

# PrepSpec User Manual

Dr. Viraja Khatu

December 2022

## 1 Introduction

PrepSpec<sup>1</sup> (developer: K. Horne) is a spectral analysis software that fits the continuum and emission lines in the input spectra with a composite model through an iterative process while correcting for any relative deviations in the calibrated wavelength and flux scales of the spectra. It outputs the model fits and the aligned spectra after correcting for the discrepancies in the relative wavelength and flux scales. Processing spectra through PrepSpec becomes important while analyzing variability in the light output from astronomical objects. An example of such a program is reverberation-mapping (RM) study of active galactic nuclei (AGN), where velocity line profiles of specific emission lines are measured from spectral data to determine changes in flux over a period of time. Measuring subtle changes in fluxes requires the spectra to be free of variations from relative differences in calibration processes, as much as possible.

In this manual, we use the long-slit, optical spectra of Markarian 142 (Mrk 142), a super-Eddington AGN in the local universe (redshift,  $z = 0.045$ ), taken with the Gemini Multi-Object Spectrograph (GMOS) on the Gemini North Telescope (Gemini) in Hawai'i (Khatu et al., 2022). The target was observed in two different configurations – B600 grating and two slits,  $0.75''$  (narrow slit) and  $5.00''$  (wide slit) – along with the comparison star (calibration star) in the same slit for accurate spectrophotometric calibration. In the long-slit mode, two chip gaps of the GMOS detector appear in the spectra. Data were collected from February through June 2019 for 33 epochs with two exposures in each configuration every night. The original purpose of the Mrk 142 Gemini campaign was to perform velocity-resolved reverberation mapping to probe the structure of the broad-line region (BLR) in a super-Eddington AGN system and hence the need for the narrow-slit data. However, due to insufficient data, we revised the aim to only determining an ultraviolet (UV) time lag for  $H\beta$  emission with concurrent photometric observations in the UV. The wide-slit data were used to correct for potential slit losses from the narrow slit. To achieve the science goal of the campaign, we processed the narrow-slit spectra through PrepSpec to obtain accurate flux measurements for the target. In what follows, we demonstrate the use of PrepSpec for the Mrk 142 RM data set (hereafter, “this work” or “example data set”).

PrepSpec comes with several dependencies. The code is written in FORTRAN and hence needs a FORTRAN compiler (`gfortran`, `ifort`, `gcc`, `g77`, or `ifort`). The recommended one is `ifort` as it compiles faster than the others and has the advantage of using `-parallel` switch that allows tuning the code on multiple cores with less efforts. PrepSpec uses PGPlot for graphing purposes.

---

<sup>1</sup>Find current version of PrepSpec at [PrepSpec link](#).

The following input files are required for **PrepSpec** in the same directory as the code.

1. Wavelength- and flux-calibrated one-dimensional (1D) spectra files: ASCII files of calibrated spectra, with three columns – wavelength [ $\text{\AA}$ ], flux density,  $1\sigma$  (where,  $\sigma$  is the standard deviation) error bars on flux density.
2. Timing file: ASCII file with two columns – filenames of the input spectra, Heliocentric Julian Date<sup>2</sup>. In this work, HJD values were calculated with the HJD calculator developed by the British Astronomical Association.

## 2 Installation

**PrepSpec** package contains the following four files.

1. *prepspec.for*: This is the main program that contains **PrepSpec** routines.
2. *misc.for*: This program contains all sub-routines that are used by *prepspec.for*.
3. *prepspec.exe*: This programs contains the executable file to run **PrepSpec**.
4. *prepspec.com*: This program contains commands that compile and link the code to generate the *prepspec.exe* file. It is required to generate a new *prepspec.com* file if the default one made available with the code is incompatible with the system. To generate *prepspec.exe*, in a Terminal window, change directory to the location of the software on the system and run `source prepspec.com`.

## 3 Preparing Your Data for PrepSpec

To prepare data for **PrepSpec**, the user must bear in mind the following points.

1. Input spectra are wavelength and flux calibrated: **PrepSpec** is specially designed to correct for shifts in wavelength and flux scales of calibrated spectra. Therefore, the input spectra must be wavelength and flux calibrated to approximately close wavelength and flux scales. Significant differences in wavelength solutions or flux levels likely indicate calibration issues that must be resolved before supplying spectra to **PrepSpec**. Similarly, no flux or saturated values (e.g., due to cosmic-ray hits or bad pixels) result in outliers in spectra and affect the modeling process in **PrepSpec**. Such outliers can be typically removed by using sigma-clipping algorithms during spectral reduction. Accurate spectrophotometric calibration is desirable for the input spectra and can be achieved by capturing a given calibration standard star in the same slit as the target. Target and calibration star in the same slit also allows optimal correction of slit losses if spectra were taken with narrow slits (see discussion on slitloss correction for Mrk 142 example spectra in this section).
2. Input spectra have no gap regions: One caveat is that **PrepSpec** is not currently designed to handle gaps in spectra. Therefore, chip gaps or regions with no data cause issues during

---

<sup>2</sup>Heliocentric Julian Date is the Julian Date of a given observation calculated by regarding that the observation was done from the center of the Sun and is typically calculated by a computer.

modeling in `PrepSpec`. The GMOS detector gaps in the example data set were replaced by median or simulated flux values (see discussion on simulated data below) before supplying them to `PrepSpec`. We recognize that replacement with simulated data is not ideal as it introduces information that is not original to the data. However, using simulated data was the only possible fix to the chip gaps in the Mrk 142 spectra. An appropriate solution to this type of data-gap issue is to take spectra with slightly different central wavelengths at the time of observations.

3. The more the number of input spectra, the more the information available to `PrepSpec` for modeling: A large number of input spectra implies that `PrepSpec` has access to information in the spectra in great details. Details about the spectra are crucial during the modeling stage. If good spectra from multiple exposures are combined or discarded, information is lost or *blurred* which limits the software from modeling the data optimally. Therefore, providing as many spectra as possible to `PrepSpec` is important.

To illustrate the data preparation process, we describe below the key steps we followed in preparing the Mrk 142 (wavelength-calibrated and extracted 1D) spectra for `PrepSpec`.

1. Correcting for GMOS chip gaps, and residuals from cosmic-ray correction and sky subtraction:

In the spectra showing strong/sharp residuals from cosmic-ray correction or sky subtraction, we replaced the values of the affected pixels either with median values from the neighboring region or with simulated data. For a region of  $<5$  pixels, we used the local median values. For a region of  $\geq 5$  pixels, we first applied linear interpolation in the region and then simulated data values assuming a Gaussian distribution of points with standard deviation equal to the noise in the interpolated values. Finally, we assigned the simulated data an uncertainty twice as high as the noise from interpolation. We used larger uncertainties for corrected regions so that they will be weighed lesser than the good data values in further analysis.

2. Recovering distorted regions in the spectra:

For a few spectra, we noticed bump-like features – “distorted” regions – that affected the shape of the spectra in those regions. We applied a recovery algorithm to achieve the original shape of the spectra in the distorted regions. Since we had taken two exposures every night, we used the other, good spectrum of a particular night with similar count levels – the reference spectrum – to recover the spectrum containing the distorted regions. We employed a low-order cubic-spline to estimate the shape of the reference spectra in the distorted regions and scaled those regions of the affected spectra to the estimated shape. For the nights when the spectra from the two exposures had slightly different count levels, we used the mean wide-slit target spectrum as the reference spectrum (also used as the reference spectrum for slitloss-correction – see below).

3. Correcting for (wavelength-dependent) slit losses:

Narrow slits, used to achieve higher resolution for the data, end up blocking some of the light from the source. Therefore, narrow-slit spectra are typically supplemented with wide-slit spectra to be able to correct for any flux losses from the narrow slits (referred to as “slit losses”). For Mrk 142, we used the wide-slit observations to correct for slit losses due to the narrow slit. We applied this correction to both the calibration star and the target spectra as follows:

- (a) We selected some of the brightest wide-slit calibration star spectra from our data set and generated a mean wide-slit calibration star spectrum – the reference calibration star spectrum.
- (b) We calculated one-to-one ratios of the reference calibration star spectrum to the narrow-slit calibration star spectra at each epoch. We then fit a cubic-spline of order 2 to the ratios to obtain a sensitivity curve across the spectrum. The user can also try using binned ratios (ratios of binned spectra) as long as the sensitivity function is able to reasonably estimate the wavelength dependence of the slit losses. For the example data set, one-to-one ratios provided a better estimation of the slit losses than binned ratios; hence, we used the former. One should expect higher (lower) slit losses on the shorter-wavelength or bluer (longer-wavelength or redder) end of the spectra due to scattering. The order of the sensitivity function should be low as we aim to fit only the overall behavior of the slit losses.
- (c) We scaled up the calibration star spectra at each epoch by dividing the spectra with the sensitivity curves determined for the respective epochs.
- (d) We repeated steps (a), (b), and (c) for the target spectra.

## 4 Get Set... Go...!

To get going with **PrepSpec**, simply type `prepspec.exe` on the command line to start working with **PrepSpec**. In some cases, the system needs the full path to the `prepspec.exe` file to start running **PrepSpec**.

### 4.1 Preparing PrepSpec for Your Data

**PrepSpec** can be run interactively or through a set of commands organized in a shell script. For beginners, the interactive way is recommended first to get an idea of how the code works. Figure 1 shows the terminal display after **PrepSpec** is launched. The initial steps to prepare the software for modeling variations in the spectra are outlined below.

1. Select the appropriate number corresponding to the target AGN from the list that appears after launching **PrepSpec** (see Figure 1). If the target AGN is not in the list of known AGN, type a number *less than -999*. The example data set is of Mrk 142. Therefore, we selected “27” at the command prompt (enclosed in a red box at the bottom of Figure 1).
2. Enter the name of the AGN at the next command prompt. Figure 2 highlights this with the first purple box. If an AGN was selected from the list, its name appears in square brackets at the prompt. Simply type `<return>` to accept the prompted name.
3. At this point, the software searches for a redshift file beginning with the name “redshift\*” (see Figure 2) in the working directory. If a redshift file is not found, a command prompt appears to enter the redshift of the selected AGN. If the selected AGN is from the available database, the redshift is known and appears in square brackets at the prompt. Enter a new value to overwrite the known one or type `<return>` to accept the known value. The example Mrk 142 spectra are in rest-frame wavelengths. Hence, we entered a redshift of “0.0” (or simply “0”!).

```

*****
Welcome to PrepSpec !
Prepares AGN Spectra for MEMECHO mapping
Keith Horne copyright 1990 ...
This Version: 2021 Jun
*****

AGN known to PREPSPEC:
 1 Arp 151      z 0.021100  [27 Mrk 142      z 0.044590]
 2 Mrk 6        z 0.018813  28 Mrk 841      z 0.036420
 3 Mrk 50       z 0.023433  29 I Zw 1      z 0.060875
 4 Mrk 290      z 0.030500  30 0007+00    z 0.315637
 5 Mrk 817      z 0.031450  31 0013+00    z 0.362352
 6 Mrk 335      z 0.025785  32 0124+14    z 0.339038
 7 Mrk 1501     z 0.089300  33 0129+14    z 0.364851
 8 NGC 3227     z 0.003500  34 0153-10    z 0.360472
 9 NGC 3516     z 0.008836  35 0247-07    z 0.352421
10 NGC 4051     z 0.002340  36 1218+02    z 0.327091
11 NGC 5548     z 0.017175  37 1224+04    z 0.357530
12 NGC 7469     z 0.015880  38 1324+03    z 0.304300
13 NGC 5273     z 0.003420  39 1346+09    z 0.336333
14 PG2130      z 0.062900  40 1435+02    z 0.304887
15 3C 120       z 0.033000  41 1438+02    z 0.375225
16 Mrk 352      z 0.014864  42 1537+05    z 0.306885
17 Mrk 926      z 0.046860  43 1609+05    z 0.338217
18 Mrk 1048     z 0.043140  44 2322-09    z 0.370985
19 Mrk 1298     z 0.061900  45 Mrk 509      z 0.034397
20 NGC 4593     z 0.008344  46 Fairall 9     z 0.045500
21 NGC 4151     z 0.003262  47 1H2106-099   z 0.026515
22 Mrk 876      z 0.129000  48 PKS 2004-447  z 0.240000
23 MCG+08-11-011 z 0.020500  49 PG 2304+043  z 0.042000
24 NGC 2617     z 0.014200  50 Mrk 279      z 0.030601
25 3C 382       z 0.057900  51 Mrk 590      z 0.026090
26 Mrk 374      z 0.020500  52 J2222+2745  z 2.805000

call system( pwd )
/Users/viraja/prepspec

Enter AGN name or number from table above.
(-N selects SDSS-RMID_N for N=0...999)
Select AGN (<-999 if not in list) [0] 
```

Figure 1: Terminal display after `PrepSpec` is launched highlighting useful information (reference information enclosed in purple boxes, information required to proceed in green boxes, and the next step at the command prompt in red boxes) for the user. The copyright and version details are shown at the top left followed by a list of 52 known AGN. The working directory is printed below the list where outputs will be saved (if outputs are written to files). The first command prompt is at the bottom of the figure. The user may enter an AGN number from the known AGN list or simply enter any number *less than -999* if the AGN is not in the list.

- The software then searches for a *Timing* file with an extension ending in “.\*jd” in the working directory. For the example data, there exists one timing file named “mrk142\_075.hjd” in the working directory as displayed in Figure 2 with a green box. We typed <return> to accept the existing timing file for the example data set. The user may type another filename to overwrite the filename option that appears in square brackets at the prompt. Providing the timing file will load the input spectra in `PrepSpec`. Figures 3 and 4 display the terminal output after the 64 Mrk 142 spectra are loaded.

By default, the software also searches for two other files containing the half-widths of typical broad absorption-line (BAL) windows in AGN. The BAL windows inform the model the velocity ranges that are affected by absorption so that they are weighed lower than the

```

Enter AGN name or number from table above.
(-N selects SDSS-RMID N for N=0...999)
Select AGN (<-999 if not in list) [0] 27
AGN name: [Mrk 142]

Look for a redshift file:
call system( head redshift* )
head: redshift*: No such file or directory
call system( \rm delete_z.lis )
rm: delete_z.lis: No such file or directory
call system( ls redshift*.dat > delete_z.lis )
ls: redshift*.dat: No such file or directory
Open(          47 , delete_z.lis, FORMATTED, OLD )
** LOADLIST aborted.          0
call system( \rm delete_z.lis )
Redshift files found:          0

Redshift: [ 4.459000063180923E-002] 0.0

call system( ls -l *.jd )
-rw-r--r-- 1 viraja staff 3682 23 Sep 18:41 mrk142_075.hjd
call system( ls *.jd > delete.lis )
call system( head delete.lis )
mrk142_075.hjd
Open(          47 , delete.lis, FORMATTED, OLD )
nlist          1
1 mrk142_075.hjd

Timing file (2-column ascii : 1=specfilename, 2=HJD) : [mrk142_075.hjd]

```

Figure 2: Terminal display after an AGN number is selected in `PrepSpec`. The annotated text follows the same color coding as Figure 1. The software requires the name of the AGN (at the top of the figure) and then its redshift. The user is required to enter a redshift filename or the redshift value for the selected AGN. `PrepSpec` needs a *Timing* file (see text for details) containing information about the spectra to be loaded. The software found one such file in its search “`mrk142_075.hjd`”. The user may enter a file from the search results or another file available in the directory that contains the required information.

rest of the spectrum during model fitting. If the files are not found, a warning message appears. However, these inputs are not required if the input spectra are devoid of absorption similar to the Mrk 142 example spectra. Figure 4 shows that `PrepSpec` did not find the files “`ps_nlr_blr.dat`” and “`ps_bal.dat`” for the example data set. In our example case, the warning message there can be safely ignored.

5. Type H or `<return>` at this point to view a list of parameters with current settings and a list of commands available to proceed in `PrepSpec`. The output is displayed in Figure 5 with a few commonly used options enclosed in purple boxes.
6. Type U to set appropriate units for flux. The options here are “FLAM” (flux density per

```

Timing file (2-column ascii : 1=specfilename, 2=HJD) : [mrk142_075.hjd] mrk142_075.hjd
Open( 47 , mrk142_075.hjd, FORMATTED, OLD )
Input list      64 of max    1000
   1  astrcgn20190206S0177tgt_p_s_l_z_s.dat 2458520.916358091
   1  1  2458520.9163580999  astrcgn20190206S0177tgt_p_s_l_z_s.dat
   2  astrcgn20190206S0178tgt_p_s_l_z_s.dat 2458520.9180463078

Timebase : HJD
----- Spectrum      1 of      300 -----
Spectrum File : astrcgn20190206S0177tgt_p_s_l_z_s.dat 2458520.916358091
Chronological time: 2458520.916358091
3-column data files.
NCOL      3 HJD COL      1 NROW      17000
Open( 49 , astrcgn20190206S0177tgt_p_s_l_z_s.dat, FORMATTED, OLD )
astrcgn20190206S0177tgt_p_s_l_z_s.dat 2458520.916358091
input HJD= 2458520.9163580999 = 2458520 + 0.91635809093713760
fixed HJD= 2458520.9163580999 = 2458520 + 0.91635809093713760
nint( 2458520.9163580999 ) = 2450000
JDSub : 2450000 JDD 0 THJD 8520.9163580909371
NWAVE= 4123 4350.0571264651189 6349.5195310604640
Dispersion(A) 0.485070944 0.485070944 kms 22.9026184 33.4295883
** DISPERSION JUMPS      0 REMAIN IN SPECTRUM      1
Bad data:      0 of 4123 0.00000000 % 
Good data: 4123 of 4123 100.000000 %
dat 1.24789596 5.98616941 sig 4.64810543E-02 0.885649860
INPUT SPECTRUM      1 STOWED SPECTRUM      1
NT      1 HJD 8520.9163580909371
NW      4123 4350.0571264651189 6349.5195310604640
----- Spectrum      2 of      300 -----
Spectrum File : astrcgn20190206S0178tgt_p_s_l_z_s.dat 2458520.9180463078
Chronological time: 2458520.9180463078
3-column data files.
NCOL      3 HJD COL      1 NROW      17000

----- Last Spectrum      64 of max      300 -----
-----
** LOST SPECTRA (NO TIME):      0 of      64
** LOST SPECTRA (NO DATA):      0 of      64
max bad pixels 50 %
** KEPT SPECTRA ( GOOD ):      64 of      64

```

Figure 3: Terminal display when spectra are being loaded in `PrepSpec`. The annotated text follows the same color coding as Figure 1. The software detected 64 input spectra that are loaded in the order in which they appear in the *Timing* file. Spectra with no data are lost by `PrepSpec`. The number of lost and retained spectra appears at the end of the loading process.

unit wavelength given by  $\text{erg s}^{-1} \text{cm}^{-2} \text{\AA}^{-1}$ ) or “MJY” (milliJanskys). A dex value must be specified for scaling the axis (default is 0). We set the units to FLAM with a dex of -15 for the example data.

7. Type W to calculate the pixel shifts (to the nearest pixel) along the wavelength axis with reference to the [O III]  $\lambda 5008$  line. This corrects for any shifts in the dispersion solutions of the calibrated input spectra suggesting whether or not there is a need to tweak the wavelength jitter in the modeling process. Figure 6 shows small pixel shifts (<7 pixels) for the Mrk 142 example spectra.
8. Type B to apply a different binning for the spectra. The user is required to specify whether to use Observed (specified by “O”) or Rest (“R”) wavelengths for the binning and the new wavelength range and pixel binning. The user may consider binning the data for high-resolution spectra if the high resolution is not required for further analysis. The user may

```

** WARNING: UNEVEN SPECTRUM NPIX      4106      4127
Nw    4123  (   4350    6349 ) astrcgn2019020650177tgt_p_s_l_z_s.dat
Nw    4124  (   4350    6349 ) astrcgn2019022650124tgt_p_s_l_z_s.dat
Nw    4122  (   4350    6349 ) astrcgn2019030350101tgt_p_s_l_z_s.dat
Nw    4123  (   4350    6349 ) astrcgn2019030950301tgt_p_s_l_z_s.dat
Nw    4125  (   4350    6349 ) astrcgn2019031650250tgt_p_s_l_z_s.dat
Nw    4120  (   4350    6349 ) astrcgn2019032650022tgt_p_s_l_z_s.dat
Nw    4124  (   4350    6349 ) astrcgn2019032750033tgt_p_s_l_z_s.dat
Nw    4126  (   4350    6349 ) astrcgn2019032950086tgt_p_s_l_z_s.dat
Nw    4125  (   4350    6349 ) astrcgn2019033150020tgt_p_s_l_z_s.dat
Nw    4123  (   4350    6349 ) astrcgn2019040350115tgt_p_s_l_z_s.dat
Nw    4124  (   4350    6349 ) astrcgn2019040450145tgt_p_s_l_z_s.dat
Nw    4126  (   4350    6349 ) astrcgn2019040550032tgt_p_s_l_z_s.dat
Nw    4123  (   4350    6349 ) astrcgn2019040650198tgt_p_s_l_z_s.dat
Nw    4125  (   4350    6349 ) astrcgn2019040750113tgt_p_s_l_z_s.dat
Nw    4106  (   4350    6349 ) astrcgn2019040850045tgt_p_s_l_z_s.dat
Nw    4125  (   4350    6349 ) astrcgn2019040950024tgt_p_s_l_z_s.dat
Nw    4124  (   4350    6349 ) astrcgn2019050250114tgt_p_s_l_z_s.dat
Nw    4125  (   4350    6349 ) astrcgn2019050750086tgt_p_s_l_z_s.dat
Nw    4127  (   4350    6349 ) astrcgn2019050850056tgt_p_s_l_z_s.dat
Nw    4126  (   4350    6349 ) astrcgn2019050950076tgt_p_s_l_z_s.dat
Nw    4125  (   4350    6349 ) astrcgn2019051350084tgt_p_s_l_z_s.dat
Nw    4126  (   4350    6349 ) astrcgn20190512450049tgt_p_s_l_z_s.dat
Nw    4127  (   4350    6349 ) astrcgn20190512500116tgt_p_s_l_z_s.dat
Nw    4121  (   4350    6349 ) astrcgn2019052650029tgt_p_s_l_z_s.dat
Nw    4126  (   4350    6349 ) astrcgn2019060150124tgt_p_s_l_z_s.dat

** Spectra with bad pixels:          0  of   64

Nt      64  of max      300 =  21.3333340 %
Nw range     4106      4127  of max     17000 =  24.2764702 %
Nw x Nt    264128  of max   5100000 =  5.17898035 %
Unrepaired dispersion jumps:        0       0

Line window half-widths:
** OPENFILE FAILED :(
** Path( ps_nlr_blr.dat ) = ps_nlr_blr.dat.dat
** Error opening file.
** ps_nlr_blr.dat
** LOADIC aborted.
NLR      500 T BLR      3000 T km/s

BAL windows from ps_bal.dat
NCOL      3 HJD COL      1 NROW      500
** OPENFILE FAILED :(
** Path( ps_bal.dat ) = ps_bal.dat.dat
** Error opening file.
** ps_bal.dat
** LOADIC aborted.
BAL windows:      0
PrepSpec > █

```

Figure 4: Terminal display after loading spectra in `PrepSpec`. The annotated text follows the same color coding as Figure 1. The software shows a message about uneven total number of pixels of the spectra, but no bad pixels are detected. The software searches for two files, “`ps_nlr_blr.dat`” and “`ps_bal.dat`”, containing velocity ranges affected by broad-absorption lines (BALs) in the spectra (see text for more details). However, BAL windows are not required if the input spectra are devoid of absorption. The software safely proceeds with assuming no BAL windows for the spectra (highlighted with a green box at the bottom of the figure).

also consider binning if a specific range of the full spectrum is needed for analysis. We discarded the ends of the Mrk 142 spectra as the calibration toward the ends is not reliable. Moreover, there is greater noise in the data toward the blue end than the rest of the spectrum. Because the blue end of the spectrum is beyond the region of interest for the Mrk 142 data analysis, we only used the region from 4430 Å to 6300 Å. We retained the bin size of 1 from the original spectra.

- At this stage, `PrepSpec` is ready to enter the modeling mode. Type `M` to do so. The output will display some pre-defined quantities for analysis and finally display a list of options that the user can set before beginning the modeling process. Typing `<return>` anytime also lists

```

PrepSpec >
A ... AGN Mrk 142
U ... Units FLAM 10^(-15)
Z ... redshift Z = 0.000000000000000
T ... Times 64 HJD 8520.91602 8635.75488 mrk142_075.hjd
W ... Wavelength alignment 2 0
B ... Bin pixels 4126 4350.37354 6349.97998
+ ... Add constant to spectra
F ... Fake error-bar spectra
E ... systematic Error 0.0000000 percent
J ... Jitter 0.0000000 pixels
C ... Continuum fits
L ... Light curves Nt = 64
S ... Spectra ISPLIT = 1
V ... Velocity at wavelength 0.000000000000000
I ... Image display
M ... Model spectral variations
D ... Dump trailed-spectrogram to disk file
Q ... Quit
PrepSpec > 

```

Figure 5: Terminal display showing available options to start processing spectra in `PrepSpec`. The annotated text follows the same color coding as Figure 1. Commonly used options are highlighted in the figure – option “U” allows setting appropriate flux units, “W” allows aligning spectra along the wavelength axis, and “B” allows binning of spectra (see text for more details). The user can enter the modeling mode with option “M”.

all available options for the user. Figure 7 shows the terminal output when we enter the modeling mode. The green boxes show some of the preset parameters in this mode. Various broad lines can be set to variable with the “L” option. This is done by assigning a negative expansion factor for specific lines. We defined the H $\beta$  and He I  $\lambda 5877$  lines for the example data set to be variable as we are interested in measuring their velocity profiles from the spectra. Thus, the two broad lines appear as “VarBLR” components in Figure 6. If the expansion factor of a line is set to 0, the line is excluded from the final model. Type “K” to adjust the velocity widths of broad and narrow lines. We used the preset values for the example data set. Typing “H” allows to display HJD on light-curve plots instead of the spectrum number. This can be useful if the user plans to write out the output at the end of the modeling process. Note that the average spectrum for the example data shows a contribution from Fe II which will be fit during iterative modeling. `PrepSpec` assumes the I Zwicky 1 template model (Véron-Cetty et al., 2001) to model the Fe II component while fitting the spectra.

## 4.2 Modeling Spectral Variations

The iterative process for modeling spectral variations in `PrepSpec` takes into account contributions from various components. The model components that can be fit include:

- Average spectrum (specified by “A”): Mean of the input spectra.

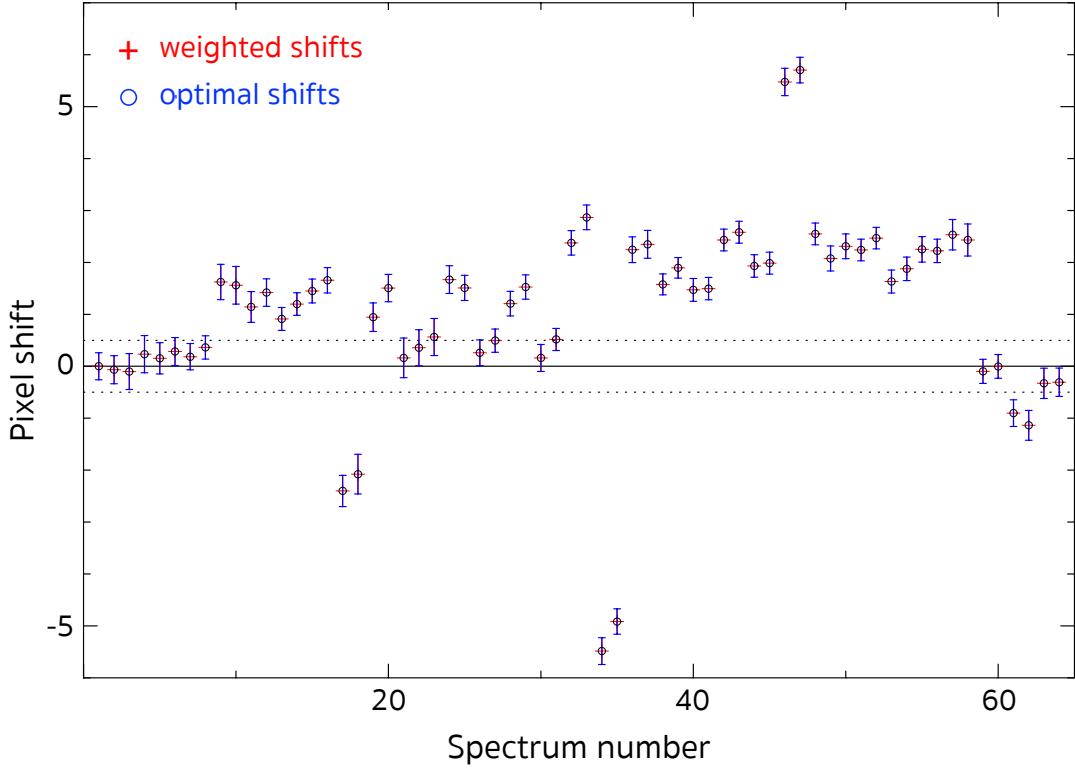


Figure 6: Pixel shifts for the 64 Mrk 142 example spectra during wavelength alignment in `PrepSpec`. The shifts are calculated to the nearest pixel using the [O III]  $\lambda 5008$  narrow line as the reference line.

- Continuum: Variations in the continuum emission from the accretion disk modelled as a polynomial defined by  $\log \lambda$  with time-dependent coefficients
- Flux jitter: Time-dependent photometric corrections, defined by a scale factor and a lower order polynomial as a function of wavelength, to minimize the scatter of narrow emission-line fluxes relative to their median.
- Broad-line variations: Variability in the broad emission-line features defined as splines. The breadth of the features is computed from the number of nodes of the piece-wise cubic spline needed to trace the broad profiles.
- Narrow-line variations: The modelled broad-line profiles are subtracted from the data values to obtain the narrow-line features, which are then estimated with spline functions.
- Wavelength jitter: Inter-spectra deviations in the calibrated wavelengths estimated to the fraction of a pixel by aligning the spectra with respect to the [O III]  $\lambda 5008$  line. The devi-

```

PrepSpec modelling.
Missing times          0 of      64 =  0.000000000 % 
Missing waves          0 of     3800 =  0.000000000 % 
Missing data           104 of   243200 =  4.27631587E-02 %
Bad wavelengths        0 of     3800 :)
Redshift check using selected Median NLR velocities.
 4960 v  6.53955030    dz  2.18135883E-05  znlr  2.1813588318764232E-005 +/- -3.33564049E-06
 5008 v -1.73968196    dz -5.80295364E-06  znlr -5.8029536376125179E-006 +/- -3.33564049E-06
 Mean v  2.39993429 +/- 5.85430098    dz  8.00531780E-06
** WARNING: NO ERROR BARS ON NLR VELOCITIES.
z  0.0000000000000000 -> 8.0053177953232080E-006 +/- 1.95278426E-05

-----
PrepSpec Modelling Nw      3800 Nt      64 Nt*Nw      243200
 Good data      243096 Bad      104 =      0 % 
 Parameters      3943 DoF      239153 =     98 % 
chimedw  0.887046337 chimedt  0.875336289 sigscale  0.875336289

M ... Model: cont  3 phot  1 shift  1 blur  1 absn  0
G ... Gaussian Prior( p(t) ) rms  20.0000000 % 
L ... Lines:      73 active  12
K ... KM/S of WinHW: NLR      500 T BLR      3000 T
 Target: NLR      0 T BLR      0 T
 V1-4: NLR 0 0 0 0 BLR 0 0 0 0
R ... Reset model
A ... Avg spectrum  1.83488882    rms  0.480267674    nfit      3800
| Fe2 0.235 = 27 % vdisp 154 +/- 0 km/s
C ... Continuum c0  0.00000000    rms  0.00000000    nfit      7
| cont( 1 ) avg 0.00000000 ) rms 0.00000000
| cont( 2 ) avg 0.00000000 ) rms 0.00000000
N ... NLR spec  0.00000000    2.86764097    npix      255 nfit      0
| NLR stiffness 1.00000000
B ... BLR spec  0.00000000    4.13558185E-02 npix      836 nfit      0
| VarBLR components: 2 of 15 npix      444
| Hb 4863 spec 0.00000000 0.1193082079 npix 201 nfit 0
| HeI 5876 5877 spec 0.00000000 1.73437726E-02 npix 243 nfit 0
J ... Joint tweak of all lightcurves.
F ... Flux jitter f0  0.00000000    rms  0.000000000    0
| mnmx 0.00000000 0.00000000 erb 0.000000000 0.000000000
W ... Wavelength jitter x0  0.00000000    rms  0.000000000    0
S ... Seeing jitter s0  0.00000000    rms  0.000000000    0
T ... Optical depth mn 0.00000000    mx  0.000000000 0 tweak F
I ... Iterate      0 AC6J
P ... Plot (or 1..8 for 1 to 8-panel plot)
V ... Velocity Profile Plots
X ... Chi^2 / ( 243200 - 3943 ) = 0.8051 => 0.9431
E ... Error bar scaling F Chi^2/dof => 0.943090677
# ... Sigma Clip      5 < -6.00000000    0 > 6.00000000
D ... Debug Diagnostics F
LW ... Line Width      2
CH ... Character Height 1.2999995
H ... HJD (vs sequence number) on lightcurve plots T
O ... Output to disk files
Q ... Quit PrepSpec Modelling

PrepSpec Modelling > 

```

Figure 7: Terminal display showing available options in the modeling mode in PrepSpec. The annotated text follows the same color coding as Figure 1. The software allows the user to set several parameters for processing the input data set. Some commonly used options are highlighted with purple boxes (see text for details). For Mrk 142 example spectra, the software has already estimated the contribution from Fe II emission (specified below the “A ... Avg spectrum” option in the figure). Furthermore, the H $\beta$  and He I lines of interest are set as variable for the model fitting process.

ations are measured by scaling the first derivative of the wavelength with the data. If spectra do not show large pixel shifts during initial wavelength alignment, this component

can be excluded from the final model.

- Seeing: Corrects for any changes in the line widths due to changes in the seeing conditions from one epoch to the next. The correction is estimated from the second derivative of the model and defined by a constant or a lower order polynomial as a function of wavelength.

Each of the model components can be fit either individually or jointly, specified by “J” at the end of a given composite model. For instance, “ACJ” performs joint fitting of the average and continuum together, or “FWJ” jointly fits the flux and wavelength deviations for the input spectra. The best practice is to start with one component, typically, the average spectrum and then add additional components to the model. Each time a new component is added, the model takes  $\sim$ 10–30 iterations to re-fit the data. In theory, the sequence of the fit components does not matter as long as the global minimum of the badness-of-fit function is attained. However, if the fit encounters a local minimum, the fit can keep oscillating in the region and may never be able to converge.

**PrepSpec** determines the best-fitting model by accessing the Bayesian Information Criterion (BIC) and reduced  $\chi^2$  ( $\chi_{\nu}^2$ ) statistics for every model. The BIC method involves maximizing the likelihood function<sup>3</sup> of a given model while balancing the function with both, the number of data points (N) and the number of parameters (P) used to define the model. BIC is given by  $\chi^2 + P \ln N$  and uses Occam penalty for complex models. Thus, the goal is to use the fewest possible parameters to appropriately describe the data while penalizing the model for the number of parameters used. **PrepSpec** further computes  $\Delta\text{BIC}$  as the difference between the BIC values of the current and previous models. The smaller the change in BIC, the closer the current model is to the best fit. The relative probability is also calculated for comparing models with different number of parameters, and a rise in the relative probability is desirable till it reaches a constant value, indicating that the model is not evolving any further. A good model yields  $\chi_{\nu}^2 \sim 1$ . A  $\chi_{\nu}^2$  smaller (larger) than the expected value of 1 implies over-fitting (under-fitting) or large (small) variance in the errors. In the **PrepSpec** modeling mode, the performance of the models can be visually inspected by typing “X”, which displays the BIC and  $\chi^2$  plots in-between iterations. Initially, fitting data interactively helps gain an understanding of which model works reasonably well for a given data set. We cover below the interactive modeling procedure for the Mrk 142 Gemini spectra.

For our example data set, we began by entering “I” for iteration. The user must provide the model at the first command prompt and the number of iterations at the next prompt every time they start the iteration process. We first performed 10 iterations with component “A” and visualized the model performance by typing “X”. The software asks for the selection of a plotting device, and here, the user can select either “X” (again!) to visualize the relative probability plots in a PGPLLOT window or “C” to save the plots as a PostScript (PS) file in the working directory. During the iteration process, it is more convenient and advisable to visualize the plots for inspection in the PGPLLOT window rather than saving and opening a PS file after every series of iterations. If the model shows convergence (both the  $\Delta\text{BIC}$  and the  $\chi^2$  ( $\chi_{\nu}^2$ ) equal to zero), the user can proceed to adding a second component to the model. If there were no convergence, the user must perform more iterations with the same model until the model converges. For the Mrk 142 data, the model converged after 10 iterations with “A”, and hence we proceeded to perform 10 iterations with the “ACJ” composite. Recollect that “ACJ” allows the joint fitting of the average spectrum and

---

<sup>3</sup>Likelihood function signifies how likely can a model, defined with a set of parameters, describe the given data. When the probability of the model describing the data is the maximum, we refer to it as the maximum likelihood function.

continuum to the input spectra. After the “ACJ” model convergence with 5 more iterations, we performed 20 iterations with “ACWJ”. The user may choose to do only 10 iterations depending on the behavior of their data. To gauge the optimum number of iterations with additional components to the model requires several runs through the entire modeling process. We performed 15 more iterations with “ACWFJ” after which the model converged. We then performed 25 iterations with “ACWFJ” and finally 50 iterations with “ACWFBJ”. The final model yielded a  $\chi^2_\nu$  of 0.780, which indicates overfitting of the data, possibly indicating inaccurate error bars larger than the scatter in the data. Figure 8 shows  $\chi^2$  and BIC plots evaluating the performance of the model. The middle panel indicates relative probability of the model risen to a constant value suggesting no further evolution in the model. We excluded fitting the seeing component to the example spectra as we found that tweaking the seeing did not improve the model performance.

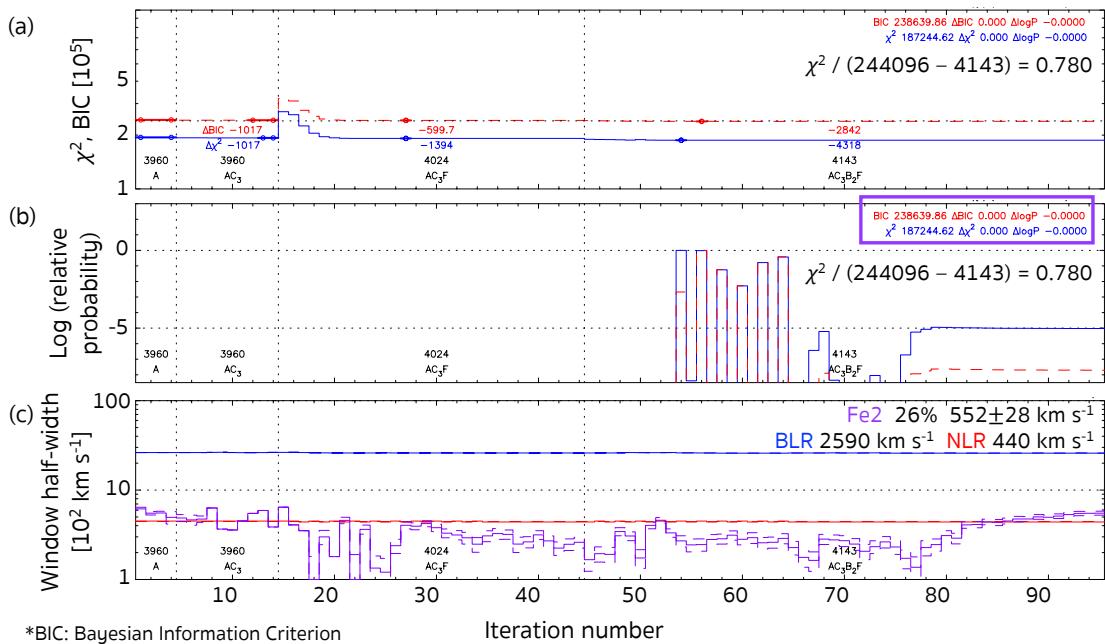


Figure 8:  $\chi^2$  and Bayesian Information Criterion (BIC) plots indicating the performance of the model through the 95 iterations during the modeling stage. Panel *a* shows the evolution in  $\chi^2$  (blue solid curve) and BIC (red dashed curve) through the iterations, whereas panel *b* shows logarithm of the relative probability of change in  $\chi^2$  (blue solid curve) and BIC (red dashed curve) as the model progresses. In panel *b*, the change in relative probability ( $\Delta \log P$ ) of 0 (purple box) indicates that the model has converged yielding a reduced  $\chi^2$  ( $\chi^2_\nu$ ) of 0.780, and there will be no further evolution in the model. Panel *c* shows the velocity half-widths of the Fe II emission (purple dashed curve) and the emission lines in the broad-line region (BLR; blue solid curve) and the narrow-line region (NLR; red solid curve).

Figure 9 displays the final model (panel *a*) and residuals (in the units of  $\sigma$ ; panel *b*) in grayscale. Typing “1” followed by a specific letter – “G” for grayscale model and “E” for residuals in the units

of  $\sigma$  – in the modeling mode allows the user to display or save a single plot. The grayscale model appears turbulent in the region blueward of  $\sim 4700 \text{ \AA}$  likely due to large variance in the region. The model is stable from the  $H\beta$  to the red end of the spectrum. The grayscale residuals for the Mrk 142 data show horizontal wiggles (dark or bright but not gray regions) prominently evident in some spectra. With visual inspection, we identified that the regions with wiggles were replaced before flux calibration by simulated data to correct for residuals from cosmic-ray correction or sky subtraction. The replacement with simulated data may have affected the performance of the model in the concerned regions. Another possible reason for such horizontal wiggles is when a higher order function is used during data reduction. In the Mrk 142 spectral reduction procedure, we used a higher order spline during flat-fielding to avoid discontinuities in the calibrated spectra across detector gaps. However, we observed no anomalous behavior in the regions with wiggles when we performed a visual inspection of the spectra processed through `PrepSpec`.

Figure 10 shows the mean (panel *a*) and root-mean-square (RMS; panel *b*) spectra displayed by typing “1” followed by “M” (for the mean) and “B” (for the RMS), respectively. The composite model for the mean spectrum shows overplotted continuum emission from the accretion disk and the Fe II pseudo continuum. The Fe II model fits less optimally indicating that a better template is required to carefully trace the sharp Fe II features in the spectra. Probably, using another Fe II template or tweaking individual lines in the existing template will result in better performance by the Fe II component. The adjusted BLR and NLR windows are indicated at the top of both the mean and the RMS plots,  $2600 \text{ km s}^{-1}$  and  $450 \text{ km s}^{-1}$  respectively. The model fit to the RMS spectrum nicely traces the  $H\beta$  and He I emission features. However, the contamination of He II  $\lambda 4687$  with Fe II emission in the region makes it difficult to estimate the shape of the He II line. The residual RMS (RMSx) approaching to zero from  $\sim 5340 \text{ \AA}$  to  $\sim 5420 \text{ \AA}$  is indicative of the detector chip gap.

In the modeling mode, `PrepSpec` also allows the user to output several other plots and data. The user can display or save to file the velocity profile plots of both the narrow and the broad lines as well as the variable broad lines by typing “V”. The user can also output data in the output mode “O”, including data for the tweaked spectra (with and without the narrow lines), the mean and RMS spectra, continuum light curves, BLR light curves (for the specified variable lines), and the line velocity profiles.

## 5 Troubleshooting

With the limited time that I spent with `PrepSpec`, I dealt with the following issues, some of which are very commonly encountered by users.

1. While preparing the `prepspec.exe` file for your specific machine, you may need certain compiler switches in the `prepspec.com` file to be turned OFF or ON based on your operating system and version. For my Mac operating system version 10.12.6, I included the `-m64` switch (in addition to the default ones) to indicate that it was a 64-bit operating system.
2. The syntax to define aliases is different for different operating systems and must be accordingly set to successfully compile the code. By default, the `prepspec.com` file contains aliases for the FORTRAN compilers `g77`, `gfortran`, and `ifort`. However, if any of the compilers is not installed on your local machine, then the user may remove the specific command for the alias.

3. Specifying an incorrect path to a certain location is one of the most common errors during programming. Thus, the appropriate path to the PGPLOT directory is required to successfully compile `PrepSpec`.
4. While loading the *Timing* file, complete names of the spectra files with their extension (e.g., “.dat” or “.txt”) must be specified. If file extensions are not included, the software reports a failure to load the data. One or more spectra can be ignored by including an “!” at the beginning of the specific row in the *Timing* file.
5. The spectral data provided to `PrepSpec` must be in ascending order of wavelengths. If the wavelengths are in random or descending order, the software attempts to predict the values where they are not in increasing order based on the values before and after that point. The predicted values can create abnormalities in the data that are undesirable.

## References

- Khatu, V. C., Gallagher, S. C., Horne, K., et al. 2022, ApJ, submitted [Manuscript #: AAS43725]
- Véron-Cetty, M. P., Véron, P., & Gonçalves, A. C. 2001, A&A, 372, 730, doi: 10.1051/0004-6361:20010489

$$\chi_{\nu}^2 = \chi^2 / (250688 - 4317) = 0.782 \quad 3917 \text{ columns} \times 64 \text{ rows}$$

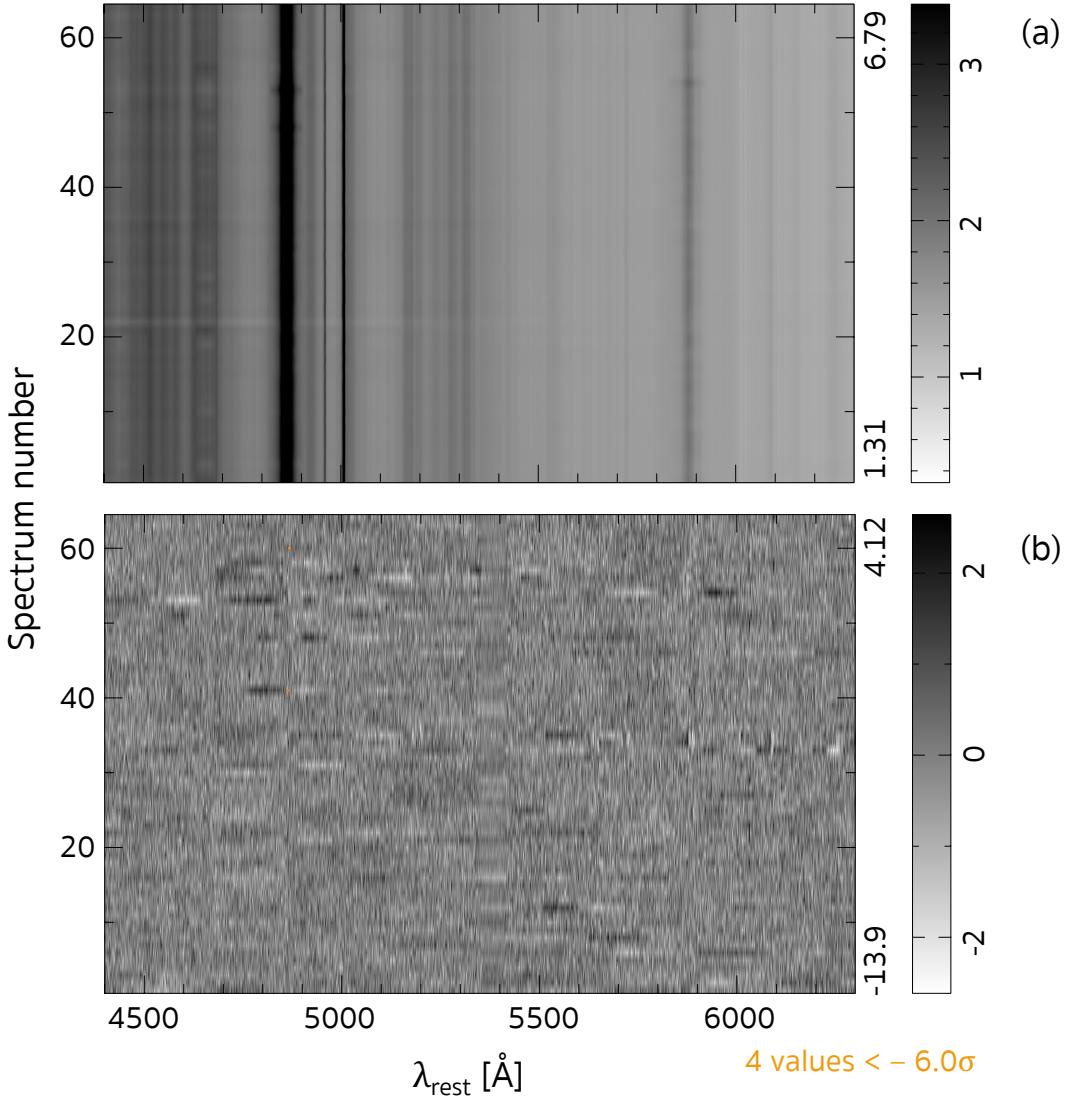


Figure 9: PrepSpec model including all components (panel *a*) and residuals (*data – model* in units of standard deviation,  $\sigma$ ; panel *b*) with a reduced  $\chi^2$  ( $\chi^2_\nu$ ) of 0.782 for 64 narrow-slit Mrk 142 spectra. In panel *a*, dark regions indicate strong emission lines of H $\beta$  and [O III], whereas the weaker He I lines and Fe II emission appears as less prominent features. In panel *b*, each row represents a single exposure spectrum (where multiple exposures at a given epoch are not yet combined). The horizontal wiggles strongly evident in some spectra are likely the result of either replacing values with simulated data in those regions or using a higher order function during flat-fielding (see text for more details). Four points where the values were less than  $6\sigma$  are indicated with orange dots in the residuals plot. The smeared region from  $\sim 5345 \text{ \AA}$  to  $\sim 5420 \text{ \AA}$  is one of the chip gaps of the GMOS detector where simulated data was added during reduction.

Fe2 30%  $577 \pm 26$  km s $^{-1}$  BLR 2600 km s $^{-1}$  NLR 450 km s $^{-1}$

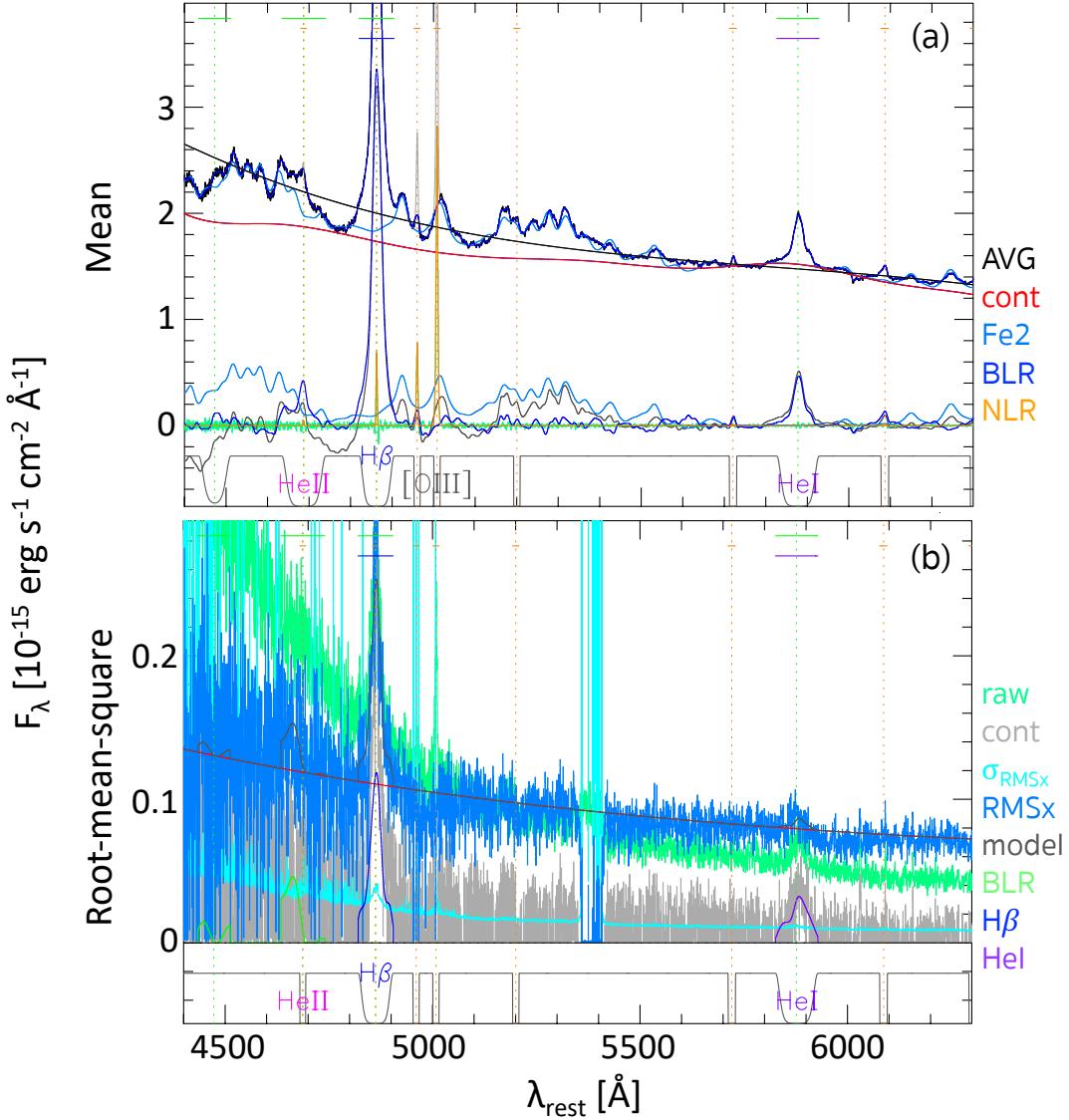


Figure 10: Mean (panel *a*) and root-mean-square (RMS; panel *b*) of 64 narrow-slit Mrk 142 Gemini spectra processed through PrepSpec showing model fits with individual components. In panel *a*, the composite model (dark blue curve), including the components shown at the bottom of the plot – average spectrum (AVG; black), continuum (cont; red), Fe II (Fe2; faint blue), broad-line region (BLR; dark blue), and narrow-line region (NLR; orange), is overlaid on the mean spectrum (black curve). The broad (narrow) emission lines are indicated with green (orange) dotted vertical lines as well as with green (orange) solid horizontal dashes. The broad lines of H $\beta$  (blue label) and He I (purple label) are marked with solid horizontal dashes. In panel *b*, model (model; dark gray curve) fit to the residual RMS spectrum (RMSx; blue curve) includes the components: continuum (cont; red curve) and BLR (green curve at the bottom of the plot). The raw RMS spectrum is the upper green curve. The BLR component comprises the broad lines of H $\beta$  (blue bump around  $\sim 4862$  Å) and He I (purple bump around  $\sim 5877$  Å) shown at the bottom of the plot. The broad He II emission feature at  $\sim 4687$  Å (box-like feature in the model) is contaminated with Fe II and hence difficult to fit given the noise in the region. The deviation in the residual RMS spectrum ( $\sigma_{\text{RMSx}}$ ; cyan curve) shows large values in the region of the GMOS detector chip gap from  $\sim 5345$  Å to  $\sim 5420$  Å.