



**Universiteit  
Leiden**  
The Netherlands

## **Pushing the characterization of exoplanet atmospheres down to temperate rocky planets in the era of JWST**

Zieba, S.

### **Citation**

Zieba, S. (2024, June 25). *Pushing the characterization of exoplanet atmospheres down to temperate rocky planets in the era of JWST*. Retrieved from <https://hdl.handle.net/1887/3765836>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3765836>

**Note:** To cite this publication please use the final published version (if applicable).

# 3

## PACMAN: A PIPELINE TO REDUCE AND ANALYZE HUBBLE WIDE FIELD CAMERA 3 IR GRISM DATA

Sebastian Zieba and Laura Kreidberg

*Published in The Journal of Open Source Software, 2022.*

## Abstract

---

The *Hubble Space Telescope* (*HST*) has become the preeminent workhorse facility for the characterization of extrasolar planets. Launched in 1990 and never designed for the observations of exoplanets, the STIS spectrograph on *HST* was used in 2002 to detect the first atmosphere ever discovered on a planet outside of our solar system (Charbonneau et al. 2002).

*HST* currently has two of the most powerful space-based tools for characterizing exoplanets over a broad spectral range: The Space Telescope Imaging Spectrograph (STIS; installed in 1997) in the UV and the Wide Field Camera 3 (WFC3; installed in 2009) in the Near Infrared (NIR). With the introduction of a spatial scan mode on WFC3 (McCullough & MacKenty 2012; Deming et al. 2012) where the star moves perpendicular to the dispersion direction during an exposure, WFC3 observations have become very efficient due to the reduction of overhead time and the possibility of longer exposures without saturation.

For exoplanet characterization, WFC3 is used for transit and secondary eclipse spectroscopy, and phase curve observations. The instrument has two different grisms: G102 with a spectral range from 800 nm to up to 1150 nm and G141 encompassing 1075 nm to about 1700 nm. The spectral range of WFC3/G141 is primarily sensitive to molecular absorption from water at approximately 1.4  $\mu\text{m}$ . This led to the successful detection of water in the atmosphere of over a dozen of exoplanets (e.g., Deming et al. 2013; Huitson et al. 2013; Fraine et al. 2014; Kreidberg et al. 2014b; Evans et al. 2016). The bluer part of WFC3, the G102 grism, is also sensitive to water and most notably led to the first detection of a helium exosphere (Spake et al. 2018).

Here we present PACMAN, an end-to-end pipeline developed to reduce and analyze *HST*/WFC3 data. The pipeline includes both spectral extraction and light curve fitting. The foundation of PACMAN has been already used in numerous publications (e.g., Kreidberg et al. 2014a, 2018a) and these papers have already accumulated hundreds of citations.

---

### 3.1 Statement of need

Exoplanet spectroscopy with Hubble requires very precise measurements that are beyond the scope of standard analysis tools provided by the Space Telescope Science Institute. The data analysis is challenging, and different pipelines have produced discrepant results in the literature (e.g., Kreidberg et al. 2019b; Teachey & Kipping 2018). To facilitate reproducibility and transparency, the data reduction and analysis software should be open-source. This will enable easy comparison between different pipelines, and also lower the barrier to entry for newcomers in the exoplanet atmosphere field.

What sets PACMAN apart from other tools provided by the community, is that it was specifically designed to reduce and fit HST data. There are several open-source tools that can fit time series observations of stars to model events like transiting exoplanets, such as EXOFASTv2 (Eastman et al. 2019), *juliet* (Espinoza et al. 2019), *allesfitter* (Günther & Daylan 2019, 2021), *exoplanet* (Foreman-Mackey et al. 2021b,a), and *starry* (Luger et al. 2019). PACMAN’s source code, however, includes fitting models that can model systematics which are characteristic to HST data, such as the orbit-long exponential ramps due to charge trapping or the upstream-downstream effect. This removes the need for the user to write these functions themselves. PACMAN will also retrieve information from the header of the FITS files, automatically detect HST orbits and visits and use this information in the fitting models.

The only other end-to-end open source pipeline specifically developed for the reduction and analysis of HST/WFC3 data is *Iraclis*<sup>1</sup> (Tsiaras et al. 2016b). Another open-source pipeline that has been for example used as an independent check of recent results presented in Mugnai et al. (2021) and Carone et al. (2021) is *CASCADE*<sup>2</sup> (Calibration of trAnsit Spectroscopy using CAusal Data). For a more detailed discussion of *CASCADE* see Appendix 1 in Carone et al. (2021).

### 3.2 Outline of the pipeline steps

The pipeline starts with the *ima* data products provided by the Space Telescope Science Institute that can be easily accessed from MAST<sup>3</sup>. These files created by the WFC3 calibration pipeline, *calwf3*, have already several calibrations applied (dark subtraction, linearity correction, flat-fielding) to each readout of the IR exposure.

In the following we highlight several steps in the reduction and fitting stages of the code which are typical for HST/WFC3 observations:

- **Wavelength calibration:** We create a reference spectrum based on the throughput of the respective grism (G102 or G141) and a stellar model. The user can decide if he or she wants to download a stellar spectrum from MAST or use a black body spectrum. This template is used for the wavelength

<sup>1</sup><https://github.com/ucl-exoplanets/Iraclis>

<sup>2</sup><https://jbouwman.gitlab.io/CASCADE/>

<sup>3</sup><https://mast.stsci.edu/search/hst>

calibration of the WFC3 spectra. We also determine the position of the star in the direct images which are commonly taken at the start of *HST* orbits to create an initial guess for the wavelength solution using the known dispersion of the grism. Using the reference spectrum as a template, we determine a shift and scaling in wavelength-space that minimizes the difference between the template and the first spectrum in the visit. This first exposure in the visit is then used as the template for the following exposures in the visit.

- **Optimal extraction and outlier removal:** PACMAN uses an optimal extraction algorithm as presented in Horne (1986) which iteratively masks bad pixels in the image. We also mask bad pixels that have been flagged by `calwf3` with data quality  $DQ = 4$  or  $512^4$ .
- **Scanning of the detector:** The majority of exoplanetary *HST*/WFC3 observations use the spatial scanning technique (McCullough & MacKenty 2012) which spreads the light perpendicular to the dispersion direction during the exposure enabling longer integration times before saturation. The *ima* files taken in this observation mode consist of a number of nondestructive reads, also known as up-the-ramp samples, each of which we treat as an independent subexposure. Figure 3.1 (left panel) shows an example of the last subexposure when using spatial scanning together with the expected position of the trace based on the direct image.
- **Fitting models:** PACMAN contains several functions to fit models which are commonly used with *HST* data. The user can fit models like in Equation 3.1 to the white light curve or to spectroscopic light curves. An example of a raw spectroscopic light curve and fitting Equation 3.1 to it, can be found in Figure 3.2. Here are some examples of the currently implemented models for the instrument systematics and the astrophysical signal:
  - systematics models:
    - \* visit-long polynomials
    - \* orbit-long exponential ramps due to charge trapping: NIR detectors like *HST*/WFC3 can trap photoelectrons (Smith et al. 2008), which will cause the number of recorded photoelectrons to increase exponentially, creating typical hook-like features in each orbit
  - astrophysical models:
    - \* transit and secondary eclipse curves as implemented in `batman`
    - \* sinusoids for phase curve fits
    - \* a constant offset that accounts for the upstream-downstream effect (McCullough & MacKenty 2012) caused by forward and reverse scanning

---

<sup>4</sup>for a list of DQ flags see <https://wfc3tools.readthedocs.io/en/latest/wfc3tools/calwf3.html>

A typical model to fit an exoplanet transit in *HST* data is the following (used, for example, by Kreidberg et al. (2014a)):

$$F(t) = T(t) (c S(t) + k t_v) (1 - \exp(-r_1 t_{\text{orb}} - r_2)), \quad (3.1)$$

with  $T(t)$  being the transit model,  $c$  ( $k$ ) a constant (slope),  $S(t)$  a scale factor equal to 1 for exposures with spatial scanning in the forward direction, and  $s$  for reverse scans,  $r_1$  and  $r_2$  are parameters to account for the exponential ramps.  $t_v$  and  $t_{\text{orb}}$  are the times from the first exposure in the visit and in the orbit, respectively.

- **Parameter estimation:** The user has different options to estimate best fitting parameters and their uncertainties:
  - least squared: `scipy.optimize`
  - MCMC: `emcee` (Foreman-Mackey et al. 2013a)
  - nested sampling: `dynesty` (Speagle 2020)
- **Multi-visit observations**
  - PACMAN has also an option to share parameters across visits.
- **Binning of the light spectrum:** The user can freely specify the bin numbers or locations. Figure 3.1 (right panel) shows the resulting 1D spectrum and a user-defined binning.

Figure 3.1 and 3.2 show some figures created by PACMAN during a run using three *HST* visits of GJ 1214 b collected in GO 13201<sup>5</sup> (Bean 2012). An analysis of all 15 visits was published in Kreidberg et al. (2014a). The analysis of three visits here using PACMAN, is consistent with the published results.

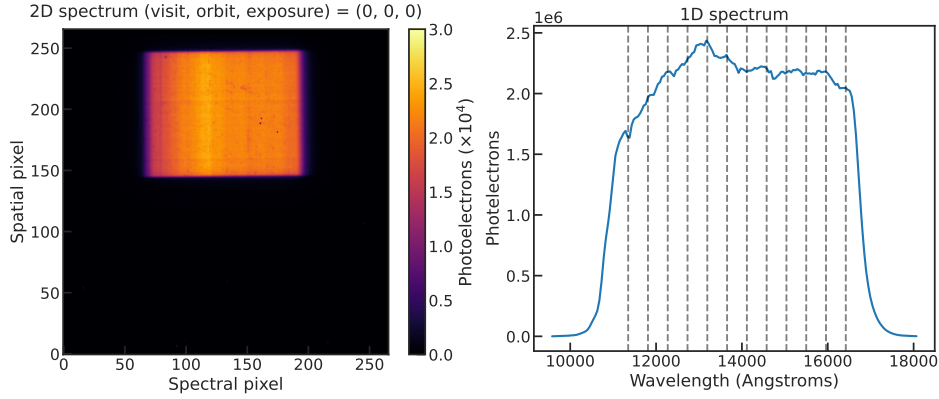
### 3.3 Dependencies

PACMAN uses typical dependencies of astrophysical Python codes: `numpy` (Harris et al. 2020), `matplotlib` (Hunter 2007a), `scipy` (Virtanen et al. 2020) and `astropy` (Astropy Collaboration et al. 2013, 2018, 2022). Other dependencies required for the fitting stage depending on the model and sampler being run are: `batman` (Kreidberg 2015), `emcee` (Foreman-Mackey et al. 2013a), `dynesty` (Speagle 2020), and `corner` (Foreman-Mackey 2016). For the barycentric correction, PACMAN accesses the API to JPL’s Horizons system<sup>6</sup>. If the user decides to use a stellar spectrum for the wavelength calibration, PACMAN will download the needed fits file from the “REFERENCE-ATLASES” high level science product<sup>7</sup> hosted on the MAST archive (STScI Development Team 2013).

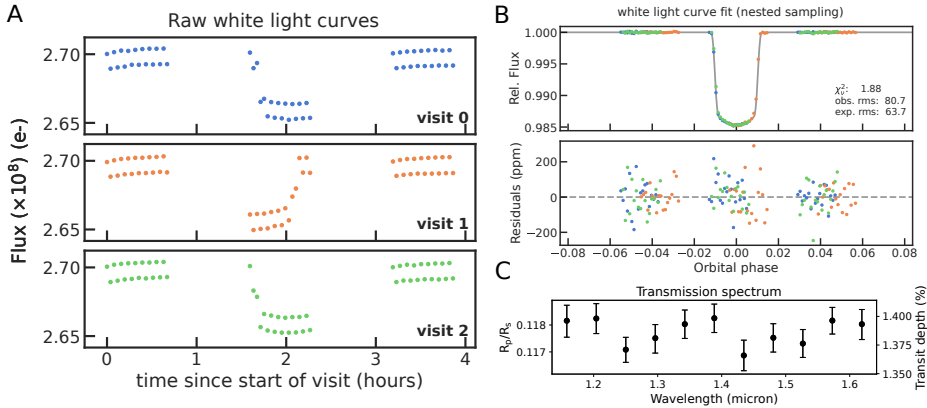
<sup>5</sup>[https://archive.stsci.edu/proposal\\_search.php?id=13021&mission=hst](https://archive.stsci.edu/proposal_search.php?id=13021&mission=hst)

<sup>6</sup>[https://ssd-api.jpl.nasa.gov/obsolete/horizons\\_batch.cgi.html](https://ssd-api.jpl.nasa.gov/obsolete/horizons_batch.cgi.html)

<sup>7</sup><https://archive.stsci.edu/hlsps/reference-atlases/cdbs/grid/>



**Figure 3.1:** *Left panel:* a typical single exposure showing the raw 2D spectrum. *Right panel:* 1D spectrum after the use of optimal extraction including vertical dashed lines showing the user-set binning to generate spectroscopic light curves.



**Figure 3.2:** *panel A:* raw white light curves for each of the three visits. One can clearly see the constant offset between two adjacent exposures due to the spatial scanning mode. *panel B:* white light curve with the best astrophysical model fit using Equation 3.1. *panel C:* the transmission spectrum after fitting 11 spectroscopic light curves revealing the flat spectrum of GJ 1214 b as published in Kreidberg et al. (2014a).

## 3.4 Documentation

The documentation for PACMAN can be found at [pacmandocs.readthedocs.io](https://pacmandocs.readthedocs.io)<sup>8</sup> hosted on ReadTheDocs<sup>9</sup>. It includes most notably, a full explanation of every parameter in the *pacman control file* (pcf), the API, and an example of how to download, reduce and analyze observations of GJ 1214 b taken with *HST*/WFC3/G141.

## 3.5 Future work

The following features are planned for future development:

- The addition of fitting models like phase curves using the open-source Python package SPIDERMAN (Louden & Kreidberg 2018).
- Orbit-long ramp fitting using the RECTE systematic model<sup>10</sup>.
- Limb darkening calculations for users wanting to fix limb darkening parameters to theoretical models in the fitting stage.
- Extension to WFC3/UVIS data reduction.

## Acknowledgements

We acknowledge B. Zawadzki for the creation of the PACMAN logo. We also acknowledge the comments and contributions by I. Momcheva to PACMAN.

---

<sup>8</sup><https://pacmandocs.readthedocs.io/en/latest/>

<sup>9</sup><https://readthedocs.org/>

<sup>10</sup><https://recte.readthedocs.io/en/latest/>