Remediation in the NOAO Pipelines

F. Valdes[1], R. Swaters[1]

National Optical Astronomy Observatories
Science Data Management

V1.0: May 11, 2011

[1]NOAO Data Products Program, P.O. Box 26732, Tucson, AZ 85732

1 Introduction

The MOSAIC/NEWFIRM pipelines attempt to remediate as many problems as possible to allow data to be processed for PIs and future archive users. This is done with verification and remediation tools, special stages that can run remediation scripts, and a calibration library that contains correct and up-to-date keyword values.

One key element of the pipeline remediation arsenal is a powerful verification and remediation tool, `fitsverify.py`, developed by the pipeline team (Rob Swaters in particular). It is driven by a customizable text file. The calibration library provides the appropriate version for the data being processed; i.e. each instrument has its own file. A sample for the CTIO MOSAIC is shown in Appendix A. The comment block describes the options provided by the program based on the fields in the verification description file including remediation strategies, value checks, and keyword presence/absence checks. In effect, essentially all the keywords expected in the data are checked and actions taken if necessary.

One of the possible actions is to run a remediation command that checks and/or fixes a problem. We have only implemented a small number of remediation strategies. The primary problem encountered in NOAO MOSAIC data is misformated filter names. The others are for MJD-OBS, TIME-OBS, DATE-OBS, OBSID, GAIN, RDNOISE, and SATURATE. Some of these – FILTER, GAIN, RDNOISE, SATURATE – interact with the pipeline calibration library. For the filter we have various checks to deal with misspelling and aliases. One of the most useful checks is against the serial number which, if present, allows ignoring the rest of the filter string which is where most errors occur.

For the NEWFIRM pipeline, the most common remediation steps include adding keywords that were not present in NEWFIRM's early days and fixing keywords related to observing recipes.

Note that if only some input data files fail all attempts at remediation are logged but the processing continues on the subset that passed. It is not uncommon for a small number of exposures to be fatally flawed and unrecoverable. These are rejected and not processed by the pipeline.

Another aspect of the pipeline is that the calibration library is used during processing stages to override values added at the telescope. This allows updating these values as information and calibrations improve. These replacements always happen and include the WCS, the gain, and the read noise, regardless of what is in the header.

A final feature of the pipelines is a standard stage that looks for an arbitrary command to run immediately after the raw data has been fetched. This has proven useful during commissioning of a new instrument, when things are changing rapidly and header problems are found. This stage is also useful when some special processing is desired, such as forcing tangent points for stacking. The special scripts are put in a configuration directory and are recognized as applying only to specific datasets; e.g. for a particular block of nights. This is rarely used but is the ultimate way to deal with problems.

All the preceding is remediation prior to standard processing. Of course some of the pipeline processing steps do result in improving the headers in the final data products. The most important example of this is the astrometric calibration which updates the WCS in the raw image by a final accurate calibration.

2   Appendix A: Sample Verification/Remediation File

```
#This file lists all the keywords that should be present in the
#headers of CTIO and KP 4 m MOSAIC data. The first column contains
#the name of the keyword.
#
#The second gives a flag indicating whether this is a top level (T)
#or subset level (S) keyword.
#
#The third column contains a string that indicates for which
#telescope this header keyword should be present, C indicating the
#CTIO 4m, K the KP 4m.
#
#The fourth column indicates what the status is if this keyword is
#missing. There are four possibilities: FITS, FATAL, ERROR, and
#WARNING. FITS indicates that this keyword is required in a correct
#fits header. FATAL means that the pipeline cannot process this file
#as is. ERROR means that the file can be processed, but that some
#modules may fail if the file is left as is. WARNING means that the
#absence of this keyword has no consequences for the pipeline.
#A fifth possible value is IGNORE, which means that this keyword
#will not be considered. RMDIATE means that remediation is requested
#even for a non FATAL/ERROR status.
#
#The fifth column contains a (regular) expression which the keyword
#should satisfy. If set to True, the contents of the keyword will be
#ignored, but remediation will be done if the keyword is missing. If
#set to False, remediation will always happen. If no remediation is
#desired at all, set the value to -. The expression may contain an
#x to refer to a float. Other types are not yet implemented.
#
#The sixth column is a reference to a possible remediation strategy.
#These strategies are listed at the bottom of this file. NO in this
#column indicates that no remediation strategy is available. A '-'
#indicates that a remediation strategy has not yet been considered for
#this keyword. UNNES means that no remediation strategy is
#necessary. A string starting with _ refers to a python method/ (in
#remediation.py) that can do the remediation, a string starting with a
#* refers to an external script that can do the remediation (the
#latter is not yet implemented), a string starting with = refers to a
#fixed numerical value.  A D as the second character in this field
#indicates that if remediation is done, the subset keywords should be
#deleted so that only the global keyword remains (this is only allowed
#if the global keyword is present).
```

```
#
#The rest of each line gives a description of the header keyword.
#
#
# GLOBAL KEYWORDS in both CTIO and KP headers
# ==========================================
#
# FITS related
# ------------
BITPIX  T CK FITS - -bits per pixel
COMMENT   T CK IGNORE - -comments
DATE  T CK FITS - -Date fits file was generated
EXTEND  T CK FITS - -File may contain extensions
NAXIS  T CK FITS - -Number of axes
NEXTEND  T CK   FATAL   (x==16) _EVAL  Number of extensions
ORIGIN  T CK FITS - -FITS file originator
SIMPLE  T CK FITS - -Fits standard
#
# IRAF related
# ------------
FILENAME T CK ERROR - NO Original host file name
IRAF-TLM T CK IGNORE - NO Time of last modification
#
# PIPELINE CRITICAL KEYWORDS - pipeline cannot run without these
# -------------------------------------------------------------
DETECTOR T CK FATAL - RDET Detector
FILTER   T CK FATAL False _Dfilter Filter
MJD-OBS  T CK FATAL (50000<x<70000) _Dmjdobs MJD of observation start
OBSID    T CK FATAL True _Dobsid Observation ID
OBSTYPE  T CK FATAL - ROBST Observation type
#
# PIPELINE IMPORTANT KEYWORDS - some modules cannot run without these
# ------------------------------------------------------------------
AIRMASS  T CK ERROR - RAIR Airmass
DEC      T CK ERROR - RCOOR DEC of observation
EXPTIME  T CK ERROR - NO Exposure time (sec)
RA       T CK ERROR - RCOOR RA of observation
RADECEQ  T CK ERROR - RCOOR Default equinox
RADECSYS T CK ERROR - RCSYS Default coordinate system
TELDEC   T CK ERROR - RCOOR DEC of telescope
TELESCOP T CK ERROR - RTEL Telescope
TELEQUIN T CK ERROR - RCOOR Equinox of tel coords
TELRA    T CK ERROR - RCOOR RA of telescope
#
# PIPELINE USEFUL KEYWORDS - information not used by the pipeline
```

```
# ----------------------------------------------------------------
DARKTIME T CK WARNING - -Dark time (sec)
DATE-OBS T CK RMDIATE True _Ddateobs Date of observation start (UTC approximate)
LSTHDR   T CK WARNING - -LST of header creation
OBSERVAT T CK WARNING - -Observatory
MJDHDR   T CK WARNING - -MJD of header creation
OBJECT  T CK WARNING - -Observation title
PIXSCAL1 T CK WARNING - -Pixel scale for axis 1 (arcsec/pixel)
PIXSCAL2 T CK WARNING - -Pixel scale for axis 2 (arcsec/pixel)
PREFLASH T CK WARNING - -Preflash time (sec)
TELRADEC T CK WARNING - -Telescope coordinate system
TIME-OBS T CK RMDIATE True _Dtimeobs Time of observation start
TIMESYS  T CK WARNING - -Time system
ZD       T CK WARNING - -Zenith distance
#
# ENGINEERING AND RELATED KEYWORDS - information not used by the pipeline
# ----------------------------------------------------------------------
ADC      T CK WARNING - -ADC Identification
AMPINTEG T CK WARNING - -Double Correlated Sample time
ARCONGI  T CK WARNING - -Waveform mode switches on
ARCONWD  T CK WARNING - -Gain selection (index into Gain Table)
CCDTEM   T CK WARNING - -CCD temperature (C)
CCDTEM2  T CK WARNING - ----
CONSWV   T CK WARNING - -Controller software version
CONTROLR T CK WARNING - -Controller identification
CORRCTOR T CK WARNING - -Corrector Identification
DECPANGL T CK WARNING - -Position angle of Dec axis (deg)
DETSIZE  T CK WARNING - -Detector size for DETSEC
DEWAR    T CK WARNING - -Dewar identification
DEWTEM   T CK WARNING - -Dewar temperature (C)
DEWTEM2  T CK WARNING - -Fill Neck temperature (C)
DEWTEM3  T CK WARNING - ----
DTACCOUN T CK WARNING - -observing account name
DTACQNAM T CK WARNING - -file name supplied at telescope
DTACQUIS T CK WARNING - -host name of data acquisition computer
DTCALDAT T CK WARNING - -calendar date from observing schedule
DTCOPYRI T CK WARNING - -copyright holder of data
DTINSTRU T CK WARNING - -instrument identifier
DTNSANAM T CK WARNING - -file name in NOAO Science Archive
DTOBSERV T CK WARNING - -scheduling institution
DTPI  T CK WARNING - -Principal Investigator
DTPIAFFL T CK WARNING - -PI affiliation
DTPROPID T CK WARNING - -observing proposal ID
DTPUBDAT T CK WARNING - -calendar date of public release
DTSITE  T CK WARNING - -observatory location
```

```
DTTELESC T CK WARNING - -telescope identifier
DTTITLE  T CK WARNING - -title of observing proposal
ENVTEM   T CK WARNING - -Ambient temperature (C)
FILPOS   T CK WARNING - -Filter position
IMAGESWV T CK WARNING - -Image creation software version
KWDICT   T CK WARNING - -Keyword dictionary
NAMPS    T CK WARNING - -Number of Amplifiers
NCCDS    T CK WARNING - -Number of CCDs
PIXSIZE1 T CK WARNING - -Pixel size for axis 1 (microns)
PIXSIZE2 T CK WARNING - -Pixel size for axis 2 (microns)
RAPANGL  T CK WARNING - -Position angle of RA axis (deg)
READTIME T CK WARNING - -unbinned pixel read time
SHUTSTAT T CK WARNING - -Shutter status
TELFOCUS T CK WARNING - -Telescope focus
TV1FOC   T CK WARNING - -North TV Camera focus position
TV2FOC   T CK WARNING - -South Camera focus position
#
# PROPOSAL RELATED KEYWORDS - information not used by the pipeline
# ----------------------------------------------------------------
OBSERVER T CK WARNING - -Observer(s)
PROPID   T CK WARNING - -Proposal identification
PROPOSAL T CK WARNING - -Proposal title
PROPOSER T CK WARNING - -Proposer(s)
#
# PIPELINE GENERATED KEYWORDS (?)
# ------------------------------
AMPNAMES T CK IGNORE - UNNES -
CALOPS   T CK IGNORE - UNNES calops string
OTFDIR   T CK IGNORE - UNNES OTF calibration directory
MAGZERO  T CK IGNORE - UNNES -
PHOTFILT T CK IGNORE - UNNES -
PHOTCOEF T CK IGNORE - UNNES -
UPARM    T CK IGNORE - UNNES -
WCSDB    T CK IGNORE - UNNES -
XTALKFIL T CK IGNORE - UNNES Crosstalk file
#
# CHECKSUM RELATED KEYWORDS - information not used by the pipeline
# ----------------------------------------------------------------
CHECKSUM T CK WARNING - -Header checksum
CHECKVER T CK WARNING - -Checksum version
DATASUM  T CK WARNING - -Data checksum
RECNO    T CK WARNING - -NOAO archive sequence number
#
#
#
```

```
# GLOBAL KEYWORDS only in CTIO headers
# ===================================
#
# ENGINEERING AND RELATED KEYWORDS - information not used by the pipeline
# ----------------------------------------------------------------------
ARCONWM  T C WARNING - --
HA   T C WARNING - --
#
# WEATHER AND RELATED KEYWORDS - information not used by the pipeline
# -----------------------------------------------------------------
AMBTEMP  T C WARNING - -Ambient temperature (degrees C)
HUMIDITY T C WARNING - -Ambient relative humidity (percent)
PRESSURE T C WARNING - -Barometric pressure (millibars)
SEEING  T C WARNING - -Tololo DIMM seeing
WEATDATE T C WARNING - -Date and time of last update
WINDDIR  T C WARNING - -Wind direction (degrees)
WINDSPD  T C WARNING - -Wind speed (mph)
#
#
#
# GLOBAL KEYWORDS only in KP headers
# ==================================
#
# ENGINEERING AND RELATED KEYWORDS - information not used by the pipeline
# ----------------------------------------------------------------------
ADCPAN1  T K WARNING - -ADC 1 prism angle
ADCPAN2  T K WARNING - -ADC 2 prism angle
ADCSTAT  T K WARNING - -ADC tracking state null-track-preset
FOCUS1  T K IGNORE  - ----
FOCUS2  T K IGNORE  - ----
FOCUS3  T K IGNORE  - ----
FOCUS4  T K IGNORE  - ----
FOCUS5  T K IGNORE  - ----
FOCUS6  T K IGNORE  - ----
FOCUS7  T K IGNORE  - ----
FOCUS8  T K IGNORE  - ----
FOCUS9  T K IGNORE  - ----
FOCUS10  T K IGNORE  - ----
FOCUS11  T K IGNORE  - ----
FOCUS12  T K IGNORE  - ----
FOCUS13  T K IGNORE  - ----
FOCUS14  T K IGNORE  - ----
FOCUS15  T K IGNORE  - ----
#
#
```

```
#
#
# KEYWORDS ONLY IN SUBSETS both in CTIO and KP data
# =================================================
#
# FITS related
# -----------
BITPIX  S CK FITS - -bits per pixel
BSCALE  S CK FITS - -REAL = TAPE*BSCALE + BZERO
BZERO   S CK FITS - ----
COMMENT  S CK IGNORE - -comments
DATE  S CK IGNORE - -Date fits file was generated
EXTNAME   S CK FITS - -Extension name
EXTVER  S CK FITS - -Extension version
GCOUNT  S CK FITS - -Only one group
INHERIT  S CK FITS - -Inherits global header
NAXIS   S CK FITS - ----
NAXIS1  S CK FITS - ----
NAXIS2  S CK FITS - ----
ORIGIN  S CK IGNORE - -FITS file originator
PCOUNT  S CK FITS - -No 'random' parameters
XTENSION S CK   FITS    - -     Image extension
#
# IRAF related
# ------------
IRAF-TLM S CK IGNORE - NO Time of last modification
#
# PIPELINE CRITICAL KEYWORDS - pipeline cannot run without these
# -------------------------------------------------------------
AMPNAME  S CK FATAL - RAMP Amplifier name
BIASSEC  S CK FATAL - RPROP Bias section
FILTER   S CK RMDIATE False _D Filter
IMAGEID  S CK FATAL - RAMP Image identification
MJD-OBS  S CK RMDIATE (50000<x<70000) _D MJD of observation start
OBSID    S CK RMDIATE True _D Observation ID
OBSTYPE  S CK IGNORE - ROBST Observation type
TRIMSEC  S CK FATAL - RPROP Trim section
#
# PIPELINE IMPORTANT KEYWORDS - some modules cannot run without these
# ------------------------------------------------------------------
DEC      S CK IGNORE - RCOOR DEC of observation
GAIN  S CK RMDIATE (x>0.1) _gain gain (e-/ADU)
RA       S CK IGNORE - RCOOR RA of observation
RADECEQ  S CK IGNORE - RCOOR Default equinox
RDNOISE  S CK RMDIATE (x>0.1) _rdnoise read noise (e-)
```

```
SATURATE S CK ERROR (1<x<65000) _saturate Maximum good data value (ADU)
#
# PIPELINE USEFUL KEYWORDS - information not used by the pipeline
# ----------------------------------------------------------------
DARKTIME S CK WARNING - -Dark time (sec)
DATE-OBS S CK RMDIATE True _D Date of observation start (UTC approximate)
EXPTIME  S CK WARNING - -Exposure time (sec)
LSTHDR   S CK IGNORE - -LST of header creation
PREFLASH S CK IGNORE - -Preflash time (sec)
TIME-OBS S CK RMDIATE True _D Time of observation start
#
# ENGINEERING AND RELATED KEYWORDS - information not used by the pipeline
# ----------------------------------------------------------------------
AMPSEC  S CK WARNING - -Amplifier section
AMPINTEG S CK IGNORE - -Double Correlated Sample time
ARCONG  S CK WARNING - -gain (e-/ADU)
ARCONGI  S CK IGNORE - -Waveform mode switches on
ARCONWD  S CK IGNORE - -Gain selection (index into Gain Table)
ATM1_1  S CK WARNING - -CCD to amplifier transformation
ATM2_2  S CK WARNING - -CCD to amplifier transformation
ATV1  S CK WARNING - -CCD to amplifier transformation
ATV2  S CK WARNING - -CCD to amplifier transformation
CCDNAME  S CK WARNING - -CCD name
CCDSEC  S CK WARNING - -CCD section
CCDSIZE  S CK WARNING - -CCD size
CCDSUM  S CK WARNING - -CCD pixel summing
CONHWV  S CK WARNING - -Controller hardware version
CONTROLR S CK IGNORE - -Controller identification
DATASEC  S CK WARNING - -Data section
DETSEC  S CK WARNING - -Detector section
DTM1_1  S CK WARNING - -CCD to detector transformation
DTM2_2  S CK WARNING - -CCD to detector transformation
DTV1  S CK WARNING - -CCD to detector transformation
DTV2  S CK WARNING - -CCD to detector transformation
OBJECT  S CK IGNORE - -Observation title
LTM1_1  S CK WARNING - -CCD to image transformation
LTM2_2  S CK WARNING - -CCD to image transformation
LTV1  S CK WARNING - -CCD to image transformation
LTV2  S CK WARNING - -CCD to image transformation
READTIME S CK IGNORE - -unbinned pixel read time
#
# PIPELINE GENERATED KEYWORDS (?)
# ------------------------------
AMPNAMES S CK   IGNORE  - UNNES -
BPM  S CK IGNORE - UNNES -
```

```
CD1_1  S CK IGNORE - UNNES -
CD1_2  S CK IGNORE - UNNES -
CD2_1  S CK IGNORE - UNNES -
CD2_2  S CK IGNORE - UNNES -
CRPIX1  S CK IGNORE - UNNES -
CRPIX2  S CK IGNORE - UNNES -
CRVAL1  S CK IGNORE - UNNES -
CRVAL2  S CK IGNORE - UNNES -
CTYPE1  S CK IGNORE - UNNES -
CTYPE2  S CK IGNORE - UNNES -
EQUINOX  S CK IGNORE - UNNES -
FLAT    S CK IGNORE - UNNES -
UPARM  S CK IGNORE - UNNES -
WAT0_001 S CK IGNORE - UNNES -
WAT1_001 S CK IGNORE - UNNES -
WAT1_002 S CK IGNORE - UNNES -
WAT1_003 S CK IGNORE - UNNES -
WAT1_004 S CK IGNORE - UNNES -
WAT1_005 S CK IGNORE - UNNES -
WAT2_001 S CK IGNORE - UNNES -
WAT2_002 S CK IGNORE - UNNES -
WAT2_003 S CK IGNORE - UNNES -
WAT2_004 S CK IGNORE - UNNES -
WAT2_005 S CK IGNORE - UNNES -
WCSASTRM S CK IGNORE - UNNES -
WCSDIM  S CK IGNORE - UNNES -
XTALKFIL S CK    IGNORE  - UNNES -
ZERO  S CK IGNORE - UNNES -
#
# CHECKSUM RELATED KEYWORDS - information not used by the pipeline
# ----------------------------------------------------------------
CHECKSUM S CK IGNORE - -Header checksum
CHECKVER S CK IGNORE - -Checksum version
DATASUM  S CK IGNORE - -Data checksum
#
#
#
# KEYWORDS ONLY IN SUBSETS only in CTIO headers
# ==============================================
#
# ENGINEERING AND RELATED KEYWORDS - information not used by the pipeline
# ----------------------------------------------------------------------
ARCONRN  S C WARNING read noise (e-)
#
#
```

```
#
# KEYWORDS ONLY IN SUBSETS only in KPNO headers
# =============================================
#
# ENGINEERING AND RELATED KEYWORDS - information not used by the pipeline
# ----------------------------------------------------------------------
FOCSTART S K IGNORE - ----
FOCSTEP  S K IGNORE - ----
FOCNEXPO S K IGNORE - ----
FOCSHIFT S K IGNORE - ----
#
#
#
#EVAL 1) Evaluate boolean expression and result is success or fail
#RAIR 1) convert ZD into AIRMASS
# 2) calculate from RA/Dec and time
#RAMP 1) Compare IMAGEID/DETECTOR and AMPNAME
#RCOOR  Use RA for TELRA etc. or the other way around
#RCSYS 1) if stable over long time, use standard value (i.e., FK5)?
# 2) Compare against previous/future value during same night/run
#RDET 1) Compare against data manager?
# 2) Compare against previous/future value during same night/run
#RFILT compare against filter names in data manager
#RMJD 1) Calculate from other time keywords
# 2) Interpolate between other exposures based on exposure time,
#    if exposures are separated by a little more than the exposure
#         time
#ROBST 1) Guess from object keyword
#RPROP 1) The values of these keywords depend on the detector (and
#    possibly on time), so these values can be looked up if
#         DETECTOR/IMAGEID and/or AMPNAME is defined
#    (This is not true if gain is user selectable)
# 2) Compare against previous/future values during the same night/run
#RTEL 1) compare again data manager?
# 2) Compare against previous/future value during same night/run
```