Chapter 6
A Statistical Analysis of the Jacobian in Retrievals of Satellite Data

Noel Cressie

Abstract Remote sensing has become an essential component of the geosciences (the study of Earth and its system components). Remote sensing measurements are almost always energies measured in selected parts of the electro-magnetic spectrum. That is, the geophysical variable of interest is only observed indirectly; a forward model relates the energies to the variable(s) of interest and other elements of the state. The first derivative of that forward model with respect to the state is known as the Jacobian. In this chapter, we review the importance of the Jacobian to inferring the state, and we use it to diagnose which state elements may be difficult to estimate. We develop the Statistical Significance Filter and flag those state elements that consistently fail to get through the filter.

6.1 Introduction

Remote sensing of the environment is a fundamentally important part of humans' quest to understand the Earth system and how the different components interact (e.g., climate, water, carbon). In the future, this knowledge may be critical to our survival. Satellite and aircraft campaigns allow a "bird's-eye view" of large parts of Earth, but not all campaigns are alike. For example, polar-orbiting satellites allow global coverage, passive instruments rely on the sun's reflected light and do not take measurements when there are clouds or when it is night, and programs such as NASA's ASCENDS will measure day or night, anywhere on the orbit track.

In this chapter, a passive instrument on a polar-orbiting satellite, namely Japan's Greenhouse Gases Observing Satellite (GOSAT), will be used as a leading example. However, the idea behind what I shall present is general and could apply to many remote sensing inversion problems involving a non-linear forward model. In such

N. Cressie (✉)
Distinguished Professor, National Institute for Applied Statistics
Research Australia (NIASRA), School of Mathematics and Applied Statistics,
University of Wollongong, Wollongong, Australia
e-mail: ncressie@uow.edu.au

problems, the goal is to infer a hidden state from energies detected by an instrument sensitive to certain known bands of the electro-magnetic spectrum.

Section 6.2 of this chapter gives a statistical framework behind the problem of uncertainty quantification of retrieved states. Section 6.3 calls out the Jacobian matrix as an important component of the retrieval algorithm and defines a unit-free Jacobian for subsequent statistical analysis. That analysis is described in Sect. 6.4, where a Statistical Significance Filter is defined. In Sect. 6.5, this methodology is applied to a number of retrievals taken over Australia, where certain state elements are flagged as being potentially difficult to estimate. The last section, Sect. 6.6, finishes with a discussion of the results obtained.

6.2   A Statistical Framework for Satellite Retrievals

The biases, variances, and mean squared prediction errors of retrievals need to be calculated in the general setting of a nonlinear forward model. The book by Rodgers (2000) has a section on error analysis, but it approaches the problem mostly from a numerical-sensitivity viewpoint. The strongly statistical viewpoint given here calculates the first two moments of a retrieval and the distribution of elements of the associated Jacobian matrix (defined below as $\mathbf{K}$). In the case where relationships are non-linear, the well known "delta method" (based on Taylor-series expansions; e.g., Meyer 1975, Chap. 10) gives *approximate* (to leading orders) biases and mean squared prediction errors of the estimators (Cressie and Wang 2013).

The n_ε-dimensional radiances $\mathbf{Y}$ are related to the n_α-dimensional state $\mathbf{X}$ through a non-linear forward model,

$$\mathbf{Y} = \mathbf{F}(\mathbf{X}) + \boldsymbol{\varepsilon}, \tag{6.1}$$

where the state vector $\mathbf{X}$ includes volume mixing ratios of CO_2 at prespecified geopotential heights, the error vector $\boldsymbol{\varepsilon} \sim \mathrm{Gau}(\mathbf{0}, \mathbf{S}_\varepsilon)$, and $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ are statistically independent. Further, there is an *a priori* assumption that

$$\mathbf{X} = \mathbf{X}_\alpha + \boldsymbol{\alpha}, \tag{6.2}$$

where $\boldsymbol{\alpha} \sim \mathrm{Gau}(\mathbf{0}, \mathbf{S}_\alpha)$. Notice that if there is consistent bias present in the retrieval, this can be accounted for by adding it to $\mathbf{X}_\alpha$, leaving the assumption, $\boldsymbol{\alpha} \sim \mathrm{Gau}(\mathbf{0}, \mathbf{S}_\alpha)$, intact. Define the matrices,

$$\mathbf{K}(\mathbf{x}) \equiv \frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}} \equiv \left(\frac{\partial F_i(\mathbf{x})}{\partial x_j} : i = 1, \ldots, n_\varepsilon; j = 1, \ldots, n_\alpha \right) \tag{6.3}$$

$$\mathbf{G}(\mathbf{x}) \equiv \{\mathbf{S}_\alpha^{-1} + \mathbf{K}(\mathbf{x})'\mathbf{S}_\varepsilon^{-1}\mathbf{K}(\mathbf{x})\}^{-1}\mathbf{K}(\mathbf{x})'\mathbf{S}_\varepsilon^{-1} \tag{6.4}$$

$$\mathbf{A}(\mathbf{x}) \equiv \mathbf{G}(\mathbf{x})\mathbf{K}(\mathbf{x}), \tag{6.5}$$

where $\mathbf{x}$ is any atmospheric state. (Recall that the true state is denoted as $\mathbf{X}$.)

The $n_\varepsilon \times n_\alpha$ matrix $\mathbf{K}(\cdot)$ is called the *Jacobian*. Partial derivatives of $\mathbf{K}(\cdot)$ represent the degree of non-linearity in the forward model. In the case of a *linear* forward model, $\mathbf{K}$ is constant, and any partial derivatives of it are zero.

An estimate of $\mathbf{X}$, sometimes called a *retrieval*, is often obtained by choosing an $\hat{\mathbf{X}}$ that allows $\mathbf{F}(\hat{\mathbf{X}})$ to be "close to" $\mathbf{Y}$, subject to smoothness conditions on $\hat{\mathbf{X}}$. This regularisation is usually defined as follows: Minimise

$$(\mathbf{Y} - \mathbf{F}(\mathbf{X}))'\mathbf{S}_\varepsilon^{-1}(\mathbf{Y} - \mathbf{F}(\mathbf{X})) + (\mathbf{X} - \mathbf{X}_\alpha)'\mathbf{S}_\alpha^{-1}(\mathbf{X} - \mathbf{X}_\alpha) \qquad (6.6)$$

with respect to $\mathbf{X}$, which results in the retrieval $\hat{\mathbf{X}}$.

The $n_\alpha \times n_\varepsilon$ matrix $\mathbf{G}(\cdot)$ represents a type of *"gain" matrix* in the relationship between retrieval $\hat{\mathbf{X}}$ and data $\mathbf{Y}$; that is,

$$\hat{\mathbf{X}} = \mathbf{X}_\alpha + \mathbf{G}(\hat{\mathbf{X}})(\mathbf{Y} - \mathbf{F}(\mathbf{X}_\alpha) - \mathbf{K}(\hat{\mathbf{X}})\mathbf{X}_\alpha) + \text{ "remainder".}$$

In the linear case, $\mathbf{G}$ is constant and the "remainder" term is zero.

The $n_\alpha \times n_\alpha$ matrix $\mathbf{A}(\cdot)$ yields the *averaging kernel matrix* in the relation between retrieval and true state; that is,

$$\hat{\mathbf{X}} = \mathbf{X}_\alpha + \mathbf{A}(\hat{\mathbf{X}})(\mathbf{X} - \mathbf{X}_\alpha) + \text{ "remainder".}$$

In the linear case, $\mathbf{A}$ is constant, the "remainder" term is $\mathbf{G}\varepsilon$, and recall that ε is independent of $\mathbf{X}$.

In this section, I discuss the bias vector and the mean-squared-prediction-error (MSPE) matrix of the retrieval, $\hat{\mathbf{X}}$. The bias vector is defined as:

$$E(\hat{\mathbf{X}} - \mathbf{X}) = E(\hat{\mathbf{X}}) - E(\mathbf{X}) = E(\hat{\mathbf{X}}) - \mathbf{X}_\alpha,$$

where recall that $\mathbf{X}_\alpha$ is the prior mean of the state vector $\mathbf{X}$.

The MSPE matrix is defined as:

$$E((\hat{\mathbf{X}} - \mathbf{X})(\hat{\mathbf{X}} - \mathbf{X})') = \text{var}(\hat{\mathbf{X}} - \mathbf{X}) + (E(\hat{\mathbf{X}}) - \mathbf{X}_\alpha)(E(\hat{\mathbf{X}}) - \mathbf{X}_\alpha)',$$

where $\text{var}(\hat{\mathbf{X}} - \mathbf{X})$ is the covariance matrix of the retrieval error, $\hat{\mathbf{X}} - \mathbf{X}$. The MSPE matrix can be a more appropriate statistical measure of uncertainty than the covariance matrix of retrieval error when there is bias present. When the bias is zero, the two measures of uncertainty are the same.

When the forward model is *linear*, it is easily seen (e.g., Rodgers 2000) that the bias vector,

$$E(\hat{\mathbf{X}} - \mathbf{X}) = \mathbf{0}. \qquad (6.7)$$

That is, in the linear case, $\hat{\mathbf{X}}$ is *unbiased*. Further, in the linear case, the MSPE matrix can be derived *exactly* and written in a number of equivalent ways. From Connor et al. (2008), Cressie and Wang (2013),

$$E((\hat{\mathbf{X}} - \mathbf{X})(\hat{\mathbf{X}} - \mathbf{X})') = E(\text{var}(\mathbf{X}|\mathbf{Y})) \equiv \hat{\mathbf{S}}, \qquad (6.8)$$

where the MSPE matrix is given by

$$\hat{\mathbf{S}} = \{\mathbf{S}_\alpha^{-1} + \mathbf{K}'\mathbf{S}_\varepsilon^{-1}\mathbf{K}\}^{-1} = (\mathbf{A} - \mathbf{I})\mathbf{S}_\alpha(\mathbf{A} - \mathbf{I})' + \mathbf{G}\mathbf{S}_\varepsilon\mathbf{G}' . \qquad (6.9)$$

When the forward model is *nonlinear*, the bias of $\hat{\mathbf{X}}$ is *nonzero*, and the equalities in (6.9) are no longer true. However, from the "delta method," Cressie et al. (2016) show that (6.7) and (6.9) hold, *to leading order*. In what follows, a leading-order analysis is carried out. This amounts to assuming the forward model to be locally linear, which is a weaker assumption than assuming global linearity, namely $\mathbf{Y} = \mathbf{c} + \mathbf{K}\mathbf{X} + \boldsymbol{\varepsilon}$, across the whole state space defined by all possible values of $\mathbf{X}$.

The locally linear forward model is derived using a Taylor-series expansion:

$$\begin{aligned}
\mathbf{Y} &= \mathbf{F}(\mathbf{X}) + \boldsymbol{\varepsilon} \\
&= \mathbf{F}(\mathbf{X}_0) + \left.\frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{X}_0} \times (\mathbf{X} - \mathbf{X}_0) + \lambda \\
&\equiv \mathbf{c}(\mathbf{X}_0) + \mathbf{K}(\mathbf{X}_0)\mathbf{X} + \lambda ,
\end{aligned}$$

where λ models the lack of fit of the local linear model (about the linearisation point $\mathbf{x} = \mathbf{X}_0$) to $\mathbf{F}(\mathbf{X})$. The linearisation point $\mathbf{X}_0$ is often chosen to be the prior mean $\mathbf{X}_\alpha$, but I want to emphasise here that it need not be.

6.3 The Jacobian Matrix and its Unit-Free Version

The Jacobian matrix is the first derivative of the n_ε-dimensional forward function vector, $\mathbf{F}(\mathbf{x})$, with respect to the n_α-dimensional state $\mathbf{x}$. From the definition given in (6.3), it is an $n_\varepsilon \times n_\alpha$ matrix. Write the matrix as (K_{ij}), and note that the units of K_{ij} are radiance (energy) per unit of state-space element j.

Define the vectors,

$$\begin{aligned}
(\sigma_{\varepsilon,1}^2, \dots, \sigma_{\varepsilon,n_\varepsilon}^2)' &\equiv \text{diag}(\mathbf{S}_\varepsilon) \\
(\sigma_{\alpha,1}^2, \dots, \sigma_{\alpha,n_\alpha}^2)' &\equiv \text{diag}(\mathbf{S}_\alpha) ,
\end{aligned}$$

where $\text{diag}(\cdot)$ is a matrix operator that extracts a vector made up of the matrix's diagonal elements. Then the *unit-free Jacobian* is defined as follows:

$$\phi_{ij} \equiv K_{ij}\sigma_{\alpha,j}/\sigma_{\varepsilon,i}\,; \ i = 1, \ldots, n_\varepsilon, j = 1, \ldots, n_\alpha\,. \tag{6.10}$$

During the retrieval, the most difficult and time-consuming part is to minimise (6.6); for example, using a Levenberg-Marquardt algorithm requires evaluation of the Jacobian matrix at each iteration of the minimisation. Let $\hat{K}_{ij}$ be a generic Jacobian element used during the retrieval. Then define the corresponding unit-free version as,

$$\hat{\phi}_{ij} \equiv \hat{K}_{ij}\sigma_{\alpha,j}/\sigma_{\varepsilon,i}\,, \tag{6.11}$$

and denote $\hat{\mathbf{\Phi}} \equiv (\hat{\phi}_{ij})$ as the $n_\varepsilon \times n_\alpha$ *unit-free Jacobian matrix.*

For satellite retrievals, the data vector $\mathbf{Y}$ can often be partitioned as

$$\mathbf{Y} = (\mathbf{Y}'_1, \ldots, \mathbf{Y}'_K)'\,,$$

where

$$\mathbf{Y}_k \equiv (Y_i : i \in \text{band}_k)'\,, \tag{6.12}$$

and $\text{band}_1, \ldots, \text{band}_K$ are mutually exclusive index sets that represent a grouping of radiances according to which bands of the electro-magnetic spectrum they belong. For example, Japan's GOSAT and NASA's Orbiting Carbon Observatory-2 (OCO-2) instruments have $K = 3$ bands, corresponding to the oxygen A band (OA), the weak carbon dioxide band (WC), and the strong carbon dioxide band (SC); our analysis in Sect. 6.5 uses data from GOSAT's three bands. Another example is from NASA's Atmospheric Infrared Sounder (AIRS) instrument flying on the Aqua satellite, which has $K = 4$ bands, corresponding to four geophysical variables, namely temperature, water vapour, ozone, and carbon dioxide.

In what follows, we abbreviate "band$_k$" to "b_k." Because the unit-free Jacobian has elements that are potentially comparable, we can partition it and analyse it in comparable ways. Recall that the index j corresponds to a given element of the state vector, for example, a water-vapour scale factor or a near-surface carbon-dioxide volume mixing ratio. Then fix the state element j, and consider the behaviour of the jth column as row i varies within individual bands. That is, for a fixed j, consider

$$\{\hat{\phi}_{ij} : i \in b_k\} \tag{6.13}$$

to be a random sample from a distribution indexed by k, for bands $k = 1, \ldots, K$.

Consequently, instead of thinking about $n_\varepsilon \cdot n_\alpha$ entries in the Jacobian, attention turns to $n_\alpha \cdot K$ *distributions*. For example, for the retrievals from GOSAT data that are being considered here, $n_\varepsilon = 2240$, $n_\alpha = 112$, and $K = 3$. Hence, the pair (j, k) indexes one of 336 possible distributions, whose mean, μ_{jk}, is of primary interest. For j a fixed element of the state vector, if $\mu_{j1} = \mu_{j2} = \cdots = \mu_{jK} = 0$, then that element is poorly determined by the data alone; see Sect. 6.4. This is a flag that says the (prior)

mean and precision of the jth state element need to be specified very carefully in the second term of (6.6) in order to obtain an acceptably precise retrieval $\hat{X}_j$.

6.4 Statistical Significance Filter

To leading order, the forward model (6.1) can be written as,

$$\mathbf{Y} = \mathbf{c} + \mathbf{K}_1 X_1 + \cdots + \mathbf{K}_{n_\alpha} X_{n_\alpha} + \boldsymbol{\varepsilon}, \tag{6.14}$$

which is a multiple-regression model with known, typically different, intercepts given by the elements of $\mathbf{c}$; known covariates $\mathbf{K}_1, \ldots, \mathbf{K}_{n_\alpha}$ (the n_α columns of $\mathbf{K}$); and unknown regression coefficients $X_1, \ldots, X_{n_\alpha}$. Clearly, if $\mathbf{K}_j$ is zero, then X_j will not be estimable. Further, if for a given j, $\{|K_{ij}| : i = 1, \ldots, n_\varepsilon\}$ are uniformly "small," then the uncertainty associated with the estimate of X_j will be large.

In the previous section, we noted that for remote sensing retrievals, the n_ε elements in $\mathbf{Y}$ can be partitioned into K bands, $\mathbf{Y}_1, \ldots, \mathbf{Y}_K$. Then write (6.14) equivalently as K equations. In obvious notation that respects the partitioning,

$$\mathbf{Y}_k = \mathbf{c}_k + \mathbf{K}_{1k} X_1 + \cdots + \mathbf{K}_{n_\alpha k} X_{n_\alpha} + \boldsymbol{\varepsilon}_k \, ; \, k = 1, \ldots, K, \tag{6.15}$$

where $\{\mathbf{K}_{jk} : j = 1, \ldots, n_\alpha\}$ are the n_α vectors corresponding to the kth band.

Clearly, if $\mathbf{K}_{jk} = \mathbf{0}$, then its unit-free version, $\boldsymbol{\Phi}_{jk}$, is also $\mathbf{0}$. Hence, the problem of whether X_j is poorly determined in the forward model (6.1) can be addressed in a statistical manner by considering the retrieval's unit-free Jacobian entries $\{\hat{\phi}_{ij} : i = 1, \ldots, n_\varepsilon\}$ as K arrays of random variables, $\{\hat{\phi}_{ij} : i \in b_k\}$, for $k = 1, \ldots, K$. If, for a fixed j, the means $\mu_{j1}, \ldots, \mu_{jk}$ of these K arrays are all zero, then X_j will be difficult to estimate.

6.4.1 Hypothesis Tests

Consider (6.13) and make the following assumption: For a given retrieval, a given state element j, and a given band k,

$$\{\hat{\phi}_{ij} : i \in b_k\} \overset{iid}{\sim} \text{Dist}(\mu_{jk}),$$

where "iid" denotes "independent and identically distributed," and "Dist(μ)" denotes a probability distribution with mean μ. For this retrieval, the idea is to flag those state elements and bands for which the null hypothesis, $H_{0,jk} : \mu_{jk} = 0$, is not rejected. In particular, failure to reject the composite hypothesis,

$$H_{0,j} : \mu_{j1} = \mu_{j2} = \cdots = \mu_{jK} = 0 , \tag{6.16}$$

implies that the jth state element will be difficult to estimate in the given retrieval.

Since the elements of $\{\hat{\phi}_{ij} : i \in b_k\}$ are considered to be a sample from a distribution with mean μ_{jk}, I shall construct a test statistic from these unit-free Jacobian values. A considerable amount of exploratory data analysis showed the common distributional assumption within the partitioned arrays to be largely correct, with occasional gross outliers that would challenge many statistical testing procedures. Those were controlled by transforming each $\hat{\phi}_{ij}$ to $|\hat{\phi}_{ij}|^{1/2}$, and the robust test statistic,

$$\tilde{\phi}_{jk} \equiv \mathrm{med}\{|\hat{\phi}_{ij}|^{1/2} : i \in b_k\} , \tag{6.17}$$

was used to test $H_{0,jk} : \mu_{jk} = 0$. The composite hypothesis test $\{H_{0,j} : j = 1, \ldots, n_\alpha\}$, where $H_{0,j}$ is given by (6.16), is then carried out using a Bonferroni adjustment (Sect. 6.4.3).

6.4.2   Distribution Theory for the Robust Test Statistic

Consider generic iid random variables $W_1, \ldots, W_m$ distributed according to a Gaussian distribution with mean μ_W and variance σ_W^2, which is written as $\mathrm{Gau}(\mu_W, \sigma_W^2)$. To test

$$H_0 : \mu_W = 0 \text{ versus } H_1 : \mu_W \neq 0 , \tag{6.18}$$

consider the robust test statistic,

$$\tilde{X} \equiv \mathrm{med}\{|W_i|^{1/2} : i = 1, \ldots, m\}. \tag{6.19}$$

I now obtain distribution theory for $\tilde{X}$ under the null hypothesis in order to carry out a significance test.

If $Y \sim \mathrm{Gau}(0,1)$, then $E(|Y|^{1/2}) = 0.82216$ and $\mathrm{var}(|Y|^{1/2}) = 0.12192$, which was derived by Cressie and Hawkins (1980). Then under $H_0 : \mu_W = 0$, $|W_i|^{1/2} \overset{.}{\sim}$ $\mathrm{Gau}(0.82216 \cdot \sigma_W^{1/2}, 0.12192 \cdot \sigma_W)$, where "$\overset{.}{\sim}$" denotes "is approximately distributed as," and the approximation is established by Cressie and Hawkins (1980). Now the distribution of the median $\tilde{X}$ from a random sample $X_1, \ldots, X_m$ of Gaussian random variables can be approximated as Gaussian with mean $E(\tilde{X}) = E(X_1)$, and variance $\mathrm{var}(\tilde{X}) = \pi \mathrm{var}(X_1)/2m$. If all these results are combined, then under the null hypothesis H_0 in (6.18),

$$\tilde{X} \overset{.}{\sim} \mathrm{Gau}(0.82216 \cdot \sigma_W^{1/2}, 0.12192 \cdot \pi \sigma_W/2m) .$$

Clearly, the alternative hypothesis H_1 in (6.18) is accepted if the test statistic $\tilde{X}$ is large. At significance level α, H_1 is accepted if

$$\tilde{X} > 0.82216 \cdot \sigma_W^{1/2} + \Phi^{-1}(1 - \alpha)(0.12192 \cdot \pi\sigma_W/2m)^{1/2}, \qquad (6.20)$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of a Gau$(0, 1)$ random variable. In practice, an estimate of σ_W will be needed.

Continuing with the same approach as above, an asymptotically unbiased, robust estimator of σ_W is used. Now, $\sigma_W = \text{var}(|W_i|^{1/2})/0.12192$, and hence $\text{var}(|W_i|^{1/2})$ can be estimated using the median absolute deviation (MAD):

$$\text{MAD} \equiv \text{med}\{||W_i|^{1/2} - \tilde{X}| : i = 1, \dots, m\}.$$

Then an asymptotically unbiased estimator of $\text{var}(|W_i|^{1/2})$ is

$$\hat{\text{var}}(|W_i|^{1/2}) = (1.4826 \cdot \text{MAD})^2,$$

from which the estimator

$$\tilde{\sigma}_W \equiv (1.4826 \cdot \text{MAD})^2/0.12192 \qquad (6.21)$$

is obtained and substituted into (6.20).

My approach to constructing this robust statistic to test whether a mean is zero, using data that may contain large, unpredictable outliers, is somewhat unusual, but it is statistically advantageous. First, the data $\{W_1, \dots, W_m\}$ are made resistant by transforming to the square-root scale where variability is dampened. Then the transformed data $\{|W_1|^{1/2}, \dots, |W_m|^{1/2}\}$ are used to define a robust test statistic, given here by the median; see (6.19). Finally, the null distribution is derived, resulting in a critical region given by (6.20) with the robust estimator (6.21) substituted in. In the next subsection, the distribution theory derived in this subsection is used in the context of multiple hypothesis testing, resulting in the *Statistical Significance Filter*.

6.4.3 Multiple Hypothesis Tests Define the Statistical Significance Filter

The elements of the unit-free Jacobian are considered as replicates within bands, which results in n_α (number of state elements) times K (number of bands) hypothesis tests of $\{H_{0jk} : \mu_{jk} = 0, \text{ for } j = 1, \dots, n_\alpha \text{ and } k = 1, \dots, K\}$. To test H_{0j} given by (6.16), jointly for $j = 1, \dots, n_\alpha$, I use a family-wise error rate of 1% and conservative Bonferroni adjustments to obtain a level of significance, $\alpha = .01/(n_\alpha \cdot K)$, that is used in each individual hypothesis test of the null hypotheses, $\{H_{0jk}\}$.

The *Statistical Significance Filter* only allows estimates $\{\tilde{\phi}_{jk}\}$ to get through the filter if $\{H_{0jk}\}$ are rejected, respectively. A given state element, j say, is flagged as problematic in a given retrieval if, simultaneously, the hypotheses $H_{0j1}, \dots, H_{0jK}$ are not rejected. If it consistently happens that under similar (or different) geophysical conditions, the jth element's bands fail to get through the Statistical Significance Filter, that element X_j is flagged as being weakly sensitive to the radiance measurements **Y**. Hence, estimation of X_j would be difficult if a very disperse prior distribution in (6.2) were chosen for it.

In the next section, I apply the Statistical Significance Filter to 30 retrievals from Japan's GOSAT instrument that measures atmospheric carbon dioxide, here over central Australia.

6.5   ACOS Retrievals of the Atmospheric State from Japan's GOSAT Satellite

Shown in Fig. 6.1 are 30 locations of retrievals from Japan's GOSAT satellite, where the ACOS (Atmospheric CO2 Observations from Space) retrieval algorithm was used. Specifically, ACOS Version B2.8 was used here, for which $n_\alpha = 112$ state elements were retrieved from $n_\varepsilon = 2240$ radiances spread roughly equally between the $K = 3$ bands, namely the OA band, the WC band, and the SC band; see Sect. 6.3. The soundings are over an arid part of Australia with uniformly high albedo, during the period from 5 June 2009–26 July 2009 (Source: CIRA, Colorado State University). The methodology and inference is illustrated on the retrieval at one of those locations, hereafter referred to as Location 1. Results from the other 29 retrievals are summarised at the end of this section.

A number of the state elements in B2.8 are functions of geopotential height, here labelled as 1 (top of atmosphere) down to 20 (surface of Earth). Figure 6.2 shows unit-free ice-cloud Jacobian values in a column of the atmosphere for Location 1; only those values that got through the Statistical Significance Filter are shown. It can be seen that for the ice-cloud variable, Jacobian values in the OA band are not statistically significant at higher altitudes in the atmospheric column, and hence they are potentially difficult to estimate. Figure 6.3 shows that the Statistical Significance Filter applied to water vapour (H_2O) in the column results in a similar set of plots. Contrast these to Fig. 6.4, which is for the all-important carbon-dioxide (CO_2) variable; only values in the SC band get through the Statistical Significance Filter.

The analysis of the retrieval for Location 1 yields non-significant Jacobian entries (i.e., forward-model derivatives near zero) *in all three bands* for the following state elements:

1, 2, 3 CO_2 values near the top of the atmosphere
21   H_2O scale factor
23   Temperature offset
105, 107, 109   Albedo slope for the three bands
110, 111, 112   Spectral dispersion offset for the three bands

This behavior is visualised in Fig. 6.5; there, a light (green) stripe in a given band for a given state element indicates that the corresponding mean is not significantly different from zero. A light stripe in every band for the given state element indicates that extra care will be needed when specifying a prior for that element. Each of the 11 elements listed above have a light stripe in every band.

The analysis was carried out on all 30 retrievals, and eight elements of the 112-dimensional state vector emerged as always having non-significant Jacobian values in all three bands for all 30 retrievals. They were:
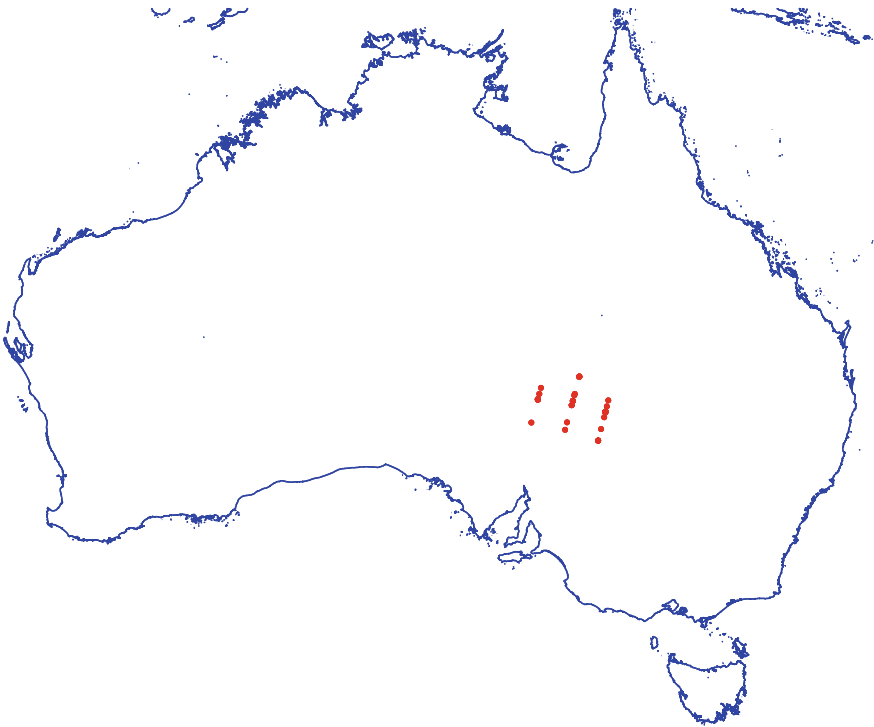
Fig. 6.1 Locations of 30 retrievals from GOSAT using the ACOS Version B2.8 retrieval: 5 June 2009–26 July 2009
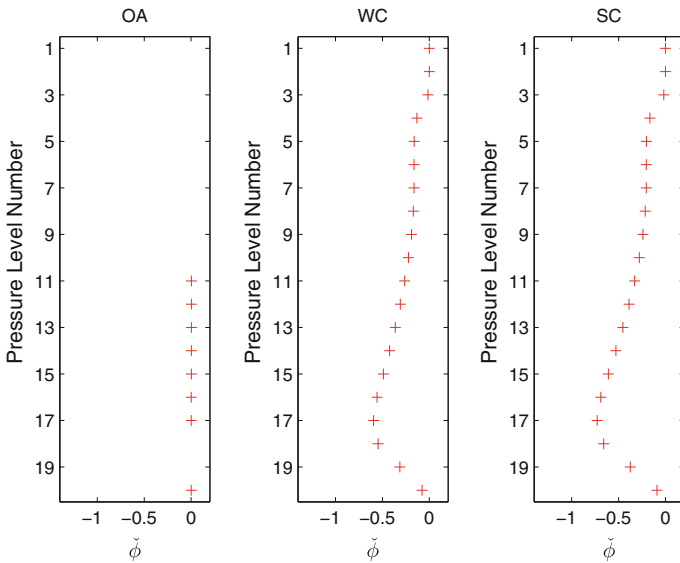
Fig. 6.2 Unit-free Jacobian ice-cloud values that pass through the statistical significance filter in the OA, WC, and SC bands. Values that did not pass through the filter are not plotted. Location 1 (out of 30 locations)
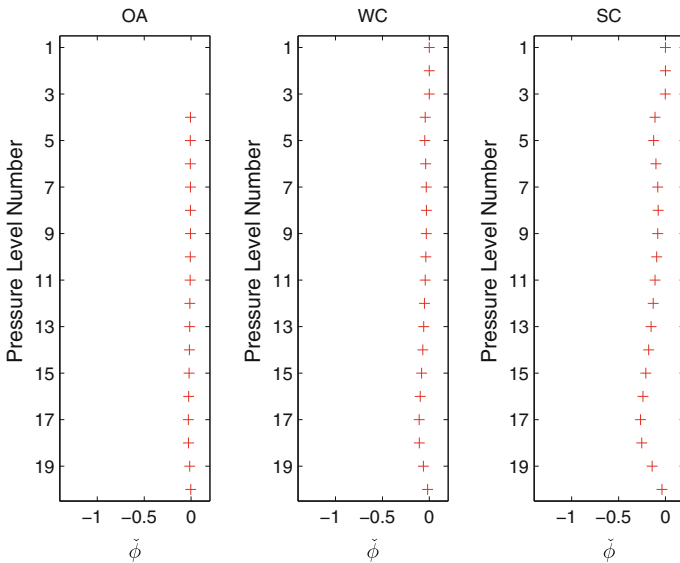
Fig. 6.3 Unit-free Jacobian H_2O values that pass through the statistical significance filter in the OA, WC, and SC bands. Values that did not pass through the filter are not plotted. Location 1 (out of 30 locations)
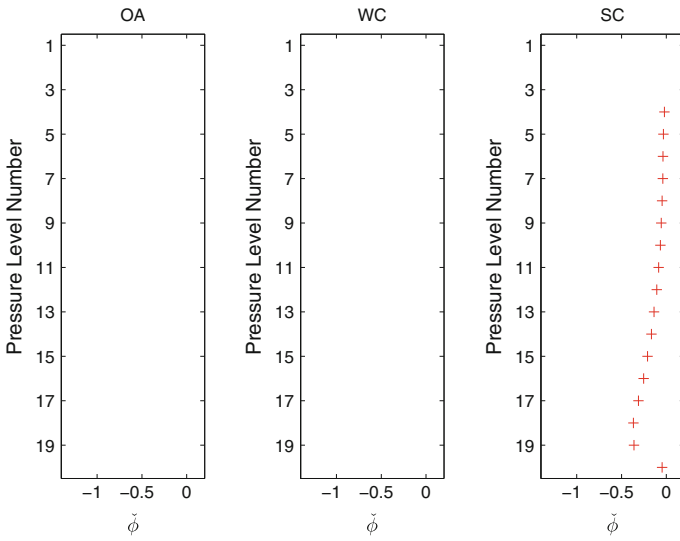
Fig. 6.4 Unit-free Jacobian CO_2 values that pass through the statistical significance filter in the OA, WC, and SC bands. Values that did not through pass the filter are not plotted. Location 1 (out of 30 locations)
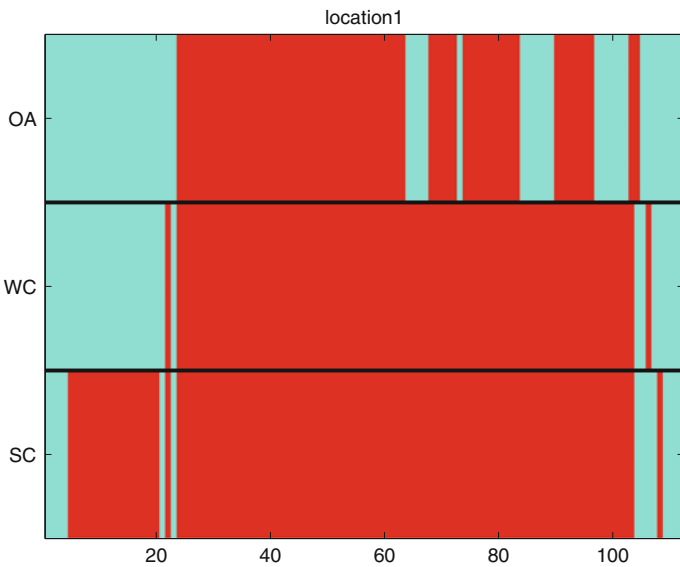
Fig. 6.5 A graphic showing which of the 112 elements of the state vector (horizontal axis) pass through the statistical significance filter (dark, red colour) and which do not (light, green colour), for "band" = OA, WC, and SC. Location 1 (out of 30 locations)

21	H_2O scale factor
23	Temperature offset
105, 107, 109	Albedo slope for the three bands
110, 111, 112	Spectral dispersion offset for the three bands

The results indicate a lack of sensitivity of these eight elements in the forward equation **F** given in (6.1), for the dry, bright, flat-terrain conditions found over central Australia. Different land surfaces and atmospheric states would almost certainly result in different elements being identified.

6.6   Discussion

The Jacobian matrix **K** is the first derivative of a vector-valued function **F(x)** of a state vector **x**. Consistently small elements in the jth column of **K** indicate that the jth element will be difficult to estimate (predict) based on data, **Y**, alone.

If prior information, as well as the data, is used to predict the state vector, this research indicates that acceptable precision for estimating this jth element may require the prior variance to be tightly constrained. For example, the element that is the H_2O scale factor is tightly constrained physically in the prior. Thus, a retrieval of that element may cause no problem, even though its column in **K** fails to get through the Statistical Significance Filter. Regarding the 20 CO_2 elements that make up the CO_2 profile in the atmospheric column, the retrievals analysed here show the importance of the strong CO_2 band (SC) to its estimation. The best result would be if all $20 \cdot 3 = 60$ hypothesis tests were rejected; at Location 1, only 17, all in the SC band, were rejected (Fig. 6.4).

Current versions of ACOS-like retrievals have between 40–50 state elements. The research presented here, on the statistical properties of the Jacobian, would allow a comparison of different versions through the behaviour of their unit-free Jacobian values. Common to all of these versions is 20 CO_2 elements, and the respective estimates of the means in each of the three bands (OA, WC, SC) can be compared across versions.

References

Connor BJ, Boesch H, Toon G, Sen B, Miller C, Crisp D (2008) Orbiting Carbon Observatory: inverse method and prospective error analysis. J Geophys Res Atmos 113:D055305

Cressie N, Hawkins DM (1980) Robust estimation of the variogram: I. J Int Assoc Math Geol 12:115–125

Cressie N, Wang R (2013) Statistical properties of the state obtained by solving a nonlinear multivariate inverse problem. Appl Stoch Models Bus Ind 29:424–438

Cressie N, Wang R, Smyth M, Miller CE (2016) Statistical bias and variance for the regularized inverse problem: Application to space-based atmospheric CO_2 retrievals. J Geophys Res Atmos 121:5526–5537. https://doi.org/10.1002/2015JDO24353

Meyer SL (1975) Data analysis for scientists and engineers. Wiley, New York

Rodgers CD (2000) Inverse methods for atmospheric sounding. World Scientific Publishing, Singapore