



5 **Figure S1** SO₂ dSCDs for the base-case at all the viewing elevation angles for the high (left) and low (right) concentration cells. Grey areas indicate under-estimation of >10% and >50% for the high and low concentration measurements, respectively. Black areas indicate over-estimation of >10% and >50% for the high and low concentration measurements, respectively.

DOAS Analysis & Fitting in DOASIS

DOAS analysis is based on the Beer-Lambert law, which describes the attenuation of light with wavelength λ emitted by a radiation source $I_0(\lambda)$ as it passes through the atmosphere along the path length s by an absorber with an absorption cross-section $\sigma(\lambda)$ and number concentration c (Honninger et al., 2004).

$$I(\lambda) = I_0 e^{-\sigma(\lambda)cs} \quad (1)$$

In order to allow interpretation of radiance measurements in the real atmosphere, equation (1) would have to be expanded to include that 1) trace gas absorption cross-sections are a function of temperature and pressure, 2) multiple absorbers are typically present in the atmosphere, and 3) Rayleigh (air molecule) and Mie (aerosol) scattering of light occurs. Determination of trace gas concentration using (1) would require quantification of all factors affecting light intensity including the effect of scattering of light, turbulence, variation in the light source, and changes in spectral sensitivity of the detector (Platt et al., 2008).

For multiple absorbers i and given temperature T

$$I(\lambda) = I_0 e^{-\sigma(\lambda,T) \int c_i(s) ds} \quad (2)$$

Given that the SCD of absorber i is $S_i = \int c_i(s) ds$, the integral can be approximated as

$$I(\lambda) = I_0 e^{-\sum \sigma_i(\lambda,T) S_i} \quad (3)$$

$$\ln(I(\lambda)) = \ln(I_0) - \sum \sigma_i(\lambda, T) S_i \quad (4)$$

The DOAS technique eliminates the problem of having to quantify all factors affecting light intensity (see above) by separating absorption structures that vary “slowly” with wavelength (σ^B) from the differential structures (σ').

$$\sigma_i(\lambda) = \sigma_i^B(\lambda) + \sigma_i'(\lambda) \quad (5)$$

The absorption bands unique to each gas are differential features while Rayleigh and Mie scattering, light source variation and spectral sensitivity are broadband features (Honninger et al., 2004).

The result is the following set of linear equations

$$\ln(I(\lambda)) = \ln(I_0) - \sum \sigma_i^B(\lambda, T) S_i - \sum \sigma_i'(\lambda, T) S_i \quad (6)$$

The first and second terms on the right-hand side of the equation are low frequency (broadband) and can be modelled as a polynomial function $P(\lambda)$ with coefficients a_j .

$$\ln(I_0) - \sum \sigma_i^B(\lambda, T) S_i = P(\lambda) = \sum_{j=0}^m a_j \lambda^j \quad (7)$$

$$\ln(I(\lambda)) = - \sum \sigma_i'(\lambda, T) S_i + P(\lambda) \quad (8)$$

S_i and a_j can be solved for using the linear least squares method.

However, changes in temperature drift cause wavelength shift (s_i) and squeeze (t_i) so that the wavelength calibration can be slightly inaccurate. Such inaccuracies are non-trivial because absorption features can be only a few detector channels wide (Kraus, 2006). The shift and squeeze parameters are non-linear, preventing a linear solution.

$$\ln(I(\lambda)) = - \sum \sigma'_i(s_i + t_i\lambda)S_i + P(\lambda) \quad (9)$$

Therefore, DOASIS uses an iterative Levenberg-Marquardt method to obtain the optimal solution. The DOAS retrieval of SCDs finds the best fit of a set of reference spectra to the measured spectrum by minimizing the cost function χ^2 .

$$\chi^2 = \int \left[\ln I(\lambda) - P(\lambda) + \sum_i \sigma'_i(s_i + t_i\lambda)S_i \right]^2 d\lambda \quad (10)$$

The first term inside the brackets includes the measured spectrum, and the other two terms include the modelled parameters and absorption cross-sections.

The Levenberg-Marquardt Method combines the gradient descent method, which tends to converge rapidly when the starting conditions are far away from the cost function minimum, and the Gauss-Newton method, which is most effective close to the optimal solution (Kraus, 2006). The methods were combined by Levenberg (1944) because the combined algorithm is more stable and converges faster than the individual methods (Platt et al., 2008). A flow chart of the steps in a DOASIS fit can be found in Fig. 7.1 in (Kraus, 2006). The parameters in the model are split into sets of linear (SCDs and polynomial) and nonlinear parameters (shift and squeeze). A first guess of the linear parameters is made, followed by an estimation of the non-linear parameters. A simple least squares method is used to solve for the linear parameters while keeping the nonlinear parameters constant. A step of the iterative algorithm then calculates the nonlinear parameters with the linear parameters kept constant. The linear parameters are now recalculated using the new nonlinear parameters, and the iterations continue until a terminal condition is reached (Kraus, 2006). If an iteration step produces a new estimate with a greater cost function, the gradient method is used to seek a better solution farther away, but if the new estimate is an improvement, the Gauss-Newton method is used. The iteration process stops when the change in the cost function is $<10^{-5}$ or maximum iteration steps have been reached. When the steps become very small, ideally the iteration is close to the cost function minimum. Note that the algorithm may not find the global optimal solution like the least squares method but rather a local minimum in the cost function. Also, the algorithm results are stable and correct if all parameters are independent, but cross-sections containing similar parameters may introduce inter-dependencies that can create instability in the retrieval (Kraus, 2006).

An offset polynomial, $O(\lambda)$ can be enabled in the DOASIS fit to compensate for local broad band structures within the fitting window (e.g., stray light).

$$\ln(I(\lambda)) = e^{-\sum \sigma'_i(\lambda,T)S_i + P(\lambda)} + O(\lambda) \quad (11)$$

An approximation that $\ln\left(1 + \frac{O(\lambda)}{I(\lambda)}\right) \approx \frac{O(\lambda)}{I(\lambda)}$ for small $\frac{O(\lambda)}{I(\lambda)}$ leads to the equation used by the software as follows:

$$\ln(I(\lambda)) = - \sum \sigma'_i(s_i + t_i\lambda)S_i + P(\lambda) + \frac{O(\lambda)}{I(\lambda)} \quad (12)$$

