AIP | Review of Scientific Instruments

UMEL: A new regression tool to identify measurement peaks in LIDAR/DIAL systems for environmental physics applications

M. Gelfusa, P. Gaudio, A. Malizia, A. Murari, J. Vega, M. Richetta, and S. Gonzalez

Citation: Review of Scientific Instruments **85**, 063112 (2014); doi: 10.1063/1.4883184
View online: http://dx.doi.org/10.1063/1.4883184
View Table of Contents: http://scitation.aip.org/content/aip/journal/rsi/85/6?ver=pdfcov
Published by the AIP Publishing

UMEL: A new regression tool to identify measurement peaks in LIDAR/DIAL systems for environmental physics applications

M. Gelfusa,[1] P. Gaudio,[1] A. Malizia,[1] A. Murari,[2] J. Vega,[3] M. Richetta,[1] and S. Gonzalez[3]
[1]*Associazione EURATOM-ENEA - University of Rome "Tor Vergata", Roma, Italy*
[2]*Consorzio RFX-Associazione EURATOM ENEA per la Fusione, I-35127 Padova, Italy*
[3]*Asociación EURATOM/CIEMAT para Fusión. Avda. Complutense, 40. 28040 Madrid, Spain*

(Received 8 April 2014; accepted 1 June 2014; published online 18 June 2014)

Recently, surveying large areas in an automatic way, for early detection of both harmful chemical agents and forest fires, has become a strategic objective of defence and public health organisations. The Lidar and Dial techniques are widely recognized as a cost-effective alternative to monitor large portions of the atmosphere. To maximize the effectiveness of the measurements and to guarantee reliable monitoring of large areas, new data analysis techniques are required. In this paper, an original tool, the Universal Multi Event Locator, is applied to the problem of automatically identifying the time location of peaks in Lidar and Dial measurements for environmental physics applications. This analysis technique improves various aspects of the measurements, ranging from the resilience to drift in the laser sources to the increase of the system sensitivity. The method is also fully general, purely software, and can therefore be applied to a large variety of problems without any additional cost. The potential of the proposed technique is exemplified with the help of data of various instruments acquired during several experimental campaigns in the field. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4883184]

I. THE RELEVANCE OF LIDAR AND DIAL MEASUREMENTS FOR ATMOSPHERIC PHYSICS

Lidar (Light Detection and Ranging) and Dial (Differential Lidar) techniques have become well established laser remote sensing methods for atmosphere exploration.[1] They are often used to probe almost any level of the atmosphere and to acquire information necessary to validate theoretical models about different topics of atmospheric physics. They can also be deployed for environment surveying, by monitoring particles and molecules. In particular in the Dial technique, the laser transmitter emits light pulses at two closely adjacent wavelengths, λ_{on} and λ_{off}, into the atmosphere; λ_{on} denotes the on-line wavelength, for which the trace gas absorption causes an additional attenuation of the radiation during transit through the atmosphere, and λ_{off} is the associated non-absorbing off-line wavelength. Comparison of the backscatter signal at both wavelengths yields the trace gas number density as a function of distance along the field-of-view of the receiving telescope.[2] From the point of view of socially relevant applications, in the last years two main fields have emerged: the detection of pollutant sources in urban and industrial areas and the early detection of wildfires in forest regions.

Most atmospheric pollutants are in a gaseous state in concentrations that vary between 0.01 and 10 ppm for molecules and between 0.01 and 10 ppb for metal vapours. Traditionally, in order to detect such substances and evaluate quantitatively their concentration, point monitoring stations are employed. These fixed monitoring stations implement different measurement techniques with integration times varying from 1 min to a few hours. Unfortunately, they can typically evaluate the concentration of the substance, for which the system has been calibrated, only in a very limited area close to the detector. So, to cover large areas, these fixed monitoring stations have to be deployed in high numbers and connected in a network. In addition to the practical difficulties and the economic implications, it is to be emphasized that no sound methodology exists to determine the optimal number of measurement points and their localization. Therefore, all these reasons motivate the development of a sensitive, flexible, and fast method with the capability of measuring concentrations of atmospheric gaseous pollutants over a wide range. With the development of reliable lasers, emitting in the appropriate range of wavelengths (see later), Lidar and Dial systems have become suitable and competitive techniques.[3,4]

Traditionally, Lidar and Dial techniques in the field of environmental monitoring have been operated in different spectral windows ranging from IR to UV. On the other hand, it is worth trying to extend the range of operation to the far infrared spectral region where there are many molecular absorption bands. Indeed, the vast majority of the substances to be detected (SOx, ammonia, ozone etc) emit and absorb in the spectral range between 9 and 11 μm. Furthermore, a wide atmospheric spectral window is located between the wavelength of 8 and 12 μm, mainly in the low troposphere where more strict should be the control of the pollution levels. This characteristic of the atmosphere allows working without significant attenuation by the main components of the atmosphere (CO_2, H_2O).

Experimental and theoretical investigations by Andreucci and Arbolino,[3] Lavrov and Vilar,[5,6] Pershin *et al.*,[7] Utkin *et al.*,[8,9] and Gaudio*et al.* [10–12] have also shown that Lidar has considerable potential as a technique for detecting the tenuous smoke plumes produced by forest fires at an early stage. For this purpose, the system is operated again in both Lidar and Dial configurations in sequence. The first

Lidar measurement is performed to evaluate the variation of the local aerosol density in the atmosphere, using a non-absorption water wavelength 10R18 (10.571 μm). If the returned signal shows a backscattering peak, the presence of a fire is probable. To confirm this hypothesis, a second Lidar measurement is carried out to reveal a second component emitted during the combustion process that will be used to elaborate the data in Dial. The chosen second component is water vapour, which is largely produced during the first combustion stage. Measuring the water concentration peak, after the detection of the aerosol density increment (compared to the standard mean atmospheric value), represents a good method to detect fires very early without paying an excessive price in terms of false alarms.

All these applications of Lidar and Dial systems rely on the capability to properly detect the backscattered peaks of radiation. In pioneering tests, this can be done manually but such a solution is not viable for systematic applications. In this respect, automatic analysis techniques can help substantially when results of very large surveys have to be analyzed or when real time alarms have to be reliably guaranteed. The rationale for the deployment of such methods is summarized in Sec. II. An original automatic detection tool, the Universal Multi Event Locator (UMEL), used to obtain the results presented in the paper, is described in detail in Sec. III. The results obtained by applying the technique to various experimental Lidar and Dial measurements in the field are reported and discussed in detail in Sec. IV. Conclusions and further developments are the subjects of Sec. V of the paper.

II. DETECTION OF PEAKS WITH AUTOMATIC MACHINE LEARNING AND STATISTICAL TOOLS

Lidar and Dial systems require prompt identification of the peaks in the return signals. This detection is complicated by the noise always present in the backscattered signals. Moreover, in the case of large surveys, the amount of data can be such that this task cannot be performed manually.

Automatic tools for detection of backscattered peaks present several advantages with respect to the manual analysis of data. An important issue to take into consideration is the cost of the analysis. The typical visual analysis of the waveforms requires a lot of manpower and, hence, it is expensive. Computer codes are much cheaper. Their price includes their design and maintenance/update but then, they can be run many times without additional costs. They, therefore, allow the analysis of large quantities of data on a low-cost basis. Another advantage of automatic techniques is their speed. Computer codes can be faster than any visual analysis carried out by experts. Visual analysis is time-intensive and the only way to speed up the process is increasing the number of experts and thus, increasing the costs. The deterministic behaviour of computer codes is also one of their benefits. The results of an analysis using a computer code are always the same, no matter the number of times that the program is executed. It is important to note that this does not imply that the codes are error-free; it does just mean that the same errors (if they occur) are made. This cannot be guaranteed by the analysis carried out by humans. For example, the same event can be located by two experts at two slightly different times and even the same expert can determine two different times for the same event. The reasons of these differences can be, among others, a different detail level in the analysis (the higher the level of detail, the longer the analysis time) or just a mistake.

To alleviate these problems of the manual analysis and to provide real time responses, some techniques for automatic detection of peaks have been reported in the literature. They basically rely on thresholding. The peaks present in the receiving channels are considered signal if their amplitude is sufficiently higher than the noise. This is a solution with significant drawbacks. First of all the relative amplitude of signal and noise does not remain necessarily constant during long measuring campaigns. Therefore, ample margins have to be foreseen for surveying, with subsequent losses of useful signals. Second the laser power is not necessarily stable and, therefore, drifts in the sources can jeopardize the choice of the threshold. These reasons have motivated the development of methods less dependent on the actual amplitude of the signals but based on their morphological aspect. One of these, UMEL, is described in Sec. III.

III. THE UMEL

In this section the mathematical background, on the data mining and statistical tools used in the rest of the paper, is provided. In Subsection III A, Support Vector Regression and its implementation by UMEL[13,14] is introduced. Subsection III B describes the details of the application of UMEL to peak detection in time series.

A. Introduction to machine learning for identification of events: UMEL

This subsection describes UMEL, a technique to locate events in waveforms and films. UMEL is a universal technique because it is independent of the type of the pattern sought (peaks, drops, or slope changes) and the type of waveforms analysed (time domain or frequency domain). UMEL is based on Support Vector Regression (SVR),[13] a version of SVM[14] for function estimation. SVR fits the training data without depending on factors such as sampling rate or noise distribution. This technique computes a fitting function and, in addition, it retrieves a list of the points from the training set that becomes Support Vectors (SVs).

In general, let us consider S training samples $(\mathbf{x}_1, y_1)$, $\ldots$, $(\mathbf{x}_S, y_S)$, $y_i = f(\mathbf{x}_i)$, where f: R $\rightarrow$ R is the function to regress.

The regression function is given by $f^*(\mathbf{x})\sum_{k=1}^{S}\gamma_k^* H(\mathbf{x}_k, \mathbf{x})$, where $H(\boldsymbol{x}_k, \boldsymbol{x})$ is a kernel function. The parameters γ_k^*, $k = 1, \ldots, S$ are determined using the solution of a quadratic optimization problem with linear constraints and, therefore, the formalism of Support Vector Machines can be used. Only a subset of the parameters γ^*_{k} is nonzero. The data points $\mathbf{x}_{\mathrm{k}}$ associated with the non-zero γ^*_{k} are called support vectors.

To perform regression, an insensitive zone, called e-tube, is defined and it provides the required level of accuracy e to approximate a function $f(x)$ by another function $f^*(x)$ such that the function $f(x)$ is situated in the e-tube of $f^*(x)$. The axis of the e-tube defines an e-approximation $f^*(x)$ of the function

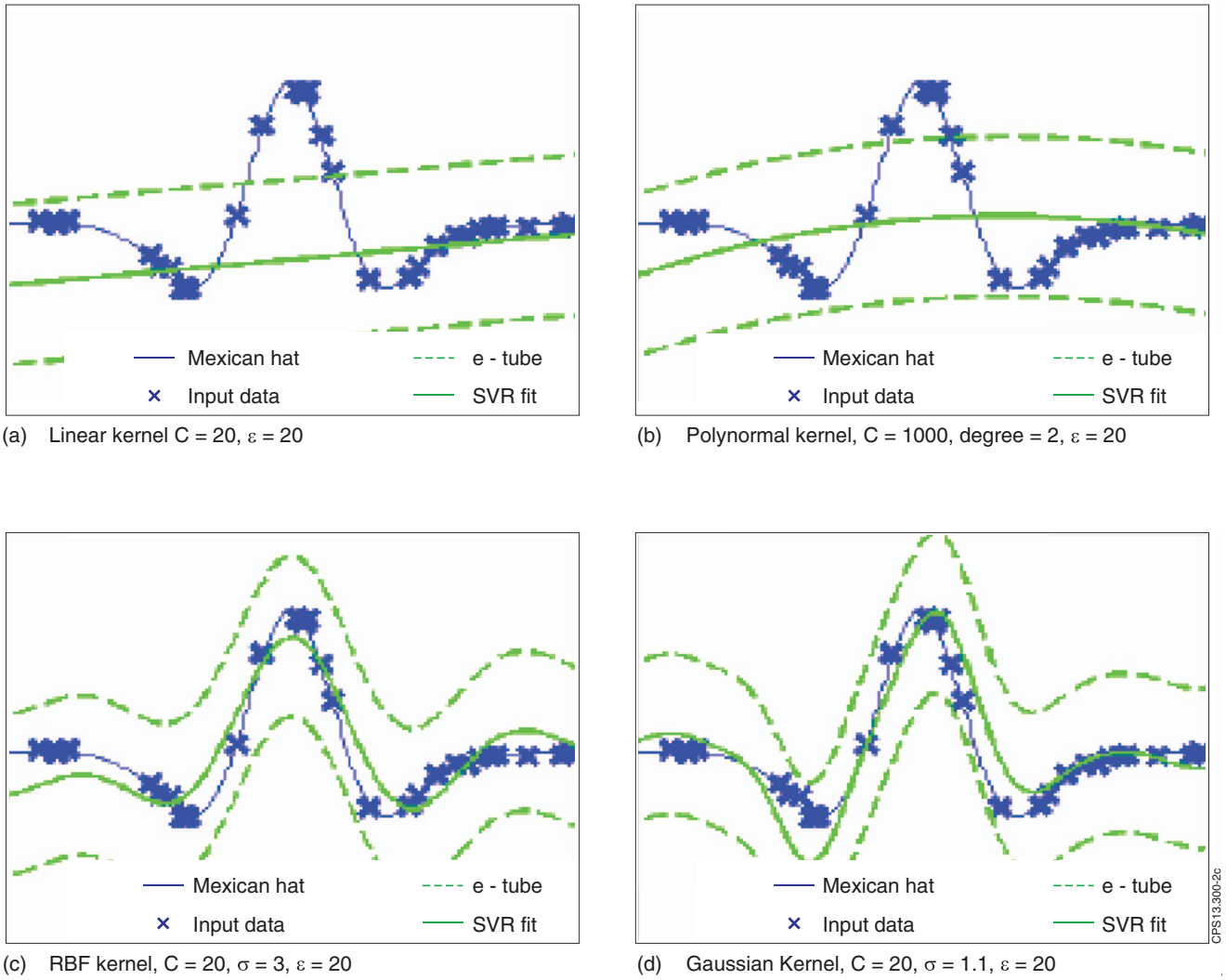

FIG. 1. SVR of the Mexican hat function using four different kernels. (a) Linear kernel, $C = 1000$, $e = 20$, (b) polynomial kernel of degree 2, $C = 1000$, $e = 20$, (c) RBF kernel, $C = 1000$, $\sigma = 3$, $e = 20$, (d) Gaussian kernel, $C = 20$, $\sigma = 1.1$, $e = 20$.

$f(x)$. Different combinations of kernels, e-values, and the regularization parameter of the optimization problem allow determining several degrees of smoothing in the estimation.

In more detail, SVR uses the e-insensitive loss function or e-tube defined as

$$|\xi|_e = \begin{cases} 0 & \text{if } |\xi| \le e \\ |\xi| - e & \text{otherwise} \end{cases}. \quad (1)$$

The goal of SVR is to find the smoothest function that fits the training data within the e-tube. To this end, the errors lower than e are not taken into consideration (the value of the e-insensitive loss function is 0 in the region (ξ [−e;+e]) but the errors higher than e are minimised. It is, therefore, possible to define the e-tube in such a way that the normal variations in the signals, including noise, remain within it and the specific events to be detected fall outside this interval.

A representative example consists of fitting data sampled from the Mexican hat. The Mexican hat is the well-known function:

$$\psi(t) = \frac{2}{\sqrt{3\sigma}\pi^{\frac{1}{4}}}\left(1 - \frac{t^2}{\sigma^2}\right)e^{-\frac{t^2}{2\sigma^2}}. \quad (2)$$

In this example, a dataset of 30 points from the Mexican hat has been randomly chosen to illustrate the use of SVR. A white Gaussian noise has been added to the samples. Four different kernels have been tested: linear, polynomial, Radial Basis Function (RBF), and Gaussian. Figure 1 shows the results obtained with the various kernels.

The regression function obtained with the linear kernel is shown in Figure 1(a). The blue line represents the Mexican hat given by Eq. (2) with $\sigma = 2$. The blue crosses indicate the data points. The green solid line is the regression function and the green dashed lines are the bounds of the e-tube. Since it uses a linear kernel, the regression function computed by SVR is a straight line. The fit obtained with the polynomial kernel of degree 2 is depicted in Figure 1(b). Figures 1(c) and 1(d) show the results of the RBF and Gaussian kernels, respectively.

UMEL can be used as an exact locator of singular points within signals. To achieve this, UMEL gives a novel interpretation of the SVs. In SVM and SVR, the complexity of the model determines the number of SVs (the higher the complexity, the larger the number of SVs). The regression of complex datasets requires large numbers of SVs. In contrast, simple datasets require smaller numbers of SVs. But the

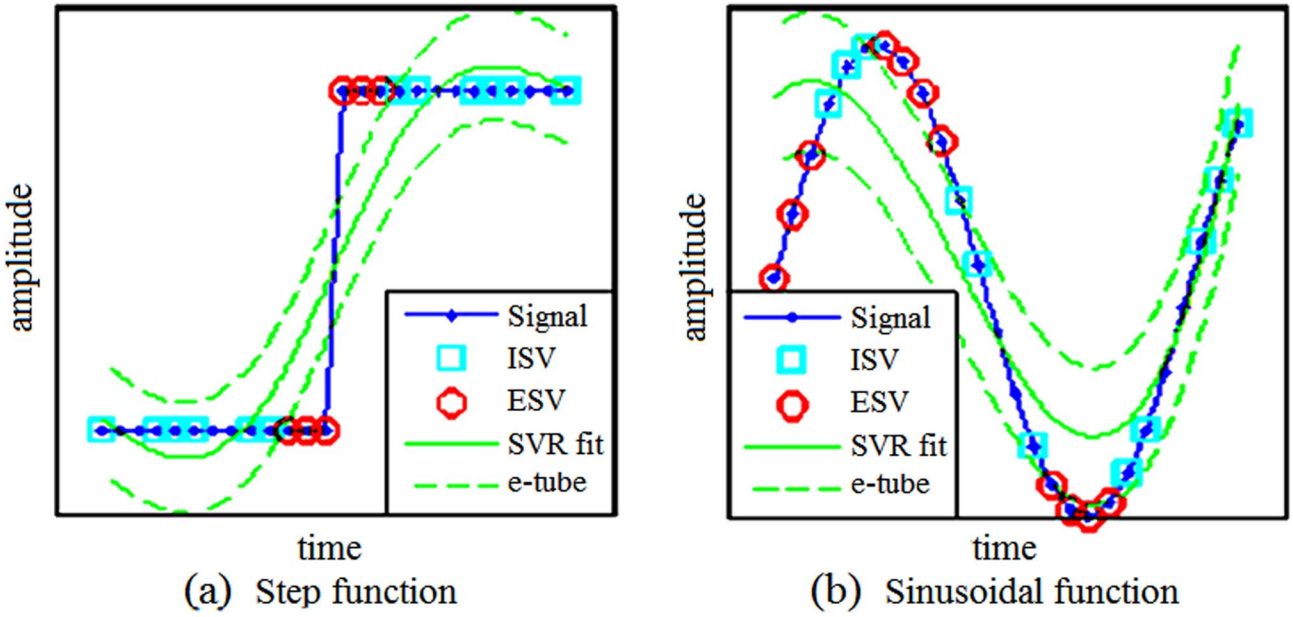

FIG. 2. [(a) and (b)] UMEL fit to a step and a sinusoidal function.

number of SVs does not depend only on the complexity of the dataset to regress. It also depends on the smoothness of the regression function. Smoother functions require fewer SVs than crispy functions. The SVs that lie on or outside the e-tube are called External Support Vectors (ESVs).[14] In contrast, the SVs within the e-tube are called Internal Support Vectors (ISV). They are defined by the relations

$$\begin{aligned} ESV \subseteq SV \ \ \forall i \in ESV, \quad |y_i - f(x_i)| \geq e \\ ISV \subseteq SV \ \ \forall i \in ISV, \quad |y_i - f(x_i)| < e \end{aligned}. \tag{3}$$

ISVs are necessary samples for the regression estimation, but they do not provide the same degree of relevance that can be assigned to ESVs. UMEL is based on a novel interpretation of ESVs: the SVs that become ESVs are the most difficult samples to regress (they cannot be fitted inside the e-tube) and these SVs provide essential information in the regression process. ESVs reveal the occurrence of special patterns inside a signal: peaks, high gradients, or segments with different morphological structure in relation to the bulk of the signal.

Figure 2 shows two examples of UMEL using a step function and a sinusoidal function. The green dashed lines delimit the e-tube. Then, the SVs within these lines are ISVs (cyan squares) and the SVs outside the e-tube are ESVs (red circles). The ESVs are clearly the most difficult samples to regress. In the case of the step function (Figure 2(a)), the samples around the step become ESVs. In the case of the sinusoidal function (Figure 2(b)), the ESVs appear at the beginning of the function and at the external points, corresponding to the maximum and the minimum of the function.

Most signals from atmospheric measurements are characterised by high frequency components in the time domain (spikes, drops, rapid slope changes, etc.). This is particularly true of Lidar measurements. Therefore, it is possible to apply UMEL to locate these events. All the mathematical details about UMEL can be found in Ref. 14.

B. Automatic localisation of peaks with UMEL

This subsection describes in detail the application of UMEL to the location of peaks in the backscattering signals of Lidar and Dial systems. The method can be applied to a wide range of signals without human intervention or modification of the code.

Three sequential tasks are carried out in the analysis: normalisation, peak location, and ESVs combination.

1. Normalisation

In order to optimise the computation of the SVR regression and allow the use of the same UMEL parameters over a wide range of cases, the signal is normalised between 0 and 1.

2. Peak location

This step locates the peaks in the backscattered signals. Figure 3 shows the location of peaks in a generic sequence of three peaks. The points of the waveform outside the e-tube become ESVs. The identification of peaks as the points above a certain threshold is not a valid strategy since the amplitude of the backscattering signals can vary from one pulse to another. The main advantage of using UMEL resides in the fact that UMEL looks for samples that do not fit a smooth regression, independently of their amplitudes.

3. ESVs combination

This step consists of the analysis of the ESVs and their use to achieve the objectives of the specific analysis problem. One typical situation is shown in Figure 3. As can be observed in Figure 4 left, more than one ESV appears on each peak of the signal. For this type of situation, the next step

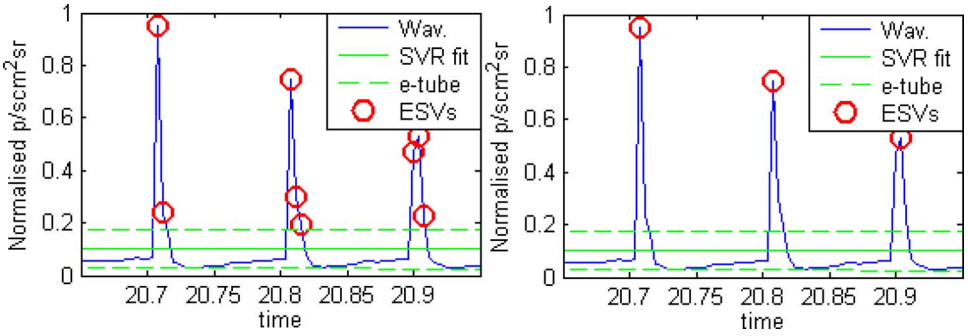

FIG. 3. Peak location and ESVs combination.

concentrates all the ESVs of each peak in a single one. Typically, in the application to peak detection in Lidar and Dial measurements, the selected point is the sample with the highest amplitude. After this task, each peak is represented by just one ESV. Figure 3 right shows the result of this step using the peaks in Figure 3 left as inputs. After this step, only one ESV remains on each peak, located at the maximum of the acquired waveform. It is worth mentioning that this is not the only viable approach. In other situations, it is more appropriate to consider the amount of ESVs in a certain region as the best indicator of the phenomenon to be detected.

IV. APPLICATION OF THE UNIVERSAL MULTI EVENT LOCATOR TO LIDAR AND DIAL MEASUREMENTS

In this section, the potential of UMEL in detecting the peaks of Lidar and Dial measurements is exemplified using data acquired during various field campaigns. The main representative cases are derived from: (a) the application of Dial systems to the estimation of pollutant concentrations in urban atmosphere (Sec. IV A) and (b) the application of Lidar measurements to the early detection of forest fires (Sec. IV B).

A. Determination of pollutant concentrations with Dial measurements

As mentioned in Sec. I, the basic concept of the remote sensing of atmospheric constituents using a differential absorption Lidar (Dial) system involves the measurement of the intensity of the backscattered radiation from a laser radar (Lidar) as a function of the frequency of the transmitted laser radiation. For the simplest case, two laser frequencies (λ_{on} and λ_{off}) are selected such that λ_{on} and λ_{off} are on-resonance and off-resonance with an absorption transition of the

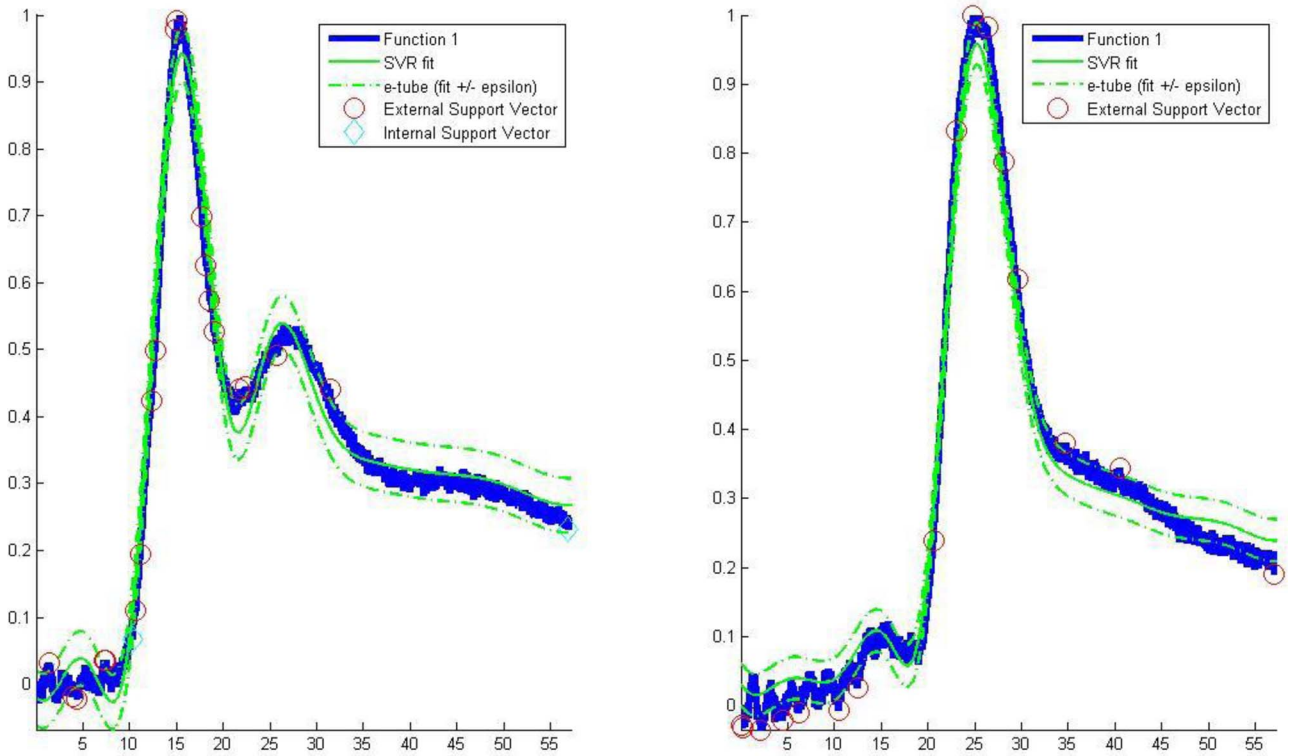

FIG. 4. Location of the laser pulses with UMEL. Laser lines of Ozone molecule. The circles indicate the laser peaks identified with the external support vectors by UMEL.

molecular specie which is being measured. The concentration of the absorbing specie may be deduced from the differential absorption of the backscattered laser radiation at wavelength λ_{on} compared to that at λ_{off}. Assuming that the atmosphere is homogeneous along the path length, the concentration for a trace gas of interest with absorption coefficients k_{on} and k_{off} is given by the dial method as

$$n(R) = \frac{1}{2(\sigma_{on} - \sigma_{off})}\left[\frac{\mathrm{d}}{\mathrm{d}R}\left\{\ln\left[\frac{P(\lambda_{off}, R)}{P(\lambda_{on}, R)}\right] - \ln\left[\frac{\beta(\lambda_{on}, R)}{\beta(\lambda_{off}, R)}\right]\right\} + k_0(\lambda_{off}, R) - k_0(\lambda_{on}, R)\right], \tag{4}$$

where the indexes *on* or *off* indicate the value referred to the absorption band or to the detuned wavelength, P is the power received by the detector, σ is the differential absorption cross section, β is the backscattered coefficient, and k_0 is the total atmospheric extinction, except the absorption due to the substance being studied.

With regard to the Dial measurements used to exemplify the potential of UMEL in this section, the used system is based on a compact CO_2 laser source operating at the wavelengths ranging from 9 to 11 μm.[15–17] This laser covers with about 60 laser lines a spectral region where there is good absorption for several atmospheric and organic molecules and it is also eye safe. The typical energy per pulse of about 200 mJ is sufficient for a range of at least 800 m. The impulse of the CO_2 is constituted by a large initial pulse and is followed by a long tail due to the presence of nitrogen in the mixture. This large initial pulse could be removed with suitable cutting systems.[18] The source is pulsed with pulse duration equal to 100 ns. The divergence of the output beam is equal to 1 mrad and the diameter of 6 mm. The maximum frequency of repetition (PRF) is 100 Hz. The envisaged application of the system is the aerial surveillance of urban areas for the detection of harmful pollutants. Therefore, automatic data analysis techniques have to be developed specifically for this purpose. Particularly delicate, from the point of view of the reliability of the measurements over long periods of time, is the stability of the laser sources, which have shown in the past a tendency to drift relative to each other. This issue, very delicate in the case of Dial systems, can be solved by a time realignment of the pulses, provided that a reliable automatic technique for the determination of the time instants of the laser pulses is available. Typically, the identification and time location of these events is achieved by means of visual analysis of the detected signals (normally waveforms of amplitude versus time). Of course, this is not possible in the case of automatic surveying systems, which cannot be manned all the time, even assuming that human operators could guarantee the required level of reliability. Therefore, UMEL has been applied to the location of these events.

The application of UMEL to the location in time of the laser pulses is illustrated in Figure 4. The location process begins with normalisation of the two Dial signals and then continues with the search of the peaks on each one. More than one ESV appears on each peak. The next step, therefore, concentrates on determining the highest ESV. In this application, a systematic analysis of large amounts of detected signals has revealed that the best sample to align the laser signals is indeed the ESV of the highest amplitude. After this selection task, each peak is represented by just one ESV. The two backscattered peaks of Figure 4 have been realigned using the described procedure and the result is shown in Figure 5. Inspection of the raising front of the pulses reveals the very good synchronicity between the two signals after application of UMEL. In general, by properly determining the level of the e-tube, it is possible to guarantee that the noise level does not affect the location of the peaks that are, therefore, all to be interpreted as laser pulses.

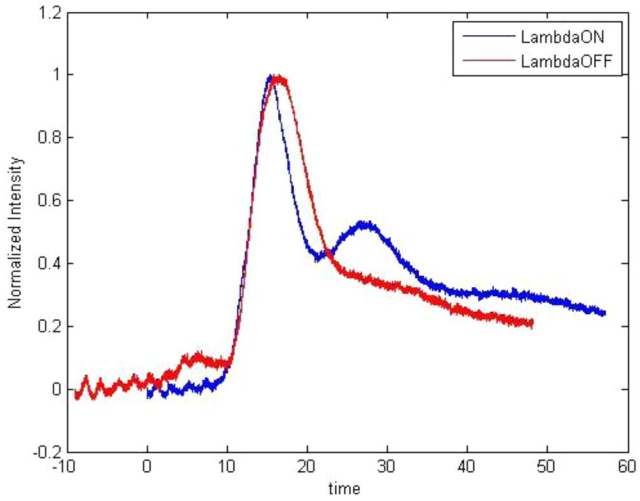

FIG. 5. Dial signals synchronized after application of UMEL.

With regard to the relevance of the correction, it is worth emphasizing that this rephasing in time of these signals is absolutely necessary to obtain realistic values of the ozone concentration in the atmosphere. Without this correction, the obtained value of the concentration would be wrong of at least two orders of magnitude. These results have been confirmed by the statistical analysis of 20 pairs of signals. In terms of the potential of UMEL in this application, compared to other solutions, it should be noticed that the final correction consists of rephasing the highest amplitude peaks of the two Dial signals. This can be performed without any information about the decay time of the peaks and their variations from laser pulse to laser pulse. The approach is, therefore, very robust and of general applicability. The time resolution is also very high, being the same as the one of the data acquisition of the system (and does not require the implementation of any additional hardware).

B. Early detection of forest fires with Lidar measurements

Forest fires can be the cause of serious environmental and economic damage. For this reason, considerable efforts have been devoted to forest protection and fire fighting. Experimental and theoretical investigations have already shown that the Lidar technique is a powerful tool to detect the tenuous smoke plumes produced by forest fires at an early stage.

Lidar is a very powerful tool for forest fire monitoring[9,11] that presents considerable advantages compared to the passive

TABLE I. Parameters of the Nd:YAG Lidar system whose measurements are analysed in this paper.

Transmitter	
Laser	
Active medium	Nd:YAG
Emission type	Pulsed (Q-Switched)
Wavelengths	1064 nm, 532 nm, 355 nm
Pulse repetition rate	10 Hz
Beam divergence	1.5 mrad
Beam waist diameter	7 mm
Pulse duration	8 ns @ 1064 nm
Pulse energy	330 mJ @ 1064 nm
Receiver	
Telescope	
Focal length	1030 mm
Primary mirror diameter	210 mm
Primary-secondary mirrors distance	820 mm
APD (model 1647 – New Focus)	
Spectral response	800 nm/1650 nm
3-db bandwidth	15 kHz–1 GHz
Peak response	0.6 A/W
NEP	1.6 pW/$\sqrt{}$Hz
Active area	0.8 mm^2

detection methods based on visible and/or infrared cameras currently in common use. Due to its very high sensitivity and spatial resolution, this active detection technique enables efficient location of small smoke plumes that originate from forest fires in the early stages of development, during both day and night and over a considerable range.

The instantaneous power of the backscattered Lidar radiation, P, is given by

$$P(\lambda, R) = P_0 \frac{A_0}{R^2} \beta(\lambda, R) \frac{c \cdot \tau}{2} \exp\left[-2 \int_0^R k(\lambda, R')\mathrm{d}R'\right], \quad (5)$$

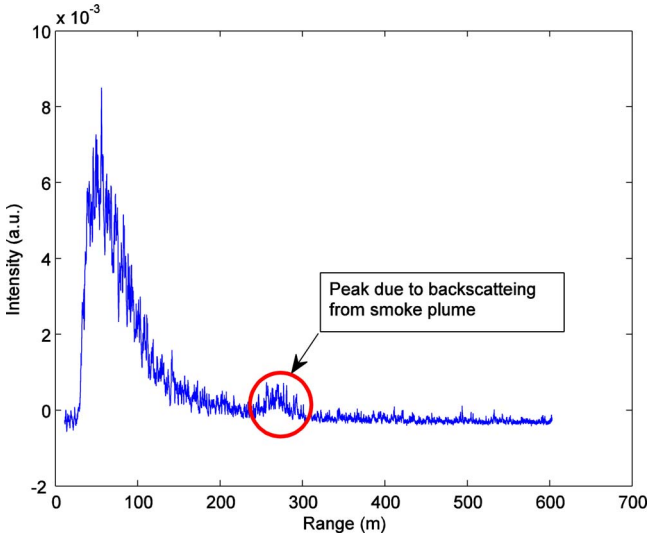

FIG. 6. Backscattered Lidar signal in the presence of smoke plume.

where P_0 is the output laser power, R is the distance of the incoming radiation, c is the speed of light, A_0 is the area of the primary mirror telescope, τ is the temporal width of the laser pulse, β is the backscattered coefficient, and k is the atmospheric extinction coefficient.

The Lidar system, for the continuous monitoring of large areas whose measurements are analysed in the rest of this section, is based on a Nd;YAG laser and a Photomultiplier detector. The choice of these components is mainly dictated by the need of developing a compact system, robust enough to guarantee continuous (24/7) operation in hostile environments. These technologies have also become relatively standard and, therefore, they can be procured at reasonable costs.[11,12,19] The overall characteristics of the systems are reported in Table I. The sensitivity of the system is typically limited by the level of noise, which is an unavoidable aspect of this type of instrumentation. Typically, indeed only the peaks of backscattered

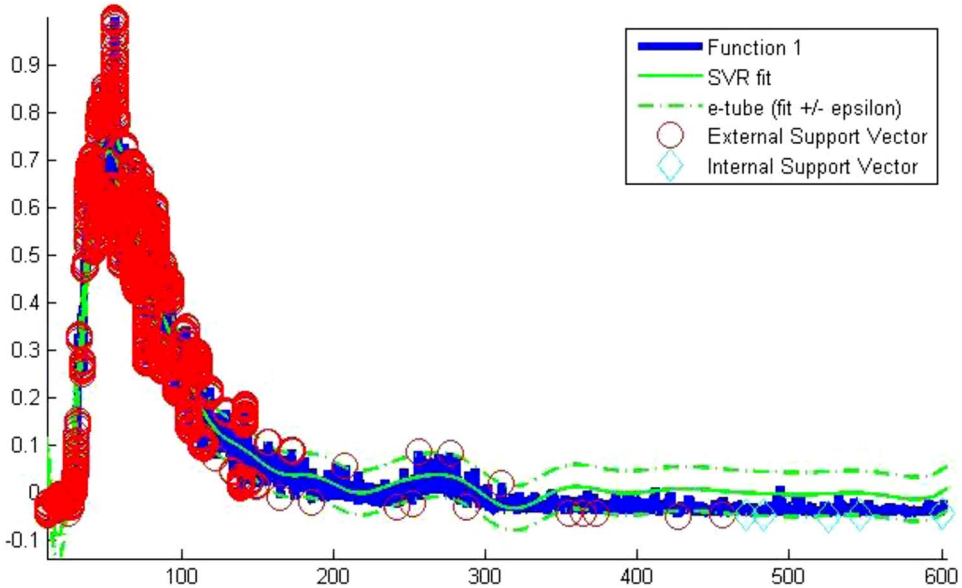

FIG. 7. UMEL application to the signal of Fig. 7. The backscattering peak is identified by the full yellow dots (all ESVs).

light which are sufficiently higher than the average noise level are considered useful signal. This approach has several limitations which motivate the development of alternative data analysis techniques, to push the boundaries of the detectable threshold. The practical implications of being able to detect smaller peaks of backscattered radiation are the capability to increase the range of the instruments and to reduce the threshold of detectable fires.

Figure 6 shows an example of a backscattered signal so small that it is smaller than the noise. It is the shape of the peak, not its amplitude, which can allow revealing the first stages of a fire. Traditional detection techniques, based on the threshold of the signal, would not permit to properly detect this case. The application of UMEL gives the fit reported in Figure 7. The peak due to the backscattering is the only part of the signal (for distances larger than 200 m which is the minimum rage of the system), which is characterised by a number of ESVs higher than 3 and located on both sides of the fit (green curve). The EVs can, therefore, be analysed and combined to identify a small signal that would go undetected otherwise. This algorithm to identify the peaks is easily implemented and it has proven to work reliably on a wide set of experimental measurements.

V. CONCLUSIONS AND FURTHER DEVELOPMENTS

In this paper the potential of UMEL to improve the measurements of Lidar and Dial systems has been documented. The potential advantages of the technique in detecting backscattering peaks are manifold. First of all, the technique is purely software and does not require any additional hardware. UMEL is also a very flexible and general method that does not depend on the knowledge of the details of the noise. Moreover, it is also independent of the sampling frequency. Therefore, the approach requires a minimum of modifications even when the hardware and/or experimental condition changes significantly. Moreover, being based on Support Vector Regression, the technique utilises the original samples of the signals and, therefore, does not cause any degradation in the time resolution of the measurements.

The effectiveness of the technique has been investigated using data of experimental campaigns in the field performed with both Lidar and Dial systems. The results are very positive both in the case of solving technical, maintenance problems and in the case of improving the performance of the systems. Particularly this second aspect, exemplified by the case of early detection of forest fires, is particularly relevant from the data analysis point of view. Indeed, UMEL was originally developed to identify automatically quite large and multiple events in the field of nuclear fusion,[20] with the aim of reducing the burden of event detection in the post-processing phase. It seems that the method is also very effective in improving the detection of single events buried in a quite high level of noise, a new field of application of the technique.

Given the positive results obtained so far, other potential applications are, therefore, being considered. In particular, UMEL could be adapted to the detection of biological agents with advanced measuring techniques such as bio-Lidar. The approach could also proved very useful in automating Lidar measurements for the detection of pollutants emitted by vehicles in urban areas.

[1] G. Fiocco and L. D. Smullin, Nature (London) **199**, 1275 (1963).
[2] R. M. Schotland, "Errors in the lidar measurements of atmospheric gases by differential absorption," J. Appl. Meteor. **13**, 71–77 (1974).
[3] F. Andreucci and M. Arbolino, "A study on forest fire automatic detection system, 2 – Smoke plume detection performance," Il Nuovo Cimento **16**(1), 51 (1993).
[4] C. Bellecci, M. Francucci, P. Gaudio, M. Gelfusa, S. Martellucci, and M. Richetta, Proc. SPIE. **5976** (2005).
[5] A. Lavrov and R. Vilar, "Application of lidar at 1.54 μm for forest fire detection," *Remote Sensing for Earth Science, Ocean and Sea Ice Applications* (SPIE, Bellingham, 1999), Vol. 3868, pp. 473–477.
[6] R. Vilar and A. Lavrov, "Estimation of required parameters for detection of small smoke plumes by lidar at 1.54 μm," Appl. Phys. B **71**(8), 225–229 (2000).
[7] S. Pershin, W. M. Hao, R. A. Susott, R. E. Babbitt, and A. Riebau, "Estimation of emission from Idaho biomass fire using compact eye-safe diode lidar," *Application of Lidar to Current Atmospheric Topics III* (SPIE, Bellingham, 1999), Vol. 3757, pp. 60–66.
[8] A. Utkin, A. Fernandes, L. Costa, R. Vilar, and F. Simoes, "Detection of small forest fire by lidar," Appl. Phys B **74**(1), 77–83 (2002).
[9] A. Utkin, A. Fernandes, R. Vilar, and A. Lavrov, "Forest fire detection by means of lidar: Forest fire research and wildland safety," in *Proceedings of the IV International Conference on Forest Fire Research* (Millpress, Rotterdam, 2002), p. 58.
[10] C. Bellecci, P. Gaudio, M. Gelfusa, T. Lo Feudo, A. Malizia, M. Richetta, and P. Ventura, "Raman water vapour concentration measurements for reduction of false alarms in forest fire detection," Proc. SPIE **7479**, 74790H (2009).
[11] C. Bellecci, L. De Leo, P. Gaudio, M. Gelfusa, S. Martellucci, T. Lo Feudo, and M. Richetta, "Evolution study of smoke backscattering coefficients in a cell by means of a compact mobile Nd:Yag lidar system," Proc. SPIE **6745**, 67451S (2007).
[12] C. Bellecci, L. De Leo, P. Gaudio, M. Gelfusa, T. Lo Feudo, S. Martellucci, and M. Richetta, "Reduction of false alarms in forest fire surveillance using water vapour concentration measurements," Opt. Laser Technol. **41**, 374–379 (2009).
[13] S. González, J. Vega, A. Murari, A. Pereira, J. M. Ramírez, S. Dormido-Canto, and JET-EFDA contributors, "Support vector machine-based feature extractor for L/H transitions in JET," *Review of Scientific Instruments, Proceedings of the 18th Topical Conference on High Temperature Plasma Diagnostics (HTPD 2010)* (Wildwood, New Jersey, 2010), Vol. 81.
[14] J. Vega, A. Murari, and S. González, "A universal support vector machines based method for automatic event location in waveforms and video-movies: Applications to massive nuclear fusion databases," Rev. Sci. Instrum. **81**(2), 023505 (2010).
[15] C. Bellecci, M. Francucci, P. Gaudio, M. Gelfusa, S. Martellucci,, T. Lo Feudo, and M. Richetta, "Application of a CO2 Dial system for infrared detection of forest fire and reduction of false alarm," Appl. Phys. B **87**, 373–378 (2007).
[16] C. Bellecci, L. Casella, S. Federico, P. Gaudio, T. Lo Feudo, S. Martellucci, and M. Richetta, "Evolution study of a water vapor plume using a mobile CO2 DIAL system," SPIE Proc. **3865**, 108–118 (1999).
[17] C. Bellecci, P. Gaudio, M. Gelfusa, T. Lo Feudo, A. Murari, M. Richetta, and L. De Leo, "In-cell measurements of smoke backscattering coefficients using a CO2 laser system for application to lidar-dial forest fire detection," Opt. Eng. **49**(12), 124302 (2010).
[18] C. Bellecci, I. Bellucci, P. Gaudio, S. Martellucci, G. Petrocelli, and M. Richetta, "Clipping the tail of a TE-CO2 laser pulse using a gas breakdown technique for high resolution chemical plume detection," Rev. Sci. Instrum. **74**, 1064–1069 (2003)
[19] P. Gaudio, M. Gelfusa, and M. Richetta, "Preliminary results of a lidar-dial integrated system for the automatic detection of atmospheric pollutants," Proc. SPIE **8534**, 853404 (2012).
[20] S. González, J. Vega, A. Murari, A. Pereira, M. Beurskens, and JET-EFDA contributors, "Automatic ELM location in JET using a Universal Multi-Event Locator," Fusion Sci. Technol. **58**(3), 755–762 (2010).