

Real-time infrared gas detection based on an adaptive Savitzky–Golay algorithm

Jingsong Li¹ · Hao Deng¹ · Pengfei Li¹ · Benli Yu¹

Received: 14 December 2014 / Accepted: 28 April 2015 / Published online: 8 May 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Based on the Savitzky–Golay filter, we have developed in the present study a simple but robust method for real-time processing of tunable diode laser absorption spectroscopy (TDLAS) signals. Our method was developed to resolve the blindness of selecting the input filter parameters and to mitigate potential signal distortion induced in digital signal processing. Application of the developed adaptive Savitzky–Golay filter algorithm to the simulated and experimentally observed signals and comparison with the wavelet-based de-noising technique indicate that the newly developed method is effective in obtaining high-quality TDLAS data for a wide variety of applications including atmospheric environmental monitoring and industrial processing control.

1 Introduction

Development of TDLAS has continued for several decades since the first demonstration of high-resolution spectroscopy with lead–tin telluride diode laser by Hinkley et al. [1]. Advantages of TDLAS include high sensitivity and selectivity, rapid response speed, and nondestructive detection. Sensors based on this technique can be tailored to determine parameters such as temperature, pressure, species concentration, or velocity. TDLAS has been extensively used in atmospheric environmental monitoring, industrial process control, medical diagnosis, military,

and public safety fields [2–4]. However, the spectral data always contain noise and disturbance signals (i.e., overlap effect), so it is significant to perform data preprocessing to obtain highly precise and accurate data. As a result, many experimental schemes or techniques have been proposed for sensitivity improvement and resolution enhancement [5–8], generally classified into two categories: software- and hardware-based techniques.

Digital filtering techniques based on software for online noise reduction or off-line data processing of recorded spectra are a better choice when temporal resolution and lower system cost are priorities. The key point is the optimization and choice of its input parameters when applying the selected digital signal processing techniques. Multi-signal averaging is a relatively simple and widely adopted method for noise suppression; however, it is time consuming and only adaptive to white noise [9]. Generally, derivative calculation is used as a resolution enhancement technique to facilitate the detection and location of poorly resolved components in the complicated spectra; however, numerical computation of the higher-order derivatives has also computational (time) costs. Among various filter techniques, wavelet transform (WT) is a powerful signal de-noising technique [14], but this method depends on more parameters, for example, mother wavelet type, thresholding method, threshold estimation, and decomposition level. Recently, the Savitzky–Golay (S–G) smoothing filter has been shown to be especially attractive since both the smoothed signal and the derivatives can be calculated in a single step [11–14], and only two parameters must be set, i.e., the width of the smoothing window and the degree of the smoothing polynomial.

Analogous to other digital signal processing techniques, the effectiveness of the S–G filter is found to be strongly dependent on the window size. Selection of the appropriate

✉ Jingsong Li
jingsong_li@ahu.edu.cn; ljs0625@126.com

¹ Key Laboratory of Opto-Electronic Information Acquisition and Manipulation of Ministry of Education, Anhui University, Hefei 230039, China

window size is essential for achieving the correct trade-off between reducing noise and avoiding bias [15]. For example, Edwards and Willson [16] have found that the optimum width of the smoothing array is 0.7 times the full width at half maximum (FWHM) of the narrowest Gaussian line of their spectra. In a similar study considering Lorentzian- as well as Gaussian-shaped lines, Enke and Nieman [17] concluded that the best signal-to-noise ratio (SNR) enhancement from a single-pass (quadratic–cubic) smoothing occurs for a smoothing array that is twice as wide as the FWHM of the peak being smoothed. In the study of Madden’ work [18], it is found that optimum width can sometimes be greater than 25 points for spectral data with 512 sampling points. Therefore, the optimum filter window width will depend on the signal features and the criteria set by the user. Moreover, an approach based on comparing the fitting residuals with the noise of the instrument was reported for selecting the optimal window size of the S–G algorithm [19]. However, in the case of non-stationary signals, the optimal window size will vary with the dynamics of the signal. Addressing this issue, the S–G filter with varying window size based on evaluation of the residuals of the smoothed data (with Gaussian lineshape) in the local region was proposed by Browne et al. [20]. This strategy was shown to be superior to fixed window S–G smoothing for a test signal at various SNR for noise removal. In the case of trace gas detection using TDLAS, the peak height or the integrated absorbance area of spectral signal is directly proportional to the gas concentration of the targeted species. Therefore, signal preservation is an important quality indicator in signal preprocessing, and this issue is often overlooked. In this work, a study of the simulated and measured TDLAS data (with Gaussian, Lorentzian, and Voigt profiles) by an adaptive S–G filter with varying window size has been conducted, in order to guide TDLAS signal preprocessing.

2 Savitzky–Golay smoothing filter

The S–G filtering technique is well known for smoothing data, so it will not be described in detail. Only some terminology and two key points considered in this work will be discussed. The main idea is similar to a moving average, but instead of just averaging the sampling points, it performs a least-squares-fit convolution procedure. The basic method of the S–G algorithm comprises the following steps: (i) data interval is selected (i.e., window size), (ii) a low-order polynomial function is fitted to the selected data interval, and (iii) the smoothed data point at the center of the selected interval is calculated from the polynomial coefficients. This smoothing process is repeated after shifting the analysis interval to the right by one sampling interval,

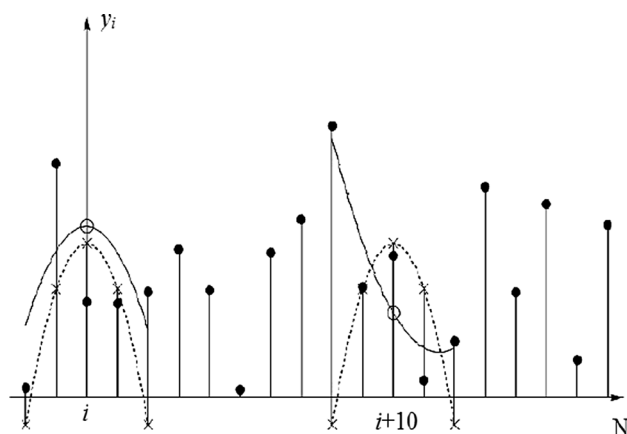


Fig. 1 Illustration of least-squares smoothing by locally fitting a low-order polynomial (solid line) to five input samples: dot denotes the raw input samples, circle denotes the least-squares smoothed samples, and x denotes the effective impulse response samples. The dotted line denotes the polynomial approximation to centered unit impulse

as depicted in Fig. 1. More detailed discussions of least-squares-fit smoothing can be found in the original paper by Savitzky and Golay [21] and the corrected versions [17, 22] as well as a review paper by Willson and Edwards [23].

Generally, the criterion to quantitatively illustrate the effectiveness of the de-noising operation is the SNR improvement, defined as follows:

$$\text{SNR(dB)} = 10 \log_{10} \left(\frac{\text{std}(\text{Signal}_{\text{noise-free}})}{\text{std}(\text{Signal}_{\text{noise-free}} - \text{Signal}_{\text{S-G-denosed}})} \right) \quad (1)$$

where std refers to the standard deviation, $S_{\text{noise-free}}$, and $S_{\text{S-G-denosed}}$ are the ideal simulated spectral signal, and S–G-filter-de-noised spectral signal, respectively. However, for real-world applications, the optimal filtering parameters cannot be directly determined from the SNR definition, since the real signal (i.e., noise-free signal) and noise source are completely unknown. To address this challenge, we proposed a varying window S–G filtering by integrating two additional criteria for TDLAS signal processing. The first criterion is to introduce a “real signal” or “noise-free signal” referred to “PolyFit” which is generated by fitting a polynomial function (initialization: polynomial order = 5, window size = 7) to a small segment (typically 50 sampling points) near the absorption peak of the raw signal, as shown in Fig. 2. The multiple linear regression analysis method is used to calculate the correlation coefficient R between the “PolyFit” and the same segment in the S–G-filter-smoothed data, instead of using SNR for assessing the optimal filtering parameters (in case of experimental data). Indeed, this condition is valid for noise reduction, while not credible for signal preservation. The second criterion is to employ a threshold “Th” defined as the difference of peak

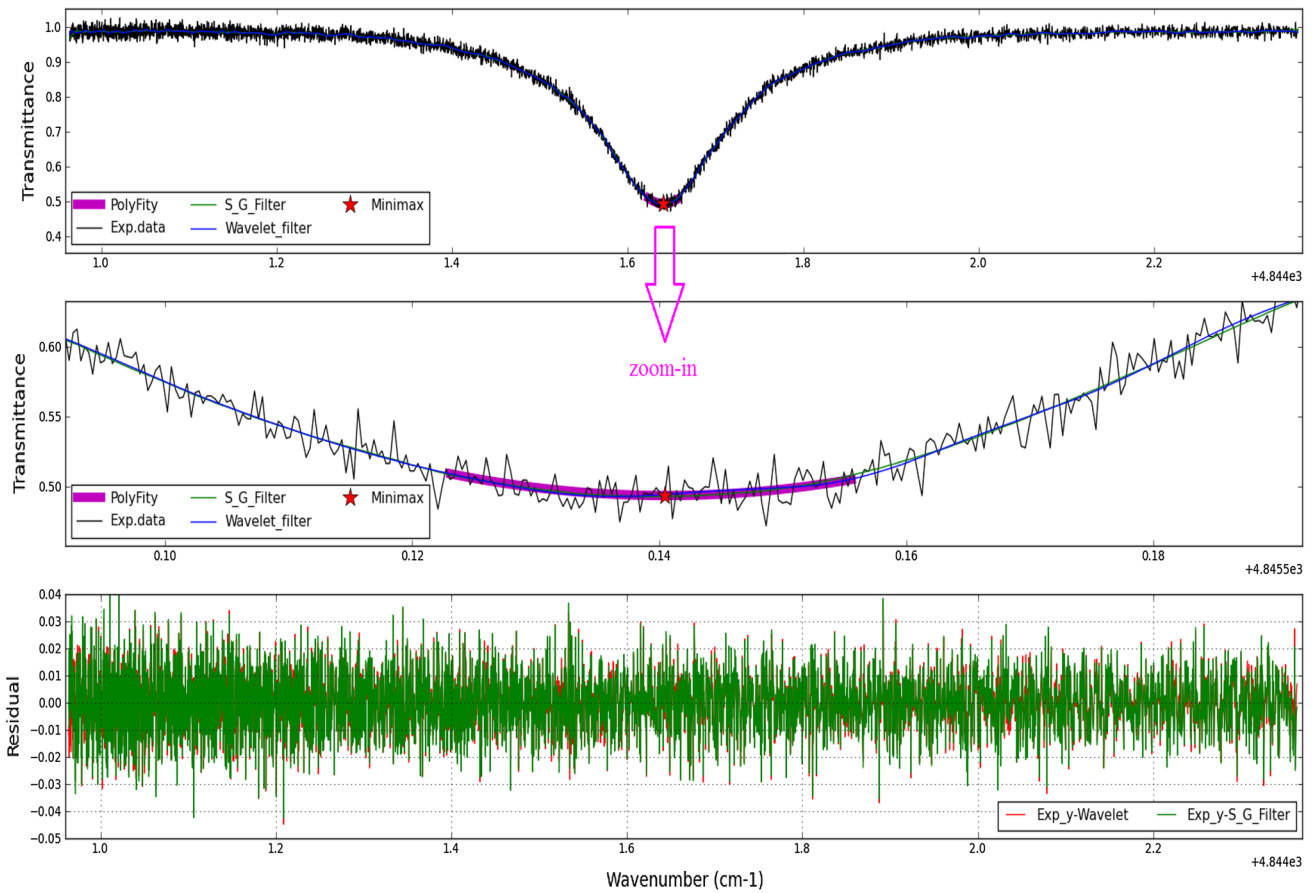
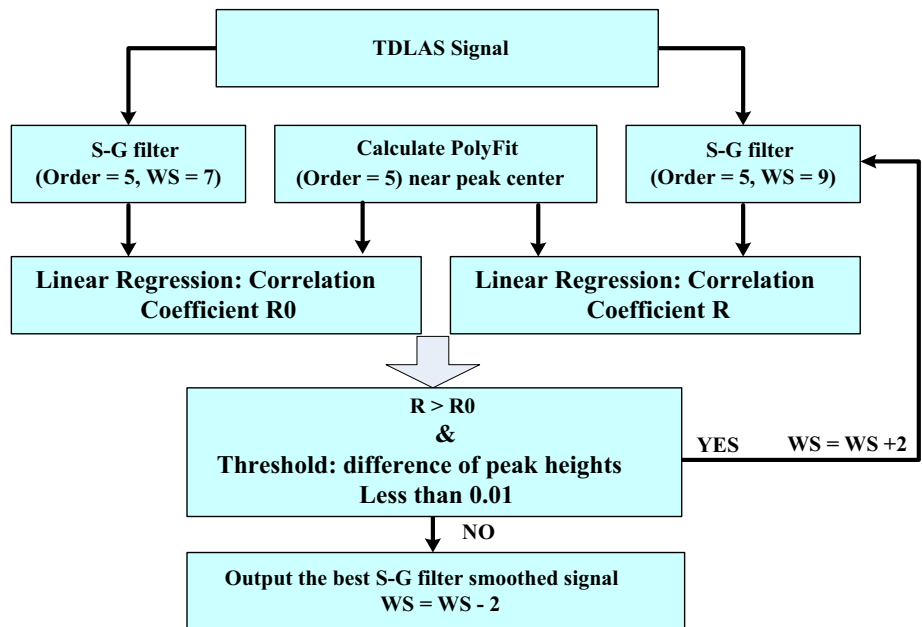


Fig. 2 Typical example of experimentally determined CO₂ absorption spectrum and corresponding results by applying S–G filter and wavelet de-noising techniques (for details see text)

Fig. 3 Flowchart of the newly developed adaptive S–G filter algorithm (for details see text)



heights between “PolyFit” and the S–G filtering smoothed data, in order to optimize filtering parameters without excessive signal distortion. The flowchart of the adaptive S–G filter algorithm is shown in Fig. 3. Note that the window size must be an odd integer number, and the polynomial order must be less than window size.

3 Parameter optimization by simulation

In order to understand the dependence of the S–G filter on its input parameters (i.e., window size and polynomial order) as well as other effects such as sampling points and signal profiles, we have performed a large numbers of spectral simulations. The simulated datasets were modeled by considering a range of undesired spectral anomalies and variations that can often occur in measured spectra, such as baseline variations, noises, and pressure effects. We first evaluated the S–G filter for the synthetic spectral signals, modified with varying magnitudes of random noise and sampling points. A computer program has been written in the numerical script language Python for the computations and signal simulations. The CO₂ spectroscopic parameters were used for simulation, which were extracted from HITRAN database [24], are compiled in Table 1. A set of given experimental conditions, such as temperature, pressure, gas concentration, and optical path length, was considered.

First, various spectral absorption signals with 1024 sampling points and different SNR have been simulated with partial signals, and the corresponding S–G filter-smoothed results are presented in Fig. 4. It can be seen that the window size must be chosen appropriately in order to preserve peak height. For a given polynomial degree, smaller window sizes will not give the best SNR; higher window size will produce a smoother result but could introduce bias of signal preservation, which in turn induces measurement errors of gas concentration. The SNR enhancement factor and the best window size as a function of polynomial order are shown in Fig. 5. This figure illustrates that the higher the polynomial order used in the S–G filter, the higher the window size needed for achieving the best SNR. On the other hand, we can see that the SNR enhancement factors are almost same for polynomial orders between 2 and 8. Furthermore, we evaluate the S–G filter by applying to the simulated signals with different sampling points, as presented in Fig. 6. Note that we found the larger the number

of total sampling points, the higher the SNR enhancement factor achieved for the same noise level. Therefore, the noise level can be more effectively reduced by increasing the number of sampling points to which the S–G filter is applied. However, one has to compromise between noise reduction and temporal resolution in real-world applications. Moreover, we found that the proposed algorithm can also construct an optimal calibration model for TDLAS spectra with different background structural characteristics (linear or nonlinear baseline drift) [25].

4 Experimental application

From the simulations discussed above, it is recommended to set the polynomial order of the S–G filter in the range of 2–8, while the window size is the primary factor strongly that limits the filtering efficiency. In order to verify this conclusion and use of the algorithm for real measured signals, various spectral data are recorded by our TDLAS system. For creating the “PolyFit,” a polynomial (order = 5, used throughout this section) function was fitted to a small segment (50 sampling points) of the original signal (4096 sampling points) near the absorption peak. The linear correlation coefficients R calculated between “PolyFit” and the same segment in the S–G filter-smoothed results are calculated to replace SNR for assessing the optimal filtering window size. In theory, the higher the R values, the smoother the S–G-filtered results. Considering the second criterion, a threshold of 0.01 is typically selected. A comparison with powerful wavelet-based de-noising technique is also conducted. As demonstrated in Fig. 2, the best S–G-filter-smoothed result is comparable to that obtained from the best wavelet filtering (where Stein thresholding policy, wavelet db10, and decomposition level 6 are used). Finally, Fig. 7 presents the values of R^2 and difference of absorption peak heights as a function of window size for polynomial order between 1 and 8. Obviously, the R^2 and difference of line peak heights shows inverse trend with the optimal window size. When the difference of line peak heights overflows the threshold, the R^2 presents decline trend or abrupt change. Figure 8 shows the parameters determined from the developed S–G filter algorithm as a function of polynomial order. Here, the SNR enhancement factor was directly calculated from the ratio of standard deviation of the segments containing no absorption baseline in the unfiltered

Table 1 Summary of spectroscopic parameters of CO₂ line pair studied in this work, data are taken from HITRAN2012 database [23]

| Species | Transition frequency (cm ⁻¹) | Line intensity at 296 K [cm ⁻¹ /cm ⁻² molecule] | Air-broadened half width (cm ⁻¹ /atm) | Self-broadened half width (cm ⁻¹ /atm) | Lower state energy (cm ⁻¹) | Coefficient of T dependence of air-broadened half width |
|-----------------|--|---|--|---|--|---|
| CO ₂ | 4845.636997 | 2.125E-22 | 0.0821 | 0.109 | 42.9225 | 0.71 |

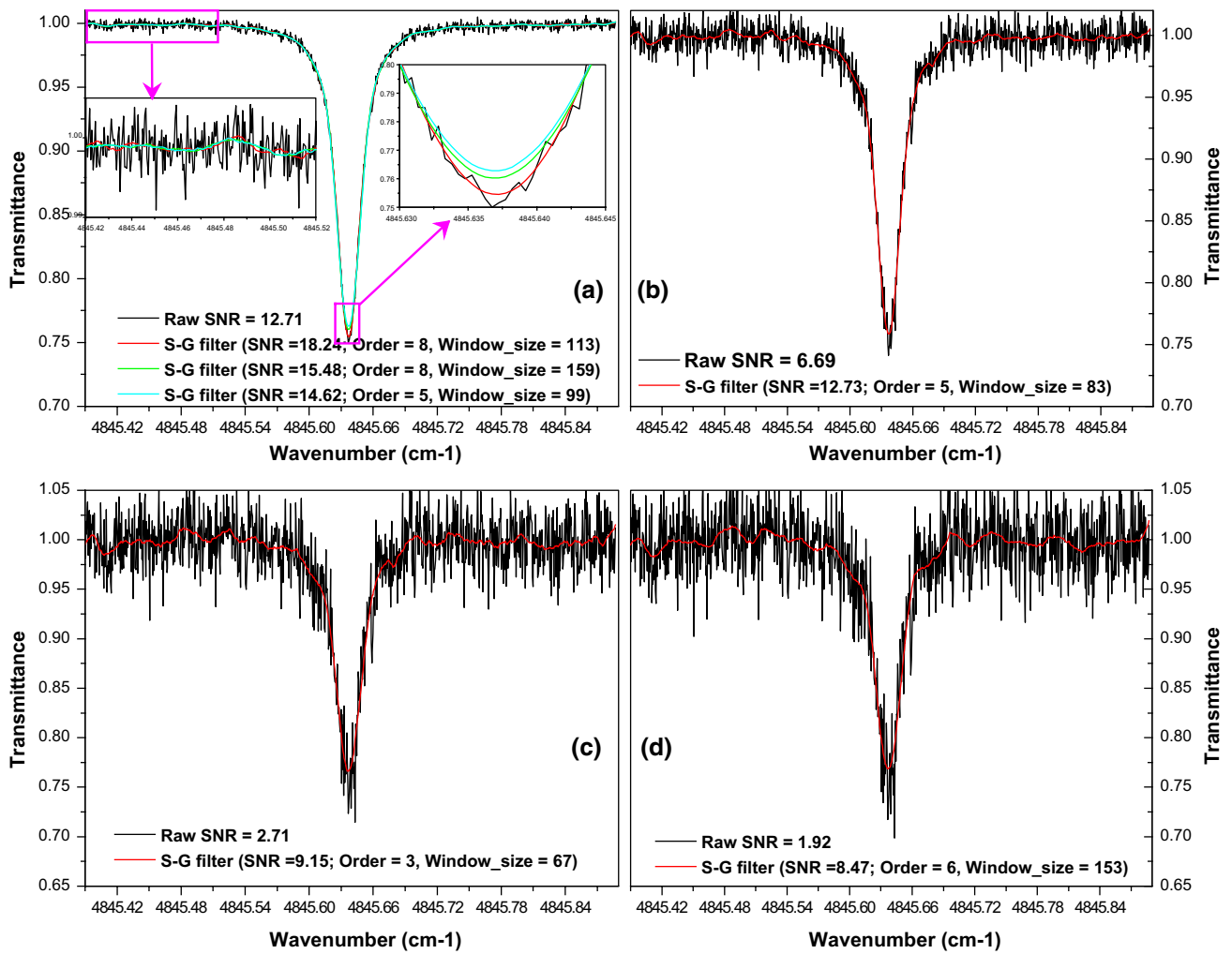


Fig. 4 Simulated spectra with different SNR (sampling points = 1024) and the corresponding S–G-filter-smoothed results

Fig. 5 SNR enhancement factor and best windows size as a function of polynomial order for the S–G filter applying to the simulated signals with different SNR (number of sampling points = 1024)

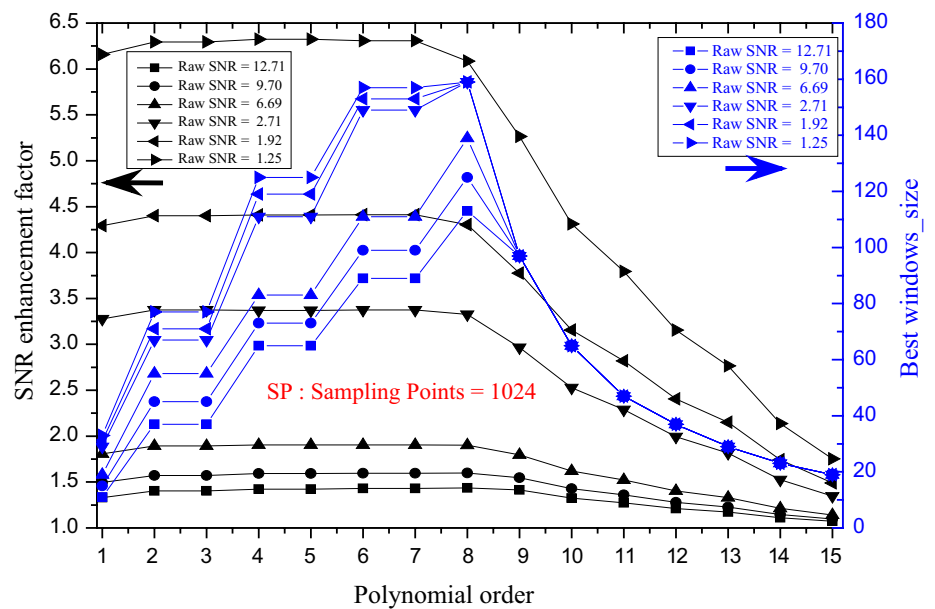


Fig. 6 SNR enhancement factor and best windows size as a function of polynomial order for the S-G filter applying to the simulated signals with different sampling points (raw SNR = 6.69)

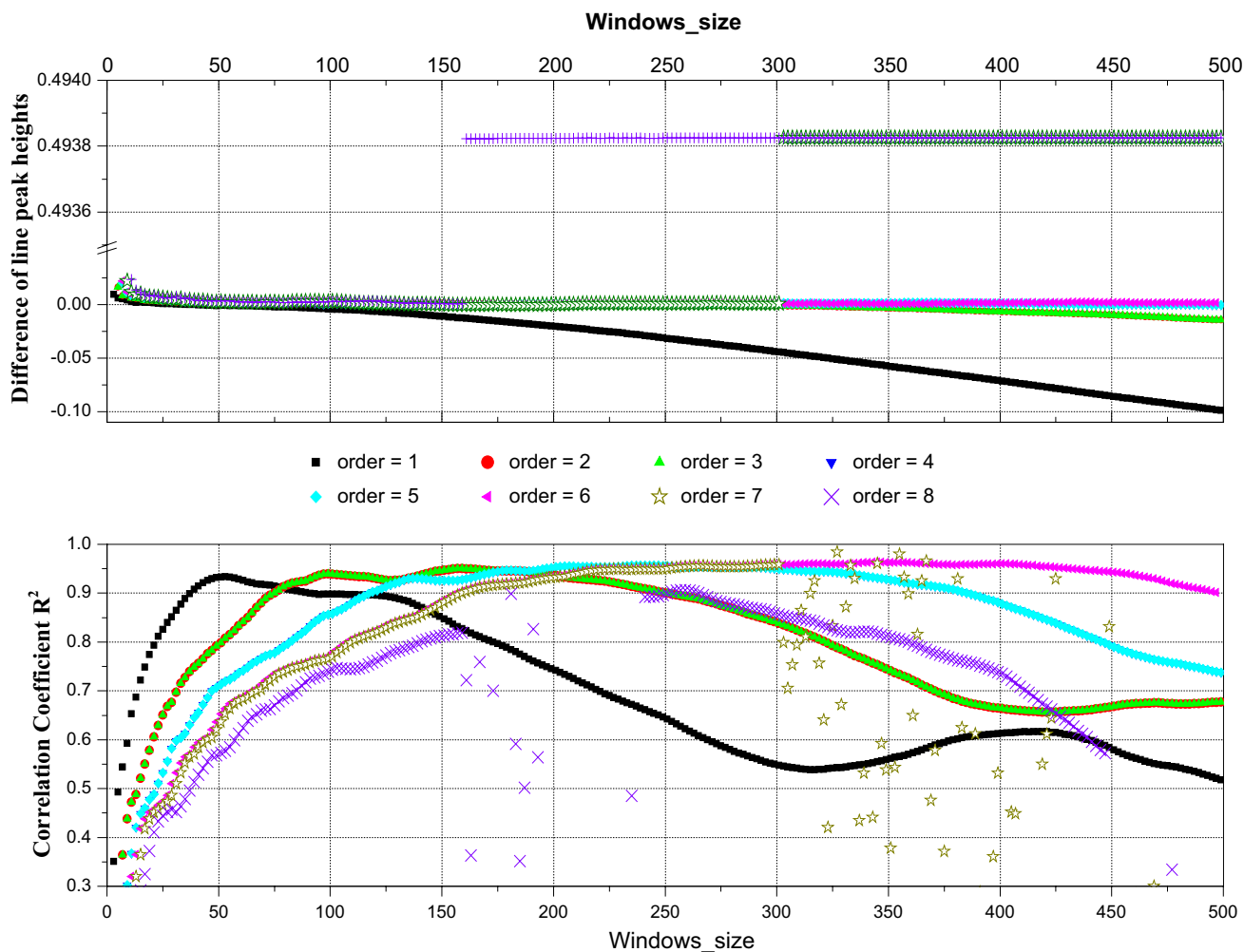
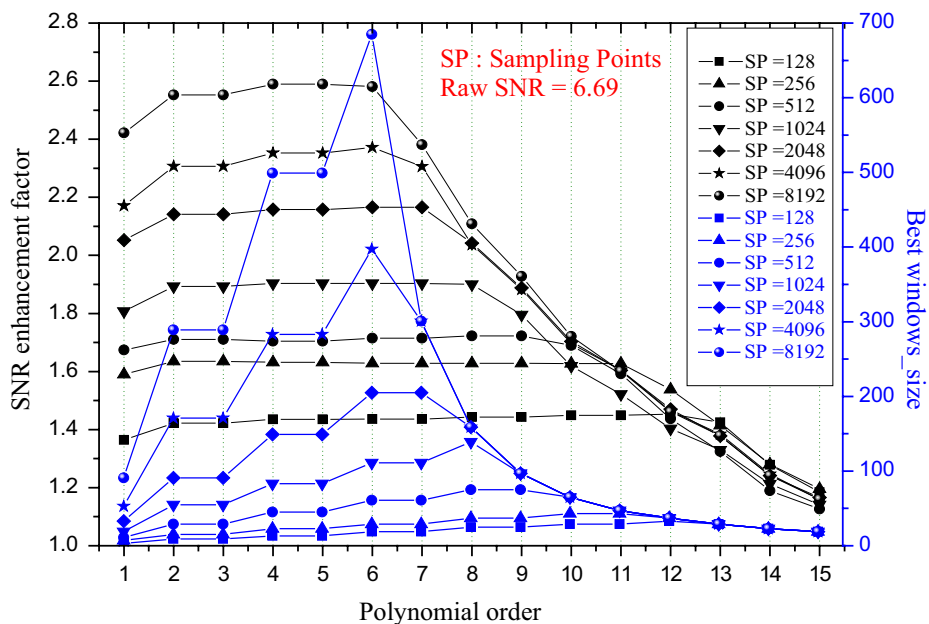


Fig. 7 Correlation coefficient R^2 and difference of line peak heights as a function of windows size

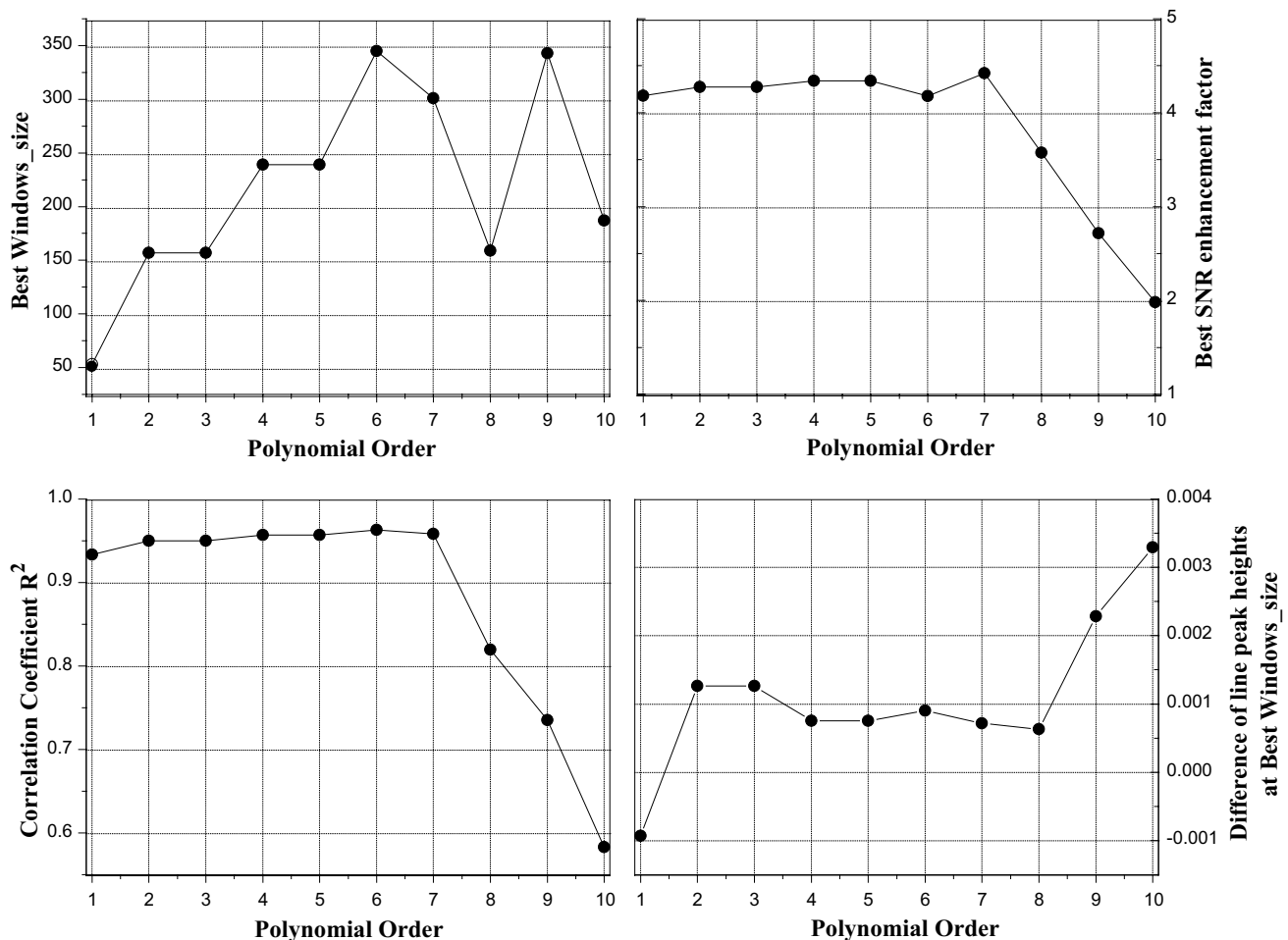


Fig. 8 Parameters determined in the developed S–G filter as a function of polynomial order. SNR is defined as the ratio of standard deviation of the segments containing no absorption in unfiltered and filtered signal

and filtered signals. The S–G-filtered results show that the highest R^2 and the best SNR enhancement factor occurred at polynomial order between 2 and 7, while the difference of absorption peak heights are within -0.001 and 0.0035 for each optimal window size, which are much less than the selected threshold of 0.01 . These results confirmedly prove that the developed algorithm is reliable for processing our experimentally measured TDLAS signals.

In order to further evaluate the suitability of the developed adaptive algorithm suitable for absorption spectra with different lineshapes, series of experimental spectra (CO_2 concentration around 1.5%) were recorded at different pressures (between a few mbar and 1 bar). We still use the standard deviation of the segments containing no absorption baseline in unfiltered and filtered signal to denote the noise level. The results are demonstrated and compared in Fig. 9. The statistical mean values are also provided as insets in the figure. Overall, the SNR

enhancement factor of 5.5 and 4.7 can be calculated from wavelet filter and the S–G filter, respectively. Based on the study of Chen et al. [11], the developed algorithms have finally been applied to a time series of CO_2 concentrations datasets. As it can be seen from Fig. 10 (upper panel), measurement precisions have been significantly improved with standard deviations of 1.01 , 0.18 , and 0.16 from raw measurements, the output of the S–G filter, and output of the wavelet filter, respectively. The Allan variance in the lower panel shows an optimal averaging time of about 200 s for the present system. The measurement precision improvement by the developed S–G filter and wavelet filter corresponds to a precision level that can be obtained by conventional 40 -s averaging. Overall, the wavelet filter demonstrated a higher ability to remove noise, but the method requires more parameters to be specified, for example, mother wavelet type, thresholding policy, threshold estimation, and decomposition level.

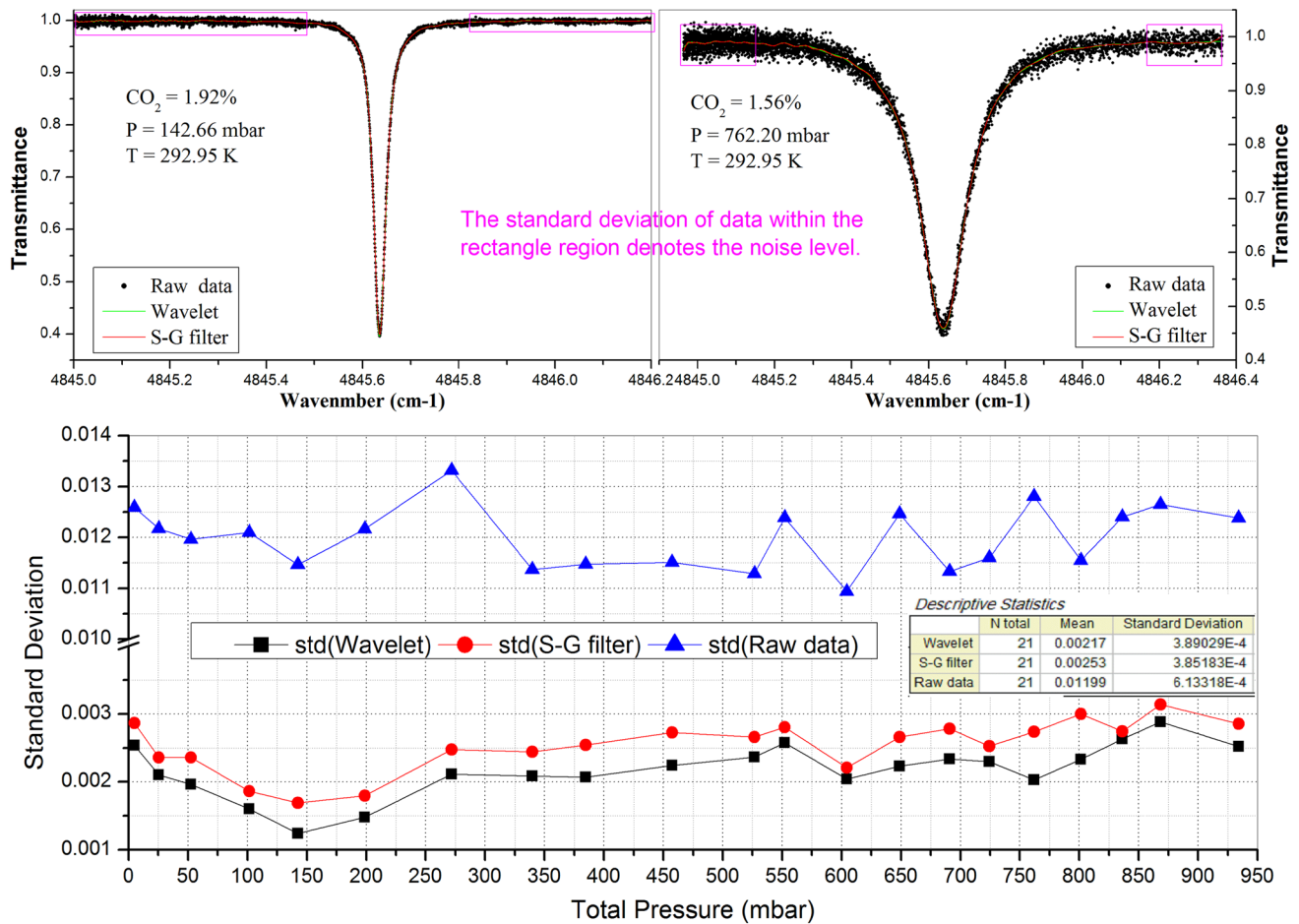


Fig. 9 Comparison of noise level (see definition at upper panel) before and after applying the S–G filter and wavelet de-noised technique to experimental spectra recorded under different pressure

On the other hand, the S–G filter shows great flexibility and has great potential for time series datasets with fast response, which are particularly attractive for TDLAS and other laser spectroscopy applications.

5 Conclusion

In this paper, we have presented a simple but robust method based on the S–G filter to smooth out noise present in our TDLAS system without distorting signals. By applying the newly developed method to both simulated and experimental spectral signals, we found that the window size is the primary factor that limits the smoothing efficiency. Comparing the results with those from the powerful wavelet transform-based filter, the developed adaptive S–G filter shows the following four advantages: (i) it can reconstruct high-quality TDLAS signal by setting only

two parameters in the S–G filter. Our results suggest that the optimal polynomial order is between 2 and 8, which is robust in most cases, while the best window size depends on the optimal polynomial order and the dynamics of signal and noise; (ii) it is very simple in theory and easy to implement because most commercial software such as ORIGIN and MATLAB include the S–G filter in their function library; (iii) it can be applied to spectral signals with any lineshape (e.g., Gaussian and Lorentzian), and there are no restrictions on the scaling of TDLAS datasets; (iv) the time cost for searching the optimal window size and outputting the best S–G-filtered result is superior than wavelet filtering technique. For these reasons, we anticipate that the developed method can be further applied to real-time smooth TDLAS spectral signals and time series concentration datasets for a wide variety of applications including atmospheric environmental monitoring and industrial processing control.

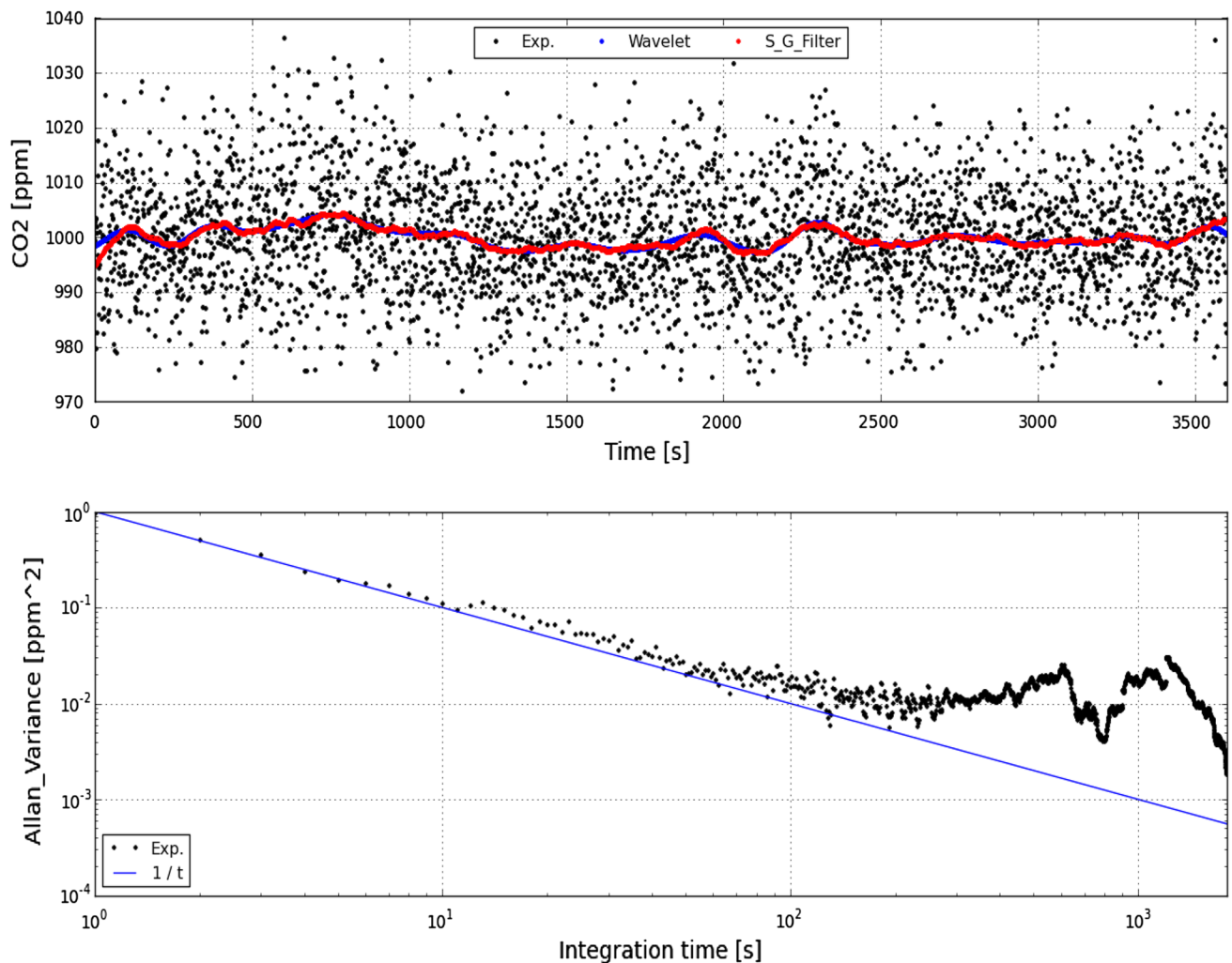


Fig. 10 Upper panel shows raw measurements of time series CO₂ concentrations with 1-Hz sampling rate and the corresponding S–G filter and wavelet filter output. Lower panel presents the Allan variance plot of raw data

Acknowledgments This work was supported in part by Anhui University personnel recruiting project of academic and technical leaders (Grant No. 10117700014), the Natural Science Fund of Anhui Province under Grant 1508085MF118, the National Natural Science Foundation of China under Grant 61440010, and the key Science and Technology Development Program of Anhui Province under Grant 1501041136. We thank two anonymous reviewers and editors for their useful comments on the manuscript. Special thanks go to Prof. A.P. Yalin (Colorado State University) for his helpful discussion and careful reading of the manuscript.

References

1. E.D. Hinkley, *Appl. Phys. Lett.* **15**, 351 (1970)
2. A. Fried, B. Henry, B. Wert, S. Sewell, J.R. Drummond, *Appl. Phys. B* **67**, 317 (1998)
3. G. Durry, J.S. Li, I. Vinogradov, A. Titov, L. Joly, J. Cousin, T. Decarpenterie, N. Amarouche, M. Liu, B. Parvitte, O. Korablev, M. Gerasimov, V. Zéninari, *Appl. Phys. B* **99**, 339 (2010)
4. J.S. Li, G. Durry, J. Cousin, L. Joly, B. Parvitte, P.H. Flamant, F. Gibert, V. Zéninari, *J. Quant. Spectrosc. Radiat. Transf.* **112**, 1411 (2011)
5. J.S. Li, B. Yu, W. Zhao, W. Chen, *Appl. Spectrosc. Rev.* **49**, 666 (2014)
6. P. Werle, *Spectrochim. Acta A* **54**, 197 (1998)
7. L. Zhang, G. Tian, J. Li, B. Yu, *Appl. Spectrosc.* **68**, 1095 (2014)
8. P. Werle, P. Mazinghi, F.D. Amato, M. De Rosa, K. Maurer, F. Slemr, *Spectrochim. Acta Part A* **60**, 1685 (2004)
9. P. Werle, R. Mucke, F. Slemr, *Appl. Phys. B* **57**, 131 (1993)
10. J.S. Li, U. Parchatka, H. Fischer, *Appl. Phys. B* **108**, 951 (2012)
11. J. Chen, P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, L. Eklundh, *Remote Sens. Environ.* **91**, 332 (2004)
12. M.A. Czarnecki, *Appl. Spectrosc.* **69**, 67 (2015)
13. R. Jiménez, M. Taslakov, V. Simeonov, B. Calpini, F. Jeanneret, D. Hofstetter, M. Beck, J. Faist, H. Van Den Bergh, *Appl. Phys. B* **78**, 249 (2004)
14. J. Luo, K. Ying, P. He, J. Bai, *Digit. Signal Process.* **15**, 122 (2005)
15. G. Glannelli, O. Altamura, *Rev. Sci. Instrum.* **47**, 32 (1976)
16. T.H. Edwards, P.D. Willson, *Appl. Spectrosc.* **28**, 541 (1974)

17. C.G. Enke, T.A. Niernan, *Anal. Chem.* **48**, 705A (1976)
18. H.H. Madden, *Anal. Chem.* **50**, 1383 (1978)
19. G. Vivó-Truyols, P.J. Schoenmakers, *Anal. Chem.* **78**, 4598 (2006)
20. M. Browne, N. Mayer, T.R.H. Cutmore, *Digit. Signal Process.* **17**, 69 (2007)
21. A. Savitzky, M.J.E. Golay, *Anal. Chem.* **36**, 1627 (1964)
22. J. Steinier, Y. Termonia, J. Deltour, *Anal. Chem.* **44**, 1906 (1972)
23. P.D. Willson, T.H. Edwards, *Appl. Spectrosc. Rev.* **12**, 1 (1976)
24. L.S. Rothman, I.E. Gordon, A. Barbe, D.C. Benner, P.F. Bernath, M. Birk et al., *J. Quant. Spectrosc. Radiat. Transf.* **110**, 533 (2009)
25. J.S. Li, B. Yu, H. Fischer, *Appl. Spectrosc.* **69**, 496 (2015)