evidence is that the system has been time- variable in the past, which means time-variable b terms, and reporting magnitudes on this 'system' is not likely to be useful. In any case, note that the color terms are dealt with correctly, in terms of the measured final colors. The present system uses instrumental magnitudes, and since we depend on the constancy of the color terms, is liable to gross errors accompanying changes in filters (such as we had last July) or other mechanical or optical changes. The system here depends on some simple linear algebra to solve for unknowns, but is correct. It *does* depend on the code doing sensible things when a measurement is missing, but that is relatively easy and is addressed below.

The choice of the transforming colors is somewhat arbitrary–indeed, there is evidence that $i' - z'$ might be the better color to use both for i' and z', instead of using $r' - i'$ for i'. In the case of the PT to USNO map, it does not matter very much, because the primary standard exposures are such that reasonably good S/N is achieved for all colors. For the patches, however, u'and z' are to be avoided, because it is common for stars to have very good S/N in g' r' and i' and be very poorly or not at all measured in u'and z'. Clearly it is OK to use $u' - g'$ for transforming u' and $i' - z'$ for transforming z', but *not* necessarily a good idea to use $i' - z'$ for transforming i'.

In any case, as important as the crossing colors, if not more so, is the color *range*, which must be determined for each color and used uniformly–either always in an *inclusive* sense, in which case a star is accepted or rejected for a given magnitude calibration if the corresponding transforming color is within some fixed range, or in an *exclusive* sense, in which a star is accepted or rejected for calibration of *all* magnitudes only if it passes *all* color cuts. The latter approach is much the safer, as it excludes strange stars of strange color which happen to have one normal color. We will define the ranges and crossing colors in the next paper.

Before outlining how these are used in the case of standards and unknowns, we call attention to a few things implicit in the notation which hide several important assumptions. Notice in the above equations that the extinction coefficients k_x and color-extinction coefficiens c_x do not have a 'p' in their subscript–we assume that the extinction coefficients measured by the PT (or, in the case of the c terms, calculated from models) are correct for all three systems–the PT, USNO, and SDSS. This is certainly not true in detail because the response functions are slightly different, but the changes are expected to be sufficiently small that they can be ignored; the expected changes can in any case easily be calculated.

So how are these used? One has a set of count rates C_{xp} for a set of standard stars of known u', g', ... z' and airmasses. One has a set, we will assume now, of known b terms and calculated c terms (which we may ignore).

Then the unknown coefficients a_{xp} and $k_x(t)$, $x = u'$, g', etc, can be solved for by standard least-squares techniques, provided that a form for $k_x(t)$ is specified with reasonable parameters. It has been noted that on good nights the extinction terms vary only slowly and, to sufficient accuracy, linearly. They nearly always decrease and the decrease is plausibly associated with the settling of aerosols stirred up during the day. The instrumental zero points, the a terms, should be accurately constant on the timescale of one night, and we will assume henceforth that *all* the variation is in the extinction. Thus for most purposes one can solve for a constant a_{xp} and the two coefficients in

$$k_x(t) = k_{xo} + \dot{k}_x(t - t_0). \tag{3}$$

This is in contrast to the current scheme of evaluating the extinction in blocks during the night and interpolating between blocks. It may be that something more complex than a linear variation is necessary, though on the basis of experience to date that seems unlikely; the data model should incorporate some somewhat more general scheme, perhaps a couple more polynomial coeffients, for the sake of flexibility. In any case, it is crucial to realize that very fast changes of extinction of necessity correspond to nonphotometric conditions, both because of the difficulty of following the changes and the fact that these fast changes *imply* anisotropy of the extinction, given that the extincting stuff is borne by the wind. The exact criterion depends on the height of the principle opacity and the wind speed, but reasonable numbers imply that substantial changes with timescales of much less than an hour almost certainly imply seriously variable photometric conditions over the sky.

If the b terms are not known a large enough block of data needs to be reduced that they can be fit for. The analysis of the PT data for the fall of 2000 and spring of 2001 indicate that though the b terms have a very large variation they are reasonably coherent in time, and fits for them should be possible which will compensate for most of the variation. It is almost certain that the c terms cannot be fit for, and if the inclusion of calculated values does not improve the fits significantly they should be turned off.

Note that the equations have the form

$$\{I + [(X - X_0)c + b]Q\}\,\mathbf{m} = [(X - X_0)c + b]\mathbf{Z} + \mathbf{i} - \mathbf{a} - [\mathbf{k_0} + \dot{\mathbf{k}}(t - t_0)]X \qquad (4a)$$

or

$$\mathbf{a} + [\mathbf{k_0} + \dot{\mathbf{k}}(t - t_0)]X = \mathbf{i} - \mathbf{m} - [(X - X_0)c + b](Q\mathbf{m} - \mathbf{Z}) \qquad (4b)$$

or

$$\mathbf{a} + [\mathbf{k_0} + \dot{\mathbf{k}}(t - t_0)]X + b(Q\mathbf{m} - \mathbf{Z}) = \mathbf{i} - \mathbf{m} - (X - X_0)c(Q\mathbf{m} - \mathbf{Z}) \qquad (4c)$$

where $\mathbf{m}$ is the vector of USNO magnitudes (u',g', etc), $\mathbf{i}$ the corresponding vector of instrumental magnitudes (just the $-2.5\log C_x$ terms above), $\mathbf{a}$ the vector of zeropoints, $\mathbf{k}$ the vector of extinction coefficients, and $\mathbf{Z}$ the vector of crossing colors. c and b are diagonal matrices whose elements are just the c_x and b_x terms; the elements are small, at most a few hundredths. The matrix Q produces colors from the vector of magnitudes, has elements all of which are 1 or -1 and has at most diagonal elements and elements one removed from the diagonal; its elements sum to zero.

These three trivial rearragements of the equations facilitate understanding the three ways they are used; namely solving for unknown magnitudes, which simply involves inverting the matrix on the left hand side of (4a) for a single star, and the two flavors of standards problems, solving by least squares for the fifteen zeropoints and extinction terms from the standards observed on a given night, and solving for those for each night in a block and the (presumed constant for the block) b-terms for a block of nights, (15*N + 5) parameters, where N is the number of nights in the block. These are best thought of by the arrangements in (4b) and (4c), respectively. In each case the unknowns in the problem are all on the left-hand sides of the equations.

For the *patch* stars (we would like to call these stars the *secondary standards*), which are unknowns for the PT and standards for the 2.5-meter, one solves (4) for $\mathbf{m}$, which is a straightforward linear problem solved by inverting the almost-identity matrix on the left or by almost instantaneous iteration if one is too lazy to do algebra. The only caveat is to remove the c and b coefficients connecting to any missing element in $\mathbf{i}$. This is equivalent to setting any missing color to the corresponding crossing color; if one wants to do better, the star can be placed on the mean color-color diagram using measurements which *are* good and the missing color replaced by the mean determined from the others. In detail, this is done by setting the b and c terms to zero, solving for the magnitudes, inferring the missing one (surely if two or more are missing, it is best to discard away the star) from the color-color relation, and solving again using that guess for the magnitude and hence colors. Whether even this is necessary or desirable depends entirely upon how often this situation occurs. Not that it is *not* necessary to do anything fancy if a good measurement is missing for u' or z'; any subset of the equations leaving out either the first or the last or both is closed; the solution for the other bands does not depend on measurements in u' or z'.

Thus the PT maps itself onto the USNO system and produces USNO magnitudes for the patch stars, over some suitable color range for both standards and patch stars, and the ranges should for the sake of consistency and algebraic stability be the same; in any case, the magnitude ranges for the patch stars cannot be *larger* than for the standards; this would involve an extrapolation of the transformations.

MTPipe-exCal-kali needs to output these USNO magnitudes, but the only consumer of these outputs in the new scheme is internal checking in this system. The magnitudes passed to Photo for its crude calibration and to NFCALIB for the final calibration should be SDSS magnitudes. We assume that the SDSS system and the USNO system are related (again over the same restricted range in color) by simple linear relationships (though the inclusion of more complex forms is trivial):

$$\begin{aligned}
u &= u' - b_u\left[(u' - g') - Z_u\right] + \delta_u \\
g &= g' - b_g\left[(g' - r') - Z_g\right] + \delta_g \\
r &= r' - b_r\left[(r' - i') - Z_r\right] + \delta_r \\
i &= i' - b_i\left[(r' - i') - Z_i\right] + \delta_i \\
z &= z' - b_z\left[(i' - z') - Z_z\right] + \delta_z
\end{aligned} \tag{5}$$

Note that here the crossing colors *may be different* from those in the PT to USNO transformation; we address this issue below; there is not really a satisfactory way to deal with this. In any case once *these* b-terms (which, of course, are very different from the b_{xp} terms for the PT–USNO transformation) are determined it is straightforward to produce SDSS magnitudes for these outputs. The zero offsets δ_x are for the sake of target selection and will be discussed below. If the crossing colors have been chosen well, they will be small, but perhaps not negligibly so. In principle they can be made zero by proper choice of the crossing colors, but this seems unwise; the crossing colors should be the overall mean colors of the transforming stars, as we will see below.

There is at this point a very important observation one must make, The transformations of equation 5, (and indeed for equation 2) are *only* applied to standard stars in a limited range of color, and are most emphatically *not* general transformation equations between the two systems in any average, compromise sense. They are in this scheme never applied to anything else, nor need they ever be within the context of the survey. If quasars or gamma-ray bursts are observed with the PT, those observations must be understood in the context of some *ad hoc* natural system for the PT, which for those objects may or may not be related to the USNO system by the linear transformations and b-terms derived for the uses outlined here.

III. CALIBRATING THE 2.5-METER

The photometric equations for the 2.5-meter are by assumption very simple:

$$\begin{aligned} u &= -2.5 \log C_u - a_u - k_u(t)X \\ g &= -2.5 \log C_g - a_g - k_g(t)X \\ r &= -2.5 \log C_r - a_r - k_r(t)X \\ i &= -2.5 \log C_i - a_i - k_i(t)X \\ z &= -2.5 \log C_z - a_z - k_z(t)X \end{aligned} \tag{6}$$

and the whole job of calibration reduces to the determination of the zeropoints a_x, given that the extinctions are determined by the PT. Note that we explicitly ignore the c terms, which would mix images in the way we wish to avoid, and assume that the standards are on the SDSS system so there are no b terms. The omission of the c terms inevitably leads to small errors, since the photometric *system* is different from the canonical one if the airmass X is different from the defining one X_0. We are reasonably confident that the zero points of the 2.5-meter are at least as stable as those for the PT (and much better than the PT was before this summer), so the only issues are the variation of the extinction terms with time, possible mismatch of the extinction terms between the USNO system and the SDSS system, and systematic inaccuracies in the patch magnitudes. These issues and how best to deal with them to produce a globally correct calibration are being considered and will be dealt with elsewhere. It is important to do the determination of the zero points correctly insofar as the use of the correct statistical weights is concerned, and this should be reviewed.

The b-terms in (5) can be determined by direct fitting in NFCALIB or iteratively looking at the regressions of 2.5-meter natural magnitudes on the colors of the input patch stars. They are expected to be stable; the USNO system appears to be homogeneous, and the response functions of the camera are extremely stable based on measurements over the years, so they need to be determined once and checked occasionally.

IV. DETERMINING THE CROSSING COLORS AND THE ZERO-POINT SHIFTS

The determination of the crossing colors and the δ_x in equation (5) is critical to another aspect of the SDSS photometry. We will be making this change in midstream, and while we have not had a proper system at all heretofore, we wish to perturb the magnitudes we produce as little as possible. We have until now been connecting disparate systems in some ill-defined mean sense, that sense clearly depending on the particular collection of stars used to tie data sets together. We would like to choose the crossing colors and the zero points so as to perturb our definitions as little as possible. The considerations are a little different for the two quantities.

Since reductions heretofore have not used b-terms at all, and hence made the implicit but incorrect assumption that the systems are identical, one must ask what one must do in order that the new reductions will best resemble the old. This is important in order that target selection not break.

Consider first the PT to USNO transformation. For a typical color, say g', the equation from (1) looks like

$$g' = -2.5 \log C_{gp} - a_{gp} - k_g X - [c_g(X - X_0) + b_{gp}]\,[(g' - r') - Z_{gp}]$$

If one takes the difference of this equation and one which is identical except that the b term is missing and averages the difference over a large number of stars, one obtains

$$\langle \delta g' \rangle = b_{gp}\,[\langle (g' - r') \rangle - Z_{gp}] \tag{7}$$

assuming that the zeropoints have been calculated correctly for the two samples. Thus the *crossing colors must be set to the mean colors of the standards used for the calibration.* If that set is different for the calibration of different colors, then the crossing colors will be different, even if the calibration *color* is the same. If this is done, and the standards used for a given night accurately reflect the population of standards used in the survey, the determination of the zero points a is independent of the correctness of the b terms. If the b terms are incorrect, however, the magnitudes of the *patch* stars will be wrong except at the crossing colors.

What about the USNO to SDSS transformation? One can make exactly the same argument to indicate that the crossing colors should be the mean colors of the patch stars. But here one must be a little careful. In the case of the primaries, the exposure times are adjusted so that the S/N of the observations is comparable and in any case high enough that external errors such as transparency variations, flat-fielding errors, etc, probably dominate the error budget, so that the errors are assumed to be and probably are the same for all primaries. This is most emphatically not the case for the patch stars, and they are properly weighted in NFCALIB. If there are nonuniform weights, the proper mean color must be the *weighted* mean color. This must be computed for a typical set. It is our understanding that the patch stars have a redder mean, which is understandable, than the standards.

This backs us up into an unfortunate corner. In order that the old reductions without b terms be correct in the mean, at least in the establishment of zero points, and reasonably robust to sample changes, the crossing colors must be determined as above.... but if the crossing colors for the USNO to SDSS transformation are different from the crossing colors for the PT to USNO transformation, there will be a *systematic* error which is of the order of

$$err = b_x * (Z_p x - Z_x)$$

This can be minimized by the (necessary) step of restricting the color ranges for the standards such that they are identical for the primaries and the patch stars–after all, the transformations to USNO are only correct for that color range. One could think of *further* restricting the red edges of the patch star colors to bring the mean colors into agreement, or sparse-sampling the red patch stars to do the same. If neither of these are done, we expect the δ_x in equation (5) to be nonzero (though not necessarily significant).

They can and must be determined in any case by taking a set of runs, reducing them with the old code and the new, and making the magnitudes of *targeted* objects agree in the mean. That it is only the determination of the zeropoints which is affected is clear, because the old calibrations for the 2.5-meter also do not use b-terms, and so the 'system' is the 2.5-meter natural system but with unsatisfactorily determined zeropoints.

V. SOME IMPORTANT CONSIDERATIONS FOR THE NEW SYSTEM

The photometric equations discussed above are the same as have been used for a very long time, certainly since the pioneering days of Johnson and Eggen. It is important to note, however, that what we are proposing is basically a new paradigm for broad-band photometry appropriate to a single stable instrument producing a huge homogeneous photometric data set, in contrast to the usual attempt to set up a system of broad astrophysical applicability. This system is most emphatically *not* reproducible in detail on other instruments. It is defined by a set of standards measured on one system and *transformed* to the new system, stars which are themselves probably never to be measured by the SDSS telescope, and stars which span a fairly narrow color range in order to facilitate the transformation. The standards are used *only* to set zero points for a system whose detailed response is defined by the survey camera, and so do not define the photometric system in at all the same way, as, say, the Johnson standards define the UBV system, or, indeed the USNO standards define the USNO system. In a very real way, the SDSS system is defined in the end by the 10^8 magnitudes the survey will produce. The secondary standards are simply a transfer of the transformed USNO standards to fainter stars, and do not predict *at all* any information about colors of objects whose nature differs from that set of stars. We may think about this either in the "theoretical" plane, in which the system is defined (but "can be understood" is better than "defined") by Mamoru Doi's detailed response measurements and the magnitudes on the SDSS system of spectrophotometrically defined objects such as +17°4708, which define the (small) departures from a real AB system, or the "observational" one, in which the system is defined by the ensemble of observations of all objects, but in particular spectrophotometrically understood ones, even if that understanding is an ensemble one, such as it is for the collection of quasars. Our goal is twofold: To ensure that the system is homogeneous, so that if an object could be moved from one part of the sky to another and its flux changed overall by some factor, all the magnitudes would change by -2.5 times the log of that factor, and that *from* spectrophotometry one can accurately predict the measured magnitudes. The former is a goal of all broadband photometry. The latter has been attempted for existing systems with varying and generally not very good success because one does not *ever* really know the response functions and the measurements made with some hardware have been transformed to a standard system defined with quite different hardware. We are in the enviable position of doing a very large survey with one instrument at one site, which defines (we hope but have every reason to believe) a very stable natural photometric system. It can be approximately used by anyone anywhere in precisely the same way as UBV is used, but that is *not* our primary concern; doing a good job internally is.

It *must* be noted, however, that knowing the response functions is key to using the system on other hardware; the standard star list is too restricted in color to define the system. We may well at some point wish to choose a set of bright stars (for the camera – say 14-18 magnitude) which we bless as standards and which could in principle be used to define the system for external use. This network would be of great utility to future larger and fainter surveys such as those to be undertaken by LSST and by the CFH megaprime, the calibration difficulties for which are only now beginning to be understood and appreciated by those concerned.