



UNIVERSITY
OF WOLLONGONG
AUSTRALIA

University of Wollongong
Research Online

National Institute for Applied Statistics Research
Australia Working Paper Series

Faculty of Engineering and Information Sciences

2016

Statistical bias and variance for the regularized-inverse problem: application to Space-based atmospheric CO₂ retrievals

N Cressie

University of Wollongong

R Wang

The Ohio State University

M Smyth

Jet Propulsion Laboratory

C E. Miller

Jet Propulsion Laboratory

Recommended Citation

Cressie, N; Wang, R; Smyth, M; and Miller, C E., Statistical bias and variance for the regularized-inverse problem: application to Space-based atmospheric CO₂ retrievals, National Institute for Applied Statistics Research Australia, University of Wollongong, Working Paper 04-16, 2016, 22.

<http://ro.uow.edu.au/niasrawp/41>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Statistical bias and variance for the regularized-inverse problem: application to Space-based atmospheric CO₂ retrievals

Abstract

Remote sensing of the atmosphere is typically achieved through measurements that are high-resolution radiance spectra. In this article, our goal is to characterize the first-moment and second-moment properties of the errors obtained when solving the regularized inverse problem associated with space-based atmospheric CO₂ retrievals, specifically for the dry air mole fraction in a column of the atmosphere. The problem of estimating (or retrieving) state variables is usually ill-posed, leading to a solution based on regularization that is often called Optimal Estimation (OE). The difference between the estimated state and the true state is defined to be the retrieval error; error analysis for OE uses a linear approximation to the forward model, resulting in a calculation where the first moment of the retrieval error (the bias) is identically zero. This is inherently unrealistic and not seen in real or simulated retrievals. Non-zero bias is expected since the forward model of radiative transfer is strongly nonlinear in the atmospheric state. In this article, we extend and improve OE's error analysis based on a first-order, multivariate Taylor-series expansion, by including the second-order terms in the expansion. Specifically, we approximate the bias through the second derivative of the forward model, which results in a formula involving the Hessian array. We propose a stable estimate of it, from which we obtain a second-order expression for the bias and mean squared prediction error of the retrieval.

NIASRA

NATIONAL INSTITUTE FOR APPLIED
STATISTICS RESEARCH AUSTRALIA



***National Institute for Applied Statistics Research
Australia***

University of Wollongong, Australia

Working Paper

04-16

Statistical Bias and Variance for the
Regularized-inverse Problem: Application to
Space-based Atmospheric CO₂ Retrievals

N. Cressie, R. Wang, M. Smyth, and C. E. Miller

This work has been submitted for publication. Copyright in this work may be transferred without further notice, and this version may no longer be accessible.

National Institute for Applied Statistics Research Australia, University of Wollongong,
Wollongong NSW 2522, Australia. Phone +61 2 4221 5076, Fax +61 2 4221 4998.
Email: karink@uow.edu.au

Statistical bias and variance for the regularized-inverse problem: Application to space-based atmospheric CO_2 retrievals

N. Cressie*, R. Wang[†], M. Smyth[‡] and C. E. Miller[‡]

Abstract

Remote sensing of the atmosphere is typically achieved through measurements that are high-resolution radiance spectra. In this article, our goal is to characterize the first-moment and second-moment properties of the errors obtained when solving the regularized inverse problem associated with space-based atmospheric CO_2 retrievals, specifically for the dry air mole fraction in a column of the atmosphere. The problem of estimating (or retrieving) state variables is usually ill-posed, leading to a solution based on regularization that is often called Optimal Estimation (OE). The difference between the estimated state and the true state is defined to be the retrieval error; error analysis for OE uses a linear approximation to the forward model, resulting in a calculation where the first moment of the retrieval error (the bias) is identically zero. This is inherently unrealistic and not seen in real or simulated retrievals. Non-zero bias is expected since the forward model of radiative transfer is strongly nonlinear in the atmospheric state. In this article, we extend and improve OE's error analysis based on a first-order, multivariate Taylor-series expansion, by inducing the second-order terms in the expansion. Specifically, we approximate the bias through the second derivative of the forward model, which results in a formula involving the Hessian array. We propose a stable estimate of it, from which we obtain a second-order expression for the bias and mean squared prediction error of the retrieval.

Keypoints

*National Institute for Applied Statistics Research Australia (NIASRA), University of Wollongong NSW 2522, Australia, and Distinguished Visiting Scientist, Jet Propulsion Laboratory, USA. Email: ncressie@uow.edu.au

[†]Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH 43210-1247, USA.

[‡]Jet Propulsion Laboratory, NASA, Pasadena, CA 91109, USA.

- The retrieval of XCO₂ is a nonlinear ill-posed inverse problem with non-zero bias.
- First-moment and second-moment statistical properties of atmospheric CO₂ retrievals from satellite remote sensing instruments are approximated.
- The approximations are assessed in a realistic simulation experiment and found to perform well.

1 Introduction

Remote sensing of the atmosphere by satellites is typically achieved through a combination of physics and statistics. The physics is captured through a forward function, yet every retrieval is recognized to have error associated with it. This article presents a statistical approach to estimating the mean and variance of the retrieval error (the retrieved state vector minus the true state vector) from radiances that are connected to the state through a forward function that is modeled using a simpler forward model. Our goal here is retrieval of atmospheric CO₂, and we work within the framework of *Rodgers (2000)*, using a nonlinear forward model and Bayesian inverse methods; this has sometimes been called Optimal Estimation (OE).

The error analysis given by *Rodgers (2000)* is a first-order analysis that proceeds as if the forward model is linear, and it always gives the result that the mean of the retrieval error (i.e., bias) is identically zero. However, because the forward model is nonlinear, there is generally a non-zero bias that we henceforth call *nonlinearity bias*.

In this article, we develop expressions that quantify the nonlinear contributions to atmospheric CO₂ retrieval errors; they are based on capturing nonlinearity through the Hessian array, which is the second derivative of the forward model. Its stable estimation is critical, and a key contribution of this article is to present a weighted estimate that preserves symmetry properties of the Hessian array. Our results apply to any retrieval obtained by solving an inverse problem through regularization, and hence they are applicable to the many remote sensing retrievals that are based on a radiative transfer function.

In the context of remote sensing of atmospheric CO₂, there have been a number of articles discussing OE of the state based on instruments such as SCIAMACHY, GOSAT, AIRS, and OCO-2; see, for example, *Böesch et al. (2006)*; *Connor et al. (2008)*; *Kuze et al. (2009)*; *Bréon and Ciais (2010)*; *Böesch et al. (2011)*; *Crisp et al. (2012)*; *O'Dell et al. (2012)*; *Cressie and Wang (2013)*. In this article, we develop statistical methodology for a second-order error analysis that recognizes nonlinearity of the forward model that connects the state of atmosphere with

high-resolution radiance spectra measured by the remote sensing instruments on board GOSAT and OCO-2 (e.g., *Crisp et al.* 2014). The full multivariate distribution could be obtained through a Markov chain Monte Carlo (MCMC) algorithm (e.g., *Tamminen and Kyrola* 2001; *Tamminen* 2004), but the computing time for one retrieval takes on the order of a day.

In Section 2, we present details of the experiment we conducted to look at the properties of the retrieval error, including bias due to nonlinearity of the forward model. Section 3 briefly reviews OE and its error analysis. Section 4 presents our statistical methodology that recognizes and estimates bias due to the inherent nonlinearity of the forward model. This is done through a second-order, multivariate Taylor-series expansion resulting in the Hessian array; using the approach of statistical estimating equations, a stable estimate of the Hessian array is derived. In Section 5, we present the results of the experiment described in Section 2. Discussion and conclusions are presented in Section 6.

2 Synthetic OCO-2 Data: A Controlled Experiment

This section describes an experiment we conducted to determine the influence of nonlinearity on the mean and variance of the retrieval error. More details are given in Section 5, where the results of the experiment are presented. The experiment involves simulations from a known nonlinear forward model, where the time atmospheric state is known (but not used in the retrieval).

Radiances are simulated that are typical of those seen by GOSAT, with an ACOS forward model (*Crisp et al.*, 2012) that emulates retrievals from OCO-2: The atmospheric state is first simulated; then the nonlinear forward model is applied (here, Version B5.0 of the ACOS/OCO-2 forward model); and finally measurement error is added to yield synthetic high-resolution radiance spectra. Critically, because in the experiment the state is known, the retrieval error can be obtained exactly. From a large number of such simulations, the distribution of the retrieval error can be obtained. In this article, our goal is to obtain estimates of the dry air mole fraction of CO_2 in a column (XCO_2), which is the key quantity used in flux inversion, along with its uncertainty quantification.

Repeating this under different scenarios, most notably location/season (varying albedo, aerosol optical depth, and vertical distributions), “aerosols,” and “clear sky,” generates scenario-specific distributions for the retrieval errors. Recall that the mean of the retrieval error is simply the bias due to the nonlinearity of the forward model; that bias must be zero if the forward model is linear.

Based on a pilot study to determine the important geophysical factors to con-

sider, the aerosol/clear sky factor emerged as the most sensitive to nonlinearity bias. Hence, we chose 18 GOSAT locations over Australia during different seasons in order to generate a variety of atmospheres (see Figure 1), and we simulated with aerosols present or not. The mean of the state was obtained from climatology, as is done for OCO-2 retrievals (*Crisp et al.*, 2014). Importantly, we generated forward, using a multivariate Gaussian distribution to obtain 700 simulations of the atmospheric state for each scenario (i.e., for a given location/season and given presence/absence of aerosols). Relative Monte Carlo accuracy to the first decimal place is achieved with this number of simulations.

For each simulated state, a radiance vector was simulated, again by generating forward from version B5.0 of ACOS/OCO-2’s full-physics forward model (*Crisp et al.*, 2014) and random noise was added to capture errors due to measurement by the instrument on board the satellite. The end result was that, for each scenario, 700 synthetic radiance vectors were available for analysis.

For each radiance vector, the next step in the experiment is to retrieve the atmospheric state, particularly the column-averaged dry-air mole fraction (XCO₂) in parts per million (ppm). To exercise strong control on the experiment, we use *exactly* the same forward model and covariances for the retrieval as were used in the simulations. This avoids confounding errors due to model misspecification with those due to nonlinearity, when attributing discrepancies found in retrieval-error properties.

Consequently, for each of the 700 radiance vectors, there is an estimated state vector obtained from the level-2 algorithm, B5.0. Although not used in the retrieval, we also have available the respective true state vectors, from which we can trivially construct, by subtraction, 700 retrieval-error vectors (= estimated state vector – true state vector). In practice, the true state vector is never known, and one must use statistical methodology to obtain properties (e.g., first and second moments, or all moments through MCMC) of the retrieval-error distribution. Hence, the distribution of retrieval error for each scenario, obtained from the simulation, represents the gold standard against which all error analyses, linear and nonlinear, can be compared.

In the next section, we establish the necessary notation to discuss the state-space model and the retrieval algorithm obtained from it. Finding the first-moment and second-moment statistical properties of the retrieval error of the multivariate state is the focus of this article; Section 5 gives the results obtained from the controlled experiment described above.

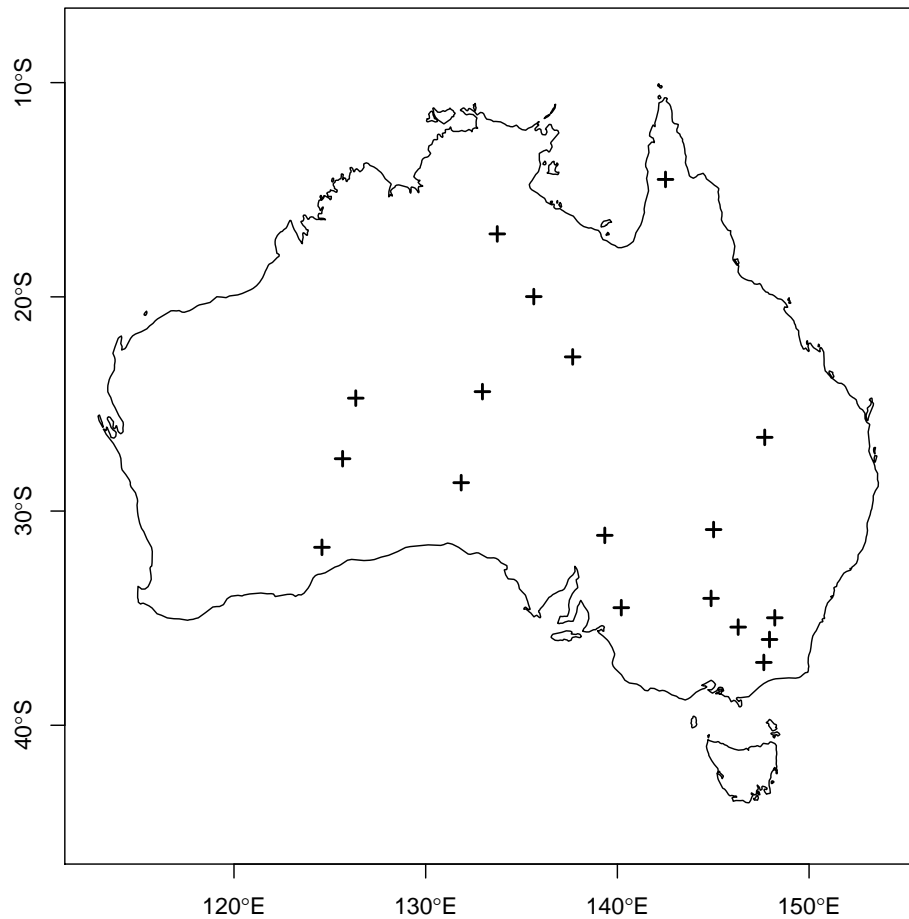


Figure 1: Map showing the 18 locations at which retrievals were performed during different seasons. At each location, 700 atmospheres and radiances were simulated using climatology, a full-physics forward model, and known measurement-errors characteristics.

3 Radiative Transfer Function and Optimal Estimation

In this section, we describe briefly the statistical model given in *Rodgers (2000)*, and we summarize the specific implementation used for the ACOS/OCO-2 retrievals. The first part of the statistical model is based on the physics of radiative transfer that links measured radiances to the physical state of a column of the atmosphere sampled by the light path. It is here that the measurement uncertainty is accounted for. The second part of the statistical model recognizes that the atmospheric state vector (e.g., CO_2 volume mixing ratios at different pressure levels, surface pressure, aerosols, albedo, and so forth) is not completely certain, and its uncertainty is described by a statistical distribution with a given mean vector and a given covariance matrix. In the terminology of OE, these are called the *prior* mean vector and the *prior* covariance matrix, respectively. In the terminology of state-space modeling, these are called the mean and covariance of *the state*, respectively.

3.1 State-Space Model

As explained above, the statistical model for retrieving the atmospheric state (\mathbf{x}) from measured radiances (\mathbf{y}) is divided into two parts. The first part is the measurement equation (relating \mathbf{y} to \mathbf{x}), and the second part is the state equation (capturing the variability of the state \mathbf{x}); see, for example, *Shumway and Stoffer (2006)*, for a description of state-space modeling and estimation. The measurement equation, which includes the (typically nonlinear) forward model $\mathbf{F}(\mathbf{x})$, is

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \boldsymbol{\varepsilon}. \quad (1)$$

In (1), $\boldsymbol{\varepsilon}$ is an n_ε -dimensional error vector that captures both measurement error in the radiances \mathbf{y} and specification error incurred by approximating the physics of radiative transfer through a generally nonlinear “working function” (or forward model) $\mathbf{F}(\cdot)$. For the remote-sensing application considered in Section 5, $n_\varepsilon = 2040$. Assume that $\boldsymbol{\varepsilon} \sim \text{Dist}(\mathbf{0}, \mathbf{S}_\varepsilon)$, where $\mathbf{S}_\varepsilon \equiv \text{cov}(\boldsymbol{\varepsilon})$ is a given $n_\varepsilon \times n_\varepsilon$ covariance matrix of measurement errors and $\text{Dist}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a generic multivariate distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; there may be other parameters that determine the distribution, but our interest centers on the first two. For example, *Dist* is often chosen to be the multivariate Gaussian (*Gau*) distribution. *Cressie and Wang (2013)* consider the general case where $E(\boldsymbol{\varepsilon}) \neq \mathbf{0}$, and they show how the data \mathbf{y} can be corrected to account for a non-zero mean. Hence, we can assume without loss of generality that $E(\boldsymbol{\varepsilon}) = \mathbf{0}$.

The state \mathbf{x} is an n_α -dimensional vector (typically, $n_\alpha \ll n_\varepsilon$), and it is not known exactly; the state equation (called the prior model in the terminology of OE) expresses this uncertainty through,

$$\mathbf{x} = \mathbf{x}_\alpha + \boldsymbol{\alpha}, \quad (2)$$

where \mathbf{x}_α is a known prior-mean specification of the state, $\boldsymbol{\alpha}$ is an error vector, and all vectors in (2) are n_α -dimensional. It is assumed that $\boldsymbol{\alpha} \sim \text{Dist}(\mathbf{0}, \mathbf{S}_\alpha)$, where $\mathbf{S}_\alpha \equiv \text{cov}(\boldsymbol{\alpha})$ is a given $n_\alpha \times n_\alpha$ covariance matrix, and note that $\text{Dist}(\cdot, \cdot)$ for $\boldsymbol{\alpha}$ may be different from that for $\boldsymbol{\varepsilon}$ in (1). For the remote sensing application considered in Section 5, $n_\alpha = 50$.

We note that there remain parameters in the forward model and variance-covariance parameters in (1) and (2) that need to be specified or estimated. A fully Bayesian approach would put prior distributions on them; in this article, we assume that the measurement and state equations are completely known, as does *Rodgers* (2000) and those that follow.

The inverse problem is to infer the state \mathbf{x} from data \mathbf{y} ; this results in the estimated state $\hat{\mathbf{x}}$, which in remote sensing is typically used to compute geophysical parameters of the atmosphere (e.g., XCO₂). The ACOS/OCO-2 algorithm uses Twomey-Tikhonov regularization (*Tikhonov*, 1963; *Twomey*, 1963) to solve the problem; see Section 3.2.

Let $\{x(p): 0 \leq p \leq P_{su}\}$ denote the CO_2 volume mixing ratios (VMRs) for every pressure level p from the surface pressure P_{su} to the top of the atmosphere. Then, for pre-determined pressure levels, $0 \leq p_1 < p_2 < \dots < p_{n_P} \leq P_{su}$, we define the n_P -dimensional subvector,

$$\mathbf{x}_P \equiv \begin{bmatrix} x(p_1) \\ \vdots \\ x(p_{n_P}) \end{bmatrix}. \quad (3)$$

We then partition the full state vector \mathbf{x} into two (or possibly more) components as:

$$\mathbf{x} \equiv \begin{bmatrix} \mathbf{x}_P \\ \mathbf{x}_Q \end{bmatrix}, \quad (4)$$

where \mathbf{x}_P is that part of the state vector pertaining to the CO_2 profile, and \mathbf{x}_Q represents all non- CO_2 state-vector elements, including the surface pressure, surface albedo, profiles of H_2O , temperature, aerosols, and so forth. Using obvious notation, we see that the dimension of \mathbf{x}_Q is $n_Q = n_\alpha - n_P$. Hence, we can partition the prior mean as,

$$\mathbf{x}_\alpha \equiv \begin{bmatrix} \mathbf{x}_{P,\alpha} \\ \mathbf{x}_{Q,\alpha} \end{bmatrix}, \quad (5)$$

and the prior covariance matrix as,

$$\mathbf{S}_\alpha \equiv \begin{bmatrix} \mathbf{S}_{PP,\alpha} & \mathbf{S}_{PQ,\alpha} \\ \mathbf{S}'_{PQ,\alpha} & \mathbf{S}_{QQ,\alpha} \end{bmatrix}. \quad (6)$$

As will be seen in Section 5, it is useful to have the definitions (3)–(6), in order to extract statistical properties of estimated CO_2 values from the statistical properties (given in Section 4) of the full state vector \mathbf{x} .

3.2 The Retrieval

Rodgers (2000) proposes an OE (often called a Bayesian) solution to the retrieval problem, which we summarize below. The *maximum a posteriori* (MAP) estimator is defined to be the state (or set of states, if it is not unique) $\hat{\mathbf{x}}$ such that the posterior distribution evaluated at $\hat{\mathbf{x}}$ satisfies:

$$P(\hat{\mathbf{x}}|\mathbf{y}) \geq P(\mathbf{x}|\mathbf{y}), \quad (7)$$

for all possible states \mathbf{x} , where $P(\mathbf{x}|\mathbf{y})$ denotes the posterior distribution of the state \mathbf{x} given the data \mathbf{y} . That is, the state $\hat{\mathbf{x}}$ in (7) is the *posterior mode*.

If the random vectors in both (1) and (2) are Gaussian then, up to an additive constant, $-2 \log P(\mathbf{x}|\mathbf{y})$ is equal to

$$(\mathbf{y} - \mathbf{F}(\mathbf{x}))' \mathbf{S}_\varepsilon^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x})) + (\mathbf{x} - \mathbf{x}_\alpha)' \mathbf{S}_\alpha^{-1} (\mathbf{x} - \mathbf{x}_\alpha), \quad (8)$$

and finding the posterior mode is equivalent to *minimizing* (8) with respect to \mathbf{x} . If the random vectors ε in (1) and α in (2) are more generically distributed according to *Dist*, the criterion (8) can still be used, since it solves the general problem of regularization formulated independently by *Tikhonov* (1963) and *Twomey* (1963).

Define the *Jacobian*,

$$\mathbf{K}(\mathbf{x}) \equiv \frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}}, \quad (9)$$

which is a matrix of order $n_\varepsilon \times n_\alpha$. Minimizing expression (8) implies that \mathbf{x} satisfies,

$$-\mathbf{K}(\mathbf{x})' \mathbf{S}_\varepsilon^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x})) + \mathbf{S}_\alpha^{-1} (\mathbf{x} - \mathbf{x}_\alpha) = \mathbf{0}, \quad (10)$$

which consists of n_α equations in n_α unknowns. The Gauss-Newton iteration scheme to solve (10) results in:

$$\mathbf{x}^{(\ell+1)} = \mathbf{x}^{(\ell)} + \{\mathbf{S}_\alpha^{-1} + (\mathbf{K}^{(\ell)})' \mathbf{S}_\varepsilon^{-1} \mathbf{K}^{(\ell)}\}^{-1} [(\mathbf{K}^{(\ell)})' \mathbf{S}_\varepsilon^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x}^{(\ell)})) - \mathbf{S}_\alpha^{-1} (\mathbf{x}^{(\ell)} - \mathbf{x}_\alpha)], \quad (11)$$

where $\mathbf{K}^{(\ell)} \equiv \mathbf{K}(\mathbf{x}^{(\ell)})$ and $\ell = 1, 2, \dots$. Equation (11) is iterated for increasing $\ell = 1, 2, \dots$ until convergence, resulting in the estimated state vector, $\hat{\mathbf{x}}$. Now (11)

can be unstable, so the OCO-2 algorithm uses a Levenberg-Marquardt approach (Levenberg, 1944; Marquardt, 1963) to solving (10).

By way of introduction to the next section, we would like to point out that the Jacobian matrix, $\mathbf{K}(\mathbf{x})$, is a first-order quantity, obtained from a first-order, multivariate Taylor-series expansion; it is the first derivative of the forward model $\mathbf{F}(\mathbf{x})$, and it captures the locally linear behavior of \mathbf{F} . However, if the Taylor-series expansion is extended to second-order, the second derivative of $\mathbf{F}(\mathbf{x})$, which is called the *Hessian array* $\{H_{ijk}(\mathbf{x})\}$, is needed to obtain a second-order approximation of the bias of the retrieval. Since the Hessian array is the second derivative of the n_ε -dimensional vector $\mathbf{F}(\mathbf{x})$, it is defined for $i = 1, \dots, n_\varepsilon$, $j = 1, \dots, n_\alpha$, and $k = 1, \dots, n_\alpha$. For the remote sensing application considered in Section 5, the number of entries in the Hessian array is $2040 \times 50 \times 50$.

4 Bias and Mean Squared Prediction Error

In general, the forward model is *nonlinear*, and a statistical error analysis of the optimally estimated state $\hat{\mathbf{x}}$ should recognize the consequences of this. In this section, the *delta method* (e.g., Meyer, 1975, Ch. 10) is used to obtain approximations for the bias vector, $E(\hat{\mathbf{x}} - \mathbf{x})$, and the mean squared prediction error (MSPE) matrix, $E\{(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})'\}$. Nonlinearity in the forward model leads to *non-zero bias* of $\hat{\mathbf{x}}$, which the OE error analysis of Rodgers (2000) neglects by assuming it is zero.

The converged solution to (11) (or the Levenberg-Marquardt version of this) satisfies:

$$\begin{aligned}\hat{\mathbf{x}} &= \mathbf{x}_\alpha + \mathbf{G}(\hat{\mathbf{x}})\{(\mathbf{y} - \mathbf{F}(\hat{\mathbf{x}})) + \mathbf{K}(\hat{\mathbf{x}})(\hat{\mathbf{x}} - \mathbf{x}_\alpha)\} \\ &= \mathbf{x}_\alpha + \mathbf{G}(\hat{\mathbf{x}})(\mathbf{F}(\mathbf{x}) - \mathbf{F}(\hat{\mathbf{x}})) + \mathbf{A}(\hat{\mathbf{x}})(\hat{\mathbf{x}} - \mathbf{x}_\alpha) + \mathbf{G}(\hat{\mathbf{x}})\boldsymbol{\varepsilon},\end{aligned}\quad (12)$$

where

$$\begin{aligned}\mathbf{G}(\mathbf{x}) &\equiv \{\mathbf{S}_\alpha^{-1} + \mathbf{K}(\mathbf{x})'\mathbf{S}_\varepsilon^{-1}\mathbf{K}(\mathbf{x})\}^{-1}\mathbf{K}(\mathbf{x})'\mathbf{S}_\varepsilon^{-1} \\ \mathbf{A}(\mathbf{x}) &\equiv \mathbf{G}(\mathbf{x})\mathbf{K}(\mathbf{x}).\end{aligned}$$

The quantities $\mathbf{G}(\mathbf{x})$ and $\mathbf{A}(\mathbf{x})$ are called the *gain matrix* and the *averaging-kernel matrix*, respectively, and in general they are functions of the true state \mathbf{x} .

4.1 MSPE Matrix

Cressie and Wang (2013) use the delta method in the multivariate setting to show that to second order,

$$\text{cov}(\hat{\mathbf{x}}) \simeq \tilde{\text{cov}}(\hat{\mathbf{x}}) \equiv \mathbf{A}(\mathbf{x}_\alpha)\mathbf{S}_\alpha\mathbf{A}(\mathbf{x}_\alpha)' + \mathbf{G}(\mathbf{x}_\alpha)\mathbf{S}_\varepsilon\mathbf{G}(\mathbf{x}_\alpha)', \quad (13)$$

and

$$\begin{aligned}\text{cov}(\hat{\mathbf{x}}, \mathbf{x}) &\simeq \tilde{\text{cov}}(\hat{\mathbf{x}}, \mathbf{x}) \equiv \mathbf{A}(\mathbf{x}_\alpha)\mathbf{S}_\alpha \\ \text{cov}(\hat{\mathbf{x}}, \boldsymbol{\varepsilon}) &\simeq \tilde{\text{cov}}(\hat{\mathbf{x}}, \boldsymbol{\varepsilon}) \equiv \mathbf{G}(\mathbf{x}_\alpha)\mathbf{S}_\varepsilon.\end{aligned}$$

When the forward model $\mathbf{F}(\cdot)$ is *linear*, $\mathbf{A}(\mathbf{x}_\alpha)$ and $\mathbf{G}(\mathbf{x}_\alpha)$ do not depend on \mathbf{x}_α , and “ \simeq ” becomes “=” in the three relations above.

The *delta method* in statistics involves a Taylor-series expansion, usually in a univariate setting, but here for a vector. We give the expansion of the retrieval $\hat{\mathbf{x}}$ about the prior mean \mathbf{x}_α , although our results are true for the expansion about any fixed vector \mathbf{x}^0 . If the prior mean \mathbf{x}_α is believed to be less reliable due to insufficient knowledge of the statistical structure of the true state, a better choice for the Taylor-series expansion might be a starting vector of the iteration that solves (10). (We note that expanding the Taylor series about the retrieved vector $\hat{\mathbf{x}}$ does not give valid results for the bias and MSPE, since this is a *random* vector. Many implementations of OE in fact do this; see *O’Dell et al. (2012)*).

To second order, the MSPE matrix can be approximated as follows: $E\{(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})'\} \simeq \tilde{\text{MSPE}}(\mathbf{x}_\alpha)$, where

$$\tilde{\text{MSPE}}(\mathbf{x}_\alpha) \equiv (\mathbf{A}(\mathbf{x}_\alpha) - \mathbf{I})\mathbf{S}_\alpha(\mathbf{A}(\mathbf{x}_\alpha) - \mathbf{I})' + \mathbf{G}(\mathbf{x}_\alpha)\mathbf{S}_\varepsilon\mathbf{G}(\mathbf{x}_\alpha)', \quad (14)$$

which is derived in *Cressie and Wang (2013)*. Once again, when the forward model $\mathbf{F}(\cdot)$ is linear, $\mathbf{A}(\mathbf{x}_\alpha)$ and $\mathbf{G}(\mathbf{x}_\alpha)$ do not depend on \mathbf{x}_α , “ \simeq ” becomes “=,” and the MSPE matrix is exactly equal to (14). Using standard matrix algebra, it is straightforward to show that (14) is equivalent to

$$(\mathbf{K}(\mathbf{x})'\mathbf{S}_\varepsilon^{-1}\mathbf{K}(\mathbf{x}) + \mathbf{S}_\alpha^{-1})^{-1} \equiv \hat{\mathbf{S}},$$

which OCO-2 calls the error covariance matrix (*O’Dell et al., 2012; Crisp et al., 2014*). Strictly speaking, $\hat{\mathbf{S}}$ should be called the *mean squared prediction error matrix*; it is easy to see that $\hat{\mathbf{S}}$ is the *error covariance matrix* only when the bias vector is zero.

4.2 Bias Vector

Because of the nonlinearity of the forward model, $\hat{\mathbf{x}}$ will be a *biased* estimate of the state vector \mathbf{x} . That is, the retrieval error, $\hat{\mathbf{x}} - \mathbf{x}$, will have a distribution that is offset from $\mathbf{0}$. To obtain the bias, $E(\hat{\mathbf{x}} - \mathbf{x})$, one can again use the delta method by expanding the mean of $\hat{\mathbf{x}}$ around its prior mean \mathbf{x}_α and keeping terms of the expansion up to second order. Then, to second order, the bias vector can

be approximated as follows: $E(\hat{\mathbf{x}} - \mathbf{x}) \simeq \tilde{\mathbf{bias}}(\mathbf{x}_\alpha)$, where

$$\begin{aligned} \tilde{\mathbf{bias}}(\mathbf{x}_\alpha) \equiv & (1/2) \begin{pmatrix} \left(\text{vec} \left(\frac{\partial \mathbf{A}(\mathbf{x}_\alpha)_{\text{1st-row}}}{\partial \mathbf{x}_\alpha} \right) \right)' \\ \left(\text{vec} \left(\frac{\partial \mathbf{A}(\mathbf{x}_\alpha)_{\text{2nd-row}}}{\partial \mathbf{x}_\alpha} \right) \right)' \\ \vdots \\ \left(\text{vec} \left(\frac{\partial \mathbf{A}(\mathbf{x}_\alpha)_{n_\alpha \text{th-row}}}{\partial \mathbf{x}_\alpha} \right) \right)' \end{pmatrix} \cdot \text{vec} \left(\tilde{\mathbf{MSPE}}(\mathbf{x}_\alpha) + 2\mathbf{S}_\alpha \mathbf{A}(\mathbf{x}_\alpha)' \right) \\ & - (1/2) \begin{pmatrix} \left(\text{vec} \left(\frac{\partial \mathbf{G}(\mathbf{x}_\alpha)_{\text{1st-row}}}{\partial \mathbf{x}_\alpha} \right) \right)' \\ \left(\text{vec} \left(\frac{\partial \mathbf{G}(\mathbf{x}_\alpha)_{\text{2nd-row}}}{\partial \mathbf{x}_\alpha} \right) \right)' \\ \vdots \\ \left(\text{vec} \left(\frac{\partial \mathbf{G}(\mathbf{x}_\alpha)_{n_\alpha \text{th-row}}}{\partial \mathbf{x}_\alpha} \right) \right)' \end{pmatrix} \cdot \text{vec} \left(\mathbf{K}(\mathbf{x}_\alpha) \cdot \tilde{\mathbf{MSPE}}(\mathbf{x}_\alpha) - 2\mathbf{S}_\varepsilon \mathbf{G}(\mathbf{x}_\alpha)' \right), \end{aligned} \quad (15)$$

which is derived in *Cressie and Wang* (2013). In (15), the ‘vec’ operator is defined as follows: For any $m \times n$ matrix $\mathbf{B} \equiv (\mathbf{B}_1, \dots, \mathbf{B}_m)$ with n -dimensional columns $\mathbf{B}_1, \dots, \mathbf{B}_m$, $\text{vec}(\mathbf{B})$ is defined as the mn -dimensional vector, $\text{vec}(\mathbf{B}) \equiv (\mathbf{B}'_1, \dots, \mathbf{B}'_m)'$. We call $\tilde{\mathbf{bias}}(\mathbf{x}_\alpha)$, defined by (15), the *nonlinearity bias*.

A consequence of the forward model $\mathbf{F}(\mathbf{x})$ being nonlinear is that $\mathbf{K}(\mathbf{x})$, $\mathbf{G}(\mathbf{x})$, and $\mathbf{A}(\mathbf{x})$ depend on the true state \mathbf{x} , and hence their derivatives with respect to \mathbf{x} are generally non-zero. When the forward model is linear, \mathbf{K} , \mathbf{G} , and \mathbf{A} are constant, and hence $E(\hat{\mathbf{x}} - \mathbf{x}) = \mathbf{0} = \tilde{\mathbf{bias}}(\mathbf{x}_\alpha)$. In the next section, we see that a non-zero value for expression (15) is directly attributable to the Hessian array, which recall is the second derivative of the forward model.

4.3 The Hessian

In the bias formula (15), the partial derivatives, $\frac{\partial \mathbf{G}(\mathbf{x})}{\partial x_k}$ and $\frac{\partial \mathbf{A}(\mathbf{x})}{\partial x_k}$, are functions of the derivatives of the Jacobian, $\left\{ \frac{\partial \mathbf{K}(\mathbf{x})}{\partial x_k} : k = 1, \dots, n_\alpha \right\}$; see the expressions just below (12). Specifically,

$$\begin{aligned} \frac{\partial \mathbf{G}(\mathbf{x})}{\partial x_k} = & - \left\{ \mathbf{S}_\alpha^{-1} + \mathbf{K}(\mathbf{x})' \mathbf{S}_\varepsilon^{-1} \mathbf{K}(\mathbf{x}) \right\}^{-1} \left[\frac{\partial \mathbf{K}(\mathbf{x})'}{\partial x_k} \mathbf{S}_\varepsilon^{-1} \mathbf{K}(\mathbf{x}) + \mathbf{K}(\mathbf{x})' \mathbf{S}_\varepsilon^{-1} \frac{\partial \mathbf{K}(\mathbf{x})}{\partial x_k} \right] \times \\ & \left\{ \mathbf{S}_\alpha^{-1} + \mathbf{K}(\mathbf{x})' \mathbf{S}_\varepsilon^{-1} \mathbf{K}(\mathbf{x}) \right\}^{-1} \mathbf{K}(\mathbf{x})' \mathbf{S}_\varepsilon^{-1} + \end{aligned}$$

$$\{\mathbf{S}_\alpha^{-1} + \mathbf{K}(\mathbf{x})' \mathbf{S}_\varepsilon^{-1} \mathbf{K}(\mathbf{x})\}^{-1} \frac{\partial \mathbf{K}(\mathbf{x})'}{\partial x_k} \mathbf{S}_\varepsilon^{-1}, \quad (16)$$

and

$$\frac{\partial \mathbf{A}(\mathbf{x})}{\partial x_k} = \frac{\partial \mathbf{G}(\mathbf{x})}{\partial x_k} \mathbf{K}(\mathbf{x}) + \mathbf{G}(\mathbf{x}) \frac{\partial \mathbf{K}(\mathbf{x})}{\partial x_k}. \quad (17)$$

Thus, calculation of the partial derivatives of the Jacobian, $\left\{ \frac{\partial \mathbf{K}(\mathbf{x})}{\partial x_k} : k = 1, \dots, n_\alpha \right\}$, is needed for calculating the nonlinearity bias given by (15). Notice that because $\mathbf{K}(\mathbf{x})$ is an $n_\varepsilon \times n_\alpha$ matrix, the array just above has $n_\varepsilon \times n_\alpha \times n_\alpha$ elements.

Now, since the Jacobian is itself a first derivative, any element in the array can be written as,

$$H_{ijk}(\mathbf{x}) \equiv \frac{\partial^2 F_i(\mathbf{x})}{\partial x_j \partial x_k}; i = 1, \dots, n_\varepsilon, j = 1, \dots, n_\alpha, k = 1, \dots, n_\alpha. \quad (18)$$

We call $\{H_{ijk}(\mathbf{x})\}$ the *Hessian array*; the subscript i ranges over n_ε elements, which for the OCO-2 instrument come from three spectral bands (Oxygen A band, Weak CO_2 band, and Strong CO_2 band); and the subscripts j and k each range over n_α elements.

Figure 2 illustrates the structure of the *Hessian array*.

An important property of the Hessian array (18) is that it is *symmetric* in j and k . Thus, any estimate of it should preserve this property, namely $H_{ijk}(\mathbf{x}) = H_{ikj}(\mathbf{x})$. We achieve this below by using the approach of statistical estimating equations.

The Hessian element $H_{ijk}(\cdot)$ can be estimated by taking a numerical derivative of the Jacobian, where the Jacobian is evaluated analytically: Let $K_{ij}(\mathbf{x})$ denote the (i, j) -th element of the $n_\varepsilon \times n_\alpha$ Jacobian matrix $\mathbf{K}(\mathbf{x})$, and define

$$\tilde{H}_{ijk}(\mathbf{x}) \equiv \frac{K_{ij}(\mathbf{x} + \Delta_k \mathbf{e}_k) - K_{ij}(\mathbf{x})}{\Delta_k}, \quad (19)$$

for $i = 1, \dots, n_\varepsilon$, and $j, k = 1, \dots, n_\alpha$. In the expression (19), \mathbf{e}_k is the vector with 1 as the k -th element and 0 everywhere else, and $\Delta_k > 0$ is a small increment. In general, symmetry of the estimate (19) does not hold; that is, $\tilde{H}_{ijk}(\cdot) \neq \tilde{H}_{ikj}(\cdot)$.

The following approach to estimating $\{H_{ijk}(\mathbf{x})\}$ uses statistical estimating equations, and it will guarantee symmetry in j and k : There are two estimating equations for $H_{ijk}(\mathbf{x})$ that follow from (19), namely

$$\Delta_k H_{ijk}(\mathbf{x}) = K_{ij}(\mathbf{x} + \Delta_k \mathbf{e}_k) - K_{ij}(\mathbf{x}),$$

and

$$\Delta_j H_{ijk}(\mathbf{x}) = K_{ik}(\mathbf{x} + \Delta_j \mathbf{e}_j) - K_{ik}(\mathbf{x}).$$

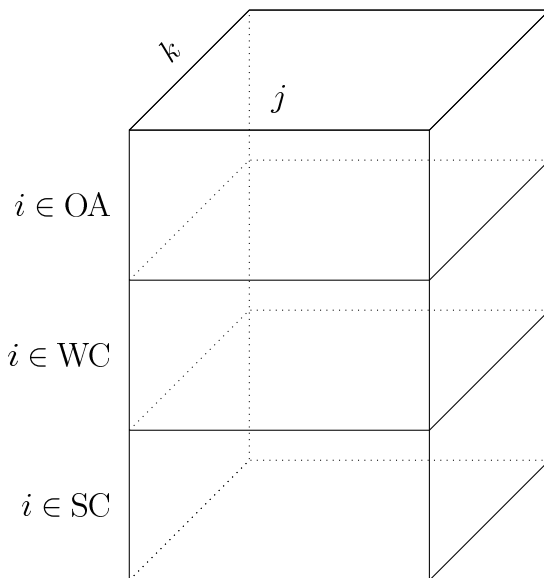


Figure 2: The form of the $n_\varepsilon \times n_\alpha \times n_\alpha$ Hessian array. For the OCO-2 algorithm, the index i ranges over the Oxygen A (OA) band, the Weak CO₂ (WC) band, and the Strong CO₂ (SC) band. For the retrievals in Section 5, $n_\varepsilon \times n_\alpha \times n_\alpha = 2040 \times 50 \times 50$; then the Hessian array has 5,100,000 entries.

Upon adding these two equations, we obtain

$$(\Delta_k + \Delta_j)H_{ijk}(\mathbf{x}) = K_{ij}(\mathbf{x} + \Delta_k \mathbf{e}_k) - K_{ij}(\mathbf{x}) + K_{ik}(\mathbf{x} + \Delta_j \mathbf{e}_j) - K_{ik}(\mathbf{x}),$$

which yields the estimate,

$$\hat{H}_{ijk}(\mathbf{x}) \equiv \tilde{H}_{ijk}(\mathbf{x}) \left(\frac{\Delta_k}{\Delta_j + \Delta_k} \right) + \tilde{H}_{ikj}(\mathbf{x}) \left(\frac{\Delta_j}{\Delta_j + \Delta_k} \right), \quad (20)$$

for $i = 1, \dots, n_\varepsilon$ and $j, k = 1, \dots, n_\alpha$. Notice that the Hessian estimate (20) is a *weighted* combination of the asymmetric estimates given by (19), and it is easy to see that $\hat{H}_{ijk}(\mathbf{x}) \equiv \hat{H}_{ikj}(\mathbf{x})$, which is the required symmetry property in j and k .

Now $\{\hat{H}_{ijk}(\mathbf{x})\}$ given by (20) is an estimate of $\left\{ \frac{\partial \mathbf{K}(\mathbf{x})}{\partial x_k} : k = 1, \dots, n_\alpha \right\}$ which, from the formulas given at the beginning of this subsection, can then be used to obtain estimates of $\frac{\partial \mathbf{G}(x)}{\partial x_k}$ in (16) and $\frac{\partial \mathbf{A}(x)}{\partial x_k}$ in (17). Finally then, the nonlinearity bias can be calculated from (15).

5 Properties of the Retrieval Error of XCO2

Section 2 describes an experiment we conducted to determine the influence of non-linearity on the first two moments of the retrieval error. Our experiment exercises “strong control.” This means that the functional form of the model and every parameter used to simulate forward from state mean \mathbf{x}_α , to obtain true state \mathbf{x} based on state covariance matrix \mathbf{S}_α , and to finally obtain radiances \mathbf{y} based on forward model $\mathbf{F}(\mathbf{x})$ and measurement-error covariance matrix \mathbf{S}_ε , is exactly the same when using the Levenberg-Marquardt algorithm to obtain the retrieval $\hat{\mathbf{x}}$. In this simulation experiment, states and measurements are realistic, and uncertainties are uniquely due to the ill-posed nature of retrieval and not due to using a misspecified forward function in the Levenberg-Marquardt algorithm.

Although \mathbf{x} is known in the experiment, it is not used in the retrieval; its role is to allow us to determine the true retrieval error, $\hat{\mathbf{x}} - \mathbf{x}$. If the simulation is repeated L times, where L is large, then in obvious notation,

$$\{\hat{\mathbf{x}}^{(l)} - \mathbf{x}^{(l)} : l = 1, \dots, L\}, \quad (21)$$

gives the *distribution* of the true retrieval error. We use $L = 700$ below, for which relative Monte Carlo accuracy to the first decimal place is achieved.

Since the forward model $\mathbf{F}(\mathbf{x})$ is nonlinear, the statistical distribution of $\hat{\mathbf{x}} - \mathbf{x}$ is not Gaussian, although its first two moments remain important for inference. It is here where an MCMC approach is a useful research tool, since the full posterior distribution is obtained. However, the excessive time taken to run the MCMC algorithm for a single retrieval (on the order of a day) means it is unimaginable that it might become an operational tool.

The methodology proposed in this article is computationally efficient enough to be made operational on selected retrievals (e.g., when aerosols are present, or for selected retrieval modes). The first moment expresses the bias of the retrieval,

$$\mathbf{bias} \equiv E(\hat{\mathbf{x}} - \mathbf{x}). \quad (22)$$

Inference on \mathbf{x} from the estimate $\hat{\mathbf{x}}$ is based on the second central moment,

$$\mathbf{cov} \equiv \text{cov}(\hat{\mathbf{x}} - \mathbf{x}). \quad (23)$$

Notice that OCO-2’s retrieval algorithm produces instead the second *non-central* moment,

$$\mathbf{MSPE} \equiv E((\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})'), \quad (24)$$

which is appropriate when $\mathbf{bias} = \mathbf{0}$. In general,

$$\mathbf{cov} = \mathbf{MSPE} - (\mathbf{bias})(\mathbf{bias})',$$

and so \mathbf{cov} can be recovered from the moments (22) and (24). In what follows in the experiment, we shall use (22) and (24) to summarize the results; these are sometimes called Figures of Merit (*Cressie and Burden, 2015*).

Apart from choosing 18 different location/seasons to generate a variety of atmospheric states, the two factors that we controlled for were the presence/absence of aerosols in the forward model (and hence in the retrieval), and the presence/absence of the nonlinearity-bias correction. The “response” of the experiment centered on the first two moments of XCO2, the column-averaged CO_2 dry air mole function. The state variable XCO2 is obtained by an appropriately weighted combination of the $n_p = 20$ VMRs of CO_2 at the 20 pressure levels given in ACOS/OCO-2’s forward model (*Crisp et al., 2014*).

In what follows in this section, we write $x \equiv$ true XCO2 value at a given sounding, which in practice is an unknown scalar. Correspondingly, we write $\hat{x}(\mathbf{y}) \equiv$ estimated XCO2 value, featuring the radiances \mathbf{y} measured at the given sounding. The theory of OE assumes that x is a random variable; and from the theory of statistical estimation, $\hat{x}(\mathbf{y})$ is also a random variable with strong statistical dependence on x . Then the retrieval error for estimating XCO2 is,

$$\hat{x}(\mathbf{y}) - x,$$

which is a random variable whose distribution is ultimately determined by the joint distribution of x and \mathbf{y} .

Recall that there are $j = 1, \dots, 18$ diverse soundings and $l = 1, \dots, 700$ simulations for each sounding, which we write as:

$$\{\hat{x}(y_j^{(l)}) - x_j^{(l)} : l = 1, \dots, 700\}; j = 1, \dots, 18.$$

At the $j = 1, \dots, 18$ soundings, we define the bias,

$$\text{bias}_j \equiv \sum_{l=1}^{700} (\hat{x}(y_j^{(l)}) - x_j^{(l)}) / 700, \quad (25)$$

and the root-mean-squared prediction error,

$$\text{rmspe}_j \equiv \left\{ \sum_{l=1}^{700} (\hat{x}(y_j^{(l)}) - x_j^{(l)})^2 / 700 \right\}^{1/2}. \quad (26)$$

These are the true biases and root-mean-squared predictor errors, which we can compute at the 18 location/seasons chosen for this experiment.

As was made clear in Section 4, for the full state vector there is a first-order result based on a linearization of the forward model, and there is a second-order result based on the delta method that gives (13) and (15). For XCO2, the nonlinear

approximations to the bias and MSPE are easily obtained as follows: Now, $XCO2 = \mathbf{w}'\mathbf{x}$ and $\hat{X}CO2 = \mathbf{w}'\hat{\mathbf{x}}$ for known weight vector \mathbf{w} . Then, using obvious notation, the second-order result is:

$$\tilde{\text{bias}}(XCO2) = \mathbf{w}'\tilde{\text{bias}}(\mathbf{x}_\alpha); \quad \tilde{\text{MSPE}}(XCO2) = \mathbf{w}'\tilde{\text{MSPE}}(\mathbf{x}_\alpha)\mathbf{w}. \quad (27)$$

From *Crisp et al.* (2012), the ACOS/OCO-2 algorithm uses

$$\hat{\text{bias}}(XCO2) = 0; \quad \hat{\text{MSPE}}(XCO2) = \mathbf{w}'\hat{\text{MSPE}}(\hat{\mathbf{x}})\mathbf{w}, \quad (28)$$

which is a result of a linear approximation to the forward model. That is, in ACOS/OCO-2's error analysis, a zero bias is obtained, and the MSPE matrix is evaluated at the retrieved state $\hat{\mathbf{x}}$.

For each of the $j = 1, \dots, 18$ soundings, we can make a comparison of the true bias, namely bias_j , with $\hat{\text{bias}}_j (= 0)$ and $\tilde{\text{bias}}_j (\equiv \tilde{\text{bias}}(XCO2_j))$. We can also make a comparison of the true root-mean-squared prediction error, namely rmspe_j , with $\hat{\text{rmspe}}_j (\equiv \{\hat{\text{MSPE}}(XCO2_j)\}^{1/2})$ and $\tilde{\text{rmspe}}_j (\equiv \{\tilde{\text{MSPE}}(XCO2_j)\}^{1/2})$. These comparisons are made through x - y plots, each of which contains 18 x - y pairs corresponding to the 18 soundings.

In Figures 3 and 4, we show only the results for the aerosols case. The clear-sky case (i.e., absence of aerosols) turns out to be less interesting because the nonlinearities in the forward model are much less pronounced. Aerosols confound the retrieval of XCO2 in a strongly nonlinear manner (*O'Dell et al.*, 2012), so it is appropriate that our comparisons focus on this case.

By comparing Fig. 3a with Fig. 3b, one can see the benefit of accounting for the nonlinearity bias when the forward model has strong nonlinearities. Statistical theory indicates that there will be very little difference in the root-mean-squared prediction errors, and this is confirmed when comparing Figures 4a and 4b.

In our experiment, we have chosen a realistic set of atmospheres with aerosols and have found that the true nonlinearity bias given by (25) is no more extreme than ± 0.2 ppm. Similar experiments conducted on earlier versions of the ACOS/OCO-2 algorithm showed much more serious biases on the order of ± 1 ppm and sometimes worse. The forward model that was behind the earlier versions was not effectively capturing the radiative transfer with more state elements (as many as 112) that were highly correlated. As a consequence, the inverse problem was more ill-posed, and the retrieval error was more biased. Based on the approach we have taken to assessing retrieval error, we have been able to quantify an improvement in the ACOS/OCO-2 algorithm.

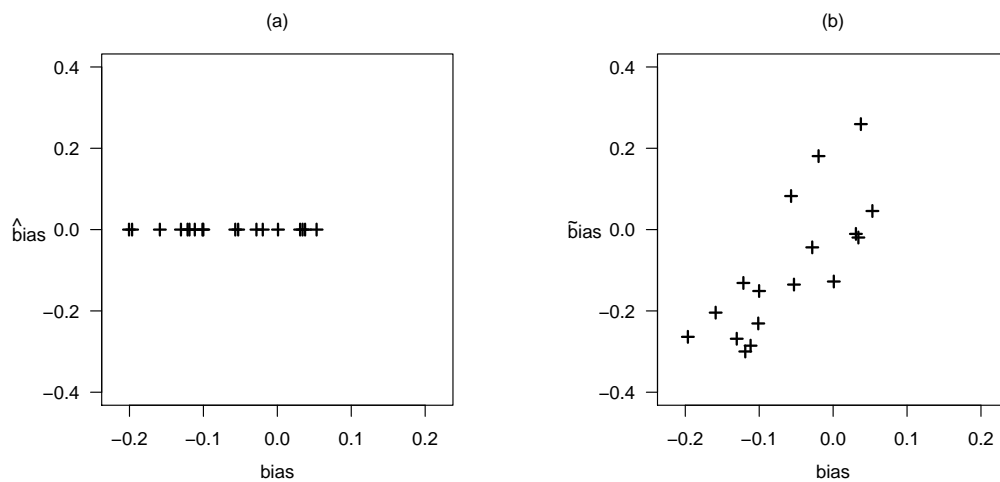


Figure 3: In the presence of aerosols, biases are compared to the true bias. For ACOS/OCO-2's error analysis, which is based on a linear approximation, $\hat{\text{bias}} = 0$ (Figure 3a). For the error analysis based on the delta method given in Section 4, $\tilde{\text{bias}}$ is plotted (Figure 3b). The horizontal axes in both plots show the true biases obtained from the first moment of the estimation error; see (25) and (26).

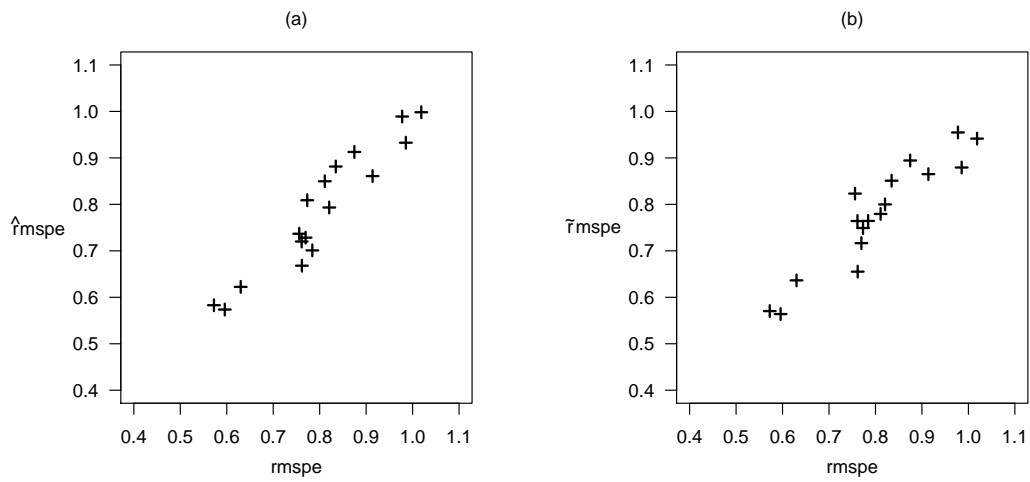


Figure 4: In the presence of aerosols, root-mean-squared prediction errors are given for ACOS/OCO-2's error analysis, $\hat{r}\text{mspe}$ (Figure 4a), and the error analysis based on the delta method given in Section 4 (Figure 4b). The horizontal axes in both plots show the true root-mean-squared prediction errors obtained from the estimation error; see (25) and (26).

6 Discussion and Conclusions

The OE approach to obtaining retrievals in remote sensing incorporates physical knowledge, applied mathematics, and statistics. It is akin to the well known approach in the signal-processing literature called state-space estimation, where both the estimated state and the true state are random, and the error analysis goes beyond a linear approximation (e.g., *Shumway and Stoffer, 2006, Ch. 6*). In this article, we have shown how statistical theory (posterior analysis, delta method, estimating equations) can be used to yield an uncertainty quantification (namely, the first two moments) of the retrieval error, by recognizing the nonlinearity of the forward model and all sources of randomness. The first two moments of the retrieval error are fundamental quantities used for CO_2 flux inversions from remote sensing data.

The nonlinearity in the forward model results in a nonlinearity bias that is estimated using a weighted estimate of the Hessian array; this approximation improves as the signal-to-noise ratio increases (*Cressie and Wang, 2013*). Using the ACOS/OCO-2 algorithm B5.0, XCO₂ retrievals in a controlled experiment (Section 2) exhibit absolute nonlinearity biases up to 0.2 ppm when aerosols are present (Section 5). For clear-sky retrievals, biases are effectively zero.

Of course, there are other sources of bias, some known for physical reasons and some described by regression relationships. Validation of the XCO₂ product is based on data from the Total Carbon Column Observing Network (TCCON) (e.g., *Wunch et al., 2011*). In fact, a TCCON “ground truth” datum is also an *estimate* of XCO₂ derived from a different nonlinear forward model, and it has its own nonlinearity bias. Our results in Section 5 indicate that under controlled conditions the nonlinearity bias of XCO₂ retrievals is within the limits of the error characteristics of TCCON. In actual retrievals, bias can arise from a number of components; our results indicate that the nonlinearity component *by itself* would not be seen by an appropriate comparison to TCCON data.

The specification of a prior distribution (usually in the form of a prior mean vector and a prior covariance matrix) for the unknown state is required for OE. The robustness of the estimates and of the associated error analysis to misspecification of the prior has not been addressed in this article. Our goal has been to work within the assumptions of OE to obtain XCO₂ estimates and their uncertainty quantification that accounts for the nonlinearity in the forward model.

Acknowledgments

The research described in this article was performed for the Orbiting Carbon Observatory Project at the Jet Propulsion Laboratory, California Institute of Tech-

nology, under contract with the National Aeronautics and Space Administration (NASA). Cressie’s and Wang’s research was supported by NASA grant NNH11-ZDA001N-OCO2.

The authors would like to thank A. Braverman, R. Castano, A. Eldering, M. Gunson, D. O’Brien, C. O’Dell, and I. Polonsky for their comments on earlier presentations of this research. We would also like to thank C. Frankenberg for his assistance in setting up the realistic simulation experiment. Finally, our sincere gratitude goes to two anonymous referees and the editor whose comments were extremely helpful in revising our article.

References

- Böesch, H., G.C. Toon, B. Sen, R.A. Washenfelder, P.O. Wennburg, M. Buchwitz, R. de Beek, J.P. Burrows, D. Crisp, M. Christi, B.J. Connor, V. Natraj, and Y.L. Yung (2006). Space-based near-infrared CO₂ measurements: Testing the Orbiting Carbon Observatory retrieval algorithm and validation concept using SCIAMACHY observations over Park Falls, Wisconsin. *Journal of Geophysical Research*, **111**, D23302, doi:10.1029/2006JD007080.
- Böesch, H., D. Baker, B. Connor, D. Crisp, and C. Miller (2011). Global characterization of CO₂ column retrievals from shortwave-infrared satellite observations of the Orbiting Carbon Observatory-2 mission. *Remote Sensing*, **3**, 270–304.
- Bréon, F.-M. and P. Ciais (2010). Spaceborne remote sensing of greenhouse gas concentrations. *Comptes Rendues Geoscience*, **342**, 412–424.
- Connor, B.J., H. Böesch, G. Toon, B. Sen, C. Miller, and D. Crisp (2008). Orbiting Carbon Observatory: Inverse method and prospective error analysis. *Journal of Geophysical Research*, **113**, D05305, doi:10.1029/2006JD008336.
- Cressie, N., and Burden, S. (2015) Figures of merit for simultaneous inference and comparisons in simulation experiments. *Stat*, **4**, 196–211.
- Cressie, N. and R. Wang (2013). Statistical properties of the state obtained by solving a nonlinear multivariate inverse problem. *Applied Stochastic Models in Business and Industry*, **29**, 424-438.
- Crisp, D., B.M. Fisher, C. O’Dell, C. Frankenberg, R. Basilio, H. Böesch, L.R. Brown, R. Castano, B. Connor, N.M. Deutscher, A. Eldering, D. Griffith, M. Gunson, A. Kuze, L. Mandrake, J. McDuffie, J. Messerschmidt, C.E. Miller, I. Morino, V. Natraj, J. Notholt, D. O’Brien, F. Oyafuso, I. Polonsky, J. Robinson, R. Salawitch, V. Sherlock, M. Smyth, H. Suto, T. Taylor, D.R. Thompson, P.O.

- Wennberg, D. Wunch, and Y.L. Yung (2012). The ACOS X_{CO_2} retrieval algorithm, Part 2: Global X_{CO_2} data characterization. *Atmospheric Measurement Techniques*, **5**, 687–707.
- Crisp, D., H. Bösch, L. Brown, R. Castano, M. Christi, B. Connor, A. Eldering, B. Fisher, C. Frankenberg, M. Gunson, J. McDuffie, C.E. Miller, V. Natraj, C. O’Dell, D. O’Brien, I. Polonski, G.B. Osterman, F. Oyafuso, M. Smyth, D. Thompson, G. Doon, and R. Spurr (2014). OCO (Orbiting Carbon Observatory)-2 Level 2 Full Physics Retrieval Algorithm Theoretical Basis, Pasadena, CA.
- Kuze, A., H. Suto, M. Nakajima, and T. Hamazaki (2009). Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring. *Applied Optics*, **48**, 6716–6733.
- Levenberg, K. (1944). A method for the solution of certain nonlinear problems in least squares. *Quarterly Journal of Applied Mathematics*, **2**, 164–168.
- Marquardt, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, **11**, 431–441.
- Meyer, S.L. (1975). *Data Analysis for Scientists and Engineers*. Wiley, New York, NY.
- O’Dell, C.W., B. Connor, H. Böesch, D. O’Brien, C. Frankenberg, R. Castano, M. Christi, D. Crisp, A. Eldering, B. Fisher, M. Gunson, J. McDuffie, C.E. Miller, V. Natraj, F. Oyafuso, I. Polonsky, M. Smyth, T. Taylor, G.C. Toon, P.O. Wennberg, and D. Wunch (2012). The ACOS CO_2 retrieval algorithm – Part 1: Description and validation against synthetic observations. *Atmospheric Measurement Techniques*, **5**, 99–121.
- Rodgers, C.D. (2000). *Inverse Methods for Atmospheric Sounding*. World Scientific Publishing, Singapore.
- Shumway, R. and D.S. Stoffer (2006). *Time Series Analysis and its Applications, with R Examples*, 2nd edn., Springer, New York, NY.
- Tamminen, J. (2004). Validation of nonlinear inverse algorithms with Markov chain Monte Carlo method. *Journal of Geophysical Research-Atmospheres*, **109**, D19303.

- Tamminen, J. and E. Kyrola (2001). Bayesian solution for nonlinear and non-Gaussian inverse problems by Markov chain Monte Carlo method. *Journal of Geophysical Research-Atmospheres*, **106**, 14377–14390.
- Tikhonov, A.N. (1963). On the solution of incorrectly stated problems and a method of regularization. *Doklady Akademii Nauk SSSR*, **151**, 501–504.
- Twomey, S. (1963). On the numerical solution of Fredholm integral equation of the first kind by the inversion of the linear system produced by quadrature. *Journal of the Association of Computing Machinery*, **10**, 97–101.
- Wunch, D., P.O. Wennberg, G.C. Toon, B.J. Connor, B. Fisher, G.B. Osterman, C. Frankenberg, L. Mandrake, C. O’Dell, P. Ahonen, S.C. Biraud, R. Castano, N. Cressie, D. Crisp, N.M. Deutscher, A. Eldering, M.L. Fisher, D.W.T. Griffith, M. Gunson, P. Heikkinen, G. Keppel-Aleks, E. Kyrö, R. Lindenmaier, R. Macatangay, J. Mendonca, J. Messerschmidt, C.E. Miller, I. Morino, J. Notholt, F.A. Oyafuso, M. Rettinger, J. Robinson, C.M. Roehl, R.J. Salawitch, V. Sherlock, K. Strong, R. Sussmann, T. Tanaka, D.R. Thompson, O. Uchino, T. Warneke, and S.C. Wofsy (2011). A method for evaluating bias in global measurements of CO₂ total columns for space. *Atmospheric Chemistry and Physics*, **11**, 12317-12337.