MIT Libraries | DSpace@MIT

MIT Open Access Articles

Deriving emissions time series from sparse atmospheric mole fractions

Massachusetts Institute of Technology

Deriving emissions time series from sparse atmospheric mole fractions

M. Rigby,[1] A. L. Ganesan,[1] and R. G. Prinn[1]

[1]   A growth-based Bayesian inverse method is presented for deriving emissions of atmospheric trace species from temporally sparse measurements of their mole fractions. This work is motivated by many recent studies that have deduced emissions using archived air samples with measurement intervals of the order of a year or longer in the early part of the record. Several techniques have been used to make this underdetermined problem invertible. These include the incorporation of prior emissions estimates, the smoothing of observations or derived emissions, the approximation of emissions time series by polynomials, or the application of regularization schemes. However, these methods often suffer from limitations, such as the unavailability of independent, unbiased priors, the emergence of unrealistic emissions fluctuations due to measurement outliers, or the subjective choice of measurement or emissions smoothing time scales. This paper presents an alternative solution that reduces the influence of potentially biased priors or measurement outliers by constraining the emissions growth rate around some growth estimate, in conjunction with the model-measurement mismatch.

Citation: Rigby, M., A. L. Ganesan, and R. G. Prinn (2011), Deriving emissions time series from sparse atmospheric mole fractions, *J. Geophys. Res.*, *116*, D08306, doi:10.1029/2010JD015401.

1.   Introduction

[2]   Measurements of atmospheric trace gas mole fractions provide invaluable information on the sources of many greenhouse gases, ozone depleting species and pollutants harmful to health. A wide range of inverse methods and atmospheric chemical transport models are commonly used to infer emissions using these observations. However, measurements are often available only at very low frequency, which can make it difficult for the investigator to derive physically realistic emissions time series. Much recent work has attempted to derive emissions time series spanning several decades using archived or firn air samples collected at one or two sites, with a frequency as low as a few measurements per decade [e.g., *Geller et al.*, 1997; *Miller et al.*, 1998; *O'Doherty et al.*, 2004; *Oram et al.*, 1998; *Weiss et al.*, 2008; *Mühle et al.*, 2009, 2010; *Montzka et al.*, 2010; *Miller et al.*, 2010; *Rigby et al.*, 2010]. A problem with inversions based on such observations is that "noisy" solutions are often obtained (i.e., the derived emissions exhibit unrealistically large temporal fluctuations) due to both the ill-conditioned nature of the inversion and measurement outliers.

[3]   Here we present a simple method for deriving emissions from a sparse atmospheric record by incorporating independent information on the evolution of the emissions processes. We reduce the scope of the problem to the one common to most of the above cited studies: the determination of global, annual release rates of long-lived trace gases that have a predominantly anthropogenic origin. However, it is anticipated that the method proposed can be more widely applied.

[4]   Several authors have addressed the issue of solution "noise" using various techniques: *Mühle et al.* [2010] derived annual emissions of the three major perfluorocarbons (CF_4, C_2F_6 and C_3F_8) from 1973–2009 using archived and in situ measurements. To remove the large fluctuations obtained following a Kalman filter-type inversion, they smoothed the a posteriori emissions using a 5 and then 3 year running mean. Many authors have used emissions polynomials to obtain solutions that did not exhibit unrealistic emissions fluctuations [e.g., *Maiss and Brenninkmeijer*, 1998; *Miller et al.*, 1998; *Mühle et al.*, 2009; *Xiao et al.*, 2007]. *Rigby et al.* [2010] and *Miller et al.* [2010] found that a Bayesian approach incorporating independent emissions estimates was able to produce a solution with little unrealistic noise, largely thanks to the availability of independent prior emissions information that agreed well with the derived values. They were able to prevent unrealistic variations in the derived emissions by constraining the solution relatively tightly to the prior. Trace gas source variability from ice core data has been inferred using piecewise cubic splines to smooth the observations [e.g., *Joos et al.*, 1999; *Enting*, 2002; *Enting et al.*, 2006]. These splines act as a low-pass filter, where the high-frequency fluctuations that are presumed to be due to noise can be removed. Other data-smoothing techniques exist, which have been used in various inversions [e.g., *Masarie and Tans*, 1995]. Another very commonly used tool is "regularization", in which the

---

[1]Center for Global Change Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

part of the solution associated with small singular values is suppressed [e.g., *Tikhonov and Arsenin*, 1977; *Hansen*, 1992; *McIntosh and Veronis*, 1993].

[5] As an alternative to these approaches, we seek a solution that (1) minimizes the influence of biased priors or measurement outliers (such as the common situation where the transport model assumes that background mole fractions have been sampled, but in reality some contamination of the observation by local sources may have occurred); (2) allows smoothing to be objectively applied, rather than arbitrarily estimated through the use of a posteriori averaging; (3) allows time-varying growth information to be included (as opposed to the implied persistence of some regularization schemes as outlined later); (4) does not impose a functional form on the observations or derived emissions; (5) does not require smoothing of the measurements, which makes it difficult to retain variations related to transport and chemistry and introduces a covariance into the measurement errors; and (6) is simple to implement based on physical reasoning. We find that the proposed growth-based inverse method satisfies these criteria for anthropogenic trace gas emissions estimation, provided that reliable prior information on the likely changes in global emission rates exists.

2. Growth-Constrained Bayesian Inversion

[6] Given the temporal sparsity of historical atmospheric measurements, the problem of deriving annual emissions over several decades is often underdetermined for at least part of the time series. Additional information or constraints are required to make the problem invertible. Since, in the case of anthropogenic gas emissions, the investigator usually has some reliable knowledge about the typical magnitude of year-to-year emissions changes, the constraint that we will use here is that an emissions time series should have a characteristic time-varying growth rate and growth uncertainty that is known a priori, independently of the observations. This approach has been modified from regularization methods used for spatial smoothing of tracers in the ocean, as well as many other problems [*McIntosh and Veronis*, 1993; *Hansen*, 1992; *Wunsch*, 2006].

[7] We relate emissions to mole fractions with the usual measurement equation as follows:

$$\mathbf{y} = \mathbf{E}\mathbf{x} + residual \qquad (1)$$

where $\mathbf{y}$ is a vector containing M measurements, $\mathbf{x}$ a vector of N emissions (here assumed to be sequential in time) and $\mathbf{E}$ is a Jacobian matrix that contains the sensitivity of concentrations in a row of $\mathbf{y}$ to a change in the emissions in the relevant row of $\mathbf{x}$. $\mathbf{E}$ can be estimated in a number of ways [e.g., *Khasibatla et al.*, 2000; *Wunsch*, 2006, section 4.2.2].

[8] We desire a solution that minimizes the difference between the model and the measurements and the deviation from some growth estimate within some specified uncertainty. These criteria are quantified in the dimensionless cost function [e.g., *Tarantola*, 2005], assuming that the probability distributions of each term are Gaussian and that the distributions are independent:

$$J = (\mathbf{y} - \mathbf{E}\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{E}\mathbf{x}) + (\mathbf{D}\mathbf{x} - \mathbf{g})^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{D}\mathbf{x} - \mathbf{g}) \qquad (2)$$

The first term represents the model-measurement mismatch, and the second is a constraint on the size of the expected changes in emissions from one time step to the next (with expected emissions growth $\mathbf{g}$ in units of [emission rate]/[time]). Note that additional terms could be incorporated that include, for example, independent prior information about absolute emission rates, or spatial gradients. Each term in the cost function is weighted by a covariance matrix: $\mathbf{R}$ is the measurement-model uncertainty covariance, and $\mathbf{S}$ is the covariance in the growth uncertainty estimate. This growth uncertainty covariance can act as a time-varying smoothing term, the size of which can be explicitly set. The operator $\mathbf{D}$ calculates the difference between elements of $\mathbf{x}$. Therefore, in the case that $\mathbf{x}$ is an emissions time series with time increasing with row index in uniform increments ΔT, $\mathbf{D}$ is given by

$$\mathbf{D} = (1/\Delta\mathrm{T})\begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & -1 & 1 \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix} \qquad (3)$$

The cost function J is minimized by the solution (differentiating with respect to $\mathbf{x}$ and equating to zero):

$$\mathbf{x} = \mathbf{P}(\mathbf{E}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y} + \mathbf{D}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{g}) \qquad (4)$$

$$\mathbf{P} = \left(\mathbf{E}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{E} + \mathbf{D}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{D}\right)^{-1} \qquad (5)$$

where $\mathbf{P}$ is the error covariance in the solution (derived as the inverse of the second derivative, with respect to $\mathbf{x}$, of the cost function).

[9] It should be noted that in the case $\mathbf{g} = \mathbf{0}$, $\mathbf{S} = \gamma\mathbf{I}$, where γ is some scalar, then the above scheme is equivalent to a regularized solution with a smoothness constraint [e.g., *Hansen*, 1992]. The value of the inclusion of a growth vector and uncertainty (rather than a single smoothing factor that would be quantified by γ) is that physically reasonable information often exists on the evolution of the growth of an anthropogenic gas. For example, it is often the case that emissions of a particular species were known to begin suddenly in a particular year when production began. Using the solution above, it is trivial to specify that growth is expected to be zero before a particular date, and positive afterward.

[10] In the above formulation, prior information is incorporated into the solution through the growth and growth uncertainty covariance terms ($\mathbf{g}$ and $\mathbf{S}$), rather than in an absolute sense as has been done in previous Bayesian inversions [e.g., *Khasibatla et al.*, 2000; *Enting*, 2002; *Tarantola*, 2005]. A key motivation for the development of the above solution was that a biased prior can often lead to a biased, and/or "noisy" emissions estimate if absolute prior emissions are used. The use of a priori information to estimate growth and growth uncertainty in the solution has the advantage that an overall bias due to an erroneous prior can be avoided, since the emissions are not "anchored" to absolute prior values. Furthermore, unphysical interannual fluctuations can be avoided through the use of an appropriately chosen growth uncertainty. Of course, if the prior growth information used in
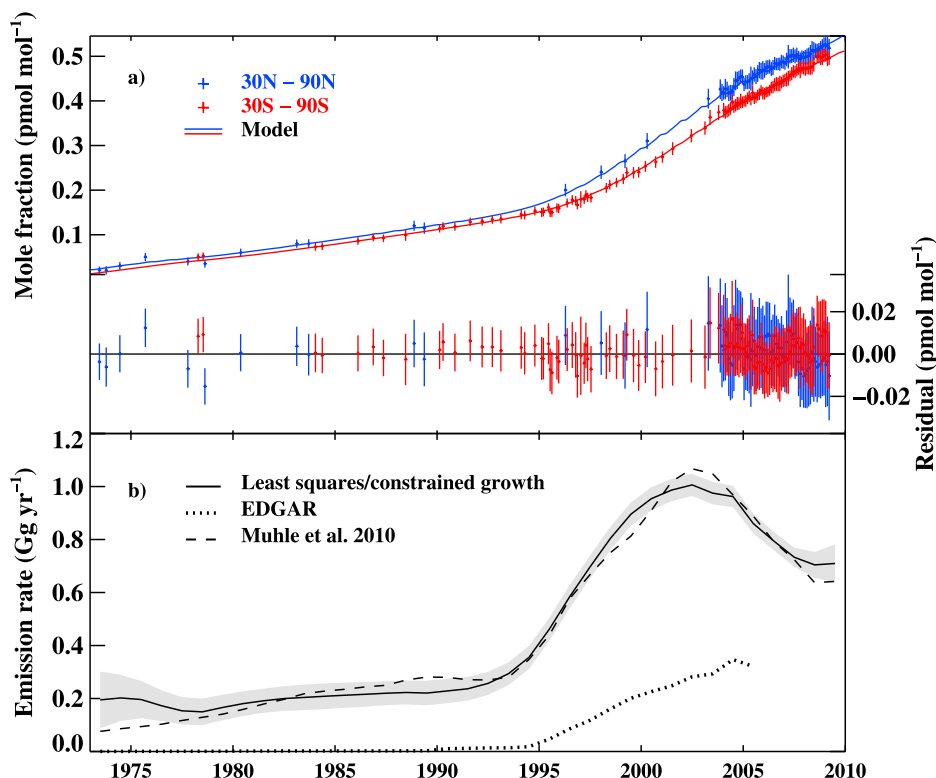
Figure 1. (a) C_3F_8 mole fractions in the extratropical Northern and Southern hemispheres (crosses) and mole fractions simulated using the derived emissions (solid lines). For clarity, the tropical measurements used in the inversion from 2006 onward are not plotted. Residuals (measurements-model) are shown below the absolute mole fractions. Error bars denote $1 - \sigma$ uncertainties. (b) Global emission rate derived using the weighted least squares method with constrained growth from 1973 to 2008 (solid line), *Mühle et al.* [2010] estimates (dashed line), and EDGAR estimates (dotted line). The shaded area shows the $1 - \sigma$ uncertainty range in the growth-constrained estimates. Model and measurement-scale uncertainties are not included in these estimates.

this scheme is strongly biased, incorrect year-to-year changes will be derived.

3. Application

[11] We applied the above method to measurements of C_3F_8, an anthropogenic compound whose annual emissions have been determined using a recursive inverse method that required a posteriori smoothing to remove solution noise [*Mühle et al.*, 2010]. Independent emissions estimates exist for this compound from the Emissions Database for Global Atmospheric Research (EDGAR) [*European Commission Joint Research Centre and Netherlands Environmental Assessment Agency*, 2009]. However, in some years they were found to be orders of magnitude smaller than those estimated using the measurements (Figure 1), making EDGAR unsuitable as a prior estimate of absolute emissions.

[12] The *Mühle et al.* [2010] measurement record was composed of air samples that were taken at Cape Grim, Tasmania, and several sites in the extratropical Northern Hemisphere (mostly at Trinidad Head, California, United States), and were combined with high-frequency Advanced Global Atmospheric Gases Experiment (AGAGE) measurements from 2006 onward at both extratropical and tropical

sites. The measurement uncertainty calculated by *Mühle et al.* [2010] was used.

[13] The sensitivity of the measurements to global emissions was estimated using the AGAGE 12-box model [*Cunnold et al.*, 1994]. This model was chosen for consistency with *Mühle et al.* [2010]. An alternative chemical transport model, incorporating interannually varying meteorology and three-dimensional transport and chemistry, for example, could readily be used with the inverse method presented. Given the very long lifetime of C_3F_8, it was considered to have no sinks in the atmosphere. For simplicity, global total emissions were solved for using the spatial distribution found by *Mühle et al.* [2010], although emissions at a higher spatial resolution could be derived, for example, through the addition of a spatial gradient term in the cost function and modification of the vector **x**. Sensitivities (matrix **E** in equation (1)) were estimated by running the model N times with global emissions independently perturbed by 1 Gg yr^{-1} for each year in each run. The columns of **E** were set to the resulting changes in mole fraction output by the model.

[14] Global, annual emissions were derived using equation (4), with the initial estimate of annual growth rate taken from EDGAR. There is little information available on

the uncertainty in the EDGAR growth rate, so we assumed that the growth uncertainty was equal to the maximum EDGAR growth. Growth uncertainty estimates could be obtained, for example, by considering the χ^2 distribution of the residuals [e.g., *Tarantola*, 2005; *Michalak et al.*, 2005]. For the period 2005–2008,where there is no EDGAR information at present, we assumed zero emissions growth a priori. The initial atmospheric mole fraction was determined by spinning up the model for 5 years to obtain a realistic vertical and interhemispheric profile, and then solving for an offset to be added to this profile in the inversion (the matrix **D** was modified accordingly so that this "initial condition" element of **x** was not included in the emissions growth constraint scheme).

[15] The observed and simulated mole fractions and estimated emissions for C_3F_8 are shown in Figure 1. We derive a set of emissions which agree well with those found by *Mühle et al.* [2010] and produced modeled mole fractions that agree well with the observations (Figure 1). The normalized posterior covariance matrix is presented in Figure S1.[1]

4. Discussion and Conclusions

[16] Given that prior estimates of emissions of many atmospheric trace species are often unavailable, or highly biased, alternative approaches are required in order to make the often underdetermined problem of emissions estimation invertible. Here we present a solution that uses prior information on emissions growth, rather than absolute emissions. This solution allows time-varying growth and growth uncertainty to be built into the inversion, based on prior expectations about the evolution of emissions.

[17] In the example above, in which emissions of C_3F_8 were estimated, EDGAR would be a poor choice of prior in a Bayesian inversion incorporating absolute emissions, since the inventory emissions are many times smaller than those derived using the measurements. *Mühle et al.* [2010] noted this limitation, and avoided using EDGAR as a prior by implementing a recursive procedure that propagated the estimated emissions forward as the recursion progressed through the data. However, their solution required a posteriori smoothing to remove unrealistic fluctuations from the derived emissions, brought about by measurement outliers. Furthermore, interannual emissions covariance information was not retained in their estimation procedure.

[18] Despite not representing the mean emission rate well, the inventory does contain some useful information on the evolution of emissions of C_3F_8: it appears that although some baseline emissions process is missing from the inventory, EDGAR does capture the onset of an emissions increase in the early 1990s (Figure 1). The onset of this emissions increase occurs during a period when there is a relatively low data density, and therefore incorporation of this information in the inversion is beneficial. The method presented in this article allows such information to be utilized in the inversion without introducing a bias in the mean emission rate and has the advantage that subjective post-inversion averaging can be avoided, as can spurious fluctuations due to measurement outliers, by the use of an explicitly chosen growth constraint. It has also been applied to the other *Mühle et al.* [2010] perfluorocarbons (CF_4 and C_2F_6) with similar results.

[19] If an excellent source of relatively unbiased, independent emissions information is thought to exist with a low estimated uncertainty, then many previous studies have shown that the Bayesian method incorporating absolute emissions information will likely lead to a satisfactory solution. However, we conclude that for many of the studies cited above, for which prior information was either not present or known to be biased [e.g., *Geller et al.*, 1997; *Miller et al.*, 1998; *Mühle et al.*, 2009, 2010], this approach, using constrained growth, can be a useful tool for deriving reasonable emissions time series in a physically justifiable way.

References

Cunnold, D. M., et al. (1994), Global trends and annual releases of CCl_3F and CCl_2F_2 estimated from ALE/GAGE and other measurements from July 1978 to June 1991, *J. Geophys. Res.*, *99*(D1), 1107–1126.

Enting, I. G. (2002), *Inverse Problems in Atmospheric Constituent Transport*, Cambridge Univ. Press, Cambridge, U. K.

Enting, I. G., C. M. Trudinger, and D. M. Etheridge (2006), Propagating data uncertainty through smoothing spline fits, *Tellus, Ser. B*, *58*, 305–309.

European Commission Joint Research Centre and Netherlands Environmental Assessment Agency (2009), Emission database for global atmospheric research (EDGAR), release version 4.0, http://edgar.jrc.ec.europa.eu, EC-JRC/PBL, Brussels.

Geller, L., et al. (1997), Tropospheric SF_6: Observed latitudinal distribution and trends, derived emissions and interhemispheric exchange time, *Geophys. Res. Lett.*, *24*(6), 675–678.

Hansen, P. C. (1992), Analysis of discrete ill-posed problems by means of the L-curve, *SIAM Rev.*, *34*(4), 561–580.

Joos, F., R. Meyer, M. Bruno, and M. Leuenberger (1999), The variability in the carbon sinks as reconstructed for the last 1000 years, *Geophys. Res. Lett.*, *26*(10), 1437–440.

Khasibatla, P., M. Heimann, P. Rayner, N. Mahowald, R. G. Prinn, and D. E. Hartley (Eds.) (2000), *Inverse Methods in Global Biogeochemical Cycles*, Geophys. Monogr. Ser., vol. 114, AGU, Washington D. C.

Maiss, M., and C. A. M. Brenninkmeijer (1998), Atmospheric SF_6: Trends, sources, and prospects, *Environ. Sci. Technol.*, *32*(20), 3077–3086.

Masarie, K. A., and P. P. Tans (1995), Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record, *J. Geophys. Res.*, *100*(D6), 11,593–11,610.

McIntosh, P. C., and G. Veronis (1993), Solving underdetermined tracer inverse problems by spatial smoothing and cross validation, *J. Phys. Oceanogr.*, *23*(4), 716–730.

Michalak, A. M., A. Hirsch, L. Bruhwiler, K. R. Gurney, W. Peters, and P. P. Tans (2005), Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions, *J. Geophys. Res.*, *110*, D24107, doi:10.1029/2005JD005970.

Miller, B. R., J. Huang, R. F. Weiss, R. G. Prinn, and P. J. Fraser (1998), Atmospheric trend and lifetime of chlorodifluoromethane (HCFC-22) and the global tropospheric OH concentration, *J. Geophy. Res.*, *103*(D11), 13,237–13,248.

Miller, B. R., et al. (2010), HFC-23 (CHF_3) emission trend response to HCFC-22 (CHClF2) production and recent HFC-23 emission abatement measures, *Atmos. Chem. Phys.*, *10*, 7875–7890.

Montzka, S. A., L. Kuijpers, M. O. Battle, M. Aydin, K. R. Verhulst, E. S. Saltzman, and D. W. Fahey (2010), Recent increases in global HFC-23 emissions, *Geophys. Res. Lett.*, *37*, L02808, doi:10.1029/2009GL041195.

---

[1]Auxiliary materials are available with the HTML.

Mühle, J., et al. (2009), Sulfuryl fluoride in the global atmosphere, *J. Geophys. Res.*, *114*, D05306, doi:10.1029/2008JD011162.

Mühle, J., et al. (2010), Perfluorocarbons in the global atmosphere: Tetrafluoromethane, hexafluoroethane, and octafluoropropane, *Atmos. Chem. Phys.*, *10*, 5145–5164.

O'Doherty, S., et al. (2004), Rapid growth of hydrofluorocarbon 134a and hydrochlorofluorocarbons 141b, 142b, and 22 from Advanced Global Atmospheric Gases Experiment (AGAGE) observations at Cape Grim, Tasmania, and Mace Head, Ireland, *J. Geophys. Res.*, *109*, D06310, doi:10.1029/2003JD004277.

Oram, D. E., W. T. Sturges, S. A. Penkett, A. McCulloch, and P. J. Fraser (1998), Growth of fluoroform (CHF_3, HFC-23) in the background atmosphere, *Geophys. Res. Lett.*, *25*(1), 35–38.

Prinn, R. G., et al. (2000), A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE, *J. Geophys. Res.*, *105*(D14), 17,751–17,792, doi:10.1029/2000JD900141.

Rigby, M., et al. (2010), History of atmospheric SF6 from 1973 to 2008, *Atmos. Chem. Phys.*, *10*, 10,305–10,320.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, Soc. for Indust. and Appl. Math., Philadelphia.

Tikhonov, A. N., and V. Y. Arsenin (1977), *Solutions of Ill-Posed Problems*, John Wiley, New York.

Weiss, R. F., J. Mühle, P. K. Salameh, and C. M. Harth (2008), Nitrogen trifluoride in the global atmosphere, *Geophys. Res. Lett.*, *35*, L20821, doi:10.1029/2008GL035913.

Wunsch, C. (2006), *Discrete Inverse and State Estimation Problems: With Geophysical Fluid Applications*, Cambridge Univ. Press, Cambridge, U. K.

Xiao, X., et al. (2007), Optimal estimation of the soil uptake rate of molecular hydrogen from the Advanced Global Atmospheric Gases Experiment and other measurements, *J. Geophys. Res.*, *112*, D07303, doi:10.1029/2006JD007241.

A. L. Ganesan, R. G. Prinn, and M. Rigby, Center for Global Change Science, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA. (aganesan@mit.edu; rprinn@mit.edu; mrigby@mit.edu)