Title:	Impact of spectral covariance on line fitting		
Date:	09/12/2014	**Issue:**	Draft 0.2
Reference:	EUCL-IPN-TN-8-001		
Custodian:	Yannick Copin (`y.copin@ipnl.in2p3.fr`)		

Authors:		**Date:**	**Signature:**
	Yannick Copin *(IPNL)*	09/12/2014	
Contributors:			
Approved by:			
Authorized by:			

Document version tracking

Issue	Date	Page	Description of changes	Comments
Draft 0.1	25/04/2014	1	Initial import from personal note.	
Draft 0.2	09/12/2014	1	Typos, font management.	

Table of contents

1 Purpose

I look for the impact of the spectral covariance on the line fitting procedure. It appears that the maximum-likelihood estimates — for all line parameters — are equally *un*biased when using the correct full-covariance χ^2 definition and the simpler pure-diagonal one (i.e. neglecting spectral correlations). However, best-fit parameter uncertainties — on line flux, position/*redshift* and width — are systematically *under*-estimated by ~ 40% when using the uncorrelated χ^2, while properly estimated when minimizing the statistically correct χ^2. The use of spectral covariance is therefore of crucial importance to derive statistically controlled spectral quantities such as redshift and line fluxes.

2 Scope

Spectral measurements in OU-SPE.

3 Applicable & Reference documents

3.1 Applicable documents

RD		Ref.	Date

3.2 Reference documents

RD		Ref.	Date

4 Acronyms

MAD	Median Absolute Deviation
ML	Maximum Likelihood
SNR	Signal-to-Noise Ratio

5 Data simulation

5.1 Intrinsic signal

The *true* simulated spectrum $\boldsymbol{S}(a,\mu,\sigma) = (S_1,\ldots,S_N)$ ($N = 32$ in this analysis) is a single Gaussian emission line, characterized by its peak amplitude $a > 0$, its mean position μ and dispersion σ (in pixel units), on a constant null continuum:

$$S_i(a,\mu,\sigma) = a\,\exp\left(-\frac{(i-\mu)^2}{2\sigma^2}\right). \tag{1}$$

5.2 Intrinsic (co)variance

The signal is simulated in the regime of *constant* normal noise[1], and the flux units are chosen such that:

$$\sigma_i = 1, \quad i = 1,\ldots,N. \tag{2}$$

The amplitude a of the emission line is therefore directly representative of its (peak) signal-to-noise ratio (SNR).

To account for *short-scale* correlations between adjacent pixels in the simulated spectrum, the intrinsic covariance matrix Σ is chosen to follow an isotropic exponential covariance function, of scale-length $\tau \geq 0$:

$$\Sigma_{ij}(\tau) = \begin{cases} \sigma_i\sigma_j\,\exp\left(-\frac{|i-j|}{\tau}\right) & \text{if} \quad \tau > 0, \\ \sigma_i\sigma_j\,\delta_{ij} & \text{if} \quad \tau = 0. \end{cases} \tag{3}$$

The limit case $\tau = 0$ corresponds to a purely diagonal covariance matrix, i.e. to the absence of correlations.

5.3 Simulated signal

A signal simulation $\boldsymbol{y}$ is the sum of the intrinsic signal $\boldsymbol{S}(a,\mu,\sigma)$ and a realization of the noise $\boldsymbol{\epsilon}(\tau)$ with the desired correlation:

$$y_i = S_i(a,\mu,\sigma) + \epsilon_i(\tau). \tag{4}$$

Intrinsic signals will be generated using the following input parameters:

- $a = 2$ (low SNR regime), 5, 10 and 20 (high SNR regime);
- Input μ is a random variable uniformly distributed in ± 1 px, to avoid any systematic sampling effect. The quoted line position is actually the offset $\delta\mu = \hat{\mu} - \mu$, where $\hat{\mu}$ is the adjusted position;
- $\sigma = 1$ (barely sampled line), 2 and 3 px (over-sampled line).

For this analysis, I initially generate $L = 1000$ *uncorrelated* noise realizations $\boldsymbol{n} = (n_1,\ldots,n_L)$ from normal distribution $\mathcal{N}(\mu = 0, \sigma^2 = 1)$. These uncorrelated noise realizations are then spectrally correlated using the targeted covariance matrix $\Sigma(\tau)$ (Eq. (3)) to produce to noise realizations $\boldsymbol{\epsilon}(\tau)$ with the desired correlation length τ:

- $\tau = 0$ (no correlation), 2 px and 5 px (strong spectral correlation[2]).

This procedure ensures that all simulated signals share the same noise realizations up to the spectral correlation (see Fig. 1): one can then directly estimate the impact of the covariance in the adjustment of the simulated spectra.

[1]This is valid if the spectral background flux per pixel — e.g. from zodiacal light — is high enough to ensure noise normality.

[2]The case $\tau = 2$ px corresponds to a correlation coefficient of $\rho = e^{-1/2} = 60\%$ between adjacent pixels, while $\tau = 5$ px gives $\rho = 82\%$.

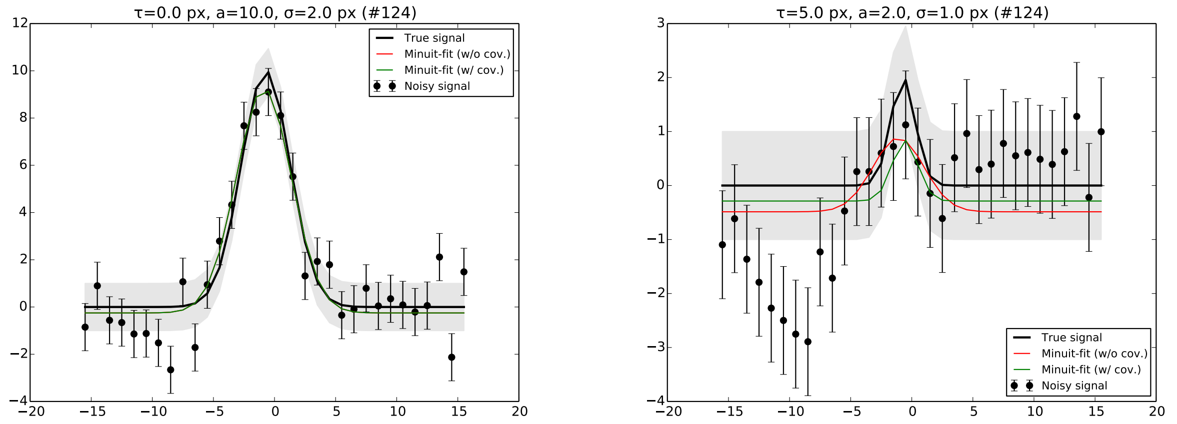

Figure 1: Example of simulated signals. *Left:* $a = 10$, $\mu = -0.71$ px and $\sigma = 2$ px, no inter-pixel correlation ($\tau = 0$); *right:* $a = 2$, $\mu = -0.71$ px and $\sigma = 1$ px, with a correlation length of $\tau = 5$ px. The two simulated signals share the same uncorrelated noise realization. The true signal and uncertainty $\boldsymbol{S} \pm \boldsymbol{\sigma}$ is represented by the *heavy line* $\pm$ *shaded area*; the actual simulated spectrum $\boldsymbol{y} \pm \boldsymbol{\sigma}$ is represented by *black symbols* $\pm$ *error bars*. The results of the adjustments using full-covariance χ^2_{Cov} (Eq. (7)) or pure-diagonal χ^2_{σ} (Eq. (9)) are displayed by the *green* and *red* lines respectively (they are naturally combined in the uncorrelated case).

6 χ^2 minimization

Since noise realizations are purely Gaussian, the maximum-likelihood (ML) parameters $\hat{\boldsymbol{\theta}}$ can be estimated from minimization of the χ^2 objective function comparing the observed signal $\boldsymbol{y}$ to the model $\boldsymbol{F}(\boldsymbol{\theta})$:

$$\hat{\boldsymbol{\theta}} = \operatorname{argmin} \chi^2(\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{\theta})). \tag{5}$$

The minimization will be performed using the `migrad` minimizer from the Minuit library, through the pyminuit interface.

6.1 Model

The adjusted model $\boldsymbol{F}(f, \mu, \sigma, b)$ is a Gaussian profile on a constant background b:

$$F_i(f, \mu, \sigma, b) = \frac{f}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(i-\mu)^2}{2\sigma^2}\right) + b. \tag{6}$$

I choose here to adjust for the physically-motivated *integrated* flux f of the line, not its peak amplitude $a = f/\sqrt{2\pi}\sigma$. Note furthermore that the model does not account for pixel-integration, and therefore is expected to poorly perform on under-sampled lines ($\sigma \lesssim 1$).

6.2 Objective functions

The objective of the analysis is to estimate the impact of the covariance use in the adjustment of spectrally correlated spectra. I therefore compare the parameters estimated by minimizing two different χ^2 objective functions:

Full-covariance: the proper χ^2 definition is presence of a covariance matrix V between the measurements $\boldsymbol{y}$ (i.e. $V_{ij} = \mathrm{Cov}(y_i, y_j)$) is:

$$\chi^2_{\mathrm{Cov}}(\boldsymbol{\theta}) = (\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{\theta}))^T \cdot \mathsf{V}^{-1} \cdot (\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{\theta})). \tag{7}$$

In the present analysis, the covariance matrix V of simulated observations $\boldsymbol{y}$ is — rather optimistically — set equal to the intrinsic covariance Σ:

$$\mathsf{V} = \Sigma. \tag{8}$$

Pure-diagonal: when neglecting the off-diagonal terms of the covariance, the previous expression only depends on the diagonal terms, i.e. variances σ_i^2:

$$\chi^2_\sigma(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{(y_i - F_i(\boldsymbol{\theta}))^2}{\sigma_i^2}. \tag{9}$$

Both objective functions supposedly follow a χ^2-distribution with $k = N - M$ degrees of freedom, where $M = 4$ is the number of adjusted parameters.

7 Results

For each $\boldsymbol{\theta} = (a, \mu, \sigma)$ input parameter set and noise correlation length τ, a dataset of $L = 1000$ simulated spectra is generated (all sharing the same uncorrelated noise realizations), and ML parameters $\hat{\boldsymbol{\theta}}$ are estimated independently from minimization of:

1. the full-covariance χ^2_{Cov} (Eq. (7)),
2. the pure-diagonal χ^2_σ (Eq. (9)).

I use the "pull" distribution, defined as:

$$p_j = \frac{\hat{\alpha}_j - \alpha}{\sigma_{\hat{\alpha}_j}}, \quad j = 1, \dots L, \tag{10}$$

where $\hat{\alpha}_j$ (resp. $\sigma_{\hat{\alpha}_j}$) is the ML estimate (resp. the estimated uncertainty) of "true" parameter α. This pull distribution has the following interesting properties:

- its mean value $\mu_p = 0$ if the parameter estimator is *unbiased*;
- its standard error $\sigma_p = 1$ if the parameter uncertainty estimate $\sigma_{\hat{\alpha}}$ is correct[3]: $\sigma_p > 1$ (resp. < 1) means that the parameter uncertainty $\sigma_{\hat{\alpha}}$ has been *under*-estimated (resp. *over*-estimated) by a factor $1/\sigma_p$;
- the pull $\chi^2_p = \sum_{j=1}^{L} p_j$ follows a χ^2 distribution with L degrees of freedom, and a goodness-of-fit (actually a parameter estimate reliability) can be estimated from its associated one-tail p-value[4].

Statistics of the adjusted parameters $\hat{\boldsymbol{\theta}}$ for the test-case $a = 10$, $\sigma = 2$ px (resulting in an input flux of $f = a\sqrt{2\pi\sigma} = 50.13$) in the moderately correlated case (correlation length of $\tau = 2$ px) are presented in Table 4:

- mean and standard deviation of parameter estimates,
- median and normalized median absolute deviation (nMAD) of parameter estimates,
- pull χ^2_p — which follows a χ^2 distribution with $L = 1000$ degrees of freedom — and associated p-value,
- mean and standard deviation of pull distribution.

As can be seen from Table 4:

[3] In presence of a bias, $\sigma_p^2 = 1 + \mu_p^2$.
[4] This is the probability for the χ^2 to reach such a high value assuming the description of the distribution is correct.

Table 4: Results for the moderately correlated case ($\tau = 2$ px), with $a = 10$ and $\sigma = 2$ px (i.e. $f = 50.13$). The pull χ^2_p has $L = 1000$ degrees of freedom, its associated p-value is quoted in percents.

	Parameter distribution				Pull distribution			
Parameter	μ	σ	Med	nMAD	χ^2_p	p [%]	μ_p	σ_p
Pure-diagonal χ^2_σ								
f	+50.923	7.236	+50.449	7.507	3117	0	+0.070	1.765
$\delta\mu$	+0.001	0.218	+0.002	0.208	2003	0	+0.006	1.416
σ	+2.015	0.227	+1.996	0.214	1706	0	$-$0.051	1.306
b	$-$0.034	0.405	$-$0.044	0.401	3459	0	$-$0.133	1.856
Full-covariance $\chi^2_{\rm Cov}$								
f	+50.611	6.837	+50.134	7.015	1083	3	+0.010	1.041
$\delta\mu$	$-$0.001	0.197	+0.000	0.195	1008	42	$-$0.001	1.005
σ	+1.997	0.198	+1.984	0.193	1119	1	$-$0.133	1.050
b	$-$0.024	0.396	$-$0.033	0.397	1036	21	$-$0.054	1.017

– $\mu_p \simeq 0$: none of the ML estimates is strongly biased, using the exact full-covariance $\chi^2_{\rm Cov}$ definition or the simpler pure-diagonal χ^2_σ one;
– $\sigma_{p,\rm Cov} \simeq 1$: the uncertainties of the ML estimates are correct when using the full-covariance $\chi^2_{\rm Cov}$ definition;
– $\sigma_{p,\sigma} > 1$: the uncertainties of the ML estimates are systematically *under*-estimated when using the pure-diagonal χ^2_σ definition: the line flux error $\sigma_{\hat{f}}$ is under-estimated by 43%, the position error $\sigma_{\hat{\delta\mu}}$ by 29%, and the line width error $\sigma_{\hat{\sigma}}$ by 23%.

The ML estimate distributions as well as the associated pull distributions are presented in Fig. 2 for the moderately correlated case ($\tau = 2$ px). Once again, it becomes apparent that while the ML parameter estimate are barely sensitive to the use or not of the full-covariance matrix, the uncertainties on ML estimates are systematically *under*-estimated by up to ~ 40% when using pure-diagonal χ^2_σ.

Fig. 3 shows the evolution of pull mean μ_p and standard deviation σ_p for different parameters as function of correlation length τ for the test-case $a = 10$, $\sigma = 2$ px. The error on the ML estimate uncertainty presumably increases steadily with τ for background level (b) and line flux (f, the two parameters being strongly correlated). On the other hand, the error on the ML estimate uncertainty of line position μ and width σ probably peaks when $\tau \sim \sigma$. This evolution with τ has to be confirmed by more exhaustive simulations.

8 Conclusions

I generated intrinsic Gaussian emission line spectra with different reasonable input parameters (peak amplitude a, mean position $\mu \simeq 0$, line width σ), and added noise realizations under the assumption of constant normal noise and varying correlation length τ.

Maximum-likelihood parameter estimates were obtained by minimizing two versions of the χ^2:
– the full-covariance $\chi^2_{\rm Cov}$, taking full account of the (supposedly known) covariance matrix;
– the pure-diagonal χ^2_σ, neglecting all off-diagonal terms (i.e. correlations) of the covariance matrix.

It appears from the pull distribution analyzes that:

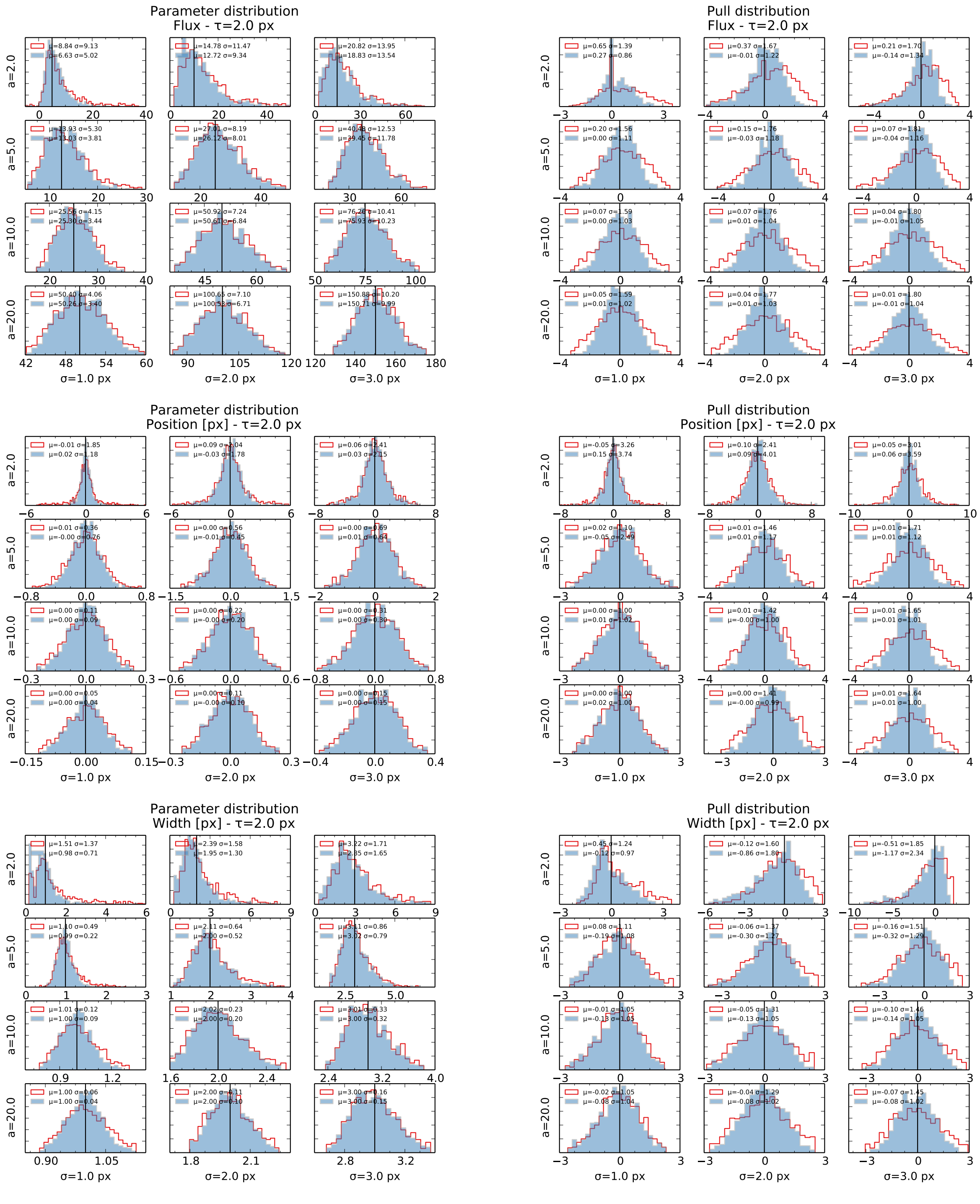

Figure 2: Results for the moderately correlated case ($\tau = 2$ px). *Left column:* parameter distribution (*from top to bottom:* flux f, position offset $\delta\mu$ and line width σ), when using full-covariance $\chi^2_{\rm Cov}$ (Eq. (7), *shaded blue*) in the line fit, or pure-diagonal χ^2_σ (Eq. (9), *red line*). *Right:* pull distributions.

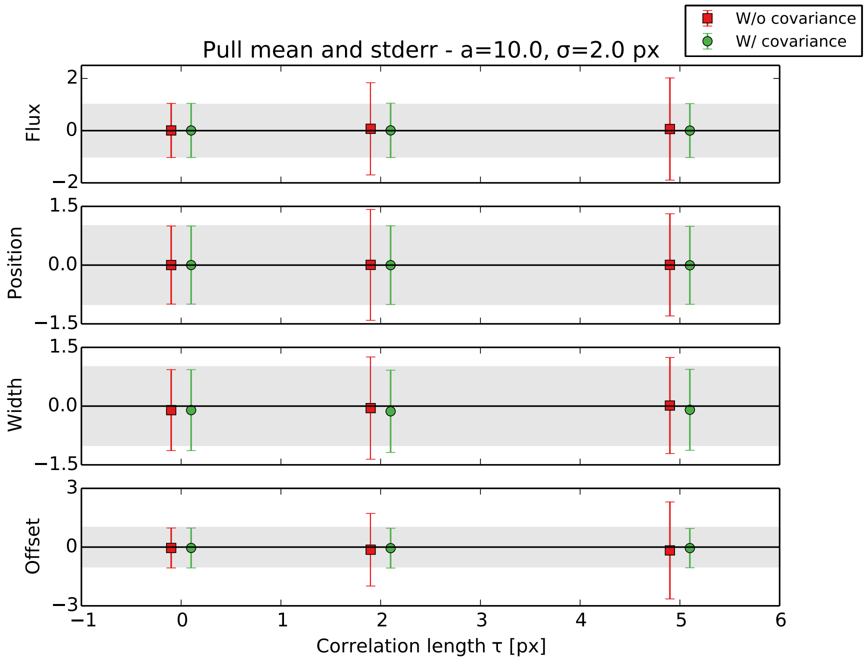

Figure 3: Evolution of pull mean and standard error as a function of correlation length τ for test-case $a = 10$ and $\sigma = 2$ px (i.e. $f = 50.13$), when using pure-diagonal χ^2_σ (Eq. (9), *red symbols*) or full-covariance χ^2_{Cov} (Eq. (7), *green symbols*) in the line fit. *From top to bottom:* flux f, position offset $\delta\mu$, line width σ and background level b. The *gray shaded area* corresponds to the ideal pull range 0 ± 1.

– the ML estimates — for all line parameters — are equally *un*biased when using the correct χ^2_{Cov} definition and the simpler χ^2_σ one;
– the ML estimate uncertainties — on line flux, position/*redshift* and width — are systematically *under*-estimated by up to 40% when using the simpler χ^2_σ, while they are correct when minimizing χ^2_{Cov};
– (to be confirmed) when using χ^2_σ, the error on flux uncertainty is increasing with correlation length τ, while error on position/redshift and line width peak at $\tau \simeq \sigma$.

The use of the full-covariance χ^2_{Cov} is therefore of crucial importance to derive statistically controlled spectral quantities such as redshift and line fluxes. This requires the precise knowledge of the spectral covariance properties of the fully-calibrated spectra, either from a proper uncertainty propagation among the successive calibration steps, or from *a posteriori* estimates on observed signals.