University of Groningen

Kapteyn Astronomical Institute

Wavelet Transform of mid-IR spectra of protoplanetary disks and water slab models

Author:
Steven Ho

Supervisor:
Prof. Dr. Inga Kamp

July 9, 2018

Abstract

Aims: We aim to use wavelet transform to de-noise a spectral emission, thus disentangling spectral features from the noise, and automatically determine the continuum location.

Methods: The multi-scale á'trous algorithm using a B_3 cubic spline was used to wavelet transform the signal. De-noising was performed with hard threshold filtering which eliminates any wavelet coefficients below a certain noise threshold.

Conclusions: The wavelet transform works well within limitations, both de-noising spectra and recovering the continuum location. However, it should first be compared to other methods to evaluate its effectiveness and efficiency.

Contents

1 Introduction **1**

1.1 Aim of this research 3

1.2 Details of datasets examined 3

2 Comparison of water lines of the disk and slab models **6**

2.1 Adjustments to data prior to comparison 6

2.2 Free parameters of the models 7

2.3 Rotational and vibrational lines intermixed with each other 10

2.4 Results of the comparison 13

3 The Wavelet Transform **14**

3.1 Theoretical introduction 14

3.2 Mathematical details 14

3.3 Choice of à'trous algorithm and B3 cubic spline wavelet 15

3.4 Code example 16

3.4.1 à'trous algorithm wavelet transform 16

3.4.2 Reconstructing the spectra from wavelet coefficients 17

4 Method **17**

4.1 Filtering and de-noising 20

4.2 Determining the location of the continuum 22

5 Results **22**

5.1 De-noising 22

5.1.1 Aaron's FLiTs models 23

5.1.2 Fred's water slab models 29

5.1.3 ProDiMo slab models 32

5.1.4 Spitzer data 35
5.2 Continuum determination 39
5.2.1 Aaron's FLiTs models 41
5.2.2 Fred's water slab models 44
5.2.3 ProDiMo slab model 45
5.2.4 Spitzer data 46

6 Discussion **51**
6.1 Viability of the Wavelet Transform 51
6.1.1 De-noising 51
6.1.2 Continuum 51
6.2 Limitations of the Wavelet Transform 51

7 Conclusion **52**

1 Introduction

Protoplanetary disks are a byproduct of star formation (Dominik, 2015). When a molecular cloud reaches a large enough critical mass, density or radius known as Jean's criterion, the molecular cloud collapses due to self-gravitation and begins the process of forming a star at its core. Due to the conservation of angular momentum, the molecular cloud is flattened from a sphere to a disk, the materials of this protoplanetary disk around the young star will eventually form planets, forming a protoplanetary system (Antonellini, 2016). This process is shown in Figure 1.

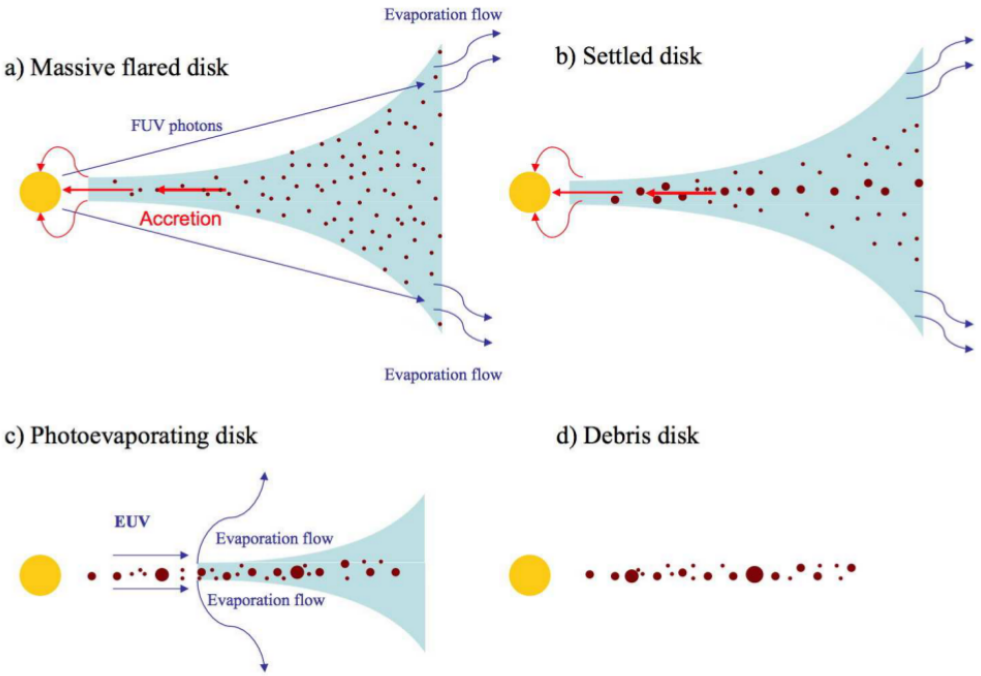

Figure 1: A molecular cloud collapses into a star and a protoplanetary disk. For a detailed explanation, refer to the source of this figure, Williams and Cieza (2011).

These disks are almost entirely gaseous, with 99% of their initial mass being gas, and the other 1% being dust. Spectral emission lines tell us the physical structure, chemical evolution, and composition of the disks (Dionatos, 2015), allowing us to understand more about the process and dynamics of planetary formation. This knowledge helps us understand our history and contribute towards our answering questions such as how the gas and dust evolve, as well as how planets and planetary systems form.

With existing data from missions such as *Spitzer* continuing to grow, and upcoming *James Webb Space Telescope (JWST)* mission, we are only looking at more and more data being collected. While the processing of observation data mostly done automatically, the analysis

of such data is still mostly performed manually. Automated methods to discover interesting sources within the wealth of data we will collect could greatly increase time efficiency.

At low spectral resolution, emission lines from molecules are usually heavily blended, making it difficult to distinguish from which molecule an emission line originates. Furthermore, spectral features are often hidden amongst the noise. This research examines the use of a wavelet transform (hereafter referred to as WT) to de-noise low resolution *Spitzer* spectral from disks, thus disentangling spectral features from noise, and determine the location of the continuum for more accurate molecule identification. If successful, the methods examined in this research could potentially be used for future missions such as the JWST mission.

As an example, the determination of the continuum is usually done manually by fitting a cubic spline through the 'valleys' of emission peaks. Finding the continuum using this method can be inaccurate in the case of when a 'forest' of underlying weak emission lines is present, as suggested by models of molecular emission. Furthermore, it can be prone to human error. An automatic method to evaluate the location of the continuum could eliminate the possibility of this error.

In Fig. 2, we note the discrepancy between the red fitted water emission lines and the blue simple disk model. We find that when fitting a disk model to protoplanetary disk spectra, the simple ProDiMo disk model underestimates water lines, though they are within the correct order of magnitude. An interesting thing to determine is to which extent the disk and slab models differ, thus whether simple slab models could be used to fit observational spectra, or if the more complicated disk models are always required.

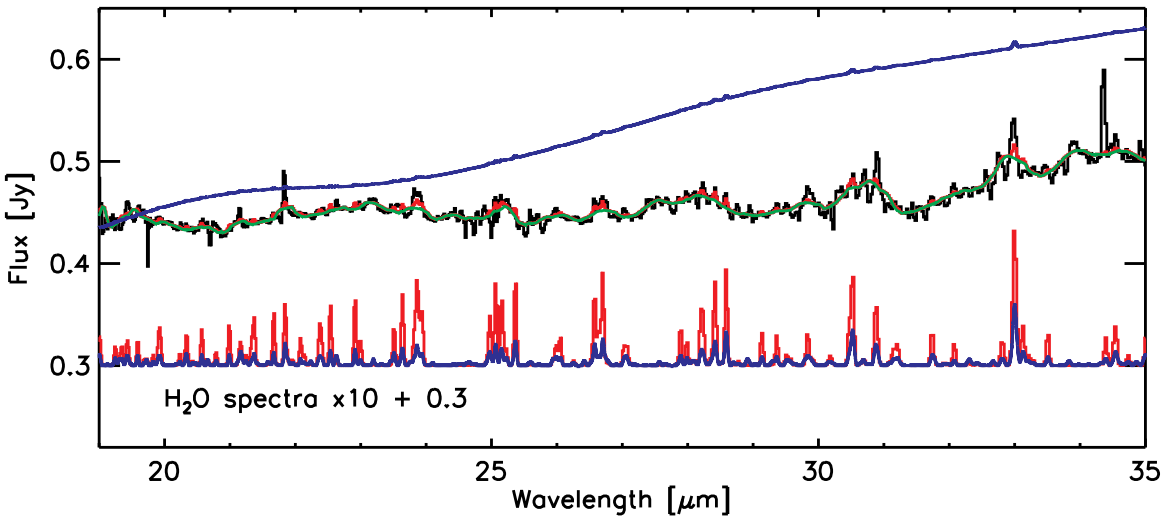

Figure 2: Black: observed Spitzer spectrum. Green: Fitting of continuum. Red: Fitting of H_2O spectrum. Blue: a ProDiMo disk model for FT Tauri with a simple escape probability fluxes. Note that they are not as accurate as the FLiTs disk models (run on top of ProDiMo models) present in this paper. Source: Fred Lahuis, private communication, FT Tau model from Garufi et al., 2014.

1.1 Aim of this research

The first goal is to understand the difference between disk and slab models and see if slab models can be used for bulk fitting of a true observational spectrum, which is similar to the disk model spectrum with temperature and column density gradients.

The second is to test the WT as an automated data analysis technique, using it to denoise spectra and locate the continuum. This is to search for weak lines in observer *Spitzer* and future *JWST* spectra.

1.2 Details of datasets examined

The datasets examined consists of three models and one actual dataset. The datasets will first be listed then briefly described:

- A continuum-subtracted, reference spectrum provided by Aaron Greenwood (see Chapter 3.4 of Greenwood, 2018) from his FLiTs (Fast Line Tracing System) models, convolved to a spectral resolution of 600. They vary in wavelength from 8-25 μm.
- Local Thermodynamical Equilibrium (LTE) water slab models provided by Fred Lahuis (Lahuis et al., in prep.), originally obtained from Pontoppidan et al. (2010). These models originally ranged in wavelength from 10-40 μm, but were filtered to 7-30 μm, and were convolved from a spectral resolution of 2260 to 600.
- ProDiMo (Protoplanetary Disk Model, presented in Woitke et al., 2009) LTE water slab models at a Hydrogen column density of 10^{22} cm^{-2} (with an inner disk water abundance of 10^{-4}, thus a water column density of 10^{18} cm^{-2}) and temperature of 450K. The wavelength was filtered to 7-30 μm, and convolved to a spectral resolution of 600.
- *Spitzer*-IRS (InfraRed Spectrograph, R = 600) spectra of T Tauri disks, consisting of the following stars: CoKu Tauri/4, DM Tauri, TW Hydrae, GM Aurigae, and UX Tauri A (Rigliaco et al., 2015).

FLiTs disk models
The FliTs disk models (full details of this parametrized disk structure, representative of a typical T Tauri star is found in Woitke et al., 2009, 2016) were run on top of an existing 2D disk structure from the thermo-chemical disk modelling code ProDiMo (Woitke et al., 2009; Kamp et al., 2010). These models have radial and vertical temperature gradients from the inner disk to the outer disk.

Fred's slab models
Fred's 1D LTE water slab models have assumed a single constant volume density of the species, and a single gas temperature. The only variables are then column density and temperature. An emitting surface area of 1 AU was assumed to obtain fluxes. These models

work well for emission lines which originate from regions of similar density and temperature, but the models break down if those lines originate from a wide region of the disk (Kamp, 2015).

ProDiMo slab model

The ProDiMo thermo-chemical disk modelling tool (Woitke et al., 2009; Kamp et al., 2010) calculates the physical, thermal, and chemical structure of protoplanetary disks using an iterative solution of the gas heating/cooling balance and chemistry Kamp (2015).

Fig. 3 shows an overview of all models of water lines across the mid-IR. Note that the FLiTs disk model has stronger lines at both ends of the wavelength range, while the simpler Fred's slab model and ProDiMo slab models. While this difference in shape is expected between the disk and two slab models, the difference in shape between Fred's slab model and the ProDiMo slab model is rather unexpected. This is examined in §2. Also note the difference in scale of the intensities.

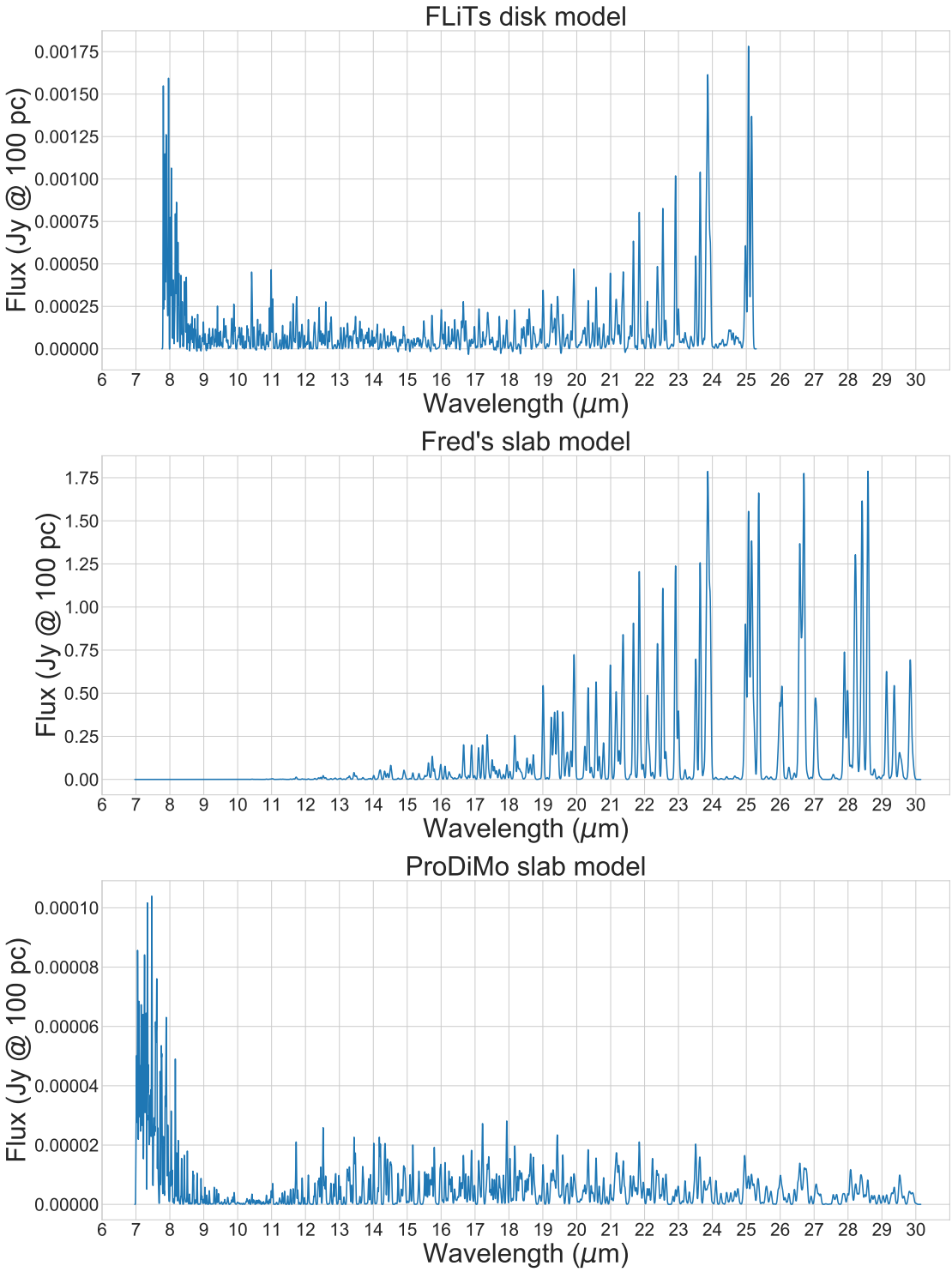

Figure 3: An overview of all water models in the mid-IR. Note the difference in scale.

2 Comparison of water lines of the disk and slab models

At a low spectral resolution of 600, where the wavelength λ examined in this study is in the mid-IR between 7-30 μm, molecular emission lines are heavily blended together.

Water emission lines of the FLiTs disk model was compared against Fred's slab model. This comparison was done to see whether we can use reference slab models when fitting observed spectra, or if disk spectra is always needed. The FLiTs model in this case represents the case where a water spectrum is calculated from a structure with a radial and vertical temperature and density gradient. In comparing these models, we can determine whether the slab model with a constant temperature and column density can match those complex spectra, which serves as a test of the method employed so far in the analysis of *Spitzer* spectra.

In our comparison of the FLiTs disk model with Fred's slab model we examine the water emission lines and find lines which are only present in one model but not the other. A closer examination of the energy levels using the HITRAN (High-resolution Transmission) 2009 database (Rothman et al., 1998, 2013), and using the identified energy level transitions from Pontoppidan et al. (2010); Daniel et al. (2011) reveal that the rotational and vibrational lines are quite intertwined with each other.

2.1 Adjustments to data prior to comparison

The flux of the ortho- and para- species of water from both the FLiTs disk models and ProDiMo models were simply added together. Note that simply combining water together in this way can cause line blends to be over-estimated (see Fig. 3.1 in §3.3 of Greenwood, 2018). Intensities of all datasets were distance-adjusted to Jy @ 100 pc. All models were convolved to a spectral resolution of 600 using the same convolution tool, as provided by Aaron Greenwood.

The intensities of both Fred's and the ProDiMo slab models were interpolated onto a new wavelength grid of 10000 points prior to convolution to avoid under-sampling.

The ProDiMo intensities were initially given in erg s^{-1} sr^{-1} cm^{-2}, but as the convolution of Fred's slab models was done in wavelength space, we obtain a factor μm from the convolution. Thus we must multiply by the following factor to convert the ProDiMo flux to Jy, distance-adjusted to 100 pc.

$$sr \ \frac{\lambda^2}{c \ d\lambda}$$

Here sr is steradian, λ is wavelength, c is the speed of light, and dλ is the width of our

wavelength bin. Here dλ at each wavelength bin is given by

$$d\lambda = \frac{\lambda}{R}$$

This comes from the definition of spectral resolution R = $\lambda/\Delta\lambda$. The spectral resolution R is set to 600 for all models. The steradian factor is

$$sr = \frac{A}{d^2}$$

Here the emitting surface area A of the water slab model was 1 AU, and the flux was distance adjusted to d = 100 pc.

2.2 Free parameters of the models

The slab models were based on an isothermal slab approximation, where the free parameters are emitting surface area A, temperature T, and Hydrogen column density N_H. Salyk et al. (2011) finds that for an LTE model, the line strength is mainly determined by the emitting surface area. Temperature and column density are around T $\approx$ 450K and $N_H \approx 10^{22}$ cm^{-2} for all sources, depending on fitting method and wavelength range. The estimates were found by fitting the H_2O emission spectra from a sample of protoplanetary disks around T Tauri stars observed by the *Spitzer*-IRS with simple LTE slab models. The fitting was performed by minimising the difference for 65 line peaks between the observed data and model. Fixing the parameters T and N_H around these estimates leaves only the emitting surface area as the only free parameter in comparing the models. Note that despite this best fit range for T and N, there is no correct choice, as the slab models are far too simple compared to the disk environment.

The emission line just under 20 micron was selected as the anchoring point when adjusting the scaling factor during the comparison of the models. This is around the middle of the overlap of the wavelength range, allowing us to see the behaviour of the two models relative to one another toward each end of the mid-IR wavelength range.

Fig. 4 shows comparison of the disk model with Fred's water slab model at a temperature of 450K with the disk model multiplied by a scaling factor of 1540. We notice that the disk model has more emission lines at the shorter wavelengths, which can be explained by the fact that the disk model contains data from a range of temperatures of the protoplanetary disk model.

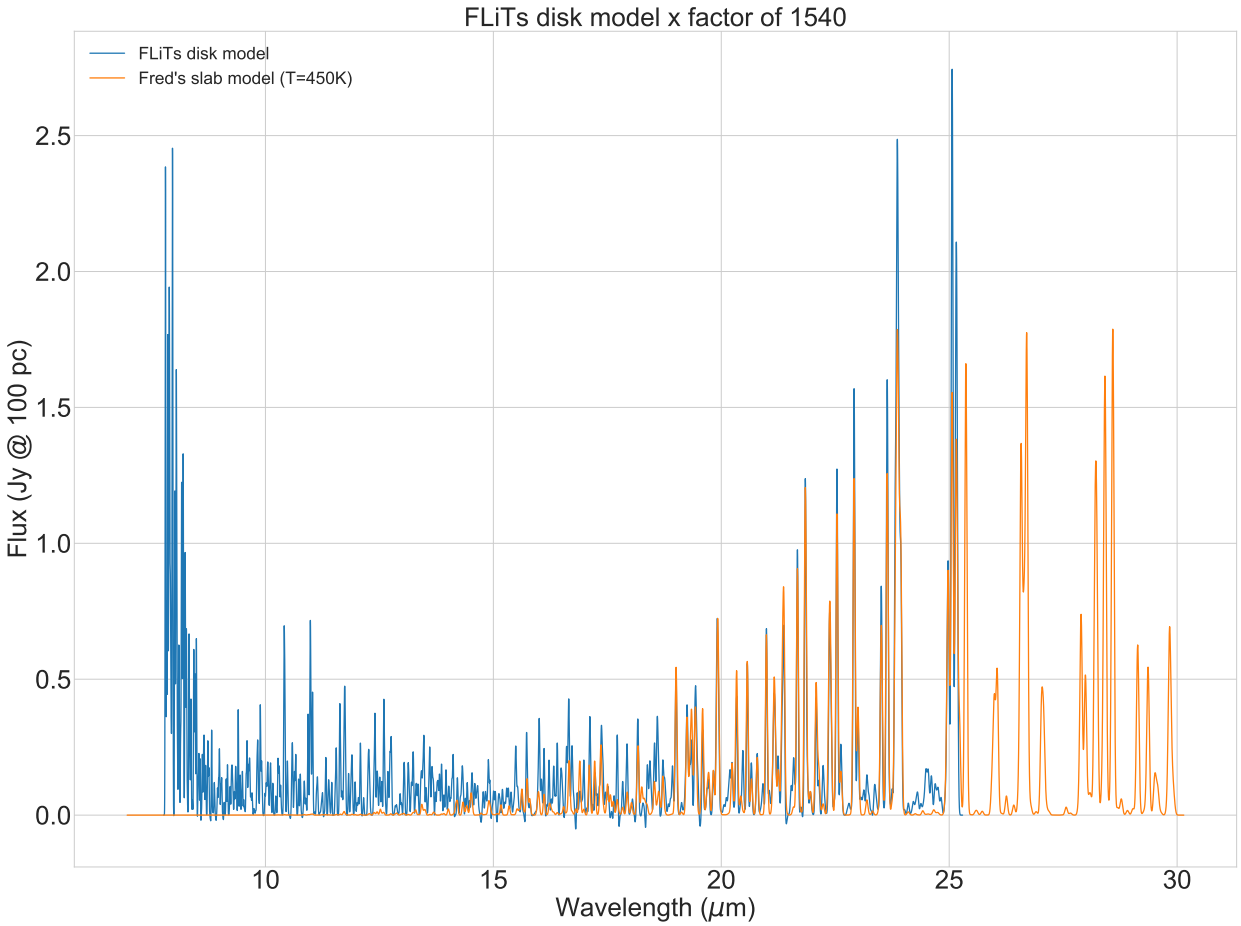

Figure 4: A comparison of the FLiTs disk model with Fred's water slab model at T = 450K, and Hydrogen column density $N_H = 10^{22}$ cm^{-2}. Note the range of the FLiTs disk model: the longer wavelengths are not covered by this model, thus the reason for no water lines in this model.

Fig. 5 examines the disk model with Fred's water slab model at a temperature of 1000K with the disk model multiplied by the same scaling factor of 1540.

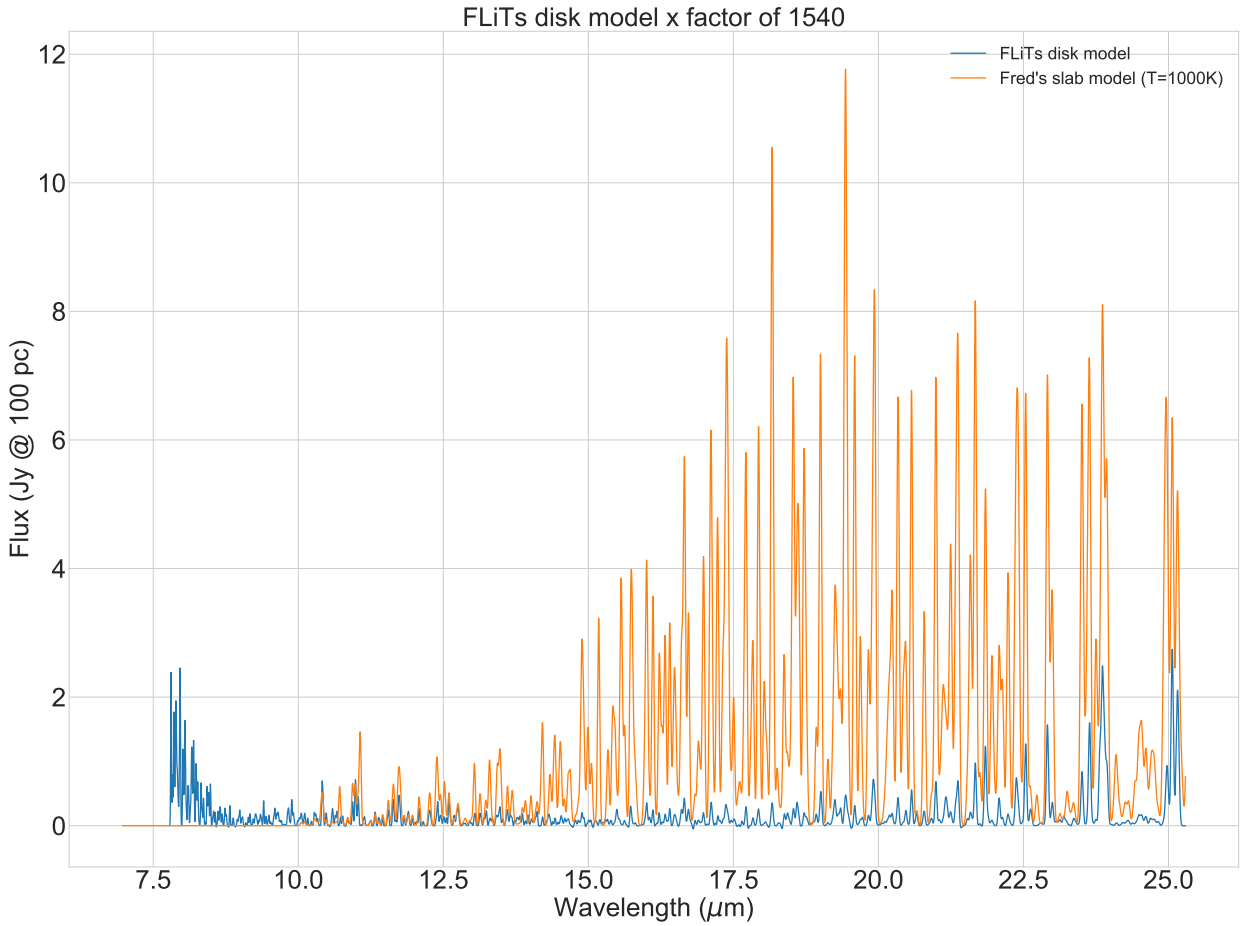

Figure 5: A comparison of the disk model with Fred's water slab model at a higher temperature of T = 1000K using the same scaling factor of multiplying the disk model by 1540 as in Fig. 4

This higher slab model temperature represents disk material closer to the star, thus it is expected to have a much smaller emitting surface area. At different wavelengths, water lines will come from different emitting regions. The higher the (upper level) excitation energy of the water line, the closer to the star the line is emitted from in the ProDiMo/FLiTs disk model. Unlike the ProDiMo/FLiTs models, slab models have the emitting surface area as a free scaling parameter. It is due to this difference that we expect to see a much higher scaling factor to match the amplitudes of the same water line.

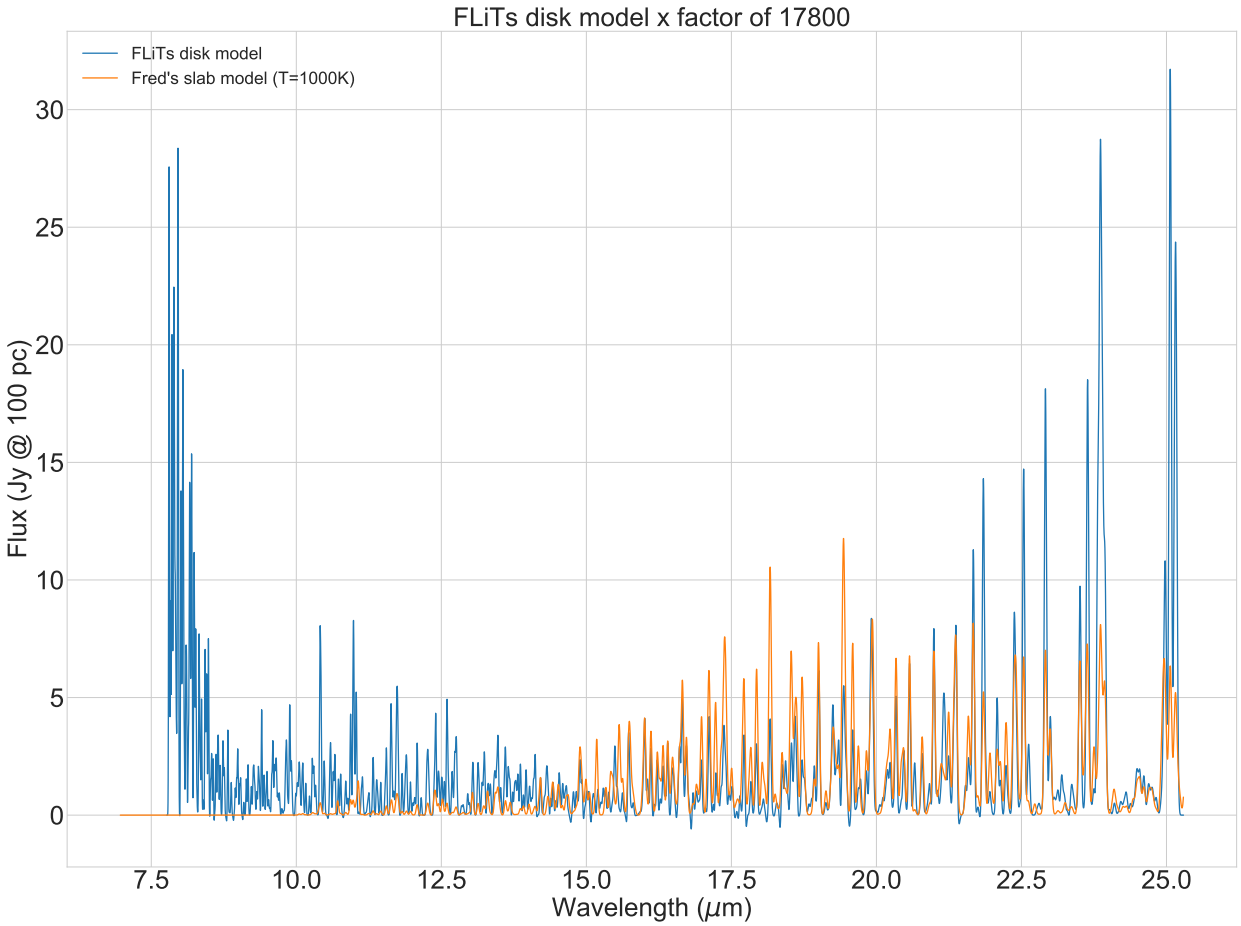

Figure 6: A comparison of the disk model with Fred's water slab model at a higher temperature of T = 1000K. It is noted that the scaling factor required to match the amplitudes of the 20 μm peak is much higher, which is required to scale the emitting surface area to the lower emitting surface area found closer to the core of protoplanetary disks.

2.3 Rotational and vibrational lines intermixed with each other

During the comparison of the disk model with Fred's slab model, intensity peaks which were only present in one model but not the other were found, see for example the closer examination around 15.5 micron in Fig. 7.

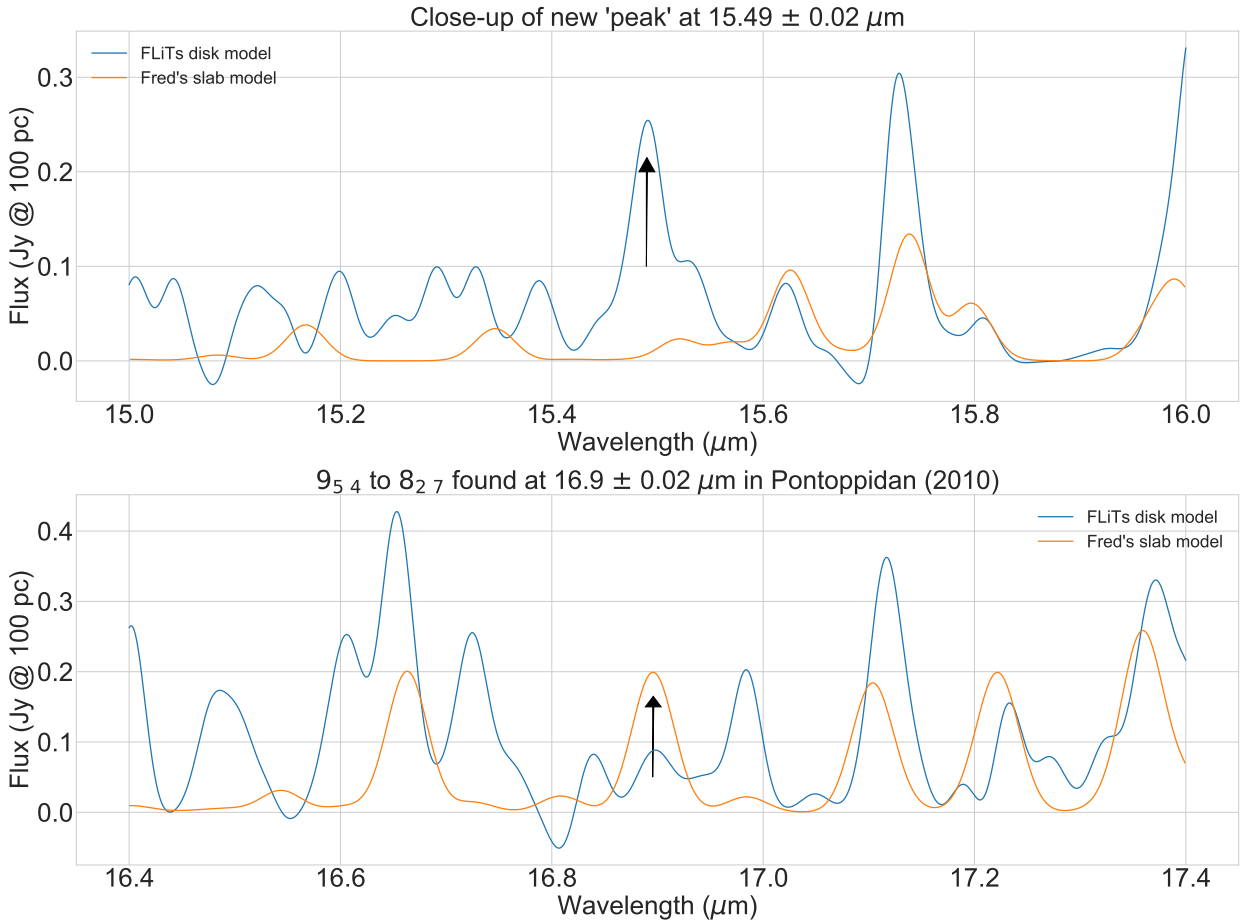

Figure 7: A close-up of 15.5 μm. An examination of the energy level transitions using Fig. 4 of Pontoppidan et al. (2010) (see Fig. 8 of this paper) and the VizieR On-line Data Catalog (Faure and Josselin (2008b), the details of which were originally published in Faure and Josselin (2008a)) reveals a discrepancy in the wavelength locations of energy level transitions. See the black arrows for the peaks referred to in the subplot titles.

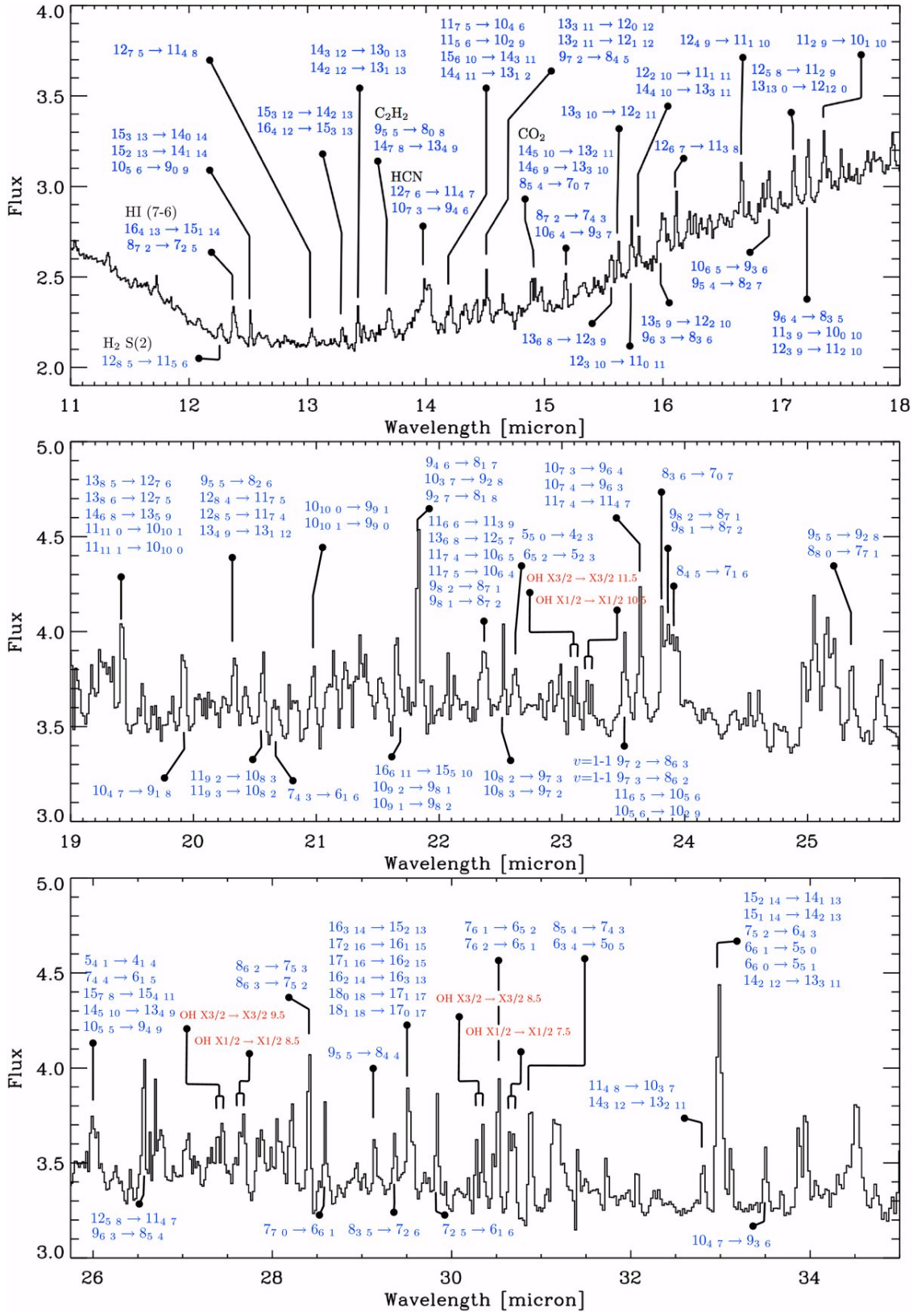

Figure 8: An identification of energy level transitions, as taken directly from Fig. 4 of Pontoppidan et al. (2010)

Energy level transitions were identified using Fig. 4 of Pontoppidan et al. (2010), shown here as Fig. 8 in this paper. The process was as follows:

- The wavelengths of the peaks were visually identified, with a margin of ± 0.07 μm.
- All energy level transitions within this range were identified using the ProDiMo models output file, which contained the following noteworthy parameters: frequency, flux (in local thermodynamic equilibrium), and Einstein A coefficients. The transition with the largest Einstein A coefficient was noted, and its upper and lower energy levels noted.
- The energies were cross checked against another table containing the collisional energy levels of water (Faure and Josselin, 2008b), from which we obtain the quantum numbers $J_{K_a K_c}$.
- This transition was then checked against Fig. 4 of Pontoppidan et al. (2010), (see Fig. 8 of this paper).

The discrepancy in the wavelength locations of energy level transitions between the two water models is due to Fred's models only containing the rotational lines, but the FLiTs disk model containing both rotational and vibrational lines. The example identified at 15.5 μm in Fig 7 is the $9_{5\ 4}$-$8_{2\ 7}$ transition but a ν=010-010 line, with an energy of the upper level E_{upper} of 3141.046191 cm^{-1}, or 4524.3 K. Pontoppidan et al. (2010) identifies this line at 16.9 μm, where we find the emission line $9_{5\ 4}$-$8_{2\ 7}$ but now ν=000-000 line, with E_{upper} of 1477.297570 cm^{-1}, or 2125.5 K.

2.4 Results of the comparison

As noted in Fig. 4, with an emission line matched in amplitude at 20 μm around the midpoint of the wavelengths of overlap, we see a trend that the disk model has stronger emission lines at shorter wavelengths, and expect the slab model to have stronger lines at longer wavelengths if the slab model still holds true. However, as previously noted in §1.2, the slab model very quickly breaks down for such a wide wavelength range (7-30 μm), in which the lines originate from a wide region of the disk, and temperature and density vary drastically (Kamp, 2015). The limitations of slab models should be kept in mind, and only be used to fit observations over a small wavelength range, in which emitting surface area, column density, and temperature do not vary. As Pontoppidan et al. (2010) has shown, line excitation is not in LTE. So simple slab models should only be used just to understand the properties of emitting atmospheres, and to investigate the variety of emissions (Salyk et al., 2011).

An interesting fact to note is that as described in §2.3, the pure rotational and ro-vibrational lines are quite intertwined with each other, and the higher vibrational lines are not much weaker than the pure rotational lines. One future possibility is to use both the information from the VizieR Online Data Catalog (Faure and Josselin, 2008b) and that of the ProDiMo results (Woitke et al., 2009; Kamp et al., 2010; Aresu et al., 2011) to distinguish

both rotational and vibrational lines. Note that the ortho and para water lines were simply added together in the FliTs disk model and the ProDiMo slab model, and the total fluxes may be slightly over-estimated (Greenwood, 2018).

When fitting observed spectra, a slab model should only be used the wavelength range is narrow, as the emissions lines all originate from levels with very similar excitation temperature. If wavelength range is broader, the excitation temperatures will vary too much and a slab model with only single temperature would not work.

3 The Wavelet Transform

3.1 Theoretical introduction

The WT is a time-frequency analysis tool similar to a Fourier Transform (FT), but decomposes the signal into finite length waves which are spatially localised. Unlike the FT, which integrates over all time and thus is only localised in frequency, the WT retains spatial information of the original signal (Daubechies, 1992).

The WT of a one-dimensional signal has a frequency and spatial location variable, and leads to a series of spectra at different resolutions, going from coarse to finer as we continue down the scales (Fligge and Solanki, 1997). Compared to the FT, this allows different parts of the spectra to be filtered individually, giving a better noise reduction.

3.2 Mathematical details

The following mathematical details behind the WT is taken from Starck and Bijaoui (1994); Starck and Murtagh (1994), previously described in Daubechies (1992); Shensa (1992).

We define original data at pixel k $\{c_o(k)\}$ as the scalar product of the function f(x) with a scaling function $\phi(x)$:

$$c_o(k) =< f(x), \phi(x-k) > \tag{1}$$

The scaling function ϕ (which corresponds to a low pass filter) is chosen so that the following dilation equation is satisfied:

$$\frac{1}{2}\phi\left(\frac{x}{2}\right) = \sum_l h(l)\phi(x-l) \tag{2}$$

Here h is a discrete low pass filter associated with the scaling function $\phi(x)$. i.e., a low-pass filtering of the data is related to another resolution level of the image (the next scale). The distance between levels increase by a factor of 2 from one scale to the next. The smoothed data $c_i(k)$ at scale i and position k is defined as:

$$c_i(k) = \frac{1}{2^i} < f(x), \phi\left(\frac{x-k}{2^i}\right) > \tag{3}$$

This is obtained by the convolution:

$$c_i(k) = \sum_i h(l) c_{i-1}(k + 2^{i-1} l) \tag{4}$$

The signal difference w_i between two consecutive resolutions is:

$$w_i(k) = c_{i-1}(k) - c_i(k) \tag{5}$$

or:

$$w_i(k) = \frac{1}{2^i} < f(x), \psi\left(\frac{x-k}{2^i}\right) > \tag{6}$$

Here ψ is the wavelet function (also known as the mother wavelet) is defined as:

$$\frac{1}{2}\psi\left(\frac{x}{2}\right) = \phi(x) - \frac{1}{2}\phi\left(\frac{x}{2}\right) \tag{7}$$

Equation 5 is the discrete wavelet transform for resolution level i, and can be seen later in §3.4.1. We choose a B-spline of order 3 (B_3) cubic spline for the scaling function ϕ for the reasons outlined in §3.3.

3.3 Choice of À'trous algorithm and B3 cubic spline wavelet

In this study, the à'trous algorithm ('with holes'), also known as the stationary wavelet transform (SWT), or the undecimated wavelet transform (UWT) is examined. A B-spline of order 3 (B_3) cubic spline wavelet is used. Starck et al. (1997) argues for the use of this algorithm and wavelet due to the following advantages, as taken verbatim from that paper:

1. The transform is carried out in direct space, and artifacts due to a periodisation do not occur.
2. The transform uses compact scaling functions.
3. The reconstruction algorithm is trivial.
4. The evolution of the transform from one scale to the next can be followed easily.
5. The transform is invariant under translation.

At each scale, the à'trous algorithm convolves the input signal with a kernel, in this case our B_3 spline, to get the smoothed array (also known as approximation coefficient, names used interchangably). The wavelet coefficient itself (or detailed coefficient), is simply the input signal minus the smoothed array, in other words the difference between the original input signal and the results of the convolution using the convolution mask.

As stated in point 3, the reconstruction algorithm is simple. The wavelet transform using the à'trous algorithm (full details of the algorithm found in Shensa, 1992) produces a set of

wavelet coefficients $\{w_j\}$ at each scale j, and a smoothed array c_p at the widest scale p which is used to reconstruct the original data c_o (Starck et al., 1995).

$$c_o = c_p + \sum_{j=1}^{p} w_j \tag{8}$$

The B_3 cubic spline is illustrated in Fig. 4, taken from Starck et al. (2015).

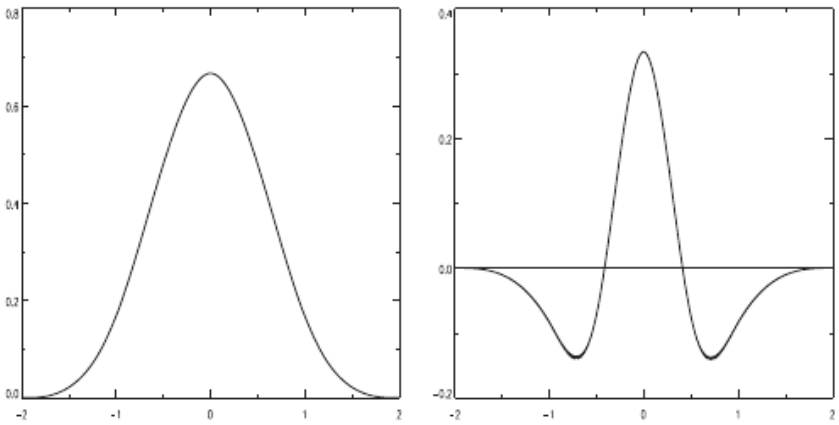

Figure 9: Left: the cubic spline scaling function ϕ, which is our initial convolution mask. Right: the wavelet function (mother wavelet) ψ. Source: Starck et al. (2015), p.55 (Fig. 3.6).

3.4 Code example

3.4.1 À'trous algorithm wavelet transform

For the first wavelet scale, in 1-D, the B_3 cubic spline convolution mask is: $\left(\frac{1}{16}, \frac{4}{16}, \frac{6}{16}, \frac{4}{16}, \frac{1}{16}\right)$. The flux of the spectrum is then convolved with this mask to give us our wavelet coefficients of the first scale. Described using Python code:

```
b3_spline = numpy.array([1/16, 4/16, 6/16, 4/16, 1/16])
c1 = scipy.signal.convolve(flux, b3_spline, mode='same')
w1 = flux - c1
```

Where c1 is the smoothed array c_1 of scale 1 (j=1). w1 is the wavelet coefficient w_1 of scale 1. Note the definition of the wavelet coefficients at scale 1 (w_1) in Eq. 5. It is simply the original signal minus the convolved signal. Note also the reconstruction described in Eq. 8. If the WT is only performed to scale 1, then simply adding c_1 and w_1 gives us our original spectrum.

For the second wavelet scale, the convolution mask is filled with zeroes (or 'holes') between the coefficients, becoming: $\left(\frac{1}{16}, 0, \frac{4}{16}, 0, \frac{6}{16}, 0, \frac{4}{16}, 0, \frac{1}{16}\right)$, and the smoothed array of the previous scale is convolved with this widened mask. Described using Python code:

```
widenend_b3_spline = numpy.insert(b3_spline, slice(1, None), 0)
c2 = scipy.signal.convolve(c1, widenend_b3_spline, mode='same')
w2 = c1 - c2
```

Here c2 is the smoothed array c_2 at scale 2, and w2 is the scale 2 wavelet coefficient w_2 of Equation 8. The wavelet coefficient for this scale w_2 is then the difference between the smoothed array of the previous scale c_1 (result of the convolution of the initial data) and the result of the convolution at this scale c_2.

The process repeats for the third wavelet scale. The convolution mask is widened with zeroes, becoming: $\left(\frac{1}{16}, 0, 0, 0, \frac{4}{16}, 0, 0, 0, \frac{6}{16}, 0, 0, 0, \frac{4}{16}, 0, 0, 0, \frac{1}{16}\right)$. Thus we convolve the signal with a progressively wider convolution mask as we continue down the wavelet scales and examine wider details within our signal.

3.4.2 Reconstructing the spectra from wavelet coefficients

As stated in point 3. by Starck et al. (1997), the reconstruction algorithm is trivial, and is simply the sum of all wavelet coefficients $\{w_j\}$ across all scales, with the smoothed array of the widest/last scale (see Eq. 8). In Python code, the reconstruction algorithm is simply:

```
reconstructed_flux = C[-1] + np.sum(W, axis=0)
```

This is Eq. 8 written down in code, where the reconstructed flux is our original data c_o. C is the set of arrays of all smoothed arrays up to the widest scale p, which is the last scale we perform the WT to, and the one we need to reconstruct the spectrum. W is the set of all wavelet coefficients $\{w_j\}$ summed up over all scales to scale p:

$$C[-1] = c_p$$

$$W = \sum_{j=1}^{p} w_j$$

4 Method

As all datasets used were noiseless and continuum-subtracted, both noise and a simple continuum had to be manually added to test the de-noising and continuum location determination of WT. Gaussian was added using Python's numpy.random.normal with a mean μ of 0, and a variable standard deviation σ which will be displayed under the relevant figures. A simple linear continuum was added by multiplying the wavelength array by 0.01 times the maximum

of the flux. This gives us a straight line which rises from the beginning of the wavelength range to 0.01 of the maximum of the flux at the longer wavelength end.

As described in §3, a spectrum was convolved with a B_3 spline to give a smoothed array, and the wavelet coefficient was found via the difference between the spectra and said smoothed array. Fligge and Solanki (1997) argues that it is unnecessary to go wider than four scales, as noise is mostly concentrated at high frequencies (see Fig. 5 of said paper for a comparison of two noise-free signals against a variety of de-noising methods.). The WT was performed up to scale 5, then the wavelet coefficients of scales 1-5 were further transformed to give coefficients of scales 11-15, 21-25, 31-35, 41-45, 51-55. Filtering was then performed on these wavelet coefficients. This multi-level approach, performed by Fligge and Solanki (1997), retains information which would otherwise by filtered out if the filtering was simply performed at the first level WT.

Fig. 10 displays the multi-level WT, decomposing the WT scales 1-5 further into scales 11-55. If the filtering were to be simply applied across scales 1-5, the signal in scale 1 would have been lost. Further WT of the scale 1 allows us to see the signal in scales 14 and 15, as we will see in Fig. 11. Similarly, we see signals in scales 24, 25, 31, 34, 35, in all scales 41-45, and 51-55.

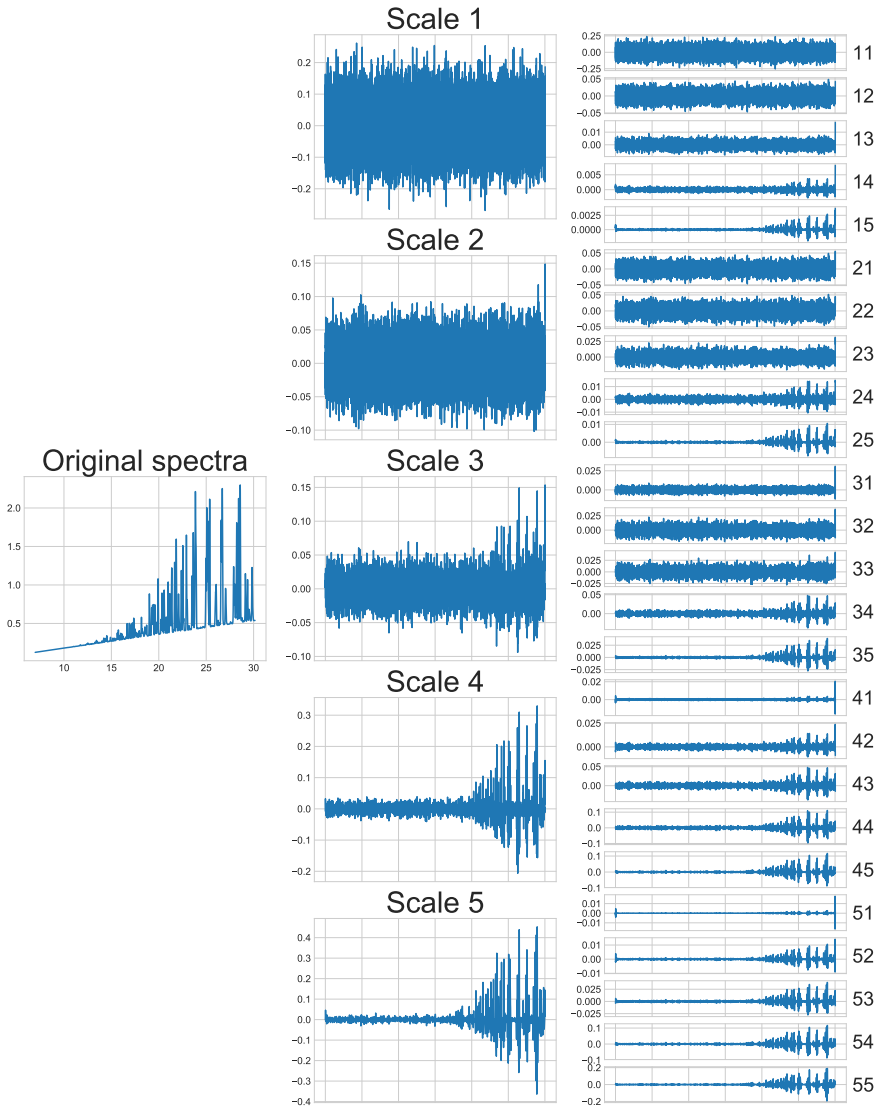

Figure 10: Multi-level WT with added noise $\sigma = 10^{-1}$

4.1 Filtering and de-noising

The filtering and de-noising of the spectra was done via a hard thresholding, where any wavelet coefficients below a certain limit is insignificant, and simply set to 0 and disregarded. Fligge and Solanki (1997) compares a number of filtering options, and finds their approach of a three-level decomposition coupled with hard thresholding to be most accurate, though computationally slightly more expensive than e.g. the adaptive filtering methods proposed by Starck and Bijaoui (1994) in de-noising and recovering weak signals.

A hard threshold value of kσ_j was set, where k was taken as 3 to give a confidence level of 99.7% (Fligge and Solanki, 1997), and σ_j is the standard deviation of the noise at scale j. For wavelet coefficients on scale j, at location (x,y):

if $|w_j(x,y)| \geq k\sigma_j$, then w_j is significant, and we keep the wavelet coefficient
if $|w_j(x,y)| < k\sigma_j$, then w_j is not significant, and we eliminate it by setting it to 0

At each wavelet scale j, we find the standard deviation of the noise σ_j using the following:

$$\sigma_j = \sigma_d \sigma_j^e$$

Here σ_d is the standard deviation of the noise in the original data, found by using Python's numpy.std() function. σ_j^e is the standard deviation of the wavelet space noise, found by generating white Gaussian noise of $\sigma = 1$, then performing a multi-level WT on that noise to obtain its wavelet coefficients at scales 11-55. This lets us observe how noise behaves across the wavelet scales Starck et al. (1995, 1997); Starck and Murtagh (1998). Thus the product of σ of the data and σ of the behaviour of noise across the wavelet scales gives us a threshold level with which we can apply our filtering.

This gives us a significance level which varies across the wavelength, and wavelet scales. Anything below this significance level is considered noise and is filtered out. This hard threshold was applied to the WT scales 11-55 and the thresholded / de-noised wavelet coefficients were recombined to give the de-noised WT scales 1-5, which were then reconstructed to give a de-noised spectra.

While not performed in Fligge and Solanki (1997, 1998), the value of σ_j was divided by 0.974, found experimentally in Starck and Murtagh (1998) to account for systematic bias, due to the fact that noise in the signal will introduce noise across all wavelet scales, and some significant wavelet coefficients above $3\sigma_j$ will be due to the noise of the signal.

As an example see Fig. 10 for a signal which has WT applied to it to give wavelet coefficients at scales 1-5, upon which a WT is applied to give wavelet coefficients at scales 11-55. The filtered wavelet coefficients at scales 11-55 is shown in Fig. 11, and are recombined to form the de-noised or filtered coefficients at scales 1-5, from which the de-noised spectrum is reconstructed.

Fig. 11 displays the filtered wavelet coefficients of Fig. 10, and the reconstruction of the spectra using these filtered wavelet coefficients.

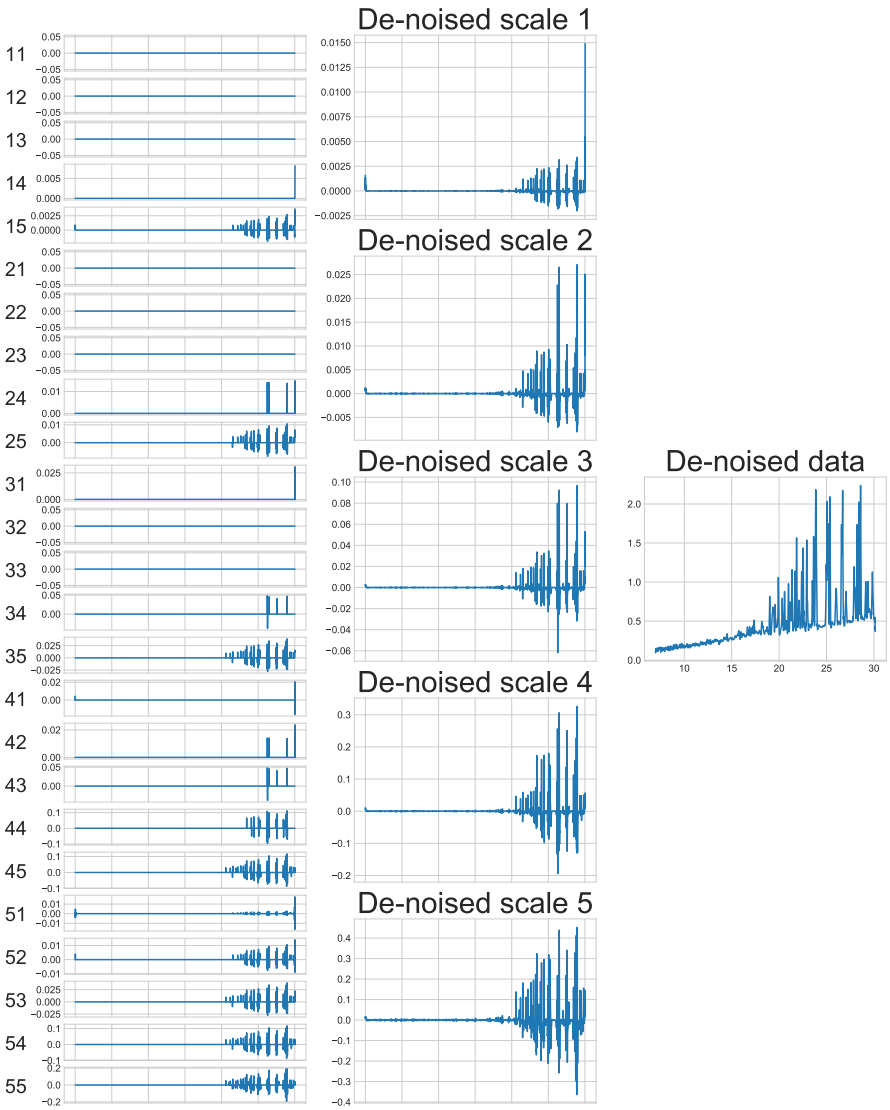

Figure 11: De-noised multi-level WT with added noise $\sigma = 10^{-1}$, with filtering applied

We see the effects of this hard thresholding in Fig. 11, where noise below this significance level is eliminated, leaving only the signal. These de-noised or filtered wavelet coefficients are then recombined back into wavelet scales 1-5, which are then used to reconstruct the de-noised spectra.

4.2 Determining the location of the continuum

Using the à'trous algorithm, the automatic determination of the continuum is straightforward, and the continuum C(λ) is simply the smoothed array $c_p(\lambda)$ at the widest scale p of Eq. 8.

At the widest scale, the smoothed array is obtained by a convolution of the spectrum by a very wide filter, in other words it contains all the information we have at a low spectral resolution (Starck et al., 1997).

Note that the continuum is not always at the widest scale, and we must manually decide to which scale p we perform the WT. In other words, it is up to us to specify how wide the bands are that we aim to detect. This information can be obtained by observing the wavelet scales of the WT of the signal (Starck et al., 1997).

5 Results

The results are presented in the following way:

In §5.1 we will apply the de-noising and filtering algorithm to the models, which will be displayed under their own subsection. We will introduce Gaussian noise of variable standard deviation σ, which will be shown as the caption under the respective figures. A range of standard deviations was used to explore the threshold where algorithm fails for each model, seen in Figs. 12, 15, 18. Then two subsequent plots for each model will demonstrate the behaviour of the WT algorithm. The continuum-subtracted *Spitzer* datasets already contain noise, thus no noise was added, however the simple continuum as described in §4 was added. In §5.2 we will examine the determination of the location of the continuum, again all models will be under their respective subsections, and no noise was added to the *Spitzer* data.

5.1 De-noising

The de-noising results are presented in the following way:

- The top left plot of each figure shows original data with added noise, with the standard deviation σ shown in the caption under each figure.
- The top right plot shows the reconstruction of the noisy data using wavelet coefficients.
- The bottom left plot shows the original spectra before introducing noise.

- The bottom right plot shows the de-noised data using filtered wavelet coefficients, using the filtering method described in §4.1.

5.1.1 Aaron's FLiTs models

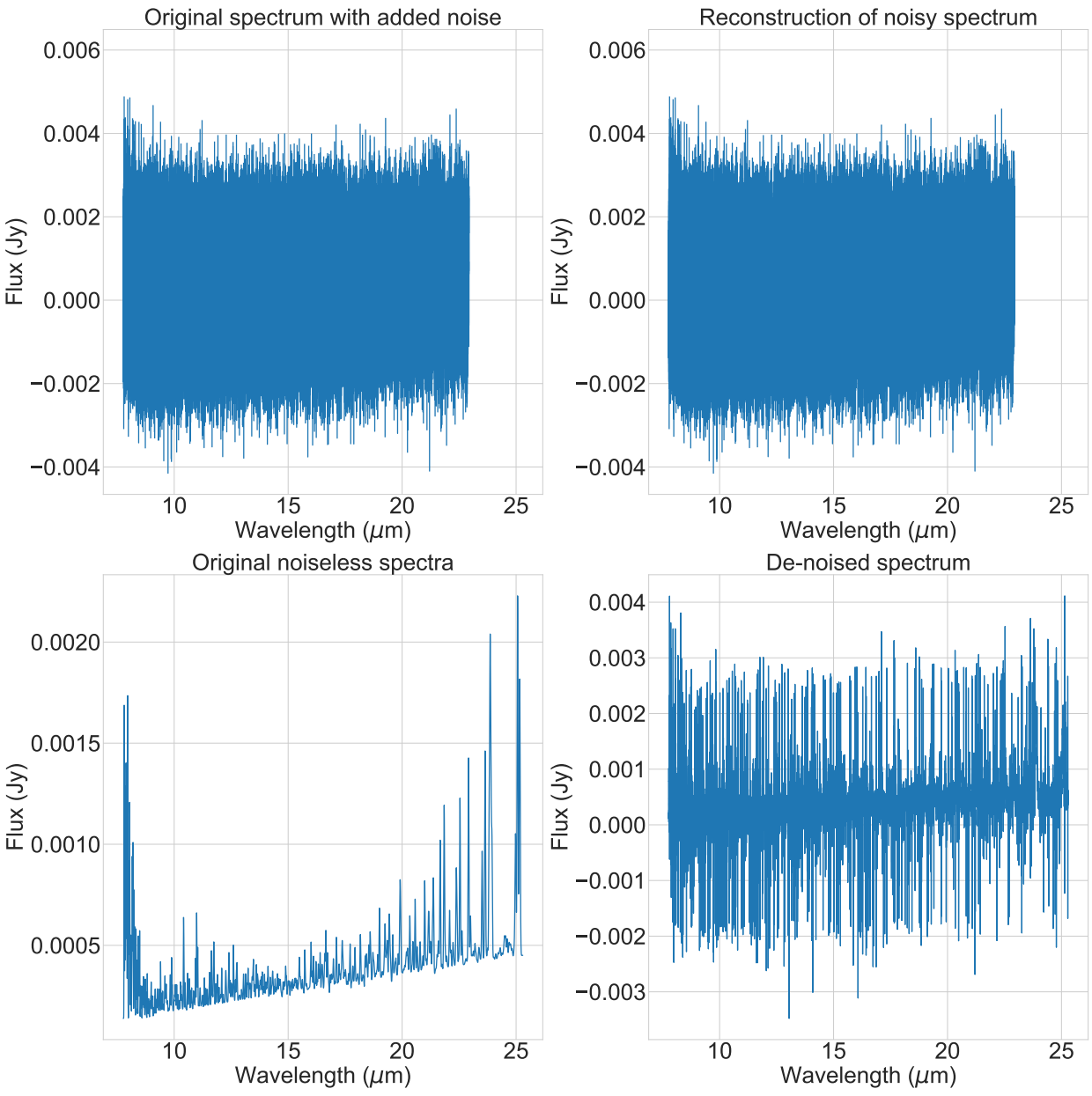

Figure 12: Aaron's FLiTs model (H2O), with added noise $\sigma = 10^{-3}$

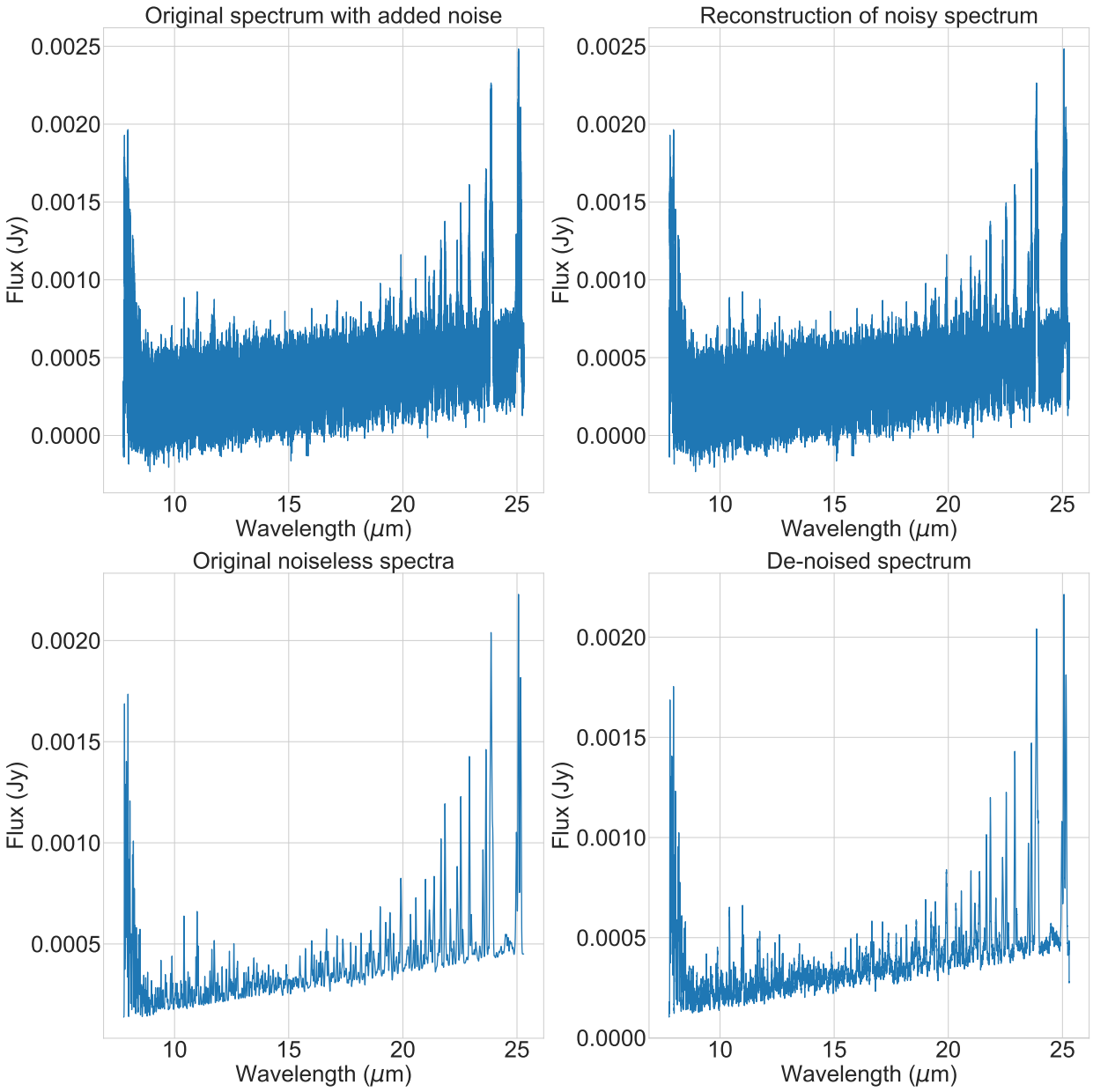

Figure 13: Aaron's FLiTs model (H2O), with added noise $\sigma = 10^{-4}$

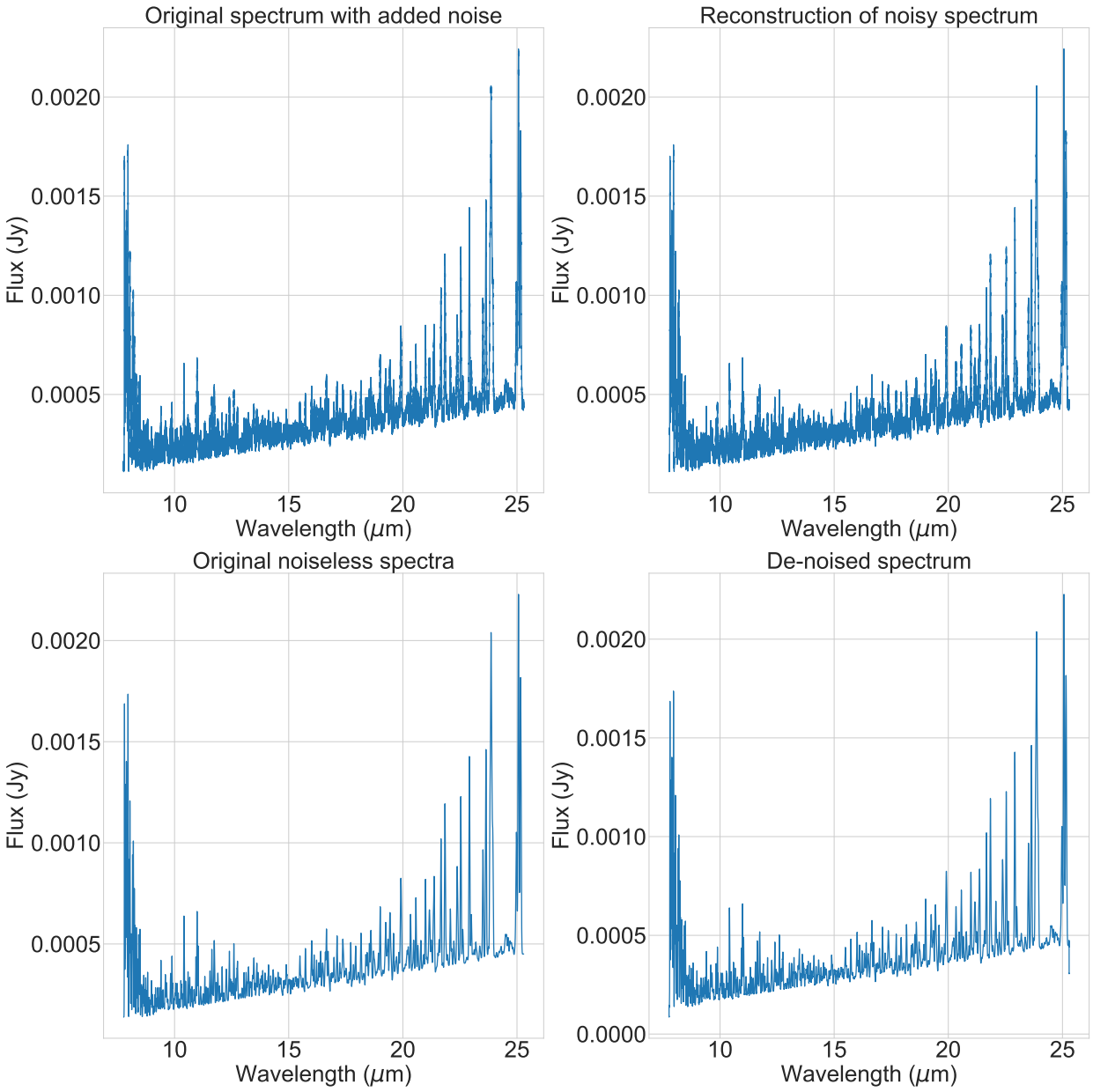

Figure 14: Aaron's FLiTs model (H2O), with added noise $\sigma = 10^{-5}$

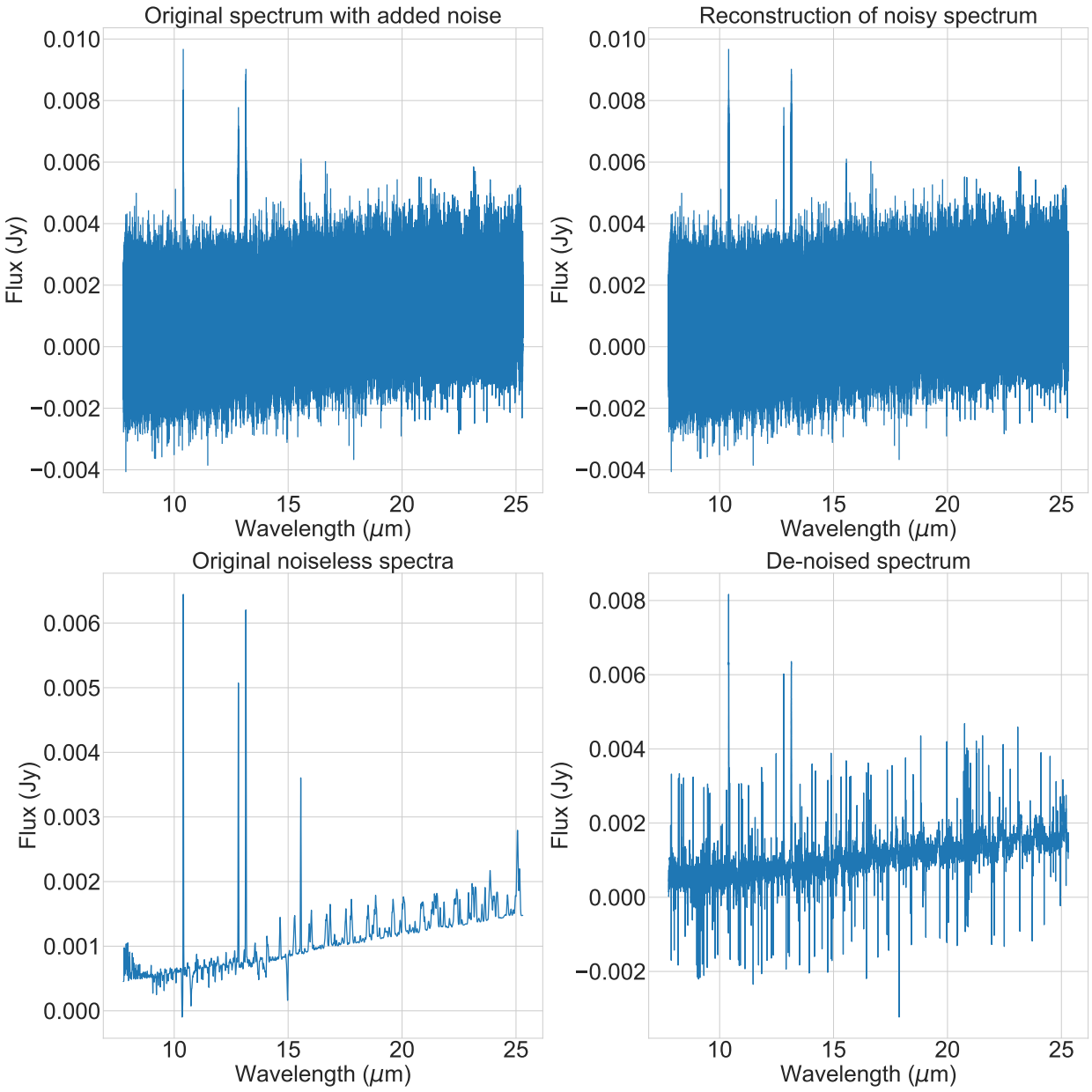

Figure 15: Aaron's FLiTs model (all molecules), with added noise $\sigma = 10^{-3}$

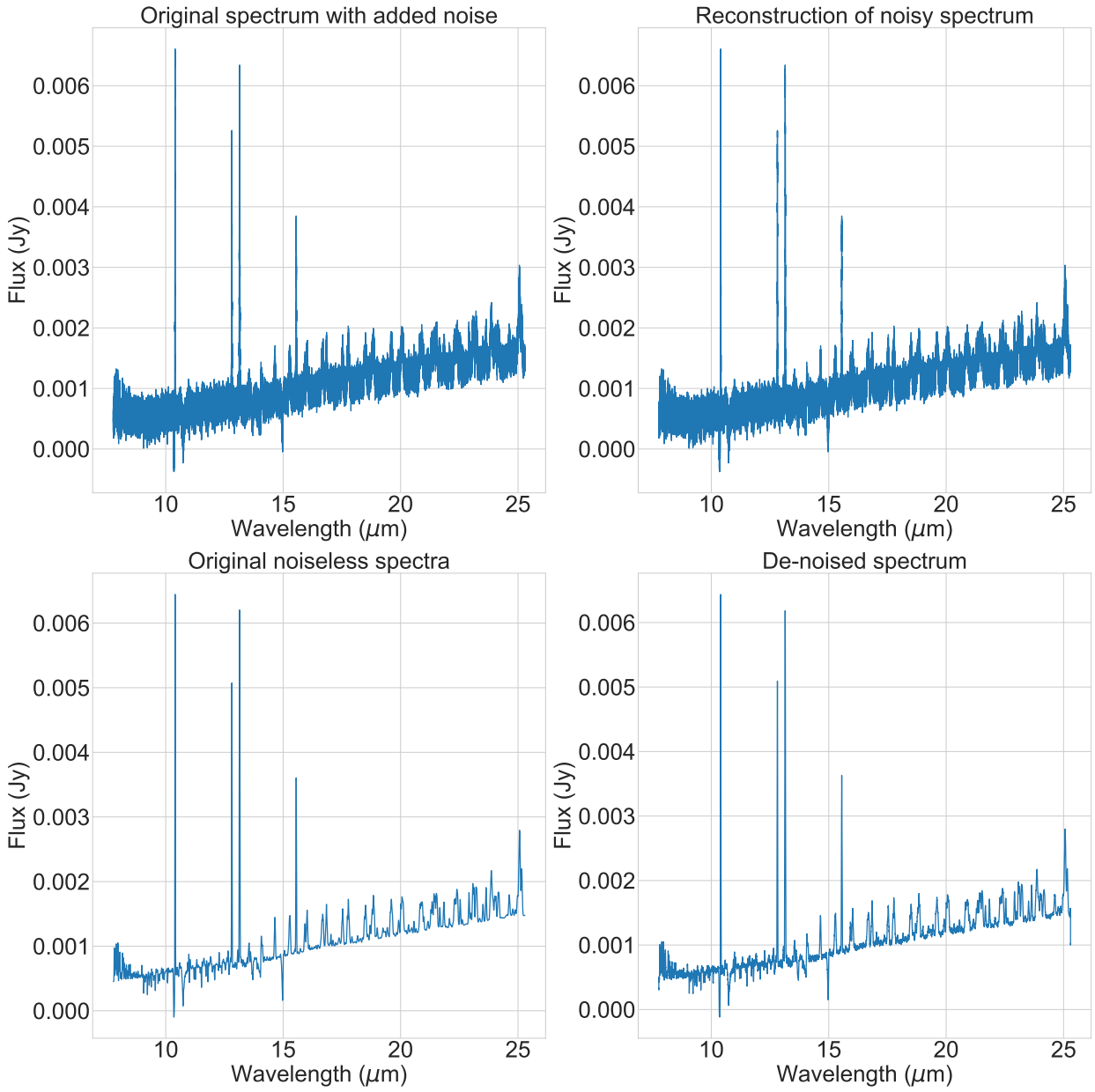

Figure 16: Aaron's FLiTs model (all molecules), with added noise $\sigma = 10^{-4}$

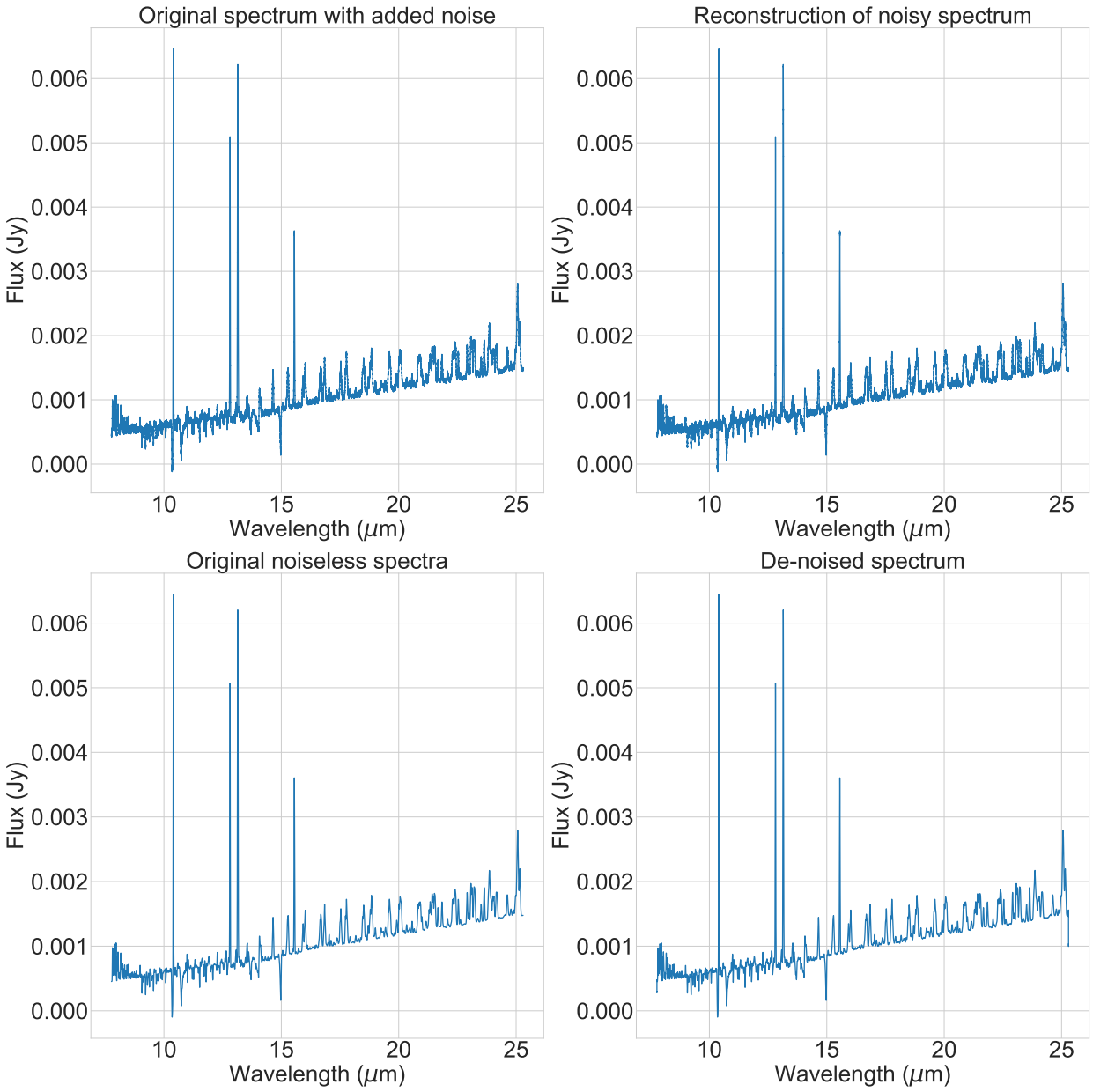

Figure 17: Aaron's FLiTs model (all molecules), with added noise $\sigma = 10^{-5}$

5.1.2 Fred's water slab models

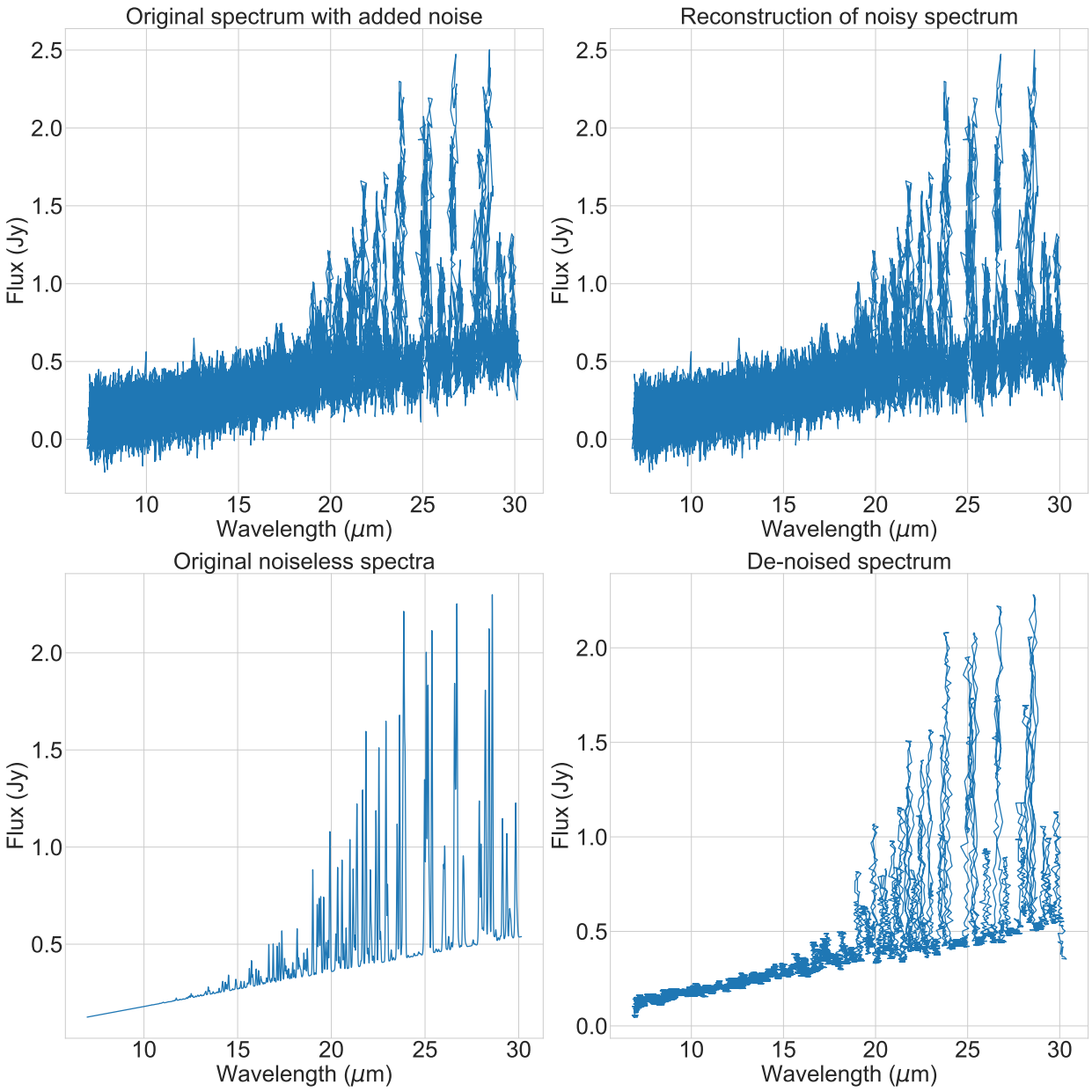

Figure 18: Fred's water slab model, with added noise $\sigma = 10^{-1}$

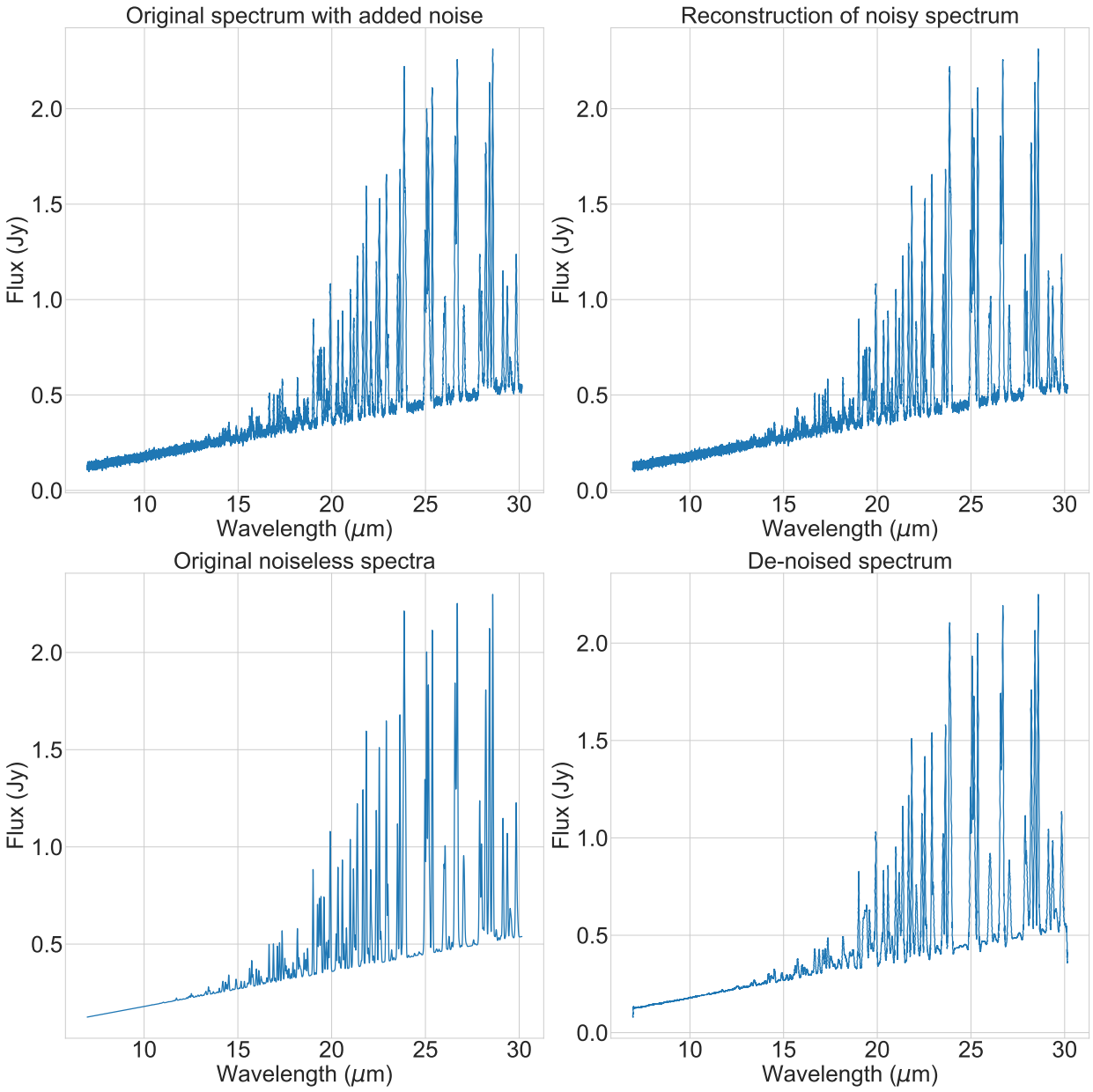

Figure 19: Fred's water slab model, with added noise $\sigma = 10^{-2}$

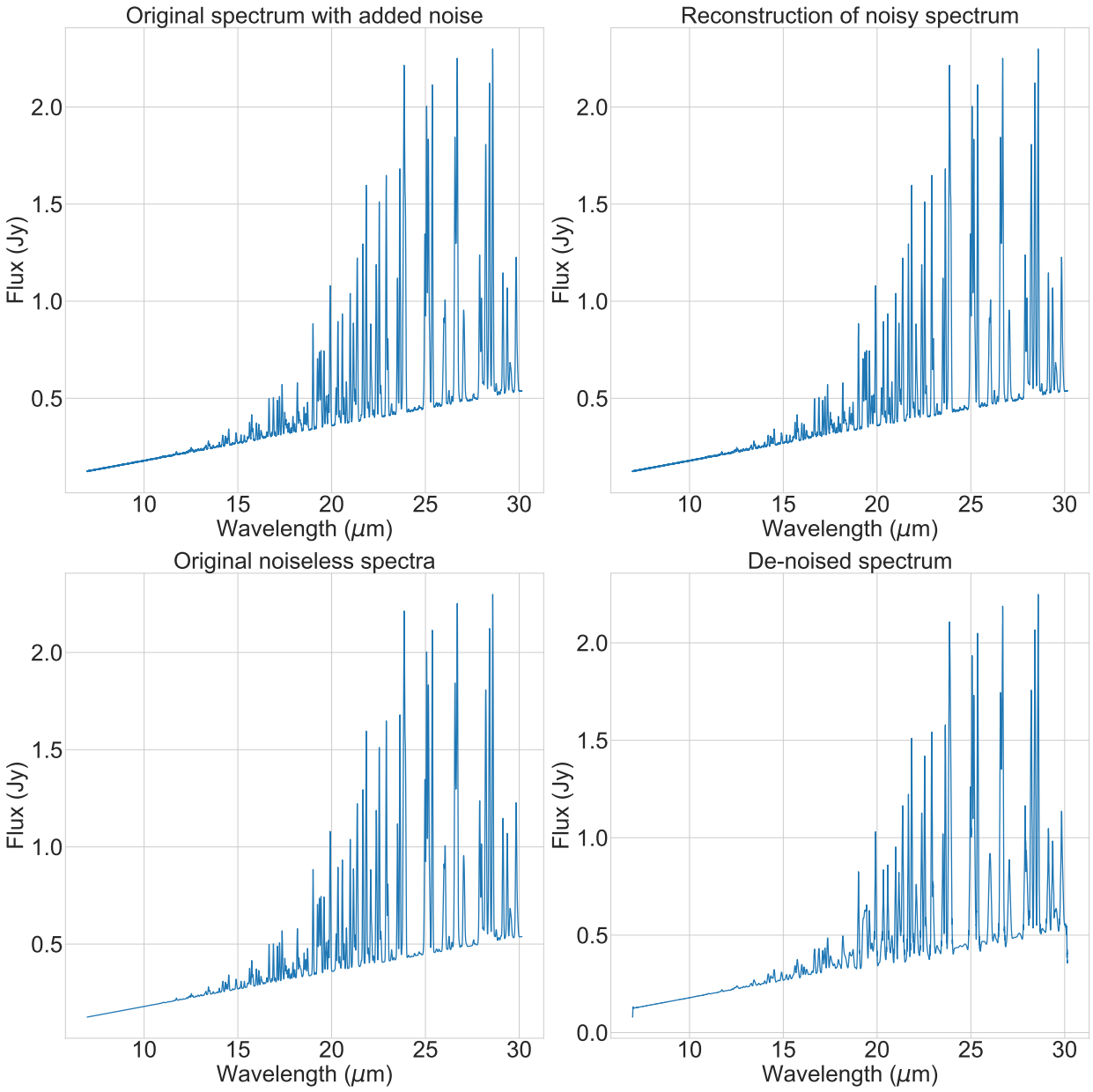

Figure 20: Fred's water slab model, with added noise $\sigma = 10^{-3}$

5.1.3 ProDiMo slab models

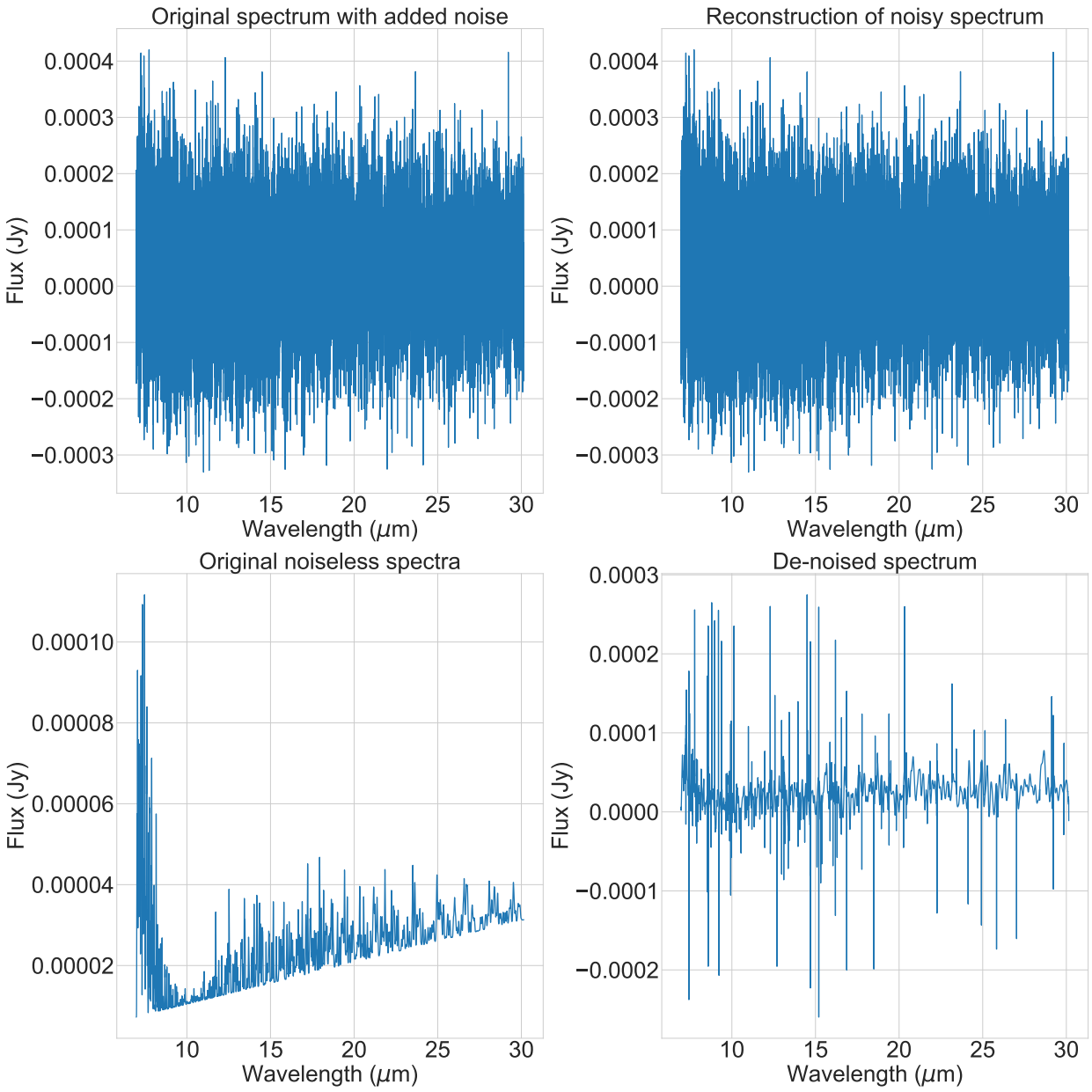

Figure 21: ProDiMo water slab model, with added noise $\sigma = 10^{-4}$

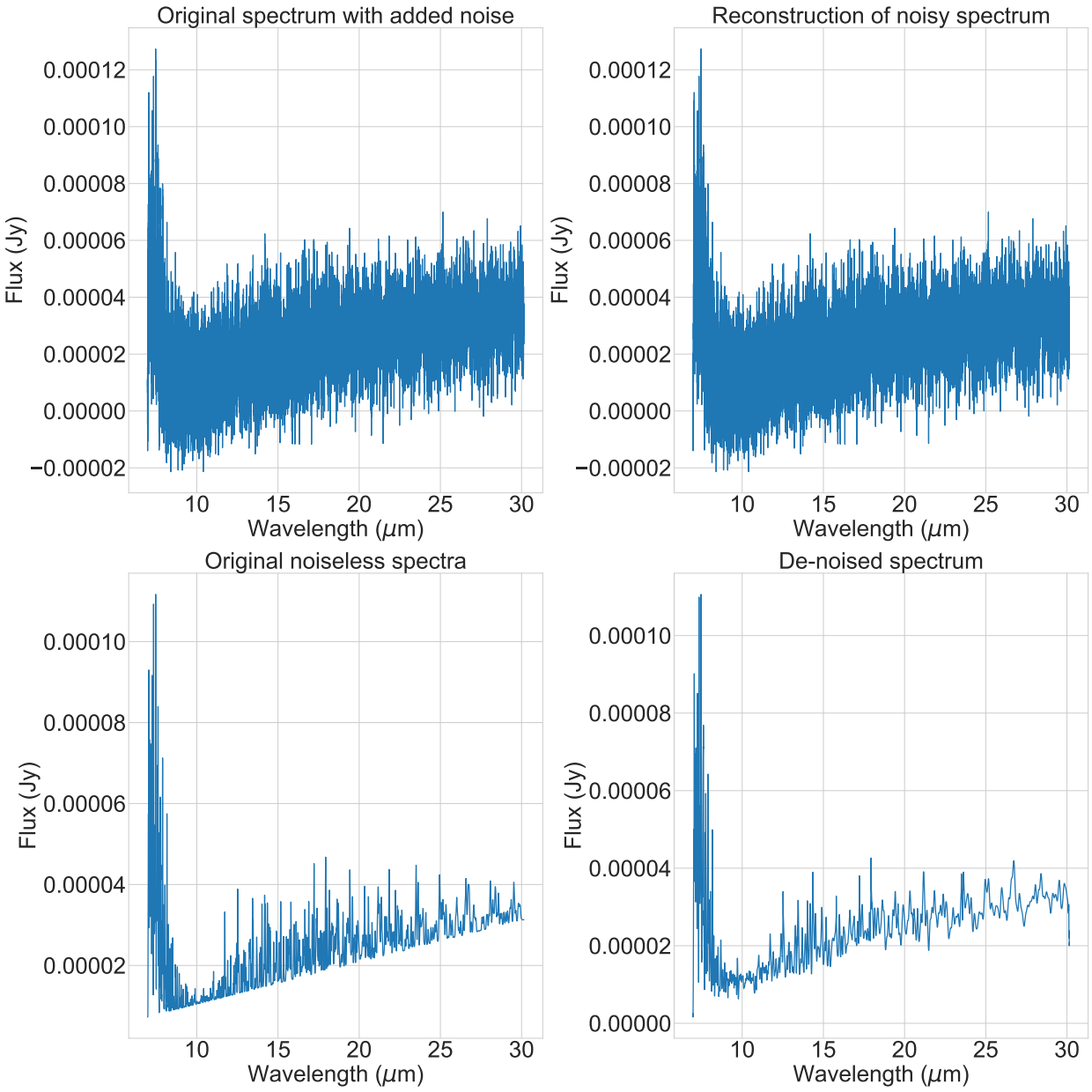

Figure 22: ProDiMo water slab model, with added noise $\sigma = 10^{-5}$

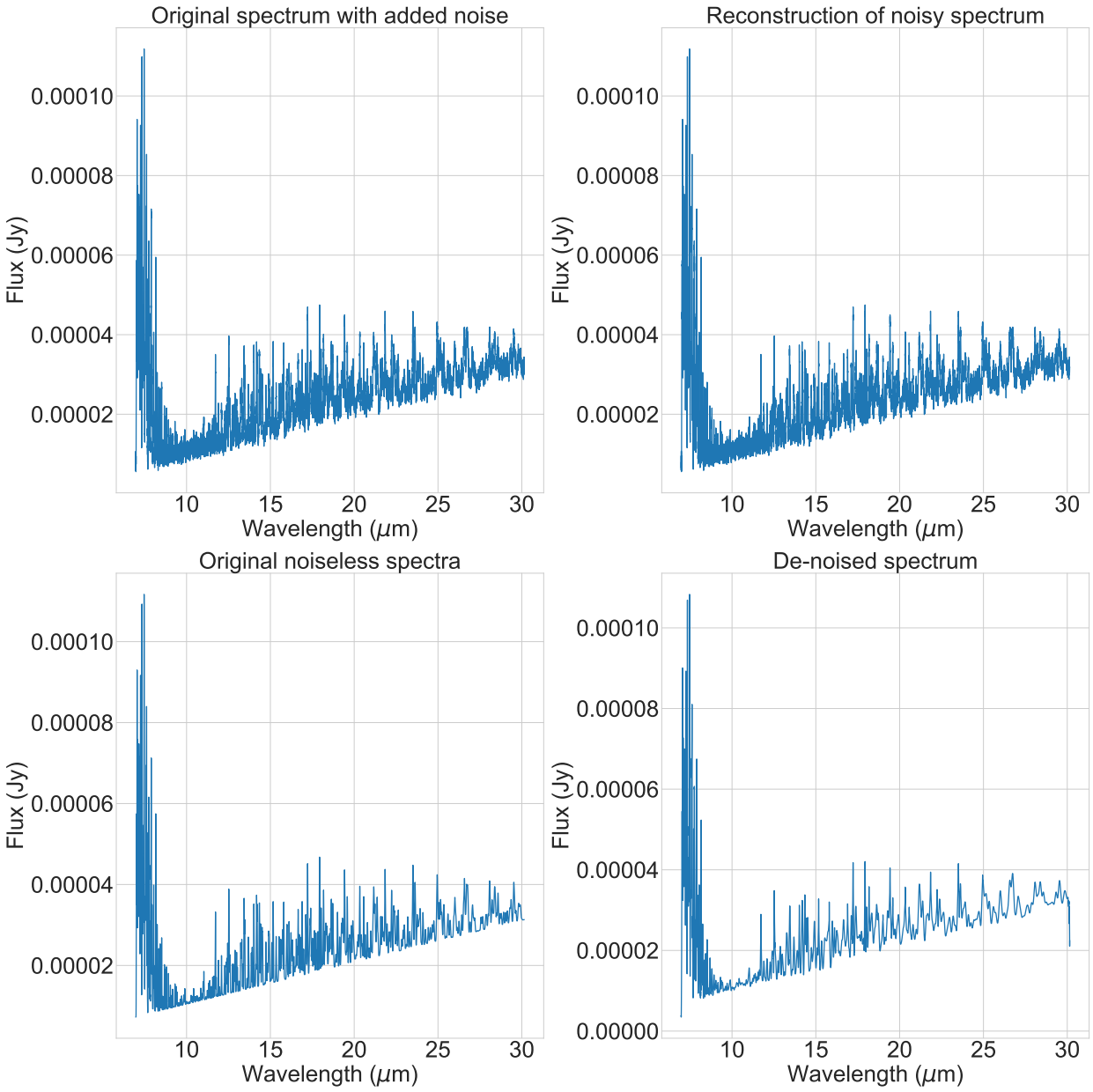

Figure 23: ProDiMo water slab model, with added noise $\sigma = 10^{-6}$

5.1.4 Spitzer data

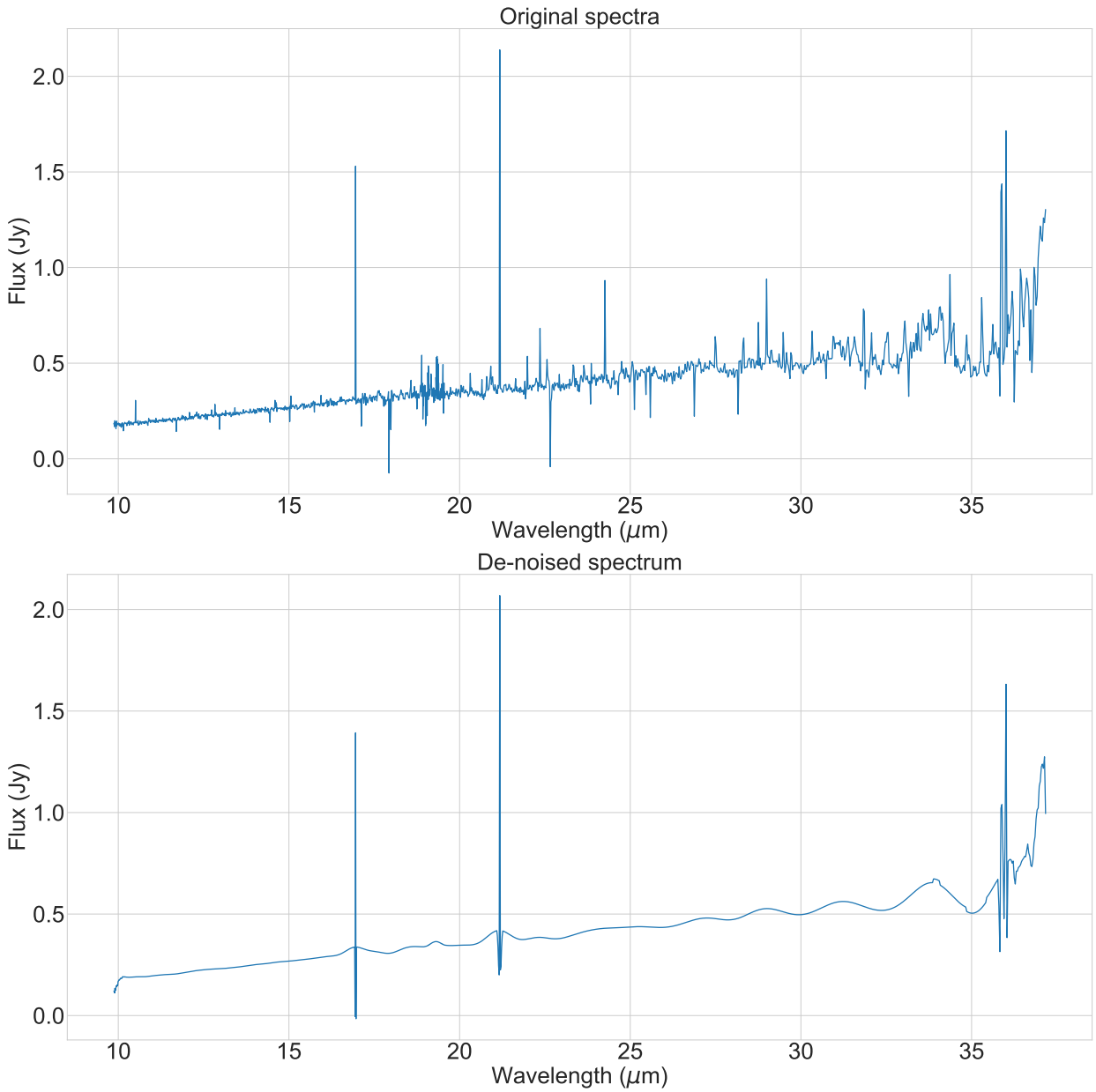

Figure 24: CoKu Tauri/4

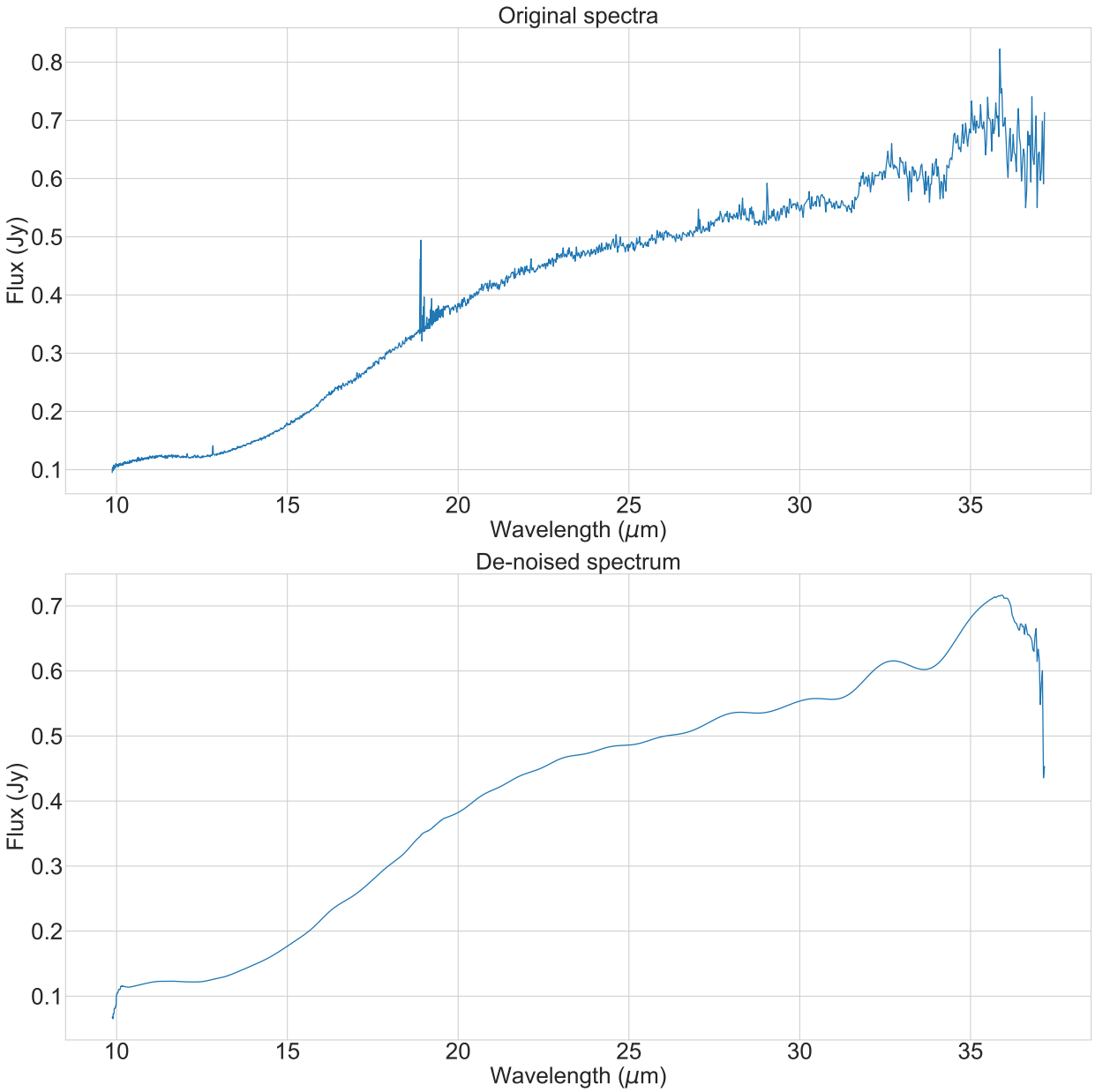

Figure 25: DM Tauri

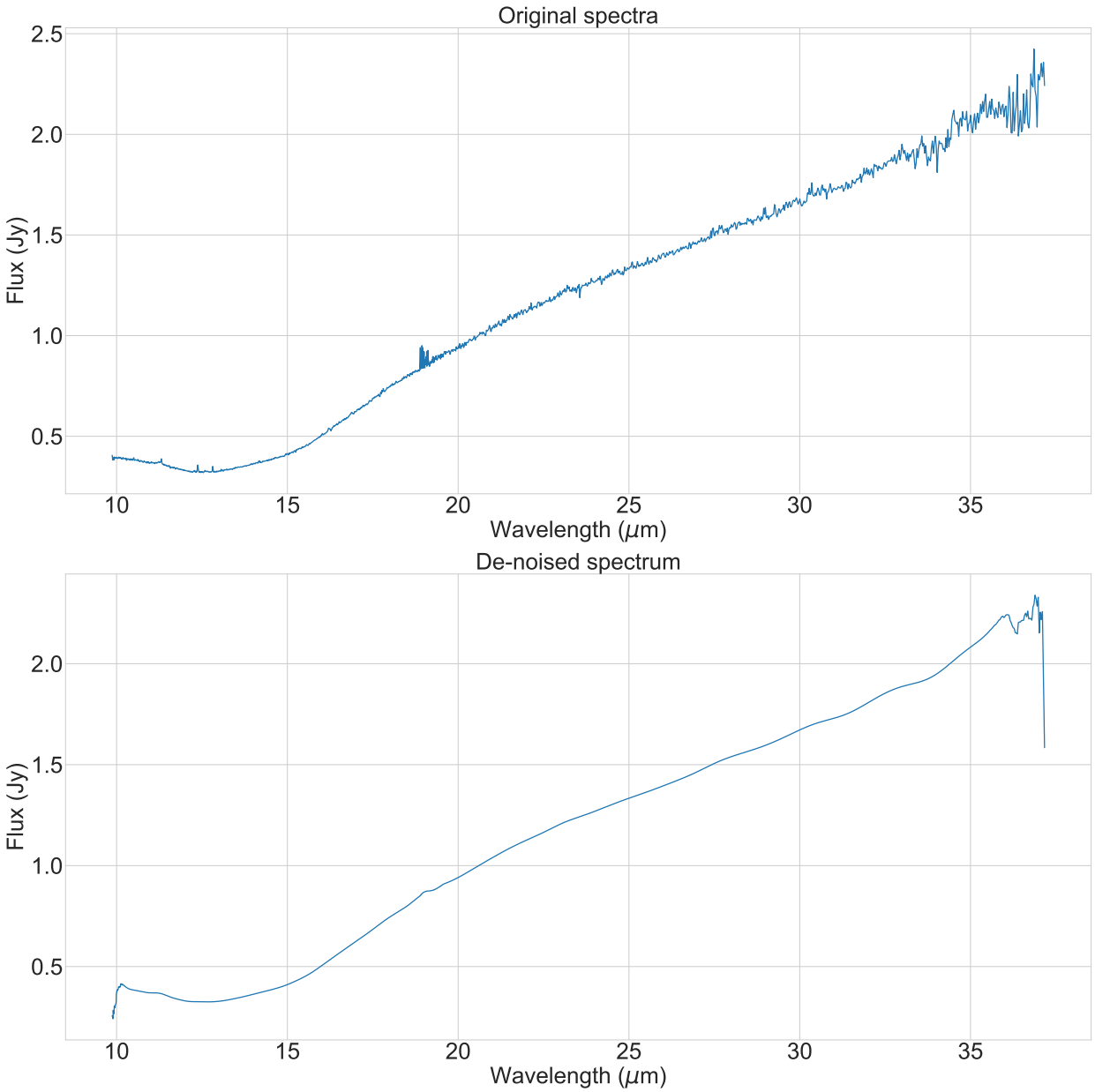

Figure 26: GM Aurigae

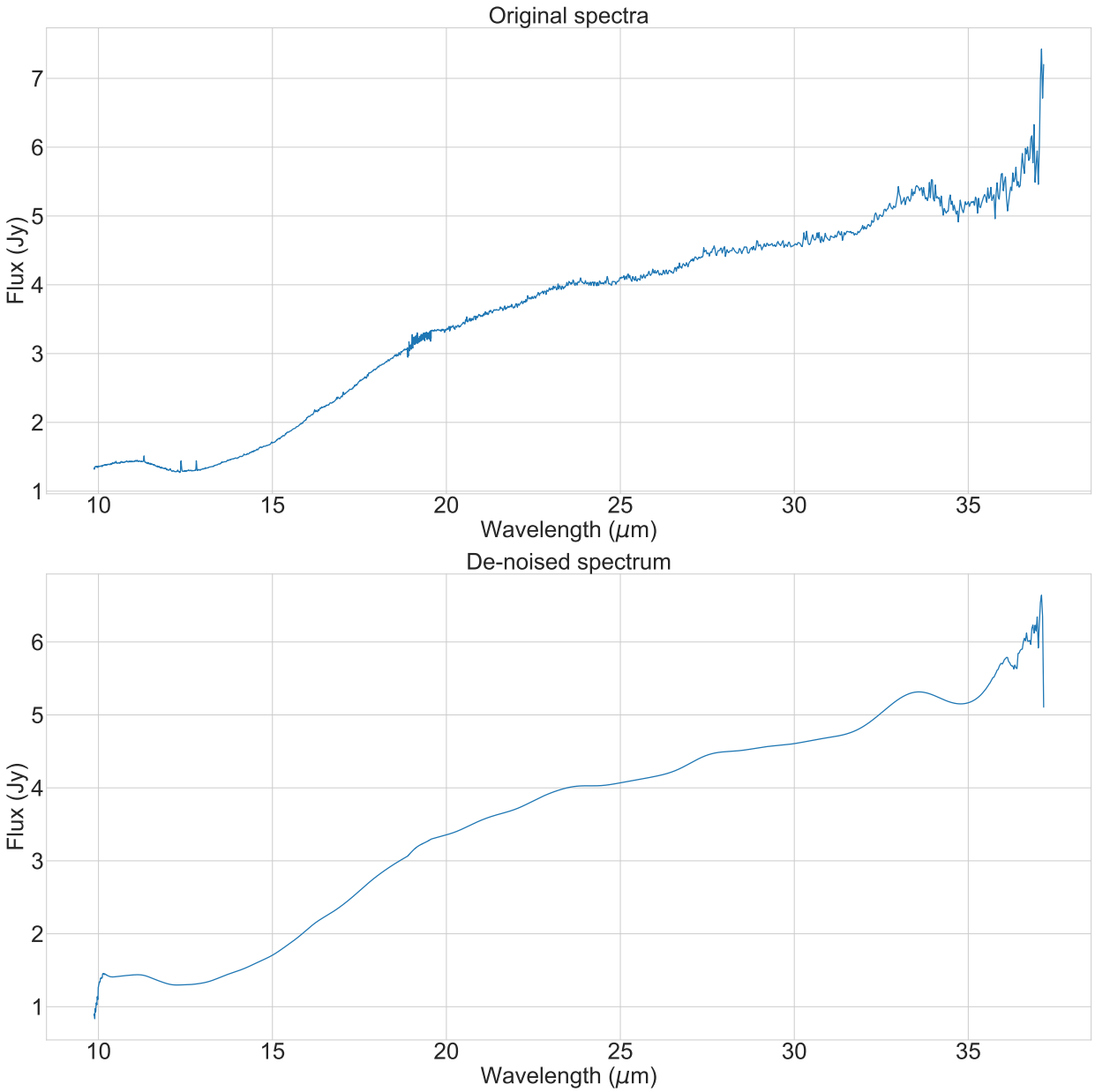

Figure 27: TW Hydrae

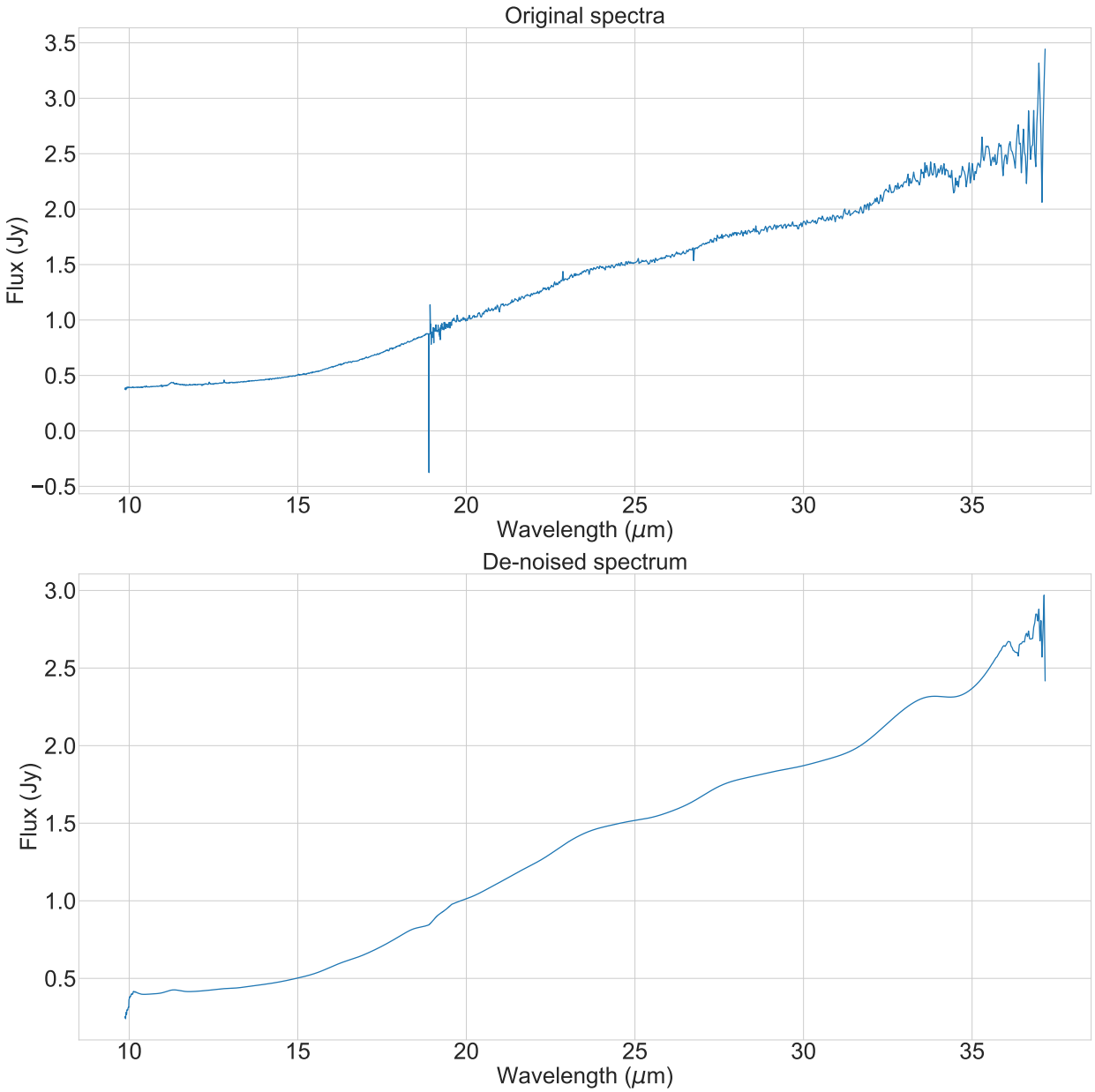

Figure 28: UX Tauri A

5.2 Continuum determination

The continuum location determination results are presented in the following way:

- The top left plot of each figure shows the added continuum.
- The top right plot shows the spectra with the added continuum.

- Under the top row will be the smoothed array c_j at scale j of Equation 1, ranging from scales j = 1 to 8.

An examination of the WT of the models at $\sigma = 10^{-3}$, 10^{-4}, 10^{-5} shows that when the signal is detectable, the smoothed array is exactly the same (see Figs. 29 & 30). Therefore, all WT were performed with added noise of only $\sigma = 10^{-6}$, unless otherwise stated in the caption, to ensure that the signal is detectable. The low value of σ is due to the low relative line strengths of some models relative to their continuum flux, see e.g. the order of the magnitude of the flux of the ProDiMo slab model in Fig 3.

5.2.1 Aaron's FLiTs models

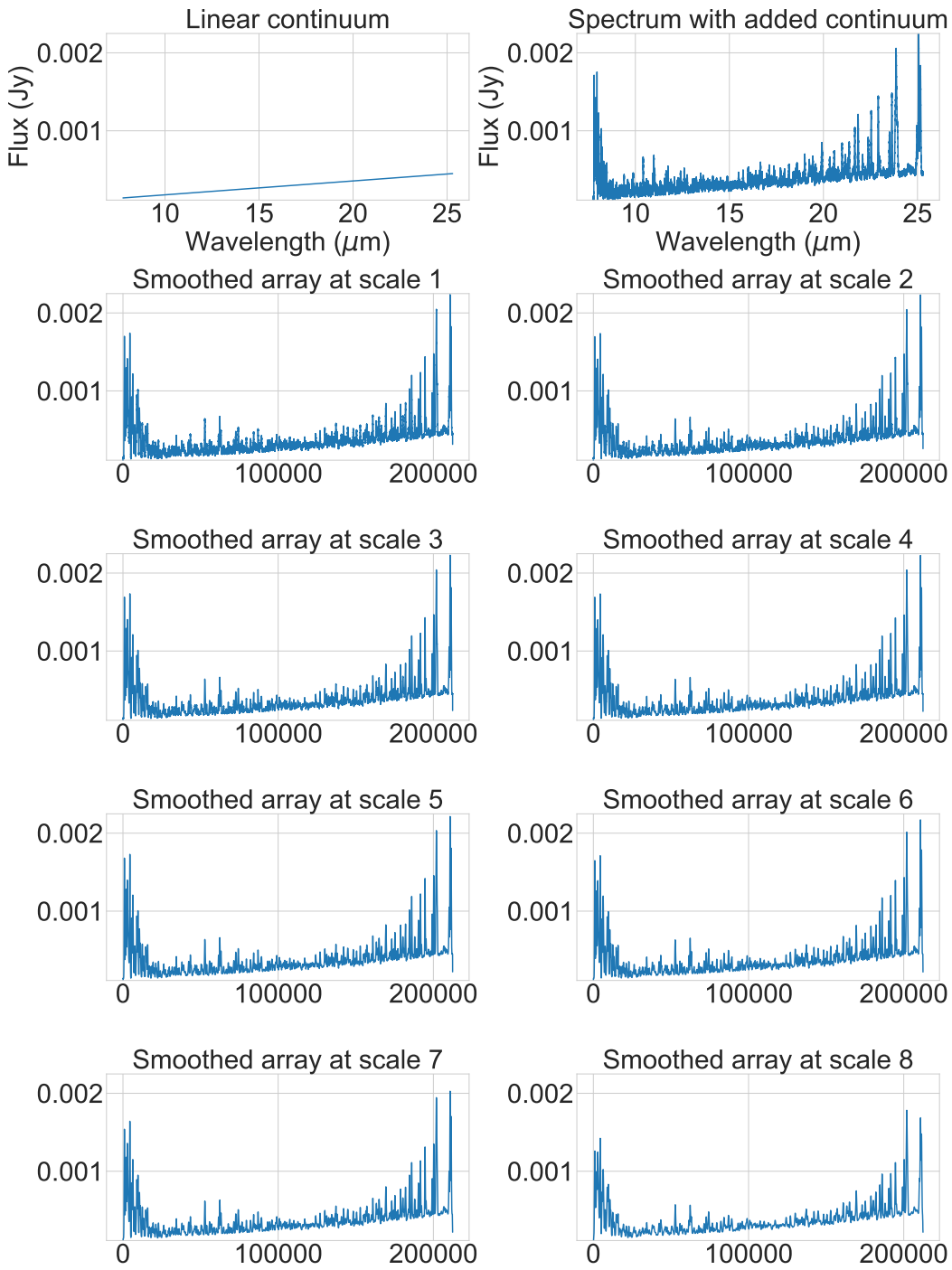

Figure 29: Aaron's FLiTs model (H2O), with added noise $\sigma = 10^{-5}$

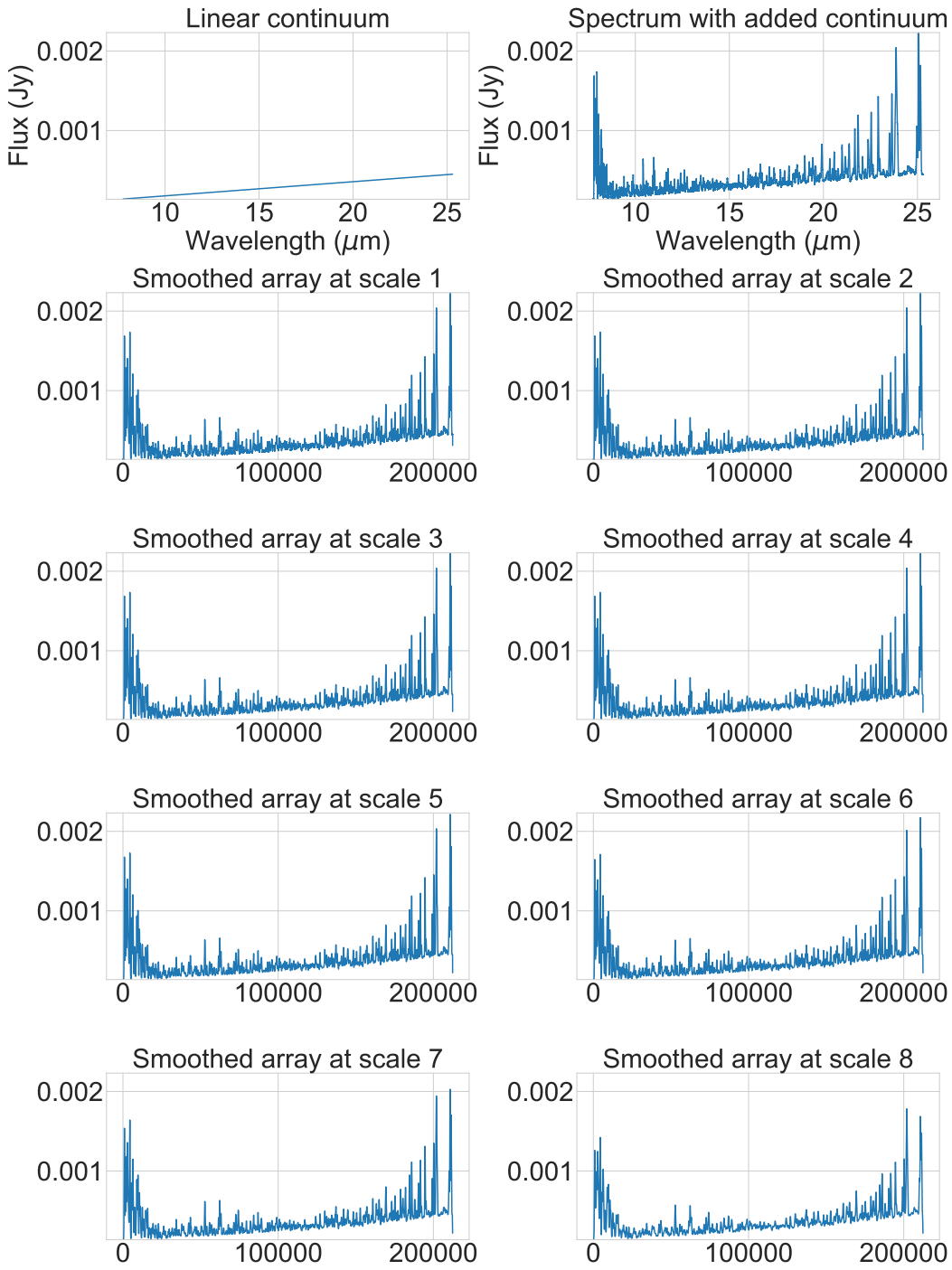

Figure 30: Aaron's FLiTs model (H2O)

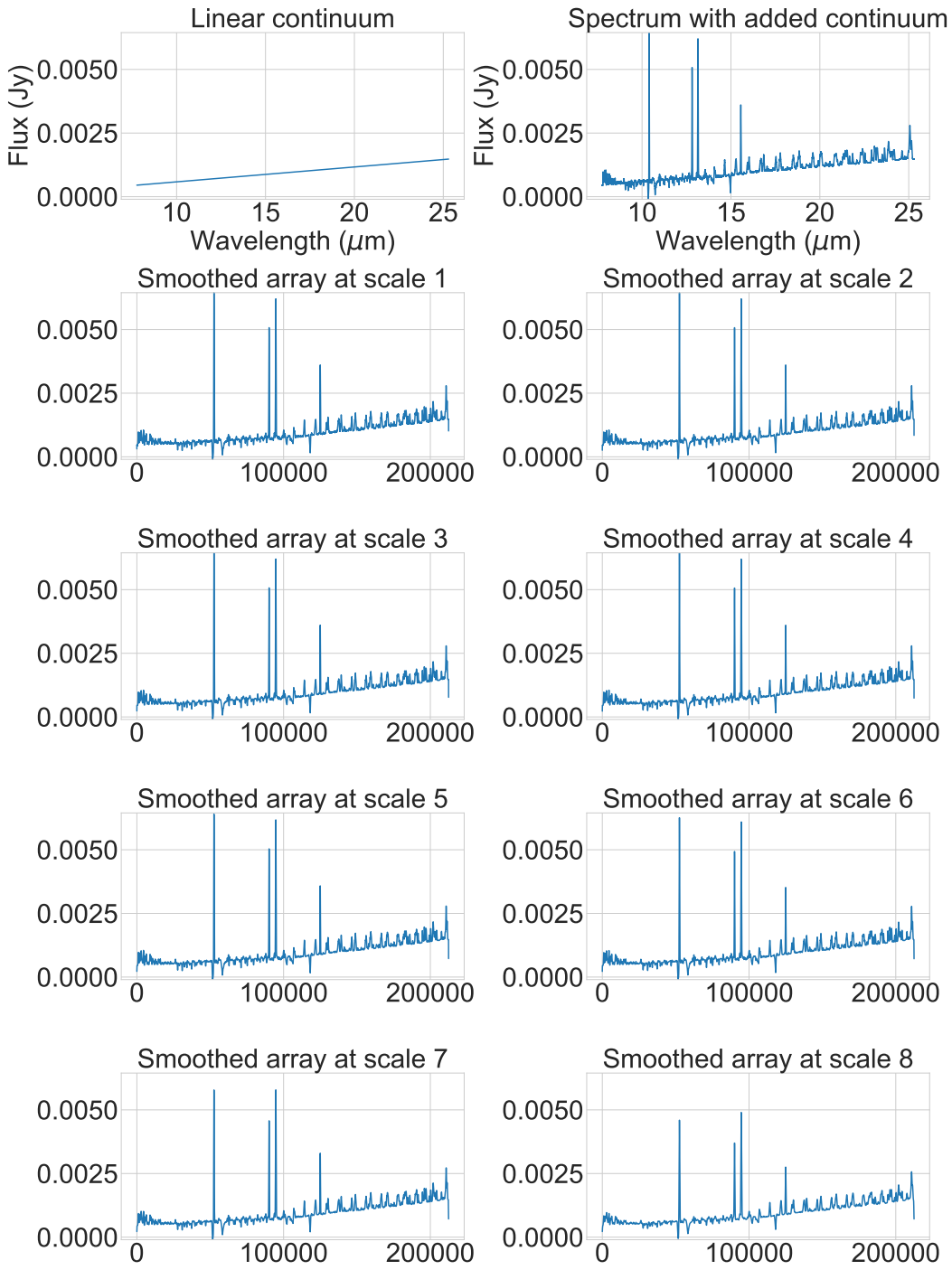

Figure 31: Aaron's FLiTs model (all molecules)

5.2.2 Fred's water slab models

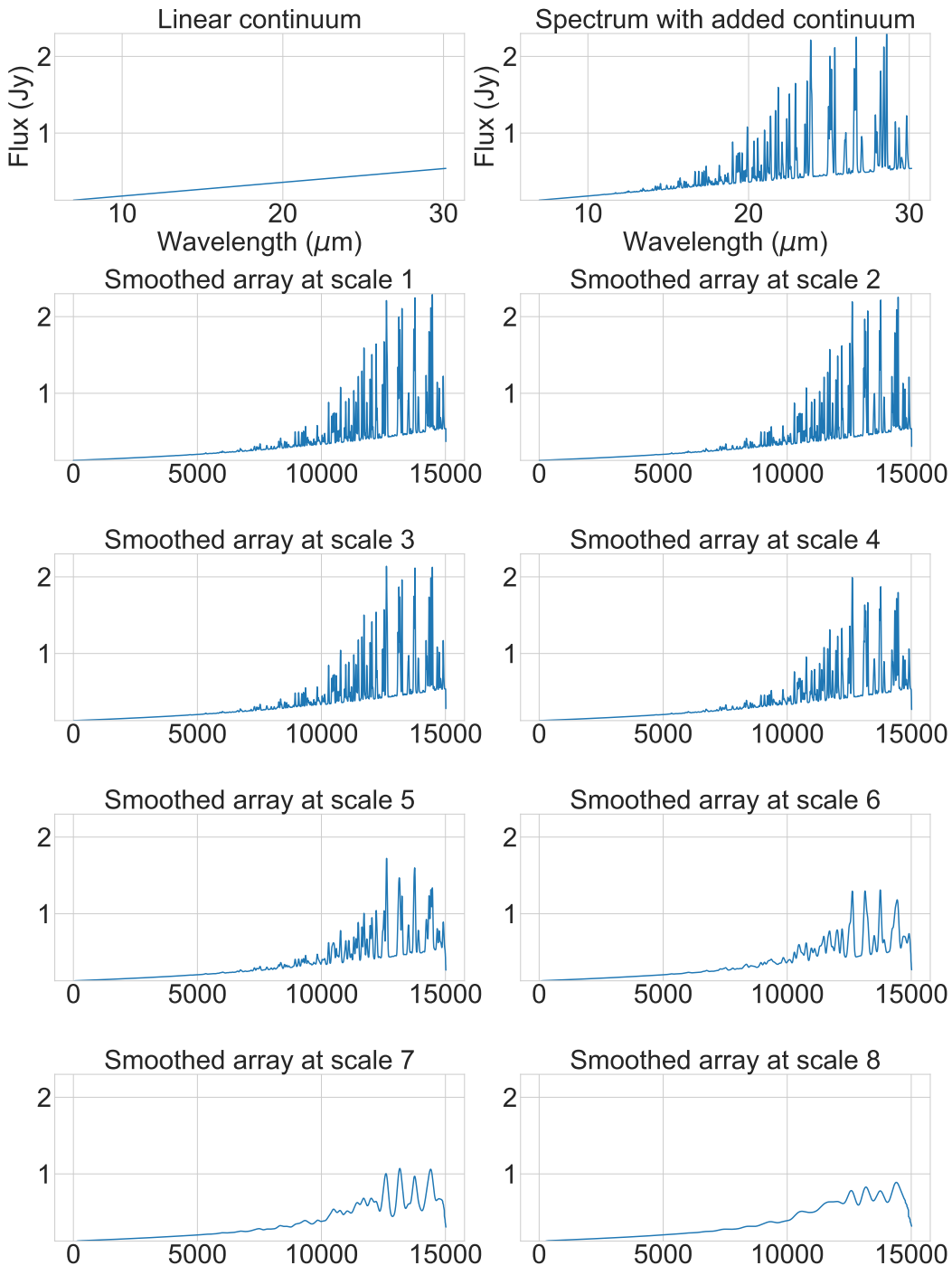

Figure 32: Fred's water slab model at T = 450K

5.2.3 ProDiMo slab model

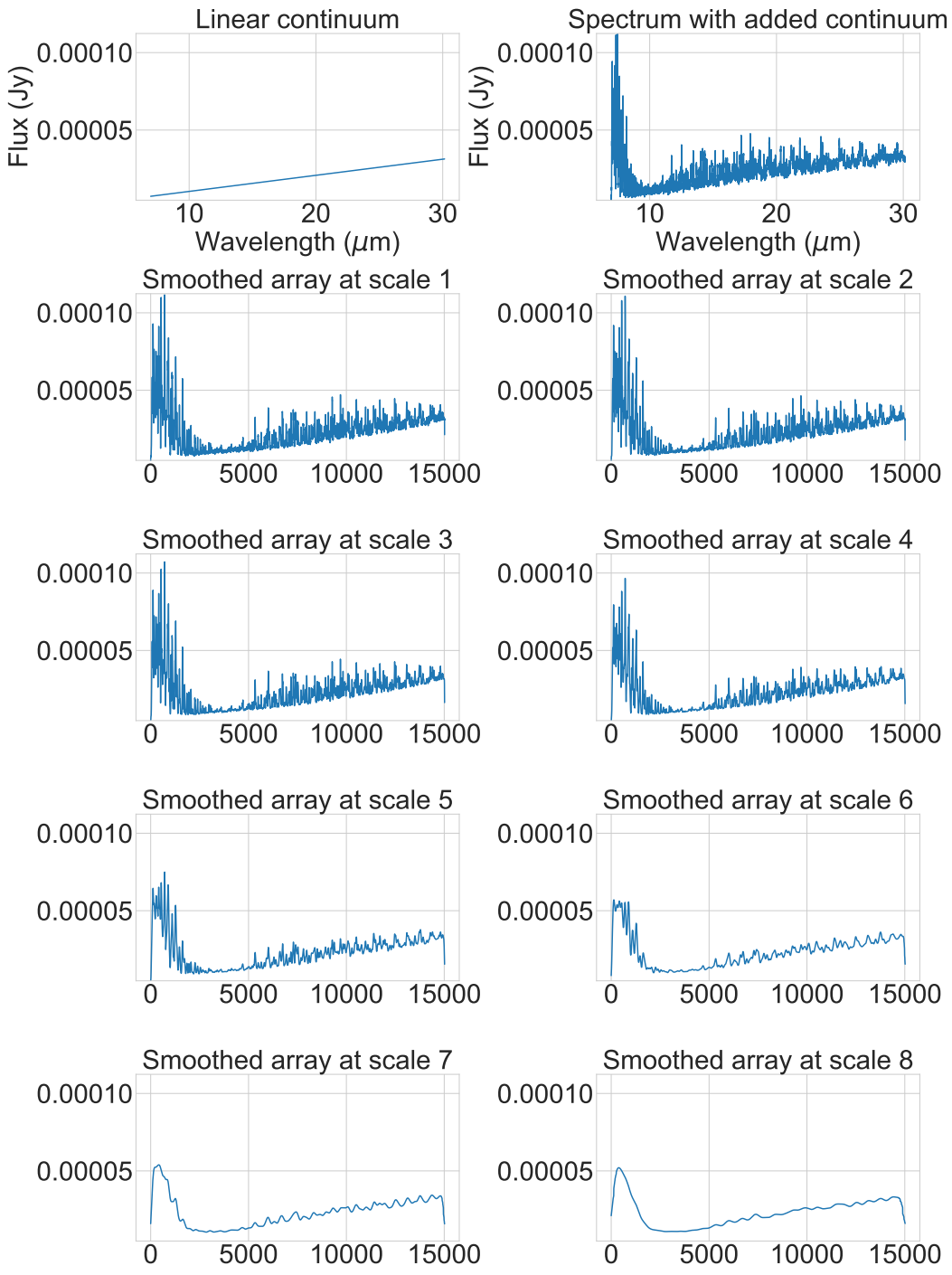

Figure 33: ProDiMo water slab model

5.2.4 Spitzer data

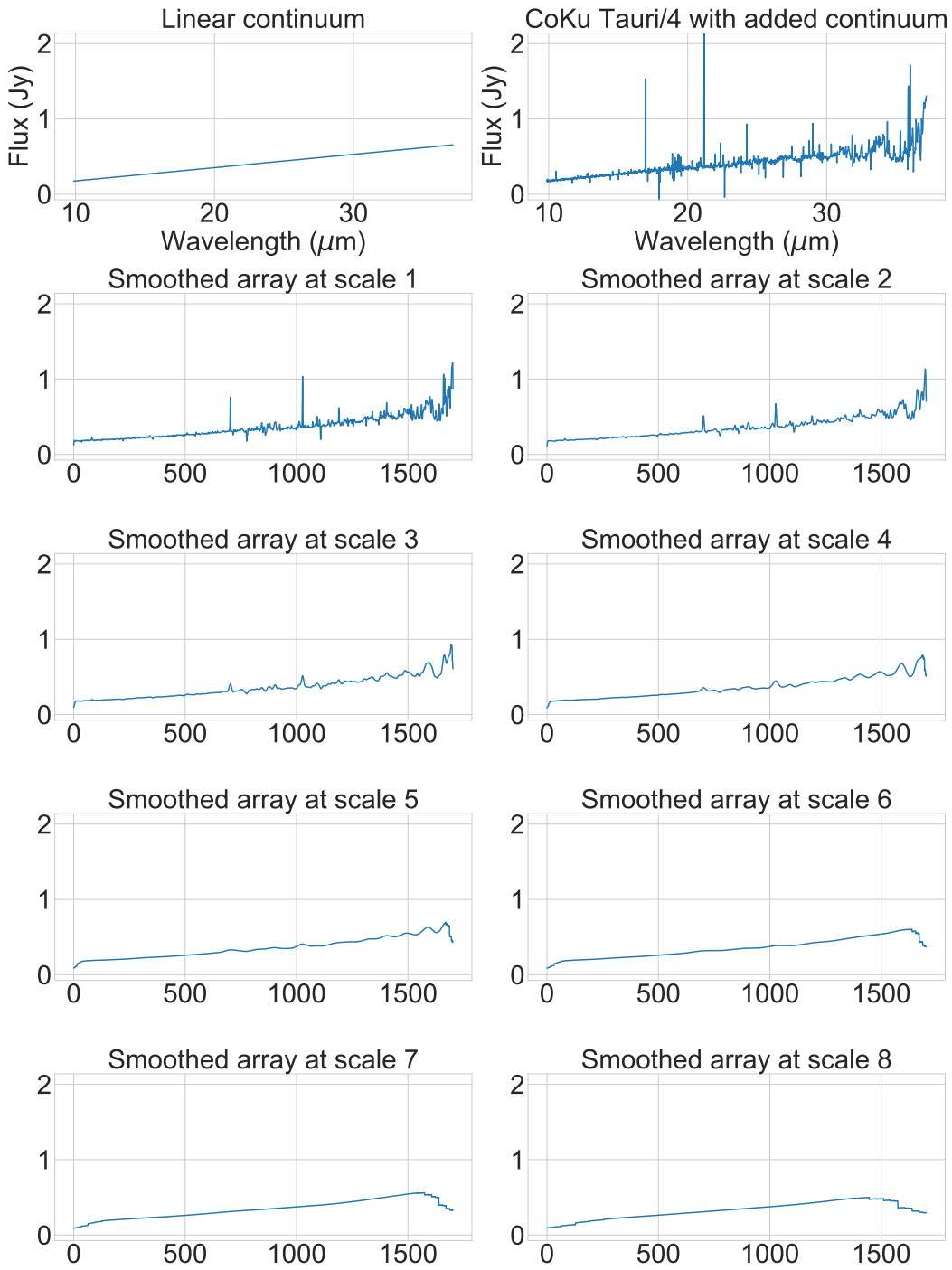

Figure 34: CoKu Tauri/4

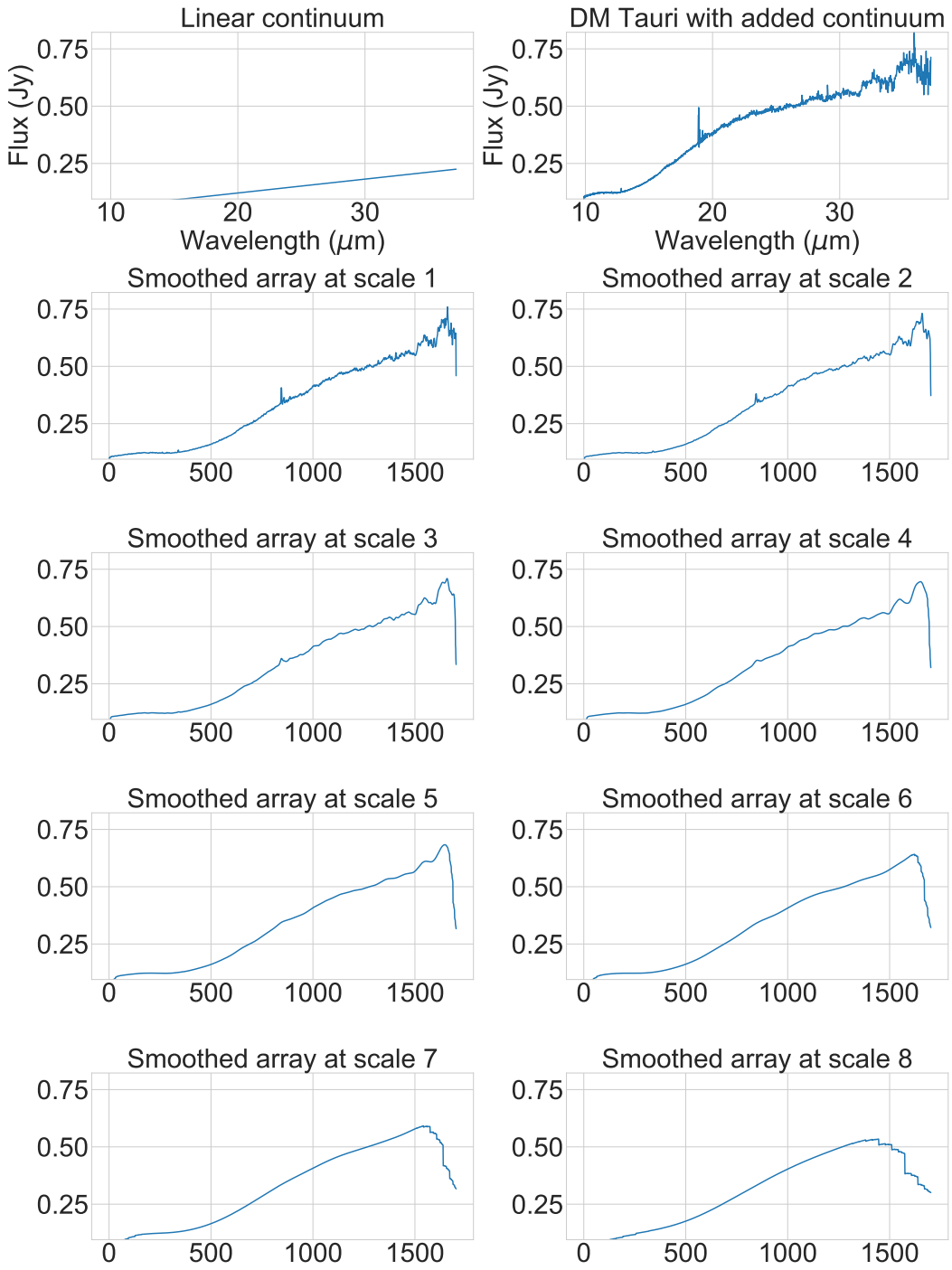

Figure 35: DM Tauri

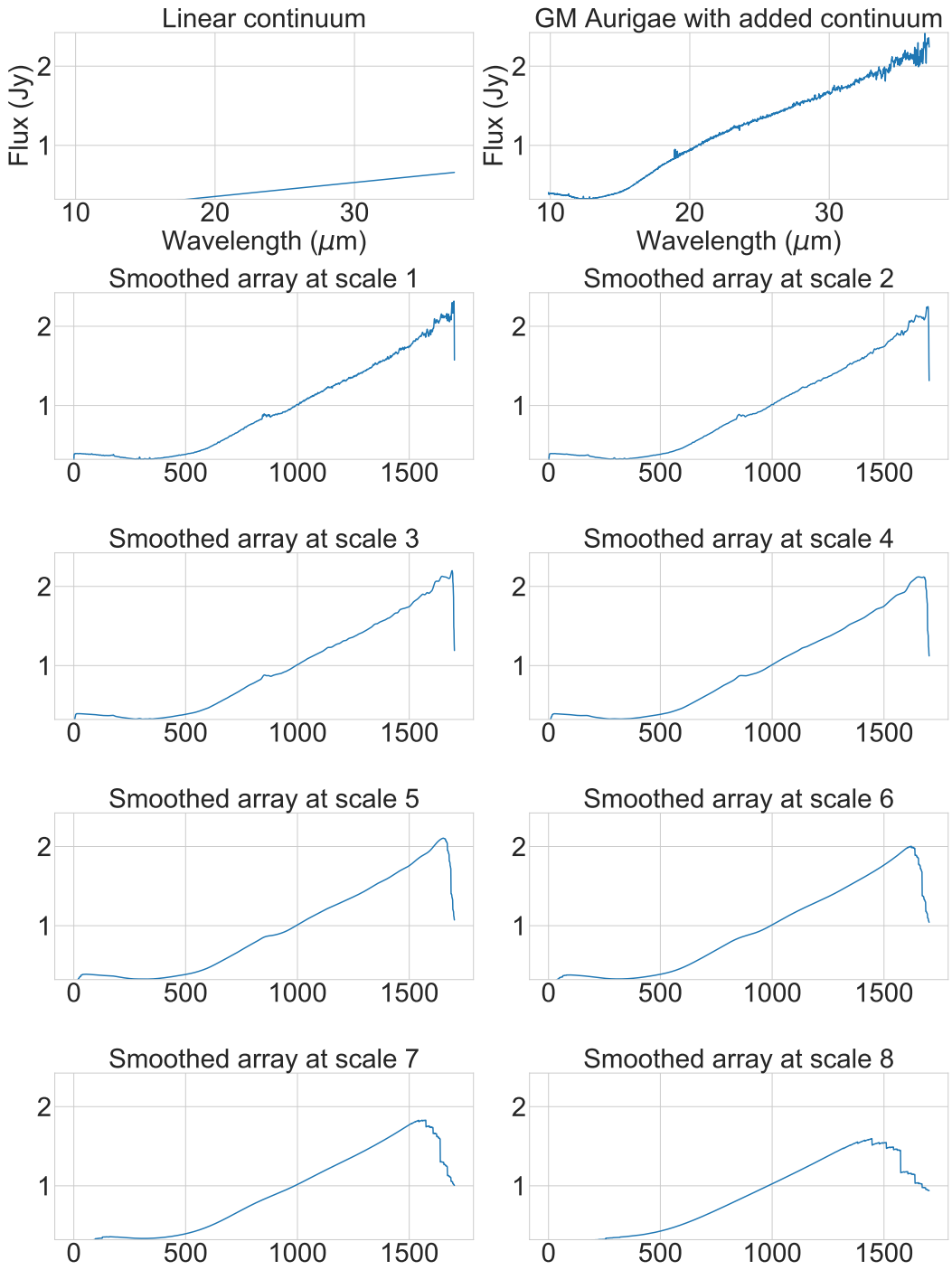

Figure 36: GM Aurigae

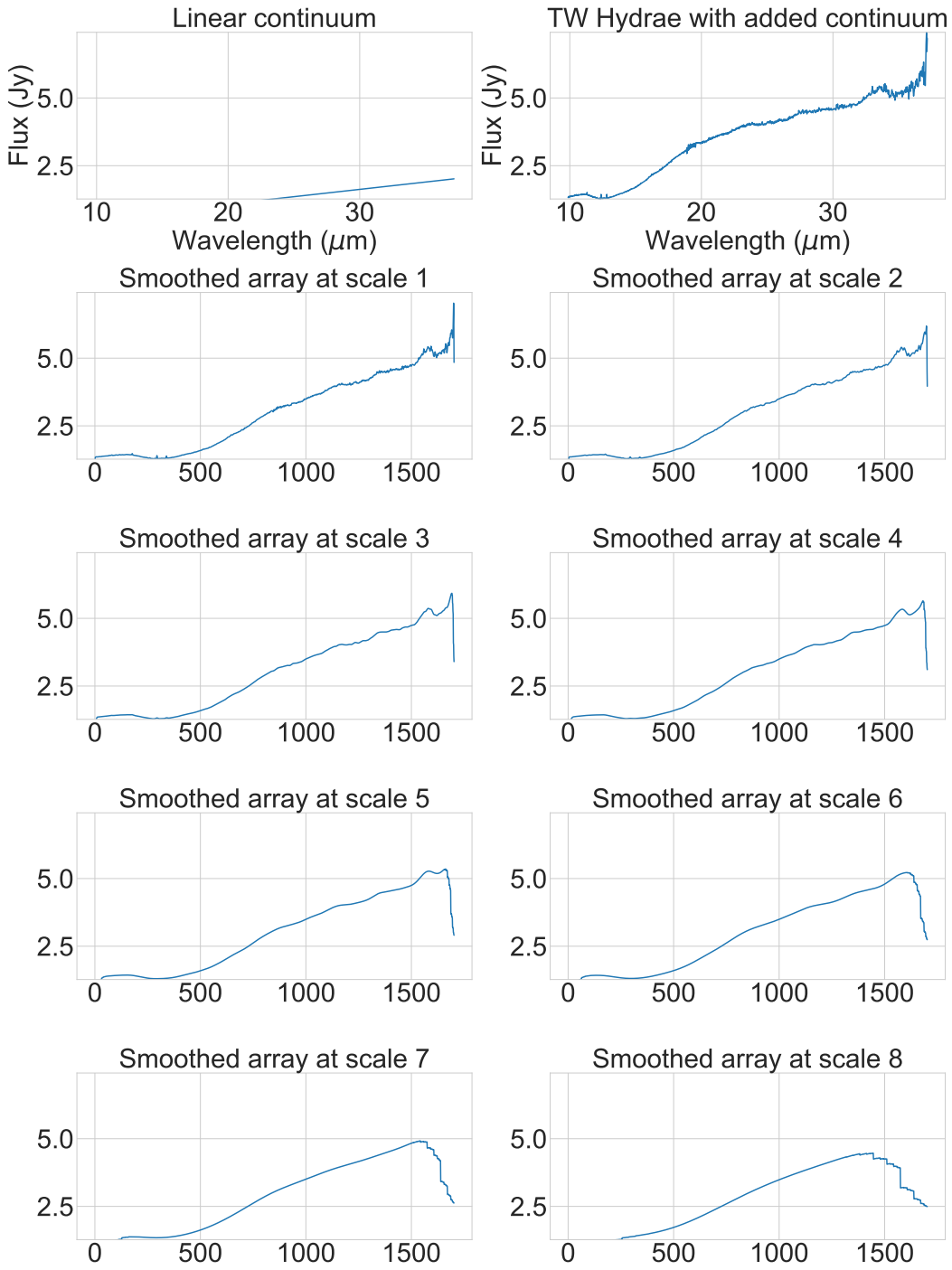

Figure 37: TW Hydrae

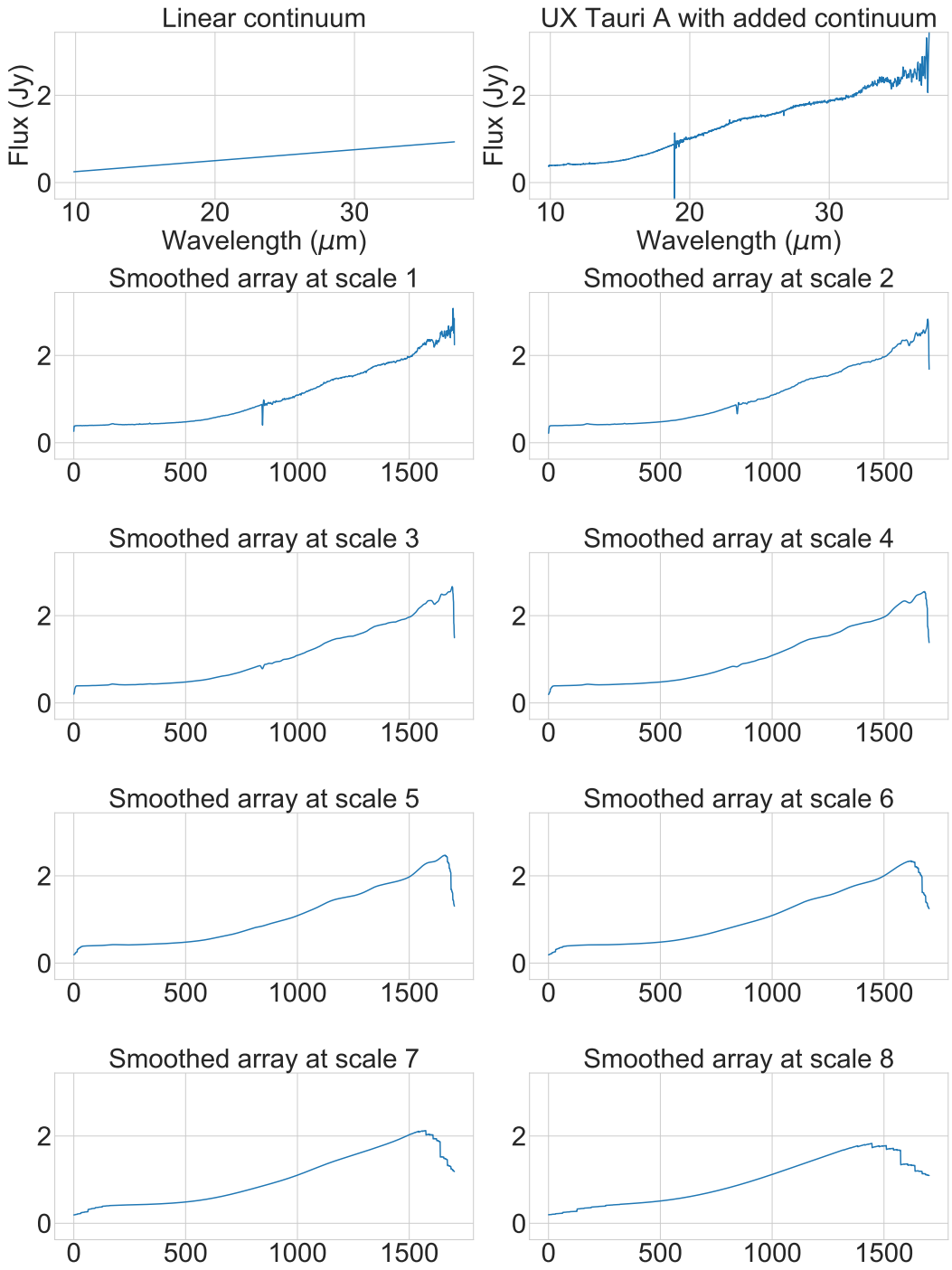

Figure 38: UX Tauri A

6 Discussion

6.1 Viability of the Wavelet Transform

6.1.1 De-noising

In Figs. 12-28 (disregarding the first plot of each model, Figs. 12, 15, 18 which only serve to show the behaviour of the WT when the SNR is too low), we see that the WT performs quite well, even when using only a two-level filtering as opposed to the more accurate three-level method as performed by Fligge and Solanki (1997, 1998).

We see that the WT de-noising breaks around $\sigma = 10^{-3}$ for Aaron's models, 10^{-1} for Fred's models, and 10^{-4} for the ProDiMo models. This is due to the weak line strengths of some models relative to the continuum flux (line-to-continuum ratio), see e.g. the order of the magnitude of the flux relative to the intensity of the weak 'forest' of the ProDiMo slab model in Fig 3. With such low intensities, the lines are only detectable with such low noise. This would also implies that the WT method will break at a lower σ if there is a forest of weak lines, which acts like a quasi-continuum.

How effective (accurate) and efficient (computationally) it is is yet to be examined. De-noising using the WT method should be compared against other de-noising methods in future studies to determine this.

6.1.2 Continuum

From Figs. 28-37, we see that the smoothed array is a combination of both the continuum and line emissions (i.e. they are entangled with one another), and the but the isolated linear continuum was never found by itself. However, the continuum is well recovered for *Spitzer* spectra.

6.2 Limitations of the Wavelet Transform

Signal-to-Noise Ratio (SNR)
The SNR must be high enough a signal is detectable above the noise. If not, the WT fails, but this is a limitation applying to all de-noising methods, and is not limited to the WT.

We obtain the line strength by using the line-to-continuum ratio, and compare that to the introduced noise σ. The line-to-continuum ratio should be above 3σ, where σ is that of our introduced noise to ensure that there is a signal within the noise.

Secondary peaks
The use of a B_3 spline wavelet function means that if there is a detection at wavelength

λ_i, then two smaller secondary peaks at opposite signs, symmetrically displaced from the detection will be observed. See Starck et al. (1997) §2.4.2 for a brief elaboration of this problem, illustrated in Fig. 6 then discussed in §2.6, of that same paper. As stated in that paper, this problem is resolved by subtracting the detected band from the original spectrum first.

7 Conclusion

When fitting observed spectra, a slab model should only be used the wavelength range is narrow enough such that the emissions lines all originate from levels with very similar excitation temperature ($<1\mu$m). If the wavelength range is broader, the excitation temperatures will vary too much and a slab model with only single temperature would not work.

De-noising using a multi-level wavelet transform methods works up to a certain level, limited by the SNR of the spectra we are analysing.

For the determination the continuum location, the continuum of all models as well as the *Spitzer* spectra appears to be well recovered via this method.

However, before we can come to a verdict whether the WT is viable or not, it should first be compared to other methods of de-noising and continuum determination to evaluate its accuracy/effectiveness and efficiency.

References

Antonellini, S. (2016). *Water in protoplanetary disks: Line flux modeling and disk structure*. PhD thesis, Kapteyn Astronomical Institute, University of Groningen, Postbus 800, 9700 AV Groningen, the Netherlands.

Aresu, G., Kamp, I., Meijerink, R., Woitke, P., Thi, W.-F., and Spaans, M. (2011). X-ray impact on the protoplanetary disks around T Tauri stars. *A&A*, 526:A163.

Daniel, F., Dubernet, M.-L., and Grosjean, A. (2011). Rotational excitation of 45 levels of ortho/para-H_2O by excited ortho/para-H_2 from 5 K to 1500 K: state-to-state, effective, and thermalized rate coefficients. *A&A*, 536:A76.

Daubechies, I. (1992). *Ten lectures on wavelets*.

Dionatos, O. (2015). Gas line observations of disks*. *EPJ Web of Conferences*, 102:00008.

Dominik, C. (2015). Disk formation and structure*. *EPJ Web of Conferences*, 102:00002.

Faure, A. and Josselin, E. (2008a). Collisional excitation of water in warm astrophysical media. I. Rate coefficients for rovibrationally excited states. *A&A*, 492:257–264.

Faure, A. and Josselin, E. (2008b). VizieR Online Data Catalog: Collisional excitation of water in warm media (Faure+, 2008). *VizieR Online Data Catalog*, 349.

Fligge, M. and Solanki, S. (1997). Noise reduction in astronomical spectra using wavelet packets. *A&AS*, 124:579–587.

Fligge, M. and Solanki, S. K. (1998). Noise Reduction in Astronomical Spectra: A New Wavelet-Based Method. 154:833.

Garufi, A., Podio, L., Kamp, I., Ménard, F., Brittain, S., Eiroa, C., Montesinos, B., Alonso-Martínez, M., Thi, W. F., and Woitke, P. (2014). The protoplanetary disk of FT Tauri: multiwavelength data analysis and modeling. , 567:A141.

Greenwood, A. (2018). *The future of protoplanetary disks models*. PhD thesis, Kapteyn Astronomical Institute, University of Groningen, Postbus 800, 9700 AV Groningen, the Netherlands.

Kamp, I. (2015). Modeling and interpretation of line observations*. *EPJ Web of Conferences*, 102:00013.

Kamp, I., Tilling, I., Woitke, P., Thi, W.-F., and Hogerheijde, M. (2010). Radiation thermo-chemical models of protoplanetary disks. II. Line diagnostics. , 510:A18.

Pontoppidan, K. M., Salyk, C., Blake, G. A., Meijerink, R., Carr, J. S., and Najita, J. (2010). A spitzer survey of mid-infrared molecular emission from protoplanetary disks. i. detection rates. *The Astrophysical Journal*, 720(1):887.

Rigliaco, E., Pascucci, I., Duchene, G., Edwards, S., Ardila, D. R., Grady, C., Mendigutía, I., Montesinos, B., Mulders, G. D., Najita, J. R., Carpenter, J., Furlan, E., Gorti, U., Meijerink, R., and Meyer, M. R. (2015). Probing stellar accretion with mid-infrared hydrogen lines. *The Astrophysical Journal*, 801(1):31.

Rothman, L. S., Gordon, I. E., Babikov, Y., Barbe, A., Chris Benner, D., Bernath, P. F., Birk, M., Bizzocchi, L., Boudon, V., Brown, L. R., Campargue, A., Chance, K., Cohen, E. A., Coudert, L. H., Devi, V. M., Drouin, B. J., Fayt, A., Flaud, J.-M., Gamache, R. R., Harrison, J. J., Hartmann, J.-M., Hill, C., Hodges, J. T., Jacquemart, D., Jolly, A., Lamouroux, J., Le Roy, R. J., Li, G., Long, D. A., Lyulin, O. M., Mackie, C. J., Massie, S. T., Mikhailenko, S., Müller, H. S. P., Naumenko, O. V., Nikitin, A. V., Orphal, J., Perevalov, V., Perrin, A., Polovtseva, E. R., Richard, C., Smith, M. A. H., Starikova, E., Sung, K., Tashkun, S., Tennyson, J., Toon, G. C., Tyuterev, V. G., and Wagner, G. (2013). The HITRAN2012 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 130:4–50.

Rothman, L. S., Rinsland, C. P., Goldman, A., Massie, S. T., Edwards, D. P., Flaud, J.-M., Perrin, A., Camy-Peyret, C., Dana, V., Mandin, J.-Y., Schroeder, J., McCann, A., Gamache, R. R., Wattson, R. B., Yoshino, K., Chance, K., Jucks, K., Brown, L. R.,

Nemtchinov, V., and Varanasi, P. (1998). The HITRAN Molecular Spectroscopic Database and HAWKS (HITRAN Atmospheric Workstation): 1996 Edition. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 60:665–710.

Salyk, C., Pontoppidan, K. M., Blake, G. A., Najita, J. R., and Carr, J. S. (2011). A spitzer survey of mid-infrared molecular emission from protoplanetary disks. ii. correlations and local thermal equilibrium models. *The Astrophysical Journal*, 731(2):130.

Shensa, M. J. (1992). The discrete wavelet transform: wedding the a trous and mallat algorithms. *IEEE Transactions on Signal Processing*, 40(10):2464–2482.

Starck, J.-L. and Bijaoui, A. (1994). Filtering and deconvolution by the wavelet transform. *Signal Processing*, 35(3):195 – 211.

Starck, J.-L. and Murtagh, F. (1994). Image restoration with noise suppression using the wavelet transform. *A&A*, 288:342–348.

Starck, J.-L. and Murtagh, F. (1998). Automatic Noise Estimation from the Multiresolution Support. *Publications of the Astronomical Society of the Pacific*, 110:193–199.

Starck, J.-L., Murtagh, F., and Bijaoui, A. (1995). Multiresolution support applied to image filtering and restoration. *Graphical Models and Image Processing*, 57:420–431.

Starck, J.-L., Murtagh, F., and Fadili, J. (2015). *Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis*. Cambridge University Press.

Starck, J.-L., Siebenmorgen, R., and Gredel, R. (1997). Spectral Analysis Using the Wavelet Transform. *The Astrophysical Journal*, 482:1011–1020.

Williams, J. P. and Cieza, L. A. (2011). Protoplanetary Disks and Their Evolution. *ARA&A*, 49:67–117.

Woitke, P., Kamp, I., and Thi, W. F. (2009). Radiation thermo-chemical models of protoplanetary disks. I. Hydrostatic disk structure and inner rim. *A&A*, 501:383–406.

Woitke, P., Min, M., Pinte, C., Thi, W.-F., Kamp, I., Rab, C., Anthonioz, F., Antonellini, S., Baldovin-Saavedra, C., Carmona, A., Dominik, C., Dionatos, O., Greaves, J., Güdel, M., Ilee, J. D., Liebhart, A., Ménard, F., Rigon, L., Waters, L. B. F. M., Aresu, G., Meijerink, R., and Spaans, M. (2016). Consistent dust and gas models for protoplanetary disks. I. Disk shape, dust settling, opacities, and PAHs. *A&A*, 586:A103.