Determination of Low-Frequency Flat-Field Structure from Photometry of Stellar Fields

Roeland P. van der Marel

September 2003

Abstract

The low-frequency content of ACS flat fields cannot be accurately determined from ground-based calibration data or internal lamp exposures. It must therefore be determined from on-orbit science data. Photometry of a stellar field that is imaged multiple times with different pointing or roll is ideal for this purpose. I describe a mathematical algorithm to use such data to determine the residual low-frequency flat-field structure (L-flat). The L-flat is expanded as a linear sum of two-dimensional basis functions. Magnitude differences between measurements for the same star at different positions on the detector constrain the coefficients of the linear sum. This reduces mathematically to an over-determined linear least-squares problem that can be solved through singular-value decomposition. This yields both the best-fitting L-flat and its formal error. A numerical FORTRAN implementation of the algorithm is presented and its accuracy verified using tests with artificially generated data sets. The software was used by Mack et al. (ACS ISR 02-08) to analyze early ACS WFC and HRC data of the globular cluster 47 Tuc. The inferred fourth-order polynomial L-flat corrections to the ground-based calibration flats were implemented in the ACS pipeline in late 2002. It is shown that newly implemented features in the software should allow further improvements in these L-flats. The methods presented here are generally applicable to other imaging instruments, and are not restricted to use on ACS data.

1. Introduction

Before the launch of ACS a significant amount of effort was spent to obtain laboratory flat fields. Longward of its ~ 3500Å optical cutoff, the Refractive Aberrated Simulator/Hubble Opto-Mechanical-Simulator (RAS/HOMS) was used in 2001 February-March with a continuum light source to produce flat fields which include both the low frequency L-flat and the high frequency pixel-to-pixel P-flat structure (Bohlin, Hartig & Martel 2001). The hope was that these LP-flats could be used for accurate calibration of on-orbit data. However, early SMOV data, both of individual spectro-photometric standards and stellar fields, suggested the presence of residual low-frequency flat-field structure, even after full pipeline calibration with the laboratory flats and correction for geometric distortion using Pydrizzle (as described in the ACS Data Handbook; Mack et al. 2002a).

There are two obvious approaches to quantify and correct for residual L-flat structure. The first approach is to use deep images of "empty" sky fields. After full reduction such images ought to be flat. Any deviations from flatness can be used directly to improve the pipeline flat fields. The second approach is to use point-source photometry, as described below. I focus here on the latter approach, in part because early in the ACS mission point-source photometry of well-chosen stellar fields is more readily available in different filters than deep sky observations. Nonetheless, it is definitely important to also construct sky-flats in as many filters as possible, and to check for consistency with the results derived from point-source photometry. Such efforts are currently in progress.

Residual L-flat structure causes stars in fully reduced images to have slightly different magnitudes when placed at different positions on the detector. I describe mathematically how this information can be used to determine the positional dependence of the residual L-flat structure. A FORTRAN implementation of the method is presented, and the software is tested on artificially generated data sets to demonstrate its accuracy and effectiveness. The application to real data and the creation of flat fields for use in the ACS pipeline was described previously in a separate ISR (Mack et al. 2002b). I describe newly implemented features of the software which should allow further refinements of the pipeline flats.

A post-hoc investigation into the scarce literature on this subject showed that the basic ideas underlying the discussion presented here were described previously in the context of other instruments or observational scenarios (Greenfield 1994; Manfroid 1995, 1996; Wild 1997; Manfroid, Selman & Jones 2001). The present paper adds to this existing body of work by providing a detailed description of the underlying mathematics and implied error analysis, a software implementation that is tested on artificial data to assess the accuracy of the method and the reliability of the error estimates, and a description of practical

applications to ACS data.

2. Mathematical Description

2.1. Problem Statement

Consider a star i, with unknown magnitude m_i. Here and henceforth, all magnitudes are assumed to be calibrated to some arbitrary photometric system. The star is observed $N_i \geq 2$ times. The telescope pointing is dithered or rotated between observations, so that in each observation the star falls on a different position (x_{ij}, y_{ij}) of a two-dimensional detector, where $j = 1, \ldots, N_i$. For each pointing, aperture photometry is performed on the fully reduced data. This yields observed stellar magnitudes o_{ij} with formal random errors e_{ij}. If a stellar field is observed, then these measurements are available for many different stars, $i = 1, \ldots, S$.

If the data were properly flat fielded, then each o_{ij} is an unbiased estimate of the unknown magnitude m_i. Thus:

$$m_i = o_{ij} \pm e_{ij}, \qquad (i = 1, \ldots, S \quad \text{and} \quad j = 1, \ldots, N_i). \tag{1}$$

For each star, the optimal estimate of m_i is then simply the weighted average of the measurements,

$$m_i = \left[\sum_{j=1}^{N_i} o_{ij}/e_{ij}^2\right] \bigg/ \left[\sum_{j=1}^{N_i} 1/e_{ij}^2\right], \qquad (i = 1, \ldots, S), \tag{2}$$

with formal random error on the weighted average

$$\Delta m_i = \left[\sum_{j=1}^{N_i} 1/e_{ij}^2\right]^{-1/2}, \qquad (i = 1, \ldots, S). \tag{3}$$

By contrast, if the data were not properly flat fielded, then equation (1) acquires an additional term that depends on the position on the chip:

$$m_i + R(x_{ij}, y_{ij}) = o_{ij} \pm e_{ij}, \qquad (i = 1, \ldots, S \quad \text{and} \quad j = 1, \ldots, N_i). \tag{4}$$

The unknown function $R(x, y)$ enters into equation (4) in an additive sense because flat fielding is a multiplicative operation whereas the magnitude scale is a logarithmic one. This is why stellar brightnesses in this context are best expressed in magnitudes, rather than intensities (counts) or fluxes (counts/sec). It yields a problem that is mathematically more easily tractable. If it is assumed that the flat field that was applied in the data reduction

is incorrect only in terms of its low-frequency content, then it makes sense to expand the function $R(x, y)$ into a linear sum of K two-dimensional basis functions $R_k(x, y)$:

$$R(x, y) = \sum_{k=1}^{K} a_k R_k(x, y), \qquad (k = 1, \ldots, K). \tag{5}$$

The basis functions are assumed to be known, and can be chosen to be polynomials or any other convenient functional form (see Section 2.4 below). The function $R(x, y)$ will henceforth be referred to as the *L-flat*. It is expressed in magnitudes. (Nothing specific is assumed about the frequency content of the L-flat, so in principle, by using a very large number of basis functions, the algorithm can determine variations at any arbitrary frequency.) The L-flat, as defined here, does not describe a property of the actual flat field. Instead, it measures the residual structure with respect to the pipeline flat field that was used to calibrate the data.

The coefficients a_k that characterize the L-flat remain to be determined. Substitution of equation (5) into equation (4) yields

$$m_i + \sum_{k=1}^{K} a_k r_{ijk} = o_{ij} \pm e_{ij}, \qquad (i = 1, \ldots, S \quad \text{and} \quad j = 1, \ldots, N_i), \tag{6}$$

where the scalars r_{ijk} are defined as

$$r_{ijk} \equiv R_k(x_{ij}, y_{ij}), \qquad (i = 1, \ldots, S \quad , \quad j = 1, \ldots, N_i \quad \text{and} \quad k = 1, \ldots, K). \tag{7}$$

The data can only determine the L-flat $R(x, y)$ up to an arbitrary additive constant. The choice of the constant (i.e., the normalization of the flat field) is up to the user, although with any given choice the zeropoint of the photometric magnitude system becomes fixed. Without loss of generality, it is imposed here that the function $R(x, y)$ averages to zero over the detector, to within an accuracy ϵ:

$$\frac{1}{P} \sum_{p=1}^{P} R(x_p, y_p) = \sum_{k=1}^{K} a_k \left[\frac{1}{P} \sum_{p=1}^{P} R_k(x_p, y_p) \right] = 0 \pm \epsilon. \tag{8}$$

The P points (x_p, y_p) are chosen to uniformly sample the detector (e.g., all the pixels on the detector, or a subsampled subset of them). The accuracy ϵ can in theory be set to zero, but for numerical reasons it is convenient to use a small but non-zero number, e.g., $\epsilon = 10^{-6}$.

2.2. Problem Solution

Division of equations (6) and (8) by their respective errors yields

$$(1/e_{ij})m_i + \sum_{k=1}^{K} a_k(r_{ijk}/e_{ij}) = (o_{ij}/e_{ij}), \qquad (i = 1, \ldots, S \quad \text{and} \quad j = 1, \ldots, N_i), \tag{9}$$

and

$$\sum_{k=1}^{K} a_k \left[\frac{1}{P\epsilon} \sum_{p=1}^{P} R_k(x_p, y_p) \right] = 0, \tag{10}$$

with a formal random error of unity for each equation, and for all i and j. This set of equations can be written as a linear matrix equation,

$$\mathbf{A}\vec{x} = \vec{b}. \tag{11}$$

The column vector $\vec{x}$ contains the unknowns. It is defined as

$$\vec{x} \equiv (m_1, \ldots, m_S, a_1, \ldots, a_K) \tag{12}$$

and has dimension $L \equiv S + K$. The column vector $\vec{b}$ contains the constraints (the observables and the adopted normalization), and is defined as

$$\vec{b} \equiv (o_{11}/e_{11}, o_{12}/e_{12}, \ldots, o_{SN_S}/e_{SN_S}, 0). \tag{13}$$

It has dimension

$$M \equiv \left(\sum_{i=1}^{S} N_i \right) + 1. \tag{14}$$

If the number of measurements N_i is the same for each star, $N_i = N$, then $M = NS + 1$. The matrix $\mathbf{A}$ has M rows and L columns. The number of unknowns L is typically much smaller than the number of constraints M and the matrix equation (11) is therefore over-determined. The coefficients of the matrix follow from equations (9)–(13). Many of the coefficients are zero, and the matrix is therefore sparse.

The formal random errors in equations (6) and (8) are uncorrelated and are approximately normally distributed. The maximum likelihood fit for the unknown quantities in the vector $\vec{x}$ is therefore the combination that minimizes the χ^2 quantity

$$\chi^2 = ||\mathbf{A}\vec{x} - \vec{b}||, \tag{15}$$

where the right hand side is the Euclidean norm of the vector $\mathbf{A}\vec{x} - \vec{b}$ in M-dimensional vector space. The vector $\vec{x}$ that minimizes this norm is called the least-squares solution of the matrix equation (11). One way to find the least-squares solution is to solve the so-called normal equations ($\mathbf{A}^T\mathbf{A}\vec{x} = \mathbf{A}^T\vec{b}$). However, it is numerically more robust to use the singular value decomposition (SVD) of the matrix $\mathbf{A}$ (e.g., Press et al. 1992). Linear algebraic theory shows that $\mathbf{A}$ can be written as the product of an $M \times L$ column-orthogonal matrix $\mathbf{U}$, an $L \times L$ diagonal matrix $\mathbf{W}$ with positive or zero elements, and the transpose of

an $L \times L$ orthogonal matrix $\mathbf{V}$. Many numerical routines exist to determine these matrices. Once they have been found, the elements x_l of the solution vector follow from

$$x_l = \sum_{i=1}^{M} \sum_{j=1}^{L} V_{lj} U_{ij} b_i / W_{jj}, \qquad (l = 1, \ldots, L). \tag{16}$$

The variances (squared formal errors) of the unknowns are

$$(\Delta x_l)^2 = \sum_{j=1}^{L} (V_{lj}/W_{jj})^2, \qquad (l = 1, \ldots, L), \tag{17}$$

and the covariances between the unknowns are

$$\mathrm{Cov}(x_l, x_k) = \sum_{j=1}^{L} V_{lj} V_{kj} / W_{jj}^2, \qquad (l = 1, \ldots, L \quad \text{and} \quad k = 1, \ldots, L). \tag{18}$$

Values of W_{jj} near zero identify singularities in the matrix problem (hence the name "singular value decomposition"). As discussed in Press et al. (1992), the appropriate thing to do for such W_{jj} is to set $1/W_{jj}$ to zero in the above equations (not infinity!). In practice one can use the criterion $W_{jj} \leq \zeta W_{\max}$, e.g., with $\zeta = 10^{-10}$, where $W_{\max}$ is the maximum of all the W_{jj}. This is equivalent to throwing away those linear combinations of constraints that affect χ^2 only at the level of the round-off error.

The solution vector $\vec{x}$ contains the coefficients a_k ($k = 1, \ldots, K$), which allows the L-flat $R(x, y)$ to be calculated at any position (x, y) using equation (5). The variance (squared formal error) of the L-flat at any position (x, y) follows upon application of the standard formulae for error propagation:

$$[\Delta R(x, y)]^2 = \sum_{i=1}^{K} \sum_{j=1}^{K} R_i(x, y) R_j(x, y) \mathrm{Cov}(a_i, a_j). \tag{19}$$

These errors measure only the propagated influence of the formal random errors e_{ij} in the stellar magnitude measurements. The same is true for the variances and covariances in equations (17) and (18). Of course, in real applications the quality of the solution is often dominated by systematic errors (e.g., the quality of the spatially dependent aperture corrections used in the stellar magnitude determinations, or the quality with which the basis functions can reproduce the true L-flat).

2.3. Alternative Approach

The CPU requirements for finding the least-squares solution of a matrix problem scale as L^3, where L is the number of unknowns. It is therefore inefficient that equation (11)

treats all the stellar magnitudes m_i $(i = 1, \ldots, S)$ as unknowns that must be solved for. After all, our primary interest is only in the coefficients a_k $(k = 1, \ldots, K)$ that characterize the L-flat. An alternative way to phrase the problem is therefore to consider observed magnitude *differences* as the fundamental constraints, and not the observed magnitudes themselves. To this end one can take equation (6) and subtract for each star the first measurement from all the other measurements. This yields:

$$\sum_{k=1}^{K} a_k(r_{ijk} - r_{i1k}) = d_{ij} \pm \delta_{ij}, \qquad (i = 1, \ldots, S \quad \text{and} \quad j = 2, \ldots, N_i), \qquad (20)$$

where

$$d_{ij} \equiv o_{ij} - o_{i1}, \qquad \delta_{ij} \equiv (e_{ij}^2 + e_{1j}^2)^{1/2}, \qquad (i = 1, \ldots, S \quad \text{and} \quad j = 2, \ldots, N_i). \qquad (21)$$

The errors e_{ij} and e_{1j} add in quadrature to form δ_{ij}. Division of equation (20) by δ_{ij} yields

$$\sum_{k=1}^{K} a_k[(r_{ijk} - r_{i1k})/\delta_{ij}] = (d_{ij}/\delta_{ij}), \qquad (i = 1, \ldots, S \quad \text{and} \quad j = 2, \ldots, N_i), \qquad (22)$$

with a formal random error equal to unity for all i and j. This set of equations, combined with the equation (10) that sets the flat-field normalization, yields another linear matrix equation:

$$\mathbf{B}\vec{y} = \vec{c}. \qquad (23)$$

The unknowns are now contained in the column vector $\vec{y}$. It is defined as

$$\vec{y} \equiv (a_1, \ldots, a_K) \qquad (24)$$

and has dimension K. The column vector $\vec{c}$ contains the constraints and is defined as

$$\vec{c} \equiv (d_{12}/\delta_{12}, d_{13}/\delta_{13}, \ldots, d_{SN_S}/\delta_{SN_S}, 0). \qquad (25)$$

It has dimension $M - S$. If the number of measurements N_i is the same for each star, $N_i = N$, then $M = (N - 1)S$. The matrix $\mathbf{B}$ has $M - S$ rows and K columns. The number of unknowns K is typically much smaller than the number of constraints $M - S$ and the matrix equation (23) is therefore over-determined. The coefficients of the matrix follow from equations (10) and (22)–(25). The matrix $\mathbf{B}$ is not sparse because none of its elements are necessarily zero.

The least-squares solution of the matrix equation (11) can be obtained through singular value decomposition in exactly the same way as described in Section 2.2. This yields the coefficients a_k $(k = 1, \ldots, K)$ and their covariances. With equations (5) and (19) these yield

the L-flat and its formal error for any position (x, y). The stellar magnitudes m_i are not obtained directly from the solution of the matrix equation. However, they can be estimated by correcting the observed magnitudes o_{ij} for the L-flat using equation (6), and taking the weighted average of the results:

$$m_i = \left[\sum_{j=1}^{N_i} \left[o_{ij} - \sum_{k=1}^{K} a_k r_{ijk} \right] \Big/ e_{ij}^2 \right] \Big/ \left[\sum_{j=1}^{N_i} 1/e_{ij}^2 \right], \qquad (i = 1, \ldots, S). \tag{26}$$

The errors in the L-flat are generally much smaller than the errors in the individual stellar magnitudes from which it was estimated. The formal error Δm_i can therefore be approximated by equation (3), which assumes that errors in the L-flat determination can be neglected.

Henceforth, the approach based on equation (11) is referred to as the direct matrix method; the approach based on equation (23) is referred to as the differential matrix method. The differential matrix method has some minor drawbacks from a mathematical viewpoint. The different observations for a star are not treated symmetrically. The first observation has a special place and is used as a reference to compare the other observations to. If the first observation has a large error e_{i1} then this propagates into all constraints through the definition of δ_{ij}. To avoid this, it is best to reorder the measurements for each star, without loss of generality, such that the first one has the smallest error: $e_{i1} \leq e_{ij}$ $(j = 2, \ldots, N_i)$.[1] An added complication from the special treatment of the first observation is that the errors in the different constraints for a given star in equation (20) cease to be uncorrelated. As a result, the least-squares solution of the matrix equation $\mathbf{B}\vec{y} = \vec{c}$ is not mathematically equivalent to the maximum likelihood solution. However, it should be very close. Measurements for different stars continue to have uncorrelated errors, and the number of different stars S is generally much larger than the number of different observations per star N_i. So the direct method is mathematically slightly more robust, but the differential method is much quicker to solve due to its lower dimensionality.

---

[1]Instead of subtracting the observation with the smallest formal error, one could also subtract the weighted average magnitude of the observations, as given by equation (2). This would have the advantage of treating all observations symmetrically. This was not explored further, given that the method described in the text already seems to work well enough (see Section 4).

2.4. Choice of Basis Functions

To approximate the L-flat with a two-dimensional polynomial of order Z I use two-dimensional basis functions R_k $(k = 1, \ldots, K)$ of the form:

$$R_k(x, y) = P_q(x')P_r(y'), \qquad k = q(Z+1) + r + 1, \qquad (q = 0, \ldots, Z \quad \text{and} \quad r = 0, \ldots, Z). \tag{27}$$

The total number of basis functions is $K = (Z + 1)^2$. The function P_n is the Legendre polynomial of order n. The quantities (x', y') are obtained from the detector coordinates (x, y) upon remapping to the interval $(-1, 1)$ using the equations

$$x' \equiv (2x/X) - 1, \qquad y' \equiv (2y/Y) - 1, \tag{28}$$

where $X \times Y$ is the size of the detector. With this choice of basis functions, the L-flat $R(x, y)$ given by equation (5) becomes an arbitrary polynomial composed of terms of the forms $x^q y^r$ with both q and r non-negative and $\leq Z$. One could also use $R_k(x, y) = x^q y^r$ as basis functions to describe the same function space, but the Legendre polynomials are numerically to be preferred due to their functional orthogonality. This orthogonality implies that all basis functions average to zero over the detector, with the exception of $R_1(x, y) = P_0(x')P_0(y') = 1$. The normalization constraint given by equation (8) therefore forces $a_1 = 0$ in the expansion of the L-flat (equation [5]) and has no effect on the other a_k $(k = 2, \ldots, K)$.

A disadvantage of polynomials is that they cannot describe L-flat variations on intermediate frequencies unless one goes to fairly high orders Z. However, polynomial fits with high orders can have disadvantages such as unphysically large excursions near the detector edges. Other sets of basis functions are therefore worth exploring. One alternative basis set has the form

$$R_k(x, y) = \begin{cases} 1, & \text{if } (x, y) \text{ in rectangle } k \text{ ;} \\ 0, & \text{otherwise .} \end{cases} \tag{29}$$

Here the detector area is divided into $K \equiv W^2$ rectangles, equal to the total number of basis functions. Each rectangle has size $[X/W] \times [Y/W]$. For a square detector with $W = 8$ this resembles a chess board. I will therefore refer to equation (29) as the "chess board basis functions". With this basis set, the L-flat $R(x, y)$ given by equation (5) becomes an arbitrary function on a "pixelized" version of the detector space (see Figure 1c below for an example).

3. Software Implementation

The algorithms were implemented into a FORTRAN program. On input, the program reads an ASCII file with stellar IDs, positions on the detector, and magnitude measurements and errors. Very small magnitude errors may not be realistic, and there is the option to add a systematic error floor in quadrature to all the photometric errors. Either all stars can be used in the analysis, or a subset can be chosen. Subsets can be selected randomly, or on the basis of stellar magnitude (e.g., only stars with average observed magnitude in a certain magnitude range) or ID number. Stars can also be removed from the sample, for example if the spread in the observed magnitudes is suspiciously large (indicating a potential problem with one of the measurements, or an intrinsically variable star). The user can select with which algorithm to solve the problem (equation [11] or [23]), and what maximum order and type of basis functions to use. For polynomial basis functions on WFC, separate basis functions are used for both CCDs (i.e., the basis functions are defined to be zero outside the area of the CCD to which they apply). There are usually stars that have been observed on both CCDs, so the relative magnitude offset between the L-flats for the two CCDs is constrained by the data. For the chess board basis functions on WFC, the camera is treated as a single square detector (the subdivision of the detector area into disjunct rectangles automatically causes the two CCDs to be treated as separate entities). The program calculates the relevant matrices and vectors. The singular value decomposition of the matrix is performed using the subroutine `svdcmp` from Numerical Recipes (Press et al. 1992). Once the solution has been found, the residuals of the data with respect to the model fit are calculated:

$$r_{ij} = o_{ij} - \left(m_i + \sum_{k=1}^{K} a_k r_{ijk} \right), \qquad (i = 1, \ldots, S \quad \text{and} \quad j = 1, \ldots, N_i). \tag{30}$$

The error in the residual is simply the error in the measurement, e_{ij}. The χ^2 of the fit to the data is therefore

$$\chi^2 = \sum_{i=1}^{S} \sum_{j=1}^{N_i} (r_{ij}/e_{ij})^2. \tag{31}$$

This can be compared to the number of degrees of freedom of the problem, $N_{\rm df}$, which is the number of constraints minus the number of unknowns. For an excellent fit one expects χ^2 to be in the range $N_{\rm df} \pm \sqrt{2N_{\rm df}}$. If $\chi^2 > N_{\rm df}$, then this may indicate that the errors in the data were underestimated (or alternatively, that the adopted set of basis functions cannot adequately reproduce the true L-flat structure). In view of this, the program offers the option to rescale all the errors by $\sqrt{\chi^2/N_{\rm df}}$. The L-flat determined by the program, as well as its formal error, are written to output images in the native IRAF format. There is an option to subsample the output images to minimize demands on memory and input/output.

The program also calculates and writes a residual image, as well as its formal error. For this, the user specifies a subdivision of the detector area into a chess board pattern (with no logical connection to the basis functions, which might also represent a chess board pattern). For each rectangle k in the pattern, the program finds the measurements for which the stellar position (x_{ij}, y_{ij}) falls inside the rectangle. The average residual for that rectangle is then calculated as

$$\langle r \rangle_k = \left[\sum_{(x_{ij},y_{ij}) \in k} r_{ij}/e_{ij}^2 \right] \Big/ \left[\sum_{(x_{ij},y_{ij}) \in k} 1/e_{ij}^2 \right], \tag{32}$$

with formal random error

$$\Delta \langle r \rangle_k = \left[\sum_{(x_{ij},y_{ij}) \in k} 1/e_{ij}^2 \right]^{-1/2}. \tag{33}$$

The residual image is valuable for visual inspection.[2] If the eye can detect low-frequency structure in the residual image, then it is likely that the set of adopted basis functions was not general enough to capture the full low-frequency content of the true L-flat.

4. Applications

4.1. Artificial Data

The most accurate way to test a newly developed algorithm is to run it on artificially generated data. Since the model from which the data are generated is fully specified in advance, one can check how well the algorithm recovers the known solution.

Artificial data are generated to loosely resemble the data actually obtained for ACS/HRC shortly after launch (e.g., Mack et al. 2002b). However, no attempt is made

---

[2]Residual images can also be used to actually solve the matrix problem through the procedure known as "fixed point iteration". For this one starts with a trial guess for the L-flat (e.g., zero). One then calculates the residual image. Subsequently, one updates the trial guess for the L-flat by adding the residual image (or a fraction thereof, for improved numerical stability). This procedure is then iterated. Convergence is reached when the residual image is numerically consistent with zero. By definition, this implies that the best L-flat solution has been found. The disadvantages of such a procedure are that it need not always converge as planned, and that it generates no formal error estimate. However, it is possible that the solution of the problem could actually be found quicker by iteration than by solving matrix equations. I didn't investigate implementation of an iterative scheme in the present context, but this may be worth considering if computational speed becomes the primary consideration for real applications. Iterative procedures have already been used successfully for the determination of flat fields in spectroscopic applications (e.g., Gilliland 1998, and references therein).

to mimic the actual properties of ACS or the astrophysical properties of a realistic stellar field in detail. The detector size is $X \times Y = 1024 \times 1024$. Magnitudes m_i $(i = 1, \ldots, T)$ are drawn randomly for $T = 2500$ stars, using the magnitude range 16–21 and a luminosity function $N(m) \, dm \propto 10^{-0.3m} \, dm$. Each star i is placed at a position (x_{i1}, y_{i1}), with each coordinate drawn at random from the range $[-300, 1324]$. If the coordinates are both in the range $[1, 1024]$, then this is interpreted to mean that the star falls on the detector in the first of a series of pointings. It is assumed that observations are performed at 8 additional pointings, stepped by either 150 or 300 pixels in the direction of position angles $45°$, $135°$, $225°$, and $315°$, measured with respect to the detector X-axis. This implies for the position of a star in each observation:

$$x_{ik} = x_{i1} + (-1)^l (150n/\sqrt{2}), \qquad y_{ik} = y_{i1} + (-1)^m (150n/\sqrt{2}), \qquad (k = 2, \ldots, 9), \quad (34)$$

where

$$k = 1 + l + 2m + 4n, \qquad (l = 1, 2 \quad , \quad m = 1, 2 \quad \text{and} \quad n = 1, 2). \tag{35}$$

The photometric random error e_{ij} for observation j of star i is specified to be

$$e_{ij} = 0.005 \times 10^{0.2(m_i - 16)}. \tag{36}$$

This ranges from 0.005 mag at $m = 16$ to 0.05 mag at $m = 21$. Magnitude measurements for the observations are assigned according to the prescription

$$o_{ij} = m_i + G(e_{ij}) + R(x_{ij}, y_{ij}), \tag{37}$$

where $G(e_{ij})$ denotes a randomly drawn Gaussian deviate from a distribution with mean zero and dispersion e_{ij}. The function $R(x, y)$ is the L-flat, which I fairly arbitrarily take to be

$$R(x, y) = 0.1 P_1(x') P_3(y') - 0.1 P_2(x') P_4(y'). \tag{38}$$

As before, the function P_n is the Legendre polynomial of order n and (x', y') are the detector coordinates (x, y) remapped to the interval $(-1, 1)$ using equation (28). Because of the orthogonality of the Legendre polynomials (see Section 2.4), the L-flat $R(x, y)$ thus defined is normalized to an average of zero, as required by convention (equation [8]). The L-flat is shown in Figure 1a. It fluctuates between -0.200 to $+0.077$ magnitudes. The artificial data are characterized by the values of o_{ij}, e_{ij}, x_{ij} and y_{ij}. The combinations of $i = 1, \ldots, T$ and $j = 1, \ldots, 9$ for which (x_{ij}, y_{ij}) does not fall on the detector area are removed from the data set. This leaves $S = 1331$ stars for which $3 \leq N_i \leq 9$ artificial measurements are available per star $(i = 1, \ldots, S)$. The artificial data is analyzed with the algorithms of Section 2 to obtain an estimate for $R(x, y)$ of the form given by equation (5). This estimate can then be compared to the actual function $R(x, y)$ in equation (38) that was used to generate the data.
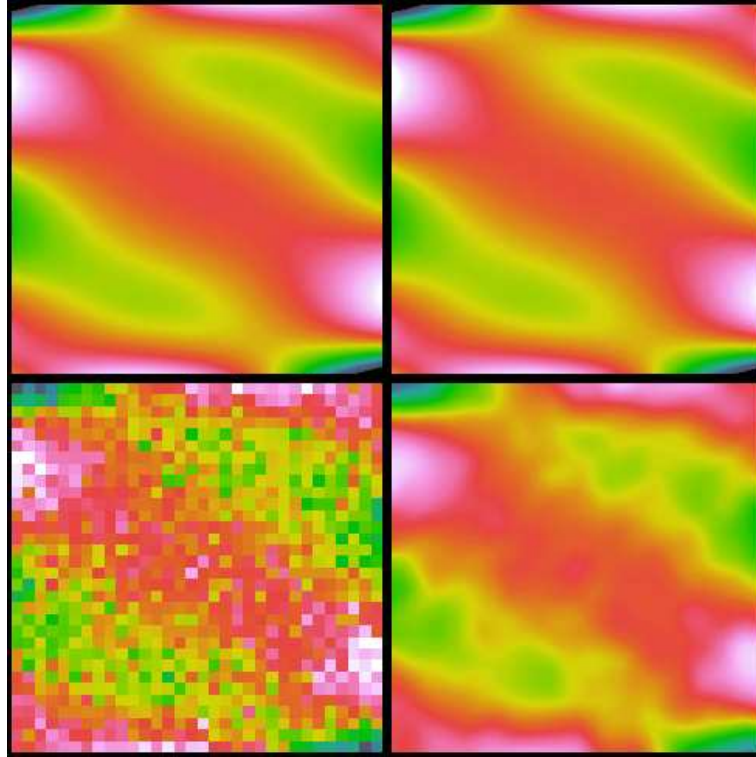
Fig. 1.— L-flats related to tests with artificial data. *(a; top left)* Input model L-flat used to generate the artificial data, given by equation (38). *(b; top right)* L-flat calculated from the artificial data using the differential matrix method and Legendre polynomial basis functions up to order $Z = 4$ (RMS difference with input model is 0.0020 mag). *(c; bottom left)* L-flat calculated from the artificial data using the differential matrix method and 32×32 chess board basis functions (RMS difference with input model is 0.0145 mag). *(d; bottom right)* result of smoothing panel (c) with a Gaussian with dispersion equal to the size of one basis function square (RMS difference with input model is 0.0055 mag). The color scheme is identical for all panels, and ranges from black (-0.15 mag) to white ($+0.075$ mag). This figure is best viewed or printed in color.

A simple consistency check is obtained by running the direct matrix method on the artificial data with either polynomial or chess board basis functions with $Z = 0$ or $W = 1$, respectively. Given the definitions in Section 2.4 this implies that there is only a single basis function that is equal to unity over the detector. The normalization constraint given by equation (8) then implies that the best-fit L-flat is identical to zero, i.e., no L-flat is invoked at all. The best-fitting stellar magnitudes m_i and their errors Δm_i, as obtained from the direct matrix method, should then be identical to the weighted averages of the observations for each star, given by equations (2) and (3). It was verified that this is indeed the case to

high accuracy.

A more sophisticated test is to try to recover the L-flat from the artificial data. First, Legendre polynomial basis functions up to order $Z = 4$ were used. For the direct method, the matrix size was 7925×1356 and the program took 3 hours and 15 minutes to complete its calculations on an Ultra 60 workstation with a 450 MHz processor (all computing speeds quoted hereafter refer to this hardware setup). For the differential method, the matrix size was 6594×25, and the program took only 2 minutes to complete its calculations. The result of the differential method is shown in Figure 1b. The result of the direct method looks the same, and is not shown. Both results are visually indistinguishable from the input image in Figure 1a that was used to create the artificial data. The RMS difference between the input image and the results of the matrix solutions are 0.0013 mag and 0.0020 mag, for the direct and differential methods respectively. The direct matrix method therefore provides the slightly more accurate solution, as expected on the basis of the arguments presented at the end of Section 2.3. However, the difference is negligible from an astrophysical viewpoint. At the level of a milli-mag many other effects are likely to dominate the error budget of a real observation. For both solutions the average formal error in the L-flat is 0.0014 mag. So for the direct matrix method the formal error provides a good estimate of the actual RMS accuracy of the solution. For the differential matrix method the actual RMS accuracy of the solution exceeds the average formal error by $\sim 50\%$. With the Legendre polynomial basis functions it is generally found that the data constrain the L-flat least accurately near the boundaries and corners of the detector. This is where the differences between the input image and the results of the matrix solutions, as well as the formal errors in the matrix solutions, tend to have their maximum values. The direct matrix method yields a best fit that reproduces the data with $\chi^2 = 6602.7$ for $N_{\mathrm{df}} = 6569$ degrees of freedom. The differential matrix method yields a best fit with $\chi^2 = 6617.5$. One expects χ^2 to be equal to N_{df} to within $\pm\sqrt{2N_{\mathrm{df}}} = \pm 114.6$ at 68.3% confidence. Hence, both χ^2 values are statistically acceptable. This is expected, given that the artificial data were generated with uncorrelated Gaussian random errors, using an L-flat that can be exactly reproduced with the adopted basis functions. The milli-mag level accuracy obtained with these tests is a measure only of the random errors in the results. In an actual application the errors will probably be dominated by systematic errors, which are more difficult to quantify (see Section 4.2 for discussion of an example case).

An additional test of the solutions obtained with the Legendre polynomial basis functions is obtained from inspection of the expansion coefficients. Let a_{qr} be the coefficient of $P_q(x')P_r(y')$ in the expansion. The input $R(x, y)$ is then characterized by $a_{13} = 0.1$, $a_{24} = -0.1$ and $a_{qr} = 0$ for all other q, r combinations. The coefficients in the solution of the direct matrix method are: $\hat{a}_{13} = 0.1014 \pm 0.0012$ and $\hat{a}_{24} = -0.1010 \pm 0.0019$. The

remaining coefficients have average and RMS spread $\hat{a}_{qr} = -0.0001 \pm 0.0012$, with an average formal error of 0.0012. The coefficients in the solution of the differential matrix method are: $\hat{a}_{13} = 0.1023 \pm 0.0013$ and $\hat{a}_{24} = -0.1029 \pm 0.0020$. The remaining coefficients have average and RMS spread $\hat{a}_{qr} = -0.0001 \pm 0.0018$, with an average formal error of 0.0013. Again, the direct matrix method provides the slightly more accurate solution.

As another test, the chess board basis functions with $W = 32$ (a 32×32 subdivision of the detector area) were used to analyze the same artificial data. For the direct method, the matrix size was 7925×2355, and the program took 11 hours and 16 minutes to complete its calculations. For the differential method, the matrix size was 6594×1024, and the program took 1 hour and 39 minutes to complete its calculations. The result of the differential method is shown in Figure 1c. The result of the direct method looks the same, and is not shown. Both results capture the essence of the input image shown in Figure 1a, apart from noise and the obvious "pixelization" of the results. The RMS difference between the input image and the results of the matrix solutions are 0.0120 mag and 0.0145 mag, for the direct and differential methods respectively. The average formal errors in the L-flats are 0.0107 mag and 0.0122 mag, respectively. The average formal errors therefore provide a reasonable estimate of the RMS accuracy of the solution. The direct method is slightly more accurate than the differential method, as was the case with the Legendre polynomial basis functions. With the chess board basis functions it is generally found that the differences between the input image and the results of the matrix solutions, as well as the formal errors in the matrix solutions, do not tend to have any particular spatial pattern. On average, the results are equally accurate anywhere on the detector. This is different than for results obtained for the Legendre polynomial basis functions, which tend to be least accurate near the detector boundaries and corners. The direct matrix method yields a best fit that reproduces the data with $\chi^2 = 5723.1$ for $N_{\mathrm{df}} = 5570$ degrees of freedom. The differential matrix method yields a best fit with $\chi^2 = 6229.2$. In both cases this is larger than what would be expected at 1σ significance. This primarily indicates that the smooth input L-flat model (equation [38]) cannot be exactly reproduced by the adopted pixelized basis functions.

The accuracy of the results obtained with the chess board basis functions can be improved significantly through application of additional smoothing with a two-dimensional Gaussian. The optimum amount of smoothing depends on the frequency content of the true L-flat, which in practice is not known. However, for the tests presented here it is found that the optimum smoothing is obtained by choosing the Gaussian dispersion equal to the size of a single basis function square. Figure 1d shows the result of applying this to Figure 1c. The RMS difference between the resulting L-flat and the input L-flat model is now only 0.0055 mag, compared to 0.0145 mag for the solution of the differential matrix method without any smoothing. Larger Gaussian dispersions give an even smoother L-flat. However, this

also smoothes away intermediate-frequency structure that is actually present in the input L-flat model. For example, using Gaussian smoothing with a dispersion equal to twice the size of a basis function square increases the RMS to 0.0080 mag. Note that any amount of smoothing increases the χ^2 of the fit because it reduces the ability of the model to fit the noise in the data. To make χ^2 useful for assessing the optimum tradeoff between fitting the data and producing an "acceptably smooth" solution one would need to add regularization terms to the definition of χ^2 (e.g., Press et al. 1992). This adds greatly to the complexity of the problem, and even if one does this, it is not straightforward to determine the optimum amount of regularization. I have therefore not explored this. This leaves open the question of how to determine the optimum amount of smoothing for real applications. This will depend on both the quality and quantity of the available data, and on the properties of the L-flat. This issue deserves further attention when the method is applied to real calibration data.

Overall, the tests with artificial data indicate that the algorithm and software work very well. The two different matrix methods produce results that are indistinguishable at astrophysically interesting levels. However, the differential matrix method is much faster, and is therefore to be preferred for real applications. It allows the use of more basis functions and more data to constrain the solution. The L-flats can be determined with very small random errors. Systematic errors are therefore likely to dominate the true error budget for realistic applications. The dominant error is the accuracy with which the basis functions can reproduce the actual L-flat. Polynomials have the disadvantage of not being able to reproduce intermediate-frequency content very well. This can be remedied by going to high polynomial orders, but then the solution can become inaccurate towards the detectors boundaries and corners. The chess board basis functions do not suffer from these shortcomings. The most general approach for real applications might therefore be to use the chess board basis functions, followed by smoothing of the solution.

4.2. Real Data

In Mack et al. (2002b) an analysis was presented of ACS/WFC data of the globular cluster 47 Tuc. Aperture photometry was performed on the pydrizzled data products from the ACS pipeline. These included flat-fielding with ground-based calibration flats, and full correction for geometric distortion. An aperture radius of 5 pixels was used, with no spatially variable aperture corrections. Jennifer Mack provided the photometry thus obtained, which I analyzed using the direct matrix method. The differential matrix method had not yet been implemented at that time. Legendre polynomial basis functions were
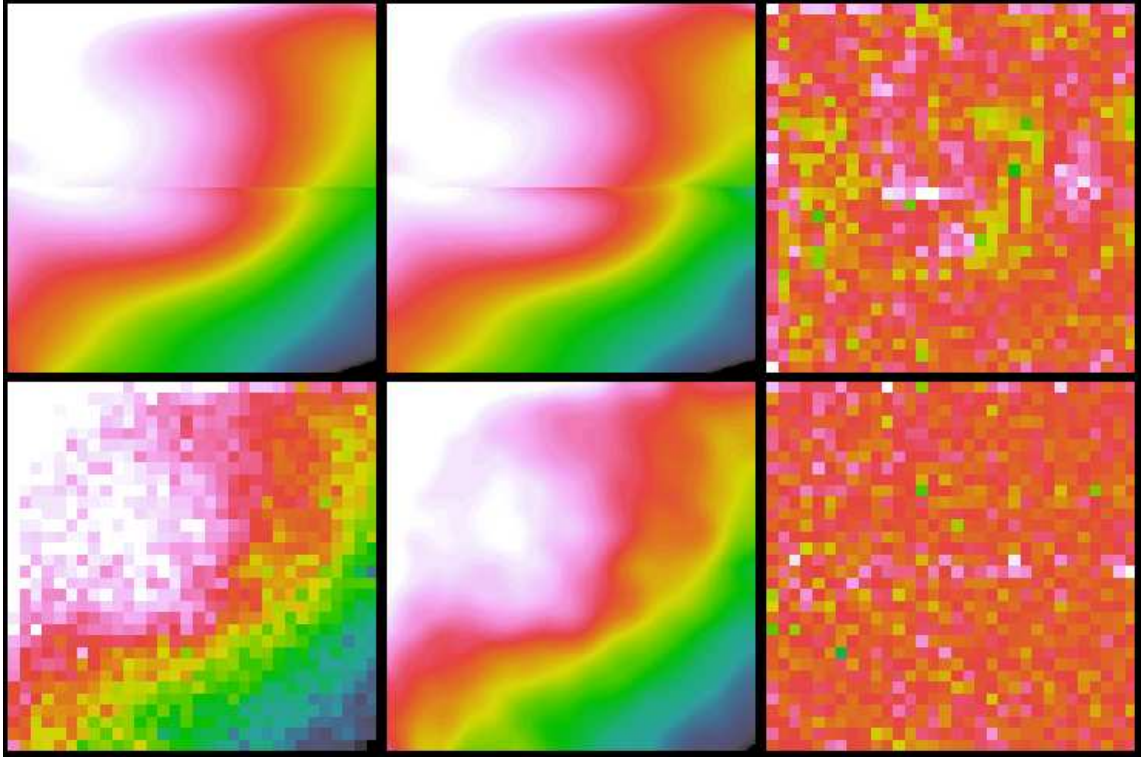
Fig. 2.— Residual L-flat for filter F606W on ACS/WFC with respect to ground-based calibration flats, inferred from photometry of the globular cluster 47 Tuc. *(a; top left)* Result presented in Mack et al. (2002b), obtained with the direct matrix method and Legendre polynomial basis functions up to order $Z = 4$. This L-flat was implemented into the ACS pipeline in late 2002, after smoothing away the discontinuity at the interchip boundary. *(b; top middle)* L-flat obtained with the differential matrix method and the same basis functions. *(c; top right)* Map of the average residuals in the stellar photometry after correction for the L-flat in panel (b). *(d; bottom left)* L-flat obtained with the differential matrix method and 32×32 chess board basis functions. *(e; bottom middle)* result of smoothing panel (d) with a Gaussian with dispersion equal to the size of one basis function square. *(f; bottom right)* Map of the average residuals in the stellar photometry after correction for the L-flat in panel (d). Comparison of the residual images in panels (c) and (f) shows that the chess board basis functions provide a less biased estimate of the true L-flat. This figure is best viewed or printed in color. The color scheme for the L-flat image panels ranges from black (-0.075 mag) to white ($+0.025$ mag). The color scheme for the residual image panels runs through the following colors: green (-0.010 mag), yellow (-0.005 mag), red (0.000 mag), pink ($+0.005$ mag) and white ($+0.010$ mag).

used up to order $Z = 4$. Computational restrictions limited the application of the method to no more than ~ 1500 stars per matrix solution. By contrast, the number of stars for which data was available was generally in the range 5000–10000. The data were therefore divided by stellar brightness into 3–6 subsets which were analyzed separately, each yielding a matrix size of $\sim 8000 \times 1500$. The L-flats inferred from the different subsets of the data were then averaged together, weighted with the formal errors. In total, the computations took some 16 hours per filter. Stars for which one or more of the measurements were potentially suspect were excluded from the fit. In practice this was done if the RMS scatter between the different measurements for a star exceeded the larger of $3\langle e \rangle_i$ and 0.05 mag, where $\langle e \rangle_i$ is the average of the formal errors e_{ij} in the measurements for a given star i. Stars were also excluded if less than three measurements were available. A more robust rejection scheme would be to iterate the matrix method, rejecting at each iteration those measurements for which the magnitude residual r_{ij} (equation [30]) exceeds some threshold (e.g., $3e_{ij}$). However, this is computationally much more expensive (because the matrix equation must be solved multiple times).

As an example of the results, the left panel of Figure 2a shows the resulting L-flat (measured with respect to the ground-based calibration flat) for the F606W filter (same as Figure 3c of Mack et al. 2002b). This took some 16 hours to calculate. There is a minor discontinuity at the interchip boundary. In Mack et al. (2002b) we smoothed out this discontinuity before delivery to the pipeline. This seems reasonable, although it is not a priori clear that the L-flat should actually be continuous at the boundary. The CCDs were cut from the same wafer, but the physical origin of the L-flat structures has not been established. Another method to enforce smoothness across the boundary is not to use separate basis functions for the two CCDs in the matrix method. I have not explored this, but this should be trivial to implement in the software.

The differential matrix method is now available as an alternative approach. All data can then be analyzed in a single run. The matrix size is 26631×50 and the result is obtained in only 2 minutes. Figure 2b shows the L-flat thus calculated. It is very similar to the result obtained with the direct matrix method. The RMS between the L-flats is only 0.0014 mag. This is a factor two larger than the average quadrature sum of the formal errors in the L-flats, which is 0.0007 mag. The largest differences (0.016 mag) between the results from the two methods occur in a very narrow strip near the boundary between the two CCDs. This is also where the formal errors in each L-flat are largest (0.005 mag). This is similar to what was seen in the tests with artificial data, and underscores the general finding that results obtained with the Legendre polynomial basis functions are most poorly constrained near the detector boundaries. The residual image of the fit is shown in Figure 2c. It reveals two problems. First, there are large residuals (up to 0.013 mag) at the boundary between

the two detectors. This indicates that the strength of the discontinuity at the interchip boundary inferred by the method is not real. Second, a coherent ring-like structure is present in the residual image at a level of ~ -0.006 mag. Both problems indicate low-level limitations of the adopted fourth-order Legendre polynomial basis function set. An analysis of sky flats (Pavlovsky et al., in progress) shows very similar residuals to those seen in Figure 2c. This confirms the results from the stellar field photometry and indicates that the current pipeline L-flats are not yet optimal.

The improvement in computational speed provided by the differential matrix method now makes it possible to use many more basis functions. With the 32×32 chess board basis the matrix size is 26631×1024, and the program completes its calculations in 6 hours and 55 minutes. The result is shown in Figure 2d. The corresponding residual image, shown in Figure 2f, shows no obvious structure. Figure 2e shows the result of smoothing the L-flat in Figure 2d with a Gaussian of dispersion equal to the size of one basis function square (see discussion in Section 4.1). Due to the lack of structure in the residual image, Figure 2e is likely to be a more accurate approximation to the true L-flat than the results in Figures 2a,b obtained with the fourth-order Legendre polynomial basis function set. This interpretation is supported by comparison of the RMS values of the residual images, which are 0.0022 mag for Figure 2f and 0.0030 mag Figure 2c.

The L-flat in Figure 2e has two main features. There is an approximately linear gradient between the top left and bottom right. In addition to that, there is an oval feature left and upwards from the detector center. Both features exist also in other ACS/WFC calibration data. For example, the diagonal between the top left and bottom right is the direction along which the pixel area varies most strongly as a result of geometric distortion. The oval feature is seen also in ground-based flats and in maps of the detector thickness (Krist 2003). The physical origin of the relations between these features and those seen in the (F606W) L-flat remain as yet unclear.

Mack et al. (2002b) provided several external tests of the accuracy of their results. This included analysis of color magnitude diagrams of 47 Tuc and comparison to preliminary sky flats. This analysis indicated that the flat field results (now implemented in the pipeline) should be accurate to $\sim 1\%$. On the other hand, subsequent tests showed that there may well be residual low-frequency errors at the $\sim 1\%$ level (Ratnatunga, Mack, Pavlovsky, priv. comm. 2003). This is confirmed by Figure 2c. These residuals can probably be calibrated using analyses such as those shown in Figures 2d–f. An important question that remains to be addressed is the importance of spatially dependent aperture corrections. In Mack et al. (2002b) we did a test for one filter by analyzing datasets obtained with either a 5 or 7 pixel radius aperture. The inferred L-flats were found to agree to 0.0015 mag RMS.

This was taken as evidence that the influence of spatially dependent aperture corrections must be small. On the other hand, subsequent work (Krist 2003; Riess 2003) has indicated that that there are in fact spatial dependencies in the aperture corrections (below $\sim 1\%$ for a 5 pixel radius aperture). These could affect the L-flat results at a low level. All of this indicates that it will be worthwhile to revisit the L-flats currently in the ACS pipeline. The new methods presented and tested here will allow an in-depth analysis of this issue, which will be the topic of a future ISR.

References

Bohlin, R. C., Hartig, G., & Martel, A. 2001, "HRC and WFC Flat Fields: Standard Filters, Polarizers and Coronagraph", ISR ACS 2001-11 (Baltimore: STScI)

Gilliland, R. 1998, "Use of FP-SPLIT Slits for Reaching High Signal-to-Noise with MAMA Detectors", ISR STIS 1998-16 (Baltimore: STScI)

Greenfield, P. 1994, in Astronomical Data Analysis Software and Systems III, A.S.P. Conference Series, Vol. 61, 1994, D. R. Crabtree, R.J. Hanisch, and J. Barnes, eds., p. 276

Mack, J., et al. 2002a, HST ACS Data Handbook, version 1.0, B. Mobasher, ed. (Baltimore: STScI)

Mack, J., Bohlin, R., Gilliland, R., van der Marel, R. P., Blakeslee, J., de Marchi, G. 2002b, "ACS L-Flats for the WFC", ISR ACS 2002-08 (Baltimore: STScI)

Manfroid, J. 1995, A&AS, 113, 587

Manfroid, J. 1996, A&AS, 118, 391

Manfroid, J., Selman, F., & Jones, H. 2001, ESO Messenger, 104, 16

Krist, J. 2003, "ACS WFC and HRC field-dependent PSF variations due to optical and charge diffusion effects", ISR ACS 2003-06 (Baltimore: STScI)

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes (Cambridge: Cambridge University Press)

Riess, A. 2003, "On-orbit Calibration of ACS CTE Corrections for Photometry", ISR ACS 2003-09 (Baltimore: STScI)

Wild, W. 1997, PASP, 109, 1269