Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Atmospheric Measurement Techniques Discussions

Estimation of background gas concentration from differential absorption lidar measurements

Peter Harris[1], Nadia Smith[1], Valerie Livina[1], Tom Gardiner[1], Rod Robinson[1], and Fabrizio Innocenti[1]
[1]National Physical Laboratory, Teddington, Middlesex TW11 0LW, UK
Correspondence to: Peter Harris (peter.harris@npl.co.uk)

Abstract. Approaches are considered to estimate the background concentration level of a target species in the atmosphere from an analysis of the measured data provided by the National Physical Laboratory's differential absorption lidar (DIAL) system. The estimation of the background concentration level is necessary for an accurate quantification of the concentration level of the target species within a plume, which is the quantity of interest. The focus of the paper is on methodologies for estimating the background concentration level and, in particular, contrasting the assumptions about the functional and statistical models that underpin those methodologies. An approach is described to characterise the noise in the recorded signals, which is necessary for a reliable estimate of the background concentration level. Results for measured data provided by a field measurement are presented, and ideas for future work are discussed.

1 Introduction

Differential absorption lidar (DIAL), which is based on the optical analogue of radar, provides the capability to measure remotely the concentration and spatial distribution of compounds in the atmosphere [Measures (1984)]. The ability to scan the optical measurement beam through the atmosphere enables pollutant concentrations to be mapped and emission fluxes to be determined. The mobile ground-based DIAL systems developed at the National Physical Laboratory (NPL) for pollution monitoring have been used worldwide for over twenty years [Milton et al. (1992); Robinson et al. (2011)]. They have been deployed for routine monitoring, emission factor studies, research investigations and targeted monitoring campaigns. A key feature of the NPL system is that it operates in both ultraviolet and infrared regions. The infrared measurements at wavelengths of around 3 μm target a range of hydrocarbon gases, including methane that has a significant background concentration level (of the order of 1.8 ppmv).

The lidar technique is based on transmitting a pulse of laser radiation into the atmosphere and measuring the light scattered by the atmosphere and returned to the system. The DIAL system extends the basic lidar technique by operating the laser alternately at two adjacent wavelengths. One of these wavelengths, termed the 'on-resonant wavelength', is chosen to be a wavelength that is absorbed by the target species, and the other, the 'off-resonant wavelength', is a wavelength that is not absorbed significantly by the target species. The two wavelengths are chosen to be close, so that parameters describing atmospheric variability, such as differences in scattering media and interference compounds, are the same for both wavelengths. Any measured difference in the returned signals is therefore due to absorption by the target species.

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Atmospheric Measurement Techniques Discussions

A critical part of the analysis of the measured data provided by the DIAL system is to estimate, and subsequently correct for, the background concentration level of the target species along the optical measurement path. This is necessary for an accurate quantification of the concentration level in the atmosphere of the target species, which may be a pollutant or greenhouse gas emission, for the purpose of source attribution and to support decisions made on the basis of that quantification. This paper is concerned with approaches to analysing the measured data provided by the DIAL system to estimate the background concentration level of the target species.

The paper is organised as follows. In section 2 methodologies for estimating the background concentration level are described and contrasted. One approach makes assumptions about the nature of the functional and statistical models used to represent the values of path integrated concentration calculated in terms of the measured data. A second approach relies upon knowledge of the statistical model for the noise in the measured data, and approaches to understand and characterise the noise are considered in section 3. Results for measured data provided by a field measurement are presented in section 4. Finally, concluding remarks and a discusion of future work are given in section 5.

2 Methodologies

2.1 Two-step linear least-squares (LLS) approach

For a given elevation angle θ, the path integrated concentration $C(r)$ of the target species at a distance r along the optical measurement path is given by

$$C(r) = \frac{1}{2\gamma} \log \frac{(S_{\text{off}}(r) - \alpha)/P_{\text{off}}}{(S_{\text{on}}(r) - \beta)/P_{\text{on}}}, \tag{1}$$

where $S_{\text{off}}(r)$ and $S_{\text{on}}(r)$ are the returned signals corresponding to the off-resonant and on-resonant wavelengths, respectively, γ is a known differential absorption coefficient, α and β represent systematic offsets of the signals due to the instrumentation, and P_{off} and P_{on} are known normalisation constants used to correct for the different laser powers at the two wavelengths. In the absence of a plume containing the target species, $C(r)$ can be modelled as

$$C(r) = A + Br, \tag{2}$$

a straight-line function of distance r with A representing a constant systematic offset and B representing the background concentration level of the target species along the measurement path. By equating the expressions (1) and (2), and re-parameterising, it follows that

$$\log \frac{S_{\text{off}}(r) - \alpha}{S_{\text{on}}(r) - \beta} = a + br, \tag{3}$$

involving unknown parameters a, b, α and β, and A and B can be recovered using the expressions

$$A = \frac{1}{2\gamma}\left(a + \log \frac{P_{\text{on}}}{P_{\text{off}}}\right), \quad B = \frac{b}{2\gamma}. \tag{4}$$

Given measured values $s_{\text{off},i}$ and $s_{\text{on},i}$ of the signals $S_{\text{off}}(r)$ and $S_{\text{on}}(r)$, respectively, at distances r_i, $i = 1, \ldots, m$, an approach often applied to determining estimates of the unknown parameters in model (3) involves the following two steps:

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Atmospheric Measurement Techniques Discussions

1. Evaluate an estimate α_0 of α as the average of measured values $s_{\mathrm{off},i}$, $i = m_{\mathrm{far}}, \ldots, m$, and similarly evaluate an estimate β_0 of β as the average of measured values $s_{\mathrm{on},i}$, $i = m_{\mathrm{far}}, \ldots, m$;

2. Evaluate estimates (a_0, b_0) of (a, b) as the solution to the linear least-squares (LLS) problem

$$\min_{(a,b)} \sum_{i=m_{\mathrm{low}}}^{m_{\mathrm{hgh}}} [y_i - (a + br_i)]^2, \quad y_i = \log \frac{s_{\mathrm{off},i} - \alpha_0}{s_{\mathrm{on},i} - \beta_0}. \tag{5}$$

The index m_{far} in step (i) is chosen empirically to define a part of the two signals for which the values of the backscattered lidar signals can be considered to have essentially fallen to zero and correspond to measurements made in the far-field. The indices m_{low} and m_{hgh} in step (ii) are chosen empirically to define that part of the signals that encompasses the backscattered lidar measurement range that is free from the effects of obstructions, and for which the signal-to-noise ratio for the values y_i, $i = m_{\mathrm{low}}, \ldots, m_{\mathrm{hgh}}$, is not too small such that those values provide useful information about the background level.

Figures 1 and 2 illustrate the application of this two-step approach to the data obtained from a field measurement. Figure 1 shows the measured values for the two signals, and indicates the positions of the windows of data used in each analysis step. The signals each comprise $m = 999$ values, and the windows are defined by setting $m_{\mathrm{low}} = 30$, $m_{\mathrm{hgh}} = 300$ and $m_{\mathrm{far}} = 500$. Since the sampling rate of the lidar data acquisition system is 40 MHz, each index step represents a distance of 3.75 m for the DIAL measurements, and therefore the length of the path is (approximately) 3.75 km, the far-field in step (i) corresponds to distances greater than 1 875 m and the data analysed in step (ii) corresponds to distances between 112.5 m and 1 125 m. Figure 2 shows the path integrated concentration data

$$C_i = \frac{1}{2\gamma} \log \frac{(s_{\mathrm{off},i} - \alpha_0)/P_{\mathrm{off}}}{(s_{\mathrm{on},i} - \beta_0)/P_{\mathrm{on}}} = \frac{1}{2\gamma} \left(y_i + \log \frac{P_{\mathrm{on}}}{P_{\mathrm{off}}} \right), \quad i = m_{\mathrm{low}}, \ldots, m_{\mathrm{hgh}}, \tag{6}$$

calculated in terms of the measured signal values and the estimates α_0 and β_0, and the straight-line model

$$C(r) = A_0 + B_0 r \tag{7}$$

with A_0 and B_0 calculated in terms of the estimates a_0 and b_0 using the expressions (4). The measured signal values corresponding to indices $i < m_{\mathrm{low}}$ represent behaviour that is inconsistent with the remainder of the signals as they are due to near-field scatter rather than the backscattered lidar return signals. For indices $i > m_{\mathrm{hgh}}$, the measured signal values are such that either the path integrated concentration values calculated in terms of them and the estimates α_0 and β_0 are undefined or they are dominated by noise and provide little useful information about the background concentration level.

The two-step LLS approach is based on two main assumptions. Firstly, an assumption is made about the consistency of a straight-line model with the data y_i, $i = m_{\mathrm{low}}, \ldots, m_{\mathrm{hgh}}$. In practice, a plume containing the target species (whose position and concentration is unknown but are to be determined) may be present and, consequently, the model may be inadequate for the data. Indeed, deviations of the data from the model may provide information about the position and path integrated concentration level of the plume. Secondly, an assumption is made about the statistical model for the data y_i, $i = m_{\mathrm{low}}, \ldots, m_{\mathrm{hgh}}$, viz., that the errors in the data are realisations of random variables that are independent and identically distributed [Mardia et al. (1979)]. It is apparent from Fig. 2 that the deviations between the path integrated concentration data and the model are far from

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Atmospheric Measurement Techniques Discussions

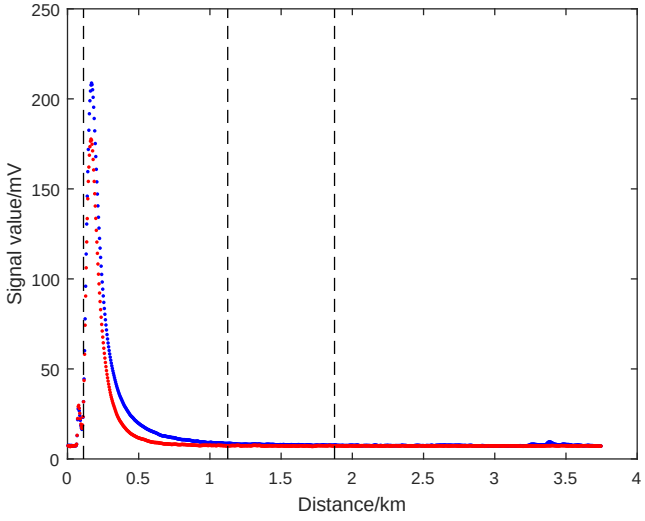

Figure 1. Measured signals (at off-resonant wavelength in blue and on-resonant wavelength in red), and the positions defined by indices m_{low}, m_{hgh} and m_{far} (vertical dashed lines, left to right) of the endpoints of the windows of data used in each step of the two-step LLS approach.

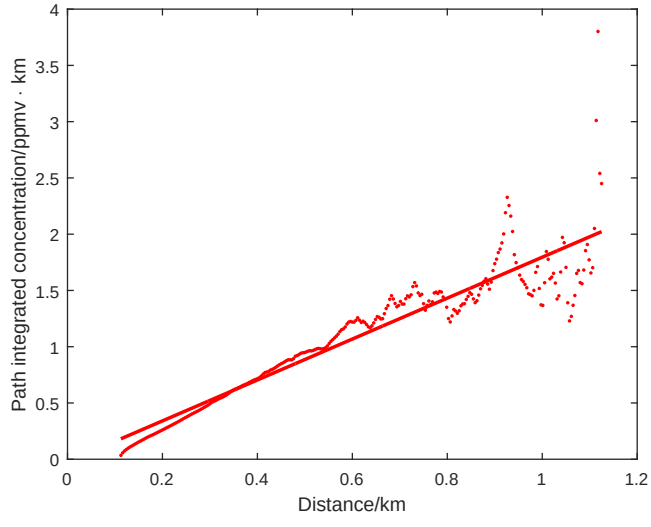

Figure 2. Path integrated concentration data and straight-line model calculated using the two-step LLS approach.

random, and consequently there is strong evidence that such an assumption about the statistical model is invalid. It follows that there can be doubt about the reliability of the estimate of the background concentration level provided by the two-step LLS approach.

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Atmospheric Measurement Techniques Discussions

2.2 One-step generalised least-squares (GLS) approach

Some information about the location of the plume, in terms of values of indices $m_{\rm s}$ and $m_{\rm e}$, where $m_{\rm s} \leq m_{\rm e}$, that define a window known to contain the plume, is often available because there can be knowledge about the location of the source of the plume relative to the location of the DIAL system. However, the reliability of this information will vary with the quality of the knowledge of the emission source location and the meteorological conditions at the time of the measurement. Given this additional information, the model (2) can be replaced by one of the form

$$C(r) = \begin{cases} A_1 + Br, & r \leq r_{m_{\rm s}}, \\ A_1 + A_2 + Br, & r \geq r_{m_{\rm e}}, \end{cases} \tag{8}$$

with B representing (as before) the background concentration level and A_2 the path integrated concentration level of the target species within the plume. (The condition $A_2 = 0$ would indicate the absence of a plume.) Use of the model (8) is a simple way to address the deficiency of the model (2) by representing the overall (macro) affect of the plume on the path integrated concentration level without attempting to model the detail (at a micro level) of how the concentration level varies within the plume.

The deficiency in the statistical model for the data y_i can be addressed by characterising the statistical properties (or probability distributions) for the errors in that data. However, such a characterisation is not straightforward because (a) those errors are obtained as a non-linear transformation (involving a quotient and a logarithm) of the errors in the measured values of the individual returned signals, and (b) they depend on the errors in the estimates (α_0, β_0) of (α, β) that, in turn, depend on the measured values for (parts of) the returned signals. However, it might be expected that the errors in the measured values of the individual returned signals themselves would be simpler to characterise and treat.

An alternative analysis approach is then based on solving a generalised least-squares (GLS) problem as follows [Forbes (1993); Forbes et al. (2002); Milton et al. (2006)]. Let $I_1 = \{m_{\rm LOW}, \ldots, m_{\rm s}\}$ and $I_2 = \{m_{\rm e}, \ldots, m_{\rm HGH}\}$ identify those parts of the measured data before and after the window containing the plume, $\boldsymbol{e}$ the vector containing $e_{{\rm off},i} = s_{{\rm off},i} - S_{{\rm off},i},\ i \in I$, and $e_{{\rm on},i} = s_{{\rm on},i} - S_{{\rm on},i},\ i \in I$, where $I = I_1 \cup I_2$, and $\boldsymbol{V}$ the covariance matrix for the complete set of data $s_{{\rm off},i},\ i \in I$, and $s_{{\rm on},i},\ i \in I$. The index $m_{\rm LOW}$ will usually be chosen equal to $m_{\rm low}$, but $m_{\rm HGH}$ can exceed $m_{\rm hgh}$ because the approach uses the measured data $s_{{\rm off},i}$ and $s_{{\rm on},i}$ directly and does not require the calculation of the data y_i. Estimates of unknown parameters a_1, a_2, b, α, β, $S_{{\rm off},i},\ i \in I$, and $S_{{\rm on},i},\ i \in I$, are obtained as the solution to the GLS problem

$$\min \boldsymbol{e}^{\top} \boldsymbol{V}^{-1} \boldsymbol{e}, \tag{9}$$

subject to the constraints

$$S_{{\rm off},i} = \alpha + (S_{{\rm on},i} - \beta)\exp(a_1 + br_i), \quad i \in I_1, \tag{10}$$

and

$$S_{{\rm off},i} = \alpha + (S_{{\rm on},i} - \beta)\exp(a_1 + a_2 + br_i), \quad i \in I_2. \tag{11}$$

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Atmospheric Measurement Techniques Discussions

The parameters A_1, A_2 and B in the model (8) are recovered from those appearing in the above formulation using

$$A_1 = \frac{1}{2\gamma}\left(a_1 + \log\frac{P_{\rm on}}{P_{\rm off}}\right), \quad A_2 = \frac{a_2}{2\gamma}, \quad B = \frac{b}{2\gamma} \tag{12}$$

(cf. expressions (4)). By using the equality constraints (10) and (11) to eliminate the parameters $S_{{\rm off},i},\ i \in I$, the problem takes the form of an unconstrained non-linear least-squares problem with unknown parameters a_1, a_2, b, α, β, and $S_{{\rm on},i},\ i \in I$, and can be solved using standard numerical algorithms [Paige (1979); Gill et al. (1981); Björck (1996)]. The problem provides (simultaneously) estimates of the unknown parameters together with the covariance matrix associated with those estimates. In terms of those estimates and uncertainties, it provides estimates and uncertainties for A_2 (the path integrated concentration level of the target species within the plume), B (the background concentration level of the target species), and the values of the returned signals at distances $r_i,\ i \in I$.

The above formulation of the GLS problem makes no assumption about the functional form of the returned signals. In an alternative formulation, the returned signal corresponding to the on-resonant wavelength is modelled as

$$S_{\rm on}(r) = \sum_{j=1}^{n} c_j \phi_j(r), \tag{13}$$

in terms of basis functions $\phi_j(r),\ j = 1, \ldots, n$, and the GLS problem is formulated as follows: determine estimates of unknown parameters a_1, a_2, b, α, β and $c_j,\ j = 1, \ldots, n$, that solve the minimisation (9) subject to the constraints

$$S_{{\rm off},i} = \alpha + \left(\sum_{j=1}^{n} c_j \phi_j(r_i) - \beta\right) \exp(a_1 + b r_i), \quad i \in I_1, \tag{14}$$

and

$$S_{{\rm off},i} = \alpha + \left(\sum_{j=1}^{n} c_j \phi_j(r_i) - \beta\right) \exp(a_1 + a_2 + b r_i), \quad i \in I_2. \tag{15}$$

For example, the signal $S_{\rm on}(r)$ (or each part of the signal before and after the window containing the plume) might be described by a cubic polynomial spline function with the basis functions $\phi_j(r)$ taking the form of B-spline basis functions of order four specified in terms of 'knots', which are the positions where the cubic polynomial pieces of the spline curve are joined [de Boor (1978); Cox (1972, 1993)]. The reason to use a parametric representation of the signal is to impose a degree of smoothness on the values of the returned signals.

A number of variants of the above formulations of the GLS problem are also considered:

1. Only the data after the window containing the plume is treated, i.e., I_1 is taken to be the empty set, and the GLS problem is solved for estimates of the parameters $a \equiv a_1 + a_2$, b, α, β, and $c_j,\ j = 1, \ldots, n$;

2. The parameters α and β are fixed, for example, to be the values α_0 and β_0 determined in terms of the data $s_{{\rm off},i}$ and $s_{{\rm on},i}$ for $i = m_{\rm far}, \ldots, m$, as in the two-step LLS approach;

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Atmospheric Measurement Techniques Discussions

3. The values α_0 and β_0 are considered to be estimates of the parameters α and β, and the minimisation (9) takes the form

$$\min \boldsymbol{e}^\top \boldsymbol{V}^{-1}\boldsymbol{e} + \lambda_\alpha(\alpha-\alpha_0)^2 + \lambda_\beta(\beta-\beta_0)^2, \tag{16}$$

where λ_α and λ_β reflect the 'quality' of the estimates α_0 and β_0.

The reasons for considering these variants of the basic problem, e.g., in which the additional terms in the minimisation (16) provide a regularisation of the minimisation (9), are discussed in section 3.

To apply the analysis approach it is necessary to know the covariance matrix $\boldsymbol{V}$ based on a characterisation of the noise in the (raw) returned signals. Obtaining such knowledge is the subject of section 3. In contrast to the two-step LLS approach, it is expected that the alternative analysis approach will provide reliable estimates of the required quantities supported by an assessment of their associated uncertainties.

3 Noise characterisation

3.1 Auto-regressive model for the noise

A characterisation of the noise in the two returned signals is undertaken in terms of estimates of the errors in the measured data for those signals. Let index $m_\mathrm{E} \geq m_\mathrm{e}$ be such that it can be stated *a priori* with confidence that the indices $m_\mathrm{E}, \ldots, m_\mathrm{HGH}$ define a part of the signals that does not include a plume. The data $s_{\mathrm{off},i}$ and $s_{\mathrm{on},i}$ within that part of the signals are smoothed (or filtered) and the residual deviations associated with the smoothed data are used as estimates of the errors in the measured data. Figure 3 illustrates the residual deviations defined by least-squares cubic polynomial spline fits to the measured data for the signals shown in Fig. 1 with $m_\mathrm{E} = 100$ and $m_\mathrm{HGH} = 500$. A cubic polynomial spline function is chosen as a flexible and empirical model to describe the underlying signals for $r_{m_\mathrm{E}} \leq r \leq r_{m_\mathrm{HGH}}$. An alternative approach to filtering the data is to use a wavelet decomposition, but the results obtained and described in section 4 are not substantially different.

Figure 4 shows the auto- and cross-correlation functions associated with the estimates of the measurement errors for the two signals. The graphs in these figures suggest that the errors are correlated, both within each signal and between the two signals. Another method for studying long-term correlations (or 'memory') in signals is detrended fluctuation analysis (DFA) [Peng et al. (1994)], which has been applied successfully to a wide range of applications in physics, medicine, climatology, and other areas. When the auto-correlation function takes the form

$$C(s) \sim s^{-\zeta}, \tag{17}$$

the corresponding detrended fluctuation function takes the form

$$F(s) \sim s^{\eta}, \tag{18}$$

where η and ζ are related by $\eta = 1 - \zeta/2$ [Kantelhardt (2001)]. Whilst Gaussian noise is usually assumed to be 'white', i.e., its auto-correlation function decays exponentially and the DFA scaling exponent is $\eta = 0.5$, in real-world systems the stochastic

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Atmospheric Measurement Techniques Discussions

variables are often described by co-called 'red' noise with $\eta > 0.5$, which means that values have memory of previous ones. The DFA scaling exponent is estimated from the linear fit to the data in a log-log plot of the fluctuation function $F(s)$. Values of η between one-half and one denote stationary noise, whereas values bigger than one denote non-stationary long-term correlated noise, with random walk noise having $\eta = 1.5$. When applied to the estimates of the measurement errors for the two signals shown in Fig. 3, the values of the DFA scaling exponents are just less than 1.5 for small lags s, indicating that the signals have appreciable short-term memory.

A (multi-output) autoregressive (AR) model is used to characterise the noise in the signals. Let $d_{\mathrm{off},i}$ and $d_{\mathrm{on},i}$ denote the residual deviations as displayed in Fig. 3. Then, the AR model takes the form

$$d_{\mathrm{off},i} + \sum_{k=1}^{q} \kappa_{1,k} d_{\mathrm{off},i-k} + \sum_{k=1}^{q} \tau_{1,k} d_{\mathrm{on},i-k} = w_{\mathrm{off},i}, \tag{19}$$

and

$$d_{\mathrm{on},i} + \sum_{k=1}^{q} \tau_{2,k} d_{\mathrm{on},i-k} + \sum_{k=1}^{q} \kappa_{2,k} d_{\mathrm{off},i-k} = w_{\mathrm{on},i}, \tag{20}$$

where q is the order of the AR model, $(w_{\mathrm{off},i}, w_{\mathrm{on},i})$ are random draws from the bivariate Gaussian distribution $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$, and $(w_{\mathrm{off},i}, w_{\mathrm{on},i})$ is assumed to be uncorrelated with $(w_{\mathrm{off},j}, w_{\mathrm{on},j})$ for $i \neq j$. (In a general form of the AR model defined above, the number of terms in each summation can be chosen to be different.) For a given choice of order q, fitting the AR model to the residual deviations provides estimates $\widehat{\kappa}_{\ell,k}$ and $\widehat{\tau}_{\ell,k}$, $\ell = 1,\ 2$, $k = 1, \ldots, q$, for the parameters $\kappa_{\ell,k}$ and $\tau_{\ell,k}$, with an associated covariance matrix, together with an estimate $\widehat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$.

Let

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{L}\boldsymbol{L}^{\top} \tag{21}$$

be the Cholesky decomposition of $\widehat{\boldsymbol{\Sigma}}$ [Golub et al. (1996)], and define

$$\boldsymbol{L}\boldsymbol{z}_i = \boldsymbol{w}_i, \tag{22}$$

where

$$\boldsymbol{w}_i = (w_{\mathrm{off},i}, w_{\mathrm{on},i})^{\top}, \quad \boldsymbol{z}_i = (z_{\mathrm{off},i}, z_{\mathrm{on},i})^{\top}. \tag{23}$$

For the residual deviations shown in Fig. 3 and the AR model of order $q = 4$ fitted to those residual deviations, Fig. 5 illustrates the transformed residual deviations $z_{\mathrm{off},i}$ and $z_{\mathrm{on},i}$, which are expected to be random draws from the bivariate Gaussian distribution $\mathrm{N}(\mathbf{0}, \boldsymbol{I})$, where $\boldsymbol{I}$ denotes the identity matrix. Figure 6 shows the auto- and cross-correlation functions associated with $z_{\mathrm{off},i}$ and $z_{\mathrm{on},i}$. The graphs in Figs. 5 and 6 suggest that the transformed residual deviations are much less correlated compared with the estimates $d_{\mathrm{off},i}$ and $d_{\mathrm{on},i}$ of the measurement errors, and resemble much more closely a white (Gaussian) noise process. Furthermore, the values of the DFA scaling exponents are approximately 0.5 for all lags s.

An approach to selecting the order q of the AR model used to characterise the noise is to obtain model fits for increasing orders, and to select that value of q for which the quality of the fit is not substantially improved using higher orders.

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Atmospheric Measurement Techniques Discussions

Figure 7 illustrates how the standard deviations and correlation coefficient derived from the covariance matrix $\widehat{\boldsymbol{\Sigma}}$ for the pairs $(w_{\mathrm{off},i}, w_{\mathrm{on},i})$ of residual deviations associated with a fitted AR model vary with the order q. It is seen that these statistics essentially saturate for $q \geq 4$, which motivates the selection of the model order made above. Furthermore, it is seen that the standard deviations for the two signals saturate to values that are quite similar, suggesting that the white noise processes underlying the noise in the two returned signals are themselves comparable.

3.2 Formulation of GLS problem with auto-regressive noise model

The estimated AR model determines the covariance matrix $\boldsymbol{V}$ in the formulation (9) of the GLS problem. However, it is unnecessary to explicitly construct the matrix $\boldsymbol{V}$, which can be a large matrix. Instead, the fitted AR model is used to define

$$f_{\mathrm{off},i} = e_{\mathrm{off},i} + \sum_{k=1}^{q} \widehat{\kappa}_{1,k} e_{\mathrm{off},i-k} + \sum_{k=1}^{q} \widehat{\tau}_{1,k} e_{\mathrm{on},i-k}, \tag{24}$$

and

$$f_{\mathrm{on},i} = e_{\mathrm{on},i} + \sum_{k=1}^{q} \widehat{\tau}_{2,k} e_{\mathrm{on},i-k} + \sum_{k=1}^{q} \widehat{\kappa}_{2,k} e_{\mathrm{off},i-k}, \tag{25}$$

for $i \in I_1' \equiv \{m_{\mathrm{LOW}}+q, \ldots, m_{\mathrm{s}}\}$ and $i \in I_2' \equiv \{m_{\mathrm{e}}+q, \ldots, m_{\mathrm{HGH}}\}$. Then, since $(f_{\mathrm{off},i}, f_{\mathrm{on},i})$, $i \in I' = I_1' \cup I_2'$, is uncorrelated with $(f_{\mathrm{off},j}, f_{\mathrm{on},j})$ for $i \neq j$ and has covariance matrix $\widehat{\boldsymbol{\Sigma}}$, the formulation (9) can be replaced by

$$\min \sum_{i \in I'} \boldsymbol{f}_i^{\top} \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{f}_i, \quad \boldsymbol{f}_i = (f_{\mathrm{off},i}, f_{\mathrm{on},i})^{\top}, \tag{26}$$

or equivalently by

$$\min \sum_{i \in I'} \tilde{\boldsymbol{f}}_i^{\top} \tilde{\boldsymbol{f}}_i \equiv \sum_{i \in I} \tilde{f}_{\mathrm{off},i}^2 + \tilde{f}_{\mathrm{on},i}^2, \tag{27}$$

where

$$\boldsymbol{L} \tilde{\boldsymbol{f}}_i = \boldsymbol{f}_i, \quad \tilde{\boldsymbol{f}}_i = (\tilde{f}_{\mathrm{off},i}, \tilde{f}_{\mathrm{on},i})^{\top}, \tag{28}$$

and the minimisation is subject to the same constraints (10) and (11) (or (14) and (15)).

For the signals shown in Fig. 1, and using an AR model of order $q = 4$ to characterise the noise in those signals as determined above, consider solving the GLS problem for the data defined by indices $I \equiv I_2 = \{m_{\mathrm{e}} = 100, \ldots, m_{\mathrm{HGH}} = 500\}$ after the window containing the plume (variant (1) in section 2.2) and for various fixed values of the parameters α and β (variant (2) in section 2.2). Let

$$S^2 = \frac{\boldsymbol{e}^{\top} \boldsymbol{V}^{-1} \boldsymbol{e}}{2|I| - p}, \tag{29}$$

here a function of α and β, be the mean-square-error evaluated at the solution, where $|I|$ is the number of indices in I and $p = 2 + n$ is the number of unknown parameters. Figure 8 shows how S^2 depends on α when $\beta = \beta_0$ and on β when $\alpha = \alpha_0$.

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

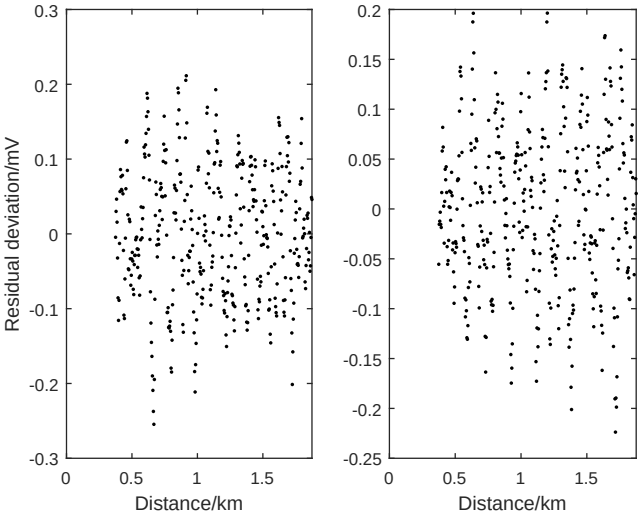

Figure 3. Residual deviations for (parts of) the measured signals shown in Fig. 1: (left) at off-resonant wavelength and (right) at on-resonant wavelength.

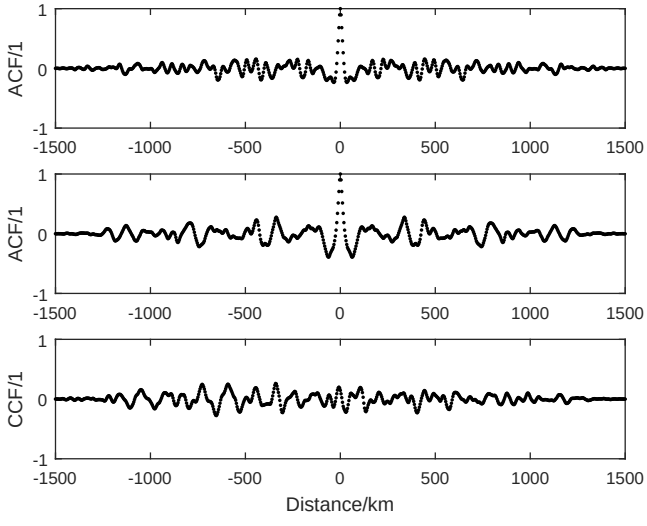

Figure 4. Properties of the residual deviations shown in Fig. 3: (top) auto-correlation function for the off-resonant wavelength residual deviations, (middle) auto-correlation function for the on-resonant wavelength residual deviations, and (bottom) cross-correlation function for the two sets of residual deviations.

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

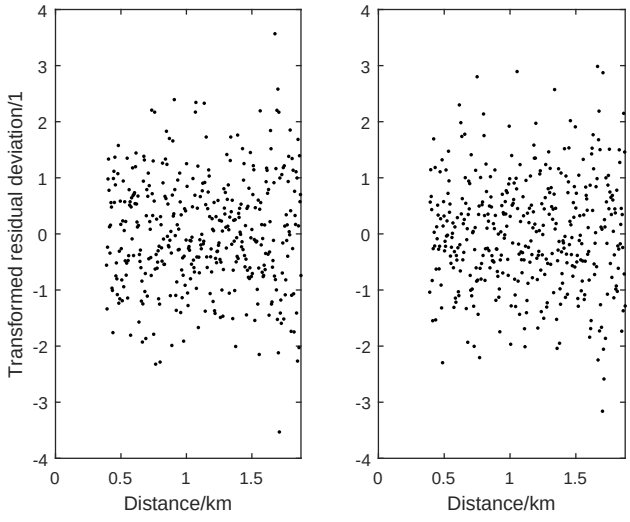

Figure 5. Transformed residual deviations for (parts of) the measured signals shown in Fig. 1 obtained using a multi-output auto-regressive model of order four for the residual deviations shown in Fig. 3: (left) at off-resonant wavelength and (right) at on-resonant wavelength.

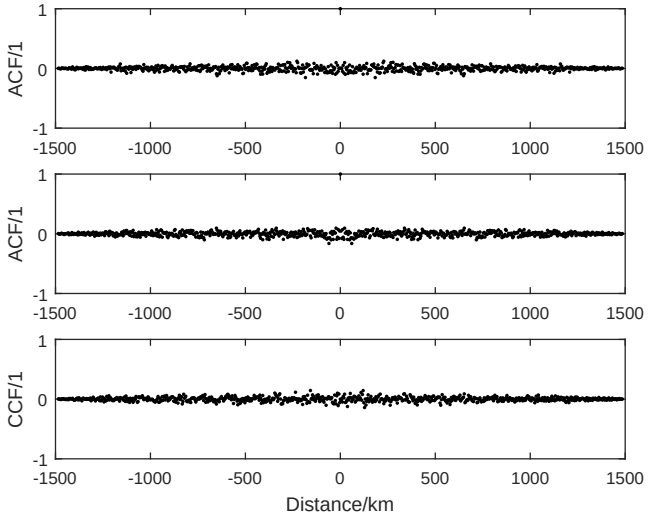

Figure 6. Properties of the transformed residual deviations shown in Fig. 5: (top) auto-correlation function for the off-resonant wavelength transformed residual deviations, (middle) auto-correlation function for the on-resonant wavelength transformed residual deviations, and (bottom) cross-correlation function for the two sets of transformed residual deviations.

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

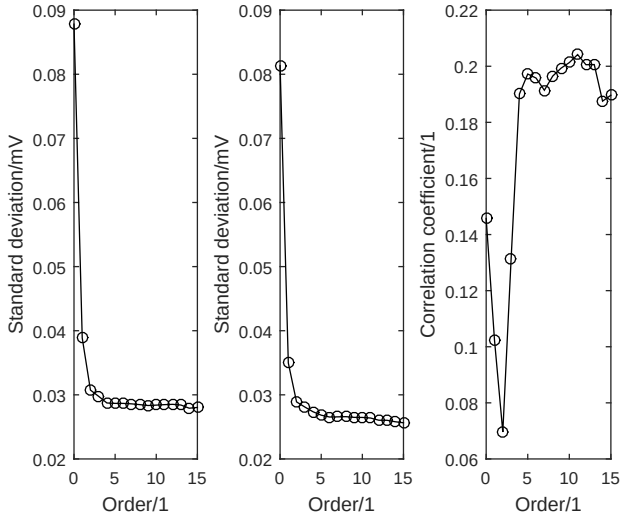

Figure 7. Values of (left) the standard deviation $\sqrt{\widehat{\mathbf{\Sigma}}_{11}}$, (middle) the standard deviation $\sqrt{\widehat{\mathbf{\Sigma}}_{22}}$, and (right) the correlation coefficient $\widehat{\mathbf{\Sigma}}_{12}/\sqrt{\widehat{\mathbf{\Sigma}}_{11}\widehat{\mathbf{\Sigma}}_{22}}$, derived from the covariance matrix $\widehat{\mathbf{\Sigma}}$ of a fitted AR noise model for different values of the order q of the model.

The graphs in the figure also show as vertical dashed lines the particular values α_0 and β_0. Although a minimum with respect to β is well-defined, and indeed corresponds to the value β_0, the same is not the case for α, since S^2 is essentially constant for values of α less than α_0 when $\beta = \beta_0$.

Figure 9 illustrates the solution to the GLS problem with unknown parameters a, b, α, β, and c_j, $j = 1, \ldots, n$. The solution is displayed by showing the path integrated concentration data calculated in terms of the estimates of α and β and the straight-line model calculated in terms of the estimates of A and B. (For comparison, the path integrated concentration data calculated in terms of the estimates of α_0 and β_0, and the linear least-squares straight-line fit to that data is also shown: cf. Fig. 2.) Figure 10 illustrates the transformed residual deviations corresponding to the solution to the GLS problem shown in Fig. 9. Figures 11 and 12 illustrate the same information but for the solution to the GLS problem with $\alpha = \alpha_0$ and $\beta = \beta_0$. It can be seen that the quality of the two solutions, both in terms of the values of S^2 and the distributions of transformed residual deviations, are comparable. However, the solutions themselves are quite different, which is a reflection of the fact that the parameter estimates are confounded with each other and are not well-defined by the data (as illustrated also in Fig. 8). It is for this reason that consideration of regularised versions of the GLS problem (variants (2) and (3) in section 2.2) are considered necessary.

Finally, Figs. 13 and 14 illustrate the solution to the GLS problem with unknown parameters a, b, α, β, and c_j, $j = 1, \ldots, n$, for the data defined by indices $I \equiv I_2 = \{m_{\mathrm{LOW}} = 30, \ldots, m_{\mathrm{HGH}} = 500\}$ that includes data before and within the window containing the plume but does not allow for the presence of the plume in the functional model. Here, particularly for data to the left of the window, the transformed residual deviations tend to be large, and suggest a lack of fit of the model, which is

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

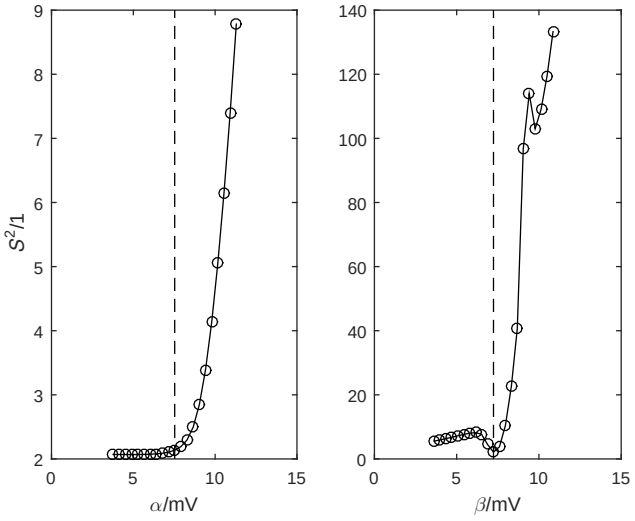

Figure 8. Values of the mean-square-error for (left) different values of α with $\beta = \beta_0$, and (right) different values of β with $\alpha = \alpha_0$, corresponding to GLS fits to data with indices $m_{\rm E}$ to $m_{\rm HGH}$. The vertical dashed lines in the two graphs correspond, respectively, to the values α_0 and β_0.

attributed to the effect of the plume. It is for this reason that adjusting the functional model for the measured data, as well as characterising the statistical model for the data, is an important part of the analysis.

4 Results

Results are presented for field measurements made for six elevation angles θ spanning the interval 3.7° to 9.7°. Figures 1 to 14 relate to the measured data corresponding to the largest elevation angle in that interval. Additionally, Figs. 15 and 16 illustrate the solution to the GLS problem with unknown parameters a, b, and c_j, $j = 1, \ldots, n$, for the data defined by indices $I = I_1 \cup I_2$ with $I_1 = \{m_{\rm LOW} = 30, \ldots, m_{\rm s} = 50\}$ and $I_2 = \{m_{\rm e} = 100, \ldots, m_{\rm HGH} = 500\}$, i.e., allowing for the presence of a plume. Here, the transformed residual deviations suggest that the noise characteristics may be different for the data before and after the window considered to contain the plume. Although the transformed residual deviations for the data before the plume appear essentially random, their magnitude is larger than would be expected.

The analysis described in sections 2 and 3 for the measured data relating to the largest elevation angle is repeated for the other elevation angles, and the results are presented in Fig. 17 and Tables 1 and 2. For all elevation angles the same window of data containing the plume, and the same order for the AR model used to characterise the noise in the returned signals, are considered. Figure 17 compares the estimates of the parameters $\kappa_{\ell,k}$ and $\tau_{\ell,k}$ (with $\kappa_{1,0}$ and $\tau_{2,0}$ set to be unity), and Table 1 compares the estimates of the elements of the covariance matrix $\widehat{\mathbf{\Sigma}}$, that jointly define the AR model used to characterise the noise in the returned signals for each elevation angle. There is generally good agreement between the estimates of $\kappa_{\ell,k}$ and $\tau_{\ell,k}$,

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

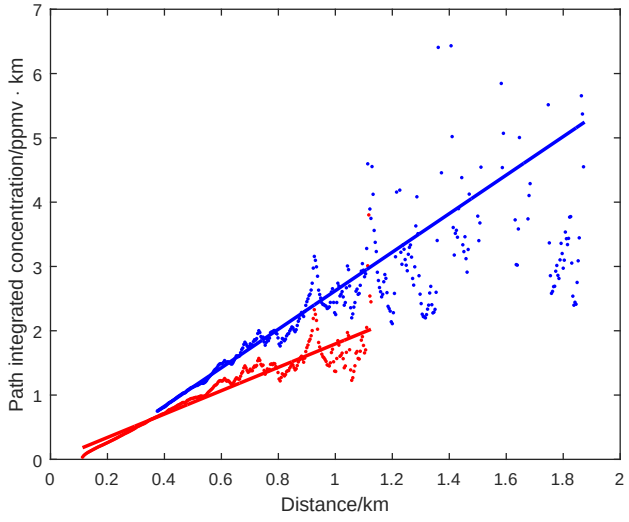

Figure 9. In red, the results of the two-step LLS approach applied to the data with indices m_{low} to m_{hgh}: the path integrated concentration data calculated using α_0 and β_0, and the linear least-squares straight-line fit to that data (cf. Fig. 2). In blue, the results of the GLS approach applied to the data with indices m_{e} to m_{HGH}: the data is calculated in terms of estimates of α and β, and the straight-line model in terms of estimates of A and B.

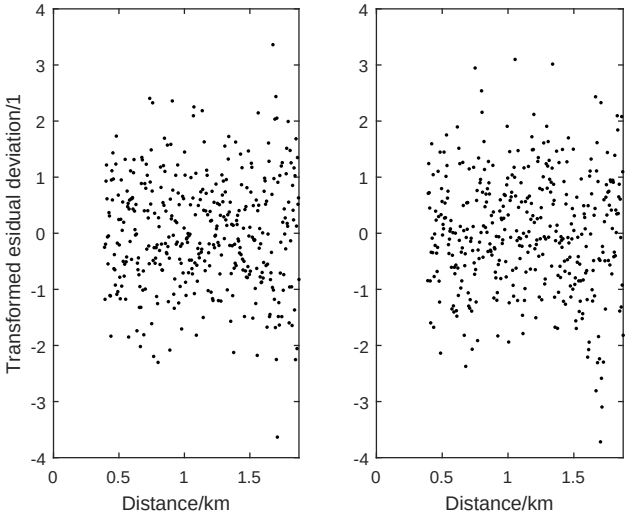

Figure 10. Values of the transformed residual deviations for the signals at (left) the off-resonant wavelength, and (right) the on-resonant wavelength, for the results of the GLS approach shown in Fig. 9.

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Atmospheric Measurement Techniques Discussions

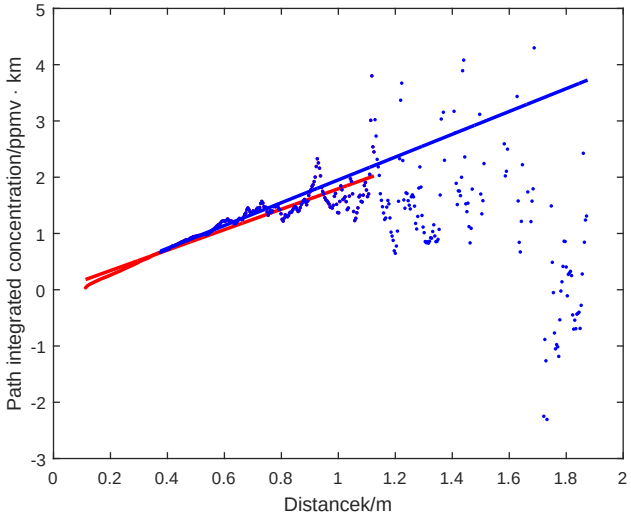

Figure 11. As Fig. 9, but with the values of α and β fixed to be α_0 and β_0 in the GLS approach.

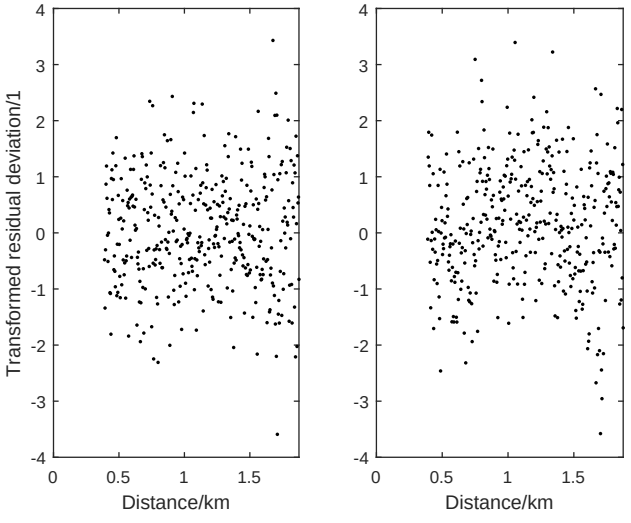

Figure 12. As Fig. 10, but for the results of the GLS approach shown in Fig. 11.

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

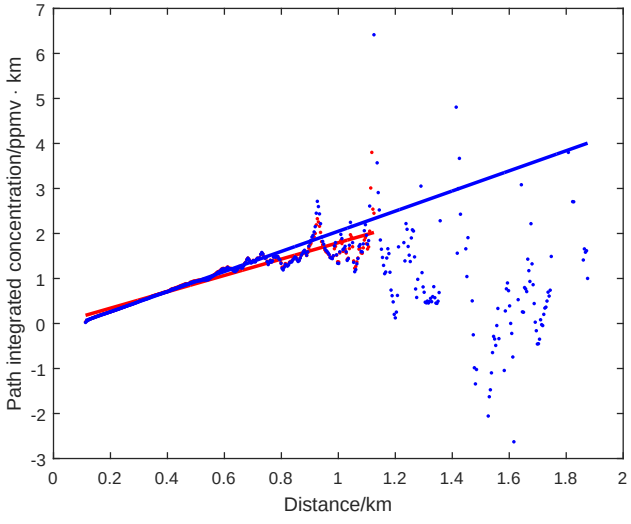

Figure 13. As Fig. 9, but for the GLS approach applied to the data with indices m_{LOW} to m_{HGH}.

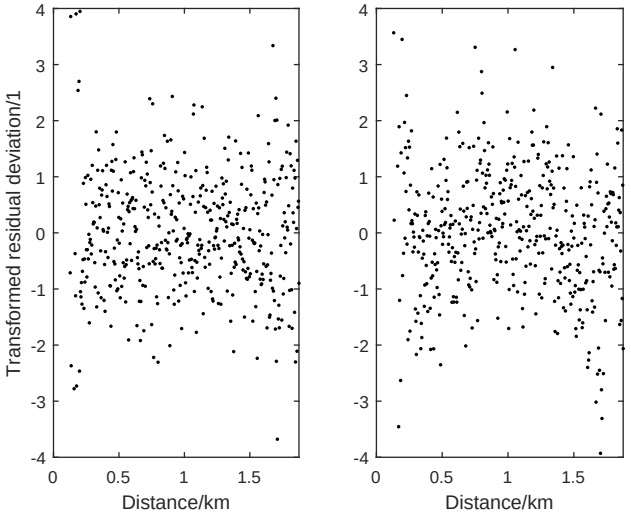

Figure 14. As Fig. 10, but for the results of the GLS approach shown in Fig. 13.

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Atmospheric Measurement Techniques Discussions

Table 1. For each elevation angle θ, estimates of the elements of $\boldsymbol{\Sigma}$ defining the AR model in a characterisation of the noise in the two returned signals.

$\theta/^\circ$	$\widehat{\Sigma}_{11}/10^{-3}\,\mathrm{mV}^2$	$\widehat{\Sigma}_{12}/10^{-3}\,\mathrm{mV}^2$	$\widehat{\Sigma}_{22}/10^{-3}\,\mathrm{mV}^2$
3.7	0.711	0.109	0.642
4.7	0.712	0.190	0.737
5.7	0.782	0.134	0.733
6.7	0.716	0.156	0.673
7.7	0.622	0.084	0.722
9.7	0.825	0.150	0.749

except that the estimates of $\kappa_{2,k}$ for the smallest elevation angle appear somewhat different from those for the other elevation angles (lower-right graph of Fig. 17).

Table 2 gives the estimates of the background concentration level B obtained in various ways and compares them for different elevation angles. In all cases, the offset parameters α and β are set to the values α_0 and β_0 calculated in terms of the signal values measured in the far-field. Estimates of B are obtained using the LLS approach, the GLS approach using data with indices $I_2 = \{m_\mathrm{e} = 100, \ldots, m_\mathrm{HGH} = 500\}$ considered to be beyond the plume, and the GLS approach using data with indices $I_1 \cup I_2$ and $I_1 = \{m_\mathrm{LOW} = 30, \ldots, m_\mathrm{s} = 50\}$, i.e., allowing for the presence of a plume. In the latter case, an estimate is also provided for the parameter A_2 representing the path-integrated concentration level of the target species within the interval defined by the indices $m_\mathrm{s} = 50$ to $m_\mathrm{e} = 100$. For all but the smallest elevation angle, the estimate of A_2 is positive, which may be taken to indicate the presence of a plume somewhere within the chosen window.

It may be expected that the results provided by the GLS approach using the data with indices I_2 are the most reliable because, firstly, data that is possibly influenced by the presence of a plume is excluded and, secondly, the same data is used for the purpose of noise characterisation. Indeed, the estimates show the least variability with elevation angle compared with the other approaches to determine estimates of the background concentration level, and particularly compared with the estimates provided by the two-step LLS approach. The results provided by the GLS approach using the data with indices $I = I_1 \cup I_2$ are, for some elevation angles, influenced appreciably by the additional data measured close to the DIAL system and before the plume. That data constitutes only a small part of the data set (about 5 %) and, as noted above, there is evidence of an inconsistency between that data and the functional and/or statistical models used in the analysis.

5 Concluding remarks and future work

This paper has been concerned with approaches to analysing the recorded measured data obtained using NPL's DIAL system to estimate the background concentration level of a target species in the atmosphere. The estimation of the background concentration level is necessary for an accurate quantification of the concentration level of the target species within a plume, which

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Atmospheric Measurement Techniques Discussions

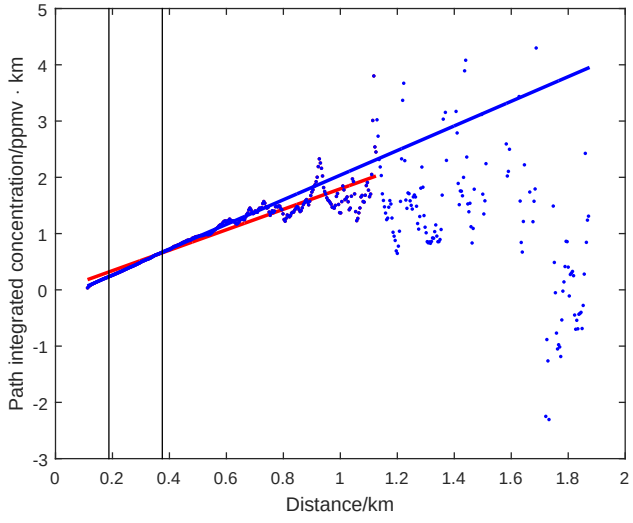

Figure 15. As Fig. 9, but for the GLS approach applied to the data with indices m_{LOW} to m_{HGH} and assuming a plume is contained in a window defined by indices m_{s} and m_{e}. The vertical lines show the endpoints of the window that is assumed to contain the plume.

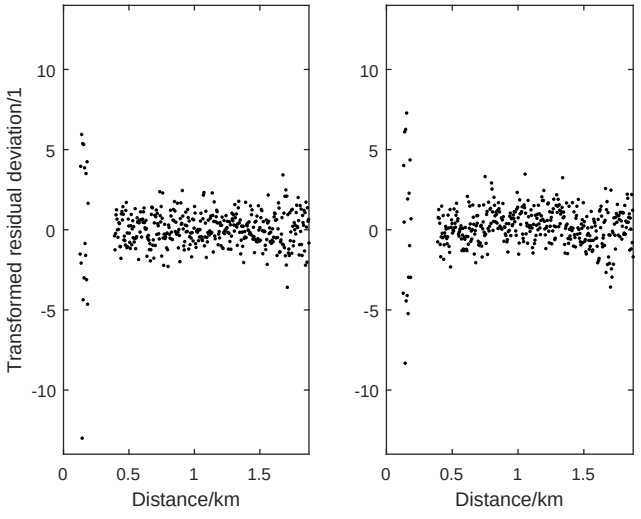

Figure 16. As Fig. 10, but for the results of the GLS approach shown in Fig. 15.

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

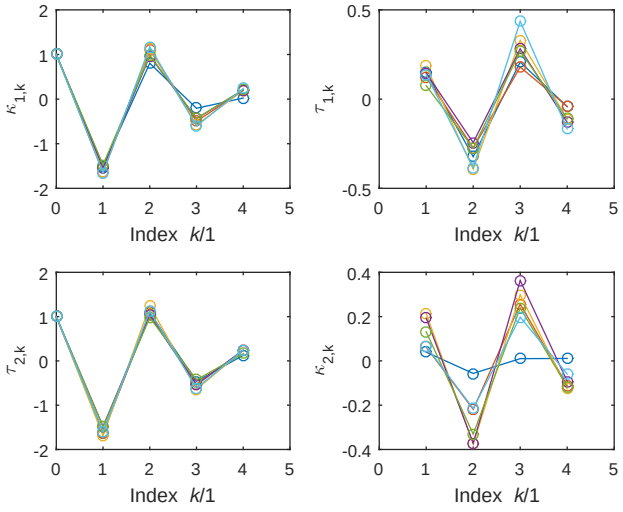

Figure 17. For each elevation angle θ, estimates of the parameters $\kappa_{1,k}$ (top left), $\tau_{1,k}$ (top right), $\tau_{2,k}$ (bottom left) and $\kappa_{2,k}$ (bottom might) defining the AR model in a characterisation of the noise in the two returned signals for different elevation anges (corresponding to the different coloured lines).

Table 2. For each elevation angle θ (column 1), estimates (α_0, β_0, b_0) provided by the two-step LLS approach (LLS, cols 2–4), estimate $\widehat{b}$ of b provided by the GLS approach with $\alpha = \alpha_0$ and $\beta = \beta_0$ for data with indices $I_2 = \{m_e = 100, \ldots, m_{HGH} = 500\}$ (GLS$_1$, col 5), and estimates $(\widehat{b}, \widehat{a}_2)$ of (b, a_2) provided by the GLS approach with $\alpha = \alpha_0$ and $\beta = \beta_0$ for data with indices $I_1 \cup I_2$ with $I_1 = \{m_{LOW} = 30, \ldots, m_s = 50\}$ (GLS$_2$, cols 6–7).

	LLS			GLS$_1$	GLS$_2$	
$\theta°$	α_0/mV	β_0/mV	b_0/ppmv	$\widehat{b}$/ppmv	$\widehat{b}$/ppmv	$\widehat{a}_2$/ppmv·km
3.7	7.716	7.645	1.822	2.020	2.440	−0.000 2
4.7	7.710	7.698	2.070	2.194	1.870	0.146 0
5.7	7.036	6.987	2.117	2.043	1.866	0.132 8
6.7	7.190	7.163	2.596	2.118	1.742	0.182 4
7.7	7.317	7.198	1.546	1.911	2.222	0.038 4
9.7	7.529	7.238	1.818	2.028	2.188	0.027 3

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Atmospheric Measurement Techniques Discussions

is the quantity of interest. The paper has focussed on methodologies for estimating the background concentration level and, in particular, contrasting the assumptions about the functional and statistical models that are part of those methodologies. It has also described an approach to characterise the noise in the recorded signals, the consideration of which is necessary to obtain a reliable estimate of the background concentration level.

The paper describes on-going work on the analysis of the measured data provided by NPL's DIAL system. Further aspects of the analysis, as well as addressing issues associated with the results presented in this paper, will be described in future publications and include:

1. The approach relies on the availability of knowledge about the location of the plume, which is then used as the basis for excluding data that can be expected to be inconsistent with the functional and statistical models used in the analysis. There is benefit in trying to refine and improve that knowledge to obtain a short window containing the plume and to increase the amount of data available to estimate the background concentration level. Similar developments could also be used to identify the presence of an unexpected plume.

2. The results presented have not included an assessment of the quality of the estimate of the background concentration level in the form of a statement of uncertainty, which is a necessary part of the quantification of the background concentration level. The assesment will need to consider not only the influence of noise in the recorded signals, which has been characterised using a particular statistical model, but also the uncertainty associated with that model, which has been estimated from an analysis of the measured data.

3. The results presented suggest that the statistical model obtained to characterise the noise in the recorded signals after the plume may provide only a partial description of the noise before the plume, in the sense that it captures the correlation structure of the noise but not its magnitude. A complete description of the noise is necessary if the part of the signals before the plume are to be used to obtain a reliable estimate of the background concentration level supported by an associated uncertainty.

4. The analysis has been applied separately and independently for each elevation angle θ. Making an assumption about the dependence of the background concentration level on θ, for example, that it is a constant or slowly-varying function of θ, would allow the measured data for the different elevation angles to be aggregated and analysed together.

5. Additional experimental data, for example, direct measurement of the background concentration level undertaken independently, such as in situations where a plume is known not to exist, would assist in the validation of the results of the analysis.

Author contributions. Tom Gardiner, Rod Robinson and Fabrizio Innocenti defined the models and provided the measured data from field measurements. Peter Harris, Nadia Smith and Valerie Livina developed the analysis approaches, implemented the approaches in software and processed the measured data to obtain numerical results. Peter Harris prepared the manuscript with contributions from all co-authors.

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Acknowledgements. The National Measurement Office of the UK's Department for Business, Innovation and Skills supported this work as part of its Innovation, Research and Development programme with additional funding support through the IMPRESS project under the European Metrology Research Programme. ©Crown Copyright, 2016.

Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2016-73, 2016
Manuscript under review for journal Atmos. Meas. Tech.
Published: 20 April 2016

Atmospheric Measurement Techniques Discussions

References

Björck A (1996) *Numerical Methods for Least Squares Problems* (Philadelphia: SIAM)

de Boor C (1978) *A Practical Guide to Splines* (New York: Springer)

Cox M G (1993) Algorithms for spline curves and surfaces *Fundamental Developments of Computer-Aided Geometric Modelling* ed L Piegl (London: Academic) 51-76

Cox M G (1972) The numerical evaluation of B-splines *J. Inst. Math. Appl.* **10** 134–49

Forbes A B (1993) Generalised regression problems in metrology *Numer. Algorithms* **5** 523–34

Forbes A B, Harris P M and Smith I M (2002) Generalised Gauss-Markov regression *Algorithms for Approximation IV* ed J Leversley et al (Huddersfield: University of Huddersfield) 270–7

Gill P E, Murray W and Wright M H (1981) *Practical Optimization* (London: Academic)

Golub G H and Van Loan C F (1996) *Matrix Computations* 3rd edn (Baltimore: Johns Hopkins University Press)

Kantelhardt J W, Koscielny-Bunde E, Rego H, Havlin S and Bunde A (2001) *Physica A* **295**

Mardia K V, Kent J T and Bibby J M (1979) *Multivariate Analysis* (London: Academic)

Measures R M (1984) *Laser Remote Sensing* (New York: Wiley)

Milton M J T, Woods P T, Jolliffe B W, Swann N R and McIlveen T J (1992) Measurements of toluene and other aromatic hydrocarbons by differential absorption lidar in the near-ultraviolet *Applied Physics B* **55** 41–5

Milton M J T, Harris P M, Smith I M, Brown A S and Goody B A (2006) Implementation of a generalized least-squares method for determining calibration curves from data with general uncertainty structures *Metrologia* **43** S291–8

Paige C C (1979) Fast numerically stable computations for generalized least squares problems *SIAM J. Numer. Anal.* **16** 165–71

Peng C-K, Buldyrev S V, Havlin S, Simons M, Stanley H E and Goldberger A L (1994) *Phys. Rev. E* **49**

Robinson R, Gardiner T, Innocenti F, Woods P and Coleman M (12011) Infrared differential absorption lidar (DIAL) measurements of hydrocarbon emissions *J. Environ. Monit.* **13**