Title:	Adaptive Emission Line Detection		
Date:	2015/04/15	**Issue:**	Draft 0.1
Reference:	EUCL-IPN-TN-8-002		
Custodian:	Yannick Copin (y.copin@ipnl.in2p3.fr)		

Authors:		**Date:**	**Signature:**
	Yannick Copin *(IPNL)*	2015/04/15	
Contributors:			
Approved by:			
Authorized by:			

Document version tracking

Issue	Date	Page	Description of changes	Comments
0.1	2015/04/15	11	First release	Very first incomplete draft

Table of contents

1 Purpose

To derive spectral quantities such as redshift, the first step in the process of precision emission line adjustment in a galaxy spectrum is to detect potential emission lines in a robust and systematic way. I present in this note a collection of algorithms developed for that purpose, including optimal spectrum combination, adaptive fitting and robust M-estimators.

2 Scope

Continuum and emission line identification for subsequent spectral measurements in OU-SPE.

Applicable work packages:

1. *Spectra Combination* Wp-4-3-07-5100 Simulation_140425
2. *Lines Identification and Redshift Measurement* Wp-4-3-07-5200
3. *Redshift Quality Determination* Wp-4-3-07-5400

3 Applicable & Reference documents

3.1 Applicable documents

RD		Ref.	Date

3.2 Reference documents

RD		Ref.	Date
	Impact of spectral covariance on line fitting	EUCL-IPN-TN-8-001	2014/09/12

4 Acronyms

ML	Maximum Likelihood
PDF	Probability Density Function
SL	Significance Level
SNR	Signal-to-Noise Ratio
TBC	To Be Completed

5 Multi-roll spectrum combination

The present analysis is performed on the `Uncontaminated` simulated spectrum sample from `Simulations_140425`[1].

Current simulation issues:
To avoid all kind of covariance-inducing resampling issues, it was agreed in the OU-SPE/SIR meeting in Marseille (2015/01/26-27) that OU-SIR output 1D-spectra should already be *regularly* resampled (either linearly or logarithmically).

1. *Spectra should be linearly sampled in wavelength.*
   In the current simulation, sampling steps vary at the 10^{-4} level around a mean value of 9.80 Å/px.
2. *All spectra* (or at least the ones originating from the same object) *should share the same wavelength sampling step.*
   Besides fluctuations mentioned above, this seems to be the case in the current simulation.
3. *All spectra* (or at least the ones originating from the same object) *should share the same starting wavelength modulo sampling step.*
   In the current simulation, this is the case *most of the time* for rolling angles of 0 and 90, with start $=$ 4.80 Å [step] (with variations up to 0.1 Å). However, spectra with rolling angle of 180, one consistently has start $\sim$ 7.4 Å [step], somehow incompatible with other orientations.
   I suggest all resampled spectra should have start $=$ 0 [step].

Furthermore,

4. *Beside spectral variance, spectra should incorporate some information about spectral covariance*, either as a full (probably sparse) covariance matrix, or as an adequate modeling of this covariance matrix (e.g. isotropic exponential covariance function scale length, see EUCL-IPN-TN-8-001).
   In the current simulation, only the variance is stored along the spectral signal. To my knowledge, it is not stated whether this variance incorporate the covariance term or not.

Modulo the current simulation issues presented above, the spectra $\boldsymbol{y}_i = y_i(\lambda)$ originating from the same target but corresponding to different roll angles are combined using a standard inverse-variance weighted average[2] (see Fig. 3 and 4):

$$y(\lambda) = \frac{1}{\sum_{i=1}^{N} 1/\sigma_i^2(\lambda)} \sum_{i=1}^{N} \frac{y_i(\lambda)}{\sigma_i^2(\lambda)}, \qquad \sigma_y^2(\lambda) = \frac{1}{\sum_{i=1}^{N} 1/\sigma_i^2(\lambda)}, \tag{1}$$

where the number N of spectra combined actually depends on wavelength λ, since the spectral coverage domain varies between the different roll exposures (see Fig. 4).

In Eq. (1), $\boldsymbol{y} = y(\lambda)$ is the maximum-likelihood (ML) estimate of the mean intrinsic signal $\boldsymbol{\mu}$ in the case where the measurements are *normally* distributed, $\boldsymbol{y}_i = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}_i^2)$. If the noise distribution is not strictly normal – e.g. contribution of a Poisson shot noise or presence of outliers from external contamination or artifacts – the above ML-estimate is no more valid, and alternative techniques (Cash statistics, pull-clipping, etc.) should be used.

6 Adaptive continuum fit

An emission line spectrum is first characterized by its continuum level, on top of which the emission lines will appear. The continuum could display some significant discontinuities (e.g. Balmer break), but it is assumed here that it is "sufficiently" smooth to be properly modeled as a low-degree polynomial. It is then standard practice to adjust a polynomial to the signal to estimate the continuum, with two caveats however:

- the degree of the continuum polynomial is not known *a priori*,
- the polynomial adjustment can be biased by the presence of emission lines.

I present here two methods to address these issues:

[1] `http://euclidsims.lambrate.inaf.it/`

[2] In the absence of information on spectral covariance, it is currently simply ignored, but all equations can be adapted for the presence of a covariance matrix. See Sect. 8 on possible ways to estimate *a posteriori* this matrix.

– an adaptive fit scheme to estimate the most adequate polynomial degree from the signal itself (Sect. 6.1),
– a robust adjustment procedure to account for the potential presence of emission lines in the spectrum (Sect. 6.2).

These methods are mostly generic and can be applied to other cases with similar issues.

6.1 Adaptive fit

To adjust n observations $\boldsymbol{y} = \{y_i\}$ with a parametric model $\boldsymbol{M}(\boldsymbol{p}) = \{M_i(\boldsymbol{p})\}$ depending on m parameters $\boldsymbol{p}$, the standard inverse-variance weighted least-square method (so-called "χ^2") correspond to the ML estimate of the parameters under the normality assumption:

$$\hat{\boldsymbol{p}} = \operatorname{argmin} \chi^2(\boldsymbol{p}) \qquad \text{with} \qquad \chi^2(\boldsymbol{p}) = \sum_{i=1}^{n} \left(\frac{y_i - M_i(\boldsymbol{p})}{\sigma_i} \right)^2 . \tag{2}$$

The goodness of the fit can then be estimated from the p-value of the χ^2 statistic with $\delta = n - m$ degrees of freedom: this gives the probability that, under the null hypothesis that the adjusted model appropriately represent the signal, such a high statistic value can be obtained just by chance. If $p < p_{\max} \ll 1$ (i.e. $\chi^2 \gg \delta$ for large enough δ), the null hypothesis can be rejected at significance level (SL) $p_{\max}$: the adjusted model does not fairly represent the signal[3]. In the opposite, if $p_{\max} < p \lesssim 1$ (i.e. $\chi^2 \sim \delta$), the null hypothesis cannot be rejected: the adjusted model appropriately describe the signal. Note however that $p \to 1$ (i.e. $\chi^2 \ll \delta$) is an indication of *over-fitting*: the model has too many degrees of freedom, or more probably the observation variances $\boldsymbol{\sigma}$ are under-estimated or do not account for a covariance term (see Sect. 8).

If one wants to compare two models adjusted to the same signal $\boldsymbol{y}$, a "simpler" one $\boldsymbol{M}_1$ depending on m_1 parameters, and a more "elaborate" (presumably better) one $\boldsymbol{M}_2$ with $m_2 > m_1$ parameters. Except under extreme cases, it is usually not enough to compare their goodness-of-fit, as both can provide equally valid fit ($p_1, p_2 > p_{\max}$) to the signal. However, in the usual case where the second model is a "extension" of the first one – i.e. same type of modeling but with more flexibility (e.g. a higher order polynomial), the $\Delta\chi^2 = \chi_1^2 - \chi_2^2 \geq 0$ difference follows a χ^2 statistic with $\Delta = m_2 - m_1$ degrees of freedom[4]. The resulting p_Δ-value can therefore be used to assess the significance of the improvement ($p_\Delta < p_{\max}$) or absence thereof ($p_\Delta \geq p_{\max}$) brought by more complex model $\boldsymbol{M}_2$ with respect to simpler model $\boldsymbol{M_1}$.

The complexity of the adjusted model can therefore increased progressively – starting e.g. from the null model $\boldsymbol{M} \equiv \boldsymbol{0}$ – until it does not significantly improve the fit at a given SL $p_{\max}$:

Algorithm 1 Generic adaptive fit procedure to adjust increasingly complex models up to maximal SL $p_{\max}$.

procedure ADAPTIVEFIT($p_{\max}$, model class $\boldsymbol{M}$)
  Compute χ_1^2 for simplest model $\boldsymbol{M}_1$
  repeat
    Compute χ_2^2 for more complex model $\boldsymbol{M}_2$
    Compute p_Δ-value of fit improvement
    if $p_\Delta < p_{\max}$ **then** ▷ The more complex model provides a significantly better fit
      $M_1 \leftarrow M_2$
    end if
  until $p_\Delta \geq p_{\max}$ or maximal model complexity is reached
  return model $\boldsymbol{M}_1$ ▷ Last significantly better model
end procedure

A low p_Δ-value means either that the null hypothesis – i.e. the signal is properly described by the simpler model without the addition of extra parameters – is false, or that it is true but an improbable event has occurred. Higher ("liberal") value of significance level $p_{\max} \gtrsim 3\%$ will therefore be more sensitive to faint features, but also to false positive detections just due to natural noise fluctuations; conversely, a lower ("conservative") $p_{\max} \leq 1\%$ will provide more robust but less sensitive detections (low test power).

[3] Under the further critical assumption that the observation variance $\boldsymbol{\sigma}$ is properly estimated, see Sect. 8.
[4] This is a direct application of the likelihood-ratio test.

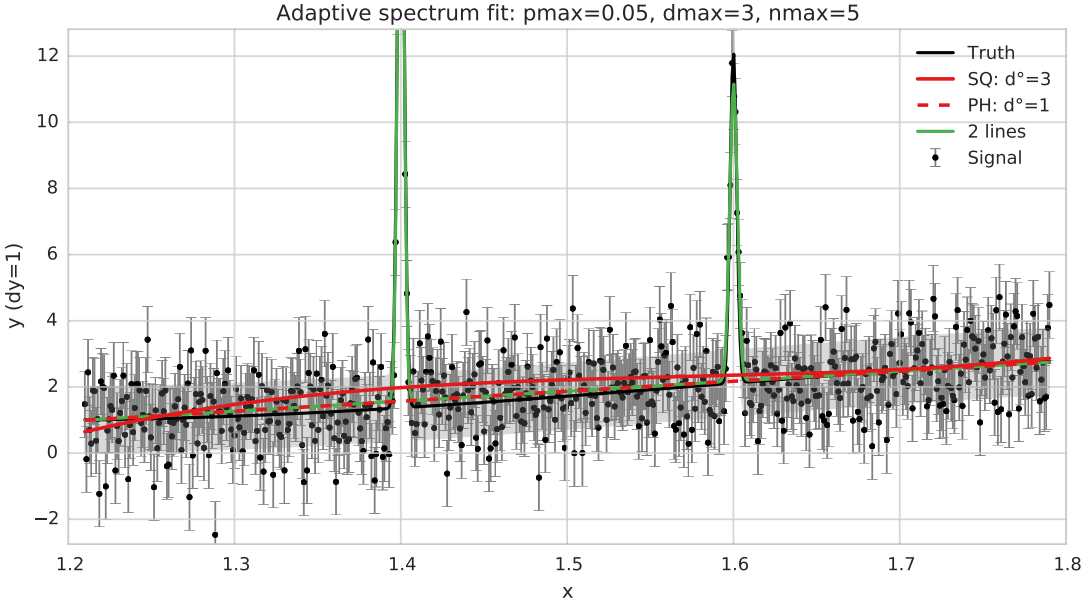

*Figure 1: Simple simulated spectrum mimicking a typical NISP merged spectrum. The intrinsic signal (*black line*) consists of a structured non-polynomial background and two Gaussian emission lines, to which was added a normal noise* $\mathcal{N}(\mu = 0, \sigma^2 = 1)$ *(*black points with error bars*). The result of the adaptive continuum fit (*red, $p_{\max} = 5\%$*) is shown for two loss functions: standard "squared" (*full line, $d = 3$*) and robust "pseudo-Huber" (*dashed line, $d = 1$*). The adaptive emission line fit detects* $n = 2$ *significant features (*green line, $p_{\max} = 5\%$*) on top of the pseudo-Huber continuum.*

Generic Algorithm 1 was implemented in the framework of `astropy.modeling`[5] to adaptively adjust the continuum of a (merged) spectrum by a polynomial, starting from the null polynomial $\boldsymbol{P} \equiv \boldsymbol{0}$ and up to maximal degree $d_{\max} = 3$; the SL is set to a not so conservative $p_{\max} = 5\%$ (see Fig. 1, as well as Fig. 3 and 4 for an illustration of the procedure).

6.2 Robust fit

Parameters minimising χ^2 such as in Eq. (2) are ML-estimates in the case of a *normal* noise distribution around intrinsic signal which the parametric model is supposed to fit. However, given the fact that the Gaussian distribution cannot account for large excursions outside its "core" domain (i.e. more than few standard deviations away from the mean), the presence of outliers can severely affects the reliability of the standard least-square parameters if not properly into account, either with an improved model (which could predict the outliers) or with a outlier-insensitive robust way to estimate the parameters.

When *least-square* adjusting a simple polynomial continuum to a spectrum, potential spectral features (e.g. emission lines) cannot be accounted for neither by the model nor by the underlying normality assumption, and may therefore significantly affect the result of the fit. An example of this effect is displayed in Fig. 1: the least-square continuum estimate (*solid red line*) is significantly "pulled" by the two emission lines and is no more a good representation of the underlying continuum.

To make the continuum adjustment less sensitive to the presence of spectral features, one can generalize Eq. (2) and use a robust M-estimator[6] based on a non-standard loss function ρ:

$$\hat{\boldsymbol{p}} = \operatorname{argmin} \sum_{i=1}^{n} \rho\left(\frac{y_i - M_i(\boldsymbol{p})}{\sigma_i}\right). \tag{3}$$

[5] http://astropy.readthedocs.org/en/latest/modeling/

[6] M-estimators are a broad class of estimators obtained as the minima of sums of functions of (weighted) residuals.

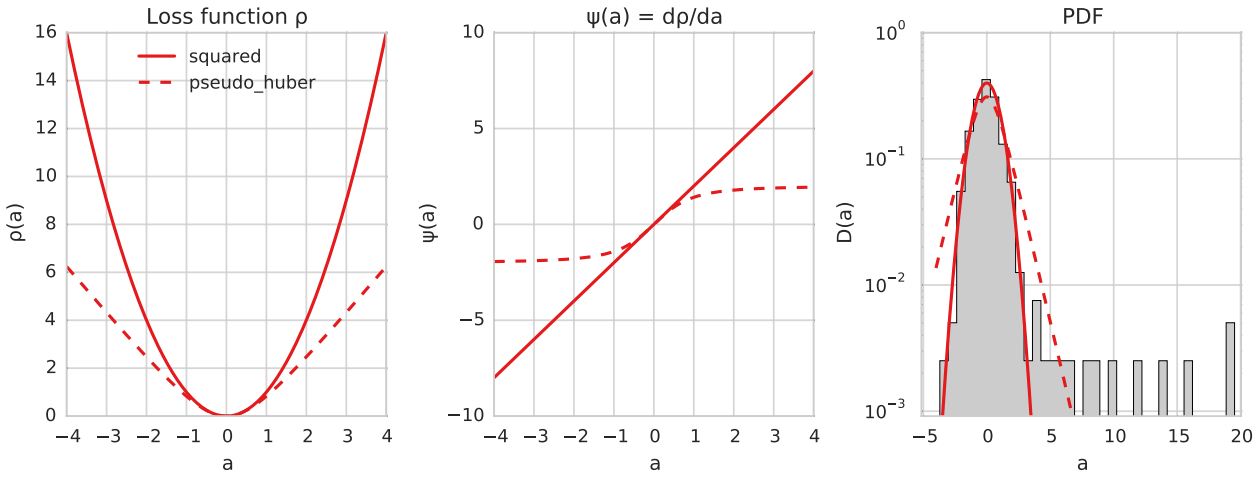

*Figure 2: Properties of two loss functions: standard "squared" (*red full line*) and robust "pseudo-Huber" (*red dashed line*).* Left panel: $\rho(a)$; central panel: $\Psi(a) = \mathrm{d}\rho/\mathrm{d}a$; right panel: *resulting PDF* $D(a) \propto e^{-\rho(a)/2}$. *This is to be compared to the heavy-tailed histogram of residuals with respect to best adaptive continuum of Fig. 1.*

Standard χ^2 uses the "squared" loss function $\rho_2 : a \mapsto a^2$, and as mentioned above, corresponds to the ML-estimate under a normality assumption. In the presence of outlying points, one can rather use a more "liberal" loss function which will reasonably account for them, e.g. the "pseudo-Huber" loss function[7] ρ_{pH} defined by:

$$\rho_{pH}(a) = 2\left(\sqrt{1+a^2} - 1\right). \tag{4}$$

Alternative loss function choices are possible – absolute errors, bi-weight, σ-clipping, etc. – but the Huber loss function is directly related to the *winsorising* technique traditionally used in robust statistics. Furthermore, the precise definition of the loss function ρ is not critical as long as it can handle outliers more appropriately than the standard squared one.

Fig. 2 displays a comparison of the two "squared" and "pseudo-Huber" loss functions. As can be seen on the *right panel*, the distribution of continuum-subtracted residuals is skewed by the presence of spectral features. Since the Huber-loss function corresponds to an underlying PDF $D(a) \propto e^{-\rho_{pH}(a)}$ with more extended tails than the normal distribution, this estimator can accommodate outlying residuals in a more robust way (see Fig. 1, *dashed red line*).

A robust fit procedure based on various loss functions, including the pseudo-Huber one, was implemented in the aforementioned `astropy.modeling` framework, and was systematically used to adaptively adjust the continuum of a spectrum by a polynomial.

7 Adaptive emission line detection

As stated in the purpose of this note, the objective of the current work is not to provide a precise emission line fit to the spectrum, including spectral covariance, Bayesian priors on joint line positions, intensities and widths, proper parameter error estimates, etc.; rather, the goal here is to provide such a fitter with pertinent initial guesses for potential emission features in the spectrum.

The same generic adaptive fit Algorithm 1 was adapted to adjust an increasing number of independent Gaussian emission lines on the continuum-subtracted (merged) spectrum, up to maximal number of $n_{\max} = 5$ (which should be enough to accommodate all strong emission lines falling simultaneously in the NISP spectral domain at any redshift). The SL is set to $p_{\max} = 5\%$ (see Fig. 1 and Fig. 3 to 5 for different applications). This liberal value should allow to detect all

[7] I choose here to work with the Huber-loss function in its "pseudo" C_∞ form rather than in its traditional C_1 form, but the practical difference is minimal.

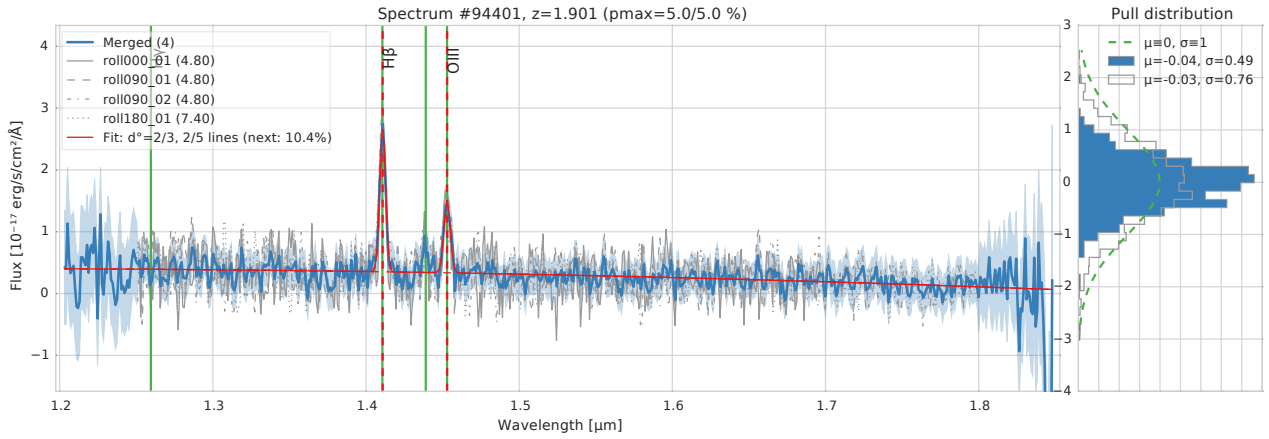

Figure 3: Example of robust adaptive line detection in spectrum #94401. Left panel: grey: *individual roll spectra;* blue: *final combined spectrum;* red: *adaptive fit (5% SL), consisting of a second-order ($d = 2$) polynomial background (adjusted with a pseudo-Huber loss function) and $n = 2$ Gaussian emission lines accounting for the $\mathrm{H}\beta$ and $[\mathrm{OIII}]\lambda5007$ lines (the $[\mathrm{OIII}]\lambda4959$ is only detected at the 10.4% SL);* green: *expected position at redshift $z = 1.901$ of major emission lines (note that this redshift is* never *used during the detection procedure).* Right panel: *pull distribution with respect to adaptive fit of individual roll spectra (*grey*) and combined spectrum (*blue*); the normal distribution $\mathcal{N}(\mu = 0, \sigma^2 = 1)$ is added for illustration (*green*).*

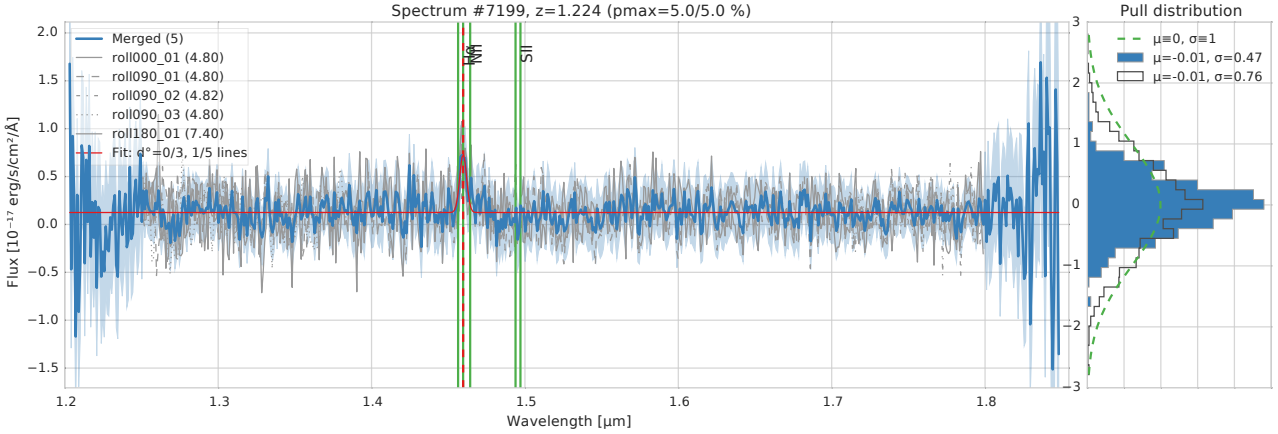

Figure 4: Same as Fig. 3 for spectrum #7199, where a single emission line is detected on top of a constant continuum ($d = 0$) at the 5% SL.

"significant" emission features in the spectrum, at the price of potential false positives. It is then left to the subsequent global adjustment procedure to decide the final significance of the detected peaks, and their probability to be genuine emission lines or just noise flukes.

7.1 Gauss-Hermite expansion

Comment:
TBC: case for Gauss-Hermite expansion, specially for the $\mathrm{H}\alpha/[\mathrm{NII}]$ complex in high SNR spectra (Fig. 6).

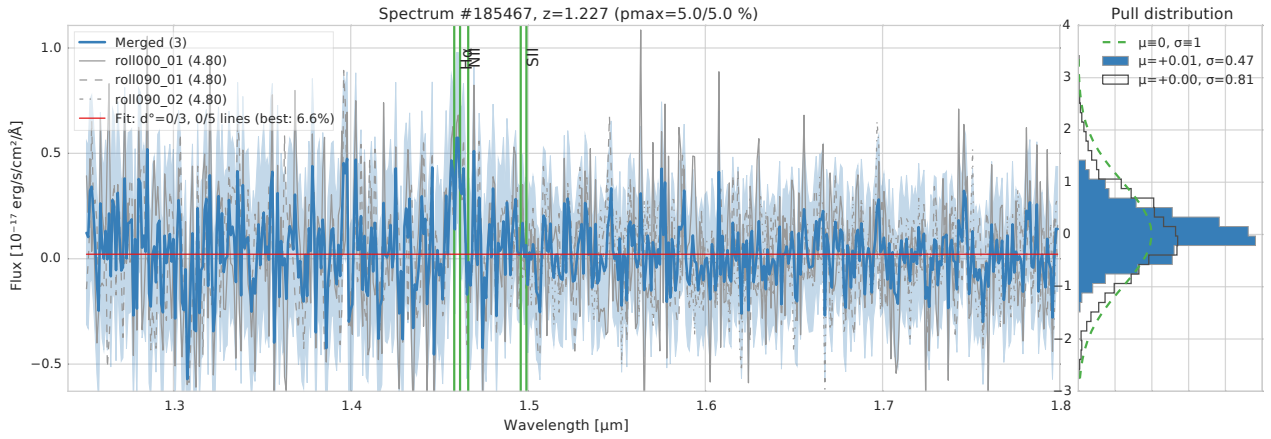

Figure 5: Same as Fig. 3 for spectrum #185467, where no significant emission line is detected at the 5% SL (highest significance feature is the $\mathrm{H}\alpha$ one at 6.6% SL).

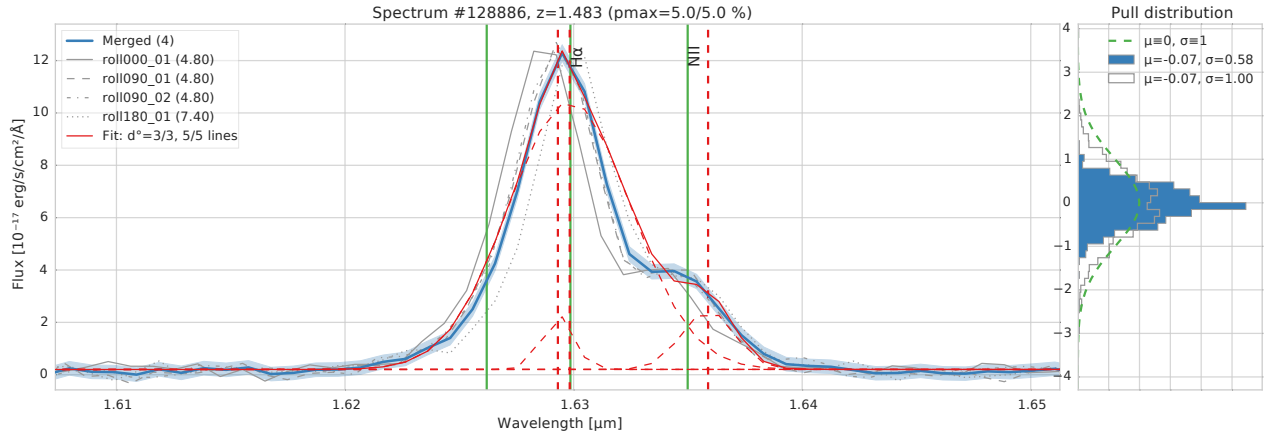

Figure 6: Same as Fig. 3 for high SNR spectrum #128886 with a particularly strong $\mathrm{H}\alpha/[\mathrm{NII}]$ blended complex distorting the line shape away from a simple Gaussian profile. As a consequence, the peak is adaptively adjusted with 3 independent normal lines (the two other lines detected are outside this zoom view), while a Gauss-Hermite expansion would be better suited.

8 Pull distribution

Comment:
TBC: The pull distribution of the final adjustment can be used to estimate *a posteriori* the validity of the (co)variance measurements.

Under the assumptions of a normally distributed noise, of a valid measurement variance estimate $\boldsymbol{\sigma^2}$ and of an "appropriate" model $\boldsymbol{M}$ (here polynomial continuum + emission lines), the distribution of the pull $\{(y_i - M_i)/\sigma_i\}$ should be $\mathcal{N}(\mu_p = 0, \sigma_p^2 = 1)$. Conversely, if this is not the case[8], a non-null mean pull $\mu_p \neq 0$ is indicative of a *biased* adjustment, while a non-unity pull variance $\sigma_p^2 > 1$ (resp. < 1) relates to an *under-* (resp. *over-*) estimation of the variance, which can then be corrected accordingly[9]. The $\sigma_p^2 > 1$ case can also point to the presence of a unaccounted-for

[8] The normality assumption is supposed to hold.
[9] Under the assumption of a null intrinsic dispersion in measurements $\{y_i\}$, due e.g. to the presence of outliers.

covariance between measurements, which can then be roughly estimated in a model-dependent way, e.g. as an isotropic exponential covariance function.

9 Tests

9.1 Individual spectra

9.2 Sample analysis