Know your SPIRou data!

Étienne Artigau for the SPIRou DRS team:

Neil Cook, Andres Carmona, Claire Moutou, Pascal Fouqué, Luc Arnold, Chris Usher, Pia Cortes Zuleta, Merwan Ould-Elhkim, Guillaume Gaisné, Xavier Delfosse, Jean-François Donati, François Bouchy, Melissa Hobson, Isabelle Boisse, Kanoa Withington
You’ve got some SPIRou data?

- Since 2019A SPIRou is the most requested CFHT instrument
- Lots of SLS and PI targets have been observed since February 2019
- You want to analyze your data but you are not familiar with the way they are obtained, reduced and distributed
- Lucky you, APERO does everything for you
Using APERO for SPIRou

Version 0.6.132 (2020-10-15)
A primer by Neil Cook
Contents

1. APERO on github
2. APERO module structure
3. APERO data structure
4. Primer on raw SPIRou data
5. Overview of the reduction process
APERO is on github

- Currently private: give Neil your github username and he will add you!
  - Repository here: https://github.com/njcuk9999/apero-drs
  - Add your name and github username here (currently 33 people):
    https://docs.google.com/spreadsheets/d/1eJNiQPIJyal_2SQnsEvcmaDVnJ00RAAslHKYjEX_nBc/edit?usp=sharing
A PipelinE to Reduce Observations

- raw science observations
- pre-processing
- calibration
- extraction
- Polarimetry
- RV calculation
- Telluric correction
The data directories

- Paths set during installation process

---

Note: in versions 0.7+ also an “assets” directory for masks and other data supplied by apero

---

- calibration
- raw science observations
- Polarimetry
- extraction
- RV calculation
- calibration
- Telluric correction
- Telluric correction
- pre-processing
The night directories

- Individual nights stored in sub-directories (directory name is not important)
  - Montreal and LAM use YYYY-MM-DD (UTC)
  - CFHT uses 19AQ08-Apr19
  - Processing is independent of sub-directory structure but structure is kept throughout (raw, tmp, reduced)
The Grand Plan

- PIs get science-ready data through the CADC archive (needs an account)
- The data package includes a number of intermediate
  - Single spectrum (all orders merged) from 0.95-2.5µm, CCF and polarization products
  - Most instrumental effects are corrected for
  - Relevant intermediate products are provided
  - Ramp and raw frames are provided
- Calibrated data are being re-processed frequently as the DRS improves
  - PI-friendly description of the updates on the SPIRou website
  - Don’t panic when your data is re-released...
4096x4096xN
- These are huge files (up to a few Gb each)!

Individual images taken every 5.572s; used to compute the per-pixel time-derivative of flux

The ramp is generally *not saved* by IR instruments
- Not saved in WIRCam
- *We save these frames in the hope of improving the per-frame filtering of data in the future*
r.fits
[ramp]

- Accumulation of signal for a bright A star over 56 frames
- Middle of $H$ band only
Raw observations

- Three fibers: two science=A,B, one reference=C
- A and B are used for polarization but are combined for intensity: AB
- APERO observation notation:
  - \{fiber AB\}_\{fiber C\}
- Many types of observations:
  - Science observations: OBJ in AB, FP or DARK in C
    - OBJ_FP, OBJ_DARK
  - Hot stars (for telluric correction): OBJ in AB, DARK in C
    - OBJ_DARK
  - Calibrations: different combinations in AB and C
    - Combinations of DARK, FLAT, FP, HC1, HC2
      - DARK_DARK, DARK_FLAT, FLAT_DARK, FLAT_FLAT, FP_FP, HC1_HC1, DARK_FP etc...
  - Sky observations, Laser Frequency comb (LFC_LFC) etc
Pre-DRS

Data cube

RAMP DATA ➔ FITS2RAMP ➔ RAW DATA
Dark pixels

Red cutoff at ~2.5 µm

\[ K \]

deep telluric absorption

\[ H \]

deep telluric absorption

\[ J \]

shallower telluric absorption

Blaze fall-off

\[ Y \]

Blaze fall-off

blue cutoff at ~0.98 µm

![Image of the page content]

**o.fits**

**[object]**

- Ext 1: The ‘slope’ image, this is the closest analog to a CCD image
- Ext 2: Intercept of the slope. This is more or less the equivalent of a bias. *Mostly for engineering.*
- Ext 3: Formal error on the slope. To be used for pixel quality assessment.
- Ext 4: Number of unsaturated readouts (normally equal to Nread). To be used for pixel quality assessment.
SPIRou Raw science observations

- H4RG detector
- Raw frame: 4096 x 4096 pixels
SPIRou Raw science observations
SPIRou Raw science observations

- Fibers A and B: OBJ
- Fibers C: DARK

- Header:
  - SBCCAS_P = pos_pk
  - SBCREF_P = pos_pk
  - TRG>Type = TARGET or SKY
  - OBJECT = Object name

- Can be hot star (telluric), science observation or sky observation

- Fibers A and B: OBJ
- Fibers C: FP

- Header:
  - SBCCAS_P = pos_pk
  - SBCREF_P = pos_fp
  - TRG_TYPE = TARGET
  - OBJECT = Object name

- Science observation
What’s in a calibration?

Quiz time: the next slides show some raw images - try to guess what is in each fiber (AB and C).

Choices are:

- FLAT
- FP
- HC

FP
- DARK

- FLAT_DARK, FLAT_FP, FLAT_HC, FLAT_FLAT
- HC_DARK, HC_FP, HC_HC, HC_FLAT
- FP_DARK, FP_FP, FP_HC, FP_FLAT
- DARK_DARK, DARK_FP, DARK_HC, DARK_FLAT

- Hint: ones in red are not used for calibration
- Note: DARK_DARK come in two flavours:
  - DARK_DARK_INT (internal darks) → just the fiber + detector
  - DARK_DARK_TEL (telescope darks) → through the cassegrain
Options:
- FLAT_DARK
- FLAT_FLAT
- HC_HC
- FP_FP
- DARK_DARK
- DARK_FP
- DARK_FLAT
Options:
- FLAT_DARK
- FLAT_FLAT
- HC_HC
- FP_FP
- DARK_DARK
- DARK_FP
- DARK_FLAT
Options:
- FLAT_DARK
- FLAT_FLAT
- HC_HC
- FP_FP
- DARK_DARK
- DARK_FP
- DARK_FLAT
Options:
- FLAT_DARK
- FLAT_FLAT
- HC_HC
- FP_FP
- DARK_DARK
- DARK_FP
- DARK_FLAT
Useful list of SPIRou raw files

<table>
<thead>
<tr>
<th>DPRTYPE</th>
<th>SBCCAS_P</th>
<th>SBCREF_P</th>
<th>SBCALI_P</th>
<th>OBSTYPE</th>
<th>TRG_TYPE</th>
<th>EXT.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DARK_DARK_INT</td>
<td>pos_pk</td>
<td>pos_pk</td>
<td>P4</td>
<td>DARK</td>
<td>CALIBRATION</td>
<td>d.fits</td>
</tr>
<tr>
<td>DARK_DARK_TEL</td>
<td>pos_pk</td>
<td>pos_pk</td>
<td>P5</td>
<td>DARK</td>
<td>CALIBRATION</td>
<td>d.fits</td>
</tr>
<tr>
<td>OBJ DARK</td>
<td>pos_pk</td>
<td>pos_pk</td>
<td>?</td>
<td>OBJECT</td>
<td>TARGET</td>
<td>o.fits</td>
</tr>
<tr>
<td>DARK_DARK_SKY</td>
<td>pos_pk</td>
<td>pos_pk</td>
<td>?</td>
<td>OBJECT</td>
<td>SKY</td>
<td>o.fits</td>
</tr>
<tr>
<td>DARK_FP_SKY</td>
<td>pos_pk</td>
<td>pos_fp</td>
<td>?</td>
<td>OBJECT</td>
<td>SKY</td>
<td>o.fits</td>
</tr>
<tr>
<td>OBJ_FP</td>
<td>pos_pk</td>
<td>pos_fp</td>
<td>?</td>
<td>OBJECT</td>
<td>TARGET</td>
<td>o.fits</td>
</tr>
<tr>
<td>FP_FP</td>
<td>pos_fp</td>
<td>pos_fp</td>
<td>?</td>
<td>ALIGN</td>
<td>CALIBRATION</td>
<td>a.fits</td>
</tr>
<tr>
<td>LFC_LFC</td>
<td>pos_rs</td>
<td>pos_rs</td>
<td>?</td>
<td>ALIGN</td>
<td>CALIBRATION</td>
<td>a.fits</td>
</tr>
<tr>
<td>FLAT_DARK</td>
<td>pos_wl</td>
<td>pos_pk</td>
<td>?</td>
<td>FLAT</td>
<td>CALIBRATION</td>
<td>f.fits</td>
</tr>
<tr>
<td>DARK_FLAT</td>
<td>pos_wl</td>
<td>pos_wl</td>
<td>?</td>
<td>FLAT</td>
<td>CALIBRATION</td>
<td>f.fits</td>
</tr>
<tr>
<td>FLAT_FLAT</td>
<td>pos_wl</td>
<td>pos_wl</td>
<td>?</td>
<td>FLAT</td>
<td>CALIBRATION</td>
<td>f.fits</td>
</tr>
<tr>
<td>HCON_DARK</td>
<td>pos_h1</td>
<td>pos_pk</td>
<td>?</td>
<td>COMPARISON</td>
<td>CALIBRATION</td>
<td>c.fits</td>
</tr>
<tr>
<td>HCTWO_DARK</td>
<td>pos_h2</td>
<td>pos_pk</td>
<td>?</td>
<td>COMPARISON</td>
<td>CALIBRATION</td>
<td>c.fits</td>
</tr>
<tr>
<td>FP_HCON</td>
<td>pos_fp</td>
<td>pos_h1</td>
<td>?</td>
<td>COMPARISON</td>
<td>CALIBRATION</td>
<td>c.fits</td>
</tr>
<tr>
<td>FP_HCTWO</td>
<td>pos_fp</td>
<td>pos_h2</td>
<td>?</td>
<td>COMPARISON</td>
<td>CALIBRATION</td>
<td>c.fits</td>
</tr>
<tr>
<td>HCON_FPA</td>
<td>pos_h1</td>
<td>pos_fp</td>
<td>?</td>
<td>COMPARISON</td>
<td>CALIBRATION</td>
<td>c.fits</td>
</tr>
<tr>
<td>HCTWO_FPA</td>
<td>pos_h2</td>
<td>pos_fp</td>
<td>?</td>
<td>COMPARISON</td>
<td>CALIBRATION</td>
<td>c.fits</td>
</tr>
<tr>
<td>DARK_HCON</td>
<td>pos_pk</td>
<td>pos_h1</td>
<td>?</td>
<td>COMPARISON</td>
<td>CALIBRATION</td>
<td>c.fits</td>
</tr>
<tr>
<td>DARK_HCTWO</td>
<td>pos_pk</td>
<td>pos_h2</td>
<td>?</td>
<td>COMPARISON</td>
<td>CALIBRATION</td>
<td>c.fits</td>
</tr>
<tr>
<td>HCON_HCON</td>
<td>pos_h1</td>
<td>pos_h1</td>
<td>?</td>
<td>COMPARISON</td>
<td>CALIBRATION</td>
<td>c.fits</td>
</tr>
<tr>
<td>HCTWO_HCTWO</td>
<td>pos_h1</td>
<td>pos_h2</td>
<td>?</td>
<td>COMPARISON</td>
<td>CALIBRATION</td>
<td>c.fits</td>
</tr>
</tbody>
</table>

- Note “DPRTYPE” set in preprocessing (so not in the raw files)
- This table can be found here: https://github.com/njcuk9999/apero-drs#71-input-files
- Raw files will have the following keys to ID the files:
  - SBCCAS_P: the position of the cassegrain fiber
  - SBCREF_P: the position of the reference fiber
  - SBCALI_P: the position of the calibration wheel
  - OBSTYPE: The observation type
  - TRG_TYPE: The target type
Overview of the reduction process
Pre-processing

- Needs to be done on every observation (science or calibration)
- Cleans data of many detector effects
Pre-processing

- Images may be flagged as corrupted at this stage ($t_{exp} < 0.1 \ t_{req}$)
- Image is rotated with respect to the raw images
- Saved in the tmp directory with “_pp” extension added

```
```
Calibration

Two types of calibration

- “Master” calibration
  - Only done once - provides a reference frame for all other calibrations
- “Night” calibration
  - Uses the calibrations taken on each night
  - Intended for science observed on the same night as those calibrations
  - May be used for a different night if calibrations fail

- Two modes:
  - “older calibration” (those taken immediately before observations started that night)
  - “closest calibration (those taken closest to an observation (either those at the start of night or end of night, depending on observation time)

- Currently the default is “closest calibrations” but this is the subject of ongoing testing
- Think about earthquake
Master calibration sequence

- **cal_preprocess**: [all files]
  - **dark_master**: [all DARK_DARK]
  - **cal_flat**: [master night FLAT_FLAT]
  - **cal_leak_master**: [all DARK_FP]

- **cal_badpix**: [master night FLAT_FLAT and DARK_DARK]

- **cal_loc**: [master night FLAT_DARK and DARK_FLAT]
  - **cal_shape**: [master night FP_FP]
  - **cal_shape_master**: [all FP_FP]

- **cal_thermal**: [master night DARK_DARK]
  - **cal_wave_master**: [master night FP_FP HC_HC]
Night calibration sequence

- **cal_preprocess** [all files]
  - **cal_badpix** [night FLAT_FLAT and DARK_DARK]
    - **cal_loc** [night FLAT_DARK and DARK_FLAT]
      - **cal_shape** [night FP_FP]
      - **cal_wave_night** [night FP_FP HC_HC]
        - **cal_thermal** [night DARK_DARK]
          - **cal_flat** [night FLAT_FLAT]
Calibration Database

Note that although only certain files are used in each calibration recipe they also rely on the output of previous steps

- All calibrations (good and bad) are saved to the reduced directory
- Calibrations that pass quality control are saved to the “calibDB”
- Calibrations that pass quality control are added to the database
  - Filename: “master_calib_SPIROU.txt”
- Other recipes use the calibration database and select based on observation time
  - $\text{MJDMID} = \text{MJDEND} - \frac{\text{EXPTIME}}{2}$
master_calib_SPIROU.txt

- Just a simple text file (for now)
- Columns:
  - Calibration key, master flag, night directory, filename, human time, unix time
  - Can remove calibration simply by commenting a line (start with a #)
- Each calibration has a unique key
  - Dark_master → DARKM
  - Bad pixel map → BADPIX
  - Master wave solution → WAVEM
- Calibrations are selected based on time in the header compared to time in this file:
  - By default the closest calibration is used
  - Can switch to older (closest in time that is older)
What happens after calibration?

- After a master sequence is processed, night calibrations can be processed.
- After night calibrations, the route forward depends on what observation was done.
- The type of observation determines how far along the pipeline we need to go.
- Types of observation:
  - Hot star (for telluric correction)
  - Spectral/RV observation
  - Polar observation

Diagram:

- raw science observations
- pre-processing
- calibration
- extraction
- Telluric correction
- RV calculation
- Polarimetry
Telluric sequence

- Hot stars are observed most nights: they are early A fast rotator stars
- If a good hot star template is available steps in green could be skipped
- Note the extraction process uses all calibration steps

**pre-processing**
[all files]

**cal_extract**
extraction
[night OBJ_FP, OBJ_DARK]

**cal_leak**
FP leakage correction
[night e2dsff]

**obj_mk_tellu**
Calculate telluric transmission
[night e2dsff]

**obj_mk_template**
Make hot star templates
[night OBJNAME]

**obj_fit_tellu**
Correct tellurics in hot stars
[night e2dsff]

**obj_mk_tellu**
Recalculate telluric transmission
[night e2dsff]
Telluric sequence

- In practice one needs many hot stars to make a good template and hot stars must be spread across airmass / time of the year (BERV) / water vapor column density
- Also obj_fit_tellu requires at least as many telluric transmissions as PCA components
- Thus for the best possible telluric database:
  - all telluric transmission should be calculated before any telluric star is corrected for tellurics
  - all hot stars should be corrected for tellurics before a hot star template is made
  - All telluric transmission should be recalculated after a template has been constructed
- All telluric outputs are saved to the telluric database (telluDB directory) similar to the calibration database

- `obj_mk_tellu` All hot stars
- `obj_fit_tellu` All hot stars
- `obj_mk_template` Once for each hot star OBJECT
- `obj_mk_tellu` All hot stars
Spectral/RV sequence

- Note the extraction process uses all calibration steps
- If a good object template is available steps in green could be skipped
- If OBJ_DARK, leak correction is not required but RV measurement may be hindered
- Similar to the hot stars for a good template many observations across time of the year (BERV) are required
Polar sequence

- **pre-processing** [all files]
- **cal_extract**
  - extraction
  - [night OBJ_FP, OBJ_DARK]
- **cal_ccf**
  - Calculate CCF
- **pol_spirou**
  - Calculate polarimetry
- **cal_leak**
  - FP leakage correction
  - [night e2dsff]
- **obj_fit_tellu**
  - Correct tellurics
  - [night e2dsff]
- **obj_mk_template**
  - Make object templates
  - [night OBJNAME]

- Treated very similar to Spectral/RV observation
- Polar recipe requires e2dsff files from A and B fibers separately
As FITS images:
- Ext 1-4: AB, A, B and C extracted spectra, 4088x49 pixels
- Ext 5-8: AB, A, B and C wavelength maps, 4088x49 pixels
- Ext 9-12: AB, A, B and C blaze maps, 4088x49 pixels
s.fits

[science]

- Stitched spectrum over the entire 0.95-2.5\textmu m domain
  - As FITS binary tables
  - Extensions for AB, A, B and C fibre configurations
- No discontinuity at the edges of order
  - Blaze-weighted mean of overlapping orders
  - Tapered weight fall-off at the end redder orders
  - By construction, cannot create artifacts smaller than 200 km/s
s.fits

[science]

- Comes into two flavours
  - Uniform in wavelength spacing (0.05 nm/pixel)
  - Uniform in velocity space (1 km/s/pixel)
- Convenient for cross-correlation and convolution over the entire domain
- Provided without telluric correction for AB, A, B and C
- Provided with telluric correction for AB
  - Deep water bands are filled with NaNs
A single *s.fits* file provides an impressive amount of information!
p.fits
[polarimetry]

- Only available for full polarimetric sequences
- Similar to e.fits in format: 4088x49 as a FITS image
  - Polarized spectrum and corresponding errors
  - Intensity spectrum and corresponding errors
  - Null spectrum
  - Blaze and wavelength spectra
Polarimetry

- From p.fits to LSD profiles (out of APERO)
- From LSD to longitudinal magnetic field
- Look for rotation period
Telluric subtraction

- oxygen
- water
- CO₂
- methane
Telluric subtraction

- Telluric lines are all over the place in the near-infrared
- The DRS provides telluric-subtracted spectra
- You must check the level of absorption at your favorite wavelength before interpreting the data!
Telluric subtraction

- Observe ~2 hot bright rapidly rotating stars every night
- Construct a library of line-of-sight absorbers
- Subtract telluric absorption from all science data
- ... does it work?
  - Yes, but it comes with some limitations
- Telluric absorption residuals increase slowly with absorption.
- In a perfect world, one would only get an increased Poisson noise due to lower transmission.
Improvement of DRS reduction over 2 years (from Guillaume Gaisné's PhD thesis)
- The fractional quality of telluric correction depends on absorption.
- For moderate absorption (0-50%), lines are subtracted to better than 1 part in 40.
  - If you have an SNR of 100 and the local absorption is <10%... you are all fine!
- Deep absorption (>50%) is more poorly corrected.
t.fits
[tellurics]

- Ext 1: Same as e.fits but after telluric correction
  - Only for AB; we assume that tellurics do not affect polarisation
- Ext 2: Recovered telluric absorption
  - NaNs where absorption is too strong for proper reconstruction
v.fits
[velocity]

- The cross-correlation function of RV data with a default CC mask
  - CCF line lists are expected to evolve
    - Ultimately one adapted to your spectrum
- Think of it as the mean line profile
  - 49+1 CCF profiles
CCF determination of RVs

- Choice of mask: best from telluric-corrected template of the same star
- Except for some targets with low SNR or small BERV excursion
- Use so-called neg masks: GL846_neg.mas
- 1 or 2 ST difference does not matter: 5 masks would be enough for all old M dwarfs
- Eder’s spirou-ccf package: see TOI-1759 paper for a description of the different steps
- Choice of line weights: line depth and mean RV dispersion around the line center within n FWHM
CCF template

- Normalize each spectrum to the continuum measured on the template.
- CCF template is the weighted mean of all spectra, where order weights combine RV content (Q factor) and dispersion in each order.
- RV is the velocity shift that best matches an individual CCF to the template CCF.
CCF outputs

- CCF gives RV, FWHM, bisector
- Allows discarding activity as a possible source of periodic variations
Line-by-line velocity

... follow-up on an idea suggested by X. Dumusque

Étienne Artigau
Back to the definition of RV measurements

Fundamental photon noise limit to radial velocity measurements

F. Bouchy, F. Pepe, and D. Queloz

Observatoire de Genève, 51 ch. des Maillettes, 1290 Sauverny, Switzerland

Received 27 March 2001 / Accepted 9 May 2001

Abstract. In the past 5 years, improvements in radial velocity measurements have led to discovery of extra-solar planets and progress in asteroseismology programs. Doppler measurements with high precision is close to the limit given by photon noise. In this paper the methodology to compute the fundamental limit of radial velocity measurement given by photon noise is presented and illustrated with a representative sample of synthetic solar-type stellar spectra. Stellar rotational broadening, instrumental spectral range as well as spectral resolution influences are also considered. This study is applied to two dedicated spectrographs in order to help the optimization of radial velocity programs. Current methods of Doppler calculation are discussed and compared.

Key words. techniques: radial velocities – instrumentation: spectrographs – stars: oscillations – stars: planetary systems
Lines?

- Domain between consecutive maxima in the spectrum
- Exactly the opposite of a CCF mask
- Typically ~10 pixels in length
- About 13,000 - 16,000 useful lines
  - Some lines are counted twice in order overlaps
Back to the definition of RV measurements

- Within one line, the velocity is simply a weighted sum of the differences between the spectrum and the template.
- The weight scales as the derivative of the spectrum square.
  - Basically the RV content of the pixel from the template.
- Bonus: by construction you get the uncertainty from the line.
- Extra-bonus: if you want, you get any higher-order derivative.
  - 2nd derivative directly linked to FWHM.
  - 3rd derivative linked to bisector tilt.
Behavior with tellurics

- Telluric residuals produce the most likely outliers
- Should be rejected by the LBL technique because of large uncertainties
- Allows getting velocities even in telluric-affected zones
- How are we doing?
**Drift correction**

- RV drifts during the night: this is the purpose of the reference fiber
- Drift is measured with respect to aft / mor calibrations
- Neutral density adds a small correction from bright to faint stars
- Computed from FP data: drift.rdb
Zero-point correction

- ZP may drift from run to run
- ZP may change with thermal cycles
- Use stable stars: common variations reflect the ZPC
- Choice of GP (but jumps are not continuous), rolling mean/median, or nightly average
Matern 3/2 vs rolling mean

© Thomas Vandalt