EFFECTIVE RESOLUTON CONCEPTS FOR LIDAR OBSERVATIONS

Marco Iarlori, Fabio Madonna, Vincenzo Rizi, Thomas Trickl and Aldo Amodeo

Paper ID: amt-2015-89

Iteration: Initial Manuscript Evaluation (Major comments included + partial Minor comments this at this review stage)

REMARKS TO THE AUTHORS

OVERALL SUMMARY

Dear authors,

The manuscript focus is two-fold: On one hand, it tackles application of low-distortion filters (particularly, the Savitzky-Golay (SG), the Gaussian, and basic cascaded and/or windowed realizations of them) to the problem of smoothing and derivation of lidar signals. On the other hand, and subsidiary to this first objective, an empirical parametric study on the spatial resolution achievable with these filter realizations is presented.

The main added value of the manuscript lies on the nice transferability, application and extension of well-known concepts carried out by the authors from the signal-processing arena (e.g., the SG dates back to 1960's) to the physical/lidar one (greenish in the sig-pro field). The authors have demonstrated hard-work in the simulation/studies presented and hence, the manuscript represents a good contribution to AMT. Consistent with this, I must say that the originality of this manuscript is somehow limited to applied-remote sensing journals of the like, as is the case of AMT.

In its present form the WRITTEN part of the manuscript is too draft. It is my feeling that -for some unknown reason- the authors are submitting a manuscript still requiring several levels of edition concerning e.g., use of English and a better structuration of paragraphs/ideas. Definitely, the manuscript is TOO LONG. The authors must do an effort towards more succinct redaction, some parts are discursive). NOTATION is un-rigorous or provisional all over the manuscript and must be cross-examined against modern standard signal processing conventions (see attached SG paper). Technical major comments are given next.

All considered, I recommend major revision along the lines suggested and an in-depth writing revision.

Level of revision:

With the present level of editing exhaustive DETAILED COMMENTS are NOT possible. The authors will find help tips from this reviewer in file annotated_manuscript.pdf.

Circled line-numbers indicate conflictive, incorrect or vaguely explained concepts that require further revision from the authors' side. Some of them are covered in the detailed comments below.

ATTACHMENTS

(1) annotated_manuscript.pdf

An annotated manuscript is attached with in-depth spelling/use-of-English correction for pp.1-5 just to help the authors on this issue.

MAJOR COMMENTS

(1) The first part of the manuscript (whole Sect. 2) and, more intensively until Eq. (7), is just a review – in summary “book” fashion – of well-known concepts on digital signal processing, which conveys no original material. Suggestions are given next to improve it in terms of notation, terminology and clearer structuration of ideas.

<u>Sect. 2/introduction.</u> Please consider slight reorganization Sect. 2, pp. 4-7 along this layout:

1. Response of linear-time-invariant (LTI) systems to *arbitrary* inputs: convolution sum,

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k)$$, Eq. (R1).

2. Causality,

$$y(n) = \sum_{k=0}^{\infty} h(k)x(n-k)$$, Eq. (R2).

3. Finite Impulse Response (FIR),

$$y(n) = \sum_{k=0}^{N} h(k)x(n-k)$$, Eq. (R3).

4. Highlight the importance of <u>linear-phase</u> FIR filters, which is what you use all over the paper. For linear-phase filters,

$$h(n) = \pm h(M-1-n),\ n = 0,1,\ldots,M-1$$, Eq. (R4).

To see 4, resort to important properties of the Fourier transform (FT), Eq. (R5a,b,c):

- A delay in time, x(t-t0) (equivalently, x(n-n0)) appears as a linear phase in the frequency domain, X(omega)e^-j*omega*t0.

$$x(t-t_0) \rightarrow X(\Omega)e^{-j\omega t_0}$$ (analog formulation), Eq. (R5a)

$$x(n-n_0) \rightarrow X(\omega)e^{-j\omega n_0}$$ (discrete formulation), Eq. (R5b)

- Real signals

$$x(n)\ real \quad \rightarrow \quad X(\omega) = X^{*}(-\omega)$$, Eq. (R5c).

5. Fourier Transform (FT) of a discrete-time aperiodic signal (your Eq.(2)).

Introduce the concept of “frequency for discrete-time signals” as

$$\omega = \frac{\Omega\left[\frac{rad}{s}\right]}{\Omega_s\left[\frac{rad}{s}\right]},$$ Eq. (R6)

where Ω is the analog frequency (continuous-time signal in [rad/s], i.e., prior to sampling) and Ω_s is the sampling frequency (in [rad/s]). Or

f (lower case) =F/F_sub_s, $f = \frac{F}{F_s}$, Eq. (R7)

Note than the highest frequency of oscillation in a discrete-time sinusoid is attained when omega=+-pi (i.e., f=+-1/2). Say that for convenience (and for practical compliance with modern signal processing tools like Matlab) the normalised frequency, $f' = \omega/\pi$, is used all over the paper (“nu” is not so common).

All frequency equations such as Eq.(5) in p.8 of the manuscript must always be formulated using ω (irrespective of the fact that your plotting variable is ω, f (bounded to +-1/2 or 0-to-0.5), or f_prime (bounded to +-1, or 0-to-1). Thus, the ideal digital differentiator is always $j\omega$.

> 6. Fig. 1 Why the negative part of the spectrum is not represented? Explain that only positive frequencies need be represented, because if x(n) is REAL - as is the case here- then X(omega) has Hermitian symmetry, $X^*(\omega) = X(-\omega)$.

(2) FIGURES. Please consider plotting $20\log_{10}|H(\omega)|$ all over most of (not all) the spectral figures of the paper instead of plain $H(\omega)$. The standard representation in signal processing is $20\log_{10}|H(\omega)|$ and $\arg[H(\omega)]$ [deg]. This will help identifying sidelobes in the STOPBAND.

The correct notation for decibel is [dB] (not db).

(3) NOTATION (not exhaustive)

"n" for sample index as x(n), not as subindex

f or $f' = \omega/\pi$ as normalised frequency

$j\omega$ as the ideal differentiator,

j as the imaginary unit

Do always write "n" (time domain) or "omega" (frequency domain) in your Eqs (revise Eq. (9)).

Suggestions: ΔR_{eff}, $\Delta R_{eff}^{SG2,4}\Big|_{Ray}$, $\Delta R_{eff}^{SG2,4}\Big|_{NRR}$, H_{-3dB} (usual notation)

R.W. Schafer, "What is a SG filter?, IEEE Trans. Signal Proc. Mag., 28, 111-117 (2011). Please consider this notation ("n" for sample index as x(n), not as subindex), frequency response in log magnitude (20log10(modulus(H), Figs. 3, 5 therein).

(2) Equivalent SG filter. The author propose cascaded SG filter combinations of the type SG2(L1) & SG4(L2). How does a single filter, for example SG2(L1+L2), compare with the cascaded solution? More generally, the authors should discuss on an equivalent SGx(y), i.e., a suitable pair of order "x" and length "y" based on nice parametric studies carried out. This is better oriented to the focus of the paper.

(3) How lidar signals are processed under a variable spatial-resolution approach? Using the same SG filter over the full range of the lidar sensor (or other remote-sensing sensor) means processing the lidar signal with the same spatial resolution (i.e., constant all over the inversion range). However, when faced with data records several km long, the rhythm of decrease of SNR forces to use much longer spatial resolutions at further ranges, where the SNR is getting lower and lower. It is difficult to believe that the same SG filter can solve the problem for the full inversion range (say 0-8 km). I advise the authors to expand/discuss this part in detriment of others.

(4) Conclusion. Just a preliminar review: Important is that –from what you have in hands- the scope of the paper at least covers: 1) an in-depth parametric study on SG and Gaussian filters and 2) Effective resolutions studies/methodologies. Both are equally important and nice to contribute, not only the second! The paper title could also be better adapter to reflect this.

DETAILED COMMENTS

p.4, lin. 8. Clarify that according to SG (1964) paper there are two big families of SG filters: “smoothers” and “n-th order differentiators” and that derivation always implies a certain level of smoothing because of the inherent construction of the SG filter.

p.4, lin. 13. Notation. Unclear indexes.

In the convolution equation, it is standard to say that the impulse response has “M” samples, the signal (x(n)), N samples, and the output (y(n)) N+M-1 samples. Please consider to include a figure sketch to help the reader (otherwise lins 21-23 becomes cumbersome…).

p.5, lin. 12. Remove “aliasing”.

“Sampling of a continuous-time aperiodic signal (e.g., the lidar signal) causes the spectrum of the discrete-time signal (i.e., $H(\omega)$) to be a continuous and periodic function of variable omega (with period 2*pi (-pi<=omega<=pi)).”

“Aliasing” usually refers to the unwanted effect of multiple spectrum folding on the frequency axis due to a too low sampling rate.

p.5, lin. 15. Normalised frequency. Please see major comments.

p.6, lins 3-5. A negative value of H… results in artifacts.

This assertion is FALSE. What causes “artifacts” of unwanted effects is the frequency content in the transient+stop-band of the spectrum, be it positive, negative, or with a given phase (in the case of a complex-valued spectrum, not your case). This part of the spectrum conveys “unwanted” leakage frequencies.

The fact that an unwanted frequency has $\arg[H(\omega_1)]$ = 0 or 180 deg (i.e., H(omega_sub_1)=+1 or -1, respectively) means that this high-frequency will show up in the grey zone of Fig. 2 without or with a sign reversal in the oscillatory behaviour of the Chirp signal.

KEY is to have low sidelobes and no discontinuities in the time domain.

p.6, lin. 8, Eq. (3). Revise/wrong. See Eq. (R8b) next.

A linear chirp is defined as $f(t) = f_0 + kt$, where f_sub_0 is the start frequency and k is the chirp rate parameter. Therefore, the chrip signal becomes,

$$x(t) = \cos\left[2\pi\left(f_0 t + \frac{k}{2}t^2\right) + \phi\right]$$, Eq. (R8a).

Sampling at $t = nT = \frac{n}{f_s}$, with T the sampling period (equivalently, $f_s = \frac{1}{T}$, the sampling frequency) the discrete signal takes the form,

$$x(n) = \cos\left[2\pi\left(\frac{f_0}{f_s}\right)n + \pi k\left(\frac{n}{f_s}\right)^2 + \phi\right]$$, Eq. (R8b).

p.8, lin. 4, Eq. (5). See annotated ms.

p.8, lins. 8 and 17. Clearly expose that the goal is to implement a low-pass-limited differentiator with the “j*omega” slope reset to zero between ω_c and $\omega = \pi$, ω_c the low-pass cut-off frequency.

p.8, lins 8-12. Slang, magazine oriented (in my opinion).

p.9. Eq. (7). Revise/wrong.

Assuming anti-symmetric unit-sample response for the desired filter, $h_d(n) = -h_d(-n), h_d(0) = 0$, gives

$H_d(\omega) = -2j\sum_{n=1}^{N} h_d(n)\sin(\omega n)$, Eq. (R9).

p.9, lins 18-20. Somehow slang and difficult to follow.

Suggestion: The SG-derivative filter can be understood as a cascaded system of two filters, the first one being a low-pass filter (LPF) acting on the raw input signal and the second one being the ideal differentiator ($H(\omega) = j\omega$) and acting on the LPF signal. Prefix “d” stands for the low-pass filter prototype associated to this first stage of the SG differentiator (inherently, low pass).

p.13, lin. 3 This is just a valid approximation. In fact there are hundred of alternative low-pass filter design approaches.

p.13, lins. 13-14. How does a SG2(L1+L2) compares with the cascaded SG2(L1)-SG4(L2)? What about SGx(y), i.e., a suitable pair of order “x” and length “y”?

p. 14, lins 13-25. This part can be moved to a discussion sub-section in Sect. 3 or simply removed. Your paper is too long.

p. 15. Unclear writing/structuration for Sect. 2.2.2.

Please distinguish between time windowing and spectral windowing.

Windowing in the time domain means multiplying x(n) by a window w(n) as indicated by Eq. (11) + handwritten block diagram. According to FT properties, multiplication in the time domain equals convolution in the frequency domain. However in Fig. 7 frequency response it is not clear if the SG-Blackman window is the result of multiplication in the time domain (correct) or the result of multiplication of a Blackman window in the frequency domain by the SG spectral response (incorrect). Please clarify.

Most readers will not understand what is “leakage”: Better introduce the problem by saying that “truncating” x(n) in the time domain is equivalent to multiplying by a rectangular window, w(n), which causes sinc=sin(x)/x (the FT of the rectangular window) to convolve in the frequency domain, i.e., many unwanted secondary lobes in the frequency response!

p.15, Eq. (12) Revise/wrong.

$w(n) = 0.42 - 0.5\cos\frac{2\pi n}{M-1} + 0.08\cos\frac{4\pi n}{M-1}$, $w(n), \quad 0 \le n \le M-1$, Eq. (R10).

Rest of detailed comments (non exhaustive): Please consider annotated manuscript.

Good luck!

NOTATION

Effective resolution concepts for lidar observations.

Marco Iarlori^1, Fabio Madonna^2, Vincenzo Rizi^1, Thomas Trickl^3 and Aldo Amodeo^2

^{1}CETEMPS/DSFC, Università Degli Studi Dell'Aquila, Via Vetoio, 67010 Coppito, L'Aquila, Italy.

^{2}Consiglio Nazionale delle Ricerche – Istituto di Metodologie per l'Analisi Ambientale CNR-IMAA, Potenza, Italy.

^{3}Karlsruhe Institute of Technology (KIT), IMK-IFU, Garmisch-Partenkirchen, Germany.

Correspondence to: Marco Iarlori (marco.iarlori@aquila.infn.it)

Abstract

Since its first establishment in 2000, EARLINET (European Aerosol Research Lidar NETwork) has been devoted to providing, through its database, exclusively quantitative aerosol properties, such as aerosol backscatter and aerosol extinction coefficients, the latter only for stations able to retrieve it independently (from Raman or High Spectral Resolution Lidars). As these coefficients are provided in terms of vertical profiles, EARLINET database must also include the details on the range resolution of the submitted data. In fact, the algorithms used in the lidar data analysis often alter the spectral content of the data, mainly working as low pass filters with the purpose of noise damping. Low-pass filters are mathematically described by the Digital Signal Processing (DSP) theory as a convolution sum. [As a consequence, this implies that each filter's output, at a given range (or time) in our case, will be the result of a linear combination of several lidar input data relative to different ranges (times) before and after the given range (time): a first hint of loss of resolution of the output signal.] The application of filtering processes will also always distort the underlying true profile whose relevant features, like aerosol layers, will then be affected both in magnitude and in spatial

extension. Thus, both the removal of noise and the spatial distortion of the true profile produce a reduction of the range resolution.

This paper provides the determination of the effective resolution (ERes) of the vertical profiles of aerosol properties retrieved starting from lidar data. Large attention has been addressed to provide an assessment of the impact of low-pass filtering on the effective range resolution in the retrieval procedure.

$$\left[\; X(w) \xrightarrow{\frac{d}{dt}} \underbrace{jwX(w)}_{\text{high pass}} \;\right]$$

FALSE

1 Introduction

Smoothing and numerical derivative are typically used in the retrieval of aerosol optical properties from lidar data ~~and both may act as low pass filter.~~ ~~Indeed, the~~ smoothing is a low pass ~~filter,~~ while the numerical derivative ~~has a low pass filter inherently associated (see Sect. 2.1)~~. For this reason, in what follows, the terms "smoothing filter" and "low pass filter" should be considered as synonymous. In particular, the smoothing is one of the operations most frequently carried out and it can be applied on the raw lidar signals as well as on final products, like the aerosol backscatter coefficient (β_a) or the aerosol extinction coefficient (α_a) (Klett, 1981; Fernald, 1984; Ansmann et al., 1992) to reduce the random noise. On the other hand, to retrieve the aerosol extinction coefficient from a Raman signal (Ansmann et al., 1992), the PBL height estimation from a Rayleigh signal (Matthias et al., 2004), ozone profiles and water vapor profiles with the DIfferential-Absorption Lidar (DIAL) technique (Wulfmeyer and Bösenberg, 1998; McGee et al., 1995), a numerical derivative is typically included in the retrieval algorithm. The application of low-pass filtering will also generate a reduction in the vertical or time resolution with respect to the unfiltered products. Moreover, there is frequently the need of comparing or combining different atmospheric variables and this requires that they are fully consistent in time and in space, which means that they must be nearly co-located, time simultaneous and with the same resolution. This latter category includes, for example, ~~the~~ retrieval of lidar ratio (S) profile or ~~the~~ comparison between the same quantity obtained by different instruments with different resolutions, like balloon-borne ozone data versus ozone lidar profiles, as pointed out by previous studies (e.g., Masci, 1999). In those cases, inconsistencies could arise if data are not compared with the same resolution. For example, to obtain a S profile, a~~n~~ high-resolution aerosol-backscatter coefficient profile ~~showing~~ having well-resolved layers, could be combined with an heavily-smoothed, low-resolution simultaneous extinction profile, where the same layers are not well resolved. This would result in a biased estimation of the actual values of the lidar ratio (see Sect 3 and Sect 3.1).

The aim of this paper is to extend the results presented in the literature (Godin at al., 1999; Beyerle and McDermid, 1999; Trickl, 2010; Leblanc et al. 2012) dealing with effective resolution (ERes) estimation for lidar products.

Although it is more common to consider the vertical range resolution for lidar profiles, the effective resolution concept can be easily generalized and extended to the time resolution. This is the case ~~for~~ of time series ~~of~~ lidar products. The application of smoothing ~~in~~ to time series ~~of~~ lidar profiles also ~~modifies~~ changes the effective time resolution of the retrieved products.

~~The~~ This paper is organized as follows. In Section 2 ~~theoretical concepts about~~ foundations on smoothing and numerical derivative on are summarized, by introducing ~~presenting~~ different ~~kinds of~~ low-pass filters realizations ~~that could be (or already are) effectively employed~~ usually used in lidar studies, ~~highlighting~~ their advantages and drawbacks. Section 3 is devoted to the ERes operative estimation based both on ~~the~~ application of the well-known Rayleigh criterion (Born and Wolf, 1999) and on ~~the quantitative~~ analysis ~~of the~~ frequency spectrum ~~removed by smoothing operations, i.e.~~ by calculating the so-called Noise Reduction Ratio (NRR) (Orfanidis, 2009). An ERes ~~operative~~ definition is ~~also provided, employing a cutoff frequency definition for a low-pass filter too~~. Finally ~~it is presented~~ a first and promising ERes estimation approach is presented based on ~~the use of~~ the so-called smoothing kernels, which are commonly adopted within the passive remote sensing scientific community (Haefele et al., 2009). Conclusions summarize the outcome of the paper and include recommendations for the lidar data analysis as well ~~as possible~~ future ~~directions~~ lines. of ~~to deepen~~ the ~~presented study.~~

2 Smoothing and Derivat~~ive~~on of a Lidar profile: the Digital Filter approach

As pointed out in previous work (Pappalardo et al. 2004; Matthias et al. 2004), the most often used algorithms to smooth or differentiate the data within the EARLINET community are those involving some kind of sliding least-square~~s~~ polynomial fitting. [ ] It has been demonstrated that (Savitzky and Golay, 1964; Madden, 1978; Schafer, 2011) the use of these algorithms is equivalent to applying a digital filter for both smoothing and derivative operations. These kinds of filters are widely known as Savitzky – Golay (SG) filters and will be discussed in some detail. Anyhow, the employment of digital filters for the above-mentioned operations is feasible also with other filter types. Without entering into details, largely discussed in several books and papers on Digital Signal Processing (DSP) (e.g. Hamming, 1989; Orfanidis, 2009), in what follows digital filters are defined by the convolution sum~~.~~,

$$y_n = \sum_{k=-N}^{N} h_k x_{n-k}, \qquad n = N+1,\ldots,n_{\max} - N + 1. \tag{1}$$

, $n = 1..n_{max}$ n-th (ok)

where x_n h is the value of the n-th point of the input signal x ~~(for example, lidar raw data or another kind of lidar-derived profiles), consisting of $n_{\max}$ points.~~ y_n is the corresponding filtered value obtained from ~~the~~ linear combination of M=2N+1 (odd) ~~x-values~~ samples of $x(n)$ centered in x_n through the coefficients h. Unless otherwise specified, the word “signal” refers to a generic input/output of ~~a~~ the filter. The Eq. (1) is a representation of the so-called Linear Time, Invariant (LTI) Finite Impulse Response (FIR) digital filter (Orfanidis, 2009). The bounds for ~~the n values~~ sample index n in Eq. (1) imply that a transient effect, ~~will emerge and cause an information loss in~~ and hence, distortion on the smoothed signal. ~~by removing 2N data points from the output.~~ In fact, this transient will normally affect the output signal removing N points at the beginning of signal x and N points at ~~the~~ its end ~~of it~~, although there are techniques (Gorry, 1990;

Otherwise, the lack of "neighbour" samples on the left (on the right) of the signal will ~~cause~~ be treated as "zeroes" and cause an initial (end))

// → New paragraph

Khan, 1987; Leach et al. 1984; Orfanidis, 2009) that are able to deal with this problem. In the study of atmospheric processes in the troposphere, which are the primary objective of EARLINET, the se transient effects could limit the ability of retrieving information ~~in the~~ PBL, which is already limited by the problem of ~~the~~ incomplete overlap between of the lidar transmitted beam and the receiver field of view, if not properly corrected (Wandinger and Ansmann, 2002). Indeed, if the spatial extension of the region of incomplete overlap is not well known, the smoothing of a profile including this region might bias the a retrieval (of the α_a profile, for example) at ~~the~~ lower ranges. // In Eq. (1), ~~The~~ coefficients h_k are the so-called impulse response of a LTI FIR filter ~~and the equation~~ (Karam et al., 1999; Hamming, 1989; Smith, 2007),

Its Direct Fourier Transform

¿Why use ν=2f? MATLAB convention?

$$H(\omega) = \sum_{k=-N}^{N} h_k e^{-i\omega k}, \qquad \left(f = \frac{\omega}{2\pi} \right) \tag{2}$$

gives

of the ~~FIR~~ filter, and in this case it is

~~is~~ the frequency response, a real function that can ~~assume~~ take both positive and negative values (Smith, 2007; Mitra, 2001; Oppenheim and Schafer 2009; Orfanidis, 2009). Because of aliasing, when working from the frequency point of view, the attention is limited to the frequency interval 0< ω=2πf ≤ π (Hamming, 1989). The latter condition could also be written as 0<ν=ω/π ≤ 1, with

Improve

usually called the

DEFINE + Eq.

ν ~~called either reduced or~~ normalized frequency: ν will be used as an independent ~~variable~~ in all the frequency response plots presented in this work. As an example, three different ~~a few~~ $H(\nu)$ curves are showed in Fig. 1 for an SG filter obtained using a 2nd degree polynomial (SG2) for different values of parameter N. //

See Word #

DSP

Equation (1) cannot be used in DSP ~~application that requires~~ real-time ~~processing~~ ~~since the~~ future

discrete-time samples to be

However,

it required

~~input data are obviously not yet~~ available, ~~but~~ because the analysis of a lidar signals is typically carried out offline (i.e. once the ~~after that~~ a whole profile has been acquired in full ~~is fully retrieved~~), "future" samples become available and, ~~in what follows is assumed to~~ therefore, it is possible to use ~~deal always with those~~ non-causal ~~(or mixed)~~ filters (Orfanidis, 2009). // As its ~~the~~ name suggests, H is a direct representation of how a filter alters the frequency content of a signal. In lidar studies, the signal data relevant features are generally confined in the lower frequency portion of the signal

Move closer to Eq.(1) + Fig.(NEW)

Terminology: DISCRETE-TIME SAMPLES

spectrum. For those frequencies that correspond to an *H* equal or close to the unity, no or slight alterations are made to an input signal (pass-band region), while those frequencies that correspond to H=0 are completely removed from it (stop-band region). A negative value of *H* corresponds to a phase shift of π in the signal output respect to the input (Beyerle and McDermid, 1999), which results in artifacts in the output signal (called also ringing or side-lobe effect).

To clarify all the above effects from the frequency point of view, let's see what happens when a low pass filter is applied to an oscillating input signal described by the following equation:

$$x_n=\cos(2t_n^{\,2})\,;\; t_n=\left(\frac{2\pi}{t_s}n\right);\; n=0,1,2,\ldots,t_s\,. \qquad (3)$$

The Eq. (3) is a representation of the so called chirp-like signal, which is useful for our scope because its spectrum contains several frequencies, starting from the DC (ν=0) toward the higher ones, as can be seen in Fig. 2. The low frequency part could be thought as the signal to preserve, while the higher frequency part represents the noise to eliminate. The result of the application of a low pass digital filter is summarized in Fig. 2. In particular, artifacts are present, showed up as waves, both poorly attenuated and inverted in sign respect to the input signal, and located where the abscissa in the smoothed signal plot is between ~3.3 and ~5.5 (corresponding to the first side lobe in the stop-band of the frequency response plot). A SG filter has been selected for this example because it is one of the most employed smoothing filter and also because it will exhibit all the above mentioned effects resulting from the smoothing process.

Both the frequency and the impulse responses of the filter contain alone a complete information and if only one of them is known the other can be retrieved exploiting the properties of the Fourier transform. This latter characteristic is useful, for example to obtain a reliable lidar-ratio estimation independently on the actual definition of ERes, as reported in Sect. 3.1. Due to the large dynamic range of a lidar profile, digital filters with a different frequency response (i.e. for example with different *N* value for SG filters of fixed polynomial order) could be applied at different altitude

→ too much focused on SG.

→ many other FIR design are possible

ranges, in order to deal properly with local values of the signal-to-noise ratio (SNR). In the following sections, some details will be given about few digital filter types that could be employed in lidar data processing. They have been selected among others because already employed in lidar studies (and in several other scientific fields) and/or they are different enough to highlight some relevant features useful for the purposes of this work. Anyhow, there are many other recipes to design efficient low pass filters (e.g. Eisele, 1998; Trickl, 2010). Therefore, the study presented in this paper could be not considered as omnicomprehensive.

The digital filter approach, as long as error propagation is concerned, enables us to estimate the random error associated to the output in a relative easy manner. In fact, the error propagation equation for the summation reported in Eq. (1) is simple, or at least rather straightforward, if compared with the covariant matrix calculations that are needed when the standard sliding least squares polynomial fitting is applied to smooth or to derive a signal. From Eq. (1), the following equation for the y_n variance could be written (Gans, 1992):

$$\sigma_{y_n}^2 = \sum_{k=-N}^{N} \left(\frac{\partial y_n}{\partial x_{n-k}} \right)^2 \sigma_{x_{n-k}}^2 = \sum_{k=-N}^{N} h_k^2 \sigma_{x_{n-k}}^2 . \qquad (4)$$

The above equation is strictly correct if no correlation exists between errors, i.e. if the covariant error matrix is diagonal for the input signal (Gans, 1992), which is a hypothesis frequently assumed in lidar studies and in many other scientific fields. That covariant error matrix should not be confused with the one that is obtained when least square calculations are concerned. The latter one is instead associated with the polynomial coefficients (Bevington and Robinson, 2003) and it is needed to assess properly the error evaluation when the standard least square approach is adopted in the smoothing/derivative process. Anyhow, if further operations are performed on a signal after the smoothing process, the error estimation must be carried out with particular attention. In fact, even if the errors of the initial input signal are uncorrelated, because of the convolution, the data or parameter errors that belong to the smoothed signal will be instead correlated (Gans, 1992).

2.1 Low pass filter and first derivative

Besides direct smoothing, the first derivative is the other operation frequently used in lidar data analysis. The frequency response of the ideal first derivative filter is (Mollova, 1999):

$$
\begin{aligned}
&H^{(1)} = i\pi\nu = \pi\nu e^{i\pi/2} \\
&\left|H^{(1)}\right| = \pi\nu
\end{aligned}
\tag{5}
$$

$H_d(\omega) = j\omega \qquad -\pi \le \omega \le \pi \qquad$ or $\qquad -\frac{1}{2} \le f \le \frac{1}{2}$

Indeed, the Eq. (5) shows a significant difference from a low pass filter: since its frequency response grows linearly with ν, the ideal derivative can be seen as a noise adding process because it amplifies high frequencies, an unwanted feature for our purposes. Therefore, to calculate the first derivative of a signal, the ideal filter could not be directly employed otherwise the output will result useless because of the embedded noise amplification. It is worth to mention that the InfoWorld's "Epic failures: 11 infamous software bugs" (Lake, 2010) reports as the most likely reason of the Mariner 1 space mission failure was caused by a not smoothed time derivative of a radius: "...Without the smoothing function, even minor variations of the speed would trigger the corrective boosters to kick in. The automobile driving equivalent would be to yank the steering wheel in the opposite direction of every obstacle in the driver's field of vision...". In Fig. 3 the chirp function of Eq. (3) is plotted along with its analytical first derivative; it helps to figure out why this amplification happens. In the same fashion of the previous example reported in Fig. 2, the "good" portion of the derivate signal is the low frequency one (for example the part corresponding to the 0÷1 interval of the Time axis), but now the high frequencies (the noisy portion) are strongly amplified respect those originally included in Eq. (3) and the higher are the frequency the higher is the amplification, as described by Eq. (5).

BANDWIDTH-LIMITED DIFFERENTIATOR (8,17)

bandwidth-limited differentiator

For this reason, to obtain a low noise first-derivative profile, a proper tradeoff has to be considered between a strictly correct derivative procedure for the whole signal and the necessary cut of high frequencies. This means that some kind of low pass filter should be applied. In other words a low pass differentiator is wanted, i.e. one whose overall frequency response can be written as $H^{(1)L}$ and that can be thought as a cascade of a low pass filter H^L and the ideal derivative $H^{(1)}$ (Luo et al., 2005 Zuo et al., 2013):

$$H^{(1)L} = H^L H^{(1)}. \tag{6}$$

The impulse response coefficients of this generic first-derivative smoothing filter can be written as $h_k^{(1)L}$. This kind of impulse response has an odd symmetry ($h_k^{(1)L} = -h_{-k}^{(1)L}, h_0^{(1)L} = 0$) (Hamming, 1989; Smith, 2007) and a frequency response that, from the Eq. (2) and using the Euler formulas, can be written as (Yunlong, 2012):

antisymmetric unit sample response

$$H^{(1)L} = \cancel{i}\sum_{k=-N}^{N} h_{-k}^{(1)L} \sin(\pi\nu k). \tag{7}$$

Revise!

non-causal

symmetric unit-sample response

where the cosine terms vanished. It is worth to recall that for low-pass filters, the terms that disappear in Eq. (2) are the sine terms, because of the even symmetry ($h_k = h_{-k}$) of their impulse responses. Thus, from the Eqs. (5), (6) and (7), the low pass filter frequency response for a generic derivative smoothing filter can be written as:

[illegible] (ok) ✓

$$H^L = \frac{H^{(1)L}}{i\pi\nu} = \frac{\sum_{k=-N}^{N} h_{-k}^{(1)L} \sin(\pi\nu k)}{\pi\nu}. \tag{8}$$

This latter equation will be useful for the determination of the effective resolution discussed in Sect. 3 (Masci, 1999; Godin, 1987). In Fig. 4, results from the Eq. (5) and from both the Eq. (7) and Eq. (8) are plotted, the latter two evaluated for a~~n~~ SG2 low-pass derivative filter. Hereafter, the smoothing portion (H^L) of a low pass derivative filter ($H^{(1)L}$) frequency response will be indicated with the letter "d" before the parent low pass (i.e. dSG2 for those in Fig. 4). In summary, the low pass filter in Eq. (8) could be considered as the measure of "how well we did" in the approximation of the first derivative of a signal (Hamming, 1989) because this equation is the ratio between the actual employed filter frequency response in Eq. (7) and the ideal one.

#But this is too messy.

N is the impulse-response half length (of the filter) | 2N+1 is the filter length
P is the polynomial order

2.2 The Savitzky – Golay Filter

~~qualitatively~~?

The SG approach allows to gain computational speed and it is relatively easier to implement than the standard least-squares calculations though, according to the theory, they would produce the same results (Savitzky and Golay, 1964). Referring to Eq. (1), the coefficients h_k ~~in~~ have to be calculated just once for fixed both N and polynomial degree (P), while using the standard least-squares smoothing a new and complete calculation of polynomial coefficients has to be done for each point of a signal, even if N and P are fixed (Press et al., 2007). It is important to note that the minimum N required to perform a meaningful smoothing is related to the chosen polynomial degree through the relation $2N>P$ (Schafer, 2011), and for $2N=P$ there is no difference between the input and the output signals (no smoothing). // With SG filters, in principle, a different and variable number of points could be used when smoothing or deriving a profile as well, as a different polynomial degree if required (Barak, 1995), and without a strong increase of the computation time. This speed enhancement, common also to the other low pass filters, is quite important especially with the introduction of the Single Calculus Chain (SCC) (D'Amico et al., 2015), a centralized calculus tool developed to perform a near real-time and fully automatic aerosol lidar data analysis within EARLINET.

→ Key is the sharp~~ones~~ness in the passband-to-stopband transition

The SG filters are popular in many scientific fields because they preserve not only the position and the area of the main signal peaks, but potentially also the higher moments. This property is connected to the flat frequency response in the pass-band as reported in Fig. 2. This feature enables a quite faithful preservation of the low frequency component of a signal, i.e., the portion of the signal to keep (Karam et al. 1999). For example, the moving/sliding average (also called box-car), which is the zero-th polynomial-order SG filter (SG0), does preserve the area (its zero-th moment) underlined by a feature in the profile (e.g., an aerosol layer). Using the SG0, the mean position (the first moment) of a symmetric layer is also preserved after the smoothing though this is not true for the standard deviation (the second moment), which could be seen as a measure of width of the layer

① first P

averaging

// New paragraph

- better structure
- bullets. properties

(Ziegler, 1981). In order to preserve the higher moments (Bromba and Ziegler, 1981), a profile can be smoothed by means of a SG filter with a higher-degree polynomial P. In fact, all the moments up to $P+1$ P will be preserved ($P= 0, 2, 4\ldots$, i.e. P_{even}, because for fixed N, the smoothed signal will not change if either P_{even} or $P_{odd} = P_{even}+1$ is used). ② Moreover, an higher P generally corresponds to an increase of the filter pass-band, but this translates in worse performances in terms of noise removal: this again suggests that a tradeoff has to be considered between a better pass-band behavior (i.e., less signal distortion) and a better noise removal (Orfanidis,2009; Savitzky and Golay, 1964; Press et al., 2007; Turton, 1992). As pointed out (see Fig. 1 and Fig. 4), also the filter radius (N) contributes to altering the frequency characteristics of a SG filter. For this reason, the SG filter pass-band depends on two parameters: the polynomial degree P and the filter radius N, as can be seen also from Eq. (10). next Beside flatness of pass-band, an SG filter has ③ a transition band which is generally smaller than other filters with similar pass-band (see Fig. 7 and Fig. 8). This is a valuable characteristic, because this means a sharp separation between pass-band and stop-band (Schafer, 2011).

The main problem with the SG filters is represented by the presence of ④ side lobes in the stop-band that in principle contaminate the output signal with the high frequency artifacts already seen in Fig. 2. Moreover, the magnitude of these side lobes is quite high respect to other low pass filters, as can be clearly observed by comparing Fig. 1, Fig. 4 and Fig. 8. In fact, if in those figures the frequency responses are examined in the stop-band regions, for SG2 filters the observed magnitude for the peak of the first side lobe is about -0.25, which implies a signal suppression of only 75%. Coupled to this poor attenuation, there is also the drawback of the negative sign, which brings to the artifacts, so SG filters do not offer a great performance in the stop-band region. // It should be noted that the above attenuation value for the first side lobe will not significant change for the SG filters we examined, anyhow it became slight worse when P increase (Schafer, 2011). Another useful property of the Savitzy – Golay recipe is that for a given P and N, also ⑤ the impulse response for the

#Move up

corresponding SG low pass derivative filter can be directly calculated (Savitzky and Golay, 1964). The SG low pass derivative filter will produce the same result for P_{odd} and the next even degree $P_{odd}+1$ (e.g. for $P=1$ and $P=2$, for $P=3$ and $P=4$ etc.) for fixed N, and the degree of flatness in the pass-band, associated to the corresponding low pass dSG filter, has to be assessed accordingly (Luo et al., 2005).

2.2.1 Cascade filters

#to basic

There are efficient recipes, that allow eliminating the side lobes (or reduce their size) in the frequency response of a low pass filters: the cascade technique is one of those. Taking advantage of the properties of the convolution in the frequency domain, two (or more) low pass filters in cascade can be easily applied to a signal. In fact, since this operation is linear in the frequency domain, the behavior of filters in a cascade can be simply expressed by the product of the single transfer functions (Das and Chakraborty, 2012), a property already used in Eq. (6). Thus, for a cascade filters the resulting impulse/frequency response can be written as:

$$\begin{aligned} h^{Lc} &= h^{L1} * h^{L2} * \ldots * h^{Ln} \\ H^{Lc} &= H^{L1} H^{L2} \ldots H^{Ln} \end{aligned} \quad . \tag{9}$$

The first equation in Eq. (9) indicates how to calculate the impulse response of a cascade filter (D'Antona and Ferrero, 2006). Cascading multiple identical low pass filters together will effectively damp the side lobes amplitude. However, a drawback of cascading identical filter consists in the reduction of the pass-band extension. In what follows, we will study how to avoid this effect and how to have more control on the cascading process when only two filters are involved. If the pass-band has to be preserved when two successive low pass filters are applied, it is useful to use a relationship among P, N and the pass-band extension. This latter parameter could be given by the location of the cutoff frequency ν_c taken at -3db level *(or $H\approx0.7$)* and for SG filters can be written as (Schafer, 2011):

$$\nu_c = \frac{P+1}{3.2N-4.6}, P = 0,2,4.... \qquad (10)$$

$N \geq 25$ and $P<N$

Just an approximation

Operatively for an efficient cascade of two smoothing filters (*L1,L2*), in Eq. (9) *L2* have to be chosen with a ν_c large enough to cover the frequency response of *L1* up to the start of its stop-band. In this respect, when SG filters are involved in the cascade, good results are obtained using two SG filters with the same *N* and Δ*P*=2. As a consequence, in the resulting cascade filter, the stop-band is much less affected by the presence of side lobe, while its pass-band will be nearly the same of the SG filter with the lower polynomial order, as can be seen in Fig. 5, and for this reason it will cause a quite similar effect on a signal for frequencies located in this latter region. In Fig. 5, the chirp function is smoothed with an SG2 and an SG4 in cascade having the same *N* and the results can be directly compared with those in Fig. 2, too. It can also be noted that in the cascade filter the pass-band is very similar to the one associated to the SG2 while the stop-band shows much less pronounced side lobes. √

→ consider to use $SG2^{L1+L2}$ instead of $SG2^{L1} * SG4^{L2}$; L

However, as a drawback, the transient zone at the start and at the end of the output signal increases and in our case they are equal to *N*(*L1*)+*N*(*L2*) at each end. This means that if *N* does not vary for the two considered smoothing filters, the loss of information at the start and at the end of the output signal is doubled compared to the case with a single filter application. // Efficient results are obtained also for the cascade between an SG*P*(*N*) filter with the corresponding dSG*P(N)*, as showed in Fig. 6. In this case for the cascade filter, the maximum absolute value of the side lobes magnitude is about 0.02, i.e. almost negligible for practical purposes, while the pass-band of the dSG2 results almost unchanged, though the difference is slightly more pronounced than in the previous case. This outcome is not so surprising because this kind of cascade also fulfills the rule of thumb for an efficient SG cascade combination, only perhaps a little relaxed. In fact, derivation implies losing a degree in the polynomial order. Therefore the dSG4 could be considered similar to an SG low pass filter based on a 3rd order polynomial and SG3=SG2. Because the above considerations, the dSG*P*

notation (?)

notation

filter and the corresponding SGP could be considered having $1<\Delta P<2$, which happens to be still good enough for our purposes. Moreover, this type of filter cascade is also computationally efficient because both the needed impulse and frequency responses are calculated simultaneously in the SG algorithm. It is worth to point out that the cascade method is useful also to design effectively high order derivative filters of a signal (Gans and Gill, 1983): for example, two consecutive stages of a first-derivative filter will lead to the second derivative of the input signal.

To summarize, operating with such cascade filters retains all the advantages of the SG filters with an added value in terms of efficiency for the high frequency damping without introducing artifacts, though the transient zone growth could be a potential problem for lidar applications. In what follows, if not otherwise stated, the value of N associated to a cascade of two smoothing filters indicates the value of N used in both the filters of the cascade and not to the overall filter radius, i.e. for a dSG4(9)·SG4(9); N for the cascade is 9, although the overall filter radius is 18.

italics

LPF filtering (in general sense)

A possible application of cascade filtering to lidar data could be the improvement of the SNR of a lidar signal for achieving a more accurate calibration of Raman/Elastic signal ratio, because of the benefit of the possible reduction in the width of the selected calibration range (Ansmann et al., 1992). This is particularly relevant for calculus routines that make use of an automatic range-finder algorithm for the normalization of the signal ratio: this requires a good SNR in order to reduce the range extension where the normalization is performed. The reduction in the width of the calibration range also reduces the normalization uncertainty and its impact on the total uncertainty budget. For this purpose a viable solution is represented by the application of the smoothing also on the signal ratio, before the retrieval of β_a. # define Then, after the processing phase another smoothing filter generally would be applied, to obtain the β_a profile with an acceptable noise level. To this end the recipe given in this section for the construction of cascade filters could be used to be sure that the second smoothing does not eliminate in the profile the details that are spatially larger than those already damped with the application of the first filter.

Better consider the simpler example of "automatic ~~range~~ smoothing of the ~~cal~~ calibration range in the Klett-Fernald algorithm for elastic signals".

¿ in FREQUENCY?

↑

2.2.2 Windowed filters

sinc in H(w) ← [of leakage/Gibbs phenomenon. # Better introduce the problem'. E.g. ~~compute~~ truncation (i.e., windowing) of x(n) using a rectangular window.

In filter design, the necessity to deal with finite length impulse response, gives rise to the so called spectral leakage (Harris, 1978) that could lead to significant undesired oscillations in the frequency response, including side lobes. In order to reduce this latter effect, tapered window functions are generally applied to suppress efficiently the oscillations in *H*. The simplicity of the design process has made this method very popular. // Each window function is a kind of the usual compromise between the requirements of higher selectivity, i.e. the narrowest the transition region and the highest suppression of undesirable spectrum, i.e. the highest stop-band attenuation (Mitra, 2001). Therefore, windows can be seen as weighting functions applied to data in order to reduce the spectral leakage associated with finite observation intervals, i.e. high frequency noise. If w_k is a tapered window function, the minimization of the ringing for a low pass filter can be obtained applying w_k to the impulse response, thus Eq. (1) could be written as:

$$y_n = \sum_{k=-N}^{N} h_k x_{n-k} = \sum_{k=-N}^{N} w_k h_k^0 x_{n-k} \, . \tag{11}$$

unclear

$$\begin{tikzcd} x(n) \arrow[r] & \otimes \arrow[r] & \boxed{h(n)} \arrow[r] & y(n) \\ & w(n) \arrow[u] & & \end{tikzcd}$$

$y(n) = [\, x(n) w(n) \,] * h(n)$

where h_k^0 are the impulse response coefficients of a generic low pass filter. Several listed window functions are reported in literature (Harris, 1978) to design a specific filter. If h_k^0 are samples of a proper optimized sinc function, Eisele has introduced a efficient window function of the Blackman–type to lidar work:

↳ Why sinc functions appears here in such ~~so~~ a disorganised way? #Previous mention by the ~~beginning~~ beginning of Sec. 2.2.2 necessary!

$$w_k = 0.42 + 0.5\cos(\pi \frac{k}{N}) + 0.08\cos(2\pi \frac{k}{N}) . \tag{12}$$

↑ revise!

The filter constructed with this window (Eisele, 1998; Trickl, 2010) does not exhibit ringing. The removal of the ringing due the window application can be observed in Fig. 7 where a Blackman–type window is applied to an SG2 filter. The side lobe disappears also in this case and the pass-band is nearly conserved, if the same *N* is used. On the other side, as can be seen from the right plot in

Fig. 7, the transition band in the SG2 filter with the Blackman–type window applied, becomes quite large causing both (see the left plot in Fig. 7) a less efficient damping of those frequencies over the pass-band and before the first side lobe of the (not windowed) SG2 (i.e. for 0.2<ν<0.3) and a slight worst preservation of frequencies in the pass-band (ν<0.2).

2.3 The Gaussian Filter

COMPLEX REASONING. Simply use basic property of FOURIER TRANSFORM.

$\frac{d}{dt} x(t) \rightarrow j\omega X(\omega)$!

The Gaussian filter (G) is another option widely adopted to smooth signals especially in image processing (Romeny, 2003). This filter is characterized by a single parameter (σ, the standard deviation), and its impulse response (a zero mean Gaussian) has the advantage that can be written analytically for both the smoothing and for the ~~low~~ derivative pass first (and, if needed, also higher orders) derivative: why low-pass?

$$h_k(\sigma) = g_k(\sigma) = (2\pi\sigma^2)^{-1/2} e^{-\frac{k^2}{2\sigma^2}};$$
$$h_k^{(1)L}(\sigma) = g_k^{(1)}(\sigma) = -\frac{k}{\sigma^2} g_k(\sigma). \qquad (13)$$

$$h_\sigma(n) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{n}{\sigma}\right)^2}, \quad -\infty < n < \infty$$
$$h'_\sigma(n) = -\frac{n}{\sigma^2} h_\sigma(n) \; ; \; \left[h'(n) = \frac{d h(n)}{dn} \right]$$

To be used as a digital filter, the Gaussian curve and its derivatives have to be sampled, as already done in writing the Eq. (13). The Fourier transform of a Gaussian function (which is Gaussian too) is everywhere non-zero and, therefore, cannot be sampled without some aliasing. The aliasing will result negligible if σ≥1 (Hale, 2011), although even a slight lower value is allowed by some authors (Romeny, 2003). Moreover, to get a usable impulse response, it must be truncated somehow. Luckily, the Gaussian curve has a quick approach to zero and for this reason it can be truncated without a strong approximation. In fact, Eq. (13) provides a value less than 0.0004 for |*k*|≥4σ. This latter condition implies that, to proper truncate the impulse response, it is sufficient to employ a value of *N* equal to 4σ (actually the nearest integer to 4σ) in Eq. (1) with no needs to go beyond this value. Because of the properties of the Gaussian function, if the above condition for σ is respected

$h_\sigma(n) \longrightarrow H_\sigma(w)$ # H is $H_\sigma(w)$

FOURIER TRANSFORM

simplify your reasoning!

and being $H^{G,\sigma}$ the frequency response of a Gaussian filter with paremeter σ, H^L could be obtained from Eq. (8), Eq. (2) and Eq. (13) (Hale, 2011):

$$H^L = \frac{H^{(1)L}}{i\pi\nu} = \frac{i\pi\nu H^{G,\sigma}}{i\pi\nu} = H^{G,\sigma}. \tag{14}$$

The Eq. (14) implies that the low pass filter (that can be indicated with dG in analogy with the denomination adopted for SG filters) embedded in a Gaussian first derivative smoothing filter with parameter σ, is indeed a Gaussian low pass filter with the same parameter. Of course, this will simplify somehow our duties when operations like the lidar ratio profile determination are performed, as will be showed in Sect. 3.1. When σ increases, the pass-band reduces its extension and provides a stronger smoothing effect, although a Gaussian filter has a transition band ~~quite~~ wider than a SG filter with a similar pass-band. A Gaussian filter is also less flat in the pass-band (van Vliet et al., 1998) than a SG filter (for $P \geq 2$), but it has also the advantage of being almost without side lobes (i.e. no artifacts in the stop-band).

in fact it is similar to the Blackman window not to SG2+BLACKMAN Sect.2

Figure 8 summarizes the performances of the low-pass filters described in this Section: filters with similar pass-bands are reported to show their differences in the whole frequency domain. It can be seen that the Gaussian filter exhibits a behavior quite similar to the SG2 with a Blackman–type window and both have no evident side lobes. Fig. 8 also clearly shows that SG2, as well as all other plain SG filters (i.e. the SG2, SG4 etc., therefore those not modified by cascading, windowing etc.), has a slight (much) better behavior in the pass-band (ν<0.1) than Gaussian/SG2 windowed filters, i.e. a more faithful signal preservation, but are heavily affected by side lobes. From the point of view of lidar studies, an application of filters (with lower) without side lobes corresponds to obtaining a well-smoothed profile without the presence of any high frequency residual (or artifacts). Both the Gaussian and the SG2 windowed filters also exhibit a much slower transition to the stop-band respect to the others, i.e. a less sharp separation between pass and stop-band. Finally the cascade filter is able to get all

#less distorsion in the passband

Compare using: $20 \log |H(w)|$ and you will see...

the attractive characteristics of the others, but it also has the described drawback of an enlarged transient (i.e. a more pronounced loss of information in the output signal, see Fig. 5).

Spatial

3. ~~The Effective~~ Resolution

Define notation (aerosol back coeff)

The investigation of the synthetic lidar data inversion (Pappalardo et al., 2004) in Fig. 9, helps to recognize the effective spatial resolution as relevant in lidar data analysis. It highlights that the effective resolution plays an important role to assess properly the problems that could arise when data with different resolution are combined. In this latter figure, the aerosol layer inserted in the true profile at 1.4-1.6 km results heavily smoothed by the low pass filter used in the retrieval. If the β_a is smoothed, the resulting lidar ratio profile is consistent with true one (see Fig. 9, central and right panel), both in value and in behavior. On the contrary, if β_a is not smoothed, the lidar-ratio profile in the layer results quite different from the synthetic one. Outside the layer the differences between the retrieved lidar-ratio profiles are less relevant because the aerosol field is nearly constant and for this reason less sensitive to the distortion effect of smoothing filter (Ziegler, 1981).

which type of filter? BOXCAR?

Too discursive/ambiguous. KEY is to begin with RESOLUTION ⟺ trade-off

Two approaches will be considered for the quantitative assessment of the ERes. // The first one is related to the distortion induced by the smoothing process on any non-trivial input signal (Enke and Nieman, 1976; Ziegler, 1981). In fact, the area preservation property (common to all the considered smoothing filters, see Sect. 2.1) implies that if the peak of a layer is reduced, its spatial width will increase and potentially could overlap with another feature present in a profile. The final result will be that it is no longer possible to distinguish one peak from another, i.e. they are no longer resolved: this means that a low pass filter reduces the vertical resolution. This latter statement naturally leads to the use of the Rayleigh criterion (Born and Wolf, 1999) for the determination the effective resolution. // The second approach is based on the removal of high frequency noise due to the smoothing operation (Gans and Gill, 1983; Orfanidis, 2009). Since high frequencies in space domain correspond to small scale details in the lidar profiles, if they are lost in a certain amount this

// New paragraph

will imply a reduction of the resolution in the output profile respect to the input one. Incidentally, it should be noted that since a smoothing filter damps effectively only high frequencies and since it is common to deal with white noise, the low frequency portion of the noise is still present in the smoothed signal, for example in the form of long wave ripples (Gans, 1992). Moreover, a link is established between the ERes estimated with each of those two approaches and the ERes evaluated via the proper cutoff frequency definition, in analogy to previous works (Godin, 1999; Masci, 1999; Beyerle and McDermid, 1999; Leblanc et al., 2012). Before discussing the two above mentioned methods, using the results of Sect. 2.1 an answer will be provided to the question about how to obtain a lidar-ratio profile that comes from aerosol extinction and backscatter profiles with the same effective resolution.

3.1 Obtaining profiles with the same effective resolution: the lidar ratio case

To retrieve the α_a profile (Ansmann, 1992) a first-derivative smoothing filter is applied. The frequency response of the embedded low-pass filter (H^L) can be found from Eq. (8), or directly with Eq. (14) if a Gaussian derivative filter is employed. The possibility to retrieve H^L, gives the solution to the problem of retrieve a consistent lidar ratio and without hypothesis or assessment about the effective resolution itself of the profiles involved: it is only needed that they share the same resolution. In fact, once H^L is known it is possible to smooth the corresponding β_a with this filter and as a result obtain both the profiles with the same effective resolution. The impulse response h_k^L of this low pass filter can be retrieved by means of what is generally called Filter Design by Frequency Sampling (Rabiner et al., 1970; Rabiner and Gold, 1975; Burrus, 2012). With this method, the frequency response H^L is sampled at a set of equally spaced frequencies. Thus, by using the Inverse Discrete Fourier Transform (IDFT), the desired filter impulse response can be determined:

$$h_k^L = IDFT[H^L(\nu_n)]. \qquad (15)$$

The resulting filter with an impulse response like Eq. (15) will have a frequency response that is exactly the same as H^L at each ν_n, so the better the original frequency response is approximated the smaller the interpolation error between them ~~is~~ (Johnson, 1989). The impulse response h_k^L, retrieved by Eq. (15), can now be used in Eq. (1) with the aerosol backscatter profile at raw resolution as the input signal. In this way β_a is smoothed with the same low-pass filter H^L applied to get the α_a profile. Even more directly the same result can be obtained by means of the IDFT only:

$$y = IDFT(XH^L) . \qquad (16)$$

Both operations written in Eq. (15) and Eq. (16) could be computed with a proper use of FFT algorithms. In Eq. (16), X is the Discrete Fourier Transform (DFT) of a generic signal which, for our purposes, will be the aerosol backscatter profile at raw resolution. Since the frequency spectrum of both the profiles has been changed by the same low pass filter, then both share the same effective resolution. To illustrate better the above concepts, in Fig. 10, a retrieval of the optical parameters are performed starting from simulated elastic/Raman lidar data (Ansmann, 1992; Pappalardo et al., 2004) with an aerosol layer 1000 m thick. The signals have been simulated for the Rayleigh signal at 351 nm and for the corresponding nitrogen Raman signal at 382 nm, without adding noise or background. Both the low pass derivative SG2 and the low pass derivative Gaussian filters are employed to retrieve the aerosol extinction profile. Then the embedded low pass filter (i.e. the dSG2 and dG) impulse response, retrieved by Eq. (13) and Eq. (15) respectively, is used to smooth the raw resolution aerosol backscatter profile. Both ~~the~~ extinction and back coeffs profiles with the same ERes are combined to get an estimation of the lidar ratio. Figure 10 shows that beside the good results for the retrieval of the lidar ratio inside the actual simulated layer, an accurate result is also obtained in the zone immediately outside the layer, i.e. where the filter distorts the profile with respect to the true layer. If the correct H^L is used to smooth the backscatter profile, this makes the information about the lidar ratio correct even at those ranges where the aerosol presence in the retrieval is only due to the distortion action of the filter. In Fig. 10, it is also shown that wrong lidar-ratio values are obtained in almost all the aerosol layers if the β_a profile is smoothed with a low-pass filter (indicated with H in Fig. 10) that is different from H^L.

3.2 The effective resolution: the Rayleigh criterion

The Rayleigh criterion is generally accepted in spectroscopy for the determination of the minimum resolvable detail (Born and Wolf, 1999). It is an empirical criterion, and states that two peaks are considered fully resolved if the drop in intensity between them is lower than 74% of the peak intensity. This is a result of the diffraction formulation that says that the imaging process is named diffraction-limited when the first diffraction minimum of the image of one source point coincides with the maximum of another. The application of Rayleigh criterion for the determination of the effective resolution could be done by analyzing the behavior of a couple of unitary pulses under the action of a low-pass filter. Operatively, two unitary pulses at fixed distance are smoothed by a low pass filter whose parameter are changed to achieve a increasing signal distortion. Increasing N for SG filters with fixed P, or σ for Gaussian filters, it is possible to find the maximum value of the filter parameter that allows to still resolve the two smoothed pulses according to the Rayleigh criterion. Then the effective resolution to be associated to that particular smoothing filter is exactly this distance. Moreover, this procedure, also known as "step function" method, has been already tested in the frame of the first EARLINET algorithm intercomparison (Pappalardo et al. 2004). An alternative approach, used in the lidar community, is based on the analysis of the full-width at half-maximum (FWHM) of a finite impulse after a smoothing procedure is applied (Leblanc et al., 2012) or to the response to a Heaviside step function (VDI(Verein Deutscher Ingenieure), 1999; Eisele and Trickl, 2005; Vogelmann and Trickl, 2008). However, with SG filters, apparently the step function procedure shows some ambiguous results as can be seen in the examples reported in Fig.

? which parameter? whose parameters? plural? with a test parameter changed

11. In fact with plain SG filters (with $P \geq 2$) it could be difficult to properly define when the Rayleigh criterion is satisfied (or not) because the occurrence of artifacts like bumps between the two peaks (Fig. 11b) and/or the displacement of the smoothed peaks from their original position (Fig. 11c). In those cases, the ratio used for the application of the Rayleigh criterion is evaluated between the intensity at the peak and the intensity at the midpoint between the peaks. Because of the artifacts, the intensity at the midpoint is not always the absolute minimum: therefore this ratio brings to a more conservative ERes estimation. Those drawbacks in the application of the Rayleigh criterion could represent a further problem caused by the presence of side lobes with significant magnitude. Instead, for filters like the Gaussian one or any filter with less important side lobes (like dSG2 and the properly built cascade filters) no major problem is observed applying the Rayleigh criterion. However, the step function method used in the case of the boxcar of width M SG0 filter leads to a first operative definition for the ERes. In fact, from Fig. 12, it should be clear that the effective resolution in this case is simply reduced by a factor of $M=2N+1$, because under the action of the SG0 all the involved data points will be equally weighted. So the ERes (ΔR_{Eff}) associated to the boxcar filter can be explicitly written as:

$$\Delta R_{Eff}^{Ray,SG0} = (2N+1)\Delta R_{raw} \,. \qquad (17)$$

It is worth to mention that for the SG0 the effective resolution is also equal to the inverse of its impulse response coefficient (multiplied by the raw resolution ΔR_{raw}), which in this case, for any given N, is a constant independent of k:

$$h_k^{SG0} = \frac{1}{(2N+1)} \,. \qquad (18)$$

To try to resolve the observed ambiguity in the application of the Rayleigh criterion to plain SG filters (with $P \geq 2$), the considerations done in Sect. 2.2.1 about the cascade filters can be exploited.

In fact, since the features of the cascade filters are constructed with our rule of thumb, it is plausible that *L1* filter shares almost the same ERes with the cascade *L1*·*L2*. For example, the ERes estimated for SG2·SG4 could be also used for approximately the SG2. Figure 13 shows the kind of effect that the cascade will produce on the central bump, making more straightforward the application of the Rayleigh criterion. In Fig. 14 there are some results of the application of the Rayleigh criterion to plain SG and the corresponding cascade filters (~~in the sense explained above~~) that show how the ERes of an SG*P* exhibits a behavior quite similar to the corresponding cascade filter. Therefore, the occurrence of artifacts seems to have a limited effect in the ERes determination (<5-10%).

Exploiting the quite evident linear relationship between the ERes and *N* that results from Fig. 14, the following equations are obtained:

$$\begin{aligned}
\Delta R_{Eff}^{Ray,SG2\cdot SG4} &= (1.17N-0.09)\Delta R_{raw} \sim \Delta R_{Eff}^{Ray,SG2} = (1.24N-0.24)\Delta R_{raw} \\
\Delta R_{Eff}^{Ray,SG4\cdot SG6} &= (0.80N-0.65)\Delta R_{raw} \sim \Delta R_{Eff}^{Ray,SG4} = (0.74N-0.48)\Delta R_{raw}\,. \\
\Delta R_{Eff}^{Ray,SG6\cdot SG8} &= (0.60N-0.78)\Delta R_{raw} \sim \Delta R_{Eff}^{Ray,SG6} = (0.62N-0.86)\Delta R_{raw}
\end{aligned} \tag{19}$$

(enumerate them) (dSG2, SG2+Block, G)

For the other filters under investigation, the application of the Rayleigh criterion does not give particular problems: the results are reported in Fig. 15.

Of course also for the filters in Fig. 15 the ERes could be written by linear fit:

$$\begin{aligned}
\Delta R_{Eff}^{Ray,dSG2} &= (1.55N+0.83)\Delta R_{raw} \\
\Delta R_{Eff}^{Ray,SG2+Blk} &= (0.80N+0.20)\Delta R_{raw}\,. \\
\Delta R_{Eff}^{Ray,G} &= (2.79\sigma-1.04)\Delta R_{raw}
\end{aligned} \tag{20}$$

For example, if an aerosol extinction profile is retrieved from a nitrogen Raman lidar signal with a raw resolution of ΔR_{raw}=15 m and by means of an SG2 derivative low pass filter (i.e. the low pass filter to consider is the dSG2) with *N*=30, its estimated ERes will be about 700 m. Because of the constraints on *N* and σ discussed in Sect 2.2 and Sect 2.3, the ERes given by Eq. (19) and Eq. (20), will be always positive and larger than the raw resolution ~~for all the low-pass filters (~~and less than

; ΔRaw,

unclear ? assumption

$(2N+1)\Delta R_{raw}$: (~~the~~ upper limit given by SG0). For example the linear fit in Eq. (19) for plain SG filters is performed with the constraint that these filters for $N=P/2$ do not smooth, therefore they do not change the vertical resolution ($\Delta R_{Eff}=\Delta R_{raw}$). It should be noted that regardless of whether (or how) the linear fit is constrained or not, the slope does not significantly change and the intercept values will have always a low impact on the ERes determination (max $\pm 1\cdot\Delta R_{raw}$, a value that could be taken as the estimation of the ERes indetermination). As the filter's parameter grows in Eq. (20) and Eq. (21), i.e. N for SG based filter and σ for Gaussian filters, the intercept values does not matter anymore in the determination of the ERes. Among the SG based filters examined, for the SG2 this is true for any N and the worst case is the SG6, where the difference in the ERes calculated with or without the intercept, becomes <10% for N>15, while with the Gaussian filter the same is obtained with σ>4.5.

~~unnecessary~~

unclear

It was a natural to adopt an operative ERes definition based on the Rayleigh criterion because of its direct relationship with the concept of resolution. Although the use of this criterion led to simple and ready-to-use linear relationships for the calculation of the ERes, no unique equation was found suitable for any given low-pass filter. In fact with the method outlined in this section, for any selected smoothing filter, the whole procedure to retrieve a relation for the ERes has to be done from scratch.

3.3 The effective resolution: the NRR criterion and the SNR Matching Criterion

~~The~~ Removal of the noise embedded in a signal is the main purpose in the application of a low-pass filter. The amount of white noise removed by a generic filter has been already explicitly assessed (Gans and Gill, 1983; Brown, 2000). In fact, in this case, the ratio between input (σ^2_{IN}) and the output (σ^2_{OUT}) noise variance ~~mean-square noise~~ values can be taken as a measure of the noise removed from an input signal after the smoothing. This quantity is also called Noise Reduction Ratio (NRR) and

noise variance, $Var[X] = E[X^2] - E[X]^2 = E[X^2]$ if zero-mean white noise

↓

[synonym of power for a random process].

0

$E[x]=0$ for white noise

depends only on the impulse response of the filter under examination (see Chapter 8.3 and Appendix A.2 in Orfanidis, 2009; Mitra, 2001):

$$\frac{\sigma_{OUT}^2}{\sigma_{IN}^2} = \sum_{k=-N}^{N} h_k^2 = NRR\,. \tag{21}$$

Using the explicit formula for the ERes associated to the SG0 filter and the Eqs. (17), (18) and Eq. (21), led to write:

$$\begin{aligned} NRR^{SG0} &= \sum_{k=-N}^{N} (h_k^{SG0})^2 = \frac{1}{(2N+1)}; \\ \Delta R_{Eff}^{SG0} &= (2N+1)\Delta R_{raw} = \frac{\Delta R_{raw}}{NRR^{SG0}}\,. \end{aligned} \tag{22}$$

From the noise reduction point of view, the Eq. (22) makes possible to infer that the ERes associated to the application of a generic low pass filter L on a signal could be written by means of the general equation:

Say clearly this is just your "Departing hypothesis"

$$\Delta R_{Eff}^{NRR,L} = \frac{\Delta R_{raw}}{NRR^{L}}\,. \tag{23}$$

Adopting a slight different point of view, a proof or at least a solid hint of the validity of Eq. (23) could be provided. Given that a low pass filter alters the SNR, it is reasonable to assume that if a given signal will emerge with the same SNR after the smoothing with different low pass filters, then those filters act on the signal in a similar fashion noise-wise. Than it could be also inferred that the filters, although different, will also have caused the same alteration of the resolution on that signal and for this reason the output profiles will have the same ERes. Operatively, the SG0 filter for different values of N is applied on a generic signal, and then the corresponding SNR of the smoothed signal is calculated. Applying on the same signal a generic low pass filter L, which will be characterized by the parameters *params* (i.e. N, P for SG based filters or σ for Gaussian filters),

NOTATION [by a set of parameters $\vec{p} = (p_1, p_2, \dots, p_n)$, e.g. $\vec{p} = (N, P)$ for SG-based filter or $\vec{p} = (\sigma)$ for a G-based filter…

Simply, you could input unitary ~~Ga~~ Gaussian noise, $\sigma_{in}=1$; $x = 1 \cdot \mathrm{randn}(1, 1000)$; and measure σ_{out} at the filter output. TIP

an optimization process can be performed to find the [N_0, *params*$_0$] couple that makes the average differences between the two SNRs as close as possible to zero (*SNR* matching criterion):

$$\begin{aligned} &\overline{\Delta SNR_{N,params}} = \overline{SNR_N^{SG0} - SNR_{params}^{L}} \\ &\Rightarrow [N_0, params_0] : \overline{\Delta SNR_{N_0,params_0}} \approx 0 \end{aligned} \quad (24)$$

Then, given the Eq. (22), finally, it can be assumed that the ERes of a generic *L* smoothing filter is:

$$\Delta R_{Eff}^{L,params_0} = (2N_0 + 1)\Delta R_{raw} . \quad (25)$$

In Fig. 16, there is an example of the similarity of the SNRs achievable using two different low pass filters. The results of the ERes obtained using the Eqs. (23), (24) and (25) for various low-pass filters can be seen in Fig. 17. The analysis of this latter figure provides a quite clear confirmation of the equivalence of the NRR and the SNR matching criterion. For this reason the Eq. (23) can be used to easily estimate the ERes for any smoothing filter, instead of the less general, and more time consuming SNR matching procedure. In fact, with the NRR criterion, the estimate of the effective resolution is based only on the impulse response of the smoothing filter employed, which is generally known or it can be anyhow calculated via Eq. (8), when necessary, as for dSGP low-pass filters. Figure 17 also shows that the ERes with the NRR criterion could be expressed by linear relationships, as happened with the application of Rayleigh criterion.

For this reason and according to the previous discussions, the results of the linear regression for the same type of smoothing filters can be explicitly written as:

$$\begin{aligned} &\Delta R_{Eff}^{NRR,SG2\cdot SG4} = (0.98N + 0.30)\Delta R_{raw} ; \Delta R_{Eff}^{NRR,SG2} = (0.89N + 0.11)\Delta R_{raw} \\ &\Delta R_{Eff}^{NRR,dSG2} = (1.61N + 1.25)\Delta R_{raw} \quad ; \Delta R_{Eff}^{NRR,SG4} = (0.57N - 0.15)\Delta R_{raw} \\ &\Delta R_{Eff}^{NRR,SG2+Blk} = (0.96N + 0.04)\Delta R_{raw} ; \Delta R_{Eff}^{NRR,SG6} = (0.42N - 0.27)\Delta R_{raw} \\ &\Delta R_{Eff}^{NRR,G} = (3.53\sigma + 0.02)\Delta R_{raw} \end{aligned} \quad . \quad (26)$$

Rayleigh ⟷ 1st zero in H(ω)
NRR ⟷ f_{-3dB} (3dB cutoff)

Clearly, as can be seen from Eq. (19), (20) and Eq. (26), some differences and similarities are evident from the comparison of the ERes estimated using the Rayleigh and NRR criterion.

1) A convenient way to summarize the results for both criteria and to understand their main differences, is to study the behavior of the frequency responses plotted in Fig. 18. From those curves, it looks that with the Rayleigh criterion, a common value of the ERes is obtained when the corresponding frequency responses of considered low pass filters share almost the same stop-band extension, while they can exhibit significant differences in the pass-band. The stop-band is easily defined by the frequencies above the value corresponding to the first zero in *H* for the filters with side lobes, (Schafer, 2011). For the Gaussian filter and the SG2 windowed filter (i.e. in case of frequency response with both a significant wider transition band and without side lobes of relevant magnitude), the stop-band starts could be taken at $\nu(H=0.1)$, like in the classical definition of the end of the transition band already used in Fig. 7. With the above definitions all the stop-band start values in the bottom plot of Fig. 18 are close each other, being comprise between about $7\cdot10^{-2}$ and about $9\cdot10^{-2}$.

2) On the contrary, the same ERes using the NRR criterion is found when the frequency responses have nearly the same pass-band (i.e. for $\nu<4\cdot10^{-2}$ in the upper plot of Fig. 18), taken as the region between the DC and the canonical definition of the cutoff frequency i.e. the frequency corresponding to -3db level, or $\nu_c= \nu(H=0.7)$. For this reason, NRR criterion tends to provide the same ERes for those smoothing filters sharing a common behavior at the lower frequencies.

As already evidenced, the distortion action of a smoothing filter is always present and its proper quantification is an outreach that should be assessed. In fact for a given amount of noise in a signal, the NRR tell us that there is a kind of saturation effect that is achieved when almost all the noise is removed. As a consequence the smoothing of a signal could not always leads to significant improvement: for example in Fig. 9 the layer structure is lost by the distortion action of the applied low pass filter. For this reason in a smoothing operation seems important to find the limit over which the (undesirable) distortion of an underlying input signal could become more relevant than the coupled (desirable) decrease of its noise level predicted by Eq. (21). Previous papers related to spectroscopic studies actually found this limit analyzing SG filters (Enke and Nieman, 1976; Ziegler, 1981; Gans and Gill, 1983; Rzhevskii and Mardilovich, 1994), and it would be interesting to apply their methods with the aim to optimize the effective resolution retrieval and more generally the whole lidar signal processing.

3.4 The Effective Resolution: the cutoff frequency

The considerations emerging from the analysis of Fig. 18 allow us to link both the approaches provided for the ERes estimation, to the cutoff frequency. The cutoff frequency associated to the frequency response of a smoothing filter can be used to estimate the effective resolution (Godin, 1987, 1999 ; Masci, 1999; Beyerle and McDermid, 1999; Leblanc et al., 2012). For this reason the following equation can be written:

$$\Delta R_{Eff}^{\nu_c} = \frac{\Delta R_{raw}}{\nu_c}, \qquad (27)$$

where ν_c is ...

Of course, the definition of effective resolution in Eq. (27) depends on the value chosen for ν_c and so on the actual pass-band (or bandwidth) definition. The Fig. 18 suggests that the proper ν_c value depends on the chosen criterion for the ERes evaluation. In order to try to find that proper value for the cutoff to be used it is useful write:

$$\nu_c = \frac{\Delta R_{raw}}{\Delta R_{Eff}}. \qquad (28)$$

In this way, once the ΔR_{Eff} is evaluated for a given low pass filter, Eq. (28) allows estimating the value of its frequency response at $\nu = \nu_c$. For example, with the NRR criterion, the cascade SG2·SG4 with N=25 will produce an $\Delta R_{Eff} \approx 25$ a.u. (ΔR_{raw}=1, from Eq. (23) or Eq. (26)), which implies a $\nu_c \approx 0.04$: thus, once estimated at that cutoff value, the frequency response relative to the above filter gives $H(\nu_c=0.04)\approx0.72$ (see Fig. 19, right panel). Indeed from Fig. 19, as far as the NRR criterion is concerned, it seems that for any given ΔR_{Eff} and for any smoothing filter (or at least within those analyzed), the values of the frequency responses at $\nu=\nu_c$ given by the Eq. (28) are quite constant and range on average between 0.65-0.72. For this reason, with the NRR criterion, if the ERes should be estimated via Eq. (27), the cutoff frequency defined as $\nu_c^{NRR}=\nu(H@-3db)$ appears the value to be chosen in this case.

Instead, for the Rayleigh approach, the ERes via Eq. (27) are close to those estimated via Eq. (19) and Eq. (20) if the cutoff frequency definition is taken as the half of frequency extension of the main lobe of the frequency response (Orfanidis, 2009), i.e. if ν_c is taken as the half of the lower frequency of the stop-band ν_{sb} (as defined in Sect. 3.3) or $\nu_c^{Ray}=\nu_{sb}/2$. This latter fact is in Fig. 20, where the values of $(2/\nu_{sb})$, plotted against ERes estimated with the Rayleigh criterion, are near the identity line for all the investigated low-pass filters. To summarize, the Eq. (27) can be rewritten for the NRR and the Rayleigh criterion as:

I would suggest to repeat Fig. 20 using the 1-st-zero criterion, instead.

$$\begin{aligned} \Delta R_{Eff}^{NRR} &\cong \frac{\Delta R_{raw}}{\nu_c^{NRR}} \cong \frac{\Delta R_{raw}}{\nu(H@-3db)} \\ \Delta R_{Eff}^{Ray} &\cong \frac{\Delta R_{raw}}{\nu_c^{Ray}} \cong \frac{2\Delta R_{raw}}{\nu_{sb}} \end{aligned} \tag{29}$$

These latter equations gives a general breath to the consideration done for the Fig. 18. Furthermore, the second formula in Eq. (29) provides a kind of general equation, or at least a rule of thumb, also for the ERes retrieval based on the Rayleigh criterion. Instead (operatively) the first one is not really needed because a general expression is already given by Eq. (23) for the ERes with the NRR criterion. It is good to precise that Eq. (29) has been obtained only using low pass filters studied in this work, and a further generalization to other filter types needs an additional analysis.

3.5 Smoothing kernels

[I would delete or reduce this section to a minimum. Non conclusive + up to a [illegible] point, speculative.]

Another approach to determine ERes, often used in the communities dealing with inverse problems applied to passive remote sensors, is based on the use of the retrieval kernels. Kernels account for the limited vertical resolution and for the sensitivity of the retrieval (in our case the smoothing is assumed as the applied retrieval) that decreases toward higher and lower altitudes depending on nadir or zenith e.g. in the case of microwave radiometers (MWR) pointing (Haefele et al., 2009). The peak of each kernel, at their associated range gate, provides the altitude of maximum sensitivity. Its full width at half maximum is typically interpreted as the value of ERes of the retrieval. A recent paper in literature (Illingworth et al., 2011) refers to the half width at half maximum as to the value of ERes of the retrieval. The calculation shown in this section agrees with the second formulation. The resolution derived from the kernel is similar in vertical shape to the resolution derived from error covariance matrices (Backus and Gilbert, 1968; Conrath, 1972).

As mentioned above, we apply the smoothing as a retrieval technique, leading to the equation:

$$y=Ax \qquad (30)$$

where x is the high resolution profile, y is the smoothed profile and A is the matrix identified by the smoothing filter. As described in Eq. (1), each smoothing procedure can be also seen as the convolution of the high resolution profile and a kernel, that, for example, in the case of a polynomial filter, is identified by the coefficients of the polynomial. Therefore the matrix A is identified by a matrix having as raw elements the coefficients of the polynomial.

To provide a quantitative comparison of the criteria mentioned above to determine the ERes with the kernels, in Fig. 21 the coefficients of the polynomial of a SG2 filter are reported for N=9 and N=19. If the half width of the two curves is calculated, a value of ERes equal to $7\Delta R_{raw}$ and $14\Delta R_{raw}$ is obtained. From Eq. (26) the same values of N are corresponding to $8\Delta R_{raw}$ and to $17\Delta R_{raw}$. From this comparison, it seems that the use of kernels provides an underestimation of the ERes with respect to that determined using the NRR criterion; the difference is larger with an increasing value of *N*. A deeper investigation is needed to learn more about this difference. Nevertheless, the use of kernels looks a promising choice to obtain a fast and automatic determination of ERes for lidar profiles, known the kernel of the applied smoothing filter. To the best of our knowledge, this is the first time this method is applied for determining the effective vertical resolution of lidar vertical profiles.

Summary and Conclusions

Make it shorter. | Say that your PARAMETRIC PLOTS / ~~SRT Ray and SRR~~ ΔR^{Ray} and ΔR^{NRR} are key contributions of the paper

NICE!

Noise ~~The removal~~ removal of noise ~~from~~ in lidar products via low-pass filters implies ~~corresponds~~ to suppress a certain amount of details in them. The smoothing operation also distorts both the magnitude and the spatial extension of the features contained in a profile. Moreover, the likely presence of several separated layers (of aerosol, ozone etc.) in a lidar profile puts ARISES the question if they are well resolved or not after the application of some kind of smoothing. Therefore, it is important to introduce the definition an effective resolution (ERes) associated to a lidar profile where a smoothing process is applied. The digital filter approach to the smoothing gives advantages respect the standard least-squares approach like:

Block I: Discuss about S6
II: " " Gaussian filters
Suggested

- A faster algorithms that are able to deal properly with the large dynamic range of a lidar signal, an interesting feature especially for the SCC algorithms (D'Amico et al., 2015).
- An easier statistical error analysis.
- Ready-to-use effective resolution definitions by an analysis of the impulse/frequency response.
- The many recipes to design efficient low pass filters in principle allow us to use the most suitable solution for any specific needs in lidar signal processing.

i.e.,

Concerning the latter point, several kinds of smoothing filters have been analyzed to also evidence the characteristics that could be useful to perform a choice among them. In fact the ERes estimation alone could not give a general guideline about why to choose a filter rather than another. Indeed, the effective resolution can be regarded as a kind of average parameter. For this reason, it that cannot take into account all the details, like the peculiar differences in the behavior in the whole frequency domain of the various filters: the analysis of other parameters is needed. If properly designed, the smoothing filters resulting from the cascade method applied to the Savitzky – Golay family seem a good choice when a lidar profile has to be smoothed. In fact it retains all the advantages of the SG smoothers while it reduces their main drawback i.e. the strong side lobe presence. Nevertheless, the cascade filters also show an enlargement of the transient zone. Other smoothing filters in our study, i.e. the Gaussian one and the SG2 with Blackman–type window, produce an even better suppression of the high frequency noise, but have a less accurate signal preservation at low frequencies and a more extended transition band. Before eventually enter in the estimation of the ERes, the possibilities given by the DSP are utilized both to further underline how relevant are the knowledge of impulse/frequency response and to solve a practical problem in lidar studies, i.e. how to calculate the lidar ratio being sure that both the required aerosol extinction and backscatter profiles have the same resolution. Then, an operative ERes estimation was determined by taking into account:

- The Rayleigh criterion, which highlights our ability to resolve (or not) close layers.
- The NRR criterion, which highlights the amount of (high frequency) noise reduction of low pass filters, and that can be seen as a measure of the spatial scales removed from a signal.

The NRR criterion underlines that with smoothing filters only the high frequency noise is efficiently removed. In fact the presence of low frequency noise will remain almost unchanged and then will still affect lidar products. The application of both the criteria, brings to a simple linear relationship between the effective resolution and the filters parameters. Different results with different criteria for the same filter have been found. Anyhow the discrepancies are limited to a maximum of ~30% in case of plain SG or Gaussian filter, while for other filters they are less pronounced (<20%) or practically not particularly relevant (<5% for dSG2 case). The investigation of the differences between the two criteria is evidenced by the analysis of the frequency responses that corresponds to a common ERes value for various smoothing filters. This latter approach permits to underline that:

Revise: I'd say in term of f_{-3dB} and $f_{FIRST ZERO}$

- The effective resolutions obtained with the Rayleigh are similar for those filters that share a comparable stop-band, while in the pass-band they could behave differently.
- The effective resolutions estimated with the NRR criterion are similar for different filters that share a similar behavior in pass-band, whose extension results also comparable.

Though feasible for any given filter, the ERes estimation based on the Rayleigh criterion shows some drawbacks and appears more elaborated respect the application of the NRR criterion. In fact for the NRR criterion, a ready-to-use equation to estimate the effective resolution was found which is directly applicable to any given ~~smoothing~~ filter. In this case the only needed input is the impulse response of the employed filter, which is always available (or determinable). For this reason the NRR approach to the ERes estimation would appears more suitable to be used as a standard for a generalized application. Moreover, the NRR criterion implies the higher uniformity in the pass-band for different filters with a common ERes, and generally the signal to preserve has interesting features that lay mainly in that portion of the frequency axis. Nevertheless, the results about the calculation of the ERes by the analysis of the 3dB cutoff frequency, allow one to obtain also for the Rayleigh criterion a specific general equation which is based only on the knowledge of the frequency response of the applied smoothing filter. Furthermore, the Rayleigh criterion measures the ability (or not) to resolve close layers, which could be a valuable feature in lidar studies. Additionally, the ERes estimated with this criterion is significantly more conservative respect the NRR criterion, at least for plain SG smoothers. Regarding the derivative process, it is fairly

slang

common the use of the SG2 low-pass first-derivative filter within EARLINET community (Pappalardo et al., 2004), whose embedded low pass is denoted with dSG2. This latter filter allows one to obtain nearly the same ERes estimation regardless the criterion chosen and, additionally, the obtained results are consistent with those given in Pappalardo et al. about the same filter. The dSG2 exhibits a quite similar behavior of a Gaussian filter with similar ERes (from the NRR point of view) in almost all the pass/transition band. Moreover, the Gaussian filters have quite better stop-band features (the absence of significant side lobes) and provide an easier way to perform correct lidar ratio calculations. Those considerations bring to the conclusion that it seems recommendable the employment of the Gaussian low-pass derivative filter to retrieve the extinction profile (and more generally anytime the first derivative of a signal is required) as long as the choice is between this filter and the widely used SG2 low pass first derivative filter. An alternative approach to the ERes assessment has also been proposed, i.e. the one based to the smoothing kernels, which produce results that are consistent with the NRR criterion although further insights are required. Anyhow, it appears a promising method that could be further developed to look at the ERes problem from a new point of view. Moreover within the lidar community, there are other approaches on the numerical derivative problem that have been proven to be effective and also other methods able to provide alternative and reasonable ERes definitions: however, the scope of this paper is not to compare all the smoothing filters applied in literature to deal with lidar profiles, but instead to provide a methodology to assess the ERes. Nevertheless, a more exhaustive comparison with other approaches such as ... (OUTLINE) for the evaluation of the ERes and smoothing filters will be likely done in future in the frame of EARLINET activities. Other promising directions for the future developments of this study could be try to obtain a more general and possibly unique rule for the effective resolution estimation and also to pursue the objective of an improvement of the lidar signal analysis. The latter objective could be achieved by means of both a deeper exploitation of DSP theory and the application and development of the smoothing optimization methods already

(18) Scope is - from what you have now -
1. Parametric study on SG and Gaussian filter
2. ERes
Both (1) and (2) are valuable!

underlined by chemical spectroscopy papers mentioned in the text (for example, Gans and Gills, 1983).

Acknowledgments. The financial support for EARLINET by the European Union under grant RICA 025991 in the Sixth Framework Programme is gratefully acknowledged. Since 2011, EARLINET has been integrated into the ACTRIS Research Infrastructure project, supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 262254.